ИНФОРМАТИКА И ЕЁ ПРИМЕНЕНИЯ, 2009. Т. 3. Вып. 3. С. 39–45

AN APPROACH TO ACTUARIAL MODELING WITH QUASI-MONTE CARLO: SIMULATION OF RANDOM SUMS DEPENDING ON STOCHASTIC FACTORS*

G. Temnov[1] and S. Kucherenko[2]

Abstract: The problem of estimating the characteristics of a random sum, when the number of summands is also random, is addressed. The considered case includes an additional stochastic factor: although the summed random variables come from a distribution of a known form, the parameters of this distribution are stochastic and can themselves be viewed as random variables (with known distributions). The Quasi-Monte-Carlo (QMC) techniques are used to handle this problem and to analyze its efficiency relative to the regular Monte-Carlo (MC) simulation methods. The typical area of the application of the investigations is actuarial practice which often deals with random sums of financial losses. Besides actuarial applications, the proposed method may be useful in application to certain problems in informatics, related to the aggregation of heavy-tailed data.

Keywords: actuarial modeling; quasi-Monte-Carlo simulation; random sums

1 Introduction

1.1 Setting up the problem

Summation of random number of random variables is a well known problem that has many applications. One of them is the so-called *loss aggregation* problem in insurance. Usually, one needs to compute with sufficient precision the cumulative distribution of the random variables (r.v.'s), having a sense of, for example, financial losses aggregated for some fixed period (normally, one year).

Specifically, we are interested in the distribution of the r.v.

$$S_N = \sum_{k=1}^{N} X_k \tag{1}$$

where N is the number of events within a selected period, generated by a process $N(t)$ of occurrences, usually called *counting process*. The crucial point is that the summed r.v.'s X_k are assumed to be mutually independent and also independent of the counting process $N(t)$.

Another important point is that usually in practical applications r.v.'s X_k having in insurance the sense of single losses can be viewed as **identically distributed** r.v.'s having a distribution $\mathbf{P}(X_k < x) =: F_X(x)$. This assumption allows to apply some particular *deterministic techniques*, making the loss aggregation a relatively simple computational task.

In the current work, the problem of loss aggregation supposing that the assumption of identical distribution of single losses X_k may be violated is addressed. As will be remarked below, deviations from the assumption of identical distribution of losses is quite a natural situation in actuarial modeling, making the application of analytical techniques impossible.

1.2 Aggregation with stochastic parameters

In practical applications, compound distributions should often be modeled with respect to the uncertainty of the parameters of initial distribution (F_X in our terms). This problem can be handled with the help of Bayesian inference. The basic idea of Bayesian modeling for taking into account parameters' uncertainty (see, e.g., [1]) is to consider the vector of the parameters (of initial distribution) as a random vector. Using some *a priori* knowledge about the distribution of this random vector, one can form its *prior distribution* probability dension function (pdf) $\pi(\theta)$. If an additional information comes into play in the form of observations $\mathbf{X}$, the *posterior distribution* pdf with respect to this information can be calculated:

$$\pi_{\Theta\,|\,\mathbf{X}}(\theta\,|\,\mathbf{x}) \propto f_{\mathbf{X}\,|\,\Theta}(\mathbf{x}\,|\,\theta)\pi(\theta)\,. \tag{2}$$

In the absence of a relevant prior information about the prior distribution, $\pi(\theta)$ can be chosen to be a uniform distribution (the case of so called *noninformative priors*).

Often, the expression (2) cannot be used for the direct computation of the posterior distribution and the

*Part of the work was performed within "PRisMa Lab" in Vienna University of Technology.

[1]Edgeworth Centre for Financial Mathematics, University College Cork, Ireland, g.temnov@ucc.ie

[2]CPSE, Imperial College, London, UK, s.kucherenko@ic.ac.uk

stochastic modeling has to be used to produce a pseudo-random sample from $\pi_{\Theta\,|\,\mathbf{X}}(\theta\,|\,\mathbf{x})$.

