June 10

18:00 – 19:30 Registration

June 11

08:30 – 09:30 Registration

09:30 – 10:00 Opening Session

10:00 – 11:00 Invited talk
Frederico Caeiro (Nova Univ. Lisbon, Portugal)
A Review and Recent Developments in Tail Inference

11:00 – 11:30 Coffee Break

11:30 – 13:00 Contributed talks – Multivariate orderings

- **M. Capaldo** (U Salerno, Italy), J. Navarro (U Murcia, Spain)
  Multivariate Gini's indices: new developments and applications
- **S. Bonnini** (U Ferrara, Italy)
  Distribution-Free Test on Treatment Effects with Multivariate Ordered Data: a Biostatistical Application
- M. Denuit, **P. Ortega-Jimenez** (U C Louvain, Belgium), C.T. Robert (U Lyon 1, France)
  Monotonicity of conditional expectations given the sum for conditionally independent risks

13:00 – 14:30 Lunch

14:30 – 16:00 Contributed talks – Censoring

- **I. Basak** (Penn State U, USA)
  Prediction for Censored Lifetimes From Weibull Distribution in Khamis and Higgins Step-Stress Model
- **Y. Lu**, M. Kateri (U Aachen, Germany)
  A Heterogeneous Step-Stress Model for Exponential Lifetimes under Type-II Censoring
- **E. Cramer** (Aachen U, Germany)
  Hybrid censored minimal repair and record data

16:00 – 16:30 Coffee Break

16:30 – 18:00 Contributed talks – Estimation

- **P. Basak** (Penn State U, USA)
  Estimation of Parameters of Mixture of Two Normals Based on Lower Record Statistics
- **N.I. Nikolov** (Sofia U, Bulgaria)
  Comparison of parameter estimation methods for a Beta-Uniform mixture model
- **M. Chacko**, A.E. Koshy (U Kerala, India)
  Estimation of multicomponent stress-strength reliability for exponentiated Gumbel distribution using progressive type II censoring

June 12

09:00 – 10:00 Invited talk

Luciano Pomatto (CalTech, USA)
New developments in the theory of stochastic orders

10:00 – 11:00 Contributed talks – Coherent systems 1

- **T.A. Mazzuchi**, S. Sarkani, L. Raubenheimer (G. Washington Univ., USA)
  Development of k-out-of-n Failure Time Distributions in Dependent Environments
- **A. Goroncy**, K. Jasinski (Torun Univ., Poland)
  Three-state discrete time k-out-of-n system and the number of components in all possible states on system failure

11:00 – 11:30 Coffee Break

11:30 – 13:00 Contributed talks – Coherent systems 2

- **M. Capaldo**, A. Di Crescenzo (U Salerno, Italy), F. Pellerey (Politecnico Torino, Italy)
  Analysis of systems with shared components, ROC distortions and lifetime distances
- J. Navarro (U Murcia, Spain), T. Rychlik (Inst. Math. Polish Acad. Sci.), **M. Szymkowiak** (Poznan U, Poland)
  Key distributions in the preservation of aging classes under the systems formation
- **K. Davies** (McMaster U, Canada), A. Dembinska (Warsaw U, Poland)
  On the Residual Lifetimes of Dependent Components Upon System Failure

13:00 – 14:30 Lunch

14:30 – 16:00 Contributed talks – Variability

- **A. Eberl**, B. Klar (Karlsruhe Inst. Tech., Germany)
  Always used, never questioned: Defining dispersion for discrete distributions
- J. Baz (U Oviedo, Spain), **F. Pellerey** (Politec. Torino, Italy), I. Diaz, S. Montes (U Oviedo, Spain)
  On effects of dependence in variability estimation
- A. Arriaza (U Cadiz, Spain), J. Navarro (U Murcia, Spain), M.A. Sordo, **A. Suárez-Llorens** (U Cadiz, Spain)
  A variance-based importance index for systems with dependent components

16:00 – 16:30 Coffee Break

16:30 – 18:00 Contributed talks – Applications

- **V. Holý**, J. Zouhar (Prague U Econ. Bus., Czechia)
  Modelling Time-Varying Rankings
- **J. Arevalillo** (UNED, Spain), J. Navarro (U Murcia, Spain)
  On stochastic orders for multivariate scale mixtures of skew normal distributions with application to assess the evolution of summer temperatures in the Iberian Peninsula
- **K. Trzcińska**, E. Zalewska (U Lodz, Poland)
  Analysis of the economic situation of people with different levels of education based on the Zenga distribution

June 13

09:00 – 10:00 Invited talk

Cécile Durot (Univ. Paris Nanterre, France)
Minimax Optimal rates of convergence in the shuffled regression, unlinked regression, and deconvolution under vanishing noise

10:00 – 11:00 Contributed talks – Order statistics 1

- **J. Navarro** (U Murcia, Spain)
  Are the order statistics ordered? A copula approach
- **T. Rychlik** (Inst. Math. Polish Acad. Sci.)
  Bounds on Variances of Generalized Order Statistics

11:00 – 11:30 Coffee Break

11:30 – 13:00 Contributed talks – Nonparametric statistics 1

- X.L. Gu, **G. Tang**, G. Yu (U Pittsburgh, USA)
  Nonparametric Estimators for A Binary Outcome Under A Monotonicity Restriction
- **M.E. Benjrada** (U Bergamo, Italy)
  Nonparametric tests in deconvolution
- **T. Lando** (U Bergamo, Italy)
  Testing second-order stochastic dominance

13:00 – 14:30 Lunch

14:30 – 16:00 Contributed talks – Distributions

- **Ç. Çetinkaya** (U Kırıkkale, Turkey)
  Order restricted inferences for $R = P(X_1 < Y < X_2)$ based on the Weibull distribution under joint progressive censoring
- **M. Neuhäuser** (Koblenz Univ., Renagen, Germany)
  Trend tests based on the ordered heterogeneity test
- **I. Birbicer**, A.I. Genc (U Cukurova, Turkey)
  Some relations for single and product moments of order statistics from K3D

16:30 – 18:00 Visit to Historical Sites of University of Coimbra

19:30 – 23:00 Conference Dinner

June 14

09:00 – 10:00 Invited talk

Anna Dembińska (Warsaw Univ. Tech., Poland)
Discrete order statistics and their applications in reliability theory

10:00 – 11:00 Contributed talks – Nonparametric statistics 2

- **O. Ozturk** (Ohio State Univ., USA), O. Kravchuk, R. Jarrett (Univ. Adelaide, Australia)
  Enhancing Statistical Inference Through Post-Stratification in Completely Randomized Designs
- **A. Kovalevskii**, M. Chebunin (Sobolev Inst. Math., Novosibirsk, Russia)
  Limit theorems for multiple orderings of multidimensional data

11:00 – 11:30 Coffee Break

11:30 – 13:00 Contributed talks – Order statistics 2

- **C. Empacher**, U. Kamps (RWTH Aachen Univ., Germany)
  Prediction of record values from several samples
- C. Empacher, U. Kamps (Aachen Univ., Germany), **A.B. Schmiedt** (Rosenheim Tech. Univ., Germany)
  Prediction intervals for future Pareto record values with applications in insurance
- **M. Bieniek** (Univ. Lublin, Poland)
  Further developments on characterizations of distributions based on regressions of GOS

13:30 – 13:15 Closing Session

A Review and Recent Developments in Tail Inference

Frederico Caeiro[a]
[a] *Center for Mathematics and Applications (CMA), NOVA University Lisbon, Portugal*

Back to schedule

Studying Extreme events can be a challenge, because these rare events can occur outside the range of available data. Consequently, statistical inference is derived from extreme observations, using an appropriate tail model. In practice, the sample size is small. The extreme-value index (EVI) measures the weight of the tail and is one of the primary parameters of extreme events. Thus, a crucial step in tail inference is the estimation of the EVI. If $\xi < 0$, the distribution function F belongs to the max-domain of attraction of the Weibull distribution, then is short-tailed i.e. F has an upper bounded support; if $\xi = 0$, F belongs to the max-domain of attraction of the Gumbel distribution and $1 - F$ has an exponential decay; If $\xi > 0$, F is heavy-tailed, i.e., F belongs max-domain of attraction of the Fréchet distribution. Consequently, $1 - F$ has a polynomial decay. Attention will be given to heavy-tailed models, which are extremely important due to the low frequency and high magnitude of extreme values.

Back to schedule

New developments in the theory of stochastic orders

Luciano Pomatto[a]
[a] *CalTech, USA*

Back to schedule

Stochastic orders are a fundamental tool in the study of decision making, in information economics, and, more generally, for the non-parametric comparison of probability distribution. In this presentation I will provide an overview of some recent findings on stochastic dominance and the Blackwell order, with implications for models of choice under risk and the comparison of experiments.

Back to schedule

Minimax Optimal rates of convergence in the shuffled regression, unlinked regression, and deconvolution under vanishing noise

Cécile Durot[a] and Debarghya Mukherjee[b]
[a] *Modal'x, Université Paris Nanterre, Nanterre, France,* [b]Boston University, USA

Back to schedule

Shuffled regression and unlinked regression represent intriguing challenges that have garnered considerable attention in many fields, including but not limited to ecological regression, multi-target tracking problems, image denoising, etc. However, a notable gap exists in the existing literature, particularly in vanishing noise, i.e., how the rate of estimation of the underlying signal scales with the error variance.

This paper aims to bridge this gap by delving into the monotone function estimation problem under vanishing noise variance, i.e., we allow the error variance to go to 0 as the number of observations increases.

Our investigation reveals that, asymptotically, the shuffled regression problem exhibits a comparatively simpler nature than the unlinked regression; if the error variance is smaller than a threshold, then the minimax risk of the shuffled regression is smaller than that of the unlinked regression. On the other hand, the minimax estimation error is of the same order in the two problems if the noise level is larger than that threshold.

Our analysis is quite general in that we do not assume any smoothness of the underlying monotone link function.

Because these problems are related to deconvolution, we also provide bounds for deconvolution in a similar context.

Through this exploration, we contribute to understanding the intricate relationships between these statistical problems and shed light on their behaviors when subjected to the nuanced constraint of vanishing noise.

Back to schedule

Discrete order statistics and their applications in reliability theory

Anna Dembińska
Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland, e-mail: anna.dembinska@pw.edu.pl

Back to schedule

Order statistics play an important role in reliability theory. They appear in a natural way when we consider lifetimes of some technical structures. Let $X_1, X_2, \ldots, X_n$ be lifetimes of n items that form a system and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics. Then it is easily seen that

- $X_{1:n}$ describes the lifetime of a series system,
- $X_{n:n}$ describes the lifetime of a parallel system,
- $X_{k:n}$ describes the lifetime of a k-out-of-n system, i.e. a system consisting of n components that functions if and only if at least k of the components work.

More generally, the lifetime of any coherent system can be expressed via order statistics, provided the random vector $(X_1, X_2, \ldots, X_n)$ is exchangeable, i.e. $(X_1, X_2, \ldots, X_n)$ has the same distribution as $(X_1, X_2, \ldots, X_n)$ for any permutation $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ of $(1, 2, \ldots, n)$. A coherent system is a technical structure such that all its components are relevant and having the property that replacing a failed component by a working one cannot cause a working system to fail. As shown in [10], for any coherent system with exchangeable lifetimes of components, the lifetime of the system T is given by

$$P(T > t) = \sum_{i=1}^{n} s_i P(X_{i:n} > t) \text{ for any } t,$$

where $\mathbf{s} = (s_1, \ldots, s_n)$ is the Samaniego signature of the system.

In my talk, I will focus on the case when the lifetimes $X_1, X_2, \ldots, X_n$ are discrete random variables. As an example of a system with discrete X_i, $i = 1, 2, \ldots, n$, we can consider a system in which the component lifetimes are the numbers of turn-on and switch-off up to the failure, or a system which performs a task repetitively and its components have certain probabilities of failure upon each cycle. Treating the discrete case is considerably more complicated than the absolutely continuous one due to non-zero probability of ties among times of component failures.

I will present methods for studying reliability properties of k-out-of-n systems composed of components with not necessarily independent and not necessarily identically distributed discrete lifetimes. I will also explain how some of these methods can be generalized to the class of coherent systems with exchangeable components.

First, I will show how to obtain a closed-form formula describing the joint probability mass function of any subset of order statistics arising from discrete samples. This formula will be a starting point to examine the following problems of interest:

- probability of a failure of a system at a given time;
- various residual lifetimes of a used system;
- residual lifetimes of components that survived system failure;
- moments of lifetime of a system;
- the distribution of the number of failed components at the time of a failure of the system;

- inference based on failure times of components observed up to and including system failure.

Discussing the above mentioned problems I will show various probabilistic properties of discrete order statistics, including characterizations and asymptotic behavior. I will also point out some open problems.

My talk will be based on a series of papers [1]-[9].

References

[1] Davies K., Dembińska A. (2019). On the number of failed components in a k-out-of-n system upon system failure when the lifetimes are discretely distributed. *Reliab. Eng. Syst. Saf.* **188**, 47–61.

[2] Davies K., Dembińska A. (2024). On the residual lifetimes of dependent components upon system failure. *Submitted.*

[3] Dembińska A. (2018). On reliability analysis of k-out-of-n systems consisting of heterogeneous components with discrete lifetimes. *IEEE Trans. Rel.* **67**, 1071–1083.

