UNIVERSITY OF THE AEGEAN

KATHOLIEKE UNIVERSITEIT LEUVEN

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

UNIVERSITY OF COPENHAGEN

5th Conference in Actuarial Science & Finance on Samos, September 4-7, 2008

5th Conference in Actuarial Science and Finance on Samos

Proceedings

Samos, March 31, 2009

PREFACE

I am very glad to open this volume with the proceedings of the last 5th Samos Conference. This time the proceedings appear in a separate volume, a fact that represents a new step to the success of the series of Samos Meetings. These proceedings reflect a small sample of the insightful discussions and diversified speculations, developed during the thirty four presentations by participants of the conference and attendants of a short course, given before the Samos Conference on "Subexponential Tails in the World of Dependence" by Qihe Tang.

I have the pleasure to report that during the closing ceremony three prizes for excellence were bestowed to the following presentations:

1. Optimal Investment Strategy to Minimize the Ruin Probability of an Insurance Company under Borrowing Constraints by Nora Mler.
2. Longevity Risk in Portfolios of Life Insurance and Annuity Liabilities: the Effect of Product Design, Product Mix and Portfolio Composition by Ralph Stevens.
3. Optimal 'per claim' reinsurance for dependent risks by Manuel Guerra.

I must express my thanks to all the factors involved in the preparation of the proceedings of the conference, starting from the authors and scientific committee. I would like to mention the significant help by Georgios Psarrakos for his work of formatting the material. I am also extremely grateful to our sponsors: University of the Aegean, Commercial Value, Samos Steamship Co., Bank of Greece Ministry of Education, Prefectural of Samos and Municipality of Karlovassi.

I take this opportunity to invite you to take part to the 6th Samos Conference, which is to be held on June 3-6, 2010 (website: http://www.actuar.aegean.gr/Samos2010/).

Karlovassi, March 28, 2009.

Dimitrios G. Konstantinides

Chairman of the Organizing Committee

SCIENTIFIC COMMITTEE

Albrecher Hansjoerg – University of Lausanne, Switzerland

Asmussen Soeren – University of Aarhus, Denmark

Drees Holger - University of Hamburg, Germany

Foss Serguei – Heriot - Watt University, United Kingdom

Cheliotis Dimitris – EURANDOM, the Netherlands

Gerber Hans – University of Lausanne, Switzerland

Goovaerts Marc – Katholieke Universiteit Leuven, Belgium, and University of Amsterdam,The Netherlands

Konstantinides Dimitrios – University of the Aegean, Greece

Makov Udi – University of Haifa, Israel

Marceau Etienne – University Laval, Canada

Mikosch Thomas – University of Copenhagen, Denmark

Morettin Pedro – University of Sao Paolo, Brazil

Ng Kai Wang – University of Hong Kong, Hong Kong

Papaioannou Takis (chair) – University of Piraeus, Greece

Samorodnitsky Gena - Cornell University, U.S.A.

Steffensen Mogens - University of Copenhagen, Denmark

Tang Qihe – University of Iowa, U.S.A.

Teugels Jef – Katholieke Universiteit Leuven, Belgium

Zariphopoulou Thaleia - University of Texas at Austin, U.S.A.

Figure 1: Photo of the participants during the third day of the conference.

CONTENTS

An extreme value approximation to the integrated tail distribution

Ants Kaasik
University of Tartu, Estonia
Email: ants.kaasik@ut.ee

Abstract

The idea of approximating the integrated tail distribution of the claim sizes with a Generalized Pareto distribution is presented in the context of the Cramér-Lunberg insurance risk model. The foundations of this semi-paramteric approach are analyzed. Simulation results for subexponential claims clearly show the superiority of the proposed approach over the empirical approach that makes use of the sample cumulative distribution function.

Keywords. Cramér-Lundberg model; Generalized Pareto distribution; Integrated tail distribution; M/G/1 queue; Subexponential distribution.

1 Introduction

In all that follows let X be a heavy-tailed random variable with support on $(0, \infty)$ and a cumulative distribution function (cdf) $F(x)$ such that $\mu = \mathbb{E}X < \infty$. Define an integrated tail distribution of F by $F^I(x) = \int_0^x (1 - F(y))dy/\mu$ and suppose we have an independent random sample from the distribution F (original data) while we are interested in the upper tail of F^I, i.e., in $F^I(x)$ as $F^I(x) \to 1$. The first concern is that we do not have any data from the integrated tail distribution.

To see why this is a realistic case, consider an $M/G/1$ queue and suppose we wish to estimate the probability of a steady-state waiting time exceeding some high value u. This probability can be expressed in terms of a random sum of independent random variables, that have the integrated tail distribution of the service time, exceeding the value u (see p. 237 in [1]). Also the infinite horizon ruin probability of a company with initial capital u can be expressed this way (the summands are then from the integrated tail distribution of the claims distribution) when the Cramér-Lundberg model is assumed. Thus should we know $F^I(x)$ for every $x > 0$, simulation can be used in the mentioned situations for estimating respectively the probability of a large delay or ruin (see e.g. [3], [4] for some effective simulation algorithms applicable for these problems). Apart from simulation also asymptotic approximations could be used, however for the heavy-tailed (subexponential) case the knowledge of the upper tail of F^I is required (see p. 296 in [1]).

While we don't have data from F^I, we can use the definition of the integrated tail distribution and replace the unknown components with their empirical counterparts, i.e. replace $F(x)$ with the empirical cumulative distribution function (ecdf) $F_n(x)$ and μ by the sample mean μ_n. This approach is used in [8], but, although theoretically justified, the approximation is not very reliable in the tail region. More precisely, let u be a high threshold. The problem is that usually we do not have too many data points in the region (u, ∞) and thus also the estimation of the tail of F^I is problematic because

$$\bar{F^I}(u) = 1 - F^I(u) = \frac{\int_u^\infty \bar{F}(y)dy}{\mu} \tag{1}$$

and in case of using the ecdf apporach clearly

$$\bar{F_n^I}(u) = \frac{\int_u^\infty \bar{F_n}(y)dy}{\mu_n} = 0 \tag{2}$$

for any $u > x_{(n)}$, where $x_{(n)}$ is the sample maximum, thus the approximation is limited to the range of the sample.

Extreme value theory offers a possible solution to a similar problem. Suppose X is in a domain of attraction of an extreme value distribution H_ξ (EV condition) i.e. $F^n(a_n x + b_n) \to H_\xi(x)$ for some $a_n > 0$, b_n (see e.g. pp. 120–123 in [5] for more details). It is known that in our case $\xi \in [0, 1)$ (otherwise either the support of X could not be $(0, \infty)$ or the mean of X could not be finite). Define

$$F_u(y) = \frac{F(u+y) - F(u)}{\bar{F}(u)}, \quad u > 0, y > 0 \tag{3}$$

the conditional distribution function of $X - u$ given that $X > u$ and

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \sigma > 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0, \sigma > 0 \end{cases} \tag{4}$$

the cdf of the generalized Pareto distribution (GPD) where $0 < y < -\sigma/\xi$ for $\xi < 0$ and $0 < y < \infty$ for $\xi \geqslant 0$. A well known result (see e.g. pp. 165–166 in [5]) states that when the EV condition holds and X has support $(0, \infty)$, the GPD is a good approximation of F_u for large values of u because

$$\lim_{u\to\infty} \sup_{x>0} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0. \tag{5}$$

This motivates the following general tail approximation (the EV condition is in nearly all cases assumed to be satisfied when we are dealing with continuous data and we assume it to be so throughout the rest of the article). Suppose we have fixed a high threshold u and our sample with size n has N data points that exceed the threshold. Find the estimates $\hat{\xi}_N$ and $\hat{\sigma}_N$ of the GPD parameters using the exceedances (each having u subtracted). Then

$$\bar{F}(u+y) = \bar{F}(u)\bar{F_u}(y) \approx \frac{N}{n}\left(1 + \frac{\hat{\xi}_N y}{\hat{\sigma}_N}\right)^{-1/\hat{\xi}_N}. \tag{6}$$

This approach was introduced and studied in detail in [9] for the case when the GPD parameters were estimated using the maximum likelihood (ML) method. Another well known general method for estimating the parameters of the GPD is the method of probability weighted moments (PWM) studied e.g. in [6]. The aim of this paper is to study the application of this approximation (6) for estimating the tail of F^I using these two methods of parameter estimation and comparing the results with the (strictly) empirical approach.

Our approximation idea is straightforward: fix a large u and estimate the GPD parameters. Based on (6), we might hope that for $y > 0$

$$\bar{F^I}(u+y) \approx \frac{N}{\hat{\mu}_n n} \int_y^\infty \left(1 + \frac{\hat{\xi}_N x}{\hat{\sigma}_N}\right)^{-1/\hat{\xi}_N} dx, \tag{7}$$

where $\hat{\mu}_n$ is the estimate of the expected value.

2 Main Result

We start this section with some definitions.

Definition 2.1. *A positive random variable X with distribution function F is called subexponential (we denote this by $F \in S$) if for all $n \in \mathbb{N}$ it holds that*

$$\lim_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\bar{F}(x)} = n, \tag{8}$$

where F^{*n} denotes the n-fold convolution of F.

Definition 2.2. *The mean residual life $a(u)$ of a positive random variable X with distribution function F at instant $u \geqslant 0$ is defined as*

$$a(u) = \frac{\int_u^\infty \bar{F}(x)dx}{\bar{F}(u)} \tag{9}$$

Remark 2.1. Class S includes many important heavy-tailed distributions, Pareto, Weibull and log-normal being the most prominent members.

It is a known fact that if (5) holds for the conditional distribution of X with $\xi \in (0, 1)$, it also holds for the tail of the integrated tail distribution (with different ξ and $\sigma(u)$) (see e.g. [2]). To see that a similar result also holds for the case when $\xi = 0$ we note that (5) is based on the fact that

$$\lim_{u\to\infty} \sup_{x>0} |\bar{F}_u(xa(u)) - \bar{G}_{\xi,1}(x)| = 0. \tag{10}$$

Denote $f_u(x) := \int_x^\infty \bar{F}_u(ya(u))dy$ and $g(x) := \int_x^\infty \bar{G}_{\xi,1}(y)dy$. The function $g(x)$ is well defined as is $f_u(x) = \int_x^\infty [\bar{F}(u + ya(u))/\bar{F}(u)]dy = \int_x^\infty \bar{F}(u + ya(u))dy/\bar{F}(u) \leq \mu/\bar{F}(u)$ as $\mu = \int_0^\infty \bar{F}(y)dy$. We have that $f_u(x) \to g(x)$ uniformly and

$$\sup_{x>0} \left| \int_x^\infty \bar{F}_u(y)dy - \int_x^\infty \bar{G}_{\xi,a(u)}(y)dy \right| = o(a(u)). \tag{11}$$

This is because $F_u(xa(u))$ can be dominated by an integrable function and thus dominated convergence applies, hence we have $f_u(0) \to g(0)$, which thanks to the Scheffe Lemma yields

$$\int_0^\infty \left| \bar{F}_u(xa(u)) - \bar{G}_{\xi,1}(x) \right| dx \to 0 \tag{12}$$

giving the uniform convergence (an analogue of 10). A change of variable in (12) proves (11). This result is the counterpart of (5) for the integrated tail distribution.

We now proceed to show that when dealing with the integrated tail distribution the tail approximation still behaves like a GPD.

Proposition 2.1. *Suppose that $\forall x > 0$ we have $\bar{F}_u(x) = \bar{G}_{\xi,\sigma}(x)$ for some $u > 0$. Then $\forall y \geqslant 0$*

$$\bar{F^I}(u + y) = \frac{\bar{F}(u)\sigma}{\mu(1-\xi)} \bar{G}_{\xi^*,\sigma^*}(y), \tag{13}$$

where $\xi^* = \xi/(1 - \xi)$ and $\sigma^* = \sigma/(1 - \xi)$.

Proof. Writing
$$\bar{F^I}(u+y) = \frac{\bar{F}(u)}{\mu} \int_y^\infty \bar{G}_{\xi,\sigma}(x)dx, \tag{14}$$
the result now follows directly for $\xi \in [0,1)$ after integration. □

Thus the following methodology suggests itself: fix a large u and estimate the parameters of the GPD using the N exceedances yielding $\hat{\xi}_N$ and $\hat{\sigma}_N$. Approximate the tail of the integrated tail distribution with GPD but with parameters $\hat{\xi}_N/(1-\hat{\xi}_N)$ and $\hat{\sigma}_N/(1-\hat{\xi}_N)$, multiplied by the ratio of $N\hat{\sigma}_N$ and $n(1-\hat{\xi}_N)\mu_n$.

3 A Simulation Study

Clearly we cannot expect the proposed approximation to give as good results as the GPD model can give in the ordinary case. However, as we shall see, the approximation still outperforms the ecdf approach. Even though we have not proven anything (and thus have no *a priori* knowledge) about the behaviour of the approximation due to the error in paramter estimation of GPD and the fact that (5) is only a limit result, we now proceed to estimate $\bar{F^I}(u)$ for large values of u with the methodology described in the end of the previous section.
Because the values of $\bar{F}(x)$ are typically rather small for arguments greater than a high threshold, we study the the supremum relative error of the approximations. Let $A^{emp}_{n,u}$ be the supremum relative error of the ecdf approach and $A^{gpd}_{n,u}$ the counterpart for the GPD approach (7); in both cases the supremum is over $u \leqslant x \leqslant x_{(n)}$. In the role of F we used either the finite-mean Pareto distribution where $\bar{F}(x) = (1+x)^{-\alpha}$ and $\alpha > 1$, the heavy-tailed Weibull distribution where $\bar{F}(x) = e^{-x^\beta}$ and $0 < \beta < 1$, or the log-normal distribution where $\bar{F}(x) = \bar{\Phi}((\ln x)/\sigma)$ and $\sigma > 0$, with some fixed values of the parameter. In each case we performed 1000 simulations and estimated $p(n,N) := \mathbb{P}(A^{gpd}_{n,u} < A^{emp}_{n,u})$, where u is such that the expected number of exceedances is N.

Table 1: Estimates of $p(n,N)$ for the Pareto case with ML

α	n	$N=50$	$N=100$	$N=150$	$N=200$
1.5	1000	0.698±0.028	0.722±0.028	0.747±0.027	0.761±0.026
	10000	0.664±0.029	0.721±0.028	0.729±0.028	0.772±0.026
	100000	0.698±0.028	0.712±0.028	0.718±0.028	0.736±0.027
2.5	1000	0.739±0.027	0.767±0.026	0.803±0.025	0.799±0.025
	10000	0.728±0.028	0.767±0.026	0.802±0.025	0.780±0.026
	100000	0.746±0.027	0.785±0.025	0.777±0.026	0.779±0.026

When the original distribution is Pareto, the integrated tail distribution is also Pareto, moreover (5) holds not only as a limit result but also for every positive u. Thus we can expect that the size of N defines the quality of the GPD approximation. Table 1 confirms this – sample size n plays no part as u is shifted to the right with the increase. In Table 2 a small dependence from n seems to exist. Surprisingly the approximation quality seems to decrease when the overall sample size increases when PWM are used. Overall impression is that the GPD outperforms the ecdf approach and the superiority seems to increase when the tail of the distribution is less heavy or more heavy respectively for the ML and PWM methods.

Table 2: Estimates of $p(n, N)$ for the Pareto case with PWM

α	n	$N = 50$	$N = 100$	$N = 150$	$N = 200$
1.5	1000	0.908±0.018	0.927±0.016	0.936±0.015	0.939±0.015
	10000	0.904±0.018	0.906±0.018	0.925±0.016	0.911±0.018
	100000	0.883±0.020	0.870±0.021	0.892±0.019	0.876±0.020
2.5	1000	0.846±0.022	0.869±0.021	0.905±0.018	0.940±0.015
	10000	0.829±0.023	0.816±0.024	0.851±0.022	0.872±0.021
	100000	0.808±0.024	0.811±0.024	0.815±0.024	0.832±0.023

Table 3: Estimates of $p(n, N)$ for the Weibull case with ML

β	n	$N = 50$	$N = 100$	$N = 150$	$N = 200$
0.25	1000	0.472±0.031	0.183±0.024	0.050±0.014	0.006±0.005
	10000	0.635±0.030	0.483±0.031	0.420±0.031	0.280±0.028
	100000	0.646±0.030	0.606±0.030	0.559±0.031	0.523±0.031
0.50	1000	0.680±0.029	0.565±0.031	0.507±0.031	0.377±0.030
	10000	0.725±0.028	0.682±0.029	0.641±0.030	0.616±0.030
	100000	0.727±0.028	0.728±0.028	0.720±0.028	0.725±0.028
0.75	1000	0.731±0.027	0.712±0.028	0.684±0.029	0.661±0.029
	10000	0.734±0.027	0.748±0.027	0.720±0.028	0.740±0.027
	100000	0.701±0.028	0.747±0.027	0.729±0.028	0.735±0.027

When the original distribution is Weibull, the integrated tail distribution is connected to a gamma distribution (see [7]). Simulation results in Table 3 show that when u is not large enough, the GPD approximation using ML fails (of course, it is of no surprise as (6) holds only for large threshold values). However, when u is set sufficiently high and the number of exceedances is big enough to properly estimate the parameters, GPD approach once again works better than the one using the ecdf. Too low threshold hinders the PWM method as well as can be seen from Table 4, however it clearly outperforms the empirical approach in the current simulation study.

For the log-normal case the GPD approach using ML performs better than the ecdf approach when u is not set too low. The difference is once again bigger for lighter tails. The situation is different when the PWM method is used as the superiority of the extreme value approach is more evident with heavier tails.

It is worth noting that in practice u is typically chosen according to the data (see e.g. Section 6.5 in [5]) and not simply as a specific quantile dependent only on the sample size. Thus one would expect the GPD apporach to perform even better when the data is carefully examined before fixing the threshold.

4 Conclusion

Estimating the distribution function of an integrated tail distribution from the original data is a hard task when we have no addtional information. The approach of using GPD approximation in the tail part remains valid, but does not produce as good results as in the usual context. When using the method of maximum likelihood for parameter estimation, the approach might fail altogether when u is not sufficiently high. For the tail sample sizes ranging from 50 to 200 the method of

Table 4: Estimates of $p(n, N)$ for the Weibull case with PWM

β	n	$N = 50$	$N = 100$	$N = 150$	$N = 200$
0.25	1000	0.783±0.026	0.766±0.026	0.696±0.029	0.597±0.030
	10000	0.780±0.026	0.743±0.027	0.724±0.028	0.659±0.029
	100000	0.737±0.027	0.712±0.028	0.682±0.029	0.683±0.029
0.50	1000	0.735±0.027	0.732±0.027	0.703±0.028	0.641±0.030
	10000	0.741±0.027	0.720±0.028	0.748±0.027	0.718±0.028
	100000	0.699±0.028	0.706±0.028	0.697±0.028	0.691±0.029
0.75	1000	0.746±0.027	0.763±0.026	0.777±0.026	0.771±0.026
	10000	0.737±0.027	0.759±0.027	0.757±0.027	0.747±0.027
	100000	0.701±0.028	0.703±0.028	0.723±0.028	0.701±0.028

Table 5: Estimates of $p(n, N)$ for the log-normal case with ML

σ	n	$N = 50$	$N = 100$	$N = 150$	$N = 200$
1	1000	0.712±0.028	0.692±0.029	0.675±0.029	0.636±0.030
	10000	0.726±0.028	0.711±0.028	0.709±0.028	0.710±0.028
	100000	0.741±0.027	0.724±0.028	0.730±0.028	0.736±0.027
2	1000	0.547±0.031	0.422±0.031	0.337±0.029	0.219±0.026
	10000	0.648±0.030	0.587±0.031	0.577±0.031	0.503±0.031
	100000	0.655±0.029	0.623±0.030	0.637±0.030	0.635±0.030

Table 6: Estimates of $p(n, N)$ for the log-normal case with PWM

σ	n	$N = 50$	$N = 100$	$N = 150$	$N = 200$
1	1000	0.783±0.026	0.774±0.026	0.823±0.024	0.817±0.024
	10000	0.740±0.028	0.747±0.027	0.764±0.026	0.772±0.026
	100000	0.706±0.028	0.732±0.027	0.729±0.028	0.745±0.027
2	1000	0.862±0.021	0.825±0.024	0.820±0.024	0.808±0.024
	10000	0.790±0.025	0.818±0.024	0.789±0.025	0.787±0.025
	100000	0.803±0.025	0.778±0.026	0.788±0.025	0.780±0.026

probability weighted moments seems to guarantee an improvement over the ecdf approach.

Acknowledgement

Partially supported by Estonian Science Foundation grants ETF 7044 and ETF 7313.

References

[1] Asmussen, S.: 2003, *Applied probability and queues. 2nd revised and extended ed.* New York, NY: Springer.

[2] Asmussen, S. and C. Klppelberg: 1996, 'Large deviations results for subexponential tails, with applications to insurance risk'. *Stochastic processes and their applications* **64**(1), 103–125.

[3] Asmussen, S. and D. P. Kroese: 2006, 'Improved algorithms for rare event simulation with heavy tails'. *Advances in Applied Probability* **38**(2), 545–558.

[4] Blanchet, J. and P. Glynn: 2008, 'Efficient rare-event simulation for the maximum of heavy-tailed random walks'. *Annals of Applied Probability* **18**, 1351–1378.

[5] Embrechts, P., T. Mikosch, and C. Klüppelberg: 1997, *Modelling extremal events: for insurance and finance*. London, UK: Springer-Verlag.

[6] Hosking, J. and J. Wallis: 1987, 'Parameter and quantile estimation for the generalized pareto distribution' *Technometrics* **29**(3), 339–349.

[7] Kaasik, A., 'Simulating the integrated tail distribution in the heavy-tailed setting'. Accepted for publication in *Proceedings of the 28th IASTED International Conference on Modelling, Identification and Control*.

[8] Kaasik, A. and K. Prna, 'Approximating the integrated tail distribution'. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **12**, 79–88.

[9] Smith, R. L.: 1987, 'Estimating Tails of Probability Distributions'. *The Annals of Statistics* **15**(3), 1174–1207.

Risk models with extremal subexponentiality

D.G. Konstantinides
Department of Statistics and Actuarial-Finance Mathematics
University of the Aegean
Karlovassi 83200, Samos, Greece
Email: konstant@aegean.gr

Abstract

In this paper we consider risk models with a heavy-tailed parametric claim distribution from the subexponential class $\mathcal{S}$ with at least two parameters. We choose a proper convergence of a parameter, that makes the tail of the claims distribution heavier or lighter and then tend it to its limitation. Finally we proceed to an appropriate functional normalization in order to keep the distributional properties.

1 Introduction.

In this paper the following problem is investigated: We consider a heavy-tailed parametric distribution from the subexponential class $\mathcal{S}$ with at least two parameters. We shall demand such a relation between the parameters, that the safety loading coefficient remains fixed. Choose a proper convergence of a parameter, that makes the tail of the claims distribution heavier and then tend it to its limitation. What happens then with the corresponding ruin probability under some special risk models?

We consider the classical risk model where the claims occur at random epochs which form a homogeneous Poisson process $\{N(t)\,,\ t \geq 0\}$ with intensity $\lambda > 0$, which is independent of the claims $Z_k\,,\ k = 1, 2, \ldots$. Using the notation of $\overline{B}(x) = 1 - B(x)$ for the tail of the claim distriburion $B(x)$, of $b(x)$ for the density and of $b = \mathbf{E}Z$ for the mean claim size, we denote the expected claim per time unit by $\rho = \lambda b$ and the

$$X(t) = \sum_{k=1}^{N(t)} Z_k$$

is the compound Poisson process representing the total claim amount accumulated until time t. Thus

$$F(x) = \frac{1}{b} \int_0^x \overline{B}(z)\, dz \tag{1}$$

is the integrated tail of the claim size distribution. In the classical risk model, the Pollaczeck-Khinchine formula takes the form

$$\psi(u) = \Big(1 - \frac{\rho}{c}\Big) \sum_{n=0}^{\infty} \Big(\frac{\rho}{c}\Big)^n \overline{F}^{n*}(u)\,, \tag{2}$$

and gives the main tool for calculation of the ruin probability.

If $F \in \mathcal{S}$, the following asymptotic formula was found in [3]

$$\psi(u) \sim \frac{\rho}{c-\rho}\overline{F}(u) = \frac{\rho}{(c-\rho)\, b} \int_u^{\infty} \overline{B}(y) dy\,, \tag{3}$$

as $u \to \infty$.

The motivation of this problem comes from the following observation. In the vicinities of the critical values of the parameter, where the ergodicity condition does not hold any more, the ruin probability jumps to 1. The practical importance of this statement is shown through the unexpected result that the ruin probability on these neighborhoods does not depend any more on the initial capital. Obviously, this fact opens a new problem of approximation of the ruin probability in these areas.

Indeed, we often deal in insurance and finance with large claims that are described by heavy-tailed distributions (Pareto, Lognormal, Weibull, Loggamma, Burr). The known results reveal only asymptotic behavior of the ruin probabilities. Numerical calculations show that the accuracy of these asymptotic formulas can be quite low, especially in the range that is relevant for practical purposes (see for example [6], [10]).

It is worthy of notice the special importance of heavy-tailed distributions, which is increasing the last years because of occasional appearance of huge claims. The problem consists in proposing other approximations that work in the area of practically significant values of the corresponding parameters and variables. To this end, the classification of the distributions describing large claims is promoted. This approach presents a new classification, since up to now all heavy-tailed distributions were considered as simple members of the subexponential class $\mathcal{S}$.

We concentrate our study on five concrete subexponential parametric families, widely used in insurance mathematics.

Example 1.1. *Pareto:*

$$\overline{B}(x) = \begin{cases} 1, & \textit{when} \quad x \leq k, \\ \frac{k^{\alpha}}{x^{\alpha}}, & \textit{when} \quad x > k, \end{cases}$$

with $\alpha > 1,\ k > 0$;

Example 1.2. *Lognormal:*

$$b(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\},$$

with μ *real number and* $\sigma > 0$*;*

Example 1.3. *Weibull:*

$$\overline{B}(x) = e^{-\nu^{\tau} x^{\tau}},$$

with $\nu > 0$, $0 < \tau < 1$*;*

Example 1.4. *Loggamma:*

$$b(x) = \frac{\alpha^p}{\Gamma(p)} \left[\ln(1+x)\right]^{p-1} (1+x)^{-\alpha-1},$$

with $\alpha > 0,\ p > 0$*;*

Example 1.5. *Burr:*

$$\overline{B}(x) = \left(\frac{\kappa}{\kappa + x^r} \right)^{\alpha},$$

with $\kappa > 0,\ r > 0,\ \alpha > 1/r$*;*

2 The Heuristics.

Firstly let us take the example of the Pareto distribution, in which the parameter α will be considered as its parameter of heavytailedness. If the other parameter k is fixed and α tends to its minimal value 1 and consequently the expectation of the claim sizes tends to infinity, and the corresponding ladder height process is not ergodic any more in the limit. In such a case the integrated tail claim distribution from (1) is meaningless and the Pollaczeck-Khinchine formula does not work. So in order to keep the balance within the chosen convergence of α, either the safety loading or in particular the mean claim

$$b = \frac{\alpha k}{\alpha - 1}$$

must be held fixed, which leads to the normalization. As a result of this, we put as our normalization condition

$$b = 1\,, \tag{1}$$

which is common for all the examples listed above. Thus we obtain $\rho = \lambda$.

In our example of the Pareto distribution the relation (1) implies that the second parameter k takes the value

$$k = \frac{\alpha - 1}{\alpha}.$$

It should be noted that this kind of ergodic control is not unique (see [6]).

Now the heavytailedness parameter converges to the limit which demonstrates its most heavily (superheavy) tailed distribution. Namely, in the example of the Pareto distribution, α tends to its least value as follows

$$\alpha \longrightarrow 1.$$

Secondly, repeating these steps for the Lognormal distribution, it follows that the normalization (1) implies

$$\mu = -\frac{\sigma^2}{2},$$

and that the value of the heavytailedness parameter σ tends as follows

$$\sigma \longrightarrow \infty,$$

caused by the need to identify the superheavy tailed distribution.

In the next example of the Weibull distribution, this pattern of the normalization (1) renders

$$\nu = \Gamma\left(1 + \frac{1}{\tau}\right),$$

where $\Gamma(.)$ denotes the Gamma function. The most heavy tailed distribution arises when the heavytailedness parameter τ tends to 0.

Further, when we consider the Loggamma distribution according to the normalization procedure (1), we take

$$\alpha = \frac{2^{1/p}}{2^{1/p} - 1}.$$

Now the heavytailedness parameter p tends to 0.

Finally, in the case of the Burr distribution the normalization (1) leads to

$$\kappa = \left(\frac{r\Gamma(\alpha)}{\Gamma(1/r)\Gamma(\alpha - 1/r)} \right)^r . \tag{2}$$

When r is fixed, the parameter of heavytailedness α tends to $1/r$.

Remark 2.1. If $B(x)$ belongs to the Pareto, Lognormal, Weibull, Loggamma or Burr distribution family, then for any $x > 0$ its tail tends to zero

$$\overline{B}(x) \to 0, \tag{3}$$

as the heavytailedness parameter tends to its limit ($\alpha \to 1$, $\sigma \to \infty$, $\tau \to 0$, $p \to 0$, $\alpha \to 1/r$ correspondingly).

Indeed, for the Pareto distribution family, we take:

$$\overline{B}(x) \sim \frac{1}{x}(\alpha - 1)^\alpha \to 0,$$

as $\alpha \to 1$ and for any $x > 0$.

For the Lognormal distribution family, it holds:

$$\overline{B}(x) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln x}{\sigma} + \frac{\sigma}{2}}^{\infty} \exp\left[-y^2/2\right] dy \to 0,$$

as $\sigma \to \infty$ and for any $x > 0$.

For the Weibull distribution family by using Stirling's formula it can be found that:

$$(\nu x)^\tau \to \left[x\Gamma\left(1 + \frac{1}{\tau}\right)\right]^\tau \sim (\sqrt{2\pi})^\tau \left(\frac{1}{\tau}\right)^{\tau/2} \frac{x^\tau}{\tau e} \to \infty,$$

as $\tau \to 0$, from which the limit follows immediately.

For the Loggamma distribution family for any $\varepsilon \in (0, x)$, the following sequence of relations is true:

$$\overline{B}(x) = \frac{1}{\Gamma(p)} \int_{A_p(x)}^{\infty} w^{p-1} e^{-w} dw \le \frac{1}{\Gamma(p)} \int_{\ln(1+x)}^{\infty} w^{p-1} e^{-w} dw \le$$

$$\le 1 - \frac{1}{\Gamma(p)(1+\varepsilon)} \int_0^{\ln(1+x)} w^{p-1} dw = 1 - \frac{\ln^p(1+x)}{p\Gamma(p)(1+\varepsilon)} \to \frac{\varepsilon}{1+\varepsilon},$$

as $p \to 0$, where

$$A_p(x) = \frac{2^{1/p}}{2^{1/p} - 1} \ln(x+1)\,.$$

In the last step we considered the well-known asymptote

$$\Gamma(\alpha) \sim \frac{1}{\alpha}, \tag{4}$$

as $\alpha \to 0$.

For the Burr distribution family for fixed r the asymptote (4) gives:

$$\kappa \sim r^r \left(\alpha - \frac{1}{r}\right)^r \to 0,$$

as $\alpha \to 1/r$, from which the limit follows immediately. $\square$

3 Superheavy Subexponential Tails

Lemma 3.1. (Tsitsiashvili-Konstantinides [12]) *In the classical risk model, if $B(x)$ belongs to the Pareto, Lognormal, Weibull, Loggamma or Burr distribution family, then for any $u > 0$, the ruin probability tends to a constant*

$$\psi(u) \to \frac{\rho}{c},$$

as the heavytailedness parameter tends to its limit ($\alpha \to 1$, $\sigma \to \infty$, $\tau \to 0$, $p \to 0$, $\alpha \to 1/r$ correspondingly).

Proof. Firstly let us take the Pareto distribution family. According to the results of the Remark 2.1 it follows that for any $\varepsilon \in (0, u)$ there is a constant $\alpha_0 > 1$ such that

$$\overline{B}(\varepsilon) \leq \frac{\varepsilon}{u - \varepsilon},$$

for any $\alpha \in (1, \alpha_0)$. So from the Pollaczeck-Khinchine formula (2), the following chain of inequalities can be taken:

$$\begin{aligned} \frac{\rho}{c} = \psi(0) \geq \psi(u) \quad &\geq \quad \frac{\rho}{c}\overline{F}(u) = \frac{\rho}{c}\left(1 - \int_0^{\varepsilon} \overline{B}(y)dy - \int_{\varepsilon}^{u} \overline{B}(y)dy\right) \\ &\geq \quad \frac{\rho}{c}[1 - \varepsilon - (u - \varepsilon)\overline{B}(\varepsilon)] \geq \frac{\rho}{c}[1 - 2\varepsilon], \end{aligned}$$

for any $\alpha \in (1, \alpha_0)$, which, due to the arbitrariness of ε, gives the desired convergence. Easily we can verify that the same argument holds for the rest of the distribution families. □

We see that the superheavy limit of the claim distribution in Remark 2.1 does not present a distribution and the superheavy limit of the ruin probability in Lemma 3.1 is not a decreasing function with respect to u. These deformations of the standard properties of the distribution and the ruin probability can be explained through an explosive behavior by the convergence to the limit. To preserve the standard properties in the course of the limit passage we apply a functional normalization. Namely we take a functional heavytailedness parameter, say $\alpha(u) > 1\,,\ \ \forall\ u \geq 0$ in the first case, such that $\alpha(u) \downarrow 1$ as $u \to \infty$.

Theorem 3.2. *If $B(x)$ belongs to the Pareto, Lognormal, Weibull, Loggamma or Burr distribution families and its heavytailedness parameter tends to its limit in the following way:*

$$\alpha(x) \downarrow 1\,, \quad [\alpha(x) - 1]\, x \to \infty\,,$$

$$\sigma(x) \to \infty\,, \quad \sigma(y) < 2\, y\,, \quad \forall\, y > 0\,,$$

$$\tau(x) \downarrow 0\,, \quad y\, \tau(y) > 1\,, \quad \forall\, y > 0\,,$$

$$p(x) \downarrow 0\,,$$

$$\alpha(x) \downarrow 1/r\,, \quad \frac{1}{r\, \alpha(x) - 1} = o\left(x^{1/r}\right)\,,$$

as $x \to \infty$ respectively, then their normalized tails tend to the following limits:

$$\overline{B}([\alpha(x) - 1]x) \sim \left(\frac{1}{x\, \alpha(x)}\right)^{\alpha(x)}\,,$$

$$\overline{B}\left(\exp\left\{\sigma(x)\,x-\frac{1}{2}\sigma^2(x)\right\}\right)\sim\frac{1}{\sqrt{2\,\pi}}\int_x^\infty\exp\left[-\frac{y^2}{2}\right]\,dy\,,$$

$$\overline{B}\left(\sqrt{\frac{\tau(x)}{2\,\pi}}\,[e\,x\,\tau(x)]^{1/\tau(x)}\right)\sim e^{-x}$$

$$\overline{B}\left(\exp\left\{\frac{x}{2^{1/p(x)}}\left(2^{1/p(x)}-1\right)\right\}-1\right)\sim\frac{1}{\Gamma[p(x)]}\int_x^\infty y^{p(x)-1}\,e^{-y}\,dy$$

$$\overline{B}([r\,\alpha(x)-1]\,(x-1)^{1/r})\sim\left(\frac{1}{x}\right)^{\alpha(x)}$$

as $x\to\infty$. Furthermore, in the classical risk model, the corresponding ruin probabilities tend to the following limits:

$$\psi([\alpha(u)-1]\,u)\sim\frac{\rho}{c-\rho}\int_u^\infty\frac{\alpha(z)-1+z\,\alpha'(z)}{(z\,[\alpha(z)-1]\,\alpha(z\,[\alpha(z)-1]))^{\alpha(z\,[\alpha(z)-1])}}\,dz\,,$$

$$\psi\left(\exp\left\{\sigma(u)\,u-\frac{1}{2}\sigma^2(u)\right\}\right)\sim$$

$$\frac{\rho}{\sqrt{2\,\pi}(c-\rho)}\int_u^\infty\frac{\sigma(z)+\sigma'(z)[z-\sigma(z)]}{\exp\left\{-z\,\sigma(z)+\frac{1}{2}\sigma^2(z)\right\}}\left[\int_{\exp\left(z\,\sigma(z)-\frac{1}{2}\sigma^2(z)\right)}^\infty\exp\left(-\frac{w^2}{2}\right)\,dw\right]\,dz\,,$$

$$\psi\left(\sqrt{\frac{\tau(u)}{2\,\pi}}[e\,\tau(u)\,u]^{1/\tau(u)}\right)\sim$$

$$\frac{\rho}{\sqrt{2\,\pi}(c-\rho)}\int_u^\infty e^{-\frac{\sqrt{\tau(z)}}{\sqrt{2\,\pi}}\,[e\,z\,\tau(z)]^{\frac{2}{\tau(z)}}}\,\tau'(z)\left\{\frac{1}{2\,\sqrt{\tau(z)}}-e\,\frac{z\tau'(z)+\tau(z)}{\tau^2(z)}\,\log[e\,z\,\tau(z)]\right\}\,dz\,,$$

$$\psi\left(\exp\left\{\frac{u}{2^{1/p(u)}}\left(2^{1/p(u)}-1\right)\right\}-1\right)\sim$$

$$\frac{\rho}{c-\rho}\int_u^\infty\frac{\left(1-2^{-1/p(z)}\right)\,\exp\left\{\left(1-2^{-1/p(z)}\right)\,z\right\}}{\Gamma\left[p\left(\exp\left[\left(1-2^{-1/p(z)}\right)\,z\right]-1\right)\right]}\int_{\exp\left[\left(1-2^{-1/p(z)}\right)z\right]-1}^\infty w^{p\left(\exp\left[\left(1-2^{-1/p(z)}\right)z\right]-1\right)}\,e^{-w}\,dw\,dz\,,$$

$$\psi\left([r\,\alpha(u)-1]\,(u-1)^{1/r}\right)\sim\frac{\rho}{c-\rho}\int_u^\infty\left(\frac{1}{[r\,\alpha(z)-1]\,(z-1)^{1/r}}\right)^{\alpha\left([r\,\alpha(z)-1]\,(z-1)^{1/r}\right)}[r\,\alpha'(z)\,(z-1)^{1/r}]\,dz\,,$$

as $u\to\infty$.

Proof. Let us start with the case of Pareto claim sizes. We look for such a nomalizing function $f[x,\alpha(x)]$, that the expression $B(x\, f[x,\alpha(x)])$ remains a distribution after the passage to the limit $x \to \infty$. This means

1. $x\, f[x,\alpha(x)] \downarrow 0$, as $x \downarrow 0$,
2. $x\, f[x,\alpha(x)] \uparrow \infty$, as $x \to \infty$.

Within this framework, we find that the function

$$f[x,\alpha(x)] = \alpha(x) - 1\,,$$

meets the requirements above and serves as candidate for the normalizing function.

Next we fix the value of $\alpha(u) = \alpha > 1$ and we shall obtain the following uniform asymptotics for the ruin probability when $u \to \infty$,

$$\lim_{u\to\infty} \sup_{\alpha>1} \left| \frac{\psi(u\, f[u,\alpha])}{\dfrac{\rho}{c-\rho}\overline{F}(u\, f[u,\alpha])} - 1 \right| = \lim_{u\to\infty} \sup_{\alpha>1} \left| \frac{\psi(u\,[\alpha-1])}{\dfrac{\rho}{c-\rho}\overline{F}(u\,[\alpha-1])} - 1 \right| = 0\,. \tag{1}$$

Indeed, as far $F(u\,[\alpha-1])$ represents a subexponential distribution function and $\dfrac{c-\rho}{2\,c} > 0$, there exists some constant $K = K\left(\dfrac{c-\rho}{2\,c}\right)$ (see for example in [2, Lemma 1.3.5]) such that for any integer $N \geq 1$ we obtain

$$\begin{aligned}
\psi(u\,[\alpha-1]) &= \frac{c-\rho}{c} \sum_{n=0}^{N-1} \left(\frac{\rho}{c}\right)^n \overline{F}^{n*}(u\,[\alpha-1]) + \frac{c-\rho}{c} \sum_{n=N}^{\infty} \left(\frac{\rho}{c}\right)^n \overline{F}^{n*}(u\,[\alpha-1]) \\
&\sim \frac{c-\rho}{c} \sum_{n=0}^{N-1} \left(\frac{\rho}{c}\right)^n n\,\overline{F}(u\,[\alpha-1]) + \frac{c-\rho}{c} \sum_{n=N}^{\infty} \left(\frac{\rho}{c}\right)^n \overline{F}^{n*}(u\,[\alpha-1]) \\
&\leq \left[\frac{\rho}{c-\rho} - \left(\frac{c}{c-\rho} + N - 1\right) \left(\frac{\rho}{c}\right)^N + \frac{c-\rho}{c} \sum_{n=N}^{\infty} K \left(\frac{\rho}{c}\left[1 + \frac{c-\rho}{2\,c}\right]\right)^n\right] \overline{F}(u\,[\alpha-1])
\end{aligned}$$

as $u \to \infty$. Therefore, for any real $M > 1$ we find

$$\lim_{u\to\infty} \sup_{\alpha\in[M,\infty)} \left| \frac{\psi(u\,[\alpha-1])}{\dfrac{\rho}{c-\rho}\overline{F}(u\,[\alpha-1])} - 1 \right| \leq \left| \left(1 + N\,\frac{c-\rho}{\rho}\right) \left(\frac{\rho}{c}\right)^N \right| + \left| \frac{2\,K\,(c-\rho)}{\rho} \left(\frac{c+\rho}{2\,c}\right)^N \right|.$$

Now we take the limit for $N \to \infty$ and the limit for $M \to 1$ and we reach the asymptotic relation (1).

Now we see that the expression $1 - \overline{B}(x\, f[x,\alpha(x)]) = 1 - \overline{B}(x\,[\alpha(x)-1])$ remains distribution after the passage to the limit because $x\,[\alpha(x)-1] \to \infty$ as $x \to \infty$ and consequently

$$\psi(u\, f[u,\alpha(u)]) \sim \frac{\rho}{c-\rho} \int_u^\infty \overline{B}(z\, f[z,\alpha(z)]) \left(z\,\frac{d(f[z,\alpha(z)])}{dz} + f[z,\alpha(z)]\right) dz\,.$$

Further, we continue with the rest cases under the following normalizing functions respectively:

$$
\begin{aligned}
f[x,\sigma(x)] &= \frac{1}{x}\,\exp\left\{\sigma(x)\,x - \frac{1}{2}\sigma^2(x)\right\}\,, \\
f[x,\tau(x)] &= \frac{1}{x}\,\sqrt{\frac{\tau(x)}{2\,\pi}}\,[e\,\tau(x)\,x]^{1/\tau(x)}\,, \\
f[x,p(x)] &= \frac{1}{x}\,\left[\exp\left\{\frac{x}{2^{1/p(x)}}\left(2^{1/p(x)}-1\right)\right\}-1\right]\,, \\
f[x,\alpha(x)] &= \frac{1}{x}\,[r\,\alpha(x)-1]\,(x-1)^{1/r}\,.
\end{aligned}
$$

It remains to repeat the uniform convergence and with the substitution, we reach the other ruin probability formulas. □

All the five examples of heavy tailed distributions outlined in the previous statements belong to the class of subexponential distributions $\mathcal{S}$. Two members of $\mathcal{S}$ which serve as exceptions to the Lemma 3.1 are given below.

Example 3.3. *The Bektander I distribution with tail*

$$\overline{B}(x) = \left[1 + \frac{2p}{\alpha}\ln(1+x)\right]\exp\left\{-p\ln^2(1+x) - (\alpha+1)\ln(1+x)\right\}$$

with $\alpha > 0$, $p > 0$, $x > 0$.

Example 3.4. *The Bektander II distribution with tail*

$$\overline{B}(x) = \frac{1}{(1+x)^{1-p}}\exp\left\{\frac{\alpha}{p}\left[1-(1+x)^p\right]\right\},$$

with $\alpha > 0$, $p \in [0,1]$, $x > 0$.

We consider again the classical risk model. In both cases, the normalization condition (1) gives

$$\alpha = 1$$

and the heavytailedness parameter tends to zero

$$p \to 0.$$

Let us denote $\psi^*(u)$ the ruin probability in the classical risk model in which the tail of claim is from the Pareto distribution: $\overline{P}(x) = 1/(1+x)^2$, for $x > 0$ (that claim distribution coincides with the Burr(2,1)).

Theorem 3.5. (Tsitsiashvili-Konstantinides [12]) *In the classical risk model, if* $B(x)$ *belongs to either Bektander I or Bektander II distribution family, then for any* $x > 0$*, its tail converges in* $\mathcal{L}^1$ *to* $\overline{P}(x)$*:*

$$\int_0^\infty |\overline{B}(x) - \overline{P}(x)|dx \to 0,$$

and for any $u > 0$*, the ruin probability tends weakly (⇒) to the function* $\psi^*(u)$*, which represents the stationary distribution tail of the waiting time in the* $M/G/1/\infty$ *queuing system with service time distribution* $P(x)$*:*

$$\psi(u) \Rightarrow \psi^*(u),$$

as the heavytailedness parameter tends to zero: $p \to 0$.

Proof. Firstly let us take the example of the Bektander I distribution. Here, for any $T > 0$

$$\int_0^\infty \left|\overline{B}(x) - \overline{P}(x)\right| dx \le \int_0^T \left|\overline{B}(x) - \overline{P}(x)\right| dx + \int_T^\infty \overline{P}(x)dx + \int_T^\infty \overline{B}(x)dx$$

$$\le \quad T \sup_{[0,T]} \overline{P}(x) \left|[1 + 2p\ln(1+x)]\exp[-p\ln^2(1+x)] - 1\right| + \frac{1}{1+T}$$

$$+ \quad \int_T^\infty \left\{\exp[-2\ln(1+x)] + 2p\ln(1+x)\exp\left[-p\ln^2(1+x) - 2\ln(1+x)\right]\right\} dx$$

$$\le \quad T \sup_{[0,T]} \left\{\left|\exp\left[-p\ln^2(1+x)\right] - 1\right| + 2p\ln(1+x)\right\} + 2\left(\frac{1}{1+T} + (1+T)e^{-T}\right).$$

Further, for any $\varepsilon > 0$ there exists a T_ε and a $p_0 > 0$ such that for any $p \in (0, p_0)$ the last expression becomes less than ε, so

$$\int_0^\infty \left|\overline{B}(x) - \overline{P}(x)\right| dx \le T \sup_{[0,T]} \left\{\left|\exp\left[-p\ln^2(1+x)\right] - 1\right| + 2p\ln(1+x)\right\} + \varepsilon \longrightarrow \varepsilon,$$

as $p \longrightarrow 0$ and thus the convergence in $\mathcal{L}^1$ for the Bektander I case is obtained. Similarly for the Bektander II case, considering the uniform convergence over any finite interval $[0, T]$

$$\exp\left(\frac{1 - e^{p\ln(1+x)}}{p}\right) \to \frac{1}{1+x},$$

as $p \to 0$, the convergence of the claim tail in $\mathcal{L}^1$ is confirmed.

In both cases the convergence of the ruin probability can be verified from a well-known result of the stability theory (see for example [5], [9]). Indeed, in classical risk model the convergence of $\overline{B}(x)$ to $\overline{P}(x)$ in $\mathcal{L}^1$ implies the convergence of the ruin probability $\psi(u)$ to the function $\psi^*(u)$, which represents the ruin probability with claim distribution $P(x)$ (or, in other words, it represents the stationary distribution of the waiting time in the $M/G/1/\infty$ queuing system with service time distribution $P(x)$). □

Remark 3.6. For the distributions from the five examples of Theorem 3.2, the convergence in $\mathcal{L}^1$ does not hold and therefore this argument from the stability theory is not applicable.

Remark 3.7. It is possible to prove that the function $\psi^*(u)$ is continuous and so the weak convergence in Theorem 3.5 can be replaced by the point wise convergence for any $u \ge 0$.

Remark 3.8. In the superheavy tail mode, the numerics become unstable, because the values of the ruin probability become too small and the precision in calculation of the intergrals fails. This observation brings up promptly the numerical issue.

4 Light Subexponential Tails

Now our interest is directed on subexponential distributions that are lying in close vicinity to the light-tailed distributions. We begin with the Pareto distribution in which the parameter α is chosen as before for the role of heavytailedness parameter. Obviously, when it tends to its limit ∞ we reach the lightest distribution tail. In the second case with the Lognormal distribution, the heavytailedness parameter σ has to tend to 0 in order to find the lightest tail. Next in the example related with the Weibull case, the heavytailedness parameter τ tends to 1. Further, in the example of the Loggamma distribution, the heavytailedness parameter p tends to ∞. In the last case with the Burr distribution, the two-dimensional heavytailedness parameter (α, r) tends to (∞, ∞). For the Bektander I, distribution the heavytailedness parameter p tends to ∞. Finally, in the example with Bektander II distribution, the heavytailedness parameter p tends to 1.

We proceed to the limit distributions.

