Statistical estimation of large claim distributions

Autor(en): **Keller, Barbara / Klüppelberg, Claudia**

Objekttyp: **Article**

Zeitschrift: **Mitteilungen / Schweizerische Vereinigung der Versicherungsmathematiker = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries**

Band (Jahr): **- (1991)**

Heft 2

PDF erstellt am: **14.09.2024**

Persistenter Link: https://doi.org/10.5169/seals-967287

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.
Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.
Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch

Barbara Keller and Claudia Klüppelberg, Zürich

Statistical Estimation of Large Claim Distributions

1 Introduction

The specification of the probability distribution for the size of a single claim is traditionally one of the major modelling problems in non-life insurance. Here one often has to face situations where very large claims occur with high probability as for instance in connection with hurricanes, fire or earthquakes. Thus one is confronted with heavy tailed distributions like *Pareto*, *Lognormal*, *Loggamma* or *Weibull* where the fit of the tail is a major problem. A wealth of interesting cases can be found in *Hogg* and *Klugman* [1984]. Another cause of interest in large claim distributions stems from the reinsurer's point of view. One of the main reinsurance problems is the calculation of a risk adequate premium rate. For this reason one has to estimate the far end of the right tail of a distribution where only very few data are available. Obviously, the relevant information for the right tail of a distribution is only contained in the upper extreme part of the sample; for this reason extreme value theory provides a natural tool in order to estimate the far end of a distribution tail. Basically, there exist two different methods following this line of argument. The so-called threshold method goes back to an idea of *Pickands* [1975] and has been developed by *Smith* [1987] culminating in a paper by *Davison* and *Smith* [1990]. They propose a nonparametric method to fit the excesses over a threshold value for distributions in the maximum domain of attraction of an extreme value law. A parametric method has been proposed by *Klüppelberg* and *Villaseñor* [1990] who suggested so-called asymptotic maximum likelihood estimators (AMLE). AMLE's are based upon the k upper order statistics, so we say AMLE of order k (AMLE(k)).

For a sample $X_1, \ldots, X_n$ from a distribution function F we denote by

$$X_{1:n} \geq \ldots \geq X_{n:n}$$

its decreasing order statistics.

In the *extended Pareto* model the tail $\overline{F} = 1 - F$ has the representation

$$\overline{F}(x) = s(x)x^{-\alpha}, \quad x \geq x_0, \quad \alpha > 0,$$

where s is a slowly varying function; i.e.

$$\lim_{x\to\infty} \frac{s(xt)}{s(x)} = 1 \quad \forall t > 0.$$

In this case the AMLE (1) for α is

$$\hat{\alpha}_1 = \frac{\log n}{\log X_{1:n}}$$

and the AMLE(k) for $k \geq 2$ is nothing else than Hill's estimator

$$\hat{\alpha}_k = \left\{ \frac{1}{k} \sum_{j=1}^{k} \log X_{j:n} - \log X_{k:n} \right\}^{-1}.$$

In the *extended Weibull* model $\overline{F}$ has the representation

$$\overline{F}(x) = \exp\{-s(x)x^{\alpha}\}, \quad x \geq x_0, \alpha > 0,$$

where s is a slowly varying function. Then the AMLE(1) for α is

$$\hat{\alpha}_1 = \frac{\log\log n}{\log X_{1:n}}$$

and the AMLE(k) for $k \geq 2$ is

$$\hat{\alpha}_k = \frac{\log\log n}{\log V_{n,k} + \log\log n}$$

with

$$V_{n,k} = \frac{1}{k} \sum_{i=1}^{k} X_{i:n} - X_{k:n}.$$

Asymptotic properties of Hill's estimator have been investigated in detail; for instance *Häusler* and *Teugels* [1985] proved the asymptotic normality of $\hat{\alpha}_k^{-1}$ when the number k of upper order statistics tends to infinity appropriately with the sample size n. An analogous result is derived in *Klüppelberg* [1991] for the AMLE(k) in the extended Weibull model.

Unfortunately, the rate of convergence in both cases is rather slow; it obviously depends on the slowly varying function s and is in general logarithmic. This is in fact disturbing for practical applications where the sample sizes may not be very large. To investigate the small sample behaviour of the AMLE(k), $k \geq 1$, a Monte Carlo simulation was conducted in *Klüppelberg* and *Villaseñor* [1990], but only distribution tails $\overline{F}(\mathrm{x}) = x^{-\alpha}$, $x \geq 1$, and $\overline{F}(x) = \exp\{-x^{\alpha}\}$, $x \geq 0$, were considered.

