

Estimation of Lifetime Distribution with Missing Censoring

Jiantian Wang
Kean University

Abstract: This paper considers the estimation of lifetime distribution based on missing-censoring data. Using the simple empirical approach rather than the maximum likelihood argument, we obtain the parametric estimations of lifetime distribution under the assumption that the failure time follows exponential or gamma distribution. We also derive the nonparametric estimation for both continuous and discrete failure distributions under the assumption that the censoring distribution is known. The loss of efficiency due to missing-censoring is shown to be generally small if the data model is specified correctly. Identifiability issue of the lifetime distribution with missing-censoring data is also addressed.

Key words: Estimable variable, missing-censoring, maximum likelihood estimation, lifetime distribution.

1. Introduction

Let T be a random variable representing the failure time of the subject under study. In many applications, such as in biological sciences and in clinical trials for cancer research, the failure time T may not be always observable due to the presence of censoring time C , which is assumed to be independent of T . Denote $F(t) = P(T < t)$ as the lifetime distribution. The estimation of $F(t)$ based on censored data is a crucial issue in survival analysis.

This estimation problem has been well-solved by the classical Kaplan-Meier estimator (Kaplan and Meier, 1958) under the usual random right censoring model. In such a data model, one observes the failure time or the censoring time, whichever comes first. However, there are some circumstances in which the censoring time can not be observed, thus resulting in the so-called missing-censoring data. A typical prototype for missing-censoring data is product warranty data. If under warranty a product experience a certain type of event, or “failure”, then from warranty report, we can have the information about the “failure”. However, for most unfailed items, the censoring time is unknown since the manufactures

likely do not know the date of the sale of the items. For them, the only information is that they have not been reported to be failed. Hu, Lawless and Suzuki (1998) give some examples there.

Several researchers have studied missing-censoring data. For instance, under the assumption that both T and C are exponential distributed, Phillips and Sweeting (1996) considered the estimation for the parameter of the exponential distribution when censoring times are subject to error. Phillips and Sweeting (2001) studied the estimation problems when censoring is missing but a concomitant variable is available. Singh (2002) considered the inferences about the lifetime distribution under the assumption that both failure and censoring times are exponentially distributed. Hu, Lawless and Suzuki (1998) considered non-parametric estimation of $F(t)$ when both T and C are discrete.

The methods used by these researchers are all based on the traditional maximum likelihood approach. However, maximum likelihood approach sometimes could be quite complicated. Besides the maximum likelihood approach may not be always applicable. As pointed out by Kalbfleisch and Prentice (1981, p12), "the maximum likelihood approach should be treated with care, since there are dangers associated with maximizing likelihood of many parameters because this technique may lead to inefficient or inconsistent estimates".

We believe that a target parameter or a target function can be consistently estimated from the observed data if and only if it is a functional of some estimable variable(s). We say a random variable X is estimable from the sample S if S is a realization of the random variable, or a function of the random variable. For instance, suppose the sample $S = \{x_1, x_2, \dots, x_n\}$ is a realization of the random variable X , then X , X^2 , and in general, $f(X)$ for given f , are all estimable from the sample S . Now consider censoring data setting. Let T and C be the failure time and censoring time respectively. With the presence of censoring, the observed data are, $\delta_i = I(T_i \leq C_i)$, and $X_i = T_i \wedge C_i$, for $i = 1, \dots, n$. From these censored data, we can see that $\delta = I(T \leq C)$, and $X = T \wedge C$, or equivalently, δ , $T\delta$, and $C(1 - \delta)$, are estimable. The target variable T is not estimable. Consequently, $E(T)$, the mean life time, can not be estimated consistently from the data.

Let $N(t) = P(T\delta > t)$ and $Y(t) = P(T\delta + C(1 - \delta) > t)$. These two functions are estimable functions from data since $T\delta$ and $T\delta + C(1 - \delta)$ are estimable. Denote $S(t) = 1 - F(t)$. Then, under the assumption that T and C are independent, we have,

$$\frac{dN(t)}{Y(t)} = \frac{dS(t)}{S(t)}.$$

Hence, $S(t)$ is estimable under the assumption of independence of T and C . If the independence assumption does not hold, then $S(t)$ will not be estimable. These

conclusions are well known results in survival analysis.

In this paper, we provide a systematic treatment for missing-censoring data by using a simple empirical approach. First, we identify the estimable variables based on observed data, and then express the target parameters or functions in terms of those estimable variables, and consequently the target parameters or functions can be estimated consistently from those expressions in a natural way. Based on this simple method, we derive the parametric estimators and their biases and variances when the lifetime distribution is exponential and gamma. Also we consider the nonparametric estimation of the lifetime distribution when failure time and censoring time are either discrete or continuous. Theoretical analysis and numerical simulations show that the loss of efficiency due to missing-censoring is generally small if the data model is specified correctly. Compared with the traditional maximum likelihood method, the approach advocated in the present paper is simpler, more straightforward, and much easier to implement.

