# Visualization tools for insurance risk processes

Krzysztof Burnecki*
Rafał Weron*

* Hugo Steinhaus Center, Wrocław University of
Technology, Poland

HSC Research Report

# Visualization Tools for Insurance Risk Processes

Krzysztof Burnecki and Rafał Weron

Hugo Steinhaus Center for Stochastic Methods, Institute of Mathematics and Computer Science, Wrocław University of Technology, 50-370 Wrocław, Poland

## 1 Introduction

This chapter develops on risk processes which, perhaps, are most suitable for computer visualization of all insurance objects. At the same time, risk processes are basic instruments for any non-life actuary – they are vital for calculating the amount of loss that an insurance company may incur. They also appear naturally in rating triggered step-up bonds, where the interest rate is bound to random changes of the companies' ratings, and catastrophe bonds, where the coupon payment amounts are conditional on the severity of catastrophic events.

A typical model for insurance risk, the so-called collective risk model, has two main components: one characterizing the frequency (or incidence) of events and another describing the severity (or size or amount) of gain or loss resulting from the occurrence of an event (Klugman, Panjer, and Willmot, 1998; Panjer and Willmot, 1992; Teugels and Sundt, 2004). Both the incidence and severity of claims are generally assumed to be stochastic and independent of each other. Together they form the backbone of a realistic risk process. Consequently, both have to be calibrated to the available historical data. All three visualization techniques discussed in Section 3:

- mean excess function,
- limited expected value function,
- and probability plot,

are relatively simple, yet, at the same time provide valuable assistance in the estimation process.

Once the stochastic models governing the incidence and severity of claims have been identified, they can be combined into the so called aggregate claim process:

$$S_t = \sum_{k=1}^{N_t} X_k, \tag{1}$$

where the claim severities are described by the random sequence $\{X_k\}$ with finite mean and the number of claims in the interval $(0, t]$ is modeled by a counting process $N_t$, often called the claim arrival process. The risk process $\{R_t\}_{t\geq 0}$ describing the capital of an insurance company is then defined as:

$$R_t = u + c(t) - S_t, \tag{2}$$

where the non-negative constant $u$ stands for the initial capital of the company and $c(t)$ is the premium the company receives for sold insurance policies.

The simplicity of the risk process (2) is only illusionary. In most cases no analytical conclusions regarding the time evolution of the process can be drawn. However, it is this evolution that is important for practitioners, who have to calculate functionals of the risk process like the expected time to ruin and the ruin probability. All this calls for efficient numerical simulation schemes (Burnecki, Härdle, and Weron, 2004) and powerful inference tools. In Section 4 we will present four such tools:

- ruin probability plot,
- density evolution plot,
- quantile lines,
- and probability gates.

All four techniques allow for immediate evaluation of the models' adequacy and the risks faced by the company based on visual inspection of the generated plots. As such they are especially useful for high-level managers who are interested in a general overview of the situation and do not need to study all the computational details underlying the final results.

To illustrate the usefulness of the presented visualization tools, throughout this chapter we will apply them to two datasets. The first one, studied in detail, is the Property Claim Services (PCS) dataset, which covers losses resulting from catastrophic events in USA. The data includes 1990-1999 market's loss amounts in USD adjusted for inflation using the Consumer Price Index. Only natural perils which caused damages exceeding 5 million dollars were taken into account, see Fig. 1. The second dataset, used here for illustration of the risk process inference tools only, concerns major inflation-adjusted Danish fire losses in profits (in Danish Krone, DKK) that occurred between 1980 and 1990 and were recorded by Copenhagen Re.

## 2 Software

Visualization tools would not be of much help without adequate software. All methods discussed in this chapter are demonstrated using two software libraries and one standalone application. The *Insurance Library* of XploRe (`www.xplore-stat.de`) is a collection of quantlets that illustrate various topics related to insurance (Čížek, Härdle, and Weron, 2005). It is accompanied by on-line, hyperlinked and freely downloadable from the web tutorials (`www.xplore-stat.de/tutorials/_Xpl_Tutorials.html`).
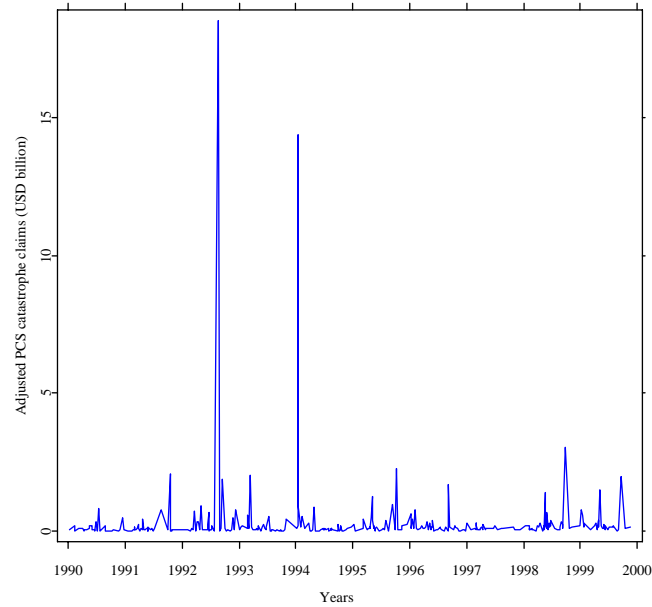
**Fig. 1.** Graph of the PCS catastrophe loss data, 1990-1999. Two largest losses in this period were caused by Hurricane Andrew (24 August 1992) and the Northridge Earthquake (17 January 1994). *From:* XploRe.

The *Ruin Probabilities* toolbox for Matlab is a set of m-functions with a graphical user interface (GUI) intended primarily for visualizing risk processes and evaluating ruin probabilities (Miśta, 2002). It can be downloaded from `www.im.pwr.wroc.pl/~hugo/stronaHSC/Podstrony/Programy.html`.

The *SDE-Solver* is a standalone application for the Windows environment. It allows for construction and visualization of solutions of stochastic differential equations (SDE's) with Gaussian, Poisson, and stable random measures (Janicki, Izydorczyk, and Gradalski, 2003); for SDE modeling concepts consult also Kloeden and Platen (1995). The graphics makes use of quantile lines and density evolution techniques and introduces an interesting concept of interactive probability gates, which give the probability that the simulated process passes through a specified interval at a specified point in time (for details see Section 4.4). More information about the software can be found at `www.math.uni.wroc.pl/~janicki/solver.html`.

## 3 Fitting Loss and Waiting Times Distributions

The derivation of loss and waiting times (inter-arrival) distributions from insurance data is not an easy task. There are three basic approaches: empirical, analytical, and moment based. The analytical approach is probably the most often used in practice and certainly the most frequently adopted in the actuarial literature. It reduces to finding a suitable analytical expression which fits the observed data well and which is easy to handle (Daykin, Pentikainen, and Pesonen, 1994).

Having a large collection of distributions to choose from, we need to narrow our selection to a single model and a unique parameter estimate. The type of the objective loss distribution (the waiting times distribution can be analyzed analogously) can be easily selected by comparing the shapes of the empirical and theoretical mean excess functions. Goodness-of-fit can be verified by plotting the corresponding limited expected value functions or making probability plots. Finally, the hypothesis that the modeled random event is governed by a certain loss distribution can be statistically tested (but is not discussed in this chapter; for a recent review of goodness-of-fit hypothesis testing see Burnecki, Misiorek, and Weron, 2005).

In the following subsections we will apply the above mentioned visual inference tools to PCS data. They will narrow our search for the optimal analytical model of the loss and waiting times distributions to one or two probability laws. Moreover, they will allow for a visual assessment of the goodness-of-fit.

