Cure-rate estimation under Case-1 interval censoring

Running title: Cure-rate under interval censoring

Arusharka Sen, Department of Mathematics and Statistics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, H3G 1M8, Canada

Fang Tan, Department of Mathematics and Statistics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, H3G 1M8, Canada

ABSTRACT. We consider nonparametric estimation of cure-rate based on mixture model under Case-1 interval censoring. We show that the nonparametric maximum-likelihood estimator (NPMLE) of cure-rate is non-unique as well as inconsistent, and propose two estimators based on the NPMLE of the distribution function under this censoring model. We present a cross-validation method for choosing a ‘cut-off’ point needed for the estimators. The limiting distributions of the latter are obtained using extreme-value theory. Graphical illustration of the procedures based on simulated data are provided.

Key-words: Case-1 interval censoring, cross-validation, cure-rate, extreme-value theory, non-homogeneous Poisson process, nonparametric maximum-likelihood estimator, strong approximation, variance-bias trade-off.

1. Introduction

Consider a sample of individuals on each of whom some sort of time-to-event data is being collected, for instance, onset time of a disease following exposure to infection, time to death under a terminal disease, time (for criminals) to re-offend after at least one offence etc. In most such cases, there may be a possibility that the individual may be immune (e.g., not catch a disease) or get cured (e.g., cured of a disease or not re-offend). This is all the more relevant when the data is subject to some kind of ‘open-ended’ censoring such as random censoring, double censoring or interval censoring, where an individual being censored (i.e., event not occurred), especially after a large amount of time, points to the possibility of cure.
In the literature, the term long-term survival has also been used for cure.

Cure is usually quantified by the probability of cure, or the cure-rate: $p = P\{X = \infty\}$, where $X$ is the time-to-event of interest. Most of the statistical literature on cure is based on one of the two following models for the ‘improper’ random variable $X$: the mixture model, in which $P\{X > t\} = p + (1 - p)S_0(t)$, $S_0(\cdot)$ being a proper survival function representing the finite part of $X$ (Berkson and Gage, 1952); and the bounded cumulative hazard (BCH) model, in which $P\{X > t\} = \exp(-\int_0^t h(s)ds)$, with $\theta := \int_0^\infty h(s)ds < \infty$, so that $p = \exp(-\theta)$ (see Tsodikov et al (2003) for an excellent review). Inference, with or without (random) censoring, has been based mostly on either the Bayesian approach (see Yin and Ibrahim (2005) and the references therein) or a semi-parametric approach (see Zhao and Zhou (2006) and the references therein).

From the non-parametric point of view, it is clear that the two models above are equivalent. Notable among the nonparametric approaches are: Laska and Meisner (1992), who consider the NPMLE of $p$ under random censoring when a number $m \geq 1$ of cures are known; Maller and Zhou (1996), who consider the value of the Kaplan-Meier distribution function at the largest datum as an estimator of $(1 - p)$ (as is well-known, the value is less than unity if the largest datum is censored — an indication of cure). See Section 2 for more comments on these two works. Another interesting paper is Betensky and Schoenfeld (2001), who consider a time-to-cure, rather than just possibility of cure, competing with time-to-event/censoring.

In this paper we study estimation of cure-rate under Case-1 interval censoring, or current-status data, using the mixture model. We have been able to trace only one paper so far under this set-up, namely Lam and Xue (2005), who work with a semi-parametric model, allowing the cure-rate to depend on covariates via a logit function. We consider only the parameters $(F, p)$, the time-to-event distribution function and the cure-rate, respectively. Of course, this is a semi-parametric model too, but one without covariates. We show that the Maller-Zhou idea does not work here and propose two estimators of $p$ based on the usual (i.e., when $p = 0$) NPMLE of $F$, as given by Groeneboom and Wellner (1992). The asymptotics of the estimators are obtained using extreme-value theory.

In Section 2, we describe the Case-1 interval censoring model with cure-rate and show that the NPMLE of $p$ is non-unique and inconsistent. We then propose the two estimators that depend on a ‘cut-off’ point. Section 3 shows how to make an optimal choice of this cut-
off point, because it involves a variance-bias trade-off as in extremal index estimation (see, for instance, Embrechts et al. (1997)). In Section 4, limiting distributions of the estimators are derived. Use of the latter to construct confidence intervals for $p$ is straightforward.