Lemma 4.1. *If $B(x)$ belongs to one of the distribution families: Pareto, Lognormal, Weibull, Loggamma, Burr or Bektander II then its tail tends to a limit distribution:*

$$\overline{B}(x) \to \overline{D}(x), \tag{1}$$

as the corresponding heavytailedness parameter reaches its limit ($\alpha \to \infty$, $\sigma \to 0$, $\tau \to 1$, $p \to \infty$, $(\alpha, r) \to (\infty, \infty)$ and $p \to 1$ respectively). Namely in the Pareto, Lognormal, Loggamma and Burr cases, the limit distribution $\overline{D}(x)$ represents a step function:

$$\overline{D}(x) = \begin{cases} 1, & 0 \le x < 1, \\ C, & x = 1, \\ 0, & 1 < x < \infty, \end{cases}$$

for some constant $C \in (0,1)$, in the Weibull and Bektander II cases it represents an exponential distribution:

$$\overline{D}(x) = e^{-x}\,.$$

Furthermore in the Pareto, Lognormal, Weibull, Loggamma, Burr and Bektander II cases, the $\mathcal{L}^1$ convergence holds:

$$\int_0^\infty \left|\overline{B}(y) - \overline{D}(y)\right| dy \to 0.$$

Proof. For the Pareto distribution family it is easy to find:

$$\overline{B}(x) = \begin{cases} 1, & 0 \le x \le 1 - \frac{1}{\alpha}, \\ \frac{1}{x^\alpha}(1-\frac{1}{\alpha})^\alpha & 1 - \frac{1}{\alpha} < x < \infty, \end{cases} \to \begin{cases} 1, & 0 \le x < 1, \\ e^{-1}, & x = 1, \\ 0, & 1 < x < \infty, \end{cases}$$

as $\alpha \to \infty$. Thus (1) holds with $C = e^{-1}$.

For the Lognormal distribution family for any $x > 0$ holds:

$$\overline{B}(x) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln x}{\sigma} + \frac{\sigma}{2}}^\infty \exp\left\{-\frac{y^2}{2}\right\} dy \to \begin{cases} 1, & 0 \le x < 1, \\ 1/2, & x = 1, \\ 0, & 1 < x < \infty, \end{cases}$$

as $\sigma \to 0$. Here the relation (1) holds with $C = 1/2$.

For the Weibull distribution family the following limit distribution can be found

$$\overline{B}(x) \longrightarrow e^{-x},$$

as $\tau \longrightarrow 1$, because $\left[\Gamma\left(1+1/\tau\right)\right]^{\tau} \longrightarrow 1$.

At the Loggamma distribution family for any $\varepsilon \in (0,x)$,

$$\overline{B}(x) = \frac{1}{\Gamma(p)} \int_{A_p(x)}^{\infty} w^{p-1} e^{-w} dw,$$

where

$$A_p(x) = \frac{2^{1/p}}{2^{1/p}-1} \ln(x+1) \sim p \frac{\ln(x+1)}{\ln 2},$$

as $p \to \infty$. Let us take

$$z = \frac{\ln(x+1)}{\ln 2}.$$

For $x < 1$, as $z < 1$ the asymptote can be found with the help of the Stirling formula

$$\begin{aligned} \overline{B}(x) &\sim \frac{1}{\Gamma(p)} \int_{pz}^{\infty} w^{p-1} e^{-w} dw \\ &\geq 1 - \frac{1}{\Gamma(p)} (pz)^{p-1} e^{-pz} pz \sim 1 - \frac{p\left(ze^{1-z}\right)^p}{\sqrt{2\pi(p-1)}} \to 1, \end{aligned} \tag{2}$$

as $p \to \infty$, because the function $w^{p-1}e^{-w}$ reaches its maximum at $w = p-1 \sim p$ and in turn the function ze^{1-z} reaches its maximum equal to 1 at $z = 1$. Furthermore, this convergence is uniform with respect to $x \in [0, 1-\varepsilon]$ for any $\varepsilon \in (0,1)$.

For $x > 1$, as $z > 1$ the asymptote can be found similarly

$$\begin{aligned} \overline{B}(x) &\sim \frac{1}{\Gamma(p)} \int_{pz}^{2p} w^{p-1} e^{-w} dw + \frac{1}{\Gamma(p)} \int_{2p}^{\infty} w^{p-1} e^{-w} dw \\ &\leq \frac{1}{\Gamma(p)} p^p e^{-pz} + \frac{2}{\Gamma(p)} \int_{p}^{\infty} \left[2(p-1)\right]^{p-1} e^{-p+1} e^{-u} du \\ &\sim \frac{p\left(ze^{1-z}\right)^p}{z\sqrt{2\pi(p-1)}} + \frac{1}{(e/2)^p \sqrt{2\pi(p-1)}} \to 0, \end{aligned}$$

as $p \to \infty$, with $u = w/2$, because the function $w^{p-1}e^{-w/2}$ reaches its maximum at $w = 2(p-1)$.

For $x = 1$ let us take

$$\overline{B}(1) \sim \frac{1}{\Gamma(p)} \int_{p}^{\infty} w^{p-1} e^{-w} dw \longrightarrow C_g\,,$$

as $p \to \infty$.

For the Burr distribution family for fixed r, it follows from (2) and Stirling's formula

$$\kappa \sim \alpha \left[\frac{r}{\Gamma(1/r)}\right]^r,$$

as $\alpha \to \infty$, from where

$$\overline{B}(x) = \left[1 + \frac{x^r}{\kappa}\right]^{-\alpha} \sim \left(1 + \frac{x^r}{\alpha}\left[\frac{1}{r}\Gamma\left(\frac{1}{r}\right)\right]^r\right)^{-\alpha} \to \exp\left\{-\left[\frac{1}{r}\Gamma\left(\frac{1}{r}\right)\right]^r x^r\right\},$$

as $\alpha \to \infty$. But

$$\left[\frac{1}{\tau}\,\Gamma\left(\frac{1}{\tau}\right)\right]^{\tau} x^{\tau} \to \begin{cases} 0, & 0 \le x < 1, \\ e^{-\gamma}, & x = 1, \\ \infty, & 1 < x < \infty, \end{cases}$$

with

$$\left[\frac{1}{\tau}\,\Gamma\left(\frac{1}{\tau}\right)\right]^{\tau} \sim \left[1 + \frac{1}{\tau}\,\Gamma'(1) + o\left(\frac{1}{\tau}\right)\right]^{\tau} \to e^{-\gamma}\,,$$

as $\tau \to \infty$, where $\gamma = 0.5772156649$ the Euler's constant, from where (1) follows with $C = \exp\left(-e^{-\gamma}\right)$

$$\lim_{\tau\to\infty}\lim_{\alpha\to\infty} \overline{B}(x) = \lim_{\tau\to\infty} \exp\left\{-\left[\frac{1}{\tau}\,\Gamma\left(\frac{1}{\tau}\right)\right]^{\tau} x^{\tau}\right\} = \begin{cases} 1, & 0 \le x < 1, \\ \exp\left(-e^{-\gamma}\right), & x = 1, \\ 0, & 1 < x < \infty, \end{cases} \tag{3}$$

and this convergence is uniform on $x \in [0, 1-\varepsilon]$ for any $\varepsilon \in (0,1)$.

Next we examine the Bektander II distribution. Again from the substitution $\alpha = 1$ it follows the limit

$$\overline{B}(x) = \frac{1}{(1+x)^{1-p}} \exp\left\{\frac{1-(1+x)^{p}}{p}\right\} \longrightarrow e^{-x},$$

as $p \to 1$.

Now, for the convergence in $\mathfrak{L}^1$, let us note that for any $\varepsilon > 0$ and all claim distributions B

$$\int_0^{\infty} \left|\overline{B}(y) - \overline{D}(y)\right| dy = 2\int_0^{1} \left[1 - \overline{B}(y)\right] dy \le 2\int_0^{1-\varepsilon} \left[1 - \overline{B}(y)\right] dy + 2\varepsilon\,. \tag{4}$$

This relation gives in the Pareto case

$$\int_0^{\infty} \left|\overline{B}(y) - \overline{D}(y)\right| dy = 2\int_{(\alpha-1)/\alpha}^{1} dy = \frac{2}{\alpha} \to 0,$$

as $\alpha \to \infty$.

In the Lognormal case, for any $\varepsilon > 0$ if we choose $\sigma_\varepsilon > 0$ such that

$$\int_0^{1-\varepsilon} \int_{-\infty}^{\frac{\ln y}{\sigma_\varepsilon} + \frac{\sigma_\varepsilon}{2}} \exp\left\{-\frac{u^2}{2}\right\} \frac{du}{\sqrt{2\pi}} dy < \varepsilon,$$

then for every $\sigma \in (0, \sigma_\varepsilon)$

$$\int_0^{\infty} \left|\overline{B}(y) - \overline{D}(y)\right| dy \le 2\int_0^{1-\varepsilon} \int_{-\infty}^{\frac{\ln y}{\sigma} + \frac{\sigma}{2}} \exp\left\{-\frac{u^2}{2}\right\} \frac{du}{\sqrt{2\pi}} dy + 2\varepsilon \le 4\varepsilon,$$

which gives the convergence in $\mathfrak{L}^1$ as $\sigma \to 0$.

We take the Weibull distribution. Let us see that for any $T > 1\,,\ \tau > 1/2$

$$\int_0^\infty \left|\overline{B}(y) - e^{-y}\right| dy = \int_0^\infty e^{-y} \left|\exp\left\{-\left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^\tau + y\right\} - 1\right| dy$$

$$\leq \left(\int_0^T + \int_T^\infty\right) \left|\exp\left\{-\left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^\tau + y\right\} - 1\right| dy$$

$$\leq T \sup_{[0,T]} \left|\exp\left\{-\left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^\tau + y\right\} - 1\right| + e^{-T}$$

$$+ \int_T^\infty \exp\left\{-\inf_{\tau\in[1/2,1]} \left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^{1/2}\right\} dy\,.$$

But $\forall \varepsilon > 0$ there exists T_ε such that

$$e^{-T_\varepsilon} + \int_{T_\varepsilon}^\infty \exp\left\{-\inf_{\tau\in[1/2,1]} \left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^{1/2}\right\} dy < \frac{\varepsilon}{2}\,.$$

As far as $y^\tau \to y$, uniformly on $[0, T_\varepsilon]$ and

$$\left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau \to 1\,,$$

as $\tau \to 1$, we can choose $\tau_\varepsilon \in (1/2, 1)$ such that $\forall \tau \in (\tau_\varepsilon, 1)$

$$\sup_{[0,T_\varepsilon]} \left|\exp\left\{-\left[\Gamma\left(1+\frac{1}{\tau}\right)\right]^\tau y^\tau + y\right\} - 1\right| < \frac{\varepsilon}{2T_\varepsilon}\,,$$

and therefore

$$\int_0^\infty \left|\overline{B}(y) - e^{-y}\right| dy < \varepsilon\,.$$

In the Loggamma case, it follows from the relation (4)

$$\int_0^\infty \left|\overline{B}(y) - \overline{D}(y)\right| dy \leq 2\int_0^{1-\varepsilon} \left[1 - \frac{1}{\Gamma(p)} \int_{A_p(y)}^\infty w^{p-1} e^{-w} dw\right] dy + 2\varepsilon \leq 4\varepsilon,$$

as $p \to \infty$. In the last inequality we used the uniform convergence on $[0, 1-\varepsilon]$ in (2).
Similarly, in the Burr distribution

$$\int_0^\infty \left|\overline{B}(y) - \overline{D}(y)\right| dy \leq 2\int_0^{1-\varepsilon} \left[1 - \overline{B}(y)\right] dy + 2\varepsilon \leq 4\varepsilon,$$

as $(\alpha, r) \to (\infty, \infty)$ or $\alpha \to \infty$, because of the uniform convergence on $[0, 1-\varepsilon]$ in (3).

We examine the Bektander II distribution. Let us notice that for any $T > 0$ and $p > 1/2$

$$\int_0^\infty \left|\overline{B}(y) - e^{-y}\right| dy \le \int_T^\infty \left|e^{-y}\right| dy + \int_T^\infty \left|\overline{B}(y)\right| dy + T \sup_{0\le y\le T} \left|\overline{B}(y) - e^{-y}\right|$$

$$\begin{aligned}
&\le \quad e^{-T} + \exp\left\{\frac{1-(1+T)^p}{p}\right\} \\
&+ \quad T\left[e^2\left(1 - \frac{1}{(1+T)^{1-p}}\right) + \sup_{0\le y\le T}\left[-1 + \exp\left\{\frac{1-(1+y)^p + py}{p}\right\}\right]\right] \\
&\le \quad e^{-T} + \exp\left\{1-(1+T)^{1/2}\right\} \\
&+ \quad T\left[e^2\left(1 - \frac{1}{(1+T)^{1-p}}\right) - 1 + \exp\left\{2\left[1-(1+T)^p + pT\right]\right\}\right],
\end{aligned}$$

where for any $\varepsilon > 0$, T_ε can be chosen such that the first two terms in the last sum are less than $\varepsilon/2$ and there exists a $p_\varepsilon > 1/2$ such that the last term in the sum become less than ε for any $p \in (p_\varepsilon, 1)$. Thus

$$\int_0^\infty \left|\overline{B}(y) - e^{-y}\right| dy \le 2\varepsilon,$$

and the convergence in $\mathcal{L}^1$ is proved. □

The stability theory with respect to $G|G|1|\infty$ systems, renders the following picture: If we have as input characteristics the distribution function of the service times (claims distribution function for our risk model) and as output characteristics the stationary distribution of the waiting times (ruin probability for our risk model), the stability means that the convergence in $\mathcal{L}^1$ of the input characteristics implies the weak convergence of the output characteristics. If in our cases the stationary distribution of waiting time is continuous, then the weak convergence is equivalent to point convergence.

Theorem 4.2. (Kalashnikov [5]) *In the classical risk model, if the claim size distribution $B(x)$ belongs to one of the following distribution families: Pareto, Lognormal, Loggamma, Burr, then the ruin probability tends to limit waiting time distribution tail in the $M/D/1$ queuing model (see [11, Th. 2.17], [1, Cor. III.3.6], [4])*

$$\psi(u) \to 1 - \left(1 - \frac{\rho}{c}\right) \sum_{n=0}^{[u]} \frac{1}{n!}\left[-\frac{\rho}{c}(u-n)\right]^n e^{-\frac{\rho}{c}(u-n)}, \quad 0 \le u < \infty,$$

as the corresponding parameter of heaviness reaches its limit ($\alpha \to \infty$, $\sigma \to 0$, $p \to \infty$ and $(\alpha, r) \to (\infty,\infty)$ respectively).

If $B(x)$ belongs to Weibull or Bektander II distribution family, the ruin probability tends to the $M/M/1$ waiting time distribution:

$$\psi(u) \longrightarrow \frac{\rho}{c} \exp\left\{-\left(1 - \frac{\rho}{c}\right) u\right\},$$

as $\tau \to 1$ and $p \to 1$ respectively.

Proof. The proof in the first case comes from a result in the stability theory of the Lindley chain as appears in [5, Th.V.5.5] (see also [9, Th.2] or [7, Th.1]). Here the classical risk model is described as queuing system $G/G/1/\infty$. The conditions of this theorem are satisfied through the appropriate choice of the test function in this chain.

For Weibull or Bektander II distributions, the ruin probability convergence follows again from the result of the stability theory [5, Th.5.3.1] (see also [9, Th.2], [7, Th.1] or [8]). The limit of the ruin probability corresponds to the waiting time distribution in $M/M/1$ queuing system (see for example [11, Th.1.15]). $\square$

Remark 4.3. We see that if $B(x)$ belongs to the Bektander I distribution family, then its tail tends to a limit equal to zero:

$$\overline{B}(x) \to 0\,,$$

as $p \to \infty$.

Namely, let us make the substitution $\alpha = 1$. Then we take

$$\overline{B}(x) = \frac{1 + 2p\ln(1+x)}{(1+x)^2}\exp[-p\ln^2(1+x)] \longrightarrow 0,$$

as $p \to \infty$.

Hence in classical risk model, if the claim size distribution $B(x)$ belongs to one of the Bektander I distribution family, the ruin probability tends to $\frac{\rho}{c}$:

$$\psi(u) \to \frac{\rho}{c},$$

as $p \to \infty$.

Indeed, for the Bektander I distribution we see that for any $\varepsilon \in (0,u)$ there is a constant $p_0 > 1$ such that

$$\overline{B}(\varepsilon) \le \frac{\varepsilon}{u-\varepsilon},$$

for any $p > p_0$. So, from Pollaczeck-Khinchine formula (2) the following chain of inequalities can be taken:

$$\begin{aligned}\frac{\rho}{c} &= \psi(0) \ge \psi(u) \ge \frac{\rho}{c}\overline{F}(u) = \frac{\rho}{c}\left(1 - \int_0^{\varepsilon}\overline{B}(y)dy - \int_{\varepsilon}^{u}\overline{B}(y)dy\right)\\ &\ge \frac{\rho}{c}[1-\varepsilon-(u-\varepsilon)\overline{B}(\varepsilon)] \ge \frac{\rho}{c}[1-2\varepsilon],\end{aligned}$$

for any $p > p_0$.

We see in this remark that the lighter limit of the Bektander I claim distribution does not represent a distribution and the lighter limit of the ruin probability is not a decreasing function with respect to u. These deformations of the standard properties of the distribution function and the ruin probability expresses a tail explosion through the convergence to the limit. As we have done in the previous section we proceed to a functional normalization. Namely we take a functional heavytailedness parameter $p(u) > 1\,,\ \ \forall u \ge 0$, such that $p(u) \to \infty$ as $u \to \infty$.

Theorem 4.4. *If $B(x)$ belongs to the Bektander I distribution family with its heavytailedness parameter $p(x) \to \infty$, then the normalized tail tend to the following limit:*

$$\overline{B}\left(\exp\left\{\sqrt{\frac{x}{p(x)}}\right\} - 1\right) \sim (1 + 2\sqrt{x\,p(x)})\exp\left\{-x - 2\sqrt{\frac{x}{p(x)}}\right\},$$

as $x \to \infty$. Further, in the classical risk model, the ruin probability tends to the following limit:

$$\psi\left(\exp\left\{\sqrt{\frac{u}{p(u)}}\right\} - 1\right) \sim \frac{\rho}{c-\rho}\int_u^\infty \frac{\sqrt{z\,p(z)} + 2\,z\,p(z)}{2\,p(z)}\left(\frac{1}{z} - \frac{p'(z)}{p(z)}\right)\exp\left\{-z - \sqrt{\frac{z}{p(z)}}\right\}dz\,.$$

as $u \to \infty$.

Proof. Indeed, again from the relation (2) and the property of the subexponentiality we find

$$\psi(u) \sim \frac{\rho}{c-\rho}\overline{F}(u) = \frac{\rho}{c-\rho}\int_u^\infty \overline{B}(y)dy\,.$$

Now we look for a normalizing function in the form $f(x, p(x))$, for which the expression $1 - \overline{B}(x\,f(x, p(x)))$ remains distribution after the passage to the limit:

$$\begin{aligned}\psi(u\,f(u,p(u))) &\sim \frac{\rho}{c-\rho}\int_{u\,f(u,p(u))}^\infty \overline{B}(y)\,dy \\ &\sim \frac{\rho}{c-\rho}\int_u^\infty \overline{B}(z\,f(z,p(z)))\left(\frac{d[f(z,p(z))]}{dz}z + f(z,p(z))\right)dz\,.\end{aligned}$$

Namely, we find the following normalizing function:

$$f(x, p(x)) = \frac{1}{x}\left(\exp\left\{\sqrt{\frac{x}{p(x)}}\right\} - 1\right).$$

And after the substitution we reach the ruin probability

$$\psi\left(\exp\left\{\sqrt{\frac{u}{p(u)}}\right\} - 1\right)$$

$$\sim \frac{\rho}{c-\rho}\int_u^\infty (1 + 2\sqrt{z\,p(z)})\exp\left\{-z - 2\sqrt{\frac{z}{p(z)}}\right\}\frac{1}{2\sqrt{p(z)}}\left(\frac{1}{\sqrt{z}} - \frac{p'(z)\sqrt{z}}{p(z)}\right)\exp\left\{\sqrt{\frac{z}{p(z)}}\right\}dz\,.$$

from where we find our result. □

References

[1] ASMUSSEN, S. 2000 *Ruin Probabilities.* World Scientific, Singapore.

[2] EMBRECHTS, P., KLUEPPELBERG, C., MIKOSCH, T. (1997) *Modelling Extremal Events for Finance and Insurance.* Springer-Verlag, Heildelberg.

[3] EMBRECHTS, P., VERAVERBEKE, N. (1982) Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Math. Econom.*, **1**, 55–72.

[4] ERLANG, A.K. 1909 Sandsynlighedsregning og telefonsamtaler. *Nyt Tidsskrift for Matematik.*, **B20** , 33–39.

[5] KALASHNIKOV, V.V. 1978 *Qualitative Analysis of the Behavior of Complex Systems by Test Functions Method.* Nauka, Moscow. (in Russian)

[6] KALASHNIKOV, V.V. 1997 *Geometric Sums: Bounds for Rare Events with Applications.* Kluwer, Dordrecht.

[7] KALASHNIKOV, V.V. 2000 The stability concept for stochastic risk models. *Working paper No. 166, Lab. of Actuar. Math.* , University of Copenhagen.

[8] KALASHNIKOV, V.V. 2001 Quantification in stochastics and the stability concept. *Lecture Notes*, University of Copenhagen.

[9] KALASHNIKOV, V.V., AND TSITSIASHVILI, G.SH. 1973 On the stability of queueing systems with respect to disturbances of their distribution functions. *Engineering Cybernetics.*, **10**, 211–217.

[10] KONSTANTINIDIS, D.G. 1999 Comparison of Ruin Probability Estimates in the Presence of Heavy Tails. *Journal Mathem. Scienc. - New York*, **93**, 552-562.

[11] PRABHU, N.U. 1998 *Stochastic Storage Processes.* Springer, New York.

[12] TSITSIASHVILI, G.SH., AND KONSTANTINIDES, D.G. 2001 Supertails in Risk Theory. *Far Eastern Mathem. J.*, **2**, 68–76.

Strong invariance principle for renewal and randomly stopped processes with applications to risk models

Nadiia Zinchenko
Department of Probability Theory & Mathematical Statistics
Kyiv National Taras Shevchenko
University, Kyiv, Ukraine
Email: znm@univ.kiev.ua

INTRODUCTION

Let $\{X, X_i, i \geq 1\}$ be i.i.d.r.v with common distribution function (d.f.) $F(x)$ and characteristic function (ch.f.) $\varphi(t)$, $EX = m$ if $E|X| < \infty$, $Var(X) = 1$ if $E|X|^2 < \infty$.

Assume that $\{Z, Z_i, i \geq 1\}$ is another sequence of i.i.d.r.v. independent of $\{X_i, i \geq 1\}$ with d.f. $F_1(x)$ and ch.f. $\varphi_1(t)$, $EZ = 1/\lambda$, $0 < \lambda < \infty$ if $E|Z| < \infty$, $Var(Z) = \tau^2$ if $E|Z|^2 < \infty$. Put

$$S(n) = \sum_{i=1}^{n} X_i, \quad Z(n) = \sum_{i=1}^{n} Z_i, \tag{1}$$

where $S(0) = 0$, $S(x) = S([x])$, $Z(0) = 0$, $Z(x) = Z([x])$ and $[a]$ is entire of $a > 0$.

Define the renewal counting process as

$$N(t) = \inf\{x \geq 0 : Z(x) > t\} \tag{2}$$

and consider the randomly stopped sum process (i.e. the superposition of random processes $S(n)$ and $N(t)$)

$$D(t) = S(N(t)) = \sum_{i=1}^{N(t)} X_i, \tag{3}$$

where renewal process $N(t)$ is defined by (2).

The main task of this paper is to study the asymptotic behavior of the random processes $D(t)$ and $N(t)$ under various assumptions on $F(x)$ and $F_1(x)$ with emphasis on the heavy-tailed cases. This problem has a deep relation with investigations of asymptotics of risk process $U(T)$ in Sparre Anderssen collective risk model

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \tag{4}$$

where: $u \geq 0$ denotes the initial capital; $c > 0$ stands for the premium income rate; i.i.d.r.v $\{X_i, i \geq 1\}$ are interpreted as claim sizes; $N(t)$ describes the claim arrival process and stands for the number of claims until time t; $\{Z_i, i \geq 1\}$ being the inter-arrival times.

In such model $S(N(t))$ is interpreted as total claim amount process and is a stochastic part of risk process.

Limit theorems for risk process such as (weak) invariance principle, which constitute the weak convergence of $U(t)$ to the Wiener process $W(t)$ with the drift (when $EX^2 < \infty$, $EZ^2 < \infty$) or to the α-stable Lévy process $Y_\alpha(t)$, lead to useful approximations of the ruin probability

$$\psi(u) = P\{\inf_{t>0} U(t) < 0\}. \tag{5}$$

Thus, in the case $EX^2 < \infty$, $EZ^2 < \infty$ one obtains the "diffusion approximation" for $\psi(u)$ as a distribution of infimum of the Wiener process (Iglehart (1969), Grandell (1991)), in the case $EX^2 = \infty$, $EZ^2 < \infty$ the ruin probability $\psi(u)$ is approximated by the distribution of infimum of the corresponding α-stable process (Furrur, Michna and Weron (1997), Furrur (1998)).

Here we deal with the other type of limit theorems, so called strong invariance principle, which is in certain sense a bridge between weak and strong convergence.

1. Strong invariance principle for the partial sums

Strong invariance principle (almost sure approximation) is an umbrella name for the class of limit theorems which ensure the possibility to construct r.v. $\{X_i, i \geq 1\}$ and Lévy process $\{Y(t), t \geq 0\}$ on the same probability space in such a way that with probability 1

$$|S(t) - mt - Y(t)| = o(r(t)) \tag{6}$$

or

$$|S(t) - mt - Y(t)| = O(r(t)) \tag{7}$$

were approximation error (rate) $r(.)$ is a non-random function depending only on assumptions posed on X.

Additional assumptions on X clear up the type of $Y(t)$ and the form of $r(.)$. Since we deal with i.i.d.r.v. it is natural to consider Wiener process $W(t)$ or α-stable Lévy process $Y_\alpha(t)$, $t \geq 0$ as an approximation process $Y(t)$ in (6), (7). If $EX^2 < \infty$ then $Y(t) = W(t)$ is a **Wiener process**.

Note that the complete solution of the problem of a.s.approximation depends not only on the distribution of r.v. $\{X_i, i \geq 1\}$ but also on a structure of the probability space, and (possibly) requires a "richer" probability space and equivalent r.v.$\{X_i^{'}, i \geq 1\}$, but for brevity we do not distinguish between r.v.$\{X_i\}$ and$\{X_i^{'}\}$ as well as their sums.

The origin of this topic goes back to the famous "Skorokhod representation" and "Skorokhod embedding scheme" (A.V. Skorokhod (1961)). Skorokhod representation allows one to study a sequence of values of the Wiener process $W(T_n)$, where T_n, $n \geq 1$, are some stopping times, instead of partial sums $S(n)$.

Based on Skorokhod embedding scheme V. Strassen (1964,1965) proved the first variant of the strong invariance principle.

In 1970 – 1995 the further investigations were carried out by a number of authors, among them Kiefer, M.Csörgő, Révész, Komlós, Major, Tusnady, Berkes, Horváth (quantile Hungerian method), Stout, Phillip, Berkes (relationship between the strong invariance principle and convergence in Prokho rov metrics), Horváth (inverse processes)

For wide references see M.Csörgő and P.Révész (1981); M. Csörgő and L.Horváth (1993), N. Zinchenko (2000).

It is obvious that using (6) with appropriate error term one can easily (almost without the proof) transfer the results about the asymptotic behavior of Lévy process $Y(t)$ or its increments on the rate of growth of partial sums and corresponding increments.

Following theorem [5,6] will serve as a good background for further investigations.

Theorem 1. It is possible to define partial sum process $\{S(t), t \geq 0\}$ and a standard Wiener process $\{W(t), t \geq 0\}$ in such a way that a.s.

$$|S(t) - mt - W(t)| = o(r(t)), \tag{8}$$

where:

- $r(T) = T^{1/p}$ iff $E\,|X|^p < \infty, p > 2$;
- $r(T) = (T \ln \ln T)^{1/2}$ iff $E\,|X|^2 < \infty$;
- right hand side of (8) is $O(\ln T)$ iff $E \exp(uX) < \infty$ for $u \in (0, u_0)$.

We shall also use concept of *a.s.approximation in a wider sense* and say, that a random process $\xi(t)$ admits the a.s. approximation by the random process $\eta(t)$, if $\xi(t)$ (or stochastically equivalent $\xi^{'}(t)$) can be constructed on the rich enough probability space together with $\eta(t), t \geq 0$, in such a way that a.s.

$$|\xi(t) - \eta(t)| = o(r_1(t)) \vee O(r_1(t)),$$

where $r_1(.)$ is again a non-random function.

2. Strong invariance principle for the sums of r.v. attracted to the stable law

Suppose that r.v. $\{X, X_i, i \geq 1\}$ are in domain of normal attraction of a stable law with $0 < \alpha < 2$, $|\beta| \leq 1$, notation $\{X_i, i \geq 1\} \in DNA(G_{\alpha,\beta})$. This means that

$$n^{-1/\alpha}\big(S(n) - a_n\big) \Rightarrow G_{\alpha,\beta}$$

where $a_n = nEX = mn$ if $1 < \alpha < 2$, $a_n = 0$ if $0 < \alpha < 1$ and $a_n = (2/\pi)\beta \log n$ if $\alpha = 1$.

Here $G_{\alpha,\beta}$ is a stable law with parameters $0 < \alpha < 2$, $|\beta| \leq 1$ and ch.f. $g_{\alpha,\beta}(u) = \exp(K(u))$

$$K(u) = K_{\alpha,\beta}(u) = -|u|\big(1 - i\beta(u/|u|)\varpi(u, \alpha)\big), \tag{9}$$

where $\varpi(u, \alpha) = \tan(\pi\alpha/2)$ if $0 < \alpha < 2$,$\alpha \neq 1$, $\varpi(u, \alpha) = -(2/\pi) \log |u|$ if $\alpha = 1$.

Now approximating process $Y(t) = Y_\alpha(t) = Y_{\alpha,\beta}(t)$, $t \geq 0$, is **stable Lévy process** with ch.f.

$$g_\alpha(t;u) = g_{\alpha,\beta}(t;u) = \exp(tK_{\alpha,\beta}(u)).$$

It occurs that the fact $\{X, X_i\} \in DNA(G_{\alpha,\beta})$ is not enough to obtain "good" error term in (6), (7), thus, certain additional assumptions are needed. We formulate them in terms of ch.f. (see Zinchenko [22, 24 - 26], Berkes et al. [2, 3], Mijnheer [16].)

Assumption (C) : *there are* $a_1 > 0, a_2 > 0$ *and* $l > \alpha$ *such that for* $|u| < a_1$

$$|f(u) - g_{\alpha,\beta}(u)| < a_2|u|^l$$

where $f(u)$ *is a ch.f. of* $(X - EX)$ *if* $1 < \alpha < 2$ *and ch.f. of* X *if* $0 < \alpha \leq 1$.

Theorem 2 [22, 24]. Put $m = EX$ for $1 < \alpha < 2$ and $m = 0$ for $0 < \alpha \leq 1$. Under assumption (C) a.s.

$$\sup_{0 \leq t \leq T} \left| S([t]) - mt - Y_{\alpha,\beta}(t) \right| = O(T^{1/\alpha - \rho_0}), \tag{10}$$

where

$$\rho_0 = \min\left(\frac{l - \alpha}{80\alpha}, \frac{2 - \alpha}{2\alpha} \right).$$

3. Strong invariance principle for counting(renewal) processes

Let $N(t)$ be counting renewal process associated with partial sums $Z(n)$. For applications it is convenient to suppose that Z_i are non-negative (non-zero) r.v.

3.a. Assumptions: $EZ^2 < \infty$**,** $EZ = 1/\lambda > 0$.

In the case $\tau^2 = varZ < \infty$ Csörgő, Horvách, Steinebach, Aalex, Deheuvels, Mason, van Zwet, see [1, 3, 4], studied a.s. approximation of the type

$$|\lambda t - N(t) - \tau\lambda^{3/2}W(t)| = o(r(t)) \vee O(r(t)) \tag{12}$$

They proved that conditions, which provide (12) and corresponding optimal errors, are the same as for $S(n)$, see Theorem 1.

3.b. Random variables $\{Z_i, i \geq 1\} \in DNA(G_{\alpha,\beta})$.

In this case we proved

Theorem 3. If $\{Z_i\}$ satisfy (C) with $1 < \alpha < 2$, then a.s.

$$\left|t\lambda - N(t) - \lambda^{1+1/\alpha} Y_{\alpha,\beta}(t)\right| = o(r_1(t)), \tag{13}$$

where $r_1(t)$ is any upper function for α-stable Levy process, for instance, $r_1(t) = t^{1/\alpha+\delta}$ for any $\delta > 0$.

4. Strong invariance principle for randomly stopped processes

Let $S(n)$, $N(t)$, $D(t) = S(N(t)) = \sum_{i=1}^{N(t)} X_i$ be as in Introduction, $Ez = 1/\lambda > 0$, $EX = m$.

Combining the results about strong invariance principle for partial sums and renewal processes it is possible to obtain a number of results about a.s. approximation of the superposition of the mentioned processes, i.e. the random sums $D(t)$. Weak convergence of the superposition of the processes was in details studied by W. Whitt (2002), D. Silvestrov (2004).

We start with the case when both $\{X, X_i, i \geq 1\}$ and $\{Z, Z_i, i \geq 1\}$ have finite moments of order grater than 2. Next theorem concerning a.s. approximation of the superposition of the random process (not obligatory connected with the partial sums) follows from results due to Csörgő and Horvách (1993).

Let $Z^*(t)$, $S^*(t)$ be two real-valued random processes, N^* – the inverse of $Z(t)$ – is defined by

$$N^*(t) = \inf\{x > 0 : Z(x) > t\}, 0 \leq t < \infty.$$

Theorem 4. Suppose that for some constants $m, \lambda > 0, \sigma > 0, \tau > 0$ a.s.

$$\sup_{0 \leq t \leq T} \left|\tau^{-1}(Z^*(t) - t/\lambda) - W_1(t)\right| = O(r(T)),$$

$$\sup_{0 \leq t \leq T} \left|\sigma^{-1}(S^*(t) - mt) - W_2(t)\right| = O(q(T)),$$

where $W_1(t)$ and $W_2(t)$ are independent Wiener processes, $q(t) \uparrow \infty$, $q(t)/t \downarrow 0$, $r(t) \uparrow \infty$ and $r(t)/t \downarrow 0$ as $t \to \infty$, then a.s.

$$\begin{gathered}\left|S^*(N^*(t)) - (m\lambda)t - \nu W(t)\right| = \\ = O\big(q(t) + r(T) + \log t\big), \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2,\end{gathered} \tag{14}$$

where $W(t)$ is a Wiener process.

Corollary 1. Let $\{X_i, i \geq 1\}$ be i.i.d.r.v., $varX_1 = \sigma^2 < \infty$, i.i.d.r.v. $\{Z_i, i \geq 1\}$ be independent of $\{X_i\}$, $0 < EZ_1 < 1/\lambda$, $\tau^2 = varZ_1 < \infty$. Mentioned r.v. can be constructed on the same probability space together with Wiener process $W(t)$ in such a way that a.s.

$$\sup_{0 \leq t \leq T} |S(N(t)) - \lambda mt - \nu W(t)| = o(r_3(T)), \tag{15}$$

$$\nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2,$$

where:

- $r_3(T) = T^{1/p}$, if $E\,|X_1|^p < \infty$, $E\,|Z_1|^p < \infty$, $p > 2$;

- $r_3(T) = T^{1/p}$, $p = \min\{p_1, p_2\}$, if $E|X_1|^{p_1} < \infty$, $E|Z_1|^{p_2} < \infty$, $p_1 > 2$, $p_2 > 2$;

- $r_3(T) = (T \ln\ln T)^{1/2}$, if $p = 2$;

- right side of (15) is $O(\ln T)$, if $E\exp(uX_1) < \infty$ and $E\exp(uZ_1) < \infty$ for all $u \in (0, u_o)$.

Developing the ideas of Csörgő and Horvách , Zinchenko (2007) proved following result

Theorem 5. Let $Z^*(t)$, $S^*(t)$ be two real-valued random processes, N^* - the inverse of $Z(t)$. Suppose that for some constants $m, \lambda > 0, \tau > 0$ a.s.

$$\sup_{0 \le t \le T} \left|\tau^{-1}(Z(t) - t/\lambda) - W_3(t)\right| = O(r(T)),$$

where $W_3(t)$ is a Wiener process, $r(t) \uparrow \infty$, $r(t)/t \downarrow 0$ as $t \to \infty$ and

$$\sup_{0 \le t \le T} \left|D(t) - mt - Y_\alpha(t)\right| = o(q(T)),$$

where $Y_\alpha(t)$ is α-stable process independent of $W_3(t)$, $q(t) \uparrow \infty$, $q(t)/t \downarrow 0$ as $t \to \infty$, then $\forall \varepsilon > 0$ a.s.

$$\left|S(N^*(t)) - (m\lambda)t - (Y_\alpha(\lambda t) - (m\lambda\tau)W_4(\lambda t))\right| =$$
$$= o(q(t)) + O\big(r(t) + \log t\big)$$
$$+ O\big((r(t) + (t \log\log t)^{1/2})^{1/(\alpha-\varepsilon)}\big),$$

where $W_4(t)$ is a Wiener process independent of $Y_\alpha(t)$.

In the case of partial sum processes, when $N^*(t) = N(t)$ is counting process, $EZ^2 < \infty$, X_i satisfying (C), $q(t) = t^{1/\alpha - \varrho}$, $\varrho > 0$, the worst estimate for $r(t)$ is $(t \log\log t)^{1/2}$. These facts lead to following

Theorem 6. Let $\{X_i, i \ge 1\}$ satisfy (C) with $1 < \alpha < 2$, and $EZ^2 < \infty$ then a.s.

$$\left|S(N(t)) - m\lambda t - Y_{\alpha,\beta}(\lambda t)\right| = o(t^{1/\alpha - \varrho_1})$$

for some $\varrho_1 = \varrho_1(\alpha, l) > 0$.

The same approach provides

Theorem 7. Let $\{X_i, i \ge 1\}$ satisfy (C) with $1 < \alpha_1 < 2$, and $\{Z_i\}$ satisfy (C) with $1 < \alpha_2 < 2$,

$$\alpha_1 \le \alpha_2$$

then a.s.

$$\left|D(t) - m\lambda t - Y_{\alpha_1,\beta}(\lambda t)\right| = o(t^{1/\alpha_1 - \varrho_2})$$

for some $\varrho_2 = \varrho_2(\alpha_1, l) > 0$.

5. Application to risk models

In the framework of collective risk model

$$D(T) = \sum_{i=1}^{N(T)} X_i = S(N(T))$$

can be interpreted as a total claim amount arising during time interval $[0, t]$, and increments

$$D(T + a_T) - D(T) = \sum_{i=N(T)+1}^{N(T+a_T)} X_i$$

as claim amounts during the time interval $[T, T + a_T]$.

In Section 4 we obtained a number of strong invariance type results for a total claim amount process $D(t)$ under various moment assumptions on claims and inter-arrival times (Theorems 4 – 7, Corollary 1). Now we shall use them for more detail investigation of asymptotics of $D(t)$ and its increments.

We shall try to give the answers on the questions:

1) What is a rate of growth of total claims when $T \to \infty$?

2) How large can be the fluctuations of the total claims/payments on the intervals whose length a_T increases, but slower than T ?

Note that question about the order of magnitude of the total amount claim process was asked in Embrechts et al. [7, section 8.5].

There is numerous results concerning asymptotic behavior of increments of the random work $S(n)$, Wiener and related processes and certain types of stable Lévy processes. We refer to [5, 6, 8, 9, 14, 15, 23] and references there.

Two approaches are applied to investigate the size of increments of partial sums. One of them is rather straightforward, it uses the properties of d.f. $F(x)$ and is based on large deviation technique; the other one - exploits strong invariance principle for partial sums.

The task of the forthcoming sections is investigation the rate of growth and magnitude of increments of random sums $D(t)$.

We shall prove certain modifications of the LIL for $D(t)$ as $t \to \infty$ under various moment assumption on $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ and study the order of magnitude of the increments $D(T + a_T) - D(T)$ over the intervals, whose length a_t grows, but not faster than T.

The main tool of investigation – application of strong invariance principle for $D(t)$ in combination with Erdös-Rényi type limit theorems for $W(t)$ or $Y_\alpha(t)$.

We shall consider separately two situations : *small* claims and *large* claims. In the actuarial mathematics individual claim sizes are usually divided in two classes, i.e. small claims and large claims, according to the tail behavior of their distribution function $F(x)$.

Claims are called *small* if $F(x)$ is light-tailed satisfying Cramér's condition

$$M(u) = E\exp(uX_1) < \infty \text{ for } u \in (0, u_0);$$

in opposite case, when moment generating function does not exist for any $u > 0$, the claims are called *large* ($F(x)$ is heavy-tailed).

But in our set up we shall distinguish two sub-classes in the class of large claims: first one includes r.v. which have finite moments of order $p > 2$, while second sub-class consists of positive r.v. in domain of normal attraction of the stable law $G_{\alpha,1}$ with $1 < \alpha < 2$, $\beta = 1$; in this case r.v. $X_i^* = -X_i$ and their sums are attracted to a stable law $G_{\alpha,-1}$ with the same $1 < \alpha < 2$ and $\beta = -1$.

It is natural to assume that inter-arrival times Z_i have distributions with tails not heavier than claims.

6. The rate of growth of random sums

In this Section we study the asymptotic behavior of $D(t)$ as $t \to \infty$ and prove generalizations of the LIL under various moment assumptions on $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$.

6.1. All random variables have finite variance.

Theorem 8 (Classical LIL for random sums). Let $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ be independent sequences of i.i.d.r.v. with $EX_1 = m < \infty$, $0 < EZ_1 = 1/\lambda < \infty$, $\sigma^2 = VarX_1 < \infty$, $\tau^2 = VarZ_1 < \infty$. Then a.s.

$$\limsup_{t\to\infty} \frac{|D(t) - m\lambda t|}{\sqrt{2t\ln\ln t}} = \nu, \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2\tau^2.$$

5.2. $\{X_i, i \geq 1\}$ are attracted to the stable law $G_{\alpha,-1}$, but $EZ_1^2 < \infty$.

When $\beta = -1$, the process $Y_\alpha(t) = Y_{\alpha,-1}(t)$ has only negative jumps. We omit β in this case.

In general it is impossible to find the exact upper bound for the growth rate of the stable process, but it can be done in particular (important for applications) mentioned case of stable processes without positive jumps ($\beta = -1$), see Zinchenko [23].

Theorem 9. Let $\{X_i, i \geq 1\}$ satisfy condition **(C)** with $1 < \alpha < 2$, $\beta = -1$ and $EZ_1^2 < \infty$. Then a.s.

$$\limsup_{t\to\infty} \frac{D(t) - m\lambda t}{t^{1/\alpha}(B^{-1}\ln\ln t)^{1/\theta}} = \lambda^{1/\alpha},$$

where $B = B(\alpha) = (\alpha-1)\alpha^{-\theta}|\cos(\pi\alpha/2)|^{1/(\alpha-1)}$, $\theta = \alpha/(\alpha-1)$.

Corollary 2. Theorems 8 and 9 are true when $N(t)$ is a homogeneous Poisson process.

6.3. $\{X_i\} \in DNA(G_{\alpha_1,-1})$ and $\{Z_i\} \in DNA(G_{\alpha_2,\beta})$ with $1 < \alpha_1 < \alpha_2 < 2$.

Theorem 10. Let $\{X_i, i \geq 1\}$ satisfy **(C)** with $1 < \alpha_1 < 2$, $\beta = -1$, and $\{Z_i, i \geq 1\}$ satisfy **(C)** with $1 < \alpha_2 < 2$, $|\beta| \leq 1$. Suppose that $\alpha_1 < \alpha_2$. Then a.s.

$$\limsup_{t\to\infty} \frac{D(t) - m\lambda t}{t^{1/\alpha_1}(B_1^{-1}\ln\ln t)^{1/\theta_1}} = \lambda^{1/\alpha_1},$$

$$B_1 = B(\alpha_1) = (\alpha_1 - 1)\alpha_1^{-\theta_1}|\cos(\pi\alpha_1/2)|^{1/(\alpha_1-1)},$$

$\theta_1 = \alpha_1/(\alpha_1 - 1)$.

7. Erdös-Rényi-Csörgő-Révész type SLLN for random sums

Here we study the magnitude of increments total claim amount process, i.e. $D(T + a_T) - D(T)$. Similar to Section 6 we explore this topic step by step.

7.1. $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ have finite variance.

In this case centered process $D(t) - m\lambda t$ can be a.s. approximated by the Wiener process with appropriate error term, whose form depends on additional moment conditions. It gives the possibility to extend the Erdös-Rényi (1970), Csörgő-Révész (1981) results about the asymptotic behavior of the increments of Wiener process on asymptotics of $D(T + a_T) - D(T)$. Notice that additional assumptions which determine the form of approximation term have impact on the length of intervals a_T, which appear in next theorems.

Theorem 11. (*Small claims, light-tailed inter-arrival times*). Let $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ be independent sequences of i.i.d.r.v., $EX_1 = m$, $var X_1 = \sigma^2$, $EZ_1 = 1/\lambda > 0$, $var Z_1 = \tau^2$,

$$E\exp(uX_1) < \infty, \quad E\exp(uZ_1) < \infty, \tag{16}$$

as $|u| < u_0$, $u_0 > 0$, function $a_T, T \geq 0$ satisfies following conditions: $0 < a_T < T$ and T/a_T does not decrease in T. Also

$$a_T/\ln T \to \infty \text{ as } T \to \infty. \tag{17}$$

Then a.s.

$$\limsup_{T\to\infty} \frac{|D(T + a_T) - D(T) - m\lambda a_T|}{\gamma(T)} = \nu, \tag{18}$$

where

$$\nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2,$$

$$\gamma(T) = \{2a_T(\ln\ln T + \ln T/a_T)\}^{1/2}.$$

Theorem 12. Let $\{X_i, i \geq 1\}$, $\{Z_i, i \geq 1\}$ and a_T satisfy all conditions of Theorem 11 with following assumption used instead of (16)

$$EX_1^{p_1} < \infty, \quad p_1 > 2, \quad EZ_1^{p_2} < \infty, \quad p_2 > 2.$$

Then (18) is true if $a_T > c_1 T^{2/p}/\ln T$ for some $c_1 > 0$, $p = \min\{p_1, p_2\}$.

7.2. On the second step we assume that i.i.d.r.v. $\{X_i, i \geq 1\}$ are attracted to an asymmetric stable law.

In this case we use Theorem 6 or Theorem 7 and variant of Erdös-Rényi-Csörgő-Révész type law for α-stable Lévy process without positive jumps (Zinchenko (1987)).

Theorem 13. Suppose that $\{X_i, i \geq 1\}$ satisfy **(C)** with $1 < \alpha_1 < 2$, $\beta = -1$, $\{Z_i, i \geq 1\}$ satisfy **(C)** with $1 < \alpha_2 < 2$, $\alpha_1 < \alpha_2$ or $EZ_1^2 < \infty$, $EX_1 = m$, $EZ_1 = 1/\lambda > 0$. Function a_T is non-decreasing, $0 < a_T < T$, T/a_T is also non-decreasing and provides ${d_T}^{-1}T^{1/\alpha-\varrho_2} \to 0$ for certain $\varrho_2 > 0$ determined by error term in strong invariance principle. Then a.s.

$$\limsup_{T\to\infty} \frac{D(T+a_T) - D(T) - m\lambda a_T}{d_T} = \lambda^{1/\alpha_1}, \tag{19}$$

where normalizing function d_T is

$$d_T = a_T^{1/\alpha_1}\{B^{-1}(\ln\ln T + \ln T/a_T)\}^{1/\theta_1},$$

$$B_1 = B(\alpha_1) = (1-\alpha_1)\alpha_1^{-\theta_1} \left|\cos(\pi\alpha_1/2)\right|^{1/(\alpha_1-1)},$$

$$\theta_1 = \alpha_1/(\alpha_1 - 1).$$

8. Conclusions

Small claims. In this case centered process $D(t) - m\lambda t$ satisfies two-sided classical LIL (Theorem 8). So, for large t we can a.s. indicate upper/lower bounds for growth of total claim amount claim process $D(t)$ as $m\lambda t \pm \nu\sqrt{2t\ln\ln t}$.

Asymptotic behavior of the increments $D(T + a_T) - D(T)$ for small claims is described by generalization of Erdös-Rényi-Csörgő-Révész type limit theorem (Theorem 11), which can be applied to increments over intervals whose length $a_T < T$, but $a_T/\ln T \to \infty$ as $T \to \infty$.

Large claims. In the case, when claims do not have finite exponential moments, but posses finite moments of order $p > 2$, process $D(t)$ also satisfies LIL and Csörgő-Révész type limit theorem, but with more restrictive conditions on intervals' length $a_T > cT^{2/p}/\ln T$ for some $c > 0$ (Theorem 12).