### 3.1 Mean Excess Function

For a claim amount random variable $X$, the mean excess function or mean residual life function is the expected payment per claim on a policy with a fixed amount deductible of $x$, where claims with amounts less than or equal to $x$ are completely ignored:

$$e(x) = \mathrm{E}(X - x|X > x) = \frac{\int_x^\infty \{1 - F(u)\}\, du}{1 - F(x)}. \tag{3}$$

In practice, the mean excess function $e$ is estimated by $\hat{e}_n$ based on a representative sample $x_1, \ldots, x_n$:

$$\hat{e}_n(x) = \frac{\sum_{x_i > x} x_i}{\#\{i : x_i > x\}} - x. \tag{4}$$

Note, that in a financial risk management context, switching from the right tail to the left tail, $e(x)$ is referred to as the expected shortfall (Weron, 2004).

When considering the shapes of mean excess functions, the exponential distribution with cumulative distribution function (cdf) $F(x) = 1 - \exp(-\beta x)$ plays a central role. It has the memoryless property, meaning that whether

the information $X > x$ is given or not, the expected value of $X - x$ is the same as if one started at $x = 0$ and calculated $E(X)$. The mean excess function for the exponential distribution is therefore constant. One in fact easily calculates that for this case $e(x) = 1/\beta$ for all $x > 0$.

If the distribution of $X$ is heavier-tailed than the exponential distribution we find that the mean excess function ultimately increases, when it is lighter-tailed $e(x)$ ultimately decreases. Hence, the shape of $e(x)$ provides important information on the sub-exponential or super-exponential nature of the tail of the distribution at hand. That is why, in practice, this tool is used not only for discovering the relevant class of claim size distributions, but also for investigating all kinds of phenomena. We will apply it to both PCS loss and waiting times data.

Mean excess functions of the well known and widely used distributional classes are given by the following formulas (selected shapes are sketched in Fig. 2, while the empirical mean excess functions $\hat{e}_n(x)$ for the PCS catastrophe data are plotted in Fig. 3):

- log-normal distribution with cdf $F(x) = \Phi\{(\log x - \mu)/\sigma\}$:

$$e(x) = \frac{\exp\left(\mu + \frac{\sigma^2}{2}\right)\left\{1 - \Phi\left(\frac{\ln x - \mu - \sigma^2}{\sigma}\right)\right\}}{\left\{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right\}} - x,$$

  where $\Phi(\cdot)$ is the standard normal (with mean 0 and variance 1) distribution function;
- Pareto distribution with cdf $F(x) = 1 - \{\lambda/(\lambda + x)\}^\alpha$:

$$e(x) = \frac{\lambda + x}{\alpha - 1}, \qquad \alpha > 1;$$

- Burr distribution with cdf $F(x) = 1 - \{\lambda/(\lambda + x^\tau)\}^\alpha$:

$$e(x) = \frac{\lambda^{1/\tau}\Gamma\left(\alpha - \frac{1}{\tau}\right)\Gamma\left(1 + \frac{1}{\tau}\right)}{\Gamma(\alpha)} \cdot \left(\frac{\lambda}{\lambda + x^\tau}\right)^{-\alpha} \cdot$$

$$\cdot \left\{1 - B\left(1 + \frac{1}{\tau}, \alpha - \frac{1}{\tau}, \frac{x^\tau}{\lambda + x^\tau}\right)\right\} - x,$$

  where $\Gamma(\cdot)$ is the standard gamma function and $B(\cdot, \cdot, \cdot)$ is the beta function;
- Weibull distribution with cdf $F(x) = 1 - \exp(-\beta x^\tau)$:

$$e(x) = \frac{\Gamma(1 + 1/\tau)}{\beta^{1/\tau}}\left\{1 - \Gamma\left(1 + \frac{1}{\tau}, \beta x^\tau\right)\right\}\exp\left(\beta x^\tau\right) - x,$$

  where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function;
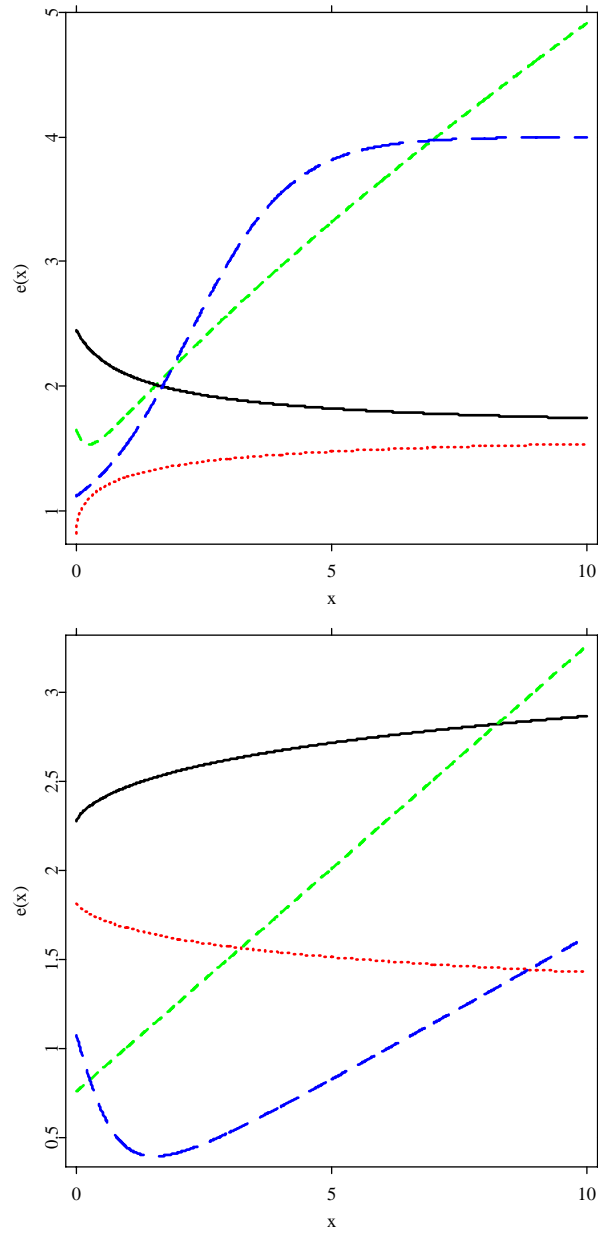
**Fig. 2.** *Top panel:* Shapes of the mean excess function $e(x)$ for the log-normal (green dashed line), gamma with $\alpha < 1$ (red dotted line), gamma with $\alpha > 1$ (black solid line) and a mixture of two exponential distributions (blue long-dashed line). *Bottom panel:* Shapes of the mean excess function $e(x)$ for the Pareto (green dashed line), Burr (blue long-dashed line), Weibull with $\tau < 1$ (black solid line) and Weibull with $\tau > 1$ (red dotted line) distributions. *From:* XploRe.
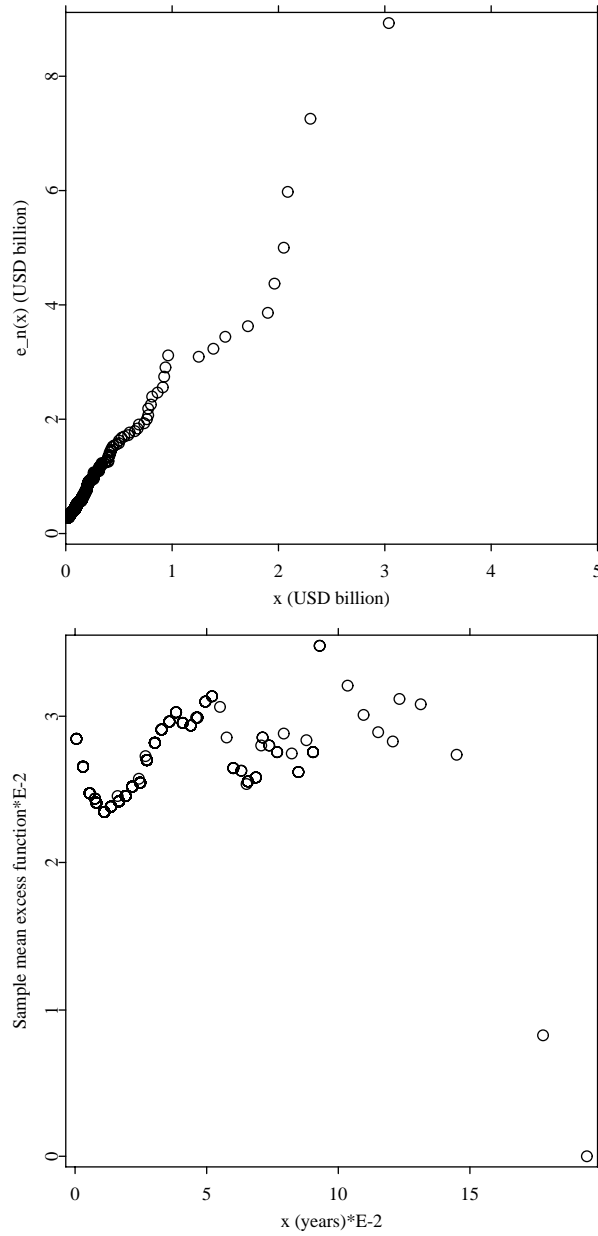
**Fig. 3.** The empirical mean excess function $\hat{e}_n(x)$ for the PCS catastrophe loss amounts in billion USD (*top panel*) and waiting times in years (*bottom panel*). Comparison with Fig. 2 suggests that log-normal, Pareto, and Burr distributions should provide a good fit for loss amounts, while log-normal, Burr, and exponential laws for the waiting times. *From:* XploRe.

