

BOOK OF ABSTRACTS



6th International Conference on Advances in Statistics

ICAS CONFERENCE
INTERNATIONAL CONFERENCE ON ADVANCES IN STATISTICS

October 16-18 2020

<http://www.icasconference.com/>

ICAS'2020

6th International Conference on Advances in Statistics

Published by the ICAS Secretariat

Editors:

Prof. Dr. Ismihan BAYRAMOGLU and Prof. Dr. Fatma NOYAN TEKELI

ICAS Secretariat

Büyükdere Cad. Ecza sok. Pol Center 4/1 Levent-İstanbul

E-mail: icasconference.academic@gmail.com

<http://www.icasconference.com>

ISBN: 978-605-65509-1-1

Copyright @ 2020 ICAS and Authors

All Rights Reserved

No part of the material protected by this copyright may be reproduced or utilized in any form or by any means electronic or mechanical, including photocopying , recording or by any storage or retrieval system, without written permission from the copyrights owners.

SCIENTIFIC COMMITTEE

Prof. Dr. Ayman BAKLEEZI

Qatar University – Qatar

Prof. Dr. Barry C. ARNOLD

University of California, Riverside – USA

Prof. Dr. George YANEV

The University of Texas Rio Grande Valley – USA

Prof. Dr. Hamparsum BOZDOGAN

The University of Tennessee – USA

Prof. Dr. Hamzeh TORABI

Yazd University – IRAN

Prof. Dr. Jorge NAVARRO

Universidad de Murcia – Spain

Prof. Dr. Jose Maria SARABIA

University of Cantabria – Spain

Prof. Dr. Leda MINKOVA

University of Sofia – Bulgaria

Prof. Dr. Narayanaswamy BALAKRISHNAN

McMaster University – Canada

Prof. Dr. Nikolai KOLEV

University of Sao Paulo – Brasil

Prof. Dr. Sarjinder SINGH

Texas A&M University-Kingsville – USA

Prof. Dr. Stelios PSARAKIS

Athens University of Economics & Finance – Greece

Dr. Ilham AKHUNDOV

University of Waterloo – Canada

Prof. Dr. Aydin ERAR

Mimar Sinan Fine Arts University – Turkey

Prof. Dr. I. Esen YILDIRIM

Marmara University – Turkey

Prof. Dr. Gulay BASARIR

Mimar Sinan Fine Arts University – Turkey

Prof. Dr. Mujgan TEZ
Marmara University – Turkey

Prof. Dr. Sahamet BULBUL
Istanbul Ayvansaray University – Turkey

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

Assoc. Prof. Dr. Esra AKDENIZ
Marmara University – Turkey

ORGANIZATION COMMITTEE

Prof. Dr. Ismihan BAYRAMOGLU
Izmir University of Economics – Turkey
Conference Chair

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

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

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

Instructor PhD Ozlem BERAK KORKMAZOGLU
Yıldız Technical University – Turkey

Dear Colleagues,

On behalf of the Organizing Committee, I am pleased to invite you to participate in **6th INTERNATIONAL E – CONFERENCE ON ADVANCES IN STATISTICS** which will be organised **fully virtual** on dates between 16-18 October 2020 .

*Conference was originally planned for April 2020 but due to the global spread of COVID-19 (Corona Virus) and The Council of Higher Education's declaration on "Measures to be Taken in Higher Education Institutions about COVID-19" (March 6, 2020) the conference is postponed to this **current date**.*

All informations are available in conference web site. For more information please do not hesitate to contact us. info@icasconference.com

We cordially invite prospective authors to submit their original papers to ICAS-2020,

- Applied Statistics
- Bayesian Statistics
- Big Data Analytics
- Bioinformatics
- Biostatistics
- Computational Statistics
- Data Analysis and Modeling
- Data Envelopment Analysis
- Data Management and Decision Support Systems
- Data Mining
- Energy and Statistics
- Entrepreneurship
- 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
- 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 **Journal of the Turkish Statistical Association**. <http://jtsa.ieu.edu.tr>

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

Conference Website : <https://icasconference.com>

E Mail: icasconference.academic@gmail.com

On behalf of Organizing Committee:

Conference Chair

Prof. Dr. İsmihan BAYRAMOĞLU

Izmir University of Economics

16 OCTOBER 2020 FRIDAY

10:30 – 11:00

OPENING CEREMONY

Professor Ismihan BAYRAMOGLU / Izmir University of

Economics - Turkey

11:00 – 11:30	Keynote Speech: Professor Serkan ERYILMAZ / Atılım University - Turkey <i>Statistical Aspects of Wind Energy</i>
----------------------	--

11:30 – 11:45	B R E A K
----------------------	------------------

SESSION A

SESSION CHAIR	Gulder KEMALBAY	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
11:45 – 12:00	Polya-Aeppli Process of Order k of Second Kind with an Application	Meglana LAZAROVA / Stefanka CHUKOVA & Leda MINKOVA
12:00 – 12:15	A Statistical Consistent Test Based on Bivariate Random Thresholds	Aysegul EREM / Ismihan BAYRAMOGLU
12:15 – 12:30	Performance Analysis of Ridge Deviance Control Charts for Monitoring Poisson Profiles	Ulduz MAMMADOVA / M. Revan OZKALE
12:30 – 12:45	Gompertz-Exponential Distribution: Record Value Theory and Applications in Reliability	Shakila BASHIR / Ahmad Mahmood QURESHI
12:45 – 13:00	On the Order Statistics of Dependent Random Variables Constructed from Bivariate Random Sequences	Ismihan BAYRAMOGLU / Omer L. GEBIZLIOGLU

13:00 – 13:30	LUNCH BREAK
----------------------	--------------------

SESSION B

SESSION CHAIR	Fatma NOYAN TEKELI	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
13:30 – 13:45	Robust Ranked Set Sampling Methods for One-Sample T-Test	Yusuf Can SEVIL / Tuğba YILDIZ
13:45 – 14:00	On Classification with Multiple Birth Support Vector Machines	Guvenc ARSLAN
14:00 – 14:15	New Mathematical Formulations for the Distributed Permutation Flowshop Scheduling Problem	Alper HAMZADAYI / Hanifi Okan ISGUDER
14:15 – 14:30	Handling Missing Values in Random Forests: An Application to Demographic Survey Data	Duygu ICEN / Ayse ABBASOGLU OZGOREN & Anil BOZ SEMERCI
14:30 – 14:45	Tail Dependence Estimation Based on Estimation of Kendall Distribution Function via Rational Bernstein Polynomials	Mahmut Sami ERDOGAN / Selim Orhun SUSAM
14:45 – 14:50 (Poster)	POT Method for Ruin Probability in Infinite Time with Non-Stationary Arrivals and Heavy Tailed Distribution Claims or Loss Models	Redhouane FRIHI / Rassoul ABDELAZIZ

14:50 – 15:00	B R E A K
----------------------	------------------

SESSION C

SESSION CHAIR	Jale ORAN	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
15:00 – 15:15	SHARP: A State-Space HAR Model Using Particle Gibbs Sampling	Aya GHALAYINI / Marwan IZZELDIN & Mike TSIONAS
15:15 – 15:30	Inflation Targeting, Credibility and Taylor Rule: The Estimation of Monetary Policy Reaction Function for the Central Bank of Turkey	Gozde YILDIRIM / Ahmet TIRYAKI
15:30 – 15:45	Machine Learning Extension of the Simulated Method of Moments for Estimation of Agent-Based Models	Jiri KUKACKA

15:45 – 16:00	The Impact of Periodicity on Volatility-Volume Relations	Sherry LUO / Zhen WEI & Marwan IZZELDIN

16:00 – 16:15	B R E A K	
----------------------	------------------	--

16:15 – 16:45	Keynote Speech: Professor Jale ORAN / Marmara University - Turkey <i>Behavioral Finance: A Retrospective</i>	
----------------------	--	--

17 OCTOBER 2020 SATURDAY

10:30 – 11:00	Keynote Speech: Professor Rza BASHIROV / Eastern Mediterranean University - North Cyprus <i>Statistical Comparison of Modelling Approaches Demonstrated for Biomedical Networks</i>	
----------------------	---	--

11:00 – 11:15	B R E A K	
----------------------	------------------	--

SESSION D

SESSION CHAIR	Guvenc ARSLAN	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
11:15 – 11:30	Bell Marginal Models for Longitudinal Count Outcomes	Hatice Tul Kubra AKDUR
11:30 – 11:45	The Interplay Between Determinism, Stochasticity And Fuzzyness Illustrated For P16-Mediated Pathway	Nimet Ilke AKCAY / Rza BASHIROV
11:45 – 12:00	Modelling Pre-service Mathematics Teachers Reasoning Under Uncertainty in the Egyptian Context	Samah Gamal Ahmed ELBEHARY
12:00 – 12:15	Hypogeometric Distribution and Related Discrete Time Point Process	Silvana PARALLOJ / Stefanka CHUKOVA & Leda MINKOVA
12:15 – 12:30	Fusion of Geometric and Texture Features for Side View Face Recognition	Salman Mohammed JIDDAH, Main ABUSHAKRA, Kamil YURTKAN

