BOOK OF ABSTRACTS

St. Petersburg | 2018

ICAS CONFERENCE
4th INTERNATIONAL CONFERENCE ON
ADVANCES IN STATISTICS

4th International Conference on Advances in Statistics

St. Petersburg | 2018

ICAS CONFERENCE

4th INTERNATIONAL CONFERENCE ON ADVANCES IN STATISTICS

MAY 11-13 2018
Original Sokos Hotel Olympia Garden – St Petersburg/Russia
http://www.icasconference.com/

ICAS'2018

4th International Conference on Advances in Statistics

St Petersburg/Russia

Published by the ICAS Secretariat

Editors:
Prof. Dr. İsmihan Bayramoğlu

Conference organised in collaboration with Smolny Institute of the Russian Academy of Education

SCIENTIFIC COMMITTEE

Prof. Dr. Barry C. ARNOLD
University of California, Riverside – USA

Prof. Dr. Gülay BAŞARIR
Mimar Sinan Fine Arts University – Turkey

Prof. Dr. İsmihan BAYRAMOGLU (BAIRAMOV)
Izmir University of Economics – Turkey

Prof. Dr. Narayanaswamy BALAKRISHNAN
Keynote Speaker / McMaster University – Canada

Prof. Dr. Hamparsum BOZDOGAN
The University of Tennessee – USA

Prof. Dr. Şahamet BULBUL
Marmara University – Turkey

Prof. Dr. Aydın ERAR
Mimar Sinan Fine Arts University – Turkey

Prof. Dr. Leda MINKOVA
Department of Probability, Operations Research and Statistics
University of Sofia "St. Kliment Ohridski

Prof. Dr. Jorge NAVARRO
Facultad de Matematicas, Universidad de Murcia – Spain

Prof Dr Sarjinder Singh
Texas A&M University-Kingsville – USA

Prof. Dr. Müjgan TEZ
Marmara University – Turkey

Prof. Dr. Nikolai KOLEV
Department of Statistics, University of Sao Paulo

Prof. Dr. İ. Esen YILDIRIM
Marmara University – Turkey

Assoc. Prof. Dr. Barıs ASIKGIL
Mimar Sinan Fine Arts University – Turkey

Assoc. Prof. Dr. Gulhayat GÖLBAŞI ŞİMŞEK
Yıldız Technical University – Turkey

Assoc. Prof. Dr. Fatma NOYAN TEKELİ
Yıldız Technical University – Turkey

Assoc. Prof. Dr. Esra Akdeniz DURAN
Istanbul Medeniyet University – Turkey

Dr. Ilham AKHUNDOV
Faculty of Mathematics University of Waterloo – Canada

ORGANIZATION COMMITTEE

Prof. Dr. İsmihan BAYRAMOGLU (BAIRAMOV)
Izmir University of Economics – Turkey
Conference Chair

Prof. Dr. Hamparsum BOZDOGAN
The University of Tennessee – USA

Assoc. Prof. Dr. Gulhayat GOLBASI SIMSEK
Yıldız Technical University – Turkey

Assoc Prof. Dr. Barıs ASIKGIL
Mimar Sinan Fine Arts University – Turkey

Assoc. Prof. Dr. Fatma NOYAN TEKELI
Yıldız Technical University – Turkey

Assist. Prof. Dr. Ibrahim GENC
Istanbul Medeniyet University – Turkey

Assist. Prof. Dr. Gulder KEMALBAY
Yıldız Technical University – Turkey

Instructor PhD Ozlem BERRAK KORKMAZOGLU
Yıldız Technical University – Turkey

Dear Colleagues,

On behalf of the Organizing Committee, I am pleased to invite you to participate in *4th INTERNATIONAL CONFERENCE ON ADVANCES IN STATISTICS* which will be held in St.Petersburg, Russia dates between 11-13 May, 2018 .

We cordially invite prospective authors to submit their original papers to ICAS-2017, Helsinki.

- Applied Statistics
- Banking, Finance, Insurance, Actuarial Sciences and Risk Management
- Bayesian Statistics
- Bioinformatics
- Biostatistics
- Clinical Trials
- Combinatorics
- Computational Statistics
- Data Analysis and Modeling
- Data Envelopment Analysis
- Data Management and Decision Support Systems
- Data Mining
- Demography
- Experimental Design
- Energy and Statistics
- Entrepreneurship
- Entropy

- Environmental and Statistics
- Fuzzy Theory and Statistical Applications
- Genetic Algorithms
- Mathematical Foundations of Statistics
- Mathematical Statistics
- Multivariate Statistics
- Neural Networks and Statistics

- Non-parametric Statistics
- Operations Research
- Optimization Methods in Statistics
- Order Statistics
- Panel Data Modelling and Analysis
- Performance Analysis in Administrative Process
- Philosophy of Statistics

- Public Opinion and Market Research
- Quality Control
- Regression Models
- Reliability Theory
- Sampling Theory
- Simulation Techniques
- Spatial Analysis
- Statistical Software
- Statistical Training
- Statistics Education
- Statistics in Social Sciences
- Stochastic Processes
- Supply Chain
- Survey Research Methodology
- Survival Analysis
- Time Series
- Water and Statistics
- Other Statistical Methods

Selected papers will be published in *Communications in Statistics-Theory and Method*, indexed by SCI-Expanded.

We hope that the conference will provide opportunities for participants to exchange and discuss new ideas and establish research relations for future scientific collaborations.

In addition to scientific program there will be also social activities including sightseeing which we hope will leave a pleasant trace on your memory.

Conference Website : http://icasconference.com
E Mail: aioc.statistics@gmail.com

On behalf of Organizing Committee:
Conference Chair
Prof. Dr. Ismihan BAYRAMOGLU,
Izmir University of Economics

10 MAY 2018 THURSDAY
18:30 – 21:00 : REGISTRATION

11 MAY 2018 FRIDAY

08:30 - 17:00 : REGISTRATION
MAIN HALL : GRAND OPENING CEREMONY
09:00 – 09:30

09:30 – 09:40	B R E A K

HALL 1 / WELCOME SPEECH

09:40 – 10:00	**PROF. DR. ISMIHAN BAYRAMOĞLU** *Conference Chair* **Department of Mathematics, Izmir University of Economics**

HALL 1 / KEYNOTE SPEAKER

10:00 – 10:40	**PROF. DR. NADEZHDA GRIBKOVA** **Mathematics and Mechanics Faculty, St.Petersburg State University**
Speech Title	Second order asymptotics for intermediate trimmed sums and L-statistics

10:40 – 11:00	C O F F E E / T E A B R E A K

HALL 1 / SESSION A

SESSION CHAIR	PROF. DR. SELAHATTIN KACIRANLAR	
TIME	**PAPER TITLE**	**PRESENTER** / CO AUTHOR
11:00 – 11:20	**Testing Performance Of Hybrid Time Series Models on Hourly Electricity Price**	<u>Büşra TAŞ</u> , Ceylan YOZGATLIGİL
11:20 – 11:40	**Systemically Important Banks of Turkey by Using Quantile Regression: A Conditional Value at Risk (CoVaR) Approach**	<u>Zehra CİVAN</u>, Gülhayat GÖLBAŞI ŞİMŞEK, Ebru ÇAĞLAYAN AKAY
11:40 – 12:00	**A Bayesian Quantile Time Series Model for Asset Returns**	<u>Gelly MITRODIMA</u> , Jim GRIFFIN
12:00 – 12:20	**Homothetic Transformation's Influence on Excess of D-optimal Designs**	Yuri D. GRIGORIEV, Viatcheslav B. MELAS, <u>**Petr V. SHPILEV**</u>

