

BOOTSTRAP CONFIDENCE INTERVALS FOR INDUSTRIAL RECURRENT EVENT DATA

Oswaldo Anacleto¹ and Francisco Louzada^{2*}

Received May 18, 2009 / Accepted June 17, 2011

ABSTRACT. Industrial recurrent event data where an event of interest can be observed more than once in a single sample unit are presented in several areas, such as engineering, manufacturing and industrial reliability. Such type of data provide information about the number of events, time to their occurrence and also their costs. Nelson (1995) presents a methodology to obtain asymptotic confidence intervals for the cost and the number of cumulative recurrent events. Although this is a standard procedure, it can not perform well in some situations, in particular when the sample size available is small. In this context, computer-intensive methods such as bootstrap can be used to construct confidence intervals. In this paper, we propose a technique based on the bootstrap method to have interval estimates for the cost and the number of cumulative events. One of the advantages of the proposed methodology is the possibility for its application in several areas and its easy computational implementation. In addition, it can be a better alternative than asymptotic-based methods to calculate confidence intervals, according to some Monte Carlo simulations. An example from the engineering area illustrates the methodology.

Keywords: industrial data, recurrent events, bootstrap, asymptotic theory, confidence intervals.

1 INTRODUCTION

In several areas, such as engineering, manufacturing and industrial reliability, we may observe recurrent event data, where the event of interest can be the repeated failures in a piece of equipment, systems which accumulate several repairs, or the number of bugs in a software under study, for instance. There are currently several models and methods developed for the analysis of such data, as described in Hougaard (2000). Approaches often used to model recurrent event data, which allow us to learn about an individual process, are those based on Poisson and renewal processes. (Cox & Isham, 1980; Cox & Lewis, 1966; Andersen, Borgan, Gill & Keiding, 1993; Lawless, 1987; Follmann & Goldberg, 1988; Prentice, Williams & Peterson, 1981).

*Corresponding author

¹Department of Mathematics and Statistics, The Open University, Milton Keynes, United Kingdom.

²Departamento de Matemática Aplicada e Estatística, Universidade de São Paulo, São Carlos, Brazil.
E-mail: louzada@icmc.usp.br

For recurrent event data, it is interesting to study the number of events that occurs over the time. Adding to that, as for each event a cost can be associated, an event cost study may be also important for the analyst. From this perspective, a mean cumulative function (MCF) for the number (or cost) of events per sample unit can be defined. Nelson (1988, 1995) presented a non parametric procedure to calculate confidence intervals for this function. However, as it relies on asymptotic distributional assumptions, the quality of their results can be affected if information about the event of interest are not largely available, which results in a small sample size of units. In this context, computer-intensive methods (Davison & Hinkley, 1999; Chernik, 2008) such as bootstrap can be used to construct confidence limits for the MCF.

In this paper, we present the estimate and confidence limits proposed by Nelson, and also introduce a bootstrap-based technique in order to obtain confidence limits for the MCF. In Section 2, we introduce the MCF estimate proposed by Nelson (1995). Two methods to calculate confidence limits for the MCF are presented in Section 3, the Nelson asymptotic procedure and our proposed technique. Section 4 presents a simulation study in order to compare the two approaches discussed in Section 3 via coverage probabilities, and in Section 5 the methodology is illustrated on a valve seats replacement data set. Some concluding remarks in Section 6 finalize the paper.

2 MODEL FORMULATION AND THE MCF ESTIMATOR

Consider a population of units which are exposed to recurrent events. Despite the occurrence of censoring, an uncensored cumulative history function $Y_i(t)$ for the cost of events is associated for each population unit i . $Y_i(t)$ denotes the cumulative cost of events on unit i up to age t . The model proposed by Nelson (1995) is a population of such uncensored cumulative functions, which extend in principle to any time of interest and does not depend on the censoring of a sample.

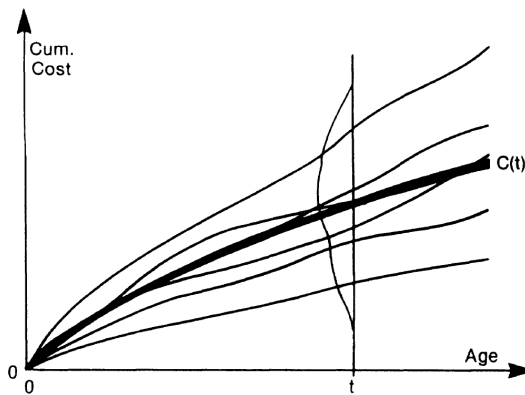


Figure 1 – Population Cumulative Cost Histories (uncensored), distribution at age t (taken from Nelson, 1995).

Figure 1 shows such functions as smooth curves for easy viewing, although each $Y_i(t)$ is better described by staircase functions due to the nature of data (Nelson, 1995). Since different units undergo different number and cost of events at different ages, there is a population distribution of cumulative cost at a given age t . It is assumed that the distribution of the cumulative cost at

any age t has a finite mean $C(t)$, which is called the population mean cumulative function per cost per unit (MCF). The MCF is represented in Figure 1 as a dark curve.

3 ESTIMATION

To estimate the population MCF, consider a sample of units which was exposed to recurrent events and their censored histories. Figure 2 shows these censored histories, where each horizontal line represents an unit cost history, each x denotes an occurrence of an event, and each dashed vertical line denotes the censoring age of a sample unit.

