On semiparametric estimation of ruin probabilities in the classical risk model

Esterina Masiello

Nonparametric Statistical Analysis of Ruin Probability

Esterina Masiello
Dipartimento di Statistica, Probabilità e Statistiche Applicate
P.le Aldo Moro 5, 00185 Rome Italy
e-mail: esterina.masiello@uniroma1.it

Abstract The ruin probability of an insurance company is a central topic in risk theory. In this paper, the classical Poisson risk model is considered and a nonparametric estimator of the ruin probability is provided. Strong consistency and asymptotic normality of the estimator are estabilished. Bootstrap confidence bands are also studied. Further, a simulation example is presented in order to investigate the finite sample properties of the proposed estimator.

Key words Poisson risk model, ruin probability, nonparametric estimation, asymptotics.

1. Introduction and preliminaries

We consider the classical Poisson risk model, where the inter-arrival times $T_1, T_2, \ldots$ form a sequence of independent random variables (r.v.'s) with common exponential distribution function (d.f.) $A(t) = P(T \leq t) = 1 - e^{-\lambda t}$ and finite mean $\mu_T = 1/\lambda$. The claim sizes $X_1, X_2, \ldots$ are positive independent and identically distributed ($i.i.d.$) r.v.'s, with common d.f. $F(x) = P(X \leq x)$, finite mean μ and finite variance σ^2. Moreover, the two sequences $\{T_i\}$ and $\{X_i\}$ are supposed to be independent. We further assume that the insurance company has an initial surplus $x > 0$ and that premiums arrive at a known steady rate, say, $c > 0$ per unit time.

Let $\Psi(x)$ denote the probability of ruin starting with initial reserve x. Exact calculation and approximation of ruin probabilities have been a great source of inspiration and technique development in actuarial mathematics since the seminal paper by Lundberg (1903). The problem is that, apart from some special cases, a general expression for Ψ does not exist. We will construct a nonparametric estimator $\Psi_n(x)$ of $\Psi(x)$ based on a sample $T_1, \ldots, T_n$ of inter-arrival times and a sample $X_1, \ldots, X_n$ of corresponding claims. 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. As remarked by Embrechts et al (1997), 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. Also Frees (1986a) and Croux and Veraverbeke (1990) have proposed a nonparametric estimator for Ψ. The estimator by Croux and Veraverbeke (1990) is a linear combination of U-statistics; they generalize techniques in Frees (1986b) whose estimator is based on the sample reuse concept. For a nonparametric approach to the estimation of Ψ see also Hipp (1989a, 1989b).

Our approach is closely connected to that developed in Pitts (1994), but the technique used to prove asymptotic results for the estimator is different from ours since that author takes a functional view of the stochastic model, following the tradition of Gill (1989), Grübel (1989) and Grübel and Pitts (1993).

Also the ideas in Frees (1986a), as well as the ideas in Harel et al. (1995), on nonparametric estimation of the renewal function are basic for our approach.

2. The nonparametric estimator

In practical applications, the parameter λ of the inter-arrival times distribution and the claim size d.f. F will not be given but have to be estimated with available data.

Let us consider a random sample of n inter-arrival times $T_1, \ldots, T_n$ with the corresponding claims $X_1, \ldots, X_n$. Suppose that the assumptions of the Cramer-Lundberg model are satisfied. If the model under consideration is stable, that is $\lambda\mu/c < 1$, it is possible to express the ruin probability Ψ as a compound geometric tail probability by mean of the so-called Pollaczeck-Khinchine formula (see, e.g. Feller (1971), Grandell (1990) or Embrechts and Veraverbeke (1982)):

$$\Psi(x) = \sum_{k=1}^{\infty} (1-\rho)\rho^k (1 - F_I^{*k}(x)) \qquad x \geq 0 \tag{1}$$

where $\rho = \lambda\mu/c$, $F_I(x) = \frac{1}{\mu}\int_0^x y\, dF(y)$ is the integrated tail distribution and F_I^{*k} denotes the k-fold convolution of the d.f. F_I. For $k = 1, 2, \ldots$, $F_I^{*k}(x) = P(Y_1 + \ldots + Y_k \leq x)$, where $Y_1, \ldots, Y_k$ are independent r.v.'s with common density function $f_I(x) = (1/\lambda)(1 - F(x))$, $x \geq 0$.

Remark 1. $F_I(x)$ is also called equilibrium distribution of the d.f. $F(x)$. In insurance, it can be interpreted as the d.f. of the amount of the drop in the surplus, or equivalently the difference between the initial surplus and the surplus at the time it first falls below the initial surplus. In other words, defined $S(y)$ as a random sum of claim amounts, F_I is the d.f. of $S(y) - cy$, given that $S(y) - cy > 0$ and $S(t) - ct \leq 0$, $0 \leq t < y$.

The Pollaczeck-Khinchine formula is not entirely satisfying as a tool for computing $\Psi(x)$ because of the infinite sum of convolution powers but it will be our key tool for estimating the ruin probability. From now on, for the sake of simplicity and without loss of generality, the rate of premium income per unit time will be taken to be equal to 1 ($c = 1$); otherwise , it is enough to refer to the transformed claims $\tilde{X}_i = X_i/c$.

Our approach is the following: we first estimate $E(T) = \mu_T$ by its maximum likelihood estimator $\overline{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$, and $E(X) = \mu$ by $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The next step consists in estimating ρ by the ratio $\hat{\rho} = \overline{X}_n/\overline{T}_n$. Finally, we replace the k-fold convolution of F_I in Eq. (1) by the following estimator (Frees, 1986b):

$$F_n^{*k}(x) = \binom{n}{k}^{-1} \sum_c I_{(Y_{i_1}+\ldots+Y_{i_k} \leq x)} \tag{2}$$

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

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

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

where $\overline{F}_n^{*k}(x) = 1 - F_n^{*k}(x)$.

Some of the properties of this estimator are shortly outlined in Conti and Masiello (2006a) and Conti and Masiello (2006b). In this paper, we obtain the asymptotic distribution of the estimator (Section 2.1) and study bootstrap confidence bands (Section 4). Moreover, in Section 5, a short simulation exemple is provided in order to illustrate the finite sample behavior of the estimator. An estimator for the p-th quantile of the ruin probability function is also proposed in Section 3, together with its main asymptotic properties.

Remark 2. The estimator $\Psi_n(x)$ is a functional of $F_n(x)$ and $\hat{\rho}$, i.e. $\Psi_n = \Phi(F_n, \hat{\rho})$. As a possible approach to the study of statistical properties of the proposed estimator, we could try to extend methods in Pitts (1994), who has studied the problem in the special case of a known value of ρ. In order to derive strong consistency and asymptotic normality results for the "output" estimator Ψ_n, it would suffice to establish analytical properties of continuity and differentiability for the functional Φ and combine them with the statistical properties of the "input" estimators F_n and $\hat{\rho}$. We will take a slightly different approach following ideas in Harel et al. (1995) in order to prove weak convergence of the stochastic process in the Skorokhod topology, but this will be discussed in the sequel.

Remark 3. Based on the estimator in (2), Frees (1986b) introduces a nonparametric estimator of the renewal function $H(t) = \sum_{k=1}^{\infty} F^{*k}(t)$ given by

$$H_n(t) = \sum_{k=1}^{m} F_n^{*k}(t)$$

where $m = m(n)$ is a positive integer depending on n, such that $m \leq n$ and $m \uparrow \infty$ as $n \uparrow \infty$. We left in the definition of our estimator $\Psi_n(x)$ an infinite sum. On this subject, observe that the results which we are going to obtain are essentially asymptotic results applied when the sample size n tends to infinity.

2.1. Basic asymptotic results

In the above section, we have provided a point estimate for $\Psi(x)$. In order to study its behavior, we now need to obtain the asymptotic distribution of the estimator $\Psi_n(x)$.

The key results of the present section are Theorem 2 and Theorem 3, where the strong consistency and the asymptotic normality of the estimator are estabilished.

To formulate our main theorems, we need some preliminary results. The asymptotic distribution of $\hat{\rho}$ is obtained first, since it will play a central role in all subsequent developments.

Theorem 1 *Under the assumptions of the Cramér-Lundberg model and if $\rho < 1$, as n goes to infinity, the following results hold:*

- $\hat{\rho} \xrightarrow{a.s.} \rho$
- $\sqrt{n}(\hat{\rho} - \rho)$ *tends in law to a normal distribution with mean zero and variance* $\sigma_\rho^2 = \frac{1}{\mu_T^2}\sigma^2 + \frac{\mu^2}{\mu_T^4}\sigma_T^2$.

