東海大學統計學研究所

碩士論文

指導教授 : 沈葆聖 教授

THE SIZE OF THE RISK SET UNDER RANDOM TRUNCATION

研究生 : 張國陽

中華民國九十二年七月

致謝

本論文之完成, 最先要感激沈葆聖教授在這兩年來多來的辛苦指導, 無論是在論文或其他課業上, 均經恩師提供寶貴的指導, 使我在學習與研究過程中獲益良多, 在此先致上最誠心的謝意。

論文口試期間, 承蒙鄭順林教授與戴政教授於百忙之中詳爲審閱, 並提供寶貴建議, 不但使我受益匪淺, 而且使得本論文的內容更加完善, 在此致上由衷的感謝, 感謝教授的評審。

再來我要感謝我女朋友張婉菁小姐, 在這段兢兢戰戰的日子中, 只有妳能當我的避風港, 讓我遠離一些狂風暴雨襲擊, 如果沒有妳, 這趟知識之旅將過得更加苦澀, 謝謝妳跟我在一起, 感謝妳。

最後我要感謝我的母親、大弟、小弟與妹妹沒有你們在經濟上的支持, 我不可能順利唸到研究所, 將來我一定用我最大的努力來報答你們, 尤其是媽媽, 您眞的辛苦了, 畢業後我一定要改善您的生活, 謝謝您把我拉拔長大。

THE SIZE OF THE RISK SET UNDER RANDOM TRUNCATION

Kuo-Yang Chang
Dept. of Statistics
Tunghai University
Taichung, 40704
Taiwan, R. O. C.

June 10, 2003

Contents

THE SIZE OF THE RISK SET UNDER RANDOM TRUNCATION

SUMMARY

Let U^* and V^* be two independent positive random variables with continuous distribution functions F and G. Let (a_f, a_g) and (b_f, b_g) denote the lower and upper boundaries of (U^*, V^*), respectively. Under left truncation, both U^* and V^* are observable only when $U^* \geq V^*$. Let $(U_1, V_1), \ldots, (U_n, V_n)$ denote the truncated sample. Let $N_F(u) = \sum_{i=1}^n I_{[U_i \leq u]}$, $N_G(v) = \sum_{i=1}^n I_{[V_i \leq v]}$, and the size of the risk set $R_n(u) = N_G(u) - N_F(u-) = \sum_{i=1}^n I_{[V_i \leq u \leq U_i]}$, where $I_{[A]}$ is the indicator function of the event A. The nonparametric maximum likelihood estimate (NPMLE) of $F(x)$ and $G(x)$ are given by

$$\hat{F}_n(x) = 1 - \prod_{u \leq x}\Big[1 - \frac{dN_F(u)}{R_n(u)}\Big],$$

and

$$\hat{G}_n(x) = \prod_{v > x}\Big[1 - \frac{dN_G(v)}{R_n(v)}\Big],$$

where $dN_F(u) = N_F(u) - N_F(u-)$ and $dN_G(v) = N_G(v) - N_G(v-)$. Let $\mathcal{K} = \{(F,G) : F(0) = G(0) = 0, \alpha(F,G) > 0\}$, where $\alpha(F,G) = \int_0^\infty G(z)dF(z) = \int_0^\infty [1 - F(z)]dG(z)$. When $(F,G) \in \mathcal{K}$ and $a_f \geq a_g$, $b_f \geq b_g$, the consistency results for the estimate $\hat{F}_n$ and $\hat{G}_n$ were proved by Woodroofe (1985). Let $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ denote the distinct ordered statistics of the sample $U_i's$. In applying $\hat{F}_n(x)$, a practical difficulty arises when $R_n(U_{(i)}) = 1$ for some $i \leq n-1$. Woodroofe (1985, Corollary 5) showed that when $(F,G) \in \mathcal{K}$, the probability $P(R_n(U_{(i)})) = 1$ for some $i \leq n-1$ converges to 0 as $n \to \infty$. In this note, we derive the exact probability of $P(R_n(U_{(1)}) = k)$ for $k = 1, \ldots, n$ and give an alternative proof of $\lim_{n\to\infty} P(R_n(U_{(1)}) = k) = 0$ for $1 \leq k < \infty$. Simulation results indicate that the probability $P(R_n(U_{(1)}) = 1)$ can be significant when $a_f - a_g$ is not sufficiently large.