- gamma distribution with cdf $F(x) = \int_0^x \beta(\beta s)^{\alpha-1} \exp(-\beta s)/\Gamma(\alpha)ds$:

$$e(x) = \frac{\alpha}{\beta} \cdot \frac{1 - F(x, \alpha + 1, \beta)}{1 - F(x, \alpha, \beta)} - x,$$

  where $F(x, \alpha, \beta)$ is the gamma cdf;
- mixture of 2 exponential distributions with cdf $F(x) = a\{1 - \exp(-\beta_1 x)\} + (1 - a)\{1 - \exp(-\beta_2 x)\}$:

$$e(x) = \frac{\frac{a}{\beta_1} \exp(-\beta_1 x) + \frac{1-a}{\beta_2} \exp(-\beta_2 x)}{a \exp(-\beta_1 x) + (1 - a) \exp(-\beta_2 x)}.$$

Comparison of Figures 2 and 3 suggests that log-normal, Pareto, and Burr distributions should provide a good fit for the loss amounts. The maximum likelihood estimates of the parameters of these three distributions are the following: $\mu = 18.4478$ and $\sigma = 1.1554$ (log-normal), $\alpha = 1.9975$ and $\lambda = 2.3248 \cdot 10^8$ (Pareto), and $\alpha = 0.4135$, $\lambda = 3.235 \cdot 10^{18}$ and $\tau = 2.4135$ (Burr). Unfortunately, the parameters of the Burr distribution imply that the first moment is infinite which contradicts the assumption that the claim amount random sequence $\{X_k\}$ has finite mean. This assumption seems natural in the insurance world since any premium formula usually includes the expected value of $X_k$. Therefore, we exclude the Burr distribution from the analysis of claim severities.

The classification of waiting times data is not that straightforward. If we discard large observations then log-normal and Burr laws should yield a good fit. However, if all waiting times are taken into account then the empirical mean excess function $\hat{e}_n(x)$ approximately forms a straight line (though somewhat oscillating) and suggests that the exponential law could be a reasonable alternative.

### 3.2 Limited Expected Value Function

The limited expected value function $L$ of a claim size variable $X$, or of the corresponding cdf $F(x)$, is defined by

$$L(x) = \mathrm{E}\{\min(X, x)\} = \int_0^x y dF(y) + x\{1 - F(x)\}, \quad x > 0. \qquad (5)$$

The value of the function $L$ at point $x$ is equal to the expectation of the cdf $F(x)$ truncated at this point. In other words, it represents the expected amount per claim retained by the insured on a policy with a fixed amount deductible of $x$. The empirical estimate is defined as follows:

$$\hat{L}_n(x) = \frac{1}{n}\left(\sum_{x_j < x} x_j + \sum_{x_j \geq x} x\right). \qquad (6)$$

In order to fit the limited expected value function $L$ of an analytical distribution to the observed data, the estimate $\hat{L}_n$ is first constructed. Thereafter one tries to find a suitable analytical cdf $F$, such that the corresponding limited expected value function $L$ is as close to the observed $\hat{L}_n$ as possible.

The limited expected value function (LEVF) has the following important properties:

(i) the graph of $L$ is concave, continuous and increasing;
(ii) $L(x) \to E(X)$, as $x \to \infty$;
(iii) $F(x) = 1 - L'(x)$, where $L'(x)$ is the derivative of the function $L$ at point $x$; if $F$ is discontinuous at $x$, then the equality holds true for the right-hand derivative $L'(x+)$.

The reason why the limited expected value function is a particularly suitable tool for our purposes is that it represents the claim size distribution in the monetary dimension. For example, we have $L(\infty) = E(X)$ if it exists. The cdf $F$, on the other hand, operates on the probability scale, i.e. takes values between 0 and 1. Therefore, it is usually difficult to see, by looking only at $F(x)$, how sensitive the price for the insurance – the premium – is to changes in the values of $F$, while the LEVF shows immediately how different parts of the claim size cdf contribute to the premium. Apart from curve-fitting purposes, the function $L$ turns out to be a very useful concept in dealing with deductibles (Burnecki, Nowicka-Zagrajek, and Wyłomańska, 2005). It is also worth mentioning, that there exists a connection between the limited expected value function and the mean excess function:

$$E(X) = L(x) + P(X > x)e(x). \tag{7}$$

The limited expected value functions (LEVFs) for all distributions considered in this chapter are given by:

- exponential distribution:

$$L(x) = \frac{1}{\beta}\left\{1 - \exp(-\beta x)\right\};$$

- log-normal distribution:

$$L(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right)\Phi\left(\frac{\ln x - \mu - \sigma^2}{\sigma}\right) + x\left\{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right\};$$

- Pareto distribution:

$$L(x) = \frac{\lambda - \lambda^\alpha(\lambda + x)^{1-\alpha}}{\alpha - 1};$$

- Burr distribution:

$$L(x) = \frac{\lambda^{1/\tau}\Gamma\left(\alpha - \frac{1}{\tau}\right)\Gamma\left(1 + \frac{1}{\tau}\right)}{\Gamma(\alpha)}\mathrm{B}\left(1 + \frac{1}{\tau}, \alpha - \frac{1}{\tau}; \frac{x^\tau}{\lambda + x^\tau}\right)$$
$$+ x\left(\frac{\lambda}{\lambda + x^\tau}\right)^\alpha;$$

