PROOF
DOI: 10.37394/232020.2021.1.5

Simulation of Extreme Insured Losses in Natural Catastrophes

VIERA PACÁKOVÁ, PAVLA JINDROVÁ
Institute of Mathematics and Quantitative Methods
Faculty of Economics and Administration
University of Pardubice
Studentská 95, 532 10 Pardubice 2
CZECH REPUBLIC

Abstract: This article aims to present the application of probability modelling and simulations based on quantile function of extreme insured losses in the world natural catastrophes based on data in time period 1970-2014, published in Swiss Re Sigma No2/2015. Quantile function provides an appropriate and flexible approach to the probability modelling needed to obtain well-fitted tails. We are specifically interested in modelling and simulations the tails of loss distributions. In a number of applications of quantile functions in insurance and reinsurance risk management interest focuses particularly on the extreme observations in the upper tail of probability distribution. Fortunately it is possible to simulate the observations in one tail of distribution without simulating the central values. This advantage will be used for estimate a few extreme high insured losses in the world's natural catastrophes in future.

Key-Words: - Extreme losses, quantile function, order statistics, simulation, Pareto distribution.

1 Introduction

The enormous impact of catastrophic events on our society is deep and long. Not only we need to investigate the causes of such events and develop plans to protect against them, but also we will have to resolve the resulting huge financial loss.

The occurrences of catastrophic events are becoming more frequent (Fig.1) and also grow indemnity of insurance and reinsurance companies at these events.

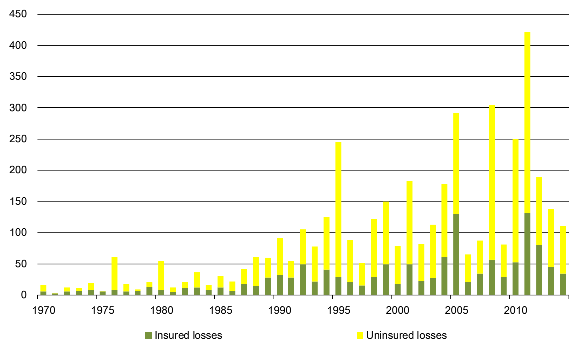

Fig.1 Number of catastrophic events, 1970-2014

Source: Swiss re economic Research&Consulting

From these facts it follows the need of knowledge the probability models for prediction of consequences of catastrophe events and thus select the best options to cover risks and correct setting premiums or reinsurance.

Developments of the financial consequences of disasters have a major impact on the global insurance market and forcing the insurance and reinsurance companies to seek for new approaches and ways to cover these risks. Raises the concern that the capacity of the world's insurance and reinsurance markets in the future will not be sufficient to cover these risks and aims to seek alternative options for their transfer.

In the modelling of extreme losses statistical methods are commonly used for inference from historical data. Different approaches had been proposed for certain circumstances, for example Extreme Value Theory, Excess over Threshold Method and other [4], [6]. We will present method for modelling and simulation based quantile function [1], [3], [5], [9].

2 Problem Formulation

Suppose losses are the independent, identically distributed (*iid*) random variables $X_1, X_2, \ldots$, with common cumulative distribution function (CDF)

$$F_X(x) = P(X \le x), \text{ where } x > 0 \qquad (1)$$

The Quantile Function, QF, denoted as $Q(p)$, expresses the *p*-quantile x_p as a function of *p*:

$x_p = Q(p)$, which is the value of x for which $p = P(X \le x_p) = F(x_p)$ [1], [3].

The definitions of the QF and the CDF can by written for any pairs of values (x, p) as $x = Q(p)$ and $p = F(x)$. These functions are simple inverses of each other, provided that they are both continuous increasing functions. Thus, we can also write $Q(p) = F^{-1}(p)$ and $F(x) = Q^{-1}(x)$.

