Journal of Statistical Planning and Inference 53 (1996) 171–195

journal of statistical planning and inference

Nonparametric estimation and consistency for renewal processes[1]

Guoxing Soon*, Michael Woodroofe

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

Received 28 October 1994; revised 21 March 1995

Abstract

In reliability or medical studies, we may only observe each ongoing renewal process for a certain period of time. When the underlying distribution F is arithmetic, Vardi (*Ann. Statist.* **10** (1982b), 772–785) developed the RT algorithm for nonparametric estimation. In this paper we extend the study to the nonarithmetic case and show that the choice of an arbitrary constant in the RT algorithm can be avoided. We also prove the strong consistency of the maximum likelihood estimators of the mean of F and F restricted to the interval $[0, b)$, if the lengths of the observation periods converge to b.

AMS classification: primary 62G05; 62F12; secondary 62D05; 60F05; 62P10; 62P20; 62P25; 62P99

Keywords: Renewal processes; Modified RT algorithm; Window censored; NPMLE; Consistency

1. Introduction

Consider a machine which fails periodically and is instantly repaired, in such a manner that the failure times form a renewal process. Suppose further that the machine was placed in service in the indefinite past, so that the process may be regarded as stationary. Interest here centers on estimating the distribution of the time between failures and its mean, when several such processes are observed over a time interval, taken to be of the form $[a, a + b]$ with $b > 0$.

Let F denote the lifetime distribution, and suppose that

$$F(0) = 0, \quad 0 < \mu_F := \int_0^\infty xF(\mathrm{d}x) < \infty. \tag{1.1}$$

[1]Research supported by NSF under DMS-9203357.

*Corresponding author. E-mail: gxsoon@stat.lsa.umich.edu.

0378-3758/96/$15.00 © 1996—Elsevier Science B.V. All rights reserved

PII S0378-3758(95)00127-1

Suppose also that F is nonarithmetic, and let

$$G(x) = \int_0^x \frac{1 - F(y-)}{\mu_F}\, dy, \quad \forall\, 0 \leqslant x < \infty.$$

See Feller (1971, Ch. XI) for the derivation of G. Further, let $X_0, X_1, \ldots$ denote independent random variables for which $X_0 \sim G$ and $X_1, X_2, \ldots \sim F$; let $0 < b < \infty$; and let

$$S_n = X_0 + X_1 + \cdots + X_n, \quad n \geqslant 0,$$

$$\tau = \tau^b = \inf\{n \geqslant 0 \colon S_n > b\},$$

$$R = R^b = S_\tau - b,$$

$$Z = Z^b = b - S_{\tau-1} \quad \text{if } \tau \geqslant 1.$$

In the machine interpretation, τ is the number of failures that occur during the observation period $[a, a+b]$, R is the residual waiting time (the time from the beginning of the study until the first breakdown), and Z is the spent waiting time (the time from the last breakdown to the end of the study). Then the process is stationary in that R^b has distribution function G for all $b > 0$. If $\tau \geqslant 1$, then the data consist of the $X_0, X_1, \ldots, X_{\tau-1}$ and Z; and if $\tau = 0$, the data consist only of the observation that $X_0 > b$. The problem is to estimate F and μ_F from data from several, similar, independent processes.

Models of this nature were studied in the pioneering paper by Vardi (1982b), who constructed an algorithm for computing an M-restricted non-parametric maximum likelihood estimator (hereafter NPMLE) of F and μ_F. Actually, Vardi studied the arithmetic case in which the relation between F and G is slightly different. The objectives of this paper are to adapt Vardi's algorithm to the nonarithmetic case and to study the asymptotic properties of the resulting estimators. It is interesting that the mean μ_F may be estimated consistently, even if the interval of the observation $[0, b]$ does not include the entire support of F.

It is known that the NPMLE based on Y or Z (or together with W) has a convergence rate $n^{1/3}$, as shown by Groeneboom (1985). However, the information they contribute cannot be ignored even if the NPMLE based on the whole observed data has a convergence rate $n^{1/2}$, this is due to the dependence of these parts of the data. One such example, based on the ages of the machines and the observed residual lifetimes, which could be right censored, was studied by Vardi (1989) and Vardi and Zhang (1992) under a random-multiplicative censoring model, in which the NPMLE for lifetime distribution was derived and its asymptotic properties were studied. In particular, the convergence rate is shown to be $n^{1/2}$, which is of higher order than $n^{1/3}$, the convergence rate of the NPMLE based only on ages or residual lifes.

Estimation problems based on combined data or incomplete data have been studied by many authors. Besides the extensively studied right censoring models (Kaplan and Meier (1958) etc.), the doubly censored model was studied by Chang and

Yang (1987), Chang (1990), Gu and Zhang (1993), the truncated model by Woodroofe (1985) and Keiding and Gill (1990), the biased sampling model by Vardi (1982a, 1985) and Gill et al. (1988), among others.

Basic notations are defined in Section 2. Since the likelihood function for the observed data may not be maximized by any proper distribution function, some generalization or restriction has to be put on the class of distribution functions under consideration. The first approach is chosen in this paper whereas the second was the choice of Vardi (1982b). In Section 3 we study both approaches. We also obtain the uniqueness of the NPMLE when the window lengths are a fixed constant. This is important for the proof of the convergence of the RT algorithm.

In Section 4 Vardi's RT algorithm is modified to compute the NPMLE of F and μ_F. A careful analysis shows that the convergence properties of the RT algorithm are preserved under such modification. Section 5 illustrates the use of the modified RT algorithm by an example. The results are compared to the results in Vardi (1982b).

Sections 6 and 7 are devoted to prove the uniform consistency of the NPMLEs. Some limit theorems for renewal processes are proved in Section 5. In Section 6 we derive the score equations satisfied by the NPMLEs and then establish the uniform consistency of these estimators for fixed window lengths.

2. Notation

The following notations are defined by assuming F is nonarithmetic.

For a single renewal process write τ for τ^b; let $Y = X_0$ and $Z = b - S_{\tau-1}$, if $\tau \geqslant 1$; and let $W = b$, if $\tau = 0$. Then the data consist of either the information that $\tau = 0$ or τ, Y, Z, X_j, $j = 1, 2, \ldots, \tau - 1$. The notation is similar to that of Vardi (1982). If there are several processes, observed over intervals $[a_i, a_i + b_i]$, $i = 1, 2, \ldots, n$, say, denote the observables by τ_i, W_i, Y_i, Z_i, and $X_{i,j}$; $j = 1, 2, \ldots, \tau_i - 1$, $i = 1, 2, \ldots, n$. Finally, let $0 < t_1 < t_2 < \cdots < t_h < \infty$ denote the distinct values assumed by W_i, Y_i, Z_i, and $X_{i,j}$, $j = 1, 2, \ldots, \tau_i - 1, i = 1, 2, \ldots, n$, and let

$$x_k = \#\{(i,j){:}\ 1 \leqslant i \leqslant n, \tau_i \geqslant 2, 1 \leqslant j \leqslant \tau_i - 1, \text{ and } X_{i,j} = t_k\},$$

$$y_k = \#\{i \leqslant n{:}\ \tau_i \geqslant 1 \text{ and } Y_i = t_k\},$$

$$z_k = \#\{i \leqslant n{:}\ \tau_i \geqslant 1 \text{ and } Z_i = t_k\},$$

and

$$w_k = \#\{i \leqslant n{:}\ \tau_i = 0 \text{ and } W_i = t_k\}$$

for $k = 1, 2, \ldots$ Further, let n_x, n_y, n_z, and n_w denote the number of X's, Y's, Z's, and W's in the data set, so that

$$n_x = \sum_{k=1}^{h} x_k, \quad n_y = \sum_{k=1}^{h} y_k, \quad n_z = \sum_{k=1}^{h} z_k, \quad n_w = \sum_{k=1}^{h} w_k,$$

and observe that $n_y + n_w = n = n_z + n_w$.

3. Characterization of the nonparametric maximum likelihood estimator

In the nonarithmetic case, write $\bar{F}(t) = 1 - F(t-)$, $0 \leqslant t < \infty$, for the survival function. Then the likelihood function is

$$\mathscr{L}(F|\text{data}) = \prod_{i=1}^{h} \left\{ [\Delta F(t_i)]^{x_i} \left[\frac{\bar{F}(t_i)}{\mu_F} \right]^{y_i} [\bar{F}(t_i)]^{z_i} \left[\frac{1}{\mu_F} \int_{t_i}^{\infty} \bar{F}(x)\,\mathrm{d}x \right]^{w_i} \right\}, \tag{3.1}$$

where $\Delta F(x) = F(x) - F(x-)$, F is right continuous.

As in the arithmetic case, three data configurations should be considered:

Configuration I: $n_w > 0$, $n_x = n_y = n_z = 0$, i.e., no failures have been observed. In this case, (3.1) is maximized by any probability function F for which $\mu_F = \infty$.

Configuration II: $n_w = 0$, $n_x + n_y + n_z > 0$. Then (3.1) becomes

$$\mathscr{L}(F|\text{data}) = \prod_{i=1}^{h} \{[\Delta F(t_i)]^{x_i} [\mu_F^{-1} \bar{F}(t_i)]^{y_i} [\bar{F}(t_i)]^{z_i}\},$$

which is identical to (2.5) of Vardi (1982b). By essentially the same arguments, it can be shown that for the purpose of maximizing $\mathscr{L}(F|\text{data})$, we can restrict our consideration to distribution functions F supported by $\{t_1, t_2, \dots, t_h\}$. It will be shown later by Lemma 3.4 that in this case, the MLE exists and is unique.