[4] Dembińska A., Eryilmaz S. (2021). Discrete time series-parallel system and its optimal configuration. *Reliab. Eng. Syst. Saf.* **215**, 107832.

[5] Dembińska A., Goroncy A. (2020). Moments of order statistics from DNID discrete random variables with application in reliability. *J. Comput. Appl. Math.* **371**, 112703.

[6] Dembińska A., Jasiński K. (2021). Maximum likelihood estimators based on discrete component lifetimes of a k-out-of-n system. *TEST* **30**, 407–428.

[7] Dembińska A., Jasiński K. (2024). Likelihood inference for geometric lifetimes of components of k-out-of-n systems. *J. Comput. Appl. Math.* **435**, 115267.

[8] Dembińska A., Jasiński K. (2024). Inference for geometric lifetimes of heterogeneous components from data colleted till failure of a k-out-of-n system. *Manuscript in preparation.*

[9] Dembińska A., Nikolov N.I., Stoimenova E. (2021). Reliability properties of k-out-of-n systems with one cold standby unit. *J. Comput. Appl. Math.* **388**, 113289.

[10] Navarro, J., Samaniego, F.J., Balakrishnan, N. and Bhattacharya, D. (2008). On the application and extension of system signatures to problems in engineering reliability. *Naval Res. Logist.* **55**, 313–327.

Back to schedule

Multivariate Gini's indices: new developments and applications

Marco Capaldo[a], Jorge Navarro[b]

[a] *Dipartimento di Matematica, Università degli Studi di Salerno*
Via Giovanni Paolo II, 132, I-84084 Fisciano (SA), Italy.

[b] *Departamento de Estadística e Investigación Operativa, Facultad de Matemática*
Universidad de Murcia, Murcia 30100, Spain.

Back to schedule

The Gini's index was defined as twice the area between the egalitarian line and the Lorenz curve and it quantifies how far a random variable and an its independent copy are. Some generalizations of the Gini's index have been introduced in the literature (see, for instance, Sections 5.3.2 and 7.4.1 in Arnold and Sarabia [1]). This talk is devoted to illustrate new kinds of multivariate Gini's indices, defined and studied in Capaldo and Navarro [2]. In the proposed settings the involved random variables are possibly dependent and not necessarily identically distributed.

The interpretation of these new indices is pointed out by discussing various results and bounds, most of them involving copulas, and by considering some illustrative examples. Aiming to apply these measures in the context of reliability theory (see Navarro [3]), we provide a suitable efficient version for any semi-coherent system. Empirical Gini's indices are also considered for data analysis.

References

[1] Arnold, B.C., Sarabia, J.M., (2018). Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics. Statistics for Social and Behavioral Sciences. Springer.

[2] Capaldo, M., Navarro, J., (submitted, 2023+). New multivariate Gini's indices. ArXiv preprint https://arxiv.org/abs/2401.01980

[3] Navarro, J., (2022). Introduction to System Reliability Theory. Springer.

Back to schedule

Distribution-Free Test on Treatment Effects with Multivariate Ordered Data: a Biostatistical Application

Stefano Bonnini[a]
[a] *University of Ferrara, Via Voltapaletto 11, 44121 Ferrara (Italy)*

Back to schedule

1 Introduction

This work consists of the application of a distribution-free test, based on the permutation approach, to a multivariate biostatistical problem. The goal of this study is to test the effect of "assisted motor activity" (AMA) on the health of patients affected by "low back pain" (LBP), "hypertension" and "diabetes". AMA is a treatment based on specific physical exercises aimed at restoring motor limitations caused by various factors. Specifically, the goal is to test whether AMA determines an improvement in the functionality and perceived health status of comorbid patients at the significance level $\alpha = 0.05$.

This is a case-control experiment with a comparative evaluation between two independent samples. The samples are a treated group of 27 patients (group 1) and a control group of 20 patients (group 0). The health status of the two groups at time t_0 (before the treatment) does not differ. They are compared at time t_1, after the treatment for group 1. The health status is measured according to 13 different binary or ordinal outcomes. The null hypothesis of the test consists of the equality in the distribution of the multivariate responses of group 1 and group 0, whereas, under the alternative hypothesis, the health status of the treated patients is better. Therefore, the alternative hypothesis is directional and we are in the presence of a multivariate stochastic dominance problem for ordinal variables. Since being jointly over 60 years old and affected by LBP may represent a risk factor, a confounding factor S is defined such as $S = 1$ if the patient is over 60 years old and affected by LBP, and $S = 0$ otherwise.

The approach proposed in this work is based on the method of Combined Permutation Test (CPT) [1, 2], which is suitable for multivariate categorical data and for the presence of the confounding factor. To avoid confounding effects by comparing similar patients in terms of the confounder, stratification of the groups and intra-stratum permutation univariate two-sample tests are carried out. Given the number of components of the multivariate response (13) and the number of strata (2), the number of such partial tests is 26. The combination of the p-values of the partial tests, according to the CPT approach, provides a test statistic suitable for the overall problem.

Section 2 is dedicated to the description of the data and of the formal problem, Section 3 includes a brief description of the proposed method, whereas in Section 4 the results are reported and commented.

2 Data and Problem

The components of the ordinal multivariate response variable are listed below:

- X_1: Self-assessment on having health issues (1-yes, 2-partial, 3-no)
- X_2: Self-assessment on the ability to perform moderate physical activity (1-no, 2-partial, 3-yes)
- X_3: Self-assessment on the difficulty in stair climbing (1-yes, 2-partial, 3-no)
- X_4: Physical performance lower than expected in the last month (1-yes, 2-no)

- X_5: Need to limit some types of activity in the last month (1-yes, 2-no)
- X_6: Physical perf. lower than expected due to emotional state in the last month (1-yes, 2-no)
- X_7: Decrease of mind concentration in the last month due to emotional state (1-yes, 2-no)
- X_8: Difficulty in daily activities due to pain in the last month (1-yes, 2-no)
- X_9: Frequency of calm and serenity in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always)
- X_{10}: Frequency of feeling full of energy in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always)
- X_{11}: Frequency of feeling discouraged and sad in the last month ((1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never)
- X_{12}: Frequency of negative effects of health and emotional state on social activities in the last month (1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never)
- X_{13}: Self-assessment of the level of stress (1-very high, 2-high, 3-average, 4-moderate, 5-very low)

Let $X_{1,sv}$ and $X_{0,sv}$ represent the vth outcome or, equivalently, the vth component of the multivariate response, in the stratum s for the treated and the control group respectively, with $s = 0, 1$ and $v = 1, \ldots, 13$. The partial problem related to the vth outcome and the stratum s consists of testing $H_{0,sv} : X_{1,sv} \overset{d}{=} X_{0,sv}$ versus $H_{1,sv} : X_{1,sv} >^d X_{0,sv}$, where $\overset{d}{=}$ and $>^d$ denote equality in distribution and stochastic dominance respectively. Such hypotheses may be written as

$$H_{0,sv} : F_{1,sv}(x) = F_{0,sv}(x), \quad \forall x \tag{1}$$

and

$$H_{1,sv} : F_{1,sv}(x) \leq F_{0,sv}(x),\ \forall x \quad \text{and} \quad \exists x : F_{1,sv}(x) < F_{0,sv}(x) \tag{2}$$

where $F_{j,sv}(x)$ denotes the cumulative distribution function of $X_{j,sv}$, with $j = 0, 1$.

Under the null hypothesis, for the vth outcome, both the intra-stratum partial null hypothesis $H_{0,1v}$ and $H_{0,0v}$ are true, thus $H_{0,v} : H_{0,1v} \cap H_{0,0v}$. Similarly, $H_{1,v} : H_{1,1v} \cup H_{1,0v}$, with obvious notation. Consequently, the overall null and alternative hypothesis of the problem can be denoted by $H_0 : \cap_{v=1}^{13} H_{0,v}$ and $H_1 : \cup_{i=1}^{13} H_{1,v}$ respectively.

3 Methodological solution

Let $f_{1j,sv}$ and $f_{0j,sv}$ denote the absolute frequency of the jth ordered category (i.e. the number of statistical units on which such a category is observed) within the stratum s for the vth variable in the treated and control group respectively. Hence, the cumulative frequencies in the treated and control group can be denoted by $F_{1j,sv} = \sum_{r=1}^{j} f_{1r,sv}$ and $F_{0j,sv} = \sum_{r=1}^{j} f_{0r,sv}$ respectively. For the partial test concerning $H_{0,sv}$ vs $H_{1,sv}$, the following Anderson-Darling type test statistic may be used

$$T_{sv} = \sum_{j=1}^{k_v - 1} (F_{0j,sv} - F_{1j,sv})[F_{\cdot j,sv}(n_s - F_{\cdot j,sv})]^{-0.5}, \tag{3}$$

where k_v is the number of ordered categories of the vth variable, $F_{\cdot j,sv} = F_{0j,sv} + F_{1j,sv}$ and $n_s = F_{\cdot k_v,sv}$. To solve the testing problem related to the vth variable, i.e. to test $H_{0,v}$ vs $H_{1,v}$, a first-level combination of the significance level functions of the partial tests of the two strata may be applied. If $L_{sv}(t_{sv}) = P[T_{sv} \geq t_{sv}|X]$ is the significance level function for the sth stratum and the vth variable given the observed dataset X, for any $t_{sv} \in \mathbb{R}$, according to the permutation distribution, a suitable combined test statistic for the vth variable is

$$T'(t_{1v}, t_{0v}) = \max[(1 - L_{1v}(t_{1v})(1 - L_{0v}(t_{0v})], \tag{4}$$

for any couple of values $(t_{1v}, t_{0v}) \in \mathbb{R}^2$.

Similarly, to solve the general multivariate problem, a second-level combination may be carried out. Let $L'_v(t'_v) = P[T_v \prime \geq t'_v|X]$ denote the significance level function of T'_v for any $t'_v \in \mathbb{R}$. The second-level combined test statistic is then

$$T''_v(t'_1, \ldots, t'_{13}) = \max[(1 - L'_1(t'_1), \cdots, (1 - L'_{13}(t'_{13})] \tag{5}$$

Finally, H_0 is rejected in favour of H_1 if the p-value of the combined test is less than or equal to the significance level $\alpha = 0.05$, formally if $L''(t - obs'') \leq \alpha$, where $L''(t'') = P[T'' \geq t''|X]$ with $t'' \in \mathbb{R}$.

Probabilities and p-values are computed according to the null permutation distributions, obtained by permuting the rows of X, because the exchangeability condition is satisfied under the null hypothesis.

4 Results and conclusions

Since the overall p-value of the CPT is equal to 0.019, then we have empirical evidence in favour of the hypothesis of significant effect of AMA on the health of patients. The proposed approach represents a valid solution for the presented testing problem, whose complexity is due to the multivariate nature of the response, the categorical data, the directional alternative hypothesis, the presence of confounders and the small sample sizes.

References

[1] Pesarin F., Salmaso L. (2010). *Permutation Tests for Complex Data.* Theory, Applications, and Software. Wiley, Chichester.

[2] Bonnini S., Corain L., Marozzi M., Salmaso L. (2014). *Nonparametric Hypothesis Testing. Rank and Permutation Methods with Applications in R.* Wiley, Chichester.

Back to schedule

Monotonicity of conditional expectations given the sum for conditionally independent risks

Michel Denuit[a], Patricia Ortega-Jiménez[a] and Christian Y. Robert[b]
[a]*ISBA (LIDAM), UCLouvain, Belgium*, [b]*ISFA, Université Lyon 1, France*

Back to schedule

Stochastic monotonicity of random variables $\{X_i\}_i^n$ given the value of their sum $S = \sum_{i=1}^n X_i$ is referred to as "Efron's monotonicity". In particular, the study of the monotonicity of conditional expectation $m_i(s) = \mathrm{E}[X_i|S = s]$ has great relevance in signal processing or risk sharing, where it is referred to as "non-sabotage" condition. In Denuit (2019), the monotonicity of $m_i(\cdot)$ is characterized by the ordering of $S + \widetilde{X_i} - X_i$ and S in the likelihood ratio order, where $\widetilde{X_i}$ stands for the size-biased transform of X_i. In this work, we provide sufficient conditions for this ordering when losses follow a common mixture model, a dependence structure often used in practice. Several examples illustrate the applicability of the results. In particular, alternative conditions are provided when considering an scale mixture model. The approach is applied to the case where losses follow a multivariate Pareto distribution of Mardia's type II and, for this particular case, the exact expression of each participant's contribution is obtained using divided differences.

References

[1] Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. *ASTIN Bulletin* **49**, 591–617.