2. Model, preliminary results and estimators

Consider a variable of interest $X$, say $X = \text{time to development of cancer following exposure to radiation}$ and an observation time $Y$, say $Y = \text{time of check-up}$. Under Case-1 interval censoring model, one observes the so-called ‘current status’ data

$$(\delta_i, Y_i), \ i = 1, 2, \ldots, n, \text{ where } \delta_i = I(X_i \leq Y_i),$$

and $Y_1, \ldots, Y_n$ are iid with distribution $G$, independent of $X_1, \ldots, X_n$ which are iid with distribution $F$. Suppose we want to estimate $F(x) = P\{X \leq x\}$. The nonparametric maximum likelihood estimator (NPMLE) is obtained by solving:

$$\max_{F} L(F_1, \ldots, F_n)$$

subject to $0 \leq F_1 \leq \ldots \leq F_n \leq 1$, \hspace{1cm} (1)

where

$$L(F_1, \ldots, F_n) = \sum_{i=1}^{n} (\delta_{[i]} \log(F_i) + (1 - \delta_{[i]}) \log(1 - F_i)),$$

and $F_i = F(Y_{(i)})$, $Y_{(i)} : \text{order-statistics for } (Y_1, \ldots, Y_n)$, $\delta_{[i]} = \text{concomitant of } Y_{(i)}, 1 \leq i \leq n$.

Solution is given by the ‘max-min’ formula of Groeneboom and Wellner (1992), namely,

$$\hat{F}_i = \max_{k \leq i} \min_{k \geq i} \sum_{j=k}^{k} \delta_{[j]}.$$

(2)

Cure-rate. Consider again $X = \text{time to cancer}$, this time with possibility of no cancer $\equiv \text{cure}$. Then $X$ can be modelled as an ‘extended’ real-valued random variable with a defective distribution, i.e.,

$$P(X = \infty) = p = \text{cure-rate} > 0$$

so that $P(X \leq t) = F_p(t) = (1 - p)F(t)$ and $P(X > t) = S_p(t) = p + (1 - p)(1 - F(t)) = p + (1 - p)S(t)$. In this case the likelihood function in Eq.(1) has to be modified as
max $L^c(p, F_1, \ldots, F_n)$ where

$$
L^c(p, F_1, \ldots, F_n) = \sum_{i=1}^{n} [\delta_i \log((1 - p)F_i) + (1 - \delta_i) \log(p + (1 - p)(1 - F_i))]
$$

subject to $0 \leq p \leq 1$, $0 \leq F_1 \leq \ldots \leq F_n \leq 1$

$$
= \sum_{i=1}^{n} [\delta_i \log(F_i) + (1 - \delta_i) \log(1 - F_i)]
$$

subject to $0 \leq p \leq 1$, $0 \leq F_1 \leq \ldots \leq F_n \leq (1 - p),

$$
(3)
$$
writing $F_i$ for $(1 - p)F_i$ in the last equality.

**Failure of NPMLE.** We state the following theorem whose proof is omitted because it is long and technical:

**THEOREM 1.** Let $L^c(p) = \max_{0 \leq F_1 \leq \ldots \leq F_n \leq (1 - p)} L^c(p, F_1, \ldots, F_n)$. Then

$$
L^c(p) = L(\hat{F}_1 \wedge (1 - p), \ldots, \hat{F}_n \wedge (1 - p)),
$$

where $\wedge$ denotes ‘minimum’ and $\hat{F}_i$, $1 \leq i \leq n$, are as in Eq.(2).

This leads to the following two observations about the NPMLE of $p$:

**REMARK 1:** NON-UNIQUENESS OF NPMLE. Obviously, $L^c(p)$ is non-increasing in $0 \leq p \leq 1$, and

$$
\sup_{0 \leq p \leq 1} L^c(p) = L(\hat{F}_1, \ldots, \hat{F}_n) = L^c(\hat{p}),
$$

for any $0 \leq \hat{p} \leq (1 - \hat{F}_n)$. Hence $\hat{p}$ is unique if and only if $(1 - \hat{F}_n) = 0 = \hat{p}$. This was also observed, in the case of random censoring, by Laska and Meisner (1992), who showed that NPMLE was unique and positive if, however, some number $m \geq 1$ of cases of cure were known. We shall explore this situation in a future paper.