Proof We may first write:

$$\hat{\rho} = f(\overline{X}_n, \overline{T}_n) = \frac{\overline{X}_n}{\overline{T}_n}$$

Take a first order Taylor expansion of $f(\overline{X}_n, \overline{T}_n)$ at the point (μ, μ_T) to obtain:

$$f(\overline{X}_n, \overline{T}_n) = f(\mu, \mu_T) + \left.\frac{\partial f}{\partial \overline{X}_n}\right|_{(\mu,\mu_T)} (\overline{X}_n - \mu) + \left.\frac{\partial f}{\partial \overline{T}_n}\right|_{(\mu,\mu_T)} (\overline{T}_n - \mu_T) + Rem$$

where Rem denotes the remainder term. It is easy to get the following relationship:

$$\sqrt{n}(\hat{\rho} - \rho) = \left(\frac{1}{\mu_T}\right) \sqrt{n}(\overline{X}_n - \mu) - \frac{\mu}{\mu_T^2}\sqrt{n}(\overline{T}_n - \mu_T) + \sqrt{n}\, Rem \quad (3)$$

The asymptotic behavior of the estimators $\overline{X}_n$ and $\overline{T}_n$ is well-known. In fact, they converge almost surely to the corresponding "true values" as the sample size n tends to infinity. Furthermore, $\sqrt{n}(\overline{X}_n - \mu)$ and $\sqrt{n}(\overline{T}_n - \mu_T)$ have normal asymptotic distributions with mean zero and variances σ^2 and σ_T^2, respectively. As a simple application of the delta method, we obtain

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \sigma_\rho^2)$$

where $\sigma_\rho^2 = \frac{1}{\mu_T^2}\sigma^2 + \frac{\mu^2}{\mu_T^4}\sigma_T^2$. □

Remark 4. The mean and the variance of $\hat{\rho}$ can be obtained by direct calculations. Taking into account that $\hat{\rho} = \overline{X}_n/\overline{T}_n$ and that $\overline{X}_n$ and $\overline{T}_n$ are independent, we can easily obtain:

$$E\left[\frac{\overline{X}_n}{\overline{T}_n}\right] = nE(\overline{X}_n)E\left(\frac{1}{\sum_{i=1}^n T_i}\right) \quad (4)$$

Since T_i is exponential with density:

$$f_T(t) = \begin{cases} 0 & \text{if } t < 0 \\ \lambda e^{-\lambda t} & \text{if } t \geq 0 \end{cases}$$

then $V = \sum_{i=1}^{n} T_i$ is Erlang distributed with density:

$$f_V(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} & \text{if } t \geq 0 \end{cases}$$

Hence, we obtain

$$\begin{aligned} E\left[\frac{1}{\sum_{i=1}^{n} T_i}\right] = E\left[\frac{1}{V}\right] &= \int_0^\infty \frac{1}{t} \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} dt \\ &= \frac{\lambda}{(n-1)!} \Gamma(n-1) \\ &= \frac{\lambda}{n-1} \end{aligned}$$

With obvious replacements, equation (4) becomes:

$$E\left[\frac{\overline{X}_n}{\overline{T}_n}\right] = \frac{n}{n-1} \frac{\mu}{\mu_T}$$

When the sample size n goes to infinity, we obtain:

$$E\left[\frac{\overline{X}_n}{\overline{T}_n}\right] \to \frac{\mu}{\mu_T}$$

so that the estimator $\hat{\rho}$ is asymptotically unbiased. Let us consider the variance:

$$\begin{aligned} Var\left(\frac{\overline{X}_n}{\overline{T}_n}\right) &= E\left[\frac{\overline{X}_n^2}{\overline{T}_n^2}\right] - \left[E\left(\frac{\overline{X}_n}{\overline{T}_n}\right)\right]^2 \\ &= E[\overline{X}_n^2] E\left[\frac{1}{\overline{T}_n^2}\right] - \left[E\left(\frac{\overline{X}_n}{\overline{T}_n}\right)\right]^2 \qquad (5) \end{aligned}$$

where $E[\overline{X}_n^2] = Var(\overline{X}_n) + [E(\overline{X}_n)]^2 = \sigma^2/n + \mu^2$ and

$$\begin{aligned} E\left(\frac{1}{\overline{T}_n^2}\right) = E\left[\frac{1}{\left(\frac{1}{n}\sum_{i=1}^{n} T_i\right)^2}\right] &= n^2 E\left(\frac{1}{V^2}\right) \\ &= n^2 \int_0^\infty \frac{1}{t^2} \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} dt \\ &= n^2 \frac{\lambda^2}{(n-1)!} \Gamma(n-2) \\ &= \frac{n^2}{(n-2)(n-1)} \frac{1}{\mu_T^2} \end{aligned}$$

Once more, with obvious replacements in equation (5), we obtain the following exact result:

$$Var\left(\frac{\overline{X}_n}{\overline{T}_n}\right) = \frac{n}{(n-2)(n-1)}\frac{\sigma^2}{\mu_T^2} + \frac{n^2}{(n-2)(n-1)^2}\frac{\mu^2}{\mu_T^2}$$

and, asymptotically,

$$\lim_{n\to\infty} nVar(\hat{\rho}) = \frac{\sigma^2}{\mu_T^2} + \frac{\mu^2}{\mu_T^2}$$

We still need some auxiliary results. Let us define

$$\widetilde{\Psi}_n(x) = \sum_{k=1}^{\infty}(1-\hat{\rho})\,\hat{\rho}^k\,\overline{F}^{*k}(x)$$

We can rewrite our "empirical" process as

$$\begin{aligned}\sqrt{n}(\Psi_n(x) - \Psi(x)) &= \sqrt{n}(\Psi_n(x) - \widetilde{\Psi}_n(x)) + \sqrt{n}(\widetilde{\Psi}_n(x) - \Psi(x)) \\ &= U_n(x) + V_n(x)\end{aligned}$$

where we have defined $U_n(x) = \sqrt{n}(\Psi_n(x) - \widetilde{\Psi}_n(x))$ and $V_n(x) = \sqrt{n}(\widetilde{\Psi}_n(x) - \Psi(x))$ and we study each of the two components separately. We begin by a couple of lemmas describing the behavior of $U_n(x)$ and $V_n(x)$, respectively.

Lemma 1 *Let us define $U_n(x) = \sqrt{n}(\Psi_n(x) - \widetilde{\Psi}_n(x))$. With the above assumptions, for each $x > 0$, as $n \to \infty$, we obtain the following result:*

$$\lim_{n\to\infty} E[U_n(x)U_n(y)] = (1-\rho)^2 \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^{k+j}\, k\, j\, Cov(\overline{F}^{*k-1}(x-Y_1), \overline{F}^{*j-1}(y-Y_1))$$

Proof First of all, we may express $U_n(x)$ in the form:

$$\begin{aligned}U_n(x) &= \sqrt{n}(\Psi_n(x) - \widetilde{\Psi}_n(x)) \\ &= \sqrt{n}\sum_{k=1}^{\infty}(1-\hat{\rho})\hat{\rho}^k(\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \\ &= \sqrt{n}\sum_{k=1}^{\infty}(1-\rho)\rho^k(\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \\ &\quad + \sum_{k=1}^{\infty} n^{1/4}[(1-\hat{\rho})\hat{\rho}^k - (1-\rho)\rho^k]n^{1/4}(\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x))\end{aligned}$$

Since the quantities $n^{1/4}[(1-\hat{\rho})\hat{\rho}^k - (1-\rho)\rho^k]$ and $n^{1/4}\left[\sup_x \left|\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)\right|\right]$ both converge in probability to zero when

n goes to infinity, the asymptotic distribution of $U_n(x)$ coincides with that of $\sqrt{n}\sum_{k=1}^{\infty}(1-\rho)\rho^k(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))$. The following relations hold:

$$E[U_n(x)U_n(y)] = (1-\rho)^2\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\rho^{k+j}E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))$$
$$\sqrt{n}(\overline{F}_n^{*j}(y)-\overline{F}^{*j}(y))]$$

