A nonparametric monotone maximum likelihood estimator of time trend for repairable systems data

Knut Heggland [1] and Bo H Lindqvist *

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract

The trend-renewal-process (TRP)is defined to be a time-transformed renewal process, where the time transformation is given by a trend function $\lambda(\cdot)$ which is similar to the intensity of a nonhomogeneous Poisson process (NHPP). A nonparametric maximum likelihood estimator of the trend function of a TRP is obtained under the often natural condition that $\lambda(\cdot)$ is monotone. An algorithm for computing the estimate is suggested and examined in detail in the case where the renewal distribution of the TRP is is a Weibull distribution. In the case where one has data from several systems, another monotone estimator of $\lambda(\cdot)$ is suggested, based on the assumption that the superposition of several TRP's can be approximated by an NHPP.

Key words: Repairable system, Trend-renewal process, Nonparametric estimation, Isotonic regression, Bootstrapping

1 Introduction

Failures of a repairable system are usually modeled by a stochastic point process in time. The most common models are the renewal process (RP), the homogeneous Poisson process (HPP), and the nonhomogeneous Poisson process (NHPP).

* Corresponding author.
Email addresses: knut.heggland@adsl.no (Knut Heggland), bo@math.ntnu.no (Bo H Lindqvist).
[1] Present address: Norwegian Computing Center, N-0314 Oslo, Norway.

As is well known, the RP model assumes what is called "perfect repair", indicating that after each failure, the system is renewed to its original condition. Of course, one would usually expect the system to improve after a repair. In fact, we may have repaired or exchanged the failed part and possibly overhauled some damaged, but not yet failed parts. There are also situations where repairing the system may cause detoriation, for example if some of the other parts were damaged in the reparation. Yet, is a complete renewal reasonable? A commonly used definition of a repairable system is that of "a system which, after failing to perform one or more of its functions satisfactorily, can be restored to fully satisfactory performance by any method, other than replacement of the entire system" (Ascher and Feingold [3]). It seems as if only a complete replacement could fully justify the renewal assumption, though.

The NHPP model, on the other hand, assumes what is called "minimal repair". After each failure and following repair, the system is in the same state as it was just prior to that failure. This is more plausible than the complete renewal assumption. One could argue that the part which has failed and is repaired or replaced is just a small part of the system, and the other parts are not affected. Yet often the replaced part is not a minor part, or the repair may affect some other parts of the system, to the better or the worse. This could mean a small jump in the intensity, in either direction.

There is thus a need for models which allow the system to deteriorate (or improve) over time, yet still allow for the possibility that the system could have a drastic increase or decrease in its failure intensity just after a repair, because of damage done, or weak spots removed. Several models have been developed for this purpose. A review is given, for example, by Pham and Wang [9].

More recently, the trend-renewal process (TRP) was introduced by Lindqvist, Elvebakk and Heggland [7]. This model contains the RP and NHPP as special cases, and in a simple manner the TRP fills the gap between the two extreme repair models. While parametric estimation for TRP's was considered in [7], the purpose of the present paper is to consider nonparametric estimation in the TRP model under the assumption of a monotonic trend in the occurrence of failures of the system. This is often a reasonable assumption in reliability applications, where, for example, a mechanical system is deteriorating due to ageing, while a software system is improving due to fault correction.

The paper is organized as follows. In Section 2 we give some basic definitions and results, including the likelihood function of TRP data. Section 3 is the main section, in which we derive the nonparametric maximum likelihood estimators. The properties of the estimators are studied via two examples in Section 4. In Section 5 we sketch the derivation of a nonparametric estimator for the case where more than one system is observed. Finally, some concluding remarks are given in Section 6.

2 Definitions and preliminaries

Consider a repairable system, observed from time $t = 0$. Let $N(t)$ be the number of failures in $(0,t]$, let T_i be the time of the i'th failure, where we define $T_0 = 0$, and let X_i be the time between failure number $i-1$ and failure number i, that is $X_i = T_i - T_{i-1}$. We assume that all repair times equal 0. This assumption is reasonable if the repair times are negligible compared to the times between failures, or if we let the time parameter be the operation time of the system. For a general treatment of repairable systems, see Ascher and Feingold [3] or Meeker and Escobar [8].

We next review the definitions of the most common point process models that are used for repairable systems, and then we define the trend-renewal process which will be the main model used in this paper.

2.1 Models for repairable systems

The homogeneous Poisson process, HPP(λ):

The process $N(t)$ is an HPP(λ) if $X_1, X_2, \ldots$ are independent and exponentially distributed with parameter λ.

The renewal process, RP(F):

The process $N(t)$ is an RP(F) if $X_1, X_2, \ldots$ are independent and identically distributed with cumulative distribution function (cdf) F, where we assume $F(0) = 0$. If F is the exponential distribution with parameter λ, then RP(F)=HPP(λ).

The nonhomogeneous Poisson process, NHPP($\lambda(\cdot)$):

Let $\lambda(t),\ t \geq 0$ be a nonnegative function, called the intensity of the process. The cumulative intensity function is then $\Lambda(t) = \int_0^t \lambda(u)du$. The process $N(t)$ is an NHPP($\lambda(\cdot)$) if the time-transformed process $\Lambda(T_1), \Lambda(T_2), \ldots$ is an HPP(1). Note that if $\lambda(t) \equiv \lambda$, then NHPP($\lambda(\cdot)$)=HPP($\lambda$).

We are now ready to define the TRP. This process is a time-transformed RP in the same way as the NHPP is a time-transformed HPP. We will use the definition given in Lindqvist et al. [7]:

The trend-renewal process, TRP($F, \lambda(\cdot)$):

Let $\lambda(t)$ be a nonnegative function on $t \geq 0$, and let $\Lambda(t) = \int_0^t \lambda(u)du$. The process $N(t)$ is a TRP($F, \lambda(\cdot)$) if the time-transformed process $\Lambda(T_1), \Lambda(T_2), \ldots$ is an RP(F), that is if $\Lambda(T_1), \Lambda(T_2) - \Lambda(T_1), \Lambda(T_3) - \Lambda(T_2), \ldots$ are independent and identically distributed with cdf F. The distribution F is called the renewal distribution, and $\lambda(\cdot)$ is called the trend function of the TRP.

Note that the representation TRP($F, \lambda(\cdot)$) is not unique, since a scale factor can be moved from F to $\lambda(\cdot)$ (or back) in the following way: Let $c > 0$ be a constant, and let $F_c(x) = F(cx)$ and $\lambda_c(t) = c^{-1}\lambda(t)$ for all x and t. Then TRP($F, \lambda(\cdot)$) = TRP($F_c, \lambda_c(\cdot)$).

It is easily seen that the TRP generalizes both the NHPP and the RP, since TRP($1 - e^{-x}, \lambda(\cdot)$) = NHPP($\lambda(\cdot)$) and TRP($F$, 1) = RP($F$).

2.2 The likelihood function for the TRP model

The conditional intensity function of a point process (Andersen et al. [1]) is defined by

$$\gamma(t) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(\text{failure in } [t, t+\Delta t) | \mathcal{F}_{t-})}{\Delta t},$$

where $\mathcal{F}_{t-}$ is the history of the process $N(t)$ up to, but not including time t. The conditional intensity function will, in general, be stochastic. For an NHPP($\lambda(\cdot)$) it is, however, nonstochastic, $\gamma(t) = \lambda(t)$. In the case of an RP(F) it is stochastic and given by $\gamma(t) = z(t - T_{N(t-)})$, where $z(\cdot)$ is the hazard rate corresponding to F, $z(t) = \frac{d}{dt}F(t)/(1 - F(t))$.

