J. Japan Statist. Soc.
Vol. 34 No. 2 2004 153–172

UNBIASED ESTIMATION OF FUNCTIONALS UNDER RANDOM CENSORSHIP

Akio Suzukawa*

This paper is intended as an investigation of estimating functionals of a lifetime distribution F under right censorship. Functionals given by $\int \varphi dF$, where φ's are known F-integrable functions, are considered. The nonparametric maximum likelihood estimator of F is given by the Kaplan-Meier (KM) estimator F_n, where n is sample size. A natural estimator of $\int \varphi dF$ is a KM integral, $\int \varphi dF_n$. However, it is known that KM integrals have serious biases for unbounded φ's. A representation of the KM integral in terms of the KM estimator of a censoring distribution is obtained. The representation may be useful not only to calculate the KM integral but also to characterize the KM integral from a point view of the censoring distribution and the biasedness. A class of unbiased estimators under the condition that the censoring distribution is known is considered, and the estimators are compared.

Key words and phrases: Censored data, Kaplan-Meier estimator, mean lifetime, product-limit estimator, survival data.

1. Introduction

Problems concerning right censored data are discussed in this paper. The Kaplan-Meier estimator (KM estimator) developed by Kaplan and Meier (1958) is the most fundamental tool to analyze such data, and it gives the nonparametric maximum likelihood estimator of lifetime distribution as shown by Kaplan and Meier (1958) and Johansen (1978).

Let $X_1, \ldots, X_n$ be i.i.d. positive random variables with a distribution function F (survival function $\bar{F} = 1 - F$). These random variables represent lifetimes and F is a lifetime distribution. Let $Y_1, \ldots, Y_n$ be i.i.d. positive random variables with a distribution function G (survival function $\bar{G} = 1 - G$) and independent of X_i's. They represent censoring time and G is a censoring distribution. In the randomly right censored data, the pairs $(X_i, Y_i), i = 1, \ldots, n$ are not observed. One observes the pairs $(Z_i, \delta_i), i = 1, \ldots, n$, where

$$Z_i = \min(X_i, Y_i) \quad \text{and} \quad \delta_i = I(X_i \le Y_i),$$

with I denoting the indicator function.

KM estimator of F is given by

$$1 - F_n(x) = \bar{F}_n(x) = \prod_{i: Z_{i:n} \le x} \left(1 - \frac{\delta_{[i:n]}}{n - i + 1}\right).$$

Received December 1, 2003. Revised July 2, 2004. Accepted July 5, 2004.

*Graduate School of Economics and Business Administration, Hokkaido University, Kita 9, Nishi 7, Kita-ku, Sapporo 060-0809, Japan.

Here, $Z_{1:n} \leq \cdots \leq Z_{n:n}$ are the ordered Z-values, where ties within lifetimes or within censoring times are ordered arbitrarily and ties among lifetimes and censoring times are treated as if the former preceed the latter, and $\delta_{[i:n]}$ is the concomitant of the i-th order statistic, that is, $\delta_{[i:n]} = \delta_j$ if $Z_{i:n} = Z_j$.

There have been several studies on the properties of the KM estimator (e.g., Efron (1967), Breslow and Crowley (1974), Gill (1980, 1983), Reid (1981), Wang (1987)). Results of studies are given in textbooks by Fleming and Harrington (1991), Andersen *et al.* (1993) and Maller and Zhou (1996). Generally, the KM estimator has asymptotically desirable properties such as uniform consistency and asymptotic normality. Although the KM estimator has negative bias, the bias converges to zero at an exponential rate as $n \to \infty$. Therefore, the bias is not so serious problem.

Incidentally, it is important to estimate not only F itself but also some functionals of F. For example, moments of lifetime, particularly mean lifetime, may be important in characterization of the lifetime distribution. In such cases, we need to estimate a functional $\int x^k dF(x)$, where k is a positive integer. To determine how far F is from a known specific distribution with density f_0, the functional $\int \log f_0(x) dF(x)$ must be estimated.

In this paper, estimation of functionals of the form $\int \varphi dF$, where φ's are known F-integrable functions, is considered. For a given φ, a natural estimator of $\int \varphi dF$ is obtained by plugging F_n into F, that is, $\int \varphi dF_n$, which is called a Kaplan-Meier integral (KM integral). Since the KM estimator F_n is a discrete (sub-)distribution, the KM integral can be expressed as

$$\int \varphi dF_n = \sum_{i=1}^{n} \{F_n(Z_{i:n}) - F_n(Z_{i:n}-)\} \varphi(Z_{i:n}) = \sum_{i=1}^{n} W_{i:n} \varphi(Z_{i:n}), \tag{1.1}$$

where for $1 \leq i \leq n$,

$$W_{i:n} = \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}}$$

is the mass attached to the i-th order statistic $Z_{i:n}$ under F_n.

It should be noted that under no censorship does the KM estimator reduce to the empirical distribution of X_i's and the KM integral is nothing but an arithmetic average of $\varphi(X_i)$'s, i.e., $n^{-1} \sum_{i=1}^{n} \varphi(X_i)$. It is strongly consistent and asymptotically normal by the strong law of large numbers and the central limit theorem. Moreover, it is unbiased. The point to emphasize here is the unbiasedness.

Let H be the distribution function of $Z_i = \min(X_i, Y_i)$, and set

$$\tau_F = \inf\{x\,;\, F(x) = 1\}, \quad \tau_G = \inf\{y\,;\, G(y) = 1\} \quad \text{and} \quad \tau = \inf\{z\,;\, H(z) = 1\}.$$

Since $H = 1 - (1-F)(1-G)$, it holds that $\tau = \min(\tau_F, \tau_G)$. The observable time $Z_i = \min(X_i, Y_i)$ never exceeds τ. Thus, if $\tau_F > \tau$ (i.e., $\tau_F > \tau_G$), then we can not obtain information about $F(x)$ for $\tau < x \le \tau_F$. In other words, we can know only about $F(x)$ for $x \le \tau$.

Throughout this paper, $\int_a^b$ means integration on a half interval $(a, b\,]$, i.e., $\int_a^b = \int_{(a,b\,]}$. For reasons mentioned above, we can not estimate $\int_0^\infty \varphi(x)dF(x)$. The estimable functional is $\int_0^\tau \varphi(x)dF(x)$. If $\tau_F \le \tau$, they are the same. On the other hand, if $\tau_F > \tau$, they are different. However, in many practical cases, $\tau_F = \tau_G = \tau = \infty$, and then these two functionals are the same.

Stute and Wang (1993) showed that under conditions in which F and G do not have jumps in common and F does not have a jump at τ,

$$\lim_{n\to\infty} \int \varphi dF_n = \int_0^\tau \varphi dF \quad \text{with probability 1.}$$

That is, $\int \varphi dF_n$ is a strongly consistent estimator of $\int_0^\tau \varphi dF$. The distributional convergence of the KM integral was investigated by Gill (1983), Schick *et al.* (1988), Yang (1994) and Stute (1995). Gill (1983) showed the distributional convergence for φ's that are nonnegative, continuous and increasing. For such a class of φ's, Schick *et al.* (1988) obtained a weak representation of the KM integral in terms of a sum of i.i.d. random variables plus a remainder. Yang (1994) extended the distributional convergence, under regularity conditions on F, to those φ's satisfying $\int \varphi^2/\bar{G}dF < \infty$. Stute (1995) obtained a representation of the KM integral as a sum of i.i.d. random variables plus a remainder that is valid under no regularity conditions on F and G.

