

Modeling the Risk Process in the XploRe Computing Environment

Krzysztof Burnecki and Rafał Weron

Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology,
Wyspiańskiego 27, 50-370 Wrocław, Poland
{burnecki,rweron}@im.pwr.wroc.pl
<http://www.im.pwr.wroc.pl/~hugo>

Abstract. A user friendly approach to modeling the risk process is presented. It utilizes the insurance library of the XploRe computing environment which is accompanied by on-line, hyperlinked and freely downloadable from the web manuals and e-books. The empirical analysis for Danish fire losses for the years 1980-90 is conducted and the best fitting of the risk process to the data is illustrated.

1 Introduction

The simulation of risk processes is vital for insurance companies. It boils down to generating aggregated claims for the calculation of losses that may occur. Since claims arrive at random times, the number of claims up to a given time is typically assumed to be driven by a stochastic process. In its simplest form the risk process utilizes the homogeneous Poisson process as the claim arrival process, however, more general processes like the non-homogeneous Poisson, the mixed Poisson, the Cox (or doubly stochastic Poisson) and the renewal process are also considered.

The risk process $\{R_t\}$ of an insurance company can be approximated by the sum of the initial capital of the company and the so-called premium function (representing income from sold insurance policies) minus the aggregated claim process (expressing liabilities resulting from claims covered by the previously sold insurance policies) [3,8,9,14]. The latter is typically modeled by a sum of random length, governed by the claim arrival point process, where the summands – representing the claim severities – form an independent sequence of positive i.i.d. random variables.

Since parameter estimation and simulation schemes can be tedious a number of computer packages have been written to automate the process. In this paper we want to present a novel solution which takes the form of a library of procedures of the XploRe system combined with on-line, hyperlinked manuals and e-books.

XploRe is a computing environment which offers a combination of classical and modern statistical procedures, in conjunction with sophisticated, interactive graphics. XploRe is also a high level object-oriented programming language. With all the standard features like recursion, local variables, dynamic data

structures, loops, and conditional execution it provides a platform for advanced statistical and econometric analysis, research, as well as teaching [10,11].

The statistical methods of XploRe are provided by various procedures and scripts called quantlets. Quantlets are combined into libraries called quantlibs. Among other these include: finance, econometrics, wavelets, generalized (partial) linear models, time series analysis and filtering, neural networks, non- and semiparametric methods, and teachware. Recent additions to this family comprise the stable distributions and insurance libraries [6]. Proprietary methods can be incorporated into XploRe, enabling the user to easily extend the environment.

One of the most outstanding features of the XploRe environment is the unique combination of computing capabilities and on-line, hyperlinked manuals and books. A variety of electronic statistical and econometric volumes are available from the XploRe web page (www.xplore-stat.de) in the html and pdf file formats. All books contain a large number of quantlets which illustrate the theoretical content. One of the newest additions is the "Statistical Tools for Finance and Insurance" e-book [6], which offers the reader over a hundred methods and procedures related to insurance. Some of these quantlets will be utilized in this paper.

The paper is organized as follows. In section 2 we briefly recall the methods of simulating two possible choices for the claim arrival process, namely the homogeneous Poisson process (HPP) and the non-homogeneous Poisson process (NHPP). In section 3 we discuss a number of claim severities distributions and present the methods for judging the goodness-of-fit. Finally, in section 4 we conduct the empirical analysis for Danish fire losses for the years 1980-90 and illustrate the best fitting risk process for two choices of the claim severities distribution.

2 Claim Arrival Process

In this section we focus on the efficient simulation of the claim arrival point process. Typically this process is simulated via the arrival times $\{T_i\}$, i.e. moments when the i th claim occurs, or the inter-arrival times (or waiting times) $W_i = T_i - T_{i-1}$, i.e. the time periods between successive claims.

2.1 Homogeneous Poisson Process

A continuous-time stochastic process $\{N_t : t \geq 0\}$ is a (homogeneous) Poisson process with intensity (or rate) $\lambda > 0$ if (i) $\{N_t\}$ is a point process, and (ii) the times between events are i.i.d. exponential random variables with intensity λ , i.e. with mean $1/\lambda$. Therefore, successive arrival times T_1, T_2, \dots, T_n of the Poisson process can be generated by a simple algorithm consisting of generating independent exponential random variables with intensity λ and taking a cumulative sum of them [3,15]. In the insurance library of XploRe this procedure is implemented in the `simHPP.xpl` quantlet.

