

NONPARAMETRIC ESTIMATION OF THE PROBABILITY OF RUIN

BY EDWARD W. FREES*

University of Wisconsin-Madison

ABSTRACT

The finite and infinite horizon time probability of ruin are important parameters in the study of actuarial risk theory. This paper introduces procedures for directly estimating these key parameters from a random sample of observations without assumptions as to the parametric form of the distribution from which the observations arise. The estimators introduced apply to most of the classical models in which ruin probabilities are used and also apply to a much broader class of models. The procedures are based on the concept of sample reuse, an old idea in statistics which is becoming more popular due to the widespread availability of high speed computers. In this paper, the almost sure consistency of the estimators is established. Further, finite sample properties of the estimators are investigated in a simulation study.

KEYWORDS

Reverse martingale, sample reuse.

1. INTRODUCTION

Let (X_i, Y_i) $i = 1, 2, \dots$ be i.i.d. random vectors with joint bivariate distribution function F . Use the random variable X_i to represent the i th claim amount and the nonnegative random variable Y_i to represent the i th interarrival time between claims. For $t > 0$, define the number of claims by time t as

$$N(t) = \sum_{k=1} I(Y_1 + \dots + Y_k \leq t)$$

where $I(\cdot)$ is the indicator function. Premiums are assumed to arrive at a known steady rate, say, P per unit time. Thus, the amount that claims exceed premiums by time t is

$$(1.1) \quad U(t) = \sum_{k=1}^{N(t)} X_k - Pt.$$

Here we interpret the sum $\sum_{k=1}^{N(t)}$ to be zero when $N(t) = 0$. We are primarily interested in the probability that $U(t)$ exceeds an initial reserve u at some time t prior to or at T , the horizon time. This probability can also be defined by

$$(1.2) \quad \psi(u, T) = P(\sup_{0 \leq t \leq T} U(t) > u).$$

* The author would like to thank James C. Hickman and Robert B. Miller for helpful comments on an earlier draft of this paper. Research partially supported by ARO Contract No. DAAG 29-80-C-0041.

We are also interested in the probability that $U(t)$ eventually exceeds an initial reserve u ,

$$(1.3) \quad \psi(u) = \psi(u, \infty) = \lim_{T \rightarrow \infty} \psi(u, T).$$

These probabilities, $\psi(u, T)$ and $\psi(u)$, are called, respectively, the finite and infinite horizon time probability of ruin.

The probability of ruin is a key parameter in the collective theory of risk and has received considerable attention over the years. For several different types of introduction to this theory, see FELLER (1971), BEARD et al. (1984), GERBER (1979) and BUHLMANN (1970). The behavior of the finite horizon time probability of ruin has been widely investigated not only because of its practical relevance but also because explicit computation of $\psi(u, T)$ is difficult except in the most trivial cases. Explicit calculations have been given for some specific forms of F , cf. SEAL (1978) and THORIN and WIKSTAD (1976). However, use of these explicit solutions has been limited due to their complexity and their dependence on a specific form of F . Because of these difficulties, papers giving approximations of $\psi(u, T)$ suggested by limit theorems (e.g., as $u \rightarrow \infty$, $T \rightarrow \infty$) are abundant in the literature. The most successful of these approximations seem to be the diffusion approximations given by SIEGMUND (1979) and applied by ASMUSSEN (1984). See LALLEY (1984) for a refinement of Siegmund's work. Another class of methods for calculating $\psi(u, T)$ is the Monte Carlo method. Surprisingly, the Monte Carlo method has received only limited attention in the risk theory literature. For some accounts, see BEARD et al. (1984) and SEAL (1978). Note that this method does not depend on a specific form of F but does depend on complete knowledge of F .