Configuration III: $n_w > 0$, $n_x + n_y + n_z > 0$, in this case there need not exist a distribution function F for which $\mathscr{L}(F|\text{data})$ is maximized. This problem can be dealt with in two different ways:

The first way is to consider only those distributions supported by large but bounded set, and maximize the likelihood $\mathscr{L}(F|\text{data})$ with respect to these distributions. This is the method used for the arithmetic case in Vardi (1982b). The adaption of it to the nonarithmetic case is straightforward. First we define

$$Q = \{\text{all distribution functions with support on } (0, \infty)\},$$

$$Q_M = \{\text{all distribution functions with support on } (0, M]\},$$

$$\mathscr{L}^* = \sup_{F \in Q} \mathscr{L}(F|\text{data}), \quad \mathscr{L}_M^* = \max_{F \in Q_M} \mathscr{L}(F|\text{data}).$$

Note we do not require distributions in Q to have finite means, because the search of MLE will reject distributions which do not have a finite mean.

Lemmas 2.1–2.4 of Vardi (1982b) and their proofs can be adapted here without much difficulty. These lemmas show that for the purpose of maximizing $\mathscr{L}(F|\text{data})$ over all F's in Q_M when $M \gg t_h$, we need only to consider those $F \in Q_M$ supported by

$\{t_1, t_2, \ldots, t_h, M\}$. Any such distribution that maximizes $\mathscr{L}(F|\text{data})$ will be called M-restricted MLE. So the problem reduces to a maximization problem with $h + 1$ variables.

Lemma 3.1. *An M-restricted MLE is a solution to the problem: Maximize*

$$\mathscr{L}(p|\text{data}) = \left(\sum_{i=1}^{h+1} t_i p_i\right)^{-(n_y+n_w)} \prod_{i=1}^{h} \left[p_i^{x_i}\left(\sum_{j=i}^{h+1} p_j\right)^{y_i+z_i}\left(\sum_{j=i}^{h+1}(t_j - t_i)p_j\right)^{w_i}\right] \tag{3.2}$$

subject to

$$\sum_{i=1}^{h+1} p_i = 1, \quad p_i \geqslant 0, \ i = 1, 2, \ldots, h+1.$$

Here $p_i = p(t_i)$, $i = 1, 2, \ldots, h + 1$, $t_{h+1} = M$. *Furthermore*, $\mathscr{L}_M^*$ *converges monotonically to* $\mathscr{L}^*$ *and*

$$\frac{\mathscr{L}_M^*}{\mathscr{L}^*} = 1 - \mathrm{O}(M^{-1}) \quad \text{as } M \to \infty.$$

Proof. Similar to the proof of Lemmas 2.1 and 2.2 of Vardi (1982b). □

The second way is to enlarge the set Q in a suitable way. This will be the approach followed by this paper. Define

$$Q_\infty = \{\tilde{F} = (F, v)\colon F \in Q, 0 \leqslant v < \infty\},$$

$$\mu_{\tilde{F}} = \mu_F + v.$$

Informally, this is the collection of all possible limits for the M-restricted MLEs as $M \to \infty$. Each $\tilde{F} \in Q_\infty$ has an infinitesimal mass at ∞ that contributes to the mean of $\tilde{F}$ by an amount v. So the mean value of $\tilde{F}$ is defined by $\mu_{\tilde{F}} = \mu_F + v$. The likelihood is defined by

$$\mathscr{L}(\tilde{F}|\text{data}) = \prod_{i=1}^{h}\left\{[\Delta F(t_i)]^{x_i}[\mu_{\tilde{F}}^{-1}\bar{F}(t_i)]^{y_i}[\bar{F}(t_i)]^{z_i}\left[\mu_{\tilde{F}}^{-1}\left(\int_{t_i}^{\infty}\bar{F}(x)\,\mathrm{d}x + v\right)\right]^{w_i}\right\}.$$

First we seek to maximize $\mathscr{L}(\tilde{F}|\text{data})$ among those $\tilde{F} = (F, v)$ with F supported by $\{t_1, \ldots, t_h\}$.

Lemma 3.2. *If* $n_y > 0$ (*Configurations* II *and* III), *then the problem of maximizing*

$$\mathscr{L}((p,v)|\text{data}) = \left(\sum_{i=1}^{h} t_i p_i + v\right)^{-(n_y+n_w)} \prod_{i=1}^{h}\left\{p_i^{x_i}\left(\sum_{j=1}^{h} p_j\right)^{y_i+z_i}\left[\sum_{j=i+1}^{h}(t_j - t_i)p_j + v\right]^{w_i}\right\} \tag{3.3}$$

subject to

$$\sum_{i=1}^{h} p_i = 1, \quad p_i \geqslant 0,\ i = 1,2, \dots, h,\ 0 \leqslant v < \infty, \tag{3.4}$$

has a solution. Here $\tilde{p} = (p, v) \in Q_\infty$, $p_i = p(t_i)$, $i = 1,2, \dots, h$ *is the mass function of the distribution* p.

Proof. Let D be the collection of (p,v)'s satisfying (3.4). Since $\mathscr{L}((p,v)|\text{data}) \geqslant 0$ is continuous in D and if $n_y > 0$, then

$$\mathscr{L}((p,v)|\text{data}) \leqslant \left(\sum_{i=1}^{h} t_i p_i + v \right)^{-n_y} \leqslant (t_1 + v)^{-n_y}$$

converges to 0 uniformly as $v \to \infty$. So the maximization problem (3.3) has a solution. □

By the theory of nonlinear programming, such solutions must satisfy the Kuhn-Tucker conditions

$$p_k \frac{\partial l(p,v)}{\partial p_k} = 0, \quad k = 1,2, \dots, h, \tag{3.5}$$

$$v \frac{\partial l(p,v)}{\partial v} = 0, \tag{3.6}$$

$$\frac{\partial l(p,v)}{\partial p_k} \leqslant 0, \quad k = 1,2, \dots, h, \tag{3.7}$$

$$\frac{\partial l(p,v)}{\partial v} \leqslant 0, \tag{3.8}$$

where

$$l(p,v) = \log \mathscr{L}((p,v)|\text{data}) - (n_x + n_z)(\textstyle\sum_{i=1}^{h} p_i - 1) \tag{3.9}$$

is the Lagrangian of $\log \mathscr{L}((p,v)|\text{data})$ and $(n_x + n_y)$ is (necessarily) the Lagrange multiplier.

Let

$$q_k(p,v) = \frac{(n_y + n_w)t_k}{\mu_{\tilde{p}}} + (n_x + n_z) - \sum_{i=1}^{k} \frac{y_i + z_i}{1 - P_{i-1}} - \sum_{i=1}^{k-1} \frac{(t_k - t_i)w_i}{\sum_{i+1}^{h}(t_j - t_i)p_j + v},$$

$k = 1, 2, \dots, h$. Then (3.5)–(3.7) can be written as

$$x_k - p_k q_k(p, v) = 0, \quad k = 1, 2, \dots, h, \tag{3.10}$$

$$\sum_{i=1}^{h} \frac{w_i v}{\sum_{j=i+1}^{h}(t_j - t_i) p_j + v} = \frac{(n_y + n_w) v}{\mu_{\tilde{p}}}, \tag{3.11}$$

$$q_k(p, v) \geqslant 0, \quad k = 1, 2, \dots, h. \tag{3.12}$$

The resulting estimator $\tilde{p}$ actually maximizes $\mathscr{L}(\tilde{F}|\text{data})$ over Q_∞, as shown by Lemma 3.3. Therefore it will be called the NPMLE.

Lemma 3.3. *If* $(p, v) \in D$ *maximizes* (3.3), *then*

$$\mathscr{L}((p, v)|\text{data}) = \lim_{M \to \infty} \mathscr{L}_M^* = \mathscr{L}^* = \mathscr{L}_\infty^* := \sup_{\tilde{F} \in Q_\infty} \mathscr{L}(\tilde{F}|\text{data}).$$

Proof. Note that for any element $\tilde{F} \in Q_\infty$ supported by t_i's, we can construct a distribution supported by t_i's and M, say F_M, in Q_M for each large M, such that F_M and its mean value converge to F and $\mu_{\tilde{F}}$. One such construction is given in Lemma 4.3. Then by the continuity of $\mathscr{L}$, $\lim_{M \to \infty} \mathscr{L}(F_M|\text{data}) = \mathscr{L}(\tilde{F}|\text{data})$. So $\lim_{M \to \infty} \mathscr{L}_M^* = \mathscr{L}_\infty^*$. Since $\mathscr{L}_M^* \leqslant \mathscr{L}^* \leqslant \mathscr{L}_\infty^*$, it follows that $\lim_{M \to \infty} \mathscr{L}_M^* = \mathscr{L}^* = \mathscr{L}_\infty^*$.