Following Nelson (1995), note that, in the representation presented in Figure 2, the units are shown in an ascending order and numbered backward according to their censoring ages. Also, these censoring ages divide the observed age range into N intervals, as well as these intervals are also numbered backward. Denote the total incremental event cost accumulated over all events in interval i on unit n by Y_{in} where $i, n = 1, 2, \dots, N$. Based on this representation, the estimate for the mean cost cumulative function at a given age t in the interval I is,

$$\begin{aligned}
C^*(t) &= \frac{1}{N} [Y_{NN} + Y_{N,N-1} + Y_{N,N-2} + \cdots + Y_{NI} + \cdots + Y_{N1}] \\
&+ \frac{1}{N-1} [Y_{N-1,N-1} + Y_{N-1,N-2} + \cdots + Y_{N-1,I} + \cdots + Y_{N-1,1}] \\
&+ \frac{1}{N-2} [Y_{N-2,N-2} + \cdots + Y_{N-2,I} + \cdots + Y_{N-2,1}] \\
&\vdots \quad \quad \quad \ddots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
&+ \frac{1}{I+1} [Y_{I+1,I+1} + Y_{I+1,I} + \cdots + Y_{I+1,1}] \\
&+ \frac{1}{I} [Y_{I,I} + \cdots + Y_{I,1}].
\end{aligned} \tag{1}$$

Note that the first row in (1) denotes the total incremental cost of N units in the interval N divided by N , the second row denotes the total incremental cost of $N - 1$ units in the interval $N - 1$ divided by $N - 1$ and so on. The sum in the last row is the total incremental cost up to age t of all I units which are exposed in interval I . It implies that each row represents the average incremental cost per unit for each interval from N up to I . Since the MCF estimate depends on the intervals which have considered for the representation in Figure 2, this estimate and also the confidence limits are conditional on the given censoring ages. Also, it is assumed that the set of units considered for the estimate are a random sample from some population, and the event histories for each unit are statistically independent of their censoring ages. To avoid consideration of ties, it is assumed the the sample ages of recurrences are known exactly and are distinct points on a continuous time scale. Note that the estimate for the mean number cumulative function is the same except that 1 is used as the cost for each event.

From the representation presented in Figure 2 and the property of the variance of a sum of random variables, Nelson (1995) derived the variance of (1), denoted by $V[C^*(t)]$. Since $V[C^*(t)]$ is the variance of a sum, it consists of the variances denoted by $V[Y_{in}]$ of all the