12:20 – 12:40	**BOUNDS IN COMBINATORIAL CENTRAL LIMIT THEOREM**	**Andrei FROLOV**
12:40 – 13:00	**DISCUSSION**	

13:00 – 14:00	**LUNCH**

HALL 1 / SESSION B

SESSION CHAIR	PROF. DR. NADEZHDA GRIBKOVA	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
14:00 – 14:20	A New Modification of Probability Paradox	**Jan NOVOTNÝ**, Jindriska SVOBODOVÁ
14:20 – 14:40	Fibonacci Sequences of Random Variables	**Ismihan BAYRAMOGLU**
14:40 – 15:00	On Some Problems of the Optimal Choice of Record Values	**Igor V. BELKOV**, Valery B. NEVZOROV
15:00 – 15:20	The Joint Distribution of Marginal Records in Extended Bivariate Random Sequences	**Gülder KEMALBAY**
15:20 – 15:40	Modeling Proportions– Simulation and Empirical Analysis	**Janne ENGBLOM**, Heli MARJANEN
15:40 – 16:00	**DISCUSSION**	

16:00 – 16:10	**BREAK**

HALL 1 / WORKSHOP I

16:10 – 16:50	**PROF. DR. ISMIHAN BAYRAMOĞLU** **Department of Mathematics, Izmir University of Economics**
Speech Title	Dependency and ageing in reliability and survival analysis

17:00 –17:45	**LIVE CONCERT by CONFERENCE PARTICIPANTS**

17:45 –19:30	**HOTEL DEPARTURE FOR BOAT TOUR (Incl into Registration Fee)**

12 MAY 2018 SATURDAY

08:30   -   17:00   : REGISTRATION

HALL 1 / WORKSHOP II

09:00 – 09:40	PROF. DR. SELAHATTIN KACIRANLAR Department of Statistics, Çukurova University
Speech Title	"Investigation Of Risk Performances Of The New Heterogeneous Estimators"

HALL 1 / SESSION C

SESSION CHAIR	PROF.DR. GÜLHAYAT GÖLBAŞI ŞİMŞEK	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
09:40 – 10:00	Artificial Mixtures for Maximum Likelihood Estimation and Their Generalizations	Alex TSODIKOV, Lyrica Xiaohong LIU, Carol TSENG
10:00 – 10:20	A New Chaotic Steganography Scheme in Spatial Domain	İdris BAYAM, Mustafa Cem KASAPBAŞI
10:20 – 10:40	Evaluation of the Proposed Recommendation System for a Turkish Construction Retail Company using Collaborative Filtering and Frequent Pattern Mining	Waleed ABDULLAH , Mustafa Cem KASAPBAŞI

10:40 – 11:00	C O F F E E / T E A B R E A K

HALL 1 / SESSION D

SESSION CHAIR	PROF. DR. ALEX TSODIKOV	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
11:00 – 11:20	A Comparison of Bayesian and Classical Approaches to Evaluate the Risk Factors for Chronic Kidney Disease in the Elderly Individuals	Elif Çiğdem ALTUNOK, Zehra EREN, Yaşar KÜÇÜKARDALI
11:20 – 11:40	A Study for Visually Comparison of Two Dendograms Using Causes of Death Statistics of Turkey	Elif Çiğdem ALTUNOK, Edis HACILAR
11:40 – 12:00	For The Development Of The Optimization Model, the Data of	Irena HARUTYUNYAN, Lilit AVSHARYAN,

	Improvement of the Quality of Life of the Population In The Republic of Artsakh	Karine HARUTYUNYAN
12:00 – 12:20	**Auxiliary Information based Control Charts for Monitoring Process Location**	**Saddam Akber ABBASI**

HALL 1 / VIDEO SESSION

TIME	PAPER TITLE	PRESENTER / CO AUTHOR
12:20 –12:40	**ANALYSIS OF RAYLEIGH EXPONENTIAL DISTRIBUTION USING THE BAYESIAN APPROXIMATION TECHNIQUE**	**Kahkashan Ateeq,** Saima Altaf , Muhammad Aslam
12:40 –13:00	DISCUSSION	

13:00 – 14:00	LUNCH

HALL 1 / SESSION E

SESSION CHAIR	DR. GELLY MITRODIMA	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
14:00 – 14:20	**Determining the Relationship among Countries' Expenditures in the Certain Areas**	Aylin ADEM, **Ali ÇOLAK,** Metin DAĞDEVİREN
14:20 – 14:40	**Modeling and extracting the term structure of interest rates: A unifying framework**	**Dario PALUMBO**
14:40 – 15:00	**Economic Growth in Turkey – a Threshold Cointegration Approach**	**Magdalena OSINSKA,** Jerzy BOEHLKE, Maciej GALECKI, Marcin FALDZINSKI
15:00 – 15:20	**DISCUSSION**	

HALL 1 / POSTER SESSION F

SESSION CHAIR	DR. Kehinde D. ILESANMİA	
15:20 – 15:40	**PAPER TITLE**	**PRESENTER / CO AUTHOR**

Statistical Analysis on the Winning Factor of NBA and How to Make Playoff	Sungin CHO, Yoon Seo JANG, **Kee-Hoon KANG**

16:00 – 16:20	C O F F E E / T E A B R E A K

13 MAY 2018 SUNDAY

HALL 1 / SESSION G

SESSION CHAIR	DR. FATMA NOYAN TEKELİ	
TIME	**PAPER TITLE**	**PRESENTER** / CO AUTHOR
09:00 – 09:20	**An Alignment Optimization for Measurement Invariance**	**Batuhan ÖZKAN ,** Fatma Noyan TEKELİ
09:20 – 09:40	**A Hybrid Seasonal Autoregressive Integrated Moving Average for the Predicting of Tourism Demand**	**Özlem B. KORKMAZOĞLU,** Gülder KEMALBAY
09:40 – 10:00	**Financial Stress Index for the South African Financial Market**	**Kehinde D. ILESANMİA,** Devi Datt TEWARIB
10:00 – 10:20	**Evaluating Conditional Cash Transfer Policies with Machine Learning Methods**	**Tzai-Shuen Chen**
10:20 – 10:40	**Extrapolative Beliefs and Exchange Rate Markets**	**May Bunsupha**

10:40 – 11:00	CLOSING

FIBONACCI SEQUENCES OF RANDOM VARIABLES

Ismihan BAYRAMOĞLU[1]

[1]Department of Mathematics, Izmir University of Economics, Izmir, Turkey
E-mail: ismihan.bayramoglu@ieu.edu.tr

Abstract

We consider a sequence of random variables constructed on the base of Fibonacci sequence of numbers. It is shown that the structure of this sequence can be determined completely by two initial absolutely continuous random variables and the members of Fibonacci sequence. We investigate the distributional and limit properties of this sequence. Some examples of Fibonacci random sequences leading to new interesting distributions are given. The graphical illustrations are provided. The R code with simulated values and graphs of Fibonacci random sequence is also given.

Key Words: Random variable, distribution function, probability density function, sequence of random variables.

References

[1] Dickson. L. E. (1966). History of the Theory of Numbers, Volume 1, New York: Chelsea.
[2] Gnedenko, B.V. (1978). The Theory of Probability, Mir Publishers, Moscow.
[3] Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Volume 2, John Wiley & Sons Inc. , New York, London, Sydney.
[4] Melham, R.S. and Shannon, A.G. (1995). A generalization of the Catalan identity and some consequences, The Fibonacci Quarterly, 33, 82--84, 1995.
[5] Ross, S. (2016). A First Course in Probability. Prentice-Hall Inc. , NJ.
[6] Skorokhod, A.V. (2005). Basic Principles and Applications of Probability Theory, Springer.

