Nonparametric estimation and bootstrap confidence intervals for the optimal maintenance time of a repairable system

Gustavo L. Gilardoni [a,*], Maristela D. de Oliveira [b], Enrico A. Colosimo [c]

[a] *Universidade de Brasília, Brazil*
[b] *Universidade Federal da Bahia, Brazil*
[c] *Universidade Federal de Minas Gerais, Brazil*

ARTICLE INFO

ABSTRACT

Consider a repairable system operating under a maintenance strategy that calls for complete preventive repair actions at pre-scheduled times and minimal repair actions whenever a failure occurs. Under minimal repair, the failures are assumed to follow a nonhomogeneous Poisson process with an increasing intensity function. This paper departs from the usual power-law-process parametric approach by using the constrained nonparametric maximum likelihood estimate of the intensity function to estimate the optimum preventive maintenance policy. Several strategies to bootstrap the failure times and construct confidence intervals for the optimal maintenance periodicity are presented and discussed. The methodology is applied to a real data set concerning the failure histories of a set of power transformers.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Consider a nonhomogeneous Poisson process (NHPP) having an increasing intensity $\lambda(t)$ and hence a convex mean function $\Lambda(t) = \int_0^t \lambda(u) \, du$, and let $N_i(t)$ be K independent realizations observed along possibly overlapping time intervals $0 \leq t \leq T_i$ ($1 \leq i \leq K$). This paper deals with the construction of bootstrap confidence intervals (CIs) for a functional $\tau = \tau[\lambda(\cdot)]$ based on the constrained nonparametric maximum likelihood estimate (NPMLE) of $\lambda(t)$. Its motivation comes from the problem of estimating the optimal periodicity τ of perfect preventive maintenance (PM) for a repairable system that is subject to minimal repair (MR) after each failure.

Repairable systems are allowed to experience more than one failure throughout their life. After each failure, a maintenance activity is necessary for the system to return to operating condition. Usually, such activity repairs just the damaged part of the equipment, leaving it in the same condition as it was just before the failure. This kind of repair is usually called MR. On the other hand, a planned maintenance activity, referred to as PM renews the entire system and brings it to *as good as new* condition. Determining the optimal maintenance time of a repairable system has been of interest since the early work of Barlow and Hunter (1960). More precisely, suppose that under MR the system failures are modeled as an NHPP with an increasing intensity $\lambda(t)$ and let C_{PM} and C_{MR} be the fixed costs of the PM and MR actions. In order to minimize the system operational cost per unit of time, PMs should be performed periodically every τ units of time, where τ is the solution of

$$\tau \lambda(\tau) - \int_0^\tau \lambda(u) \, du = C_{PM}/C_{MR} \qquad (1)$$

---

* Correspondence to: Departamento de Estatística, Campus Darcy Ribeiro, Asa Norte, Brasília, DF 70910-900, Brazil. Tel.: +55 61 3107 3956.
  E-mail addresses: gilardon@gmail.com, gilardon@unb.br (G.L. Gilardoni), marridias@gmail.com (M.D. de Oliveira), enricoc@est.ufmg.br (E.A. Colosimo).

(cf. Barlow and Hunter, 1960 or Gilardoni and Colosimo, 2007). Since the intensity $\lambda(t)$ is typically unknown, in practice one obtains an estimate $\hat{\lambda}(t)$ based on the failure histories of one or more realizations of the system under consideration and uses it to compute an estimate of τ by solving (1) with $\hat{\lambda}$ instead of λ. Gilardoni and Colosimo (2007, 2011) consider both parametric and nonparametric estimation of τ. For the former, they use a power law process (PLP) intensity $\lambda(t) = (\beta/\theta) \, (t/\theta)^{\beta-1}$ (cf. Crow, 1974 or Rigdon and Basu, 2000). On the other hand, their nonparametric approach first pools or superimposes the K failure histories into one realization using the *Total Time on Test* (TTT) transform. Based on this one realization, they follow Cowling et al. (1996) to bootstrap the failure times in the TTT scale and compute a kernel estimate and bootstrap confidence bands for the intensity of the superimposed process. These confidence bands are then transformed back to the original time scale using the TTT inverse transformation. Finally, the bands for the intensity of the original process are plugged into (1) to obtain the desired intervals for τ. However, this approach has the drawback that the monotonicity restriction on $\lambda(t)$, without which (1) may not have a solution, is not taken into account by the kernel estimates used in the process.

The unconstrained NPMLE of the mean function $\Lambda(t)$ is the Nelson–Aalen estimate, i.e., the step function having jumps of size (1/number of realizations at risk) at each observed event time (cf. Aalen, 1978, Ramlau-Hansen, 1983 or Aalen et al., 2008). On the other hand, NPMLEs of the intensity function of an NHPP subject to monotonicity constraints have been considered in Boswell (1966) and Bartoszynski et al. (1981). When only one realization is observed, Boswell (1966) showed that the constrained NPMLE of $\lambda(t)$ is the right derivative of the greatest convex minorant (GCM) of the unconstrained NPMLE. Later, Bartoszynski et al. (1981) obtained an expression for the constrained NPMLE when several realizations of the NHPP are observed along overlapping time intervals. Recently, Gilardoni and Colosimo (2011) showed that the TTT transform provides the right connection between the algorithm of Boswell and that of Bartoszynski et al. More precisely, they showed that the algorithm to obtain the constrained NPMLE in the many realizations setup can be described as follows. First, the many realizations are superimposed in the TTT scale. Based on this one realization, one then uses Boswell's algorithm to estimate the intensity of the superimposed process. Finally, to obtain the constrained NPMLE of the intensity of the original process, one goes back to the original time scale using the inverse of the TTT transformation.

Based on the constrained NPMLE of the intensity function, we propose several strategies to obtain bootstrap samples of τ. On the one hand, we can work in the original time scale and draw a bootstrap sample of the systems by drawing randomly with replacement K of the observed systems, say $N_1^*(t), \ldots, N_K^*(t)$, based on which we can then compute the constrained NPMLE $\hat{\lambda}^*(t)$ and plug it into (1) to get a bootstrap sample τ^* of the estimate of the optimal maintenance time. On the other hand, we can first superimpose the systems in the TTT scale and use Boswell's algorithm to compute the constrained NPMLE of the intensity of the superimposed process, say $\hat{\lambda}_S(s)$. Then, by analogy to Cowling et al. (1996), we can generate an NHPP with intensity $\hat{\lambda}_S(s)$ and use the simulated failure times and again Boswell's algorithm to compute a bootstrap sample $\hat{\lambda}_S^*(s)$. Finally, $\hat{\lambda}_S^*(s)$ is transformed back into the original time scale to get an estimate $\hat{\lambda}^*(t)$ of the intensity of the original process, which is plugged into (1) to get the desired bootstrap sample τ^*. In a sense, the question between these two strategies is whether to bootstrap first and pool second or *vice versa*.

Most of the literature on repairable systems considers a parametric approach using the PLP intensity. Recent examples are Yu et al. (2007, 2008), Pan and Rigdon (2009) and Ryan et al. (2011). Further developments regarding optimal maintenance, which will not be pursued here, can be found in Gerstack (1977), Sloan (2008), Colosimo et al. (2010) and Tsai et al. (2011). From a purely mathematical point of view, the problem of finding the NPMLE of the intensity of an NHPP under a monotonicity restriction is formally the same as that of estimating a density function under the same kind of constraint, which is solved in the fundamental work of Barlow et al. (1972). Although it is not obvious, these two problems are also equivalent to that of estimating a monotonic hazard function based on right censored data, which is addressed by Tsai (1988). Hall and Huang (2001) follow a different approach to constrained estimation which is based on kernel estimates and hence does not rely upon maximum likelihood.

The rest of this paper is organized as follows. Sections 2 and 3 contain brief reviews of the derivation of the constrained NPMLE, respectively, of the determination of the optimal periodicity of PM. Section 4 describes in detail the strategies used to generate the bootstrap samples of τ, including the two mentioned above and three others that are slight modifications of the second one. Section 5 describes a Monte Carlo experiment that was run to study the performance of the nonparametric bootstrap CIs and to compare them with the parametric CI based on the PLP and maximum likelihood asymptotics. Section 6 gives an application of the methodology to a real data set consisting of the failure histories of 40 electrical power transformers from a Brazilian electrical power company (cf. Fig. 1 and Gilardoni and Colosimo, 2007, 2011; Tsai et al., 2011). Section 7 contains some concluding remarks and, finally, we include in an Appendix a proof of the form of the constrained NPMLE estimate of the intensity function that is self-contained, in the sense that it only requires knowledge of some elementary properties of convex sets.