Back to schedule

Prediction for Censored Lifetimes From Weibull Distribution in Khamis and Higgins Step-Stress Model

Indrani Basak[a]
[a] *Penn State Altoona, 3000 Ivyside Park, Altoona, PA 16601, USA*

Back to schedule

We consider the problem of prediction of lifetimes of units from the Weibull distribution which are censored under a simple step-stress testing experiment in this article. We considered progressive Type-II censoring as the form of censoring. Suppose a sample of n experimental units are placed on a simple step-stress life test at an initial stress level of s_1 and the stress level is changed to s_2 at a pre-fixed time τ. Then, the progressive Type-II censoring is implemented in this experimental setting in the following manner. At the stress-level s_1 and at the time of the first failure, R_1 of the $n-1$ surviving units are randomly removed from the experiment. At the time of the second failure, R_2 of the $n-2-R_1$ surviving units are randomly removed from the experiment, and similarly the test continues until time τ. Let N_1 be the random number of units that fail at stress level s_1 and $R^{(1)} = \sum_{i=1}^{n_1} R_i$ be the total number of the censored units at stress level s_1 where n_1 denotes the observed value of N_1. Then, after time τ (at stress level s_2), at the time of the (n_1+1)-th failure, R_{n_1+1} of the $n-n_1-R^{(1)}-1$ surviving units are randomly removed from the experiment. At the time of the (n_1+2)-th failure, R_{n_1+2} of the $n-n_1-R_{(1)}-R_{n_1+1}-2$ surviving units are randomly removed from the experiment, and similarly the test continues at the stress level s_2. Let $R^{(2)} = \sum_{I=n_1+1}^{m} R_i$ be the total number of the censored units at stress level s_2 for a fixed value of m, the total number of observations. Then, let $N_2 = m - N_1$ denotes the random number of units that fail at stress level s_2. With m, R_i $(i = 1, 2, \ldots, m-1)$ fixed in advance, the test continues until the m-th failure at which time all the remaining $n - m - R^{(1)} - \sum_{i=n_1+1}^{m-1} R_i$ surviving units are removed. Note that $n = n_1 + n_2 + R^{(1)} + R^{(2)}$ where n_2 denotes the observed value of N_2. If $R_1 = \cdots = R_{n_1} = 0$, $R_{n_1+1} = \cdots = R_m = 0$, then $n = m$ which corresponds to the complete sample situation. If $R_1 = \cdots = R_{n_1} = 0$, $R_{n_1+1} = \cdots = R_{m-1} = 0$ and $R_m = n - m$, then it corresponds to the conventional Type-II right censoring scheme. Note that the life-testing experiment is terminated when the m-th failure occurs. With these notations, we will observe the following progressively censored data:

$$\mathbf{t} = (t_1, \ldots, t_{n_1}, t_{n_1+1}, \ldots, t_m)$$

with $t_1 < \ldots < t_{n_1} < \tau \le t_{n_1+1} < \ldots < t_m$.

Here, $\mathbf{t}$ is the observed values of the variable $\mathbf{T} = (T1, \ldots, T_{N_1}, T_{N_1+1}, \ldots, T_m)$ denoting the m Type-II progressively right censored order statistics from a population with pdf $g(t) = g(t;\theta)$ where $(t;\theta) \in D = (\mathbb{R}^+)^m\Omega$. Also, $\theta = (\mu_1, \mu_2, \sigma)$ and Ω is a 3-dimensional parametric space.

Cumulative Exposure Model (CEM) is the most popular model for analyzing step-stress data. In case of the Weibull distribution, the CEM becomes quite complicated. Due to this reason, Khamis and Higgins (1998) proposed a step-stress model based on the hazard functions and we will use that model in this article. Khamis and Higgins [1] proposed the following step-stress model for the Weibull distribution:

$$h(y) = \begin{cases} h_1(y) = \frac{\beta}{\theta_1} y^{\beta-1} & \text{for} \quad 0 < y < \tau \\ h_2(y) = \frac{\beta}{\theta_2} y^{\beta-1} & \text{for} \quad \tau < y < \infty. \end{cases} \tag{1}$$

The corresponding survival functions are:

$$\overline{F}(y) = \begin{cases} \overline{F}_1(y) &= e^{-\frac{y^\beta}{\theta_1}} \quad \text{for} \quad 0 < y < \tau \\ \\ \overline{F}_2(y) &= e^{-\frac{y^\beta - \tau^\beta}{\theta_2} - \frac{\tau^\beta}{\theta_1}} \quad \text{for} \quad \tau < y < \infty \end{cases} \tag{2}$$

and will be referred to as Khamis-Higgins model in this article.

In this article, we have derived two kinds of predictors – the Maximum Likelihood Predictors (MLP) and the Conditional Median Predictors (CMP) for the survival times of unit $T_{j:R_i}$ which is the j-th order statistic out of a sample of size R_i; $j = 1, 2, \ldots, R_i$; $i = 1, 2, \ldots, m$ from the Weibull distribution which are progressively Type II censored under a simple step-stress model. These are derived in each of the following three cases:

Case 1: $(1 \leq n_1 \leq m - 1$ and $i = 1, \ldots, n_1)$

Case 2: $(n_1 = m$ and $i = 1, \ldots, n_1)$

Case 3: $(1 \leq n_1 \leq m - 1$ and $i = n_1 + 1, \ldots, m)$

We will assume that, at the stress level s_1, lifetimes have the distribution Weibull (β, θ_1) with the shape and scale parameters β and θ_1 and at the stress level s_2, lifetimes have the distribution Weibull (β, θ_2) with the shape and scale parameters β and θ_2. We found that the prediction methods become complicated for the original Weibull distribution and therefore we will be working with the variable $T = \ln Y$ where the hazard function and survival function of the variable Y is given by (1) and (2) respectively. The probability density function $g(t)$ and the cumulative distribution function $G(t)$ of T are given as

$$g(t) = \begin{cases} g_1(t) &= \frac{1}{\sigma} e^{\frac{t-\mu_1}{\sigma}} e^{-e^{\frac{t-\mu_1}{\sigma}}} \quad \text{for} \quad t < \ln \tau \\ \\ g_2(y) &= \frac{1}{\sigma} e^{-[e^{\frac{t-\mu_2}{\sigma}} - e^{\frac{\ln \tau - \mu_2}{\sigma}} + e^{\frac{\ln \tau - \mu_1}{\sigma}}]} \quad \text{for} \quad \ln \tau < y < \infty. \end{cases}$$

and

$$G(t) = \begin{cases} G_1(t) &= 1 - e^{-e^{\frac{t-\mu_1}{\sigma}}} \quad \text{for} \quad t < \ln \tau \\ \\ G_2(y) &= 1 - e^{-[e^{\frac{t-\mu_2}{\sigma}} - e^{\frac{\ln \tau - \mu_2}{\sigma}} + e^{\frac{\ln \tau - \mu_1}{\sigma}}]} \quad \text{for} \quad \ln \tau < y < \infty. \end{cases}$$

where $\mu_1 = \frac{1}{\beta} \ln \theta_1$, $\mu_2 = \frac{1}{\beta} \ln \theta_2$ and $\sigma = \frac{1}{\beta}$.

It is noted that these two predictors are quite easy to compute. These two prediction methods are numerically illustrated using simulation studies along with generating Mean Squared Prediction Error (MSPE) and Prediction Intervals (PI). We then compare the MLP and the CMP with respect to MSPE and PI. We generated progressive Type-II censored data under the step-stress setting and computed the values of MLP and CMP of $T_{j:R_i}$, having observed T. Then, we carried out a numerical study to compare the performances of these MLP and CMP in terms of their MSPEs. Derivations of the MSPEs of the MLP and CMP for the Extreme Value distribution are complicated and so we used simulations to get those. Moreover, using simulation studies, standard errors of these $T_{j:R_i}$ were generated and the prediction intervals were constructed for each of the predictors MLP and CMP in each situation.

It is found that the predicted values for CMP are generally closer to the actual values than the corresponding predicted values for MLP. Simulation studies show that the prediction method using CMP yield the closest prediction result particularly for larger sample size, larger number of uncensored observations and for delayed censoring scheme as long as the number of predicted observations is not small. It is also observed that the bias for the predicted MLP values decreases when the sample size increases. The MLPs have smaller MSPEs than the CMPs. But the ratio of MSPEs of CMPs to the MSPEs of MLPs becomes closer to 1, in general, as sample size increases.

References

[1] Khamis, I. H. and Higgins, S. H. (1998). A New Model for Step-Stress Testing, *IEEE Transactions on Reliability* **47**, 131.-134.

Back to schedule

A Heterogeneous Step-Stress Model for Exponential Lifetimes under Type-II Censoring

Yao Lu[a], Maria Kateri[a]
[a] *Institute of Statistics, RWTH Aachen University, Pontdriesch 14-16, 52056, Aachen, Germany*

Back to schedule

Accelerated life testing (ALT) is widely implemented to investigate lifetime performance within a condensed timeframe by imposing higher stress levels, inducing failures much earlier than under normal operating conditions. Step-stress ALT (SSALT) is a special type of ALT in which the stress level applied to the tested units is incrementally adjusted at pre-defined time points throughout the experiment. Statistical models for SSALT experiments, assuming a homogeneous population and considering various lifetime distributions or censoring schemes, have been thoroughly discussed in the literature. However, SSALT for a heterogeneous population has received little attention so far, especially in cases where the group membership is unknown and the testing units form disjoint groups. This is a realistic scenario that occurs during a test in practice, e.g., when the testing units separate into groups based on their response, corresponding to different quality levels. Therefore, a heterogeneous SSALT (hSSALT) model with two stress levels (simple hSSALT) for Type-II censored exponential lifetimes is introduced, under the cumulative exposure (CE) assumption. To capture the underlying heterogeneity, a mixture model approach is employed and the EM algorithm is adapted for the maximum likelihood estimation of the model's parameters. The associated asymptotic and bootstrap confidence intervals are also provided. The validity of the proposed model and its advantage over the standard SSALT model in the presence of heterogeneity, are proved and demonstrated via simulation studies.

Acknowledgements

The first author gratefully acknowledges the financial support provided by the Federal Ministry of Education and Research (BMBF) through the projects BALd with Grant Nr. 03XP0320A.

References

[1] Ateya S. F. (2014). Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. *Statistical Papers* **55**, 311—325.

[2] Balakrishnan N., Kundu D., Ng K. T. and Kannan, N. (2007). Point and interval estimation for a simple step-stress model with Type-II censoring. *Journal of Quality Technology* **39(1)**, 35—47.

[3] Bordes L. and Chauveau D. (2016). Stochastic EM algorithms for parametric and semi-parametric mixture models for right-censored lifetime data. *Computational Statistics* **31**, 1513—1538.

[4] Dempster A. P., Laird N. M. and Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1—38.

Back to schedule

Hybrid censored minimal repair and record data

Erhard Cramer[a]
[a] *Institute of Statistics, RWTH Aachen*

Back to schedule

Recently, Berzborn, Cramer [2] applied hybrid censoring schemes to minimal repair and record data using the idea of *samples with replacement* proposed in Epstein [4]. In particular, they utilized the refined modularization/decomposition approach to conduct statistical inference (see [3, 5]). For a recent review on hybrid censoring, see Balakrishnan et al. [1]. In particular, the resulting decomposition can be used to derive the exact (conditional) distribution of the MLE for exponentially distributed lifetimes and leads to compact expressions of the MLE's density function in terms of gamma distributions. Furthermore, it turns out that the distributions have point masses at infinitely many points.

Applications of the approach to likelihood inference, Bayesian estimation, prediction, and Fisher information are presented for minimal repair data/record data.

References

[1] Balakrishnan, N., Cramer, E., Kundu, D. (2023) Hybrid Censoring Know-How – Designs and Implementations. Academic Press, Cambridge.

[2] Berzborn, M., Cramer, E. (2024) Inference for Type-I and Type-II hybrid censored minimal repair and record data. *J. Statist. Theory Practice* **17**, 1-38.

[3] Cramer, E. (2024) Structure of hybrid censoring schemes and its implications, Metrika, to appear.

[4] Epstein, B. (1954) Truncated life tests in the exponential case. *Ann. Math. Stat.* **25**, 555-564.

[5] Górny, J., Cramer, E. (2018) Modularization of hybrid censoring schemes and its application to unified progressive hybrid censoring. *Metrika*, **81**, 173-210.

Back to schedule

Estimation of Parameters of Mixture of Two Normals Based on Lower Record Statistics

Prasanta Basak[a]
[a] *Penn State Altoona, 3000 Ivyside Park, Altoona, PA 16601, USA*

Back to schedule

Record statistics can be viewed as order statistics coming from a sample whose size is determined by the values and the order of occurrence of the observations. The study of record statistics was introduced by Chandler [3]. Since then, a large number of publications on records have appeared. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with absolutely continuous distribution function (cdf) F and probability density function (pdf) f. For $n \geq 1$, we denote the order statistics of $X_1, X_2, \ldots, X_n$ by $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Define

$$L(1) = 1, \quad L(n+1) = \min\{j : j > L(n), X_j > X_{j-1,j-1}\} \quad \text{and} \quad X(n) = X_{L(n),L(n)}, n \geq 1.$$

The sequence $\{X(n)\}(\{L(n)\})$ is called upper record statistics (times). For a comprehensive discussion of records, see Arnold et al. [2], Ahsanullah [1] and Nevzorov [9]. The lower record times and lower record statistics can be obtained from upper record times and upper record values simply by replacing the original sequence of random variables by $\{-X_i, i \geq 1\}$ or (if $P(X_i > 0) = 1$) by $\{1/X_i, i \geq 1\}$.

In this article, we deal with the problem of estimating the parameters of a mixture of two normal distributions based on record statistics. The maximum likelihood and Bayes' methods of estimation are used for this purpose. Statistical models and methods for survival data and other time-to-event data are extensively used in many fields, including the biomedical sciences, engineering, the environmental sciences, economics, actuarial sciences, management and social sciences. Mixtures of two life distributions occur when two different causes of failure are present, each with the same parametric form of life distributions. In recent years, the finite mixture of life distributions have proven to be of considerable interest both in terms of their methodological development and practical applications, see for example, McLachlan [7], McLachlan and Peel [8], Everitt and Hand [4], and Tittertington, Smith and Makov [10].