The approximations of $\psi(u, T)$ and $\psi(u)$ given in this paper are different in nature from those sketched above and are inspired by the concept of *sample reuse*. Sample reuse is an old idea in statistics, popularized in the nineteen-forties by HOEFFDING (1948) and more recently by EFRON (cf. 1982). The idea for our applications is as follows. Consider the random variable, $Z = \sup_{0 \leq t \leq T} U(t)$. Since the distribution of $U(t)$ is completely determined by F , then the distribution of Z is completely determined by F . Note that from (1.2), $\psi(u, T) = 1 - P(Z \leq u)$. By the usual multivariate Glivenko-Cantelli Theorem, knowledge of (X_i, Y_i) $i = 1, 2, \dots$ completely determines F and thus, for a sufficiently large sample size n , from $\{(X_i, Y_i)\}_{i=1}^n$ we can build a good approximation to F , the usual (multivariate) empirical distribution function. This raises the natural question of how to build a reasonable estimator of $\psi(u, T)$ based on a random sample of size n , $\{(X_i, Y_i)\}_{i=1}^n$. The purposes of this paper are to argue that this is an important question and to develop estimators of $\psi(u, T)$ (and $\psi(u)$) that possess desirable properties.

In Section 2, estimators of $\psi(u, T)$ and $\psi(u)$ are introduced. The almost sure (a.s.) consistency of these estimators is proved in Section 3. A small simulation study is given in Section 4. Because of the nature of the approximations, the assumptions made in our development are different and more general than the

usual risk theory assumptions. In Section 5 we discuss these differences and make other concluding remarks.

2. ESTIMATORS

Let $\{(X_i, Y_i)\}_{i=1}^n$ be a random sample from a population with distribution function F . Let $\{a_{1j}, a_{2j}, \dots, a_{nj}\} = A_j$ be the j th permutation of $\{1, 2, \dots, n\}$, $j = 1, \dots, n!$. We intend to *reuse* the sample by considering reordered pairs $\{(X_{a_{ij}}, Y_{a_{ij}})\}_{i=1}^n$, $j = 1, \dots, n!$

Define $S_n = Y_1 + \dots + Y_n$. For the j th permutation, the number of claims by time t is

$$\begin{aligned} NA_j(t) &= \sum_{k=1}^n I(Y_{a_{1j}} + \dots + Y_{a_{kj}} \leq t), & t \leq S_n \\ &= n, & t > S_n, \end{aligned}$$

and the amount that claims exceeds premiums by time t is

$$UA_j(t) = \sum_{k=1}^{NA_j(t)} X_{a_{kj}} - Pt.$$

For the j th permutation, we have ruin if

$$(2.1) \quad I(\sup UA_j(t) > u) = 1$$

where the supremum is over the set $\{t: 0 \leq t \leq \min(S_n, T)\}$. Note that to compute the function in (2.1) one does not need to evaluate $UA_j(t)$ at all $t \in [0, \min(S_n, T)]$ but only at the random time points $0, Y_{a_{1j}}, Y_{a_{1j}} + Y_{a_{2j}}, \dots, Y_{a_{1j}} + \dots + Y_{a_{nj}}$. The first type of estimator we consider is the average over all permutations,

$$(2.2) \quad \psi_n(u, T) = (n!)^{-1} \sum_p I(\sup UA_j(t) > u)$$

and

$$(2.3) \quad \psi_n(u) = \psi_n(u, S_n)$$

where \sum_p is the sum over all permutations of $\{1, 2, \dots, n\}$. Note that we could alternatively define

$$\psi_n(u) = \psi_n(u, \infty) = \lim_{T \rightarrow \infty} \psi_n(u, T).$$

The consistency of these estimators is provided in the following

THEOREM 2.1. *Suppose that $EY > 0$. Then, for each T ,*

$$(2.4) \quad \lim_{n \rightarrow \infty} \psi_n(u, T) = \psi(u, T) \quad \text{a.s.}$$

and

$$(2.5) \quad \lim_{n \rightarrow \infty} \psi_n(u) = \psi(u) \quad \text{a.s.}$$

By the SLLN, the requirement $EY > 0$ ensures that $S_n \rightarrow \infty$ a.s. The proof, given in Section 3, is based on the idea that ψ_n can be shown to be a reverse martingale plus negligible terms. We remark that the estimators ψ_n defined in (2.2) and (2.3) each require the evaluation of $n!$ indicators of ruin, an extensive amount of computations even for moderate sample sizes (say, $n \geq 10$). Because of this computational difficulty, we now introduce a *sample reuse* estimator of the probability of ruin. The sample reuse methodology, popularized by EFRON (cf. 1982), is also computer-intensive but does not require a prohibitive amount of computation.