Since the expected value of the homogeneous Poisson process $EN_t = \lambda t$, it is natural to define the premium function as a linear function of the form $c(t) = (1 + \theta)\mu\lambda t$. Here $\mu = EX_k$ is the expected value of the claim size and $\theta > 0$ is the relative safety loading which "guarantees" survival of the insurance company. With such a choice of the risk function we obtain the classical form of the risk process [3,8,9]:

$$R_t = u + (1 + \theta)\mu\lambda t - \sum_{i=1}^{N_t} X_i . \tag{1}$$

The nonnegative constant u represents the initial capital of the company and $\{X_i\}$ is the i.i.d. claim severities sequence.

2.2 Non-homogeneous Poisson Process

In real life situations the homogeneous Poisson process may be too simplistic. In order to obtain a more reasonable description of reality we might want to include the possibility of a variable (eg. increasing) size of the portfolio of insurance contracts or seasonality in the number of claims. For modeling such phenomena the non-homogeneous Poisson process (NHPP) is much better. To distinguish it from the HPP we denote it by M_t . The NHPP can be thought of as a Poisson process with a variable intensity defined by the deterministic intensity (rate) function $\lambda(t)$. Note that the increments of a NHPP do not have to be stationary. In the special case when $\lambda(t)$ takes the constant value λ , the NHPP reduces to the homogeneous Poisson process with intensity λ .

The simulation of a NHPP is slightly more complicated than in the homogeneous case. The first approach, the so-called "integration method", is based on the observation that for a NHPP with rate function $\lambda(t)$ the increment $M_t - M_s$, $0 < s < t$, is distributed as a Poisson random variable with intensity $\bar{\lambda} = \int_s^t \lambda(u)du$ [9]. Hence, the distribution function F_s of the waiting time W_s is given by:

$$\begin{aligned} F_s(t) &= P(W_s \leq t) = 1 - P(W_s > t) = 1 - P(M_{s+t} - M_s = 0) = \\ &= 1 - \exp\left(-\int_s^{s+t} \lambda(u)du\right) = 1 - \exp\left(-\int_0^t \lambda(s+v)dv\right) . \end{aligned} \tag{2}$$

If the function $\lambda(t)$ is such that we can find an explicit formula for the inverse F_s^{-1} then for each s we can generate a random quantity X with the distribution F_s by using the inverse transform method. Otherwise, we have to apply numerical schemes and the algorithm becomes relatively slow.

The second approach, known as the "thinning" or "rejection method", is based on the following observation [2,15]. Suppose that there exists a constant $\bar{\lambda}$ such that $\lambda(t) \leq \bar{\lambda}$ for all t . Let $T_1^*, T_2^*, T_3^*, \dots$ be the successive arrival times of a HPP with intensity $\bar{\lambda}$. If we accept the i th arrival time with probability $\lambda(T_i^*)/\bar{\lambda}$, independently of all other arrivals, then the sequence T_1, T_2, \dots of the accepted arrival times (in ascending order) forms a sequence of the arrival times

of a NHPP with rate function $\lambda(t)$. The resulting algorithm [3] is implemented in the insurance library of XploRe in the `simNHPP.xpl` quantlet.

Since the mean-value function, i.e. the expected value of the process M_t , is given by:

$$EM_t = \int_0^t \lambda(s)ds , \tag{3}$$

in the non-homogeneous case the premium function is typically defined as $c(t) = (1 + \theta)\mu \int_0^t \lambda(s)ds$ and the risk process takes the form:

$$R_t = u + (1 + \theta)\mu \int_0^t \lambda(s)ds - \sum_{i=1}^{M_t} X_i . \tag{4}$$

3 Claim Severities Distribution

The derivation of claim size distributions from the loss data could be considered to be a separate discipline in its own [7]. The objective is to find a distribution function F which fits the observed data in a satisfactory manner. The approach most frequently adopted in insurance is to find a suitable analytic expression which fits the observed data well and which is easy to handle, see e.g. [5].

The claim distributions, especially describing property losses, are usually heavy-tailed. Note, that in the actuarial literature such distributions are defined as having tails heavier than exponential. The lognormal, Pareto, Burr and Weibull distributions are typical candidates considered in applications [13]. Let us briefly recall them.