We denoted a set of ordered sampling data of losses by

$$x_{(1)}, x_{(2)}, ..., x_{(r)}, ..., x_{(n-1)}, x_{(n)}.$$

The corresponding random variables are being denoted by

$$X_{(1)}, X_{(2)}, ..., X_{(r)}, ..., X_{(n-1)}, X_{(n)}.$$

Thus $X_{(n)}$ for example is the random variable representing the largest observation of the sample of *n* observations. The *n* random variables are referred as the *n* order statistics. These statistics play a major role in modelling with quantile distribution $Q(p)$.

Consider first the distribution of the largest observations on $X_{(n)}$ with distribution function denoted $F_{(n)}(x) = p_{(n)}$. By [1], [5], [7], [9] the probability

$$F_{(n)}(x) = p_{(n)} = P(X_{(n)} \le x)$$

is also probability that all *n* independent observations on *X* are less than or equal to this value *x*, which for each one is *p*. By the multiplication law of probability

$$p_{(n)} = p^n \text{ so } p = p_{(n)}^{1/n} \text{ and } F(x) = p = p_{(n)}^{1/n}.$$

Inverting $F(x)$, to get the quantile function, we have

$$Q_{(n)}(p_{(n)}) = Q(p_{(n)}^{1/n}) \quad (2)$$

So the quantile function of the largest observation is thus found from the original quantile function in very simple of calculation.

For the general *r*-th order statistic $X_{(r)}$ by so called *The order statistics distribution rule* [1] we get the result: If a sample of *n* observations from a distribution with quantile function $Q(p)$ is ordered, then the quantile function of the distribution of the *r*-th order statistic is given by

$$Q_{(r)}(p_{(r)}) = Q(BETAINV(p_{(r)}, r, n-r+1)) \quad (3)$$

$BETAINV(\cdot)$ is a standard function in packages such as Excel. Thus, the quantiles of the order statistics can be evaluated directly from the distribution $Q(p)$ of the data.

By the Uniform transformation rule (Gilchrist) if *U* has a uniform distribution then the variable *X*, where $x = Q(u)$ has a distribution with quantile function $Q(p)$. Thus data and distributions can be visualized as g enerated from the uniform distribution by transformation $Q(\cdot)$, where $Q(p)$ is the quantile function.

The uniform transformation rule shows that the values of *x* from any distribution with quantile function $Q(p)$ can be simulated as

$$x_i = Q(u_i), \quad i = 1, 2, ..., n$$

where $u_1, u_2, ..., u_n$ are simulated from uniform distribution on t he interval [0, 1]. The non-decreasing nature of $Q(\ )$ ensures the proper ordering of the *x*.

The quantile function thus provides the natural way to simulate values for those distributions for which it is an explicit function of *p*.

2.1 Simulation of extremes

Quantile function allows simulating the observations in the upper tail of distribution without simulating the central values.

Consider the right-hand tail. The distribution of the largest observation has been shown to be $Q(p^{1/n})$. Thus by [1], [5], [9] the largest observation can be simulated as $x_{(n)} = Q(u_{(n)})$, where $u_{(n)} = v_n^{1/n}$ and v_n is a random number from interval [0, 1]. If we now generate a set of transformed variables by

$$\begin{aligned} u_{(n)} &= v_n^{1/n} \\ u_{(n-1)} &= (v_{n-1})^{\frac{1}{n-1}} \cdot u_{(n)} \\ u_{(n-2)} &= (v_{n-2})^{\frac{1}{n-2}} \cdot u_{(n-1)} \\ &\vdots \end{aligned} \quad (4)$$

where the v_i, $i = n, n-1, n-2, ...$ are simply simulated set of independent random uniform variables, not ordered in any way. It will be seen from their definitions that u_i, $i = n, n-1, n-2, ...$ form a decreasing series of values and $u_{(i-1)} < u_{(i)}$.