2. Nonparametric inference for an increasing intensity function

There are basically two ways to observe data from a repairable system, depending on whether data collection ceases after a specified number of failures n or at some predetermined time T. These sampling schemes are said to be *failure*, respectively, *time truncated* (cf. Rigdon and Basu, 2000). Although many inference procedures require only slight adjustments depending on which sampling scheme was selected, the distinction is crucial for us because the constrained NPMLE does not exist in the failure truncated case (cf. Bartoszynski et al., 1981). Hence, from now on we will consider just the time truncated case.
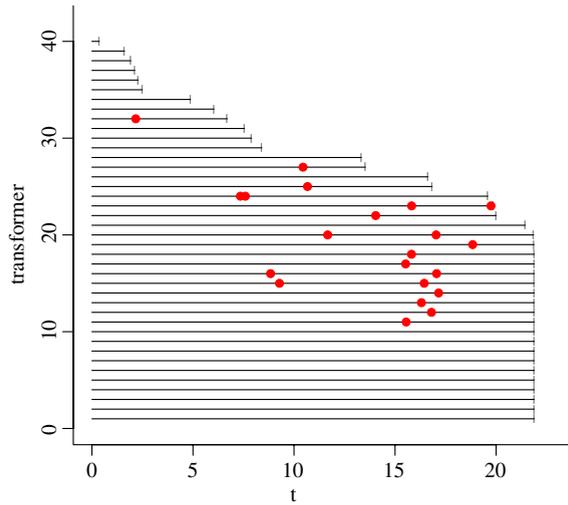
Fig. 1. Failure histories of electrical power transformers. Dots represent failure times. Time unit is 1000 h.

Suppose first that only one system is observed and let $N(t) = \sum_{j=1}^{n} I(t \geq t_j)$ be the number of failures up to time t, where $0 < t_1 < \cdots < t_n < T$ are the observed failure times. Since for a time truncated system the *at risk* process is $Y(t) = I \, (0 \leq t \leq T)$, it follows from Aalen (1978, Section 5) that the log-likelihood is

$$\ell(\lambda) = \int_0^T \log Y(s) \, dN(s) + \int_0^T \log \lambda(s) \, dN(s) - \int_0^T \lambda(s) Y(s) \, ds$$

$$= \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \lambda(u) \, du, \tag{2}$$

where we have omitted an additive constant that does not depend on $\lambda(\cdot)$ (see also Rigdon and Basu, 2000). The uncon-strained NPMLE of $\Lambda(t)$, in the sense of Scholz (1980), is the step function $\tilde{\Lambda}(t) = N(t)$. Since its derivative is almost every-where zero, the unconstrained NPMLE of $\lambda(t)$ does not exist. However, depending on the type of restriction imposed on $\lambda(t)$, the constrained NPMLE might exist. Indeed, Boswell (1966) showed that the maximum of (2) among increasing intensities is attained for a right continuous step function $\hat{\lambda}(t)$ which has jumps at a subset of the t_i's and such that $\hat{\lambda}(0) = 0$ and

$$\hat{\lambda}(t_j) = \max_{1 \leq h \leq j} \min_{j \leq k \leq n+1} \frac{k - h}{t_k - t_h}, \tag{3}$$

where we put $t_0 = 0$ and $t_{n+1} = T$. An alternative representation of the constrained NPMLE is $\hat{\lambda}(t) = \hat{\Lambda}'(t + 0)$, where $\hat{\Lambda}(t) = \sup\{f(t) : f \text{ is convex and } f(u) \leq \tilde{\Lambda}(u) \text{ for all } u\}$ is the GCM of $\tilde{\Lambda}(t)$ (see Fig. 2(a)). In Appendix we give a direct proof of both (3) and the GCM representation using only some elementary properties of convex sets. In practice, $\hat{\lambda}(t)$ is computed using the following algorithm:

- Set $i_0 = 0$;
- Repeat until $i_{E+1} = n + 1$:
  Set i_{h+1} to be the index which minimizes the slopes between
  $(t_{i_h}, i_h - 1)$ and $(t_i, i - 1)$ $(i = i_h + 1, \ldots, n + 1)$;
- The constrained NPMLE is then given by $\hat{\lambda}(t) = (i_{j+1} - i_j)/(t_{i_{j+1}} - t_{i_j})$ whenever $t_{i_j} < t \leq t_{i_{j+1}}$. $\left. \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \right\}$ (4)

Suppose now that we observe K independent realizations of the same NHPP truncated at possibly different times $T_1, \ldots, T_K$. Let the ith realization be $N_i(t) = \sum_{j=1}^{n_i} I(t \geq t_{ij})$, where t_{ij} is the time of the jth failure of the ith system. The log-likelihood is now

$$\ell(\lambda) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log \lambda(t_{ij}) - \sum_{i=1}^{K} \int_0^{T_i} \lambda(u) \, du. \tag{5}$$

Define the TTT transformation R by $R(t) = \sum_{i=1}^{K} \min(t, T_i)$, and its generalized inverse by $R^{-1}(s) = \inf\{t : R(t) \geq s\}$. Gilardoni and Colosimo (2011) showed that (5) is equivalent to

$$\ell(\lambda) = c + \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log \lambda_S(s_{ij}) - \int_0^S \lambda_S(s) \, ds, \tag{6}$$
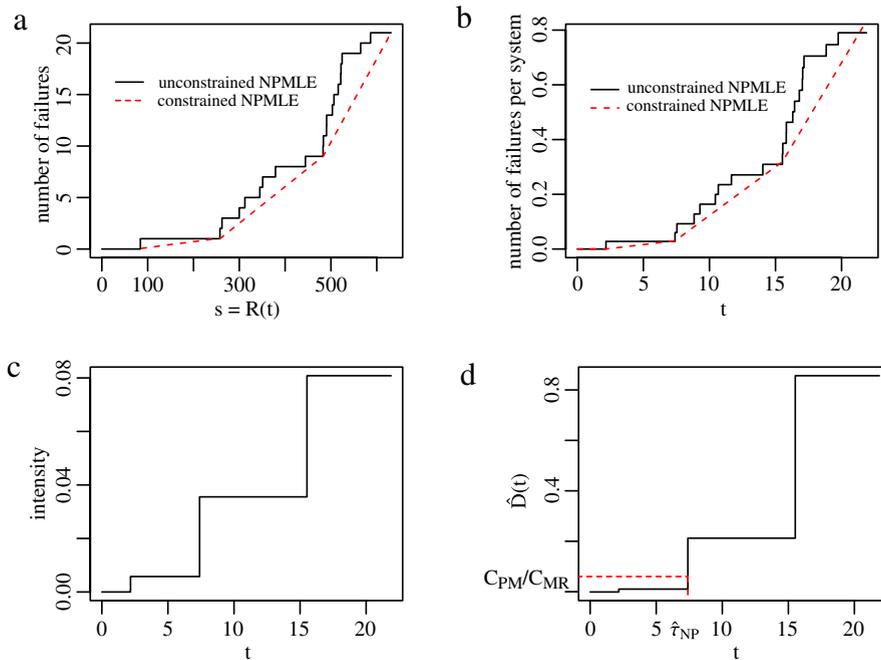
Fig. 2. NPMLEs of (a) the mean function $\Lambda_S(s)$ for the single realization of the superimposed process in the TTT scale; (b) the mean function $\Lambda(t)$ of the original process; (c) the intensity $\lambda(t)$ in the original time scale and (d) $D(t)$ (cf. Eq. (8)) and optimal PM time.

where $s_{ij} = R(t_{ij})$, $\lambda_S(s) = \lambda[R^{-1}(s)]$ and $S = R(\max\{T_1, \ldots, T_K\}) = \sum_{i=1}^{K} T_i$. Noting that $\lambda(t)$ is increasing if and only if $\lambda_S(s)$ is too, and comparing (2) with (6), one concludes that the computation of the constrained NPMLE in the many realizations setup can be reduced to the single realization setup discussed before. Essentially, one defines the superimposed process $N_S(s) = \sum_{i=1}^{K} N_i[R^{-1}(s)]$ and its intensity $\lambda_S(s) = \lambda[R^{-1}(s)]$, uses the algorithm (4) to compute the constrained NPMLE $\hat{\lambda}_S(s)$, and then goes back to the original time scale by letting $\hat{\lambda}(t) = \hat{\lambda}_S[R(t)]$. This process is illustrated in Fig. 2(a)–(c) for the power transformers data set.

