XVIII-th International Summer Conference on Probability and Statistics (ISCPS-2018)

Workshop on Branching Processes and Applications

Workshop on Biostatistics

Seminar on Statistical Data Analysis

Conference Proceedings and Abstracts

Edited by E. Stoimenova and M. Slavtchova-Bojkova

Pomorie, Bulgaria, 25 – 30 June 2018

This volume contains a collection of abstracts and short contributed papers delivered by the participants at the XVIII-th International Summer Conference on Probability and Statistics held in Pomorie, Bulgaria, from 25 June to 30 June 2018.

Co-organizers of this joint event were:

- Bulgarian Statistical Society
- Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences
- Faculty of Mathematics and Informatics of the University of Sofia "St. Kliment Ohridski"

The Conference took jointly with the Workshop on Branching Processes and Applications, Workshop on Biostatistics and Seminar on Statistical Data Analysis.

©Eugenia Stoimenova, Maroussia Slavtchova-Bojkova, Editors

Publisher: *Institute of Mathematics and Informatics*
Bulgarian Academy of Sciences
Acad. G.Bontchev str., bl.8, 1113 Sofia, BULGARIA

ISBN 978-954-8986-50-2

Contents

KEYNOTE INVITED SPEAKERS

On general models of exceedance statistics based on random intervals and ordered random variables

Ismihan Bayramoglu
Izmir University of Economics
Izmir, Turkey
e-mail: ismihan.bayramoglu@ieu.edu.tr

We consider some models of exceedance statistics defined as a number of observations exceeding a random threshold formed by random intervals and ordered random variables. We are interested in finite and asymptotic distributions of such statistics. The results concern ordinary order statistics, record values, generalized order statistics. Many of the results presented for here are derived for absolutely continuous distributions, however we also review results obtained for distributions having countable number of points of discontinuity. Exceedance and threshold models are useful in many fields of applications such as hydrology, ecology, engineering, economy and finance.

References

[1] Erem, A.and Bayramoglu, I. (2017) Exact and asymptotic distributions of exceedance statistics for bivariate random sequences. Statistics & Probablity Letters. 125, 181-188.

[2] Bairamov, I. and Tanil, H. (2007) On distributions of exceedances based on generalized order statistics. Communications in Statistics—Theory and Methods, 36: 1479–1491.

[3] Bairamov, I. and Khan, M.K. (2007) On exceedances of record and order statistics. Proceedings of the American Mathematical Society. Volume 135, Number 6, pp. 1935-1945.

[4] Bairamov I. (2006) Advances in Exceedance Statistics Based on Ordered Random Variables. Recent Developments in Ordered Random Variables. Eds. Mohammad Ahsanullah and M.Z. Raqab. Nova Science Publishers. New York.

[5] Bairamov I.G. and Eryılmaz S. (2006) Spaciings, exceedances and concomitants in progressive type II censoring scheme. Journal of Statistical Planning and inference. 136, pp. 527-536.

Vaccination in epidemic modelling: stochastic monotonicity and continuity properties

Frank Ball[1], M. González[2], R. Martínez[2],
Maroussia Slavtchova–Bojkova[3,4] ,
[1] *School of Mathematical Science, the University of Nottingham, UK*
[2] *University of Extremadura, Badajoz, Spain*
[3] *Sofia University, Faculty of Mathematics and Informatics,*
[4] *Institute of Mathematics and Informatics, BAS, Sofia, Bulgaria*
e-mail: [3,4]bojkova@fmi.uni-sofia.bg

One of the key tools in controlling the spread of an epidemic is vaccination. There have been numerous mathematical studies of its effect on disease dynamics, but the majority of such studies consider only the situation where a specified fraction of the population is vaccinated prior to an outbreak. In Ball et al. [1], a framework was developed for analyzing time–dependent vaccination policies for epidemics which are modeled by a Crump–Mode–Jagers branching process. Stochastic monotonicity and continuity results for a wide class of functions (e.g., duration and total size of an outbreak) defined on such a branching process were proved, leading to optimal vaccination schemes to control corresponding functions of epidemic outbreaks.

In this work, we study the extension of those results to the framework of a very general SIR epidemic model. In addition to standard SIR and SEIR models, this framework includes as special cases both households and network models. It also allows for a very general model for vaccine action. The results are illustrated by a simulated example.

Acknowledgements This work is supported by the financial funds of the Ministerio de Economía y Competitividad (grant MTM2015–70522–P), the Junta de Extremadura (grant IB16099) and the Fondo Europeo de Desarrollo Regional, and the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria (grant DFNI–I02/17).

References

[1] Ball F., González M., Martínez R. and Slavtchova–Bojkova M. (2014). Stochastic monotonicity and continuity properties of functions defined on Crump–Mode–Jagers branching processes, with application to vaccination in epidemic modelling. *Bernoulli.* Vol. **20**, no. 4, pp. 2076–2101.

Some contributions concerning two-sex branching populations

Manuel Molina[1], Manuel Mota[2]
[1,2]*Department of Mathematics and Institute of Advanced Scientific Computation, University of Extremadura*
Badajoz, Spain
e-mail: [1]mmolina@unex.es, [2]mota@unex.es

Branching processes were especially developed to model biological phenomena. In the last years, in order to describe the demographic dynamics of biological populations with sexual reproduction, several classes of two-sex branching processes have been introduced and some theory about them developed. In these processes, the population consists of two disjoint types of individuals, females and males, and two important biological phases are considered, the mating phase where the couples are formed, and the reproduction phase in which each couple produces new individuals according to a given offspring probability distribution. In this talk, we review such classes of processes, we comment the main theoretical contributions obtained, and we discuss some possible questions for research. Special attention is devoted to the class of population-dependent two-sex processes with random mating introduced in *Jacob et al. (Bernoulli, 23(3), 2017, 1737-1758).*

Comparisons of distribution functions and moments of mixtures of ordered families of distributions

Tomasz Rychlik
Institute of Mathematics, Polish Academy of Sciences
Warsaw, Poland
e-mail: trychlik@impan.pl

Abstract

We compare two mixtures of arbitrary family of stochastically ordered distribution functions with respect to fixed mixing distributions. Under the assumption that the first mixture distribution is known, we establish optimal lower and upper bounds on the latter mixture distribution function, and present single families of ordered distributions which attain the bounds uniformly at all real arguments. Furthermore, we determine sharp upper and lower bounds on the differences between the expectations of the mixtures expressed in various scale units. We also present sharp lower and upper bounds on the ratios of various dispersion measures of the mixtures.

1 Introduction

Let $\{F_\theta\}_{\theta\in\mathbb{R}}$ be an arbitrary family of stochastically ordered distribution functions. i.e., ones that satisfy

$$\theta_1 < \theta_2 \Rightarrow F_{\theta_1}(x) \geq F_{\theta_2}(x), \qquad x \in \mathbb{R}. \tag{1}$$

We assume that it is not known. We further introduce two distribution functions S and T which are assumed to be known. We do not impose any restrictions on them. They may have discrete and/or continuous components, and supports concentrated on possibly partially or even fully different subsets of the real line. The purpose of the paper is to compare the mixture distribution functions

$$G(x) = \int_{\mathbb{R}} F_\theta(x) S(d\theta), \tag{2}$$

$$H(x) = \int_{\mathbb{R}} F_\theta(x) T(d\theta), \tag{3}$$

and their expectations and dispersion measures.

Precisely, in Section 2 we determine sharp lower and upper bounds on distribution functions (3) under the constraint that condition (2) is satisfied for an arbitrarily fixed G. The bounds are attained uniformly for any G, i.e., there exist single families $\{\underline{F}_\theta\}_{\theta\in\mathbb{R}}$ and $\{\overline{F}_\theta\}_{\theta\in\mathbb{R}}$ such that (2) holds, and the lower and upper bounds on (3) are attained by $\{\underline{F}_\theta\}_{\theta\in\mathbb{R}}$ and $\{\overline{F}_\theta\}_{\theta\in\mathbb{R}}$, respectively, for every real argument x. In Section 3, we determine the greatest possible lower and upper deviations of $\mathbb{E}Y = \int_{\mathbb{R}} xH(dx)$ from $\mathbb{E}X = \int_{\mathbb{R}} xG(dx)$, measured in various scale units $(\mathbb{E}|X-\mathbb{E}X|^p)^{1/p} = \left(\int_{\mathbb{R}} |x-\mathbb{E}X|^p G(dx)\right)^{1/p}$, $p \geq 1$, based on central absolute moments of (2). In Section 4, we present the lower and upper bounds on the ratios of general dispersion measures (including variance and mean absolute deviation from the mean as special cases) of the mixtures. We illustrate general achievements for exemplarily chosen mixing distributions in Section 5.

We emphasize that all the bounds presented in Sections 2–4 are optimal. The attainability conditions are rather complicate, and we do not present them here. We also omit the proofs. The results of Sections 2 and 3 were comprehensively presented in [3]. Proposition 3 of Section 3 was earlier established by means of different tools in [2]. The results of Section 4 come from [1].

2 Distribution functions of mixtures

Theorem 1. *Let S, T, and G be fixed distribution functions on $\mathbb{R}$. Then for every family $\{F_\theta\}_{\theta\in\mathbb{R}}$ of stochastically ordered distribution functions* (cf (1)) *satisfying* (2)*, the following inequalities hold*

$$\underline{H}(G(x)) \leq \int_{\mathbb{R}} F_\theta(x) T(d\theta) \leq \overline{H}(G(x)), \qquad x \in \mathbb{R}, \tag{4}$$

where $\underline{H}, \overline{H} : [0,1] \mapsto [0,1]$ are the greatest convex minorant and the smallest concave majorant, respectively, of the set

$$\mathcal{U} = \mathcal{U}(S,T) = \{(u,v) = (S(\theta), T(\theta)) : \ \theta \in \mathbb{R}\}. \tag{5}$$

Moreover, the bounds in (4) *are optimal and uniformly attainable which means that there exist stochastically ordered families $\{\underline{F}_\theta\}_{\theta\in\mathbb{R}}$ and $\{\overline{F}_\theta\}_{\theta\in\mathbb{R}}$*

satisfying (2) *such that for every real* x *yields*

$$\begin{aligned} \underline{H}(G(x)) &= \int_{\mathbb{R}} \underline{F}_\theta(x) T(d\theta), \\ \overline{H}(G(x)) &= \int_{\mathbb{R}} \overline{F}_\theta(x) T(d\theta). \end{aligned}$$

Remark 1. Set (5) may have very different forms for various S and T. When both S and T are continuous, (5) is represented by a single curve joining $(0,0)$ and $(1,1)$ containing possibly vertical and/or horizontal line segments if the supports of S and/or T have some gaps. If S and T are discrete, it consists of separate points. In general, set (5) consists of at most countably many pieces. For each pair of pieces, one is located right and above the other. Only one coordinate of the left lower end-point of the former piece may be equal to the respective coordinate of the right upper end-point of the latter one. The pieces of (5) are either curves or separate points. The curves may contain horizontally and vertically oriented line segments as well as some their pieces can represent the graphs of strictly increasing functions. Each curve contains the point with minimal values of both coordinates, but it does not necessarily contain the right upper end-point.

Remark 2. The points of the graphs of the greatest convex minorant and the smallest concave majorant of (5) either coincide with the points of the set, or belong to straight line segments joining the elements of the set (more precisely, the limiting values of the right upper ends of the set pieces may also serve as the left and right ends of the line segments). By definition of $\mathcal{U}(S,T)$, the greatest convex minorant and the smallest concave majorant are non-decreasing functions. They are obviously continuous as well. Accordingly, compositions $\underline{H} \circ G$ and $\overline{H} \circ G$ are well-defined functions on the whole real axis, non-decreasing, right-continuous, and have all values in $[0,1]$.

They are not necessarily proper distribution function, though. We express the necessary conditions using the following notation: two intervals $\Theta_1, \Theta_2 \subset \mathbb{R}$ satisfy $\Theta_1 \prec \Theta_2$ if every element of Θ_1 is less than any element of Θ_2. Accordingly, $\underline{H} \circ G$ and $\underline{H} \circ G$ are distribution functions iff

$$\forall\, \Theta_1 \prec \Theta_2 \quad S(\Theta_1) = 1 \;\Rightarrow\; T(\Theta_2) = 0, \tag{6}$$

$$\forall\, \Theta_1 \prec \Theta_2 \quad S(\Theta_2) = 1 \;\Rightarrow\; T(\Theta_1) = 0, \tag{7}$$

respectively (for simplicity of notation, we use the same symbol for the distribution function and respective probability distribution).

3 Expectations of mixtures

In this section we examine variations of the expectations of mixtures under stochastic ordering of mixed distributions. Suppose that we consider a mixture of arbitrary unknown family of ordered distribution functions, and the actual mixing distribution is S. The resulting random variable X has an unknown distribution function (2). However, we assume that the mixing distribution is T, different from S. This generates a random variable Y with different distribution function (3) and expectation $\mathbb{E}Y$. Our purpose is to evaluate the maximal possible differences between the assumed and actual expectations of mixtures $\mathbb{E}Y - \mathbb{E}X$. This is measured in various scale units $\sigma_p = (\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}$, $p \geq 1$, generated by the central absolute moments of the actual mixture variable X. The bounds depend merely on the mixing distributions S and T, and are valid for all possible $\{F_\theta\}_{\theta\in\mathbb{R}}$, and resulting mixture distributions G and H. The only restriction is that X has a finite pth raw moment $\mathbb{E}|X|^p$ of chosen order $p \geq 1$, and Y has a finite expectation. Similarly, under condition that actual mixture variable X has a bounded support, we determine upper and lower bounds on $\mathbb{E}Y - \mathbb{E}X$ gauged in the scale units $\sigma_\infty = \operatorname{ess\,sup} |X - \mathbb{E}X|$.

Proposition 1. *Let X and Y be random variables with distribution functions which are the mixtures of some stochastically ordered families of distributions $\{F_\theta\}_{\theta\in\mathbb{R}}$ with respect to fixed distribution functions S and T, respectively. Assume that X has a finite pth moment for some fixed $1 < p < \infty$.*
(i) Suppose that $S \not\geq T$ satisfy (6)*, and denote by $\underline{h}$ the right-continuous version of the derivative of the greatest convex minorant of set $\mathcal{U}(S,T)$ defined in* (5)*. If $\int_0^1 \underline{h}^{p/(p-1)}(u)\,du$ is finite, then for all mixture distribution functions* (2) *for which $\mathbb{E}|X|^p < \infty$ we have*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}} \leq \left(\int_0^1 |\underline{h}(u) - C_p|^{p/(p-1)} du \right)^{(p-1)/p}, \tag{8}$$

where $\underline{h}(0) < C_p < \underline{h}(1)$ *uniquely solves the equation*

$$\int_{\{\underline{h}(u)<c\}} [c - \underline{h}(u)]^{1/(p-1)} du = \int_{\{\underline{h}(u)>c\}} [\underline{h}(u) - c]^{1/(p-1)} du. \qquad (9)$$

(ii) Let $S \not\leq T$ *satisfy* (7)*, and* $\overline{h}$ *denote the right derivative of the smallest convex majorant of* (5)*. If* $\int_0^1 \overline{h}^{p/(p-1)}(u)\, du < \infty$*, then for all mixture distributions* (2) *with finite pth moment yields*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}} \geq -\left(\int_0^1 |\overline{h}(u) - c_p|^{p/(p-1)} du \right)^{(p-1)/p},$$

where $\overline{h}(0) > c_p > \overline{h}(1)$ *is determined by equation* (9) *with* $\underline{h}$ *replaced by* $\overline{h}$*.*

We briefly discuss the assumptions of Proposition 1(i). Condition $S \geq T$ would imply $G \geq H$ for any stochastically ordered $\{F_\theta\}_{\theta \in \mathbb{R}}$, and $\mathbb{E}Y - \mathbb{E}X \leq 0$ in consequence. Here we consider only sharp positive bounds, and non-positive ones are studied in another paper which is under preparation yet. Assumption (6) guarantees that $\underline{H}$ is a distribution function supported on $[0, 1]$. Otherwise it is easy to construct G with a bounded support (and clearly all finite moments) and H with infinite expectation. Assumption $\mathbb{E}|X|^p < \infty$ is necessary for defining the scale unit, and $\int_0^1 \underline{h}^{p/(p-1)}(u) du < \infty$ ensures finiteness of bound in (8). Note that increasing p, we make more stringent requirements on moments of X, and at the same time we weaken conditions on the values of $\underline{h}$. The assumptions of Proposition 1(ii) can be justified in a similar way. Formulae of Proposition 1 significantly simplify for $p = 2$.

Corollary 1. *Assume notation of Proposition* 1*, and put* $p = 2$*.*
(i) If (6)*,* $S \not\geq T$*, and* $\int_0^1 \underline{h}^2(u) du < \infty$ *hold, then*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\mathbb{V}ar\, X}} \leq \sqrt{\int_0^1 \underline{h}^2(u) du - 1}.$$

(ii) Under assumptions (7)*,* $S \not\leq T$*, and* $\int_0^1 \overline{h}^2(u) du < \infty$

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\mathbb{V}ar\, X}} \geq -\sqrt{\int_0^1 \overline{h}^2(u) du - 1}.$$

Proposition 2. *Let X, Y, and $\underline{h}$, $\overline{h}$ be defined as in Proposition 1, and suppose that X has a finite mean.*
(i) If (6) *holds, $S \not\geq T$, and $\underline{h}$ is bounded, then*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq \frac{\underline{h}(1) - \underline{h}(0)}{2},$$

where $\underline{h}(1)$ is the left derivative of $\underline{H}$ at 1.
(ii) Under (7), *$S \not\leq T$, and boundedness of $\overline{h}$, we have*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \geq -\frac{\overline{h}(0) - \overline{h}(1)}{2},$$

with $\overline{h}(1)$ denoting the left derivative of $\overline{H}$ at 1.