- Weibull distribution:

$$L(x) = \frac{\Gamma\left(1+1/\tau\right)}{\beta^{1/\tau}}\Gamma\left(1+\frac{1}{\tau},\beta x^\alpha\right) + xe^{-\beta x^\alpha};$$

- gamma distribution:

$$L(x) = \frac{\alpha}{\beta}F(x,\alpha+1,\beta) + x\left\{1 - F(x,\alpha,\beta)\right\};$$

- mixture of 2 exponential distributions:

$$L(x) = \frac{a}{\beta_1}\left\{1 - \exp\left(-\beta_1 x\right)\right\} + \frac{1-a}{\beta_2}\left\{1 - \exp\left(-\beta_2 x\right)\right\}.$$

From the curve-fitting point of view the use of the LEVFs has the advantage, compared with the use of the cdfs, that both the analytical and the corresponding observed function $\hat{L}_n$, based on the observed discrete cdf, are continuous and concave, whereas the observed claim size cdf $F_n$ is a discontinuous step function. Property (iii) implies that the limited expected value function determines the corresponding cdf uniquely. When the limited expected value functions of two distributions are close to each other, not only are the mean values of the distributions close to each other, but the whole distributions as well.

Since the LEVF represents the claim size distribution in the monetary dimension, it is usually used exclusively for analyzing price data. In Figure 4 we depict the empirical and analytical LEVFs for the two distributions suggested by the mean excess function to best fit the PCS catastrophe loss amounts. We can see that the Pareto law is definitely superior to the log-normal one.

### 3.3 Probability plot

The purpose of the probability plot is to graphically assess whether the data comes from a specific distribution. It can provide some assurance that this assumption is not being violated, or provide an early warning of a problem with our assumptions.

The probability plot is constructed in the following way. First, the observations $x_1, ..., x_n$ are ordered from the smallest to the largest: $x_{(1)} \le ... \le x_{(n)}$. Next, they are plotted against their observed cumulative frequency, i.e. the points (pluses in Figs. 5-8) correspond to the pairs $(x_{(i)}, F^{-1}([i - 0.5]/n))$, for $i = 1, ..., n$. If the hypothesized distribution $F$ adequately describes the data, the plotted points fall approximately along a straight line. If the plotted points deviate significantly from a straight line, especially at the ends, then the hypothesized distribution is not appropriate.

In Figure 5 the Pareto probability plot of the PCS loss data is displayed. Apart from the two very extreme observations (corresponding to Hurricane Andrew and Northridge Earthquake) the points more or less constitute a
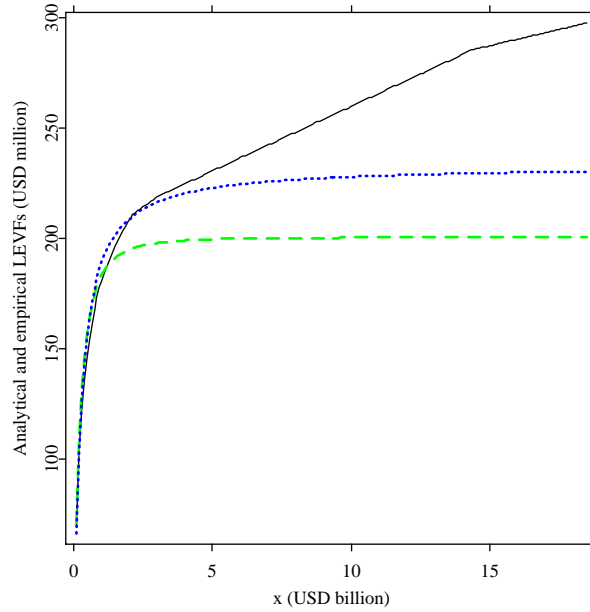
**Fig. 4.** The empirical (black solid line) and analytical limited expected value functions (LEVFs) for the log-normal (green dashed line) and Pareto (blue dotted line) distributions for the PCS loss catastrophe data. *From:* XploRe.

straight line, validating the choice of the Pareto distribution. In Figures 6 and 7 the log-normal probability plots of the PCS data are presented. To this end we applied the standard normal probability plots to the logarithms of the losses and waiting times, respectively. Figure 6 suggests that the log-normal distribution for the loss severities is not the best choice, whereas Figure 7 justifies the use of that particular distribution to the waiting times data. Figure 8 depicts the exponential probability plot of the latter dataset. We can see that the exponential distribution is not a very good candidate for the underlying distribution of the waiting times data – the points deviate from a straight line in the far end. Nevertheless, the deviation is not that large and the exponential law may be an acceptable model.

The presented probability plots suggest that, as far as the loss amounts are concerned, the Pareto law provides a much better fit than the log-normal distribution. In fact, apart from the two very extreme observations (Hurricane Andrew and Northridge Earthquake), it does provide a very good fit. Note, that the procedure of trimming the top 1-5% of the data before calibration is known as robust estimation and leads to very similar conclusions (see a recent paper by Chernobai et al., 2006).

From the probability plots we also infer that the waiting times data can be described by the log-normal and – to some degree – the exponential dis-
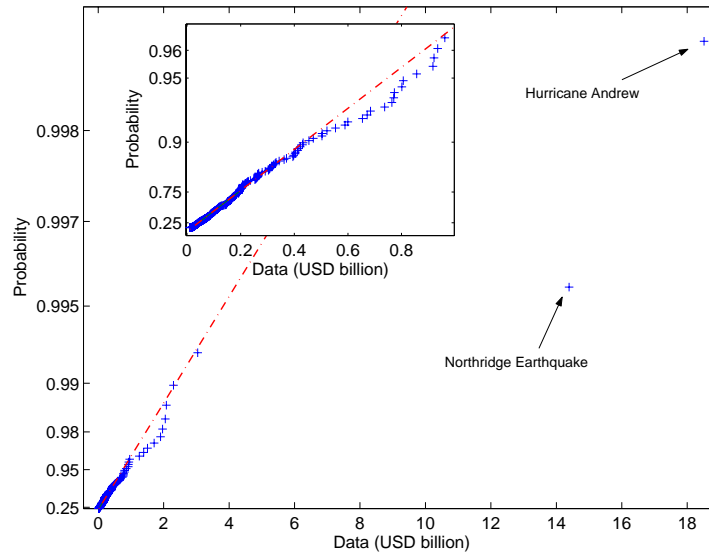
**Fig. 5.** Pareto probability plot of the PCS loss data. Apart from the two very extreme observations (Hurricane Andrew and Northridge Earthquake) the points (pluses) more or less constitute a straight line, validating the choice of the Pareto distribution. The inset is a magnification of the bottom left part of the original plot. *From:* Ruin Probabilities Toolbox.
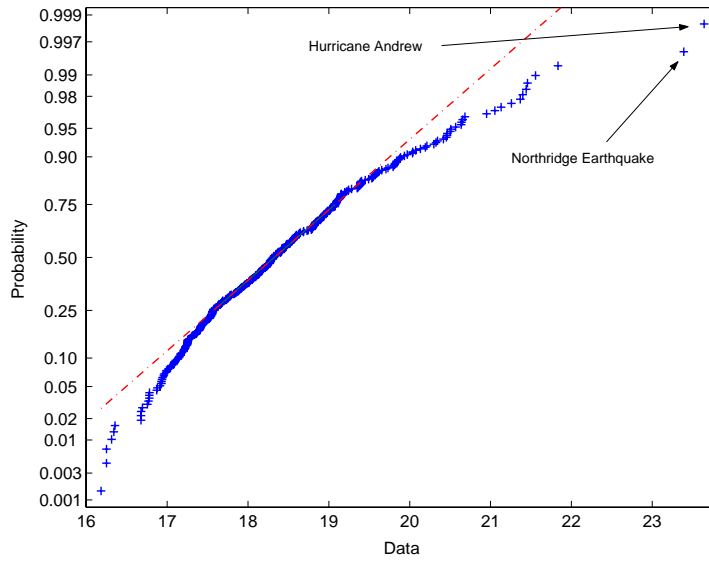