A STUDY FOR VISUALLY COMPARISON OF TWO DENDOGRAMS USING CAUSES OF DEATH STATISTICS OF TURKEY

Elif Çiğdem ALTUNOK[1], Edis HACILAR[2],

[1] Yeditepe University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Istanbul, Turkey. ecaltunok@yeditepe.edu.tr
[2] Yeditepe University, Faculty of Medicine, Phase III student Istanbul, Turkey. edishacilar@gmail.com

Abstract

The researcher should know the importance of fully exploring the features of data before statistical summarization and statistical tests. If the researcher cannot be able to see the nature of the data, research will result with mistaken conclusions. The primary tools of exploratory data analysis are graphics and summary statistics that convert a confusing amount of numbers into pictures and few descriptive numbers that are easily assimilated and understood [1]. Hierarchical Cluster Analysis is a widely used family of unsupervised statistical methods for classifying a set of items into some hierarchy of clusters (groups) according to the similarities among the items [2]. In clustering, clusters are often computed incrementally. In the beginning each object forms its own cluster, and then, step-by-step, the pair of clusters that is closest according to some distance measure is joined. A binary tree called dendrogram, where the leaves represent elements and each inner node of the tree represents a cluster containing the leaves in its sub tree, naturally represents such a hierarchical clustering. Pairs of dendrograms of the same data stemming from different clustering algorithms or parameter settings can be compared visually using tanglegrams [3]. The tanglegram function allows the visual comparison of two dendrograms, from different algorithms or experiments, by facing them one in front of the other and connecting their labels with lines. In this study, it is aimed to present tanglegram as an alternative and powerful comparison tool to explore the two dendograms with an application. The 2016 causes of death statistics of Turkey were taken from TUIK research for the application. R 3.4.4 and dendextend is an R package was used. Using hierarchical clustering analysis, we compared dendograms, which were evaluated from two clustering algorithms single vs complete linkage. The entanglement function, which measures the quality of the tanglegram, was calculated and Cophenetic and Baker correlation coefficients were obtained. As a result, after demonstrating dendograms, tanglegram was illustrated for two different algorithms. The entanglement function corresponds to a good alignment layout and there was a strong relationship between two clustering algorithms. It was shown that two clustering algorithms support each other with few differences. In conclusion, it was found that clusters are validated, and can be used for conclusions. Tanglegram can be used as sensitivity and replicability analysis for researchers who are interested in validating their hierarchical clustering results.

Key Words: *Hierarchical Cluster Analysis, Dendogram, Tanglegram, Entanglement Function*

References

1] LeBlanc D., (2004) Statistics: Concepts and Applications for Science, Canada, Johns &Barlett.

[2] Galili T., dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015 Nov 15;31(22):3718-20.

[3] Martin Nöllenburg, Markus Völker, Alexander Wolff, and Danny Holten, Drawing Binary Tanglegrams: An Experimental Evaluation, Drawing Binary Tanglegrams: An Experimental Evaluation

A COMPARISON OF BAYESIAN AND CLASSICAL APPROACHES TO EVALUATE THE RISK FACTORS FOR CHRONIC KIDNEY DISEASE IN THE ELDERLY INDIVIDUALS

Elif Çiğdem ALTUNOK[1], Zehra Eren[2] Yaşar KÜÇÜKARDALI[2],

[1] Yeditepe University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Istanbul, Turkey. ecaltunok@yeditepe.edu.tr
[2] Yeditepe University, Faculty of Medicine, Department of Internal Medicine, Istanbul, Turkey. zeren@yeditepe.edu.tr ,yasarkardali@yeditepe.edu.tr

Abstract

Chronic kidney disease (CKD) is serious health problem in general population with an increasing incidence and prevalence among the elderly [1]. Poor outcomes of CKD include progression to end-stage kidney failure and complications of decreased kidney function, such as hypertension, anaemia, reduced quality of life etc. The high prevalence of CKD in this population indicates a need for greater awareness regarding the risks of CKD and aggressive management for CKD prevention. Because of that reasons advanced statistical approaches should be performed to predict early recognition [2]. The classical assessment of the risk factors of a disease in a study depends on calculating p-values. The publication of studies hinges upon p values that play a deciding role in whether the data are thought to reflect an actual difference, or random happenstance. However, just because a measure is ubiquitous does not necessarily mean that it is the best measure. In particular, Bayesian methods, and Bayes factors, have been suggested as an excellent alternative to overcome some of the shortcomings of classical tests (frequentist) and the associated p-values [3]. The aim of this study is to compare the Bayesian approach to statistics and to contrast it with the frequentist approach. In this study, data were taken from 8-year single-centre, cohort study consisting of 612 people living in a nursing home from 2005–2013 [4]. Standard demographic, clinical and physiological data were collected and outcome variable glomerular filtration rate (GFR or eGFR) was calculated which is a diagnostic tool for CKD. SPSS 25, AMOS were used for statistical evaluations. For classical approach two independent samples t-tests were used to evaluate the differences in means between the two groups. The Chi-square test was used to compare the frequencies of the groups. Multiple logistic regression analysis was used to evaluate the independent factors of CKD. Age OR= 0.95, 95% CI (0.93–0.96), female sex OR=3.32, 95% CI (2.25–0.91), hypertension OR=2.13, 95% CI (1.44–3.161), congestive heart disease OR=1.56, 95% CI (1.02–2.38) and coronary artery disease (OR 1.84, 95% CI 1.20–2.81) were significantly associated with CKD. For Bayesian approach the models have been estimated using Markov chain Monte Carlo methods with Gibbs sampling. Bayes factors were calculated using the BIC method. Significance level for computing credible intervals was specified 95%, tolerance value was 0,0001. Iteration was specified 2000 and 10000 samples simulated to posterior distribution. Bayes factors were interpreted by commonly used thresholds to define significance of evidence. It was found that Bayesian approach supports the decisions of classical approach. Bayesian methods are more flexible and their results more clinically interpretable, but they require more careful development and specialized software. Using these high evidences, early recognition of CKD might improve drug dosages, treatment of CKD-related comorbidities and renal management to prevent the loss of kidney function.

Key Words: _Bayesian approach, Frequentist approach, Bayesian factors, Odds Ratios, Chronic kidney disease_

References

[1] Magnason RL, Indridason OS, Sigvaldason H, Sigfusson N, Palsson R. Prevalence and progression of CRF in Iceland: a population-based study. Am J Kidney Dis 2002; 40: 955–963.
[2] Kault D, Kault S (2015) From P-Values to Objective Probabilities in Assessing Medical Treatments. PLoS ONE 10(11): e0142132. doi:10.1371/journal.pone.0142132
[3] Jarosz, Andrew F. and Wiley, Jennifer (2014) "What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors," The Journal of Problem Solving: Vol. 7: Iss. 1
[4] Zehra Eren, Yasar Küçükardalı, Mehmet Akif Öztürk, Betül Küçükardalı, Elif Çiğdem Kaspar, and Gülçin Kantarcı,(2015) Geriatr Gerontol Int; 15: 715–720

THE JOINT DISTRIBUTION OF MARGINAL RECORDS IN EXTENDED BIVARIATE RANDOM SEQUENCES

Gülder KEMALBAY[1]

[1]Yıldız Technical University, Faculty of Art & Science, Department of Statistics,
E-mail: kemalbay@yildiz.edu.tr

Abstract

In this study, we are dealing with marginal records in extended bivariate sequences and interested in the joint distributions of marginal record times and values. For this purpose, some distributional properties of upper record vectors are provided. The obtained probability mass function of record times and cumulative density function of record values come into prominence while predicting future records based on the past observations. Assuming record in rainfall intensity and rainfall depth as an example of bivariate record data, we can predict the next record value of rainfall depth given the record value of rainfall intensity having observations up to present time. This prediction is crucial for reducing the risk and preventing the extreme flooding events. However, we provide some numerical and graphical examples for underlying distributions including also independence case.