3. Optimal maintenance time

As before, let $N(t)$ be the number of failures up to time t and suppose that under MR, $N(t)$ is an NHPP with intensity $\lambda(t)$ and mean function $\Lambda(t) = \int_0^t \lambda(u)\,du$. Since preventive maintenance can only be justified by the fact of system degradation over time, we also assume that $\lambda(t)$ is increasing and hence that $\Lambda(t)$ is convex. Suppose now that the system will be in operation for T units of time, with T large, and assume that it is decided to perform PM every τ units of time. In order to compute the expected cost $C_{(0,T]}(\tau)$ of this maintenance strategy, write $(0, T] = (0, \tau] \cup (\tau, 2\tau] \cup \cdots \cup ((m-1)\tau, m\tau] \cup (m\tau, T]$, where $m = [T/\tau]$ is the largest integer smaller than or equal to T/τ, and let C_{PM} and C_{MR} be the fixed costs of performing a PM, respectively, an MR action. Now, since a PM performed at $t = k\tau$ renews the process, one expects to have $EN(\tau) = \Lambda(\tau)$ failures in each interval $(k\tau, (k+1)\tau]$. Hence, the expected cost of operating the system in that interval is $C_{MR}\,\Lambda(\tau) + C_{PM}$ and the total expected cost over $(0, T]$ is

$$C_{(0,T]}(\tau) = m\{C_{MR}\,\Lambda(\tau) + C_{PM}\} + C_{MR}\,\Lambda(T - m\tau),$$

where the last term is the cost associated with the failures in the interval $(m\tau, T]$. Dividing by T and taking limits as $T \to \infty$ we get the expected cost per unit of time over an infinite horizon

$$H(\tau) = \lim_{T \to \infty} \frac{C_{(0,T]}(\tau)}{T} = \frac{C_{MR}\,\Lambda(\tau) + C_{PM}}{\tau}. \tag{7}$$

Therefore, provided that $\lambda(t)$ is continuous, the optimal PM periodicity is obtained by solving

$$D(\tau) = \tau\lambda(\tau) - \Lambda(\tau) = \frac{\tau^2}{C_{MR}} H'(\tau) + \frac{C_{PM}}{C_{MR}} = \frac{C_{PM}}{C_{MR}}. \tag{8}$$

More generally, if $\lambda(t)$ is a step function, as will be the case later on, the optimal periodicity τ_{OPT} is any time such that $D(\tau) < C_{PM}/C_{MR}$ for $\tau < \tau_{OPT}$ and $D(\tau) > C_{PM}/C_{MR}$ for $\tau > \tau_{OPT}$.

Once we have estimates $\hat{\lambda}(t)$ and $\hat{\Lambda}(t) = \int_0^t \hat{\lambda}(u)\, du$, the natural estimate of the optimal PM periodicity is obtained by minimizing $\hat{H}(\tau) = [C_{MR}\,\hat{\Lambda}(\tau) + C_{PM}]/\tau$. Since parametric estimates of the intensity function will typically be continuous, the optimal τ will be in this case the solution of $\hat{D}(\tau) = \tau\hat{\lambda}(\tau) - \hat{\Lambda}(\tau) = C_{PM}/C_{MR}$. On the other hand, as mentioned above, some care is needed for the constrained NPMLE discussed in Section 2 because $\hat{D}(\tau)$ will be discontinuous. More precisely, using the notation in (4), suppose that the constrained NPMLE of $\lambda(t)$ in the original time scale is given by $\hat{\lambda}(\tau) = (i_{j+1} - i_j)/(t_{i_{j+1}} - t_{i_j})$ for $t_{i_j} < \tau < t_{i_{j+1}}$, where $0 = i_0 < i_1 < \cdots < i_E < i_{E+1} = n + 1$ are the indices of the extreme points of the GCM. Then it is easy to check that $\hat{\Lambda}(\tau) = i_j + (\tau - t_{i_j})(i_{j+1} - i_j)/(t_{i_{j+1}} - t_{i_j})$ whenever $t_{i_j} < \tau < t_{i_{j+1}}$. Hence, $\hat{D}(\tau) = \tau\hat{\lambda}(\tau) - \hat{\Lambda}(\tau) = t_{i_j}(i_{j+1} - i_j)/(t_{i_{j+1}} - t_{i_j}) - i_j$ is also a step function. It follows that in order to minimize $\hat{H}(\tau)$, we search for the first t_{i_j} such that $\hat{D}(t_{i_h}) < C_{PM}/C_{MR}$ for $h < j$ and $\hat{D}(t_{i_h}) > C_{PM}/C_{MR}$ for $h \geq j$ (see Fig. 2(d)). If no such t_{i_j} exists, *i.e.*, if $\hat{D}(t_{E+1}) \leq C_{PM}/C_{MR}$, all we can conclude is that the optimal periodicity is either greater than the maximum truncation time or is infinite. We also note here that in theory the optimal periodicity may not be unique if there is a t_{i_j} for which $\hat{D}(t_{i_j})$ is exactly equal to C_{PM}/C_{MR}, but this event has probability zero under the NHPP model.

4. Bootstrap methods

We consider five different strategies to bootstrap the estimates of the optimal periodicity. The first one, denoted below by **boot.sys**, bootstraps the systems or realizations in the original time scale as if they were clustered data (cf. Field and Welsh, 2007). More precisely, suppose that the data consists of failure times t_{ij} and truncation times T_i ($1 \leq j \leq n_i$, $1 \leq i \leq K$). We first sample with replacement K integers between 1 and K, say $i_1^*, \ldots, i_K^*$. Based on the set of (possibly repeated) failure times $t_{i_h^*j}$ and truncations $T_{i_h^*}$, we compute the TTT transform and the estimates $\hat{\lambda}_S^*(s)$ and $\hat{\lambda}^*(t)$ of the intensities in the TTT and original time scale and, finally, we use $\hat{\lambda}^*(t)$ to compute a bootstrap estimate $\hat{\tau}^*$ of the optimal periodicity. Here, as well as in the rest of this section, the expressions in Section 2 must be adjusted to account for the possibility of repeated failure times. For instance, the numerator in (3) should be $f_k - f_h$, where f_i is the number of failures up to and including t_i.

All other strategies resample failure times from the superimposed process in the TTT time scale. The second one, denoted by **boot+n+t**, generates TTT failure times $s_1^*, \ldots, s_{n^*}^*$ by simulating an NHPP whose intensity function is the constrained NPMLE $\hat{\lambda}_S(s)$ computed with the original data set. Recall that to generate an NHPP with intensity $\hat{\lambda}_S(s)$ and mean function $\hat{\Lambda}_S(s)$ truncated in $(0, S)$, one first generates $n^* \sim \text{Poisson}(\hat{\Lambda}_S(S))$ and then obtains the TTT failure times as the order statistics of a size n^* i.i.d. random sample from the cdf $\hat{\Lambda}_S(s)/\hat{\Lambda}_S(S)$, $0 < s < S$, where $S = \sum_{i=1}^K T_i$ is the truncation time of the superimposed process. Once these failure times are obtained, we use (3) or (4) to compute $\hat{\lambda}_S^*(s)$, then the inverse TTT transform to get $\hat{\lambda}^*(t)$ and finally the procedure described at the end of Section 3 to get a resample $\hat{\tau}^*$ of the estimate of the optimal periodicity.

The last three strategies are variants of the one just described. More precisely, the third one, **boot−n+t**, proceeds exactly as does **boot+n+t** except for the fact that n^* is kept fixed and equal to n, the number of failures in the original data set. The fourth one, **boot+n−t**, differs from **boot+n+t** only in that, instead of resampling the TTT failure times $s_1^*, \ldots, s_{n^*}^*$ from the cdf $\hat{\Lambda}(s)/\hat{\Lambda}(S)$, they are the order statistics of a size n^* random sample with replacement from the set of original TTT failure times $\{s_{ij} : 1 \leq i \leq K,\ 1 \leq j \leq n_i\}$. Finally, the last strategy **boot−n−t** both keeps $n^* = n$ fixed and resamples from the original TTT failure times. We note that in a somewhat different context, namely that of bootstrapping a kernel estimate of $\lambda(t)$, Cowling et al. (1996) consider resampling procedures which are analogous to the strategies **boot+n+t**, **boot−n+t** and **boot−n−t**.

The strategies **boot+n+t** and **boot−n+t** require to sample from the cdf $\hat{\Lambda}_S(s)/\hat{\Lambda}_S(S)$ ($0 < s < S$). This can be done easily: since $\hat{\lambda}_S(s)$ is a step function, the corresponding pdf $\hat{\lambda}_S(s)/\hat{\Lambda}_S(S)$ can be written as a mixture of uniform distributions.