To show the first equality, let $p^M = \{p_1^M, \dots, p_h^M, p_{h+1}^M\}$ be the M-restricted MLE for fixed large M, and denote

$$v^M = M p_{h+1}^M, \qquad \mu^M = \sum_{i=1}^{h+1} p_i^M t_i,$$

where $t_{h+1} = M$. Consider the sequence $\tilde{p}^M = (\{p_1^M, \dots, p_h^M\}, v_M)$ where M runs through large integers. Let $\tilde{p}^* = (\{p_1^*, \dots, p_h^*\}, v^*)$ be any limit point of the sequence, then $v^* < \infty$. For otherwise $\mu^* = \lim_{M \to \infty} \mu^M = \infty$,

$$\mathscr{L}_M^* = \mathscr{L}(p^M|\text{data}) \leqslant (\mu^M)^{-n_y} \to 0 \quad \text{as } M \to \infty,$$

contradicting the monotonicity of $\mathscr{L}_M^*$. So $p_{h+1}^M \to 0$, implying $\sum_{i=1}^{h} p_i^* = 1$. Thus $\tilde{p}^* \in D$. And the conclusion follows from

$$\lim_{M \to \infty} \mathscr{L}_M^* = \lim_{M \to \infty} \mathscr{L}(p^M|\text{data}) = \mathscr{L}(\tilde{p}^*|\text{data}) \leqslant \mathscr{L}((p, v)|\text{data}) \leqslant \mathscr{L}_\infty^*. \qquad \square$$

Kuhn–Tucker conditions are necessary conditions for optimality. If they have a unique solution, then the solution must be optimal. This is indeed the case when $n_w = w_h$, which includes the case of fixed window lengths:

Lemma 3.4. *If* $n_w = w_h < n$, *then* (3.5)–(3.8) *subject to constraints* (3.4) *have a unique solution. Consequently, the problem of maximizing* (3.3) *subject to* (3.4) *has a unique solution.*

Proof. It follows from $n_w = w_h$ that $w_k = 0$; $x_k + y_k + z_k \geqslant 1$, $k = 1,2,\ldots,h-1$. So, q_k and (3.11) reduce to

$$q_k(\tilde{p}) = \frac{(n_y + n_w)t_k}{\mu_{\tilde{p}}} + (n_x + n_z) - \sum_{i=1}^{k} \frac{y_i + z_i}{1 - P_{i-1}}, \tag{3.13}$$

$$\frac{v}{\mu_{\tilde{p}}} = \frac{n_w}{n_y + n_w}. \tag{3.14}$$

Suppose $\tilde{p}^i = (p^i, v^i)$, $i = 1,2$ are two solutions, and let $\mu_{\tilde{p}^i}$, $i = 1,2$ be their corresponding mean values.

Suppose that $\mu_{\tilde{p}^1} < \mu_{\tilde{p}^2}$. Then

$$q_1(\tilde{p}^1) = \frac{(n_y + n_w)t_1}{\mu_{\tilde{p}^1}} + (n_x + n_z) - (y_1 + z_1) > q_1(\tilde{p}^2) \geqslant 0.$$

We claim $p_1^1 \leqslant p_1^2$.

Case I: $x_1 > 0$. Then $p_1^1 = x_1/q_1(\tilde{p}^1) < x_1/q_1(\tilde{p}^2) = p_1^2$.

Case II: $x_1 = 0$. By (3.10) and $q_1(\tilde{p}^1) > 0$ we have $p_1^1 = 0 \leqslant p_1^2$.

Assume $p_i^1 \leqslant p_i^2$ for $i = 1,2,\ldots,k$, $1 \leqslant k \leqslant h-1$. Then $q_{k+1}(\tilde{p}^1) > q_{k+1}(\tilde{p}^2) \geqslant 0$. Similar arguments show that $p_{k+1}^1 \leqslant p_{k+1}^2$. Thus we conclude $p_i^1 \leqslant p_i^2$ for $i = 1,2,\ldots,h$, consequently $p_i^1 = p_i^2$ for $i = 1,2,\ldots,h$ and $\mu_{p^1} = \mu_{p^2}$. It then follows from (3.14) that

$$\mu_{\tilde{p}^1} = \frac{n_y + n_z}{n_y}\mu_{p^1} = \frac{n_y + n_z}{n_y}\mu_{p^2} = \mu_{\tilde{p}^2},$$

a contradiction. Similarly, $\mu_{\tilde{p}^1} > \mu_{\tilde{p}^2}$ cannot hold. Thus $\mu_{\tilde{p}^1} = \mu_{\tilde{p}^2}$.

Now we can prove $p_i^1 = p_i^2$, $i = 1,2,\ldots,h$. If not, let k be the first i such that $p_i^1 \neq p_i^2$. Without loss of generality, assume $p_k^1 < p_k^2$, we will show that this leads to $p_i^1 \leqslant p_i^2$, $i = k, k+1, \ldots, h$.

If $k = h$, there is nothing to prove. So we assume $k < h$. Then it is easy to see that $q_{k+1}(\tilde{p}^1) \geqslant q_{k+1}(\tilde{p}^2)$.

Case I: $x_{k+1} > 0$, then $p_{k+1}^1 = x_{k+1}/q_{k+1}(\tilde{p}^1) \leqslant x_{k+1}/q_{k+1}(\tilde{p}^2) = p_{k+1}^2$.

Case II: $x_{k+1} = 0$, $k+1 < h$. By assumption, $y_{k+1} + z_{k+1} > 0$, thus $q_{k+1}(\tilde{p}^1) > q_{k+1}(\tilde{p}^2) \geqslant 0$, $p_{k+1}^1 = 0 \leqslant p_{k+1}^2$.

Using induction arguments we see $p_i^1 \leqslant p_i^2$, $i = k, k+1, \ldots, h-1$.

If $x_h + y_h + z_h > 0$, the above proof remains valid for $k = h-1$. If $x_h + y_h + z_h = 0$, then $q_h(\tilde{p}^1) = n(t_h - t_{h-1})/\mu_{\tilde{p}^1} + q_{h-1}(\tilde{p}^1) > 0$, implying $p_h^1 = 0 \leqslant p_h^2$.

Thus $1 = \sum_{k=1}^{h} p_k^1 < \sum_{k=1}^{h} p_k^2 = 1$, a contradiction. □

4. The modified RT algorithm

Vardi (1982b) studied the same problem for the arithmetic case. By introducing an artificial incomplete data problem which has the identical likelihood as the observed data, he developed the RT algorithm for finding the M-restricted MLE by using the EM algorithm. Note that the likelihood function (3.2) differs from (2.17) of Vardi (1982b) only in the last term: If $t_j - t_i$ in (3.2) is replaced by $t_j - t_i + 1$, it will coincide with (2.17) of Vardi (1982b). So by replacing $t_k - t_i + 1, t_j - t_i + 1$ with $t_k - t_i, t_j - t_i$, the RT algorithm of Vardi can be used to find the M-restricted MLE for nonarithmetic cases. The resulting algorithm will again be referred to as the RT algorithm.

To derive an algorithm for finding NPMLE, let $M \to \infty$, $p_{h+1} \to 0$ and $Mp_{h+1} \to v$ in the RT algorithm. The resulting algorithm will be called the modified RT algorithm. The description of this algorithm and its properties are given below.

Step A: Start with an initial estimate $\{p_k^{\text{old}}\}_{k=1}^h$ and v^{old} satisfying

$$p_k^{\text{old}} > 0, \quad k = 1, 2, \ldots, h, \quad \sum_{k=1}^{h} p_k^{\text{old}} = 1,$$

$$v^{\text{old}} > 0 \text{ (can be set at 0 if } n_w = 0).$$

Step B: Evaluate, for $k = 1, 2, \ldots, h$, let

$$r_k = x_k + p_k^{\text{old}} \sum_{i=1}^{k} \left\{ \frac{y_i + z_i}{\sum_{j=i}^{h} p_j^{\text{old}}} + \frac{(t_k - t_i)\, w_i}{\sum_{j=i}^{h} (t_j - t_i)\, p_j^{\text{old}} + v^{\text{old}}} \right\}$$

and

$$r_{h+1} = v^{\text{old}} \sum_{i=1}^{h} \frac{w_i}{\sum_{j=i}^{h} (t_j - t_i) p_j^{\text{old}} + v^{\text{old}}}.$$

Solve the equation

$$\sum_{k=1}^{h} \frac{r_k t_k}{(n_x + n_z)\mu + (n_y + n_w) t_k} = 1 - \frac{r_{h+1}}{n_y + n_w}$$

for μ. Denote the solution by μ^{new}.

Step C: Set

$$p_k^{\text{new}} = \frac{r_k \mu^{\text{new}}}{(n_x + n_z)\mu^{\text{new}} + (n_y + n_w) t_k}, \quad k = 1, 2, \ldots, h,$$

and

$$v^{\text{new}} = \frac{r_{h+1} \mu^{\text{new}}}{n_y + n_w}.$$

Step D: If $(p^{\text{new}}, v^{\text{new}})$ is sufficiently close to $(p^{\text{old}}, v^{\text{old}})$, then stop. Otherwise return to step B with $(p_k^{\text{new}}, v^{\text{new}})$ replacing $(p_k^{\text{old}}, v^{\text{old}})$.

The initial values could be set at $p_k^{\text{old}} = 1/h$, $k = 1, 2, \ldots, h$; $v^{\text{old}} = n_w/n$.

Remark 4.1. If $n_y = 0$ (Configuration I), there is no need to go through the calculations.

An analogue of Theorem 1 of Vardi (1982b) for the modified RT algorithm is

Theorem 4.2. *If $n_y \geqslant 1$ (configurations II and III), then the modified RT algorithm converges monotonically in likelihood to a fixed point (p^*, v^*) which satisfies the Kuhn–Tucker conditions* (3.5)–(3.8).