In fact, values $u_{(i)}$ form an ordering sequence from a uniform distribution. Notice that once $u_{(n)}$ is obtained, the relations have the general form

$$u_{(m)} = (v_m)^{\frac{1}{m}} \cdot u_{(m+1)}, \quad m = n, n-1, n-2, ...$$

The order statistics for the largest observations on *X* are then simulated by

$$x_{(n)} = Q(u_{(n)})$$
$$x_{(n-1)} = Q(u_{(n-1)}) \quad (5)$$
$$x_{(n-2)} = Q(u_{(n-2)})$$
$$\vdots$$

In most simulation studies m samples of n observations are generated and the sample analyzes *repeated m* times to give an overall view of their behavior. A technique that is sometimes used as an alternative to such simulation called a profile. Such a set of ideal observations could be provided by the median rankits, M_r, $r = 1, 2, ..., n$.

2.2 Pareto distribution

Modelling of the tail of the loss distributions in general insurance is one of the problem areas, where obtaining a good fit to the extreme tails is of major importance. Thus is of particular relevance in non-proportional reinsurance if we are required to choose or price a high-excess layer [2], [8], [12].

The Pareto distribution is often used as a model for insurance losses needed to obtain well-fitted tails.

The Pareto cumulative distribution function of the losses X_a that exceed known threshold a is [7], [8], [10]:

$$F_a(x) = p = 1 - \left(\frac{a}{x}\right)^b, \quad x \geq a \quad (6)$$

The quantile function QF we can derive by inverting this CDF in the form

$$Q(p) = \frac{a}{(1-p)^{1/b}} \quad (7)$$

The parameter b is the Pareto parameter and we need it estimate it, the most often by maximum likelihood method in the form [7], [9]

$$\frac{n}{\sum_{i=1}^{n} \ln\left(\frac{X_{a,i}}{a}\right)} \quad (8)$$

3 Problem Solution

The publication [11], Swiss Re Sigma No2/2015 in Table 10, page 41, provides data about the 40 most costly insurance losses (1970- 2015). These data are the basis for our analysis. These values are ranging from 3410 to 78638 million USD in 2014 prices.

We want to verify whether the 2-parameter Pareto distribution defined by (6) fits the data adequately by selecting Goodness-of-Fit Tests [1], [2], [9]. The first step is parameters estimation by maximum likelihood method [2], [8], [12]. The estimated parameters of the fitted distribution are shown in Table 1. In our parameters markers by (6) or (7) *est a* = 3410 and *est b* = 1.04777.

Table 1 Parameters of Fitted Distribution

Pareto (2-Parameter)
shape = 1.04777
lower threshold = 3410.0

Source: Output from Statgraphics Centurion XV

The Table 2 shows the results of test run to determine whether the most costly insured losses can be adequately fit by a 2-parameter Pareto distribution (6).

Since the smallest P-value = 0.858776 amongst the tests performed is greater than or equal to 0.05 we cannot reject the idea that losses comes from a 2-parameter Pareto distribution with 95% confidence.

Table 2 Results of Kolmogorov-Smirnov Test

	Pareto (2-Parameter)
DPLUS	0.0576431
DMINUS	0.0955203
DN	0.0955203
P-Value	0.858776

Source: Output from Statgraphics Centurion XV

We can also by Quantile plot and Quantile-Quantile or Q-Q plot assess visually how well the 2-parameter Pareto distribution with parameters (Table 1) fits the data.

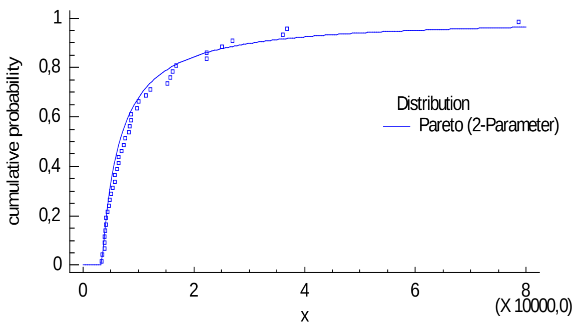

Fig.2 Quantile plot

Source: Output from Statgraphics Centurion XV

The Quantile Plot (Fig.2) shows the fraction of observations at or below x, together with the cumulative distribution function of the fitted distribution. To create the plot, the data are sorted from smallest to largest and plotted at the coordinates. Ideally, the points will lie close to the line for the fitted distribution, as is the case in the plot above.