**REMARK 2:** NON-CONSISTENCY OF NPMLE. Note that by Eq.(2),

$$
\hat{F}_n = \max_{i \leq n} \frac{\sum_{j=i}^{n} \delta_j}{n - i + 1},
$$

so that $\hat{F}_n = 1$ if and only if $\delta_{[n]} = 1$. Thus for $0 < p < 1$ and any $0 < \varepsilon < p$,

$$
P\{ |\hat{F}_n - (1 - p)| > \varepsilon \} \geq P\{ \hat{F}_n = 1 \} = P\{ \delta_{[n]} = 1 \} = (1 - p)E(F(Y_{(n)})) \rightarrow (1 - p)F(\tau_G),
$$
where $\tau_G = \sup\{y | G(y) = 1\}$. Hence $\hat{F}_n$ is not a consistent estimator of $(1 - p)$. This is in stark contrast to the case of random censoring where the former was shown to be in fact $\sqrt{n}$-consistent (asymptotically normal) by Maller and Zhou (1996).

**The proposed estimators.** Let us look at

$$\hat{F}_n = \max_{i \leq n} \frac{\sum_{j=1}^{n} \delta[j]}{n - i + 1} = \max_{x \leq Y^{(n)}} \frac{\sum_{j=1}^{n} \delta_j I(Y_j \geq x)}{\sum_{j=1}^{n} I(Y_j \geq x)}.$$  

Thus $\hat{F}_n$ is the maximum of the tail-averages of the concomitants, $\delta_{[i]}$, $1 \leq i \leq n$. Hence consider the ratio empirical process

$$p_{1n}(x) := \frac{\sum_{j=1}^{n} \delta_j I(Y_j \geq x)}{\sum_{j=1}^{n} I(Y_j \geq x)} \to p_1(x) := (1 - p) \frac{\int_{x}^{\infty} FdG}{\int_{x}^{\infty} dG}$$

almost surely for each $x \geq 0$ as $n \to \infty$. Moreover, note that

$$p_1(x) \uparrow (1 - p) \text{ as } x \uparrow \infty$$

and

$$p_{2n}(x) := \max_{y \leq x} p_{1n}(y) \to (1 - p) \max_{y \leq x} \frac{\int_{y}^{\infty} FdG}{\int_{y}^{\infty} dG} \uparrow (1 - p) \text{ as } x \uparrow \infty$$

These observations lead us to the following:

**Estimator-1.** Define

$$\hat{p}_{1n} = p_{1n}(x_n) = \frac{\sum_{j=1}^{n} \delta_j I(Y_j \geq x_n)}{\sum_{j=1}^{n} I(Y_j \geq x_n)}$$

i.e., tail-average at a suitable sequence $x_n \uparrow \infty$ of ‘cut-off’ points.

Figure 1 gives a sample-plot of $p_{1n}(i) \equiv p_{1n}(Y^{(i)}) = \sum_{j=i}^{n} \delta_{[j]}/(n - i + 1)$ against $1 \leq i \leq n$, for $p = 0.3$, $n = 100$. It is seen that for $i \approx 55$, $p_{1n}(i) \approx 0.7 = (1 - p)$. For comparison, a sample-plot for another sample with $p = 0$ (i.e., no cure) is also given.

**Estimator-2.** Define

$$\hat{p}_{2n} = p_{2n}(x_n) = \max_{y \leq x_n} p_{1n}(y),$$

i.e., partial maximum of the tail-averages (rather than the global maximum $\hat{F}_n$ which is inconsistent).

Figure 2 gives a sample-plot of $p_{2n}(i) = \max_{k \leq i} \sum_{j=k}^{n} \delta_{[j]}/(n - i + 1)$ against $1 \leq i \leq n$, for the same sample as in Figure 1. $p_{2n}(\cdot)$ looks more stable than $p_{1n}(\cdot)$, as is to be expected.

The choice of $x_n$ for a given sample of size $n$ is discussed in the next section.
Figure 1: sample plot of $p_{1n}(i)$ vs. $i$: $F = \text{Exp} (2)$, $G = \text{Exp} (1)$, $n = 100$, and $p = 0.3$ (solid line), $p = 0$ (broken line).
Figure 2: sample plot of $p_{2n}(i)$ vs. $i$: $F = \text{Exp} (2)$, $G = \text{Exp} (1)$, $n = 100$, and $p = 0.3$ (solid line), $p = 0$ (broken line).
3. Choice of cut-off point