Distributional convergence under continuous F and G is presented here. When F and G are continuous, under $\int \varphi^2/\bar{G}dF < \infty$,

$$n^{1/2}\left(\int \varphi dF_n - \int_0^\tau \varphi dF\right) \xrightarrow{\text{d}} N(0, \sigma^2) \quad \text{as } n \to \infty,$$

where

$$\sigma^2 = \int_0^\tau \varphi^2/\bar{G}dF - \left(\int_0^\tau \varphi dF\right)^2 - \int \left(\int_x^\tau \varphi dF\right)^2 \frac{\bar{F}(x)}{\{\bar{H}(x)\}^2} dG(x). \tag{1.2}$$

We are concerned with a bias of $\int \varphi dF_n$ in estimating $\int_0^\tau \varphi dF$. Mauro (1985) showed that for nonnegative φ's,

$$\text{Bias}\left(\int \varphi dF_n\right) = E\left[\int \varphi dF_n\right] - \int_0^\tau \varphi dF \le 0.$$

Zhou (1988) obtained a lower bound of the bias for nonnegative and continuous φ's:

$$-\int \varphi(x)\{H(x)\}^n dF(x) \le \text{Bias}\left(\int \varphi dF_n\right) \le 0.$$

Stute (1994) derived an expansion of the bias and showed that the bias decreases to zero exponentially fast as $n \to \infty$ if φ is bounded and vanishes right of some

$T < \tau$. Since the KM estimator itself is a KM integral for $\varphi =$ indicator, this result is a generalization of the result that the bias of the KM estimator itself converges to zero at an exponential rate, which is mentioned above. Moreover Stute (1994) pointed out that the bias may decrease to zero at a rate slower than $n^{-1/2}$ when $0 \leq \varphi(x) \uparrow \infty$ as $x \to \infty$ and censoring is heavy. In estimating mean lifetimes, $\varphi(x) = x$ has this property. Thus, it is important to reduce the bias. Stute and Wang (1994) proposed a jackknife modification of $\int \varphi dF_n$. They showed, by a simulation study, that the jackknifing may lead to a reduction of the bias. However, they also reported that the jackknifing leads to an increase in variance.

In this paper, 'unbiased' estimation of $\int_0^\tau \varphi dF$ in the case where the censoring distribution G is known is considered. A situation in which G is known is not so unreal. Two examples are considered.

One example is a clinical trial (see Figure 1). The beginning time of the study is set to zero, and entry of patients into the study is received until time $a > 0$. Let E_i be the entry time of the i-th patient. It seems reasonable to suppose that E_i is uniformly distributed on $[0, a]$. After additional follow-up of $b - a$ (> 0) time units, the data are analyzed. Then, the censoring time for the i-th patient is $b - E_i$, which is uniformly distributed on $[b - a, b]$. Thus, in this case, the censoring distribution is completely known.

Another example is the two-competing risks problem. A system consists of two sub-systems, A and B. Failure of either sub-system A or B damages the whole system. The observable time is the failure time of the whole system, which is a minimum of failure times of A and B. We are interested in failure time of sub-system A. Then, the failure time of B is censoring time. Supposing the failure time distribution of B is known experientially, the censoring distribution is known.

The question is whether the information that G is known is useful in estimating $\int_0^\tau \varphi dF$. We should notice that the likelihood based on the censored data can be decomposed into two parts. One part depends only on F, and the KM estimator is derived by maximizing it. Another depends only on G. Thus, the maximum likelihood estimator of F is given by the KM estimator even if G is known. This fact suggests that the known G is not useful asymptotically.

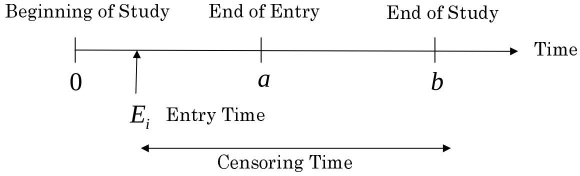

Figure 1. Entry time and censoring time in a clinical trial.

However, there is no conclusive proof that it is not useful in the case of finite samples. As we shall see later in the next section, when G is known, $\int_0^\tau \varphi dF$ can be estimated without bias. The purpose of this study is to obtain reasonable unbiased estimators of $\int_0^\tau \varphi dF$ by using the information that G is known.

2. Representations of Kaplan-Meier integrals

In this section, a representation of $\int \varphi dF_n$ in terms of a KM estimator of G is given. The representation will suggest us an unbiased estimator of $\int_0^\tau \varphi dF$ when G is known.

The KM estimator of G is given by

$$1 - G_n(y) = \bar{G}_n(y) = \prod_{j=1}^{n} \left(1 - \frac{1 - \delta_{[j:n]}}{n - j + 1}\right)^{I(Z_{j:n} \le y)}. \tag{2.1}$$

Proposition 1. *For any F-integrable function $\varphi : [0, \infty) \to R$,*

$$\int \varphi dF_n = n^{-1} \sum_{i=1}^{n} \frac{\delta_i \varphi(Z_i)}{\bar{G}_n(Z_i -)} \quad \textit{and} \quad \int \varphi dG_n = n^{-1} \sum_{i=1}^{n} \frac{(1 - \delta_i)\varphi(Z_i)}{\bar{F}_n(Z_i)}.$$

Proof. Only the first equation is shown. The second equation can be shown similarly. Since the KM integral $\int \varphi dF_n$ is given by (1.1), it suffices to show that

$$\delta_{[i:n]} / \bar{G}_n(Z_{i:n} -) = W_{i:n}, \quad i = 1, \dots, n. \tag{2.2}$$

If $\delta_{[i:n]} = 0$, both sides are surely zero. Thus, we should consider only the case of $\delta_{[i:n]} = 1$.

From (2.1), we have

$$\begin{aligned}\bar{G}_n(Z_{i:n} -) &= \prod_{j=1}^{n} \left(1 - \frac{1 - \delta_{[j:n]}}{n - j + 1}\right)^{I(Z_{j:n} < Z_{i:n})} = \prod_{j=1}^{n} \left(\frac{n - j}{n - j + 1}\right)^{(1 - \delta_{[j:n]}) I(Z_{j:n} < Z_{i:n})} \\ &= \left\{\prod_{j=1}^{n} \left(\frac{n - j}{n - j + 1}\right)^{I(Z_{j:n} < Z_{i:n})}\right\} \times \left\{\prod_{j=1}^{n} \left(\frac{n - j}{n - j + 1}\right)^{\delta_{[j:n]} I(Z_{j:n} < Z_{i:n})}\right\}.\end{aligned}$$

From this, when $\delta_{[i:n]} = 1$,

$$\begin{aligned}\frac{1}{\bar{G}_n(Z_{i:n} -)} = &\left\{\prod_{j=1}^{n} \left(\frac{n - j + 1}{n - j}\right)^{I(Z_{j:n} < Z_{i:n})}\right\} \\ &\times \left\{\prod_{j=1}^{n} \left(\frac{n - j + 1}{n - j}\right)^{\delta_{[j:n]} I(Z_{j:n} < Z_{i:n})}\right\}.\end{aligned} \tag{2.3}$$

Suppose that there are $k+l+1$ $(k,l \geq 0)$ ties at $Z_{i:n}$ as

$$Z_{i-k-1:n} < \underbrace{Z_{i-k:n} = \cdots = Z_{i-1:n}}_{k} = Z_{i:n} = \underbrace{Z_{i+1:n} = \cdots = Z_{i+l:n}}_{l} < Z_{i+l+1:n}.$$

Then the first term of (2.3) is

$$\prod_{j=1}^{i-k-1} \left(\frac{n-j+1}{n-j} \right) = \frac{n}{n-i+k+1}.$$

Since ties among lifetimes and censoring times are treated as if the former precedes the latter, $\delta_{[i-k:n]} = \cdots = \delta_{[i-1:n]} = 1$. Thus,

$$\begin{aligned} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} &= \left\{ \prod_{j=1}^{i-k-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} \right\} \times \left\{ \prod_{j=i-k}^{i-1} \left(\frac{n-j}{n-j+1} \right) \right\} \\ &= \left\{ \prod_{j=1}^{i-k-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} \right\} \times \frac{n-i+1}{n-i+k+1}. \end{aligned}$$

Therefore, the second term of (2.3) is

$$\prod_{j=1}^{i-k-1} \left(\frac{n-j+1}{n-j} \right)^{\delta_{[j:n]}} = \frac{n-i+k+1}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}}.$$

As a consequence, if $\delta_{[i:n]} = 1$

$$\frac{1}{\bar{G}_n(Z_{i:n}-)} = \frac{1}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} = W_{i:n}. \qquad \square$$

The representations given by Proposition 1 may be useful from the viewpoint of calculation of KM integrals. The KM integral $\int \varphi dF_n$ $(\int \varphi dG_n)$ can be calculated by using the opposite side KM-survival function $\bar{G}_n$ $(\bar{F}_n)$. KM-survival functions can be easily obtained by almost any statistical software.