Key Words: *extended sequence, bivariate records, copula, record time, record value*

References

[1] Ahsanullah, M. (1992). Record values of independent and identically distributed continuous random variables. Pak. J. Statist, 8(2), 9-34.

[2] Ahsanullah, M. (1995). Record statistics. Nova Science Publishers, New York.

[3] Chandler, K. N. (1952). The distribution and frequency of record values. Journal of the Royal Statistical Society. Series B (Methodological), 220-228.

[4] Nagaraja, H. N. (1988). Record values and related statistics-a review. Communications in Statistics-Theory and Methods, 17(7), 2223-2238.

[5] Nevzorov, V. B. (1988). Records. Theory of Probability & Its App., 32(2), 201-228.

[6] Nevzorov, V. B. (2001). Records: Mathematical Theory. Translation of Mathematical Monographs, vol. 194. American Mathematical Society, Providence, RI.

A HYBRID SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE FOR THE PREDICTING OF TOURISM DEMAND

Özlem Berak Korkmazoğlu[1], Gülder KEMALBAY[2],

[1]Department of Statistics, Yıldız Technical University, Istanbul, Turkey, berak@yildiz.edu.tr
[2]Department of Statistics, Yıldız Technical University, Istanbul, Turkey, kemalbay@yildiz.edu.tr

Abstract

This study aims to propose a hybrid model that combines two common methods for predicting the tourism demand: the seasonal auto-regressive integrated moving average method (SARIMA) and Artificial Neural Network (ANN). For this purpose, the Box–Jenkins methodology is applied and several alternative specifications are tested. A methodology based on integrating the data obtained from autoregressive integrated moving average model in the artificial neural network model to predict the number of monthly tourist arrivals. The results indicate that the hybrid models outperform either of the models used separately. This methodology may become a powerful decision-making tool at other inspection facilities of tourism demands.

Key Words: Tourism Forecasting, Hybrid Model, Seasonal Adjustment, ARIMA, ANN.

References

[1] Aslanargun, A., Mammadov, M., Yazici, B., & Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. Journal of Statistical Computation and Simulation, 77(1), 29-53.
[2] Cadenas, E., & Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. Renewable Energy, 35(12), 2732-2738.
[3] Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence, 23(4), 586-594.
[4] Shahrabi, J., Hadavandi, E., & Asadi, S. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. Knowledge-Based Systems, 43, 112-122.
[5] Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism management, 29(2), 203-220.

AN ALIGNMENT OPTIMIZATION FOR MEASUREMENT INVARIANCE

Batuhan ÖZKAN[1] , Fatma NOYAN TEKELİ [2]

[1]Yıldız Technical University, Faculty of Art & Science, Department of Statistics,
batuhan81@gmail.com

[2]Yıldız Technical University, Faculty of Art & Science, Department of Statistics, fnoyan@yildiz.edu.tr

Abstract

Multi item questionnaires are often used to investigate scores on hidden factors such as human values, attitudes and behaviours. Such researches often involve a comparison between certain groups of individuals over time, at one or more points. Significant comparisons of means or relationships between constructs across groups require equivalent measures of these structures. **MEASUREMENT INVARIANCE IS THE DEGREE TO WHICH THE MEASUREMENT MODEL OF A LATENT VARIABLE IS THE SAME ACROSS GROUPS INVOLVED IN THE ANALYSIS.** Multi group confirmatory factor analysis is the most commonly used technique for evaluating measurement invariance (MI). However, when many groups are taken into account, the measurement equality or invariance tests often fail. Asparouhov and Muthén (2014) presented a new method for measurement invariance, called alignment method. In this study, we will discuss two methods for investigating measurement invariance for using monte carlo simulation. Then, we will use this two methods for investigate the measurement invariance of Retail Service Quality Scale across different retailers.

Key Words: measurement invariance, alignment method; multiple group Confirmatory; Service Quality

References

[1] Marsh, H. W., Guo, J., Nagengast, B., Parker, P. D., Asparouhov, T., Muthén, B., & Dicke, T. (2016, accepted). What to do when scalar invariance fails: The extended alignment method for multigroup factor analysis comparison of latent means across many groups. Structural Equation Modeling: A Multidisciplinary Journal.
[2] Muthén, B. & Asparouhov, T. (2016). Recent methods for the study of measurement invariance with many groups: Alignment and random effects.
[3] Asparouhov T. & Muthén, B. (2014). Multiple-group factor analysis alignment. Structural Equation Modeling: A Multidisciplinary Journal, 21:4, 495-508.
[4] Muthén, B. & Asparouhov T. (2014). IRT studies of many groups: The alignment method. Frontiers in Psychology, Volume 5, DOI: 10.3389/fpsyg.2014.00978
[5] Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54:4, 557-585.

SYSTEMICALLY IMPORTANT BANKS OF TURKEY BY USING QUANTILE REGRESSION: A CONDITIONAL VALUE AT RISK (COVAR) APPROACH

Zehra CİVAN[1], Gülhayat GÖLBAŞI ŞİMŞEK[2], Ebru ÇAĞLAYAN AKAY[3]

[1]Department of Statistics, Yıldız Technical University, Istanbul, Turkey, civanzh@gmail.com
[2]Department of Statistics, Yıldız Technical University, Istanbul, Turkey, gulhayat@yildiz.edu.tr
[3]Department of Econometrics, Marmara University, Istanbul, Turkey, ecaglayan@marmara.edu.tr

Abstract

Systemic risk, one of the most discussable and worked subjects after the global financial crisis of 2008, was studied on the behalf of banks operating in Turkey. The aim of this study to analyze the banks in Turkey in terms of systemic risk and to identify systemically important banks of Turkey by using Conditional Value-at-Risk (CoVaR).

One of the measurement methods of systemic risk, CoVaR [1] has been applied by the way of quantile regression [2] in this study. The quarterly and yearly publicly announced financial indicators of the banks were used. During the process of evaluating the contribution of the financial institutions to the financial system's systemic risk, it has been estimated value-at-risk (VaR) and CoVaR of these banks by using quantile regression taking into consideration the growth rate of return of assets of each bank, macro-economic variables of financial system and banking variables. Afterwards, their contribution to systemic risk was estimated separately.

Key Words: *Systemic risk, systemically important bank, conditional value at risk, value at risk, quantile regression*

References

[1] Adrian, T., & Brunnermeier, M. K., (2014). CoVaR, *Federal Reserve Bank of New York Staff Reports*, No:348.

[2] Koenker, R., & Bassett, G., (1978). Regression Quantiles, *Econometrica*, 46 (1), 33-50.