We obtain the conditional intensity function of a TRP($F, \lambda(\cdot)$) as follows:

$$\begin{aligned}
\gamma(t) &= \lim_{\Delta t \to 0} \frac{\mathrm{P}(\text{failure in } [t, t+\Delta t) | \mathcal{F}_{t-})}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathrm{P}(\text{failure in RP}(F) \text{ in } [\Lambda(t), \Lambda(t+\Delta t)) | \mathcal{F}_{t-})}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathrm{P}(\text{failure in RP}(F) \text{ in } [\Lambda(t), \Lambda(t) + \Delta\Lambda(t)) | \mathcal{F}_{t-})}{\Delta\Lambda(t)} \frac{\Delta\Lambda(t)}{\Delta t} \\
&= z(\Lambda(t) - \Lambda(T_{N(t-)})) \cdot \lim_{\Delta t \to 0} \frac{\Lambda(t+\Delta t) - \Lambda(t)}{\Delta t} \\
&= z(\Lambda(t) - \Lambda(T_{N(t-)}))\lambda(t). \qquad (1)
\end{aligned}$$

Consider now a point process $N(t)$, observed from time $t = 0$ to time $t = \tau$, with corresponding failure times $T_1, T_2, \ldots, T_{N(\tau)}$ and conditional intensity function $\gamma(t)$. The likelihood function of the process is then given by (Andersen et al. [1])

$$L = \{\prod_{i=1}^{N(\tau)} \gamma(T_i)\} e^{-\int_0^\tau \gamma(u)du}. \quad (2)$$

The likelihood function of a TRP is obtained by substituting (1) into (2), giving

$$L = \{\prod_{i=1}^{N(\tau)} z(\Lambda(T_i) - \Lambda(T_{i-1}))\lambda(T_i)\} e^{-\sum_{i=1}^{N(\tau)} \int_{T_{i-1}}^{T_i} z(\Lambda(u)-\Lambda(T_{i-1}))\lambda(u)du} \cdot e^{-\int_{T_{N(\tau)}}^{\tau} z(\Lambda(u)-\Lambda(T_{N(\tau)}))\lambda(u)du}.$$

By making the substitution $v = \Lambda(u) - \Lambda(T_{i-1})$ and taking the log we get the log likelihood function

$$l = \ln L = \sum_{i=1}^{N(\tau)} \{\ln(z(\Lambda(T_i) - \Lambda(T_{i-1}))) + \ln(\lambda(T_i)) - \int_0^{\Lambda(T_i)-\Lambda(T_{i-1})} z(v)dv\} - \int_0^{\Lambda(\tau)-\Lambda(T_{N(\tau)})} z(v)dv. \quad (3)$$

3 Nonparametric estimation of $\lambda(\cdot)$

Often, we know little about the shape of $\lambda(\cdot)$, or we may suspect that it has a form which is difficult to describe as a single functional expression. Rather than assuming a pre-given functional form of $\lambda(\cdot)$, we may wish to let its shape be suggested by the data only.

Looking at the likelihood function (3), however, we see that the naïve non-parametric MLE of $\lambda(\cdot)$ is

$$\hat{\lambda}(t) = \sum_{i=1}^{N(\tau)} \delta(t - T_i),$$

where $\delta(\cdot)$ is the Dirac delta function. This estimate gives an infinite value to the likelihood function, but will rarely be particulary useful.

To obtain estimates which may be of some use, we obviously need to restrict the class of functions to which $\lambda(\cdot)$ may belong. In many practical cases, we will see that the systems we consider tend to fail more often when their operational time increases, or that they at least do not seem to improve over time. This suggests that many systems have an increasing, or at least nondecreasing, failure intensity. On the other hand some systems, for example software systems, could be expected to have a nonincreasing failure intensity.

We will therefore restrict ourself to the case where $\lambda(t)$ is monotone everywhere and continuous except at a finite number of points. We will first assume that $\lambda(\cdot)$ is nondecreasing. The case where it is nonincreasing is very similar, though, and will be considered at the end of the next subsection.

3.1 A monotone nonparametric maximum likelihood estimator (NPMLE) of $\lambda(\cdot)$

Bartozyński et al. [4] derived a nondecreasing NPMLE of $\lambda(\cdot)$ for an NHPP. We extend this approach to the TRP.

More specifically, we seek to maximize the log likelihood function l given in (3) under the condition that $\lambda(\cdot)$ belongs to the class of nonnegative, nondecreasing functions on $[0, \tau]$. The optimal $\lambda(t)$ must then consist of step functions closed on the left with no jumps except at some of the failure time points.

To see this, suppose $\bar{\lambda}(t)$ is a nondecreasing function which maximizes l in (3). Look at the interval $[T_{k-1}, T_k)$ for some fixed k, $1 \leq k \leq n+1$, where we define $T_{n+1} = \tau$. Let $v = \int_{T_{k-1}}^{T_k} \bar{\lambda}(u)du$. If we now choose $\lambda(t) \equiv v/(T_k - T_{k-1})$ on $[T_{k-1}, T_k)$, then obviously $\int_{T_{k-1}}^{T_k} \lambda(u)du$ also equals v, leaving all the terms of l unchanged except the term $\ln \lambda(T_{k-1})$. But clearly $\lambda(T_{k-1}) \geq \bar{\lambda}(T_{k-1})$ with equality only if $\bar{\lambda}(t) \equiv \lambda(t)$ on $[T_{k-1}, T_k)$. So unless $\bar{\lambda}(t)$ is constant on every interval of the form $[T_{k-1}, T_k)$, it is possible to increase l without violating the nondecreasing property.

Now let $n = N(\tau)$, let $\lambda_i = \lambda(T_i)$, $i = 0, 1, \ldots, n$, let $X_i = T_i - T_{i-1}$, $i = 1, 2, \ldots, n$ and let $X_{n+1} = \tau - T_n$. The problem of maximizing l is then simplified to the problem of maximizing

$$l' = \sum_{i=1}^{n} \{\ln z(\lambda_{i-1}X_i) + \ln \lambda_i - \int_0^{\lambda_{i-1}X_i} z(v)dv\} - \int_0^{\lambda_n X_{n+1}} z(v)dv, \quad (4)$$

subject to $0 \leq \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_n$. (See also [4].)

To proceed from here, we need to make some assumptions concerning $z(\cdot)$. Unfortunately, unless we choose $z(\cdot)$ to be constant, thus assuming that the process is an NHPP, it turns out to be difficult to obtain maximum likelihood estimates of both $z(\cdot)$ and $\lambda(\cdot)$ simultaneously. In some cases, however, if we choose a parametrization of $z(\cdot)$, it might be possible to obtain estimates by an iteration technique: One starts by guessing the values of the parameters, then maximizes the log likelihood function given these values to obtain an estimate of $\lambda(\cdot)$, and an estimate of the parameters is then obtained by maximizing the log likelihood function using the estimate of $\lambda(\cdot)$. The process is continued until convergence.