12:30 –13:30	LUNCH BREAK
--------------	--------------------

SESSION E

SESSION CHAIR	Gulhayat GOLBASI SIMSEK	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
13:30 – 13:45	Different Similarity Measures for the Clustering of Time Series	Yamina KHEMAL-BENCHEIKH / Assia BOUIZANE
13:45 – 14:00	A Method for Constructing and Interpreting Some Weighted Premium Principles	Gema FIGUEIRAS / Antonia CASTAÑO-MARTÍNEZ & Fernando LÓPEZ-BLAZQUEZ & Miguel A. SORDO
14:00 – 14:15	A Bayesian Model of COVID19 New Cases	Marta SANCHEZ-SANCHEZ / Alfonso SUÁREZ-LLORENS & Ángel BERIHUETE
14:15 – 14:20 (POSTER)	A Survey Of Groundwater Quality in Suburb of Ulaanbaatar City, Mongoliahydrochemical Investigation of Groundwater in 14 Th Khoroo of Khan-Uul District Using Multivariate Statistical Techniques	Enkhbayar JAMSRANJAV / Dagvasuren GANBOLD & Gerelt-Od DASHDONDOG & Munkhtsetseg ZORIGT

14:20 – 14:30	BREAK
---------------	--------------

14:30 – 15:00	Keynote Speech: Professor Agamirza BASHIROV / Eastern Mediterranean University - North Cyprus <i>Wide Band Noises: Theory and Applications</i>
---------------	--

SESSION F

SESSION CHAIR	Aysegul EREM	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
15:00 – 15:15	Evaluating the Effects of Outliers in Bootstrap	Ugur BINZAT / Engin YILDIZTEPE

15:15 – 15:30	An Alternative P Chart For Monitoring High-Quality Processes Based on Improved Estimator	Senem SAHAN VAHAPLAR / Ozlem EGE ORUC
15:30 – 15:35 (Poster)	Estimating the Gini Index for Income Loss Distributions Under Random Censoring	Bari AMINA / Abdelaziz RASSOUL & Ould Rouis HAMID

15:35 – 16:15	B R E A K
----------------------	------------------

SESSION G

SESSION CHAIR	Umut UYAR	
TIME	PAPER TITLE	PRESENTER / CO AUTHOR
16:15 – 16:30	Regression Discontinuity Design in the Analysis of South African Social Development Praxis	Doug ENGELBRECHT / Joshua ENGELBRECH
16:30 – 16:45	Assessing The Impact of Microfinance: Findings From a Survey of Microfinance Participants in Akole Taluka of Maharashtra, India	Amita YADWADKAR
16:45 – 17:00	Currency Devaluation Versus Tariff – A Trade War Simulation	Jen-CHI CHENG / Bryce ENGELLAND
17:00 – 17:15	Effects of The Epidemic on The Bist Network Structure	Deniz SUKRUOGLU
17:15 – 17:30	Predictive Power of Exchange Rates and Interest Rates for Capacity Utilization and Real Sector Confidence in Turkey	Sıtkı SONMEZER / Ismail Erkan CELIK

POLYA-AEPPLI PROCESS OF ORDER K OF SECOND KIND WITH AN APPLICATION	
<i>Stefanka CHUKOVA¹, Meglena LAZAROVA², Leda MINKOVA³</i>	16
A STATISTICAL CONSISTENT TEST BASED ON BIVARIATE RANDOM THRESHOLDS	
<i>Aysegul EREM¹, Ismihan BAYRAMOGLU²</i> ,	17
PERFORMANCE ANALYSIS OF RIDGE DEVIANCE CONTROL CHARTS FOR MONITORING POISSON PROFILES	
<i>Ulduz MAMMADOVA¹, M. Revan ÖZKALE²</i> ,	19
GOMPERTZ-EXPONENTIAL DISTRIBUTION: RECORD VALUE THEORY AND APPLICATIONS IN RELIABILITY	
<i>Shakila Bashir¹, Ahmad Mahmood Qureshi²</i> ,	21
ON THE ORDER STATISTICS OF DEPENDENT RANDOM VARIABLES CONSTRUCTED FROM BIVARIATE RANDOM SEQUENCES	
<i>Ismihan BAYRAMOGLU¹, Omer L. GEBIZLIOGLU²</i> ,	22
ROBUST RANKED SET SAMPLING METHODS FOR ONE-SAMPLE T-TEST	
<i>Yusuf Can SEVİL¹, Tuğba YILDIZ²</i> ,	23
ON CLASSIFICATION WITH MULTIPLE BIRTH SUPPORT VECTOR MACHINES	
<i>Güvenç ARSLAN¹</i>	24
NEW MATHEMATICAL FORMULATIONS FOR THE DISTRIBUTED PERMUTATION FLOWSHOP SCHEDULING PROBLEM	
<i>Alper HAMZADAYI¹, Hanifi Okan İŞGÜDER²</i> ,	25
HANDLING MISSING VALUES IN RANDOM FORESTS: AN APPLICATION TO DEMOGRAPHIC SURVEY DATA	
<i>Duygu İÇEN¹ – Ayşe ABBASOĞLU ÖZGÖREN² – Anıl BOZ SEMERCİ³</i>	26
TAIL DEPENDENCE ESTIMATION BASED ON ESTIMATION OF KENDALL DISTRIBUTION FUNCTION VIA RATIONAL BERNSTEIN POLYNOMIALS	
<i>Mahmut Sami ERDOĞAN¹, Selim Orhun SUSAM²</i> ,	27
POT METHOD FOR RUIN PROBABILITY IN INFINITE TIME WITH NON-STATIONARY ARRIVALS AND HEAVY TAILED DISTRIBUTION CLAIMS FOR LOSS MODELS	
<i>Frihi, Redhouane¹, Rassoul, Abdelaziz²</i>	28
STATISTICAL COMPARISON OF MODELLING APPROACHES DEMONSTRATED FOR BIOMEDICAL NETWORKS	
<i>Rza BASHIROV¹</i>	29
BELL MARGINAL MODELS FOR LONGITUDINAL COUNT OUTCOMES	
<i>Hatice Tul Kubra AKDUR¹</i>	31
THE INTERPLAY BETWEEN DETERMINISM, STOCHASTICITY AND FUZZYNESS ILLUSTRATED FOR P16-MEDIATED PATHWAY	
<i>Nimet İlke AKÇAY¹, Rza BASHIROV²</i> ,	32
MODELLING PRE-SERVICE MATHEMATICS TEACHERS REASONING UNDER UNCERTAINTY IN THE EGYPTIAN CONTEXT	

<i>Samah Gamal Ahmed ELBEHARY</i>	33
HYPOGEOMETRIC DISTRIBUTION AND RELATED DISCRETE TIME POINT PROCESS	
<i>Stefanka CHUKOVA¹, Leda MINKOVA, Silvana PARALLOJ²</i>	35
DIFFERENT SIMILARITY MEASURES FOR THE CLUSTERING OF TIME SERIES	
<i>Yamina KHEMAL-BENCHEIKH¹, Assia BOUIZANE²</i>	36
A METHOD FOR CONSTRUCTING AND INTERPRETING SOME WEIGHTED PREMIUM PRINCIPLES	
<i>Antonia CASTAÑO-MARTÍNEZ¹, Fernando LÓPEZ-BLAZQUEZ², Gema PIGUEIRAS³, Miguel A. SORDO⁴</i>	37
A BAYESIAN MODEL OF COVID19 NEW CASES	
<i>Marta SANCHEZ-SANCHEZ¹, Alfonso Suárez-Llorens², Ángel Berihuete³</i>	38
A SURVEY OF GROUNDWATER QUALITY IN SUBURB OF ULAANBAATAR CITY, MONGOLIA HYDROCHEMICAL INVESTIGATION OF GROUNDWATER IN 14 TH KHOROO OF KHAN-UUL DISTRICT USING MULTIVARIATE STATISTICAL TECHNIQUES	
<i>Dagvasuren GANBOLD¹, Gerelt-Od DASHDONDOG², Munkhtsetseg ZORIGT³, Enkhbayar JAMSRANJAV³</i>	39
WIDE BAND NOISES: THEORY & APPLICATIONS.....	
<i>A.E. Bashirov</i>	41
EVALUATING THE EFFECTS OF OUTLIERS IN BOOTSTRAP	
<i>Uğur BİNZAT¹, Engin YILDIZTEPE²</i> ,	42
AN ALTERNATIVE P CHART FOR MONITORING HIGH-QUALITY PROCESSES BASED ON IMPROVED ESTIMATOR	
<i>Senem SAHAN VAHAPLAR¹, Ozlem EGE ORUC²</i> ,	43
ESTIMATING THE GINI INDEX FOR INCOME LOSS DISTRIBUTIONS UNDER RANDOM CENSORING	
<i>Bari Amina, RASSOUL Abdelaziz, Ould Rouis Hamid</i>	44
FUSION OF GEOMETRIC AND TEXTURE FEATURES FOR SIDE VIEW FACE RECOGNITION.....	45
Salman Mohammed JIDDAH 1 , Main ABUSHAKRA 2 , Kamil YURTKAN3	