The sample from posterior distribution can therefore be used to get the sample from the corresponding compound distribution. Specifically, if the $g(z\,|\,\theta)$ is the pdf of the compond distribution (r.v. S in our terms) given a value of the parameter θ, then the corresponding *full predictive distribution* is defined as $h(z\,|\,\mathbf{X}) = \int g(z\,|\,\theta)\pi(\theta\,|\,\mathbf{X})d\theta$ (which is the weighted average with the respect to the distribution of θ as an r.v.).

Obviously, this task is not compatible with *deterministic techniques* (i.e., the ones based on analytical representations of distribution functions), as each time the realization of the predictive distribution is modeled, different values of the parameters $\boldsymbol{\theta}$ and λ should be used. Thus, one needs to use MC simulation methods. The common scheme for modeling of the predictive distribution using MC method could be as follows:

1. Simulate the realization of the severity parameters' vector $\boldsymbol{\theta}$ and frequency λ from their joint distribution $\pi(\boldsymbol{\gamma})$ where $\boldsymbol{\gamma} = (\theta, \lambda)$.
2. Given $\boldsymbol{\theta}$ and λ generate yearly losses, i.e., (i) generate the number N of yearly losses $N \sim \text{Pois}(\lambda)$ and (ii) generate the sample $\{X_j\} = (X_1, \ldots, X_N)$.
3. Given N and $\{X_j\}$ calculate the annual loss $S = \sum\limits_{i=1}^{N} X_i$.
4. Repeat Steps 1–3 K times to get: $\{S\}_{j=1}^{K}$.
5. Estimate α-quanile, $\widehat{Q}_B$, of annual loss ($Q_\alpha \sim \text{sort}(S_j)[\alpha]$).

However, there are at least two aspects in this context that make the use of regular MC simulation techniques very time-demanding. These aspects are (i) actuarial losses are often *heavy tailed* and (ii) one usually needs to estimate a sufficiently high quantile of the aggregate loss distribution. In the case of operational risk measurement, the rules prescript to estimate the α-quantile of aggregate loss called *Value-at-Risk*, i.e., $VaR_\alpha := \sup\limits_x(x :\ F(x) < \alpha)$, at the level $\alpha = 0.999$. Thus, the precision of the modeled predictive distribution should be high enough to obtain a reliable estimate of the upper quantile. In practical applications dealing with operational risk modeling, it is rather usual that not less than $K = 10^6$ repetitions should be made to ensure necessary precision for 0.999-quantile estimate (see, e.g., [1, 2]).

That motivated authors' search for techniques that would reduce the number of claimed repetitions in the modeling scheme above and lead to QMC methods.

2 Methods

2.1 Deterministic techniques

A short comment on deterministic approaches should be made. Some of the deterministic techniques are based on a passage from probability distributions to characteristic functions or probability generating functions (pgf). This approach is applicable to the task (1) in the case of iid summands which will be summarized below.

For a random variable N taking only nonnegative integer values, consider the pgf $P_N(z) = \mathbf{E}[z^N] = \sum\limits_{n=0}^{\infty} \mathbf{P}[N = n]z^n$ which is defined and analytic at least for $|z| \le 1$. Considering the power series expansion of this function $P_N(z) = \sum\limits_{n=0}^{\infty} p_n z^n$, one is able to retrieve the distribution $\mathbf{P}[N = n] = p_n$ for $n \ge 0$ by calculating the coefficients of $P_N(z)$. Denote the pgf of a compound sum of the form (1) by $P_S(z)$ (*considering integer valued loss sizes*) and using the independence assumption and find

$$P_S(z) = \mathbf{E}[z^S] \sum_{k=1}^{\infty} \mathbf{P}[N = k] P_X(z)^k \,.$$

One has the well known representation

$$P_S(z) = P_N(P_X(z)) \tag{3}$$

where $P_N(z)$ is the pgf of the distribution of loss occurrences and $P_X(z)$ corresponds to loss sizes. For Poisson distributed occurrences,

$$P_S(z) = \exp(\lambda(P_X(z) - 1)) \,.$$

Exactly the same representation is valid in terms of characteristic function (chf).