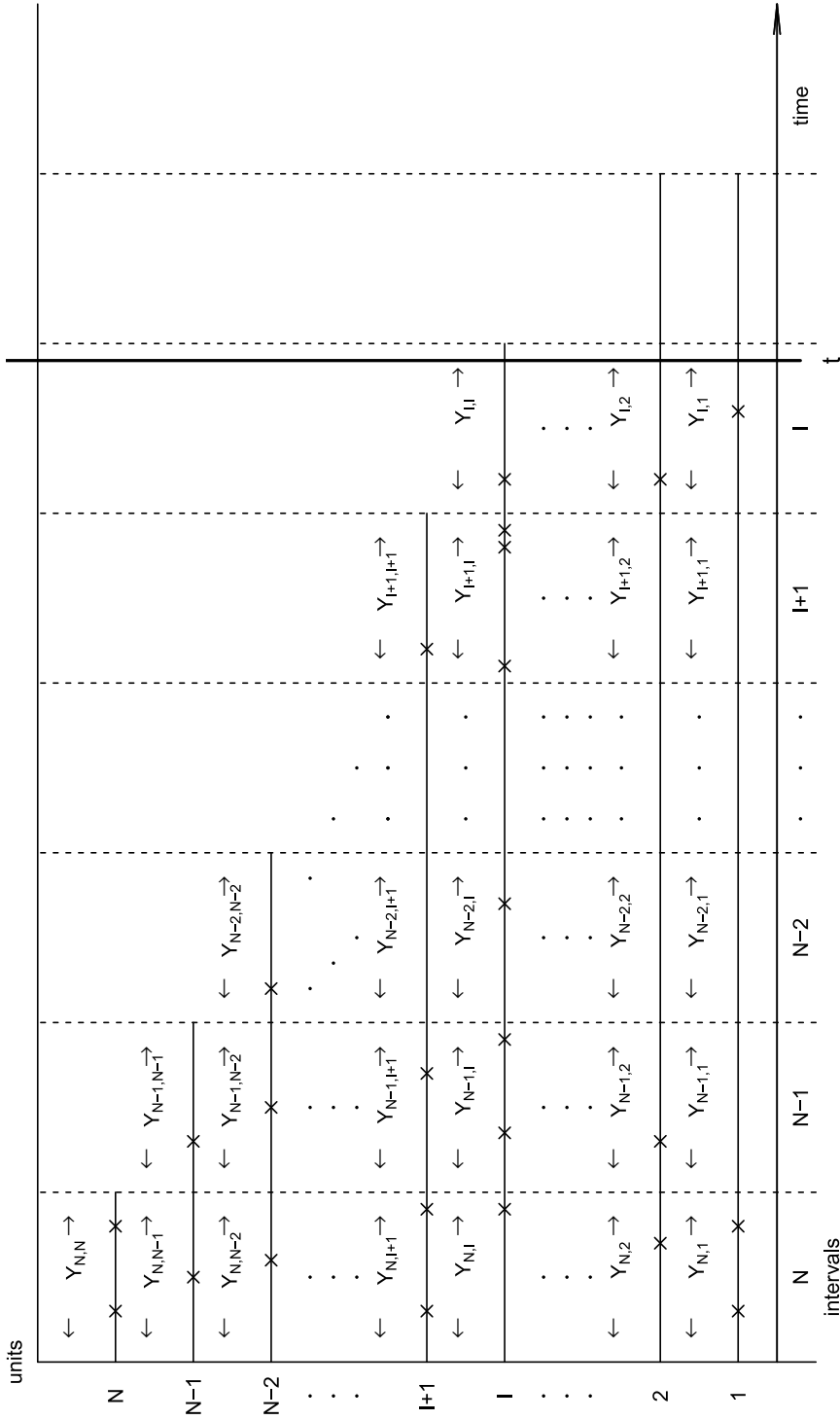


Figure 2 – A diagram of sample of units which was exposed to recurrent events and their censored histories (Nelson, 1995).

$$\begin{aligned}
 & + \frac{2}{(N-1)(N-4)} \sum_{n=1}^{N-4} V(Y_{N-1,n}, Y_{N-4,n}) \\
 & \quad \vdots \quad \quad \quad \vdots \\
 & + \frac{2}{(N-1)(I+1)} \sum_{n=1}^{I+1} V(Y_{N-1,n}, Y_{I+1,n}) \\
 & + \frac{2}{(N-1)I} \sum_{n=1}^I V(Y_{N-1,n}, Y_{I,n}) \tag{2} \\
 & \text{-----} \\
 & \quad \vdots \quad \quad \quad \vdots \\
 & \text{-----} \\
 & \frac{2}{(I+1)I} \sum_{n=1}^I V(Y_{I+1,n}, Y_{I,n}).
 \end{aligned}$$

The first block of terms consists of the individual variances of each of the Y_{in} in (1), the second block consists of the covariances between incremental costs in interval $i = N$ and those of each in subsequent intervals $i = N - 1, i = N - 2, \dots, i = I$. The third block of terms consists of the covariances between incremental costs in interval $i = N - 1$ and those of each in subsequent intervals $i = N - 2, i = N - 3, \dots, i = I$ and so on until the last block, which consists of the covariances between incremental costs in interval $i = I + 1$ and those in the interval $i = I$ up to age t .

Since the variances appearing in the first row of the first block in (2) are N independent observations from the same incremental cost distribution for the interval N , we have

$$V(Y_{NN}) = V(Y_{N,N-1}) = \dots = V(Y_{N1}) = V(Y_{Nn}).$$

Hence, the sum of the first row of the first block in (2) is $NV(Y_{Nn})$. By this reasoning, the $N - 1$ variances in the second row of the first block have a common value $V(Y_{N-1,n})$ and a sum of $(N - 1)V(Y_{N-1,n})$ and so on. Similarly, the covariance terms can be combined, since the covariance terms in a sum in a single row of (2) are all equal. For instance, the first row of the second block has $N - 1$ covariances with a common value $V(Y_{Nn}, Y_{N-1,n})$ and a sum of $(N - 1)V(Y_{Nn}, Y_{N-1,n})$. Hence, the variance of the $C^*(t)$ can be simplified as

$$\begin{aligned}
 V[C^*(t)] &= \frac{1}{N} V(Y_{Nn}) + \frac{1}{N-1} V(Y_{N-1,n}) + \frac{1}{N-2} V(Y_{N-2,n}) + \dots + \frac{1}{I} V(Y_{In}) \\
 &+ \frac{2}{N} [V(Y_{Nn}, Y_{N-1,n}) + V(Y_{Nn}, Y_{N-2,n}) + \dots + V(Y_{Nn}, Y_{I,n})] \\
 &+ \frac{2}{N-1} [V(Y_{N-1n}, Y_{N-2,n}) + \dots + V(Y_{N-1n}, Y_{I,n})] + \dots \\
 &\dots + \frac{2}{I+1} [V(Y_{I+1n}, Y_{I,n})]. \tag{3}
 \end{aligned}$$

For the estimation of the terms in (3), consider the i sample incremental costs $Y_{ii}, Y_{ii-1}, \dots, Y_{i1}$ observed in interval i . These observed costs are a random sample from the incremental cost distribution of the interval i . Thus, their sample variance,

$$V^*(Y_{In}) = + \frac{1}{I} \sum_{n=1}^i (Y_{in} - \bar{Y}_i)^2 \quad (4)$$

is an unbiased estimate of the population variance $V(Y_{in})$. Also, the population covariance $V(Y_{in}, Y_{jn})$ is estimated by the sample covariance,

$$V^*(Y_{IN}, Y_{IJ}) = + \frac{1}{J-1} \sum_{n=1}^J (Y_{in} - \bar{Y}_i)(Y_{jn} - \bar{Y}_j). \quad (5)$$

$$J < I.$$

The inclusion of (4) and (5) into (3) provides an unbiased estimate for the true variance $V[C^*(t)]$ (Nelson, 1995).

