Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 119 (2004) 171–189

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Non-parametric estimation under progressive censoring

Laurent Bordes*

Équipe de Mathématiques Appliquées, Université de Technologie de Compiègne, B.P. 529, 60205 Compiègne Cedex, France

Received 16 May 2001; accepted 7 August 2002

Abstract

We consider non-parametric estimation of cumulative hazard functions and reliability functions of progressively type-II right censored data. As shown in the book of Balakrishnan and Aggarwala (Progressive Censoring, Birkhäuser, Basel, 2000), many results of classical order statistics can be generalized to this kind of statistics. These authors proposed also many inferential methods for parametric models. In this paper we show that non-parametric maximum likelihood estimators (NPMLE) may also be derived under such censoring schemes. These estimators are obtained in a reliability context but they can also be extended to arbitrary continuous distribution functions. Since the large sample properties of the NPMLE depend on counting processes based upon generalized order statistics that are generated by progressive censoring, we need to establish some basic properties of these processes (e.g. martingales properties and weak consistency). Finally, the non-parametric estimator of the reliability is compared with two parametric estimators for a real data set and additionally, some Monte-Carlo simulations are provided.

MSC: Primary: 62G30; 62N01; Secondary: 62G05; 62N05

Keywords: Generalized order statistics; Non-parametric estimation; Type II progressive censoring; Martingales; Counting processes; Reliability

1. Introduction

Order statistics are widely used in statistical modeling and inference. In a paper presenting a unified approach for many models, based on order statistics and record values, Kamps (1995) proposed a generalized form of the joint distribution of n ordered

* Fax: +33-3-44-23-44-77.
E-mail address: laurent.bordes@utc.fr (L. Bordes).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00414-7

random variables. One of the models included in this general setup is the Type II progressive censoring scheme as defined by Balakrishnan and Aggarwala (2000).

This scheme of censoring appears to be of great importance in planning duration experiments in reliability studies. In many industrial experiments involving lifetimes of machines or units, experiments have to be terminated early or the number of experiments must be limited due to a variety of circumstances (e.g. when expensive items must be destroyed, when experiments are time-consuming and expensive, etc.). In addition, some lifetests require removals of functioning test specimens to collect degradation related information to failure time data. The samples that arise from such experiments are called censored. The planning of experiments with the aim of reducing both the number of failures and the total duration of the experiment leads naturally to the well-known type-I and type-II right censoring schemes. For many references and historical notes on this subject we refer to Balakrishnan and Aggarwala (2000). The progressive censoring scheme, as will be introduced hereafter, has the same objectives, but it is constructed with the aim of moderating the loss of information by reducing the number of failures with respect to the full sample approach. Montanari and Cacciari (1988) reported results of progressively censored data aging tests on XLPE-insulated cable models under combined thermal-electrical stresses. In this experiment, live specimens were removed at selected times and/or at the time of breakdowns. The progressive censoring sampling plans by Montanari and Cacciari (1988) are considered in Balasoorya et al. (2000) with the aim of defining optimal sampling plans under Weibull estimation.

In the classical type-II right censoring scheme only the first m failures are observed for a sample of size n whereas for a progressively censored sample, the loss of information of $n-m$ durations (as for the type-II plan) is organized sequentially as it is described subsequently. Let $X_1,\dots,X_n$ be independent and identically distributed random lifetimes of n items. A type-II progressively right censored sample may be obtained in the following way: at the time of the first failure, denoted with $X_{1:m:n}$, r_1 surviving items are removed at random from the $n-1$ remaining surviving items, at the time of the next failure, denoted with $X_{2:m:n}$, r_2 surviving items are removed at random from the $n-r_1-2$ remaining items, and so on. At the time of the mth failure, all the remaining $r_m=n-m-r_1-\cdots-r_{m-1}$ surviving items are censored. Therefore, a progressively type-II right censoring scheme is specified by integer numbers n, m and $r_1,\dots,r_{m-1}$ with the constraints $n-m-r_1-\cdots-r_{m-1}\geqslant 0$ and $n\geqslant m\geqslant 1$.

Remark 1. If $r_1=\cdots=r_{m-1}=0$ we get the usual type-II right censored sample; thus we observe the first m order statistics $X_{1:n},\dots,X_{m:n}$ and the $n-m$ remaining times are right censored by $X_{m:n}$. If moreover $m=n$, the usual order statistic is obtained.

In their book, Balakrishnan and Aggarwala (2000) developed many parametric methods to analyze progressively type-II right censored data. Our aim here is to develop a non-parametric approach for this kind of censored data in order to estimate both reliability and cumulative hazard functions. In Section 2 we show how to derive NPMLE for both the cumulative hazard function and the reliability function for lifetime distribution. A simple transformation enables therefore to yield an estimator for any continuous distribution function. Section 3 is devoted to several known results for progressively censored samples. We also give in this section an important martingale property of the counting process based upon progressively censored data. In Section 4 we give assumptions and preliminary results on the asymptotic behavior of a progressively censored sample. Finally, the main asymptotic results are obtained by standard methods for counting processes (Lenglart inequality, Rebolledo theorem); see Andersen et al. (1993). Section 5 contains an application to a data set constructed from Nelson (1982, p. 105, Table 1.1) and some simulations illustrating the behavior of the reliability estimator for moderate sample size. Concluding remarks are given in Section 6.

2. Non-parametric estimation

We consider a type-II progressively censored sample $X_{1:m:n},\dots,X_{m:m:n}$ observed from a sample $X_1,\dots,X_n$ of independent and identically distributed non-negative random variables with distribution function F, reliability function R, density function f, hazard rate function λ and cumulative hazard function Λ. From Balakrishnan and Aggarwala (2000), the joint density of the progressively censored sample (p.c.s.) $(X_{1:m:n},\dots,X_{m:m:n})$ under the scheme $(r_1,\dots,r_m)$ where $r_1 \geqslant 0,\dots,r_{m-1} \geqslant 0$ and $r_m = n-m-r_1-\cdots-r_{m-1} \geqslant 0$ is given by

$$f_{X_{1:m:n},\dots,X_{m:m:n}}(x_1,\dots,x_m) = \prod_{i=1}^{m} \alpha_i^m f(x_i) R^{r_i}(x_i) 1(0 < x_1 < \cdots < x_m), \tag{1}$$

where $\alpha_i^m = \sum_{j=i}^{m} r_j + m - i + 1$ and $1(\cdot)$ is the set indicator function. It follows that if $(x_1,\dots,x_m)$ is an observation of $(X_{1:m:n},\dots,X_{m:m:n})$ the likelihood function is proportional to

$$\mathscr{L}_{(x_1,\dots,x_m)}(\lambda) = \prod_{i=1}^{m} f(x_i) R^{r_i}(x_i)$$

and then, the log-likelihood function, up to an additive (but non informative) term, is equal to

$$l_{(x_1,\dots,x_m)}(\lambda) = \sum_{i=1}^{m} (\log \lambda(x_i) - (r_i+1)\Lambda(x_i)).$$

