**Secondary Publication**

**Markovitch, Natalia M.; Krieger, Udo R.**

**Nonparametric Estimation of the Renewal Function by Empirical Data**

Date of secondary publication: 17.04.2026

Accepted Manuscript (Postprint), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-114764x

**Primary publication**

Markovitch, Natalia M.; Krieger, Udo R. (2006): Nonparametric Estimation of the Renewal Function by Empirical Data, in: Stochastic models : affiliated publication of the Institute for Operations Research and the Management Sciences, London: Taylor & Francis, Vol. 22, No. 2, pp. 175–199, doi: 10.1080/15326340600648922.

**Legal Notice**

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.

The license information is available online:

https://creativecommons.org/licenses/by-nc/4.0/legalcode

# NONPARAMETRIC ESTIMATION OF THE RENEWAL FUNCTION BY EMPIRICAL DATA

**Natalia M. Markovich** □ *Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*

**Udo R. Krieger** □ *Information Systems and Applied Computer Science, Otto-Friedrich University, Bamberg, Germany*

□ *We consider an estimate of the renewal function (rf) $H(t)$ using a limited number of independent observations of the interarrival times for an unknown interarrival-time distribution (itd). The nonparametric estimate is derived from the rf-representation as a series of distribution functions (dfs) of consecutive arrival times using a finite summation and approximations of the latter by empirical dfs. Due to the limited number of observed interarrival times, the estimate is accurate just for closed time intervals $[0, t]$. An important aspect is given by the selection of an optimal number of terms $k$ of the finite sum. Here two methods are proposed: (1) an a priori choice of $k$ as function of the sample size $l$ which provides almost surely (a.s.) the uniform convergence of the estimate to the rf for light- and heavy-tailed itds if the time interval is not too large, and (2) a data-dependent selection of $k$ by a bootstrap method. To evaluate both the efficiency of the estimate and the selection methods of $k$, a Monte Carlo study is performed.*

**Keywords** Bootstrap method; Nonparametric estimate; Renewal function.

**Mathematics Subject Classification** 62G05; 60K05; 62M09.

## 1. INTRODUCTION

Renewal processes have a wide range of applications in the warranty control, in the reliability analysis of technical systems and, particularly, of telecommunication networks such as high-speed packet-switched networks like the Internet. Normally, measurement facilities count the events of interest, e.g., the number of requested and transferred Web pages, incoming or outgoing calls, frames, packets or cells in consecutive time intervals of fixed length. It is important for planning and control purposes to estimate the related traffic load in terms of the mean numbers of counted events and their variances in these intervals. In such applications the renewal function (rf) $H(t)$ constitutes the basic characteristic of an underlying renewal process since by means of this function the expectation $H(t) = \mathbb{E}(N_t)$ and variance $\mathrm{var}(N_t)$ of the number of arrivals $N_t$ of the relevant events before a fixed time instant $t$ can be calculated. It is well known that the variance of the number of arrivals $N_t$ is equal to

$$\mathrm{var}(N_t) = 2\int_0^t H(t-\tau)dH(\tau) + H(t) - H^2(t)$$

(Ref.[12], p. 98, (2.3.7)). To estimate the rf, several realizations of the counting process may be required, e.g., the observations of the number of calls within several days. Here we estimate the rf using interarrival times between events of only one realization of the process. Let $F(t) = \mathbb{P}\{\tau_n < t\}$ with $F(0+) = 0$ denote the common distribution function (df) of the i.i.d. interarrival times $\{\tau_n, n = 1, 2, \dots\}$ of these events. The renewal counting process $\{N_t, t \geq 0\}$ denotes the number of events before time $t$, $N_t = \max\{n : t_n < t\}$ for $t \geq 0$, where $t_n = \sum_{i=1}^n \tau_i$, $t_0 = 0$ are the arrival times.

The rf $H(t)$ is expressed by

$$H(t) = \mathbb{E}(N_t) = \sum_{n=1}^{\infty} \mathbb{P}\{t_n < t\} = \sum_{n=1}^{\infty} F^{*n}(t) \quad (1)$$

for $t \geq 0$ where $F^{*n}$ denotes the $n$-fold recursive Stieltjes convolution of $F$.

Several rf-estimation methods have been developed for a known interarrival-time distribution. Unfortunately, explicit forms of the rf are obtained only in rare cases, for example, if the interarrival times have a uniform distribution, or for the wide class of matrix-exponential distributions (exponential and Erlang distributions belong to this class) (Ref.[2]). Therefore, several attempts have been made to evaluate the rf computationally (Refs.[4,6,9,17,25]).

If the mean $\mu$ and variance $\sigma^2$ of $F$ are finite, then the rf $H(t)$ may be approximated for large $t$ by an expression

$$H(t) = \frac{t}{\mu} + \frac{\sigma^2}{2\mu^2} - \frac{1}{2} + \circ(1)$$

widely used in the literature (Ref.[12]). In Ref.[4] an alternative asymptotic expression is stated for the case that the Laplace-Stieltjes transform (LST) of $F(t)$ is a rational function. Such estimates do not perform well for small times $t$ relative to $\mu$, which is especially important for the warranty control of devices (Ref.[10]).

In practice, it is a more realistic situation that the distribution is unknown or that just general information describing it is available. The restoration of the df or the probability density function (pdf), if the latter exists, may become complicated if the distributions of the random variables (rvs) are heavy-tailed. This means that those distributions have heavier tails than an exponential one (Refs.[7,8,13,18,19]). Weibull distributions with a shape parameter less than one and Pareto distributions provide examples of such pdfs. Heavy-tailed distributions often arise in practice, for example, in insurance and queueing or in the characterization of World Wide Web (WWW) traffic (Ref.[20]).

In this paper, we propose to estimate the rf without any information about the form of the underlying distribution and we use only an empirical sample $T^l = \{\tau_n, n = 1, 2, \ldots, l\}$ of the nonnegative i.i.d. interarrival times between events of the size $l$. The stated nonparametric estimate is related to a histogram-type estimate where the unknown probabilities $\mathbb{P}\{t_n < t\}$ in (1) are replaced by the corresponding empirical dfs and a limited number of terms $k$ is used in the summation. A similar nonparametric estimate

$$H_l(t, k) = \sum_{n=1}^{k} F_l^{(n)}(t) \tag{2}$$

was proposed by Frees[10,11] and further investigated in Ref.[24]. These authors have used

$$F_l^{(n)}(t) = \binom{l}{n}^{-1} \sum_{c} \theta(t - (\tau_{i_1} + \cdots + \tau_{i_n}))$$

as estimate of the arrival-time distribution. Here $\theta(t) = \left\{\begin{smallmatrix} 1, & t \geq 0 \\ 0, & t < 0 \end{smallmatrix}\right.$ and $\sum_c$ denotes the sum over all $\binom{l}{n}$ distinct index combinations $\{i_1, i_2, \ldots, i_n\}$ of length $n$. The $U$-statistic $F_l^{(n)}(t)$ is a minimum-variance unbiased estimator of $F^{*n}(t)$. In contrast to our estimate (4), which uses just one combination of adjacent interarrival times, the computation of the $H_l(t, k)$ is awkward.

The accuracy of such types of estimates depends on $k$. Frees has obtained the uniform consistency of $H_l(t, k)$ a.s. on compact intervals $[0, t], 0 \leq t < \infty$, under the assumptions that $k = l$ and $F(t)$ has a positive mean and finite variance and the asymptotic normality of $H_l(t, k)$ for each fixed point $t$ under some moment conditions (Ref.[11]). However, the data-dependent selection of $k$ (which is important for moderate samples) was not considered in Refs.[10,11,24]. In Ref.[14] the convergence of a non-computational empirical rf

$$H_{l_g}(t) = \sum_{n=1}^{\infty} \widehat{F}_l^{*n}(t) \tag{3}$$

on $\mathbb{R}$ as $l \to \infty$ has been proved. Here, $\widehat{F}_l^{*n}(t)$ is the $n$-fold convolution of the empirical df $F_l(t)$ based on the sample $T^l$.

In our consideration, we use an unbiased estimate of $F^{*n}(t)$, but its variance is not minimal. We compensate this inaccuracy by the data-dependent selection of $k$ and the use of larger samples. An a priori choice of $k$ as a function of the sample size $l$ is proposed to obtain a.s. the uniform convergence of the estimate to the rf as $l \to \infty$ for light- and heavy-tailed pdfs. The bootstrap method is applied for a data-dependent choice of $k$.

The paper is organized as follows. In Section 2 the histogram-type estimate of the rf is investigated. Theorems 2.1.1., 2.1.2, and 2.1.3, related to the a.s. uniform convergence of this estimate to the rf are stated. The selection of the parameter $k$ by the bootstrap method is described for different time intervals. Section 3 contains a simulation study about the accuracy of the proposed estimate for different itds and different values of $k$ selected by the bootstrap method. A comparison with Frees' estimate is given. Finally, the findings are summarized in the conclusion. In the Appendix, the proofs of the theorems of Section 2 are presented.

## 2. HISTOGRAM-TYPE ESTIMATE OF THE RENEWAL FUNCTION

We consider the estimate of the rf $H(t)$ which was first introduced in Ref.[16]. Let $[r]$ denote the integer part of a number $r$. Inserting the empirical mean for $\mathbb{E}(N_t)$, we replace the df $\mathbb{P}\{t_n < t\}$ by the empirical df $F_{l_n}(t) = \frac{1}{l_n}\sum_{i=1}^{l_n}\theta(t - t_n^i)$. It is its unbiased estimate and $t_n^i = \sum_{q=1+n(i-1)}^{n \cdot i} \tau_q$, $i = 1, \ldots, l_n, l_n = \left[\frac{l}{n}\right]$, $n = 1, \ldots, k$, are the observations of the rv $t_n$. Then one can estimate the renewal function $H(t)$ based on the samples of independent renewal-time observations $t_1 = \{t_1^1, \ldots, t_1^{l_1}\}, \ldots,$ $t_k = \{t_k^1, \ldots, t_k^{l_k}\}$ by:

$$\widetilde{H}(t, k, l) = \sum_{n=1}^{k} \frac{1}{l_n} \sum_{i=1}^{l_n} \theta(t - t_n^i). \tag{4}$$

Note that $\widetilde{H}(t, k, l) = k$ holds for $t \in [t_{\max}(k), \infty)$ where $t_{\max}(k) = \max_{1 \le n \le k} \max_{1 \le i \le l_n} t_n^i$ and $k$ is some fixed number.