A random variable X is said to follow a finite mixture distribution with k components, if the pdf of X can be written in the form

$$f(x) = \sum_{i=1}^{k} p_i f_i(x), \tag{1}$$

where $p_i \geq 0$ (known as the ith mixing proportion) such that $\sum_{i=1}^{k} p_i = 1$ and f_i is a density function (known as the ith component of the mixture), $i = 1, 2, \ldots, k$.

The pdf and the cdf of a mixture of two normal distributions with same variances are given respectively by

$$\left.\begin{aligned} f(x; p, \mu_1, \mu_2, \sigma) &= p \cdot f_1(x; \mu_1, \sigma) + (1-p) \cdot f_2(x, \mu_2, \sigma) \\ F(x; p, \mu_1, \mu_2, \sigma) &= p \cdot F_1(x; \mu_1, \sigma) + (1-p) \cdot F_2(x, \mu_2, \sigma) \end{aligned}\right\} \tag{2}$$

where $f_i(x; \mu_1, \sigma)$, $i = 1, 2$, is the pdf pf the normal distribution function given by

$$f_i(x; \mu_1, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma^2}\right] \tag{3}$$

and $F_i(x; \mu_1, \sigma)$ is the corresponding cdf of normal distribution. Here, we will denote the vector of parameters of the mixture model by $\psi = (p, \mu_1, \mu_2, \sigma)$.

The mixtures of normal distribution have many important applications. Details on mixtures of normal distributions can be found in Everitt and Hand [4], Titterington et al. [10] and Lindsay [5]. Inferences based on mixtures of normal distributions are considered by many authors – the above mentioned citations have a number of references. We derive the maximum likelihood estimator of ψ. We obtain Bayes' estimates of ψ using the approximate form of Lindley [6] based on lower record statistics from the distribution given in equation (2). The Bayes' estimates are computed and compared with their corresponding maximum likelihood estimates based on Monte Carlo simulation study.

References

[1] Ahsanullah, M. (1995). *Record Statistics*. Nova Science Publishers, Inc. Commack, NY.

[2] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*. John Wiley & Sons, New York.

[3] Chandler, (1952), Record statistics, *Annals of Statistics* **17**, 722–740.

[4] Everitt, B. S., and Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman and Hall, New York.

[5] Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol 5, Institute of Mathematical Statistics, 1995.

[6] Lindley, D. V. (1990). *Approximate Bayesian methods*. Trabajos de Estadistica, 31, pp. 223–237.

[7] McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. Mercel Dekke, New York.

[8] McLachlan, G. J. and Peel, K. D. (1988). *Mixture Models: Inference and Applications to Clustering*. Mercel Dekker, New York.

[9] Nevzorov, V. B. (2000). *Records: Mathematical Theory. American Mathematical Association*. Providence, Rhode Island.

[10] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

Back to schedule

Comparison of parameter estimation methods for a Beta-Uniform mixture model

Nikolay I. Nikolov[a,b]

[a] *Sofia University "St. Kliment Ohridski", Sofia 1504, Bulgaria*
[b] *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria*
email: n.nikolov@math.bas.bg

Back to schedule

The Benjamini-Hochberg (B-H) procedure, introduced in [1], is a method for controlling the false discovery rate in the problem of large-scale multiple testing, common in biomedical and genomic research. A Beta-Uniform mixture is suggested in the literature for approximating the distribution of the tests p-values before the B-H adjustment, see for example [3] and [4]. In this talk, different methods for estimating the unknown parameters of the proposed mixture are studied. To evaluate the maximum-likelihood (ML) estimations, an expectation-maximization algorithm is derived. As a more robust estimation technique, the maximum product of spacings (MPS) method is considered. For finding the MPS estimations, an iterative minorization-maximization algorithm, see [2], is tailored to fit the MPS framework under a general mixture model. In addition, the method of moments is applied and compared to the ML and MPS approaches. The quality of the proposed estimation procedures is measured by their bias and mean squared error, computed via numerical simulations for various combinations of the sample size, proportion parameter in the mixture and shape parameters of the beta distribution.

The talk is based on a joint work with Dean Palejev (Sofia University & Institute of Mathematics and Informatics, Bulgarian Academy of Sciences).

References

[1] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 57(1), 289–300.

[2] Hunter, D. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statistician*, Vol. 58(1), 30–37.

[3] Ji, Y., Wu, C., Liu, P., Wang, J., and Coombes, K. R. (2005). Applications of beta-mixture models in bioinformatics. *Bioinformatics*, Vol. 21(9), 2118–2122.

[4] Palejev, D., and Savov, M. (2021). On the convergence of the Benjamini–Hochberg procedure. *Mathematics*, Vol. 9(17), 2154.

Back to schedule

Estimation of multicomponent stress-strength reliability for exponentiated Gumbel distribution using progressive type II censoring

Manoj Chacko[a] and Ashly Elizabeth Koshy[a]
[a] *University of Kerala, Thiruvanathapuram, India*

Back to schedule

In this paper, the stress-strength reliability, $R_{s,k}$, of a multicomponent s-out-of-k system for exponentiated Gumbel distribution using progressive type II censoring scheme is considered. The maximum likelihood estimation procedure and Bayesian estimation for $R_{s,k}$ are discussed. Bayes estimators are evaluated under both symmetric and asymmetric loss functions using Markov chain Monte Carlo (MCMC) method. The asymptotic, percentile bootstrap and highest posterior density (HPD) confidence intervals for $R_{s,k}$ are obtained. A simulation study is carried out for assessing the efficiency of the estimators developed in this paper. A real data is also examined for illustrative purpose.

Keywords: Exponentiated Gumbel distribution; stress–strength reliability; progressive type II censoring; maximum likelihood estimation; Bayesian estimation; MCMC method

Back to schedule

Development of k-out-of-n Failure Time Distributions in Dependent Environments

Thomas A. Mazzuchi[a], Shahram Sarkani[a], and Lizanne Raubenheimer[b]
[a] *Department of Engineering Management and Systems Engineering, George Washington University, Washington DC, USA*, [b] *Department of Statistics, Rhodes University,Makhanda, South Africa*

Back to schedule

In reliability theory, a k-out-of-n system is a system which functions if k out of n components function (referred to as a G system) or a system that will fail if k of n components fails (referred to as an F system). The use of k-out-of-n systems has become increasingly popular, especially for fault tolerant systems, and has applications in military, communications, electronics, automotive, and data processing systems to name a few. The failure time distribution for a k-out-of-n system can be constructed as the distribution for the $(n + 1 - k)$th order statistic. This is easily constructed when components are independent; however, it is well known that many of these systems are subject to common failure environments. Modeling common failure environments can also be a difficult task, however, the Marshall-Olkin Multivariate Exponential distribution has been used effectively in this capacity. We thus combine the two models and develop expressions for the time to failure of a k-out-of-n system (or the $n + 1 - k$th order statistic distribution) using a shock model approach. When n is even moderately large this can pose a combinatorial challenging problem for the general Marshall-Olkin model. In most k-out-of-n systems, however, identical components are assumed, and we may use the symmetric or Binomial Failure Rate variate of the Marshall-Olkin Model to further reduce complexity.

Back to schedule

Three-state discrete time k-out-of-n system and the number of components in all possible states on system failure.

Agnieszka Goroncy[a] and Krzysztof Jasiński[a]
[a] *Nicolaus Copernicus University, Toruń, Poland*

Back to schedule

We consider a three-state **k**-out-of-n system composed of components which lifetimes are modelled by independent and identically distributed discrete random variables. Such system and its components can function perfectly or totally fail, but can also enter a partial performance state, when they may not fail completely but their efficiency is reduced. Based on the definition introduced by Huang et al. (2000) we focus on the random vector representing the numbers of components in each state, for which we derive its distribution. We illustrate the theoretical results with numerical examples concerning systems with components with geometrically distributed lifetimes following the Markov degradation process.

References

[1] Eryilmaz S. (2015). On the number of remaining components in three-state k-out-of-n system. *Oper Res Lett* 43:616–621.

[2] Erylimaz S., Xie M. (2014). Dynamic modeling of general three-state k-out-of-n:G systems: Permanent based computational results. *J Computat Appl Math* 272:97–106.

[3] Goroncy A., Jasiński K. (2024). Discrete time three-state **k**-out-of-n system's failure and numbers of components in each state. Under review.

[4] Huang J, Zuo MJ, Wu Y (2000). Generalized multi-state k-out-of-$n : G$ systems. *IEEE Trans Reliab* 49: 105-111.

[5] Tian Z, Zuo MJ, Yam RCM (2009). Multi-state k-out-of-n systems and their performance evaluation. *IIE Trans* 41: 32-44.

Back to schedule

Analysis of systems with shared components, ROC distortions and lifetime distances

Marco Capaldo[a], Antonio Di Crescenzo[a], Franco Pellerey[b]

[a] *Dipartimento di Matematica, Università degli Studi di Salerno Via Giovanni Paolo II, 132, I-84084 Fisciano (SA), Italy.*

[b] *Dipartimento di Scienze Matematiche, Politecnico di Torino Corso Duca degli Abruzzi, 24, I-10129 Torino, Italy.*

Back to schedule

Distortion and copula functions play a special role for the detection of the system's reliability from their components' reliability, by taking into account the system's structure. Along this line, in Capaldo et al. [1] we define (i) a new distortion function related to the ROC curve and (ii) new families of distortions in order to deal with both series and parallel structures. We study several pairs of reliability systems with one or more shared components, in the case in which their lifetimes are independent and identically distributed or independent but not identically distributed. The dependence that arises from sharing components is often described by Marshall-Olkin copulas (see Li and Pellerey [3]). Moreover, the lifetimes analysis is enriched by considering some distance measures related to the Gini's mean difference and its new recent generalizations (see Capaldo and Navarro [2] and references therein).

References

[1] Capaldo, M., Di Crescenzo, A., Pellerey, F., (submitted, 2024+). Mean distances and dependence structures for lifetimes of systems with shared components.

[2] Capaldo, M., Navarro, J., (submitted, 2023+). New multivariate Gini's indices. ArXiv preprint https://arxiv.org/abs/2401.01980

[3] Li, X., Pellerey, F., (2011). Generalized Marshall–Olkin distributions and related bivariate aging properties. Journal of Multivariate Analysis **102**(10), 1399-1409.

Back to schedule

Key distributions in the preservation of aging classes under the systems formation

Jorge Navarro[a], Tomasz Rychlik[b] and Magdalena Szymkowiak[c]
[a] *Universidad de Murcia, Campus de Espinardo, Murcia, 30100, Spain,*
[b] *Polish Academy of Sciences, Śniadeckich 8, 00656 Warsaw, Poland,*
[c] *Poznan University of Technology, Plac Marii Skłodowskiej-Curie 5, 60965 Poznań, Poland*

Back to schedule

We show that some distributions play a central role in the preservation of aging classes under the formation of semicoherent (or mixed) systems. Therefore, if an aging class of the component lifetimes is preserved in a system for the key distributions, then it is preserved for all distributions from the class. In the main aging classes (i.e., Increasing/Decreasing Failure Rate, Increasing/Decreasing Failure Rate Average, New Better/Worse than Used), the exponential distribution is crucial becasue it represents units without aging (with the lack of memory property). In other classes (e.g., Increasing/Decreasing Density) the uniform distribution is the key one. These distributions lead to mathematical properties that can also be used to determine whether an aging class of components lifetimes is inherited by a specific system. The results stated here can be applied to systems with independent or dependent identically distributed components.

References

[1] Arnold B.C., Rychlik T. and Szymkowiak M. (2022). Preservation of distributional properties of component lifetimes by system lifetimes. *Test* **31**, 901–930.

[2] Navarro J. (2022). *Introduction to System Reliability Theory.* Springer.

[3] Rychlik T. and Szymkowiak M. (2022). Signature conditions for distributional properties of system lifetimes if component lifetimes are iid exponential. *IEEE Trans. Reliab.* **71**, 590–602.

[4] Rychlik T. and Szymkowiak M. (2023). Preservation of transform order properties of component lifetimes by system lifetimes, submitted.

Back to schedule

On the Residual Lifetimes of Dependent Components Upon System Failure

Katherine Davies[a], and Anna Dembińska[b]
[a] *Department of Mathematics & Statistics, McMaster University, Hamilton ON Canada L8S 4K1*
[b] *Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland*

Back to schedule

In this talk, I consider coherent systems composed of identical yet possibly dependent components whose dependence structures are modeled via copulas. The main focus is on residual lifetimes of components that survived the failure of the system. I provide general formula describing the joint distribution of these residual lifetimes. I then look more closely at this distribution in the special case of Clayton copula and standard exponential marginals. Finally I present some numerical results and observations.

Back to schedule

Always used, never questioned: Defining dispersion for discrete distributions

Andreas Eberl[a] and Bernhard Klar[a]
[a]*Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany*

Back to schedule

Measures of dispersion are among the first topics covered in any introductory statistics course and are routinely used in all areas of application. The crucial defining property of a dispersion measure τ requires that it preserves some corresponding stochastic order $\preceq$ in the sense that $\tau(F) \leq \tau(G)$ whenever $F \preceq G$. The usual choice for the stochastic order underlying dispersion measures is the so-called dispersive order, which requires that all interquantile ranges of G are larger than the corresponding ranges of F. This is the strongest commonly used order of dispersion and is preserved by all well-known measures of dispersion. However, it is not meaningfully applicable to discrete distributions. This means that there is no theoretical guarantee that classical dispersion measures actually quantify the dispersion of a distribution. In this talk, we examine this problem in more detail and propose a solution in the form of a discrete dispersive order. Its construction is directly informed by key properties of the original dispersive order and its meaningfulness is illustrated using prototypical examples. After establishing a number of parallels between the original dispersive order and its discrete analogue, the compatibility of the discrete order with a number of popular dispersion measures is examined. This ensures that the application of these measures to discrete distributions is meaningful, with the notable exception being the interquantile range. Finally, the behaviour of our proposed discrete order on well-known families of discrete distributions is analyzed.