Let us consider the expected value on the right hand side of this last expression, to obtain

$$\begin{aligned}E[\sqrt{n}(\ \overline{F}_n^{*k}\ (x)-\overline{F}^{*k}(x))\sqrt{n}(\overline{F}_n^{*j}(y)-\overline{F}^{*j}(y))] &= n\{E[\overline{F}_n^{*k}(x)\overline{F}_n^{*j}(y)]\\ &-\overline{F}^{*j}(y)\overline{F}^{*k}(x)-\overline{F}^{*k}(x)\overline{F}^{*j}(y)+\overline{F}^{*k}(x)\overline{F}^{*j}(y)\}\\ &= nE[\overline{F}_n^{*k}(x)\overline{F}_n^{*j}(y)]-n\overline{F}^{*j}(y)\overline{F}^{*k}(x)\end{aligned}$$

since $\overline{F}_n^{*k}(x)$ is an unbiased estimator of $\overline{F}^{*k}(x)$. As pointed out by Frees (1986b), this estimator is a U-statistic and thus it is easy to establish that for each $k\geq 1$ and for each $x\geq 0$, $F_n^{*k}(x)\overset{a.s.}{\to}F^{*k}(x)$ as $n\to\infty$. Moreover, it is uniformly almost surely consistent, that is, for every $k\geq 1$, as n goes to infinity

$$\sup_x|F_n^{*k}(x)-F^{*k}(x)|\overset{a.s.}{\to}0$$

We thus obtain:

$$\begin{aligned}E[\overline{F}_n^{*k}(x)\overline{F}_n^{*j}(y)] &= E\left[\binom{n}{k}^{-1}\sum_{c}I_{(Y_{i_1}+\ldots+Y_{i_k}>x)}\binom{n}{j}^{-1}\sum_{c}I_{(Y_{i_1}+\ldots+Y_{i_j}>y)}\right]\\ &= E\left[\binom{n}{k}^{-1}\binom{n}{j}^{-1}\sum_{\underline{c}_k\in C_{n,k}}\sum_{\underline{c}_j\in C_{n,j}}I_{(Y_{i_1}+\ldots+Y_{i_k}>x)}\right.\\ &\qquad\left.I_{(Y_{i_1}+\ldots+Y_{i_j}>y)}\right]\end{aligned}$$

where $C_{n,k}$ is the set of the combinations of n elements of class k and $C_{n,j}$ is the set of the combinations of n elements of class j; $\underline{c}_k$ and $\underline{c}_j$ are two subsets of $\{1,\ldots,n\}$ that have $l\leq\min(k,j)$ elements in common. Suppose $k>j$, then it is possible to write $\underline{c}_j=(\underline{c}_{j,k},\underline{c}_{j,\overline{k}})$, where $\underline{c}_{j,k}$ is the set of the elements of $\underline{c}_j$ which are in $\underline{c}_k$. In general, $\underline{c}_{j,k}$ is composed by l elements ($l=0,1,\ldots,j$). Hence, we have

$$E[\overline{F}_n^{*k}\ (x)\ \overline{F}_n^{*j}(y)] = \binom{n}{k}^{-1}\binom{n}{j}^{-1}\sum_{\underline{c}_k\in C_{n,k}}\sum_{l=0}^{j}\sum_{\underline{c}_{j,k}\in C_{k,l}}\sum_{\underline{c}_{j,\overline{k}}\in C_{n-k,j-l}}$$
$$E\ \ [I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{i_{l+1}}+\ldots+Y_{i_k}>x)}I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{h_{l+1}}+\ldots+Y_{h_j}>y)}]\quad(6)$$

where $\underline{c}_{j,k}$ is a combination without replacement of l elements of $\underline{c}_j$ and $\underline{c}_{j,\overline{k}}$ is a combination without replacement of $j-l$ elements of $\{1,\ldots,n\}\backslash\,\underline{c}_j$; moreover,

$n-k$ is not smaller than $j-l$. Using the independence of Y_i's, we easily have the following relationship:

$$
\begin{aligned}
& E[I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{i_{l+1}}+\ldots+Y_{i_k}>x)} I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{h_{l+1}}+\ldots+Y_{h_j}>y)}] \\
& = E[E[I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{i_{l+1}}+\ldots+Y_{i_k}>x)} I_{(Y_{i_1}+\ldots+Y_{i_l}+Y_{h_{l+1}}+\ldots+Y_{h_j}>y)} | Y_{i_1}, \ldots, Y_{i_l}]] \\
& = E[\overline{F}^{*k-l}(x-Y_{i_1}-\ldots-Y_{i_l})\overline{F}^{*j-l}(y-Y_{i_1}-\ldots-Y_{i_l})]
\end{aligned}
$$

from which it follows that (6) takes the form

$$
\binom{n}{k}^{-1}\binom{n}{j}^{-1} \sum_{\underline{c}_k \in C_{n,k}} \sum_{l=0}^{j} \sum_{\underline{c}_{j,k} \in C_{k,l}} \sum_{\underline{c}_{j,\overline{k}} \in C_{n-k,j-l}} E[\overline{F}^{*k-l}(x-Y_{i_1}-\ldots-Y_{i_l})
\overline{F}^{*j-l}(y-Y_{i_1}-\ldots-Y_{i_l})]
$$

Finally, with obvious replacements, we obtain the following expression:

$$
\begin{aligned}
E[U_n(x)U_n(y)] = (1-\rho)^2 \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^{k+j} \Bigg\{ n \binom{n}{k}^{-1}\binom{n}{j}^{-1} \sum_{\underline{c}_k \in C_{n,k}} \\
\sum_{l=0}^{j} \sum_{\underline{c}_{j,k} \in C_{k,l}} \sum_{\underline{c}_{j,\overline{k}} \in C_{n-k,j-l}} E[\overline{F}^{*k-l}(x-Y_{i_1}-\ldots-Y_{i_l}) \\
\overline{F}^{*j-l}(y-Y_{i_1}-\ldots-Y_{i_l})] - n\overline{F}^{*j}(y)\overline{F}^{*k}(x) \Bigg\}
\end{aligned}
$$

Now, let n go to infinity to obtain

$$
\begin{aligned}
\lim_{n\to\infty} E[U_n(x)U_n(y)] = (1-\rho)^2 \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^{k+j}\, k\, j \\
Cov(\overline{F}^{*k-1}(x-Y_1), \overline{F}^{*j-1}(y-Y_1))
\end{aligned} \tag{7}
$$

□

Lemma 2 *Let us define $V_n(x) = \sqrt{n}(\widetilde{\Psi}_n(x) - \Psi(x))$. Under the same assumptions as in Lemma 1, for each $x > 0$, as $n \to \infty$, we obtain*

$$
\lim_{n\to\infty} E[V_n(x)V_n(y)] = \left(\frac{\sigma^2}{\mu_T^2} + \frac{\mu^2}{\mu_T^4}\sigma_T^2\right) \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_k^{'}(\rho) f_j^{'}(\rho) \overline{F}^{*k}(x)\overline{F}^{*j}(y)
$$

where $f_l^{'}(\rho)$ denotes the first derivative of the function $f_l(\rho) = (1-\rho)\rho^l$.

Proof We have first

$$
\begin{aligned}
V_n(x) &= \sqrt{n}(\widetilde{\Psi}_n(x) - \Psi(x)) \\
&= \sqrt{n} \sum_{k=1}^{\infty} [(1-\hat{\rho})\hat{\rho}^k - (1-\rho)\rho^k] \overline{F}^{*k}(x) \\
&= \sqrt{n} \sum_{k=1}^{\infty} [f_k(\hat{\rho}) - f_k(\rho)] \overline{F}^{*k}(x)
\end{aligned}
$$

where $f_k(t) = (1-t)t^k$. Taking a first order Taylor expansion of $f_k(\hat{\rho})$ at the point ρ, we obtain the following relationship:

$$V_n(x) = \sum_{k=1}^{\infty} [\sqrt{n}(\hat{\rho} - \rho) f_k^{'}(\rho) + \sqrt{n} Rem_k] \overline{F}^{*k}(x)$$

with $\sup_x \sqrt{n} |\sum_{k=1}^{\infty} Rem_k \overline{F}^{*k}(x)| \xrightarrow{p} 0$ when $n \to \infty$. Hence, $V_n(x)$ possesses the same asymptotic distribution as $\sqrt{n}(\hat{\rho} - \rho) \sum_{k=1}^{\infty} f_k^{'}(\rho) \overline{F}^{*k}(x)$.

Let us identify the covariance:

$$E[V_n(x)V_n(y)] = E\left[\left(\sqrt{n}(\hat{\rho} - \rho) \sum_{k=1}^{\infty} f_k^{'}(\rho) \overline{F}^{*k}(x)\right)\left(\sqrt{n}(\hat{\rho} - \rho) \sum_{j=1}^{\infty} f_j^{'}(\rho) \overline{F}^{*j}(y)\right)\right]$$

Taking into account relation (3), we have:

$$\begin{aligned} E[V_n(x)V_n(y)] &= E\left[\left(\sqrt{n}\frac{1}{\mu_T}(\overline{X}_n - \mu) - \sqrt{n}\frac{\mu}{\mu_T^2}(\overline{T}_n - \mu_T)\right)^2 \right. \\ &\quad \left. \left(\sum_{k=1}^{\infty} f_k^{'}(\rho)\overline{F}^{*k}(x)\right)\left(\sum_{j=1}^{\infty} f_j^{'}(\rho)\overline{F}^{*j}(y)\right)\right] \\ &= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_k^{'}(\rho) f_j^{'}(\rho)\overline{F}^{*k}(x)\overline{F}^{*j}(y)\left\{\frac{1}{\mu_T^2}E[\sqrt{n}(\overline{X}_n - \mu)]^2 - \frac{2\mu}{\mu_T^3}\right. \\ &\quad \left. E[\sqrt{n}(\overline{X}_n - \mu)]E[\sqrt{n}(\overline{T}_n - \mu_T)] + \frac{\mu^2}{\mu_T^4}E[\sqrt{n}(\overline{T}_n - \mu_T)]^2\right\} \end{aligned}$$

Letting n go to infinity and taking into account that $E(\sqrt{n}(\overline{X}_n - \mu)) = 0$ and $E(\sqrt{n}(\overline{T}_n - \mu_T)) = 0$, we obtain:

$$\lim_{n\to\infty} E[V_n(x)V_n(y)] = \left(\frac{\sigma^2}{\mu_T^2} + \frac{\mu^2}{\mu_T^4}\sigma_T^2\right)\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_k^{'}(\rho) f_j^{'}(\rho)\overline{F}^{*k}(x)\overline{F}^{*j}(y) \quad (8)$$

□

We are now in a position to establish the main results of the present paper in the following two theorems.