POLYA-AEPPLI PROCESS OF ORDER K OF SECOND KIND WITH AN APPLICATION

*Stefanka CHUKOVA*¹, *Meglana LAZAROVA*², *Leda MINKOVA*³

¹School of Mathematics and Statistics, Victoria University of Wellington, New Zealand,
e-mail: stefanka.chukova@vuw.ac.nz

²Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Bulgaria
e-mail: mevlana.laz@tu-sofia.bg

³Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Bulgaria
e-mail: leda@fmi.uni-sofia.bg

Abstract

In this paper we propose and study the so called Polya-Aeppli process of order k of second kind. Firstly the process is defined using the probability generating function followed by its definition as a birth process. The distribution of the related counting process is presented by recursion formulae. The Polya-Aeppli process of order k of second kind is considered within the framework of the risk process and corresponding probability of ruin is studied. Using a simulation some interesting results for the probability of ruin are obtained. Also a comparison between the Polya-Aeppli process of order k and Polya-Aeppli process of order k of second kind is discussed.

Key Words: *Polya-Aeppli distribution; distributions of order k ; compound distributions; ruin probability.*

References

- [1] Aki S, Kuboku H, Hirano K (1984), On a discrete distributions of order k , Ann. Inst. Statist. Math. 36, Part A, 431-440.
- [2] Balakrishnan N, Kourtas M (2002), Runs and Scans with Applications, Wiley Series in Probability and Statistics
- [3] Charalambides Ch. (1986), On discrete distributions of order k , Ann. Inst. Statist. Math., 38, Part A, 557-568.
- [4] Chukova S, Minkova L (2015), Polya-Aeppli of order k Risk Model, Commun. Statist – Simulation and Computation, 44, 551-564.

Acknowledgement

The second and the third authors are totally supported by Grant DN 12/11/20 Dec. 2017 of the Ministry of Education and Science of Bulgaria.

A Statistical Consistent Test Based on Bivariate Random Thresholds

Aysegul EREM¹, Ismihan BAYRAMOGLU²,

¹Department of Basic Sciences and Humanities, Faculty of Arts and Sciences, Cyprus International University, Lefkosa, Mersin 10, Turkey (aysegulerem@gmail.com)

²Department of Mathematics, Izmir University of Economics, Izmir, Turkey (ismihan.bayramoglu@ieu.edu.tr)

Let $Z_1 = \{(X_k^{(1)}, Y_k^{(1)}), k = 1, 2, \dots, n\}$ be a sequence of independent random variables with joint cumulative distribution function (cdf) $F(x, y) = C_1(F_X(x), F_Y(y))$, where $C_1(F_X(x), F_Y(y))$, $(u, v) \in [0, 1]^2$ is a connecting copula and $F_X(x)$ and $F_Y(y)$ are the marginal cdf's of X and Y , respectively. Furthermore, let $Z_2 = \{(X_k^{(2)}, Y_k^{(2)}), k = 1, 2, \dots, n\}$ be another sequence of independent random variables with joint cdf $G(x, y) = C_2(F_X(x), F_Y(y))$, where $C_2(F_X(x), F_Y(y))$, $(u, v) \in [0, 1]^2$ is a connecting copula and $F_X(x)$ and $F_Y(y)$ are the marginal cdf's of X and Y , respectively. Let $f(x, y) = \frac{\delta^2 F(x, y)}{\delta x \delta y}$, $g(x, y) = \frac{\delta^2 G(x, y)}{\delta x \delta y}$, $f_X(x) = \frac{dF_X(x)}{dx}$ and $f_Y(y) = \frac{dF_Y(y)}{dy}$. We assume that Z_1 and Z_2 are independent and we call them the training samples from populations with joint cdf's of F and G , respectively.

Let $Z = \{(X_k, Y_k), k = 1, 2, \dots, n\}$ be a sequence of independent random variables with joint cdf $H(x, y) = C(F_X(x), F_Y(y))$, where $C(F_X(x), F_Y(y))$, $(u, v) \in [0, 1]^2$ is a connecting copula. Recall Z_1 and Z_2 are training samples and Z is control sample.

We propose a consistent test for testing the distribution of control sample Z , based on bivariate order statistics of training samples. The probability of type I, type II errors and probability of making no decisions under null and alternative hypotheses are calculated. The consistency of the proposed test is discussed under particular conditions. Furthermore, an unbiased and consistent estimator is proposed for probability of making no decision. Moreover a simulation study is performed for showing the consistency of the test for some well-known copulas such as independent, Clayton, Gumbel, Frank and Farlie-Gumbel-Morgenstein copulas.

Key Words: *Bivariate two sample problem; bivariate order statistics; copula; hypothesis test; consistent tests*

References

- [1] Barakat, H. M. (1999). On moments of bivariate order statistics. *Annals of the Institute of Statistical Mathematics*, 51(2), 351-358.
- [2] Bairamov, I. G. and Petunin, Y. I. (1991). Statistical tests based on training samples. *Cybernetics and Systems Analysis*, 27(3), 408-413.
- [3] Erem, A. (2019). Bivariate two sample test based on exceedance statistics. *Communications in Statistics - Simulation and Computation*, DOI:10.1080/03610918.2018.1520868.
- [4] Kemalbay, G. and Bayramoglu, I. (2015). Joint distribution of new sample rank of bivariate order statistics. *Journal of Applied Statistics*, 42(10), 2280-2289.
- [5] Stoimenova, E. (2011). The power of exceedance-type tests under lehmann alternatives. *Communications in Statistics-Theory and Methods*, 40(4), 731-744.
- [6] Stoimenova, E. and Balakrishnan, N. (2011). A class of exceedance-type statistics for the two-sample problem. *Journal of Statistical Planning and Inference*, 141(9), 3244-3255.

PERFORMANCE ANALYSIS OF RIDGE DEVIANCE CONTROL CHARTS FOR MONITORING POISSON PROFILES

Ulduz MAMMADOVA¹, M. Revan ÖZKALE²,

¹Author's contact address and e-mail: ulduzozel@gmail.com

²Author's contact address and e-mail: mrevan@cu.edu.tr

Abstract

Multicollinearity is one of the common problems in industrial datasets caused by the absence of the independency among predictors. This problem can be solved by using biased estimators like ridge estimator. Since the value of the ridge estimator depends on the biasing constant, several number of techniques were studied in the literature to find a better way of calculating the biasing constant.

In this study, we constructed Shewhart, CUSUM, and EWMA control charts based on deviance residual and compared each of them to the corresponding control chart based on the ridge deviance residual. We used some of the existing techniques proposed by [1], [2], and [3] to obtain biasing constant and calculated iterative ridge estimator for monitoring Poisson profiles under multicollinearity. The performance of each control chart is measured by using the average run length criterion and compared through a simulation study. The simulations are performed under different scenarios where different types and sizes of shifts, degrees of multicollinearity, sample size, and the number of predictors are considered.

The inspection of the simulation results revealed that the proposed ridge deviance-based control charts show better performance than the deviance-based control charts when the sample size is large. When sample size is small, the ridge deviance-based control charts are not efficient as in case where sample size is large. Also, the performances of the control charts are not affected by the collinearity level among the predictors.

Key Words: *profile monitoring; deviance control charts; Poisson regression; ridge estimators; biasing constant.*

References

- [1] Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge regression: some simulations. *Communications in Statistics Theory and Methods* 4 105-123.
- [2] Kibria BMG (2003) Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation* 32 (2) 419-435.
- [3] Alkhamisi MA, Shukur G (2007) A Monte Carlo study of recent ridge parameters. *Communications in Statistics-Simulation and computation* 36 535-547.
- [4] Agresti A, Amiri A, Niaki STA (2014) A new link function in generalized linear model-based control charts to improve monitoring of two-stage processes with Poisson response. *International Journal of Advanced Manufacturing Technology* 72 1243-1256.
- [5] Amiri A, Koosha M, Azhdari A (2011) Profile monitoring for Poisson responses. *Proceedings of the Industrial Engineering and Engineering Management Conference* 1481-1484.
- [6] Amiri A, Koosha M, Wang G (2015) Phase I monitoring of generalized linear model-based regression profiles. *Journal of Statistical Computation and Simulation* 85 2839-2859.