Back to schedule

On effects of dependence in variability estimation

Juan Baz[a], Franco Pellerey[b], Irene Díaz[c] and Susana Montes[a]

[a] *Department of Statistics and I.O. and Didactics of Mathematics, University of Oviedo, Spain,*
[b] *Dipartimento di Scienze Matematiche, Politecnico di Torino, Italy,*
[c] *Department of Computer Science, University of Oviedo, Spain*

Back to schedule

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ be two random samples that refer to a specific quantitative attribute in two different populations, say X and Y. Conditions on X and Y, i.e. on the marginal distributions of $\mathbf{X}$ and $\mathbf{Y}$, such that the estimators of different variability measures referring to X and Y are ordered in the usual stochastic order or in the increasing convex order, rather than only in the simple expectation order, have been described in Baz et al. (2024a). In there, the same dependence between the components of the two samples $\mathbf{X}$ and $\mathbf{Y}$ was mainly assumed (with the independence as a special case).

In this talk we consider the case where different structures of dependence exist between the components of $\mathbf{X}$ and between the components of $\mathbf{Y}$. For this case, conditions such that estimators of variability measures are stochastically ordered are described, pointing out the effects of the dependence in such estimators. The results presented here, mainly based on stochastic comparisons among the vectors $\mathcal{X} = \{|X_i - X_j|\}_{i,j\in\{1,\ldots,n\}}$ and $\mathcal{Y} = \{|Y_i - Y_j|\}_{i,j\in\{1,\ldots,n\}}$, also generalize in the multivariate setting some known results dealing with the comparison between absolute differences of identically distributed random variables.

References

[1] Baz, J., Pellerey, F., Diaz, I. and Montes, S. (2024a). Stochastic ordering of variability measure estimators. *Statistics*, **58**(1), 26–43.

[2] Baz, J., Pellerey, F., Diaz, I. and Montes, S. (2024b). On effects of dependence in variability estimation. *Submitted for publication*

Back to schedule

A variance-based importance index for systems with dependent components

Antonio Arriaza[a], Jorge Navarro[b], Miguel A. Sordo[a] and Alfonso Suárez-Llorens[a]
[a] *Universidad de Cádiz, Facultad de Ciencias*, [b] *Universidad de Murcia, Facultad de Matemáticas*

Back to schedule

Our work proposes a variance-based measure of importance for coherent systems with dependent and heterogeneous components. The particular cases of independent components and homogeneous components are also considered. We model the dependence structure among the components by the concept of copula. The proposed measure allows us to provide the best estimation of the system lifetime, in terms of the mean squared error, under the assumption that the lifetime of one of its components is known. We include theoretical results that are useful to calculate a closed-form of our measure and to compare two components of a system. We also provide some procedures to approximate the importance measure by Monte Carlo simulation methods. Finally, we illustrate the main results with several examples.

Back to schedule

Modelling Time-Varying Rankings

Vladimir Holý[a], Jan Zouhar[a]
[a]*Prague University of Economics and Business, Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic*

Back to schedule

1 Motivation

A recent survey of the ranking literature conducted by Yu et al. [4] highlights the absence of consideration for the time perspective in rankings and emphasizes the need for research in this direction. In response to this call, we introduce our paper Holý and Zouhar [2], which aims to address this gap and contribute to the limited body of literature on time-varying ranking data. In contrast to existing models for time variation in rankings, our approach endeavors to offer a flexible tool for modeling time-varying ranking data akin to the autoregressive moving average (ARMA) model employed for continuous variables.

2 Dynamic Score-Driven Ranking Model

Let us consider a set of N items $\mathcal{Y} = \{1, \ldots, N\}$. Our main object of interest is a complete permutation of this set, i.e. a ranking, $y_t = (y_t(1), \ldots, y_t(N))$ at time $t = 1, \ldots, T$. Element $y_t(i)$ represents the rank given to item i at time t, while r_t^{th} represents the item with rank r at time t.

We assume that a random permutation Y_t follows the Plackett–Luce distribution. See, e.g., Plackett [3] for more details. The probability of a complete ranking y_t is given by

$$\mathrm{P}\left[Y_t = y_t | f_t\right] = \prod_{r=1}^{N} \frac{\exp f_{r_t^{\text{th}},t}}{\sum_{s=r}^{N} \exp f_{s_t^{\text{th}},t}}, \tag{1}$$

where $f_t = (f_{1,t}, \ldots, f_{N,t})'$ are the worth parameters of the items at time t.

Let us assume that individual worth parameters linearly depend on covariates $x_1, \ldots, x_M$ and evolve over time according to the recursion

$$f_{i,t} = \omega_i + \sum_{j=1}^{M} \beta_j x_{i,t,j} + \alpha \nabla_i \left(f_{t-1} | y_{t-1}\right) + \varphi f_{i,t-1}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2}$$

where ω_i is the individual fixed effect of item i, β_j is the regression parameter on x_j, α is the score parameter, φ is the autoregressive parameter, $x_{i,t,j}$ is the value of x_j for item i at time t, and $\nabla_i \left(f_{t-1} | y_{t-1}\right)$ is the lagged score given by

$$\nabla_i \left(f_t | y_t\right) = 1 - \sum_{r=1}^{y_t(i)} \frac{\exp f_{i,t}}{\sum_{s=r}^{N} \exp f_{s^{\text{th}},t}}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{3}$$

The score is the gradient of the log-likelihood function and represents the direction for improving the fit of the distribution to a specific observation. Our model belongs to the class of score-driven models, also known as generalized autoregressive score (GAS) models and dynamic conditional score (DCS) models, introduced by Creal et al. [1].

The parameters of the model are estimated by the maximum likelihood method. The standard errors of the parameters are obtained by the empirical Hessian of the log-likelihood.

Table 1: Selected estimates for the Ice Hockey World Championships data (1998–2019).

	Mean-Reverting	Static	Random Walk
Home ice ($\hat{\beta}$)	0.227	0.171	0.099
	(0.258)	(0.262)	(0.188)
Score parameter ($\hat{\alpha}$)	0.392***		0.343***
	(0.083)		(0.058)
Autoregressive parameter ($\hat{\varphi}$)	0.506***		
	(0.149)		
log-likelihood	-611.195	-625.800	-625.425
AIC	1274.391	1299.600	1300.851

3 Application to Ice Hockey Rankings

We demonstrate the use of our model using data on the results of the Ice Hockey World Championships. The permutation y_t represents the final ranking of teams for year t. We also include a dummy variable indicating the host country of the championships in a given year. We compare the static, mean-reverting, and random walk specifications of the model. Main results are summarized in Table 1.

References

[1] Creal D., Koopman S. J., and Lucas A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, **28**(5), 777–795. doi: 10.1002/jae.1279

[2] Holý V. and Zouhar J. (2022). Modelling Time-Varying Rankings with Autoregressive and Score-Driven Dynamics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **71**(5), 1427–1450. doi: 10.1111/rssc.12584

[3] Plackett R. (1975). The Analysis of Permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **24**(2), 193–202. doi: 10.2307/2346567

[4] Yu P., Gu h., and Xu H. (2019). Analysis of Ranking Data. *Wiley Interdisciplinary Reviews: Computational Statistics*, **6**(11), 1483. doi: 10.1002/wics.1483

Back to schedule

On stochastic orders for multivariate scale mixtures of skew normal distributions with application to assess the evolution of summer temperatures in the Iberian Peninsula

Jorge M Arevalillo[a,b] and Jorge Navarro[c]
[a] *aUC3M-Santander Big Data Institute, Madrid Street 135, 28903 Getafe, Madrid, Spain*, [b] *Department of Statistics and Operational Research, UNED, Juan del Rosal 10, 28040 Madrid, Spain*, [c] *Department of Statistics and Operational Research, Facultad de Matem ´aticas, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain*

Back to schedule

One of the main characteristics of temperature data is that they exhibit asymmetries since the records collected by stations correspond to maximum daily temperature values. Hence, temperature data are far away from normality so non-normal multivariate distributions must be employed to model and capture their non-normal features. Scale mixtures of skew normal (SMSN) distributions are flexible models that account for asymmetry and tail weight behavior simultaneously. This work is concerned with the stochastic comparison of vectors belonging to the scale mixtures of skew normal family. Several stochastic orders are proposed to carry out tail weight stochastic comparisons of SMSN vectors. We also investigate the relationships between the proposed orders and the non-normality parameters of some popular distributions within the SMSN family. Our theoretical results are useful to address the stochastic comparison of extreme temperature records; we illustrate them with an application to a real study about summer temperatures in the Iberian Peninsula during the last century by using the multivariate skew-t distribution. The conclusions of this application will shed light on the evolution of extreme summer temperature records for such a long period.

The data used in this work come from the Spanish TEmperature At Daily scale (STEAD) dataset [2] which can be downloaded from the Spanish CSIC repository located at the website: *https://digital.csic.es/handle/10261/177655*. All the theoretical details about the findings presented in this work appear in a recent contribution by both authors [1].

Funding

This research was supported by SocialProbing project (TED2021-131264B-I00), which has been funded by MCIN/AEI /10.13039/501100011033 and the European Union-NextGenerationEU/PRTR. JMA also acknowledges.

References

[1] Jorge M Arevalillo, Jorge Navarro (2024). Assessment of extreme records in environmental data through the study of stochastic orders for scale mixtures of skew normal vectors. *Environmental and Ecological Statistics* **31**, 151-–179.

[2] Roberto Serrano-Notivoli, Santiago Beguería, and Martín de Luis (2019). STEAD: a highresolution daily gridded temperature dataset for Spain. *Earth System Science Data* **11**, 1171—1188.

Back to schedule

Analysis of the economic situation of people with different levels of education based on the Zenga distribution

Kamila Trzcińska[a], Elizabeta Zalewska[b]
[a]*Department of Statistical Methods, University of Lodz, 41 Rewolucji 1905 r. St., 92-2014 Lodz, Poland,*
[b]*Department of Statistical Methods, University of Lodz, 41 Rewolucji 1905 r. St., 92-2014 Lodz, Poland*

Back to schedule

Household income is one of the most important economic categories depending on various factors, in particular on the level of education of the head of the household. Therefore, the question arises how the level of education affects the distribution of household income. A lot of research has been directed at describing empirical distributions by using a theoretical model. In the literature there are proposals for various types of mathematical functions. In 2010 Zenga [1] proposed a new three-parameter model for economic size distribution which possesses interesting statistical properties that can be used to model income, wealth and financial variables.
The probability density function $f(x;\mu,\alpha,\theta)$, $(\mu > 0, \alpha > 0, \theta > 0)$ of the Zenga distribution for non-negative variables has the form:

$$f(x;\mu,\alpha,\theta) = \begin{cases} \dfrac{1}{2\mu B(\alpha;\theta)} \left(\dfrac{x}{\mu}\right)^{-3/2} \displaystyle\int_0^{\frac{x}{\mu}} k^{\alpha-1/2}(1-k)^{\theta-2}dk & \text{for } 0 < x < \mu \\ \dfrac{1}{2\mu B(\alpha;\theta)} \left(\dfrac{\mu}{x}\right)^{3/2} \displaystyle\int_0^{\frac{\mu}{x}} k^{\alpha-1/2}(1-k)^{\theta-2}dk & \text{for } x > \mu \end{cases} \tag{1}$$

where $B(\alpha;\theta)$ is the beta function.

In this model μ is the scale parameter and is equal to the expected value, α and θ are shape parameters that the inequality depends on. Studies performed in various countries show that the Zenga distribution exhibits high conformance to the empirical distributions of incomes [1, 2, 3]. The parameters estimates of the Zenga distribution were obtained by means of the D'Addario's invariants methods.

The Zenga distribution was applied to the estimation of point and synthetic inequality measures. We analyzed the distribution of household income and income inequality, taking into account the education level of the head of the household using the Zenga model and the decomposition of the Theil coefficient, which is an important and current socio-economic problem. Household income data from the European Central Bank (ECB)-Eurosystem Household Finance and Consumption (HFCS) wave 2021 were used to describe and analyze the economic situation of household income of people with different levels of education in selected European countries with different economic models. The empirical analysis results unveiled both commonalities and significant distinctions among the countries. Conducted empirical research on various samples significantly complements the existing knowledge on this topic.

References

[1] Zenga, M.M., Pasquazzi, L., Zenga, Ma. (2010). First Applications of a New Three Parameter Distribution for Non-Negative Variables. *Statistica & Applicazioni* **X(2)**, 131–149.

[2] Trzcińska, K. (2022). An Analysis of Household Income in Poland and Slovakia Based on Selected Income Models. *Folia Oeconomica Stetinensia* **22(1)**, 287–301.

[3] Ćwiek, M., Trzcińska, K. (2023). Assessment of goodness of fit of income distribution in France and Germany based on the Zenga distribution. *Quality& Quantity*. **57(5)**,4013–4027.

Back to schedule

Are the order statistics ordered? A copula approach.