Theorem 2 *Under the same assumptions as in the previous lemmas,*

$$\sup_x |\Psi_n(x) - \Psi(x)| \xrightarrow{a.s.} 0$$

as n tends to infinity.

Proof First observe that

$$\begin{aligned}\sup_x |\Psi_n(x) - \Psi(x)| &= \sup_x \left| \sum_{k=1}^{\infty} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \right. \\ &\quad \left. + (\hat{\rho} - \rho) \sum_{k=1}^{\infty} f_k'(\rho) \overline{F}^{*k}(x) \right| \\ &\leq \sup_x \left| \sum_{k=1}^{\infty} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \right| \\ &\quad + \sup_x \left| (\hat{\rho} - \rho) \sum_{k=1}^{\infty} f_k'(\rho) \overline{F}^{*k}(x) \right|\end{aligned}$$

Consider that

$$\begin{aligned}\sum_{k=1}^{\infty} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) &= \sum_{k=1}^{L} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \\ &\quad + \sum_{k=L+1}^{\infty} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x))\end{aligned}$$

Now, take into account that for every $\epsilon > 0$, there exists $L = L_\epsilon$ such that $\sum_{k=L+1}^{\infty} (1-\rho)\rho^k < \epsilon$ and write

$$\begin{aligned}\sup_x \left| \sum_{k=L+1}^{\infty} (1-\rho)\rho^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \right| &\leq \sum_{k=L+1}^{\infty} (1-\rho)\rho^k \\ &\quad \sup_x \left| (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \right| \\ &\leq \sum_{k=L+1}^{\infty} (1-\rho)\rho^k \\ &\leq \epsilon\end{aligned}$$

On the basis of this last relationships, also taking into account that $|\hat{\rho} - \rho| \overset{a.s.}{\to} 0$ and that $\sup_x \left| \overline{F}_n^{*k}(x) - \overline{F}^{*k}(x) \right| \overset{a.s.}{\to} 0$, the statement of the theorem easily follows. □

Theorem 3 *Let D denote the set of right-continuous functions with left-hand limits, endowed with the Skorokhod topology. For each $x > 0$, as $n \to \infty$, the sequence of stochastic processes $Z_n(x) = \sqrt{n}(\Psi_n(x) - \Psi(x))$ converges weakly in the space D to a Gaussian process Z with mean function zero and covariance kernel given by (17).*

Proof For the sake of clarity, the proof is split into three steps. In the first step we show that the finite-dimensional distributions of $Z_n(\cdot)$ converge in law to a multivariate normal distribution while, in the second step, using arguments in

Harel et al. (1995), we state the weak convergence of the process. In the last step, calculations for the identification of the covariance kernel of the process are showed.

Step 1 (Convergence of the finite-dimensional distributions). We just consider one-dimensional distributions, since the same reasoning applies also to the multidimensional case. We need to show that, if t is a continuity point of the distribution of Z,

$$\lim_{n\to\infty} P(Z_n \leq t) = P(Z \leq t)$$

Just recall that

$$\begin{aligned} Z_n\ (x) &= U_n(x) + V_n(x) \\ &= (1-\rho)\sqrt{n}\sum_{k=1}^{\infty}\rho^k(\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) + \sqrt{n}(\hat{\rho}-\rho)\sum_{k=1}^{\infty} f_k^{'}(\rho)\overline{F}^{*k}(x) \end{aligned} \quad (9)$$

From Lemma 2 and Theorem 1, $V_n(x) = \sqrt{n}(\hat{\rho}-\rho)\sum_{k=1}^{\infty} f_k^{'}(\rho)\overline{F}^{*k}(x)$ converges in law to a normal distribution with mean zero and variance σ_V^2 obtained by (8) simply taking $x = y$. Then, we can write

$$P(V_n(x) \leq t) = P(N(0, \sigma_V^2) \leq t) = \Phi(t/\sigma_V)$$

where, with obvious notation, $\Phi(x) = P(N(0,1) \leq x)$.

Let us now study the term $U_n(x)$. To simplify the notation, we will consider from now on, throughout this step, $F^{*k}(x)$ instead of the tail $\overline{F}^{*k}(x)$, since the same arguments apply. For every $L \geq 1$, we have that

$$\begin{aligned} \widetilde{U}_n(x) = (1-\rho)\sum_{k=1}^{L}\rho^k\sqrt{n}(F_n^{*k}(x) - F^{*k}(x)) \\ +(1-\rho)\sum_{k=L+1}^{\infty}\rho^k\sqrt{n}(F_n^{*k}(x) - F^{*k}(x)) \end{aligned}$$

Once more to simplify the notation, let us take $a_k = (1-\rho)\rho^k\sqrt{n}(F_n^{*k}(x) - F^{*k}(x))$. For fixed x, the event $(\widetilde{U}_n(x) \leq t)$ can be written as

$$(\widetilde{U}_n(x) \leq t) = \left(\sum_{k=1}^{L} a_k \leq t - \sum_{k=L+1}^{\infty} a_k\right)$$

For every $\epsilon > 0$ and $L \geq 1$, we will have

$$\begin{aligned} (\widetilde{U}_n(x) \leq t) = &\left\{\left(\sum_{k=1}^{L} a_k \leq t - \sum_{k=L+1}^{\infty} a_k\right)\bigcap\left(\left|\sum_{k=L+1}^{\infty} a_k\right| \leq \epsilon\right)\right\} \\ &\bigcup\left\{\left(\sum_{k=1}^{L} a_k \leq t - \sum_{k=L+1}^{\infty} a_k\right)\bigcap\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right)\right\} \end{aligned}$$

The following chain of inequalities holds true:

$$
\begin{aligned}
P(\widetilde{U}_n(x) \le t) &= P\left(\sum_{k=1}^{\infty} a_k \le t\right) \\
&\le P\left(\sum_{k=1}^{L} a_k \le t - \sum_{k=L+1}^{\infty} a_k, \left|\sum_{k=L+1}^{\infty} a_k\right| \le \epsilon\right) \\
&\quad + P\left(\sum_{k=1}^{L} a_k \le t - \sum_{k=L+1}^{\infty} a_k, \left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \\
&\le P\left(\sum_{k=1}^{L} a_k \le t - \sum_{k=L+1}^{\infty} a_k, -\epsilon \le \sum_{k=L+1}^{\infty} a_k \le \epsilon\right) \\
&\quad + P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \\
&\le P\left(\sum_{k=1}^{L} a_k \le t + \epsilon, -\epsilon \le \sum_{k=L+1}^{\infty} a_k \le \epsilon\right) \\
&\quad + P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \\
&\le P\left(\sum_{k=1}^{L} a_k \le t + \epsilon\right) + P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \qquad (10)
\end{aligned}
$$

Let us consider the first term in the summation in the expression (10); we have

$$
\begin{bmatrix} a_1 \\ \dots \\ a_L \end{bmatrix} = \begin{bmatrix} (1-\rho)\rho\sqrt{n}(F_n^{*1}(x) - F^{*1}(x)) \\ \dots \\ (1-\rho)\rho^L\sqrt{n}(F_n^{*L}(x) - F^{*L}(x)) \end{bmatrix}
$$

that is L linear combinations of U-statistics. The joint distribution is multinormal as an application of Hoeffding theorem (Serfling (1980)). For every $L \ge 1$, $\sqrt{n}\sum_{k=1}^{L} a_k$ converges in law, as n goes to infinity, to a normal distribution with mean zero and variance

$$
\sigma_L^2 = (1-\rho)^2 \sum_{k=1}^{L}\sum_{j=1}^{L} \rho^{k+j} k\, j Cov(F^{*k-1}(x - Y_1), F^{*j-1}(x - Y_1)) \quad (11)
$$

obtained from (7) by letting $y = x$. We can finally state

$$
\lim_{n\to\infty} P\left(\sum_{k=1}^{L} a_k \le t + \epsilon\right) = P(N(0, \sigma_L^2) \le t + \epsilon) = \Phi((t + \epsilon)/\sigma_L)
$$