The errors of the estimation arise both from the approximation of $H(t)$ in (4) by a finite sum and the approximation of $\mathbb{P}\{t_n < t\}$ by the empirical df $F_{l_n}(t)$:

$$\|H(t) - \widetilde{H}(t, k, l)\| = \left\| \sum_{n=k+1}^{\infty} \mathbb{P}\{t_n < t\} + \sum_{n=1}^{k}(\mathbb{P}\{t_n < t\} - F_{l_n}(t)) \right\|. \tag{5}$$

By this formula one can see that $\widetilde{H}(t, k, l)$ as well as the estimator (2) are biased since $k$ is limited.

A rough upper bound of the bias is given by

$$\text{bias}(t,k,l) = H(t) - \mathbb{E}\widetilde{H}(t,k,l) = \sum_{n=k+1}^{\infty} \mathbb{P}\{t_n < t\}$$
$$\leq \sum_{n=k+1}^{\infty} (F(t))^n = \frac{(F(t))^{k+1}}{1-F(t)}. \tag{6}$$

For small $t$ $F(t)$ is generally small and $F(t) < 1$, thus, this error is small.

To provide a good approximation of $\mathbb{P}\{t_n < t\}$ by the empirical df, according to the Glivenko-Cantelli theorem sufficiently large values $l_n$ should be used, i.e., $k < l$. Note that $l_k = 1$ for $\frac{l}{2} < k \leq l$, i.e., the sample $t_k$ contains only one point. Therefore, it is reasonable to take $k \leq \frac{l}{2}$. In the following, we provide an optimized estimate of $k$. On the other hand, to provide a good approximation of $H(t)$ by means of $\widetilde{H}(t,k,l)$ in general, the value of $k$ should be large enough. Therefore, the estimate $\widetilde{H}(t,k,l)$ is sensitive to the choice of $k$ and the length of the estimation interval $[0,t]$. Obviously, the estimate $\widetilde{H}(t,k,l)$ may only be accurate within the interval $[0, t_{\max}(k)]$ since the sample size $l$ is limited.

### 2.1. Convergence of the Histogram-Type Estimate

Now we investigate the convergence of the estimate (4) to the rf in the metric of the space $C$ of continuous functions. To estimate the risk (5), one is interested in the $t$-regions $[0,t] \subseteq [0, t_{\max}(k)]$. It is the main problem to estimate the systematic error $\sum_{n=k+1}^{\infty} \mathbb{P}\{t_n < t\}$. To estimate it, one needs some information about the df $F(t)$ of the rv $\tau$ and then one can use precise large deviation results for $\mathbb{P}\{t_n < t\}$. Such a principal information may be the existence of the moment generating function (Cramér's condition) (Ref.[21]). The rv $\tau$ satisfies Cramér's condition if there exists $\theta > 0$ such that $\mathbb{E}(e^{\theta\tau}) < \infty$. The Cramér's condition is equivalent to an exponential decay rate of $1 - F(t)$ and it is satisfied for light-tailed distributions. The Cramér's condition provides the existence of all moments of the rv $\tau$.

**Theorem 2.1.1.** *Let $\{\tau_1, \ldots, \tau_l\}$ be a sequence of i.i.d. rvs and $t \in [0, t_{max}(k)]$. We suppose that $\mathbb{E}|\tau_i|^m < \infty$ for some integer $m \geq 1$, $\mathbb{E}\tau_i = \mu$, $\text{var}(\tau_i) = \sigma^2$, and that the parameter $k$ obeys*

$$k = k(l) \sim l^{\rho} \quad (as \quad l \to \infty), \;\; 0 < \rho < 1/3. \tag{7}$$

*Then*

$$\mathbb{P}\Big\{\omega : \overline{\lim_{l\to\infty}} \sup_t \big|H(t) - \widetilde{H}(t,k,l)\big| = 0\Big\} = 1$$

*holds.*

The rate of this uniform convergence may be proved for the class $\widetilde{S}$ of itds such that

$$1 - F(t) \geq \exp(-\nu t)$$

for any $t \in [0, T]$ and some $\nu > 0$. We assume, without loss of generality, that $[0, T] = [0, 1]$. The class $\widetilde{S}$ includes, for example, the exponential distribution and the Weibull distribution with a shape parameter larger than one. Hence, it follows for the estimate of the right-hand side of (6):

$$\frac{(F(t))^{k+1}}{1 - F(t)} \leq \frac{(1 - \exp(-\nu t))^{k+1}}{\exp(-\nu t)}.$$

Then, for $F(t) \in \widetilde{S}$ the error of an approximation by (4) in the metric of $C$ is estimated by

$$\sup_t |H(t) - \widetilde{H}(t, k, l)| \leq \sup_t \left( \frac{(1 - \exp(-\nu t))^{k+1}}{\exp(-\nu t)} + \left| \sum_{n=1}^{k} (\mathbb{P}\{t_n < t\} - F_{l_n}(t)) \right| \right). \quad (8)$$

**Theorem 2.1.2.** *If $\{\tau_1, \ldots, \tau_l\}$ is an i.i.d. sample with the df $F \in \widetilde{S}$, $t \in [0, 1]$ and the parameter $k = c \cdot l^\rho$ $(c = c(\nu, \alpha, \rho) > 0)$, $0 < \rho < 1/3 - (2/3)\alpha$, $0 < \alpha < 1/2$, then the asymptotic rate of convergence of the estimate $\widetilde{H}(t, k, l)$ to $H(t)$ is given by the expression*

$$\mathbb{P}\left\{\omega : \overline{\lim_{l \to \infty}} \sup_t l^\alpha \left| H(t) - \widetilde{H}(t, k, l) \right| \leq c_1 \right\} = 1,$$

*where $c_1$ is a constant that is independent of $l$.*

Then the following confidence interval is derived for the rf.

**Corollary 2.1.1.** *If the assumptions of Theorem 2.1.2 hold, then the following inequalities hold with a probability of at least $1 - \chi$, $0 < \chi < 1$:*

$$\widetilde{H}(t, k, l) - D \leq H(t) \leq \widetilde{H}(t, k, l) + D,$$

*where*

$$D = l^{-\alpha} + k\sqrt{-\frac{k \ln(\chi/2)}{2l}}.$$

In practice, interarrival times are often described by distributions with heavy tails (Refs.[5,13]). Two classes of heavy-tailed distributions are well known: the distributions with regularly varying tails where $1 - F(t) = t^{-\alpha}L(t)$, $t > 0$, $\alpha > 0$, and $L$ is a slowly varying function, and the subexponential distributions, i.e., the distributions with the property: for any $\varepsilon > 0$ there exists $T = T(\varepsilon, F)$ such that for any $t > T$, $1 - F(t) > \exp(-\varepsilon t)$. It is the specific feature of heavy-tailed distributions that they do not satisfy Cramér's condition.

If $t$ is not too large, an approximation of $\mathbb{P}\{t_n > t\}$ by the tail of the standard normal distribution $\overline{\Phi}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{y^2}{2}}\, dy$ is used for heavy-tailed distributions, namely,

$$\mathbb{P}\{t_n > t\} \sim \overline{\Phi}\left(\frac{t}{\sqrt{n}}\right) \tag{9}$$

(this means that $\lim_{n\to\infty} \sup_{0<t<\frac{c_n}{h_n}} \left| \frac{\mathbb{P}\{t_n>t\}}{\overline{\Phi}\left(\frac{t}{\sqrt{n}}\right)} - 1 \right| = 0$) if $t \in \left(0, \frac{c_n}{h_n}\right)$ for any choice of the sequence $h_n \to \infty$ as $n \to \infty$ (Ref.[18]).

Several threshold sequences $c_n$ are proposed by different authors. For example, for a Weibull distribution with a shape parameter $0 < \alpha \le 0.5$, $c_n/h_n \sim n^{1/(2-\alpha)}$ and for $0.5 < \alpha < 1$, $c_n/h_n \sim n^{2/3}$, for distributions with regularly varying tails and $\alpha > 2$, $c_n/h_n$ may be $\sim n^{0.5} \ln^{0.5} n$ (Ref.[19]). Hence, the following theorem can be proved.

**Theorem 2.1.3.** *If* $\{\tau_1, \ldots, \tau_l\}$ *is a sequence of i.i.d. rvs with the heavy-tailed df* $F(t)$, $t \in \left(0, \min\left(t_{max}(k), \frac{c_k}{h_k}\right)\right]$, *and the parameter* $k$ *obeys* (7), *then*

$$\mathbb{P}\left\{\omega : \overline{\lim_{l\to\infty}} \sup_t \left|H(t) - \widetilde{H}(t,k,l)\right| = 0\right\} = 1$$

*holds.*

The theorems determine the values of $k$ as functions of the sample size $l$. These values $k$ are given only up to a rough asymptotic equivalence. For instance, $k$ can be multiplied by any positive constant and the theorems remain valid. In practice, one needs exact optimal values of $k$ which are adapted to the empirical data. Therefore, we subsequently consider a data-dependent selection of $k$.