The aim of this paper is to investigate in more detail the influence of the slowly varying function s on the AMLE (k). For this reason one has to simulate the upper k order statistics of extended Pareto and extended Weibull distributions. Basically, one meets two different kinds of problems. First of all we are mainly interested in the k upper order statistics rather than the whole sample. Secondly, we are faced with simulation methodology for heavy tailed distributions in cases where the usual inversion method or a simple rejection method do not apply.

Our paper is organized as follows: In the next section we describe the simulation method which deals with the two problems mentioned above. Here we would like to thank Richard Smith for his helpful comments on these problems. In section 3 we present the simulation results showing the small sample properties of AMLE's for tail estimation. In section 4 we use our simulation results for an indication of the influence of the slowly varying function s to different reinsurance treaties.

2 The simulation method

To investigate properties of the AMLE's by Monte Carlo simulation one must keep in mind that for these estimators one only needs the upper k order statistics, and these upper order statistics should come from the far end of the tail. Hence it would be inefficient to generate first a whole sample and sort it only to use finally the upper few percent of the sample. Methods for the generation of an ordered sample have been considered by several authors, see e.g. *Gerontides* and *Smith* [1982] and references therein.

However, it would be convenient to simulate directly only the few upper order statistics needed for the AMLE's. This can be done as follows: We fix a high threshold x_T such that $F(x_T)$ is for instance approximately 0.90, 0.95 or 0.99 ensuring that the generated random numbers come from the last 10, 5 or 1 percent of our distribution. Assume for the moment we want to generate a sample $(X_1, \ldots, X_n)$ from F; then the event $X_i > x_T$ would be a success with probability $P(X_i > x_T) = \overline{F}(x_T)$ and the event $X_i \leq x_T$ would be a failure with probability $P(X_i \leq x_T) = F(x_T)$. Thus the number K of successes is a binomial variable with parameters n and $p=\overline{F}(x_T)$. Now let n be fixed and $y > 0$, then for k variables, say $X_1, \dots, X_k$ we have

$$
\begin{aligned}
&P[X_1 \le x_T + y, \dots, X_k \le x_T + y] \\
&\quad = P[X_1 \le x_T + y, \dots, X_k \le x_T + y | X_1 \le x_T, \dots, X_k \le x_T] \\
&\qquad \cdot P[X_1 \le x_T, \dots, X_k \le x_T] \\
&\qquad + P[X_1 \le x_T + y, \dots, X_k \le x_T + y | X_1 > x_T, \dots, X_k > x_T] \\
&\qquad \cdot P[X_1 > x_T, \dots, X_k > x_T] \\
&\quad = 1 \cdot P[X_1 \le x_T, \dots, X_k \le x_T] \\
&\qquad + \frac{P[X_1 \in (x_T, x_T + y], \dots, X_k \in (x_T, x_T + y]]}{P[X_1 > x_T, \dots, X_k > x_T]} \\
&\qquad \cdot P[X_1 > x_T, \dots, X_k > x_T].
\end{aligned}
$$

Furthermore,

$$
P[X_1 > x_T, \dots, X_k > x_T] = \binom{n}{k} p^k (1-p)^{n-k}
$$

and

$$
\frac{P[X_1 \in (x_T + y), \dots, X_k \in (x_T + y)]}{P[X_1 > x_T, \dots, X_k > x_T]} = \Big(\frac{F(x_T + y) - F(x_T)}{1 - F(x_T)}\Big)^k,
$$

where

$$
F_T(y) := \frac{F(x_T + y) - F(x_T)}{1 - F(x_T)}
$$

defines a distribution function on $(0, \infty)$.
Thus we shall simulate random variables $> x_T$ as follows:

- Fix a high threshold x_T and generate a random variable k from a binomial distribution with parameters n and $p=\overline{F}(x_T)$.
- Generate k independent random variables with distribution F_T.

This method has also the advantage that for x_T large enough the density f_T of F_T is monotone or even convex on $(0, \infty)$ although the original density of F can exhibit fluctuations caused by the slowly varying function s. This is due to the fact that we only consider so-called normalized slowly varying functions s and in this case the function $s(x)x^\varrho$ is asymptotically monotone for each $\varrho \ne 0$. Considering only normalized slowly varying function is not a serious restriction, since any slowly varying function is asymptotic to a normalized one (see *Bingham, Goldie, Teugels* [1987]).