5. Monte Carlo simulation

In this section we describe some Monte Carlo simulations in order to compare the five resampling methods. Failure times were generated for $36 = 3 \times 2 \times 2 \times 3$ different scenarios. These were defined considering (i) three parametric models, (ii) either $K = 40$ or $K = 80$ systems, (iii) either equal or unequal truncation times, and (iv) truncation times giving on average roughly 0.5, 1.0 or 2.0 expected failures per system.

For the parametric models, we used a PLP intensity $\lambda(t) = (\beta/\theta)\,(t/\theta)^{\beta-1}$, a log-linear intensity $\lambda(t) = \exp\{\alpha + \beta t\}$, and an increasing bounded intensity

$$\lambda(t) = \alpha \left[1 - \frac{1}{\sqrt{1 + t/\beta}} \right]. \tag{9}$$

The parameter values were set at $\beta = 2$ and $\theta = 24$ for the PLP, $\alpha = -4.5$ and $\beta = 0.1$ for the log-linear, and $\alpha = 0.6$ and $\beta = 80$ for the bounded intensity model. These are roughly the MLEs obtained by fitting each of these models to the power transformers data set. In all cases we used a cost ratio $C_{PM}/C_{MR} = 1/16$, giving actual optimal periodicities of preventive

Table 1
Simulation results for PLP intensity $\lambda(t) = \beta t^\beta / \theta^\beta$, $\beta = 2$, $\theta = 24$.

Point Est.		Equal truncation times						Unequal truncation times					
		K = 40			K = 80			K = 40			K = 80		
	T $E(N) \approx$	17 20	24 40	34 80	17 40	24 80	34 160	T_{tr} 20	$1.36\,T_{\mathrm{tr}}$ 40	$1.92\,T_{\mathrm{tr}}$ 80	T_{tr}^* 40	$1.36\,T_{\mathrm{tr}}^*$ 80	$1.92\,T_{\mathrm{tr}}^*$ 160
$\hat\tau_{\mathrm{PLP}}$	MRE	0.049	0.023	0.014	0.025	0.013	0.005	0.043	0.028	0.011	0.022	0.016	0.006
	IQR.RE	0.195	0.141	0.137	0.121	0.099	0.093	0.189	0.155	0.145	0.121	0.109	0.099
$\hat\tau_{\mathrm{NP}}$	MRE	0.052	0.051	0.059	0.024	0.050	0.025	0.085	0.091	0.078	0.042	0.044	0.034
	IQR.RE	0.535	0.531	0.531	0.437	0.434	0.436	0.582	0.568	0.573	0.469	0.464	0.466
90% CI													
Parametric-PLP	MRL	0.430	0.352	0.331	0.285	0.239	0.231	0.444	0.382	0.355	0.295	0.263	0.249
	$1 - \hat\alpha$	0.911	0.896	0.884	0.917	0.896	0.896	0.908	0.896	0.892	0.895	0.893	0.906
Boot+n+t	MRL	1.059	1.058	1.056	0.905	0.914	0.897	1.197	1.161	1.133	1.006	0.964	0.956
	$1 - \hat\alpha$	0.897	0.882	0.881	0.973	0.969	0.974	0.848	0.849	0.868	0.959	0.961	0.964
Boot−n+t	MRL	0.969	1.013	1.040	0.855	0.887	0.885	1.103	1.111	1.109	0.960	0.940	0.944
	$1 - \hat\alpha$	0.893	0.880	0.883	0.973	0.969	0.972	0.844	0.849	0.868	0.957	0.960	0.963
Boot+n−t	MRL	1.001	0.987	0.980	0.870	0.878	0.860	1.118	1.084	1.048	0.958	0.925	0.920
	$1 - \hat\alpha$	0.787	0.773	0.775	0.880	0.859	0.860	0.740	0.733	0.771	0.852	0.845	0.856
Boot−n−t	MRL	0.887	0.928	0.956	0.822	0.855	0.842	1.003	1.030	1.022	0.907	0.899	0.896
	$1 - \hat\alpha$	0.765	0.767	0.766	0.867	0.853	0.857	0.722	0.724	0.765	0.842	0.841	0.854
Boot.sys	MRL	0.974	0.977	0.965	0.863	0.867	0.850	1.107	1.076	1.032	0.950	0.916	0.898
	$1 - \hat\alpha$	0.781	0.768	0.766	0.875	0.854	0.857	0.736	0.729	0.763	0.850	0.843	0.852

Results based on $M = 3000$ simulations. In all cases the bootstrap sample size is $B = 1000$. MRE, IQR.RE, MRL and $(1 - \hat\alpha)$ stand respectively for the median and interquartile range of the relative error, the median length of the 90% CI (i.e. length divided by actual τ) and the observed coverage. T_{tr} are the truncation times of the power transformers data set. T_{tr}^* is T_{tr} replicated twice.

maintenance equal to $\tau = 6.00$, 8.04 and 5.99 for, respectively, the PLP, loglinear and bounded intensities. We note that although other bounded intensities have been considered previously in the literature, e.g., $\lambda(t) = \alpha[1 - \exp(-t/\beta)]$ by Pulcini (2001), they are difficult to simulate from. This is because, as mentioned above, simulating an *NHPP* truncated in $(0, T)$ requires sampling from the cdf $\Lambda(t)/\Lambda(T)$ $(0 < s < S)$. This is typically achieved by sampling a Uniform(0, 1) random variate u and solving the equation $\Lambda(t)/\Lambda(T) = u$, which does not have an explicit solution for Pulcini's intensity. On the other hand, in the case of the bounded intensity (9), the mean number of failures is $\Lambda(t) = \alpha[t - 2\beta\sqrt{1 + t/\beta}]$ and hence the equation $\Lambda(t)/\Lambda(T) = u$ has for $0 < t < T$ the unique solution $t = u\Lambda(T)/\alpha + 2\beta + 2\sqrt{\beta}\sqrt{u\Lambda(T)/\alpha + 2\beta}$.

For each of the three parametric models and $K = 40$ systems, we considered three scenarios each for the two cases of equal and unequal truncation times. Equal truncation times were set to be $T = 17$, 24, and 34 for the PLP, $T = 17$, 24, and 30 for the loglinear, and $T = 17$, 24, and 36 for the bounded intensity function. As mentioned before, these values were set in such a way that the expected failures per system are roughly 0.5, 1.0, and 2.0. Unequal truncation times were set to be a multiple of the truncation times in the power transformers data set. The multiplication constant was set to be $c = 1.00$, 1.36, and 1.92 for the PLP, $c = 1.00$, 1.25, and 1.60 for the loglinear, and $c = 1.00$, 1.42, and 2.07 for the bounded intensity case. Again, these values were set to have on average roughly 0.5, 1.0, and 2.0 expected failures per system. Finally, we also considered scenarios with $K = 80$ systems by replicating twice the truncation times considered before when $K = 40$.

The failure times for each of the 36 scenarios were simulated $M = 3000$ times. For each simulation, the estimated optimal periodicity was resampled $B = 1000$ for each of the five strategies discussed in the previous section. Bootstrap 90% CIs were defined by the 5% and 95% quantiles of the bootstrap sample. Then, for each scenario and each strategy, we computed the observed coverage $1 - \hat\alpha$ and the median relative CI length (MRL, i.e., median CI length divided by the actual τ). We note that since $\hat\tau$ may be infinite with positive probability no matter the scenario considered, the average CI length actually estimates ∞. Hence, the median length is more appropriate than the average length as a measure of efficiency. For comparison, we also computed the parametric CI assuming a PLP process. To do this, we reparametrized the PLP intensity in terms of $\eta = \log\beta$ and $\nu = \log\theta$, computed MLEs $\hat\eta$ and $\hat\nu$ by maximizing numerically the PLP likelihood and obtained the estimate of the optimal periodicity as

$$\hat\tau_{\mathrm{PLP}} = \hat\theta\left[\frac{C_{\mathrm{PM}}}{(\hat\beta - 1)C_{\mathrm{MR}}}\right]^{1/\hat\beta} = e^{\hat\nu}\left[\frac{C_{\mathrm{PM}}}{(e^{\hat\eta} - 1)C_{\mathrm{MR}}}\right]^{e^{-\hat\eta}} \tag{10}$$

(cf. Gilardoni and Colosimo, 2007). Finally, in order to build CIs, the asymptotic variance of $\hat\tau_{\mathrm{PLP}}$ was obtained by using the Delta Method and the (η, ν) observed information matrix.