We are now looking for a discrete measure on the set $(x_1,\dots,x_m)$, under which the log-likelihood function is maximized. The discrete measure is called $\hat{\lambda} = \sum_{i=1}^{m} \hat{\lambda}_i \delta_{x_i}$, where δ_{x_i} is the Dirac measure at point x_i. We have

$$l_{(x_1,\dots,x_m)}(\hat{\lambda}) = \sum_{i=1}^{m} \left(\log \hat{\lambda}_i - (r_i+1) \sum_{j=1}^{i} \hat{\lambda}_j \right).$$

Now, seen as a function of $(\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ the above quantity has a gradient equal to

$$\nabla l_{(x_1,\ldots,x_m)}(\hat{\lambda}) = \begin{pmatrix} \dfrac{\partial}{\partial \hat{\lambda}_1} l_{(x_1,\ldots,x_m)}(\hat{\lambda}) \\ \vdots \\ \dfrac{\partial}{\partial \hat{\lambda}_m} l_{(x_1,\ldots,x_m)}(\hat{\lambda}) \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\hat{\lambda}_1} - \displaystyle\sum_{i=1}^{m} (r_i + 1) \\ \vdots \\ \dfrac{1}{\hat{\lambda}_m} - \displaystyle\sum_{i=m}^{m} (r_i + 1) \end{pmatrix}.$$

We take for $\hat{\lambda}$ the one which maximizes $l_{(x_1,\ldots,x_m)}(\hat{\lambda})$ and therefore the one which solves $\nabla l_{(x_1,\ldots,x_m)}(\hat{\lambda}) = 0$. We get

$$\hat{\lambda}_k = \frac{1}{\alpha_k^m} \quad (k = 1, \ldots, m).$$

An estimator of λ may be obtained by smoothing $\hat{\lambda}$. Finally, the corresponding estimator for Λ is $\hat{\Lambda}$ defined by

$$\hat{\Lambda}(t) = \hat{\lambda}([0,t]) = \sum_{k=1}^{m} \frac{1}{\alpha_k^m} 1(x_k \leqslant t).$$

Using the standard relation between the reliability function and the cumulative hazard function, we get $\hat{R}$, the product limit estimator of R (see, e.g., Andersen et al., 1993)

$$\hat{R}(t) = \prod_{\{1 \leqslant i \leqslant n; x_i \leqslant t\}} \frac{\alpha_i^m - 1}{\alpha_i^m}.$$

Remark 2. If $r_i = 0$ for $1 \leqslant i \leqslant m$, then, $m = n$ and it is easy to see that

$$\hat{R}(t) = 1 - \hat{F}(t), \quad \text{where } \hat{F}(t) = \sum_{\{1 \leqslant i \leqslant n; x_i \leqslant t\}} 1/n,$$

i.e. $\hat{F}$ is the standard empirical distribution function for a full sample.

Remark 3. If $(X_{i:m:n})_{1 \leqslant i \leqslant m}$ is a p.c.s. from an arbitrary continuous distribution function F on $\mathbb{R}$, then, for $X_{i:m:n}^+ = \exp(X_{i:m:n})$ we have that $(X_{i:m:n}^+)_{1 \leqslant i \leqslant m}$ is a p.c.s. from distribution function $F^+ = F \circ \log$. It results that if $\widehat{R^+}$ is the reliability estimator of $R^+ = 1 - F^+$, then a natural estimator for F is given by

$$\hat{F}(x) = 1 - \widehat{R^+}(\exp(x)),$$

for $x \in \mathbb{R}$.

3. Some basic results

3.1. Simulation algorithm and a Markov property

Balakrishnan and Aggarwala (2000, p. 34) proposed (e.g.) the following algorithm to simulate a type II p.c.s. of size m, from a distribution function F:

1. Simulate m independent and identically distributed (i.i.d.) exponential random variables $Z_1, \ldots, Z_m$ with mean 1;

2. Set for $i = 1, \ldots, m$

$$Y_{i:m:n} = \sum_{j=1}^{i} Z_j/\alpha_j^m. \tag{2}$$

   Then $(Y_{i:m:n})_{i=1,\ldots,m}$ is a p.c.s. from an exponential distribution function with mean 1;
3. Set for $i = 1, \ldots, m$

$$X_{i:m:n} = F^{-1}(1 - \exp(-Y_{i:m:n})) = \Lambda(Y_{i:m:n}),$$

   where F^{-1} is taken in the generalized inverse sense, and Λ is the cumulative hazard rate function.

Remark 4. Steps 2 and 3 in the above algorithm will be of particular interest in the sense that if $(X_{i:m:n})_{i=1,\ldots,m}$ is a p.c.s. from distribution function F, $(\tilde{X}_{i:m:n})_{i=1,\ldots,m}$ where $\tilde{X}_{i:m:n} = \Lambda(X_{i:m:n})$ is a p.c.s. from an exponential distribution function with mean 1.

Another important result is the following.

Proposition 1 (Markov property). *Let $(X_{i:m:n})_{i=1,\ldots,m}$ a p.c.s. with underlying distribution function F and density f. Given $X_{1:m:n} = x_1, \ldots, X_{i:m:n} = x_i$, the random variables $X_{i+1:m:n}, \ldots, X_{m:m:n}$ are jointly distributed as a p.c.s. of size α_{i+1}^m with density function g defined by*

$$g(x) = \frac{f(x)}{1 - F(x_i)} 1(x \geqslant x_i).$$

Proof. See Balakrishnan and Aggarwala (2000, Theorems 2.4 and 2.5, pp. 14–15).

3.2. A martingale approach

Let N be the counting process defined by

$$N(t) = \sum_{i=1}^{m} 1(X_{i:m:n} \leqslant t),$$

and $\mathscr{F}^N = (\mathscr{F}_t^N)_{t \geqslant 0}$ be the natural filtration generated by N and $(r_n)_{n \geqslant 1}$, thus $\mathscr{F}_t^N = \sigma\{X_{i:m:n}, r_i;\ X_{i:m:n} \leqslant t\}$.

Proposition 2. *The process M defined on $[0, +\infty)$ by*

$$M(t) = N(t) - \int_0^t Y(s)\lambda(s)\,\mathrm{d}s, \tag{3}$$

where

$$Y(s) = \sum_{i=1}^{m} (r_i + 1)1(X_{i:m:n} \geqslant s),$$

is a martingale with respect to the filtration $\mathscr{F}^N$.