When $EX_1^2 = \infty$, we consider the positive r.v. in domain of normal attraction of asymmetric stable law $G_{\alpha,1}$ with $1 < \alpha < 2$, $\beta = 1$. For instance, it can be r.v. with Pareto type tails with corresponding $1 < \alpha < 2$. If these claim sizes' distributions additionally satisfy condition (**C**), then process $D^*(t) = -D(t) = -\sum_{i=1}^{N(t)} X_i$ obeys the modification of the LIL (Theorems 9, 10), i.e.

$$\limsup_{t\to\infty} \frac{m\lambda t - D(t)}{t^{1/\alpha}(B^{-1}\ln\ln t)^{1/\theta}} = \lambda^{1/\alpha},$$

and its increments – Theorem 13 . Note that in this case limit results have one-side form.

REFERENCES

References

[1] Alex, M., Steinebach, J. *Invariance principles for renewal processes and some applications.* Teor. Imovirnost. ta Matem. Statyst, **50**, (1994), 22–54.

[2] Berkes, I, Dehling, H., Dobrovski, D. and Philipp, W.*A strong approximation theorem for sums of random vectors in domain of attraction to a stable law*, Acta Math. Hung., **48**, (1986), N 1-2, 161–172.

[3] Berkes, I., Dehling, H. *Almost sure and weak invariance principle for random variables attracted by a stable law*, Probab. Theory and Related Fields, **38**, (1989), N 3, 331–353.

[4] Csörgő, M., Horváth, M., Steinebach. J., *Invariance principles for renewal processes,* Ann. Prob., **15**, (1987), 1441–1460.

[5] Csörgő, M., Révész, P., *Strong Approximation in Probability and Statistics,* Acad. Press., New York (1981).

[6] Csörgő, M., Horváth, L., *Weighted Approximation in Probability and Statistics,* J.Wiley, New York (1993).

[7] Embrechts, P., Klüppelberg, C., Mikosch, T., *Modelling Extremal Events,* Springer - Verlag, Berlin, (1997).

[8] Erdös, P., Rényi,A., *On a new law of large numbers,* J. Analyse Math, **23**, (1970), 103–111.

[9] Frolov, A., *On Erdös-Rényi Laws for renewal processes,* Teor. Imovir. ta Matem. Statyst., **68**, (2003), 164–173.

[10] Furrer, H., *Risk processes pertubed by α-stable Lévy motion,* Scand. Actuarial. J., (1998), no. 1, 59–74.

[11] Furrur, H., Michna, Z., Weron, A., *Stable Lévy motion approximation in collective risk theory,* Insurance Math. Econ., **20** , (1997), 97–114.

[12] Grandell J.,*Aspects of Risk Theory,* Springer, Berlin, (1991).

[13] Iglehart, D.L.,*Diffusion approximation in collective risk theory,* J. Appl. Probab., **6**, 285–292.

[14] Lanzinger, H., Stadmüller, U., *Maxima of increments of partial sums for certain subexponential distributions*, Stoch.Proc.Appl., **86**, (2000), 323 –343.

[15] Lin, Z.Y., Lu,C.R. *Strong Limit Theorems*, Kluwer Science Press, Hong Kong, (1992).

[16] Mijnheer, J., *Limit theorems for sums of independent random variables in the domain of attraction of a stable law: a survey* , Teor.Imovirnost. ta Matem. Statyst, **53**, (1995), 109–115; English transl. in Theory Probab. Math. Statist., **53**, (1995).

[17] Silvestrov, D., *Limit Theorems for Randomly Stopped Stochastic Processes,* Springer-Verlag, London, (2004).

[18] Skorokhod, A. V. *Studies in the Theory of Random Processes,* Kiev Univ, (1961).

[19] Strassen, V., An invariance principle for the LIL. Z. Wahr. verw. Geb., **3**, (1964), 211–226.

[20] Strassen, V., *Almost sure behavior for sums of independent r.v. and martingales,* Pros. 5th Berkeley Symp.**2**, (1967), 315–343.

[21] Whitt, W., *Stochastic-Processes Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer-Verlag, New York, (2002).

[22] Zinchenko, N., *The strong invariance principle for sums of random variables in the domain of attraction of a stable law,* Teor. Veroyatnost. i Primenen., **30**, (1985), 131–135; English transl. in Theory Probab. Appl.**30**, (1985).

[23] Zinchenko, N., *Asymptotics of increments of the stable processes with the jumps of one sign,* Teor. Veroyatnost. i Primenen., **32**, (1987), 793–796; English transl. in Theory Probab. Appl., **32**, (1987).

[24] Zinchenko, N., *Generalization of the strong invariance principle for multiple sums of random variables in the domain of attraction of a stable law,* Teor. Imovirnost. ta Mat. Statist., **57**, (1997), 31–40; English transl. in Theory Probab. Math. Statist. **58**, (1998).

[25] Zinchenko, N., *Skorokhod representation and strong invariance principle,* Teor. Imovirnost. ta Mat. Statist., **63**, (2000), 51–63; English transl. in Theory Probab. Math. Statist. **63**, (2001).

[26] Zinchenko, N., *The strong invariance principle for renewal and randomly stopped processes,* Theory of Stoch.Processes, **13(29)**, no 4, (2007), 233–245.

On the use of the Jensen difference in actuarial science

Athanasios Sachlas and Takis Papaioannou
asachlas@unipi.gr takpap@unipi.gr
Department of Statistics & Insurance Science, University of Piraeus

Abstract

In mathematics and statistics there exist many divergences (see for example Read and Cressie (1988), Liese and Vajda (1987), Pardo (2006)). One of them, which has a special appeal since it originates from Shannon's entropy (a well known index of diversity) and its concavity property, is *Jensen's difference* as it was called by Burbea and Rao (1982). The Jensen difference is given by

$$J(\boldsymbol{p},\boldsymbol{q}) \quad \equiv \quad H\left(\tfrac{1}{2}(\boldsymbol{p}+\boldsymbol{q})\right) - \tfrac{1}{2}\left[H(\boldsymbol{p}) + H(\boldsymbol{q})\right],$$

where $H(\boldsymbol{p}) = -\sum_i p_i \ln p_i$ is the Shannon entropy between the probability vectors $\boldsymbol{p} = (p_1,\ldots,p_n)^T$ and $\boldsymbol{q} = (q_1,\ldots,q_n)^T$. Continuing our research on the properties and the use of divergence and information measures in the actuarial field, in the present paper, we investigate the properties of the Jensen difference in the case of non-probability vectors. This appears in actuarial graduation. We also investigate the use of Jensen's difference in minimum discrimination information, MDI, and in the problem of determining a client's loss distribution (Brockett, 1991).

Keywords. Jensen difference, divergence measures, graduation, loss distribution

1 Introduction

The dominating notion in Information Theory is *Shannon's entropy* given by

$$H(X) = -\sum_x p(x)\ln p(x) \;\text{ or }\; H(X) = -\int f(x)\ln f(x)\,dx$$

depending on whether the random variable X is discrete or continuous with distributions $p(x)$ or $f(x)$, respectively. In the latter case, $H(X)$ is also called *differential entropy*. This measure quantifies the expected uncertainty related with the result of an experiment. In other words it provides information for the predictability of the result of a random variable X. The bigger the entropy the less concentrated the distribution of X and thus an observation of X provides a little information.

Shannon (1948) used entropy to compare distributions while Kullback and Leibler (1951) introduced the discrimination information function (Kullback - Leibler directed divergence) between two distributions. Since then statisticians and econometricians have developed information indices for categorical data analysis, regression modelling, canonical correlation, simultaneous equations, time series, testing distributional hypotheses, and describing income disparity, to name a few. Information measures are usually grouped into two categories: entropy type and divergence type measures. The most well known representative of the first group is Shannon's entropy while for the latter group the Kullback - Leibler directed divergence, the power and Jensen divergences.

In sciences there exist many measures of divergence. The most well known is the Kullback - Leibler directed divergence. Since the appearance of this divergence, a large number of measures of divergence have been proposed, with several properties, axiomatic or not, and several attempts for their use and establishment have been made in several fields. Besides the basic properties, sampling distributions of sample estimates of the various measures of divergence have been studied, mostly asymptotic distributions, and this has led to tests of hypotheses, mostly goodness of fit and confidence intervals for the parameters of the statistical model. Attempts for the unification of the measures and the theory have also been made with main representative the ϕ-divergence

$$I_X^C(f_1, f_2) = I_X^{\phi}(f_1, f_2) = \int \phi\left(\frac{f_1(x)}{f_2(x)}\right) f_2(x)\, dx$$

of Csiszar and later the almost equivalent family of power divergences. Here $f_1(x)$, $f_2(x)$ are pdf's and $\phi(\cdot)$ is a convex function satisfying certain conditions. There exist, however, divergences which cannot be obtained from the ϕ-divergence (or the power divergence). One of them is Jensen's difference given above in the abstract, which is also known as Sibson's information radius [cf. Sibson (1969)] and is a simple special case of divergences originating from ϕ-entropies, where now ϕ is a concave function. For details see Mathai and Rathie (1975), Burbea and Rao (1982), Pardo et al. (1995) and Pardo (2006) and references cited there in.

The properties, both population and sampling, of these divergences in various statistical problems have not been fully studied or developed. We intend to do this in our future research. Jensen's difference has a special appeal because of its simplicity in terms of entropies. The advantages and disadvantages over existing divergences is an open problem and requires a comparative study. In actuarial science, while entropy or the principle of maximum entropy are considerably used as it is demonstrated in Section 2 below, recent and modern divergences are not much in use.

The first major use of a measure of divergence in actuarial science is in Brockett's (1991) work, who used the Kullback - Leibler divergence. In general, the Kullback - Leibler divergence is the first, and sometimes the only, measure of divergence used for the solution of actuarial problems. This gave us the motivation to examine the role and performance of other divergences and especially that of Jensen's difference in actuarial science.

In this paper after surveying in Section 2 the use of information theory in actuarial science we present in Section 3 two actuarial problems involving divergences. One of them is the determination of a client's loss distribution and the second is the graduation of mortality rates. Both of them have been presented and solved by means of the Kullback-Leibler divergence, in the seminal paper of Brockett (1991). In this paper the emphasis is on the Jensen difference which we study in detail in Section 4. A special feature of our approach is the use of non-probability vectors which appear in the actuarial graduation problem but may appear in many other situations. In Section 5 we give a numerical example concerning the determination of a clients loss distribution and graduation while in Section 6 we give concluding results.

2 Use of information theory in actuarial science

Information theory is related to actuarial science through the use of information measures for the treatment and solution of actuarial problems. In general terms we can categorize the use of information theory into three categories: through entropy, through the Kullback-Leibler divergence or relative entropy and through other measures.

Entropy, as a measure of uncertainty and information, is useful for studying and evaluating actuarial models. A well known method of estimating probability models is the method of maximum entropy (ME). In this method, starting with some moments, which may provide the only available information for the model, the model which maximizes the entropy is selected. This method is widely used in several sciences such as economics, accounting, biology, medicine, ecology, actuarial science etc. (Kapur, 1989).

Berliner and Lev (1978) derived the Poisson and Pareto distributions for insurance problems using ME principle and showed by an application of Bayes' theorem that the ME principle also leads to estimates of the parameters of the density function. His conclusion was that there are many possibilities of applying the ME principle in insurance and it represents a vast and interesting hunting ground for actuaries. Many of the procedures developed by actuaries in a more or less empirical way for treatment of special problems may have a simple interpretation through the concept of ME.

Haberman, in his comments on a paper by Moore (1980), suggests the use of the ME principle as a method of choosing the distribution in a wide range of circumstances where the normal distribution is used. Another actuarial area of potential use of ME is the estimation of the largest or smallest members of a set (i.e. extreme values) which could have an application in such diverse fields as the estimation of very large insurance claims and the prediction of stock market peaks (or troughs) in the long-term interest rate.

Kapur (1989) presents some examples of applications of the ME principle in insurance. Especially, he employs the ME procedure to find the probability distribution of the number of claims on an insurance company in a time interval, to find the distribution of catastrophic events etc.

Levin and Tchernitser (2003) choose a probability distribution for the stochastic variance (SV) that provides the most uncertain outcomes given only information about the average value using the ME principle. In the case of one asset, the Maximum Entropy principle gives a broad class of pure jump Generalized Gamma processes for the SV.

The ME principle, when applied to credit risk, leads to models containing the minimal assumptions coherent with the available information (Brunel, 2004). These models are called minimal models and constitute a reference point when we design a new model. She applied this approach to the choice of the loss distribution of credit portfolios, asset backed securities and to the distribution of recovery rates, and she showed how to use it for challenging the assumptions of any model in these areas.

Darooneh (2004) uses the ME principle for pricing the non-life insurance. Specifically, he employs the ME procedure in order to calculate the price density (the insurance price is defined through to a probability function), assuming that the average of the market's wealth is constant. Finally, Luthi and Doege (2005) discuss the family of entropy based risk measures.

The Kullback-Leibler directed divergence was first introduced in actuarial problems as an information theoretic method for actuarial graduation by Brockett and Zhang (1986). More specifically, Zhang and Brockett (1987) tried to construct a smooth series of n death probabilities $\{v_x\}$, $x = 1, 2, \ldots, n$ which is as close as possible to the observed series $\{u_x\}$, in the sense of Kullback-Leibler divergence, subject to three mathematical and two actuarial constraints.

Brockett (1991) gives a very good description of the use of information theory in actuarial science. He describes the use of the Kullback-Leibler divergence for model selection and how information theory unifies and extends certain Bayesian methods used in actuarial science. He also presents the loglinear model, and its special case the logit model which have applications in several aspects of actuarial science as a consequence of information theoretic modelling. He also describes the use of information theory in the determination of a client's loss and the adjustment of mortality tables.

Xu et al. (1998) calculate upper and lower bounds on the stop-loss premium, i.e. the expected payment by the reinsurer, when the claim distribution is unknown but assumed to be in the proximity of the empirical distribution of past claims. The ”distance” from the observed empirical distribution is measured by the I-divergence, i.e. the Kullback-Leibler information number. This ”distance” is then used to determine the bounds.

Relative entropy is commonly used to price risky financial assets in incomplete markets, while distortion is widely used to price insurance risks and in risk management (Reesor and McLeish, 2002). One can obtain the conditional tail expectation (CTE) or Tail-VaR, a well-known risk measure, via the minimization of the Kullback-Leibler divergence between the distorted and original distributions subject to the constraint that the entire distribution is supported in the tail $\{x; x > VaR_\alpha\}$. Relative entropy provides an easy way of constructing new coherent risk measures by prescribing new sets of moments constraints.

Information measures were used to establish a prior distribution for the dispersion parameter λ of the exponential dispersion model. Landsman and Makov (1998) used the maximum entropy principle and Landsman and Makov (1999) minimized the Fisher information. This latter criterion was also used in Landsman and Makov (2001) to establish a prior distribution for λ in conjunction with knowledge on the probability that a claim exceeds a certain threshold, thus allowing for information on tail behavior to affect the premium. Promislow and Young (2000) develop equitable credibility premiums using an entropy loss function, instead, so that a measure of the relative difference between the charged premium and the true premium is minimized in place of the usual squared error.

Finally, Sachlas and Papaioannou (2008a, 2008b) present the use of the Cressie and Read divergence in the problem of actuarial graduation and suggest the minimization of the divergence subject to an additional to those suggested by Zhang and Brockett (1987) constraint.

3 Actuarial problems

Two actuarial problems that can be solved via information theoretic methods are the determination of a client’s loss distribution and the graduation of mortality rates (Brockett, 1991).

3.1 Determination of a client’s disability distribution

Most insurance companies adopt a reference or a standard distribution for losses. Here we are concerned with the distribution of the duration, say in days, of a disability. This distribution might not be immediately applicable to a particular client’s situation. So it is common to make adjustments in order to reflect the known characteristics of the client. Particularly, for the determination of the distribution of a client’s disability duration with expectation μ different from that of a standard table, which is the less distinguishable from the standard table, we can minimize a measure of divergence

$$D(\boldsymbol{p}, \boldsymbol{q})$$

subject to

$$\sum_{i=1}^{\omega} p_i = 1 \text{ and } \sum_{i=1}^{\omega} x_i p_i = \mu,$$

where q_i is the known probability of the disability having a duration of x_i days, obtained from a reference table, $\sum_{i=1}^{\omega} q_i = 1$, p_i is the unknown probability of a duration of x_i days to be developed for the particular client and $x_1, x_2, \ldots, x_\omega$ are ω discrete times of interest and given in the standard table. The first constraint is used in order the p_i's to form a probability distribution.

Brockett (1991) describes the use of Minimum Discrimination Information for this problem. Specifically, he minimizes the Kullback - Leibler divergence, i.e.

$$\sum_{i=1}^{\omega} p_i \ln\left(\frac{p_i}{q_i}\right)$$

subject to the two above mentioned constraints. We note that Brockett solves the above minimization problem via its unconstrained dual convex programming problem.

3.2 Graduation through divergences

A common problem for an actuary is the description of the actual but unknown mortality pattern of a population. In order to achieve this the actuary calculates from raw data crude mortality rates, death probabilities or forces of mortality. Because these entities form an irregular series the actuary revises the initial estimates with the aim of producing smoother estimates, with a procedure called *graduation*. There are several methods of graduation classified into parametric curve fitting and non-parametric smoothing methods. A very good reference book for graduation is that of London (1985).

Brockett and Zhang (1986) were the first to propose the use of information theoretic ideas in graduation. Zhang and Brockett (1987) tried to construct a smooth series of n death probabilities $\{v_x\}$, $x = 1, 2, \ldots, n$ which is as close as possible to the observed series $\{u_x\}$ and in addition they assumed that the true but unknown underlying mortality pattern is (i) smooth, (ii) increasing with age x, i.e. monotone, (iii) more steeply increasing in higher ages, i.e. convex. They also assumed that (iv) the total number of deaths in the graduated data equals the total number of deaths in the observed data, and (v) the total age of death in the graduated data equals the total age of death in the observed data. By total age of death we mean the sum of the product of the number of deaths at every age by the corresponding age. The last two constraints imply that the average age of death is required to be the same for the observed and graduated mortality data. In the sequel and in this section, when we use $x = 1, 2, \ldots, n$ we shall mean the corresponding ages $x_1, x_2, \ldots, x_n$.

Mathematically the five constraints are written as follows: (i) $\sum_x \left(\Delta^3 v_x\right)^2 \leq M$, where M is a predetermined positive constant and $\Delta^3 v_x = -v_x + 3v_{x+1} - 3v_{x+2} + v_{x+3}$; (ii) $\Delta v_x \geq 0$, where $\Delta v_x = v_{x+1} - v_x$; (iii) $\Delta^2 v_x \geq 0$, where $\Delta^2 v_x = v_x - 2v_{x+1} + v_{x+2}$, (iv) $\sum_x l_x v_x = \sum_x l_x u_x$, where l_x is the number of people at risk in the age x; and (v) $\sum_x x l_x v_x = \sum_x x l_x u_x$. In matrix notation the constraints can be written as: (i) $(\boldsymbol{Av})^T (\boldsymbol{Av}) = \boldsymbol{v}^T \boldsymbol{A}^T \boldsymbol{Av} \leq M$, where $\boldsymbol{A}$ is an $(n-3) \times n$ matrix with rows of the form (0, −1, 3, −3, 1, 0, . . ., 0); (ii) $\boldsymbol{Bv} \geq \boldsymbol{0}$, where $\boldsymbol{B}$ is an $(n-1) \times n$ matrix with rows of the form (0, −1, 1, 0, 0, . . ., 0); (iii) $\boldsymbol{Cv} \geq \boldsymbol{0}$, where $\boldsymbol{C}$ is an $(n-2) \times n$ matrix with rows of the form (0, 1, −2, 1, 0, . . ., 0); (iv) $\boldsymbol{d}^T \boldsymbol{v} = \boldsymbol{d}^T \boldsymbol{u}$, where $\boldsymbol{d} = (l_x, l_{x+1}, ..., l_{x+n-1})^T$; and (v) $\boldsymbol{e}^T \boldsymbol{v} = \boldsymbol{e}^T \boldsymbol{u}$, where $\boldsymbol{e} = (xl_x, (x+1)l_{x+1}, ..., (x+n-1)l_{x+n-1})^T$, respectively. For more details see Zhang and Brockett (1987). It is easy to see that the constraints (i) - (v) may be written in the form of $g_i(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^T \boldsymbol{D}_i \boldsymbol{v} + \boldsymbol{b}_i^T \boldsymbol{v} + c_i \leq 0$, $i = 1, 2, ..., m$, where, for each i, $\boldsymbol{D}_i$, $\boldsymbol{b}_i$, c_i are a positive semidefinite matrix and constants, respectively easily written down from (i) - (v) and in this case we have $m = 2(n+1)$ constraints, where n is the number of ungraduated values.

In order to obtain the graduated values, Brockett (1991) minimize the Kullback-Leibler divergence between the crude death probabilities $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ and the new death probabilities $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)^T$,

$$D^{KL}(\boldsymbol{v}, \boldsymbol{u}) = \sum_x v_x \ln \frac{v_x}{u_x},$$

subject to the constraints (i) - (v).

It is easily seen that the mortality rates (death probabilities) $\boldsymbol{u}$ and $\boldsymbol{v}$ are not probability vectors since we have $\sum_{x=1}^{n} u_x > 1$ and $\sum_{x=1}^{n} v_x > 1$. To solve this problem, Sachlas and Papaioannou (2008a, 2008b), as a byproduct, investigated the properties of measures of divergence in the case of non-probability vectors, concluding that under some conditions these can be used as proper divergence measures and proposed the use of an extra constraint in the minimization problem, i.e.

$$\text{(vi)} \sum_{x=1}^{n} v_x = \sum_{x=1}^{n} u_x$$

Constraint (vi) can be written in vector notation and thus in the form of $g(\boldsymbol{v})$ as $\mathbf{1}^T\boldsymbol{v} - \mathbf{1}^T\boldsymbol{u} = \mathbf{0}$, where $\mathbf{1} = (1, 1, ..., 1)^T$.

Sachlas and Papaioannou (2008a, 2008b) also presented a new and unifying way to obtain the graduated values v_x. This is by minimizing the Cressie-Read divergence

$$D^{CR}(\boldsymbol{v}, \boldsymbol{u}) = \frac{1}{r(r+1)} \sum_x v_x \left[\left(\frac{v_x}{u_x}\right)^r - 1\right]$$

for given r subject to constraints (i) - (v) and/or (vi), i.e. $\boldsymbol{v} \geq \mathbf{0}$ and $g(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^T\boldsymbol{D}_i\boldsymbol{v} + \boldsymbol{b}_i^T\boldsymbol{v} + c_i \leq 0$, $i = 1, 2, ..., m+1$, where $m = 2(n+1)$. The minimization is done for various values of the parameter r and the objective is to find the best r for graduation purposes.

4 The Jensen difference

As stated in the Introduction in mathematics and statistics there exist many divergences between discrete or continuous distributions (see for example Read and Cressie (1988), Liese and Vajda (1987), Mathai and Rathie (1975)). One of them, which has a special appeal since it originates from Shannon's entropy and the concavity property, is *Jensen's difference* as it was called by Burbea and Rao (1982). The Jensen difference between probability vectors is given by

$$J(\boldsymbol{p}^*, \boldsymbol{q}^*) \equiv H\left(\tfrac{1}{2}(\boldsymbol{p}^* + \boldsymbol{q}^*)\right) - \tfrac{1}{2}\left[H(\boldsymbol{p}^*) + H(\boldsymbol{q}^*)\right],$$

where $H(\boldsymbol{p}^*) = -\sum_i p_i^* \ln p_i^*$ is the Shannon entropy between the probability vectors $\boldsymbol{p}^* = (p_1^*, \ldots, p_n^*)^T$ and $\boldsymbol{q}^* = (q_1^*, \ldots, q_n^*)^T$.

The Jensen difference is a natural measure of divergence between the probability vectors $\boldsymbol{p}^*$ and $\boldsymbol{q}^*$ as it satisfies the two basic properties of a divergence measure. It is nonnegative and vanishes if and only if $\boldsymbol{p}^* = \boldsymbol{q}^*$. An interesting property of $J(\boldsymbol{p}^*, \boldsymbol{q}^*)$ is that considered as a function of $(\boldsymbol{p}^*, \boldsymbol{q}^*)$ is convex.

4.1 The Jensen difference without probability vectors

In this section we will derive some properties of the Jensen difference when we have non-probability vectors as is the case in actuarial graduation. This also supplements our previous work on the properties of divergence measures without probability vectors.

Definition 1. *We define as*

$$J(\boldsymbol{p},\boldsymbol{q}) \equiv H\left(\tfrac{1}{2}(\boldsymbol{p}+\boldsymbol{q})\right) - \tfrac{1}{2}\left[H(\boldsymbol{p})+H(\boldsymbol{q})\right], \tag{1}$$

the Jensen difference between the non-probability vectors $\boldsymbol{p} = (p_1,\ldots,p_n)^T$ *and* $\boldsymbol{q} = (q_1,\ldots,q_n)^T$, *where* $\boldsymbol{p} \geq \boldsymbol{0}$, $\boldsymbol{q} \geq \boldsymbol{0}$ *and* $\sum_i p_i \neq 1$ *and* $\sum_i q_i \neq 1$. *The* $H(\boldsymbol{p}) = -\sum_i p_i \ln p_i$ *is the Shannon entropy.*

Now we have to see if this measure has information theoretic and divergence properties. For convenience we will assume in the sequel that $\sum_i p_i = \sum_i q_i$.

Lemma 1. *If* $\sum_i p_i = \sum_i q_i$, *then for the Jensen difference involving non-probability vectors* $\boldsymbol{p},\boldsymbol{q}$, *it holds that*

$$J(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{\sum_i p_i} J(\boldsymbol{p}^*,\boldsymbol{q}^*),$$

where $J(\boldsymbol{p}^*,\boldsymbol{q}^*)$ *is the Jensen difference between the two probability vectors* $\boldsymbol{p}^*,\boldsymbol{q}^*$.

Proposition 1. *Let* $\sum_i p_i = \sum_i q_i$. *Then* $J(\boldsymbol{p},\boldsymbol{q}) \geq 0$ *with equality if and only if* $\boldsymbol{p} = \boldsymbol{q}$, *where* $\boldsymbol{p}$ *and* $\boldsymbol{q}$ *are non-probability vectors.*

Definition 2. *(Bivariate Shannon entropy) Let* $p(x,y)$ *be a bivariate measure (non - probability function) associated with two discrete variables* X, Y *in* R^2 *for which it holds* $\sum_x \sum_y p(x,y) \neq 1$. *We define the Shannon entropy involving a bivariate non-probability function* p *as*

$$H_{X,Y}(p) = -\sum_x \sum_y p(x,y) \ln p(x,y).$$

Definition 3. *(Conditional Shannon entropy) For the discrete variables* X, Y *and the bivariate non-probability function* $p(x,y)$, *as given above let* $f(x) = \sum_y p(x,y)$, $h(y|x) = \frac{p(x,y)}{f(x)}$, $g(y) = \sum_x p(x,y)$, *and* $r(x|y) = \frac{p(x,y)}{g(y)}$, $i = 1,2$. *We set*

$$H_{Y|X=x}(h) = \sum_y h(y|x) \ln h(y|x), H_{X|Y=y}(r) = \sum_x r(x|y) \ln r(x|y)$$

and define

$$H_{Y|X}(h) = E_X\left[H_{Y|X=x}(h)\right] = \sum_x f(x) \sum_y h(y|x) \ln h(y|x),$$
$$H_{X|Y}(r) = E_Y\left[H_{X|Y=y}(r)\right] = \sum_y g(y) \sum_x r(x|y) \ln r(x|y).$$

Definition 4. *(Bivariate Jensen difference) Let* $p_i(x,y)$, $i = 1,2$, *be two bivariate measures (non-probability functions) associated with two discrete variables* X, Y *in* R^2 *for which it holds* $\sum_x \sum_y p_i(x,y) \neq 1$. *We define the Jensen difference between two bivariate non-probability functions* p_1, p_2 *as*

$$\begin{aligned} J_{X,Y}(p_1,p_2) &= H\left(\tfrac{1}{2}(p_1+p_2)\right) - \tfrac{1}{2}\left[H(p_1)+H(p_2)\right] \\ &= -\sum_x \sum_y \tfrac{1}{2}(p_1(x,y)+p_2(x,y)) \ln(\tfrac{1}{2}(p_1(x,y)+p_2(x,y))) \\ &\quad -\tfrac{1}{2}\left[-\sum_x \sum_y p_1(x,y) \ln p_1(x,y) - \sum_x \sum_y p_2(x,y) \ln p_2(x,y)\right]. \end{aligned}$$

Definition 5. *(Conditional Jensen difference) For the discrete variables X, Y and the bivariate non - probability functions $p_i(x,y)$, $i = 1,2$, as given above let $f_i(x) = \sum_y p_i(x,y)$, $h_i(y|x) = \frac{p_i(x,y)}{f_i(x)}$, $g_i(y) = \sum_x p_i(x,y)$, and $r_i(x|y) = \frac{p_i(x,y)}{g_i(y)}$, $i = 1,2$. We set*

$$\begin{aligned} J_{Y|X=x}(h_1,h_2) &= H\left(\tfrac{1}{2}(h_1+h_2)\right) - \tfrac{1}{2}\left[H(h_1)+H(h_2)\right] \\ &= -\sum_x \sum_y \tfrac{1}{2}(h_1(y|x)+h_2(y|x)) \ln(\tfrac{1}{2}(h_1(y|x)+h_2(y|x))) \\ &\quad -\tfrac{1}{2}\left[-\sum_x \sum_y h_1(y|x) \ln h_1(y|x) - \sum_x \sum_y h_2(y|x) \ln h_2(y|x)\right] \end{aligned}$$

and define

$$\begin{aligned} J_{Y|X}(h_1,h_2) &= E_X\left[J_{Y|X=x}(h_1,h_2)\right] \\ &= -\sum_x f_1(x) \sum_y \tfrac{1}{2}(h_1(y|x)+h_2(y|x)) \ln(\tfrac{1}{2}(h_1(y|x)+h_2(y|x))) \\ &\quad -\tfrac{1}{2}\left[-\sum_x f_1(x) \sum_y h_1(y|x) \ln h_1(y|x)\right. \\ &\qquad \left. -\sum_x f_1(x) \sum_y h_2(y|x) \ln h_2(y|x)\right]. \end{aligned}$$

Similarly we define $J_{X|Y=y}(r_1,r_2)$ and $J_{X|Y}(r_1,r_2)$.

Proposition 2. *(Strong Additivity) Let p_1, p_2 be two bivariate non-probability functions associated with two discrete variables X, Y in R^2 as in Definition 5. Then*

$$J_{X,Y}(p_1,p_2) = J_X(f_1,f_2) + J_{Y|X}(h_1,h_2) = J_Y(g_1,g_2) + J_{X|Y}(r_1,r_2),$$

where the functions f_i, h_i, g_i, r_i, $i = 1,2$ are as in Definition 5.

For weak additivity we have the following proposition.

Proposition 3. *(Weak additivity) If $h_i(y|x) = g_i(y)$ and thus $p_i(x,y) = f_i(x)g_i(y)$, $i = 1,2$, we have that the random variables X^*, Y^*, which are the "standardized" values of X, Y, are independent, and*

$$J_{X,Y}(p_1,p_2) = J_X(f_1,f_2) + J_Y(g_1,g_2).$$

Proposition 4. *(Maximal information and sufficiency) Let $Y = T(X)$ be a measurable transformation of X, then*

$$J_X(p_1,p_2) \geq J_Y(g_1,g_2),$$

with equality if and only if Y is "sufficient", where $p_i = p_i(x)$, $g_i = g_i(y)$, $i = 1,2$.

For proofs of the previous results see Sachlas and Papaioannou (2009)

We have already seen that the Jensen difference $J(\boldsymbol{p},\boldsymbol{q})$ for non-probability vectors, under some conditions is nonnegative, additive and invariant under sufficient transformations. It also shares the property of maximal information. So, we can regard $J(\boldsymbol{p},\boldsymbol{q})$ as a measure of divergence, provided that $\sum\limits_i p_i = \sum\limits_i q_i$.

Since the Jensen difference can be considered as a measure of divergence we can use it in order to graduate actuarial entities, in the way we described in Section 3.2. This involves convex minimization with constraints and can be done using standard routines. It is of interest to examine the Lagrangian dual. In the sequel we derive the Lagrangian dual of the Jensen difference minimization problem.

4.2 Lagrangian duality for the Jensen difference

The quadratically constrained Jensen difference problem is defined as finding $\boldsymbol{x} \in R^n$ which solves the primal problem

$$(P) \qquad \min -\sum_{j=1}^{n} \tfrac{1}{2}(x_j + d_j)\ln\left(\tfrac{1}{2}(x_j+d_j)\right) + \tfrac{1}{2}\left[\sum_{j=1}^{n} x_j \ln x_j + \sum_{j=1}^{n} d_j \ln d_j\right]$$

subject to

$g_i(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{D}_i\boldsymbol{x} + \boldsymbol{b}_i^T\boldsymbol{x} + c_i \leq 0$, $i = 1,2,\ldots,m$, $\boldsymbol{x} \geq \boldsymbol{0}$,

where $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)^T$ is a given vector with strictly positive components, $\boldsymbol{D}_i$ is a given positive semi-definite matrix for each i, $\boldsymbol{b}_i \in R^n$ and c_i are given constants not both equal zero.

In the sequel, we will try to derive a dual representation of the primal problem (P) by means of Lagrangian duality. This will be done by using a simple decomposition argument to convert problem (P) to an equivalent convex program with linear and quadratic constraints.

Because $\boldsymbol{D}_i$ is a semipositive definite $n \times n$ matrix, we can express it as $\boldsymbol{D}_i = \boldsymbol{A}_i^T\boldsymbol{A}_i$, where $\boldsymbol{A}_i$ is an $n_i \times n$ matrix and n_i is the rank of $\boldsymbol{D}_i$, $i = 1,2,\ldots,m$. In this case the constraints can be written as $g_i(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}_i^T\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i^T\boldsymbol{x} + c_i$. Defining the new variables $\boldsymbol{u}_i = \boldsymbol{A}_i\boldsymbol{x}$, $\boldsymbol{u}_i \in R^{n_i}$, $i = 1,2,\ldots,m$, the problem (P) is equivalent to the following convex program with linear equality and quadratic inequality constraints:

$$(P^*) \qquad \min_{\boldsymbol{x},\boldsymbol{u}_i} -\sum_{j=1}^{n} \tfrac{1}{2}(x_j + d_j)\ln\left(\tfrac{1}{2}(x_j+d_j)\right) + \tfrac{1}{2}\left[\sum_{j=1}^{n} x_j \ln x_j + \sum_{j=1}^{n} d_j \ln d_j\right]$$

subject to

$\frac{1}{2}\boldsymbol{u}_i^T\boldsymbol{u}_i + \boldsymbol{b}_i^T\boldsymbol{x} + c_i \leq 0$, $\boldsymbol{A}_i\boldsymbol{x} = \boldsymbol{u}_i$, $\boldsymbol{u}_i \in R^{n_i}$, $i = 1,2,\ldots,m$, $\boldsymbol{x} \geq \boldsymbol{0}$.

Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_m)^T \in R^{n_i} \times \ldots \times R^{n_m}$, $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m)^T \in R^{n_i} \times \ldots \times R^{n_m}$ and $N = n_1 + \ldots + n_m$.

Theorem 1. *The Lagrangian dual problem of* (P) *is given by*

$$(D) \qquad \sup_{\boldsymbol{\lambda}\in R_+^m, \boldsymbol{y}_i \in R^{n_i}} \left\{ -\sum_{j=1}^{n} \frac{d_j}{2e^{2s_j}-1}\left[e^{2s_j}\ln\left(\frac{d_j e^{2s_j}}{2e^{2s_j}-1}\right) - \frac{1}{2}\ln\left(\frac{d_j}{2e^{2s_j}-1}\right) - s_j\right] \right.$$
$$\left. -\frac{1}{2}\sum_{i=1}^{m}\frac{\|\boldsymbol{y}_i\|^2}{\lambda_i} + \boldsymbol{\lambda}^T\boldsymbol{c} + \boldsymbol{d}^T\boldsymbol{z} \right\},$$

where $\boldsymbol{z}^T = (\ln d_1, \ln d_2, \ldots, \ln d_n)$

Theorem 2. *(a) If* (P) *is feasible then* $\inf(P)$ *is attained and* $\min(P) = \sup(D)$. *Moreover, if there exists an* $\boldsymbol{x} \in R^n$ *satisfying* $\boldsymbol{x} > 0$, $g_i(\boldsymbol{x}) < 0$, $i = 1,\ldots,m$, *then* $\sup(D)$ *is attained and* $\min(P) = \max(D)$.

(b) If $\boldsymbol{x}^*$ solves the primal problem (P) and $\boldsymbol{y}_i^* \in R^{n_i}$, $\boldsymbol{\lambda}^* \in R_+^m$ solve the dual problem (D), then

$$x_j^* = \frac{d_j}{2\exp\left\{2\sum_{i=1}^{m}(\lambda_i^\star \boldsymbol{b}_i^T + \boldsymbol{y}_i^{\star T}\boldsymbol{A}_i)_j\right\} - 1}.$$

For the proof of Theorems 1 and 2 see Sachlas and Papaioannou (2009)

5 Numerical Illustration

5.1 Determination of a client's disability distribution with Jensen's difference

In this section we use the Jensen difference to determine the loss distribution that meets the special characteristics of a client than the reference table that the insurance company uses. The data that we will use come from Bowers et al. (1997, Table 13.2). It is a standard table with mean duration of 31.35 days, given in the second column of Table 1. It is easy to notice that$\sum_{i=1}^n q_i = 1$. Suppose that we have a client with expected duration of $\mu = 21$ and we want to construct a duration table for this particular client which is the least distinguishable from the standard one. This problem was also solved by Brockett (1991) by minimizing the Kullback-Leibler divergence between the unknown probabilities for the client and the corresponding probabilities of the standard table subject to the constraints $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n x_i p_i = 21$. His results are shown in the third column of Table 1.

Our approach to this problem is the minimization of another divergence - the Jensen difference - subject to the same constraints, i.e.

$$min\ J(\boldsymbol{p},\boldsymbol{q}) = H\left(\tfrac{1}{2}(\boldsymbol{p}+\boldsymbol{q})\right) - \tfrac{1}{2}\left[H(\boldsymbol{p}) + H(\boldsymbol{q})\right]$$

subject to

$$\sum_{i=1}^n p_i = 1 \text{ and } \sum_{i=1}^n x_i p_i = 21.$$

Our results are shown in the fourth column of Table 1. We notice that the two methods find almost the same results. We repeated the procedure with $\mu = 17$ and $\mu = 38$, respectively. The results are again similar.

5.2 Actuarial graduation with Jensen's difference

For illustration, we will use a data set of death probabilities coming from London (1985, p. 20). It consists of 15 death probabilities belonging to ages 70 to 84 (computed from a total of 2073 observations). These data set was graduated by London (1985) by graphic means and a linear transformation of the graduated values and by Brockett (1991) via the minimization of the Kullback-Leibler divergence subject to constraints (i) - (v).

We graduated the crude values via the minimization of the Jensen difference. The minimization was conducted subject to constraints (i) - (v), proposed by Brockett (1991), the additional constraint

$$(vi)\quad \sum_{x=1}^n v_x = \sum_{x=1}^n u_x$$

that Sachlas and Papaioannou (2008a) proposed and finally subject to constraints (i) - (iii) and (vi). The relevant results are presented along with the raw data in Table 2(a).

		$\mu = 21$		$\mu = 17$		$\mu = 38$	
x	Standard	K-L	Jensen	K-L	Jensen	K-L	Jensen
1	0.03500	0.05081	0.05298	0.06022	0.06588	0.02777	0.02810
2	0.03474	0.04968	0.05151	0.05832	0.06285	0.02775	0.02806
3	0.03349	0.04717	0.04865	0.05486	0.05830	0.02694	0.02721
4	0.03318	0.04604	0.04724	0.05303	0.05563	0.02687	0.02712
5	0.03195	0.04367	0.04459	0.04982	0.05164	0.02606	0.02627
6	0.03160	0.04254	0.04324	0.04808	0.04927	0.02595	0.02614
7	0.03040	0.04031	0.04079	0.04513	0.04576	0.02514	0.02530
8	0.03002	0.03921	0.03951	0.04348	0.04366	0.02499	0.02513
9	0.02885	0.03712	0.03725	0.04077	0.04057	0.02419	0.02430
10	0.02701	0.03423	0.03423	0.03725	0.03675	0.02280	0.02289
11	0.02530	0.03159	0.03147	0.03404	0.03332	0.02150	0.02157
12	0.02370	0.02915	0.02894	0.03111	0.03023	0.02028	0.02033
13	0.02222	0.02692	0.02664	0.02846	0.02747	0.01915	0.01917
14	0.02083	0.02485	0.02452	0.02603	0.02497	0.01808	0.01808
15	0.01953	0.02295	0.02258	0.02381	0.02272	0.01706	0.01706
16	0.01831	0.02120	0.02080	0.02178	0.02067	0.01611	0.01609
17	0.01772	0.02021	0.01978	0.02057	0.01943	0.01570	0.01567
18	0.01662	0.01867	0.01823	0.01882	0.01770	0.01483	0.01479
19	0.01611	0.01783	0.01737	0.01780	0.01668	0.01447	0.01442
20	0.01510	0.01646	0.01600	0.01628	0.01520	0.01366	0.01361
21	0.01465	0.01573	0.01526	0.01541	0.01435	0.01334	0.01328
22	0.01374	0.01453	0.01408	0.01410	0.01309	0.01260	0.01254
23	0.01334	0.01390	0.01344	0.01336	0.01237	0.01232	0.01225
24	0.01295	0.01329	0.01283	0.01265	0.01170	0.01204	0.01196
25	0.01214	0.01227	0.01183	0.01158	0.01068	0.01136	0.01129
26	0.01180	0.01175	0.01132	0.01098	0.01012	0.01112	0.01104
27	0.01106	0.01085	0.01044	0.01004	0.00924	0.01050	0.01041
28	0.01076	0.01039	0.00999	0.00953	0.00877	0.01028	0.01020
31	0.06361	0.05873	0.05634	0.05233	0.04811	0.06206	0.06145
38	0.04832	0.04014	0.03846	0.03346	0.03099	0.04947	0.04886
45	0.03753	0.02805	0.02698	0.02188	0.02062	0.04032	0.03976
52	0.02980	0.02004	0.01943	0.01462	0.01414	0.03360	0.03312
59	0.02399	0.01452	0.01424	0.00991	0.00989	0.02839	0.02800
66	0.01939	0.01056	0.01051	0.00674	0.00699	0.02408	0.02380
73	0.01586	0.00777	0.00787	0.00464	0.00502	0.02067	0.02051
80	0.01300	0.00573	0.00592	0.00320	0.00363	0.01778	0.01773
87	0.01077	0.00427	0.00451	0.00223	0.00266	0.01546	0.01552
91	0.12561	0.04690	0.05025	0.02362	0.02894	0.18531	0.18697

Table 1: Loss distribution determination through Kullback-Leibler divergence and Jensen's difference

Table 2: Several graduations through Jensen difference

x	u_x	v_x (5 constraints)	v_x (6 constraints)	v_x (4 constraints)
70	0.044	0.062	0.054	0.059
71	0.084	0.066	0.061	0.064
72	0.071	0.071	0.068	0.069
73	0.076	0.075	0.075	0.073
74	0.040	0.080	0.082	0.078
75	0.104	0.086	0.089	0.085
76	0.160	0.093	0.097	0.092
77	0.058	0.099	0.104	0.098
78	0.110	0.106	0.112	0.105
79	0.093	0.113	0.119	0.112
80	0.139	0.131	0.138	0.132
81	0.154	0.156	0.159	0.157
82	0.183	0.182	0.180	0.184
83	0.206	0.209	0.201	0.212
84	0.239	0.238	0.222	0.242

(a) Graduated values

	5 constraints	6 constraints	4 constraints
S	0.000199	0.0002	0.0002
F	16.62	16.70	16.93
Deviance	16.40	16.89	16.48
log-likelihood	-713.12	-713.37	-713.16
χ^2	16.59	16.68	16.93

(b) Smoothness and goodness of fit values

The results appear nearly equivalent to those presented by London and Brockett. The differences are small. The value of the smoothness measure and the goodness of fit measures, i.e. F, deviance, log-likelihood and χ^2, are given in Table 2(b). The numerical investigation of Sachlas and Papaioannou (2008a), with same data set, compared the graduations made by London (1985), Brockett (1991), the minimization of the Cressie-Read power divergence and the minimization of the Jensen difference, with the conclusion that the overall winner is the graduation through the minimization of the Jensen difference subject to constraints (i) - (v), as judged by S and F.

6 Conclusions

In this paper we studied the use of Jensen's difference in actuarial science. Because mortality rates are not probability vectors, and in order to use $J(\boldsymbol{p},\boldsymbol{q})$ for graduation purposes, we investigated the properties of the Jensen difference in the case of non-probability vectors. We show that, under some conditions it is nonnegative, additive and invariant under sufficient transformations. It also shares the property of maximal information. So, we can regard $J(\boldsymbol{p},\boldsymbol{q})$ as a measure of divergence, provided that $\sum_i p_i = \sum_i q_i$, and use it for graduation.

We also provided Lagrangian duality results for the problem of minimizing the Jensen difference subject to quadratic and linear inequality constraints. Especially, we derived the Lagrangian dual problem of minimizing the Jensen difference, which proved to be unconstrained, and the solution of the dual problem. These results are important in actuarial science and especially in the problem of graduation of mortality rates were mortality rates do not constitute probability vectors.

The numerical investigation indicated that the use of Jensen's divergence in actuarial problems such as the loss distribution determination and the graduation is comparable to other divergences used to these problems. Especially in the problem of graduation the minimization of the Jensen difference between the crude and graduated rates seems to be the best ”divergence” method.

References

[1] Berliner, B. and Lev, B. (1980). On the use of maximum entropy concepts in insurance, *Proceedings of the 21th of the International Congress of Actuaries*, T1, 47-61.

[2] Bowers,N.L.,Jr., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd ed., Schaumburg, Ill.: Society of Actuaries.

[3] Brockett, P.L. (1991). Information Theoretic Approach to Actuarial Science: A Unification and Extension of Relevant Theory and Applications, *Transactions of the Society of Actuaries*, **43**, 73-114.

[4] Brockett, P.L., Zhang, J. (1986). Information Theoretical Mortality Graduation, *Scandinavian Actuarial Journal*, 131-140.

[5] Brunel, V. (2004). Minimal Models for Credit Risk: An Information Theory Aproach, Working problem, http://vivienbrunel.free.fr/WorkingPapers/Entropy.pdf.

[6] Burbea, J., Rao, C.R. (1982). On the Convexity of Some Divergence Measures Based on Entropy Functions, *IEEE Transactions on Information Theory*, vol 28, no 3, 489-495.

[7] Cressie, N.A.C., Read, T.R.C. (1984). Multinomial Goodness - of - Fit Tests, *Journal of Royal Statistical Society*, B, Vol. 46, No. 3, 440-464.

[8] Darooneh, A.H. (2004). Non-life insurance pricing: multiagent model, *Insurance: Mathematics and Economics*, **24**, 23-29.

[9] Kapur, J.N. (1989). *Maximum-Entropy Models in Science and Engineering*. Wiley, New York.

[10] Kullback, S. (1959). *Information Theory and Statistics*, John Wiley & Sons, New York.

[11] Landsman, L.M. and Makov, U.E. (1998). Exponential dispersion models and credibility, *Scandinavian Actuarial Journal*, **1**, 89-96.

[12] Landsman, L. and Makov, U. (1999). Credibility evaluation for the exponential dispersion family, *Insurance: Mathematics and Economics*, **24**, 23-29.

[13] Landsman, L. and Makov, U. (2001). On Credibility Evaluation and the Tail Area of the Exponential Dispersion Family, *Insurance: Mathematics and Economics*, **27**, 277-283.

[14] Levin, A. and Tchernitser, A. (2003) Multifactor stochastic variance models in risk management: Maximum entropy approach and Levy processes, In *Handbook of Heavy Tailed Distributions in Finance*, (S.T. Rachev, ed), 443-480, Elsevier Science B.V., Amsterdam.

[15] Liese F., Vajda, I. (1987). *Convex Statistical Distances*, B. G. Teubner, Leipzig.

[16] Moore, P.G. (1980). Are objectivism and subjectivism compatible concepts?, *Journal of the Institute of Actuaries*, **107**, 11-47.

[17] Reesor, R.M. and McLeish, D.L. (2002). Risk, Entropy, and the Transformation of Distributions, *Bank of Canada Working Paper*.

[18] Pardo, L, Morales, O. and Taneja, I.J. (1995). Generalized Jensen difference divergence measures and Fisher measure of information, *Kybernetes*, , Vol. 24, No. 2, 15-28.

[19] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*, Chapman and Hall, London.

[20] Promislow, S.D. and Young, V.R. (2000). Equity and Exact Credibility, *ASTIN Bulletin*, **30(1)**, 3-11.

[21] Read, T.R.C., Cressie, N.A.C. (1988). *Goodness - of - Fit Statistics for Discrete Multivariate Data*, Springer - Verlag, New York.

[22] Sachlas, A.P., Papaioannou, T. (2008a). On a minimization problem involving divergences and its applications, *Advances in Data Analysis: Theory and Applications to Reliability and Inference, Data Mining, Bioinformatics, Lifetime Data and Neural Networks*, (C. H. Skiadas, Ed.), 83-97, Birkhauser, Boston.