Key words: risk set;truncated data.

CHAPTER 1. INTRODUCTION

Let U^* and V^* be two independent positive random variables with continuous distribution functions F and G. Let (a_f, a_g) and (b_f, b_g) denote the lower and upper boundaries of (U^*, V^*), respectively. Under left truncation, both U^* and V^* are observable only when $U^* \geq V^*$. Truncated data occur in astronomy,(e.g., Lynden-Bell (1971), Woodroofe (1985)), epidemiology, biometry (e.g., Wang, Jewell and Tsai (1986), Tsai, Jewell and Wang (1987)) and possibly in other field such as economics. The follwing examples describes situtations where the models of left truncation are appropriate.

Example 1 (retirement data):

Channing House is a retirement center located in Palo Alto, California. Data on ages at death of 462 individuals (97 males and 365 females), who were in residence during the period January 1964 to July 1975, has been reported by Hyde (1980). The life lengths in this data set are left-truncated because an individual must survive to a sufficient age to enter the retirement community. The truncation variable V^*, is then the potential patient's age at entry, and the target variable U^*, is the patient's age at death. Obviously we can only observe (U^*, V^*) if $U^* \geq V^*$.

Example 2 (AIDS blood-transfusion data):

The blood transfusion related AIDS data given by Kalbfleisch and Lawless (1989). They gives infection times V^*, in months with 1 representing January 1978, incubation times T in months, and age in years for 34 'children' aged 0 to 4 years, 120 'adults' aged 5 to 59 years, and 141 'elderly' aged 60 and over, who were infected by contaminated blood transfusions and developed AIDS by 1 July 1986. Let $U^* = 102 - T$. The truncation effect comes from the fact that we only observed over the period $(0, 102]$. An individual is observed if and only if $T + V^* \leq 102$ or $V^* \leq U^*$.

Let $(U_1, V_1), \ldots, (U_n, V_n)$ denote the truncated sample. Define

$N_F(u) = \sum_{i=1}^{n} I_{[U_i \leq u]}$, $N_G(v) = \sum_{i=1}^{n} I_{[V_i \leq v]}$, and the size of the risk set $R_n(u) =$

$N_G(u) - N_F(u-) = \sum_{i=1}^n I_{[V_i \le u \le U_i]}$, where $I_{[A]}$ is the indicator function of the event A. The nonparametric maximum likelihood estimate (NPMLE) of $F(x)$ and $G(x)$ (see Wang (1987)) are given by

$$\hat{F}_n(x) = 1 - \prod_{u \le x}\Big[1 - \frac{dN_F(u)}{R_n(u)}\Big],$$

and

$$\hat{G}_n(x) = \prod_{v > x}\Big[1 - \frac{dN_G(v)}{R_n(v)}\Big],$$

where $dN_F(u) = N_F(u) - N_F(u-)$ and $dN_G(v) = N_G(v) - N_G(v-)$. Let $\mathcal{K} = \{(F, G) : F(0) = G(0) = 0, \alpha(F, G) > 0\}$, where $\alpha(F, G) = \int_0^\infty G(z)dF(z) = \int_0^\infty [1 - F(z)]dG(z)$. The justifications of using $\hat{F}_n(x)$ and $\hat{G}_n(x)$ are given as follows,

$F^*(x) = P(U_i \le x) = P(U^* \le x | U^* \ge V^*) = P(V^* \le U^* \le x)/\alpha(F, G) = \int_0^x G(u)dF(u) = G(x)dF(x)/\alpha(F, G)$