TESTING PERFORMANCE OF HYBRID TIME SERIES MODELS ON HOURLY ELECTRICITY PRICE

Büşra Taş[1], Ceylan Yozgatlıgil[2]

[1]Middle East Technical University, Statistics, Ankara 06800, Turkey
busratas@metu.edu.tr

[2]Middle East Technical University, Statistics, Ankara 06800, Turkey
ceylan@metu.edu.tr

Abstract

Electricity price forecasting is very important in a competitive market. Decision makers highly benefit from accurate forecasting. There should be a balance between electricity production and consumption since electricity cannot be stored. Shocks to demand or supply affect the electricity prices. Therefore, electricity prices show high volatility. In addition, it may have multiple levels of seasonality. Thus it makes forecasting very difficult with conventional methods. Time series generally have both linear and nonlinear patterns. Zhang [1] proposed a hybrid methodology which combines linear and nonlinear components. According to hybrid methodology, predicted values of a time series can be obtained from summation of linear component and nonlinear component. In this study, hybrid models are constructed with SARIMA, TBATS and Neural Network models for analysis of hourly electricity prices in Turkey. Using a hybrid model can give better results in forecasting. Both linear and nonlinear parts of the time series can be modeled by this approach. The data set used in this study is hourly electricity demand (in MWh) and price (in TL/MWh) of Turkey. Series is from 1st January 2012 to 15 January 2018. This time period equivalent to 52944 hours. There are multiple seasonality in the series. In the first hybrid model, Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used to capture the linear behavior of the electricity price series. However, nonlinear patterns cannot be modeled by SARIMA models. In the second hybrid model, TBATS model which is introduced by De Livera et al [2] is used to model linear structure. TBATS model uses exponential smoothing and also allows for automatic Box-Cox transformation and ARMA errors. In both hybrid models, NARX Neural Network is used to model the nonlinearity in the series. Residuals of the SARIMA model and TBATS model are used as output variable in the Nonlinear Autoregressive Model with Exogenous Inputs (NARX). Electricity demand is used as exogenous variable in NARX model. Number of hidden neurons and number of delays are determined according to have a well performed network. SARIMA and TBATS modeling is implemented in R and NARX model is built using Neural Network Time Series Tool in MATLAB. After building these models, one week ahead forecast values are obtained which is equal to 168 hours. Summation of forecast values of linear and nonlinear components is the final forecast value of the electricity price. Forecast performances are compared with RMSE and MAPE values. 4 th International Conference on Advances in Statistics MAY 11-13, 2018, St. PETERSBURG, RUSSIA As a result, more accurate forecasts are obtained by hybrid methodology than using only individual models. The best model is the hybrid model which is the combination of SARIMA and NARX NN model. Therefore, hybrid models are effective to forecast hourly electricity price which shows high volatility and multiple seasonality.

Key Words: Electricity price forecasting; Time series analysis; Hybrid method; Neural Network; TBATS

References

[1] Zhang, G.P. (2003), Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing (50), 159-175.

[2] De Livera, A.M., Hyndman, R.J., & Snyder, R. D. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing, Journal of the American Statistical Association, 106(496), 1513-1527.

[3] Aladag, C.H., Egrioglu, E., Kadilar, C. (2009), Forecasting nonlinear time series with a hybrid methodology. Applied Mathematic Letters (22), 1467-1470.

[4] Khashei, M., Bijari, M. (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing (11), 2664-2675.

[5] Weron, R. (2014), Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting (30), 1030-1081.

A NEW CHAOTIC STEGANOGRAPHY SCHEME
IN SPATIAL DOMAIN

İdris Bayam[1], *Mustafa Cem Kasapbaşı[2]*,

[1] Istanbul Commerce University, Küçükyalı E5 Kavşağı İnönü Cad. No: 4, Küçükyalı 34840, İstanbul
idrisbayam@gmail.com
[2] Istanbul Commerce University, Küçükyalı E5 Kavşağı İnönü Cad. No: 4, Küçükyalı 34840, İstanbul
mckasapbasi@ticaret.edu.tr

Abstract

Steganography is an art of concealing information/message in a medium, so that it can not be noticed by unintended parties. There are many studies offering variety of techniques in the literature utilizing LSB image steganography but enhanced steganographic schemes are needed for improving quality of stego-image as well as computational performance. In this study a new steganographic scheme is presented which utilizes different chaotic maps namely Logistic map, Tent map, Quadratic Map, Bernoulli Map, Sine Map, Chebyshev map in different combinations to select the pixel location for hiding data. The message is compressed before embedding in the cover image so that the capacity of embedding is improved. Quality of the stego image is assessed with not only by statistical values like PSNR, MSE and histogram but also in terms of correlation, entropy, homogeneity, contrast, and energy. It is understood from the statistical results that the proposed scheme is strong and secure for hiding the information in a digital medium. The resultant stego image is almost identical to the original image that is deduced from variety of statistical analyses.

Key Words: Steganography, Chaotic map, spatial domain steganography, PSNR, data compression, statistical analysis

EVALUATION OF THE PROPOSED RECOMMENDATION SYSTEM FOR A TURKISH CONSTRUCTION RETAIL COMPANY USING COLLABORATIVE FILTERING AND FREQUENT PATTERN MINING

Waleed ABDULLAH[1], Mustafa Cem KASAPBAŞI[2],

[1] Istanbul Commerce University, küçükyalı E5 kavşağı inönü Cad.No 4 34840 küçükyalı,Istanbul
waleedabdullah928@gmail.com
[2] Istanbul Commerce University, küçükyalı E5 kavşağı inönü Cad.No 4 34840 küçükyalı,Istanbul,
mckasapbasi@ticaret.edu.tr

Abstract

In this new era of E-Commerce, recommendation systems are mainly requirement of every e-commerce website. Accuracy and efficiency of these systems are the core concern of business. To measure these factors, we have performed analysis on some of the popular techniques. In this study half a million transactions of Turkish Private Construction Retail company were used amongst 1023 products. A detail evaluation of item-item collaborative filtering (CF) and frequent pattern mining (FPM) has been carried out using Cosine, Jaccard and Pearson similarity functions for CF and Apriori , FPGrowth algorithm for FPM respectively. Initially, the similarity matrices are calculated with raw data later, after adding new augmented attributes to the data model similarity matrices are calculated again. K nearest neighbor (KNN) algorithm is applied to propose the recommendations regarding calculated similarity matrices. Results has shown the significant improvement shift of precision score in Cosine and Jaccard of 0.05 and 0.2 respectively by using our proposed data model. An other recommendation comparison is carried out to utilize FPM using WEKA Software and GraphLab Library. Results indicates that Jaccard similarity and FP-Growth algorithm were the best among our analysis.

Key Words: Collaborative Filtering; Frequent pattern Mining; Recommendation system; K nearest neighbor (KNN); E-Commerce

HOMOTHETIC TRANSFORMATION'S INFLUENCE ON EXCESS OF D-OPTIMAL DESIGNS

Yuri D. Grigoriev[1], Viatcheslav B. Melas[2] and Petr V. Shpilev[3],

[1] St. Petersburg State Electrotechnical University, e-mail: yuri_grigoriev@mail.ru
[2] St.Petersburg State University, e-mail: vbmelas@post.ru
[3] St.Petersburg State University, e-mail: p.shpilev@spbu.ru

Abstract

The problem of searching non-singular optimal designs with the minimal number of support points is quite important since the use of such designs allows decreasing experimental expenses. Many works were devoted to the study of this problem, see, e. g., [1]. In pioneer paper [2] de la Garza shown that d-optimal designs are always saturated for polynomial regression models, i. e. the number n of support points of these designs coincides with the number p of unknown parameters of the regression model. On the other hand, for nonlinear in parameters models, cases in which optimal designs arise with the number of support points n > p are not rare. The series of papers [3], [4] and [5] consider the question on transferring de la Garza result to nonlinear models. Most authors are concentrated their attention on models with one explaining variable whereas many regression models used in practice are multidimensional. These models are much more difficult to study. To a large extent, this is related to the fact that Chebyshev systems of functions do not exist in this case see [3]. In our recent paper [6] we proposed to call such cases the *excess phenomenon*, and the corresponding designs *excess designs*.

