ФИЛИАЛ МОСКОВСКОГО ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА
имени М.В. ЛОМОНОСОВА в городе ТАШКЕНТ

ТРУДЫ

республиканской научно-практической конференции

СТАТИСТИКА

и ее применения



17-18 октября 2019 года
Филиал Московского Государственного Университета
имени М.В. Ломоносова в городе Ташкент

M.V.LOMONOSOV MOSCOW STATE UNIVERSITY,
TASHKENT BRANCH

PROCEEDINGS

of republican scientific-applied conference

STATISTICS

and its applications



17-18 October, 2019
M.V.Lomonosov Moscow State University, Tashkent Branch

УДК 519.2

Труды республиканской научно-практической конференции "СТАТИСТИКА и ее применения-2019". Под редакцией профессора А.А.Абдушукурова - Ташкент, Филиал МГУ, 2019, 417 стр.

Организационный комитет:

Кудрявцев В.Б. - Председатель, руководитель филиала МГУ, академик;

Сайдаматов Э.М. - Заместитель председателя, исполнительный директор филиала МГУ, доцент;

Маманазаров А.Б. - Заместитель председателя, заместитель руководителя филиала МГУ, доцент;

Дергач П.С. - Заместитель председателя, 1-й заместитель руководителя филиала МГУ;

Строгалов А.С. - Заместитель председателя, доцент кафедры МаТИС МГУ имени М.В.Ломоносова;

Абдушукуров А.А. - Профессор кафедры ПМиИ филиала МГУ (Организатор);

Программный комитет:

Каршиев Т.О. - Заместитель директора по учебной и научной работе, доцент;

Базаров Т.Ю. - Заместитель руководителя филиала МГУ, профессор;

Мирзаев Ш.Ш. - Заместитель директора по работе с молодҷжью, доцент;

Рыжов А.П. - Заместитель декана по научной работе факультета ПМиИ, профессор;

Носов М.В. - Заместитель декана по учебной работе факультета ПМиИ, старший научный сотрудник;

Аюпов Ш.А. - Директор Института Математики АН РУз, академик;

Алимов Ш.А. - Профессор кафедры "Дифференциальные уравнения и математическая физикањ НУУз, академик;

Садуллаев А.С. - Заведующий кафедрой "Математический анализ"НУУз, академик;

Чилин В.И. - Профессор кафедры "Алгебра и функциональный анализ НУУз;

Зупаров Т.М. - Профессор кафедры "Теория вероятностей и математическая статистика НУУз;

Гафуров М.У. - Профессор кафедры ПМиИ;

Бешимов Р.Б. - Заведующий кафедрой "Геометрия и топология НУУз;

Алоев Р.Д. - Профессор кафедры "Вычислительная математика и информационные системы НУУз;

Турсунов Г.Т. - Доцент кафедры "Теория вероятностей и математическая статистика НУУз;

Сагидуллаев К.С. - Заведующий кафедрой "Естественные науки Технический институт Еджу;

Нагорный А.М. - Доцент кафедры ПМиИ;

Альхамов Р.Р. - Доцент кафедры ПМиИ;

Ибрагимова Д.Э. - Доцент кафедры ПМиИ;

Исраилова Д.М. - Доцент кафедры ПМиИ;

Шералиев Ш.Н. - И.о. зам.декана факультета ПМиИ;

Нурмухамедова Н.С. - И.о.доцента кафедры "Эконометрика и экономическое моделирование НУУз;

Какаджанова Л.Р. - PhD докторант кафедры "Теория вероятностей и математическая статистика НУУз;

Секции конференции

1. Математическая статистика.
2. Вероятность и смежные вопросы.
3. Дискретная и непрерывная математика;
4. Математические - статистические методы в экономике, медицине, естествознании и методике преподавания.
5. Фундаментальная информатика и информационные технологии.

ISBN 978-9943-305-86-1

Спонсоры:
Конференцию поддержал Республиканский фонд фундаментальных исследований (Грант Ф4-40). Конференцию поддержала Страховая компания ООО **"APEX Insurance"**.

UDK 519.2

Proceedings of of republican scientific-applied conference "STATISTICS and its applications-2019". Editor Professor A.A.Abdushukurov - Tashkent, NUUz, 2019, 417 pages.

Organizing Committee:

Kudryavtsev V.B. - Chairman, Head of the MSU Branch, academician;

Saydamatov E.M. - Vice-Chairman, Executive Director of the MSU Branch, Associate professor;

Mamanazarov A.B. - Vice-Chairman, Deputy Head of the MSU Branch, Associate professor;

Dergach P.S. - Vice-Chairman, Deputy Head of the MSU Branch, Associate professor;

Strogalov A.S. - Vice-Chairman, Associate professor, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University

Abdushukurov A.A. - Professor of Dpt. Appl. Math. & Informatics, MSU Branch (Organizer);

Program Committee:

Karshiyev T.O. - Deputy Director of Academic and Scientific Work of faculty of Appl. Math. & Informatics, Associate professor;

Ryjov A.P. - Vice-dean of Scientific Work of faculty of Appl. Math. & Informatics, professor;

Nosov M.V. - Vice-dean of Academic Work of faculty of Appl. Math. & Informatics, Senior researcher;

Ayupov Sh.A. - Director of Institute of Mathematics AS RUz, academician;

Alimov Sh.A. - Professor of Dpt. Dif. Eq. & Math. Physics, NUUz, academician;

Sadullayev A.S. - Head of Dpt. Math Analysis, NUUz, academician;

Chilin V.I. - Professor of Dpt. Algebra & Functional Analysis, NUUz;

Zuparov T.M. - Professor of Dpt. Prob. Theory & Math. Stat., NUUz;

Gafurov M.U. - Professor of Dpt. Appl. Math. & Informatics, MSU Branch;

Beshimov R.B. - Head of Dpt. Geometry & Topology, NUUz;

Aloev R.D. - Professor of Dpt. Comp. Math. & Inform. Syst., NUUz;

Tursunov G.T. - Associate professor of Dpt. Prob. Theory & Math. Stat., NUUz;

Sagidullayev K.S. - Head of Dpt. Nat. Sciences, Technical Institute Yodju;

Nagorniy A.M. - Associate professor of Dpt. Appl. Math. & Informatics, MSU Branch;

Alkhamov R.R. - Associate professor of Dpt. Appl. Math. & Informatics, MSU Branch;

Ibragimova D.E. - Associate professor of Dpt. Appl. Math. & Informatics, MSU Branch;

Israilova D.M. - Associate professor of Dpt. Appl. Math. & Informatics, MSU Branch;

Sheraliyev Sh.N. - Acting of vice-dean of faculty of Appl. Math. & Informatics;

Nurmukhamedova N.S. - Acting of Associate professor Dpt. Econometrics & Economical Modeling, NUUz;

Kakadjanova L.R. - PhD student of Dpt. Prob. Theory & Math. Stat., NUUz;

Section of conference

1. Mathematical Statistics.
2. Probability and related problems.
3. Discrete and continuous mathematics;
4. Mathematical-Statistical methods in economics, medicine, natural sciences and teaching methodology.
5. Fundamental informatics and information technology;.

ISBN 978-9943-305-86-1

Sponsors:
Conference was supported by Foundation for Fundamental Researches Republic of Uzbekistan (Grant F4-40). Conference was supported by the Insurance company **"APEX Insurance"** LLC.

ПРЕДИСЛОВИЕ

Конференция "СТАТИСТИКА и ее применения"была приурочена к Всемирному Дню Статистики, который отмечается каждый год 20 октября, начиная с 2010 года во многих странах в соответствии с Резолюцией Генеральной Ассамблеи ООН. Конференция призвана объединить ученых не только Республики Узбекистан, которые занимаются актуальными вопросами статистики и ее практического применения, а также и других стран. I научно-практическая конференция "СТАТИСТИКА и ее применения"проводилась 17-18 октября 2012 года, в материалы которой были включены 86 статей. 17-18 октября 2013 года была организована II научно-практическая конференция "СТАТИСТИКА и ее применения". Конференция была приурочена к 22-годовщине Независимости Республики Узбекистан, , а также тому, что 2013 год был объявлен Всемирным Годом Статистики. Материалы конференции включили в себя 87 статей. III научно-практическая конференция "СТАТИСТИКА и ее применения"проводилась 16-17 октября 2015 года, и в материалах конференции было опубликовано 110 статей.

25-27 апреля 2016 года был проведен совместный научно-практический семинар "СТАТИСТИКА и ее применения"кафедры "Теория Вероятностей и Математическая Статистика" НУУз и отдела "Биостатистика" Университета Флориды США с участием профессоров Сомнат Датта, Сусмита Датта и их учеников. IV конференция "СТАТИСТИКА и ее применения"проводилась 19-20 октября 2017 года. В материалах этой конференции были представлены 90 статьей. Настоящая, V конференция "СТАТИСТИКА и ее применения"проводится 17-18 октября 2019 года в Филиале МГУ в г. Ташкенте и она приурочена также и 60 летию проф. А.А.Абдушукурова, профессора филиала и основного организатора этих конференций. Труды данной конференции включили в себя 72 работ не только ученых из Узбекистана но также и США, Индии и России. Оргкомитет выражает глубокую благодарность всем участникам конференции любезно представившим свои научные работы в данный сборник.

Оргкомитет

PREFACE

The conference "STATISTICS and its applications" was dedicated to the World Day of Statistics which celebrated on October 20 every year, since 2010 in many countries according to the UN General Assembly Resolution. Conference is aimed to collaboration of scientists not only of the Republic of Uzbekistan, which are engaged in current issues of statistics and its practical application, but also from other countries. I scientific-applied conference "STATISTICS and its applications" was held on 17-18 October, 2012 and materials of the conference include 86 articles. On 17-18 October, 2013 was organized II scientific-applied conference "STATISTICS and its applications". This conference was dedicated to the 22nd anniversary of Independence of the Republic of Uzbekistan, , also that 2013 year declared the International Year of Statistics. Materials of this conference consists of 87 articles. III scientific-applied conference "STATISTICS and its applications"was held on 16-17 October, 2015 and materials of the conference include 110 articles.

25-27 April 2016 was organizes jointly scientific-applied seminar "STATISTICS and its applications"of the Department of "Probability Theory and Mathematical Statistics"of National University of Uzbekistan and Department of "Biostatistics"of University of Florida of USA with participating of professors Somnath Datta and Susmita Datta and their graduate students.

IV scientific-applied conference "STATISTICS and its applications" organized on 19-20 October, 2017. Materials of this conference include 90 papers. This, V conference "STATISTICS and its applications"organized on October 17-18, 2019 in Tashkent Branch of MSU and it dedicated also to 60^{th} Birthday of Prof. A.A.Abdushukurov, the Principle Organizer of these conferences. Proceedings of this conference include 72 papers of scientists not only from Uzbekistan, also from USA, India and Russia. The organizing committee expresses its deep appreciation to all the participants of the conference who kingly submit their articles to this book.

Organizing Committee



60-year Jubilee of Professor Abdurakhim Abdushukurov

To the 60-years Jubilee of Prof. Dr. Abdurakhim Abdushukurov

Today we celebrating the 60^{th} birthday of Famous Scientist on Mathematical Statistics of Professor Abdushukurov Abdurakhim Ahmedovich. After graduating from Mathematics Faculty of Tash SU {Tashkent State University} in 1981, A.A. Abdushukurov was directed to Mathematics Institute of Academy of Sciences of Republic of Uzbekistan (AS RUz). There he worked as a senior laboratory worker, junior scientific researcher and scientific researcher of Mathematical Statistics department. Since 1990 up to 2017 A.A.Abdushukurov has been worked in National University of Uzbekistan (NUUz, former Tash SU) in following positions: senior lecturer (1990-1992), associate professor and doctorant (1992-1995), vice-dean (1996-1997), associate professor (1997-2002), vice-dean (2002-2003), associate professor (2003-2005), professor (2005-2006), head of Dpt. *Probability Theory and Mathematical Statistics* (2006-2010),vice-rector (2010), head of Dpt. *Of Probability Theory and Mathematical Statistics* (2010-2016), dean of *Mathematics Faculty* (2016), rector (2016) and head of Dpt. *Probability Theory and Mathematics Statistics* (2016-2017). Since 2018 he worked as professor of Dpt. *Applied Mathematics and Informatics* of Tashkent Branch of Moscow State University named after M.V.Lomonosov.

On March 15, 1989 he defended candidates dissertation (PhD) in physical-mathematical sciences on the speciality Probability Theory and Mathematical Statistics on the subject *"About effectiveness of statistical estimates by censored observations and new estimates in the proportional hazards model"/,* through specialized council of Mathematics Institute of AS RUz.

On July 7, 2005 he defended senior doctorates dissertation (DSc) on the speciality Probability Theory and Mathematical Statistics on the subject *"Nonparametric estimates of functionals from unknown distributions by incomplete observations and their statistical analysis"* through specialized council of NUUz.

Prof. A.A. Abdushukurov's scientific interests include such problems of mathematical statistics as parametrical, semiparametrical and nonparametrical statistical estimation by incomplete observations. He proposed and studied new functionals of unknown sub-distributions and using them built nonparametric estimates in generalized models of incomplete observations. These models include several models of random censoring and competing risks. Estimates are studied as in fixed as well as in random sample sizes. While studying estimates properties, along with analytical methods of probability theory and mathematical statistics, he used asymptotical theory of empirical processes,

method of martingales, strong and weak approximations and U-statistics.

Prof. A.A. Abdushukurov's main results are followings:

He proposed basis functionals of exponential, product and power structures and studied their properties. Using these basis functionals he built nonparametric estimates in generalized models of incomplete observations. Set properties of uniform strong consistency with speed of convergence Ц laws of iterated logarithm type, both in homogenous and in variable random censoring are investigated. Proved generalized Gaussian approximations in weak and strong forms with optimal speeds of approximation. He proved that three classes of estimates converge to the same Gaussian process.

He found and studied similar properties of nonparametric estimates with random Poisson size of sample and in Bayesian estimating using a priori Dirichlet distribution. Estimates on their own are new. He is co-author of celebrated ACL (Abdushukurov-Cheng-Lin)-estimate of survival function.

He proposed and systematically studied nonparametric estimates of multivariate survival functions, when incomplete data are non-homogenous, dependent and censored random vectors.

In Cox's regression model with random censoring from both sides he built and studied three classes of parametric-nonparametric estimates for basis functional of survival function.

Using method of strong approximation in several models of random censoring he obtained presentations for log likelihood ratio statistics and its truncated analogy through stochastic integrals from two-parametrical Wiener processes.

Majority of these results are essential investments to the theory of nonparametric statistics of incomplete observations. They can be applied to the analysis of survival data in insurance, demography, biology and astronomy.

A.A.Abdushukurov is often cited scientist in Google-Scholar and Scopus.

Prof. A.A. Abdushukurov reads lectures and conducts practical lessons for students in the following courses: *"Modern problems of Applied Mathematics"* , *"Probability Theory and Applied Statistics"* , *"Mathematical Statistics"* , it "Multivariate Statistical Analysis" , *"Statistics of Incomplete Observations"* and special courses on *"Non-classic Models of Statistics"* , *"Regression Analysis"* and *"Statistical Hypothesis Testing"* for Master's students using latest advances in the field of mathematical statistics.

Under his supervision there were more than 40 Master's dissertations and 4 PhD dissertations have defended on Probability Theory and Mathematical Statistics. Currently under Prof. A.A. Abdushukurov's supervision completed his dissertations 5 of PhD students.

Since 2006 up to present time prof. A.A.Abdushukurov guided of 3-th Scientific Grant of Republic of Uzbekistan in domain of Mathematical Statistics.

Prof. A.A. Abdushukurov is the author more than 330 of scientific and methodical works (monographs and Text books (15), papers in scientific journals and books (more than 200). He is principal organizer of scientific-applied conference *"STATISTICS and its application"* (2012, 2013, 2015, 2017, 2019).

Prof. A.A. Abdushukurov actively participates in many international scientific conferences, among which one can note the following:

I - *Congress of Bernoulli Society of Statistics* (Tashkent, 1986);

- V -, VII -, VIII - *Vilnius Conferences on Pr. The. & Math. Stat.* (Lithuania, 1989, 1998, 2002);

- IV, XIII - *Islamic Countries Conf. on Stat Sciences* (Lahore, Pakistan, Indonesia, Bogor, 1994,2015);

- VII - *Belarusian Mathematical Congress* (Minsk, Belorussia, 1997);

- 22*nd European Meeting of Statistics* (Vilnius, Lithuania, 1998);

- *Each Year Russian Workshop on Stochastic Methods* (Russia, since 1995);

- *International Seminar On Stability Problems for Stochastic Models* (Russia, Hungaria, since 2011);

- *Advances in Statistical Inferential Methods* (Almaty, Kazakhstan, 2003);

- *Bernoulli Society EAPR Conference* (Hong - Kong, 2003);

- *FAM*: (Krasnoyarsk, Russia, 2008, 2009, 2010-2015);

- *AMSA* (Russia, 2011, 2013, 2015, 2017);

- *CDAM* (Minsk,Belorusia, 1995, 2016, 2019);

- Prof. A.A. Abdushukurov's some important publications (total of 70):

- *Communications in Statistics. Theory & Methods.* (USA, 1998, 2012, 2015);

- *Sankhiya: The Indian Journal of Statistics* (India,1987);

- *Journal of the Italian Stat. Society.* (Italy,1997);

- *Pakistan J. of Statistics* (Pakistan,1996);

- *Theory Probab. Appl-ns*: (Russia,1998, 2000);

- *Statistical Methods of Estimation and Hypothesis Testing* (Perm State University, Russia, 1996, 1999, 2001, 2002, 2005, 2006-2012, 2015, 2018, 2019);

- *Intern. J. Sci. Advanced Technology* (2012);

- J. of Mathematical Sciences (USA, Springer 2005, 2006, 2007, 2012, 2013, 2017, 2019);

- *Applied Mathematics* (USA, 2015);

- *J. of Siberian Federal University. Mathematics & Physics* (Russia, 2013, 2014, 2015, 2018);

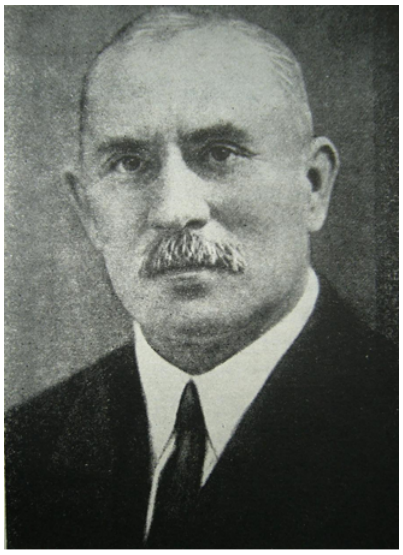
- *International J. of Innovative Research in Science, Engineering and Technology* (USA, 2015);
- *Intern. J. Innovation in Science and Mathematics* (2014);
- *Open Science J. of Statistics and Application* (USA, 2015);
- *New Trends in Mathematical Sciences* (Turkey, 2014, 2015);
- *Statistics of Incomplete Observations: Asymptotic Theory of Nonclassical Models.* Tashkent. 2009. University Press. 267 pages .
- *Industrial Mathematics. Mathematical Methods of Investigations* (Russia, 2014,2016);
- *International Encyclopedia of Statistical Sciences.* Springer. 2011. Part 14;
- *Proceedings of Internat. Workshop AMSA (Applied Methods of Statistical Analysis, Simulation and Statistical Inference). Novosibirsk State Technical University.* Novosibirsk. (Russia, 2011, 2013, 2015, 2019);
- *Estimators of Unknown Distributions from Incomplete Observations and Its Properties.* LAMBERT Academic Publishing. 2011. 299 p. (Germany);
- *Asymptotic Analysis of Kernel Estimators Connected with Mean Value.* LAMBERT Academic Publishing. 2011. 132 p. (Germany);
- *Estimation of Multivariate Distributions and Its Mixtures from Incomplete Data.* LAMBERT Academic Publishing. 2011. 123 p. (Germany);
- *Investigations of Power Estimates of Survival Function in Regression Models.* LAMBERT Academic Publishing. 2012. 89 p. (Germany);
- *Local Asymptotic Normality of Statistical Tests.* LAMBERT Academic Publishing. 2012. 136 p. (Germany);
- *Sequential Estimation by Intervals of Fixed Width,* LAMBERT Academic Publishing 2017. 89 p.(Germany);
- *Asymptotical Properties of Statistical Experiments.* LAMBERT Academic Publishing 2019. 150 p. (Germany);
- *Far East J. Theoretical Statistics* (India, 2013);
- *Proceeding of International Conference CDAM* (*Computer Data Analysis and Modeling*). Belorussian State University. Minsk (1995, 2016, 2019).

We would like to say a special thank to our supervisor, Prof. A.A. Abdushukurov. His support, guidance and overall insights in scientific field have made this an inspiring experience for us, for his patience, motivation and immense knowledge. We wish to Prof. A.A.Abdushukurov good health and great successes in their scientific investigations.

Organizing Committee and Students

О работах В.И. Романовского в области математической статистики (к 140-летию В.И. Романовского)

А.А. Абдушукуров
Филиал Московского Государственного Университета им. М.В. Ломоносова в г. Ташкент
a_abdushukurov@rambler.ru



Введение. В этом году, 5 декабря исполняется 140 лет со дня рождения замечательного математика Всеволода Ивановича Романовского (1879-1954), крупнейшего специалиста в области теории вероятностей и математической статистики, одного из основателей Среднеазиатского Университета (ныне Национального Университета Узбекистана) и высшего образования в республиках Средней Азии и, в частности, в Узбекистане. Начало теоретических и прикладных исследований по теории вероятностей и математической статистике в Узбекистане тесным образом связано с жизнью и деятельностью В.И. Романовского. Будучи учеником А.А. Маркова и представителем Петербургской математической школы, он является основоположником Ташкентской математической школы. Он опубликовал свыше 180 научных работ в разных областях математики. Несмотря на то, что в начале своей научной деятельности он занимался дифференциальными уравнениями, а также вопросами алгебры и теории чисел, подавляющее большинство работ В.И. Романовского относится к теории вероятностей и математической статистике.

Исследования в области математической статистики. В.И. Романовский начал заниматься проблемами, связанными с математической статистикой, приблизительно с 1910 г. В Варшаве была опубликована его работа *"Закон больших чисел и теория Якоба Бернулли"*. В 1916 г. на заседании общества естествоиспытателей Варшавского Университета он сделал доклад о своем исследовании, посвященном обобщению кривых Пирсона. Дальнейшая его работа в этом направлении была связана с Ташкентом. Уже 1919-1920 г.г. он начал читать курс лекций по математической статистике в Среднеазиатском университете. Интерес В.И. Романовского к ма-

тематической статистике обусловлен появлением в ней достижений представителей английских и немецких школ. Важное значение для понимания творчества В.И. Романовского в области математической статистики занимает его статья *"Статистическое мировоззрение"* (1921). В ней он не только подвел итоги развития этой науки, но, в сущности, составил план своей дальнейшей работы. В 1925 г. В.И. Романовский отправился в зарубежную командировку, которая имела важное значение для его научной деятельности в области математической статистики. Он лично познакомился с известными специалистами в области математической статистики Карлом Пирсоном и Рональдом Фишером. К. Пирсон (1857-1936) - английский математик, биолог, профессор Лондонского университетского колледжа, автор теории корреляции, критериев хи-квадрат проверки соответствия результатов экспериментов гипотетическим предположениям. Он ввел кривые распределения (*"кривые Пирсона"*), включающие в себя многие известные распределения, и развил теорию метода моментов для оценивания неизвестных параметров генеральной совокупности. В 1900 г. Пирсон основал журнал *"Biometrica"*, постоянным читателем которого был и В.И. Романовский и, в котором, начиная с 1923 г. , печатались и его статьи (1923,1924,1927,1933,1936 г.г.). Следует отметить, что этот журнал и в настоящее время является одним из престижных изданий в области математической и прикладной статистики. Другим английским ученым, также оказавшим большое влияние на В.И. Романовского, был Р. Фишер (1890-1962), математик и генетик, автор многих трудов по математической статистике, в частности, один из основателей современной теории эффективности статистических оценок, в которой *"информация Фишера"* занимает ключевое место. Оценку исследований Фишера Романовский дал в своей рецензии на его книгу "Статистические методы для исследователей"(1934 г.), рассмотрев ее как "замечательное явление в статистической литературе". Установленные во время командировки В.И. Романовского связи с К. Пирсоном и Р. Фишером, с основателями современной математической статистики, и другими иностранными учеными, в дальнейшем не прерывались. Однако, следует отметить, что в 30-годы прошлого века, когда началась борьба "за материалистическую диалектику в математике всей силой обрушившаяся на математическую статистику, поддерживать связи с иностранными учеными было нелегко. Были попытки *"перестройки"* математической статистики в духе идеологии коммунистов. В этой связи стоит упомянуть Всесоюзное совещание по математической статистике, проходившее осенью 1948 г. в Ташкенте. Участники этого совещания должны

были обсудить развитие идей К. Пирсона и Р. Фишера, представлявших собой как *"преклонение перед капитализмом"*. На совещании В.И. Романовский выступил с двумя докладами, и его выступления были сопровождены бурными дискуссиями. В.И. Романовскому, долгие годы сотрудничавшему с К. Пирсоном и Р. Фишером и в своих исследованиях придерживавшемуся англо-американских методов, по тем временам пришлось *"признать идеологические ошибки"*. Сейчас все это кажется абсурдным, так как нельзя представить современную математическую статистику без базовых идей и методов Пирсона и Фишера.

В своих исследованиях Романовский обращается к работам Стьюдента, Пирсона и Фишера. Переходя к критериям оценки статистик, он указывает, что по Фишеру они могут быть построены для отыскания некоторых неизвестных параметров, являющихся характеристикой генеральной совокупности и бывают трех типов: *критерий состоятельности, критерий эффективности* и *критерий достаточности*. В.И. Романовский выделяет замечание Фишера о методе моментов Пирсона: *"этот способ представляет собой одно из применений метода состоятельности"*. Останавливаясь на анализе критерия эффективности и отыскания эффективных статистик, Романовский предлагает, следуя Фишеру, эффективную статистику, соответствующую исследуемому параметру. Для этого он применяет метод наибольшего правдоподобия и приводит его доказательство. Затем он приводит и доказывает следующую теорему, весьма важную, по его словам, для всего учения Фишера: *"Если некоторая статистика построена при помощи уравнения наибольшего правдоподобия и для выборок большого объема имеет распределение, близкое к нормальному, то она эффективна"*. Используя терминологию современной математической статистики, в этой теореме речь пойдет о *"наилучшей асимптотической нормальности оценок наибольшего правдоподобия"*, что является одним из важных свойств оценок.

В.И. Романовский обращает свое внимание на проблему оценки неизвестных параметров с использованием следующих трех способов: классический способ апостериорных вероятностей, способ доверительных интервалов Неймана. Разъяснив сущность идей Фишера и Неймана, Романовский переходит к изложению своих собственных результатов, относящихся к обобщению закона больших чисел, свободных от априорных вероятностей.

В.И. Романовский впервые дал аналитический вывод законов распределения критериев t и Z Стьюдента-Фишера, эмпирических коэффициентов

регрессий; им была создана более общая теория критерия хи-квадрат Пирсона. Из работ по математической статистике следует упомянуть также и следующие:

- *"Математическая статистика"* (1938);
- *"Статистические задачи, связанные с цепями Маркова"* (1940);
- *"Основные понятия и задачи математической статистики"* (1941);
- *"Об индивидуальных выводах в статистике"* (1942).

Интенсивно занимаясь научной деятельностью в области математической статистики в течение полувека, В.И. Романовский, особенно с 30-х годов, уделяет большое внимание применению статистических методов и в производстве. К числу его работ можно отнести следующие:

- *"О новейших методах математической статистики, применяемых в полевом опыте"* (1934);
- *"Математическая статистика и индустрия"* (1934);
- *"Об одном методе индустриального контроля"* (1936);
- *"О распределении цели"* (1950).

Следующие работы В.И. Романовского, посвященные статистическому контролю качества готовой продукции, представляют значительный интерес:

- *"Оценка качества продукции, основанная на текущих малых выборках"* (1947);
- *"О статистическом контроле хода производства при помощи упорядоченных выборок"* (1950);
- *"О приемочном статистическим контроле"* (1951);
- *"О статистических методах контроля производства и качества продукции"* (1951).

В рамках одной небольшой статьи сделать обзор о научном наследии В.И. Романовского, хотя бы в области математической статистики, практически невозможно. Однако, следует отметить, что важной заслугой В.И.Романовского является то, что он подвел строгую математическую базу под математическую статистику. Продолжая традиции своего учителя, последователи В.И. Романовского проводят исследования по современным направлениям математической статистики, сотрудничая с ведущими учеными и научными центрами всего мира. Так, первым собранием Общества имени Бернулли по математической статистике и теории вероятностей был Конгресс, который проходил в г. Ташкенте в сентябре 1986г. под почетным председательством академика Ю.В. Прохорова, Ташкентским Оргкомитетом под председательством академика С.Х. Сираждинова, Международ-

ным Оргкомитетом под руководством профессора П. Ревеза (Венгрия) и Международным Программным Комитетом под председательством профессора К. Крикеберга (Франция). В конгрессе участвовало более 1000 иностранных и отечественных ученых. На церемонии закрытия Конгресса Общества имени Бернулли его президент К. Хейде (Австралия) подчеркнул, что успешно прошедший и хорошо организованный конгресс в Ташкенте оказался плодотворным с научной точки зрения, и дал целостное представление о нынешнем состоянии вероятностно-статистической науки в единстве с ее практическими применениями.

Учитывая все более возрастающий интерес к развитию и применению статистических методов, начиная с 2012 года в Узбекистане регулярно (2012, 2013, 2015, 2017, 2019) проводится научно-практическая конференция *"СТАТИСТИКА и ее применения"* с участием зарубежных ученых, специалистов. Эта конференция приурочена также и ко Всемирному Дню Статистики - 20 октября.

ЛИТЕРАТУРА

1. Абдушукуров А.А. Из ранней истории развития статистики. Матер. научно-практ. конф. "СТАТИСТИКА и ее применения". Ташкент. 2012. 17-18 октября с.5-20.

2. Абдушукуров А.А. Из истории развития статистики в Узбекистане. Матер. научно-практ. конф. "СТАТИСТИКА и ее применения". Ташкент. 2015. 16-17 октября с.7-19.

3. Боголюбов А.Н., Матвиевская Г.П. Всеволод Иванович Романовский. 1879-1954. М.:Наука 1997. 157 стр.

4. Романовский Всеволод Иванович. Материалы к библиографии ученых Узбекистана. Ташкент ФАН. 1979. 30 стр.

Cramer-Rao Inequality Revisited for Randomly Censored Data

B.L.S. PRAKASA RAO
CR Rao Advanced Institute of Research in Mathematics, Statistics and Computer Science, Hyderabad 500046, India

Abstract. *Cramer-Rao inequality for randomly censored data was derived by Abdushukurov and Kim (J. Soviet. Math. (1987) pp. 2171-2185). Cramer-Rao type integral inequalities for randomly censored data were derived in Prakasa Rao (1995) extending the work of Abdushukurov and Kim (1987) to obtain lower bounds for the Bayesian risk. As an application of Walker's inequality which is an improvement of the Cauchy-Schwartz inequality, Prakasa Rao (2018) obtained improved versions of Cramer-Rao type inequality for randomly censored data. We present a short overview of these results.*

1 Introduction

Suppose that ψ and η are random variables defined on a probability space with distribution functions $F(x;\theta)$ and $G(x;\theta)$ respectively where $\theta \in \Theta \subset R$. Further suppose that $\psi_i, 1 \leq i \leq n$ are independent and identically distributed (i.i.d.) as the random variable ψ and $\eta_i, 1 \leq i \leq n$ be i.i.d. random variables distributed as the random variable η. Define $\zeta_i = \min(\psi_i; \eta_i)$ and $\delta_i = I(\psi_i \leq \eta_i), 1 \leq i \leq n$ where $I(A)$ denotes the indicator function of a set A. Suppose that ψ_i and η_i are not observable but $(\zeta_i; \delta_i)$ is observable for $1 \leq i \leq n$. The data $(\zeta_i; \delta_i), 1 \leq i \leq n$ is termed as *randomly censored data* and is common in studies in survival analysis and reliability. It is assumed that the random variables $\psi_i, 1 \leq i \leq n$ are independent of the *censoring* random variables $\eta_i, 1 \leq i \leq n$.

Abdushukurov and Kim (1987) obtained a lower bound for the variance of an unbiased estimator of the parameter θ based on the data (ζ, δ). They have derived analogues of the Cramer-Rao and Bhattacharya (1946) bounds in the uncensored case. Wyckoff and Engelhardt (1980) derived Cramer-Rao lower bounds for estimators based on data obtained from Type II censoring. Eubank and La Riccia (1982) and Crow and Shimi (1982) studied the problem for Type I censoring. In a recent work, Walker (2017) obtained an improved version of the Cauchy-Schwartz inequality. Cramer-Rao type integral inequalities for uncensored data in the Bayesian framework were derived in Prakasa Rao (1991,1992) among others. Cramer-Rao type integral inequalities for randomly censored data were derived in Prakasa Rao (1995) extending the work of

Abdushukurov and Kim (1987) to obtain lower bounds for the Bayesian risk. We give a short review of some of these results.

2 Preliminaries

In the sequel, we assume that the distribution functions $F(x;\theta)$ and $G(x;\theta)$ are absolutely continuous with densities $f(x;\theta)$ and $g(x;\theta)$ respectively. It is easy to see that the random variables $\zeta_i, 1 \leq i \leq n$ are independent and identically distributed random variables with the distribution function

$$H(x,\theta) = 1 - \overline{F}(x,\theta)\overline{G}(x,\theta)$$

where $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$. Suppose that the densities f and g are differentiable with respect to θ up to k times. Let $f^{(i)}$ denote the i-th derivative f with respect to θ and $g^{(i)}$ denote the i-th derivative of g with respect to θ: Let

$$\lambda^{(i)}(x,\theta) = \frac{f^{(i)}(x,\theta)}{f(x,\theta)} \text{ and } \mu^{(i)}(x,\theta) = \frac{g^{(i)}(x,\theta)}{g(x,\theta)}.$$

Suppose the following regularity conditions hold:

(C_1) The derivatives $f^{(i)}(x,\theta)$ and $g^{(i)}(x,\theta)$ with respect to θ exist for $i = 1, ..., k$ and

$$\int_{-\infty}^{\infty} \left| f^{(i)}(x,\theta) \right| dx < \infty \text{ and } \int_{-\infty}^{\infty} \left| g^{(i)}(x,\theta) \right| dx < \infty.$$

Let $\phi(\theta)$ be a real-valued function of the parameter θ and suppose that $\phi(\theta)$ is differentiable k times with respect to θ. Let $\phi^{(i)}(\theta)$ denote the i-th derivative of $\phi(\theta)$ with respect to θ. Suppose that $\widehat{\phi}(\zeta,\delta)$ is an estimator of the function $\phi(\theta)$ based on the observation (ζ,δ). Let $E_\theta[\widehat{\phi}(\zeta,\delta)] = \gamma(\theta)$.

Observe that

$$\int_{R\times\{0,1\}} \widehat{\phi}(x,y)h(x,y,\theta)d\nu(x,y) = \gamma(\theta), \tag{2.1}$$

where

$$h(x,y,\theta) = \left[\overline{G}(x,\theta)f(x,\theta)\right]^y \left[\overline{F}(x,\theta)g(x,\theta)\right]^{1-y}$$

is the joint density function of the random vector (ζ,δ) and $v(x;y)$ is appropriate σ-finite measure on $R \times \{0,1\}$.

(C_2) Suppose the integral on the left side of the equation (2.1) is differentiable k-times with respect to θ under the integral sign.

In particular, it follows that

$$\int_{R\times\{0,1\}} \widehat{\phi}(x,y)h^{(i)}(x,y,\theta)d\nu(x,y) = \gamma^{(i)}(\theta), \quad i=1,...,k \tag{2.2}$$

where

$$h^{(i)}(x,y,\theta) = \frac{\partial^i}{\partial\theta^i}h(x,y,\theta).$$

(C_3) Suppose that $E_\theta\left|\lambda^{(i)}(\psi,\theta)\right|^2 < \infty$ and $E_\theta\left|\mu^{(i)}(\eta,\theta)\right|^2 < \infty$. Consider the random variable

$$S_i(\zeta,\delta,\theta) = \sum_{j=0}^{i}\binom{i}{j}\left[\delta\frac{f^{(i)}(\zeta,\theta)\overline{G}^{(i-j)}(\zeta,\theta)}{f(\zeta,\theta)\overline{G}(\zeta,\theta)} + (1-\delta)\frac{g^{(i-j)}(\zeta,\theta)\overline{F}^{(j)}(\zeta,\theta)}{g(\zeta,\theta)\overline{F}(\zeta,\theta)}\right]$$

where

$$\overline{F}^{(i)}(x,\theta) = \tfrac{\partial^i}{\partial\theta^i}\overline{F}(x,\theta) \text{ and } \overline{G}^{(i)}(x,\theta) = \tfrac{\partial^i}{\partial\theta^i}\overline{G}(x,\theta).$$

Define

$$\mathbf{S}^T(\zeta,\delta,\theta) = (S_1(\zeta,\delta,\theta),...,S_k(\zeta,\delta,\theta))$$

where $\mathbf{S}^T$ denotes the transpose of the vector S.

(C_4) Suppose the matrix

$$I(\theta) = E_\theta\left[S(\zeta,\delta,\theta)S^T(\zeta,\delta,\theta)\right] = ((I_{ij}(\theta)))_{k\times k}$$

is positive definite where

$$I_{ij}(\theta) = S_i(\zeta,\delta,\theta) =$$

$$= \sum_{l=0}^{i}\sum_{m=0}^{j}\binom{i}{l}\binom{j}{m}\left[\int_{-\infty}^{\infty}\frac{f^{(l)}(x,\theta)f^{(m)}(x,\theta)\overline{G}^{(i-l)}(x,\theta)\overline{G}^{(j-m)}(x,\theta)}{f(x,\theta)\overline{G}(x,\theta)}dx + \right.$$

$$\left. + \int_{-\infty}^{\infty}\frac{g^{(i-l)}(x,\theta)g^{(j-m)}(x,\theta)\overline{F}^{(l)}(x,\theta)\overline{F}^{(m)}(x,\theta)}{g(x,\theta)\overline{F}(x,\theta)}dx\right].$$

(C_5) The set $\{x : \overline{G}(x,\theta)f(x,\theta) = 0\} \cup \{x : \overline{F}(x,\theta)g(x,\theta) = 0\}$ does not depend on the parameter θ.

Let us first consider the case $k = 1$. The following theorem is due to Abdushukurov and Kim (1987) generalizing the Cramer-Rao inequality for the randomly censored data.

Theorem 2.1. *Suppose the conditions $(C_1)-(C_5)$ hold and the estimator $\widehat{\phi}(\zeta,\delta)$ is an unbiased estimator of the function $\phi(\theta)$. Then*

$$Var\left(\widehat{\phi}(\zeta,\delta)\right) \geq \tfrac{\left[\phi^{(1)}(\theta)\right]^2}{I(\theta)}. \tag{2.3}$$

Remarks: A Bayesian version of the result stated above was obtained in Prakasa Rao (1995). An improved version of this result will be derived later as an application of the improved Cauchy-Schwartz inequality due to Walker (2017).

The following theorem, again due to Abdushukurov and Kim (1987), generalizes Theorem 2.1 to the case $k > 1$ and it is an analogue of the Bhattacharya bound in the uncensored case. They have also characterized the class of distributions for which equality is attained in the equation (2.3) and in the inequality (2.4) given below.

Theorem 2.2. *Suppose the conditions* $(C_1) - (C_5)$ *hold and the estimator* $\widehat{\phi}(\zeta, \delta)$ *is an unbiased estimator of the function* $\phi(\theta)$. *Then*

$$Var\left(\widehat{\phi}(\zeta,\delta)\right) \geq \sum_{i,j=1}^{k} \phi^{(i)}(\theta)\phi^{(j)}(\theta)\widetilde{I}_{ij}(\theta) \tag{2.4}$$

where

$$((\widetilde{I}_{ij}(\theta)))_{k\times k} = ((I_{ij}(\theta)))^{-1}_{k\times k}.$$

3 Improved Cauchy-Schwartz inequality

Walker (2017) obtained an improved version of the Cauchy-Schwartz inequality which implies the following probabilistic version.

Theorem 3.1. (Walker's inequality) *If* X *and* Y *are random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *with finite second moments, then*

$$|E(XY)|^2 \leq E(X^2)E(Y^2) - \left(|E(X)|\sqrt{Var(Y)} - |E(Y)|\sqrt{Var(X)}\right)^2. \tag{3.1}$$

In particular, if the random variable Y has mean zero but positive variance, then it follows that

$$|E(XY)|^2 \leq E(X^2)E(Y^2) - |E(X)|^2 E(Y^2) = Var(X)E(Y^2). \tag{3.2}$$

Hence

$$E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)}$$

and we have the following corollary.

Corollary 3.2. *If* X *and* Y *are random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *with finite second moments and if* $E(Y) = 0$, *then*

$$E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)}. \tag{3.2}$$

Walker (2017) indicated that the inequality derived in Theorem 3.1 is an improved inequality over the Cauchy-Schwartz inequality. For instance if Y is a standard normal random variable and X is a random variable with mean μ and variance σ^2, then the Cauchy-Schwarz inequality shows that

$$[E(XY)]^2 \leq E(X^2)E(Y^2) = \mu^2 + \sigma^2$$

where as the improved Cauchy-Schwartz inequality implies that

$$[E(XY)]^2 \leq \sigma^2$$

which is a strict improvement over the earlier inequality.

4 Cramer-Rao inequalities for randomly censored data

The following lemma is proved in Abdushukurov and Kim (1987).

Lemma 4.1. *Suppose the conditions* (C_1) *and* (C_5) *hold. Then, almost everywhere on* $R \times \{0,1\}$*, there exist derivatives* $h^{(1)}(x,y,\theta)$ *for all* $\theta \in \Theta$ *and for* $i = 1, ..., k$

(4.1)

$$h^{(i)}(x,y,\theta) = h(x,y,\theta) \sum_{j=0}^{i} \left(\begin{array}{c} i \\ j \end{array} \right) [\, y \frac{f^{(i)}(x,\theta)\overline{G}^{(i-j)}(x,\theta)}{f(x,\theta)\overline{G}(x,\theta)} +$$

$$+(1-y)\frac{g^{(i-j)}(x,\theta)\overline{F}^{(j)}(x,\theta)}{g(x,\theta)\overline{F}(x,\theta)}\Bigg].$$

Let

$$S_{1i}(\zeta,\delta,\theta) = \textstyle\sum_{j=0}^{i} \left(\begin{array}{c} i \\ j \end{array} \right) \left[\delta \frac{f^{(i)}(\zeta,\theta)\overline{G}^{(i-j)}(\zeta,\theta)}{f(\zeta,\theta)\overline{G}(\zeta,\theta)}\right],\ i = 1, ..., k \tag{4.2}$$

and

$$S_{2i}(\zeta,\delta,\theta) = \textstyle\sum_{j=0}^{i} \left(\begin{array}{c} i \\ j \end{array} \right) \left[(1-\delta) \frac{g^{(i-j)}(\zeta,\theta)\overline{F}^{(j)}(\zeta,\theta)}{g(\zeta,\theta)\overline{F}(\zeta,\theta)}\right],\ i = 1, ..., k. \tag{4.3}$$

Define

$$\mathbf{S}^{(1)}(\zeta,\delta,\theta) = (S_{11}(\zeta,\delta,\theta), ..., S_{1k}(\zeta,\delta,\theta))$$

and

$$\mathbf{S}^{(2)}(\zeta,\delta,\theta) = (S_{21}(\zeta,\delta,\theta), ..., S_{2k}(\zeta,\delta,\theta)).$$

Note that

$$\mathbf{S}(\zeta, \delta, \theta) = S^{(1)}(\zeta, \delta, \theta) + S^{(2)}(\zeta, \delta, \theta).$$

We have obtained an improved version of the inequality derived in Theorem 2.1 due to Abdushukurov and Kim (1987). Let

$$F^*(x, \theta) = P_\theta(\zeta \leq x, \delta = 1) = \int_{-\infty}^{x} \overline{G}(t, \theta) f(t, \theta) dt$$

and

$$G^*(x, \theta) = P_\theta(\zeta \leq x, \delta = 0) = \int_{-\infty}^{x} \overline{F}(t, \theta) g(t, \theta) dt.$$

Observe that

$$F^*(x, \theta) + G^*(x, \theta) = H(x, \theta) = 1 - \overline{F}(x, \theta)\overline{G}(x, \theta), \ x \in R$$

and

$$F^*(\infty, \theta) + G^*(\infty, \theta) = 1$$

where

$$F^*(\infty, \theta) = \lim_{x \to \infty} F(x, \theta) = P_\theta(\delta = 1) = E_\theta(\delta)$$

and

$$G^*(\infty, \theta) = \lim_{x \to \infty} G(x, \theta) = P_\theta(\delta = 0) = E_\theta(1 - \delta).$$

Let us first consider the case $k = 1$. Suppose that $\widehat{\phi}(\zeta, \delta)$ is an estimator of the parametric function $\phi(\theta)$ and suppose that

$$E_\theta \left[\widehat{\phi}(\zeta, \delta) \right] = \gamma(\theta), \ \theta \in \Theta.$$

Note that

$$\int_{R \times \{0,1\}} \widehat{\phi}(x, y) h(x, y, \theta) d\nu(x, y) = \gamma(\theta).$$

Then, by Lemma 4.1, it followes that

$$S(\zeta, \delta, \theta) = S_{11}(\zeta, \delta, \theta) + S_{21}(\zeta, \delta, \theta)$$

where

$$S^{(1)} = S_{11} = \frac{h^{(1)}(\zeta, \delta, \theta)}{h(\zeta, \delta, \theta)} = \delta \left[\frac{f^{(1)}}{f} + \frac{\overline{G}^{(1)}}{\overline{G}} \right]$$

and

$$S^{(2)} = S_{21} = (1 - \delta) \left[\frac{g^{(1)}}{g} + \frac{\overline{F}^{(1)}}{\overline{F}} \right].$$

Note that

$$E_\theta\left[S_{11}(\zeta,\delta,\theta)\right] = \int_{-\infty}^{\infty}\left[\frac{f^{(1)}(x,\theta)}{f(x,\theta)} + \frac{\overline{G}^{(1)}(x,\theta)}{\overline{G}(x,\theta)}\right]dF^*(x,\theta) = \tag{4.4}$$

$$= \int_{-\infty}^{\infty} f^{(1)}(x,\theta)\overline{G}(x,\theta)dx + \int_{-\infty}^{\infty}\overline{G}^{(1)}(x,\theta)f(x,\theta)dx = \frac{\partial}{\partial\theta}E_\theta[\delta].$$

Similarly, we get that

$$E_\theta[S_{21}(\zeta,\delta,\theta)] = \frac{\partial}{\partial\theta}E_\theta[1-\delta]. \tag{4.5}$$

Combining the above equations, it follows that

$$E_\theta[S(\zeta,\delta,\theta)] = \frac{\partial}{\partial\theta}E_\theta[\delta] + \frac{\partial}{\partial\theta}E_\theta[1-\delta] = 0. \tag{4.6}$$

This equation was derived in Abdushukurov and Kim (1987). We include it here for completeness. Applying Corollary 3.1 with $Y = S(\zeta,\delta,\theta)$ and $X = \widehat{\phi}(\zeta,\delta) - \phi(\theta)$, it follows that

$$E_\theta\left([\widehat{\phi}(\zeta,\delta) - \phi(\theta)]^2\right) \geq \left(E_\theta[\widehat{\phi}(\zeta,\delta) - \phi(\theta)]\right)^2 + \frac{|E(XY)|^2}{E(Y^2)} \tag{4.7}$$

Observe that

$$E(XY) = E_\theta\left[\left(\widehat{\phi}(\zeta,\delta) - \phi(\theta)\right)S(\zeta,\delta,\theta)\right] = \tag{4.8}$$

$$E_\theta\left[\widehat{\phi}(\zeta,\delta)S(\zeta,\delta,\theta)\right] = \gamma^{(1)}(\theta).$$

Hence we obtain the following inequality for randomly censored data improving the inequality proved by Abdushukurov and Kim (1987).

Theorem 4.2. *Suppose the conditions* (C_1) *to* (C_5) *hold. Then, for all* $\theta \in \Theta$,

$$E_\theta\left(\left[\widehat{\phi}(\zeta,\delta) - \phi(\theta)\right]^2\right) \geq \left(E_\theta\left[\widehat{\phi}(\zeta,\delta) - \phi(\theta)\right]\right)^2 + \frac{\left[\gamma^{(1)}(\theta)\right]^2}{E_\theta[S^2(\zeta,\delta,\theta)]}. \tag{4.9}$$

We will now consider the case $k > 1$. In addition to the conditions $(C_1)-(C_5)$, suppose the following condition holds.

(C_6) Suppose the random function $S_i(\zeta,\delta,\theta)$ are linearly independent. Define the function

$$Q(\zeta,\delta,\theta) = \widehat{\phi}(\zeta,\delta) - \gamma(\theta) - \sum_{i=1}^{k} q_i S_i(\zeta,\delta,\theta)$$

where $q_i,\ 1 \leq i \leq k$ are to be chosen.

The following result was proved in Prakasa Rao (2018c) following Walker's inequality. For details, see Prakasa Rao (2018c).

Theorem 4.3. *Suppose the conditions* (C_1) *to* (C_6) *hold for all* $\theta \in \Theta$. *Then, for any estimator* $\widehat{\phi}(\zeta,\delta)$ *of the parametric function* $\phi(\theta)$ *with expectation* $\gamma(\theta)$,

$$(4.10)\quad E_\theta\left(\left[\widehat{\phi}(\zeta,\delta)-\phi(\theta)\right]^2\right)\geq[\phi(\theta)-\gamma(\theta)]^2+\sum_{i,j=1}^k\gamma^{(i)}(\theta)\gamma^{(j)}(\theta)\widetilde{I}_{ij}(\theta)$$

where

$$((\widetilde{I}_{ij}(\theta)))_{k\times k}=((I_{ij}(\theta)))^{-1}_{k\times k}.$$

This result generalizes the result of Abdushukurov and Kim (1987) from unbiased estimators to arbitrary estimators by the obvious relation.

$$E_\theta\left(\left[\widehat{\phi}(\zeta,\delta)-\phi(\theta)\right]^2\right)=\left(E_\theta\left[\widehat{\phi}(\zeta,\delta)-\phi(\theta)\right]\right)^2+Var_\theta\left(\widehat{\phi}(\zeta,\delta)\right).$$

5 Cramer-Rao type integral inequalities for randomly censored data

Suppose $\widehat{\theta}$ is the Bayes estimator of the scalar parameter $\theta\in\Theta$ given the observations $Z_i=(\zeta_i,\delta_i)\,,\,1\leq i\leq n$ with respect to the squared error loss function corresponding to the prior density $\lambda(\cdot)$ for θ. Then

$$(5.1)\qquad \widehat{\theta}=\int_\Theta\theta h_{\theta|Z}(\theta\,|z)d\theta$$

where

$$(5.2)\qquad h_{\theta|Z}(\theta\,|z)=\frac{\prod_{i=1}^n h(z_i,\theta)\lambda(\theta)}{\int_\Theta\prod_{i=1}^n h(z_i,\theta)\lambda(\theta)d\theta}$$

is the posterior density. The following result is due to Prakasa Rao (1995). For details, see Prakasa Rao (1995)

Theorem 5.1. *Suppose that the classical regularity conditions hold for the validity of Cramer-Rao inequality for the family of conditional densities $h_{\theta|Z}(\theta|z),\theta\in\Theta$ for every $z\in R\times\{0,1\}$. Then*

$$(5.3)\qquad E\left[\widehat{\theta}-\theta\right]^2\geq\frac{1}{F+nE(I(\theta))}$$

where

$$(5.4)\qquad I(\theta)=E_{Z|\theta}[\frac{d\log h(Z_i,\theta)}{d\theta}]^2$$

and

$$(5.5)\qquad F=E\left[\frac{d\log\lambda(\theta)}{d\theta}\right]^2.$$

Lower bounds for the posterior risk and lower asymptotic minimax bounds were obtained in Prakasa Rao (1995)extending the work of Weinstein and Weiss (1985), Babrovsky and Zakai (1976), Borovkov and Sakhanenko (1980) and Prakasa Rao (1991, 1992).

Acknowledgment: This work was done under the "Indian National Science Academy Senior Scientist"scheme at the CR Rao Advanced Institute of Mathematics, Statistics and Computer Science, Hyderabad 500046, India. It is a great pleasure for me to contribute this paper to the Proceedings of the Conference to be organized in honour of 60th birthday of Professor A.A. Abdushukurov.

REFERENCES

1. Abdushukurov, A.A. and Kim, L.V. (1987). Lower Cramer-Rao and Bhattacharya bounds for randomly censored data, *J. Soviet. Math.*, 2171-2185.

2. Babrovsky, B.Z. and Zakai, M. (1976). A lower bound on the estimation for certain diffusion processes, *IEEE Trans. Inform. Theory*, IT-22, 45-52.

3. Bhattacharya, A. (1946). On some analogues of the amount of information and their use in statistical estimation, *Sankhya*, 8, 1-14.

4. Borovkov, A.A. and Sakhanenko, A.I. (1980). On estimates for the average quadratic risk, *Probab. Math. Stat.*, 1, 185-195 (in Russian).

5. Crow, L.H. and Shimi, I.N. (1982). Estimation using type I censoring for a class of life time distributions, *J. Statist. Plan. Inference*, 4, 91-97.

6. Eubank, R.L. and La Riccia, V.N. (1982). Location and scale parameter estimation for randomly censored data, *Commun. Statist. A- Theory and Methods*, 11, 2869-2888.

7. Prakasa Rao, B.L.S. (1991). On Cramer-Rao type integral inequalities, *Calcutta Statistical Association Bulletin*, 40, 183-205.

8. Prakasa Rao, B.L.S. (1992). Cramer-Rao type integral inequalities for functions of multidimensional parameter, *Sankhya Series* A, 54, 53-73.

9. Prakasa Rao, B.L.S. (1995). Remarks on Cramer-Rao type integral inequalities for random censored data, In *Analysis of Censored Data*, Eds. H.L. Koul and J. V. Deshpande, Institute of Mathematical Statistics, Lecture Notes-Monograph Series No.27, pp. 163- 175.

10. Prakasa Rao, B.L.S. (2018a). Improved Cramer-Rao type integral inequalities or Bayesian Cramer-Rao lower bounds, *Journal of Indian Society for Probability and Statistics* , 19, 1-7.

11. Prakasa Rao, B.L.S. (2018b). Improved sequential Cramer-Rao type integral inequality, *Sequential Analysis*, 37, 59-68.

12. Prakasa Rao, B.L.S. (2018c). Improved Cramer-Rao inequality for randomly censored data, *Journal of Iranian Statistical Society*, 17.

13. Walker, S.G. (2017). A self-improvement to the Cauchy-Schwartz inequality, *Statist. Probab. Lett.*, 122, 86-90.

14. Weinstein, E. and Weiss, A. (1985). Lower bounds on the mean square estimation error, *Proc. of IEEE*, 73, 1433-1434.

15. Wyckoff, J. and Engelhardt, M. (1980). Cramer-Rao lower bounds for estimators based on censored data, *Commun. Statist. A- Theory Methods*, 5, 1385-1399.

Some Recent Statistical Approaches to Personalized Medicine

Somnath Datta[1], Chathura Siriwardhana[2], K.B. Kulasekera[3]

[1] Department of Biostatistics, University of Florida, Gainesville, FL 32603, USA

[2] Department of Quantitative Health Sciences, University of Hawaii John A. Burns School of Medicine, Honolulu, HI 96813, USA

[3]Department of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY 40202, USA

Summary:

We present a review of some recent results developed by our team in selecting an optimal treatment given a patient's clinical characteristic. In essence, we build a machine learning algorithm that can be trained on the data obtained from a randomized trial. This learner can be useful for allocating future patients of a similar trial to appropriate treatment arms in order to optimize their treatment outcomes. We start from the simplest case of two treatment arms and extend it to multiple treatments. Construction of a score function to summarize the patient level covariates is presented by employing a single index model. We show how different criteria, such as maximization of conditional expectations of outcomes, conditional quantiles of outcomes, and certain probability of dominance can be used in constructing the optimal treatment rule. We derive some theoretical consistency properties and present simulation and real data results for our procedures. The paper then considers the more challenging problem of selecting an optimal treatment when patient welfare (or progression of cure) is measured in more than one ways. Thus one needs to deal with optimizing multivariate responses in order to choose an optimal treatment. We show how a rank aggregation procedure offers a practical solution to this problem that has excellent performances under various scenarios. We then consider a slightly different setting of a crossover design. Once again, we show how to produce an optimal treatment selection strategy.

1. Background and Theoretical Strategies. Personalized medicine is often synonymous with optimal treatment selection. These termed have been used a lot lately in various areas of medical and pharmaceutical research. Statisticians and data scientist can play an important role in this development in various aspects of subset identification and development of patient specific treatment allocations. Mathematically and statistically these developments present an opportunity for us in finding novel research directions. Here we present a

selective summary of research our team has conducted over the past five years. Besides our own papers, some landmark statistical papers in this area include Qian and Murphy (2011), Cai et al. (2011), Zhang et al. (2012), Zhao et al. (2012), Lizotte et al. (2012), Schulte et al. (2014), Laber et al. (2014), Ertefaie et al. (2016), Fu et al. (2016), and Butler et al. (2018).

1.1 Two treatments, single response.

In the traditional clinical trials, a new treatment is compared to an old treatment. The goal was to show that the expected response for individual receiving this treatment is higher (without loss of generality, we assume throughout that larger response represents better outcome for a patient) than those receiving the old treatment. To demonstrate this, patients taking part in the trial are randomized into two treatment arms and the sample means in the two groups are compared in terms of an overall effect size. Thus if the new treatment passes this test it is supposed to be a better option for all patients. In contrast, under personalized medicine, we are willing to accept that depending on individual characteristics, the new drug may be better for certain patients, whereas the opposite is true for the remaining patients.

To proceed under these scenarios we need to think of potential or counterfactual outcomes as in the framework of causal inference. Imagine a triplet (X, Y_1^*, Y_2^*) for an individual patient, where X is a vector of covariates representing patient characteristics, and Y_j^* would be the response if the patient receives treatment j, $j = 1, 2$. Letting $\Delta = E(Y_2^*|S) - E(Y_1^*|S)$, where S is a lower dimensional mapping of X (which is often very high dimensional) representing a sufficient or practical summary of patient characteristics, treatment 2 will be optimal for this individual patient if $\Delta > 0$, otherwise it will be treatment 1.

Of course, estimating this Δ from existing trial data is multi-faceted challenge. We will outline our approach to it and the related statistical estimation problems in Section 2. We conclude this section by introducing three other related settings and our approach to optimality in each setting.

1.2. Multiple treatments, single response.

Consider a $(K + 1)$ tuple of covariate and responses under the k treatments, $(X, Y_1^*, \cdots, Y_K^*)$. The optimal treatment in this case corresponds to

$$\ell^* = \text{argma}_{\ell} \text{x} \quad E(Y_\ell^*|S).$$

Note that for $K = 2$, $\ell^* = I(\Delta > 0) + 1$ (Siriwardhana et al., 2019a).

1.3 Multiple treatments, multiple responses.

In many applications, patient recovery may be measured on multiple outcomes leading to multivariate responses $\boldsymbol{Y}_k = (Y^*_{k1,} \cdots, Y^*_{kJ})$, where J denotes the number of outcomes. As for example, in a clinical trials involving HIV patients $J = 3$, if we want to use (i) CD4 counts, (ii) CD8 counts, and (iii) survival as indicators of better treatment outcomes.

Let $\mu_{jk}(s_j) = E(Y^*_{jk}|S_j = s_j)$ be the expected response for the jth outcome under treatment k, given an appropriate covariate score, and let $\boldsymbol{\mu_k}(\boldsymbol{s}) = (\mu_{1k}(s_1), \cdots, \mu_{Jk}(s_1))^T$, $\boldsymbol{s} = (s_1, \cdots, s_J)^T$ is a realization of the vector of scores $\boldsymbol{S} = \boldsymbol{S}(X) = (S_1(X), \cdots, S_J(X))^T$. Let $\boldsymbol{v}_j(\boldsymbol{s}) = (v_{j1}(\boldsymbol{s}), \cdots, v_{j1}(\boldsymbol{s}))$, where $v_{jk}(\boldsymbol{s})$ is the rank of $\mu_{jk}(s_j)$ amongst $\{\mu_{j1}(s_j), \cdots, \mu_{jK}(s_j)\}$, from largest to the smallest. The next step is to aggregate the vector of ranks in order to find the "best" treatment overall. The aggregate rank using a procedure outlined in Pihur et al. (2007, 2009) minimizes an objective function of the form

$$\rho(\boldsymbol{v}) = \sum_{j=1}^{J} w_j \gamma(\boldsymbol{v}, \boldsymbol{v}_j(\boldsymbol{s}))$$

over all possible rankings, and where w_j are user selectable weights indicating the relative importance of the various components of the response vector in determining patients' well being. Here γ is an appropriate distance measure on the set of ranks. In Siriwardhana et al. (2019b), we used a weighted Spearman's Footrule distance advocated by Pihur et al. (2007). Finally, the optimal treatment corresponds to the one which has the smallest aggregate rank.

1.4. Crossover trials.

Unlike randomized trials, in a crossover design, the same individual receives both treatments. However, the true counterfactual response is marred by the presence of a carryover effect. Assuming there is no sequencing effect, a patient receiving treatment 1 first and treatment 2 next would have a response $(Y^*_1, Y^*_2 + \delta_{12}(X)) = (R_1, R_2)$, say, where δ_{12} represents a carryover effect which may depend on the patient covariate X. Likewise, the observed response will be of the form $(Y^*_2, Y^*_1 + \delta_{21}(X)) = (W_1, W_2)$, say, for a patient receiving treatment 2 first followed by treatment 1. Assuming that the treatment sequences are randomly assigned, we will have

$$E(R_2 - W_1|S(X)) = E(\delta_{12}|S).$$

In turn this implies $\mu_2(s) - \mu_1(s) = E(R_2|S) - E(R_1|S) - E(\delta_{12}|S)$. An hence an optimal treatment can be selected by estimating these quantities using a

fraction of the data. Alternative strategies are provided in Siriwardhana et al. (2019c) which may be more efficient.

2. Model for Estimation and Implemental Strategies.

Let Y_{jki} be the response corresponding to the jth outcome measure for ith patient receiving treatment k in the randomized trail which will form our training data. We assume we have a sample of size n_k; i.e., $1 \leq i \leq n_k$, where $n_k/(\sum_{k=1}^{K} n_k) \to c_k \in (0,1)$. We consider a single index model connecting the response and the covariate:

$$Y_{jki} = g_{jk}(\beta_{jk}^T X_{ki}) + \epsilon_{jki},\ i = 1, \cdots, n_k.$$

The unknown link functions g and the parameters β are estimated following the procedure developed in Hristache et al. (2001). For any specific value x of the covariate define the following functions

$$\widehat{U}_{jk}(x) = \widehat{g}_{jk}(\beta_{jk}^T x) - \max_{m \neq k} \{\widehat{g}_{jm}(\beta_{jm}^T x)\}\ , 1 \leq k \leq K,$$

$$\widehat{U}_j(x) = \max_k \{\widehat{U}_{jk}(x),$$

$$\widehat{\delta}_j(x) = \arg\max_k \{\widehat{U}_{jk}(x)\},$$

and

$$\widehat{S}_j(x) = (\widehat{U}_j(x),\ \widehat{\delta}_j(x)),\ 1 \leq j \leq J. \tag{1}$$

Given a new patient we covariate $X = x_0$, we find the optimal treatment for her/him following the procedure provided in the previous procedure where the set of theoretical conditional expectations $\mu_{jk}(s_{j0})$ are replaced by there estimate $\widehat{\mu}_{jk}(\widehat{s}_{j0})$. Here $\widehat{s}_{j0}$ is given by (1) with $x = x_0$, and $\widehat{\mu}$ is obtained by kernel regression

$$\widehat{\mu}_{jk}(s_j) = \frac{\sum_{i=1}^{n_k} Y_{jki} H(\frac{u_j - \widehat{U}_j(X_{ki})}{h_k}) I(\widehat{\delta}_j(X_{ki}) = d_j)}{\sum_{i=1}^{n_k} H(\frac{u_j - \widehat{U}_j(X_{ki})}{h_k}) I(\widehat{\delta}_j(X_{ki}) = d_j)}, \text{ for } s_j = (u_j, d_j),$$

where H is a kernel (e.g., density) function and h_k are bandwidths.

See Siriwardhana et al. (2016, 2019a) for the theoretical consistency of $\widehat{\mu}$ and that of the corresponding optimal treatment selector. In addition, extensive simulation studies were also conducted. In practice one could use a Gaussian

kernel and determine h by the method suggested in Wand and Jones (1995) for kernel smoothing.

3. A Numerical Illustration.

We experimented with a dataset obtained from the ACTG-175 Clinical Trial (Hammer et al., 1996) involving HIV-patients. Four antiviral treatments ($K = 4$) were compared with respect to both T cell CD4 and CD8 counts ($J = 2$). We train our optimal treatment selector on a random subsets of 200 patients which was then used to determine the optimal treatment allocation for the remaining 1336 patients. Table 1 shows the allocation.

Table 1: The optimal treatment allocations for 1336 patients.

Weights	Optimal Treatment Allocation			
$w_1 = w_{CD4}\,(w_2 = w_{CD8} = 1 - w_1)$	Arm-1	Arm-2	Arm-3	Arm-4
1	4	731	426	175
0.8	4	731	426	175
0.6	11	814	368	143
0.5	132	867	261	76

As expected, the allocations were far from being uniform. Treatment 2 (a combination of two drugs) appears 2 be the most promising options for majority of the patients whereas Treatment 1 is the least appropriate option, which was an old drug. Siriwardhana et al. (2019a) also found a way to quantify the average gain in response by using the optimal treatment compared to what they had received as part of the randomized allocation. See Siriwardhana et al. (2019b) for further details.

4. Discussion.

Optimal treatment selection within the concept of personalized medicine (van't Veer and Bernards, 2008; Vazquez, 2013) is a fertile area of research only for the drug makers (e.g., chemist, pharmacist and so on) but also for the statisticians in guiding such developments and the physicians. We present some of our contributions in this growing area of active research. We have not considered the situation where some responses are missing or censored. Reweighting is an option to extend our techniques to such situations. See Siriwardhana (2016) for some early development. Another useful extension will be to train such an optimal selector on existing data from non-randomized trials. Once again the proposed methods need to be adjusted to reflect the potential bias arising from the selection sampling underlying such data. A possible strategy with be to inversely reweight a patient's contribution by its propensity or stratification score in estimating the various components of the

optimal treatment allocation strategy. We will explore the details elsewhere.

We have motivated the optimal selection in terms of conditional means. However, in practice conditional quantiles could also be used and may be preferred for certain types of responses. We (Siriwardhana et al., 2019a, 2019b, 2019c) have developed another measure which we refer to as the "probability of dominance". In our simulation studies it seems to perform better than strategies based on comparing conditional means (details not shown).

REFERENCE

1. Butler, E. L., Laber, E. B., Davis, S. M., Kosorok, M. R. (2018). Incorporating patient preferences into estimation of optimal individualized treatment rules. Biometrics, 74, 18-26.

2. Cai, T., Tian, L., Wong, P. H. , et al. (2011). Analysis of randomized comparative clinical trial data for personalized treatment selections. Biostatistics, 12, 270-282.

3. Ertefaie, A., Wu, T., Lynch, K. G., Nahum-Shani, I. (2016). Identifying a set that contains the best dynamic treatment regimes. Biostatistics, 17, 135-148.

4. Fu, H., Zhou, J., Faries, D. E. (2016). Estimating optimal treatment regimes via subgroup identification in randomized control trials and observational studies. Stat Med, 35, 3285-3302.

5. Hristache, M., Juditsky, A., Polzehl, J., Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. Ann. Stat., 29, 1537-1566.

6. Laber, E. B., Lizotte, D. J., Ferguson, B. (2014). Set-valued dynamic treatment regimes for competing outcomes. Biometrics, 70, 53-61.

7. Lizotte, D. J., Laber, E. B. (2016). Multi-objective Markov decision processes for data-driven decision support. J. Mach. Learn. Res., 17, 1-28.

8. Pihur, V., Datta, S., Datta, S. (2007). Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. Bioinformatics, 23, 1607-1615.

9. Pihur, V., Datta, S., Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. BMC Bioinformatics, 10, 62.

10. Qian, M., Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. Ann Stat, 39, 1180-1210.

11. Schulte, P. J., Tsiatis, A. A., Laber, E. B., Davidian, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. Stat Sci, 29, 640-661.

12. Siriwardhana, C., Datta, S., Kulasekera, K. B. (2019c). Personalized treatment selection using data from crossover designs with carry over effects. Statistics in Medicine, in press

13. Siriwardhana, C., Datta, S., Kulasekera, K. B. (2019b). Personalized plans with multiple outcome measures. Journal of Biopharmaceutical Statistics, in press.

14. Siriwardhana, C., Zhao, M., Datta, S., Kulasekera, K. B. (2019a). A probability based method for selecting the optimal personalized treatment from multiple treatments. Statistical Methods in Medical Research, in press. Epub ahead of print doi: 10.1177/0962280217735701.

15. Siriwardhana, C. (2016). Semi-parametric Methods for Personalized Treatment Selection and Multistate Models. Doctoral Dissertation, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA

16. van't Veer, L. J. Bernards, R. (2008). Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature, 452, 564-570.

17. Vazquez, A. (2013). Optimization of personalized therapies for anticancer treatment. BMC Syst. Biol., 7.

18. Wand, M. P., Jones, M. C. (1995) Kernel Smoothing. Chapman & Hall London.

19. Zhang, B., Tsiatis, A. A., Laber, E. B, et al. (2012). A robust method for estimating optimal treatment regimes. Biometrics, 68, 1010-1018.

20. Zhao, Y., Donglin, Z,, Rush, A. J., et al. (2012). Estimating individualized treatment rules using outcome weighted learning. J. Am. Stat. Assoc. 107, 1106-1118.

Machine Learning Methods and Time Series Forecasting

Milovan Krnjajić
krnjajic@yahoo.com

Abstract. *Methods of machine learning (ML) have been successfully applied in diverse areas such as computer vision, speech recognition, language translation, games, and many others. While ML methods have shown superb performance in a number of classification and prediction tasks, their application in more traditional settings, such as forecasting of time series (TS), hasn't received widespread attention. There have been a few attempts to assess relative predictive accuracy of ML methods versus time series models, with growing empirical evidence in support of TS models. Presence of time-dependence in time series data imposes limitations on the development and use of ML methods, including training and estimation of their ability to generalize. Without going into a comprehensive comparison of performance, this article addresses a preceding question of principal suitability of ML methods for analysis and forecasting of time-dependent data.*

Keywords: Machine learning, time series, forecasting.

Introduction

Success in applying machine learning algorithms in a variety of applications lead to proposing ML techniques for more traditional tasks such as forecasting for econometrics or financial time series, see for example [6]. There have been some reports on the predictive performance of ML methods on time-dependent data but no widespread published studies comparing relative merits of ML methods versus time series models. However, [7] along with [2] performed comprehensive studies in comparing forecasting perfomance of ML methods with TS models. In particular, Makridakis et al. (2018) report results of comprehensive comparison of eight ML methods (such as neural nets, support vector machines, kernel regression methods) with eight traditional time series models (such as simple exponential smooting, Holt's method, ARIMA). Empirical evidence, based on forecasts on over one thousand time series used for M3-Competition, shows that time series models outperform, on average, the ML techniques, in terms of predictive accuracy. A recent free online book on time series models refers to use of neural nets for time series forecasting [5].

More comprehensive studies and empirical evidence are needed to establish a definitive answer regarding relative merits of ML vs time-series forecasting, so it may be informative to examine suitability of ML techniques beforehand. This paper addresses several issues and problems facing ML methods, which may stand in the way of their more successful application in forecasting.

Remainder of the paper is divided in two sections with main focus on (a) discusing issues stemming from the assumptions about data – how time-dependent data impose constraints on the utility of cross-validation as the main tool for training, hyper parameter optimization and estimation of the future generalization performance and (b) inherent limitations of some ML methods to extrapolate, along with the lack of intepretation of ML results and difficulties of uncertainty quantification.

ML Methods and Time Dependent Data

General examples of the application of ML methods are pattern recognition, classification and clustering. In a typical supervised learning scenario, the data are provided as input to a ML methods as a set of pairs of features and responses. The learning is based on the ability of ML method to generalize the feature – response pairs so as to recognize similar patterns in future (test) data. The mechanism of supervised learning has a firm theoretical foundation based on the work of V. Vapnik [10].

Time-dependent data and statistical learning

Vapnik used statistical framework of minimizing expected empirical loss to describe a general model of supervised learning from observed examples for cases of pattern recognition, regression and density estimation. Such a model of learning from examples includes:

- generator of i.i.d. random vectors, x, from a fixed and unknown $P(x)$.
- supervisor, returns responses y_i for inputs x_i according to $P(y|x)$.
- learning machine capable of implementing a set of functions (hypotheses) $h(x, \alpha), \alpha \in \mathcal{A}$.

The goal here is to choose such an h which best predicts the responses. Selection of h is made based on a training set of i.i.d. observation pairs $(x_1, y_1), ..., (x_N, y_N)$ from $P(x, y)$ by minimizing expected empirical loss $\mathcal{L}, computed as functional, R(\alpha) = \int \mathcal{L}(y, h(x, \alpha))dP(x, y)$, where

$\mathcal{L}(y, h(x, \alpha)) = (y - h(x, \alpha))^2$, for regression. Distribution $P(x, y)$ is unknown and all information is in the training set.

Note that it is already apparent here that the time-dependence of data points in a time series, violates basic i.i.d. assumptions of the theory. In principle, this is not necessarily a show stopper for learning from non-i.i.d. data as it is possible that despite the assumptions not being satisfied, the learning theory may hold approximately. However, it is not clear to what degree the i.i.d. requirement affects the ability of a supervised learning method to generalize so we examine this problem in practical terms.

Cross-validation with time-dependent data

When developing an ML solution it is important to acknowledge the need for a strict division of the available data in two non-overlapping parts. One part is used for training, which also includes optimization of hyper-parameters of the ML method. Training process usually proceeds by partitioning the training dataset for cross-validation (CV) in search of the optimal values for the hyperparameters of the ML method.

Now, note that the partitioning is done by random selection of training points which is alright when the data are i.i.d. However, for time-dependent sequential data, such a procedure leads to training on the future points and choosing of the hyper-parameter values based on how well the method fits the past points. This is a flagrant case of data snooping which must not be alowed, see [1]. One way to avoid this problem is to use only sequential partitioning following the time scale, but this severely limits the extent and flexibility of CV procedures, especially for smaller data sets. However, ignoring this issue leads to selecting suboptimal values for hyperparameters which in turn results in weaker generalization.

A similar problem (of data snooping) occurs for time series when estimating the future performance of the trained model on a holdout set since this also requires using the CV. The solution again is to respect the timing order and use a limited CV as in the case of training and hyper-parameter optimization. On the other hand, proceeding with applying the CV as if the data were i.i.d. will tend to unrealistically overestimate the future performance. Note that a correctly formed holdout set can contain only points which are more recent than the latest in the training set.

Format of input data for ML methods

A time series is simply a vector of values, whereas a conventional input to ML algorithm is in the form of pairs (x_i, y_i), where x_i are (d-dimensional) vectors of features and y_i are (usually, one-dimensional) responses.

In order to learn from a series of time-dependent points it's necessary to emulate the timing information by stacking the shifted time-series. Assume a TS, $y_1, y_2, ..., y_N$, of (say, monthly) observations of sales of a product. In order to train an ML algorithm on a TS of length N, allowing k past values to learn from, we need to form a $(N-k) \times n$ matrix and a $k \times 1$ response vector.

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & \cdots & y_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N-k} & y_{N-k+1} & \cdots & y_{N-1} \end{bmatrix} \rightarrow \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_N \\ \\ \end{bmatrix}$$

Note that packing of the TS into a suitable ML input increases the training set from N to $k \times (N-k)$ points without providing new information. For $k = 12$ and $N = 96$ this is about a tenfold increase.

Forecasting for time horizons larger than one (as shown above) requires new repackings of the data and separate training for each horizon. A number of other approaches to adapting the time series for ML input have been developed, some of which are described in [3], for example.

Feature selection

ML datasets are usually of high dimension which may lead to overfitting based on the spurious corelations that may exist between response and the features as well as among the features themelves. A way forward is to reduce dimension of the data, which requires applying a procedure for feature selection, see for example [4]. This is a general problem present also with i.i.d. data, but the resolution of it is more difficult when the data are time-dependent as they require additional feature selection techniques, see for example [9]. Methods of feature selection based on using CV are limited in their application to time series data for the same reasons as already noted at the beginning of the section.

Strenghts and Limitations of ML Methods

Machine learning methods implement algorithms capable of uncovering general non-linear relationships between features and response. Their flexibility in

handling almost any kind of data and ability to uncover patterns are among their main advantages over statistical models which are limited by the forms of parametric families. Next, we briefly comment on neural networks (NNs) as one of the earliest developed ML techniques since they epitomize properties of other ML approaches.

ML methods as non-linear regression

Neural nets have been successfuly used in many applications since they can approximate any reasonabe function of covariates with desired accuracy, and are not constrained to specific parametric formulation of the model. In effect, NNs can implement approximations of any unknown non-linear regression functions.

NNs work well with high-dimensional data and can capture important interactions between features which may not be easy or even possible to do using standard parametric models. More recent NN architectures (RNN, LSTM) handle sequential data and have been successfully applied in applications to speech recognition, language machine translation. They are also used for forecasting but with limited success, see [7].

Complexity and heuristic properties of NN architectures, along with difficulties in training, and the lack of interpretability of ML results are among the main disadvantages for using NNs. The vast literature on statistical machine learning describes properties of ML methods in depth, notably [8], which is one of the well known references.

Lack of extrapolation capability of some ML methods

It is important to note that some of the ML methods routinely used for classification and regression tasks, such as tree-based methods (random forests, gradient boosted machines, for example) by their construction can't extrapolate beyond observed data ranges, as described in [8]. Note that NNs have no such problem. Not being able to handle even the trend, these methods aren't directly usable for time series forecasting. Attempting to process the data and remove trend may not be useful as the information is lost and there is no parametric model to excplicitely tie together quantities of interest such as trend, seasonality and cycles.

Uncertainty Qauntification

Utility of point forecasts is enhanced by providing uncertainty bands, such as 95% prediction intervals, for example. Classical inference based on statistical

parametric models provides uncertainty estimates derived using asymptotic approximations and normality assumptions, which in case of Bayesian models is obtained by MCMC simulations from a joint posterior distribution.

The only way that the non-parametric ML methods may quantify uncertainty of the predictions is by using simulations suitable for the concrete application setting. For example, a trained model may be run on the bootstrapped versions of the holdout dataset in order to assess variability of predictions on future datasets. The simulation may also be repeated using a number of trained versions of the method, in some cross-validatory fashion, for a more comprehensive study of uncertanty. This works well for i.i.d. case, but the time-dependent data are again haunted by the same limited use of cross-validation and bootstrapping schemes as indicated in the first section.

To work well, most ML methods require large amounts of data which in turn may lead to extensive pre-processing in order to comply with input requirements for a specific ML algorithms (for example, some do not handle non-numeric or missing data). Moreover, larger data sets add to the computational burden which can be substantial, especially for ML methods that search extensively in hyper-parameter space in order to find an optimal subset of values. This holds for the i.i.d. data as well, but the situation is more difficult with time series data since standard resampling and CV schemes must be upgraded to respect time-dependence.

Conclusion

Understanding relative advantages or disadvantages of ML methods vs time series models is most important when developing a system for practical forecasting at a large scale. ML methods are appealing as having impressive performance in many applications and look as reasonable candidates to be included as a part of the forecasting solution. However, the empirical evidence so far indicates a subpar predictive accuracy of the ML methods when forecasting time series, which has motivated this brief examination of possible reasons for such performance.

The main impediment to successful use of ML methods for time series forecasting appears to stem from the very time-dependence of the data, which limits applicability of cross-validation and random sampling schemes as the main tools/procedures underlying the ML training and testing methodology. On the other hand, time series models may perform better simply because they are developed to specifically handle features of time-dependent data, which

reflects favorably on their ability to make more accurate predictions, as shown in a number of empirical studies.

REFERENCES

1. Y. Abu-Mostafa, M. Ismail, H. Li (2012) Learning From Data, AMLBook (2012)
2. A. Atiya, A. Gayar, N. El-Shishiny (2010) An Empirical Comparison of ML Models for TS Forecasting. Econ. Rev. 2010, 29(5-6)
3. G. Bontempi, S. Taieb, Y. Borgne, (2013) Machine Learning Strategies for Time Series Forecasting, Lecture Notes in Business Information Processing, Jan 2013.
4. J.Fan, Q.Shao, W.Zhou (2018) Are discoveries spurious? Ann Stats, 2018 Jun, 46(3)
5. Hyndman R, Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2
6. B. Krollner, B. Vanstone, G. Finnie, (2010) Financial TS Forecasting with ML Techniques, ESANN 2010 Proc. European Symposium on Artif. Neural Nets.
7. Makridakis S, Spiliotis E, Assimakopoulos V, (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 13(3)
8. T. Hastie, R. Tibshirani, J. Friedman (2016) ESL, Springer, 2nd ed. 2016
9. M. Tsagris, V. Lagani, I. Tsamardinos, (2018) Feature selection for high-dimensional temporal data, BMC Bioinformatics, 2018, vol 19 (17)
10. V. Vapnik, (1999) An Overview of Statistical Learning Theory, IEEE Transactions on neural networks, vol. 10, no. 5, Sep 1999

Статистика нечисловых данных как часть современной математической статистики

А.И.Орлов
Институт высоких статистических технологий и эконометрики Московского государственного технического университета им. Н.Э. Баумана
prof-orlov@mail.ru

История, современное состояние и перспективы развития математической статистики как научной, практической и учебной дисциплины заслуживают обсуждения. В книге [1] выделено четыре этапа развития этой дисциплины: до 1900 г. - описательная статистика, с 1900 по 1933 гг. - параметрическая статистика, с 1933 по 1979 гг. - непараметрическая статистика, с 1979 г. по настоящее время - статистика нечисловых данных. Названия этапам даны по наиболее ярким новым явлениям в развитии математической статистики. Согласно [2] публикации по статистике нечисловых данных составляют 63% от всех публикаций по математической статистике (в разделе "Математические методы исследования"журнала "Заводская лаборатория. Диагностика материалов"). Вместе с тем распространенные учебники по математической статистике основное внимание продолжают уделять параметрической статистике, т.е. отстают от передовых научных исследований на столетие. Этот естественный консерватизм делает необходимым обсуждение статистики нечисловых данных как части современной математической статистики.

Сорок лет развития статистики нечисловых данных

В настоящее время статистика нечисловых данных - одна из четырех основных областей математической статистики, выделенных по виду данных (наряду со статистикой чисел, многомерным статистическим анализом, статистикой временных рядов и случайных процессов). Статистика нечисловых данных делится на статистику в пространствах общей природы и разделы, посвященные конкретным типам нечисловых данных (статистика интервальных данных, статистика нечетких множеств, статистика бинарных отношений и др.). Научные результаты, полученные в рамках статистики в пространствах общей природы, могут быть использованы для конкретных видов данных (например, непараметрические оценки плотности). Следовательно, статистика в пространствах общей природы - центральная часть математической статистики, а включающая ее статистика нечисловых данных - основная область математической статистики.

Первоначально использовался термин "статистика объектов нечисловой природы". Он впервые появился в 1979 г. в нашей монографии [3] для обозначения совокупности некоторых полученных в ней научных результатов. В том же году в статье [4] нами была развернута программа построения этой новой области статистических методов, приведены первоначальные формулировки ряда основных теорем. К тридцатилетию новой области вышел и первый учебник по нечисловой статистике [5]. Здесь использован термин "нечисловая статистика". Он представляется слишком кратким, в то время как исходный термин "статистика объектов нечисловой природы слишком тяжеловесным. В настоящей статье будем называть рассматриваемую область прикладной статистики "статистикой нечисловых данных". Такое название в наилучшей степени отражает ее содержание. Все три термина (статистика объектов нечисловой природы, статистика нечисловых данных, нечисловая статистика) - синонимы.

Новая парадигма математических методов исследования

Появление и развитие статистики нечисловых данных знаменует переход к новой парадигме математической статистики [6].

Парадигма научная (от греч. paradeigma – пример, образец) – совокупность научных достижений, признаваемых всем научным сообществом в тот или иной период времени и служащих основой и образцом новых научных исследований. Понятие парадигмы получило широкое распространение после выхода в свет книги американского историка науки Т. Куна кСтруктура научных революцийњ (1962).

Математические методы исследования используются для решения практических задач с давних времен. В Ветхом Завете рассказано о весьма квалифицированно проведенной переписи военнообязанных (Четвертая книга Моисеева "Числа"). В первой половине XX в. была разработана классическая парадигма методов обработки данных, полученных в результате измерений (наблюдений, испытаний, анализов, опытов). Математические методы исследования, соответствующие классической парадигме, широко используются. Со стороны может показаться, что в этой области основное давно сделано, современные работы направлены на мелкие усовершенствования. Однако это совсем не так. Новая парадигма математических методов исследования принципиально меняет прежние представления. Она зародилась в 1980-х гг., но была развита в серии наших монографий и учебников уже в XXI в.

Типовые исходные данные в новой парадигме – объекты нечисловой природы (элементы нелинейных пространств, которые нельзя складывать и умножать на число, например, множества, бинарные отношения), а в старой – числа, конечномерные векторы, функции [6]. Ранее (в классической старой парадигме) для расчетов использовались разнообразные суммы, однако объекты нечисловой природы нельзя складывать, поэтому в новой парадигме применяется другой математический аппарат, основанный на расстояниях между объектами нечисловой природы и решении задач оптимизации.

Изменились постановки задач анализа данных. Старая парадигма исходит из идей начала XX в., когда К. Пирсон предложил использовать четырехпараметрическое семейство распределений для описания распределений реальных данных. В это семейство как частные случаи входят, в частности, подсемейства нормальных, экспоненциальных, Вейбулла-Гнеденко, гамма-распределений. Сразу было ясно, что распределения реальных данных, как правило, не входят в семейство распределений Пирсона (об этом говорил, например, академик С.Н. Бернштейн в 1927 г. в докладе на Всероссийском съезде математиков). Однако математическая теория параметрических семейств распределений (методы оценивание параметров и проверки гипотез) оказалась достаточно интересной, и именно на ней до сих пор основано преподавание во многих вузах. Итак, в старой парадигме основной подход к описанию данных - распределения из параметрических семейств, а оцениваемые величины – их параметры, в новой парадигме рассматривают произвольные распределения, а оценивают - характеристики и плотности распределений, зависимости, правила диагностики и др. Центральная часть теории – уже не статистика числовых случайных величин, а статистика в пространствах произвольной природы.

В старой парадигме источники постановок новых задач - традиции, сформировавшиеся к середине XX века, а в новой - современные потребности математического моделирования и анализа данных (XXI век), т.е. запросы практики. Конкретизируем это общее различие. В старой парадигме типовые результаты - предельные теоремы, в новой - рекомендации для конкретных значений параметров, в частности, объемов выборок (см., например, [7]). Изменилась роль информационных технологий – ранее они использовались в основном для расчета таблиц (в частности, информатика находилась вне математической статистики), теперь же они - инструменты получения выводов (имитационное моделирование, датчики псевдослучайных чисел, методы размножение выборок, в т.ч. бутстреп, и др.).

Вид постановок задач приблизился к потребностям практики – при анализе данных от отдельных задач оценивания и проверки гипотез перешли к статистическим технологиям (технологическим процессам анализа данных). Выявилась важность проблемы кстыковки алгоритмовњ - влияния выполнения предыдущих алгоритмов в технологической цепочке на условия применимости последующих алгоритмов. В старой парадигме эта проблема не рассматривалась, для новой – весьма важна.

Если в старой парадигме вопросы методологии моделирования практически не обсуждались, достаточными признавались схемы начала XX в., то в новой парадигме роль методологии (учения об организации деятельности) является основополагающей. Резко повысилась роль моделирования – от отдельных систем аксиом произошел переход к системам моделей. Сама возможность применения вероятностного подхода теперь – не кналичие повторяющегося комплекса условийњ (реликт физического определения вероятности, использовавшегося до аксиоматизации теории вероятностей А.Н. Колмогоровым в 1930-х гг.), а наличие обоснованной вероятностно-статистической модели. Если раньше данные считались полностью известными, то для новой парадигмы характерен учет свойств данных, в частности, интервальных и нечетких. Изменилось отношение к вопросам устойчивости выводов – в старой парадигме практически отсутствовал интерес к этой тематике, в новой разработана развитая теория устойчивости (робастности) выводов по отношению к допустимым отклонениям исходных данных и предпосылок моделей.

Статистика нечисловых данных развивается в соответствии с новой парадигмой математических методов исследования.

Различные виды нечисловых данных

Типичный исходный объект в прикладной статистике - это выборка, т.е. совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике элементы выборки - это числа. В многомерном статистическом анализе - вектора. А в статистике нечисловых данных элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной (линейной) структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков, в том числе результаты кодировки объектов с помощью заданного перечня категорий (градаций);

- упорядочения (ранжировки) экспертами объектов экспертизы - образцов продукции (при оценке еҷ технического уровня, качества и конкурентоспособности)), ее характеристик, заявок на проведение научных работ (при проведении конкурсов на выделение грантов) и т.п.;

- классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);

- толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ, оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки;

- результаты парных сравнений или контроля качества продукции по альтернативному признаку (кгоденњ - кбракњ), т.е. последовательности из 0 и 1;

- множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;

- слова, предложения, тексты;

- графы;

- вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности организации или анкета эксперта, в которой ответы на часть вопросов носят качественный характер, а на часть - количественный;

- ответы на вопросы экспертной, медицинской, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д.

Все средства измерения имеют погрешности. Однако до недавнего времени это очевидное обстоятельство никак не учитывалось в статистических процедурах. Только с конца 1970-х годов начала развиваться статистика интервальных данных, в которой предполагается, что исходные данные - это не числа, а интервалы. Статистику интервальных данных можно рассматривать как часть интервальной математики. Выводы в ней часто принципиально отличны от классических.

Интервальные данные можно рассматривать как частный случай нечетких множеств. Если характеристическая функция нечеткого множества равна 1 на некотором интервале и равна 0 вне этого интервала, то задание такого нечеткого множества эквивалентно заданию интервала. С ме-

тодологической точки зрения важно, что *теория нечетких множеств в определенном смысле сводится к теории случайных множеств*. Цикл соответствующих теорем приведен, например, в монографии [5].

Исторический путь статистической науки

Развитие статистических методов проанализировано в [1, 2]. Дадим здесь краткую сводку, позволяющую выявить роль статистики нечисловых данных.

К 60-м годам XX в. в нашей стране сформировалась научно-практическая дисциплина, которую называем классической математической статистикой. Новое поколение училось теории по фундаментальной монографии шведского математика Г. Крамера, написанной в военные годы и впервые изданной на русском языке в 1948 г. Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Как реакция на уход в математику выделилась новая научная дисциплина - прикладная статистика. В базовом учебнике по прикладной статистике [8] в качестве рубежа, когда это стало очевидным, указан 1981 г. С этого времени линии развития математической статистики и прикладной статистики разошлись. Первая из этих дисциплин полностью ушла в математику, перестав интересоваться практическими делами. Вторая [8] позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов, обследований.

Вполне естественно, что в прикладной статистике стали развиваться новые математические методы и модели. Необходимость их развития вытекает из потребностей конкретных прикладных исследований. Это математизированное ядро прикладной статистики назовем теоретической статистикой. Тогда под собственно прикладной статистикой следует понимать обширную промежуточную область между теоретической статистикой и применением статистических методов в конкретных областях. В нее входят, в частности, вопросы формирования вероятностно-статистических моделей и выбора конкретных методов анализа данных (т.е. методология прикладной статистики и других статистических методов), проблемы разработки и применения информационных статистических технологий, организации сбора и анализа данных, т.е. разработки статистических технологий.

Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика – часть математики, изучающая статистические структуры. Сама по себе не дает рецептов анализа статистических

данных, однако разрабатывает методы, полезные для использования в теоретической статистике.

2. Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных.

3. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных. Она включает в себя методологию статистических методов, вопросы организации выборочных исследований, разработки статистических технологий, создания и использования статистических программных продуктов.

4. Применение статистических методов в конкретных областях (в экономике и менеджменте – эконометрика, в биологии – биометрика, в химии – хемометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, истории, и т.д.).

Часто позиции 2 и 3 вместе называют прикладной статистикой. Иногда позицию 1 именуют теоретической статистикой. Эти терминологические расхождения связаны с тем, что описанное выше развитие рассматриваемой научно-прикладной области не сразу, не полностью и не всегда адекватно отражается в сознании специалистов. Так, до сих пор выпускают учебники, соответствующие старой парадигме - уровню представлений середины XX века.

Отметим, что математическая статистика, как и теоретическая с прикладной, заметно отличается от ведомственной науки органов официальной государственной статистики. ЦСУ, Госкомстат, Росстат применяли и применяют лишь проверенные временем приемы XIX века. Возможно, следовало бы от этого ведомства полностью отмежеваться и сменить название научной дисциплины, например, на "Анализ данных". В настоящее время компромиссным самоназванием является термин "статистические методы".

Во второй половине 80-х годов развернулось общественное движение, имеющее целью создание профессионального объединения статистиков. Аналогами являются британское Королевское статистическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации оказалась парализованной в результате развала СССР. Среди стран СНГ наибольшую активность в настоящее время проявляют узбекские исследователи, регулярно проводящие представительные конференции по статистике и ее применениям.

Статистика в пространствах произвольного вида

В чем принципиальная новизна статистики нечисловых данных? Для классической математической статистики характерна операция сложения. При расчете выборочных характеристик распределения (выборочное среднее арифметическое, выборочная дисперсия и др.), в регрессионном анализе и других областях этой научной дисциплины постоянно используются суммы. Математический аппарат - законы больших чисел, Центральная предельная теорема и другие теоремы - нацелены на изучение сумм. Принципиально важно, что в статистике нечисловых данных нельзя использовать операцию сложения, поскольку элементы выборки лежат в пространствах, где нет операции сложения. Методы обработки нечисловых данных основаны на принципиально ином математическом аппарате - на применении различных расстояний в пространствах объектов нечисловой природы.

Следует отметить, что в статистике нечисловых данных одна и та же математическая схема может с успехом применяться во многих прикладных областях, для анализа данных различных типов, а потому ее лучше всего формулировать и изучать в наиболее общем виде, для объектов произвольной природы.

Кратко рассмотрим несколько идей, развиваемых в статистике нечисловых данных для элементов выборок, лежащих в пространствах произвольного вида. Они нацелены на решение классических задач описания данных, оценивания, проверки гипотез - но для неклассических данных, а потому неклассическими методами.

Первой обсудим проблему определения средних величин [9]. В рамках теории измерений удается указать вид средних величин, соответствующих тем или иным шкалам измерения. Теория измерений [5], в середине XX в. рассматривавшаяся как часть математического обеспечения психологии, к настоящему времени признана общенаучной дисциплиной.

В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Теоретическое среднее определяется как решение задачи минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве

до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для получения эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке (примером является медиана Кемени [5, 9]). При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственными элементами рассматриваемого пространства, а являться некоторыми множествами таких элементов. Они могут оказаться и пустыми. Тем не менее удалось сформулировать и доказать законы больших чисел для средних величин, определенных указанным образом, т.е. установить сходимость (в специально определенном смысле) эмпирических средних к теоретическим [5, 9].

Оказалось, что методы доказательства законов больших чисел допускают существенно более широкую область применения, чем та, для которой они были разработаны. А именно, удалось изучить [5] асимптотику решений экстремальных статистических задач, к которым, как известно, сводится большинство постановок прикладной статистики. В частности, кроме законов больших чисел установлена и состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок. К настоящему времени подобные оценки изучены также и в статистике интервальных данных.

В статистике в пространствах произвольного вида большую роль играют непараметрические оценки плотности распределения вероятностей, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В статистике нечисловых данных предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в том числе в дискретных пространствах [5]. В частности, доказана их состоятельность, изучена скорость сходимости и установлен (для ядерных оценок плотности) примечательный факт совпадения наилучшей скорости сходимости в произвольном пространстве с той, которая имеет быть в классической теории для числовых случайных величин.

Дискриминантный, кластерный, регрессионный анализы в пространствах произвольной природы основаны либо на параметрической теории - и тогда применяется подход, связанный с асимптотикой решения экстремальных статистических задач - либо на непараметрической теории - и тогда используются алгоритмы на основе непараметрических оценок плотности.

Для анализа нечисловых, в частности, экспертных данных весьма важ-

ны методы классификации [5]. Интересно движение мысли в обратном направлении - наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, именно в рамках статистики объектов нечисловой природы (а не, скажем, многомерного статистического анализа). Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа). Аналогичным образом задачи многомерного шкалирования, т.е. визуализации данных, также естественно отнести к статистике объектов нечисловой природы. Важны методы оценки истинной размерности признакового пространства.

Отметим несколько конкретных научных результатов математической теории классификации. В задачах диагностики (дискриминантного анализа), как следует из леммы Неймана-Пирсона, целесообразно строить алгоритмы на основе отношения непараметрических оценок плотностей распределения вероятностей, соответствующих классам. Установлено, что наилучшим показателем качества алгоритма диагностики является прогностическая сила [5]. Устойчивость классификации относительно выбора метода кластер-анализа обосновывает вывод о реальности кластеров [5].

Для проверки гипотез в пространствах нечисловой природы могут быть использованы статистики интегрального типа, в частности, типа омега-квадрат. Отметим, что предельная теория таких статистик, построенная первоначально в классической постановке, приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида [5], поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством.

ЛИТЕРАТУРА

1. Лойко В.И., Луценко Е.В., Орлов А.И. Высокие статистические технологии и системно-когнитивное моделирование в экологии. – Краснодар: КубГАУ, 2019. – 258 с.

2. Орлов А.И. Развитие прикладной статистики и математических методов исследования в России (2006 - 2015) // Материалы республиканской научно-практической конференции "Статистика и ее применения - 2017". Под ред. проф. А.А. Абдушукурова. - Ташкент: НУУз, 2017. - С. 13-32.

3. Орлов А.И. Устойчивость в социально-экономических моделях. – М. : Наука, 1979. – 296 с.

4. Орлов А.И. Статистика объектов нечисловой природы и экспертные оценки / Экспертные оценки. Вопросы кибернетики. Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме кКибернетикаң, 1979. С. 17-33.

5. Орлов А.И. Организационно-экономическое моделирование. Часть 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 544 с.

6. Орлов А.И. Новая парадигма математической статистики // Материалы республиканской научно-практической конференции кСтатистика и еҷ применения – 2012ҳ. Под редакцией профессора А.А. Абдушукурова. – Ташкент: НУУз, 2012. – С.21-36.

7. Орлов А.И. О развитии асимптотических методов статистического контроля // Материалы республиканской научно-практической конференции "Статистика и ее применения - 2015". Под редакцией профессора А.А. Абдушукурова. - Ташкент: НУУз, 2015. - С. 30-40.

8. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.

9. Орлов А.И. Средние величины и законы больших чисел в топологических пространствах с показателями различия // Статистика и еҷ применения. Материалы республиканской научно-практической конференции (Ташкент, 17-18 октября 2013 г.). – Ташкент: Изд-во кУниверситетҳ, 2013. – С.30-37.

On a Probability of the Event: Each Cell Contains Even Number of Particles

F.A.Abdushukurov, A.N.Chuprunov
achuprunov@mail.ru

Abstract. We consider the probability $P(n,N)$ of the event: under allocation of $2n$ distingusning particles by N different cells each cell contains even number of particles and the probability $P'(n,N)$ of the event: under allocation of $2n$ non-distingusning particles by N different cells each cell contains even number of particles. For various type of the convergence $N,n\to\infty$ we study assimptotical bihavior of $P(n,N)$ and $P'(n,N)$.

Key words: allocation scheme, Gaussian random variable, Poisson limit theorem, Berry-Essen inequelity, limit theorem, local limit theorem.

Mathematical Subject Classification: 60C05, 60F05, 62G30

INTRODUCTION AND MAIN RESULTS

Let n,N be integer numbers. A homogeneous allocation scheme of n distinguishable particles by N different cells is named the random variables $\eta_1,\dots,\eta_N$, with the joint distribution defined by formula

$$\mathbf{P}\{\eta_1=k_1,\dots\eta_N=k_N\}=\frac{n!}{k_1!k_2!\cdots k_N!}\left(\frac{1}{N}\right)^n,$$

where $k_1,k_2,\dots k_N$ are nonnegative integer number such that $k_1+k_2+\cdots+k_N=n$. Many papers deal with limit theorems for allocation scheme of distinguishable particles by different cells (see (Kolcnin, Sevastianov and Chistiakov, [1]) and its references).

Denote by $P(n,N)$ the probability of the event: in allocation scheme of $2n$ distinguishable particles by N different cells each cell contains even number of particles.

Observe that

$$P(n,N)=\sum_{k_i\in\mathbf{N},\ 1\le i\le N,\ k_1+k_2+\cdots+k_N=n}\frac{(2n)!}{(2k_1)!(2k_2)!\cdots(2k_N)!}\frac{1}{N^{2n}}$$

$$=\sum_{k_i\in\mathbf{N},\ 1\le i\le N,\ k_1+k_2+\cdots+k_N=n}\frac{\frac{\alpha^{2k_1}}{(2k_1)!}\frac{\alpha^{2k_2}}{(2k_2)!}\cdots\frac{\alpha^{2k_N}}{(2k_N)!}}{\frac{(\alpha N)^{2n}}{(2n)!}}=$$

$$=\left(\frac{ch(\alpha)}{e^{\alpha}}\right)^N\sum_{k_i\in\mathbf{N},\ 1\le i\le N,\ k_1+k_2+\cdots+k_N=2n}\frac{\frac{\alpha^{2k_1}}{ch(\alpha)(2k_1)!}\frac{\alpha^{2k_2}}{ch(\alpha)(2k_2)!}\cdots\frac{\alpha^{2k_N}}{ch(\alpha)(2k_N)!}}{e^{-\alpha N}\frac{(\alpha N)^{2n}}{(2n)!}}.$$

Therefore, using Stirling formula for estimation of $(2n)!$, for $\alpha = \frac{2n}{N}$ we obtain

$$P(n,N) = (1+o(1))\sqrt{4\pi n}\left(\frac{1+e^{-2\alpha}}{2}\right)^N \mathbf{P}(\xi_1+\cdots+\xi_N = 2n), \qquad (1.1)$$

where $\xi_1, \xi_2, \ldots \xi_N$ are independent random variables with the distribution

$$\mathbf{P}(\xi_i = 2k) = \frac{\alpha^{2k}}{(2k)!ch(\alpha)}, \;\; k = 0,1,2\ldots,$$

or

$$P(n,N) = (1+o(1))\sqrt{4\pi n}\left(\frac{1+e^{-2\sqrt{\alpha_1}}}{2}\right)^N \mathbf{P}(\xi_1'+\cdots+\xi_N' = n), \qquad (1.2)$$

where $\xi_1', \xi_2', \ldots \xi_N'$ are independent random variables with the distribution

$$\mathbf{P}(\xi_i' = k) = \frac{\alpha_1^k}{(2k)!ch(\sqrt{\alpha_1})}, \quad k = 0,1,2\ldots, \quad \alpha_1 = \alpha^2.$$

Using various local limit theorems for an estimation of $\mathbf{P}(\xi_1+\cdots+\xi_N = 2n)$ or $\mathbf{P}(\xi_1'+\cdots+\xi_N' = n)$ we obtain various limit theorems for $P(n,N)$.

Observe that the expectation and the variance of ξ_i (see (2.6), (2.7)) are

$$e(\alpha) = \alpha th(\alpha), \quad \sigma^2(\alpha) = \frac{\alpha^2}{ch^2(\alpha)} + \alpha th(\alpha).$$

Using Poisson limit theorem for an estimation of the probability $\mathbf{P}(\xi_1+\cdots+\xi_N = 2n)$ we obtain the following theorem.

Theorem 1.1. *Let $n, N \to \infty$ such that the family of the numbers*

$$N\frac{\alpha^2}{2} = \lambda \qquad (1.3)$$

is bounded. Then we have

$$P(n,N) = (1+o(1))\sqrt{4\pi n}e^{-4n}\left(e^{-\lambda}\frac{\lambda^n}{n!} + O\left(\frac{1}{N}\right)\right). \qquad (1.4)$$

In the next theorem for an estimation of the probability $\mathbf{P}(\xi_1+\cdots+\xi_N = 2n)$ we use the local limit theorem from (Timashev, A.N., [3]).

Theorem 1.2. *Let $0 < \alpha' < \alpha'' < \infty$. Let $n, N \to \infty$ such that $\alpha' < \alpha < \alpha''$. Then we have*

$$P(n,N)=(1+o(1))\sqrt{\frac{1}{\frac{\alpha}{ch^2(\alpha)}+th(\alpha)}}\left(\frac{1+e^{-2\alpha}}{2}\right)^N e^{-NC(\alpha)}, \tag{1.5}$$

where

$$C(\alpha)=\frac{\alpha e^{-2\alpha}}{2(\alpha+sh(\alpha)ch(\alpha))}$$

In the next theorem for an estimation of the probability $\mathbf{P}(\xi_1+\cdots+\xi_N=2n)$ we use the local limit theorems from (Kolchin, A. V. and Kolchin, V. F., [4]).

Theorem 1.3. *Let $n,N\to\infty$ such that $\alpha\to 0$ and $N\alpha^2\to\infty$. Then we have*

$$P(n,N)=\sqrt{\frac{1}{2\alpha}}\left(\frac{1+e^{-2\alpha}}{2}\right)^N\left(e^{-NC(\alpha)}+o(1)\right). \tag{1.6}$$

In the next theorem we use a local limit theorem tor the sum $\xi_1+\cdots+\xi_N$ as $\alpha\to\infty$ (Theorem 2.2).

Theorem 1.4. *Let $n,N\to\infty$ such that $\alpha\to\infty$. Then we have*

$$P(n,N)=\left(\frac{1+e^{-2\alpha}}{2}\right)^N\left((1+\cos(2\pi nth(\alpha)))e^{-NC(\alpha)}+o(1)\right). \tag{1.7}$$

Theorem 1.5. *Let $n,N\to\infty$ such that $Ne^{-2\alpha}\to\beta$, $0\le\beta<\infty$. Then we have*

$$P(n,N)=\frac{e^{\beta}(1+\cos(2\pi nth(\alpha))+o(1))}{2^N}. \tag{1.8}$$

Theorem 1.6. *Let $n,N\to\infty$ such that $ne^{-2\alpha}\to\beta$, for some $\beta<\infty$. Then we have*

$$P(n,N)=\frac{1+\cos(2\pi\beta)+o(1)}{2^N}. \tag{1.9}$$

Allocation scneme of n non-distingushing paticles by N different cells are the random variables $\eta_1,\ldots,\eta_N$ with the joint distribution

$$\mathbf{P}\{\eta_1=k_1,\ldots\eta_N=k_N\}=\frac{1}{C_{n+N-1}^{N-1}},$$

where $k_1, k_2, \dots k_N$ are nonnegative integer numbers such that $k_1 + k_2 + \cdots + k_N = n$.

Denote by $P'(n, N)$ the probability of the event: under allocation of $2n$ nondistingusning particles by N different cells each cell contains even number of particles. We have

$$P'(n, N) = \sum_{k_i \in \mathbf{N},\ 1 \le i \le N,\ k_1 + k_2 + \cdots + k_N = n} \frac{1}{C_{2n+N-1}^{N-1}} = \frac{C_{n+N-1}^{N-1}}{C_{2n+N-1}^{N-1}}.$$

Theorem 1.7. *Let $n, N \to \infty$ such that $\frac{N^2}{4n} \to \beta$, where $0 \le \beta < \infty$. Then we have*

$$P'(n, N) = (1 + o(1))\frac{e^{\beta}}{2^{N-1}}. \tag{1.10}$$

Rmark 1. Consider N independent allocation schemes of $2n$ nondistingusning particles by 2 different cells. Let A_i, $1 \le i \le N$, be the event: 1th cell cotaines even number of particles from ith scheme. Then $\cap_{i=1}^{N} A_i$, $1 \le i \le N$, is the event: 1th cell cotaines even number of particles from ith scheme for all $1 \le i \le N$. Denote $P^*(n, N) = \mathbf{P}\left(\cap_{i=1}^{N} A_i\right)$. We have

$$P^*(n, N) = \left(\frac{n+1}{2n+1}\right)^N = \frac{1}{2^N}\left(1 + \frac{1}{2n+1}\right)^N = \frac{1}{2^N} e^{\frac{N}{2n}(1+o(1))}, \quad \text{as} \quad n \to \infty.$$

LOCAL LIMIT THEOREMS

We will prove local limit theorems for sums $S_N = \xi_1 + \cdots + \xi_N$ which we will use.

Theorem 2.1. *Let $n, N \to \infty$ such that the famely of numbers*

$$N\frac{\alpha^2}{2} = \lambda,$$

is bounded. Then we have

$$\mathbf{P}(\xi_1 + \cdots + \xi_N = 2n) = e^{-\lambda}\frac{\lambda^n}{n!} + O\left(\frac{1}{N}\right). \tag{2.1}$$

Proof. We have

$$\mathbf{P}(\xi_1 + \cdots + \xi_N = 2n) =$$

$$= \mathbf{P}(\xi_1 + \cdots + \xi_N = 2n, \max_{1 \le i \le N} \xi_i \le 2) + \mathbf{P}(\xi_1 + \cdots + \xi_N = 2n, \max_{1 \le i \le N} \xi_i \ge 4) \tag{2.2}$$

Since

$$\mathbf{P}(\xi_1+\cdots+\xi_N=2n,\max_{1\le i\le N}\xi_i\ge 4)\le \mathbf{P}(\max_{1\le i\le N}\xi_i\ge 4)\le \sum_{i=1}^{N}\mathbf{P}(\xi_i\ge 4)=$$

$$=N\sum_{k=2}^{\infty}\frac{\alpha^{2k}}{(2k)!ch(\alpha)}=N\frac{\alpha^4}{4}\left(\frac{1}{6ch(\alpha)}+o(1)\right)=\frac{\lambda^2}{N}\left(\frac{1}{6ch(\alpha)}+o(1)\right)=O\left(\frac{1}{N}\right),$$

it holds

$$\mathbf{P}(\xi_1+\cdots+\xi_N=2n,\max_{1\le i\le N}\xi_i\ge 4)=O\left(\frac{1}{N}\right). \tag{2.3}$$

From Theorem 4.2 on p. 178 in [2] one has

$$\mathbf{P}(\xi_1+\cdots+\xi_N=2n,\max_{1\le i\le N}\xi_i\le 2)=\mathbf{P}(I_{\{\xi_1=2\}}+\cdots+I_{\{\xi_N=2\}}=n)=$$

$$=e^{-\lambda}\frac{\lambda^n}{n!}+O\left(N\left(\frac{\alpha^2}{2ch(\alpha)}\right)^2\right)=e^{-\lambda}\frac{\lambda^n}{n!}+O\left(\frac{1}{N}\right). \tag{2.4}$$

From (2.2), (2.3) and (2.4) it follows (2.1). □

We will denote: $\xrightarrow{d}$ is a convergence in distribution, γ is a gaussian random variable with the variance one and the expectation zero, Φ is a distribution function of γ, $S_N=\xi_1+\cdots+\xi_N$. Then $N\sigma^2(\alpha)$ is a variance of S_N.