$R(x) = P(V_i \le x \le U_i) = P(V^* \le x \le U^* | U^* \ge V^*) = [\alpha(F, G)]^{-1}G(x)[1 - F(x)]$

Hence,$\frac{dF^*(X)}{R_n(X)} = \frac{dF(x)}{1-F(x)} = d\Lambda(x)$,where $\Lambda(x) = \int_0^x \frac{dF(u)}{1-F(u)}$

Note that

$$1 - F(x) = \prod_{u \le x}\Big[1 - d\Lambda(u)\Big]$$

$$= \prod_{u \le x}\Big[1 - \frac{dF(u)}{1 - F(u_-)}\Big] = \prod_{u \le x}\Big[1 - \frac{dF^*(u)}{R(u)}\Big]$$

Now $dF^*(u)$ can be consistently estimated by $\frac{dN_F(u)}{n} = \frac{N_F(u) - N_F(u-)}{n}$ and R(u) can be consistently estimated by $\frac{R_n(u)}{n} = \frac{\sum_{i=1}^n I_{[V_i \le u \le U_i]}}{n}$. This justified the use of $\hat{F}_n(x)$ and $\hat{G}_n(x)$.

When $(F, G) \in \mathcal{K}$ and $a_f \ge a_g$ $b_f \ge b_g$, the consistency results for the estimate $\hat{F}_n$ and $\hat{G}_n$ were proved by Woodroofe (1985). Let $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ denote the distinct ordered statistics of the sample $U_i's$. Note that $R_n(x)/n$ is a consistent estimator of $R(x) = [\alpha(F,G)]^{-1}G(x)[1-F(x)]$. The $R(x)$ is not monotone in x and converge to zero if $G(x) \to 0$ or $F(x) \to 1$. Especially, in applying $F_n(x)$, a practical difficulty arises when $R_n(U_{(i)}) = 1$ for some $i \leq n-1$. Since $R_n(U_{(i)}) = 1$ for some $i \leq n-1$, then $F_n(U_{(i)}) = 1$. This is a disturbing property of the estimators. Furthermore, even when $R_n(U_{(i)}) > 1$ for all $i \leq n-1$, $\hat{F}_n$ may still be a very poor estimator (badly biased and large variance) of F for moderate sample sizes because of the small risk set size $R_n(x)$ for x near a_g (see Woodroofe 1985 Lemma 2, and a simulation study of Lai and Ying (1991, pages 440-441).

Woodroofe (1985, Corollary 5) showed that when $(F,G) \in \mathcal{K}$,

(i) $P(R_n(U_{(i)}) = 1 \; for \; some \; i \leq n-1)$ converges to 0 as $n \to \infty$ and

(ii) $\min\{R_n(U_{(i)}) : 1 \leq i \leq (1-\epsilon)n\} \to \infty$ in probability as $n \to \infty$ for all ϵ, $0 < \epsilon < 1$.

The proof of the second assertion (ii) is given by Woodroofe (1985, page 172). The proof of the first assertion is only briefly described by Woodroofe (1985). We now give the detailed proof of the first assertion.

proof of the first assertion:

First, we show that $R_n(U_{(i)}) = 1$ implies that $R_n(V_{(i+1)}) = 1$, where $V_{(i+1)}$ denote the $(i+1)^{th}$ order statistic of the sample $V_i's$. Note that

$$\sum_{j=1}^{n} I_{[V_j \leq U_{(i)} \leq U_j]} = \sum_{j=i}^{n} I_{[\tilde{V}_{(j)} \leq U_{(i)} \leq U_{(j)}]}, \tag{1.1}$$

where $\tilde{V}_{(j)}$ is the concomitant of $U_{(j)}$. Hence, (1.1) $= 1$ implies that $U_{(i)} < \tilde{V}_{(j)}$ for $j = i+1, \ldots, n$ and $U_{(i)} > \tilde{V}_{(j)}$ for $j = 1, \ldots, i-1$. Hence, $V_{(i+1)} = \tilde{V}_{(k)}$ for some $k \in \{j = i+1, \ldots, n\}$. Thus, we have