Let $B = B(n)$ be a positive integer depending on n such that $B \rightarrow \infty$ as $n \rightarrow \infty$. Based on the observed sample $\{(X_i, Y_i)\}_{i=1}^n$, we draw B independent (conditional on $\{(X_i, Y_i)\}_{i=1}^n$) realizations of ψ_n in the following two steps. For $b = 1, \dots, B$,

Step 1. Make n draws without replacement from $\{(X_i, Y_i)\}_{i=1}^n$ to get (X_i^{*b}, Y_i^{*b}) for $i = 1, \dots, n$.

Step 2. Define

$$N^{*b}(t) = \sum_{k=1}^n I(Y_1^{*b} + \dots + Y_k^{*b} \leq t)$$

and compute

$$\psi_n^{*b}(u, T) = I\left(\sup_{0 \leq t \leq \min(S_n, T)} \left(\sum_{k=1}^{N^{*b}(t)} X_k^{*b} - Pt\right) > u\right).$$

The sample reuse estimates are defined by

$$(2.6) \quad \psi_n^*(u, T) = B^{-1} \sum_{b=1}^B \psi_n^{*b}(u, T)$$

and

$$(2.7) \quad \psi_n^*(u) = \psi_n^*(u, S_n).$$

The consistency of these estimators is provided in the following

THEOREM 2.2. *Suppose $EY > 0$ and*

$$(2.8) \quad \log n = O(B^{1/2}).$$

Then, for each T ,

$$(2.9) \quad \lim_{n \rightarrow \infty} \psi_n^*(u, T) = \psi(u, T) \quad \text{a.s.}$$

and

$$(2.10) \quad \lim_{n \rightarrow \infty} \psi_n^*(u) = \psi(u) \quad \text{a.s.}$$

The proof of theorem 2.2 is given in Section 3. The condition on B in (2.8) guarantees that B grows sufficiently quickly to achieve a.s. convergence. In Section 4 we investigate the finite sample properties of ψ_n^* in a small simulation study.

3. PROOFS

Define $X_{1n}, X_{2n}, \dots, X_{nn}$ to be the order statistics of X_1, X_2, \dots, X_n and let Y_n be the claims time associated with $X_{in}, i = 1, \dots, n$. Let $G_n = \sigma((X_{in}, Y_{in}), i = 1, \dots, n, (X_i, Y_i), i > n)$ for $n \geq 1$, a nonincreasing sequence of sub σ -fields. Because all the arguments follow easily for the case $T = \infty$, we only give proofs for the a.s. consistency of $\psi_n = \psi_n(u, T)$. We now define a version of ψ_n and show that it is a reverse martingale. Later we show that this version is close to ψ_n in the appropriate sense.

For the j th permutation of $\{1, 2, \dots, n\}$, let

$$NA'_j(t) = NA_j(t) + \sum_{k>n} I(S_k \leq t)$$

be a version of the number of claims by time $t, j = 1, \dots, n!$ With $a_{kj} = k$ for $k > n$, define

$$(3.1) \quad \psi'_n(u, T) = (n!)^{-1} \sum_p I\left(\sup_{0 \leq t \leq T} \sum_{k=1}^{NA'_j(t)} X_{a_{kj}} - Pt > u\right).$$

We have the following property for this version of the probability of ruin.

LEMMA 3.1. For each $u, T, (\psi'_n(u, T), G_n)$ is a reverse martingale.

PROOF. It is easy to see that $\psi'_n = \psi'_n(u, T)$ is G_n -measurable and G_n -integrable. With $U(t)$ defined as in (1.1), we have the relation

$$(3.2) \quad \psi'_n = E\{I(\sup_{0 \leq t \leq T} U(t) > u) | G_n\}.$$

Thus,

$$\begin{aligned} E\{\psi'_n | G_{n+1}\} &= E\{E(I(\sup_{0 \leq t \leq T} U(t) > u) | G_n) | G_{n+1}\} \\ &= E\{I(\sup_{0 \leq t \leq T} U(t) > u) | G_{n+1}\} = \psi'_{n+1}. \end{aligned} \quad (\text{Q.E.D.})$$