The key element in the proof of Theorem 1 of Vardi (1982b) is to use the properties of the EM algorithms to show that the RT algorithm increases monotonically in likelihood. The RT algorithm for the nonarithmetic case preserves this property and similar convergence results can be stated. The proof (Section 3 of Vardi), however, needs certain modifications. Specifically, part of the incomplete data, A_w, should be redefined as

$$\{[R_{n_y+1} v_{n_y+1} - 1] \wedge \bar{w}_1, \ldots, [R_{n_y+n_w} v_{n_y+n_w} - 1] \wedge \bar{w}_{n_w}\},$$

where R_i's are uniform on $(0, 1)$, v_i's follow the length-biased distribution of p. See Vardi (1982b) for details. Consequently, in the E-step of the EM algorithm, the expected log-likelihood for the complete data given the observed incomplete data becomes

$$\mathcal{Q}_M(p | p^{\text{old}}) = \sum_{i=1}^{h+1} \tilde{\xi}_i \log p_i + \sum_{i=1}^{h+1} \tilde{\eta}_i \log \frac{t_i p_i}{\sum_{j=1}^{h+1} t_j p_j},$$

where

$$\tilde{\xi}_k = x_k + p_k^{\text{old}} \sum_{i=1}^{k} \frac{z_i}{\sum_{j=1}^{h+1} p_j^{\text{old}}},$$

$$\tilde{\eta}_k = p_k^{\text{old}} \sum_{i=1}^{k} \frac{y_i}{\sum_{j=1}^{h+1} p_j^{\text{old}}} + p_k^{\text{old}} \sum_{i=1}^{k} \frac{(t_k - t_i) w_i}{\sum_{j=1}^{h+1} (t_j - t_i) p_j^{\text{old}}}.$$

For the modified RT algorithm, since $(p, v) \in D$ is no longer a proper distribution, we cannot derive it directly from the EM algorithm. In order to prove Theorem 4.2, first we show that the modified RT algorithm is indeed monotone in likelihood:

Lemma 4.3. *If in the modified RT algorithm, $(p^{\text{new}}, v^{\text{new}}) \neq (p^{\text{old}}, v^{\text{old}})$, then $\mathcal{L}((p^{\text{new}}, v^{\text{new}}) |$* data) $> \mathcal{L}((p^{\text{old}}, v^{\text{old}}) |$ data).

Proof. For any (p, v) satisfying (3.4), let $t_{h+1} = M \gg t_h$, and

$$p_{M,i} = \frac{p_i}{1 + v/M}, \quad i = 1, 2, \ldots, h; \quad p_{M,h+1} = \frac{v/M}{1 + v/M}.$$

For sufficiently large M, $p_M = (p_{M,1}, p_{M,2}, \dots, p_{M,h+1}) \in Q_M$, $p_{M,i} \to p_i$, $i = 1, 2, \dots, h$, and $\mu_{p_M} \to \mu_p + v$ as $M \to \infty$. Assume $(p^{\text{old}}, v^{\text{old}})$ satisfies (3.4). Applying a single iteration of the RT algorithm to p_M^{old}, a single iteration of the modified RT algorithm to $(p^{\text{old}}, v^{\text{old}})$ we get p_M^{new} and $(p^{\text{new}}, v^{\text{new}})$ (resp.). By Lemma 1 of Dempster et al. (1977),

$$\mathscr{L}(p_M^{\text{new}}|\text{data}) - \mathscr{L}(p_M^{\text{old}}|\text{data}) \geqslant \mathscr{Q}_M(p_M^{\text{new}}|p_M^{\text{old}}) - \mathscr{Q}_M(p_M^{\text{old}}|p_M^{\text{old}}).$$

Let $M \to \infty$, it is easy to show from the algorithm that

$$p_{M,k}^{\text{new}} \to p_k^{\text{new}}, \ k = 1, 2, \dots, h; \quad Mp_{M,h+1}^{\text{new}} \to v^{\text{new}},$$

$$\mathscr{L}((p^{\text{new}}, v^{\text{new}})|\text{data}) - \mathscr{L}((p^{\text{old}}, v^{\text{old}})|\text{data}),$$

$$\geqslant \mathscr{Q}((p^{\text{new}}, v^{\text{new}})|(p^{\text{old}}, v^{\text{old}})) - \mathscr{Q}((p^{\text{old}}, v^{\text{old}})|(p^{\text{old}}, v^{\text{old}})),$$

where

$$\mathscr{Q}((p, v)|(p^{\text{old}}, v^{\text{old}})) = \sum_{i=1}^{h} \tilde{\xi}_i \log p_i + \sum_{i=1}^{h} \tilde{\eta}_i \log \frac{t_i p_i}{\sum_{j=1}^{h} t_j p_j + v} + \tilde{\eta}_{h+1} \log \frac{v}{\sum_{j=1}^{h} t_j p_j + v},$$

$$\tilde{\xi}_k = x_k + p_k^{\text{old}} \sum_{i=1}^{k} \frac{z_i}{\sum_{j=i}^{h} p_j^{\text{old}}},$$

$$\tilde{\eta}_k = p_k^{\text{old}} \sum_{i=1}^{k} \frac{y_i}{\sum_{j=i}^{h} p_j^{\text{old}}} + p_k^{\text{old}} \sum_{i=1}^{k} \frac{(t_k - t_i) w_i}{\sum_{j=i}^{h} (t_j - t_i) p_j^{\text{old}} + v^{\text{old}}},$$

for $k = 1, 2, \dots, h$, and

$$\tilde{\eta}_{h+1} = v^{\text{old}} \sum_{i=1}^{h} \frac{w_i}{\sum_{j=i}^{h} (t_j - t_i) p_j^{\text{old}} + v^{\text{old}}}.$$

It can be shown that $\mathscr{Q}((p, v)|(p^{\text{old}}, v^{\text{old}}))$, as a function of (p, v), subject to the constraints (3.4), is maximized uniquely at $(p^{\text{new}}, v^{\text{new}})$. Thus

$$\mathscr{L}((p^{\text{new}}, v^{\text{new}})|\text{data}) > \mathscr{L}((p^{\text{old}}, v^{\text{old}})|\text{data})$$

if $(p^{\text{new}}, v^{\text{new}}) \neq (p^{\text{old}}, v^{\text{old}})$. □

Using Convergence Theorem A of Zangwill (1969, p. 91) we can show that the modified RT algorithm converges to a fixed point. Eqs. (3.5) and (3.6) are rewritings of the algorithm for fixed points. Eqs. (3.7) and (3.8) can be proved by negation arguments, similar to that of Vardi. Details are omitted here.

5. An example

We illustrate the use of the modified RT algorithm by an example. The example comes from Vardi (1982b). For the purpose of comparison, we reproduce Table 1 of Vardi here together with the estimates given by the modified RT algorithm. In the table, for $M = 10^2, 10^3, \dots, 10^6$, p_M is the M-restricted estimates. The last column, $M = \infty$, corresponds to the (unrestricted) RT estimates given by the modified RT algorithm. It conforms to the trend in the M-restricted estimates. The last two rows give the estimates of the residual mean above 19 and the estimates of the mean.

6. Preparatory limit theorems for renewal processes

Consider a delayed renewal process starting at 0 with residual life distribution G as initial distribution and with F as lifetime distribution. Thus the process is stationary. See Feller (1971, Ch. XI) for details. Let $X_0, X_1, X_2, \dots$ be the lifetimes, then $X_0 \sim G, X_1, X_2, \dots \sim F$ and $X_0, X_1, X_2, \dots$ are independent. Let

$$V(x) = \begin{cases} \sum_{k=0}^{\infty} F^{\star k}(x) & \text{if } x \geqslant 0, \\ 0 & \text{else,} \end{cases} \tag{6.1}$$

where $\star$ stands for convolution. Then V defines a measure on $\mathbb{R}$. $U = G\star V$ is the renewal measure. It gives the expected number of failures,

$$E(\tau^x) = U(x) = \frac{x}{\mu_F} \quad \forall x \in (0, \infty).$$

Table 1
A comparison of the M-restricted and unrestricted estimates

Data					Estimates $p_M(t)$					
t_i	x_i	y_i	z_i	w_i	$M = 10^2$	$M = 10^3$	$M = 10^4$	$M = 10^5$	$M = 10^6$	$M = \infty$
3	0	1	0	0	0	0	0	0	0	0
7	1	0	0	0	0.1082	0.1098	0.1101	0.1101	0.1101	0.1104
8	0	0	1	0	0	0	0	0	0	0
9	2	0	0	0	0.2361	0.2411	0.2417	0.2418	0.2418	0.2428
10	0	1	0	0	0	0	0	0	0	0
13	1	0	0	0	0.1307	0.1354	0.1360	0.1361	0.1361	0.1371
14	0	0	1	0	0	0	0	0	0	0
16	1	0	0	0	0.1592	0.1692	0.1705	0.1707	0.1707	0.1728
17	0	0	0	1	0	0	0	0	0	0
19	2	0	0	0	0.3303	0.3393	0.3411	0.3413	0.3413	0.3369
M	0	0	0	0	0.0355	0.0053	†ε_1	ε_2	ε_3	0
					$v = 3.550$	5.174	5.405	5.427	5.430	5.942
					$\mu_M = 16.954$	19.027	19.328	19.359	19.362	19.848

†$10^2\varepsilon_3 \approx 10\varepsilon_2 \approx \varepsilon_1 \approx 5.4 \times 10^{-4}$.