Proof. Define G_k and g_k by

$$G_k(t)=P(X_{k:m:n}-X_{k-1:m:n}\leqslant t|\mathscr{F}^N_{X_{k-1:m:n}})=\int_0^t g_k(s)\,\mathrm{d}s,\qquad t\geqslant 0,$$

where $\mathscr{F}^N_{X_{k-1:m:n}}$ is the filtration stopped at time $X_{k-1:m:n}$ (we put $X_{0:m:n}\equiv 0$). By Theorem 11 of Aven and Jensen (1999, p. 55), the intensity of N is given by

$$\tilde{\lambda}(t)=\sum_{k=1}^m \frac{g_k(t-X_{k-1:m:n})}{\int_{t-X_{k-1:m:n}}^{+\infty} g_k(s)\,\mathrm{d}s}1(X_{k-1:m:n}<t\leqslant X_{k:m:n}).$$

Denote by $\stackrel{\mathscr{L}}{=}$ the equality in law of two random variables. From Proposition 1, given $X_{k-1:m:n}=x$, we have

$$X_{k:m:n}\stackrel{\mathscr{L}}{=}\min_{1\leqslant j\leqslant \alpha_k^m} Y_j,$$

where the Y_j's are α_k^m i.i.d. random variables with density function g defined by

$$g(t)=\frac{f(t)}{R(x)}1(t\geqslant x).$$

Then, noting $x=X_{k-1:m:n}$, we have

$$\begin{aligned}1-G_k(t)&=P(X_{k:m:n}-X_{k-1:m:n}>t|\mathscr{F}^N_{X_{k-1:m:n}})\\&=P\left(\min_{1\leqslant j\leqslant \alpha_k^m} Y_j-x\geqslant t\right)=\left(\int_{t+x}^{+\infty} g(s)\,\mathrm{d}s\right)^{\alpha_k^m}=\left(\frac{R(x+t)}{R(x)}\right)^{\alpha_k^m}.\end{aligned}$$

We have for $t\geqslant 0$

$$g_k(t)=\alpha_k^m f(x+t)\frac{R(x+t)^{\alpha_k^m-1}}{R(x)^{\alpha_k^m}},$$

then, for $X_{k-1:m:n}<t\leqslant X_{k:m:n}$, we have

$$\frac{g_k(t-X_{k-1:m:n})}{\int_{t-X_{k-1:m:n}}^{+\infty} g_k(u)\,\mathrm{d}u}=\alpha_k^m\lambda(t).$$

It is now easy to show that if Y denotes the process

$$Y(t)=\sum_{i=1}^m (r_i+1)1(X_{i:m:n}\geqslant t),$$

then, for $t\geqslant 0$, we have

$$\tilde{\lambda}(t)=Y(t)\lambda(t).$$

Consequently the process $M(t)=N(t)-\int_0^t Y(s)\lambda(s)\,\mathrm{d}s$ is an $\mathscr{F}^N$-martingale. □

Remark 5. Proposition 2 makes it possible to retrieve the previous cumulative hazard estimator. Using the following classical heuristic

$$\mathrm{d}N(t)\approx Y(t)\lambda(t)\,\mathrm{d}t$$

we get by neglecting the martingale part in (3), an estimator $\hat{\Lambda}$ of the cumulative hazard function Λ

$$\hat{\Lambda}(t) = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)}, \quad t \geqslant 0,$$

which is the same as $\hat{\Lambda}$ of Section 2.

It is also clear that the properties of the estimators will strongly depend on properties of processes N and Y or their normalized versions $N^{(m)} = N/m$ and $Y^{(m)} = Y/m$.

4. Asymptotics for the estimators

The main purpose of this section is to study the asymptotic behavior of the processes $\hat{\Lambda}^{(m)}$ and $\hat{R}^{(m)}$. This is achieved by using the now classical martingale and counting processes approach developed in Andersen et al. (1993). These methods require some knowledge about uniform consistency of the processes $N^{(m)}$ and $Y^{(m)}$. The next subsection contains preliminary results on $N^{(m)}$ and $Y^{(m)}$. This part enables us to understand that under the assumptions below, a p.c.s. with distribution function F behaves asymptotically like the usual order statistic of a m-sample with distribution function $1-(1-F)^{r+1}$ (r is given in A2 below; note that the result holds for finite m when the r_i's are constant equal to r).

Let us consider the following assumptions. In the sequel, all limits are taken with respect to m tending to infinity.

A1. $\sup_{m\geqslant 1} r_m \leqslant K < +\infty$;
A2. $\sum_{i=1}^m r_i/m \to r$;
A3. τ is a real number such that $F(\tau) < 1$.

Remark 6. In the sequel we assume that the sequence $(r_i)_{i\geqslant}$ is deterministic. However, it could be a random sequence (this is the case in our simulation results) and then, results involving $(r_i)_{i\geqslant 1}$ should be understood in the almost sure sense. This is also the reason why the r_i's are included in the filtration of the previous section.

4.1. About the processes $N^{(m)}$ and $Y^{(m)}$

Lemma 1. *Let* $(r_m)_{m\geqslant 1}$ *satisfy* A1–A2 *and let* α *be a real such that* $0<\alpha<1$. *Then*

$$\sup_{1\leqslant j\leqslant [\alpha m]} \left| \frac{\alpha_j^m}{(r+1)(m-j+1)} - 1 \right| \to 0.$$

Proof. It is sufficient to show that

$$\sup_{0\leqslant j\leqslant [\alpha m]} \left| \frac{1}{m-j} \sum_{i=j+1}^m r_i - r \right| \to 0.$$

Let $\varepsilon > 0$ be a real number and put $u_m = m^{-1}\sum_{i=1}^m r_i - r$. Then there exist:

(i) m_0 such that if $m, j \geqslant m_0$, then $|u_m - u_j| < \frac{\varepsilon(1-\alpha)}{2\alpha}$ (Cauchy);
(ii) m_1 such that if $m \geqslant m_1$, then $|u_m| < \varepsilon/2$ (A2);
(iii) m_2 such that if $m \geqslant m_2$, then $j/(m-j) \leqslant \varepsilon/4(K+1)$ if $0 \leqslant j \leqslant m_0$.