Since X_i and Y_i are independent and $\bar{H}$ is the survival function of $Z_i = \min(X_i, Y_i)$, it holds that $\bar{H}(z) = \bar{F}(z)\bar{G}(z)$. Denote the empirical survival function of Z_i's by $\bar{H}_n$, i.e., $\bar{H}_n(z) = n^{-1} \sum_{i=1}^{n} I(z < Z_i)$. This is an estimator of $\bar{H}$. Using the representations of Proposition 1, we can easily show the well-known identity $\bar{F}_n \bar{G}_n = \bar{H}_n$ as is shown in Theorem 9.1 of Maller and Zhou (1996).

For any fixed $z \geq 0$, let $\varphi(x) = \bar{G}_n(x-)I(x > z)$ and $\varphi(x) = \bar{F}_n(x)I(x > z)$ in the first and the second equations of Proposition 1, respectively. Then, we have

$$\int_z^\infty \bar{G}_n(x-)dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_i I(Z_i > z) \quad \text{and}$$

$$\int_z^\infty \bar{F}_n(x)dG_n(x) = \frac{1}{n} \sum_{i=1}^{n} (1-\delta_i) I(Z_i > z).$$

The sum of the right-hand sides is $\bar{H}_n(z)$. On the other hand, the sum of the left-hand sides is $-\bar{F}_n(\infty)\bar{G}_n(\infty)+\bar{F}_n(z)\bar{G}_n(z)$ by the integration by parts formula as can be seen in Theorem A.1.2 of Fleming and Harrington (1991). The first term is zero since one of $\bar{F}_n(\infty)$ and $\bar{G}_n(\infty)$ is surely zero. (If the largest observation is censored, then $\bar{F}_n(\infty) > 0$ but $\bar{G}_n(\infty) = 0$.)

Define a functional of G as

$$S_{0n}(G) = n^{-1}\sum_{i=1}^{n} \delta_i\varphi(Z_i)/\bar{G}(Z_i-). \tag{2.4}$$

Then, by Proposition 1, the KM integral $\int \varphi dF_n$ is given by $S_{0n}(G_n)$ (i.e., plugging of G_n into $S_{0n}(G)$), which is a strongly consistent estimator of $\int_0^\tau \varphi dF$ as shown by Stute and Wang (1993). It, however, has a bias as mentioned above. If G is known, there is no need to plug G_n into $S_{0n}(G)$. In fact, $S_{0n}(G)$ is unbiased since

$$E[S_{0n}(G)] = E[\delta_i\varphi(Z_i)/\bar{G}(Z_i-)] = \int_0^\tau \bar{G}(x-)\varphi(x)/\bar{G}(x-)dF(x) = \int_0^\tau \varphi dF.$$

We call $S_{0n}(G)$ a 'simple' unbiased estimator. We can say that the bias of the KM integral is due to the plugging of G_n into $S_{0n}(G)$.

Variance of the simple unbiased estimator is given by

$$n \times \mathrm{Var}[S_{0n}(G)] = \int_0^\tau \frac{\{\varphi(x)\}^2}{\bar{G}(x-)}dF(x) - \left\{\int_0^\tau \varphi dF\right\}^2.$$

Although no explicit expressions of exact variance of the KM integral have been obtained yet, its asymptotic variance is given by σ^2 of (1.2). We can easily see that $S_{0n}(G)$ has asymptotically larger variance than KM integral. (This will be reconfirmed in the next section.) This means that plug-in of G_n asymptotically decreases variance. Intuitively, it seems strange. An interpretation of the unnaturalness will be also given in the next section.

3. A class of unbiased estimators

In this section, unbiased estimation of $\theta = \int_0^\tau \varphi dF$ in the case where G is known will be discussed. We consider a class of estimators

$$\hat{\theta}_n(\varphi_1, \varphi_0) = n^{-1}\sum_{i=1}^{n} \{\delta_i\varphi_1(Z_i) + (1-\delta_i)\varphi_0(Z_i)\}, \tag{3.1}$$

where $\int |\varphi_1|dG < \infty$ and $\int |\varphi_0|dG < \infty$, and φ_1 and φ_0 are independent of the unknown F.

If we choose the two functions as

$$\varphi_1(z) = \varphi(z)/\bar{G}_n(z-) \quad \text{and} \quad \varphi_0(z) \equiv 0,$$

then, by Proposition 1, $\hat{\theta}_n(\varphi_1, \varphi_0)$ is the KM integral. However it is biased.

PROPOSITION 2. *For any lifetime distribution* F, $\hat{\theta}_n(\varphi_1, \varphi_0)$ *defined by* (3.1) *is unbiased estimators of* $\theta = \int_0^\tau \varphi dF$ *if* φ_1 *and* φ_0 *satisfy the two conditions*

$$\bar{G}(z-)\varphi_1(z) + \int_0^{z-} \varphi_0(y)dG(y) = \varphi(z) \quad \text{for any} \quad 0 < z \leq \tau \tag{3.2}$$

and

$$\int_0^{\tau_G} \varphi_0(y)dG(y) = 0 \quad \text{if} \quad \tau_G = \tau. \tag{3.3}$$

PROOF. The bias of $\hat{\theta}_n(\varphi_1, \varphi_0)$ can be expressed as

$$\begin{aligned}
&E[\hat{\theta}_n(\varphi_1, \varphi_0)] - \theta \\
&= E[\delta_i\varphi_1(Z_i) + (1-\delta_i)\varphi_0(Z_i)] - \theta \\
&= \int_0^\tau \bar{G}(z-)\varphi_1(z)dF(z) + \int_0^\tau \bar{F}(y)\varphi_0(y)dG(y) - \int_0^\tau \varphi(z)dF(z) \\
&= \int_0^\tau \bar{G}(z-)\varphi_1(z)dF(z) + \int_0^\tau \left\{\int_y^\tau dF(z) + \bar{F}(\tau)\right\}\varphi_0(y)dG(y) - \int_0^\tau \varphi(z)dF(z) \\
&= \int_0^\tau \left\{\bar{G}(z-)\varphi_1(z) + \int_0^{z-} \varphi_0(y)dG(y) - \varphi(z)\right\} dF(z) \\
&\quad + \bar{F}(\tau)I(\tau_G = \tau)\int_0^{\tau_G} \varphi_0(y)dG(y).
\end{aligned}$$

In order for $\hat{\theta}_n(\varphi_1, \varphi_0)$ to be unbiased, without being dependent on F, the integrand in the first term must be zero and the integral in the second term must be zero when $\tau_G = \tau$. □

A trivial solution of the unbiasedness conditions in Proposition 2 is obtained by letting $\varphi_0 \equiv 0$, which always satisfies (3.3). Then, by (3.2), φ_1 is given by $\varphi_1(z) = \varphi(z)/\bar{G}(z-)$, which gives the simple unbiased estimator $S_{0n}(G)$ defined by (2.4).

More general solutions for the unbiasedness conditions are given by the following proposition.

PROPOSITION 3. *Assume that* γ *is a function on* $[0, \tau]$, *which satisfies*

$$\int_0^{z-} \frac{|\gamma(y)|}{\bar{G}(y)} dG(y) < \infty \quad \text{for any} \quad 0 < z \leq \tau \tag{3.4}$$

and

$$\gamma(\tau_G)\{G(\tau_G) - G(\tau_G-)\} = 0 \quad \text{if} \quad \tau_G = \tau. \tag{3.5}$$

Define

$$\varphi_1(z) = \frac{\varphi(z)}{\bar{G}(z-)} - \int_0^{z-} \frac{\gamma(y)}{\bar{G}(y)} dG(y) \quad \text{and} \quad \varphi_0(z) = \gamma(z) - \int_0^{z-} \frac{\gamma(y)}{\bar{G}(y)} dG(y).$$

Then, these functions fulfill the unbiasedness conditions (3.2) *and* (3.3).