Lemma 1. *Let $n,N\to\infty$ such that $\alpha\to\infty$. Then we have*

$$\frac{S_N-Ne(\alpha)}{\sqrt{N}\sigma(\alpha)}\xrightarrow{d}\gamma. \tag{2.5}$$

Proof. Observe that

$$\mathbf{E}\xi_i^{[k]}=\mathbf{E}\xi_i(\xi_i-1)\dots(\xi_i-k+1)=\frac{\alpha^k(ch(\alpha))^{(k)}}{ch(\alpha)}.$$

Therefore one has

$$\mathbf{E}\xi_i=\mathbf{E}\xi_i^{[1]}=\frac{\alpha sh(\alpha)}{ch(\alpha)},\ \ \mathbf{E}\xi_i^2=\mathbf{E}\xi_i^{[2]}+\mathbf{E}\xi_i^{[1]}=\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)},$$

$$\mathbf{E}\xi_i^3=\mathbf{E}\xi_i^{[3]}+3\mathbf{E}\xi_i^{[2]}+\mathbf{E}\xi_i^{[1]}=\frac{\alpha^3 sh(\alpha)}{ch(\alpha)}+3\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)},$$

$$\mathbf{E}\xi_i^4=\mathbf{E}\xi_i^{[4]}+6\mathbf{E}\xi_i^{[3]}+7\mathbf{E}\xi_i^{[2]}+\mathbf{E}\xi_i^{[1]}=\alpha^4+6\frac{\alpha^3 sh(\alpha)}{ch(\alpha)}+7\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}. \tag{2.6}$$

Cosequently, the variance of ξ_i is equel

$$\sigma^2(\alpha)=\mathbf{E}\xi_i^{[2]}+\mathbf{E}\xi_i^{[1]}-(\mathbf{E}\xi_i^{[1]})^2=\alpha^2+\alpha th(\alpha)-(\alpha th(\alpha))^2=\frac{\alpha^2}{ch^2(\alpha)}+\alpha th(\alpha). \tag{2.7}$$

Using (2.6) we have

$$\mathbf{E}(\xi_i-\mathbf{E}\xi_i)^4=\mathbf{E}\xi_i^4-4\mathbf{E}(\xi_i^3)(\mathbf{E}\xi_i)+6(\mathbf{E}\xi_i^2)(\mathbf{E}\xi_i)^2-3(\mathbf{E}\xi_i)^4=$$

$$=\alpha^4+6\frac{\alpha^3 sh(\alpha)}{ch(\alpha)}+7\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}-$$

$$-4\left(\frac{\alpha^3 sh(\alpha)}{ch(\alpha)}+3\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}\right)\left(\frac{\alpha(sh(\alpha))}{ch(\alpha)}\right)+6\left(\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}\right)\left(\frac{\alpha(sh(\alpha))}{ch(\alpha)}\right)^2-$$

$$-3\left(\frac{\alpha(sh(\alpha))}{ch(\alpha)}\right)^4=\alpha^4\left(1-4\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2+6\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2-3\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^4\right)+$$

$$+\alpha^3\left(6\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)-12\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)+6\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^3\right)+\alpha^2\left(7-4\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2\right)+$$

$$+\frac{\alpha sh(\alpha)}{ch(\alpha)}=\alpha^4\left(1-\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2\right)\left(1+3\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2\right)+$$

$$+6\alpha^3\frac{(sh(\alpha))}{ch(\alpha)}\left(\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2-1\right)+4\alpha^2\left(1-\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2\right)+3\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}=$$

$$=\frac{\alpha^4}{ch^2(\alpha)}\left(1+3\left(\frac{(sh(\alpha))}{ch(\alpha)}\right)^2\right)-6\frac{(sh(\alpha))}{ch(\alpha)}\frac{\alpha^3}{ch^2(\alpha)}+4\frac{\alpha^2}{ch^2(\alpha)}+3\alpha^2+\frac{\alpha sh(\alpha)}{ch(\alpha)}. \tag{2.8}$$

Sibce

$$\frac{\alpha^k}{ch^2(\alpha)}\to 0,\ \ th(\alpha)\to 1,\quad \text{as}\ \ \alpha\to\infty,$$

for all fixed $k\in\mathbf{N}$, then from (2.8) it follows that

$$\mathbf{E}(\xi_i-\mathbf{E}\xi_i)^4=3\alpha^2(1+o(1)),\quad \text{as}\ \ \alpha\to\infty. \tag{2.9}$$

Morother from (2.7) it follows

$$\sigma^2(\alpha)=\alpha(1+o(1)),\quad \text{as}\ \ \alpha\to\infty. \tag{2.10}$$

Using (2.9) and (2.10) for the estimation of left hand Berry-Essen type inequality we obtain

$$\sup_{t\in\mathbf{R}}\left|\P\left(\frac{S_N-Ne(\alpha)}{\sqrt{N}\sigma(\alpha)}<t\right)-\Phi(t)\right|<2c\left(\frac{N\mathbf{E}(\xi_i-\mathbf{E}\xi_i)^4}{N^2\sigma^4(\alpha)}\right)^{\frac{1}{2}}=2c\left(\frac{3(1+o(1))}{N}\right)^{\frac{1}{2}}, \tag{2.11}$$

for $\alpha\to\infty$. Here c is a constant in Berry-Essen inequality From (2.11) it follows (2.5). $\square$

We will denote: $\phi(t)=\frac{ch(\alpha e^{it})}{ch(\alpha)}$ is a characteristic function of ξ_i, $\phi^c(t)=\phi(t)e^{-ite(\alpha)}$, $\phi_N(t)$ is a characteristic function of $\frac{S_N-Ne(\alpha)}{\sqrt{N}\sigma(\alpha)}$.

Theorem 2.2. *Let $n,N\to\infty$ such that $\alpha\to\infty$. Then we have*

$$\sqrt{N}\sigma(\alpha)\mathbf{P}\{S_N=2n\}-\frac{1+\cos(2\pi nth(\alpha))}{\sqrt{2\pi}}e^{-\frac{(2n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}}\to 0. \tag{2.12}$$

uniformly by $n=0,1,2,\dots$.

Proof. Denote

$$z=\frac{2n-Ne(\alpha)}{N\sigma^2(\alpha)}.$$

Such as in Gnedenko local limit theorem we represent the probability $\mathbf{P}\{S_N=2n\}$ as the integral

$$\mathbf{P}\{S_N=2n\}=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-it\sqrt{N}\sigma(\alpha)z}(\phi^c(t))^N dt.$$

After the replacement $x=t\sigma(\alpha)\sqrt{N}$ we obtain

$$\sigma(\alpha)\sqrt{N}\mathbf{P}\{S_N=2n\}=\frac{1}{2\pi}\int_{-\pi\sigma(\alpha)\sqrt{N}}^{\pi\sigma(\alpha)\sqrt{N}}e^{-ixz}\left(\phi^c\left(\frac{x}{\sigma(\alpha)\sqrt{N}}\right)\right)^N dx.$$

Let $0<\varepsilon<1$, $B>0$. Since

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ixz}e^{-\frac{x^2}{2}}dx,$$

we have

$$R_N=2\pi\left(\sigma_N\mathbf{P}\{S_N=n\}-\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right)=I_1+I_2+I_3+I_4+I_5, \tag{2.13}$$

where

$$I_1=\int_{|x|<B}e^{-ixz}\phi_N(x)dx-\int_{|x|<B}e^{-ixz}e^{-\frac{x^2}{2}}dx,$$

$$I_2 = -\int_{|x|>B} e^{-ixz} e^{-\frac{x^2}{2}}\,dx,$$

$$I_3 = \int_{B<|x|\le\varepsilon\sqrt{N}\sigma(\alpha)} e^{-ixz}\left(\phi^c\left(\frac{x}{\sqrt{N}\sigma(\alpha)}\right)\right)^N dt+$$

$$+\int_{(\pi-\varepsilon)\sqrt{N}\sigma(\alpha)<|x|\le\pi\sqrt{N}\sigma(\alpha)-B} e^{-ixz}\left(\phi^c\left(\frac{x}{\sqrt{N}\sigma(\alpha)}\right)\right)^N dt,$$

$$I_4 = \int_{\varepsilon\sqrt{N}\sigma(\alpha)<|x|\le(\pi-\varepsilon)\sqrt{N}\sigma(\alpha)} e^{-ixz}\left(\phi^c\left(\frac{x}{\sqrt{N}\sigma(\alpha)}\right)\right)^N dx,$$

$$I_5 = \int_{\pi\sqrt{N}\sigma(\alpha)-B<|x|<\pi\sqrt{N}\sigma(\alpha)} e^{-ixz}\left(\phi^c\left(\frac{x}{\sqrt{N}\sigma(\alpha)}\right)\right)^N dx$$

Observe that

$$\phi_N(x) = \phi_N(\pi - x), \quad x \in \mathbf{R}. \tag{2.14}$$

Therefore

$$I_1 = \int_{|x|<B} e^{-ixz}\left(\phi_N(x) - e^{-\frac{x^2}{2}}\right)dx$$

and by Lemma 1 for all fixed $B > 0$ one has

$$I_1 \to 0. \tag{2.15}$$

Since

$$|I_2| < \int_{|x|>B} e^{-\frac{x^2}{2}}\,dx,$$

then we have

$$I_2 \to 0, \quad \text{as} \quad B \to \infty. \tag{2.16}$$

Since

$$\cos(x) - 1 \le -\frac{x^2}{2} + \frac{x^4}{24} \le -\frac{11x^2}{24} \quad |x| \le 1,, \quad e^x - 1 \le xe^x, \quad x \ge 0,$$

by (2.14) we obtain

$$|I_3| \le 2\int_{B<|x|\le\varepsilon\sqrt{N}\sigma(\alpha)} \left|e^{-ixz}\left(\phi^c\left(\frac{x}{\sqrt{N}\sigma(\alpha)}\right)\right)^N\right| dt \le$$

$$\le 2\int_{B<|t|\le\varepsilon\sqrt{N}\sigma(\alpha)} \left(e^{\alpha\left(\cos\frac{t}{\sqrt{N}\sigma(\alpha)}-1\right)}\right)^N \left|\frac{1+e^{-2\alpha e^{i\frac{t}{\sqrt{N}\sigma(\alpha)}}}}{1+e^{-2\alpha}}\right|^N dt \le$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\alpha N\frac{11\cdot 16t^2}{24N\alpha(1+o(1))}}\right)\left(\frac{1+e^{-2\alpha\cos\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)}}{1+e^{-2\alpha}}\right)^N dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\left(1+\frac{e^{-2\alpha\cos\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)}-e^{-2\alpha}}{1+e^{-2\alpha}}\right)^N dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\exp\left(N\ln\left(1+\frac{e^{-2\alpha\cos\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)}-e^{-2\alpha}}{1+e^{-2\alpha}}\right)\right)dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\exp\left(Ne^{-2\alpha}\frac{e^{2\alpha\left(1-\cos\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)\right)}-1}{1+e^{-2\alpha}}\right)dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\exp\left(Ne^{-2\alpha}e\frac{2\alpha\left(1-\cos\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)\right)}{1+e^{-2\alpha}}\right)dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\exp\left(Ne^{-2\alpha}e\frac{2\alpha\frac{1}{2}\left(\frac{t}{\sqrt{N}\sigma(\alpha)}\right)^2}{1+e^{-2\alpha}}\right)dt\leq$$

$$\leq 2\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)\exp\left(e\frac{4\alpha e^{-2\alpha}}{(1+o(1))(1+e^{-2\alpha})}\right)dt\leq$$

$$\leq 2(1+o(1))\int_{B<|t|\leq\varepsilon\sqrt{N}\sigma(\alpha)}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)dt\leq 2(1+o(1))\int_{B<|t|}\left(e^{-\frac{22t^2}{3(1+o(1))}}\right)dt.$$

Therefore we have

$$|I_3|\to 0,\quad \text{as}\quad B\to\infty. \tag{2.17}$$

Since

$$|\phi_N(x)|\leq\frac{ch(\alpha\cos(\varepsilon))}{ch(\alpha)},\quad \varepsilon\leq x\leq\pi-\varepsilon,$$

one has

$$|I_4|\leq 2\pi\sqrt{N}\sigma(\alpha)\left(\frac{ch(\alpha\cos(\varepsilon))}{ch(\alpha)}\right)^N\leq(1+o(1))\pi\sqrt{n}\left(\frac{ch(\alpha\cos(\varepsilon))}{ch(\alpha)}\right)^N.$$

Since

$$\left(\frac{ch(\alpha\cos(\varepsilon))}{ch(\alpha)}\right)^N=e^{(\cos(\varepsilon)-1)\alpha N}\left(\frac{1+e^{-2\alpha\cos(\varepsilon)}}{1+e^{-2\alpha}}\right)^N=$$

$$= e^{(\cos(\varepsilon)-1)n}\left(1+\frac{e^{-2\alpha\cos(\varepsilon)}-e^{-2\alpha}}{1+e^{-2\alpha}}\right)^N \le e^{(\cos(\varepsilon)-1)n}\exp\left(N\frac{e^{-2\alpha\cos(\varepsilon)}-e^{-2\alpha}}{1+e^{-2\alpha}}\right)\le$$

$$\le e^{(\cos(\varepsilon)-1)n}\exp\left(Ne^{-2\alpha}\frac{e^{-2\alpha(1-\cos(\varepsilon))}-1}{1+e^{-2\alpha}}\right)\le e^{(\cos(\varepsilon)-1)n}\exp\left(Ne^{-2\alpha}\frac{e^{2\alpha(1-\cos(\varepsilon))}-1}{1+e^{-2\alpha}}\right)\le$$

$$\le e^{(\cos(\varepsilon)-1)n}\exp\left(Ne^{-2\alpha}\frac{e^{2\alpha(1-\cos(\varepsilon))}2\alpha(1-\cos(\varepsilon))}{1+e^{-2\alpha}}\right)\le$$

$$\le e^{(\cos(\varepsilon)-1)n}\exp\left(4ne^{-2\alpha\cos(\varepsilon)}(1-\cos(\varepsilon))\right)\le q^n,\ \ \alpha>2(2\cos(\varepsilon))^{-1},$$

where

$$q=\exp\left(\cos(\varepsilon)-1)(1-4e^{-2})\right),$$

we obtain

$$|I_4|\le(1+o(1))\pi\sqrt{n}q^n.$$

Therefore it holds

$$I_4\to 0. \tag{2.18}$$

Using (2.14) and replacement of the variable $\frac{u}{\sqrt{N}\sigma(\alpha)}=\pi-\frac{x}{\sqrt{N}\sigma(\alpha)}$ we obtain

$$I_5=\int_{|u|<B}e^{-ixz}\left(\phi\left(-\frac{u}{\sqrt{N}\sigma(\alpha)}\right)e^{-ie(\alpha)\left(\pi-\frac{u}{\sqrt{N}\sigma(\alpha)}\right)}\right)^N du=$$

$$=\int_{|u|<B}e^{-ixz}e^{-ie(\alpha)\pi N}\phi_N(-u)du=\int_{|u|<B}e^{-ixz}e^{-ie(\alpha)\pi N}\left(\phi_N(-u)-e^{-\frac{u^2}{2}}\right)du+$$

$$+\int_{-\infty}^{\infty}e^{-ixz}e^{-ie(\alpha)\pi N}e^{-\frac{u^2}{2}}du-\int_{|u|\ge B}e^{-ixz}e^{-ie(\alpha)\pi N}e^{-\frac{u^2}{2}}du=e^{-\frac{z^2}{2}-ie(\alpha)\pi N}+o(1)=$$

$$=e^{-\frac{z^2}{2}}e^{-2\pi nith(\alpha)}+o(1). \tag{2.19}$$

Using (2.15), (2.16), (2.17), (2.18) and (2.19) for the estimation of summands in right hand of (2.13), we obtain

$$\sqrt{N}\sigma(\alpha)\mathbf{P}\{S_N=2n\}-\frac{1+e^{-2\pi nith(\alpha)}}{\sqrt{2\pi}}e^{-\frac{(2n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}}\to 0.$$

uniformly by $n=0,1,2,\dots$. Therefore

$$Re\left(\sqrt{N}\sigma(\alpha)\mathbf{P}\{S_N=2n\}-\frac{1+e^{-2\pi nith(\alpha)}}{\sqrt{2\pi}}e^{-\frac{(2n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}}\right)=$$

$$=\sqrt{N}\sigma(\alpha)\mathbf{P}\{S_N=2n\}-\frac{1+\cos(2\pi nth(\alpha))}{\sqrt{2\pi}}e^{-\frac{(2n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}}\to 0.$$

This follows (2.12). □

PROOFS OF THEOREMS

Proof of Theorem 1.1. Observe that (1.3) follows $\alpha \to 0$. Therefore we have

$$\left(\frac{1+e^{-2\alpha}}{2}\right)^N = \left(1+\frac{e^{-2\alpha}-1}{2}\right)^N = \exp\left(N\ln\left(1+(-2\alpha+2\alpha^2+O(\alpha^3))\right)\right) =$$

$$= \exp\left(N\left((-2\alpha+2\alpha^2+O(\alpha^3)) - \frac{1}{2}(-2\alpha+2\alpha^2+O(\alpha^3))^2 + O(\alpha^3)\right)\right) =$$

$$= \exp\left(-2N\alpha + o(1)\right) = \exp\left(-4n\right)(1+o(1)). \tag{3.1}$$

Using (3.1) and (2.1) in (1.1), we obtain (1.4). □

Proof of Theorem 1.2. Observe that ξ_i' sutisfies the condition: $\mathbf{P}(\xi_i' = k) > 0$, $k = 0, 1, 2 \ldots$. Therefore by Theorem 2.2.7 from [3]. we have

$$\sqrt{2\pi N}\sigma(\alpha)\mathbf{P}(\xi_1' + \cdots + \xi_N' = n) = \exp\left(-\frac{(n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}\right)(1+o(1))$$

and

$$\frac{(n-Ne(\alpha))^2}{2N\sigma^2(\alpha)} = NC(\alpha). \tag{3.2}$$

Using (3.2) and (2.7) in (1.2), we obtain (1.5). □

Proof of Theorem 1.3. Since ξ_i' sutisfies the condition (A_0) $\mathbf{P}(\xi_i' = 0) > 0$, $\mathbf{P}(\xi_i' = 1) > 0$ and $N\alpha_1 \to \infty$ we can use Theorem 5 from [4]. By Theorem 5 from [4] it holds

$$\sqrt{2\pi N}\sigma(\alpha)\mathbf{P}(\xi_1' + \cdots + \xi_N' = n) = \exp\left(-\frac{(n-Ne(\alpha))^2}{2N\sigma^2(\alpha)}\right) + o(1). \tag{3.3}$$

Observe that as $\alpha \to 0$ we have

$$ch(\alpha) = 1 + o(1), \quad th(\alpha) = \alpha(1+o(1)). \tag{3.4}$$

Using (3.2), (3.3), (3.4) in (1.1), we obtain (1.6). □

Proof of Theorem 1.4. Since $\alpha \to \infty$ we have

$$\frac{\alpha}{ch(\alpha)} = o(1), \quad th(\alpha) = 1 + o(1)). \tag{3.5}$$

Using (3.2), (3.5) and (2.12) in (1.1), we obtain (1.7). □

PROOF OF THEOREM 1.5. Since $Ne^{-2\alpha} \to \beta$, $0 \le \beta < \infty$, we have

$$\left(1+e^{-2\alpha}\right)^N \to e^{\beta}, \;\; NC(\alpha) \to 0.$$

Therefore Theorem 1.4 follows Theorem 1.5. □

PROOF OF THEOREM 1.6. We have

$$\alpha \to \infty, \;\; Ne^{-2\alpha} = n\frac{e^{-2\alpha}}{\alpha} \to 0,$$

$$\frac{(n-Ne(\alpha))^2}{2N\sigma^2(\alpha)} = (1+o(1))2ne^{-4\alpha} \to 0$$

and

$$\cos(2\pi nth(\alpha)) = \cos\left(2\pi n\left(1+\frac{e^{-2\alpha}}{1+e^{-2\alpha}}\right)\right) = \cos\left(2\pi ne^{-2\alpha}(1+o(1))\right) =$$

$$= \cos(2\pi\beta) + o(1).$$

Therefore Theorem 1.5 follows Theorem 1.6. □

PROOF OF THEOREM 1.7. Using Stirling formula we obtain

$$P'(n,N) = \frac{C_{n+N-1}^{N-1}}{C_{2n+N-1}^{N-1}} = \frac{(n+N-1)!}{(2n+N-1)!}\frac{(2n)!}{n!} =$$

$$= \frac{\sqrt{n+N-1}(n+N-1)^{n+N-1}\sqrt{2n}(2n)^{2n}}{\sqrt{2n+N-1}(2n+N-1)^{2n+N-1}\sqrt{n}n^n}(1+o(1)) =$$

$$= \frac{(n+N-1)^{n+N-1}(2n)^{2n+N-1}}{(2n+N-1)^{2n+N-1}2^{N-1}n^{n+N-1}}(1+o(1)) =$$

$$= \frac{1}{2^{N-1}}\frac{\exp\left((n+N-1)\ln\left(1+\frac{N-1}{n}\right)\right)}{\exp\left((2n+N-1)\ln\left(1+\frac{N-1}{2n}\right)\right)}(1+o(1)) =$$

$$= \frac{1}{2^{N-1}}\frac{\exp\left(n\left(1+\frac{N-1}{n}\right)\left(\frac{N-1}{n}-\frac{1}{2}\left(\frac{N-1}{n}\right)^2+O\left(\left(\frac{N-1}{n}\right)^3\right)\right)\right)}{\exp\left(2n\left(1+\frac{N-1}{2n}\right)\left(\frac{N-1}{2n}-\frac{1}{2}\left(\frac{N-1}{2n}\right)^2+O\left(\left(\frac{N-1}{2n}\right)^3\right)\right)\right)}(1+o(1)) =$$

$$= \frac{1}{2^{N-1}}\exp\left(\frac{\frac{1}{4}(N-1)^2}{n}+nO\left(\left(\frac{N-1}{n}\right)^3\right)\right)(1+o(1)) = \frac{e^{\beta}}{2^{N-1}}(1+o(1)).$$

So (1.10) is proved. □

Authors thank professor I.R. Kaymov for an important remark.

REFERENCE

1. Kolchin V. F., Sevast'ynov B. P., Chistiakov V. P. *Random allocation.* – Scripta Series in Mathematics. V. H. Winston& Sons, Washington, DC, 1978.

2. Borovkov, A. A. *Probability theory.* Universitext. Springer, London, 2013.

3. Timashev, A. N. *Asimptotic expansions in probabilistic combinatorics.* – M.: TVP. 2011.

4. Kolchin, A. V. and Kolchin, V. F. On transition of distributions of sums of independent identically distributed random variables from one lattice to another in the generalised allocation scheme. *Discrete Math. Appl.*, **16**, No. 6, 527–540 (2006); translation from *Diskret. Mat.*, **18**, No. 4, 113–127 (2006).

On the distribution of sign test for symmetry.

Koldanov Petr Alexandrovich, Koldanov Alexander Petrovich
NRU Higher School of Economics
pkoldanov@hse.ru

0.1 Introduction

The [1] gives the asymptotic distribution of modified sign test statistic for symmetry hypothesis testing with respect to uknown shift parameter. Such test is based on the statistic

$$T_n = \sum_{i=1}^{n} I(X(i) - \overline{X} > 0) \tag{1}$$

where $X_1, \ldots, X_n$ is a sample from distribution with density $f(x)$ and finite variance, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. It was shown that for $n \to \infty$ the distribution of statistic (1) converges to normal distribution with variance which depend from $f(x)$ for the case of truth the symmetry hypothesis. Also it was shown that asymptotic distribution of the statistic (1) differ from asymptotic distribution of statistic

$$T = \sum_{i=1}^{n} I(X(i) - \mu > 0) \tag{2}$$

with known shift parameter μ. It follows that statistical test based on statistic (1) has not distribution-free property unlike usual sign test based on statistic (2). It is obvious for any sample size n distribution of statistic (2) is binomial. The question of statistic (1) distribution investigation under finite sample size for different functions $f(x)$ is an open. The aim of the paper is to find the exact distribution of statistic (1) for the case of the sample of size n from normal distribution $N(\mu, \sigma^2)$. The exact distribution of statistic (1) is obtained for the sample sizes $n = 3, 4, 5$. For the sample sizes $n = 6, 7$ the distribution of statistic (1) is obtained by numerical calculations of the orthant probabilities from one side and by simulations from the other side. Moreover it is shown that obtained distribution is close to the asymptotic distribution from [1].

0.2 Distribution of statistic

Asymptotic distribution. In [1] it is proved that for sample $X(1), \ldots, X(n)$ from continuous distribution $F(x)$ with density $f(x)$, expectation μ and

variance σ^2 under $n \to \infty$ the distribution of the $\frac{T_n - np}{\sqrt{n}}$ is normal with expectation $E\left(\frac{T_n-np}{\sqrt{n}}\right) = 0$ and variance $\sigma_1^2 = D\left(\frac{T_n-np}{\sqrt{n}}\right) = pq + \sigma^2 f^2(\mu) + 2f(\mu)\left[\int_{-\infty}^{\mu} x dF(x) - p\mu\right]$ where $p = F(\mu), \quad q = 1 - p$. In the considered case

$$p = q = \frac{1}{2}, \quad D\left(\frac{T_n - n/2}{\sqrt{n}}\right) = \sigma_1^2 = \frac{1}{4} - \frac{1}{2\pi}$$

Exact distribution. Let

$$X(1), X(2), \ldots, X(n)$$

be a sample from continuous distribution with density $f(x)$.

Let

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$

be the corresponding order statistics.

It is easy to see that

$$T_n = 1 + \sum_{i=2}^{n-1} I(X_{(i)} > \overline{X}) \tag{3}$$

Since $X(1), X(2), \ldots, X(n)$ is independent identically distributed random variables then

$$P(T_n = k) = P(X_{(n-k)} < \overline{X} < X_{(n-k+1)}) = C_n^k p_{0\ldots01\ldots1} \tag{4}$$

where

$$p_{0\ldots01\ldots1} = P(X(1) < \overline{X}, X(2) < \overline{X}, \ldots, X(n-k) < \\ < \overline{X}, X(n-k+1) > \overline{X}, \ldots, X(n) > \overline{X})$$

Case $n = 3$**.** Define random variables $X(1) - \overline{X}, X(2) - \overline{X}, X(3) - \overline{X}$. It is obvious from (3)

$$T_3 = \begin{cases} 1, & P(T = 1) = p_1 \\ 2, & P(T = 2) = p_2 \end{cases}$$

Then $p_1 + p_2 = 1$. Since under the assumption $P(X(i) > \overline{X}) = \frac{1}{2}$ one has $ET_3 = \frac{3}{2}$ then $p_1 + 2p_2 = \frac{3}{2}$ or $p_1 = p_2 = \frac{1}{2}$. From [1] one has $p_1 = p_2 = 0.4830663$.

Note 1. *Let us consider the sample of size n from normal distribution $N(\mu, \sigma^2)$ and let*

$$p(i_1, i_2, \ldots, i_n) = P(i_1(X(1) - \overline{X}) > 0, i_2(X(2) - \overline{X}) > 0, \ldots, i_n(X(n) - \overline{X}) > 0)$$

where $i_k \in \{-1, 1\}, k = 1, \dots, n.$

One has

$$p(i_1, i_2, \dots, i_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \int_{\substack{i_1(x(1)-\overline{x})>0 \\ \dots \\ i_n(x(n)-\overline{x})>0}} \cdots$$

$$\int \exp\left(-\frac{\sum_{i=1}^{n}(x(i)-\mu)^2}{2\sigma^2}\right) dx(1)\dots dx(n) == p(j_1, j_2, \dots, j_n),$$

where $j_k = -i_k, k = 1, \dots, n.$

It is easy to see $p_{l_1,l_2,\dots,l_n} = p(2l_1-1, 2l_2-1, \dots, 2l_k-1)$, *where* $l_k \in \{0, 1\}, k = 1, \dots, n$

Therefore

$p_{l_1,l_2,\dots,l_n} = p_{r_1,r_2,\dots,r_n}$ *if for* $l_i = 1 \to r_i = 0$ *and for* $l_i = 0 \to r_i = 1; i = 1, \dots, n$

Case $n = 4$**.** Let $X(1), X(2), X(3), X(4) - N(\mu, \sigma^2)$

It is obvious from (3),(4) that

$$T_4 = \begin{cases} 1, & P(T_4 = 1) = 4p_{0001} \\ 2, & P(T_4 = 2) = 6p_{0011} \\ 3, & P(T_4 = 3) = 4p_{0111} \end{cases}$$

It follows from note 1 that

$$p_1 = P(T_4 = 1) = P(T_4 = 3)$$

or

$$T_4 = \begin{cases} 1, & p_1 \\ 2, & p_2 \\ 3, & p_1 \end{cases}$$

It is obvious that $2p_1 + p_2 = 1$.

One has

$$E(T_4^2) = 10p_1 + 4p_2$$

and

$$E(T_4^2) = \sum_{i=1}^{4} EI(X(i) > \overline{X})^2 + \sum_{i<j} P(X(i) > \overline{X}, X(j) > \overline{X}) =$$

$$= 2 + 12P(X(1) > \overline{X}, X(2) > \overline{X})$$

As follows from [2]

$$P(X(1)-\overline{X}>0, X(2)-\overline{X}>0)=\frac{1}{4}+\frac{1}{2\pi}\arcsin\left(\frac{E(X(1)-\overline{X})(X(2)-\overline{X})}{D((X(1)-\overline{X}))}\right)$$

Therefore

$$P(X(1)-\overline{X}>0, X(2)-\overline{X}>0)=\frac{1}{4}-\frac{1}{2\pi}\arcsin\left(\frac{1}{3}\right)=0.1959133$$

One has

$$\begin{cases} 2p_1+p_2=1 \\ 10p_1+4p_2=4.35096 \end{cases}$$

Therefore

$$T_4=\begin{cases} 1, & P(T_4=1)=0.1754797 \\ 2, & P(T_4=2)=0.6490406 \\ 3, & P(T_4=3)=0.1754797 \end{cases} \tag{5}$$

for the case of sample of size 4 from normal distribution.

Note *According to numerical computations in R [4] one has* $p_{0011}=0.1081741$ *and* $P(T_4=2)=0.6490446$. *From [1] one has* $P(T_4=1)=P(T_4=3)=0.167168, P(T_4=2)=0.661804$.

Case $n=5$**.** The distribution of statistic T_5 calculated by three different method is presented in the following table.

Table 1.Distribution of T_5

Probabilities/Approaches	Exact	Orthant	Asymptotic
p_1	0.04892345	0.04892432	0.04973218
p_2	0.4510766	0.4503176	0.4495306
p_3	0.4510766	0.4511604	0.4495306
p_4	0.04892345	0.04892291	0.04973223

It is interesting to note that $p_1\neq p_4, p_2\neq p_3$ for calculating distribution of T_5 through orthant probabilities of multivariate normal distribution.

Case $n=6$**.** One has

$$T_6=\begin{cases} 1, & p_1 \\ 2, & p_2 \\ 3, & p_3 \\ 4, & p_2 \\ 5, & p_1 \end{cases}$$

and

$$2p_1+2p_2+p_3=1 \tag{6}$$

Besides

$$ET_6^2 = p_1 + 4p_2 + 9p_3 + 16p_2 + 25p_1 = 26p_1 + 20p_2 + 9p_3$$

and

$$ET_6^2 = 3 + \frac{30}{4} + \frac{30}{2\pi}\arcsin\left(\rho(X_i - \overline{X}, X_j - \overline{X})\right)$$

Since

$$\rho(X(1) - \overline{X}, X(2) - \overline{X}) = \frac{E(X(1) - \overline{X})(X(2) - \overline{X})}{D((X(1) - \overline{X}))} = -\frac{1}{5}$$

then

$$26p_1 + 20p_2 + 9p_3 = \frac{42}{4} + \frac{30}{2\pi}\arcsin\left(-\frac{1}{5}\right) \tag{7}$$

Let us consider

$$ET_6^4 = 626p_1 + 272p_2 + 81p_3$$

One has

$$\begin{cases} 2p_1 + 2p_2 + p_3 = 1 \\ 26p_1 + 20p_2 + 9p_3 = \frac{42}{4} + \frac{30}{2\pi}\arcsin\left(-\frac{1}{5}\right) \\ 626p_1 + 272p_2 + 81p_3 = ET^4 \end{cases}$$

$$ET_6^4 = 360 * E(I(X_1 > \overline{X})I(X_2 > \overline{X})I(X_3 > \overline{X})I(X_4 > \overline{X})) +$$

$$+720 * E(I(X_1 > \overline{X})^2 I(X_2 > \overline{X})I(X_3 > \overline{X})) + 90E(I(X_1 > \overline{X})^2 I(X_2 > \overline{X})^2) +$$

$$+120 * E(I(X_1 > \overline{X})^3 I(X_2 > \overline{X})) + 6E(I(X_1 > \overline{X})^4) =$$

$$= 360P(X_1 > \overline{X}, X_2 > \overline{X}, X_3 > \overline{X}, X_4 > \overline{X}) + 720\left(\frac{1}{8} - \frac{3}{4\pi}\arcsin\left(\frac{1}{5}\right)\right) +$$

$$+210\left(\frac{1}{4} - \frac{1}{2\pi}\arcsin\left(\frac{1}{5}\right)\right) + 3$$

For calculation $P(X_1 > \overline{X}, X_2 > \overline{X}, X_3 > \overline{X})$ the following formula is used [3]

$$2p_{+++} = p_{+\bullet+} + p_{++\bullet} + p_{\bullet++} - \frac{1}{2} \tag{8}$$

For calculation of

$$P(X_1 > \overline{X}, X_2 > \overline{X}, X_3 > \overline{X}, X_4 > \overline{X}) = 0.0187$$

computer program R has to be used.

Therefore $ET_6^4 = 110.8695$

Finally one has

$$p_1 = 0.0070625, p_2 = 0.24175, p_3 = 0.502375 \tag{9}$$

The distribution of statistic T_6 calculated by four different methods is presented in the following table.

Table 2. Distribution of T_6.

Probabilities/Approaches	(9)	Orthant	Simulation	Asymptotic
p_1	0.0070625	0.01167561	0.01141	0.0137768
p_2	0.24175	0.2218076	0.22101	0.2159234
p_3	0.502375	0.530119	0.53147	0.5403608
p_4	0.24175	0.2232135	0.22451	0.2159234
p_5	0.0070625	0.01152457	0.0116	0.0137768

It is interesting to note that $p_1 \neq p_5, p_2 \neq p_4$ for calculating distribution of T_5 through orthant probabilities of multivariate normal distribution.

Case $n = 7$ In the same way one can obtain the distribution of statistic T_7. One has

$$p_1 = 0.00202, p_2 = 0.0891125, p_3 = 0.4088674 \tag{10}$$

The distribution of statistic T_7 calculated by three different methods is presented in the following table.

Table 3. Distribution of T_7.

Probabilities/Approaches	(10)	Simulation	Asymptotic
p_1	0.00202	0.003	0.003672918
p_2	0.0891125	0.092	0.08529012
p_3	0.4088674	0.392	0.4110005
p_4	0.4088674	0.425	0.4110005
p_5	0.0891125	0.085	0.08529012
p_6	0.00202	0.003	0.003672918

0.3 Conclusion

Exact distribution of the sign test statistic (1) is founded in the paper for $n = 3, 4, 5$. This result is obtained using representation of orthant probabilities through correlations. The orthant probabilities could be used to find the distribution of statistic T_n for the case $n \geq 4$. This is done in the paper for the case $n = 4, 5, 6, 7$. The comparisons of obtained results with asymptotic

distribution are presented. It is interesting to note that obtained numerical results are very close to known asymptotic distribution which is far from binomial distribution.

REFERENCE

1. GASTWIRTH, JOSEPH L. (1971), "On the Sign Test for Symmetry,"Journal of the American Statistical Association, 66, 821-823.
2. Cramer H. Mathematical methods of statistics, Princeton University Press, Ninth Edi- tion 1962.
3. DUSTIN CARTWRIGHT. DISCRETIZING GAUSSIAN MODELS, 2008.
4. Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2017). mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-6.

Koldanov Petr Alexandrovich (Nijni Novgorod, Russia), Koldanov Alexandr Petrovich (Nijni Novgorod, Russia)

On the distribution of sign test for symmetry.

Abstract. *Exact distribution of the sign test statistic for the case of unknown mean for sample of sizes 3, 4, 5 from normal distribution is founded. Moreover for the sample of sizes 6 and 7 the distribution is obtained by numerical algorithms for calculation of orthant probabilities of elliptically contoured distribution. The results of comparison of the distribution with known asymptotic distribution is presented. It is shown that the degree of convergence to the asymptotic distribution is very rapid.*

Key words:Symmetry hypothesis, sign test, normal distribution, probability of sign coincidence.

Случайно-множественная интерпретация мер значимости ассоциативных правил

Д.В.Семенова[1,2]
[1] Сибирский федеральный университет
[2] Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого МЗ РФ
DVSemenova@sfu-kras.ru

В настоящее время актуальны задачи, связанные с необходимостью обработки больших массивов данных с целью поиска скрытых закономерностей, установления и выявления новых знаний, которые впоследствии могут быть использованы экспертами прикладной области. Подобные задачи, как правило, имеют комбинаторный характер и заключаются в поиске зависимостей между связанными событиями в виде, доступном интерпретации человеком. Извлечение полезных знаний невозможно без хорошего понимания сути данных, а успешный анализ требует их качественной предобработки и тщательного выбора модели для интерпретации зависимостей, которые могут быть обнаружены. Средством описания причинно-следственных закономерностей в многомерных данных, представленных $0,1$-матрицей кобъект – признакнь, служат ассоциативные правила [1, 5, 6, 9, 13, 15, 16], отражающие какие признаки появляются вместе и насколько часто это происходит. Таким образом, ассоциативные правила представляют собой механизм нахождения логических закономерностей между связанными элементами (событиями или объектами). Алгоритмы поиска ассоциативных правил основаны на анализе формальных понятий [3, 10, 14].

Анализируемое множество данных будем описывать бинарным контекстом $\mathcal{K} = \langle \mathcal{G}, \mathcal{M}, \mathcal{I} \rangle$, где $\mathcal{G}$ — множество объектов, $\mathcal{M}$ — множество признаков, а $\mathcal{I} \subseteq \mathcal{G} \times \mathcal{M}$ — отношение инцидентности такое, что $(g, f) \in \mathcal{I}$ тогда и только тогда, когда объект g обладает свойством f (или признак f присущ объекту g). Ассоциативным правилом r на множестве признаков $\mathcal{M}$ контекста $\mathcal{K} = \langle \mathcal{G}, \mathcal{M}, \mathcal{I} \rangle$, называется такая упорядоченная пара $r = (X, Y)$, где $X, Y \subseteq \mathcal{M}$. Используется следующая условная форма записи ассоциативного правила r: $X \to Y$, где X – причина/посылка, Y – следствие/заключение [1].

В настоящее время практическое применение ассоциативных правил ограничивается проблемой размерности. Имеющиеся методы и средства анализа данных могут приводить к значительному числу искомых ассоциативных правил. Это создает проблему их верификации и снижает степень

доверия пользователя к полученным результатам. Одним из подходов к решению данной проблемы является фильтрация ассоциативных правил на основе мер значимости, которая производится путем генерации частых множеств алгоритмом Apriori или его разновидностями [1, 5, 9, 12].

Под мерой значимости ассоциативного правила $r = (X, Y)$ будем понимать числовую функцию $\mu : 2^{\mathcal{M}} \times 2^{\mathcal{M}} \to \mathbb{R}$, «измеряющую» важность/значимость/интерес ассоциативного правила r контекста $\mathcal{K}$.

В интеллектуальном анализе данных выделяют традиционные и специфические меры значимости [7, 8]. Традиционные меры значимости служат фильтром для выбора подгруппы потенциально интересных ассоциативных правил. Специфические меры значимости могут использоваться в качестве второго фильтра, чтобы сохранить только те правила, которые действительно интересны как для пользователя, так и для сферы их применения. Ассоциативные правила позволяют выявить в данных закономерности между взаимосвязанными событиями. Таким образом, задача поиска ассоциативных правил для заданного контекста заключается в нахождении множества всех ассоциативных правил, меры значимости которых выше заданных соответствующих пороговых значений.

В общем случае задача поиска ассоциативных правил в бинарном контексте $\mathcal{K}$ формулируется следующим образом.

Для заданного бинарного контекста $\mathcal{K} = \langle \mathcal{G}, \mathcal{M}, \mathcal{I} \rangle$, с пороговыми значениями $\{\theta_1, \ldots, \theta_k\}$, которые соответствуют мерам значимости ассоциативных правил $\{\mu_1(r), \ldots, \mu_k(r)\}$, найти множество $\mathcal{AR}$ всех ассоциативных правил $X \to Y$, $X, Y \in \mathcal{M}$, удовлетворяющих условиям:

$$\begin{cases} \mu_1(X \to Y) \geq \theta_1; \\ \ldots \\ \mu_k(X \to Y) \geq \theta_k, \end{cases} \tag{11}$$

где $\{\mu_1(r), \ldots, \mu_k(r)\}$ — меры значимости ассоциативных правил, а $\{\theta_1, \ldots, \theta_k\}$ — соответствующие им пороговые значения.

Задача носит перечислительный комбинаторный характер. В худшем случае число всех ассоциативных правил в заданном контексте $\mathcal{K} = \langle \mathcal{G}, \mathcal{M}, \mathcal{I} \rangle$ экспоненциально зависит от $|\mathcal{M}|$. Поэтому все существующие алгоритмы ее решения неполиномиальные по времени.

Тип модели при решении задачи поиска ассоциативных правил зависит от используемых в (11) мер значимости (см. табл. 1). Наиболее широкое применение для решения задачи поиска ассоциативных правил в рамках модели типа кподдержка-достоверностьь получили алгоритмы Apriori, Close,

MClose [1, 3, 5, 15, 16]. Поддержка $\delta(X \to Y)$ задает частоту встречаемости признаков $X \cup Y$ в рассматриваемом контексте $\mathcal{K} = \langle \mathcal{G}, \mathcal{M}, \mathcal{I} \rangle$ и определяется формулой $\delta(A \to B) = |(X \cup Y)'|/|G|$. Здесь применяются обозначения отображения Галуа как оператора вида [10, 14]: $Z' = \{g \in \mathcal{G} : \forall m \in Z, (g, m) \in \mathcal{I}\}$, и Z' определяет множество всех объектов общих для всех признаков из $Z \subseteq \mathcal{M}$. Известно, что этому отображению свойственна антимонотонность [1]. Поэтому поддержка тоже антимонотонная. Если $Z = Z''$, то Z называется замкнутым множеством относительно оператора к″нь. Достоверность $\gamma(X \to Y) = \delta(X \to Y)/\delta(X)$ определяет отношение числа объектов, обладающих всеми признаками из $X \cup Y$, к числу объектов, которым свойственны только признаки X.

В настоящее время активно изучаются модели типа кподдержка-достоверность-зависимостьњ [2], в которых помимо поддержки и достоверности вводится специальная мера, являющейся еще одним параметром для выбора или отклонения правила. Следует заметить, что в большинстве случаев меры значимости основаны на анализе статистических свойств данных.

Все последние десятилетия и до настоящего времени проблеме исследования вероятностных свойств мер значимости ассоциативных правил уделяется особое внимание, поскольку фильтрация на основе мер значимости позволяет снизить вычислительные затраты при поиске ассоциативных правил. В таблице 1 приведен далеко неполный перечень мер, используемых в ассоциативном анализе [7, 8].

С другой стороны, математической моделью сложных объектов и систем, у которых число описываемых признаков конечно и появление любого из них представляется как случайное событие, рассматриваются конечные случайные множества [4]. Под конечным случайным множеством понимается случайный элемент со значениями из совокупности всех подмножеств некоторого конечного множества $\mathcal{M}$. При этом множество M будем называть *базовым множеством*, а значения конечного случайного множества образуют полную решетку относительно теоретико-множественного включения и теоретико-множественных операций объединения и пересечения. Если число признаков равно N, то возможное число различных состояний объекта равно 2^N. Если каждому из этих состояний сопоставить некоторую вероятность появления, то набор всех 2^N вероятностей будет задавать при определенных условиях распределение, называемое распределением вероятностей конечного случайного множества. Вероятностные распределения конечных случайных множеств позволяют дать сжатое описание неструк-

Таблица 1: Вероятностная и алгебраическая интерпретация мер значимости ассоциативного правила

Мера	**Алгебраическая интерпретация**	**Вероятностная интерпретация**	**Автор**
Поддержка	$\delta(X \to Y)$	$\mathbf{P}(XY)$	Agrawal, 1993
Достоверность	$\gamma(X \to Y) = \frac{\delta(X \to Y)}{\delta(X)}$	$\mathbf{P}(Y\|X)$	Agrawal, 1993
Покрытие	$Coverage(X \to Y) = \delta(X)$	$\mathbf{P}(X)$	Agrawal и Srikant, 1993
Распространенность	$Prev(X \to Y) = \delta(Y)$	$\mathbf{P}(Y)$	Agrawal и Srikant, 1993
Память	$Rec(X \to Y) = \frac{\delta(X \to Y)}{\delta(Y)}$	$\mathbf{P}(X\|Y)$	Getoor, 2001
Специфичность	$Specif(X \to Y) = \frac{\delta\left(\overline{X} \to \overline{Y}\right)}{\delta\left(\overline{X}\right)}$	$\mathbf{P}\left(Y^c\|X^c\right)$	Tan, 2004
Точность	$Acc(X \to Y) = \delta(X \to Y) + \delta\left(\overline{X} \to \overline{Y}\right)$	$\mathbf{P}(XY) + \mathbf{P}\left(X^cY^c\right)$	Holte, 1993
Интересность	$Lift(X \to Y) = \frac{\delta(X \to Y)}{\delta(X)\delta(Y)}$	$\frac{\mathbf{P}(Y\|X)}{\mathbf{P}(Y)}$	Brin, 1997
Влияние	$Lev(X \to Y) = \frac{\delta(X \to Y)}{\delta(X)} - \delta(X)\delta(Y)$	$\mathbf{P}(Y\|X) - \mathbf{P}(X)\mathbf{P}(Y)$	Piatetsky-Shapiro, 1991
Добавочная ценность	$AV(X \to Y) = \frac{\delta(X \to Y)}{\delta(X)} - \delta(Y)$	$\mathbf{P}(Y\|X) - \mathbf{P}(Y)$	Tan, 2004
Фактор определенности	$CF(X \to Y) = \frac{\gamma(X \to Y) - \delta(Y)}{1 - \delta(Y)}$	$\frac{\mathbf{P}(Y\|X) - \mathbf{P}(Y)}{1 - \mathbf{P}(Y)}$	Berzal, 2002
Пятецкий-Шапиро	$PS(X \to Y) = \delta(XY) - \delta(X)\delta(Y)$	$\mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y)$	Piatetsky-Shapiro, 1991

турированных данных, для которых размер каждой транзакции (множества событий, произошедших одновременно) не является фиксированным [4]. В данной работе меры анализируются с точки зрения теории конечных случайных множеств.

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — исходное вероятностное пространство. Пусть $K_{\mathcal{M}}$ — конечное случайное множество, заданное на базовом множестве $|\mathcal{M}| = N < \infty$, где $\mathcal{M} \subset \mathcal{F}$. Распределением вероятностей конечного случайного множества $K_{\mathcal{M}}$ называется набор из 2^N значений вероятностной меры $\mathbf{P}$, определенной на системе событий $2^{\mathcal{M}}$.

Для полного определения распределения вероятностей конечного случайного множества $K_{\mathcal{M}}$ достаточно любого одного из шести типов распреде-

лений I — VI рода [4]. В работе рассматривается распределения I и II рода.

Распределение вероятностей I рода конечного случайного множества $K_{\mathcal{M}}$ — функция множества, определенная на системе несовместных событий вида $\{K_{\mathcal{M}} = X\} = \left(\left(\bigcap_{x\in X} x\right)\bigcap\left(\bigcap_{v\in X^c} v^c\right)\right)$, где $X^c = \mathcal{M}\setminus X$, $x^c = \Omega\setminus x$. Это набор из 2^N вероятностей: $p(X) = \mathbf{P}\big(\{K_{\mathfrak{X}} = X\}\big)$. Распределение вероятностей I рода конечного случайного множества $K_{\mathfrak{X}}$ удовлетворяет условию $\sum\limits_{X\subseteq\mathfrak{X}} p(X) = 1$.

Распределение вероятностей II рода конечного случайного множества $K_{\mathcal{M}}$ — функция множества, определенная на системе совместных событий вида $\{K_{\mathcal{M}} \supseteq X\} = \left(\bigcap_{x\in X} x\right)$. Это набор из 2^N вероятностей: $p_X = \mathbf{P}(K_{\mathcal{M}} \supseteq X)$. Распределение вероятностей II рода удовлетворяет системе неравенств Фреше для вероятностей пересечений событий [4]:

$$\max\left\{0, \sum_{x\in X}\mathbf{P}(x) - |X| + 1\right\} \le \mathbf{P}\left(\bigcap_{x\in X} x\right) \le \min_{x\in X}\mathbf{P}(x), \qquad (12)$$

где $\mathbf{P}(x) = \mathbf{P}(\{x \in K_{\mathcal{M}}\})$, $x \in X$, $X \subseteq K_{\mathcal{M}}$, — вероятности самих событий.

Каждую строку контекста $\mathcal{K} = \langle\mathcal{G}, \mathcal{M}, \mathcal{I}\rangle$ будем рассматривать как бинарное представление значения некоторого конечного случайного множества $K_{\mathcal{M}}$. Тогда строки контекста $K_1, K_2, \ldots, K_{|\mathcal{G}|}$ — случайная выборка, состоящая из $|\mathcal{G}|$ независимых одинаково распределенных случайных множеств с распределением $\{p_X, X \subseteq \mathcal{M}\}$. Оценка вероятности II рода для событий вида $\{K_{\mathcal{M}} \supseteq X\}$ очевидным образом выражается через отображение «$'$»: $\tilde{p}_X = X'/|\mathcal{G}|$, для всех $X \subseteq \mathcal{M}$ и совпадает с определением поддержки множества $\delta(X)$ в алгебраической интерпретации. Таким образом, набор $\{\delta(X), X \subseteq \mathcal{M}\}$ является непараметрической оценкой распределения вероятностей II рода $\{p_X, X \subseteq \mathcal{M}\}$ конечного случайного множества $K_{\mathcal{M}}$.

Таким образом, для случайно-множественного анализа всех мер значимости из таблицы 1 достаточно знать $\{\tilde{p}_X, X \subseteq \mathcal{M}\} = \{\delta(X), X \subseteq \mathcal{M}\}$. Отсюда следует, что для ассоциативного правила $X \to Y$ меры поддержка $\delta(X \to Y)$, покрытие $Coverage(X \to Y)$ и распространенность $Prev(X \to Y)$ совпадают со значениями соответствующих вероятностей II рода: p_{XY}, p_X, p_Y. Более того, зная только поддержки $\delta(x) = p_x$ каждого отдельного признака $x \in \mathcal{M}$ с использованием формулы (12) нетрудно получить верхнюю и нижнюю границу значений поддержки любого ассо-

циативного правила $X \to Y$.

Для анализа мер значимости из таблицы 1, которые выражаются через условные вероятности, возникает необходимость рассматривать помимо всего базового множества событий некоторую его часть $\mathfrak{Y} \subset \mathcal{M}$. При этом приходится оперировать двумя случайными множествами[1] $K_{\mathcal{M}}$ и $K_{\mathfrak{Y}}$, распределения которых связаны аналитическими формулами [4]: $p\left(Y/\!/\mathfrak{Y}\right) = \sum_{Z \subseteq \mathcal{M} \setminus \mathfrak{Y}} p\left(Y + Z/\!/\mathcal{M}\right),\ p_{Y/\!/\mathfrak{Y}} = p,\ \forall\, Y \subseteq \mathfrak{Y}$. Фиксация некоторого множества признаков $\mathfrak{Y} \subset \mathcal{M}$ в качестве причины/следствия некоторого ассоциативного правила приводит к $2^{|\mathfrak{Y}|}$ условным распределениям случайных множеств $\left\{p_{X/\!/\mathcal{M}\setminus\mathfrak{Y}|Y/\!/\mathfrak{Y}}, X \subseteq \mathcal{M} \setminus \mathfrak{Y}\right\},\ Y \subseteq \mathfrak{Y}$. Меры значимости достоверность $\gamma(X \to Y)$, память $Rec(X \to Y)$, специфичность $Specif(X \to Y)$ совпадают с условными вероятностями соответствующих событий. Нетрудно показать, что равенство этих мер значимости единице приводит к вырожденным условным распределениям, что указывает на вложенную структуру зависимости [4]. Например, для строгого ассоциативного правила из равенства $\gamma(X \to Y) = 1$ следует, что $X \subset Y$.

Мера Пятецкого-Шапиро из табл. 1 совпадает с одной из важнейших характеристик конечного случайного множества — арной ковариаций множества событий. Арная ковариация «определяет, насколько вероятнее (чаще) или невероятнее (реже) наступают рассматриваемые события по сравнению с независимыми событиями, имеющими те же вероятности» [4]. Для множества $\mathcal{M}$, состоящего из N событий, существует набор из $2^N - N - 1$ арных ковариаций, которые определяются распределением вероятностей II рода конечного случайного множества: $\mathrm{Kov}_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) - \prod_{x \in X} p_x, \quad X \subseteq \mathfrak{X}, \quad |X| > 1$. Арная ковариация множества событий обращается в ноль, когда события из X независимы; больше нуля, когда события из множества X «статистически притягиваются»; меньше нуля, когда события из множества X «статистически отталкиваются» [4]. Данная интерпретация позволяет статистически оценить зависимость между причиной и следствием ассоциативного правила.

Случайно-множественный подход к анализу мер значимости ассоциативных правил рассматривает, с точки зрения практики, «худший» случай,

[1]В ситуациях, когда рассматриваются два и более конечных случайных множества, определенных на различных базовых множествах событий, будем использовать следующее обозначение $Y/\!/\mathfrak{Y}$. Это обозначение указывает, что рассматривается некоторое подмножество Y из множества случайных событий $\mathfrak{Y}$.

когда мы исследуем всч множество ассоциативных правил, мощность которого экспоненциально зависит от $|\mathcal{M}|$. Однако, подобный анализ показал, что некоторые алгоритмы поиска ассоциативных правил успешно работают только для конечных случайных множеств с известной структурой зависимости (например, Close и MClose ищут строгие ассоциативные правила, что соответствует вложенной структуре зависимостей конечного случайного множества событий). И ассоциативный анализ и теория конечных случайных множеств — это технологии обработки $0,1$-матриц, которые направлены на извлечение и сжатое представление закономерностей в многомерных данных.

ЛИТЕРАТУРА

1. Быкова, В. В. О неизбыточном представлении минимаксного базиса строгих ассоциативных правил / В.В. Быкова, А.В. Катаева // Красноярск: ПДМ. СФУ. — 2017. №36. С. 113–126.

2. Городецкий В.И., Самойлов В.В. Ассоциативный и причинный анализ и ассоциативные байесовские сети // Тр. СПИИРАН. 2009. Вып. 9. С. 13Ц65.

3. Кузнецов, С. О. Автоматическое обучение на основе анализа формальных понятий / С. О. Кузнецов // Автоматика и телемеханика. – 2001. — №10. — С. 3–27.

4. Семенова, Д. В. Рекуррентное построение распределений вероятностей конечных случайных множеств: монография / Д. В. Семенова, Н. А. Лукьянова. — Красноярск: Сиб. федер. ун-т, 2018. Ц– 136 с.

5. Adamo J.-M. Data Mining for Association Rules and Sequential Patterns. / Adamo J.-M. // Springer Science+Business Media New York, 2001. — pp. 254.

6. Agrawal R. Fast Algorithms for Mining Association Rules in Large Databases. / R. Agrawal and R. Srikant // In Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. – pp. 487–499, 1994.

7. Berzal F. Measuring the accuracy and interest of association rules: a new framework, / F. Berzal, I. Blanco, D. Sanchez, and M.-A. Vila // Intelligent Data Analysis, vol. 6, no. 3, pp. 221–235, 2002.

8. Geng L. Interestingness Measures for Data Mining: A Survey / L. Geng, H. J. Hamilton // ACM Computing Surveys, Vol. 38, No. 3, Article 9, Publication date: September 2006.

9. Brin S. Beyond market baskets: generalizing association rules to correlations / S. Brin, R. Motwani, C. Silverstein // Proc. of the ACM SIGMOD Intern. Conf. on Management of Data. 1997. P. 255Ц264.

10. Ganter, B. Formal Concept Analysis: Mathematical Foundations / B. Ganter and R. Wille // Springer, 1999. — P. 284.

11. Pasquier N. Generating a condensed representation for association rules. / N. Pasquier, Y. Bastide, R. Taoui and L. Lakhal // Intelligent Inform. Systems. — 2005, vol. 24, no. 1, P. 29–60.

12. Piatetsky–Shapiro G. Discovery, analysis and presentation of strong rules / Knowledge Discovery in Databases. AAAI Press/MIT Press. 1991. P. 229Ц248

13. Tan P. Interestingness Measures for Association Patterns: A Perspective. / P. Tan and V. Kumar // In Proceedings of the Workshop on Postprocessing in Machine Learning and Data Mining, New York, USA, 2000.

14. Wille, R. Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts / R. Wille // Ordered Sets. — Boston: Reidel, 1982. — P. 445–470.

15. Zaki M. J. Efficient Algorithms for Mining Closed Itemsets and Their Lattice Structure. / M. J. Zaki, Ch.-J. Hsiao IEEE Trans. on Knowledge and Data Engineering. 2005, vol. 17, no. 4, P. 462–478.

16. Zhang, C. Association rule mining / C. Zhang, S. Zhang // Springer, 2002. — P. 240.

Semenova Daria (Krasnoyarsk, Russia)

Interpretation of interestingness measures for association rules by use theory of finite random sets

Abstract. *Association rules mining is an important topic in the domain of data mining and knowledge discovering. Some papers have presented several interestingness measure methods; the most typical are Support, Confidence, Lift, Improve, and so forth. The paper proposes a analysis of known*

interestingness measures for association rules by use theory of finite random sets.

Key words: association rules, interestingness measures, finite random set

The Probabilities of Large Deviations Associated with Multinomial Distribution

Sh.Mirakhmedov
Institute of Mathematics. Academy of Sciences of Uzbekistan
shmirakhmedov@yahoo.com

Abstract. Let $\eta = (\eta_1, ..., \eta_N)$ be a multinomial random vector with parameters $n = \eta_1 + ... + \eta_N$and $p_m > 0$,$m = 1, ..., N$,$p_1 + ... + p_N = 1$, such that $N \to \infty$and $\max p_m \to 0$, as $n \to \infty$. Main results of the paper are the probability of large deviation theorems for the class of statistic $R_N(\eta) = h_1(\eta_1) + ... + h_N(\eta_N)$, where h_ms are real-valued functions defined on non-negative integers. These theorems are applied to derive new large deviation results for the chi-squared statistic, the log-likelihood ratio statistic and the empty boxes statistic.

Key words and phrases. Chi-squared statistic, Empty boxes statistic, Log-likelihood ration statistic, large deviations, multinomial distribution, Poisson distribution.

1. Introduction. Let $\eta = (\eta_1, ..., \eta_N)$ be a random vector having a multinomial distribution$M(n, p_1, ..., p_N)$, where $n = \eta_1 + ... + \eta_N$, $p_1 + ... + p_N = 1$,$p_m > 0$,$m = 1, ..., N$, and let

$$R_N = \sum_{m=1}^{N} h_m(\eta_m), \tag{1.1}$$

where $h_1(x), ..., h_N(x)$ are real-valued functions on non-negative real line. The random variables (r.v.) of the form (1.1) are of interest, as they arise in the problems of probability theory, in the context of urn models, and of mathematical statistics associated with grouped data. The most important special cases of R_N are

$$\chi^2_N = \sum_{m=1}^{N} \frac{(\eta_m - np_m)^2}{np_m}, \Lambda_N = 2\sum_{m=1}^{N} \eta_m \ln \frac{\eta_m}{np_m} \text{ and } \mu_0 = \sum_{m=1}^{N} I\{\eta_m = 0\}, \tag{1.2}$$

where I$\{A\}$is the indicator function of A. These are known as Pearson's chi-square statistics, log-likelihood ratio (LR) statistics and empty boxes (EB) statistics, respectively. The chi-square and LR statistics play leading role in testing goodness of fit on grouped data, whereas EB statistics mostly used in connection with occupancy problems. In the present paper we

study$P\left\{\tilde{R}_N \geq x_N\right\}$, where $\tilde{R}_N$ is standardized version of $R_N$and$x_N \to \infty$, in a triangular array situation, when $N = N(n) \to \infty$and $p_m = p_m(n) \to 0$,$m = 1, 2, ..., N$, as $n \to \infty$.

Statistics (1.1) has been considered in the literature. Among others, we refer to Holst (1972), Morris (1975), Mirakhmedov (1987), Sirajdinov et al (1989), Ivchenko and Mirakhmedov (1992) and Mirakhmedov et al (2014).

2. Results. Let $\xi_1, ..., \xi_N$be independent r.v.s, where $\xi_m \sim Poi(np_m)$,$p_{\max} = \max\limits_{1\leq k\leq N} p_k$,

$p_{\min} = \min\limits_{1\leq k\leq N} p_k$,$\lambda_n = n/N$. We assume that $p_{\max} = o(1)$, $N \to \infty$, as $n \to \infty$. Set

$$g_m(\xi_m) = h_m(\xi_m) - Eh_m(\xi_m) - \gamma_n(\xi_m - np_m), \gamma_n = n^{-1}\sum_{m=1}^{N} cov\left(h_m(\xi_m), \xi_m\right),$$

$$A_N = \sum_{m=1}^{N} Eh_m(\xi_m), \tilde{\sigma}_N^2 = \sum_{m=1}^{N} Varh_m(\xi_m), \sigma_N^2 = \sum_{m=1}^{N} Var\, g_m(\xi_m) = \tilde{\sigma}_N^2 - n\gamma_n^2. \tag{2.1}$$

Note that under the below conditions (i) and (ii) $ER_N = A_N + o(N)$ and $VarR_N = \sigma_N^2(1 + o(1))$.

Theorem 2.1. Let

(i) $E\left|h_m(\xi_m)\right|^s = O\left(Eh_m^s(\xi_m)\right)$,

(ii) $E\left|h_m(\xi_m)\right|^s (\xi_m - np_m)^2 = O\left(s^a(np_m)^b Eh_m^s(\xi_m)\right)$

with $a = a_1$ and $b = b_1$for all $m \in \mathcal{N} \subseteq (1, 2, ..., N)$, and with $a = a_2$ and $b = b_2$for all $m \in (1, 2, ..., N)\backslash\mathcal{N}$, where some a_i and $b_i$are non-negative, and integer s: $3 \leq s \leq k = o(p_{\max}^{-1})$,

(iii) for each $m = 1, ..., N$ there exist $\omega_m^2 > 0$and a sequence $\{V_n > 0\}$such that

$$|\mathcal{C}_k(h_m(\xi_m))| \leq (k!)^2 V_n^{k-2}\omega_m^2, \tag{2.2}$$

for all k:

$$3 \leq k = o\left(\min(K_n(a_1, b_1), K_n(a_2, b_2), p_{\max}^{-1}\right), \tag{2.3}$$

where $K_n(a, b) = \left(n^{1-b}p_{\max}^{-b}\right)^{1/(\bar{a}+1)}$, $\bar{a} = \max(1, a)$. Then for all $x_N \geq 0$ such that

$$x_N = o\left(\min\left(W_N^{1/3}, K_n^{1/2}(a_1, b_1), K_n^{1/2}(a_2, b_2), p_{\max}^{-1/2}\right)\right), \tag{2.4}$$

where $W_N = V_n^{-1}\sigma_N \min\left(1, \sigma_N^2/\tilde{\omega}_N^2\right)$,$\tilde{\omega}_N^2 = \omega_1^2 + ... + \omega_N^2$, it holds

$$P\left\{R_N > x_N\sigma_N + A_N\right\} = \left(1 - \Phi(x_N)\right)\left(1 + o(1)\right) \tag{2.5}$$

and

$$P\left\{R_N < -x_N\sigma_N + A_N\right\} = \Phi(-x_N)\left(1 + o(1)\right). \tag{2.6}$$

Let us consider application of Theorem 2.1 to the statistics (1.2).

2.1. Chi-square statistics. Set $\nabla_n = \max\left(1, (np_{\min})^{-1}\right)$. For the χ^2_Nstatistics $h_n(u) = (u - np_m)^2/np_m$,

$$\tilde{\sigma}^2_N = \sum_{m=1}^{N} \frac{1}{np_m} + 2N, \sigma^2_N = \frac{1}{\lambda_n}\sum_{m=1}^{N}\left(\frac{1}{Np_m} - 1\right) + 2N = \tilde{\sigma}^2_N - N\lambda_n^{-1}. \tag{2.7}$$

Theorem 2.2. For all$x_N \geq 0$ such that $x_N = o\left(\min\left(\left(\sigma^3_N/\tilde{\sigma}^2_N\nabla_n\right)^{1/3}, n^{1/6}, p_{\max}^{-1/4}\right)\right)$the relations (2.5) and (2.6) are hold true with $R_N = \chi^2_N$, $A_N = N$, and σ^2_N is defined in (2.7).

Corollary 2.1. Let $\lambda_n \geq C > 0$and $C_1 \leq Np_{\min} \leq Np_{\max} \leq C_2N^{1/3}$, then for all $x_N \geq 0$ such that $x_N = o(N^{1/6})$ the relations (2.5) and (2.6) are valid with$R_N = \chi^2_N$, $A_N = N$and σ^2_Ndefined in (2.7).

Corollary 2.2. Let$p_m = N^{-1}\left(1 + \delta(n)d_{m,n}\right)$, $m = 1, ..., N$, where $\delta(n)to0$, $d_{1,n} + ... + d_{N,n} = 0$, and $d^2_{1,n} + ... + d^2_{N,n} = C_0N$. Then for arbitrary λ_n and $x_N \geq 0$ such that $x_N = o\left(N^{1/6}\min(1, \lambda_n^{2/3})\right)$ the relations (2.5) and (2.6) are valid with$R_N = \chi^2_N$, $A_N = N$and $\sigma^2_N = 2N$.

Remarks 2.2. For $x_N \to \infty$the relation (2,5) can be written in the form $\log P\left\{\chi^2_N > x_N\sigma_N + N\right\} = -x^2_N/2(1 + o(1))$. From Kallenberg (1985, Eq. (2.17)) such kind relation follows if $\log N = o(x^2_N)$, $x_N = o(N^{1/6})$ and $N = o(n^{3/8})$. Last condition excludes the case recommended by Mann and Wald (1942) , who obtained the relation$N = cn^{2/5}$ concerning the optimal choice of the number of groups in chi-square goodness of fit test.

2.2. Likelihood ratio statisticsΛ_N. Here we restrict attention to the case $\lambda_n \to \infty$; the case $\lambda_n \to \lambda \in (0, \infty)$was given by Sirajdinov et al (1989).

Theorem 2.3. Let$\lambda_n \to \infty$and $C_3 \leq Np_{\min} \leq Np_{\max} \leq C_4\lambda_n$. Then for all$x_N \geq 0$such that$x_N = o\left(\min(N^{1/6}, p_{\max}^{-1/4})\right)$ it holds the relations (2.5) and (2.6), with $R_N = \Lambda_N$, $A_N = N$and $\sigma^2_N = 2N$.

Corollary 2.3. Let $\lambda_n \to \infty$and$p_m = N^{-1}\left(1 + o(1)\right)$, $m = 1, ..., N$**,** then for all $x_N \geq 0$ such that $x_N = o(N^{1/6})$the relations (2.5) and (2.6)are valid with $R_N = \Lambda_N$,$A_N = N$and $\sigma^2_N = 2N$.

Remarks 2.3. Let $\lambda_n \to \infty$,$Np_{\max} = O(\min(N^{1/3}, \lambda_n))$, $x_N \to \infty$and$x_N = o(N^{1/6})$. Then from (2.1) it follows that$\log P\left\{\Lambda_N > x_N\sqrt{2N} + N\right\}$ == $-x^2_N/2\left(1 + o(1)\right)$. From Kallenberg (1985, Eq. (2.13)) such relation follows under additionally conditions$N = o\left(n^{3/7}\right)$and $\log N = o(x^2_n)$.

2.3 The Empty Boxes Statistics. The μ_0 is statistics (1.1) with $h_m(x) = \mathrm{I}\{x = 0\}$ and

$$A_N = \sum_{m=1}^{N} \exp\{-np_m\}, \gamma_n = -\sum_{m=1}^{N} p_m \exp\{-np_m\},$$

$$\sigma_N^2 = \sum_{m=1}^{N} \exp\{-np_m\}(1 - \exp\{-np_m\}) - n\gamma_n^2. \tag{2.8}$$

Theorem 2.4. For all$x_N \geq 0$ such that $x_N = o\left(\min\left(\sigma_N N^{-1/3}, (np_{\max}^2)^{-1/2}, p_{\max}^{-1/2}\right)\right)$the relations (2.5) and (2.6) are hold true with $R_N = \mu_0$, $A_N$and σ_N^2 are defined in (2.8).

Corollary 2.4. If (i) there exist positive constants C_5,C_6 and C_7 such that $Np_{\max} \leq C_5$and

$C_6 \leq \lambda_n \leq C_7$.or (ii) $p_1 = ... = p_N = N^{-1}$and $\lambda_n \to \infty$ but $\lambda_n - \ln N \to -\infty$, or $\lambda_n \to 0$but $N\lambda_n^2 \to \infty$. Then for all$x_N \geq 0$ such that$x_N = o\left(N^{-1/6}\right)$the relations (2.5) and (2.6) are hold true with $R_N = \mu_0$,$A_N$and σ_N^2 defined in (2.8).

References

1. Holst L. (1972). Asymptotic normality and efficiency for certain goodness-of fit tests. Biometrica, v.59,137-145.

2. Morris C. (1975).Central limit theorems for multinomial sums. Ann Stat. v.3, p.165-188

3. Ivchenko G.I. and Mirakhmedov S. A.(1992), On limit theorems for decomposable statistics and efficiency of the corresponding statistical tests. Discrete Math Appl. v.2, 547-562.

4. Ivchenko G.I. and Mirakhmedov S.A.,(1995). Large deviations and intermediate efficiency of the decomposable statistics in multinomial scheme. Math. Methods in Statist. v.**4**, p.294-311.

5. Mann H.B. and Wald A. (1942). On the choice of the number of intervals in the application of the chi-square test. Ann. Math. Statist. v.13, 306-317.

6. Mirakhmedov S.A. (1987). Approximation of the distribution of multi-dimensional randomized divisible statistics by normal distribution.(Multinomial scheme). Theory Probabl. Appl., 32, p.696-707.

7. Mirakhmedov S.M .,Jamalamadaka S.R. and Ibrahim B.M. (2014), On Edgeworth Expansions in Generalized Urn Models. J Theor Probab 27:725–753. DOI 10.1007/s10959-012-0454-z

8. Sirajidinov S.Kh, Mirakhmedov S.A. and Ismatullaev S.A., (1989). Large deviation probabilities for decomposable statistics in a polynomial scheme. Theory Probabl. Appl. v.34, p.706-719

Statistical Inferences with Competing Risks

A.A.Abdushukurov
Moscow State University named after M.V.Lomonosov, Tashkent Branch
abdushukurov1710@gmail.com, a_ abdushukurov@rambler.ru

Abstract: The main goal of this paper is to review of some approaches of specification of competing risks model and basic nonparametric estimation results in the presence of censoring variables.

Key words: competing risks, survival function, hazard rate, random censoring, uniform consistency, strong approximation.

1. Introduction. Competing risks theory is the subdiscipline of survival analysis. The therm *"competing risks"* follows from the trials in which subject (a physical system or an individual) is exposed to two or more mutually exclusive causes of failure or death. Hence subjects eventual failure or death can be attributed to precisely one of the causes. These problems arise quite frequently in reliability life testing, demography and in medical treatments. There exists enormous literature on competing risks analysis (see, for example [2,3,10,22,23,25-29]).

2. Competing Risks Model (GRM) Specification. In survival analysis there are situations where several causes of failure are possible, but only the occurrence of the first of them can be observed. Let's consider an individual with the survival time X be at risk of dying from *"cause of death"* $J_d \in J = 1, ..., k$. Interest then focuses on the joint distribution of (X, J_d), which could be completely specified through the cause-specific hazards (or transition intensities). Then we can discuss competing risks within the framework of multi-state models (see, [10]), which has *"one transient state 0: alive"* and k absorbing states $j = 1, \dots k$ (see, Figure 1), corresponding to *"death from cause* j ".

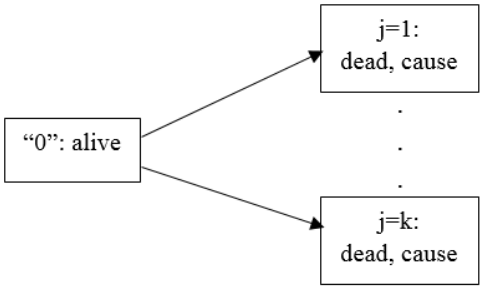


Figure 1. The competing risks multi-state model

Through $h_j(t)$ we define a cause-specific hazard function for the j^{th} cause the rate of occurrence of j^{th} failure:

$$h_j(t) = \lim_{\Delta t \downarrow 0} \frac{P(X < t + \Delta t, J_d = j / X \geq t)}{\Delta t}, \qquad j = 1, 2 \ldots k. \tag{2.1}$$

Corresponding cumulative incidence function, also referred to as subdistribution function is

$$F_j(t) = P(X \leq t,\ J_d = j) \quad j = 1, 2 \ldots k. \tag{2.2}$$

Described process is Markovian, and hazard function $h_j(t)$ denote the transition intensities from state 0 to state j, $j = 1, ..., k$. The total hazard $h(t)$ and the overall survival function $S(t) = P(X>t) = 1 - F(t)$ for X are defined, respectively, in terms of the specific hazards as follows:

$$h(t) = \lim_{\triangle t \to 0} \frac{P(X < t + \triangle t / X \geq t)}{\triangle t} = \sum_{j=1}^{k} h_j(t),$$

$$S(t) = exp\left\{ - \int_0^t h(u)du \right\} = exp\left\{ - \sum_{j=1}^{k} \int_0^t h_j(u)du \right\}. \tag{2.3}$$

The distribution function (d.f.) for X is obtained from the subdistribution function as

$$F(t) = \sum_{j=1}^{k} F_j(t). \tag{2.4}$$

The second, failure time approach to competing risks one supposes the existence of k failure times $X_1, ..., X_k$ for each individual and for observation is available the smallest failure time and corresponding cause $J_d : X = \min(X_1, ..., X_k)$, $J_d = arg\{\min(X_1, ..., X_k)\}$.

Our interest focuses on the joint survival distribution

$$Q(t_1, ..., t_k) = P(X_1 > t_1, ..., X_k > t_k). \tag{2.5}$$

Then, the marginal survival distribution of X, is $S(t) = Q(t, ..., t)$ and the cause-specific hazards are given by

$$h_j(t) = -\frac{\partial log Q(t_1, \ldots, t_k)}{\partial t_j} \, | t_1 = \cdots = t_k = t \,, \quad j = 1, 2, \ldots k \tag{2.6}$$

It follows from (2.1) and (2.3) that the survival function $S(t)$ can be factorized in to k functions $S_j^*(t) = exp\left\{-\int_0^t h_j(u)\,du\right\}$ as

$$S(t) = \prod_{j=1}^{k} S_j^*(t), \tag{2.7}$$

where functions $S_j^*(t)$ are not the survivor functions of any observable random variables (r.v.-s) and moreover, $S_j^*(t) \neq 1 - F_j(t) = S_j(t)$. In classical survival analysis the survival function $S(t)$ often used to describe r.v. X, but in the competing risks framework, it seems more suitable to use the cumulative incidence functions $F_j(t)$ instead of cause-specific survival function $\tilde{S}_j(t) = P(X > t,\ J_d = j) = p_j - F_j(t)$, where $p_j = P(J_d = j) = \lim\limits_{t\to\infty} F_j(t)$, and consequently we have alternative decomposition of $S(t)$: $S(t) = \sum_{j=1}^{k} \tilde{S}_j(t)$. This shows that $S_j^*(t)$ and $\tilde{S}_j(t)$ are different functions. However, in the case of *"independent competing risks"* (i.e. when the failure times $X_1, \ldots, X_k$ are independent) the joint survival distribution (2.5) defined as

$$Q(t_1, \ldots, t_k) = \prod_{j=1}^{k} P(X_j > t_j) = \prod_{j=1}^{k} S_j^*(t_j), \tag{2.8}$$

and hence $S_j^*(t) = P(X_j > t) = exp\left\{-\int_0^t h_j(u)\,du\right\}$ is the survival function of r.v. X_j, $j = 1, \ldots, k$. Thus, in general dependent competing risks situation we have three cause-specific survival-like functions $S_j^*(t)$, $S_j(t)$ and $\tilde{S}_j(t)$, which satisfy some of the mathematical properties of survival function, but they are not proper survival function (see, [25]). Authors [10,25] also investigated some problems arising in dependent CRM with not identifiability properties of joint distributions. They noted that joint d.f. (or survival function) of r.v.-s $X_1, ..., X_k$ is not identifiable based unique on the observed data. Indeed, two joint d.f.-s may be result of into to same marginal for (X, J_d). Moreover, for two different joint distributions, the likelihood functions of the cause-specific hazards can be same. Only under strong assumptions such as independence of risks the multivariable distribution is identifiable. In nonparametric estimation problems in CRM instead of independence assumption on competing risks one can consider the property of *"quasindependence"* . Authors [22,23] have considered condition an independence for the diagonal of the joint distribution of the vector $(X_1, ..., X_k)$ through of equalities:

$$\lambda_j(t) = h_j(t) \text{for all} \, t \geq 0, i = 1, ..., k, \tag{2.9}$$

where

$$\lambda_j(t) = \lim_{\Delta t \downarrow 0} \frac{P(X_j < t + \Delta t / X_j \geq t)}{\Delta t} \quad , \quad j = 1, 2, \ldots k \tag{2.10}$$

is the hazard function of r.v. X_j (see, also [11,14]). Thus, the assumption that risks are independent can be relaxed to weaker condition (2.9). Moreover, Horvath [19] dropped the independence conditions in CRM under some necessary identifiability conditions and demonstrated that asymptotic properties of nonparametrical estimates of survival functions of risks still hold.

3. Estimation and Asymptotical Results for Competing Risks. At present there are a number of papers on nonparametrical estimation in CRM. The weak approximation of the life expectancy process under CRM is studied by Yang [29]. Langberg et.al. [23] proposed the strong consistent estimators for marginal survival functions of risks. The strong approximation results for the product-limit and exponential-hazard vector processes in CRM was proved by Burke et.al [12,13]. Csorgo S. [15] proved universal Gaussian approximation for these processes uniformly up to same large order statistics in the sample, with the approximation rates depending on the order of these statistics (see also [24]). In [1-3,11,14] authors have considered the nonparametrical estimation oh hazard rates and density functions of competing risks. In [2,3] Abdushukurov has considered the nonparametrical Bayesian estimation of survival functions in CRM. In [2,3,7,8] Abdushukurov proposed a new estimators of relative-risk power types for survival functions of risks and proved several asymptotic results for estimates. These estimators are have some advantages with respect to product-limit and exponential-hazard estimators. In [2,3] author established the local approximate normality results for likelihood ratio statistics in CRM. In particular, in proportional hazards model (PHM) of competing risks Abdushukurov [1] proposed a new estimators of survival functions and showed that they are asymptotically effective with respect to other estimators. Extention of PHM under dependent competing risks has been considered by authors [21].

4. Competing Risks under Censoring. In many practical situations one can consider a population in which each subject (physical system or individual) exposing to k mutually exclusive competing risks besides may be not always completely observable. On other words, minima $X = min(X_1, ..., X_k)$ of competing risks may be also subjected to a random censoring. For example, if subject comes under observation some random time L after the time origin, then r.v. X is censored from the left and observed only in the case of $X \geq L$.

this phenomenon is called a delayed entry. There is a close connection wetween the CRM and right censoring of competing risks by r.v. Y , where X is observed only if $X \leq Y$. When competing risks are censored from both sides , we observe X only if $L \leq X \leq Y$ and we observe $j-th$ risks X_j only under validity of product event $\{J_d = j\} \cap \{L \leq X \leq Y\}$.

Hence, presence of censoring variables apart from the competing risks created a more difficulties in problems of nonparametrical estimation of functionals of survival functions of risks. These problems were considered in the case of left censoring by authors [4-6,16-20] .In [6] author considered product-limit and exponential-hazard type two estimators for conditional survival functions of risks. He proved strong approximation results for estimators with best rate O ($n^{-\frac{1}{2}}logn$) (with respect to rate $\Theta(n^{-\lambda})$ with $\lambda = 45000$ in [20]).

In monographies [2,3] author proposed and investigated a number of new estimators in several models of incomplete data. He considers statistical models, where competing risks subjected to random censoring from the one side (left or right) or from both sides. In these models: it proposed a basical functionals of exponential, product and power structures and investigated its properties; by using these functionals the nonparametrical estimators of survival functions of competing risks constructed in generalized incomplete sample models; for nonparametrical estimators its strong uniform consistency properties proved and strong approximation by sequence of Gaussian processes with optimal rate also established ; author also constructed and investigated a new estimators when the censoring is variable and a dependent. Some of these results were considered in case when sample size is random.

The property of local approximaty normality of likelihood ratio statistics when competing risks are random censored from both sides proved by Abdushukurov and Nurmuhammedova [9].

Remark. It is necessary to note that among an enormous works devoting to competing risks theory sometimes one can meet such, authors of which not distinguished the competing risks from censoring variables. So, in order to analysing some results of monograph [2] , author [18] admits confusion. Without taking into account the role of censoring variables , he asserts that : "*... classical CRM and the CRM under random cencoring from both sides are coincides...*"

This assertsion is equivalent to the statement that " *there are not differences between the complete and censored from both sides datas*" , which certainly is not true. Indeed, the great differencies between the classical CRM and CRM under presence of several schemas of censorship has been demonstrated by many

of authors including also a deeper results [2,3].It is also necessary to note that in [18] the authors opinion in abstract and in content of paper are different. This makes difficult to understanding the meaning of it.

REFERENCE

1. Abdushukurov A.A. On Efficiency of Statistical Estimators under Random Censoring of Observations and a New Estimators in Proportional Hazards Model. PhD. 1988. Tashkent. 151 pages (In Russian).

2. Abdushukurov A.A. Statistics of Incomplete Observations: Asymptotic Theory of Estimation for Nonclassical Models: National University of Uzbekistan. University press. Tashkent. 2009. 269 pages (In Russian).

3. Abdushukurov A.A. Estimates of Unknown Distributions from Incomplete Observations and Their Properties. LAMBERT Academic Publishing. Germany. 2011. 299 pages (In Russian).

4. Abdushukurov A.A. Competing Risks Model by Random Censoring from the Left. Pakistan J.Statist. 1996. V.121(1). p. 31-40.

5. Abdushukurov A.A. Karimkulov G.D Nonparametric Kernel Estimation of Density Function under Random Censorship from Both Sides. Pakistan J. Statist. 1996. v.12(1). p. 23-29.

6. Abdushukurov A.A. Competing Risks Model with Random Censoring on the Left. J.Italian Statist. Soc. 1997. v.6 N.1 p.1-21.

7. Abdushukurov A.A. Nonparametric Estimation of the Distribution Function Based on Relative-Risk Function. Commun. Statist: Theory & Methods 1998. v.27 N.8. p.1991-2012

8. Abdushukurov A.A. On Nonparametric Estimation of Reliability Indices by Censored Samples. Theory Rrobab. Appl 1999. V.43 n.1 p.116-122.

9. Abdushukurov A.A., Nurmuhamedova N.S. Local Approximate Normality of Likelihood Ratio Statistics in Competing Risks Model under Random Censorship from Both Sides. Far East J. Theory. Statist. 2013. v.42. N.2. p. 107-122.

10. Andersen P.K Competing Risks as a Multi-state Model. Stat.Meth.& Med.Res. 2002. v.11. p.203-2015.

11. Burke M.D. Approximation of Some Hazard Rate Estimators in a Competing Risks Model Stoch Proc. Appl. 1983. v.14. p. 157-174.

12. Burke M.D., Csörgő S., Horvâth L. Strong Appoximation of Some Biometric Estimates under Random Censorship . J.Wahr verw. Geb. 1981. v.56. p. 87-112.

13. Burke M.D, Csörgő S, Horvâth L. A Correction to and Improvement of "Strong Approximation of Some Biometric Estimates under Random Censorship". Probab. Theory Relat.Fields 1988. v.79. p. 51-57.

14. Burke M.D., Horvath L. Density and Failure Rate Estimation in a Competing Risks Model Sankhya. 1984, v.46. Ser. A. p. 135-154.

15. Csörgő S. Universal Gaussian Approximations under Random Censorship. Ann.Statist. 1996 v.24 N.6 p.2744-2778.

16. Csörgő S., Horvâth L. Random censorship from the Left. Carleton Math. Lect Note. 1981. N.30 p.175-183.

17. Dhar D. Estimation of Survival Function under Random Censorship. Calcutta Statistical Association Bulletin. 1992. v.42. N.165-166. p. 75-86.

18. Formanov Sh.K About of One Model of Statistical Structure. Uzbek Math. J. 2015. N.4. p.154-159. (In Russian)

19. Horvâth L. Dropping Continuity and Independence Asymptions in Random Censorship Models. Carleton Math Lect. Note. 1981. N.30, p. 225-239.

20. Horvâth L. On Random Censorship from Both Sides. Statistics. 1984. v.15. N.4 p.581-594.

21. Kochar S.C., Proschan F. Independence of Time and Cause Failure in the Multiple Dependent Competing Risks Model. Statist.Sinica 1991.v.1.p.295-299.

22. Langberg N., Proschan F., Quinzi A.J. Converting Dependent Models Into Independents Ones, Preserving Essential Features. Ann.Probab. 1978.v.6.p.174-181.

23. Langberg N., Proschan F., Quinzi A.J. Estimating Dependent Life Length, with Application to the Theory of Competing Risks. Ann.Statist .1981.v.9.p.157-167.

24. Njamen-Njomen D.A., Ngatchou-Wandji J. Nelson-Aalen and Kaplen-Meier Estimators in Competing Risks. Applied Mathematics.2014.v.5. p.765-776.

25. Porta N., Gomez G., Calle M.L. The Role of Survival Functions in Competing Risks. DR 2008106.28, May. 2008. 21 pages.

26. Sun Y., Tiwari R.C. Competing Cause Specific Hazard Rates of a Competing Risks Model with Censored Data. Analysis of Censored Data IMS Lecture Notes-Monograph Series.1995.v.27.p.255-270.

27. Tsiatis A.A. A Non Identifiability Aspects of the Problem of Competing Risks. Proceedings of the National Academy of Sciences in USA. 1975.v.72.p.20-22.

28. Tsiatis A.A. Competing Risks. Encyclopedia of Biostatistics. 1998.v.1.p.824-834.

29. Yang G. Life Expectancy under Random Censorship. Stoch.Proc.Appl. 1977.v.6.p.33-39.

Оценка среднего значения числа выходов случайного блуждания за подвижную линейную границу

М.У.Гафуров
Филиал Московского Государственного Университета им. М.В. Ломоносова в г. Ташкент

Данная работа посвящена дальнейшему анализу теорем Клесова, Гута-Штейнебаха, Хе-Кси и др. об асимптотических оценках среднего значения числа выходов случайного блуждания за линейную границу с малым параметром:

Пусть $\{\zeta_n,\, n \geq 1\}$ последовательность случайных величины (с.в), $\nu(\varepsilon) = \sum\limits_{n=1}^{\infty} I(|\zeta_n| > \varepsilon)$, где $I(A)$ - индикатор событья A, $\varepsilon > 0$ - произвольный параметр. $\nu(\varepsilon)$ характеризует число выходов траектории $\{\zeta_n,\, n \geq 1\}$ за границу $\pm\varepsilon$.

В 1947 г. Хсу и Роббинс [1] ввели определение комплектной сходимости $\{\zeta_n,\, n \geq 1\}$ к нулю.

Определение [1]. Последовательность с.в. $\{\zeta_n,\, n \geq 1\}$ сходится к нулю комплектно, если $E\nu(\varepsilon) < \infty$, $\forall\varepsilon > 0$. В случае $\zeta_n = \frac{S_n}{n}$, где $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$, $\{\xi_n,\, n \geq 1\}$ - последовательность независимых, одинокого распределенных с.в. результаты [1] и Эрдеша [2] позволяют сформулировать знаменитую теорему Хсу-Роббинс-Эрдеша (ХРЭ), которой в этом году исполняется равно 70 лет.

Теорема (ХРЭ). $E\nu(\varepsilon) < \infty\ \forall\varepsilon > 0 \Leftrightarrow E\xi_1 = 0,\ E\xi_1^2 < \infty$.

Отметим, что если $\varepsilon \downarrow 0$, то $E\nu(\varepsilon) \uparrow \infty$, и следовательно, представляет интерес задача нахождения точного роста $E\nu(\varepsilon)$ при $\varepsilon \downarrow 0$. Первая попытка решения этой задачи осуществлено в 1970 г. Сливкой и Северо [3] для случая, когда с.в. ξ_1 имеет нормальное распределение с параметрами 0 и σ^2.

Ими установлено, что $\forall\varepsilon > 0$

$$\frac{\sigma^2}{\varepsilon^2} - 1 \leq E\nu(\varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Отсюда, при $\varepsilon \downarrow 0$

$$\varepsilon^2 E\nu(\varepsilon) \to \sigma^2 \tag{1}$$

В 1975 г. Хейди [4] доказал изящную теорему о том, что соотношение (1) справедливо и в общем случае независимых, одинокого распределенных с.в. с $E\xi_1 = 0$ и $E\xi_1^2 < \infty$.

Теорема (Хейди). $E\xi_1 = 0$, $E\xi_1^2 < \sigma^2$, тогда $\lim\limits_{\varepsilon \downarrow 0} \varepsilon^2 E\nu\left(\varepsilon\right) = \sigma^2$.

Дальнейшие обобщения теоремы Хейди и связанные с ней в идейном отношении близкие результаты полученные в различных направлениях, приведены в работах Чэна [5], Гута, Спатару [6], Ланзингера и Штадтмюллера [7], Гафурова [8], Сираждинова и Гафурова [9] (более подробную информацию по литературе до 1987 г. можно найти в [9]), Клесова [10], [11], Булдыгина, Клесова и Штейнебаха [12], Гута и Штейнебаха [13] и др.

Список публикаций, основанных на результатах [4], можно продолжить, он достаточно длинный и географически широкий.

Однако, естественно возникающая проблема об оценка скорости сходимости по ε в теореме Хейди до 1994 г. оставалось не исследованной. Первый результат по данной проблеме установлен Клесовым [14].

Теорема (Клесов). Пусть $\{\xi_n,\, n \geq 1\}$ последовательность независимых, одинаково распределенных с.в. с нулевым математическим ожиданием и $E\xi_1^2 = \sigma^2 > 0$, $E|\xi_1|^3 < \infty$. Тогда при $\varepsilon \downarrow 0$

$$\varepsilon^2 E\nu\left(\varepsilon\right) - \sigma^2 = o\left(\varepsilon^{1/2}\right). \tag{2}$$

Этот результат вызвал определенный интерес среди специалистов Швеции, Германии, Румынии, Китая, Узбекистана и др., которые занимались дальнейшим уточнением и ослаблением моментных условий [14], [6].

Отметим следующий результат Гута и Штейнебаха [13], где условие $E|\xi_1|^3 < \infty$, $0 < \delta \leq 1$.

Теорема (Гут, Штейнебах). Пусть $\{\xi_n,\, n \geq 1\}$- последовательность независимых, одинаково распределенных с.в. $0 < \delta \leq 1$. Если $E\xi_1 = 0$, $E\xi_1^2 = \sigma^2 > 0$ и $E|\xi_1|^{2+\delta} < \infty$, $0 < \delta \leq 1$, то при $\varepsilon \downarrow 0$

$$\varepsilon^2 E\nu(\varepsilon) - \sigma^2 = \begin{cases} O(\varepsilon),\ \delta = 1, \\ \\ o(\varepsilon^{\delta}),\ 0 < \delta < 1. \end{cases}$$

Рассмотрим следующие характеристики

$$\rho = \sup_{z>0} \left\{ \left| \int\limits_{-z}^{z} u^3 dP\left(\xi_1 < u\right) \right| + z \int\limits_{|u|>z} u^2 dP\left(\xi_1 < u\right) \right\},$$

$$\lambda_\delta = \sup_{z>0} z^\delta \int\limits_{|u|>z} u^2 dP\left(\xi_1 < u\right),$$

введенные Эссеном [15].

Отметим, что условие $\rho < \infty$ может выполняться и в случае $E|\xi_1|^3 = +\infty$, а условие $\lambda_\delta < \infty$ влечет лищь $E\xi_1^2 < \infty$.

Один из основных результатов данной работы сводится к следующей теореме.

Теорема. Если $E\xi_1 = 0$, $E\xi_1^2 = \sigma^2 > 0$, то

$$\varepsilon^2 E\nu(\varepsilon) - \sigma^2 = \begin{cases} O(\varepsilon), \ \rho < \infty, \\ \\ o(\varepsilon^\delta), \ \lambda_\delta < \infty, \ 0 < \delta < 1. \end{cases}$$

Доказательство утверждения теоремы и другие связанные с ними результаты будут опубликованы в ближайшее время.

ЛИТЕРАТУРА

1. P.L.Hcu, H.Robbince (1947) Complete convergence and the law of large numbers. Proc. Nat. Acad. Sci. USA, 2, pp. 25-31.

2. P.Erdos (1949) On a theorem of Hsu and Robbins. Ann. Math. Statist. 20, 2, pp.286-291; Remark on my paper YOn a theorem of Hsu and RobbinsY, Ann. Math. Statist. 21, 1950, p. 138.

3. J.Slivka, N.Severo (1970) On the strong law of large numbers, Proc. Amer. Math. Soc., 24, p. 729-734.

4. C.C.Heyde (1975) A supplement to strong law of large numbers. J. Appl. Probab. 12, 1, pp. 903-907.

5. R.Chen (1976) A remark on the strong law of large numbers. Proc. Amer. Math. Soc., 61, 1, pp. 112-116.

6. A.Gut, A.Spataru (2000) Precise asymptotics in the Baum-Katz and Davis law of large numbers. J. Math. Appl. 248, pp. 233-246.

7. H.Lanzinger, U.Stadtmuller (2004) Refined Baum-Katz laws for weighted sums of i.i.d. random variables. Stat. Probab. Let. 69, pp. 357-368.

8. М.У.Гафуров (1981) Оценки для вероятностных характеристик выхода случайного блуждания за криволинейную границу. Диссертация на соиск. учен. степени доктора физ.-мат. Наук. Ташкент, 250 стр.

9. С.Х.Сираждинов, М.Гафуров (1987) Метод рядов в граничных задачах для случайных блужданий. Ташкент, Из-во. Фан. 138 стр.

10. O.Klesov (1985) Strong law of large numbers for multiple sums of i.i.d. random variables. Math. Notes. pp. 1006-1014.

11. O.Klesov (1995) Complete convergence of randomly indexed sums of random variables. J.Math.Sci., 76, 2, pp. 2241-2249.

12. V.Buldygin, O.Klesov and J.Steinebach (2007) Precise asymptotics over a small parameter for a series of large deviation probabilities. Theor. Of Stochas. Processes. 13,29, №1-2, pp. 44-56.

13. A.Gut, J.Steinebach (2012) Convergence rates in Precise asymptotics. J.Math.Appl., 390, pp. 1-14.

14. O.Klesov (1994) On a convergence rate in a theorem of Heyde. Theor.Probab. and Math. Statist. 49, pp. 83-87.

15. C.Essen (1968) On the remainder term in the central limit theorem, Arkiv for Matematic Band. 2, №2, pp. 7-15.

Центральная предельная теорема для линейных процессов, порожденных последовательностью независимых случайных величин

Т.М.Зупаров
Tm_zuparov@rambler.ru

Пусть $\{\xi_{kn}, k \in \mathbb{Z}, k \le n, n \ge 1\}-$ последовательность серий случайных величин, $M\xi_{kn} = 0,\ D\xi_{kn} = \sigma_{kn}^2 < \infty,\ k = 0, \pm 1, ...$ удовлетворяет следующему условию (A):

(A) для любого $\varepsilon > 0$ справедливо соотношение $\sup\limits_{k\in\mathbb{Z}} P\{|\xi_{kn}| > \varepsilon\} \xrightarrow[n\to\infty]{} 0.$

Определение 1. Если ряд $X_{kn} = \sum_{i=-\infty}^{\infty} a_{in}\xi_{k-i,n}$ сходится с вероятностью 1, тогда последовательность случайных величин $\{X_{kn}, k \in \mathbb{Z}\}$ называется линейным процессом с коэффициентами $\{a_{kn}, k \in Z, n \ge 1\}$, порожденным инновационным процессом $\{\xi_{kn}, k \in Z\}$.

Предположим, что коэффициенты линейного процесса удовлетворяют условию

(B) $\sum_{j=-\infty}^{\infty} a_{jn}^2 < \infty$, $\sup\limits_{k\in\mathbb{Z}} \frac{b_{kn}^2}{B_n^2} \le C$, где $C-$ число не зависящий от n, где $b_{kn} = a_{1-k,n} + a_{2-k,n} + ... + a_{n-k,n}$, и $B_n^2 = \sum_{k\in\mathbb{Z}} b_{kn}^2\sigma_{kn}^2 < \infty,\ \forall n = 1, 2, 3, ...$

Работа посвящена доказательству центральной предельной теоремы для линейного процесса, порожденного последовательностью независимых в каждой серии случайных величин с конечными вторыми моментами, удовлетворяющих условию равномерной бесконечной малости. С помощью представления суммы n первых членов линейного процесса в виде ряда получен аналог теоремы Линдеберга-Феллера для линейных процессов.

Первый результат в этом направлении принадлежит И.А.Ибрагимову (см. [1], гл. XVIII). В этой работе доказана центральная предельная теорема, в случае, когда линейный процесс порожден последовательностью независимых одинаково распределенных случайных величин, коэффициенты которого не зависят от n и удовлетворяют условиям $\sum_{i=-\infty}^{\infty} a_i^2 < \infty$, и $B_n^2 \to \infty$ (при $n \to \infty$). Работа P.C.B.Philips и V.Solo [2] посвящена асимптотическому анализу распределения суммы линейного процесса, порожденного независимой одинаково распределенной последовательностью случайных величин с коэффициентами не зависящими от n. В этой работе, авторы, используя БН- разложение (определения БН- разложения см в [3]) линейного процесса, доказали центральную предельную теорему и принцип инвариантности при выполнения условий $\sum_{k=1}^{\infty} k^2 a_k^2 < \infty$ и $a_{kn} = 0,\ k > 0$. В работе [4] доказана центральная предельная теорема и получена нерав-

номерная оценка скорости сходимости для линейных процессов, порожденных независимыми и одинаково распределенными центрированными случайными величинами с единичными дисперсиями, удовлетворяющих условиям $\sum_{k=-\infty}^{\infty} a_{kn}^2$ и $\sup\limits_k \left|\frac{b_{kn}}{B_n}\right| = o(1)$ при $n \to \infty$.

В работе [5] изучен асимптотический анализ распределения суммы первых n членов линейного процесса, порожденного последовательностью $\{\xi_{kn}, k \in \mathbb{Z}\}$, удовлетворяющему следующему (более сильному, чем условие (A)) условию (A*):

(A*) для любого $\varepsilon > 0$ справедливо соотношение $P\left\{\sup\limits_{k\in\mathbb{Z}} |\xi_{kn}| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$.

В этой работе, при довольно жестких условиях на коэффициенты линейного процесса, с помощью БН-разложения доказан критерий слабой сходимости нормированной суммы линейного процесса к предельному распределению и как следствие этого результата получен аналог теоремы Линдеберга – Феллера для линейных процессов, порожденных $\varphi-$ перемешивающейся инновационной последовательностью. Справедливости центральной предельной теоремы для линейных процессов, порожденных не обязательно независимой последовательностью случайных величин, посвящены статьи [6] - [9].

В данной работе, с помощью представления первых n членов линейного процесса в виде $S_n = \sum_{k=-\infty}^{\infty} b_{kn}\xi_{kn}$, доказана следующая центральная предельная теорема.

Теорема. Пусть выполнены условия (A) и (B). Тогда для справедливости соотношения

$$\frac{1}{B_n}\sum_{k=1}^{n} X_{kn} \Rightarrow \Phi(x),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du$, необходимо и достаточно выполнение условия Линдеберга: для любого $\varepsilon > 0$ при $n \to \infty$

$$L_n(\varepsilon) = \sum_{k=-n}^{n} E\xi_{kn}^2 I\left\{|\xi_{kn}| > \varepsilon\right\} \to 0,$$

где $I\{A\}-$ индикатор события A.

ЛИТЕРАТУРА

1. И. А.Ибрагимов, Ю. В.Линник. Независимые и стационарно связанные величины. - М.:Наука, 1965, 524 стр.

2. P. C. W.Phillips, V.Solo. Asymptotics for linear processes, Ann. Statist., v.20, No2, 1992, p.971 - 1001.

3. Beveridje S., and Nelson, C. R. A new approach to decomposition of economic time series into permanent and transitory components wiyth particular attention to measurement of the "business cycle". Journal of monetary Economics, 7, p.151-174.

4. Т. М.Зупаров, Ш. Т.Зупаров. Центральная предельная теорема для линейных процессов. УзМУ хабарлари 2010, No3, с.79 - 86.

5. Т. М.Зупаров. Асимптотический анализ распределения суммы линейного процесса. Материалы республиканской научной конференции "Современные проблемы теории вероятностей и математической статистики Ташкент, 2019, с.22 - 24.

6. Peligrad M. and Utev S. Central limit theorem for linear processes. Ann. Probab., 1997, 25(1), p.443 - 456.

7. Peligrad M. and Utev S. Central limit theorem for stationary linear processes. Ann. Probab., 2006, 34(4), p.1608 - 1626.

8. Wu W. B. Central limit theorem for functional of linear processes and their applications. Statist, Sinica, 2002, 12, p.635 - 649.

9. Wu W. B. and Min W. On linear processes with dependent innovations. Stohastic Process. Appl, 115, p.939 - 958.

Об оценках объема нулей голоморфной функции, зависящей от комплексного параметра

А. Садуллаев[1], А.М. Кытманов[2]
[1]Национальный университет Узбекистана
[2] Сибирский федеральный университет
sadullaev@mail.ru, kytmanov49@gmail.com

Аннотация: В работе для голоморфной функции $f(\sigma, z)$, $\sigma \in \mathbb{C}^m$, $z \in \mathbb{C}^n$, дается равномерная по σ оценка объема нулей относительно множества $\left\{z \ : f\left(\sigma^0, z\right) = 0\right\}$. Такие оценки очень полезны в вопросах изучения осцилляторных интегралов $[J(\lambda, \sigma) = \int_{\mathbb{R}^n} a(x, \sigma)\, e^{i\lambda\Phi(x,\sigma)} dx,]$ при $\lambda \to \infty$. Здесь $a(x, \sigma) \in C_0^\infty(\mathbb{R}^n \times \mathbb{R}^m)$ – так называемая амплитудная функция и $\Phi(x, \sigma)$ – функция фазы.

Ключевые слова: Подготовительная теорема Вейерштрасса, аналитическое множество, обыкновенные точки, объем аналитического множества, теорема Виртингера

Если функция $f(\sigma, z)$ – голоморфна от переменных $\sigma \in \mathbb{C}^m$ и $z \in \mathbb{C}$ в окрестности нуля $(0, 0)$, $f(0, z) /\equiv 0$, то согласно подготовительной теоремы Вейерштрасса, в некоторой окрестности $U' \times V'M0$ функция $f(\sigma, z)$ представляется в виде

$$f(\sigma, z) = \left[z^k + c_1(\sigma) z^{k-1} + ... + c_k(\sigma)\right] \varphi(\sigma, z), \qquad (1)$$

где $k \geq 0$, $c_j(\sigma) \in O(V')$, $\varphi(\sigma, z) \in O(U' \times V')$ и $\varphi(\sigma, z) \neq 0$.

Поэтому, число нулей функции $f\left(\sigma^0, z\right)$ в окрестности V' по переменной z равно k для любой фиксированной $\sigma^0 \in V'$. В случае, когда $f(0, z) \equiv 0$, можно ожидать представление

$$f(\sigma, z) = \left[c_0(\sigma) z^k + c_1(\sigma) z^{k-1} + ... + c_k(\sigma)\right] \varphi(\sigma, z), \qquad (2)$$

где $k \geq 0$, $c_j(\sigma) \in O(V')$, $\varphi(\sigma, z) \in O(U' \times V')$ и $\varphi(\sigma, z) \neq 0$, и доказать, что число нулей функции $f\left(\sigma^0, z\right)$ в V' по переменной z не превосходит k для любой фиксированной $\sigma^0 \in V'$. При $m = 1$ такое представление действительно имеет место, ибо мы можем заранее выразить функцию $f(\sigma, z)$, $f(0, z) \equiv 0$, в виде $f(\sigma, z) = \sigma^p \tilde{f}(\sigma, z)$, $p > 0$, $\tilde{f}(0, z) /\equiv 0$ и получит требуемое утверждение.

Ситуация значительно осложняется, когда $m > 1$, $f(0, z) \equiv 0$. Как показывает пример Осгуда, в этом случае обобщенная подготовительная теорема Вейерштрасса (2) не всегда имеет место. Более того в окрестности точки

$0 \in \mathbb{C}^m$ число нулей функции $f\left(\sigma^0, z\right)$, априори, может неограниченно возрастать при подходе к $\sigma = 0$. Чтобы наглядно представить ситуацию, мы приведем, несколько видоизменив, пример Осгуда.

Пример Осгуда [1]. Берем в единичном круге $|\xi| < 1$ функцию $q(\xi) = \xi + \sum_{k=2}^{\infty} \xi^{k!}$. Она голоморфно не продолжается вне единичной окружности, т.е. каждая точка окружности $|\xi| = 1$ является особой точкой для $q(\xi)$. В некоторой окрестности нуля функция $z = q(\xi)$ имеет обратную, $\xi = q^{-1}(z)$. Функция трех переменных $f(\sigma_1, \sigma_2, z) = \sigma_1 q^{-1}(z) - \sigma_2$ - голоморфна в некоторой окрестности $(0,0,0) \in \mathbb{C}^3$ и $f(0,0,z) \equiv 0$.

Предположим теперь, что в некоторой окрестности нуля f представляется в виде (2). Тогда ее нули

$$c_0(\sigma_1, \sigma_2) z^k + c_1(\sigma_1, \sigma_2) z^{k-1} + ... + c_k(\sigma_1, \sigma_2) = 0$$

образуют псевдоалгебраическое множество. С другой стороны нулевым множеством функции f является график $z = q\left(\frac{\sigma_2}{\sigma_1}\right)$, а функция $z = q\left(\frac{\sigma_2}{\sigma_1}\right)$ не может аналитически продолжаться вне конуса $\left|\frac{\sigma_2}{\sigma_1}\right| < 1$.
Имеет место

Теорема 1. Если функция $f(\sigma, z) -$ голоморфна от переменных $\sigma \in \mathbb{C}^m$ и $z \in \mathbb{C}$ в окрестности замыкания поликруга $W = U(0,r) \times \{|z| < r\} \subset \mathbb{C}^{m+1}$, $r > 0$, то число корней $N_f(\sigma)$, с учетом кратностей, функции $f(\sigma, z)$ по переменной z при фиксированном σ равномерно ограничена относительно параметра $\sigma \in U$. т.е. $\exists M < \infty :\ \ N_f(\sigma) \leq M$ для любого фиксированного $\sigma \in U :\ \ f(\sigma,\, z) \not\equiv 0$.

Теорема 1 означает, что число нулей функции $f\left(\sigma^0, z\right)$ равномерно ограничено на открытом множестве

$$U \backslash \Im_f,\ \ N_f(\sigma) \leq C\ \forall \sigma \notin \Im_f,$$

где $\Im_f = \left\{\sigma^0 \in U :\ \ f\left(\sigma^0, z\right) \equiv 0\right\}$, $C-$ константа, не зависящая от σ, а $N_f(\sigma) -$ число нулей с учетом кратностей, функции $f(\sigma, z)$ по переменной z, при фиксированном $\sigma \in U$. В дальнейшем нам удобно считать $N_f\left(\sigma^0\right) = -1$, если $f\left(\sigma^0,\, z\right) \equiv 0$. Тогда Теорема 1 означает, что $\exists M < \infty :\ \ N_f(\sigma) \leq M\ \forall \sigma \in U$.

Справедлив также многомерный аналог Теоремы 1. В связи с тем, что его доказательство существенно отличается от доказательства Теоремы 1, хотя использует некоторые еч моменты, мы приведем его отдельно

Теорема 2. Пусть функция $f(\sigma, z) -$ голоморфна от переменных $z \in \mathbb{C}^n$ и голоморфно зависит от параметра $\sigma \in \mathbb{C}^m$ в окрестности замыкании

некоторого поликруга $W = U \times V \subset \mathbb{C}^{m+n}$. Тогда существует число $M < \infty$ такое, что

$$Vol_{2n-2}\left(Z_{f(\sigma^0,z)} \bigcap (U \times V)\right) \leq M, \ \ \forall \sigma^0 \in U: \ \ f\left(\sigma^0, z\right) \not\equiv 0, \qquad (3)$$

где $Z_{f(\sigma^0,\,z)} = \left\{z \in V: \ \ f\left(\sigma^0, z\right) = 0\right\}$ – аналитическое множество коразмерности 1.

Отметим, что равенства типа (3) очень полезны в вопросах изучении осцилляторных интегралов

$$J\left(\lambda, \sigma\right) = \int_{\mathbb{R}^n} a\left(x, \sigma\right) \, e^{i\lambda\Phi(x,\sigma)} dx,$$

при $\lambda \to \infty$. Здесь $a\left(x, \sigma\right) \in C_0^\infty\left(\mathbb{R}^n \times \mathbb{R}^m\right)$ – так называемая амплитудная функция и $\Phi\left(x, \sigma\right)$ – функция фазы (см. [2-4]).

ЛИТЕРАТУРА

1. W. Osgood, Lehrbuch der Funktionentheorie, Bd.II, Teubner, Leipzig, 1929, page 90.
2. V. I. Arnol'd, S. M. Gusein-Zade, A. N. Varchenko, Singularities of differentiable maps. Vol. I. The classification of critical points, caustics and wave fronts. Monographs in Mathematics, 82. Birkhduser Boston, Inc., Boston, MA, 1985.
3. I. A. Ikromov, Damped oscillatory integrals and maximal operators. Mat. Zametki 78 (2005), no. 6, 833-852; translation in Math. Notes 78 (2005), no. 5-6, 773-790.
4. I.A. Ikromov, Sh.A. Muranov, Estimates of oscillatory integrals with a damping factor, Math. Notes, V.104(2018), no.2, 218-230.
5. A. Sadullaev, Criteria algebraicity of analytic sets, Functional analysis and its application, V.6(1), (1972), 85-86.
6. M. Herve, Several Complex Variables, Oxford University Press, Bombay, 1963.
7. B.V. Shabat, Introduction to Complex Analysis, Part II, AMS, 1992.
8. Кытманов А.М., Интеграл Бохнера-Мартинелли и его применения. Новосибирск, Наука, 1991.

9. Е.М. Чирка, Комплексные Аналитические Множества. Москва, Наука, 1985.

A. Sadullaev, A.M. Kytmanov

On estimates of the volume of zeros of a holomorphic function depending on a complex parameter

For holomorphic function $f(\sigma, z)$, $\sigma \in \mathbb{C}^m$, $z \in \mathbb{C}^n$, we give an uniform by σ estimates of the zero set $\{z \ : f(\sigma, z) = 0\}$, $\sigma - fixed$. Such kind of estimates are very useful in the study of oscillatory integrals
$J(\lambda, \sigma) = \int_{\mathbb{R}^n} a(x, \sigma)\, e^{i\lambda\Phi(x,\sigma)} dx$, as $\lambda \to \infty$.
Here $a(x, \sigma) \in C_0^\infty(\mathbb{R}^n \times \mathbb{R}^m)$ is, so called amplitude function and $\Phi(x, \sigma)$ is faze function.

Key words. Weierstrass preparation theorem, analytic set, regular points, volium of analytic set, Wirtinger theorem

RQ-система M|M|1 с отрицательными заявками

М.И.Гаман, С.П.Моисеева
НИ Томский государственный университет
gamanm@yandex.ru

Введение

В настоящее время в мировой литературе находят широкое отражение многочисленные исследования по теории массового обслуживания и ее приложениям, в частности в теории телетрафика и телекоммуникационных систем. В монографиях российских и зарубежных ученых дается подробный обзор современных приложений моделей массового обслуживания в области телекоммуникаций, современных компьютерных сетей и информационных систем. В качестве математических моделей телекоммуникационных систем и компьютерных сетей связи используются различные классы систем массового обслуживания в том числе наибольшей популярностью пользуются модели систем с повторными вызовами (Retrial Queueing). В работах Р.И. Вилкинсона, Дж.В. Коэна, К. Весоловски, А.С. Альфа, Г. Гоштони, Б.Д. Чоя , Геленбе, Степанов С.Н., Дудина А.Н. и др. показано, что модели с повторными вызовами (RQ-системы) широко применяются для проектирования и оптимизации информационно-коммуникационных систем различного уровня (локальных , глобальных), цифровых сетей связи, управляемых протоколами случайного множественного доступа, а также сетей сотовой связи, вычислительных кластеров и др. [1-10]. RQ-системы — это системы, содержащие источники повторных вызовов. В данных системах массового обслуживания поступающие заявки не теряются при отсутствии доступных для обслуживания приборов (например, занятых или сломанных). Наличие источника повторных вызовов означает, что когда заявка застала прибор занятым, она уходит в состояние ожидания прибора, а затем снова пытается занять прибор спустя некоторое случайное время. В реальной жизни источник повторных вызовов имеется, например, у call-центров, когда оператор занят и абонент пытается до него дозвониться. Исследованию RQ-систем посвящено большое количество работ, наиболее обширный обзор значимых результатов представлен в монографиях Artalejo J. R., Gomez-Corral A. G.I. Falin, J.G.C. Templeton [5-6]. Имеющиеся на сегодняшний день научные публикации в данной области предлагают достаточно много различных моделей и подходов к их анализу. Конфигурация современных RQ систем учитывает различные особенности, такие как:

нетерпеливость заявок, различные протоколы доступа, наличие конфликтов, СМО с отрицательными заявками. Из современных исследователей можно назвать Дудина А.Н., Назарова А.А., Морозова Е.В., Dragieva, V., Phung-Duc T., Choi D., Krishnamoorthy A., Kim B. и др.

Рассмотрение отрицательных заявок в системе был предложен Э. Геленбе. Также системы, содержащие отрицательные заявки, принято называть G-системами или G-сетями [11-12]. Особенностью отрицательных заявок является то, что они не требуют обслуживания, оказывая прямое врияние на функционирование системы. В случае попадания отрицательной заявки в систему, она уничтожает положительные заявки. Заявки могут уничтожаться как на приборе, так и в источнике повторных вызовов. В случае уничтожения заявок в источнике повторных вызовов, может уничтожаться одна заявка, в некоторых случаях уничтожается группа заявок, а в некоторых случаях уничтожаются все заявки. В данной работе рассматривается ситуация, когда при попадании отрицательной заявки в систему данная заявка покидает систему, в случае, когда система свободна, а в случае, когда система занята, отрицательная заявка покидает систему, уничтожая при этом одну заявку в источнике повторных вызовов. RQ-системам с отрицательными заявками уделяли такие исследователи как Дж. Арталехо, Я.Шин, В. Анисимов [13, 14].

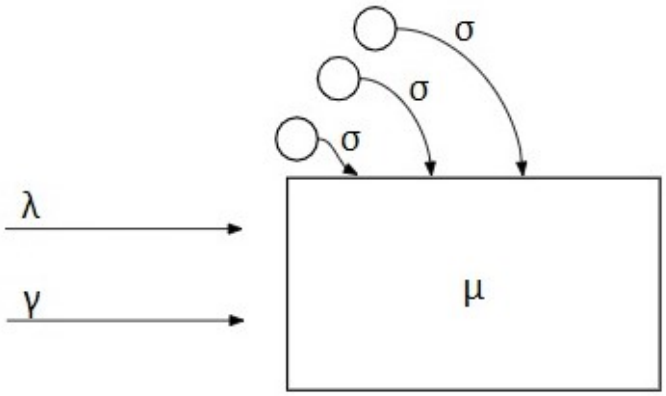
Несмотря на то, что в реальных системах имеется несколько обслуживающих приборов, в данной работе рассматривается классическая задача, где в системе находится один обслуживающий прибор. Рассматривается система M|M|1, в которой на вход поступает простейший поток заявок с постоянной интенсивностью, время обслуживания имеет экспоненциальное распределение.