As far as the second term in expression (10) is concerned, by applying Chebyshev inequality we obtain

$$P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) = P\left(\left|\sum_{k=L+1}^{\infty}(1-\rho)\rho^k\sqrt{n}(F_n^{*k}(x) - F^{*k}(x))\right| > \epsilon\right)$$
$$\leq \frac{1}{\epsilon^2}E\left[\left(\sum_{k=L+1}^{\infty}(1-\rho)\rho^k\sqrt{n}(F_n^{*k}(x) - F^{*k}(x))\right)^2\right]$$

Letting $T_n^k(x) = \sqrt{n}(F_n^{*k}(x) - F^{*k}(x))$, the last expression becomes

$$P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}(1-\rho)^2\sum_{k=L+1}^{\infty}\sum_{j=L+1}^{\infty}\rho^{k+j}E[T_n^k(x)T_n^j(x)]$$
$$= \frac{1}{\epsilon^2}(1-\rho)^2\sum_{k=L+1}^{\infty}\sum_{j=L+1}^{\infty}\rho^{k+j}\sigma_n^{kj}$$

where, following the same lines as in the proof of Lemma 1, we easily obtain:

$$\sigma_n^{kj} = n\binom{n}{k}^{-1}\binom{n}{j}^{-1}\sum_{\underline{c}_k\in C_{n,k}}\sum_{l=0}^{j}\sum_{\underline{c}_{j,k}\in C_{k,l}}\sum_{\underline{c}_{j,\overline{k}}\in C_{n-k,j-l}}$$
$$E[F^{*k-l}(x - Y_{i_1} - \ldots - Y_{i_l})F^{*j-l}(x - Y_{i_1} - \ldots - Y_{i_l})]$$
$$-nF^{*j}(x)F^{*k}(x)$$

Let n go to infinity to obtain:

$$\overline{\lim_{n\to\infty}} P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}(1-\rho)^2\sum_{k=L+1}^{\infty}\sum_{j=L+1}^{\infty}\rho^{k+j}\sigma^{kj}$$

Hence

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{\infty} a_k \leq t\right) \leq \Phi((t+\epsilon)/\sigma_L) + \frac{1}{\epsilon^2}(1-\rho)^2\sum_{k=L+1}^{\infty}\sum_{j=L+1}^{\infty}\rho^{k+j}\sigma^{kj}$$

Letting now L tend to infinity and ϵ tend to zero in such a way that the quantity $\frac{1}{\epsilon^2}(1-\rho)^2\sum_{k=L+1}^{\infty}\sum_{j=L+1}^{\infty}\rho^{k+j}\sigma^{kj}$ tends to zero, we finally obtain

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{\infty} a_k \leq t\right) \leq \Phi(t/\sigma_\infty) \qquad (12)$$

where σ_∞ is obtained from (11) by letting L tend to infinity.

To prove the reverse inequality, let us consider the following relationships

$$
\begin{aligned}
P\left(\sum_{k=1}^{\infty} a_k \le t\right) &= P\left(\sum_{k=1}^{L} a_k \le t - \sum_{k=L+1}^{\infty} a_k\right) \\
&\ge P\left(\sum_{k=1}^{L} a_k \le t - \sum_{k=L+1}^{\infty} a_k, \left|\sum_{k=L+1}^{\infty} a_k\right| \le \epsilon\right) \\
&\ge P\left(\sum_{k=1}^{L} a_k \le t - \epsilon, \left|\sum_{k=L+1}^{\infty} a_k\right| \le \epsilon\right) \\
&\ge P\left(\sum_{k=1}^{L} a_k \le t - \epsilon\right) + P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| \le \epsilon\right) - 1 \\
&= P\left(\sum_{k=1}^{L} a_k \le t - \epsilon\right) - P\left(\left|\sum_{k=L+1}^{\infty} a_k\right| > \epsilon\right)
\end{aligned}
$$

Note that the last two relationships are easily obtained on the basis of the following simple consideration. Let A and B be two sets; it is well known that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, from which we obtain

$$
\begin{aligned}
P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\
&\ge P(A) + P(B) - 1
\end{aligned}
$$

and $P(A) + P(B) - 1 \le P(A \cap B) \le \min(P(A), P(B))$.

As before, it is now easy to prove that

$$
\lim_{n\to\infty} P\left(\sum_{k=1}^{\infty} a_k \le t\right) \ge \Phi(t/\sigma_\infty) \tag{13}
$$

and from (12) and (13) the relationship

$$
\lim_{n\to\infty} P\left(\sum_{k=1}^{\infty} a_k \le t\right) = \Phi(t/\sigma_\infty)
$$

The same technique applies, with small changes, to prove the convergence of all finite-dimensional distributions of $Z_n(\cdot)$, and this concludes Step 1.

Step 2 (Weak convergence of the process). We want to state weak convergence of the process in the space D of right-continuous functions with left-hand limits, endowed with the Skorokhod topology. To this purpose, we follow ideas in Harel et al (1995); the same reasoning applied by these authors to the empirical renewal process may in fact be applied in this specific context.

The starting point is the well known fact that the empirical process $\sqrt{n}(F_n - F)$ converges in distribution to $B^\circ \circ F$ in the space D. Here, B° denotes the Brownian bridge and "$\circ$" denotes composition of functions. We want to deduce from this the corresponding result for $\sqrt{n}(\Psi_n - \Psi)$.

The main problem is that Z_n is a nonlinear function of the empirical process. The solution to such a problem is found in the following lines, which provide calculations for a linearization of the stochastic process $Z_n(x) = \sqrt{n}(\Psi_n(x) - \Psi(x))$. Once more, let us consider the following decomposition of the process $Z_n(x)$:

$$\begin{aligned} Z_n(x) &= U_n(x) + V_n(x) \\ &= \sqrt{n}(1-\hat{\rho}) \sum_{k=1}^{\infty} \hat{\rho}^k (\overline{F}_n^{*k}(x) - \overline{F}^{*k}(x)) \\ &\quad + \sqrt{n} \sum_{k=1}^{\infty} [(1-\hat{\rho})\hat{\rho}^k - (1-\rho)\rho^k] \overline{F}^{*k}(x) \end{aligned}$$

We just take in consideration the first term in the summation, since the proof of weak convergence of the second term is trivial. Moreover, to simplify the notation, let us consider F^{*k} instead of $\overline{F}^{*k}$ and, therefore,

$$\begin{aligned} \widetilde{U}_n &= \sqrt{n} \sum_{k=1}^{\infty} (1-\hat{\rho})\hat{\rho}^k (F_n^{*k} - F^{*k}) \\ &= \sqrt{n}(F_n - F) * \sum_{k=1}^{\infty} (1-\hat{\rho})\hat{\rho}^k (F_n^{*k-1} + F_n^{*k-2} * F \\ &\quad + \cdots + F_n * F^{*k-2} + F^{*k-1}) \\ &= \sqrt{n}(F_n - F) * \sum_{k=1}^{\infty} (1-\hat{\rho})\hat{\rho}^k F_n^{*k} * \sum_{k=1}^{\infty} (1-\hat{\rho})\hat{\rho}^k F^{*k} \\ &= \sqrt{n}(F_n - F) * \overline{\Psi}_n * \overline{\overline{\Psi}}_n \qquad (14) \\ &= \sqrt{n}(F_n - F) * \overline{\overline{\Psi}}_n * \overline{\overline{\Psi}}_n + \sqrt{n}(F_n - F) * (\overline{\Psi}_n - \overline{\overline{\Psi}}_n) * \overline{\overline{\Psi}}_n \end{aligned}$$

Observe that, by (14),

$$\sqrt{n}(\overline{\Psi}_n - \overline{\overline{\Psi}}_n) = \widetilde{U}_n = \sqrt{n}(F_n - F) * \overline{\Psi}_n * \overline{\overline{\Psi}}_n$$

We easily obtain:

$$\begin{aligned} \widetilde{U}_n &= \sqrt{n}(F_n - F) * \overline{\overline{\Psi}}_n * \overline{\overline{\Psi}}_n + \sqrt{n}(F_n - F) * \overline{\Psi}_n * \overline{\overline{\Psi}}_n * (F_n - F) * \overline{\overline{\Psi}}_n \\ &= \sqrt{n}(F_n - F) * \overline{\overline{\Psi}}_n * \overline{\overline{\Psi}}_n + \sqrt{n}(F_n - F) * (F_n - F) * \overline{\Psi}_n * \overline{\overline{\Psi}}_n * \overline{\overline{\Psi}}_n \end{aligned}$$

Let $G = \overline{\overline{\Psi}}_n * \overline{\overline{\Psi}}_n$. We finally obtain:

$$\widetilde{U}_n = \sqrt{n}(F_n - F) * G + \sqrt{n}(F_n - F)^{*2} * \overline{\Psi}_n * G$$

We have decomposed the process into two terms: a first term which is linear in the empirical process and a second term which can be shown to be negligible as $n \to \infty$, simply following the same lines as in Harel et al. (1995). By Theorem 3.2 in Harel et al (1995), we can definitely state weak convergence of the process.