Jorge Navarro[a]
[a] *Facultad de Matemáticas, Universidad de Murcia, 30100 Murcia, Spain*

Back to schedule

If $X_1, \ldots, X_n$ are the random variables representing some sample values, the associated (increasing) ordered values $X_{1:n} \leq \cdots \leq X_{n:n}$ are know as *order statistics*. Several properties for them in the case of IID (independent and identically distributed) samples can be seen in [1].

If $X_1, \ldots, X_n$ represent the component lifetimes of a system, then the ordered values $X_{1:n} \leq \cdots \leq X_{n:n}$ represent the lifetimes of k-out-of-n systems (systems that works when at least k components work). In this case, the Xs can be dependent and they can be heterogeneous, that is, they are not ID.

In both cases, to get stochastic comparisons for them is a relevant topic. The main stochastic orders are the stochastic (st), hazard rate (hr) and likelihood ratio (lr) orders. The main properties for them can be seen in [5].

In some case the order statistics are ordered (as expected). However, this is not always the case. In this talk we will present conditions to get stochastic comparisons in these orders for the order statistics. The conditions will depend on the copula C that models the dependence between $X_1, \ldots, X_n$. The conditions will be connected with dependence properties of C. The talk is based on the results published in [2, 3, 4].

Acknowledgement

JN thanks the partial support of Ministerio de Ciencia e Innovación of Spain under grant PID2022-137396NB-I00 and grant TED2021-129813A-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union 'NextGenerationEU'/PRTR.

References

[1] Arnold B.C., Balakrishnan N., Nagaraja, H.N. (2008). *A First Course in Order Statistics.* SIAM.

[2] Navarro J. (2022). *Introduction to System Reliability Theory.* Springer.

[3] Navarro J., Durante F., Fernández-Sánchez J. (2021) Connecting copula properties with reliability properties of coherent systems. *Applied Stochastic Models in Business and Industry* **37**, 496–512.

[4] Navarro J., Torrado N., del Águila Y. (2018). Comparisons between largest order statistics from multiple-outlier models with dependence. *Methodology and Computing in Applied Probability* 20, 411–433.

[5] Shaked M., Shanthikumar, J.G. (2007). *Stochastic Orders.* Springer, New York.

Back to schedule

Bounds on Variances of Generalized Order Statistics

Tomasz Rychlik[a]
[a] *Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00 656 Warsaw, Poland*

Back to schedule

For an arbitrary number of generalized order statistic and arbitrarily fixed its parameters we set out sharp upper and lower bounds on the variance of generalized order statistic valid for all baseline distributions with finite second moments. The bounds are measured in the scale units being the variances of baseline distributions. We also discuss the implications of these results for the specific submodels of the generalized order statistics model: classic order statistics, record values and progressively censored type II order statistics.

Back to schedule

Nonparametric Estimators for A Binary Outcome Under A Monotonicity Restriction

X.L. Gu[a], Gong Tang[a] and G. Yu[a]
[a] University of Pittsburgh, 130 DeSoto Street, Department of Biostatistics, Pittsburgh, PA 15261 USA

Back to schedule

1 Overview

In randomized controlled clinical trials, the post-stratification causal framework is used to study the causal relationship among the treatment, an intermediate outcome and the long-term outcome. It requires modeling the distribution of the intermediate outcome given a baseline covariate under a monotonicity assumption, that is, responders to the control would respond to the experimental treatment [1]. Estimation with a discrete covariate under the monotonicity assumption is well understood. However, regression analysis with a binary intermediate outcome and a continuous covariate under the monotonicity restriction has not been established. Here we proposed a few nonparametric methods to predict the binary potential outcomes given a continuous covariate under the monotonicity restriction, including the logistic regression with P-splines and a weighted least square kernel estimator. We examined the asymptotic properties and finite-sample performance of the proposed kernel estimator and generalized it to settings with multiple continuous covariates in an application to neoadjuvant trials on early-stage breast cancer.

2 Main Results

Consider variables (X, Z, S), where $Z, S \in \{0, 1\}$ and X is a continuous variable. The observed data are $\{x_i, z_i, s_i;\ \ i = 1, 2, \ldots, n\}$.We are interested in the following nonparametric regression:

$$\pi_z(x) = Pr[S = 1|X = x, Z = z] \tag{1}$$

subject to the restriction: $\pi_0(x) \leq \pi_1(x)$, for all x.

If you need to state theorems or other mathematical bodies of the sort, please use the predefined environments:

Theorem 1 *Under some regularity conditions, for any x, consider the following weighted least square kernel estimates:*

$$(\tilde{\pi}_0(x), \tilde{\pi}_1(x)) = \text{argmin}_{\beta_0,\beta_1} \sum_{i:z_i=0} (s_i - \beta_0)^2 K_h(x_i - x) + \sum_{i:z_i=1} (s_i - \beta_1)^2 K_h(x_i - x), \tag{2}$$

where $K_h(t) = \frac{1}{h}K(\frac{t}{h}), K(\cdot)$ is a kernel function. Denote:

$$\hat{\pi}_z(x) = \frac{\sum_{i:z_i=z} K_h(x_i - x)s_i}{\sum_{i:z_i=z} K_h(x_i - x)}, \quad z = 0, 1. \tag{3}$$

With the restriction: $\pi_0(x) \leq \pi_1(x)$, the weighted least square kernel estimates are:

$$\tilde{\pi}_z(x) = \begin{cases} \hat{\pi}_z(x) & \text{if } \hat{\pi}_0(x) \leq \hat{\pi}_1(x) \\ \hat{\pi}_x & \text{otherwise} \end{cases}$$

where $\hat{\pi}(x) = \frac{\sum_{i=1}^n K_h(x_i-x)s_i}{\sum_{i=1}^n K_h(x_i-x)}$ is the pooled Nadaraya-Watson estimator.

References

[1] Shepherd B.E., Gilbert P.B., Jemiai Y., Rotnitzky A. (2006). Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics* **62**,2:332–342.

Back to schedule

Nonparametric tests in deconvolution

Mohammed E. Benjrada[a]
[a] *Department of Management, Economics and Quantitative Methods, University of Bergamo (UNIBG), Italy*

Back to schedule

Introduction

We consider the problem of estimation and tests using data that are contaminated by additive noise $\{\varepsilon_i\}_{i=1}^n$. Actually, due to the nature of the experimental environment or the measuring tools, the random process $\{X_i\}_{i=1}^n$ is not available for direct observation. Instead of X_i, we only observe the random variables Y_i given by

$$Y_i \triangleq X_i + \varepsilon_i, \quad i = 1, \ldots, n. \tag{1}$$

Model (1) is called a convolution and the problem of estimating with this model occurs in various domains. This model has been studied in Experimental Sciences. For example, Biological Organisms (see [9]); Communication Theory (see [6]) and Applied Physics (see [10]).

The literature abounds of work devoted to the study of the p.d.f. in convolution problems. [12] proposed a consistent estimator for the density based on grouped data for some cases of error density. [8] considered the estimation of the multivariate probability density functions under some structures of dependence. [11] used the Moving Polynomial Regression (MPR) to smooth the empirical distribution function estimator. [7] considered the asymptotic uniform confidence bands.

We study the deconvolution under the assumption that characteristic function ϕ_ε of the measurement error $\{\varepsilon_i\}_{i=1}^n$ decays algebraically at infinity i.e:

$$|t|^\beta |\phi_\varepsilon(t)| \underset{|t|\to+\infty}{\to} \beta_1 \text{ for some } \beta > 0 \text{ and } \beta_1 > 0.$$

Here, the error is called ordinary smooth. The parameter β is called the order of the noise density $f_\varepsilon(x)$. Actually, it has a direct impact on the rate of convergence of the function to be estimated. The ordinary smooth distribution covers in particular the case of Gamma, Double Exponential, and Symmetric Gamma densities $f_\varepsilon(x)$.

The goal here is to test the concavity of the function to be estimated. It is important to test this, for example, if the target function is a concave distribution function F_X, then, the distribution function of the observed variable $\{Y_i = X_i + \varepsilon_i\}_{i=1}^n$ can be non-concave. Take the example where X_i and ε_i are uniform then F_X is concave but F_Y is triangular (not concave).

To test this, we measure the distance between the estimator of the target function and its least concave majorant (LCM for short). We give the definition of the LCM operator.

Definition 1 *Given a convex interval $I \subseteq \mathbb{R}^+$, the LCM over I is the operator $\mathcal{M}_I : \ell_I^\infty \to \ell_I^\infty$ that maps each function $\theta \in \ell_I^\infty$ to $\mathcal{M}_I\theta(x)$ where*

$$\mathcal{M}_I\theta(x) := \inf\{\theta(x) : \theta \text{ is a concave function lies in } \ell_I^\infty \text{ and } \theta \le \theta \text{ on } I\}, \, x \in I.$$

For convenience, we write $\mathcal{M}$ to refer to $\mathcal{M}_{\mathbb{R}+}$

To have the asymptotic distribution of the test statistic, one should use the well-known Functional Delta Method (see, e.g., van der Vaart and Wellner [3]). This approach requires the operator under study to be Hadamard differentiable. However, this approach (Functional Delta

Method) is more widely applicable under Hadamard directional differentiability, as demonstrated by Shapiro [5] (see Thm. 2.1). The following is the definition of a Hadamard directionally differentiable map, according to Shapiro ([4], [5]) and Bonnans and Shapiro [2].

Definition 2 *Let A and B be topological vector spaces of finite dimension over $\mathbb{R}$ provided respectively with the norms $\|\cdot\|_A$ and $\|\cdot\|_B$. A map $\phi : A \to B$ is said to be Hadamard directionally differentiable at $\theta \in A$ tangentially to $A_0 \in A$ if there exists a map $\phi'_\theta : A_0 \to B$ such that*

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + t_n g_n)}{t_n} - \phi'_\theta(g) \right\|_B \to 0,$$

for all $g \in A_0$ and all $g_1, g_2, \ldots \in A$ and $t_1, t_2, \ldots \in \mathbb{R}^+$ such that $\|g_n - g\|_A \to 0$ and $t_n \downarrow 0$. The map $\phi'_\theta(g)$ is called the Hadamard directional derivative of ϕ at θ tangentially to A_0

References

[1] Brendan K. and Zheng F (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *E. J. of Stat.* **11**, 3841–3870.

[2] Bonnans J.F. and Shapiro A. (2000). *Perturbation Analysis of Optimization Problems.* Springer, New York.

[3] Van Der Vaart A. W and Wellner J. A. (1997). *Weak convergence and empirical processes: with applications to statistics.* Springer, New York.

[4] Shapiro A. (1990). On concepts of directional differentiability. *J. Optim. Theory Appl.* **66**, 477–487.

[5] Shapiro A. (1991). Asymptotic analysis of stochastic programs. *Ann. Oper. Res.***30**, 169–186.

[6] Wise G., Traganitis A., and Thomas J. (1977). The estimation of a probability density function from measurements corrupted by Poisson noise. *IEEE Trans. Inf. Theory* **23**, 764–766.

[7] Fan J. (1991). Asymptotic Normality for Deconvolution Kernel Density Estimators. *Sankhyā (A)* **53**, 97–110.

[8] Masry E. (1991) Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stoch. Process. Appl.* **53**, 53–74.

[9] Medgyessy P. (1977). *Decomposition of Superposition of Density Functions and Discrete Distributions.* John Wiley & Sons, New York.

[10] Snyder D. L., Miller M. I., and Schultz T. J. (1988). Constrained probability density estimation from noisy data. *Proc. 22nd Annu. Conf. Inf. Sci. Syst.*, 170–172.

[11] Sarda P. and Vieu Ph. (1991). Smoothing parameter selection in hazard estimation. *Stat. Probab. Lett.* **11**, 429–434.

[12] Zhang C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, 806—831.

Back to schedule

Distribution-Free Test on Treatment Effects with Multivariate Ordered Data: a Biostatistical Application

Tomasso Lando[a]

[a] *Department of Management, Economics and Quantitative Methods, University of Bergamo (UNIBG), Italy*

Back to schedule

Given samples from two non-negative random variables, we propose a new class of nonparametric tests for the null hypothesis that one random variable dominates the other with respect to second-order stochastic dominance. These tests are based on the Lorenz P-P plot (LPP), which is the composition between the inverse unscaled Lorenz curve of one distribution and the unscaled Lorenz curve of the other. The LPP exceeds the identity function if and only if the dominance condition is violated, providing a rather simple method to construct test statistics, given by functionals defined over the difference between the identity and the LPP. We determine a stochastic upper bound for such test statistics under the null hypothesis, and derive its limit distribution, to be approximated via bootstrap procedures. We also establish the asymptotic validity of the tests under relatively mild conditions, allowing for both dependent and independent samples. Finally, finite sample properties are investigated through simulation studies.