Proposition 3. *Take X, Y, and $\underline{h}$, $\overline{h}$ defined above and suppose that X has a bounded support.*
(i) If (6) *holds, $S \not\geq T$, and $\int_0^1 \underline{h}(u)du < \infty$, we have*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\operatorname{ess\,sup}|X - \mathbb{E}X|} \leq 1 - 2\underline{H}\left(\frac{1}{2}\right).$$

(ii) Assume (7), *$S \not\leq T$, and $\int_0^1 \overline{h}(u)du < \infty$. It follows that*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\operatorname{ess\,sup}|X - \mathbb{E}X|} \geq 1 - 2\overline{H}\left(\frac{1}{2}\right).$$

4 Dispersions of mixtures

We construct a general class of dispersion measures on a basis of loss functions $\varrho : \mathbb{R} \mapsto \mathbb{R}_+$ used in estimation of location parameters. We assume that ϱ is non-increasing on $\mathbb{R}_-$, vanishing at 0, and non-decreasing on $\mathbb{R}_+$. Then the dispersion measure of random variable X with some distribution function F associated with the loss function ϱ is defined as

$$\sigma(X, \varrho) = \inf_{\mu \in \mathbb{R}} \mathbb{E}\varrho(X - \mu).$$

In particular, if $\varrho(x) = |x|$, then $\sigma(X, \varrho) = \mathbb{E}\left|X - F^{-1}\left(\frac{1}{2}\right)\right|$. If $\varrho(x) = x^2$, this is just $\mathbb{V}ar\, X$. For asymmetric Linux loss functions $\varrho(x) = e^{ax} - 1 - ax$ with $a \neq 0$, we have $\sigma(X, \varrho) = \mathbb{E}e^{aX} - a\mathbb{E}X$.

Proposition 4. *Let X and Y have distribution functions* (2) *and* (3) *with fixed S and T, and arbitrary stochastically ordered family $\{F_\theta\}_{\theta\in\mathbb{R}}$ (see* (1)*). If $0 < \sigma(X,\varrho) < +\infty$ then*

$$\min\{\overline{h}(1), \underline{h}(0)\} \le \frac{\sigma(Y,\varrho)}{\sigma(X,\varrho)} \le \max\{\overline{h}(0), \underline{h}(1)\}.$$

Moreover, if ϱ is continuous at 0 and positive for $x \neq 0$, then the bounds are sharp.

Observe that the above evaluations are common for various ϱ.

5 Example

We consider a pair of apparently similar symmetric unimodal mixing distributions supported on $[0,1]$. One is

$$S(\theta) = \begin{cases} \frac{3}{2}\theta^2, & 0 < \theta < \frac{1}{2}, \\ 1 - \frac{3}{2}(1-\theta)^2, & \frac{1}{2} \le \theta < 1, \end{cases}$$

which has a symmetric triangular density, and an atom $\frac{1}{4}$ at the mode $\frac{1}{2}$. The other one

$$T(\theta) = \begin{cases} 4\theta^3, & 0 < \theta < \frac{1}{2}, \\ 1 - 4(1-\theta)^3, & \frac{1}{2} \le \theta < 1, \end{cases}$$

has quadratic density of different forms on $\left(0, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, 1\right)$, and a sharp peak at $\frac{1}{2}$. Then

$$\begin{aligned} \mathcal{U}(S,T) &= \left\{ \left(u, \frac{8}{3}\sqrt{\frac{2}{3}} u^{3/2} \right) : \ 0 \le u < \frac{3}{8} \right\} \\ &\cup \left\{ \left(u, 1 - \frac{8}{3}\sqrt{\frac{2}{3}} (1-u)^{3/2} \right) : \ \frac{5}{8} \le u < 1 \right\}. \end{aligned}$$

One can check that the greatest convex minorant of the set and its derivative have the forms

$$\underline{H}(u) = \begin{cases} \frac{8}{3}\sqrt{\frac{2}{3}}u^{3/2}, & 0 \le u \le \frac{9}{8} - \frac{3}{4}\sqrt{2}, \\ (2\sqrt{2}-2)\left(u - \frac{5}{8}\right) + \frac{1}{2}, & \frac{9}{8} - \frac{3}{4}\sqrt{2} \le u \le \frac{5}{8}, \\ \frac{4}{3}(u-1)+1, & \frac{5}{8} \le u \le 1, \end{cases}$$

$$\underline{h}(u) = \begin{cases} 4\sqrt{\frac{2u}{3}}, & 0 \le u \le \frac{9}{8} - \frac{3}{4}\sqrt{2}, \\ 2\sqrt{2}-2, & \frac{9}{8} - \frac{3}{4}\sqrt{2} \le u < \frac{5}{8}, \\ \frac{4}{3}, & \frac{5}{8} \le x < 1, \end{cases}$$

respectively. Similarly, we verify that $\overline{H}(u) = 1 - \underline{H}(1-u)$, and $\overline{h}(u) = \underline{h}(1-u)$. By simple calculations we obtain

$$\begin{aligned} -\frac{2}{3} &\le \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \le \frac{2}{3}, \\ -\sqrt{4\sqrt{2} - 5\frac{7}{12}} &\le \frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\mathbb{V}ar\, X}} \le \sqrt{4\sqrt{2} - 5\frac{7}{12}} \approx 0.27115, \\ -\frac{\sqrt{2}-1}{2} &\le \frac{\mathbb{E}Y - \mathbb{E}X}{\operatorname{ess\,sup}|X - \mathbb{E}X|} \le \frac{\sqrt{2}-1}{2} \approx 0.20711, \\ 0 &\le \frac{\sigma(Y,\varrho)}{\sigma(X,\varrho)} \le \frac{4}{3}. \end{aligned}$$

In particular, the last inequalities hold for the variances and mean absolute deviations from the medians.

References

[1] Miziuła, P. (2017). Comparison of dispersions of mixtures of ordered distributions. *Statistics* Vol. **51**, 862–877.

[2] Miziuła, P., Solnický, R. (2018). Sharp bounds on change in expected values and variances for single risk analysis in the flood catastrophe model. *Scand. Actuar. J.* Vol. **2018**, 64–75.

[3] Rychlik, T. (2018). Sharp bounds on distribution functions and expectations of mixtures of ordered families of distributions, under review.

Branching regression models with random effects: penalized inference and related issues

Anand N. Vidyashankar
Department of Statistics,
Volgenau School of Engineering,
George Mason University
Fairfax VA, 22030, USA
e-mail: avidyash@gmu.edu

Branching processes (BP) and their variants are frequently used to model sizes of evolving populations in a variety of scientific fields such as criminology, biochemical processes, finance, insurance, and epidemiology. In these examples, it is useful to relate a functional of the branching process to a covariate set (which is typically large), and study their effect on the evolutionary structure of the process. Motivated by such applications, we introduce a class of stochastic processes referred to as Branching regression models with random effects that provide a phenomenological description of the underlying dynamics and investigate the inferential issues; specifically, penalized inference. In the process, we also address the role of model selection uncertainty in the resulting inference.

Keywords: Penalized inference, Oracle property, Branching regression models with random effects

Estimation in a class of controlled branching processes

George P. Yanev
The University of Texas Rio Grande Valley
Edinburg, Texas, USA
e-mail: george.yanev@utrgv.edu

Controlled branching processes constitute a large class of stochastic processes and emigration. Consider a Galton-Watson branching process with non-stationary immigration (BPNI). Assume that every individual from the n-th generation, independently from the others, is detected of having certain condition with probability p and remains undetected with probability 1-p. The detected individuals form the n-th generation of another branching process called partially observed branching process (POBP). Due to the sampling procedure, the detected and undetected individuals might end up having different offspring distributions. The resulting POBP is a multiple controlled process with non-stationary immigration and state-dependent emigration. The POBP is observed, while the original process BPNI is not.

We construct nonparametric estimators for the offspring mean of the original BPNI based on data from POBP only. That is, assuming binomially distributed subset of observations in each generation of BPNI. The proposed estimators are related to those used when the underlying process is fully observed taking into account the dependence structure induced by the partial observation (detection) procedure. Under certain conditions on the first two moments of the immigration, we show strong consistency and asymptotic normality of the estimators in all three modes of criticality.

The talk is based on joint work with I. Rahimov.

BIOSTATISTICS
(INVITED SESSION)

Session 1: Statistics for Precision medicine

Precision medicine is an emerging approach for disease prevention and treatment that takes into account people's individual variations in genes, environment, demographic characteristics, physiology, family health history and lifestyle. The Precision Medicine Initiative encompasses: identifying genetic profiles associated with a disease or disorder; development of new targeted treatments; identifying subgroups of patients who are likely to benefit from a given treatment; and clinical decision-making based on patient characteristics. In this session, we give an overview how statistical science can further the goals of precision medicine, and we introduce some areas of statistical methodology that currently are experiencing dramatic growth by pursuing these goals. This session will also present some recent research.

Statistical framework and theory of optimal treatment decisions

Eva Petkova
New York University
New York, USA
e-mail: eva.petkova@nyumc.org

In this talk we will give background for the goals of precision medicine and will formalize clinical decision-making. We will introduce the notion of potential outcomes and will propose criteria for optimality of treatment decision rules. Here we will discuss the challenges of determining optimal treatment decision rules and will argue that statistical methodology is essential. A general statistical framework for optimal treatment decisions will be discussed.

Keywords: precision medicine, potential outcomes, value of treatment decisions

Regression-based methods for estimation of optimal treatment decisions

Thaddeus Tarpey
Wright State University
Dayton, Ohio, USA
e-mail: thaddeus.tarpey@wright.edu

In this presentation, we will consider optimal treatment decision rules from a regression perspective. In regression models for the outcome, if there is a non-zero interaction between treatment and a predictor, that predictor is called an "effect modifier". Identification of such effect modifiers is crucial as we move towards precision medicine, i.e., optimizing individual treatment assignment based on patient measurements assessed when presenting for treatment. In most settings, there will be several baseline predictor variables that could potentially modify the treatment effects. Here we will propose optimal methods of constructing a composite variable (defined as a linear combination of pre-treatment patient characteristics) in order to generate an effect modifiers. Linear and nonparametric regression will be considered. We will illustrate the methodology using data from a clinical trial for the treatment of depression.

Use of complex functional data for precision medicine

Todd Ogden
Columbia University, New York
New York, USA
e-mail: to166@columbia.edu

In many areas of medicine, imaging data and other very high dimensional data are available as predictors in regression models, e.g., to predict a patient's treatment outcome based on brain imaging data obtained at baseline. Obtaining meaningful models in such problems requires some form of dimension reduction while taking into account the structure of the data, a primary goal of functional data analysis. This talk will discuss various functional data analytic approaches for fitting such models.

Session 2: Joint modeling of multiple outcomes

A common objective in longitudinal studies is to characterize the relationship between several longitudinal response processes, such as, for example, blood pressure, lipids and time to stroke in patients with cardiovascular disease; or alcohol craving and number of drinks per day in treatment of alcohol abuse; level of hopelessness, helplessness and lack of interest in depression. Considerable recent interest has focused on the so-called joint models, where models for the outcomes are taken to depend on a common set of latent random effects. This session will review the rationale for and some development of joint models.

Joint analysis of survival and longitudinal non-survival data: current methods and issues

Gleb Haynatzki
University of Nebraska MC,
Omaha, Nebraska, USA
e-mail: ghaynatzki@unmc.edu

Joint modeling underlying the relationships between survival time variables (where the event could be, eg, death or dropout) and multivariate longitudinal non-survival variables (eg, biomarkers) could lead to improved estimates (eg, by reducing bias) and to improved efficiency of longitudinal studies (eg, clinical trials). Methods for the joint modeling of survival and non-survival longitudinal variables in the same study have been well developed for a single survival outcome and a single non-survival longitudinal variable. However, longitudinal studies typically collect data on more than one non-survival longitudinal variables, and incorporating information from all data sources could improve the predictive capability of statistical models. We will review currently available methods of such joint modelling, including distributional and modelling assumptions, the association structures, estimation approaches, and software that can be used to carry out the methods, as well as applications of the methodologies.

Keywords: joint modeling, survival time, longitudinal non-survival, estimation, statistical models

Joint analysis of repeatedly observed outcomes in medical research

Ralitza Gueorguieva
Yale University, New Haven, Connecticut, USA
e-mail: ralitza.gueorguieva@yale.edu

Joint modeling of variables of different types is often necessary to evaluate predictor effects on multiple outcomes, often repeatedly measured over time. In particular, correlated probit models are used for simultaneous analysis of continuous and ordinal/binary measures; mixture models are helpful in dealing with zero-inflated data; joint models of frequency and intensity measures evaluate different aspects of substance use behaviors. Advantages of joint modeling include assessment of overall treatment effects, better control of type I error, characterization of the correlation structure among the outcomes and potential efficiency gains for the parameter estimates. We will consider several traditional and novel models for joint analysis of categorical and continuous outcomes and will discuss their advantages and disadvantages. Maximum-likelihood estimation and a Bayesian approach to model fitting, together with software implementations will be briefly considered. Simulation results evaluating efficiency gains of simultaneous analysis compared to separate analysis will be presented. Potential bias under miss-specification of the random effects structure will also be discussed. Data on frequency and intensity of drinking, and on the relationship between the use of traditional tobacco products and e-cigarettes will illustrate the methods and discussion.

Correlated probit analysis of two longitudinal ordinal outcomes

Denitsa Grigorova
Bocconi University, Milan, Italy
e-mail: denitsa.grigorova@unibocconi.it

Correlated probit models (CPMs) are widely used for modeling of ordinal data or joint analyses of ordinal and continuous data which are common outcomes in medical studies. When we have clustered or longitudinal data CPMs with random effects are used to take into account the dependence between clustered measurements. When the dimension of the random effects is large, finding of the maximum likelihood estimates (MLEs) of the model parameters via standard numerical approximations is computationally cumbersome or in some cases impossible. EM algorithms for one ordinal longitudinal variable [1] and for one ordinal and one continuous longitudinal variable [2] are recently developed. ECM algorithm for MLEs of CPM for two longitudinal ordinal variables will be presented. The algorithm is applied to estimation of CPM for the longitudinal ordinal outcomes self-rated health and categorized body mass index from the Health and Retirement Study. Results from fitting the model to the data and also results from some simulation studies will be reported.

References

[1] Grigorova, D., Gueorguieva, R. (2013). Implementation of the EM algorithm for maximum likelihood estimation of a random effects model for one longitudinal ordinal outcome. *Pliska Stud. Math. Bulgar.* Vol. **22**, pp. 41-56.

[2] Grigorova, D., Gueorguieva, R. (2016). Correlated probit analysis of repeatedly measured ordinal and continuous outcomes with application to the Health and Retirement Study. *Statistics in Medicine*. Vol. **35**, no. 23, pp. 4202-4225.

SHORT COMMUNICATIONS

Almost sure behaviour of maxima from strictly stationary processes with application in insurance

Anna Dembińska
Warsaw University of Technology
Warsaw, Poland
e-mail: dembinsk@mini.pw.edu.pl

Abstract

Assume that a sequence of observations forms a strictly stationary process. We describe the almost sure asymptotic behaviour of partial maxima arising from this process. More precisely, we show that such maxima converge almost surely to some random variable and describe the law of the limiting variate. We prove that also proportions of observations in the sample that fall near the current maximum are almost surely convergent and again we establish the distribution of the random variable appearing in the limit. Finally, we apply the presented results in insurance to examine the long-term behaviour of the number of near-maximum insurance claims.

1 Introduction

Let $\mathbb{X} = (X_n, n \geq 1)$ be a sequence of random variables (rv's) and $M_n = \max(X_1, X_2, \ldots, X_n)$ be the maximum of the sample $(X_1, X_2, \ldots, X_n)$. Our aim is to describe the almost sure behaviour of $(M_n, n \geq 1)$, as $n \to \infty$, under the assumption that $\mathbb{X}$ forms a strictly stationary process. For this purpose we need the concept of conditional right endpoint of the support of an rv given a sigma-field.

Definition 1.1. *Suppose X is an rv on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}$ is a sigma-field with $\mathcal{G} \subseteq \mathcal{F}$. Then the (possibly) extended rv $\underline{Q}_1$ satisfying the following three conditions*

(i) *$\underline{Q}_1$ is $\mathcal{G}$-measurable,*

(ii) *$\mathbb{P}(X \leq \underline{Q}_1 | \mathcal{G}) = 1$ almost surely,*

(iii) *for any $\mathcal{G}$-measurable extended rv Q_1 such that $\mathbb{P}(X \leq Q_1|\mathcal{G}) = 1$ almost surely, we have $\mathbb{P}(Q_1 \geq \underline{Q}_1)$,*

is called a conditional right endpoint of the support of rv X with respect to $\mathcal{G}$ and is denoted by $\gamma_1(X|\mathcal{G})$.

A proof of the existence of $\gamma_1(X|\mathcal{G})$ for any rv on $(\Omega, \mathcal{F}, \mathbb{P})$ and for any sigma-field $\mathcal{G} \subseteq \mathcal{F}$ together with presentation of properties of $\gamma_1(X|\mathcal{G})$ can be found in [1].

To formulate our main results we will also need the notion of invariant set with respect to a sequence of rv's.