PROOF OF THEOREM 2.1. From (3.2), we have that $E\psi'_n(u, T) = \psi(u, T)$. From lemma 3.1 and the (reverse) martingale convergence theorem, it is easy to show that

$$(3.3) \quad \lim_{n \rightarrow \infty} \psi'_n(u, T) = \psi(u, T) \quad \text{a.s.}$$

Define the stopping time

$$\tau = \inf \left\{ n \geq 1: \sum_{k=1}^n (X_k - PY_k) > u \right\}.$$

Now, it is easy to see that $\psi_n(u, T)$ and $\psi'_n(u, T)$ differ only on the set $\{n < \tau < \infty\}$. Thus

$$0 \leq \psi_n(u, T) - \psi'_n(u, T) = I(n < \tau < \infty) \rightarrow 0 \quad \text{a.s.}$$

This and (3.3) are sufficient for the proof. (Q.E.D.)

PROOF OF THEOREM 2.2. Denote $\psi_n^* = \psi_n^*(u, T)$. From (2.4), a sufficient condition for (2.9) is, for $\varepsilon > 0$,

$$\sum_n P(|\psi_n^* - \psi_n| > \varepsilon) < \infty.$$

We show

$$(3.4) \quad \sum P(\psi_n^* - \psi_n > \varepsilon) < \infty,$$

the proof of the other inequality being similar.

Define $F_n = \sigma((X_i, Y_i), i = 1, \dots, n)$ for $n \geq 1$, a nondecreasing sequence of sub σ -fields. Now, from the condition (2.8), there exists a positive constant K such that, for sufficiently large n ,

$$(3.5) \quad n^{-1} \geq \exp(-KB^{1/2}).$$

From the Markov inequality, with $\delta > \varepsilon^{-1}$, we have

$$P(\psi_n^* - \psi_n > \varepsilon) \leq \exp\{-\varepsilon\delta KB^{1/2}\} E\{\exp(\delta KB^{1/2}(\psi_n^* - \psi_n))\}.$$

From (3.5), $\exp\{-\varepsilon\delta KB^{1/2}\}$ is summable. Thus, to prove (3.4), we need only show

$$(3.6) \quad \sup_n E\{\exp(sB^{1/2}(\psi_n^* - \psi_n))\} < \infty$$

where $s = \delta K$. From (2.6), since conditionally on F_n , ψ_n^* is the mean of a binomial random variable, we have

$$\begin{aligned} E\{\exp(sB^{1/2}(\psi_n^* - \psi_n))\} &= E\{(1 - \psi_n) \exp\{-s\psi_n B^{-1/2}\} \\ &\quad + \psi_n \exp\{s(1 - \psi_n) B^{-1/2}\}\}^B \\ &= E(1 + s^2 \psi_n(1 - \psi_n)/(2B) + O(B^{-3/2}))^B \\ &\leq (1 + s^2/(8B) + O(B^{-3/2}))^B \end{aligned}$$

by a Taylor-series expansion. This is sufficient for (3.6) and hence (2.9). (Q.E.D.)

4. SIMULATION

In this section, finite sample properties of the sample reuse estimators introduced in Section 2 were investigated. A simple example was used so that calculation of exact probabilities of ruin and comparison with other studies were possible. Claim amounts were assumed to be exponentially distributed with mean 1. The claims were assumed to arrive as a Poisson process with intensity $\rho = 0.8$, i.e., interarrival times are independent and exponentially distributed with mean 1.25. The claims amount and arrival times were assumed to be independent and premiums arrive with unit intensity ($P = 1$). Under these assumptions, it can be shown that

$$(4.1) \quad \psi(u) = 0.8 \exp\{-0.2u\},$$

thus giving an easy expression for exact values of $\psi(u)$. Furthermore, in a recent study, ASMUSSEN (1984) has provided exact values of $\psi(u, T)$ and several popular approximations of $\psi(u, T)$ for various values of T .

The bias (BIAS) and root mean square error (RMSE) were used to judge the performance of the estimators. All computations were done on a VAX 11/750 owned and operated by the Department of Statistics at the University of Wisconsin-Madison. The IMSL Fortran subroutines produced the random deviates.