4 CONFIDENCE INTERVALS FOR MCF

In this Section, the usual procedure to calculate confidence limits for the Mean Cumulative Function are presented, as well as an alternative based on bootstrap techniques.

4.1 Confidence Intervals Based on Asymptotic Theory

Suppose that N cumulative history functions for cost represented in Figure 2 are a simple random sample from a infinite population. At a given time t , the estimator of the Mean Cumulative Function estimator is given by (3). Since this estimator is the sample mean considering censored histories, by the central-limit theorem (Lehmann, 1999), $C^*(t)$ has a normal distribution with Mean $C(t)$ (the mean cumulative function at the time t) and variance $V^*[C^*(t)]$ (Nelson, 1995). Hence, the two sided normal approximate $(100 - \alpha)\%$ confidence interval for $C(t)$ is given by,

$$C^*(t) \pm K_\alpha * \{V^*[C^*(t)]\}^{1/2} \quad (6)$$

where $V^*[C^*(t)]$ is the $V[C^*(t)]$ estimator and K_α is the $\alpha/2$ standard normal percentile.

This procedure are based on a sample of units, in which the asymptotic based confidence intervals presented here can not perform well if the size sample is small. In this context, computer-intensive methods such as bootstrap can be used to construct confidence intervals for the Mean Cumulative Function.

4.2 Confidence Intervals Based on the Bootstrap

The bootstrap is a computer-intensive method which can be used to obtain confidence intervals for quantities of interest (Efron, 1979). According to Moretti & Mendes (2003), this technique is especially useful for dealing with statistical problems involving a small sample size and those involving estimators whose distribution (exact or asymptotic) has not yet been obtained. The

basic idea is to consider the observed data as a population, and then samples from this population are obtained based on a sampling scheme with replacement from the original sample. If this procedure is repeated several times, different values of the quantities of interest can be obtained, thus providing an empirical distribution of this quantity. Based on this idea, it is possible to construct the percentile confidence intervals (Efron & Tibshirani, 1993; Davison & Hinkley, 1999; Chernik, 2008; Souza, Souza & Staub, 2009) for the MCF, resampling the original set of units exposed to recurrent events and calculating the mean cumulative function estimate for each sample available.

Along these lines, an algorithm to obtain the $100(1 - \alpha)\%$ percentile bootstrap confidence intervals for the MCF is given by the following steps:

Step 1: From the original dataset, obtain B resamples of units based on a sampling with replacement scheme;

Step 2: To each of the B resamples, calculate the Mean Cumulative Function estimate;

Step 3: Based on the estimates obtained from the resamples of the original dataset, calculate the $\frac{\alpha}{2}$ and $(1 - \frac{\alpha}{2})$ percentiles from the empirical distributions for each recurrent time for the units from the original dataset, provided for the B sets of estimatives calculated from the B resamples.

A program to calculate the bootstrap confidence intervals for the MCF is available from the authors. An implementation of the variance estimate of (3) as well as asymptotic confidence limits are provided by the SAS software.

5 A SIMULATION STUDY

In order to assess the efficiency and have a comparison of the confidence intervals provided by the asymptotic theory and the bootstrap, as well as verifying the sample size influence in these methods, a simulation study was performed to check the coverage probability and the mean range of the confidence intervals developed here.

The study considered the sample sizes of 10, 30 and 100. For each sample size, four scenarios based on the parameter settings for the data generation were considered: the number of events in each sample unit was generated from a Poisson distribution with means 2 and 5, and the recurrence times were generated from an Weibull distribution with scale parameter 1000 and shape parameter equal to 1 and 3, assuming the biggest time generated for each unit as a censored event. We considered then four different scenarios: Scenario 1 (Poisson distribution with mean equals to 2 and Weibull distribution with shape equals 1), Scenario 2 (Poisson distribution with mean equals to 2 and Weibull distribution with shape equals 3), Scenario 1 (Poisson distribution with mean equals to 5 and Weibull distribution with shape equals 1) and Scenario 1 (Poisson distribution with mean equals to 5 and Weibull distribution with shape equals 3). It was not considered the presence of ties. We studied the behavior of the 90% confidence limits for the mean number cumulative function, which is a MCF particular case.

To determine the coverage probability, it was first generated an original data set and their MCF estimate was calculated. Then, 399 samples was generated considering the same specifications that it was used to generate the original data set. Then, the number B of resamples was set at

399. According to Hall (1986) this number of replications is enough to obtain a critical level of 0.05 from the 0.95 percentile of the empirical distribution of the test statistics. To set up the Monte Carlo simulation, this procedure was repeated 399 times. The MCF estimate was calculated for each of the samples. In order to calculate the coverage probability, it was necessary to set percentiles, since the recurrence times from the 399 samples and the original sample were generated from a probability distribution, thus varying from sample to sample. The percentiles 10, 25, 50, 75 and 90. were chosen. With this, for each considered percentile it was verified whether the confidence intervals of the 399 samples covered the MCF estimate obtained in the original sample. If not, it was also verified if the related percentile MCF estimate lied above the upper limit or below the lower limit.

The results for all scenarios considered are presented in Table 1. It contains, for each verified percentile, the time related to the percentile in the original sample, and, for the two methods considered, the coverage probability, the average range of the interval and the standard deviation of the interval range in the 399 samples apart from the original sample. The relative difference of these quantities between the methods is also presented, always considering the quantity provided by the asymptotic method in the denominator.