Back to schedule

Order restricted inferences for $R = P(X_1 < Y < X_2)$ based on the Weibull distribution under joint progressive censoring

Çağaatay Çetinkaya[a]
[a] Department of Actuarial Sciences, Kırıkkale University, 71450, Kırıkkale, Turkey

Back to schedule

Let Y represent the strength variable effected by the stress variables X_1 and X_2. Assume that three samples are independent. Thus, $R = P(X_1 < Y < X_2)$ denotes the reliability where the strength Y should not only be greater than stress X_1 but also smaller than stress X_2 [3]. Estimation of R has extensive applications in various areas since it provides a useful measure of differences between or among populations. In this study, the Weibull distributions with common shape parameters and different scale parameters are assumed to be the underlying distributions of the components. These three components are assumed to be performed under joint progressive type-II censoring scheme introduced by Balakrishnan et al. [1]. Further, a natural constraint on the scale parameters such as $\lambda_1 < \lambda_2 < \lambda_3$ considered. Thus, the inference of R is obtained under jointly progressive censored data under order-restricted scale parameters. The maximum likelihood estimations are obtained from the findings of the generalized isotonic regression problem defined by Brunk et al. [2]. Additionally, the Bayesian estimation is obtained under gamma-Dirichlet prior distribution performing the importance sampling algorithm. The approximate confidence intervals, including asymptotic and highest posterior density intervals, are also derived for R.

References

[1] Balakrishnan N., Su F., and Liu K.Y. (2015). Exact likelihood inference for k exponential populations under joint progressive type-II censoring. *Communications in Statistics-Simulation and Computation* **44(4)**, 902–923.

[2] Brunk H., Bartholomew D., and Bremner J. (1972). *Statistical Inference under Order Restrictions*. John Wiley & Sons, New York.

[3] Singh N. (1980). On the estimation of $Pr(X_1 < Y < X_2)$. *Communications in Statistics-Theory and Methods* **9(15)**, 1551–1561.

Back to schedule

Trend tests based on the ordered heterogeneity test

M. Neuhäuser
[a] *Koblenz University of Applied Sciences, Dept. of Mathematics and Technology, RheinAhrCampus, Remagen, Germany*

Back to schedule

Tests for trend are common in practice when several independent groups with increasing doses are investigated for example in clinical trials or agricultural field studies. In order to construct a trend test one can combine a non-directional heterogeneity test with the rank-order information under the alternative (Rice and Gaines, 1994). Neuhäuser and Hothorn (2006) proposed two modifications of this ordered heterogeneity test procedure. On the one hand, the maximum correlation out of the $2^{k-1} - 1$ possibilities under the alternative can be used instead of a single ordering such as $(1, 2, \ldots, k)$, where k denotes the number of groups. On the other hand, the mean ranks of the groups rather than the sample means can be used in order to determine the observed ordering of the groups. These two modifications can increase the power of the ordered heterogeneity test. Moreover, the modified ordered heterogeneity tests are quite robust and can detect all patterns that are possible under the alternative with relatively high power.

The modified ordered heterogeneity tests can be extended to many trend test situations when using permutation or bootstrap for inference. In particular, we shall present results for non-normal distributions and heterogeneous variances.

References

[1] Neuhäuser M., Hothorn L.A. (2006). A robust modification of the ordered-heterogeneity test. *Journal of Applied Statistics* **33**, 721–727.

[2] Rice W.R., Gaines, S.D. (1994). The Ordered-Heterogeneity Family of Tests. *Biometrics* **50**, 746—752.

Back to schedule

Some relations for single and product moments of order statistics from K3D

İsmet Birbiçer[a], Ali İ. Genç[a]
[a] *Department of Statistics, Cukurova University, Turkey*

Back to schedule

The four-parameter kappa distribution, as defined by Hosking [1], includes four distinct extreme value models and has extensive applications in various fields such as environmental sciences and hydrology. As a special case, the three-parameter kappa distribution, also referred to as K3D, is previously overlooked in the literature until Jeong [2] undertook a comprehensive study on it. In this talk, we consider the K3D for order statistics point of view and obtain moments of order statistics from it. Moreover, we derive some useful recurrence relations for calculating both single and product moments of order statistics. As an application of these relations, we use these calculations to derive the best linear unbiased estimator (BLUE) for estimating the location and scale parameters of the distribution. Finally, we demonstrate a representative data fitting to see the performance of the BLUE's versus the maximum likelihood estimates.

References

[1] Hosking, J. R. (1994). The four-parameter kappa distribution. *IBM Journal of Research and Development* **38**, 251—258.

[2] Jeong, B.Y., Murshed, M. S., Seo, Y.A., Park, J.S. (2014). A three-parameter kappa distribution with hydrologic application: a generalized Gumbel distribution. *Stochastic Environmental Research and Risk Assessment* **28**, 2063–2074.

Back to schedule

Enhancing Statistical Inference Through Post-Stratification in Completely Randomized Designs

Omer Ozturk[a], Olena Kravchuk[b] and Richard Jarrett[b]
[a] *The Ohio State University, Department of Statistics, Columbus, OH, USA*, [b] *School of Agriculture, Food & Wine, University of Adelaide, SA, Australia*

Back to schedule

In this study, we introduce a novel approach to statistical inference for Completely Randomized Design (CRD) by employing post-stratification of experimental units. After the experiment is completed, our method involves randomly pairing experimental units (EUs) subjected to two different treatments, identified as treatment h and treatment h'. Each pair of EUs is ranked based on their inherent variation. Ranking process is conducted blindly to the treatment assignments to prevent bias. Subsequently, these ranked sets are divided into two distinct categories: one where the higher rank corresponds to treatment h and the lower rank to treatment h', and the other with the reverse treatment allocation. This post-stratification is performed for all possible pairs of treatments (h and h', $h < h'$). By ranking within each pair, we induces positively correlated judgment order statistics between response variables from EUs receiving treatments h and h', thereby facilitating a significant reduction in the variance of the estimated contrast parameter ($\Delta = \mu_h - \mu_{h'}$), where μ_h and $\mu_{h'}$ represent the mean of treatments h and h', respectively. Our findings demonstrate a substantial variance reduction in the estimation of the contrast parameter. We also develop a multiple comparison procedure for pairwise differences of contrast parameters.

Back to schedule

Limit theorems for multiple orderings of multidimensional data

Artem Kovalevskii[a], Mikhail Chebunin[b]
[a]*Sobolev Instutute of Mathematics, 4 Koptjuga, Novosibirsk 630090, Russia,* [b]*Karlsruhe Institute of Technology, 76131, Karlsruhe, Germany*

Back to schedule

1 Introduction

We study multiple orderings of multidimensional data, that is, independent copies of a random vector are sequentially ordered in ascending order of several of its components. The result is a sequence of vectors of higher dimension, consisting of induced order statistics (concomitants) corresponding to different orderings. Our Lemma 1 deals with weak convergence of random fields under consideration to corresponding Gaussian random fields. Lemma 2 describes weak convergence of the process of sequential sums of random vectors under different orderings to a centered Gaussian process of corresponding dimension. Next, we assume a linear relationship of the components, use standard least squares estimates to compute regression residuals, that is, the differences between response values and the predicted ones by the linear model. We prove Theorem 3 about weak convergence of the process of sums of of regression residuals under the necessary normalization to a centered Gaussian process.

Regression analysis deals with models for the linear dependence of one variable (response) on other variables (regressors). However, standard regression analysis methods do not include methods of detecting that the proposed linear model is incorrect entirely. If the model is incorrect then it must either be completely discarded or substantially modified. This is a change points situation of presence of structural breaks in a data. MacNeill (1978) proposed a change point test for time series, and Bishoff (1998) significantly relaxed the assumptions of MacNeill. An analysis of results in this direction can be found in Csorgo and Horváth (1997, Chapters 2 and 3) and MacNeill et al. (2020). Kovalevskii (2020) proposed tests for matching of regression models using data ordering by one of the regressors. We propose a statistical test that uses multiple ordering of data. Proofs can be found in Chebunin and Kovalevskii (2021).

2 Induced order statistics

Let $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, 2 \ldots$, be the independent copies of a random vector $(\mathbf{X}, \mathbf{Y})$ such that $\mathbf{X} = (X^{(1)}, \ldots, X^{(d_1)})$ takes values in $[0,1]^{d_1}$, $\mathbf{Y}$ takes values in $\mathbf{R}^{d_2}$. The distribution function (copula) of $\mathbf{X}$ is $C(\mathbf{u}) = \mathbf{P}(\mathbf{X} \leq \mathbf{u}) = \mathbf{P}\left(X^{(1)} \leq u^{(1)}, \ldots, X^{(d_1)} \leq u^{(d_1)}\right)$, $\mathbf{u} = (u^{(1)}, \ldots, u^{(d_1)}) \in [0,1]^{d_1}$.

We assume that there is copula density $c(\mathbf{u})$, that is, $C(\mathbf{u}) = \int_{\mathbf{v} \leq \mathbf{u}} c(\mathbf{v}) d\mathbf{v}$, $\mathbf{v} \in [0,1]^{d_1}$. We define the weak convergence of random fields in the d_1-dimensional analog of Skorohod metrics, see Straf (1972). Davydov & Zitikis (2008) proposed a method of proving this convergence. We use the symbol $\Rightarrow$ to denote the weak convergence of random fields in the sense that has been mentioned above. We use the same symbol for the weak convergence of random variables and the weak convergence of stochastic processes in the Skorohod topology. Let

$$\mathbf{Q}_n(\mathbf{u}) = \sum_{j=1}^{n} \mathbf{Y}_j \mathbf{1}\left(\mathbf{X}_j \leq \mathbf{u}\right) = \sum_{j=1}^{n} \mathbf{Y}_j \mathbf{1}\left(X_j^{(1)} \leq u^{(1)}, \ldots, X_j^{(d_1)} \leq u^{(d_1)}\right), \ \mathbf{u} \in [0,1]^{d_1},$$

$$\mathbf{m}(\mathbf{u}) = \mathbf{E}(\mathbf{Y} \mid \mathbf{X} = \mathbf{u}), \quad \mathbf{u} \in [0,1]^{d_1},$$

and

$$\mathbf{f}(\mathbf{u}) = \int_{\mathbf{0}}^{\mathbf{u}} \mathbf{m}(\mathbf{v}) c(\mathbf{v}) d\mathbf{v},$$

$$\sigma^2(\mathbf{u}) = \mathbf{E}\left\{(\mathbf{Y} - \mathbf{m}(\mathbf{X}))^T(\mathbf{Y} - \mathbf{m}(\mathbf{X})) \mid \mathbf{X} = \mathbf{u}\right\}$$

be the conditional covariance matrix of $\mathbf{Y}$ and $\sigma(\mathbf{u})$ be the positive definite matrix such that $\sigma(\mathbf{u})^T \sigma(\mathbf{u}) = \sigma^2(\mathbf{u})$. The following Lemma 1 generalizes the result of the first part of Theorem 2.1(1) by Davydov and Egorov (2000) to random fields.

Lemma 3 *If* $\mathbf{E}\|\mathbf{Y}\|^2 < \infty$ *then* $\widetilde{\mathbf{Q}}_n = \frac{\mathbf{Q}_n - \mathbf{f}}{\sqrt{n}} \Rightarrow \mathbf{Q}$*, a centered Gaussian field with covariance*

$$K(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{E}\mathbf{Q}^T(\mathbf{u}_1)\mathbf{Q}(\mathbf{u}_2) = \int_{\mathbf{0}}^{\min(\mathbf{u}_1, \mathbf{u}_2)} \sigma^2(\mathbf{v}) c(\mathbf{v}) d\mathbf{v}$$

$$+ \int_{\mathbf{0}}^{\min(\mathbf{u}_1, \mathbf{u}_2)} \mathbf{m}^T(\mathbf{v})\mathbf{m}(\mathbf{v}) c(\mathbf{v}) d\mathbf{v} - \int_{\mathbf{0}}^{\mathbf{u}_1} \mathbf{m}^T(\mathbf{v}) c(\mathbf{v}) d\mathbf{v} \int_{\mathbf{0}}^{\mathbf{u}_2} \mathbf{m}(\mathbf{v}) c(\mathbf{v}) d\mathbf{v}, \qquad \mathbf{u}_1,\ \mathbf{u}_2 \in [0,1]^{d_1}.$$

Denote $X_{n,1}^{(k)} \le X_{n,2}^{(k)} \le \cdots \le X_{n,n}^{(k)}$, $1 \le k \le d_1$, the order statistics of the k-th column of matrix X, and $\mathbf{Y}_{n,1}^{(k)}, \mathbf{Y}_{n,2}^{(k)}, \ldots, \mathbf{Y}_{n,n}^{(k)}$ the corresponding values of the vectors $\mathbf{Y}_i$. The random vectors $\left(\mathbf{Y}_{n,i}^{(k)}, i \le n\right)$ are called induced order statistics (concomitants). Let $\mathbf{e}_{k,t} = (1, \ldots, 1, t, 1, \ldots, 1)$ the vector in $[0,1]^{d_1}$ with k-th coordinate being t and other coordinates being 1.