Постановка задачи

Рассмотрим однолинейную RQ-систему (рис. 1), на вход которой поступает простейший поток заявок с интенсивностью λ. Будем называть заявки данного потока положительными. Когда положительная заявка поступает на прибор, то занимают его в случае, если он свободен. Время обслуживания каждой положительной заявки распределено по экспоненциальному закону с параметром μ. В случае, если прибор занят, заявка переходит в источник повторных вызовов (орбита), где осуществляет случайную задержку продолжительностью, распределенной по экспоненциальному закону с параметром σ. После случайной задержки на орбите заявка вновь обращается к прибору с повторной попыткой его захвата. Если прибор свободен, заявка

занимает его и обслуживается, в противном случае заявка возвращается на орбиту для реализации следующей задержки.

Также в систему поступает простейший поток отрицательных заявок с интенсивностью γ. Если отрицательная заявка застает прибор свободным, то она уходит из системы. Если отрицательная заявка застает прибор занятым, то она и идет в источник повторных вызовов и уничтожает одну случайную заявку.



1-рисунок

Пусть $i(t)$ - число заявок в источнике повторных вызовов, $k(t)$ - определяет состояние прибора следующим образом: Обозначим $P\{k(t) = k, i(t) = i = P(k, i, t)\}$ Ц вероятность того, что прибор в момент времени t находится в состоянии k и в источнике повторных вызовов находится i заявок. Процесс $\{k(t), i(t)\}$ изменения состояний во времени является марковским.

Ставится задача нахождения распределения вероятностей числа заявок на орбите.

Система дифференциальных уравнений Колмогорова

Для заданных вероятностей $P_k(i,t)$ составим систему дифференциальных уравнений Колмргорова, которая в стационарном режиме будет иметь следующий вид:

$$\begin{cases} -(\lambda + i\sigma)\Pi_0(i) + \mu\Pi_1(i) = 0 \\ \\ -(\lambda + \mu + \gamma)\Pi_1(i) + \lambda\Pi_0(i) + \lambda\Pi_1(i-1) + (i+1)\sigma\Pi_0(i+1) + \gamma\Pi_1(i+1) = 0 \end{cases} \tag{1}$$

Введем производящие функции вида:

$$F_0(z) = \sum_{i=0}^{\infty} z^i \Pi_0(i) \text{ и } F_1(z) = \sum_{i=0}^{\infty} z^i \Pi_1(i),$$

с помощью которых составим систему для нахождения функций $F_0(z)$ и $F_1(z)$ следующего вида:

$$\begin{cases} F_1(z) = F_0(z)(\dfrac{\lambda}{-\lambda z + \mu + \gamma}) \\ \\ \sigma F_0^{'}(z) = (\lambda + \mu + \gamma)F_1(z) - \lambda F_0(z) - \lambda z F_1(z) - (\gamma/z)F_0(z) \end{cases} \tag{2}$$

Решив данную систему, получили функции $F_0(z)$ и $F_1(z)$:

$$F_0(z) = C \left(\frac{z}{-\lambda z + \mu + \gamma} \right)^{\frac{-\lambda\gamma}{\sigma(\mu+\gamma)}} \tag{3}$$

$$F_1(z) = \frac{\lambda}{-\lambda z + \mu + \gamma} C \left(\frac{z}{-\lambda z + \mu + \gamma} \right)^{\frac{-\lambda\gamma}{\sigma(\mu+\gamma)}} \tag{4}$$

В данном случае константа C имеет вид:

$$C = \left(\frac{1}{-\lambda + \mu + \gamma} \right)^{\frac{\lambda\gamma}{\sigma(\mu+\gamma)}} \frac{-\lambda + \mu + \gamma}{\mu + \gamma} \tag{5}$$

Таким образом, учитывая то, что $F(z) = F_0(z) + F_1(z)$, производящая функция имеет следующий вид:

$$F(z) = \frac{-\lambda + \mu + \gamma}{\mu + \gamma} \left(\frac{1}{-\lambda + \mu + \gamma} \right)^{\frac{\lambda\gamma}{\sigma(\mu+\gamma)}} \left(\frac{z}{-\lambda z + \mu + \gamma} \right)^{\frac{-\lambda\gamma}{\sigma(\mu+\gamma)}} \left(1 + \frac{\lambda}{-\lambda + \mu + \gamma} \right) \tag{6}$$

Обозначим $\frac{\lambda}{\mu+\gamma} = k$, тогда производящая функция будет иметь вид:

$$F(z) = (1-k)\frac{1-kz}{(1-k)z)}^{\frac{\lambda\gamma}{\sigma k}} \frac{1-k(z-1)}{1-kz} \tag{7}$$

Таким образом, получили допредельное выражение для производящей функции числа заявок на орбите. Полученная производящая функция позволяет определить основные характеристики исследуемого процесса i.

ЛИТЕРАТУРА

1. Wilkinson R.I. Theories for toll traffic engineering in the USA // The Bell System Technical Journal. 1956. V. 35. No. 2. P. 421Ц507.

2. Cohen J.W. Basic problems of telephone trafic and the influence of repeated calls // Philips Telecommunication Review. 1957. V. 18. No. 2. P. 49Ц100.

3. Gosztony G. Repeated call attempts and their efect on trafic engineering // Budavox Telecommunication Review. 1976. V. 2. P. 16Ц26.

4. Elldin A., Lind G. Elementary Telephone Trafic Theory. Stockholm: Ericsson Public Telecommunications, 1971.

5. Artalejo J.R., Gomez-Corral A. Retrial Queueing Systems. A Computational Approach. Stockholm: Springer, 2008.

6. Falin G.I., Templeton J.G.C. Retrial queues. London: Chapman & Hall, 1997.

7. Artalejo J.R., Falin G.I. Standard and retrial queueing systems: A comparative analysis // Revista Matematica Complutense. 2002. V. 15. P. 101Ц129.

8. Roszik J., Sztrik J., Kim C. Retrial queues in the performance modelling of cellular mobile networks using MOSEL // International Journal of Simulation. 2005. No. 6. P. 38Ц47.

9. Kuznetsov D.Yu., Nazarov A.A. Analysis of non-Markovian models of communication networks with adaptive protocols of multiple random access // Automation and Remote Control. 2001. No. 5. P. 124Ц146.

10. Aguir S., Karaesmen F., Askin O.Z., Chauvet F. The impact of retrials on call center performance // OR Spektrum. 2004. No. 26. P. 353Ц376.

11. Gelenbe E. Random neural networks with positive and negative signals and product form solution // Neural Computation. 1989. vol. 1. no. 4. pp. 502-511. DOI: 10.1162/neco.1989.1.4.502

12. Gelenbe E. Product-form queueing networks with negative and positive customers // Journal of Applied Probability. 1991. vol. 28. pp. 656-663. DOI: 10.2307/3214499

13. Shin Y. W. Multi-server retrial queue with negative customers and disasters // Queueing Systems. 2007. vol. 55. no. 4. pp. 223-237. DOI: 10.1007/s11134-007-9018-9

14. Anisimov V.V., Artalejo J.R. Analysis of Markov multiserver retrial queues with negative arrivals // Queueing Systems. 2001. vol. 39. no. 2/3. pp. 157-182. DOI: 10.1023/A:1012796517394

Gaman Margarita Igorevna, Moiseeva Svetlana Petrovna (Tomsk, Russia)

M|M|1 System with negative requests

Abstract. *We consider a single-line RQ-system with negative requests, the input of which receives a simple flow of applications with a constant intensity, the service time has an exponential distribution. The orbital delay time also has an exponential distribution. Each request that gets into the source of repeated calls tries to occupy the device at random intervals. A negative request, coming into the system with a free service device, leaves it. If the device is busy, the request also leaves the system, thus destroying one request from orbit. The problem of finding the probability distribution of the number of requests in orbit is posed. To solve the problem, the differential equations of Kolmogorov in the stationary mode are composed. The generating function is also found.*

Key words: RQ-system, orbit, source of repeated calls, negative requests

Поток повторных обращений в двухфазных системах массового обслуживания с неограниченным числом приборов при нестационарном режиме работы

М.А.Шкленник, А.Н.Моисеев
НИ Томский государственный университет
shklennikm@yandex.ru

Введение

Системы массового обслуживания с неограниченным числом приборов широко используются в качестве математических моделей различных социально-демографических, экономических и других процессов [1-3]. Системы с возможностью обращения заявки для повторного обслуживания — системы с обратной связью — используются для описания процессов дообслуживания при решении задач в теории телетрафика [4]. Однако, большинство работ, посвященных исследованию потоков в таких системах проводится в предположении, что система работает в стационарном режиме [5-7].

В данной работе проведено исследование потока повторных обращений в двухфазной системе массового обслуживания с неограниченным числом приборов на каждой фазе и обратной связью на второй фазе при нестационарном режиме работы.

Математическая модель

Рассмотрим двухфазную систему массового обслуживания с неограниченным числом приборов на каждой фазе и возможностью повторного обращения ко второй фазе. На вход системы поступает простейший поток заявок с интенсивностью λ. Каждая поступающая заявка мгновенно занимает свободный прибор на первой фазе, где обслуживается в течение случайного времени. Завершив обслуживание на первой фазе системы, заявка с вероятностью r_1 может перейти на вторую фазу (это обращение будем считать повторным), или с вероятностью $(1 - r_1)$ может покинуть систему (завершить свое обслуживание). Время обслуживания заявки на второй фазе системы является случайной величиной, имеющей отличные от времени обслуживания на первой фазе параметры распределения. Завершив обслуживание на второй фазе системы, каждая заявка может с вероятностью r_2 может вернуться на вторую фазу системы для следующего повторного обслуживания, или с вероятностью $(1 - r_2)$ может покинуть систему.

В работе ставится задача исследования r-потока — потока повторных обращений в системе с момента начала ее функционирования, то есть нахождения распределения вероятностей числа событий r-потока, наступивших за определенный интервал времени, начиная с $t_0 = 0$. При этом мы предполагаем, что в начальный момент времени система пуста и в ней нет обслуживаемых заявок.

R-поток в системе $M/M/\infty$

Рассмотрим описанную выше систему массового обслуживания, предположив, что время обслуживания заявок является экспоненциально распределенной случайной величиной. Пусть параметр экспоненциального распределения на первой фазе равен μ_1, а на второй фазе - μ_2. Для решения поставленной задачи будем использовать метод производящей функции. Рассмотрим трехмерный марковский процесс $\{i_1(t), i_2(t), n(t)\}$, где $i_1(t)$ Ц число приборов, занятых обслуживанием заявок на первой фазе в момент времени t; $i_2(t)$ Ц число приборов второй фазы, занятых обслуживанием заявок в момент времени t; $n(t)$ Ц число событий r-потока, то есть число обращений заявок ко второй фазе системы за интервал времени $[0, t]$. Для распределения вероятностей исследуемого процесса $P(i_1, i_2, n, t) = P\{i_1(t) = i_1, i_2(t) = i_2, n(t) = n\}$ запишем дифференциальное уравнение Колмогорова

$$\frac{\partial P(i_1, i_2, n, t)}{\partial t} = -(\lambda + i_1\mu_1 + i_2\mu_2)P(i_1, i_2, n, t) + \lambda P(i_1 - 1, i_2, n, t) +$$

$$+(i_1 + 1)\mu_1 r_1 P(i_1 + 1, i_2 - 1, n - 1, t) + (i_1 + 1)\mu_1(1 - r_1)P(i_1 + 1, i_2, n, t) +$$

$$+i_2\mu_2 r_2 P(i_1, i_2, n - 1, t) + (i_2 + 1)\mu_2(1 - r_2)P(i_1, i_2 + 1, n, t). \qquad (1)$$

с начальными условиями

$$P(i_1, i_2, n, 0) = \begin{cases} 1, \ i_1 = i_2 = n = 0, \\ \\ 0, \ i_1 \neq 0, \ i_2 \neq 0, \ n \neq 0. \end{cases} \qquad (2)$$

Определим трехмерную производящую функцию в виде

$$F(x, y, z, t) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \sum_{n=0}^{\infty} x^{i_1} y^{i_2} z^n P(i_1, i_2, n, t).$$

Учитывая свойства производящей функции, несложно из уравнения (1) получить дифференциальное уравнение в частных производных для производящей функции

$$\frac{\partial F(x,y,z,t)}{\partial t}+\mu_1\left(x-r_1yz-(1-r_1)\right)\frac{\partial F(x,y,z,t)}{\partial x}+$$

$$+\mu_2\left(y-r_2yz-(1-r_2)\right)\frac{\partial F(x,y,z,t)}{\partial y}=\lambda(x-1)F(x,y,z,t) \qquad (3)$$

с начальным условием

$$F(x,y,z,0)\equiv 1. \qquad (4)$$

Решая уравнение (3) с начальным условием (4), находим вид производящей функции для трехмерного процесса $\{i_1(t),i_2(t),n(t)\}$

$$F(x,y,z,t)=\exp\left\{\frac{\lambda r_1(z-1)}{1-r_2z}t+\frac{\lambda(x-1)}{\mu_1}\left(1-e^{-\mu_1 t}\right)-\frac{\lambda r_1(z-1)}{\mu_1(1-r_2z)}\left(1-e^{-\mu_1 t}\right)+\right.$$

$$+\frac{\lambda r_1 z}{\mu_2(1-r_2z)-\mu_1}\cdot\left(y-\frac{1-r_2}{1-r_2z}\right)\left(1-e^{-\mu_1 t}\right)-$$

$$\left.-\frac{\lambda r_1\mu_1 z}{\mu_2(1-r_2z)\left(\mu_2(1-r_2z)-\mu_1\right)}\cdot\left(y-\frac{1-r_2}{1-r_2z}\right)\left(1-e^{-\mu_2(1-r_2z)t}\right)\right\}.$$

Тогда положив $y=1,z=1$, получим производящую функцию для одномерного процесса $\{n(t)\}$

$$F(z,t)=\exp\left\{\lambda r_1(z-1)\left[\frac{t}{1-r_2z}-\frac{(\mu_2-\mu_1)\left(1-e^{-\mu_1 t}\right)}{\mu_1\left(\mu_2(1-r_2z)-\mu_1\right)}+\right.\right.$$

$$\left.\left.+\frac{\mu_1 r_2 z\left(1-e^{-\mu_2(1-r_2z)t}\right)}{\mu_2(1-r_2z)^2\left(\mu_2(1-r_2z)-\mu_1\right)}\right]\right\}.$$

Учитывая связь производящей функции с характеристической функцией, можем записать выражение для характеристической функции процесса $\{n(t)\}$

$$H(u,\,t)=\exp\left\{\lambda r_1(e^{ju}-1)\left[\frac{t}{1-r_2e^{ju}}-\frac{(\mu_2-\mu_1)\left(1-e^{-\mu_1 t}\right)}{\mu_1\left(\mu_2(1-r_2e^{ju})-\mu_1\right)}+\right.\right.$$

$$\left.\left.+\frac{\mu_1}{\mu_2}\cdot\frac{r_2e^{ju}}{(1-r_2e^{ju})^2}\cdot\frac{\left(1-e^{-\mu_2(1-r_2e^{ju})t}\right)}{\mu_2(1-r_2e^{ju})-\mu_1}\right]\right\}.$$

R-поток в системе $M/GI/\infty$

Найдем характеристическую функцию распределения вероятностей r-потока для системы, в которой время обслуживания заявок задано произвольной функцией распределения $B_1(x)$ на первой фазе, и $B_2(x)$ - на второй.

Для решения поставленной задачи будем использовать метод предельной декомпозиции [8]. Входящий поток заявок, интенсивность которого равна λ, разделим на N независимых потоков по полиномиальной схеме с равными вероятностями. Интенсивность каждого такого потока будет равна λ/N. Так как число приборов на фазах системы неограниченно, определим для заявок каждого из полученных потоков единственную однолинейную двухфазную линию обслуживания с отказами. В этой системе заявки, поступающие в период времени, когда хотя бы одна фаза линии занята, не обслуживаются. При $N \to \infty$ вероятностью потерь заявок можно пренебречь, и тогда характеристики исходной системы будут сходиться к суммарным характеристикам совокупности N однолинейных двухфазных СМО.

В однолинейной системе рассмотрим многомерный случайный процесс $\{n(t, N)\}$, описывающий число повторных обращений, реализованных за время t в системе. Введем в рассмотрение дополнительные переменные: процесс $k(t)$ Ц состояние линии обслуживания, то есть если k-я фаза линии занята, то $k(t) = k, k = 1, 2$; и если линия свободна, то $k(t) = 0$; $z(t)$ Ц длина интервала времени от момента t до момента окончания текущего обслуживания, если линия занята.

Процесс $\{k(t), n(t, N), z(t)\}$ является марковским. Распределение вероятностей этого процесса обозначим следующим образом: $P_0(n, t, N) = P\{k(t) = 0, n(t, N) = n\}$ Ц вероятность того, что в момент времени t линия свободна и к этому моменту к системе повторно обратились n заявок; $Pk(n, t, z, N) = P\{k(t) = k, n(t, N) = n, z(t) < z\}, k = 1, 2$ Ц вероятность того, что в момент времени t занята k-я фаза системы, к этому моменту к системе повторно обратилось n заявок, и остаточное время обслуживания заявки, находящейся в системе, меньше z.

Определим частичные характеристические функции в виде

$$h_0(u, t, N) = \sum_{n=0}^{\infty} e^{jun} P_0(n, t, N);\ h_k(u, t, z, N) = \sum_{n=0}^{\infty} e^{jun} P_k(n, t, z, N),\ k = 1, 2.$$

Тогда, составив систему дифференциальных уравнений Колмогорова для распределения вероятностей, получим систему дифференциальных урав-

нений для характеристических функций

$$\frac{\partial h_0(u,t,N)}{\partial t} = (1-r_1)\frac{\partial h_1(u,t,0,N)}{\partial z} + (1-r_2)\frac{\partial h_2(u,t,0,N)}{\partial z} - \frac{\lambda}{N}h_0(u,t,N), \tag{8}$$

$$\frac{\partial h_1(u,t,z,N)}{\partial t} - \frac{\partial h_1(u,t,z,N)}{\partial z} = -\frac{\partial h_1(u,t,0,N)}{\partial z} + \frac{\lambda}{N}B_1(z)h_0(u,t,N), \tag{9}$$

$$\frac{\partial h_2(u,t,z,N)}{\partial t} - \frac{\partial h_2(u,t,z,N)}{\partial z} = -\frac{\partial h_2(u,t,0,N)}{\partial z} + r_1 B_2(z)e^{ju}\frac{\partial h_1(u,t,0,N)}{\partial z} + r_2 B_2(z)e^{ju}\frac{\partial h_2(u,t,0,N)}{\partial z}, \tag{10}$$

с начальными условиями

$$h_0(u,0,N) \equiv 1;\ h_1(u,0,z,N) = h_2(n,0,z,N) \equiv 0.$$

Решая систему уравнений (8)-(10) с заданными начальными условиями, найдем вид функций $h_0(u,t,N)$, $h_1(u,t,z,N)$ и $h_2(u,t,z,N)$. Тогда, характеристическая функция $h(u,t,N)$ случайного процесса $\{n(t,N)\}$ в двухфазной однолинейной системе с повторными обращениями

$$h(u,t,N) = h_0(u,t,N) + h_1(u,t,\infty,N) + h_2(u,t,\infty,N) =$$

$$= 1 + \frac{1}{N}\left(e^{ju}-1\right)\left(r_1\int\limits_0^t f_1(u,x)dx + r_2\int\limits_0^t f_2(u,x)dx\right) + o(N^{-2}),$$

где

$$f_1(u,x) = \lambda B_1(x);\ f_2(u,x) = \frac{\lambda r_1 e^{ju}}{2\pi}\int\limits_{-\infty}^{+\infty}\frac{b_1^*(\alpha)b_2^*(\alpha)\left(1-e^{-j\alpha x}\right)}{j\alpha\left(1-r_2e^{ju}b_2^*(\alpha)\right)}d\alpha,$$

$$b_k^*(\alpha) = \int\limits_0^\infty e^{j\alpha x}d_xB_k(x), k = 1,2.$$

В силу независимости однолинейных систем, рассмотренных выше, характеристическая функция $H(u,t)$ процесса $\{n(t)\}$, описывающего поток повторных обращений в двухфазной системе с неограниченным числом линий, определяется выражением

$$H(u,t) = \lim_{N\to\infty} h(u,t,N)^N.$$

С учетом выражения для функций $f_1(u,x)$ и $f_2(u,x)$ окончательно получаем

$$H(u,\,t) = \exp\left[\lambda r_1(e^{ju}-1)\left\{\int\limits_0^t B_1(x)dx + \right.\right.$$

$$\left.\left. + \frac{r_2 e^{ju}}{2\pi}\int\limits_{-\infty}^{+\infty} \frac{b_1^*(\alpha)b_2^*(\alpha)}{j\alpha\left(1-r_2e^{ju}b_2^*(\alpha)\right)}\left(t-\frac{1-e^{-j\alpha t}}{j\alpha}\right)d\alpha\right\}\right].$$

Заключение

Итак, в результате проведенного исследования получены выражения для характеристических функций распределения вероятностей числа повторных обращений заявок за интервал времени $[0,t]$ в двух типах СМО - в системе с экспоненциально распределенным временем обслуживания $M/M/\infty$ и в системе с произвольно распределенным временем обслуживания $M/GI/\infty$. По виду полученных выражений можно утверждать, что исследуемый r-поток не является пуассоновским.

Используя свойства характеристических функций несложно найти все основные вероятностные характеристики исследуемого потока.

ЛИТЕРАТУРА

1. Жидкова Л.А., Моисеева С.П. Математическая модель потоков покупателей двухпродуктовой торговой компании в виде системы массового обслуживания с повторными обращениями к блокам. // Известия Томского политехнического университета. Том 322, є6, 2013. Томск: Изд-во ТПУ, 2013. с.5 Ц 9.

2. Shklennik M., Moiseeva S., Moiseev A. Optimization of Two-Level Discount Values Using Queueing Tandem Model with Feedback //CCIS. 2018. Vol. 912. P. 321-332.

3. Моисеева С.П., Захорольная И.А. Математическая модель параллельного обслуживания кратных заявок с повторными обращениями // Автометрия Ц Новосибирск: Изд-во СО РАН, 2011г. Т. 47. є6. С. 51-58.

4. Coffman E.G., Kleinrock L.: Feedback queueing models for time-shared systems. Journal of the Assoeiation for Computing Machinery, Vol. 15, No. 4, October 1968, pp. 549-576.

5. Морозова А. С., Моисеева С. П. Исследование потока повторных обращений в бесконечнолинейной СМО с повторным обслуживанием // Вестник Томского государственного университета. Ц 2005. Ц є 287. Ц С. 46Ц51.

6. Моисеева С.П., Шкленник М.А., Набокова О.О. Исследование потоков в двухфазной бесконечнолинейной СМО с повторными обращениями методом предельной декомпозиции //Информационные технологии и математическое моделирование (ИТММ-2017) : материалы XVI Междунар. конф. имени А. Ф. Терпугова, 29 сент. - 3 окт. 2017 г. Ч. 1. Томск: Изд-во НТЛ, 2017. С. 108-114.

7. 7. Pekoz, E.A., Joglekar, N.: Poisson traffic flow in a general feedback. J. Appl. Probab. 39(3), 630Ц636 (2002)

8. Назаров А.А., Моисеева С.П., Морозова А.С. Исследование СМО с повторным обращением и неограниченным числом обслуживающих приборов методом предельной декомпозиции // Вычислительные технологии. 2008. Т. 13. є 55. С.88 Ц 92.

Shklennik Maria Alexandrovna (Tomsk, Russia), Moiseev Alexander Nikolaevich (Tomsk, Russia)

Repeat flow in infinite-server queueing tandem

Abstract. *The paper presents two types of mathematical models of queueing tandem with an unlimited number of servers and feedback: $M/M/\infty$ and $M/GI/\infty$. The service times of primary and repeated customers have different parameters. The expression for the characteristic probability distribution function of the number of repeated customers during fixed time period is found.*

Key words: queueing tandem, repeat flow, characteristic function

Исследование бесконечнолинейных СМО с групповым обслуживанием

А.А.Бояркина
НИ Томский государственный университет
boyarkina.93@inbox.ru

Введение

Системы с групповым обслуживанием находят применение для решения широкого спектра научно-прикладных задач. Разнообразие проблем, стоящих перед исследователями, привело к разработке самых различных модификаций таких СМО - со всевозможными типами входящих потоков и дисциплинами обслуживания, приоритетами, конечной очередью, ненадчжным прибором и др. Считается, что начало исследованиям СМО с групповым обслуживанием положено в работах Н. Бейли и Ф. Даунтона [1,2]. Дж. Миллер [3] впервые рассмотрел СМО с групповым поступлением и групповым обслуживанием, в которой длины групп заявок, поступающих в систему или на обслуживающий прибор, являются дискретными случайными величинами. Для произвольных распределений входящего потока и времени обслуживания он нашчл условие эргодичности очереди. При пуассоновском поступлении заявок методом вложенных цепей Маркова Дж. Миллер исследовал время ожидания и период занятости, а для случая экспоненциального обслуживания - характеристики стационарного распределения очереди. Н. Жайсвал [4] предложил модель СМО, в которой длина группы заявок, поступающей на обслуживание, зависит от числа заявок, уже обслуживаемых прибором, кроме того, прибор может обслуживать группы нулевой длины. Для исследования этой системы он также использовал метод вложенных цепей Маркова. П. Финч [5] исследовал СМО с конечной очередью и экспоненциальным обслуживанием группами фиксированной длины. Дж. Гойял в [6] методом производящих функций исследовал систему Hm/M/1 с гиперэкспоненциальным входящим потоком и экспоненциальным обслуживанием группами переменной длины. Результаты этих и многих других работ изложены в монографиях Дж. Коэн [7] и Дж. Темплетоном в соавторстве с М. Шодри [8]. Далее многие авторы обращались к исследованию СМО с групповым обслуживанием заявок, привнося огромное количество модификаций. Обзор современного состояния проблемы исследования рассмотрен в работах [9,10]. Говоря об эволюции методов исследований, необходимо заметить, что в большинстве работ используются тесно связанные между собой методы вложенных цепей Маркова и

производящих функций. Хорошо известно, что относительная их простота нивелируется ограниченными возможностями анализа характеристик системы. Метод введения дополнительной переменной лишҷн указанного недостатка, но высокая сложность реализации этого метода для систем с групповым обслуживанием обуславливает весьма редкое его применение. К настоящему времени наиболее изученными являются однолинейные СМО с групповым обслуживания различной конфигурации. Для многоканальных и бесконечнолинейных немарковских СМО общего подхода не выработано.

В данной статье демонстрируется применение метода динамического просеивания для решения задачи исследования бесконечнолинейных систем массового обслуживания с групповым обслуживанием заявок.

Постановка задачи

Пусть S Ч открытая система массового обслуживания с неординарным входящим потоком и групповым обслуживанием (Рис.1). Из неординарности входящего потока и группового обслуживания следует, что исследуемая система обладает отличительной особенностью. Данная особенность заключается в том, что поступившая на обслуживание группа заявок после обслуживания в том же составе покидает систему.

Рассматриваемая система массового обслуживания имеет неограниченное число обслуживающих приборов и является системой без отказов в обслуживании.

На вход системы поступает простейший пуассоновский неординарный поток заявок (группы заявок). Интенсивность прихода групп λ. Размер группы дискретная случайная величина, заданная рядом распределения $P\{\xi = i\} = v_i$.

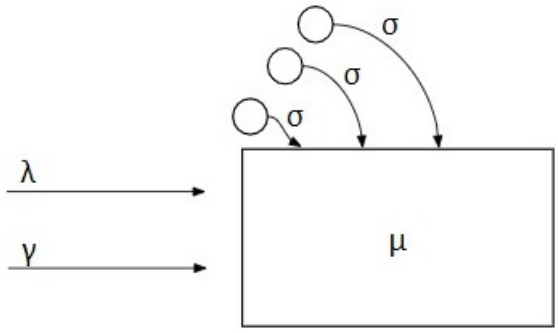
Входящая группа заявок занимает любые из свободных приборов, где обслуживается в течении случайного времени $\tau > 0$ с функцией распределения вероятностей $B(x) = P\{\tau < x\}$, а по окончании обслуживания вся группа освобождает все приборы одновременно.

Ставится задача исследования числа занятых приборов в системе. Сложность аналитического исследования числа заключается в том, обязательное условие, что все заявки группы покидают систему одновременно. Следовательно, возникает необходимость отслеживать также и число поступивших групп. Для решения задачи предлагается использовать авторский метод динамического просеивания [11,12] в применении к ресурсным системам [13]. Указанный подход в сочетании с асимптотическими методами в даль-

нейшем позволит исследовать более сложные системы с коррелированными входящими потоками.

Пусть в начальный момент времени t_0 система пуста. Обозначим $K(t)$ Ч число занятых групп в системе в момент времени t, $V(t) = \sum_{i=1}^{K(t)} v_i$ Ч суммарный объем занятых приборов в момент времени $t \geq t_0$.

Необходимо найти стационарное распределение вероятностей двумерного случайного процесса $\{K(t), V(t)\}$.



1-рисунок

Ввиду того, что рассматриваемый двумерный процесс не является марковским, будет использоваться модифицированный метод динамического просеивания, который позволяет учитывать занятые приборы группами, которые находятся в системе к моменту времени t (прошли процедуру просеивания).

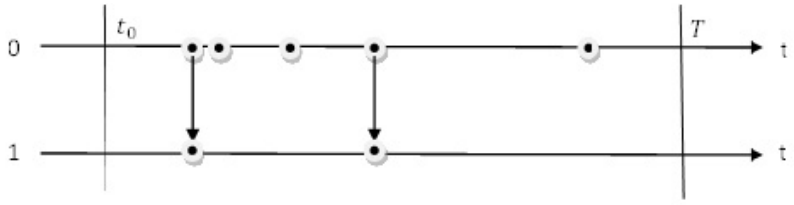
Зафиксируем некоторый момент времени $T > t_0$. Отметим моменты времени наступления событий входящего потока на оси 0 (Рис.2). Далее будем учитывать объемы занятых приборов для групп, которые к моменту времени T не закончат свое обслуживание.

Пусть функция $S(t, T)$ определяет динамическую вероятность просеивания на нижнюю ось, согласно формуле:

$$S(t, T) = 1 - B(T - t) \tag{1}$$

Таким образом, $S(t, T)$ Ч вероятность того, что группа поступившая в систему в момент времени $t < T$, к моменту времени T не закончит обслуживание и будет занимать v приборов.

Следовательно, с вероятностью $1 - S(t, T)$ группа закончит свое обслуживание и освободит приборы, которые занимала при поступлении в систему, поэтому в просеянном потоке она не рассматривается.



2-рисунок

Обозначим $n(t)$ Ч число групп событий просеянного потока , наступивших на промежутке , $W(t)$ Ч общее число занимаемых приборов в просеянном потоке.

Между процессами $n(t)$ и $K(t)$ установлена связь. А именно утверждается, что в момент времени $t = T$ законы распределения вероятностей данных случайных величин совпадают:

$$P\{K(t) = m\} = P\{n(t) = m\} \tag{2}$$

для любых m.

Из данной формулы очевидно следует равенство (3).

$$P\{K(t) = m, V(t) = V\} = P\{n(t) = m, W(t) = V\} \tag{3}$$

Отметим, что в теории равенство (3) называется основной формулой метода динамического просеивания для ресурсных систем массового обслуживания с неограниченным числом приборов.

В данной статье рассматривается применение метода для исследования бесконечнолинейных СМО с групповым обслуживанием.

Исследование бесконечнолинейных СМО с групповым обслуживанием

Для рассматриваемой системы с групповым обслуживанием и дискретным временем построим двумерный случайный процесс $\{n(t), W(t)\}, n, w = 0, 1, ..., \infty$.

Введем обозначение распределение вероятностей:

$$P\{n(t) = n, W(t) = w\} = P(n, w, t) \tag{4}$$

для $n, w = \overline{0, \infty}$. Тогда по формуле полной вероятности запишем равенства:

$$P(n, w, t + \Delta t) = P(n, w, t)(1 - \lambda\Delta t) + P(n, w, t)\lambda\Delta t(1 - S(t)) +$$

$$+ \sum_{k=0}^{w} P(n - 1, w - k, t)v_k\lambda\Delta tS(t) \tag{5}$$

Откуда получаем систему дифференциальных уравнений Колмогорова

$$\frac{\partial P(n,w,t)}{\partial t} = \lambda S(t)\left(\sum_{k=0}^{w} P(n-1,w-k,t)v_k - P(n,w,t)\right) \tag{6}$$

С начальным условием

$$P(n,w,t_0) = \begin{cases} 1, & \text{если } n=0, \\ 0, & \text{иначе,} \end{cases} \tag{7}$$

Введем двумерную производящую функцию вида:

$$F(x,y,t) = \sum_{n=0}^{\infty}\sum_{w=0}^{\infty} x^n y^w P(n,w,t) \tag{8}$$

С учетом производящей функции СДУ (6) перепишется в виде

$$\frac{\partial F(x,y,t)}{\partial t} = \lambda S(t)\left(\sum_{n=0}^{\infty}\sum_{w=0}^{\infty}\sum_{k=0}^{w} x^n y^w P(n-1,w-k,t)v_k - F(x,y,t)\right) =$$

$$= \lambda S(t)\left(x\sum_{n=0}^{\infty}\sum_{w=0}^{\infty}\sum_{k=0}^{w} x^{n-1}y^{w-k}y^k P(n-1,w-k,t)v_k - F(x,y,t)\right) =$$

$$= \lambda S(t)\left(x\sum_{k=0}^{\infty} y^k v_k \sum_{n=0}^{\infty}\sum_{w=k}^{\infty} x^{n-1}y^{w-k} P(n-1,w-k,t) - F(x,y,t)\right) =$$

$$= \lambda S(t)F(x,y,t)(xf(y)-1) \tag{9}$$

где $f(y) = \sum\limits_{k=0}^{\infty} y^k v_k$, которая является производящей функцией числа заявок в группе.

Таким образом, СДУ (6) окончательно примет вид:

$$\frac{\partial F(x,y,t)}{\partial t} = \lambda S(t)F(x,y,t)(xf(y)-1) \tag{10}$$

с начальным условием

$$F(x,y,t_0) = 1 \tag{11}$$

Нетрудно показать, что решение уравнения (10) примет вид:

$$F(x,y,t) = exp\left\{\int_{t_0}^{T} S(u)du(xf(y)-1)\right\} \tag{12}$$

При $t = T$, $t_0 \to -\infty$ для производящей функции двумерного процесса числа групп, находящихся в системе и суммарного объема занятых ресурса приборов в стационарном режиме получим

$$F(x, y) = exp\{\lambda b(xf(y) - 1)\} \tag{13}$$

где

$$b = \int_0^\infty (1 - B(\tau))d\tau \tag{14}$$

Производящая функция суммарного количества занятых приборов в стационарном режиме имеет вид

$$F(y) = exp\{\lambda b(f(y) - 1)\} \tag{15}$$

которая совпадает с результатами, полученными в работах [2].

Очевидно, полагая y = 1, получим производящую функцию распределения вероятностей числа находящихся в системе групп в стационарном режиме

$$F(x) = exp\{\lambda b(x - 1)\} \tag{13}$$

которая совпадает с производящей функцией пуассоновского распределения вероятностей с параметром λb, что не противоречит известным результатам для системы M/GI/∞, полученным ранее.

Заключение

В данной работе была исследована математическая модель бесконечнолинейных систем с групповым обслуживанием и неординарным потоком требований в дискретном времени.

Посредством модифицированного метода динамического просеивания были получены основные характеристики системы. Предлагаемый метод позволит в дальнейшем исследовать системы с более сложными моделями потоков пачек (полумарковскими и марковски модулированнным).

ЛИТЕРАТУРА

1. Bailey N. T. J. On Queueing Processes with Bulk Service // Journal of the Royal Statistical Society, Ser. B. - 1954. - Vol. 16, No 1. - Pp. 80-87.

2. Тихоненко О. М. Модели массового обслуживания в системах обработки информации / О. М. Тихоненко. Ц Минск : Университетское, 1990. Ц 191 с.

3. Rupert G., Miller J. A Contribution to the Theory of Bulk Queues // Journal of the Royal Statistical Society. Ser. B. - 1959. - Vol. 21, No 2. - Pp. 320-337.

4. Jaiswal N. A Bulk-Service Queueing Problem with Variable Capacity // Journal of the Royal Statistical Society. Ser. B. - 1961. - Vol. 23, No 1. - Pp. 143-148.

5. Finch P. D. On the Transient Behavior of a Queueing System with Bulk Service and Finite Capacity // Annals of Mathematical Statistics. - 1962. -Vol. 33, No 3. - Pp. 973-985.

6. Goyal J. K. Queues with Hyper-Poisson Arrivals and Bulk Exponential Service //Metrika. - 1967. - Vol. 11, No 1. - Pp. 157-167.

7. Cohen J. The Single Server Queue. - Amsterdam: North Holland, 1969. J. K. Queues with Hyper-Poisson Arrivals and Bulk Exponential Service //Metrika. - 1967. - Vol. 11, No 1. - Pp. 157-167.

8. Chaudhry M. L., Templeton J. G. C. A First Course in Bulk Queues. - New York: Wiley, 1983

9. Першаков Н. В., Самуйлов К. Е. Системы M/G/1 с групповым обслуживанием. Часть I: обзор методов анализа вероятностных характеристик // Вестник РУДН. Серия .Математика. Информатика. Физика.. - 2009. - Т. 1. - С. 32-44.

10. Першаков Н. В., Самуйлов К. Е. Системы M/G/1 с групповым обслуживанием. Часть I: обзор методов анализа вероятностных характеристик // Вестник РУДН. Серия .Математика. Информатика. Физика.. - 2009. - Т. 2. - С. 43-53.

11. Назаров А. А. Методы асимптотического анализа в теории массового обслуживания / А. А. Назаров, С. П. Моисеева. - Томск : Изд-во НТЛ, 2006. - 112 с.

12. Моисеев Назаров Бесконечнолинейные системы и сети

13. Лисовская Е. Ю. Асимптотический анализ второго порядка для бесконечнолинейной системы массового обслуживания с рекуррентным входящим потоком требований случайного объема / Е. Ю. Лисовская, С. П. Моисеева // Актуальные проблемы прикладной математики и

информационных технологий - аль-Хорезми 2016 : труды международной конференции (9-10 ноября 2016 г., Ташкент). - Ташкент, 2016. - С. 205-208.

Boyarkina Anna Aleksandrovna (Tomsk, Russia)

Research of endless linear QS with group service

Abstract. *We Consider a QS with an extraordinary input stream and group service. The queuing system has an unlimited number of service devices and is a denial-of-service system. At the input of the system, the simplest stream of applications as part of groups of random size is received. The group received by the service in the system knows any free devices. Group service time is exponential. At the end of the service, the group releases all the devices that it occupied. To study the system under consideration, an analytical solution was obtained based on the dynamic sieving method.*

Key words: QS with an extraordinary input stream, group service,dynamic sieving method

Прикладные вопросы классификации с использованием гребневой регрессии

Кононова Н.В.[1] **Мангалова Е.С.**[2] **Чубарова О.В.**[1]

[1] Сибирский федеральный университет 660041, Красноярский край, г. Красноярск, пр. Свободный, 79

[2] ООО "АрДиСайнс 660017, Россия, г. Красноярск, ул. Кирова, д.19, оф.29

E-mail koplyarovanv@mail.ru

Введение. Острый панкреатит остается важной проблемой ургентной абдоминальной хирургии. На протяжении последнего десятилетия по данным разных авторов больные острым панкреатитом составляют 3-10% от общего числа пациентов хирургического профиля [1]. Заболеваемость острым панкреатитом во всем мире варьирует от 4,9 до 73,4 случаев на 100000 населения и по темпам роста опережает все другие неотложные заболевания органов брюшной полости [1]. В течение нескольких лет в ряде регионов РФ эта патология занимает первое место в структуре острой хирургической заболеваемости органов брюшной полости. Летальность при остром панкреатите, как в России, так и за рубежом остачтся высокой (от 12% до 75%) [1].

В соответствии с национальными клиническими рекомендациями по диагностике и лечению острого панкреатита, важным моментом является стратификация степени тяжести заболевания, что является достаточно сложной задачей. Неправильная, субъективная трактовка симптомов, результатов лабораторных и инструментальных методов обследования может привести к недооценке тяжести панкреатита у определенной категории пациентов, а это чревато назначением недостаточно эффективного лечения, что может привести к прогрессированию заболевания.

Широкое применение на госпитальном этапе получила оценка тяжести состояния больного, проводимая с использованием интегральных шкал, которых по различным источникам насчитывается более 20. Однако большинство из них включают сложные критерии, такие как параметры электролитного и газового состава крови, которые не всегда доступны в условиях приемного отделения городских и районных стационаров [1].

Применение в клинической практике современных цифровых технологий позволяет расширить возможности прогнозирования течения и исхода острого панкреатита путем применения современных математических алгоритмов, построенных на основе анализа цифровых баз данных. Таким

образом, актуальной становится задача разработки алгоритмов и создания программных продуктов, которые могут выполнять роль помощников врача на этапе принятия решения, уменьшая риски субъективного восприятия врачом информации о пациенте.

Цель настоящей работы: на основе анализа данных о пациентах выявить значимые показатели с учетом современных представлений о патогенезе и этиологии острого панкреатита для объективного определения тяжести заболевания и прогнозирования исхода.

Общая постановка задачи. Имеется множество историй болезни пациентов с диагнозом острый панкреатит с выставленными степенями тяжести заболевания (классами). Это множество будем называть выборкой. Требуется построить алгоритм, способный классифицировать вновь поступившего пациента с диагнозом острый панкреатит, то есть определить степень тяжести заболевания.

Для решения задачи классификации в первую очередь необходимо выявить значимые показатели (признаки) состояния пациента. Наличие в данных неинформативных (не значимых) признаков приводит к снижению точности решений. Отбор признаков позволит по минимально возможному набору показателей выявить степень тяжести, а значит, назначить своевременно корректное лечение.

Объект исследования: 130 электронных карт стационарного больного, извлеченных из баз данных электронной медицинской системы qMS историй болезни с диагнозом острый панкреатит с 2015 по 2017 г. В исходной выборке содержалось 26 показателей состояния пациента: 11 показателей развернутого, биохимического анализа крови, 4 показателя анализа мочи, 8 показателей ультразвукового исследования (УЗИ) органов брюшной полости, температура пациента, характер кишечной перистальтики, наличие вздутия живота. Семь показателей являлись категориальными переменными (принимали значение 0, 0.5 либо 1). Основная часть категориальных переменных имела значения 0 и 1, то есть наличие или отсутствие признака, а один — вздутие, принимал значения 0, 0.5 и 1, где 0 — отсутствие признака (живот не вздут), 1 — признак присутствует (живот вздут) и 0.5 — характеристика переменной типа живот умеренно вздут, подвздут и т.п. Для каждого пациента выборки экспертом была поставлена оценка тяжести по шкале от 1 до 3.

Приведение значений показателей состояния пациента к равномерному распределению. В исходную выборку входят переменные, измеренные в разных шкалах, некоторые из них являются категориальными,

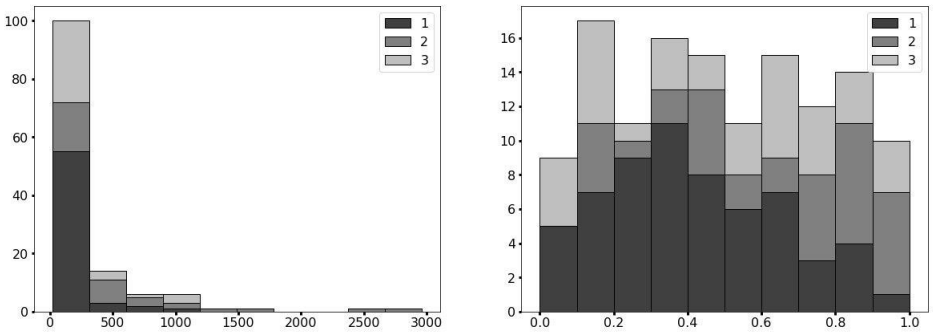


Рис. 1: Гистограмма признака Амилаза до преобразования (слева) и после (справа)

как следствие, размах значений некоторых также существенно разнится. Для приведения всех переменных к одинаковым единицам измерения служит процедура нормализации [2].

Помимо того, что многие переменные выражаются в разных единицах, некоторые из них имеют также явно выраженные асимметричные распределения (например, амилаза, см. рис. 1). Выполняя процедуру нормализации в случае распределения, представленного на рис. 1, получим существенное сгущение точек в одной (левой) части области значений признака и значительно меньшее количество в оставшейся части. Такая особенность частоты признака приводит к тому, что при исследовании на значимость, он может покинуть набор значимых признаков, являясь на самом деле значимым. Произойти это может именно потому, что на реальных выборках может чаще встречаться более легкая форма заболевания, где значения признаков меняются в небольших интервалах отклонения от нормы. Гораздо реже встречаются пациенты, у которых результат анализов настолько сильно отличается от нормы или от умеренного отклонения от нее, что выглядит на гистограмме как некий выброс или ошибочное значение. Снизить риск исключения значимого признака с такой особенностью распределения позволяет квантильное преобразование исходных значений (метод обратного преобразования Кокса-Бокса) [3].

Отбор значимых признаков и проведение классификации. В случае большого числа признаков технология решения задачи классификации начинается с отбора значимых признаков и с сужения пространства. В случае обработки медицинских данных стоит предполагать достаточно сильную, почти функциональную связь между признаками, поскольку все они связаны в единую биологическую систему, организм, где изменение одного элемента (органа) влечет за собой изменения в работе остальных для компенсации пагубных влияний. Поэтому предположение о мультиколлинеар-

ности признаков имеет место при обработке данных медицинских анализов и характерных признаков состояния больного.

Одним из эффективных методов понижения размерности и борьбы с мультиколлинеарностью признаков является гребневая регрессия [3, 4]. Гребневую регрессию используют, если имеет место: переизбыточность данных; коррелированность независимых переменных (мультиколлинеарность); сильные различия собственных значений характеристического уравнения или близость к нулю некоторых из них.

Дана выборка $V = \{x_i^j, y_i, j = \overline{1,N}, i = \overline{1,s}\}$, s — объем выборочных данных, N — количество признаков, x^j – значения j –ой независимой переменной (признака), y — значения зависимой переменной (принадлежность классу).

Используем линейную модель: $y = f(x,\beta)$, где f — линейный оператор (линейная функциональная зависимость), β — параметры модели.

Полагаем, что вектор коэффициентов линейной регрессионной модели β находится методом наименьших квадратов:

$$\sum_{i=1}^{n}(f(x,\beta)-y_i)^2 -> \min_{\beta_i}. \tag{1}$$

Аналитическое решение данной задачи: $\beta = (X^TX)^{-1}X^TY$, однако при вырожденности матрицы X^TX решение оказывается не единственным, а при ее плохой обусловленности Ц не устойчивым. Поэтому вводится регуляризация по параметру β, например, L_2. Тогда приходим к следующей задаче минимизации:

$$Q(\beta) = Y - X\beta^2 + \lambda\beta^2 -> \min_{\beta}, \tag{2}$$

где $\lambda > 0$ параметр регуляризации. Регуляризованное решение метода наименьших квадратов выглядит следующим образом:

$$\beta = (X^TX + \lambda I)^{-1}X^TY. \tag{3}$$

Увеличение параметра λ приводит к уменьшению нормы вектора параметров и повышению эффективности размерности признакового пространства.

В качестве критерия качества использовался показатель AUC (Area Under Curve) - площадь под кривой ROC. Значение AUC может меняться от 0 до 1, но, как правило, говорят об изменениях от 0.5 до 1 (кидеальнаянь модель).

Число разбиений составило 10000 сгенерированных по методу Монте-Карло наборов данных, сформированных таким образом, чтобы тестовая и обучающая части повторяли соотношения долей объектов разных классов.

Каждая история болезни содержала признак кметка классань в постановке двуальтернативной гипотезы кодин-против-всехнь, то есть для задачи 3 тяжесть против 1 и 2, если тяжесть была 3 ставилась метка 1, если тяжесть была 1 или 2, то ставилась метка 0. Аналогично формировался признак для задачи 1-я тяжесть против 2 и 3. Для каждой задач проводилась процедура понижения размерности. Далее в статье подробно рассматривается технология использования линейной регрессии с параметром регуляризации для понижения размерности в задаче 3 тяжесть против 1 и 2, с последующим приведением численных результатов по обеим задачам.

Настройка регуляризующего параметра осуществлялась расчетом по сетке с неравномерным шагом в диапазоне значений [0:1000]. С каждым значением регуляризующего параметра строилась гребневая регрессия. Если с увеличением значения параметра точность увеличивалась, следовательно, продолжало выбираться следующее большее значение. В варианте незначительного увеличения точности возвращалось предыдущее значение (см. табл. 1).

Таблица 2: Настройка регуляризующего параметра

Значения параметра регуляризации λ	AUC
0	0.86476
1	0.88878
5	0.90580
10	0.91203
100	0.92251
200	0.92318
1000	0.92370

Из табл. 1 видно, что для работы с гребневой регрессией целесообразно остановиться на значении λ равного 200, дальнейшее увеличение значительного повышения точности не дает. Далее с целью понижения размерности признакового пространства применялась гребневая регрессия в комбинации с методом последовательного сокращения (алгоритм Del)[3, 4]. Алгоритм Del представляет собой метод упрощения структуры регрессионной модели. Основная идея последовательного сокращения: элементы модели, которые оказывают малое влияние на ошибку, можно исключить из модели без значительного ухудшения качества. Следуя концепции алгоритма Del, из структуры гребневой регрессии удалялись по очереди признаки, коэффициенты которых были наименьшими; если при этом точность улучшалась или не ухудшалась, признак удалялся. В табл. 2 представлена очередность удаления признаков и значения AUC для варианта третья степень тяжести против первой и второй.

В результате значимыми, с позиции решения задачи классификации 3 степень тяжести против 1 и 2, стали признаки: температура x_1, головка x_2 и тело x_3 поджелудочной железы , контур поджелудочной железы x_4, гемоглобин x_5, глюкоза x_6, моноциты x_7, натрий x_8, скорость оседания эритроцитов x_9, перистальтика x_{10}, наличие жидкости в брюшной полости x_{11}, хвост поджелудочной железы x_{12}.

На оценке коэффициентов множественной линейной зависимости признаков был сформирован некоторый показатель, представляющий собой результат работы линейной модели, связывающий значимые параметры для третьего класса. Линейная зависимость описывается уравнением, коэффициенты которой оценивались по обучающей выборке:

$$y(x_1, ..., x_{12}) = 0.0252x_1 + 0.0094x_2 + 0.009x_3 + 0.0138x_4 - 0.0175x_5 +$$

$$+0.019x_6 + 0.093x_7 - 0.0113x_8 + 0.0337x_9 - 0.0355x_{10} + 0.0709x_{11} + 0.0061x_{12} + 0.2314. \quad (4)$$

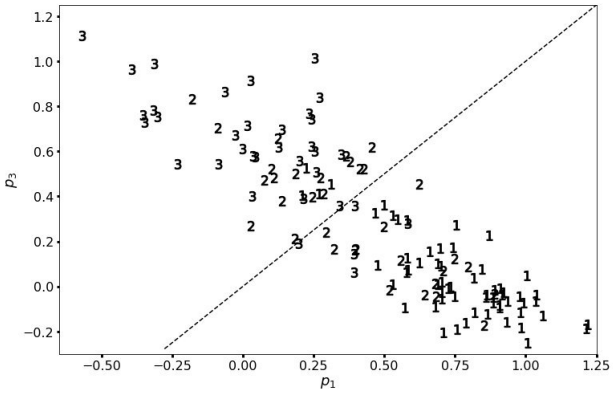
Та же процедура была проведена для варианта 1 степень тяжести против 2 и 3, в результате набор значимых признаков расширился и изменился: температура x_1, наличие дополнительных образований x_2, гемоглобин x_3, тромбоциты x_4, лейкоциты x_5, амилаза x_6, калий x_7, мочевина x_8, натрий x_9, перистальтика x_{10}, наличие жидкости в брюшной полости x_{11}, хвост поджелудочной железы x_{12}, АСТ x_{13}. Полученная линейная зависимость признаков имеет следующий вид:

$$y(x_1, ..., x_13) = -0.033x_1 - 0.029x_2 + 0.033x_3 - 0.019x_4 - 0.034x_5 - 0.043x_6 +$$

$$+0.035x_7 - 0.016x_8 + 0.026x_9 + 0.05x_{10} - 0.128x_{11} - 0.027x_{12} - 0.036x_{13} + 0.51. \quad (5)$$

На основании полученных линейных зависимостей, формировалась обучающая выборка для построения разделяющей поверхности между первым и третьим классом. Обучение классификатора проводилось методом скользящего экзамена с использованием линейной структуры. Результаты представлены на рис. 4, где цифрой "3"обозначены точки, относящиеся к третьей степени тяжести, которую поставил врач, "1" – точки первой степень тяжести, "2" – второй. Пунктиром нарисована разделяющая линия Ц классификатор.

Все точки, расположенные левее нее, алгоритм отнес к третьей степени тяжести тяжесть, все что правее Ц характеризует решение классификатора о принадлежности точки к первой степени.



4 - Результаты классификации.

Средний процент ошибки относительно неверно классифицированных точек первого и третьего классов составил 8.46.

Выводы. В работе рассматривается процедура обработки данных историй болезни для отбора значимых признаков и построения классификатора для определения легкой и тяжелой степени тяжести острого панкреатита по данным лабораторных и инструментальных видов исследования в первые сутки заболевания. Приведены численные исследования опробованных методов, процедур настройки коэффициентов линейной зависимости значимых признаков заболевания, обучения классификатора. Разработанный метод классификации с использованием гребневой регрессии показал свою эффективность для задачи отбора значимых признаков с показателем среднего числа правильных ответов свыше 92% относительно экспертной оценки (на примере оценки степени тяжести острого панкреатита). Опробованная процедура построения классификатора может быть использована

в качестве одного из элементов математического ядра системы поддержки принятия врачебных решений.

Предложенная модель классификатора позволяет быстро и эффективно определить тяжесть патологического процесса, что в дальнейшем позволит сделать выбор рациональной стартовой терапии, оценить необходимость оперативного вмешательства и в случаи тяжелой степени, назначить усиленную терапию. Что прогнозируемо снизит процент осложнений острого панкреатита, соответственно уменьшит частоту летальных исходов.

ЛИТЕРАТУРА

1. Литвин А. А., Жариков О. Г., Ковалев В. А.Система поддержки принятия решений в прогнозировании и диагностике инфицированного панкреонекроза. // Врач и информационные технологии. Ц 2012. - є2. стр. 54Ц63.

2. Вадзинский Р.Н. Справочник по вероятностным распределениям. - СПб.: Наука, 2001, 295 с.

3. Стрижов, В. В. Методы выбора регрессионных моделей / В.В. Стрижов, Е.А. Крымова - М.: Вычислительный центр РАН, 2010 Ц 60 с.

4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Книга 2. М.: Финансы и статистика, 1986. Ч 351 с.

Kononova Nadezhda (Krasnoyarsk, Russia), Mangalova Ekaterina (Krasnoyarsk, Russia), Chubarova Olesya (Krasnoyarsk, Russia)

APPLIED CLASSIFICATION PROBLEMS USING RIDGE REGRESSION

Abstract. The purpose of this study is to improve the treatment results for patients with acute pancreatitis by improving the severity objectification of acute pancreatitis. To achieve this purpose data mining methods were used, statistical classification in particular. The study was based on a retrospective analysis of 130 cases of acute pancreatitis. To solve the given classification problem ridge regression was used in combination with the sequential reduction algorithm. The result of the study was the classifier that allows to distinguish three degrees of acute pancreatitis severity, to better determine treatment tactics. During validation the method of classifying the severity of acute pancreatitis using ridge regression has proven to be effective.

Key words: *acute pancreatitis, severity, classifier, ridge regression, AUC (Area Under Curve*

Таблица 3: Очередность удаления признаков

Признак λ	Значение AUC
Дополнительные образования	0.9443
Лимфоциты	0.9465
Калий	0.9488
Эхогенность	0.9515
Вздутие живота	0.9537
Билирубин	0.9549
Амилаза	0.9554
Структура	0.9560
Мочевина	0.9563
Эозинофилы	0.9565
Тромбоциты	0.9567
АСТ	0.9570
Эритроциты	0.9571
Лейкоциты	0.9579

Корреляционно-регрессионный анализ показателей заболеваемости и смертности населения красноярского края по причине социально значимых заболеваний

В.С.Дансорунова[1], Н.А.Лукьянова[1,2]

[1] Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ,

[2] Институт математики и фундаментальной информатики Сибирский федеральный университет

dansorunova_victoria@mail.ru, nalukyanovakrsk@gmail.com

Актуальность. Социально значимые заболевания (СЗЗ) – это болезни, которые возникают и распространяются в социально-экономических условиях и приносят ущерб обществу и требуют социальной защиты населения. Основным признаком и проблемой этих заболеваний – это способность к массовому распространению среди населения. Согласно перечню (утв. постановлением Правительства РФ от 1 декабря 2004 г. №715 с изменениями и дополнениями от 12 июля 2012 г.) [1] к СЗЗ относятся: туберкулез, инфекции, передающиеся преимущественно половым путем, гепатит В, гепатит С, болезнь, вызванная вирусом иммунодефицита человека (ВИЧ), злокачественные новообразования, сахарный диабет, психические расстройства и расстройства поведения, болезни, характеризующиеся повышенным кровяным давлением.

В работе проведено статистическое исследование взаимосвязи показателей смертности и общей заболеваемости населения Красноярского края в связи с заболеваниями системы кровообращения, злокачественными новообразованиями (ЗНО), вирусными гепатитами (ВГ) и туберкулезом.

Туберкулез это широко распространенное во всем мире инфекционное заболевание человека, а также других представителей живых организмов, вызываемое микобактериями, так называемыми палочками Коха. Туберкулез чаще всего поражает дыхательную систему, а именно легкие, но также он может поражать другие органы (например, печень или селезенку) или системы органов (например, костную систему). Передается это заболевание воздушно-капельным путем. Обычно после инфицирования микобактериями заболевание протекает в бессимптомной, скрытой форме. Однако, примерно в одном из десяти случаев заболевание может перейти в активную форму.

Злокачественные новообразования являются важной медико-социальной проблемой. Социальная значимость заболеваемости злокачественными но-

вообразованиями связана с различными этиологическими факторами, и сложностью профилактики.

Вирусный гепатит (ВГ) имеет широкое распространение, характеризуются неблагоприятной тенденцией к росту заболеваемости, нередкими случаями хронизации. Данная патология и ее серьезные осложнения в МКБ занимают лидирующее положение в структуре смертности, наряду, с сердечно-сосудистыми заболеваниями, онкологией, сахарным диабетом и хроническими заболеваниями легких. Высокие показатели заболеваемости и болезненности от данной патологии влекут за собой значительный ущерб на медицинском и на социально-экономическом уровнях. Красноярский край показателен в отношении гепатита. В рамках Сибирского федерального округа он занимает второе место по заболеваемости ВГ.

На протяжении многих лет уровень заболеваемости населения в Российской Федерации болезнями системы кровообращения (БСК) стабильно сохраняется одним из самых высоких среди других классов болезней. На сегодняшний момент в списке заболеваемости населения Красноярского края БСК занимают вторую строчку. Заболеваемость и смертность населения по причине БСК принадлежат к важным статистическим показателям, которые характеризуют санитарное положение в Красноярском крае. Показатели смертности населения от болезней системы кровообращения стойко сохраняют за собой первое место

Цель работы: оценка тесноты и существенности связи между показателями смертности и заболеваемости населения Красноярского края по причине болезней системы кровообращения, онкологических заболеваний, туберкулеза и вирусного гипатита.

Заболеваемость – медико-статистический показатель, определяемый отношением числа заболеваний с впервые установленным диагнозом к средней численности населения, проживающей на данной территории.

Болезненность (распространенность) – медико-статистический показатель, характеризующийся совокупностью заболеваний с определенным диагнозом в анализируемой популяции, зарегистрированный в данном календарном году, вне зависимости от первичного диагностирования и времени выявления.

Смертность – статистический показатель, рассчитывающийся, как отношение числа умерших от конкретной причины среди населения данной территории к среднегодовой численности населения на данной территории.

Материалы и методы. Информационную базу для данного исследования составили данные ежегодных санитарно-эпидемиологических отчетов

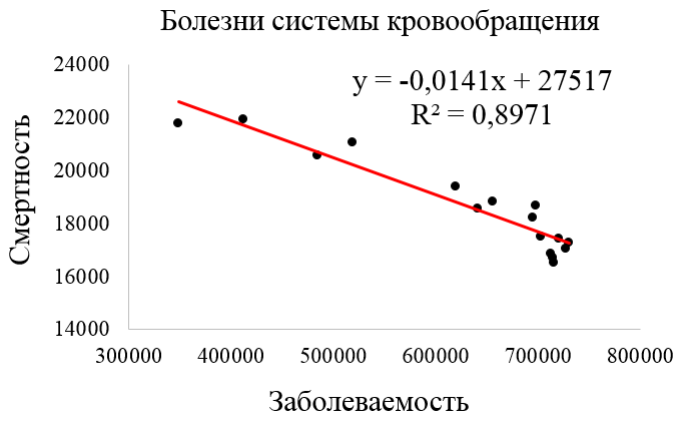
Красноярского краевого медицинского информационного аналитического центра [2], Федерального центра мониторинга противодействия распространению туберкулеза в РФ [3], Московского научно-исследовательского онкологического института имени П.А. Герцена – филиал «Национального медицинского исследовательского центра радиологии» Министерства здравоохранения РФ [4]. Методы исследования: корреляционный анализ, регрессионный анализ [5]. Обработка информации проводилась с использованием Microsoft Excel.

Результаты и обсуждение. На основе статистических данных по показателям заболеваемости и смертности от болезней системы кровообращения в Красноярском крае с 2002 по 2017 год [2] был проведен анализ динамики исследуемых показателей. Показатели заболеваемости БСК имеют растущую тенденцию, а показатели смертности - убывающую (рисунок 1). Результаты корреляционного анализа позволяют говорить, что между объясняемой переменной (смертность) и влияющего на него фактора (заболеваемость) имеется сильная отрицательная связь с коэффициентом корреляции $R = -0,947$, на уровне значимости 0,05 и со значением коэффициента детерминации $R^2 = 0,897$. Связь показателей хорошо описывает линейная регрессия (рисунок 2).

Результаты корреляционного анализа показали, что увеличение показателя заболеваемости в Красноярском крае влечет снижение показателя смертности, а это косвенно свидетельствует о своевременной ранней диагностике заболеваемости, и как следствие, снижении количества летальных исходов.

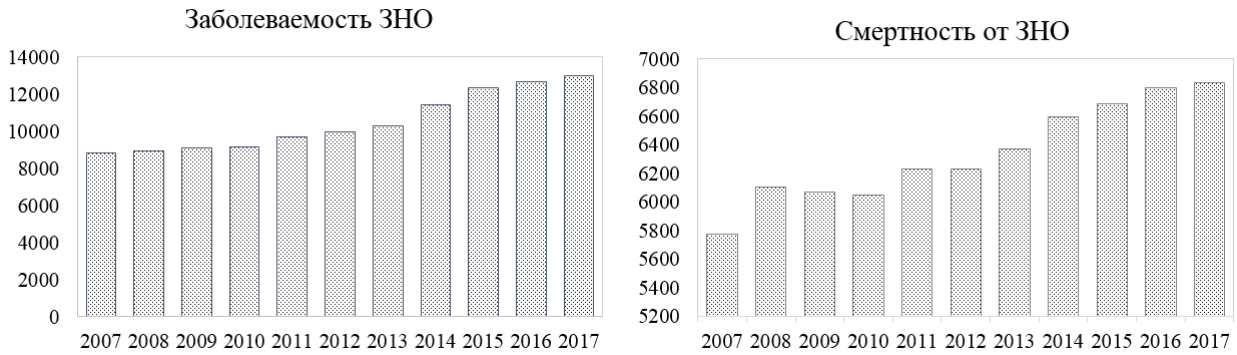


1-рисунок Динамика показателей показателей заболеваемости и смертности от БСК (абс. число) в Красноярском крае за 2002-2017 гг.

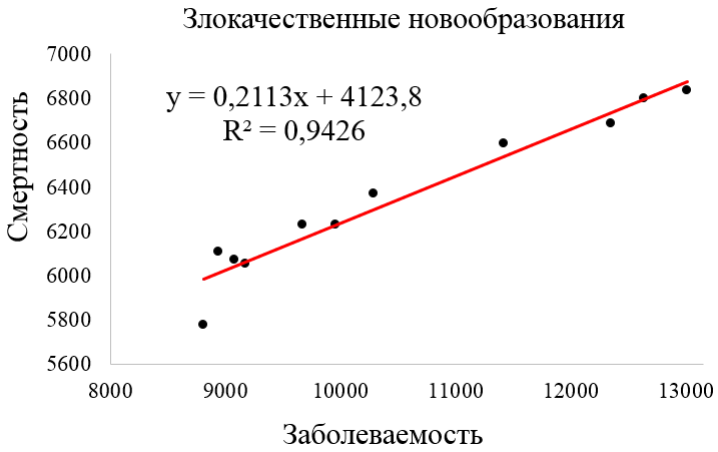


2-рисунок Диаграмма рассеяния показателей заболеваемости и смертности от БСК (абс. число) в Красноярском крае за 2002-2017 гг.

На основе данных ежегодных санитарно-эпидемиологических отчетов Московского научно-исследовательского онкологического института имени П.А. Герцена [5] за период с 2007 по 2017 год проведен корреляционный анализ показателей смертности и заболеваемости злокачественными заболеваниями в Красноярском крае. Показатели заболеваемости ЗНО и смертности от ЗНО за 2007-2017 гг. имеют растущую тенденцию (рисунок 3). Наблюдается сильная положительная связь с коэффициентом корреляции $R = 0,97$, на уровне значимости 0,05. Взаимосвязь показателей хорошо описывает линейная регрессия (рисунок 4). Коэффициент детерминации показывает, что 94% вариации объясняемой переменной (смертность) учтено в модели и обусловлено влиянием на нее фактора (заболеваемость), включенного в модель. Отсюда следует, что вслед за увеличением уровня заболеваемости повышается и показатель смертности, соответственно, можно судить о том, что онкология все еще остается на первых строчках списка заболеваний, уносящих жизни населения Красноярского края. Положительная тенденция зависимости смертности от заболеваемости может быть связана со слишком поздним обнаружением опухолевых образований, несовершенством технического и медикоментозного оснащения медицинского учреждения, сложности самого лечения, требующего постоянного наблюдения и точной диагностики.



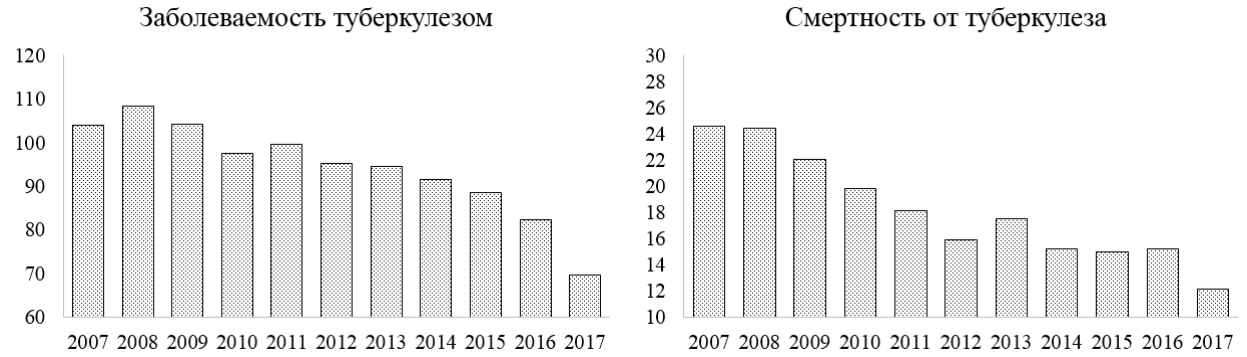
3-рисунок Динамика показателей заболеваемости и смертности от ЗНО (абс. число) в Красноярском крае за 2007-2017 гг.



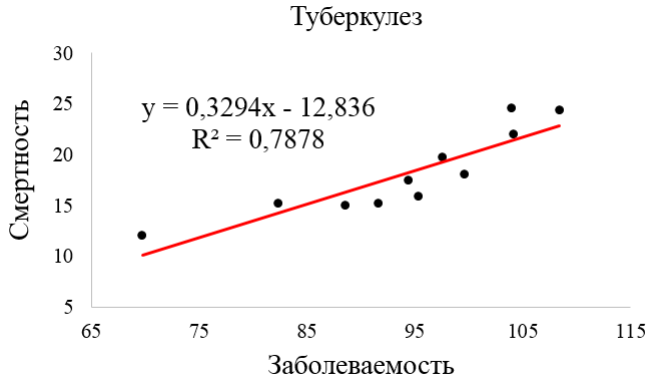
4-рисунок Диаграмма рассеяния показателей заболеваемости и смертности от ЗНО (абс. число) в Красноярском крае за 2007-2017 гг.

На основе статистических данных [3, 4] за период с 2007 по 2017 год проведен анализ показателей смертности и заболеваемости туберкулезом в Красноярском крае. Динамика показателей заболеваемости и смертности от туберкулеза за 2007-2017 гг. имеют убывающую тенденцию (рисунок 5). Показатель заболеваемости туберкулезом в Красноярском крае на 2007г. составлял 103,9 на 100 тыс. населения, за 10 лет он снизился на 33% и в 2017г. составил уже 69,7 на 100 тыс. населения. Показатель смертности от туберкулеза снизился на 50,9%, в 2007г. он составлял 24,6 на 100 тыс. населения, а в 2017г. уже 12,1 на 100 тыс. населения, т.е. можно сказать с полной уверенностью, что показатель смертности за промежуток времени с 2007-2017гг. стал меньше в 2 раза. Наблюдается сильная положительная связь с коэффициентом корреляции $R = 0,89$, на уровне значимости 0,05. Взаимосвязь показателей описывает линейная регрессия (рисунок 6). Коэффициент детерминации показывает, что 79% вариации объясняемой

переменной (смертность) учтено в модели и обусловлено влиянием на нее фактора (заболеваемость), включенного в модель. С учетом динамики показателей в исследуемый период времени видно, что произошло снижение заболеваемости туберкулезом и как следствие снижение смертности. Снижение показателя общей заболеваемости туберкулезом происходит на фоне улучшения организации профилактических осмотров на туберкулез и снижения доли запущенных форм туберкулеза среди впервые выявленных пациентов.



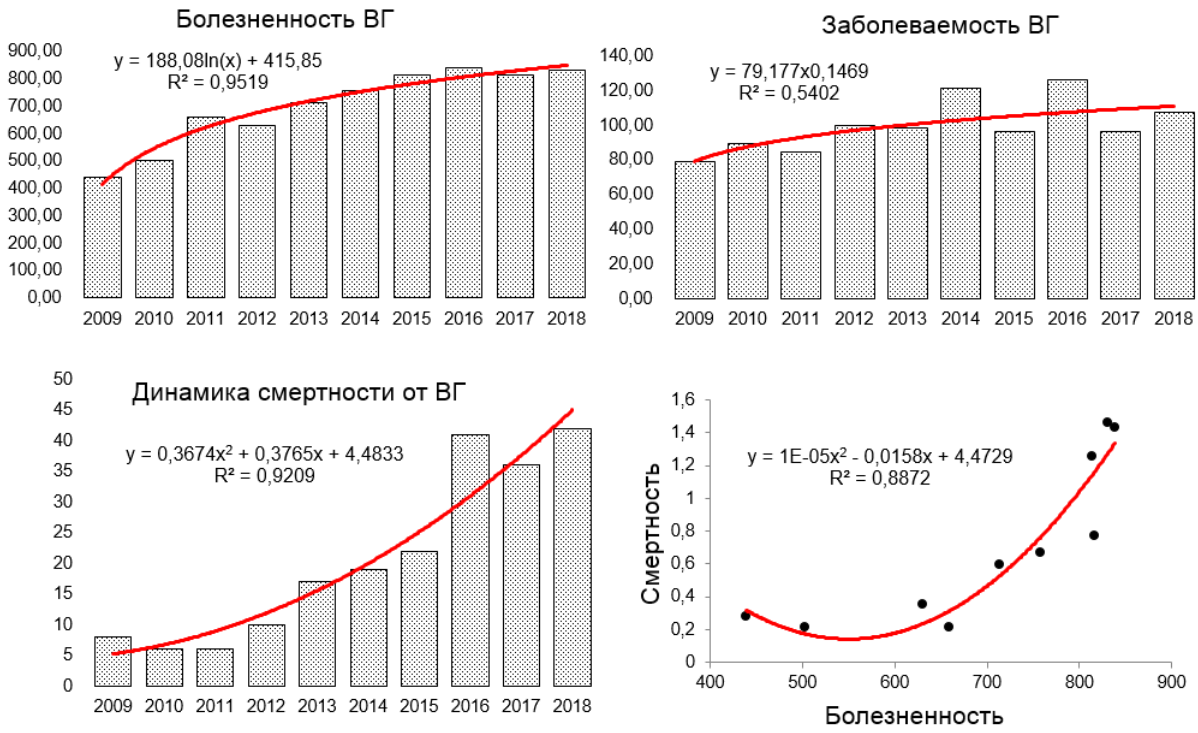
5-рисунок Динамика показателей заболеваемости и смертности от туберкулеза (на 100 тыс. населения) в Красноярском крае за 2007-2017 гг.



6-рисунок Диаграмма рассеяния показателей смерности и заболеваемости туберкулезом в Красноярском крае за 2007-2017 гг.

Проанализированы показатели заболеваемости, распространенности и смертности по причине вирусного гепатита по Красноярскому краю за период 2009-2018 гг. В многолетней динамике заболеваемости ВГ за период 2009-2018 гг. отмечается стабильно высокий уровень показателя заболеваемости. Показатель заболеваемости ВГ за 10 лет имеет плавную тенденцию к росту, аппроксимируется степенным законом $y = 79,177x^{0,147}$, ($R^2 = 0.54$). Пик заболеваемости в крае приходился на 2014 г. и 2016 г., где показатель заболеваемости, соответственно, был равен 121,35 и 125,94 на

100 тыс. чел. Показатели болезненности и смертности от ВГ в период с 2009-2018 гг. имеют также растущую тенденцию. Показатель болезненности аппроксимируется логарифмическим законом $y = 188,08\ln(x) + 415,85,\ (R^2 = 0.95)$, показатель смертности – квадратическим законом $y = 0,2324x^2 + 2,078x,\ (R^2 = 0.913)$. Показатель смертности в течение 10 изучаемых лет также растет. Наблюдается увеличение показателей более чем в 5 раз на начало и конец изучаемого периода. Проведенный корреляционно-регрессионный анализ для показателей болезненности и смертности по причине ВГ выявил сильную положительную связь между показателями: коэффициент детерминации $R^2 = 0,8872$. Связь показателей хорошо описывает квадратическая регрессия $y = 10^{-5}x^2 - 0,0158x + 4,4729$ (рисунок 7).



7-рисунок Динамика показателя болезненности ВГ (на 100 тыс. населения), показателя заболеваемости ВГ (на 100 тыс. населения), показателя смертности ВГ (умершие – абс.значения), диаграмма рассеяния показателей смерности и болезненности за 2009-2018 гг.

ЛИТЕРАТУРА

1. Баянова Т. А. Социально значимые заболевания. Эпидемиология и профилактика инфекционных заболеваний: туберкулез, ВИЧ-инфекция, вирусные гепатиты В, С : учеб. пособие / Т. А. Баянова, А. Д. Ботвинкин, Н. Ю. Куприянова. Иркутск: ИГМУ, 2015. 49 с.

2. Официальный сайт Красноярского краевого медицинского информационно-аналитического центра. [Электронный ресурс]. - Режим доступа: https://www.kmiac.ru/ (дата обращения: 15.09.2019).

3. Официальный сайт Федерального центра мониторинга противодействия распространению туберкулеза в РФ. [Электронный ресурс]. - Режим доступа: https://mednet.ru/informatizatsiya/tsentr-monitoringa-tuberkuleza (дата обращения: 16.09.2019).

4. Статистический сборник. ФГБУ "Национальный медицинский исследовательский центр радиологии"Министерства здравоохранения РФ. [Электронный ресурс]. - Режим доступа: https://nmicr.ru/nauka/nashi-izdaniya/statisticheskiy-sbornik/index.php (дата обращения: 16.09.2019).

5. Юнкеров В. И., Григорьев С. Г. Математико-статистическая обработка данных медицинских исследований. СПб.: ВМедА., 2002. 266 с.

Dansorunova Viktoria (Krasnoyarsk, Russia), Lukyanova Natalia (Krasnoyarsk, Russia)

Correlation regression analysis of morbidity and mortality of population of krasnoyarsk region due to socially significant diseases.

Abstract. *In the article, a statistical research of the relationship between mortality rates and the general morbidity of the population of the Krasnoyarsk Territory due to circulatory system diseases, oncological diseases, viral hepatitis, and tuberculosis is carried out.*

Key words: *applied statistics, correlation, regression analysis*

Нечеткая модель для классификации степени тяжести панкреатита

А.И.Ионкина[1], **Е.Е.Голденок**[1,2]

[1] Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ,

[2] Институт математики и фундаментальной информатики Сибирский федеральный университет

E-mail: zolotoponyuk@gmail.com, egoldenok@gmail.com

Острый панкреатит (ОП) — острое (хирургическое) заболевание, характеризуется развитием отчка поджелудочной железы. Число больных с острым панкреатитом ежегодно растет, что может быть связано с различными обстоятельствами и причинами. Возраст пиковой заболеваемости составляет от 45 до 55 лет у мужчин и от 55 до 65 лет у женщин. Уровень летальности остается неизменно высоким. Частота: 10-20 случаев на 100000 жителей в год [1, 2, 3]. Важным показателем смертности является степень тяжести течения ОП, объективизация которого имеет большое практическое значение в диагностике заболевания [3]. Таким образом, задача постановки диагноза с математической точки зрения может быть сведена к задаче классификации. Целью данной работы является разработка нечеткой модели классификации степени тяжести ОП для поддержки принятия врачебных решений.

Одно из главных преимуществ нечетких моделей по сравнению с остальными математическими моделями, применяемыми для решения задач медицинской диагностики, состоит в том, что для их применения можно довольствоваться значительно меньшими объемами информации о системе, то есть рассматривать сложную систему как черный ящик, для которого характерна недостаточность информации о происходящих внутри него физических явлениях; не известны законы, преобразующие входные переменные в выходные.

Обозначим $\mathfrak{X}$ — множество пациентов с панкреатитом. Каждый пациент $X \in \mathfrak{X}$ характеризуется вектором своих значений $(x_1, \ldots, x_n)$ — результатами клинических, лабораторных и инструментальных исследований. Для каждого пациента $X \in \mathfrak{X}$ необходимо поставить диагноз — соответствующую степень тяжести панкреатита: 1 — легкая, 2 — средняя или 3 — тяжелая. В задачах классификации предполагается, что существует целевая функция $y : \mathfrak{X} \to \{1, 2, 3\}$, значения которой $y(X_i)$ известны только на конечном подмножестве пациентов $\mathbf{X} = \{X_1, \ldots, X_l\}$, образующих обуча-

ющую выборку. Необходимо по выборке $\mathbf{X}$ восстановить зависимость $y(X)$, причѣм не только на объектах обучающей выборки, но и на всѣм множестве $\mathfrak{X}$ [4].

Для решения задачи классификации степени тяжести панкреатита использовать модель Мамдани [4]. В рамках модели Мамдани пациент X_i рассматривается как система типа MISO, в которой входами являются все всевозможные результаты обследования, а выход всего один – степень тяжести заболевания $Y_j,\ j = 1, 2, 3$, соответственно Y_1 – легкая, Y_2 – средняя, Y_3 – тяжелая [3].

Информационную базу исследования составили обезличенные данные о 96 пациентах с ОП различной степени тяжести панкреатита. На основе существующих медицинских шкал оценки тяжести панкреатита, таких как SAPS, Ranson и клинико-лабораторных критериев – признаков ССВО, было отобрано 25 количественных признаков. Эти признаки (для трех групп пациентов с разной степенью тяжести ОП) были визуализированы с помощью коробчатых диаграмм, в качестве наиболее информативных выбраны те признаки, для которых очевидна бóльшая разделимость по классам. Анализ показал, что в качестве входных переменных можно использовать следующие показатели клинических и биохимических анализов крови: гемоглобин, лейкоциты, глюкоза, мочевина; и данных аускультации: систолическое артериальное давление (САД). В работе определены лингвистические оценки переменных и необходимые для их формализации функции принадлежности. Первоначально база знаний была сформирована из 30 коньюктивных правил. Далее были исключены противоречивые и избыточные правила, в результате получена база из 19 правил лингвистически неполная, зато численно полная и непротиворечивая. Предложенная нечеткая модель реализована в пакете визуального моделирования Simulink с системой нечеткой логики Fuzzy Logic Toolbox прикладной программы MatLab. Выбор вида функций принадлежностей, их параметров и выбор используемых операторов производился методом проб и ошибок, корректировался на основе данных обучающей выборки. Построенная учебная нечеткая модель определения степени тяжести панкреатита, в принципе, может быть частью системы поддержки принятия решений врача на этапе диагностики степени тяжести острого панкреатита.

Выводы. В случае ненастраиваемых нечетких моделей и регуляторов влияние используемых операторов проявляется намного сильнее, поскольку выбор неподходящих операторов невозможно компенсировать ничем. С учетом этого, был использован метод проб и ошибок, исследованы харак-

теристики модели для различных комбинаций операторов и принято решение об использовании оператора PROD, потому что при его использовании классификация обучающей выборки давала наименьшее число ошибок. Сформулированная база правил является лингвистически неполной, что, в принципе, не является недостатком модели, численно полной и непротиворечивой. Недостатком модели является то, что она не поддается более тонкой настройке средствами пакета Fuzzy Logic Tol lBox. Для более тонкой настройки модели в дальнейшем предлагается использовать нейронечеткие модели.

ЛИТЕРАТУРА

1. Акимов, А. А. Острый панкреатит в молодом возрасте / А. А. Акимов, С. Н. Стяжкина, А. А. Валинуров, [и др.] // Здоровье и образование в XXI веке. – 2018. – Т. 20, №4. - С. 18–21.

2. Попов А. Е. Острый панкреатит. Современные принципы диагностики и лечения: учебно-методическое пособие. Красноярск: КрасГМУ, 2018. 46 с.

3. Подолужный В. И. Острый панкреатит: современные представления об этиологии, патогенезе, диагностике и лечении. Фундаментальная и клиническая медицина, 2017. Т. 2, № 4. с. 62–71.

4. Пегат А. Нечеткое моделирование и управление. [Электронный ресурс]. А. Пегат пер. с англ. А. Г. Подвесовского, Ю. В. Тюменцева – 3-е изд. (эл.). М.: БИНОМ. Лаборатория знаний, 2015. 798 с.

5. Штовба С. Д. Проектирование нечетких систем средствами MATLAB. Москва: Горячая линия – Телеком, 2007. 288 с.

6. Sivanandam S. N. Introduction to Fuzzy Logic using MATLAB. / S. N. Sivanandam, S. Sumathi, S. N. Deepa. – Heidelberg: Springer-Verlag, 2007. p. 430

Ionkina Anastasia (Krasnoyarsk, Russia), Goldenok Elena (Krasnoyarsk, Russia)

Fuzzy model for classification of severity of pancreatitis

Abstract. *In the article is presented the fuzzy mamdani model for classification of severity of pancreatitis. A fuzzy rule base is formulated. The formulated rule base is linguistically incomplete, numerically complete and*

consistent. The proposed fuzzy classification model is implemented in the Fuzzy Logic TollBox package.

Key words: *fuzzy, fuzzy mamdani model, classification, rule base, linguistic variable*

Компонентный анализ смертности населения для оценки динамики ожидаемой продолжительности жизни

А.А.Миронова, В.О.Кобаненко, К.М.Параскевопуло, Р.Б.Курбанисмаилов

Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ

2800817@mail.ru, kobanenko1999@bk.ru, darkslave9595@gmail.com, krasgmu05@mail.ru

Ожидаемая продолжительность жизни (ОПЖ) – показатель, который отражает сколько в среднем будут жить люди, родившиеся в этом году, если смертность при этом будет неизменной [1], он независим от особенностей возрастной структуры и удобен для анализа в динамике [2]. Этот показатель крайне важен для оценки демографического состояния в стране [3]. Компонентный анализ проводится с целью разложить разницу ОПЖ между началом и концом выбранного периода на отдельные составляющие, которые соответствуют вкладу изменения смертности различных возрастных групп или от отдельных причин смерти в динамику этого показателя.

Цель работы: определение причин смерти в половозрастных группах, которые вносят наибольший вклад в изменение ОПЖ на примере Красноярского края.

Материалы и методы. Для выполнения компонентного анализа влияния изменения смертности в разных возрастных группах и по различным причинам смерти на динамику ожидаемой продолжительности жизни была использована методика E.E. Arriaga [4] по формуле:

$$eff_x = \left[\frac{I_x^1}{100000} \cdot \left(\frac{L_x^2}{I_x^2} - \frac{L_x^1}{I_x^1}\right)\right] + \left[\frac{T_{x+1}^2}{100000} \cdot \left(\frac{I_x^1}{I_x^2} - \frac{I_{x+1}^1}{I_x^2}\right)\right],$$

где

eff_x – вклад изменения смертности в динамику ОПЖ;

I_x^1 – порядок вымирания гипотетического поколения новорожденных в начальном году;

L_x^2 – число человеко-лет, прожитых гипотетическим поколением в возрасте от 1 года и далее в конечном году;

I_x^2 – порядок вымирания гипотетического поколения новорожденных в конечном году;

L_x^1 – число человеко-лет, прожитых гипотетическим поколением от возраста x до следующего возрастного интервала в начальном году;

T^2_{x+1} – число человеко-лет, прожитых гипотетическим поколением в возрастном интервале $x+1$ конечного года;

I^1_{x+1} – порядок вымирания гипотетического поколения новорожденных в возрастном интервале $x+1$ начального года.

Выбор методики обусловлен тем, что при вычислении повозрастных компонент их сумма равна фактическому изменению ОПЖ в изучаемый период времени [5].

Вычисления выполнены с использованием баз данных зарегистрированных случаев смерти в Красноярском крае с 1999 по 2017 год и баз данных численности населения с 1999 по 2018 год.

Введем обозначения: h – возрастной интервал, Σ – сумма компонент.

Компонентный анализ динамики ОПЖ выполнен в трех периодах, которые были выбраны на основе изменений общего коэффициента смертности в Красноярском крае с 1999 по 2017 г. Первый период с 1999 по 2004 г. – период роста смертности (коэффициент общей смертности вырос на 12,1%), второй период с 2004 по 2007 г. – период резкого снижения смертности (коэффициент общей смертности снизился на 14,6%) и третий период с 2007 по 2017 г. – период медленного снижения смертности (коэффициент общей смертности снизился на 8,9%).

Результаты и обсуждение. Компонентный анализ показал, что в период с 1999 по 2017 г. наблюдается положительный вклад в динамику ОПЖ из-за изменения смертности во всех возрастных группах, кроме интервала 35-39 лет среди женщин (таблица 3). Наибольший вклад в ОПЖ населения вносит снижение младенческой смертности (таблица 1).

Результаты компонентного анализа по причинам смерти всего населения Красноярского края представлены в таблице 1, мужского населения Красноярского края – в таблице 2, женского населения – в таблице 3.

В период с 1999 по 2017 г. самый большой отрицательный вклад в ОПЖ вносит изменение смертности от болезней нервной системы (VI класс по МКБ–10). Изменение смертности от данной причины снижает ОПЖ всего населения на 0,2709 года (у мужчин на 0,2786 года, а у женщин на 0,2199 года). Изменение смертности от данной причины в первых двух периодах вносит положительный вклад, а в третьем периоде наибольший отрицательный вклад в динамику ОПЖ всего населения – 0,3706 года.

Наибольший положительный вклад в динамику ОПЖ можно наблюдать за счет снижения смертности от болезней системы кровообращения (IX класс по МКБ–10). Изменение смертности от данной причины повышает ОПЖ всего населения на 3,9031 лет (у мужчин – на 3,1570 лет, а у женщин

– на 4,4021 лет). Рост ОПЖ за счет изменения смертности от данной причины происходил во всех трех периодах, но наибольший положительный вклад наблюдался в третьем периоде – 3,1445 года.

Наибольший рост ОПЖ мужчин Красноярского края можно наблюдать за счет изменения смертности от болезней системы кровообращения (IX), он составляет 2,6868 лет. Изменение смертности от данной причины вносит отрицательный вклад в динамику ОПЖ в первом периоде, а во втором и третьем – положительный.

Наибольший отрицательный вклад в динамику ОПЖ мужской части населения составляет 0,3996 года, который наблюдается за счет изменения смертности от болезней нервной системы (VI). Данная причина вносит отрицательный вклад только в последний период (медленного снижения смертности), что говорит о том, что ситуация со смертностью мужчин от болезней нервной системы стала хуже, по сравнению с тем, что было в предыдущих периодах.

Наибольший положительный вклад в динамику ОПЖ женской части населения Красноярского края наблюдается за счет изменения смертности от болезней системы кровообращения (IX), он составляет 3,3646 лет. Рост ОПЖ женщин за счет изменения смертности от данной причины происходил во всех трех периодах.

Наибольший отрицательный вклад в динамику ОПЖ женщин Красноярского края можно наблюдать за счет изменения смертности от внешних причин заболеваемости и смертности (XX), который составляет 0,4675 года в последнем периоде. Вклад изменения смертности от данной причины изменяется в каждом из периодов. В первом периоде вклад в динамику ОПЖ женской части населения отрицательный, во втором периоде положительный, а в третьем снова отрицательный.

Выводы. Компонентный анализ смертности населения в половозрастных группах, на примере Красноярского края, позволяет выделить причины смерти, которые вносят наибольший положительный или отрицательный вклад в динамику ОПЖ населения и целенаправленно планировать управляющие решения в системе здравоохранения.

ЛИТЕРАТУРА

1. Леманова П. В. Социальная политика в управлении развитием человеческого капитала: учеб. пособие. М.: Издательский дом Академии Естествознания, 2016. 228 с.

2. Кулак А. Г. Теоретические основы статистической оценки факторов, вызывающих потери продолжительности жизни. Вестник кафедры статистики Российского экономического университета имени Г.В. Плеханова. 2018. с. 145–148.

3. Колпаков С. Л., Яковлев А. А. О методологии оценки эпидемиологической ситуации. Эпидемиология и инфекционные болезни, 2015. № 20 (4). с. 34–39.

4. Наркевич А. Н., Миронова А. А. Вклад случаев смерти от различных причин в ожидаемую продолжительность жизни населения Красноярского края и его динамика. Медицинский альянс. 2019. №1. с. 71–81.

5. Arriaga E. E. Measuring and explaining the change in life expectancies. Demography. 1984. Vol. 21, №1. p. 83–96.

6. Миронова А. А., Наркевич А. Н., Виноградов К. А., Курбанисмаилов Р. Б. Декомпозиция динамики ожидаемой продолжительности жизни населения Красноярского края. Социальные аспекты здоровья населения [сетевое издание] 2019. №65(2). URL: http://vestnik.mednet.ru/content/view/1061/30/ DOI: 10.21045/2071-5021-2019-65-2-7

Mironova Alena (Krasnoyarsk, Russia), Kobanenko Vladislav (Krasnoyarsk, Russia), Paraskevopoulo Konstantin (Krasnoyarsk, Russia), Kurbanismailov Renat (Krasnoyarsk, Russia)

Component analysis of mortality to assess the dynamics of life expectancy

Abstract. *The article presents the contribution of changes in mortality in different age groups and from various causes in the dynamics of life expectancy.*

Key words: *applied statistics, component analysis, life expectancy*

Таблица 4: Повозрастные компоненты смертности и компоненты причин смерти всего населения Красноярского края 1999-2017 гг.

h	1999-2005	2005-2007	2007-2017	1999-2017	Класс МКБ	1999-2005	2005-2007	2007-2017	1999-2017
0	0,6477	0,0916	0,4540	1,2721	I	0,0194	0,0433	-0,0605	0,0988
1-4	0,1009	0,0263	0,0803	0,2195	II	0,1110	-0,0038	0,0110	0,2011
5-9	0,0592	0,0142	0,0161	0,0973	III	0,0003	-0,0051	0,0001	-0,0021
10-14	0,0163	-0,0292	0,0325	0,0199	IV	-0,0416	0,0652	0,0170	0,0098
15-19	0,0497	0,0778	0,0677	0,2060	V	0,0075	0,0411	0,0593	0,1379
20-24	-0,0015	0,1521	0,2588	0,4134	VI	0,0226	0,0331	-0,3706	-0,2709
25-29	-0,1963	0,1812	0,4062	0,3615	VIII	0,0030	-0,0031	0,0012	0,0029
30-34	-0,2101	0,2522	0,2705	0,2884	IX	0,0759	0,8923	3,1445	3,9031
35-39	-0,1566	0,3189	0,0767	0,2350	X	-0,0047	0,1397	0,5957	0,8347
40-44	-0,2070	0,3966	0,2167	0,3994	XI	-0,3625	0,2204	0,1240	-0,0414
45-49	-0,1366	0,4061	0,3486	0,6241	XII	-0,0049	-0,0122	-0,0077	-0,0152
50-54	-0,1338	0,4227	0,3959	0,6905	XIII	0,0028	0,0063	-0,0070	0,0024
55-59	-0,0571	0,3108	0,3383	0,5991	XIV	0,0108	0,0048	-0,0276	0,0120
60-64	-0,0494	0,3701	0,2060	0,5528	XV	0,0088	-0,0089	0,0109	0,0167
65-69	0,0577	0,1732	0,2445	0,4841	XVI	0,2881	0,0720	0,1200	0,5038
70-74	-0,0038	0,1771	0,2911	0,4455	XVII	0,1056	-0,0217	0,1640	0,2932
75-79	-0,0292	0,0882	0,2448	0,2656	XVIII	0,1564	0,4456	0,1404	0,9694
80-84	0,0417	0,0345	0,1356	0,1919	XX	-0,5703	0,4812	0,2601	0,8238
85-89	0,0246	0,0013	0,0563	0,0702	XXI	–	–	–	–
90+	0,0119	0,0112	0,0341	0,0437	XXII	–	–	–	–
Σ	-0,1717	3,4768	4,1748	7,4800	Σ	-0,1717	2,3902	4,1748	7,4800

Таблица 5: Повозрастные компоненты смертности и компоненты причин смерти мужского населения Красноярского края 1999-2017 гг.

h	1999-2005	2005-2007	2007-2017	1999-2017	Класс МКБ	1999-2005	2005-2007	2007-2017	1999-2017
0	0,6183	0,0699	0,0433	1,3135	I	0,0425	0,1707	0,0024	0,2403
1-4	0,1094	-0,0088	-0,0038	0,2039	II	0,1002	0,0675	0,0630	0,2684
5-9	0,0450	0,0180	-0,0051	0,0933	III	0,0078	0,0000	-0,0064	0,0028
10-14	0,0155	-0,0316	0,0652	0,0315	IV	-0,0236	0,0174	0,0027	-0,0057
15-19	0,0834	0,0978	0,0411	0,2505	V	0,0079	0,0794	0,0780	0,1709
20-24	-0,0138	0,2301	0,0331	0,5725	VI	0,0397	0,0247	-0,3996	-0,2786
25-29	-0,2931	0,2814	–	0,4923	VIII	0,0056	-0,0003	0,0010	0,0071
30-34	-0,3138	0,3930	-0,0031	0,4010	IX	-0,1741	0,9329	2,6868	3,1570
35-39	-0,1662	0,4010	0,8923	0,3874	X	-0,0530	0,3274	0,6880	0,9182
40-44	-0,2368	0,5088	0,1397	0,5249	XI	-0,3022	0,2306	0,1437	0,0266
45-49	-0,1619	0,4930	0,2204	0,7924	XII	-0,0093	0,0045	-0,0062	-0,0123
50-54	-0,1334	0,4780	-0,0122	0,8052	XIII	0,0022	0,0048	-0,0063	0,0016
55-59	-0,0537	0,3414	0,0063	0,6384	XIV	0,0186	0,0207	-0,0225	0,0319
60-64	-0,0574	0,3558	0,0048	0,5149	XV	–	–	–	–
65-69	-0,0176	0,1441	-0,0089	0,3503	XVI	0,2870	0,0476	0,1751	0,5538
70-74	-0,0117	0,1110	0,0720	0,2917	XVII	0,0518	0,0366	0,1477	0,2456
75-79	-0,0180	0,0353	-0,0217	0,1339	XVIII	0,2403	0,6695	0,1118	1,1482
80-84	0,0230	-0,0001	0,4456	0,0672	XX	-0,8200	1,2873	0,9203	1,4465
85-89	0,0036	-0,0012	–	0,0244	XXI	–	–	–	–
90+	0,0004	0,0043	0,4812	0,0331	XXII	–	–	–	–
Σ	-0,5786	3,9214	2,3902	7,9223	Σ	-0,5786	3,9214	4,5795	7,9223

Таблица 6: Повозрастные компоненты смертности и компоненты причин смерти женского населения Красноярского края 1999-2017 гг.

h	1999-2005	2005-2007	2007-2017	1999-2017	Класс МКБ	1999-2005	2005-2007	2007-2017	1999-2017
0	0,6766	0,1164	0,3470	1,1965	I	-0,0305	0,0433	-0,1374	0,2403
1-4	0,0878	0,0703	0,0680	0,2344	II	0,0998	-0,0038	-0,0447	0,2684
5-9	0,0768	0,0090	0,0080	0,1003	III	-0,0108	-0,0051	0,0075	0,0028
10-14	0,0171	-0,0253	0,0126	0,0049	IV	-0,0635	0,0652	0,0362	-0,0057
15-19	0,0050	0,0501	0,0914	0,1475	V	0,0024	0,0411	0,0272	0,1709
20-24	0,0128	0,0484	0,1375	0,1997	VI	-0,0019	0,0331	-0,2876	-0,2786
25-29	-0,0789	0,0401	0,2117	0,1626	VIII	-0,0003	-0,0031	0,0014	0,0071
30-34	-0,0687	0,0329	0,1572	0,1125	IX	0,3930	0,8923	3,3646	3,1570
35-39	-0,1362	0,1686	-0,0302	-0,0050	X	0,0800	0,1397	0,4241	0,9182
40-44	-0,1617	0,1959	0,1429	0,1661	XI	-0,4235	0,2204	0,1006	0,0266
45-49	-0,1133	0,2267	0,1974	0,3091	XII	0,0037	-0,0122	-0,0084	-0,0123
50-54	-0,1432	0,2623	0,3016	0,4163	XIII	0,0034	0,0063	-0,0076	0,0016
55-59	-0,0855	0,2053	0,3041	0,4227	XIV	0,0001	0,0048	-0,0270	0,0319
60-64	-0,0902	0,3210	0,2090	0,4546	XV	0,0210	-0,0089	0,0244	–
65-69	0,1527	0,1629	0,2196	0,5708	XVI	0,2854	0,0720	0,0497	0,5538
70-74	0,0457	0,2361	0,3424	0,6294	XVII	0,1744	-0,0217	0,1817	0,2456
75-79	0,0008	0,1655	0,3267	0,4677	XVIII	0,0145	0,4456	0,1671	1,1482
80-84	0,0612	0,0823	0,2444	0,3570	XX	-0,2121	0,4812	-0,4675	1,4465
85-89	0,0487	0,0047	0,0915	0,1261	XXI	–	–	–	–
90+	0,0276	0,0172	0,0216	0,0564	XXII	–	–	–	–
Σ	0,3351	2,3902	3,4044	6,1297	Σ	0,3351	2,3902	3,4044	7,9223

Вычисление отношения шансов для многомерных статистических данных

О.Д.Мельникова[1], Н.А.Лукьянова[2,3]

[1] Сибирский государственный университет науки и технологий имени академика М.Ф.Решетнева,

[2] Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ,

[3] Институт математики и фундаментальной информатики Сибирский федеральный университет

melok325@gmail.com, nalukyanovakrsk@gmail.com

Современная медицина должна быть персонализированной (оценка персональных данных человека) и превентивной (профилактическая, предупредительная), предиктивной (оценка различных факторов риска развития заболевания человека) и партисипативной (активное участие пациента в сохранении его здоровья) – 4-П медицины. Фактор риска - это не причинно-следсвенная связь заболевания, а это индикатор по которому можно оценивать, прогнозировать и предвидеть, классифицировать и тд. развитие и течение заболевания. В настоящее время актуальным является прогнозирование неблагоприятного течения и исхода беременности и родов для своевременного лечения и направления женщин в медицинские организации, где им будет оказана качественная медицинская помощь.

Данная статья является продолжением исследования связи между исходом и фактором риска клинического течения беременности и родов [1]. Информационную базу исследования составили результаты показателей перинатального мониторинга Красноярского края. Методы исследования в работе: аналитические, статистические. В исследование были включены случаи течения беременности с известным исходом 119938, в том числе с неблагоприятным исходом беременности и родов - «Преждевременные роды на 28-36 неделе» и с благоприятным исходом беременности и родов – «Срочные роды 37-40 недель» (таблица 1).

Исследовано 65 факторов риска (ФР), в том числе социальные, анамнестические, клинико-лабораторные, показатели здоровья, наличие заболеваний и др. Для ретроспективного сравнения влияния определенного фактора риска на две группы лиц в исследовании типа случай-контроль используется показатель отношения шансов [2].

Таблица 7: Исходы беременности (срочные, преждевременные роды) абс., %

Исходы родов	Число, чел.	Доля, %
«Срочные роды 37-40 недель»	114769	95,7
«Преждевременные роды на 28-36 неделе»	5169	4,3
Всего	119938	100

Целью работы является разработка алгоритма и программы вычисления показателей отношения шансов для многомерных статистических данных.

Следует отметить, что есть всевозможные онлайн калькуляторы расчета данного показателя, но только для одного фактора риска. Отсутствие бесплатного, универсального и общедоступного программного обеспечения для проведения вычисления многомерной оценки отношения шансов делает задачу разработки программы актуальной. Среда программной разработки: Mathcad 15.

Формула для расчета отношения шансов для одного фактора риска:

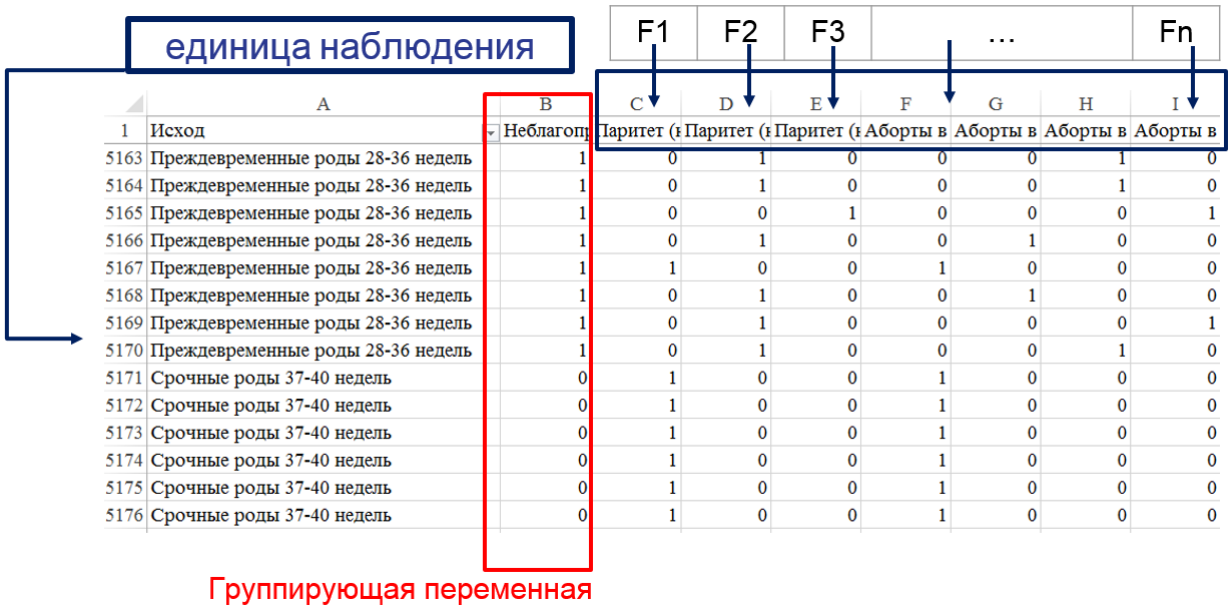
$$OR = \frac{\text{Шанс найти фактор риска в основной группе}}{\text{Шанс найти фактор риска в контрольной группе}} = \frac{A/C}{B/D}$$

где A, B, C, D представлены в таблице 2.

Таблица 8: Таблица сопряженности для расчета отношения шансов для одного фактора

	Исход есть (1) (основная группа)	Исхода нет (0) (контрольная группа)	Всего
Фактор риска есть (1)	A	B	A+B
Фактор риска отсутствует (0)	C	D	C+D
Всего	A+C	B+D	A+B+C+D

Разработка проводилась поэтапно. На этапе предобработки формируется матрица исходных данных $S_{m\times(n+1)}$, где m – количество наблюдаемых единиц (в нашем случае, 119938), n – количество исследуемых факторов риска F_n (n– величина, изменяющаяся по запросу пользователя, в нашем случае, $n = 65$). Первый столбец является группирующей бинарной переменной по неблагоприятным исходам беременности и родов - «Преждевременные роды на 28-36 неделе» – 1 и с благоприятным исходом беременности и родов – «Срочные роды 37-40 недель» – 0 (рисунок 1).



	A	B	C	D	E	F	G	H	I
1	Исход	Неблагоп	Паритет (1	Паритет (2	Паритет (3	Аборты в	Аборты в	Аборты в	Аборты в
5163	Преждевременные роды 28-36 недель	1	0	1	0	0	0	1	0
5164	Преждевременные роды 28-36 недель	1	0	1	0	0	0	1	0
5165	Преждевременные роды 28-36 недель	1	0	0	1	0	0	0	1
5166	Преждевременные роды 28-36 недель	1	0	1	0	0	1	0	0
5167	Преждевременные роды 28-36 недель	1	1	0	0	1	0	0	0
5168	Преждевременные роды 28-36 недель	1	0	1	0	0	1	0	0
5169	Преждевременные роды 28-36 недель	1	0	1	0	0	0	0	1
5170	Преждевременные роды 28-36 недель	1	0	1	0	0	0	1	0
5171	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0
5172	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0
5173	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0
5174	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0
5175	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0
5176	Срочные роды 37-40 недель	0	1	0	0	1	0	0	0

1-рисунок Входные данные

Далее, в вспомогательном программной модуле производится сортировка данных по группирующей переменной и подсчет количества возможных исходов по каждому фактору риска. Результатом вспомогательного модуля является матрица $M_{4\times n}$, в которой 4 фиксированные строки:

первая: A– количество исходов при наличии ФР,

вторая: B – количество отсутствия исхода при наличии ФР,

третья: C – количество исхода при отсутствии ФР,

четвертая: D – количество отсутствия исхода при отсутствии ФР.

На втором этапе разработан алгоритм вычисления многомерной оценки отношения шансов (алгоритме 1), который программно реализован в Mathcad.

Алгоритм 1. Алгоритм вычисления многомерной оценки отношения шансов

Вход: Матрица $M_{4\times n}$ (n количество факторов риска), α – уровень значимости

Выход: Матрица $v_{6\times n}$

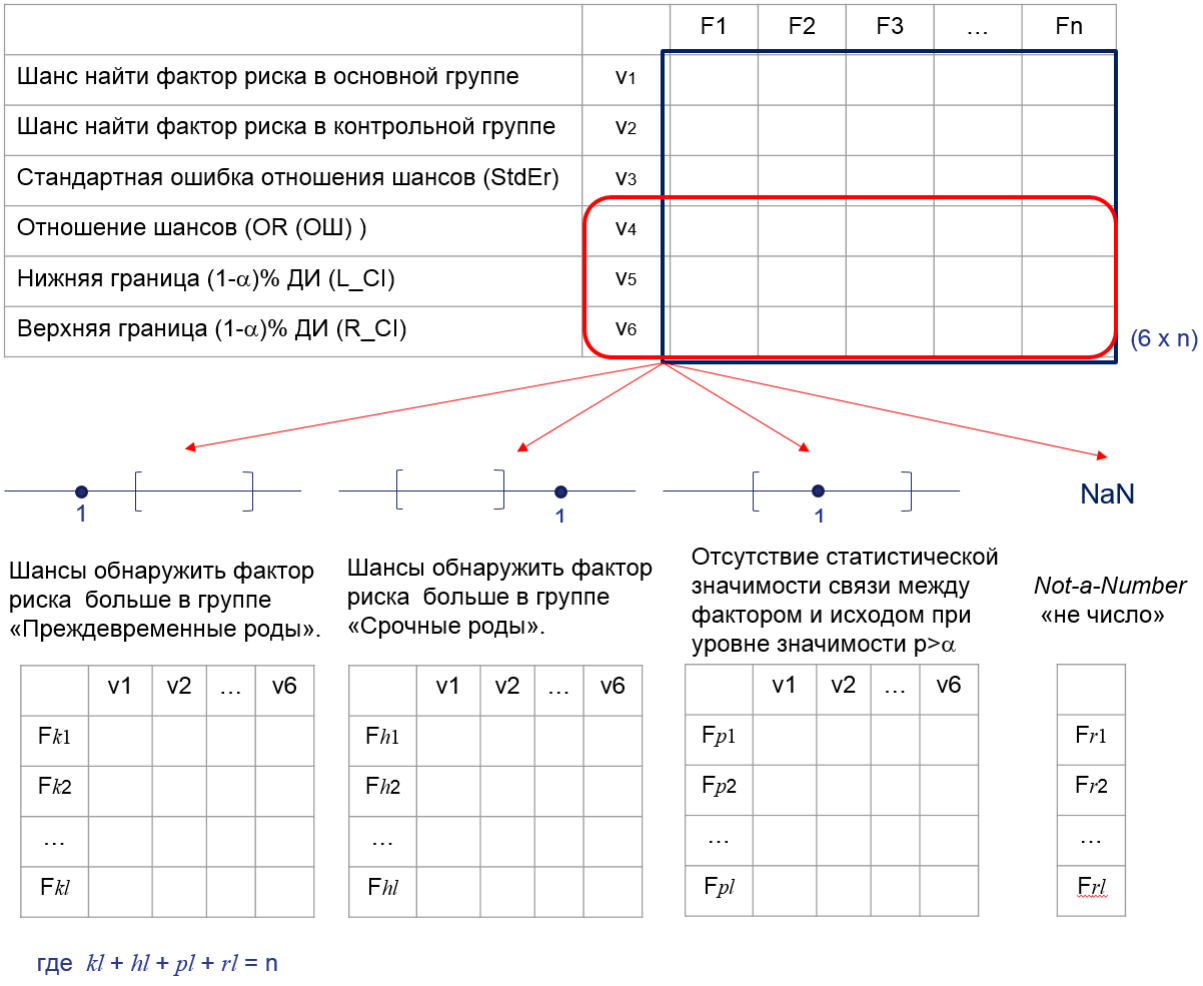
1: t=qnorm(1-α/2; 0; 1) ▷ квантиль нормального распределения
2: Для j=0 .. cols(M)-1 **выполнить**
3: $A = M_{0,j}$ ▷ количество исходов при наличии ФР
4: $B = M_{1,j}$ ▷ количество отсутствия исхода при наличии ФР
5: $C = M_{2,j}$ ▷ количество исхода при отсутствии ФР
6: $D = M_{3,j}$ ▷ количество отсутствия исхода при отсутствии ФР
7: Если $A \cdot B \cdot C \cdot D \neq 0$ **тогда**
8: $v_{0,j} = \dfrac{A}{C}$ ▷ Шанс найти фактор риска в основной группе (OFR(1))
9: $v_{1,j} = \dfrac{B}{D}$ ▷ Шанс найти фактор риска в контрольной группе (OFR(0))
10: $v_{2,j} = \sqrt{\dfrac{1}{A} + \dfrac{1}{B} + \dfrac{1}{C} + \dfrac{1}{D}}$ ▷ Стандартная ошибка отношения шансов
11: $v_{3,j} = \dfrac{v_{0,j}}{v_{1,j}}$ ▷ Отношение шансов (OR)
12: $v_{4,j} = e^{\ln v_{3,j} - t\cdot v_{2,j}}$ ▷ Нижняя граница ($1-\alpha$)% ДИ (L_CI)
13: $v_{4,j} = e^{\ln v_{3,j} + t\cdot v_{2,j}}$ ▷ Верхняя граница ($1-\alpha$)% ДИ (R_CI)
14: иначе
15: **Если** $C \neq 0$ **тогда** $v_{0,j} = \dfrac{A}{C}$
16: **иначе** $v_{0,j}$ =«NaN» **Конец условия**
17: **Если** $D \neq 0$ **тогда** $v_{1,j} = \dfrac{B}{D}$
18: **иначе** $v_{1,j}$ =«NaN» **Конец условия**
19: **иначе** $v_{2,j}$ =«NaN»
20: **Если** $C \cdot B \neq 0$ **тогда** $v_{3,j} = 0$
21: **иначе** $v_{3,j}$ =«NaN» **Конец условия**
22: **иначе** $v_{4,j}$ =«NaN»
23: **иначе** $v_{5,j}$ =«NaN»
24: Конец цикла
25: Вывод v ▷ Вывод матрицы результатов по каждому фактору риска
26: Конец работы.

В результате работы программного модуля по алгоритму 1 получаем по всем исследуемым факторам риска матричный отчет (рисунок 2):
– шанс найти фактор риска в основной группе (OFR(1)),

– шанс найти фактор риска в контрольной группе (OFR(0)),
– стандартная ошибка отношения шансов (StdEr),
– отношение шансов (OR),
– нижняя граница доверительного интервала ($1-\alpha$)% ДИ (CI_L),
– верхняя граница ($1-\alpha$)% ДИ (CI_R).

На третьем этапе проводится тестирование алгоритмов и программы, с последующим проведением численной обработки исследуемого материала (рисунок 3).

Импорт исходных данных и экспорт результатов в виде файлов txt- и xlsx-формата.