In present work we provided the analytical solution of the problem of finding the dependence between the number of the locally optimal design support points and the lengths of the design intervals for the Cobb-Douglas model which is used in microeconomics. The saturated optimal designs were constructed in explicit form. To find the excess optimal designs one could use various numerical methods. The main idea of our paper was to show how a homothety of the design space *X* influences the form of an optimal design. We suppose that the approach proposed in our work will be useful for distinguishing classes of models for which the homothety transformation leads to excess designs, and classes of models in which this property does not take place.

Key Words: *Excess design; Locally D-Optimal Designs; Homothetic transformation; Cobb-Douglas model;*

References

[1] Pukelsheim, F. (2006) *Optimal Design of Experiments*. SIAM, Philadelphia.

[2] de la Garza, A. (1954) Spacing of information in polynomial regression, *Ann. Math. Statist.*, 25:123-130.

[3] Dette, H. and Melas, B. (2011) A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.*, 39(2):1266--1281.

[4] Yang, M. and Stufken, J. (2009) Support points of locally optimal designs for nonlinear modelswith two parameters. *Ann. Statist.*, 37:518-541.

[5] Yang, M. and Stufken, J. (2012) Identifying locally optimal designs for nonlinear models: a simple extension with profound consequences. *Ann. Statist.*, 40(3):1665-1681.

[6] Grigoriev, Yu. D., Melas, V. B., Shpilev, P. V.(2017), Excess of locally D-optimal designs and homothetic transformations. *Vestnik St. Petersburg University: Mathematics*, 50(4): 329-336.

INVESTIGATION OF RISK PERFORMANCES OF THE NEW HETEROGENEOUS ESTIMATORS

Selahattin KAÇIRANLAR[1], Nimet ÖZBAY[2],

[1]Çukurova University, Faculty of Science and Letters, Department of Statistics, Adana, Turkey
skacir@cu.edu.tr
[2] Çukurova University, Faculty of Science and Letters, Department of Statistics, Adana, Turkey
nturker@cu.edu.tr

Abstract

Homogeneous and heterogeneous minimum mean square error estimators have considerable attention in the literature. Farebrother [1] offered an adaptive form of the minimum homogeneous mean square error estimator. Stahlecker and Trenkler [4] put forward a minimum heterogeneous estimator incorporating some prior information. Then, Tracy and Srivastava [5] recommended an adaptive form of the Stahlecker and Trenkler's heterogeneous estimator by regulating this estimator as a convex combination of the ordinary least squares estimator and its mean vector. Shrinkage estimators can be handled in the context of homogeneous and heterogeneous minimum mean square error estimators. In a Bayesian point of view, Lindley [5] derived an estimator by shrinking the ordinary least squares estimator toward a nonzero prior mean which is named as Lindley's mean correction. In this study, we develop two methods for proposing two new shrinkage estimators. At first, we make use of the Lindley's mean correction in the heterogeneous minimum mean square error estimator of Stahlecker and Trenkler [4] and offer a new adaptive form of this estimator. Afterwards, the Lindley's mean correction is used as a mean vector for the unknown regression parameter and we propose a new shrinkage estimator by the way of Bayesian consideration. In order to analyze risk properties of our new estimators we prefer well known quadratic loss function as well as extended balanced loss function of Shalabh et. al [3]. Risk performances of the new estimators are examined via an extensive Monte Carlo experiment. The numerical outcomes show that our new methods are preferable to the old ones.

Key Words: *Extended balanced loss function; Heterogeneous minimum mean square error estimator; Quadratic loss function; Risk performance; Shrinkage estimator*

References

[1] Farebrother RW (1975). The minimum mean square error linear estimator and ridge regression, Technometrics, 17:127-128.
[2] Lindley DV (1962). Discussion of Professor Stein's Paper, Journal of the Royal Statistical Society Ser. B, 24:285-287.
[3] Shalabh, Toutenburg H, Heumann C (2009). Stein-rule estimation under an extended balanced loss function, Journal of Statistical Computation and Simulation, 79:1259-1273.
[4] Stahlecker P, Trenkler G (1985). On heterogeneous versions of the best linear and the ridge estimator. Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications, 301-322, Department of Mathematical Sciences, University of Tampere, Finland.
[5] Tracy DS, Srivastava AK (1994). Comparision of operational variants of best homogeneous and heterogeneous estimators in linear regression. Communications in Statistics Theory and Methods, 23:2313-2322.

ON SOME PROBLEMS OF THE OPTIMAL CHOICE
OF RECORD VALUES

Igor V. BELKOV[1], Valery B. NEVZOROV[2]

[1] Department of Mathematics and Mechanics, St-Petersburg State University, St-Petersburg, Russia,
igor.belkov@gmail.com
[2] Department of Mathematics and Mechanics, St-Petersburg State University, St-Petersburg, Russia,
valnev@mail.ru

Abstract

Independent random variables $X_1, X_2,..., X_n$ having $U([0,1])$-uniform distribution and upper and lower record values in this set are considered. We study the problem how to maximize (taking into account some consecutively observed values $x_1, x_2,..., x_k$ of these X's) the expectation of sums of records in this sequence under the optimal choice of the corresponding value x_k of X_k (instead of X_1) as the initial record value. The following questions are discussed and the considered problems are solved.

1) How to maximize the expectation of the sum of upper record values amongst these X's?
2) How to maximize the expected value of the sum of lower records?
3) What is the maximal possible value in the suggested scheme of the expectations of sums and differences of upper and lower records?
4) What is the way to obtain the maximal mean values of the numbers of upper and lower records?

**Key Words:** record times, record values, expected number of records, uniform distribution, optimal choice problem

References

[1] Belkov IV, Nevzorov VB (2017) Zap. Nauchn. Sem. POMI 466:30-37 (in Russian).
[2] Belkov IV, Nevzorov VB (2018) Vestnik SPbSU. Mathematics, Mechanics, Astronomy 5(63):2 (in Russian, in print).

SECOND ORDER ASYMPTOTICS FOR INTERMEDIATE TRIMMED SUMS AND L-STATISTICS

Nadezhda GRIBKOVA

St. Petersburg State University, Mathematic and Mechanic Faculty, 199034, Universitetskaya nab. 7/9, St. Petersburg, Russia, n.gribkova@spbu.ru

Abstract

The class of *L*-statistics is one of the most commonly used classes in statistical inferences. There is an extensive literature on asymptotic properties of *L*-statistics, but its part relating to large deviations is not so vast. We can mention a few of highly sharp results on this topic for *L*-statistics with smooth weight functions ([1], [6] and references therein). As to the trimmed *L*-statistics, the first – and up to the recent time the single – result on probabilities of large deviations was obtained in [2], but under some strict and unnatural conditions. Recently, the latter result was strengthened in [4], where a different approach, other than in [2], was proposed and implemented; this approach allowed us to establish in [3]-[5] a number of new results on large and moderate deviations under quite mild and natural conditions.

Some of our recent results from [3]-[5] will be presented in the talk. We will also discuss our approach for studying asymptotic properties of trimmed *L*-statistics based on a stochastic approximation with use of Winsorized observations.

Key Words: *large deviations; moderate deviations; L-statistics; intermediate trimmed mean; asymptotic normality.*

References

[1] Bentkus,V., Zitikis, R. (1990) Probabilities of large deviations for *L*-statistics, Lithuanian Mathematical Journal, 30: 215–222.