[23] Sachlas, A.P., Papaioannou, T. (2008b). Divergences without probability vectors and their applications, Under revision.

[24] Sachlas, A.P., Papaioannou, T. (2009). Jensen's divergence in Actuarial Science. Paper under preparation.

[25] Sibson, R. (1969). Information radius, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 14, 149-160.

[26] Xu, L., Bricker, D.L. and Kortanek, K.O. (1987). Bounds for stop-loss premium under restrictions on I-divergence, *Insurance: Mathematics and Economics*, **23**, 119-139.

[27] Zhang, J., Brockett, P.L. (1987). Quadratically Constrained Information Theoretic Analysis, *SIAM Journal of Applied Mathematics*, Vol. 47, No. 4, 871-885.

Robust prediction problem for periodically correlated stochastic sequences

Mikhail Moklyachuk
Department of Probability Theory and Mathematical Statistics
Kyiv National Taras Shevchenko University
Kyiv 01601, Ukraine
Email: mmp@univ.kiev.ua

Abstract

We deal with the problem of optimal linear estimation of the functional

$$A\zeta = \sum_{j=0}^{\infty} a(j)\zeta(j)$$

which depends on the unknown values of a periodically correlated (cyclostationary) discrete time stochastic process $\zeta(j)$ from observations of the process $\zeta(n) + \theta(n)$ for $n = -1, -2, \ldots,$ where $\theta(n)$ is uncorrelated with $\zeta(n)$ periodically correlated (cyclostationary) sequence.

keywords:Periodically correlated stochastic sequence; stationary stochastic sequence; robust linear prediction; observations with noise; mean square error; least favorable spectral densities; minimax spectral characteristic.

AMS Subject Classifications. Primary: 60G10, 62M20, 62P05, Secondary: 60G35, 93E10, 93E11

1 Introduction.

Traditional methods of solution of the linear extrapolation, interpolation and filtering problems for stationary stochastic processes may be employed under the condition that spectral densities of processes are known exactly (see, for example, selected works of A. N. Kolmogorov (1992), survey article by T. Kailath (1974), books by Yu. A. Rozanov (1990), N. Wiener (1966), A. M. Yaglom (1987)). In practice, however, complete information on the spectral densities is impossible in most cases. To find a solution to the problem one finds parametric or nonparametric estimates of the unknown spectral densities or selects these densities by other reasoning. Then applies the classical estimation method provided that the estimated or selected densities are the true one. This procedure can result in significant increasing of the value of error as K. S. Vastola and H. V. Poor (1983) have demonstrated with the help of some examples. This is a reason to search estimates which are optimal for all densities from a certain class of the admissible spectral densities. These estimates are called minimax since they minimize the maximal value of the error. A survey of results in minimax (robust) methods of data processing can be found in the paper by S. A. Kassam and H. V. Poor (1985). The paper by Ulf Grenander (1957) should be marked as the first one where the minimax approach to extrapolation problem for stationary processes was proposed. J. Franke (1984, 1985, 1991), J. Franke and H. V. Poor (1984) investigated the minimax extrapolation and filtering problems for stationary sequences with the help of convex optimization methods. This approach makes it possible to find equations that determine the least favorable spectral densities for various classes of densities. In the papers by M. P. Moklyachuk (1994, 1997, 1998, 2000, 2001), M. P. Moklyachuk and A. Yu. Masyutka (2005, 2006) the minimax approach to extrapolation, interpolation and filtering problems are investigated for functionals which depend on the unknown values of stationary processes and sequences.

In this article we deal with the problem of optimal linear estimation of the functional $A\zeta = \sum_{j=0}^{\infty} a(j)\zeta(j)$ which depends on the unknown values of a periodically correlated (cyclostationary) discrete time stochastic process $\zeta(j)$ from observations of the process $\zeta(n)+\theta(n)$ for $n=-1,-2,\ldots$, where $\theta(n)$ is uncorrelated with $\zeta(n)$ periodically correlated (cyclostationary) discrete time stochastic process. The problem is reduced to the corresponding problem of optimal linear estimation of the functional $A\vec{\xi}$ for multidimensional stationary stochastic process with the help of approach proposed by E. G. Gladyshev (1961). Formulas are obtained for calculation the mean square error and the spectral characteristic of the optimal linear prediction of the functional $A\vec{\xi} = \sum_{j=0}^{\infty} \vec{a}(j)\vec{\xi}(j)$ which depends on the unknown values of a multidimensional stationary stochastic process $\vec{\xi}(j)$ from observations of the process $\vec{\xi}(n)+\vec{\eta}(n)$ for $n=-1,-2,\ldots$ under the condition that the spectral density matrix $F(\lambda)$ of the process $\vec{\xi}(n)$ and the spectral density matrix $G(\lambda)$ of the process $\vec{\eta}(n)$ are known. The least favorable spectral densities and the minimax spectral characteristic of the optimal linear prediction of the functional $A\vec{\xi}$ are found for classes $D = D_F \times D_G$ of spectral densities under the condition that spectral density matrices $F(\lambda)$ and $G(\lambda)$ are not known, but classes $D = D_F \times D_G$ of admissible spectral densities are given.

2 Generating stationary sequences.

Definition 1. *A discrete time stochastic process $\zeta(n)$ with zero mean, $E\zeta(n) = 0$, and finite variance, $E|\zeta(n)|^2 < \infty$, is called cyclostationary or periodically correlated (PC) with period T if the correlation function of the process has the property*

$$E\zeta(n)\overline{\zeta(m)} = K_\zeta(n,m) = K_\zeta(n+T, m+T)$$

for every $n, m \in Z$.

If $\zeta(n)$ is PC with period T, then the function $K_\zeta(n+p,n)$ is T-periodic in n for every $p \in Z$, and we can apply the discrete Fourier transform and get the relations

$$K_\zeta(n+p,m) = \sum_{j=0}^{T-1} e^{2\pi ijn/T} a_j(n), \tag{1}$$

$$a_j(p) = \frac{1}{T}\sum_{n=0}^{T-1} e^{-2\pi ijn/T} K_\xi(n+p,p) \tag{2}$$

Let

$$H_\zeta(n) = \overline{sp}\{\zeta(k), k = n, n-1, \ldots\}$$
$$H_\zeta = \overline{sp}\{\zeta(k), k \in Z\}$$

If $\zeta(n)$ is PC with period T, then the mapping $V:\zeta(n)\to\zeta(n+T)$, $n\in Z$, extends to a unitary operator

$$V:H_{\zeta}\to H_{\zeta}$$

The operator V is called T-shift operator of $\zeta(n)$. Let U be a unitary T-th root of V, such that $U^T=V$ on H_{ζ}. Then $p(n)=U^{-n}\zeta(n)$ is a T-periodic sequence and $\zeta(n)=U^np(n)$.
Define

$$\xi^q(n)=U^n\left(\frac{1}{T}\sum_{n=0}^{T-1}e^{-2\pi jnq/T}p(n)\right),q=0,1,\dots,T-1. \tag{3}$$

Here $\xi^q(n)$ is a T-dimensional stationary sequence and the following representation holds true

$$\zeta(n)=\sum_{q=0}^{T-1}e^{2\pi inq/T}\xi^q(n),\ n\in Z. \tag{4}$$

The sequence $\xi^q(n)$ is called **generating sequence** of $\zeta(n)$. In this case the correlation function

$$K_{\zeta}(n+p,n)=\sum_{j=0}^{T-1}e^{2\pi ijn/T}\left(\sum_{k=0}^{T-1}e^{2\pi ikp/T}(\xi^k(p),\xi^{k-j}(0))\right), \tag{5}$$

where

$$(\xi^k(p),\xi^{k-j}(0))=E\xi^k(p)\overline{\xi^{k-j}(0)}.$$

Therefore we can represent $a_j(p)$ in the form

$$a_j(p)=\sum_{k=0}^{T-1}e^{2\pi ikp/T}(\xi^k(p),\xi^{k-j}(0))=$$

$$=\int_{-\pi}^{\pi}e^{ips}\sum_{k=0}^{T-1}\Gamma^{k,k-j}(ds-2\pi k/T) \tag{6}$$

or

$$a_j(p)=\int_{-\pi}^{\pi}e^{ips}\gamma_j(ds), \tag{7}$$

where

$$\gamma_j(ds)=\sum_{k=0}^{T-1}\Gamma^{k,k-j}(ds-2\pi k/T).$$

The matrix $[\Gamma^{k,j}(ds)]$ is the spectral measure of $\xi^q(n)$.

The family of measures $\gamma_j(ds)$ is called spectrum of $\zeta(n)$.

Proposition 1. (Gladyshev E. G.) *A sequence $\zeta(n)$ is PC with period T iff there is a T-dimensional stationary sequence $\xi^q(n)$ such that $\zeta(n)$ has representation (4)*

$$\zeta(n)=\sum_{q=0}^{T-1}e^{2\pi inq/T}\xi^q(n),\ n\in Z.$$

3 Estimates of functionals of periodically correlated discrete time stochastic process.

Consider the problem of optimal linear estimation of the functional $A\zeta = \sum_{j=0}^{\infty} a(j)\zeta(j)$ which depends on the unknown values of a periodically correlated (cyclostationary) discrete time stochastic process $\zeta(j)$ from observations of the process $\zeta(n)+\theta(n)$ for $n<0$, where $\theta(n)$ is uncorrelated with $\zeta(n)$ periodically correlated (cyclostationary) sequence.

The periodically correlated (cyclostationary) discrete time stochastic process $\zeta(j)$ can be represented in the form

$$\zeta(n) = \sum_{q=0}^{T-1} e^{2\pi inq/T}\xi^q(n),\ \ n \in Z.$$

where $\xi^q(n)$ is a T-dimensional stationary sequence (generating sequence of $\zeta(n)$).

The matrix spectral measure $[\Gamma^{k,j}]$ of $\xi^q(n)$ and the spectrum γ_j of $\zeta(n)$ are related by the formulas

$$\gamma_j(ds) = \sum_{k=0}^{T-1} \Gamma^{k,k-j}(ds - 2\pi k/T).$$

Using these relations we come to conclusion that the problem of estimation of the functional

$$A\zeta = \sum_{j=0}^{\infty} a(j)\zeta(j)$$

may be reduced to the problem of estimation of the functional

$$A\zeta = \sum_{j=0}^{\infty} a(j)\zeta(j) = \sum_{j=0}^{\infty} a(j) \sum_{q=0}^{T-1} e^{2\pi inq/T}\xi^q(j) =$$

$$= \sum_{j=0}^{\infty}\sum_{q=0}^{T-1} a_q(j)\xi^q(j) = \sum_{j=0}^{\infty} \vec{a}(j)\vec{\xi}(j) = A\vec{\xi},$$

which depends on values of multidimensional stationary stochastic process $\vec{\xi}(j) = \{\xi^q(j)\}_{q=0}^{T-1}$ with spectral matrix measure $[\Gamma^{k,j}(ds)]$.

Consider the problem of optimal linear estimation of the functional $A\vec{\xi} = \sum_{j=0}^{\infty} \vec{a}(j)\vec{\xi}(j)$ which depends on the unknown values of a multidimensional stationary sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^{T}$ with the spectral density matrix $F(\lambda) = \{f_{kl}(\lambda)\}_{k,l=1}^{T}$ from observations of the sequence $\vec{\xi}(n) + \vec{\eta}(n)$ for $n<0$, where $\vec{\eta}(n) = \{\eta_k(n)\}_{k=1}^{T}$ is uncorrelated with $\vec{\eta}(n)$ multidimensional stationary sequence with the spectral density matrix $G(\lambda) = \{g_{kl}(\lambda)\}_{k,l=1}^{T}$.

Let the vector sequence $\vec{a}(j)$, which determines the functional $A\vec{\xi}$, satisfies conditions

$$\sum_{j=0}^{\infty}\sum_{k=1}^{T} |a_k(j)| < \infty, \sum_{j=0}^{\infty} (j+1) \sum_{k=1}^{T} |a_k(j)|^2 < \infty. \tag{8}$$

and let the spectral densities $F(\lambda) = \{f_{kl}(\lambda)\}_{k,l=1}^{T}$ and $G(\lambda) = \{g_{kl}(\lambda)\}_{k,l=1}^{T}$ satisfy the minimality condition

$$\int_{-\pi}^{\pi} Tr\left[(F(\lambda)+G(\lambda))^{-1}\right] d\lambda < \infty. \tag{9}$$

Denote by $L_2(F)$ the Hilbert space of vector-valued functions $\varphi(\lambda) = \{\varphi_k(\lambda)\}_{k=1}^{T}$ integrable with respect to a measure with the density $F(\lambda) = \{f_{kl}(\lambda)\}_{k,l=1}^{T}$:

$$\int_{-\pi}^{\pi} \varphi(\lambda)F(\lambda)\varphi^*(\lambda)d\lambda$$

$$= \int_{-\pi}^{\pi} \sum_{k,l=1}^{T} \varphi_k(\lambda)\overline{\varphi_l(\lambda)}\, f_{kl}(\lambda)d\lambda < \infty.$$

Denote by $L_2^-(F)$ the subspace in $L_2(F)$ generated by functions $e^{ij\lambda}\delta_k$, $\delta_k = \{\delta_{kl}\}_{l=1}^{T}$, $k = \overline{1,T}$, $j \in Z, j < 0$, where $\delta_{kk} = 1$, $\delta_{kl} = 0$ for $k \neq l$.

A linear estimate $\hat{A}\vec{\xi}$ of the functional $A\vec{\xi}$ from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ for $j < 0$ has the form

$$\hat{A}\vec{\xi} = \int_{-\pi}^{\pi} h(e^{i\lambda})\left(Z^{\xi}(d\lambda) + Z^{\eta}(d\lambda)\right)$$

$$= \int_{-\pi}^{\pi} \sum_{k=1}^{T} h_k(e^{i\lambda})(Z_k^{\xi}(d\lambda) + Z_k^{\eta}(d\lambda)),$$

where $Z^{\xi}(\Delta) = \left\{Z_k^{\xi}(\Delta)\right\}_{k=1}^{T}$ and $Z^{\eta}(\Delta) = \left\{Z_k^{\eta}(\Delta)\right\}_{k=1}^{T}$ are orthogonal random measures of the sequences $\vec{\xi}(j)$ and $\vec{\eta}(j)$ correspondingly, and $h(\lambda) = \{h_k(\lambda)\}_{k=1}^{T}$ is the spectral characteristic of the estimate $\hat{A}\vec{\xi}$. The function $h(e^{i\lambda}) \in L_2^-(F+G)$.

The mean square error $\Delta(h; F, G)$ of the estimate $\hat{A}\vec{\xi}$ is calculated by the formulae

$$\Delta(h;F,G) = E\left|A\vec{\xi} - \hat{A}\vec{\xi}\right|^2 =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[A(e^{i\lambda}) - h(e^{i\lambda})\right] F(\lambda) \left[A(e^{i\lambda}) - h^*(e^{i\lambda})\right] d\lambda +$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi} h(e^{i\lambda})\, G(\lambda)\, h^*(e^{i\lambda})d\lambda, \quad A(e^{i\lambda}) = \sum_{j=0}^{\infty} \vec{a}(j)e^{ij\lambda}.$$

The spectral characteristic $h(F, G)$ of the optimal linear estimate of $A\vec{\xi}$ minimizes the mean square error

$$\Delta(F,G) = \Delta(h(F,G);F,G)$$

$$= \min_{h \in L_2^-(F+G)} \Delta(h;F,G) = \min_{\hat{A}\vec{\xi}} E\left|A\vec{\xi} - \hat{A}\vec{\xi}\right|^2. \tag{10}$$

The optimal estimate $\hat{A}\vec{\xi}$ is a solution of the optimization problem (10). With the help of the Hilbert space projection method proposed by A.N.Kolmogorov we can find that

$$h(F,G) = \left[A(e^{i\lambda})\,F(\lambda) - C(e^{i\lambda})\right]\,[F(\lambda)+G(\lambda)]^{-1}$$

$$= A(e^{i\lambda}) - \left[A(e^{i\lambda})\,G(\lambda) + C(e^{i\lambda})\right]\,[F(\lambda)+G(\lambda)]^{-1}, \tag{11}$$

$$\Delta(F,G) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[A(e^{i\lambda})\,G(\lambda) + C(e^{i\lambda})\right]\,[F(\lambda)+G(\lambda)]^{-1}\times$$

$$F(\lambda)\,[F(\lambda)+G(\lambda)]^{-1}\left[A(e^{i\lambda})\,G(\lambda) + C(e^{i\lambda})\right]^{*} d\lambda +$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left[A(e^{i\lambda})\,F(\lambda) - C(e^{i\lambda})\right][F(\lambda)+G(\lambda)]^{-1}\times$$

$$\times G(\lambda)\,[F(\lambda)+G(\lambda)]^{-1}\left[A(e^{i\lambda})\,G(\lambda) + C(e^{i\lambda})\right]^{*} d\lambda =$$

$$= \langle \vec{c},\, B\vec{c}\rangle + \langle \vec{a}\,,\, R\vec{a}\rangle, \tag{12}$$

where

$$C(e^{i\lambda}) = \sum_{j=0}^{\infty} \vec{c}(j)e^{ij\lambda}, \vec{c} = \{\vec{c}(k)\}_{k=0}^{\infty} = B^{-1}D\vec{a},$$

$\vec{a} = \{\vec{a}(k)\}_{k=0}^{\infty}$, $\langle a\,,\, b\rangle$ is the scalar product; B , D , R are matrices composed with block-matrices of dimension $T\times T$:

$$B(k\,,\,j) = \frac{1}{(2\pi)^T}\int_{-\pi}^{\pi} \left[(F(\lambda)+G(\lambda))^{-1}\right]^{T} e^{i(j-k)\lambda}d\lambda,$$

$$D(k\,,\,j) = \frac{1}{(2\pi)^T}\int_{-\pi}^{\pi} \left[F(\lambda)\,(F(\lambda)+G(\lambda))^{-1}\right]^{T} e^{i(j-k)\lambda}d\lambda,$$

$$R(k\,,\,j) = \frac{1}{(2\pi)^T}\int_{-\pi}^{\pi} \left[F(\lambda)\,(F(\lambda)+G(\lambda))^{-1}G(\lambda)\right]^{T} e^{i(j-k)\lambda}d\lambda,$$

$$k, j = 0, 1, \ldots.$$

Theorem 1. *Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^{T}$, $\vec{\eta}(j) = \{\eta_k(j)\}_{k=1}^{T}$ be uncorrelated stationary stochastic sequences with densities $F(\lambda) = \{f_{kl}(\lambda)\}_{k,l=1}^{T}$, $G(\lambda) = \{g_{kl}(\lambda)\}_{k,l=1}^{T}$, which satisfy the minimality condition (9). The spectral characteristics $h(F,G)$ and the mean square error $\Delta(F,G)$ of the optimal linear estimate of the functional $A\vec{\xi}$ on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(n)+\vec{\eta}(n), n<0$, are calculated by the formulas (11), (12).*

Corollary 1. *Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^{T}$ be a stationary stochastic sequence with the density $F(\lambda) = \{f_{kl}(\lambda)\}_{k,l=1}^{T}$, which satisfies the minimality condition*

$$\int_{-\pi}^{\pi} Tr\left[(F(\lambda))^{-1}\right] d\lambda \,<\, \infty.$$

The spectral characteristics $h(F)$ and the mean square error $\Delta(F)$ of the optimal linear estimate of the functional $A\vec{\xi}$ on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(j)$, $j<0$, are calculated by the formulas

$$h(F) = A(e^{i\lambda}) - C(e^{i\lambda})\,[F(\lambda)]^{-1}, \tag{13}$$

$$\Delta(F) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C(e^{i\lambda})\,[F(\lambda)]^{-1}C^*(e^{i\lambda})d\lambda = \\ = \left\langle B^{-1}\vec{a}, \vec{a}\right\rangle, \tag{14}$$

where $C(e^{i\lambda}) = \sum\limits_{j=0}^{\infty}\vec{c}(j)e^{ij\lambda}$, $\vec{c}=B^{-1}\vec{a}$, B is a matrix composed with block-matrices of dimension $T\times T$:

$$B(k,j) = \frac{1}{(2\pi)^T}\int_{-\pi}^{\pi}\left[(F(\lambda))^{-1}\right]^T e^{i(j-k)\lambda}d\lambda,$$
$$k,j=0,1,\dots$$

Let the sequence $\vec{\xi}(t)$ admits the canonical moving average representation

$$\vec{\xi}(j) = \sum_{u=-\infty}^{j} d(j-u)\vec{\varepsilon}(u), \tag{15}$$

where $d(k)=\{d_{ij}(k)\}_{i=\overline{1,T}}^{j=\overline{1,m}}$ is a matrix function and $\vec{\varepsilon}(u)=\{\varepsilon_k(u)\}_{k=1}^{m}$ is a multidimensional stationary stochastic sequence with uncorrelated values (white noise): $E\varepsilon_k(u)=0,\ E\left|\varepsilon_k(u)\right|^2 = 1,\ k=\overline{1,m},\ \ E\varepsilon_i(t)\overline{\varepsilon_j(s)}=0,\ t\neq s$. In this case the spectral density matrix $F(\lambda)=\{f_{ij}(\lambda)\}_{i,j=1}^{T}$ of the sequence $\vec{\xi}(t)$ admits the canonical factorization:

$$F(\lambda) = \varphi(\lambda)\,\varphi^*(\lambda), \quad \varphi(\lambda) = \sum_{k=0}^{\infty} d(k)e^{-ik\lambda}. \tag{16}$$

If the sequence $\vec{\xi}(t)$ admits the canonical moving average representation (15), then the optimal estimate of the functional $A\vec{\xi}=\sum_{j=0}^{\infty}\vec{a}(j)\vec{\xi}(j)$ from observations of the sequence $\vec{\xi}(n)$ for $n<0$ is determined by the spectral characteristic $h(F)\in L_2^-(F)$ that minimizes the mean square error

$$\Delta(h(F),F) = \min_{h\in L_2^-(F)}\Delta(h,F) = \left\|Ad\right\|^2, \tag{17}$$

where

$$\left\|Ad\right\|^2 = \sum_{k=0}^{\infty}\|(Ad)_k\|^2,$$

$$(Ad)_k = \sum_{l=k}^{\infty}\vec{a}(l)d(l-k).$$

Note, that $\left\|Ad\right\|^2<\infty$ under condition (1). The spectral characteristic $h(F)$ is calculated by the formula

$$h(F) = A(e^{i\lambda}) - r(e^{i\lambda})\psi(\lambda), \tag{18}$$

$$r(e^{i\lambda}) = \sum_{k=0}^{\infty} (Ad)_k e^{ik\lambda},$$

where $\psi(\lambda) = \{\psi_{ij}(\lambda)\}_{i=\overline{1,m}}^{j=\overline{1,T}}$ is a matrix function which satisfies the equation

$$\psi(\lambda)\varphi(\lambda) = I_m,$$

I_m is the identity matrix of order m.

For the functional $A_N\vec{\xi} = \sum_{j=0}^{N} \vec{a}(j)\vec{\xi}(j)$ the value of the mean square error and the spectral characteristic of the optimal estimate are determined by the following formulas

$$\Delta_N(h(F), F) = \|A_N d\|^2; \tag{19}$$

$$h(F) = A_N(e^{i\lambda}) - r_N(e^{i\lambda})\psi(\lambda), \tag{20}$$

$$r_N(e^{i\lambda}) = \sum_{k=0}^{N} (A_N d)_k e^{ik\lambda},$$

where

$$A_N(e^{i\lambda}) = \sum_{j=0}^{N} \vec{a}(j)e^{ij\lambda}, \quad \|A_N d\|^2 = \sum_{k=0}^{N} \|(A_N d)_k\|^2,$$

$$(A_N d)_k = \sum_{l=k}^{N} \vec{a}(l)d(l-k), k = \overline{0,N}.$$

As a corollary we can get the following formulas for calculation the mean square error of the optimal estimate $\widehat{\xi}_k(j)$ of the unknown values $\xi_k(j)$, $k = \overline{1,T}$, $j = \overline{0,N}$:

$$E\left|\xi_k(j) - \widehat{\xi}_k(j)\right|^2 = \sum_{l=0}^{j} \|d_k(l)\|^2, \tag{21}$$

where $d_k(l)$ is the k-th row of the matrix $d(l)$ determined by the factorization (16) of the spectral density matrix $F(\lambda)$.

Theorem 2. *Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^{T}$, $E\vec{\xi}(j) = 0$, be a stationary stochastic sequence that admits the canonical moving average representation (15) with the spectral density matrix $F(\lambda)$ that admits the canonical factorization (16) and let condition (8) be satisfied. The value of the mean-square error $\Delta(h(F), F)$ of the optimal linear estimate of the functional $A\vec{\xi}$ from observations of the sequence $\vec{\xi}(n)$ for $n < 0$ can be calculated by formula (17) (by formula (19) if the functional $A_N\vec{\xi}$ is estimated). The spectral characteristics $h(F)$ of the optimal linear estimate can be calculated by formula (18) (by formula (20) if the functional $A_N\vec{\xi}$ is estimated).*

4 Minimax-robust method of analysis.

The proposed formulas may be employed only under the condition that spectral densities of stochastic sequences are known. In the case where the densities $F(\lambda)$, $G(\lambda)$ are not known exactly, but a set $D = D_F \times D_G$ of possible spectral densities is given, the minimax (robust) approach to estimation of functionals of the unknown values of stationary sequences is reasonable. Instead of searching an estimate that is optimal for a given spectral densities $F(\lambda)$ and $G(\lambda)$

we find an estimate that minimizes the mean square error for all spectral densities from a given class simultaneously.

Definition 2. For a given class of pairs of spectral densities $D = D_F \times D_G$ spectral densities $F^0(\lambda) \in D_F,\ G^0(\lambda) \in D_G$ are called the least favorable in $D = D_F \times D_G$ for the optimal linear estimation of the functional $A\vec{\xi}$ if the following relations holds true:

$$\Delta\left(F^0, G^0\right) = \Delta\left(h\left(F^0, G^0\right); F^0, G^0\right)$$
$$= \max_{(F,G)\in D} \Delta\left(h\left(F, G\right); F, G\right).$$

Definition 3. For a given class of pairs of spectral densities $D = D_F \times D_G$ the spectral characteristic $h^0(\lambda)$ of the optimal linear estimate of the functional $A\vec{\xi}$ is called minimax (robust) if

$$h^0(\lambda) \in H_D = \bigcap_{(F,G)\in D} L_2^-\left(F+G\right),$$

$$\min_{h\in H_D}\max_{(F,G)\in D} \Delta\left(h; F, G\right) = \max_{(F,G)\in D} \Delta\left(h^0; F, G\right).$$

Lemma 1. *Spectral densities $F^0(\lambda),\ G^0(\lambda)$ are the least favorable in $D = D_F \times D_G$ for the optimal linear estimation of the functional $A\vec{\xi}$, if the Fourier coefficients of the matrix functions*

$$(F^0(\lambda) + G^0(\lambda))^{-1},$$
$$F^0(\lambda)(F^0(\lambda) + G^0(\lambda))^{-1},$$
$$F^0(\lambda)(F^0(\lambda) + G^0(\lambda))^{-1}G^0(\lambda)$$

determine the matrices B^0, D^0, R^0, which give a solution to the extremum problem

$$\max_{(F,G)\in D}\left\langle B^{-1}D\vec{a}, D\vec{a}\right\rangle + \left\langle \vec{a}, R\vec{a}\right\rangle =$$
$$= \left\langle (B^0)^{-1}D^0\vec{a}, D^0\vec{a}\right\rangle + \left\langle \vec{a}, R^0\vec{a}\right\rangle. \quad (22)$$

Minimax spectral characteristics $h^0 = h(F^0, G^0)$ is calculated by formula (11) if the condition $h(F^0, G^0) \in H_D$ holds true.

The least favorable spectral densities $F^0(\lambda),\ G^0(\lambda)$ and the minimax (robust) spectral characteristics $h^0 = h\left(F^0,\ G^0\right)$ form a saddle point of the function $\Delta\left(h;\ F,\ G\right)$ on the set $H_D \times D$. The saddle point inequalities

$$\Delta\left(h^0; F, G\right) \le \Delta\left(h^0; F^0, G^0\right) \le \Delta\left(h; F^0, G^0\right)$$
$$\forall h \in H_D \quad \forall F \in D_F \quad \forall G \in D_G$$

hold when $h^0 = h(F^0, G^0)$, $h\left(F^0,\ G^0\right) \in H_D$ and $\left(F^0,\ G^0\right)$ is a solution to the conditional extremum problem

$$\sup_{(F,G)\in D} \Delta\left(h\left(F^0, G^0\right); F, G\right) = \Delta\left(h\left(F^0, G^0\right); F^0, G^0\right), \quad (23)$$

$$\Delta\left(h\left(F^0, G^0\right); F, G\right) =$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(A(e^{i\lambda})\, G^0(\lambda) + C^0(e^{i\lambda})\right)\left(F^0(\lambda) + G^0(\lambda)\right)^{-1} F(\lambda)\times$$

$$(F^0(\lambda)+G^0(\lambda))^{-1}(A(e^{i\lambda})\,G^0(\lambda)+C^0_N(e^{i\lambda}))^*d\lambda$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi}(A(e^{i\lambda})\,F^0(\lambda)-C^0(e^{i\lambda}))(F^0(\lambda)+G^0(\lambda))^{-1}\times$$
$$\times G(\lambda)\,(F^0(\lambda)+G^0(\lambda))^{-1}(A(e^{i\lambda})\,G^0(\lambda)+C^0(e^{i\lambda}))^*d\lambda.$$

Lemma 2. *Let $\left(F^0,\ G^0\right)$ be a solution to extremum problem (22). Spectral densities $F^0\left(\lambda\right),\ G^0\left(\lambda\right)$ are the least favorable in the class $D=D_F\times D_G$, v $h^0=h(F^0,G^0)$ is minimax for the optimal estimate of $A\vec{\xi}$, if $h(F^0,G^0)\in H_D$.*

Taking into account relations (1)–(16), it is possible to verify the following lemmas.

Lemma 3. *The spectral density matrix $F^0(\lambda)\in D$ is the least favorable in the class D for the optimal linear estimation of the functional $A\vec{\xi}$ from observations of the sequence $\vec{\xi}(j)$ for $j<0$ if $F^0(\lambda)$ admits the canonical factorization*

$$F^0(\lambda)=\left(\sum_{k=0}^{\infty}d^0(k)e^{-ik\lambda}\right)\cdot\left(\sum_{k=0}^{\infty}d^0(k)e^{-ik\lambda}\right)^*, \tag{24}$$

where $d^0=\left\{d^0(k):k=0,1,\dots\right\}$ is a solution to the conditional extremum problem

$$\|Ad\|^2\to\max, \tag{25}$$

$$F(\lambda)=\left(\sum_{k=0}^{\infty}d(k)e^{-ik\lambda}\right)\cdot\left(\sum_{k=0}^{\infty}d(k)e^{-ik\lambda}\right)^*\in D.$$

The sequence $\vec{\xi}(j)$ in this case admits the canonical moving average representation

$$\vec{\xi}(j)=\sum_{u=-\infty}^{j}d(j-u)\vec{\varepsilon}(u). \tag{26}$$

The minimax spectral characteristic $h^0=h(F^0)$ of the optimal estimate of the functional is calculated by the formula (18) under the condition $h(F^0)\in H_D$.

Lemma 4. *The spectral density $F^0(\lambda)\in D$ is the least favorable in the class D for the optimal linear estimation of the functional $A_N\vec{\xi}$ from observations of the sequence $\vec{\xi}(j)$ for $j<0$ if $F^0(\lambda)$ admits the canonical factorization*

$$F^0(\lambda)=\left(\sum_{k=0}^{N}d^0(k)e^{-ik\lambda}\right)\cdot\left(\sum_{k=0}^{N}d^0(k)e^{-ik\lambda}\right)^*, \tag{27}$$

where $d^0=\left\{d^0(k):0\leq k\leq N\right\}$ is a solution to the conditional extremum problem

$$\|A_Nd\|^2\to\max, \tag{28}$$

$$F(\lambda)=\left(\sum_{k=0}^{N}d(k)e^{-ik\lambda}\right)\cdot\left(\sum_{k=0}^{N}d(k)e^{-ik\lambda}\right)^*\in D.$$

The sequence $\vec{\xi}(j)$ in this case admits the canonical moving average representation of the order N:

$$\vec{\xi}(j) = \sum_{u=j-N}^{j} d(j-u)\vec{\varepsilon}(u). \tag{29}$$

The minimax spectral characteristics $h^0 = h(F^0)$ of the optimal estimate of the functional is calculated by the formula (20) under the condition $h(F^0) \in H_D$.

5 Least favorable spectral densities in the class D_0.

Consider the problem for the set of spectral density matrices

$$D_0 = \left\{ F(\lambda) \colon \ \frac{1}{(2\pi)^T} \int_{-\pi}^{\pi} F(\lambda) d\lambda = P \right\}.$$

With the help of the Lagrange multipliers method we can find that solutions to the conditional extremum problem (23) that determine the least favorable density matrix $F^0(\lambda) \in D_0$ of the maximal rank is of the form

$$F^0(\lambda) = \vec{\beta} \, (\sum_{k=0}^{\infty} (Ad)_k e^{ik\lambda}) \cdot (\sum_{k=0}^{\infty} (Ad)_k e^{ik\lambda})^* \, \vec{\beta}^*. \tag{30}$$

The unknown $\vec{\beta} = (\beta_1, \ldots, \beta_T)^T$, $d = \{d(k) : k = 0, 1, \ldots\}$ are determined by the canonical factorization (16) of the density matrix $F^0(\lambda)$, condition (25) and the condition

$$\frac{1}{(2\pi)^T} \int_{-\pi}^{\pi} F(\lambda) d\lambda = P. \tag{31}$$

For solutions $d = \{d(k) : k = 0,\, 1,\, \ldots\}$ to the system of equations

$$(Ad)_k = \vec{c} d^*(k), \quad k \geq 0, \quad \vec{c} = (c_1, \ldots, c_T), \tag{32}$$

such that

$$\|d\|^2 = \sum_{k=0}^{\infty} \|d(k)\|^2 = \sum_{k=0}^{\infty} \sum_{i,j} |d_{ij}(k)|^2 = P,$$

the following equality holds true

$$F(\lambda) = (\sum_{k=0}^{\infty} d(k) e^{-ik\lambda})(\sum_{k=0}^{\infty} d(k) e^{-ik\lambda})^* =$$

$$\vec{\beta}(\sum_{k=0}^{\infty} (Ad)_k e^{ik\lambda})(\sum_{k=0}^{\infty} (Ad)_k e^{ik\lambda})^* \vec{\beta}^*.$$

Denote by ν_0^P the maximal value of $\|Ad\|^2$, where d are solutions to equation (32), that satisfy condition (31) and determine the canonical factorization (16) of the density matrix $F(\lambda)$, $F(\lambda) \in$

D_0. Denote by μ_0^P the maximal value of $\|Ad\|^2$, where d, $\|d\|^2 = P$, determine the canonical factorization (16) of the density matrix $F^0(\lambda)$ of the form (30). If there exists a solution d^0 to the equation (22) such that $\left\|d^0\right\|^2 = P$ and $\nu_0^P = \mu_0^P = \left\|Ad^0\right\|^2$, then the least favorable in D_0 is the density matrix

$$F^0(\lambda) = (\sum_{k=0}^{\infty} d^0(k)e^{-ik\lambda}) \cdot (\sum_{k=0}^{\infty} d^0(k)e^{-ik\lambda})^*. \tag{33}$$

The stationary sequence $\vec{\xi}(j)$ in this case admits the moving average representation (26). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional $A\vec{\xi}$ is calculated by formula (18) since the functions $A(\lambda)$ and $r(\lambda)$ are bounded and $h(F^0) \in H_D$.

The following theorem holds true.

Theorem 3. *If there exists a solution*

$$d^0 = \left\{d^0(k) : k = 0,\, 1,\, \ldots\right\}$$

to the equation (32) that satisfy condition (31) and such that $\nu_0^P = \mu_0^P = \left\|Ad^0\right\|^2$, then the least favorable in D_0 for the optimal linear estimation of the functional $A\vec{\xi}$ is the density matrix (33). The corresponding stationary sequence $\vec{\xi}(j)$ in this case admits the moving average representation (26). If $\nu_0^P < \mu_0^P$, then the density matrix (30), that admits the canonical factorization (33), is the least favorable in D_0. The vector $\vec{\beta}$ and the sequence $d^0 = \left\{d^0(k) : k = 0,\, 1,\, \ldots\right\}$ are determined by conditions (25), (31). The minimax spectral characteristics $h(F)$ of the optimal linear estimate is calculated by formula (18).

For the functional $A_N\vec{\xi}$ the density matrix (16) is of the form

$$F^0(\lambda) = \vec{\beta}\, (\sum_{k=0}^{N} (A_N d)_k e^{ik\lambda}) \cdot (\sum_{k=0}^{N} (A_N d)_k e^{ik\lambda})^*\, \vec{\beta}^* \tag{34}$$

In this case the equality holds true

$$r_N(e^{i\lambda}) r_N(e^{i\lambda})^* = (\sum_{k=0}^{N} (A_N d)_k e^{ik\lambda})(\sum_{k=0}^{N} (A_N d)_k e^{ik\lambda})^* =$$

$$= (\sum_{k=0}^{N} (\tilde{A}_N d)_k e^{ik\lambda})(\sum_{k=0}^{N} (\tilde{A}_N d)_k e^{ik\lambda})^*,$$

where

$$(\tilde{A}_N d)_k = \sum_{u=0}^{k} \vec{a}(N-k+u) d(u),\ 0 \le k \le N.$$

For these reason all solutions $d = \{d(k) : 0 \le k \le N\}$ to equations

$$(A_N d)_k = \vec{c} d(k)^*, 0 \le k \le N, \vec{c} = (c_1, \ldots, c_T); \tag{35}$$

$$(\tilde{A}_N d)_k = \vec{b} d(k)^*, 0 \le k \le N, \vec{b} = (b_1, \ldots, b_T), \tag{36}$$

such that $\|d\|^2 = P$ the equality holds true

$$F(\lambda) = (\sum_{k=0}^{N} d(k)e^{-ik\lambda})(\sum_{k=0}^{N} d(k)e^{-ik\lambda})^*$$

$$= \vec{\beta}\, r_N(e^{i\lambda}) r_N(e^{i\lambda})^* \vec{\beta}^*.$$

Denote by ν_0^{NP} the maximal value of $\|A_N d\|^2 = \left\|\tilde{A}_N d\right\|^2$, where d are solutions to equations (35), (36) that satisfy condition (31) and determine the canonical factorization (16) of the density matrix $F^0(\lambda)$. Denote by μ_0^{LP} the maximal value of $\|A_N d\|^2$, where d satisfy condition (31) and determine the canonical factorization (16) of the density matrix $F^0(\lambda)$ with $F^0(\lambda)$ of the form (34). If there exists a solution d^0 to equation (35), or equation (36), that satisfy condition (31) and such that $\left\|d^0\right\|^2 = P$ and $\nu_0^{LP} = \mu_0^{LP} = \left\|A_N d^0\right\|^2$, then the least favorable in D_0 is the density matrix

$$F^0(\lambda) = (\sum_{k=0}^{N} d^0(k)e^{-ik\lambda})(\sum_{k=0}^{N} d^0(k)e^{-ik\lambda})^*. \tag{37}$$

The stationary sequence $\vec{\xi}(j)$ in this case admits the moving average representation (29).

The following theorem holds true.

Theorem 4. *If there exists a sequence*

$$d^0 = \left\{d^0(k) : k = 0, 1, \ldots, N\right\}$$

that satisfy condition $\left\|d^0\right\|^2 = P$ *and such that* $\nu_0^{LP} = \mu_0^{LP} = \left\|A_L d^0\right\|^2$, *then the least favorable in* D_0 *for the optimal linear estimation of the functional* $A_N\vec{\xi}$ *is the density matrix (37). The corresponding stationary sequence* $\vec{\xi}(j)$ *in this case admits the moving average representation (29). If* $\nu_0^{LP} < \mu_0^{LP}$, *then the density matrix (34), that admits the canonical factorization (27), is the least favorable in* D_0. *The vector* $\vec{\beta}$ *and the sequence* $d^0 = \left\{d^0(k) : k = 0, 1, \ldots, N\right\}$ *are determined by conditions (28), (31). The minimax spectral characteristics* $h(F)$ *of the optimal linear estimate of the functional* $A_L\vec{\xi}$ *is calculated by formula (20).*

6 Conclusion.

We propose formulas for calculation the mean square error and the spectral characteristic of the optimal linear prediction of the unknown value of the functional $A\xi = \sum_{j=0}^{\infty} \vec{a}(j)\vec{\xi}(j)$ which depends on the unknown values of a multidimensional stationary stochastic process $\vec{\xi}(j)$ from observations of the process $\vec{\xi}(n) + \vec{\eta}(n)$ for $n = -1, -2, \ldots$ under the condition that spectral density matrices $F(\lambda)$ and $G(\lambda)$ of the signal process $\vec{\xi}(n)$ and the noise process $\vec{\eta}(n)$ are known. Formulas are proposed that determine the least favorable spectral densities and the minimax-robust spectral characteristics of the optimal linear prediction of the functional for concrete classes $D = D_F \times D_G$ of spectral densities under the condition that spectral density matrices $F(\lambda), G(\lambda)$ are not known, but classes $D = D_F \times D_G$ of admissible spectral densities are given. These formulas give us a possibility to solve the corresponding estimation problem for periodically correlated (cyclostationary) discrete time stochastic processes.

References

[1] Gladyshev, E. G., *Periodically correlated random sequences*, Soviet Math. Dokl., **2** (1961), 385 – 389.

[2] Franke, J., *On the robust prediction and interpolation of time series in the presence of correlated noise*, J. Time Series Analysis, **5** (1984), no. 4, 227 – 244.

[3] Franke, J., *Minimax robust prediction of discrete time series*, Z. Wahrsch. Verw. Gebiete. **68** (1985), 337 – 364.

[4] Franke, J. and Poor, H. V., *Minimax–robust filtering and finite–length robust predictors*, In Robust and Nonlinear Time Series Analysis (Heidelberg, 1983) Lecture Notes in Statistics, Springer-Verlag, **26** (1984), 87 – 126.

[5] Franke, J., *A general version of Breiman's minimax filter*, Note di Matematica, **11** (1991), 157 – 175.

[6] Grenander, U., *A prediction problem in game theory*, Ark. Mat., **3** (1957), 371 – 379.

[7] Kailath, T., *A view of three decades of linear filtering theory*, IEEE Trans. on Inform. Theory, **20** (1974), no. 2, 146 – 181.

[8] Kolmogorov, A. N., *Selected works of A. N. Kolmogorov. Vol. II: Probability theory and mathematical statistics.*, Ed. by A. N. Shiryayev. Mathematics and Its Applications. Soviet Series. 26. Dordrecht etc.: Kluwer Academic Publishers. 1992.

[9] Kassam, S. A. and Poor, H. V. *Robust techniques for signal processing: A survey*, Proc. IEEE, **73** (1985), no. 3, 433 – 481.

[10] Makagon, A., *Theoretical prediction of periodically correlated sequences*, Probab. Math. Stat., **19** (1999), no. 2, 287 – 322.

[11] Moklyachuk, M. P., *Stochastic autoregressive sequence and minimax interpolation*, Theor. Probab. and Math. Stat., **48**, (1994), 95 – 103.

[12] Moklyachuk, M. P., *Estimates of stochastic processes from observations with noise*, Theory Stoch. Process., **3(19)** (1997), no.3-4, 330 – 338.

[13] Moklyachuk, M. P., *Extrapolation of stationary sequences from observations with noise*, Theor. Probab. and Math. Stat., **57** (1998), 133 – 141.

[14] Moklyachuk, M. P., *Robust procedures in time series analysis*, Theory Stoch. Process., **6(22)** (2000), no.3-4, 127 – 147.

[15] Moklyachuk, M. P., *Game theory and convex optimization methods in robust estimation problems*, Theory Stoch. Process., **7(23)** (2001), no.1-2, 253-264.

[16] Pshenichnyi, B. N., *Necessary conditions for an extremum*, 2nd ed. Moscow, "Nauka", 1982.

[17] Rozanov, Yu. A., *Stationary stochastic processes*, 2nd rev. ed. Moscow, "Nauka", 1990. (English transl. of 1st ed., Holden-Day, San Francisco, 1967)

[18] Vastola, K. S. and Poor, H. V., *An analysis of the effects of spectral uncertainty on Wiener filtering*, Automatica, **28** (1983), 289 – 293.

[19] Wiener, N., *Extrapolation, interpolation, and smoothing of stationary time series. With engineering applications*, Cambridge, Mass.: The M. I. T. Press, Massachusetts Institute of Technology. 1966.

[20] Yaglom, A.M., *Correlation theory of stationary and related random functions. Vol. I: Basic results.* Springer Series in Statistics. New York etc.: Springer-Verlag. 1987.

[21] Yaglom, A. M., *Correlation theory of stationary and related random functions. Vol. II: Supplementary notes and references.* Springer Series in Statistics. New York etc.: Springer-Verlag. 1987.

Proximity of barrier option prices in discrete and continuous time

Olena Soloveiko and Georgiy Shevchenko
osoloveyko@univ.kiev.ua zhora@univ.kiev.ua
Department of Probability Theory & Mathematical Statistics
Kyiv National Taras Shevchenko University, Ukraine

Abstract

We estimate the difference between barrier option prices in a continuous time market model and in a discrete time binomial market model. As an auxilliary result, we estimate the difference between barrier option prices in a continuous time market and in a discrete time Gaussian market.

1 Introduction

A barrier option is a derivative with a payoff that depends on the fact whether asset price crosses certain level during certain time interval. Thus, payment for barrier option depends on the behavior of the price asset during all the time interval, i.e. barrier option is a particular case of exotic option.

The simplest barrier options are calls and puts that are knocked out or knocked in by the underlying asset itself. The payoff of a knock-out option is made if underlying asset price does not cross the barrier, such options are of two types: if asset price does not cross the barrier below, then such an option is called "up-and-out", if from above – "down-and-out". Payoff of a knock-in option is made if underlying asset price crosses the barrier, they also are of two types accordingly: "up-and-in" and "down-and-in". Altogether there are eight types of barrier options.

For example, the payoff function of up-and-in option is given by

$$C = \begin{cases} (S_T - K)^+, & \text{if } \max_{0\le t\le T} S_t \ge H, \\ 0 & \text{else,} \end{cases}$$

where H is a barrier level ($H > S_0$ and $H > K$), K is a strike price. Payoffs for the rest options are determined in the same way. Barrier options are among the most popular path-dependent option traded in exchanges worldwide and also over-the-counter markets.

The problem of pricing and hedging barrier option in the models with continuous time is rather complete, and analytical formulae for the prices of such options are known only in the simplest cases. Therefore the problem of asymptotic estimation of the prices of such options arises. The simplest asymptotic methods is the method of time discretization, which can be described in the following way. Time interval is divided into m equal parts and now the asset price model with discrete time is considered. In such a formulation we can approximately calculate option price using Monte Carlo simulations, modelling the path of the underlying asset price. From the other side, the opposite problem could arise: let we have analytical formula for option price in continuous time model. Then the demand may come to estimate the price of the option with payoff realized when the asset price crosses the barrier level, and this price is observed only in certain time moments (for example, daily when stock exchange is closing).

From the practical point of view, when we approximately estimate the price of the option it is important to know the quality of such an estimation, i.e. the order of the error.

In [1] authors introduce a simple continuity correction for approximate pricing of discrete barrier option. Their method uses formulae for the prices of continuous barrier options but shifts the barrier to correct for discrete monitoring. Compared with using the unadjusted continuous price, their formula reduces the error from $O(\frac{1}{\sqrt{m}})$ to $o(\frac{1}{\sqrt{m}})$, as the number of monitoring points m increases. The correction is justified both theoretically and experimentally.

Theorem 1.1. *[1] Let $V(H)$ be the price of a continuously monitored knock-in or knock-out down call or put with barrier H, and let $V_m(H)$ be the price of the corresponding discrete monitored barrier option. Then*

$$V_m(H) = V(He^{\pm\beta\sigma\sqrt{T/m}}) + o(\frac{1}{\sqrt{m}}),$$

where $+$ applies if $H > S_0$, and $-$ applies if $H < S_0$, $\beta = -\zeta(1/2)/\sqrt{2\pi} \approx 0.5826$, with ζ the Riemann zeta function.

The paper [6] extends an approximation by Broadie et al. in [1] for discretely monitored barrier options by covering more cases and giving a simpler proof. The paper [4] also continues the work of Broadie and determine formulae to estimate the price of discrete up-and-out/in calls, down-and-out/in puts and double barrier option. The methods used here lead to slightly different barrier correction formulae. In [2] the rate of convergence for lookback options and other exotic options is obtained.

The model considered in [8] investigates the rate of convergence of option price in discrete market, but this price is not fair in the sense that it might be not unique. Discrete market, generated by the increments of geometric Brownian motion, is not complete, so there are many “fair prices”. Thus it would be better to have result for convergence of the unique price in complete market. That because in our work we consider discrete binomial market and investigate the rate of convergence of fair price of barrier option in such market to correspondent price on continuous market. We have proved that the rate of convergence is $O(\ln n/\sqrt{n})$, where n is the number of periods in the binomial market.