Let us define

$$v_{j,m} = \frac{1}{m-j}\sum_{i=j+1}^{m} r_i - r,$$

we have

$$v_{j,m} = u_m + \frac{j}{m-j}(u_m - u_j).$$

Then, for $m \geqslant \max(m_0, m_1, m_2)$, using (i)–(iii), we have:

- if $0 \leqslant j \leqslant m_0$:

$$|v_{j,m}| \leqslant \varepsilon/2 + \left(\frac{j}{m-j}\right) 2(K+1) \leqslant \varepsilon,$$

- if $m_0 < j \leqslant [\alpha m]$:

$$|v_{j,m}| \leqslant \varepsilon/2 + \frac{[\alpha m]}{m - [\alpha m]}\,\frac{\varepsilon(1-\alpha)}{2\alpha} \leqslant \varepsilon.$$

The lemma is proved. □

Let us now remark that by inverting steps 2 and 3 of the algorithm of Section 3.1 we have the following almost sure representation result:

$$Y_{i:m:n} = \Lambda(X_{i:m:n}) = \sum_{j=1}^{i} \frac{Z_j}{\alpha_j^m}, \quad i = 1, \ldots, m,$$

where the Z_j's are i.i.d. exponentially distributed random variables with mean 1. Introducing the random variables $(\tilde{Y}_{i:m:n})_{1 \leqslant i \leqslant m}$, where for $1 \leqslant i \leqslant m$

$$\tilde{Y}_{i:m:n} = \sum_{j=1}^{i} \frac{Z_j}{(r+1)(m-j+1)},$$

we get the following lemma.

Lemma 2. *Under* A1–A3, *for* $0 \leqslant \alpha < 1$, *we have*

$$\sup_{1 \leqslant i \leqslant [\alpha m]} |\tilde{Y}_{i:m:n} - Y_{i:m:n}| \to 0, \quad a.s.$$

Proof. Note that for $1 \leqslant i \leqslant [\alpha m]$, we have

$$|\tilde{Y}_{i:m:n} - Y_{i:m:n}| \leqslant \sum_{j=1}^{i} \left| \frac{1}{(r+1)(m-j+1)} - \frac{1}{\alpha_j^m} \right| Z_j$$

$$\leqslant \sup_{1 \leqslant j \leqslant [\alpha m]} \frac{1}{\alpha_j^m} \left| \frac{\alpha_j^m}{(r+1)(m-j+1)} - 1 \right| \sum_{k=1}^{[\alpha m]} Z_k$$

$$\leqslant \sup_{1 \leqslant j \leqslant [\alpha m]} \left| \frac{\alpha_j^m}{(r+1)(m-j+1)} - 1 \right|$$

$$\times \frac{[\alpha m]}{(m - [\alpha m])} \times \frac{1}{[\alpha m]} \sum_{k=1}^{[\alpha m]} Z_k.$$

The above inequality together with

- $\sup_{1 \leqslant j \leqslant [\alpha m]} \left| \frac{\alpha_j^m}{(r+1)(m-j+1)} - 1 \right| \to 0$ for all $0 \leqslant \alpha < 1$ from Lemma 1,
- $\frac{[\alpha m]}{(m - [\alpha m])} \leqslant \frac{\alpha}{(1-\alpha)} < +\infty$,
- $\frac{1}{[\alpha m]} \sum_{k=1}^{[\alpha m]} Z_k \to 1$ a.s. from the strong law of large numbers,

give the expected convergence result

$$\sup_{1 \leqslant i \leqslant [\alpha m]} |\tilde{Y}_{i:m:n} - Y_{i:m:n}| \to 0, \quad \text{a.s.} \qquad \square$$

Lemma 3. *Under* A1 *and* A3 *we have* $\lim_{m \to +\infty} P(X_{m:m:n} > \tau) = 1$.

Proof. By A3 we have $\Lambda(\tau) < +\infty$ and by A1

$$P(X_{m:m:n} > \tau) = P(Y_{m:m:n} > \Lambda(\tau)) = P\left(\sum_{j=1}^{m} \frac{Z_j}{\alpha_j^m} > \Lambda(\tau) \right)$$

$$\geqslant P\left(\sum_{j=1}^{m} \frac{Z_j}{m-j+1} > (K+1)\Lambda(\tau) \right) = P(Z_{m:m} > (K+1)\Lambda(\tau))$$

$$= 1 - (1 - \exp(-(K+1)\Lambda(\tau)))^m = 1 - F^{m(K+1)}(\tau) \to 1,$$

where $Z_{m:m} = \max_{1 \leqslant i \leqslant m} Z_i$. $\square$

Proposition 3. *Under* A1–A3 *we have*

$$\sup_{0\leqslant t\leqslant\tau} |N^{(m)}(t)-(1-R^{r+1}(t))| \to 0, \quad a.s., \tag{4}$$

$$\sup_{0\leqslant t\leqslant\tau} |Y^{(m)}(t)-(r+1)R^{r+1}(t)| \to 0, \quad a.s. \tag{5}$$

Proof. Let us show (4). We have $N^{(m)}(t)-(1-R^{r+1}(t))=I^{(m)}(t)+II^{(m)}(t)$, where

$$I^{(m)}(t)=\frac{1}{m}\sum_{i=1}^{m}(1(Y_{i:m:n}\leqslant\Lambda(t))-1(\tilde{Y}_{i:m:n}\leqslant\Lambda(t)))$$

and

$$II^{(m)}(t)=\frac{1}{m}\sum_{i=1}^{m}1(\tilde{Y}_{i:m:n}\leqslant\Lambda(t))-(1-R^{r+1}(t)).$$

Using the fact that the $\tilde{Y}'_{i:m:n}s$ have the same joint distribution as an order statistic of a m-sample exponentially distributed with mean $1/(r+1)$, it results that $II^{(m)}(t)$ is the centered empirical distribution function of a sample of m exponential random variables with mean $1/(r+1)$. Then the Glivenko–Cantelli theorem (see e.g. Shorack and Wellner, 1986) gives

$$\sup_{0\leqslant t\leqslant\Lambda^{-1}(\tau)} |II^{(m)}(t)| \to 0, \quad \text{a.s.}$$

Let ε be a real number in $(0,1)$. We have

$$I^{(m)}(t)\leqslant\left|\frac{1}{m}\sum_{i=1}^{[(1-\varepsilon)m]}(1(Y_{i:m:n}\leqslant\Lambda(t))-1(\tilde{Y}_{i:m:n}\leqslant\Lambda(t)))\right|+2\varepsilon=III^{(m)}(t)+2\varepsilon.$$

Moreover, for $\alpha=\varepsilon/(2(r+1))$ we have, for all $1\leqslant i\leqslant[(1-\varepsilon)m]$:

$$\begin{aligned}&|1(Y_{i:m:n}\leqslant\Lambda(t))-1(\tilde{Y}_{i:m:n}\leqslant\Lambda(t))|\\ &\quad\leqslant 1(|\tilde{Y}_{i:m:n}-\Lambda(t)|\leqslant\alpha)+1(|\tilde{Y}_{i:m:n}-Y_{i:m:n}|>\alpha),\end{aligned}$$

and then

$$III^{(m)}(t)\leqslant\frac{1}{m}\sum_{i=1}^{m}1(|\tilde{Y}_{i:m:n}-\Lambda(t)|\leqslant\alpha)+1\left(\sup_{1\leqslant i\leqslant[(1-\varepsilon)m]}|\tilde{Y}_{i:m:n}-Y_{i:m:n}|>\alpha\right).$$

By applying Lemma 2 and using again the strong law of large numbers, the right-hand side of the above inequality converges almost surely to

$$\exp(-(r+1)(\Lambda(t)-\alpha))-\exp(-(r+1)(\Lambda(t)+\alpha))\leqslant\varepsilon$$

and the convergence is uniform for $t\in[0,\tau]$. Result (4) is proved.