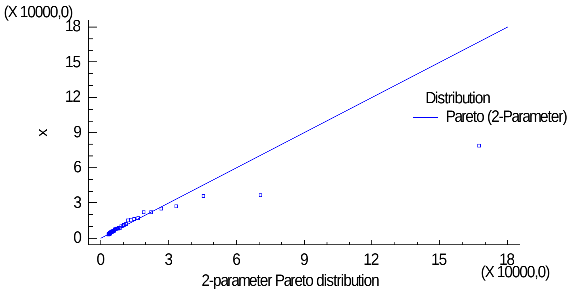

Fig.3 Quantile-Quantile plot

Source: Output from Statgraphics Centurion XV

The Quantile-Quantile plot (Fig.3) shows the fraction of observations at or below *x* plotted versus the equivalent percentiles of the fitted distribution. The fitted Pareto distribution has been used to define the *x*-axis. The fact that the points lie close to the diagonal line confirms the fact that the Pareto distribution provides good fit for the data.

Table 3 Quantiles of fitted Pareto distribution

Lower Tail Area (<=)	*Pareto (2-Parameter)*
0.50	6607.83
0.75	12804.5
0.90	30701.5
0.95	59492.8
0.99	276417

Source: Output from Statgraphics Centurion XV

The Table 3 contains the selected quantiles of Pareto distribution, which is well fitted model for the most costly insured catastrophe losses.

If will not change conditions of the occurrence of these events on the globe, will not change even their distribution. Then 50% of the most costly insurance losses will exceed 6607.83 million USD, 10% will exceed 30701.5 million USD, 1% will exceed 276417 million USD.

Knowing the probability model and its parameters, we can use quantile function (7) and by simulation procedure described in 2.1 we can find five the highest values at 40 most costly insurance losses.

Table 4 Process of simulation $Q(u)$

v	n	v^1/n	u	$Q(u)$
0.13549	40	0.964494	0.951257	60956.83
0.33132	39	0.972073	0.924691	40243.93
0.25384	38	0.971348	0.891923	28507.97
0.99347	37	0.970585	0.891765	28468.24
0.18092	36	0.969781	0.850404	20903.36

Source: Own calculation by (4) and (5)

The steps of simulation presents Table 4 and possible the highest five values (in million USD) in the world natural catastrophes we can find in the last column denoted as Q(u). So the highest simulated loss is 82 421.36 million USD, the second highest is 48123.9 million USD etc.

Two last columns in Table 5 show the boundaries for each order statistic. For example the highest possible insured loss is with probability 0.95 from 24 991.87 million USD to 18 066 831.58 million USD and 0.5% of losses may even exceed the value of 18 066 831.58 million USD.

Visualized results of the simulation process we can see at Fig.4.

Table 5 Quantiles of selected order statistics

Q(BETAINV(0.5))	*Q(BETAINV(0.995))*	*Q(BETAINV(0.005))*
164 921.29	18 066 831.58	24 991.87
70 901.33	993 661.54	18 346.01
45 453.34	318 235.24	15 002.34
33 581.25	163 674.11	12 884.65
26 690.92	103 499.70	11 390.54

Source: Own calculation by (3)