2 Main result

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with filtration $\{\mathcal{F}_t, t \geq 0\}$, $\{W_t, t \geq 0\}$ is standard $\mathcal{F}_t$-Brownian motion on it. Consider Black–Scholes financial market model, where we have two assets: riskless (bond), whose price at the moment t equals

$$B_t = B_0 \exp\Big\{\int_0^t r_s ds\Big\},$$

and a risky asset (stock), whose price is

$$S_t = S_0 \exp\Big\{ \int_0^t \mu_s ds + \sigma W_t \Big\},$$

where W_t is standard Brownian motion defined before. Volatility $\sigma > 0$ is assumed to be constant. For simplicity, we assume that P itself is a martingale measure for discounting process of risky asset price, i.e. $\mu_t = r_t - \sigma^2/2$.

Besides, we demand the interest rate r_t to be Lipschitz continuous, i.e. for every $t, s \in [0, T]$

$$|r_t - r_s| \le C|t - s|, \tag{2.1}$$

where C is a constant.

In the market with continuous time the fair option price is defined as the expectation of discounting payoff for the option given martingale measure. Let I_A denote the indicator of an event A, $M_T = \max\{S_t, t \in [0, T]\}$, $m_t = \min\{S_t, t \in [0, T]\}$. Then, for instance, European up-and-out call option price is given by

$$V(H) = E\Big(\exp\Big\{ - \int_0^T r_t dt \Big\} (S_T - K)^+ I_{\{M_T < H\}} \Big),$$

where $K > 0$ is a strike price, $H > S_0$ is a barrier, and European down-and-in put option price is given by

$$V(H) = E\Big(\exp\Big\{ - \int_0^T r_t dt \Big\} (K - S_T)^+ I_{\{m_T \le H\}} \Big),$$

where $H < S_0$ is a barrier. In Merton's paper [7] an explicit form for the price of knock-out call option is established, when the risk-neutral interest rate r is constant.

Now consider a binomial market model with discrete time, which is constructed as follows. Divide time interval $[0, T]$ into $n \ge 1$ parts, define $\Delta = \frac{T}{n}$, $t_i = i\Delta$, $i = 0, \dots, n$. Let ξ_i, $i = 0, \dots, n-1$ be independent identically distributed random variables, such that $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$. The risky asset price in the binomial market model is defined as

$$S^b_{t_i} = S_0 \exp\Big\{ \sum_{j=0}^{i-1} (\mu_j \Delta + \sigma \xi_j \sqrt{\Delta}) \Big\}, i = 1, \dots, n;$$

on $[t_i, t_{i+1})$ we put $S^b_t = S^b_{t_i}$, and set the interest rate to be equal to r_{t_i}. Instead of Brownian motion, the role of "random driver" of financial market in the binomial model is played by a random walk $\{\Xi_i\}$, defined as

$$\Xi_i := \sum_{j=0}^{i-1} \xi_j.$$

An analogue of European up-and-out call option in the binomial model has the payoff function $(S^b_T - K)^+ I_{\{M^b_T < H\}}$, consequently, the price is

$$V^b_n(H) = E\Big(\exp\Big\{ - \sum_{i=0}^{n-1} r_{t_i} \Delta \Big\} (S^b_T - K)^+ I_{\{M^b_T < H\}} \Big),$$

where $M_T^b = \max_{0\le i\le n} S_{t_i}^b = \max_{t\in[0,T]} S_t^b$.

The following is the main result about convergence of price in binomial model to the one in continuous model.

Theorem 2.1. *The difference of European up-and-out call options fair prices in discrete binomial and continuous models under the assumption* (2.1) *satisfies*

$$V(H) - V_n^b(H) = O\Big(\frac{\ln n}{\sqrt{n}}\Big),\ n\to\infty.$$

In order to prove this result, we will use another approximation result for convergence of European up-and-out call options fair prices in discrete and continuous models.

In the discrete time market define a discretized version of S:

$$S_{t_i}^d = S_0 \exp\Big\{\sum_{j=0}^{i-1}(\mu_j\Delta + \sigma Z_j\sqrt{\Delta})\Big\}, i = 1,\dots,n,$$

where $Z_j = (W_{t_{j+1}} - W_{t_j})/\sqrt{\Delta}$, and consider European up-and-out call option with a payoff $(S_T^d - K)_+ I_{\{M_T^d<H\}}$, where $M_T^d = \max_{0\le i\le n} S_{t_i}^d$. Its fair price is

$$V_n^d(H) = E\bigg(\exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\}(S_T^d - K)^+ I_{\{M_T^d<H\}}\bigg).$$

Theorem 2.2. *The difference of European up-and-out call options fair prices in discrete and continuous models under the assumption* (2.1) *satisfies*

$$V_n(H) - V(H) = O\Big(\frac{1}{\sqrt{n}}\Big),\ n\to\infty.$$

We prove theorem 2.2 first.

Proof. In the following C will denote a generic positive constant, which may depend only on σ, the Lipschitz continuity parameter of r_t, H, K, S_0, i.e. the inputs of our problem.

Whereas joint distribution of $\Delta W_1,\dots,\Delta W_m$ is same as of $W_{t_1},\dots,W_{t_m}$, we can suppose that $W_n = W_{t_n}$, it makes our calculations easier.

For any process $Y(t)$ define $\tau(x,Y) := \inf\{t\ge 0 : Y(t)\ge x\}$, i.e. $\tau(H,S) = \tau_H$ is the first moment t, in which S_t reaches the level H, and $\tau(H,S_n) = \tau_H^n$ is the first moment t_i, in which $S_{t_i}^d$ reaches the level H.

Consider the difference of European up-and-out call options fair prices in discrete and continuous models to find the rate of convergence of their prices in the models.

$$V_n(H) - V(H) = \exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\} E\Big(\big((S_n^d - K)_+ - (S_T - K)_+\big) I_{\{\tau_H^n > T\}}\Big) +$$
$$\exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\} E\Big((S_T - K)^+ \big(I_{\{\tau_H^n > T\}} - I_{\{\tau_H > T\}}\big)\Big) +$$
$$E\big((S_T - K)^+ I_{\{\tau_H^n > T\}}\big)\Bigg(\exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\} - \exp\Big\{-\int_0^T r_t dt\Big\}\Bigg) =$$
$$\delta_1^n + \exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\} E\Big((S_T - K)^+ \big(I_{\{\tau_H^n > T\}} - I_{\{\tau_H > T\}}\big)\Big) + \delta_2^n,$$

where, using the assumption about the function r,

$$|\delta_2^n| \le (H - K)\left|\Bigg(\exp\Big\{-\int_0^T r_t dt\Big\} - \exp\Big\{-\sum_{i=0}^{n-1} r_{t_i}\Delta\Big\}\Bigg) P(\tau_H^n > T)\right| \le$$
$$C\left|\exp\Big\{-\int_0^T r_t dt + \sum_{i=0}^{m-1} r_{t_i}\Delta\Big\} - 1\right|.$$

For any x inequality $|e^x - 1| \le |x| e^{|x|}$ is true. So, we obtain

$$|\delta_2^n| \le C\left| -\int_0^T r_t dt + \sum_{i=0}^{n-1} r_{t_i}\Delta\right| \le C\left|\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} (r_t - r_{t_i}) dt\right| \le$$
$$C\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} |r_t - r_{t_i}|\, dt \le C\Delta \le C\sqrt{\Delta}.$$

To estimate the value of δ_1^n we should notice that S is a solution of stochastic differential equation $dS_t = r_t dt + \sigma dW_t$, and S_n^d is Euler approximation for the value S_T of such a solution. Then, using well known estimation for mathematical expectation of their difference (see, for ex., [9]) and whereas r is bounded, we have that

$$|\delta_1^n| \le CE\left|S_T - S_n^d\right| \le C\sqrt{\Delta}.$$

Therefore we obtain

$$|V_n^d(H) - V(H)| \le C\left|E\Big((H - K)\big(I_{\{\tau_H^n > T\}} - I_{\{\tau_H > T\}}\big)\Big)\right| + C\sqrt{\Delta} =$$
$$C\left|E\big(I_{\{\tau_H^n > T\}} - I_{\{\tau_H > T\}}\big)\right| + C\sqrt{\Delta}.$$

To estimate the first item, we will use the Girsanov formula. For this we will need the following notation. For a bounded measurable function $g : [0, T] \to R$ define $W_t(g) = W_t + \int_0^t \frac{g(s)}{\sigma} ds$ is Brownian motion with the drift g/σ,

$$\mathcal{E}(g) = \exp\Big\{-\int_0^T \frac{g(s)}{\sigma} dW_s - \frac{1}{2}\int_0^T \frac{g^2(s)}{\sigma^2} ds\Big\}$$

is the density of the martingale measure P^g for $W(g)$, $E^g(\cdot) = E(\mathcal{E}(g)\cdot)$ is the mathematical expectation given this measure; for the process X we denote

$$X^* = \sup_{[0,T]} X_t, X_n^* = \max_{1\le i\le n} X_{t_i};$$

and also denote

$$\mu_t^n = \sum_{i=1}^n \mu_{t_i} I_{t\in[t_i,t_{i+1})}, \quad L = \frac{1}{\sigma}\ln(H/S_0), \quad S'_n = \max_{1\le i\le n} S_i^d.$$

According to the Girsanov theorem $W(g)$ and $\mathcal{E}^{-1}(g)$ have the same joint distributions with respect to P^g, the same as W and $\mathcal{E}(-g)$ have with respect to P. So, we can write

$$\begin{gathered}\Big|E\big(I_{\{\tau_H^n>T\}} - I_{\{\tau_H>T\}}\big)\Big| = \big|E(I_{\{S^*\ge H\}}) - E(I_{\{S'_n\ge H\}})\big| = \\ \big|E^{\mu}\big(\mathcal{E}^{-1}(\mu)I_{\{(W(\mu))^*\ge L\}}\big) - E^{\mu^n}\big(\mathcal{E}^{-1}(\mu^n)I_{\{(W(\mu^n))_n^*\ge L\}}\big)\big| = \\ \big|E\big(\mathcal{E}(-\mu)I_{\{W^*\ge L\}}\big) - E\big(\mathcal{E}(-\mu^n)I_{\{W_n^*\ge L\}}\big)\big| \le \\ E\big(\left|\mathcal{E}(-\mu) - \mathcal{E}(-\mu^n)\right| I_{\{W^*\ge L\}}\big) + E\big(\mathcal{E}(-\mu^n)\left|I_{\{W^*\ge L\}} - I_{\{W_n^*\ge L\}}\right|\big) = \epsilon_1^n + \epsilon_2^n.\end{gathered}$$

We estimate the first item. Let us mention that $\mathcal{E}(-\mu)$ is the value of the solution of stochastic differential equation $dX_t = \frac{\mu}{\sigma}dW_t$, and $\mathcal{E}(-\mu^n)$ is the vlaue of Euler approximation for it, so by using the standard estimate mentioned above we obtain that

$$\epsilon_1^n \le E\big(\left|\mathcal{E}(-\mu) - \mathcal{E}(-\mu^n)\right|\big) \le C\sqrt{\Delta}.$$

To estimate ϵ_2^n at once, we note that $I_{\{W^*\ge L\}} - I_{\{W_n^*\ge L\}} = I_{\{W^*\ge L, W_n^*<L\}}$, i.e. for estimating mathematical expectation we can suppose that W_{t_i} is bounded from above by some constant L. Adding up the terms for $\mathcal{E}(-\mu^n)$, we obtain

$$\begin{gathered}\int_0^T \frac{\mu_t^n}{\sigma}dW_t = \frac{1}{\sigma}\sum_{i=0}^{n-1}\mu_{t_i}(W_{t_{i+1}} - W_{t_i}) = \\ \frac{1}{\sigma}\Big(\mu_{t_{n-1}}W_{t_n} - \sum_{i=1}^{n-1} W_{t_i}(\mu_{t_i} - \mu_{t_{i-1}})\Big),\end{gathered}$$

whence, using Lipschitz condition for μ, we can write

$$\epsilon_2^n \le CE\big(\exp\{C\sup_{1\le i\le n}(W_{t_i})_-\}I_{\{W^*\ge L, W_n^*<L\}}\big),$$

here $x_- = -\min\{x,0\}$. Denote $Z = \sup_{1\le i\le n}(W_{t_i})_-$. Random variables $A_n = \{W^* \ge L, W_n^* < L\}$ and $B_x = \big\{\sup_{1\le i\le n}(W_{t_i})_- > x\big\}$ are negatively correlated. We can be convinced in it directly, writing simultaneous densities of the distribution W_{t_i}, $i = 1,\dots,n$ and W^*. But intuitively it is obvious, if we look at conditional probabilities of this values given $W_n^* = y$, $y \in (x, L)$ than for every y their negative correlation is obvious.

Writing now

$$E\big(e^{CZ}I_A\big) = E\bigg(\Big(1+\int_0^\infty Ce^{Cx}I_{\{Z>x\}}dx\Big)I_{A_n}\bigg)$$
$$= P(A_n) + C\int_0^\infty e^{Cx}P(\{Z>x\}\cap A_n)dx \leq$$
$$P(A_n) + C\int_0^\infty e^{Cx}P(B_x)P(A_n)dx = P(A_n)E(e^{CZ}) \leq CP(A_n).$$

Finally we note that one of the result of the paper [2] contains the fact that $P(A_n) \sim C\sqrt{\Delta}$ when $n\to\infty$, hence $\epsilon_2^n \leq C\sqrt{\Delta}$.

So, we have

$$\Big|V_n^d(H) - V(H)\Big| \leq C\Delta = O\Big(\frac{1}{\sqrt{n}}\Big).$$

$\square$

We can prove theorem 2.1 using the result of theorem 2.2 .

Proof. Taking into account the result of theorem 2.2, it is enough to prove that $V_n^d(H) - V_n^b(H) = O(\ln n/\sqrt{n})$, $n\to\infty$.

It is clear that

$$|V_n^d(H) - V_n^b(H)| \leq C\left|E\big((S_T^d-K)^+I_{\{M_T^d<H\}}\big) - E\big((S_T^b-K)^+I_{\{M_T^b<H\}}\big)\right|. \tag{2.2}$$

Now we apply the result of Komlós, Major and Tusnády [5]. It says that for any given $\lambda>0$ it is possible to construct independent random variables $\eta_i \overset{d}{=} \xi_i$ and independent standard random variables ζ_i, $0\leq i\leq n-1$, such that for some positive constants K

$$P\Big(\max_{0\leq i\leq n-1}|S_i - T_i| > K\ln n + x\Big) \leq Ke^{-\lambda x}, \tag{2.3}$$

where

$$S_i = \sum_{j=0}^{i}\eta_j, \quad T_i = \sum_{j=0}^{i}\zeta_j.$$

Note that (2.3) implies $E(\max_{0\leq i\leq n-1}|S_i - T_i|^2) \leq C\ln^2 n$. Indeed, denoting $R = \max_{0\leq i\leq n-1}|S_i - T_i|$, we have

$$E(R^2) \leq (2K+2)^2\ln^2 n + E(R^2 I_{\{R>(2K+2)\ln n\}})$$
$$\leq C\ln^2 n + \int_0^\infty P(R^2 > (2K+2)^2\ln^2 n + x)dx$$
$$\leq C\ln^2 n + \int_0^\infty P(R > (K+1)\ln n + x/2)dx$$
$$\leq C\ln^2 n + Kn^{-\lambda}\int_0^\infty e^{-\lambda\sqrt{x/2}}dx \leq C\ln^2 n.$$

In the following we will assume without loss of generality $K\lambda > 1/2$.

As long as $\{\xi_i,\ i=0,\dots,n-1\} \overset{d}{=} \{\eta_i,\ i=0,\dots,n-1\}$ and $\{Z_i,\ i=0,\dots,n-1\} \overset{d}{=} \{\zeta_i,\ i=0,\dots,n-1\}$, in order to estimate the difference $|V_n^d(H)-V_n^b(H)|$ we can assume that $\xi_i=\eta_i$ and $Z_i=\zeta_i$, because this will not change the expectations in (2.2). Now write

$$|V_n^d(H)-V_n^b(H)| \le C(I_1+I_2),$$

where

$$\begin{aligned}
I_1 &= \big|E([(S_T^d-K)_+ - (S_T^b-K)_+]I_{\{M_T^b<H\}})\big| \\
&\le E(|(S_T^d-K)_+ - (S_T^b-K)_+|I_{\{M_T^b<H\}}) \le E(|S_T^d-S_T^b|I_{\{M_T^b<H\}}), \\
I_2 &= \big|E((S_T^d-K)_+[I_{\{M_T^d<H\}} - I_{\{M_T^b<H\}}])\big| \\
&\le CE\big(|I_{\{M_T^d<H\}} - I_{\{M_T^b<H\}}|\big) \\
&\le C\big(P(M_T^d<H, M_T^b\ge H) + P(M_T^d\ge H, M_T^b<H)\big).
\end{aligned}$$

Processes S^d and S^b are of the form S_0e^x, hence from inequality $|e^x-e^y| \le (e^x+e^y)|x-y|$ we obtain

$$I_1 \le CE\left(|S_T^b+S_T^d|\sigma\sqrt{\Delta}\left|\sum_{j=0}^{n-1}(Z_j-\xi_j)\right|I_{\{M_T^b<H\}}\right)$$

Using the Cauchy–Bunyakovsky inequality, we get:

$$I_1 \le C\sigma\sqrt{\Delta}\big(E(|S_T^d+S_T^b|^2 I_{\{M_T^b<H\}})\big)^{1/2} \times \left(E\Big[\sum_{j=0}^{n-1}\big(Z_j-\xi_{j+l}\big)\Big]^2\right)^{1/2}.$$

Now

$$\begin{aligned}
E(|S_T^d+S_T^b|^2 I_{\{M_T^b<H\}}) &\le 2E\big[\big((S_T^d)^2+(S_T^b)^2\big)I_{\{M_T^b<H\}}\big] \\
&\le C\big(E(S_0\exp\{2CT+2\sigma TW_T\}) + H^2\big) \le C,
\end{aligned}$$

as $\exp\{\sigma TW_T\}$ is integrable, and $|\mu_t|$ is bounded. On the other hand, as it was pointed above,

$$E\Big[\sum_{j=0}^{n-1}\big(Z_j-\xi_{j+l}\big)\Big]^2 \le C\ln^2 n,$$

thus we have

$$I_1 \le C\sqrt{\Delta}\ln n \le C\frac{\ln n}{\sqrt{n}}.$$

Now turn to I_2. Both probabilities are estimated in a similar manner, so we will estimate only the first one. Write

$$\begin{aligned}
&P(M_T^d<H, M_m^b\ge H) \\
&\le P(H-\delta\le M_T^d<H, M_T^b\ge H) + P(M_T^d<H-\delta, M_T^b\ge H) \\
&\le P(H-\delta\le M_T^d<H) + P(M_T^d<H-\delta, M_T^b\ge H) =: P_1+P_2.
\end{aligned}$$

It is easy to see that M_T^d has a bounded density, so $P_1 \leq C\delta$.

Now we observe that $P(M_T^d < H-\delta, M_T^b \geq H)$ implies that for some i $S_T^d < H-\delta < H \leq S_T^b$, so, by taking logarithms, we have

$$\sqrt{\Delta}\sum_{j=0}^{i-1}(\xi_j - Z_j) > C\delta,$$

which implies

$$\sum_{j=0}^{i-1}(\xi_j - Z_j) > C\delta\sqrt{n}.$$

Now take $\delta = 2K\ln n/\sqrt{n}$. With this choice we have from (2.3) $P_2 \leq Cn^{-\lambda K} \leq C\ln n/\sqrt{n}$. Summing up, we have $I_2 \leq C\ln n/\sqrt{n}$, and the assertion of the theorem follows. □

3 Modelling

As in [8], we give an example showing how fast the price in discrete binomial model converges to correspondent price in continuous model.

Consider the drift function of the form:

$$\mu_t = \begin{cases}\mu_1, & 0 \leq t < T/2,\\ \mu_2, & T/2 \leq t \leq T.\end{cases}$$

This function (and corresponding interest rate r_t) does not satisfy the condition of continuity (2.1), which we have impose on it. But, if we track the proof process of the theorem 2.1, it is not difficult to see, that it is enough to have the condition (2.1) fulfilling for $t = t_i, s \in [t_i, t_{i+1})$, that is true for such a function.

According to [3] we have that for Brownian motion X_t with initial value x and constant drift coefficient μ simultaneous density of the distribution of maximum M_t on interval $[0,t]$, of the points T_t of the maximum and of the values X_t is given by

$$P(X_t \in dz, M_t \in dy, T_t \in ds) = \frac{(y-x)(y-z)}{\pi\sqrt{s^3(t-s)^3}}\times$$
$$\exp\Big(-\frac{(y-x)^2}{2s} - \frac{(y-z)^2}{2(t-s)} - \mu(x-z) - \frac{\mu^2 t}{2}\Big)dz\,dy\,ds =: f_{t,x,\mu}(z,y,s)dz\,dy\,ds$$

when $x \leq y$ and $z \leq y$; when $x > y$ or $z > y$ it equals to zero. Noting $\nu(T) = \exp\left\{-\int_0^T r_t dt\right\} = \exp\Big\{-\frac{T}{2}(\mu_1+\mu_2+\sigma^2)\Big\}$ and using the fact, that $Z_t = \frac{1}{\sigma}\ln S_t$ is a Brownian motion with the drift $\nu_1 = \frac{\mu_1}{\sigma}$ on $[0,\frac{T}{2}]$ and $\nu_2 = \frac{\mu_2}{\sigma}$ on $[\frac{T}{2},T]$, we can get the European up-and-out call option fair price as

$$V(H) = E\Big(\nu(T)(S_T - K)^+ I_{\{\tau(H,S)>T\}}\Big) = \nu(T)E\Big((S_T - K)^+ I_{\{\sup_{[0,T]} S_t<H\}}\Big) =$$

$$\nu(T)E\Big(E\big((S_T - K)_+ I_{\left\{\sup_{[\frac{T}{2},T]} S_t<H\right\}} \,\big|\, \mathcal{F}_{\frac{T}{2}}\big) I_{\left\{\sup_{[0,\frac{T}{2}]} S_t<H\right\}}\Big) =$$

$$\nu(T)E\Big(\int_0^{\frac{T}{2}} \int_{Z_{\frac{T}{2}}}^{\frac{1}{\sigma}\ln H} \int_{-\infty}^{y} (e^{\sigma z} - K)_+ f_{\frac{T}{2},Z_{\frac{T}{2}},\nu_2}(z,y,s)dz\,dy\,ds\; I_{\left\{\sup_{[0,\frac{T}{2}]} S_t<H\right\}}\Big) =$$

$$\nu(T)\int_0^{\frac{T}{2}} \int_{Z_0}^{\frac{1}{\sigma}\ln H} \int_{-\infty}^{v} f_{\frac{T}{2},Z_0,\nu_1}(x,v,u)\times$$

$$\int_0^{\frac{T}{2}} \int_{x}^{\frac{1}{\sigma}\ln H} \int_{-\infty}^{y} (e^{\sigma z} - K)_+ f_{\frac{T}{2},x,\nu_2}(z,y,s)dz\,dy\,ds\,dx\,dv\,du.$$

The last integral is rather difficult to calculate because of its high dimension. Nevertheless, integrals in y and v can be evaluated in closed form, with the use of the standard normal distribution function; we do not give the result of this integrating — formulae are very intricate — and give only the final estimation for the integral.

Let take the following meanings of parameters: $S_0 = 100$, $\sigma = 0.1$, $K = 100$, $H = 105$, $T = 0{,}2$, $\mu_1 = 0{,}1$, $\mu_2 = 0{,}2$. Then with accuracy 10^{-4}

$$V(H) = 0{,}4744.$$

To estimate the order of the rate of convergence for the option prices with discrete time, using Monte Carlo simulations for the estimation of mathematical expectation, we will generate 100000 trajectories of asset price (50000 trajectories for $m = 1000, 2000$). The results we have got are noted in the table 1. We should note that the option prices with discrete time are bigger and decreasing when size of partition increasing. This property is natural, because in the case when the quantity of the points in our division increases, the moment set in which we examine does asset price cross given level or not also increases. There is no clear evidence however from this data whether our estimate for the rate of convergence is sharp. Nevertheless, we believe it is sharp, because this rate is the best rate of *strong* approximation of Wiener process by a binomial random walk; but we cannot claim the sharpness, because our result concerns weak convergence.

4 Conclusions

We have proved that barrier option fair prices in discrete binomial Black–Scholes model with non-constant drift coefficient converges to corresponding price in continuous model, and the rate of convergence could be estimated as $O(\frac{\ln n}{\sqrt{n}})$, where n is the number of operational moments in the discrete binomial model.

References

[1] Broadie, M., Glasserman, P., Kou, S.G. *A continuity correction for discrete barrier options*, Math. Finance, 1997, v:7, pp. 325-349.

n	$V_n^b(H)$	$V_n^b(H) - V(H)$	$(V_n^b(H) - V(H))n^{1/2}/\ln n$
10	0.5675	0.0932	0.1279
20	0.5001	0.0257	0.0883 (0.0383)
50	0.5772	0.1028	0.1858
100	0.5438	0.0694	0.1507
200	0.4991	0.0237	0.0658
500	0.4921	0.0177	0.0638
1000	0.4906	0.0162	0.0742
2000	0.4785	0.0041	0.0241

Table 1: Price $V(H)$ of the European up-and-out call option in continuous model and the price $V_n^b(H)$ of the same option in discrete binomial model.

[2] Broadie, M., Glasserman, P., Kou, S.G. *Connecting discrete and continuous path-dependent options*, Finance and Stochastics, 1999, v:3, pp. 55-82.

[3] Csáki, E., Földes A., Salminen P., *On the joint distribution of the maximum and its location for a linear diffusion*, Ann. Inst. Henri Poincaré, Probab. Stat. 1987, v:23, pp. 179–194.

[4] Hŏrfelt, P. *Extension of the corrected barrier approximation by Broadie, Glasserman, and Kou*, Finance and Stochastics, 2003, v:7, pp. 231-243.

[5] Komlós, J., Major, P., Tusnády, G. *An approximation of partial sums of independent RV's and the sample DF. II*, Z. Wahrscheinlichkeitstheor. Verw. Geb. 34, 33-58 (1976).

[6] Kou, S.G. *On pricing of discrete barrier options*, Statistika Sinica, 2003, v:13, pp. 955-964.

[7] Merton, R.C. *Theory of rational option pricing*, Bell J. Econom. Manage. Sci., 1973, v:4, pp. 141-183.

[8] Soloveiko, O., Shevchenko, G., *The rate of convergence of barrier option price with non-constant drift in discrete time to ones in continuous time*, Accepted to Probability Theory and Mathematical Statistics, 2008.

[9] Kloeden, P.E., Platen, E. Numerical solution of stochastic differential equations, Springer, Berlin, 1992.

Complex risk measures in portfolio optimization

Efim Bronshtein and Julia Kurelenkova
Ufa State Aviation technical University, Russia
e-mail: bro-efim@yandex.ru

Abstract

We investigate the portfolio optimization problem using complex risk measures. In particular, we propose risk measures, composed of Value-at-Risk, Conditional Value-at-Risk, average, standard deviation, average of absolute deviation and their modifications. We analyze the received portfolios, constructed from Russian and World stocks on the different time horizons. The practical recommendations for described approaches are given. The results can be compared with Basel-2.

Keywords.: optimal stocks portfolios; risk measures.

Portfolio optimization and risk measures

One of the main problems of financial analysis is the market risk measuring. We consider this problem in the context of stocks portfolio optimization.

The foundations of modern investment theory were layd in 1952 by H. Markovitz. Risk obtained mathematical definition in the investment theory, what made it possible to construct mathematical models of portfolio optimization problem. Then the risk of the portfolio was determined as standard deviation. Since then appeared many approaches for risk measurement.

We define the daily (week, month etc.) portfolio profit as follows:

$$r_t = \frac{P_{t+1} - P_t}{P_t} \tag{1}$$

where r_t - portfolio profit, P_t - portfolio cost in the time moment t.

In the financial world nowadays Value-at-Risk (VaR) has become one of the most used and important measures of risk. The VaR of the portfolio at the confidence level $\alpha \in (0;1)$ is equal to such value that the probability of the event "profit X will not be less than VaR" is more than α:

$$VaR_\alpha(X) = \max\{\xi | p(X \geq \xi) \geq \alpha\} \tag{2}$$

The VaR has some undesirable features. The main is that VaR ignores the distribution of portfolio returns beyond its VaR. That is, VaR ignores the magnitude of the worst cases.

In [4] a set of axioms was proposed as the key properties to be satisfied by "coherent measure of risk":

$$1. \quad \text{Monotonicity:} \quad \text{if} \quad X \geq Y \quad \text{then} \quad \rho(X) \leq \rho(Y); \tag{3}$$

$$2. \quad \text{Positive homogeneity:} \quad \text{if} \quad \lambda \geq 0 \quad \text{then} \quad \rho(\lambda X) = \lambda\rho(X); \tag{4}$$

$$3.\quad \text{Translational invariance :} \quad \rho(X + a) = \rho(X) - a; \tag{5}$$

$$4.\quad \text{Subadditivity:} \quad \rho(X + Y) \leq \rho(X) + \rho(Y) \tag{6}$$

The VaR is not a coherent measure of risk since it is not subadditive in general case. The VaR of a portfolio with two stocks may be larger than the sum of the VaRs of the stocks in the portfolio.

The alternative risk measure that overcomes the described features is the Conditional Value-at-Risk ($CVaR$).

The $CVaR$ is an expected value of portfolio profit X under the condition $X \leq VaR_\alpha(X)$.

$$CVaR_\alpha(X) = E(X|X \leq VaR_\alpha(X)). \tag{7}$$

In other words, $CVaR_\alpha(X)$ is an average value of $(1 - \alpha) * 100\%$ of the smallest profits.

In [4] were proposed risk measures

$$VaR_{E,\alpha} = VaR_\alpha(X - E(X)), \tag{8}$$

$$CVaR_{E,\alpha} = CVaR_\alpha(X - E(X)), \tag{9}$$

where $E(X)$ is an average value.

We also analyse the following modifications of VaR and $CVaR$:

$$VaR_{Me,\alpha} = VaR_\alpha(X - Me(X))and \tag{10}$$

$$CVaR_{Me\alpha} = CVaR_\alpha(X - Me(X)), \tag{11}$$

where $Me(X)$ is the median.

We propose new complex risk measures (CRM), integrating VaR, $CVaR$, standard deviation and average of absolute deviation:

$$CRM_1 = CVaR_\alpha(X) - \beta E|X - E(X)|, \tag{12}$$

$$CRM_2 = VaR_\alpha(X) + CVaR_\alpha(X) - \beta E|X - E(X)|. \tag{13}$$

where $\beta \geq 0$.

Method of research

The described risk measures were used for the stocks portfolio optimization. Russian, World and mixed stocks portfolios were compared by the profitability for the following period. The historical data includes up to 5 years of stocks costs On the Russian stock market we analyzed stocks of the following companies: Sberbank, Gazprom, Norilskii Nikel, Aeroflot, Mosenergo, MTS, Baltika, Salavatnefteorgsintez, GAZ. On the World stock market we analyzed the stocks of companies General Electric, Microsoft, IBM, Citigroup, Pfizer, American International Group, Bank of America, Johnson&Johnson, Wal-Mart Stores, Exxon Mobil.

Each of the risk measures was analyzed for the following parameters:

- The dependence of optimal portfolio profit on parameter β for some measures;
- Analysis of diversification measures for optimal portfolios;
- The optimal portfolio structure and profit for different historical periods (from 1 month to 5 years).

We used the following method for portfolio optimization. On the first step we generate the random portfolio i.e. the vector $(a_1, a_2, \ldots, a_n)$, $a_i \geq 0, \sum_{a_i}^{n} a_i = 1$ where a_i is the parth of i-th company stocks in portfolio. We calculate the basic value of optimization criteria (one of the described above risk measures). Then we increase by turn the part of each stock with $a_i < 1$ in a portfolio (accordingly we reduce parts $a_j > 0, j \neq i$) and define the best direction for continuation of our calculations. The process continues while the improvement becomes less than a defined level.

Then we repeat the generating of random portfolios (first step) until the necessary accuracy level will be reached.

The efficient usage of the proposed risk measures requires the correct choice of parameter α. We analyzed optimal portfolios for $\alpha = 0,95$.

Some results of research

We investigated the dependence on coefficient β of optimal portfolio profit for measures CRM_1 and CRM_2. It was determined that for the high values of β (approximatly $\beta = 50$) the profit of optimal portfolios becomes stable. That is, the further increasing of β will not influence on the result of optimization. That is why we use $\beta = 50$.

We define a diversification measure of portfolio as the maximal stock parts in optimal portfolio. The lower is diversification measure, the more diversified is portfolio.

The most diversified optimal portfolios were obtained for $CVaR_{Me}$ risk measure.

The profit of optimal portfolios depends on the stock market situation. We observed the cases when all the risk measures have got the positive profit and vice versa.

Let us analyse some examples.

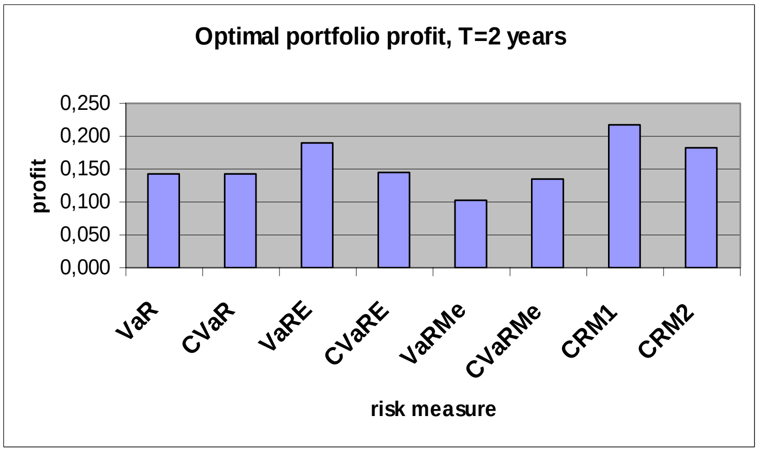
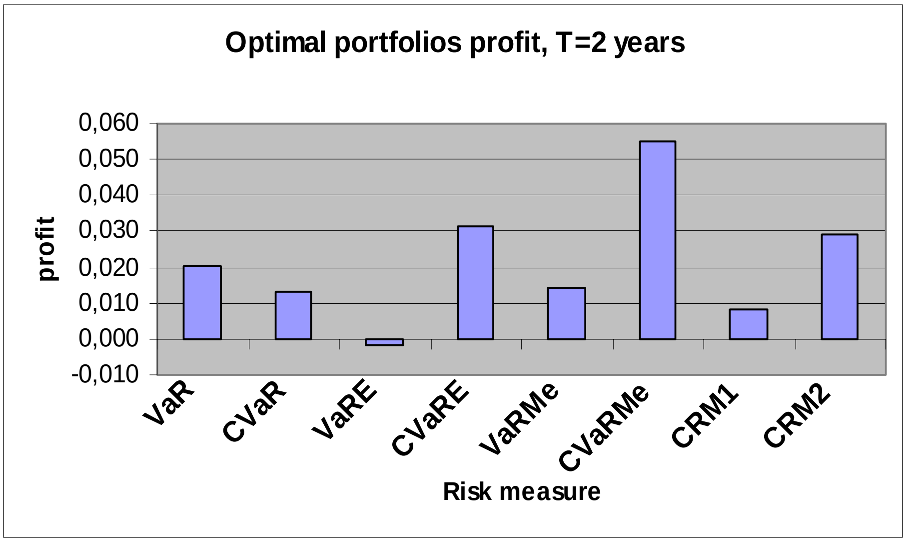


Figure 1. Optimal portfolio profit, Russian stock market, hystorical period 2 years.

We see that the most profitable optimal portfolios were obtained with using VaR_E, CRM_1, CRM_2.

Figure 2. Optimal portfolio profit, World stock market, hystorical period 2 years.



For World stock market the best results gave measures $CVaR_{Me}, CVaR_E, CRM_2$.
We also investigated mixed stock portfolios, consisted of Russian and World stocks

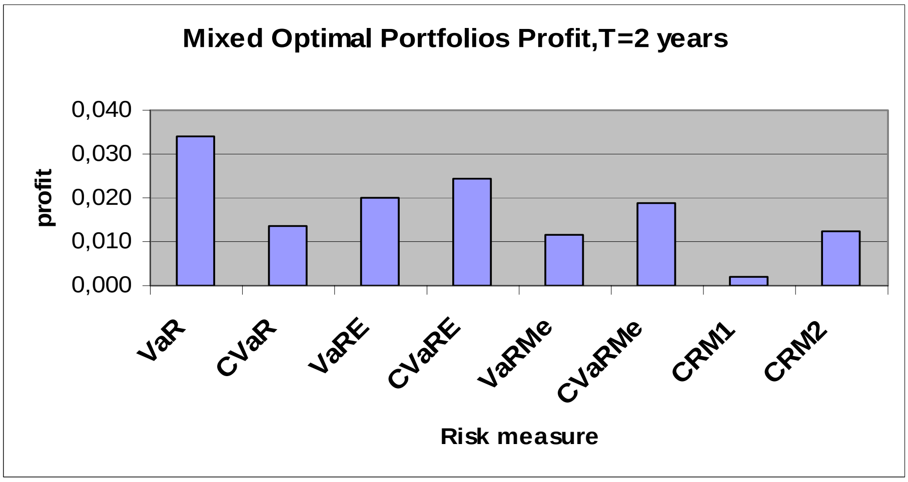


Figure 3. Mixed optimal portfolio profit, hystorical period 2 years

Here the best result show $VaR, CVaR_E, VaR_E, CVaR_{Me}$.
The structure of mixed optimal portfolio for $CVaR_{Me}$ has the following structure:

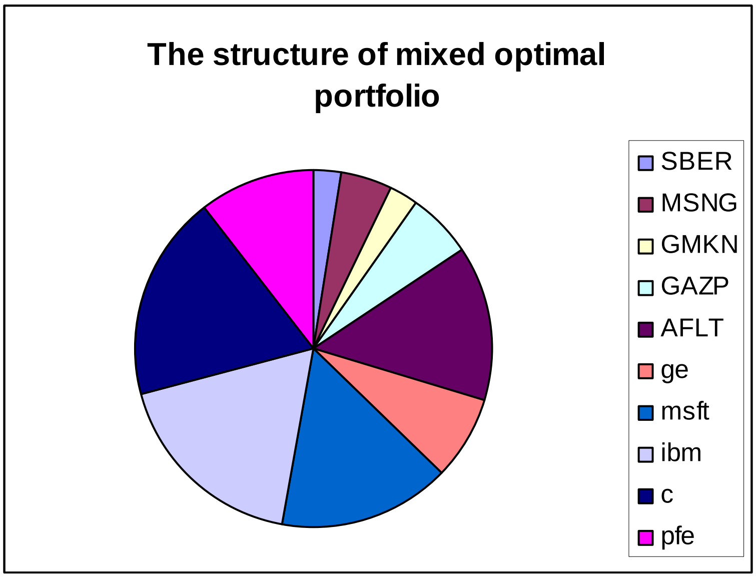


Figure 4. The structure of mixed optimal portfolio.

Conclusions

- Introdused risk measures can be applied equally with traditional measures. But the most profitable mixed optimal portfolios were obtained with VaR.
- The main part in the structure of mixed optimal portfolio take the World stocks. Aeroflot is the leader among Russian stocks.
- The most profitable optimal portfolios for the Russian stocks were obtained with VaR_E, CRM_1, CRM_1 risk measures.
- For the World stocks the most profitable optimal portfolios were obtained with $CVaR$-based risk measures and CRM_2.
- The World optimal portfolios were more diversificated in the short time periods.
- The most preferable historical period varies for different risk measures. Generally it was equal to two years.
- On the short time periods CRM_1 is more efficient than CRM_2.

Acknowledgement.
The research supported by Russian Foundation for Basic Researches (project 07-06-00021)

References

[1] *Bronshtein E.M., Kurelenkova Y.V.* Comparison of optimal investment portfolios, minimizing different risk measures // Economic Analysis: theory and practice. 2006; 3 (in Russian).

[2] *Bronshtein E.M., Kurelenkova Y.V.* How to measure the risk // Sequrity Market. 2006; 24 (in Russian).

[3] *Menshikov I.S., Shelagin D.A.* Market risks: models and methods // Computation Center RAS, 2000 (in Russian).

[4] *Artzner Ph., Delbaen F., Eber J.-M., Health D.* Thinking Coherently // Risk. 10, November, 1997.

[5] *Uryasev S.* Conditional Value-at-Risk: Optimization Algorithms and Applications // Financial Eng. News. 14, February, 2000.

[6] *Rockafellar R.T, Uryasev S., Zabarankin M.* Deviation Measures in Risk Analysis and Optimization // Research Report 2002-7, ISE Dept. Univ. Florida.

Stochastic vortices in periodically reclassified populations - An application to pension funds

Gracinda Rita Diogo Guerreiro
Dpt. of Mathematics
UNL - FCT
grg@fct.unl.pt

João Tiago P. Nunes Mexia
Dpt. of Mathematics
UNL - FCT
jtm@fct.unl.pt

Abstract

In our model we consider open populations divided into a finite number of sub-populations. These populations play a relevant part in many problems. For instance, we may consider the drivers which are clients of an insurance company. According to their records they are placed in one of the *Bonus-Malus* classes. A similar example is given by the clients of a bank each of which is placed in a *Credit-Rating* level. Another example can be given by the population within a Pension Plan. The beneficiaries of the fund can be grouped into subgroups, namely, "Active", "Retired", "Invalid" and even more.

It is easily seen that to manage these populations it is very important to have information about the relative sizes of the sub-populations.

With our model, we obtain limit results for these relative sizes considering the possibilities of entrances, reclassifications and departures of population elements.

We also consider that new elements entering the population are subject to an initial classification and so they can be initially placed into any of the sub-populations.

Our treatment will be based on finite, discrete parameter, homogeneous Markov chains. We consider the possibility of more than one transient class as well as more than one recurrent class.

The stability of relative sizes of sub-populations, despite entrances, departures and reallocations, shows the existence of a structure. We call these structures stochastic vortices.

We will not develop this aspect in here, but during our study we noted that an interesting problem occurs when, for the one step transition matrix of a recurrent class, we have more than one module 1 eigenvalues. Then, from the Frobenius theorem, it may be shown that there is a limit cycle for the transition probabilities between states in that class. Nevertheless, under general conditions the relative sizes of the corresponding sub-populations will be stable.

Keywords: Markov Chains; Stochastic Vortices.

1 Introduction

Let us consider an open population divided into sub-populations, whose elements are subject to periodic reclassifications. In each reclassification, the population elements can be placed into any of the sub-populations. Besides this, we assume that new elements entering the population are subject to an initial classification and likewise, they can be initially placed into any of the sub-populations. These populations play a relevant part in many problems. For instance, we may consider the drivers which are clients of an Insurance Company. In each annuity, according to their records they are reclassified and placed in one of the Bonus-Malus classes. A similar example is given by the clients of a Bank Institution, each of which is placed in a Credit-Rating level. The list of examples can, very easily, be extended. At the end of the presentation we will see an application of the model to the population of Pension Fund Beneficiaries. It is easily seen that to manage these populations it is very important to have information about the relative sizes of the sub-populations. Using the stochastic vortices model, we obtain limit results for these relative sizes assuming that: a) entries, reallocations and departures occur at equally spaced times; b) probabilities of reallocation of the population elements are stable; c) entries are given by Poisson distributed random variables; d) transition probabilities between sub-populations are stable. Our treatment will be based on finite, discrete parameter, homogeneous Markov chains, considering the possibility of more than one transient class and more than one recurrent class. However, in this presentation, we will foccus our attention in the transient classes. The stability of relative sizes of sub-populations, despite entrances, departures and reallocations, shows the existence of a structure. We call these structures stochastic vortices. An interesting problem occurs when, for the one step transition matrix of a recurrent class, we have more than one module 1 eigenvalues. Then, from the Frobenius theorem, it may be shown that there is a limit cycle for the transition probabilities between states in that class. Nevertheless, under general conditions the relative sizes of the corresponding sub-populations will still be stable.

2 Population Structure

In our study will consider:

- populations divided into k sub-populations, corresponding to k Markov chain states, grouped into w communication classes;
- k_d^+ transient states grouped into d transient classes, with k_j , $j = 1, \dots, d$ states;
- $k - k_d^+$ recurrent states grouped into r recurrent classes, with k_{d+j} , $j = 1, \dots, r$ states.

The one step transition matrix of the Markov between sub-populations will be

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_{1,1} & \dots & \boldsymbol{P}_{1,d} & \boldsymbol{P}_{1,d+1} & \dots & \boldsymbol{P}_{1,w} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P}_{d,1} & \dots & \boldsymbol{P}_{d,d} & \boldsymbol{P}_{d,d+1} & \dots & \boldsymbol{P}_{d,w} \\ \boldsymbol{P}_{d+1,1} & \dots & \boldsymbol{P}_{d+1,d} & \boldsymbol{P}_{d+1,d+1} & \dots & \boldsymbol{P}_{d+1,w} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P}_{w,1} & \dots & \boldsymbol{P}_{w,d} & \boldsymbol{P}_{w,d+1} & \dots & \boldsymbol{P}_{w,w} \end{bmatrix} \tag{1}$$

with $\boldsymbol{P}_{i,j}$ the $k_i \times k_j$ sub-matrix of the transition probabilities between the states of the transient classes with index i, $i = 1, \dots, d$ [recurrent class with index $i-d$, $i = d+1, \dots, w$] and the states of the transient classes with index j , $j = 1, \dots, d$ [recurrent class with index $j - d$, $j = d + 1, \dots, w$]

We order the classes in such a way that

$$\begin{cases} \boldsymbol{P}_{l,h} = 0 & , \ l > h \\ \boldsymbol{P}_{l,h} = 0 & , \ h \neq l \ , \ l = d + 1, \dots w \end{cases} \tag{2}$$

The one step transition matrix can be then written as

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_{1,1} & \dots & \boldsymbol{P}_{1,d} & \boldsymbol{P}_{1,d+1} & \dots & \boldsymbol{P}_{1,w} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \dots & \boldsymbol{P}_{d,d} & \boldsymbol{P}_{d,d+1} & \dots & \boldsymbol{P}_{d,w} \\ \boldsymbol{O} & \dots & \boldsymbol{O} & \boldsymbol{P}_{d+1,d+1} & \dots & \boldsymbol{O} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \dots & \boldsymbol{O} & \boldsymbol{O} & \dots & \boldsymbol{P}_{w,w} \end{bmatrix} \tag{3}$$

To lighten the writing, from now on we will put

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{K} & \boldsymbol{U} \\ \boldsymbol{O} & \boldsymbol{V} \end{bmatrix} \tag{4}$$

with

- $\boldsymbol{K}$ - the $k_d^+ \times k_d^+$ transition matrix between transient states;
- $\boldsymbol{U}$ - the $k_d^+ \times (k - k_d^+)$ transition matrix between the transient and the recurrent states;
- $\boldsymbol{O}$ - the $(k - k_d^+) \times k_d^+$ null matrix;
- $\boldsymbol{V}$ - the $(k - k_d^+) \times (k - k_d^+)$ transition matrix between the recurrent states

Lemma 1 *The n-th step transition matrix will be*

$$\boldsymbol{P}^n = \begin{bmatrix} \boldsymbol{K}^n & \boldsymbol{U}_n \\ \boldsymbol{O} & \boldsymbol{V}^n \end{bmatrix} \tag{5}$$

with

$$\boldsymbol{U}_n = \boldsymbol{U}_{n-1} \cdot \boldsymbol{V} + \boldsymbol{U} \cdot \boldsymbol{K}^{n-1} \quad , n \in \mathbb{N} \setminus \{1\} \tag{6}$$

Proof 1 *Since the Markov chain is homogeneous we will have $P(n) = P^n$ and the thesis is easily established through mathematical induction.* □

3 Stochastic Vortices Model

3.1 Entrances in the Population

In this section we will make several assumptions:

- The number of elements entering the population in each time period, E_i , $i \in \mathbb{N}_0$, will be independent and Poisson distributed with mean λ_i , $i \in \mathbb{N}_0$,

$$E_i \sim P(\lambda_i) \, , \, i \in \mathbb{N}_0 \tag{7}$$

- The mean values λ_i , $i \in \mathbb{N}_0$, will be given by

$$\lambda_i = a + b\,\theta^i \ , \ \theta > 0 \ , \ i \in \mathbb{N}_0 \tag{8}$$

  Note that this is quite a general assumption. For instance, for $a = 0$ we can obtain $\lambda_i = b\theta^i$, which represents a population with a geometric growth. We can also obtain $\lambda_i = a(1 - e^{-\delta\, i})$, if $b = -a$ and $\theta = e^{-\delta}$, which represents a population with an asymptotic growth;

- New elements entering the population are subject to an initial classification. Elements entered in the i-th time period will be allocated to the different sub-populations according to the components of $\boldsymbol{c}_i$, $i \in \mathbb{N}_0$. We will also consider the sub-vector $\boldsymbol{t}_i$ $[\boldsymbol{r}_i]$ of $\boldsymbol{c}_i$, whose components are the probabilities for entering in transient [recurrent] states, thus $\boldsymbol{c}_i^T = \left[\begin{array}{c|c} \boldsymbol{t}_i & \boldsymbol{r}_i \end{array} \right]^T$. Each component of the probabilities of initial classification vector will be given by

$$c_{i,j} = c_{0,j} + e_j\,\gamma^i \ , \ i \in \mathbb{N}_0 \ , \ j = 1, \ldots, k.$$

  with

  - $\boldsymbol{c}_0 \rightarrow$ stable probability vector;
  - $\boldsymbol{e} \rightarrow \sum_{j=1}^{k} e_j = 0$, $i \in \mathbb{N}_0$ and $|e_j| < |c_{0,j}|$, $j = 1, \ldots, k$;
  - $0 < \gamma < 1$.