**Fig. 6.** Log-normal probability plot of the PCS loss data. The $x$-axis corresponds to logarithms of the losses. The deviations from the straight line at both ends question the adequacy of the log-normal law. *From:* Ruin Probabilities Toolbox.
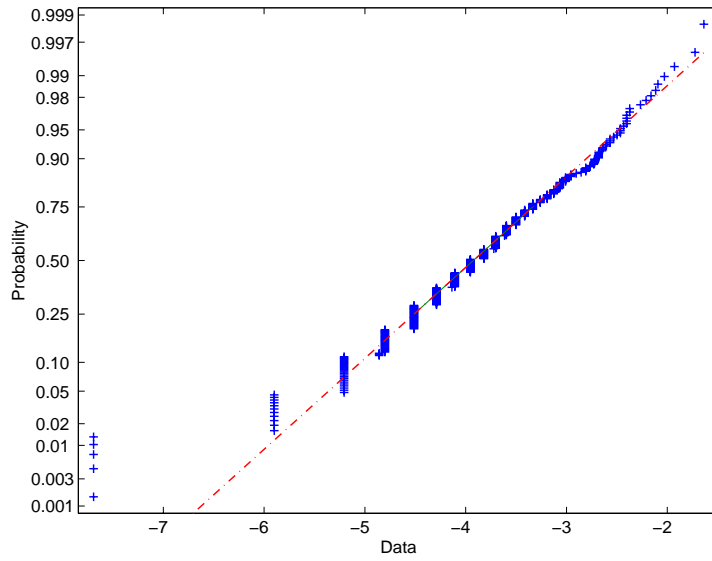
**Fig. 7.** Log-normal probability plot of the PCS waiting times data. The $x$-axis corresponds to logarithms of the losses. *From:* Ruin Probabilities Toolbox.
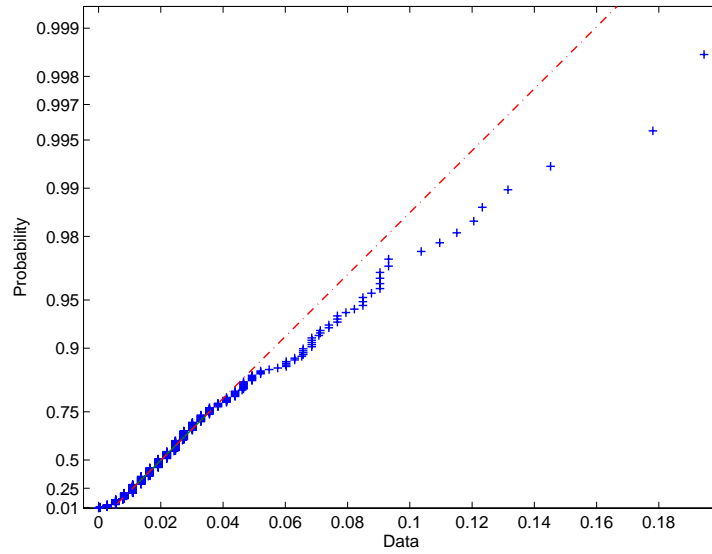


**Fig. 8.** Exponential probability plot of the PCS waiting times data. The plot deviates from a straight line in the far end. *From:* Ruin Probabilities Toolbox.

tribution. The maximum likelihood estimates of the parameters of these two distributions are given by: $\mu = -3.9431$ and $\sigma = 0.9685$ (log-normal) and $\beta = 34.572$ (exponential).

## 4 Risk Process and its Visualization

The risk process (2) is a sum of the initial capital and the premium function minus the aggregate claim process governed by two random phenomena – the severity and incidence of claims. In many practical situations it is reasonable to consider the counting process $N_t$ (responsible for the incidence of events) to be (i) a renewal process, i.e. a counting process with inter-arrival times being i.i.d. positive random variables with mean $1/\lambda$, and (ii) independent of the claim severities $\{X_k\}$. In such a case, the premium function can be defined in a natural way as $c(t) = (1 + \theta)\mu\lambda t$, where $\mu$ is the expectation of $X_k$'s and $\theta > 0$ is the relative safety loading on the premium which "guarantees" survival of the insurance company. The (homogeneous) Poisson process (HPP) is a special case of the renewal process with exponentially distributed waiting times.

Two standard ways of presenting sample trajectories of the risk process are displayed in Figs. 9 and 10. Here, we use the Danish fire losses dataset, which can be modeled by a log-normal claim amount distribution with parameters $\mu = 12.6795$ and $\sigma = 1.4241$ (obtained via maximum likelihood) and a HPP counting process with monthly intensity $\lambda = 4.81$. The company's initial capital is assumed to be $u = 10$ million DKK. In Figure 9 five real (discontinuous) sample realizations of the resulting risk process are presented, whereas in Figure 10 the trajectories are shown in a continuous fashion. The latter way of depicting sample realizations seems to be more illustrative. Notice also that one of the trajectories falls below zero indicating a scenario leading to a bankruptcy of the insurance company.

### 4.1 Ruin Probability Plots

In examining the nature of the risk associated with a portfolio of business, it is often of interest to assess how the portfolio may be expected to perform over an extended period of time. In this context the ruin theory (Grandell, 1991) comes in handy. Ruin theory is concerned with the excess of the income $c(t)$ (with respect to a portfolio of business) over the outgo, or claims paid, $S(t)$. This quantity, referred to as insurer's surplus, varies in time. Specifically, ruin is said to occur if the insurer's surplus reaches a specified lower bound, e.g. minus the initial capital. This can be observed in Figure 10 where the time of ruin is denoted by a star. One measure of risk is the probability of such an event, clearly reflecting the volatility inherent in the business. In addition, it can serve as a useful tool in long range planning for the use of insurer's funds.
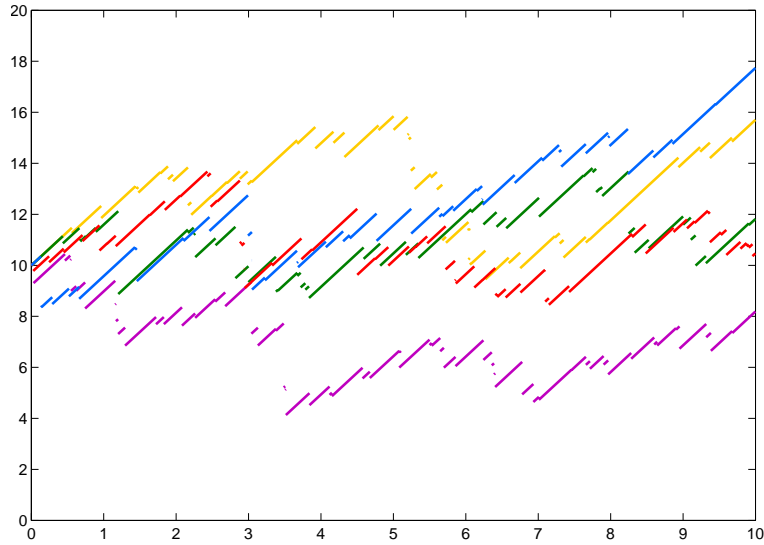
The ruin probability in finite time $T$ is given by