Consider

\[ \hat{p}_{1n} - (1 - p) = (\hat{p}_{1n} - p_1(x_n)) + (p_1(x_n) - (1 - p)) \]
\[ = \frac{\sum_{j=1}^{n}[\delta_j I(Y_j \geq x_n) - (1 - p)(\int_{x_n}^{\infty} Fd\bar{G}(x_n))I(Y_j \geq x_n)]}{n\bar{G}(x_n)} \cdot \frac{\bar{G}(x_n)}{n^{-1}\sum_{j=1}^{n} I(Y_j \geq x_n)} \]
\[= - (1 - p) \int_{x_n}^{\infty} (1 - F)d\bar{G}/\bar{G}(x_n) \]
\[= A_n(x_n)C_n(x_n) - B_n(x_n), \text{ say, (4)} \]

where \( \bar{G}(x) = \int_{x}^{\infty} dG = 1 - G(x) \). Now

\[ n\bar{G}(x_n)(\text{var } A_n(x_n)) = (1 - p) \int_{x_n}^{\infty} Fd\bar{G}/\bar{G}(x_n) - [(1 - p)(\int_{x_n}^{\infty} Fd\bar{G}/\bar{G}(x_n))]^2 \to p(1 - p) \] (5)
as \( x_n \to \infty \).

Further, \( C_n(x_n) = O_P(1) \) (see Shorack and Wellner (1986), p.415) and \( B_n(x_n) = o(1) \) as \( x_n \to \infty \). Hence from Eq.(4),

\[ \hat{p}_{1n} - (1 - p) = (\hat{p}_{1n} - p_1(x_n)) + (p_1(x_n) - (1 - p)) = O_P((n\bar{G}(x_n))^{-1/2}) + o(1), \]
as \( x_n \to \infty \).

**Variance-bias trade-off.** Thus we have the following trade-off: as \( n \to \infty \), we must have \( x_n \uparrow \infty \) (so that the bias \(-B_n(x_n) \to 0\) and also \( \bar{G}(x_n) \to 0\)), but slowly enough so that \( n\bar{G}(x_n) \to \infty \) (i.e., \( \text{var } (A_n(x_n)) \to 0\)). A similar phenomenon occurs in the case of the Hill estimator of extremal index in extreme value theory (see Embrechts et al, 1997, p.341).

In view of Eq.(4)–(5), optimal order of \( x_n \uparrow \infty \) could be determined by minimizing, with respect to \( x \), the function

\[ M_n(x) = \frac{(p(1 - p)/n\bar{G}(x)) + (1 - p)^2(\int_{x}^{\infty} (1 - F)d\bar{G}/\bar{G}(x))^2}{n^{-1}}. \]

**Example 1.** Let \( F, G \) be Exponential (\( \lambda \)) and Exponential (\( \mu \)) distributions, respectively, i.e., \( \bar{F}(x) = 1 - F(x) = \exp(-\lambda x) \), \( \bar{G}(x) = 1 - G(x) = \exp(-\mu x) \). Then we have

\[ M_n(x) \]
\[ = \frac{(p(1 - p)/n\bar{G}(x)) + (1 - p)^2(\int_{x}^{\infty} (1 - F)d\bar{G}/\bar{G}(x))^2}{n^{-1}} \]
\[= n^{-1}p(1 - p)\exp(\mu x) + ((1 - p)\mu/(\lambda + \mu))^2\exp(-2\lambda x), \]
and \((d/dx)(M_n(x)) = 0\) gives
\[
n^{-1}p(1 - p)\mu \exp(\mu x) = ((1 - p)\mu/(\lambda + \mu))^22\lambda \exp(-2\lambda x),
\]
or
\[
x_n = (\mu + 2\lambda)^{-1}\log\left(\frac{(1 - p)\mu}{2p\lambda(\lambda + \mu)^2n}\right).
\]
Thus \(nG(x_n) = c(p, \lambda, \mu)n^{2\lambda/(\mu+2\lambda)}\), which shows that the optimal rate of convergence, \((nG(x_n))^{1/2} = O(n^{\lambda/(\mu+2\lambda)})\), is much slower than \(\sqrt{n}\).

**Cross-validation.** Eq.(4)–(5) also suggest that we could make a data-driven choice of \(x_n\), say \(\hat{x}_n\), as the minimizer of
\[
\hat{M}_n(x) := \hat{\text{var}}(A_n(x)) + \hat{B}_n^2(x)
\]
with respect to \(x\), where \(\hat{\text{var}}(A_n(x))\) and \(\hat{B}_n(x)\) denote suitable estimators of \(\text{var}(A_n(x))\) and \(B_n(x)\), respectively.