The machine which works at time b has a life X_{τ^b}. It is known that the distribution of X_{τ^b} is length-biased if $b \to \infty$. Generally,

Lemma 6.1. *For all* $b > 0$,

$$P(X_{\tau^b} \geqslant x, \tau^b \geqslant 1) = \frac{b \wedge x}{\mu_F} \bar{F}(x) + \frac{1}{\mu_F} \int_{b \wedge x}^{b} \bar{F}(t)\,\mathrm{d}t, \quad \forall x > 0, \tag{6.2}$$

$$P(X_{\tau^b} \leqslant x, \tau^b \geqslant 1) = \frac{1}{\mu_F} \int_0^x (s \wedge b) F(\mathrm{d}s), \quad \forall x > 0, \tag{6.3}$$

$$R^b \sim G, \tag{6.4}$$

$$P(Z^b \leqslant z, \tau^b \geqslant 1) = G(z), \quad \forall\, 0 \leqslant z < b, \tag{6.5}$$

where R^b, Z^b *are defined in the Introduction.*

Proof. Write τ, R, Z for τ^b, R^b, Z^b (resp.). Then

$$\begin{aligned}
P(Z \geqslant z,\ R \geqslant y,\ \tau \geqslant 1) &= \sum_{k=1}^{\infty} P(Z \geqslant z,\ R \geqslant y,\ \tau = k) \\
&= \sum_{k=1}^{\infty} P(S_{k-1} \leqslant b - z, S_{k-1} + X_k \geqslant b + y) \\
&= \sum_{k=0}^{\infty} \int_{y+z}^{\infty} G\star F^{\star k}[b + y - s, b - z] F(\mathrm{d}s) \\
&= \int_{y+z}^{\infty} U[b + y - s, b - z] F\,(\mathrm{d}s) \\
&= \frac{1}{\mu_F} \int_{y+z}^{\infty} ((s - y - z) \wedge (b - z)) F(\mathrm{d}s) \\
&= \frac{1}{\mu_F} \int_{y+z}^{y+b} (1 - F(s-))\,\mathrm{d}s \\
&= G(y + b) - G(y + z)
\end{aligned}$$

for all $y, z > 0$. Observe that

$$P(\tau \geqslant 1) = P(X_0 \leqslant b) = G(b),$$

$$X_\tau = Z + R.$$

Relations (6.4) and (6.5) follow by letting $z = 0$ and $y = 0$. We omit the details, since the result is known. See Feller (1971).

Proof of (6.3): If F has a continuous density, then

$$P(X_\tau \leqslant x, \tau \geqslant 1) = P(Z + R \leqslant x, \tau \geqslant 1)$$

$$= \int_0^{b \wedge x} \int_0^{x-z} \frac{\mathrm{d}^2}{\mathrm{d}z\,\mathrm{d}y} P(Z \geqslant z, R \geqslant y, \tau \geqslant 1)\,\mathrm{d}y\,\mathrm{d}z$$

$$= \frac{1}{\mu_F} \int_0^{b \wedge x} \int_0^{x-z} \frac{\mathrm{d}}{\mathrm{d}y} F(y+z)\,\mathrm{d}y\,\mathrm{d}z$$

$$= \frac{1}{\mu_F} \int_0^{b \wedge x} [F(x) - F(z)]\,\mathrm{d}z$$

$$= \frac{1}{\mu_F} \int_0^{b \wedge x} zF(\mathrm{d}z) + [F(x) - F(b \wedge x)](b \wedge x)$$

$$= \frac{1}{\mu_F} \int_0^{x} (s \wedge b) F(\mathrm{d}s).$$

The general case may then be obtained by a transparent smoothing agrument.

Eq. (6.2) follows from (6.3) and $P(\tau \geqslant 1) = G(b)$. □

Proposition 6.2. *If $b > 0$ and h is a measurable function defined on $[0, b]$, then*

$$E\left(\sum_{k=1}^{\tau^b - 1} h(X_k)\right) = \frac{1}{\mu_F} \int_0^b (b-t) h(t) F(\mathrm{d}t)$$

with summation convention $\sum_{k=1}^{i} = 0$ for $i \leqslant 0$.

Proof. Since $\tau > k$ iff $S_k \leqslant b$,

$$E \sum_{k=1}^{\tau - 1} h(X_k) = \sum_{k=1}^{\infty} \int_{S_k \leqslant b} h(X_k)\,\mathrm{d}P$$

$$= \sum_{k=1}^{\infty} \int_0^b h(t) G \star F^{\star(k-1)}(b-t) F(\mathrm{d}t)$$

$$= \int_0^b h(t) U(b-t) F(\mathrm{d}t)$$

$$= \frac{1}{\mu_F} \int_0^b (b-t) h(t) F(\mathrm{d}t). \qquad \square$$

For $0 \leqslant x \leqslant b$, let $h(t) = \{t \leqslant x\}$, then

Corollary 6.3. *For* $0 \leqslant x \leqslant b$,

$$E\left(\sum_{k=1}^{\tau^b - 1} \{X_k \leqslant x\}\right) = \frac{1}{\mu_F} \int_0^x (b - t) F(\mathrm{d}t).$$

Returning to the estimation problem, suppose that n renewal processes are observed in intervals $[a_i, a_i + b_i]$, $i = 1, 2, \ldots, n$. For $0 \leqslant x < \infty$, define

$$F_X^*(x) = \frac{1}{n} \sum_{k: t_k \leqslant x} x_k, \quad S_X^*(x) = \frac{1}{n} \sum_{k: t_k \geqslant x} x_k,$$

$$F_Y^*(x) = \frac{1}{n} \sum_{k: t_k \leqslant x} y_k, \quad S_Y^*(x) = \frac{1}{n} \sum_{k: t_k \geqslant x} y_k,$$

$$F_Z^*(x) = \frac{1}{n} \sum_{k: t_k \leqslant x} z_k, \quad S_Z^*(x) = \frac{1}{n} \sum_{k: t_k \geqslant x} z_k,$$

$$F_W^*(x) = \frac{1}{n} \sum_{k: t_k \leqslant x} w_k, \quad S_W^*(x) = \frac{1}{n} \sum_{k: t_k \geqslant x} w_k.$$

If f is a bounded function on $[0, b]$, let $\| f \|_0^b = \sup_{0 \leqslant x \leqslant b} |f(x)|$. For $b > 0$ define

$$F_X(x) = F_X(x; b) = \frac{1}{\mu_F} \int_0^x (b - t) F(\mathrm{d}t),$$

$$S_X(x) = S_X(x; b) = F_X(b) - F_X(x-)$$

over interval $0 \leqslant x \leqslant b$.

Theorem 6.4. *If F is nonarithmetic, $b_1 = b_2 = \cdots = b$, then as $n \to \infty$,*

$$\frac{n_x}{n} \to \frac{b}{\mu_F} - G(b), \tag{6.6}$$

$$\frac{n_y}{n} = \frac{n_z}{n} - G(b), \tag{6.7}$$

$$\bar{\tau}_n =: \frac{\sum_{i=1}^n \tau_i}{n} = \frac{n_x + n_y}{n} \to \frac{b}{\mu_F}, \tag{6.8}$$

$$\frac{n_w}{n} \to 1 - G(b), \tag{6.9}$$

$$\| F_X^* - F_X \|_0^b \to 0, \tag{6.10}$$

$$\| F_Y^* - G \|_0^b \to 0, \tag{6.11}$$

$$\|F_Z^* - G\|_0^b \to 0, \tag{6.12}$$

$$\|S_X^* - S_X\|_0^b \to 0, \tag{6.13}$$

$$\|S_Y^* - (G(b) - G)\|_0^b \to 0, \tag{6.14}$$

$$\|S_Z^* - (G(b) - G)\|_0^b \to 0, \tag{6.15}$$

$$\|S_W^*(x) - (1 - G(b))\|_0^b \to 0. \tag{6.16}$$

Proof. Relation (6.7) is clear from the Strong Law of Large Numbers, since $n_y = n_z = \#\{i\colon \tau_i \geqslant 1\}$ and $P(\tau_i \geqslant 1) = G(b)$, and (6.9) and (6.16) then follow from $n_y + n_w = n$. Similarly, (6.8) follows from $E(\tau) = b/\mu_F$, and (6.6) from (6.7) and (6.8). A lemma is needed for (6.10)–(6.15).

Lemma 6.5. *Suppose that $\phi, \phi_1, \phi_2, \ldots$ are nondecreasing, right continuous functions defined on a compact interval $I \subset \mathbb{R}$, that $\phi_n(x) \to \phi(x)$ as $n \to \infty$ for $x \in I$, and that ϕ is bounded over I. Further assume that $\phi_n(x-) \to \phi(x-)$ at discontinuity points of ϕ, then $\phi_n(x) \to \phi(x)$ uniformly for $x \in I$.*

Proof of Lemma 6.5. Without loss of generality, suppose that $I = [0, b]$ for some $b > 0$. For any $\varepsilon > 0$, there is a partition of I, say by $0 = \xi_0 < \xi_1 < \cdots < \xi_N = b$ such that $|\phi(\xi_i-) - \phi(\xi_{i-1})| \leqslant \varepsilon$ for all $i = 1, 2, \ldots, N$. See Lemma 1, Billingsley (1968, p. 110). By the assumptions, for any $i = 1, 2, \ldots, N$, $|\phi_n(\xi_{i-1}) - \phi(\xi_{i-1})| \leqslant \varepsilon$, $|\phi_n(\xi_i-) - \phi(\xi_i-)| \leqslant \varepsilon$ for large n. Fix $x \in [\xi_{i-1}, \xi_i)$, $1 \leqslant i \leqslant N$, then $|\phi_n(x) - \phi(x)| \leqslant |\phi_n(\xi_i-) - \phi(\xi_i-)| + |\phi(\xi_i-) - \phi(\xi_{i-1})| + |\phi_n(\xi_{i-1}) - \phi(\xi_{i-1})| \leqslant 3\varepsilon$. This proves the uniform convergence. □

Proof of Theorem 6.4 (*continued*). For (6.10)–(6.12), we first prove their pointwise versions. Let $0 \leqslant x \leqslant b$. Then

$$nF_X^*(x) = \sum_{k: t_k \leqslant x} x_k = \sum_{i=1}^{n} \sum_{j=1}^{\tau_i - 1} \{X_{i,j} \leqslant x\},$$

$$nF_Y^*(x) = \sum_{k: t_k \leqslant x} y_k = \sum_{i=1}^{n} \{Y_i \leqslant x, \tau_i \geqslant 1\},$$

$$nF_Z^*(x) = \sum_{k: t_k \leqslant x} z_k = \sum_{i=1}^{n} \{Z_i \leqslant x, \tau_i \geqslant 1\}$$

are sums of i.i.d. random variables. Applying the Strong Law of Large Numbers again we have the pointwise a.s. convergences. Suppose $K \subset I$ is countable and dense, then with probability 1 the pointwise convergences hold for all $x \in K$, since the limiting functions are right continuous, therefore this can be extended to $x \in I$. Note that all the empirical processes are monotone and their pointwise limits are bounded, by Lemma 6.5 we have the desired uniform a.s. convergences. Similarly we can prove (6.13)–(6.15). □

Remark 6.6. In Theorem 6.4 the condition $b_1 = b_2 = \cdots = b_n$ can be replaced by $b_n \to b$ by using Kolmogorov's Strong Law of Large Numbers.