For sample sizes bigger than 30, in all the scenarios, it was verified that the bootstrap method and the asymptotic method provides similar coverage probabilities. Besides, the results indicate that the coverage probabilities are underestimated in the smallest percentiles and overestimated in the biggest quantiles, and it tends to decrease as the sample size increases as well as the average confidence intervals and its standard deviations do. However, for the sample size 10, the bootstrap method provides smaller confidence intervals average ranges as well as smaller standard deviations of these ranges. It indicates that, since the asymptotic methods requires a sufficiently large sample for developing inferences, the bootstrap method can be used as an alternative approach to provide confidence limits for the mean cumulative function in presence of small samples.

6 THE VALVE SEATS REPLACEMENT DATA

The presented methodology was applied to a real dataset provided by Nelson (1995). The data is the valve seats replacement over the time in 41 engines in a fleet. In this case, the recurrent event is the valve seats replacements in each of the engines. The interest relies on verifying if the replacement rate increases with engine age (in days).

The confidence limits obtained via asymptotic theory and the bootstrap method are presented in the Figure 3, as well as in Table 1. It is verified that both methods indicate that the replacement rate is constant over the time. Besides, it is also verified that the confidence limits becomes bigger as the age increases, since information about the valve seats replacements decreases over the time. However, the asymptotic procedure leads to negative lower confidence limits, which is impossible from the practical point of view. This problem is overcome by considering our bootstrap procedure. Also, for approximately 92% of the recurrence times the bootstrap procedure produces confidence interval ranges approximately 50% smaller than the asymptotic

Table 1 – Coverage Probabilities and Range of Asymptotic and Bootstrap Confidence Intervals.

Scenario 1: Poisson (2) and Weibull (1)											
			Asymptotic Method			Bootstrap Method			Relative Difference (%) (asym-boot)/asym		
Percentile	Time	Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range		
			average	std. dev.		average	std. dev.		average	std. dev.	
S A M	Q10	46	0.99	0.57	0.10	0.99	0.42	0.08	0.00	0.25	0.24
	Q25	92	0.99	0.88	0.19	0.99	0.75	0.18	0.00	0.15	0.09
	Q50	246	0.95	1.14	0.30	0.94	1.01	0.28	0.01	0.12	0.07
	Q75	1,001	0.91	1.29	0.33	0.86	1.15	0.29	0.06	0.10	0.10
	Q90	1,523	0.86	1.37	0.37	0.83	1.22	0.32	0.04	0.11	0.15
P L E	Q10	70	0.99	0.34	0.04	0.99	0.31	0.05	0.00	0.08	-0.02
	Q25	177	0.99	0.51	0.06	0.99	0.48	0.07	-0.01	0.05	-0.15
	Q50	410	0.99	0.66	0.09	0.99	0.63	0.09	0.00	0.04	0.00
	Q75	818	0.94	0.75	0.10	0.94	0.73	0.11	0.00	0.03	-0.08
	Q90	1,434	0.88	0.83	0.12	0.85	0.79	0.12	0.03	0.05	0.05
I Z E	Q10	76	0.98	0.19	0.02	0.99	0.19	0.02	-0.01	0.02	-0.27
	Q25	191	0.98	0.29	0.02	0.98	0.28	0.02	0.00	0.02	-0.15
	Q50	483	0.99	0.37	0.03	0.97	0.37	0.03	0.02	0.01	-0.15
	Q75	876	0.92	0.42	0.03	0.91	0.41	0.04	0.01	0.02	-0.22
	Q90	1,364	0.87	0.46	0.03	0.86	0.44	0.04	0.01	0.02	-0.27

Table 1 (continuation) – Coverage Probabilities and Range of Asymptotic and Bootstrap Confidence Intervals.

Scenario 2: Poisson (2) and Weibull (1)													
			Asymptotic Method				Bootstrap Method				Relative Difference (%) (asym-boot)/asym		
Percentile	Time	Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range	
			average	std. dev.		average	std. dev.		average	std. dev.			
S 10	Q10	0.99	0.56	0.09	0.99	0.41	0.08	0.99	0.41	0.08	0.00	0.26	0.14
	Q25	0.99	0.88	0.18	0.99	0.74	0.17	0.99	0.74	0.17	0.01	0.16	0.04
	Q50	0.95	1.15	0.30	0.94	1.01	0.28	0.94	1.01	0.28	0.01	0.13	0.06
A 30	Q75	0.91	1.29	0.34	0.86	1.15	0.30	0.86	1.15	0.30	0.06	0.11	0.11
	Q90	0.86	1.38	0.38	0.82	1.23	0.32	0.82	1.23	0.32	0.05	0.11	0.15
	Q10	0.99	0.33	0.04	0.99	0.30	0.04	0.99	0.30	0.04	0.00	0.08	0.00
E 100	Q25	0.99	0.50	0.06	0.99	0.47	0.07	0.99	0.47	0.07	0.00	0.06	-0.15
	Q50	0.98	0.65	0.10	0.99	0.63	0.10	0.99	0.63	0.10	-0.01	0.04	-0.06
	Q75	0.92	0.75	0.10	0.92	0.73	0.11	0.92	0.73	0.11	0.00	0.02	-0.06
S I Z E	Q90	0.87	0.83	0.13	0.83	0.79	0.12	0.83	0.79	0.12	0.05	0.05	0.03
	Q10	0.99	0.18	0.01	0.99	0.18	0.01	0.99	0.18	0.01	0.00	0.00	0.00
	Q25	0.98	0.27	0.02	0.97	0.27	0.02	0.97	0.27	0.02	0.01	0.00	0.00
E	Q50	0.97	0.37	0.03	0.90	0.37	0.03	0.90	0.37	0.03	0.07	0.00	0.00
	Q75	0.92	0.42	0.03	0.88	0.42	0.03	0.88	0.42	0.03	0.04	0.00	0.00
	Q90	0.86	0.45	0.03	0.89	0.45	0.03	0.89	0.45	0.03	-0.04	0.00	0.00