2-рисунок Вывод результатов по каждому фактору

По исследуемому статистическому материалу с помощью разработанной программы установлено 40 факторов повышающих риск неблагоприятного исхода преждевременные роды, для которых выявленная связь с исходом статистически значима при $p<0{,}05$. Шансы обнаружить фактор риска

больше в группе «Преждевременные роды». Фактор имеет прямую связь с вероятностью наступления исхода.

Для 18 факторов ДИ включает 1, поэтому делаем вывод об отсутствии статистической значимости связи между фактором и исходом при уровне значимости p>0,05.

Для 3 факторов OR<1, значит, шансы обнаружить фактор риска больше в контрольной группе «Срочные роды». Фактор имеет обратную связь с вероятностью наступления исхода.

По 4 факторам рассчитать OR не было возможности, потому что в эмпирических данных показатель наличия фактора риска в основной группе «Преждевременные роды» – A=0.

Из данных факторов риска можно выделить те факторы, которые при своем влиянии увеличивают риск развития преждевременные роды:
1) Преждевременные роды: двое и более ($OR = 7,63$, ДИ 95% [5,94; 9,80], $\chi^2 = 351,55$)
2) Положительная реакция на волчаночный антикоагулянт: да ($OR = 7,40$, ДИ 95% [2,02; 27,59], $\chi^2 = 12,6$)
3) Эндокринопатии: сахарный диабет ($OR = 6,62$, ДИ 95% [5,01; 8,75], $\chi^2 = 232,89$)
4) Сердечно-сосудистые заболевания матери: хроническая артериальная гипертензия 3-й стадии ($OR = 5,56$, ДИ 95% [2,57; 12,07], $\chi^2 = 23,89$)
5) Вредные привычки у матери: злоупотребление алкоголем, наркотиками ($OR = 4,92$, ДИ 95% [3,60; 6,71], $\chi^2 = 124,16$)
6) Сердечно-сосудистые заболевания матери: хроническая артериальная гипертензия 2-й стадии ($OR = 4,06$, ДИ 95% [2,97; 5,54], $\chi^2 = 90,83$)
7) Пороки развития матки: есть ($OR = 3,71$, ДИ 95% [2,36; 5,84], $\chi^2 = 37,21$)
8) Смерть в неонатальном периоде: 2 случая и более ($OR = 3,71$, ДИ 95% [1,83; 7,51], $\chi^2 = 15,21$)

Выводы. Разработанный программный комплекс позволяет оптимизировать расчеты, используя статистические данные в матричном виде. Результаты работы программы вычисления многомерной оценки отношения шансов согласуются с результатами, полученными при вычислении в пакете Excel в работе [1].

ЛИТЕРАТУРА

1. Мельникова О. Д., Шиманова Ю. В. Оценка статистической значимости выявленной связи между исходом и фактором риска клинического течения беременности и родов. Фестиваль молодежной науки – 2018 : сб. материалов Всерос. Студенческой науч.-практ.конф. с междунар.участием, посвящ. 110-летию со дня рождения проф. П.Г. Подзолкова. Красноярск, 16-20 апр.2018 г. Красноярск: тип. КрасГМУ. 2018. с.86–87.

2. Медик В. А., Юрьев В. К. Общественное здоровье и здравоохранение: учеб. пособие для мед. вузов. М.: ГЭОТАР-Медиа, 2012. 608 с.

Melnikova Olga (Krasnoyarsk, Russia), Lukyanova Natalia (Krasnoyarsk, Russia)

Calculation odds ratio for multidimensional statistical data

Abstract. *A tool has been developed for calculating odds ratio indicators for multidimensional statistical data, with the help of which the influence of biomedical, age-gender and socially-professional factors on the risk of premature birth is analyzed.*

Key words: *applied statistics, odds ratio, risk factors*

Статистический анализ заболеваемости вирусным гепатитом в Красноярском крае

Д.И.Одарич [1], Н.А.Лукьянова [1,2], М.Г.Семенова [1]

[1] Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ,

[2] Институт математики и фундаментальной информатики Сибирский федеральный университет

E-mail: oddaria94@gmail.com, nalukyanovakrsk@gmail.com, semyonovamariya1095@gmail.com

Одной из самых серьезных проблем общественного здравоохранения является распространенность вирусным гепатитом (ВГ). Высокие показатели заболеваемости от данной патологии и ее серьезные осложнения в структуре смертности представляют проблему как на медицинском, так и на социально-экономических уровнях. В период 2000-2018гг. наблюдается тенденция снижения роста заболеваемости острым вирусным гепатитом в Российской Федерации. Причиной этому послужило успешное проведение профилактических мер, в том числе вакцинации населения против гепатитов А и В и контролируемая ревакцинация для групп риска. Динамика хронической формы данного заболевания представляет обратную картину. В рамках Сибирского федерального округа Красноярский край занимает второе место по заболеваемости гепатитом и является территорией риска. В целом, показатели заболеваемости, болезненности и смертности по краю от вирусного гепатита в период с 2009-2018 гг. имеют растущую тенденцию, поэтому проведение статистического анализа заболеваемости по отдельным муниципальным образованиям (МО) Красноярского края является актуальной задачей.

Цель работы: комплексная оценка заболеваемости и эпидемиологической обстановки по вирусному гепатиту в Красноярском крае.

Материалы и методы. Информационную базу исследования составили ежегодные официальные статистические материалы, подготовленные Красноярским краевым медицинским информационным аналитическим центром (ККМИАЦ) [1].

Комплексная оценка эпидемиологической ситуации по ВГ в Красноярском крае проводилась с использованием метода [2, 3, 4], основанного на ранжировании отдельных показателей (от минимума к максимуму) по территориям и во времени, с последующим суммированием полученных ран-

гов и вычислением коэффициента наглядности:

$$\text{Кн} = \left(1 - \frac{S_x - S_p}{S_x - S_y}\right) \cdot 100\%, \tag{1}$$

где S_x - наихудшая сумма мест; S_p - сумма мест у конкретного объекта; S_y - наилучшая сумма мест.

Если ранжирование проводится по территориям за n_t количество лет наблюдения, то в этом случае наихудшая сумма мест определяется по формуле:

$$S_x = n_{\text{показателей}} \cdot n_t \cdot n_{\text{территорий}}, \tag{2}$$

где $n_{\text{показателей}}$ — число эпидемиологических показателей, взятых для анализа; n_t – количество лет наблюдения, $n_{\text{территорий}}$ – количество исследуемых территорий, взятых для ранжировки. Наилучшая сумма мест в этом случае находится по формуле

$$S_y = n_{\text{показателей}} \cdot n_t. \tag{3}$$

Если же ранжирование проводится по временному ряду с n_t количеством лет наблюдения для исследования одной териитории, то в этом случае наихудшая сумма мест определяется по формуле:

$$S_x = n_{\text{показателей}} \cdot n_t, \tag{4}$$

а наилучшая сумма мест в этом случае соответствует числу показателей, взятых для анализа

$$S_y = n_{\text{показателей}}. \tag{5}$$

Для оценки заболеваемости и эпидемиологической ситуации возникает потребность разработать оценочную шкалу эпидемиологической обстановки (ЭО) на исследуемых территориях. Для этого недостаточно вычисления интегрального эпидемиологического показателя, необходимо произвести дополнительные статистическую обработку данных. Немалая часть экспертов сходятся во мнении, что единой, верной методологии оценки эпидемиологической ситуации не существует. В исследовании была применена методика для разработки шкалы оценки ЭО [5], основанная на вычислении доверительного медианного интервала [6, 7]. Расчет доверительных границ медианы с доверительной вероятностью 95% позволяет использовать этот интервал как инструмент градации.

Пусть задана выборка $X_n = \{x_1, \ldots, x_n\}$ некоторой случайной величины X. При $5 \leq n \leq 100$ доверительный интервал для медианы определяется

порядковыми статистиками

$$L \leq Me \leq U, \;\; L = x_k, \; U = x_{n-k+1}, \tag{6}$$

где

$$k = \left[\frac{1}{2}\left(n - z_{(1-\alpha/2)}\sqrt{n}\right)\right],$$

n - объем выборки, $z_{(1-\alpha/2)}$ - значение нормального распределения для выбранной доверительной вероятности (при $\alpha = 0.05 \;\; (P = 95\%)$ $z_{(1-\alpha/2)} = 1.96$), k - номер порядковой статистики, L и U - нижняя и верхняя граница доверительного интервала соответственно.

Алгоритм оценки эпидемиологической ситуации по методике [5]:

- Изучение фактического распределения по всем объектам (МО Красноярского края), ранжирование статистических показателей;

- Создание и градация оценочной шкалы с использованием доверительного интервала медианы;

- Оценка заболеваемости и эпидемической обстановки на рассматриваемом объекте по следующим критериям - благополучная, неустойчивая, неблагополучная, чрезвычайно неблагополучная.

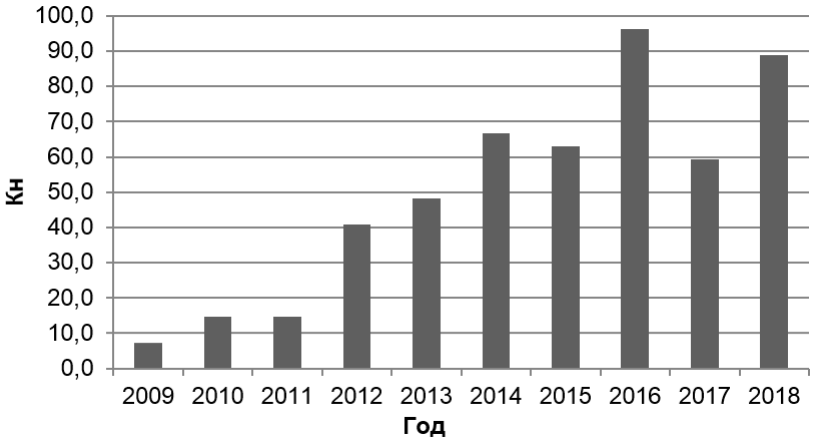
Результаты и обсуждение.

На основе данных ККМИАЦ [1]: численность населения по Красноярскому краю за 2009-2018 гг.; количество всего заболевших вирусным гепатитом по Красноярскому краю за 2009-2018 гг.; количество заболевших вирусным гепатитом впервые в жизни по Красноярскому краю за 2009-2018 гг.; количество умерших по причине вирусного гепатита по Красноярскому краю за 2009-2018 гг. были вычислены коэффициенты заболеваемости $k_{\text{забол.}}$, болезненности $k_{\text{болез.}}$, смертности $k_{\text{смерт.}}$ на 100 тыс. населения и проранжированы по возрастанию (таблица 1), а также расчитаны интегральные показатели в виде суммы мест рангов (ΣИП) и коэффициента наглядности (Кн, %) по формуле (1). В разрезе Красноярского края согласно (4) и (5) $S_x = 30; S_y = 3$, потому как период времени исследования 10 лет, а количество эпидемиологических показателей $n_{\text{показателей}} = 3$.

Рассчитанный интегральный показатель позволил наглядно представить ситуацию по вирусному гепатиту в Красноярском крае за 2009-2018гг. Таким образом видно, что самым неблагоприятным был 2016 г., в котором интегральный показатель в виде коэффициента наглядности достигает 96,3%, а самым благоприятным – 2009 г. с Кн=7,4%.

Таблица 9: Расчет интегральных показателей в виде суммы мест (ΣИП) и коэффициента наглядности (Кн, %) на примере Красноярского края за 2009-2018гг.

Год	$k_{\text{забол.}}$	ранг	$k_{\text{смерт.}}$	ранг	$k_{\text{болез.}}$	ранг	ΣИП	Кн, %
2009	79,18	1	0,2868	3	438,58	1	5	7,4
2010	89,46	3	0,2122	2	503,22	2	7	14,8
2011	84,44	2	0,2121	1	658,62	4	7	14,8
2012	99,60	7	0,3523	4	629,90	3	14	40,7
2013	98,44	6	0,5972	5	713,02	5	16	48,1
2014	121,35	9	0,6660	6	757,81	6	21	66,7
2015	96,34	5	0,7696	7	815,84	8	20	63,0
2016	125,94	10	1,4303	9	838,45	10	29	96,3
2017	96,06	4	1,2520	8	813,27	7	19	59,3
2018	107,35	8	1,4601	10	830,73	9	27	88,9



2-рисунок Интегральный показатель в виде коэффициента наглядности (Кн) по Красноярскому краю

Оценка эпидемиологической ситуации по отдельным МО Красноярского края и их группировка по степени напряженности эпидемиологической си-

туации с учетом интегрального показателя и коэффициента наглядности представлены в таблице 2.

Материалами исследования являются данные ККМИАЦ [1]: распределение численности населения Красноярского края по муниципальным образованиям за 2009–2017гг.; распределение больных с данным диагнозом, установленным впервые в жизни 2009–2017гг.; распределение зарегистрированных заболевших ВГ (всего) по муниципальным образованиям Красноярского края за 2009–2017гг. На данном этапе исследования используются только 2 эпидемиологических показателя – заболеваемость и болезненность, по той причине, что летальный исход у больных ВГ зачастую возникает не от названной инфекции непосредственно, а от тяжелых осложнений (гепатоцеллюлярная карцинома и цирроз печени), возникающих в процессе болезни. Значения наихудшей и наилучшей сумм рангов согласно (2) и (3) получились следующие: $S_x = 990$; $S_y = 18$.

Расчет интегрального показателя в виде суммы рангов (ΣИП) и коэффициента наглядности (Кн, %) по 55 МО Красноярского края за 2009-2017гг. представлен в таблице 2. Наихудший показатель ΣИП выявлен в Лесосибирске и составил 953. Показатель в Красноярске (административный центр Красноярского края) составил 706. Наилучший показатель ΣИП регистрируется в Шарыповском районе – это 85.



2-рисунок Оценочная шкала эпидемиологической ситуации по ВГ за 2009-2017 гг. по 55 территориях Красноярского края, Me – медиана, L и U - нижняя и верхняя граница доверительного интервала соответственно (6)

Медиана фактического распределения Me=465, для которой 95% доверительный интервал (6) составил [376, 579], длина интервала 203. Коли-

чество территорий с ΣИП выше медианного значения составляет 27 территорий. Составлена оценочная шкала эпидемиологической обстановки по вирусному гепатиту за 2009-2017гг. по 55 муниципальным образованиям Красноярского края, которая подразделяется на 4 группы (рисунок 2): благополучная, неустойчивая, неблагополучная, чрезвычайно неблагополучная. Градацию эпидемиологической обстановки по отдельным территориям проводим по интегральному показателю в виде суммы мест по всем МО Красноярского края (55 территорий) за 2009-2017гг. (таблица 3).

– Благополучные территории (35 субъектов): Шарыповский, Назаровский, Краснотуранский, Казачинский, Минусинский, Богучанский, Саянский, Тюхтетский, Манский, Мотыгинский, Большемуртинский, Нижнеингашский, Ермаковский, Северо-Енисейский, Рыбинский, Сухобузимский, Ирбейский, Ачинский, Бирилюсский, Канский, Абанский, Туруханский, Идринский, Боготол, Иланский, Курагинский, Козульский, Новоселовский, Каратузский, Кежемский, Шушенский, Емельяновский, Уярский, Балахтинский, Большеулуйский.

– Территории с неустойчивой эпидемиологической ситуацией по ВГ (12 субъектов): Дзержинский, Пировский, Тасеевский, Шарыпово, Минусинск, Назарово, Ужурский, Красноярск, Ачинск, Сосновоборск, Березовский, Эвенкийский.

– Территории неблагополучные (8 субъектов): Бородино, Канск, Норильск, Партизанский, Енисейский, Таймырский, Дивногорск, Лесосибирск.

Чрезвычайно неблагополучных территорий не выявлено.

Выводы. Проведенный с использованием интегрального показателя анализ позволил оценить эпидемиологическую ситуацию по вирусному гепатиту в Красноярском крае в ее многолетней динамике и выделить районы и группы районов с разной степенью их эпидемиологического благополучия, что в дальнейшем позволит увеличить финансирование таких ключевых звеньев здравоохранения, как профилактика, диагностика и лечение вирусного гепатита на этих территориях. Как следствие, поспособствовать снижению смертности как от патологии непосредственно, так и от его летальных осложнений. Подводя итог, можно сказать, что в городах краевого значения с большей плотностью населения, эпидемиологическая ситуация по вирусному гепатиту до сих пор остается напряженной. Что, несомненно, наносит, ущерб стране, как на медицинском, так и на социально-экономическом уровнях.

ЛИТЕРАТУРА

1. Официальный сайт Красноярского краевого медицинского информационно-аналитического центра. [Электронный ресурс]. - Режим доступа: https://www.kmiac.ru/ (дата обращения: 15.09.2019).

2. Астафьев В. А, Савилов Е. Д, Чемезова Н. Н, Степаненко Л. А. Оценка заболеваемости вирусным гепатитом С в Иркутской области по интегральному эпидемиологическому показателю. Сиб. мед. журн. (Иркутск). 2012. №6. с. 120–122.

3. Савилов Е. Д., Мамонтова Л. М., Астафьев В. А., Жданова С. Н. Применение статистических методов в эпидемиологическом анализе. М: МЕДпресс-информ, 2004. 112 с.

4. Савилов Е. Д., Астафьев В. А., Жданова С. Н. Эпидемиологический анализ: Методы статистической обработки материала. М.: Наука-Центр, 2011. с. 76–91.

5. Колпаков С. Л., Яковлев А. А. О методологии оценки эпидемиологической ситуации. Эпидемиология и инфекционные болезни, 2015. № 20 (4). с. 34–39.

6. ГОСТ Р ИСО 16269-7. Статистическое представление данных. Медиана. Определение точечной оценки и доверительных интервалов. 2004. 8 с.

7. Наркевич А. Н., Наркевич А. А., Виноградов К. А. Интервальная оценка медианы и ее автоматизация. Врач и информационные технологии, 2013. № 4. с. 40–49

Odarich Daria (Krasnoyarsk, Russia), Lukyanova Natalia (Krasnoyarsk, Russia), Semenova Maria (Krasnoyarsk, Russia)

Statistical analysis of the viral hepatitis disease in Krasnoyarsk area

Abstract. *In the article is presented the data of the complex epidemiological analysis of viral hepatitis in Krasnoyarsk area by the integral index.*

Key words: *applied statistics, confidence interval median, integral index*

Таблица 10: Расчет интегрального показателя в виде суммы рангов (ΣИП) и коэффициента наглядности (Кн, %) по 55 МО Красноярского края за 2009-2017гг.

Территория	ΣИП	Кн, %	Ранг	Территория	ΣИП	Кн, %	Ранг
Шарыповский	85	6,89	1	Новоселовский	465	45,99	28
Назаровский	132,5	11,78	2	Каратузский	472	46,71	29
Краснотуранский	146,5	13,22	3	Кежемский	489	48,46	30
Казачинский	222	20,99	4	Шушенский	531	52,78	31
Минусинский	224,5	21,24	5	Емельяновский	547	54,42	32
Богучанский	225,5	21,35	6	Уярский	551,5	54,89	33
Тюхтетский	236	22,43	7	Большеулуйский	571	56,89	34
Мотыгинский	250	23,87	8	Балахтинский	577	57,51	35
Саянский	251	23,97	9	Дзержинский	579	57,72	36
Большемуртинский	253	24,18	10	Тасеевский	637	63,68	37
Нижнеингашский	253	24,18	10	Пировский	641,5	64,15	38
Манский	254	24,28	12	Шарыпово	664	66,46	39
Ермаковский	271	26,03	13	Минусинск	677	67,80	40
Рыбинский	289	27,88	14	Назарово	685	68,62	41
Сухобузимский	292,5	28,24	15	Ужурский	696	69,75	42
Северо-Енисейский	302,5	29,27	16	Красноярск	706	71,26	43
Ирбейский	347	33,85	17	Ачинск	715	71,71	44
Ачинский	355	34,67	18	Сосновоборск	763	76,65	45
Канский	367	35,91	19	Березовский	773	77,67	46
Бирилюсский	376	36,83	20	Эвенкийский МР	775	77,88	47
Абанский	393,5	38,63	21	Бородино	789	79,32	48
Иланский	428	42,18	22	Канск	846	85,19	49
Боготол	433,5	42,75	23	Норильск	848	85,39	50

Таблица 11: Градация территорий по благополучию эпидемиологической ситуации

Эпидемиологическая ситуация	Критерии для оценки ЭС
Благополучная	Сумма ранговых мест $[18, 579]$
Неустойчивая	Сумма ранговых мест $(579, 783]$
Неблагополучная	Сумма ранговых мест $(783, 985]$
Чрезвычайно неблагополучная	Сумма ранговых мест $(985, 990]$

Анализ влияния территориального признака на заболеваемость и ожидаемую продолжительность жизни населения Красноярского края по причине болезней системы кровообращения

Е.А.Савкина [1], **Н.А.Лукьянова** [1,2]

[1] Красноярский государственный медицинский университет имени профессора В.Ф.Войно-Ясенецкого МЗ РФ,

[2] Институт математики и фундаментальной информатики Сибирский федеральный университет

asoletr@gmail.com, nalukyanovakrsk@gmail.com

Одной из самых актуальных проблем практического здравоохранения и научно-исследовательской медицины являются болезни системы кровообращения (БСК). Эти болезни на сегодняшний момент занимают вторую строчку в списке заболеваемости Красноярского края. Заболеваемость и смертность населения по причине БСК принадлежат к важным статистическим показателям, которые характеризуют санитарное положение в Красноярском крае. В структуре смертности они на протяжении последних 10 лет прочно занимают первое место. Поэтому показатели смертности и заболеваемости по причине БСК нуждаются в постоянном мониторинге и принадлежат к важным статистическим показателям, характеризующим оценку диагностики, профилактики и лечения БСК. Объект исследования – болезни системы кровообращения населения Красноярского края. Предмет исследования – показатели заболеваемости и смертности населения по причине болезней системы кровообращения. В современных условиях на данные показатели также могут влиять и физико-географические, и экологические, и социально-экономические факторы. В связи с этим возникает необходимость провести оценку влияния территориального фактора на исследуемые показатели.

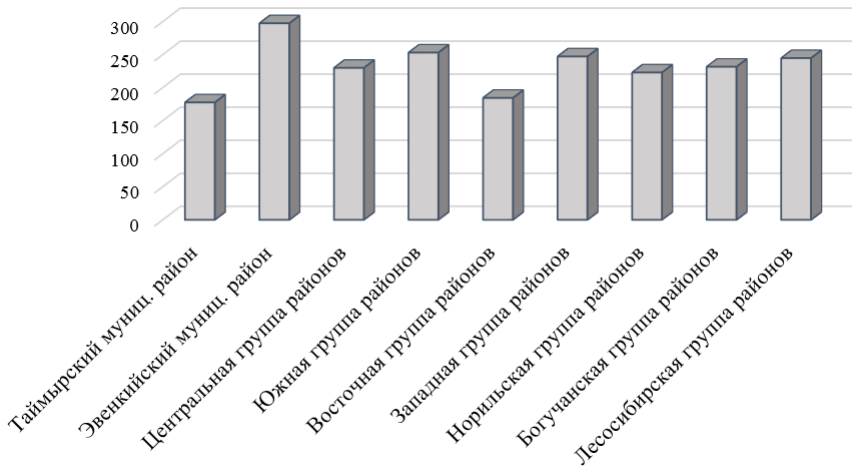
Материалы и методы. В работе использовалась информационная база, которую составляли статистические данные Федеральной службы государственной статистики РФ и Красноярского краевого медицинского информационного аналитического центра [2]. Для обработки данных применялись прикладные пакеты программы Microsoft Excel, Statistica 6.0 [3]. Методы исследования: аналитические, статистические [4, 5, 6].

Оценка влияния территориального признака на показатель заболеваемости. Проведение оценки влияния территориального признака по показателю общей заболеваемости по причине БСК за 2017 год (на 1000

населения) осуществлялось для следующих четырех переменных:

1) все население;
2) дети;
3) подростки;
4) взрослые.

Распределение групп районов Красноярского края по показателю общей заболеваемости по причине БСК (на 1000 населения) за 2017 год показано на рисунке 1.



1-рисунок Заболеваемость по группам районов Красноярского края (на 1000 населения) за 2017 год

Проведение оценки влияния территориального признака проводилось в пакете Statistica при помощи дисперсионного анализа по непараметрическому критерию Краскела-Уоллиса, поскольку распределения показателей общей заболеваемости; общей заболеваемости детей; общей заболеваемости подростков; общей заболеваемости взрослых имеют ненормальное распределение. Для анализа рассматривается один воздействующий фактор – территориальный признак, который имеет 8 групп исследования, по которым распределены 56 муниципальных образований Красноярского края. Здесь 7 групп это Центральная группа районов (15 субъектов), Южная группа районов (8 субъектов), Восточная группа районов (8 субъектов), Западная группа районов (13 субъектов), Норильская группа районов (2 субъекта), Богучанская группа районов (3 субъекта), Лесосибирская группа районов (5 субъектов), а восьмая группа для данного исследования представляет объединение Таймырского муниципального района и Эвенкийского муниципального района (группе дано название Таймыр-Эвенкия).

Общая заболеваемость по все четырем переменным (все население; дети; подростки; взрослые) на территории Красноярского края однород-

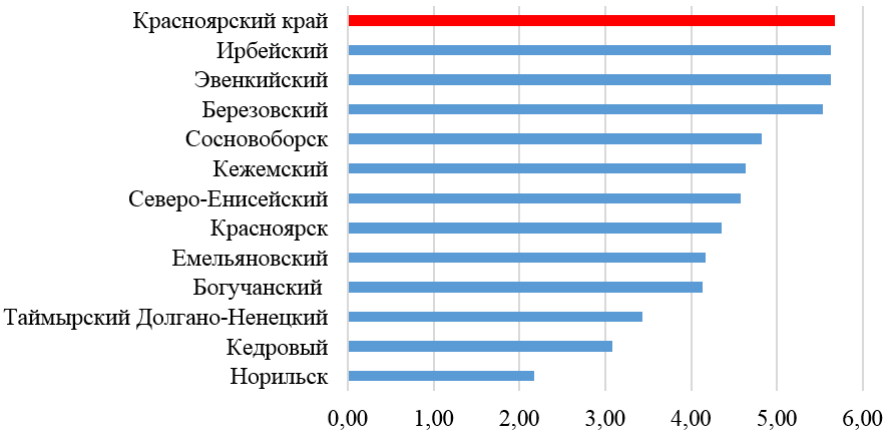
на. Результаты непараметрического дисперсионного анализа по критерию Краскела-Уоллиса по уровню значимости для каждой группы показателей следующие: по общей заболеваемости $p = 0,558$, общей заболеваемости детей $p = 0,403$, общей заболеваемости подростков $p = 0,051$ и общей заболеваемости взрослых $p = 0,548$. Каждое из значений уровня p оказалось больше уровня значимости 0,05, следовательно, можно сказать, что основная гипотеза о случайности различий между выборками принимается. Иными словами, территориальный признак не оказывает влияния на показатели общей заболеваемости, общей заболеваемости детей, общей заболеваемости подростков и общей заболеваемости взрослых .

Оценка влияния территориального признака на ожидаемую продолжительность жизни. При анализе популяционного здоровья используется показатель ожидаемой продолжительности жизни (ОПЖ). Ожидаемая продолжительность жизни – медико-демографический показатель, характеризующий возрастную структуру смертности и численности населения. ОПЖ показывает, какова была бы средняя продолжительность жизни новорожденного, если бы и далее повозрастная смертность оставалась такой же, как на данный момент времени. ОПЖ рассчитывается путем построения таблиц смертности и таблиц дожития [1].

Смертность на территории Красноярского края не однородна. В связи с этим возникает необходимость анализа на краевом уровне. Была выдвинута гипотеза о том, что территориальные различия в уровне ожидаемой продолжительности жизни (ОПЖ) определяются смертностью от болезней системы кровообращения. Для подтверждения гипотезы был проведен анализ влияния территориального фактора на исследуемый показатель за 2017 год.

Анализ ОПЖ предусматривает дифференциацию по полу, типу населения (городское, сельское), территориям и исследование динамики. Проведен анализ ожидаемой продолжительности жизни по городским округам и муниципальным образованиям Красноярского края и анализ на основе статистических данных ККМИАЦ о численности населения и количество умерших от болезни системы кровообращения по 57 территориям (города и районы) края за 2017 год.

Ранжирование территорий по уровню смертности населения по причине БСК в Красноярском крае в 2017 году позволило выделить 12 территорий края (рисунок 2), где значение коэффициента смертности меньше, чем средний показатель по краю, что нельзя сказать про остальные 45 субъектов края.



2-рисунок Ранговое распределение территорий Красноярскому краю по показателю смертности по причине БСК в 2017 году

Для оценки влияния территориального признака на ожидаемую продолжительность жизни определяется однородность совокупности при помощи коэффициента вариации равного 11,37%. Полученное значение меньше 30%, поэтому совокупность однородна [6] по признаку ожидаемой продолжительности жизни по причине БСК. По результатам проверки распределения на нормальность по критерию Колмогорова-Смирнова нулевая гипотеза о нормальном распределении муниципальных районов Красноярского края по ожидаемой продолжительности жизни принимается при уровне значимости 0,05.

Проведем группировку районов Красноярского края по территориальному признаку. Рассмотрено 8 групп, для каждой из групп исследования вычисляется среднее значение ОПЖ (таблица 1).

Рассчитана межгрупповая дисперсия – $\delta^2 = 14,05$. Общая дисперсия изучаемого признака равна $\sigma^2 = 37,33$. Далее найден эмпирический коэффициент детерминации, который широко применяется в статистическом анализе и является показателем, представляющим долю межгрупповой дисперсии в дисперсии результативного признака и характеризует силу влияния группировочного признака на образование общей вариации [6].

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{14,05}{37,3} = 0,377.$$

Эмпирический коэффициент детерминации показывает долю вариации результативного признака под влиянием факторного признака, таким образом, на 37,7% вариация ожидаемой продолжительности жизни обусловлена различиями по территориальному признаку, а на 62,3% другими факторами.

Таблица 12: Расчет среднего значения ОПЖ по причине БСК за 2017 год по группам районов Красноярского края

Группа районов (ГР)	количество субъектов	среднее значение ОПЖ
Центральная ГР	15	52,30
Южная ГР	8	54,15
Восточная ГР	8	53,15
Западная ГР	13	56,07
Норильская ГР	2	53,42
Богучанская ГР	3	53,66
Лесосибирская ГР	6	56,38
Таймыр-Эвенкия	2	41,25

Эмпирическое корреляционное отношение показывает тесноту связи между статистическими данными:

$$\eta = \sqrt{0.377} = 0.614$$

По шкале Чеддока территориальный признак влияет на ожидаемую продолжительность жизни заметно [6].

Выводы. По результатам исследования не обнаружено влияния территориального признака на показатель общей заболеваемости по причине БСК, что может говорить о массовости явлений возникновения и распространения данной патологии среди населения. Вариация ожидаемой продолжительности жизни по причине болезни системы кровообращения на 37,7% обусловлена различиями по территориальному признаку, по шкале Чеддока связь заметная.

ЛИТЕРАТУРА

1. МР 2.1.10.0082-13. Методы оценки медико-демографической ситуации на популяционном уровне: Методические рекомендации. М.: Феде-

ральный центр гигиены и эпидемиологии Роспотребнадзора, 2014. 31 с.

2. Официальный сайт Красноярского краевого медицинского информационно-аналитического центра. [Электронный ресурс]. - Режим доступа: https://www.kmiac.ru/ (дата обращения: 15.09.2019).

3. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М.: Медиа Сфера, 2003. 312 с.

4. Петри А., Сэбин К. Наглядная статистика в медицине / пер. с англ. В.П. Леонова. М.: ГЭОТАР-МЕД, 2003. 144 с.

5. Юнкеров В. И., Григорьев С. Г. Математико-статистическая обработка данных медицинских исследований. СПб.: ВМедА., 2002. 266 с.

6. Балинова В. С. Статистика в вопросах и ответах: Учеб.пособие. М.: ТК.Велби, Изд-во Проспект, 2004. 344 с.

Savkina Evgeniya (Krasnoyarsk, Russia), Lukyanova Natalia (Krasnoyarsk, Russia)

Analysis of influence geographic position on the morbidity and life expectancy of the population of the Krasnoyarsk area due to diseases of the circulatory system

Abstract. *In the article are presented the results of a statistical study of the influence of a territorial attribute on indicators of population morbidity and life expectancy due to diseases of the circulatory system for 2017 using the empirical determination coefficient and correlation ratio.*

Key words: *applied statistics, empirical determination coefficient, correlation ratio*

Полиномиальный алгоритм поиска приближенного решения в графе с ограничением

А.А. Солдатенко
Сибирский федеральный университет, г. Красноярск, Россия
glinckon@gmail.com

В работе рассматривается задача о ресурсоограниченном кратчайшем пути (Resource Constrained Shortest Path problem, RCSP). Данная задача является NP-трудным расширением хорошо известной задачи о кратчайшем пути и имеет многие реальные приложения, например мультисервисные телекоммуникационные сети, логистические и дорожные сети, и пр. Традиционно задача RCSP формулируется в терминах теории графов и целочисленного линейного программирования. В теоретико-графовой формулировке рассматриваемая сеть представляется ориентированным графом, вершины которого соответствуют узлам сети, а дуги — каналам связи. Предполагается, что всякая дуга обладает основным весом, например стоимостью, а также некоторыми дополнительными весами, отражающими потребности в ресурсах, которые нужны для передвижения по этой дуге. Требуется найти кратчайший путь между двумя заданными узлами сети, удовлетворяющий заданным ограничениям на итоговые ресурсные затраты, которые необходимы для прохождения этого пути. Известно, что даже при одном ограничении задача RCSP является NP-трудной [1].

В настоящее время выделяют три класса методов и соответствующих им алгоритмов, способных находить точное или приближенное решение задачи RCSP: методы ранжирования путей [2], методы маркировки вершин [1, 3, 4], методы лагранжевой релаксации [5, 6]. Первые два класса основаны на теоретико-графовой постановке задачи, в то время как методы третьего класса исходят из постановки задачи RCSP на языке целочисленного линейного программирования. К сожалению, большинство существующих алгоритмов решения задачи RCSP медленно работают на сетях большой размерности, для многих современных приложений совершенно неприемлемы временные характеристики данных алгоритмов и их многочисленных усовершенствованных версий [3]. Поэтому остается актуальной разработка приближенных алгоритмов решения задачи RCSP, способных быстро находить решение на сетях большой размерности [3, 6].

Сформулируем задачу RCSP в терминах теории графов. Пусть сеть описана взвешенным ориентированным графом (далее просто графом) $G = (V, E)$ без кратных дуг и петель, в котором каждая вершина $v \in V$ пред-

ставляет узел сети, а каждая дуга $e \in E$ — канал связи между соответствующими узлами сети, при этом $n = |V|$ и $m = |E|$. Считаем, что на множестве дуг графа $G = (V, E)$ задана функция $w(e)\colon E \to \mathbb{R}^+$, ставящая в соответствие каждой дуге $e \in E$ ее вес $w(e) > 0$. Пусть для вершин $s, d \in V$ в графе G существует путь P, идущий от вершины s к вершине d. Полагаем, что вес этого (s, d)-пути P вычисляется как сумма весов всех входящих в него дуг:

$$w(P) = \sum_{e \in P} w(e), \tag{1}$$

т. е. функция $w(e)$ является аддитивной. Если $w(e)$ — стоимость передвижения по дуге e, то $w(P)$ можно интерпретировать как стоимость прохождения (s, d)-пути P в графе G.

Пусть для каждой дуги графа G также заданы функции $r_i(e)\colon E \to \mathbb{R}^+$, $i = 1, \ldots, k$, отражающие ресурсные потребности, которые необходимы для передвижения по этой дуге, и всегда $r_i(e) > 0$. Предполагается, что все эти функции аддитивные, т. е. для любого (s, d)-пути P верны равенства

$$r_i(P) = \sum_{e \in P} r_i(e), \ i = 1, \ldots, k. \tag{2}$$

Кроме того, определены величины $R_i \in \mathbb{R}^+$, $i = 1, \ldots, k$, определяющие ресурсные ограничения рассматриваемой мультисервисной сети. Всякий (s, d)-путь P называется допустимым, если он удовлетворяет ресурсным ограничениям

$$r_i(P) \le R_i, \ i = 1, \ldots, k. \tag{3}$$

Требуется найти (s, d)-путь P, который минимизирует целевую функцию (1) и удовлетворяет ограничениям (3). Такой путь определяет оптимальное решение задачи RCSP и называется ресурсноограниченным кратчайшим (s, d)-путем в графе $G = (V, E)$. Всякий допустимый (s, d)-путь можно рассматривать в качестве приближенного решения данной задачи. В настоящей работе предлагается модификация алгоритма RevTree, предложенного в работе [7], которая сохраняет свойства исходного алгоритма, однако на практике требует значительно меньше времени для нахождения решения.

Под размерностью задачи RCSP понимаются значения n и m — число вершин и дуг графа $G = (V, E)$ соответственно, а под ее параметрами — значения функций $w(e), r_i(e)$. Для краткости вместо $r_i(e)$ и R_i будем писать $r(e)$ и R. Опишем алгоритм RevTree для случая, когда имеется только один ресурс R_i и $w(e), r_i(e)$ — положительные вещественнозначные функции.

Алгоритм RevTree состоит из двух фаз, на каждой из которых однократно выполняется известный алгоритм Дейкстры [8]. На первой фазе, исходя из функций $r(e)$, $e \in E$, вычисляется дерево путей минимального веса с корнем в вершине d. Это дерево определяет для каждой вершины $u \in V$ такой (u, d)-путь, вес которого указывает минимальный ресурс для прохождения (u, d)-пути. Обозначим (u, d)-путь через P_2.

Пусть P_1 — путь из стартовой вершины s в текущую вершину v, найденный как некоторое допустимое решение задачи RCSP для вершин s и v. Согласно формуле (2), для прохождения этого (s, v)-пути затрачен ресурс величины $r(P_1)$. Множество $\Gamma(v)$ определяет возможные направления дальнейшего движения по дугам графа G из вершины v. Тогда для перемещения из вершины v в вершину $u \in \Gamma(v)$ необходимо выполнение условия

$$r(P_1) + r(v, u) + \pi(u) \leq R, \tag{4}$$

где $\pi(u) = \sum\limits_{e \in P_2} r(e)$. Условие (4) гарантирует, что путь (P_1, e, P_2), где $e = (v, u) \in E$, является допустимым решением задачи RCSP.

На второй фазе вновь выполняется алгоритм Дейкстры, но только с усеченными окрестностями вершин. Усечение окрестности $\Gamma(v)$ для текущей вершины $v \in V$ осуществляется следующим образом: если для $u \in \Gamma(v)$ нарушается условие (4), то она удаляется из $\Gamma(v)$. Таким образом, алгоритм RevTree уменьшает мощность множеств $\Gamma(v)$ с учетом ресурсного ограничения R, что позволяет находить допустимое решение задачи RCSP, если оно существует.

Как было показано в работе [7], алгоритм RevTree для нахождения решения задачи RCSP затрачивает $\mathcal{O}(n^2)$ времени и $\mathcal{O}(n)$ памяти, при этом решение отклоняется от оптимального решения задачи RCSP не более чем на величину $\varepsilon = \lambda_{\max}/\lambda_{\min} - 1$, где $\lambda_{\min} = \min_{e \in E} (r(e)/w(e)) > 0$, $\lambda_{\max} = \max_{e \in E}(r(e)/w(e)) > 0$.

В настоящей работе предлагается алгоритм RevTreeShort, который снижает трудоемкость первой фазы алгоритма RevTree. В алгоритме RevTree на первой фазе находится полное дерево путей минимального веса, вычисляемых по весовой функции $r(e), e \in E$. Ускорение данной фазы без потери точности решение может быть достигнуто, путем вычисления не полного дерева путей, а до тех пор, пока не будет включена стартовая вершина s. Если в процессе выполнения второй фазы будет затронута вершина v для которой не найдено значение минимального веса $\pi(v)$, то первая фаза возобновляется до тех пор, пока в дерево путей не будет включена вершина

Таблица 13: Результаты сравнения алгоритмов RevTree и RevTreeShort

Название графа	n	m	Время работы алгоритма RevTree, с	Время работы алгоритма RevTreeShort, с
G_1	500	3923	2,404	1,910
G_2	1000	16345	2,988	0,837
G_3	1500	36655	94,849	55,151
DC	9559	29818	330,098	203,964
AK	69082	157662	6584,240	1198,150

v. Для возобновления первой фазы требуется множество метоков всех вершин графа, которые используются в алгоритме Дейкстры. Хранение этой информации требует $\mathcal{O}(n)$ памяти. Алгоритм RevTreeShort не превосходит RevTree по сложности, поскольку даже в худшем случае первая фаза будет совпадать с аналогичной фазой алгоритма RevTree. В предложенной модификации сохранена основная логика алгоритма RevTree, поэтому оценки полученные в работе [7] могут быть применены к алгоритму RevTreeShort.

Для оценки результативности алгоритма RevTreeShort проведены вычислительные эксперименты на компьютере с процессором Intel Core i7-7700K Processor (8 MB Cache, 3,60 ГГц) и ОЗУ объемом 16 Гбайт. Осуществлялось сравнение программной реализации алгоритмов RevTree и RevTreeShort по времени работы. Эксперименты проводились на случайно сгенерированных графах G_1–G_3 и реальных дорожных сетях DC, AK из базы DIMACS для последовательности из 1000 случайно сгенерированных (s, d)-запросов. Для случайной генерации графов применялся метод Ваксмена с параметрами $\alpha = 0{,}15$, $\beta = 0{,}25$, который традиционно используется для генерации графов топологически схожих с реальными компьютерными сетями [9]. Результаты экспериментов представлены в таблице.

В работе предложена модификация полиномиального приближенного алгоритма RevTree, которая дает существенный эффект по времени. Алгоритм RevTreeShort требует $\mathcal{O}(n^2)$ времени и всегда находит допустимое решение задачи, если оно существует, отклоняясь от оптимального решения на величину, не превышающую $\varepsilon = \lambda_{\max}/\lambda_{\min} - 1$, зависящую от исходного графа. Целесообразны дальнейшие исследования данного алгоритма, с целью увеличения области его применения.

ЛИТЕРАТУРА

1. Joksch H. C. The shortest route problem with constraints. Math. Analysis Appl., 1966. V. 14. P. 191–197.

2. Di Puglia Pugliese L., Guerriero F. A survey of resource constrained shortest path-problems: exact solution approaches. Networks, 2013. P. 183–200.

3. Zhu X., Wilhelm W. E. A three-stage approach for the resource-constrained shortest path as a sub-problem in column generation. Computers & Operations Research, 2012. V. 39. Iss. 2. P. 164–178.

4. Dumitrescu I., Boland N. Improved preprocessing, labeling and scaling algorithms for the weight-constrained shortest path problem. Networks, 2003. V. 42. P. 135–153.

5. Jepsen M., Petersen B., Spoorendonk S., Pisinger D. A branch-and-cut algorithm for the capacitated profitable tour problem. Discrete Optimization, 2014. V. 14. P. 78–96.

6. Horvath M., Kis T. Solving resource constrained shortest path problems with LP-based methods. Computers & Operations Research, 2016. V. 73. P. 150–164.

7. Солдатенко А. А. Приближенный алгоритм поиска оптимального маршрута в сети с ограничением. ПДМ. Приложение, 2019. № 12, с. 186Ц191.

8. Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. Л., Штайн К. Алгоритмы. Построение и анализ. М.: Вильямс, 2018. 1328 с.

9. Pathan A. K., Monowar M. M., Khan S. Simulation Technologies in Networking and Communications: Selecting the Best Tool for the Test. CRC Press, 2017. 648 p.

Aleksandr Soldatenko (Krasnoyarsk, Russia)

Polynomial approximation algorithm for search shortest path in graph with restrictions

Abstract. *In the paper, we consider the Resource Constrained Shortest Path problem (RCSP). This problem is NP-hard extension of a well-known shortest path problem in the directed graph* $G = (V, E)$*. In the RCSP problem each arc*

e from E has a cost $w(e)$ and additional weight functions $r_i(e)$, $i = 1, \ldots, k$, which specifying its requirements from a finite set of resource. The RCSP problem has various practical applications, including design and operation of multi-service network. Nowadays, multi-service networks grow at a rapid pace. Therefore, it is relevant to search for a new approximation algorithms that can solve the RCSP problem quickly. A polynomial time modification of existing approximation algorithm is presented in the paper as well as results of computational experiments, which justify the effectiveness of proposed modification.

Key words: *resource constrained shortest path, graph-based algorithm, optimal routing*

On a number of empty cells in allocation scheme with even number of particles in each cell

F.A.Abdushukurov, A.N.Chuprunov
achuprunov@mail.ru

Abstract. We consider the random variable $\mu_0(2n, N)$ that is a number of empty cells in allocation scheme of $2n$ distinguishable particles by N different cells under the condition: each cell contains even number of particles. For various type of the convergence $N, n \to \infty$ we prove the convergence of $\mu_0(2n, N)$.

Key words: allocation scheme, Gaussian random variable, Poisson limit theorem

Mathematical Subject Classification: 60C05, 60F05, 62G30

Let n, N be integer numbers. A homogeneous allocation scheme of n distinguishable particles by N different cells is named the random variables $\eta'_1, \dots, \eta'_N$, with the joint distribution defined by formula

$$\mathbb{P}\{\eta'_1 = k_1, \dots \eta'_N = k_N\} = \frac{n!}{k_1!k_2! \cdots k_N!} \left(\frac{1}{N}\right)^n, \tag{1}$$

where $k_1, k_2, \dots k_N$ are nonnegative integer number such that $k_1+k_2+\cdots+k_N = n$. Many papers deal with limit theorems for allocation scheme of distinguishable paticles by different cells (see (Kolcnin, Sevastianov and Chistiakov, [1]) and its references).

We will consider the random variables $\eta_1, \dots, \eta_N$, with the joint distribution defined by formula

$$\mathbb{P}\{\eta_1 = 2k_1, \dots \eta_N = 2k_N\} =$$

$$= \mathbb{P}\{\eta'_1 = 2k_1, \dots \eta'_N = 2k_N \mid \eta'_i \text{ takes even values}, 1 \le i \le N\}, \tag{2}$$

where $k_1, k_2, \dots k_N$ are nonnegative integer number such that $k_1+k_2+\cdots+k_N = n$.

Observe that the scheme (2) is an allocation scheme of $2n$ distinguishable particles by N different cells under the condition: each cell contains even number of particles.

Denote by $\stackrel{d}{=}$ an equality of distributions. We will consider the random variables

$$\mu_0(2n, N) = \sum_{i=1}^{N} I_{\{\eta_i=0\}} \quad \text{and} \quad \mu'_0(n, N) = \sum_{i=1}^{N} I_{\{\eta'_i=0\}}.$$

Observe that $\mu_0'(n,N)$ and $\mu_0(2n,N)$ are numbers of empty cells in scheme (1) and scheme (2), correspondingly.

Denote: δ_l is a number of particles which throw with a moment of filling of $l-1$ cells to a moment of filling of l cells in the scheme (2), δ_l' is a number of particles which throw with a moment of filling of $l-1$ cells to a moment of filling of l cells in the scheme (1), $1<l\le N$, $\delta_1=1$, $\delta_1'=1$. Then the random variables

$$\nu_k=\delta_1+\delta_2+\cdots+\delta_k, \quad \text{and} \quad \nu_k'=\delta_1'+\delta_2'+\cdots+\delta_k'$$

are minimal numbers of particles under which at first exactly k cells is occupated in scheme (2) and scheme (1), correspondingly. The random variables ν_k' and δ_l' was defined in [1] (see p. 17).

Then we have (see (33) in p. 17, [1])

$$\mathbf{P}(\nu_k'\le n)=\mathbf{P}(\mu_0'(n,N)\le N-k) \quad \text{and} \quad \mathbf{P}(\nu_k\le 2n)=\mathbf{P}(\mu_0(2n,N)\le N-k).$$

Denote: A is an event such that 2 particles allocate one cell from the set S and $|S|=N-(l-1)$, B is an event such that 2 particles allocate one cell. Then it holds

$$\mathbf{P}(A)=\frac{N-(l-1)}{N^2}, \quad \mathbf{P}(B)=\frac{N}{N^2}=\frac{1}{N}$$

and $A\subset B$. Therefore we have

$$\mathbf{P}(A\mid B)=\frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)}=\frac{\frac{N-(l-1)}{N^2}}{\frac{1}{N}}=\frac{N-(l-1)}{N}$$

and

$$\mathbf{P}(\delta_l=2m)=\left(\frac{l-1}{N}\right)^{m-1}\left(1-\frac{l-1}{N}\right)=\mathbf{P}(\delta_l'=m), \quad m=0,1,2\ldots.$$

Consequently, $\mu_0'(n,N)\overset{d}{=}\mu_0(2n,N)$ and we can use the limit theorems, which obtained for $\mu_0'(n,N)$ in [1] to formulate limit theorems for $\mu_0(2n,N)$.

Denote: $\alpha=\frac{2n}{N}$, $\mathbf{E}\mu_0$ is an expectation of $\mu_0(2n,N)$, $\mathbf{D}\mu_0$ is a variance of $\mu_0(2n,N)$.

We say that N,n belong to right domain if $n,N\to\infty$ such that

$$\alpha\to\infty \quad \text{and} \quad \mathbf{E}\mu_0\to\lambda<\infty.$$

From Theorem 1 on p.13 in [1] it follows

Theorem 1. *In the right domain we have*

$$\lim_{N\to\infty} \mathbf{P}(\mu_0(2n,N)=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,\dots.$$

We say that N,n belong to central domain if $n,N\to\infty$ such that

$$c_1<\alpha<c_2, \quad \text{for some} \quad 0<c_1<c_2<\infty.$$

From Theorem 1 on p. 22 in [1] it follows

Theorem 2. *In the central domain we have*

$$\lim_{N\to\infty} \mathbf{P}\left(\frac{\mu_0(2n,N)-\mathbf{E}\mu_0}{\sqrt{\mathbf{D}\mu_0}}<t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-u^2/2}du, \quad t\in\mathbf{R}.$$

We say that N,n belong to left intermediate domain if $n,N\to\infty$ such that

$$\alpha\to 0 \quad \text{and} \quad \frac{N\alpha^2}{2}\to\infty.$$

From Theorem 3 on p. 30 in [1] it follows

Theorem 3. *In the left intermediate domain we have*

$$\lim_{N\to\infty} \mathbf{P}\left(\frac{\mu_0(2n,N)-\mathbf{E}\mu_0}{\sqrt{\mathbf{D}\mu_0}}<t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-u^2/2}du, \quad t\in\mathbf{R}.$$

We say that N,n belong to right intermediate domain if $n,N\to\infty$ such that

$$\alpha\to\infty \quad \text{and} \quad \mathbf{E}\mu_0\to\infty.$$

From Theorem 2 on p. 28 in [1] it follows

Theorem 4. *In the right intermediate domain we have*

$$\lim_{N\to\infty} \mathbf{P}\left(\frac{\mu_0(2n,N)-\mathbf{E}\mu_0}{\sqrt{\mathbf{D}\mu_0}}<t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-u^2/2}du, \quad t\in\mathbf{R}.$$

We say that N,n belong to left domain if $n,N\to\infty$ such that

$$\frac{N\alpha^2}{2}\to\lambda<\infty.$$

From Theorem 1 on p. 31 in [1] it follows

Theorem 5. *In the left domain we have*

$$\lim_{N\to\infty} \mathbf{P}\left(\mu_0(2n,N) - (N-2n) = k\right) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots.$$

So in Theorem 1 and Theorem 5 we have a convergence to a Poisson random variable, in Theorem 2, Theorem 3 and Theorem 4 we have a convergence to a gaussian random variable.

REFERENCE

1. Kolchin V. F., Sevast'ynov B. P., Chistiakov V. P. *Random allocation.* – Scripta Series in Mathematics. V. H. Winston& Sons, Washington, DC, 1978.

2-local derivations on Virasora algebra

Sh.A.Ayupov [1], B.B.Yusupov [2]

[1] V.I.Romanovskiy Institute of Mathematics, Uzbekistan Academy of Sciences,

[2] National University of Uzbekistan

[1] *sh_ayupov@mail.ru,* [2] *baxtiyor_yusupov_93@mail.ru;*

In 1997, Šemrl [10] introduced 2-local derivations and 2-local automorphisms on algebras. A map $\Delta : \mathcal{L} \to \mathcal{L}$ (not necessarily linear) on an algebra $\mathcal{L}$ is called a *2-local derivation* if, for every $x, y \in \mathcal{L}$, there exists a derivation $D_{x,y} : \mathcal{L} \to \mathcal{L}$ such that $D_{x,y}(x) = \Delta(x)$ and $D_{x,y}(y) = \Delta(y)$. For a given algebra $\mathcal{L}$, the main problem concerning these notions is to prove that they automatically become a derivation (respectively, an automorphism) or to give examples of local and 2-local derivations or automorphisms of $\mathcal{L}$, which are not derivations or automorphisms, respectively. Solution of such problems for finite-dimensional Lie algebras over algebraically closed field of zero characteristic were obtained in [1, 2, 3] and [5]. Namely, in [3] it is proved that every 2-local derivation on a semi-simple Lie algebra $\mathcal{L}$ is a derivation and that each finite-dimensional nilpotent Lie algebra with dimension larger than two admits 2-local derivation which is not a derivation. In [1] the authors have proved that every local derivation on semi-simple Lie algebras is a derivation and gave examples of nilpotent finite-dimensional Lie algebras with local derivations which are not derivations. Concerning 2-local automorphism, Chen and Wang in [5] prove that if $\mathcal{L}$ is a simple Lie algebra of type A_l, D_l or $E_k, k = 6, 7, 8)$ over an algebraically closed field of characteristic zero, then every 2-local automorphism of $\mathcal{L}$, is an automorphism. Finally, in [2] Ayupov and Kudaybergenov generalized this result of [10] and proved that every 2-local automorphism of a finite-dimensional semi-simple Lie algebra over an algebraically closed field of characteristic zero is an automorphism. Moreover, they show also that every nilpotent Lie algebra with finite dimension larger than two admits 2-local automorphisms which is not an automorphism. In [4] authors have proved that every anti-symmetric biderivation on a simple generalized Witt algebra W is an inner derivation.

In [11] the authors studied 2-local derivations of infinite-dimensional Lie algebras over a field of characteristic zero and proved that all 2-local derivations of the Witt algebra as well as of the positive Witt algebra and the classical one-sided Witt algebra are (global) derivations.

In the present paper we study derivations and 2-local derivations on 2-local derivations on Virasora algebra.

A derivation on a Lie algebra $\mathcal{L}$ is a linear map $D : \mathcal{L} \to \mathcal{L}$ which satisfies the Leibniz law, that is,

$$D([x, y]) = [D(x), y] + [x, D(y)]$$

for all $x, y \in \mathcal{L}$. The set of all derivations of $\mathcal{L}$ which respect to the commutation operation is a Lie algebra and it is denoted by $Der(\mathcal{L})$.

Let $A = \mathbb{C}[x, x^{-1}]$ be the algebra of Laurent polynomials in one variable. The Lie algebra of derivations

$$Der(A) = span\{f(x)\frac{d}{dx} : f \in \mathbb{C}[x, x^{-1}]\}$$

is called *Witt algebra* and denoted by W. We consider an infinite-dimensional simple Witt algebra W over a field of characteristic zero. W is a Lie algebra which has a basis $\{e_i : e_i = x^{i+1}\frac{d}{dx}, i \in \mathbb{Z}\}$ and the multiplication rule [7]

$$[e_i, e_j] = (j - i)e_{i+j},\ i, j \in \mathbb{Z}.$$

Recall that a map $\Delta : \mathcal{L} \to \mathcal{L}$ (not liner in general) is called a *2-local derivation* if for every $x, y \in \mathcal{L}$, there exists a derivation $D_{x,y} : \mathcal{L} \to \mathcal{L}$ (depending on x, y) such that $\Delta(x) = D_{x,y}(x)$ and $\Delta(x) = D_{x,y}(y)$.

Definition 1. A central extension of a Lie algebra G is a Lie algebra $\mathcal{L}$ with a subalgebra U such that U is central and the quotient of $\mathcal{L}$ by U is G. Using short exact sequences, we have

$$0 \to U \to \mathcal{L} \to G \to 0.$$

Theorem 2. *The Witt algebra W has a unique nontrivial one-dimensional central extension $Vir = W \oplus \mathbb{C}\bar{c}$ up to Lie algebra isomorphism. This extension has a basis $\{c\} \cup \{e_n | n \in \mathbb{Z}\}$ where $c \in \mathbb{C}\bar{c}$, such that the following relations are satisfied:*

$$[e_i, c] = 0, \quad \text{for} \ \ i \in \mathbb{Z}$$

$$[e_i, e_j] = (i - j)e_{i+j} + \delta_{i,-j}\frac{1}{12}(i^3 - i)c \quad \text{for} \ \ i, j \in \mathbb{Z},$$

where $\delta_{i,-j}$-Kronecker symbol [9].

The extended Lie algebra Vir is called the Virasoro algebra.

Theorem 3. *Let D be a derivation on Vir. Then exists an element $a \in Vir$ such that*

$$D = L_a.$$

Theorem 4. *Let Vir be the Virasora algebra over a field of characteristic zero. Then any 2-local derivation on Vir is a derivation.*

REFERENCE

1. Ayupov Sh.A., Kudaybergenov K.K., Local derivations on finite dimensional Lie algebras, Linear Algebra and its Applications, **493**, 381–398 (2016).
2. Ayupov Sh.A., Kudaybergenov K.K., Local automarphisms on finite dimensional Lie algebras, Linear Algebra and its Applications, **507** , 121-131 (2016).
3. Ayupov Sh.A., Kudaybergenov K.K., Rakhimov I.S., 2-Local derivations on finite-dimensional Lie algebras, Linear Algebra and its Applications, **474**, 1-11 (2015).
4. Chen Z., Biderivations and linear commuting maps on simple generalized Witt algebras over a field, Electronic Journal of Linear Algebra, Volume **31**, 1–12 (2016)
5. Chen Z., Wang D., 2-Local automorphisms of finite-dimensional simple Lie algebras, Linear Algebra and its Applications, **486**, 335–344 (2015).
6. Goze M., Khakimdjanov Yu., Nilpotent Lie algebras, Kluwer Academic Pulisher, Dordrecht, 1996.
7. Kac V., Raina A., Bombay lectures on highest weight representations of infinite-dimensional Lie algebras, World Sci.Singapore, 1987.
8. Khakimdjanova K., Khakimdjanov Yu., Sur une classe d'algebres de Lie de dimension infinie, Comm. Algebra, Vol.**29**(1), 177–191 (2001).
9. Patera J., Zassenhaus H., The Higher Rank Virasoro Algebras, Commun. Math. Phys. **136**, 1-14 (1991).
10. Šemrl P., Local automorphisms and derivations on $B(H)$, Proc. Amer. Math. Soc., **125**, 2677–2680 (1997).
11. Ayupov Sh.A., Yusupov B.B., 2-Local derivations of infinite-dimensional Lie algebras, Journal of Algebra and Its Applications, DOI: 10.1142/S0219498820501005.

Ayupov Sh.A. (81, Mirzo Ulughbek street, 100170 , Tashkent, Uzbekistan), Yusupov B.B. (4, University street, 100174, Tashkent, Uzbekistan)

2-local derivations on Virasora algebra

Abstract. *The present paper is devoted to study 2-local derivations on infinite-dimensional Lie algebras over a field of characteristic zero. We prove that all 2-local derivations on the Virasoro algebra is a derivation.*

Key words: *Lie algebras, Witt algebra, Virasoro algebra, derivation, 2-local derivation.*

Software, algorithms and methods of data encryption based on national standards

R.D.Aloev [1], M.M.Nurullaev[2] A.A.Abdiraximov[1]

[1] National University of Uzbekistan named after Mirzo Ulugbek

[2] Bukhara Engineering Technological Institute

aloevr@mail.ru, mirxon@mail.ru

Abstract The article provides a brief description of the software of the algorithms and methods of data encryption based on national standards developed by the authors, which is designed to create encryption keys, private and public keys of electronic digital signature, create and confirm authenticity of digital signature, hashing, encrypting and simulating data using the algorithms described in State Standards of Uzbekistan. It can be used in telecommunications networks, public information systems, government corporate information systems by embedding into application applications that store, process and transmit information that does not contain information related to state secrets, as well as in the exchange of information and ensuring the legal significance of electronic documents. The software includes the following functional components: a dynamically loadable library that implements a biophysical random number sensor; dynamic library that implements cryptographic algorithms in accordance with the State Standards of Uzbekistan; module supporting the work with external devices; installation module that provides the installation of cryptography service provider in the appropriate environment of operation (environment).

Keywords: tools of cryptographic protection of information, data encryption algorithm, hash function, encryption key.

1. Introduction The *software* provides the creation of private and public electronic digital signature keys and encryption keys; creation and confirmation of authenticity of electronic digital signature according to the algorithms described in [2, 7]; the formation of derived encryption keys used by data encryption algorithms described in [4, 5]; work with key information stored on external media; hashing of memory areas and other data according to the algorithms described in [3, 6]; encryption of memory areas and other data in accordance with the data encryption algorithms described in [4, 5].

The cryptography service provider provides support for identifiers of algorithms and parameters for compatibility with third-party crypto-providers in terms of the ability to work with public-key certificates issued by third-party

registration centers, provided they use the cryptographic algorithms described in [2, 3, 4, 5, 6, 7]. The cryptography service provider provides the ability to work with digital certificates of public keys, which are structured binary in ASN.1 format, conforming to ITU-T X.509 v.3 standard and IETF RFC 5280, RFC 3739 Recommendations. The cryptography service provider provides work with external key carriers such as USB-flash, eToken Aladdin (eToken PRO 72K (JAVA)). As part of CIPF - CSP, modules are provided that provide for calling cryptographic functions through the Microsoft CryptoAPI 2.0 interface when running under Microsoft operating systems.

In accordance with the functional purpose of a cryptography service provider, it generates public and private electronic digital signature keys, hash keys, functional keys and encryption keys for use, respectively, in the algorithms described in [2, 3, 4, 5, 6, 7]. In applications where the cryptography service provider will be integrated, an appropriate key manufacturing and distribution system is provided, which will be based on the key generation functions of the cryptography service provider. Cryptography service provider allows the use of a multi-level key protection model using random and derivative keys of key protection. Protection of keys is carried out on the basis of cryptographic transformations in accordance with the PKCS 5 standard [9] using the State Standards of Uzbekistan [3], [4] or the interstate standard [5]. A cryptography service provider provides storage of key information on a hard magnetic disk drive (HDD) and/or external key storage devices, such as USB-flash, eToken Aladdin (eToken PRO 72K (JAVA)). A cryptography service provider provides work with key containers containing: signature keys, encryption keys and additional information necessary to ensure the cryptographic protection of keys and ensure control of their integrity. To protect key information from substitution and/or distortion during its storage on HDD and external key carriers, as well as during distribution, key information is supplied with a checksum. In order to ensure the safe use of the cryptography service provider installed on a PC, organizational measures are provided, as well as software and hardware methods and means of protecting information are used to ensure the secrecy of secret keys located in the PC's memory during the operation of cryptography service provider, as well as service cryptography service provider parameters stored on the hard disk. Cryptography service provider contains a component that allows you to verify the operability of the cryptographic algorithms implemented in it. Functioning is carried out on the basis of test examples. To ensure the safe use of an application with a built-in ICS-CSP, mechanisms are provided to control the integrity of cryptography

service provider libraries. In cryptography service provider, a biophysical sensor of random numbers is used to generate random binary sequences, which implements the mechanism for generating secret digital signature keys, encryption keys, initialization vectors using various algorithms.

The paper briefly describes the software necessary for the operation of the tools of cryptographic protection of information (TCPI) - data encryption cryptography service provider which is developed by the authors based on the national standards. Furthermore, here is given the purpose, the capabilities of the TCPI - CSP software and its main characteristics, the limitation of the using area of this software. One of the following operating systems is necessary in order to perform the TCPI - CSP: Microsoft Windows XP (32 bit) Professional SP3; Microsoft Windows Vista (32 bit) Ultimate SP2; Microsoft Windows 7 (32 bit) Ultimate; Microsoft Windows Server 2003 (32 bit) Enterprise Edition R2 SP2; Microsoft Windows Server 2008 (32 bit) Enterprise Edition SP2. The original programming languages for the TCPI - CSP are C and C ++. TCPI - CSP performs the following main functions:

- generation of encryption keys for the data encryption algorithm O‘z DSt 1105: 2009 [1] and the algorithm GOST 28147-89 [6];
- encryption of RAM areas and other data in accordance with the data encryption algorithm O‘z DSt 1105: 2009 [1] and the algorithm GOST 28147-89 [6];
- generation of keys for implementing and verifying the electronic digital signature (EDS) using algorithms 1 and 2 of O‘z DSt 1092: 2009 [3] and the algorithm of GOST R 34.10-2001 [8];
- hashing of memory areas and other data according to algorithm 1 with the parameter p = 256 O‘z DSt 1106: 2009 [2] and the algorithm GOST 34.11-94 [7];
- formation and verification of the EDS result in accordance with O‘z DSt 1092: 2009 [3] algorithms 1 and 2 and GOST R 34.10-2001 [8];

work with key information stored on external media. TCPI - CSP can be used as a default crypto-provider for the Windows operating system. TCPI - CSP supports the cryptographic algorithms of the Republic of Uzbekistan and Russia as well as some of the common cryptographic algorithms used in Windows OS, such as RSA, 3DES, SHA-1 etc.

2. Encryption function

The encryption function is a cryptographic algorithm, which is a bijective mapping from a finite set of plaintext to a finite set of encrypted texts, in which the mapping function depends on a secret parameter called a key.

The encryption function uses to encrypt and decrypt information.

The encryption function in accordance with [1] can use cryptographic keys of length **256** or **512** bits for encrypting and decrypting data blocks of length **256** bits.

The encryption function is used for cryptographic protection of data storing and transmitting in computer networks, telecommunications, in separate computing systems or in computers of enterprises, organizations, and institutions.

In symmetric cryptosystems, data exchange takes place in three stages:

the sender of the message sends the encryption key (or/and functional key) to the recipient via a secure channel that is known only to them;

the sender, using the encryption key and the function key, converts the original data into encrypted data and sends them to the recipient via the communication channel;

the recipient, having received the encrypted data, decrypts it with the help of an encryption key and a function key. Both sides may use these keys several times.

In addition to data protection, the encryption function can be used to protect the symmetric keys themselves as they are transmitted over unprotected communication channels. In this case, the transmitted symmetric key is encrypted with some other key, called the security key.

The encryption function [1] contains two modes: electronic codebook mode (ECM);

block chaining mode (BCM)

The electronic codebook mode is an encryption mode in which all plaintext blocks are encrypted independently of one another on the same key, in accordance with the data encryption algorithm.

ECM mode is usually used when encrypting symmetric keys.

T**he block chaining mode** of encryption is a mode in which each encrypted (decrypted) depends on the previous block of an encrypted (decrypted) block. For the first block, the initialization vector is used as the previous block. If the last block of text is not complete, it is supplemented to the required length. This procedure is called padding. BCM mode is usually used when encrypting data.

The purpose of these functions, functionality of operation algorithms are presented in the document "O‘z DSt 1105: 2009. State standard of Uzbekistan. Information technology. CRYPTOGRAPHIC PROTECTION OF INFORMATION. Data encryption algorithm".

3. Signature function

The electronic digital signature function is used to generate and confirm the authenticity of an electronic digital signature (EDS) under a given message (electronic document) transmitted over unprotected public telecommunications channels. Upon receipt of the message, the recipient can verify the integrity of the message transmitted by the sender and verify the authenticity of the sender's authorship. EDS is an electronic analog of a written signature and therefore an EDS can be used by the recipient or a third party to verify that the message was actually signed by the sender. To describe the formation and confirmation of the authenticity of a digital signature, two algorithms are used (Algorithm 1, Algorithm 2). Algorithm 1 is considered in two basic modes:

– without session key;

– with a session key.

Algorithm 2 is used in the classical (without session key) mode. Algorithm 1 provides a backup path for detecting a fake digital signature by introducing a session key procedure used in the process of authenticating the authenticity of a digital signature to the EDS generation process. The composition and purpose of this function, functionality, and functioning algorithm are presented in the document "O‘z DSt 1092: 2009. State standard of Uzbekistan. Information technology. CRYPTOGRAPHIC PROTECTION OF INFORMATION. Processes of formation and verification of electronic digital signature".

4. Functional restrictions on the use of TCPI - CSP

Cryptographic interface TCPI - CSP is implemented in accordance with the CSP standard, which is applicable only in the Windows operating system and is not applicable in other operating systems such as Linux, Mac, Unix, etc. Since cryptographic algorithms of the Republic of Uzbekistan are implemented in TCPI - CSP, which are not recognized by standard Windows tools, for embedding TCPI in typical Windows applications (MS Outlook, Internet Explorer, VPN, etc.) requires changes to the standard OS software (advapi32.dll, cryptsp.dll, crypt32.dll, inetcomm.dll, schannel.dll, secur32.dll, mailcomm.dll, etc.).

5. The used algorithms and methods

When implementing cryptographic algorithms of the Republic of Uzbekistan, including the implementation of operations with big numbers, the source texts of programs from the OpenSSL the software package was used. When implementing cryptographic algorithms of the Republic of Uzbekistan, all simple numbers that are part of the parameters of the algorithms are checked for simplicity using the standard procedures contained in the package OpenSSL.

Encryption of symmetric keys using symmetric keys

To encrypt symmetric keys using symmetric keys, an DEA is used in ECM mode. To produce imitation protection, a hash function from cryptographic standards of Uzbekistan in 256 bit mode is used. The first 4 bytes of the key's hash function is the simulated prefix for the encrypted key.

Encryption of symmetric keys using asymmetric keys

When encrypting symmetric keys using asymmetric keys, the following algorithm is used. Using the Diffie-Hellman method, a common key of the form axy is formed, where x, y are private keys, the common key size is 256 and 64 bytes for algorithms 1 and 2, respectively. This shared key is used instead of a password to generate a shared symmetric key and initialization vector in accordance with PKCS#5. The original symmetric key is encrypted in BCM mode using a common symmetric key and an initialization vector. The size of the encrypted key is 64 bytes. Imitation protection provides padding size of 32 bytes.

An algorithm of generation and verification of EDS

Processes of formation and verification of electronic digital signatures in TCPI - CSP supports both cryptographic algorithms of the Republic of Uzbekistan [3] and the cryptographic standard of the Russian Federation (GOST R 34.10-2001) [8]. The function of EDS (formation and verification) is designed to ensure the reliability of the transmitted and received information and confirm its authorship.

6. Conclusion

So, the work is devoted to a brief description of the CSP software, which is designed to create encryption keys, private and public keys of an electronic digital signature, creating and authenticating EDS, hashing, encrypting and simulating data using the algorithms described in the State Standards of Uzbekistan. It can be used in telecommunications networks, public information systems, government corporate information systems by embedding into application applications that store, process and transmit information that does not contain information related to state secrets, as well as in the exchange of information and ensuring the legal significance of electronic documents. CSP

includes the following functional components: a dynamically loadable library that implements a biophysical sensor of random numbers; dynamic library that implements cryptographic algorithms in accordance with the State Standards of Uzbekistan; module supporting the work with external devices; installation module that provides the installation of CSP in the appropriate environment of operation (environment).

CSP provides the creation of private and public EDS keys and encryption keys; creation and confirmation of authenticity of EDS according to the algorithms described in [2, 7]; the formation of derived encryption keys used by data encryption algorithms described in [4, 5]; work with key information stored on external media; hashing of memory areas and other data according to the algorithms described in [3, 6]; encryption of memory areas and other data in accordance with the data encryption algorithms described in [4, 5].

CSP provides support for identifiers of algorithms and parameters for the implementation of compatibility with third-party cryptographic providers in terms of the ability to work with public-key certificates issued by third-party registration centers, provided they use the cryptographic algorithms described in [2, 3, 4, 5, 6, 7]. The cryptography service provider provides the ability to work with digital certificates of public keys, which are structured binary in ASN.1 format, conforming to ITU-T X.509 v.3 standard and IETF RFC 5280, RFC 3739 Recommendations. CSP provides work with external key carriers such as USB-flash, eToken Aladdin (eToken PRO 72K (JAVA)). As part of CSP, modules are provided that provide for calling cryptographic functions through the Microsoft CryptoAPI 2.0 interface when running under Microsoft operating systems. In accordance with the functional purpose, CSP provides the generation of public and private EDS keys, hashing keys, functional keys and encryption keys for use, respectively, in the algorithms described in [2, 3, 4, 5, 6, 7]. In the application where the CSP will be integrated, an appropriate system for the manufacture and distribution of keys will be provided, which will be based on the functions of generation of CSP keys. CSP allows the use of a multi-level key protection model using random and derivative keys of key protection. Protection of keys is carried out on the basis of cryptographic transformations in accordance with the PKCS #5 standard [9] using the State Standards of Uzbekistan [3], [4] or the interstate standard [5]. CSP provides storage of key information on a hard disk drive (HDD) and/or external key storage devices, such as USB-flash, eToken Aladdin (eToken PRO 72K (JAVA)). CSP provides work with key containers containing: signature keys, encryption keys and additional information necessary to ensure the cryptographic protection

of keys and ensure control of their integrity. To protect key information from substitution and/or distortion during its storage on HDD and external key carriers, as well as during distribution, key information is supplied with a checksum. In order to ensure the safe use of the CSP installed on the PC, organizational measures are provided, as well as software and hardware methods and means of protecting information are used to ensure that secret keys stored in the PC's memory during operation of the CSP are kept secret, as well as the CSP service parameters stored on the hard drive. CSP contains a component that allows you to verify the operation of cryptographic algorithms implemented in it. Functioning is carried out on the basis of test examples. To ensure the safe use of an application with a built-in CSP, mechanisms are provided to control the integrity of the CSP libraries. In CSP, a biophysical random number sensor is used to generate random binary sequences that implement the mechanism for generating secret digital signature keys, encryption keys, initialization vectors using various algorithms.

References

1. O‘z DSt 1105: 2009. State standard of Uzbekistan. Information technology CRYPTOGRAPHIC PROTECTION OF INFORMATION. Data encryption algorithm.

2. O‘z DSt 1106: 2009. State standard of Uzbekistan. Information technology CRYPTOGRAPHIC PROTECTION OF INFORMATION. A Hash function.

3. O‘z DSt 1092: 2009. State standard of Uzbekistan. Information technology CRYPTOGRAPHIC PROTECTION OF INFORMATION. Processes of formation and verification of electronic digital signature.

4. PKCS #5 v2.0: Password-Based Cryptography Standard. RSA Laboratories. March 25, 1999. [Electronic resource]. - Access mode: http://www.rsa.com/rsalabs/node.asp?id=2127.

5. Expansion of PKCS #11 for the use of Russian cryptographic algorithms. Moscow, 2008.

6. GOST 28147-89 - Information processing systems. Cryptographic protection. An algorithm of cryptographic transformation.

7. GOST R 34.11-94 - Information technology. Cryptographic protection of information. A Hash function.

8. GOST R 34.10-2001 - Information technology. Cryptographic protection of information. Processes of formation and verification of electronic digital signature.

9. RFC 4357 Additional Cryptographic Algorithms for Use with GOST 28147-89, GOST R 34.10-94, GOST R 34.10-2001, and GOST R 34.11-94 Algorithms.

HITTING TIMES AND SINGULAR DISTRIBUTIONS OF CIRCLE MAPS WITH A BREAK

A.Dzhalilov [1], J.Karimov [2]

[1] Turin Polytechnic University in Tashkent

[2] National University of Uzbekistan named after Mirzo Ulugbek, Turin Polytechnic University in Tashkent

adzhalilov21@gmail.com, jkarimov0702@gmail.com

In this paper we study the limit behaviour of distribution function of hitting times with respect to Lebesgue measure for circle homeomorphism with a single break point. Consider orientation preserving circle homeomorphism f with "golden"irrational rotation number $\rho = \rho(f) = \frac{\sqrt{5}-1}{2}$ i.e. the continued fraction expansion of ρ is $\rho = [1, 1, \dots, 1, \dots]$. Set $\frac{p_n}{q_n} = [1, 1, \dots, 1]$, $n \geq 1$. The numbers q_n, $n \geq 1$ satisfies difference equation $q_{n+1} = q_n + q_{n-1}$, $q_0 = 1$, $q_1 = 1$. Take an arbitrary point x_0 of the unit circle $S^1 = \mathbb{R}^1/\mathbb{Z}^1 \simeq [0, 1)$. We denote $x_i = f^i(x_0)$, $i \geq 1$, where f^n denotes the $n-$th iteration of f. Let $\Delta_0^{(n)}(x_0)$ be a closed segment connecting the points x_0 and x_{q_n}. Set $\Delta_i^{(n)}(x_0) = f^i(\Delta_0^{(n)}(x_0))$, $i \geq 1$. Fix $c \in (0, 1)$. For every $n \geq 1$ uniquely define the points c_n by relation $|[x_0, c_n]| = c \cdot |[x_0, x_{q_n}]|$, where $|\cdot|$ denotes the length of the interval. We denote by $I_{n,c}$ interval $[x_0, c_n)$ with $x_0 = 0$ (see Picture 1).



1-picture

Define the first return time function $R_{n,c} : I_{n,c} \to \mathbb{N}$:

$$R_{n,c}(x) = \min\{j \geq 1 : T^j x \in I_{n,c}\}.$$

Using the structure of dynamical partitions (see [1], [2]) we get:

$$R_{n,c}(x) = \begin{cases} q_{n+1}\ ,\ x \in [x_{-q_{n+1}}, c_n] \\ q_{n+2}\ ,\ x \in [x_0, f^{-q_{n+2}} c_n] \\ q_{n+3}\ ,\ x \in [f^{-q_{n+2}} c_n, x_{-q_{n+1}}] \end{cases} \tag{1}$$

Now we define the hitting time function $E_{n,c} : S^1 \to N$:

$$E_{n,c}(x) = \min\{j \geq 1 : T^j x \in I_{n,c}\},\ \ x \in S^1.$$

Using (1) we obtain that $E_{n,c}(x)$ get values from 1 to q_{n+3}. Now we define the normalized hitting time function $\overline{E}_{n,c}(x)$:

$$\overline{E}_{n,c}(x) = \frac{1}{q_{n+3}} E_{n,c}(x).$$

However on S^1 we have two natural probability measures: Lebesgue measure ℓ and probability invariant measure μ. We consider the following distribution functions by measures μ and ℓ :

$$F_n(t) = \mu\left(x \in S^1 : \overline{E}(x) \leq t\right), \ \ \forall t \in \mathbb{R}^1;$$

$$\Phi_n(t) = \ell\left(x \in S^1 : \overline{E}(x) \leq t\right), \ \ \forall t \in \mathbb{R}^1.$$

E. de Faria and Z. Coelho in [3] established that limit behaviour of the sequence $\{F_n(t), \ \ n = 1, 2, ...\}$, depends on rotation number ρ, namely either it is uniform distribution, or continuous piecewise linear distribution on $[0, 1]$. For each $k > 1$ the corresponding sequence of distribution functions $\{F_{n_s}(t), \ \ s = 1, 2, ...\}$ converges to either distribution for random variable of $\xi \equiv 1$, or step function with two discontinuous points.

For linear rotation by irrational angle Lebesgue measure is unique probability invariant measure. A.Dzhalilov in [4] studied the limit behaviour of distribution function with respect to Lebesgue measure for critical circle maps with "golden mean"irrational rotation number.

Obviously, that function $\overline{E}_{n,c}(x)$ will be a random variable taking values in $[0, 1]$. We denote by $\Phi_{n,c}(t)$ distribution function of $\overline{E}_{n,c}(x)$.

Lemma 1. *The distribution function of the normalized entrance function* $\overline{E}_{n,c}(x)$ *has the following form:*

$$\Phi_{n,c}(t) = \begin{cases} 0, \text{ if } t \le 0 \\ \sum\limits_{i=q_{n+1}-m}^{q_{n+1}-1} |A_i^{(n)}| + \sum\limits_{j=q_{n+2}-m}^{q_{n+2}-1} |B_j^{(n)}| + \sum\limits_{k=q_{n+3}-m}^{q_{n+3}-1} |C_k^{(n)}|, \\ \text{if } mq_{n+3}^{-1} \le t \le (m+1)q_{n+3}^{-1},\ 1 \le m \le q_{n+1} \\ \sum\limits_{i=0}^{q_{n+1}-1} |A_i^{(n)}| + \sum\limits_{j=q_{n+2}-m}^{q_{n+2}-1} |B_j^{(n)}| + \sum\limits_{k=q_{n+3}-m}^{q_{n+3}-1} |C_k^{(n)}|, \\ \text{if } mq_{n+3}^{-1} \le t \le (m+1)q_{n+3}^{-1},\ q_{n+1} \le m \le q_{n+2} \\ \sum\limits_{i=0}^{q_{n+1}-1} |A_i^{(n)}| + \sum\limits_{j=0}^{q_{n+2}-1} |B_j^{(n)}| + \sum\limits_{k=q_{n+3}-m}^{q_{n+3}-1} |C_k^{(n)}|, \\ \text{if } mq_{n+3}^{-1} \le t \le (m+1)q_{n+3}^{-1},\ q_{n+2} \le m \le q_{n+3} \\ 1, \text{ if } t \ge 1. \end{cases}$$

Now we formulate the main result of our work.

Theorem 1. *Let* $f \in C^{2+\varepsilon}(S^1 \setminus \{x_0\})$ *be circle homeomorphism with a single break point* x_0 *and "golden mean"irrational rotation number. Let* $c \in [\rho, 1)$ *and* $\left\{\Phi_{n,c}(t)\right\}_{n=1}^{\infty}$ *be the sequence of distribution functions with respect to Lebesgue*

measure on the circle, corresponding to the first normalized hitting time on $I_{n,c}$. *Then*

1) for all $t \in R^1$ *there exist the finite limit*

$$\lim_{n\to\infty} \Phi_{n,c}(t) = \Phi_c(t),$$

and where $\Phi_c(t) = 0$, *if* $t \le 0$, *and* $\Phi_c(t) = 1$, *if* $t > 1$;

2) $\Phi(t)$ *is a strictly increasing on* $[0,1]$ *and continuous distribution function on* R^1 .

3) $\Phi(t)$ *is singular on* $[0,1]$, *i.e.* $\frac{d\Phi^{(1)}(t)}{dt} = 0$ *a.e.*

REFERENCE

1. Ya.G.Sinai, Topics in Ergodic Theory, Princeton Mathematical Series, 44, Princeton University Press, Princeton, NJ. 1994.

2. D.H.Kim, B.K.Seo, The waiting time for or rational rotations, Nonlinearity 16, 2003, p. 1861-1868.

3. Z. Coelho, E. de Faria, Limit laws of entrance times for homeomorphisms, Israel J. Math, 93, 1996, p. 93-112.

4. A. A. Dzhalilov, Limiting Laws for Entrance Times of Critical Mappings of a Circle, Theoret. and Math. Phys, 138:2, 2004, p. 190-207.

Akhtam Dzhalilov (Tashkent, Uzbekistan), Javlon Karimov (Tashkent, Uzbekistan)

Hitting times and singular distributions of circle maps with a break

Abstract. *In present work we study the limit behaviour of hitting times of circle homeomorphisms with a single break a point.*

Key words: *circle homeomorphism, hitting time, break point, dynamical partition.*

Исследование адаптивного алгоритма классификации объектов по данным с пропусками

А.В.Медведев[1], **Д.А.Мелех**[2], **Н.А.Сергеева**[2], **О.В.Чубарова**[1]
[1] Институт космических и информационных технологий
Сибирский федеральный университет,
[2] Институт математики и фундаментальной информатики
Сибирский федеральный университет
sergena@outlook.com

Серьезным вопросом при проектировании реальных систем поддержки принятия решений является вопрос обработки поступающих для обработки данных, которые содержат пропуски. Причины пропусков могут быть разными и это не всегда техническая ошибка или влияние случайных возмущений на систему телеметрии. Иногда значение признака может быть объективно неизвестным. Примером подобной ситуации могут быть результаты ультразвукового обследования органов человека, когда невозможно диагностировать значение того или иного признака из-за состояния самого органа. В этом случае значение признака останется незаполненным, и такой объект необходимо обработать математическим алгоритмами для каких-либо диагностических целей. Проблема заполнения пропущенных значений ставится в задачах предобработки данных давно и имеет множество вариантов решения, которые предлагают заполнение пропущенных значений разными методами [1, 3].

В настоящей работе предлагается альтернативный подход, который предполагает подойти к решению проблемы пропусков со стороны перестроения алгоритмов обработки данных, оставляя измерения признаков как есть, то есть сохраняя пропущенные значения незаполненными. Реагировать на наличие пропусков будут сами алгоритмы через предлагаемые правила их адаптивной перестройки. Исследование подхода адаптивной перестройки алгоритмов предлагается для задачи классификации объектов, измерения признаков которых содержат пропуски [2, 3]. Для построения адаптивных алгоритмов классификации применяется модификация многомерной непараметрической оценки вероятности Розенблатта-Парзена [4]

Обозначим множество объектов $\{O_i, i = 1, \ldots, s\}$, где s – объем выборки с известным набором признаков $\{p_i, i = 1, \ldots, n\}$, измеренных в абсолютной шкале: n_1, и ранговой: n_2, $n_1 + n_2 = n$. Известен класс каждого объекта $O_i \in Z_l, l = \overline{1, L}\}$, L – число классов. Измерения признаков для каждого объекта являются набором значений $\{\left(z_i, x_i^j\right), i = \overline{1, s}, j = \overline{1, n}\}$, где x_i^j

– значение p_j-го признака у O_i-го объекта, z_i – номер класса, s_l – число объектов класса Z_l, $\sum_{l=1}^{L} s_l = s$, предполагается, что измерения признаков объекта содержат пропуски.

Необходимо построить алгоритм классификации, оперирующий данными с пропусками без этапа их заполнения, разработать процедуры настройки параметров.

Для оценивания степени принадлежности классу объекта O_t используется только набор тех признаков, у которых имеются исходные значения p. Для этого на каждом t-ом такте работы алгоритма пересортировывается вся обучающая выборка относительно имеющегося набора признаков p у O_t, предъявленного для классификации. Вследствие этого изменится объем обучающей выборки: $\sum_{l=1}^{L} s_l^t = s^t \leq s$, где s_l^t – объем выборки в классе Z_l, s^t – общий объем выборки после отбора непустых признаков. Изменение набора признаков для оценивания схожести с объектами каждого класса повлечет изменение числа абсолютных и ранговых признаков n_1^t и n_2^t, соответсвенно.

Алгоритм классификации состоит из трех основных этапов.

Шаг 1:

Для каждого класса введем следующую величину:

$$\alpha_l^t = \sum_{i=1}^{s_l^t} \prod_{i_1=1}^{n_1^t} \Phi\left(\frac{x_t^{j_1} - x_i^{j_1}}{C^{j_1}}\right) \prod_{j_2=1}^{n_2^t} 1(x_t^{j_2}, x_i^{j_2}), \quad l = \overline{1, L}, \tag{1}$$

конструктивно повторяющую многомерную непараметрическую оценку плотности вероятности Розенблатта-Парзена.

Шаг 2:

Расчет нормированных оценок β_l^t на основе вычисленных на шаге 1 α_l^t:

$$\beta_l^t = \alpha_l^t \Big/ \sum_{l=1}^{L} \alpha_l^t. \tag{2}$$

Шаг 3: Решающее правило: чем вероятнее принадлежность объекта к классу, тем ближе к 1 его значение относительной оценки принадлежности:

$$O_t \in Z_k | \quad \beta_k^t = \max_{l=\overline{1,L}} (\beta_l^t). \tag{3}$$

В качестве колоколообразной функции $\Phi(\cdot)$ в (1) могут быть использованы треугольное ядро, усеченная парабола, гауссово ядро, $\cos(\cdot)$, функция

Соболева и другие для признаков в абсолютной шкале. Для ранговых признаков используется индикатор равенства типа Кронекера. Параметр размытости представляет собой вектор (по числу количественных признаков). Оптимизационная процедура настройки параметров алгоритма проводится по числу точек k под колоколообразной функцией.

Для каждого нового объекта Q_t, предъявленного для классификации, объем обучающей выборки для вычисления соотношения (1) будет меняться из-за различного набора незаполненных признаков. Таким образом, будет меняться число слагаемых для расчета внутри одного класса объектов. Этот факт отражает адаптивный характер вычислительной процедуры, которая для каждого вновь поступающего на классификацию объекта использует заново сформированную обучающую выборку из исходной [5].

В качестве критерия качества классификатора используется оценка площади под ROC-кривой (AUC).

В рамках численного эксперимента проводилось исследование влияния объема пропусков на качество классификации. Численный эксперимент заключался в расчѣте разброса точности классификации в зависимости от доли пропусков в классе и по всему признаку. Доля пропусков по признаку характеризует объем пропусков, сгенерированный случайным образом в каждом признаке по всем классам, а доля пропусков в классе накладывает требование на объем пропусков по каждому классу отдельно. В рамках данного эксперимента эти показатели были сведены к одному общему параметру Ц доля пропусков.

Без нарушения общности и простоты интерпретации результатов на начальной стадии численных экспериментов была сформирована выборка с двумя признаками в абсолютной шкале, двумя ранговыми и двумя равномощными классами, которые в исходном признаковом пространстве не пересекаются. В качестве показателя точности рассматривалась метрика AUC, объѣм выборки $s = 500$ объектов. По условиям эксперимента пропуски допускались только в ранговых признаках (как правило именно этот тип признаков сложно заполнить при традиционном подходе).

Основной процесс данного эксперимента заключается в проведении k итераций расчѣта точности классификации при условии, что выбранная доля пропусков постоянна, но при этом с каждой итерацией пропуски заново случайно перераспределяются в выборке. Данный процесс повторяется для каждого выбранного показателя доли смеси r_{mix} между пропусками: в результате генерации значения этого показателя пропуски у ранговых признаков объекта могли отсутствовать, мог быть пропускв одном признаке

или в обоих.

Для генерации числа пропусков в классе было взято долевое соотношение от 0.1 до 0.5 с шагом дискретизации 0.1. С увеличением доли пропусков и при увеличении доли смеси между самими пропусками пропорционально снижается точность классификации, а также увеличивается разброс (рисунок 1, 2, 3).

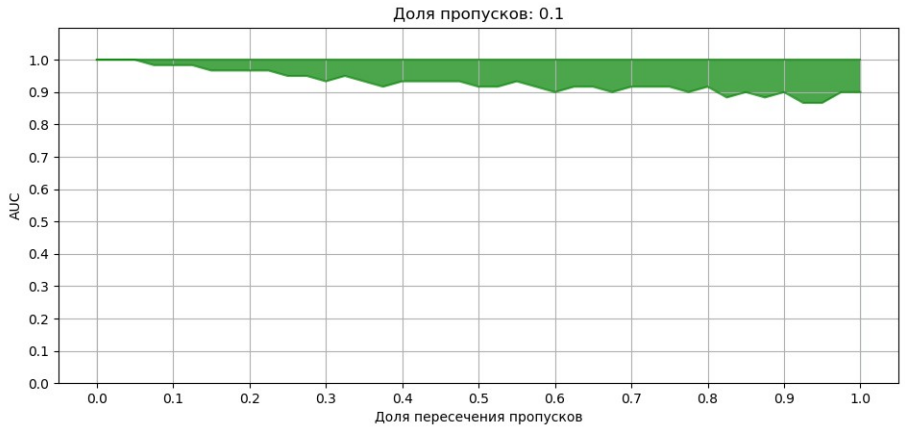


Рис. 2: Разброс точности при изменении доли смеси пропусков, доля пропусков 0.1

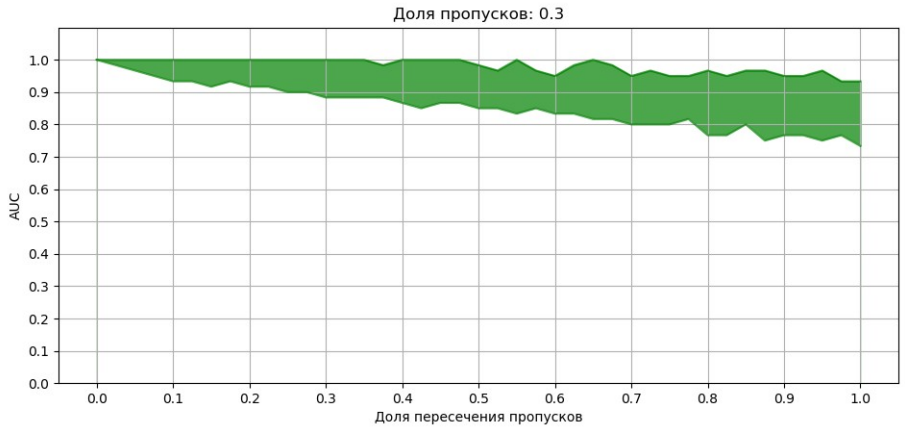


Рис. 3: Разброс точности при изменении доли смеси пропусков, доля пропусков 0.3

Доля пересечения пропусков показывает долю объектов, среди которых было пропущенны значения у обоих ранговых признаков, то есть склассификация свелась только к пространству абсолютных признаков. Пропуски генерировались у всего датасета, как у объектов выделенных в тестовую часть, так и у объектов обучающей выборки.

Выводы. Проведено численное исследование адаптивного алгоритма классификации объектов при наличии пропусков в данных без этапа их

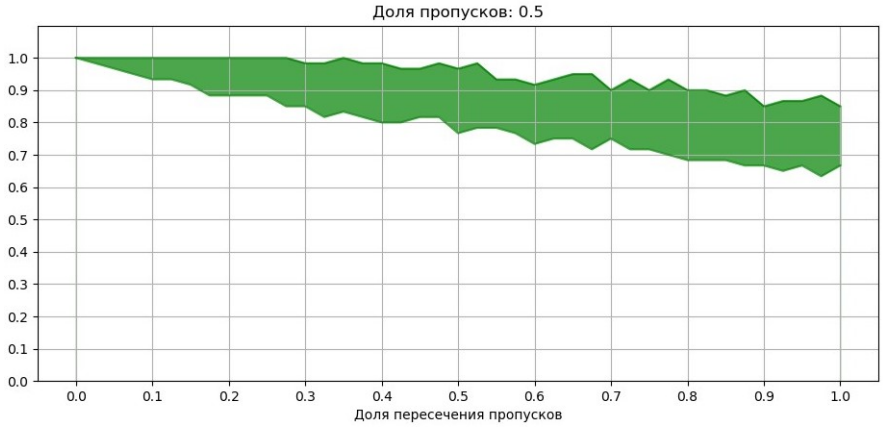


Рис. 4: Разброс точности при изменении доли смеси пропусков, доля пропусков 0.5

заполнения. Численное исследование проведено на имитационных данных, когда классы не пересекались. При этом пропуски генерировались только на ранговых признаках. Свыше 30% пропусков в точность классфиикации значительно падает, что накладывает ограничения на применение алгоритма на практике. В дальнейшем изучение особенностей поведения алгоритма требует сравнения с традиционными подходами.

ЛИТЕРАТУРА

1. Garc?a-Laencina P. J. Pattern classification with missing data: a review / P. J. Garc?a-Laencina, L. Sancho-Gomez, A. R. Figueiras-Vidal // Neural Comput & Applic. Ч 2009. Ч DOI 10.1007/s00521-009-0295-6, Springer-Verlag.

2. Li Y. Y. Classification with missing data in a wireless sensor network / Y. Y. Li, L. E. Parker // IEEE SoutheastCon 3-6 April Ч 2008. Ч P. 533 Ч 538.

3. Zhang S. Missing data imputation by utilizing information within incomplete instances / S. Zhang, Z. Jin, X. Zhu // Journal of Systems and Software, Volume 84, Issue 3, March. Ч 2011. Ч P. 452-459.

4. Hardle W. Applied Nonparametric Methods / W. Hardle, O. Linton // Cowles foundation for research in economics at Yale university, New Haven, Connecticut, 1994. p. 44

5. Медведев А. В. О задачеи классификации объектов по данным с пропусками / А. В. Медведев, Д. А. Мелех, Н. А. Сергеева, О. В. Чубарова // Информационные технологии и математическое моделирование

(ИТММ-2019): Материалы XVIII Международной конфреенции им. А. Ф.Терпугова (26-30 июня 2019 г.) – Томск: Изд-во НТЛ, 2019 – Ч. 1 – 430 с.

Medvedev Alexander (Krasnoyarsk, Russia), Melekh Daniil (Krasnoyarsk, Russia), Sergeeva Natalya (Krasnoyarsk, Russia), Chubarova Olesya (Krasnoyarsk, Russia)

Research of adaptive algorithm for classification with missing data objects

Abstract. *The article proposes a study of the adaptive adjustment algorithm approach for the classification of objects with missing data measurements.*

Key words: *classification, missing data, non-parametric estimation, adaptive algorithm.*

Об управлении процессом, описывающим двумерным нормальным распределением при помощи контрольных карт

С.А.Ахмедов, К.С. Аблазова
Андижанский Государственный Университет имени З.М. Бабура, Андижан, Узбекистан
s.a.axmedov@mail.ru, k.s.ablazova@gmail.com

Пусть производственный процесс описываются нормальной системой случайных величин: $(X,\, Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy})$. Мы говорим, что процесс действует в статистически управляемом состоянии, если источником изменчивости являются только обычные причины. В этом случае с течением времени эмпирические распределения описываются предсказуемым распределением. Если имеют место особые причины вариаций, то результат процесса не является стабильном во времени. На практике мало пользы от предсказаний, основанных на данных, собираемых от процесса, который не стабилен и непредсказуем во времени. Поэтому возникает проблема «держать» производственный процесс статистически управляемом состоянии в течении времени производственных операций.

Имеются различные методы решения проблемы. В тезисе мы опишем метод решения задачи основанных статистических инструментов так называемые контрольные карты (КК) (см. [1], [2]).

Для того чтобы график двухмерной плотности нормальной случайной величины $(X,\, Y)$ с течением времени сохранило свою симметричность и «крутость» необходимо, что их ортогональные сечения относительно осей OX и OY должны удовлетворят этим требованиям. Для проверки их мы можем использовать выборочный коэффициент асимметрии и эксцесса случайных величин X и Y. Независимость или линейную зависимость X и Y можно проверить с помощью выборочного коэффициента корреляции.

Если известно связь между X и Y (это можно также проверить при помощи КК) для управления процесса достаточно пять КК [две КК для асимметрии (a-карты), две КК для эксцесса (γ-карты), и одна КК для коэффициента корреляции (ρ-карта), проверяющие независимость или линейную зависимость]. Для введения КК на производстве необходимо определить их характеристики:

LCL – нижняя контрольная граница;

UCL – верхняя контрольная граница;

CL – центральная линия.

В основе расчета LCL и UCL лежат знание распределения статистического показателя. Для КК асимметрии и эксцесса статистическим показателем является выборочная дисперсия, а для КК определяющие независимость или линейную зависимость статистический показатель есть выборочный коэффициент корреляции CL обычно дается техническим допуском или определяется при помощи предварительных выборок. Для нас достаточно определить LCL и UCL.

Совершенствование процесса с помощью КК есть итерационная процедура. В определенные моменты берут мгновенные выборки и вычисляют значения статистического показателя. Если эти значения попадут в интервал (LCL, UCL), то считается, что процесс в статистически управляемом состоянии, в противном случае говорят, что процесс вышел из управляемого состояния и требуется подналадка процесса.

Имеет место следующая

Теорема. Пусть процесс находиться в статистически управляемом состоянии, тогда статистической достоверностью можно найти LCL и UCL для a-карты, γ-карты, и ρ-карты.

Если введено КК на производстве, то из текущего процесса берем мгновенные выборки и вычисляем значения статистического показателя. То на графике (в диаграмме) мы можем следить за процессом по правилом проведенных выше.

Доказательство теоремы основано на четыре леммы.

ЛИТЕРАТУРА

1. Миттаг Й., Ринне Х. Статические методы обеспечения качества, – М.: "Машиностроение" , 1995.

2. Адлер Ю.П., Максимова О.В., Шпер В.Л. Контрольные карты Шухарта в России и за рубежом: краткий обзор современного состояния (статистические аспекты). Журнал "Стандарты и качество" , 2011, No. 8., pp. 1-40.

S.A.Akhmedov, K.S.Ablazova

On process control, described by two-dimensional normal distribution using control charts.

Abstract. *In this thesis the rule to control the process using control charts is given, when the process described by two-dimensional normal distribution.*

Key words: *two-dimensional normal distribution, skewness, excess, correlation coefficient, control chart.*

The Fokas' method for inverse source problem for heat equation on simple metric graph

M.R.Eshimbetov, G.Khudayberganov, Z.A.Sobirov
National University of Uzbekistan named after Mirzo Ulugbek
mr.eshimbetov92@gmail.com, gkhudaiberg@mail.ru, sobirovzar@gmail.com

Consider simple metric graph Γ with three finite bonds connected on the point O. The point O called to be a vertex of the graph. We label the bonds of the graph as B_j,$j = \overline{1,3}$. Define coordinate x_j on the bond B_j for $j = \overline{1,3}$ corresponding it to the intervals $(0, L_j)$. At each bond the vertex point O has a coordinate 0. Further we will use x instead of x_j.

In each bond of the graph consider the heat transfer equation

$$V_t^{(j)}(x,t) = V_{xx}^{(j)}(x,t) + f^{(j)}(x,t),\ j = \overline{1,3} \tag{1}$$

with initial conditions

$$V^{(j)}(x,0) = V_0^{(j)}(x),\ \ x \in \overline{B_j},\ j = \overline{1,3} \tag{2}$$

and boundary conditions

$$V^{(j)}(x,t)|_{x=L_j} = 0, \quad j = \overline{1,3}. \tag{3}$$

Moreover, we need to define the following gluing conditions for connectivity of the graph

$$V^{(1)}(0,t) = V^{(2)}(0,t) = V^{(3)}(0,t), \sum_{j=1}^{3} \delta_j^2 V_x^{(j)}(0,t) = 0. \tag{4}$$

The last conditions usually called continuity and flux conservation (Kirchhoff) conditions on branching point of the graphs. We suppose, that initial data are smooth enough functions and they satisfies the conditions (3)-(4).

We consider the problem: find functions $\left\{V^{(j)}(x,t),\ f^{(j)}(x,t),\ j = \overline{1,3}\right\}$ that satisfy the equation (1), initial and boundary conditions (2)-(4) and additional conditions at the ends of vertices

$$V_x^{(j)}(x,t)|_{x=L_j} = h_1^{(j)}(t).$$

The solution of the inverse problem for the heat dispersion equation is found by the Fokas method [1,2].

REFERENCES

1. Fokas A. S., Unified Approach to Boundary Value Problems. University of Cambridge Cambridge, United Kingdom: 2008., p: 352.

2. Khudayberganov G., Sobirov Z.A., Eshimbetov M.R., Unified Transform method for the Schrödinger Equation on a Simple Metric Graph. Journal of Siberian Federal University. Mathematics and Physics 2019, 12(4), p: 412–420.

Mardonbek R. Eshimbetov, Gulmirza Khudayberganov, Zarifboy A. Sobirov (Tashkent, Uzbekistan)

The Fokas' method for inverse source problem for heat equation on simple metric graph

Abstract. *We considered initial-boundary value problem for heat equation on simple metric graph with three finite bonds. We find solution of the problem adapting so called Fokas' method to the case of metric graph.*

Key words: *Heat equation, metric graphs, branched structures, Fokas' method, unified transformation, Fourier transformation, initial problem, boundary value problem, inverse problem.*

Об одном случайном процессе с переключениями

В.Р.Ходжибаев
Наманганский инженерно-строительный институт,
vkhodjibayev@mail.ru

Пусть $\left\{\xi_i^{(j)}\right\}_{i=1}^{\infty}$, $\left\{\eta_i^{(j)}\right\}_{i=1}^{\infty}$, $j=1,2$ взаимно независимые последовательности независимых случайных величин, одинаково распределенных внутри каждой последовательности, $a_1 = E\xi_1^{(1)} > 0$, $a_2 = E\xi_2^{(2)} < 0$, $\mathrm{P}\left(\eta_1^{(j)} \geq 0\right) = 1$, $j=1,2$. Обозначим

$$S_0^{(j)} = 0, \quad S_k^{(j)} = \xi_1^{(j)} + \xi_2^{(j)} + ... + \xi_k^{(j)},$$

$$\eta_j(t) = \max\left\{k : \eta_1^{(j)} + \eta_2^{(j)} + ... + \eta_k^{(j)} < t\right\},$$

$$\xi_j(t) = S_{\eta_j(t)}^{(j)}, \quad j=1,2.$$

Случайные процессы $\xi_j(t)$, $j=1,2$ называются обобщенными процессами восстановления. Если случайные величины $\eta_1^{(j)}$ показательно распределены, то процессы $\xi_j(t)$ являются однородными, т.е. обобщенными пуассоновскими процессами.

Для произвольного $b>0$ определим случайные величины

$$\tau_+ = \inf\left\{t \geq 0 : \ \xi_1(t) \geq b\right\}, \ \tau_- = \inf\left\{t \geq 0 : \ \xi_2(t) \leq -b\right\}.$$

Рассмотрим две независимые между собой последовательности $\tau_+,\ \tau_1,\ \tau_3, ...,\ \tau_{2k+1}, ...$ и $\tau_-,\ \tau_2,\ \tau_4, ...,\ \tau_{2k}, ...$ независимых одинаково распределенных внутри каждой последовательности случайных величин, и пусть

$$\theta_k = \tau_1 + \tau_2 + ... + \tau_k, \ \theta_0 = 0.$$

Случайный процесс $X(t)$ строится следующим образом: $X(0) = 0$, при $k = 0,\ 1,\ 2, ...$

$$X(t) = \begin{cases} \min\left\{\xi_1(t-\theta_{2k}),\ b\right\}, & \theta_{2k} < t \leq \theta_{2k+1}, \\ \max\left\{b + \xi_2(t-\theta_{2k+1},\ 0)\right\}, & \theta_{2k+1} < t \leq \theta_{2k+2}. \end{cases}$$

Процесс $X(t)$ совпадает с $\xi_1(t)$ до тех пор пока впервые не будет достигнут уровень b в момент времени τ_1 и полагаем $X(\tau_1) = b$. При $t > \tau_1$ в качестве

приращений процесса $X(t)$ используются приращения процесса $\xi_2(t)$ до тех пор, пока впервые после $\theta_1 = \tau_1$ не будет достигнут нулевой уровень в момент времени θ_2, и в этот момент приращение процесса переключается на приращение процесса $\xi_1(t)$, $X(\theta_2) = 0$. Дальнейшие переключения на процессы $\xi_2(t)$ и $\xi_1(t)$ происходят в моменты $\theta_3,\ \theta_4, ...,$ поочередных достижений уровней b и 0, $X(\theta_{2k+1}) = b,\ \ X(\theta_{2k}) = 0$. Значить,

$$X(t) = \xi_1(t - \theta_{2k}) \quad \text{при} \quad \theta_{2k} \le t < \theta_{2k+1},$$
$$X(t) = b + \xi_2(t - \theta_{2k+1}) \quad \text{при} \quad \theta_{2k+1} \le t < \theta_{2k+2},\ k = 0,\ 1,\ 2, ...$$

В данной работе найдено преобразование Лапласа-Стилтьеса стационарного распределения случайного процесса $X(t)$. Эта задача для случая $P\left(\eta_1^{(j)} = 1\right) = 1,\ \ j = 1, 2$ решалась в [1] и здесь используются методы этой работы.

Из определения ясно, что $X(t)$ является регенерирующим случайным процессом с непрерывным временем. Моменты $\theta_2,\ \theta_4, ...,\ \theta_{2k}, ...$ (последовательные моменты достижений нулевого уровня процессом $X(t)$) являются моментами регенерации, т.к. с вероятностью единица процесс $X(t)$ попадает в нуль и $E(\tau_+ + \tau_-) < \infty$. Времена между моментами регенерации являются независимыми и одинаково распределенными случайными величинами, $E\theta_2 = E\tau_+ + E\tau_- < \infty$. Вероятностные характеристики траектории случайного процесса $X(t)$ между моментами регенерации являются одинаковыми.

Регенерирующие процессы имеют многочисленные применения в различных областях прикладной математики. Библиографические сведения можно найти в [1].

Известно (см., например, [2, гл. 21, теорема 5.1]), что если промежутки времени между моментами регенерации имеют конечное среднее, то существует стационарное распределение для регенерирующего процесса $X(t)$.

Обозначим

$$\varphi_j(\lambda) = Ee^{\lambda\xi_1^{(j)}},\quad f_j(u) = Ee^{-u\eta_1^{(j)}},\quad j = 1, 2,$$
$$\Psi(u,\ \lambda) = \int_0^\infty e^{-ut}Ee^{\lambda X(t)}dt, \lim_{t\to\infty} P(X(t) < x) = F(x),$$
$$W(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x}dF(x).$$

Далее используется факторизация

$$1 - z\varphi_j(\lambda) = R_-^{(j)}(z,\ \ ,\lambda)\, R_+^{(j)}(z,\ \lambda),\ \ |z| < 1,\ \ Re\lambda = 0,$$

введенная в [2].

Отметим, что при отображении $z = f_j(u)$ множество $\{u : u \geq 0\}$ переходит в множество $\{z : 0 < z \leq 1\}$, а при малых $\delta > 0$ множество $\{u : 0 \leq u < \delta\}$ переходит внутрь множества $\{z : 1 - \gamma < z \leq 1\}$ при некотором $\gamma > 0$. Поэтому при $u > 0$ имеет место факторизации

$$1 - f_j(u)\,\varphi_j(u) = R_-^{(j)}(f_j(u), \lambda) R_+^{(j)}(f_j(u), \lambda)$$

со всеми известными свойствами их компонент.

Известно, что ([2]), $(R_+^{(2)})^{-1}(1,\, \lambda) = \frac{Ee^{\lambda S_2}}{1-p_2}$, $\quad (R_-^{(1)})(1,\, \lambda) = \frac{Ee^{\lambda S_1}}{1-p_1}$,
где

$$S_1 = \inf_{k\geq 0} S_k^{(1)}, \quad p_1 = P(S_1 < 0) < 1, \quad S_2 = \sup_{k\geq 0} S_k^{(2)}, \quad p_2 = P(S_2 < 0) < 1.$$

Теорема. *Для преобразования Лапласа – Стилтьеса стационарного распределения процесса $X(t)$ справедливо следующее представление:*

$$W(\lambda) = \frac{E\eta_1^{(1)} Ee^{\lambda S_1} \lim\limits_{u\to 0}\left[(R_+^{(1)})^{-1}(f_1(u),\lambda)\right]^{[0,b)}}{\bar{a}(1-p_1)} + \frac{E\eta_1^{(2)} Ee^{\lambda S_2} e^{\lambda b} \lim\limits_{u\to 0}\left[(R_-^{(2)})^{-1}(f_2(u),\lambda)\right]^{(-b,0]}}{\bar{a}(1-p_2)},$$

в котором $u > 0, \quad Re\lambda = 0.$

ЛИТЕРАТУРА

1. Лотов В. И. О случайном блуждании с переключениями. Сибирские электронные математические известия, 2018. № 15. с.1320-1331.

2. Borovkov A. A. Probability Theory. London: Springer-Verlag, 2013. MR 3086572.

V.R.Xodjibayev (Namangan, Uzbekistan)

On the Generalized Renewal Processes with Switchings

Abstract. *We find the Laplace-Stielties transforms of stationary distribution of a stochastic process with switchings between two generalized renewal processes while achieving regulatory barriers.*

Key words: *generalized renewal process, regenerative process, stationary distributions.*

Infinite dimensional Leibniz algebras constructed by $\mathfrak{sl}_2$ Lie algebras

T.K.Kurbanbaev
V.I.Romanovskiy Institute of Mathematics, Uzbekistan Academy of Sciences,
tuuelbay@mail.ru;

Leibniz algebras were introduced and developed by J.-L.Loday [6]. They are a generalization of Lie algebras, removing the restriction that the product is anti-commutative or that the square of an element is zero.

For Lie algebras, it is known that an arbitrary finite-dimensional Lie algebra over a field of characteristic zero decomposes into the semidirect sum of the maximal solvable ideal and its semisimple subalgebra (Levi's Theorem, [4]). A similar result is also true for Leibniz algebras, namely, a finite-dimensional Leibniz algebra decomposes into the semidirect sum of a maximal solvable ideal and a semisimple Lie subalgebra (Levi's Theorem, [3]). A complete description of semisimple finite-dimensional Lie algebras over a field of characteristic zero is known [3, 4]. Therefore, the study of finite-dimensional Leibniz algebras is reduced to the study of solvable ones.

A non-Lie Leibniz algebra L contains the non-trivial ideal, generated by the squares of elements of the algebra L (denoted by $\mathcal{I}$ and usually called the *Leibniz kernel*), i.e., $\mathcal{I} = \langle [x, x] \mid x \in L \rangle$. The ideal $\mathcal{I}$ plays an important role in the theory of Leibniz algebras since it determines the (possible) non-Lie property of a Leibniz algebra. Moreover, this ideal belongs to the right annihilator of L and it is the minimal ideal with the property that the quotient algebra $L/\mathcal{I}$ is a Lie algebra.

Note that in [2] and [5] the infinite-dimensional Leibniz algebra constructed by Witt and Virasoro algebras is considered. This paper is devoted to the study of the infinite-dimensional Leibniz algebra constructed by the Lie algebra $\mathfrak{sl}_2$ and simple $\mathfrak{sl}_2$-modules.

Definition 1.1. An algebra $(\mathcal{L}, [\cdot, \cdot])$ over a field $\mathbb{K}$ is called a *Leibniz algebra* if it is defined by the Leibniz identity

$$[x, [y, z]] = [[x, y], z] - [[x, z], y], \text{ for all } x, y \in L.$$

For a Leibniz algebra $\mathcal{L}$ we consider the natural homomorphism onto the quotient Lie algebra $\mathcal{L}/\mathcal{I}$, which is called the *corresponding Lie algebra* to the Leibniz algebra $\mathcal{L}$ (in some references this algebra is called the *liezation* of $\mathcal{L}$).

Now we present a construction of Leibniz algebras by a given Lie algebra and its right module.

Let $(\mathfrak{g}, [-,-])$ be a Lie algebra and let V be a right $\mathfrak{g}$-module. We equip the vector space $Q(\mathfrak{g}, V) = \mathfrak{g} \oplus V$ with the multiplication $(-,-)$ in the following way:

$$(x+v, y+w) := [x,y] + v \star y, \qquad x, y \in \mathfrak{g},\ v, w \in \mathcal{V}, \tag{1}$$

where $\star$ is the action of $\mathfrak{g}$ on V. Then $Q(\mathfrak{g}, V)$ is a Leibniz algebra and from (1) we get $(\mathfrak{g}, V) = (V, V) = 0$.