PROOF. For any $0 < z \le \tau$,

$$\bar{G}(z-)\varphi_1(z) = \varphi(z) - \bar{G}(z-)\int_0^{z-} \frac{\gamma(y)}{\bar{G}(y)} dG(y),$$

$$\int_0^{z-} \varphi_0(y)dG(y) = \int_0^{z-} \gamma(y)dG(y) - \int_0^{z-} \left\{ \int_0^{y-} \frac{\gamma(x)}{\bar{G}(x)} dG(x) \right\} dG(y)$$

$$= \int_0^{z-} \gamma(y)dG(y) - \int_0^{z-} \frac{\gamma(x)}{\bar{G}(x)} \left\{ \int_x^{z-} dG(y) \right\} dG(x)$$
$$= \int_0^{z-} \gamma(y)dG(y) - \int_0^{z-} \frac{\gamma(x)}{\bar{G}(x)} \{\bar{G}(x) - \bar{G}(z-)\} dG(x)$$
$$= \bar{G}(z-)\int_0^{z-} \frac{\gamma(x)}{\bar{G}(x)} dG(x).$$

By adding both sides, we can see that (3.2) is satisfied. By the same calculation, we get

$$\int_0^{\tau_G} \varphi_0(y)dG(y) = \gamma(\tau_G)\{G(\tau_G) - G(\tau_G-)\}.$$

From assumption (3.5), this becomes zero if $\tau_G = \tau$. □

For any γ with (3.4), assumption (3.5) is always fulfilled if G does not have a jump at τ_G. Thus, when G is continuous, assumption (3.5) is unnecessary.

Substituting φ_1 and φ_0 of Proposition 3 into (3.1), we obtain a class of unbiased estimators

$$(3.6) \qquad U_n(\gamma; G) = n^{-1}\sum_{i=1}^{n} \left\{ \delta_i \frac{\varphi(Z_i)}{\bar{G}(Z_i-)} + (1-\delta_i)\gamma(Z_i) - \int_0^{Z_i-} \frac{\gamma(y)}{\bar{G}(y)} dG(y) \right\}.$$

The first term is nothing but the simple unbiased estimator $S_{0n}(G)$, and its expectation is θ. Expectation of the remainder is zero since expectations of the second and the third terms are the same. Thus, the estimator $U_n(\gamma; G)$ can be interpreted as

simple unbiased estimator + mean-zero variate.

If we set $\gamma \equiv 0$, which always fulfills (3.4) and (3.5), then $U_n(0; G) = S_{0n}(G)$.

In the previous section, we have seen that the KM integral is obtained by plugging G_n into $S_{0n}(G)$, i.e., $S_{0n}(G_n) = \int \varphi dF_n$. This can be also expressed as $U_n(0; G_n) = \int \varphi dF_n$. It is generalized as follows.

PROPOSITION 4. *For any* γ, $U_n(\gamma; G_n) = \int \varphi dF_n$.

PROOF. Since $n^{-1}\sum_{i=1}^{n}\delta_i\varphi(Z_i)/\bar{G}_n(Z_i-)=\int\varphi dF_n$, we have

$$U_n(\gamma;G_n)=\int\varphi dF_n+n^{-1}\sum_{i=1}^{n}(1-\delta_i)\gamma(Z_i) - n^{-1}\sum_{i=1}^{n}\int_0^{Z_i-}\frac{\gamma(y)}{\bar{G}_n(y)}dG_n(y). \tag{3.7}$$

By Proposition 1, the second term of (3.7) is

$$n^{-1}\sum_{i=1}^{n}(1-\delta_i)\gamma(Z_i)=n^{-1}\sum_{i=1}^{n}\frac{(1-\delta_i)\bar{F}_n(Z_i)\gamma(Z_i)}{\bar{F}_n(Z_i)}=\int\bar{F}_n\gamma dG_n.$$

The third term of (3.7) is

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\int_0^{Z_i-}\frac{\gamma(y)}{\bar{G}_n(y)}dG_n(y) &= \int_0^{\infty}\left\{n^{-1}\sum_{i=1}^{n}I(y<Z_i)\right\}\frac{\gamma(y)}{\bar{G}_n(y)}dG_n(y) \\ &= \int_0^{\infty}\bar{H}_n(y)\frac{\gamma(y)}{\bar{G}_n(y)}dG_n(y)=\int\bar{F}_n\gamma dG_n. \end{aligned}$$

The last equality follows from $\bar{H}_n=\bar{F}_n\bar{G}_n$. □

We are interested in an optimal choice of γ in $U_n(\gamma;G)$ in the sense that its variance is minimized. Variance of $U_n(\gamma;G)$ and the optimal choice are given by the next two propositions. In order to make the result brief, we assume that F and G are continuous.

PROPOSITION 5. *Assume that F and G are continuous. Then, under* (3.4),

$$\int_0^{\tau}\frac{\{\varphi(x)\}^2}{\bar{G}(x)}dF(x)<\infty \tag{3.8}$$

and

$$\int_0^{\tau}\bar{F}(x)\{\gamma(x)\}^2dG(x)<\infty, \tag{3.9}$$

the variance of $U_n(\gamma;G)$ is given by

$$\begin{aligned} n\times\mathrm{Var}[U_n(\gamma;G)] &= \int_0^{\tau}\frac{\{\varphi(x)\}^2}{\bar{G}(x)}dF(x)-\left\{\int_0^{\tau}\varphi(x)dF(x)\right\}^2 \\ &+\int_0^{\tau}\bar{F}(x)\{\gamma(x)\}^2dG(x)-2\int_0^{\tau}\frac{\gamma(x)}{\bar{G}(x)}\left\{\int_x^{\tau}\varphi(t)dF(t)\right\}dG(x). \end{aligned} \tag{3.10}$$

PROOF. Let

$$V_i(\gamma;G) = \delta_i \frac{\varphi(Z_i)}{\bar{G}(Z_i)} + (1-\delta_i)\gamma(Z_i) - \int_0^{Z_i} \frac{\gamma(y)}{\bar{G}(y)} dG(y), \quad i = 1, \dots, n.$$

Then, from (3.6), $U_n(\gamma;G) = n^{-1}\sum_{i=1}^n V_i(\gamma;G)$, and

$$\begin{aligned} n \times \mathrm{Var}[U_n(\gamma;G)] &= \mathrm{Var}[V_i(\gamma;G)] = E[\{V_i(\gamma;G)\}^2] - \{E[V_i(\gamma;G)]\}^2 \\ &= E[\{V_i(\gamma;G)\}^2] - \left\{\int_0^\tau \varphi(x)dF(x)\right\}^2. \end{aligned} \tag{3.11}$$

The second moment of $V_i(\gamma;G)$ is

$$\begin{aligned} &E[\{V_i(\gamma;G)\}^2] \\ &= E\left[\delta_i \left\{\frac{\varphi(Z_i)}{\bar{G}(Z_i)}\right\}^2\right] + E\left[(1-\delta_i)\{\gamma(Z_i)\}^2\right] + E\left[\left\{\int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right\}^2\right] \\ &\quad - 2E\left[\frac{\delta_i\varphi(Z_i)}{\bar{G}(Z_i)} \int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right] - 2E\left[(1-\delta_i)\gamma(Z_i)\int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right]. \end{aligned} \tag{3.12}$$

The third term of (3.12) is calculated as

$$\begin{aligned} E\left[\left\{\int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right\}^2\right] &= E\left[\left\{\int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right\}\left\{\int_0^{Z_i} \frac{\gamma(y)}{\bar{G}(y)}dG(y)\right\}\right] \\ &= \int_0^\tau \left\{\int_0^z \int_0^z \frac{\gamma(x)\gamma(y)}{\bar{G}(x)\bar{G}(y)}dG(x)dG(y)\right\} dH(z) \\ &= \int_0^\tau \int_0^\tau \bar{H}(x \vee y)\frac{\gamma(x)\gamma(y)}{\bar{G}(x)\bar{G}(y)}dG(x)dG(y) \\ &= 2\int_0^\tau \int_0^\tau I(x<y)\bar{H}(y)\frac{\gamma(x)\gamma(y)}{\bar{G}(x)\bar{G}(y)}dG(x)dG(y) \\ &= 2\int_0^\tau \bar{F}(y)\gamma(y)\left\{\int_0^y \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right\}dG(y) \\ &= 2E\left[(1-\delta_i)\gamma(Z_i)\int_0^{Z_i} \frac{\gamma(x)}{\bar{G}(x)}dG(x)\right]. \end{aligned}$$