In Table 1 we give the results of the performance of the sample reuse estimator of the infinite horizon time probability of ruin. The tables give the criteria for ruin probabilities $\psi(u) = 1\%$, 5% , 10% , 40% and for sample sizes $n = 25, 50, 100, 150, 200$. The ruin probabilities were chosen to represent a range which is typically of interest to the actuarial community. The sample sizes were selected to represent small and moderate numbers of claims. We used $B = 100$ sample replications to compute the estimators for each simulation run.

In Table 2 we give the results of the performance of the sample reuse estimator of the finite horizon time probability of ruin. The number of sample replications are as in Table 1. To make our study comparable to the study of ASMUSSEN (1984), the level of initial reserve yielding $\psi(u) = 0.8\%$ and horizon times $T = 13.8, 41.3, 68.8, 96.4$ were selected. For this level of reserve and these times T , the exact probabilities $\psi(u, T)$ were taken from ASMUSSEN (1984). Because $\psi(u)$ is small, we used larger sample sizes, $n = 100, 150, 200, 250, 300$.

TABLE 1
SAMPLE REUSE ESTIMATOR OF $\psi(u)$

$\psi(u)$	n	BIAS	RMSE
1%	25	-0 01000	0 01000
	50	-0 01000	0 01000
	100	-0 00865	0 01366
	150	-0 00780	0 01331
	200	-0 00310	0 07100
5%	25	-0 04905	0 05037
	50	-0 04000	0 08653
	100	-0 02430	0 18292
	150	-0 01120	0.16296
	200	-0 01430	0 09928
10%	25	-0 08825	0 12187
	50	-0 05965	0 15995
	100	-0.03795	0 18523
	150	-0 00190	0 16258
	200	-0.01245	0 14340
40%	25	-0 13520	0 35943
	50	-0.08230	0 31779
	100	-0 04175	0 26445
	150	0 01305	0.27088
	200	0 00565	0.21420

TABLE 2
 SAMPLE REUSE ESTIMATOR OF $\psi(u, T)$

T	$\psi(u, T)$	n	BIAS	RMSE
13.8	0.00007	100	-0.00002	0.00071
		150	-0.00007	0.00007
		200	0.00003	0.00141
		250	0.00003	0.00099
		300	-0.00002	0.00071
41.3	0.00145	100	-0.00090	0.00436
		150	-0.00090	0.00316
		200	0.00005	0.01236
		250	0.00070	0.00821
		300	-0.00080	0.00295
68.8	0.00338	100	-0.00277	0.00509
		150	-0.00228	0.00617
		200	-0.00028	0.03057
		250	0.00107	0.01931
		300	-0.00118	0.00785
96.4	0.00491	100	-0.00441	0.00614
		150	-0.00376	0.00730
		200	0.00120	0.07068
		250	0.00135	0.02797
		300	-0.00140	0.01366

As was expected, the performance of the infinite time estimator was better the closer $\psi(u)$ was to 50%. In Table 1, the asymptotic theory comes quickly into play when $\psi(u) = 40\%$ as evidenced by the decreasing BIAS and RMSE terms with increasing sample size. For $\psi(u) = 5\%$, we see some levelling off of the RMSE term from $n = 100$ to $n = 150$. The smallest probabilities showed no evidence of levelling off for the sample sizes considered. Some attempts were made to increase B , the sampling replication number, but this did not seem to affect the performance of the estimators. See EFRON (1982) for further discussion of the selection of B .

Perhaps the most interesting fact of the simulation study was that the performance of the finite time estimators improves as T decreases. In Table 2, the asymptotic theory comes quickly into play when $T = 13.8$. For $T = 41.3$, we see the decreasing trend in the BIAS and RMSE terms beginning at a larger sample size $n = 100$. The intuition is as follows. For $T = 13.8$, on the average it requires about 11 observations to check for ruin by time 13.8 (since $11 \times 1.25 \times 1 = 13.75$). With a sample size of $n = 100$, we have approximately 9 independent and identical realizations of an indicator of ruin by time $T = 13.8$. Repeating this reasoning for $T = 41.3$, the reader can verify that we have only approximately 3 independent and identical realizations of this indicator of ruin. While the goal of the estimators introduced in Section 2 is to reuse sampling information, by increasing the horizon time T we increase the number of observations necessary to check for ruin and thus reduce the number available for resampling.