Definition 1.2. *A set $A \in \mathcal{F}$ is called invariant with respect to the sequence $\mathbb{X} = (X_n, n \geq 1)$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if there exists a Borel set $B \in \mathbb{R}^{\mathbb{N}}$ such that*

$$A = \{\omega \in \Omega : \ (X_i(\omega), X_{i+1}(\omega), \ldots) \in B\} \text{ for any } i \geq 1.$$

It is well known that the collection of all invariant sets with respect to the sequence of rv's $\mathbb{X} = (X_n, n \geq 1)$ is a sigma-field. We will denote this sigma-field by $\mathcal{I}^{\mathbb{X}}$.

2 Main results

The sequence of partial maxima from any strictly stationary discrete-time process is almost surely convergent. The distribution of the limiting rv is given in the following theorem.

Theorem 2.1. *Let $\mathbb{X} = (X_n, n \geq 1)$ be a strictly stationary sequence on $(\Omega, \mathcal{F}, \mathbb{P})$ with an arbitrary univariate cumulative distribution function. Then*

$$M_n \xrightarrow{a.s.} \gamma_1(X_1|\mathcal{I}^{\mathbb{X}}) \quad as \quad n \to \infty,$$

where $\xrightarrow{a.s.}$ denotes the almost sure convergence.

An analogues result holds for proportions of observations in the sample that fall near the current maximum but this time we need to add some assumptions concerning the distribution of $\mathbb{X}$.

Theorem 2.2. *Let $a > 0$ and $K_n(a) = \#\{i \in \{1, \ldots, n\};\ X_i \in (M_n - a, M_n]\}$ be the number of near-maximum observations. Then, under the assumptions of Theorem* 2.1*, if moreover*

$$\mathbb{P}\left(\gamma_1(X_1|\mathcal{I}^{\mathbb{X}}) = X_1|\mathcal{I}^{\mathbb{X}}\right) = 0 \text{ almost surely} \tag{1}$$

and

$$\mathbb{P}\left(\gamma_1(X_1|\mathcal{I}^{\mathbb{X}}) = X_1 + a|\mathcal{I}^{\mathbb{X}}\right) = 0 \text{ almost surely}, \tag{2}$$

we have

$$K_n(a)/n \xrightarrow{a.s.} \mathbb{P}\left(X_1 > \gamma_1(X_1|\mathcal{I}^{\mathbb{X}}) - a|\mathcal{I}^{\mathbb{X}}\right) \quad as \quad n \to \infty.$$

Proofs of Theorems 2.1 and 2.2 can be found in [1] and [2], respectively.

3 An application in insurance

Let us consider the following model of insurance claims introduced by Li and Pakes [3] and studied also by Hashorva [4]. Assume that $X_1, X_2, \ldots$ represent claim sizes and that these claims occur at the event times of an orderly point process $(N(t), t \geq 0)$ taking on values in the set of non-negative integers and such that $N(0) = 1$. We are interested in the number of claims that at time t exceed the random level $M_{N(t)} - a$:

$$\mathcal{K}_t(a) = \#\{i \in \{1, \ldots, n\};\ X_i \in (M_{N(t)} - a, M_{N(t)}]\}.$$

Asymptotic properties of $\mathcal{K}_t(a)$ provide information on the long-term behaviour of the considered model of insurance claims.

Theorem 3.1. *Under the assumptions of Theorem* 2.1*, if moreover*

- *(1) and (2) hold,*
- $\mathbb{X} = (X_n, n \geq 1)$ *and* $(N(t), t \geq 0)$ *are independent,*
- $N(t)/t \xrightarrow{a.s.} Z$*, where Z is an rv such that* $\mathbb{P}(Z > 0) = 1$*,*

we have, as $t \to \infty$,

$$\mathcal{K}_t(a)/t \xrightarrow{a.s.} Z\mathbb{P}\left(X_1 > \gamma_1(X_1|\mathcal{I}^{\mathbb{X}}) - a|\mathcal{I}^{\mathbb{X}}\right).$$

For a proof of Theorem 3.1 and some other results on asymptotic properties of $\mathcal{K}_t(a)$ we refer the reader to [2].

References

[1] Buraczyńska, A., Dembińska, A. (2018). A strong ergodic theorem for extreme and intermediate order statistics. *J. Math. Anal. Appl.*, Vol. **468**, pp. 382-399.

[2] Buraczyńska, A., Dembińska, A. (2018). The long-term behaviour of the number of near-maximum insurance claims. *Manuscript*.

[3] Li, Y., Pakes, A. (2001). On the number of near-maximum insurance claims. *Insurance Math. Econom.*, Vol. **28**, pp. 309-323.

[4] Hashorva, E. (2003). On the number of near-maximum insurance claims under dependence. *Insurance Math. Econom.*, Vol. **32**, pp. 37-49.

On moments of discrete order statistics with applications in reliability

Agnieszka Goroncy[1], Anna Dembińska[2]
[1]*Nicolaus Copernicus University*
Toruń, Poland
[2]*Warsaw University of Technology*
Warsaw, Poland
e-mail: [1]gemini@mat.umk.pl, [2]dembinsk@mini.pw.edu.pl

Abstract

We consider a set of possibly dependent and not necessarily identically distributed discrete random variables. For such setting we compute moments of respective order statistics and present applications of this result to reliability theory. Namely we develop a method to establish expectation of lifetimes of coherent systems consisting of possibly dependent and nonhomogeneous components. We focus on the case of systems with multivariate geometrically distributed components' lifetimes.

1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of random variables and by $X_{1:n}, \ldots, X_{n:n}$ denote the respective order statistics. We assume that variables X_i, $i = 1, \ldots, n$ are arbitrarily dependent and non necessarily identically distributed (DNID, for short). Moreover, we focus on the case when the parent distribution of the considered sequence is discrete. Such restrictions are motivated by the need to determine the expectation of lifetime of reliability system in the case when e.g. the system operates in the discrete cycles or when the system's components are exposed to breakdowns that occur in discrete times.

For analogous results on the moments of discrete order statistics from the model with independent and non-identically distributed variates the reader is referred to Davies and Dembińska [1].

Formally, let us consider a coherent system that consists of n elements, where $T_1, T_2, \ldots, T_n$ coincide with the components' lifetimes. By T denote the whole system's lifetime. Let the random variables

$T_1, T_2, \ldots, T_n$ be DNID and take values in the set of non-negative integers. By $P_1, \ldots, P_s$ we denote the minimal path sets of such system and by $C_1, \ldots, C_m$, its minimal cut sets. Note that the system's lifetime is given by

$$T = \max_{1\le j\le s} \min_{i\in P_j} T_i = \min_{1\le j\le m} \max_{i\in C_j} T_i, \tag{1}$$

(see Navarro et al. [5]). Therefore the calculation of the expectation of (1) reduces to computation of the expectation of properly chosen minima (maxima).

2 Main results

First we present a formula describing single moments of order statistics from DNID discrete random variables $X_1, X_2, \ldots, X_n$, assuming that they take values in the set of non-negative integers. We have

$$EX_{r:n}^p = \sum_{i=0}^{p-1} \binom{p}{i} \sum_{m=0}^{\infty} m^i \sum_{s=0}^{r-1} \sum_{\mathcal{P}_s} P\left({}^{(j_1,j_2,\ldots,j_n)}B_s^n\right), \tag{2}$$

for $1 \le r \le n$ and $p = 1, 2, \ldots$, where

$${}^{(j_1,j_2,\ldots,j_n)}B_s^n = \left(\bigcap_{l=1}^{s} \{X_{j_l} \le m\}\right) \cup \left(\bigcap_{l=s+1}^{n} \{X_{j_l} > m\}\right),$$

and the summation over $\mathcal{P}_s$ is executed over all permutations $(j_1, j_2, \ldots, j_n)$ of the set $(1, 2, \ldots, n)$, for which $j_1 < j_2 < \ldots < j_n$ and $j_{s+1} < j_{s+2} < \ldots < j_n$. We agree that sets $\mathcal{P}_0, \mathcal{P}_n$ consist of only one permutation $(j_1, j_2, \ldots, j_n) = (1, 2, \ldots, n)$.

Further, let us consider a coherent system which components' lifetimes $T_1, T_2, \ldots, T_n$ take values in the set of non-negative integers and $P_1, \ldots, P_s$ are minimal path sets of this system. Moreover, let us restrict our attention to the case when components' lifetimes have the multivariate geometric distribution (see Esary and Marshall [4]) with parameters θ_I, $I \subset J$, where J is the class of all nonempty subsets of the set $\{1, 2, \ldots, n\}$. Then using (2) we can show that the expected lifetime of

the system T is given by

$$ET = \sum_{j=1}^{s}(-1)^{j+1} \sum_{1\le k_1<\ldots<k_j\le s} \frac{1}{1-\prod\limits_{\substack{\mathcal{J}\subset\{1,\ldots,n\}\\ \mathcal{J}\cap\bigcup\limits_{l=1}^{j} P_{k_l}\neq\emptyset}} \theta_{\mathcal{J}}}. \tag{3}$$

A detailed presentation of the above results and their proofs, along with a discussion on the subject can be found in Dembińska and Goroncy [3].

3 Applications

In this section we present an application of main results to a specific coherent system. We choose a system from the class of k-out-of-n systems, which are often investigated in the reliability theory. Some properties of k-out-of-n systems in the case of DNID components' lifetimes having the joint discrete distributions are given in Dembińska [2].

Let us consider a system consisting of three components, which lifetimes correspond to random variables T_1, T_2, T_3, and assume that the joint distribution of (T_1, T_2, T_3) is multivariate geometric with parameters $p_{\mathcal{J}}$, $\emptyset \neq \mathcal{J} \subset \{1,2,3\}$, where $p_{\mathcal{J}} \in (0,1)$ if $\mathcal{J} = \{1\},\{2\},\{3\},\{1,2,3\}$ and $p_{\mathcal{J}} = 1$ otherwise. It means that each of the system's component is exposed to a failure in discrete cycles, which it survives with probability p_i, and does not survive with probability $1-p_i$, $i=1,2,3$, respectively. Moreover, all three components are simultaneously subject to breakdowns in cycles, which all of them survive with probability $p_{\{1,2,3\}}$ and do not survive with probability $1-p_{\{1,2,3\}}$. We also assume that the system functions as long as at least two of its components work (2-out-of-3 system). Our aim is to find the expectation of system's lifetime T. We first write down all the minimal path sets of this system, which are $P_1 = \{1,2\}$, $P_2 = \{1,3\}$, $P_3 = \{2,3\}$. Note that there are three single sums of path sets P_i, $i=1,\ldots,3$, three double sums of path sets $P_i \cup P_j$, $i \neq j$, $i,j \in \{1,2,3\}$ and one sum of all path sets $P_1 \cup P_2 \cup P_3$. Therefore,

using (3) we obtain

$$\begin{aligned} ET &= \sum_{j=1}^{3}(-1)^{j+1} \sum_{1\le k_1<\ldots<k_j\le 3} \frac{1}{1-\prod\limits_{\substack{\mathcal{J}\subset\{1,2,3\}\\ \mathcal{J}\cap\bigcup_{l=1}^{j} P_{k_l}\neq\emptyset}} \theta_{\mathcal{J}}} \\ &= \frac{1}{1-p_1p_2\theta}+\frac{1}{1-p_1p_3\theta}+\frac{1}{1-p_2p_3\theta}-\frac{2}{1-p_1p_2p_3\theta}, \end{aligned}$$

where $\theta = p_{\{1,2,3\}}$. Having specific values of the distribution parameters, p_1, p_2, p_3 and θ, we can use the above formula to compute the expected lifetime of the system. Applications of (3) to other reliability systems (i.e. to 2-out-of-4 and a classical bridge system) can be found in [3].

Acknowledgements

The research was supported by the Polish National Science Center Grant no. 2015/19/B/ST1/03100.

References

[1] Davies K., Dembińska, A. (2018). Computing moments of discrete order statistics from non-identical distributions. *J. Comput. Appl. Math.*, Vol. **328**, pp. 340-354.

[2] Dembinska, A. (2018). On reliability analysis of k-out-of-n systems consisting of heterogeneous components with discrete lifetimes. *IEEE Transactions on Reliability*, to appear.

[3] Dembińska A., Goroncy A. (2018). Moments of discrete lifetimes of coherent systems with possibly dependent and nonhomogeneous components. *Manuscript*.

[4] Esary J.D., Marshall A.W. (1973). Multivariate geometric distributions generated by a cumulative damage process, *Naval Postgraduate School Technical Report* NPS55EY73041A.

[5] Navarro J., Ruiz J.M., Sandoval C.J. (2007). Properties of coherent systems with dependent components. *Commun. Stat.-Theory Methods*, Vol. **36**, pp. 175-191.

Analysis G-network with restart at a non-stationary regime and its application

D. Kopats[1], V. Naumenko[2], M. Matalytski[3]
[1,2,3]*Grodno State Univercity of Yanka Kupala*
Grodno, Belarus
e-mail: [1]dk80395@mail.ru, [2]victornn86@gmail.com,
[3]m.matalytski@gmail.com

Abstract

This article discusses the question of restarting the script, when a restart is used by many users of the information network, which can be modeled as a G-network. Negative customer simulate the crash of the script and the resending of the request.

1 Introduction

In everyday life there are many situations when an impatient client after a certain waiting time refuses to wait longer for his job to finish cancels the task and restarts it. For example, downloading from the Internet is the most widely known situation in which is restart can be useful. There are also many other cases. Although restarting is often a simple solution, it can also negatively impact the system to which it is applied since restarting can actually mean increasing the load on the system, thereby exacerbating the problem that it must solve. In this regard it is often necessary to carefully select the restart interval. In recent years, some aspects of the restart problem have been studied. In [1, 2] a stochastic restart model is proposed to minimize the shutdown time. The probability of completing the job during a restart was maximized in [3]. In these works authors considered an individual user sending independent tasks that must be completed in accordance with some distribution of the completion time. Restart has therefore been well studied for scenarios where only one user applies this technique. In this article we consider the case when a restart is applied by several users in one or more shared resources. As a model we will use G-networks with signals. The signals in our model are restarted, as they delete a random job in the queue. The restarted job is returned to the network system queue as other types

of requests and processed with a different processing intensity during restart. In our model, we will use phase-type distributions (PH) for the distribution of service time [4] to be able to reflect the characteristics of real systems. Such distribution is more general than the exponential distribution. The queueing theory with signals has received considerable attention since the original article on positive and negative customers[5], published by E. Gelenbe 26 years ago. Traditional simulation systems are used to represent competition among customers for some resources. The client tasks are moved from the server to the server, they wait for service, but do not interact with each other. Signals are used to change these rules. In G-network with signals, clients are allowed to change signals at the end of their service and the signals interact when they arrive to the queue with client tasks already present in the queue. In addition, signals are never stay to the queue. They try to interact with the tasks of the clients and immediately disappear. For the first time in [6] signals were introduced as a negative customers. Negative customer removes the positive upon its arrival in the queue. A negative customer is never placed in the queue. Positive are ordinary customers that are queued and received service or deleted by negative customers. Under typical assumptions for Markov Queueing System (QS) (incoming Poisson flow for both types of customers, exponential service time for positive customers, Markov routing of customers, open topology, independence) E. Gelenbe proved that in such network the stationary states probabilities have a multiplicative form. Here we present new result for G-networks, when the effect of the signals restarts the servecing of customers. The service time according to the PH distributions, which depend on the type of customer. The analysis will be carried out at a non-stationary regime.

2 Main result

We will consider an open G-network with n QS. In the network seviced r types of positive customers and one type of negative customer (also referred to as signals, the effect of which is to restart one customer in the queue). Simple flow of customers of type c arrives to the system S_i from the external environment (QS S_0) with the rate λ^+_{0ic}, additional simplest signal flow also arrives with the rate $\lambda^-_{0i}, i = \overline{1,n}, c = \overline{1,r}$ .We assume that all customer flows arriving to QS are independent.

The service discipline of positive customers are processor sharing (PS). Suppose, that service time is distributed by a phase basis. Let $\mu_{ic}^{(h)}$ the rate of service customer of type c in the i-th QS in phase $h, i = \overline{1,n}, c = \overline{1,r}, h = \overline{1,H}$. Matrix of probability transitions H_{ic} descripe the phase changes of customers of type c in queue i. Without loss of generality, we assume that phase-type distributions that describe the service time follows the rules: - the initial state has index 1; - the output state has index 0. Thus, service for the customer of type c in i-th QS is a transition from state 1 to state 0 with the following matrix H_{ic}. We have:

$$\forall i, c, h \sum_{q=0}^{H} H_{ic}(h, q) = 1 \tag{1}$$

According to the Markov transition matrix a customer of type c of phase h at the moment of completing its service in the i-th QS can join the j-th QS as a positive customer of type s with probability p_{icjs}^{+}. It can also leave the network with probability $p_{icis}^{+} = 0$. Suppose that a customer can not return to the queue that it just left: for all i, c, and s. Naturally, we have:

$$\forall i, c \sum_{j=1}^{n} \sum_{s=1}^{r} p_{icjs}^{+} + p_{i0c} = 1 \tag{2}$$

Network state descripted by the vector $\vec{k}(t) = (k_1, k_2, ..., k_n, t)$ (dimension $n \times r \times H$), where component $\vec{k}_i$ indicates the state of the i-th QS. The state of the i-th QS compose of component $\vec{k}_{ic}^{(H)}, i = \overline{1,n}, c = \overline{1,r}, h = \overline{1,H}$. In addition, note that $|\vec{k}_i|$ - the total number of customers in the i-th QS. It is clear that $\vec{k}(t)$ the Markov chain.