Now an obvious choice of \(\hat{\text{var}}(A_n(x))\) is
\[
\hat{\text{var}}(A_n(x)) = \frac{p_{2n}(x)(1 - p_{2n}(x))}{\sum_{j=1}^{n} I(Y_j \geq x)}, \quad (6)
\]
where we have used \(p_{2n}(\cdot)\) in view of its stability, as is evident from Figure-2. The choice of \(\hat{B}_n(x)\), however, is not clear in general. Let us therefore consider the special case of the Koziol–Green model of censoring:

**Assumption A.1.** \(1 - F(x) = (1 - G(x))^\alpha\) for some \(\alpha > 0\).

Under A.1, we have
\[
B_n(x) = -(1 - p)(1 - G(x))^\alpha/(\alpha + 1)
\]
\[
E(1 - \delta) = p + (1 - p)/(\alpha + 1)
\]
whence \((1 - p)/(\alpha + 1) = E(1 - \delta) - p\) \(\quad (8)\)
and \(\alpha = E(\delta)/[E(1 - \delta) - p]. \quad (9)\)

We then replace \(E(\delta)\) by \(\bar{\delta}_n := n^{-1}\sum_{i=1}^{n} \delta_i\) and \((1 - p)\) by
\[
\bar{p}_{2n} := n^{-1}\sum_{i=1}^{n} p_{2n}(Y_i) = \int p_{2n}(x)dG_n(x), \quad (10)
\]
where $G_n(\cdot)$ is the empirical distribution function of $Y_1, \ldots, Y_n$. This is motivated as follows: for $y \geq 0$,

$$
\int_y^\infty p_{2n}(x) dG_n(x) \approx (1 - p) \int_y^\infty \left[ \int_x^\infty F dG(x) \right] dG(x) \approx (1 - p) [1 - (\alpha + 1)^{-2}(1 - G(y))^\alpha],
$$

which has bias of a smaller order than $p_{2n}(y)$; to a first approximation, we let $y = 0$ to get $\bar{p}_{2n}$.

Thus by Eq.(6–10), we arrive at the following cross-validation function:

$$
\hat{M}_n^1(x) = \frac{p_{1n}(x)(1 - p_{1n}(x))}{\sum_{j=1}^{n} I(Y_j \geq x)} + (\bar{p}_{2n} - \bar{\delta}_n)^2 \left[ n^{-1} \sum_{j=1}^{n} I(Y_j \geq x) \right]^{2\hat{\alpha}}, \quad (11)
$$

where $\hat{\alpha} = \bar{\delta}_n / (\bar{p}_{2n} - \bar{\delta}_n)$, which could be minimized with respect to $x$ to obtain $\hat{x}_n$.

In general, motivated by Eq.(10) we could estimate the bias, $B_n(x) = (1 - p) \int_x^\infty F dG(x) - (1 - p)$, by $\hat{B}_n(x) := p_{2n}(x) - \bar{p}_{2n}$. This leads to another cross-validation function

$$
\hat{M}_n^2(x) = \frac{p_{1n}(x)(1 - p_{1n}(x))}{\sum_{j=1}^{n} I(Y_j \geq x)} + (p_{2n}(x) - \bar{p}_{2n})^2 \quad (12)
$$

Figure 3 gives sample-plots of $\hat{M}_n^l(i) \equiv \hat{M}_n^l(Y(i))$, $l = 1, 2$. Both the curves exhibit clear convex shapes with unique minima. However, $\hat{M}_n^1(\cdot)$ shows a spurious minimum at the upper extreme, which must be discarded. Further, the respective minimizers are seen to underestimate $(1 - p)$, so there appears to be scope for improvement here.
Figure 3: sample-plot of $\hat{M}_n(i)$ vs. $i$: $p = 0.3$, $F = \text{Exp (2)}$, $G = \text{Exp (1)}$, $n = 100$, $\hat{M}_n^1(\cdot)$ (solid line: minimizer $i_1 = 58$, ignoring $i = 100$, $p_{2n}(58) = 0.651$), $\hat{M}_n^2(\cdot)$ (broken line: minimizer $i_2 = 47$, $p_{2n}(47) = 0.611$)

4. Limiting distributions.

Eq.(5) suggests that $\hat{p}_{1n}$ would require a random norming, namely $\left(\sum_{j=1}^{n} I(Y_j \geq x_n)\right)^{1/2}$, for asymptotic normality. We establish this, as well as the limiting distribution of $\hat{p}_{2n}$, using the asymptotic theory of sample extremes. To this end, assume
Assumption A.2. $G(\cdot)$ belongs to the maximum domain of attraction of an extreme-value distribution $G_e(\cdot)$, i.e., there exist sequences of constants $a_n > 0$, $b_n$, $n \geq 1$, such that $G^n(a_n x + b_n) \to G_e(x)$, or equivalently $nG(a_n x + b_n) \to -\log(G_e(x))$, as $n \to \infty$, for each $x \in \mathbb{R}$.