A common method to simulate arbitrary densities is the so-called rejection method (see e.g. *Morgan* [1984], p. 98ff). Unfortunately, no appropriate enveloppe exists in case of our examples. All we know is that for large x_T our densities $f_T(y)$ are positive and monotone for all $y > 0$. So we decided to use a combination of inversion and rejection determining the inversion intervals by a *Newton-Raphson* iteration as proposed by *Devroye* [1984].
In a first step we divide the interval $[0, \infty)$ into subintervals $x_0 < x_1 < \ldots$, defining

$$x_{n+1} = x_n + \frac{\overline{F}_T(x_n)}{f_T(x_n)}, \quad n \in \mathbb{N}_0, \ x_0 = 0.$$

Note that the sequence $\{x_n\}_{n \in \mathbb{N}_0}$ is the sequence iterated by the *Newton-Raphson* method to solve iteratively the equation $F_T(x) = 1$. Since F_T has unbounded support and f_T is positive on $(0, \infty)$, the sequence $\{x_n\}_{n \in \mathbb{N}_0}$, increases to ∞ as $n \to \infty$. Thus for any $u \in (0, 1)$, the solution of $F(x) = u$ certainly belongs to one of the intervals $[x_n, x_{n+1})$ for $n \in \mathbb{N}_0$.
In a second step, the random numbers with density f_T are generated by a rejection method. Suppose the uniform random number u is such that the solution $F(x) = u$ falls into the interval $[x_n, x_{n+1})$ for some $n \in \mathbb{N}_0$. The corresponding random number x is then generated by a rejection method on the interval $[x_n, x_{n+1})$ where the enveloppe $e_n(x)$ is chosen to be a rectangle on $[x_n, x_{n+1})$, where $e_n(x) = f_T(x_n)$ for all $x \in [x_n, x_{n+1})$.
Thus our simulation method is a rejection method where the enveloppe $e(x)$ is a step function with representation $e(x) = \sum_{n=0}^{\infty} f_T(x_n) I_{[x_n, x_{n+1})}(x)$.

3 The simulation result

In order to study the influence of the slowly varying function s on the AMLE's we let s depend on a second parameter β and considered for the same α-value different β-values.
As extended Pareto distribution we chose

$$\overline{F}(x) = (\log x)^{\beta} x^{-\alpha}, \quad x \geq x_0, \ \alpha > 0, \ \beta \in \mathbb{R}.$$

For the range of α we took $\alpha = 1.5(0.5)4.0$ and for β we considered the values $\beta = -2.0(0.5)2.0$.

As extended Weibull distribution we simulated

$$\overline{F}(x) = x^{\beta} e^{-x^{\alpha}}, \quad x \geq x_0,\ \alpha \in (0,1],\ \beta \in \mathbb{R},$$

including the extended exponential distribution as a limit for heavy tailed distributions. As α-values in our study we took $\alpha = 0.1(0.3)1.0$ and for β again the range $\beta = -2.0(0.5)2.0$.