The signals do not stay in the network. Upon arrival in the queue, the signal interacts with the selected customer and then disappears instantly. If on arrival, the queue is already empty, it also instantly disappears. The selected customer is randomly selected according to the state-dependent distribution that simulates the PS service discipline. In the state $\vec{k}_i$, the probability of selecting of the customer is equal to $\frac{|\vec{k}_{ic}^{(H)}|}{|\vec{k}_i|} u(|\vec{k}_i|)$, and the signal is triggered with probability $\alpha_{ic}^{(H)}$. The result is a request restart: this customer (it has class c and phase h) is routed as a customer of type s in phase 1 with probability β_{ics}. Suppose that for all c, $\beta_{icc} = 0$. We

have:

$$\forall c \sum_{s=1}^{r} \beta_{ics} = 1. \tag{3}$$

Please note that at the end of the service we do not allow the customer to become a signal. In our model this is not required, because we do not want to represent a join restart of a group of customers.
We introduce some notation. Let $I_{ic}^{(h)}$ be a zero dimensional $n \times r \times h$ vector, with the exception of a component with the number $H(r(i-1)+c)$ 1;$P(\vec{k},t)$ - nonstationary probability distribution of the network state $(\vec{k},t)$, if it exists at time t; $u(x)$ is the Heaviside function. Let $(\vec{k}_i + I_{ic}^{(h)})$ (respectively $(\vec{k}_i - I_{ic}^{(h)})$, the state of the i-th QS, obtained by adding (correspondingly decreasing) one customer of type c at the maintenance phase h. Note that the intensity of service of customers of type c at phase h in the queue i. Since the servicing discipline under consideration is a processor partition it $M_{ic}^{(h)}(\vec{k}_i)$ can be written as a function $\mu_{ic}^{(h)}$:

$$M_{ic}^{(h)}(\vec{k}_i) = \mu_{ic}^{(h)} \frac{|\vec{k}_{ic}^{(H)}|}{|\vec{k}_i|} u(|\vec{k}_i|) \tag{4}$$

Since the selection of customes according to the PS discipline, the probability of restarting a customer of type c in phase h, when the QS is in the state $\vec{k}_i$, is

$$Q_{ic}^{(h)}(\vec{k}_i) = \alpha_{ic}^{(h)} \frac{|\vec{k}_{ic}^{(H)}|}{|\vec{k}_i|} u(|\vec{k}_i|). \tag{5}$$

The nonstationary probabilities of the states considered of the network in this case will satisfy the Kolmogorov different-difference equations (DDE) system

$$\frac{dP(\vec{k},t)}{dt} = -\left[\sum_{i=1}^{n}\sum_{c=1}^{r} \lambda_{0ic}^{+} + \sum_{h=1}^{H} M_{ic}^{(h)}(\vec{k}_i) u(k_{ic}^{(h)})\right] P(\vec{k},t) +$$

$$+ \sum_{i=1}^{n}\sum_{c=1}^{r} \lambda_{0ic}^{+} u(|k_{ic}^{(1)}|) P(\vec{k}_i - I_{ic}^{(1)}, t) +$$

$$+ \sum_{i=1}^{n}\sum_{c=1}^{r}\sum_{h=1}^{H} M_{ic}^{(h)}(\vec{k}_i + I_{ic}^{(h)}) p_{i0c} H_{ic}(h,0) P(\vec{k}_i + I_{ic}^{(h)}, t) +$$

$$+\sum_{i=1}^{n}\sum_{c=1}^{r}\sum_{h=1}^{H}\sum_{q=1}^{H} M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)}-I_{ic}^{(q)})H_{ic}(h,q)u(|k_{ic}^{(h)}|)P(\vec{k}_i+I_{ic}^{(h)}-I_{ic}^{(q)},t)+$$

$$+\sum_{i,j=1}^{n}\sum_{c,s=1}^{r}\sum_{h=1}^{H} M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})p_{icjs}^{+}u(|k_{js}^{(1)}|)H_{ic}(h,0)P(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(1)},t)+$$

$$+\sum_{i=1}^{n}\sum_{c=1}^{r}\sum_{h=1}^{H}\lambda_{0i}^{-}Q_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})\sum_{s=1}^{r}\beta_{ics}u(|k_{js}^{(1)}|)P(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(1)},t). \quad (6)$$

System of DDE (6) can be represented as:

$$\frac{dP(\vec{k},t)}{dt}=-\Lambda(\vec{k},t)P(\vec{k},t)+\sum_{i,j=1}^{n}\sum_{c,s=1}^{r}\sum_{h,q=1}^{H}T_{icqjsh}(\vec{k})P(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(q)},t), \quad (7)$$

where

$$\Lambda(\vec{k},t)=\left[\sum_{i=1}^{n}\sum_{c=1}^{r}\lambda_{0ic}^{+}+\sum_{h=1}^{H}M_{ic}^{(h)}(\vec{k}_i)u(k_{ic}^{(h)})\right],$$

$$T_{ichjsq}(\vec{k})=\delta_{0i}\delta_{0c}\delta_{q1}\delta_{h0}\lambda_{0ic}^{+}u(|k_{ic}^{(1)}|)+\delta_{j0}\delta_{s0}\delta_{q0}M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})p_{i0c}H_{ic}(h,0)+$$

$$+\delta_{ji}\delta_{sc}\delta_{qh}M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})p_{i0c}H_{ic}(h,0)+\delta_{q1}M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)}-I_{ic}^{(q)})H_{ic}(h,q)u(|k_{ic}^{(h)}|)+$$

$$+\delta_{q1}\sum_{i,j=1}^{n}\sum_{c,s=1}^{r}\sum_{h=1}^{H}M_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})p_{icjs}^{+}u(|k_{js}^{(1)}|)H_{ic}(h,0)P(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(1)},t)+$$

$$+\delta_{q1}\delta_{h1}\lambda_{0i}^{-}Q_{ic}^{(h)}(\vec{k}_i+I_{ic}^{(h)})\beta_{ics}u(|k_{js}^{(1)}|)$$

Let $P_q(\vec{k},t)$ be approximation of $P(\vec{k},t)$ at q-th iteration, $P_{q+1}(\vec{k},t)$ solution of (7) obtained by successive approximations. Can be shown

$$P_{q+1}(\vec{k},t)=e^{-\Lambda(\vec{k})t}(P(\vec{k},0)+\int_0^t e^{\Lambda(\vec{k})x}\sum_{i,j=1}^{n}\sum_{c,s=1}^{r}\sum_{h,q=1}^{H}T_{icqjsh}(\vec{k})\times$$

$$P_q(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(q)},x)dx$$

As an initial approximation we take the stationary distribution $P_0(\vec{k},t)=$ $=P(\vec{k})=\lim\limits_{t\to\infty}P_q(\vec{k},t)$ which satisfies the relation

$$\Lambda(\vec{k})P(\vec{k})=\sum_{i,j=1}^{n}\sum_{c,s=1}^{r}\sum_{h,q=1}^{H}T_{icqjsh}(\vec{k})P_q(\vec{k}_i+I_{ic}^{(h)}-I_{js}^{(q)}). \quad (8)$$

The following theorems are valid for successive approximations.
Theorem 1. Sequential approximations $P_q(\vec{k},t), q = 0,1,2,...$ converge for $t \to \infty$ to a stationary solution of the system of equations (7).
Theorem 2. The sequence constructed according to scheme (8), for any zero-th approximation bounded in $P_0(\vec{k},t)$ converge for $q \to \infty$ to a unique solution of the system of equations (7).
Theorem 3. Each successive approximation $P_q(\vec{k},t), q \geq 1$, is representable in the form of a convergent power series

$$P_q(\vec{k},t) = \sum_{l=0}^{\infty} d_{ql}^{+-} t^l, \tag{9}$$

whose coefficients satisfy the recurrence relations:

$$d_{ql}^{+-}(\vec{k}) = -\frac{(\Lambda(\vec{k}))^l}{l!}(P(\vec{k},0) + \sum_{u=0}^{l-1} \frac{(-1)^u u!}{\Lambda(\vec{k})^{u+1}} D_{qu}^{+-}(\vec{k})), l \geq 0$$

$$D_{ql}^{+-}(\vec{k}) = \sum_{i,j=1}^{n} \sum_{c,s=1}^{r} \sum_{h,q=1}^{H} T_{icqjsh}(\vec{k}) d_{ql}^{+-}(\vec{k_i} + I_{ic}^{(h)} - I_{js}^{(q)}). \tag{10}$$

Theorem 1-3 are proofed similary as in [8].

References

[1] Moorsel Aad P.A.,Wolter K. (2004). *Analysis and algorithms for restart. Quantitative Evaluation of Systems : Proc. Of the 1st Intern. Conf., The Netherlands* Univercity of Twente; Ed. T.A. Henzinger, Enschede, pp. 195-204.

[2] Moorsel Aad P.A.,Wolter K. (2006). Analysis of restart mechanisms in software systems *IEEE Transactions on Software Engineering*. Vol. **32**, pp. 547-558.

[3] Moorsel Aad P.A.,Wolter K. (2004).Optimal restart times for moments of completion time *IEEE Proceedings Software.*. Vol. **151**, pp. 219-223.

[4] Moorsel F. (1981) Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. *IEEE Proceedings Software.* The Johns Hopkins University Press, Massachusetts.

[5] Gelenbe E. (1991) Product-form queuing networks with negative and positive customers. *Journal of Applied Probability.* Vol. **28**, pp. 656-663.

[6] Gelenbe E. (1993) G-networks with triggered customer movement. *Journal of Applied Probability.* Vol. **30**, pp. 742-748.

[7] Matalytski M., Naumenkoi V. (2016) Stochastic network with nonstandart moving of customers *Publishing Office of GrSU*, Grodno

[8] Matalytski M., Kopats D. (2017) Finding expected revenues in G-network with signals and customers batch removal *Probability in the Engineering and Informational Sciences* Vol. **31**, pp. 561-575.

Degree-degree dependencies and clustering in configuration graphs

Marina Leri[1], Yuri Pavlov[2]
[1,2]*Institute of Applied Mathematical Research*
of the Karelian Research Centre of the Russian Academy of Sciences
Petrozavodsk, Russia
e-mail: [1]leri@krc.karelia.ru, [2]pavlov@krc.karelia.ru

Abstract

We consider power-law configuration graphs in both random and non-random environment, meaning that the parameter of vertex degree distribution is either a fixed value or a random variable. We study such properties of our models as assortativity and clustering.

Random graphs are widely used for modeling complex communication networks. Transport, electricity, telephone and social networks, along with the main worldwide web – the Internet are just some examples. Numerous observations of real networks showed (see e.g. [3]) that vertex degrees can be viewed as independent identically distributed random variables following a common power-law distribution. One of the best known types of network models are the so called configuration graphs, for the first time introduced in [2]. Each vertex degree of a configuration graph is equal to the number of incident semiedges. All semiegdes are numbered in an arbitrary order. The sum of vertex degrees of any graph has to be even. If otherwise, one semiedge is added to an equiprobably chosen vertex. The graph is constructed by equiprobably joining all semiedges one to another to form edges. Obviously, such a graph may have loops and multiple edges. Let ξ be a random variable equal to the degree of an arbitrary vertex. For modeling complex networks the authors of [7] suggested using configuration graphs, where

$$\mathbf{P}\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \qquad \tau > 0,\ k = 1, 2, \ldots, \tag{1}$$

which agrees with an important feature of real networks – power-law node degree distribution. In [7] and in many following works the structure of configuration graphs is studied both by asymptotical methods through consideration of its limit behaviour as the number of graph vertices tends

to infinity and by simulation modeling. It is well known [3] that for a vast number of real networks the parameter τ of the distribution (1) lies in the interval $(1, 2)$. Attempts have recently been made to use configuration graphs for modeling forest fires [4, 5], where the parameter τ turned out to be greater than 2. There are also some works (see e.g. [1]), where the vertex distribution can change with the growth of the number of graph vertices or even be random.

Real networks are often described using various numerical characteristics that show both local and global network properties. Naturally, these characteristics are to be taken into account when constructing models so that not only the node degree distribution would fit (1), but other network properties would also fit in. The best known among such characteristics are assortativity and clustering coefficients.

Assortativity coefficient A is meant for estimating correlation between the degrees of incident vertices, wherefore in [8] the authors proposed to use Pearson correlation coefficient for this purpose. It is obvious that if vertices with high degrees mostly connect to vertices also having high degrees, then the assortativity coefficient will be positive, otherwise it will be negative.

To estimate the degree of the graph clusterization we used the following global clustering coefficient C [6]:

$$C = \frac{3 \times \text{number of graph triangles}}{\text{number of connected triples of vertices}}, \tag{2}$$

where a "connected triple" means a single vertex connected by edges to two others [6]. Loops are not counted and multiple edges are considered as one. The value of the global clustering coefficient can be interpreted as the probability that in an equiprobably chosen pair of connected edges having a common vertex the other ends of the edges are also connected by an edge.

Below are the results of our study of the characteristics of the coefficients A and C in the considered configuration model. The results were obtained for two cases of changing the parameter τ: firstly, for different fixed values of τ and, secondly, when τ was a random variable following either uniform or truncated normal distribution on a predefined interval (a, b). The dependencies of the coefficients A and C on the number of graph vertices N and the parameter τ were found by a simulation technique. The number of graph vertices ranged from 1000 to 10000 with a

step of 500. In the case of a fixed τ its values ranged from 1.1 to 2.5 with a step of 0.1. For each pair (N, τ) 100 simulations were run. Thus, the total number of generated graphs was 28500. Based on the obtained results we derived the following regression dependence of the coefficient A on N and τ:

$$A = -0.58N^{-0.888+0.896/\tau} \tag{3}$$

with the determination coefficient $R^2 = 0.93$.

In the case of random τ we considered three intervals (a, b). The first interval $(1, 2)$ corresponds to a well-known property of communication networks. The second interval $(2, 3)$ is connected with forest fire modeling and the third one $(1, 3)$ is a generalization of the first two. For each interval the same 19 values of N as in the case of fixed τ were taken, and the same number of graphs (100 graphs) was generated. Hence, 1900 simulations were run for uniform and normal distributions of τ each. For the truncated normal distribution the expectation of ξ at each interval (a, b) was defined as the middle value $(a+b)/2$ and the standard deviation $\sigma = (b - a)/6$ in accordance with the three-sigma rule. In what follows let us denote by $\tau \sim \mathbf{U}(a, b)$ and $\tau \sim \mathbf{N}(a, b)$ the cases where the random variable τ follows either uniform or truncated normal distribution on the interval (a, b), respectively. We obtained the following dependencies of A on N:

$$\begin{aligned}
\tau \sim \mathbf{U}(1,2): &\qquad \mathbf{A} = -0.54 \cdot N^{-0.16} &&(R^2 = 0.99);\\
\tau \sim \mathbf{U}(2,3): &\qquad \mathbf{A} = -0.32 \cdot N^{-0.46} &&(R^2 = 0.94);\\
\tau \sim \mathbf{U}(1,3): &\qquad \mathbf{A} = -0.22 \cdot N^{-0.08} &&(R^2 = 0.99);\\
\tau \sim \mathbf{N}(1,2): &\qquad \mathbf{A} = -0.24 \cdot N^{-0.12} &&(R^2 = 0.99);\\
\tau \sim \mathbf{N}(2,3): &\qquad \mathbf{A} = -1.02 \cdot N^{-0.64} &&(R^2 = 0.94);\\
\tau \sim \mathbf{N}(1,3): &\qquad \mathbf{A} = -0.28 \cdot N^{-0.23} &&(R^2 = 0.99).
\end{aligned}$$

For the clustering coefficient C and fixed values of τ we derived the following regression relation:

$$C = 0.457N^{-1.215+1.09/\tau}, \qquad R^2 = 0.92. \tag{4}$$

For a random τ we obtained the following dependencies:

$$\begin{aligned}
\tau \sim \mathbf{U}(1,2): &\qquad \mathbf{C} = 0.38 \cdot N^{-0.38} &\qquad (R^2 = 0.99);\\
\tau \sim \mathbf{U}(2,3): &\qquad \mathbf{C} = 0.6 \cdot N^{-0.84} &\qquad (R^2 = 0.99);\\
\tau \sim \mathbf{U}(1,3): &\qquad \mathbf{C} = 0.2 \cdot N^{-0.38} &\qquad (R^2 = 0.98);\\
\tau \sim \mathbf{N}(1,2): &\qquad \mathbf{C} = 0.33 \cdot N^{-0.38} &\qquad (R^2 = 0.99);\\
\tau \sim \mathbf{N}(2,3): &\qquad \mathbf{C} = 0.44 \cdot N^{-0.85} &\qquad (R^2 = 0.99);\\
\tau \sim \mathbf{N}(1,3): &\qquad \mathbf{C} = 0.21 \cdot N^{-0.48} &\qquad (R^2 = 0.98).
\end{aligned}$$

We hope that these results will help in constructing models of specific networks in the form of configuration graphs with vertex degree distribution (1) by choosing the best fitting fixed values of the parameter τ or by finding the distribution of a random τ fitting the real values of the assortativity and clustering coefficients of these networks.

The study was carried out under state order to the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research KarRC RAS) and partly supported by the Russian Foundation for Basic Research, grant 16-01-00005a. The research was carried out using the equipment of the Core Facility of the Karelian Research Centre of the Russian Academy of Sciences.

References

[1] Biaconi G., Barabasi A.-L. (2001). Bose-Einstein condensation in complex networks. *Physical Review Letters*. Vol. **86(24)**, pp. 5632-5635.

[2] Bollobas B. A. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*. Vol. **1(4)**, pp. 311-316.