Suppose now that $b_1, b_2, \ldots,$ are i.i.d. realizations of a distribution function B, $B(0) = 0$, and that they are independent of the renewal processes. Let $\mu_B = \int_0^\infty \bar{B}(x)\,dx > 0$. Then we have an analogue of Theorem 6.4:

Theorem 6.7. *Suppose $B_n \to B$ in $[0, b]$, then*

$$\frac{n_x}{n} \to \frac{\mu_B}{\mu_F} - \frac{1}{\mu_F}\int_0^\infty \bar{F}(t)\bar{B}(t)\,dt, \tag{6.17}$$

$$\frac{n_y}{n} = \frac{n_z}{n} \to \frac{1}{\mu_F}\int_0^\infty \bar{F}(t)\bar{B}(t)\,dt, \tag{6.18}$$

$$\frac{n_w}{n} \to 1 - \frac{1}{\mu_F}\int_0^\infty \bar{F}(t)\bar{B}(t)\,dt, \tag{6.19}$$

$$S_X^*(x) \to \frac{1}{\mu_F}\int_x^\infty \bar{B}(t)[\bar{F}(x) - \bar{F}(t)]\,dt, \tag{6.20}$$

$$S_Y^*(x),\ S_Z^*(x) \to \frac{1}{\mu_F}\int_x^\infty \bar{B}(t)\bar{F}(t)\,dt, \tag{6.21}$$

$$S_W^*(x) \to \frac{1}{\mu_F}\int_x^\infty \bar{F}(t)[\bar{B}(x) - \bar{B}(t)]\,dt, \tag{6.22}$$

and the convergences in (6.20)–(6.22) *are uniform in* $[0, b]$.

Proof. From the proof of Theorem 6.4, we know that

$$E((\tau_1 - 1) \vee 0 \mid b_1 = b) = \frac{b}{\mu_F} - G(b).$$

Taking expectation on both sides, we have

$$\begin{aligned} E((\tau_1 - 1) \vee 0) &= E\left(\frac{b_1}{\mu_F} - G(b_1)\right) \\ &= \frac{\mu_B}{\mu_F} - \frac{1}{\mu_F}\int_0^\infty \int_0^b \bar{F}(t)\,dt\,B(db) \\ &= \text{RHS of (6.17)}, \end{aligned}$$

thus (6.17) holds. Assertions (6.18)–(6.22) may be proved similarly. □

Remark 6.8. Theorem 6.7 can be put in a more general setting. Suppose B is a distribution function with $B(0) = 0$, and that $\mu_B = \int_0^\infty \bar{B}(x)\,dx > 0$. Define $B_n(x) = \#\{i \leqslant n: b_i \leqslant x\}/n$ for $x \geqslant 0$, $n = 1, 2, \ldots$ If $B_n \to B$ in $[0, b]$, then Theorem 6.7 holds. The proof again involves Kolmogorov's Strong Law of Large Numbers.

7. Consistency

Suppose a sequence of stationary renewal processes are observed in windows $[a_i, a_i + b_i]$ respectively, $i = 1, 2, \ldots$ Further assume that the selection of a_i, b_i's are independent of the processes. Let $\tilde{p}^n = (p^n, v_n)$ be the estimator generated by the modified RT algorithm based on the observations from the first n processes (hereafter the RT estimator), or $\tilde{p}^n = (p^n, v_n)$ is the NPMLE. Define

$$H_n(x) = \sum_{k: t_{n,k} \leqslant x} p_k^n, \quad x \geqslant 0,$$

$$\mu_n = \mu_{p^n} = \sum t_{n,k} p_k^n, \quad \tilde{\mu}_n = \mu_n + v_n,$$

where $\{t_{n,k}\}$ are ordered distinct observations from X, Y, Z, W of the first n renewal processes. See section 2 for notations. In this section, when the discussion is restricted to fixed n we will suppress the index "n" in some notations. For example, $t_{n,k}$ may be written as t_k. By Theorem 4.2, $\tilde{p}^n$ satisfies Kuhn–Tucker conditions (3.10)–(3.12). First we write these conditions in terms of the estimated distribution function H_n and the estimated mean value $\tilde{\mu}_n$.

Lemma 7.1. *If $(H_n, \tilde{\mu}_n)$ is the NPMLE (or the RT estimator) of F and μ_F, then it satisfies the* integrated score equation

$$S_X^*(x) + S_Y^*(x) + S_Z^*(x) + S_W^*(x) + \int_{[0,x)} \frac{\bar{H}_n(x)}{\bar{H}_n(t)}\, d(F_Y^*(t) + F_Z^*(t))$$
$$+ \int_{[0,x)} \frac{(x-t)\bar{H}_n(x) + \tilde{\mu}_n - \int_0^x \bar{H}_n(u)\,du}{\tilde{\mu}_n - \int_0^t \bar{H}_n(u)\,du}\, dF_W^*(t)$$
$$= \frac{1}{\tilde{\mu}_n}\left(x\bar{H}_n(x) + \tilde{\mu}_n - \int_0^x \bar{H}_n(u)\,du\right) + \frac{n_x + n_z}{n}\bar{H}_n(x). \tag{7.1}$$

Proof. Let $P_k = \sum_{i=1}^k p_i$, $k = 1, 2, \ldots$ For a fixed $0 < x < t_h$ let r be the smallest integer k for which $t_k \geqslant x$. Summing (3.10) for $k \geqslant r$ and (3.11), dividing the sum by n, then using the definition of q_k, we get

$$S_X^*(x) = \frac{\int_x^\infty t H_n(dt) + v_n}{\tilde{\mu}_n} + \bar{\tau}_n \bar{H}_n(x) - \frac{1}{n}\sum_{k=r}^h p_k \sum_{i=1}^k \frac{y_i + z_i}{1 - P_{i-1}}$$
$$- \frac{1}{n}\left\{\sum_{k=r}^h p_k \sum_{i=1}^{k-1} \frac{w_i(t_k - t_i)}{\sum_{j=i}^h (t_j - t_i)p_j + v} + \sum_{i=1}^h \frac{w_i v}{\sum_{j=i+1}^h (t_j - t_i)p_j + v}\right\}. \tag{7.2}$$

Here

$$\frac{1}{n}\sum_{k=r}^{h} p_k \sum_{i=1}^{k} \frac{y_i + z_i}{1 - P_{i-1}} = \frac{1}{n}\left\{\sum_{i=1}^{r-1}\sum_{k=r}^{h} + \sum_{i=r}^{h}\sum_{k=i}^{h}\right\}\frac{y_i + z_i}{1 - P_{i-1}} p_k$$

$$= \int_{[0,x)} \frac{\bar{H}_n(x)}{\bar{H}_n(t)} \mathrm{d}(F_Y^*(t) + F_Z^*(t)) + S_Y^*(x) + S_Z^*(x),$$

$$\frac{1}{n}\left\{\sum_{k=r}^{h} p_k \sum_{i=1}^{k-1} \frac{w_i(t_k - t_i)}{\sum_{j=i}^{h}(t_j - t_i)p_j + v} + \sum_{i=1}^{h} \frac{w_i v}{\sum_{j=i+1}^{h}(t_j - t_i)p_j + v}\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{r-1} \frac{\sum_{k=r}^{h}(t_k - t_i)p_k + v}{\sum_{j=i}^{h}(t_j - t_i)p_j + v} w_i + \sum_{k=r}^{h} w_k\right\}$$

$$= S_W^*(x) + \int_{[0,x)} \frac{(x - t)\bar{H}_n(x) + \int_x^\infty \bar{H}_n(u)\,\mathrm{d}u + v_n}{\int_t^\infty \bar{H}_n(u)\,\mathrm{d}u + v_n} \mathrm{d}F_W^*(t)$$

and

$$\int_x^\infty \bar{H}_n(s)\,\mathrm{d}s + v_n = \tilde{\mu}_n - \int_0^x \bar{H}_n(s)\,\mathrm{d}s.$$

Substituting these relations into (7.2) yields (7.1). □

Letting $x > t_h$ in the above equation, or by using (3.11), we have

$$\int_0^\infty \frac{v_n}{\tilde{\mu}_n - \int_0^t \bar{H}_n(u)\,\mathrm{d}u} \mathrm{d}F_W^*(t) = \frac{v_n}{\tilde{\mu}_n}. \tag{7.3}$$

By Helly's Selection Theorem, there is a subsequence $H_{n_k}(x)$ and a nonincreasing right continuous function $H(x)$ with values in $[0, 1]$ for which $\lim_{k\to\infty} H_{n_k}(x) = H(x)$ at all continuity points x of H. Both n_k and H may depend on ω, and we cannot assume (at this point) that H is a distribution function. We may also assume $\tilde{\mu}_{n_k}$, μ_{n_k} and v_{n_k} converge to limits $\tilde{\mu}_H$, μ_H and v_H, where $0 \leqslant \mu_H + v_H = \tilde{\mu}_H \leqslant \infty$ (otherwise we can always take a further subsequence). Taking limit along the subsequence $\{n_k, k = 1, 2, \dots\}$ and using (7.1) and (7.3), we may obtain