### 2.2. Selection of *k* by a Bootstrap Method

Using empirical data, the automatic choice of $k$ can be done by minimizing the bootstrap estimate of the mean squared error of $H(t)$ for fixed $t$, i.e.,

$$MSE(t,k,l) = \mathbb{E}(\widetilde{H}(t,k,l) - H(t))^2 \to \min_k .$$

The bootstrap estimate is obtained by drawing $B$ re-samples with replacement from the original data set $T^l$. Some observations from $T^l$ may appear more than once while others do not appear at all.

Considering the solution of several problems of statistics like the choice of smoothing parameters of the kernel estimator for a pdf, a nonparametric regression or Hill's estimate of a tail index, it is recommended to use smaller re-samples of the size $l_1 < l$. It is the goal to avoid the situation when the bootstrap estimate of the bias is equal to zero regardless of the non-zero true bias of the estimator (Ref.[15]).

The bootstrap estimate of the rf that is constructed from $T^l$ by some of the re-samples $T^*_{l_1} = \{\tau^*_1, \dots, \tau^*_{l_1}\}$ of size $l_1$ is given in a way similar to (4) by

$$\widetilde{H}^*(t, k_1, l_1) = \sum_{n=1}^{k_1} \frac{1}{l_n^1} \sum_{i=1}^{l_n^1} \theta(t - t_n^{*i}), \quad l_n^1 = [l_1/n], \ t_n^{*i} = \sum_{q=1+n(i-1)}^{ni} \tau^*_q.$$

The values $l_1$ and $l$ may be related by

$$l_1 = l^{\beta}, \quad 0 < \beta < 1. \tag{10}$$

The values $k_1$ and $k$ are related by

$$k = k_1(l/l_1)^{\alpha}, \quad 0 < \alpha < 1. \tag{11}$$

The problem arises which $\alpha$ and $\beta$ should be taken. Considering the related problem to choose the smoothing parameter in the case of a nonparametric pdf estimator or nonparametric regression, Hall[15] has derived by means of an asymptotic theory that $\beta = 1/2$ leads to the most accurate results. Regarding the bootstrap estimation of the parameter of Hill's estimate $\alpha = 2/3$ has been recommended.

The bias and the variance of $\widetilde{H}^*(t, k_1, l_1)$ are given by

$$b^*(t, l_1, k_1) = \mathbb{E}\{\widetilde{H}^*(t, k_1, l_1) \mid T^l\} - \widetilde{H}(t, k, l) \tag{12}$$

and

$$\text{var}^*(t, l_1, k_1) = \mathbb{E}\{\widetilde{H}^*(t, k_1, l_1)^2 \mid T^l\} - \left(\mathbb{E}\left\{\widetilde{H}^*(t, k_1, l_1) \mid T^l\right\}\right)^2, \tag{13}$$

respectively. Here, $T^l$ is fixed and the expectation is calculated over all theoretically possible re-samples $T^*_{l_1}$ of $T^l$ with the size $l_1$. Then the bootstrap estimate of the $MSE(t, k, l)$ is determined by

$$\begin{aligned} MSE^*(t, k_1, l_1) &= \mathbb{E}\{(\widetilde{H}^*(t, k_1, l_1) - \widetilde{H}(t, k, l))^2 \mid T^l\} \\ &= (b^*(t, l_1, k_1))^2 + \text{var}^*(t, l_1, k_1). \end{aligned}$$

In the following, we will show that minimizing $MSE^*(t, k_1, l_1)$ by $k_1$ is as awkward as the calculation of the estimator (2). The problem arises from the calculation of the subsequently considered statistic $F_{l^n}(t)$. It coincides with the statistic $\widehat{F}_l^{*n}(t)$, see (3), and it is close to the $U$-statistic $F_l^{(n)}(t)$. We note the difference that $F_{l^n}(t)$ is calculated over all combinations of $n$ observations with possible repetitions. Namely, there are $l$, $l^2, \ldots, l^{k_1}$ re-samples from $T^l$ of sizes $1, 2, \ldots, k_1$, respectively. Then

$$\mathbb{E}\{\widetilde{H}^*(t, k_1, l_1) \mid T^l\} = \mathbb{E}\left\{ \sum_{n=1}^{k_1} \frac{1}{l_n^1} \sum_{i=1}^{l_n^1} \theta(t - t_n^{*i}) \mid T^l \right\}$$
$$= \sum_{n=1}^{k_1} \frac{1}{l^n} \sum_{i_1=1}^{l} \cdots \sum_{i_n=1}^{l} \theta(t - (\tau_{i_1} + \cdots + \tau_{i_n})) = \sum_{n=1}^{k_1} F_{l^n}(t) \tag{14}$$

holds. Hence, by (12) and (14) one can see that the bias of the bootstrap approach does not depend on $l_1$, i.e.,

$$b^*(t, l, k_1) = \sum_{n=1}^{k_1} F_{l^n}(t) - \widetilde{H}(t, k, l). \tag{15}$$

When re-samples of size $l$ are used, then the bias of the bootstrap

$$b^*(t, l, k) = \sum_{n=1}^{k} F_{l^n}(t) - \widetilde{H}(t, k, l)$$

may be close to 0 for sufficiently large $l$ (since $F_{l^n}$ and $F_{l_n}$ may not differ so much), but for $k = 1$ it is equal to 0 since $l^n = l_n$ regardless of the true bias of $\widetilde{H}(t, 1, l)$ (see (6)).

Independent of the values $k_1$ and $l_1$ the bootstrap variance is equal to 0. This property follows from (13), (14) and the expression

$$\mathbb{E}\{\widetilde{H}^*(t, k_1, l_1)^2 \mid T^l\}$$
$$= \sum_{n=1}^{k_1} \sum_{m=1}^{k_1} \frac{1}{l^{n+m}} \sum_{i_1=1}^{l} \cdots \sum_{i_n=1}^{l} \sum_{j_1=1}^{l} \cdots$$
$$\cdots \sum_{j_m=1}^{l} \theta(t - \max(\tau_{i_1} + \cdots + \tau_{i_n}, \tau_{j_1} + \cdots + \tau_{j_m})).$$

Hence, the bootstrap estimate of the $MSE(t,k,l)$ is given by

$$MSE^*(t,k_1,l) = b^*(t,l,k_1)^2. \tag{16}$$

However, it is a problem that the statistic $F_{l^n}(t)$ cannot be computed easily. The idea to use the empirical estimate of $b^*(t,l,k_1)$

$$\hat{b}^*(t,l_1,k_1) = \frac{1}{B}\sum_{b=1}^{B} H^b(t,k_1,l_1) - \widetilde{H}(t,k,l)$$

(where $B$ denotes the number of $l_1$-sized re-samples) instead of the actual bootstrap bias may give rough results. Here, we denote by $H^b(t,k_1,l_1)$ the estimate (4) constructed by some re-sample. By the same reasons we suggest to minimize in practice the empirical estimate of $MSE^*(t,k_1,l)$

$$\widehat{MSE}^*(t,k_1,l_1) = \hat{b}^*(t,l_1,k_1)^2 + \widehat{\text{var}}^*(t,l_1,k_1), \tag{17}$$

where

$$\widehat{\text{var}}^*(t,l_1,k_1) = \frac{1}{B-1}\sum_{b=1}^{B}\left(H^b(t,k_1,l_1) - \frac{1}{B}\sum_{b=1}^{B} H^b(t,k_1,l_1)\right)^2$$

is an empirical estimate of the bootstrap variance. All possible values of $k_1$ should be examined, where $k_1$ is an integer in the interval $[1,[l_1/2]]$.

The estimate $H_l(t,k)$ requires $A_1 = \sum_{n=1}^{k}\binom{l}{n}(n+1) = 2^l(1+l/2) - \sum_{n=k+1}^{l}\binom{l}{n} - \sum_{n=k+1}^{l} n\binom{l}{n} - 1$ operations, whereas $\widetilde{H}(t,k,l)$ requires $A_2 = \sum_{n=1}^{k}\left[\frac{l}{n}\right](n+1)$ operations. The selection of $k$ in $\widetilde{H}(t,k,l)$ by means of the empirical bootstrap method (i.e., the minimization of (17)) requires additionally to $A_2$ $A_3 = \sum_{k_1=1}^{[l_1/2]}\left(B\sum_{n_1=1}^{k_1}\left[\frac{l_1}{n_1}\right](n_1+1)+6\right)$ operations. Note, that

$$S_k = \sum_{n=1}^{k}\frac{1}{n} = c + \psi(k) + \frac{1}{k},$$

where $c \approx 0.5772$ is Euler's constant, $\psi(z) = \Gamma'(z)/\Gamma(z)$, $\Gamma(\cdot)$ is the Gamma-function (Ref.[23]). We suppose for simplicity that $[l/n] = l/n$, $[l_1/n_1] = l_1/n_1$, $[l_1/2] = l_1/2$ and $k = [l/2] = l/2$. By Ref.[23] it holds

$$\sum_{n=0}^{l/2}\binom{l}{n} = 2^{l-1} + \frac{1+(-1)^l}{4}\binom{l}{l/2},$$

hence, we have

$$A_1 = 2^{l-1}(1+l) + \frac{1+(-1)^l}{4}\binom{l}{l/2} - 1 - \sum_{n=l/2+1}^{l} n\binom{l}{n},$$

$$A_2 = l(l/2 + S_{l/2}), \quad A_3 = Bl_1\left(l_1(2+l_1)/8 + \sum_{k_1=1}^{l_1/2} S_{k_1}\right) + 3l_1.$$

Hence, for example, when $k = 10,\ l = 20$ we have $A_1 = 5.86 \cdot 10^6$, $A_2 = 258.579$, $A_1/A_2 = 2.266 \cdot 10^4$. Let $B = 50,\ l_1 = 6 \approx l^{0.6}$ then $A_3 = 3.118 \cdot 10^3$ and $A_1/(A_2 + A_3) = 1.735 \cdot 10^3$.