**Fig. 9.** Discontinuous visualization of the trajectories of a risk process. The initial capital $u = 10$ million DKK, the relative safety loading $\theta = 0.05$, the claim size distribution is log-normal with parameters $\mu = 12.6795$ and $\sigma = 1.4241$, and the driving counting process is a HPP with monthly intensity $\lambda = 4.81$. *From:* Ruin Probabilities Toolbox.
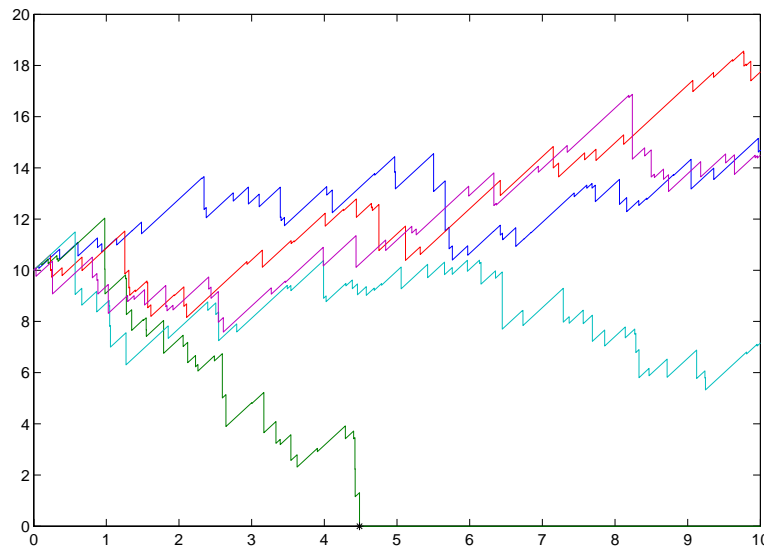


**Fig. 10.** Alternative (continuous) visualization of the trajectories of a risk process. The bankruptcy time is denoted by a star. The parameters of the risk process are the same as in Fig. 9. *From:* Ruin Probabilities Toolbox.
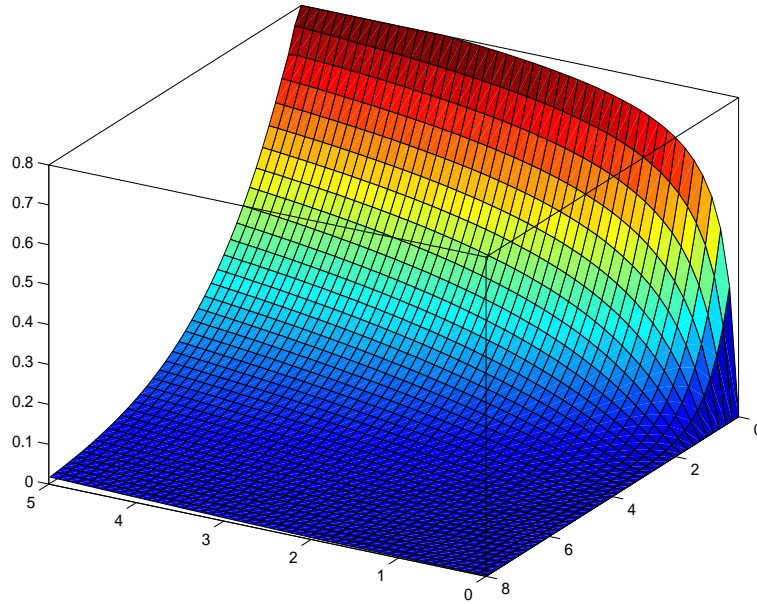
**Fig. 11.** Ruin probability plot with respect to the time horizon $T$ (left axis, in months) and the initial capital $u$ (right axis, in million DKK). The relative safety loading $\theta = 0.15$; other parameters of the risk process are the same as in Fig. 9. *From:* Ruin Probabilities Toolbox.

$$\psi(u, T) = \mathrm{P} \left( \inf_{0 < t < T} \{R_t\} < 0 \right). \tag{8}$$

Most insurance managers will closely follow the development of the risk business and increase the premium if the business behaves badly. The planning horizon may be thought of as the sum of the following: the time until the risk business is found to behave "badly", the time until the management reacts, and the time until a decision of a premium increase takes effect (Grandell, 1991).

In Figure 11 a 3-dimensional (3D) visualization of the ruin probability with respect to the initial capital $u$ varying from 0 to 8 million DKK and the time horizon $T$ changing from 0 to 5 months is depicted. The remaining parameters of the risk process are the same as those used in Figs. 9 and 10, only the relative safety loading was raised from $\theta = 0.05$ to $\theta = 0.15$. We can see that the ruin probability increases with the time horizon and decreases as the initial capital grows.

The ruin probability in finite time can always be computed directly using Monte Carlo simulations. Naturally, the choice of the intensity function and the distribution of claim severities heavily affects the simulated values and the ruin probability. Graphical tools that are presented below can help in
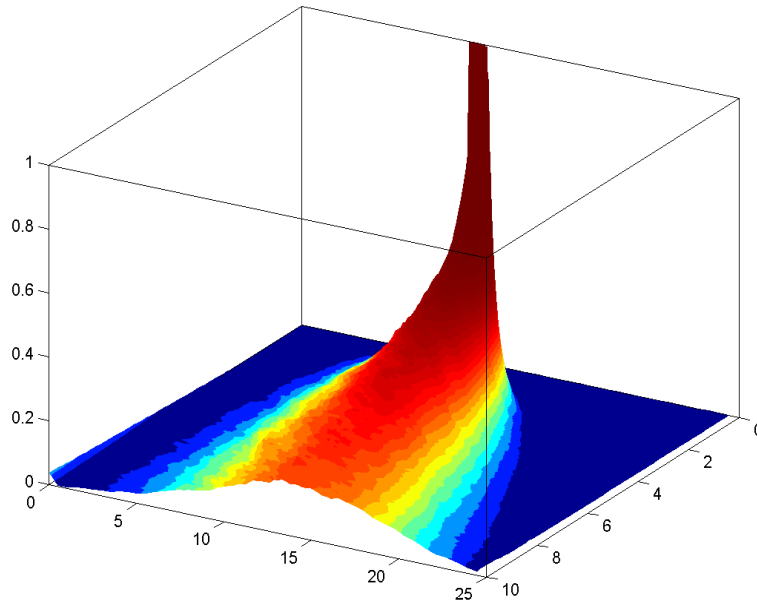
**Fig. 12.** 3-dimensional visualization of the density evolution of a risk process with respect to the risk process value $R_t$ (left axis) and time $t$ (right axis). The parameters of the risk process are the same as in Fig. 11. *From:* Ruin Probabilities Toolbox.

assessing whether the choices made result in a reasonable risk process and, hence, greatly reduce the time needed to construct an adequate model.

### 4.2 Density Evolution

Density evolution plots (and their 2-dimensional projections) are a visually attractive method of representing the time evolution of a process. At each time point $t = t_0, t_1, ..., t_n$, a density estimate of the distribution of process values at this time point is evaluated, see Fig. 12 (the same parameters of the risk process as in Fig. 11 were used). Then the densities are plotted on a grid of $t$ values. The obtained 3-dimensional surface can be further rotated to better present the behavior of the process in a particular interval.