The rest of the paper is organized as follows. In Section 2, we consider the parametric estimation of the lifetime distribution when it is exponential and gamma. Nonparametric estimation of the lifetime distribution is discussed in Section 3. Section 4 provides some computational simulations assessing the efficiency of the corresponding parametric estimators with missing-censoring data. A short discussion is presented in Section 5 while some theoretical derivations are deferred to Appendix.

2. Parametric Estimation of Lifetime Distribution

In this section, we consider the parametric estimation of the lifetime distribution under the parametric assumption for the lifetime distribution and the censoring distribution. We begin with the case when both T and C are exponential and then the case when one is exponential while the other is gamma. General parametric estimation problem is also addressed.

2.1 Both T and C are exponential

Assume that both the failure time T and the censoring time C follow exponential distributions with mean $1/\lambda$ and $1/\phi$, respectively. Suppose that T_i is the failure time for product unit i in n manufactured units and C_i is the censoring time for that unit. When there is no censoring, the observed data are T_i , $i = 1, \dots, n$, and the estimable variable is T . With the presence of the usual random censoring, the observed data are $\delta_i = I(T_i \leq C_i)$, and T_i or C_i , if $\delta_i = 1$ or $\delta_i = 0$, for $i = 1, \dots, n$. Thus the estimable variables are $\delta = I(T \leq C)$, $T\delta$, and $C(1 - \delta)$. When censoring is missing, the observed data are T_i if $\delta_i = 1$, and nothing excepting the fact that $T_i > C_i$ when failure time is censored. Hence,

the estimable variables for missing-censoring data are δ and $T\delta$.

By some calculations, we have, $E(T) = 1/\lambda$, $E(\delta) = \lambda/(\lambda + \phi)$, $E(T\delta) = \lambda/(\lambda + \phi)^2$, and $E\{C(1 - \delta)\} = \phi/(\lambda + \phi)^2$. Thus, $\lambda = 1/E(T)$, $\lambda = E(\delta)/\{E(T\delta) + E\{C(1 - \delta)\}\} = E(\delta)/E(X)$, where $X = \min(T, C)$, and $\lambda = E^2(\delta)/E(T\delta)$. Consequently, when there is no censoring, the corresponding consistent estimator for the parameter λ is,

$$\hat{\lambda}_0 = \frac{n}{\sum_{i=1}^n T_i}. \quad (2.1)$$

With the presence of censoring, the estimator is,

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n \{T_i \delta_i + C_i(1 - \delta_i)\}} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n X_i}. \quad (2.2)$$

While when censoring is missing, the estimator is,

$$\hat{\lambda}_2 = \frac{(\sum_{i=1}^n \delta_i)^2}{n \sum_{i=1}^n T_i \delta_i}. \quad (2.3)$$

As shown in Appendix A, the asymptotic biases and variances of these estimators are:

$$\begin{aligned} \text{bias}(\hat{\lambda}_0) &= \lambda/n, \text{ and } \text{var}(\hat{\lambda}_0) = \lambda^2/n, \\ \text{bias}(\hat{\lambda}_1) &= \lambda/n, \text{ and } \text{var}(\hat{\lambda}_1) = \lambda(\lambda + \phi)/n, \\ \text{bias}(\hat{\lambda}_2) &= (\lambda + \phi)/n, \text{ and } \text{var}(\hat{\lambda}_2) = \lambda(\lambda + 2\phi)/n. \end{aligned}$$

Remark. These estimators and their asymptotic biases and variances were derived in Philips and Sweeting (2001) by using maximum likelihood argument and Fisher information. The derivation in this paper is much more straightforward.

2.2 T is Gamma and C is exponential

Now consider the situation when T follows a gamma distribution with density function $f(t) = \lambda^\alpha t^{\alpha-1} e^{-\lambda t} / \Gamma(\alpha)$ and censoring time C follows an exponential distribution with mean $1/\phi$.

Some calculations lead to, $E(T) = \alpha/\lambda$, $E(T^2) = \alpha(\alpha + 1)/\lambda^2$, $E(\delta) = \lambda^\alpha/(\lambda + \phi)^\alpha$, $E(T\delta) = \alpha/(\lambda + \phi)E(\delta)$, $E\{C(1 - \delta)\} = \phi^{-1} - \phi^{-1}E(\delta) - E(T\delta)$, and $E(T^2\delta) = \{(\alpha + 1)/(\lambda + \phi)\}E(T\delta)$.