It remains to show (5). Note that

$$Y^{(m)}(t)=\frac{1}{m}\sum_{i=1}^{m}(r_i-r)+(r+1)(1-N^{(m)}(t))+\frac{1}{m}\sum_{i=1}^{m}(r_i-r)1(X_{i:m:n}\leqslant t).$$

From A3 and (i), respectively, the first and second term on the right hand side of the above equality converge, respectively, to 0 and $(r+1)R^{r+1}(t)$ (a.s. and uniformly in $t \in [0,\tau]$). Let $\varepsilon > 0$ be a real number and choose $\theta \in (0,\tau)$, such that $(K+r)(1-R^{r+1}(\theta)) \leqslant \varepsilon$. For the third term, denoted $IV^{(m)}(t)$, observe that

$$|IV^{(m)}(t)| \leqslant (K+r)N^{(m)}(\theta) + \sup_{\theta \leqslant t \leqslant \tau} \left| \frac{1}{N(t)} \sum_{i=1}^{N(t)} (r_i - r) \right|.$$

From (i), on a set of probability 1, we have $\lim_{m\to+\infty} (K+r)N^{(m)}(\theta) \leqslant \varepsilon$ and $N(\theta) \to +\infty$, and then, on the same set, we have

$$\lim_{m\to+\infty} \sup_{0\leqslant t\leqslant \tau} |IV^{(m)}(t)| \leqslant \varepsilon.$$

Since ε is arbitrary we get

$$\sup_{0\leqslant t\leqslant\tau} |IV^{(m)}(t)| \to 0, \quad \text{a.s.},$$

which achieves the proof of (5). $\square$

4.2. Asymptotics

We now present the main results concerning the asymptotic behavior of the estimators.

Theorem 1 (weak consistency). *Suppose that* A1–A3 *are satisfied. Then*

(i) $\sup_{0\leqslant s\leqslant\tau} |\hat{\Lambda}(s) - \Lambda(s)| \xrightarrow{P} 0,$

(ii) $\sup_{0\leqslant s\leqslant\tau} |\hat{R}(s) - R(s)| \xrightarrow{P} 0.$

Proof.

(i) Firstly note that for $t \in [0,\tau]$ we have

$$\hat{\Lambda}(t) - \Lambda(t) = \int_0^t \frac{J(s)}{Y(s)} M(\mathrm{d}s) + \int_0^t (1-J(s))\lambda(s)\,\mathrm{d}s,$$

where $J(t) = 1(Y(t) > 0)$. Following the lines of Andersen et al. (1993, p. 190) we get, with both the Lenglart inequality and Proposition 3, for $\varepsilon > 0$,

$$\lim_{m\to+\infty} P\left(\sup_{t\in[0,\tau]} \left| \int_0^t \frac{J(s)}{Y(s)} M(\mathrm{d}s) \right| > \varepsilon \right) = 0.$$

On the other hand, we have

$$\sup_{t\in[0,\tau]} \left| \int_0^t (1-J(s))\lambda(s)\,\mathrm{d}s \right| \leqslant \Lambda(\tau) - \Lambda(\tau \wedge X_{m:m:n}) \xrightarrow{P} 0, \qquad (6)$$

from Lemma 3. So we have proved that

$$\sup_{0\leqslant s\leqslant\tau} \left|\hat{\Lambda}(s) - \Lambda(s)\right| \xrightarrow{P} 0.$$

(ii) Denote $R^*(t)=\int_0^t J(s)\lambda(s)\,\mathrm{d}s$. By the Duhamel equation (see Andersen et al., 1993, pp. 90–91) we have

$$\frac{\hat{R}(t)}{R^*(t)}-1=-\int_0^t \frac{\hat{R}(s-)J(s)}{R^*(s)Y(s)}M(\mathrm{d}s)$$

which converges uniformly to 0 in probability on $[0,\tau]$ by the Lenglart inequality and Proposition 3. Furthermore, it is easy to see that

$$\left|\frac{R(t)}{R^*(t)}-1\right|\leqslant (R(\tau))^{-1}\int_0^\tau (1-J(s))\lambda(s)\,\mathrm{d}s.$$

Hence from (6) we get

$$\sup_{t\in[0,\tau]}\left|\frac{R(t)}{R^*(t)}-1\right|\xrightarrow{P}0,$$

which achieves the proof of (ii). $\square$

Theorem 2 (weak convergence). *Suppose that* A1–A3 *hold and let B be the standard Brownian Motion on* $[0,+\infty)$. *Then we have*

(i)

$$\sqrt{m}(\hat{\Lambda}(t)-\Lambda(t))\xrightarrow{\mathscr{D}}B\circ v(t)\quad \textit{in } D[0,\tau],$$

where the covariance function v is defined by

$$\operatorname{cov}(B\circ v(s),B\circ v(t))=v(s\wedge t)=\frac{1-R^{r+1}(s\wedge t)}{(r+1)^2R^{r+1}(s\wedge t)},$$

for $s,t\in[0,\tau]$.

(ii) *In addition, v may be consistently (uniformly) estimated on* $[0,\tau]$ *by*

$$\hat{v}(t)=m\int_0^t \frac{\mathrm{d}N(s)}{Y^2(s)}.$$

(iii)

$$\sqrt{m}(\hat{R}(t)-R(t))\xrightarrow{\mathscr{D}}R(t)B\circ v(t)\quad \textit{in } D[0,\tau],$$

with $\operatorname{cov}(R(t)B\circ v(t),R(s)B\circ v(s))=R(t)R(s)v(s\wedge t)$ *for* $s,t\in[0,\tau]$.

(iv)

$$\sup_{0\leqslant s\leqslant\tau}\left(\frac{m}{\hat{v}(s)}\right)^{1/2}\frac{|\hat{R}(s)-R(s)|}{\hat{R}(s)}\xrightarrow{\mathscr{D}}\sup_{0\leqslant s\leqslant 1}|B(s)|.$$

Remark 7. Results of Theorems 1 and 2 ((i), (ii) and (iii)) allow us to construct pointwise confidence intervals for Λ and R. By result (iv) in Theorem 2 we can construct Gill type (see, e.g., Fleming and Harrington, 1991, p. 240) confidence bands. However, better confidence band results should be obtained by using the method of Hall and Wellner (1980).