[2] Callaert, H., Vandemaele, M. and Veraverbeke, N. (1982) A Cramér type large deviations theorem for trimmed linear combinations of order statistics, Communications in Statistics – Theory and Methods, 11: 2689–2698.

[3] Gribkova, N.V. (2017) Cramér-type moderate deviations for intermediate trimmed means, Communications in Statistics – Theory and Methods, 46: 11918–11932.

[4] Gribkova N.V. (2017) Cramér type large deviations for trimmed *L*-statistics, Probability and Mathematical Statistics, 37: 101–118.

[5] Gribkova, N.V. (2016) Cramér type moderate deviations for trimmed *L*-statistics, Mathematical Methods of Statistics, 25: 313-322.

[6] Vandemaele, M. Veraverbeke, N. (1982) Cramér type large deviations for linear combinations of order statistics, Annals of Probabability, 10: 423–434.

ARTIFICIAL MIXTURES FOR MAXIMUM LIKELIHOOD ESTIMATION AND THEIR GENERALIZATIONS

Alex TSODIKOV[1], Lyrica Xiaohong LIU[2], and Carol TSENG[3],

[1]University of Michigan, School of Public Health, Department of
Biostatistics, 1415 Washington Heights, Ann Arbor, MI 48109,
tsodikov@umich.edu
[2]Amgen South San Francisco, One Amgen Center Dr. Thousand Oaks,
California 91320, lyricaliu@gmail.com
[3]H2O Clinical, LLC, 200 International Circle, STE 5888, Hunt Valley, MD
21030 carol.tseng@h2oclinical.com.

Abstract

We offer an approach of representing a complicated likelihood for a statistical model as a marginal one based on artificial missing data. If artificial missing data were observed, we would have the so-called complete-data likelihood, whose choice is not unique and is to some extent up to us. When choosing the form of the complete-data likelihood, simplification is sought in the likelihood factorization at the complete data level. Semiparametric survival models and models for categorical data are used as an example. A generalization of the approach serves a situation when the model at the complete data level is not a legitimate probability model or if it does not exist at all. The method is used to formulate and provide an estimating procedure for a novel Copula-based semiparametric model for multivariate survival data that combines features of Gamma and positive stable shared frailty models. The method is applied to analyse data on time to blindness in diabetic retinopathy patients.

Key Words: *Biostatistics; Semiparametric multivariate survival models; Maximum Likelihood Estimation; Diabetic retinopathy*

A BAYESIAN QUANTILE TIME SERIES MODEL FOR ASSET RETURNS

Gelly Mitrodima[1], Jim Griffin[2]

[1]Department of Statistics, London School of Economics, Columbia House, Houghton Street, London WC2A 2AE, E.Mitrodima@lse.ac.uk
[2]University of Kent, United Kingdom

Abstract

The conditional distribution of asset returns has been widely studied in the literature using a wide range of methods that usually model its conditional variance. However, empirical studies show that the returns of most assets display time-dependence beyond volatility, and there is difficulty with fitting their extreme tails. Our aim is to study the time variation in the shape of the return distribution by jointly modelling a finite collection of quantiles over time under a Bayesian nonparametric framework. Formal Bayesian inference on quantile is challenging since we need access to both the quantile function and its inverse. We employ a flexible Bayesian implementation of a conditional transformation model and we propose a novel class of Bayesian nonparametric priors for quantiles. This allows fast and efficient Markov chain Monte Carlo (MCMC) methods to be applied for posterior simulation and forecasting. Under this Bayesian nonparametric framework, we avoid strong parametric assumptions about the underlying distribution, and so we obtain a model that is flexible about the shape of the distribution. We show that the proposed model can be used to define a stationary process of distributions. In our empirical exercise, we find that the model fits the data well, offers robust results, and acceptable forecasts for a sample of stock, index, and commodity returns.

A NEW MODIFICATION OF PROBABILITY PARADOX

Jan NOVOTNÝ[1], Jindriska SVOBODOVÁ[2]

[1]Faculty of Education, Masaryk University, Porici 7, 60300 Brno, novotny@physics.muni.cz
[2]Faculty of Education, Masaryk University, Porici 7, 60300 Brno, svobodova@ped.muni.cz

---

Abstract

This article describes a new modified probability paradox Janosik the Robber. It uses one Slovak legend of hero, who took from rich people and money was given to the poor. That statistic and probability paradox is discussed in context with another related paradoxes. A computer application that facilitates investigation of that paradox is presented. The implementation and use of the application are explained. Paradoxes can play a useful role in the classroom for fruitful discussions and provoke deeper thinking about the nonintuitive probabilistic ideas. The students could be encouraged to think about how to design their own real experiment to simulate that paradox situation. That paradox problem and his simulation is suitable for introductory science or statistics lecture. The students responses is presented, too.

Key Words: *probability paradoxes, statistics for sciences, education*

References

[1] Gardner, M.(1981). Aha! Gotcha, paradoxes to puzzle and delight. W.H. Freeman Co.
[2] Hald, A., 1998. A History of Mathematical Statistics. Wiley, New York.
[3] Novotný, J., Svobodová J. (2014) Jak pracuje věda, Brno, Masarykova univerzita, 2014.

STATISTICAL ANALYSIS ON THE WINNING FACTOR OF NBA AND HOW TO MAKE PLAYOFF

Sungin Cho[1], Yoon Seo Jang[1], <u>Kee-Hoon Kang[2]</u>

[1]Department of Statistics, Hankuk University of Foreign Studies, Yongin 17035, Korea.
[2]Department of Statistics, Hankuk University of Foreign Studies, Yongin 17035, Korea. e-mail: khkang@hufs.ac.kr

Abstract

In addition to professional sports such as soccer, baseball and basketball, archery, fencing, and athletics are also judged as records. By collecting and analysing these records, you can identify the factors of victory. Owing to advances in technology and increased interest in sports, statistics are frequently used in a variety of sports as well as baseball represented by the sabermetrics. In this study, we try to analyse the two research themes using data obtained from the NBA in America. The first topic is the analysis of victory factor of the game. We analyse the NBA data from 2014 to 2016 and compare it with the previous research results of Oliver (2004). The second one is the analysis of playoff entry factors, a common goal of all teams in the NBA. For this, seasonal team data from 2006 to 2016 were used to analyse. The most important factor for victory is earning points aggressively. Variables for representing mistakes also appeared to be important, but above all, scoring was a priority. Unlike previous studies by Oliver (2004), the importance of free throw is very low. In the past, most of the teams relied heavily on the center, and the center had a very low free throw success rate, and the win was depending on how well they put free throws. In modern basketball, however, the game is played around guard and forward. Most guards and forwards lead to scoring with a high free throw success rate of over 95%. Therefore, it seems that the preceding factors such as fouls and turnovers, which are direct causes of free throws, are more important than free throws. It has been shown that reducing the number of mistakes and the success of defensive play are the most important factors in entering the playoffs. In other words, building teamwork to determine mistakes and defences was identified as an important factor.

Key Words: Correlation analysis; linear discriminant analysis; logistic regression; random forest; variable selection.

References

[1] Hu F, Zidek JV (2004) Lecture Notes-Monograph Series, 385-395.
[2] Maymin P (2013) MIT Sloan Sports Analytics Conference.
[3] Simonoff JS (1998) Smoothing Methods in Statistics, Springer-Verlag, New York.
[4] Oliver D. (2004) Basketball on Paper, Brassey's inc., Dulles, Virginia.