Thus, the third term and the fifth term of (3.12) are canceled. The first term is $\int_0^\tau \frac{\{\varphi(x)\}^2}{\bar{G}(x)}dF(x)$, the second term is $\int_0^\tau \bar{F}(x)\{\gamma(x)\}^2 dG(x)$, and the fourth term is $-2\int_0^\tau \frac{\gamma(x)}{\bar{G}(x)}\left\{\int_x^\tau \varphi(t)dF(t)\right\}dG(x)$. The result follows from (3.11). □

PROPOSITION 6. *Assume F and G are continuous. Define a function γ_{opt} on $[0,\tau]$ by*

$$\gamma_{\mathrm{opt}}(x) = \{\bar{H}(x)\}^{-1}\int_x^\tau \varphi(t)dF(t). \tag{3.13}$$

Then, for any γ *with* (3.4), (3.8) *and* (3.9), *it holds that*

$$\mathrm{Var}[U_n(\gamma; G)] \geq \mathrm{Var}[U_n(\gamma_{\mathrm{opt}}; G)].$$

Proof. From (3.10),

$$\begin{aligned} n\,\mathrm{Var}[U_n(\gamma; G)] &= \int_0^\tau \frac{\{\varphi(x)\}^2}{\bar{G}(x)} dF(x) - \left\{\int_0^\tau \varphi(x) dF(x)\right\}^2 \\ &\quad + \int_0^\tau \bar{F}(x)\{\gamma(x)\}^2 dG(x) - 2\int_0^\tau \bar{F}(x)\gamma(x)\gamma_{\mathrm{opt}}(x) dG(x) \\ &= \int_0^\tau \frac{\{\varphi(x)\}^2}{\bar{G}(x)} dF(x) - \left\{\int_0^\tau \varphi(x) dF(x)\right\}^2 \\ &\quad - \int_0^\tau \bar{F}(x)\{\gamma_{\mathrm{opt}}(x)\}^2 dG(x) + \int_0^\tau \bar{F}(x)\{\gamma(x) - \gamma_{\mathrm{opt}}(x)\}^2 dG(x) \\ &= n\mathrm{Var}[U_n(\gamma_{\mathrm{opt}}; G)] + \int_0^\tau \bar{F}(x)\{\gamma(x) - \gamma_{\mathrm{opt}}(x)\}^2 dG(x). \end{aligned}$$

The last term is nonnegative. □

The function γ_{opt} defined by (3.13) is an optimal choice of γ in the class of unbiased estimators $U_n(\gamma; G)$. From Theorem 1.1 of Stute (1995), we can see that

$$U_n(\gamma_{\mathrm{opt}}; G) = \int \varphi dF_n + o_p(n^{-1/2}).$$

Thus, the optimal unbiased estimator $U_n(\gamma_{\mathrm{opt}}; G)$ is asymptotically equivalent to the KM integral. This is natural because the KM integral is a maximum likelihood estimator and it is asymptotically efficient as has been shown by Schick *et al.* (1988).

At the last of the previous section, it has been stated that the simple unbiased estimator $S_{0n}(G) = U_n(0; G)$ has asymptotically larger variance than the KM integral. This is reconfirmed by Proposition 6 since $\mathrm{Var}[U_n(0; G)] > \mathrm{Var}[U_n(\gamma_{\mathrm{opt}}; G)]$ and the KM integral is asymptotically equivalent to the optimal unbiased estimator $U_n(\gamma_{\mathrm{opt}}; G)$.

From Proposition 4, it can be seen that $U_n(\gamma_{\mathrm{opt}}; G_n) = \int \varphi dF_n$. Thus, it should be considered that the KM integral is a plug-in of G_n not into $S_{0n}(G) = U_n(0; G)$ but into $U_n(\gamma_{\mathrm{opt}}; G)$. If we regard the KM integral as a plug-in of G_n into $S_{0n}(G) = U_n(0; G)$, then the plug-in results in a decrease in variance. This is unnatural. It is intuitively natural that the KM integral is a plug-in of G_n into $U_n(\gamma_{\mathrm{opt}}; G)$ and the plug-in asymptotically does not have any influence on variance, since G_n is consistent to G. It, however, causes a bias in the case of finite samples.

When G is known, it seems reasonable to suppose that $U_n(\gamma_{\mathrm{opt}}; G)$ is better than the KM integral because $U_n(\gamma_{\mathrm{opt}}; G)$ is unbiased and both estimators are asymptotically equivalent. However, the optimal function γ_{opt} defined by (3.13) depends on the unknown F, and hence $\gamma_{\rm opt}$ is unknown. This is an important problem practically.

One approach to solving this problem may be to estimate $\gamma_{\rm opt}$. However, this is putting the cart before the horse. Since

$$\gamma_{\rm opt}(0) = \{\bar{H}(0)\}^{-1}\int_0^\tau \varphi(t)dF(t) = \int_0^\tau \varphi(t)dF(t) = \theta,$$

if $\gamma_{\rm opt}(x)$ can be appropriately estimated by a function $\hat{\gamma}_{\rm opt}(x)$, we can immediately obtain an estimate of θ by $\hat{\gamma}_{\rm opt}(0)$. Moreover, if the estimated optimal function $\hat{\gamma}_{\rm opt}$ depends on the data, then $U_n(\hat{\gamma}_{\rm opt}; G)$ is not necessarily unbiased.

Thus, it is important to choose a suitable γ so that the variance of $U_n(\gamma; G)$ does not increase greatly compared with $\gamma_{\rm opt}$, and the choice must be independent of the data.

If there is no censorship at all, i.e., $\delta_i = 1$ and $Z_i = X_i$ for all i, then the KM integral reduces to $n^{-1}\sum_{i=1}^n \varphi(X_i)$, which is an expectation of $\varphi(X)$ with respect to the empirical distribution of X_i's. On the other hand, in this case, $U_n(\gamma; G)$ reduces to

$$n^{-1}\sum_{i=1}^n \left\{\frac{\varphi(X_i)}{\bar{G}(X_i-)} - \int_0^{X_i-} \frac{\gamma(y)}{\bar{G}(y)}dG(y)\right\}.$$

It is natural to expect that this is also in agreement with $n^{-1}\sum_{i=1}^n \varphi(X_i)$ under no censorship.

PROPOSITION 7. *Assume that G is continuous and its density g is positive on $(0,\tau)$ and that φ is differentiable on $(0,\tau)$. Define, for $z \in (0,\tau)$,*

$$\tilde{\gamma}(z) = \frac{\varphi(z)}{\bar{G}(z)} + \frac{G(z)}{g(z)}\frac{d\varphi(z)}{dz}. \tag{3.14}$$

Then, for any $0 < x < \tau$,

$$\frac{\varphi(x)}{\bar{G}(x)} - \int_0^x \frac{\tilde{\gamma}(y)}{\bar{G}(y)}dG(y) = \varphi(x). \tag{3.15}$$

Hence, $U_n(\tilde{\gamma}; G)$ reduces to $n^{-1}\sum_{i=1}^n \varphi(X_i)$ when there is no censorship at all.

