Forecasting Extremes

Valéria Skřivánková [1]

Abstract

The paper deals with a special type of extreme values – with records. For analyzing and forecasting record values, the distribution of the maxima, the limit distribution of the number of records and the expected record time are studied. For the longest run of records interval estimation is derived. The considered methods to real non-life insurance data are applied.

Keywords

Extreme values, records, record times, return period, limit theorems, risk management, non-life insurance data

1 Introduction

Most of publications about Extreme Value Theory (EVT) deal mainly with two basic methods of registration of extreme values: Block Maxima and Peaks-over-Threshold. The block maxima method is used for example in hydrology, by weather forecasting and so on. The peaks-over-threshold method is an important tool for analyzing and forecasting extreme financial and insurance data. For risk management in these areas it is necessary to determine correctly the special level (threshold) above or below which the ruin probability will be too high (see [3], [4], [6], [7], [8]).

Some insurance portfolios seem to have a tendency to occasionally include large claims that can jeopardize the solvency of the company (major accidents such as earthquakes, floods, hurricanes, windstorms, airplane accidents and so on). An insurance company will always safeguard itself against portfolio contamination caused by extreme claims at least by excess-of-loss reinsurance contract where the reinsurer pays for the claim over a given retention.

Financial time-series consist of speculative prices of assets such as stocks, foreign currencies or commodities. Risk management at a commercial bank is intended to guard against risk of loss due to fall in prices of financial assets held or issued by the bank. The Value-at-Risk (VaR) of a portfolio is essentially the level below which the future portfolio will drop with only a small probability. VaR is one of the important risk measures that have been used by investors to predict the impact of unfavorable random events.

In this paper we will concentrate on records as a special type of extreme values. First we explain: What is a record in the context of extreme value theory?

2 Basic terms and properties

Consider a sequence of independent identically distributed (iid) random variables $X_1, X_2, \ldots, X_n, \ldots$ with continuous distribution function F.

[1] Doc. RNDr. Valéria Skřivánková, CSc., Institute of Mathematical Science, P. J. Šafarik University in Košice, Jesenná 5., 041 54 Košice, SR ; e-mail : valeria.skrivankova@upjs.sk

Denote the maximum of these random variables as

$$M_1 = X_1 ,$$
$$M_n = \max \{ X_1, X_2, \ldots , X_n \}, \text{ for } n \geq 2 . \qquad (1)$$

A record occurs when there is a jump in the sequence { Mn }, more precisely Xn is a record if and only if

$$X_n > M_{n-1} , \quad \text{where} \quad M_{n-1} = \max\{ X_1, X_2, \ldots , X_{n-1} \}. \qquad (2)$$

By definition we take $X1$ as a record.

The times $T_1 < T_2 < \ldots$ when the jumps in { M_n}occur at random are called **record times** of {X_n}. Define

$$T_1 = 1 ,$$
$$T_n = \min\{k > T_{n-1} : X_k > M_{k-1}\} \qquad (3)$$

Further define the **record counting process** { R_n}, n = 1,2, 3, ...(the total number of records among $X_1, X_2, \ldots, X_n$) as

$$R_1 = 1,$$
$$R_n = 1 + \sum_{k=2}^{n} I_k , \qquad (4)$$

where I_k is the **indicator** of the event { $X_k > M_{k-1}$ }, so that $I_k = 1$ if X_k is a record and $I_k = 0$ if X_k is not a record.

Under the condition that all X_i, i = 1, 2, ..., n are different, we can order these random variables in n! ways and get (n -1)! possible favorable cases for the first n – 1 random variables . So for the probability that X_n is a record we get

$$p_{\acute{n}} = P(X_n \text{ is a record}) = \frac{(n-1)!}{n!} = \frac{1}{n} , \quad \text{for } n = 1, 2, 3, \ldots \qquad (5)$$

3 How many records do we expect?

We denoted the total number of records in a sequence $\{X_1, X_2, \ldots, X_n\}$ as R_n . The expected number of records we get using (5)

$$E(R_n) = \sum_{k=1}^{n} I_k \cdot p_k = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \cdot \qquad (6)$$

Because the indicators I_k , $2 \leq k \leq n$ are independent random variables (for proof see [1], p.88), we can easily show that the variance of the total number of records is

$$D(R_n) = D(1 + \sum_{k=2}^{n} I_k) = \sum_{k=2}^{n} D(I_k) = \sum_{k=2}^{n} \left(\frac{1}{k} - \frac{1}{k^2} \right) \cdot \qquad (7)$$

Remark.

