

Nonparametric Statistical Analysis of Ruin Probability under Conditions of "Small" and "Large" Claims

Una Analisi Statistica Nonparametrica della Probabilità di Rovina

Pier Luigi Conti, Esterina Masiello

Dipartimento di Statistica, Probabilità e Statistiche Applicate

Università di Roma "La Sapienza"

e-mail: esterina.masiello@uniroma1.it

Riassunto: Il presente lavoro si propone di sviluppare una analisi statistica nonparametrica della probabilità di rovina. Viene quindi introdotto uno stimatore della probabilità di rovina che possa essere applicato in casi generali in cui non si fa alcuna assunzione sulla distribuzione degli ammontari delle richieste alla compagnia di assicurazione e sul parametro della distribuzione dei tempi inter-arrivo. Per tale stimatore, si ottengono risultati di consistenza e convergenza debole a un processo Gaussiano.

Un approccio differente viene poi considerato per il caso in cui le richieste siano modellate da distribuzioni con code pesanti. L'idea è approssimare la probabilità di rovina Ψ mediante il Pareto bound introdotto da Willmot (1994) e fare inferenza su tale approssimazione.

Keywords: Poisson risk model, probability of ruin, nonparametric estimation, asymptotics, heavy-tailed distribution.

1. Introduction

The interest of actuarial mathematics in ruin probabilities dates back to the work of Lundberg (1903). Since this starting point, a number of contributions has been given, the main problem being the difficulty in finding a general expression for the probability of ruin Ψ (except for a few special cases).

We consider here the classical risk model, where claims occur as a Poisson process. The claim sizes $X_1, X_2, \dots$ are positive independent and identically distributed (*i.i.d.*) random variables (r.v.s), having common distribution function (d.f.) $F(x) = P(X_1 \leq x)$, finite mean $\mu = E(X_1)$ and finite variance $\sigma^2 = Var(X_1)$. The inter-arrival times $T_1, T_2, \dots$ are *i.i.d.* exponentially distributed, with finite mean $\mu_T = E(T_1) = 1/\lambda$ and d.f. $A(t) = P(T_1 \leq t) = 1 - e^{-\lambda t}$. Moreover, the two sequences $\{X_i\}$ and $\{T_i\}$ are independent. We further assume that the insurance company has an initial surplus $x > 0$ and that the premium income is linear in time with rate c . With no loss of generality, from now on we assume that $c = 1$.

The function of interest is $\Psi(x)$, the probability that the company will be ruined if the claims will exceed its available resources. If $\lambda\mu < 1$, it is possible to express Ψ as a compound geometric tail probability by mean of the so-called Pollaczek-Khinchine formula:

$$\Psi(x) = \sum_{k=1}^{\infty} (1 - \rho) \rho^k \bar{F}_I^{*k}(x) \quad x \geq 0 \quad (1)$$

where $\rho = \lambda\mu$, $\bar{F}_I(x) = 1 - F_I(x)$, $F_I(x) = \frac{1}{\mu} \int_0^x y dF(y)$ is the integrated tail distribution and F^{*k} denotes the k -fold convolution of the d.f. F_I .

2. A nonparametric estimator for the ruin probability

In this section, we deal with the construction of a nonparametric estimator $\Psi_n(x)$ of $\Psi(x)$ based on a sample $T_1, \dots, T_n$ of inter-arrival times and a sample $X_1, \dots, X_n$ of corresponding claims. This problem has already been studied by Pitts (1994), in the special case of a known value of ρ . We generalize here her results to the more realistic case of an unknown value of ρ . Our only assumption is the stability of the model: $\rho < 1$. The d.f. of inter-arrival times and claims are both supposed unknown. More specifically, the inter-arrival time distribution is assumed to be exponential with unknown parameter λ . As far as the claim distribution is concerned, no specific parametric model is assumed. Such a statistical problem has been considered by different authors but literature has mostly been concerned with asymptotic expansions; few results focus on statistical estimation of Ψ . Clearly, one could work out a parametric estimation procedure or use nonparametric methods. As remarked in Pitts (1994), the latter approach is particularly interesting in view of its applicability.