We do not provide here a detailed proof since it proceeds in the same way as the proofs of Theorem 3.2 and Theorem 4.1 in Harel et al (1995).

Step 3 (Identification of the covariance kernel). To complete the proof of the theorem, we just need to evaluate the covariance kernel of the stochastic process. It is awkward and we need some other preliminary notations to make it readable. Observe that

$$\begin{aligned} E[Z_n(x)Z_n(y)] &= E[(U_n(x)+V_n(x))(U_n(y)+V_n(y))] \\ &= E[U_n(x)U_n(y)] + E[U_n(x)V_n(y)] \\ &\quad + E[V_n(x)U_n(y)] + E[V_n(x)V_n(y)] \end{aligned} \tag{15}$$

and consider each of the four components separately. We have already computed the first and the last term on the right hand side of (15) (see Lemma 1 and Lemma 2); let us compute $E[U_n(x)V_n(y)]$

$$\begin{aligned} E[U_n(x)V_n(y)] &= E\Bigg[\Bigg((1-\rho)\sum_{k=1}^{\infty}\rho^k\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))\Bigg) \\ &\qquad \Bigg(\sqrt{n}\Bigg(\frac{1}{\mu_T}(\overline{X}_n-\mu)-\frac{\mu}{\mu_T^2}(\overline{T}_n-\mu_T)\Bigg)\sum_{j=1}^{\infty}f_j^{'}(\rho)\overline{F}^{*j}(y)\Bigg)\Bigg] \\ &= \frac{1}{\mu_T}(1-\rho)\sum_{k=1}^{\infty}\rho^k\sum_{j=1}^{\infty}f_j^{'}(\rho)\overline{F}^{*j}(y)E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x)) \\ &\qquad \sqrt{n}(\overline{X}_n-\mu)] - \frac{\mu}{\mu_T^2}(1-\rho)\sum_{k=1}^{\infty}\rho^k\sum_{j=1}^{\infty}f_j^{'}(\rho)\overline{F}^{*j}(y) \\ &\qquad \times E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))\sqrt{n}(\overline{T}_n-\mu_T)] \end{aligned} \tag{16}$$

Taking into account that $E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))\sqrt{n}(\overline{T}_n-\mu_T)] = 0$, we obtain that (16) is equal to

$$\frac{(1-\rho)}{\mu_T}\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\rho^k f_j^{'}(\rho)\overline{F}^{*j}(y)E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))\sqrt{n}(\overline{X}_n-\mu)]$$

The computation of this last expression is not difficult. First of all, we have

$$\begin{aligned} E[\sqrt{n}(\overline{F}_n^{*k}(x)-\overline{F}^{*k}(x))\sqrt{n}(\overline{X}_n-\mu)] &= nE[\overline{F}_n^{*k}(x)\overline{X}_n-\overline{F}_n^{*k}(x)\mu \\ &\quad -\overline{F}^{*k}(x)\overline{X}_n+\overline{F}^{*k}(x)\mu] \\ &= nE[\overline{F}_n^{*k}(x)\overline{X}_n]-n\mu\overline{F}^{*k}(x) \end{aligned}$$

Furthermore

$$\begin{aligned} E[\overline{F}_n^{*k}(x)\overline{X}_n] &= E\left[\binom{n}{k}^{-1}\sum_{\underline{c}_k\in C_{n,k}}I_{(Y_{i_1}+\ldots+Y_{i_k}>x)}\frac{1}{n}\sum_{j=1}^{n}X_j\right] \\ &= \frac{1}{n}\frac{1}{\binom{n}{k}}E\left[\sum_{j=1}^{n}X_j\sum_{\underline{c}_k\in C_{n,k}}I_{(Y_{i_1}+\ldots+Y_{i_k}>x)}\right] \end{aligned}$$

Since $C_{n,k} = C^j_{n,k} \bigcup C^{\bar{j}}_{n,k}$, where $C^j_{n,k}$ is the set of all $\binom{n-1}{k-1}$ combinations of n elements of class k containing j and $C^{\bar{j}}_{n,k}$ is the set of all $\binom{n-1}{k}$ combinations of n elements of class k that does not contain j, we obtain:

$$
\begin{aligned}
E[\overline{F}^{*k}(x)\overline{X}_n] &= \frac{1}{n}\frac{1}{\binom{n}{k}} E\left[\sum_{j=1}^{n} X_j \left(\sum_{\underline{c}_k \in C^j_{n,k}} I_{(Y_{i_1}+\ldots+Y_{i_{k-1}}+Y_j>x)}\right.\right.\\
&\quad \left.\left. + \sum_{\underline{c}_k \in C^{\bar{j}}_{n,k}} I_{(Y_{i_1}+\ldots+Y_{i_{k-1}}+Y_{i_k}>x)}\right)\right]\\
&= \frac{1}{n}\frac{1}{\binom{n}{k}} \sum_{j=1}^{n} \left\{ \sum_{\underline{c}_k \in C^j_{n,k}} E[X_j I_{(Y_{i_1}+\ldots+Y_{i_{k-1}}+Y_j>x)}] \right.\\
&\quad \left. + \sum_{\underline{c}_k \in C^{\bar{j}}_{n,k}} E[X_j I_{(Y_{i_1}+\ldots+Y_{i_{k-1}}+Y_{i_k}>x)}] \right\}\\
&= \frac{1}{n}\frac{1}{\binom{n}{k}} \sum_{j=1}^{n} \left\{ \binom{n-1}{k-1} E[X_j \overline{F}^{*k-1}(x-Y_j)] + \binom{n-1}{k} \mu \overline{F}^{*k}(x) \right\}
\end{aligned}
$$

We can definitely state that

$$
\begin{aligned}
E[U_n(x)V_n(y)] = \frac{(1-\rho)}{\mu_T} \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^k f_j'(\rho)\overline{F}^{*j}(y)\\
\left\{ \frac{1}{\binom{n}{k}} \sum_{h=1}^{n} \left[\binom{n-1}{k-1} E[X_h \overline{F}^{*k-1}(x-Y_h)] \right.\right.\\
\left.\left. + \binom{n-1}{k} \mu \overline{F}^{*k}(x) \right] - n\mu\overline{F}^{*k}(x) \right\}
\end{aligned}
$$

By adding all components together in (15), we obtain the following expression:

$$
\begin{aligned}
E[Z_n(x)Z_n(y)] = (1-\rho)^2 \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^{k+j} \Big\{ n \binom{n}{k}^{-1} \binom{n}{j}^{-1} & \sum_{\underline{c}_k \in C_{n,k}} \sum_{l=0}^{j} \sum_{\underline{c}_{j,k} \in C_{k,l}} \\
& \sum_{\underline{c}_{j,\overline{k}} \in C_{n-k,j-l}} E[\overline{F}^{*k-l}(x - Y_{i_1} - \ldots - Y_{i_l}) \\
& \overline{F}^{*j-l}(y - Y_{i_1} - \ldots - Y_{i_l})] - n\overline{F}^{*j}(y)\overline{F}^{*k}(x) \Big\} \\
& + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_k'(\rho) f_j'(\rho) \overline{F}^{*k}(x) \overline{F}^{*j}(y) \\
& \times \left\{ \frac{1}{\mu_T^2} E[\sqrt{n}(\overline{X}_n - \mu)]^2 + \frac{\mu^2}{\mu_T^4} E[\sqrt{n}(\overline{T}_n - \mu_T)]^2 \right\} \\
& +2 \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} (1-\rho)\rho^k \frac{1}{\mu_T} f_j'(\rho) \overline{F}^{*j}(y) \left\{ \frac{1}{\binom{n}{k}} \sum_{h=1}^{n} \left[\binom{n-1}{k-1} \right.\right. \\
& \left.\left. E[X_h \overline{F}^{*k-1}(x - Y_h)] + \binom{n-1}{k} \mu \overline{F}^{*k}(x) \right] - n\mu\overline{F}^{*k}(x) \right\}
\end{aligned}
$$

Now, let n go to infinity to finally obtain the following asymptotic expression:

$$
\begin{aligned}
E[Z(x)Z(y)] = (1-\rho)^2 & \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^{k+j}\, k\, j\, Cov(\overline{F}^{*k-1}(x - Y_1), \overline{F}^{*j-1}(y - Y_1)) \\
& + \left(\frac{\sigma^2}{\mu_T^2} + \frac{\mu^2}{\mu_T^4}\sigma_T^2 \right) \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_k'(\rho) f_j'(\rho) \overline{F}^{*k}(x) \overline{F}^{*j}(y) \\
& +2 \frac{1}{\mu_T}(1-\rho) \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \rho^k f_j'(\rho) \overline{F}^{*j}(y) \\
& \times Cov(X_1, \overline{F}^{*k-1}(x - Y_1))
\end{aligned} \tag{17}
$$

□

Remark 5. The estimator for the probability of ruin described in Section 2 has been constructed by replacing the true k-fold convolution of the claim size d.f. F by the estimator introduced by Frees (1986b). Instead of doing this, one could estimate F by the empirical d.f. $\tilde{F}_n^{*1}(x) = F_n^{*1}(x)$ and then define recursively estimates of $F^{*k}(x)$ by the relationship

$$\tilde{F}_n^{*k}(x) = \int \tilde{F}_n^{*k-1}(x-u)d\tilde{F}_n^{*1}(u)$$

that is considering the k-fold convolution of the empirical d.f.. As observed by Frees, although $\tilde{F}_n^{*k}(x)$ is a biased estimate of $F^{*k}(x)$, it has the possible advantage of being the nonparametric maximum likelihood estimator.