## 3. A SIMULATION STUDY

To evaluate the performance of the bootstrap approach, we have to investigate the influence of the values $\alpha$ and $\beta$ in (10) and (11) by a Monte-Carlo simulation. The following values $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ are used for a $\beta \in \{0.3, 0.5, 0.7\}$. Let $[0, T]$ be the interval of the estimation. The provided times are given by $T \in \{0.5, 2.5, 4.5, 6.5, 8.5, 10\}$. We generate samples with the known pdfs

$$f_1(t) = \begin{cases} \lambda \exp(-\lambda t), & t \geq 0, \\ 0, & t < 0, \end{cases}$$

of an exponential distribution with the parameter $\lambda = 1$,

$$f_2(t) = \begin{cases} t^{s-1}\lambda^{-s}\exp(-t/\lambda)/\Gamma(s), & t > 0, \\ 0, & t \leq 0, \end{cases}$$

of a Gamma distribution $Ga(\lambda, s)$ with the parameters $s = 2$ and $\lambda = 1$,

$$f_3(t) = \begin{cases} st^{s-1}\exp(-t^s), & t > 0, \\ 0, & t \leq 0, \end{cases}$$

of a Weibull distribution with $s = 0.5$. The latter distribution is heavy-tailed and subexponential. It is one of the most interesting distributions in reliability engineering where the pdf is singular.

For $f_1(t)$ and $f_2(t)$ the rfs are determined by $H_1(t) = \lambda t$ and $H_2(t) = 0.5 \cdot (t - 0.5 + 0.5\exp(-2t))$, respectively. For $f_3(t)$ an explicit form of the rf is unknown. Therefore, we use the results of a numerical approximation by Xie's RS-method as $H_3(t)$ since this method provides rather accurate results for a known pdf and a correctly selected step size $h = \frac{t}{N}$. Here $N$ is the number of points inside the interval $[0, t]$ (Ref.[25]). Strictly speaking,

$H(i)$ is recursively calculated by

$$H(i) \approx \frac{F1(i) + \sum_{j=1}^{i-1} F(i-j)(H(j) - H(j-1)) - F0H(i-1)}{1 - F0},$$

where $0 = z_0 < z_1 < \cdots < z_N = t$ and $H(i) = H\left(\frac{i \cdot t}{N}\right)$, $F(i) = F\left(\frac{(i+0.5)t}{N}\right)$, $F0 = F\left(0.5\frac{t}{N}\right)$, $F1(i) = F\left(\frac{it}{N}\right)$ are used.

In Tables 1 and 2 one can see the bias and mean squared error of the estimate (4) calculated by 200 repeated samples for the given $f_1(t)$ and $f_2(t)$. The parameter $k$ in (4) is determined by the bootstrap method. The sample size $l = 50$ and the number $B = 50$ of bootstrap re-samples were taken. In order to understand the results of Tables 1–3 and Figures 1 and 2 may be helpful. The values $\overline{MSE}$ and $\overline{BIAS}$ are the averages of absolute values of *MSE* and *BIAS* over all different $T$ for each fixed tuple $(\alpha, \beta)$. In Figures 1 and 2, the left figures correspond to Table 1 and the right figures to Table 2. In Table 3, one can see the corresponding minimal values of $\overline{MSE}$ and $\overline{BIAS}$ for a Gamma and an exponential distribution.

From Tables 1–3 and Figures 1 and 2 one can see that

- $\alpha = 0.7$, $\beta = 0.3$ provide the minimal $\overline{MSE}$;
- the better trade-off between the averages $\overline{MSE}$ and $\overline{BIAS}$ is provided by $\alpha = 0.7$, $\beta \in \{0.3, 0.5\}$;
- the mean squared error increases if the time interval $[0, T]$ of the estimation is extended.

One can observe that the bias and, consequently, the *MSE* arising from an exponential distribution are worse than those of a Gamma distribution, especially for large $T$. Obviously, the difference stems from the fact that an exponential tail is heavier than a Gamma tail. By our experience we know that the bias and the *MSE* of the heavy-tailed Weibull distribution $f_3(t)$ are worse than those of the exponential distribution $f_1(t)$. These results do not contradict the theory. Namely, the central limit theorem is fulfilled for heavy-tailed distributions just for relatively small time intervals (see (9)). The width of these intervals is established asymptotically and depends on the shape parameters which determine how heavy the tail of the distribution is. This theoretical fact is partly supported by our experiment with moderate samples since we have considered the same time intervals $[0, T]$ both for $f_1(t)$ and $f_2(t)$.

In Figures 3–5, Xie's estimate and the histogram-type estimate are shown for the pdf $f_3(t)$ and the pdf

$$f_4(t) = \begin{cases} \frac{c}{\rho}\left(\frac{\rho}{\rho + t}\right)^{c+1}, & t > 0, \\ 0, & t \leq 0, \end{cases}$$

**TABLE 1** Quality of estimate (4): Gamma ($s = 2$, $\lambda = 1$, $\mathbb{E}\tau = 2$), sample size $l = 50$

| | | $\beta = 0.3$ | | $\beta = 0.5$ | | $\beta = 0.7$ | |
|---|---|---|---|---|---|---|---|
| T | α | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ |
| 0.5 | 0.1 | −26.99 | 17.5 | 82.3 | 21.36 | 45.3 | 17.18 |
| | 0.2 | 26.3 | 20.06 | −16.99 | 19.09 | −13.7 | 16.74 |
| | 0.3 | 23.3 | 15.67 | 18.3 | 19.05 | 29.3 | 19.21 |
| | 0.4 | 59.3 | 23.45 | −26.7 | 18.03 | 14.3 | 17.74 |
| | 0.5 | −31.57 | 19.62 | 39.3 | 17.09 | −23.7 | 18.17 |
| | 0.6 | 14.3 | 18.14 | 50.43 | 23.12 | 14.3 | 19.82 |
| | 0.7 | −68.7 | 19.59 | 32.3 | 20.51 | −36.7 | 15.91 |
| 2.5 | 0.1 | −280 | 210 | −270 | 200 | −220 | 260 |
| | 0.2 | −530 | 170 | −270 | 230 | −230 | 200 |
| | 0.3 | −110 | 290 | −390 | 230 | −260 | 230 |
| | 0.4 | −140 | 250 | −120 | 250 | −98.68 | 250 |
| | 0.5 | 76.86 | 250 | 36.35 | 270 | −40.05 | 280 |
| | 0.6 | 64.45 | 230 | 83.03 | 270 | −320 | 250 |
| | 0.7 | −68.84 | 260 | −76.26 | 280 | −68.47 | 270 |
| 4.5 | 0.1 | −1840 | 910 | −1660 | 750 | −570 | 690 |
| | 0.2 | −1400 | 1130 | −1310 | 650 | −760 | 660 |
| | 0.3 | −240 | 800 | −1010 | 850 | −670 | 620 |
| | 0.4 | 260 | 670 | −200 | 610 | 24.27 | 880 |
| | 0.5 | 120 | 690 | 99.1 | 710 | −190 | 760 |
| | 0.6 | −26.1 | 600 | 230 | 890 | −130 | 780 |
| | 0.7 | 110 | 630 | −1040 | 670 | −350 | 720 |
| 6.5 | 0.1 | −4700 | 4360 | −3090 | 2560 | −1300 | 1570 |
| | 0.2 | −3970 | 4770 | −2680 | 2800 | −1630 | 1570 |
| | 0.3 | −330 | 1540 | 2330 | 3310 | −1500 | 1770 |
| | 0.4 | −280 | 1600 | −320 | 1410 | −880 | 1820 |
| | 0.5 | −870 | 1260 | −390 | 1330 | −540 | 1940 |
| | 0.6 | 410 | 1340 | −540 | 1500 | −57.84 | 1800 |
| | 0.7 | −7.029 | 1270 | 260 | 1510 | 360 | 1410 |
| 8.5 | 0.1 | −7070 | 11100 | −6160 | 7920 | −2870 | 4130 |
| | 0.2 | −8920 | 18220 | −6060 | 10400 | −2620 | 4320 |
| | 0.3 | −1720 | 3530 | −5280 | 11640 | −2470 | 4350 |
| | 0.4 | −1140 | 3570 | −710 | 2480 | −2540 | 5870 |
| | 0.5 | 290 | 2220 | −170 | 2730 | −840 | 1900 |
| | 0.6 | −130 | 2150 | −170 | 2040 | −1940 | 5630 |
| | 0.7 | 150 | 2430 | 320 | 2360 | 280 | 2210 |
| 10 | 0.1 | −11340 | 18530 | −8180 | 14530 | −5550 | 7060 |
| | 0.2 | −17940 | 46720 | −7290 | 16290 | −4260 | 8400 |
| | 0.3 | −4560 | 8850 | −10660 | 29870 | −4670 | 1120 |
| | 0.4 | −4370 | 8770 | −1120 | 3290 | −4750 | 13830 |
| | 0.5 | −340 | 3360 | −1620 | 4510 | −6370 | 18730 |
| | 0.6 | 87.66 | 3240 | 300 | 2750 | −3920 | 11810 |
| | 0.7 | −240 | 2680 | −570 | 3200 | 450 | 3180 |