3.2 Assuming a Weibull distribution for F

We illustrate the above procedure by an important, yet relatively simple example. Let then $z(\cdot)$ be the hazard rate corresponding to a Weibull distribution,

$$z(t) = ab(at)^{b-1} \text{ for } a, b > 0$$

Here we put $a \equiv 1$ due to the nonuniqueness property of the TRP (Section 2.1).

Our problem then is to maximize (4), which now becomes

$$\begin{aligned} l'(b, \lambda_0, \ldots, \lambda_n) &= \sum_{i=1}^{n} \{\ln[b(\lambda_{i-1}X_i)^{b-1}] + \ln\lambda_i - (\lambda_{i-1}X_i)^b\} - (\lambda_n X_{n+1})^b \\ &= n\ln b + \sum_{i=1}^{n}\{(b-1)\ln\lambda_{i-1} + (b-1)\ln X_i + \ln\lambda_i - \lambda_{i-1}^b X_i^b\} - \lambda_n^b X_{n+1}^b \\ &= n\ln b + (b-1)\sum_{i=1}^{n}\ln X_i + (b-1)\ln\lambda_0 - \lambda_0^b X_1^b \\ &\quad + \sum_{i=1}^{n-1}\{b\ln\lambda_i - \lambda_i^b X_{i+1}^b\} + \ln\lambda_n - \lambda_n^b X_{n+1}^b. \end{aligned} \tag{5}$$

Now suppose that the value of b is known. Let $D_i = X_{i+1}^b$, $i = 0, 1, \ldots, n$. Let further $C_0 = (b-1)/b$, $C_n = 1/b$ and $C_i = 1$ for $i = 1, 2, \ldots, n-1$. Let $a_i = \lambda_i^b$, $i = 0, 1, \ldots, n$. Then the problem of maximizing (5) (for given b) simplifies to

$$\max_{a_0, a_1, \ldots, a_n} \sum_{i=0}^{n}\{C_i \ln a_i - D_i a_i\}$$

subject to $0 \le a_0 \le \ldots \le a_n$. (Note that with the possible exception of C_0, which may be zero or negative, and D_n, which may be zero, all C's and D's are positive.)

The solution to this problem is given by

$$a_0 = a_1 = \ldots = a_{k_1} = \min_{0 \le t \le n} \frac{\sum_{j=0}^{t} C_j}{\sum_{j=0}^{t} D_j} \tag{6}$$

$$a_{k_1+1} = a_{k_1+2} = \ldots = a_{k_2} = \min_{k_1+1 \le t \le n} \frac{\sum_{j=k_1+1}^{t} C_j}{\sum_{j=k_1+1}^{t} D_j} \tag{7}$$

$$\vdots$$

where k_1 is the value of t at which the minimum of (6) is attained, $k_2 > k_1$ is the value of t at which the minimum of (7) is attained, etc. Continue this way until $k_l = n$

The problem is in fact a special case of the isotonic regression problem, see for instance Robertson et al. [10, p. 39], while the method of solution leading to the above is a special case of the "Minimum lower sets Algorithm", [10, pp. 24-25].

Some remarks are appropriate at this stage. If there are more than one t that all give the same minimum value, choose the biggest. This will not affect the solution, but it will probably be computionally faster, and lead to strict inequalities $a_{k_i} < a_{k_i+1}$. If $b < 1$, a_0 will attain a negative value. Since we have required $a_0 \ge 0$, the optimum value of a_0 in this case will be 0. This, unfortunately, gives $\max l' = \infty$. To see that the estimator obtained is the MLE, impose the artificial constraint $a_0 \ge \delta > 0$ upon the problem. This will cause $a_0 = \delta$, thus ensuring a finite likelihood, but it will not affect any of the other a's if we choose δ small enough. Then, in the end, let $\delta \to 0$. If the system is observed to the time of the n'th failure, $\tau = T_n$, then $D_n = 0$ and $a_n = \infty$, also giving a infinite value of the likelihood function. In this case, we may impose $a_n \le M < \infty$, and let M approach infinity in the end.

We now return to the problem of maximizing (5) when also b is unknown. The iteration scheme goes as follows: Start by letting $b = b^{(1)}$, where $b^{(1)}$ is some initial choice. Calculate the C_i and D_i, then use (6-7) to find $a_0^{(1)}, a_1^{(1)}, \ldots, a_n^{(1)}$. This gives us our first estimate of $\lambda(\cdot)$, $\lambda_i^{(1)} = (a_i^{(1)})^{1/b^{(1)}}$. Now use the fact that the process $Y_i = \Lambda(T_i)$, $i = 1, \ldots, n$ is by assumption an RP(F) to find a new estimate $b^{(2)}$ of b. To do this, define first

$$Y_i^{(1)} = \sum_{j=0}^{i-1} \lambda_j^{(1)} X_{j+1}, \quad i = 1, 2, \ldots, n+1$$

Under the assumption that this process is a (censored) RP(F), where $F(x) = 1 - \exp(-x^b)$, the likelihood function is

$$L = \{\prod_{i=1}^{n} f(Y_i^{(1)} - Y_{i-1}^{(1)})\}[1 - F(Y_{n+1}^{(1)} - Y_n^{(1)})]$$

and the log-likelihood function is

$$
\begin{aligned}
l &= \sum_{i=1}^{n}\{\ln f(Y_i^{(1)} - Y_{i-1}^{(1)})\} + \ln[1 - F(Y_{n+1}^{(1)} - Y_n^{(1)})] \\
&= \sum_{i=1}^{n}\{\ln f(\lambda_{i-1}^{(1)} X_i)\} + \ln[1 - F(\lambda_n^{(1)} X_{n+1})] \\
&= \sum_{i=1}^{n}\{\ln b + (b-1)\ln[\lambda_{i-1}^{(1)} X_i] - [\lambda_{i-1}^{(1)} X_i]^b\} - [\lambda_n^{(1)} X_{n+1}]^b \\
&= n\ln b + (b-1)\sum_{i=1}^{n}\ln[\lambda_{i-1}^{(1)} X_i] - \sum_{i=1}^{n+1}[\lambda_{i-1}^{(1)} X_i]^b \qquad (8)
\end{aligned}
$$

which, except for a constant term, is (5) using $\lambda_0^{(1)}, \lambda_1^{(1)}, \ldots, \lambda_n^{(1)}$.

A new estimate $b^{(2)}$ is then obtained by maximizing (8) (numerically) with respect to b. (If $\lambda_0^{(1)} = 0$, then the process may be viewed as if it started at time $t = T_1$ instead of at time $t = 0$, and we may let the summation start at $i = 2$, thus avoiding any problems. Also, if $\lambda_n = \infty$, then $\tau = T_n$ and $\lambda_n X_{n+1} = 0$.) Then use this $b^{(2)}$ to obtain a new estimate $\lambda_i^{(2)}$, $i = 0, 1, \ldots, n$ of $\lambda(\cdot)$, then use this to find the estimate $b^{(3)}$ of b, and continue the process until the difference between two consecutive estimates of b is smaller than a given ϵ.

Because of the term $(b-1)\ln\lambda_0$ in (5), which can be made arbitrarily large by choosing both b and λ_0 close to 0, there is a small problem concerning the convergence of the iteration. As long as b is less than 1 and $\lambda_0 = 0$ for every iteration, we may simply ignore the troublesome term, and the iteration will converge, because every time we obtain a new estimate of b (or $\lambda(\cdot)$), this estimate is the MLE conditioned upon the previous estimate of $\lambda(\cdot)$ (or b). Thus with every new iteration, the unconditional log likelihood function (5) will increase. And as long as b stays above 1 for every iteration, there is no term in (5) that may cause problems, and the iteration converges by the same argument as above.