Gompertz-Exponential Distribution: Record Value Theory and Applications in Reliability

Shakila Bashir¹, Ahmad Mahmood Qureshi²

¹Department of Statistics Forman Christian College (a Chartered University) Lahore, Pakistan

E-mail: shakilabashir@fccollege.edu.pk

²Department of Mathematics Forman Christian College (a Chartered University) Lahore, Pakistan

E-mail: mahmoodqureshi@fccollege.edu.pk

Abstract

The continuous probability distributions have great importance in the field of transportation (they are used to estimate how funds can be allocated to improve roads, railways, bridges, waterways, airports etc.), reliability engineering (to check the reliability/performance of a product or even to check the reliability of a system, failure chances etc.), the newly derive model is applied on the data sets : failure and service times for a particular model aircraft windshield; strengths of 1.5cm glass fibres. The Gompertz exponential (GoE) distribution is derived using Gompertz G generator. Various statistical properties of the model derived and discussed in details. The parameters of the model are estimated using maximum likelihood estimation method. The upper record values from the GoE distribution have also been introduced with various properties. Moreover, applications of the GoE distributions has been provided in the field of reliability to check the performance of some transportation related parts and the suggested model provides better fit than the existing well-known models. Finally, a simulation study is carried out. Random numbers of size 50 are generated 15 times for GoE distribution and upper records has been noted.

Key Words: *Gompertz family of distributions; exponential; MLE; GoE; reliability*

On the Order Statistics of Dependent Random Variables Constructed from Bivariate Random Sequences

Ismihan BAYRAMOGLU¹, Omer L. GEBIZLIOGLU²,

¹Department of Mathematics, Izmir University of Economics, Izmir, Turkey

²Department of International Trade and Finance Kadir Has University, Istanbul, Turkey

Abstract

We consider new model of random variables constructed on the base of bivariate random sequences and study the distributions of their order statistics. The method of derivations of distributions are based on the model of bivariate binomial distribution. This model allows some interesting extensions that are important in reliability analysis and actuarial sciences. Some examples for special distributions are provided and the application to complex system of n components with two subcomponents per component is described. The reliability function and the mean residual life function are also studied.

References

- [1] Aitken, A.C. and Gonin, H.T. (1935) On fourfold sampling with and without replacement, Proc. Roy. Soc. Edinburgh 55, 114--125, 1935.
- [2] Bairamov, I. and Elmastas Gultekin, O. (2010) Discrete distributions connected with bivariate binomial. Hacettepe Journal of Mathematics and Statistics. Volume 39 (1), 109 -120.
- [3] Bairamov, I., Ahsanullah, M. and Akhundov, I. (2002) A residual life function of a system having parallel or series structures. Journal of Statistical Theory and Applications.1(2):119-32.
- [4] Bairamov,I. and Arnold, B.C. (2008) On the residual life lengths of the remaining components in an $(n-k+1)$ - out-of- n system. Statistics and Probability Letters. 78(8):945-52.
- [5] Barlow, R. E. and Proschan, F. (1975) Statistical theory of reliability and lifetesting.New York:Holt, Rinehart & Winston.
- [6] Bayramoglu, I. and Kemalbay G. (2013) Some novel discrete distributions under fourfold sampling schemes and conditional bivariate order statistics. Journal of Computational and Applied Mathematics. 248, 1-14.

Robust Ranked Set Sampling Methods for One-Sample t-test

Yusuf Can SEVİL¹, Tuğba YILDIZ²,

¹Department of Statistics; Dokuz Eylül University; İzmir, Turkey e-mail: yusuf.sevil@ogr.deu.edu.tr

² Department of Statistics; Dokuz Eylül University; İzmir, Turkey e-mail: tugba.ozkal@deu.edu.tr

Abstract

Ranked set sampling (RSS) was proposed by McIntyre [1] as an alternative method to SRS. Also, modified versions of RSS have been discussed in the literature. Among these modified versions, L-RSS [2], truncation-based RSS [3] and robust extreme RSS [4] are robust methods against outliers.

One-sample t-test is the most popular and commonly used procedure to test whether a population mean is significantly different from some hypothesized value. However, it is well known that the one-sample t-test is very sensitive to outliers.

In this study, we investigated the performance of the one-sample t-test based on different ranked set sampling schemes included robust methods in terms of both Type-I error probability and power.

Key Words: *Robust ranked set sampling; one-sample t-test; outliers; type-I error probability; power*

References

- [1] McIntyre GA (1952) A method for selective sampling, using ranked sets. Australian Journal of Agricultural Research, 3:395-390.
- [2] Al-Nasser AD (2007) L ranked set sampling: A generalization procedure for robust visual sampling. Communications in Statistics-Simulation and Computation 36:33-43.
- [3] Al-Omari AI, Raqab MZ (2013) Estimation of the population mean and median using truncation-based ranked set samples. Journal of Statistical Computation and Simulation 83:1453-1471.
- [4] Al-Nasser AD, Mustafa AB (2009) Robust extreme ranked set sampling. Journal of Statistical Computation and Simulation 79:859-867.

ON CLASSIFICATION WITH MULTIPLE BIRTH SUPPORT VECTOR MACHINES

Güvenç ARSLAN¹

¹Kırıkkale University, guvenc.arslan@gmail.com

Abstract

Support vector machines are used with success in many different areas and applications. Thus, one may find many new variants of this learning algorithm. One of these recent variants is called the Multiple Birth Support Vector Machine, which was introduced by Yang et al. in 2013 [1]. An important difference of Multiple Birth Support Vector Machines is that non-parallel hyper-planes are used with separate optimization problems for each class category. In this study, we investigate a novel approach for classification that uses Multiple Birth Support Vector Machines. A novel clustering algorithm is used as an intermediate step before applying the Multiple Birth Support Vector Machines. The first results show that the proposed approach may be useful in developing new classification algorithms, which are expected also to be effective for big data applications.

***Key Words:** Classification; support vector machines; multiple birth; clustering*

References

[1] Yang Z-X, Shao Y-H, Zhang X-S (2013) Neural Comput & Applic 22 (Suppl 1):153-161. (Times New Roman, 12pt, Maximum 6 references)

NEW MATHEMATICAL FORMULATIONS FOR THE DISTRIBUTED PERMUTATION FLOWSHOP SCHEDULING PROBLEM

Alper HAMZADAYI¹, Hanifi Okan İŞGÜDER²,

¹Department of Industrial Engineering, Van Yüzüncü Yıl University, 65080, Van, Turkey

E-mail address: alperhamzadayi@yyu.edu.tr

²Department of Statistics, Dokuz Eylül University, 35210, Izmir, Turkey

E-mail address: okan.isguder@deu.edu.tr

Abstract

The distributed permutation flowshop scheduling problem (DPFSP) has recently occurred as a generalization of the regular flowshop scheduling problem where several factories are available for processing the jobs. The DPFSP dealing with real-life applications has attracted attention of the researchers for almost a decade. In the current literature, the studies carried out on this problem have been generally intended for developing approximation algorithms. In this paper, two new mathematical models are developed by drawing inspiration from the formulations developed for the multiple-travelling salesman problem (mTSP). Two newly developed mathematical models are compared in detail with each other and the best mathematical models given by Naderi and Ruiz [1] by using the 84 data sets available in the literature. The results of the experiment have revealed that the new mathematical models have outperformed substantially when compared to the other methods.

Key Words: *Distributed flowshop problem, Mixed integer linear programming*

References

[1] Naderi B, Ruiz R (2010). The distributed permutation flowshop scheduling problem. *Computers and Operations Research*, 37, 754–768.

HANDLING MISSING VALUES IN RANDOM FORESTS: AN APPLICATION TO DEMOGRAPHIC SURVEY DATA

Duygu İÇEN¹ – Ayşe ABBASOĞLU ÖZGÖREN² – Anıl BOZ SEMERCI³

¹ Hacettepe University, Department of Statistics, duyguicn@hacettepe.edu.tr

² Hacettepe University Institute of Population Studies, Department of Demography, ayseabs@gmail.com

³ Hacettepe University, Department of Business Administration, annilboz@gmail.com

Abstract

The purpose of this study is to examine how missing values should be handled when a classification is made with random forest algorithm to the most recent Turkey Demographic and Health Survey (2018 TDHS) data. The main idea of ensemble learning methods is to create a better model, each solving the same problem, with more accurate and reliable predictions or decisions than using a single model [1]. As being one of the ensemble methods, Random Forest (RF) is developed by Leo Breiman in 2001 and has been increasingly used in the field of data science since then [2]. Some important advantages of the random forest method are it handles a large number of input variables also it is speediness [3]. The inevitable problem of the data scientist is that s/he faces missing values in almost all areas of science. We first focus on the Turkey Demographic and Health Survey (2018 TDHS) data that has some missing values. We use different imputation methods for the missing values of this data [4]. Finally, the best imputation method for 2018 TDHS data is determined in the classification problem using Random forests.