**TABLE 2** Quality of estimate (4): Exp ($\lambda = 1$, $\mathbb{E}\tau = 1$), sample size $l = 50$

| | | $\beta = 0.3$ | | $\beta = 0.5$ | | $\beta = 0.7$ | |
|---|---|---|---|---|---|---|---|
| T | α | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ |
| 0.5 | 0.1 | −170 | 120 | −280 | 130 | −220 | 130 |
| | 0.2 | −250 | 140 | −120 | 130 | −100 | 140 |
| | 0.3 | 63.17 | 180 | −110 | 160 | −220 | 140 |
| | 0.4 | −170 | 130 | −180 | 180 | −87 | 170 |
| | 0.5 | 38.92 | 140 | 160 | 160 | −41.12 | 150 |
| | 0.6 | −94.96 | 200 | 96.71 | 160 | −21.37 | 130 |
| | 0.7 | −14.29 | 160 | 39.46 | 160 | 83.21 | 170 |
| 2.5 | 0.1 | −5580 | 4620 | −4160 | 2920 | −2760 | 2190 |
| | 0.2 | −4290 | 4430 | −4750 | 3780 | −3100 | 2250 |
| | 0.3 | −810 | 2240 | −4600 | 4100 | −3300 | 2809 |
| | 0.4 | −1760 | 1990 | −520 | 1705 | −2060 | 2540 |
| | 0.5 | −300 | 1500 | −360 | 1800 | −1620 | 2520 |
| | 0.6 | −240 | 1780 | −91.67 | 1780 | −1480 | 2340 |
| | 0.7 | 200 | 1900 | −310 | 1790 | 10.83 | 1980 |
| 4.5 | 0.1 | −13190 | 24050 | −11850 | 21340 | −8880 | 12660 |
| | 0.2 | −14970 | 35240 | −12300 | 27260 | −8190 | 14160 |
| | 0.3 | −5920 | 11690 | −11990 | 34250 | −11140 | 23220 |
| | 0.4 | −6370 | 12122 | 8570 | 7344 | 15500 | 24025 |
| | 0.5 | −1530 | 4914 | 8980 | 8064 | −8940 | 23720 |
| | 0.6 | −340 | 5970 | −690 | 5530 | −10160 | 28880 |
| | 0.7 | −340 | 6050 | −580 | 4760 | −840 | 4750 |
| 6.5 | 0.1 | −29770 | 93881 | −17910 | 53084 | −16540 | 45496 |
| | 0.2 | −43990 | 196426 | −25250 | 96100 | −17600 | 56929 |
| | 0.3 | −29560 | 98470 | −25900 | 121243 | −20780 | 79919 |
| | 0.4 | −30890 | 104522 | −10460 | 38887 | −25240 | 111020 |
| | 0.5 | −16080 | 37133 | −7080 | 31541 | −24060 | 112225 |
| | 0.6 | −5660 | 11534 | −3200 | 15951 | −24000 | 110622 |
| | 0.7 | −970 | 7691 | −1290 | 12454 | −3420 | 20050 |
| 8.5 | 0.1 | −47860 | 232324 | −28200 | 116417 | −24310 | 98658 |
| | 0.2 | −65030 | 422890 | −44620 | 242950 | −29930 | 148687 |
| | 0.3 | −54830 | 301401 | −51080 | 325356 | −35690 | 206116 |
| | 0.4 | −54720 | 300523 | −27000 | 143110 | −40360 | 266565 |
| | 0.5 | −43610 | 194481 | −24240 | 127377 | −4220 | 279417 |
| | 0.6 | −29300 | 97281 | −12490 | 54102 | −39760 | 269464 |
| | 0.7 | −15120 | 34040 | −8290 | 39720 | −16810 | 89880 |
| 10 | 0.1 | −60990 | 372954 | −46240 | 246810 | −32180 | 152412 |
| | 0.2 | −8000 | 160000 | −65010 | 447293 | −43990 | 287296 |
| | 0.3 | −70040 | 490560 | −75810 | 598302 | −48620 | 345162 |
| | 0.4 | −70030 | 490420 | −55120 | 370272 | −51640 | 424061 |
| | 0.5 | −60020 | 360600 | −52790 | 348808 | −56520 | 387048 |
| | 0.6 | −48530 | 239806 | −30320 | 163944 | −55660 | 451449 |
| | 0.7 | −35700 | 137418 | −28160 | 150466 | −17800 | 141676 |

**TABLE 3** The minimal values min $\overline{MSE}$ and min $\overline{BIAS}$ of averages $\overline{MSE}$ and $\overline{BIAS}$ calculated by Tables 1, 2 and corresponding $\alpha$ and $\beta$

| | | |
|---|---|---|
| | Gamma distribution | |
| min $\overline{MSE}$ | 0.121 | $(\alpha, \beta) = (0.7, 0.3)$ |
| min $\overline{BIAS}$ | 0.011 | $(\alpha, \beta) = (0.7, 0.3)$ |
| | Exponential distribution | |
| min $\overline{MSE}$ | 3.121 | $(\alpha, \beta) = (0.7, 0.3)$ |
| min $\overline{BIAS}$ | 0.6445 | $(\alpha, \beta) = (0.7, 0.5)$ |

of a Pareto distribution with the parameters $c = 0.5$, $\rho = 0.5$ at the time interval $[0, 5]$, as well as for the exponential pdf with $\lambda = 1$. The sample size $l = 100$ is used. The parameter $k$ has been selected by the bootstrap method with parameters $(\alpha, \beta) \in \{(0.7, 0.3);\ (0.7, 0.5);\ (0.7, 0.7);\ (0.1, 0.5);\ (0.4, 0.5)\}$. Regarding the bootstrap the number $B = 50$ of re-samples was taken. In Figures 4 and 5, the lines corresponding to $(\alpha, \beta) \in \{(0.7, 0.3);\ (0.7, 0.5);\ (0.7, 0.7)\}$ coincide with each other. In Figure 3, the line corresponding to $(\alpha, \beta) = (0.7, 0.7)$ differs from those lines with $(\alpha, \beta) \in \{(0.7, 0.3);\ (0.7, 0.5)\}$ approximately at the interval $[2.3, 3]$. One can see that the curves corresponding to $k$ selected by bootstrap with $(\alpha, \beta) \in \{(0.7, 0.3);\ (0.7, 0.5)\}$ for $f_3(t)$ and with $(\alpha, \beta) \in \{(0.7, 0.3);\ (0.7, 0.5);\ (0.7, 0.7)\}$ for the Pareto and exponential pdf's are closer to the true rf than all other curves. The line with $(\alpha, \beta) = (0.4, 0.5)$ is better than that with $(\alpha, \beta) = (0.1, 0.5)$, especially for the Pareto pdf. The figures support our previous conclusion regarding the prevalence of $\alpha = 0.7$ and $\beta \in \{0.3, 0.5\}$.

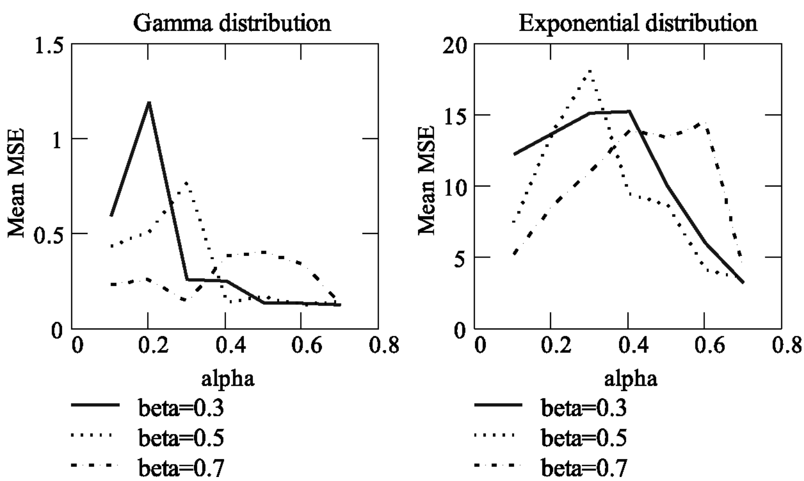


**FIGURE 1** The averages of the *MSE*s from Tables 1, 2 over different $T$ for fixed $\alpha = 0.1(0.1)0.7$, $\beta \in \{0.3, 0.5, 0.7\}$ and for a Gamma distribution (left) and an exponential distribution (right).

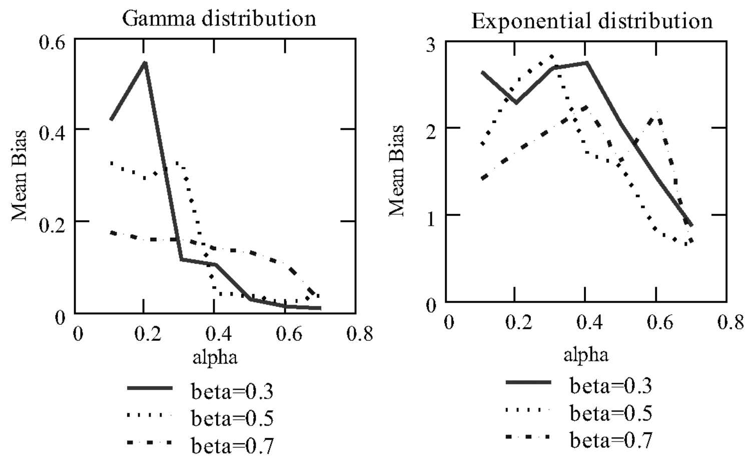


**FIGURE 2** The averages of absolute values of *BIAS*s in Tables 1, 2 over different $T$ for fixed $\alpha = 0.1(0.1)0.7$, $\beta \in \{0.3, 0.5, 0.7\}$ and for a Gamma distribution (left) and an exponential distribution (right).