However, if b from one iteration to the next switches from a value below 1 to a value above 1 (or opposite), we should suddenly include (or exclude) the term $(b-1)\ln\lambda_0$ in the log likelihood. This could easily lead to a relatively huge change in the estimates, and b might drop below (or rise above) 1 again. And the iteration procedure could start to oscillate between values of b on both sides of 1.

To avoid this, we suggest the following modification: Start by assuming that $\lambda_0 = 0$, and ignore the term $(b-1)\ln\lambda_0$. Go through the iteration. If the value of b when it converges is below 1, stop, and use the estimates found. If not, start again, this time including the term $(b-1)\ln\lambda_0$. If the iteration converges, use the estimates now found, if not, use the previous found values.

3.2.1 Nonincreasing $\lambda(\cdot)$

In some cases, we would expect $\lambda(\cdot)$ to be nonincreasing rather than nondecreasing. An example could be software reliability; when a software error is detected and corrected, we would not expect it to show up again, and the system should be improving over time.

The method of obtaining a nonincreasing MLE of $\lambda(\cdot)$ is very similar to the method just described for nondecreasing MLE. The estimator will still consist of step functions with jumps only at some of the failure times, but it will be left continuous instead of right continuous. Otherwise the algorithm suggested will be almost the same:

Let $\lambda_i = \lambda(T_i)$ and $X_i = T_i - T_{i-1}$, $i = 1, 2, \ldots, n+1$, where T_{n+1} is defined as τ. Looking at (3), it is obvious that the optimum value of $\lambda(t)$ is zero for $t > T_{N(\tau)}$, i.e. $\lambda_{n+1} = 0$. Thus the problem of maximizing (3) simplifies to the problem of maximizing

$$l' = \sum_{i=1}^{n} \{\ln z(\lambda_i X_i) + \ln \lambda_i - \int_0^{\lambda_i X_i} z(v)dv\}$$

subject to $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$.

By assuming that $z(\cdot)$ is the hazard rate of a Weibull RP, we may proceed in the same manner as we did earlier. We wish to maximize

$$l'(b, \lambda_1, \ldots, \lambda_n) = n \ln b + (b-1) \sum_{i=1}^{n} \ln X_i + \sum_{i=1}^{n} \{\ln \lambda_i^b - \lambda_i^b X_i^b\}$$

which actually is a simpler problem than the one we faced when $\lambda(\cdot)$ was nondecreasing; this time there is no choice of b or λ_i that may give an infinite likelihood. We may then use the same iteration scheme as in the previous section to obtain estimates of b and λ_i, the only difference being that we have to solve the problem

$$\max_{a_1, a_2, \ldots, a_n} \sum_{i=1}^{n} \{\ln a_i - D_i a_i\}$$

subject to $a_1 \geq \ldots \geq a_n \geq 0$. (Where $D_i = X_i^b$ and $a_i = \lambda_i^b$ as before, while $C_i = 1$, $i = 1, 2, \ldots, n$.) The solution to this problem is similar to the one for the nondecreasing case: e

$$a_1 = a_2 = \ldots = a_{k_1} = \max_{1 \leq t \leq n} \frac{t}{\sum_{j=1}^{t} D_j} \tag{9}$$

$$a_{k_1+1} = a_{k_1+2} = \ldots = a_{k_2} = \max_{k_1+1 \leq t \leq n} \frac{t - k_1}{\sum_{j=k_1+1}^{t} D_j} \tag{10}$$

$$a_{k_2+1} = a_{k_2+2} = \ldots = a_{k_3} = \max_{k_2+1 \le t \le n} \frac{t - k_2}{\sum_{j=k_2+1}^{t} D_j} \tag{11}$$

$$\vdots$$

where k_1 is the value of t at which the maximum of (9) is attained, $k_2 > k_1$ is the value of t at which the maximum of (10) is attained, $k_3 > k_2$ is the value of t at which the maximum of (11) is attained, etc. Continue this way until $k_l = n$.

4 Examples

Example 1: A simulated TRP

We simulated 200 failure times from a TRP with $\lambda(t) = 0.01t^{0.5}$ and $F(x) = 1 - \exp(-x^3)$, and used both parametric estimation, assuming $\lambda(t) = \alpha\beta t^{\beta-1}$ and $F(x) = 1 - \exp(-x^b)$, and the nonparametric method suggested in the previous section to estimate F and $\lambda(\cdot)$.

The parametric maximum likelihood estimates with approximate standard deviations were computed using the approach of [7] and are given in Table 1.

It is now of interest to compare the estimates of $\lambda(\cdot)$ and b obtained by the parametric and nonparametric methods, respectively.

The "nonparametric" estimate of b using the method of Section 3 is 3.269. Estimates of bias and standard deviation of the estimator will be computed by the method of bootstrapping in Section 4.1.

In order to compare the two estimates of $\lambda(\cdot)$ in a fair way, we scale the parametric estimates of F and $\lambda(\cdot)$ to make the expectations corresponding to the parametric and the "nonparametric" estimates of F the same. As explained in Section 2.1, a TRP$(F, \lambda(\cdot))$ with

$$F(x) = 1 - e^{-x^b}$$

and

$$\lambda(t) = \alpha\beta t^{\beta-1}$$

has the same distribution as TRP$(F_c, \lambda_c(\cdot))$ where

$$F_c(x) = F(cx) = 1 - e^{-(cx)^b}$$

and

$$\lambda_c(t) = c^{-1}\lambda(t) = c^{-1}\alpha\beta t^{\beta-1}$$

for any $c > 0$.

If the expectation of F is μ, then the expectation of F_c is clearly μ/c. So if we want the expectation of the scaled distribution F_c to equal a given number m, typically the expectation of the nonparametric estimate of F, we have to choose $c = \mu/m$.

By the standard formula for the expectation of the Weibull distribution we find the expectation of the parametric estimate of F to be $\Gamma(1/3.024 + 1)$, while the expectation of the "nonparametric" estimate is $\Gamma(1/3.269 + 1)$. To scale the parametric estimate we thus choose

$$c = \frac{\Gamma(1/3.024 + 1)}{\Gamma(1/3.269 + 1)} = 0.9963$$

In Figure 1 we plot the nonparametric estimate of $\lambda(\cdot)$ together with the (scaled) parametric estimate $\hat{\lambda}(t) = c^{-1}\hat{\alpha}\hat{\beta}t^{\hat{\beta}-1}$, and in addition we plot the corresponding cumulative functions. As we see, there is a good correspondence between the parametric and nonparametric estimates.

To check the convergence of the nonparametric method, we used three different starting values of b, 0.1, 1 and 10 respectively. In all three cases both b and $\lambda(\cdot)$ converged to the same values. Using a convergence criterion that the difference between to consecutive values of b should be less than 10^{-6} they all converged within 8 iterations.

Example 2: U.S.S Halfbeak data

Meeker and Escobar [8, Table 16.4] display 71 times of unscheduled maintenance actions for the U.S.S. Halfbeak number 4 main propulsion diesel engine. Figure 2 shows the plot of cumulative failure number against time (in thousands of hours of operation).