$$\sum_{j=1}^{n} I_{[\tilde{V}_{(j)} \leq V_{(i+1)} \leq U_{(j)}]} = \sum_{j=1}^{n} I_{[\tilde{V}_{(j)} \leq \tilde{V}_{(k)} \leq U_{(j)}]}$$

$$= \sum_{j=1}^{i} I_{[\tilde{V}_{(j)} \leq \tilde{V}_{(k)} \leq U_{(j)}]} + I_{[\tilde{V}_{(k)} \leq \tilde{V}_{(k)} \leq U_{(k)}]} = 0 + 1 = 1$$

Now, Let $X_j = 1/U_j$ and $Y_j = 1/V_j$. Hence, we have

$$R_n(U_{(i)}) = R_n(V_{(i+1)}) = \sum_{j=1}^{n} I_{[V_j \le V_{(i+1)} \le U_j]}$$

$$= \sum_{j=1}^{n} I_{[\frac{1}{U_j} \le \frac{1}{V_{(i+1)}} \le \frac{1}{V_j}]} = \sum_{j=1}^{n} I_{[X_j \le \frac{1}{V_{(i+1)}} \le Y_j]}$$

$$= \sum_{j=1}^{n} I_{[X_j \le Y_{(n-i)} \le Y_j]}.$$

Hence, $R_n(U_{(i)}) = 1$ for $i = 1, \ldots, n-1$ is equivalent to $R_n(Y_{(n-i)}) = 1$ for $i = 1, \ldots, n-1$. The first assertion then follows from the second assertion by letting $\epsilon = 1/n$.

For randomly censored data, Maller and Zhou (1993) gived necessary and sufficient conditions for the probability that the largest censored data is zero. Note that the Kaplan-Meier estimator (1958) for the survival function of randomly censored time-to-event data is improper when the largest observation is censored. In this note, we derive the exact probability of $P(R_n(U_{(1)}) = k)$ for $k = 1, \ldots, n$. Motivated by Maller and Zhou (1993) we give an alternative proof of $\lim_{n\to\infty} P(R_n(U_{(1)}) = k) = 0$ for $1 \le k < \infty$. Simulation results indicate that the probability $P(R_n(U_{(1)}) = 1)$ can be significant when $a_f - a_g$ is not sufficiently large.

CHAPTER 2. THE PROBABILITY $P(R_n(U_{(1)}) = k)$

The following Lemma dervies the exact probability $P(R_n(U_{(1)}) = k)$

Lemma 2.1:

For $k = 1, \ldots, n$,

$$P(R_n(U_{(1)}) = k) = \int_{\max\{a_f, a_g\}}^{b_f} P(B(x) = k-1) dF_{U_{(1)}}(x),$$

where $B(x)$ is binomial random variable with parameter $n-1$ and the probability of success $p(x) = P(V_i < x < U_i)/P(U_i > x) = P(V^* < x < U^*)/P(U^* > x, U^* > V^*)$, and $F_{U_{(1)}}(x) = P(U_{(1)} \leq x) = [P(U_i \leq x)]^n = [P(U^* \leq x | U^* \geq V^*)]^n$.

proof:

Let F_u and G_v denote the distribution function of U_i and V_i, respecitvely. That is, $F_u(x) = P(U_i \leq x) = P(U^* \leq x | U^* \geq V^*)$ and $G_v(x) = P(V_i \leq x) = P(V^* \leq x | U^* \geq V^*)$ For $j = 1, \ldots, n$, let $D_j = \{(U_i, V_i) : i = 1, \ldots, j-1, j+1, \ldots, n\}$ denote the set of the observations when $U_{(j)}$ is deleted from the sample. Given $k = 1, \ldots, n$,