Concerning the calculation of compound distributions **with fixed parameters**, deterministic methods are well developed and include, besides the techniques based of pgf and chf, also, e.g., recursive techniques related to Panjer recursion. For detailed discussion of deterministic techniques in actuarial modeling, see, e.g., [3] and [4]. Deterministic methods are usually more effective than the ones based on MC modeling from the point of precision and speed of calculations. For comparison of the effectiveness of different techniques, see, e. g., [2].

However, when parameters of the distribution of random summands are also random coming from posterior distribution (2) as in the case we deal with, deterministic techniques cannot be used. Indeed, in this case, the pgf $P_X(z)$ in right-hand side of (3) will have no closed form, as it will depend on the distribution of random parameters, which cannot be included into pgf explicitly. Therefore, to handle this problem, one has to turn to MC simulation methods.

2.2 Monte Carlo modeling in heavy tailed cases

Handling heavy tailed distributions, which implies simulation of rare and severe events, has been a challenging task in applied statistics (see, e. g., [5] or [6] for detailed overviews).

Methods usually proposed for solving this problem allow to reduce the computational effort within using the standard MC modeling. According to [5], algorithms involving order statistics and methods using importance sampling are among the most effective techniques for handling random sums. Moreover, various *variance reduction* techniques can be used to increase the efficiency of MC simulation (see, e. g., [7, 8]). However, in certain cases when the resulting distribution depends on random factors, standard variance reduction methods cannot be used, as these techniques rely on explicit representations of distributions used for the modeling. That is indeed the case of our problem, as we need to calculate the random sum of varying volume, and, in addition, the distribution of each variable includes random parameters.

2.3 Quasi-Monte Carlo in risk management

From the previous subsections, motivation for QMC becomes apparent: Clearly, simulation techniques remain a basic tool for modeling compound sums of the form (1) when the distributions involved have stochastic parameters, but dealing with heavy-tailed distributions one needs to handle the variance of simulations in one or another way, and QMC is one of the effective and stable methods to reduce the variance of simulations.

However, QMC methods are not widely used in modeling random sums, as there are certain natural restrictions for using QMC for that particular task. Application of QMC in risk management was studied [9], including of problem of the summation of random variables. In [9], the high-dimensional Sobol' sequences were apply to the problem of risk aggregation for a portfolio of individual losses, when the dimension of the portfolio is fixed. Thus, the problem reduces to the summation of a specified (fixed) number of random variables. It relates the methodology described in [9] to this work, but there are two specific aspects:

(1) in our case, the number of random variables to be summed up is a (discrete) random variable itself; and

(2) we are interested particularly in the summation of heavy tailed r.v.'s.

These aspects motivated the authors to make a separate study in order to find out the efficiency in the QMC scheme in the frame of this problem.

3 Using Quasi-Monte Carlo for the Random Loss Aggregation

In the present section, it will be discussed how Sobol' sequence can be used for the summation of random number of random variables.

Constructing Sobol' sequence. The Sobol' sequence is one of the standard quasi-random sequences and is widely used in QMC applications. The construction of Sobol' sequence will not be described here, referring to [10, 11] and related works for technical details.

3.1 Role of independence

Obviously, the advantage of MC techniques in application to a statistical problem is that it allows to model *independent* random variables. As already mentioned, the assumption of independence plays a major role in the modeling of compound distributions.

Recall that the cumulative probability distribution (cdf) of the compound sum is

$$\begin{aligned} F_S(x) &= \mathbf{P}(S \le x) \\ &= \sum_{k=1}^{\infty} \mathbf{P}[N=k]\mathbf{P}\left(X_1 + \cdots + X_k \le x\right) \\ &= \sum_{k=1}^{\infty} p_k F_X^{*k}(x) \end{aligned} \quad (4)$$

where $F_X^{*k}(x)$ is the k-fold convolution of the pdf with itself, i.e.,

$$F_X^{*k}(x) = \int_0^x F_X^{*k}(x-u)\, dF_X(u)\,, \quad (5)$$

and $F_X^0(x) \equiv 1$ $(x > 0)$.