Table 1 (continuation) – Coverage Probabilities and Range of Asymptotic and Bootstrap Confidence Intervals.

Scenario 3: Poisson (5) and Weibull (1)													
			Asymptotic Method				Bootstrap Method				Relative Difference (%) (asym-boot)/asym		
Percentile	Time	Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range	
			average	std. dev.		average	std. dev.		average	std. dev.			
S A M	Q10	41	0.99	0.77	0.17	0.99	0.63	0.15	-0.01	0.17	0.13		
	Q25	140	0.98	1.20	0.28	0.95	1.05	0.24	0.03	0.12	0.11		
	Q50	501	0.94	1.64	0.41	0.91	1.45	0.39	0.03	0.12	0.05		
	Q75	1,067	0.88	1.93	0.46	0.83	1.72	0.43	0.06	0.11	0.06		
	Q90	1,523	0.82	2.12	0.50	0.76	1.87	0.43	0.07	0.12	0.13		
P L E	Q10	58	0.99	0.46	0.05	0.99	0.43	0.06	0.00	0.06	-0.12		
	Q25	183	0.95	0.70	0.09	0.97	0.67	0.10	-0.02	0.05	-0.08		
	Q50	548	0.91	0.96	0.13	0.90	0.93	0.13	0.01	0.04	0.00		
	Q75	1,074	0.90	1.14	0.15	0.88	1.10	0.16	0.02	0.04	-0.05		
	Q90	1,634	0.89	1.27	0.19	0.89	1.20	0.18	0.00	0.05	0.03		
S I Z E	Q10	107	0.99	0.27	0.02	0.99	0.26	0.03	0.00	0.03	-0.23		
	Q25	261	0.97	0.41	0.03	0.97	0.40	0.03	0.00	0.03	-0.08		
	Q50	561	0.92	0.55	0.04	0.94	0.54	0.05	-0.02	0.02	-0.18		
	Q75	1,103	0.90	0.65	0.05	0.93	0.64	0.06	-0.03	0.01	-0.22		
	Q90	1,642	0.92	0.71	0.05	0.92	0.70	0.07	0.00	0.01	-0.36		

Table 1 (continuation) – Coverage Probabilities and Range of Asymptotic and Bootstrap Confidence Intervals.

Scenario 4: Poisson (5) and Weibull (3)												
			Asymptotic Method				Bootstrap Method				Relative Difference (%) (asyp-boot)/asyp	
Percentile	Time	Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range		Cov. Prob.	CI Range
			average	std. dev.		average	std. dev.		average	std. dev.		
S A M	Q10	430	0.99	0.78	0.15	0.99	0.64	0.15	0.00	0.19	0.06	0.06
	Q25	574	0.99	1.20	0.27	0.99	1.04	0.25	0.00	0.13	0.09	0.09
	Q50	854	0.98	1.63	0.41	0.93	1.42	0.36	0.05	0.13	0.13	0.13
	Q75	1,037	0.92	1.98	0.49	0.89	1.76	0.46	0.03	0.12	0.06	0.06
	Q90	1,192	0.89	2.22	0.53	0.84	1.95	0.48	0.06	0.12	0.10	0.10
P L E	Q10	477	0.99	0.44	0.05	0.99	0.41	0.06	0.00	0.07	-0.11	-0.11
	Q25	666	0.95	0.69	0.08	0.95	0.65	0.08	0.00	0.06	0.02	0.02
	Q50	855	0.91	0.96	0.12	0.89	0.93	0.13	0.02	0.04	-0.09	-0.09
	Q75	1,062	0.90	1.15	0.15	0.86	1.11	0.16	0.04	0.03	-0.08	-0.08
	Q90	1,216	0.82	1.28	0.16	0.82	1.22	0.16	0.00	0.04	0.00	0.00
S I Z E	Q10	414	0.99	0.23	0.02	0.99	0.22	0.02	0.00	0.02	-0.20	-0.20
	Q25	598	0.94	0.37	0.02	0.95	0.36	0.03	-0.01	0.01	-0.38	-0.38
	Q50	829	0.87	0.53	0.04	0.86	0.52	0.04	0.01	0.02	-0.19	-0.19
	Q75	1,044	0.89	0.64	0.05	0.86	0.63	0.06	0.03	0.01	-0.35	-0.35
	Q90	1,204	0.88	0.71	0.05	0.87	0.69	0.06	0.01	0.02	-0.13	-0.13