From the figure it seems fairly obvious that the system is deteriorating, and a nondecreasing $\lambda(\cdot)$ is a fair assumption, even if we might suspect that $\lambda(t)$ could be decreasing for $t > 22$. A parametric estimation using the power law function $\lambda(t) = \alpha\beta t^{\beta-1}$, on the other hand, does not (judging from Figure 2) seem to be able to give a very good fit to this set of data. We will, however, use this parametric model anyway for comparison with the nonparametric approach. The parametric MLE's are given in Table 2.

The "nonparametric" estimate of b is 0.937, giving a scale factor (see Example 1)

$$c = \frac{\Gamma(1/0.762 + 1)}{\Gamma(1/0.937 + 1)} = 1.1408.$$

Plots of both the scaled parametric and the nonparametric estimates of $\lambda(t)$ and $\Lambda(t)$ are found in Figure 3. This time, however, the "nonparametric" estimate of is so far from the parametric estimate that it probably is rather meaningless to compare the different estimates.

The problem is of course that the parametric power law model does not fit the data sufficiently well, as is indicated by the right panel plot of Fig. 3. A standard confidence interval for b based on Table 2 (estimate $\pm$ 2 $\times$ standard deviation) does not include the value 1, hence we reject a null hypothesis of NHPP. On the other hand, the "nonparametric" estimate of b is close to 1 (0.937) and hence indicates that an NHPP model still may be appropriate, which is also reasonable for this type of data. Likewise is an increasing trend function reasonable and a possible conclusion from this example is that nonparametric estimation of the trend function and parametric estimation of the renewal distribution is a reasonable approach, while a power law trend function does not fit and hence makes the estimate of b also suspicious.

4.1 Bootstrap estimates of bias and standard deviation

Example 1 (continued)

A parametric bootstrapping (Efron and Tibshirani [6]) was performed by assuming that the failure times are time-truncated at the largest observation which is 931.92. In order to obtain a bootstrap sample we then first drew a sufficiently large number n of observations $Y_1^*, Y_2^*, \ldots, Y_n^*$ from the Weibull cdf $F(x) = 1 - \exp(-x^{3.269})$. Then we used the transformation $T_j^* = \hat{\Lambda}^{-1}(\sum_{i=1}^{j} Y_i^*)$, where $\hat{\Lambda}(\cdot)$ is our nonparametric estimate of $\Lambda(\cdot)$, to obtain a sequence of failure times $T_1^*, T_2^*, \ldots, T_n^*$ of which the ones below 931.92 were used as our first bootstrap sample. This procedure was repeated 50 times, to give us 50 bootstrap samples.

For each of the bootstrap samples we went through the same estimation procedure as we did with the original sample, to obtain the bootstrap estimates $\hat{b}_i^*$ and $\hat{\lambda}_i^*(\cdot)$, $i = 1, 2, \ldots, 50$.

The bootstrap estimate of the standard deviation of $\hat{b}$ is 0.209, which may be compared to the normal theory estimate of the standard deviation in the parametric case, which is 0.162 (Table 1). The estimates are reasonably close, and we should also expect the former to be larger because of a less specified model.

The mean of the bootstrap estimates, on the other hand, is 3.578, which could indicate that our estimation is somewhat biased. The bootstrap estimate of bias is thus $3.578 - 3.269 = 0.309$, which is rather large, in fact it is almost a factor 1.5 times the estimated standard deviation. If we try to remove the bias by subtraction, we get the bias-corrected estimate to be $3.269 - 0.309 = 2.960$. This is very close to both the "true value" of 3, and the estimated value in the parametric case of 3.024. See for example [6] for a discussion on bias corrections.

Figure 4 shows the original estimate of $\Lambda(t)$ along with the estimated curves from the 50 bootstrap samples, giving an indication of the variability of $\hat{\Lambda}(\cdot)$. (We have not scaled these estimates to make the expectation of all the bootstrap estimates of F the same, as this would probably move some of the variation from $\Lambda(\cdot)$ to the scale parameter c.) The figure seems to indicate a relatively low degree of variation, slightly increasing with increasing t. Also, our original estimate seems to be placed neatly in the middle of all the bootstrap estimates, so we do not believe there is any particular bias in this case.

Example 2 (continued)

In this example we have data from a real system, and although we assumed during the estimation that the renewal distribution is Weibull, we should suspect that it could exhibit some non-Weibull behaviour. To weaken the effect that a non-Weibull distribution may have on bootstrap samples, we chose not to draw the Y^*'s from the Weibull distribution $F(x) = 1 - \exp(-x^{0.937})$, but rather from the empirical distribution function

$$\hat{F}(x) = \frac{\text{number of } \hat{Y}_i \leq x}{n}$$

where $\hat{Y}_i = \hat{\Lambda}(T_i) - \hat{\Lambda}(T_{i-1})$, $i = 1, 2, \ldots, n$, and $n = 71$ is the total number of failure times.

Otherwise we proceeded in exactly the same manner as in Example 1. The estimated standard deviation of $\hat{b}$ is now 0.133, while the mean of the bootstrap b's is 1.016, giving an estimate of the bias, $1.016 - 0.937 = 0.079$. Since this is small compared to the estimated standard deviation we do not perform a bias correction (see [6]).

Figure 5 shows the original estimate $\hat{\Lambda}(\cdot)$ and the bootstrap simulations of $\Lambda(\cdot)$, and in this example the variability is rather large, particularly in the area where $\hat{\Lambda}(\cdot)$ starts to increase sharply. The variability is larger than in Figure 4. This is probably due to having only 71 failure times, compared to 200 in Example 1, and presumably also due to the fact that we consider a real system where certain changes apparently took place around $t = 19$.

Figure 5 also indicates that $\hat{\Lambda}(\cdot)$ perhaps is underestimating the true $\Lambda(\cdot)$, since most of the bootstrap curves apparently are below the NPMLE $\hat{\Lambda}(\cdot)$. A bootstrap evaluation of the standard deviation and bias of $\hat{\Lambda}(\cdot)$ is possible, but somewhat involved. We would have to calculate $\hat{\Lambda}^*_{(\cdot)}(t)$ for several values of t, and then use interpolation to approximate $\hat{\Lambda}^*_{(\cdot)}(\cdot)$ between these values. We did this for only one value of t, $t = 19.067$, which is where the sharp increase in $\hat{\Lambda}(t)$ starts. The bootstrap estimate of the standard deviation of $\hat{\Lambda}(19.067)$ is 4.286, and the mean of the bootstrap estimates at this point is 16.122. Compared to $\hat{\Lambda}(19.067) = 17.228$ we see that the estimate of bias is $16.122 - 17.228 = -1.116$ which is relatively small compared to the estimated standard deviation. The apparent underestimation could be due to forcing $\hat{\lambda}(\cdot)$ to be nondecreasing and hence $\hat{\Lambda}(\cdot)$ to be convex, while Figure 2 indicates that $\lambda(\cdot)$ might be decreasing for large t.