Proof. (i) Using the decomposition

$$\sqrt{m}(\hat{\Lambda}(t)-\Lambda(t))=\sqrt{m}\int_0^t \frac{J(s)}{Y(s)}M(\mathrm{d}s)+\sqrt{m}\int_0^t (1-J(s))\lambda(s)\,\mathrm{d}s$$

we get for $\varepsilon>0$

$$P\left(\sup_{t\in[0,\tau]}\sqrt{m}\left|\int_0^t (1-J(s))\lambda(s)\,\mathrm{d}s\right|>\varepsilon\right)\leqslant P(X_{m:m:n}<\tau)\to 0,$$

from Lemma 3. It follows that the processes

$$\sqrt{m}(\hat{\Lambda}(t)-\Lambda(t)) \quad \text{and} \quad \sqrt{m}\int_0^t \frac{J(s)}{Y(s)}M(\mathrm{d}s),$$

are asymptotically equivalent on $[0,\tau]$. Using Proposition 3 and following the lines of Andersen et al. (1993, p. 191), we get by Rebolledo's theorem that

$$\sqrt{m}\int_0^t \frac{J(s)}{Y(s)}M(\mathrm{d}s)\xrightarrow{\mathscr{D}} B\circ v(t) \quad \text{in } D[0,\tau],$$

where B is a standard Brownian Motion on $[0,+\infty)$ and the covariance function v is defined by

$$\operatorname{cov}(B\circ v(s),B\circ v(t))=v(s\wedge t)=\int_0^{s\wedge t}\frac{\lambda(u)}{(r+1)R^{r+1}(u)}\,\mathrm{d}u$$
$$=-\frac{1-R^{-(r+1)}(s\wedge t)}{(r+1)^2},$$

for $s,t\in[0,\tau]$, if:

(a) $\langle\sqrt{m}\int_0^{\cdot}\frac{J(s)}{Y(s)}M(\mathrm{d}s)\rangle(t)\xrightarrow{P} v(t)$, for all $t\in[0,\tau]$, and

(b) for all $\varepsilon>0$, $\langle\sqrt{m}\int_0^{\cdot}\frac{J(s)}{Y(s)}1(\frac{J(s)\sqrt{m}}{Y(s)}>\varepsilon)M(\mathrm{d}s)\rangle(t)\xrightarrow{P}0$, for all $t\in[0,\tau]$.

Condition (a) is satisfied since we have

$$\left\langle\sqrt{m}\int_0^{\cdot}\frac{J(s)}{Y(s)}M(\mathrm{d}s)\right\rangle(t)=\int_0^t\frac{J(s)\lambda(s)}{Y^{(m)}(s)}\,\mathrm{d}s,$$

and the strong uniform convergence of $Y^{(m)}$ towards $(r+1)R^{r+1}$, which is uniformly bounded away from 0 on $[0,\tau]$ (see Proposition 3). Moreover, for $\eta>0$ we have

$$P\left(\left\langle\sqrt{m}\int_0^{\cdot}\frac{J(s)}{Y(s)}1\left(\frac{J(s)\sqrt{m}}{Y(s)}>\varepsilon\right)M(\mathrm{d}s)\right\rangle(t)>\eta\right)$$
$$=P\left(\int_0^t\frac{J(s)}{Y^{(m)}(s)}1\left(\frac{J(s)}{\sqrt{m}\varepsilon}>Y^{(m)}(s)\right)\lambda(s)\,\mathrm{d}s>\eta\right)$$
$$\leqslant P(Y^{(m)}(\tau)<m^{-1/2}\varepsilon^{-1})\to 0,$$

by Proposition 3.

(ii) Write

$$\hat{v}(t)=\frac{1}{m}\int_0^t J(s)\left(\frac{m}{Y(s)}\right)^2 M(\mathrm{d}s)+\int_0^t J(s)\frac{m}{Y(s)}\lambda(s)\,\mathrm{d}s.$$

Consistency of $\hat{v}$ follows from an application of Lenglart's inequality (see e.g. Andersen et al., 1993) and our Proposition 3, to the first term on the right-hand side of the above equation (then we get its uniform convergence in probability to 0) and we show, using once again Proposition 3, that the second term converges uniformly to v on $[0,\tau]$.

(iii) The weak convergence of $\sqrt{m}(\hat{R}-R)$ in $D[0,\tau]$ follows immediately from compact differentiability of product integral (see Andersen et al., 1993, Proposition II.8.7), the functional δ-method (see Andersen et al., 1993, Theorem II.8.1) and the weak convergence of $\sqrt{m}(\hat{\Lambda}-\Lambda)$ in $D[0,\tau]$ obtained in (i) of this theorem.

(iv) Finally, from (iii) and Slutsky's Lemma we have the following weak convergence result

$$\left(\frac{m}{\hat{v}(s)}\right)^{1/2}\frac{\hat{R}(s)-R(s)}{\hat{R}(s)}\xrightarrow{\mathscr{D}} B(s).$$

By continuity of $x\to\sup_{s\in[0,\tau]}|x(s)|$ on the space of continuous functions and the continuous mapping theorem (see e.g. Shorack and Wellner, 1986), we get the expected result. □

Remark 8. Let $\hat{F}$ be the estimator of an arbitrary distribution function F (with density f) defined in Remark 3. Suppose that A1–A2 are satisfied and that τ is a real number such that $F(\tau)<1$. Let B be the standard Brownian Motion on $[0,+\infty)$. Then we have the uniform weak consistency of $\hat{F}$ on $(-\infty,\tau]$, $\sqrt{m}(\hat{F}(t)-F(t))\xrightarrow{\mathscr{D}}(1-F(t))B\circ v(t)$ in $D(-\infty,\tau]$, where

$$v(t)=\frac{1-(1-F(t))^{r+1}}{(r+1)^2(1-F(t))^{r+1}},$$

and v may be consistently estimated on $(-\infty,\tau]$ by

$$\hat{v}(t)=m\int_0^{\exp(t)}\frac{N^+(\mathrm{d}s)}{(Y^+(s))^2},$$

where $N^+(s)=\sum_{i=1}^m 1(X_{i:m:n}\leqslant\log(s))$ and $Y^+(s)=\sum_{i=1}^m (r_i+1)1(X^+_{i:m:n}\geqslant\log(s))$.