PROOF. Substituting (3.14) into the integral in (3.15), we have

$$\begin{aligned}\int_0^x \frac{\tilde{\gamma}(y)}{\bar{G}(y)}dG(y) &= \int_0^x \frac{\varphi(y)}{\{\bar{G}(y)\}^2}dG(y) + \int_0^x \frac{G(y)}{\bar{G}(y)}\frac{d\varphi(y)}{dy}dy \\ &= \int_0^x \frac{\varphi(y)}{\{\bar{G}(y)\}^2}dG(y) + \frac{G(x)}{\bar{G}(x)}\varphi(x) - \int_0^x \frac{\varphi(y)}{\{\bar{G}(y)\}^2}dG(y) \\ &= -\varphi(x) + \frac{\varphi(x)}{\bar{G}(x)}.\end{aligned}$$

□

We shall now carefully examine the difference between γ_{opt} and $\tilde{\gamma}$. Assume that F is continuous and that its hazard function λ_F is positive on $(0,\tau)$. Let $\gamma^*_{\mathrm{opt}} = \bar{G}\gamma_{\mathrm{opt}}$. Then, for $z \in (0,\tau)$,

$$\bar{F}(z)\gamma^*_{\mathrm{opt}}(z) = \int_z^\tau \bar{F}(x)\varphi(x)\lambda_F(x)dx.$$

Differentiating this, we obtain the expression

$$\gamma_{\mathrm{opt}}(z) = \frac{\gamma^*_{\mathrm{opt}}(z)}{\bar{G}(z)} = \frac{1}{\bar{G}(z)}\left\{\varphi(z) + \frac{1}{\lambda_F(z)}\frac{d\gamma^*_{\mathrm{opt}}(z)}{dz}\right\}. \tag{3.16}$$

On the other hand, $\tilde{\gamma}$ defined by (3.14) can be written as

$$\tilde{\gamma}(z) = \frac{1}{\bar{G}(z)}\left\{\varphi(z) + \frac{G(z)}{\lambda_G(z)}\frac{d\varphi(z)}{dz}\right\},$$

where $\lambda_G = g/\bar{G}$. From (3.16), we can see that $\gamma_{\mathrm{opt}}(z)$ is determined by two unknown functions, $d\gamma^*_{\mathrm{opt}}/dz$ and λ_F, except for the known functions φ and $\bar{G}$. In $\tilde{\gamma}$, these unknown functions are replaced by known functions $d\varphi/dz$ and λ_G/G, respectively.

First we shall concentrate on the replacement of $d\gamma^*_{\mathrm{opt}}/dz$ by $d\varphi/dz$. We can write as

$$\gamma^*_{\mathrm{opt}}(z) = E\left[\varphi(X) \mid X > z\right] - \bar{F}(\tau)/\bar{F}(z).$$

If $\tau = \tau_F$ (i.e., $\tau_F \le \tau_G$) and the lifetime X has a property of lack of memory as

$$E\left[\varphi(X) \mid X > z\right] = \varphi(z) + \text{constant},$$

then $d\gamma^*_{\mathrm{opt}}/dz = d\varphi/dz$. Although the lack of memory does not hold in general, it seems reasonable to replace $d\gamma^*_{\mathrm{opt}}/dz$ by $d\varphi/dz$ since the lifetime distribution is completely unknown.

Next we consider replacement of λ_F. Let

$$q(z) = \Pr\{\delta = 1 \mid Z = z\} = \frac{\lambda_G(z)}{\lambda_F(z) + \lambda_G(z)},$$

which has been called a censoring pattern function by Suzukawa and Taneichi (2000). Then, λ_F can be expressed as

$$\lambda_F(z) = \frac{1 - q(z)}{q(z)}\lambda_G(z).$$

In this expression, λ_G is known, but the censoring pattern function q is unknown. Thus, we have to consider with what kind of function q should be replaced.

In $\tilde{\gamma}$, q is replaced by $G/(1+G)$, which is increasing and not greater than $1/2$. Thus, this replacement may have validity when the censoring pattern is increasing with observable time and censoring is not so heavy (censoring proportion being less than 1/2). However, there is no necessity of replacing q by $G/(1+G)$. It must be noted that this replacement excludes important cases in which the censoring pattern is constant (i.e., $q(z)$ is independent of z).

We consider replacement of q by a constant. The censoring pattern function q is constant if and only if hazards $\lambda_F(z)$ and $\lambda_G(z)$ are proportional as

$$\lambda_G(z)/\lambda_F(z) = \alpha,$$

where $\alpha > 0$ is a constant. Substituting this and $d\gamma^*_{\mathrm{opt}}(z)/dz = d\varphi(z)/dz$ into (3.16), we obtain the function

$$\bar{\gamma}_\alpha(z) = \frac{1}{\bar{G}(z)}\left\{\varphi(z) + \frac{\alpha}{\lambda_G(z)}\frac{d\varphi(z)}{dz}\right\}. \tag{3.17}$$

In the following section, unbiased estimators $U_n(0;G)$, $U_n(\tilde{\gamma};G)$ and $U_n(\bar{\gamma}_\alpha;G)$ in the case of estimation of mean lifetimes, i.e., $\varphi(x) = x$, are compared.

4. Estimation of mean lifetimes

In this section, the case where censoring distribution is known and it is an exponential distribution with hazard $\lambda > 0$, i.e., $\bar{G}(y) = \exp(-\lambda y)$, is considered. The purpose here is to estimate the mean lifetime $\mu \equiv \int x dF(x)$.

The simple unbiased estimator of μ is given by

$$\hat{\mu}_{0n} \equiv U_n(0;G) = n^{-1}\sum_{i=1}^{n}\delta_i Z_i/\bar{G}(Z_i) = n^{-1}\sum_{i=1}^{n}\delta_i Z_i\, e^{\lambda Z_i}.$$

Substituting $\varphi(z) = z$ and $\bar{G}(y) = \exp(-\lambda y)$ into (3.14) and (3.17), we have

$$\tilde{\gamma}(z) = \{z + \lambda^{-1}(1 - e^{-\lambda z})\}e^{\lambda z} \quad \text{and} \quad \bar{\gamma}_\alpha(z) = (z + \alpha\lambda^{-1})e^{\lambda z}.$$

Using these functions as γ, we obtain unbiased estimators

$$\tilde{\mu}_n \equiv U_n(\tilde{\gamma};G) = n^{-1}\sum_{i=1}^{n}\left\{Z_i + (1-\delta_i)\lambda^{-1}(e^{\lambda Z_i} - 1)\right\} \quad \text{and}$$

$$\bar{\mu}_n(\alpha) \equiv U_n(\bar{\gamma}_\alpha;G) = n^{-1}\sum_{i=1}^{n}\left\{\lambda^{-1}(1-\alpha\delta_i)e^{\lambda Z_i} + \lambda^{-1}(\alpha - 1)\right\}.$$

The main concern is to compare estimators $\hat{\mu}_{0n}$, $\tilde{\mu}_n$ and $\bar{\mu}_n(\alpha)$. It is noted that $\hat{\mu}_{0n}$ and $\tilde{\mu}_n$ are unique, but $\bar{\mu}_n(\alpha)$ depends on the choice of the constant $\alpha > 0$.

In order to evaluate variances of these unbiased estimators, it is assumed that the lifetime distribution is exponential with hazard one; $\bar{F}(x) = e^{-x}$. Under this assumption, a true value of μ is one, and expectations of above three estimators are one.

From (3.13), the optimal γ is given by $\gamma_{\text{opt}}(z) = (z+1)e^{\lambda z}$. The optimal unbiased estimator is given by

$$\hat{\mu}_{n,\text{opt}} \equiv U_n(\gamma_{\text{opt}}; G) = n^{-1} \sum_{i=1}^{n} \left\{ \lambda^{-1}(1-\lambda\delta_i)e^{\lambda Z_i} + \lambda^{-1}(\lambda - 1) \right\}.$$

We can see that $\bar{\mu}_n(\lambda) = \hat{\mu}_{n,\text{opt}}$. The optimal choice of the constant α in $\bar{\mu}_n(\alpha)$ is λ.

The condition (3.8) for existence of variance is $\int_0^\infty x^2 e^{(\lambda-1)x} dx < \infty$, which is equivalent to $\lambda < 1$. For $\tilde{\gamma}, \bar{\gamma}_\alpha$ and γ_{opt} defined above, condition (3.9) is satisfied if $\lambda < 1$. Thus, if $\lambda < 1$, variances of all estimators can be obtained by (3.10). Let $q = \Pr\{\delta_i = 0\}$ (proportion of censoring), then $q = \lambda/(1+\lambda)$. Thus, the condition $\lambda < 1$ is equivalent to $q < 1/2$ (censoring proportion less than 1/2). We say that censoring is heavy when q is near 1/2 (i.e., λ is near one) and is light when q is near zero (i.e., λ is near zero).