Key Words: *Random Forest, Missing Values, 2018 TDHS Data*

References

- [1] Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning Data Mining Inference and prediction. Springer.
- [2] Hapfelmeier, A (2012) Analysis of missing data with random forests, “http://edoc.ub.uni-muenchen.de/15058/1/Hapfelmeier_Alexander.pdf”, (2020-08-05).
- [3] Tang, F, Ishwaran, H (2017) Random forest missing data algorithms. Stat. Anal. Data Min.10, 363–377.
- [4] R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>.

TAIL DEPENDENCE ESTIMATION BASED ON ESTIMATION OF KENDALL DISTRIBUTION FUNCTION VIA RATIONAL BERNSTEIN POLYNOMIALS

Mahmut Sami ERDOĞAN¹, Selim Orhun SUSAM²,

¹Department of Statistics, Istanbul Medeniyet University and e-mail: sami.erdogan@medeniyet.edu.tr

²Department of Econometrics, Munzur University and e-mail: orhunsusam@munzur.edu.tr

Abstract

This study introduces the estimation of Kendall distribution function via rational Bernstein polynomials as an alternative to the methods used in the literature. The proposed estimation method here has many advantages such as providing better control of the shape of the curve. The estimation of the lower and upper tail dependence is focused on using the new estimator with rational Bernstein polynomials. A designed Monte Carlo study is used to measure the performance of the new method. The simulation study indicates that the proposed estimator here is preferable according to ASE performance. Moreover, as a special case, the rational estimation method reduces to Bernstein polynomial estimation method where all the weights are equal. The new estimator is performed to estimate the dependence coefficient of real data set as example.

Key Words: *Copula; Kendall function; Archimedean copula; rational Bernstein polynomials; tail dependence.*

POT Method for Ruin Probability in Infinite time with Non-Stationary Arrivals and Heavy Tailed Distribution Claims for loss models

Frihi, Redhouane¹., Rassoul, Abdelaziz²,

Blida 1 University, BP 270 road soumaa, Blida, Algeria¹.
National Higher School of Hydraulic, Guerouaou, BP 31, Blida,
09000, Algeria².

¹rfrihi@yahoo.fr

²r.rassoul@ensh.dz

Abstract

In this paper, we obtain infinite-horizon ruin probability asymptotics for risk processes with claims of Heavy Tailed Distribution for non-stationary arrivals processes. For this purpose, we introduce the Peak Over threshold (POT) method with large initial reserves and infinite variance in infinite time. We give some examples of the ruin probability for non-stationary arrivals processes, like Hawkes process and Cox process. Our approach is based on the result of Balkema et al.(1974) Klppelberg et al. (1996), Assmusen (2000) and Zhu (2013).

Key Words: *Extremes values; Heavy-tailed distribution; Generalised Pareto Distribution (GPD); Ruin probability; Peak Over Threshold method.*

References

- [1] Asmussen, S.(2000). Ruin Probabilities. World Scientific, Singapore, New Jersey, London, Hong Kong .
- [2] Balkema, A. de Haan, L. (1974). Residual Life Time at Great Age.The Annals of Probability.
- [3] De Haan L and Ferreira A. Extreme value theory: an introduction. 1st ed. New York: Springer Science+ Business Media, 2007, p.141.
- [4] Caeiro,F and M. Ivette Gomes. Threshold Selection in Extreme Value. Analysis Extreme Value Modeling and Risk Analysis: Methods and Applications,2015, p.69-82.
- [5] Klüppelberg, C. and Stadtmüller, U. (1996), Ruin Probabilities in the Presence of Heavy-Tails and Interest Rates,SCAND. ACTUAR. J, 49,49-58.
- [6] Scarrott C and MacDonald A. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT: Stat J 2012; 10(1): 33-60
- [7] Zhu, L. Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. Insurance: Mathematics and Economics 53 (2013) 544–550.

STATISTICAL COMPARISON OF MODELLING APPROACHES DEMONSTRATED FOR BIOMEDICAL NETWORKS

Rza BASHIROV¹

¹Department of Mathematics, Faculty of Arts and Sciences, Eastern Mediterranean University
Famagusta, North Cyprus, Mersin-10, Turkey
e-mail: rza.bashirov@emu.edu.tr

Abstract

Randomness and uncertainty are two major problems one faces while modeling non-linear dynamics of biomolecular systems. Randomness in biology arises in the form of stochastic timing of molecular events induced by spontaneous fluctuations of small number molecules, unavoidable noise in an experiment or intrinsic noise [1-3]. Kinetic parameters are, on the other hand, uncertain due to incomplete knowledge, which is usually expressed by qualitative expressions such as “is almost disrupted” or “is faster than” [4]. Measurement error and natural variation are among factors leading to uncertainty in biomolecular systems. Deterministic, stochastic and fuzzy approaches are used for modelling complex biomolecular networks, but there is no consensus among researchers regarding which approach can be used when. Given a biomolecular network, it is a cumbersome task to decide which modelling approach is the optimal, providing the closest approximation of the underlying biomolecular system.

The present work explores fuzzy stochastic Petri nets to cope with random timing of molecular events and deal with uncertainty of reaction rates. The approach is demonstrated through a case study of simulation-based prediction of optimal drug combinations for Spinal Muscular Atrophy [4]. We develop hybrid model of underlying biological network, comprising both continuous and discrete components, validate the model through existing empirical data and perform simulations to predict optimal drug combinations. Based on simulation results we identify drug combinations holding promise for the potential beneficial therapeutic effects on the Spinal Muscular Atrophy patients. To explore the strength of the approach and establish its feasibility we use SPSS Statistics Software Package and perform statistical analysis of the simulation results obtained for deterministic, pure stochastic and fuzzy stochastic models. Statistical analysis based on Friedman test and Wilcoxon Signed Rank Test reveals that the three approaches lead to substantially different results, though for some drug combinations they behave similarly. This fact clearly indicates that fuzzy stochastic modelling is superior to the deterministic and pure stochastic approaches for the proposed case study, while allows for the creation of the closest approximation of the underlying biological network through dealing with both randomness and uncertainty.

Key Words: *Statistical analysis; Fuzzy stochastic Petri net; Quantitative modelling; Spinal Muscular Atrophy*

References

- [1] Bashirov R, Akçay Nİ (2018) Stochastic simulation-based prediction of the behavior of the p16-mediated signaling pathway. *Fundamenta Informaticae* 160:167-179.
- [2] Tüysüz F, Kahraman C (2010) Modeling a flexible manufacturing cell using stochastic Petri nets with fuzzy parameters. *Expert Systems with Applications* 37(5):3910-3920.
- [3] Liu F, Heiner M, Yang M (2016) Fuzzy stochastic Petri nets for modelling biological systems with uncertain kinetic parameters. *PLoS ONE* 11(2):e0149674. doi:10.1371/journal.pone.0149674.
- [4] Bashirov R, Duranay R, Şeytanoğlu A, Mehraei M, Akçay Nİ (2019) *Turkish Journal of Electrical Engineering and Computer Sciences* 27:4009-4022.

Bell Marginal Models for Longitudinal Count Outcomes

Hatice Tul Kubra AKDUR¹

¹Department of Statistics, Faculty of Science, Gazi University, Ankara, TURKEY.
e-mail: haticesenol@gazi.edu.tr

Abstract

Correlated count responses are usually observed in clinical, economical and biological researches. The main assumption is to assume that a dependency structure exists between observations in the same experimental unit or cluster while no dependency exists between observation from different experimental units. Longitudinal count models are generally modelled through the use of generalized estimating equations firstly introduced by Liang and Zeger (1986) [1] (GEEs). Popular count marginal models are usually based on poisson and negative binomial distribution. In this study, a new marginal model is introduced and developed for longitudinal count responses based on bell distribution [2]. Bell distribution and its related regression model have been recently proposed by Castellares et al. [2] for count dataset. Although the bell distribution does not contain a dispersion parameter, it can model overdispersion. It indicates that this is more practical and useful than the negative binomial distribution. Real data application is presented to illustrate the new marginal model. The working covariance model selection method the “quasi-likelihood under the independence model criterion” (QIC) is utilized for the application [3]. The parameter estimations of the bell marginal model based on GEEs are obtained by geeM R package [4]. Some diagnostic measures are also provided for the bell marginal model.

Key Words: *Bell distribution; Generalized Estimating Equations; count outcomes; quasi-likelihood; correlation.*

References

- [1] Liang KY, Zeger, SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.
- [2] Castellares F, Ferrari SL, Lemonte AJ. (2018). On the Bell distribution and its associated regression model for count data. *Applied Mathematical Modelling*, 56, 172-185.
- [3] Pan W (2001). Akaike’s information criterion in generalized estimating equations. *Biometrics*, 57:120–125.
- [4] McDaniel LS, Henderson NC, Rathouz PJ (2013). Fast pure R implementation of GEE: application of the matrix package. *The R journal*, 5(1), 181.