Once the assumption of independence is dropped, the representation of the compound sum (4) does not reduce to the sum of convolutions (5) any longer. In case of the summation of dependent sequences, in order to calculate the cdf $F_S(s) = \mathbf{P}\left(X_1 + \cdots + X_d \le s\right)$ we would have to deal with the integrals of the form $\int\limits_{\Omega_s := [u_1 + \cdots + u_d \le s]} dF(u_1, \ldots, u_d)$ instead of multiple convolutions.

Independence and multidimensional Sobol' sequences. Clearly, it is already the scheme of the construction of the low-discrepancy sequences that claims the QMC sequences to be dependent. However, if one uses the elements of the sequences from different dimensions, their relation would "imitate" relation between independent random variables.

The issues related to the independence of different dimensions of Sobol' sequences were discussed in [9] where the tests based on rank correlations were used. For brevity, the results of independence tests are not indicated here, but refer to [9] and related literature stating that spatial distribution of multidimensional Sobol' sequences relates to the distribution of independent random variables.

Note that the notions of "randomness" is understood in the studied case certainly not in its usual way. The observed "distributions" of low-discrepancy sequences would not be empirical probability distributions in its general sense. Nevertheless, observing the spatial structure formed by the sequencies, one is able to judge how good is the resulting "imitation" of the independence between values in simulated sequences, due to independence between different dimensions in QMC.

Summarizing the paragraph, note that the right approach for modeling the sums of independent random variables would be to use different (sequential) dimensions for the generation of each of the r.v.'s

3.2 Quasi-Monte Carlo: the modeling set

In operational risk framework, before modeling of the compound sums, one should use historical data to estimate (single-loss) severity and frequency distributions. The historical data of operational risk losses is classified by business lines (BL), and this division is important for the calculation of regulatory capital for OpRisk. Particularly, according to the Basel II recommendations, the Value-at-Risk (VaR) estimators, defining the regulatory capital, should be done for each BL, $L_1, \ldots, L_K$, separately.

In order to estimate the severity and frequency distributions using historical data $\{X_i^{L_j}\}$ in an i.i.d.-case, standard methods such as maximum likelihood estimate (MLE) can be used to estimate the parameters of the distribution $F(x) = \mathbf{P}\left(X_1^{L_j} < x\right)$. In the case of uncertain parameters, the distribution parameters are themselves random variables; hence, the "parameters of the parameters" should be estimated. As soon as the estimates are found, the scheme outlined here can be applied to each BL, using either the regular MC, or QMC techniques.

The following notation for Sobol' sequences is used: **Sobol(i, n, d)=:**$S_n^d(i)$ in a unit hypercube $[0,1]^d$ (here, i is the initial index, n is the length of the sequence, and d is the number of dimensions). Then, if the inverse pdf of an r.v. X, F^{-1}, is known analytically, the realizations of X can be generated via $F^{-1}\left(S_n^d(i)\right)$.

Then, the whole algorithm can be represented in the following way. Suppose for simplicity that there are only one parameter of the severity distribution and one frequency parameter (which is often the case in applications).

I. Simulate an N-length sequence of the severity parameter θ by $I_\theta = F_\theta^{-1}\left(S_n^{d_\theta}(i_1)\right)$ with **fixed** d_j (i. e., using a one-dimensional QMC sequence), and a chosen initial index i_1; an N-length sequence of the frequency parameter λ by $I_\lambda = F_\lambda^{-1}\left(S_n^{d_\lambda}(i_1)\right)$ likewise.

II. Simulate N yearly frequencies using λ_j from I_λ, the quantile function of the Poisson distribution and a new QMC-sequence $\{\Lambda_j\} \sim qpois_{\lambda_j}^{-1}\left(S_n^{d_\Lambda}(i_\Lambda)\right)$ with fixed d_Λ and a chosen index i_Λ.