[3] Hofstad R. (2017). *Random Graphs and Complex Networks. Vol. 1.* Cambridge University Press, Cambridge.

[4] Leri M., Pavlov Yu. (2016). Forest fire models on configuration random graphs. *Fundamenta Informaticae*. Vol. **145(3)**, pp. 313-322.

[5] Leri M., Pavlov Yu. (2014). Power-law random graphs' robustness: link saving and forest fire model. *Austrian Journal of Statistics*. Vol. **43(4)**, pp. 229-236.

[6] Newman M.E.J. (2003). The structure and function of complex networks. *SIAM Rev*. Vol. **45(2)**, pp. 167-256.

[7] Reittu H., Norros I. (2004). On the power-law random graph model of massive data networks. *Performance Evaluation*, Vol. **55(1-2)**, pp. 3-23.

[8] Watts P.J., Strogatz S.H. (1998). Collective dynamics of "small world" networks. *Nature*. Vol. **393(6684)**, pp. 440-442.

Generating Goldbach numbers at random: a convergence in mean

Ljuben Mutafchiev
American University in Bulgaria and
Institute of Mathematics of the Bulgarian Academy of Sciences
Blagoevgrad and Sofia, Bulgaria
e-mail: ljuben@aubg.edu

Abstract

We consider a model of generating random Goldbach numbers that do not exceed a certain large positive even integer. We show that these random numbers, appropriately normalized, converge in mean to the maximum of two independent and uniformly distributed random variables.

Let Σ_{2n} be the set of all partitions of the even integers from the interval $(4, 2n], n > 2$, into two odd prime parts (summands). We select a partition from the set Σ_{2n} uniformly at random (i.e., with probability $1/|\Sigma_{2n}|$). The number $2G_n$, partitioned by this selection, is called a Goldbach number. In [3] we have proved that G_n/n converges weakly, as $n \to \infty$, to the maximum M of two independent random variables which are uniformly distributed in the interval $[0, 1]$. The proof is based on generating functions and on a classical Tauberian theorem due to Hardy, Littlewood and Karamata (see, e.g., [4, Theorem 8.7]). Since G_n/n is bounded for all $n > 2$, well known properties of weak convergent sequences of random variables imply that the moments of G_n/n tend to the moments of the random variable M (see, e.g., [1, p. 316]).

In this work, we use probabilistic tools to establish a stronger result on the convergence of the sequence $\{G_n/n\}_{n>2}$. Further on, by $\mathbb{E}$ we denote the expectation with respect to the Lebesgue measure on the interval $[0, 1]$.

Theorem 1. *We have*

$$\lim_{n\to\infty} \mathbb{E}\left(|G_n/n - M|\right) = 0,$$

that is, $\{G_n/n\}_{n>2}$ converges in mean to the random variable M, whose probability density function equals $2x$ if $0 \le x \le 1$, and zero otherwise.

The proof of Theorem 1 is based on a comparison between the asymptotic properties of the sequences $\{G_n/n\}_{n>2}$ and $\{T_n/n\}_{n>2}$, where $T_n = \max\{k, l\}$ and k and l are two integers chosen uniformly at random and independently among the integers in the interval $(2, n]$. Furthermore, we also apply Slutsky's theorem on the modes of convergence of sequences of random variables (see, e.g., [1, p. 318]).

Our study is related to the famous Goldbach binary conjecture formulated in 1742. In the context of the model of generating random Goldbach numbers, it can be stated as follows: for all $n > 2$, the support of the random variable $2G_n$ is the set of all even integers from the interval $(4, 2n]$. This problem still remains unsolved.

Remark. In his invited address at the 1912 International Congress of Mathematicians, held in Cambridge, Edmund Landau [2] presented a survey on the developments in the theory of prime numbers and the Riemann zeta function. Besides this he mentioned four specific problems about primes which he considered as "unattackable at the present state of science". The second problem in his list was the Goldbach binary problem. A comprehensive survey on current studies of Landau's open problems and several interesting historical remarks concerning the Goldbach conjecture are given by János Pintz in [5].

References

[1] Grimmett G., Stirzaker D. (1995). *Probability and Random Processes*. Oxford Univ. Press, Oxford.

[2] Landau E. (1912). Gelöste und ungelöste Probleme aus der Theorie der Primzahlverteilung und der Riemanschen Zetafunktion. *Jahresber. Deutsche Math. Ver.* Vol. **21**, pp. 208–228.

[3] Mutafchiev L. (2015). A note on Goldbach partitions of large even integers. *Electr. J. Combin.* Vol. **22**, #P1.40.

[4] Odlyzko A. M. (1995). Asymptotic enumeration methods. *In: Handbook of Combinatorics, Vol. II (R. Graham, M. Grötschel and L. Lovász, Eds.)*. Elsevier, Amsterdam.

[5] Pintz J. (2009). Landau's problems on primes. *Journal de Théorie des Nombres de Bordeaux*. Vol. **21**, pp. 357–404.

Multivariate regular variation and the power series distributions

A. L. Yakymiv
Steklov Mathematical Institute of Russian Academy of Sciences
Moscow, Russia
e-mail: arsen@mi.ras.ru

Abstract

We give a definition of regular variation of multiple power series at some boundary point of it's convergence. As an application, we formulate limit theorem for multivariate power series distributions.

1 Introduction

Let $(a(i) \geq 0,\ i \in Z_+^n \equiv (0, 1, 2 \dots)^n)$ be any multiple sequence such that the power series

$$B(x) = \sum_{i \in Z_+^n} a(i)x^i \equiv \sum_{i_1,\dots,i_n \in Z_+} a(i_1, \dots, i_n)x_1^{i_1} \dots x_n^{i_n}$$

converges for $x = (x_1, \dots, x_n) \in [0, 1)^n$ (it is excluded the trivial case $a(i) \equiv 0,\ i \in Z_+^n$). For $x \in (0, 1)^n$, we say that the random variable (r.v.) ξ_x has the power series distribution $B(x)$ if

$$\mathsf{P}\{\xi_x = i\} = \frac{a(i)x^i}{B(x)}.$$

for an arbitrary $i \in Z_+^n$. Such distributions were introduced in Noack's fundamental work [5] (1950). Begining from Kolchin [6] (1968), they are widely used in the general allocation scheme and in probabilistic combinatorics. Bibliography one can see in Johnson N.L. et al [3, 4] (1997,2005) and in Kolchin [7] (1999). Further, the definition of regular variation of multiple power series at some boundary point of it's convergence and also the definition of one-sided weak oscillation of it's cofficients at infinity are given. Using these definitions, we give in the next section a limit theorem for the r.v. having the multiple power series distribution. This theorem generalises the recent one-dimensional Timashev's result [13] (2018).

2 Main Result

Definition 1. *Let an arbitrary sequence* $b = b(k) = (b_1(k), \ldots, b_n(k)) \in (0,\infty)^n,\ k \in N$ *be given with* $b_j = b_j(k) \to \infty,\ \forall j = 1, \ldots, n$ *as* $k \to \infty$. *We say that* $B(x)$ *is regularly varying as* $x \uparrow \mathbf{1}$ *along* $b = b(k)$ *if*

$$\frac{B(\exp(-\lambda/b))}{B(\exp(-\mathbf{1}/b))} \to \Psi(\lambda) \in (0,\infty)$$

for any fixed $\lambda = (\lambda_1, \ldots, \lambda_n) \in (0,\infty)^n$ *as* $k \to \infty$. *Here* $\mathbf{1} = (1, \ldots, 1),\ \lambda/b = (\lambda_1/b_1, \ldots, \lambda_n/b_n)$ *and* $\exp(y) = (\exp(y_1), \ldots, \exp(y_n))$ *for* $y = (y_1, \ldots, y_n) \in R^n$.

Definition 2. *We also say that the sequence* $(a(i) \ge 0,\ i \in Z_+^n)$ *with* $(a(i) > 0,\ i \in Z_+^n,\ i_j > C > 0,\ \forall j = 1, \ldots, n)$ *be one-sided weakly oscillatory at infinity along the sequence* $b = b(k)$ *if, for an arbitrary sequence* $z_j = z_j(k) > 1, z_j = 1 + o(1)$ *and for every* $j = 1, \ldots, n$ *one from the next inequalities*

$$\liminf_{k\to\infty} a(b_1, \ldots, b_{j-1}, z_j b_j, b_{j+1}, \ldots, b_n)/a(b) \ge 1;$$

$$\limsup_{k\to\infty} a(b_1, \ldots, b_{j-1}, z_j b_j, b_{j+1}, \ldots, b_n)/a(b) \le 1.$$

takes place. Here $a(b) = a([b]) \equiv a([b_1], \ldots, [b_n])$ *for* $b \notin Z_+^n,\ b \in R_+^n$.

Theorem 1. *Let* $B(x)$ *be regularly varying as* $x \uparrow \mathbf{1}$ *along the sequence* $b = b(k)$ *and an arbitrary vector* $u \in (0,\infty)^n$ *be fixed. Then for* $x = \exp(-u/b)$

$$\xi_x(\mathbf{1} - x) \xrightarrow{d} \eta(u),$$

as $k \to \infty$ *were r.v.* $\eta(u)$ *has the Laplace transform* $\Psi((\lambda+\mathbf{1})u)/\Psi(u)$. *It follows from (1) that the function* $\Psi(\lambda)$ *is the Laplace transform of some* σ*-finite measure* ν *on* R_+^n. *Let this measure* ν *be absolutely continuous in* $(0,\infty)^n$ *with continuous density* $\varphi(\cdot)$. *Also assume that the sequence* $(a(i),\ i \in Z_+^n)$ *be one-sided weakly oscillatory at infinity along the sequence* $b = b(k)$. *Then the distribution of r.v.* $\eta(u)$ *is also absolutely continuous in* $(0,\infty)^n$ *with the density*

$$\psi_u(y) = \frac{\varphi(y/u)e^{-(y,\mathbf{1})}}{\prod_{j=1}^n u_j \Psi(u)}, \quad \forall y \in (0,\infty)^n.$$

At the same time

$$\frac{\mathsf{P}\{\xi_x = [y/(\mathbf{1}-x)]\}}{\prod_{j=1}^n (1-x_j)} \to \psi_u(y)$$

as $k \to \infty$ *uniformly by* $y \in K$*, for an arbitrary fixed compact* $K \subset (0,\infty)^n$*.*

Note that for $\nu(\partial R_+^n) = 0$, it follows from this theorem that the density $\psi_u(y)$ has the Laplace transform $\Psi((\lambda+\mathbf{1})u)/\Psi(u)$. The proof of the theorem 1 uses the corresponding tauberian theorem from [15].

3 Example

Definition 3. *We say that* $B(x)$ *is regularly varying for* $x \uparrow \mathbf{1}$ *if*

$$\frac{B(\exp(-\lambda/b))}{B(x)} \to \Psi(\lambda) \in (0,\infty)$$

as $x \to \mathbf{1}, x \in (0,1)^n$ *for any fixed* $\lambda = (\lambda_1,\dots,\lambda_n) \in (0,\infty)^n$ *where* $b = -\mathbf{1}/\ln x \equiv -(1/\ln x_1,\dots,1/\ln x_n)$*.*

Definition 4. *We also say that the sequence* $(a(i) \geq 0,\ i \in Z_+^n)$ *with* $(a(i) > 0,\ i \in Z_+^n,\ i_j > C > 0,\ \forall j = 1,\dots,n)$ *be one-sided weakly oscillatory at infinity if, for an arbitrary sequence* $z_j = z_j(k) > 1, z_j = 1 + o(1)$ *and for every* $j = 1,\dots,n$ *one from the next inequalities*

$$\liminf a(b_1,\dots,b_{j-1},z_jb_j,b_{j+1},\dots,b_n)/a(b) \geq 1;$$

$$\limsup a(b_1,\dots,b_{j-1},z_jb_j,b_{j+1},\dots,b_n)/a(b) \leq 1.$$

takes place as $\min_{l=1,\dots,n} b_l \to +\infty$*.*

The next corrolary of theorem 1 holds.

Corollary 1. *If* $B(x)$ *be regularly varying for* $x \uparrow \mathbf{1}$ *then*

$$\xi_x(\mathbf{1}-x) \xrightarrow{d} (\gamma_1,\dots,\gamma_n).$$

as $x \uparrow \mathbf{1}$*, where r.v.* $\gamma_1,\dots,\gamma_n$ *are independent and r.v.* γ_j *has* Γ*-distribution with some parameter* $\alpha_j \geq 0$*, for any* $j = 1,\dots,n$*. Moreover,*

if the sequence $(a(i) \geq 0,\ i \in Z_+^n)$ *be one-sided weakly oscillatory at infinity and* $\alpha_j > 0,\ \forall j = 1, \ldots, n$, *then*

$$\frac{\mathsf{P}\{\xi_x = [y/(\mathbf{1} - x)]\}}{\prod_{j=1}^n (1 - x_j)} \to \frac{\prod_{j=1}^n y_j^{\alpha_j - 1} e^{-y_j}}{\prod_{j=1}^n \Gamma(\alpha_j)}$$

as $x \uparrow \mathbf{1}$ *uniformly by* $y \in K$, *for an arbitrary compact* $K \subset (0, \infty)^n$.

(For convenience, we say that r.v. γ_j has Γ-distribution with parameter 0, if $\mathsf{P}\{\gamma_j = 0\} = 1$). It suffices to apply the theorem 1 for every sequence $b = b(k)$ with $b_j = b_j(k) \to \infty,\ \forall j = 1, \ldots, n$ as $k \to \infty$.

4 On some definitions of multivariate regularly varying functions

Definition 3 goes back to Bajašanski and Karamata [1] (1968/69). Definition 1 was used by Diamond [2] (1987) and by Omey in his book [8] (1989) for $n = 2$ and by the author in [15] (2016) for an arbitrary natural n. A short rewiev on different definitions of multivariate regularly varying functions one can find in author's book [14, capter 1.9] (2005). A little later, Resnick [9, 10] (2007, 2008) considered some very interesting classes of functions. In paticular, they allow to obtain the next asymptotics:

$$t\mathsf{P}\{(X/a(t) \geq x, Y/b(t) \geq y)\} \to \nu_0(x, y) \equiv \nu([x, \infty] \times [y, \infty]),\ \forall x, y > 0,$$

where r.v. X and Y have applications and concrete intepretation in preferential attachment networks, see Resnick et al [11, 12] (2015, 2016). The authors of last two papers say that that the regular variation is nonstandard, if $a(t)$ and $b(t)$ have different order at infinity. For instance, such is the regular variation in the last relation, and such are the most classes of functions defined in already mentioned papers [2, 8, 15].

References

[1] Bajašanski B., Karamata J. 1968/69. Regularly varying functions and the principle of equicontinuity. *Publ. Ramanujan Inst.*. Vol. **1**, pp. 235-246.

[2] Diamond Ph. Slowly varying functions of two varibles and a Tauberian theorem for the double Laplace transform.- Applicable Analysis, 1987, v. 23, p. 301–318.

[3] Johnson N.L., Kemp A.V., Balakrishnan N. (1997) *Discrete multivariate distributions*. A Wiley-Interscience Publication. (John Wiley and Sons), New York.

[4] Johnson N.L., Kemp A.V., Kotz S. (2005). *Univariate discrete distributions*. Third edition. A Wiley-Interscience Publication. (John Wiley and Sons), Hoboken.

[5] Noack A. (1950). A class of random variables with discrete distribution. *Annals of Mathematical Statistics*. Vol. **21**, pp. 127-132.

[6] Kolchin V.F. (1968). A certain class of limit theorems for conditional distributions (Russian). *Litovsk. Mat. Sb.* Vol. **8**, pp. 53-63.

[7] Kolchin V.F. (1999). *Random graphs*. Cambridge University Press, Cambridge.

[8] Omey E. (1989). *Multivariate regular variation and application in probability theory*. Electica, Brussel.

[9] Resnick S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

[10] Resnick S. (2008). *Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws*. *Stochastics* Vol. **80**, no. 2-3, pp. 269-298.

[11] Resnick S., Samorodnitsky G. (2015). Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks. *Extremes*, Vol. **80**, no. 3, pp. 269-298.

[12] Resnick S., Samorodnitsky G., Towsley D., Davis R., Willis A., Wan Ph. (2016). Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *J. Appl. Probab.*, Vol. **53**, no. 1, pp. 146-161.

[13] Timashev A.N. (2018). Limit theorems for the power series distribution with a finite convergence radius (Russian). *Teor. Veroyatn. Primen.* Vol. **63**, pp. 57-69.

[14] Yakimiv A.L. (2005). *Probabilistic applications of Tauberian theorems*. Modern Probability and Statistics, VSP, Leiden, Boston.

[15] Yakymiv A.L. (2016). A Tauberian theorem for multiple power series, *Sb. Math.*, Vol. **207**, pp. 287-313.

Stochastic times and defautable derivatives

Tsvetelin S. Zaevski[1], Ognyan Kounchev[2]
[1,2]*Institute of Mathematics and Informatics,*
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: [1]t_s_zaevski@math.bas.bg, [2]kounchev@math.bas.bg

Abstract

The purpose of this paper is to present a scheme for deriving the partial differential equation (PDE) for the price of a defaultable derivative. The asset price is modeled by a geometric Brownian motion, stopped in a stochastic time. The filtration is enlarged in a proper way to incorporate it in the information flow.