Note that if $V \star \mathfrak{g} = V$, then V is nothing else but the ideal $\mathcal{I}$ of the Leibniz algebra $Q(\mathfrak{g}, V)$.

The map $\mathcal{I} \times \mathcal{L}/\mathcal{I} \longrightarrow \mathcal{I}$ defined as $(v, \overline{x}) \longrightarrow [v, x]$, $v \in \mathcal{I}$, $x \in \mathcal{L}$ endows $\mathcal{I}$ with a structure of a right $\mathcal{L}/\mathcal{I}$-module (it is well-defined due to $\mathcal{I}$ being in the right annihilator). Thus, for a given Leibniz algebra $\mathcal{L}$ we have a Lie algebra $\mathcal{L}/\mathcal{I}$ and its right module $\mathcal{I}$.

The main goal of this paper is to describe Leibniz algebras such that its corresponding Lie algebra is a given Lie algebra $\mathfrak{g}$ and the ideal $\mathcal{I}$ is a given right $\mathfrak{g}$-module V.

Let $\{e, f, h\}$ be a basis of $\mathfrak{sl}_2$ in which it admits the following products:

$$[e,f] = h, \ \ [e,h] = 2e, \ \ [f,h] = -2f.$$

This simple 3-dimensional Lie algebra is denoted by $\mathfrak{sl}_2$ and the basis $\{e, f, h\}$ is called canonical basis. Note that any three-dimensional simple Lie algebra is isomorphic to $\mathfrak{sl}_2$. The representation of $\mathfrak{sl}_2$ is determined by the images $E,\ F,\ H$ of the basis elements e, f, h and we have

$$[E,H] = 2E, \quad [F,H] = -2F, \quad [E,F] = H.$$

Theorem 1 [4]. *For each integer $m = 0, 1, 2, \ldots$ there exists one and, in the sense of isomorphism, only one irreducible $\mathfrak{sl}_2$-module $V(m+1)$ of dimension $m+1$. The module $V(m)$ has a basis $\{x_0, x_1, \ldots, x_m\}$ such that the representing transformations E, F and H corresponding to the canonical basis $\{e, f, h\}$ are given by:*

$$H(x_k) = (m-2k)x_k, \ \ k = 0, \ldots, m,$$

$$F(x_m) = 0, \ \ F(x_k) = x_{k+1}, \ \ k = 0, \ldots, m-1,$$

$$E(x_0) = 0, \ \ E(x_k) = -k(m+1-k)x_{k-1}, \ \ k = 1, \ldots, m.$$

Let V be a $\mathfrak{sl}_2$-module (not necessarily finite-dimensional). For $\lambda \in \mathbb{C}$ denote

$$V_\lambda = \{v \in V \ : \ H(v) = \lambda \cdot v\}.$$

The number λ is called a *weight* and the space V_λ is called the corresponding *weight space*. The module V is called a *weight module* provided that

$$V = \bigoplus_{\lambda \in \mathbb{C}} V_\lambda.$$

For a weight $\mathfrak{sl}_2$-module V we define the *support* $\mathrm{supp}V$ of V as follows:

$$\mathrm{supp}(V) = \{\lambda \in \mathbb{C} \ : \ V_\lambda \neq 0\}.$$

If V is a weight $\mathfrak{sl}_2$-module and $\mu \in \mathrm{supp}V$ is such that $\mu + 2 \notin \mathrm{supp}V$, then the weight μ is called a *highest weight* and any nonzero $v \in V_\mu$ is called a *highest weight* vector. A module, generated by a highest weight vector, is called a *highest weight* module.

Fix now $\xi \in \mathbb{C}/2\mathbb{Z}$ and $\tau \in \mathbb{C}$. Consider the vector space $V(\xi, \tau)$ with the basis $\{v_\mu \ : \ \mu \in \xi\}$. Consider the linear operators E, F and H on $V(\xi, \tau)$ defined as follows:

$$\begin{aligned} H(v_\mu) &= \mu v_\mu, \\ F(v_\mu) &= v_{\mu-2}, \\ E(v_\mu) &= \tfrac{1}{4}(\tau - (\mu+1)^2) v_{\mu+2}. \end{aligned} \tag{2}$$

The module $V(\xi, \tau)$ is called the *dense module* with with support ξ.

Lemma 1 [7]. *The formulae* (2) *define on* $V(\xi, \tau)$ *the structure of a dense* $\mathfrak{sl}_2$*-module with support* ξ.

Theorem 3 (Structure of $V(\xi, \tau)$) [7]. *Let* $\xi \in \mathbb{C} \setminus 2\mathbb{Z}$ *and* $\tau \in \mathbb{C}$. *The module* $V(\xi, \tau)$ *is simple if and only if* $\tau \neq (\mu+1)^2$.

Let us fix now $\xi \in \mathbb{C}/2\mathbb{Z}$ and $\tau \in \mathbb{C}$. Let $V = V(\xi, \tau)$ be a simple dense module with the basis $\{v_\mu \ : \ \mu \in \xi\}$. Then by formulae (2) we have the following brackets:

$$\begin{aligned} [v_\mu, h] &= \mu v_\mu, \\ [v_\mu, f] &= v_{\mu-2}, \\ [v_\mu, e] &= \tfrac{1}{4}(\tau - (\mu+1)^2) v_{\mu+2}, \end{aligned} \tag{3}$$

Let $\mathcal{L}$ be Leibniz algebra such that its corresponding $\mathfrak{g} = \mathfrak{sl}_2$ algebra and the right $\mathfrak{sl}_2$-module $V = V(\xi, \tau)$ is the ideal $\mathcal{I}$. We equip the direct sum of the vector spaces $\mathcal{L} = \mathfrak{g} \oplus \mathcal{I}$ with a product $[-,-]$ satisfying the conditions:

$$[V, \mathfrak{g}] = V \star \mathfrak{g}, \quad [\mathfrak{g}, V] = [V, V] = 0,$$

where the product $V \star \mathfrak{g}$ follows from (3).

Thus, in order to completely describe the Leibniz algebra $\mathcal{L}$ we have to clarify the product $[\mathfrak{g}, \mathfrak{g}]$.

Theorem 4. *Let $\mathcal{L}$ be a Leibniz algebra and let $\mathcal{I}$ be the ideal generated by squares in $\mathcal{L}$. We assume that the quotient $\mathcal{L}/\mathcal{I}$ is isomorphic to the Lie algebra $\mathfrak{sl}_2$. Then there exists a basis $\{e, f, h, v_\mu, \mu \in \xi\}$ of $\mathcal{L}$ such that the non zero products of the basis vectors in $\mathcal{L}$ are represented as follows*:

$$[e,h] = 2e, \qquad [f,h] = -2f, \quad [e,f] = h,$$

$$[h,e] = -2e, \qquad [h,f] = 2f, \quad [f,e] = -h,$$

$$[v_\mu, h] = \mu v_\mu,$$

$$[v_\mu, f] = v_{\mu-2},$$

$$[v_\mu, e] = \tfrac{1}{4}(\tau - (\mu+1)^2) v_{\mu+2},$$

where the omitted products are equal to zero, $\xi \in \mathbb{C}/2\mathbb{Z}$, $\tau \in \mathbb{C}$.

Reference

1. Barnes D.W., On Levi's theorem for Leibniz algebras, *B.Aust. Math. Soc.*, **86 (2)** (2012), 184-185.
2. Camacho L.M., Omirov B.A., Kurbanbaev T.K., Leibniz algebras constructed by Witt algebras, *Linear and Multilinear algebra* (2018) doi.org/10.1080/03081087.2018.1480704
3. Humphreys J.E., Introduction to Lie algebras and Representation theory, *Springer-Verlag New York*, 1972.
4. Jacobson N., Lie algebras, 340p. *Interscience Publishers, Wiley, New York*, (1962).

5. Kurbanbaev T.K. Leibniz algebras corresponding to the Virasoro algebra, *Uzbek Mathematical Journal*, 2018, No 1, pp.115-120.

6. Loday J.-L., Une version non commutative des algèbres de Lie: les algèbres de Leibniz, *Enseign. Math.*, **19** (1993), 269-293.

7. Mazorchuk V. Lectures on $\mathfrak{sl}_2(\mathbb{C})$-modules, *World Scientific Publishing,* 2009, P.276.

Kurbanbaev T.K. (Tashkent, Uzbekistan)

Infinite dimensional Leibniz algebras constructed by $\mathfrak{sl}_2$ Lie algebras

Abstract. *We study Leibniz algebras are constructed by the Lie algebra $\mathfrak{sl}_2$ and the dense $\mathfrak{sl}_2$-modules*

Key words: *Lie algebra; Leibniz algebra; simple $\mathfrak{sl}_2$-module*

Моделирования рынка страховых услуг

Р.Т.Турсунов[1], А.А.Алимов[2]
[1] Ташкентский филиал Российского экономического Университета им. Г.В. Плеханова,
[2] Национальный Университет Узбекистана имени Мирзо Улугбека
rasul_tursunov@yahoo.com, akram_alimov@mail.ru

Реальные преобразования, происходящие сегодня во всех сферах экономико-политической жизни республики, создают объективные предпосылки для возрастания роли страхования, как одной из составляющих финансовой системы экономики. Страховой рынок является одним из важнейших сегментов рыночной экономики, роль которого, в социально-экономическом развитии государства, трудно переоценить.

В соответствии "Стратегии действий по пяти приоритетным направлениям развития Республики Узбекистан в 2017-2021 годах"целенаправленно осуществляется меры по дальнейшему развитию страхового рынка страны [1].

Страховыми компаниями удается стремительно наращивать объемы страховых операций в основном по накопительным видам страхования жизни. Главным стимулом данного роста являются созданные правительством страны благоприятные налоговые условия для физических лиц, пользующихся страховыми услугами по долгосрочному страхованию жизни. При этом следует отметить, что в 2018 году ставка налога на доход физических лиц была снижена до 12%, что может оказать влияние на тенденции и динамики объемов страховых премий в отрасли страхования жизни в последующие годы.

Основными показателями страхового рынка являются объем страховых премий и выплат. Рост объема страховых премий показывает расширение деятельности страховых компании в сфере страхования, что положительно влияет на имидж кредитной организации, укрепление ее стабильности и финансовой устойчивости, а это, в свою очередь, способствует расширению клиентской базы как на уровне одного страхового компания, так и на уровне всей страховой системы.

В целом по объему премий явными лидерами рынка страховых услуг со значительным отрывом от последующих компаний являются АО "Узагросугурта"(14,6%), НКЭИС "Узбекинвест"(13,9%) и ООО "Gross Insurance"(12,6%). Последующими двумя компаниями, занимающими 4- и

5-позиции в отрасли являются АО "Страховая компания Кафолат"(9,6%) и СО ООО "Alfa Invest"(6,8%). Доли остальных компаний в отрасли составляют ниже 6,0% [2].

Известно, что соотношения страховых выплат к страховым премиям определяет уровень убыточности страховых премий как на уровне одного страхового компания, так и на уровне всей страховой системы.

Целью данной работы является выявление зависимости между объемом уплаченных страховых выплат и объемом страховых премий, собранных страховыми компаниями. Для проведения количественной оценки влияния объема страховых премий на страховых выплат использован метод корреляционно-регрессионного анализа.

В качестве эндогенной (зависимой) переменной выбран объем страховых выплат. Это связано с тем, что именно данный показатель является основным важным индикатором работы страховых компаний и характеризует обоснованность всех проводимых кампаниями расходов для достижения максимума прибыли. Экзогенной (независимой) переменной в данном исследовании будет объем страховых премий.

Для проведения исследования были использованы данные за 15 лет (период с 2003 по 2018 гг.), представленные на сайте Министерства Финансов Республики Узбекистан [3].

Сначала определим корреляцию между объемов страховых премий и выплат. По вычисленной коэффициента корреляции сделан вывод о прямой сильной связи между ними, так как значение этого коэффициента находится в пределах $+0,8 < r < +1$.

Следующий шаг - подобрать функцию, наилучшим образом описывающую зависимость Y от , чтобы отклонения реальных значений параметра от модельных были малы. В данном случае можно выдвинуть гипотезы о линейном или нелинейном характере зависимости между переменными: влияние совокупных объемов страховых премий на величину совокупного объема выплат (Y) в текущих ценах описывается линейной либо нелинейной функцией. Этот вывод подтверждается характером разброса точек на корреляционном поле.

По результатам проведенного исследования построены линейная, полиномиальная, степенная, логарифмическая и экспоненциальная регрессионные модели рынка страховых услуг (рис. 1, 2).

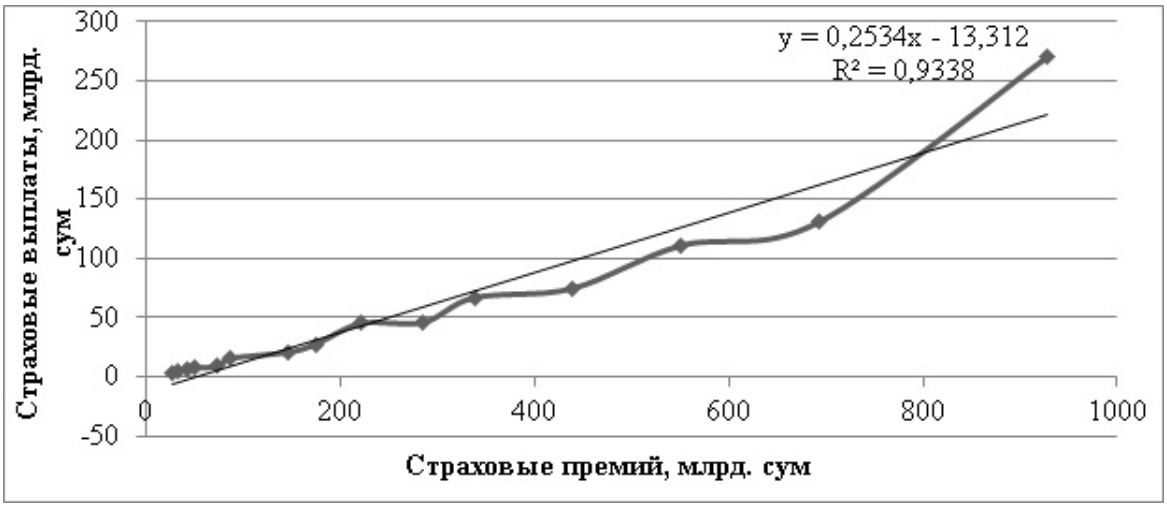


Рис. 1 - Линейная регрессионная модель

Оценка качества построенной линейной регрессионной модели - основных характеристик исследуемых переменных уравнения парной регрессии Y от X и его параметров (рис. 1) показывает:

Коэффициент корреляции между страховыми выплатами и премий равна 0,9663, а коэффициент детерминации равна 0,9338. Это означает, что 93,38 % вариации страховых выплат объясняется вариацией страховых премий. С увеличением страховых премий на 1 млрд. сум страховые выплаты уменьшается в среднем на 13,312 млрд. сум.

Фактическое $F-$ значение Фишера $F = 138,38$. Сравнив его с критическим значением $F-$ критерия, определяемым по заданным уровню значимости $\alpha = 0,01$ и числу степеней свободы $\kappa_1 = 2-1 = 1$ и $\kappa_2 = 15-2 = 13$ составляет $F = 9,074$. Таким образом, на основе $F > F$ выявленная зависимость величины страховых выплат от объемов страховых премий за 2003-2017 гг., носит неслучайный характер.

Средняя стандартная ошибка рассчитанной на основе линейной модели составляют 0,5573 или ошибка аппроксимации 55,73%.

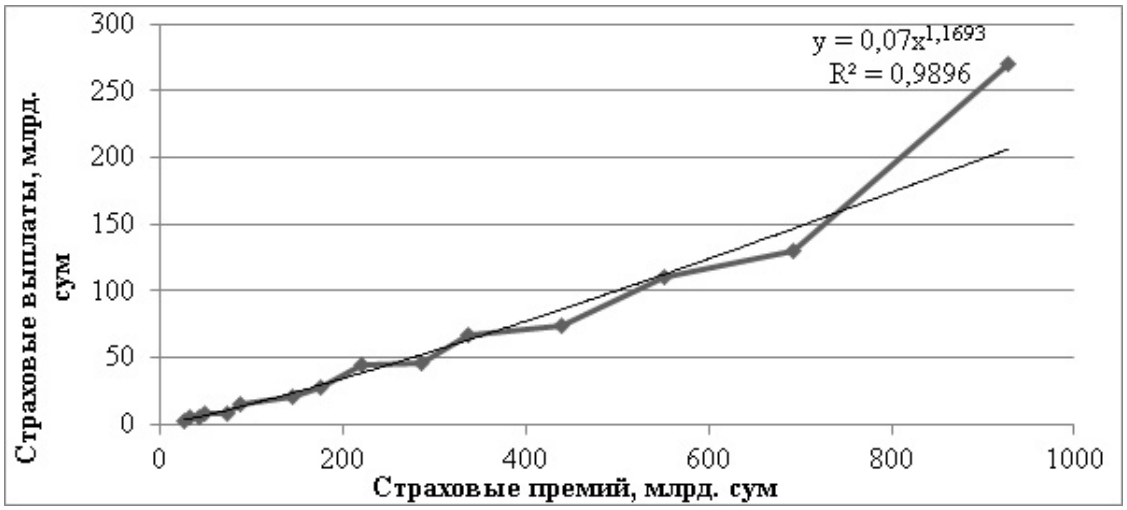


Рис. 2 - Степенная регрессионная модель

В построенной степенной регрессионной модели - основных характеристик исследуемых переменных уравнения парной регрессии Y от X и его параметров (рис. 2) показывает:

В степенной регрессионной модели параметр b=1,1693 является коэффициентом эластичности. Этот коэффициент показывает что, если объем страховой премий изменится 1%, то страховые выплаты в среднем изменится (увеличится) на 1,1693%. Коэффициент детерминации очень высокая, равна 0,9896. Это означает, что 98,96% вариации страховых выплат объясняется вариацией страховых премий.

Фактическое $F-$ значение Фишера $F = 1237$, которое намного больше его критического значения $F-$ критерия ($F = 9,074$) по уровню значимости $\alpha = 0,01$. Что доказывает сильной зависимости величины страховых выплат от объемов страховых премий за 2003-2017 гг.

Средняя стандартная ошибка рассчитанной на основе степенной регрессионной модели составляют 0,1132 или ошибка аппроксимации равна 11,32%, которая 5 раз лучше, чем линейной модели.

ЛИТЕРАТУРА

1. Указ Президента Республики Узбекистан "Стратегии действий по пяти приоритетным направлениям развития Республики Узбекистан в 2017-2021 годах"№УП-4947, 07.02.2017 г.

2. Обзор страхового рынка Узбекистана. [Электронный ресурс] - Режим доступа https://www.insurance.uzreport.uz/ (25.12.2018).

3. Официальный сайт Министерство Финансов Республики Узбекистан. [Электронный ресурс] - Режим доступа https://www.mf.uz/(18.01.2019).

R.T.Tursunov (Tashkent Branch of Russian Economic University named after G.V. Plekhanov), A.A.Alimov (National University of Uzbekistan after Mirza Ulugbek.)

Modelling of insurances market

Abstract. *In the article by means of correlation and regression analysis there revealed interrelation between total volume of insurance premiums and insurance payments of insurance market of the Republic of Uzbekistan. It is proved that the most optimum is sedate regression model.*

Key words: *the insurance market, insurance awards, insurance payments, a loss ratio of insurance premiums, correlation.*

Вероятность попадания в ноль ветвящихся процессов иммиграцией с τ типами частиц

Р.И.Ибрагимов, А.Машраббоев
Наманганский Государственный Университет

Пусть $Z_t = (Z_{t_1}, Z_{t_2}, ..., Z_{t_\tau})$ -ветвящийся процесс с иммиграцией, заданный производящей функцией плотностей

$$f^i(s) = \sum_{k\geq 0} P_k^i S^k = \sum_{k_1,...,k_\tau \geq 0} P^i_{k_1,...,k_\tau} S^{k_1,...,k_\tau}, \sum_{k_1,...,k_\tau \geq 0} P_{k_1,..,k_\tau} = 0 \qquad (1)$$

(вероятность превращения одной частицы типа i, $i = \overline{1,\tau}$ в совокупности частиц $k = (k_1, k_2, ...k_\tau)$ за время $\Delta t \to 0$ равна $P^i_{k_1,...,k_\tau}(t+\Delta t)$, если $(k_1, k_2, ..., k_\tau) \neq (0, 0, .., l_i, 0, 0, ..., 0)$, при $k = (0, 0, .., l_i, 0, 0, ..., 0)$, $1 + P^i_{k_1,...,k_\tau}(\Delta t + o(\Delta t))$ и

$$g(s) = \sum_{k\geq 0} g_k S^k = \sum_{k_1,...,k_\tau \geq 0} g_{k_1,...,k_\tau} S^{k_1,...,k_\tau}, \sum_{k\geq 0} g_k = 0 \qquad (2)$$

вероятность иммиграции $k = (k_1, k_2, ...k_\tau)$- частиц за время $\Delta t \to 0$ при $k \neq (0, 0, ..., 0)$ равна $g_k \Delta t + o(\Delta t)$, при $k = (0, 0, ..., 0)$ равна $1 + g_0 \Delta t + o(\Delta t)$ (см. [I])).

Положим

$$a^i_j = \left.\frac{\partial f^i(s)}{\partial s_j}\right|_{s=1}, \lambda_i = \left.\frac{\partial g(s)}{\partial s_i}\right|_{s=1}, b^i_{jl} = \left.\frac{\partial^2 f^i(s)}{\partial s_j \partial s_l}\right|_{s=1}, b = \sum_{i,j,l} v_l b^i_{jl} u^i u^l,$$

$$\theta = 2\sum_{i=1}^{\tau} \frac{\lambda_i u^i}{b}, i, j, l = \overline{1,\tau},$$

где $u = (u^1, u^2, ..., u^\tau)$, $v = (v_1, v_2, ..., v_\tau)$ – собственные векторы матрицы $\left\|a^i_j\right\|$,

$$\sum_i u^i = \sum_i u^i v_i = 1, u^i > 0, v_i > 0, i, j = \overline{1,\tau},$$

Через τ обозначим продолжительность жизни процесса z_t. Имеет место следующее утверждение:

Теорема 1. Если

$$\lambda_i = \frac{\partial G_i(s)}{\partial s}|_{s=1} < \infty$$

и

$$1 - F_i(t,s) = \frac{u^i \sum_{k=1}^{\tau} v_k(1-s_k)}{1 + \frac{B_i t}{2} \sum_{k=1}^{\tau} v_k(1-s_k)}(1 - \alpha_i(t,s)),$$

где

$$\alpha_i(t,s) \to 0, 1 - t \to \infty, |s| \leq 1, j = 1, \tau,$$

то

$$P\{\tau_i = \infty\} = \lim_{k\to\infty} u_i(k) = \frac{1}{(1-q_0)A_i'(1)}.$$

ЛИТЕРАТУРА

1. Севастьянов Б.А. Ветвящиеся процессы. М.: Наука, 1971.
2. Бадалбаев И.С., Машраббаев А. Асимптотика вероятности попадания в нуль ветвящихся процессов с иммиграцией. Докл.АН 1984. №1. стр. 5-8.
3. Ибрагимов Р., Машраббаев А. ВЕРОЯТНОСТЬ ПОПАДАНИЯ В НУЛЬ ВЕТВЯШИХСЯ ПРОЦЕССОВ С ИММИГРАЦИЕЙ С τ ТИПАМИ ЧАСТИЦ. Материалы Республиканской научной конференции. Ташкент. 2019 г. стр.35-36. НУУз.

Предельная теорема для ветвящегося процесса Беллмана-Харриса с иммиграцией

А.Машраббоев, Р.Полванов, А.Рахманов

Пусть задан ветвящийся процесс Беллмана-Харриса с двумя типами частиц. В этой модели имеются частицы типа T_1 и T_2.

Каждая из частиц типа $T_i, i = 1, 2$, имеет случайную продолжительность жизни τ^i с функцией распределения $P\{\tau^i < t\} = G^i(t)$. В конце своей жизни частица превращается в некоторое количество частиц типа T_i и T_2 нулевого возраста.

Через $P^i(\alpha)$ обозначим вероятность превращения частиц типа $T_i, i = 1, 2$, в совокупность частиц, состоящие из α_j частиц типа $T_j, j = 1, 2$, где $\alpha = (\alpha_1, \alpha_2)$.

Введем следующие обозначения:

$$s = (a_1, s_2), s^\alpha = s_1^{\alpha_1} \cdot s_2^{\alpha_2}, \partial s^\alpha = \partial s_1^{\alpha_1} \partial s_2^{\alpha_2}, |n| = n_1 + n_2,$$

$$s \cdot y = \sum_{i=1}^{2} s_i y_i, [x]^{[k]} = x(x-1)...(x-k+1).$$

Если $s_i \le y_i (s_i < y_i)$, то $s \le y (s < y)$, если $i = P, P+1, ..., m$, то будем писать $i = P + m$.

Введем производящие функции распределения частиц непосредственных потомков частиц типа $T_i : f^i(s) = \sum_\alpha P^i(\alpha) s^\alpha, i = 1, 2$.

Пусть вектор $z_i(t) = (z_{1i}(t), z_{i2}(t))$ означает, что в момент t существуют z_{ij} частиц типа $T_j, j = 1, 2$, в предположении, что в начальный момент $t = 0$ была одна частица типа $T_i, i = 1, 2$.

Введем производящие функции $F^i(t, s) = \sum_\alpha P\{z_i(t) = \alpha\} s^\alpha, i = 1, 2$. Известно [1], что $F^i(t, s)$ удовлетворяют следующей системе интегральных уравнений:

$$F^i(t, s) = s_i(1 - b_1^i(t)) + \int_0^t f^i(F(t-u, s)) dG^i(u), i = 1, 2,$$

$$F^0(t, s) = (1 - b_1^0(t)) + \int_0^t F^0(t-u, s) f^0(F(t-u, s)) dG^0(u).$$

Пусть $d_{ij} = \frac{\partial f^i}{\partial s_j} = a_j < \infty, j = 1, 2$, где $\frac{\partial f^0}{\partial s_j} = a_j < \infty, j = 1, 2$.

$D = \|d_{ij}\|$- матрица математических ожиданий, перронов корень матрицы $D : \rho = 1$.

В этом случае ветвящийся процесс с иммиграцией называется критическим.

Пусть

$$F^i(t,s) = s + (1-s)^{1+v} L_i(\frac{1}{1-s}), o \le v \le 1,$$

где $L_i(t)$ - медленно меняющаяся функция;

$$A_i(s) = \int e^{-st} d\gamma_i(t)$$

и

$$U_i(t) = P\{\tau_i > t | Z_i(-1) = 0, Z_i(0) > 0\}$$

$\lambda_i = v_i'(0)$ - вероятность иммиграции k частиц.

Теорема 1. При выше приведенных условиях верно $A_i'(0) < \infty$ при $v_i = \frac{\lambda_i, u_i}{p_i} = \frac{2\sum_{k=1}^{\tau} \lambda_k^i u_i^k}{b_i} > 1$.

Теорема 2. При выше приведенных условиях верно: $A_i'(0) = \infty$ и $\theta_i = \frac{(\lambda_i u_i)}{p_i} < 1$.

ЛИТЕРАТУРА

1. Севастьянов Б.А. Ветвящиеся процессы. М.: Наука, 1971.
2. Феллер В. Введение в теорию вероятностей и еҷ приложения. Т-1,2. М.:Мир, 1967, 528 с., 738 с.
3. Бадалбаев И.С., Машраббоев А. Периоды жизни ветвящегося процесса Беллмана-Харриса с иммиграцией. // Вероятностные распределения и математическая статистика Т.: "Фан 1986, 60-82 с.
4. Машраббоев А., Ибрагимов Р. Предельная теорема для ветвящегося процесса Беллмана-Харриса с иммиграцией. // Материалы. - Международной Ферганской конференции "Предельные теоремы теории вероятностей и их приложения (г. Фергана, 10-12 мая 2005 г.) Ташкент 2005. с. 130-131.

Асимптотические задачи последовательного точечного оценивания функционалов от неизвестной функции распределения

Г.Г.Рахимова
Национальный университет Узбекистана имени Мирзо Улугбека
rakhimova.gulnoza92@gmail.com

1. Введение. В отличие от классических методов математической статистики, в которых число производимых наблюдений фиксируется заранее, методы последовательного анализа характеризуются тем, что момент прекращения наблюдений (момент остановки) является случайным и определяется в зависимости от значений наблюдаемых данных и от принятой меры оптимальности построенной статистической оценки в изучаемой статистической модели. В простом статистическом примере изложим основные асимптотические задачи последовательного точечного оценивания.

Рассмотрим на вероятностном пространстве $(\Omega, \mathrm{F}, \mathrm{P})$ последовательность $\xi_1, \xi_2, ..., \xi_n, ...$ независимых одинаково распределенных случайных величин со средним θ и дисперсией σ^2. Неизвестное среднее θ оценим эмпирическим средним $\theta_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i$.

Предположим, что при точечном оценивании θ статистикой θ_n функция потерь состоит из квадратичной ошибки и стоимости выборки, то есть $L_n = A\left(\theta_n - \theta\right)^2 + \varepsilon \cdot n$, где A известная весовая постоянная, ε - стоимость единицы выборки. Тогда риск оценивания θ статистикой θ_n равен

$$R_n(\sigma) = E\left(L_n\right) = \frac{A\sigma^2}{n} + \varepsilon \cdot n.$$

Если σ^2 известна, то риск является минимальным при $n = n_0(\varepsilon) = \inf\left(n \geq 1 : n \geq \sigma \cdot \sqrt{\frac{A}{\varepsilon}}\right)$ и $\min\limits_{n\geq 1} R_n(\sigma) = R_{n_0(\varepsilon)}(\sigma) = 2\varepsilon n_0(\varepsilon) = 2\sigma \cdot \sqrt{A\varepsilon}$, то есть θ_n является оценкой минимального риска для θ.

Если σ^2 неизвестна, то как доказано еще Леманом [1] для любой оценки θ_n среднего θ не существует выборочной процедуры с фиксированным неслучайным объёмом выборки, минимизирующей риск одновременно для всех σ^2. В этом случае задача планирования выборочной процедуры фактически сводится к задаче выбора правила или момента остановки $N(\varepsilon) = \inf\left(n \geq 1 : n \geq \sigma_n \cdot \sqrt{\frac{A}{\varepsilon}}\right)$, где σ_n^2 состоятельная оценка для σ^2.

Если в этой выборочной процедуре $\lim\limits_{\varepsilon\to 0}\frac{R^*_{N(\varepsilon)}(\sigma)}{R_{n_0(\varepsilon)}(\sigma)} = 1$, где $R^*_{N(\varepsilon)}(\sigma) = A\left(\theta_{N(\varepsilon)} - \theta\right)^2 + \varepsilon E\left(N(\varepsilon)\right)$, то оценка $\theta_{N(\varepsilon)}$ называется оценкой с асимпто-

тически минимальным риском для θ. Заметим, что асимптотическая оптимальность оценки $\theta_{N(\varepsilon)}$ является по отношении к её непоследовательному аналогу $\theta_{n_0(\varepsilon)}$, основанному оптимальному объёму выборки $n_0(\varepsilon)$.

Содержательный обзор результатов, полученных в этом направлении приведены в работах [2] и [3].

2. Основные результаты. Пусть $\xi_1, \xi_2, ..., \xi_n$ независимые одинаково распределенные случайные величины с неизвестной функцией распределения $F(x)$, $x \in R_1, \in \mathbf{F}$, где $\mathbf{F}$ семейство функций распределений, удовлетворяющих определенным условиям регулярности. Для оценки функционала $\theta(F)$ от неизвестной функции распределения $F(x)$ рассмотрим статистику $\theta_n(F) = \theta_n(\xi_1, \xi_2, ..., \xi_n)$. Например, $\theta(F)$ может быть плотностью вероятности, функцией регрессии или характеристической функцией и так далее. Обозначим через $\varphi(y)$, $y \geq 0$ положительную, возрастающую функцию с $\varphi(0) = 0$, через $\alpha_n = \alpha_n(F) = |\theta_n(F) - \theta(F)|$ абсолютную погрешность оценивания и через ε стоимость единицы выборки.

Определение 1. Величина $L_n = L_n(\varepsilon, F) = \varphi(\alpha_n) + \varepsilon \cdot n$ называется функцией потерь оценивания $\theta(F)$ статистикой $\theta_n(F) = \theta_n(\xi_1, \xi_2, ..., \xi_n)$.

Например, если $\varphi(y) = y$, то L_n является линейной функцией потерь и если $\varphi(y) = y^2$, то L_n является квадратичной функцией потерь.

Введем условие:

(A_1) для всех $n \geq 1$, $\beta_n(F) = E(\varphi(\alpha_n))$ существует, убывает по n и $\lim\limits_{n\to\infty} \beta_n(F) = 0$.

Определение 2. Величина

$$R_n(\varepsilon, F) = E(L_n(\varepsilon, F)) = \beta_n(F) + \varepsilon \cdot n$$

называется функцией риска оценивания $\theta(F)$ статистикой $\theta_n(F) = \theta_n(\xi_1, \xi_2, ..., \xi_n)$.

Поскольку величина $\beta_n(F)$ убывает, а $\varepsilon \cdot n$ возрастает по n, то существует такое оптимальное значение $n_0 = n_0(\varepsilon, F)$, что

$$R_{n_0}(\varepsilon, F) = \inf_{n \geq n_0} R_n(\varepsilon, F). \tag{1}$$

Определение 3. Если для оценки $\theta_n(F) = \theta_n(\xi_1, \xi_2, ..., \xi_n)$ выполняется (1), то она называется оценкой минимального риска.

Здесь следует отметить, что оптимальный объём выборки $n_0 = n_0(\varepsilon, F)$ зависит от неизвестной ф. р. $F(x)$ через $\beta_n(F)$, следовательно, как отмечено во введении только выборка со случайным объёмом в виде момента остановки может обеспечить выполнение (1), то есть оценка минимального

риска $\theta_n(F)$ может быть построен, когда n случайный момент остановки, причем эти результаты будут асимптотическими при $\varepsilon > 0$ и $\varepsilon \to 0$.

Для нахождения этого момента остановки введем условие:
(A_2) существуют такие постоянные $0 < m \leq 1$ и $0 < \beta_0(F) < \infty$, $F \in \mathbf{F}$, что $n^m \beta_n(F) \to \beta_0(F)$ при $n \to \infty$.

Если $\varphi(y) = y$, то $m = \frac{1}{2}$ и если $\varphi(y) = y^2$, то $m = 1$.

Из этого условия получаем соотношение $R_n(\varepsilon, F) = \beta_n(F) + \varepsilon \cdot n \cong n^{-m}\beta_0(F) + \varepsilon \cdot n$ при $n \to \infty$, здесь символ $a_n \cong b_n$ означает асимптотическое равенство, то есть $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 1$.

Через $R_n^*(\varepsilon, F) = n^{-m}\beta_0(F) + \varepsilon \cdot n$ обозначим функцию риска, асимптотически равную функции риска $R_n(\varepsilon, F)$.

Определение 4. Если для оценки $\theta_n(F) = \theta_n(\xi_1, \xi_2, ..., \xi_n)$ существует такое значение $n_0 = n_0(\varepsilon, F)$, что выполняется следующее соотношение

$$R_{n_0}^*(\varepsilon, F) = \min_{n\geq 1} R_n^*(\varepsilon, F) \qquad (2),$$

то $R_n(\varepsilon, F)$ называется асимптотически минимальной функцией риска, величина $n_0 = n_0(\varepsilon, F)$ называется асимптотически оптимальным объёмом выборки.

Лемма. *Если выполнены условия* (A_1) *и* (A_2), *то асимптотически оптимальный объём выборки равен* $n_0 = n_0(\varepsilon, F) = \inf\left(n \geq 1 : n \geq \left(\frac{m\beta_0(F)}{\varepsilon}\right)^{\frac{1}{m+1}}\right)$ *и минимальное значение функции риска* $R_n^*(\varepsilon, F) = n^{-m}\beta_0(F) + \varepsilon \cdot n$ *равен* $R_{n_0}^*(\varepsilon, F) = \left(\varepsilon^m \left(m + m^{-m}\right)\beta_0(F)\right)^{\frac{1}{1+m}}$.

Оптимальный объём выборки равен $n_0 = n_0(\varepsilon, F)$ и минимальное значение функции риска $R_{n_0(\varepsilon,F)}^*(\varepsilon, F)$ обладают тем недостатком, что они определяется функционалом $\beta_0(F)$ от неизвестной функции распределения $F(x)$, поэтому необходимо оценить $\beta_0(F)$. Предположим, что существует состоятельная положительная оценка $b_n = b(\xi_1, \xi_2, ..., \xi_n)$ для $\beta_0(F)$, $F \in \mathbf{F}$, $n \geq 1$, такая, что выполняется условие: (A_3) существуют число $\alpha > m + \gamma$ и целое число n_0 такие, что для любого $\delta{>}0$ и $n \geq n_0$ $P\{|b_n - \beta_0(F)| \geq \delta\} \leq \frac{c(\alpha,\delta)}{n^{1+\alpha}}$, где $0 < c(\alpha, \delta) < \infty$ постоянная.

Исходя из вида асимптотически оптимального объёма выборки $n_0 = n_0(\varepsilon, F)$, определим момент остановки $N(\varepsilon)$ следующим образом:

$$N(\varepsilon) = \inf\left(n \geq 1 : n \geq \left(\frac{m(b_n + n^{-\gamma}}{\varepsilon}\right)^{\frac{1}{m+1}}\right) = \inf\left(n \geq 1 : n^{m+1} \geq \frac{m(b_n + n^{-\gamma}}{\varepsilon}\right).$$

Теорема 1. *Если выполнены условия* (A_1),(A_2) *и* (A_3), *то справедливы следующие утверждения:*

1) Для любого $\varepsilon > 0$ $P\{N(\varepsilon) < \infty\} = 1$;
2) $N(\varepsilon) \to \infty$ с вероятностью 1 при $\varepsilon \to 0$;
3) $\lim\limits_{\varepsilon\to 0} E(N(\varepsilon)) = \infty$;
4) $\frac{N(\varepsilon)}{n_0(\varepsilon,F)} \to 1$ с вероятностью 1 при $\varepsilon \to 0$.

Теорема 2. *Если выполнены условия* (A_1),(A_2), (A_3) *и для некоторого* $\varepsilon_0 > 0$ *ряд* $\sum_{m=1}^{\infty} \sup\limits_{0\le\varepsilon\le\varepsilon_0} P\left\{\frac{N(\varepsilon)}{n_0(\varepsilon,F)} \ge m\right\}$ *сходится, тогда* $\lim\limits_{\varepsilon\to 0} \frac{E(N(\varepsilon))}{n_0(\varepsilon,F)} = 1$.

3. Пример. В этом разделе проверим выполнение условий регулярности (A_1),(A_2) и (A_3) в частном случае, когда функционал $\theta(F)$ от неизвестной функции распределения $F(x)$ равен самой функции распределения, то есть $\theta(F) = F(x)$.

Пусть $\xi_1, \xi_2, ..., \xi_n$ независимые случайные величины с неизвестной функции распределения $F(x), x \in R_1$. Оценим её эмпирической функцией распределения $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\xi_i < x)$, где $I(A)$ индикатор события A.

В качестве функции потерь оценивания неизвестной функции распределения $F(x)$ эмпирической функцией распределения $F_n(x)$ примем следующую квадратичную функцию потерь:

$$L_n = L_n(\varepsilon, F) = (F_n(x) - F(x))^2 + \varepsilon \cdot n,$$

где ε стоимость одного элемента выборки.

Математическое ожидание функции потерь, то есть функция риска оценивания неизвестной функции распределения $F(x)$ статистикой $F_n(x)$ равно следующему $R_n(\varepsilon, F) = M(L_n(\varepsilon, F) = M(F_n(x) - F(x))^2 + \varepsilon \cdot n$.

Поскольку функция $\varphi(y) = y^2$, $\alpha_n = |F_n(x) - F(x)|$ и $\beta_n(F)) = \frac{F(x)(1-F(x))}{n}$, то $\lim\limits_{n\to\infty}(n\beta_n(F)) = \lim\limits_{n\to\infty}\left(n\frac{F(x)(1-F(x))}{n}\right) = \lim\limits_{n\to\infty}(F(x)(1 - F(x))) = F(x)(1 - F(x)) = \beta_0(F)$ и условия (A_1), (A_2) выполняются при $m = 1$ и $\beta_0(F) = F(x)(1 - F(x))$.

Минимум по n функции риска $R_n(\varepsilon, F) = \frac{F(x)(1-F(x))}{n} + \varepsilon \cdot n$ достигается при значении n равном $n_0 = n_0(\varepsilon, F) = \left[\left(\frac{F(x)(1-F(x))}{\varepsilon}\right)^{\frac{1}{2}}\right] = \inf\left(n \ge 1 : n \ge \left(\frac{F(x)(1-F(x))}{\varepsilon}\right)^{\frac{1}{2}}\right)$ и минимальное значение функции риска равно $R_{n_0}(\varepsilon, F) = \min\limits_{n\ge 1} R_n(\varepsilon, F) = 2\left(\varepsilon F(x)(1 - F(x))\right)^{\frac{1}{2}}$.

Функция $R_{n_0}(\varepsilon, F)$ является минимальной функцией риска, а значение $n_0 = n_0(\varepsilon, F)$ является оптимальным объёмом выборки. Поскольку оптимальный объём выборки $n_0 = n_0(\varepsilon, F)$ и минимальное значение функции риска $R_{n_0}(\varepsilon, F)$ зависят от неизвестного $\beta_0(F) = F(x)(1 - F(x))$,

то необходимо оценить $\beta_0(F)$. В качестве оценки неизвестного $\beta_0(F) = F(x)(1-F(x))$ возмем статистику $b_n = F_n(x)(1-F_n(x))$, $n \geq 1$ и покажем выполнение условия (A_3). Для этого преобразуя, получим неравенство

$$|b_n - \beta_0(F)| = |F_n(x)(1-F_n(x)) - F(x)(1-F(x))| =$$

$$= \left|(F_n(x) - F(x)) - (F_n^2(x) - F^2(x))\right| =$$

$$= |(F_n(x) - F(x))(1 - F_n(x) - F(x))| \leq |F_n(x) - F(x)| \leq \sup_{-\infty<x<\infty} |F_n(x) - F(x)|.$$

Воспользуемся следующим неравенством Дворецкого-Кифера-Вольфовича: для любого числа $y > 0$ существует такое постоянное число c, что справедливо следующее неравенство $P\{\sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)| \geq \frac{y}{\sqrt{n}}\} \leq ce^{-2y^2}$.

Используя эти две неравенства, для произвольного числа $\alpha > 0$ получим

$$P\{|b_n - \beta_0(F)| \geq \delta\} = P\{|F_n(x)(1-F_n(x)) - F(x)(1-F(x))| \geq \delta\} \leq$$

$$\leq P\left\{\sup_{0<x<\infty} |F_n(x) - F(x)| \geq \delta\right\} = P\left\{\sup_{0<x<\infty} |F_n(x) - F(x)| \geq \frac{\delta\sqrt{n}}{\sqrt{n}}\right\} \leq$$

$$\leq ce^{2n\delta^2} < \frac{c}{n^{1+\alpha}}$$

То есть условие (A_3) выполняется.

Исходя из вида оптимального объёма выборки $n_0 = n_0(\varepsilon, F)$, момент остановки $N(\varepsilon)$ введем следующим образом:

$$N(\varepsilon) = \left[\left(\frac{F_n(x)(1-F_n(x))}{\varepsilon}\right)^{\frac{1}{2}}\right] = \inf\left(n \geq 1 : n \geq \left(\frac{F_n(x)(1-F_n(x))}{\varepsilon}\right)^{\frac{1}{2}}\right).$$

Свойства этого момента остановки приведены в следующей теореме.

Теорема 3. *Справедливы следующие утверждения:*

1) Для любого $\varepsilon > 0$ $P\{N(\varepsilon) < \infty\} = 1$;
2) $N(\varepsilon) \to \infty$ с вероятностью 1 при $\varepsilon \to 0$;
3) $\lim\limits_{\varepsilon\to 0} E(N(\varepsilon)) = \infty$;
4) $\frac{N(\varepsilon)}{n_0(\varepsilon,F)} \to 1$ с вероятностью 1 при $\varepsilon \to 0$.
5) $\lim\limits_{\varepsilon\to 0} \frac{E(N(\varepsilon))}{n_0(\varepsilon,F)} = 1$.

ЛИТЕРАТУРА

1. Lehmann E. L. "Notes on the theory of estimation 1951, Berkely, Univ. of Calif. Press.

2. Ghosh M., Mukhopadhyay N., Sen P. K. "Sequential estimation New York, Wiley, 1996.

3. Sen P. K. "Sequential nonparametrics: Invariance principles and statistical inference New York, Wiley, 1981.

G.G.Rakhimova (Tashkent, Uzbekistan)

Asymptotic problems of sequential point estimation of functionals of an unknown distribution function

Abstract. *In this paper consider sequential point estimation of fairly wide class of functionals of an unknown distribution function. Regularity conditions are provided that provide the property of asymptotically minimality of risk function for a sufficiently wide class of estimates and loss functions. The asymptotic properties of the stopping moment are studied.*

Key words: *stopping moment, loss function, risk function, asymptotically minimality, asymptotic efficiency.*

Асимптотика степенных оценок для условных функций распределения при случайном цензурировании справа

Ф.А.Абдикаликов, П.Б.Насыров, Р.П.Сейтахов
Каракалпакский государственный университет
abdikalikovf@mail.ru

1. Непараметрическая регрессионная модель при случайном цензурировании справа. Пусть случайная величина (с.в.) Z, например время жизни пациента, находится под влиянием с.в. X, означающей дозу лекарства. В индустриальных опытах с.в. Z соответствует времени безотказной работы испытываемого технического устройства, подвергаемого давлению, напряжению, температуре или ещё каким-нибудь другим факторам, описываемым величиною X. Рассмотрим ситуацию, когда с.в. Z подвергается случайному цензурированию справа с.в. Y. Пусть $\left\{(Z_i, Y_i)\,,\ i=\overline{1,n}\right\}$- независимые реализации вектора (Z,Y) в n экспериментах и $0 = x_0 \leq x_1 \leq ... \leq x_n \leq 1$ - фиксированные точки дизайна ковариаты X. Наблюдается следующая выборка объема n : $S_1^{(n)} = \left\{(\xi_i, \delta_i, X_i)\,,\ i=\overline{1,n}\right\}$, где $\xi_i = \min(Z_i, Y_i)$ и $\delta_i = I(Z_i \leq Y_i)\,,\ i=\overline{1,n}$ индикаторы соответствующих событий. Пусть F_{x_i}, G_{x_i} и H_{x_i} означают соответственно условные функции распределения (ф.р.) с.в. Z_i, Y_i и ξ_i при заданном $X = x_i$. Тогда легко видеть, что $H_{x_i}(t) = 1 - (1 - F_{x_i}(t))(1 - G_{x_i}(t))$. Далее, для фиксированной точки дизайна $x \in [0,1]$, обозначим через F_x, G_x и H_x соответственно ф.р. с.в. Z_x, C_x и $\xi_x = \min(Z_x, C_{\ x})$ и пусть $\delta_x = I(Z_x \leq C_x)$. В данной модели регрессии задача состоит в оценивании условной ф.р. F_x интересующих нас с.в. Z_x по выборке $S_1^{(n)}$, в которой они наблюдаемы лишь при $\delta_x = 1$. Пусть ф.р. F_x является непрерывной. Для построения оценки для F_x введём вспомогательные субраспределения:

$$\begin{aligned}
&\tilde{H}_x(t) = P(\xi_x \leq t, \delta_x = 1) = \int_0^t (1 - G_x(u-))dF_x(u),\\
&\tilde{\tilde{H}}_x(t) = P(\xi_x \leq t, \delta_x = 0) = \int_0^t (1 - F_x(u))dG_x(u),\\
&\tilde{H}_x(t) + \tilde{\tilde{H}}_x(t) = H_x(t)\,.
\end{aligned} \tag{1}$$

Введём также и соответствующие интегральные функции интенсивности (и.ф.и.):

$$\begin{aligned}\tilde{\Lambda}_x(t) &= \int\limits_0^t \frac{dF_x(u)}{1-F_x(u)} = \int\limits_0^t \frac{d\tilde{H}_x(u)}{1-H_x(u-)},\\ \tilde{\tilde{\Lambda}}_x(t) &= \int\limits_0^t \frac{dG_x(u)}{1-G_x(u-)} = \int\limits_0^t \frac{d\tilde{\tilde{H}}_x(u)}{1-H_x(u-)},\end{aligned} \tag{2}$$

для которых справедливо равенство $\tilde{\Lambda}_x(t) + \tilde{\tilde{\Lambda}}_x(t) = \int\limits_0^t \frac{dH_x(u)}{1-H_x(u-)} = \Lambda_x(t)$, т.е. имеем и.ф.и., соответствующую ф.р. H_x. Построим следующие оценки типа Стоуна [5] для (1) формулами

$$\tilde{H}_{xh}(t) = \sum_{i=1}^n \omega_{ni}\left(x; h_n\right) I\left(\xi_i \le t, \delta_i = 1\right),$$

$$\tilde{\tilde{H}}_{xh}(t) = \sum_{i=1}^n \omega_{ni}\left(x; h_n\right) I\left(\xi_i \le t, \delta_i = 0\right),$$

$$H_{xh}(t) = \tilde{H}_{xh}(t) + \tilde{\tilde{H}}_{xh}(t) = \sum_{i=1}^n \omega_{ni}\left(x; h_n\right) I\left(\xi_i \le t\right),$$

где

$$\omega_{ni}\left(x; h_n\right) = C_n^{-1}\left(x; h_n\right) \int\limits_{x_{i-1}}^{x_i} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right) dy,\ i = \overline{1,n},$$

$$C_n\left(x; h_n\right) = \int\limits_0^{x_n} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right) dy,\ x \in [0,1].$$

Аналогично оценим и и.ф.и. (2) статистиками

$$\tilde{\Lambda}_{xh}(t) = \int\limits_0^t \frac{d\tilde{H}_{xh}(u)}{1-H_{xh}(u-)},\ \tilde{\tilde{\Lambda}}_{xh}(t) = \int\limits_0^t \frac{d\tilde{\tilde{H}}_{xh}(u)}{1-H_{xh}(u-)},$$

$$\Lambda_{xh}(t) = \tilde{\Lambda}_{xh}(t) + \tilde{\tilde{\Lambda}}_{xh}(t) = \int\limits_0^t \frac{dH_{xh}(u)}{1-H_{xh}(u-)}.$$

В [1] построена следующая оценка для F_x, являющаяся обобщённым аналогом степенной оценки из [4]

$$F_{xh}\left(t\right) = 1 - \left[1 - H_{xh}\left(t\right)\right]^{\mathbb{R}_{xh}(t)},\ t \ge 0, \tag{3}$$

где $\mathbb{R}_{xh}(t) = \tilde{\Lambda}_{xh}(t)\left(\Lambda_{xh}(t)\right)^{-1}$ - оценка для отношения интенсивностей $\mathbb{R}_x(t) = \tilde{\Lambda}_x(t)\left(\Lambda_x(t)\right)^{-1}$. Оценка (3) определена на всей прямой, имеет скачки во всех наблюдаемых вариантах $\xi_{(1)} \le \xi_{(2)} \le \ldots \le \xi_{(n)}$ и при

$\omega_{ni}(x;h_n) = 1/n$ совпадает со степенной оценкой [3,4], а при отсутствии цензурирования, т.е. при $\mathbb{R}_x(t) \equiv \mathbb{R}_{xh}(t) \equiv 1$ совпадает с оценкой Стоуна [5]. Далее нами будут использованы следующие обозначения:

(I) $\underline{\Delta}_n = \min\{(x_i - x_{i-1}), 1 \le i \le n\}$, $\overline{\Delta}_n = \max\{(x_i - x_{i-1}), 1 \le i \le n\}$.

(II) $\|\pi\|_2^2 = \int\limits_{-\infty}^{\infty} \pi^2(y)dy$, $m_1(\pi) = \int\limits_{-\infty}^{\infty} y\pi(y)dy$, $m_2(\pi) = \int\limits_{-\infty}^{\infty} y^2\pi(y)dy$.

Рассмотрим условия:

(У1) При $n \to \infty$, $x_n \to 1$, $\overline{\Delta}_n = O(n^{-1})$, $\overline{\Delta}_n - \underline{\Delta}_n = o(n^{-1})$;

(У2) Ядро π является плотностью с компактным носителем $[-M, M]$, при некотором $M > 0$, $m_1(\pi) = 0$ и π удовлетворяет условию Липщица: $|\pi(y) - \pi(y')| \le C_\pi |y - y'|$, $y, y' \in [-M, M]$;

(У3) Производные $\ddot{H}_x(t) = \frac{\partial^2}{\partial x^2}H_x(t)$, $H''_x(t) = \frac{\partial^2}{\partial t^2}H_x(t)$, $\dot{H}'_x(t) = \frac{\partial^2}{\partial x \partial t}H_x(t)$ Ц существуют и непрерывны для $(t,x) \in [\tau, T] \times [0,1]$, где $\sup\{t \ge 0 : H_x(t) = 0\} = \tau_{H_x} < \tau$, $T < T_{H_x} = \inf\{t \ge 0 : H_x(t) = 1\}$.

Переходим к асимптотическим свойствам оценки (3) при $n \to \infty$ (т.е.$h_n \to 0$). Имеет место утверждение об аппроксимации разницы $F_{xh} - F_x$ суммой случайных функций с оценкой скорости сходимости.

Теорема 1 [1]. Пусть выполнены условия (У1), (У2) и для функций H_x, $\tilde{H}_x$ условие (У3) в $[\tau, T]$, при $n \to \infty$, $h_n \to 0$, $\frac{\log n}{nh_n} = o(1)$, $\frac{nh_n^5}{\log n} = O(1)$. Тогда

$$F_{xh}(t) - F_x(t) = \sum_{i=1}^{n} \omega_{ni}(x;h_n)\Psi_{tx}(\xi_i, \delta_i) + R_n(t,x),$$

где

$$\begin{aligned}\Psi_{tx}(\xi_i,\delta_i) &= (1 - F_x(t))\left\{\int\limits_0^t \frac{I(\xi_i \le u) - H_x(u)}{(1-H_x(u))^2} d\tilde{H}_x(u) + \right. \\ &+ \left. \frac{I(\xi_i \le t, \delta_i = 1) - \tilde{H}_x(t)}{1 - H_x(t)} - \int\limits_0^t \frac{I(\xi_i \le u, \delta_i = 1) - \tilde{H}_x(u)}{(1-H_x(u))^2} dH_x(u)\right\},\end{aligned}$$

и при $n \to \infty$, $\sup\limits_{\tau \le t \le T} |R_n(t,x)| \overset{\text{п.н.}}{=} O\left((nh_n)^{-3/4}(\log n)^{3/4}\right)$.

Следствие 1. В условиях теоремы 1 следует равномерно сильная состоятельность оценки (3) при $n \to \infty$:

$$\sup_{\tau \le t \le T} |F_{xh}(t) - F_x(t)| \overset{\text{п.н.}}{=} O\left((nh_n)^{-1/2}(\log n)^{1/2}\right).$$

Асимптотическая нормальность оценок (3) содержится в следующем утверждении.

Теорема 2 [1]. Пусть выполнены условия (У1), (У2), а функции H_x и $\tilde{H}_x$ удовлетворяют условием (У3) при $\tau \le t \le T$. Тогда

(A) если $nh_n^5 = o(1)$ и $(nh_n)^{-1}(\log n)^3 = o(1)$ при $n \to \infty$, то

$$(nh_n)^{1/2}\left(F_{xh}(t) - F_x(t)\right) \overset{D}{\Rightarrow} N\left(0; \tilde{\sigma}_x^2(t)\right);$$

(B) если $h_n = Cn^{-1/5}$ при некотором $C > 0$, то при $n \to \infty$

$$(nh_n)^{1/2}\left(F_{xh}(t) - F_x(t)\right) \overset{D}{\Rightarrow} N\left(\tilde{a}_x(t); \tilde{\sigma}_x^2(t)\right),$$

где

$$\tilde{a}_x(t) = \frac{1}{2}\left(1 - F_x(t)\right)\left\{\int_0^t \frac{\ddot{H}_x(u)d\tilde{H}_x(u)}{\left(1 - H_x(u)\right)^2} + \int_0^t \frac{d\ddot{\tilde{H}}_x(u)}{1 - H_x(u)}\right\} m_2(\pi)C^{5/2},$$

$$\tilde{\sigma}_x^2(t) = \|\pi\|_2^2\left(1 - F_x(t)\right)^2 \int_0^t \frac{d\tilde{H}_x(u)}{\left(1 - H_x(u)\right)^2}.$$

Теперь сформулируем утверждение о слабой сходимости процесса $\left\{\mathbb{W}_{nx}(t) = (nh_n)^{1/2}(F_{xh}(t) - F_x(t))\right\}$ в пространстве Скорохода $\mathbb{D}[\tau, T]$ к гауссовским процессам. Рассмотрим $\mathbb{W}_x(t)$- гауссовский процесс с нулевым средним и ковариационной функцией при $s, t \in [\tau, T]$:

$$Q_{st} = \|\pi\|_2^2\left(1 - F_x(t)\right)\left(1 - F_x(s)\right) \int_0^{\min(t,s)} \frac{d\tilde{H}_x(u)}{\left(1 - H_x(u)\right)^2},$$

а также гауссовский процесс $\tilde{\mathbb{W}}_x(t)$ со средним $\tilde{a}_x(t)$ и ковариацией Q_{st}.

Теорема 3 [1]. Пусть выполнены условия теоремы 2. Тогда

$(A)'$ в условиях (A): $\mathbb{W}_{nx}(\cdot) \overset{D}{\Rightarrow} \mathbb{W}_x(\cdot)$ в $\mathbb{D}[\tau, T]$;

$(B)'$ в условиях (B): $\mathbb{W}_{nx}(\cdot) \overset{D}{\Rightarrow} \tilde{\mathbb{W}}_x(\cdot)$ в $\mathbb{D}[\tau, T]$.

2. Частично-информативная регрессионная модель случайного цензурирования справа. В этой модели наблюдается выборка $S_2^{(n)} = \left\{(\xi_i, \delta_i, X_i),\ i = \overline{1,n}\right\}$, где $\xi_i = \min(Z_i, Y_{1i}, Y_{2i})$ и индикаторная с.в.

$$\delta_i = \begin{cases} -1, & Y_{2i} \le \min(Z_i, Y_{1i}), \\ 0, & Y_{1i} \le \min(Z_i, Y_{2i}), \\ 1, & \mathrm{Z}_i \le \min(Y_{1i}, Y_{2i}), \end{cases}$$

и следовательно интересующая нас с.в. Z_i наблюдается лишь в случае $\delta_i = 1$. В выборке $S_2^{(n)}$ компоненты независимых векторов $(Z_1, Y_{11}, Y_{21}), ..., (Z_n, Y_{1n}, Y_{2n})$ условно независимы при заданном $X_i = x \in [0,1]$. Заметим, что в выборке $S_2^{(n)}$ с.в. Z_x цензурируются с.в. $\min(Y_{1x}, Y_{2x})$ справа и поэтому $H_x(t) = P(\xi_x \le t) = = 1 - (1 - K_x(t))(1 - G_{2x}(t))$, где $K_x(t) = P(\min(Z_x, Y_{1x}) \le t) = 1 - (1 - F_x(t))(1 - G_{1x}(t))$.

В данной модели цензурирование справа является частично-информативным, т.е. существует число $\beta_x > 0$ такое, что $1 - G_{1x}(t) = (1 - F_x(t))^{\beta_x}$, $t \ge 0$. В рассматриваемой модели оценку для F_x по выборке $S_2^{(n)}$ построим следующим образом [2]:

$$1 - F_{xh}^0(t) = [1 - H_{xh}(t)]^{\gamma_{xh}\mathbb{R}_{xh}^0(t)}, \ t \ge 0, \tag{4}$$

где $\gamma_{xh} = p_{xh}^{(1)}(p_{xh}^{(1)} + p_{xh}^{(0)})^{-1}$, $p_{xh}^{(m)} = \sum\limits_{i=1}^{n} \omega_{ni}(x; h_n) I(\delta_i = m)$, $m = -1, 0, 1$; $\mathbb{R}_{xh}^0(t) = \tilde{\Lambda}_{xh}(t)(\Lambda_{xh}(t))^{-1}$, $\Lambda_{xh}(t) = \tilde{\Lambda}_{xh}(t) + \tilde{\tilde{\Lambda}}_{xh}(t)$, $H_{xh}(t) = \tilde{H}_{xh}(t) + \tilde{\tilde{H}}_{xh}(t)$, $\tilde{H}_{xh}(t) = \sum\limits_{i=1}^{n} \omega_{ni}(x; h_n)(\xi_i \le t, \delta_i \ne -1)$, $\tilde{\tilde{H}}_{xh}(t) = \sum\limits_{i=1}^{n} \omega_{ni}(x; h_n) I(\xi_i \le t, \delta_i = -1)$.

Пусть $\tilde{H}_x^{(1)}(t) = P(\xi_x \le t, \delta_x = 1)$, $\tilde{H}_x^{(0)}(t) = P(\xi_x \le t, \delta_x = 0)$. Тогда $\tilde{H}_x(t) = P(\xi_x \le t, \delta_x \ne -1) = \tilde{H}_x^{(0)}(t) + \tilde{H}_x^{(1)}(t)$. В формулу оценки (4) входит статистика γ_{xh}, являющаяся оценкой $\gamma_x = (\beta_x + 1)^{-1} = p_x^{(1)}(p_x^{(1)} + p_x^{(0)})^{-1}$, $p_x^{(0)} = \tilde{H}_x^{(0)}(\infty) = \beta_x p_x^{(1)}$,$p_x^{(1)} = \tilde{H}_x^{(1)}(\infty) = \int\limits_0^\infty (1 - F_x(u))^{\beta_x}(1 - G_{2x}(u-))dF_x(u)$, $p_x^{(0)} + p_x^{(1)} = 1 - p_x^{(-1)}$, $p_x^{(-1)} = \tilde{\tilde{H}}_x(\infty)$.

В этом моделей наряду с (У1), (У2), также рассматриваются условия

(У3)' $\ddot{H}_x(t)$, $\ddot{\tilde{H}}_x(t)$, $H''_x(t)$, $\tilde{H}''_x(t)$, $\dot{H}'_x(t)$, $\dot{\tilde{H}}'_x(t)$ существуют и непрерывны для $(t, x) \in [\tau, T] \times [0, 1]$;

(У4)' $\ddot{p}_x^{(m)}$, $m = 0, 1$, существуют и непрерывны для $x \in [0, 1]$.

Теорема 4 [2]. Пусть выполнены условия (У1), (У2), а для функций H_x, $\tilde{H}_x$, $p_x^{(m)}$, $m = 0, 1$, условия (У3)', (У4)' и при $n \to \infty$, $h_n \to 0$, $\frac{\log n}{nh_n} = o(1)$, $\frac{nh_n^5}{\log n} = O(1)$. Тогда для $\tau \le t \le T$:

$$F_{xh}^0(t) - F_x(t) = \sum_{i=1}^{n} \omega_{ni}(x; h_n)\Psi_{tx}^0(\xi_i, \delta_i) + R_n^0(t, x),$$

где

$$\Psi_{tx}^{0}\left(\xi_i,\delta_i\right)=\gamma_x(1-F_x(t))\left\{\int\limits_0^t\frac{(I(\xi_i\le u)-H_x(u))d\tilde{H}_x(u)}{\left(1-H_x(u)\right)^2}+\right.$$

$$+\frac{(I(\xi_i\le t,\delta_i\ne -1)-\tilde{H}_x(t))}{1-H_x(t)}-\int\limits_0^t\frac{(I(\xi_i\le u,\delta_i\ne -1)-\tilde{H}_x(u))dH_x(u)}{\left(1-H_x(u)\right)^2}-$$

$$\left.-\frac{1}{\left(p_x^{(1)}\right)^2}\log(1-F_x(t))\left[p_x^{(0)}I(\delta_i=1)-p_x^{(1)}I(\delta_i=0)\right]\right\},$$

и при $n\to\infty$, $\sup\limits_{\tau\le t\le T}\left|R_n^0(t,x)\right|\overset{\text{п.н.}}{=}O\left((nh_n)^{-3/4}(\log n)^{3/4}\right)$.

Следствие 2. В условиях теоремы 4 следует равномерно сильная состоятельность оценки (4) при $n\to\infty$:

$$\sup_{\tau\le t\le T}\left|F_{xh}(t)-F_x(t)\right|\overset{\text{п.н.}}{=}O\left((nh_n)^{-1/2}(\log n)^{1/2}\right).$$

Теорема 5 [2]. Пусть выполнены условия (У1), (У2), а функции H_x,$\tilde{H}_x$ и $p_x^{(m)}$, $m=0,1$, удовлетворяют условиям (У3)', (У4)' при $\tau\le t\le T$. Тогда

(A) если $nh_n^5=o(1)$ и $(nh_n)^{-1}(\log n)^3=o(1)$ при $n\to\infty$, то

$$(nh_n)^{1/2}\left(F_{xh}^0(t)-F_x(t)\right)\overset{D}{\Rightarrow}N\left(0;\sigma_x^{*^2}(t)\right);$$

(B) если $h_n=Cn^{-1/5}$ при некотором $C>0$, то при $n\to\infty$

$$(nh_n)^{1/2}\left(F_{xh}^0(t)-F_x(t)\right)\overset{D}{\Rightarrow}N\left(a_x^*(t);\sigma_x^{*^2}(t)\right),$$

где

$$a_x^*(t)=\frac{1}{2}\gamma_x\left(1-F_x(t)\right)\left\{\int\limits_0^t\frac{\ddot{H}_x(u)d\tilde{H}_x(u)}{\left(1-H_x(u)\right)^2}+\int\limits_0^t\frac{d\ddot{\tilde{H}}_x(u)}{1-H_x(u)}-\right.$$

$$\left.-\frac{1}{\left(p_x^{(1)}\right)^2}\log(1-F_x(t))\left[p_x^{(0)}\ddot{p}_x^{(1)}-p_x^{(1)}\ddot{p}_x^{(0)}\right]\right\}m_2(\pi)C^{5/2},$$

$$\sigma_x^{*^2}(t)=\|\pi\|_2^2\,\gamma_x^2(1-F_x(t))^2\left\{\int\limits_0^t\frac{d\tilde{H}_x(u)}{\left(1-H_x(u)\right)^2}+\frac{p_x^{(0)}(p_x^{(0)}+p_x^{(1)})}{\left(p_x^{(1)}\right)^3}\log^2(1-F_x(t))\right\}.$$

Рассмотрим гауссовский процесс $\mathbb{W}_x^0(t)$ с нулевым средним и ковариационной функцией при $s,t \in [\tau,T]$:

$$Q_{st}^0 = \|\pi\|_2^2 \gamma_x^2 (1-F_x(t))(1-F_x(s)) \left\{ \int\limits_0^{\min(t,s)} \frac{d\tilde{H}_x(u)}{(1-H_x(u))^2} + \right.$$

$$\left. + \frac{p_x^{(0)}(p_x^{(0)}+p_x^{(1)})}{(p_x^{(1)})^3} \log(1-F_x(t)) \log(1-F_x(s)) \right\},$$

а также гауссовский процесс $\tilde{\mathbb{W}}_x^0(t)$ со средним $a_x^*(t)$ и ковариацией Q_{st}^0. Пусть $\{\mathbb{W}_{nx}^0(t) = (nh_n)^{1/2}(F_{xh}^0(t) - F_x(t)),\ \tau \le t \le T\}$.

Теорема 6 [2]. Пусть выполнены условия теоремы 5. Тогда

$(A)'$ в условиях (A): $\mathbb{W}_{nx}^0(\cdot) \overset{D}{\Rightarrow} \mathbb{W}_x^0(\cdot)$ в $\mathbb{D}[\tau,T]$;

$(B)'$ в условиях (B): $\mathbb{W}_{nx}^0(\cdot) \overset{D}{\Rightarrow} \tilde{\mathbb{W}}_x^0(\cdot)$ в $\mathbb{D}[\tau,T]$.

Замечание. Следует отметить, что использование оценки (4), учитывающую информативность случайного цензурирования справа вместо оценки (3) является оправданным. Оценка (4) является асимптотически эффективной по сравнению с оценкой (3) так как имеет место неравенство $\|\pi\|_2^2 \gamma_x (1-F_x(t))^2 \log^2(1-K_x(t)) < \sigma_x^{*^2}(t) < \tilde{\sigma}_x^2(t), \tau \le t \le T$.

ЛИТЕРАТУРА

1. Абдушукуров А.А., Абдикаликов Ф.А. Оценивание условной функции распределения при непараметрической регрессии с фиксированным дизайном. Сб. Статистические методы оценивания и проверка гипотез. Пермь. Пермский госуниверситет. - 2011. вып. 23. - С. 181-189.

2. Абдушукуров А.А., Абдикаликов Ф.А. Полупараметрическое оценивание условной функции выживания в частично- информативной регрессионной модели случайного цензурирования справа. Уз.Мат. журнал. - 2012. - No 3. - С. 9-21.

3. Абдушукуров А.А. Оценки неизвестных распределений по неполным наблюдениям и их свойства. LAMBERT Academic Publishing. - 2011.

4. Abdushukurov A.A. Nonparametric estimation of the distribution function based on relative risk function. Commun. Statist.: Th.and Meth. -

1998. - Vol. 27, - No.8, - P.1991-2012.

5. Stone C.J. Consistent nonparametric regression. Ann. Statist. - 1977. - Vol. 5. - P.595-645.

Свойства оценок условных функций распределения при случайном цензурировании наблюдений с двух сторон

Ф.А.Абдикаликов, И.Ж.Турсынбекова, Л.Ж.Реимова
Каракалпакский государственный университет
abdikalikovf@mail.ru

1. Модель непараметрической регрессии при случайном цензурировании с двух сторон. Пусть случайная величина (с.в.) Z находится под влиянием другой с.в. X. В медико-биологических исследованиях Z соответствует времени жизни индивидума, X - дозе лекарства, а в индустриальных опытах эти с.в. соответствуют времени безотказной работы технического устройства и напряжению (давлению и температуре), которому подвергается это устройство. На практике, как правило с.в. Z полностью ненаблюдаема и может подвергаться случайному цензурированию с двух сторон с.в. L и Y. Предполагается, что с.в. $\{L, Z, Y\}$ - независимы в совокупности при заданной ковариате $X = x$. Пусть $\left\{(L_k, Z_k, Y_k), k = \overline{1,n}\right\}$ - независимые реализации с.в. $\{L, Z, Y\}$. Схема цензурирования такова, что в n-м шаге эксперимента наблюдается выборка $S^{(n)} = \left\{\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}, X_i\right), i = \overline{1,n}\right\}$, где $\xi_i = \max\{L_i, \min(Z_i, Y_i)\}$, $\chi_i^{(0)} = I(\min(Z_i, Y_i) < L_i)$, $\chi_i^{(1)} = I(L_i \le Z_i \le Y_i)$, $\chi_i^{(2)} = I(L_i \le Y_i < Z_i)$. Пусть F_{x_i}, G_{x_i}, K_{x_i} и H_{x_i} - условные функций распределения (ф.р.) с.в. Z_i, Y_i, L_i и ξ_i при заданной ковариате $X_i = x_i$ являются непрерывными. Тогда легко видеть, что $H_{x_i}(t) = K_{x_i}(t)\left[1 - (1 - G_{x_i}(t))(1 - F_{x_i}(t))\right]$, $t \ge 0$. Рассмотрим случай, когда x_i являются фиксированными точками дизайна из $[0,1] : 0 = x_0 \le x_1 \le ... \le x_n \le 1$. Далее для упрощения записей x_i обозначим через x. В рассматриваемой модели интересующие нас с.в. Z_x наблюдаемы только при $\chi_x^{(1)} = 1$ и задача состоит в оценивании условной функции выживания $1 - F_{\tau x}(t) = P(Z_x > t / Z_x \ge \tau)$, $t \ge \tau \ge 0$ по выборке $S^{(n)}$. Для построения оценки для $1 - F_{\tau x}$ введём вспомогательные субраспределения:

$$T_x^{(0)}(t) = P(Z_x \wedge Y_x < L_x, L_x \le t) = \int_0^t [1 - (1 - G_x(u))(1 - F_x(u))] dK_x(u),$$

$$T_x^{(1)}(t) = P(L_x \le Z_x \le Y_x, Z_x \le t) = \int_0^t K_x(u)(1 - G_x(u))\, dF_x(u),$$

$$T_x^{(2)}(t) = P(L_x \le Y_x < Z_x, Y_x \le t) = \int_0^t K_x(u)(1 - F_x(u))\, dG_x(u).$$

Легко проверить, что $T_x^{(0)}(t)+T_x^{(1)}(t)+T_x^{(2)}(t)=H_x(t),\ t\geq 0$. Рассмотрим вероятность $q_x(t)=P(L_x\leq t\leq Z_x\wedge Y_x)=K_x(t)-H_x(t),\ t\geq 0$, а также интегральные функции интенсивности (и.ф.и.) для ф.р. $F_{\tau x}$:

$$\Lambda_{\tau x}^{(1)}(t)=\int\limits_{\tau}^{t}\frac{dF_{\tau x}(u)}{1-F_{\tau x}(u)}=\int\limits_{\tau}^{t}\frac{dF_x(u)}{1-F_x(u)}=\int\limits_{\tau}^{t}\frac{dT_x^{(1)}(u)}{q_x(u)},\ \ t\geq\tau\geq 0.$$

Введём также левосторонюю и.ф.и. для ф.р. K_x:

$$\Lambda_{\tau x}^{(0)}(t)=\int\limits_{t}^{+\infty}\frac{dK_x(u)}{K_x(u)}=\int\limits_{t}^{+\infty}\frac{dT_x^{(0)}(u)}{H_x(u)},\ \ t\geq\tau>0.$$

Число $\tau=\tau(K_x,G_x,F_x)$ выбирается из условий:

$$\begin{cases}\inf\limits_{t\in[\tau,+\infty]\cap Sp(N_x)}\{K_x(t)(1-N_x(t))\}>0,\\ \Gamma_{\tau x}^{(0)}\cap\Gamma_{\tau x}^{(1)}\neq\varnothing,\end{cases}$$

где

$$N_x(t)=P(\min(Z_x,Y_x)\leq t)=1-(1-G_x(t))(1-F_x(t)),$$

$$Sp(N_x)=\{t\geq 0:0<N_x(t)<1\},$$

$$\Gamma_{\tau x}^{(0)}=\left\{t\geq\tau:0<\Lambda_{\tau x}^{(0)}(t)<\infty\right\},\ \ \Gamma_{\tau x}^{(1)}=\left\{t\geq\tau:0<\Lambda_{\tau x}^{(1)}(t)<\infty\right\}.$$

Для субраспределений $T_x^{(m)},\ m=0,1,2$; ф.р. H_x и вероятности q_x рассмотрим следующую оценку Стоуна [4] с весами типа Гессера-Мюллера [5]:

$$T_{xh}^{(m)}(t)=\sum_{i=1}^{n}\omega_{ni}(x;h_n)\,I\left(\xi_i\leq t,\chi_i^{(m)}=1\right),\ \ m=0,1,2,$$

$$H_{xh}(t)=T_{xh}^{(0)}(t)+T_{xh}^{(1)}(t)+T_{xh}^{(2)}(t)=\sum_{i=1}^{n}\omega_{ni}(x;h_n)\,I(\xi_i\leq t),$$

$$q_{xh}(t)=K_{xh}(t)-H_{xh}(t),\ \ K_{xh}(t)=\exp\left\{\int\limits_{t}^{+\infty}\frac{dT_{xh}^{(0)}(u)}{H_{xh}(u)}\right\},\ \ t\geq\tau,$$

где $\omega_{ni}(x;h_n)=\frac{1}{C_n(x;h_n)}\int\limits_{x_{i-1}}^{x_i}\frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy,\ i=1,...,n,\ C_n(x;h_n)=\int\limits_{0}^{x_n}\frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy.$

Здесь π известная функция плотности (ядро) и $\{h_n,\, n \geq 1\}$ - последовательность положительных чисел такая, что $h_n \to 0$ при $n \to \infty$. Аналогично, оценкой и.ф.и. $\Lambda_{\tau x}^{(1)}$ является

$$\Lambda_{\tau xh}^{(1)}(t) = \int\limits_{\tau}^{t} \frac{dT_{xh}^{(1)}(u)}{q_{xh}(u)},\ t \geq \tau.$$

В [1] построена следующая оценка для $1 - F_{\tau x}$, являющаяся обобщчнным аналогом степенной оценки Абдушукурова из [2] на случай регрессионной модели:

$$1 - F_{\tau xh}(t) = \left[\frac{q_{xh}(t)}{q_{xh}(\tau)}\right]^{\mathbb{R}_{\tau xh}(t)},\ t \geq \tau, \tag{1}$$

где $\mathbb{R}_{\tau xh}(t) = \Lambda_{\tau xh}^{(1)}(t)\left[-\int\limits_{\tau}^{t} \frac{dq_{xh}(u)}{q_{xh}(u)}\right]^{-1}$ - соответствующая оценка функции $\mathbb{R}_{\tau x}(t) = \Lambda_{\tau x}^{(1)}(t)\left[-\int\limits_{\tau}^{t} \frac{dq_{x}(u)}{q_{x}(u)}\right]^{-1},\ t \geq \tau.$

Далее нами будут использованы следующие обозначения:

(I) $\underline{\Delta}_n = \min\left\{(x_i - x_{i-1}), 1 \leq i \leq n\right\}, \overline{\Delta}_n = \max\left\{(x_i - x_{i-1}), 1 \leq i \leq n\right\}.$

(II) $\|\pi\|_2^2 = \int\limits_{-\infty}^{\infty} \pi^2(y)dy,\ m_1(\pi) = \int\limits_{-\infty}^{\infty} y\pi(y)dy,\ m_2(\pi) = \int\limits_{-\infty}^{\infty} y^2\pi(y)dy.$

Рассмотрим условия:

(У1) При $n \to \infty$, $x_n \to 1$, $\overline{\Delta}_n = O(n^{-1})$, $\overline{\Delta}_n - \underline{\Delta}_n = o(n^{-1})$;

(У2) Ядро π является плотностью с компактным носителем $[-M, M]$, при некотором $M > 0$, $m_1(\pi) = 0$ и π удовлетворяет условию Липщица: $|\pi(y) - \pi(y')| \leq C_\pi |y - y'|$, $y, y' \in [-M, M]$;

(У3) Производные второго порядка $\ddot{H}_x(t) = \frac{\partial^2}{\partial x^2}H_x(t)$, $H''_x(t) = \frac{\partial^2}{\partial t^2}H_x(t)$, $\dot{H}'_x(t) = \frac{\partial^2}{\partial x \partial t}H_x(t)$, $\ddot{T}_x^{(m)}(t) = \frac{\partial^2}{\partial x^2}T_x^{(m)}(t)$, $T_x^{(m)''}(t) = \frac{\partial^2}{\partial t^2}T_x^{(m)}(t)$, $\dot{T}_x^{(m)'}(t) = \frac{\partial^2}{\partial x \partial t}T_x^{(m)}$, $m = 0, 1$, существуют и непрерывны для $(t, x) \in [\tau, T] \times [0, 1]$.

Теперь приведем асимптотические свойства оценки (1) при $n \to \infty$ (т.е. $h_n \to 0$).

Имеет место теорема об асимптотическом представлении оценок (1) в виде взвешенной суммы с.в. с оценкой остаточного члена.

Теорема 1 [1]. Пусть выполнены условия (У1), (У2), а для функций H_x, $T_x^{(0)}$, $T_x^{(1)}$ условие (У3) и при $n \to \infty$, $h_n \to 0$, $\frac{\log n}{nh_n} = o(1)$, $\frac{nh_n^5}{\log n} = O(1)$. Тогда для $\tau \leq t \leq T$:

$$F_{\tau xh}(t) - F_{\tau x}(t) = \sum_{i=1}^{n} \omega_{ni}(x; h_n)\, \Psi_{\tau tx}\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\right) + R_n(t, x),$$

где

$$\Psi_{\tau tx}\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\right) = \left(1 - F_{\tau x}\left(t\right)\right) \cdot$$

$$\cdot \left\{ \int\limits_{\tau}^{t} \frac{\left(\left(I\left(\xi_i \le u\right) - H_x\left(u\right)\right) - \lambda_{xh}\left(u\right)\right) dT_x^{(1)}\left(u\right)}{\left(K_x\left(u\right) - H_x\left(u\right)\right)^2} + \right.$$

$$+ \frac{\left(I\left(\xi_i \le t,\, \chi_i^{(1)} = 1\right) - T_x^{(1)}\left(t\right)\right)}{K_x\left(t\right) - H_x\left(t\right)} - \frac{I\left(\xi_i \le \tau,\, \chi_i^{(1)} = 1\right) - T_x^{(1)}\left(\tau\right)}{K_x\left(\tau\right) - H_x\left(\tau\right)} +$$

$$\left. + \int\limits_{\tau}^{t} \frac{\left(I\left(\xi_i \le u,\, \chi_i^{(1)} = 1\right) - T_x^{(1)}\left(u\right)\right) d\left(K_x\left(u\right) - H_x\left(u\right)\right)}{\left(K_x\left(u\right) - H_x\left(u\right)\right)^2} \right\},$$

$$\lambda_{xh}\left(t\right) = K_x\left(t\right) \sum_{i=1}^{n} \left\{ \int\limits_{t}^{+\infty} \frac{\left(I\left(\xi_i \le u\right) - H_x\left(u\right)\right) dT_x^{(0)}\left(u\right)}{H_x^2\left(u\right)} - \right.$$

$$\left. - \frac{I\left(\xi_i \le t,\, \chi_i^{(0)} = 1\right) - T_x^{(0)}\left(t\right)}{H_x\left(t\right)} - \int\limits_{t}^{+\infty} \frac{\left(I\left(\xi_i \le u,\, \chi_i^{(0)} = 1\right) - T_x^{(0)}\left(u\right)\right) dH_x\left(u\right)}{H_x^2\left(u\right)} \right\},$$

и при $n \to \infty$

$$\sup_{\tau \le t \le T} \left|R_n\left(t, x\right)\right| \overset{\text{п.н.}}{=} O\left(\left(nh_n\right)^{-3/4} \left(\log n\right)^{3/4}\right).$$

Следствие 1. В условиях теоремы 1 следует равномерно сильная состоятельность оценки (1) при $n \to \infty$:

$$\sup_{\tau \le t \le T} \left|F_{\tau xh}(t) - F_{\tau x}(t)\right| \overset{\text{п.н.}}{=} O\left(\left(nh_n\right)^{-1/2} \left(\log n\right)^{1/2}\right).$$

2. Модель регрессии при информативном случайном цензурировании с двух сторон. В этой модели ф.р. F_x, K_x и G_x являются непрерывными и удовлетворяют следующим степенным представлениям при всех $t \ge 0$:

$$\begin{cases} 1 - G_x\left(t\right) = \left(1 - F_x\left(t\right)\right)^{\theta_x}, \\ \\ K_x\left(t\right) = \left(N_x\left(t\right)\right)^{\beta_x}, \end{cases} \tag{2}$$

где θ_x и β_x – положительные неизвестные параметры. Рассматриваемая модель регрессии (2) характеризуется следующей теоремой.

Теорема 2 [3]. Равенства (2) имеют место тогда и только тогда, когда с.в. ξ_x и вектор $(\chi_x^{(0)}, \chi_x^{(1)}, \chi_x^{(2)})$ независимы при заданной ковариате $X = x$. В рассматриваемой модели оценку для $1 - F_x$ по выборке $S^{(n)}$ построим следующим образом [3]:

$$1 - \tilde{F}_{xh}(t) = \left\{1 - [H_{xh}(t)]^{\lambda_{xh}}\right\}^{\gamma_{xh}}, \quad t \geq 0, \tag{3}$$

где $\lambda_{xh} = 1 - p_{xh}^{(0)}$, $\gamma_{xh} = p_{xh}^{(1)}(1 - p_{xh}^{(0)})^{-1}$ и $p_{xh}^{(m)} = \sum\limits_{i=1}^{n} \omega_{ni}(x; h_n)\chi_i^{(m)}$, $m = 0, 1, 2$.

В данном моделе наряду с (У1), (У2), также рассматриваются условия

(У3)* $\dot{H}_x(t)$, $\ddot{H}_x(t)$ – существуют и непрерывны для $(t, x) \in [\tau, T] \times [0, 1]$;

(У4)* $\dot{\theta}_x$, $\dot{\beta}_x$, $\ddot{\theta}_x$, $\ddot{\beta}_x$ существуют и непрерывны для $x \in [0, 1]$. Обозначим $\left\|\dot{H}\right\| = \left\{\sup\left|\dot{H}_x(t)\right|,\ (t, x) \in [\tau, T] \times [0, 1]\right\}$, $\left\|\ddot{H}\right\| = \left\{\sup\left|\ddot{H}_x(t)\right|,\ (t, x) \in [\tau, T] \times [0, 1]\right\}$, $\left\|\dot{p}_x^{(m)}\right\| = \left\{\sup\left|\dot{p}_x^{(m)}\right|,\ x \in [0, 1]\right\}$, $\left\|\ddot{p}_x^{(m)}\right\| = \left\{\sup\left|\ddot{p}_x^{(m)}\right|,\ x \in [0, 1]\right\}$, $m = 0, 1; r^{-1} = \left\{\sup\left[(H_x(t))^{p_x^{(0)}} - H_x(t)\right]^{-1},\ t \in [\tau, T]\right\}$.

Справедлива теорема об экспоненциальной оценке и равномерной сильной состоятельности оценки (3).

Теорема 3 [3]. Пусть выполнены условия (У1), (У2), а для функций H_x, $p_x^{(0)}$, $p_x^{(1)}$ условия (У3)*, (У4)*, $r > 0$ и $h_n \to 0$.

(A) Для $\varepsilon > 0$ и достаточно большого n, пусть $\min\left(\frac{1}{p_x^{(0)}}, \frac{1}{1-p_x^{(0)}}, \frac{1}{p_x^{(1)}}\right) \geq \max\left(\frac{r}{r+8}, \frac{\varepsilon}{\varepsilon+3}\right) \geq \frac{\varepsilon r}{\varepsilon r+18} \geq \geq 4\max\left\{r^{-1}\left(\left\|\dot{H}\right\|\overline{\Delta}_n + \left\|\ddot{H}\right\| m_2(\pi) h_n^2\right), \left[r\left(1 - p_x^{(0)}\right)\right]^{-1} \cdot \left(\left\|\dot{p}_x^{(0)}\right\|\overline{\Delta}_n + \left\|\ddot{p}_x^{(0)}\right\| m_2(\pi) h_n^2\right), \left(p_x^{(1)}\right)^{-1}\left(\left\|\dot{p}_x^{(1)}\right\|\overline{\Delta}_n + \left\|\ddot{p}_x^{(1)}\right\| m_2(\pi) h_n^2\right)\right\}$.

Тогда

$$P\left(\sup_{\tau \leq t \leq T}\left|\tilde{F}_{xh}(t) - F_x(t)\right| > \varepsilon\right) \leq \frac{4C_1}{r}\left[\frac{(r+8)}{H_x(\tau)} e^{-C_2\left(\frac{rH_x(\tau)}{r+8}\right)^2 nh_n} + e^{-C_2\left(\frac{r}{4}\right)^2 nh_n} + \right.$$

$$\left. + \frac{3}{2\varepsilon} e^{-C_2\left(\frac{\varepsilon r}{6}\right)^2 nh_n}\right] + 4\left[e^{-C_0\left(\frac{rp_x^{(0)}}{r+8}\right)^2 nh_n} + e^{-C_0\left(\frac{\varepsilon r\left(1-p_x^{(0)}\right)}{\varepsilon r+18}\right)^2 nh_n} + e^{-C_0\left(\frac{\varepsilon p_x^{(1)}}{\varepsilon+3}\right)^2 nh_n}\right],$$

для некоторых абсолютных постоянных C_0, C_1 и C_2.

(B) Если при $n \to \infty$, $\frac{nh_n^5}{\log n} = O(1)$, то $\sup\limits_{\tau \le t \le T} \left| \tilde{F}_{xh}(t) - F_x(t) \right| \overset{\text{п.н.}}{=} O\left(\left(\frac{\log n}{nh_n} \right)^{1/2} \right)$.

Следующий результат состоит в асимптотическом представлении оценок (3) в виде взвешенной суммы с оценкой остаточного члена.

Теорема 4 [3]. Пусть выполнены условия (У1), (У2), а для функций H_x, $p_x^{(0)}$,$p_x^{(1)}$ условия (У3)*, (У4)*, $r > 0$ и при $n \to \infty$, $h_n \to 0$, $\frac{nh_n^5}{\log n} = O(1)$. Тогда для $t \in [\tau, T]$:

$$\tilde{F}_{xh}(t) - F_x(t) = \sum_{i=1}^{n} \omega_{ni}(x; h_n) \tilde{\Psi}_{tx}\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\right) + \tilde{q}_n(t, x), \quad (4)$$

где

$$\tilde{\Psi}_{tx}\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\right) = (1 - F_x(t)) \left\{ p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \cdot \right.$$

$$\cdot (I(\xi_i \le t) - H_x(t)) - \left[\frac{p_x^{(1)}}{\left(1 - p_x^{(0)}\right)^2} \log\left[1 - (H_x(t))^{1-p_x^{(0)}}\right] + \right.$$

$$\left. + \frac{p_x^{(1)}}{1 - p_x^{(0)}} H_x(t) \log H_x(t) \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \right] \left(\chi_i^{(0)} - p_x^{(0)} \right) -$$

$$\left. - \frac{1}{1 - p_x^{(0)}} \operatorname{lo} g\left[1 - (H_x(t))^{1-p_x^{(0)}}\right] \left(\chi_i^{(1)} - p_x^{(1)} \right) \right\},$$

и при $n \to \infty$, $\sup\limits_{\tau \le t \le T} |\tilde{q}_n(t, x)| \overset{\text{п.н.}}{=} O\left(\frac{\log n}{nh_n} \right)$. Следует отметить, что скорость аппроксимации в теореме 4 является оптимальной, чем в теореме 1.

Следствие 2. Из теоремы 4 при достаточно больших n следует, что $(nh_n)^{1/2} \left(\tilde{F}_{xh}(t) - F_x(t) \right)$ аппроксимируется суммой в (4) умноженной на $(nh_n)^{1/2}$ со скоростью $(nh_n)^{-1/2} \log n$.

С учетом этого установим асимптотическую нормальность оценок (3).

Теорема 5 [3]. Пусть выполнены условия (У1), (У2), а функций H_x, $p_x^{(0)}$, $p_x^{(1)}$ удовлетворяют условиям (У3)*, (У4)* и $r > 0$. Тогда

(A) если $nh_n^5 = o(1)$ и $(nh_n)^{-1/2} \log n = o(1)$, тогда при $n \to \infty$ и $t \in [\tau, T]$ $(nh_n)^{1/2} (\tilde{F}_{xh}(t) - F_x(t)) \overset{D}{\Rightarrow} N\left(0, \sigma_x^2(t)\right)$;

(B) если $h_n = Cn^{-1/5}$ при некотором $C > 0$, тогда при $n \to \infty$ и $t \in [\tau, T]$ $(nh_n)^{1/2} (\tilde{F}_{xh}(t) - F_x(t)) \overset{D}{\Rightarrow} N\left(a_x(t), \sigma_x^2(t)\right)$, где

$$a_x(t) = \frac{1}{2}(1 - F_x(t)) \left\{ p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \ddot{H}_x(t) - \right.$$

$$- \left[\frac{p_x^{(1)}}{\left(1-p_x^{(0)}\right)^2} \log \left[1-\left(H_x(t)\right)^{1-p_x^{(0)}}\right] + \right.$$

$$\left. + \frac{p_x^{(1)}}{1-p_x^{(0)}} H_x(t) \log H_x(t) \left[\left(H_x(t)\right)^{p_x^{(0)}} - H_x(t)\right]^{-1} \right] \ddot{p}_x^{(0)} -$$

$$- \frac{1}{1-p_x^{(0)}} \text{lo}\, g \left[1-\left(H_x(t)\right)^{1-p_x^{(0)}}\right] \ddot{p}_x^{(1)} \Big\} m_2(\pi) C^{5/2} ,$$

$$\sigma_x^2(t) = \|\pi\|_2^2 \left(1-F_x(t)\right)^2 \Big\{ A_x^2(t) H_x(t)(1-H_x(t)) + B_x^2(t) p_x^{(0)}(1-p_x^{(0)}) +$$

$$+ C_x^2(t) p_x^{(1)}(1-p_x^{(1)}) - 2B_x(t) C_x(t) p_x^{(0)} p_x^{(1)} \Big\} ,$$

Пусть $\{\tilde{\mathbb{W}}_{nx}(t) = (nh_n)^{1/2}(\tilde{F}_{xh}(t) - F_x(t)),\ \tau \le t \le T\}$ и $\tilde{\mathbb{W}}_x(t)$ – гауссовский процесс с нулевым средним и ковариационной функцией при $s,t \in [\tau,T]$: $\tilde{Q}_{st} = \|\pi\|_2^2 \left(1-F_x(s)\right)\left(1-F_x(t)\right)\{A_x(s)A_x(t)\left[H_x(min(s,t) - H_x(s)H_x(t)\right] + +B_x(s)B_x(t)p_x^{(0)}(1 - p_x^{(0)}) + C_x(s)C_x(t)p_x^{(1)}(1 - p_x^{(1)}) - 2B_x(min(s,t)C_x(min(s,t)\}$,где $A_x(t) = p_x^{(1)}\left[\left(H_x(t)\right)^{p_x^{(0)}} - H_x(t)\right]^{-1}$, $B_x(t) = -\left[\frac{p_x^{(1)}}{1-p_x^{(0)}} C_x(t) + \frac{A_x(t)}{1-p_x^{(0)}} H_x(t)\log H_x(t)\right]$, $C_x(t) = -\frac{1}{1-p_x^{(0)}} \log\left[1-\left(H_x(t)\right)^{1-p_x^{(0)}}\right]$, а также гауссовский процесс $\tilde{\mathbb{W}}_x^*(t)$ со средним $a_x(t)$ и ковариацией $\tilde{Q}_{st}$.

Теорема 6 [3]. Пусть выполнены условия Теоремы 5. Тогда

$(A)'$ в условиях (A): $\tilde{\mathbb{W}}_{nx}(\cdot) \overset{D}{\Rightarrow} \tilde{\mathbb{W}}_x(\cdot)$ в $\mathbb{D}[\tau,T]$;

$(B)'$ в условиях (B): $\tilde{\mathbb{W}}_{nx}(\cdot) \overset{D}{\Rightarrow} \tilde{\mathbb{W}}_x^*(\cdot)$ в $\mathbb{D}[\tau,T]$.

ЛИТЕРАТУРА

1. Абдушукуров А.А., Абдикаликов Ф.А. Непараметрическое оценивание условной функции выживания в регрессионной модели случайного цензурировании с двух сторон. // Труды меж. конф. XVI - ФАМЭБ, Красноярск. - 2012. - С. 27-29.

2. Абдушукуров А.А. Оценки неизвестных распределений по неполным наблюдениям и их свойства. LAMBERT Academic Publishing. - 2011.

3. Абдикаликов Ф.А., Абдушукуров А.А. Полупараметрическое оценивание условной функции выживания в информативной регрессионной модели случайного цензурирования с двух сторон. // Сб. Статистические методы

оценивания и проверка гипотез. - Пермь. Пермский госуниверситет. - 2012. вып. 24. -С. 145-162.

4. Stone C.J. Consistent nonparametric regression. //Ann. Statist. - 1977. - Vol. 5. - P.595-645.

5. Gasser T., Muller H.G. Kernel estimation of regression function. Lecture Notes in Mathematics. Springer. New York. - 1977.

On kernel estimates of distribution function and density in proportional hazard model

S.B.Kozimov
Moscow State University named after M.V.Lomonosov,
Tashkent Branch
kozimovsardorbegzodovich@mail.ru

Let X and Y be two independent random variables with continuous distribution functions F and G respectively. Suppose that pair (F, G) follows to proportional hazards model (PHM), that is there exist a positive number β such,that

$$1 - G(t) = (1 - F(t))^{\beta}, t \in R. \tag{1}$$

Note, that name of this model follows from proportionality of corresponding cumulative hazard functions $\Lambda_F = -log(1 - F)$ and $\Lambda_G = -log(1 - G)$, that is $\Lambda_G = \beta\Lambda_F(t)$. One of the important property of PHM is its characterization by independency of random variables Z=min(X,Y) and $\delta = I(X \leq Y)$, where I(.) is indicator function (for details, see [1-5]). Let $f = F'$- density function of random variable X. Consider statistical model of censoring of random variable X from the right by random variable Y, where by observations available a sample $C^{(n)} = \{(Z_i, \delta_i), i = 1, ..., n\}$, where $Z_i = min(X_i, Y_i)$, $\delta_i = I(X_i \leq Y_i)$ and (X_i, Y_i) - independent replicas of pair. According to (1), in samples $C^{(n)}$ subsamples $(Z_1, ..., Z_n)$ and $(\delta_1, ..., \delta_n)$ are independent. Hence empirical estimates $H_n(t)$ and p_n of $H(t) = P(Z \leq t)$ and $p = P(X \leq Y) = E\delta$ are also independent, where

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i \leqslant t), p_n = \frac{1}{n}\sum_{i=1}^{n} \delta_i. \tag{2}$$

It is easily seen that in PHM $H(t) = 1 - (1 - F(t))(1 - G(t)) = (1 - F(t))^{\beta+1}$ and $p = \frac{1}{\beta+1}$. Consequently, $1-F(t) = (1-H(t))^p$, $t \in R$. Using this expression of F(t) one can construct estimator for it as

$$F_n(t) = 1 - (1 - H_n(t))^{p_n}. \tag{3}$$

In statistical literature estimate (3) is called as semiparametric estimate of Abdushukurov-Cheng-Lin(ACL - estimator) and used by many authors(in details, see [4]).Estimate (3) submit a form

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} w_{ni} I(Z_i \leq t). \tag{4}$$

where

$$w_{ni} = n\left\{\left(1 - \frac{i-1}{n}\right)^{p_n} - \left(1 - \frac{i}{n}\right)^{p_n}\right\}$$

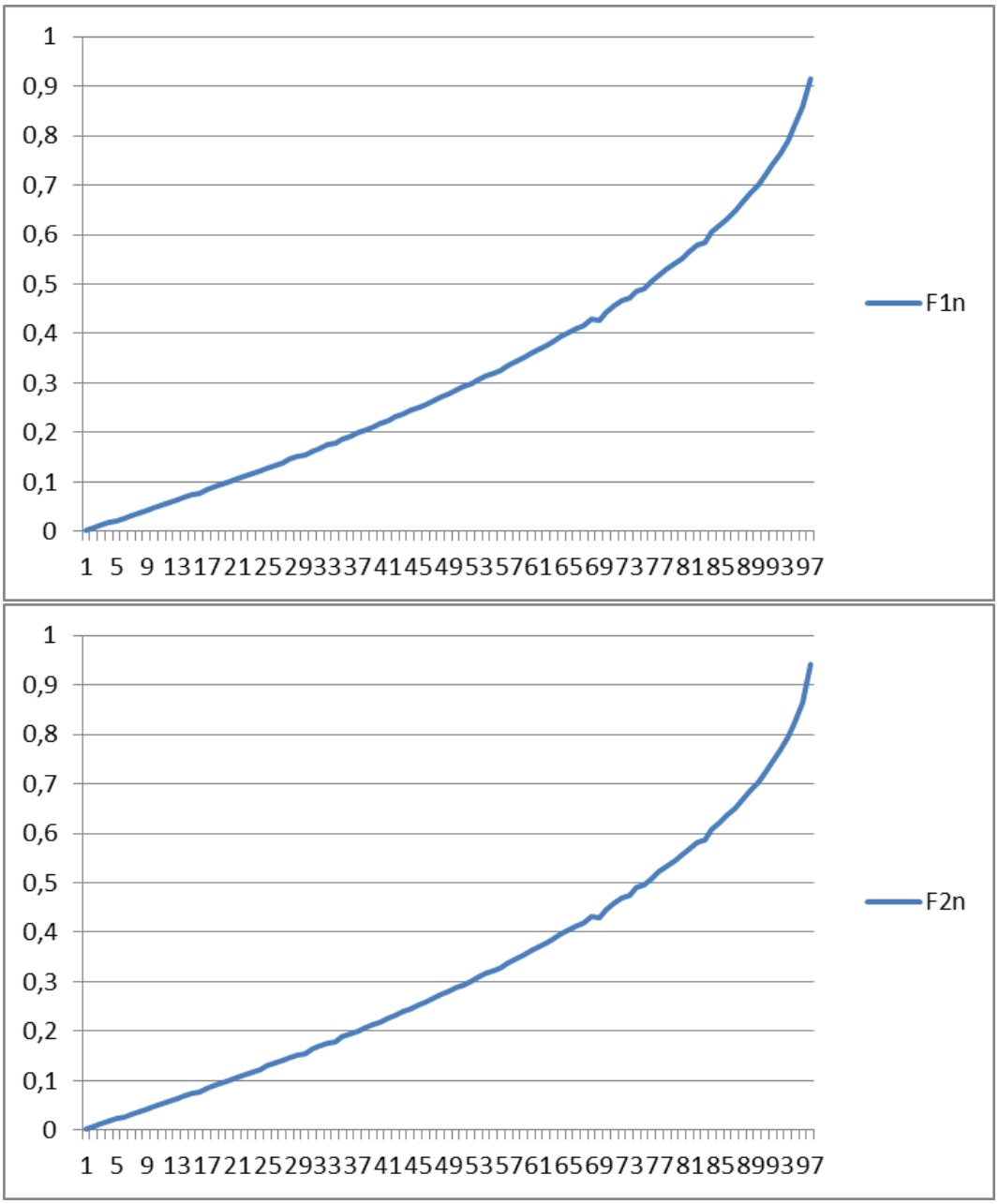
jumps of estimate (3) in points Z_{ni}, i=1,...,n and $Z_1 \leq ... \leq Z_n$ - order statistics from subsample $(Z_1, ..., Z_n)$. In this paper we consider following kernel estimates of distribution function F(t) and density f(t), $t \in R$, at actual drawing of kernel and "window width"sequence:

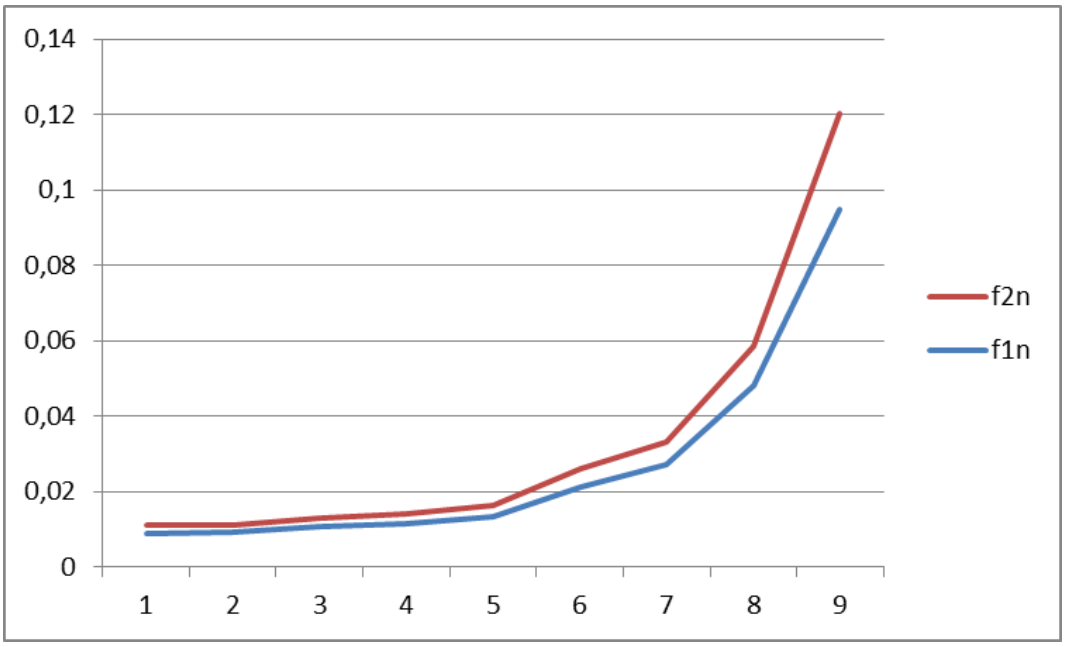
$$F_{1n}(t) = (1 - (1 - H_n^*(t)))^{p_n}, F_{2n}(t) = \int_{-\infty}^{\infty} K(\frac{t-u}{\mu(n)})dF_n(u),$$

$$f_{1n}(t) = \frac{dF_{1n}(t)}{dt} = p_n h_n(t)(1 - H_n^*(t))^{p_{n-1}},$$

$$f_{2n}(t) = \frac{dF_{2n}(t)}{dt} = \int_{-\infty}^{\infty} \frac{1}{\mu(n)} K\left(\frac{t-u}{\mu(n)}\right) dF_{2n}(u)$$

where kernel $k(t) = K'(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ - density of standard normal distribution,"window width"$\mu(n) = n^{-\frac{1}{5}}$,$h_n(t) = \frac{d}{dt}H_n^*(t)$ and $H_n^*(t) = \int_{-\infty}^{\infty} K(\frac{t-u}{\mu(n)})dH_n(u)$, $t \in R$,corresponding kernel estimate for distribution function H(t). Note that $F_{1n}(t)$, $F_{2n}(t)$ are smoothed estimates for F(t) and $H_n^*(t)$ is smoothed estimate for H(t) respectively. Using of this smoothed estimates for continuous distribution function F(t) and H(t) is more reasonable, than using of discrete estimates $F_n(t)$ and $H_n(t)$. It should be pointed out that estimates $F_{1n}(t)$ and $f_{1n}(t)$ proposed and investigated in works [2,3], whatever estimates $F_{2n}(t)$ and $f_{2n}(t)$ in [3-5]. Using data of size n=97 (Channing House Date) from [1], we build mentioned below plots of all of four estimates. From the plots we can see ,that for estimates $F_{1n}(t), F_{2n}(t)$ they almost entirely same, although according to the formulas they are significantly different.


1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1 5 9 13172125293337414549535761656973778185899397
F1n
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1 5 9 13172125293337414549535761656973778185899397
F2n



REFERENCE

1. Abdushukurov A.A. Statistics of incomplete observations. 2009. Tashkent. University press. 269p.(In Russian).

2. Abdushukurov.A.A. , Estimating of the probability density and insensity function of the Koziol-Green model of random censoring. Sankhya: Indian J. Statist.1987. Ser A. v.48.p 150 -168.

3. Abdushukurov A.A. Sagidullayev K.C. Kernel estimation of derivatives of distribution in proportional Razards model. In: Statistical Methods of Estimation and Hypothesis Testing. Perm. 2013. V.25.p 4-9.(In Russian).

4. *Csörgö* S. Estimating in the proportional hazards model of random censorship. Statistics. 1988. V19.N3.p 437-463.

5. *Csörgö* S. Mielniezuk J. Density estimating in the simple proportional hazards model. Statist.Probab.Letters.1988.v6.p 419-426.

Анализ данных типа времени жизни и сравнение непараметрических оценок функции распределения в модели случайного цензурирования справа

Л.Р.Махмутова
Филиал МГУ имени М.В. Ломоносова в городе Ташкенте
liliyamakhmutova17@gmail.com

1. Введение

В результате возникшего интереса к прогнозированию во многих прикладных областях (страхование, медицина, экономика, инженерия) часто объектом исследования выступает время до наступления некоторого события. Это время является неотрицательной случайной величиной (с.в.). Исследование и оценивание различных показателей таких с.в. производится с помощью методов анализа данных типа времени жизни.

Основной проблемой в этой области является неполнота, то есть цензурированность данных. Примером является выборка, элементами которой являются продолжительности жизни больных. В частности, данные продолжительности жизни больных из центра уединения Channing House в г. Пало-Альто в Калифорнии в виде вариационного ряда является цензурированной справа выборкой. Здесь возможны два случая: смерть больного или отсутсвие данных о его наблюдении начиная с некоторого момента времени (например, по причине выздоровления пациента или по причине окончания наблюдения). Второй случай и описывает цензурирование.

В данной статье предполагается независимость интересующей нас с.в. от мешающей (цензурирующей) с.в., а также проведен сравнительный анализ трех непараметрических оценок функции распределения (ф.р.): множительная оценка Каплана-Мейера, экспоненциальная оценка Альтшулера-Бреслоу и степенная оценка отношения рисков Абдушукурова в модели случайного цензурирования справа (см. подробно в [1]).

2. Основные характеристики данных типа времени жизни и их оценки

Пусть ξ – интересующая нас с.в. с ф.р. $F(x)$, плотностью $f(x)$ и функцией выживания (ф.в.) $S(x) := 1 - F(x)$.

Функция интенсивности (ф.и.) отказов (смертей) с.в. ξ определяется формулой:

$$\lambda(x) := \lim_{\Delta x \to 0+} \frac{P(x \le \xi \le x + \Delta x | \xi > x)}{\Delta x} = -\frac{d}{dx} \log S(x).$$

Интегральной функцией интенсивности (и.ф.и.) отказов является функция:

$$\Lambda(x) := \int_{-\infty}^{x} \lambda(u)du,$$

откуда следует:

$$S(x) = \exp(-\Lambda(x)).$$

Справедливо более общее представление Гилла для ф.р. (см [1]):

$$S(x) = \exp(-\Lambda^c(x)) \prod_{u \le x} (1 - \Delta\Lambda(u)), \tag{1}$$

где $\Lambda^c(x)$ – абсолютно непрерывная компонента и.ф.и. $\Lambda(x)$ и $\Delta\Lambda(u) := \Lambda(u) - \Lambda(u-)$.

Если $\{X_k, k \in \{1, ..., n\}\}$ независимые наблюдения над с.в. ξ, то эмпирическая ф.р. определяется формулой:

$$F_n(x) := \frac{1}{n} \sum_{k=1}^{n} I(X_k \le x),$$

где $I(\cdot)$ – индикаторная функция, удовлетворяет представлению (1):

$$1 - F_n(x) = \exp(\sum_{u \le x} \Delta \log(1 - F_n(u))) = \prod_{u \le x} (1 - \Delta\Lambda_n(u)),$$

где скачок эмпирической оценки $\Lambda_n(u)$ равен

$$\Delta\Lambda_n(u) = \frac{F_n(u) - F_n(u-)}{1 - F_n(u-)}.$$

Известно, что эмпирическая ф.р. является довольно хорошей оценкой ф.р. и, в частности, она равномерно сильно-состоятельная оценка [2]. Однако недостатком является то, что она дискретная. Поскольку чаще всего встречаются именно непрерывные ф.р., то оценивать их лучше с помощью непрерывных статистик. В этой связи вводятся так называемые ‘‘*сглаженные*” эмпирические оценки. В этом случае эмпирическую ф.р. F_n ‘‘сглаживают” при помощи некоторого ядра (весовой функцией) G, которая обладает некоторыми свойствами гладкости [2].

Рассмотрим ядерную оценку ф.р. F с.в. ξ, определяемую формулой:

$$F_n^*(x) := \int_{-\infty}^{x} G(\frac{x-u}{h(n)}) dF_n(u) = \frac{1}{n} \sum_{i=1}^{n} G(\frac{x - X_i}{h(n)}).$$

Здесь $\{h(n), n \geq 0\}$ – "*ширина окна*" (то есть положительная последовательность чисел, удовлетворяющая: $h(n) \to 0$, $nh(n) \to \infty$ при $n \to \infty$).

Ядерную оценку функции плотности (*оценку Парзена-Розенблатта*) с.в. ξ определим формулой:

$$f_n^*(x) := \frac{d}{dx}F_n^*(x) = \frac{1}{nh(n)}\sum_{i=1}^{n} w(\frac{x - x_i}{h(n)}),$$

где $w(x)$ – также весовая функция и $w(x) = G'(x)$. Стоит отметить, что ядерные оценки ф.р. и плотности также являются равномерно сильными состоятельными оценками при некоторых условиях на G, w и $h(n)$ (см. [3]). Аналогично оценивается и и.ф.и Λ, и ее "сглаженная" оценка имеет вид:

$$\Lambda_n^*(x) := \int_{-\infty}^{x} \frac{dF_n^*(u)}{1 - F_n^*(u)} = \int_{-\infty}^{x} \frac{f_n^*(u)du}{1 - F_n^*(u)}.$$

Тогда для ф.и. $\lambda(x)$ имеем оценку:

$$\lambda_n^*(x) := \frac{d}{dx}\Lambda_n^*(x) = \frac{f_n^*(x)}{1 - F_n^*(x)}.$$

3. Модель случайного цензурирования справа

Пусть ξ – интересующая нас с.в. с ф.р. $F(x)$ и η – цензурирующая с.в. с ф.р. $G(y)$. Вместо ξ наблюдаем пару (Z,δ), где $Z = \min(\xi,\ \eta)$ и $\delta = I(Z = \xi) = I(\xi \leq \eta)$. С.в. ξ и η предполагаются независимыми и наблюдается выборка $C^{(n)} = \{(Z_i,\ \delta_i),\ i \in 1, ..., n\}$, где $Z_i = \min(X_i, Y_i)$, $\delta_i = I(X_i \leq Y_i)$, где (X_i, Y_i) – независимые реализации пары (ξ, η).

Задача состоит в оценке ф.р. $F(x)$ по выборке $C^{(n)}$ при наличии мешающей ф.р. $G(y)$. Будем считать, что имеем дело с непараметрическим случаем, то есть семейства, в которые входят $F(x)$ и $G(y)$, нам заранее неизвестны.

Определим одну частную модель случайного цензурирования справа. Специальная модель случайного цензурирования справа, в которой существует число $\alpha > 0$ и $\Lambda_G(x) = \alpha\Lambda_F(x)$, где $\Lambda_G(x)$ и $\Lambda_F(x)$ – и.ф.и, соответствующие ф.р. $F(x)$ и $G(y)$, называются *моделью пропорциональных интенсивностей* (МПИ) Козела-Грина. Важной характеристикой МПИ является тот факт, что в МПИ с.в. Z и δ независимы. Это свойство облегчает последующий анализ оценок по выборке $C^{(n)}$ в МПИ.