Furthermore, it is a V-statistic and thus it is closely related to the U-statistic $F_n^{*k}(x)$. Remind that U-statistics and V-statistics have the same asymptotic properties.

3. Quantile estimation

In this section, we propose an estimator for the p-th quantile of the ruin probability function. For $0 \leq p < 1$, define the p-th quantile ξ_p of $\Psi(\cdot)$ as

$$\xi_p = \Psi^{-1}(p) = \inf\{x \geq 0 : \Psi(x) \geq p\}$$

A natural estimate $\widehat{\xi}_p$ of ξ_p can be defined as

$$\widehat{\xi}_p = \widehat{\Psi}_n^{-1}(p) = \inf\{x \geq 0 : \Psi_n(x) \geq p\}$$

The main asymptotic properties of the proposed estimator are presented in the following two theorems.

Theorem 4 *Assume that ξ_p, $0 \leq p < 1$, is uniquely defined, i.e. for every $\epsilon > 0$, $\Psi(\xi_p + \epsilon) < \Psi(\xi_p) = p < \Psi(\xi_p - \epsilon)$. Then, as $n \to \infty$*

$$\widehat{\xi}_p \overset{a.s.}{\to} \xi_p$$

Proof $\Psi(x)$ is a decreasing function and the equation $\Psi(x) = p$ possesses a single root, namely ξ_p. From the uniqueness of ξ_p, for every $\epsilon > 0$

$$\delta_\epsilon = \min\{p - \Psi(\xi_p + \epsilon), \Psi(\xi_p - \epsilon) - p\} > 0$$

We note that

$$\begin{aligned} Pr\{\Psi_n(\xi_p - \epsilon) - p > 0\} &= Pr\{\Psi(\xi_p - \epsilon) - \Psi_n(\xi_p - \epsilon) < \Psi(\xi_p - \epsilon) - p\} \\ &\geq Pr\{\Psi(\xi_p - \epsilon) - \Psi_n(\xi_p - \epsilon) < \delta_\epsilon\} \\ &\geq Pr\{|\Psi_n(\xi_p - \epsilon) - \Psi(\xi_p - \epsilon)| < \delta_\epsilon\} \qquad (18) \end{aligned}$$

Taking into account that Ψ_n is a consistent estimator of Ψ, for every $\delta_\epsilon > 0$, (18) converges to 1 as $n \to \infty$. Therefore

$$Pr\{\Psi_n(\xi_p - \epsilon) - p > 0\} \to 1 \qquad (n \to \infty)$$

In the same way we can show that

$$Pr\{p - \Psi_n(\xi_p + \epsilon) < 0\} \to 0 \qquad (n \to \infty)$$

Then, as $n \to \infty$

$$Pr\{\Psi_n(\xi_p + \epsilon) < p < \Psi_n(\xi_p - \epsilon)\} \to 1 \qquad \forall \epsilon > 0$$

which implies that

$$Pr\{\xi_p + \epsilon > \widehat{\xi}_p > \xi_p - \epsilon\} \to 1$$

as $n \to \infty$, $\forall \epsilon > 0$. $\square$

Theorem 5 *Let us denote by $G^{'}$ the first derivative of the function G. Assume that $G^{'}$ exists and that $G^{'}(\xi_p) \neq 0$. Under this hypothesis*

$$\sqrt{n}(\widehat{\xi}_p - \xi_p) \xrightarrow{d} N(0, \sigma_\xi^2) \tag{19}$$

where σ_ξ^2 is given by (21).

Proof $\Psi^{-1}(p) = \xi_p$ is root of the equation $G(x) = \Psi(x) - p = 0$ and $\widehat{\Psi}_n^{-1}(p) = \widehat{\xi}_p$ is root of the equation $G_n(x) = \Psi_n(x) - p = 0$. By a first order Taylor expansion of $G_n(\widehat{\xi}_p)$ at the point ξ_p, we obtain

$$G(\xi_p) = 0 = G_n(\widehat{\xi}_p) = G_n(\xi_p) + G_n^{'}(\xi_p^*)(\widehat{\xi}_p - \xi_p)$$

where ξ_p^* lies in the interval having extremes ξ_p and $\widehat{\xi}_p$. Hence

$$\sqrt{n}(\widehat{\xi}_p - \xi_p) = -G_n^{'}(\xi_p^*)^{-1}\sqrt{n}(G_n(\xi_p) - G(\xi_p))$$

Let us first prove that

$$G_n^{'}(\xi_p^*) = \Psi_n^{'}(\xi_p^*) \xrightarrow{a.s.} G^{'}(\xi_p) = \Psi^{'}(\xi_p) \tag{20}$$

Because of $\widehat{\xi}_p \xrightarrow{a.s} \xi_p$, if $\widehat{\xi}_p \in (\xi_p - \epsilon, \xi_p + \epsilon)$ we have

$$G_n^{'}(\xi_p - \epsilon) < G_n^{'}(\xi_p^*) < G_n^{'}(\xi_p + \epsilon)$$

By the law of large number

$$G_n^{'}(\xi) = \Psi_n^{'}(\xi) = \sum_{k=1}^{\infty}(1-\hat{\rho})\hat{\rho}^k[\overline{F}_n^{*k}(\xi)]^{'} \xrightarrow{a.s.} G^{'}(\xi) = \Psi^{'}(\xi) = \sum_{k=1}^{\infty}(1-\rho)\rho^k[\overline{F}^{*k}(\xi)]^{'}$$

then $G_n^{'}(\xi) \xrightarrow{a.s.} G^{'}(\xi)$ for all ξ and, for all sufficiently large n we have:

$$G^{'}(\xi_p - \epsilon) < G_n^{'}(\xi_p^*) < G^{'}(\xi_p + \epsilon)$$

which implies $G_n^{'}(\xi_p^*) \xrightarrow{a.s} G^{'}(\xi_p)$ as n goes to infinity. Since

$$\begin{aligned}\sqrt{n}(G_n(\xi_p) - G(\xi_p)) &= \sqrt{n}(\Psi_n(\xi_p) - p - \Psi(\xi_p) + p) \\ &= \sqrt{n}(\Psi_n(\xi_p) - \Psi(\xi_p))\end{aligned}$$

and because of (20), we obtain the relationship

$$\sqrt{n}(\widehat{\xi}_p - \xi_p) = -G^{'}(\xi_p)^{-1}\sqrt{n}(\Psi_n(\xi_p) - \Psi(\xi_p))$$

Taking into account the fundamental result in Theorem 3 and by a simple application of the delta method, (19) easily follows. In particular, we obtain the following expression for the variance:

$$\sigma_\xi^2 = G^{'}(\xi_p)^{-2}\Bigg\{(1-\rho)^2\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\rho^{k+j}\,k\,j\,Cov(\overline{F}^{*k-1}(\xi_p-Y_1),\overline{F}^{*j-1}(\xi_p-Y_1))$$
$$+\left(\frac{\sigma^2}{\mu_T^2}+\frac{\mu^2}{\mu_T^4}\sigma_T^2\right)\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}f_k^{'}(\rho)f_j^{'}(\rho)\overline{F}^{*k}(\xi_p)\overline{F}^{*j}(\xi_p)$$
$$+2\frac{1}{\mu_T}(1-\rho)\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\rho^k f_j^{'}(\rho)\overline{F}^{*j}(\xi_p)Cov(X_1,\overline{F}^{*k-1}(\xi_p-Y_1))\Bigg\} \qquad (21)$$

□

4. Bootstrap confidence region

It is possible to construct a confidence region for the unknown probability of ruin Ψ. The development of such a confidence region follows that in Grübel and Pitts (1993) and Pitts (1994).

The interest in studying bootstrap confidence regions comes from the complicated structure of the limiting Gaussian process Z in Theorem 3, that does not allow to construct confidence regions for $\Psi(x)$ in any easy way.