5 Nonparametric estimation in the case of more than one system

Suppose we have data from more than one system, where systems are assumed to be independent and follow identical TRP$(F, \lambda(\cdot))$ laws. We are then faced with the problem that the superposition of several trend-renewal processes is not, in general, another trend-renewal process. Although it is still true that the MLE of $\lambda(\cdot)$, when restricted to a monotone function, must consist of step functions closed on the left with jumps only at some of the failure time points, the method suggested for the case where we only had one system is not applicable. However, if we have many systems, all observed on the same time interval $(0, \tau]$ we might be able to obtain a useful estimate of $\lambda(\cdot)$, based on the convergence theorem of superimposed renewal processes (Drenick [5]). This estimate will not, however, be the exact MLE of $\lambda(\cdot)$, except for the special case where all the processes are NHPP's.

Indeed, Drenick [5] showed that the superposition of n independent renewal processes in time equilibrium is a homogenous Poisson process under certain mild conditions when $n \to \infty$. One might therefore argue that the superposition of a large number of RP's, all observed over a long interval of time, at least could be approximated by an HPP. Although there are some problems here, for example that the asymptotic HPP often will not be visible until after some time $t > 0$ (see e.g. Arjas et al. [2]), we may use this approximation to find a reasonable nondecreasing nonparametric estimator of $\lambda(\cdot)$ in the case where we have more than one system.

Suppose we have n independent TRP$(F, \lambda(\cdot))$'s observed on $(0, \tau]$. Let m_j be the number of failures on system j, let T_{ij}, $i = 1, 2, \ldots, m_j$, $j = 1, 2, \ldots, n$ be the time of the i'th failure on system number j, and let τ_k, $k = 1, 2, \ldots, s$ where $s = \sum_{j=1}^{n} m_j$ be the (ordered) failure times of the superimposed process.

Since the time-transformed processes $\Lambda(T_{1j}), \Lambda(T_{2j}), \ldots, \Lambda(T_{m_j j})$, $j = 1, 2, \ldots, n$ all are RP(F)'s, we will assume that the superposition of these transformed processes can be approximated by an HPP with intensity n/μ, that is $\Lambda(\tau_1), \Lambda(\tau_2), \ldots, \Lambda(\tau_s)$ is (approximately) an HPP(n/μ). (Where μ is the expectation of F, $\mu = \int_0^\infty x dF(x)$.) Then from the definition of an NHPP we see that the superimposed TRP may be approximated with an NHPP($\bar{\lambda}(\cdot)$), where $\bar{\lambda}(t) = n\lambda(t)/\mu$. We will proceed as if this holds exactly.

We start by finding a nondecreasing estimate of $\bar{\lambda}(\cdot)$. If we accept the assumption that the superimposed process is an NHPP($\bar{\lambda}(\cdot)$), then the log likelihood function is

$$l = \sum_{k=1}^{s} \ln \bar{\lambda}(\tau_k) - \int_0^{\mathcal{T}} \bar{\lambda}(u)du$$

The nondecreasing sequence $\hat{\bar{\lambda}}_1, \hat{\bar{\lambda}}_2, \ldots, \hat{\bar{\lambda}}_s$ which maximizes this can be found as in (6)-(7). We will regard this estimate as the true value of $\bar{\lambda}(\cdot)$ when estimating F below.

Now choose a parametric model for F. We want to estimate $\lambda(\cdot)$ and the parameters of F. Note that by assumption

$$\lambda(t) = \bar{\lambda}(t)\frac{\mu}{n}$$

or

$$\lambda_k = \bar{\lambda}_k\frac{\mu}{n} \qquad k = 0, 1, \ldots, s$$

If we knew the value of μ, we could estimate $\lambda(\cdot)$ by $\hat{\lambda}(t) = \hat{\bar{\lambda}}(t)\mu/n$, and make the transformation

$$Y_{ij} = \hat{\Lambda}(T_{ij}) = \int_0^{T_{ij}} \hat{\lambda}(u)du = \int_0^{\tau_v} \hat{\lambda}(u)du = \sum_{k=0}^{v-1} \hat{\lambda}_k[\tau_{k+1} - \tau_k] = \frac{\mu}{n}\sum_{k=0}^{v-1} \hat{\bar{\lambda}}_k[\tau_{k+1} - \tau_k] \quad (12)$$

for $i = 1, 2, \ldots, m_j$ and $j = 1, 2, \ldots, n$. (Choosing v to make $\tau_v = T_{ij}$.) Then for all j, we could let

$$Y_{m_j+1,j} = \hat{\Lambda}(\tau) = \int_0^{\tau} \hat{\lambda}(u)du = \frac{\mu}{n}\sum_{k=0}^{s-1}\{\hat{\bar{\lambda}}_k[\tau_{k+1} - \tau_k]\} + \frac{\mu}{n}\hat{\bar{\lambda}}_s[\tau - \tau_s] \qquad (13)$$

The processes $Y_{1j}, Y_{2j}, \ldots$ would be assumed to be RP(F)'s, their total likelihood function would be

$$L = (\prod_{j=1}^{n}\{\prod_{i=1}^{m_j} f(X_{ij})\}[1 - F(X_{m_j+1,j})])$$

and their log likelihood function

$$l = \sum_{j=1}^{n} (\sum_{i=1}^{m_j} \{\ln f(X_{ij})\} + \ln[1 - F(X_{m_j+1,j})]) \tag{14}$$

where $X_{ij} = Y_{ij} - Y_{i-1,j}$, $i = 1, 2, \ldots, m_j + 1$, $j = 1, 2, \ldots, n$.

This could then be maximized to obtain estimates of the parameters of F.

Unfortunately, the value of μ is unknown, which unables us to compute X_{ij}. However, as μ depends only upon the same parameters as do f and F, it may still be possible to obtain the MLE's of the parameters.

For all i, j define

$$X_{ij}^{(0)} = \frac{1}{\mu} X_{ij} \tag{15}$$

$X_{ij}^{(0)}$ may be computed by letting $\mu = 1$ in (12) and (13), and is thus a function of only known observations and estimates. Inserting (15) in (14) gives

$$l = \sum_{j=1}^{n} (\sum_{i=1}^{m_j} \{\ln f(\mu X_{ij}^{(0)})\} + \ln[1 - F(\mu X_{m_j+1,j}^{(0)})])$$

which may then be maximized to obtain estimators for the parameters of F.

Having thus found an estimate $\hat{F}$ of F, we may estimate μ by

$$\hat{\mu} = \int_0^\infty (1 - \hat{F}(x))dx$$

The estimate of $\lambda(t)$ is then finally

$$\hat{\lambda}(t) = \hat{\hat{\lambda}}(t) \frac{\hat{\mu}}{n}$$

To illustrate we will consider the same example as in the previous sections, where F is the Weibull density function; $F(x) = 1 - \exp(-x^b)$, with expectation $\Gamma(\frac{1}{b} + 1)$. The log likelihood function to maximize will then depend only upon b,

$$l = \sum_{j=1}^{n} (\sum_{i=1}^{m_j} \{\ln[b(\mu X_{ij}^{(0)})^{b-1} \exp(-[\mu X_{ij}^{(0)}]^b)]\} + \ln[\exp(-[\mu X_{m_j+1,j}^{(0)}]^b)]$$

$$= \sum_{j=1}^{n} (\sum_{i=1}^{m_j} \{ \ln b + (b-1) \ln(\mu X_{ij}^{(0)}) - [\mu X_{ij}^{(0)}]^b \} - [\mu X_{m_j+1,j}^{(0)}]^b)$$

$$= \sum_{j=1}^{n} (m_j \ln b + (b-1) \sum_{i=1}^{m_j} \ln(\Gamma(\frac{1}{b}+1) X_{ij}^{(0)}) - \sum_{i=1}^{m_j+1} [\Gamma(\frac{1}{b}+1) X_{ij}^{(0)}]^b) \quad (16)$$

If $\hat{b}$ maximizes (16), then our estimate of $\lambda(t)$ is

$$\hat{\lambda}(t) = \Gamma(\frac{1}{\hat{b}}+1) \frac{\hat{\hat{\lambda}}(t)}{n}$$

6 Concluding remarks

The aim of this paper has been to develop useful nonparametric estimation techniques for analysis of repairable systems, particularly when the systems are modelled by trend-renewal processes.