It is well-known that, under A.2, \[ \sum_{j=1}^{n} I(Y_j \geq a_n x + b_n) \] converges weakly to a non-homogeneous Poisson process with mean-function \[ \Lambda(x) = -\log G_e(x) \]. It turns out that \[ \sum_{j=1}^{n} \delta_j I(Y_j \geq a_n x + b_n) \] converges to an (independently) thinned version of this process.

Lemma 1. With $\overset{d}{\to}$ denoting weak convergence in the space $D(\mathbb{R})$ of right-continuous functions on $\mathbb{R}$ with left-limits, we have, as $n \to \infty$,

(a) \( N_n(x) := \sum_{j=1}^{n} I(Y_j \geq a_n x + b_n) \overset{d}{\to} N(x) \equiv N([x, \infty)) \), a Poisson process with mean \( \Lambda(x) = -\log G_e(x) \);

(b) \( (N_{1n}(x), N_{0n}(x)) \overset{d}{\to} (N_1(x), N_0(x)) \), where \( N_{1n}(x) := \sum_{j=1}^{n} \delta_j I(Y_j \geq a_n x + b_n) \), \( N_{0n}(x) := \sum_{j=1}^{n} (1-\delta_j) I(Y_j \geq a_n x + b_n) \), and \( N_1(x), N_0(x) \) are independent Poisson processes with mean-functions \( \mu_1(x) := (1-p)\Lambda(x) \), \( \mu_0(x) := p\Lambda(x) \) respectively.

(c) Further, \( N_1(x) \overset{d}{=} \sum_{j=1}^{N(x)} \eta_j \), and \( N_0(x) \overset{d}{=} \sum_{j=1}^{N(x)} (1-\eta_j) \), where \( (\eta_1, \eta_2, \ldots) \) are iid Bernoulli \( 1-p \), independent of \( N(\cdot) \), and \( N(\cdot) \) is the Poisson process defined in Part (a) above.

Proof:

(a) This is a classical result. For a proof see, for instance, Embrechts et al. (1997).

(b) First, consider weak convergence of \( N_{1n}(x) \) alone. It is enough to verify convergence of the finite-dimensional distributions \( (N_{1n}(x_1), \ldots, N_{1n}(x_k)), \ k \geq 1 \) (see, for instance, Karr (1991), Theorem 1.21, p.14). For the sake of convenience let us consider just two points,
\((x_1, x_2)\) with \(x_1 < x_2\). Then with \(i = \sqrt{-1}\) and any real numbers \(t_1, t_2\),
\[
E[\exp(it_1 N_{1n}(x_1) + it_2 N_{1n}(x_2))] = (E[\exp(i \{t_1 I(Y_1 \ge a_n x_1 + b_n) + t_2 I(Y_1 \ge a_n x_2 + b_n)\})])^n
\]
\[
= \left( (p + (1 - p) \int_0^\infty (1 - F) dG + (1 - p) \int_{a_n x_1 + b_n}^{a_n x_2 + b_n} FdG) + e^{it_1} (1 - p) \int_{a_n x_1 + b_n}^{a_n x_2 + b_n} FdG \right)^n
\]
\[
= \left[ 1 + n^{-1} n \tilde{G}(a_n x_2 + b_n) \left\{ (1 - p) (e^{it_1} - 1) \int_{a_n x_1 + b_n}^{a_n x_2 + b_n} FdG / \tilde{G}(a_n x_2 + b_n) \right\} + (1 - p) (e^{it_1 + it_2} - 1) \int_{a_n x_2 + b_n}^{\infty} FdG / \tilde{G}(a_n x_2 + b_n) \right]^n
\]
\[
\to \exp \left( (1 - p) \Lambda(x_2) \left\{ (e^{it_1} - 1) (\Lambda(x_2) \Lambda^{-1}(x_2) - 1) + (e^{it_1 + it_2} - 1) \right\} \right),
\]
whence the result. Note that here we have used the fact that as \(n \to \infty\), \((a_n x + b_n) \to \tau_G\), so that \(\int_{a_n x + b_n}^{\infty} FdG / \tilde{G}(a_n x + b_n) \to 1\). The joint weak convergence of \((N_{1n}(x), N_{0n}(x))\), as well as their asymptotic independence, follow by exactly similar arguments.