Lemma 4 *If* $\mathbf{E}\|\mathbf{Y}\|^2 < \infty$*,* $\mathbf{m} \equiv \mathbf{0}$ *then* $\widetilde{\mathbf{Z}}_n = \frac{\mathbf{Z}_n}{\sqrt{n}} \Rightarrow \mathbf{Z}$*, a centered Gaussian* $(d_1 \times d_2)$*-dimensional process with covariance matrix function* $\mathbf{E}\mathbf{Z}^T(t_1)\mathbf{Z}(t_2) = (K(\mathbf{e}_{k_1,t_1}, \mathbf{e}_{k_2,t_2}))_{k_1,k_2=1}^{d_1}$*,*

$$K(\mathbf{e}_{k_1,t_1}, \mathbf{e}_{k_2,t_2}) = \mathbf{E}\mathbf{Q}^T(\mathbf{e}_{k_1,t_1})\mathbf{Q}(\mathbf{e}_{k_2,t_2}) = \int_{\mathbf{0}}^{\min(\mathbf{e}_{k_1,t_1}, \mathbf{e}_{k_2,t_2})} \sigma^2(\mathbf{v}) c(\mathbf{v}) d\mathbf{v}.$$

Let $(\mathbf{X}_i, \xi_i, \eta_i) = (X_{i1}, \ldots, X_{id_1}, \xi_{i1}, \ldots, \xi_{i,d_2-1}, \eta_i)$ be independent and identically distributed random vector rows, $i = 1, \ldots, n$. All components of a raw can be dependent and $X_{i1}, \ldots, X_{id_1}$ have a copula (so their marginal distributions are uniform on [0, 1]) with a density $c(\mathbf{v})$. Rows $(\mathbf{X}_i, \xi_i, \eta_i)$ form matrix (X, ξ, η). We assume a linear regression hypothesis H_0:

$$\eta_i = \xi_i \theta + \varepsilon_i = \sum_{j=1}^{d_2-1} \xi_{ij}\theta_j + \varepsilon_i, \tag{1}$$

$\{\varepsilon_i\}_{i=1}^n$ and $\{(\mathbf{X}_i, \xi_i)\}_{i=1}^n$ are independent, $\mathbf{E}\,\varepsilon_1 = 0$, $\mathbf{Var}\,\varepsilon_1 > 0$. Vector $\theta = (\theta_1, \ldots, \theta_{d_2-1})^T$ and constant $\mathbf{Var}\,\varepsilon_1$ are unknown. We consider d_1 orderings of rows of the matrix (X, ξ, η) in acsending order of columns of X. The result of d_1 orderings is a sequence of d_1 matrices $(X^{(j)}, \xi^{(j)}, \eta^{(j)})$ with rows $(\mathbf{X}_i^{(j)}, \xi_i^{(j)}, \eta_i^{(j)}) = (X_{i1}^{(j)}, \ldots, X_{id_1}^{(j)}, \xi_{i1}^{(j)}, \ldots, \xi_{i,d_2-1}^{(j)}, \eta_i^{(j)})$, $j = 1, \ldots, d_1$. Let $\widehat{\theta}$ be LSE: $\widehat{\theta} = (\xi^T\xi)^{-1}\xi^T\eta$. It does not depend on the order of rows. Let $h^{(j)}(x) = \mathbf{E}\{\xi_1 | X_{1j} = x\}$ be conditional expectations, $L^{(j)}(x) = \int_0^x h^{(j)}(s)\, ds$ be *induced theoretical generalised Lorentz curves* (see Davydov and Egorov (2000)), $b_j^2(x) = \mathbf{E}\left((\xi_1 - h^{(j)}(x))^T(\xi_1 - h^{(j)}(x)) \mid X_{1j} = x\right)$ be matrices of conditional covariances. Let $G = \mathbf{E}\xi_1^T\xi_1$. Then

$$\int_0^1 \left(b_j^2(x) + (h^{(j)}(x))^T h^{(j)}(x)\right) dx = G$$

for any $j = 1, \dots, d_2 - 1$. Let $\widehat{\varepsilon}_i^{(j)} = \eta_i^{(j)} - \xi_i^{(j)}\widehat{\theta}$ be regression residuals, $\widehat{\Delta}_k^{(j)} = \sum_{i=1}^{k} \widehat{\varepsilon}_i^{(j)}$ be its partial sums, $\widehat{\Delta}_0^{(j)} = 0$. Let $\widehat{Z}_n^{(j)} = \{\widehat{Z}_n^{(j)}(t),\, 0 \le t \le 1\}$ be a piecewise linear random function with nodes

$$\left(\frac{k}{n},\ \frac{\widehat{\Delta}_k^{(j)}}{\sqrt{n\mathbf{Var}\varepsilon_1}}\right),\quad k = 0,\ 1, \dots,\ n.$$

The next theorem generalizes Theorem 1 in Kovalevskii (2020) to the multidimensional case.

Theorem 5 *If matrix G exists and is non-degenerate and H_0 is true then $\widehat{Z}_n \Longrightarrow \widehat{Z}$. Here $\widehat{Z}$ is a centered d_1-dimensional Gaussian process with continuous a.s. sample paths and covariance matrix function* $\widehat{K}(s,t) = \left(\widehat{K}_{ij}(s,t)\right)_{i,j=1}^{d_1}$,

$$\widehat{K}_{ij}(s,t) = \mathbf{P}(X_{1i} \le s, X_{1j} \le t) - L^{(i)}(s)G^{-1}(L^{(j)}(t))^T, \quad s,t \in [0,1].$$

We discuss applications of this result to testing the hypothesis of linear dependence.

Acknowledgement The work is supported by the Mathematical Center in Akademgorodok under the agreement N. 075-15-2022-281 with the Ministry of Science and Higher Education of the Russian Federation.

References

[1] Bischoff, W., 1998. *A functional central limit theorem for regression models*, Ann. Stat. **26**, 1398–1410. MR1647677

[2] Chebunin, M. G., and Kovalevskii, A.P., 2021. *Asymptotics of sums of regression residuals under multiple ordering of regressors*, Siberian Electronic Mathematical Reports **18**(2), 1482–1492.

[3] Csorgo, M., Horváth, L., 1997. *Limit Theorems in Change-Point Analysis*, NY: Wiley. MR2743035

[4] Davydov, Y., Egorov, V., 2000. *Functional limit theorems for induced order statistics*, Mathematical Methods of Statistics **9** (3), 297–313. MR1807096

[5] Davydov, Y., Zitikis, R., 2008. *On weak convergence of random fields*, Annals of the Institute of Statistical Mathematics **60**, 345–365. MR2403523

[6] Kovalevskii, A. P., 2020. *Asymptotics of an empirical bridge of regression on induced order statistics*, Siberian Electronic Mathematical Reports **17**, 954–963. MR4217195

[7] MacNeill, I. B., 1978. *Limit processes for sequences of partial sums of regression residuals*, Ann. Prob. **6**, 695–698. MR0494708

[8] MacNeill, I.B., Jandhyala, V.K., Kaul, A., Fotopoulos, S.B., 2020. *Multiple change-point models for time series*, Environmetrics **31** (1), e2593. MR4061127

[9] Straf, M.L., 1972. *Weak convergence of stochastic processes with several parameters*, Proc. Sixth Berkely Symp. Math. Statist. Prob. **2**, 187–222. MR0402847

Back to schedule

Prediction of record values from several samples

Christina Empacher[a] and Udo Kamps[a]
[a] *Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany*

Back to schedule

The prediction of record values based on a single sample has been studied extensively in the literature (cf. e.g., [2, 5, 6]). The case of two observed samples of records is considered for example in [1] and [7], where a record of one sample is predicted based on data of the other sample.

In our talk, we discuss another approach to prediction with several samples. We predict future records for each sample simultaneously and the prediction is based on information from observed records of all samples.

Known point prediction procedures in the one-sample case, namely maximum likelihood, maximum observed likelihood and maximum product of spacings prediction ([4, 8, 9]), are adapted to the described setting of several samples and compared to the one sample case. Additionally, methods for interval prediction are given. These prediction methods are studied in case of an underlying Pareto distribution. Finally, the point predictors are applied to data from athletics, which may be assumed to follow a Pareto distribution ([3]).

References

[1] Ahmadi J., Balakrishnan N. (2013). On the nearness of record values to order statistics from Pitman's measure of closeness. *Metrika* **76(4)**, 521–541.

[2] Ahsanullah M. (1980). Linear prediction of record values for the two parameter exponential distribution. *Annals of the Institute of Statistical Mathematics* **32(3)**, 363–368.

[3] Empacher C., Kamps U., Volovskiy G. (2023). Statistical prediction of future sports records based on record values. *Stats* **6(1)**, 131–147.

[4] Kaminsky K.S., Rhodin L.S. (1985). Maximum likelihood prediction. *Annals of the Institute of Statistical Mathematics* **37(3)**, 507–517.

[5] Madi M.T., Raqab M.Z. (2004). Bayesian prediction of temperature records using the Pareto model. *Environmetrics* **15(7)**, 701–710.

[6] Navarro J. (2022). Prediction of record values by using quantile regression curves and distortion functions. *Metrika* **85(6)**, 675–706.

[7] Singh S., Tripathi Y.M., Wu S.-J. (2017). Bayesian estimation and prediction based on lognormal record values. *Journal of Applied Statistics* **44(5)**, 916–940.

[8] Volovskiy G., Kamps U. (2020a). Maximum observed likelihood prediction of future record values. *Test* **29(4)**, 1072–1097.

[9] Volovskiy G., Kamps U. (2020b). Maximum product of spacings prediction of future record values. *Metrika* **83(7)**, 853–868.

Back to schedule

Prediction intervals for future Pareto record values with applications in insurance

Christina Empacher[a], Udo Kamps[a] and Anja B. Schmiedt[a]
[a] *Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany*

Back to schedule

Statistical analysis of record values is of interest in various fields, such as actuarial science, environmental studies and sports. For basic references on record values, we refer to the monographs [1, 2]. In this presentation, applications from insurance are considered, where prediction of extreme claims or losses is crucial for pricing and quantitative risk modeling (cf. [2, 4, 5]).

Based on record values in a sequence of iid random variables, several authors study point prediction of future record values, e.g., [6, 7, 8, 9, 10, 11]. Here, the problem of predicting a future record value (such as a future record claim) based on a sequence of previously observed record values is addressed by means of prediction intervals.

There are former results on this topic; for an underlying Pareto distribution, respective exact and approximate prediction intervals for future upper record values (see [9, 12, 13]) are summarized and modified and new ones are developed based on a point predictor via the method of maximum product of spacings (see [11]).

In a simulation study, these prediction intervals are evaluated and compared regarding coverage frequency and length. Several prediction methods are applied to real data sets from the insurance industry, which turn out to perform well.

The use of the Pareto distribution is discussed along with the common situation of a fairly small number of observed record values. In order to increase the number of observations, the application of k-th record values (cf. [14]) is examined as an option for statistical analyses to predict, e.g., second largest record claims. The methods seem to be able to capture the magnitude of future record claims, even for small numbers of record observations (see [15]).

Keywords: record values, interval prediction, Pareto distribution, real insurance data sets

References

[1] Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (1998). *Records*. Wiley, New York.

[2] Nevzorov, V.B. (2001). *Records: Mathematical Theory*. American Mathematical Society, Providence

[3] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. Springer, Berlin.

[4] Klüppelberg, C., Straub, D., andWelpe, I.M. (2014). *Risk - A Multidisciplinary Introduction*. Springer.

[5] McNeil, A.J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton University Press.

[6] Awad, A.M., and Raqab M.Z. (2000). Prediction intervals for the future record values from exponential distribution: comparative study. *J. Stat. Comput. Simul.* **65(1–4)**, 325-–340.

[7] Basak, P., and Balakrishnan, N. (2003). Maximum likelihood prediction of future record statistics. In: Lindqvist, B., and Doksum, K. (eds.) Chapter 11: Mathematical and Statistical Methods in Reliability. World Scientific Publishing, New York, pp 159—175.

[8] Raqab, M.Z. (2007). Exponential distribution records: different methods of prediction. In: Ahsanullah, M., and Raqab, M.Z. (eds.) Chapter 16: Recent Developments in Ordered Random Variables. Nova Science Publishers, Hauppauge, pp 239—251.

[9] Raqab, M.Z., Ahmadi, J., and Doostparast, M. (2007). Statistical inference based on record data from Pareto model. *Statistics* **41(2)**, 105—118.

[10] Volovskiy, G., and Kamps, U. (2020a). Maximum observed likelihood prediction of future record values. *TEST* **29(4)**, 1072-–1097.

[11] Volovskiy, G., and Kamps, U. (2020b). Maximum product of spacings prediction of future record values. *Metrika* **83(7)**, 853–868.

[12] Asgharzadeh, A., Abdi, M., and Kus, C. (2011). Interval estimation for the two-parameter Pareto distribution based on record values. *Selcuk J. Appl. Math.*, 149–161.

[13] Empacher, C., Kamps, U., and Volovskiy, G. (2023). Statistical prediction of future sports records based on record values. *Stats* **6(1)**, 131–147.

[14] Dziubdziela, W., and Kopocinski, B. (1976). Limiting properties of the k-th record values. *Applicationes Mathematicae* **15(2)**, 187-–190.

[15] Empacher, E., Kamps, U., and Schmiedt, A.B. (2024). Prediction intervals for future Pareto record claims. *Submitted.*

Back to schedule

Further developments on characterizations of distributions based on regressions of GOS

Mariusz Bieniek[a]
[a] *Institute of Mathematics, Maria Curie-Skłodowska University, Lublin, Poland*

Back to schedule

We study the problem of the characterization of probability distributions by a regression function of non-adjacent generalized order statistics (GOS). For any fixed continuous and strictly increasing function $h : (\alpha, \beta) \to \mathbb{R}$ and for GOS based on a continuous distribution function F we define the regression function of GOS by

$$\xi(x) = \mathbb{E}\left(h(X_*^{(s)}) \;\middle|\; X_*^{(r)} = x\right), \quad x \in (\alpha, \beta), \tag{1}$$

where $s > r \geq 1$. We prove the uniqueness of the characterization and we show necessary and sufficient conditions for a function $\xi : (\alpha, \beta) \to \mathbb{R}$ to be a regression of the form (1) for some continuous distribution F.

Back to schedule