Введем некоторые оценки ф.р. F по выборке $C^{(n)}$, следуя [1]:

1. Множительная оценка Каплана-Мейера:

$$F_n^{KM}(x) = 1 - \prod_{u \le x} (1 - \Delta\Lambda_{1n}(u)) = 1 - \prod_{u \le x} (1 - \frac{H_{1n}(u) - H_{1n}(u-)}{1 - H_{1n}(u-)}); \tag{2}$$

2. Экспоненциальная оценка Альтшулера-Бреслоу:

$$F_n^{AB}(x) = 1 - e^{-\Lambda_{1n}(x)} = e^{-\sum_{u \le x} \frac{H_{1n}(u) - H_{1n}(u-)}{1 - H_{1n}(u-)}}; \tag{3}$$

3. Степенная оценка отношения рисков Абдушукурова:

$$F_n^{RR}(x) = 1 - (1 - H_n(x))^{\frac{\Lambda_{1n}(x)}{\Lambda_n(x)}}; \tag{4}$$

Здесь: $1 - H_{1n}(u-) = \frac{1}{n} \sum_{j=1}^{n} I(z_j \ge u)$, $1 - H_n(x) = \frac{1}{n} \sum_{j=1}^{n} I(z_j > x)$, $H_{1n}(u) = \frac{1}{n} \sum_{j=1}^{n} \delta_j I(z_j \le u)$, $H_{1n}(u) - H_{1n}(u-) = \frac{\delta_j I(z_j = u)}{n}$,

$$\Lambda_{1n}(x) = \int_{-\infty}^{x} \frac{d\, H_{1n}(u)}{1 - H_{1n}(u-)} = \frac{1}{n} \sum_{j=1}^{n} \frac{\delta_j I(z_j \le x)}{\frac{1}{n} \sum_{i=1}^{n} I(z_i \ge z_j)} = \sum_{j=1}^{n} \frac{\delta_j I(z_j \le x)}{\sum_{i=1}^{n} I(z_i \ge z_j)},$$

$$\Lambda_n(x) = \int_{-\infty}^{x} \frac{d\, H_n(u)}{1 - H_n(u-)} = \sum_{j=1}^{n} \frac{I(z_j \le x)}{\sum_{i=1}^{n} I(z_i \ge z_j)}.$$

Качество оценок (2), (3) и (4) будем сравнивать также и построением доверительных полос для ф.р. F. Используем данные объема $n = 97$ из [1] (Channing House data). Согласно этим данным для ф.р. $F(x)$ будем строить доверительную полосу:

$$\left[M_n^-(x),\ M_n^+(x)\right],\quad 0 \le x \le T = 1128,$$

где в качестве $F_n^*(x)$ использованы оценки $F_n^{AB}(x)$, $F_n^{KM}(x)$ и $F_n^{RR}(x)$:

$$M_n^-(x) = F_n^*(x) - n^{-\frac{1}{2}} (1 - F_n^*(x)) (d_n^{\frac{1}{2}}(T) + 1,37 \frac{d_n(x)}{d_n(T)}), \tag{5}$$

$$M_n^+(x) = \frac{F_n^*(x) + n^{-\frac{1}{2}} (d_n^{\frac{1}{2}}(T) + 1,37 \frac{d_n(x)}{d_n(T)})}{1 + n^{-\frac{1}{2}} (d_n^{\frac{1}{2}}(T) + 1,37 \frac{d_n(x)}{d_n(T)})}, \tag{6}$$

$$d_n(x) = n \sum_{j=1}^{n} \frac{\delta_j I(z_j \le x)}{[\sum_{i=1}^{n} I(z_i \ge z_j)]^2}.$$

4.Постановка задачи и формулировка основных результатов

4.1 Проверка гипотезы о справедливости МПИ

Пусть $p = M\delta_i = P(x_i \leq y_i)$. $H_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(z_i \leq x)$, $p_n = \frac{\sum_{i=1}^{n}\delta_i}{n} = \frac{\nu_n}{n} = \frac{46}{97} \approx 0.47$. Проверим следующую гипотезу: H_0: $\{z_i$ и δ_i независимые$\}$. Напомним, что при выполнении H_0 мы будем иметь дело с МПИ Козела-Грина, так как именно в этом случае z_i и δ_i независимы, то есть $H_1(x) = p \times H(x)$.

Используем следующую статистику для ее проверки (см. [4]):

$$\Delta_n = \sqrt{\frac{n}{p_n(1-p_n)}} \sup_{|x|<\infty} |H_{1n}(x) - p_n H_n(x)|,$$

где

$$H_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_i I(z_i \leq x).$$

Учтем тот факт, что (см. [4]):
$\lim_{n\to\infty} P(\Delta_n \leq y|\ H_0) = K(y)$ – распределение Колмогорова.
Имеем: $\Delta_n = 0,394119724 < 1,36$ (1,36 – значение распределения Колмогорова при уровне значимости 0,05).

Следовательно, гипотеза принимается. Таким образом, данные Channing House отвечают МПИ, и это облегчает исследование вышеприведенных трех оценок.

4.2 Сравнение трех непараметрических оценок Проведем теперь сравнительный анализ трех основных непараметрических оценок функции распределения: оценка Каплана-Мейера, оценка Альтшулера-Бреслоу и оценка Абдушукурова. Будем строить их графики согласно формулам (2), (3) и (4), а также их доверительные полосы, согласно формулам (5), (6).

5 Заключение Была проверена и установлена справедливость гипотезы H_0: $\{z_i$ и δ_i независимые$\}$ для данных из Channing House. Согласно графикам, оценки Альтшулера-Бреслоу, Каплана-Мейера и Абдушукурова в целом близки друг к другу. Однако F_n^{KM} имеет существенный недостаток: если последнее наблюдение цензурировано, то оценка не достигает единицы. По аналогии с F_n^{KM}, F_n^{AB} имеет схожий недостаток: принимает единичное значение только при $t \to +\infty$. F_n^{RR} лишена недостатка такого неопределенного поведения и обращается в единицу до $t = +\infty$ в точке $Z_{(n)} = max\{Z_i, i = \overline{1,n}\}$. У оценок не была выявлена сильная вычислительная погрешность.

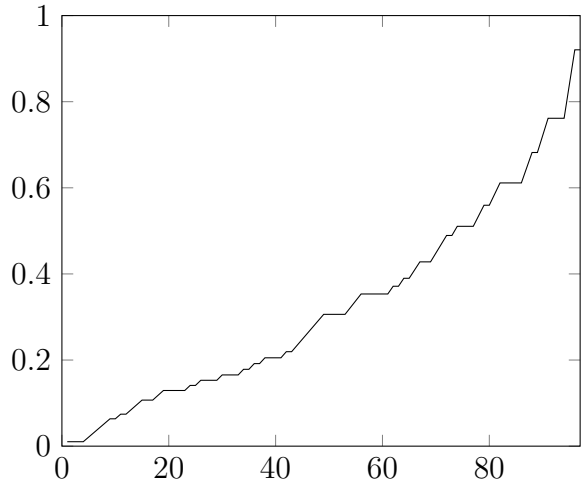


Рис. 5: Оценка Каплана-Мейера

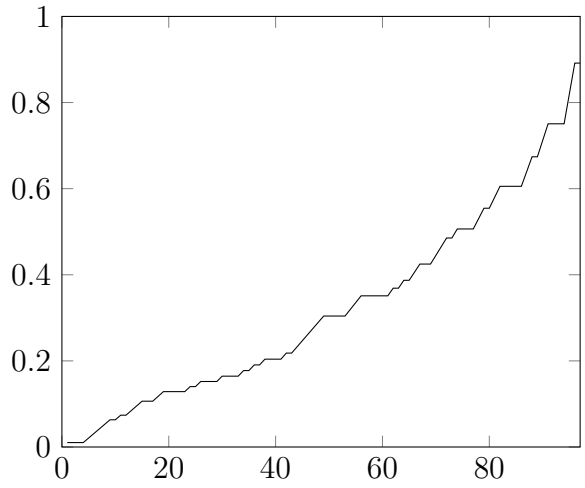


Рис. 6: Оценка Альтшулера-Бреслоу

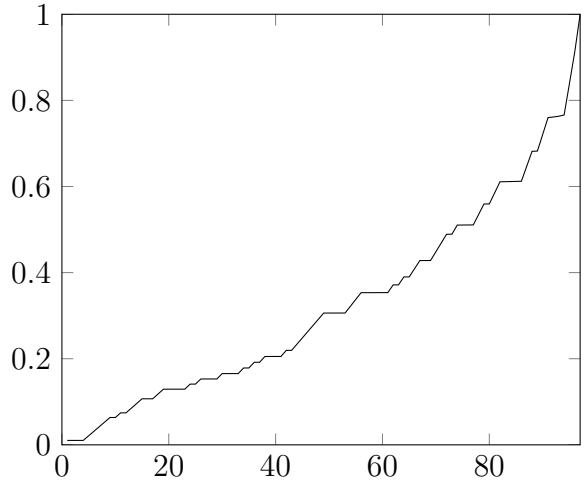


Рис. 7: Оценка Абдушукурова

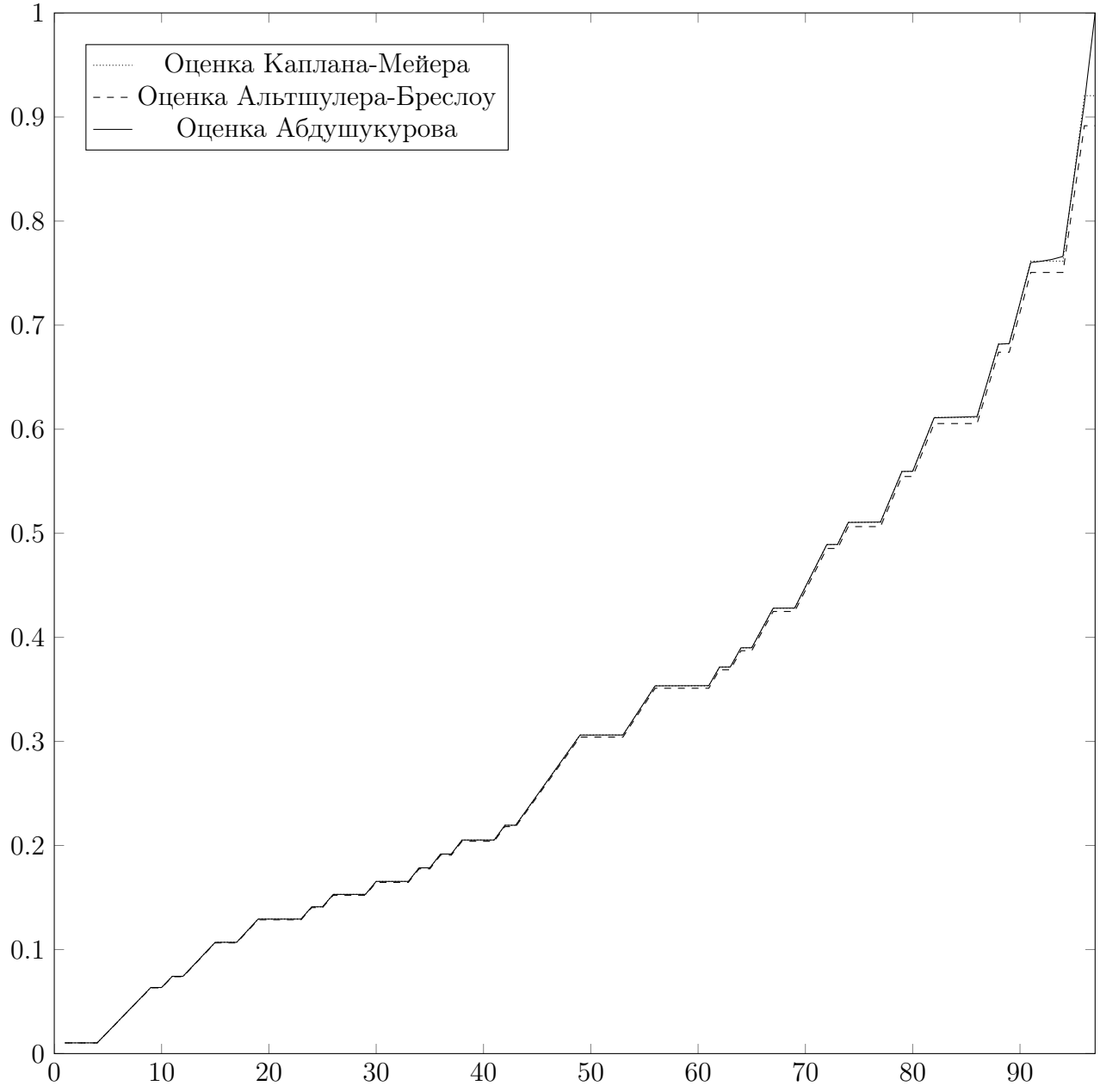


Рис. 8: Сводный график оценок

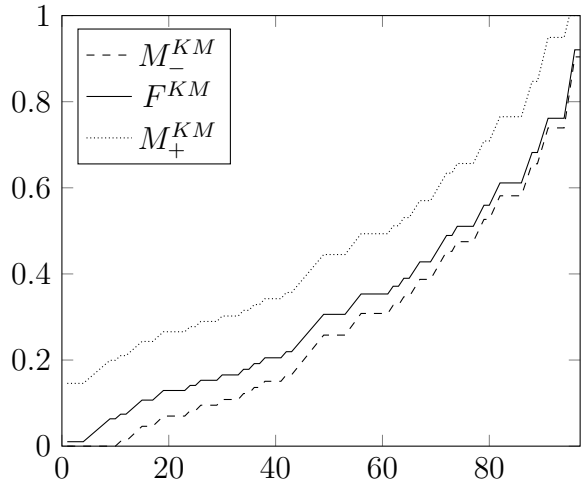


Рис. 9: Оценка Каплана-Мейера с доверительной полосой

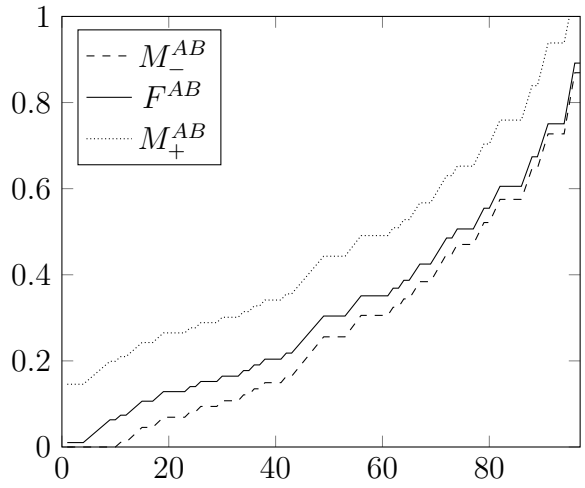


Рис. 10: Оценка Альтшулера-Бреслоу с доверительной полосой

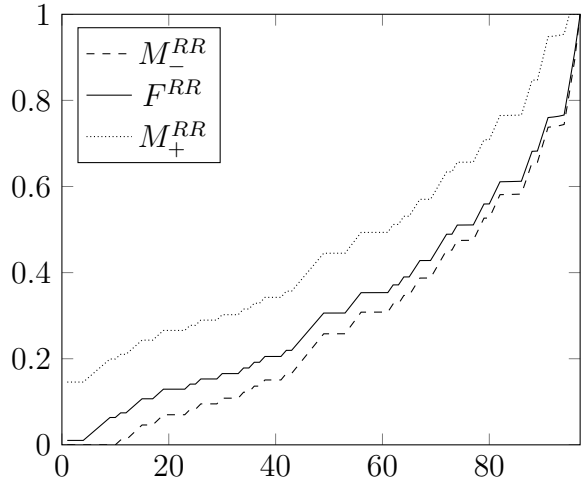


Рис. 11: Оценка Абдушукурова с доверительной полосой

ЛИТЕРАТУРА

1. Абдушукуров А.А. Статистика неполных наблюдений. Ташкент, «Университет» 2009. 269с.

2. Боровков А.А. Математическая статистика. Москва, «Наука» 1984. 472с.

3. Абдушукуров А.А., Сагидуллаев К.С. Аппроксимация сглаженных эмпирических процессов. Известия ВУЗов, 2003. №1-2, с. 51–59.

4. Abdushukurov A.A., Kakadjanova L.R. A class of special empirical processes of independence. J. Siberian Federal University: Math & Physics, 2015. v. 8, №2., p. 125–133.

Makhmutova L.R. (Tashkent, Uzbekistan)

Survival data analysis and comparison of nonparametrical estimators of cumulative distribution function in right random censoring model

Abstract. *In this paper we make survival data analysis (i.e. check whether the data satisfy Koziol-Green proportional hazards model). We also make a comparative analysis of three estimators of cumulative distribution function (Kaplan-Meier estimator, Altschuler-Breslow estimator, and Abdushukurov estimator).*

Key words: *Random right censoring, Koziol-Green proportional hazards model, Kaplan-Meier estimator, Altschuler-Breslow estimator, Abdushukurov estimator.*

Класс оценок многомерной плотности с оцениваемым параметром

К.С.Сагидуллаев
Технический институт Еджу в Ташкенте
khalmurza@mail.ru

Пусть $\{\xi_1, \xi_2, ...\}$ – последовательность независимых $k-$ мерных случайных векторов с общей абсолютно – непрерывной функцией распределения (ф.р.) $F(x)$ и плотностью $f(x)$, $x \in R^k$. Через $F_n(x)$, $x \in R^k$, обозначим эмпирическую ф.р. (э.ф.р.), построенную по выборке $\xi^{(n)} = (\xi_1, \xi_2, ..., \xi_n)$: $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\xi_i < x)$, $x \in R^k$. Пусть $\left\{\Delta_n(x) = \sqrt{n}\left(F_n(x) - F(x)\right), \quad x \in R^k\right\}$ – эмпирический процесс. Предположим, что ф.р. F задана с точностью до некоторого неизвестного параметра $\theta = (\theta_1, \theta_2, ..., \theta_s) \in \Theta \subseteq R^s$ и $F(x;\theta_0) = F(x)$ при истинном значении $\theta_0 = (\theta_{10}, \theta_{20}, ..., \theta_{s0})$ параметра θ, где θ_0 – внутренняя точка Θ. Последовательность оценок $\{\theta_n, n \geq 1\}$ такова, что имеет место представление
(А)

$$\sqrt{n}\left(\theta_n - \theta_0\right) = n^{-1/2}\sum_{i=1}^{n}\lambda\left(\xi_i;\theta_0\right) + \varepsilon_n, \tag{1}$$

где $\lambda(\cdot\,;\,\theta_0)$ – измеримая s – мерная векторзначная функция такая, что

$$E_{\theta_0}\lambda\left(\xi_i;\theta_0\right) = 0,$$

и матрица

$$E_{\theta_0}\left[\lambda\left(X_i;\theta_0\right)^T \cdot \lambda\left(\xi_i;\theta_0\right)\right] = \Im\left(\theta_0\right)$$

– конечная неотрицательно определена и $\varepsilon_n \xrightarrow{P} 0$, $n \to \infty$.
Например, если θ_n – оценка максимального правдоподобия, то

$$\lambda\left(x;\theta\right) = \frac{\partial \log f\left(x;\theta\right)}{\partial\theta} = \left(\frac{\partial \log f\left(x;\theta\right)}{\partial\theta_1}, ..., \frac{\partial \log f\left(x;\theta\right)}{\partial\theta_s}\right)^T,$$

и $\Im(\theta)$ – информационная матрица Фишера. Следует отметить, что представление (1) имеет место для достаточно широкого класса асимптотически нормальных оценок (подробно, см. [1]).
Обозначим $f(x;\theta_0) = f(x)$, $x \in R^k$. Для плотности $f(x)$ рассмотрим следующий класс непараметрических оценок:

$$f_n(x) = \int_{R^k}\psi_n(x;y)\,dF_n(y), \tag{2}$$

где $\left\{\psi_n(x;y),\ x,y\in R^k,\ n\geq 1\right\}$ последовательность измеримых по Борелью функций на квадрате $B^{(2k)}=B^k\times B^k$, где $B^k=(a,b)^k$. Выбирая функции $\psi_n(\cdot\,;\,\cdot)$ в (2) подходящим образом, получаем различные непараметрические оценки (см. [2]):

(a) При $\psi_n(x;y)=\left(\prod_{m=1}^k\mu_m(n)\right)^{-1}\cdot K\left(\frac{x-y}{\mu(n)}\right)$ получаем класс ядерных оценок типа Парзена-Розенблата, где $\mu(n)=(\mu_1(n);\mu_2(n);...;\mu_k(n))$, $\mu_m(n)\to 0$ и $n\mu_m(n)\to\infty$, – последовательность "ширины окна"и $\left\{K(x),\ x\in B^k\right\}$ – некоторая плотность;

(b) Пусть $\{q_j(u),\ j\geq 0,\ u\in(a;b)\}$ – полная ортонормированная последовательность функций на $(a;b)$. Для заданного вектора $I=(i_1,i_2,...,i_k)^T$ от целых неотрицательных чисел система $\left\{Q_I(x)=\prod_{j=1}^k q_{i_j}(x_j),\ \ x\in B^k\right\}$ из $k-$ мерных функций является ортонормальной. Определим последовательность функций

$$\left\{\psi_n(x;y)=\prod_{\bar\gamma\geq I}^{\gamma(n)}\varphi_{\bar\gamma}(x)\varphi_{\bar\gamma}(y),\ \ (x;y)\in B^{(2k)}\right\},\tag{3}$$

где $\bar\gamma=(\gamma(n),...,\gamma(n))^T$, $\gamma(n)\uparrow\infty$ при $n\to\infty$ и неравенство $I\leq\bar\gamma$ понимается покомпонентно. Класс оценок (2) с ядром (3) называется классом проекционных оценок.

Введём в рассмотрение процесс с оцениваемым параметром $\left\{\beta_n(x)=\sqrt{n}\left(f_n(x)-f_n^*(x;\theta_n)\right),\ \ x\in B^k,\ \ n\geq 1\right\}$, где $f_n^*(x;\theta)=\int_{R^k}\psi_n(x;y)dF(y;\theta),\ \ x\in B^k$.

В качестве аппроксимирующей для $\beta_n(x)$ рассмотрим процесс

$$G(x;n)=\int_{R^k}\psi_n(x;y)dR(y;n),$$

где $R(y;n)=n^{-1/2}K(y;n)-\nabla_\theta F(y;\theta_0)^T\cdot\int_{R^k}\lambda(z;\theta_0)\,dn^{-1/2}K(z;n)$.

Здесь $\nabla_\theta F(y;\theta)=\left(\frac{\partial F(y;\theta)}{\partial\theta_1},...,\frac{\partial F(y;\theta)}{\partial\theta_s}\right)^T$, а двухпараметрический гауссовский процесс (процесс Кифера) $K(\cdot\,;\,\cdot)$ имеет нулевое среднее и ковариационную структуру при $x,y\in R^k$:

$$E_{\theta_0}K(x;n)\cdot K(y;m)=\min(n;m)\cdot\left\{F(\min(x;y))-F(x)F(y)\right\},$$

где $\min(x,y)=(\min(x_1,y_1),...,\min(x_k,y_k))$.

Следующая теорема дачт возможность аппроксимации процесса $\beta_n(x)$ последовательностью гауссовских процессов $G(x;n)$.

Теорема. Пусть $\{f(\cdot;\ \theta),\ \theta \in \Theta\}$ регулярное семейство плотностей и для оценки θ_n параметра θ справедливо представление (1). Тогда при $n \to \infty$

$$\sup_{x \in B^k} |\beta_n(x) - G(x;n)| = o_p(1).$$

ЛИТЕРАТУРА

1. Burke M.D., Csörgő M., Csörgő S., Révész P. Approximations of the empirical process when parameter are estimated. // Ann.Probab.1978.v.7.N.5.pp.790-810.

2. Хашимов Ш.А. О непараметрической оценке многомерной плотности распределений. Ташкент. ФАН. 1985.с.197-207.

On estimation of bivariate survival function under right random censorship

A.A.Abdushukurov [1], R.S.Muradov [2],
[1] Moscow State University Tashkent Branch, Uzbekistan,
[2] Namangan Institute of Engineering & Technology,
Uzbekistan, ISI & ISEC, Kolkata, India
abdushukurov1710@gmail.com,rustamjonmuradov@gmail.com

1. Introduction. The problem of estimation of multivariate distribution (or survival) function from incomplete data is considered with beginning of 1980's (Campbell (1981), Campbell & Földes (1982), Hanley & Parnes (1983), Horváth (1983), Tsay, Leurgang & Crowley (1986), Burke (1988), Dabrowska (1988, 1989), Gill (1992), Huang (2000), Abdushukurov(2004) etc.)(see, [1-20]). In the special bivariate case, there are the numerous examples of paired data representing the times to death of individuals (twins or married couples), the failure times of components of system and others which subject to random censoring. At present time there are several approaches to estimation of survival functions of vectors of life times. However, some of these estimators either are inconsistent or not fully defined in range of joint survival functions and hence not applicable in practice. In this work we present estimators for bivariate survival function and present some large sample properties of estimators.

2. Random right censoring model. Let $\mathbb{X} = \{X_i = (X_{1i}, X_{2i})\}_{i=1}^{\infty}$ −a sequence of independent and identically distributed (i.i.d.) two-dimensional random vectors with a common continuous survival function $F(s,t) = P(X_{11} > s, X_{21} > t)$, $(s,t) \in \overline{R}^{+2} = [0,\infty)\times[0,\infty)$. This sequence is censored from the right by sequence $\mathbb{Y} = \{Y_i = (Y_{1i}, Y_{2i})\}_{i=1}^{\infty}$of i.i.d. random vectors with survival function $G(s,t) = P(Y_{11} > s, Y_{21} > t)$, $(s,t) \in \overline{R}^{+2}$. The observation is available the sample $\mathbb{V}^{(n)} = \{(Z_i,\ \Delta_i),\ 1 \le i \le n\}$, where $Z_i = (Z_{1i}, Z_{2i})$, $\Delta_i = (\delta_{1i}, \delta_{2i})$, $Z_{ki} = \min(X_{ki}, Y_{ki})$,$\delta_{ki} = I(Z_{ki} = X_{ki})$, $k = 1,2$, and $I(\cdot)$is indicator. The problem is consist in estimating of F from the sample $\mathbb{V}^{(n)}$ under nuisance G. Let $H(s,t) = P(Z_{1i} > s, Z_{2i} > t)$, $(s,t) \in \overline{R}^{+2}$ and the sequences $\mathbb{X}$ and $\mathbb{Y}$ are independent. Then $H(s,t) = F(s,t)\,G(s,t)$, $(s,t) \in \bar{R}^{+2}$. In this paper, we using exponential-hazard, product-limit and relative-risk power types functionals from [2] construct the corresponding estimates of three types for F. However, unlike of [2] in the empirical estimates the upper index of summation, n we replace by the Poisson random variable (r.v.) μ_n with expectation $E\mu_n = n$. Such amounts arise in

the insurance as the size of group insurance payments by an insurance company to customers in connection with an insured event. Following [2], we introduce some auxiliary estimated functionals for $(x,y) \in \bar{R}^{+2}$:

$$M(x,y) = P(Z_{11} \le x,\, Z_{21} > y), \;\; N(x,y) = P(Z_{11} > x,\, Z_{21} \le y),$$

$$\bar{M}(x,y) = P(Z_{11} \le x,\, Z_{21} > y,\, \delta_{11} = 1), \;\; \bar{N}(x,y) = P(Z_{11} > x,\, Z_{21} \le y,\, \delta_{21} = 1),$$

$$\Lambda_1(x,y) = \int_0^x \frac{M(ds,y)}{H(s-,y)}, \qquad \Lambda_2(x,y) = \int_0^y \frac{N(x,dt)}{H(x,t-)}, \tag{2.1}$$

$$\bar{\Lambda}_1(x,y) = \int_0^x \frac{\bar{M}(ds,y)}{H(s-,y)}, \qquad \bar{\Lambda}_2(x,y) = \int_0^y \frac{\bar{N}(x,dt)}{H(x,t-)},$$

$$\Lambda(x,y) = \Lambda_1(x,0) + \Lambda_2(x,y), \;\; \bar{\Lambda}(x,y) = \bar{\Lambda}_1(x,0) + \bar{\Lambda}_2(x,y),$$

$$\Lambda^c(x,y) = \Lambda_1^c(x,0) + \Lambda_2^c(x,y), \;\; \bar{\Lambda}^c(x,y) = \bar{\Lambda}_1^c(x,0) + \bar{\Lambda}_2^c(x,y),$$

where

$$\Lambda_1^c(x,y) = \Lambda_1(x,y) - \sum_{s \le x} \Lambda_1(\triangle\, s, y), \;\; \Lambda_1(\triangle\, s, y) = \Lambda_1(s,y) - \Lambda_1(s-,y),$$

$$\Lambda_2^c(x,y) = \Lambda_2(x,y) - \sum_{t \le y} \Lambda_2(x, \triangle\, t), \;\; \Lambda_2(x, \triangle\, t) = \Lambda_2(x,t) - \Lambda_2(x,t-),$$

and similarly defined $\bar{\Lambda}_1^c$ and $\bar{\Lambda}_2^c$. To construct estimates for F we estimate the functionals (2.1). At first, we introduce the following empirical estimates from the sample $\mathbb{V}^{(n)}$of the first four probabilities in (2.1):

$$H_n(x,y) = \frac{1}{n} \sum_{i=1}^{n} I(Z_{1i} > x,\, Z_{2i} > y),$$

$$M_n(x,y) = \frac{1}{n} \sum_{i=1}^{n} I(Z_{1i} \le x,\, Z_{2i} > y),$$

$$N_n(x,y)=\frac{1}{n}\sum_{i=1}^{n} I\left(Z_{1i}>x,\ Z_{2i}\le y\right), \tag{2.2}$$

$$\bar{M}_n(x,y)=\frac{1}{n}\sum_{i=1}^{n} I\left(Z_{1i}\le x,\ Z_{2i}>y,\ \delta_{1i}=1\right),$$

$$\bar{N}_n(x,y)=\frac{1}{n}\sum_{i=1}^{n} I\left(Z_{1i}>x,\ Z_{2i}\le y,\ \delta_{2i}=1\right),$$

Let $\{\mu_n,\ n\ge 1\}$- be a sequence of Poisson r.v-s. with parameter $E\mu_n=n$, and independent of the pair $(\mathbb{X};\mathbb{Y})$. Along with estimates (2.2), we propose also their analogues $H_n^*,\ M_n^*,\ N_n^*,\ \bar{M}_n^*,\ \bar{N}_n^*$ obtained from them by replacing the upper limit of summation n by r.v. μ_n. However, it should be noted that these estimates have the disadvantage that they can be greater than 1. In fact, for example, for

$$H_n^*(x,y)=\frac{1}{n}\sum_{i=1}^{\mu_n} I\left(Z_{1i}>x,\ Z_{2i}>y\right),$$

we have

$$P\left(H_n^*(0,0)>1\right)=P\left(\mu_n>n\right)=\sum_{m=n+1}^{\infty}\frac{n^m e^{-n}}{m!}>0.$$

To avoid this disadvantage, as is done in the one-dimensional case in [2], we consider the following truncated versions $H_n^*,\ M_n^*,\ N_n^*,\ \bar{M}_n^*,\bar{N}_n^*$:

$$H_n^0(x,y)=1-\left(1-H_n^*(x,y)\right)I\left(H_n^*(x,y)\le 1\right)=$$

$$=\begin{cases} H_n^*(x,y), \text{ if } H_n^*(x,y)\le 1, \\ \\ 0, \text{ if } H_n^*(x,y)>1, \end{cases}$$

and similarly constructed the estimates $M_n^0,\ N_n^0,\ \bar{M}_n^0,\ \bar{N}_n^0$. Using the latests we now construct the corresponding estimates for the functionals in (2.1):

$$\Lambda_{1n}(x,y)=\int_0^x\frac{M_n^0(ds,y)}{H_n^0(s-,y)},\quad \Lambda_{2n}(x,y)=\int_0^y\frac{N_n^0(x,dt)}{H_n^0(x,t-)},$$

$$\bar{\Lambda}_{1n}(x,y)=\int_0^x\frac{\bar{M}_n^0(ds,y)}{\bar{H}_n^0(s-,y)},\quad \bar{\Lambda}_{2n}(x,y)=\int_0^y\frac{\bar{N}_n^0(x,dt)}{\bar{H}_n^0(x,t-)}, \tag{2.3}$$

$$\Lambda_n(x,y) = \Lambda_{1n}(x,0) + \Lambda_{2n}(x,y),\ \bar{\Lambda}_n(x,y) = \bar{\Lambda}_{1n}(x,0) + \bar{\Lambda}_{2n}(x,y).$$

Consider also the relative-risk function $R(x,y) = \frac{\bar{\Lambda}(x,y)}{\Lambda(x,y)}$ and its estimator $R_n(x,y) = \frac{\bar{\Lambda}_n(x,y)}{\Lambda_n(x,y)}$. Now we propose the following three estimates of $F(x,y)$ of exponential, product and power structures by using the estimates (2.3) (for details, see formulas (1.5.1) from [2]):

$$F_{1n}(x,y) = \exp\left\{-\bar{\Lambda}_n(x,y)\right\} = \exp\left\{-\left(\bar{\Lambda}_{1n}(x,0) + \bar{\Lambda}_{2n}(x,y)\right)\right\},$$

$$F_{2n}(x,y) = \prod_{s\le x}\left(1 - \bar{\Lambda}_{1n}(\triangle\, s, 0)\right)\prod_{t\le y}\left(1 - \bar{\Lambda}_{2n}(x, \triangle\, t)\right), \qquad (2.4)$$

$$F_{3n}(x,y) = [H_n(x,y)]^{R_n(x,y)}.$$

Let $\Delta_n = \left[0, Z_1^{(n)}\right] \times \left[0, Z_2^{(n)}\right] \cap \Delta$, where $Z_k^{(n)} = \max(Z_{k1}, ..., Z_{kn}),\ k = 1, 2.$ The following theorem states the asymptotic equivalence of estimates (2.4).

Theorem. *For all* $(x,y) \in \Delta_n$ *:*
(I) $0 \le F_{1n}(x,y) - F_{2n}(x,y) = \mathrm{O}_p\left(\frac{1}{n}\right)$.
If the survival function G *is also continuous on* Δ_n*, then*
(II) $|F_{1n}(x,y) - F_{3n}(x,y)| = \mathrm{O}_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right)$.
It is easy to see that, from (I) and (II) we also obtain

$$|F_{3n}(x,y) - F_{2n}(x,y)| = \mathrm{O}_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right).$$

REFERENCE

1. Abdushukurov A.A., Nonparametrical estimators of survival function on the plane by censored observations. Uzbek mathematical journal. 2004, Vol. 2, pp. 3-11.(In Russian)
2. Abdushukurov A.A., Statistics of incomplete observations. NUUz. University press, Tashkent, 2009. (In Russian)
3. Abdushukurov A.A., Estimates of unknown distributions from incomplete observations and their properties. LAMBERT Academic Publishing, Germany, 2011.(In Russian)
4. Abdushukurov A.A., Estimation of joint survival function from censored observations. Zavod. Lab. Diagn. Mater.. 2016, Vol. 82, pp. 80-84.(In Russian)

5. Abdushukurov A.A., Muradov R.S., On the estimates of the distribution function in a random censorship model. Zavod. Lab. Diagn. Mater.. 2014, Vol. 80, pp. 62-67.(In Russian)

6. Abdushukurov A.A., Muradov R.S., On Estimation of Conditional Distribution Function under Dependent Random Right Censored Data. Journal of Siberian Federal University. 2014, Vol. 7, pp. 409-416.

7. Abdushukurov A.A., Muradov R.S., Estimation of survival functions from dependent random right censored data. Intern. J. Innovation in Science and Mathematics. 2014, Vol. 2, pp. 280-287.

8. Abdushukurov A.A., Muradov R.S., Some algebraic properties of Archimedean copula functions and their applications in the statistical estimation of the survival function. New Trends in Math. Sci.. 2016, Vol. 4, 2016, pp. 88-98.

9. Burke M.D., Estimation of a bivariate distribution function under random censorship. Biometrika. 1988, Vol. 75, pp. 379-382.

10. Campbell G., Nonparametric bivariate estimation with randomly censored data. Biometrika. 1981, Vol. 68, pp. 417-422.

11. Campbell G., Földes A. , Lange sample properties of nonparametric bivariate estimators with censored data. Colloquia Mathematica-Societatis János Bolyai. 1982, Vol. 32, pp. 103-122.

12. Dabrowska D.M., Kaplan Meier estimate on the plane. Ann. Statist.. 1988, Vol. 16, pp. 1475-1489.

13. Dabrowska D.M., Kaplan Meier estimate on the plane: weak convergence, LIL and the bootstrap J. Multivar. Anal.. 1989, Vol. 29, pp. 308-325.

14. Gill R.D., Multivariate Survival Analysis I. Theory Probab. Appl.. 1992, Vol. 37, pp. 19-35.(In Russian)

15. Gill R.D., Multivariate Survival Analysis II. Theory Probab. Appl.. 1992, Vol. 37, pp. 307-328.(In Russian)

16. Hanley J.A., Parnes M.N., Nonparametric estimation of a multivariate distribution in the presence of censoring. Biometrica. 1983, Vol. 39, pp. 129-139.

17. Horváth L., The rate of strong uniform consistency for the multivariate product-limit estimator. J. Multivar. Anal.. 1983, Vol. 13, pp. 202-209.

18. Huang Y., Two sample multistate accelerated sojourn times model. J.A.S.A.. 2000, Vol. 95, pp. 619-627.

19. Muradov R.S., Abdushukurov A.A., Estimation of multivariate distributions and there mixes by incomplete data. LAMBERT Academic Publishing, Germany. 2011.(In Russian)

20. Tsai W.Y., Leurgans S., Crowley J., Nonparametric estimation of a bivariate survival function in the presence of censoring. Ann. Statist.. 1986, Vol. 14, pp. 1351-1365.

A.A.Abdushukurov (Tashkent, Uzbekistan), R.S.Muradov (Namangan, Uzbekistan & Kolkata, India),

On estimation of bivariate survival function under right random censorship

Abstract. *At present time there are several approaches to estimation of survival functions of vectors of life times. However, some of these estimators either are inconsistent or not fully defined in range of joint survival functions and hence not applicable in practice. In this paper we consider three types of estimates of exponential-hazard, product-limit and relative-risk power structures for the bivariate survival function, when replacing the number of summands in empirical estimates with a sequence of Poisson random variables. It is shown that these estimates are asymptotically equivalent.*

Key words: *Bivariate survival function, Poisson random variables, empirical estimates.*

Potts model on a Cayley Tree in the Presence of Three Competing Interactions

A.M.Rahmatullaev
Institute of Mathematics, Tashkent, Uzbekistan
anasbek.rahmatullaev@gmail.com

We will consider a semi-infinite Cayley tree Γ_+^2 of second order, i.e.an infinite graph without cycles with 3 edges issuing from each vertex except for x^0 which has only 2 edges.

Let denote the Cayley tree as $\Gamma_+^2 = (V, L)$, where V is the set of vertices and L is the set of edges. Two vertices x and y, $x, y \in V$ are called *nearest-neighbors* if there exists an edge $l \in \Lambda$ connecting them, which is denoted by $l = \langle x, y\rangle$. The distance $d(x, y), x, y \in V$, on the Cayley tree Γ_+^2, is the number of edges in the shortest path from x to y. For a fixed $x^0 \in V$ we set

$$W_n = \{x \in V | d(x, x^0) = n\}, \quad V_n = \{x \in V | d(x, x^0) \le n\}$$

and L_n denotes the set of edges in V_n. The fixed vertex x^0 is called the 0-th level and the vertices in W_n are called the n-th level. For the sake of simplicity we put $|x| = d(x, x^0)$, $x \in V$. Two vertices $x, y \in V$ are called *the next-nearest-neighbors* if $d(x, y) = 2$. The next-nearest-neighbor vertices x and y are called *prolonged next-nearest-neighbors* if $|x| \neq |y|$ and is denoted by $\widetilde{\rangle x, y\langle}$. Three vertices x, y and z are called a triple of neighbors, and they denoted by $\langle x, y, z\rangle$ if $\langle x, y\rangle$ and $\langle y, x\rangle$ are nearest-neighbors. The triple of vertices $x, y, z \in V$ is called *two level* and denoted $\overline{\langle x, y, z\rangle}$ if $y \in W_n$ and $x, z \in W_{n+1}$ for some n.

For the 3-state Potts model with spin values in $\Phi = \{1, 2, 3\}$, the relevant Hamiltonian with competing nearest-neighbor, prolonged next-nearest-neighbor interactions and two-level triple-neighbor triple-neighbor interactions has the form

$$H(\sigma) = -J_1 \sum_{\langle x,y\rangle \in L} \delta_{\sigma(x)\sigma(y)} - J_p \sum_{\widetilde{\rangle x,y\langle}} \delta_{\sigma(x)\sigma(y)} - J_t \sum_{\overline{\langle x,y,z\rangle}} \delta_{\sigma(x)\sigma(y)\sigma(z)}, \qquad (1)$$

where $J_1, J_p, J_t \in R$ are coupling constants with $J_1 > 0, J_p < 0$ and $\langle x, y\rangle$ stands for nearest-neighbors vertices and $\widetilde{\rangle x, y\langle}$ stands for next-nearest-neighbor interaction restricted to the sites belonging to the same branch and $\overline{\langle x, y, z\rangle}$ two-level triple-neighbor interactions.

In this paper we consider $\delta_{\sigma(x)\sigma(y)\sigma(z)}$ as follows

$$\delta_{\sigma(x)\sigma(y)\sigma(z)} = \begin{cases} 1 & \text{if } \sigma(x)=\sigma(y)=\sigma(z) \\ 0 & \text{otherwise} \end{cases}$$

The case $J_t = 0$ was considered in [1]. Follow the papers [2] to produce the recurrent equations, we consider the relation of the partition function on V_n to the partition function on subsets of V_{n-1}. Let $Z^{(n)}(i_1, i_0, i_2)$ be a partition function on V_n where the spin in the root x^0 is i_0 and the two spins in the proceeding x^1 and x^2 are i_1 and i_2, respectively. As shown in [2] one can select only five independent variables $Z^{(n)}(1,1,1)$, $Z^{(n)}(2,1,2)$, $Z^{(n)}(1,2,1)$, $Z^{(n)}(2,2,2)$, $Z^{(n)}(3,2,3)$ and with the introduction of new variables

$$u_1^{(n)} = \sqrt{Z^{(n)}(1,1,1)},\ u_2^{(n)} = \sqrt{Z^{(n)}(2,1,2)},$$
$$u_3^{(n)} = \sqrt{Z^{(n)}(1,2,1)},\ u_4^{(n)} = \sqrt{Z^{(n)}(2,2,2)},\ u_5^{(n)} = \sqrt{Z^{(n)}(3,2,3)},$$

straightforward calculations show that

$$u_1^{(n+1)} = ac(bu_1^{(n)} + 2u_2^{(n)})^2,$$

$$u_2^{(n+1)} = (bu_3^{(n)} + u_4^{(n)} + u_5^{(n)})^2,$$

$$u_3^{(n+1)} = (u_1^{(n)} + (b+1)u_2^{(n)})^2, \tag{2}$$

$$u_4^{(n+1)} = ac(u_3^{(n)} + bu_4^{(n)} + u_5^{(n)})^2,$$

$$u_5^{(n+1)} = (u_3^{(n)} + u_4^{(n)} + u_5^{(n)})^2,$$

and the total partition function is given in terms of (u_i) by

$$Z^{(n)} = (u_1^{(n)} + 2u_2^{(n)})^2 + 2(u_3^{(n)} + u_4^{(n)} + u_5^{(n)})^2, \tag{3}$$

where β is the inverse temperature and $a = \exp(\beta J_1), b = \exp(\beta J_p)$ and $c = exp(\beta J_t/2)$. We note that, in the paramagnetic phase (high symmetry phase), $u_1 = u_4$ and $u_2 = u_3 = u_5$. For discussing the phase diagram, the following choice of reduced variables is convenient:

$$x = \frac{2u_2 + u_3 + u_5}{u_1 + u_4}, \quad y_1 = \frac{u_1 - u_4}{u_1 + u_4}, \quad y_2 = \frac{u_2 - u_3}{u_1 + u_4}, \quad y_3 = \frac{u_2 - u_5}{u_1 + u_4}. \tag{4}$$

The variable x is just a measure of the frustration of the nearest-neighbor bonds and is not an order parameter like y_1, y_2 and y_3. In this case the basic equations have following form:

$$x' = \frac{1}{2acD}[P(y_1, y_2, y_3) + ((b+1)x + 2 - y_1 - by_2 - y_3)^2],$$

$$y_1' = \frac{2}{D}(b + x)(by_1 + y_2 + y_3),$$

$$y_2' = -\frac{1}{acD}[y_1 + by_2 - y_3][2 + (b+1)x - (b-1)(y_2 - y_3)], \tag{5}$$

$$y_3' = -\frac{1}{acD}(b-1)(y_2 - y_3)[2 + (b+1)x - 2y_1 - (b+1)(y_2 + y_3)],$$

where

$$D = (b + x)^2 + (by_1 + y_2 + y_3)^2,$$
$$P(y_1, y_2, y_3) = 3y_1^2 + (4b^2 - 4b + 3)y_2^2 + (3b^2 - 4b + 4)y_3^2 +$$
$$+2(2b + 1)y_1y_2 + 2(b + 2)y_1y_3 - (2b^2 - 7b + 2)y_2y_3.$$

Theorem. If there exist such $a > 0, b > 0, c > 0$ that the basic equation (3) has more than one positive invariant points than takes place phase transition for the model (1).

REFERENCE

1. N. Ganikhodjaev and A. Rahmatullaev, Journal of Physics: Conference Series 435 012035 (2013)
2. N.N. Ganikhodjaev, F.M. Mukhamedov and C.H. Pah, Phys. Lett. A **373**, 33 (2008)

$m-$Выпуклые функции

С.М.Ражабов
Ташкентский Государственный университет Узбекского языка и литературы
rajabovs90@mail.ru

Пусть $u(x)$ дважды гладкая вещественная функция в области $D \subset \mathbb{R}^n$, $u \in C^2(D)$. Рассмотрим матрицу производных второго порядка $D^2u = \left(\frac{\partial^2 u}{\partial x_j \partial\, x_k}\right)$. После подходящего ортонормального преобразования матрица она преобразуется в диагональную форму

$$\left(\frac{\partial^2 u}{\partial x_j \partial\, x_k}\right) \to \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{pmatrix},$$

где $\lambda_j = \lambda_j(x) \in \mathbb{R}-$собственные значения матрицы D^2u. Тогда сумма $H_k(u) = \sum_{1\le j_1<...<j_k\le n} \lambda_{j\,1}... \lambda_{j_k}$ называется гессианом размерности k вектора собственных значений $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$.

Определение 1. Дважды гладкая функция $u \in C^2(D)$ называется $m-$*выпуклой* в $D \subset \mathbb{R}^n$, $u \in m - cx(D)$, если

$$H_k(\lambda(x)) \ge 0,\ \forall x \in D,\ \ k = 1, ..., m$$

Связь $m-cx$ функций с sh_m функциями. Мы вложим $\mathbb{R}^n$ в $\mathbb{C}^n$, $\mathbb{R}^n(x) \subset \mathbb{C}^n(z = x + iy)$, как вещественное подпространство.

Теорема 1. Дважды гладкая функция $u(x) \in C^2(D)$, $D \subset \mathbb{R}^n$, является $m - cx$ в D тогда и только тогда когда функция $u^c(z) = u^c(x + iy) = u(x)$ является sh_m в области $D \times \mathbb{R}^n(y)$.

Напомним, дважды гладкая функция $w(z) \in C^2(G)$, $G \subset \mathbb{C}^n$, называется sh_m функцией, если $(dd^c w)^k \wedge \beta^{n-k} \ge 0,\ \ k = 1, ..., m$. Доказана, что $w(z) \in C^2(G)$является $sh_m(G) \Leftrightarrow$ когда $dd^c w \wedge dd^c v_1 \wedge ... \wedge dd^c v_{m-1} \wedge \beta^{n-m} \ge 0\ \ \forall\ v_1, ..., v_{m-1} \in sh_m(G) \bigcap C^2(G)$.

Теорема 2. Дважды гладкая функция $u(x)$, $x \in D$, является $m - cx(D)$ тогда и только тогда, когда в $D \times \mathbb{R}^n$ выполняется следующее усло-

вие

$$\forall\, v_1, ..., v_{m-1} \in m{-}cx\,(D) \cap C^2\,(D)\,, \quad dd^c u^c \wedge dd^c v_1^c \wedge ... \wedge dd^c v_{m-1}^c \wedge \beta^{n-m} \geq 0 \quad (1)$$

Общее определение $m-cx$ функций. Соотношение (1) позволяет нам определить $m - cx$ функции в классе L^1_{loc} функций.

Определение 2. Функция $u \in L^1_{loc}\,(D)$ называется $m - cx$ в области $D \subset \mathbb{R}^n$, если она:

1. полунепрерывна сверху в $D \subset \mathbb{R}^n$,т.е.

$$\overline{\lim_{x \to x^0}}\, u(x) = \lim_{\varepsilon \to 0} \sup_{B(x^0, \varepsilon)} u(x) \leq u(x^0)$$

2. для любых дважды гладких $m - cx\,(D)$ функций $v_1\,,\, ...\,, v_{m-1}$ поток $dd^c u^c \wedge dd^c v_1^c \wedge ... \wedge dd^c v_{m-1}^c \wedge \beta^{n-m}$, определяемый как

$$[dd^c u^c \wedge dd^c v_1^c \wedge ... \wedge dd^c v_{m-1}^c \wedge \beta^{n-m}]\,(\omega) =$$

$$= \int u\, dd^c v_1^c \wedge ... \wedge dd^c v_{m-1}^c \wedge \beta^{n-m} \wedge dd^c \omega, \qquad \omega \in F^{0,0}\,(D \times \mathbb{R}^n)$$

положителен в $D \times \mathbb{R}^n$.

Сформулируем следующих важных свойств $m - cx$ функций**:**

1) $u_j(x) \in m - cx(D), \quad a_j \in R_+ \quad (j = 1, 2, ..., N) \quad \Rightarrow$

$$a_1 u_1(x) + a_2 u_2(x) + ... + a_N u_N(x) \in m - cx(D);$$

2) $u_j(x) \in m - cx(D)$, $u_j\,(x) \geq u_{j+1}\,(x)$, $(j = 1, 2, ...) \Rightarrow$

$$\lim_{j \to \infty} u_j(x) \in m - cx(D);$$

3) $u_j(x) \in m - cx(D)$, $(j = 1, 2, ...)$, $u_j(x) u\,(x)$, тогда $u\,(x) \in m - cx\,(D)$;

4) $cx = n - cx \subset ... \subset 1 - cx = sh$;

Докажем последнее свойство. В случае m=1 очевидно. Покажем что, если m=n, тогда $u(x)$есть выпуклая функция, $u \in cx(D)$. Рассмотрим полином

$$\prod_{i=1}^{n} (t + \lambda_i) = t^n + H_1(\lambda) t^{n-1} + ... + H_n(\lambda)\,, \quad t \in \mathbb{R}.$$

Здесь $H_k(\lambda) \geq 0,\ \ k = 1, ..., n$. Отсюда вытекает что, все корни полинома отрицательно, $-\lambda_i \leq 0,\ \ i = 1, ..., n$. Значит $\lambda_i \geq 0,\ \ i = 1, ..., n$ т.е.$u \in cx(D)$.

ЛИТЕРАТУРА

1. Zbigniew Blocki Weak solutions to the complex hessian equation, Ann. Inst. Fourier, Grenoble 55, 5, 2005, 1735–1756

2. Neil S. Trudinger, Xu-Jua Wang,Hessian Measures I, Topological Methods in Analysis, Journal of the Juliuusz Schauder Center, V. 10,1997,225-239

3. Neil S. Trudinger, Xu-Jua Wang,Hessian Measures II, Annals of Mathematics, V.150, 1999, 579-604

Suyunjon Rajabov (Tashkent, Uzbekistan)

Title

Abstract. *In this paper the criterion of m-convexity functions was found.*

Key words: $m - cx$ *function, cx function,* sh_m *function*

Моделирование задачи комплекса очистки загрязненных вод

Н.У.Утеулиев [1], Б.Н.Бегилов [2], Г.М.Джайков [3]

[1,2,3] Нукусский филиал Ташкентского университета информационных технологии имени Мухаммеда аль-Хорезмий

bbegilov@gmail.com

При формировании задачи оптимизации водоохранных мероприятий существенное значение имеет выбор модели комплекса водоохранных мероприятий, устанавливающей зависимость между величиной средств, выделяемых водопользователю для очистки и качественным составом загрязненных вод перед сбросом в водный объект. Учитывая это, важной задачей является задача оптимизации развития совокупности комплексов очистки загрязненных вод, связанных единой гидрографической сетью при заданных ограничениях на качество вод в водном объекте.

В данной работе рассматривается стохастическая эколого-экономическая модель, отражающая процесс поиска наилучшего способа очистки с минимальными затратами. Приводится алгоритм решения предложенной задачи.

Теперь остановимся на проблеме, состоящей в распределении некоторого объема загрязненных вод сброса среди технологических схем очистки. Под технологической схемой очистки понимают набор некоторых устройств очистки, расположенных последовательно. Часть сброса, отведенная под очистку по некоторой схеме, в дальнейшем не должна попадать на устройства чистки другой схемы.

Для каждого варианта технической реализации способа очистки можно установить связь между расходом очищаемых вод и соответствующими затратами на их очистку, а также качественные характеристики очистки по различным схемам для контроля видов примесей. По имеющимся литературным источникам [3], зависимости, отражающие связь затрат водопользователя на реализацию водоохранных мероприятий с объѣмом загрузки технологических схем очистки являются вогнутыми функциями и хорошо аппроксимируются степенными функциями с дробным показателем. Кроме того, коэффициенты целевой функции могут зависеть от случайных факторов, тогда получим стохастическую эколого-экономическую модель комплекса очистки загрязненных вод.

Задача состоит в выборе такой допустимой комбинации технологий и такого распределения объѣма загрязненных вод между ними, чтобы затра-

ты на очистку были минимальными и при этом концентрации примесей после очистки у контрольного створа не превышали заданных предельно допустимых значений. В реальности задача водоочистки решается для некоторого водоочистного комплекса при наличии нескольких водосбросов и концентрация контролируемых примесей у контрольного створа зависит от качества проведения водоохранных мероприятий на всех объектах водосброса.

Введем следующие обозначения: - q_k объем загрязненных вод, поступающих на комплекс очистки на k- том водосбросе ; - x_{ki} объем загрязненных вод, который будет обрабатываться по iЦ той технологической схеме очистки на k- том водосбросе. Таким образом, $X_k = \{x_{ki}\}, i = \overline{1,N}$, вектор распределения всего объъма загрязненных вод между технологиями очистки на k - том водосбросе; x_{ki}^- и x_{ki}^+ - величины, ограничивающие производственные мощности i - той технологической схемы при очистке сброса объема q_k; C_{0kj}, C_{kj} - концентрации примесей j -го вида соответственно до очистки и предельно-допустимые концентрации (ПДК) после очистки; P_{kij} - степень очистки j-той примеси по i-той технологии очистки на k- том водосбросе; $A_{ki}(\theta), \lambda_{ki}$- коэффициенты аппроксимации функции затрат на очистку загрязненных вод по i -той технологической схеме на k- том водосбросе, θ- случайное число ($A_{ki}(\theta) \geq 0; \quad 0 \leq \lambda_{ki} \leq 1, \quad k = \overline{1,L},\ i = \overline{1,N}$). Определяем стохастическую эколого-экономическую модель, отражающую процесс поиска наилучшего способа очистки с минимальными затратами:

$$F(x,\theta) = \sum_{k=1}^{L}\sum_{i=1}^{N} MA_{ki}(\theta)x_{ki}^{\lambda_{ki}} \to \min, \tag{1}$$

$$\sum_{i=1}^{N} B_{kij}x_{ki} \leq C_{kj}, \quad k = \overline{1,L};\ j = \overline{1,M}, \tag{2}$$

$$\sum_{i=1}^{N} x_{ki} = q_k,\ k = \overline{1,L}, \tag{3}$$

$$x_{ki}^- \leq x_{ki} \leq x_{ki}^+, \quad A_{ki} \geq 0, \quad 0 \leq \lambda_{ki} \leq 1, \quad k = \overline{1,L},\ i = \overline{1,N}. \tag{4}$$

Остановимся на экономическом содержании модели (1)-(4). Целевая функция (1) с математическим ожиданием M отражает средние суммарные затраты водопользователя на проведение водоочистных мероприятий. Очевидно, что зависимости, отражающие связь затрат на реализацию очистки с объъмом загрузки технологий очистки являются вогнутыми функциями.

Минимизация целевой функции даже на выпуклой области представляет повышенную сложность. Ограничения (2)-(4) задают область поиска наилучшего решения. Ограничение (2) означает, что у контрольного створа концентрация j- того вида примеси не должна превышать допустимую.

Коэффициент "очищаемости"обозначенный B_{kij}, вычисляется следующим образом:

$$B_{kij} = C_{0kj}(1 - P_{kij}). \tag{5}$$

Ограничение (3) требует, чтобы вес объем k- того сброса на том водосбросе был распределен по технологическим схемам. Ограничение (4) означает, что объчм загрязненных вод, направленных на очистку, не должен превышать производственной мощности технологических схем. Для решения данной задачи (1)-(4) применяем алгоритм, предложенный в работе [2]. Построим функцию Лагранжа

$$L(x, u, \theta) = \sum_{k=1}^{L} (\sum_{i=1}^{N} (MA_{ki}(\theta)x_{ki}^{\lambda_{ki}} - u_k x_{ki}) + u_k q_k). \tag{6}$$

Если нам известно u^* - оптимальное решение двойственной задачи к задаче (1)-(4), то можно рассматривать еч эквивалентную постановку:

$$L(x, u^*, \theta) \to \min. \tag{7}$$

где $x \in X$, X -множество векторов, удовлетворяющих условиям (2),(4). Из задачи (7) получаем, с учетом (6), k-независимых подзадач для каждого водосбора

$$\sum_{i=1}^{N} (MA_{ki}(\theta)x_{ki}^{\lambda_{ki}} - u_k^* x_{ki}) \to \min. \tag{8}$$

где $x \in X$.

Задача (8) является задачей геометрического программирования с вогнутой нелинейной целевой функцией и она может быть решена методом условного градиента.

Таким образом, при известном решении u^* происходит декомпозиция задачи (1)-(4). То же самое будет и при произвольном u, только тогда полученное при этом решение задачи (8) не будет оптимальным решением задачи (1)-(4).

На самом деле точное оптимальное решение двойственной задачи неизвестно. Однако, можно построить процедуру определения u^*, которая имеет вид:

В качестве начального приближения u^0 берем произвольное число. Пусть нам известно u^s, s=0,1,... Тогда определяем x^s , решая задачу

$$\sum_{i=1}^{N} (MA_{ki}(\theta)x_{ki}^{\lambda_{ki}} - u_k^s x_{ki}) \to \min . \tag{9}$$

где $x \in X$.

Новое значение x^{s+1} определяется с помощью формулы

$$u_k^{s+1} = \max\{0, \quad \rho_s(q_k - \sum_{i=1}^{N} x_{ki}^s)\}. \tag{10}$$

где ρ_s- некоторый шаговый множитель. Следует отметить, что $q_k - \sum_{i=1}^{N} x_{ki}^s$ в (10) является обобщенным градиентом целевой функции $\varphi(u) = \min_{x \in X} L(x, u, \theta)$ двойственной задачи. При сделанных предположениях эта функция будет недифференцируемой при тех us, когда хотя бы одна из задач (9) имеет более одного оптимального решения. Сама процедура (10) будет методом обобщенно-градиентного спуска [1] для решения двойственной задачи. Поэтому выбор шагового множителя ρ_s может быть основан на следующих условиях обеспечивающих сходимость метода:

$$\rho_s \geq 0, \quad \rho_s \to 0 \quad s \to \infty, \quad \sum_{s=0}^{\infty} \rho_s = \infty, \quad \sum_{s=0}^{\infty} \rho_s^2 < \infty.$$

Этим условиям удовлетворяет, например следующее значение шагового множителя $\rho_s = \frac{1}{(1+s)^\alpha}$, $\frac{1}{2} < \alpha \leq 1$.

ЛИТЕРАТУРА

1. Ермольев Ю. М. Методы стохастического программирования. М: Наука, 1976. 176с.

2. Ляшенко И. Н., Михалевич М. В., Утеулиев Н. У. Методы эколого-экономического моделирования. Н: Билим, 1994. 236с.

3. Тукалевский С. Л. О решении многоэкстремальной нелинейной задачи распределительного типа с использованием двойственных оценок. Методы решения экстремальных задач и смежные вопросы, 1989. с.39–44.

Uteiev Nietbay Uteulievich (Nukus, Uzbekistan), Begilov Bakhadir Nazarbaevich (Nukus, Uzbekistan), Djaykov Gafur Muratbaevich (Nukus, Uzbekistan)

Modeling the Problem of the Complex of Purification of Contaminated Waters

Abstract. *In this paper we consider a stochastic environmental and economic model that reflects the process of finding the best cleaning method with minimal cost. An algorithm for solving the proposed problem is given.*

Key words: *Wastewater, technological schemes, approximation, stochastic environmental-economic model, maximum permissible concentration*

Wavelet estimation of distribution function in Koziol-Green model of random censorship

B.B.Muratova
Qarshi Engineering - Economics Institute

1. Introduction

The wavelets in curve estimation problems of statistics are well known tool (see, [1-7]). However, there is not much researches on wavelet estimation of curves under random censoring models, Authors [7] investigated the asymptotic expression of the MISE for the estimator of the density function under the Koziol-Green proportional hazards model (PHM) of random censorship. In this paper we focus on nonlinear wavelets of distribution function in PHM by using two approaches.

2. PHM and estimators.

Let $X_1, ..., X_n$ be independent random variables (r.v-s) with a absolute continuous distribution function (d.f) F. These are censored on the right by the independent r.v.-s $Y_1, ..., Y_n$ with d.f. G so, that the observation available are the pairs (Z_k, Δ_k), where $Z_k = \min(X_k, Y_k)$, $\Delta_k = I(Z_k = X_k)$, $1 \leq k \leq n$ and $I(\cdot)$ is the indicator function. Assume that censored X sequence and the censoring Y sequence are independent and hence $1 - H(t) = P(Z > t) = (1 - F(t))(1 - G(t))$, $t \in R$. The simple PHM is an special semiparametric model in which there exist a positive constant β, the censoring parameter, such that

$$1 - G(t) = (1 - F(t))^{\beta}, \quad t \in R. \tag{1}$$

Under (1), the expected proportion $p = P(\delta = 1)$ of uncensored observations is $p = 1/(1 + \beta)$ and from (1)

$$1 - F(t) = (1 - H(t))^{p}, \quad t \in R. \tag{2}$$

Note that in PHM r.v.-s $\{Z_1, ..., Z_n\}$ and $\{\Delta_1, ..., \Delta_n\}$ are independent. Therefore, probabilities $H(t)$ and p can be estimated by their empiricals as

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_k \leq t) \text{ and } p_n = \frac{1}{n}\sum_{i=1}^{n} \Delta_k. \tag{3}$$

Then the corresponding estimator of F is easy can be obtained from (2) by plugin in the empirical (3) as follows

$$F_n^{ACL}(t) = 1 - (1 - H_n(t))^{p_n}, t \in R. \tag{4}$$

The estimator (4) is well-known Abdushukurov-Cheng-Lin estimator and there are enormous literature, where PHM has been very frequently used for constructing and investigating the estimators of several functionals of d.f. F. In this paper we propose the wavelet analogue of estimator (4).

Let $h(t) = H'(t)$ be density function of d.f. $H(t)$. Let φ and Ψ be "father"and "mother"wavelets are bounded and compactly supported, and, for some $r \geq 1$, we have

$$\mu_k = \int y^k \Psi(y)\, dy = 0, \quad 0 \leq k \leq r-1,$$

$$\mu_r = r! k_0 \neq 0, \quad k_0 = (r!)^{-1} \int y^r \Psi(y)\, dy.$$

Let's introduce the functions

$$\varphi_j(x) = q^{1/2} \varphi(q\ x - j), \quad \Psi_{i,j}(x) = q_i^{1/2} \Psi(q_i\ x - j), \quad x \in R,$$

with $q > 0,\ j \in \mathbb{Z}$ and $q_i = q\ 2^i,\ i \in \mathbb{N} \bigcup \{0\}$. For wavelets φ and Ψ we have

$$\int \varphi_{j_1}(x)\, \varphi_{j_2}(x)\, dx = \delta_{j_1,j_2}, \ \int \Psi_{i_1,j_j}(x)\, \Psi_{i_2,j_2}(x)\, dx =$$

$$= \delta_{i_2,i_2} \cdot \delta_{j_1,j_2}, \quad \int \varphi_{j_1}(x)\, \Psi_{i,j_2}(x)\, dx = 0,$$

where $\delta_{i,j}$ denotes the Kronecker delta. Furthermore, an arbitrary square-integrable function $h(t)$ $(h \in \mathbb{L}_2(R))$ may be expanded in a generalized Fourier series of the form

$$h(t) = \sum_{j \in \mathbb{Z}} b_j \varphi_j(t) + \sum_{i=0}^{\infty} \sum_{j \in \mathbb{Z}} b_{ij} \Psi_{i,j}(t), \tag{5}$$

where

$$b_j = \int \varphi_j(x)\, dH(x), \quad b_{i,j} = \int \Psi_{i,j}(x)\, dH(x),$$

and the series in (5) converges in the $\mathbb{L}_2(R)$ space. Then using the unbiased estimates

$$\widehat{b}_j = \frac{1}{n} \sum_{k=1}^{n} \varphi_j(Z_k), \quad \widehat{b}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \Psi_{i,j}(Z_k)$$

of coefficients b_j and $b_{i,j}$ respectively we introduce the wavelet estimator of $h(t)$ as

$$h_n(t) = \sum_{j \in \mathbb{Z}} \widehat{b}_j \varphi_j(t) + \sum_{i=0}^{l-1} \sum_{j \in \mathbb{Z}} \widehat{b}_{ij} I\left(\left|\widehat{b}_{ij}\right| > \delta\right) \Psi_{i,j}(t), \tag{6}$$

where $\delta > 0$ and $l \geq 1$ smoothing parameter. Then two estimators of d.f. F follows from (2) as

$$\widehat{F}_n(t) = 1 - [1 - \int_{-\infty}^{t} h_n(s)ds]^{p_n}, t \in R, \tag{7}$$

$$\widetilde{F}_n(t) = p_n \int_{-\infty}^{t} h_n(s)[1 - H_n(s)]^{p_n - 1}ds, t \in R. \tag{8}$$

We present some asymptotic (as $n \to \infty$) results for estimators (7) and (8): strong uniform consistency and asymptotic normality.

REFERENCE

1. Abdushukurov A.A., Sagidullaev K.S., Muratova B.B. Estimation of distribution function in proportional hazards model by wavelet function. // Mater.of Republ. scient.-appl.conf. "STATISTICS and its application ". October 19-20, 2017. Tashkent. NUUz. p.108-110.

2. Csörgő S., Faraway J.J. The paradoxical nature of the proportional hazards model of random censorship. Statistics. 1998. v.31. p.67-78.

3. Donoho D.L., Johnstone I.M. Ideal spatial adaptation by wavelet shrinkage. Biometrika. 1994. v.81. N.3. p. 425-455.

4. Donoho D.L., Johnstone I.M. Adapting to unknown smoothnees via wavelet shrinkage. J.A.S.A. 1995. v.90. p. 1200-1224.

5. Donoho D.L., Johnstone I.M., Kerkyacharian G., Picard D. Density estimation by wavelet thresholding. Ann. Statist. 1996. v.23. p. 508-539.

6. Hall P., Patil P. Formulas for mean integrated squared error of nonlinear wavelet-based density estimators. Ann. Statist. 1995. v.23.p. 905-928.

7. Rodriguez-Casal A., de Una-Alvarez J. Nonlinear wavelet density estimation under the Koziol-Green model. Reports in Statistics and Operations Research. Report 02-01. 24.05.2002.

Локальная асимптотическая нормальность неоднородных статистических экспериментов в модели конкурирующих рисков при гибридном случайном цензурировании справа

Н.С.Нурмухамедова
Национальный университет Узбекистана
rasulova_ nargiza@mail.ru

1. Введение.

Статистика отношения правдоподобия (СОП) играет ключевую роль в теории принятия решений. Например, при проверке простой гипотезы H_0 против сложной альтернативы H_1 о неизвестном законе распределения критерии, построенные на основе СОП согласно фундаментальной лемме Неймана-Пирсона является равномерно - наиболее мощными при любом объъме n наблюдений (см. [5,6]). Довольно интересные задачи возникают, когда альтернативы H_1 зависят от n и являются кблизкимиь к H_0, т.е. $H_1 = H_{1n} \to H_0$ при $n \to \infty$. В таких ситуациях проявляются асимптотические свойства СОП, полезные в теории оценивания неизвестных параметров и проверки гипотез. К таким свойствам относятся локальная и равномерная асимптотическая нормальность (ЛАН и РАН) СОП. В работах авторов [1-3] свойства ЛАН для СОП установлены в некоторых моделях неполных наблюдений, получаемых случайным цензурированием наблюдений в модели конкурирующих рисков (МКР). В данной работе также исследованы свойства ЛАН для СОП при специальном гибридном цензурировании справа в МКР, где исходные наблюдения неоднородны.

2. Неоднородная МКР и гибридной цензурирование справа.

Следует отметить, что в классической МКР предполагается однородность наблюдаемых случайных величин (с.в) и сопровождающих их событий. В настоящей работы мы откажемся от условия однородности экспериментов и предположим, что исходные наблюдения образуют выборку $\overline{Z} = (Z_1, \ldots, Z_n)$ из независимых векторов $Z_j = \left(X_j, \overset{(1)}{j}, \ldots \overset{(k)}{j}\right)$ с совместным распределением

$$Q_{j\theta}\left(x,\ y^{(1)}, \ldots y^{(k)}\right) = P_{j\theta}\left(X_j < x,\ \delta_j^{(1)} = y^{(1)}, \ldots, \delta_j^{(k)} = y^{(k)}\right),$$

где $x \in R^1$, $y^{(i)} \in \{0;1\}$, $\delta_j^{(i)} = I\left(A_j^{(i)}\right)$, $i = \overline{1,k}$- индикаторы событий $\left\{A_j^{(i)}, \ldots, A_j^{(k)}\right\}$ образующих полную группу при каждом j и Θ открытой множество в R^1. Пусть $H_j(x;\theta) = P_{j\ \theta}(X_j < x)$ и $H_j^{(i)}(x;) = P_{j\theta}(X < x,\ \delta_j^{(i)} = 1)$-маргинальные распределения с.в. X_j и пар $\left(X_j, \overset{(i)}{j}\right)$, $i = \overline{1,k}$.

Так как $\delta_j^{(1)}+\cdots+\delta_j^{(k)}=1$, следовательно для всех (x;θ)$\in R^1\times\Theta$: $H_j^{(1)}(x;)+\ldots+H_j^{(k)}(x;)=H_j(x;)$ при каждом j.

Предположим, что в МКР субраспределения $H_j^{(i)}$ абсолютно непрерывны и имеют плотности $h_j^{(i)}$, тогда $h_j^{(1)}(x;\theta)+...+h_j^{(k)}(x;\theta)=\frac{\partial H_j(x;\theta)}{\partial x}=h_j(x;\theta)$ для всех $(x;\theta)\in R^1\times\Theta$. Введем также и функции интенсивности $\lambda_j^{(i)}(x;\theta)=h^{(i)}(x;\theta)/(1-H_j(x;\theta))$, $\lambda_j(x;\theta)=h_j^{(i)}(x;\theta)/(1-H_j(x;\theta))=\lambda_j^{(1)}(x;\theta)+...+\lambda_j^{(k)}(x;\theta)$, $(x;\theta)\in R^1\times\Theta$. Пусть $X_{1n}<X_{2n}<...<X_{nn}$ - вариационный ряд из с.в. $\{X_1,...,X_n\}$. Через $\{\delta_{jn}^{(1)},...,\delta_{jn}^{(k)},j=\overline{1,n}\}$ обозначим индикаторы в выборке , соответствующие вариантам $\{X_{jn},j=\overline{1,n}\}$. Пусть

$$\mathbb{Z}^{(m)}=(\mathcal{Z}_{1n},...,\mathcal{Z}_{mn}),\ m=1,...,n,\ \mathbb{Z}^{(0)}=\mathcal{Z}_{0n}=0, \tag{1}$$

где $\mathcal{Z}_{jn}=(X_{jn},\delta_{jn}^{(1)},...,\delta_{jn}^{(k)})$. Через $(\mathcal{Y}^{(n)},\mathcal{U}^{(n)},Q_\theta^{(n)})$ обозначим последовательность статистических экспериментов, порожденную наблюдениями (1). Следовательно,$\mathcal{Y}^{(n)}=\{\mathcal{X}_j\times\{0,1\}^{(k)}\}^{(n)}$, $\mathcal{U}^{(n)}=\sigma(\mathcal{Y}^{(n)})$ и $Q_\theta^{(n)}$ - распределение на $(\mathcal{Y}^{(n)},\mathcal{U}^{(n)})$, являющееся n-кратным произведением "одномерных"распределений $Q_{j\theta}$. Семейство мер $\{Q_\theta^{(n)},\theta\in\Theta,n\geq 1\}$ является абсолютно непрерывным относительно меры $\nu^{(n)}=\nu_1\times...\times\nu_n$, где $d\nu_m=dx_m\times\varepsilon_{y_m^{(1)}}\times...\times\varepsilon_{y_m^{(k)}}$, и $\varepsilon_{y_m^{(i)}}$ -считающие меры, сосредоточенные в точке $y_m^{(i)}\in\{0,1\}$, $i=\overline{1,k}$; $m=\overline{1,n}$ и ее плотность задается формулой

$$q_n(\mathbb{Z}^{(n)};\theta)=\frac{dQ_\theta^{(n)}(\mathbb{Z}^{(n)})}{dv^{(n)}(\mathbb{Z}^{(n)})}=\prod_{m=1}^{n}\prod_{i=1}^{k}[h_m^{(i)}(x_m;\theta)]^{y_m^{(i)}}. \tag{2}$$

В дальнейшем рассмотрим такую статистическую схему, согласно которой X_j подвергается цензурированию справа случайной величиной $T_{rn}=T\wedge X_{rn}$, где T-фиксированный момент времени и X_{rn}-член вариационного ряда порядка $r\,(0<r<n)$. Рассматриваемую схему наблюдений назовем гибридной $(n,r,T)^k$-моделью. Очевидно эта модель содержит в себе МКР и в частных случаях при $(n,n,T)^k=(n,T)^k$ и $(n,r,\infty)^k=(n,r)^k$ получаем МКР при цензурировании по типу I и II соответственно. Такого рода неоднородные эксперименты могут быть проведены для сохранения больше ресурсов, останавливая испытания n различных объектов с продолжительностями безотказной работы X_j и подвергаемых k конкурирующим рискам либо в заранее фиксированный момент времени T либо при осуществлении r-выхода из строя в момент X_{rn}. Следует отметить, что в $(n,r,T)^k$-модели

число наблюдаемых X_j-х является случайной величиной

$$\tau = \begin{cases} \sup\{m : X_{mn} \le T,\ m \le r\},\ \text{и } X_{1n} \le T, \\ \\ 0,\ \text{в противном случае.} \end{cases}$$

Таким образом, наблюдению доступен вектор $\left(Z^{(\tau)}, \tau\right)$, где $Z^{(\tau)} = (Z_{1n}, ..., Z_{\tau n})$. Совместное распределение вектора $\left(Z^{(\tau)}, \tau\right)$ задаётся плотностью

$$p_n\left(Z^{(\tau)}, \theta\right) = \frac{n!}{(n-\tau)!} \prod_{l=1}^{\tau} \prod_{i=1}^{k} \left\{ \left[h^{(i)}(x_{in}; \theta)\right]^{y_{l\,n}^{(i)}} \right\} [1 - H(t_{rn}; \theta)]^{n-\tau} \cdot$$

$$\cdot I\left(x_{1n} < ... < x_{rn},\ \tau \le r\right), \qquad (4)$$

где $t_{rn} = x_{rn} \wedge T$ и $\prod_{l=1}^{0} \equiv 1$.

3. ЛАН статистического эксперимента Пусть $(\mathcal{Y}^*, \mathcal{U}^*) = \prod_{j=1}^{\infty} (\mathcal{Y}^{(j)}, \mathcal{U}^{(j)})$ и $\mathcal{P}_\theta$ - вероятностное распределение на $(\mathcal{Y}^*, \mathcal{U}^*)$, являющееся произведением мер Q_θ и M_θ оператор математического ожидания относительно $\mathcal{P}_\theta$. Через $U_{n,m}$ обозначим σ - алгебры, порожденные векторами (1). Момент остановки (3) является измеримой относительно неубывающей по m последовательности σ - алгебр $\{U_{n,m}, 1 \le m \le n\}$. Через $\mathcal{P}_{n,\theta}$ обозначим сужение меры $\mathcal{P}_\theta$ на σ - алгебру U_{n,τ_n}. Далее нас будет интересовать свойство ЛАН семейства вероятностных мер $\{\mathcal{P}_{n,\theta}, \theta \in \Theta\}$, соответствующего цензурированной на уровне τ_n справа выборке . Пусть $\mathcal{P}_{n,\theta}^{(m)}$ - сужение меры $\mathcal{P}_\theta$ на $U_{n,m}$. Тогда совместная плотность вектора согласно (4) задается формулой

$$p_n(z^{(m)}; \theta) = \frac{n!}{(n-m)!} \prod_{l=1}^{m} \prod_{i=1}^{k} \{[h^{(i)}(z_{ln}; \theta)]^{y_{ln}^{(i)}}\} [1 - H(z_{mn}; \theta)]^{n-m}, \qquad (5)$$

где $z^{(m)} = (z_{1n}, ..., z_{mn})$ - реализация вектора . Носителем плотности (5) является множество $N_{n,m} = \{z^{(m)} : 0 < z_{1n} < ... < z_{mn} < \infty\}$. Пусть $u_\theta(z_{mn}/U_{n,m-1})$ - условная плотность Z_{mn} относительно σ - алгебры $U_{n,m-1}$. Тогда нетрудно видеть, что согласно (5) для $z_{mn} > Z_{m-1,n}$:

$$u_\theta(z_{mn}/U_{n,m-1}) = (n-m+1) \prod_{i=1}^{k} \left\{ \frac{[h^{(i)}(z_{mn}; \theta)]^{y_{mn}^{(i)}} [1 - H(z_{mn}; \theta)]^{n-m}}{[1 - H(z_{m-1,n}; \theta)]^{n-m+1}} \right\}. \qquad (6)$$

С учетом (5) и (6), имеем

$$p_n(z^{(m)};\theta) = p_n(z^{(m-1)};\theta)u_\theta(z_{mn}/U_{n,m-1}) = \prod_{l=1}^{m} u_\theta(z_{ln}/U_{n,l-1}). \qquad (7)$$

Пусть $\theta_0 \in \Theta$ - фиксированное значение параметра θ. Для заданного , определим последовательность $\theta_n = \theta_0 + un^{-1/2} \in \Theta$. Зададимся СОП для $m = \overline{1,n}$ согласно (7):

$$\frac{d\mathcal{P}^{(m)}_{n,\theta_n}}{d\mathcal{P}^{(m)}_{n,\theta_0}} = l_{n,m}(u) = \frac{p_n(z^{(m)};\theta_n)}{p_n(z^{(m)};\theta_0)} = \prod_{l=1}^{m} \frac{u_{\theta_n}(z_{ln}/U_{n,l-1})}{u_{\theta_0}(z_{ln}/U_{n,l-1})}. \qquad (8)$$

Для доказательства свойства ЛАН для СОП (8) введем условия регулярности при $j = \overline{1,n}$

(У1) Носители $\{N^{(j)}_{h^{(i)}} = \{x \geq 0 : h^{(i)}_j(x;\theta) > 0\}, i = \overline{1,k}\}$ - независят от θ;

(У2) Плотности $\{h^{(i)}_j, i = \overline{1,k}\}$ - непрерывно дифференцируемы по θ и производные $\left\{\frac{\partial h^{(i)}_j(x;\theta)}{\partial\theta}, i = \overline{1,k}\right\}$ - равномерно ограничены для всех $(x;\theta) \in R^+ \times U_{\theta_0}$, где U_{θ_0} - окрестность θ_0;

(У3) Функции $\left\{\frac{\partial \log \lambda^{(i)}_j(x;\theta_0)}{\partial\theta}, i = \overline{1,k}\right\}$- дифференцируемы;

(У4) Существует число $\Delta > 0$ такое, что для всех $i = \overline{1,k}$: $M_{\theta_0}\left|\frac{\partial \log \lambda^{(i)}_j(X;\theta_0)}{\partial\theta}\right|^{2+\Delta} < \infty$, $M_{\theta_0}\left|\frac{\partial \log h^{(i)}_j(X;\theta_0)}{\partial\theta}\right|^{2+\Delta} < \infty$;

(У5) Существует число $\beta \in (0,1]$ такое, что при $n \to \infty$, $\frac{\tau_n}{n} \overset{Q_{\theta_0}}{\to} \beta$ и для всех $i = \overline{1,k}$: $I^{(i)}_{\beta j} = I^{(i)}_{\beta j}(\theta_0) = \int\limits_0^{H^{-1}(\beta;\theta_0)} \left(\frac{\partial \log \lambda^{(i)}_j(x;\theta_0)}{\partial\theta}\right)^2 dH_j(x;\theta_0) > 0$, где $H^{-1}_j(\beta;\theta_0)$ - квантиль уровня β распределения $H_j(x;\theta_0)$;

(У6) При каждом $m = \overline{1,n}$ функции $\int\limits_{z_{m-1,n}}^{\infty} u(z/U_{n,m-1})d\gamma(z)$ - дифференцируемы по θ в точке θ_0 под знаком интеграла;

(У7) При каждом $u \in R^1$

$$\lim_{n\to\infty} M_\theta \left\{ \sup_{|\theta-\theta_0|\leq \frac{u}{\sqrt{n}}} \frac{1}{n} \sum_{m=1}^{n} \int\limits_{z_{m-1,n}}^{+\infty} \left[\frac{\partial}{\partial\theta}\{u_\theta(z/U_{n,m-1})\}^{1/2} - \frac{\partial}{\partial\theta_0}\{u_{\theta_0}(z/U_{n,m-1})\}^{1/2}\right]^2 d\gamma(z) \right\} = 0,$$

где $\frac{\partial}{\partial\theta_0}\{u_{\theta_0}(z/U_{n,m-1})\}^{1/2} = \left(\frac{\partial}{\partial\theta}\{u_\theta(z/U_{n,m-1})\}^{1/2}\right)_{\theta=\theta_0}$.

Отметим, что условия (У1) - (У3) являются обычными условиями регулярности, (У4) обеспечит условие Ляпунова для установления ЛАН для

СОП, (У5) дает ограничение на рост τ_n и требует положительность информаций Фишера $I_{\beta j}^{(i)}$, а условия (У6) и (У7) требуют непрерывность по θ интегралов от условных плотностей.

Исследуем асимптотические свойства статистик $\{\omega_{n,\tau_n}, n \geq 1\}$, где . Пусть $I_{n\beta} = \sum_{j=1}^{n} \sum_{i=1}^{k} I_{\beta j}^{(i)}$. Справедливо утверждение.

Теорема. При справедливости условий (У1) - (У7) при каждом $u \in R^1$ для СОП имеет место представление

$$l_{n,\tau_n}(u) = \exp\{uI_{n\beta}^{1/2}\omega_{n,\tau_n} - \frac{u^2}{2}I_{n\beta} + R_n(u)\}, \tag{9}$$

где при $n \to \infty$, $R_n(u) \overset{Q_{\theta_0}^{(n)}}{\to} 0$, $L(\omega_{n,\tau_n}/Q_{\theta_0}^{(n)}) \to L(\zeta_0)$, и $\zeta_0 \overset{D}{=} N(0,1)$. Из представления (9) следует свойство ЛАН семейства распределений $\{Q_{\theta_0}^{(n)}, \theta \in \Theta\}$ в точке θ_0.

ЛИТЕРАТУРА

1. Абдушукуров А.А. Статистика неполных наблюдений. Ташкент: Университет. 2009. 296 с.

2. Abdushukurov A.A., Nurmukhamedova N.S. Local approximate normality of likelihood ratio statistics in competing risks model under random censorship from both sides. // Far East J. of Theoretical Statistics. 2013. v.42. N.2.p.107-122.

3. Abdushukurov A.A., Nurmukhamedova N.S. Locally asymptotically normality of the family of distributions by incomplete observations. // Journal of Siberian Federal University. Mathematics & Physics. 2014. v.7. N.2. p.141-154.

4. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. М.: Наука. 1979.527с.

5. Леман Э. Проверка статистических гипотез. М.: Наука. 1964. 408 с.

6. Русас Дж. Контигуальность вероятностных мер. М. : Мир. 1975.254с.

On Empirical Process of Independence with Estimated Parameter and its Application to Testing of Proportional Hazards Model

L.R.Kakadjanova
National University of Uzbekistan
leyla_tvms@rambler.ru

Abstract. We consider the convergence of empirical processes of independence of random elements and events indexed by functions that depend on an estimated parameter and show that its application to Proportional hazard model.

Keywords: empirical process, independence, metric entropy, Glivenko-Cantelly and Donsker theorems, proportional hazards

1. Introduction

Folowing [1-2] consider a sequence of experiments in which observed data consist of pairs $\{(X_k, A_k), k \geq 1\}$, where X_k be independent and identically distributed random elements in a measurable space $(\mathfrak{X}, \mathfrak{B})$ with probability law $\mathbb{P}$ and A_k are events with common probability $p = \mathbb{P}(A_k) \in (0,1)$. Let $\delta_k = I(A_k)$ be an indicator of the event A_k. At the $n-$ th stage of experiments observed data is $\mathbb{S}^{(n)} = \{(X_k, \delta_k), 1 \leq k \leq n\}$. Each pair (X_k, δ_k) induced a statistical model with sample space $\mathfrak{X} \otimes \{0,1\}$ with σ-algebra $\mathcal{G}$ of sets $B \times D$ and distribution $\mathbb{Q}^*(\cdot)$ on $(\mathfrak{X} \otimes \{0,1\}, \mathcal{G})$:

$$\mathbb{Q}^*(B \times D) = \mathbb{P}(X_k \in B, \delta_k \in D), B \in \mathfrak{B}, D \subset \{0,1\}.$$

Our interests are focused on submeasures $\mathbb{Q}_m(B) = \mathbb{Q}^*(B \times \{m\})$, $m = 0,1$ and $\mathbb{Q}(B) = \mathbb{Q}_0(B) + \mathbb{Q}_1(B) = \mathbb{Q}^*(B \times \{0,1\})$, $B \in \mathfrak{B}$. From practical point of view, it is important to know the occurrence of hypothesis $\mathcal{H}$ about independence of random elements X_k and event A_k for each $k \geq 1$. In order to verify this we use the signed measure $\Lambda(B) = \mathbb{Q}_1(B) - p\mathbb{Q}(B)$, $B \in \mathfrak{B}$, where $p = \mathbb{Q}_1(\mathfrak{X})$. Then validity of $\mathcal{H}$ is equivalent to the equality $\Lambda(B) = 0$ for any $B \in \mathfrak{B}$. For constructing the test statistics for testing the hypothesis $\mathcal{H}$ we introduce the empirical analogue of measure $\Lambda(B)$ for any $B \in \mathfrak{B}$ as $\Lambda_n(B) = \mathbb{Q}_{1n}(B) - p_n\mathbb{Q}_n(B)$, where

$$p_n = \mathbb{Q}_{1n}(\mathfrak{X}), \mathbb{Q}_n(B) = \mathbb{Q}_{0n}(B) + \mathbb{Q}_{1n}(B),$$

$$\mathbb{Q}_{0n}(B) = \frac{1}{n}\sum_{k=1}^{n}(1-\delta_k)I(X_k \in B), \mathbb{Q}_{1n}(B) = \frac{1}{n}\sum_{k=1}^{n}\delta_k I(X_k \in B) \qquad (1)$$

- empirical counterparts of p, $\mathbb{Q}_(B)$ and $\mathbb{Q}_m(B), m = 0,1$, respectively. According to the strong law of large numbers (SLLN) for each $B \in \mathfrak{B}$ at

$n \to \infty$, $\Lambda_n(B) \overset{\text{a.s.}}{\to} \Lambda(B)$ and consequently under validity of $\mathcal{H}$, $\Lambda_n(B) \overset{\text{a.s.}}{\to} 0$. Then naturally becomes investigation of limit behaviors of normalized signed empirical process $\{\chi_n = a_n(\Lambda_n(B) - \Lambda(B)), B \in \mathcal{J}\}$, indexed by certain class $\mathcal{J}$ of sets from $\mathfrak{B}$, where $\{a_n, n \geq 1\}$ is a (possible random) sequence of positive numbers. In the previous works [1,2] we investigated the specially normalized empirical process of independence indexed by the class $\mathcal{F}$ of measurable functions f on $\mathfrak{X}$, which coincides with χ_n when $f = I(\cdot)$ is an indicator function.

2. Empirical Process of Independence Indexed by Class of Measurable Functions

For a signed measure G and class $\mathcal{F}$ of Borel measurable functions $f : \mathfrak{X} \to \mathbb{R}$ introduce the integral

$$Gf = \int_{\mathfrak{X}} f dG, f \in \mathcal{F}.$$

Let's introduce the following $\mathcal{F}$-indexed extensions of empiricals (1) for $f \in \mathcal{F}$:

$$\mathbb{Q}_{0n} f = \frac{1}{n} \sum_{k=1}^{n} (1 - \delta_k) f(X_k), \quad \mathbb{Q}_{1n} f = \frac{1}{n} \sum_{k=1}^{n} \delta_k f(X_k),$$

$$\mathbb{Q}_n f = \mathbb{Q}_{0n} f + \mathbb{Q}_{1n} f = \frac{1}{n} \sum_{k=1}^{n} f(X_k), \tag{2}$$

and $\Lambda_n f = \mathbb{Q}_{1n} f - p_n \mathbb{Q}_n f$, where $p_n = \mathbb{Q}_{1n} 1 = \mathbb{Q}_{1n}(\mathfrak{X}) = \frac{1}{n} \sum_{k=1}^{n} \delta_k$. Observe that formulas (1) are special cases of (2) when $\mathcal{F} = \{I(B), B \in \mathcal{J}\}$. We define $\mathcal{F}$-indexed empirical process $G_n : \mathcal{F} \to \mathbb{R}$ as

$$f \mapsto G_n f = \sqrt{n} (\mathbb{Q}_n - \mathbb{Q}) f = n^{-1/2} \sum_{k=1}^{n} (f(X_k) - \mathbb{Q} f), f \in \mathcal{F}. \tag{3}$$

Here $G_n f = G_{0n} f + G_{1n} f$ with subempirical processes

$$G_{jn} f = \sqrt{n} (\mathbb{Q}_{jn} - \mathbb{Q}_j) f, \quad j = 0, 1, f \in \mathcal{F}. \tag{4}$$

For a given f by SLLN and central limit theorem (CLT)

$$(a) \quad \mathbb{Q}_n f \xrightarrow[n \to \infty]{a.s.} \mathbb{Q} f \text{ as } \mathbb{Q}|f| < \infty; \tag{5}$$

$$(b) \quad G_n f \Rightarrow Gf \overset{d}{=} N\left(0, \sigma^2_{\mathbb{Q}}(f)\right), \quad n \to \infty \text{ as } \mathbb{Q} f^2 < \infty, \tag{6}$$

where $\sigma_{\mathbb{Q}}^2(f) = \mathbb{Q}(f - \mathbb{Q}f)^2$.

Note that the uniformly variants for special classes $\mathcal{F}$ of measurable functions of statements (5) and (6) have a very solid theory (see, for example, [3,10]). There are different extensions of classical theorems of Glivenko-Cantelli and Donsker for $\mathcal{F}$-indexed empirical processes (3) under certain conditions on the set $\mathcal{F}$ of measurable functions. These conditions ensure the convergence of $n^{-1/2}\|G_n f\|_{\mathcal{F}} = \sup\left\{n^{-1/2}\left|G_n f\right|, f \in \mathcal{F}\right\}$ either in probability or almost surely to zero. These classes $\mathcal{F}$ are called the weak or strong Glivenko-Cantelli classes respectively. Donsker-type theorems provide a general conditions on $\mathcal{F}$ in order to get a weak convergence

$$G_n f \Rightarrow Gf \quad \text{in} \quad l^{\infty}(\mathcal{F}), \tag{7}$$

where $l^{\infty}(\mathcal{F})$ is a space of all bounded functions $f : \mathfrak{X} \to \mathbb{R}$ equipped with the supremum-norm $\|.\|_{\mathcal{F}}$ (see, [8], p.81). Class $\mathcal{F}$ for which holds the convergence (7) is called a Donsker class. The limiting field $\{Gf,\ f \in \mathcal{F}\}$ in (7) is called a $\mathbb{Q}$-Brownian bridge. It is a tight Borel measurable element of $l^{\infty}(\mathcal{F})$ and is a Gaussian field with zero mean and covariance function

$$\operatorname{cov}(Gf, Gg) = \mathbb{Q}fg - \mathbb{Q}f\mathbb{Q}g, \quad f, g \in \mathcal{F}. \tag{8}$$

Remind that $\mathbb{Q}$-Brownian bridge $\{Gf, f \in \mathcal{F}\}$ can be represented in distribution though $\mathbb{Q}$-Brownian sheet $\{W(f), f \in \mathcal{F}\}$ with zero mean and covariance

$$\operatorname{cov}(W(f), W(g)) = \mathbb{Q}fg, \quad f, g \in \mathcal{F}, \tag{9}$$

by distributional equality

$$Gf \stackrel{d}{=} W(f) - W(1)\,\mathbb{Q}f, \quad f \in \mathcal{F}. \tag{10}$$

By SLLN under conditions $\mathbb{Q}_j|f| < \infty,\ j = 0, 1$ for given f:

$$\Lambda_n f \underset{n\to\infty}{\overset{a.s.}{\to}} \Lambda f \overset{under\ \mathcal{H}}{=} 0. \tag{11}$$

Moreover, for a given f variable $\sqrt{n}(\Lambda_n - \Lambda) f$ being a linear functional of subempirical processes (4) under conditions $\mathbb{Q}_j f^2 < \infty,\ j = 0, 1$, have a limiting normal distribution $N\left(0, \sigma_{\mathbb{Q}}^2(f)\right)$. In [1] the authors have proved uniform SLLN and CLT for a specially normalized empirical $\mathcal{F}$-indexed process

$$\left\{\Delta_n f = \left(\frac{n}{p_n(1-p_n)}\right)^{1/2}(\Lambda_n - \Lambda) f, \ \ f \in \mathcal{F}\right\}, \tag{12}$$

and showed that the limiting distribution is $\mathbb{Q}$-Brownian bridge $\{Gf,\ f \in \mathcal{F}\}$ with covariance (8).

3. Asymptotical results

Let $\mathcal{L}_q(\mathbb{Q})$ - the space of functions $f : \mathfrak{X} \to \mathbb{R}$ with the norm

$$\|f\|_{\mathbb{Q},q} = \left(\mathbb{Q}|f|^q\right)^{1/q} = \left\{\int\limits_{\mathfrak{X}} |f|^q d\mathbb{Q}\right\}^{1/q}.$$

To prove the $\mathcal{F}$ - uniform variants of theorems of Glivenko-Cantelli and Donsker we define the complexity or entropy of class $\mathcal{F}$. To determine the entropy it is necessary to define a concept of $\varepsilon-$ brackets. $\varepsilon-$ bracket in $\mathcal{L}_q(\mathbb{Q})$ is a pairs of functions $\varphi, \psi \in \mathcal{L}_q(\mathbb{Q})$ such that $\mathbb{Q}\left(\varphi(X) \le \psi(X)\right) = 1$ and $\|\psi - \varphi\|_{\mathbb{Q},q} \le \varepsilon$, i.e. $\mathbb{Q}(\psi - \varphi)^q \le \varepsilon^q$. Function $f \in \mathcal{F}$ is in the (or covered by) bracket $[\varphi, \psi]$, if $\mathbb{Q}\left(\varphi(X) \le f(X) \le \psi(X)\right) = 1$. Note that the functions φ and ψ may not belong to the class $\mathcal{F}$, but they must have finite norms. Bracketing (or covering) number $N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ is the minimum number of $\varepsilon-$ brackets in $\mathcal{L}_q(\mathbb{Q})$ needed to cover $\mathcal{F}$ (see [8,9]):

$$N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right) = \min \begin{cases} k: \text{ for some } f_1, ..., f_k \in \mathcal{L}_q\left(\mathbb{Q}\right), \\ \mathcal{F} \subset \bigcup\limits_{i,j} [f_i, f_j] : \|f_j - f_i\|_{\mathbb{Q},q} \le \varepsilon. \end{cases}$$

Number $H_q\left(\varepsilon\right) = \log N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ is called the metric entropy with bracketing of the class $\mathcal{F}$ in $\mathcal{L}_q(\mathbb{Q})$. Number $H_{jq}\left(\varepsilon\right) = \log N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}_j\right)\right), j = 0, 1$ denotes the metric entropy of a class $\mathcal{F}$ in $\mathcal{L}_q(\mathbb{Q}_j)$, $j = 0, 1$, respectively. To prove the weak convergence of $\mathcal{F}-$ indexed empirical processes (12) we introduce the integral of the metric entropy with bracketing as

$$J^{(q)}_{j[]}\left(\delta\right) = J_{j[]}\left(\delta; \mathcal{F}; \mathcal{L}_q\left(\mathbb{Q}_j\right)\right) = \int\limits_0^\delta \left(H_{jq}\left(\varepsilon\right)\right)^{1/2} d\varepsilon, j = 0, 1, \text{ for } 0 < \delta < 1.$$

Recall that numbers $N_{[]}\left(\cdot\right)$ converge to $+\infty$ at $\varepsilon \downarrow 0$. However, it necessary for Donsker's theorem that they converge not very fast to $+\infty$. This speed is measured by the integrals $J^{(q)}_{j[]}\left(\delta\right)$ (see [8,9]).

The following theorem shows validity of Glivenko-Cantelli type theorem for the process $\{\Delta_n f,\ f \in \mathcal{F}\}$. Here $*$ - means a.s. convergence by outer probability.

Theorem 3.1. [1] *Let the class* $\mathcal{F}$ *such that*

$$N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_1\left(\mathbb{Q}_j\right)\right) < \infty, j = 0, 1. \tag{13}$$

Then under validity of the hypothesis $\mathcal{H}$ *and at* $n \to \infty$

$$\left\| n^{-1/2} \Delta_n f \right\|_{\mathcal{F}}^{*} \overset{a.s.}{\to} 0. \tag{14}$$

To prove the weak convergence of the process (12) to a Gaussian process, we first investigate the limiting properties of two-dimensional empirical field $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g)\,,\ f, g \in \mathcal{F}\}$, where $\mathbb{A}_n f = = n^{1/2}\left(\mathbb{Q}_n - \mathbb{Q}\right) f$ and $\mathbb{A}_{1n} g = n^{1/2}\left(\mathbb{Q}_{1n} - \mathbb{Q}_1\right) g$.

Theorem 3.2. [1] *Let the class* $\mathcal{F}$ *such that*

$$\mathcal{F} \subset \mathcal{L}_2(\mathbb{Q}_j) \text{ and } J_{j[]}^{(2)}(1) < \infty,\ j = 0, 1. \tag{15}$$

Then for $n \to \infty$ *sequence* $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g)\,,\ f, g \in \mathcal{F}\}$ *of* $\mathcal{F} \to \mathbb{R}^2$ *maps weak converge in* $l^{\infty}(\mathcal{F}) \times l^{\infty}(\mathcal{F})$ *to the two-dimensional Gaussian field* $\{(\mathbb{A} f, \mathbb{A}_1 g)\,,\ f, g \in \mathcal{F}\}$ *with zero mean and covariance structure for* $f, g \in \mathcal{F}$*:*

$$\begin{gathered} \mathbb{E}\left(\mathbb{A} f \cdot \mathbb{A} g\right) = \mathbb{Q} f g - \mathbb{Q} f \mathbb{Q} g, \\ \mathbb{E}\left(\mathbb{A}_1 f \cdot \mathbb{A}_1 g\right) = \mathbb{Q}_1 f g - \mathbb{Q}_1 f \mathbb{Q}_1 g, \\ \mathbb{E}\left(\mathbb{A} f \cdot \mathbb{A}_1 g\right) = \mathbb{Q}_1 f g - \mathbb{Q} f \mathbb{Q}_1 g. \end{gathered} \tag{16}$$

Remark. In the last formula of covariance in (16), in particular at $g \equiv 1$ we have $\mathbb{Q}_1 1 = p$ and

$$\mathbb{E}\left(\mathbb{A} f \cdot \mathbb{A}_1 1\right) = \mathbb{Q}_1 f - p\, \mathbb{Q} f,\ f \in \mathcal{F}. \tag{17}$$

Hence, when hypothesis $\mathcal{H}$ is valid, then the covariance (17) is equal to zero for all $f \in \mathcal{F}$. Thus, under the hypothesis $\mathcal{H}$, the Brownian bridge $\{\mathbb{A}\ f,\ f \in \mathcal{F}\}$ and r.v. $\mu_0 = \mathbb{A}_1 1$ with a normal distribution $\mathcal{N}\left(0, p\left(1 - p\right)\right)$ are independent.

In study of process (12) basic property of weak convergence to a $\mathbb{Q}-$ Brownian bridge is contained in the following statement.

Theorem 3.3. [1] *Under the conditions of Theorem 3.2 for* $n \to \infty$

$$\Delta_n f \Rightarrow \Delta f \text{ in } l^{\infty}(\mathcal{F}), \tag{18}$$

where $\{\Delta f,\ f \in \mathcal{F}\}$ *is a Gaussian field with zero mean and under validity of the hypothesis* $\mathcal{H}$*, it coincides in distribution with* $\mathbb{Q}-$ *Brownian bridge.*

Now we consider the convergence of empirical process of independence (12) when the class $\mathcal{F}$ of measurable functions $f_{\theta,\eta} : \mathfrak{X} \to \mathbb{R}$ indexed by sets Θ and $\mathcal{K}$: $\mathcal{F} = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{K}\}$, where η is an estimated parameter. Let η_n be an estimator of η is a random element with values in $\mathcal{K}$ defined on the same probability space as $X_1, ..., X_n$ and $\eta_0 \in \mathcal{K}$ is a fixed element, which is limit in probability of the sequence η_n. In several applications it is interesting to prove that, as $n \to \infty$,

$$\sup_{\theta\in\Theta} |\Delta_n(f_{\theta,\eta_n} - f_{\theta,\eta_0})| \xrightarrow{p} 0. \tag{19}$$

We say that η_n is consistent for η_0, if

$$\sup_{\theta\in\Theta} \left|\Delta_n(f_{\theta,\eta_n} - f_{\theta,\eta_0})^2\right| \xrightarrow[n\to\infty]{p} 0. \tag{20}$$

Theorem 3.4. *If conditions (15) and (20) holds, then (19) is valid.*

The result (19) can be applicated to the estimaton of the functional $\theta \mapsto \Delta f_{\theta,\eta}$. When the parameter η is unknown, we can replace it by an estimator η_n an use the estimator $\Delta_n f_{\theta,\eta_n}$. This result helps to derive the limiting behaviour of this estimator by using the decomposition

$$(\Delta_n f_{\theta,\eta_n} - \Delta f_{\theta,\eta_0}) = \Delta_n(f_{\theta,\eta_n} - f_{\theta,\eta_0}) + \Delta_n f_{\theta,\eta_0} + \sqrt{\frac{n}{p_n(1-p_n)}}\Delta_n(f_{\theta,\eta_n} - f_{\theta,\eta_0}). \tag{21}$$

In right side of decomposition (21) the first term converges to zero by (19), the second term will converge to a Gaussian process by theorem 3 and the third term converges to zero by (20).

4. Application to Proportional Hazard Model

Let's consider a right random censoring model, where $X_i = \min\{T_i, C_i\}$ and $A_i = \{T_i \le C_i\}$. Here r.v.-s T_i and C_i denotes life times and censoring times, mutually independent with common continuous distribution functions F and G respectively ($F(0) = G(0) = 0$). Then in observing data $\mathbb{S}^{(n)} = \{(X_i, \delta_i), 1 \le i \le n\}$ with $\delta_i = I(A_i)$, r.v.-s of interest T_i are observed when A_i occurs, i.e. $\delta_i = 1$. Observe that the X_i have common distribution function $H = 1 - (1 - F)(1 - G)$ and subdistributions defined as

$$\mathbb{Q}_0(B) = \mathbb{P}(X_k \in B, \delta_k = 0) = \mathbb{P}(C_k \in B \cap [0, T_k)) \int_B (1 - F(t))G(dt),$$

$$\mathbb{Q}_1(B) = \mathbb{P}(X_k \in B, \delta_k = 1) = \mathbb{P}(T_k \in B \cap [0, C_k]) \int_B (1 - G(t))F(dt). \tag{22}$$

Now consider simple proportional hazards model (PHM) or Koziol-Green model which has been found in statistical literature to be very useful and interesting in practice (see, for example, [1] and many recently papers). In PHM assume the parametric relation

$$1 - G = (1 - F)^{\beta} \text{ for some } \beta > 0. \tag{23}$$

Taking into (23), it is easy see that $1 - F = (1 - H)^p$, where $p = \frac{1}{1+\beta} = \mathbb{P}(A_k)$. One of basic properties of PHM that (23) holds when r.v.-s X_k and δ_k are independent and only this case. Such characterization of PHM plays a basic role in constructing and investigating of estimators of many functionals of distribution F. The following sufficient maximum likelihood estimator of F was first introduced and studied by Abdushukurov (1984) in PHM:

$$F_n(t) = 1 - (1 - H_n(t))^{p_n}, \tag{24}$$

where $H_n(t) = \frac{1}{n}\sum\limits_{k=1}^{n} I(X_k \leq t)$ and $p_n = \frac{1}{n}\sum\limits_{k=1}^{n} \delta_k$ are independent empirical estimators of $H(t)$ and p respectively.

There are many papers devoted to statistical analysis of F_n. These papers are concerned with the superiority of methods for estimation and the testing in PHM are based on F_n rather than the product-limit estimator of Kaplan-Meier. Hence the question arises as to when the advantages of PHM can be used. In other words, there is now a need for testing of validity of PHM, i.e., for the composite hypothesis described by relation (23). But this relation is equivalent to hypothesis $\mathcal{H}$ on independence of r.v.-s $(X_1, ..., X_n)$ and $(\delta_1, ..., \delta_n)$. Let us consider following special empirical process of (12), when $f = I(\cdot)$:

$$\Delta_n(t) = \left(\frac{n}{p_n\,(1 - p_n)}\right)^{1/2} (H_{1n}(t) - p_n H_n(t))\,,\ -\infty < t < \infty, \tag{25}$$

where $H_{1n}(t) = \frac{1}{n}\sum\limits_{k=1}^{n} I(X_k \leq t, \delta_k = 1)$. Then we have next consequence from Theorem 3.3: if $\mathcal{H}$ holds then as $n \to \infty$

$$\Delta_n\,(\cdot) \Rightarrow \mathbb{B}\,(H\,(\cdot))\,, \tag{26}$$

where $\{\mathbb{B}(y), 0 \leq y \leq 1\}$ is a Brownian bridge. Several statistics for testing $\mathcal{H}$ may be considered based on convergence (26). Moreover, by Theorem 4 one can also consider more general classes of statistics using $\mathcal{F}-$ indexed processes by suitable choosing of functions $f_{\theta,\eta}$ that are more flexible in applications than (25).

Acknowledgements

I would like sincerely congratulate my supervisor professor A.A. Abdushukurov with the 60-year anniversary and to thank him for his invaluable help and support in my scientific work.

REFERENCE

1. Abdushukurov A.A., Kakadjanova L.R. A class of special empirical process of independence. *J. of Siberian Federal Univ. Math. Phys.* 2015 Vol. **8**, N.2., pp. 1-9.

2. Abdushukurov A.A., Kakadjanova L.R. Sequential empirical process of independence. *J. of Siberian Federal Univ. Math. Phys.* 2018 Vol. **11**, N.5., pp. 634-643.

3. Alexander K.S. Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* 1984 Vol. **12**, N.4., pp. 1041-1067.

4. Dudley R.M. Central limit theorems for empirical measures. *Ann. Probab.* 1978 Vol. **6**, pp. 899-929.

5. Gaensler P., Stute W. Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab.* 1979 Vol. **7**, N.2., pp. 193-243.

6. Gine E., Zinn J. Some limit theorems for empirical processes. *Ann. Probab.* 1984 Vol. **12**, N.4., pp. 929-989.

7. Shorack G.R., Wellner J.A. *Empirical processes with applications to statisticsk*. John Wiley&Sons. 1986

8. Van der Vaart A.W., Wellner J.A. *Weak convergence and empirical processes*. Springer. 1986

9. Van der Vaart A.W. *Asymptotic Statistics*. Cambridge University Press, Cambridge. 1998

10. Van der Vaart A.W., Wellner J.A. IMS Lecture Notes-Monograph Series. Asymptotics: Particless, Processes and Inverse Problems. 2007 Vol. **55**, pp. 234-252.

Kakadjanova Leyla Reshitovna (Tashkent, Uzbekistan)

On empirical process of independence with estimated parameter and its application to testing of proportional hazards model

Abstract. *We consider the convergence of empirical processes of independence of random elements and events indexed by functions that depend on an estimated parameter and show that its application to Proportional hazard model.*

Key words: *empirical process, independence, metric entropy, Glivenko-Cantelly and Donsker theorems, proportional hazards*

Estimation of Conditional Distribution and Mean Residual Life Function under Dependent Right Random Censored Data

A.A.Abdushukurov[1], N.T.Dushatov[2]
[1]Moscow State University, Tashkent Branch,
[2] National University of Uzbekistan, Tashkent, Uzbekistan
a_ abdushukurov@rambler.ru, n_ dushatov@rambler.ru/headrd/

Introduction. In such research areas as bio-medicine, engineering, insurance, social sciences, . . ., researchers are interested in positive variables, which are expressed as a time until a certain event. For example, in medicine the survival time of individual, while in industrial trials, time until break down of a machine are non-negative random variables (r.v.-s) of interest. But in such practical situations, the observed data may be incomplete, that is censored. This is the case, for example, in medicine when the event of interest-death due to a given cause and the censoring event is death due to other cause. In industrial study, it may occur that some piece of equipment is taken away (that is censored) because it shows some sign of future failure. Moreover, the r.v.-s of interest (lifetimes, failure times) and censoring r.v.-s usually can be infuenced by other variable, often called prognostic factor or covariate. In medicine, dose of a drug and in engineering some environmental conditions (temperature, pressure, . . .) are influenced to the observed variables. The basic problem consist in estimation of distribution of lifetime by such censored dependent data. The aim of paper is considering this problem in the case of right random censoring model in the presence of covariate.

Let's consider the case when the support of covariate C is the interval $[0,1]$ and we describe our results on fixed design points $0 \leq x_1 \leq x_2 \leq ... \leq x_n \leq 1$ at which we consider responses (survival or failure times) $X_1, ..., X_n$ and censoring times $Y_1, ..., Y_n$ of identical objects, which are under study. These responses are independent and nonnegative r.v.-s with conditional distribution function (d.f.) at x_i,$F_{x_i}(t) = P(X_i \leq t/C_i = x_i)$. They are subjected to random right censoring, that is for X_i there is a censoring variable Y_i with conditional d.f. $G_{x_i}(t) = P(Y_i \leq t/C_i = x_i)$ and at $n-$th stage of experiment the observed data is

$$S^{(n)} = \{(Z_i, \delta_i, C_i), 1 \leq i \leq n\},$$

where $Z_i = min(X_i, Y_i), \delta_i = I(X_i \leq Y_i)$ with $I(A)$ denoting the indicator of event A. Note that in sample S^n r.v. X_i is observed only when $\delta_i = 1$. Commonly, in survival analysis to assume independence between the r.v.-s X_i and Y_i conditional on the covariate C_i. But, in some practical situations, this

assumption does not hold. Therefore, in this article we consider a dependence model in which dependence structure is described through copula function. So let

$$S_x(t_1, t_2) = P(X_x > t_1, Y_x > t_2),\ t_1, t_2 \geq 0,$$

the joint survival function of the response X_x and the censoring variable Y_x at x. Then the marginal survival functions are ${S_x}^X(t) = 1 - F_x(t) = S_x(t, 0)$ and ${S_x}^Y(t) = 1 - G_x(t) = S_x(0, t), t \leq 0$. We suppose that the marginal d.f.-s F_x and G_x are continuous. Then according to the Theorem of Sclar (see, [1]), the joint survival function $S_x(t_1, t_2)$ can be expressed as

$$S_x(t_1, t_2) = C_x(S_x^X(t_1), S_x^X(t_2))\ ,\ t_1, t_2 \geq 0, \tag{1}$$

where $C_x(u, v)$ is a known copula function depending on x, S_x^X and S_x^Y in a general way. It is necessary to note that in the case of no covariates, this idea first was considered by Zeng and Klein [2] and proposed copula-graphic estimator. Rivest and Wells [3] investigated copula- graphic estimator and derived a closed form expression for estimator when the joint survival function (1) is modeled an Archimedean copula. The copula-graphic estimator is then shown to be uniformly consistent and asymptotically normal. Note that the copula-graphic estimator is equivalent to the product-limit estimator of Kaplan and Meier [4] when the survival and censoring times are assumed to be independent. Braekers and Veraverbeke [5] extend copula- graphic estimator to the fixed design regression case and show that estimator has an asymptotic representation and a Gaussian limit. We consider other estimator of d.f. F_x which had a simpler form than copula-graphic estimator and it is also equivalent to the usual relative-risk power estimator [11,12] under independent censoring case. We study the large sample properties of estimator proposed and present result of uniform normality with the same limiting Gaussian process as for copula-graphic estimator.