A 2-dimensional projection of the density evolution plot (see Fig. 13) reveals equivalent information to that represented by the quantile lines (see Section 4.3). Only in this case the presented visual information is more attractive to the eye, however, not that rigid. Depending on the discretization of the time and process value intervals and the kernel density estimate used, slightly different density evolution plots can be obtained (Janicki and Izydorczyk, 2001; Miśta, 2002).
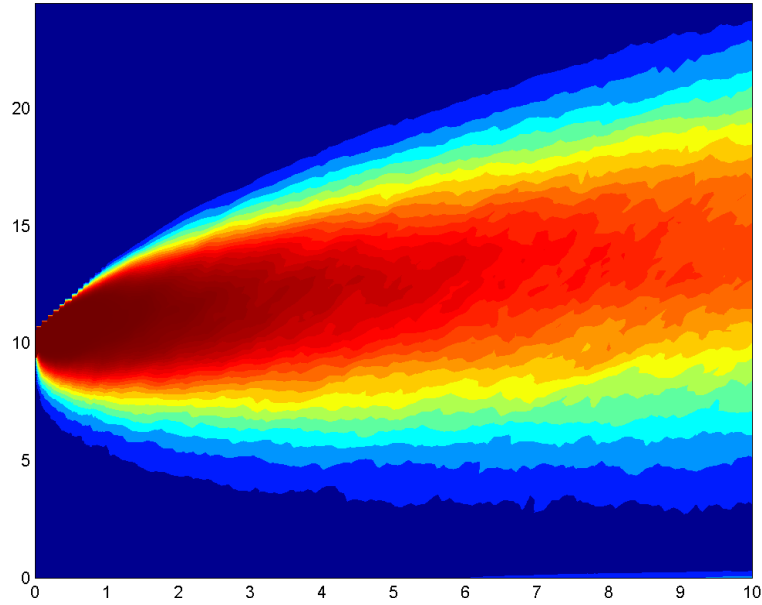
**Fig. 13.** 2-dimensional projection of the density evolution depicted in Fig. 12. *From:*
Ruin Probabilities Toolbox.

### 4.3 Quantile Lines

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 empirical distribution function (edf). Recall, that for a sample of
observations $\{x_1, \ldots, x_n\}$ the edf is defined as:

$$F_n(x) = \frac{1}{n}\#\{i : x_i \leq x\}, \tag{9}$$

i.e. it is a piecewise constant function with jumps of size $1/n$ at points $x_i$
(Burnecki, Misiorek, and Weron, 2005).

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 Janicki and Weron (1994). In Figures 14, 15, and
16 they visualize the evolution of the risk process.

Quantile lines can be also a useful tool for comparing two processes, see
Fig. 14. It depicts quantile lines and two sample trajectories of the risk process
and its diffusion approximation; consult Burnecki, Miśta, and Weron (2005)
for a discussion of different approximations in the context of ruin probability.
The parameters of the risk process are the same as in Fig. 11. We can see that
the quantile lines of the risk process and its approximation coincide. This
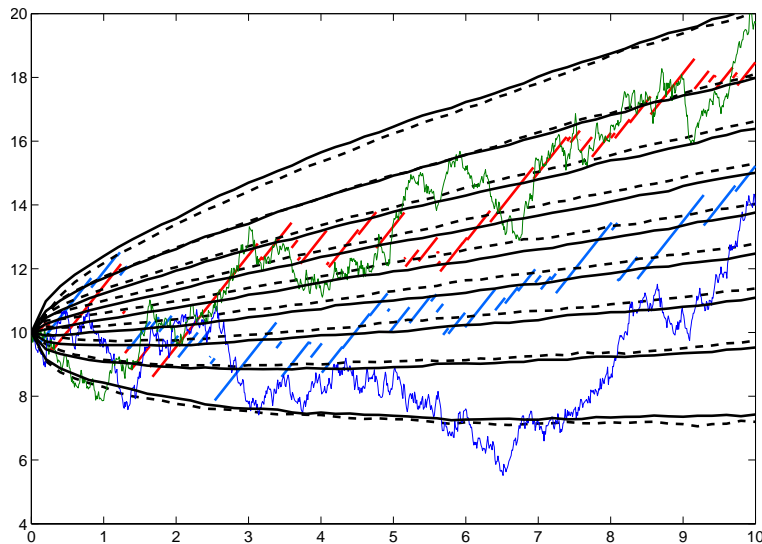justifies the use of the Brownian approximation for such parameters of the
risk process.

**Fig. 14.** A Poisson-driven risk process (discontinuous thin lines) and its Brownian motion approximation (continuous thin lines). The quantile lines allow for an easy and fast comparison of the processes. The thick solid lines represent the sample 0.1, ..., 0.9-quantile lines based on 10000 trajectories of the risk process, whereas thick dashed lines correspond to their approximation counterparts. The parameters of the risk process are the same as in Fig. 11. *From:* Ruin Probabilities Toolbox.