The choice of i_Λ can be arbitrary in the initial simulation set, but one repeats the simulation cycle to obtain a different estimate of the quantile, then the initial index of next simulation should be chosen such that the QMC sequence used in the previous cycle is not used again.

III. Simulate N sequences of yearly losses $\{X\}_1, \ldots, \{X\}_N$ by

1. $\{X_j^{(1)}\}_{i \le \Lambda_1} = F_X^{-1}\left(S_1^{d_i}\right)$
2. $\{X_j^{(2)}\}_{i \le \Lambda_2} = F_X^{-1}\left(S_2^{d_i}\right)$
...
N. $\{X_j^{(N)}\}_{i \le \Lambda_N} = F_X^{-1}\left(S_N^{d_i}\right)$

Here, d_i changes sequentially in each case (i. e., from 1 to Λ_i for each i).

IV. Loss aggregation.

Consequently, one gets n sums $S_j = \sum\limits_{i=1}^{N_j} X_i^{(j)}$ $(j = 1, \ldots, N)$ which are the realizations of the aggregate loss that we are interested in. Find the quantile:

- put the obtained sample in increasing order to get the order statistics $L_{1:n} \le \cdots \le L_{n:n}$ where $L_{1:n}$ denotes the smallest of the n simulations and $L_{n:n}$ the biggest simulated loss; and
- the element at position $[\alpha n + 1]$ of the ordered sample, where $[\cdot]$ denotes rounding downwards, is the estimator of the quantile (i.e., of VaR) to the level α (e.g., choose $\alpha = 0.999$).

Note that in the above construction, two fundamental properties of the multidimensional QMC sequence have been used: (1) the projection of multidimensional QMC into lower dimensions is again a low-discrepancy sequence; and (2) in each multidimensional sequence, the elements corresponding to different dimensions are independent. That is, taking the elements of Sobol'

sequence sequentally from different dimensions, one keeps the properties of a low-discrepancy sequence, still advancing in having the properties of the QMC sequence.

As a result of the above simulation cycle, a single estimate of the 0.999-quantile of the aggregate loss distribution is obtained. To obtain a different realization of 0.999-quantile, one may repeat the whole cycle, correspondingly changing the initial indices i_1, i_λ, i_θ, etc.

3.3 Results and rates of convergence

Next, overview the obtained results. The *SobolSeq* generator (see www.broda.co.uk, 2009, for the full reference) was used for the QMC simulation, while the general modeling was made in SPLus. We were interested in estimating the 0.999-quantile of the aggregate loss distribution.

To analyze the rate of convergence to the true quantile, the performance of the algorithm was traced for a range of the number of simulation N from 10^5 to $2 \cdot 10^6$, for both QMC and pseudo-random realizations.

Remark 3.1. The benchmark for the true quantile we are referring to can be estimated using, e. g., MC Markov Chain (MCMC) modeling, which is often used as an efficient tool for taking parameter uncertainty into account (see, e. g., [1] or [12]). Using MCMC, it is possible to model the so-called full predictive distribution of aggregate losses with random parameters (though modeling with MCMC is as time demanding as regular MC). The true quantile that can be used as a benchmark to estimate the absolute error of MC and QMC modeling is the quantile of the full predictive distribution.

Note that 10^6 simulations required 525 s of the CPU on a Pentium 2.0 GHz, 1 GB memory, for the whole scheme described above including generation of Sobol' sequence and loss aggregation. The same scheme with the MC requires roughly the same time for the same number of simulations.

The model used for the modeling set was GPD-Poisson, i. e., single losses are supposed to have Generalized Pareto distribution (GPD)

$$G(x) = 1 - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}$$

while the number of yearly losses follows a Poisson process with intensity λ.

Among GPD parameters, the location parameter μ is usually fixed (it plays a role of the threshold), while shape and scale parameters (ξ, σ) are the ones to be estimated. As Maximum likelihood estimation is often used, it is natural to assume that the parameters' vector (ξ, σ) has the bivariate normal distribution and also assume normal distribution for the intensity λ. Furthermore, the following parameters for the distributions of the model parameters were used (the choice of the parameters' values reflects typical values in operational risk framework):

- the threshold $\mu = 7000$;
- vector of mean values for the parameters $(\xi, \sigma) = (1, 12\,000)$, vector of their variance values $(0.18, 1645.0)$, and the covariance value 0.64; and
- mean and variance of the Poisson intensity λ distribution was taken as $(12, 1.7)$.