1 Introduction

Defautable derivatives are a modification of ordinary (European) derivatives. The main difference between them is that the ordinary derivatives pay at fixed maturity date, whereas defautables pay at some stochastic (stopping) time. Since every fixed moment may be a stopping time, the ordinary derivatives are a subset of the defautable. There are many studies for this kind of derivatives – we refer to Jarrow and Turnbull [10], Hull and White [9], Cooper and Martin [5], Duffie and Singleton [7, 8], Bielecki et al. [2], and Bielecki and Rutkowski [3].

From the classical work of Black and Scholes [4] the most popular approach is to model the basic asset by a geometric Brownian motion. Despite the shortcomings of the assumption for Gaussian increments it has major advantages – it is very suitable for work and, more important, by its infinitesimal generator the results can be easily generalized for the larger class of Lévy processes or more general class of the Feller processes. For these reasons we shall work in the Gaussian framework too.

There are two main modes for introducing a random event – first, to introduce independently of the asset prices random time and then enlarge by it the information flow (the filtration) or second, to use some random time, known w.r.t. the current information flow – a stopping time. We shall present the first technique in that study. The assumption imposed

by us is that the stopping time is totally inaccessible with an absolutely continuous compensator which shall allow us to derive the PDE for the asset price.

A traditional choice of the stopping time distribution is the exponential one, because it has some very nice properties – it is positive and memoryless. We shall begin with this distribution and later shall essentially generalize by introducing different dependencies. Hence, our assumptions are that the asset price moves as a geometric Brownian motion before the stopping time, in which it falls with 100η percents and stops; here η is deterministic or stochastic.

The fundamental theorem of asset pricing says that the market is arbitrage-free if there exists a martingale measure. That is why, together with building of a realistic real-world model, we derive the risk neutral measures too.

One of the main contributions of this article is deriving of the corresponding to the Black and Scholes [4] and Merton [12] equations – for different assumptions we obtain PDEs (9) and (13).

The article is organized as follows. In section 2 we state the general model of weak path dependent derivatives and in section 3 we examine the defautables as a particular case.

2 Definitions

Let S_t be a Feller Markov price process and τ be a totally inaccessible stopping time under the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, Q)$ which satisfies the usual conditions, the filtration is right continuous and complete, see Dellacherie [6] or Protter [13]. The measure Q is risk-neutral, i.e. the discount price processes have to be martingales. Now we shall work under this measure, but later in section 3 we shall use both of real world measure P and risk-neutral Q. For simplicity, we suppose that the risk-free rate of return is the constant r. Let us denote by $\mathcal{B}$ the differential operator

$$\mathcal{B}f\left(t,x\right) = f_t\left(t,x\right) + \mathcal{A}f\left(t,x\right) - rf\left(t,x\right), \tag{1}$$

where $\mathcal{A}$ is the infinitesimal generator of the Feller semigroup. Let N_τ be the amount payable at the event moment. Now we assume that the price of the defaultable derivative can be presented as

$$Y_t = A(t, S_t)\,\Phi_t + e^{r(t-\tau)} N(\tau, S_\tau)\,\Lambda_t, \tag{2}$$

where Λ_t and Φ_t are the indicator processes $\Lambda_t = I_{\{\tau \le t\}}$ and $\Phi_t = 1 - \Lambda_t$. Let λ_t be the hazard rate of the stopping time, i.e. the compensator of Λ_t is given by

$$\int_0^t \lambda \Phi_{u^-} du. \tag{3}$$

Using the Itô product formula and the value of the quadratic variation $[A(\cdot, S), \Phi]_t = (A(\tau, S_\tau) - A(\tau, S_{\tau^-}))\,\Lambda_t$ we obtain equality[1]

$$e^{-rt} Y_t = Y_0 + \int_0^t e^{-ru} \Phi_{u^-} \mathcal{B} A(u, S_{u^-})\, du + e^{-r\tau} (N_\tau - A(\tau, S_\tau))\,\Lambda_t + M_t. \tag{4}$$

Hence, since the discount derivative process has to be a martingale, we can formulate the following proposition

Proposition 1. *The price of the defaultable derivative can be presented as a sum of the corresponding European derivative price, $A(t, S_t)$, plus the price of derivative which pays amount of $N_\tau - A(\tau, S_\tau)$ at the event time if it occurs before the maturity T.*

Another important proposition for stopping times is placed bellow and shall be used later.

Proposition 2. *Let τ be a stopping time with compensator A and U be a $\mathcal{F}_\tau$-measurable random variable. Let β_t be a process such that $\beta_\tau = E[U|\,\mathcal{F}_{\tau^-}]$. Then the compensator of $U\Lambda_t$ is*

$$\int_0^t \beta_u dA_u.$$

[1]From now we shall use the notation M_t for a suitable $\mathcal{F}_t$-martingales, the form of which are not important – in different equations M_t will be used for different martingales.

3 Brownian filtration enlarged with a stopping time

3.1 Constant intensity

Let us have a Brownian motion w.r.t. the real-world measure P, and independent from it exponentially distributed random time τ with intensity λ. It plays the role of the default time. First, we have to enlarge the filtration in some way (we can use progressive, minimal, or initial enlargement) to make the random time to be stopping – for example see Mansuy and Yor [11] or Aksamit and Jeanblanc [1]. The hazard rate of this stopping time is the constant λ. Let the asset price, S_t, before the default coincide with the Geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

and after the default the asset price falls with 100η percents and stops, i.e. for $t \geq \tau$, $S_t = (1-\eta) X_\tau$. Note that defined in that way the asset price at the event moment is proportional to the value just before it. So, the price processes can be written as

$$\begin{aligned} S_t = &\quad X_t \Phi_t + (1-\eta) X_\tau \Lambda_t & (5) \\ Y_t = &\quad A(t, X_t) \Phi_t + N(\tau, (1-\eta) X_t). & (6) \end{aligned}$$

Since $S_\tau = (1-\eta) X_\tau$ and equation (3) is in force, the conditions of proposition 2 are satisfied and therefore the compensator of $(1-\eta) X_\tau \Lambda_t$ is

$$\int_0^t \Phi_{u^-} \lambda (1-\eta) X_{u^-} du. \qquad (7)$$

First, we have to find the asset price dynamic under the risk-neutral measure – i.e. the value of μ. We shall change the measure with accordance to the Girsanov theorem, see [1]. We can freely change μ and λ to μ^Q and λ^Q. So, under the measure Q, the Itô product formula leads us to

$$e^{-rt}S_t = S_0 + \int_0^t e^{-ru}\Phi_{u^-}X_{u^-}\left(\mu^Q - \lambda^Q\eta - r\right)du + M_t^Q$$

and therefore $\mu^Q = \lambda^Q\eta + r$. In that way we find the set of risk-neutral measures – all constants μ^Q and $\lambda^Q > 0$, such that $\mu^Q = \lambda^Q\eta + r$, lead to a risk-neutral measure.

We continue with pricing of the derivative. Using $N_\tau = N\left(\tau, (1-\eta)X_\tau\right)$, equation (4), and proposition 2 we obtain equation

$$\begin{aligned} &e^{-rt}Y_t = Y_0 + M_t^Q \\ &\int_0^t e^{-ru}\Phi_{u^-}\left(\mathcal{B}^Q A\left(u, X_{u^-}\right) + \lambda^Q\left(N\left(u, (1-\eta)X_{u^-}\right) - A\left(u, X_{u^-}\right)\right)\right)du. \end{aligned} \tag{8}$$

Since $\mu^Q = \lambda^Q\eta + r$ under the risk-neutral measure we obtain the following theorem:

Theorem 1. *If the intensity and loss rates are constants then the price of defautable derivative solves the boundary value problem for a PDE*

$$\begin{aligned} &-\left(r+\lambda^Q\right)A\left(t,x\right) + A_t\left(t,x\right) + \left(r+\lambda^Q\eta\right)xA_x\left(t,x\right) \\ &+\frac{1}{2}\sigma^2x^2A_{xx}\left(t,x\right) + \lambda^Q N\left(t,(1-\eta)x\right) = 0 \\ &A\left(T,x\right) = M\left(x\right). \end{aligned} \tag{9}$$

The function $M(\cdot)$ determines the maturity payment if the default is not occur before the moment T.

3.2 Stochastic intensity

An interesting question is what changes if we remove the assumption that the rate of loss, η, and the default intensity, λ, are constants. We can remove the first restriction in two ways – 1. setting that the loss rate is a random variable independent of X_t or 2. a function of τ and X_τ. We shall combine both assumptions. Let the random variable ξ with

distribution $\nu(dy)$ lead to 1. and $\eta_\tau = \eta(\tau, X_\tau)$ lead to 2. Of course, we have to state that ξ has support $[0,1]$ and $\eta(\cdot,\cdot) : R^+ \times R^+ \to [0,1]$. Let also λ_t be the hazard intensity process. Hence, after the default the asset price has to be

$$S_\tau = (1 - \eta(\tau, X_\tau)\xi) X_\tau. \tag{10}$$

When we change the measure using the general Girsanov theorem, see again Aksamit and Jeanblanc [1], we reach the dynamic

$$dX_t = (\mu + \sigma\psi_t) X_t dt + \sigma X_t B_t^Q,$$

Also, the stopping time has a Q-hazard rate $\lambda_t^Q = (1+\gamma_t)\lambda_t$ and the random variable ξ has, an equivalent to $\nu(dy)$, Q-distribution $\nu^Q(dy)$. Since

$$E^Q\left[(1 - \eta(\tau, X_\tau)\xi) X_\tau | \mathcal{F}_{\tau^-}\right] = X_\tau \left(1 - \eta(\tau, X_\tau) \int_0^1 y\nu^Q(dy)\right), \tag{11}$$

the Q-compensator of $(1 - \eta(\tau, X_\tau)\xi) X_\tau \Lambda_t$ has to be

$$\int_0^t \lambda_u^Q \Phi_{u^-} X_{u^-} \left(1 - \eta(u, X_{u^-}) \int_0^1 y\nu^Q(dy)\right) du$$

and therefore the risk-neutral condition is

$$\mu_t^Q \equiv \mu + \sigma\psi_t = r + (1+\gamma_t)\lambda_t \eta(t, X_t) \int_0^1 y\nu^Q(dy). \tag{12}$$

Of course, if λ_t^Q is a constant and 1. $\nu^Q(dy) \equiv \nu(dy) = \delta(y-\eta)dy$ and $\eta(t,x) \equiv 1$ or 2. $\nu^Q(dy) \equiv \nu(dy) = \delta(y-1)dy$ and $\eta(t,x) \equiv \eta$, equation (12) leads to $\mu^Q = \lambda^Q \eta + r$. Now, calculating the conditional expectation

$$E^Q\left[N(\tau, (1 - \eta(\tau, X_\tau)\xi) X_\tau) | \mathcal{F}_{\tau^-}\right] = \int_0^1 N(\tau, (1 - \eta(\tau, X_\tau) y) X_\tau) \nu^Q(dy)$$

and assuming that $\mu_t^Q, \psi_t, \gamma_t$, and λ_t are functions of τ and X_τ, we see that the boundary value problem from theorem 1 turns to a new one:

Theorem 2. *In the case of stochastic intensity and loss rates the price of defautable derivative solves the boundary value problem for a PDE*

$$\begin{aligned}
&-\left(r+\left(1+\gamma\left(t,x\right)\right)\lambda\left(t,x\right)\right)A\left(t,x\right)+A_t\left(t,x\right)+\frac{1}{2}\sigma^2x^2A_{xx}\left(t,x\right)\\
&+\left(r+\left(1+\gamma\left(t,x\right)\right)\lambda\left(t,x\right)\eta\left(t,x\right)\int\limits_0^1 y\nu^Q\left(dy\right)\right)xA_x\left(t,x\right)\\
&+\left(1+\gamma\left(t,x\right)\right)\lambda\left(t,x\right)\int\limits_0^1 N\left(t,\left(1-\eta\left(t,x\right)y\right)x\right)\nu^Q\left(dy\right)=0\\
&A\left(T,x\right)=M\left(x\right),
\end{aligned} \tag{13}$$

where μ_t^Q is the drift w.r.t. the risk-neutral measure obtained by the risk-neutral condition (12)*, and ψ_t and γ_t are the transition between the measures terms. It is assumed that they are functions of τ and X_τ.*

Acknowledgements

The author TZ was supported partially by project I 02/19, while OK was partially supported by project DH 02-13 with Bulgarian NSF.

References

[1] Aksamit, A. and Jeanblanc, M. (2017). *Enlargement of Filtration with Finance in View.* Springer Briefs in Quantitative Finance. Springer International Publishing.

[2] Bielecki, T., Jeanblanc, M., and Rutkowski, M. (2009). *Credit Risk Modeling.* Osaka University CSFI Lecture Notes Series. Osaka University Press.

[3] Bielecki, T. and Rutkowski, M. (2014). *Credit Risk: Modeling, Valuation and Hedging.* Springer Finance. Springer Berlin Heidelberg.

[4] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659.

[5] Cooper, I. and Martin, M. (1996). Default risk and derivative products. *Applied Mathematical Finance*, 3(1):53–70.

[6] Dellacherie, C. (1972). *Capacités et processus stochastiques*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag.

[7] Duffie, D. and Singleton, K. (1997). An econometric model of the term structure of interest-rate swap yields. *The Journal of Finance*, 52(4):1287–1321.

[8] Duffie, D. and Singleton, K. (1999). Modeling term structures of defaultable bonds. *The Review of Financial Studies*, 12(4):687–720.

[9] Hull, J. and White, A. (1995). The impact of default risk on the prices of options and other derivative securities. *Journal of Banking & Finance*, 19(2):299–322.

[10] Jarrow, R. and Turnbull, S. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1):53–85.

[11] Mansuy, R. and Yor, M. (2006). *Random Times and Enlargements of Filtrations in a Brownian Setting*, volume 1873 of *Lecture Notes in Mathematics*. Springer.

[12] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144.

[13] Protter, P. (2005). *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.

Approximation of stochastic processes by entire functions of an exponential type and Levitan polynomials

Dmytro Zatula
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
e-mail: dm_zatula@univ.kiev.ua

Abstract

In the following we produce the estimates of distributions of Hölder semi-norms of sample functions of stochastic processes, defined on an infinite interval. Approximations of stochastic processes by entire functions of an exponential type are obtained using these estimates. Also, on the basis of the convergence of Levitan polynomials to entire functions, the approximations of stochastic processes by Levitan polynomials are established.

1 Introduction

Let $(\mathbb{T}, \rho)$ be some metric space. Consider a random process $X = \{X(t), t \in \mathbb{T}\}$ such that the upper limit of the value $\sup\limits_{0<\rho(t,s)\leq\varepsilon} |X(t) - X(s)|$, divided by some function $f(\varepsilon)$, does not exceed 1 as $\varepsilon \to 0$ for all $t, s \in \mathbb{T}$. If it holds then the function f is a modulus of continuity for the process X.

A space of functions with moduli of continuity $f(\varepsilon)$ is the Hölder space and the functional

$$\sup_{\substack{0<\rho(t,s)\leq\varepsilon \\ t,s\in\mathbb{T}}} \frac{|X(t) - X(s)|}{f(\rho(t,s))}$$

is a semi-norm in this space. In the following we produce the estimates of distributions of Hölder semi-norms of sample functions of stochastic processes $X = \{X(t), t \in \mathbb{T}\}$ belonging to $\mathbb{F}_\psi(\Omega)$ spaces.

Such estimates and assumptions under which semi-norms of sample functions of processes from $\mathbb{F}_\psi(\Omega)$ spaces, defined on a compact space,

satisfy the Hölder conditions were obtained in [6]. Moduli of continuity and similar results were provided for Gaussian processes, defined on a compact, by Dudley [2]. Buldygin and Kozachenko [1] generalized Dudley's results to random processes belonging to Orlicz spaces. Kozachenko et al. [5] studied the Lipschitz continuity of generalized sub-Gaussian processes and provided estimates for distributions of Lipschitz norms of such processes. But all these problems were not considered yet for processes, defined on an infinite interval. Only for $L_p(\Omega)$ processes, defined on an infinite interval, estimates for distributions of semi-norms of these processes and assumptions under which semi-norms of sample functions of these processes satisfy Hölder conditions were obtained by Zatula [7].

The Hölder continuity of random processes applies to the study of the rate of approximation of random functions by trigonometric polynomials. In particular, Kamenshchikova and Yanevich [3] investigated an approximation of stochastic processes belonging to spaces $L_p(\Omega)$ by trigonometric sums in the space $L_q[0, 2\pi]$.

2 $\mathbb{F}_\psi(\Omega)$ spaces of random variables

Definition 1. ([4]) Let $\psi(u) > 0$, $u \geq 1$ be some increasing function such that $\psi(u) \to \infty$ as $u \to \infty$. A random variable ξ belongs to $\mathbb{F}_\psi(\Omega)$ space if

$$\sup_{u \geq 1} \frac{(\mathsf{E}|\xi|^u)^{1/u}}{\psi(u)} < \infty.$$

Properties of random variables and processes from $\mathbb{F}_\psi(\Omega)$ spaces were considered in detail in [4]. Henceforth we will consider the spaces $\mathbb{F}_\psi(\Omega)$, which have the following property.