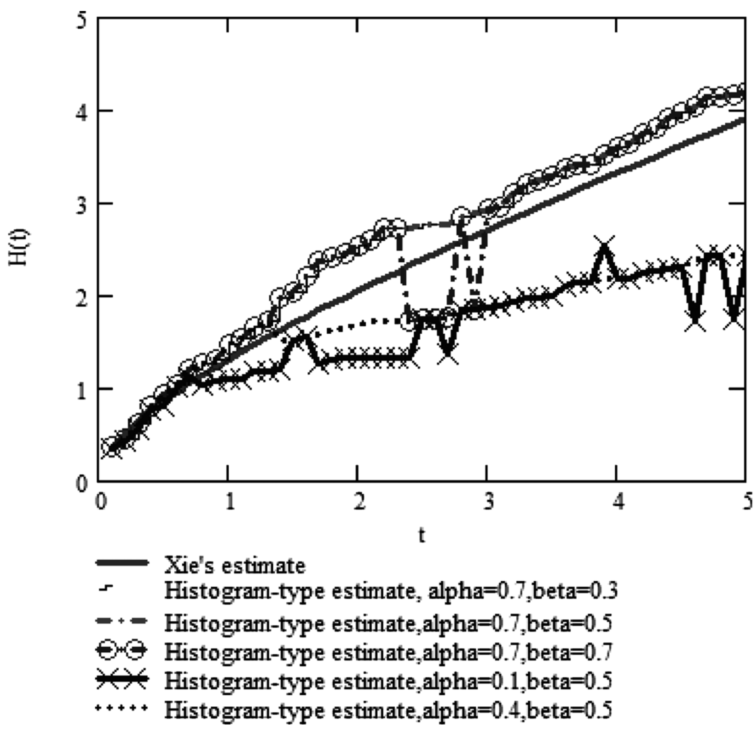


**FIGURE 3** Estimation of the renewal function of a Weibull distribution: $\widetilde{H}(t, k, l)$ against $t$.

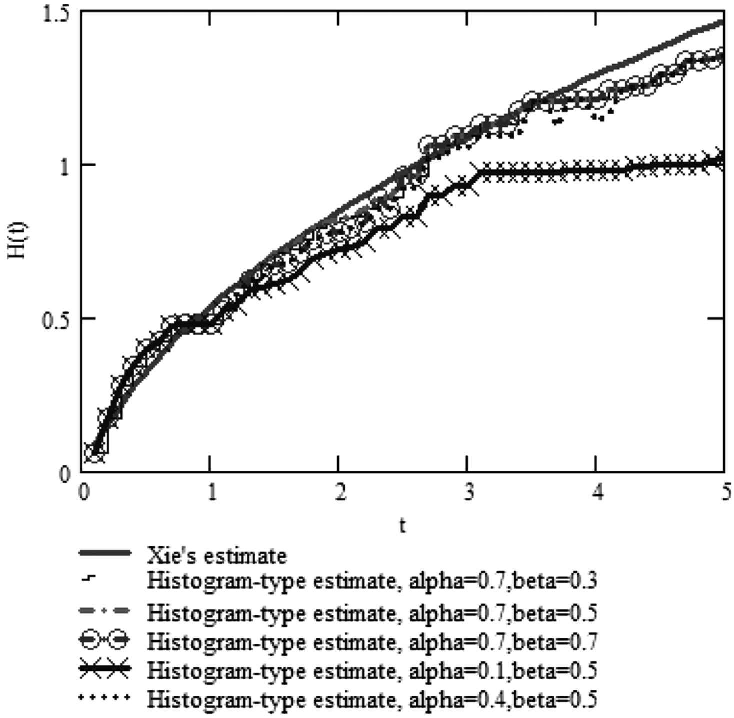


**FIGURE 4** Estimation of the renewal function of a Pareto distribution: $\widetilde{H}(t, k, l)$ against $t$.

The figures also illustrate the following phenomenon. Referring to formula (4) one can see that for some fixed $t$ the value of $\widetilde{H}(t, k, l)$ may not change any more as $k$ increases. For example, we have for $f_3(t)$ (Figure 3) and $t = 3$ that $k \in \{3; 3; 26; 29; 36\}$ for $(\alpha, \beta) \in \{(0.1, 0.3); (0.4, 0.5); (0.7, 0.7); (0.7, 0.3); (0.7, 0.5)\}$, respectively. It reflects the situation when the corresponding $\{t_n^i\}$ are larger than $t$ and the corresponding terms in sum (4) are equal to 0.

The second part of the simulation study relates to the comparison with Frees' tables presented in Ref.[10]. For this purpose samples of the log normal distribution with density $f(x) = (x\sigma\sqrt{(}2\pi))^{-1} \exp(-(\log x - \mu)^2/(2\sigma^2))$, where $\mu = 0$ and $\sigma^2 = 1$ and of the Weibull distribution with $s = 3$ (not heavy-tailed Weibull) have been generated. The Gamma distribution Ga(1, 0.55) presented in Ref.[10] was not considered. Since the generator of Gamma random variates (Ref.[11]) used in Ref.[10] is not reliable for small samples, it influences adversely on the accuracy of the results of a simulation study.

The provided times are given by $T \in \{0.25, 0.5, 0.75, 1.0, 1.25\}$ and $l \in \{10, 15, 20, 25, 30, 100\}$ are the samples sizes. As in Ref.[10] two characteristics, the bias and the mean squared error of the estimates,

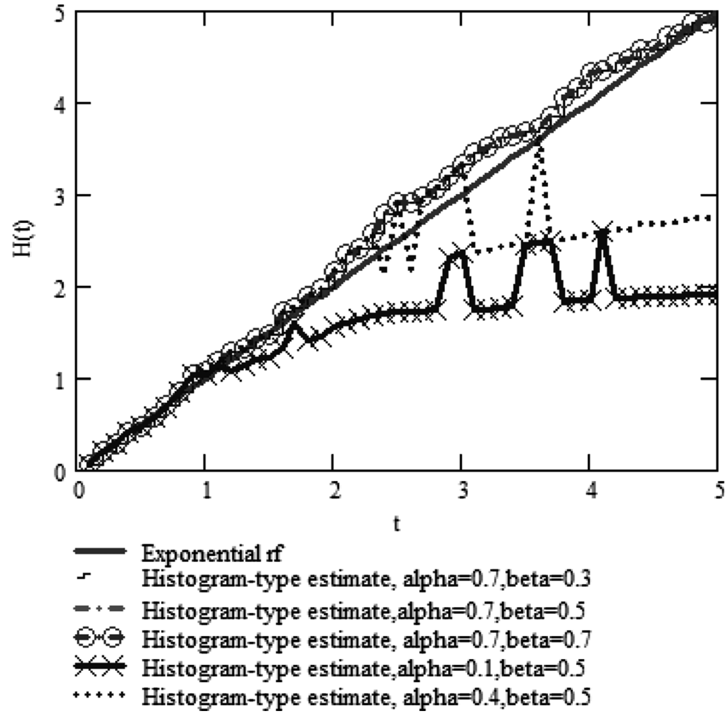


**FIGURE 5** Estimation of the renewal function of an exponential distribution: $\widetilde{H}(t, k, l)$ against $t$.

were calculated over 500 Monte Carlo repetitions and $H(t)$ is the true rf. For the fixed number of points $T$ and for the mentioned distributions the latter was taken from those tables presented in Ref.[3]. The results of the calculation are presented in Tables 5 to 8. Frees' results are included here, where $H_{3n}(t)$ denotes the estimate (2). Since (2) requires much computational effort, only $k \in \{5, 10\}$ and $l \leq 30$ were considered. Considering (4), the parameter $k$ is calculated by the bootstrap method, i.e., by minimizing (17) by $k_1$, where $l_1$ and $k_1$ are related to $l$ and $k$ by the formulas (10) and (11) with the parameters $\alpha = 0.7$ and $\beta = 0.3$. The number $B = 50$ of bootstrap re-samples was taken. The Tables 5 to 8 show

**TABLE 4** Comparison of $H_{3n}(t)$ for the sample size 30 and $\widetilde{H}(t)$ for the sample size 100 regarding averages of the $MSE$ and the $|BIAS|$ for different distributions

| Distribution | $\overline{MSE}_{H_{3n}}/\overline{MSE}_{\widetilde{H}}$ | $\overline{\lvert BIAS\rvert}_{H_{3n}}/\overline{\lvert BIAS\rvert}_{\widetilde{H}}$ |
|---|---|---|
| Log normal | 2.62 | 2.7 |
| Weibull | 2.25 | 1.69 |

**TABLE 5** Part I: Log normal ($\mu = 0,\ \sigma = 1,\ \mathbb{E}\tau = \exp(1/2)$)

| | | $H_{3n}(t)$ $(k = 5)$ | | $H_{3n}(t)$ $(k = 10)$ | | $\widetilde{H}(t, k, l)$ $(k_{boot})$ | |
|---|---|---|---|---|---|---|---|
| Size | T | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ |
| 10 | 0.25 | −1 | 77 | −1 | 77 | 44.24 | 77.25 |
| | 0.5 | 149 | 255 | 149 | 255 | 55.69 | 270 |
| | 0.75 | 128 | 422 | 128 | 422 | 43.35 | 610 |
| | 1 | 163 | 659 | −163 | 659 | 63.68 | 800 |
| | 1.25 | 117 | 918 | 117 | 918 | −51.65 | 1220 |
| 15 | 0.25 | 1 | 54 | 1 | 54 | 10.21 | 53.47 |
| | 0.5 | 123 | 179 | 123 | 179 | 59.68 | 190 |
| | 0.75 | 99 | 296 | 99 | 296 | −33.54 | 360 |
| | 1 | 133 | 443 | 123 | 443 | −100 | 620 |
| | 1.25 | 93 | 619 | 93 | 619 | 35.89 | 890 |

that for all estimates

(a) the mean squared error increases as $T$ increases;
(b) for any fixed $T$ the mean squared error decreases as $l$ becomes larger;
(c) the bias does not exhibit a stable behavior.