ANALYSIS OF RAYLEIGH EXPONENTIAL DISTRIBUTION USING THE BAYESIAN APPROXIMATION TECHNIQUE

Kahkashan Ateeq[1], Saima Altaf[2] and Muhammad Aslam[3]

[1]Department of Statistics, The Women University Multan, Pakistan: kahkashanateeq98@yahoo.com
[2]Department of Statistics, Bahauddin Zakariya University, Multan 60800, Pakistan: nodupk@yahoo.com
[3]Department of Statistics, Bahauddin Zakariya University, Multan 60800, Pakistan: aslamasadi@bzu.edu.pk

Abstract

As complexities, diversities and variations exist in our real world, different statistical distributions are derived to model them. Still, there are many important circumstances where it is difficult to model the real life data as they apparently do not follow any standard probability distribution. Because of this, efforts have always been extended for the development and advancement of generalized statistical models.

In this study, we introduce a generalization, named as the Rayleigh exponential distribution using the Transformed-Transformer method [1]. This method is elaborated briefly. A random variable *T* follows the Rayleigh distribution, is transformed through the function of cumulative distribution function of exponential distribution. We have also showed that in some real life phenomena, this distribution performs well than some other existing distributions.

In the current study, we have also explored the Rayleigh exponential distribution in Bayesian Paradigm using noninformative priors. In Bayesian study, probability distribution is assigned not only to the observed data but also to the unknown parameters. Then a posterior distribution is obtained which contains all the probabilistic information about parameters.

The chief objective of this study is to compare classical and Bayesian estimation techniques. The expression for the estimators of unknown parameters of said distribution are obtained using maximum likelihood method and through Bayesian approach under five different loss functions, which are square error, weighted, quadratic, precautionary and modified II loss functions.

Selection of prior is mandatory in Bayesian frame work. Sometimes prior distribution has less or no information about the unknown parameters as compared to the likelihood function. In such situations, it is better to use noninformative priors. Jeffreys' and uniform priors are considered as noninformative priors which we have used in this article. Noninformative priors are used to derive the posterior distribution of parameters and then the posterior estimates are obtained.

A complete implementation of Lindley approximation technique for the estimation of Bayes estimators [2] is given. Simulation technique is used to compare the performance of the Bayes estimates under noninformative priors with the maximum likelihood estimates obtained through theoretical results. Random samples have been generated from Rayleigh exponential distribution using the technique proposed by Alzaatreh, Lee et al. (2013) [1]. The frequentist and Bayes estimators are compared through risk functions. Results show that the Bayes estimators using uniform as well as Jeffreys' priors have smaller risk than those of the maximum likelihood estimators. Performance of quadratic loss function is best for one parameter and for the other one, modified II loss function performs best.

A real life application is analyzed for the performance of classical and Bayes estimators.

It is concluded that, when we compare the maximum likelihood estimators with the Bayes estimators using Lindley's approximation in term of their risk functions, the Bayes estimators using noninformative priors, perform better than the maximum likelihood estimators. However, for large sample sizes, the Bayesian and classical estimates become closer in terms of risk function.

Key Words: Rayleigh distribution, Lindley's approximation, Bayesian analysis, square error loss function, Transformed-Transformer method

References

[1]Alzaatreh A, Lee C, Famoye F (2013)A new method for generating families of continuous distributions Metron 71(1): 63-79.

[2]Lindley DV (1980) Approximate bayesian methods Trabajos de estadística y de investigación operativa 31(1): 223-245.

AUXILIARY INFORMATION BASED CONTROL CHARTS FOR MONITORING PROCESS LOCATION

Saddam Akber ABBASI

Department of mathematics, Statistics and Physics, Qatar University, Doha, Qatar
e-mail: sabbasi@qu.edu.qa

Abstract

Control chart acts as the most important tool for the monitoring of process parameters. Although firstly proposed for manufacuring industry, control charts are recently used in a number of fields such as chemical laboratories [1], nuclear engineering [2], health sciences[3] etc. Efficient control charts are always desirable for the detection of abnormal variations in process parameters at an early stage. Control charts typically work in two phases: Phase I (retrospective phase) and Phase II (prospective phase). Phase I involves the estimation of parameters and control limits based on a clean historical set of samples whereas Phase II charts are applied for the online monitoring of processes. Recently, control charts based on auxiliary information have been proposed for the monitoring of process location in Phase II (cf. [4, 5]). These charts have been shown to be more efficient as compared to usual Shewhart type location charts. As per our knowledge, no work has been done to investigate the performance of auxiliary information based location charts in Phase I. The use of auxiliary information can help in increasing the preciseness of parameter estimates and in return increasing the efficiency of the control chart procedures in Phase I. In this study, we propose and investigate the performance of auxiliary information based charts for the monitoring of process location in Phase I. Assuming bivariate normality of the variables, the control limit structures are designed for the proposed chart. Moreover, the performance of the charts is investigated and compared with the usual Shewhart $\bar{X}$ chart considering localized mean disturbances in the Phase I dataset. Probability to signal is used as a performance measure in this study following [6]. The comparisons revealed that that the auxilairy information based location chart outperforms the usual Shewhart $\bar{X}$ chart in terms of detection of contaminations in Phase I data. A real life example is also provided to illustrate the application of the proposed chart. This study will help quality practitioners to choose an efficent chart for the monitoring of process location in Phase I.

References

[1] Abbasi, S. A. (2016) Exponentially moving average control chart and two component measurement error. Quality and Reliability Engineering International 32:2, 499-504.

[2] Hwang S. L., Lin J. T., Liang G. F., Yau Y. J., Yenn T. C. and Hsu C. C. (2008). Application control chart concepts of designing a pre-alarm system in the nuclear power plant control room. Nuclear Engineering and Design. 238:12, 3522–3527.

[3] Woodall W. H. (2006). The use of control charts in health-care and public-health surveillance. Journal of Quality Technology. 38:2, 89–104.

[4] Riaz, M. (2008). Monitoring process mean level using auxiliary information. Statistica Neerlandica, 62:4, 458–481.

[5] Abbasi, S. A. and Riaz, M. (2016). On dual use of auxiliary information for efficient monitoring. Quality and Reliability Engineering International, 32:2, 705-714.

[6] Abbasi, S. A., Riaz, M., Miller, A. And Ahmad, S. (2015) On the performance of Phase I dispersion control charts for process monitoring. Quality and Reliability Engineering International, 31:8, 1705-1716.

BOUNDS IN COMBINATORIAL CENTRAL LIMIT THEOREM

Andrei FROLOV[1],

[1]Author's contact address and e-mail: Dept. of Math. & Mechanics, St.Petersburg State University, Universiteskii pr. 28, Stary Peterhof, St.Petersburg, Russia; a.frolov@spbu.ru

Abstract

We discuss Esseen type bounds for the remainder in a combinatorial central limit theorem (CLT). We start with the case of non-independent random variables. No moment assumptions are assumed. Next, we turn to the Esseen bounds in a combinatorial CLT for independent random variables with finite moments of order p>2. We also show that the combinatorial CLT may hold while variations are infinite. The case of random combinatorial sums is mentioned. Applications to moderate deviations of combinatorial sums are discussed as well.

Key Words: combinatorial central limit theorem; Esseen inequality; moderate deviations; combinatorial sum

References

[1] Frolov A.N. (2014) Esseen type bounds of the remainder in a combinatorial CLT. J. Statist. Planning and Inference 149, 90-97.

[2] Frolov A.N. (2015) Bounds of the remainder in a combinatorial central limit theorem.

Statist. Probab. Letters 105, 37-46.

[3] Frolov A.N. (2017) On Esseen type inequalities for combinatorial random sums. Communications in Statistics -Theory and Methods. 46 (12), 5932-5940.