Definition 2. ([6]) Let ξ_1, ..., ξ_n be random variables belonging to the space $\mathbb{F}_\psi(\Omega)$. $\mathbb{F}_\psi(\Omega)$ space has property Z if there are monotone non-decreasing function $z(x) > 0$, monotone increasing function $U(n)$ and a real number $x_0 > 0$ such that for any sequence of random variables $(\xi_k,\ k = \overline{1,n})$ from $\mathbb{F}_\psi(\Omega)$ space, $\forall x > x_0$ and for all $n \geq 2$ the following inequality is performed:

$$\mathsf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| > x \cdot \max_{1 \leq k \leq n} \|\xi_k\|_\psi \cdot U(n)\right\} \leq \frac{1}{n} \exp\{-z(x)\}.$$

3 Approximation of stochastic processes belonging to $\mathbb{F}_\psi(\Omega)$ spaces

Theorem 1. *Consider a separable random process* $X = \{X(t), t \in [0,\infty)\}$ *belonging to* $\mathbb{F}_\psi(\Omega)$ *space, which has the property* Z *with functions* $U(n)$ *and*
$z(x)$ *for* $x_0 > 0$. *Let* $[0,\infty) = \bigcup\limits_{i=0}^{\infty} A_i$, *where* $A_i = [a_i, a_{i+1}]$, $\{a_i, i = 0,1,...,\infty\}$ *is some increasing sequence,* $a_0 = 0$. *Denote* $\alpha_i = a_{i+1} - a_i$, $D_i = [a_i, a_{i+1}+\theta]$, $\theta \in \left(0, \min\limits_{i\geq 0}\alpha_i\right)$. *Assume that there are monotonically increasing continuous functions* $\sigma_i = \{\sigma_i(h),\ h \geq 0\}$ *such that* $\sigma_i(0) = 0$, $i = 0,1,...$ *and* $\forall i = 0,1,...:$

$$\sup_{\substack{\rho(t,s)\leq h\\ t,s\in D_i}} \|X(t) - X(s)\|_\psi \leq \sigma_i(h), \qquad 0 < h < \alpha_i + \theta.$$

Let $N_i(\varepsilon)$ *be metric massivenesses of intervals* D_i, $i = 0,1,...$ *with respect to the metric* $\rho(t,s) = |t-s|$, $t,s \in [0,\infty)$. *Also let* $\varepsilon_0 = \min\limits_{i\geq 0}\left\{\sigma_i^{(-1)}(\alpha_i + \theta)\right\}$, *where* $\sigma_i^{(-1)}(h)$ *is an inverse function to a function* $\sigma_i(h)$, $i = 0,1,...,$ *and* $\forall i = 0,1,...:$

$$g_{B,i}(\varepsilon) = \int\limits_0^{\sigma_i(\varepsilon)} U(B^2 N_i^2(\sigma_i^{(-1)}(t)))dt < \infty; f_{B,i}(\varepsilon) = \int\limits_0^{\sigma_i(\varepsilon)} U(BN_i(\sigma_i^{(-1)}(t)))dt.$$

Denote $w_{B,i}(t,s) = (6+4\sqrt{2})f_{B,i}(|t-s|)+(5+2\sqrt{6})g_{B,i}(|t-s|)$, $t,s \in D_i$, *and* $w_B(t,s)$ *is such a function that* $w_B(t,s) = \{w_{B,i}(t,s)\,|\,t,s \in A_i$ *or*
$\min\{t,s\} \in A_i$, $\max\{t,s\} \in A_{i+1}\}$.

Then for all $x > x_0$, $\theta \in \left(0, \min\limits_{i\geq 0}\alpha_i\right)$ *and* $\varepsilon \in (0, \min\{\varepsilon_0, \theta\})$ *under the condition that* $\sum\limits_{i=0}^{\infty}\frac{1}{\alpha_i} < \infty$ *the following inequality holds true:*

$$\mathsf{P}\left\{\sup_{\substack{0<|t-s|\leq\varepsilon\\ t,s\in[0,\infty)}} \frac{|X(t)-X(s)|}{w_B(t,s)} > x\right\} \leq \frac{4\varepsilon B(2B+1)}{B^2-1}\cdot\exp\{-z(x)\}\cdot\sum_{i=0}^{\infty}\frac{1}{\alpha_i+\varepsilon}.$$

Theorem 2. *Let all the assumptions of Theorem* 1 *be fulfilled and the process* $\{X(t),\, t \in [0,\infty)\}$ *is uniformly continuous on the interval* $[0,\infty)$. *Assume that the function* $w_B(t,s)$ *is such that the following inequality holds*
$\forall \varepsilon \in (0, \min\{\varepsilon_0, \theta\})$*:*

$$\sup_{\substack{0<|t-s|\le\varepsilon \\ t,s\in[0,\infty)}} |X(t)-X(s)| \le w_{B,\varepsilon} \cdot \sup_{\substack{0<|t-s|\le\varepsilon \\ t,s\in[0,\infty)}} \frac{|X(t)-X(s)|}{w_B(t,s)},$$

where $w_{B,\varepsilon} = \max\limits_{\substack{0<|t-s|\le\varepsilon \\ t,s\in[0,\infty)}} w_B(t,s)$. *Then for* $x > 3x_0 \cdot w_{B,\frac{1}{\gamma}}$, $\gamma \in \left(0, \frac{1}{\min\{\varepsilon_0,\theta\}}\right)$ *and under the condition that* $\sum\limits_{i=0}^{\infty} \frac{1}{\alpha_i} < \infty$ *the following holds*

$$\mathsf{P}\left\{ \inf_{F\in\{f \mid f(t)-X(t)\in\mathbb{F}_\psi(\Omega)\}} \sup_{t\in[0,\infty)} \|X(t)-F(t)\|_\psi > x \right\} \le$$

$$\le \frac{4B(2B+1)}{\gamma(B^2-1)} \cdot \exp\left\{ -z\left(\frac{x}{3w_{B,\frac{1}{\gamma}}}\right)\right\} \cdot \sum_{i=0}^{\infty} \frac{1}{\alpha_i + \frac{1}{\gamma}},$$

where $f(t)$ *are entire functions of an exponential type* $\le \gamma$ $(\gamma > 0)$.

References

[1] Buldygin V.V., Kozachenko Yu.V. (2000). *Metric Characterization of Random Variables and Random Processes*. American Mathematical Society, Providence, Rhode Island.

[2] Dudley R.M. (1973). Sample functions of the Gaussian processes. *Annal. of Probability*. Vol. **1**, no. 1, pp. 3-68.

[3] Kamenshchikova O.E., Yanevich T.O. (2011). An approximation of $L_p(\Omega)$ processes. *Theor. Probab. & Math. Statist.* Vol. **83**, pp. 71-82.

[4] Kozachenko Yu.V., Mlavets' Yu.Yu. (2013). The Banach spaces $\mathbb{F}_\psi(\Omega)$ of random variables. *Theor. Probab. & Math. Statist.* Vol. **86**, pp. 105-121.

[5] Kozachenko Yu., Sottinen T., Vasylyk O. (2011). Lipschitz conditions for $\mathrm{Sub}_{\varphi}(\Omega)$-processes and applications to weakly self-similar processes with stationary increments. *Theor. Probab. & Math. Statist.* Vol. **82**, pp. 57-73.

[6] Kozachenko Yu.V., Zatula D.V. (2015). Lipschitz conditions for stochastic processes in the Banach spaces $\mathbb{F}_{\psi}(\Omega)$ of random variables. *Theor. Probab. & Math. Statist.* Vol. **91**, pp. 43-60.

[7] Zatula D. (2015). Estimates for the distribution of semi-norms of $L_p(\Omega)$ processes in Hölder spaces. *Journal of Applied Mathematics and Statistics*. Vol. **2**, no. 1, pp. 9-20.

ABSTRACTS

Generating representative subsample

Dimitar Atanasov
New Bulgarian University
Sofia, Bulgaria
e-mail: datanasov@nbu.bg

Empirical analyses usually estimate the effect of some variables, considered as independent, on one (or more), considered as dependend. It is rare situation to have the whole population available for the analysis. Then, usually a subsample is chosen and the conclusions are generalized to the whole population. This approach is based on the concept that the sample represents, in some meaning the population. If a sample is large enough and the selection process is random, usually the sample is considered to be representative one. However in some cases the sample should satisfy some restrictions on selected characteristics, which are previously known for the population. Here we propose an approach to this problem as a solution of a task of integer linear programming.

Transience and recurrence of Markov processes with constrained local time

Adam Barker
Department of Mathematics and Statistics,
University of Reading, UK
email: adam.barker@pgr.reading.ac.uk

We study the problem of a Markov process conditioned so that its local time must grow slower than a prescribed function. Building upon recent work on Brownian motion with constrained local time in [1], we study the problem for a large class of Markov processes.

We find a necessary and sufficient condition for transience/recurrence of the conditioned process, and also explicitly determine the distribution of the conditioned (inverse) local time. In the transient case, we explicitly determine the law of the conditioned Markov process. In the recurrent case, we further determine the "entropic repulsion envelope", which formally characterises how the process is affected by "entropic forces" (the tendency of a system to increase its entropy).

This work is theoretical, but is related to problems in polymer physics in which a long polymer chain is modelled by a random process which is in some sense "weakly self-avoiding" (constraining the local time corresponds to "weak" self-avoidance). The talk provides a brief overview of some relevant random polymer models.

References

[1] Kolb M., Savov M. (2016). Transience and recurrence of a Brownian path with limited local time. *Ann. Probab.*, Vol. **44**, no. 6, pp. 4083–4132.

Selective inference on a tree of hypotheses: new error rates and controlling strategies

Marina Bogomolov[1], Christine B. Peterson,
Yoav Benjamini, Chiara Sabatti
Technion–Israel Institute of Technology
Haifa, Israel
e-mail: [1]marinabo@technion.ac.il

In modern statistical challenges we are often presented with a set of families of hypotheses, which are organized hierarchically in a tree structure. Each family is selected and tested only if all its ancestor hypotheses are rejected. We address the situation where the p-values for parent hypotheses are dependent on the p-values for the hypotheses within the families they index. We formulate a general class of error rates addressing selective inference on families which are organized hierarchically in a tree structure and propose a hierarchical testing procedure with a guaranteed control of such error rates.

Statistical analysis of vibrodiagnostics data via time-varying dynamical models

Marek Brabec
Institute of Computer Science,
The Czech Academy of Sciences,
Pod Vodárenskou vezi 2, 182 07
Praha 8, Czech Republic
e-mail: mbrabec@cs.cas.cz

In this talk, we will present a semiparametric modeling framework for detailed analysis of blade-tip-timing (BTT) data from vibrodiagnostic measurements. The data come as times when individual blades of industrial turbine pass one or more sensors mounted on the turbine coating. Our model is formulated as a time-varying extension of standard (spectral-based) approach build on a generalized additive model (GAM) with penalized splines and autocorrelated process in the residuals. Penalized components are used to test whether the coefficients of important trigonometric terms are constant or not. If they are not, their trajectory is extracted and time-varying amplitudes and phases are derived. We explore several strategies for modeling the dynamic process in the residuals. While traditional discrete-time (ARMA) formulation is possible, continuous-time approximation of diffusion type can be advantageous computationally. We use quasilikelihood approach in high-frequency asymptotics settings to estimate diffusion sub-model parameters and to compare them across various turbine working conditions. This work is supported from the Strategy AV21 of the Czech Academy of Sciences.

Configuration graphs with bounded number of edges

I. Cheplyukova[1], Yu. Pavlov[2]
[1,2]*Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences*
Petrozavodsk, Russia
e-mail: [1]chia@krc.karelia.ru, [2]pavlov@krc.karelia.ru

We consider configuration random graphs with N vertices. The degrees of vertices are independent identically distributed random variables. Each vertex degree is equal to the number of numbered semiedges of this vertex. If the sum of vertex degrees is odd we introduce an auxiliary vertex of degree one. Graph is constructed by joining all semiedges pairwise equiprobably to form edges. Let the random variables ξ which is equal to degree of any vertex in the graph has unknown distribution. Assume that we know only the limit behaviour of the tail of this distribution as $k \to \infty$:

$$\mathbf{P}\{\xi = k\} = \frac{d}{k^g (\ln k)^h},$$

where $d > 0, g \geq 1, k \geq 0, g + h > 1$.

We consider two types of configuration graphs. One of them is a subset of graphs where the number of edges is equal to n and in other subset the number of edges at most n. We obtained the limit distributions of the maximum vertex degree and the number of vertices with a given degree in these conditional configuration graphs under various types of behaviour N and n tending to infinity.

Acknowledgements

The study was partially supported by the Russian Foundation for Basic Research, grant 16-01-00005.

Reconstruction of the modified Langevin equation from p-order persistent time series

Zbigniew Czechowski
Institute of Geophysics, Polish Academy of Sciences
Warsaw, Poland
e-mail: zczech@igf.edu.pl

Modeling lies at the heart of time series analysis. The aim of the modeling is to understand the underlying mechanism that generates the observed data and to forecast future behavior. Non-regularity of many natural processes induces us to treat the registered data as realizations of stochastic processes and requires using nonlinear models. Unfortunately, nonlinear methods are much less understood than classical methods for linear cases (e.g., ARMA models). The Langevin equation introduces nonlinearity in drift and diffusion terms and leads to a wide class of distributions; from Gaussian to inverse-power. Moreover, following the correspondence of the Langevin and the Fokker-Planck equation, the procedure of reconstruction of the Langevin equation was introduced (Siegert et al. 1998, Friedrich et al. 2011). Both the models, linear ARMA and nonlinear Langevin, have their merits and demerits. ARMA models can describe Markov time series of order m, but the linearity is limiting their usefulness. On the other hand, the Langevin model is nonlinear, but it can describe only Markov processes (of order 1).
In this work we introduce the generalized discrete Langevin equation for some class of non-Markov processes, namely for persistent time series of order p. For persistent processes non-local effects must be considered. We assume that the next state of the process is dependent not only on the present state but also on signs of p previous jumps. To this aim, the standard discrete Langevin equation is modified by introducing a new random function which determines the sign of the diffusion term. The standard procedure (Siegert et al.. 1998, 2011) of reconstruction of the Langevin equation from time series leads to the proper estimation of the diffusion function but to the wrong reconstruction of the drift function in the case of the modified equation. To estimate the deviation in the drift we propose a new reconstruction procedure which was tested on many time series generated by the modified Langevin equation with different

drift and diffusion functions and different parameters of persistence. This work is a significant extension of our previous approach (Czechowski 2016) in which persistent processes of order p = 1 were taken into account. The generalization opens a wide possibilities of nonlinear modeling od data in which persistence and antipersistence of different orders can be mixed in a time series under investigation.

Acknowledgements

This work has been financed by the project of the National Science Centre (contract No. 2016/21/B/ST10/02998).

References

[1] Czechowski Z. (2016). Reconstruction of the modified discrete Langevin equation from persistent time series. *Chaos*. Vol. **26**, pp. 053109-1-053109-8.

[2] Friedrich R., Peinke J., Sahimi M., Reza Rahimi Tabar M. (2011). Approaching complexity by stochastic methods: From biological systems to turbulence. *Physics Reports* . Vol. **506**, pp. 87-162.

[3] Siegert S., R. Friedrich, and J. Peinke. (1998). Analysis of data sets of stochastic systems. *Phys. Lett. A*. Vol. **243**, pp. 275-280.

Finite source retrial queue with server vacations

Velika Dragieva
University of Forestry, Sofia, Bulgaria
e-mail: dragievav@yahoo.com

The paper considers a queueing system with one server that serves a finite number of customers. The customer, arriving at a moment of a busy server joins a virtual waiting room called orbit and after an exponentially distributed time interval repeats his/her attempt for service. After each service the server with a certain probability goes on vacation, whose duration follows an arbitrary probability distribution, different from the distribution of the service times. Applying the discrete transformations method we derive formulas for the joint steady state distribution of the server state and the orbit size, and the main macro characteristics of the system performance. Numerical examples illustrating the dependence of these characteristics on the system input parameters are presented.

On multivatiate modifications of Cramer-Lundberg risk model

Pavlina K. Jordanova[1], Milan Stehlík[2,3]
[1] *Shumen University, Shumen, Bulgaria*
[2] *Department of Applied Statistics and Linz Institute of Technology, Johannes Kepler University in Linz, Austria*
[3] *Department of Statistics, University of Valparaiso, Chile*
e-mail: [1]pavlina_kj@abv.bg, [2,3]milan.stehlik@jku.at

The paper considers very general multivariate modifications of Cramer-Lundberg risk model. The claims can be of different types and can arrive in groups. The groups arrival processes within a type have constant intensities. The counting groups processes are dependent multivariate compound Poisson processes of type I. We allow empty groups and show that in that case we can find stochastically equivalent Cramer-Lundberg model with non-empty groups.

The investigated model generalizes the risk model with common shocks, the Poisson risk process of order k, the Poisson negative binomial, the Polya-Aeppli of order k among others. All of them with one or more types of polices.

The relations between the numerical characteristics and distributions of the components of the risk processes are proven to be corollaries of the corresponding formulae of the Cramer-Lundberg risk model.

Generalized expected discounted penalty function for a class of two-sided jumps Lévy risk processes

Ekaterina Kolkovska
Centro de Investigacion en Matematicas
Guanajuato, Mexico
e-mail: todorova@cimat.mx

We give an explicit expression for the Generalized expected discounted penalty function corresponding to a class of two-sided jumps Lévy processes whose positive jumps have a rational Laplace transform. The resulting expression is given in terms of known functions depending only on the parameters of the process.