The plots summarizing the results of the 0.999-quautile modeling illustrating the convergence to the true quantile are presented in Figs. 1 and 2. Figure 1 gives the picture on the log-log scale for the whole range of the simulation numbers, while Fig. 2 concentrates on the segment of the larger values of N on a regular scale where local smooth trend lines show the rate of convergence.

Interpretation of the results indicated on the plots should be made with respect to specific properties of multidimensional Sobol' sequences. Analyzing empirical errors of numerical integration via QMC, [13] shows that actual rates of convergence may differ sufficiently for different types of QMC sequences, depending on the dimension. Main theoretical result on integration errors with QMC known as the *Koksma–Hlawka inequality* states

$$\left| \int_{I^s} f(x)dx - \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right| \le V(f) D_N$$

where D_N is the discrepancy and s is the dimension of the integration domain I^s. The bound on the discrepancy of a random sequence indicates $N^{-1/2}$, suggesting that a sequence with smaller discrepancy could give smaller errors. For low-discrepancy sequences, $D_N \sim N^{-\alpha}$ with α depending on a particular type of sequence. Empirically studying the power α of the rate $N^{-\alpha}$, [13] states quite wide range from 1 to 0.45 depending on the dimension of the function.

In the studied case, the dimensions are "floating," as each of the random sums consists of random number of variables and is therefore generated via a QMC sequence of different (random) dimensions. That fact can at least partially explain the changing rate of convergence to the true quantile observed in Fig. 1. At the same time, as seen from Fig. 2, the rate of convergence for QMC sequences is more stable than the one via pseudo-random numbers simulation, for higher values N corresponding

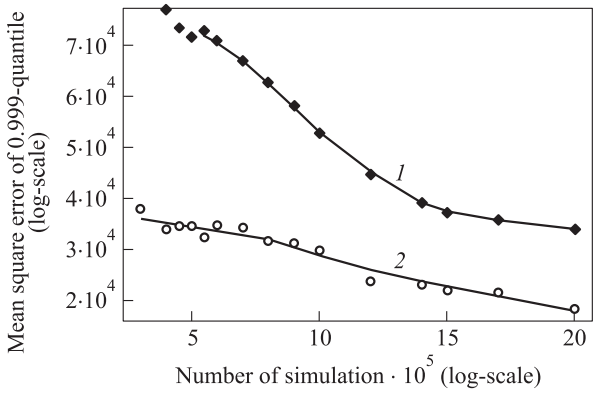

Figure 1 Comparison of the performance of MC (points) with the one obtained by QMC (solid line) for the aggregate–loss distribution of the generalized Pareto

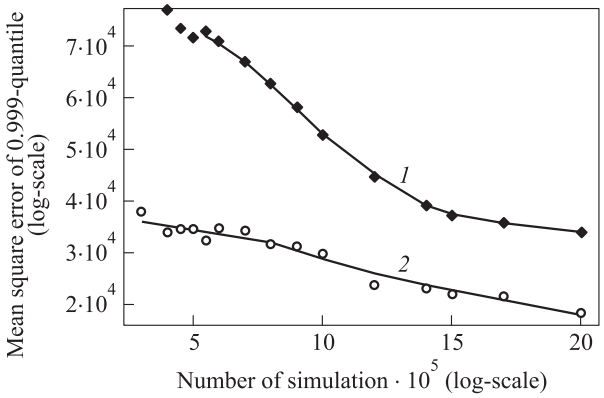

Figure 2 Comparison of the MC precision (*1*) with the precision of QMC (*2*) in terms of the mean square error for the aggregate–loss distribution of the generalized Pareto

to higher precision. As estimated by ordinary linear regression, for the range of higher N starting from $8 \cdot 10^5$, the rate of convergence via QMC sequences is $N^{-0.8}$.