**TABLE 6** Part II: Log normal ($\mu = 0,\ \sigma = 1,\ \mathbb{E}\tau = \exp(1/2)$)

| | | $H_{3n}(t)$ $(k = 5)$ | | $\widetilde{H}(t, k, l)$ $(k_{boot})$ | |
|---|---|---|---|---|---|
| Size | T | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ | *BIAS* $\cdot 10^4$ | *MSE* $\cdot 10^4$ |
| 20 | 0.25 | −10 | 38 | −52.57 | 36.04 |
| | 0.5 | 94 | 124 | −22.15 | 130 |
| | 0.75 | 95 | 218 | 59.99 | 280 |
| | 1 | 141 | 329 | −61.07 | 390 |
| | 1.25 | 126 | 452 | 55.8 | 600 |
| 25 | 0.25 | 16 | 29 | 16.62 | 29.71 |
| | 0.5 | 111 | 101 | −37.62 | 100 |
| | 0.75 | 123 | 188 | 30.21 | 200 |
| | 1 | 179 | 277 | 110 | 340 |
| | 1.25 | 185 | 384 | −120 | 520 |
| 30 | 0.25 | 9 | 23 | 15.63 | 26.05 |
| | 0.5 | 79 | 80 | −31.47 | 77.29 |
| | 0.75 | 67 | 151 | −29.83 | 150 |
| | 1 | 96 | 218 | −66.82 | 320 |
| | 1.25 | 112 | 305 | −100 | 430 |
| 100 | 0.25 | n.a. | n.a. | −49.98 | 7.425 |
| | 0.5 | n.a. | n.a. | −24.53 | 27.44 |
| | 0.75 | n.a. | n.a. | −16.75 | 52.76 |
| | 1 | n.a. | n.a. | −21.61 | 88.89 |
| | 1.25 | n.a. | n.a. | −20.67 | 120 |

**TABLE 7** Part I: Weibull ($s = 3,\ \mathbb{E}\tau = 0.89$)

| Size | T | $H_{3n}(t)$ $(k=5)$ BIAS $\cdot 10^4$ | $H_{3n}(t)$ $(k=5)$ MSE $\cdot 10^4$ | $H_{3n}(t)$ $(k=10)$ BIAS $\cdot 10^4$ | $H_{3n}(t)$ $(k=10)$ MSE $\cdot 10^4$ | $\widetilde{H}(t,k,l)$ $(k_{boot},(\alpha,\beta)=(0.5,0.7))$ BIAS $\cdot 10^4$ | $\widetilde{H}(t,k,l)$ $(k_{boot},(\alpha,\beta)=(0.5,0.7))$ MSE $\cdot 10^4$ |
|---|---|---|---|---|---|---|---|
| 10 | 0.25 | −12 | 15 | −12 | 15 | −8.295 | 13.79 |
| | 0.5 | 9 | 98 | 9 | 98 | −19.33 | 100 |
| | 0.75 | 16 | 238 | 16 | 238 | 7.938 | 260 |
| | 1 | −8 | 318 | −8 | 318 | 16.52 | 370 |
| | 1.25 | −35 | 307 | −35 | 307 | 34.9 | 440 |
| 15 | 0.25 | −17 | 9 | −17 | 9 | −6.965 | 9.494 |
| | 0.5 | −13 | 71 | −13 | 71 | 28.48 | 79.1 |
| | 0.75 | 36 | 177 | 36 | 177 | −40.44 | 190 |
| | 1 | 9 | 211 | 9 | 211 | 9.58 | 260 |
| | 1.25 | −20 | 219 | −20 | 219 | 71.88 | 320 |

Comparing $\widetilde{H}(t,k,l)$ with $H_{3n}(t)$ one may conclude that

(a) biases and mean squared errors of $\widetilde{H}(t,k,l)$ and $H_{3n}(t)$ are comparable for the same sample sizes;
(b) increasing the sample size provides better accuracy regarding $\widetilde{H}(t,k,l)$ as shown in Table 4, where $\overline{MSE}_{H_{3n}}$, $\overline{MSE}_{\widetilde{H}}$ and $\overline{|BIAS|}_{H_{3n}}$, $\overline{|BIAS|}_{\widetilde{H}}$ are

**TABLE 8** Part II: Weibull ($s = 3,\ \mathbb{E}\tau = 0.89$)

| Size | T | $H_{3n}(t)$ $(k=5)$ BIAS $\cdot 10^4$ | $H_{3n}(t)$ $(k=5)$ MSE $\cdot 10^4$ | $\widetilde{H}(t,k,l)$ $(k_{boot})$ BIAS $\cdot 10^4$ | $\widetilde{H}(t,k,l)$ $(k_{boot})$ MSE $\cdot 10^4$ |
|---|---|---|---|---|---|
| 20 | 0.25 | −12 | 7 | −2.307 | 7.818 |
| | 0.5 | 1 | 53 | 12.61 | 59.16 |
| | 0.75 | 43 | 130 | 47.86 | 130 |
| | 1 | 34 | 174 | −14.42 | 200 |
| | 1.25 | 35 | 171 | 25.26 | 230 |
| 25 | 0.25 | −13 | 6 | 16.46 | 6.059 |
| | 0.5 | −12 | 38 | 13.47 | 48.87 |
| | 0.75 | 1 | 104 | −92.19 | 120 |
| | 1 | −12 | 130 | 12.23 | 160 |
| | 1.25 | 13 | 125 | −14.9 | 200 |
| 30 | 0.25 | −14 | 5 | −2.973 | 5.554 |
| | 0.5 | −39 | 31 | −22.65 | 38.53 |
| | 0.75 | −35 | 90 | −25.33 | 92.53 |
| | 1 | −46 | 114 | −58.66 | 130 |
| | 1.25 | −47 | 111 | 16.28 | 160 |
| 100 | 0.25 | n.a. | n.a. | 8.271 | 1.582 |
| | 0.5 | n.a. | n.a. | 6.621 | 11.29 |
| | 0.75 | n.a. | n.a. | −7.625 | 29.3 |
| | 1 | n.a. | n.a. | 22.16 | 46.03 |
| | 1.25 | n.a. | n.a. | −61.99 | 67.3 |

averages of the $MSE$ and the $|BIAS|$ over different $T$. The averaging was provided using the results of Tables 6 and 8 when the sample size is equal to 30 in the case of $H_{3n}$ and equal to 100 regarding $\widetilde{H}$.

## 4. CONCLUSIONS AND DISCUSSION

In this paper we have developed and investigated a nonparametric histogram-type estimate of the renewal function (rf) which does not require any knowledge about the form of the interarrival-time distribution.

Due to the limited number of empirical data the histogram-type estimate (as well as Frees' estimate (2)) can be applied for closed time intervals $[0, t]$ with a relative small $t$. Compared to Frees' estimate $F_l^{(n)}(t)$ of the arrival-time distribution $F^{*n}(t)$ in (2), the estimate (4) proposed here uses a simpler and rougher estimate of $F^{*n}(t)$. The rf $H(t)$ is approximated by a finite sum of estimates of the arrival-time distributions with $k$ terms. The parameter $k$ is selected to compensate the error of the risk function. The estimate (4) may be computed for sufficiently large $l$ and $k$ which is not realistic for (2).

The Theorems 2.1.1 and 2.1.3 state both for heavy- and light-tailed interarrival-time distributions (itds) those values of the parameter $k$ as functions of the sample size which provide a.s. the uniform convergence of the histogram-type estimate (4) to the true rf for sufficiently small $t$. It is proved that a smaller value of $k$ ($k < l$) than in Ref.[11] is sufficient to get a reliable estimate of the rf. In Theorem 2.1.2 the rate of the uniform convergence and a confidence interval of the rf for the specific class of itds with an exponential decay rate of the tails are presented. But these theorems determine $k$ only up to a rough asymptotic equivalence. Such a value $k$ does not depend on the empirical data. This feature may influence the accuracy of the estimation.

To estimate $k$ by samples of a moderate size, the bootstrap method is used. Following Hall's idea Ref.[15], a smaller re-sample size $l_1$ (and $k_1$, respectively) is used to avoid the situation where the bootstrap estimate of the bias is equal or close to zero regardless of the true bias of the estimate. Then the bootstrap estimate of the mean squared error $MSE(t, k, l) = \mathbb{E}\{(\widetilde{H}(t, l, k) - H(t))^2\}$, i.e., $\mathbb{E}\{(\widetilde{H}(t, l_1, k_1) - \widetilde{H}(t, l, k))^2 \mid T^l\}$, is minimized by $k_1$, where $\widetilde{H}(t, l_1, k_1)$ is the rf estimate derived from one of the re-samples. The relevant relationships between $l$ and $l_1$ as well as between $k$ and $k_1$ are found by a Monte Carlo study.

It has been shown that the bias and the mean squared error of the histogram-type estimate (4) with a bootstrap selected $k$ are comparable with those characteristics of Frees' estimate (2) for samples of the same sizes. For larger samples the mean squared error of $\widetilde{H}(t, k, l)$ is less than that of Frees' estimate.

The amount of operations required by (4) with the bootstrap selection of $k$ is much less than that of Frees' estimate.

In conclusion, a reliable computationally tractable histogram-type estimate of the renewal function has been investigated that is also applicable to heavy-tailed interarrival-time distributions.

## APPENDIX

*Proof of Theorem* 2.1.1. By (5) we get for $0 \le t \le t_{\max}(k)$:

$$\sup_t \left|H(t) - \widetilde{H}(t,k,l)\right| \le \sup_t \sum_{n=k+1}^{\infty} \mathbb{P}\{t_n < t\} + k \max_{1\le n\le k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)|.$$

Under the conditions of the theorem it follows by well-known results that

$$\mathbb{P}\left\{\frac{t_n - n\mu}{\sigma\sqrt{n}} < t\right\} = \Phi(t) + \sum_{i=1}^{m-2} \frac{Q_i(t)}{n^{i/2}} + o\left(n^{-(m-2)/2}\right) \qquad (18)$$

holds uniformly in $t \in (-\infty, \infty)$, where $Q_i$ are expressions involving the density function $\varphi(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$ of the standard normal df $\Phi$, the Hermite polynomials and semi-invariants of $\tau_i$ (Ref.[8], Theorem 2.3.2, p. 85; Ref.[21]). Particularly,

$$Q_1(x) = \varphi(x)\frac{1-x^2}{6}\frac{\mathbb{E}(\tau_1-\mu)^3}{\sigma^2}$$

and for $m = 3$

$$\mathbb{P}\left\{\frac{t_n - n\mu}{\sigma\sqrt{n}} < t\right\} = \Phi(t) + \frac{\mathbb{E}(\tau_1-\mu)^3}{6\sigma^3\sqrt{n}}(1-t^2)\frac{1}{\sqrt{2\pi}}\exp\{-t^2/2\} + o\left(\frac{1}{\sqrt{n}}\right)$$

holds.