(c) The representations of \((N_1(x), N_0(x))\) are obvious. □

Next note that
\[
p_{1n}(x_n) = \sum_{j=1}^{n} \delta_j I(Y_j \ge x_n) / \sum_{j=1}^{n} I(Y_j \ge x_n) = N_{1n}^\prime(x_n^\prime) / N_n(x_n^\prime),
\]
where \(x_n^\prime = (x_n - b_n) / a_n\). Therefore, in addition to the weak convergence in Lemma 1, we need strong approximation by a Poisson process. This follows in a straightforward way from Einmahl (1997) and is stated below:

**Theorem 2.** Under A.2, on some probability space one can construct the random variables \((\delta_i, Y_i), i = 1, 2, \ldots,\) and a sequence of Poisson processes \(N_n^\prime = (N_{1n}^\prime, N_{0n}^\prime)\) on \(\mathbb{R} \times \mathbb{R}\), where for each \(n \ge 1\), \(N_{1n}^\prime, N_{0n}^\prime\) are independent with mean-functions \(\mu_1(x), \mu_0(x)\), respectively, such that as \(n \to \infty\),
\[
\sup_{x : 0 < G(x) < 1} |N_{1n}(x) - N_{1n}^\prime(x)| \to 0,
\]
\[
\sup_{x : 0 < G(x) < 1} |N_{0n}(x) - N_{0n}^\prime(x)| \to 0.
\]

**Proof:** Follows by arguments similar to the proof of Corollary 2.6, p.37, of Einmahl (1997) □

We are now ready to state the limiting distributions of our estimators. In Theorem 3 below, by ‘lim’ we mean limit in distribution.

13
Theorem 3. Under A.2, if \( n \bar{G}(x_n) \to \infty \) as \( n \to \infty \), then

(a) \( \Lambda(x_n') \to \infty \), where \( x_n' = (x_n - b_n)/a_n \);

(b) let
\[
Z_{1n} = \frac{\left( \sum_{j=1}^{n} I(Y_j \geq x_n) \right)^{1/2}(p_{1n}(x_n) - (1 - p))}{\sqrt{p(1 - p)}},
\]
then
\[
\lim_{n \to \infty} Z_{1n} = \lim_{n \to \infty} \sqrt{N(x_n')} \left[ \sum_{j=1}^{N(x_n')} \eta_j \right] / \sqrt{p(1 - p)} = \text{Normal (0, 1)},
\]
where \( (\eta_1, \eta_2, \ldots) \) are iid Bernoulli \( (1 - p) \) as in Lemma 1, Part (c);

(c) let
\[
Z_{2n} = \frac{\left( \sum_{j=1}^{n} I(Y_j \geq x_n) \right)^{1/2}(p_{2n}(x_n) - (1 - p))}{\sqrt{p(1 - p)}},
\]
then
\[
\lim_{n \to \infty} Z_{2n} = \lim_{n \to \infty} \sqrt{N(x_n')} \sup_{x \leq x_n'} \left[ \sum_{j=1}^{N(x)} \eta_j \right] / \sqrt{p(1 - p)} = \text{half-Normal (0, 1)},
\]
where ‘half-Normal’ \( (0, 1) \) is the distribution of \( | \text{Normal (0, 1)} | \).

Proof:

(a) Since extreme-value distributions are all continuous, the convergence \( |G^n(a_n x + b_n) - G_0(x)| \to 0 \) is uniform in \( x \). Now \( n \bar{G}(x_n) \to \infty \Rightarrow G^n(x_n) = G^n(a_n x_n' + b_n) \to 0 \), hence \( G_0(x_n') \to 0 \). The result follows because \( \Lambda(x_n') = - \log G_0(x_n') \).