Tables 1–3 show the results of the Monte Carlo study. MRE, IQR.RE, MRL, and $(1 - \hat\alpha)$ stand, respectively, for the median and interquartile range of the relative error, the median relative length of the 90% CI (i.e., length divided by actual τ), and the observed coverage. T_{tr} are the truncation times of the power transformers data set. T_{tr}^* is T_{tr} replicated twice. For completeness, we also included in these tables information about the parametric estimate $\hat\tau_{\mathrm{PLP}}$ and the nonparametric estimate of τ. In this case, for each point estimate we report the median and the interquartile range of the relative estimation error (i.e., of the difference between the estimate and the actual τ divided by τ).

Table 2

Simulation results for loglinear intensity $\lambda(t) = \exp\{\alpha + \beta t\}$, $\alpha = -4.5$, $\beta = 0.1$.

Point Est.		Equal truncation times						Unequal truncation times					
		$K = 40$			$K = 80$			$K = 40$			$K = 80$		
	T	17	24	30	17	24	30	T_{tr}	$1.25\,T_{\mathrm{tr}}$	$1.60\,T_{\mathrm{tr}}$	T_{tr}^*	$1.25\,T_{\mathrm{tr}}^*$	$1.60\,T_{\mathrm{tr}}^*$
	$E(N) \approx$	20	40	80	40	80	160	20	40	80	40	80	160
$\hat{\tau}_{\mathrm{PLP}}$	MRE	−0.163	−0.272	−0.272	−0.186	−0.288	−0.274	−0.236	−0.290	−0.251	−0.259	−0.305	−0.247
	IQR.RE	0.276	0.111	0.132	0.182	0.072	0.095	0.157	0.128	0.149	0.103	0.085	0.104
$\hat{\tau}_{\mathrm{NP}}$	MRE	−0.028	−0.035	−0.023	−0.024	−0.036	−0.037	−0.008	−0.018	−0.014	−0.014	−0.011	−0.015
	IQR.RE	0.549	0.541	0.558	0.469	0.472	0.443	0.608	0.576	0.599	0.506	0.487	0.479
90% CI													
Parametric-PLP	MRL	0.519	0.251	0.255	0.370	0.166	0.180	0.337	0.271	0.276	0.223	0.183	0.196
	$1 - \hat{\alpha}$	0.650	0.167	0.159	0.524	0.018	0.023	0.428	0.177	0.243	0.191	0.018	0.067
Boot+n+t	MRL	1.055	1.062	1.063	0.929	0.947	0.935	1.155	1.160	1.150	1.030	1.019	0.999
	$1 - \hat{\alpha}$	0.976	0.977	0.974	0.973	0.974	0.976	0.971	0.975	0.976	0.975	0.975	0.972
Boot−n+t	MRL	0.981	1.032	1.046	0.892	0.929	0.925	1.089	1.120	1.137	0.997	1.000	0.990
	$1 - \hat{\alpha}$	0.961	0.970	0.972	0.960	0.968	0.972	0.960	0.968	0.973	0.966	0.971	0.969
Boot+n−t	MRL	1.003	0.997	1.007	0.875	0.901	0.892	1.078	1.079	1.081	0.968	0.956	0.955
	$1 - \hat{\alpha}$	0.881	0.882	0.880	0.910	0.907	0.920	0.854	0.866	0.870	0.905	0.908	0.906
Boot−n−t	MRL	0.930	0.967	0.981	0.848	0.880	0.885	1.024	1.048	1.069	0.943	0.938	0.947
	$1 - \hat{\alpha}$	0.957	0.874	0.871	0.895	0.903	0.918	0.832	0.860	0.861	0.898	0.901	0.903
Boot.sys	MRL	0.987	0.986	0.993	0.870	0.894	0.884	1.071	1.068	1.068	0.962	0.959	0.947
	$1 - \hat{\alpha}$	0.876	0.877	0.871	0.910	0.902	0.916	0.852	0.864	0.859	0.900	0.900	0.898

Results based on $M = 3000$ simulations. In all cases the bootstrap sample size is $B = 1000$. MRE, IQR.RE, MRL and $(1 - \hat{\alpha})$ stand respectively for the median and interquartile range of the relative error, the median relative length of the 90% CI (i.e. length divided by actual τ) and the observed coverage. T_{tr} are the truncation times of the power transformers data set. T_{tr}^* is T_{tr} replicated twice.

Table 3

Simulation results for bounded intensity $\lambda(t) = \alpha[1 - (1 + t/\beta)^{-0.5}]$, $\alpha = 0.6$, $\beta = 80$.

Point Est.		Equal truncation times						Unequal truncation times					
		$K = 40$			$K = 80$			$K = 40$			$K = 80$		
	T	17	24	36	17	24	36	T_{tr}	$1.42\,T_{\mathrm{tr}}$	$2.07\,T_{\mathrm{tr}}$	T_{tr}^*	$1.42\,T_{\mathrm{tr}}^*$	$2.07\,T_{\mathrm{tr}}^*$
	$E(N) \approx$	20	40	80	40	80	160	20	40	80	40	80	160
$\hat{\tau}_{\mathrm{PLP}}$	MRE	0.062	0.048	0.035	0.041	0.034	0.025	0.068	0.050	0.027	0.044	0.035	0.025
	IQR.RE	0.197	0.157	0.124	0.130	0.102	0.091	0.196	0.154	0.130	0.127	0.104	0.095
$\hat{\tau}_{\mathrm{NP}}$	MRE	0.044	0.062	0.070	0.037	0.038	0.030	0.089	0.074	0.079	0.029	0.054	0.050
	IQR.RE	0.538	0.564	0.541	0.467	0.466	0.455	0.606	0.575	0.568	0.467	0.469	0.477
90% CI													
Parametric-PLP	MRL	0.445	0.364	0.319	0.296	0.243	0.221	0.466	0.380	0.336	0.306	0.258	0.238
	$1 - \hat{\alpha}$	0.921	0.885	0.884	0.923	0.885	0.885	0.902	0.896	0.904	0.894	0.880	0.894
Boot+n+t	MRL	1.073	1.103	1.089	0.928	0.938	0.925	1.234	1.196	1.155	1.036	1.000	0.997
	$1 - \hat{\alpha}$	0.895	0.901	0.880	0.975	0.971	0.972	0.856	0.873	0.881	0.969	0.968	0.972
Boot−n+t	MRL	0.974	1.052	1.064	0.878	0.911	0.911	1.135	1.139	1.135	0.986	0.976	0.984
	$1 - \hat{\alpha}$	0.895	0.897	0.878	0.970	0.969	0.971	0.853	0.870	0.881	0.966	0.967	0.970
Boot+n−t	MRL	1.027	1.045	1.010	0.896	0.901	0.893	1.153	1.119	1.086	1.000	0.969	0.952
	$1 - \hat{\alpha}$	0.792	0.788	0.765	0.876	0.867	0.868	0.749	0.759	0.761	0.858	0.860	0.859
Boot−n−t	MRL	0.918	0.982	0.988	0.837	0.875	0.871	1.041	1.048	1.049	0.939	0.943	0.935
	$1 - \hat{\alpha}$	0.770	0.777	0.760	0.861	0.859	0.864	0.730	0.748	0.756	0.848	0.854	0.858
Boot.sys	MRL	1.009	1.019	0.990	0.886	0.898	0.879	1.149	1.105	1.067	1.004	0.966	0.943
	$1 - \hat{\alpha}$	0.787	0.783	0.775	0.871	0.864	0.863	0.744	0.757	0.752	0.855	0.855	0.859

Results based on $M = 3000$ simulations. In all cases the bootstrap sample size is $B = 1000$. MRE, IQR.RE, MRL and $(1 - \hat{\alpha})$ stand respectively for the median and interquartile range of the relative error, the median relative length of the 90% CI (i.e. length divided by actual τ) and the observed coverage. T_{tr} are the truncation times of the power transformers data set. T_{tr}^* is T_{tr} replicated twice.

The observed coverage and MRL for the case of $K = 40$ equally truncated systems are shown in Fig. 3. Below we summarize the main findings. Of course, our conclusions are limited to the scenarios considered above.