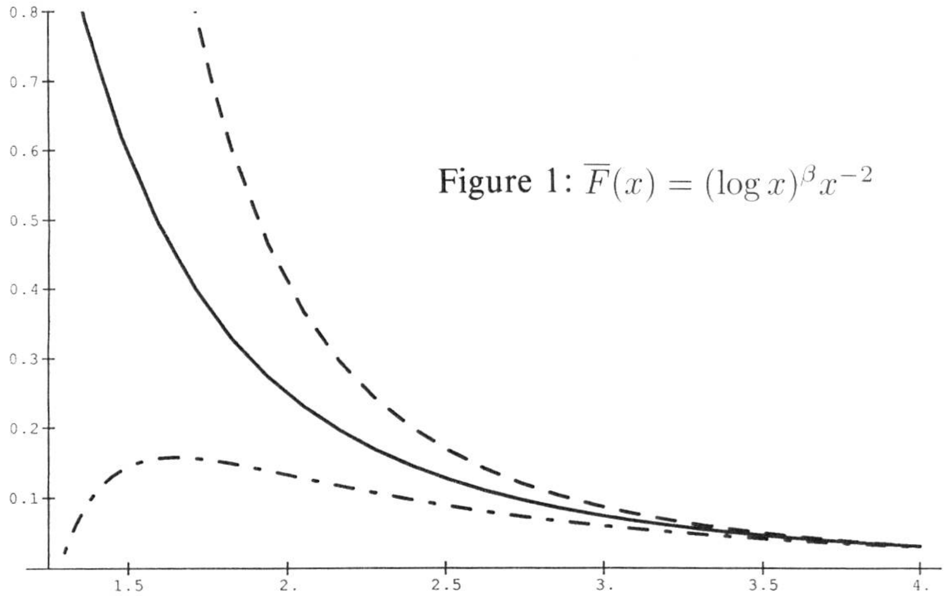

Figure 1: $\overline{F}(x) = (\log x)^{\beta} x^{-2}$

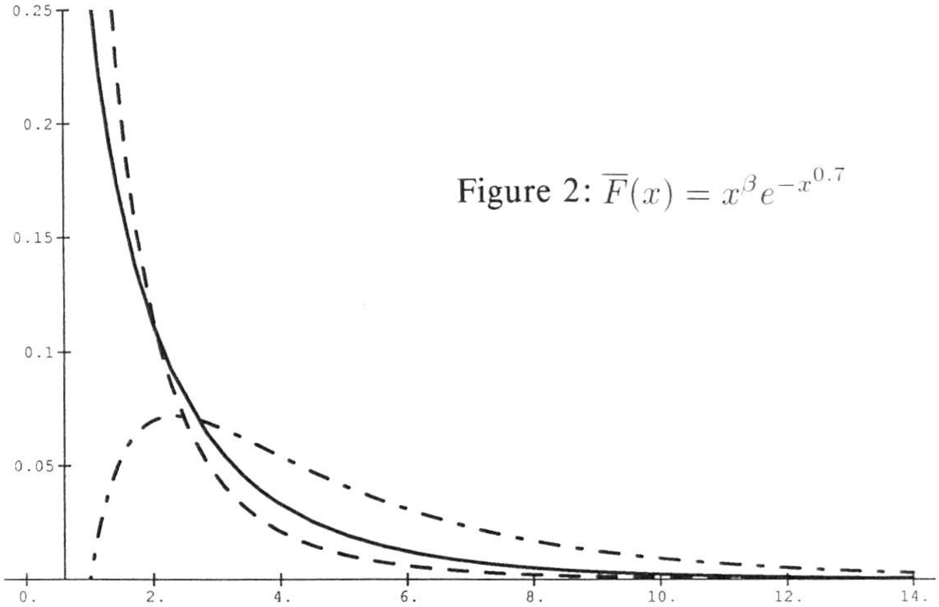

Figure 2: $\overline{F}(x) = x^{\beta} e^{-x^{0.7}}$

To give an idea of how the slowly varying functions affect a pure Pareto or Weibull distribution we give examples of the densities in Figure 1 and 2. In both figures, a plain line denotes $\beta = 0$ (i.e. a pure Pareto or, respectively, Weibull density), a dashed dotted line is plotted in case $\beta = 0.5$ an a dashed line marks $\beta = -0.5$. Whereas in a neighborhood of the left endpoint of the distribution the value of β even changes the shape of the distribution, its influence vanishes going to the right tail. Nevertheless, depending on the size n of the sample and especially on the number of observations falling into the last 10, 5 or 1 percent of the distribution the value of β also can have a certain effect on the AMLE's for α. We shall discuss this point in more details later.
As sample sizes we have considered 1000, 2000, 5000 and 10000. Obviously, the AMLE's are better for larger samples since the influence of β decreases the farther to the right the upper order statistics move. In this paper, however, we want to stress their small sample properties and hence we present here only the results for a sample size $n = 1000$. Remember that from this size n of the complete sample we only consider the upper few extremes to estimate the parameter α. For purpose of illustration we have limited ourselves in this paper in the *Pareto* case to $\alpha = 2$ and $\alpha = 4$ and in the *Weibull* case to $\alpha = 0.4$ and $\alpha = 0.7$. In both cases we considered $\beta = \pm 0.5$ (the case $\beta = 0$ can be found in *Klüppelberg* and *Villaseñor* [1990]). For more detailed results of our simulation study we refer to *Keller* [1991].
As explained in section 2 we chose a threshold x_T such that all random variables determining the AMLE come from the far end of the distribution tail. It turned out that for a sample size of $n = 1000$ at least 5 percent of the distribution should be considered. Thus we chose for x_T two different values, namely x_T such that $F(x_T) = 0.9$ and $F(x_T) = 0.95$. Then the number k is the realization of a binomial random variable which caused the specific values for k. In the following tables we give the simulated values of $\hat{\alpha}_k$. Note that for each combination of α and β we have simulated 50 runs and we also have calculated the sample variance (SV) and the sample mean square error ($SMSE$).