If the random variable X is normally distributed with mean μ and variance σ^2 , then the distribution of $Y = e^X$ is lognormal. Its distribution function (d.f.) is given by $F_{LN}(x) = \Phi([\ln x - \mu]/\sigma)$, where $\Phi(x)$ is the standard normal d.f., and in the insurance library of XploRe is implemented in the `cdfln.xpl` quantlet. The lognormal law is very useful in modeling of claim severities. It has a thick right tail and fits many situations well.

One of the most frequently used analytic claim size distributions is the Pareto law which is defined by $F_P(x) = 1 - (\lambda/[\lambda + x])^\alpha$ with $\alpha, \lambda > 0$, see the `cdfPareto.xpl` quantlet. The first parameter controls the thickness of the tail: the smaller the α , the heavier the tail. Empirical experience has shown that the Pareto formula is often an appropriate model for the claim size distribution, particularly when exceptionally large claims may occur [4,7]. However, there is a need to find heavy tailed distributions which offer yet greater flexibility. Such flexibility is offered by the Burr distribution, which is just a generalization of the Pareto law. Its distribution function is given by $F_B(x) = 1 - (\lambda/[\lambda + x^\tau])^\alpha$ with all three parameters (α , λ , and τ) being positive real constants, see the `cdfBurr.xpl` quantlet.

Another frequently used analytic claim size distribution is the Weibull distribution which is defined by $F_W(t) = 1 - e^{-\lambda t^\alpha}$ with $\alpha, \lambda > 0$. Observe, that the Weibull distribution is a generalization of the exponential law. In XploRe it is implemented in the `cdfWeibull.xpl` quantlet.

Once the distribution is selected, we must obtain parameter estimates. In what follows we use the moment and maximum likelihood estimation approaches. The next step is to test whether the fit is adequate. This is usually done by comparing the fitted and empirical distribution functions. More precisely, by checking whether values of the fitted distribution function at sample points form a uniform distribution. In the next section we apply the χ^2 , the Kolmogorov-Smirnov (KS), the Cramer-von Mises (CM), and the Anderson-Darling (AD) test statistics [1,5,16]. Generally, the smaller the value of the statistics, the better the fit. These test statistics are implemented in the quantlets `chi2stat.xpl`, `kstat.xpl`, `cmstat.xpl`, and `adstat.xpl`, respectively [6].

4 Empirical Analysis

We conducted empirical studies for Danish fire losses recorded by Copenhagen Re. The data, see Fig. 1, concerns major Danish fire losses in Danish Krone (DKK), occurred between 1980 and 1990 and adjusted for inflation. Only losses in profits connected with the fires were taken into consideration. In order to calibrate the risk process we had to fit both the distribution function F of the incurred losses $\{X_i\}$ and the claim arrival process. First we studied the

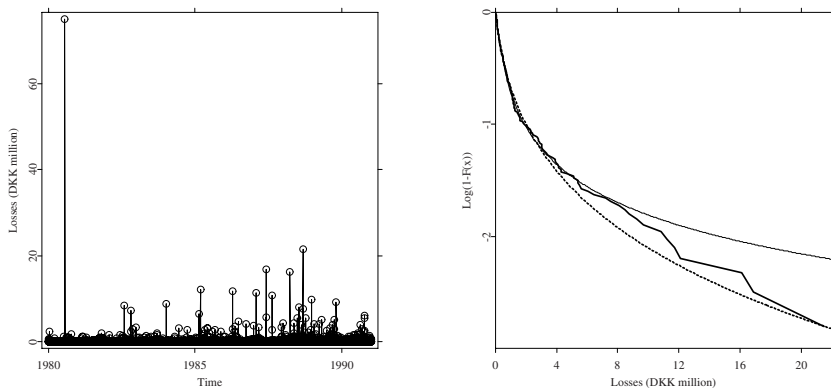


Fig. 1. *Left panel:* Illustration of the major Danish fire losses in millions of Danish Krone (DKK), occurred between January 1, 1980 and December 31, 1990 and adjusted for inflation. *Right panel:* Logarithm of the right tails of the empirical claim sizes distribution function $F(x)$ (thick solid line) together with lognormal (dotted line) and Burr (thin solid line) fits

loss sizes. We fitted lognormal, Pareto, Burr, and Weibull distributions using the `estln.xpl`, `estPareto.xpl`, `estBurr.xpl`, and `estWeibull.xpl` quantlets, respectively. The results of the parameter estimation and test statistics are