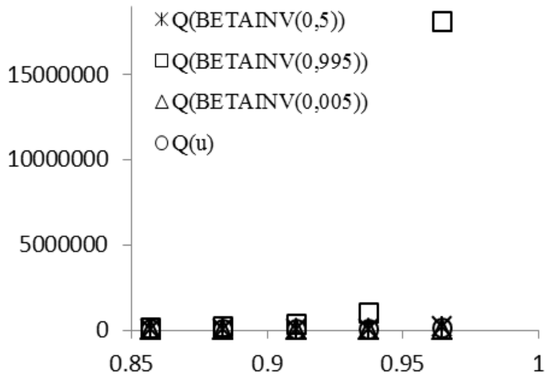

Fig.4 Graphical result of simulation of 5 the largest most costly insurance losses

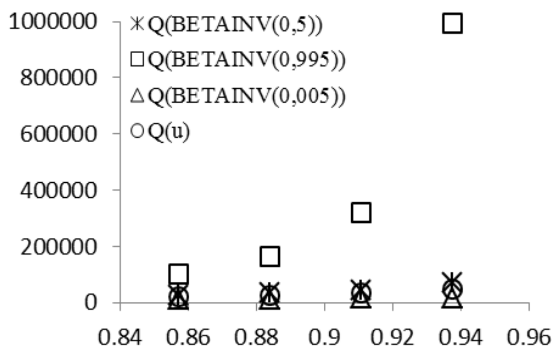

Fig.5 Graphical result of simulation of 4 the largest most costly insurance losses

4 Conclusion

The results of the analysis based on data of extreme insured losses in the world natural catastrophes in time period 1970-2014 are alarming.

Are justified concerns that the capacity of the world's insurance and reinsurance markets in the future will not be sufficient to cover these risks. It is high time for humanity to start emphatically remove the causes of the occurrence of catastrophes and their consequences.

References:

[1] W. G. Gilchrist, *Statistical Modelling with Quantile Functions*, Chapman & Hall/CRC, London 2000.

[2] R. J. Gray, S. M. Pitts, *Risk Modelling in General Insurance*. Cambridge University Press, 2012, ch. 2.

[3] P. Jindrová, Ľ. Sipková, Statistical Tools for Modeling Claim Severity. In: *European Financial Systems 2014*. Proceedings of the 11th International Scientific Conference. Lednice, June 12-13, 2014. Brno: Masaryk University, 2014, pp. 288-294.

[4] A. J. McNeil, A., *Estimating the Tails of Loss Severity Distributions using Extreme Value Theory*. ETH Zentrum, Zürich 1996. [online]. Available on: https://www.casact.org/library/astin/vol27no1/117.pdf

[5] V. Pacáková, B. Linda, Simulations of Extreme Losses in Non-Life Insurance. *E+M Economics and Management*, Volume XII, 4/2009.

[6] V. Pacáková, L. Kubec, Modelling of catastrophic losses. *Scientific Papers of the University of Pardubice*, Series D, Vol. XIX, No. 25 (3/2012), pp 125-134.

[7] V. Pacáková, J. Gogola, Pareto Distribution in Insurance and Reinsurance. Conference proceedings from 9th international scientific conference *Financial Management of Firms and Financial Institutions*, VŠB Ostrava, 2013. pp. 298-306.

[8] V. Pacáková, D. Brebera, Loss Distributions in Insurance Risk management. In: *Recent Advances on Economics and Business Administration*, Proceedings of the International Conference on E conomics and Statistics (ES 2015), INASE Conference, Vienna, March 15-17, 2015. pp. 17-22.

[9] V. Pacáková, D. Brebera, Loss Distributions and Simulations in General Insurance and Reinsurance. *International Journal of Mathematics and Computers in Simulation*. NAUN, Volume 9, 2015, pp. 159-167.

[10] Probability Distributions, On-line Manuals, StatPoint, Inc., 2005.

[11] Natural catastrophes and man-made disasters in 2014, *SIGMA* No2/2015, Swiss Re [online]. Available on: http://media.swissre.com/documents/sigma2_2015_en_final.pdf

[12] Y. K. Tse, *Nonlife Actuarial Models*. Cambridge: University Press, 2009.

[13] Swiss Re Economic Research & Consulting and Cat Perils. http://www.swissre.com/sigma/