5. DISCUSSION

The estimators introduced in Section 2 are defined in terms of the classical method of collecting data, i.e., based on a random sample of size n . However, it is easy to modify these estimators for other methods of data collection and retain their important statistical properties. For example, suppose an insurance company would like to estimate the probability of ruin based on one year of observed data. Then the sample size itself is random. However, it is easy to see that the usual results on random change of time (cf. CSÓRGO and RÉVÉSZ, 1980, theorem 7.1.1) can be applied to preserve the properties stated in theorems 2.1 and 2.2. In the risk and queuing literature, this change of time is usually referred to as a transition to operational time.

The most important drawback of the estimators ψ_n and ψ_n^* is that they rely heavily on the independence of the bivariate pairs (X_i, Y_i) . While this assumption is used in most models constructed to calculate the probability of ruin, other models such as a model which uses the mixed Poisson process for the claims number process (cf. SEAL and GERBER, 1984) do not.

Except for the important assumption of independence, the estimators ψ_n and ψ_n^* and their properties are relatively free from assumptions when compared to other estimators of the probability of ruin. Both estimators are non parametric in the sense that they do not assume a particular parametric form nor knowledge of the distribution function F . While we have primarily concerned ourselves with the continuous time case the estimator is valid in the discrete time case, i.e., when Y_i is equal to some constant K with probability one. In the assumptions we have not precluded the case $P(Y_i = 0) > 0$, thus allowing for the possibility of multiple claims at any point in time. Further, we have not assumed independence between the claim arrival time and the claim amount. If this assumption is made, we conjecture that a different estimation procedure can be constructed that uses the data more efficiently, in some sense. We leave this as an open question for future research.

The sample reuse estimator ψ_n^* is computationally similar to the usual Monte Carlo procedure. However, it is different philosophically in that the Monte Carlo procedure assumes knowledge of the underlying distribution function while the sample reuse does not. This similarity suggests that dynamic factors in ruin probabilities such as interest, inflation, economic cycles, etc., that have been incorporated in Monte Carlo methods (cf. BEARD et al., 1984, Chapter 7) may be incorporated in sample reuse estimators. We leave this as an open area for future research.

REFERENCES

- ASMUSSEN, S (1984) Approximations for the Probability of Ruin within Finite Time. *Scandinavian Actuarial Journal* 31-57
- BEARD, R, PENTIKAINEN, T and PESONEN, E (1984) *Risk Theory The Stochastic Basis of Insurance* Chapman and Hall New York
- BUHLMANN, H (1970) *Mathematical Methods in Risk Theory* Springer-Verlag, Heidelberg

- CSORGO, M. and RÉVÉSZ, P. (1981) *Strong Approximations in Probability and Statistics*. Academic Press. New York
- EFRON, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM: Philadelphia.
- FELLER, W. (1971) *An Introduction to Probability Theory and Its Applications*, Vol 2. Wiley: New York.
- GERBER, H. (1979) *An Introduction to Mathematical Risk Theory* Irwin Homewood, Illinois
- HOEFFDING, W. (1948) A Class of Statistics with Asymptotically Normal Distributions *Annals of Mathematical Statistics* **19**, 293-325.
- LALLEY, S. (1984) Limit Theorems for First-Passage Times in Linear and Non-Linear Renewal Theory *Advances in Applied Probability* **16**, 766-803
- SEAL, H. (1978) *Survival Probabilities* Wiley New York.
- SEAL, H. and GERBER, H. (1984) Mixed Poisson Processes and the Probability of Ruin. *Insurance Mathematics and Economics* **3**, 189-190
- SIEGMUND, D. (1979) Corrected Diffusion Approximations in Certain Random Walk Problems *Advances in Applied Probability* **11**, 701-719
- THORIN, O. and WIKSTAD, N. (1976) Calculation of Ruin Probabilities When the Claim Distribution is Lognormal *Astin Bulletin* **9**, 231-246

EDWARD W. FREES

*School of Business, University of Wisconsin-Madison, 1155 Observatory Drive,
Madison, WI 53706, USA*