THE INTERPLAY BETWEEN DETERMINISM, STOCHASTICITY AND FUZZYNESS ILLUSTRATED FOR P16-MEDIATED PATHWAY

Nimet İlke AKÇAY¹, Rza BASHIROV²,

¹Faculty of Medicine, Eastern Mediterranean University, Famagusta, North Cyprus, Mersin-10, Turkey

²Department of Applied Mathematics and Computer Science, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, North Cyprus, Mersin-10, Turkey

Abstract

Quantitative modeling of complex biological networks at the molecular level is a challenging task because such systems are inherently random and contain uncertain kinetic parameters. Stochastic and fuzzy models are used to deal with the randomness and the vagueness in biological systems [1,2]. On the other hand, the level of stochasticity in a biological system depends on molecular density such that a system becomes more deterministic and less stochastic with increase of the number of molecules. It is rather hard, therefore, to select an optimal modelling approach by balancing between deterministic, stochastic and fuzzy behavior of the system. Neither there exists a common voice among researchers regarding which mechanism can be used for determining the best approach for biological networks.

In the present work, we propose a novel approach for selection of the best modelling framework, which is based upon statistical comparison of the deterministic, pure stochastic and fuzzy stochastic models. The approach is exemplified through a case study of p16-mediated pathway. We create Hybrid Petri net model of underlying biological network, validate the model, perform deterministic [3], stochastic [4] and fuzzy stochastic simulations and use Kruskal-Wallis tests to statistically compare the simulations via data sets collected from three modelling approaches. Statistical analysis reveals that there is a significant difference between deterministic, pure stochastic, and fuzzy stochastic simulation data sets, and that pure stochastic and fuzzy stochastic approaches are the optimal modelling choices for the case study, describing best the behavior of the underlying biological network. The proposed approach can be easily adapted to other biomolecular networks.

Key Words: *Statistical analysis; Quantitative modelling; Stochastic simulation; Fuzzy logic; p16-mediated pathway*

References

- [1]. Bashirov R, Duranay R, ŞEYTANOĞLU A, Mehraei M, Akçay N. Exploiting stochastic Petri nets with fuzzy parameters to predict efficient drug combinations for Spinal Muscular Atrophy. *Turkish J Electr Eng Comput Sci.* 2019;27(5):4009–22.
- [2]. Liu F, Heiner M, Yang M. Fuzzy Stochastic Petri Nets for Modeling Biological Systems with Uncertain Kinetic Parameters. 2016;1–19.
- [3]. Akçay NI, Bashirov R, Tüzmen Ş. Validation of signalling pathways: Case study of the p16-mediated pathway. *J Bioinform Comput Biol.* 2015;13(2).
- [4]. Bashirov R, Akçay NI. Stochastic Simulation-based Prediction of the Behavior of the p16-mediated Signaling Pathway. *Fundam Informaticae.* 2018;160(1–2):167–79.

Modelling Pre-service Mathematics Teachers Reasoning Under Uncertainty in the Egyptian Context

Samah Gamal Ahmed ELBEHARY

Assistant lecturer at Faculty of Education, Tanta University, Egypt

Ph.D. Students at Graduate School for International Development and Cooperation, Hiroshima University, Japan

dr.samahelbe7ary@yahoo.com dr.samahelbe7ary@gmail.com

Abstract

Probability involves substantial knowledge that is often manipulated in our daily life, needed in everyday settings for all citizens in decision-making situations. It, too, helps to form a specific type of reasoning of which we can formally structure our vague thought regarding random phenomena and overcome our deterministic thinking. Thus, it has been recognized by educational authorities in many countries and included in their official curricula. Despite that, several issues were declared regarding deficiencies of probability education. One such problem that is concerned here is the inadequate preparation of mathematics teachers to teach probability. Accordingly, this investigation aimed at exploring: how do pre-service mathematics teachers (PSMTs) reason under uncertainty? That may give insights towards an efficient pedagogical preparation that matches their thinking trajectories. Thus, to approach such an area, three different probabilistic contexts of giving birth, throwing a die, and weather predictability were adapted; they meet PSMTs, school curriculum, and pupils' viewpoints of the most commonly utilized setting of probability, respectively. Later, forty-eight PSMTs were asked to determine the probability of (a) giving birth to a girl, (b) getting number 5 in an experiment of throwing a die; besides, (c) interpret the meaning of a 60% chance of rain. They were selected in light of their availability and willingness to participate; and (b) prior knowledge of theoretical, experimental, and conditional probability concepts. Hence, their numerical answers and given argumentations were analyzed inductively through NVivo software and following Thomas's steps [1]. As a result, four major categories that express PSMTs reasoning under uncertainty were inferred and coded under the terms of Mathematically (**M**), Subjectively (**S**), Outcome, and Intuitively oriented thinkers (**I**). Furthermore, the distribution of such manners of reasoning has altered depending upon the context. While in the context of giving birth **M**, **S**, and **O** were distributed by a percentage of 29.4 %, 60.3 %, and 10.3 %, respectively, type **S** reasoning hasn't emerged in the context of throwing a die of which PSMTs reasoning diverged between **M** and **O** by a percentage of 73.5% and 26.5%, sequentially. Besides, type **I** reasoning has arisen only in the context of weather predictability by a percentage of 25% compared to 14.6%, 10.4%, and 50% for **M**, **S**, and **O**, respectively. Beyond that, the investigation revealed several cognitive biases that are shared by PSMTs and negatively affect their reasoning. For example, **M** thinkers shared the *equiprobable bias* of which they thought that all possible outcomes of the sample space are equally likely to occur [2], [3]; moreover, while some **S** thinkers yielded the *prediction bias* of which their prediction has the meaning of exact prediction [4], **O** thinkers confuse *causality and conditionality* [5], [6]. These findings reflect the value of admitting the contextual differences to select the appropriate probability interpretation that could model a phenomenon. Consequently, it may serve as a basis for teacher educators; it can help them to utilize relevant activities that may exhibit whether PSMTs reasoning processes are consistent with the normative theories or not.

Key Words: Reasoning under uncertainty; Pre-service mathematics teachers; Probability conceptions; Cognitive biases.

References

- [1] Thomas DR. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2):237-246.
- [2] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- [3] Lecoutre M. P., & Fischbein E. (1998). Evolution avec l'âge de "misconceptions" dans les intuitions probabilistes en France et en Israël [Evolution with age of probabilistic intuitions in France and Israel]. *Recherches en Didactique des Mathématiques*, 18, 311-332.
- [4] Savard, A. (2014). Developing probabilistic thinking: What about people's conceptions? In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives*. (Vol. 2, pp. 283-298). New York, NY: Springer Science and Business Media.
- [5] Borovnik, M. (2012). Multiple perspectives on the concept of conditional probability. University of Klagenfurt. *Avances de Investigacion en Education Matematica*, 2, 5-27.
- [6] Díaz, C., Batanero, C., & Contreras, J, M. (2010). Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa*, 26 (2), 149-162.

HYPOGEOMETRIC DISTRIBUTION AND RELATED DISCRETE TIME POINT PROCESS

Stefanka CHUKOVA¹, Leda MINKOVA, Silvana PARALLOJ²

¹School of Mathematics and Statistics, Victoria University of Wellington, New Zealand,
e-mail: stefanka.chukova@vuw.ac.nz

²Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Bulgaria
e-mail: sparalloi@yahoo.com, leda@fmi.uni-sofia.bg

Abstract

In this paper we propose and study a new distribution, called the hypogeometric distribution, which is a sum of independent geometrically distributed variables with different parameters. Also, we propose and study a discrete time point process based on this distribution. As an example, we focus on a particular form of this process. Also, we show that this type of processes could be used as an appropriate tool to model arrivals with increasing or decreasing time trends. Some possible extensions of this work are also included in the paper.

Key Words: *geometric distribution, hypogeometric distribution, waiting time, counting process*

Acknowledgement

The research of the Silvana Paralloj was partially supported by Grant 801043, 26.03.2020 of Sofia University.

DIFFERENT SIMILARITY MEASURES FOR THE CLUSTERING OF TIME SERIES

Yamina KHEMAL-BENCHEIKH¹, Assia BOUIZANE²

¹ Fundamental and Numerical Mathematics Laboratory LMFN
Department of Mathematics- Faculty of Sciences
UFA Setif1 El-Bez 19000 Algeria
bencheikh-00@univ-setif.dz

² Department of Mathematics- Faculty of Sciences
UFA Setif1 El-Bez 19000 Algeria
bouizaneassia1996@gmail.com

Abstract

The processing of sequential data represents an important part of the problems tackled in machine learning. Classical clustering methods are not always adapted to the characteristics of these temporal data (sequentiality, relationships between learning examples ...). Clustering aims to divide a set into different homogeneous groups, in the sense that the data of each subset share common characteristics, which most often correspond to proximity (or similarity) criteria which are defined by introducing measures of similarity between objects. The contribution presented in this work focuses on the choice of the “right” similarity measure because it defines the level of similarity between objects. The paper begins with a description of the different approaches to clustering. A review of the different similarity measures treated in the literature is also given. The IMs-DTW measurement was selected using the K-medoids algorithm to cluster the countries most affected by the Covid-19 Coronavirus pandemic. This study was completed by a hierarchical clustering for the purpose of method comparison.