Lemma 7.2. *Suppose $B_n(t) = \sum_{i=1}^{n}\{b_i \leqslant t\}/n \to B(t)$ as $n \to \infty$. With the notation of the previous paragraph,*

$$\frac{\bar{F}(x)}{\mu_F}\int_x^\infty \bar{B}(t)\,\mathrm{d}t + \frac{\bar{B}(x)}{\mu_F}\int_x^\infty \bar{F}(t)\,\mathrm{d}t + \frac{2\bar{H}(x)}{\mu_F}\int_0^x \frac{\bar{F}(t)}{\bar{H}(t)}\bar{B}(t)\,\mathrm{d}t$$

$$+ \int_{[0,x)} \frac{(x - t)\bar{H}(x) + \tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t}{\tilde{\mu}_H - \int_{[0,t)} \bar{H}(u)\,\mathrm{d}u} \frac{\int_t^\infty \bar{F}(u)\,\mathrm{d}u}{\mu_F} \mathrm{d}B(t)$$

$$= \frac{1}{\tilde{\mu}_H}\left(\tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t + x\bar{H}(x)\right) + \frac{\mu_B}{\mu_F}\bar{H}(x) \tag{7.4}$$

and

$$\int_0^\infty \frac{\nu_H}{\tilde{\mu}_H - \int_0^t \bar{H}(u)\,\mathrm{d}u} \frac{\int_t^\infty \bar{F}(u)\,\mathrm{d}u}{\mu_F}\,\mathrm{d}B(t) = \frac{\nu_H}{\tilde{\mu}_H}. \tag{7.5}$$

Proof. First we show that $\bar{H}(x) > 0$ for sufficiently small $x > 0$. For $t \leqslant x$, we have

$$\bar{H}_n(x) \leqslant \bar{H}_n(t),$$

$$(x-t)\bar{H}_n(x) \leqslant \int_t^x \bar{H}_n(u)\,\mathrm{d}u \leqslant \int_t^\infty \bar{H}_n(u)\,\mathrm{d}u + \nu_n \leqslant \tilde{\mu}_n - \int_0^t \bar{H}_n(u)\,\mathrm{d}u,$$

$$x\bar{H}_n(x) + \tilde{\mu}_n - \int_0^x \bar{H}_n(u)\,\mathrm{d}u = \tilde{\mu}_n - \int_0^x (\bar{H}_n(t) - \bar{H}_n(x))\,\mathrm{d}t \leqslant \tilde{\mu}_n.$$

So, by (7.1)

$$\begin{aligned} 1 + \frac{n_x + n_z}{n}\bar{H}_n(x) &\geqslant \text{RHS of (7.1)} = \text{LHS of (7.1)} \\ &\geqslant S_X^*(x) + S_Y^*(x) + S_Z^*(x) + S_W^*(x) \\ &\to 1 + \frac{\mu_B}{\mu_H} > 1, \quad \text{as } n \to \infty,\ x \to 0+. \end{aligned}$$

So $\bar{H}(x) > 0$ for sufficiently small $x > 0$. It then follows from Fatou's Lemma that $\tilde{\mu}_H \geqslant \mu_H \geqslant \int_0^\infty \bar{H}(x)\,\mathrm{d}x > 0$.

Since H is nonincreasing, its continuity points are dense in $\mathbb{R}$, so $H_{n_k}(x) \to H(x)$ a.e. in Lesbegue measure. Note that all the integrands in (7.1) and (7.5) are bounded (by 2), the limits of S_Y^* and S_Z^* are continuous distribution functions, and the integrands corresponding to $\mathrm{d}F_W^*$ are continuous in variable t. So (7.4) holds. Note that when $\tilde{\mu}_H = \infty$, the arguments follow with the interpretation $\infty/\infty = 1$. □

We complete the proof by showing that $\tilde{\mu}_H = \mu_F$, $H(x) = F(x)$ for $0 \leqslant x < b^*$ is the unique solution to (7.4) and (7.5), where $b^* = \sup\{b\colon B(b) < 1\}$. It is easy to see that μ_F and F is a solution, since

$$\begin{aligned} &\frac{2\bar{H}(x)}{\mu_F}\int_0^x \frac{\bar{F}(t)}{\bar{H}(t)}\bar{B}(t)\,\mathrm{d}t \\ &\quad + \int_{[0,x)} \frac{(x-t)\bar{H}(x) + \tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t}{\tilde{\mu}_H - \int_{[0,t)} \bar{H}(u)\,\mathrm{d}u} \frac{\int_t^\infty \bar{F}(u)\,\mathrm{d}u}{\mu_F}\,\mathrm{d}B(t) \end{aligned}$$

$$= \frac{2\bar{F}(x)}{\mu_F}\int_0^x \bar{B}(t)\,dt + \frac{B(x)}{\mu_F}\int_x^\infty \bar{F}(t)\,dt + \frac{\bar{F}(x)}{\mu_F}\int_{[0,x)}(x-t)\,dB(t)$$

$$= \frac{B(x)}{\mu_F}\int_x^\infty \bar{F}(t)\,dt + \frac{x\bar{F}(x)}{\mu_F} + \frac{\bar{F}(x)}{\mu_F}\int_0^x \bar{B}(t)\,dt$$

for any real x, so (7.4) holds. Letting $x \to \infty$ yields (7.5).

Lemma 7.3. *$F \ll H$ in $[0, b^*)$ and*

$$\beta(x) := \frac{\gamma(x)}{\int_x^\infty \bar{B}(t)\,dt} \in \frac{dF}{dH}, \quad 0 \leqslant x < b^*, \tag{7.6}$$

defines a version of the Radon–Nikodym derivatives, where

$$\gamma(x) = \frac{\mu_F}{\tilde{\mu}_H}x + \mu_B - 2\int_0^x \frac{\bar{F}(t)}{\bar{H}(t)}\bar{B}(t)\,dt - \int_0^x \frac{\int_t^\infty \bar{F}(u)\,du}{\tilde{\mu}_H - \int_0^t \bar{H}(u)\,du}(x-t)\,dB(t).$$

Proof. First we assume that F, H and B have continuous derivatives f, h and b in $[0, b^*)$. Differentiating (7.4) and solving for f/g we get the lemma easily. For the general case we can use integration by parts formula to rewrite (7.4) as

$$\int_0^x \int_t^\infty \bar{B}(u)\,du\,F(dt) = \int_0^x \gamma(t) H(dt).$$

From here we see $F \ll \int_0^{\cdot}\int_t^\infty \bar{B}(u)\,du\,\bar{F}(dt) \ll H$ and (7.6). Details are omitted here. □

Let $\Delta = H - F$, then

Corollary 7.4. *$\Delta \ll H$ in $[0, b^*)$ and*

$$1 - \beta(x) = \left(\int_x^\infty \bar{B}(t)\,dt\right)^{-1}\left\{\left(1 - \frac{\mu_F}{\tilde{\mu}_H}\right)x + 2\int_0^x \frac{\bar{B}(t)}{\bar{H}(t)}\Delta(t)\,dt\right.$$
$$\left. + \int_{[0,x)} \frac{x-t}{\tilde{\mu}_H - \int_0^t \bar{H}(u)\,du}\left(\mu_F - \tilde{\mu}_H - \int_0^t \Delta(u)\,du\right)dB(t)\right\}$$
$$\in \frac{d\Delta}{dH}, \quad 0 \leqslant x < b^*. \tag{7.7}$$

First we show that if the mean value is consistent then the survival function is also consistent.

Proposition 7.5. *If the pair $(H, \tilde{\mu}_H)$ satisfies (7.4) and $\tilde{\mu}_H = \mu_F$, then $H(x) = F(x)$ for $0 \leqslant x < b^*$.*

Proof. If $H \neq F$ in $[0, b^*)$, then $0 \leqslant x_0 := \inf\{0 \leqslant x < b^*: H(x) \neq F(x)\} < b^*$, and $\Delta(x) = 0$ for $0 \leqslant x \leqslant x_0$. By (7.7)

$$\begin{aligned}\Delta(x) = H(x) - F(x) &= \int_0^x [1 - \beta(t)] H(\mathrm{d}t) \\ &= \int_0^x \left(\int_t^\infty \bar{B}(w)\,\mathrm{d}w \right)^{-1} \left\{ 2 \int_0^t \frac{\bar{B}(u)}{\bar{H}(u)} \Delta(u)\,\mathrm{d}u \right. \\ &\quad \left. - \int_{[0,t)} \frac{t-u}{\tilde{\mu}_H - \int_0^u \bar{H}(w)\,\mathrm{d}w} \int_0^u \Delta(w)\,\mathrm{d}w\,\mathrm{d}B(u) \right\} H(\mathrm{d}t) \\ &= \int_0^x K(x,t) \Delta(t)\,\mathrm{d}t,\end{aligned}$$

where

$$\begin{aligned}K(x,t) = &\frac{2\bar{B}(t)}{\bar{H}(t)} \int_t^x \frac{H(\mathrm{d}u)}{\int_u^\infty \bar{B}(w)\,\mathrm{d}w} \\ &- \int_t^x \frac{1}{\int_u^\infty \bar{B}(w)\,\mathrm{d}w} \int_t^u \frac{u-v}{\tilde{\mu}_H - \int_0^v H(w)\,\mathrm{d}w} B(\mathrm{d}v) H(\mathrm{d}u).\end{aligned}$$

If $H(x_0+) = 0$, then for all $x_0 < x < b^*$, $H(x) = 1$ and $H(x) - F(x) = \Delta(x) = \int (x_0, x] [1 - \beta(t)] H(\mathrm{d}t) = 0$, contradicting the definition of x_0. Thus $H(x_0) < 1$.