- As should be expected, the parametric CI is the best one under the PLP model. Indeed, in all 12 PLP scenarios, the parametric CI gave observed coverages between 0.884 and 0.917, hence quite close to the 90% nominal coverage, while the corresponding CI lengths were roughly between one-half and one-fourth those of the best nonparametric CI. The results in Table 3 indicate that the parametric CI also performed well for the bounded intensity scenarios. For some choice of parameter values, almost all observed failures are located in a time period where the PLP and the bounded intensities are very similar. Our results in the next section suggest that this happens too for the power transformer application. However, the parametric CI performed badly for the loglinear model, sometimes with almost null observed coverages.
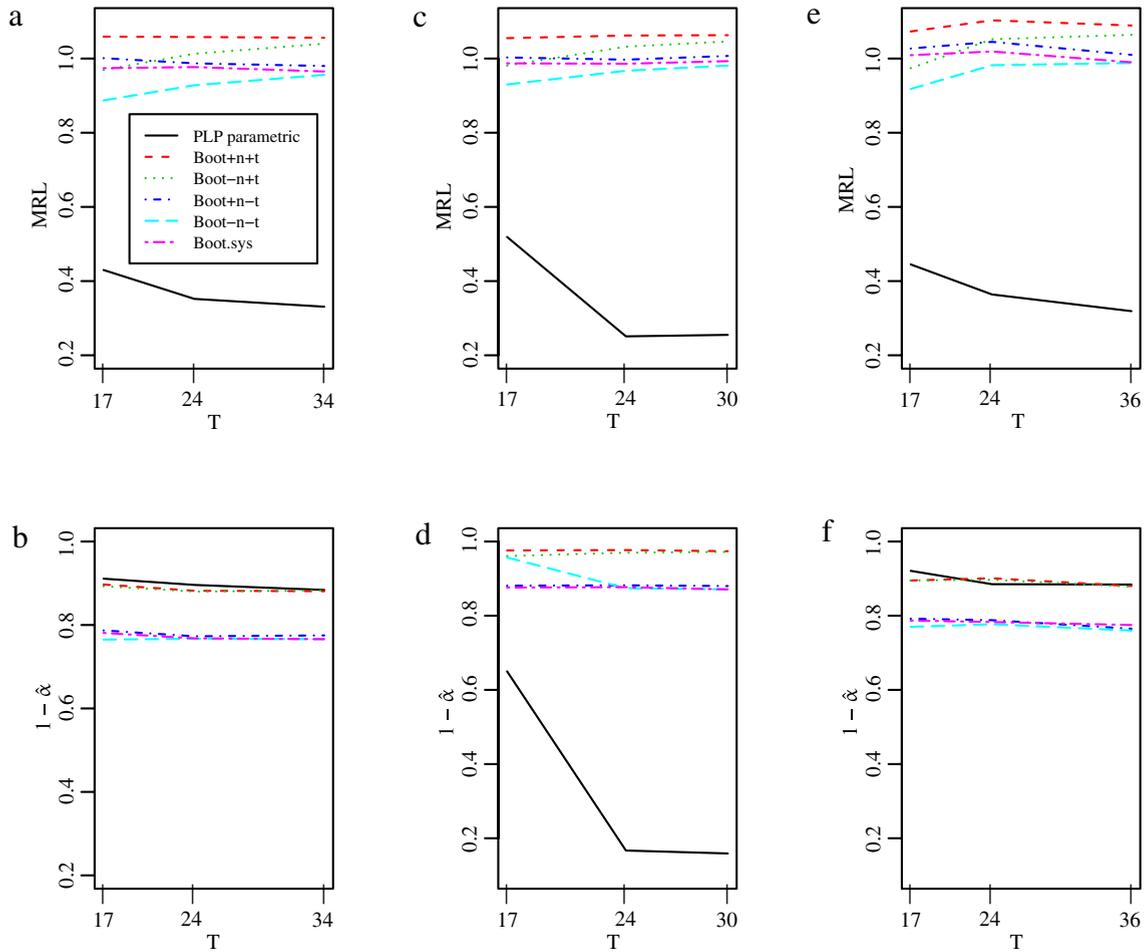
Fig. 3. Median relative length and descriptive confidence levels for $K = 40$ and equal truncation times: (a) and (b) are for the PLP, (c) and (d) for the log-linear and (e) and (f) for the bounded intensity model. Nominal coverage is 90%.

- Although resampling the systems in the original time scale (i.e., **boot.sys**) may be acceptable for large K, it tends to behave always worst than, for instance, **boot−n+t**.
- Increasing the truncation times has little effect, if any, on the performance of the nonparametric point and interval estimates. This is probably due to the facts that (i) nonparametric estimates behave locally and (ii) the estimation of τ depend on the values of $\hat{\lambda}(t)$ for $t < \tau$. Of course, this is true so long as the truncation times are greater than the optimal τ.
- Across the different simulation scenarios, the bootstrap strategies **boot+n+t** or **boot−n+t** perform at least as well, and in some cases much better, than any of the other three (see, for instance, the PLP case with $K = 40$ systems). In other words, resampling the failure times from the cdf $\hat{\Lambda}_S(s)/\hat{\Lambda}_S(S)$ is better than resampling from the actual times. Finally, among the two strategies that resample from the cdf $\hat{\Lambda}_S(s)/\hat{\Lambda}_S(S)$, **boot−n+t** seems to behave slightly better than **boot+n+t**.

6. Application

Here we consider the data set consisting of $n = 21$ failure times for $K = 40$ power transformers (see Fig. 1). The electric company operating the transformers provided the cost ratio $C_{PM}/C_{RM} = 1/15$ necessary to obtain τ in (1).

NPMLEs of the mean function $\Lambda(t)$ and the intensity $\lambda(t)$ are shown in Fig. 2(b) and (c). The corresponding estimates $\hat{D}(t)$ and $\hat{\tau}_{NP} = 7.396$ are shown in Fig. 2(d). We also adjusted the three parametric models discussed previously by numerically maximizing the corresponding likelihood functions. The MLEs were $\hat{\theta} = 24.366$, $\hat{\beta} = 1.995$, and $\hat{\tau}_{PLP} = 6.286$ for the PLP; $\hat{\alpha} = -4.505$, $\hat{\beta} = 0.094$, and $\hat{\tau}_{loglin} = 8.586$ for the loglinear; and $\hat{\alpha} = 0.561$, $\hat{\beta} = 73.138$, and $\hat{\tau}_{bounded} = 6.140$ for the bounded intensity model. Fig. 4(a) shows the unconstrained NPMLE estimate of $\Lambda(t)$ along with the three parametric estimates and suggests that the PLP and bounded intensity models adjust better to the data than the loglinear model. Fig. 4(b)
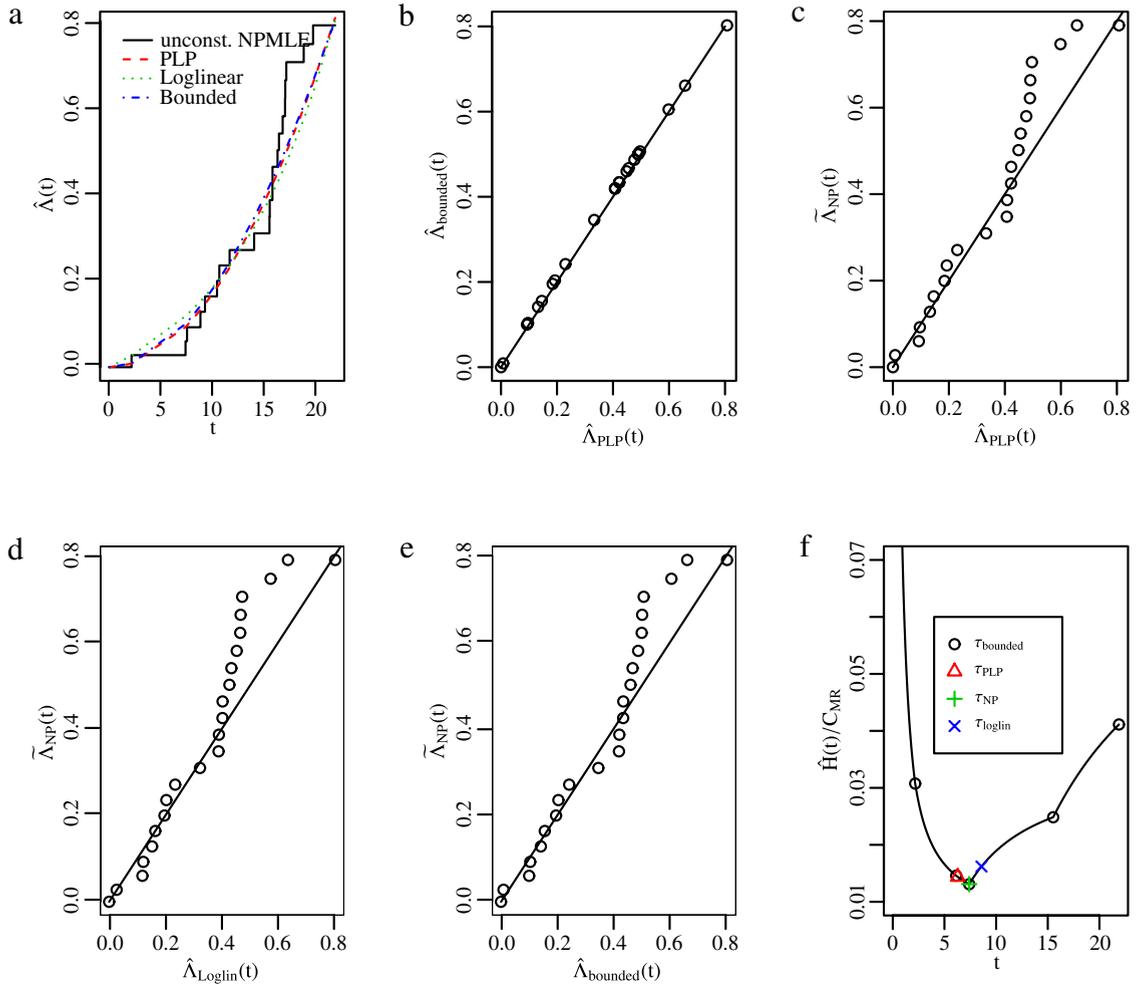
Fig. 4. Estimates of $\Lambda(t)$ for the observed failure times: (a): NPMLE, PLP, loglinear, and bounded model estimates against time, (b): PLP against bounded model estimates, (c)–(e): parametric against NPMLE estimates, (f): nonparametric estimate of $H(t)/C_{MR}$, the mean cost per unit of time measured in C_{MR} monetary units.