To calculate the expected value E(R_n) for large n in (6), we usually estimate the sum on the right hand side

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \varepsilon + \ln n , \qquad (8)$$

where $\varepsilon \cong 0.57721$ is the well known Euler constant (see [1]). The following table shows that the expected number of records in a sequence of random variables grows very slowly.

$n = 10^k$, $k =$	1	2	3	4	5	6	7	8	9
$E(R_n)$	2.9	5.2	7.5	9.8	12.1	14.4	16.7	19.0	21.3
$\sqrt{D(R_n)}$	1.2	1.9	2.4	2.8	3.2	3.6	3.9	4.2	4.4

Table1: Mean and standard deviation of the total number of records in the sequence $\{X_1 X_2, ..., X_n\}$

4 Limit theorems for records

To characterize the limit distribution and the speed of the convergence we formulate the central limit theorem (CLT) and the strong law of large numbers (SLLN) for records ([3],[8]).

Theorem 1. (limit results for the frequency of records)

Let $\{X_n\}$ be a sequence of iid random variables with continuous distribution function F. Let R_n be the total number of records among $X_1, X_2, ..., X_n$. Then for $n \to \infty$ hold

CLT : $$\frac{R_n - \ln n}{\sqrt{\ln n}} \xrightarrow{d} N(0,1) \qquad (9)$$

(it means that the random variable on the left hand side in (9) converges in distribution and has a N(0,1) limit distribution), and

SLLN : $$\frac{R_n}{\ln n} \to 1 \quad \text{with probability 1.} \qquad (10)$$

Theorem 2.(limit results for the growth of records)

Let $\{X_n\}$ be a sequence of iid random variables with continuous distribution function F and with record times $\{T_n\}$. Then for $n \to \infty$ hold

CLT : $$\frac{\ln T_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1) \ , \qquad (11)$$

SLLN : $$\frac{\ln T_n}{n} \to 1 \quad \text{with probability 1.} \qquad (12)$$

Remark.

In summary, by (10) the number of records is roughly of the order ln n. So records become more and more unlikely for large n.

On the other hand , by (12) the record times T_n grow roughly exponentially (like $\exp\{n\}$) and thus the period between two successive records becomes bigger and bigger.

5 The return period

Consider a sequence of iid random variables $\{X_n\}$ with continuous distribution function F and Bernoulli trials in which the probability of success will be defined as

$$p = P(X_n \text{ is a record}) \ , n = 1, 2, 3, ...$$

Let the waiting time $W(u)$ for the next record be discrete and let it can be represented by the time of the first accidence of the threshold $u = M_{n-1}$. Then $W(u)$ is a geometric distributed random variable and the probability of the waiting time is

$$P(W(u) = k) = p.(1-p)^k \ , k = 1,2,3, ... \qquad (13)$$

The expected waiting time $W(u)$ is called the **return period** and we can determine it as

$$E(W(u)) = \sum_{k=1}^{\infty} k \cdot p(1-p)^k = p(1-p) \cdot \sum_{k=1}^{\infty} k(1-p)^k = \frac{1}{p} \cdot \quad (14)$$

Remark.

Relation (14) says that the expected waiting time for the following record and the expected waiting time between two records is p^{-1}. If the number of records has Poisson distribution with parameter λ, the waiting time is exponential with parameter $\delta = 1/\lambda$.

The longest success run. Denote the length of the longest success-run (run of records) in a sequence of iid random variables $\{X_1, X_2, ..., X_n\}$as Z_n and the probability of success as p

$$P(X=1) = p, \quad p \in (0,1),$$
$$P(X=0) = q = 1-p.$$

A run of 1s of length j in $\{X_1, X_2, ..., X_n\}$ is defined as a subsequence $\{X_{i+1}, X_{i+2}, ..., X_{i+j}\}$ such that

$$Xi = 0, \quad X_{i+1} = X_{i+2} = ... = X_{i+j} = 1, \quad X_{i+j+1} = 0,$$

and formally set $X_0 = 0$, $X_{n+1} = 0$.

The question is: *How long is the longest run of 1s in* $X_1, X_2, ..., X_n$ *?*

Rényi in [4] proved a result on the almost sure growth in the following limit theorem.

Theorem 3.

For every fixed $p \in (0,1)$ with probability 1 holds

$$\lim_{n\to\infty} \frac{Z_n}{\ln n} = -\frac{1}{\ln p} \cdot \quad (15)$$

Theorem 4.