A monotone NPMLE of $\lambda(\cdot)$ was developed in Section 3 for data given from a single system. The estimator was computed assuming that the renewal distribution F was a Weibull distribution function, and as shown by the examples it performed well when compared to the parametric MLE. The Weibull distribution was chosen for the renewal distribution because it is important in reliability analyses, and because it gives a mathematically simple model. In addition it implies a smooth connection with the exponential distribution which is the distribution which is inherent in an NHPP assumption. It will, though, be of interest to see how the suggested algorithm can be adapted to other distributions, such as the gamma or log-normal distributions.

In the case where we have data from more than one system, the suggested estimator (Section 5) was based on the approximation of the superposition of several TRP's by an NHPP. This is of course exactly true only if the original processes are NHPP's. It would be interesting to use better approximations, which may correct for some of the facts that we only observe the systems for a finite time range, that the processes are not necessarily in equilibrium, or that we have only a finite, and sometimes even a small, number of systems. There exist results in the literature regarding superposition of renewal processes which might be modified, see e.g. Arjas et al. [2].

The nonparametric estimators that we have suggested are based on the assumption that the trend function $\lambda(\cdot)$ is monotone. While this is reasonable (or at least not unreasonable) in many situations, a slightly more general estimator could perhaps be more useful. In Example 2, we did suspect that $\lambda(t)$ could be decreasing, not increasing for larger t. A natural extension could be to allow $\lambda(t)$ to be nondecreasing in one interval, and nonincreasing in another. We would have to face the problem of estimating at what point t $\lambda(t)$ changes from nondecreasing to nonincreasing, as well as how to "tie" the two intervals together. More generally, it is of course possible to consider nonparametric estimation of $\lambda(\cdot)$ without monotonicity restrictions. The classical approach here is to use kernel methods in a similar manner to what is done in probability density estimation.

References

[1] Andersen P, Borgan Ø, Gill R, Keiding N. Statistical models based on counting processes. New York: Springer, 1992.

[2] Arjas E, Hansen CK, Thyregod P. Heterogeneous part quality as a source of reliability improvement in repairable systems. Technometrics 1991;33:1-12.

[3] Ascher H, Feingold H. Repairable systems reliability: modeling, inference, misconceptions and their causes. New York: Marcel Dekker, 1984.

[4] Bartozyński R, Brown BW, McBride CM, Thompson JR. Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. Annals of Statistics 1981;9(5):1050-1060.

[5] Drenick RF. The failure law of complex equipment. J. Soc. Indust. Appl. Math. 1960;8:680-690.

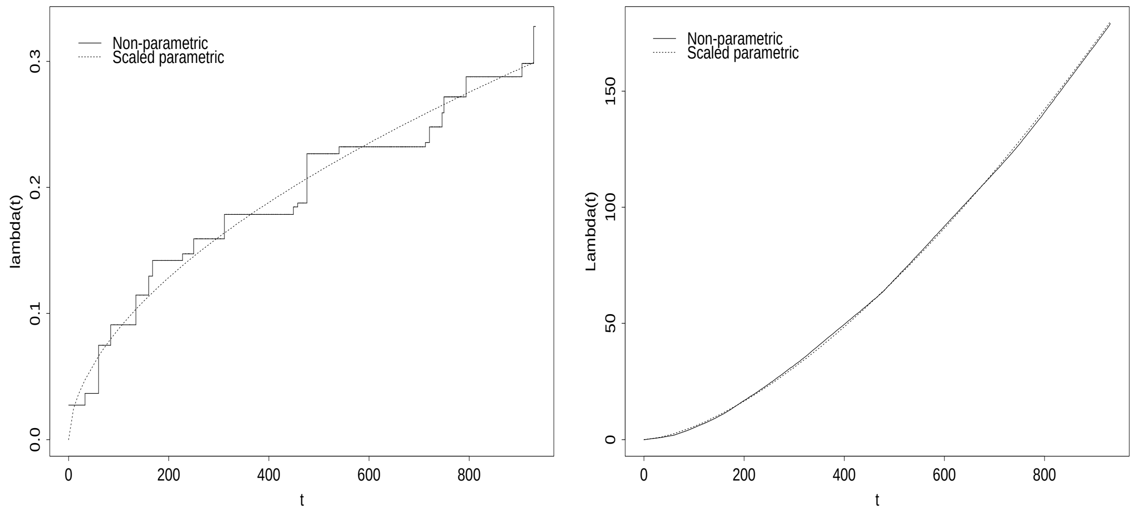
[6] Efron B, Tibshirani R. An introduction to the bootstrap. Boca Raton, Florida: Chapman & Hall/CRC, 1993.

[7] Lindqvist BH, Elvebakk G, Heggland K. The trend-renewal process for statistical analysis of repairable systems. Technometrics 2003;45:31-44.

[8] Meeker WQ, Escobar LA. Statistical methods for reliability data. New York: Wiley, 1998.

[9] Pham H, Wang H. Imperfect maintenance. European Journal of Operations Research 1996;94:425-428.

[10] Robertson T, Wright FT, Dykstra RL. Order restricted statistical inference. New York: Wiley, 1988.



,

Fig. 1. Parametric and nonparametric estimates of $\lambda(t)$ (left panel) and $\Lambda(t)$ (right panel), a simulated TRP