5. Example and Monte-Carlo simulations

5.1. An example

Here we compare our non-parametric estimator with two parametric ones, proposed by Balakrishnan and Aggarwala (2000). The comparison is done on a data set extracted

Table 1
Progressively censored sample generated from times to break-down data on insulating fluid tested at 34 kV by Nelson (1982)

i	1	2	3	4	5	6	7	8
$x_{i:8:19}$	0.19	0.78	0.96	1.31	2.78	4.85	6.5	7.35
r_i	0	0	3	0	3	0	0	5

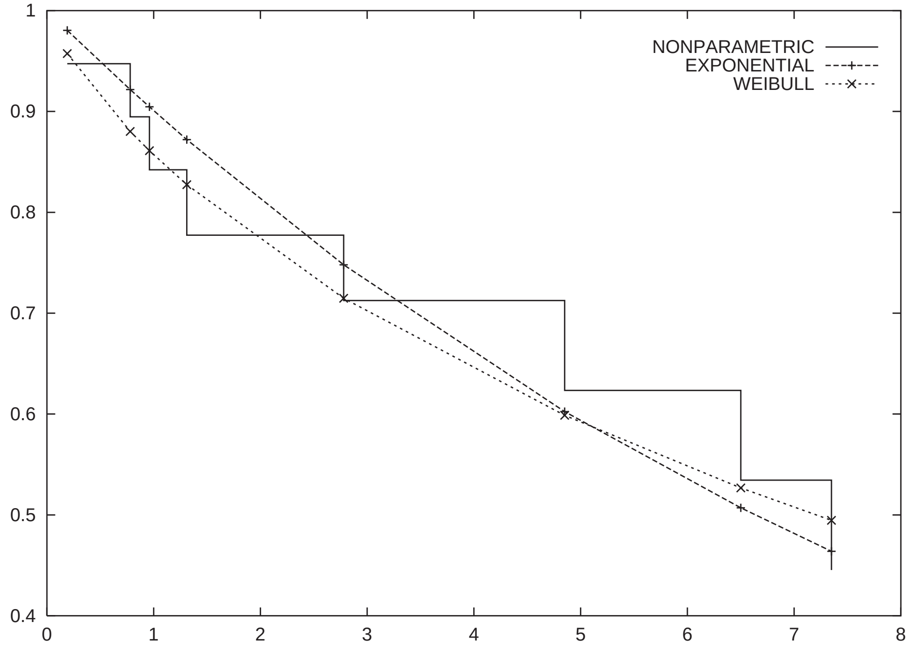

Fig. 1. Three estimators of the reliability function.

(see Table 1) by Viveros and Balakrishnan (1994) from a sample by Nelson (1982, p. 105, Table 1.1: 3 values between 2.78 and 4.85 and 8 values greater than 7.35 are omitted which gives the total number of 11 progressively censored items).

Fig. 1 gives a comparison of our *non-parametric* estimator with both an *exponential* estimator (see Balakrishnan and Aggarwala, 2000, p. 95) and a *Weibull* estimator (see Balakrishnan and Aggarwala, 2000, p. 102; this estimator is denoted Weibull, but estimation is done via an extreme value estimation based on a log-time transformation). It is interesting to note that our estimator can help to *choose* between the two parametric estimators. For example we note that the non-parametric estimator seems closer to the Weibull estimator than to the exponential one.

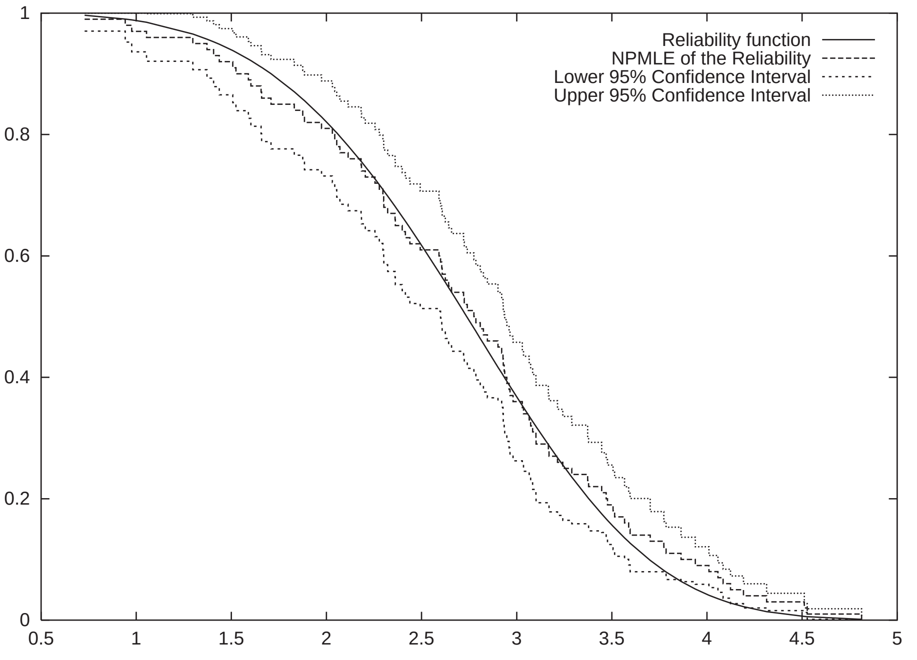

Fig. 2. Uncensored data, $m = 100$.

5.2. Monte-Carlo simulations

The following Monte-Carlo simulations have been obtained for two kinds of data:

(i) *uncensored data* (u.d.): $r_1 = \cdots = r_m = 0$;
(ii) *progressively censored data* (p.c.d.): the r_i's are i.i.d. with $P(r_i = 0) = \cdots = P(r_i = 4) = 0.2$, then $E(r_i) = 2$ and since $n = r_1 + \cdots + r_m + m$, we have $E(n) = 3m$.

We simulated lifetimes with Weibull distribution function F defined by

$$F(t) = (1 - \exp(-(t/3)^4))1(t \geqslant 0).$$

Each of the four figures contains:

(i) $R(t) = 1 - F(t)$ the true reliability function;
(ii) $\hat{R}(t)$ for m u.d. or p.c.d;
(iii) 95% pointwise confidence interval for the reliability.

The comparison of Fig. 3 with Figs. 2 and 4 shows the poor performance of the NPMLE for the reliability function under progressive censoring in estimating small reliability values. This result is not surprising since this sequential censoring scheme makes the probability of removing large values in the initial sample larger than that of removing small ones.