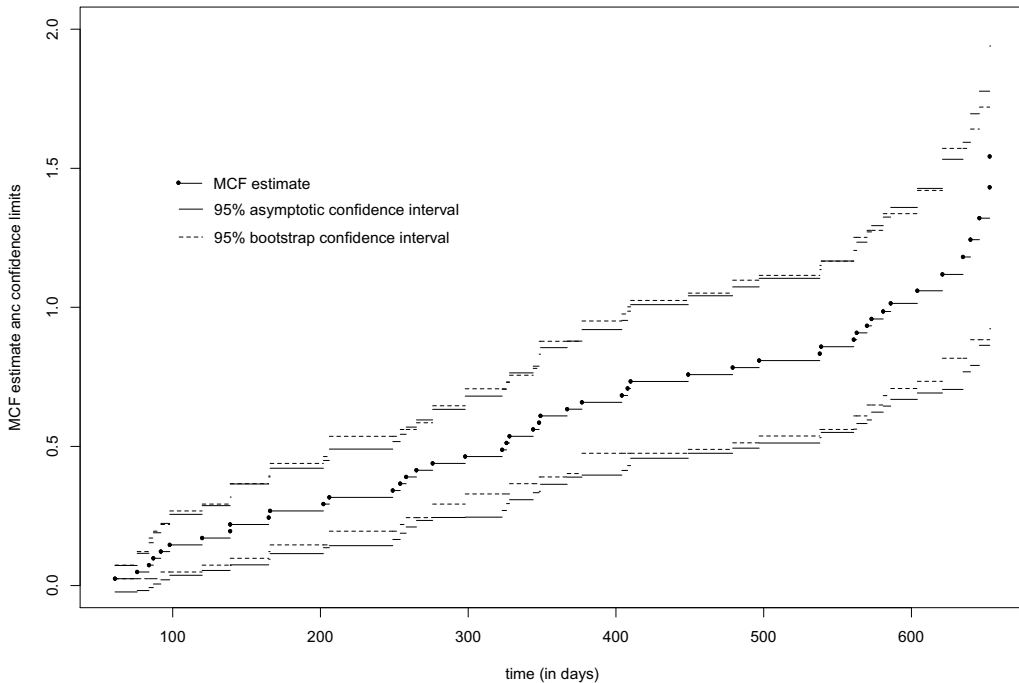


Figure 3 – 95% Asymptotic and bootstrap Confidence Intervals for the MCF.

one. Even though these results are not conclusive, they provide an indication of the advantage of the practical use of the bootstrap confidence interval procedure over the asymptotic method.

7 CONCLUDING REMARKS

In this work, we presented the estimate and confidence intervals based in the asymptotic theory proposed by Nelson (1995) for the Mean Cumulative Function, using non parametric methodology for recurrent events data. Also, it was presented an implementation of the bootstrap technique for the construction of confidence limits for the MCF. These two procedures were applied in a real dataset. One of the advantages of the proposed methodology presented here is the possibility for its application in several areas, its easy computational implementation.

Our simulation results suggest that the confidence intervals based on the two procedures are similar to moderate and large sample sizes. However, for small sample sizes, the bootstrap method provides smaller confidence intervals ranges as well as smaller standard deviations of these ranges. Hence, the bootstrap method can be used as an alternative approach to provide confidence intervals for the Mean Cumulative Function, in particular when there are restrictions regarding the availability of information about the event under study.

We only considered the percentile bootstrap method to develop the confidence intervals for the MCF, since such method is the most straightforward one. Also, we keep the non-parametric nature of you the MCF estimator, which is non-parametric. However, other bootstrap schemes,

Table 2 – 95% asymptotic and bootstrap Confidence Intervals for the MCF.

Recurrence time	Asymptotic			Bootstrap		
	MCF	CI 95% (Lower limit)	CI 95% (Upper limit)	MCF	CI 95% (Lower limit)	CI 95% (Upper limit)
61	0.0244	-0.0234	0.0722	0.0244	0.0244	0.0732
76	0.0488	-0.0180	0.1155	0.0488	0.0244	0.1220
84	0.0732	-0.0075	0.1539	0.0732	0.0244	0.1707
87	0.0976	0.0056	0.1895	0.0976	0.0244	0.1951
92	0.1220	0.0205	0.2234	0.1220	0.0488	0.2195
98	0.1463	0.0368	0.2559	0.1463	0.0488	0.2683
120	0.1707	0.0541	0.2873	0.1707	0.0732	0.2927
139	0.1951	0.0723	0.3179	0.1951	0.0976	0.3659
139	0.2195	0.0741	0.3649	0.2195	0.0976	0.3659
165	0.2439	0.0943	0.3936	0.2439	0.1220	0.3902
166	0.2683	0.1149	0.4217	0.2683	0.1463	0.4390
202	0.2927	0.1359	0.4494	0.2927	0.1463	0.4634
206	0.3171	0.1434	0.4908	0.3171	0.1951	0.5366
249	0.3415	0.1655	0.5174	0.3415	0.1951	0.5366
254	0.3659	0.1879	0.5438	0.3659	0.2195	0.5610
258	0.3902	0.2107	0.5697	0.3902	0.2439	0.5610
265	0.4146	0.2339	0.5954	0.4146	0.2439	0.5854
276	0.4390	0.2448	0.6332	0.4390	0.2927	0.6463
298	0.4634	0.2459	0.6809	0.4634	0.3293	0.7073
323	0.4878	0.2700	0.7056	0.4878	0.3293	0.7073
326	0.5122	0.2944	0.7300	0.5122	0.3293	0.7317
328	0.5366	0.3086	0.7646	0.5366	0.3659	0.7561
344	0.5610	0.3335	0.7885	0.5610	0.3659	0.7805
348	0.5854	0.3388	0.8319	0.5854	0.3902	0.8780
349	0.6098	0.3641	0.8554	0.6098	0.3902	0.8780
367	0.6341	0.3897	0.8786	0.6341	0.4024	0.8780
377	0.6585	0.3969	0.9202	0.6585	0.4756	0.9512
404	0.6835	0.4138	0.9533	0.6829	0.4756	0.9762
408	0.7085	0.4311	0.9860	0.7079	0.4756	1.0019
410	0.7335	0.4574	1.0096	0.7329	0.4756	1.0250
449	0.7585	0.4754	1.0417	0.7579	0.4893	1.0513
479	0.7835	0.4937	1.0734	0.7829	0.5131	1.0976
497	0.8085	0.5123	1.1047	0.8079	0.5375	1.1153
538	0.8335	0.5314	1.1357	0.8329	0.5610	1.1512
539	0.8585	0.5507	1.1664	0.8579	0.5610	1.1662
561	0.8835	0.5625	1.2045	0.8829	0.6098	1.2518
563	0.9085	0.5826	1.2345	0.9079	0.6098	1.2518