- The one-step transition matrix is given by (4) so the n-th step transition matrix is given by (5).

3.2 Expected Sub-Populations Dimension

Let $\boldsymbol{N}_i$ represent the number of elements arriving at each sub-population in time period i , $i \in \mathbb{N}_0$. Recalling that E_i, number of elements arriving at the population in time period i, is Poisson distributed with mean λ_i, it is easy to prove that

$$\boldsymbol{N}_i \sim \mathcal{P}(\lambda_i\, \boldsymbol{c}_i) \quad , \quad i \in \mathbb{N}_0.$$

After n time periods, the elements entered in the i-th time period have been subject to $n-i$ reclassifications. In this way, for the number of elements in each population in period n, that arrived to population in time period i, $N_{i,n}$, we can prove that

$$\boldsymbol{N}_{i,n} \sim \mathcal{P}(\lambda_i\, \boldsymbol{c}_i\, \boldsymbol{P}^{n-i}) \ , \ i \in \mathbb{N}_0.$$

Finally, for the total number of elements in each sub-population in time period n, we will have

$$\boldsymbol{N}_n^{++} = \sum_{i=0}^{n} N_{n,i} \sim \mathcal{P}(\boldsymbol{\lambda}_n^{++})$$

with

$$\boldsymbol{\lambda}_n^{++\,T} = \left[\underbrace{\sum_{i=0}^{n} \lambda_i\, \boldsymbol{t}_i^T\, \boldsymbol{K}^{n-i}}_{\mathbf{1}} \;\middle|\; \underbrace{\sum_{i=0}^{n} \lambda_i\, \boldsymbol{t}_i^T\, \boldsymbol{U}_{n-i} + \sum_{i=0}^{n} \lambda_i\, \boldsymbol{r}_i^T\, \boldsymbol{V}^{n-i}}_{\mathbf{2}} \right]$$

where:

1 - Expected Dimension of Sub-Populations corresponding to transient states;

2 - Expected Dimension of Sub-Populations corresponding to recurrent states.

3.3 Asymptotic Results for Transient States

For the purpose of this presentation it is sufficient to foccus our attention in the transient states. In this sub-section we will analyse the existence of stochastic vortices in the transient states, which will imply a stable limit dimension of those sub-populations.

Let us assume that $\boldsymbol{P}$ is a $k \times k$ diagonilizable matrix.
Under very general conditions, see Schott [6], we will have:

$$\boldsymbol{P} = \sum_{j=1}^{k} \eta_j\, \boldsymbol{\alpha}_j\, \boldsymbol{\beta}_j^T \tag{9}$$

where η_j $\left[\boldsymbol{\alpha}_j\ ,\ \boldsymbol{\beta}_j^T\right]$; $j = 1, \ldots, k$ are the eigenvalues [left and right eigenvectors] of $\boldsymbol{P}$.

Now, the n-th step transition matrix, $\boldsymbol{P}(n)$ will be, see also Schott [6], the *n-th* power of $\boldsymbol{P}$ and so,

$$\boldsymbol{P}(n) = \boldsymbol{P}^n = \sum_{j=1}^{k} \eta_j^n \, \boldsymbol{\alpha}_j \, \boldsymbol{\beta}_j^T \tag{10}$$

For the sub-matrices $\boldsymbol{P}_{l,h}, l = 1, \ldots, w, h = 1, \ldots, w$ of the transition matrix $\boldsymbol{P}$, considering $k_i^+ = \sum_{j=1}^{i} k_j$, we will have

$$\boldsymbol{P}_{l,h} = \sum_{j=k_{l-1}^+ + 1}^{k_h^+} \eta_j \, \boldsymbol{\alpha}_{j,l} \, \boldsymbol{\beta}_{j,h}^T \quad , \; l = 1, \ldots, w \; , \; h = 1, \ldots, w \tag{11}$$

as well as

$$\boldsymbol{P}_{l,h}(n) = \sum_{j=k_{l-1}^+ + 1}^{k_h^+} \eta_j^n \, \boldsymbol{\alpha}_{j,l} \, \boldsymbol{\beta}_{j,h}^T \quad , \; l \leq h \; , \; h = 1, \ldots, d \tag{12}$$

From Parzen [4], we know that the transition probabilities between the transient states tend to zero, as $n \to +\infty$, so

$$\lim_{n \to +\infty} p_{l,h}(n) = 0 \; , \; l = 1, \ldots, d \; , \; h = 1, \ldots, d \tag{13}$$

and so, from (12) and (13) we can conclude that for the transient states we have
$|\eta_j| < 1 \; , \; j = 1, \ldots, k_d^+$.

For vortices based on the transient states, we will only consider the $\boldsymbol{K}$ sub-matrix of (4) given by

$$\boldsymbol{K} = \sum_{j=1}^{k_d^+} \eta_j \, \boldsymbol{\alpha}_j \, \boldsymbol{\beta}_j^T \tag{14}$$

where the $\boldsymbol{\alpha}_j \; \left[\boldsymbol{\beta}_j^T\right] \; , \; j = 1, \ldots, k_d^+$ are the left [right] eigenvectors of $\boldsymbol{K}$.

Proposition 1 *With* $\lambda_i = a + b\,\theta^i \; , \; 0 < \theta < 1, a, b \in \mathbb{R}$ *and* $\boldsymbol{K}$ *a diagonizable matrix, the limit dimension for the sub-populations in the transient states is given by*

$$\begin{aligned} \boldsymbol{\lambda}_\infty^{+\,T} &= \lim_{n \to +\infty} \boldsymbol{\lambda}_n^{+\,T} = \lim_{n \to +\infty} \sum_{i=0}^{n} (a + b\,\theta^i)\, \boldsymbol{t}_i^T \, \boldsymbol{K}^{n-i} = \\ &= \begin{cases} a\, \boldsymbol{t}_0^T \, (\boldsymbol{I}_{k_d^+} - \boldsymbol{K})^{-1} & , \; 0 < \theta < 1 \\ \\ (a+b)\, \boldsymbol{t}_0^T \, (\boldsymbol{I}_{k_d^+} - \boldsymbol{K})^{-1} & , \; \theta = 1 \end{cases} \end{aligned}$$

Proof 2 *With* $\lambda_i = a + b\,\theta^i$, $i \in \mathbb{N}_0$, *we will have*

$$\begin{aligned}
\boldsymbol{\lambda}_n^{+T} &= \sum_{i=0}^{n} \lambda_i\, \boldsymbol{t}_i^T\, \boldsymbol{K}^{n-i} = \sum_{i=0}^{n} \lambda_{n-i}\, \boldsymbol{t}_{n-i}^T\, \boldsymbol{K}^i = \\
&= \sum_{i=0}^{n} (a + b\theta^{n-i})\, (\boldsymbol{t}_0^T + \boldsymbol{e}^T\, \gamma^{n-i})\, \boldsymbol{K}^i = \\
&= \sum_{i=0}^{n} \Big[a\, \boldsymbol{t}_0^T + a\, \boldsymbol{e}^T\, \gamma^{n-i} + b\, \boldsymbol{t}_0^T\, \theta^{n-i} + b\, \boldsymbol{e}^T\, (\theta\,\gamma)^{n-i} \Big] \boldsymbol{K}^i.
\end{aligned}$$

In order to simplify calculation, we will present it by parts:

1. $\displaystyle\lim_{n\to+\infty} a\, \boldsymbol{t}_0^T \sum_{i=0}^{n} \boldsymbol{K}^i = a\, \boldsymbol{t}_0^T \sum_{i=0}^{+\infty} \boldsymbol{K}^i = a\, \boldsymbol{t}_0^T (\boldsymbol{I} - \boldsymbol{K})^{-1}$,

since $\boldsymbol{K}$ *is the transient states matrix and likewise* $|\eta_j| < 1$, $j = 1, \ldots, k_d^+$.

2. $\displaystyle\lim_{n\to+\infty} \gamma^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\gamma}\right)^i = \lim_{n\to+\infty} \gamma^n \sum_{i=0}^{n} \sum_{j=1}^{k_d^+} \left(\frac{\eta_j}{\gamma}\right)^i \boldsymbol{\alpha}_j\, \boldsymbol{\beta}_j^T =$

$$= \sum_{j=1}^{k_d^+} \left[\lim_{n\to+\infty} \gamma^n \sum_{i=0}^{n} \left(\frac{\eta_j}{\gamma}\right)^i \right] \boldsymbol{\alpha}_j\, \boldsymbol{\beta}_j^T = 0,$$

since that

$$\lim_{n\to+\infty} \gamma^n \sum_{i=0}^{n} \left(\frac{\eta_j}{\gamma}\right)^i = \lim_{n\to+\infty} \frac{\gamma^{n+1} - \eta_j^{n+1}}{\gamma - \eta_j} = 0,$$

with $\gamma \neq \eta_j$, *once that* $0 < \gamma < 1$ *and* $|\eta_j| < 1$, $j = 1, \ldots, k_d^+$. *In this way,*

$$\lim_{n\to+\infty} a\, \boldsymbol{e}^T\, \gamma^{n-i} \sum_{i=0}^{n} \boldsymbol{K}^i = \lim_{n\to+\infty} a\, \boldsymbol{e}^T\, \gamma^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\gamma}\right)^i = 0$$

Note that, if $\gamma = \eta_j'$ *we still have*

$$\lim_{n\to+\infty} \gamma^n \sum_{i=0}^{n} \left(\frac{\eta_j}{\gamma}\right)^i = \lim_{n\to+\infty} \gamma^n (n+1) = 0, \text{ once that } 0 < \gamma < 1.$$

3. $\displaystyle\lim_{n\to+\infty} \theta^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\theta}\right)^i$

This case resumes to one of the previous according to $\theta = 1$ *or* $0 < \theta < 1$.

Then,

$$\lim_{n\to+\infty} b\, \boldsymbol{t}_0^T\, \theta^{n-i} \sum_{i=0}^{n} \boldsymbol{K}^i = \lim_{n\to+\infty} b\boldsymbol{t}_0^T\, \theta^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\theta}\right)^i = \begin{cases} b\, \boldsymbol{t}_0^T\, (\boldsymbol{I}-\boldsymbol{K})^{-1} & , \theta = 1 \\ 0 & , 0 < \theta < 1 \end{cases}$$

4. $\lim_{n\to+\infty} (\theta\,\gamma)^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\theta\,\gamma}\right)^i = 0$

Note that this case equals case 2 since that $\theta\,\gamma \neq 1$. In this way,

$$\lim_{n\to+\infty} b\, \boldsymbol{e}^T\, (\theta\,\gamma)^{n-i} \sum_{i=0}^{n} \boldsymbol{K}^i = \lim_{n\to+\infty} b\, \boldsymbol{e}^T\, (\theta\,\gamma)^n \sum_{i=0}^{n} \left(\frac{\boldsymbol{K}}{\theta\,\gamma}\right)^i = 0$$

The rest of the proof is easy to establish. □

The limit relative dimensions will then be stable, for $0 < \theta \leq 1$:

$$\pi_{\infty,j} = \frac{\lambda_{\infty,j}^{+}}{\sum_{j=1}^{k_d^+} \lambda_{\infty,j}^{+}} \quad , \quad j = 1, \ldots, k_d^+$$

Thus, a stochastic vortice is established in the transient states, for $0 < \theta \leq 1$.

If $\theta > 1$, the sub-populations dimension will grow to $+\infty$, proporcionally to θ^{n+1} and we may establish the next proposition.

Proposition 2 *With $\lambda_i = a + b\,\theta^i$, $\theta > 1, a, b \in \mathbb{R}$ and $\boldsymbol{K}$ a diagonizable matrix, the limit dimension for the sub-populations in the transient states is given by*

$$\boldsymbol{\lambda}_{\infty,\theta}^{+\,T} = \lim_{n\to+\infty} \frac{\boldsymbol{\lambda}_n^{+\,T}}{\theta^{n+1}} = \frac{1}{\theta}\, b\, \boldsymbol{t}_0^T \left(\boldsymbol{I}_{k_d^+} - \frac{\boldsymbol{K}}{\theta}\right)$$

The proof is similar to the previous one.

The limit relative dimensions will be stable, as $n \to +\infty$, and will be given by:

$$\pi_{\infty,\theta,j} = \frac{\lambda_{\infty,\theta,j}^{+}}{\sum_{j=1}^{k_d^+} \lambda_{\infty,\theta,j}^{+}} \quad , \quad j = 1, \ldots, k_d^+$$

so a stochastic vortice is established in the transient states, even in the cases where $\theta > 1$.

3.4 Confidence Intervals

For the total number of elements in each sub-population, we proved that

$$N_{n,j}^{+} \sim \mathcal{P}(\lambda_{n,j}^{+}).$$

For large populations, we can obtain the confidence interval for $\lambda_{n,j}^{+}$:

$$\left[N_{n,j}^{+} - z_{q/2}\sqrt{N_{n,j}^{+}}\,;\; N_{n,j}^{+} + z_{q/2}\sqrt{N_{n,j}^{+}}\right]\;,\; j = 1, \dots, k_d^{+}$$

with $z_{q/2} = \mathbb{P}\left[Z_j \leq \frac{z}{2}\right] = 1 - \frac{q}{2}$, where $Z_j \sim \mathcal{N}(0,1)$.

4 An Application to Pension Funds

4.1 Population

Consider the population of a Pension Fund Beneficiaries, for which we considered the following sub-populations:

- Active Employees [$\boldsymbol{a}$] [Ages 20 to 64]
- Retired [$\boldsymbol{r}$] [Ages 65 to 106]
- Disabled [$\boldsymbol{i}$] [Ages 21 to 106]
- Survival Beneficiaries
  - Sons [$\boldsymbol{f}$] [Ages 0 to 24]
  - Spouses [$\boldsymbol{c}$] [Ages 26 to 106]
  - Spouses with Sons [$\boldsymbol{c}^{+}$] [Ages 28 to 52]

4.2 One Step Transition Matrix

The one-step transition matrix was built using the tables:

- TV 73-78 [Mortality Table]
- EKV 80 [Disability Table]
- PCR Turnover [Turnover Table]
- Portuguese Remaridation Rates

and considering a Plan with the relations between sub-populations expressed in the one-step transition matrix :

$$P = \begin{bmatrix} \boldsymbol{P}_{a,a} & \boldsymbol{P}_{a,i} & \boldsymbol{P}_{a,r} & \boldsymbol{P}_{a,c^+} & \boldsymbol{P}_{a,c} & \boldsymbol{P}_{a,f} & \boldsymbol{p}_{a,s} \\ \boldsymbol{O} & \boldsymbol{P}_{i,i} & \boldsymbol{O} & \boldsymbol{P}_{i,c^+} & \boldsymbol{P}_{i,c} & \boldsymbol{P}_{i,f} & \boldsymbol{p}_{i,s} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{P}_{r,r} & \boldsymbol{P}_{r,c^+} & \boldsymbol{P}_{r,c} & \boldsymbol{P}_{r,f} & \boldsymbol{p}_{r,s} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{P}_{c^+,c^+} & \boldsymbol{P}_{c^+,c} & \boldsymbol{P}_{c^+,f} & \boldsymbol{p}_{c^+,s} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{P}_{c,c} & \boldsymbol{O} & \boldsymbol{p}_{c,s} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{P}_{f,f} & \boldsymbol{p}_{f,s} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Considering that each age of population elements corresponds to a state, we get a total of 305 states in the Markov chain.

For the initial classification probabilities we used a fixed vector estimated from data of some Portuguese Pension Funds.

4.3 Entrances in the Population

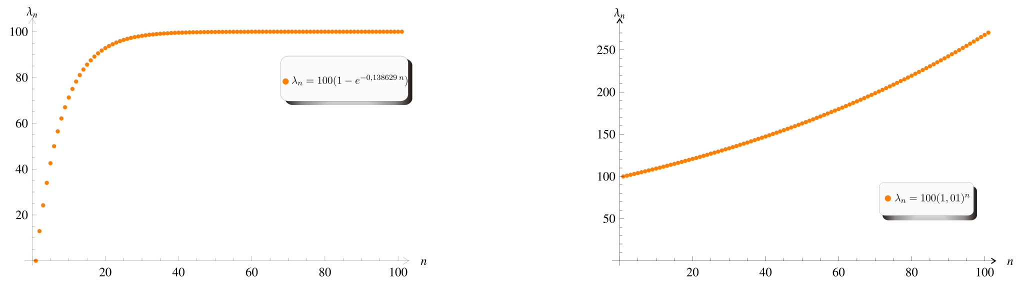
To implement the model we considered different scenarios for entrances in the population. In order to analyse the evolution of the population according to strong differences in the rates of entrances, we will consider two very different scenarios. In the first one we will assume an asymptotic growth in the number of new elements arriving the population, which is represented by $\lambda_i = \kappa(1 - e^{-\beta\, i})$, $i \in \mathbb{N}_0$. The second scenario will represent a population where the rate of new elements as geometric growth, which is represented by $\lambda_i = \rho\, r^i$, $i \in \mathbb{N}_0$. Note that both are particular cases of $\lambda_i = a + b\theta^i$, $i \in \mathbb{N}_0$. Parameters will be chosen in such a way that the first scenario corresponds to a case where $0 < \theta < 1$ and in the second one, θ will be greater than 1.
When choosing this scenarios we are trying to illustrate some different possibilities for the entrances in the population and how the model responds to those differences. We did not concern, at this moment, if any of those rates of entrances is realistic to a pension fund population.

Figure 1 illustrates the rate of entrances in both scenarios during a period of 100 years.

Table 1: Expected Entrances in Population

Scenario 1	**Scenario 2**
$\lambda_n = 100(1 - e^{-0{,}138629\, n})$	$\lambda_n = 100(1,01)^n$

Figure 1: Expected Entrances in Population - Scenarios 1 and 2



4.4 Sub-Population Dimension

Figure 2 illustrates the expected dimension of the total population, in both scenarios, over a period of 100 years. As we may see, for the first scenario the expected number of elements tend to stabilize after a certain period of time and in the second one the dimension of the total population keeps on growing over time.

Figure 2: Expected Dimension of Population - Scenarios 1 and 2

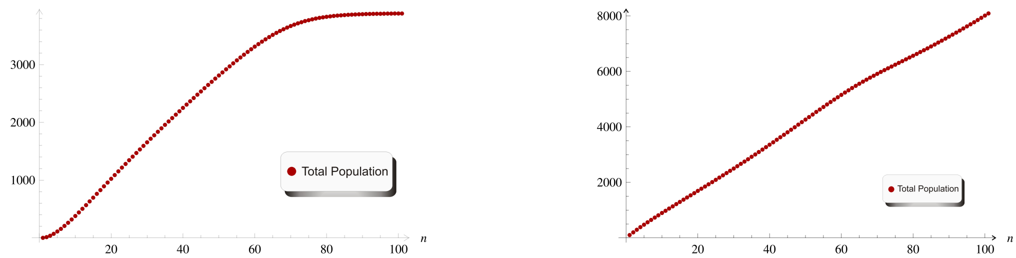
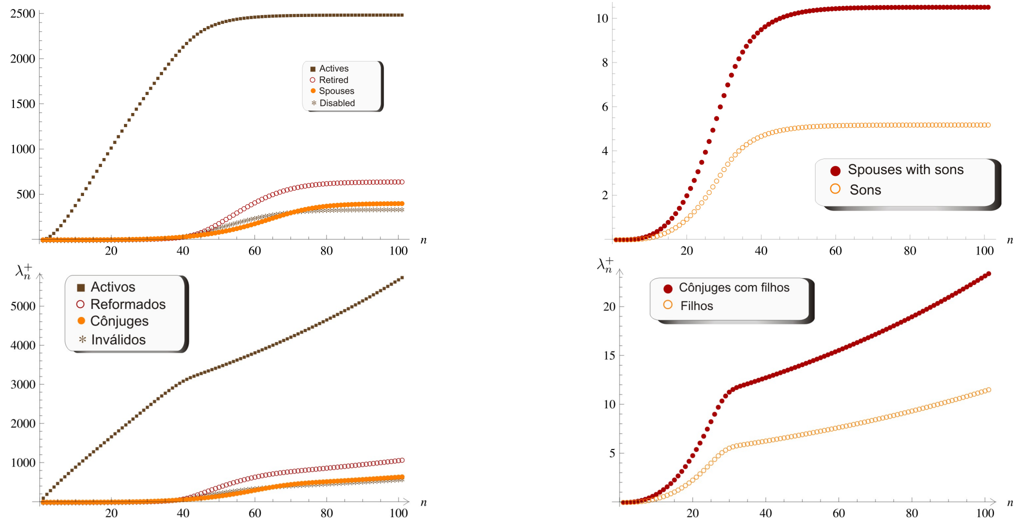


Figure 3 illustrates the sub-populations dimension over the same period of time. As we may see, the evolution of the sub-populations, in both scenarios, are similar to the evolution of the total population.

Figures 4 and 5 illustrate the evolution of the expected relative dimensions of sub-populations in scenarios 1 and 2, respectively. As we may note, despite the continuous growth of the expected absolute dimensions in scenario 2, the relative dimensions will also tend to be stable, which corresponds to the existence of a stochastic vortice in this population. This

Figure 3: Expected Dimension of Sub-Populations - Scenarios 1 and 2



existence of a stochastic vortice is verified in both scenarios.

Figure 4: Expected Relative Dimension of Sub-Populations - Scenario 1

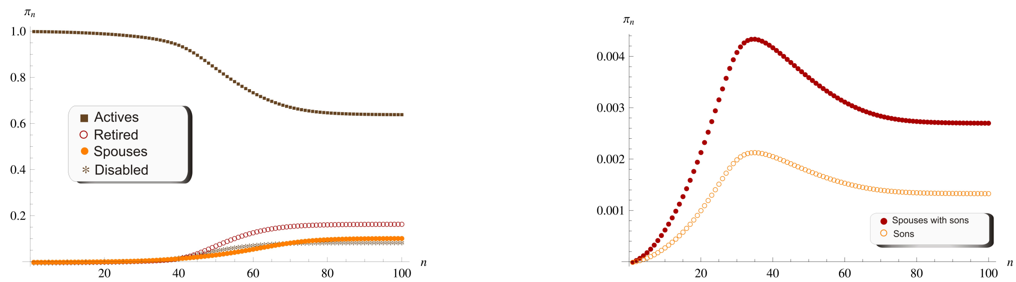


Table 2 illustrates the estimated dimensions of the sub-populations in both scenarios in a long time perspective.

4.5 Confidence Bands

Figure 6 illustrates some confidence intervals for the estimated dimensions calculated over a period of 100 years for three of the sub-populations - Spouses, Retired, Active Employees - and also for the Total Population. These confidence bands refer to the estimated dimensions in scenario 1.

Figure 5: Expected Relative Dimension of Sub-Populations - Scenario 2

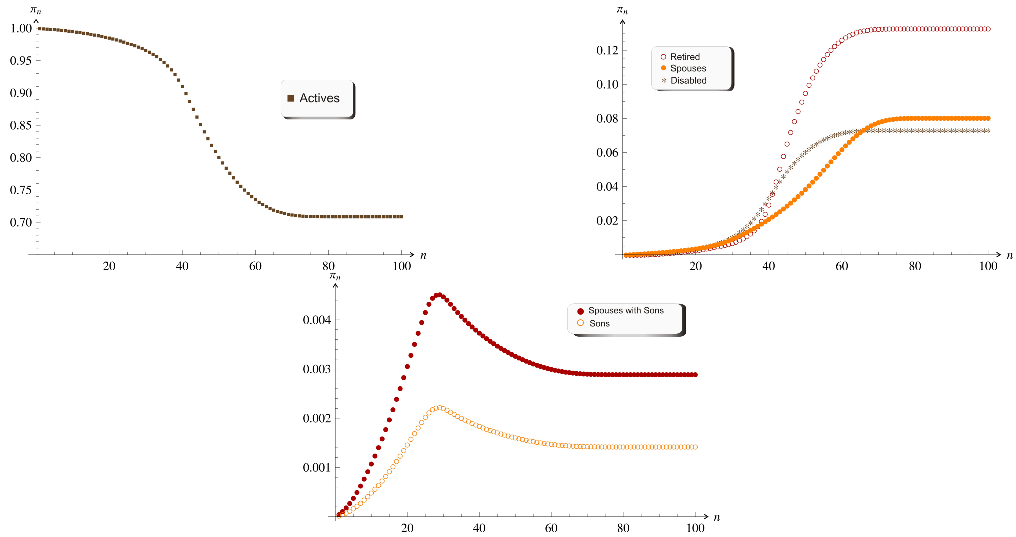
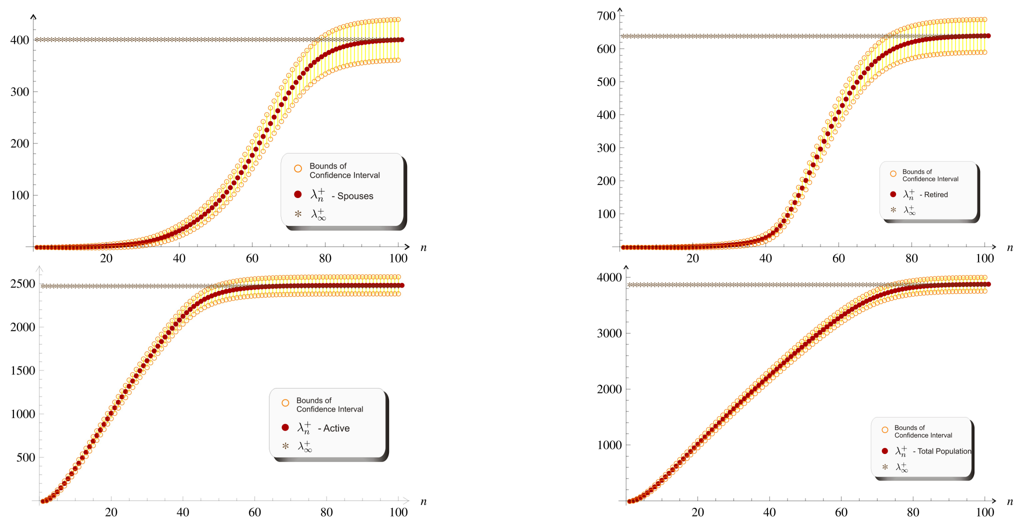


Figure 6: Confidence Bands



5 Conclusions

The stochastic vortices model is a useful model in the study of populations subject to periodic reclassifications. It is a model easy to implement and produces good results in a variety of situations where closed models fail. The fact that the model admits possibilities of entrances, departures and initial classification implies that the model produces results more representative of reality.

Table 2: Expected Dimension of sub-populations in long run perspective

Sub-Populations	Scenario 1		Scenario 2
	λ_{∞}^{+}	$\pi_{\infty,j}$	$\pi_{\infty,j}$
Actives	2.485,79	0,6392	0,7090
Disabled	342,29	0,0880	0,0735
Retired	642,27	0,1651	0,1328
Spouses with Sons	10,53	0,0027	0,0029
Spouses	403,22	0,1037	0,0804
Sons	5,20	0,0013	0,0014
Total	**3889,30**	**1,0000**	**1,0000**

References

[1] Gantmacher, F.; *The theory of matrices*, Vol.II, Chelsea, New York, 1960.

[2] Guerreiro, G. & Mexia, J., *Stochastic Vortices in periodically reclassified populations*, Discussiones Mathematicae, Probability and Statistics, accepted for publication.

[3] Guerreiro, G. & Mexia, J., *An Alternative Approach to Bonus Malus*, Discussiones Mathematicae, Probability and Statistics, 24, pp. 197 - 213 , 2004

[4] Parzen, E., *Stochastic Processes*, Holden Day, San Francisco, 1962

[5] Rolski, T.,Schmidli, H.,Schmidt, V. & Teugels, J., *Stochastic processes for insurance and finance*, John Wiley & Sons, 1999

[6] Schott, J.R., *Matrix Analysis for Statistics*, New York, John Wiley and Sons, Inc., 1997

[7] Winklevoss, H.; Pension Mathematics with numerical illustrations, 2^{nd} ed., Pension Council of the Wharton School of the University of Pennsylvania, 1993.

Subexponential Tails in the World of Dependence - Short Course in the 5th Conference in Actuarial Science & Finance on Samos

Qihe Tang
Department of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242, USA
E-mail: qtang@stat.uiowa.edu

September 1-3, 2008

Abstract

The study of subexponentiality has formed a basis of studies concerning tail probabilities. However, the mainstream of this study has been restricted to independent cases and available methods can hardly be used to deal with dependent cases. Motivated by applications in many applied fields including insurance and finance, we propose to promote this theory to dependent cases.

(1) A remarkable feature of subexponentiality is that the sum and maximum of independent, identically distributed, subexponential random variables are tail asymptotic, known as the principle of a single big jump. We shall consider to what extent the underlying random variables can be dependent while this principle still holds.

(2) We shall also study the tail behavior of the product of dependent subexponential random variables. The product is more intractable than the sum due to the very definition of subexponentiality. We shall consider how the product inherits the tail behavior of its factors.

(3) We shall derive explicit asymptotic formulas for the tail probabilities of weighted sums of finite or infinite terms, where the random weights are dependent and are allowed to be dependent on the primary subexponential random variables.

1 Introduction

The study of the tail probabilities of quantities of certain stochastic structure such as sums of random variables is of fundamental interest in many applied fields including insurance and finance. However, it is usually not possible to get closed-form expressions except for very ideal cases. People have to add certain regular conditions on the underlying distributions so as to alternatively derive asymptotic formulas. Furthermore, practical studies show that empirical data in applied fields, particularly in insurance and finance, are usually heavy tailed. Subexponentiality emerged from this trend four decades ago.

The concept of subexponentiality was independently introduced by Chistyakov (1964) and Chover *et al.* (1973) in the context of branch processes. An early textbook treatment was given by Athreya and Ney (1972). The potential role of subexponential distributions within queueing theory was recognized by Pakes (1975) and Borovkov (1976, 1984). The importance of the subexponential class as a useful class of heavy-tailed distributions in the context of applied probability, in general, and insurance mathematics, in particular, was realized by Teugels (1975). Reviews of subexponential distributions can be found in Bingham *et al.* (1989), Embrechts *et al.* (1997), Rolski *et al.* (1999), Asmussen (2000), Rachev (2003), and Resnick (2007), among others.

Through the participation of many people in theoretical and applied probability, a beautiful picture for the theory, methodologies, and applications of subexponentiality is coming into being. A rather complete list of references on subexponentiality can be retrieved from MathSciNet (an official web site of American Mathematical Society) by inputting appropriate key words such as 'subexponential', 'heavy tail', 'regular variation', etc. in the search item 'title'. From the list we see many major figures in the worldwide community of theoretical and applied probability working on the topic. Subexponentiality has important applications in various areas, particularly in insurance and finance, as summarized by Embrechts *et al.* (1997).

As in most other branches of probability theory, mainstream study of subexponentiality has been restricted to independent cases. We plan to promote subexponential theory to dependent cases. As the tail behavior of more complicated stochastic models can usually be reduced to the tail behavior of sums or products or sums of products of random variables, we are going to raise some questions which we think are interesting and useful, but challenging as well. Very often, questions related to subexponentiality look simple and elementary but to solve them could be very hard or impossible. More precisely, we shall study the asymptotic tail behavior of

- sums of dependent random variables;
- products of dependent random variables;
- finite sums with dependent random weights; and
- infinite sums with dependent random weights.

The scenarios for independent and dependent cases may be sometimes similar sometimes different. However, most available methods developed in the study of subexponentiality can hardly be used to deal with dependent cases. It is likely that techniques dealing with independent and dependent cases are completely different.

Example 1 Consider a discrete-time risk model in which the surplus of the insurance company is invested into a risky asset that generates a random, possibly negative, return rate in each period. Denote by $A_n \in (-\infty, \infty)$ the net income (the total premium income minus the total claim amount) within period n and by $R_n \in (-1, \infty)$ the random return rate in period n, $n = 1, 2, \ldots$. Let the initial surplus be $x \geq 0$. Hence, if we assume that the net income A_n is calculated at the end of period n then the surplus, denoted by S_n, accumulated till the end of period n satisfies the recursive equation

$$S_0 = x \geq 0, \quad S_n = (1 + R_n) S_{n-1} + A_n, \qquad n = 1, 2, \ldots.$$

Now write

$$X_n = -A_n, \quad Y_n = \frac{1}{1 + R_n}, \qquad n = 1, 2, \ldots.$$

The random variable X_n is the net payout within period n and the random variable Y_n is the discount factor from time n to time $n - 1$, $n = 1, 2, \ldots$. In the terminology of Norberg (1999) as well as Tang and Tsitsiashvili (2003b, 2004), we call X_n, $n = 1, 2, \ldots$, the insurance risks and Y_n, $n = 1, 2, \ldots$, the financial risks.

The discounted value of the surplus process is defined as

$$\widetilde{S}_0 = x, \quad \widetilde{S}_n = S_n \prod_{j=1}^{n} Y_j, \qquad n = 1, 2, \ldots.$$

It is easy to see that $\widetilde{S}_n$ can also be expressed as

$$\widetilde{S}_0 = x, \quad \widetilde{S}_n = x - \sum_{i=1}^{n} X_i \prod_{j=1}^{i} Y_j, \qquad n = 1, 2, \ldots.$$

See Nyrhinen (1999, 2001) and Tang and Tsitsiashvili (2003b, 2004). □

Example 2 The study of tail probabilities is of fundamental interests in financial risk management as most risk measures are calculated through tail probabilities. Consider the sum

$$S_n = X_1 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are n dependent risks. We shall focus on the asymptotic behavior of the tail probability $\Pr(S_n > x)$ for large x. This is relevant for applications in financial risk management. As an example, let us look at calculation of the Conditional Tail Expectation (CTE) of S_n. For an overview of risk measures, see Dhaene *et al.* (2006) and McNeil *et al.* (2005). The confidence level α should be chosen to be close to 1 (typically $\alpha = 95\%$ or 99%), indicating that the corresponding Value at Risk (VaR) or Quantile $Q_\alpha[S_n]$ should be large. Note that

$$\begin{aligned}\mathrm{CTE}_\alpha[S_n] &= \mathrm{E}[S_n|S_n > Q_\alpha[S_n]] \\ &= Q_\alpha[S_n] + \frac{\mathrm{E}\left[(S_n - Q_\alpha[S_n])_+\right]}{\Pr(S_n > Q_\alpha[S_n])} \\ &= \inf\{x : \Pr(S_n > x) \le 1-\alpha\} + \frac{\int_{Q_\alpha[S_n]}^\infty \Pr(S_n > x)\,\mathrm{d}x}{\Pr(S_n > Q_\alpha[S_n])}. \qquad (1)\end{aligned}$$

Therefore, if we have a transparent and efficient approximation for $\Pr(S_n > x)$ for large x, then by plugging it into the right-hand side of (1) we can immediately obtain an efficient approximation for $\mathrm{CTE}_\alpha[S_n]$.

As commented by Bingham *et al.* (2003), "the empirical evidence suggests that most financial data show both pronounced asymmetry and much heavier tail behaviour than is consistent with normality". We may assume that the random variables $X_1, \ldots, X_n$ follow subexponential distributions and then pursue some efficient approximations as desired.

Based on the fact that Tail VaR preserves the convex order, it is possible to construct upper and lower bounds for this risk measure by using the concept of comonotonicity; see Dhaene *et al.* (2002a, 2002b) and references therein. Their approach is generalized to the class of concave distortion risk measures; see Dhaene *et al.* (2006). In doing so, for the upper bound it is hard to capture the impact of any subtle dependence information. The lower bound on the other hand, which is based on a conditioning technique, leads to much better performance in this respect. Unfortunately, until now the lower bound approach can only be applied to sums of lognormal random variables, with some extensions to the class of multivariate elliptical distributions.

To extend the study to more useful scenarios, we shall continue to consider the tail behavior of the weighted sum

$$S_n^{(w)} = w_1X_1 + \cdots + w_nX_n,$$

where $w_1, \ldots, w_n$ are n other positive (random or nonrandom) variables representing weights of the primary random variables $X_1, \ldots, X_n$.

Our results allow immediate applications to financial risk management. As an example, we consider the capital allocation $\mathbf{w} \in R_+^n$ among n risky assets with returns $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$. Write $\mathbf{X} = -\mathbf{Y}$ and introduce a portfolio risk measure ρ. Then, any risk-averse investor will choose a portfolio solution of the following optimization problem:

$$\mathbf{w} = \arg\min_{\mathbf{w}\in R_+^n} \rho\left(\mathbf{w}^\top\mathbf{X}\right) \qquad \text{such that} \quad \mathbf{w}^\top\mathbf{e} = 1.$$

It is most likely that we can not get an exact solution. However, a good asymptotic formula for the tail probability of $S_n^{(w)}$ enables us to derive an explicit asymptotic solution. $\square$

2 A Brief Review on Subexponentiality

Hereafter, all limit relationships are according to $x \to \infty$ unless stated otherwise, the symbol $\sim$ means that the quotient of both sides tends to 1, the symbol $\lesssim$ means that the upper limit of the quotient is not more than 1, and the symbol $\gtrsim$ is understood in a similar way. By saying that a distribution F is supported on $[0, \infty)$ we mean $F(0-) = 0$ and $\overline{F}(x) = 1 - F(x) > 0$ for all x. If F on $[0, \infty)$ has a finite mean $\mu > 0$, then its equilibrium distribution is defined as

$$F_e(x) = \frac{1}{\mu} \int_0^x \overline{F}(y)\,\mathrm{d}y, \qquad x \geq 0.$$

The moment generating function at some real number r of a measure ν on $[0, \infty)$ is defined to be

$$\hat{\nu}(r) = \int_{0-}^{\infty} \mathrm{e}^{rx} \nu(\mathrm{d}x).$$

2.1 Definition of Subexponentiality

A distribution F on $[0, \infty)$ is said to be subexponential, written as $F \in \mathcal{S}$, if the relation

$$\lim_{x \to \infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} = 2 \tag{2}$$

holds, where $F^{2*}(x) = \int_{0-}^{\infty} \overline{F}(x-y)F(\mathrm{d}y)$, $x \geq 0$, denotes the convolution of F with itself. Table 1.2.6 of Embrechts *et al.* (1997) shows that the subexponential class contains a lot of popular distributions such as Pareto, Lognormal, heavy-tailed Weibull, and Loggamma distributions.

Examples for Subexponential Distributions
(F = distribution function, f = density function)

- Lognormal: for $-\infty < \mu < \infty$ and $\sigma > 0$,

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\{-(\ln x - \mu)^2/(2\sigma^2)\};$$

- Pareto: for $\alpha > 0,\ \kappa > 0$,

$$\overline{F}(x) = \left(\frac{\kappa}{\kappa + x}\right)^{\alpha};$$

- Burr: for $\alpha > 0,\ \kappa > 0,\ \tau > 0$,

$$\overline{F}(x) = \left(\frac{\kappa}{\kappa + x^{\tau}}\right)^{\alpha};$$

- Benktander-type I: for $\alpha > 0,\ \beta > 0,$

$$\overline{F}(x) = (1 + 2(\beta/\alpha)\ln x)\exp\{-\beta(\ln x)^2 - (\alpha+1)\ln x\};$$

- Benktander-type II: for $\alpha > 0,\ 0 < \beta < 1,$

$$\overline{F}(x) = e^{\alpha/\beta}x^{-(1-\beta)}\exp\{-\alpha x^{\beta}/\beta\};$$

- Weibull: for $c > 0,\ 0 < \tau < 1,$

$$\overline{F}(x) = \exp\{-cx^{\tau}\};$$

- Loggamma: for $\alpha > 0,\ \beta > 0,$

$$f(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1};$$

A remarkable feature of subexponentiality is that, for independent and identically distributed (i.i.d.) random variables $X_1,\ X_2,\ \ldots$ with common distribution $F \in \mathcal{S}$, the relation

$$\Pr\left(\sum_{k=1}^{n} X_k > x\right) \sim n\overline{F}(x) \tag{3}$$

holds for all $n = 1, 2, \ldots$; see Embrechts *et al.* (1979). Since $\Pr\left(\max_{1\le k\le n} X_k > x\right) \sim n\overline{F}(x)$, relation (3) means that, in the tail asymptotic sense, the maximum of the first n random variables exhausts the sum. This explains why subexponentiality is useful in modeling heavy-tailed distributions. In extreme value theory, it is generally accepted that an extreme event happens mainly due to a single unusually large input to the stochastic system, known as the principle of a single big jump.

This feature of subexponentiality is especially relevant nowadays, in view of enormous insurance claims made in the aftermath of September 11, 2001 attacks, the 2004 Indian Ocean Tsunami, the 2005 Hurricane Katrina, the 2007 Californian fires, and the most recent 2008 Sichuan earthquake.

2.2 Related Classes of Distributions

It is well known that each subexponential distribution F is long tailed, written as $F \in \mathcal{L}$, in the sense that the relation

$$\overline{F}(x + a) \sim \overline{F}(x)$$

holds for all real numbers $a > 0$; see Chistyakov (1964) or Lemma 1.3.5(a) of Embrechts *et al.* (1997). One easily sees that, for every distribution $F \in \mathcal{L}$, there is some function $a(\cdot) : [0, \infty) \to [0, \infty)$ such that the following items hold simultaneously:

$$a(x) \to \infty, \qquad a(x) = o(x), \qquad \overline{F}(x \pm a(x)) \sim \overline{F}(x). \tag{4}$$

People have introduced some subclasses of $\mathcal{S}$ in various situations in order to overcome technical difficulties due to excessive generality of the concept of subexponentiality. The class $\mathcal{S}^*$ was introduced by Klüppelberg (1988) and is characterized by the relation

$$\int_0^x \overline{F}(x-y)\overline{F}(y)\mathrm{d}y \sim 2\mu\overline{F}(x)$$

with $0 < \mu < \infty$ being the mean of F supported on $[0, \infty)$. It is known that if $F \in \mathcal{S}^*$ then both $F \in \mathcal{S}$ and $F_e \in \mathcal{S}$. This class is marginally smaller than $\mathcal{S}$ but it enjoys several nicer properties than $\mathcal{S}$. Taking full advantages of these properties of $\mathcal{S}^*$, Foss and Zachary (2003) and Foss *et al.* (2005) studied the probability that a random walk crosses a high boundary on a random time interval and they derived an elegant asymptotic formula which holds uniformly over all stopping times and a wide class of nonlinear boundaries. This result significantly extends the original work of Asmussen (1998). Nevertheless, their proof relies heavily on the i.i.d. assumption and leaves little room for extension to dependent cases.

Another commonly used subclass of $\mathcal{S}$ is the intersection $\mathcal{L} \cap \mathcal{D}$, where $\mathcal{D}$ is the class of distributions with dominatedly-varying tails (with dominated variation) characterized by the relation

$$\limsup_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} < \infty \qquad \text{for all } 0 < y < 1.$$

Clearly, if $F \in \mathcal{D}$ then it holds for every $y > 0$ that

$$0 < \liminf_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} \leq \limsup_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} < \infty. \tag{5}$$

In particular, $\mathcal{L} \cap \mathcal{D}$ contains all distributions with regularly-varying tails characterized by the relation

$$\lim_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} = y^{-\alpha}, \qquad y > 0,$$

for some $\alpha \geq 0$. Denote by $F \in \mathcal{R}_{-\alpha}$ the regularity property above, so that $\mathcal{R}$ is the union of all $\mathcal{R}_{-\alpha}$ over the range $0 \leq \alpha < \infty$.

Thanks to the well-developed Karamata theory, it is usually much easier to handle distributions from $\mathcal{R}$ than to handle other subexponential distributions. Results similar to but weaker than Karamata theorems are also available for distributions from $\mathcal{D}$. To see this, we further introduce two indices of a distribution. Let F be a distribution on $[0, \infty)$ and write $f(x) = \left(\overline{F}(x)\right)^{-1}$, which is a positive and non-decreasing function on $(-\infty, \infty)$. Let J_F^+ be the infimum of those J for which there exists a constant $A = A(J)$ such that for each $\Lambda > 1$, the relation

$$\frac{f(\lambda x)}{f(x)} \leq A(1 + o(1))\lambda^J$$

holds uniformly in $\lambda \in [1, \Lambda]$, and let J_F^- be the supremum of those J for which there exists a constant $B = B(J)$ such that for each $\Lambda > 1$, the relation

$$\frac{f(\lambda x)}{f(x)} \geq B(1 + o(1))\lambda^J$$

holds uniformly in $\lambda \in [1, \Lambda]$. The quantities J_F^+ and J_F^- define the upper and lower Matuszewska indices of the function $f(x)$, respectively. By Theorem 2.1.5 and Corollary 2.1.6 of Bingham *et al.* (1989), they coincide with

$$J_F^+ = \inf\left\{-\frac{\log \overline{F}_*(y)}{\log y} : y > 1\right\}, \quad J_F^- = \sup\left\{-\frac{\log \overline{F}^*(y)}{\log y} : y > 1\right\},$$

where

$$\overline{F}_*(y) = \liminf_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)}, \quad \overline{F}^*(y) = \limsup_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)}.$$

Following Tang and Tsitsiashvili (2003a), we simply call J_F^+ and J_F^- the upper and lower Matuszewska indices of F, respectively.

Clearly, if $F \in \mathcal{D}$ then $J_F^+ < \infty$, and if $F \in \mathcal{R}_{-\alpha}$ then $J_F^- = J_F^+ = \alpha$. From Proposition 2.2.1 of Bingham *et al.* (1989), we see that, for any $p_1 < J_F^-$ and $p_2 > J_F^+$, there are positive constants C and D such that the two-sided inequality

$$C^{-1}y^{-p_2} \leq \frac{\overline{F}(xy)}{\overline{F}(x)} \leq Cy^{-p_1} \tag{6}$$

holds for all $xy \geq x \geq D$. In particular, from (6) one easily sees that if $F \in \mathcal{D}$ then the relation

$$x^{-p} = o(\overline{F}(x)) \tag{7}$$

holds for all $p > J_F^+$. For more details of (6) and (7), see e.g. Lemma 3.5 of Tang and Tsitsiashvili (2003a) and its proof.

Using (6) and (7), research related to distributions from $\mathcal{L} \cap \mathcal{D}$ is usually doable under less restrictive conditions. There are essential difficulties in extending the discussion to broader subclasses of $\mathcal{S}$ to include lognormal and Weibull distributions.

As an extremal situation of $\mathcal{R}_{-\alpha}$, the class $\mathcal{R}_{-\infty}$ of distributions with rapidly-varying tails (with rapid variation) is characterized by the relation

$$\lim_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} = 0 \qquad \text{for all } y > 1.$$

We remark that $\mathcal{R}_{-\infty}$ is a broad class containing both heavy-tailed and light-tailed distributions and that it well complements the class $\mathcal{R}_{-\alpha}$ with $\alpha = \infty$. Rapid variation has been investigated by many researchers in applied probability since de Haan (1970). Recent applications of the class $\mathcal{R}_{-\infty}$ can be found in Tang and Tsitsiashvili (2004) and Barbe and McCormick (2008), among others.

In the context of ruin theory, Konstantinides *et al.* (2002) first introduced the class $\mathcal{A}$. In terms of the lower Matuszewska index, we can restate its definition as that a distribution F on $[0,\infty)$ belongs to the class $\mathcal{A}$ if F is subexponential and has a lower Matuszewska index $0 < J_F^- \leq \infty$. Clearly, the condition $0 < J_F^- \leq \infty$ is equivalent to the condition

$$\limsup_{x\to\infty} \frac{\overline{F}(vx)}{\overline{F}(x)} < 1 \qquad \text{for some } v > 1, \tag{8}$$

which is really a mild restriction in view of the fact that it is fulfilled by almost all useful distributions with infinite supports. From this point of view, the class $\mathcal{A}$ almost coincides with the class $\mathcal{S}$.

2.3 Two Examples

The definition of subexponentiality opens a natural way to derive explicit asymptotic estimates for sums of random variables in general probabilistic models. Let us look at the following two examples for some insights.