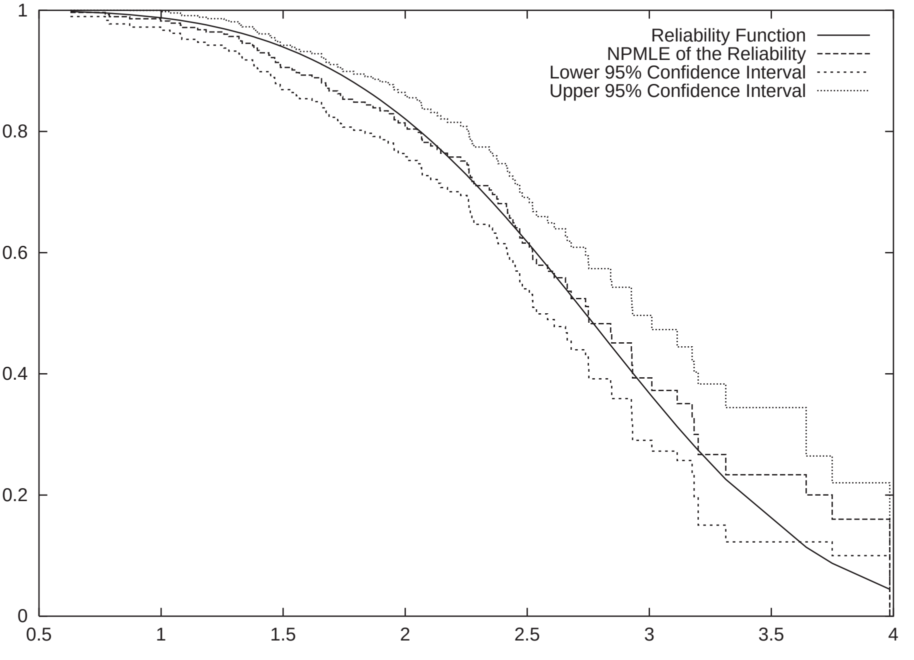

Fig. 3. Progressively censored data, $m = 100$, $E(n) = 3m$.

Otherwise, it is interesting to note that estimation (from a confidence interval point of view) works well for moderate to high reliability values. For such reliability values, we note that performances of the NPMLE under progressive censoring (see Fig. 3) are closer to those of the NPMLE for 300 uncensored data (see Fig. 4) than those of the NPMLE for 100 uncensored data (see Fig. 2).

6. Concluding remarks

We proposed a non-parametric estimator for the reliability function. In fact, it is a Kaplan–Meier type estimator, in the sense that a p.c.s may be viewed as a sample of size n in which m failures are observed whereas other observations are exactly censored by the times-to-failure $(X_{i:m:n})_{i=1,\dots,m}$. However, this fact makes the study of our estimators non obvious in the sense that the classical assumption of independence between times-to-failure and censoring times is not true here.

Another issue of this work could be another way to bootstrap an empirical distribution function: obtain N sub-p.c.s from an initial sample by choosing randomly several sequences $(r_i^{(j)})_{i=1,\dots,m_j}^{j=1,\dots,N}$ (satisfying A1–A2), for each sequence $(r_i^{(j)})_{i=1,\dots,m_j}$ define the corresponding estimator $\hat{R}^{(j)}$, then estimate F by

$$\hat{F}(t) = 1 - \frac{1}{N}\sum_{j=1}^{N} \hat{R}^{(j)}(t).$$

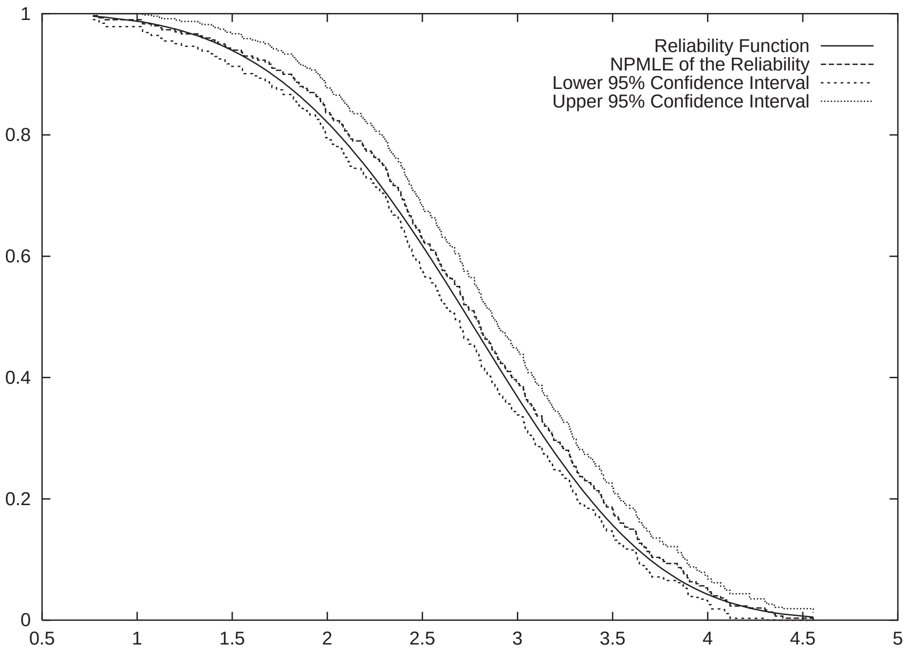

Fig. 4. Uncensored data, $m = 300$.

Finally, note that from Proposition 3 there is another "natural" estimator of the distribution function F. Indeed, from this proposition we have

$$\lim_{n\to+\infty} EN^{(m)}(t) = 1 - (1-F)^{r+1}(t).$$

Then, defining $\hat{r} = \sum_{i=1}^{m} r_i/m$, a consistent (under A1–A3) estimator of F is given by

$$\tilde{F}(t) = 1 - (1 - N^{(m)}(t))^{1/(\hat{r}+1)}.$$

Acknowledgements

Thanks to Sergio Alvarez with whom I had many constructive discussions on progressive censoring. Many thanks to an anonymous Referee who gave the interpretation at the beginning of Section 4 and suggested to improve the proofs of Section 4.1 by showing Lemma 2.

References

Andersen, P.K., Borgan, O., Gill, R., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, New York.

Aven, T., Jensen, U., 1999. Stochastic Models in Reliability. Springer, Berlin.

Balakrishnan, N., Aggarwala, R., 2000. Progressive Censoring. Birkhäuser, Basel.

Balasoorya, U., Saw, S.L.C., Gadag, V., 2000. Progressively censored reliability sampling plans for the Weibull distribution. Technometrics 42, 160–167.

Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.

Hall, W.J., Wellner, J.A., 1980. Confidence bands for a survival curve from censored data. Biometrika 67, 133–143.

Kamps, U., 1995. A concept of generalized order statistics. J. Statist. Plann. Inference 48, 1–23.

Montanari, G.C., Cacciari, M., 1988. Progressively-censored aging tests on XLPE-insulated cable models. IEEE Trans. Electrical Insul. 23, 365–372.

Nelson, W., 1982. Applied Life Data Analysis. Wiley, New York.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Viveros, R., Balakrishnan, N., 1994. Interval estimation of life characteristics from progressively censored data. Technometrics 36, 84–91.