Table 1: Extended Pareto distribution

$\overline{F}(x) = (\log x)^{\beta} x^{-\alpha}, \quad \alpha = 4$								
		$\beta = -0.5$				$\beta = +0.5$		
		$\hat{\alpha}$	SV	$SMSE$		$\hat{\alpha}$	SV	$SMSE$
$F(x_T) \approx 0.9$	$k = 70$	4.83868	0.38731	0.93151	$k = 45$	3.74773	0.24786	0.34633
$F(x_T) \approx 0.95$	$k = 38$	4.70853	0.47951	0.84582	$k = 25$	3.96149	0.57748	0.57344
$\overline{F}(x) = (\log x)^{\beta} x^{-\alpha}, \quad \alpha = 2$								
		$\beta = -0.5$				$\beta = +0.5$		
		$\hat{\alpha}$	SV	$SMSE$		$\hat{\alpha}$	SV	$SMSE$
$F(x_T) \approx 0.9$	$k = 62$	2.47625	0.08420	0.26579	$k = 79$	1.83236	0.03806	0.07850
$F(x_T) \approx 0.95$	$k = 30$	2.45991	0.19067	0.35643	$k = 21$	2.00963	0.18435	0.17989

Table 2: Extended Weibull distribution

$\overline{F}(x) = x^{\beta} e^{-x^{\alpha}}, \quad \alpha = 0.4$								
		$\beta = -0.5$				$\beta = +0.5$		
		$\hat{\alpha}$	SV	$SMSE$		$\hat{\alpha}$	SV	$SMSE$
$F(x_T) \approx 0.9$	$k = 113$	0.53337	0.00069	0.01592	$k = 111$	0.34936	0.00011	0.00345
$F(x_T) \approx 0.95$	$k = 42$	0.49236	0.00119	0.00801	$k = 32$	0.33619	0.00020	0.00523
$\overline{F}(x) = x^{\beta} e^{-x^{\alpha}}, \quad \alpha = 0.7$								
		$\beta = -0.5$				$\beta = 0.5$		
		$\hat{\alpha}$	SV	$SMSE$		$\hat{\alpha}$	SV	$SMSE$
$F(x_T) \approx 0.9$	$k = 91$	0.80246	0.00251	0.00982	$k = 56$	0.64046	0.00096	0.00612
$F(x_T) \approx 0.95$	$k = 27$	0.76395	0.00496	0.00764	$k = 36$	0.63214	0.00129	0.00771

We conclude this section with a brief discussion of our simulation results:

- Whereas for an undisturbed Pareto and Weibull distribution $\hat{\alpha}_1$ was surprisingly good (see *Klüppelberg* and *Villaseñor* [1990]), in the extended case it may not be appropriate even for larger samples.
- The quality of $\hat{\alpha}_k$, $k \geq 2$, depends rather on the proportion of α and β than on their absolute values: the smaller α is with respect to $|\beta|$, the worse is the effect of β on the estimate of α, an effect which can be smoothened in choosing a higher threshold x_T.
- On the other hand, if α is rather dominant, as e.g. for $\alpha = 0.7$ in the *Weibull* case, then $\hat{\alpha}$ is not seriously affected by the function s, in our case by the value of β.
- A negative β may have a worse effect on $\hat{\alpha}_k$ than a positive β.
- The sample variance SV is in general larger for a higher threshold. This is caused by the fact that the number k is smaller.
- In general, the estimate of α can be improved by choosing a higher threshold x_T. But then the probability of obtaining random numbers $> x_T$ becomes very small, so the sample size n would have to be increased.

4 An application to reinsurance

A general reinsurance treaty can be represented as a function of the order statistics of the sample of claimsizes; for more details see *Kremer* [1985] and the references therein. We restrict ourselves here to three examples:

- the largest claims reinsurance

$$LCR = \sum_{i=1}^{k} X_{i:n}, \quad k \in \{1, \ldots, n\},$$

where the k largest claims are covered,
- the excess of loss with priority P,

$$XLR = \sum_{i=1}^{n} (X_{i:n} - P)^+$$

where the excess over a fixed threshold P is covered,

- the $ECOMOR$

$$ECOMOR = \sum_{i=1}^{k}(X_{i:n} - X_{k:n}), \quad k \in \{1, \ldots, n\}$$

where the excess over the k-largest claim is covered.