Table 1. Parameter estimates and test statistics for the fire loss amounts

d.f.:	Lognormal	Pareto	Burr	Weibull	Gaussian
Para-	$\mu = 12.704$	$\alpha = 2.4189$	$\alpha = 0.8935$	$\alpha = 0.6963$	$\mu = 4.7332e5$
meters:	$\sigma = 1.4271$	$\lambda = 1.0261e6$	$\lambda = 1.1219e7$	$\lambda = 8.9740e-5$	$\sigma = 6.7224e5$
			$\tau = 1.2976$		
χ^2	56.109	73.879	48.493	129.24	592.98
KS	0.0373	0.0397	0.0413	0.0783	0.2433
CM	0.1687	0.2878	0.1438	1.5245	6.3655
AD	1.0533	2.7712	0.8221	10.638	78.410

presented in Table 1. For illustration purposes we also added parameter estimates for the Gaussian distribution. The lognormal distribution with parameters $\mu = 12.7036$ and $\sigma = 1.4271$ and the Burr distribution with $\alpha = 0.8935$, $\lambda = 1.1219 \cdot 10^7$ and $\tau = 1.2976$ produced the best results, see Fig. 1. Hence, we chose them for further analysis. Next, we fitted the claim arrival process.

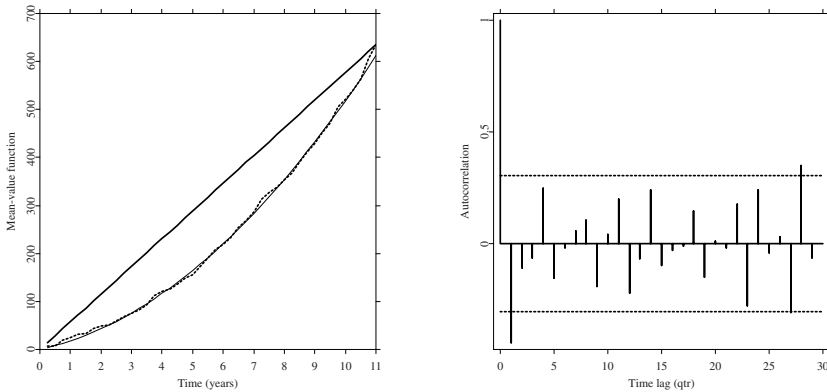


Fig. 2. *Left panel:* The aggregate number of losses of the fire data (dotted line), the HPP (thick solid line), and the NHPP (thin solid line). *Right panel:* Autocorrelation function of the differenced quarterly fire losses data revealing no seasonality or significant dependencies. Horizontal dashed lines represent the 95% confidence intervals for white noise

We started the analysis with the homogeneous Poisson process N_t with intensity λ_1 . Studies of the quarterly numbers of losses and the interoccurrence times of the catastrophes led us to the conclusion that the homogeneous Poisson process

with the annual intensity $\lambda_1 = 57.72$ gave the best fit. However, as we can see in Fig. 2, the fit is not very good suggesting that the HPP is too simplistic and causing us to consider the NHPP.

The data reveals no seasonality but a clear increasing trend can be observed in the number of quarterly losses, see Fig. 2. We tested different exponential and polynomial functional forms, but a simple linear intensity function $\lambda_2(s) = c + ds$ yielded the best fit. Applying a least squares procedure we arrived at the following values of the parameters: $c = 13.97$ and $d = 7.57$. Both choices of the intensity function, λ_1 and $\lambda_2(s)$, are illustrated in Fig. 2, where the accumulated number of fire losses and mean-value functions for all 11 years of data are depicted. The simulation results are presented in Fig. 3. We consider a

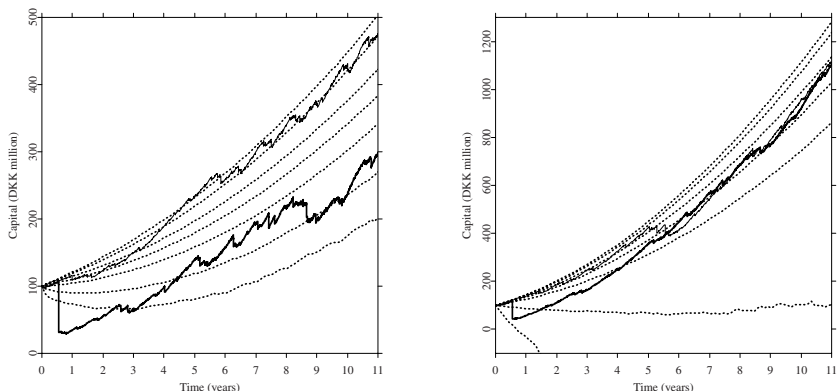


Fig. 3. Simulation results for a non-homogeneous Poisson process with lognormal (*left panel*) and Burr (*right panel*) claim sizes. The thick solid line is the "real" risk process, i.e. a trajectory constructed from the historical arrival times and values of the losses. The thin solid line is a sample trajectory. The dotted lines are the sample 0.01, 0.05, 0.25, 0.50, 0.75, 0.95, 0.99-quantile lines. Clearly, the Burr distribution describes the extreme losses much better