For x around x_0, $H(x) < 1$ and $\tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t = v_H + \int_x^{b^*} \bar{H}(t)\,\mathrm{d}t \geqslant \int_x^{b^*} \bar{H}(t)\,\mathrm{d}t > 0$. So we can find $x_0 < x_1 < b^*$, $M > 1$ such that

$$\bar{H}(x),\ \tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t,\ \int_x^\infty \bar{B}(t)\,\mathrm{d}t > M^{-1}$$

for $x_0 < x < x_1$.

Now we see that $K(x,t)$ is bounded in $x_0 \leqslant x \leqslant x_1$, $0 \leqslant t \leqslant x$,

$$|K(x,t)| \leqslant 2M^2 + x_1 M =: M_1$$

and for $x_0 < x^* = \min\{x_1, x_0 + M_1^{-1}/2\}$

$$\begin{aligned}\sup_{x_0 \leqslant z \leqslant x^*} |\Delta(z)| &\leqslant \sup_{x_0 \leqslant z \leqslant x^*} \left| \int_{x_0}^z K(z,t) \Delta(t)\,\mathrm{d}t \right| \\ &\leqslant \left(\sup_{x_0 \leqslant z \leqslant x^*} |\Delta(z)| \right) M_1 (x - x_0) \\ &\leqslant 0.5 \left(\sup_{x_0 \leqslant z \leqslant x^*} |\Delta(z)| \right).\end{aligned}$$

Thus $\sup_{x_0 \leqslant z \leqslant x^*} |\Delta(z)| = 0$, which implies $\Delta(x) = 0$ for $0 \leqslant x \leqslant x^*$, contradicting the definition of x_0. □

Theorem 7.6. *Suppose F is continuous in $[0, b^*)$ and satisfies* (1.1). *Let F_n and $\tilde{\mu}_n$ be either the RT estimator or the NPMLEs of F and μ_F. If $\tilde{\mu}_n \to \mu_F$ then*

$$\sup_{0 \leqslant x \leqslant b^*} |F_n(x) - F(x)| \to 0.$$

Proof. By Proposition 7.5, $H_n \to F$ weakly. Since F is continuous, the pointwise convergence holds with probability 1. By Lemma 6.5 the uniform convergence follows. □

So the consistency of the NPMLEs and the RT estimators of F, μ_F depends on the consistency of $\tilde{\mu}_n$. We will show that this is true when b_i's are a fixed constant.

Assume $b_1 = b_2 = \cdots = b$. Then $\bar{B}(x) = I(x \leqslant b)$, $H(x) = 1$ for $x \geqslant b$, and (7.4), (7.7) become

$$\frac{b-x}{\mu_F}\bar{F}(x) + \frac{1}{\mu_F}\int_x^{\infty}\bar{F}(t)\,\mathrm{d}t + \frac{2\bar{H}(x)}{\mu_F}\int_0^x \frac{\bar{F}(t)}{\bar{H}(t)}\,\mathrm{d}t$$
$$= \frac{1}{\tilde{\mu}_H}\left(\tilde{\mu}_H - \int_0^x \bar{H}(t)\,\mathrm{d}t + x\bar{H}(x)\right) + \frac{b}{\mu_F}\bar{H}(x), \tag{7.8}$$

$$1 - \beta(x) = \frac{(1 - \mu_F/\tilde{\mu}_H)x + 2\int_0^x (\Delta(t)/\bar{H}(t))\,\mathrm{d}t}{b - x} \in \frac{\mathrm{d}\Delta}{\mathrm{d}H}$$

for $0 \leqslant x < b$. Since $w_1 = w_2 = \cdots = w_{h-1} = 0$ and $w_h = n_w$, (3.14) holds. Taking limit along n_k, we get

$$\frac{v_H}{\tilde{\mu}_H} = 1 - G(b). \tag{7.9}$$

Combining with $\tilde{\mu}_H = \int_0^b \bar{H}(x)\,\mathrm{d}x + v_H$ yields

$$\frac{\tilde{\mu}_H}{\mu_F} = \frac{\int_0^b \bar{H}(x)\,\mathrm{d}x}{\int_0^b \bar{F}(x)\,\mathrm{d}x}. \tag{7.10}$$

Now we are ready to prove

Proposition 7.7. *If the pair $(H, \tilde{\mu}_H)$ satisfies* (7.8) *and* (7.9), *then $\tilde{\mu}_H = \mu_F$.*

Proof. Observe that $1 - \beta(x) < 0$ iff

$$\left(1 - \frac{\mu_F}{\tilde{\mu}_H}\right)x + 2\int_0^x \frac{\Delta(t)}{\bar{H}(t)}\,\mathrm{d}t < 0. \tag{7.11}$$

By (7.4), H and Δ are right continuous at 0. If $\tilde{\mu}_H < \mu_F$, then (7.11) < 0 for all sufficiently small x, say $0 < x \leqslant \delta$, since the integral in (7.11) is of smaller order of magnitude than the first term as $x \to 0$. If $1 - \beta(x) > 0$ for some $0 \leqslant x < b$, then

$$0 < x_1 := \inf\{0 \leqslant x < b : 1 - \beta(x) > 0\} < b,$$

and $1 - \beta(x) \leqslant 0$ for all $0 \leqslant x \leqslant x_1$. Note that $\Delta(0) = 0$, therefore $\Delta(x) \leqslant 0$ for all such x. If $H(x_1) < 1$, then $\bar{H}(x) > 0$ for $x_1 \leqslant x \leqslant x_1 + \delta_1$ for sufficiently small $\delta_1 > 0$. For such x,

$$\int_0^x \frac{\Delta(t)}{\bar{H}(t)} \mathrm{d}t \leqslant \int_{x_1}^x \frac{\Delta(t)}{\bar{H}(t)} \mathrm{d}t \leqslant \frac{x - x_1}{\bar{H}(x_1 + \delta_1)}.$$

So

$$\left(1 - \frac{\mu_F}{\tilde{\mu}_H}\right) x + 2 \int_0^x \frac{\Delta(t)}{\bar{H}(t)} \mathrm{d}t \leqslant \left(1 - \frac{\mu_F}{\tilde{\mu}_H}\right) x_1 + \frac{2(x - x_1)}{\bar{H}(x_1 + \delta_1)} < 0$$

for all $x_1 < x < x_1 + \delta_2$ for some $0 < \delta_2 < \delta_1$. This contradicts the definition of x_1. On the other hand, if $H(x_1) = 1$, then $H(x) = 1$ for $x_1 < x < b$, and

$$\Delta(x) = \int_0^x [\beta(t) - 1] H(\mathrm{d}t) = \Delta(x_1) \leqslant 0,$$

$$1 - \beta(x) \leqslant \frac{x}{b - x}\left(1 - \frac{\mu_F}{\tilde{\mu}_H}\right) < 0,$$

i.e., $1 - \beta(x) < 0$ and $\Delta(x) \leqslant 0$ for $0 \leqslant x \leqslant b$. By (7.10),

$$\frac{\mu_F - \tilde{\mu}_H}{\mu_F} = \frac{\int_0^b \Delta(x)\,\mathrm{d}x}{\int_0^b \bar{F}(x)\,\mathrm{d}x} \leqslant 0,$$

which contradicts the assumption. Similarly we can show that $\tilde{\mu}_H > \mu_F$ cannot hold. Thus $\tilde{\mu}_H = \mu_F$. □

Remark 7.8. The same proof goes through if we only assume $b_n \to b$.

Theorem 7.9. *Suppose $b_n \to b > 0$, F is continuous in $[0, b)$ and satisfies* (1.1), *then with probability* 1,

$$\tilde{\mu}_n \to \mu_F,$$

$$\sup_{0 \leqslant x < b} |F_n(x) - F(x)| \to 0,$$

where F_n, $\tilde{\mu}_n$ are either the RT estimators or NPMLEs of F, μ_F. In the case $b_n = b$ for all n, the NPMLEs and the RT estimators coincide.

Acknowledgement

The authors thank the two referees for their careful reading and helpful comments.

References

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Billingsley, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

Chang, M.N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391–404.

Chang, M.N. and G.L. Yang (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15**, 1536–1547.

Dempster A.P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc.* **39**, 1–37.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd ed. Wiley, New York.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd ed. Wiley, New York.

Gill, R.D., Y. Vardi and J.A. Wellner (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069–1112.

Gill, R.D. (1989). Non and semi-parametric maximum likelihood estimators and the von Mises method (part I). *Scand. J. Statist.* **16**, 97–128.

Groeneboom, P. (1985). Estimating a monotone density. In: L.M. Le Cam and R.A. Olshen, Eds., *Proc. Berkeley Conf. in Honor of Jersy Neyman and Jack Kiefer*, Vol. 2. Wadsworth, Belmont, CA, 539–555.

Gu, M.G. and C.-H. Zhang (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611–624.

Kaplan, E.L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

Keiding, N. and R.D. Gill (1990). Random truncation models and Markov processes. *Ann. Statist.* **18**, 582–602.

Trai, W.-Y. and J. Crowley (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Statist.* **13**, 1317–1334.

Vardi, Y. (1982a). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10**, 616–620.

Vardi, Y. (1982b). Nonparametric estimation in renewal processes. *Ann. Statist.* **10**, 772–785.

Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13**, 178–203.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density. Technical Report, AT&T Statistical Research.

Vardi, Y. and C.-H. Zhang (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022–1039.

Wang. M.-C. (1987). Product limit estimates: a generalized maximum likelihood study. *Comm. Statist. Theory Methods* **16**, 3117–3132.

Wang, M.C., N.P. Jewell and W.Y. Tsai (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14**, 1597–1605.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163–177.

Wu, C.F.J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95–103.

Zangwill, W.I. (1969). Nonlinear Programming: a unified approach. Prentice-Hall.