Example 3 (Geometric Sums) Consider the random sum

$$S_\tau = \sum_{i=1}^{\tau} X_i,$$

where $X_1, X_2, \ldots$ are i.i.d. random variables with common distribution $F \in \mathcal{S}$, while τ is an integer-valued nonnegative random variable independent of $X_1, X_2, \ldots$ and having a finite moment generating function at a neighborhood of zero. It is well known that the ultimate maximum of a random walk with negative drift can be expressed as this random sum with τ geometrically distributed; see Feller (1971) and Kalashnikov (1997). Recall a classical result that, for every $\varepsilon > 0$, there exists some absolute constant $C_\varepsilon > 0$ such that the inequality

$$\Pr(S_n > x) \leq C_\varepsilon (1+\varepsilon)^n \overline{F}(x) \tag{9}$$

holds for all $n = 1, 2, \ldots$ and $x \geq 0$; see e.g. Theorem 1.3.5(c) of Embrechts *et al.* (1997). Hence, an application of dominated convergence gives

$$\Pr(S_\tau > x) = \sum_{n=1}^{\infty} \Pr(S_n > x) \Pr(\tau = n) \sim \mathrm{E}\tau \overline{F}(x). \tag{10}$$

Example 4 (Gerber-Shiu Functions) This example is the germ of the recent work of Tang and Wei (2009). Consider the compound Poisson risk model in which the surplus process of the insurance company is modelled as

$$P_t = x + ct - \sum_{i=1}^{N_t} X_i, \qquad t \geq 0,$$

where $x \geq 0$ is the initial capital, $c > 0$ is the constant premium rate, $\{X_1, X_2, \ldots\}$ denote sizes of successive claims forming a sequence of i.i.d. positive random variables with common continuous distribution F and finite mean μ, and these claims arrive according to a homogeneous Poisson process $\{N_t, t \geq 0\}$ with intensity $\lambda > 0$. As usual, we assume the safety loading condition $\lambda\mu < c$.

Define the time of ruin as

$$T = \inf\{t \geq 0 : P_t < 0\},$$

where $\inf \varnothing = \infty$ by convention. For a nonnegative constant δ and a bivariate nonnegative function $w(\cdot, \cdot)$ on $[0, \infty)^2$, the well-known Gerber-Shiu expected discounted penalty function is defined as

$$\phi_\delta(x) = \mathrm{E}\left[\mathrm{e}^{-\delta T} w\left(P_{T-}, |P_T|\right) 1_{(T<\infty)} \Big| P_0 = x \right].$$

When $\delta = 0$ and $w(\cdot, \cdot) \equiv 1$, the Gerber-Shiu function reduces to the ruin probability

$$\psi(x) = \Pr\left(T < \infty | P_0 = x\right).$$

The following technique of using renewal theory in the study is due to Gerber and Shiu (1998). Assume $\delta \geq 0$ and let $\rho = \rho(\delta) \geq 0$ solve the generalized Lundberg equation

$$\delta + \lambda - cs = \lambda \hat{F}(-s). \tag{11}$$

With this ρ, we have the defective renewal equation

$$\phi_\delta(x) = \frac{c\rho - \delta}{c\rho} \int_0^x \phi_\delta(x - y) g(y) \mathrm{d}y + h(x), \tag{12}$$

where

$$\begin{aligned} g(y) &= \frac{\lambda\rho}{c\rho - \delta} \int_y^\infty \mathrm{e}^{-\rho(z-y)} F(\mathrm{d}z), \qquad y \geq 0, \\ h(x) &= \frac{\lambda}{c} \int_x^\infty \mathrm{e}^{-\rho(y-x)} \int_y^\infty w(y, z-y) F(\mathrm{d}z) \mathrm{d}y. \end{aligned}$$

As remarked by Gerber and Shiu (1998), when $\delta = 0$ (so $\rho = 0$) we have the understanding

$$\frac{\delta}{\rho} = \lim_{s \to 0+} \frac{cs - \lambda\left(1 - \hat{F}(-s)\right)}{s} = c - \lambda\mu. \tag{13}$$

Hence for this case, relation (12) is simplified to

$$\phi_0(x) = \frac{\lambda\mu}{c} \int_0^x \phi_0(x - y) g(y) \mathrm{d}y + h(x),$$

where

$$\begin{aligned} g(y) &= \frac{1}{\mu}\overline{F}(y), \qquad y \geq 0, \\ h(x) &= \frac{\lambda}{c}\int_x^\infty \int_y^\infty w(y, z-y)F(\mathrm{d}z)\mathrm{d}y. \end{aligned}$$

It is easy to see that $g(\cdot)$ is a standard probability density function on $(0,\infty)$ for both $\delta > 0$ and $\delta = 0$. Write its corresponding cumulative distribution function as G. For simplicity, we assume $w(\cdot,\cdot) = w_1(\cdot)w_2(\cdot)$ with $w_1(\cdot) : [0,\infty) \to [0,\infty)$ non-increasing and $w_2(\cdot) : [0,\infty) \to [0,\infty)$ non-decreasing. Then, we have

$$\begin{aligned} h'(x) &= -\frac{\lambda}{c}w_1(x)\int_x^\infty w_2(z-x)F(\mathrm{d}z) + \frac{\rho\lambda}{c}\int_x^\infty \mathrm{e}^{-\rho(y-x)}\int_y^\infty w_1(y)w_2(z-y)F(\mathrm{d}z)\mathrm{d}y \\ &= -\frac{\lambda}{c}w_1(x)\int_x^\infty w_2(z-x)F(\mathrm{d}z) + \frac{\rho\lambda}{c}\int_x^\infty \int_x^z \mathrm{e}^{-\rho(y-x)}w_1(y)w_2(z-y)\mathrm{d}yF(\mathrm{d}z) \\ &\leq -\frac{\lambda}{c}w_1(x)\int_x^\infty w_2(z-x)F(\mathrm{d}z) + \frac{\rho\lambda}{c}w_1(x)\int_x^\infty \int_x^\infty \mathrm{e}^{-\rho(y-x)}w_2(z-x)\mathrm{d}yF(\mathrm{d}z) \\ &= 0, \end{aligned}$$

where the last-but-one line is due to the monotonicity of $w_1(\cdot)$ and $w_2(\cdot)$. Therefore, $h(\cdot)$ is non-increasing on $(0,\infty)$. Introduce a measure H on $(0,\infty)$ such that $H((x,\infty)) = h(x)$. Therefore, we may rewrite (12) as

$$\phi_\delta(x) = \int_{0-}^x h\,(x-y)\,V(\mathrm{d}y) = \overline{H * V}(x) - h(0)V((x,\infty)), \tag{14}$$

where

$$V(\cdot) = \sum_{n=0}^\infty \left(\frac{c\rho - \delta}{c\rho}\right)^n G^{*n}(\cdot). \tag{15}$$

Theorem 1 *Assume $\hat{w}_2(\varepsilon) < \infty$ for some $\varepsilon > 0$.*

(1) For $\delta > 0$, further assume $F \in \mathcal{S}$. Then,

$$\lim_{x\to\infty}\frac{\phi_\delta(x)}{\overline{F}(x)} = \frac{\lambda}{\delta}w_1(\infty)w_2(\infty).$$

(2) For $\delta = 0$, further assume $F_e \in \mathcal{S}$. Then,

$$\lim_{x\to\infty}\frac{\phi_0(x)}{\overline{F_e}(x)} = \frac{\lambda\mu}{c-\lambda\mu}w_1(\infty)w_2(\infty).$$

The following lemma is from Rogozin and Sgibnev (1999):

Lemma 1 *For some distribution $F \in \mathcal{S}$ and two other (possibly defective or excessive) distributions F_1 and F_2 on $[0,\infty)$ such that $k_i = \lim_{x\to\infty}\overline{F_i}(x)/\overline{F}(x)$ exists and is finite, $i = 1, 2$, it holds that*

$$\lim_{x\to\infty}\frac{\overline{F_1 * F_2}(x)}{\overline{F}(x)} = k_1F_2(\infty) + k_2F_1(\infty).$$

Proof of Theorem 1. We are to apply Lemma 1 to (14).

Step 1: Expressing G in F. When $\delta > 0$, we have

$$\begin{aligned}\overline{G}(x) &= \frac{\lambda\rho}{c\rho-\delta}\int_x^\infty\int_y^\infty \mathrm{e}^{-\rho(z-y)}F(\mathrm{d}z)\mathrm{d}y\\ &= \frac{\lambda}{c\rho-\delta}\int_x^\infty\left(1-\mathrm{e}^{-\rho(z-x)}\right)F(\mathrm{d}z)\\ &= \frac{\lambda\rho}{c\rho-\delta}\int_0^\infty\overline{F}(z+x)\mathrm{e}^{-\rho z}\mathrm{d}z\\ &\sim \frac{\lambda}{c\rho-\delta}\overline{F}(x).\end{aligned}$$

When $\delta = 0$ (so $\rho = 0$), we simply have

$$\overline{G}(x) = \frac{1}{\mu}\int_x^\infty\overline{F}(y)\mathrm{d}y = \overline{F_e}(x).$$

Thus, for each case, by closure of the class $\mathcal{S}$ under tail equivalence, we have $G \in \mathcal{S}$.

Step 2: Expressing V in G. Apply to (15) the dominated convergence theorem guaranteed by (9). When $\delta > 0$, we have

$$\begin{aligned}V((x,\infty)) &= \sum_{n=1}^\infty\left(\frac{c\rho-\delta}{c\rho}\right)^n\overline{G^{*n}}(x)\\ &\sim \overline{G}(x)\sum_{n=1}^\infty n\left(\frac{c\rho-\delta}{c\rho}\right)^n\\ &\sim \frac{\frac{c\rho-\delta}{c\rho}}{\left(1-\frac{c\rho-\delta}{c\rho}\right)^2}\frac{\lambda}{c\rho-\delta}\overline{F}(x)\\ &= \frac{\lambda}{\delta^2}\overline{F}(x).\end{aligned}$$

When $\delta = 0$, we have

$$V((x,\infty)) \sim \overline{G}(x)\sum_{n=1}^\infty n\left(\frac{\lambda\mu}{c}\right)^n = \frac{\lambda\mu}{c}\overline{F_e}(x).$$

Step 3: Expressing h in G. We have

$$\begin{aligned}h(x) &= \frac{\lambda}{c}\int_x^\infty \mathrm{e}^{-\rho(y-x)}w_1(y)\left(w_2(0)\overline{F}(y)+\int_0^\infty\overline{F}(z+y)w_2(\mathrm{d}z)\right)\mathrm{d}y\\ &= \frac{\lambda}{c}\int_x^\infty \mathrm{e}^{-\rho(y-x)}w_1(y)\left(\int_{0-}^\infty\overline{F}(z+y)w_2(\mathrm{d}z)\right)\mathrm{d}y\\ &\sim \frac{\lambda}{c}\int_x^\infty \mathrm{e}^{-\rho(y-x)}w_1(y)\left(\int_{0-}^\infty w_2(\mathrm{d}z)\right)\overline{F}(y)\mathrm{d}y\\ &= \frac{\lambda}{c}w_2(\infty)\int_0^\infty \mathrm{e}^{-\rho y}w_1(x+y)\overline{F}(x+y)\mathrm{d}y,\end{aligned}$$

where in the third step we used the dominated convergence theory guaranteed by $F \in \mathcal{L}$ and $\hat{w}_2(\varepsilon) < \infty$. Note that $w_1(\infty) \in [0, \infty)$ is well defined. Thus, when $\delta > 0$,

$$\lim_{x\to\infty} \frac{h(x)}{\overline{F}(x)} = \frac{\lambda}{c} w_1(\infty) w_2(\infty) \lim_{x\to\infty} \int_0^\infty \mathrm{e}^{-\rho y} \frac{\overline{F}(x+y)}{\overline{F}(x)} \mathrm{d}y = \frac{\lambda}{c\rho} w_1(\infty) w_2(\infty),$$

while when $\delta = 0$,

$$\lim_{x\to\infty} \frac{h(x)}{\overline{F_e}(x)} = \frac{\lambda\mu}{c} w_1(\infty) w_2(\infty).$$

Step 4: Conclusion. To derive an explicit asymptotic formula for $\phi_\delta(x)$, we notice that, when $\delta > 0$,

$$V(\infty) = \sum_{n=0}^{\infty} \left(\frac{c\rho - \delta}{c\rho} \right)^n = \frac{c\rho}{\delta},$$

while when $\delta = 0$,

$$V(\infty) = \sum_{n=0}^{\infty} \left(\frac{\lambda\mu}{c} \right)^n = \frac{c}{c - \lambda\mu}.$$

For both $\delta > 0$ and $\delta = 0$, a simple application of Lemma 1 to relation (14) completes the proof of Theorem 1. □

By Theorem 1(2) with $\delta = 0$ and $w(\cdot, \cdot) \equiv 1$, under $F_e \in \mathcal{S}$ we have

$$\lim_{x\to\infty} \frac{\psi(x)}{\overline{F_e}(x)} = \frac{\lambda\mu}{c - \lambda\mu},$$

which was first obtained by Embrechts and Veraverbeke (1982) (see also Theorem 1.3.6 of Embrechts *et al.* (1997)).

3 Tail Behavior of Sums

Let $X_1, \ldots, X_n$ be nonnegative random variables with distributions $F_1, \ldots, F_n$, respectively. By saying that $(X_1^*, \ldots, X_n^*)$ is an independent copy of $(X_1, \ldots, X_n)$ we mean that $(X_1^*, \ldots, X_n^*)$ and $(X_1, \ldots, X_n)$ are two independent random vectors with the same marginal distributions and the components of $(X_1^*, \ldots, X_n^*)$ are independent. Write S_n and S_n^* their sums, respectively.

First, assume that $X_1, \ldots, X_n$ are independent. If $F_k \in \mathcal{R}$ for all $k = 1, \ldots, n$, then the distribution of S_n belongs to $\mathcal{R}$ and the relation

$$\Pr(S_n > x) \sim \sum_{k=1}^{n} \overline{F_k}(x) \tag{16}$$

holds; see page 278 of Feller (1971) or Lemma 1.3.1 of Embrechts *et al.* (1997). Relation (16) succeeds (3) to work for non-identical cases. A close look at the original proof tells that this result can be extended to the class $\mathcal{L} \cap \mathcal{D}$; see Cai and Tang (2004). Actually, relation (16) even holds for the whole subexponential class $\mathcal{S}$ as long as one of $\overline{F_1}, \ldots, \overline{F_n}$ dominates all of the others; see Embrechts and Goldie (1980) and Cline (1986).

We consider to what extent the random variables $X_1, \ldots, X_n$ can be dependent while relation (16) remains valid. The context of this section is based on Ko and Tang (2008), Chen and Yuen (2009), and Geluk and Tang (2009).

In the first result below we consider the case $F_k \in \mathcal{D} \cap \mathcal{L}$ for all $k = 1, \ldots, n$. As for the dependence structure between $X_1, \ldots, X_n$, we assume that:

The relation

$$\lim_{x_i \wedge x_j \to \infty} \Pr\left(X_i > x_i \mid X_j > x_j \right) = 0 \qquad \text{Assumption A}$$

holds for all $1 \leq i \neq j \leq n$.

Assumption A allows a wide range of dependence structures. For example, recall that an n-dimensional distribution is called a Farlie-Gumbel-Morgenstern (FGM) distribution if it has the form

$$F(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_k(x_k) \left(1 + \sum_{1 \leq i < j \leq n} a_{ij} \overline{F_i}(x_i) \overline{F_j}(x_j) \right), \tag{17}$$

where $F_1, \ldots, F_n$ are the corresponding marginal distributions and a_{ij} are real numbers fulfilling certain requirements so that $F(x_1, ..., x_n)$ is a proper n-dimensional distribution. We refer the reader to Kotz *et al.* (2000) for a general account on multivariate FGM distributions. Clearly, if the random variables $X_1, \ldots, X_n$ follow a joint n-dimensional FGM distribution (17), then for all $1 \leq i \neq j \leq n$, the random variables X_i and X_j follow the joint distribution

$$F_{ij}(x_i, x_j) = F_i(x_i) F_j(x_j) \left(1 + a_{ij} \overline{F_i}(x_i) \overline{F_j}(x_j) \right),$$

so that Assumption A is satisfied.

Intuitively, Assumption A requires that the dependence structure between X_1 and X_2 should not be too strongly positive. Hence, it excludes extremely positive dependence structures such as comonotonicity.

Remark 1 For simplicity, let F_1 and F_2 be absolutely continuous and let X_1 and X_2 be dependent according to a copula $C(u_1, u_2)$ for $(u_1, u_2) \in [0, 1]^2$. Thus, the joint distribution of X_1 and X_2 is given by

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2));$$

see, e.g. page 15 of Nelsen (2006). Let $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$, so that they are two uniform random variables following the joint distribution $C(u_1, u_2)$. Under Assumption

A, it is clear that X_1 and X_2 are tail independent in the sense that the (upper) tail dependence measure, defined by

$$\chi = \lim_{u\to 1} \Pr\left(U_2 > u | U_1 > u\right),$$

is equal to 0. □

Theorem 2 *If $F_k \in \mathcal{D} \cap \mathcal{L}$ for all $k = 1, \ldots, n$ and Assumption A holds, then relation (16) holds.*

Proof of Theorem 2. Clearly,

$$\Pr\left(S_n > x\right) \geq \Pr\left(\max_{1\leq i\leq n} X_i > x\right) \sim \sum_{i=1}^{n} \overline{F_i}(x). \tag{18}$$

On the other hand, choose a function $a(\cdot) : [0, \infty) \to [0, \infty)$ such that the items in (4) hold for all F_1, ..., F_n. We find that

$$\begin{aligned}\Pr\left(S_n > x\right) &\leq \Pr\left(\max_{1\leq i\leq n} X_i > x - a(x)\right) + \Pr\left(S_n > x, \max_{1\leq i\leq n} X_i \leq x - a(x)\right) \\ &= I_1(x) + I_2(x).\end{aligned}$$

For $I_1(x)$, we have

$$I_1(x) \sim \sum_{i=1}^{n} \overline{F_i}(x - a(x)) \sim \sum_{i=1}^{n} \overline{F_i}(x).$$

To deal with $I_2(x)$, without loss of generality we assume $n \geq 2$. By Assumption A,

$$\begin{aligned}I_2(x) &= \Pr\left(S_n > x, \max_{1\leq i\leq n} X_i > \frac{x}{n}, \max_{1\leq i\leq n} X_i \leq x - a(x)\right) \\ &\leq \sum_{i=1}^{n} \Pr\left(S_n - X_i > a(x), X_i > \frac{x}{n}\right) \\ &\leq \sum_{i=1}^{n} \sum_{j:1\leq j\leq n, j\neq i} \Pr\left(X_j > \frac{a(x)}{n-1}, X_i > \frac{x}{n}\right) \\ &\leq o(1) \sum_{i=1}^{n} \sum_{j:1\leq j\leq n, j\neq i} \Pr\left(X_i > \frac{x}{n}\right) \\ &\leq o(1) \sum_{i=1}^{n} \overline{F_i}(x),\end{aligned}$$

where the last step is due to the property of the class $\mathcal{D}$ as described in (5). We conclude that

$$\Pr\left(S_n > x\right) \lesssim \sum_{i=1}^{n} \overline{F_i}(x).$$

This ends the proof of Theorem 2. □

The assumption $F_k \in \mathcal{D} \cap \mathcal{L}$ for all $k = 1, \ldots, n$ in Theorem 2 indicates that their tails behave essentially like power functions. Hence, some important subexponential distributions such as lognormal and Weibull distributions are unfortunately excluded.

We then attempt to establish relation (16) for the case of subexponential marginal distributions. In doing so, however, we have to strengthen the assumption of dependence from Assumption A to the following:

There exist positive constants x_0 and c such that the inequality

$$\Pr(X_i > x_i \mid X_j = x_j \text{ with } j \in J) \le c\overline{F_i}(x_i) \qquad \text{Assumption B}$$

holds for all $1 \le i \le n$, $\emptyset \ne J \subset \{1, \ldots, n\}\backslash\{i\}$, $x_i > x_0$, and $x_j > x_0$ with $j \in J$.

When x_j is not a possible value of X_j, i.e. $\Pr(X_j \in \Delta) = 0$ for some open set Δ containing x_j, the conditional probability in Assumption B is simply understood as 0.

This dependence structure is related to the so-called negative (or positive) regression dependence introduced by Lehmann (1966). In particular, it is easy to check that this assumption is still satisfied if the random variables $X_1, \ldots, X_n$ follow a joint n-dimensional FGM distribution (17) whose marginal distributions F_k for $k = 1, \ldots, n$ are absolutely continuous. Obviously, Assumption B implies Assumption A.

Theorem 3 *Let $X_1, \ldots, X_n$ be n nonnegative random variables with distributions $F_1, \ldots, F_n$, respectively. If $F_i * F_j \in \mathcal{S}$ for all $1 \le i, j \le n$ and Assumption B holds, then relation (16) holds.*

Note that, due to the fact that the class $\mathcal{S}$ is not closed under convolution (see Leslie (1989)), the condition $F_i * F_j \in \mathcal{S}$ for all $1 \le i \ne j \le n$ in Theorem 3 is necessary. Note also that $F_i * F_i \in \mathcal{S}$ implies $F_i \in \mathcal{S}$; see Embrechts *et al.* (1979). In the proof below, we write $S_{n,k} = S_n - X_k$ and $S^*_{n,k} = S^*_n - X^*_k$ for every $k = 1, \ldots, n$.

Lemma 2 *If $F_k \in \mathcal{L}$ for all $k = 1, \ldots, n$ and Assumption B holds, then there exist positive constants x_0 and d_n such that the inequality*

$$\Pr(S_{n,k} > x \mid X_k = x_k) \le d_n \Pr\left(S^*_{n,k} > x\right) \tag{19}$$

holds for all $k = 1, \ldots, n$, $x > x_0$, and $x_k > x_0$.

Proof of Lemma 2. We proceed the proof by induction in n. For $n = 2$, the statement follows directly from Assumption B. Assume that the statement holds for $n - 1$. To prove it for n, without loss of generality we only show (19) for $k = n$. Clearly,

$$\begin{aligned}
&\Pr\left(S_{n-1} > x \mid X_n = x_n\right) \\
=& \Pr\left(S_{n-1} > x, \min_{1\le i\le n-1} X_i \le x_0 \,\middle|\, X_n = x_n\right) + \Pr\left(S_{n-1} > x, \min_{1\le i\le n-1} X_i > x_0 \,\middle|\, X_n = x_n\right) \\
=& J_1(x) + J_2(x).
\end{aligned}$$

By our inductive assumption, it holds for all $x_n > x_0$ and all large x that

$$\begin{aligned}
J_1(x) &\le \sum_{i=1}^{n-1} \Pr\left(S_{n-1,i} > x - x_0 \mid X_n = x_n\right) \\
&\le d_{n-1} \sum_{i=1}^{n-1} \Pr\left(S^*_{n-1,i} > x - x_0\right) \\
&\lesssim d_{n-1}(n-1)\Pr\left(S^*_{n-1} > x\right),
\end{aligned}$$

where at the last step we used Lemma 4.2 of Ng *et al.* (2002), which says that the convolution of long-tailed distributions is still long tailed. For $J_2(x)$, because of Assumption B, by conditioning also on the random variables $X_2, \ldots, X_{n-1}$, we have

$$J_2(x) \le c\Pr\left(X_1^* + X_2 + \cdots + X_{n-1} > x, \min_{2\le i\le n-1} X_i > x_0 \,\middle|\, X_n = x_n\right).$$

Repeating this procedure by conditioning on $X_1^*, X_3, \ldots, X_{n-1}$, we further have

$$J_2(x) \le c^2\Pr\left(X_1^* + X_2^* + X_3 + \cdots + X_{n-1} > x, \min_{3\le i\le n-1} X_i > x_0 \,\middle|\, X_n = x_n\right).$$

In this way, we eventually obtain that

$$J_2(x) \le c^{n-1}\Pr\left(S^*_{n-1} > x\right).$$

From these estimates we conclude that relation (19) holds with $k = n$. □

Lemma 3 *Let $F_i * F_j \in \mathcal{S}$ for all $1 \le i, j \le n$. Then, for every function $a(\cdot) : [0,\infty) \to [0,\infty)$ with $a(x) \to \infty$ and for every $1 \le j \le n$,*

$$\Pr\left(S_n^* > x, a(x) < X_j^* \le x\right) = o(1)\sum_{k=1}^{n} \overline{F_k}(x). \tag{20}$$

Proof of Lemma 3. Note that

$$\begin{aligned}
&\Pr\left(S_n^* > x, a(x) < X_j^* \le x\right) \\
\le& \int_0^x \Pr\left(x - y < S^*_{n,j} \le x\right) F_j(\mathrm{d}y) + \Pr\left(S^*_{n,j} > x\right)\overline{F_j}(a(x)) \\
=& \Pr\left(S_n^* > x\right) - \Pr\left(S^*_{n,j} \vee X_j^* > x\right) + \Pr\left(S^*_{n,j} > x\right)\overline{F_j}(a(x)).
\end{aligned}$$

Therefore, relation (20) follows by using Theorem 3 of Geluk and De Vries (2006), which shows that under the current conditions, the distribution of $S^*_{n,j}$ is subexponential and its tail is asymptotic to the sum of the tails of all individual summands. $\square$

Proof of Theorem 3. Note that the condition $F_k \in \mathcal{L}$ for all $k = 1, \dots, n$ suffices for the proof of (18). We only need to establish the asymptotic relation

$$\Pr(S_n > x) \lesssim \sum_{k=1}^{n} \overline{F_k}(x). \tag{21}$$

As before, choose a function $a(\cdot) : [0, \infty) \to [0, \infty)$ such that the items in (4) hold for all $F_1, \dots, F_n$. For all large x,

$$\begin{aligned}
\Pr(S_n > x) &\le \Pr\left(\max_{1\le k\le n} X_k > x - a(x)\right) + \Pr\left(S_n > x, \max_{1\le k\le n} X_k \le x - a(x)\right) \\
&\le \sum_{k=1}^{n} \Pr(X_k > x - a(x)) + \Pr\left(S_n > x, a(x) < \max_{1\le k\le n} X_k \le x - a(x)\right) \\
&\lesssim \sum_{k=1}^{n} \overline{F_k}(x) + \sum_{k=1}^{n} \Pr(S_n > x, a(x) < X_k \le x - a(x)) \\
&= \sum_{k=1}^{n} \overline{F_k}(x) + \sum_{k=1}^{n} \int_{a(x)}^{x-a(x)} \Pr(S_{n,k} > x - y | X_k = y) F_k(\mathrm{d}y).
\end{aligned}$$

Using Lemmas 2 and 3, the last term above is bounded by

$$\begin{aligned}
d_n \sum_{k=1}^{n} \int_{a(x)}^{x-a(x)} \Pr\left(S^*_{n,k} > x - y\right) F_k(\mathrm{d}y) &= d_n \sum_{k=1}^{n} \Pr\left(S^*_n > x, a(x) < X^*_k < x - a(x)\right) \\
&= o(1) \sum_{k=1}^{n} \overline{F_k}(x).
\end{aligned}$$

This proves relation (21). $\square$

Recently, Asmussen and Rojas-Nandayapa (2008) studied the tail probability $\Pr(S_n > x)$ for a special case with dependent lognormal marginals, i.e. $X_k = e^{Y_k}$ for $k = 1, \dots, n$ with $(Y_1, \dots, Y_n)^\top$ following a multivariate normal distribution with mean vector $(\mu_1, \dots, \mu_n)^\top$ and covariance matrix $(\sigma_{ij})_{n\times n}$. Note that in this case each X_k has a tail

$$\overline{F_k}(x) = \overline{F}(x; \mu_k, \sigma_{kk}) \sim \frac{\sqrt{\sigma_{kk}}}{\sqrt{2\pi}\log x} \exp\left\{-\frac{(\log x - \mu_k)^2}{2\sigma_{kk}}\right\}.$$

Their result is

$$\Pr(S_n > x) \sim m_n \overline{F}(x; \mu, \sigma^2), \tag{22}$$

where

$$\sigma^2 = \max_{1\le k\le n} \sigma_{kk}, \quad \mu = \max_{k:\sigma_{kk}=\sigma^2} \mu_k, \quad m_n = \#\{k : \sigma_{kk} = \sigma^2, \mu_k = \mu\}.$$

It is easy to see that the right-hand side of (22) is asymptotically equal to $\sum_{k=1}^{n} \overline{F_k}(x)$.

Closely related discussions can also be found in Albrecher *et al.* (2006), Foss and Richards (2008), and Gao *et al.* (2009).

These results demonstrate that the heavy-tailed marginal distributions eliminate the strength of the dependence of the summands. Heuristically, the more heavy-tailed the marginal distributions are the more robust the asymptotic relation (16) is with respect to the underlying dependence.

A more interesting question would be how to capture the strength of the dependence of the summands when studying the tail probability of the sum. So far there are some discussions in the literature; see Wüthrich (2003), Alink *et al.* (2004), Resnick (1987, 2004), Barbe *et al.* (2006), Malevergne and Sornette (2006), among others. Most of these works use copulas to describe the dependence of the random variables and derive an asymptotic formula which is usually not transparent.

4 Tail Behavior of Products

This section is based the recent work of Jiang and Tang (2009). Let X and Y be two nonnegative random variables with distributions F and G, respectively, and let H be the distribution of the product

$$Z = XY. \tag{23}$$

To avoid triviality, we assume that X and Y are not degenerate at 0. The product in (23) is one of basic elements in stochastic modelling. To describe its tail behavior is usually the core of the study of the tail behavior of quantities containing products of random variables. Cline and Samorodnitsky (1994) explained motivations of this study in infinite variance regression, infinite variance time series, and infinitely divisible stochastic processes.

There are several papers studying the subexponentiality of H for the case that X and Y are independent. Breiman (1965) proved that if $F \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$ and $\mathrm{E}Y^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$ then $\overline{H}(x) \sim \mathrm{E}Y^{\alpha}\overline{F}(x)$. Actually, the proof is a simple application of the dominated convergence theorem guaranteed by relation (6) and the definition of the class $\mathcal{R}$. Several refined versions of Breiman's theorem are given by Denisov and Zwart (2007). Embrechts and Goldie (1980) proved that if $F \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$ and either $\overline{G}(x) = o(\overline{F}(x))$ or $G \in \mathcal{R}_{-\alpha}$ then $H \in \mathcal{R}_{-\alpha}$. Cline and Samorodnitsky (1994) extended the scope to the class $\mathcal{S}$. Further extensions can be found in Tang (2006b, 2008).

In most practical situations, however, the random variables X and Y have to be dependent. For example, let X be the amount transferred by an investor from a bond to a stock at the beginning of a time period and let P_t, $t \geq 0$, be the price process of the stock. With $Y = P_1$, the product Z defined in (23) represents the accumulated value of this transaction at the end of the period. If the investor believes that an increase of the stock price is more likely then he will be willing to increase the transferred amount. In this example, X and Y seem to be positively dependent.

For another example, let $\delta > 0$ be a constant force of compound interest. Suppose that a claim of size X comes at a random time τ. With $Y = \mathrm{e}^{-\delta\tau}$, the present value of this claim is given by Z defined in (23). In this situation, it is more reasonable to assume that the claim size X and its waiting time τ are positively dependent, hence that X and Y are negatively dependent.

To our surprise, we have seen no discussion in the literature focusing on the product of dependent random variables. Available methodologies developed in the references cited above can hardly be used to deal with dependent cases. An obvious reason is that the relation

$$\overline{H}(x) = \int_0^\infty \overline{F}(x/y)G(\mathrm{d}y), \qquad x \geq 0,$$

which is the starting point of those works, does not hold any more.

We assume that X and Y follow a generalized Farlie-Gumbel-Morgenstern (FGM) distribution of the form

$$\Pi_\theta\left(x,y\right) = F(x)G(y)\left[1 + \theta A\left(F(x)\right)B\left(G(y)\right)\right], \tag{24}$$

where the parameter θ, taking values from $[-1,1]$, and the kernels $A(\cdot)$ and $B(\cdot)$, with $A(1) = B(1) = 0$, are suitably chosen such that $\Pi_\theta(\cdot,\cdot)$ is a proper bivariate distribution. Clearly, F and G are the marginal distributions of $\Pi_\theta(\cdot,\cdot)$. The flexibility in choosing the parameter and kernels provides an easy manner to construct bivariate distributions with a variety of dependence structures. This model was originally introduced by Morgenstern (1956) for $A\left(u\right) = B\left(u\right) = 1-u$ and investigated by Gumbel (1960) for exponential marginal distributions. The subsequent generalization to the current form (24) is due to Farlie (1960). For more recent discussions on generalized FGM distributions, the reader is referred to Huang and Kotz (1999), Bairamov *et al.* (2001), Amblard and Girard (2002), and Rodríguez-Lallena and Úbeda-Flores (2004), among others.

We are interested in the question how to capture the impact of the dependence of X and Y in this model on the tail behavior of their product Z. We shall derive an explicit asymptotic formula for the tail probability of Z. In comparison to the asymptotic formula for the independence case, ours contains an extra factor representing the impact of the dependence of X and Y.

By Theorem 2.3 of Rodríguez-Lallena and Úbeda-Flores (2004) (see also discussions of Amblard and Girard (2002)), for $\{\Pi_\theta(\cdot,\cdot) : \theta \in [-1,1]\}$ to be a family of proper bivariate distributions, it is sufficient that the following conditions hold simultaneously:

(b1) $uA\left(u\right)$ and $uB\left(u\right)$ are absolutely continuous on $[0,1]$,

(b2) $\left|\mathrm{d}\left(uA\left(u\right)\right)/\mathrm{d}u\right| \vee \left|\mathrm{d}\left(uB\left(u\right)\right)/\mathrm{d}u\right| \leq 1$ almost everywhere for $u \in [0,1]$, and (b3) $|uA(u)| \vee |uB(u)| \leq u \wedge (1-u)$ for all $u \in [0,1]$.

From now on, we assume that (b1)-(b3) hold. Denote by λ_A and λ_B the left derivatives of $A(u)$ and $B(u)$ at $u=1$. By (b2), we have $|\lambda_A| \vee |\lambda_B| \leq 1$. Introduce

$$F_A(x) \triangleq F(x)\left[1 - A(F(x))\right], \quad G_B(y) \triangleq G(y)\left[1 - B(G(y))\right], \qquad x, y \geq 0, \tag{25}$$

which are two proper distributions supported on $[0,\infty)$ because of (b1)-(b3). Then, introduce two random variables X_A and Y_B distributed by F_A and G_B, respectively. We write $\mu_Y(\alpha) = \mathrm{E}Y^\alpha$ and $\mu_{Y_B}(\alpha) = \mathrm{E}Y_B^\alpha$ for $0 \leq \alpha < \infty$.

Denote by

$$\widehat{y} = \sup\{y | G(y) < 1\} \in (0, \infty]$$

the right endpoint of the support of G with mass $\widehat{p} = \Pr(Y = \widehat{y})$. Note that $\widehat{p} = 0$ when $\widehat{y} = \infty$.

Lemma 4 *Let Y and Y_B be two nonnegative random variables with distributions G and G_B introduced in (25), respectively, where $B(\cdot)$ is a function satisfying (b1)-(b3).*

(i) For every $0 \leq \alpha < \infty$, if $\mu_Y(\alpha) < \infty$ then $\mu_{Y_B}(\alpha) < \infty$;

(ii) If Y has finite moments of all orders $0 \leq \alpha < \infty$ then

$$\lim_{\alpha \nearrow \infty} \frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)} = 1 + (1-\widehat{p})\frac{B(1-\widehat{p})}{\widehat{p}}. \tag{26}$$

Proof of Lemma 4. Regardless of $\widehat{y} < \infty$ or $\widehat{y} = \infty$, we have

$$\lim_{y \nearrow \widehat{y}} \frac{\overline{G_B}(y)}{\overline{G}(y)} = \lim_{y \nearrow \widehat{y}} \left(1 + \frac{G(y)B(G(y))}{\overline{G}(y)}\right) = 1 + (1-\widehat{p})\frac{B(1-\widehat{p})}{\widehat{p}} < \infty. \tag{27}$$

Therefore, $\mu_Y(\alpha) < \infty$ implies $\mu_{Y_B}(\alpha) < \infty$. This proves (i).

To prove (ii), arbitrarily choose y_1 and y_2 such that $0 < y_1 < y_2 < \widehat{y} \leq \infty$. Then, as $\alpha \nearrow \infty$,

$$\frac{\int_0^{y_1} \overline{G}(y) y^{\alpha-1} \mathrm{d}y}{\int_{y_1}^{\widehat{y}} \overline{G}(y) y^{\alpha-1} \mathrm{d}y} \leq \frac{\int_0^{y_1} \overline{G}(y) y^{\alpha-1} \mathrm{d}y}{\int_{y_1}^{y_2} \overline{G}(y) y^{\alpha-1} \mathrm{d}y} \leq \frac{\int_0^{y_1} \overline{G}(y) y^{\alpha-1} \mathrm{d}y}{\overline{G}(y_2) y_1^{\alpha-1}(y_2 - y_1)} \to 0.$$

Similarly,

$$\lim_{\alpha \nearrow \infty} \frac{\int_0^{y_1} \overline{G_B}(y) y^{\alpha-1} \mathrm{d}y}{\int_{y_1}^{\widehat{y}} \overline{G_B}(y) y^{\alpha-1} \mathrm{d}y} = 0.$$

Hence as $\alpha \nearrow \infty$,

$$\frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)} = \frac{\int_0^{\widehat{y}} \overline{G_B}(y) y^{\alpha-1} \mathrm{d}y}{\int_0^{\widehat{y}} \overline{G}(y) y^{\alpha-1} \mathrm{d}y} \sim \frac{\int_{y_1}^{\widehat{y}} \overline{G_B}(y) y^{\alpha-1} \mathrm{d}y}{\int_{y_1}^{\widehat{y}} \overline{G}(y) y^{\alpha-1} \mathrm{d}y}.$$

Moreover, it follows from (27) that

$$\lim_{y_1 \nearrow \widehat{y}} \frac{\int_{y_1}^{\widehat{y}} \overline{G_B}(y) y^{\alpha-1} \mathrm{d}y}{\int_{y_1}^{\widehat{y}} \overline{G}(y) y^{\alpha-1} \mathrm{d}y} = 1 + (1-\widehat{p})\frac{B(1-\widehat{p})}{\widehat{p}}.$$

Finally, by a standard argument on upper and lower bounds, we obtain relation (26). This proves (ii). $\square$

As before, denote by (X^*, Y^*) the independent version of (X, Y) and by H^* the distribution of X^*Y^*. We shall focus on describing the limit

$$\lim_{x\to\infty} \frac{\overline{H}(x)}{\overline{H^*}(x)} = \lim_{x\to\infty} \frac{\Pr(XY > x)}{\Pr(X^*Y^* > x)} \tag{28}$$

under the assumption that $F \in \mathcal{R}_{-\alpha}$ for some $0 \le \alpha \le \infty$. This limit, if it exists, provides a natural measure for the impact of the underlying dependence structure. For this purpose, if $\mu_Y(\alpha) < \infty$ (hence $\mu_{Y_B}(\alpha) < \infty$ by Lemma 4(i)), we define

$$I_\theta(\alpha) = 1 - \theta\lambda_A \left(\frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)} - 1 \right). \tag{29}$$

In this case, we have $0 \le I_\theta(\alpha) \le 2$ since, by (b3),

$$\left| \frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)} - 1 \right| \le \sup_{0\le u\le 1} \left| \frac{uB(u)}{1-u} \right| \le 1, \tag{30}$$

where and throughout, by convention, $B(u)/(1-u) = -\lambda_B$ when $u = 1$. If $\mu_Y(\alpha) < \infty$ for all $0 \le \alpha < \infty$, then by Lemma 4(ii), the function $I_\theta(\alpha)$ converges as $\alpha \nearrow \infty$ with

$$I_\theta(\infty) = 1 - \theta\lambda_A (1 - \widehat{p}) \frac{B(1-\widehat{p})}{\widehat{p}}, \tag{31}$$

which further equals $1 + \theta\lambda_A\lambda_B$ (hence is well defined) for $\widehat{p} = 0$.

Recall that $a(x) \sim b(x)$ stands for $a(x)/b(x) \to 1$. Occasionally, to make some formulas look stringent, we still write $a(x) \sim c_\theta b(x)$ even if the involved coefficient c_θ could equal 0 for certain values of the parameter θ, but its exact meaning should be $a(x)/b(x) \to c_\theta$ in this case. Now we are ready to give the following result, which shows that the quantity $I_\theta(\alpha)$ defined by (29) and (31) is the desired limit in (28):

Theorem 4 *Recall (23) in which X and Y follow a generalized FGM distribution of the form (24) with $F \in \mathcal{R}_{-\alpha}$ for some $0 \le \alpha \le \infty$. Then, the relation*

$$\overline{H}(x) \sim I_\theta(\alpha)\, \overline{H^*}(x) \tag{32}$$

holds with $I_\theta(\alpha)$ given by (29) and (31) under one of the following two sets of conditions:

(i) $0 \le \alpha < \infty$ and $\mu_Y(\alpha + \varepsilon) < \infty$ for some $\varepsilon > 0$;

(ii) $\alpha = \infty$ and $\overline{G}(ux) = o\left(\overline{H^*}(x)\right)$ for all $u > 0$.

For case (i), by Breiman's theorem, relation (32) can be rewritten as

$$\overline{H}(x) \sim ((1 + \theta\lambda_A)\, \mu_Y(\alpha) - \theta\lambda_A \mu_{Y_B}(\alpha))\, \overline{F}(x). \tag{33}$$

For case (ii), the condition on G is automatic if $\widehat{y} < \infty$. Furthermore, Corollary 2.1 of Tang (2006b) shows that this condition on G holds if $\overline{G}(vx) = o\left(\overline{F}(x)\right)$ for some $v > 0$ or $\overline{G}(vx) = o\left(\overline{G}(x)\right)$ for some $v > 1$.

Remark 2 One should be aware that $I_\theta(\alpha)$ could equal 0 for certain rare cases. For example, let $A(u) = 1 - u$, $1/2 \le p_0 = G\{0\} < 1$, and

$$B(u) = \begin{cases} (1-p_0)/p_0, & 0 \le u \le p_0; \\ (1-u)/u, & p_0 < u \le 1. \end{cases}$$

Then, (b1)-(b3) hold but $I_\theta(\alpha) = 0$ for all relevant values of $\alpha \in [0, \infty]$ when $\theta = -1$.

Recall that, in case $I_\theta(\alpha) = 0$, relation (32) should be understood as $\overline{H}(x) = o\left(\overline{H^*}(x)\right)$. We shall not repeat this explanation in the sequel but we should always bear it in mind. Nevertheless, from (29) and (30), for $I_\theta(\alpha) = 0$ it is necessary that $|\theta| = 1$, $|\lambda_A| = 1$, and $\sup_{0\le u\le 1} |uB(u) / (1-u)| = 1$. For all other cases we have $I_\theta(\alpha) > 0$. □

Proof of Theorem 4. In view of (24), we have

$$\Pi_\theta(\mathrm{d}s, \mathrm{d}t) = (1+\theta) F(\mathrm{d}s)G(\mathrm{d}t) - \theta F_A(\mathrm{d}s)G(\mathrm{d}t) - \theta F(\mathrm{d}s)G_B(\mathrm{d}t) + \theta F_A(\mathrm{d}s)G_B(\mathrm{d}t). \quad (34)$$

Therefore,

$$\overline{H}(x) = (1+\theta)\overline{H^*}(x) - \theta \Pr(X_A^* Y^* > x) - \theta \Pr(X^* Y_B^* > x) + \theta \Pr(X_A^* Y_B^* > x). \quad (35)$$

By Breiman's theorem, for case (i) the relation $\overline{G}(ux) = o\left(\overline{H^*}(x)\right)$ still holds for all $u > 0$. Thus, for both cases, by Lemma 3.2 of Tang (2006b) there is a function $a(\cdot) : [0, \infty) \to [0, \infty)$ such that $a(x) = o(x)$ and $\overline{G}(a(x)) = o\left(\overline{H^*}(x)\right)$. For the second term on the right-hand side of (35), with $\widehat{a}(x) = a(x) \wedge \widehat{y}$ we have

$$\Pr(X_A^* Y^* > x) = \int_0^{\widehat{a}(x)} \overline{F_A}(x/t)G(\mathrm{d}t) + o\left(\overline{H^*}(x)\right) = (1 - \lambda_A + o(1))\overline{H^*}(x). \quad (36)$$

Now we deal with the third term on the right-hand side of (35) for the cases $0 \le \alpha < \infty$ and $\alpha = \infty$ separately. When $0 \le \alpha < \infty$, by Lemma 4(i) and Breiman's theorem,

$$\lim_{x\to\infty} \frac{\Pr(X^* Y_B^* > x)}{\overline{H^*}(x)} = \frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)}. \quad (37)$$

When $\alpha = \infty$, by Lemma 3.1(ii) of Tang (2006a), it holds for every $y \in (0, \widehat{y})$ that

$$\frac{\Pr(X^* Y_B^* > x)}{\overline{H^*}(x)} \sim \frac{\int_{x/\widehat{y}}^{x/y} \overline{G_B}(x/s)F(\mathrm{d}s)}{\int_{x/\widehat{y}}^{x/y} \overline{G}(x/s)F(\mathrm{d}s)}.$$

This, together with relation (27), the arbitrariness of y, and a standard argument on upper and lower bounds, gives that

$$\lim_{x\to\infty} \frac{\Pr(X^* Y_B^* > x)}{\overline{H^*}(x)} = 1 + (1 - \widehat{p}) \frac{B(1-\widehat{p})}{\widehat{p}}.$$

Recalling (26), this indicates that relation (37) still holds for $\alpha = \infty$. To deal with the fourth term on the right-hand side of (35), notice that, by (27), $\overline{G_B}(a(x)) = O\left(\overline{G}(a(x))\right) = o\left(\overline{H^*}(x)\right)$. Using the ideas in relations (36) and (37), in turn,

$$\begin{aligned}\frac{\Pr\left(X_A^* Y_B^* > x\right)}{\overline{H^*}(x)} &= (1+o(1)) \frac{\int_0^{\widehat{a}(x)} \overline{F_A}(x/t) G_B(\mathrm{d}t)}{\int_0^{\widehat{a}(x)} \overline{F}(x/t) G_B(\mathrm{d}t)} \frac{\Pr\left(X^* Y_B^* > x\right)}{\overline{H^*}(x)} \\ &\to (1-\lambda_A) \frac{\mu_{Y_B}(\alpha)}{\mu_Y(\alpha)}. \end{aligned} \tag{38}$$

Plugging relations (36)-(38) into relation (35), we obtain relation (32) for both cases of Theorem 4. □

Let $\{(X,Y),(X_n,Y_n), n=1,2,\ldots\}$ be a sequence of i.i.d. random pairs with nonnegative components. Now we turn to study the random difference equation

$$S_0 = 0, \qquad S_n = (S_{n-1} + X_n) Y_n, \qquad n = 1,2,\ldots. \tag{39}$$

Iterating (39) yields that

$$S_n = \sum_{i=1}^n X_i \prod_{j=i}^n Y_j =_D \sum_{i=1}^n X_i \prod_{j=1}^i Y_j, \qquad n=1,2,\ldots, \tag{40}$$

where $=_D$ denotes equality in distribution. Recalling Example 1, the sum on the right-hand side of (40) can be interpreted as discounted aggregate losses by time n of an insurer in a stochastic economic environment.

In the sequel, for each $i = 1,2,\ldots$, we write H_i^* as the distribution of $X_i^* \prod_{j=1}^i Y_j^*$. Note that $H_1^* = H^*$.

Theorem 5 *Consider the random difference equation (39). Assume that (X,Y) follows a generalized FGM distribution of the form (24) with $F \in \mathcal{S} \cap \mathcal{R}_{-\alpha}$ for some $0 \le \alpha \le \infty$. Then for each $n = 1,2,\ldots$, the relation*

$$\Pr(S_n > x) \sim I_\theta(\alpha) \sum_{i=1}^n \overline{H_i^*}(x) \tag{41}$$

holds with $I_\theta(\alpha)$ given by (29) and (31) under one of the following two sets of conditions:

(i) $0 \le \alpha < \infty$ and $\mu_Y(\alpha+\varepsilon) < \infty$ for some $\varepsilon > 0$;

(ii) $\alpha = \infty$ and there are functions $a(\cdot)$ and $b(\cdot) : [0,\infty) \to [0,\infty)$ such that the following items hold simultaneously:

(c1) $b(x) = o(x)$,

(c2) $a(x) = o(b(x))$,

(c3) $\overline{G}(a(x)) = o\left(\overline{H^*}(x)\right)$, and

(c4) $\overline{H^*}(x - b(x)) \sim \overline{H^*}(x)$.