Key Words: *Clustering ; Time Series ; Covid-19 Corona virus Pandemic.*

References

- [1] Aach, John and Church, G. M. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics*, 17(6) : 495-508.
- [2] Agrawal, R. Faloutsos, C and A. Swami, A. (1993). Efficient Similarity Search in Sequence Databases. In *Proceeding of the Fourth International Conference on Foundations of Data Organization and Algorithms*, Chicago. Also in *Lecture Notes in Computer Science 730*, Springer Verlag, 69-84.
- [3] Dilmi, M. D., Barthes, L., Mallet, C. (2018). Iterative Multi-Scale Dynamic Time Warping (IMs-DTW): Estimation de la dissimilarité entre des séries temporelles de précipitation et application à la classification des événements pluvieux». *Présentation SAMA 13 mars 2018*, Paris-Saclay.
- [4] Halkidi, M. Batistakis, Y and Vazirgiannis, M. (2001). On Clustering Validation Techniques». *Intelligent Information Systems Journal*, Kluwer Publishers, vol.17, n°2-3, pp. 107-145.
- [5] <https://coronavirus.politologue.com/> (2020)
- [6] <https://data.europa.eu/euodp/fr/data/dataset/covid-19-coronavirus-data> (2020)

A METHOD FOR CONSTRUCTING AND INTERPRETING SOME WEIGHTED PREMIUM PRINCIPLES

Antonia CASTAÑO-MARTÍNEZ¹, Fernando LÓPEZ-BLAZQUEZ², Gema PIGUEIRAS³, Miguel A. SORDO⁴

^{1,3,4} Department of Statistics and Operation Research, University of Cádiz: antonia.castano@uca.es; gema.pigueiras@uca.es; mangel.sordo@uca.es

² Department of Statistics and Operation Research, University of Sevilla: lopez@us.es

We present a method for constructing and interpreting weighted premium principles. The method is based on modifying the underlying risk distribution in such a way that the risk-adjusted expected value (or premium) is greater than the expected value of some conveniently chosen function of claims, which defines the insurer's perception of the risk. Under some assumptions on the function of claims, the method produces distortion premium principles. We provide several examples under different assumptions on the claim arrival process and different functions of claims, including record claims and k th record claims.

Key Words: *Weighted premium principle; distortion premium principle; equilibrium distribution; order statistics; record value*

A Bayesian model of COVID19 new cases

Marta SANCHEZ-SANCHEZ¹, Alfonso Suárez-Llorens², Ángel Berihuete³

Facultad de Ciencias, Universidad de Cádiz,
Avda. Rep. Saharaui s/n, 11510 Puerto Real, Cádiz, Spain
[1marta.sanchez@uca.es](mailto:marta.sanchez@uca.es), [2alfonso.suarez@uca.es](mailto:alfonso.suarez@uca.es), [3angel.berihuete@uca.es](mailto:angel.berihuete@uca.es)

Abstract

Given the expected rate of infections by using the Gompertz model and previous knowledges available about the coronavirus SARS-COV-2, we will obtain a prior distribution for the underlying parameters. The likelihood function will be based on a non-homogenous Poisson process by using the Gompertz curve. After that, we will obtain the Bayesian posterior model in order to forecast the number of new cases of COVID19 in near future time intervals from a Bayesian perspective. Finally, by using the official information given by the governments of different regions in Spain, we will show a particular example of forecasts in different times of the pandemic.

Key Words: *Bayesian inference, Gompertz curve, COVID19, nonhomogeneous Poisson process, prior distribution.*

A survey of groundwater quality in suburb of Ulaanbaatar city, MongoliaHydrochemical investigation of groundwater in 14th khoroo of Khan-Uul district using multivariate statistical techniques

*Dagvasuren GANBOLD*¹, *Gerelt-Od DASHDONDOG*², *Munkhtsetseg ZORIGT*³,
*Enkhbayar JAMSRANJAV*³

¹Mongolian Criminology Association, Mongolia, e-mail: gdagvasuren@gmail.com

² Division of Water Resources and Water Utilization Institute of Geography & Geoecology, Mongolian Academy of Sciences, Mongolia, e-mail: gereltodd@mas.ac.mn

³Mongolian Crimino

³Department of Applied Mathematics, SEAS, NUM, e-mail: munkhtsetseg.z@seas.num.edu.mn,
enkhbayar.j@seas.num.edu.mn

Abstract

Groundwater is the most important natural resource required for drinking to many people around the world, especially in rural areas. For 14th khoroo of Khan-Uul district of Ulaanbaatar is only one khoroo with a pronounced five species of livestock (horse, camel, cattle, sheep and goats) and agriculture plays important role in the region, faces an increasing rural population, a growing livestock and agriculture sector that increase in the demand for residential and agricultural water. This research was focused on evaluation of groundwater quality in that area, through physicochemical analysis with view to determining is suitability for drinking and irrigation purposes. During the sampling periods, a total of 51 groundwater samples were collected from 46 deep wells and 5 shallow wells of researchable area, between October and November 2019. The depths of the wells were varied from 22 to 102 m. Samples were analysed for complete physicochemical characterization of drinking water from groundwater sources in suburb of Ulaanbaatar. To ensure the suitability of groundwater in 14th khoroo for drinking purposes, the hydrochemical parameters were compared with the guideline recommended by the World Health Organization (WHO) and the National Standard (MNS 0900:2018). Results found that the trends of cations and anions are $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Na}^{++} > \text{K}^{+}$ and $\text{HCO}_3^{-} > \text{SO}_4^{2-} > \text{Cl}^{-}$ respectively, and $\text{Ca}^{2+} - \text{Mg}^{2+} - \text{HCO}_3^{-}$ is the dominant groundwater type. Parameters of drinking water quality were within the National Drinking Water Standard and WHO guideline values in groundwater samples except for calcium maximum measured value (116.2 mg/l), magnesium (54.1 mg/l), NO_3 (70.0 mg/l), TDS (691 mg/l, WHO (2011) guideline value). From regression and correlation analysis it was found that, EC, Ca and Mg were correlated with each other's; which might helpful for site specific monitoring of groundwater quality.

We proposed to identify that hydro-chemical characteristics and spatial distribution of the aquifer using the GIS and statistical approaches in the Ulziit area of the Ulaanbaatar, capital city of Mongolia. There 51 groundwater samples were gathered from the wells and analyzed cluster and component analysis. Samples were included 8 anions and cations (Na^{+} , K^{+} , Ca^{2+} , Mg^{2+} , Cl^{-} , SO_4^{-} , NO_3^{-} , HCO_3^{-}), EC and depth of the wells. As a result, dominated groundwater type is $\text{HCO}_3 - \text{CA}^{2+} - \text{Na}$ for the all samples. However, the first cluster is defined anions and cations are permissible level compared to other two clusters. The Mg^{2+} in water samples of the second cluster were slightly higher than the permissible level. For the third cluster, parameters of Mg ²⁺ and Ca^{2+} of permissible limit of the drinking water were not set. The findings of this study suggest that the need for the extended assessment with the coverage of all types of drinking water sources for drinking water supply in Ulaanbaatar.

Key Words: *Groundwater., Multivariate., Ulaanbaatar City., Multivariate.*

References

- [1] [Adhikary, P.P., Chandrasekharan, H., Chakraborty, D., Kumar, B., Yadv, B.R., \(2009\) Statistical approaches for hydrogeochemical characterization of ground water in West Delhi, India. Environ. Monit. Acces.](#)
- [2] [Ahmed, N., Bodrud-Doza, M., Islam, S.M.D.U., Choudry, M.A., Muhib, M.I., Zahid, A., Hossain, S., Moniruzzaman, M., Deb, N., Bhuiyan, M.A.Q Hydrogeochemical evaluation and statistical analysis of groundwater of Sylhet, north-eastern Bangladesh. Acta Geochim.](#)
- [3] [Bodrud-Doza, M., Islam, A.R.M.T., Ahmed, F., Das, S., Saha, N., Rahman, M.S., 2016. Characterization of groundwater quality using water evaluation indices, multivariate statistics and geostatistics in central Bangladesh. Water Sci. 30 \(1\), 19–40.](#)
- [4] [Mohan, R., Singh, A.K., Tripathi, J.K., Chowdhary, G.C., 2000. Hydrochemistry and quality assessment of groundwater in Naini industrial area, Allahabad district, Uttar Pradesh. J. Geol. Soc. India 55, 77–89.](#)
- [5] [Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists. Statistics in practice. Wiley, Chichester.](#)
- [6] [WHO, 2011. WHO guidelines for drinking-water quality, 4th ed. World Health Organization, Geneva.](#)

WIDE BAND NOISES: THEORY & APPLICATIONS

A.E. Bashirov

agamirza.bashirov@emu.edu.tr

Abstract

Kalman filter was discovered by Rudilf Kalman in 1960 for a discrete linear system disturbed by independent white noises. In 1961, it was modified to continuous time systems by Kalman and Bucy. Afterwards, an increasing demand to Kalman type filters in the areas such as signal processing, control systems, guidance, navigation, forecasting, robotics, GPS pushed to create its different modifications. In such a way, the extended and unscented Kalman filters for nonlinear systems, the hybrid Kalman filter for continuous signals with discrete observations, the Kalman filter for colored noises, the Kalman filter for infinite dimensional systems, etc. were created.