Weibull accelerated failure time frailty models and extensions

Adams Kusi-Appiah[1], Gleb Haynatzki[2]
[1,2] *University of Nebraska*
Omaha, Nebraska, USA
e-mail: [2]ghaynatzki@unmc.edu

Clustered or grouped survival data frequently occur in longitudinal studies (clinical trials and cohort studies). Data within clusters may be more strongly correlated than between clusters. The natural approach is to incorporate random effects (or shared frailties) to account for within-cluster homogeneity in outcomes. Different probability distributions have been proposed for the frailties, most popular being the Gamma distribution, which leads to closed-form representation of the survival and hazard functions. Others are the inverse Gaussian, lognormal, and positive stable distributions. Also, frailty models have been considered in the Cox proportional hazards (PH) semiparametric regression framework. The Weibull (parametric) regression can be considered as an attractive alternative to the Cox PH semiparametric regression in modeling survival data. Weibull regression may be expressed as PH model or accelerated failure time model. While for some frailty distributions, e.g., Gamma, one can obtain tractable integration, however, no closed-form expressions for the marginal survival and hazard function can be derived for some other frailty distributions, e.g., normal. We will review some of the main frailty distributions under the Weibull accelerated failure time regression model. We will discuss some of the existing estimation techniques, as well the analytical and computational challenges of these methods. Illustrations with real-life data will be explored.

Bivariate noncentral Pólya-Aeppli disribution

Meglena D. Lazarova[1], Leda D. Minkova[2]
[1]*Faculty of Applied Mathematics and Informatics*
Technical University of Sofia, Sofia, Bulgaria
[2]*Faculty of Mathematics and Informatics*
Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
e-mail: [1]meglena.laz@tu-sofia.bg, [2]leda@fmi.uni-sofia.bg

In this paper we consider a random variable which is a sum of Poisson distributed and Pólya-Aeppli distributed variables. We called the resulting distribution a Noncentral Pólya-Aeppli distribution. Then by the trivariate reduction method we introduce a bivariate Noncentral Pólya-Aeppli distribution. For the univariate and the bivariate distribution we give some essential properties and derive the probability mass function, recursion formulas, correlation structure and moments. We also obtain a moment estimation of the parameters and illustrate a numerical example.

References

[1] Minkova L.D. (2002). A generalization of the classical discrete distributions, *Communications in Statistics-Theory and Methods,***31**, 871–888.

[2] Ong S.H. and Lee P.A. (1979). The non-central negative binomial distribution, *Biom. Journal,* **21**, 611–628.

Explosion of a reaction-diffusion equation perturbed by a fractional Brownian motion

José Alfredo López-Mimbela
Centro de Investigación en Matemáticas
Guanajuato, Mexico
e-mail: jalfredo@cimat.mx

We provide conditions implying finite-time blowup of positive weak solutions to the SPDE

$$\begin{aligned} du(t,x) &= \left[\Delta_\alpha u(t,x) + Ku(t,x) + u^{1+\beta}(t,x)\right] dt + \mu u(t,x)\, dB_t^H, \\ u(0,x) &= f(x),\ x \in \mathbb{R}^d,\ t \geq 0, \end{aligned} \tag{1}$$

where $\alpha \in (0,2]$, $K \in \mathbb{R}$, $\beta > 0$, $\mu \geq 0$ and $H \in [\frac{1}{2}, 1)$ are constants, Δ_α is the fractional power $-(-\Delta)^{\alpha/2}$ of the Laplacian, (B_t^H) is a fractional Brownian motion with Hurst parameter H, and $f \gneq 0$ is a bounded measurable function. To achieve this we investigate the a.s. finiteness of exponential functionals of the form

$$\int_{r_0}^{\infty} \left[\frac{e^{Ks+\mu B_s^H}}{s^{d/\alpha}}\right]^{\beta} ds$$

with $r_0 > 0$. Moreover, we obtain bounds for the explosion times of the equation above in the cases $K < 0$ and $K > 0$. Setting $K = \mu = 0$ in (1) we recover the classical deterministic result proved by Sugitani (1975) that, if $d \leq \alpha/\beta$, then the equation $\frac{\partial}{\partial t}u(t,x) = \Delta_\alpha u(t,x) + u^{1+\beta}(t,x)$ in $\mathbb{R}^d$ possesses no non-trivial positive global solutions.

Consecutive subordination of Poisson processes and gamma processes

Penka Mayster
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia,Bulgaria
e-mail: penka.mayster@math.bas.bg

Subordination by Bochner is a randomization of the time parameter in stochastic processes. The time process is supposed to be a subordinator - Lévy process with non decreasing sample paths. The subordinated process is defined as $Y(t) = X(T(t))$. We consider the sequences of consecutive subordination of Poisson processes and Gamma processes. A Poisson process subordinated by a Gamma process is a Negative Binomial process. The Negative Binomial process subordinated by a Poisson process is a Pólya-Aeppli process. A Gamma process subordinated by a Poisson process is a Tweedie process - Poisson exponential Lévy process. We prove that the Tweedie process is stable by iteration. The Pólya-Aeppli process is stable by subordination with a Tweedie process. The transition probabilities of subordinated processes are expressed by special functions - modified Bessel function and confluent hypergeometric function. The asymptotic behavior is considered.

The problem

How does this additional randomness will be accumulated after several consecutive subordinations of Poisson processes and Gamma processes.

References

[1] Mayster P. I.(2018). Cosecutive subordination of Poison processes and Gamma processes . *Compt. Rend. Acad. Bulg. Sci.*. Vol. **71**, no. 5, pp. 603-611.

A point process characterisation of extreme temperatures: an application to South African data

Murendeni Maurel Nemukula[1], Caston Sigauke[2]
[1] *University of Limpopo, Polokwane, Limpopo Province, South Africa*
[2] *University of Venda, Thohoyandou, South Africa*
e-mail: [1]murendeni.nemukula@gmail.com

The point process modelling approach is considered a more elegant alternative of extreme value analysis. This is because of its capability in modelling both the frequency and intensity rates of the occurrence of extremes. In this paper we demonstrate the use of the point process modelling approach in which stationary and non-stationary models are used in modelling Average Maximum Daily Temperature (AMDT) in South Africa. A penalized cubic smoothing spline function is used for non-linear detrending of the data and determining a fixed threshold above which excesses are extracted and used. An extremal mixture model is then fitted to determine a threshold in which a boundary corrected kernel density is fitted to the bulk model and a generalized Pareto distribution fitted to the tail of the distribution. These data exhibits properties of short-range dependence and strong seasonality, leading to declustering. An interval estimator method is used to decluster data for the purpose of fitting point process models to cluster maxima. The models that are used in this paper are nested and likelihood ratio tests are conducted using the deviance statistic. The tests support the fit of the stationary point process model. We further fitted the stationary GPD and used the formal tests which are the Cramér-von Mises test and the Anderson-Darling test to diagnose fit. These tests and the diagnostic plots support fit of the stationary GPD to cluster maxima. The stationary point process model was used with the reparameterization approach to determine frequency of the occurrence of extremely hot days, which are found to be 15 times per year. The modelling framework and results of this paper are important to Eskom, South Africa's power utility company. This is based on the fact that it is during the non-winter period that they plan for the maintenance of the power plants.

Keywords: extreme value theory, point process, temperature.

Parameter estimation for Normal-Inverse Gaussian distributions

Metodi Nikolov
Faculty of Mathematics and Informatics, Sofia University
Sofia, Bulgaria
e-mail: metodin@fmi.uni-sofia.bg

The Normal-Inverse Gaussian distribution was first described by Barndorff-Nielsen as Normal variance mixture with an Inverse Gaussian mixing distribution. In this talk we will discuss different parametrizations of the distribution and the estimation of its parameters in various settings: maximum likelihood, method of moments, Bayesian. Some new algorithms will be presented.

References

[1] Ole E. Barndorff-Nielsen. Normal Inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, Mar 1997.

[2] Ole E Barndorff-Nielsen and S Levendorskii. Feller processes of normal inverse Gaussian type. *Quantitative Finance*, 1(3):318–331, 2001.

[3] Andrew Gelman. *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)*, volume 1. Chapman and Hall/CRC, Nov 2013.

[4] Matthew D. Homan and Andrew Gelman. The No-U-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, January 2014.

[5] Dimitris Karlis. An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters*, 57(1):43–52, Mar 2002.

[6] Bianchi Fabozzi Ratchev, Kim. Financial models with L'evy processes and volatility clustering, Jun 2010.

Some properties of Lee distance in two-sample location problem

Nikolay I. Nikolov
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: n.nikolov@math.bas.bg

There are various applications of metrics on permutations in nonparametric statistics (see, e.g.,[2, 3]). In this study, Critchlow's unified approach in [1] is applied to Lee distance in the case of two-sample location problem. Some properties of the obtained rank statistic are studied under the null hypothesis. The joint distribution of the statistics based on Hamming distance and Lee distance is given recursively. Its asymptotic behavior is studied for large sample sizes. The power and robustness of the considered test statistic are compared to those of the most popular rank tests for two-sample problems.

Acknowledgements
The research was partially supported by the National Science Fund of Bulgaria under Grant I02/19 and by the Support Program of Bulgarian Academy of Sciences for Young Researchers under Grant 17-95/2017.

References

[1] Critchlow D. E. (1986). A Unified Approach to Constructing Nonparametric Tests, Tech. Report. No. **86-15**, Dept. of Statistics, Purdue University, Indiana.

[2] Diaconis P. (1988). *Group Representations in Probability and Statistics*. IMS Lecture Notes - Monograph Series, Vol. **11**, Hayward, Carifornia.

[3] Marden J. I. (1995). *Analyzing and Modeling Rank Data*. Monographs on Statistics and Applied Probability, No. **64**, Chapman & Hall.

Fractional diffusion of variable order and anomalous aggregation phenomenon

Mladen Savov
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: mladensavov@math.bas.bg

In this talk we consider fractional diffusions of variable order which may be thought of as a particle moving in diverse porous milieu. Linking them to general semi-Markov processes we discuss the long-term behaviour of these diffusions. In particular depending on the variable order of the diffusion, we prove rigorously an anticipated anomalous behaviour, that is the domination of the time the fractional diffusion spends in regions where the order of the diffusion is minimal over the time spent elsewhere. Under some conditions and depending on the variable order of the diffusion we also demonstrate that the probability to find the particle in regions where the order of the diffusion is minimal converges to one as time increases. The main techniques are classical laws of the iterated logarithm for general Levy processes. We discuss some open problems and technical difficulties.

Joint work with Bruno Toaldo (Naples)

Acknowledgements
The research was partially supported by the National Science Fund of Bulgaria under Grant I02/19.

Modelling of cancer evolution with multi-type age-dependent branching processes

Maroussia Slavtchova-Bojkova[1,2] , Kaloyan Vitanov[1]
[1] *Sofia University, Faculty of Mathematics and Informatics,*
[2] *Institute of Mathematics and Informatics, BAS*
Sofia, Bulgaria
e-mail: [1]bojkova@fmi.uni-sofia.bg

In this talk a novel multi-type decomposable branching process model of a cell population, with n types, $n > 2$, is considered ([2]). We are exploring two important consequences arising from the biological nature of cancer: first, the metastatic process could involve more than one human organ and second, if transported into a completely different environment (organ) a cancerous cell may change its lifespan and division characteristics. The presented work is taking into account these facts in cancer development and is generalising the results in [1] considering the presence of mutation probabilities among $n > 2$ cell types as well as heavy-tailed (and other) distributions as models of cancer cell lifespan.

Acknowledgements The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and partially supported by and Ministerio de Economia y Competitividad and the FEDER through the Plan Nacional de Investigacion Cientifica, Desarrollo e Innovacion Tecnologica, grant MTM2015-70522-P, the Junta de Extremadura (grant GR15105), Spain.

References

[1] Slavtchova-Bojkova, M., Trayanov, P., Dimitrov, S. (2017). Branching processes in continuous time as models of mutations: Computational approaches and algorithms. Computational Statistics and Data Analysis, http://dx.doi.org/10.1016/j.csda.2016.12.013l.
[2] Vitanov, K. N., Slavtchova-Bojkova M. (2018). Multitype branching processes in continuous time as models of cancer. Ann. Sofia Univ., Fac. Math. and Inf., 104, 2017, 221–228.

Diagnostic for Monte Carlo simulations of multitype branching processes with power series offspring distributions

Ana Staneva
Sofia University "St. Kliment Ohridski"
Sofia, Bulgaria
e-mail: anastaneva@gmail.com

In our study we consider the parametric estimation of the multitype discrete-time branching processes with individual distribution from the class of the power series offspring distributions. The problem of obtaining an approximation of the posterior distribution is studied via the Bayesian approach with Monte Carlo simulations. We make conclusions about the convergence of the MCMC chains as well as statistical inference on the basis of the output data set of the MCMC simulations. We consider several examples with computational results and simulations in the software environment R.

On the properties of a class of nonparametric tests based on the number of exceedances

Eugenia Stoimenova
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: jeni@math.bas.bg

Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ be random samples from continuous distribution functions F and G, respectively. Denote the ordered X's and Y's by $X_{(1)} < \cdots < X_{(m)}$, and $Y_{(1)} < \cdots < Y_{(n)}$, respectively. For $0 \leq r < n$, define the exceedance statistics based on thresholds from the both samples.

$$\begin{array}{lllll} A_s = & \text{the number of} & Y\text{-observations} & \text{larger than} & X_{(m-s)}, \\ B_r = & \text{the number of} & X\text{-observations} & \text{smaller than} & Y_{(1+r)}. \end{array} \quad (1)$$

These statistics are potentially useful for testing whether the two random samples are from the same population. For example classic precedence test [1] is based on the number of observations in the X-sample that are smaller than $Y_{(r)}$. Large values of this statistic lead to rejection of the null hypothesis about equality of the two distributions. Precedence tests are useful in life-testing experiments where in the data become available naturally in order of size. The experiment is terminated after a certain number of failures.

In the general situation of testing equality of two distribution, the precedences and exceedances are both necessary with respect to thresholds from both samples. The number of exceedances in the Y-sample with respect to a threshold from the X-sample could be used along with the number of precedences in the X-sample with respect to a threshold from the Y-sample. Several rank tests are proposed using the number of exceedances. We discuss here distribution properties of such type tests statistics and consistency of the tests against specific alternatives.

Acknowledgements

The research was partially supported by the National Science Fund of Bulgaria under Grant DH 02-13.

References

[1] Stoimenova, E., & Balakrishnan, N. (2017). Sidak-type tests for the two-sample problem based on precedence and exceedance statistics. *Statistics*, Vol. **51**, 247–264.

Statistical inferences for parameters of multitype branching processes with power series offspring distributions

Vessela Stoimenova[1], Ana Staneva[2], Dimitar Atanasov[3]
[1,2]*Sofia University "St. Kliment Ohridski", Sofia, Bulgaria*
[3]*New Bulgarian University, Sofia, Bulgaria*
e-mail: [1]stoimenova@fmi.uni-sofia.bg, [3]datanasov@nbu.bg

Branching processes have numerous applications in different scientific and practical areas, many of them involving multitype modeling. Statistical estimation of the process' characteristics is an important issue in their study. The present work considers some aspects of the simulation, visualization and statistical estimation of the parameters of multitype discrete time branching processes with power series offspring distributions. We consider numerical results for simulated data samples of different types - samples over family trees and over generation sizes - and compare estimators.

Acknowledgements

The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and partially supported by the financial funds allocated to the Sofia University "St. Kl. Ohridski", grant No 80-10-222/2018.

Modelling Electromagnetic showers with branching processes

Assen Tchorbadjieff[1], Penka Mayster[2]
[1,2]*Institute of Mathematics and Informatics,*
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: [1]assen@math.bas.bg, [2]penka.mayster@math.bas.bg

We consider a system in which there are many particles moving trough trajectories induced by random in time and location interactions of particles with either homogeneous or changing medium. Their number and position in space $R^3 \otimes R^3 \otimes R_+$ may be described only by branching processes with trajectory of the ancestral line of any particle and prescribed by some random measure. For more comprehensive solution we selected a model derived from electromagnetic cascades expansion through the matter, mainly resembling cosmic rays cascades in atmosphere. Following the most properties of the process we defined analytical solution and compared its results with specially designed stochastic software simulations. The number of particles by generations constitutes a multi-type Galton-Watson process.

Poisson random measures and Sevastyanov branching processes

Nikolay M. Yanev
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: yanev@math.bas.bg

The research continues the investigations started in [1] on critical Sevastyanov branching processes with non-homogeneous Poisson immigration. Now the local intensity of the Poisson random measure is a regularly varying function and some new asymptotic results are obtained.

Acknowledgements. The paper is supported by the NFSR of the Min. Edu. Sci. of Bulgaria, grant No. DFNI-I02/17.

References

[1] K. V. Mitov, N. M. Yanev.; Non-homogeneous Poisson Immigration. Proceedings of the Steklov Institute of Mathematics, Vol. 282, 2013, 181-194.

Statistical modeling of the process syneresis of the production of yogurt with water extract of Rosa canina

Maya Zhelyazkova
Department of Mathematics and Informatics, Sofia University
Sofia, Bulgaria
e-mail: zhelyazkova@fmi.uni-sofia.bg

The role of antioxidants on the health of the human body begins to be explored in depth in the second half of the 20-th century with the discovery of the so-called free radicals. In this paper we explore the effect of Rosa canina as a natural antioxidant. The aim of the study is to find the dependency between percent water extract of Rosa canina and the amount of separated whey in preparation of yogurt. The process of separating the whey is crucial for good quality yogurt-texture, acidity, formation of the flavor of the milk. The results of statistical analysis show that 5 % water extract of the fruits of Rosa canina does not affect significantly the amount of separated whey in the process syneresis.

References

[1] Pagano,M., Gauvreau,K. (1993). *Principles of Biostatistics*. Duxbury Press, Belmont, California

[2] Maya Zhelyazkova, Ira Taneva (2016). Statistical modeling of the process syneresis of the production of yogurt with water extract of Rosa canina. *Journal of Pharmacognosy and Phytochemistry*. **5(2)**, pp. 204-206