Under $\lambda < 1$, variances of the unbiased estimators are given by

(4.1)

$$n \times \text{Var}[\hat{\mu}_{0n}] = 2(1-\lambda)^{-3} - 1, \quad n \times \text{Var}[\tilde{\mu}_n] = (1-\lambda)^{-1} + \lambda(1+\lambda)^{-1},$$
$$n \times \text{Var}[\bar{\mu}_n(\alpha)] = (1-\lambda)^{-1}\left\{1 + (\alpha-\lambda)^2/\lambda\right\}, \quad n \times \text{Var}[\hat{\mu}_{n,\text{opt}}] = (1-\lambda)^{-1}.$$

It holds that for any $0 < \lambda < 1$

$$\text{Var}[\hat{\mu}_{n,\text{opt}}] < \text{Var}[\tilde{\mu}_n] < \text{Var}[\hat{\mu}_{0n}].$$

Thus, the estimator $\tilde{\mu}_n$ improves the simple estimator $\hat{\mu}_{0n}$. Figure 2 shows variances of these estimators. The degree of improvement is so remarkable that censoring becomes heavy. This fact can also be seen from

light-censoring comparison:

$$\lim_{q \searrow 0} \frac{\text{Var}[\tilde{\mu}_n]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = 1, \quad \lim_{q \searrow 0} \frac{\text{Var}[\hat{\mu}_{0n}]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = 1, \tag{4.2}$$

heavy-censoring comparison:

$$\lim_{q \nearrow 1/2} \frac{\text{Var}[\tilde{\mu}_n]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = 1, \quad \lim_{q \nearrow 1/2} \frac{\text{Var}[\hat{\mu}_{0n}]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = \infty. \tag{4.3}$$

On the other hand, $\bar{\mu}_n(\alpha)$ does not always improve $\hat{\mu}_{0n}$. We are interested in the choice of α such that $\bar{\mu}_n(\alpha)$ improves $\hat{\mu}_{0n}$. At least, the optimal choice $\alpha = \lambda$ improves it. It can be easily seen that

$$\text{Var}[\bar{\mu}_n(\alpha)] < \text{Var}[\hat{\mu}_{0n}] \Leftrightarrow 0 < \alpha < \lambda + \sqrt{\frac{2\lambda}{(1-\lambda)^2} + \lambda^2 - 2\lambda}.$$

Figure 3 shows the region of (λ, α) in which $\bar{\mu}_n(\alpha)$ improves $\hat{\mu}_{0n}$. This region is under the boundary curve. The optimal choice $\alpha = \lambda$ is in this region. For

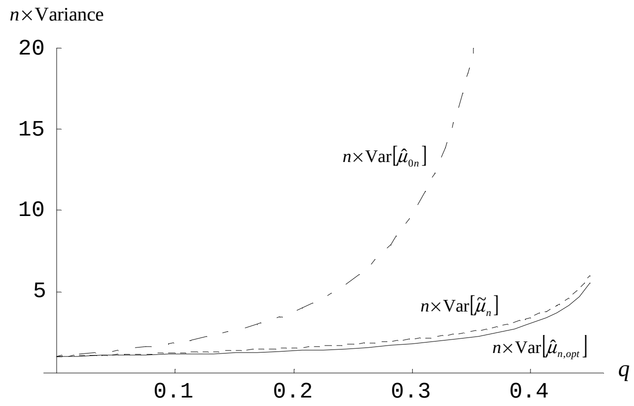

Figure 2. Variances of unbiased estimators $\hat{\mu}_{0n}$, $\tilde{\mu}_n$ and $\hat{\mu}_{n,\text{opt}}$.

example, if $\lambda = 1/2$ (i.e., censoring proportion $q = 1/3$), then $\text{Var}[\bar{\mu}_n(\alpha)] < \text{Var}[\hat{\mu}_{0n}]$ for $0 < \alpha < 2.30$. From this figure, it can be seen that if censoring is heavy (λ is near one), $\bar{\mu}_n(\alpha)$ improves $\hat{\mu}_{0n}$ for any choice of α. On the other hand, when censoring is light, there is no improvement for almost all α. An interesting choice of α is $\alpha = 0$, which always improves $\hat{\mu}_{0n}$. The estimator is

$$\bar{\mu}_n(0) = n^{-1} \sum_{i=1}^{n} \left\{ \lambda^{-1}(e^{\lambda Z_i} - 1) \right\},$$

and its variance is given by $n \times \text{Var}[\bar{\mu}_n(0)] = (1+\lambda)/(1-\lambda)$.

We are also interested in whether $\bar{\mu}_n(\alpha)$ improves $\tilde{\mu}_n$. At least, the optimal choice $\alpha = \lambda$ improves $\tilde{\mu}_n$ since $\bar{\mu}_n(\lambda) = \hat{\mu}_{n,\text{opt}}$. We can see that

$$\text{Var}[\bar{\mu}_n(\alpha)] < \text{Var}[\tilde{\mu}_n] \quad \Leftrightarrow \quad \lambda\left(1 - \sqrt{\frac{1-\lambda}{1+\lambda}}\right) < \alpha < \lambda\left(1 + \sqrt{\frac{1-\lambda}{1+\lambda}}\right).$$

Figure 4 shows the region of (λ, α) in which $\bar{\mu}_n(\alpha)$ improves $\tilde{\mu}_n$. This region is the inner part of the boundary curve, and the optimal choice $\alpha = \lambda$ is in this region. In the neighborhood of the optimal line $\alpha = \lambda$, $\bar{\mu}_n(\alpha)$ has smaller variance than $\tilde{\mu}_n$. For example, if $\lambda = 1/2$, then the optimal α is $1/2$ and $\text{Var}[\bar{\mu}_n(\alpha)] < \text{Var}[\tilde{\mu}_n]$ for $0.22 < \alpha < 0.78$. This figure also shows that $\text{Var}[\bar{\mu}_n(0)] > \text{Var}[\tilde{\mu}_n]$ unless $\lambda = 0$ (no censoring). Namely, $\bar{\mu}_n(0)$ is inferior to $\tilde{\mu}_n$, though it improves $\hat{\mu}_{0n}$.

Figure 5 shows variances of unbiased estimators $\bar{\mu}_n(1/2)$, $\tilde{\mu}_n$ and $\hat{\mu}_{n,\text{opt}}$. The estimator $\bar{\mu}_n(1/2)$ attains an optimal value at $q = 1/3$, and it improves $\tilde{\mu}_n$ in the neighborhood of $q = 1/3$. However, it has undesirable properties

$$\lim_{q \searrow 0} \frac{\text{Var}[\bar{\mu}_n(1/2)]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = \infty \quad \text{and} \quad \lim_{q \nearrow 1/2} \frac{\text{Var}[\bar{\mu}_n(1/2)]}{\text{Var}[\hat{\mu}_{n,\text{opt}}]} = 1 + (1/2)^2 > 1.$$

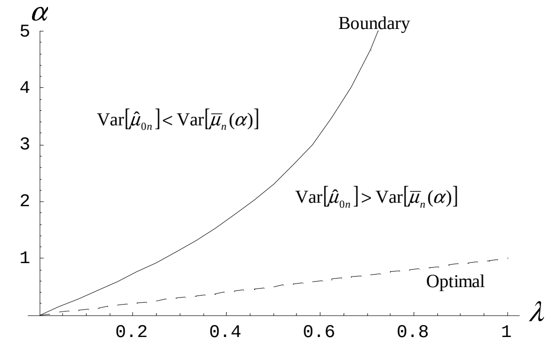

Figure 3. Region of (λ, α) in which $\bar{\mu}_n(\alpha)$ improves $\hat{\mu}_{0n}$.

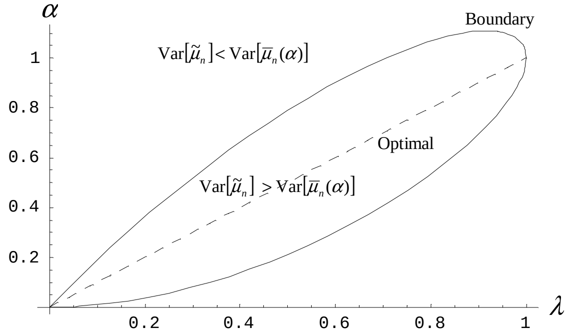

Figure 4. Region of (λ, α) in which $\bar{\mu}_n(\alpha)$ improves $\tilde{\mu}_n$.