Construction of estimator and asymptotic results

Assume that at the fixed design value $x \in (0, 1), C_x$ in (1) is Archimedean copula, i.e.

$$S_x(t_1, t_2) = \varphi_x^{[-1]}(\varphi_x(S_x^X(t_1)) + \varphi_x(S_x^Y(t_2))),\ t_1, t_2 \geq 0, \tag{2}$$

where, for each $x, \varphi_x : [0, 1] \to [0, +\infty]$ is a known continuous, convex, strictly

decreasing function with $\varphi_x = 0$. $\varphi_x^{[-1]}$ is a pseudo-inverse of φ_x (see, [1]) and given by

$$\varphi_x^{[-1]}(s) = \begin{cases} \varphi_x^{-1}(s), & 0 \le s \le \varphi_x(0), \\ 0, & \varphi_x(0) \le s \le \infty. \end{cases}$$

We assume that copula generator function φ_x is strict, i.e. $\varphi_x(0) = \infty$ and hence $\varphi_x^{[-1]} = \varphi_x^{-1}$. From (2), it follows that

$$P(Z_x > t) = 1 - H_x(t) = \overline{H_x(t)} = S_x{}^Z(t) = S_x(t,t) = \\ = \varphi_x^{-1}(\varphi_x(S_x^X(t)) + \varphi_x(S_x^Y(t))),\ t \ge 0, \tag{3}$$

Let $H_x^{(1)}(t) = P(Z_x \le t, \delta_x = 1)$ be a subdistribution function and $\Lambda_x(t)$ is crude hazard function of r.v. X_x subjecting to censoring by Y_x,

$$\Lambda_x(dt) = \frac{P(X_x \in dt, X_x \le Y_x)}{P(X_x \ge t, Y_x \ge t)} = \frac{H_x^{(1)}(dt)}{S_x^Z(t-)}. \tag{4}$$

From (4) one can obtain following expression of survival function S_x^X :

$$S_x^X(t) = \varphi_x^{-1}[-\int_0^t S_x^Z(u-)\varphi_x^{'}(S_x^Z(u))d\Lambda_x(u)] = \\ = \varphi_x^{-1}[-\int_0^t \varphi_x^{'}(S_x^Z(u))dH_x^{(1)}(u)],\ t \ge 0, \tag{5}$$

(see, for example, [3,5]). In order to constructing the estimator of S_x^X according to representation (5), we introduce some smoothed estimators of $S_x^Z, H_x^{(1)}$ and regularity conditions for them. Similarly to Breakers and Veraverbeke [5], we will also use the Gasser-Müller weights

$$w_{ni}(x, h_n) = \frac{1}{q_n(x, h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n}\pi(\frac{x-z}{h_n})dz,\ i = 1, ..., n, \tag{6}$$

with

$$q_n(x, h_n) = \int_0^{x_n} \frac{1}{h_n}\pi(\frac{x-z}{h_n})dz,$$

where $x_0 = 0$, π is a known probability density function (kernel) and $\{h_n, n \ge$

1} is a sequence of positive constants, tending to zero as $n \to \infty$, called bandwidth sequence. Let's introduce the weighted estimators of H_x, S_x^Z and $H_x^{(1)}$ respectively as

$$\begin{aligned} H_{xh}(t) &= \sum_{i=1}^{n} w_{ni}(x, h_n) I(Z_i \le t), \\ S_{xh}^Z(t) &= 1 - H_{xh}(t), \\ H_{xh}^{(1)}(t) &= \sum_{i=1}^{n} w_{ni}(x, h_n) I(Z_i \le t, \delta_i = 1). \end{aligned} \tag{7}$$

Then pluggin in (5) estimators (7) Abdushukurov and Muradov [10] obtained following estimator of S_x^X :

$$S_{xh}^X(t) = 1 - F_{xh}(t) = \varphi_x^{-1}[-\int_0^t \varphi_x^{'}(S_x^Z(u)) dH_x^{(1)}(u)], \ t \ge 0, \tag{8}$$

Remark that in the case of no covariate, estimator (8) reduces to estimator first obtained by Zeng and Klein [2]. In the case of the independent copula $\varphi(y) = -logy$, Zeng and Klein estimate reduces to a exponential-hazard estimate (see, [8,9]). Also it is well-known that under independent censoring case Kaplan-Meier's product-limit estimator and exponential-hazard estimators are asymptotical equivalent. Therefore, we will show that estimator (8) and copula-graphic estimator of Breakers and Veraverbeke have the same asymptotic behaviours.

In this work we propose the next analogue of estimator introduced in [7] in the existence of covariate :

$$\widehat{S}_{xh}^Z(t) = \varphi_x^{-1}[\varphi(S_{xh}^Z(t)) \cdot \mu_{xh}(t)] = 1 - \widehat{F}_{xh}(t), \tag{9}$$

where

$$\mu_{xh}(t) = \varphi(S_{xh}^X(t))/\varphi(\widetilde{S}_{xh}^Z(t)),$$

$$\varphi(S_{xh}^X(t)) = -\int_0^t \varphi_x^{'}(S_{xh}^Z(u)) dH_{xh}^{(1)}(u),$$

$$\varphi(\widetilde{S}^Z_{xh}(t)) = -\int_0^t \varphi_x^{'}(S^Z_{xh}(u))dH_{xh}(u).$$

Note that in the case of no covariate, estimator (9) reduces to estimator proposed in [7]. In the case of the independent copula estimator of authors [7] reduces to a relative risk power estimator of Abdushukurov [11,12], which have a some peculiarities with respect to product-limit and exponential-hazard estimators. In order to investigate the estimate (9) we introduce some conditions. For the design points $x_1, ..., x_n$, denote

$$\underline{\Delta_n} = \min_{1\le i\le n}(x_i - x_{i-1}),\ \overline{\Delta_n} = \max_{1\le i\le n}(x_i - x_{i-1}).$$

For the kernel π, let

$$\|\pi\|_2^2 = \int_{-\infty}^{\infty}\pi^2(u)du,\ m_\nu(\pi) = \int_{-\infty}^{\infty}u^\nu\pi(u)du,\ \nu = 1,2,\ \|\pi\|_\infty = \sup_{u\in R}\pi(u).$$

Moreover, we use next assumptions on the design and on the kernel function:
(A1) As $n\to\infty, x_n\to 1, \underline{\Delta_n} = O(\frac{1}{n}), \overline{\Delta_n} - \underline{\Delta_n} = o(\frac{1}{n})$.
(A2) π is a probability density function with compact support $[-M, M]$ for some $M > 0$, with $m_1(\pi) = 0$ and $|\pi(u) - \pi(u^{'})| \le C(\pi)|u - u^{'}|$, where $C(\pi)$ is some constant.

Let $T_{H_x} = inf t \ge 0 : H_x(t) = 1$. Then $T_{H_x} = min(T_{F_x}, T_{G_x})$. For our results we need some smoothnees conditions on functions $H_x(t)$ and $H_x^{(1)}(t)$. We formulate them for a general (sub)distribution function $N_x(t), 0\le x\le 1, t\in R$ and for a fixed $T > 0$.
(A3) $\frac{\partial}{\partial x}N_x(t) = \dot{N}_x(t)$ exists and is continuous in $(x,t)\in[0,1]\times[0,T]$.
(A4) $\frac{\partial}{\partial x}N_x(t) = N_x^{'}(t)$ exists and is continuous in $(x,t)\in[0,1]\times[0,T]$.
(A5) $\frac{\partial^2}{\partial x^2}N_x(t) = \ddot{N}_x(t)$ exists and is continuous in $(x,t)\in[0,1]\times[0,T]$.
(A6) $\frac{\partial^2}{\partial t^2}N_x(t) = N_x^{''}(t)$ exists and is continuous in $(x,t)\in[0,1]\times[0,T]$.
(A7) $\frac{\partial^2}{\partial x\partial t}N_x(t) = \dot{N}^{'}{}_x(t)$ exists and is continuous in $(x,t)\in[0,1]\times[0,T]$.
(A8) $\frac{\partial\varphi_x(u)}{\partial u} = \varphi_x^{'}(u)$ and $\frac{\partial^2\varphi_x(u)}{\partial u^2} = \varphi_x^{''}(u)$ are Lipschitz in the $x-$direction with a bounded Lipschitz constant and $\frac{\partial^3\varphi_x(u)}{\partial u^3} = \varphi_x^{'''}(u)$ exists and is continuous in $(x,u)\in[0,1]\times(0,1]$.

It is clear that for existence of right hand side of representation (5) we must require the conditions (A 4) for functions $H_x(t)$ and $H_x^{(1)}(t)$ in $[0,1]\times[0,T]$ with $T < T_{H_x}$ and existence of $\varphi_x^{'}(u)$ on $[0,1]\times(0,1]$.

We derive an almost sure representation result with rate.

Theorem 1. Assume (A1), (A2), $H_x(t)$ and $H_x^{(1)}(t)$ satisfy (A5)-(A7) in $[0,T]$ with $T < T_{H_x}$, φ_x satisfies (A8) and $h_n \to 0, \frac{logn}{nh_n} \to 0, \frac{nh_n^5}{logn} = O(1)$. Then, as $n \to \infty$,

$$\widehat{F}_{xh}(t) - F_x(t) = \sum_{i=1}^{n} w_{ni}(x, h_n)\Psi_{tx}(Z_i, \delta_i) + r_n(t),$$

where

$$\Psi_{tx}(Z_i,\delta_i) = \frac{-1}{\varphi_x^{'}(S_x^X(t))}[\int_0^t \varphi_x^{''}(S_x^Z(u))(I(Z_i \le u) - H_x(u))dH_x^{(1)}(u) -$$
$$-\varphi_x^{'}(S_x^Z(t))(I(Z_i \le t, \delta_i = 1) - H_x^{(1)}(t)) -$$
$$-\int_0^t \varphi_x^{''}(S_x^Z(u))(I(Z_i \le u, \delta_i = 1) - H_x^{(1)}(u))dH_x(u)],$$

and

$$\sup_{0\le t\le T} |r_n(t)| \overset{a.s.}{=} O((\frac{\log n}{nh_n})^{3/4}).$$

The weak convergence of the empirical process $(nh_n)^{1/2}\{\widehat{F}_{xh}(\cdot) - F_x(\cdot)\}$ in the space $l^\infty[0,T]$ of uniformly bounded functions on $[0,T]$, endowed with the uniform topology is the contents of the next theorem.

Theorem 2. Assume (A1), (A2), $H_x(t)$ and $H_x^{(1)}(t)$ satisfy (A5)-(A7) in $[0,T]$ with $T < T_{H_x}$, and that φ_x satisfies (A8).

(I) If $nh_n^5 \to 0$ and $\frac{(logn)^3}{nh_n} \to 0$, then, as $n \to \infty$,

$$(nh_n)^{1/2}\{\widehat{F}_{xh}(\cdot) - F_x(\cdot)\} \Rightarrow \mathbf{W}_x(\cdot) \ in \ l^\infty[0T].$$

(II)If $h_n = Cn^{-1/5}$ for some $C > 0$, then, as $n \to \infty$,

$$(nh_n)^{1/2}\{\widehat{F}_{xh}(\cdot) - F_x(\cdot)\} \Rightarrow \mathbf{W}_x^*(\cdot) \ in \ l^\infty[0T],$$

where $\mathbf{W}_x(\cdot)$ and $\mathbf{W}_x^*(\cdot)$ are Gaussian processes with means

$$E\mathbf{W}_x(t) = 0, E\mathbf{W}_x^*(t) = a_x(t),$$

and same covariance

$$Cov(\mathbf{W}_x(t), \mathbf{W}_x^*(s)) = Cov(\mathbf{W}_x^*(t), \mathbf{W}_x^*(s)) = \Gamma_x(t,s),$$

with

$$a_x(t) = \frac{-C^{5/2} m_2(\pi)}{2\varphi_x^{'}(S_x^X(t))} \int_0^t [\varphi_x^{''}(S_x^Z(u)) \ddot{H}_x(u) dH_x^{(1)}(u) - \varphi_x^{'}(S_x^Z(u)) d\ddot{H}_x^{(1)}(u)\,],$$

and

$$\Gamma_x(t,s) = \frac{\|\pi\|_2^2}{\varphi_x^{'}(S_x^X(t))\varphi_x^{'}(S_x^X(s))} \{ \int_0^{\min(t,s)} (\varphi_x^{'}(S_x^Z(z)))^2 dH_x^{(1)}(z) +$$

$$+ \int_0^{\min(t,s)} [\varphi_x^{''}(S_x^Z(w)) S_x^Z(w) + \varphi_x^{'}(S_x^Z(w))] \int_0^w \varphi_x^{''}(S_x^Z(y)) dH_x^{(1)}(y) dH_x^{(1)}(w) +$$

$$+ \int_0^{\min(t,s)} \varphi_x^{''}(S_x^Z(w)) \int_w^{\max(t,s)} (\varphi_x^{''}(S_x^Z(y)) S_x^Z(y) + \varphi_x^{'}(S_x^Z(y))) dH_x^{(1)}(y) dH_x^{(1)}(w) -$$

$$- \int_0^t [\varphi_x^{''}(S_x^Z(y)) S_x^Z(y) + \varphi_x^{'}(S_x^Z(y))] dH_x^{(1)}(y) \cdot$$

$$\cdot \int_0^s [\varphi_x^{''}(S_x^Z(w)) S_x^Z(w) + \varphi_x^{'}(S_x^Z(w))] dH_x^{(1)}(w) \}.$$

Using the statistics (9) we can introduce corresponding estimator of the mean conditional residual lifetime function

$$\mu_x(t) = (1 - F_x(t))^{-1} \int_t^{+\infty} (1 - F_x(s)) ds, t > 0,$$

as follows:

$$\widehat{\mu}_{xh}(t) = (1 - \widehat{F}_x(t))^{-1} \int_t^{+\infty} (1 - \widehat{F}_x(s)) ds, t > 0. \qquad (10)$$

We have proved an uniform consistency result for estimator (10).
Let's introduce the weighted uniform measure

$$\varepsilon_{xh}(F_x) = \sup_{0 \le t < \infty} \chi(F_x(t)) \left| \widehat{\mu}_{xh}(t) - \mu_x(t) \right|,$$

where weight function $\chi : [0,1] \to \overline{R}^+$ satisfies the following conditions:

(A9) Function χ is measurable and for every $\eta > 0$:

$$\sup\{\chi(u) : u \in [0, 1-\eta]\} < \infty;$$

(A10) Function $\chi^*(u) = \chi(u)/(1-u)$ is nondecreasing in a neighborhood of 1;
(A11) For $T_x = \{\sup\ t : \ S_x^X(t) > 0\}$, let

$$\int_{-\infty}^{T_x} \{(1-F_x(t))^{-1}[\int_t^{T_x} \chi(F_x(s))ds]\}dF_x(t) < \infty.$$

Theorem 3. Assume conditions of theorem 1, (A9)-(A11) and $\mu_x(0) < \infty$. Then for $n \to \infty$, $\varepsilon_{xh}(F_x) \xrightarrow{P} 0$.
We discuss also the weak convergence of normed process

$$(nh_n)^{1/2}(\widehat{\mu}_{xh}(t) - \mu_x(t))$$

to the Gaussian process.

REFERENCE

1. A.A. Abdushukurov Survival function estimation from fixed design regression model in the presence of dependent random censoring. // Proceeding of the Internat.Workshop "AMSA-2019 18-20 September, 2019. p. 85-89.

2. M.Zeng, J.P.Klein Estimates of marginal survival for dependent competing risks based on an assumed copula, Biometrika, 1995, 82, pp.127-138.

3. L.P.Rivest, M.T.Wells A martingall approach to the copula-graphic estimator for the survival function under dependent censoring, Journal of Multivariate Analysis, 2001, 79, pp.138-155.

4. E.L.Kaplan, P.Meier Nonparametric estimation from incomplete observations, Journal of American Statistical Association, 1958, 53, pp.457-481.

5. R.Breakers, N.Veraverbeke A copula-graphic estimator for the conditional survival function under dependent censoring, The Canadian Journal of Statistics, 2005, 33, No.3, pp.429-447.

6. T.R.Fleming, D.P.Harrington Counting Processes and Survival Analysis, Wiley, New York, 1991.

7. R.S.Muradov, A.A.Abdushukurov Estimation of multivariate distributions and its mixtures, LAMBERT Academic Publishing, 2011 (in Russian).

8. A.A.Abdushukurov, R.S.Muradov Estimation of survival and mean residual life functions from dependent random censored data, New Trends in Mathematical Sciences, 2014, 2, No.1, pp.35-48.

9. I.Van Keilegom, N.Veraverbeke Estimation and the bootstrap with censored data in fixed design nonparametric regression, Ann. Inst. Statist. Math.,1997, 49, No.3, pp.467-491.

10. Abdushukurov A.A., Muradov R.S. On estimation of conditional distribution function under dependent random right censored data. J. Siberian Federal Univ.: Math.Phys., 2014. V.7.Issue 4. pp.409-416.

11. Abdushukurov A.A. Nonparametric estimation of distribution function based on relative risk function. Commun. Statist.: Theory and Methods. 1998. V.27. No.8. pp.1991-2012.

12. Abdushukurov A.A. On nonparametric estimation of reliability indices by censored samples. Theory Probab. Appl. 1999. V.43.No.1.pp.3-11.

Статистический анализ основных показателей сельского хозяйства

Р.Ибрагимов, А.Машраббоев, Ю.Ташмирзаев

Наманганский Государственный Университет

Цель данной работы определить статистические параметры связи посевной площади сельскохозяйственных культур с продукцией сельского хозяйства в сумах а также растеневодства и животноводства в Узбекистане в период с 2000 г. до 2018 г. в виде эмпирических статистических характеристик приведена основные показатели сельского хозяйства за 18 лет (таблица1.)

Таблица 1.

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Посевная площадь сельскохозяйственных культур, тыс. Гектар - X	3778,3	3444,5	3540,8	3790,1	3695,7	3647,5	3637,4	3560,3	3609,7
Продукция сельского хозяйства, млрд. Сум - Y	1387,2	2104,8	3255,3	4083,3	4615,8	5978,3	7538,8	9304,9	11310,7
в том числе:									
Растениеводства - t	696,8	1086,0	1648,8	2102,0	2432,1	3323,1	4215,0	5170,1	6400,1
Животноводства - z	690,4	1018,8	1606,5	1981,3	2183,7	2655,2	3323,8	4134,8	4910,6
Темпы роста производства продукции сельского хозяйства, в процентах к предыдущему году - v	103,1	104,2	106,0	107,3	108,9	105,4	106,7	106,1	104,5

2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018*)**
3608,6	3708,4	3601,6	3628,1	3658,6	3678,2	3694,2	3706,7	3474,5	3396,0
13628,6	30856,7	45285,9	55750,0	66435,3	81794,3	99604,6	115599,2	148199,3	192699,2
8089,0	18119,0	25874,6	30592,3	36237,4	43194,3	55429,2	61755,1	83303,4	102495,1
5539,6	12737,7	19411,3	25157,7	30197,9	38600,0	44175,4	53844,1	64895,9	90204,1
105,8	106,3	106,2	107,2	106,6	106,3	106,1	106,3	101,0	100,2

Используя данные таблицы 1 приведем статистичекий анализ в виде таблицы 2.

Таблице 2.

X			
Среднее	3619,222222		
Стандартная ошибка	24,29430604		
Медиана	3609		
Мода	3609		
Стандартное отклонение	103,0720113		
Дисперсия выборки	10623,83951		
Эксцесс	-0,27188727		
Асимметричность	-0,480745958		
Интерваль	394		
Минимум	3396		
Максимум	3790		
Сумма	65146		
Счет	18		
Доверительный интервал	3619,222222	±	47,61684
Уровень надежности(95,0 %)	127,809294		

Y			
Среднее	46920,94444		
Стандартная ошибка	13311,81598		
Медиана	12470		
Мода	9305		
Стандартное отклонение	56477,25211		
Дисперсия выборки	3189680005		
Эксцесс	0,367971689		
Асимметричность	1,224476089		
Интервал	191312		
Минимум	1387		
Максимум	192699		
Сумма	844577		
Счет	18		
Доверительный интервал	46920,94444	±	26091,15933
Уровень надежности(95,0 %)	70031,79261		

T			
Среднее	19634,3889		
Стандартная ошибка	5836,50865		
Медиана	5225		
Мода	4135		
Стандартное отклонение	24762,2091		
Дисперсия выборки	613166997		
Эксцесс	1,48437346		
Асимметричность	1,529116		
Интервал	89514		
Минимум	690		
Максимум	90204		
Сумма	353419		
Счет	18		
Доверительный интервал	19634,3889	±	11439,55695
Уровень надежности(95,0 %)	30705,1392		

V			
Среднее	106		
Стандартная ошибка	0,538631095		
Медиана	107		
Мода	106		
Стандартное отклонение	2,2852182		
Дисперсия выборки	5,222222222		
Эксцесс	0,344952467		
Асимметричность	-0,363110724		
Интервал	10		
Минимум	101		
Максимум	111		
Сумма	1908		
Счет	18		
Доверительный интервал	106	±	1,055717
Уровень надежности(95,0 %)	2,833670568		

Z			
Среднее	22714,33333		
Стандартная ошибка	6429,451287		
Медиана	7245		
Мода	5170		
Стандартное отклонение	27277,85162		
Дисперсия выборки	744081189,2		
Эксцесс	1,58218552		
Асимметричность	1,478249961		
Интервал	101797		
Минимум	698		
Максимум	102495		
Сумма	408858		
Счет	18		
Доверительный интервал	22714,33333	±	12601,7245
Уровень надежности(95,0 %)	33824,53601		

Определим зависимость этих выборок вычислением коэффициента корреляции по формуле

$$R_{xy} = \frac{\sum_{i=1}^{18} x_i y_i - 18\bar{x}\,\bar{y}}{18\sigma_x\sigma_y},$$

где $\overline{x}$ и $\overline{y}$ средние, а σ_x, σ_y среднеквадратические отклонения x и y.
Имеем

$$R_{xy} = \frac{3017232178 - 18 * 3618 * 46921}{18 * 103 * 56477} = -0,36722$$

$$R_{yz} = 0,991, R_{yt} = 0,994, R_{yv} = 0,27. \quad (1)$$

Из этих показателей следует, что зависимость между x и y слабая т.е.рост x и y почти несравнимы. Зависимость от z и t линейна, что естественно. Зависимость от v почти несравнима, т.е. они почти независимы. Воспользуясь результатами таблицы 2 а также (1) по формуле

$$y_x = R_{xy}\frac{\sigma_y}{\sigma_x}x + \bar{y} - R_{xy}\frac{\sigma_y}{\sigma_x}$$

определим уравнения зависимости (уравнения регрессии) между y, z и t.

$$y_z = 2,052z + 46919$$

$$y_t = 2,199t + 46919$$

Аналогично можно вычислить зависимость x от z и t.

Заключение:

1) При анализе статистических показателей можно использовать таблицу 2.

2)Зависимость посевных площадей сельскохозяйственных культур от сумм продукции сельскохозяйственных культур достаточно мала.

$G_k^{(2)}-$ Периодические основные состояния для модели ПОТТС-SOS на дереве Кэли

М.М.Рахматуллаев [1], М.А.Расулова [2]

[1] Наманганский Государственный университет,

[2] Наманганский академический лицей МВД РУз

mrahmatullaev@rambler.ru, m_rasulova_a@rambler.ru

Каждой мере Гиббса сопоставляется одна фаза физической системы. Если существует более чем одна мера Гиббса, то говорят, что существуют фазовые переходы. При низких температурах, периодическому основному состоянию соответствует периодическая мера Гиббса (см.[1],[2]). Поэтому, естественно возникает задача описания периодических основных состояний.

Пусть $\tau^k = (V, L), k \geq 1$ есть дерево Кэли порядка k, где $V-$ множество вершин, $L-$ множество ребер τ^k. Известно, что τ^k можно представить как G_k- свободное произведение $k+1$ циклических групп второго порядка(см.[1]).

Через $S_1(x)$ обозначим множество всех ближайших соседей точки $x \in G_k$, т.е. $S_1(x) = \{y \in G_k : \langle x, y\rangle\}$.

Мы рассматриваем модели, где спин принимает значения из множества $\Phi = \{0, 1, 2, ..., m\}$, $m \geq 1$. Тогда конфигурация σ на V определяется как функция $x \in V \mapsto \sigma(x) \in \Phi$; множество всех конфигураций совпадает с $\Omega = \Phi^V$.

Определим G_k^*-периодическую конфигурацию $\sigma(x)$, которая является инвариантной относительно подгруппы $G_k^* \subset G_k$ конечного индекса, т.е. $\sigma(yx) = \sigma(x)$ для любых $x \in G_k, y \in G_k^*$. Конфигурация, которая является инвариантной относительно всех переходов, называется трансляционно-инвариантной.

В работе рассматривается случай $m = 2$, то есть $|\Phi| = 3$.

Гамильтониан модели Поттс-SOS имеет вид

$$H(\sigma) = J \sum_{\langle x,y\rangle \in L} |\sigma(x) - \sigma(y)| + J_p \sum_{\langle x,y\rangle \in L} \delta_{\sigma(x)\sigma(y)}, \qquad (1)$$

где $J, J_p \in R-$ константы не равны нулю.

Пусть $M-$ множество единичных шаров с вершинами в V. Мы назовем сужение конфигурации σ на шаре $b \in M$ ограниченной конфигурацией σ_b.

Определим энергию конфигурации σ_b на b следующим образом:

$$U(\sigma_b) = \frac{1}{2}(J \sum_{\substack{\langle x,y\rangle: \\ x,y\in b}} |\sigma(x) - \sigma(y)| + J_p \sum_{\substack{\langle x,y\rangle: \\ x,y\in b}} \delta_{\sigma(x)\sigma(y)}). \tag{2}$$

Символом c_b обозначим центр единичного шара b, и пусть $B_t = \{x \in S_1(c_b) : \varphi_b(x) = t$ для всех $t \in \Phi$.

Легко доказать следующую лемму.

Лемма. *Для каждой конфигурации φ_b мы имеем*

$$U(\varphi_b) \in \{U_{i,n} : i = 0, 1, 2, ..., k+1, n = 0, ..., k+1-i\},$$

где

$$U_{i,n} = \frac{J}{2}(k + 1 + n - i) + \frac{J_p}{2}i. \tag{3}$$

Определение. Конфигурация φ называется основным состоянием для гамильтониана H, если

$$U(\varphi_b) = \min\{U_{i,n} : i = 0, 1, 2, ..., k+1, n = 0, ..., k+1-i\}$$

для любого $b \in M$.

Обозначим $C_{i,n} = \{\varphi_b : U(\varphi_b) = U_{i,n}\}$ и

$$A_{\xi,\eta} = \{(J, J_p) \in R^2 : U_{\xi,\eta} = \min\{U_{i,n} : i = 0, 1, 2, ..., k+1, n = 0, ..., k+1-i\}\}. \tag{4}$$

Цель этой работы при $|A| = k + 1$ описать H_A-периодические основные состояния для модели Поттс-SOS на дереве Кэли порядка $k \geq 2$.

С помощью (4) найдем следующие множества:

$$A_{0,0} = \{(J, J_p) \in R^2 : J_p \geq J \geq 0\},\ A_{0,k+1} = \{(J, J_p) \in R^2 : J_p \geq 2J, J \leq 0\}.$$

Пусть $A \subset \{1, 2, ..., k+1\}$. $H_A = \{x \in G_k : \sum_{j\in A} w_j(x) -$ четно$\}$, где $w_j(x)-$ число a_j в слове x. H_A- является нормалным делителем индекса 2.

Рассмотрим фактор группу $G_k/H_A = \{H_A, G_k \setminus H_A\}$. Для простоты обозначим $H_0 = H_A, H_1 = G_k \setminus H_A$. H_A-периодические конфигурации имеют следующий вид:

$$\varphi(x) = \begin{cases} l, \text{ если } x \in H_0, \\ \\ s, \text{ если } x \in H_1, \end{cases}$$

где $l, s \in \Phi$.

Теорема. *Пусть* $|\Phi| = 3$ *и* $|A| = k+1$. *Для модели Поттс-SOS справедливы следующие утверждения:*
1) на множестве $A_{0,0}$ *существуют четыре* H_A-*периодического основного состояния, которые являются* $G_k^{(2)}-$*периодическими (но не являющимся трансляционно-инвариантными);*
2) на множестве $A_{0,k+1}$ *существуют два* H_A-*периодического основного состояния, являющихся* $G_k^{(2)}-$*периодическими (но не являющихся трансляционно-инвариантными).*

Доказательство. Докажем сначала 1). Рассмотрим φ_1:

$$\varphi_1(x) = \begin{cases} l, \text{ если } x \in H_0, \\ \\ s, \text{ если } x \in H_1, \end{cases}$$

где $|l-s| = 1, l, s \in \Phi$.

Пусть $c_b \in H_0$, тогда $\varphi_{1,b}(c_b) = l, |B_l| = 0, |B_s| = k+1$; следовательно, $\varphi_{1,b} \in C_{0,0}$.

Пусть $c_b \in H_1$, тогда $\varphi_{1,b}(c_b) = s, |B_l| = k+1, |B_s| = 0$; следовательно, $\varphi_{1,b} \in C_{0,0}$.

Значит, на множестве $A_{0,0}$ данная периодическая конфигурация φ_1 является $G_k^{(2)}-$периодическим основным состоянием. Количество таких конфигураций равно 4.

Докажем 2). Рассмотрим φ_2:

$$\varphi_2(x) = \begin{cases} l, \text{ если } x \in H_0, \\ \\ s, \text{ если } x \in H_1, \end{cases}$$

где $|l-s| = 2, l, s \in \Phi$.

Пусть $c_b \in H_0$, тогда $\varphi_{2,b}(c_b) = l, |B_l| = 0, |B_s| = k+1$; следовательно, $\varphi_{2,b} \in C_{0,k+1}$.

Пусть $c_b \in H_1$, тогда $\varphi_{2,b}(c_b) = s, |B_l| = k+1, |B_s| = 0$; следовательно, $\varphi_{2,b} \in C_{0,k+1}$.

Значит, на множестве $A_{0,k+1}$ данная периодическая конфигурация φ_2 является $G_k^{(2)}-$периодическим основным состоянием. Количество таких кон-

фигураций равно 2.

ЛИТЕРАТУРА

1. Rozikov U. A. Gibbs measures on Cayley trees. World scientific, 2013.

2. Рахматуллаев М. М., Расулова М. А. Периодические и слабо периодические основные состояния для модели Поттса с конкурирующими взаимодействиями на дереве Кэли. Математические труды, 2015. том 18, № 2, с. 112–132.

Rakhmatullaev Muzaffar Mukhammadjanovich (Namangan, Uzbekistan), Rasulova Mukhayo (Namangan, Uzbekistan)

$G_k^{(2)}-$periodic ground states for the Potts-SOS model on the Cayley tree

Abstract. *We consider the Potts-SOS model on the Cayley tree of order k with $k \geq 2$. We describe the sets of $G_k^{(2)}-$periodic ground states.*

Key words: *Cayley tree, Potts-SOS model, periodic ground states*

Convergence rates in the law of large numbers for random variables with multidimensional indices

M.K. Holmurodov, R. Mamadjanov
Namangan State University

Let Z_+^d, where $d \geq 2$ is an integer, denote the positive integer d-dimensional lattice points and let $\left\{X_n, n \in Z_+^d\right\}$ be a set of independent identically distributed random variables. The notation $m \prec n$, where $m = (m_1, m_2, ..., m_d)$ and $n = (n_1, n_2, ..., n_d) \in Z_+^d$ means that $m_i \leq n_i, i = 1, 2, ..., d, |n|$ is used for $\prod_{i=1}^d n_i$and $n \to \infty$ is to be interpreted as $n_i \to \infty, i = 1, 2, ..., d$.

Define $d\,(x) = card\left\{n \in Z_+^d,\ |n| = [x]\right\}$ and $M(x) = \sum_{k=1}^{[x]} d(k)$. We have $M(x) = O\left(x\left(\log^+ x\right)^{d-1}\right)$ and $d\,(x) = o(x^\delta)$ as $x \to \infty$ for any $\delta > 0$, where for x> 0. For x > 0 we set log+ x = max{1, log x}.

Throughout the paper we assume that the random variables $X_n, n \in Z^d$are independent and $EX_n = 0, EX_n^2 = 1, EX_n^2 = \sigma_n^2 < \infty, n \in Z^d$. For $n \geq 1$, define the partial sums

$$S(n) = \sum_{k \prec n} X_k.$$

The distribution of S(n) has been studied by Smythe [9], Gut [1-5], Sirajidinov , Gafurov[7], Klesov[8], Holmurodov [6] etc.

Theorem 1. Let > 1, and suppose that {X_k, k Z} are i.i.d. random variables with E(X) = 0 and partial sums . The following are equivalent:

$$E\exp\{(\log|X|)^{\alpha/\gamma}\}\,|X|^{(2l-1)/\gamma}\,(\log^+|X|)^{\alpha+d-1} < \infty;$$

$$\sum_n \exp\{(\log|n|)^\alpha\} \cdot \frac{(\log|n|)^{\alpha-1}}{|n|^{2(\ell-1)}} P(|S_n| > |n|^\gamma \varepsilon) < \infty \quad for\ \ all\ \ \varepsilon > 1;$$

$$\sum_n \exp\{(\log|n|)^\alpha\} \cdot \frac{(\log|n|)^{\alpha-1}}{|n|^{2(\ell-1)}} P(\max_{k \leq n}|S_k| > |n|^\gamma \varepsilon) < \infty \quad for\ \ all\ \ \varepsilon > 1;$$

$$\sum_n \exp\{(\log|n|)^\alpha\} \cdot \frac{(\log j)^{\alpha-1}}{j^{2(\ell-1)}} P(\sup_{j \leq |k|}|S_k/|k|^\gamma| > \varepsilon) < \infty \quad for\ \ all\ \ \varepsilon > 1.$$

This result generalizes results of work [5].($\gamma = 1, \ell = 2$).

Theorem 2. Let 0 < < 1, and suppose that {X_k, k Z} are i.i.d. random variables with E(X) = 0 and partial sums +. The following are equivalent:

$$E\exp\{|X|)^{\alpha/\gamma}\}\,|X|^{(\alpha-1)/\gamma}\,(\log^+|X|)^{\alpha+d-1} < \infty;$$

$$\sum_n \exp\{|n|^\alpha\} \cdot |n|^{\alpha-2} P(|S_n| > |n|^\gamma \varepsilon) < \infty \quad for \ \ all \ \ \varepsilon > 1;$$

$$\sum_n \exp\{|n|^\alpha\} \cdot |n|^{\alpha-2} P(\max_{k \le n} |S_k| > |n|^\gamma \varepsilon) < \infty \quad for \ \ all \ \ \varepsilon > 1;$$

$$\sum_n \exp\{j^\alpha\} \cdot j^{\alpha-2} P(\sup_{j \le |k|} |S_k/|k|^\gamma| > \varepsilon) < \infty \quad for \ \ all \ \ \varepsilon > 1.$$

Remark 3.1 The equivalence with respect to the moment assumptions should be interpreted as in our earlier results. This result generalizes results of work [5].($\gamma = 1$).

Once again, we refer to the original source Gut and Stadtmьller (2011c) for proofs and further details.

The proofs of the theorems amount to rather straightforward generalizations of those in [2, 3] and are omitted, except for the following extension of Lemma 2.1.

Lemma 3.1 For any random variable X and $\gamma > 0$,

$$E \exp\{(\log^+ |X/\gamma|)^\alpha\} (\log^+ |X|)^{d-1} < \infty$$

$$\Leftrightarrow \textstyle\sum_n \exp\{(\log |n|)^\alpha\} \frac{(\log |n|)^{\alpha-1}}{|n|} P(|X| > |n|\gamma) < \infty;$$

$$E \exp\{|X/\gamma|^\alpha\} (\log^+ |X|)^{d-1} < \infty \Leftrightarrow \sum \exp\{|n|^\alpha\} \cdot |n|^{\alpha-1} P(|X| > |n|\gamma| < \infty.$$

The basis of the proof of the lemma is, again, partial summation, together with the fact that terms with equisized indices are equal, viz., we may write

$$\sum_n ... = \sum_{j=1}^{\infty} \sum_{|n|=j} d(j)...,$$

where

$$d(j) = Card \ \ \{k : |k| = j\}, \quad j \ge 1.$$

Using this device the sums in the lemma turn into

$$\textstyle\sum_n \exp\{|n|)^\alpha\} \cdot |n|)^{\alpha-1} P(|X| > |n| = \sum_{j=1}^{\infty} d(j) \exp\{j^\alpha\} \cdot j^{\alpha-1} P(|X| > j),$$

$$\textstyle\sum_n \exp\{(\log |n|^\alpha\} \frac{(\log^+ |n|)^{\alpha-1}}{|n|} P(|X| > |n| = \sum_{j=1}^{\infty} d(j) \exp\{(\log j)^\alpha\} \frac{(\log j)^{\alpha-1}}{j} P(|X| > j).$$

respectivly, after it remains to connect these sums of the respective tail probabilities to the appro-priate moment (cf. [4], Section 2.12, [6]- [9]).

In order to do so we also need the quantity

$$M(j) = Card\ \{k : |k| = j\}\ \ (= \sum_{k=1}^{j} d(k))\ \ j \geq 1,$$

with its asymptotics

$$\frac{M(j)}{j(\log j)^{d-1}} \to \frac{1}{(d-1)!}\quad as\quad j \to \infty.$$

For details concerning these number theoretical matters we refer to [5].

REFERENCES

1. Gut A. Marsinkiwiz laws and convergence rates in the law of large numbers for random variables with multidimensional indices // Ann. Probability. – 1978.– v.6. –P. 469-482.
2. Gut A. Convergence rates for probabilities of moderate deviations for sums of random variables with multidimensional indices // Ann. Probability. – 1980. – V.8. – P. 298-313.
3. Cut.A.Convergance rates in the central limit theorem for multidimensionally indexed random variables. U.U.D.M.report.2007.
4.Gut, A. Probability: A Graduate Course, Corr. 2nd printing. Springer-Verlag, New York.2007.
5.Gut A., Stadtmuller. On the Hsu-Robbins-Erdos -Spitzer-Baum-Katz theorem for random fields. Uppsala Universitet.U.U.D.M.2010,11. - P. 21.
6.M.K. Holmurodov. On probabilistic one-sided deviations for sums of independent random variables with multidimensional indices. Fan. Tashkent. 1985. 207-215.
7. S.Kh.Sirajidinov , M.U.Gafurov. The Method of Series in Limit Prombles for Random Walks. Fan. Tashkent. 1987.
8. Klesov O.I. Strong law of large numbers for random fields: Dis. Candidate. Physics and Mathematics. Kiev: 1981.123s.
9. Smythe, R. Strong laws of large numbers for r-dimensional arrays of random vari-ables. Ann. Probab.1973. 1, 164-170.

Математическое моделирование показателей иммунной защиты организма, свертывающей системы крови и липидного обмена в реабилитации больных церебральным атеросклерозом

Ф.М. Холмуродов, С.А.Болтабоев

Наманганский государственный университет

Цель исследования. Математическое моделирование изменения показателей иммунной защиты организма, свертывающей системы крови и липидного обмена у больных церебральным атеросклерозом.

Материал и методы. Под наблюдением находилось 40 больных церебральным атеросклерозом в среднем возрасте 53 года. Из них 16 (40%) мужчин и 24 (60%) женщин. Давность заболевания от 1 года до 10 лет.

Клинико-функциональное состояние больных до и после лечебно-реабилитационных программ оценивалось по состоянию клеточного и гуморального иммунитета [- изучалось содержание в сыворотки крови СД3-лимфоцитов (Жондал), СД4 и СД8 (Моретта), СД19 (Е.Р.Кудрявцева) и G(IgG), A(IgA), M(IgM) иммуноглобулинов (Манчини)] и следующими биохимическими методами исследования: - ПТИ (Туголукова), фибриноген (Рутберга); - [обшей холестерин-ОХС (Илька), триглицериды-Тг, ?-липопротеиды (ХС-ЛПОНП, ХС-ЛПНП) и ?-липопротеиды (ХС-ЛПВП) (Брунштейн-Самай)].

С целью решения поставленной задачи разработана 5-ти этапная комплекс лечебно-реабилитационная программа больных рассчитанная на 2 года. Все исследуемые группы больных в качестве базисной терапии включали следующие процедуры: - иммунокорректор-тимоптин; - ловастатин; - витамин-Е; - лечебных танцевальных физических упражнений в виде танцев например: андижанская полка, тановар. Нагрузку увеличивали день за днем постепенно, учитывая индивидуальные особенности, подготовленности и здоровье больных в течении 2 лет; - антисклеротичешская фитотерапия (в течении 2 лет); - йода-бромная бальнеотерапия: йода-бромная минеральная ванна назначена при температуре воды 37^o, продолжительность 10 минут, через день, на курс 8 процедур (курс повторили вначале 2-го года).

Результаты. До лечения у всех больных выявлено Т и В иммунодефицитное состояние (в виде снижения клеточной и повышения гуморальной иммунной защиты организма), гиперкоагуляция, гиперхолистеринемия и гиперлипидемия.

После проводимых лечебно-реабилитационных мероприятий отмечался достоверное повышение средних показателей клеточной [СД3-18,5%, СД4-19,5%, СД8-23,4%, СД19-52,9% (P<0,01)] и достоверное снижение средних показателей гуморальной [IgA-37,3%, IgM-41,2%, IgG-40,4% (P<0,05)] иммунной защиты организма. А со стороны показателей свертывающей системы крови и липидного спектра обмена произошла достоверное снижение средних показателей соответственно: ПТИ-15,6%, фибриноген-26,3%, ОХС–21,5%, Тг–27,5% (P<0,01), ХС-ЛПОНП–33,3%, ХС-ЛПНП–19,7% (P<0,05) и достоверное повышение средних показателей ?-липопротеидов ХС-ЛПВП–189,3% (P<0,001).

Статистические методы анализа данных применяются практически во всех областях деятельности человека.

Целесообразно выделить три вида научной и прикладной деятельности в области статистических методов анализа данных (по степени специфичности методов, сопряженной с погруженностью в конкретные проблемы):

а) разработка и исследование методов общего назначения, без учета специфики области применения;

б) разработка и исследование статистических моделей реальных явлений и процессов в соответствии с потребностями той или иной области деятельности;

в) применение статистических методов и моделей для статистического анализа конкретных данных.

Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В линейной множественной регрессии $\widehat{y}_x = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ параметры при x называются коэффициентами "чистой"регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

Рассмотрим линейную модель множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon. \qquad (1)$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных $\widehat{y}$ минимальна:

$$\sum_i (y_i - \widehat{y}_{x_i})^2 \quad \to \quad \min. \qquad (2)$$

Как известно из курса математического анализа, для того чтобы найти экстремум функции нескольких переменных, надо вычислить частные производные первого порядка по каждому из параметров и приравнять их к нулю.

Итак, имеем функцию $m+1$ аргумента:

$$S\left(a,\ b_1,\ b_2,\ ...,\ b_m\right) = \sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right)^2.$$

Находим частные производные первого порядка:

$$\begin{cases} \frac{\partial S}{\partial a} = -2\sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ \frac{\partial S}{\partial b_1} = -2\sum x_1\left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ ..... \\ \frac{\partial S}{\partial b_m} = -2\sum x_m\left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0. \end{cases}$$

После элементарных преобразований приходим к системе линейных нормальных уравнений для нахождения параметров линейного уравнения множественной регрессии (1):

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 + ... + b_m\sum x_m = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + ... + b_m\sum x_1x_m = \sum yx_1, \\ ..... \\ a\sum x_m + b_1\sum x_1x_m + b_2\sum x_2x_m + ... + b_m\sum x_m^2 = \sum yx_m. \end{cases}$$

Метод наименьших квадратов применим и к уравнению множественной регрессии в стандартизированном масштабе:

$$t_y = \beta_1t_{x_1} + \beta_2t_{x_2} + ... + \beta_mt_{x_m} + \varepsilon,$$

где $t_y,\ t_{x_1},\ ...,\ t_{x_m}$ – стандартизированные переменные: $t_y = \frac{y-\bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i-\bar{x}_i}{\sigma_{x_i}}$, для которых среднее значение равно нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β_i – стандартизированные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор x_i изменится на одну единицу при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов кчистойњ регрессии, которые несравнимы между собой.

Используя методов наименьших квадратов, вычислены коэффициенты и получены математические модели.

Математическая модель коэффициента иммунной защиты организма, свертывающей системы крови и липидного обмена определяется следующим образом: $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_7x_7 + a_8x_8 + a_9x_9 + a_{10}x_{10}$

где, y - иммунодефицитное состояние (в виде снижения клеточной и повышения гуморальной иммунной защиты организма);

x_1 – показатель СДЗ;x_2- средний показатель гуморальной иммунной защиты организма; x_3 - ПТИ; x_4 - фибриноген (сред); x_5 - ОХС; x_6 –Тг;

x_7 - ХС-ЛПОНП; x_8 - ХС-ЛПНП; x_9 - ХС-ЛПВП; x_{10}, a_0 , a_1, ..., a_{10} - постоянные числа.

Для определения коэффициентов a_0, a_1, ..., a_{10} используем метод наименьших квадратов математической статистики.

1. Показатель иммунодефицитного состояния:

y= 23,456+ 4,1601x_1+ 0,3024 x_2+ 3,06x_3+ 8,32x_4+ 0,1910 x_5+ 0,235 x_6+ 15,5 x_7+2,6 x_8+ 2,5 x_9+19,3 x_{10}

1. Показатель свертывающей системы крови и липидного спектра обмена:

y= 8,8523– 6,96x_1+ 0,31 x_2+ 9,18x_3+ 0,19 x_4+ 5,16 x_5+ 0,15 x_6+15,5 x_7+ +2,6 x_8+ 2,5 x_9+19,3 x_{10}

Табличное значение F-критерия при пятипроцентном уровне значимости ($\alpha = 0,05$, $k_1 = 10$, $k_2 = 40 - 10 - 1 = 29$): $F_{\text{B01;}} = 2,19$. Так как $F_{\text{D0:B}} = 28,76 > F_{\text{B01;}} = 2,19$, то уравнение признается статистически значимым.

Выводы. Таким образом, на основании выше изложенного можно заключить, что проводимое в комплексе лечебно-реабилитационная программа

оказывает иммунокоррегирующий, гипокоагуляционный, гипохолестеринемический и гиполипидемический эффект у больных церебральным атеросклерозом.

ЛИТЕРАТУРА

1. Болтабоев С.А. Рахимов Б.Х. Нишонов Г.Х. Внутривенное лазерное облучение крови у больных церебральным атеросклерозом. //Материалы V международной медико-фармацевтической конференции студентов и молодых учёных (82-й ежегодный научный форум). Всеукраинский медицинский журнал молодых учёных. 2008., вип. 10.

2. Новиков Д.А., Новочадов В.В. Статистические методы в медико-биологическом эксперименте (типовые случаи). - Волгоград: Издательство ВолГМУ, 2005. – 87 с.

Об одной реализации шара Ли

Г.Худайберганов, Ж.Ш.Абдуллаев
Национальный университет Узбекистана имени Мирзо Улугбека
gkhudaiberg@mail.ru, jonibekabdullayev1201@gmail.com

Рассмотрим $n-$мерное комплексное пространство $\mathbb{C}^n$, точками которого являются упорядоченные наборы n комплексных чисел $z=(z_1,z_2,...,z_n)$.

Область $\Re_{IV}^n$ (Шар Ли) образована $n-$мерными комплексными векторами z, удовлетворяющими условиям

$$\Re_{IV}^n=\left\{ z\in\mathbb{C}^n \ : \ |zz'|^2-2\overline{z}z'+1>0,\,|zz'|<1\right\}, \tag{1}$$

где $z'-$ транспонированный вектор $z=(z_1,z_2,...,z_n)$.

Это классическая область четвертого типа (по классификации Э. Картана [**1**]) или называется шаром Ли. Граница Шилова (остов) Γ_{IV}^n для области $\Re_{IV}^n$ определяется следующим образом:

$$\Gamma_{IV}^n=\left\{ z\in\mathbb{C}^n \ : \ |zz'|^2-2\overline{z}z'+1=0,\,|zz'|=1\right\}. \tag{2}$$

Трубой будущего в $\mathbb{C}^{n+1}$ называется неограниченная область, имеющая вид

$$\tau^+(n)=\left\{w\in\mathbb{C}^{n+1}:(Imw_{n+1})^2>(Imw_1)^2+...+(Imw_n)^2\,,Imw_{n+1}>0\right\}.$$

Граница $\partial\tau^+(n)$ области $\tau^+(n)$ состоит из

$$\partial\tau^+(n)=\left\{w\in\mathbb{C}^{n+1}:(Imw_{n+1})^2=(Imw_1)^2+...+(Imw_n)^2\,,Imw_{n+1}>0\right\}$$

и остова

$$\Gamma_{\tau^+(n)}=\left\{w\in\mathbb{C}^{n+1}: \ \ Imw_1=\ ...=Imw_n=Imw_{n+1}=0\right\}=\mathbb{R}^{\mathrm{n+1}},$$

на котором граница вырождается.

Наша цель – найти отображение в явном виде, устанавливающее биголоморфную эквивалентность областей $\Re_{IV}^n$ и $\tau^+(n-1)$.

В работах [2-6] рассмотрет вопрос существование биголоморфной эквивалентности $\tau^+(n)$ и классической области четвертого типа, (т.е. шаром Ли $\Re_{IV}^{n+1}$).

Теорема 1. *Отображение* $\Phi:\ \mathbb{C}_z \to \mathbb{C}_w$ *определяемое соответствиями*

$$w_k = \frac{-2iz_k}{\sum\limits_{k=1}^{n-1} z_k^2 + (z_n - i)^2},\ \ k = 1,\ ...,\ n-1,\ \ w_n = \frac{2\,(z_n - i)}{\sum\limits_{k=1}^{n-1} z_k^2 + (z_n - i)^2} - i, \quad (3)$$

биголоморфно отображает область $\Re_{IV}^n$ *на* $\tau^+(n-1)$, *при этом* Γ_{IV}^n *переходит в* $\Gamma_{\tau^+(n-1)}$.

Из (3) найдем обратное отображение $\Psi = \Phi^{-1}:\ \mathbb{C}_w \to \mathbb{C}_z$ каторе определяется как

$$z_k = \frac{-2iw_k}{\sum\limits_{k=1}^{n-1} w_k^2 - (w_n + i)^2},\ \ k = 1,\ ...,\ (n-1),\ \ z_n = i - \frac{2\,(w_n + i)}{\sum\limits_{k=1}^{n-1} w_k^2 - (w_n + i)^2}. \quad (4)$$

Мы назавем преобразовании (3) «*обобщенным преобразованием Кэли*».

Теперь вычислим Якобианы преобразования (3) и (4). Введем обозначение

$$W = \sum_{k=1}^{n-1} w_k^2 - (w_n + i)^2. \quad (5)$$

Заметим, что

$$W = -4Z^{-1}, \text{ где } Z = \sum_{k=1}^{n-1} z_k^2 + (z_n - i)^2.$$

Теорема 2. *Якобианы преобразования* Φ *вида (3) и* Φ^{-1} *вида (4) вычисляются по формулам*

$$J_\Phi(z) = 2^n (-i)^{n+1} Z^{-n}.$$

и

$$J_{\Phi^{-1}}(w) = -2^n (-i)^{n+1} W^{-n}.$$

Известно, что всякое биголоморфное отображение $\Phi:\ D \to G$ устанавливает групповой изоморфизм $AutD$ и $AutG$ по формуле

$$\widetilde{\Phi}: \varphi \to \Phi \circ \varphi \circ \Phi^{-1},\ \varphi \in AutD, \quad (6)$$

т.е., изоморфность групп $AutD$ и $AutG$ необходима для голоморфной эквивалентности областей D и G(но не достаточно) (см. [7, с. 62]).

Тем самым, используя теорему 1 и с помощью формулы (6) мы получим транзитивную группу голоморфных автоморфизмов области $\tau^+(n-1)$ (Напомним, что группа автоморфизмов области называется транзитивной, если существует автоморфизм, переводящий любую точку A из области в любую другую ее точку).

С помощью преобразования (3) и автоморфизма Φ_A шара Ли, переводящий точку A в (0, 0, ..., 0), определим следующее отображение:

$$\psi_B = \Phi \circ \Phi_A \circ \Phi^{-1}, \text{ где } B = \Phi(A).$$

Ясно, что ψ_B является автоморфизмом области $\tau^+(n-1)$, переводящий точку $B \in \tau^+(n-1)$ в точку $(0, 0, ..., i)$.

Таким образом, при помощи отображения (3) выписывается группа автоморфизмов области $\tau^+(n-1)$.

ЛИТЕРАТУРА

1. Хуа Локен. Гармонический анализ функций многих комплексных переменных в классических областях. – М.: ИЛ, 1959. – 163 с.
2. E. Cartan Sur les domaines bornes homogenes de l 'espace de n variables complexes, Abh. Math. Sern. Univ. Hamburg 11(1935),116-162.
3. Зигель К Автоморфные функции нескольких комплексных переменных. М.: ИЛ, 1954. – 168 с.
4. Пятецкий-Шапиро И.И. Геометрия классических областей и теория автоморфных функций. М.: ИЛ., 1961 г. 192 с.
5. Владимиров В.С., Сергеев А.Г. Комплексный анализ в трубе будущего// Современные проблемы математики. Фундаментальные направления. – М.: ВИНИТИ, 1985. – Т. 8. – С. 191–266.
6. Г. Худайберганов, А. М. Кытманов, Б. А. Шаимкулов. Комплексный анализ в матричных областях. Монография. Красноярск: Сибирский федеральный ун-т, 2017. – 297 с.
7. Шабат Б. В. Введение в комплексный анализ. Ч.2. М.:Наука. 3-е изд., 1985 г.

Khudayberganov G. Doctor of Physical and Mathematical Sciences, Professor. National University of Uzbekistan named after Mirzo Ulugbek,

Department of Mathematical Analysis, Faculty of Mathematics. 4 University str, Tashkent , Uzbekistan.,

Abdullayev J. Doctor of Philosophy (PhD) . National University of Uzbekistan named after Mirzo Ulugbek, Department of Mathematical Analysis, Faculty of Mathematics. 4 University str, Tashkent , Uzbekistan.,

About realization of Lie ball

Abstract. *Our task is to find an explicit mapping that establishes a biholomorphic equivalence of the domains* $\Re_{IV}^n$ *and* $\tau^+(n-1)$.

Key words: *classical domains, Lie ball, future tube, Shilov's boundary.*

О преподавании элементов теории вероятностей в академических лицеях, колледжах и школах

А.К.Аманов, Г.М.Рахимов
Ташкентский медицинский лицей
gairat48@gmail.com

Основной задачей колледжей, лицеев и школ является воспитание, социально-педагогическая поддержка становления и развития высоконравственного, ответственного, творческого, инициативного, компетентного, конкурентно-способного гражданина Республики Узбекистан.

Включение в систему образования таких учреждений нового типа, как колледжи, лицеи и школ вызвало необходимость пересмотра отдельных частных методик преподавания дисциплин и, в частности, более тщательной и детальной проработки вопросов методики преподавания математики. В программе по математике для школ, лицеев и колледжей республики Узбекистан установлены определенные часы по изучению раздела элементов теории вероятностей.

Заметим, что одним из важнейших факторов, влияющих на ход и результат учебно-воспитательного процесса, являются методы и формы обучения.

Процесс внедрения элементов теории вероятностей в курс математики является делом достаточно специфическим и трудным. Для усвоения начал теории вероятностей необходим предварительный запас идей, представлений, привычек, коренным образом отличающихся от тех, которые развиваются у обучающихся при традиционном обучении.

Согласно подходам, изложенным в подавляющем большинстве методической литературы, считается, что главным при изучении данного курса должен стать практический опыт учащихся, поэтому обучение желательно начинать с вопросов, в которых требуется найти решение поставленной проблемы на фоне реальной ситуации. Необходимо отметить, что в процессе обучения не следует доказывать все теоремы, так как на это тратиться большое количество времени.

Целью исследования является разработка методики преподавания элементов теории вероятности академических лицеях, колледжах и школах.

Изучение курса обычно начинают с изучения основ комбинаторики, причем параллельно должна изучаться теория вероятностей, так как комбинаторика используется при подсчете вероятностей. Комбинаторика раздел математики, который изучает различные комбинации и перестановки предметов. Начинать обучение комбинаторике целесообразно с реше-

ния простых комбинаторных задач методом перебора. Перед тем как дать ученикам формулу следует поставить какую-либо проблемную задачу.

После изучения основных формул комбинаторики можно давать учащимся задачи на вычисление вероятности, для решения которых необходимо применять комбинаторные формулы. Далее нужно перейти к теории вероятностей.

Преподавание элементов теории вероятностей требует от преподавателя кардинального изменения стиля работы. Оно подразумевает организацию дискуссий на уроке, интенсивную устную работу, а также расширение собственного кругозора в областях других наук, таких, как биология, география, история, литература и многих других, в дополнение к сложившимся методам и подходам к обучению. При этом задачи, стоящие в учебнике рядом, не аналогичны, и решение одной из задач не означает, что будет с легкостью решена следующая. Поэтому главным условием роста профессионализма преподавателя в области преподавания теории вероятностей является изменение технологии преподавательской деятельности при преподавании этого курса.

Несмотря на трудности, которые возникают у преподавателей, учащихся преподавание и изучение элементов "Теории вероятностей"является насущной необходимостью, обусловленной реалиями современного мира.

ЛИТЕРАТУРА

1. Авдеева Н.Н. О статистическом образовании в школе / Н.Н. Авдеева // Математика в школе. 1973. №3. С.4-8.

2. Виленкин, Н.Я. Комбинаторика / Н.Я. Виленкин, А.Н. Виленкин, П.А. Виленкин. М.: МЦНМО, 2006. 400 с.

3. Гнеденко Б.В. Политехнические аспекты преподавания математики в средней школе / Б.В. Гнеденко // На путях обновления школьного курса математики. М., 1978.

4. Мостеллер, Ф. Пятьдесят занимательных вероятностных задач с решениями / Ф. Мостеллер. М.: Наука, 1975. 112 с.

5. Мирзаахмедов М.А., Исмоилов Ш.Н., Аманов А.К. Математика. Алоебра и начала анализа. 2018, ООО "Zamln Nashr (Языки издания: o'zbek, казах, тажик, киргиз, туркман, каракалпак).

Некоторые особенности кредитной системы обучения

У.Ш.Якубова [1], Н.Т.Парпиева [2]

[1] Ташкентский Государственный Экономический Университет

[2] Ташкентский Государственный Педагогический Университет

u.yakubova@tsue.uz, nparpieva@mail.ru

Аннотация: В статье рассмотрены организационная структура учебного процесса в условиях кредитной системы, что из себя представляет система ECTS, отмечены свойства ECTS, параметры ECTS, основные документы ECTS, рамочные условия по применению кредитной технологии в вузах Узбекистана.

Ключевые слова: кредитная система обучения, кредитная технология, высшее образование, квалификация, стратегия процесса, Европейская система зачетного перевода

Социально-экономическое развитие нашей страны обуславливает коренное совершенствование высшего образования. Повышается значимость подготовки кадров, создаются условия для переподготовки специалистов высшего образования на уровне международных стандартов. Исходя из естественных требований общественной жизни и экономики, одной из основных задач модернизации системы высшего образования является внедрение в обучение современных форм и технологий на основе изучения международного опыта. [1, с.534]

Рассмотрим организационную структуру учебного процесса в условиях кредитной системы обучения. Основополагающим документом, который определяет стратегию этого процесса, является Конвенция "О признании квалификаций, относящихся к высшему образованию в Европейском регионе обнародованную ЮНЕСКО в апреле 1997 года в Лиссабоне.

Как известно, все 29 стран, подписавшие Болонскую декларацию в 1999 году, обязаны перейти на двухуровневую систему высшего образования. Важнейшей составляющей общеевропейской системы образования являются кредиты. Внедрение системы кредитов по типу ECTS – европейской системы зачетного перевода – рассматривается в качестве средства поддержки крупномасштабной студенческой мобильности.

Европейская Система Перевода и накопления Кредитов, ECTS, была разработана Комиссией Европейского Сообщества, чтобы обеспечить общие процедуры гарантии академического признания учебы за границей. ECTS

обеспечивает способ измерения и сравнения уровня достижений в образовании, а также передачу их от одного учреждения к другому. ECTS система основана на принципе взаимного доверия между учреждениями высшего образования. ECTS-кредит – это числовое значение, которое предназначено для описания общих затрат времени студентов на изучение материала учебного курса. Они отражают суммарную работу студента над дисциплиной на лекциях и семинарах, практических и лабораторных занятиях, самостоятельную работу – в библиотеке и дома, которые направлены на овладение предметом. Перевод (1989) способствует мобильности и гибкости; определяет относительную ценность системы. Накопление(1995) обеспечивает непрерывность и связь между ступенями обучения; определяет абсолютную ценность системы; можно пользоваться в процессе обучения в течение всей жизни.

Необходимо отметить три свойства ECTS: система акцентирована на студенте и позволяет описать всю совокупность траекторий обучения, определяя каждую составляющую в кредитах; сопоставимость и совместимость программ подготовки и обучения; привлекательность системы для студентов других континентов.

Двумя тесно взаимосвязанными параметрами ECTS являются: необходимый объем учебной нагрузки для достижения целей конкретной программы обучения; результаты обучения акцентируются на знаниях и компетенциях. Оба данных параметра тесно взаимосвязаны.

Основные документы ECTS – это само предложение об обучении, т.е. информационный каталог всех предметов и информационные брошюры; личные анкеты кандидатов; договор об обучении; рейтинговый лист. Информационный каталог содержит четыре основные рубрики:

Информация об институте;

Информация о программах;

Список UE (единиц обучения);

Информация для встречи студентов.

ECTS единица учёта основывается на доверии:

Признается везде;

Позволяет обмен;

ECTS единица измерения:

Является конвертируемой;

Снабжает общими сведениями;

Накапливается.

Рамочные условия по применению кредитной технологии в вузах Узбекистана:

1 семестр = 30 кредитов, 1 год = 60кредитов.

Степень Бакалавра требует 4-х лет обучения:

4х60=240 кредитов.

Степень Магистра требует 2-х лет обучения:

2х60=120 кредитов.

1 кредит = 30 академических часов;

1 семестр = 30 час * 30 кредитов = 900 часов;

1 год = 900*2 = 1800 часов.

Практика = 5 кредитов.

Дипломная работа = 10 кредитов.

Студент получает кредиты только в результате успешного завершения учебной дисциплины, т.е. кредит не присуждается просто за посещение занятий – студент должен выполнить все требования и успешно пройти все формы оценивания. Количество кредитов не зависит от оценки - "стоимость"курса в кредитах предопределена заранее.

Несомненно, самая известная форма интеграции высшего образования – это мобильность студентов, т.е. увеличение числа студентов, обучающихся за границей. За последние 50 лет темпы увеличения потоков студентов превысили темпы распространения самого высшего образования. По данным ЮНЕСКО, уровень международной мобильности студентов вырос за последние 35 лет более, чем на 300

Целью кредитной технологии обучения являются обеспечение академической мобильности субъектов образовательного процесса, а также гармонизация отечественной системы образования с международным образовательным пространством.

А задачами кредитной технологии обучения можно назвать унификацию объема знаний, максимальную индивидуализацию обучения и усиление роли самостоятельной работы.

В процесс подготовки входят проведение семинаров для студентов и преподавателей, составление и утверждение Рабочих планов перехода академии на кредитную систему, разработка Временных положений об организации учебного процесса для внедрения кредитной технологии, составление нового академического календаря учебного процесса, пересмотр структуры факультетов, а также составление информационных каталогов (в отдельных случаях регламента обучения).

Для реализации проекта необходимо проведение семинаров для студентов и преподавателей, работа с консультантами и сотрудниками деканата, разработка памяток и буклетов для студентов и преподавателей, составление и издание методических разработок, каталога курсов, и других материалов по распространению правил кредитной системы. Сам процесс внедрения проекта состоит из составления примерных индивидуальных учебных планов студентов с помощью консультантов, проведения лекторами презентации курсов и записи студентами на выбранные курсы, а также составление индивидуальных расписаний занятий студентов.

Результатами внедрения можно назвать следующие: создание нормативно – правовой базы для перехода; проведение функциональных и структурных изменений; обеспечение нелинейной организации учебного процесса; повышение ответственности студентов и создание конкурентной среды среди преподавателей за качество обучения; повышение основных показателей учебной работы, такие как посещаемость, успеваемость; обеспечение тесной связи студента и преподавателя, основанной на взаимном доверии и взаимном уважении; изменение мотивов и правил работы для преподавателей.

Кредит – количественный показатель, определяющий, какую часть объема четырех годовой нагрузки студент выполняет в данном курсе (дисциплине);

Оценка – качественный показатель, определяющий, насколько успешно студент освоил программу курса, выполнив определенный в кредитах объем нагрузки.

ШКАЛА ОЦЕНКИ ЗНАНИЙ СТУДЕНТОВ КЕФ:

- Вариация баллов 0-20
- Отрицательный результат 0-10
- Положительный результат 10-20
- Неудовлетворительно 0-5

(повторное изучение дисциплины)

- Неудовлетворительно 5-10

(возможна дополнительная пересдача)

- Удовлетворительно 10-12
- Хорошо 12-16
- Отлично 16-20

И наконец, о самостоятельной работе. Самостоятельная работа обучающегося (СРО) – работа по определенному перечню тем, отведенных на самостоятельное изучение, обеспеченных учебно-методической литературой

и рекомендациями, контролируемые в виде тестов, контрольных работ, коллоквиумов, рефератов, сочинений и отчетов.

При кредитной системе обучения самостоятельная работа обучающегося подразделяется на две составляющие:

- Самостоятельная работа обучающегося, выполняемая под руководством преподавателя по графику;
- Работа (задание), выполняемая полностью самостоятельно обучающимся.

Средний балл успеваемости (Grade Point Average - GPA) – это средневзвешенная оценка уровня учебных достижений обучающегося за один учебный год по выбранной программе (отношение суммы произведений кредитов на цифровой эквивалент баллов оценки итогового контроля по дисциплинам к общему количеству кредитов за текущий период обучения).

ЛИТЕРАТУРА

1. Ходжаниязов С. У. Практическая реализация опыта европейской кредитной технологии обучения в условиях Узбекистана // Молодой ученый. – 2012. – №5. – С. 534-536.

Yakubova U.Sh., ParpievaN.T.

Some features of credit system

Abstract. *The article examines the organizational structure of the training process in the conditions of the credit system, what the ECTS system is, the properties of ECTS, ECTS parameters, the main documents of ECTS, the framework conditions for the use of credit technology in Uzbekistan's universities.*

Key words: *credit education system, credit technology, high education, qualification, strategy of process, European credit transfer system.*

The scheme of business processes LMS NUUz

B.B.Elov[1], M.H.Primova[2]
[1]Tashkent State University of Uzbek Language and Literature
[2] National University of Uzbekistan
ebb@mail.ru

Abstract. This paper describes the LMS process, which are designed specifically for the National University of Uzbekistan to support the traditional educational process. We used BPMN methodology for describing business processes of LMS NUUz.

Keywords.component; LMS; curriculum; educational process; educational programs.

I. INTRODUCTION

The National University of Uzbekistan used Learning Management System (LMS) - learning support system that can be used by all participants in the educational process.

The system is developed, taking into account the specifics of the educational process in our university.

The LMS NUUz each student has the opportunity to receive personalized information about an individual curriculum, academic achievement (electronic record book) and individual schedule [1]. All materials published on the websites of the teacher studied or passed courses available to students at any given time.

The teacher can use the LMS to accommodate NUUz materials for classes, management worksheet, exchange messages with the students to use the test to verify the knowledge acquired intermediate and final control, organize, through the LMS NUUz delivery of projects and homework.

The LMS NUUz implemented services [1]:

- the possibility of self planning their educational trajectory using module "Elective courses". Using these module, each student has the opportunity to apply not only to the discipline of the variable part of the curriculum their educational program, but also the discipline of other programs;

- to improve the quality of teaching at the University, as well as to obtain feedback from the students developed a special module, "assess their courses."At the end of each module, each student creates an individual profile of the teachers with the evaluation criteria. All the results of the survey are anonymous.

LMS system (Learning Management System) - a system of on-line support of the educational process NUUz. It contains disciplines working curricula attached to them by students. In each discipline a teacher has the ability to:

- to place training materials;
- supervise knowledge through testing;
- Maintain gradebook and report on the learning activities of students;
- Collect and store the independent work of students (essays, term papers and dissertations, projects, etc.);
- Meet and consult with students.
- The system recorded all students NUUz.
- The system is connected with the base from which it imported all approved by the working curriculum subjects with designated students to study it.
- If discipline is a discipline of choice or electives, students are attached to it manually by an administrator LMS system.

Implementation of the requirements for modern educational institutions from the management of the university, its customers (in this case by the client understands students - individuals involved in the educational process of high school, and their agents - both physical and legal persons) and the educational environment. It is achieved by controlling the educational process in accordance with certain concepts and principles [2].

LMS is based on three concepts of the educational process: modularity, individuality, encapsulation.

The principles laid down in the LMS, implemented in the technology subsystems and processes.

The composition process is typical and may be different in the implementation of the volume or the representation of one or another direction [3].

Here we restrict ourselves to the study of processes of the LMS NUUz.

II. THE SCHEME OF BUSINESS PROCESSES LMS

As realisations of various processes are interconnected, all processes are grouped in some independent tasks:

1. Planning
- The task "Formation of educational programs
- The task "Administrative curriculum
- The task "Schedule formation
- The task "Financial planning".

2. Accounting
- The task "Accounting of clients

- The task "Accounting of resources of educational process
- The task "Accounting of an education-methodical material
- The task "Accounting of training courses".

3. Reporting

- The task "Accounting of results of educational process
- The task "Formation of diplomas and the academic certificates
- The task "DocumentЎs flow".

In this paper, we consider only the selected task. Within the framework of the planning of the educational process is carried out:

- Formation of the educational programs of the university;
- Select by students of training courses;
- Formation of the university administrative training plan for the implementation of educational programs in accordance with the material and technical base and Teaching Staff;
- Creation of a "template"schedules to distinguish the use of scarce resources;
- Formation of a week schedule, taking into account the current state of the material and technical resources (MTR) and Teaching Staff for each student individually.

Within the framework of the account of clients, material and technical resources, of teaching staff, training and methodological support, the results of the educational process is carried out:

- Accounting logistic support training process;
- Accounting of teaching staff;
- Accounting for customers of the university;;
- Accounting modules of the curriculum;
- Individual accounting of the educational process results.

Within the framework of planning, accounting and analysis of financial flows and the performance is carried out:

- Planning of revenue from tuition fees;
- Fixing and analysis payment training;
- Fixing the financial obligations of the university;
- Decision on the execution of the financial obligations of the university;
- Fixing performance of the financial obligations of the university;
- Formation of the financial budget of the university;
- Monitoring performance of expenditure and revenue side of the budget;

Within the framework of reporting institution of higher education outside certification organizations, both public and foreign made:

- Formation of reports for submission to the state controlling authorities;

- Formation of reports for submission to the international supervisory bodies (AACSB, ECTS, etc.);
- Formation of academic documents for customers.

Within the framework of formation of an information portal to provide information to the participants of the educational process is carried out:

- Publication of information on existing educational programs of the university and the current curriculum;
- Publication of information on students' ratings, their scientific and practical and project activities within the walls of the university;
- Publication of the results of the educational process;
- Publication of general university normative, reference and other information.

A. *Processes of a task "Formation of educational programs"*

Function (task) is necessary for the formation of the educational university programs, taking into account resource constraints and in order to solve educational problems. On the basis of the educational program, formed for the duration of study, carried out the operational plan for each period of the educational program. Below we describes on detail the tasks and subtasks specified in the figures 1-9.

B. *Processes of a task "Administrative curriculum"*

Administrative curriculum is used to negotiate the implementation of several training programs at a total resource base (teachers, students, material and technical resources). Each training program has its own unique content and forms of training. Administrative curriculum (hereinafter AC) is prepared for each training program for the period defined by the schedule of the educational process (SEP) (a document that fixed the dates of beginning and end of the administrative period of planning, training mode - training, weekends, holidays, training time).

Below we describes on detail the tasks and subtasks specified in the figures 10-12.

CONCLUSION

In this paper we construct following process diagrams of LMS NUUz:

- Function "Formation of educational programs"
- Function "Administrative curriculum".

Similarly, we can construct other schemes of LMS NUUz processes.

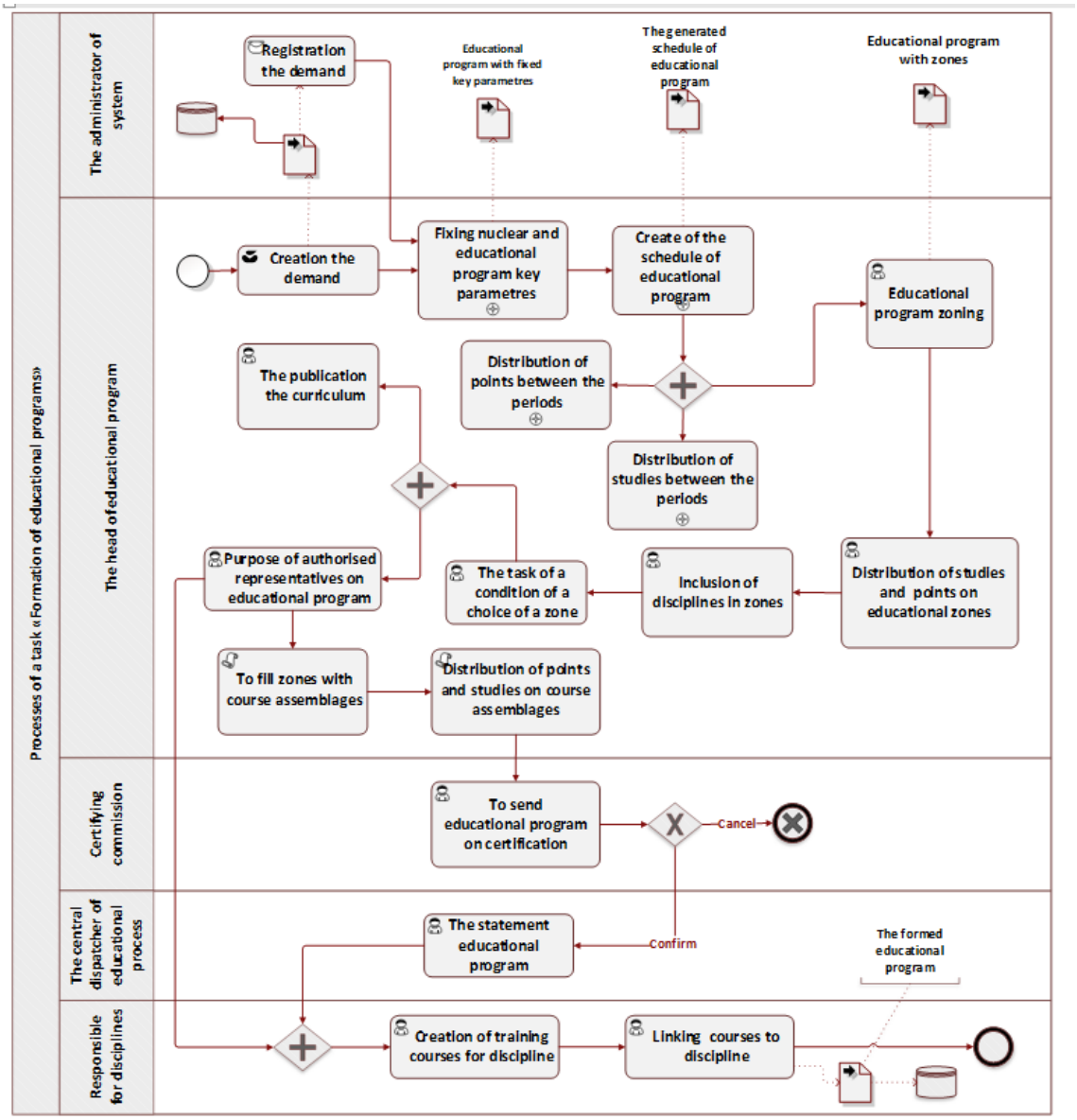


Figure 1. Processes of a task «Formation of educational programs»

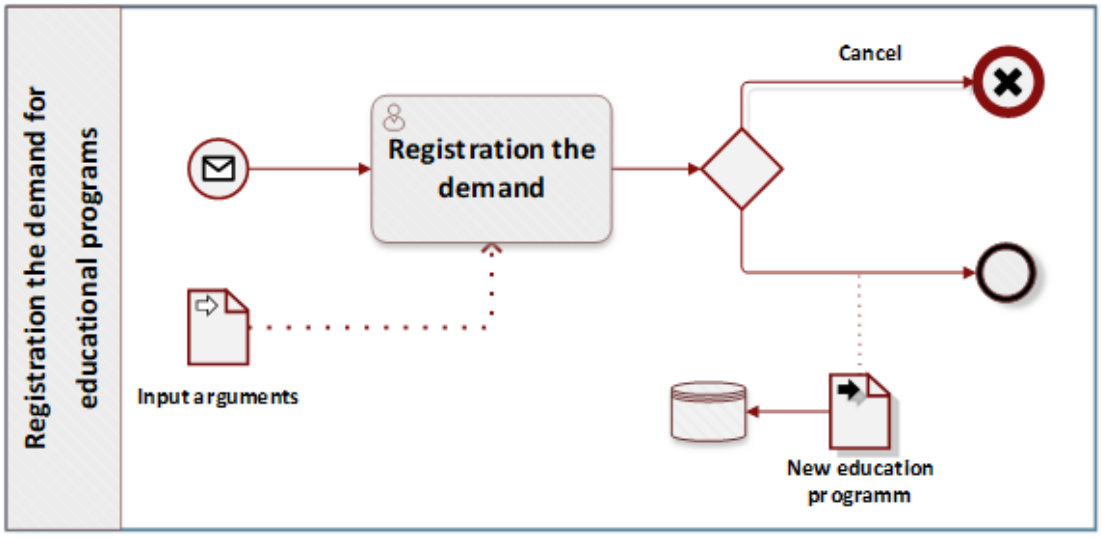


Figure 2. Formation of educational programs

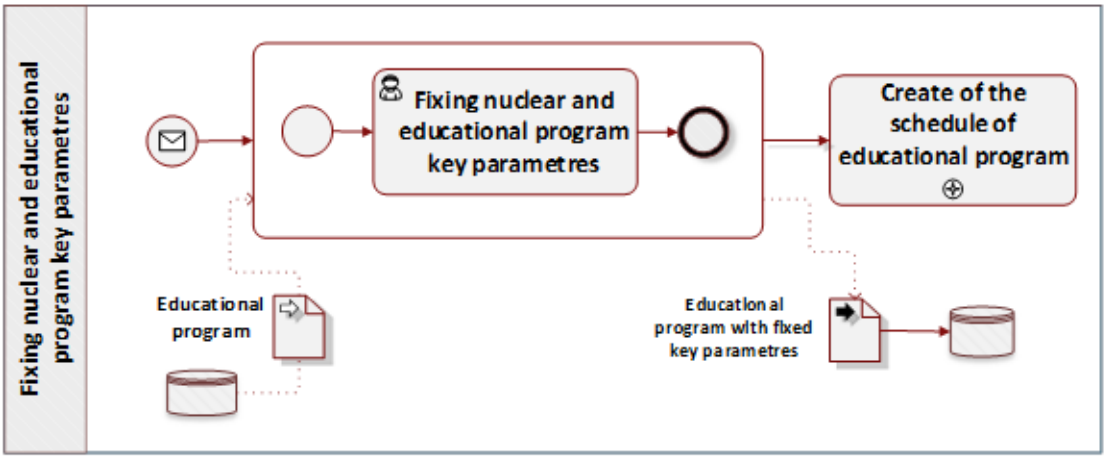


Figure 3. Fixing nuclear and educational program key parametres

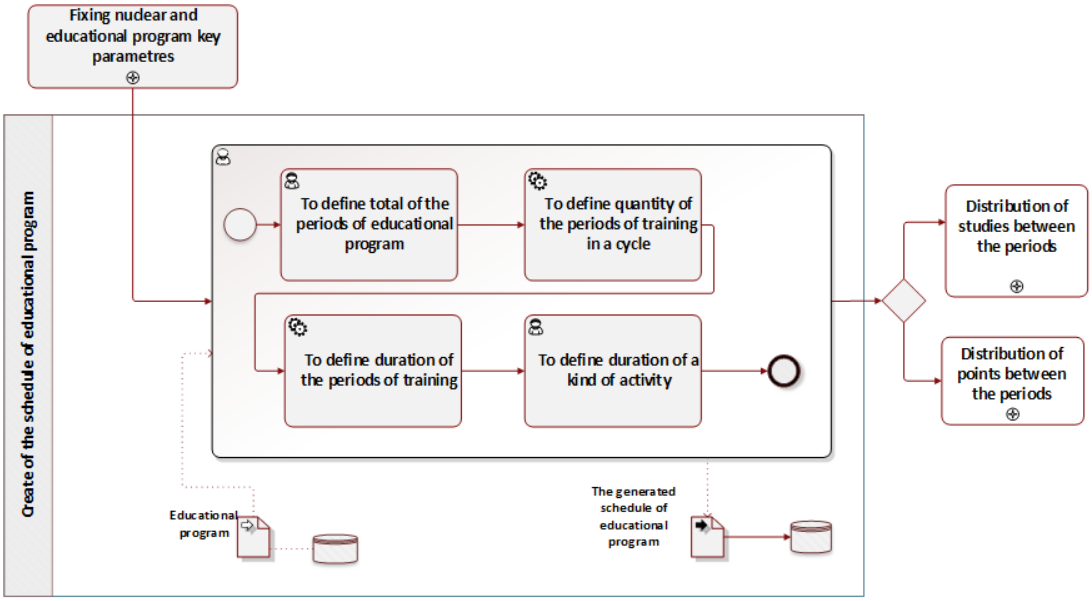


Figure 4. Create of the schedule of educational program

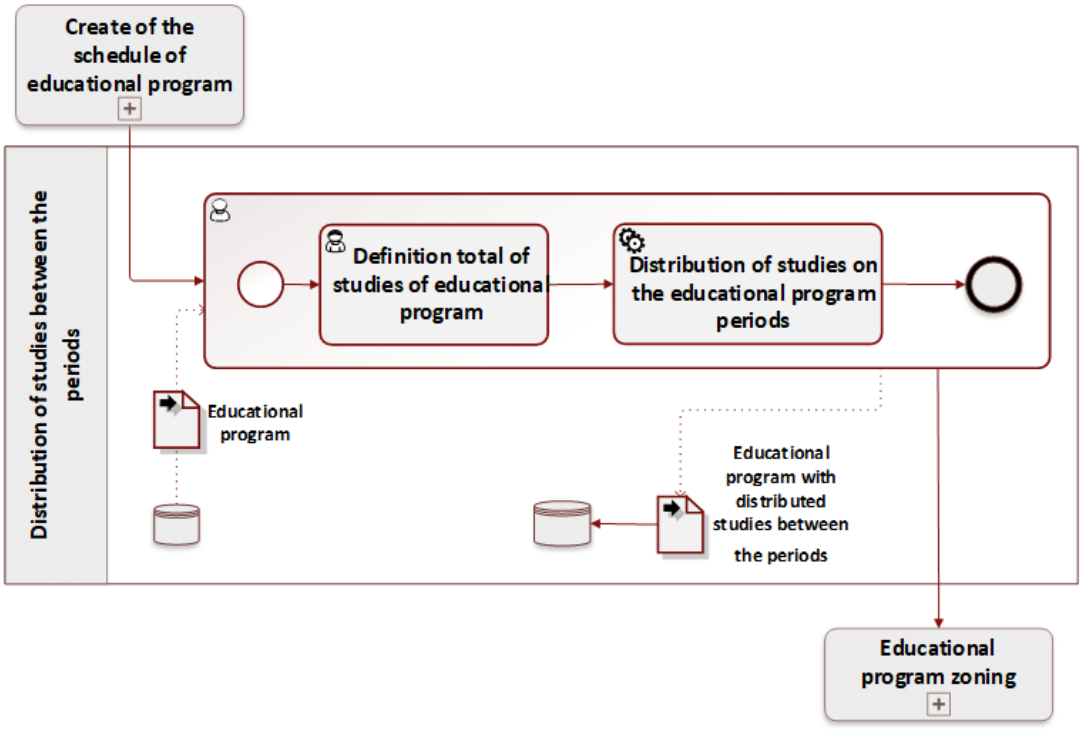


Figure 5. Distribution of studies between the periods

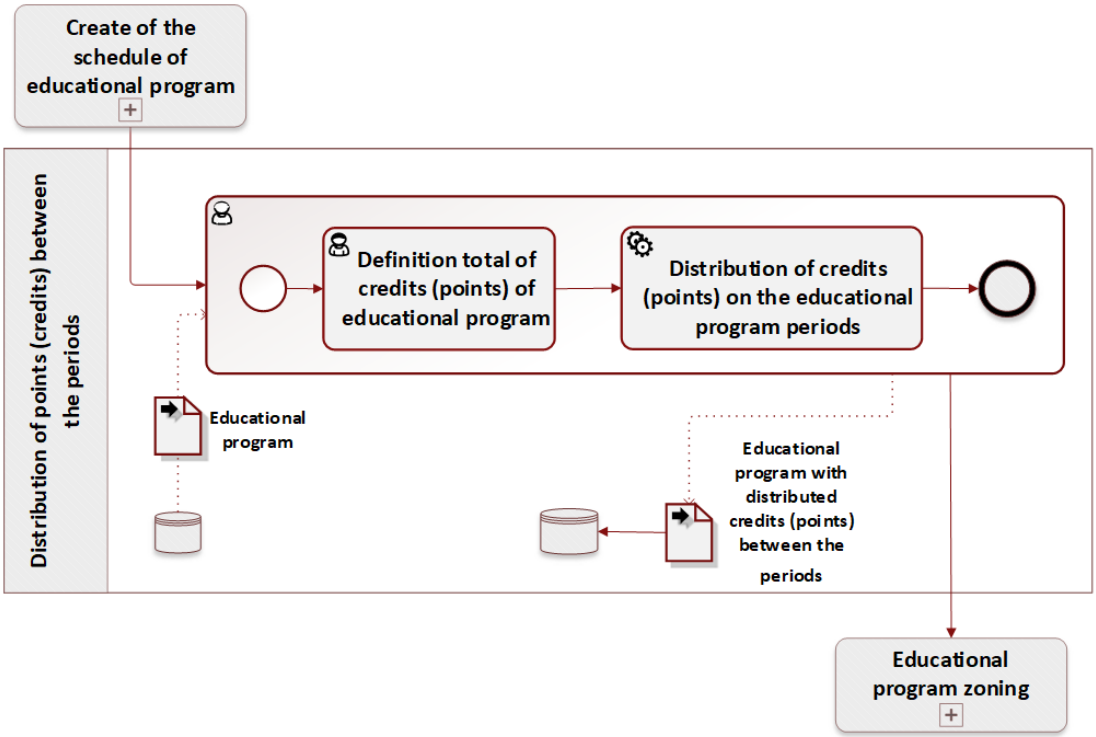


Figure 6. Distribution of points (credits) between the periods

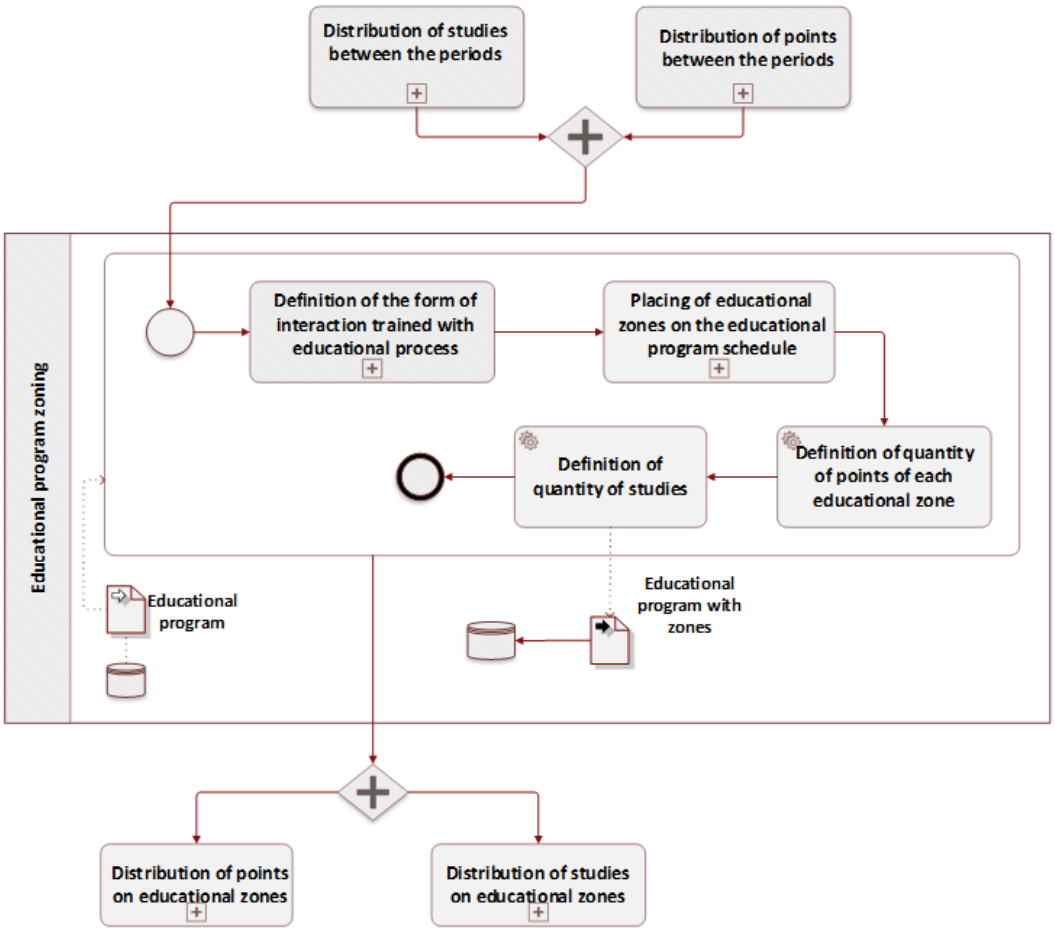


Figure 7. Educational program zoning

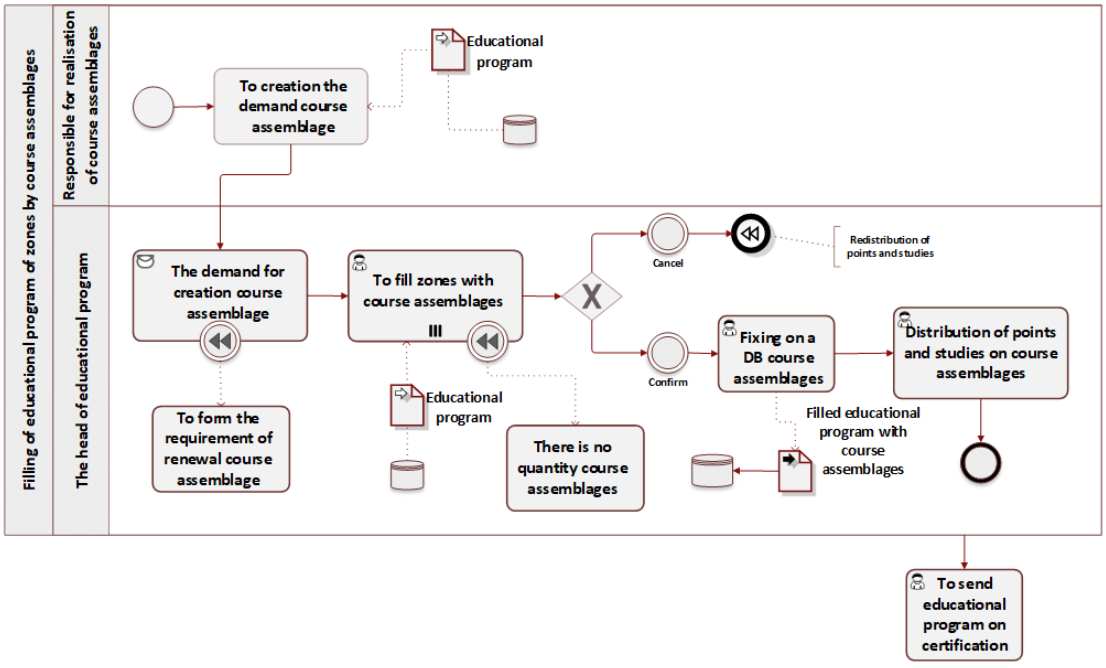


Figure 8. Filling of educational program of zones by course assemblages

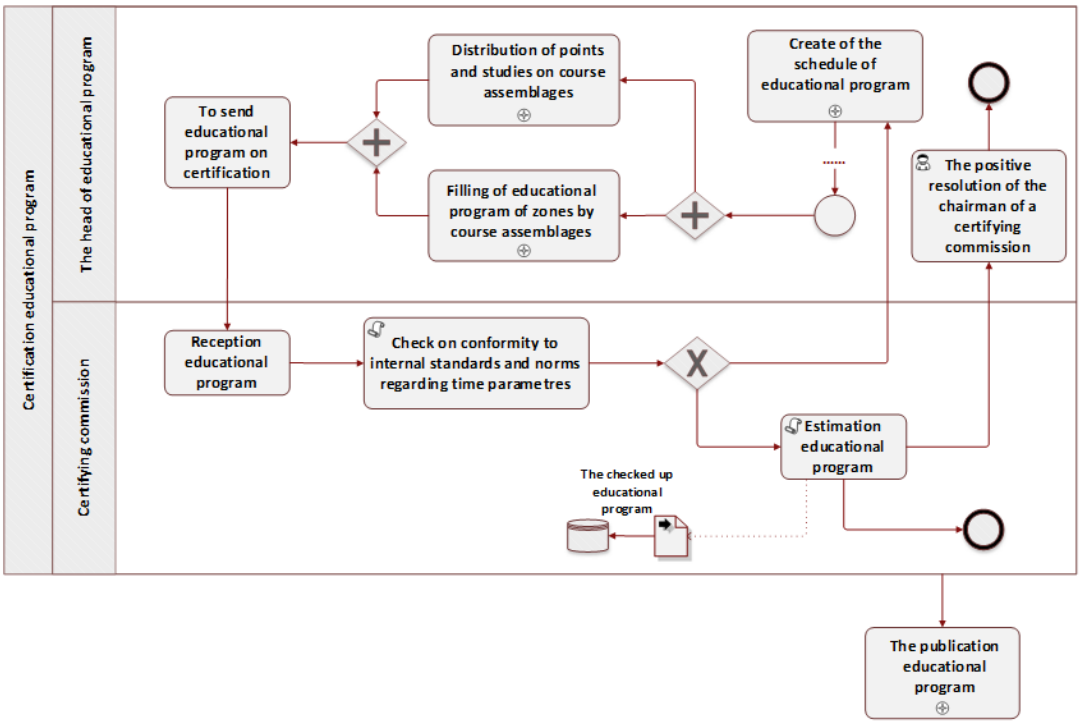


Figure 9. Certification educational program

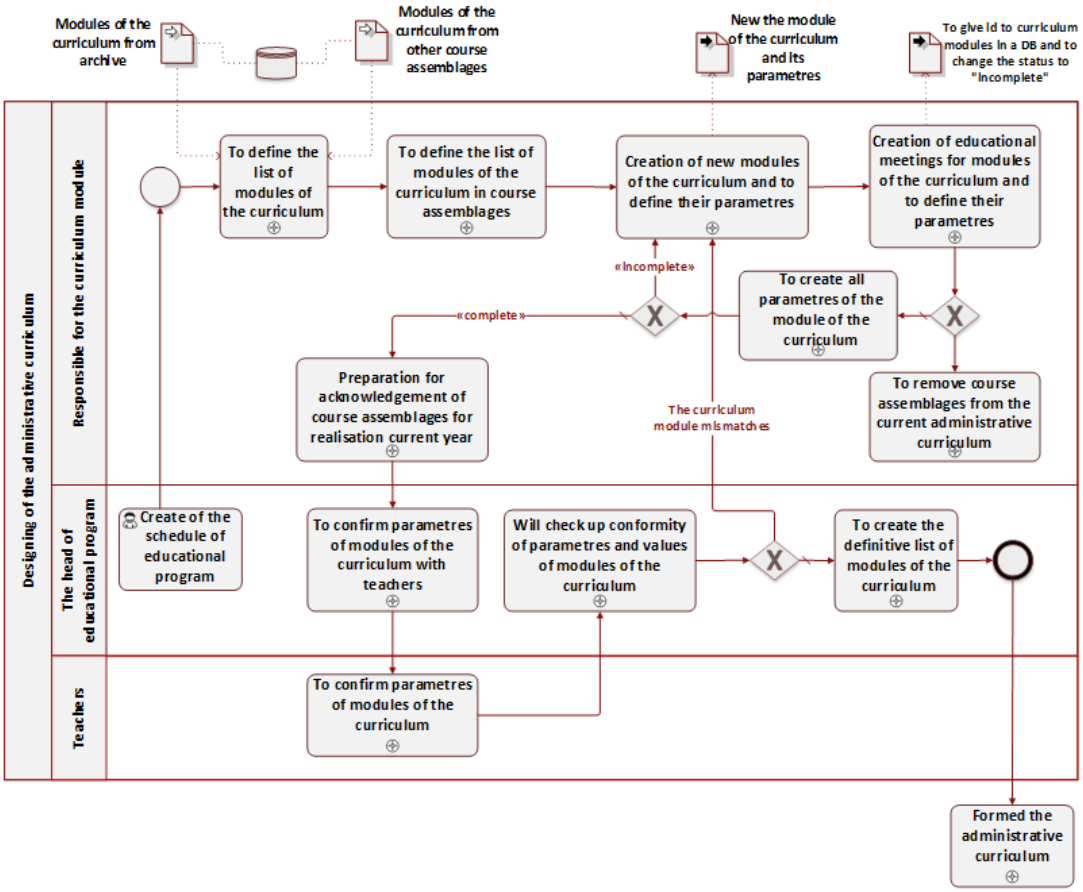


Figure 10. Designing of the administrative curriculum

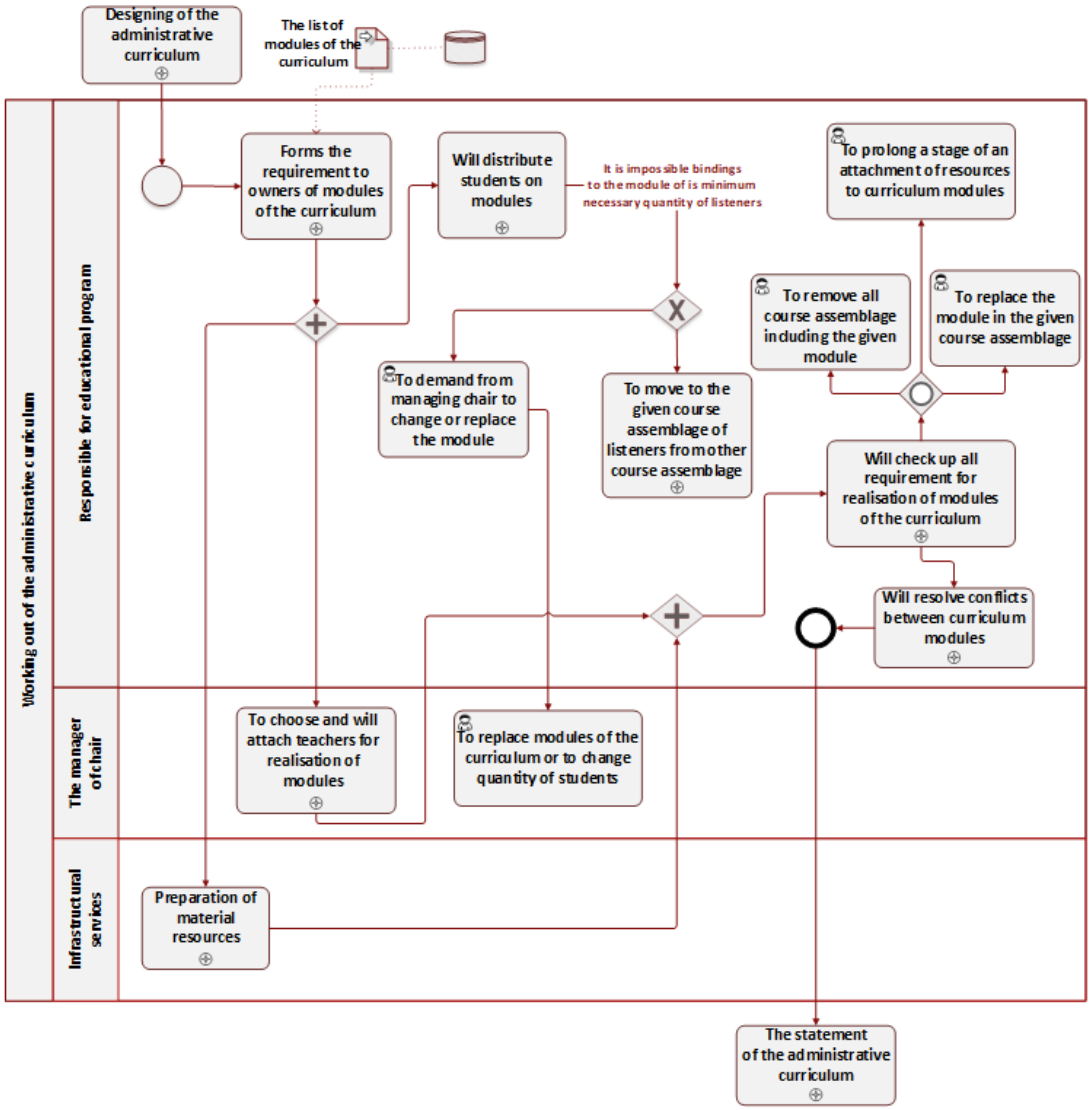


Figure 11. Working out of the administrative curriculum

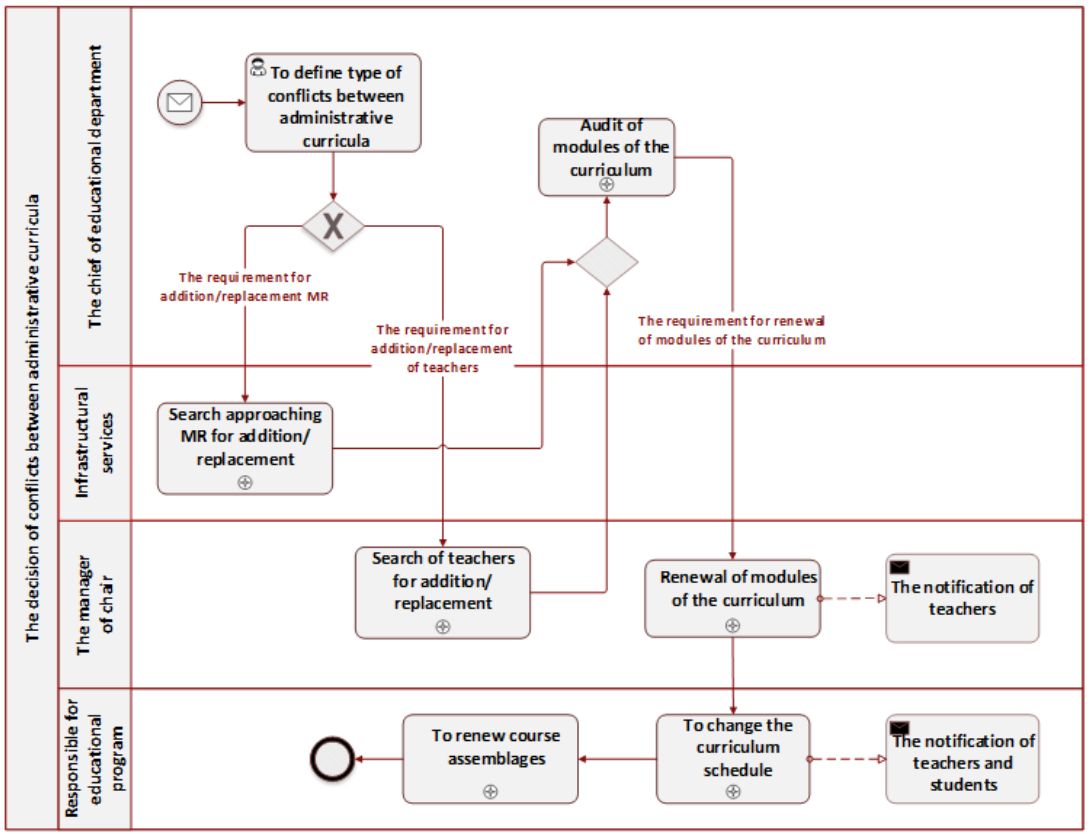


Figure 12. The decision of conflicts between administrative curricula

Reference

1. Elov B. Portal national university of Uzbekistan. International Journal of Applied Engineering Research and Development(IJAERD) ISSN(P): 2250-1584; ISSN(E): 2278-9383 Vol. 6, Issue 2, Apr 2016, 33-38 TJPRC Pvt. Ltd.

2. Nor Azura Adzharuddin, Lee Hwei Ling, Member, IACSIT. Learning Management System (LMS) among University Students: Does It Work? International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 3, No. 3, June 2013.

3. Nouran M. Radwan, M. Badr Senousy. Approaches for Managing Uncertainty in Learning Management Systems. Egyptian Computer Science Journal Vol. 40 No.2 May 2016 ISSN-1110-2586.

Некоторые неравенство для моментов ветвящихся случайных процессов с иммиграцией

Я.М.Хусанбаев, Х.Э.Кудратов

Пусть $\{\xi_{k,i}, k, i \geq 1\}$ и $\{\varepsilon_k, k \geq 1\}$-две независимые совокупности независимых, неотрицательных, принимающих целые значения и одинаково распределенных случайных величин. Рассмотрим ветвящийся процесс с иммиграцией (см. [1]) $X_k, k \geq 0$ определенную следующими рекуррентными соотношениями:

$$X_0 = \eta,\ X_n = \sum_{j=1}^{X_{n-1}} \xi_{n,j} + \varepsilon_n$$

где $\eta-$ некоторая неотрицательная, принимающая целые значения случайная величина, независящая от семейство случайных величин $\{\xi_{k,i}, \varepsilon_k, k, i \geq 1\}$.

Обозначим через $\mathcal{F}_k - \sigma-$ алгебру, порожденную случайными величинами $X_0, X_1, ..., X_k$. В дальнейшим знак $E(\xi|\mathcal{F}_n)$ будет обозначать условное математическое ожидание случайной величины ξ относительно алгебры $\mathcal{F}_n$.

Введем следующие обозначения:

$$m = E\xi_{1,1},\ \gamma_p = E\xi_{1,1}^p,\ \theta_p = E \mid \xi_{1,1} - m \mid^p$$

$$\lambda = E\varepsilon_1,\ \tau_p = E\varepsilon_1^p,\ \delta_p = E \mid \varepsilon_1 - \lambda \mid^p$$

$$\nu = E\eta,\ \nu_p = E\eta^p$$

Справедливы следующие теоремы.

Теорема 1. Пусть $m < \infty, \nu < \infty$. Тогда для любого $1 < p \leq 2$ имеет место неравенство

1. если $m \neq 1$, то

$$E|X_n - E(X_n|\mathcal{F}_{n-1})|^p \leq 2^{p-1}[2\theta_p(m^{n-1}\nu + \frac{m^{n-1}-1}{m-1}\lambda) + \delta_p]$$

2. если $m = 1$, то

$$E|X_n - E(X_n|\mathcal{F}_{n-1})|^p \leq 2^{p-1}[2\theta_p(\nu + n\lambda) + \delta_p]$$

Теорема 2. Пусть $\gamma_{\frac{p}{2}} < \infty, \theta_p < \infty, \tau_{\frac{p}{2}} < \infty, \delta_p < \infty, \nu_p < \infty$ при некотором фиксированном $p > 2$. Тогда справедливы следующие неравенства:

1. если $2^{\frac{p}{2}-1}\gamma_{\frac{p}{2}} \neq 1$, то

$$E|X_n - E(X_n|\mathcal{F}_{n-1})|^p \leq 2^{p-1}(C(p)\theta_p((2^{\frac{p}{2}-1}\gamma_{\frac{p}{2}})^{n-1}\nu_p +$$

$$+\frac{2^{\frac{p}{2}-1}\tau_{\frac{p}{2}}}{2^{\frac{p}{2}-1}\gamma_{\frac{p}{2}}-1}[(2^{\frac{p}{2}-1}\gamma_{\frac{p}{2}})^{n-1}-1])+\delta_p)$$

2. если $2^{\frac{p}{2}-1}\gamma_{\frac{p}{2}}=1$, то

$$E|X_n-E(X_n|\mathcal{F}_{n-1})|^p\leq 2^{p-1}(C(p)\theta_p((\nu_p+2^{\frac{p}{2}-1}\tau_{\frac{p}{2}}n+\delta_p)$$

где $C(p)$- постоянное число, зависящее только от p.

ЛИТЕРАТУРА

1. Athreya K.B., Ney P.E. Branching processes, Springer - Verlag, Berlin. 1972.287p.

2. Петров В.В. Предельные теоремы для сумм независимых случайных величин. Москва, 1987.

Аппроксимации функционалов некоторых статистик - мер неравенства

Ш.А.Исматуллаев, Ш.Ш.Шарахметов
Специализированная школа им.Аль-Хорезми, Ташкентский Государственный Экономический Университет

Введённая К. Джини [4] средняя разность Δ =E|X-Y|, характеризующая разброс значений вариант X и У друг относительно друга, представляет значительный практический и теоретический интерес. Наряду с Δ, часто рассматривается, так называемый, коэффициент Джини G – отношение средней разности Δ к общему математическому ожиданию вариант m=EX=EY, т.е. G= $\frac{\Delta}{2\mathrm{m}}$, который не зависит от масштаба значений варианты (см. подробно в [2]). Этот нормированный вариант средней разности широко используется в экономико-статистических исследованиях как мера неравенства. Знак модуля, участвующий в формулах определяющих среднюю разность Джини, вносит известные трудности при ее использовании в выборочной теории. Поэтому, в частности, задача нахождения точных границ для коэффициента Джини и других мер неравенства представляет значительный интерес в экономических и статистических исследованиях и смежных вопросах (см.[1], [2]). Функционал inf E |X-Y|, названный К.Джини кпростым индексом различиянь, используется также в метрической теории, теории вероятностей и функциональном анализе (см. [1]).

В настоящем сообщении находятся точные границы аппроксимации для функционала γ_n (см. ниже). Сформулируем основные полученные результаты. Часть из них была анонсирована в [5].

Теорема 1. Имеет место соотношение:

$$\underset{F}{Sup}\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|\ d\ F(x)\ d\ F(y)}{\sqrt{\int_{-\infty}^{\infty}x^2\ d\ F\,(x)-\left(\int_{-\infty}^{\infty}x\ d\ F(x)\right)^2}}=\frac{2}{3}\sqrt{3}\ \ ;$$

где Sup берется по всем функциям распределения F с $\int_{-\infty}^{\infty}\mathrm{x}^2\ \mathrm{dF(x)}<\infty$. При этом супремум достигается на [0,1]-равномерной функции распределения.

Теорема 2. Для функционала γ_n=$Sup\ \frac{\mathrm{E}|X-Y|}{\sqrt{DX}}$ имеет место соотношение

$$Sup\ \frac{\mathrm{E}|X-Y|}{\sqrt{DX}}\ =\ \frac{2}{3}\ \sqrt{3\ (1-\frac{1}{n^2})}\ ;$$

где Sup берется по всем независимым одинаково распределенным дискретным с.в. X и Y, принимающим n значений, и таким что DX=$DY<\infty$.

Следствие 1. Для функционала γ_n справедливы соотношения:
1). $\gamma_2 = 1$, $\gamma_3 = \frac{4}{9}\sqrt{6}$, $\gamma_4 = \frac{\sqrt{5}}{2}$, ...
2). $1 \le \gamma_n < \frac{2}{3}\sqrt{3}$,
3). $\gamma_n < \gamma_{n+1}$, $\gamma_n \uparrow \frac{2}{3}\sqrt{3}$.

В доказательстве основных результатов главное место занимает следующее утверждение.

Лемма 1. Пусть X и Y- независимые одинаково распределенные с.в. с распределениями

$P\ (X=a_k) = P\ (Y=a_k) = p_k,\ k=1,2,\ \dots,\ n.$

Тогда

$Sup\ \frac{\mathrm{E}|X-Y|}{\sqrt{D}} = \frac{2}{3}\sqrt{3\ (1 - \frac{1}{n^2})}$;

где Sup берется по всем с.в. X и Y, принимающим n значений и таким, что

$EX = EY = m,\ Ex^2 = Ey^2 = d,\ D = d - m^2 > 0.$

При этом, рассматриваемый супремум достигается на распределении

$P\ (\ X = \frac{\sqrt{3D}}{\sqrt{n^2-1}}\ (\ 2k - n - 1) + m) = \frac{1}{n}\ ;\ k=1,2,\ \dots\ ,n.$

Доказательство.

Обозначим $\Delta\ (a,p) = \mathrm{E}|X-Y|$. Не нарушая общности будем считать, что $a_1 < a_2 < \dots < a_n$. Ясно, что тогда

$\Delta\ (a,p) = \sum_{1 \le i,\ j \le n} |\ a_j - a_i|\ p_i p_j = 2\sum_{1 \le i < j \le n} (\ a_j - a_i) p_i p_j.$

Отсюда, учитывая что $a_j - a_i = \sum_{k=i}^{j-1} (a_{k+1} - a_k)$, получим следующее представление

$\Delta(a,p) = \sum_{k=1}^{n-1} (a_{k+1} - a_k)\ \alpha_k\ ,\ \text{где}\ \ \alpha_k = 2\sum_{i=1}^{k} p_i \sum_{j=k+1}^{n} p_j\ .\qquad \langle 1 \rangle$

Дальнейшее доказательство проведем в два этапа: на первом этапе для произвольных $p_1,\ \dots\ ,\ p_n$ найдем

$$\Delta(p) = \max_{a} \sum_{k=1}^{n-1} (a_{k+1} - a_k)\alpha_k,$$

где max вычисляется по всем $a_1,\ \dots\ ,\ a_n$, удовлетворяющим условиям

$\sum_{i=1}^{n} a_i p_i = m,\quad \sum_{i=1}^{n} {a_i}^2\ p_i = d;$

а на втором этапе найдем $\Delta_n = \max_{p} \Delta\ (p)$, по всем $p_1 > 0,\ \dots\ ,\ p_n > 0$, таким, что $p_1 + \dots\ + p_n = 1$.