Hence, $\mathbb{P}\{t_n < t\}$ is defined like the right-hand side of (18) with the replacement of $t$ by $(t - n\mu)/(\sigma\sqrt{n})$.

Hence, $\sum_{n=1}^{\infty} \mathbb{P}\{t_n < t\}$ converges for $m \ge 3$ and

$$\sum_{n=k+1}^{\infty} \mathbb{P}\{t_n < t\} \le c, \quad c > 0 \qquad (19)$$

holds. If

$$k \max_{1\le n\le k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| \le \eta$$

holds for any constant $\eta > 0$, then at $t \in [0, t_{\max}(k)]$

$$\sup_t |H(t) - \widetilde{H}(t,k,l)| \le c + \eta$$

follows. Hence:

$$\mathbb{P}\left\{ \sup_t |H(t) - \widetilde{H}(t,k,l)| > c + \eta \right\} < \mathbb{P}\left\{ k \max_{1\le n\le k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| > \eta \right\}.$$

The right-hand side may be estimated using the asymptotical estimate of the convergence rate of the empirical df to the true df (Ref.[22])

$$\mathbb{P}\left\{ \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| > \eta \right\} \le 2\exp(-2l_n\eta^2), \tag{20}$$

which is satisfied for sufficiently large $l$. Then it follows

$$\mathbb{P}\left\{ \sup_t |H(t) - \widetilde{H}(t,k,l)| > c + \eta \right\} < 2\exp\left(-2\frac{l}{k^3}\eta^2\right) = P(\eta, l, k).$$

Since $k \sim l^\rho$, $0 < \rho < 1/3$, holds, the series $\sum_{l=1}^{\infty} P(\eta, l, k)$ converges at least for one $\eta > 0$, and according to the Borel-Cantelli lemma the assertion of the theorem follows.

*Proof of Theorem* 2.1.2. Using (8) we have for $t \in [0,1]$:

$$\begin{aligned} &\sup_t |H(t) - \widetilde{H}(t,k,l)| \\ &\quad \le (1-\exp(-v))^{k+1}\exp(v) + k \max_{1\le n\le k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| \end{aligned}$$

and for $\alpha > 0$:

$$\begin{aligned} &l^\alpha \sup_t \left|H(t) - \widetilde{H}(t,k,l)\right| \\ &\quad \le l^\alpha(1-\exp(-v))^{k+1}\exp(v) + l^\alpha k \max_{1\le n\le k} \sup_t \left|\mathbb{P}\{t_n < t\} - F_{l_n}(t)\right|. \end{aligned}$$

Since $k = c(v)\cdot l^\rho$, where $\rho < 1/3 - (2/3)\alpha$, $0 < \alpha < 0.5$, $\rho > 0$, for sufficiently large $l$ and the corresponding $c(v)$ we get

$$l^\alpha(1-\exp(-v))^{k+1}\exp(v) \le 1,$$

therefore, if

$$l^\alpha k \max_{1\le n\le k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| \le \eta$$

for any constant $\eta > 0$, then it follows

$$l^\alpha \sup_t |H(t) - \widetilde{H}(t, k, l)| \leq 1 + \eta.$$

Hence,

$$\mathbb{P}\left\{ l^\alpha \sup_t |H(t) - \widetilde{H}(t, k, l)| > 1 + \eta \right\}$$
$$< \mathbb{P}\left\{ l^\alpha k \max_{1 \leq n \leq k} \sup_t |\mathbb{P}\{t_n < t\} - F_{l_n}(t)| > \eta \right\}.$$

Using (20) we have:

$$\mathbb{P}\left\{ l^\alpha \sup_t |H(t) - \widetilde{H}(t, k, l)| > 1 + \eta \right\} < 2 \exp\left( -2 \frac{l}{k^3} \left( \frac{\eta}{l^\alpha} \right)^2 \right) = P(\eta, l, k). \tag{21}$$

Since $k = c(v) \cdot l^\rho$ and $\alpha + 1.5\rho < 0.5$ holds, the series $\sum_{l=1}^\infty P(\eta, l, k)$ converges at least for one $\eta > 0$, and according to the Borel–Cantelli lemma the assertion of the theorem holds.

*Proof of Corollary* 2.1.1. Let the right-hand side of (21) be equal to $0 < \chi < 1$:

$$2 \exp\left( -2 \frac{l}{k^3} \left( \frac{\eta}{l^\alpha} \right)^2 \right) = \chi.$$

Hence, we have

$$\eta = k l^\alpha \sqrt{ - \frac{k \ln(\chi/2)}{2l} }.$$

Then one gets the level of the confidence interval $D = (1 + \eta) l^{-\alpha}$.

*Proof of Theorem* 2.1.3. Since for $t \in \left(0, \frac{c_k}{h_k}\right)$ expression (9) is valid for sufficiently large $n$,

$$\sum_{n=k+1}^\infty \mathbb{P}\{t_n < t\} \sim \sum_{n=k+1}^\infty \Phi\left( \frac{t}{\sqrt{n}} \right)$$

holds. The expansion at the right-hand side converges. Therefore, $\sum_{n=k+1}^\infty \mathbb{P}\{t_n < t\} < c$ follows, where $c$ is a constant. The rest of the proof is similar to the proof of Theorem 2.1.1.

## ACKNOWLEDGMENT

The authors are grateful to the reviewers and editors for their constructive comments that supported them to improve the quality of the paper.

## REFERENCES

1. Ahrends, J.H.; Dieter, U. Computer methods for sampling from Gamma, Beta, Poisson and Binomial distributions. Computing **1974**, *12*, 223–246.
2. Asmussen, S. Renewal theory and queueing algorithms for matrix-exponential distributions. In *Matrix-Analytic Methods in Stochastic Models*; Chakravarthy, S.R., Alfa, A.A., Eds.; New York, 1996; 313–341.
3. Baxter, L.A.; McConalogue, D.J.; Scheuer, E.M.; Blischke, W.R. On the tabulation of the renewal function. Technometrics **1982**, *24* (2), 640–648.
4. Chaudhry, M.L. On computations of the mean and variance of the number of renewals: a unified approach. Journal of the Operational Research Society **1995**, *46*, 1352–1364.
5. Chistyakov, V.P. A theorem on sums of independent positive random variables and its applications to branching random processes. Theory Probab. Appl. **1964**, *9*, 640–648.
6. Deligönül, Z.S. An approximate solution of the integral equation of renewal theory. J. Appl. Prob. **1985**, *22*, 926–931.
7. Devroye, L.; Györfi, L. *Nonparametric Density Estimation, the $L_1$ View*; John Wiley: New York, 1985.
8. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events*; Springer: Berlin, 1997.
9. Feller, W. On the integral equation of renewal theory. Ann. Math. Statist. **1941**, *12*, 243–267.
10. Frees, E.W. Warranty analysis and renewal function estimation. Nav. Res. Logist. Quart. **1986**, *33*, 361–372.
11. Frees, E.W. Nonparametric renewal function estimation. Ann. Statist. **1986**, *14*, 1366–1378.
12. Gnedenko, B.W.; Belyayev, Yu.K.; Solovyev, A.D. *Mathematical Methods of Reliability Theory*; Academic Press: New York, 1969.
13. Goldie, C.M.; Klüppelberg, C. Subexponential distributions. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*; Adler, R., Feldman, R., Taqqu, M.S. Eds.; Birkhäuser: Boston, 1998; 435–459.
14. Grübel, R.; Pitts, S.M. Nonparametric estimation in renewal theory 1: the empirical renewal function. Ann. Statist. **1993**, *21* (3), 1431–1451.
15. Hall, P. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of Multivariate Analysis **1990**, *32*, 177–203.
16. Markovitch, N.M.; Krieger, U.R. Estimating basic characteristics of arrival processes in advanced packet-switched networks by empirical data. In *Proceedings of First IEEE/Popov Workshop on Internet Technologies and Services*, October 25–28; Moscow, Russia, 1999; 70–78.
17. McConalogue, D.J. Numerical treatment of convolution integrals involving distributions with densities having singularities at the origin. Comm. in Statistics **1981**, *Series B10*, 265–280.
18. Mikosch, T.; Nagaev, A.V. Large deviations for heavy-tailed sums with applications to insurance. Extremes **1998**, *1*, 81–110.
19. Mikosch, T. Regular variation, subexponentiality and their applications in probability theory. Technical Report 99-013, University of Groningen, 1999.
20. Nabe, M.; Murata, M.; Miyahara, H. Analysis and modelling of World Wide Web traffic for capacity dimensioning of Internet access lines. Performance Evaluation **1998**, *34*, 249–271.
21. Petrov, V.V. *Sums of Independent Random Variables*; Springer: New York, 1975.
22. Prakasa Rao, B.L.S. *Nonparametric Functional Estimation*; Academic Press: Orlando, Florida, 1983.
23. Prudnikov, A.P.; Brychkov, Yu.A.; Marichev, O.I. Integrals and expansions. *Elementary Functions*; Nauka: Moscow, 1981.
24. Schneider, H.; Lin, B.-S.; O'Cinneide, C. Comparison of nonparametric estimators for the renewal function. Appl. Statist. **1990**, *39* (1), 55–61.
25. Xie, M. On the solution of renewal-type integral equations. Commun. Statist.-Simula. **1989**, *18* (1), 281–293.