We use the simulated upper order statistics to demonstrate how minor changes of the tail of the claimsize distribution can affect substantially the reinsurance sums of the different treaties. Again we consider extended Pareto and extended Weibull claims. As priority P in the XLR treaty we took the threshold value x_T. Since k is a binomial random number, depending on the threshold value x_T, for each simulation run we obtain a different k. To make our simulations results comparable, in the following table 3 we give the simulated values of the normalized sums $LCR' = \frac{1}{k}LCR$, $XLR' = \frac{1}{k}XLR$ and $ECOMOR' = \frac{1}{k}ECOMOR$ (note that for XLR we took k as the number of positive summands). The first line in the table refers to a threshold value of $x_T = 0.9$ and the second to $x_T = 0.95$ respectively. Again for each value we have simulated 50 runs and we also have calculated the sample variance which is given in brackets.

Table 3: Reinsurance treaties

	$\overline{F}(x) = (\log x)^{\beta} x^{-\alpha}, \quad \alpha = 4$					
	$\beta = -0.5$		$\beta = 0$		$\beta = +0.5$	
LCR'	2.60887	(0.01092)	2.54804	(0.00798)	2.98632	(0.02064)
	3.28768	(0.03600)	3.28216	(0.04326)	3.27456	(0.07521)
XLR'	0.57225	(0.01092)	0.63282	(0.00798)	0.84563	(0.02064)
	0.74408	(0.03600)	0.81562	(0.04326)	0.91578	(0.07521)
$ECOMOR'$	0.56525	(0.01078)	0.62685	(0.00800)	0.83425	(0.01996)
	0.72893	(0.03540)	0.78329	(0.04084)	0.88720	(0.07726)
	$\overline{F}(x) = (\log x)^{\beta} x^{-\alpha}, \quad \alpha = 2$					
	$\beta = -0.5$		$\beta = 0$		$\beta = +0.5$	
LCR'	6.41125	(0.51304)	8.13463	(0.57903)	9.95120	(1.84292)
	9.56956	(2.68687)	11.92534	(4.35227)	17.59069	(38.72920)
XLR'	2.81069	(0.51304)	3.99429	(0.57903)	5.48002	(1.84292)
	4.33282	(2.68687)	5.52946	(4.35227)	9.72377	(38.72920)
$ECOMOR'$	2.78566	(0.50882)	3.95658	(0.57497)	5.42160	(1.86342)
	4.29479	(2.66879)	5.37459	(4.27340)	9.54652	(38.72069)
	$\overline{F}(x) = x^{\beta} e^{-x^{\alpha}}, \quad \alpha = 0.7$					
	$\beta = -0.5$		$\beta = 0$		$\beta = +0.5$	
LCR'	4.36832	(0.03420)	6.26720	(0.07316)	10.60423	(0.21219)
	6.36357	(0.08872)	8.90124	(0.29695)	11.29170	(0.29102)
XLR'	1.67259	(0.03420)	2.45763	(0.07316)	3.37025	(0.21219)
	2.05822	(0.08872)	2.67477	(0.29695)	3.42321	(0.29102)
$ECOMOR'$	1.65454	(0.03321)	2.42113	(0.06892)	3.32042	(0.21841)
	2.01071	(0.08988)	2.59020	(0.26120)	3.35104	(0.27059)
	$\overline{F}(x) = x^{\beta} e^{-x^{\alpha}}, \quad \alpha = 0.4$					
	$\beta = -0.5$		$\beta = 0$		$\beta = +0.5$	
LCR'	9.41175	(0.70861)	36.95900	(12.80458)	75.13442	(41.49711)
	16.25182	(4.65231)	28.34684	(16.46001)	122.95194	(100.50392)
XLR'	5.80624	(0.70861)	21.12145	(12.80458)	41.73969	(41.49711)
	9.11398	(4.65231)	21.39967	(16.46001)	52.24463	(100.50392)
$ECOMOR'$	5.78060	(0.71422)	20.71072	(12.99890)	41.32434	(41.66830)
	9.02422	(4.60467)	21.10413	(15.98780)	51.37023	(99.54538)

Acknowledgement

The author would like to thank Hanspeter Schmidli for pointing out an error in the first version of the paper.