$$P(R_n(U_{(1)}) = k) =$$

$$\sum_{j=1}^{n} P(U_j < \min_{s \in D_j} U_s, \quad \max_{s=i_1,\ldots,i_{k-1}} V_s < U_j < \min_{s \in D_j, s \neq i_1,\ldots,i_{k-1},} V_s, \ for \ i_1, \ldots, i_{k-1} \in D_j)$$

$$= nP(\max_{s=i_1,\ldots,i_{k-1}} V_s < U_1 < \min_{s=i_1,\ldots,i_{k-1}} U_s, \ U_1 < \min_{s \in D_1, s \neq i_1,\ldots,i_{k-1},} V_s, \ for \ i_1, \ldots, i_{k-1} \in D_1)$$

$$= n \binom{n-1}{k-1} \int_{\max\{a_f, a_g\}}^{b_f} [P(V_i < x < U_i)]^{k-1} [P(V_i > x)]^{n-k} dF_u(x)$$

$$= \int_{\max\{a_f, a_g\}}^{b_f} \binom{n-1}{k-1} \left[\frac{P(V_i < x < U_i)}{P(U_i > x)}\right]^{k-1} \left[\frac{P(V_i > x)}{P(U_i > x)}\right]^{n-k} n[P(U_i > x)]^{n-1} dF_u(x). \tag{2.1}$$

Since $P(V_i > x) = P(U_i > V_i > x)$, we have $P(V_i > x)/P(U_i > x) = P(V_i > x | U_i > x) = 1 - P(V_i < x | U_i > x)$

$= 1 - P(V_i < x < U_i)/P(U_i > x)$. Note that $n[P(U_i > x)]^{n-1}dF_u(x) = dF_{U_{(1)}}(x)$, where $F_{U_{(1)}}(x) = P(U_{(1)} \leq x)$. Hence, (2.1) can be expressed as

$$\int_{\max\{a_f,a_g\}}^{b_f} \binom{n-1}{k-1} [p(x)]^{k-1}[1-p(x)]^{n-k} dF_{U_{(1)}}(x)$$

$$= \int_{\max\{a_f,a_g\}}^{b_f} P(B(x) = k-1) dF_{U_{(1)}}(x).$$

The proof is completed.

The following Lemma shows that $\lim_{n\to\infty} P(R_n(U_{(1)}) = k) = 0$ for $1 \leq k < \infty$.
Lemma 2.2:

Suppose that $(F, G) \in \mathcal{K}$ then $\lim_{n\to\infty} P(R_n(U_{(1)}) = k) = 0$ for $1 \leq k < \infty$.

proof:

First, we consider the case $a_f > a_g$ and $k = 1$. Note that $a_f > a_g$ implies that $\alpha(F, G) > 0$. From Lemma 2.1, when $a_f > a_g$, we obtain

$$P(R_n(U_{(1)}) = 1) = n \int_{a_f}^{b_f} [P(V_i > x)]^{n-1} dF_u(x) \leq n[P(V_i > a_f)]^{n-1}.$$

Since $P(V_i > a_f) < 1$, we have $n[P(V_i > a_f)]^{n-1} \to 0$, as $n \to \infty$. the proof is completed.

Next, we consider the case $a_f \leq a_g$.

Define $\tilde{U}_i = b_f - U_i$ and $\tilde{V}_i = b_f - V_i$. Let F_1 and F_2 denote the distribution function of $\tilde{V}_i$ and $\tilde{U}_i$. When $a_f \leq a_g$, $P(R_n(U_{(1)}) = 1)$ can be written as

$$P(R_n(U_{(1)}) = 1) = n \int_0^{b_f - a_g} [F_1(x)]^{n-1} dF_2(x)\circ$$

Since $F_1(x) < 1$ for $x < b_f - a_g$, according to Lemma 2.3 of Maller and Zhou (1993),

$$\lim_{n\to\infty} n \int_0^{b_f - a_g} [F_1(x)]^{n-1} dF_2(x) = L \;\; if\ and\ only\ if \;\; \lim_{x \uparrow b_f - a_g} \frac{1 - F_2(x)}{1 - F_1(x)} = L.$$