In our approach, the key tool for estimating the ruin probability is the Pollaczek-Khinchine formula. We first estimate $E(T) = \mu_T$ by its maximum likelihood estimator $\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_i$, and $E(X) = \mu$ by $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. The next step consists in estimating ρ by the ratio $\hat{\rho} = \bar{X}_n / \bar{T}_n$. Finally, we replace the k -fold convolution of F in Eq. (1) by the following estimator (Frees, 1986):

$$F_n^{*k}(x) = \binom{n}{k}^{-1} \sum_c I_{(X_{i_1} + \dots + X_{i_k} \leq x)}$$

where $\{i_1, i_2, \dots, i_k\}$ is a subset of size k of $\{1, 2, \dots, n\}$ and $\sum_c$ is the sum over all $\binom{n}{k}$ different combinations of $\{i_1, i_2, \dots, i_k\}$.

We have easily estimated the quantities at the right hand side of Eq. (1) by the corresponding sampling counterparts. It seems therefore natural to estimate $\Psi(x)$ by

$$\Psi_n(x) = \sum_{k=1}^{\infty} (1 - \hat{\rho}) \hat{\rho}^k \bar{F}_n^{*k}(x)$$

The asymptotic behavior of the proposed estimator has been investigated. Our main results are summarized in Proposition 2.1.

Proposition 2.1 *Under the assumption $\rho < 1$, as $n \rightarrow \infty$, we have:*

- (i) $\sup_x |\Psi_n(x) - \Psi(x)| \xrightarrow{a.s.} 0$
- (ii) *the sequence of stochastic processes $\{\sqrt{n}(\Psi_n(x) - \Psi(x)); x \geq 0\}_{n \geq 1}$ converges weakly (in the Skorokhod topology in $D[0, \infty]$) to a Gaussian process Z with mean function 0 and covariance kernel given by:*

$$E[Z(x)Z(y)] = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (1 - \rho)^2 \rho^{k+j} k j \text{Cov}(\bar{F}^{*k-1}(x - X_1), \bar{F}^{*j-1}(y - X_1))$$

$$\begin{aligned}
& + \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} f'_k(\rho) \bar{F}^{*k}(x) f'_j(\rho) \bar{F}^{*j}(y) \left(\frac{1}{\mu_T^2} \sigma^2 + \frac{\mu^2}{\mu_T^4} \sigma_T^2 \right) \\
& + 2 \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (1-\rho) \rho^k \frac{1}{\mu_T} f'_j(\rho) \bar{F}^{*j}(y) \text{Cov}(X_1, \bar{F}^{*k-1}(x - X_1))
\end{aligned}$$

This result can be used to construct confidence bands for $\Psi(x)$, based on resampling techniques.

3. Ruin probability under "high claims"

What happens to the ruin probability of an insurer who is incurring "large claims" modeled by distributions with heavy tails, which do not have exponential moments?

The estimator proposed in the above section, which naturally arises for the ruin probability, is essentially based on the empirical distribution function F_n and this involves a bad accuracy on the tails. As a consequence, when dealing with heavy-tailed distributions, we are forced to resort to a different approach. The basic idea is to resort to an appropriate approximation of $\Psi(x)$ ("for x large") and estimate this approximation.

The main result in this direction is the celebrated Lundberg's inequality. Assume that the Cramér-Lundberg condition holds, i.e. there exists a positive constant R (the so-called adjustment coefficient) such that:

$$\int_0^{\infty} e^{Rx} \bar{F}(x) dx < \infty$$

Then, Lundberg's inequality states the following exponential bound:

$$\Psi(x) \leq e^{-Rx} \quad \forall x \geq 0 \quad (2)$$

The problem is that for many practical distributions, Lundberg's inequality is not available because of the non-existence of the corresponding adjustment coefficients. For example, exponential bounds are not applicable to long tail distributions such as subexponential distributions or distributions which only have a finite number of moments.

Moreover, in many applications claims are mostly modeled by heavy-tailed distributions, such as Pareto, log-gamma, lognormal or heavy-tailed Weibull. In this case we meet a condition of "high claims". A "high claim" may be represented by an extremal event, such as hurricane, storm, forest fire, earthquake and natural catastrophes. It is difficult to predict and it is not generally caused by man-made disasters. Moreover, such extremal events occur with relatively small probability but have a significant influence on the behavior of the whole model.