hypothetical scenario where the insurance company insures losses resulting from fire damage. The company's initial capital is assumed to be $u = 100$ million kr and the relative safety loading used is $\theta = 0.5$. We chose two models of the risk process whose application is most justified by the statistical results described above: a non-homogeneous Poisson process with lognormal claim sizes and a non-homogeneous Poisson process with Burr claim sizes.

In both subplots of Fig. 3 the thick solid line is the "real" risk process, i.e. a trajectory constructed from the historical arrival times and values of the losses. The thin solid line is a sample trajectory. The dotted lines are the sample 0.01, 0.05, 0.25, 0.50, 0.75, 0.95, 0.99-quantile lines based on 10000 trajectories of

the risk process. Recall that the function $\hat{x}_p(t)$ is called a sample p -quantile line if for each $t \in [t_0, T]$, $\hat{x}_p(t)$ is the sample p -quantile, i.e. if it satisfies $F_n(x_p-) \leq p \leq F_n(x_p)$, where F_n is the sample distribution function. Quantile lines are a very helpful tool in the analysis of stochastic processes. For example, they can provide a simple justification of the stationarity (or the lack of it) of a process, see [12]. In Fig. 3 they visualize the evolution of the density of the risk process. Clearly, if claim severities are Burr distributed then extreme events are more probable to happen than in the lognormal case, for which the historical trajectory falls even outside the 0.01-quantile line. This suggests that Burr distributed claim sizes are more adequate for modeling the "real" risk process.

Acknowledgements. The authors thankfully acknowledge the support of the State Committee for Scientific Research (KBN) Grant No. PBZ-KBN 016/P03/99.

References

1. D'Agostino, R.B., Stephens, M.A.: Goodness-of-Fit Techniques. Marcel Dekker, New York (1986)
2. Bratley, P., Fox, B.L., Schrage, L.E.: A Guide to Simulation. Springer-Verlag, New York (1987)
3. Burnecki, K., Härdle, W., Weron, R.: An Introduction to Simulation of Risk Processes. In: Teugels, J., Sundt, B. (eds.): Encyclopedia of Actuarial Science. Wiley, Chichester (to appear)
4. Burnecki, K., Kukla, G.: Pricing of Zero-Coupon and Coupon CAT Bonds. *Appl. Math. Model.* **30(3)** (2003) 315–324
5. Burnecki, K., Kukla, G., Weron, R.: Property Insurance Loss Distributions. *Physica A* **287** (2000) 269–278
6. Cizek, P., Härdle, W., Weron, R. (ed.): Statistical Tools for Finance and Insurance. Springer, Heidelberg (2004)
7. Daykin, C.D., Pentikainen, T., Pesonen, M.: Practical Risk Theory for Actuaries. Chapman&Hall, London (1994)
8. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events. Springer, Berlin (1997)
9. Grandell, J.: Aspects of Risk Theory. Springer, New York (1991)
10. Härdle, W., Hlavka, Z., Klinke, S. (ed.): XploRe Application Guide. Springer, Heidelberg (2000)
11. Härdle, W., Klinke, S., Müller, M. (ed.): XploRe Learning Guide. Springer, Heidelberg (2000)
12. Janicki, A., Weron, A.: Simulation and Chaotic Behavior of α -Stable Stochastic Processes. Marcel Dekker, New York (1994)
13. Panjer, H.H., Willmot, G.E.: Insurance Risk Models. Society of Actuaries, Schaumburg (1992)
14. Rolski, T., Schmidli, H., Schmidt, V., Teugels, J.L.: Stochastic Processes for Insurance and Finance. Wiley, Chichester (1999)
15. Ross, S.: Simulation. 3rd edn. Academic Press, San Diego (2001)
16. Stephens, M.A.: EDF Statistics for Goodness-of-Fit and Some Comparisons. *Journal of the American Statistical Association* **69** (1974) 730–737