This is equivalent to

$$\lim_{x\downarrow a_g}\frac{F_u(x)}{G_v(x)}=\lim_{x\downarrow a_g}\frac{\int_{a_g}^x G(z)dF(z)}{\int_{a_g}^x G(z)dF(z)+G(x)\bar{F}(x)}=\frac{1}{1+\frac{G(x)\bar{F}(x)}{\int_{a_g}^x G(z)dF(z)}}=L.$$

Note that

$$\frac{G(x)\bar{F}(x)}{\int_{a_g}^x G(z)dF(z)}=\frac{\bar{F}(x)}{\int_{a_g}^x \frac{G(z)}{G(x)}dF(z)}\geq\frac{\bar{F}(x)}{\int_{a_g}^x dF(z)}=\frac{\bar{F}(x)}{F(x)-F(a_g)}.$$

Since $\lim_{x\downarrow a_g} F(x)-F(a_g)=0$ and $\lim_{x\downarrow a_g}\bar{F}(x)=\bar{F}(a_g)>0$, we have

$\lim_{x\downarrow a_g}\frac{F_u(x)}{G_v(x)}=0$.

Hence, $\lim_{n\to\infty}P(R_n(U_{(1)})=1)=0$.

Next, we consider the case $k>1$. From Lemma 2.1, we have

$$\begin{aligned}
&P(R_n(U_{(1)})=k)\\
&=n\binom{n-1}{k-1}\int_{\max\{a_f,a_g\}}^{b_f}[P(V_i<x<U_i)]^{k-1}[P(V_i>x)]^{n-k}dF_u(x)\\
&\leq n\binom{n-1}{k-1}\int_{\max\{a_f,a_g\}}^{b_f}[1-F_u(x)]^{k-1}[1-G_v(x)]^{n-k}dF_u(x)\\
&=n\binom{n-1}{k-1}\int_{\max\{a_f,a_g\}}^{b_f}[F_2(b_f-x)]^{k-1}[F_1(b_f-x)]^{n-k}dF_2(b_f-x)\\
&=n\binom{n-1}{k-1}\int_{0}^{b_f-\max\{a_f,a_g\}}[F_2(y)]^{k-1}[F_1(y)]^{n-k}dF_2(y).
\end{aligned}$$

Since $F_1(x)<1$ for $x<b_f-\max\{a_f,a_g\}$, according to (2.9) and (2.10) of Maller and Zhou (1993),

$$\lim_{n\to\infty}n\binom{n-1}{k-1}\int_{0}^{b_f-\max\{a_f,a_g\}}[F_2(y)]^{k-1}[F_1(y)]^{n-k}dF_2(y)=L^k$$

if and only if

$$\lim_{x\uparrow b_f-\max\{a_f,a_g\}}\frac{1-F_2(x)}{1-F_1(x)}=L.$$

This is equivalent to $\lim_{x\downarrow\max\{a_f,a_g\}} \frac{F_u(x)}{G_v(x)} = L$. Similar to the argument for $k = 1$, we have $\lim_{x\downarrow\max\{a_f,a_g\}} F(x) - F(\max\{a_f, a_g\}) = 0$ and $\lim_{x\downarrow\max\{a_f,a_g\}} \bar{F}(x) = \bar{F}(\max\{a_f, a_g\}) > 0$. The proof is completed.