Note that not only the rate of convergence can play a role, but also the absolute value of the error. According to [13], for integration problems with discontinuous in high dimensions, the following result is valid

$$|\,\text{error}\,| = C_s N^{-s/(2s-1)} \tag{6}$$

where C_s changes depending of the particular type of the low discrepancy sequence and the number of dimensions. Thus, the rate of convergence much better than that of a random sequence cannot be expected; however, the precision still can be improved regarding the constant C_s.

In our case, as illustrated by both Figs. 1 and 2, the coefficient C_s is significantly lower than the one associated with the random sequence, allowing to require much lower number of simulations for obtaining the same precision of 0.999-quantile, regarding Sobol' sequence.

References

1. *Shevchenko, P.* 2008. Estimation of operational risk capital charge under parameter uncertainty. *J. Operational Risk* 3(1):51–63.
2. *Temnov, G., and R. Warnung*. 2008. A comparison of loss aggregation methods for operational risk. *J. Operational Risk* 3(1):3–23.
3. *Klugman, S., H. Panjer, and G. Willmot*. 2004. *Loss models. From data to decisions*. Hoboken, NJ, USA: Wiley–Interscience. 2nd ed.
4. *Panjer, H.* 2006. *Operational risk. Modelling analytics*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience. 2nd ed.
5. *Asmussen, S., K. Binswanger, and B. Hojgaard.* 2000. Rare event simulation for heavy tailed distributions. *Bernoulli* 6(2):303–22.

6. *Huang, Z., and P. Shahabuddin*. 2004. A unified approach for finite-dimensional, rare-event Monte Carlo simulation. *2004 Winter Simulation Conference Proceedings.*
7. *Glasserman, P.* 2003. *Monte Carlo methods in financial engineering*. Springer.
8. *Robert, C., and G. Casella* 2004. *Monte Carlo statistical methods*. Springer Texts in Statistics, New York. 2nd ed.
9. *Prakash, S.* 2005. *On the use of high dimensional quasi-random sequences for risk measurement*. Master Thesis, ETH Zurich.
10. *Sobol', I. M.* 1976. Uniformly distributed sequences with additional uniformity properties. *USSR J. Comput. Math. Phys.* 16(5):236–42.
11. *Sobol', I. M.* 1998. On quasi-Monte Carlo integrations. *Math. Computers Simulation* 47:103–12.
12. *Shevchenko, P., and G. Temnov*. 2009 (in press.). Modelling operational risk data reported above a time varying threshold. *J. Operational Risk*.
13. *Morokoff, W., and R. Caflisch*. 1995. Quasi-Monte Carlo integration. *J. Comput. Physics* 122(2):218–30.

ПОДХОД К АКТУАЛЬНОМУ МОДЕЛИРОВАНИЮ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДА КВАЗИ-МОНТЕ-КАРЛО ДЛЯ СЛУЧАЙНЫХ СУММ, ЗАВИСЯЩИХ ОТ СТОХАСТИЧЕСКИХ ФАКТОРОВ

Г. Темнов[1], С. Кучеренко[2]

[1]Корский университет, Ирландия, g.temnov@ucc.ie
[2]Империал Колледж, Лондон, Великобритания, s.kucherenko@ic.ac.uk

Аннотация: Рассматривается задача оценивания характеристик случайной суммы, в которой число слагаемых также случайно. Рассматриваемый случай включает дополнительный случайный фактор: хотя тип распределения слагаемых известен, параметры этого распределения рассматриваются как случайные величины с известным распределением. Рассматриваемая задача решается с помощью метода квази-Монте-Карло. Анализируется эффективность данного подхода по сравнению с обычным методом Монте-Карло. Рассматриваемые методы имеют применение в актуарной практике, а также при решении некоторых задач информатики, связанных с агрегированием данных с тяжелыми хвостами.

Ключевые слова: актуарное моделирование, метод квази-Монте-Карло, случайные суммы