Let $z \in \mathbb{R}$ and define $R_n(z) = P(\sqrt{n}||\Psi_n - \Psi|| \leq z)$, where $||f|| = sup_t|f(t)|$ is the supremum norm. If known, it could be used to form a confidence region for Ψ. Recall that $\sqrt{n}(\Psi_n - \Psi) \xrightarrow{d} Z$ (Theorem 3) and define, for $z \in \mathbb{R}$, $R_Z(z) = P(||Z|| \leq z)$. By the continuous mapping theorem (see, for example, Pollard (1984)), $R_n(z) \to R_Z(z)$ as $n \to \infty$, for all continuity points z of R_Z.

Using the α-quantile $q(\alpha)$ of R_Z we could obtain asymptotic confidence regions for Ψ. The problem is that the structure of Z is complicated and it is not easy to see how the α-quantile $q(\alpha)$ of R_Z can be obtained. This is the reason for which we construct Bootstrap confidence regions as a good alternative to classical asymptotic confidence regions.

We adopt a bootstrap scheme which consists in generating m_n bootstrap inter-arrival time samples $(T_{1,j}^*, \ldots, T_{n,j}^*)$, $(j = 1, \ldots, m_n)$, and m_n independent bootstrap claim samples $(X_{1,j}^*, \ldots, X_{n,j}^*)$, $(j = 1, \ldots, m_n)$. We then define the m_n quantities $V_{n,j}^* = \sqrt{n}||\Psi_{n,j}^* - \Psi_n||$, $(j = 1, \ldots, m_n)$, where Ψ^* runs through the empirical d.f.s obtained by taking samples of size n from $\hat{F}_n$.

The bootstrap estimator $\hat{R}_n$ of the d.f. of $\sqrt{n}||\Psi_n - \Psi||$ is defined as the empirical d.f. for the m_n values of $\sqrt{n}||\Psi_{n,j}^* - \Psi_n||$. In symbols:

$$\hat{R}_n(x) = m_n^{-1}\sum_{j=1}^{m_n} I_{(-\infty,x]}(\sqrt{n}||\Psi_{n,j}^* - \Psi_n||)$$

Let $\hat{q}_n(\alpha)$ be the α-quantile of $\hat{R}_n$

$$\hat{q}_n(\alpha) = \hat{R}_n^{-1}(\alpha) = \inf\{x : \hat{R}_n(x) \geq \alpha\}$$

It is possible to show that confidence regions constructed from $\hat{q}_n(\alpha)$ give asymptotically correct coverage probabilities. In practice, $\hat{R}_n$ and hence $\hat{q}_n(\alpha)$ are approximated by Monte Carlo methods.

Remark 1. The bootstrap confidence band obtained here is of constant width since we have considered the supremum norm. It would be more appropriate to have non-constant width simultaneous confidence bands for d.f. estimators and this is an area for further investigation. One way to achieve this aim would be to use weighted spaces of functions, as in Pitts (1994), as one setting in which non-constant width bootstrap confidence bands can be obtained.

5. Simulation example

We illustrate the finite sample behavior of the proposed estimator $\Psi_n(x)$ for sample size $n = 100$ by the following example.

Example Both claim amount and inter-arrival times distributions are exponential with mean value equal to 1 and 10, respectively, so that $\rho = 0.1$. The premium rate c equals 1 and an interval of values for the initial surplus x is considered. A sample of size $n = 100$ has been generated from F and this has been used to construct our estimate of the ruin probability function Ψ_n.

Figure 1 shows the resulting ruin function estimate Ψ_n together with the “true” Ψ and an approximate 95% bootstrap confidence region obtained for the unknown Ψ. The band is calculated from $m_n = 200$ bootstrap repetitions. The “true” Ψ is computed using the closed form expression available for Ψ when dealing with the Erlang model.

In the example, the distribution of $\Psi_n(x)$ is quite well concentrated around the “true” value $\Psi(x)$.

Some more considerations are in order. We used a “truncated” version of the estimator:

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

where the parameter m, as in Frees (1986b), is a positive integer depending on n, such that $m \leq n$ and $m \uparrow \infty$ as $n \uparrow \infty$. A choice for the parameter m has to be done.

In many situations, this choice is dictated by practical considerations and often simply $m = n$. However, the quantity $(1 - \rho)\rho^k$ dies out quickly as k approaches infinity as well as the convolution $F^{*k}(x)$ (see Frees (1986b)). This implies that typically m can be small compared to the sample size.

This is important since the amount of computations increases quickly as m increases and the numerical computation of our estimator is cumbersome even for small values of m. This is the reason for which the number of Monte Carlo

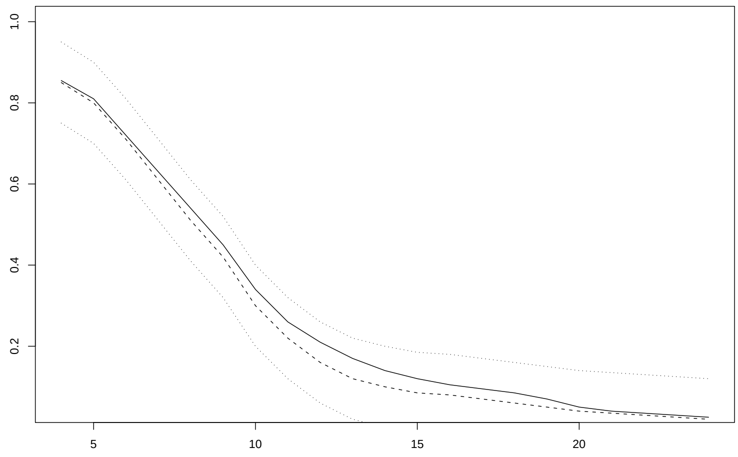

Figure 1: *Estimate Ψ_n (long dashes), "true" Ψ (full line) and bootstrap confidence band (dotted line)*

simulations in the example is small. A larger scale Monte Carlo study with 500 or more trials may improve the results but is a computer time consuming task.

References

1. Asmussen S (2000) *Ruin probability*. World scientific, Singapore .
2. Billingsley P (1968) *Convergence of probability measures*. John Wiley, New York.
3. Bühlmann H (1970) *Mathematical methods in risk theory*. Springer-Verlag, Berlin.
4. Cramér H (1930) *On the mathematical theory of risk*. Skandia Jubilee Volume, Stockholm.
5. Cramér H (1955) *Collective risk theory*. The Jubilee Volume of Skandia Insurance Company, Stockholm.
6. Conti PL, Masiello E (2006a) Nonparametrical statistical analysis of ruin probability under conditions of "small" and "large" claims. Atti della XLIII Riunione Scientifica SIS, Torino, 14-16 giugno 2006.
7. Conti PL, Masiello E (2006b) Nonparametrical statistical analysis of ruin probability under conditions of "small" and "large" claims. COMPSTAT 2006 - Proceedings in Computational Statistics, 17th Symposium Held in Rome, Italy, A. Rizzi and M. Vichi (eds).
8. Croux K, Veraverbeke N (1990) Non-parametric estimators for the probability of ruin. Insurance: Math. Econom. **9**:127–130
9. Frees EW (1986a) Nonparametric estimation of the probability of ruin. Astin Bulletin **16**:81–90
10. Frees EW (1986b) Nonparametric renewal function estimation. The Annals of Statistics **14**:1366–1378
11. Gerber H (1979) *An introduction to mathematical risk theory*. S.S. Huebner Foundation Monographs, University of Pennsylvania
12. Gill RD (1989) Non and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). Scand. J. Statist. **16**:97–128
13. Grandell J (1991) *Aspects of risk theory*. Springer-Verlag, New York.
14. Grübel R (1989) Stochastic models as functionals: Some remarks on the renewal case. J. Appl. Probab. **26**:296–303
15. Grübel R, Pitts SM (1993) Nonparametric estimation in renewal theory I: The empirical renewal function. Ann. Statist. **21(3)**:1431–1451
16. Harel M, O' Cinneide CA, Schneider H (1995) Asymptotics of the sample renewal function. Journal of mathematical analysis and applications **189**:240–255
17. Hipp C (1989a) Efficient estimators for ruin probabilities. Proc. Fourth Prague Symp. on Asymptotic Statistics. Ed. P. Mandl and M. Huskova. Charles University, Prague, 259–268
18. Hipp C (1989b) Estimators and bootstrap confidence intervals for ruin probabilities. Astin Bulletin **19**:57–70
19. Lundberg F (1903) Approximerad Framställning av Sannolikhetsfunktionen. Aterförsäkring av Kollektivrisker. II Almqvist & Wiksell, Uppsala
20. Lundberg F (1926) Försäkringsteknisk Riskutjämmong. F. Englunds Boktryckeri AB, Stockholm
21. Pitts SM (1994) Nonparametric estimation of compound distributions with applications in insurance. Ann. Inst. Statist. Math. **46(3)**:537–555
22. Pollard D (1984) *Convergence of stochastic processes*. Springer, New York
23. Serfling RJ (1980) *Approximation theorems of mathematical statistics*. John Wiley, New York
24. Van der Vaart AW, Wellner JA (1996) *Weak convergence and empirical processes*. Springer