Refinements and generalizations of the inequality (2) have been considered by several authors. A simple approach to the problem consists in relaxing the Cramér-Lundberg condition by considering functions different from the exponential function. Following this idea, Willmot (1994) replaces the exponential term by a thicker tail, such as a Pareto-type tail (see also Willmot and Lin (2000)). In such a way, he obtains the so-called Pareto bound which, applied to the ruin probability $\Psi(x)$, becomes:

$$\Psi(x) \leq (1 + kx)^{-m} \quad (3)$$

where $k > 0$ satisfies the relationship:

$$\int_0^\infty (1 + kx)^m dF(x) = \frac{1}{\rho}$$

and $m > 0$ is the number of moments of the distribution of the X_i 's supposed to exist. The moment based bound introduced by Willmot (1994) can be considered as a useful approximation of $\Psi(x)$. We make inference on this approximation. The problem we study is the construction of nonparametric point and interval estimates for the bound $(1 + kx)^{-m}$ at the right hand side of relation (3).

Again, we do not assume any specific parametric model for the distribution of the claims, while T_i 's are assumed *i.i.d.* exponential r.v.s with unknown parameter λ . Let $T_1, \dots, T_n$ and $X_1, \dots, X_n$ be two samples of inter-arrival times and claims, respectively. Our idea is to estimate $F(x)$ by the corresponding empirical d.f. $F_n(x)$ and define k_n as the solution to the equation

$$\frac{1}{n} \sum_{i=1}^n (1 + kX_i)^m - \frac{1}{\hat{\rho}} = 0$$

Now, it seems natural to consider the following estimate for the Pareto upper bound $(1 + kx)^{-m}$ at the right hand side of relation (3)

$$(1 + k_n x)^{-m} \tag{4}$$

Our main results are asymptotic results about the behavior of the estimator (4). They are stated in Proposition 3.1.

Proposition 3.1 *Under the assumption $\rho < 1$, as the sample size n goes to infinity, the following results hold true.*

- (i) $(1 + k_n x)^{-m} \xrightarrow{a.s.} (1 + kx)^{-m} \quad \forall x \geq 0.$
- (ii) *The sequence of r.v.s $\sqrt{n}((1 + k_n x)^{-m} - (1 + kx)^{-m})$ converges in law to a normal distribution with mean 0 and variance $\bar{\sigma}^2 = (m^2(1 + kx)^{-2(m+1)}x^2)\sigma_k^2$.*
- (iii) *Let $\bar{\sigma}_n^2 = (m^2(1 + k_n x)^{-2(m+1)}x^2)\hat{\sigma}_k^2$, then $\bar{\sigma}_n^2 \xrightarrow{a.s.} \bar{\sigma}^2$.*
- (iv) *Let $q_{\alpha/2}$ be the $\alpha/2$ -th quantile of the normal standard distribution. A confidence interval for $(1 + kx)^{-m}$, with asymptotic size $(1 - \alpha)$, is given by*

$$[(1 + k_n x)^{-m} - q_{\alpha/2}\bar{\sigma}_n/\sqrt{n}, (1 + k_n x)^{-m} + q_{\alpha/2}\bar{\sigma}_n/\sqrt{n}]$$

References

- Frees E. W. (1986) Nonparametric renewal function estimation, *The Annals of Statistics*, 14, 1366–1378.
- Lundberg F. (1903) Approximerad Framställning av Sannolikhetsfunktionen, *Aterförsäkring av Kollektivrisker.*, II, Almqvist & Wiksell, Uppsala.
- Pitts S. M. (1994) Nonparametric estimation of compound distributions with applications in insurance, *Ann. Inst. Statist. Math.*, 46, No. 3, 537–555.
- Willmot G. E. (1994) Refinements and distributional generalizations of Lundberg's inequality, *Insurance: Math. Econom.*, 15, 49–63.
- Willmot G. E., Lin X. S. (2000). *Lundberg approximations for compound distributions with insurance applications*, Springer-Verlang, Berlin.