Claudia Klüppelberg
Department of Mathematics
ETH Zürich
CH-8092 Zürich

Barbara Keller
Department of Mathematics
ETH Zürich
CH-8092 Zürich

References

Bingham, N.H./Goldie, C.M./Teugels, J.L. (1987): Regular Variation (Encyclopedia of mathematics and its applications v. 27). Cambridge University Press.

Davison, A.C./Smith, R.L. (1990): Models for exceedances over high thresholds (with discussion). J.R. Statist. Soc. B 52, 393–442

Devroye, L. (1984): Random variable generation for unimodal and monotone densities. Computing 32, 43–68.

Gerontides, L./Smith, R.L. (1982): Monte Carlo generation of order statistics from general distributions. Appl. Statist. 31, 238–243.

Häusler, E./Teugels, J.L. (1985): On asymptotic normality of Hill's estimator for the exponent of regular variation. Ann. Statist. 13, 743–756.

Hogg, R.V./Klugmann, S.A. (1984): Loss Distribution. Wiley, New York.

Keller, B. (1991): Simulation verallgemeinerter Pareto und Weinbullverteilungen und Berechnung Asymptotischer Maximum Likelihood Schätzer. Diplomarbeit ETH Zürich.

Klüppelberg, C./Villaseñor, J.A. (1990): Estimation of distribution tails – a parametric approach. Preprint.

Klüppelberg, C. (1991): On the asymptotic normality of parameter estimates for heavy Weibull-like tails. Preprint.

Kremer, E. (1985): Einführung in die Versichungsmathematik (Studia mathematica: Script, 7). Vanderhoeck & Rupprecht, Göttingen

Morgan, B.J.T. (1984): Elements of Simulation. Chapman and Hall, London.

Pickands, J. (1975): Statistical inference using extreme order statistics. Ann. Stat. 3, 113–131.

Smith, R.L. (1987): Estimating tails of probability distributions. Ann. Stat. 15, 1174–1207.

Summary

Heavy tailed distribution functions F are commonly used to model large claims in non-life insurance. A statistical method to estimate the far end of distribution, i.e. $F(x)$ for $x \geq x_T$, x_T large, is given by the so-called asymptotic maximum likelihood estimators. We investigate their properties for small samples in an Monte Carlo simulation. In particular, we study their robustness with respect to minor extensions of the model. Moreover, we introduce a combined inversion-rejection method to simulate random numbers only from the far end of a distribution which seems to be ideal for longtailed distributions. Finally, we apply our simulation results to certain reinsurance problems.

Zusammenfassung

Langschwänzige Verteilungen F dienen üblicherweise der Modellierung von grossen Schäden in der Sachversicherung. Eine statistische Methode, um das obere Ende $F(x)$, $x \geq x_T$, x_T gross, zu schätzen, sind die sogenannten Asymptotischen Maximum Likelihood-Schätzer. Wir untersuchen ihre Eigenschaften für kleine Stichproben mittels einer Simulationsstudie. Insbesondere studieren wir ihre Robustheit gegenüber geringfügigen Modellerweiterungen. Wir stellen ausserdem eine kombinierte Inversions-Verwerfungsmethode vor, die speziell Zufallsvariablen nur aus dem oberen Ende einer Verteilung simuliert und für langschwänzige Verteilungen besonders geeignet erscheint. Schliesslich wenden wir unsere Simulationsergebnisse auf verschiedene Vertragsformen in der Rückversicherung an.

Résumé

La construction de modèles pour les sinistres de montant très élevé dans l'assurance non-vie fait appel à des familles de fonctions de répartition $F(x)$ qui donnent un poids important aux grandes valeurs de x. Une méthode statistique permettant d'estimer la partie supérieure de $F(x)$, $x \geq x_T$, x_T grand, est caractérisée par les estimations asymptotiques de vraisemblance maximale. Nous analysons les propriétés de telles estimations pour de petits échantillons à l'aide de simulations. En particulier, nous étudions leur robustesse dans le cas d'extensions minimes du modèle. De plus, nous introduisons une procédure combinant les méthodes d'inversion et de rejet pour simuler des nombres aléatoires se rapportant uniquement à la partie supérieure de la distribution. Les résultats des simulations sont appliqués à différents problèmes de réassurance.