На обоих этапах, соответствующие задачи на экстремум решаются методом множителей Лагранжа.

Обозначим $\mathcal{L}\ (a,p)$ функцию Лагранжа

$\mathcal{L}\ (a,p) = \sum_{k=1}^{n-1} (a_{k+1} - a_k)\ \alpha_k - \lambda_1\ (\sum_{i=1}^{n} a_i p_i - m) - \lambda_2\ (\sum_{i=1}^{n} {a_i}^2 p_i - d),$

и найдем стационарные точки, как корни следующей системы уравнений

$$\frac{\partial\ \mathcal{L}\ (\mathrm{a},\mathrm{p})}{\partial a_1} = -\alpha_1 - \lambda_1\ p_1 - \ 2\lambda_2\ a_1 p_1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial a_1} = -\alpha_1 - \lambda_1 p_1 - 2\lambda_2 \alpha_1 p_1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial a_k} = \alpha_{k-1} - \alpha_k - \lambda_1 p_1 - 2\lambda_2 \alpha_k p_k = 0, ...n-1, \qquad (2)$$

$$\frac{\partial \mathcal{L}}{\partial a_n} = \alpha_{n-1} - \lambda_1 p_1 - 2\lambda_2 \alpha_n p_n = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = m - \sum_{i=1}^{n} a_i p_i = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = d - \sum_{i=1}^{n} a_i^2 p_i = 0.$$

Отсюда находим

$$0 = \sum_{k=1}^{n} \frac{\partial\ \mathcal{L}\ (a,p)}{\partial\ a_k} = -\alpha_1 + \sum_{k=2}^{n-1} (\alpha_{k-1} - \ \alpha_k) + \ \alpha_{k-1} - \lambda_1 \sum_{k=1}^{n} p_k -$$

$- \ 2\ \lambda_2\ \sum_{k=1}^{n} a_k p_k = 0\text{- }\lambda_1\text{- }2\ \lambda_2 m,$
$0 = \sum_{k=1}^{n} a_k \frac{\partial\ \mathcal{L}\ (a,p)}{\partial\ a_k} = \sum_{k=1}^{n-1} (a_{k+1} - a_k)\alpha_k \text{ -}\lambda_1 \sum_{k=1}^{n} a_k p_k\text{-}$
$-\lambda_2 \sum_{k=1}^{n} {a_k}^2 p_k = \Delta\ (p) - \ \lambda_1\ m\ \text{-}\ \lambda_2 d.$
Следовательно,

$$\lambda_1 = -2\lambda_2 m, \qquad (3)$$

$$\Delta(p) = \lambda_1 m + 2\lambda_2 d = 2\lambda_2 D. \qquad (4)$$

Теперь найдем $\Delta(p)$ традиционным способом: сначала вычислим стационарные точки $\Delta(a,\ p)$, а затем его значения в этих точках.

Из системы $\langle 2 \rangle$ находим a_k, при этом учитываем соотношение $\langle 3 \rangle$:

$$a_1 = -\frac{\alpha_1 + \lambda_1 p_1}{2\ \lambda_2 p_1} \ = \frac{-\alpha_1 + \ 2\lambda_2 m p_1}{2\ \lambda_2 p_1},$$

а для $k=2, \ldots, n\text{-}1$

$$a_k = \frac{\alpha_{k-1} - \alpha_k - \lambda_1 p_k}{2\lambda_2 p_k} = \frac{(2\sum_{i=1}^{k-1} p_i \sum_{j=k}^{n} p_j - 2\ \sum_{i=1}^{k} p_i \sum_{j=k+1}^{n} p_j + 2\lambda_2 m\ p_k)}{2\ \lambda_2\ p_k} =$$

$$= \frac{(2p_k(\sum_{i=1}^{k-1} p_i - \sum_{j=k+1}^{n} p_j) + 2\lambda_2 m p_k\ )}{2\ \lambda_2\ p_k} = \frac{1}{\lambda_2}\left(\sum_{i=1}^{k-1} p_i - \sum_{j=k+1}^{n} p_j\right) + m,$$

$$a_n = \frac{\alpha_{n-1} + 2\lambda_2 m p_n}{2\lambda_2 p_n}.$$

Таким образом, координаты стационарных точек имеют вид

$$a_k = \frac{1}{\lambda_2}\left(\sum_{i=1}^{k-1} p_i - \sum_{j=k+1}^{n} p_j\right) + m; \quad k = 1, \ldots,\ n. \qquad \langle 5 \rangle$$

Подставляя их в выражение для $\Delta(a,\ p)$, получим

$$\Delta\,(p) = \max_a \sum_{k=1}^{n-1} (a_{k+1} - a_k)\,\alpha_k = \frac{1}{\lambda_2}\sum_{k=1}^{n-1} (p_{k+1} + p_k)\,\alpha_k\ . \qquad \langle 6 \rangle$$

Теперь, приравнивая правые части соотношений $\langle 4 \rangle$ и $\langle 6 \rangle$, определяем λ_2 :

$$\lambda_2 = \sqrt{\frac{1}{2D}\sum_{k=1}^{n-1} (p_{k+1} + p_k)\,\alpha_k \quad}\ . \qquad \langle 7 \rangle$$

Подставив это выражение λ_2 в $\langle 6 \rangle$, получим

$$\Delta\,(p) = \sqrt{2D\sum_{k=1}^{n-1} (p_{k+1} + p_k)\,\alpha_k \qquad}\ . \qquad \langle 8 \rangle$$

Переходим ко второму этапу доказательства леммы 1.

Введем функцию Лагранжа

$$\mathcal{L}\,(p) = \sum_{k=1}^{n-1} (p_{k+1} + p_k)\,\alpha_k - \lambda_3\left(\sum_{i=1}^{n} p_i - 1\right),$$

и для нахождения стационарных точек распишем ее в развернутом виде:

$$\mathcal{L}\,(p) = 2\,(p_1 + p_2)\,p_1\,(p_2 + \cdots + p_n) + 2\,(p_2 + p_3)\,(p_1 + p_2)\,(p_3 + \cdots + p_n) + \cdots +$$

$$+2\,(p_{k-2} + p_{k-1})\,(p_1 + \cdots + p_{k-2})\,(p_{k-1} + p_k + \cdots + p_n) +$$

$$+2\,(p_{k-1} + p_k)\,(p_1 + \cdots + p_{k-1})\,(p_k + \cdots + p_n) +$$

$$+2\,(p_k + p_{k+1})\,(p_1 + \cdots + p_k)\,(p_{k+1} + \cdots + p_n) + \cdots +$$

$$+2\left(p_{n-1}+p_n\right)\left(p_1+\cdots+p_{n-1}\right)p_n-\lambda_3\left(p_1+\cdots+p_n-1\right).$$

Отсюда, последовательно имеем

$$0=\frac{\partial\mathcal{L}\left(p\right)}{\partial p_1}=2\left(2p_1+p_2\right)\left(p_2+\cdots+p_n\right)+2\left(p_2+p_3\right)\left(p_3+\cdots+p_n\right)+\cdots+$$

$$+2\left(p_{n-1}+p_n\right)p_n-\lambda_3=2\left[\left(p_1+\cdots+p_n\right)^2-{p_1}^2-\frac{1}{2}\lambda_3\right],$$

$$0=\frac{\partial\mathcal{L}\left(p\right)}{\partial p_k}=2\left(p_1+p_2\right)p_1+2\left(p_2+p_3\right)\left(p_1+p_2\right)+\cdots+$$

$$+2\left(p_{k-2}+p_{k-1}\right)\left(p_1+\cdots+p_{k-2}\right)+2\left(p_1+\cdots+p_{k-1}\right)\cdot$$

$$\cdot\left(p_{k-1}+2p_k+p_{k+1}+\cdots+p_n\right)+$$

$$+2\left(p_{k+1}+\cdots+p_n\right)\left(p_1+\cdots+p_{k-1}+2p_k+p_{k+1}\right)+$$

$$+2\left(p_{k+1}+p_{k+2}\right)\left(p_{k+2}+\cdots+p_n\right)+2\left(p_{n-1}+p_n\right)p_n-\lambda_3=$$

$$=2\left[\left(p_1+\cdots+p_n\right)^2-{p_k}^2-\frac{1}{2}\lambda_3\right],$$

$$0=\frac{\partial\mathcal{L}\left(p\right)}{\partial p_n}=2\left(p_1+p_2\right)p_1+2\left(p_2+p_3\right)\left(p_1+p_2\right)+$$

$$+2\left(p_3+p_4\right)\left(p_1+p_2+p_3\right)+\cdots+2\left(p_1+p_2+\cdots+p_{n-1}\right)\left(p_{n-1}+2p_n\right)-\lambda_3=$$

$$=2\left[\left(p_1+\cdots+p_n\right)^2-{p_n}^2-\frac{1}{2}\lambda_3\right].$$

Из этих равенств следует, что

$$\frac{1}{2}\lambda_3=1-{p_1}^2=1-{p_2}^2=\cdots=1-{p_n}^2,$$

и, следовательно, $p_1=p_2=\cdots=p_n=\frac{1}{n}$.
При этих значениях $p_i,\ i=1,\ \ldots,\ n,$ как нетрудно убедиться, имеем

$$\sum_{k=1}^{n-1}\left(p_{k+1}+p_k\right)\alpha_k=\frac{2}{3}\frac{n^2-1}{n^2}.$$

Подставляя это выражение в $\langle 7\rangle$, последовательно из $\langle 8\rangle$ и $\langle 5\rangle$, находим

$$\lambda_2=\sqrt{\frac{n^2-1}{3Dn^2}},\qquad\qquad \Delta_n=2\frac{\sqrt{3}}{3}\sqrt{D\left(1-\frac{1}{n^2}\right)},$$

$$a_k=\sqrt{\frac{3D}{n^2-1}}\left(2k-n-1\right)+m. \qquad \langle 9\rangle$$

Таким образом, лемма 1 доказана.

На самом деле мы доказали теорему 2 в дополнительных условиях: для с.в. X и Y с фиксированными математическим ожиданием и дисперсией

$$EX = EY = m, \qquad DX = DY = D.$$

Следующая лемма позволяет перейти к общему случаю.

Лемма 2.

Пусть X и Y – н.о.р.с.в. с конечной дисперсией. Тогда, для произвольных $a,\ b > 0$, существуют н.о.р.с.в. X_1 и Y_1 такие, что

$$\mathrm{EX}_1 = \mathrm{EY}_1 = a, \qquad D\mathrm{X}_1 = D\mathrm{Y}_1 = b \qquad \text{и} \qquad \frac{\mathrm{E}\,|\mathrm{X}-\mathrm{Y}|}{\sqrt{DX}} = \frac{1}{b}\mathrm{E}\,|\mathrm{X}_1-\mathrm{Y}_1|\ .$$

Доказательство:

Действительно, достаточно взять

$$\mathrm{X}_1 = \frac{\sqrt{b}}{\sqrt{DX}}\,(\mathrm{X}-\mathrm{EX})+a, \qquad \mathrm{Y}_1 = \frac{\sqrt{b}}{\sqrt{D\mathrm{Y}}}\,(\mathrm{Y}-\mathrm{EY})+a. \qquad \langle 10\rangle$$

Доказательство теоремы 2: Пусть X и Y – н.о.р.с.в. с $\mathrm{EX}^2 = \mathrm{EY}^2 < \infty$.Преобразуем их по формулам $\langle 10\rangle$. Ясно, что с.в. X_1 и Y_1 удовлетворяют условиям леммы 1. Поэтому справедливо неравенство

$$\frac{1}{b}\mathrm{E}\,|\mathrm{X}_1-\mathrm{Y}_1| \le \frac{2}{3}\sqrt{3\left(1-\frac{1}{n^2}\right)}\,, \qquad \langle 11\rangle$$

причем равенство в нем достигается на распределениях

$$\mathrm{P}\left(\mathrm{X}_1 = \frac{\sqrt{3b}}{\sqrt{n^2-1}}\,(2k-n-1)+a\right) = \frac{1}{n}\ . \qquad \langle 12\rangle$$

В силу леммы 2 и неравенства $\langle 11\rangle$, будем иметь

$$\frac{\mathrm{E}\,|\mathrm{X}\ -Y\ |}{\sqrt{DX}} = \frac{1}{b}\mathrm{E}\,|\mathrm{X}_1-\mathrm{Y}_1| \le \frac{2}{3}\sqrt{3\left(1-\frac{1}{n^2}\right)};$$

причем равенство достигается на распределениях $\langle 12\rangle$ или, что тоже самое,

$$\mathrm{P}\left(\mathrm{X} = \frac{\sqrt{3DX}}{\sqrt{n^2-1}}\,(2k-n-1)+\mathrm{EX}\right) = \frac{1}{n}.$$

Теорема 2 доказана.

Д о к а з а т е л ь с т в о теоремы 1 следует из теоремы 2 и из определения математического ожидания от неотрицательных с.в. (как интеграла Лебега от неотрицательной функции) как предела простых с.в. (см., напр., [4], гл.II, §6).

Кроме того, следующее, легко проверяемое, соотношение

$$\frac{2}{3}\sqrt{3} = \frac{\int_0^1 \int_0^1 |x-y|\, dxdy}{\sqrt{\int_0^1 x^2 dx - \left(\int_0^1 x dx\right)^2}},$$

показывает достижимость супремума на равномерном на [0,1] распределении.

Выражаем благодарность М.Т. Бакоеву за полезные обсуждения и помощь в подсчете коэффициента Джини для конкретных распределений.

ЛИТЕРАТУРА

1. Золотарев В.М. Современная теория суммирования независимых случайных величин. – М.: Наука, 1986.

2. Кендалл М.Дж., Стьюарт А. Теория распределений. - М.: Наука, 1966.

3. Ширяев А.Н. Вероятность. – М.: Наука, 1980.

4. Gini C. Variabilita e Mutabilita contributo allo studio delle distribuzioni e relazioni statistiche. Studi Economico-Giuridici della R.Universita di Cagliari. 1912.

5. Sharakhmetov. Sh. Sh., Ismatullaev Sh. A. Sharp bounds for the Gini coefficient. / Proceedings of the 5^{th} Uzbekistan – Indonesia International Joint Conference on Globalization, Economic Development, and Nation Character Building, October 29-30, Indonesia, 2015, p.p.37-39.

$\Psi-$ Гармоническая мера

К.К.Кулдашев
Национальный университет Узбекистана
qobil2407@mail.ru

Пусть $D \subset \mathbb{R}^n$ регулярное ограниченное область, т.е. существует $\rho(x) \in sh(D) : \rho(x) < 0, \lim\limits_{x\to\partial D} \rho(x) = 0$ и $E \subset D -$ произвольное множество. $\psi(x)$ некоторая непрерывная, строго отрицательная функция на D.

Рассмотрим класс функций

$$\mathcal{U}(E, D, \psi) = \Big\{u(x) \in sh(D) : u(x)|_E \le \psi(x),\ u(x)|_D < 0\Big\}$$

и положим

$$\omega(x, E, D, \psi) = \sup\{u(x) : u(x) \in \mathcal{U}(E, D, \psi)\},$$

$$\omega^*(x, E, D, \psi) = \overline{\lim_{y\to x}}\, \omega(y, E, D, \psi).$$

Определение 1. $\omega^*(x, E, D, \psi)$ называется ψ-гармонической мерой множества E относительно области D.

Определение 2. Точка $x^0 \in K$ называется глобально ψ-регулярной точкой компакта, если $\omega^*(x^0, K, D, \psi) = \psi(x^0)$. Компакт называется ψ-регулярным компактом, если он ψ-регулярен в любой своей точке. Множество E называется локально ψ-регулярным в точке $x^0 \in \bar{E}$, если для любого $r > 0$ функция $\omega^*(x^0, \overline{E \cap B(x^0, r)}, D, \psi) = \psi(x^0)$.

Отметим, что $\omega^*(x, E, D, -1)$ совпадает с гармонической мерой теории потенциала, т.е. $\omega^*(x, E, D, -1) = \omega^*(x, E, D)$ [1, 2]. Функция $\omega^*(x, E, D, \psi)$ удовлетворяет многим свойствам функции $\omega^*(x, E, D)$.

1. $\omega^*(x, E, D, \psi) \in sh(D) \cap h(D \setminus \overline{E})$.
2. Если $\psi_1 \le \psi_2$, то $\omega(x, E, D, \psi_1) \le \omega(x, E, D, \psi_2),\ \forall x \in D$.
3. Если $E_1 \subset E_2 \subset D_1 \subset D_2$, то для любого $x \in D_1$ имеет место неравенства
$$\omega(x, E_2, D_2, \psi) \le \omega(x, E_2, D_1, \psi) \le \omega(x, E_1, D_1, \psi).$$
4. $\omega^*(x, E, D, \psi)$ либо нигде равна нулю, либо тождественно равна нулю. $\omega^*(x, E, D, \psi) \equiv 0$ тогда и только тогда, когда E полярное в D.
5. Если $E \subset\subset D$, то $\lim\limits_{x\to\partial D} \omega^*(x, E, D, \psi) = 0$.

6. Пусть $E = \bigcup\limits_{j=1}^{\infty} E_j$, то для любого $x \in D$ имеет место неравенство

$$\omega^*(x, E, D, \psi) \geq \sum_{j=1}^{\infty} \omega^*(x, E_j, D, \psi).$$

7. Если $\psi \in sh(D)$, то $\psi(x) \leq \omega^*(x, E, D, \psi)$ и $\psi(x)|_{E^0} \equiv \omega^*(x, E, D, \psi)|_{E^0}$

8. Если $U \subset D$ открытое множество, $\psi \in sh(D)$ и $U = \bigcup\limits_{j=1}^{\infty} K_j$, где $K_j \subset K_{j+1}^0$ компакты, то $\omega^*(x, K_j, D, \psi) \downarrow \omega^*(x, U, D, \psi)$.

9. Если $E \subset D$ произвольное множество, то существует убывающая последовательность открытых множеств, $U_j \supset E, U_j \supset U_{j+1}$ такая, что

$$(\lim_{j\to\infty} \omega(x, U_j, D, \psi))^* = \omega^*(x, E, D, \psi).$$

10. **a)** Для любого компакта $K \subset D$ имеет место неравенство

$$-\min_K \psi(x) \cdot \omega^*(x, K, D) \leq \omega^*(x, K, D, \psi) \leq -\max_K \psi(x) \cdot \omega^*(x, K, D).$$

b) Точка $x^0 \in K$ является локально ψ-регулярной тогда и только тогда, когда она является локально регулярной точкой компакта K.

11. Если $K - \psi$-регулярный компакт, то $\omega^*(x, K, D, \psi) \equiv \omega(x, K, D, \psi) \in C(\overline{D})$ и $\omega^*|_{\partial D} = 0$

ЛИТЕРАТУРА

1. Садуллаев А. Теория плюрипотенциала. Применения. Германия: Palmarium Academic Publishing, 2012. 64 с.

2. Садуллаев А. Плюрисубгармонические меры и емкости на комплексных многообразиях. АН СССР: Успехи математических наук, 1981, № 4, 35-105 с.

Qobiljon Quldashev (Tashkent, Uzbekistan)

ψ - harmonic measure

Abstract. *In this thesis we define ψ - harmonic measure.*

Key words: *ψ - harmonic measure, harmonic measure.*

Maximality Criteria for Subharmonic and Plurisubharmonic functions

Sh.Ibragimov
National University of Uzbekistan
ibragimovshohrux@gmail.com

Abstract. *It is known that there are several maximality criteria for subharmonic and plurisubharmonic functions. In this paper, we mention another criteria, more simple than others.*

Key words: *Harmonic function, Subharmonic function, Plurisubharmonic function, Maximal function, Complex line.*

Let us given the domain $D \subset \mathbb{C}^n$.

Definition. A function $u \in psh(D)$ $(u \in sh(D))$ is called **maximal** in the domain D if the following maximal principle holds: For any function $v \in psh(D)$ $(v \in sh(D))$, $\liminf_{z\to\partial D}(u(z) - v(z)) \geq 0$ implies $u(z) \geq v(z)$, $\forall z \in D$.

The condition $\liminf_{z\to\partial D}(u(z) - v(z)) \geq 0$ in the above definition means that for arbitrary fixed $\forall \varepsilon > 0$ there exists a compact $K_\varepsilon \subset\subset D$ such that $v(z) \leq u(z) + \varepsilon$, $\forall z \in D \setminus K_\varepsilon$.

Next lemma is useful in many applications.

Lemma. *The followings are equivalent:*

(1) $u \in psh(D)$ $(u \in sh(D))$ is maximal in D;

(2) For any domain $G \subset\subset D$ and for any function $v \in psh(G)$ $(v \in sh(G))$: $\liminf_{z\to\partial G}(u(z) - v(z)) \geq 0$ *implies* $u(z) \geq v(z), \forall z \in G$;

(3) For any domain $G \subset\subset D$ and for any function $v \in psh(G)$ $(v \in sh(G))$: $u|_{\partial G} \geq v|_{\partial G}$ *implies* $u(z) \geq v(z), \forall z \in G$.

For a proof see A.Sadullaev[2].

It is well-known that the equation $\Delta u = 0$ defines harmonic functions on the complex plane $\mathbb{C}$.

Theorem. *If u is harmonic in a domain $D \subset \mathbb{C}$, then $u(z) \geq v(z)$, $\forall z \in D$ for any function $v \in sh(D)$, such that* $\liminf_{z\to\partial D}(u(z) - v(z)) \geq 0$. *In other words, the harmonic function u has the maximality property in the class of sh functions.*

Next theorem is maximality criteria for locally bounded *psh* functions which is considered by E.Bedford and B.A.Taylor [1].

Theorem. *A function $u \in psh(D) \cap L^\infty_{loc}(D)$ is maximal if and only if $(dd^c u)^n = 0$, where $(dd^c *)^n$ is the complex Monge Ampere operator.*

Motivating above-mentioned facts and taking the advice from academician A.Sadullaev, it is considered to be true the following hypothesis.

Proposition. *Let u be plurisubharmonic function in $D \subset \mathbb{C}^n$ ($u \in psh(D)$). Assume for any $z^0 \in D$, there exists a complex line $l \subset D$ ($z^0 \in l$), such that u is harmonic in $l \cap D$. Then u is maximal in D.*

REFERENCE

1. E. Bedford and B. A.Taylor, A new capacity for plurisubharmonic functions. Acta Math, 1982. No. 1-2, 1-40.**149**

2. A.Sadullaev, On maximal plurisubharmonic functions. Complex Analysis and Potential Theory, November 2012. p. 211–216.

Singular probability measure of circle homeomorphisms with several critical points

U.A. Safarov
Turin polytechnic university in Tashkent
safarovua@mail.ru

We consider orientation preserving circle homeomorphisms f with lift $F : R^1 \to R^1$, which is continuous, strictly increasing and $F(x+1) = F(x) + 1$, for any $x \in R^1$. The most important arithmetic characteristic of the homeomorphism f of the unit circle S^1 is the rotation number

$$\rho_f = \lim_{n\to\infty} \frac{F^n(x)}{n} (mod 1),\ x \in R^1.$$

Here and below, F^n denotes the nth iterate of the function F. The rotation number is rational if and only if f has periodic points. Denjoy proved, that if f is a circle diffeomorphism with irrational rotation number $\rho = \rho(f)$ and $\log Df$ is of bounded variation, then f is topologically conjugate to the pure rotation $f_\rho : x \to x + \rho(mod 1)$; that is, there exists an essentially unique homeomorphism φ of the circle with $\varphi(f(x)) = f_\rho(\varphi(x)),\ x \in S^1$ (see.[1]).

Since the conjugating map φ and the unique f-invariant measure μ_f are related by $\varphi(x) = \mu_f([0, x])$ (see.[1]), regularity properties of the conjugating map φ imply corresponding properties of the density of the absolutely continuous invariant measure μ_f. The problem of relating the smoothness of φ was studied extensively by M.Herman, Y. Katznelson and D. Ornstein, K. Khanin and Ya. Sinai.

A natural extension of circle diffeomorphisms are the class of **critical** circle homeomorphisms.

Definition 1. *The point $x_{cr} \in S^1$ is called non-flat critical point of a homeomorphism f with order $d \in R^1$, $d > 2$, if for a some $\delta-$ neighborhood $U_\delta(x_{cr})$, the homeomorphism f can be written as $f(x) = \phi(x)|\phi(x)|^{d-1} + f(x_{cr})$, where ϕ is a C^3 local diffeomorphism with $\phi(x_{cr}) = 0$.*

Yoccoz in [2] generalized Denjoy's classical result, a critical circle homeomorphism with irrational rotation number is topologically conjugate to an irrational rotation.

The singularity of invariant probability measure for the class critical circle homeomorphisms was shown by J. Graczyk and G. Swiatek in [3]. They proved that if f is C^3 smooth circle homeomorphisms with finitely many critical points of integer orders and an irrational rotation number of bounded type, then the conjugating map φ is a singular function.

Now we formulate our main result.

Theorem *Let f be a C^2 circle homeomorphism with irrational rotation number ρ. Suppose that the homeomorphism has critical points $x_{cr}^{(k)}$, $k = \overline{1,m}$ with orders $d_k > 2$. Then the f-invariant probability measure μ_f is singular w.r.t. Lebesgue measure ℓ on the circle S^1.*

REFERENCE

1. I.P. Cornfeld , S.V. Fomin and Ya.G. Sinai : *Ergodic Theory*, Springer Verlag, Berlin (1982).

2. J.C. Yoccoz: Il n'a a pas de contre-exemple de Denjoy analytique. C. R. Acad. Sci., Paris, Ser.I,Math. **298**(7),1984, P.141-144.

3. J. Graczyk ,G. Swiatek: Singular measures in circle dynamics. Commun. Math. Phys., **157**, P.213-230, 1993.

Safarov Utkir Avlayarovich (Tashkent, Uzbekistan)

Singular probability measure of circle homeomorphisms with several critical points

Abstract. *We prove that, invariant probability measure of a circle homeomorphism with several critical points is singular w.r.t. Lebesgue measure.*

Key words: *rotation number, critical cpoint, invariant measure, singular*

Двувыборочные критерии согласия при случайном цензурировании справа

А.Л.Абдулвахидов [1] **Д.Г.Захидов** [2]
[1]Андижанских Государственный университет
[2]Филиал Ташкентского Государственного Аграрного Университета в г. Андижан

1. Введение

Двувыборочным задачам при случайном цензурировании наблюдений посвящено много работ. В них предложены статистики для проверки гипотезы $H_0: \ F_1 = F_2 = F$, где $F_1(t)$ и $F_2(t)$, $t\epsilon R$, - функции распределения (ф.р.) двух генеральных совокупностей и F – вообще говоря, неизвестная ф.р.. Эти статистики получаются в основном заменой эмпирических оценок ф.р. F_1 и F_2 на их аналоги при случайном цензуриравании. Так, в работах [1,2] в модели случайного цензурирования справа для $F_k, k = 1, 2$, использована *PL*- оценка, которая обладает рядом недостатков по сравнению, скажем, со степенной оценкой отношения рисков А.Абдушукурова [3,4]. В данной работе нами использованы степенные оценки для ф.р. $F_k, k = 1, 2$, для построения непараметрических критериев проверки гипотезы H_0 в случае цензурирования наблюдений справа.

2. Постановка задачи и основные результаты

Пусть $X_{11}, X_{12}, \ldots, X_{1n_1}$ и $X_{21}, X_{22}, \ldots, X_{2n_2}$ – две независимые повторные выборки обьёмов n_1и n_2 с непрерывными ф.р. $F_1(t)$ и $F_2(t)$, $t\epsilon R$, соответственно. Предположим, что эти выборки цензурированы справа независящими от них с.в. $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ и $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$соответственно с непрерывными ф.р. $G_1(t)$ и $G_2(t)$, $t\epsilon R$. Наблюдаются две выборки $C^{(n_k)} = \left\{(Z_{kj}, \delta_{kj}) \ \ j = \overline{1, n_k}\right\}$, где $Z_{kj} = min(X_{kj}, Y_{kj})$, $\delta_{kj} = I(X_{kj} \le Y_{kj})$ и $I(A)$- индикатор события A; здесь, а также всюду в дальнейшем $k = 1, 2$. Заметим, что с.в. Z_{kj} имеют общую ф.р. $H_k(t) = 1 - \overline{F_k(t)} \cdot \overline{G_k(t)}$, где $\overline{F_k} = 1 - F_k$, $\overline{G_k} = 1 - G_k$. Предположим, что $\lim_{n_1,n_2} \frac{n_k}{n} = q_k > 0$ где $n = n_1 + n_2$ объём объединённой выборки. Пусть $H_k^{(1)}(t) = P(Z_{kj} \le t, \delta_{kj} = 1) = \int_{-\infty}^{t} (1 - G_k(u)) dF_k(u)$ субраспределения. Определим эмпирические оценки для H_kи $H_k^{(1)}$:

$$H_{kn_k}(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} I(Z_{kj} \le t), \ H_{kn_k}^{(1)}(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \delta_{kj} I(Z_{kj} \le t).$$

Введём соответствующие эмпирические процессы $\alpha_{kn_k}(t) = \sqrt{n_k}\left(H_{kn_k}(t) - H_k(t)\right)$ и $\alpha_{kn_k}^{(1)}(t) = \sqrt{n_k}\left(H_{kn_k}^{(1)}(t) - H_k^{(1)}(t)\right)$. Функции

$H(t) = q_1 H_1(t) + q_2 H_2(t)$ и $H_N(t) = \frac{n_1}{n} H_{1n_1}(t) + \frac{n_2}{n} H_{2n_2}(t)$ являются соответственно теоретической и эмпирической ф.р. объединённой выборки $\left\{\left[Z_{1i}, i = \overline{1, n_1}\right], \quad \left[Z_{2j}, j = \overline{1, n_2}\right]\right\}$ объёма n и $N = (n_1; n_2)$. Пусть $\rho(t) = F_1(t) - F_2(t)$. Тогда гипотеза H_0 равносильна равенству $\rho(t) = 0$ для всех $t \epsilon R$. Введём оценку для $\rho(t)$: $\rho_N(t) = F_{1n_1}(t) - F_{2n_2}(t), t \epsilon R$, где $F_{kn_k}(t)$ – степенные оценки для $F_k(t)$ автора [3,4]:

$$F_{kn_k}(t) = 1 - [1 - H_{kn_k}(t)]^{R_{kn_k}(t)}, \ t \epsilon R, \tag{1}$$

где $R_{kn_k}(t) = \Lambda^{(1)}_{kn_k}(t) [\Lambda_{kn_k}(t)]^{-1}$ и

$$\Lambda_{kn_k}(t) = \int_{-\infty}^{t} \frac{dH_{kn_k}(u)}{1 - H_{kn_k}(u-)},$$

$$\Lambda^{(1)}_{kn_k}(t) = \int_{-\infty}^{t} \frac{dH^{(1)}_{kn_k}(u)}{1 - H_{kn_k}(u-)},$$

оценки соответствующие интегральным функциям интенсивности

$$\Lambda_k(t) = -\log(1 - H_k(t)) \ , \ \Lambda^{(1)}_k(t) = -\log(1 - F_k(t)) \ .$$

Обозначим:

$$d^*(t) = (q_1^*)^{-1} \cdot d_1(t) + (q_2^*)^{-1} \cdot d_2(t),$$

где $q_k^* = \frac{n_k}{n}$ и

$$d_k(t) = \int_{-\infty}^{t} \left[\left(\overline{F_k(u)}\right)^2 \overline{G_k(u)}\right]^{-1} dF_k(u).$$

Определим нормированный процесс

$$\pi_N(t) = \sqrt{n} \rho_N(t) \left[\overline{F_N(t)}\right]^{-1},$$

где

$$\overline{F_N(t)} = \frac{n_1}{n} \overline{F_{1n_1}(t)} + \frac{n_2}{n} \overline{F_{2n_2}(t)}$$

оценка для $\overline{F(t)} = 1 - F(t)$ по обьединенной выборке.

Пусть $\{W_N(t), \quad t \epsilon [0, \infty)\}$ последовательность винеровских процессов с нулевым средним и ковариацией $EW_N(t) W_N(s) = d^*(min(t, s))$. Приведём результат об аппроксимации последовательности нормированых процессов $\pi_N(t)$ последовательностью винеровских процессов $W_N(t)$ на интервале $(-\infty, \mathrm{T}]$, где

$$\mathrm{T} = \min(T_1, T_2), \quad T_k < T_{H_k} = inf\left\{t \epsilon R^1 : H_k(t) = 1\right\} .$$

Теорема. При справедливости гипотезы H_0, и $n_1, n_2 \to \infty$ имеем

$$\sup_{t\leq T} |\pi_n(t) - W_N(t)| \xrightarrow{p} 0. \tag{2}$$

Результат (2) при подходящем выборе числа T можно использовать для построения статистик для проверки гипотезы H_0. Пусть $\left\{W^{\circ}(y),\ \ 0 \leq y \leq 1\right\}$ – стандартный винеровский процесе на $[0,1]$. Тогда из (2), в частности следует, что

$$\overset{\circ}{\pi}_N(t) = [d^*(T)]^{-1/2} \cdot \pi_N(t) \overset{D}{\Rightarrow} W^{\circ}\left(\frac{d(t)}{d(T)}\right), \tag{3}$$

где $\overset{D}{\Rightarrow}$ означает сходимость по распределению и $d(t) = q_1^{-1}d_1(t) + q_2^{-1}d_2(t)$. Теперь (3), в свою очередь, влечет сходимость

$$K_N = \sup_{t\leq T} \left|\overset{\circ}{\pi}_N(t)\right| \overset{D}{\Rightarrow} \sup_{0\leq y\leq 1} \left|W^{\circ}(y)\right| = K^{\circ},$$

т.е. статистики K_N можно рассматривать в качестве аналога статистики Колмогорова для проверки справедливости H_0.

Литература

1. Абдушукуров А.А. О некоторых вариантах статистик двух выборочных критериев при случайном цензурировании //Сб. <<Асимптотические методы в теории вероятностей и математической статистике>> Ташкент. Фан. 1988. с. 3-18

2. Schumacher M. Two-sample tests of Cramer-von Mises and Kolmogorov-Smirnov-type for randomly censored data. //Internat. Statist. Rewiew. 1984 v.52 N.3. p.263-281.

3. Абдушукуров А.А. Статистика неполных наблюдений. Ташкент – Университет. 2009. 269 стр.

4. Abdushukurov A.A. Nonparametric estimation based on incomplete observations. // Internat. Enciclopedia of Statistical Sciences. Springer 2011. Part 14. p.462-464.

Полупараметрическая квантильная оценка процентной остаточной продолжительности жизни в информативной модели конкурирующих рисков

Ф.М.Холмуродов
Наманганский государственный университет
hfarhod@bk.ru

1. Введение и постановка задачи.

Рассмотрим информативную модель конкурирующих рисков, следуя работам [1,2]. Пусть $\{(X_i, Y_{1i}, Y_{2i}), i = 1, 2, ...\}$ последовательность независимых и одинаково распределенных (н.о.р.) случайных векторов с независимыми неотрицательными компонентами X_i, Y_{1i} и Y_{2i} с непрерывными функциями распределения (ф.р.) $F(t) = P(X_{1i} \leq t), G_k(t) = P(Y_{ki} \leq t), k = 1, 2; t \in R^+$ $k = 1, 2; t \in R^+$, соответственно. Предполагается, что пара (F, G_1) отвечает модели пропорциональных интенсивностей (МПИ), согласно которой существует число $\beta > 0$ и для всех $t \in R^+$ имеет место равенство

$$1 - G_1(t) = (1 - F(t))^{\beta}, \tag{1}$$

что эквивалентно пропорциональности интегральных функций интенсивностей (и.ф.и.) $[-\log(1 - G_1(t))] = [-\beta \log(1 - F(t))]$. Одним из важных свойств МПИ является то, что равенство (1) эквивалентно независимости случайных величин (с.в.) $\xi_i = \min(X_i, Y_{1i})$ и$I(\xi_i = X_i)$, где $I(A)$-индикатор события A. Статистическая модель такова, что наблюдается выборка $S^{(n)} = \{(z_i, \delta_i), i = 1, ..., n\}$, где $z_i = \min(X_i, Y_{1i}, Y_{2i})$ и

$$\delta_i = \begin{cases} -1, Y_{2i} \leq \min(X_i, Y_{1i}), \\ 0, Y_{1i} \leq \min(X_i, Y_{2i}), \\ 1, X_i \leq \min(Y_{1i}, Y_{2i}). \end{cases}$$

В этой модели интерес представляют с.в. X_i,которые наблюдаемы лишь в случае $\delta_i = 1$ и в остальных случаях цензурируется мешающей парой с.в. (Y_{1i}, Y_{2i}). В данной информативной модели конкурирующих рисков при оценивании ф.р. F определим субраспределения $\tilde{H}(t) = P(z_i \leq t, \delta_i \neq -1)$ и $\tilde{\tilde{H}}(t) = P(z_i \leq t, \delta_i \neq -1)$, и их эмпирические оценки $\tilde{H}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(z_i \leq t, \delta_i \neq -1)$,$\tilde{\tilde{H}}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(z_i \leq t, \delta_i \neq -1)$. Также определим ф.р. $H(t) = P(z_i \leq t)$ и $K(t) = 1 - (1 - F(t))(1 - G_1(t)) =$

$1-(1-F(t))^{\frac{1}{\gamma}}$, $t \in R^+$, где $\gamma = \frac{1}{1+\beta} = P(\delta_i \leq 1, \delta_i \neq -1) = \frac{P_1}{(P_0+P_1)}$, $P_k = P(\delta_i = k), k = 0,1$. Очевидно, естественными оценками Hи γ являются $H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(z_i \leq t) = \tilde{H}_n(t) + \tilde{\tilde{H}}_n(t)$ и $\gamma_n = P_{1n}/(P_{0n}+P_{1n})$, где $P_{kn} = \frac{1}{n}\sum_{i=1}^{n} I(\delta_i = k), k = 0,1$. Далее предположим, что $p_0 + p_1 \neq 0$. В работе [1] была предложена следующая полупараметрическая степенная оценка для $F(t)$:

$$F_n(t) = 1 - [1 - K_n(t)]^{\gamma_n} = 1 - [1 - H_n(t)]^{\gamma_n R_n(t)}, t \in R^+, \tag{2}$$

где

$$R_n(t) = \frac{\Lambda_{1n}(t)}{\Lambda_n(t)}, \Lambda_{1n}(t) = \int_0^t \frac{d\tilde{H}_n(u)}{1 - H_n(u-)}, \Lambda_{2n}(t) = \int_0^t \frac{dH_n(u)}{1 - H_n(u-)},$$

$$\Lambda_n(t) = \Lambda_{1n}(t) + \Lambda_{2n}(t).$$

Определим квантильную функцию ф.р. F:

$$Q(y) = F^{-1}(y) = \inf\left\{t \in R^+ : F(t) \geq y\right\}, 0 < y < 1, \tag{3}$$

и ее оценку

$$Q_n(y) = {F_n}^{-1}(y) = \inf\left\{t \in R^+ : F_n(t) \geq y\right\}, 0 < y < 1, \tag{4}$$

получаемую использованием статистики (2).

С учётом информативности рассматриваемой модели, формулы (3) и (4) можем представить в виде

$$Q(y) = Q^K(1 - (1-y)^{1/\gamma}), 0 < y < 1, \tag{5}$$

$$Q_n(y) = {Q_n}^{K_n}(1 - (1-y)^{1/\gamma_n}), 0 < y < 1, \tag{6}$$

где $Q^K(S) = \inf\{t \in R^+ : K(t) \geq s\}$, и $Q_n^{K_n}(s) = \inf\{t \in R^+ : K_n(t) \geq s\}$, $0 < s < 1$-квантильные функции, соответствующие ф.р. K и K_n. В работе [2] были исследованы свойства квантильной оценки (6) и построены аналогичные задачи для другого функционала распределения p-процентной остаточной продолжительности жизни.

2. Основные результаты

Для $0 < p < 1$ введём $(1-p)-$ процентную – остаточную продолжительность жизни $\pi(p;t) = Q(1 - p(1 - F(t))) - t, t > 0, 0 < p < 1$. Естественной оценкой $\pi(p;t)$ является $\pi_n(p;t) = Q_n(1 - p(1 - F_n(t))) - t, t > 0, 0 < p < 1$, где $F_n(t)$-оценка для $F(t)$, определяемая формулой (2).

Установлены следующие асимптотические свойства оценки $\pi_n(p;t)$. Пусть существует плотность $f = F'$.

Теорема 1. Пусть $t > 0$ и $0 < p < 1$ фиксированы. Если Q непрерывна в точке $1-p(1-F(t))$и $Q\left(1-p(1-F(t))\right)$ является единственным решением s неравенства $F(s-) \le 1-(1-F(t)) \le F(s)$, то при $n \to \infty$, $\pi_n(p;t) \xrightarrow{p} \pi(p;t)$.

Определим нормированную функцию $r_n(p;t) = n^{\frac{1}{2}} f(t+\pi(p;t))(\pi_n(p;t) - \pi(p;t))$.

Теорема 2. Пусть $t > 0$ и $0 < p < 1$ фиксированы так, что $H(t+\pi(p,t)) < 1$, а функция $f(Q(\cdot))$-положительна и конечна в точке $1-p(1-F(t))$.

Тогда при $n \to \infty$ $r_n(p;t) \xrightarrow{D} N(0;\sigma^2(p;t))$, где

$$\sigma^2(p;t) = p^2(1-F(t))^2\left[d(t+\pi(p;t)) - d(t)\right], d(t) = \left\{\gamma^2 a(t) + \frac{\gamma(1-\gamma)}{p_0+p_1}\Lambda_1^2(t)\right\},$$

$$a(t) = \int_0^t \frac{d\tilde{H}(u)}{(1-H(u))^2}, \Lambda_1(t) = -\log(1-K(t)).$$

При $\lambda > 0$ определим функции:

$$B_n^{\pm}(\lambda;p;t) = \pi_n(p;t) \pm \lambda n^{-\frac{1}{2}}\sigma_n(p;t),$$

где $\sigma_n^2(p;t) = p^2(1-F_n(t))\left[d_n(t+\pi_n(p;t)) - d_n(t)\right]$, и $d_n(t) = \left\{\gamma_n^2 \int_0^t \frac{d\tilde{H}_n(u)}{(1-H_n(u))^2} + \frac{\gamma_n(1-\gamma_n)}{p_{0n}+p_{1n}}\Lambda_{1n}^2(t)\right\}$, соответствующие оценки для $\sigma^2(p;t)$ и $d(t)$.

В следующем утверждении предложена доверительная полоса для $\pi(p;t)$.

Теорема 3. Если $0 < \alpha < 1$, то в условиях теорем 1 и 2, $\lim\limits_{n\to\infty} P(B_n^-(\lambda_\alpha;p;t) \le \pi(p;t) \le B_n^+(\lambda_\alpha;p;t)) = 1-\alpha$, где $\lambda_\alpha > 0$ решение уравнения $\Phi(\lambda_\alpha) = 1 - \frac{\alpha}{2}$ и $\Phi(t)$- стандартное нормальное распределение.

Следует отметить, что теоремы 1-3 обобщают соответствующие теоремы из [3], доказанные при $\beta = 0$.

ЛИТЕРАТУРА

1.*Абдушукуров А.А., Махмудова Д.* Полупараметрическое оценивание функции распределения в информативной модели конкурирующих рисков.//Стат. методы оценивания и проверки гипотез. Пермь.2010.N.21.с.98-106.

2.*Абдушукуров А.А., Холмуродов Ф.М.* Полупараметрическая оценка квантиля в информативной модели случайного цензурирования. // Статистические методы оценивания и проверки гипотез./ Пермь.Гос.Ун-т.-Пермь.2011.Вып.23.-с.144-151.

3.*Абдушукуров А.А., Холмуродов Ф.М.* Квантильная оценка процентной остаточной продолжительности жизни по неполным наблюдениям.//Матер. респуб. научно-практ. конф. << Статистика и еҷ применения >> 19-20 октября, 2017г. Ташкент НУУз. с.122-124.

On a martingale generated by interval maps

A. Begmatov
National University of Uzbekistan, Tashkent Branch
abdumajidb@gmail.com

Let $\mathcal{A}$ be an alphabet with $d \geq 2$ symbols. Consider the partition of I into d subintervals indexed by $\mathcal{A}$, that is, $\mathcal{P} = \{I_\alpha,\ \alpha \in \mathcal{A}\}$. Let $f : I \to I$ be a bijection. We say that the triple $(f, \mathcal{A}, \mathcal{P})$ is a **generalized interval exchange map** with d intervals (for short g.i.e.m.), if $f|_{I_\alpha}$ is an orientation-preserving homeomorphism for all $\alpha \in \mathcal{A}$.

Let $f : I \to I$ be a g.i.e.m. with alphabet $\mathcal{A}$ and π_0, $\pi_1 : \mathcal{A} \to \{1, ..., d\}$, be bijections such that $\pi_0(\alpha) < \pi_0(\beta)$, iff $I_\alpha < I_\beta$, and $\pi_1(\alpha) < \pi_1(\beta)$, iff $f(I_\alpha) < f(I_\beta)$. We call pair $\pi = (\pi_0, \pi_1)$ the **combinatorial data** associated to the g.i.e.m. f. When appropriate we will also use the notation $\pi = (\pi(1), \pi(2), ..., \pi(d))$ for the combinatorial data of f. We always assume that the pair $\pi = (\pi_0, \pi_1)$ is **irreducible**, that is, for all $j \in \{1, ..., d-1\}$ we have: $\pi_0^{-1}(1, ..., j) \neq \pi_1^{-1}(1, ..., j)$. We say that g.i.e.m. f has **cyclic permutation**, if $\pi_0(\{1, 2, ..., d\}) = \{j+1, ..., d, 1, ..., j\}$, for some $1 \leq j \leq d-1$.

Let $\pi = (\pi_0, \pi_1)$ be the combinatorial data associated to the g.i.e.m f. For each $\varepsilon \in \{0, 1\}$, denote by $\alpha(\varepsilon)$ the last symbol in the expression of π_ε, that is $\alpha(\varepsilon) = \pi_\varepsilon^{-1}(d)$. Let us assume that the intervals $I_{\alpha(0)}$ and $f(I_{\alpha(1)})$ have different lengths. Then the g.i.e.m. f is called **Rauzy-Veech renormalizable**(renormalizable, for short). If $|I_{\alpha(0)}| > |f(I_{\alpha(1)})|$ we say that f is renormalizable of **type 0**. When $|I_{\alpha(0)}| < |f(I_{\alpha(1)})|$ we say that f is renormalizable of **type 1**. Let $I^{(1)}$ be the subinterval of I obtained by removing the shorter of intervals $|I_{\alpha(0)}|$ and $|f(I_{\alpha(1)})|$:

$$I^{(1)} = \begin{cases} I \setminus f(I_{\alpha(1)}), & \text{if type 0,} \\ I \setminus I_{\alpha(0)}, & \text{if type 1.} \end{cases}$$

The **Rauzy-Veech induction** of f is the first return map $R(f)$ to the subinterval $I^{(1)}$. We want to see $R(f)$ is again g.i.e.m. with the same alphabet $\mathcal{A}$. For this we need to associate to this map an $\mathcal{A}$ - indexed partition of its domain. Denote by $I_\alpha^{(1)}$ the subintervals of $I^{(1)}$. Let f be renormalizable of type

0. Then the domain of $R(f)$ is the interval $I^{(1)} = I \setminus f(I_{\alpha(1)})$ and we have

$$R(f)(x) = \begin{cases} f(x), & \text{if } \ x \in I_\alpha^{(1)} \ \text{ and } \alpha \neq \alpha(1), \\ f^2(x), & \text{if } \ x \in I_{\alpha(1)}^{(1)}. \end{cases}$$

If f is renormalizable of type 1, the domain of $R(f)$ is the interval $I^{(1)} = I \setminus I_{\alpha(0)}$ and we have

$$R(f)(x) = \begin{cases} f(x), & \text{if } \ x \in I_\alpha^{(1)} \ \text{ and } \alpha \neq \alpha(0), \\ f^2(x), & \text{if } \ x \in I_{\alpha(0)}^{(1)}. \end{cases}$$

It is easy to see, that $R(f)$ is a bijection on $I^{(1)}$ and an orientation-preserving homeomorphisms on each $I_\alpha^{(1)}$. Moreover, the alphabet $\mathcal{A}$ for f and $R(f)$ remains the same.

The triple $(R(f), \mathcal{A}, \mathcal{P}^1)$ is called the **Rauzy-Veech renormalization** of f. We say that a g.i.e.m. f is **infinitely renormalizable**, if $R^n(f)$ is well defined for every $n \in \mathbf{N}$.

Let $(f, \mathcal{A}, \mathcal{P})$ be a g.i.e.m. with d intervals and $\mathcal{P} = \{I_\alpha : \alpha \in \mathcal{A}\}$ be the initial $\mathcal{A}$- indexed partition of I. Suppose that f is infinitely renormalizable. Let $I^{(n)}$ be the domain of $R^n(f)$. Note that $I^{(n)}$ is the nested sequence of subintervals, with the same left endpoint of I. We want to construct the dynamical partition of I associated to the domain of $R^n(f)$.

Notice that $R(f)$ is g.i.e.m. with d intervals and the intervals $I_\alpha^{(1)}$ generate an $\mathcal{A}$- indexed partition of $I^{(1)}$, denoted by $\mathcal{P}^1$. By induction one can check, that $R^n(f)$ is g.i.e.m. with d intervals. Let $\mathcal{P}^n = \{I_\alpha^{(n)} : \alpha \in \mathcal{A}\}$ be the $\mathcal{A}$- indexed partition of $I^{(n)}$, generated by $R^n(f)$. We call $\mathcal{P}^n$ the **fundamental partition** and $I_\alpha^{(n)}$ the **fundamental segments** of rank n.

Since $R^n(f)$ is the first return map for f to the interval $I^{(n)}$, each fundamental segment $I_\alpha^{(n)} \in \mathcal{P}^n$ returns to $I^{(n)}$ under certain iterates of the map f. Until returning, these intervals will be in the interval $I \setminus I^{(n)}$ for some time. Consequently the system of intervals (their interiors are mutually disjoint)

$$\mathbf{P}_n = \left\{ f^i(I_\alpha^{(n)}),\ 0 \leq i \leq q_\alpha^n - 1,\ \alpha \in \mathcal{A} \right\}$$

cover the whole interval and form a partition of I, where q_α^n is the first return time for $I_\alpha^{(n)}$ to the interval I.

The system of intervals $\mathbf{P}_n$ is called the **n-th dynamical partition** of I. The dynamical partitions $\mathbf{P}_n$ are refined with increasing n, where $\mathbf{P}_{n+1} \supset \mathbf{P}_n$ means that any element of the preceding partition is a union of a number of elements of the next partition, or belongs to the next partition. Denote by $\mathbf{P}^{pr}_{n+1}$ the system of preserved intervals of $\mathbf{P}_n$. More precisely, if $R^n f$ has type ε

$$\mathbf{P}^{pr}_{n+1} = \left\{ f^i(I^{(n)}_\alpha),\ 0 \le i \le q^n_\alpha - 1, \quad \text{for } \alpha \neq \alpha(\varepsilon) \right\},\ \ \varepsilon = 0 \vee 1.$$

Let $\mathbf{P}^{tn}_{n+1} := \mathbf{P}_{n+1} \setminus \mathbf{P}^{pr}_{n+1}$ be the set of elements of $\mathbf{P}_{n+1}$ which are properly contained in some element of $\mathbf{P}_n$. Therefore if $R^n f$ has type 0

$$\mathbf{P}^{tn}_{n+1} = \bigcup_{i=0}^{q^n_{\alpha(0)}-1} \left\{ f^i \left(I^{(n)}_{\alpha(0)} \setminus f^{q^n_{\alpha(1)}} I^{(n)}_{\alpha(1)} \right) \right\} \bigcup \bigcup_{i=q^n_{\alpha(1)}}^{q^n_{\alpha(1)}+q^n_{\alpha(0)}-1} \left\{ f^i(I^{(n)}_{\alpha(1)}) \right\},$$

and if $R^n f$ has type 1

$$\mathbf{P}^{tn}_{n+1} = \bigcup_{i=0}^{q^n_{\alpha(1)}-1} \left\{ f^i \left(f^{-q^n_{\alpha(1)}}(I^{(n)}_{\alpha(0)}) \right) \right\} \bigcup \bigcup_{i=0}^{q^n_{\alpha(1)}-1} \left\{ f^i \left(I^{(n)}_{\alpha(1)} \setminus f^{-q^n_{\alpha(1)}}(I^{(n)}_{\alpha(0)}) \right) \right\}.$$

So, the partition $\mathbf{P}_{n+1}$ consists of the *preserving* elements of $\mathbf{P}_n$ and the *images of two (new) intervals* which are arise for defining $R^{n+1}(f)$.

Let $g : I \to I$ be a function of class $\mathbf{L}_p(I, d\ell),\ p > 1$. For simplicity nation we use $L^{(n)}$ to denote an interval $f^i(I^{(n)}_\alpha)$ of the dynamical partition $\mathbf{P}_n(f)$. Define a sequence of random variables $\{\Phi_n\}$ on the interval I as:

$$\xi_n(x) = \frac{|g(L^{(n)})|}{|L^{(n)}|}, \quad \text{if} \quad x \in L^{(n)}.$$

Theorema. *Let $g \in \mathbf{L}_p(I, d\ell),\ p > 1$. Then the sequence of random variables $\{\xi_n(x),\ n \ge 1\}$ generate a finite martingale with respect to the dynamical partition $\boldsymbol{P}_n(f)$.*

References

1. A. Begmatov and K. Cunha: *On the convergence of renormalizations of piecewise smooth homeomorphisms on the circle*, https://arXiv:1807.09159v1.

2. Y. Katznelson and D. Ornstein: *The absolute continuity of the conjugation of certain diffeomorphisms of the circle*. Ergod. Theor. Dyn. Sys., **9**, 681-690, (1989).

3. S. Marmi, P. Moussa, J.-C. Yoccoz: *Linearization of generalized interval exchange maps*. Ann. of Math. 176, 1583-1646, (2012).

4. M. Viana: *Dynamics of interval exchange transformations and Teichmüller flows*. Lecture notes of graduate courses taught at IMPA in 2005 and 2007.

Оценивания параметра показательного распределения в модели пропорциональных интенсивностей

Абдужалилова Г.С.

Метод максимального правдоподобия является одним из самых широко используемых методов оценивания. Оценки полученные по методу максимального правдоподобия, при условиях регулярности обладают хорошими свойствами: асимптотическая нормальность, асимптотическая эффективность. Однако, в некоторых случаях уравнение максимального правдоподобия трудно решается в явном виде. Обычным ответом на такие трудности, является использование альтернативных методов. В данной работе рассмотрим метод характеристических функций. Метод характеристических функции для независимых и одинаково распределенных случайных величин была изучена авторами [2-9]. В случаи цензурированной выборки метод характеристических функций был использован в [1]. Данная работа посвящена оцениванию параметра показательного распределения в модели пропорциональных интенсивностей случайного цензурирования справа.

Рассмотрим одну специальную модель случайного цензурирования справа, называемую моделью пропорциональных интенсивностей(МПИ). Определим МПИ, следуя [2]. Пусть независимые с.в. X и Y имеют ф.р. $F(x;\theta)$ и $G(y;\theta)$. Наблюдению доступна выборка $S^{(n)} = \{((Z_i, \delta_i),\ i = \overline{1,n}\}$, где $Z_i = min\{X_i, Y_i\}$,$\delta_i = I(Z_i = X_i) = I(X_i \leq Y_i)$. Легко видеть, что с.в. Z_i имеет ф.р.

$$H(x;\theta) = 1 - (1 - F(x;\theta))(1 - G(x;\theta)). \tag{1}$$

В МПИ предполагается, что существует не зависящее от θ число $\beta > 0$ такое, что для всех x :

$$1 - G(x;\theta) = (1 - F(x;\theta))^{\beta}.$$

(1) можно записать как

$$1 - H(x;\theta) = (1 - F(x;\theta))^{\beta+1}, \quad \beta > 0,$$

и плотность $H(x;\theta)$

$$h(x;\theta) = (\beta + 1)(1 - F(x;\theta))^{\beta} f(x;\theta). \tag{2}$$

Рассмотрим задачу оценивания параметра θ показательного распределения в МПИ:

$$f(x;\theta) = \begin{cases} \theta e^{-\theta x}, & x > 0, \quad \theta > 0, \\ 0, & x \leq 0, \end{cases}$$

$$h(x;\theta) = (\beta+1)\theta e^{-(\beta+1)x\theta}, \quad \theta > 0, \quad \beta > 0, \quad x > 0.$$

Рассмотрим случаи, когда β известно и неизвестно.

1) β **известно.** Находим весовую функцию через дельта функцию $\delta(t)$

$$\omega_\theta(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\frac{\partial \log h(x;\theta)}{\partial\theta}e^{-itx}dx = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}(\frac{1}{\theta}-(\beta+1)x)e^{-itx}dx =$$

$$= \frac{1}{\theta}\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}e^{-itx}dx - (\beta+1)\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}xe^{-itx}dx = \frac{1}{\theta}\delta(t) + \frac{(\beta+1)}{i}\delta'(t).$$

Составим уравнение $\int\limits_{-\infty}^{+\infty}\omega_\theta(t)(\varphi_n(t)-\varphi_\theta(t))dt = 0$, где

$$\varphi_\theta(t) = M_\theta e^{itZ} = \int\limits_{-\infty}^{+\infty}e^{itx}h(x;\theta)dx = \int\limits_{0}^{+\infty}(\beta+1)\theta e^{-(\beta+1)x\theta}dx =$$

$$= -(\beta+1)\theta\int\limits_{0}^{+\infty}e^{-((\beta+1)\theta-it)x}dx = \frac{(\beta+1)}{-(\beta+1)\theta-it}\,e^{-((\beta+1)\theta-it)x}\Big|_0^{+\infty} = \frac{(\beta+1)\theta}{(\beta+1)\theta-it}$$

-х.ф. показательного распределения с параметром $(\beta+1)\theta$. Решаем уравнение используя свойства дельта функции:

$$\int\limits_{-\infty}^{+\infty}\omega_\theta(t)(\varphi_n(t)-\varphi_\theta(t))dt = \int\limits_{-\infty}^{+\infty}(\frac{1}{\theta}\delta(t)+\frac{(\beta+1)}{i}\delta'(t))\varphi_n(t)dt -$$

$$-\int\limits_{-\infty}^{+\infty}(\frac{1}{\theta}\delta(t)-(\beta+1)\delta'(t))\varphi_\theta(t)dt = \frac{1}{\theta}\int\limits_{-\infty}^{+\infty}\delta(t)\varphi_n(t)dt + \frac{(\beta+1)}{i}\int\limits_{-\infty}^{+\infty}\delta'(t)\varphi_n(t)dt -$$

$$-\frac{1}{\theta}\int\limits_{-\infty}^{+\infty}\delta(t)\frac{(\beta+1)}{(\beta+1)\theta-it}dt - \frac{(\beta+1)}{i}\int\limits_{-\infty}^{+\infty}\delta'(t)\frac{(\beta+1)}{(\beta+1)\theta-it}dt =$$

$$= \frac{1}{\theta}\varphi_n(0) + \frac{(\beta+1)}{i}\varphi'_n(0) - \frac{1}{\theta}\left(\frac{(\beta+1)}{(\beta+1)\theta-it}\right)_{t=0} - \frac{(\beta+1)}{i}\left(\frac{(\beta+1)}{(\beta+1)\theta-it}\right)'_{t=0} =$$

$$= \frac{1}{\theta} + (\beta+1)\overline{z} - \frac{1}{\theta} - \frac{1}{\theta} = 0$$

и получаем оценку $\widehat{\theta} = \frac{1}{(\beta+1)\overline{z}}$, где $\overline{z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

2) β **неизвестно.** В этом случае оцениваем β :

$$P_\theta(Z_i = X_i) = \frac{1}{1+\beta} \sim \overline{\delta} = \frac{1}{n}\sum_{i=1}^{n} \delta_i.$$

Тогда оценка имеет следующий вид: $\widehat{\theta} = \frac{\overline{\delta}}{\overline{z}}$.

Следует отметить (см. [2]), что обе оценки $\frac{1}{(\beta+1)\overline{z}}$ и $\frac{\overline{\delta}}{\overline{z}}$ являются также и оценками методов максимального правдоподобия и моментов.

Теперь исследуем свойства этих оценок. Рассмотрим случай, когда β неизвестно.

Оценка $\widehat{\theta} = \frac{\overline{\delta}}{\overline{z}}$ обладает свойствами:

а) асимптотическая несмещенность;

б) состоятельность;

в) асимптотическая эффективность.

ЛИТЕРАТУРА

1. Абдушукуров А.А., Кадыров О. Методы характеристических функций в задачах статистического оценивания по цензурированным выборкам. Сб. "Статистические методы оценивания и проверки гипотез". Пермский Гос. Ун-т. Пермь. 2001. с. 4-12.

2. Абдушукуров А.А. Статистика неполных наблюдений. Ташкент. Университет. 2009. с.269.

3. Bryant J.L., Paulson A.S. Some comments on characteristic function-based estimators. Sanhya. Ser. A. 1979. V.41. p.109-116.

4. Csörgő S. Limit behavior of the empirical characteristic function. Ann. Probab. 1981. V.9. N.1. p.130-144.

5. Csörgő S. Estimating characteristic functions, under random censorship. Carleton Math. Lect/ Note. 1981. N.30. P. 301-315.

6. Feververger A., McDunnough PH. One the efficiency empirical characteristic function procedures. J. Royal Statist. Soc. Ser. B. 1981. V. 43. N.1. P 20-27.

7. Feververger A., McDunnough PH. One some Fourier methods for inference. J.A.S.A. 1981. V.76. N.374. p.379-387.

8. Feververger A., McDunnough PH. One statistical transform methods and their efficiency. Canadian J. Statist. 1984. V.12. N.4. p. 303-317.

9. Heathcote C.R. The integrated squared error estimation of parameters. Biometrika. 1978. V.64. p.255-264.

Структура времени вселенной событий и множества достижимости

С.С.Саитова

Национальный университет Узбекистана имени Мирзо Улугбека,

Время, на данный момент является основным параметром во вселенной событий. И это же время, оно рассматривалось как некий одномерный прямолинейный параметр. Пытаясь вникнуть в теорию струн квантовой физики, я пришла к следующему заключению: время не одномерно, не линейно, и тем более, не прямолинейно. Своей многомерной структурой время заполняет пространство событий, и каждая его составляющая влечет или идет с тем или иным событием вселенной квантовой физики.

Определив составляющую(ие) того или иного события или череды событий мы сможем получить информацию об этом событии, точнее состоянии вселенной во время этого события. Т.е. череда событий приведшая именно к этому событию и есть составляющие того момента времени, когда событие случилось. Достаточно выбрать интересующее нас событие, и все было до, есть сейчас, и будет после задано в его моменте времени.

Струны квантовой вселенной соединяют те или иные составляющие многомерного времени, и поэтому могут информацию о любом параметре вселенной в любой момент времени.

Событие и его момент времени неотделимы. В нужном месте, в нужное время не два требования, а одно требование определенного события.

К данной гипотезе применимы результаты из работ [1]-[7].

Итак, допустим, что мы находимся в некой исходной точке пространства событий (предполагается, что данное пространство имеет необходимую математическую структуру) и делаем выбор, т.е. допускаем какие-то изменения, которые можно выразить векторными полями. Среди них имеются так называемые векторные поля Киллинга (см. [1]-[7]). Именно они являются объектом моих исследований.

Траектории (см. [1], [2], [4]) заданных векторных полей, это "путь"по которому мы идем, сделав выбор. Всевозможные траектории исходной точки, образуют так называемую орбиту этой точки. Если финальная точка принадлежит орбите, мы достигаем цели. Очевидно, что в этом случае время может быть любое.

С другой стороны, в моих исследованиях выделены также так называемые множества T-достижимости (см. [7]), т.е. некие подмножества орбиты

с заданным "промежутком"времени. И опять, если финальная точка есть точка множества Т-достижимости, то мы достигаем "цель"в заданное "время".

Отдельное исследование множеств Т-достижимости показало, что **в некоторых случаях множество Т-достижимости совпадает со всей орбитой.**

ЛИТЕРАТУРА

1. Нарманов А.Я., Саитова С.С. О геометрии векторных полей Киллинга // Доклады АН РУз. 2013, №5. с.3-5.

2. Саитова С.С. О коммутирующих векторных полях Киллинга //Узб.Мат.Журнал. 2013, №1. с.109-117.

3. Нарманов А.Я., Саитова С.С. Foliations defined by closed differential 1-form // International Journal of Geometry. 2014, vol.3, №1. p.37-43.

4. Нарманов А.Я., Саитова С.С. О геометрии векторных полей Киллинга // Дифференциальные уравнения. 2014, №12. с.1-8.

5. Нарманов А.Я., Саитова С.С. О геометрии множества достижимости векторных полей Киллинга // Доклады АН РУз. 2016, №1. с.7-9.

6. Саитова С.С. О геометрии орбит векторных полей Киллинга // Узб.Мат.Журнал, 2016, №4. с.49-58.

7. Нарманов А.Я., Саитова С.С. О геометрии множества достижимости векторных полей // Дифференциальные уравнения. 2017, №3, том 53. с.321-328.

Носители весовой функции Грина для строго плюрисубгармонического веса

Н.Х.Нарзиллаев
Национальный университет Узбекистана
n.narzillaev@nuu.uz

Пусть $K \subset \mathbb{C}^n$ компактное множество. $\psi(z)$ некоторая ограниченная функция на K. Рассмотрим класс функций

$$L(K,\psi) := \{u(z) \in L,\ u(z)|_K \leq \psi(z)\}$$

где $L = \big\{u(z) \in psh(\mathbb{C}^n) :\ u(z) \leq C_u + \ln^+|z|,\ z \in \mathbb{C}^n\big\}$.

Определение. $V^*(z,K,\psi) = \overline{\lim\limits_{w\to z}} V(w,K,\psi)$ называется функцией Грина с весом $\psi(z)$ компакта K, где $V(z,K,\psi) := \sup\{u(z) : u(z) \in L(K,\psi)\}$, $z \in \mathbb{C}^n$.

В работе А. Садуллаева [1] рассмотрен случай, когда $K \subset \mathbb{C}^n-$ компакт и весовая функция $\psi(z), z \in K$, продолжается на все $\mathbb{C}^n$ как строго плюрисубгармоническая функция из класса L, т.е

$$\exists\tilde{\psi} \in L \cup C^2, dd^c\tilde{\psi} > 0, \tilde{\psi}|_K = \psi. \tag{2}$$

Проблема.(Садуллаев) [1] Верно ли, что $supp\Big(\big(dd^cV^*(z,K,\psi)\big)^n\Big) = K$ для любой строго плюрисубгармонической функции $\psi(z)$?

М.А. Алан привел контрпример этой проблемы и получил следующую теорему, которая отвечает на проблему:

Теорема.[2] Пусть $K-$ компакт в $\mathbb{C}^n$ и $\psi(z) \in L^+$ и строго плюрисубгармоническая функция в окрестности K. Тогда

$$supp\Big(\big(dd^cV^*(z,K,\psi)\big)^n\Big) = K.$$

Мы представляем следующее следствие этой теоремы:

Следствие. Пусть $K-$ компакт в $\mathbb{C}^n$ и $\psi(z)$ строго плюрисубгармоническая функция в окрестности K. Тогда существует число $\alpha > 0$, такое, что

$$supp\Big(\big(dd^cV^*(z,K,\alpha\psi)\big)^n\Big) = K.$$

ЛИТЕРАТУРА

1. Sadullaev A. Pluriregular compacts in $\mathbb{P}^n$. USA: American Journal of Mathematics, Contemporary mathematics, № 662. p. 145–156.

2. M.A. Alan, Weighted regularity and two problems of Sadullaev on weighted regularity, Springer, Complex Analysis and its Synergies 5:8, 2019.

Nurbek Narzillaev (Tashkent, Uzbekistan)

Supports weighted green function for strictly plurisubharmonic weight

Abstract. *We obtain the following corollary of this theorem M.A.Alan [3].*

Key words: *Weighted Green function*

Об одной задаче R - перемешивания для порядковых статистик

Б.Б.Халхаджаев, А.Эшкабилов
Ташкентский институт инженеров железнодорожного транспорта
xalxadjaev@tashiit.uz, sherzod.yusupov@tashiit.uz

Пусть

$$X_1, X_2, ..., X_n$$

независимая выборка объема n, взятая из генеральной совокупности с общей функцией распределения $F(x) = P(\xi < x)$. Пусть

$$\xi_1{}^{(n)} \leq \xi_2{}^{(n)} \leq ... \leq \xi_n{}^{(n)}$$

вариационный ряд, построенный по выборке $X_1, X_2, ..., X_n$. Далее пусть

$$G_{kn}(x) = P(\xi_k{}^{(n)} < x)$$

законом распределения k-го $(k = 1, 2, ..., n)$ члена $\xi_k{}^{(n)}$ этого ряда. Общее условие принадлежности закона $F(x)$ к области $k(n)$ притяжения, удовлетворяющего условиям

$$k = k(n) \to \infty, \frac{k(n)}{n} \to 0, (n \to \infty)$$

нормального закона, можно сформулировать в виде следующего предложения.

Теорема 1. (Н.В.Смирнов [5]) Для того, чтобы при $n \to \infty$, $k(n) = o(n)$ в заданных последовательностях $a_n > 0$ и b_n, $n = 1, 2, ...$ выполнялось соотношение

$$G_{kn}(a_n x + b_n) = P\{\frac{\xi_k{}^{(n)} - b_n}{a_n}\} \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt,$$

необходимо и достаточно, чтобы равномерно по x

$$u_n(x) = \frac{nF(a_n x + b_n) - k(n)}{\sqrt{k(n)}} \to x, \quad n \to \infty.$$

В работе [5] Н.В.Смирнов, несколько модифицируя рассуждения Б.В.Гнеденко ([6]) установил условия, гарантирующие принадлежность области $k(n)$ притяжения к нормальному закону. Обозначим $\frac{k}{n}$ через λ_{kn} .

Теорема 2. (Н.В.Смирнов [5]) Для того, чтобы закон $F(x)$ при условиях

$k(n) \to \infty$, $\lambda_{kn} \to 0$ испытывал $k(n)$ притяжение к нормальному закону, необходимо и достаточно, чтобы при любых x

$$u_n(x) = \frac{F(a_n x + b_n) - \lambda_{kn}}{\lambda_{kn}} \sqrt{k} \to x, \quad n \to \infty,$$

где b_n определяется как наименьшее x , удовлетворяющее условию

$$F(x) = F(x - 0) \leq \lambda_{kn} \leq F(x + 0),$$

а a_n как наименьшее y при данном b_n , удовлетворяющее неравенствам

$$F(y_n + b_n - 0) \leq \lambda_{kn}(1 + \frac{1}{\sqrt{k}}) \leq F(y_n + b_n + 0).$$

Основным результатом является следующая теорема.

Теорема 3. Пусть при $n \to \infty$, $k(n) = o(n)$ и при некоторой последовательности чисел $a_n > 0$ и b_n выполняется условие, что при $n \to \infty$

$$u_n(x) = \frac{F(a_n x + b_n) - \lambda_{kn}}{\lambda_{kn}} \sqrt{k} \to x,$$

тогда для любого события $A \in F$, $P(A) > 0$ при $n \to \infty$

$$\lim_{n \to \infty} P\{\frac{\xi_k{}^{(n)} - b_n}{a_n} < x/A\} = \Phi(x)$$

Литература

1. *Азларов Т.А., Джамирзаев А.А., Мамуров И.Н.* Предельные теоремы для распределении членов вариационного ряда при случайном объеме выборки,Узб. мат.журнал, 1991 г.

2. *Гнеденко Б.В.* Курс теории вероятностей, М.: "Наука 1988 г.

3. *Дейвид Г.* Порядковые статистики, М.:"Наука 1979 г.

4. *Джамирзаев А.А.* О свойстве перемешивание в смысле А.Реньи для статистики критерии однородности, Пермь, Пермский гос унив-т, вып 21, 2008 г., с. 33-37.

5. *Смирнов Н.В.* О сходимости к нормальному закону распределений членов вариационного ряда, Изв. АН УзССР, Сер. Физ.-мат.наук, No 3, 1966 г., с.24-32.

6. *Гнеденко Б.В.* Sur la distribution limite du terme maximum d'une serie aleatoire, Annale of matematies, v. 44, No 3, 1943, p.423-453.

The use of gravitational forces as a parameter of time

[1]**Kadirov R. Kh.,** [2,1]**Nazarov A. I.,** [2,2]**Ibrokhimov B. S.**
[1]Scientific and Practical Center of Dentistry and Oral and Maxillofacial Surgery of the Tashkent State Dental Institute.
[2]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi [1]*k-r-x@mail.ru;*
[2,1]*alisher.nazarov.1958@mail.ru;* [2,2]*b.ibrohimov@tuit.uz*

Annotation. The paper substantiates the use of gravitational forces of the solar system as a parameter of time and highlights the advantages of this approach. On the example of the problem of predicting atmospheric air pressure the procedures increasing the uniformity of the variational data series, which increases the efficiency of model estimates are proposed. These procedures allow the dynamic forecasting problem to be reduced to statistical.

Key words: forecasting methods, influence of gravitational forces, environmental processes, least squares method.

Introduction

When predicting various natural processes, the mathematical apparatus of the theory of differential, integral equations, time series, etc. are actively used. The basis of these theories is the time parameter t, while initially it is assumed that the unit of time is equal, with minutes being added to hours, hours to days, days to weeks, months, years. In fact, this position levels the difference not only between date and time, but also between adjacent time intervals.

Main part

It is known that the functioning of natural systems substantially depends on the time of the study: morning or evening, winter or summer. In this context, we can conclude that the time parameter t:

- levels the difference between time and date;
- describes only the quantitative component of a unit of time and does not reflect its qualitative component.

The aim of the study is to justify the use of the gravitational forces of the celestial bodies of the solar system as a parameter of time and to show the advantages of their use in predicting natural phenomena (atmospheric air pressure).

If in mathematics the time parameter t is used both to describe the time of day ($Time$), as well as to describe the time of year ($Data$), then in astronomy

these factors are clearly separated. In particular, the time of day t is determined by the rotation of the Earth around its axis w, and Data - by the coordinates (x, y) of the location of the center of the Earth O in its orbit L with respect to the Sun (Fig. 1). Since the cut from S to $O(x, y)$ is the distance between the celestial bodies, according to NewtonЎ¯s law it can be replaced by the force F that arises between them.

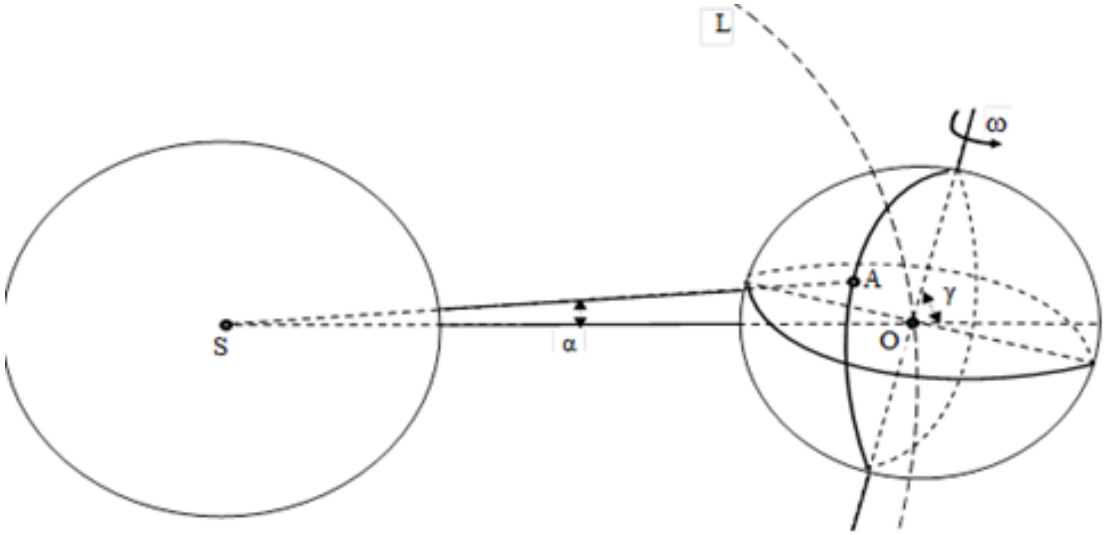


Рис. 12: The motion of the earth around the sun

Since similar interactions take place between the Earth and the Moon, as well as with the planets of the solar system: Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, the time parameter can be represented in the form (1):

$$t = Function(w, F_i) \tag{1}$$

Calculation of gravitational forces can be carried out using tidal potential, developed in the theory of tides [**?**].

In favor of this approach, a number of advantages characteristic of gravitational forces can be attributed. In particular, they are as follows:

- there is a property of mass and it affects any object that has mass;
- fully describe the lunar, terrestrial, solar and other cosmic cycles [**?**];
- allow to take into account the influence of any planet of the solar system on the Earth;
- can be calculated with a high degree of accuracy for any point of the globe and at any time;

- give an unambiguous correspondence for any point of the globe, taking into account the year, season, day, hour

It should be noted that with the help of tidal potential, the prediction of sea tides is carried out quite accurately. Differences in atmospheric air pressure can also be considered as tides of the air ocean, but it was not possible to obtain effective prediction of their dynamics based on tidal potential. This circumstance served as the basis for the statement about the leading role of thermodynamic processes in the dynamics of the atmosphere.

From our point of view, effective forecasts were not obtained due to the above specifics of using the time parameter in mathematics. We show this by the example of solving the problem of forecasting atmospheric air pressure in the city of Tashkent. The source material for the study was the statistical data of the meteorological station.

It is known that the weather depends on incoming solar energy, the circulation of air masses in the atmosphere and the nature of the underlying surface of a given place, on the latitude and elevation of the place above sea level, the features of the terrain and vegetation, the presence of glaciers and snow cover, and the degree of atmospheric pollution. In this case, the rotation of the Earth around its axis and the revolution of the Earth around the Sun determine the daily and annual variations of weather on Earth.

All these factors can be roughly correlated in the following categories:

1. Category "Place of observation". The intensity and nature of atmospheric phenomena depends on the specific characteristics of the geographic observation point $M(i)$.

2. Category "Time". Atmospheric air pressure changes during the day t.

3. Category "Date". Atmospheric air pressure depends on the time of the year $Data$.

4. Category "Sun". The atmospheric pressure is influenced by the intensity β and the incidence angle α of the sunbeam on the earth's surface at the observation point.

5. Category YGravity forcesY. Atmospheric air pressure depends on the mass distribution in the solar system relative to the observation point F_i.

Thus, to obtain effective forecasts on the dynamics of atmospheric air pressure (ADB), it is necessary to construct a regression equation in the form of:

$$ADB = f(M(i), t, Data, \beta, \alpha, F_i) \tag{2}$$

In this case, the mathematical model of the ADB forecast will depend on many factors, which in turn are continuously changing in time and space. This leads to violations of the requirements of the least squares method [?, ?]. To increase the uniformity of the initial data sample when solving the problem of forecasting ADB, the following procedures are proposed.

Factor "Place of observation". If observations recorded at one geographical point are selected from the initial meteorological data array, the resulting data array will be homogeneous with respect to the specific features of the observation site. In this case, we can assume that the observed changes in atmospheric air pressure are due to factors of other categories. Thus, the factor YPlace of observation¥ from the category of variables go into the category of constant values and the original model (2) is converted to:

$$ADB = f(t, Data, \beta, \alpha, F_i) \tag{3}$$

Using this procedure, we increase the uniformity of the initial data array by excluding from consideration characteristics that describe the specific features of the observation site. Moreover, we implicitly agree with the fact that for each geographical point its own predictive model of ADB will be built.

Factor YTime of day¥ and YTime of year¥. As you know, time is determined by the rotation of the Earth around its axis w, and Date is determined by the coordinates (x, y) of the location of the center of the Earth O in its orbit L with respect to the Sun. Moreover, seasonality is described by the angle of inclination of the axis of rotation of the Earth to the ecliptic γ.

The task is complicated by the fact that the parameters w, γ, and $O(x, y)$ vary continuously and simultaneously. This leads to a change in the angle of incidence of the sun's beam α and to a change in the distance of the geographical point A relative to the center of the Sun S (Fig. 1).

Let the position of the Earth relative to the Sun depicted in Fig. 1 corresponds to 12 noon on January 1, 2002. One year corresponds to 1 revolution of the Earth in its orbit relative to the Sun. Hence, it is precisely that this position of the Earth in relation to the Sun will be celebrated every year on January 1 at 12 oŸclock in the afternoon. We form an array of data corresponding to this position by years (Table 14):

A feature of this data set is the fact that the Earth is in one position with respect to the Sun. This means that factors such as the rotation of the Earth

N	ADB	Time (t)	Date	Year
1	723	12	1st January	2002
2	720	12	1st January	2003
3	726	12	1st January	2004
...	...	...	...	...
k	734	12	1st January	2011

Таблица 14: The values of ADB in Tashkent on January 1, 12 noon.

around its axis α, the coordinates of the Earth in orbit with respect to the Sun $O(x, y)$, and the inclination of the axis of rotation of the Earth to the ecliptic plane γ and the angle of incidence of the sun's beam α are constant. Note that in this case, the YYearY parameter, which describes the number of revolutions that the Earth made around the Sun from the moment taken as the initial one, acts as the data sample size. With such data filtering, we find that the parameters $t = const, Data = const, \alpha = const$.

In this case, the original model (3) is converted to the form:

$$ADB = f(\beta, F_i) \tag{4}$$

Here a few words should be said about such a factor as solar activity β. With this method of generating data, the Earth is in one position with respect to the Sun, and this ensures that the angle of incidence of the sun's beam α at the geographic point under study is constant. However, in different years, different intensities of solar radiation can be observed, which have a significant effect on the flow of air masses.

Some researchers believe that this periodicity is due to thermonuclear processes occurring on the sun. Others - it is believed that the change in solar activity is due to the gravitational impact of the planets of the solar system [**?**, **?**]. If we adhere to this point of view, then in our case the influence of solar activity on ADB is already taken into account by the gravitational forces of the planets included as independent variables. Thus, it can be stated that the use of procedure 2 allows the parameters of the categories "Time "Date"and "Sun"to be transferred to the category of constants and, accordingly, the task from form (4) is converted to:

$$ADB = f(F_i) \tag{5}$$

In this case, the change in the ADB values will be due to the gravitational influence of other planets and the Moon on the Earth due to their redistribution in the solar system.

The main advantage of this approach is that the forecast is based on the gravitational forces of the planets, the Sun and the Moon, which can be calculated with high accuracy, since the orbits of celestial bodies (except Pluto) are known.

The proposed procedure allows us to increase the uniformity of the initial data sample, but at the same time forces us to build a predictive model for each point in time and each day of the year separately.

In principle, the amount of computation in constructing mathematical models is not a problem for modern computer technology.

Factor "Gravity forces."At the previous stages, using the procedures Yplace of observation¥, Ytime of day¥ and Ytime of year¥, the initial problem (2) was reduced to the form (5), i.e. to the problem of accounting for the redistribution of the gravitational forces of the solar system.

It is known that tidal potential is the result of the total effect of waves of various lengths. Parameter **H** - determines the type of function. A complete and purely harmonic decomposition of this function into spherical harmonics can be made.

Such decomposition was first made by A.T. Dudson in 1921. Dudson decomposition contains 386 waves. In 1971, D. K. Cartwright published a more complete expansion of the tidal potential, containing 550 waves [**?**, **?**]. Each of the 550 waves can be considered in projections onto the parallel, meridian and vertical in their kinematic characteristics. In total, 44550 characteristics (7 planets, the Moon, the Sun $*$ 3 projections $*$ 3 kinematic characteristics $*$ 550 waves) can be included in the consideration.

The construction of regression models was carried out on the basis of only those characteristics of gravitational waves of the tidal potential that had a reliable correlation with the values of the ADB ($p < 0.05$). Moreover, a large number of independent variables led to a violation of the MNC assumptions about rank. In this regard, gravitational waves were combined into 6 groups depending on the projection (mer, par, wer) and the nature of the correlation connectionv$(r > 0, r < 0) : S1(mer, r > 0), S2(mer, r < 0), S3(par, r > 0), S4(par, r < 0), S5(ver, r > 0), S6(ver, r < 0)$.

The predictive rules for ADB were built in the form of a simple regression from each of these amounts. When constructing simple regression equations, the parameters of the model were imposed with the condition of their efficiency not lower than $p < 0.05$ according to the Student criterion. The arithmetic mean of simple regressions was taken as the final result.

On average for months, the standard deviations of model values from real measurements were 2.7 - 3.0, 0 mm Hg.

The main advantages of the proposed forecasting approach are the following:

- allows the problem of forecasting the dynamics of a parameter of the process under study to be solved by methods of mathematical statistics;
- allows you to make hourly forecasts;
- the absence of extrapolation procedures in this methodology allows one to obtain forecasts whose accuracy does not depend on the forecasting period.

The technique was tested in predicting atmospheric pressure and air temperature for different cities of the world, geomagnetic activity indices for the Moscow observatory [?, ?].

We believe that this approach can be used for a wide range of tasks. In particular, it can be the subject of interest for its effectiveness in predicting the seismic activity of the earth's crust, solar activity, groundwater level, frequency of various diseases, epidemics.

References

1. Cupronickel P. Physics and dynamics of planets. - M.: Mir Publishing House, 1975.

2. Bolshakov V.A. A new concept of the orbital theory of paleoclimate. M .: 2003 256 s. Faculty of Geography, Moscow State University.

3. Kremer N.Sh. Probability Theory and Mathematical Statistics: A Textbook for High Schools. - M., UNITY-DANA, 2001.

4. Mkhitaryan V.S., Troshin L.I., Adamova E.V., Shevchenko K.K., Bambaeva N.Ya. Probability Theory and Mathematical Statistics / Moscow International Institute of Econometrics, Computer Science, Finance and Law. - M., 2002

5. Mushailov B.R., Rykhlova L.V. On the dynamic effects of large planets on the parameters of the solar cycle. Proceedings of GAI. T. 70. 2002.

6. Sh.F. Makhtiev, V.E. Khanin, T.A. Ismail-Zade, E.N. Khalilov YThe regularity of the spatio-temporal distribution of volcanic eruptions¥ / Scientific

discovery No. 239 of 10/15/2003 issued by Mezhdunar. Acad. authors æ scientific discoveries and images., Intern. Ass. authors of scientific discoveries and RANS.

7. Cartwright D.E., Tayler R.I. New Computations of the Tide-generating Poten-tial. Geophys. J. Roy. Astron. Soc., 23, 45-47 (1971)

8. Cartwright D., Edden A. Corrected tables of tidal harmonics. J. Geophys. Res. 33, No. 3, 253-263 (1973).

9. Kadirov R.KH., Nazarov A.I., Ibrohimov B.S. Prediction of Solar Activity on the basis of Redistribution of Masses of the Solar System. International Journal of ADBanced Research in Science, Engineering and Technology. OF IJARSET, ISSN: 2350-0328. Volume 4, Issue 10, 2017 / - P. 486-4692.

10. Kadirov R.H., Saipov A., Nazarov A.I. Modeling of natural processes. IV International conference YIndustrial technolofies and enægineering¥ ICITE - 2017. - T.2. Shimkent. P.276-282



ПРИЛОЖЕНИЯ

APPENDIX

Организация Объединенных Наций A/69/L.72

Генеральная Ассамблея

Distr.: Limited
27 May 2015
Russian
Original: English

Шестьдесят девятая сессия
Пункт 13(a) повестки дня
Комплексное и скоординированное осуществление решений крупных конференций и встреч на высшем уровне Организации Объединенных Наций в экономической, социальной и смежных областях и последующая деятельность в связи с ними

Австралия, Австрия, Албания, Аргентина, Армения, Барбадос, Беларусь, Бельгия, Болгария, Босния и Герцеговина, бывшая югославская Республика Македония, Венгрия, Вьетнам, Гватемала, Германия, Греция, Доминиканская Республика, Израиль, Индия, Ирак, Ирландия, Исландия, Испания, Италия, Казахстан, Катар, Кипр, Китай, Латвия, Литва, Люксембург, Маврикий, Малайзия, Мальта, Марокко, Мексика, Мозамбик, Монголия, Науру, Нидерланды, Норвегия, Палау, Папуа — Новая Гвинея, Польша, Португалия, Республика Корея, Российская Федерация, Румыния, Самоа, Сенегал, Сербия, Сингапур, Словакия, Словения, Соединенное Королевство Великобритании и Северной Ирландии, Сомали, Таджикистан, Таиланд, Турция, Украина, Филиппины, Финляндия, Франция, Хорватия, Черногория, Чешская Республика, Швейцария, Швеция, Эквадор, Эстония, Эфиопия и Япония: проект резолюции

Всемирный день статистики

Генеральная Ассамблея,

ссылаясь на свою резолюцию 64/267 от 3 июня 2010 года, в которой 20 октября 2010 года было объявлено Всемирным днем статистики, проводимым под общим лозунгом «Отметим многочисленные достижения официальной статистики!»,

ссылаясь также на принятое Статистической комиссией на ее сорок пятой сессии решение 45/113[1], в котором Комиссия одобрила предложение о проведении 20 октября 2015 года Всемирного дня статистики, а также доклад Генерального секретаря по вопросу о ходе подготовки к его проведению[2],

[1] *Официальные отчеты Экономического и Социального Совета, 2014 год, Дополнение № 4* (E/2014/24), глава I.B.
[2] E/CN.3/2015/28*.

Просьба отправить на вторичную переработку

вновь подтверждая свои резолюции 53/199 от 15 декабря 1998 года и 61/185 от 20 декабря 2006 года о провозглашении международных годов и резолюцию 1980/67 Экономического и Социального Совета от 25 июля 1980 года о международных годах и годовщинах, в частности пункты 1–10 приложения к ней, касающиеся согласованных критериев провозглашения международных годов, а также пункты 13 и 14, в которых говорится, что международный день или год не следует провозглашать до принятия основных мер, необходимых для его организации и финансирования,

признавая существование официальной статистики на протяжении длительного времени и то содействие, которое Организация Объединенных Наций оказывает в этой области с момента создания в 1947 году Статистической комиссии, которой было поручено способствовать развитию национальных статистических систем и улучшению сопоставимости представляемых ими статистических данных, координировать работу специализированных учреждений в области статистики, развивать центральные статистические службы Секретариата, консультировать органы Организации Объединенных Наций по общим вопросам, касающимся сбора, анализа и распространения статистической информации, и способствовать общему улучшению статистических данных и статистических методов,

признавая основополагающее значение устойчивого национального статистического потенциала для своевременной подготовки надежных статистических данных и показателей, предназначенных для измерения прогресса отдельно взятой страны,

учитывая исключительно важное значение качественной официальной статистической информации для целей анализа и принятия обоснованных решений директивного характера в поддержку усилий по обеспечению устойчивого развития,

ссылаясь на резолюцию 2006/6 Экономического и Социального Совета от 24 июля 2006 года под названием «Укрепление статистического потенциала», в которой Совет призвал государства-члены, систему Организации Объединенных Наций, международные учреждения, страны-доноры и организации-доноры поддерживать национальные усилия по созданию и укреплению национального потенциала в области статистики, в частности в развивающихся странах,

ссылаясь также на свою резолюцию 68/261 от 29 января 2014 года, в которой Генеральная Ассамблея одобрила Основополагающие принципы официальной статистики[3], принятие которых стало важной вехой в кодификации и поощрении основных ценностей в области официальной статистики,

отмечая, что первый Всемирный день статистики, отмечавшийся 20 октября 2010 года, прошел с огромным успехом — соответствующие мероприятия были организованы в более чем 130 государствах-членах и по меньшей мере 40 международными и региональными организациями и структурами,

[3] *Официальные отчеты Экономического и Социального Совета, 1994 год, Дополнение № 9* (E/1994/29), пункт 59.

отмечая также, что в 2015 году отмечается 200-летие со дня рождения Джорджа Буля, чья работа по вопросам применения принципов логики в качестве одной из форм алгебры, носящей в настоящее время его имя, лежит в основе современных компьютерных наук,

отмечая далее, что некоторые государства-члены и регионы уже по достоинству оценивают роль официальной статистики и осуществляют различные национальные и региональные инициативы, такие как проведение месячников, недель и дней статистики, а также специальных мероприятий, и приветствуя предоставляемую ими поддержку и их готовность к обеспечению координации таких мероприятий, проводимых под эгидой Организации Объединенных Наций,

1. *постановляет* объявить 20 октября 2015 года вторым Всемирным днем статистики под общим лозунгом «Повышение качества данных, повышение качества жизни»;

2. *постановляет также* отмечать Всемирный день статистики 20 октября каждые пять лет;

3. *предлагает* всем государствам-членам, членам специализированных учреждений, организациям системы Организации Объединенных Наций и другим международным и региональным организациям, а также представителям гражданского общества, включая неправительственные организации, такие как научно-исследовательские институты, средства массовой информации и все производители и пользователи официальной статистики, надлежащим образом отмечать Всемирный день статистики;

4. *подчеркивает*, что расходы на все мероприятия, которые могут быть проведены в связи с осуществлением настоящей резолюции, должны покрываться за счет добровольных взносов;

5. *просит* Генерального секретаря принять необходимые меры для проведения Организацией Объединенных Наций в 2015 году Всемирного дня статистики и довести настоящую резолюцию до сведения всех государств-членов и членов специализированных учреждений, а также организаций системы Организации Объединенных Наций.

United Nations A/69/L.72

General Assembly

Distr.: Limited
27 May 2015

Original: English

Sixty-ninth session
Agenda item 13 (a)
Integrated and coordinated implementation of and follow-up to the outcomes of the major United Nations conferences and summits in the economic, social and related fields

Albania, Argentina, Armenia, Australia, Austria, Barbados, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, China, Croatia, Cyprus, Czech Republic, Dominican Republic, Ecuador, Estonia, Ethiopia, Finland, France, Germany, Greece, Guatemala, Hungary, Iceland, India, Iraq, Ireland, Israel, Italy, Japan, Kazakhstan, Latvia, Lithuania, Luxembourg, Malaysia, Malta, Mauritius, Mexico, Mongolia, Montenegro, Morocco, Mozambique, Nauru, Netherlands, Norway, Palau, Papua New Guinea, Philippines, Poland, Portugal, Qatar, Republic of Korea, Romania, Russian Federation, Samoa, Senegal, Serbia, Singapore, Slovakia, Slovenia, Somalia, Spain, Sweden, Switzerland, Tajikistan, Thailand, the former Yugoslav Republic of Macedonia, Turkey, Ukraine, United Kingdom of Great Britain and Northern Ireland and Viet Nam: draft resolution

World Statistics Day

The General Assembly,

Recalling its resolution 64/267 of 3 June 2010, in which 20 October 2010 was designated as World Statistics Day under the general theme "Celebrating the many achievements of official statistics",

Recalling also decision 45/113 of the Statistical Commission at its forty-fifth session,[1] in which the Commission endorsed the proposal to celebrate 20 October 2015 as World Statistics Day, and the report of the Secretary-General to the Commission on the preparations,[2]

Reaffirming its resolutions 53/199 of 15 December 1998 and 61/185 of 20 December 2006 on the proclamation of international years, and Economic and Social Council resolution 1980/67 of 25 July 1980 on international years and anniversaries, particularly paragraphs 1 to 10 of its annex on the agreed criteria for their proclamation, as well as paragraphs 13 and 14, stating that an international day

[1] *Official Records of the Economic and Social Council, 2014, Supplement No. 4* (E/2014/24), chap. I.B.
[2] E/CN.3/2015/28.

Please recycle

or year should not be proclaimed before the basic arrangements for its organization and financing have been made,

Recognizing the long history of official statistics and the facilitating role played by the United Nations since the creation in 1947 of the Statistical Commission, which was tasked with promoting the development of national statistics and improving their comparability, coordinating the statistical work of specialized agencies, developing central statistical services of the Secretariat, advising the organs of the United Nations on general questions relating to the collection, analysis and dissemination of statistical information and promoting the improvement of statistics and statistical methods generally,

Acknowledging the fundamental importance of sustainable national statistical capacity to produce reliable and timely statistics and indicators measuring a country's progress,

Bearing in mind the critical role of high-quality official statistical information in analysis and informed policy decision-making in support of sustainable development,

Recalling Economic and Social Council resolution 2006/6 of 24 July 2006, entitled "Strengthening statistical capacity", in which the Council called upon Member States, the United Nations system, international agencies, donor countries and organizations to support national efforts in building and strengthening national statistical capacity, in particular of developing countries,

Recalling also its resolution 68/261 of 29 January 2014, in which the General Assembly endorsed the Fundamental Principles of Official Statistics[3] as a milestone in the codification and promotion of basic values in official statistics,

Noting that the first World Statistics Day, held on 20 October 2010, was an overwhelming success, with activities being organized in more than 130 Member States and by at least 40 international and regional organizations and entities,

Noting also that 2015 marks the bicentenary of the birth of George Boole, whose work on the application of the principles of logic as a form of algebra underpins modern computer science and which form of algebra today bears his name,

Noting further that some Member States and regions already celebrate official statistics through a variety of national and regional initiatives, such as designated months, weeks, days or events, and welcoming their support and willingness to coordinate such events under the auspices of the United Nations,

1. *Decides* to designate 20 October 2015 as the second World Statistics Day under the general theme "Better data, better lives";

2. *Also decides* to celebrate World Statistics Day every five years on 20 October;

3. *Invites* all Member States, members of the specialized agencies, organizations of the United Nations system and other international and regional organizations, as well as representatives of civil society, including non-governmental organizations, such as research institutions, media and all producers and users of official statistics, to observe World Statistics Day in an appropriate manner;

[3] *Official Records of the Economic and Social Council, 1994, Supplement No. 9* (E/1994/29), para. 59.

4. *Stresses* that the cost of all activities that may arise from the implementation of the present resolution should be met from voluntary contributions;

5. *Requests* the Secretary-General to take the necessary measures for the observance by the United Nations of World Statistics Day in 2015 and to bring the present resolution to the attention of all Member States and members of the specialized agencies and United Nations organizations.

THE WORLD OF STATISTICS

RECOGNIZING THE CONTRIBUTIONS OF STATISTICS TO SOCIETY WORLDWIDE



www.worldofstatistics.org



МОНУМЕНТ:
"СТАТИСТИКА-ЗНАНИЕ-СИЛА"
г. Омск (Россия)
MONUMENT:
"STATISTICS-KNOWLEDGE-POWER"
Omsk (Russia)

РАЗМЫШЛЕНИЯ О МАТЕМАТИКЕ СЛУЧАЯ
(Прим.отв.ред.)

В настоящее время основы Теории Вероятностей и Математической Статистики изучаются во всех ВУЗах и на всех специальностях, включая и гуманитарные, и современного специалиста, в какой бы области он ни работал, уже невозможно себе представить без соответствующей вероятностной подготовки. Вот как образно писал о пользе изучения Теории Вероятностей и Математической Статистики замечательный венгерский математик и педагог Альфред Реньи (1921-1969):

"Изучение теории вероятностей благоприятно сказывается на характере учащихся, например, развивает смелость, поскольку позволяет понять, что при определенных обстоятельствах неудачи можно просто отнести к случайностям и, следовательно потерпев неудачу, отнюдь не следует отказываться от борьбы за достижение намеченной цели. Люди, стоящие на низкой ступени развития, склонны к чрезмерной подозрительности: какая бы беда ни приключилась с ними, они склонны приписывать ее чьему - то злому умыслу, даже если подобные утверждения лишены малейших оснований. Объясняется это тем, что примитивные люди не знакомы с таким понятием, как случайность. Преподавание теории вероятностей может принести несомненную пользу, поскольку позволяет окончательно порвать с пережитками магического мышления каменного века. Изучая теорию вероятностей, люди становятся более снисходительными и терпимыми к окружающим и, следовательно, с большей легкостью вписываются в жизнь общества".

СЛУЧАЙНОСТЬ И ЗАКОНОМЕРНОСТЬ
(Стихи проф. Г.И.Ивченко)

Случайно мы встречаемся на жизненном пути,
Случайностью решается, куда и с кем идти.
И то, что представляется перстом закономерности,
В реальности является лишь играми случайности.
Как часто убеждаемся мы в той закономерности,
Что то, в чем мы нуждаемся, подвластно ей-случайности,
И нет необычайности в наличии дилемм,
Поскольку путь случайности не вычислен никем.
И где мы все окажемся - Творцу лишь это ведомо:
Что нам случайным кажется, все знает Он заведомо,

Для нас же - случай "править бал", недаром он - Величество,
В его руках "девятый вал", кто жертвы, их количество.
"Несчастный случай" часто нам проблему "быть-не быть"решает,
"Счастливый случай" - в светлый храм нам щедро двери раскрывает.
Ряд случая явлений - вот нашей жизни вехи,
Он - импульс обновлений, он дарит нам успехи.
Но пусть бесспорен тезис тот, что все определяет случай,
Пассивно человек не ждет: ведь разум дан ему могучий,
Стремится снизить он урон от случай последствий.
Проникнуть в суть, познать закон и связь причин и следствий.
И открывается уму закономерность случая,
И уж не столь темна ему действительность кипучая,
Приоткрываются пред ним Природы тайн покровы,
Ее законы - света гимн - служить ему готовы,
И хаос - ужас всех богов - пред ним снимает маску,
И дух, свободный от оков, в быль претворяет сказку.
Так что же - случай иль закон в основе всех явлений?
И есть ли вообще резон таких сопоставлений?
Ведь просвещенному уму нам мир есть море Случая,
И чтобы плавать по нему, должны быть все обучены,
Лишь Случая постигнув суть, его закономерности,
Мы свой прокладываем путь, страхуясь от случайности.
И нам оружие дает в борьбе с плешивым Случаем
Наука, что вперед ведет и знает все о случае.

МАТЕМАТИЧЕСКИЕ АФОРИЗМЫ

- Нельзя быть настоящим математиком, не будучи немного поэтом. (**Карл Вейерштрасс**)

- Математика-это наука о хитроумных операциях, производимых по специально разработанным правилам над специально придуманными понятиями. Ясно, что особенно важная роль при этом отводится придумыванию новых понятий. Запас интересных теорем в математике быстро исчерпался бы, если бы их приходилось формулировать лишь при помощи понятий, содержащихся в аксиомах. (**Юджин Пол Вигнер**)

- Никакой достоверности нет в науках там, где нельзя приложить ни одной из математических наук, и в том, что не имеет связи с математикой (**Леонардо Да Винчи**)

- Если мы действительно что-то, знаем, то это мы знаем благодаря изучению математики. (**Пьер Гассенди**)

- Математики похожи на французов: что бы вы ни сказали, они все переведут на собственный язык. Получится нечто противоположное. (**Иоганн Вольфганг Гете**)

- Математика является учением об отношениях между формулами, лишенными какого бы то н было содержания. (**Давид Гильберт**)

- Если теорему так и не смогли доказать, она остановится аксиомой. (**Евклид**)

- В каждой естественной науке заключено столько истины, сколько в ней математики. (**Иммануил Кант**)

- Доказательство называется строгим, если таковым его считает большинство математиков. (**Моррис Клайн**)

- Всякий знает, что такое кривая, пока не выучится математике, что вконец запутается в бесконечных исключениях. (**Феликс Клейн**)

- Математические науки, естественные науки и гуманитарные науки могут быть названы, соответственно, науками сверхъестественными, естественными и неестественными. (**Лев Давидович Ландау**)

- Мнимые числа-это прекрасное и чудесное убежище божественного духа, почти что сочетание бытия с небытием. (**Готфрид Вильгельм Лейбниц**)

- Женщине -католичке уже позволено избегать беременности при помощи математики, но строго-настрого запрещено прибегать к химии или физике. (**Генри Луис Менкен**)

- Легче найти квадратуру круга, чем перехитрить математика. (**Огастес де Морган**)

- Математическая истина, независимо от того, в Париже или в Тулузе, одна и та же. (**Блез Паскаль**)

- В математике нет символов для неясных мыслей. (**Анри Пуанкаре**)

- Чистая математика- это такой предмет, где мы не знаем, о чем мы говорим, и не знаем, истинно ли то, что мы говорим. (**Бертран Рассел**)

- Если бы я только имел теоремы! Тогда я бы мог бы достаточно легко найти доказательства. (**Бернхард Риман**)

- Подобно тому, как все искусства тяготеют к музыке, все науки стремятся к математике. (**Джордж Сантаяна**)

- Математика может открыть определенную последовательность даже в хаосе. (**Гертруда Стайн**)

- Математики похожи на влюбленных - достаточно согласиться с простейшим утверждением математика, как он выведет следствие, с которым вновь прийдется согласится, а из этого следствия - еще одно. (**Бернард Ле Бовье**)

- Разве вы не математик? Нет. Тогда мне не о чем с вами говорить Я разговариваю лишь с теми, кто владеет методом математического анализа. (**Анатоль Франс**)

- Математика похожа на мельницу: если вы засыпаете в нее зерна пшеницы, то получите муку, если же засыпаете отруби, отруби и получите. (**Андру Филлинг Хакси**)

- Из дома реальности легко забрести в лес математики, но лишь немногие способны вернуться обратно. (**Хуго Штейнгаус**)

- Легкость математики основана на возможности чисто логического ее построения, трудность, отпугивающая многих, - на невозможности иного изложения. (**Хуго Штейнгаус**)

- Между духом и материей посредничает математика. (**Хуго Штейнгаус**)

- Законы математики, имеющие какое - либо отношение к реальному миру, ненадежны; а надежные математические законы не имеют отношения к реальному миру. (**Алберт Эйнштейн**)

- Математик уже кое-что может, но, разумеется, не то, что от него хотят получить в данный момент. (**Альберт Эйнштейн**)

- Математика - это единственный метод водить самого себя за нос. (**Альберт Эйнштейн**)

- С тех пор как за теорию относительности принялись математики, я ее уже сам больше не понимаю. (**Альберт Эйнштейн**)

- Существует поразительная возможность овладеть предметом математически, не поняв существа дела. (**Альберт Эйнштейн**)

- Я с дрожью ужаса отворачиваюсь от ваших несчастных проклятых функций, у которых нет производных. (**Шарль Эрмит**)

СОДЕРЖАНИЕ

А.А. Абдушукуров О работах В.И. Романовского в области математической статистики (к 140-летию В.И. Романовского) 14

B.L.S. PRAKASA RAO Cramer-Rao Inequality Revisited for Randomly Censored Data 19

Somnath Datta, Chathura Siriwardhana, K.B. Kulasekera Some Recent Statistical Approaches to Personalized Medicine..... 29

Milovan Krnjajić Machine Learning Methods and Time Series Forecasting 36

А.И.Орлов Статистика нечисловых данных как часть современной математической статистики..... 43

F.A.Abdushukurov, A.N.Chuprunov On a Probability of the Event: Each Cell Contains Even Number of Particles 54

P.A.Koldanov, A.P.Koldanov On the distribution of sign test for symmetry..... 67

Д.В.Семенова Случайно-множественная интерпретация мер значимости ассоциативных правил 74

Sh.Mirakhmedov The Probabilities of Large Deviations Associated with Multinomial Distribution 83

A.A.Abdushukurov Statistical Inferences with Competing Risks..... 88

М.У.Гафуров Оценка среднего значения числа выходов случайного блуждания за подвижную линейную границу 96

Т.М.Зупаров Центральная предельная теорема для линейных процессов, порожденных последовательностью независимых случайных величин 100

А.Садуллаев, А.М.Кытманов Об оценках объема нулей голоморфной функции, зависящей от комплексного параметра 103

М.И.Гаман, С.П.Моисеева RQ-система M|M|1 с отрицательными заявками 107

М.А.Шкленник, А.Н.Моисеев Поток повторных обращений в двухфазных системах массового обслуживания с неограниченным числом приборов при нестационарном режиме работы 113

А.А.Бояркина Исследование бесконечнолинейных СМО с групповым обслуживанием 120

Н.В.Кононова, Е.С.Мангалова, О.В.Чубарова Прикладные вопросы классификации с использованием гребневой регрессии 128

В.С.Дансорунова, Н.А.Лукьянова Корреляционно-регрессионный анализ показателей заболеваемости и смертности населения красноярского края по причине социально значимых заболеваний 138

А.И.Ионкина, Е.Е.Голденок Нечеткая модель для классификации степени тяжести панкреатита 146

А.А.Миронова, В.О.Кобаненко, К.М.Параскевопуло, Р.Б.Курбанисмаилов Компонентный анализ смертности населения для оценки динамики ожидаемой продолжительности жизни 150

О.Д.Мельникова, Н.А.Лукьянова Вычисление отношения шансов для многомерных статистических данных 157

Д.И.Одарич, Н.А.Лукьянова, М.Г.Семенова Статистический анализ заболеваемости вирусным гепатитом в Красноярском крае 164

Е.А.Савкина, Н.А.Лукьянова Анализ влияния территориального признака на заболеваемость и ожидаемую продолжительность жизни населения Красноярского края по причине болезней системы кровообращения 172

А.А. Солдатенко Полиномиальный алгоритм поиска приближенного решения в графе с ограничением 178

F.A.Abdushukurov, A.N.Chuprunov On a number of empty cells in allocation scheme with even number of particles in each cell 184

Sh.A.Ayupov, B.B.Yusupov 2-local derivations on Virasora algebra 188

R.D.Aloev, M.M.Nurullaev, A.A.Abdiraximov Software, algorithms and methods of data encryption based on national standards 192

A.Dzhalilov, J.Karimov Hitting Times and Singular Distributions of Circle Maps with a Break 201

А.В.Медведев, Д.А.Мелех, Н.А.Сергеева, О.В.Чубарова Исследование адаптивного алгоритма классификации объектов по данным с пропусками 205

С.А.Ахмедов, К.С. Аблазова Об управлении процессом, описывающим двумерным нормальным распределением при помощи контрольных карт 211

M.R.Eshimbetov, G.Khudayberganov, Z.A.Sobirov The Fokas' method for inverse source problem for heat equation on simple metric graph 213

В.Р.Ходжибаев Об одном случайном процессе с переключениями ... 215

T.K.Kurbanbaev Infinite dimensional Leibniz algebras constructed by $\mathfrak{sl}_2$ Lie algebras 218

Р.Т.Турсунов, А.А.Алимов Моделирования рынка страховых услуг 223

Р.И.Ибрагимов, А.Машраббоев Вероятность попадания в нуль ветвящихся процессов с иммиграцией с τ типами частиц 228

А.Машраббоев, Р.Полванов, А.Рахманов Предельная теорема для ветвящегося процесса Беллмана-Харриса с иммиграцией 230

Г.Г.Рахимова Асимптотические задачи последовательного точечного оценивания функционалов от неизвестной функции распределения 232

Ф.А.Абдикаликов, П.Б.Насыров, Р.П.Сейтахов Асимптотика степенных оценок для условных функций распределения при случайном цензурировании справа 238

Ф.А.Абдикаликов, И.Ж.Турсынбекова, Л.Ж.Реимова Свойства оценок условных функций распределения при случайном цензурировании наблюдений с двух сторон 246

S.B.Kozimov On kernel estimates of distribution function and density in proportional hazard model 254

Л.Р.Махмутова Анализ данных типа времени жизни и сравнение непараметрических оценок функции распределения в модели случайного цензурирования справа 258

К.С.Сагидуллаев Класс оценок многомерной плотности с оцениваемым параметром 267

A.A.Abdushukurov, R.S.Muradov On estimation of bivariate survival function under right random censorship 270

A.M.Rahmatullaev Potts model on a Cayley Tree in the Presence of Three Competing Interactions 276

С.М.Ражабов $m-$Выпуклые функции 279

Н.У.Утеулиев, Б.Н.Бегилов, Г.М.Джайков Моделирование задачи комплекса очистки загрязненных вод 282

B.B.Muratova Wavelet estimation of distribution function in Koziol-Green model of random censorship 287

Н.С.Нурмухамедова Локальная асимптотическая нормальность неоднородных статистических экспериментов в модели конкурирующих рисков при гибридном случайном цензурировании справа 290

L.R.Kakadjanova On Empirical Process of Independence with Estimated Parameter and its Application to Testing of Proportional Hazards Model .. 295

A.A.Abdushukurov, N.T.Dushatov Estimation of Conditional Distribution and Mean Residual Life Function under Dependent Right Random Censored Data 304

Р.Ибрагимов, А.Машраббоев, Ю.Ташмирзаев Статистический анализ основных показателей сельского хозяйства 313

М.М.Рахматуллаев, М.А.Расулова $G_k^{(2)}-$ Периодические основные состояния для модели ПОТТС-SOS на дереве Кэли 318

M.K. Holmurodov, R. Mamadjanov Convergence rates in the law of large numbers for random variables with multidimensional indices 322

Ф.М. Холмуродов, С.А.Болтабоев Математические моделирование показателей иммунной защиты организма, свертывающей системы крови и липидного обмена в реабилитации больных церебральным атеросклерозом 328

Г.Худайберганов, Ж.Ш.Абдуллаев Об одной реализации шара Ли 330

А.К.Аманов, Г.М.Рахимов О преподавании элементов теории вероятности в академических лицеях, колледжах и школах 334

У.Ш.Якубова, Н.Т.Парпиева Некоторые особенности кредитной системы обучения 336

B.B.Elov, M.H.Primova The scheme of business processes LMS NUUz. 341

Я.М.Хусанбаев, Х.Э.Кудратов Некоторые неравенство для моментов ветвящихся случайных процессов с иммиграцией 354

Ш.А.Исматуллаев, Ш.Ш.Шарахметов Аппроксимации функционалов некоторых статистик - мер неравенства 356

К.К.Кулдашев $\Psi-$ Гармоническая мера 363

Sh.Ibragimov Maximality Criteria for Subharmonic and Plurisubharmonic functions 365

U.A. Safarov Singular probability measure of circle homeomorphisms with several critical points 367

А.Л.Абдулвахидов, Д.Г.Захидов Двувыборочные критерии согласия при случайном цензурировании справа 369

Ф.М.Холмуродов Полупараметрическая квантильная оценка процентной остаточной продолжительности жизни в информативной модели конкурирующих рисков 372

A. Begmatov On a martingale generated by interval maps 376

Абдужалилова Г.С. Оценивания параметра показательного распределения в модели пропорциональных интенсивностей 379

С.С.Саитова Структура времени вселенной событий и множества достижимости 383

Н.Х.Нарзиллаев Носители весовой функции Грина для строго плюрисубгармонического веса 385

Б.Б.Халхаджаев, А.Эшкабилов Об одной задаче R - перемешивания для порядковых статистик..... 387

Kadirov R.Kh., Nazarov A.I., Ibrokhimov B.S. The use of gravitational forces as a parameter of time..... 389