(b) Note that
\[
\lim_{n \to \infty} \frac{\left( \sum_{j=1}^{n} I(Y_j \geq x_n) \right)^{1/2}(p_{1n}(x_n) - (1 - p))}{\sqrt{p(1 - p)}}
= \lim_{n \to \infty} (N_{1n}(x_n') + N_{0n}(x_n'))^{1/2} \left[ \frac{N_{1n}(x_n')}{N_{1n}(x_n') + N_{0n}(x_n')} - (1 - p) \right] / \sqrt{p(1 - p)}
= \lim_{n \to \infty} (N_{1n}'(x_n') + N_{0n}'(x_n'))^{1/2} \left[ \frac{N_{1n}'(x_n')}{N_{1n}'(x_n') + N_{0n}'(x_n')} - (1 - p) \right] / \sqrt{p(1 - p)},
\]
by Theorem 2. The result now follows using the representation in Lemma 1, Part (c), and the random central limit theorem, since \((\eta_1, \eta_2, \ldots)\) are iid Bernoulli \((1 - p)\), independent of \(N(\cdot)\), and further, by Part (a) above, \(\Lambda(x'_n) \to \infty\), \(N(x'_n)/\Lambda(x'_n) \xrightarrow{P} 1\), as \(n \to \infty\).

(c) This result too follows as in Part (b) above, by noting that

\[
\lim_{n \to \infty} \sqrt{N(x'_n)} \sup_{x \leq x'_n} \left[ \frac{\sum_{j=1}^{N(x)} \eta_j}{N(x)} - (1 - p) \right] / \sqrt{p(1 - p)} = \lim_{n \to \infty} \sqrt{n} \sup_{m \geq n} \left[ \frac{\sum_{j=1}^{m} \eta_j}{m} - (1 - p) \right] / \sqrt{p(1 - p)}.
\]

Weak convergence of the sequence on right-hand-side to the half-Normal distribution is established in Robbins et al (1968) (see also Stute (1983) for a generalization to \(M\)-estimators).

**Remark 1.** Figures 4 and 5 give histograms of \(Z_{1n}\) and \(Z_{2n}\), respectively, based on 5000 samples each. Either of \(Z_{1n}\) and \(Z_{2n}\) may easily be used to construct confidence intervals for \((1 - p)\). However, note that limiting variance of \(Z_{1n} = 1 > 1 - 2\pi^{-1} = \) limiting variance of \(Z_{2n}\). Hence the latter may be a better choice. On the other hand, Figure-5 shows that the convergence of \(Z_{2n}\) to the half-Normal distribution is not very good.
Figure 4: histogram of studentized $Z_{1n}$: $p = 0.3$, $F = \text{Exp} (2)$, $G = \text{Exp} (1)$, $n = 100$, based on 5000 samples
Figure 5: histogram of studentized $Z_{2n}$: $p = 0.3$, $F = \text{Exp (2)}$, $G = \text{Exp (1)}$, $n = 100$, based on 5000 samples

Acknowledgements: This research was supported by a Discovery grant of NSERC, Canada, to A.Sen.

REFERENCES

Berkson, J. & Gage, R.P. (1952). Survival curve for cancer patients following treatment. *J. Amer. Statist. Assoc.* **47**, 501-515.
Betensky, R.A. & Schoenfeld, D.A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics* **57**, 282–286.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling extremal events. For insurance and finance.* Springer-Verlag, Berlin.

Groeneboom, P. & Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation.* Birkhäuser Verlag, Basel.

Karr, A.F. (1991). *Point processes and their statistical inference*, 2nd ed. Marcel Dekker, New York.

Lam, K. F. & Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika* **92**, 573–586.

Laska, E.M. & Meisner, M.J. (1992). Nonparametric estimation and testing in a cure rate model. *Biometrics* **48**, 1223-1234.

Maller, R.A. & Zhou, X. (1996). *Survival analysis with long-term survivors.* Wiley, Chichester.

Robbins, H., Siegmund, D. & Wendel, J. (1968). The limiting distribution of the last time $s_n \geq n\varepsilon$. *Proc. Nat. Acad. Sci.* **61**, 1228–1230.

Shorack, G.R. & Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

Stute, W. (1983). Last passage times of $M$-estimators. *Scand. J. Statist.* **10**, 301–305.

Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *J. Amer. Statist. Assoc.* **98**, 1063–1078.

Yin, G. & Ibrahim, J.G. (2005). Cure rate models: a unified approach. *Canad. J. Statist.* **33**, 559–570.

Zhao, X. & Zhou, X. (2006). Proportional hazards models for survival data with long-term survivors. *Statist. Probab. Lett.* **76**, 1685–1693.