The first property is a serious problem. Therefore, to use $\bar{\mu}_n(\alpha)$, it is necessary to choose α carefully. A natural estimator of q is $\hat{q} = 1 - \sum_{i=1}^{n} \delta_i/n$. It seems reasonable to use an α that is near $\hat{q}/(1-\hat{q})$. However, it must be noted that $\bar{\mu}_n(\hat{q}/(1-\hat{q}))$ is not unbiased any longer.

Generally, $\tilde{\mu}_n$ does not cause a great increase in variance compared with $\hat{\mu}_{n,\mathrm{opt}}$ for all $q < 1/2$. Moreover, from (4.2) and (4.3), we can say that it is equivalent to $\hat{\mu}_{n,\mathrm{opt}}$ under both light-censoring and heavy-censoring conditions.

We shall investigate whether these unbiased estimators are more desirable than the KM mean $\hat{\mu}_n^{\mathrm{KM}} \equiv \int x dF_n(x)$. We shall now examine the mean squared errors (MSEs) of these estimators. For unbiased estimators, their MSEs are given by their variances (4.1). The exact MSE of the KM mean has not been obtained yet. It was investigated here by simulations.

Figure 6 shows MSEs of $\tilde{\mu}_n$, $\hat{\mu}_{n,\mathrm{opt}}$ and $\hat{\mu}_n^{\mathrm{KM}}$. When $n = 10$ and censoring is heavy, $\hat{\mu}_n^{\mathrm{KM}}$ has a smaller MSE than $\tilde{\mu}_n$ and $\hat{\mu}_{n,\mathrm{opt}}$. In this case, $\hat{\mu}_n^{\mathrm{KM}}$ has a smaller variance, though it has a bias. Generally, $\hat{\mu}_n^{\mathrm{KM}}$ has a smaller variance in the case of small sample size. However, it has a negative bias. As the sample size becomes

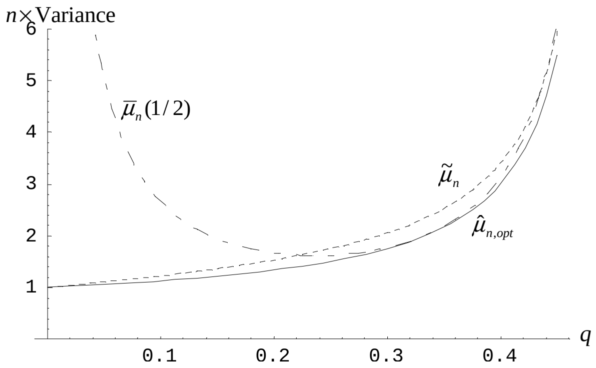

Figure 5. Variances of unbiased estimators $\bar{\mu}_n(1/2)$, $\tilde{\mu}_n$ and $\hat{\mu}_{n,\mathrm{opt}}$.

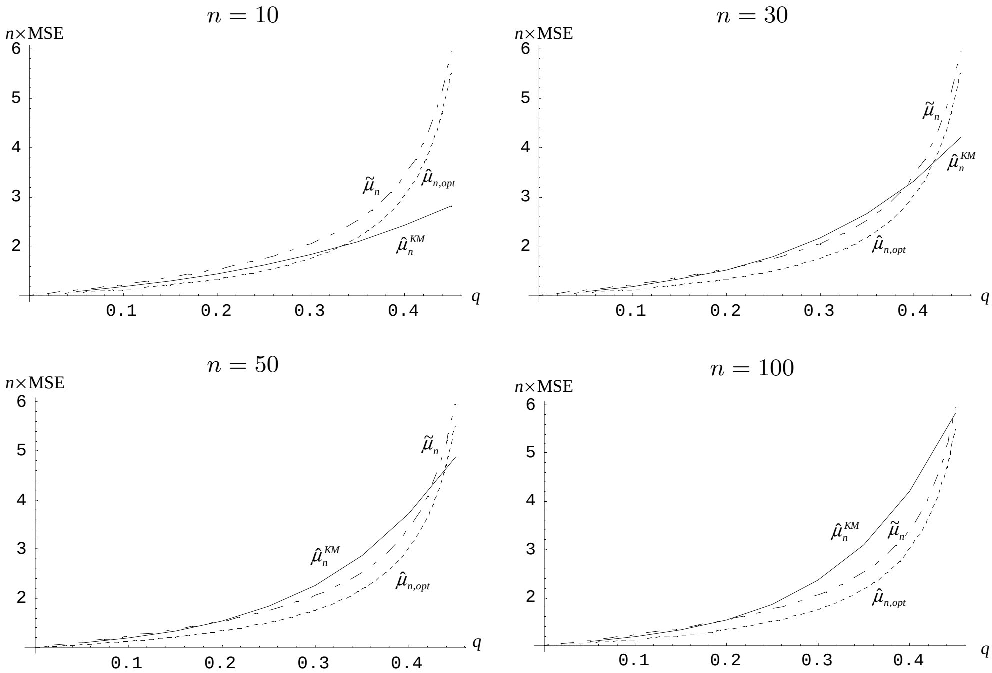

Figure 6. Mean squared errors (MSEs) of mean lifetime estimators $\tilde{\mu}_n$, $\hat{\mu}_{n,\mathrm{opt}}$ and $\hat{\mu}_n^{\mathrm{KM}}$.

large, differences in variances become small, but the bias of $\hat{\mu}_n^{\mathrm{KM}}$ still remains. The reason for the larger MSE of $\hat{\mu}_n^{\mathrm{KM}}$ in the case of $n = 100$ is that its bias still remains. Although $\hat{\mu}_n^{\mathrm{KM}}$ is asymptotically unbiased, its convergence is slow. It also seems natural that $\hat{\mu}_n^{\mathrm{KM}}$ is not greatly improved, since it is a nonparametric maximum likelihood estimator. In many practical situations, sample size is not so large and the censoring proportion is between 0.1 and 0.4. In such situations, not only $\hat{\mu}_n^{\mathrm{KM}}$ but also $\tilde{\mu}_n$ may be taken into consideration.

Acknowledgements

I gratefully acknowledge helpful discussions with Professor Yoshihide Kakizawa on several points in the paper. I thank also the referee for helpful comments.

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.*, **2**, 437–453.

Efron, B. (1967). The two sample problem with censored data, *Proc. Fifth Berkely Symp. Math. Statist. Prob.*, **4**, 831–853.

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

Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Math. Centre Tracts, **124**, Mathematisch Centrum, Amsterdam.

Gill, R. D. (1983). Large sample behavior of the product limit estimator on the whole line, *Ann. Statist.*, **11**, 49–58.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator, *Scand. J. Statist.*, **5**, 195–199.

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

Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*, Wiley, New York.

Mauro, D. (1985). A combinatoric approach to the Kaplan-Meier estimator, *Ann. Statist.*, **13**, 142–149.

Reid, N. (1981). Influence functions for censored data, *Ann. Statist.*, **9**, 78–92.

Schick, A., Susarla, V. and Koul, H. (1988). Efficient estimation of functionals with censored data, *Statist. and Decisions*, **6**, 349–360.

Stute, W. (1994). The bias of Kaplan-Meier integrals, *Scand. J. Statist.*, **21**, 475–484.

Stute, W. (1995). The central limit theorem under random censorship, *Ann. Statist.*, **23**, 422–439.

Stute, W. and Wang, J. L. (1993). The strong law under random censorship, *Ann. Statist.*, **21**, 1591–1607.

Stute, W. and Wang, J. L. (1994). The jackknife estimate of a Kaplan-Meier integral, *Biometrika*, **81**, 602–606.

Suzukawa, A. and Taneichi, N. (2000). Estimation of survival function based on modeling of censoring pattern, *J. Jpn. Statist. Soc.*, **30**, 177–196.

Wang, J. (1987). A note on the uniform consistency of the Kaplan-Meier estimator, *Ann. Statist.*, **15**, 1313–1316.

Yang, S. (1994). A central limit theorem for functionals of the Kaplan-Meier estimator, *Statist. Probab. Letters*, **21**, 337–345.

Zhou, M. (1988). Two-sided bias bound of the Kaplan-Meier estimator, *Probab. Theory Rel. Fields*, **79**, 165–173.