shows the parametric estimates of $\Lambda(t)$ for the observed failure times for the PLP and bounded intensity models. Since the points lie almost exactly on the line $y = x$, this suggests that for the range of values considered in this data set the two models are to a large extent indistinguishable. We note that this finding is consistent with the simulation results reported in the previous section.

The maximized log-likelihoods were $\hat{\ell}_{PLP} = -87.671$, $\hat{\ell}_{loglin} = -88.913$, and $\hat{\ell}_{bounded} = -87.639$. Since all three models have exactly two parameters, the Bayesian information criterion (BIC) to compare them differs from the log-likelihoods by the same constant. Hence, assuming equal prior probabilities for the three models and diffuse priors for the parameters, we can interpret the likelihoods as being approximately proportional to the posterior probabilities of the models. These would give probabilities 0.431, 0.124, and 0.445 for the PLP, loglinear and bounded models, respectively. Hence, even if the PLP and bounded intensity models adjust better, it is not possible to totally dismiss the loglinear model. Moreover, the diagnostic plots of the parametric estimates $\hat{\Lambda}_{PLP}(t_{ij})$, $\hat{\Lambda}_{bounded}(t_{ij})$ and $\hat{\Lambda}_{loglin}(t_{ij})$ against the unconstrained NPMLE estimate, shown in Fig. 4(c)–(e), leave some doubt about the quality of the three parametric fits.

The parametric 90% CI based on the Delta Method and (10) was $5.203 < \tau < 7.593$, which is based on an estimated error equal to 0.115 for $\log \hat{\tau}_{PLP}$. On the other hand, the 90% CI using the **boot−n+t** was $5.235 < \tau < 10.894$. We note that the difference between the two CIs seems to be almost exclusively along the upper tail of the sampling distributions, in the sense that the two lower limits are quite similar. Given the small number of failures and the previous observations about the fit of the three parametric models, we feel that the parametric CI may be misleading in the sense that it may fail to capture all the uncertainty in the data.

To conclude this section, we note that the constrained NPMLE of $\Lambda(t)$ can be plugged into (7) to obtain a nonparametric estimate of the expected cost per unit of time. Note that since $\hat{\Lambda}(t)$ is piecewise linear, the estimate $\hat{H}_{NP}(t)$ will be piecewise

hyperbolic (see Fig. 4(f)). For the estimates considered above, the estimated costs per unit of time (1000 h), measured in C_{MR} monetary units, are 0.0131 for $\hat{\tau}_{NP} = 7.396$, 0.0144 for $\hat{\tau}_{PLP} = 6.286$, 0.0162 for $\hat{\tau}_{loglin} = 8.586$, and 0.0146 for $\hat{\tau}_{bounded} = 6.140$. Even if the differences between the costs may seem small, one should consider here that it may have a large financial impact for a company that operates many transformers over a very long horizon.

7. Final remarks

This paper presented a bootstrap nonparametric approach to constructing CIs for functionals of the intensity of an NHPP subject to a monotonicity constraint. When several realizations of the NHPP are observed, our results suggest that one should first pool or superimpose the realizations using the TTT time transform and then bootstrap the one realization of this superimposed process. In a sense, this is consistent with the fact that, in the context of kernel estimation of the intensity function, there is substantial evidence that one should pool first and estimate second and not *vice versa* (see, for instance, Chiang et al., 2005 or Gilardoni and Colosimo, 2011).

Among the four different bootstrap strategies that resample the TTT superimposed failure times, we conclude that one should always resample the times from the cdf $\hat{\Lambda}_S(s)/\hat{\Lambda}_S(S)$ ($0 < s < S$), although whether resampling the total number of failures n^* from a Poisson distribution with mean $\hat{\Lambda}_S(S)$ or keeping it fixed and equal to the observed number of failures has little impact on the length and coverage of the CIs. Based on this, we found in Section 5 that the strategy **boot−n+t** was marginally better than **boot+n+t**, but any of those two performed much better than the other three bootstrap procedures studied. In Section 6, we found a nonparametric CI for the optimal maintenance time for a real data set, improving, in our view, on some parametric CIs that have been reported previously in the literature. We note that one of the advantages of the TTT transform is that the corresponding bootstrap CIs are quite fast to compute (about 13 s on R running under Linux on a 2.40 GHz Intel core i5 processor for the transformers data set). Finally, as a side note, in Section 5 we introduced a new bounded intensity model for the NHPP which, unlike other such intensities, is easy to simulate from.

Acknowledgments

Appendix. Proof of (3)

Here we prove that the constrained NPMLE of $\lambda(t)$ is given by (3). We begin by reviewing some basic facts about convex sets (cf. Ferguson, 1967 or Rao, 1967). The *convex envelope* $\mathcal{C}(A)$ of a set A is the set of all points x that are convex linear combinations (clc) of elements of A, i.e., all x such that there exist an integer r, $a_1, \ldots, a_r \in A$, and nonnegative weights $w_1, \ldots, w_r$ with $\sum_{i=1}^r w_i = 1$ such that $x = \sum_{i=1}^r w_i a_i$. An element of $\mathcal{C}(A)$ is said to be an *extreme point* if it cannot be written as a clc of any two different elements of $\mathcal{C}(A)$. The *boundary* of $\mathcal{C}(A)$ is the set formed by all extreme points and all those points which are the clc of at most two elements of $\mathcal{C}(A)$.

Given an increasing intensity $\lambda(t)$, define $\lambda^*(t) = \lambda(\max\{t_i : t_i \leq t\})$. Since $L(\lambda) \leq L(\lambda^*)$, this implies that the constrained NPMLE $\hat{\lambda}(t)$ ought to be a step function. It is also easy to prove that $\int_0^T \hat{\lambda}(u)\,du = n$ (see, for instance, Gilardoni and Colosimo, 2011, Proposition 3).

Let $A = \{(t, y) \in \mathbb{R}^2 : y \geq \tilde{\Lambda}(t)\}$ be the epigraph of $\tilde{\Lambda}(t)$ and denote by $(t_1, 0) = (t_{i_1}, i_1 - 1)$, $(t_{i_2}, i_2 - 1), \ldots$, $(t_{i_{E+1}}, i_{E+1} - 1) = (T, n)$ the extreme points of $\mathcal{C}(A)$, ordered so that $1 = i_0 < i_1 < \cdots < i_E < i_{E+1} = n + 1$. To prove that (3) gives the constrained NPMLE of $\lambda(t)$, note that with this notation (3) is equivalent to

$$\hat{\lambda}(t) = \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}} \tag{A.1}$$

(i.e. the slope of the straight line through $(t_{i_j}, i_j - 1)$ e $(t_{i_{j+1}}, i_{j+1} - 1)$) whenever $t_{i_j} \leq t < t_{i_{j+1}}$. Hence, we need to show that for any increasing step function $\lambda(t)$ we have that $\ell(\hat{\lambda}) \geq \ell(\lambda)$.