Wide band noise filter takes place among the modifications of the Kalman filter as well. The actuality of wide band noises were noticed by Fleming and Rishel. Presently, there are two approaches to wide band noises. Weare going to discuss one of these methods which is based on a distributed delay of white noises.

Evaluating the Effects of Outliers in Bootstrap

Uğur BİNZAT¹, Engin YILDIZTEPE²,

¹Department of Statistics, Dokuz Eylul University, Izmir, Turkey; ugur.binzat@deu.edu.tr

²Department of Statistics, Dokuz Eylul University, Izmir, Turkey; engin.yildiztepe@deu.edu.tr

Abstract

Bootstrap is a common method to derive statistical inferences using sample data repeatedly. Despite being a strong tool in statistics, bootstrap might not work well in the presence of outliers. While constructing bootstrap confidence intervals, bootstrap resamples may have more outliers than the original sample and this leads to tail problems [1]. Using bootstrap with robust estimators may be a solution, but in some cases removing observations from the original data is not preferred. In this case, the bootstrap process can be robustified [2]. Robust bootstrap is an active research area within the literature. In this study, the performance of bootstrap methods with data including outliers is evaluated via various simulation designs. The sample mean, correlation coefficient, and regression coefficients were estimated by bootstrap methods for random samples containing different proportions of outliers. In the study, contaminated normal distribution and manually added outliers were used to generate the samples with outliers. Symmetrical and asymmetrical cases were also evaluated. In the asymmetrical cases, MSE values obtained by the bootstrap estimates from samples containing outliers were found up to 350 times higher than cases without outliers according to the simulation results. All computations were performed in R statistical programming language.

Key Words: *Bootstrap; Outliers; Parameter estimation; Simulation*

References

- [1] Singh, K. (1998). Breakdown theory for bootstrap quantiles. *The Annals of Statistics*, 26(5), 1719-1732.
- [2] Amado, C., Bianco, A. M., Boente, G. L., & Pires, A. M. (2014). Robust bootstrap: an alternative to bootstrapping robust estimators. *REVSTAT*, 12(2), 169-197.

AN ALTERNATIVE P CHART FOR MONITORING HIGH-QUALITY PROCESSES BASED ON IMPROVED ESTIMATOR

Senem SAHAN VAHAPLAR¹, Ozlem EGE ORUC²,

¹Dokuz Eylul University, Department of Statistics, İzmir / Turkey, senem.sahan@deu.edu.tr

²Dokuz Eylul University, Department of Statistics, İzmir / Turkey, ozlem.ege@deu.edu.tr

Abstract

Recently, manufacturing processes with very low rejection rate are regularly observed in practice. This type of process is known as high-quality process, for which conventional Shewhart 3-sigma p control chart shows serious drawbacks as the true rate of nonconforming items is small. In this study, we proposed an alternative p chart based on improved p estimator by using coefficient of variation as prior information, which can provide an improvement over the usual p chart. The proposed p chart is compared with both traditional p chart and Cornish–Fisher corrected p chart. The benefits of using the improved p estimator for monitoring high-quality processes is illustrated with simulated data.

Key Words: *improved estimation; p control chart; false alarm risk; proportion of nonconforming; statistical process control*

References

- [1] Joeke S, Barbosa EP (2013) An improved attribute control chart for monitoring nonconforming proportion in high quality processes, *Control Engineering Practice* 21:407-412.
- [2] Searls DT (1964) The utilization of a known coefficient of variation in the estimation procedure, *Journal of the American Statistical Association* 59:1225-1226.
- [3] Ho LL, Quinino RdC, Suyama E, Lourenço RP (2012) Monitoring the conforming fraction of high-quality processes using a control chart p under a small sample size and an alternative estimator, *Statistical Papers* 53:507-519.
- [4] Wang H (2009) Comparison of p control charts for low defective rate, *Computational Statistics & Data Analysis* 53: 4210-4220.
- [5] Xie M, Goh TN, Kuralmani V (2002) *Statistical Models and Control Charts for High-Quality Processes*, Springer: Boston, MA.
- [6] Chen G (1998) An improved p chart through simple adjustments, *Journal of Quality Technology* 30(2):142-151.

Estimating the Gini index for income loss distributions under random censoring

Bari Amina, RASSOUL Abdelaziz, Ould Rouis Hamid

LRDSI Laboratory, blida 1 University

Abstract

The Gini index is one of the most widely used inequality indices, it has the distinction of being derived from the Lorenz curve, but generally it is estimated assuming that complete and unbiased samples are available. In this paper, we make use of the extreme value theory and Kaplan-Meier estimator to construct a new estimator of the Gini index when the data are censored, and we study the asymptotic normality property. We show the performance of our proposed estimator by some results of simulations.

Keywords : Gini index, random censoring, loss distributions, Kaplan-Meier estimator.

References :

- Atkinson, A.B., (1970), On the measurement of inequality, *Journal of Economic Theory*, Vol. 2, pp. 244-263
- Beirlant, J., Guillou, A., Dierckx, G. and Fils-Villetard, A., 2007. Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* 10, 151-174.
- Bonetti M, Gigliarano C, Muliere P (2009) The Gini concentration test for survival data. *Lifetime Data Anal* 15:493–518.
- Cowell, F. A. and Flachaire, E. (2007). "Income Distribution and Inequality Measurement: The Problem of Extreme Values," *Journal of Econometrics*, vol. 141, pp. 1044-1072.
- Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *American Statistical Association*, 9, 209-219.
- Tse S.M. (2006) Lorenz curve for truncated and censored data, *Annals of the Institute of Statistical Mathematics*, 58, 675–686.

FUSION OF GEOMETRIC AND TEXTURE FEATURES FOR SIDE VIEW FACE RECOGNITION

Salman Mohammed JIDDAH 1 , Main ABUSHAKRA 2 , Kamil YURTKAN 3

1 Computer Engineering Department, Cyprus International University, 21809666@student.ciu.edu.tr

2 Computer Engineering Department, Cyprus International University, maanamen@gmail.com

3 Computer Engineering Department, Cyprus International University, kyurtkan@ciu.edu.tr

Abstract

Biometric recognition systems have been getting a lot of attention in both academia and the industrial sector. One of such aspects of biometrics attracting interest is side-view face recognition, where the side-view of the face is the only known information about the subjects. Currently, face recognition algorithms achieved significant success and there are recent products including face recognition ability. However, there are still challenges when the cases are in real time and if there are limited samples about the subjects. Having only side views of a face is currently a challenge for face recognition algorithms. Our study embarks on contributing to the research of side-view face biometrics by proposing the fusion of geometric and texture features of the side-view face. Local Binary Pattern (LBP) is used for the extraction of texture features. The Laplacian filter is employed in order to extract the geometric features. Both features are fused in the feature extraction level. Experiments are carried out separately before the fusion of the features in order to observe and compare the effect of the fusion on the performance of side-view face recognition. Support Vector Machine (SVM) has been employed as the classifier. The training of the system was done using the histograms of the texture and geometric features extracted and labelled for every individual subject in the dataset. All experiments are completed on the National Cheng Kung University (NCKU) faces dataset.

Key Words: Face recognition; side view; biometric; feature extraction; fusion.

References

[1] Joshi, Abhilasha, Neetu Sood, and Indu Saini. "Analysis of Side View Face Recognition Using Gabor Filter and PCA Techniques." *Journal of Image Processing & Pattern Recognition Progress* 4.2 (2017): 21-27.

[2] National Cheng Kung University (NCKU) Robotics Face Dataset, 2002, [online]

Available:

http://robotics.csie.ncku.edu.tw/Databases/FaceDetect_PoseEstimate.htm#Our_Database_.

[3] D. Huang, C. Shan, M. Ardebilian, Y. Wang, and L. Chen, “Local Binary Patterns and Its Application to Facial Image Analysis : A Survey,” pp. 1–17.

[4] Jiddah, Salman Mohammed, and Kamil Yurtkan. “Fusion of geometric and texture features for ear recognition.” 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2018.