From Lemma 2.1,

$$P(R_n(U_{(1)}) = 1) = n \int_{\max\{a_f,a_g\}}^{b_f} [P(V_i > x)]^{n-1} dF_u(x) \leq n[P(V_i > \max\{a_f, a_g\})]^{n-1}.$$

Hence, when $a_f > a_g$ and $[P(V_i > a_f)]^{n-1}$ is sufficiently small, the probability $P(R_n(U_{(1)}) = 1)$ is negligible. However, when $a_f \leq a_g$, this probabiltiy can be signifcant. Similarly, for $k > 1$,

$$P(R_n(U_{(1)}) = k) \leq n\binom{n-1}{k-1} \int_{\max\{a_f,a_g\}}^{b_f} [P(V_i > x)]^{n-k} dF_u(x)$$

$$\leq n\binom{n-1}{k-1}[P(V_i > \max\{a_f, a_g\})]^{n-k}.$$

Hence, for small values of k, when $a_f \leq a_g$, this probabiltiy can be signifcant.

CHAPTER 3. SIMULAITON RESULTS AND DISCUSSION

Next, simulation study was conducted to investigate the probability

$P(R_n(U_{(1)}) = 1)$. The U^*'s are left-truncated Weibull distributed:

$U^* \sim LW(a_f, \delta)$, that is $F(x) = 1 - e^{-(x-a_f)^\delta}$ for $x \geq a_f$ with varying parameters $a_f = 0.3(0.3)0.9$ and $\delta = 0.25, 1.0, 4.0$. The V^*'s are uniform distributed: $V^* \sim U(a_g, \delta_g)$ with $a_g = 0.5$ and $b_g = 5.0$. Sample sizes are chosen as $n = 10, 25$ and 50. The replication is 10000 times. Table 1 lists the probability $P(R_n(U_{(1)}) = 1)$.

Table 1. Simulation results of $P(R_n(U_{(1)}) = 1)$:$V_i \sim U(0.5, 5.0)$

n	δ	$a_f = 0.3$	$a_f = 0.6$	$a_f = 0.9$
10	0.25	0.105	0.248	0.117
25	0.25	0.049	0.141	0.005
50	0.25	0.037	0.050	0.000
10	1.00	0.098	0.071	0.018
25	1.00	0.033	0.014	0.000
50	1.00	0.018	0.002	0.000
10	4.00	0.015	0.002	0.000
25	4.00	0.003	0.000	0.000
50	4.00	0.001	0.000	0.000

Simulation results indicate that when $n = 50$ or ($n = 25$ and $a_f - a_g = 0.6$), the probability $P(R_n(U_{(1)}) = 1)$ is close to zero. However, when $n = 10$, $a_f - a_g = -0.2, 0.1$ and $\delta = 0.25, 1$, the probability is larger than 0.05. Lai and Ying (1991) suggest a solution to this problem by a slight modification of the NPMLE $\hat{F}_n$ where deaths are ignored when the risk set is small. Their estimator is given by

$$\tilde{F}_n(x) = 1 - \prod_{u \leq x} \Big[1 - I_{[R_n(u) \geq cn^p]} \frac{dN_F(u)}{R_n(u)} \Big],$$

where $c > 0$ and $0 < p < 1$. This estimator $\tilde{F}_n(u)$ is asymptotically equivalent to the NPMLE $\hat{F}_n$. For finite sample, further investigation is needed to compare the two estimators.

REFERENCES

Hyde, J. (1980), Survival analysis with incomplete observation. In Biostatistics Casebook, Miller R. G., Efron B., Brown B. W. and Moses, L. E., eds. New York:John Wiley and Sons, 31-46.

Kalbfleisch, J. D., and Lawless, J. F. (1989), Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association* **84**, 360-372.

Lai, T. L. and Ying, Z. (1991), Estimation a ditribution function with truncated and censored data. *Ann. Statist.* **19**, No. 1, 417-442.

Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Not. R. Astr. Soc.* **155**, 95-118.

Maller, R. A. and Zhou, S. (1993) The probability that the largest observation is censored. *Journal of Applied Probability* **30**, 602-615.

Tsai, W.-Y., Jewell, N. P. and Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74** 883-886.

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

Wang, M.-C. Product-limit estimates: a generalized maximum likelihood study. Communi. in Statist., Part A- Theory and Methods, **1987**, *6*, 3117-3132.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics* **13**, 163-167.