The proof of this theorem is omitted. Similar results but only for the independent case can be found in Tang and Tsitsiashvili (2003, 2004). The following corollary and example show two special cases of Theorem 5(ii):

Corollary 1 *Consider the random difference equation (39) with (X, Y) following a generalized FGM distribution of the form (24). If $F \in \mathcal{S} \cap \mathcal{R}_{-\infty}$ and $\widehat{p} > 0$ (hence $\widehat{y} < \infty$), then for each $n = 1, 2, \ldots,$*

$$\Pr(S_n > x) \sim \left(1 - \theta\lambda_A (1 - \widehat{p}) \frac{B(1 - \widehat{p})}{\widehat{p}}\right) \sum_{i=1}^{n} \widehat{p}^i \overline{F}\left(x/\widehat{y}^i\right). \tag{42}$$

Proof of Corollary 1. By Corollary 2.5 of Cline and Samorodnitsky (1994) and Lemma 2.2 of Tang and Tsitsiashvili (2004), we have $H_i^* \in \mathcal{S} \cap \mathcal{R}_{-\infty}$ for each $i = 1, \ldots, n$. Then, it is easy to see that all conditions of Theorem 5(ii) hold. Hence by the dominated convergence theorem,

$$\overline{H_i^*}(x) \sim \widehat{p}^i \overline{F}\left(x/\widehat{y}^i\right), \qquad i = 1, \ldots, n.$$

Plugging these asymptotic relations to (41) gives (42). □

Example 5 Consider the random difference equation (39) with (X, Y) following a generalized FGM distribution of the form (24). Let F and G be two lognormal distributions with parameters $\left(\mu_1, \sigma_1^2\right)$ and $\left(\mu_2, \sigma_2^2\right)$, respectively. Then, the product X^*Y^* also follows a lognormal distribution with parameters $\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$. From Section 3.3 of Embrechts *et al.* (1997) we know that $\overline{H^*}(x - b(x)) \sim \overline{H^*}(x)$ holds for every $b(x) = o(x/\ln x)$. Choose

$$a(x) = x^{\sigma_2\left(\sigma_1^2+\sigma_2^2\right)^{-0.5}+\varepsilon} \quad \text{and} \quad b(x) = x^{\sigma_2\left(\sigma_1^2+\sigma_2^2\right)^{-0.5}+2\varepsilon}$$

for some $\varepsilon > 0$ such that $\sigma_2\left(\sigma_1^2 + \sigma_2^2\right)^{-0.5} + 2\varepsilon < 1$. Direct calculation shows that

$$\frac{\overline{G}(a(x))}{\overline{H^*}(x)} \sim \left(1 + \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_2}\varepsilon\right) \exp\left\{\frac{(\ln x - \mu_1 - \mu_2)^2}{2\left(\sigma_1^2 + \sigma_2^2\right)} - \frac{(\ln a(x) - \mu_2)^2}{2\sigma_2^2}\right\} = o(1).$$

Thus, all conditions of Theorem 5(ii) hold. For this case, relation (41) becomes

$$\Pr(S_n > x) \sim (1 + \theta\lambda_A\lambda_B) \sum_{i=1}^{n} \overline{\Phi}\left(\frac{\ln x - \mu_1 - i\mu_2}{\sqrt{\sigma_1^2 + i\sigma_2^2}}\right),$$

where Φ denotes the standard normal distribution. □

The following result shows that confining the discussions to the regular variation case we can enhance the asymptotic relation (41) to be uniform for all $n = 1, 2, \ldots$:

Theorem 6 *Consider the random difference equation (39). Assume that (X, Y) follows a generalized FGM distribution of the form (24) with $F \in \mathcal{R}_{-\alpha}$ for some $0 < \alpha < \infty$, $\mu_Y(\alpha) < 1$, and $\mu_Y(\alpha + \varepsilon) < \infty$ for some $\varepsilon > 0$. Then, the relation*

$$\Pr(S_n > x) \sim (1 - \mu_Y^n(\alpha)) \frac{(1 + \theta\lambda_A)\mu_Y(\alpha) - \theta\lambda_A \mu_{Y_B}(\alpha)}{1 - \mu_Y(\alpha)} \overline{F}(x) \tag{43}$$

holds uniformly for all $n = 1, 2, \ldots$, i.e.,

$$\lim_{x\to\infty} \sup_{n\geq 1} \left| \frac{\Pr(S_n > x)}{(1 - \mu_Y^n(\alpha)) \overline{F}(x)} - \frac{(1 + \theta\lambda_A)\mu_Y(\alpha) - \theta\lambda_A \mu_{Y_B}(\alpha)}{1 - \mu_Y(\alpha)} \right| = 0.$$

The proof of this theorem is also omitted.

The uniformity of this result has significant theoretical and practical interests. For example, introduce τ to be a positive integer-valued random variable independent of the sequence $\{(X, Y), (X_i, Y_i), i = 1, 2, \ldots\}$. Then, under the conditions of Theorem 8, a straightforward application of the uniformity of relation (43) gives that

$$\Pr(S_\tau > x) \sim \left(1 - \mathrm{E}e^{\tau \ln \mu_Y(\alpha)}\right) \frac{(1 + \theta\lambda_A)\mu_Y(\alpha) - \theta\lambda_A \mu_{Y_B}(\alpha)}{1 - \mu_Y(\alpha)} \overline{F}(x).$$

Moreover, taking $n = \infty$ into relation (43) yields that

$$\Pr(S_\infty > x) \sim \frac{(1 + \theta\lambda_A)\mu_Y(\alpha) - \theta\lambda_A \mu_{Y_B}(\alpha)}{1 - \mu_Y(\alpha)} \overline{F}(x),$$

which corresponds to a special case of Theorem 1 of Grey (1994).

5 Tail Behavior of Finite Sums with Random Weights

In this section we propose to consider, as a special sum of dependent random variables, the following randomly weighted sum:

$$S_n^{(w)} = W_1 X_1 + \cdots + W_n X_n, \tag{44}$$

where $X_1, \ldots, X_n$, called the primary random variables, are nonnegative, i.i.d., with common distribution $F \in \mathcal{S}$, while $W_1, \ldots, W_n$, called the weights, are nonnegative and random. The dependence of this model comes from the weights and the relationship between the primary random variables and the weights.

There are two extreme cases. The first case is that the weights are also mutually independent and are independent of the primary random variables. Then, $S_n^{(w)}$ is reduced to an independent sum. Using some ideas in Cline and Samorodnitsky (1994) and Tang (2006b, 2008), it is easy to establish the subexponentiality of the products $W_1 X_1, \ldots, W_n X_n$, hence to get

$$\Pr\left(S_n^{(w)} > x\right) \sim \sum_{k=1}^{n} \Pr(W_k X_k > x). \tag{45}$$

The second case is that the weights are mutually 'perfectly' dependent but are still independent of the primary random variables. Here the 'perfect' dependence, which is called comonotonicity by people working in financial risk management (see Dhaene *et al.* (2002a, 2002b)), means that the weights can be expressed as non-decreasing functions of a common random variable, say W. For this case, by taking conditional expectation on the common random variable W, it is not difficult to establish relation (45) again.

The analysis for the two cases above indicates that the asymptotic relation (45) is very stable with respect to the dependence of the weights. The two theorems below demonstrate this again.

Theorem 7 *If $F \in \mathcal{S}$ and the nonnegative random variables $W_1,\ldots, W_n$ are bounded from above, then relation (45) holds.*

An important feature of Theorem 7 is that it does not require any specific information about the dependence structure of $W_1,\ldots, W_n$.

The following lemma is from Tang and Tsitsiashvili (2003a, Proposition 5.1):

Lemma 5 *Let $\{X_k, k = 1,\ldots,n\}$ be n i.i.d. random variables with common distribution $F \in \mathcal{S}$. Then for any fixed $0 < a \le b < \infty$, $k = 1,\ldots,n$, the relation*

$$\Pr\left(\sum_{k=1}^{n} w_k X_k > x\right) \sim \sum_{k=1}^{n} \overline{F}(x/w_k)$$

holds uniformly for $(w_1,\ldots,w_n) \in [a,b]^n$.

Proof of Theorem 7. Without loss of generality, we assume that the random variables $W_1,\ldots, W_n$ are bounded from above by 1.

We derive the lower bound for (45). By Bonferroni's inequality,

$$\begin{aligned}
\Pr\left(S_n^w > x\right) &\ge \Pr\left(\bigcup_{k=1}^{n} (W_k X_k > x)\right)\\
&\ge \sum_{k=1}^{n} \Pr\left(W_k X_k > x\right) - \sum_{1\le i\ne j\le n} \Pr\left(W_i X_i > x, W_j X_j > x\right)\\
&\ge \sum_{k=1}^{n} \Pr\left(W_k X_k > x\right) - \sum_{1\le i\ne j\le n} \Pr\left(W_i X_i > x\right)\Pr\left(X_j > x\right)\\
&\sim \sum_{k=1}^{n} \Pr\left(W_k X_k > x\right).
\end{aligned}$$

To obtain a corresponding upper bound, first we assume that the random weights involved are positive. Let $0 < \varepsilon < 1$ be arbitrarily fixed. For any set $\mathbb{I} \subset \{1,\ldots,n\}$, denote

$$\Delta_{\mathbb{I}}(\varepsilon) = \left(W_i \le \varepsilon \text{ whenever } i \in \mathbb{I},\ W_j > \varepsilon \text{ whenever } j \notin \mathbb{I}\right).$$

We derive

$$\begin{aligned}
&\Pr\left(S_n^w > x\right) \\
&= \Pr\left(\sum_{k=1}^n W_k X_k > x, W_j > \varepsilon \text{ for all } j\right) + \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \Pr\left(\sum_{k=1}^n W_k X_k > x, \Delta_{\mathbb{I}}(\varepsilon)\right) \\
&= I_1(x) + I_2(x). \qquad (46)
\end{aligned}$$

Clearly, by Lemma 5,

$$I_1(x) \sim \sum_{k=1}^n \Pr\left(W_k X_k > x, W_j > \varepsilon \text{ for all } j\right).$$

For $I_2(x)$, by Lemma 5 again we have

$$\begin{aligned}
I_2(x) &\le \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \Pr\left(\varepsilon\sum_{k\in\mathbb{I}} X_k + \sum_{k\notin\mathbb{I}} W_k X_k > x, \Delta_{\mathbb{I}}(\varepsilon)\right) \\
&\sim \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \left(\sum_{k\in\mathbb{I}} \Pr\left(\varepsilon X_k > x, \Delta_{\mathbb{I}}(\varepsilon)\right) + \sum_{k\notin\mathbb{I}} \Pr\left(W_k X_k > x, \Delta_{\mathbb{I}}(\varepsilon)\right)\right) \\
&= I_{21}(x) + I_{22}(x).
\end{aligned}$$

For $I_{21}(x)$, we have

$$\begin{aligned}
I_{21}(x) &= \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) \sum_{k\in\mathbb{I}} \frac{\Pr\left(\varepsilon X_k > x, W_k > \varepsilon\right)}{\Pr\left(W_k > \varepsilon\right)} \\
&\le \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) \sum_{k\in\mathbb{I}} \frac{\Pr\left(W_k X_k > x\right)}{\Pr\left(W_k > \varepsilon\right)} \\
&\le \frac{1}{\Pr\left(\min_{k=1,\dots,n} W_k > \varepsilon\right)} \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) \sum_{k\in\mathbb{I}} \Pr\left(W_k X_k > x\right).
\end{aligned}$$

We turn to $I_{22}(x)$.

$$\begin{aligned}
I_{22}(x) &= \sum_{k=1}^n \sum_{\phi\neq\mathbb{I}\subset\{1,\dots,n\}, k\notin\mathbb{I}} \Pr\left(W_k X_k > x,\ \Delta_{\mathbb{I}}(\varepsilon)\right) \\
&= \sum_{k=1}^n \Pr\left(W_k X_k > x,\ W_k > \varepsilon,\ W_i \le \varepsilon \text{ for some } i = 1,\dots,n\right).
\end{aligned}$$

Substituting $I_{21}(x)$ and $I_{22}(x)$ into $I_2(x)$ and then substituting $I_1(x)$ and $I_2(x)$ into (46), we obtain that

$$
\begin{aligned}
\Pr\left(S_n^w > x\right) &\lesssim \sum_{k=1}^n \Pr\left(W_k X_k > x, W_j > \varepsilon \text{ for all } j\right) \\
&\quad + \frac{1}{\Pr\left(\min_{k=1,\ldots,n} W_k > \varepsilon\right)} \sum_{\phi \neq \mathbb{I} \subset \{1,\ldots,n\}} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) \sum_{k \in \mathbb{I}} \Pr\left(W_k X_k > x\right) \\
&\quad + \sum_{k=1}^n \Pr\left(W_k X_k > x,\ W_k > \varepsilon,\ W_i \leq \varepsilon \text{ for some } i = 1,\ldots,n\right) \\
&\leq \frac{1}{\Pr\left(\min_{k=1,\ldots,n} W_k > \varepsilon\right)} \sum_{\phi \neq \mathbb{I} \subset \{1,\ldots,n\}} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) \sum_{k \in \mathbb{I}} \Pr\left(W_k X_k > x\right) \\
&\quad + \sum_{k=1}^n \Pr\left(W_k X_k > x\right).
\end{aligned}
$$

Since $\lim_{\varepsilon \to 0_+} \Pr\left(\Delta_{\mathbb{I}}(\varepsilon)\right) = 0$ for any $\mathbb{I} \subset \{1,\ldots,n\}$, it follows that

$$\Pr\left(S_n^w > x\right) \lesssim \sum_{k=1}^n \Pr\left(W_k X_k > x\right).$$

Now we consider the general case where the random weights may take value 0 with positive probability. For an arbitrary set $\mathbb{J} \subset \{1,\ldots,n\}$, denote

$$D_{\mathbb{J}} = \left(W_i = 0 \text{ whenever } i \in \mathbb{J},\, W_j > 0 \text{ whenever } j \notin \mathbb{J}\right).$$

Then,

$$
\begin{aligned}
\Pr\left(S_n^w > x\right) &= \sum_{\mathbb{J} \subsetneq \{1,\ldots,n\}} \Pr\left(\sum_{k=1}^n W_k X_k > x, D_{\mathbb{J}}\right) \\
&= \sum_{\mathbb{J} \subsetneq \{1,\ldots,n\}} \Pr\left(\sum_{k \notin \mathbb{J}} W_k X_k > x, D_{\mathbb{J}}\right) \\
&\sim \sum_{\mathbb{J} \subsetneq \{1,\ldots,n\}} \sum_{k \notin \mathbb{J}} \Pr\left(W_k X_k > x, D_{\mathbb{J}}\right) \\
&= \sum_{k=1}^n \sum_{\mathbb{J} \subset \{1,\ldots,n\}, k \notin \mathbb{J}} \Pr\left(W_k X_k > x, D_{\mathbb{J}}\right) \\
&= \sum_{k=1}^n \Pr\left(W_k X_k > x\right).
\end{aligned}
$$

This ends the proof of Theorem 7. □

Theorem 8 *If $F \in \mathcal{A}$ and there is a positive function $a(\cdot)$ with $a(x) = o(x)$ and $a(x) \to \infty$ such that for each $k = 1,\ldots,n$,*

$$\Pr(W_k > a(x)) = o(\Pr(W_k X > x)),$$

then relation (45) holds.

Lemma 6 *Let $\{X_k, k = 1, \ldots, n\}$ be n i.i.d. nonnegative random variables with common distribution $F \in \mathcal{A}$. Then,*

$$\lim_{x\to\infty} \sup_{0<r_1,\ldots,r_n\le 1} \frac{\Pr(r_1X_1 + \cdots + r_nX_n > x)}{\Pr(r_1X_1 > x) + \cdots + \Pr(r_nX_n > x)} \le 1. \tag{47}$$

Proof of Lemma 6. Without loss of generality, we assume $r_1 = 1$. Similarly as in the proof of Theorem 7, for any $\varepsilon > 0$ and $\mathbb{I} \subset \{2, \ldots, n\}$, write

$$\Delta_{\mathbb{I}} = (r_i \le \varepsilon \text{ whenever } i \in \mathbb{I} \text{ and } r_j > \varepsilon \text{ whenever } j \notin \mathbb{I}).$$

Clearly,

$$\lim_{x\to\infty} \sup_{0<r_2,\ldots,r_n\le 1} = \lim_{x\to\infty} \max_{\mathbb{I}} \sup_{\Delta_{\mathbb{I}}} = \max_{\mathbb{I}} \lim_{x\to\infty} \sup_{\Delta_{\mathbb{I}}}.$$

Applying Lemma 5, we obtain

$$\limsup_{x\to\infty} \sup_{\varepsilon\le r_2,\ldots,r_n\le 1} \frac{\Pr(X_1 + r_2X_2 + \cdots + r_nX_n > x)}{\Pr(X_1 > x) + \Pr(r_2X_2 > x) + \cdots + \Pr(r_nX_n > x)} = 1.$$

For a nonempty set $\mathbb{I} \subset \{2, \ldots, n\}$, applying inequality (6), we have for some $0 < p < J_F^-$ and $C_1 > 0$,

$$\begin{aligned}
\limsup_{x\to\infty} \sup_{\Delta_{\mathbb{I}}} &\le \limsup_{x\to\infty} \frac{\Pr\left(X_1 + \varepsilon \sum_{i\in\mathbb{I}} X_i + \sum_{j\notin\mathbb{I}} X_j > x\right)}{\Pr(X_1 > x) + \sum_{j\notin\mathbb{I}} \Pr(X_j > x)} \\
&= \limsup_{x\to\infty} \frac{\Pr(X_1 > x) + \sum_{i\in\mathbb{I}} \Pr(\varepsilon X_i > x) + \sum_{j\notin\mathbb{I}} \Pr(X_j > x)}{\Pr(X_1 > x) + \sum_{j\notin\mathbb{I}} \Pr(X_j > x)} \\
&\le 1 + C_1\varepsilon^p \|\mathbb{I}\| \\
&\le 1 + C_1\varepsilon^p n.
\end{aligned}$$

This proves that

$$\lim_{x\to\infty} \sup_{0<r_2,\ldots,r_n\le 1} \le 1 + C_1\varepsilon^p n.$$

Hence, (47) follows. □

Proof of Theorem 8. We derive the lower bound for (45). By Bonferroni's inequality,

$$\begin{aligned}
\Pr(S_n^w > x) &\ge \Pr\left(\bigcup_{k=1}^n (W_kX_k > x)\right) \\
&\ge \sum_{k=1}^n \Pr(W_kX_k > x) - \sum_{1\le i\ne j\le n} \Pr(W_iX_i > x, W_jX_j > x).
\end{aligned}$$

Since for any $i \neq j$,

$$\begin{aligned}
\Pr(W_i X_i > x, W_j X_j > x) &\leq \Pr(W_i X_i > x, W_j X_j > x, W_j \leq a(x)) + \Pr(W_j > a(x)) \\
&\leq \Pr(W_i X_i > x) \Pr\left(X_j > \frac{x}{a(x)}\right) + \Pr(W_j > a(x)) \\
&= o\left(\Pr(W_i X_i > x) + \Pr(W_j X_j > x)\right).
\end{aligned}$$

This proves that

$$\Pr(S_n^w > x) \gtrsim \sum_{k=1}^{n} \Pr(W_k X_k > x).$$

Then we aim to obtain a corresponding upper bound. First we assume that the random weights involved are positive. We derive

$$\begin{aligned}
\Pr(S_n^w > x) &\leq \Pr\left(\sum_{k=1}^{n} W_k X_k > x\right) \\
&\leq \Pr\left(\sum_{k=1}^{n} W_k X_k > x, W_k \leq a(x)\right) + \sum_{k=1}^{n} \Pr(W_k > a(x)) \\
&= \Pr\left(\sum_{k=1}^{n} \frac{W_k}{W_{(n)}} X_k > \frac{x}{W_{(n)}}, W_{(n)} \leq a(x)\right) + o\left(\sum_{k=1}^{n} \Pr(W_k X_k > x)\right),
\end{aligned}$$

where $W_{(n)} = \max\{W_k; k = 1, \ldots, n\}$. Applying Lemma 6 we have

$$\begin{aligned}
\Pr\left(\sum_{k=1}^{n} \frac{W_k}{W_{(n)}} X_k > \frac{x}{W_{(n)}}, W_{(n)} \leq a(x)\right) &\lesssim \sum_{k=1}^{n} \Pr\left(\frac{W_k}{W_{(n)}} X_k > \frac{x}{W_{(n)}}, W_{(n)} \leq a(x)\right) \\
&\leq \sum_{k=1}^{n} \Pr(W_k X_k > x).
\end{aligned}$$

Finally, for the general case where the random weights may take value 0 with positive probability, the extension can be done by following the last step of the proof of Theorem 7. $\square$

All these results demonstrate that the heavy tails of the primary random variables eliminate the strength of the dependence of the weights.

6 Tail Behavior of Infinite Sums with Random Weights

A challenging question is how to extend the results of the previous section to the infinite sum

$$S^{(w)} = \sum_{k=1}^{\infty} W_k X_k, \tag{48}$$

where the primary random variables X_1, X_2, $\ldots$ are nonnegative, i.i.d., with common distribution $F \in \mathcal{S}$, while the weights W_1, W_2, $\ldots$ are nonnegative and random/nonrandom.

For the case that the weights are nonrandom, the heavy-tail behavior of this sum has been studied by many people including Rootzén (1986), Davis and Resnick (1988), Zerner (2002), and Chen *et al.* (2005). For this case, the distribution of $S^{(w)}$ is of interest because the marginal distribution of any stationary linear process can be represented as the distribution of a sum of the form (48). Linear processes, however, are basic in classical time series analysis. For example, every stationary causal ARMA process is linear with weights which decay exponentially fast to zero; see Sections 3.1 and 13.3 of Brockwell and Davis (1991) and Section 7.1 of Embrechts *et al.* (1997).

Let us concentrate on the case of random weights. For the multivariate counterpart in which $F \in \mathcal{R}$ and the sequences $\{X_1, X_2, \ldots\}$ and $\{W_1, W_2, \ldots\}$ are independent, Resnick and Willekens (1991) obtained an important tail asymptotic formula for $S^{(w)}$ under some moment summability conditions on the weights.

The sum $S^{(w)}$ in (48) with random weights is closely related to the stochastic integral

$$Z_t = \int_{0-}^{t} \mathrm{e}^{-R_s} \mathrm{d}P_s, \qquad t \geq 0, \tag{49}$$

where $\{P_t, t \geq 0\}$ and $\{R_t, t \geq 0\}$ are two stochastic processes fulfilling certain requirements so that Z_∞ is well defined. For example, if $\{P_t, t \geq 0\}$ is a compound renewal process with i.i.d. nonnegative innovations X_1, X_2, $\ldots$ which consecutively arrive at the moments $0 < \tau_1 < \tau_2 < \cdots$, then Z_t in (49) is equal to

$$Z_t = \sum_{k=1}^{\infty} X_k \mathrm{e}^{-R_{\tau_k}} 1_{(\tau_k \leq t)}, \qquad t \geq 0,$$

which corresponds to (48) with dependent weights $W_k = \mathrm{e}^{-R_{\tau_k}} 1_{(\tau_k \leq t)}$ for $k = 1, 2, \ldots$. When $\{P_t, t \geq 0\}$ and $\{R_t, t \geq 0\}$ in (49) are two independent Lévy processes, Gjessing and Paulsen (1997) gave a wealth of examples showing the exact distribution or the asymptotic tail probability of Z_∞. Related discussions can also be found in Dufresne (1990), Paulsen (1993, 1997), and Nilsen and Paulsen (1996), among others. However, we notice that none of these references paid special attention to the important case that $\{P_t, t \geq 0\}$ has heavy-tailed jumps.

The sum $S^{(w)}$ in (48) is also closely related to the stochastic difference equation

$$S_n = (X_n + S_{n-1})\, Y_n, \qquad n = 1, 2, \ldots, \tag{50}$$

where (X_n, Y_n), $n = 1, 2, \ldots$, form a sequence of random vectors fulfilling certain conditions so that S_n converges in distribution to some random variable S_∞. Iterating (50) gives $S_n = \sum_{k=1}^{n} X_k \prod_{i=1}^{k} Y_i$, which suggests that the weak limit of (50), if it exists, must be

$$S_\infty =_d \sum_{k=1}^{\infty} X_k \prod_{i=1}^{k} Y_i. \tag{51}$$

This corresponds to (48) with weights $W_k = \prod_{i=1}^{k} Y_i$ for $k = 1, 2, \ldots$.

Assume that (X_n, Y_n), $n = 1, 2, \ldots$, form a sequence of i.i.d. random vectors with generic random vector (X, Y). For this case, it is well known that S_n weakly converges to some random variable S_∞ as $n \to \infty$ provided that $-\infty \le \mathrm{E} \ln Y < 0$ and $\mathrm{E} \ln^+ X < \infty$; see Theorem 1.6 of Vervaat (1979). Using the Cramér condition and some other technical assumptions, Goldie (1991) derived a celebrated formula that

$$\Pr(S_\infty > x) \sim C_+ x^{-\alpha}, \tag{52}$$

where α and C_+ are some positive constants, with α solving $\mathrm{E}|Y|^\alpha = 1$ but C_+ having a rather involved form. See also Kesten (1973, 1974), Grey (1994), and Tang and Tsitsiashvili (2004), among others.

Still assume that (X_n, Y_n), $n = 1, 2, \ldots$, are i.i.d. random vectors. We propose to study the tail behavior of S_∞ in the following steps. Assume that X follows $F \in \mathcal{S}$ and that Y takes value in $[0, 1]$ but is not degenerate at 0 or 1. First, by (51) it holds for every n that

$$S_\infty =_d S_n + \widetilde{S}_\infty \prod_{i=1}^{n} Y_i, \tag{53}$$

where $\widetilde{S}_\infty$ is an independent copy of S_∞. Next, look for an independent and nonnegative random variable X^* such that the quotient $\Pr(X^* > x)/\Pr(X > x)$ tends to a positive constant and that

$$(X + X^*)Y \le_d X^*. \tag{54}$$

For example, let $F \in \mathcal{S} \cap \mathcal{R}_{-\infty}$ and $\hat{p} = \Pr(Y = 1) \in [0, 1)$. Choose some constant $a > 0$ such that $(1 + a)\hat{p} < a$. For some large $b > 0$ such that $a\overline{F}(b) < 1$, construct a distribution F^* such that

$$\overline{F^*}(x) = \begin{cases} 1, & \text{when } x < b, \\ a\overline{F}(x), & \text{when } x \ge b. \end{cases}$$

Then, introduce a random variable X^* distributed by F^* and independent of (X, Y). It is easy to verify that this random variable X^* satisfies the requirements above.

Relation (54) gives the following, in turn:

$(X_1 + X^*)Y_1 \le_d X^*$,

$(X_2 + X^*)Y_2 \le_d X^*$, and

$X_1 Y_1 + X_2 Y_1 Y_2 + X^* Y_1 Y_2 \le_d X^*$.

Hence, $S_1 \le_d X^*$ and $S_2 \le_d X^*$. Continuing the procedure, we see that the inequality $S_n \le_d X^*$ holds for all n. Hence, $\widetilde{S}_\infty =_d S_\infty \le_d X^*$. Substituting this into (53) yields that

$$S_\infty \le_d S_n + X^* \prod_{i=1}^{n} Y_i.$$

This is a key point in the derivation as its right-hand side is a finite weighted sum, which enables us to use Theorem 7 to derive an upper estimate as

$$\Pr\left(S_\infty > x\right) \lesssim \sum_{k=1}^{n} \Pr\left(X_k \prod_{i=1}^{k} Y_i > x\right) + \Pr\left(X^* \prod_{i=1}^{n} Y_i > x\right).$$

Upon some mild technical assumptions on Y, it is possible to show that the last term of the above is asymptotically negligible for all large n. Hence,

$$\Pr\left(S_\infty > x\right) \lesssim \sum_{k=1}^{\infty} \Pr\left(X_k \prod_{i=1}^{k} Y_i > x\right) = \int_{0-}^{\infty} \overline{F}(x\mathrm{e}^t)\mathrm{d}m_t,$$

where m_t denotes the expected number of sums $\sum_{i=1}^{k} - \log Y_i$ falling into the interval $[0, t]$. The corresponding asymptotic lower bound can be established similarly. Therefore,

$$\Pr(S_\infty > x) \sim \int_{0-}^{\infty} \overline{F}(x\mathrm{e}^t)\mathrm{d}m_t. \tag{55}$$

This formula looks totally different from Goldie's (1991) formula (52). Cases for which the renewal function m_t, $t \geq 0$, assumes an explicit form can be found on pages 88 and 148 of Asmussen (2003).

7 A Further Example

The example given in this section will help us understand how the results in the previous sections work in applied fields. This example comes from a newly rising interdisciplinary area of mathematical finance and actuarial science. It describes a stochastic economic environment in which an insurer makes risk-free and/or risky investments. We shall derive an asymptotic formula for the finite-time ruin probability. Such a result is important for the insurer to determine an optimal investment strategy.

Consider an insurance business commencing at 0 with initial wealth $x \geq 0$. The cash flow of premiums less claims is modeled by a compound Poisson process of the form

$$P_t = ct - \sum_{k=1}^{N_t} X_k, \tag{56}$$

where $c \geq 0$ is a constant premium rate, $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. claim sizes with common distribution $F \in \mathcal{S}$, and $\{N_t, t \geq 0\}$, independent of $\{X_1, X_2, \ldots\}$, is a homogeneous Poisson process with intensity $\lambda > 0$ and arrival times $0 < \tau_1 < \tau_2 < \cdots$.

Suppose that the insurer makes risk-free and/or risky investments. The wealth process S_t, $t \geq 0$, starting from $S_0 = x$, evolves according to the stochastic differential equation

$$\mathrm{d}S_t = \mathrm{d}P_t + S_{t-}\mathrm{d}R_t, \tag{57}$$

where $\{R_t, t \geq 0\}$ is a stochastic process describing returns on investments. For explanations of this model, see Paulsen (2002) and references therein.

For simplicity, we assume that $\{R_t, t \geq 0\}$ is a continuous semimartingale with $R_0 = 0$ and that $\{P_t, t \geq 0\}$ and $\{R_t, t \geq 0\}$ are independent. Then, the explicit solution to equation (57) is given by

$$S_t = \mathrm{e}^{\delta_t} \left(x + \int_0^t \mathrm{e}^{-\delta_s} \mathrm{d}P_s \right), \tag{58}$$

where $\mathrm{e}^{\delta_t} = \exp\left\{R_t - \frac{1}{2}[R,R]_t\right\}$ denotes the Doléans-Dade exponential of R_t; see e.g. page 328 of Protter (2005). In terms of this exponential, the interpretation of the process $\{R_t, t \geq 0\}$ is that one dollar invested at time 0 will be worth e^{δ_t} dollars at time t. We shall start with (58) instead of (57), as done by many people including Kalashnikov and Norberg (2002).

Using (56) and (58), we have

$$\mathrm{e}^{-\delta_t} S_t = x + c \int_0^t \mathrm{e}^{-\delta_s} \mathrm{d}s - \sum_{k=1}^{\infty} X_k \mathrm{e}^{-\delta_{\tau_k}} 1_{(\tau_k \leq t)}. \tag{59}$$

We are going to study the finite-time ruin probability defined as

$$\psi(x, T) = \Pr\left(\inf_{0 < t \leq T} S_t < 0 \mid S_0 = x \right), \qquad T > 0.$$

Consider the Black-Scholes type market consisting of a risk-free bond with a constant force of compound interest $\delta > 0$ and a risky stock with a price process e^{L_t}, $t \geq 0$, where $\{L_t, t \geq 0\}$ is a general stochastic process commencing at 0. We need to assume that $\delta_t \geq -c$ for some $c = c_T > 0$ and all $0 < t \leq T$. This assumption is satisfied if the insurer periodically (e.g. daily, not continuously) rebalances his portfolio so as to invest a constant fraction $\pi \in [0, 1)$ of his wealth in the stock and keep the remaining wealth in the bond. Actually, for this investment strategy, the overall accumulation factor over one period is

$$\mathrm{e}^{\delta_1} = (1 - \pi)\mathrm{e}^{\delta} + \pi \mathrm{e}^{L_1}.$$

Denote by $[t]$ the integer part and by $\{t\}$ the decimal part of t. Then, the overall accumulation factor from time 0 to time t is

$$\mathrm{e}^{\delta_t} = \left[\prod_{k=1}^{[t]} \left((1-\pi)\mathrm{e}^{\delta} + \pi \mathrm{e}^{L_k - L_{k-1}} \right) \right] \left((1-\pi)\mathrm{e}^{\{t\}\delta} + \pi \mathrm{e}^{L_t - L_{[t]}} \right), \tag{60}$$

so that $\delta_t \geq -c$ holds for some $c > 0$ and all $0 < t \leq T$. This periodically constant portfolio seems more reasonable than the constant portfolio proposed by Emmer *et al.* (2001) and Emmer and Klüppelberg (2004).

From (59) we have

$$\Pr\left(\sum_{k=1}^{\infty} X_k \mathrm{e}^{-\delta_{\tau_k}} 1_{(\tau_k \le T)} - c\int_0^T \mathrm{e}^{-\delta_s}\mathrm{d}s > x\right) \le \psi(x,T) \le \Pr\left(\sum_{k=1}^{\infty} X_k \mathrm{e}^{-\delta_{\tau_k}} 1_{(\tau_k \le T)} > x\right). \tag{61}$$

The two terms in the bracket on the left-hand side of (61) are not independent. However, with a mild technical assumption on $\{\delta_t, t \ge 0\}$ we may show that the impact of the term $c\int_0^\infty \mathrm{e}^{-\delta_s}\mathrm{d}s$ is asymptotically negligible. Indeed, if we assume the periodically constant portfolio above, then by (60),

$$c\int_0^T \mathrm{e}^{-\delta_s}\mathrm{d}s \le \frac{c}{1-\pi}\int_0^T \left((1-\pi)\mathrm{e}^{\delta}\right)^{-[s]}\mathrm{d}s = C_T < \infty.$$

Hence, it suffices to estimate the probability on the right-hand side of (61).

It is well known that, for the Poisson process $\{N_t, t \ge 0\}$, the conditional distribution of $(\tau_1, \ldots, \tau_n)$ given $N_T = n$ is identical to the distribution of T multiples of the order statistics of random variables $U_1, \ldots, U_n$ being i.i.d., uniformly distributed on $(0,1)$, and independent of $\{X_1, X_2, \ldots\}$; see e.g. Theorem 2.3.1 of Ross (1983). Therefore,

$$\Pr\left(\sum_{k=1}^{\infty} X_k \mathrm{e}^{-\delta_{\tau_k}} 1_{(\tau_k \le T)} > x\right) = \sum_{n=1}^{\infty} \Pr\left(\sum_{k=1}^{n} X_k \mathrm{e}^{-\delta_{TU_k}} > x\right) \Pr\left(N_T = n\right). \tag{62}$$

This tempts us to employ the idea used in deriving (10). To overcome the difficulty due to the dependence of $\mathrm{e}^{-\delta_{TU_1}}, \ldots, \mathrm{e}^{-\delta_{TU_n}}$, we consider a general weighted sum of the form (44), where X_k, $k = 1, 2, \ldots$, are i.i.d. nonnegative random variables with common distribution $F \in \mathcal{S}$, and W_k, $k = 1, 2, \ldots$, are positive and uniformed bounded random variables independent of X_k, $k = 1, 2, \ldots$. Following the discussions in Section 6, it is not difficult to prove that relation (45) holds. Furthermore, following the original proof of inequality (9) with some extra efforts, we can prove the following: for every $\varepsilon > 0$, there exists some absolute constant $C_\varepsilon > 0$ irrespective to the distributional information of W_1, $W_2, \ldots$ such that the inequality

$$\Pr\left(S_n^{(w)} > x\right) \le C_\varepsilon (1+\varepsilon)^n \sum_{k=1}^{n} \Pr\left(W_k X_k > x\right)$$

holds for all $n = 1, 2, \ldots$ and all $x \ge 0$.

These preliminary results enable us to apply the dominated convergence theorem to prove that the right-hand side of (62) is asymptotic to $\sum_{n=1}^{\infty} n \Pr\left(X_1 \mathrm{e}^{-\delta_{TU_1}} > x\right) \Pr\left(N_T = n\right)$. In conclusion, we have

$$\psi(x,T) \sim \lambda \int_0^T \Pr\left(X_1 \mathrm{e}^{-\delta_t} > x\right)\mathrm{d}t.$$

Acknowledgments. I would like to thank all the participants of this short course for their helpful feedback. In particular, I would like to thank Dimitrios Konstantinides for his warm invitation and excellent organization and thank Jan Dhaene for his stimulating discussions.

References

[1] Albrecher, H.; Asmussen, S.; Kortschak, D. Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes* **9** (2006), 107–130.

[2] Alink, S.; Löwe, M.; Wüthrich, M. V. Diversification of aggregate dependent risks. *Insurance Math. Econom.* **35** (2004), 77–95.

[3] Amblard, C.; Girard, S. Symmetry and dependence properties within a semiparametric family of bivariate copulas. *J. Nonparametr. Stat.* **14** (2002), no. **6**, 715–727.

[4] Asmussen, S. Subexponential asymptotics for stochastic processes: extremal behavior, stationary distributions and first passage probabilities. *Ann. Appl. Probab.* **8** (1998), 354–374.

[5] Asmussen, S. *Ruin Probabilities.* River Edge, NJ, 2000.

[6] Asmussen, S. *Applied Probability and Queues.* Second edition. Springer-Verlag, New York, 2003.

[7] Asmussen, S.; Rojas-Nandayapa, L. Asymptotics of sums of lognormal random variables with Gaussian copula. *Statist. Probab. Lett.* **78** (2008), no. **16**, 2709–2714.

[8] Athreya, K. B.; Ney, P. E. *Branching Processes.* Springer-Verlag, New York, 1972.

[9] Bairamov, I.; Kotz, S.; Bekci, M. New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics. *J. Appl. Stat.* **28** (2001), no. **5**, 521–536.

[10] Barbe, Ph.; Fougères, A.-L.; Genest, C. On the tail behavior of sums of dependent risks. *Astin Bull.* **36** (2006), 361–373.

[11] Barbe, Ph.; McCormick, W. P. Asymptotic expansions for infinite weighted convolutions of rapidly varying subexponential distributions. *Probab. Theory. Relat. Fields* **141** (2008), no. **1-2**, 155–180.

[12] Bingham, N. H.; Goldie, C. M.; Teugels, J. L. *Regular Variation.* Cambridge University Press, Cambridge, 1989.

[13] Bingham, N. H.; Kiesel, R.; Schmidt, R. A semi-parametric approach to risk management. *Quantitative Finance* **3** (2003), no. **6**, 426–441.

[14] Borovkov, A. A. *Stochastic Processes in Queueing Theory.* Springer-Verlag, New York-Berlin, 1976.

[15] Borovkov, A. A. *Asymptotic Methods in Queuing Theory.* John Wiley & Sons, Ltd., Chichester, 1984.

[16] Breiman, L. On some limit theorems similar to the arc-sin law. *Theor. Probability Appl.* **10** (1965), 323–331.

[17] Brockwell, P. J.; Davis, R. A. *Time Series: Theory and Methods.* Second edition. Springer-Verlag, New York, 1991.

[18] Cai, J.; Tang, Q. On max-sum equivalence and convolution closure of heavy-tailed distributions and their applications. *J. Appl. Probab.* **41** (2004), 117–130.

[19] Chen, Y.; Ng, K. W.; Tang, Q. Weighted sums of subexponential random variables and their maxima. *Adv. in Appl. Probab.* **37** (2005), 510–522.

[20] Chen, Y.; Yuen, K. C. Sums of pairwise quasi-asymptotically independent random variables with consistent variation. *Stoch. Models* **25** (2009), **no. 1**, 76–89.

[21] Chistyakov, V. P. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory Prob. Appl.* **9** (1964), 640–648.

[22] Chover, J.; Ney, P.; Wainger, S. Functions of probability measures. *J. Analyse Math.* **26** (1973), 255–302.

[23] Cline, D. B. H. Convolution tails, product tails and domains of attraction. *Probab. Theory Relat. Fields* **72** (1986), 529–557.

[24] Cline, D. B. H.; Samorodnitsky, G. Subexponentiality of the product of independent random variables. *Stochastic Process. Appl.* **49** (1994), 75–98.

[25] Davis, R.; Resnick, S. Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. *Stochastic Process. Appl.* **30** (1988), 41–68.

[26] Denisov, D.; Zwart, B. On a theorem of Breiman and a class of random difference equations. *J. Appl. Probab.* **44** (2007), no. **4**, 1031–1046.

[27] Dhaene, J.; Denuit, M.; Goovaerts, M. J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom.* **31** (2002a), no. **1**, 3–33.

[28] Dhaene, J.; Denuit, M.; Goovaerts, M. J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom.* **31** (2002b), no. **2**, 133–161.

[29] Dhaene, J.; Vanduffel, S.; Goovaerts, M. J.; Kaas, R.; Tang, Q.; Vyncke, D. Risk measures and comonotonicity: a review. *Stoch. Models* **22** (2006), no. **4**, 573–606.

[30] Dufresne, D. The distribution of a perpetuity, with applications to risk theory and pension funding. *Scand. Actuar. J.* **1990**, 39–79.

[31] Embrechts, P.; Goldie, C. M. On closure and factorization properties of subexponential and related distributions. *J. Austral. Math. Soc. Ser. A* **29** (1980), 243–256.

[32] Embrechts, P.; Goldie, C. M.; Veraverbeke, N. Subexponentiality and infinite divisibility. *Z. Wahrsch. Verw. Gebiete* **49** (1979), no. **3**, 335–347.

[33] Embrechts, P.; Klüppelberg, C; Mikosch, T. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin, 1997.

[34] Embrechts, P.; Veraverbeke, N. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1** (1982), no. **1**, 55–72.

[35] Emmer, S.; Klüppelberg, C. Optimal portfolios when stock prices follow an exponential Lévy process. *Finance Stoch.* **8** (2004), 17–44.

[36] Emmer, S.; Klüppelberg, C.; Korn, R. Optimal portfolios with bounded capital at risk. *Math. Finance* **11** (2001), 365–384.

[37] Farlie, D. J. G. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47** (1960), 307–323.

[38] Feller, W. *An Introduction to Probability Theory and Its Applications.* Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[39] Foss, S.; Palmowski, Z.; Zachary, S. The probability of exceeding a high boundary on a random time interval for a heavy-tailed random walk. *Ann. Appl. Probab.* **15** (2005), 1936–1957.

[40] Foss, S.; Richards, A. On sums of conditionally independent subexponential random variables. (2008), to appear.

[41] Foss, S.; Zachary, S. The maximum on a random time interval of a random walk with long-tailed increments and negative drift. *Ann. Appl. Probab.* **13** (2003), 37–53.

[42] Gao, X.; Xu, H.; Ye, D.; Ye, D. Asymptotic behaviors of tail density for sum of correlated lognormal variables. Int. J. Math. Math. Sci. (2009), to appear.

[43] Geluk, J. L.; De Vries, C. G. Weighted sums of subexponential random variables and asymptotic dependence between returns on reinsurance equities. *Insurance Math. Econom.* **38** (2006), no. **1**, 39–56.

[44] Geluk, J.; Tang, Q. Asymptotic tail probabilities of sums of dependent subexponential random variables. J. Theoret. Probab. (2009), to appear.

[45] Gerber, H. U.; Shiu, E. S. W. On the time value of ruin. *N. Am. Actuar. J.* **2** (1998), no. 1, 48–78.

[46] Gjessing, H. K.; Paulsen, J. Present value distributions with applications to ruin theory and stochastic equations. *Stochastic Process. Appl.* **71** (1997), 123–144.

[47] Goldie, C. M. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** (1991), 126–166.

[48] Grey, D. R. Regular variation in the tail behaviour of solutions of random difference equations. *Ann. Appl. Probab.* **4** (1994), 169–183.

[49] Gumbel, E. J. Bivariate exponential distributions. *J. Amer. Statist. Assoc.* **55** (1960), 698–707.

[50] de Haan, L. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tracts, Amsterdam, 1970.

[51] Huang, J. S.; Kotz, S. Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. *Metrika* **49** (1999), no. **2**, 135–145.

[52] Jiang, J.; Tang, Q. The product of two dependent nonnegative random variables with regularly- or rapidly-varying tails. *Adv. in Appl. Probab.* (2009), submitted.

[53] Kalashnikov, V. *Geometric Sums: Bounds for Rare Events with Applications.* Kluwer Academic Publishers Group, Dordrecht, 1997.

[54] Kalashnikov, V.; Norberg, R. Power tailed ruin probabilities in the presence of risky investments. *Stochastic Process. Appl.* **98** (2002), 211–228.

[55] Kesten, H. Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** (1973), 207–248.

[56] Kesten, H. Renewal theory for functionals of a Markov chain with general state space. *Ann. Probability* **2** (1974), 355–386.

[57] Klüppelberg, C. Subexponential distributions and integrated tails. *J. Appl. Probab.* **25** (1988), 132–141.

[58] Ko, B.; Tang, Q. Sums of dependent nonnegative random variables with subexponential tails. *J. Appl. Probab.* **45** (2008), no. **1**, 85–94.

[59] Konstantinides, D.; Tang, Q.; Tsitsiashvili, G. Estimates for the ruin probability in the classical risk model with constant interest force in the presence of heavy tails. *Insurance Math. Econom.* **31** (2002), no. **3**, 447–460.

[60] Kotz, S.; Balakrishnan, N.; Johnson, N. L. *Continuous Multivariate Distributions.* Second edition. Wiley-Interscience, New York, 2000.

[61] Lehmann, E. L. Some concepts of dependence. *Ann. Math. Statist.* **37** (1966), 1137–1153.

[62] Leslie, J. R. On the nonclosure under convolution of the subexponential family. J. Appl. Probab. 26 (1989), no. 1, 58–66.

[63] Malevergne, Y.; Sornette, D. *Extreme Financial Risks. From Dependence to Risk Management.* Springer-Verlag, Berlin, 2006.

[64] McNeil, A. J.; Frey, R.; Embrechts, P. *Quantitative Risk Management.* Princeton University Press, Princeton, NJ, 2005.

[65] Morgenstern, D. Einfache Beispiele zweidimensionaler Verteilungen. (German) *Mitteilungsbl. Math. Statist.* **8** (1956), 234–235.

[66] Nelsen, R. B. *An Introduction to Copulas* (Second edition). Springer, New York, 2006.

[67] Ng, K. W.; Tang, Q.; Yang, H. Maxima of sums of heavy-tailed random variables. Astin Bull. 32 (2002), no. 1, 43–55.

[68] Nilsen, T.; Paulsen, J. On the distribution of a randomly discounted compound Poisson process. *Stochastic Process. Appl.* **61** (1996), 305–310.

[69] Norberg, R. Ruin problems with assets and liabilities of diffusion type. *Stochastic Process. Appl.* **81** (1999), 255–269.

[70] Nyrhinen, H. On the ruin probabilities in a general economic environment. *Stochastic Process. Appl.* **83** (1999), no. **2**, 319–330.

[71] Nyrhinen, H. Finite and infinite time ruin probabilities in a stochastic economic environment. *Stochastic Process. Appl.* **92** (2001), no. **2**, 265–285.

[72] Pakes, A. G. On the tails of waiting-time distributions. *J. Appl. Probability* **12** (1975), 555–564.

[73] Paulsen, J. Risk theory in a stochastic economic environment. *Stochastic Process. Appl.* **46** (1993), 327–361.

[74] Paulsen, J. Present value of some insurance portfolios. *Scand. Actuar. J.* **1997**, 11–37.

[75] Paulsen, J. On Cramér-like asymptotics for risk processes with stochastic return on investments. *Ann. Appl. Probab.* **12** (2002), 1247–1260.

[76] Protter, P. *Stochastic Integration and Differential Equations.* Springer-Verlag, Berlin, 2005.

[77] Rachev, S. T. *Handbook of Heavy Tailed Distributions in Finance.* Elsevier, 2003.

[78] Resnick, S. I. *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York, 1987.

[79] Resnick, S. The extremal dependence measure and asymptotic independence. *Stoch. Models* **20** (2004), 205–227.

[80] Resnick, S. I. *Heavy-tail Phenomena.* Springer, New York, 2007.

[81] Resnick, S. I.; Willekens, E. Moving averages with random coefficients and random coefficient autoregressive models. *Comm. Statist. Stochastic Models* **7** (1991), 511–525.

[82] Rodríguez-Lallena, J. A.; Úbeda-Flores, M. A new class of bivariate copulas. *Statist. Probab. Lett.* **66** (2004), no. **3**, 315–325.

[83] Rogozin, B. A.; Sgibnev, M. S. Banach algebras of measures on the line with given asymptotics of distributions at infinity. *Siberian Math. J.* **40** (1999), no. **3**, 565–576.

[84] Rolski, T.; Schmidli, H.; Schmidt, V.; Teugels, J. *Stochastic Processes for Insurance and Finance.* John Wiley & Sons, Ltd., Chichester, 1999.

[85] Rootzén, H. Extreme value theory for moving average processes. *Ann. Probab.* **14** (1986), 612–652.

[86] Ross, S. M. *Stochastic Processes.* John Wiley & Sons, Inc., New York, 1983.

[87] Tang, Q. On convolution equivalence with applications. *Bernoulli* **12** (2006a), 535–549.

[88] Tang, Q. The subexponentiality of products revisited. *Extremes*, **9** (2006b), 231–241.

[89] Tang, Q. From light tails to heavy tails through multiplier. Extremes, 11 (2008), no. 4, 379–391.

[90] Tang, Q.; Tsitsiashvili, G. Randomly weighted sums of subexponential random variables with application to ruin theory. *Extremes* **6** (2003a), 171–188.

[91] Tang, Q.; Tsitsiashvili, G. Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks. *Stochastic Process. Appl.* **108** (2003b), 299–325.

[92] Tang, Q.; Tsitsiashvili, G. Finite- and infinite-time ruin probabilities in the presence of stochastic returns on investments. *Adv. in Appl. Probab.* **36** (2004), 1278–1299.

[93] Tang, Q.; Wei, L. Asymptotic aspects of the Gerber-Shiu function in the renewal risk model using Wiener-Hopf factorization and convolution equivalence. *Insurance Math. Econom.* (2009), submitted.

[94] Teugels, J. L. The class of subexponential distributions. *Ann. Probability* **3** (1975), 1000–1011.

[95] Vervaat, W. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* **11** (1979), 750–783.

[96] Wüthrich, M. V. Asymptotic value-at-risk estimates for sums of dependent random variables. *Astin Bull.* **33** (2003), 75–92.

[97] Zerner, M. P. W. Integrability of infinite weighted sums of heavy-tailed i.i.d. random variables. *Stochastic Process. Appl.* **99** (2002), 81–94.