For every fixed $\varepsilon > 0$ and $r \in N$ with probability 1 the length of the longest run of 1s in $X_1, X_2, ..., X_n$ falls into the interval $< \alpha_n, \beta_n >$, where

$$\alpha_{\acute{n}} = \left[\frac{\ln(nq) - \ln_3(nq) - \varepsilon}{-\ln p} \right] - 1 , \quad (16)$$

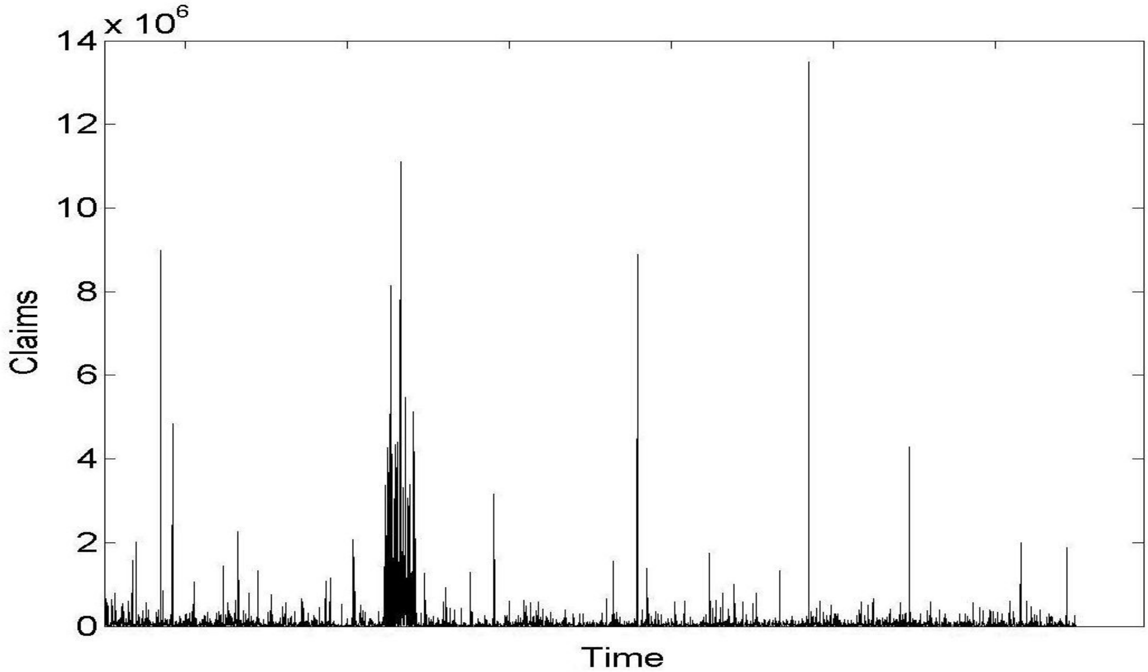
$$\beta_n = \left[\frac{\ln_1(nq) + ... + \ln_r(nq) + \varepsilon \ln_r(nq)}{-\ln p} \right] , \quad (17)$$

and $[x]$ denote the integer part of x, $\ln_0(x) = x$, $\ln_1(x) = \max(0, \ln x)$, $\ln_k(x) = \max(0, \ln_{k-1}(x))$.

6 Real data analysis

Using the above considered properties of records, we analyzed 3742 real non-life insurance data (over time interval 3.1.2006 – 4.12.2007). Figure 1 gives the visualization of the time series. The basic statistical characteristics are given in the Table 2.

Figure 1: Graph of the time series



Date of occurance	Record value
3.1.2006	56 800
4.1.2006	93 801
6.1.2006	556 381
8.2.2006	662 345
17.2.2006	798 733
31.3.2006	1 567 400
20.4.2006	2 021 825
6.7.2006	8 995 000
18.1.2007	11 099 861
10.7.2007	13 500 010