Now we return to the PCS dataset. To study the evolution of the risk process we simulate sample trajectories and compute quantile lines. We consider a hypothetical scenario where the insurance company insures losses resulting from catastrophic events in the United States. The company's initial capital is assumed to be $u = 100$ billion USD and the relative safety loading used is $\theta = 0.5$. We choose two different models of the risk process whose application is justified by the statistical analysis, see Section 3: a homogeneous Poisson process (HPP) with log-normal claim sizes and a renewal process with Pareto claim sizes and log-normal waiting times. The results are presented in Fig. 15. 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 different shapes of the "real" risk process in the subplots are due to the different forms of the premium function $c(t)$. The thin solid line is a sample trajectory. The dotted lines are the sample 0.001, 0.01, 0.05, 0.25, 0.50, 0.75, 0.95, 0.99, 0.999-quantile lines based on 3000 trajectories of the risk process. The quantile lines visualize the evolution of the density of the risk process. Clearly, if claim severities are Pareto distributed then extreme events are more probable to happen than in the log-normal case, for which the historical trajectory falls even outside the 0.001-quantile line. Figure 15 suggests that the second model (Pareto dis-
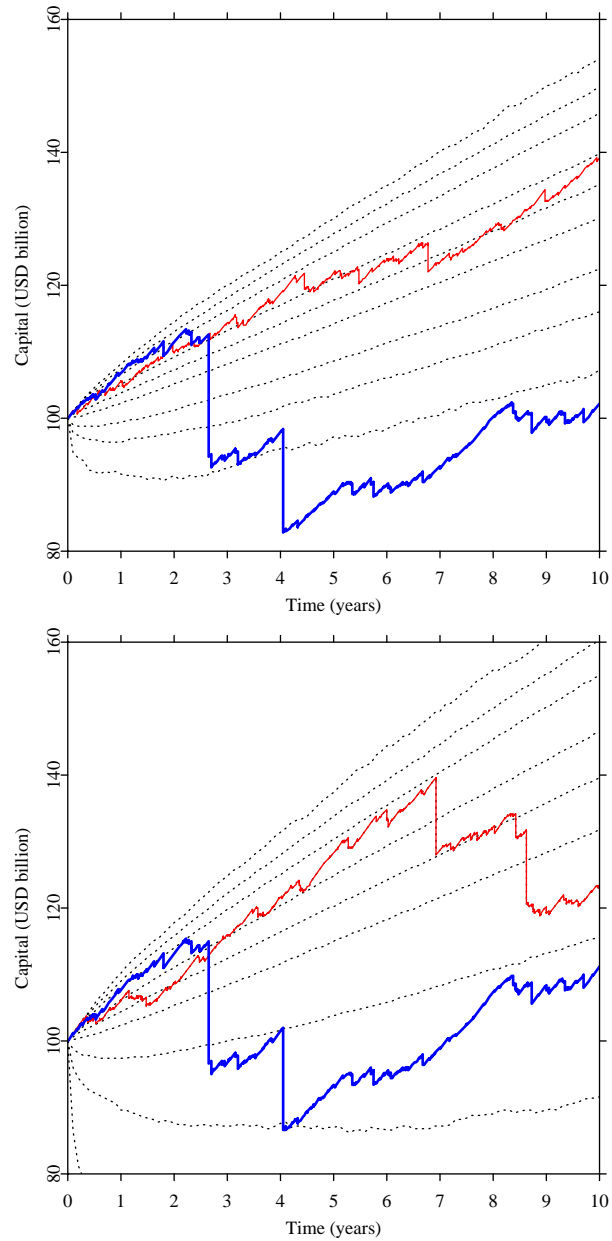
**Fig. 15.** The PCS data simulation results for a homogeneous Poisson process with log-normal claim sizes (*top panel*) and a renewal process with Pareto claim sizes and log-normal waiting times (*bottom panel*). The dotted lines are the sample 0.001, 0.01, 0.05, 0.25, 0.50, 0.75, 0.95, 0.99, 0.999-quantile lines based on 3000 trajectories of the risk process. *From:* XploRe.
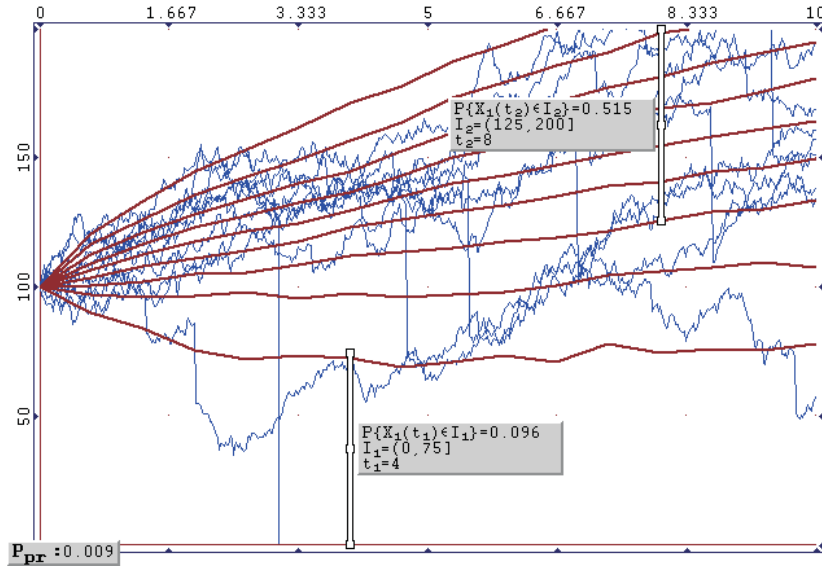
**Fig. 16.** "Probability gates" are an interactive graphical tool used for determining the probability that the process passes through a specified interval. The $0.1, ..., 0.9$-quantile lines (thick red lines) are based on 1000 simulated trajectories (thin blue lines) of the risk process originating at $u = 100$ billion USD. The parameters of the $\alpha$-stable Lévy motion approximation of the risk process were chosen to comply with PCS data. *From:* SDE-Solver.

tributed claim sizes and log-normal waiting times) yields a reasonable model for the "real" risk process.

### 4.4 Probability gates

"Probability gates" are an interactive graphical tool implemented in SDE-Solver. They can be of invaluable assistance in real-time analysis of the risk process and its models. A "probability gate" gives the so-called cylindrical probability $P\{X_{t_0} \in (a, b]\}$ that the simulated process $X_t$ passes through a specified interval $(a, b]$ at a specified point in time $t_0$ (Janicki and Izydorczyk, 2001; Janicki, Izydorczyk, and Gradalski, 2003). In Figure 16 two probability gates are defined; Monte Carlo simulations can be used to obtain the probability estimates. One yields the probability of the risk process falling below 75 billion after 4 years, i.e. $P\{R_4 \in (0, 75]\} = 0.096$. The other yields the probability of the risk process being in the range $(125, 200]$ billion after 8 years, i.e. $P\{R_8 \in (125, 200]\} = 0.515$. Additionally, product probabilities of the process passing through all defined gates are provided. In the above example the product probability of the risk process first falling below 75 billion (after 4 years) and then recovering to over 125 billion but less than 200 billion (after 8 years) is equal to: $P\{R_4 \in (0, 75], R_8 \in (125, 200]\} = 0.009$. The

parameters of the $\alpha$-stable Lévy motion approximation (Furrer, Michna, and Weron, 1997) of the risk process were chosen to comply with the catastrophic data example.

# References

Burnecki, K., Härdle, W., and Weron, R. (2004). Simulation of risk processes, *in* J. Teugels, B. Sundt (eds.) *Encyclopedia of Actuarial Science*, Wiley, Chichester.

Burnecki, K., Misiorek, A., and Weron, R. (2005). Loss Distributions, *in* P. Cizek, W. Härdle, R. Weron (eds.) *Statistical Tools for Finance and Insurance*, Springer, Berlin, 289–317.

Burnecki, K., Miśta, P., and Weron, A. (2005). Ruin Probabilities in Finite and Infinite Time, *in* P. Cizek, W. Härdle, R. Weron (eds.) *Statistical Tools for Finance and Insurance*, Springer, Berlin, 341–379.

Burnecki, K., Nowicka-Zagrajek, J., and Wyłomańska, A. (2005). Pure Risk Premiums under Deductibles, *in* P. Cizek, W. Härdle, R. Weron (eds.) *Statistical Tools for Finance and Insurance*, Springer, Berlin, 427–452.

Čižek, P., Härdle, W., and Weron, R. (eds.) (2005). *Statistical Tools for Finance and Insurance,* Springer, Berlin.

Chernobai, A., Burnecki, K., Rachev, S.T., Trück, S., and Weron, R. (2006). Modelling catastrophe claims with left-truncated severity distributions, *Computational Statistics*, forthcoming.

Daykin, C.D., Pentikainen, T., and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*, Chapman, London.

Grandell, J. (1991). *Aspects of Risk Theory*, Springer, New York.

Furrer, H., Michna, Z., and Weron, A. (1997). Stable Lévy motion approximation in collective risk theory, *Insurance: Mathematics and Economics* **20**: 97-114.

Janicki, A. and Izydorczyk, A. (2001). *Computer Methods in Stochastic Modeling*, WNT, Warszawa (in Polish).

Janicki, A., Izydorczyk, A., and Gradalski, P. (2003). Computer Simulation of Stochastic Models with SDE-Solver Software Package, *Lecture Notes in Computer Science* **2657**: 361-370.

Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behavior of $\alpha$-Stable Stochastic Processes*, Marcel Dekker.

Kloeden, P.E. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin.

Klugman, S.A., Panjer, H.H., and Willmot, G.E. (1998). *Loss Models: From Data to Decisions*, Wiley, New York.

Miśta, P. (2002). *Computer methods of ruin probability approximation and its control via optimal reinsurance policies*, Fundacja "Warta", Warsaw (in Polish).

Panjer, H.H. and Willmot, G.E. (1992). *Insurance Risk Models*, Society of Actuaries, Chicago.

Teugels, J. and Sundt, B. (2004). *Encyclopedia of Actuarial Science*, Wiley, Chichester.

Weron, R. (2004). Computationally Intensive Value at Risk Calculations, *in* J. E. Gentle, W. Härdle, Y. Mori (eds.) *Handbook of Computational Statistics*, Springer, Berlin, 911–950.

# HSC Research Report Series 2006

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html