It follows from (A.1) and (2) that

$$\ell(\hat{\lambda}) = \sum_{i=1}^n \log \hat{\lambda}(t_i) - \int_0^{t_{n+1}} \hat{\lambda}(u)\,du$$

$$= \sum_{j=0}^E (i_{j+1} - i_j) \log \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}} - \sum_{j=0}^E (t_{i_{j+1}} - t_{i_j}) \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}}$$

$$= \sum_{j=0}^E (i_{j+1} - i_j) \left[\log \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}} - 1 \right]. \tag{A.2}$$

On the other hand, for any increasing step function $\lambda(t)$,

$$\ell(\lambda) = \sum_{i=1}^{n} [\log \lambda_i - (t_{i+1} - t_i)\lambda_i] = \sum_{j=0}^{E} \sum_{i=i_j}^{i_{j+1}-1} [\log \lambda_i - (t_{i+1} - t_i)\lambda_i], \tag{A.3}$$

where we have defined $\lambda_i = \lambda(t_i)$ $(0 \leq \lambda_1 \leq \cdots \leq \lambda_n)$.

To finish the proof we will show that every term in (A.2) is not less than the corresponding term in (A.3). To do this, note that since the indices i_j correspond to extreme points of $\mathcal{C}(A)$, for any t_i such that $t_{i_j} \leq t_i < t_{i_{j+1}}$ (equivalently, for any i such that $i_j \leq i < i_{j+1}$) we must have that the point $(t_i, i-1)$ lies above or over the straight line through the points $(t_{i_j}, i_j - 1)$ and $(t_{i_{j+1}}, i_{j+1} - 1)$. Otherwise, at least one of the points $(t_{i_j}, i_j - 1)$ or $(t_{i_{j+1}}, i_{j+1} - 1)$ would not be extreme. Hence, we conclude that

$$t_i - t_{i_j} \leq (t_{i_{j+1}} - t_{i_j}) \frac{i - i_j}{i_{j+1} - i_j}, \tag{A.4}$$

whenever $i_j \leq i < i_{j+1}$. Taking this observation into account, a generic term in (A.3) must satisfy

$$\sum_{i=i_j}^{i_{j+1}-1} [\log \lambda_i - (t_{i+1} - t_i)\lambda_i] = \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - (t_{i_{j+1}} - t_{i_j})\lambda_{i_{j+1}-1} + \sum_{i=i_j}^{i_{j+1}-1} (t_i - t_{i_j})(\lambda_i - \lambda_{i-1})$$

$$\leq \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - (t_{i_{j+1}} - t_{i_j}) \left[\lambda_{i_{j+1}-1} - \sum_{i=i_j}^{i_{j+1}-1} \frac{i - i_j}{i_{j+1} - i_j}(\lambda_i - \lambda_{i-1}) \right]$$

$$= \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - \frac{t_{i_{j+1}} - t_{i_j}}{i_{j+1} - i_j} \left[(i_{j+1} - i_j)\lambda_{i_{j+1}-1} - \sum_{i=i_j}^{i_{j+1}-1} (i - i_j)(\lambda_i - \lambda_{i-1}) \right]$$

$$= \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - \frac{t_{i_{j+1}} - t_{i_j}}{i_{j+1} - i_j} \left[(i_{j+1} - i_j)\lambda_{i_{j+1}-1} - \sum_{i=i_j}^{i_{j+1}-1} \left\{ (i - i_j)\lambda_i - (i - 1 - i_j)\lambda_{i-1} - \lambda_{i-1} \right\} \right]$$

$$= \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - \frac{t_{i_{j+1}} - t_{i_j}}{i_{j+1} - i_j} \left[(i_{j+1} - i_j)\lambda_{i_{j+1}-1} - (i_{j+1} - 1 - i_j)\lambda_{i_{j+1}-1} + \lambda_{i_j-1} + \sum_{i=i_j}^{i_{j+1}-1} \lambda_{i-1} \right]$$

$$= \sum_{i=i_j}^{i_{j+1}-1} \log \lambda_i - \frac{t_{i_{j+1}} - t_{i_j}}{i_{j+1} - i_j} \sum_{i=i_j}^{i_{j+1}-1} \lambda_i = \sum_{i=i_j}^{i_{j+1}-1} \left[\log \lambda_i - \frac{t_{i_{j+1}} - t_{i_j}}{i_{j+1} - i_j} \lambda_i \right]$$

$$\leq \sum_{i=i_j}^{i_{j+1}-1} \left[\log \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}} - 1 \right] = (i_{j+1} - i_j) \left[\log \frac{i_{j+1} - i_j}{t_{i_{j+1}} - t_{i_j}} - 1 \right],$$

as was to be shown. Here, the first inequality follows from (A.4) and the fact that $\lambda_i - \lambda_{i-1} \geq 0$, while the second inequality follows from the fact that $\log x - ax \leq \log(1/a) - 1$ whenever $x, a > 0$.

References

Aalen, O., 1978. Nonparametric inference for a family of counting processes. The Annals of Statistics 6 (4), 701–726.

Aalen, O.O., Borgan, Ø., Gjessing, H.K., 2008. Survival and Event History Analysis: A Process Point of View. Springer, Berlin.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression. Wiley, London.

Barlow, R., Hunter, L., 1960. Optimum preventive maintenance policies. Operations Research 8 (1), 90–100.

Bartoszynski, R., Brown, B.W., McBride, C.M., Thompson, J.R., 1981. Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. The Annals of Statistics 9 (5), 1050–1060.

Boswell, M.T., 1966. Estimating and testing trend in a stochastic process of Poisson type. The Annals of Mathematical Statistics 37 (6), 1564–1573.

Chiang, C.T., Wang, M.C., Huang, C.Y., 2005. Kernel estimation of rate function for recurrent event data. Scandinavian Journal of Statistics 32, 77–91.

Colosimo, E.A., Santos, W.B., Gilardoni, G.L., Motta, S.B., 2010. Optimal maintenance time for repairable systems under two types of failures. Communications in Statistics—Theory and Methods 39, 1289–1298.

Cowling, A., Hall, P., Phillips, M.J., 1996. Bootstrap confidence regions for the intensity of a Poisson point process. Journal of the American Statistical Association 91 (436), 1516–1524.

Crow, L.R., 1974. Reliability analysis for complex systems. In: Proschan, F., Serfling, J. (Eds.), Reliability and Biometry. pp. 379–410.

Ferguson, T.S., 1967. Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

Field, C.A., Welsh, A.H., 2007. Bootstrapping clustered data. Journal of the Royal Statistical Society. Series B 69, 369–390.

Gerstack, I.B., 1977. Models of Preventive Maintenance. North Holland, Amsterdam.

Gilardoni, G.L., Colosimo, E.A., 2007. Optimal maintenance time for repairable systems. Journal of Quality Technology 39 (1), 48–53.

Gilardoni, G.L., Colosimo, E.A., 2011. On the superposition of overlapping Poisson processes and nonparametric estimation of their intensity function. Journal of Statistical Planning and Inference 171, 3075–3083.

Hall, P., Huang, L.S., 2001. Nonparametric kernel regression subject to monotonicity constraints. Annals of Statistics 29 (3), 624–647.

Pan, R., Rigdon, S.E., 2009. Bayes inference for general repairable systems. Journal of Quality Technology 41 (1), 82–94.

Pulcini, G., 2001. A bounded intensity process for the reliability of repairable equipment. Journal of Quality Technology 33 (4), 480–492.

Ramlau-Hansen, H., 1983. Smoothing counting process intensities by means of kernel functions. Annals of Statistics 11, 453–466.

Rao, C.R., 1967. Linear Statistical Inference and its Applications. Wiley, New York.

Rigdon, S.E., Basu, A.P., 2000. Statistical Methods for the Reliability of Repairable Systems. Wiley, New York.

Ryan, K.J., Hamada, M.S., Reese, C.S., 2011. A Bayesian hierarchical PLP model for multiple repairable systems with application to supercomputer reliability. Journal of Quality Technology 43, 209–223.

Scholz, F.W., 1980. Towards a unified definition of maximum likelihood. The Canadian Journal of Statistics 8 (2), 193–203.

Sloan, T., 2008. Simultaneous determination of production and maintenance schedules using in-line equipment condition and yield information. Naval Research Logistics 55 (2), 116–129.

Tsai, W.Y., 1988. Estimation of the survival function with increasing failure rate based on left truncated and right censored data. Biometrika 75 (2), 319–324.

Tsai, T.R., Liu, P.H., Lio, Y.L., 2011. Optimal maintenance time for imperfect maintenance actions on repairable product. Computers & Industrial Engineering 60, 744–749.

Yu, J.W., Tian, G.L., Tang, M.L., 2007. Predictive analyses for nonhomogeneuous Poisson process with power law using Bayesian approach. Computational Statistics & Data Analysis 51.

Yu, J.W., Tian, G.L., Tang, M.L., 2008. Statistical inference and prediction for the Weibull process with incomplete observations. Computational Statistics & Data Analysis 52.