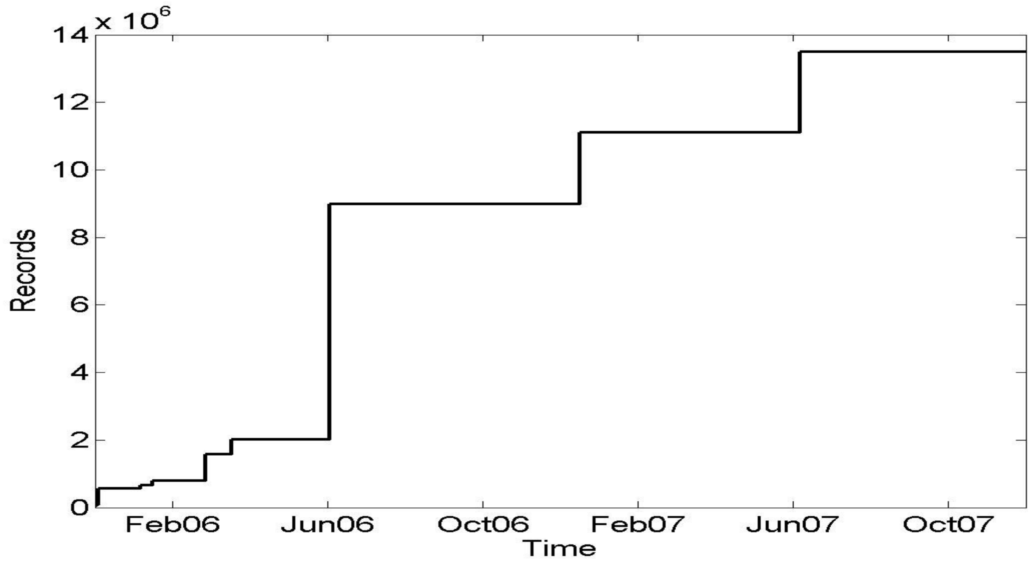
Table 3: Records

We found 10 record claims in the analyzed time series. The record times and the record values are given in the Table 3. The graphical representation of the records is situated on the Figure 2.

Count	3 742
Median	24 413
Mean	114 253
Variance	$2{,}78 \cdot 10^{11}$
Standard deviation	526 985,4
Coeff. of variation	4,61
Min	509
Max	13 500 010
Range	13 499 501

Table 2: Statistical characteristics

Figure 2: Records



We calculate now the expected number and the variance of records and determined the length of the longest success-run.:

$$E(R_n) = \sum_{k=1}^{3742} \frac{1}{k} = 8{,}8, \qquad D(R_n) = \sum_{k=1}^{3742} \left(\frac{1}{k} - \frac{1}{k^2} \right) = 7{,}2$$

To calculate the approximately length of the longest run we need to determine the probability of success. In our case $p = 10/3742 \cong 0{,}003$. From here using (15) we get

$$Z_n = -\frac{\ln 3742}{\ln 0{,}003} \approx 1{,}4.$$

Detailed statistical analysis of considered non-life insurance data can be found in [8].

7 Conclusion

Managing risk of the portfolio with extreme values including records requires knowledge about the probability distribution or at least their statistical characteristics. The paper gives some results for the expected number of records, record times, limit results for the frequency and the growth of records and interval estimation for the length of the longest run of records. The considered methods are applied analyzing real non-life insurance iid data. The new trends in extreme value theory admit dependence of data and study the measure of this dependence.

Acknowledgement : This work was supported by Slovak grant foundation under the grants 1/0131/09 and 1/0724/08.

References

[1] ANDĚL J.: *Matematika náhody.* Matfyzpress, Praha, 2000.

[2] BEIRLANT J.at al.: *Statistics of Extremes.* Wiley, New York, 2004.

[3] EMBRECHTS P. at al: *Modelling extremal events.* Springer, Berlin, 1997.

[4] HORÁKOVÁ G., KOZUBÍK A.: *Calculation of the expected claims for the catastrophic events.* J. Information, Control and Management Systems, Žilina,Vol.3, No.1(2005),9-16.

[5] RÉNYI A.: *Probability theory.* Akadémia Kiadó, Budapest, 1970.

[6] SKŘIVÁNKOVÁ V., TARTAĽOVÁ A.: *Catastrophic risk management in non-life insurance.* E+M Economics and Management, 2/2008, 65-72.

[7] URBANÍKOVÁ M.: *Probability of ruin.* In: Applied Mathematics and Informatics at Universities, Bratislava, 2002, 87-91.

[8] ZIMANOVÁ J.: *Štatistická analýza extrémnych hodnôt.* Diplomová práca na PF UPJŠ, Košice, 2008.

Summary

Predpoveď extrémnych hodnôt

Príspevok sa zaoberá so špeciálnym typom extrémnych hodnôt – s rekordami. Kvôli analýze a prognóze rekordov sa študuje rozdelenie rekordu jako maxima , celkový počet rekordov a očakávaný rekordný čas. Výsledky sú aplikované pri spracovaní reálnych dát z oblasti neživotného poistenia.