Table 2 (continuation) – 95% asymptotic and bootstrap Confidence Intervals for the MCF.

Recurrence time	Asymptotic			Bootstrap		
	MCF	CI 95% (Lower limit)	CI 95% (Upper limit)	MCF	CI 95% (Lower limit)	CI 95% (Upper limit)
570	0.9335	0.5955	1.2716	0.9329	0.6491	1.2768
573	0.9585	0.6232	1.2938	0.9579	0.6491	1.2768
581	0.9849	0.6451	1.3246	0.9829	0.6829	1.3369
586	1.0143	0.6692	1.3593	1.0092	0.7082	1.3369
604	1.0597	0.6920	1.4275	1.0387	0.7340	1.4202
621	1.1185	0.7048	1.5323	1.0841	0.8170	1.5715
635	1.1810	0.7685	1.5936	1.1429	0.8170	1.5715
640	1.2435	0.7911	1.6959	1.2054	0.8840	1.6409
646	1.3205	0.8635	1.7774	1.2679	0.8840	1.7201
653	1.4316	0.9232	1.9399	1.3449	0.9702	2.1313
653	1.5427	0.9079	2.1774	1.4560	0.9702	2.1313

such as the normal, percentile t and the pivotal method (Davison & Hinkley, 1999; Chernik, 2008), can also be considered in the context of obtaining confidence intervals for the MCF.

ACKNOWLEDGMENTS

We would like to thank the referees for their very useful comments which substantially improved the paper. The research was partially supported by the Brazilian Organizations CAPES and CNPQ.

REFERENCES

- [1] ANDERSEN PK, BORGAN O, GILL RD & KEIDING. 1993. *Statistical Models Based on Counting Processes*. New York: Springer.
- [2] COX DR & LEWIS PAW. 1966. *The Statistical Analysis of Series of Events*. London: Methuen.
- [3] COX DR & ISHAM V. 1980. *Point Processes*. London: Chapman and Hall.
- [4] CHERNICK MR. 2008. *Bootstrap Methods: A Guide For Practitioners and Researches*. New York: Wiley.
- [5] DAVISON AC & HINKLEY DV. 1997. *Bootstrap Methods and Their Application*. Cambridge University Press.
- [6] EFRON B. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7: 1–26.
- [7] EFRON B & TIBSHIRANI RJ. 1993. *An introduction to the bootstrap*. New York: Chapman & Hall.
- [8] FOLLMANN DA & GOLDBERG MS. 1988. Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, 30: 389–396.

- [9] HALL P. 1986. On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, **14**: 125–129.
- [10] HOUGAARD P. 2000. *Analysis of Multivariate Survival Data*. New York: Springer-Verlag.
- [11] LAWLESS JF. 1987. Regression methods for Poisson process model. *Journal of the American Statistical Association*, **82**: 808–815.
- [12] LEHMANN EL. 1999. *Elements of Large-sample Theory*. New York: Springer-Verlag.
- [13] MORETTI AR & MENDES BVM. 2003. Sobre a precisão das estimativas de máxima verossimilhança nas distribuições bivariadas de valores extremos. *Pesquisa Operacional [online]*, **23**(2): 301–324.
- [14] NELSON W. 1988. Graphical Analysis of System Repair Data. *Journal of Quality Technology*, **20**: 24–35.
- [15] NELSON W. 1995. Confidence Limits for Recurrence Data-Applied to Cost or Number of Product Repair. *Technometrics*, **37**: 147–157.
- [16] PRENTICE RL, WILLIAMS BJ & PETERSON AV. 1981. On the regression analysis of multivariate failure time data. *Biometrika*, **68**: 373–379.
- [17] SOUZA MO, SOUZA GS & STAUB RB. 2009. Influência de variáveis contextuais em medidas não-paramétricas de eficiência: uma aplicação com métodos de reamostragem. *Pesquisa Operacional [online]*, **29**(2): 289–302.