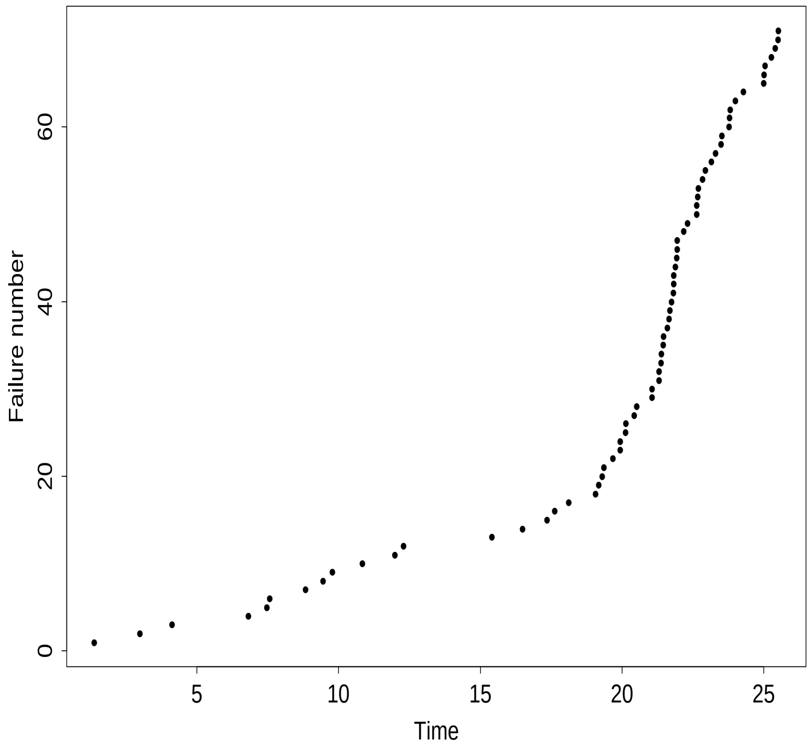
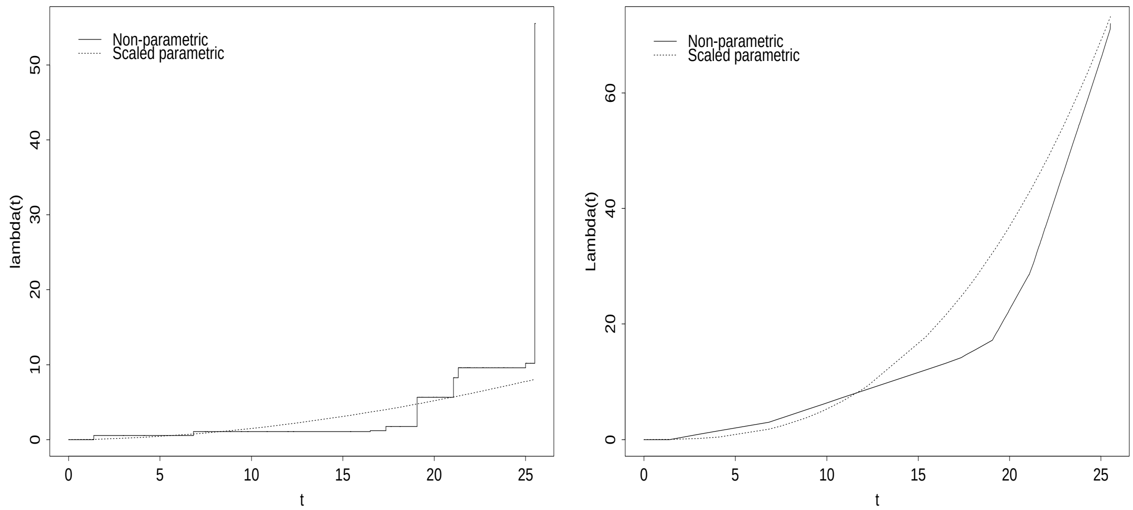


Fig. 2. Failure number against time in thousands of hours of operation, U.S.S. Halfbeak data



,

Fig. 3. Parametric and nonparametric estimates of $\lambda(t)$ (left panel) and $\Lambda(t)$ (right panel), USS Halfbeak data

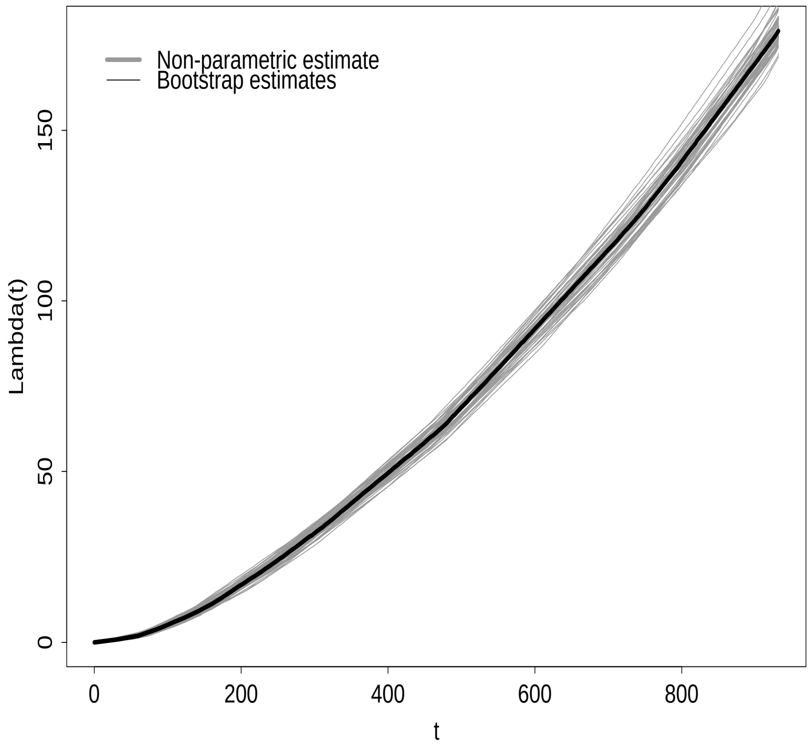


Fig. 4. Original and bootstrap estimates of $\Lambda(t)$, a simulated TRP

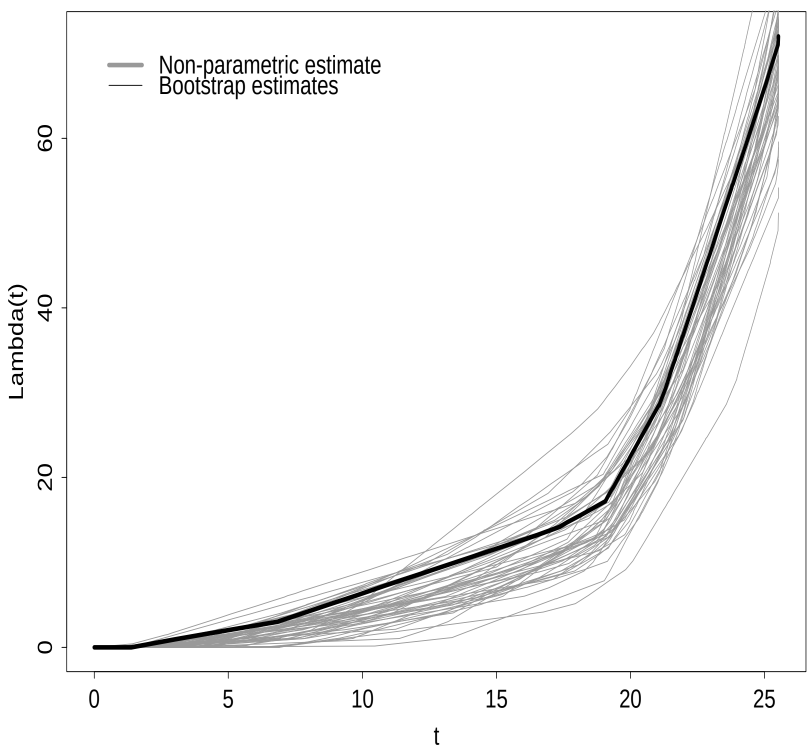


Fig. 5. Nonparametric original and bootstrap estimates of $\Lambda(t)$, U.S.S. Halfbeak data

Table 1
The parametric maximum likelihood estimates (and estimated standard deviations), a simulated TRP

$\hat{\alpha}$	$\hat{\beta}$	$\hat{b}$
0.00448	1.550	3.024
(0.00107)	(0.035)	(0.162)

Table 2
The parametric maximum likelihood estimates (and estimated standard deviations), U.S.S. Halfbeak data

$\hat{\alpha}$	$\hat{\beta}$	$\hat{b}$
0.00936	2.808	0.762
(0.01225)	(0.402)	(0.071)