From the expressions

$$\alpha = \frac{E^2(T)}{E(T^2) - E^2(T)}, \quad \text{and} \quad \lambda = \frac{E(T)}{E(T^2) - E^2(T)},$$

come out the parametric estimators for α and λ when there is no censoring as,

$$\hat{\alpha}_0 = \frac{U_1^2}{nU_2 - U_1^2}, \quad \text{and} \quad \hat{\lambda}_0 = \frac{nU_1}{nU_2 - U_1^2}, \tag{2.4}$$

where $U_1 = \sum_{i=1}^n T_i$, and $U_2 = \sum_{i=1}^n T_i^2$.

From the expressions

$$\alpha = \frac{E^2(T\delta)}{E(\delta)E(T^2\delta) - E^2(T\delta)},$$

$$\phi = \frac{1 - E(\delta)}{E(T\delta) + E\{C(1 - \delta)\}} = \frac{1 - E(\delta)}{E(X)},$$

$$\lambda = \frac{E(\delta)E(T\delta)}{E(\delta)E(T^2\delta) - E^2(T\delta)} - \phi,$$

we can see that the reasonable estimators for α and λ under usual random censoring model can be

$$\hat{\alpha}_1 = \frac{W_1^2}{W_0W_2 - W_1^2}, \tag{2.5}$$

and

$$\hat{\lambda}_1 = \frac{W_0W_1}{W_0W_2 - W_1^2} - \frac{\sum_{i=1}^n (1 - \delta_i)}{\sum_{i=1}^n X_i}, \tag{2.6}$$

where $W_0 = \sum_{i=1}^n \delta_i$, $W_1 = \sum_{i=1}^n T_i\delta_i$, and $W_2 = \sum_{i=1}^n T_i^2\delta_i$.

For the case when censoring times are missing, the corresponding estimators for α and λ can be derived from the expressions

$$\alpha = \frac{E^2(T\delta)}{E(\delta)E(T^2\delta) - E^2(T\delta)} \quad \text{and} \quad \lambda = \alpha \frac{E(\delta)^{\frac{\alpha+1}{\alpha}}}{E(T\delta)},$$

as

$$\hat{\alpha}_2 = \frac{W_1^2}{W_0W_2 - W_1^2}, \tag{2.7}$$

and

$$\hat{\lambda}_2 = \frac{\hat{\alpha}}{\sqrt[n]{n}} \frac{W_0^{\frac{\hat{\alpha}+1}{\hat{\alpha}}}}{W_1}. \tag{2.8}$$

The biases and variances of these estimators, even though can be derived without much difficulty, are cumbersome and thus of little theoretical interest. In Section 4, we just use simulations to compare the efficiency of these estimators.

2.3 T is exponential and C is gamma

For the situation where T follows an exponential distribution with density function $f(t) = \lambda \exp(-\lambda t)$ and C follows a gamma distribution with density function $g(t) = \phi(\phi t)^{k-1} e^{-\phi t} / \Gamma(k)$, again by some calculations, we have,

$$\begin{aligned} E(\delta) &= 1 - \left(\frac{\phi}{\lambda + \phi}\right)^k, \\ E(T\delta) &= \frac{1}{\lambda} \left\{ 1 - \left(\frac{\phi}{\lambda + \phi}\right)^k \left(1 + k \frac{\lambda}{\lambda + \phi}\right) \right\}, \\ E(T^2\delta) &= \frac{2}{\lambda^2} \left\{ 1 - \left(\frac{\phi}{\lambda + \phi}\right)^k \frac{(k+2)(k+1)\lambda^2 + 2(k+2)\lambda\phi + 2\phi^2}{(\lambda + \phi)^2} \right\}. \end{aligned}$$

Replace $E(\delta)$, $E(T\delta)$ and $E(T^2\delta)$ by W_0/n , W_1/n and W_2/n respectively, we can get a nonlinear system about unknown parameters λ , k and ϕ . Such nonlinear system usually can only be solved numerically, and the solution gives the estimators to the parameters when censoring is missing. However, due to the lack of closed forms for these estimators, inferences concerning the resulted estimators could be elusive.

2.4 Identifiability of the lifetime distribution under parametric assumption

Assume that both failure time T and censoring time C follow some distributions with some unknown parameters. Let $f(t|\theta_1)$ be the density function of T and $K(t|\theta_2)$ be the survival function of C respectively, where θ_1 and θ_2 are unknown parameter vectors. Suppose there are totally m unknown parameters. With missing-censoring data, the estimable variables are $\delta = I(T \leq C)$ and $V = T\delta$, and thus, the estimable quantities are $E(\delta)$ and $E(V^k) = E(T^k\delta)$, $k = 1, 2, \dots$. Denote $\phi_k(\theta) = \int_0^\infty t^k f(t; \theta_1) K(t; \theta_2) dt$, $k = 0, 1, \dots, m$. The quantities ϕ_k can be estimated consistently by $W_k = (\sum_{i=1}^n T_i^k \delta_i) / n$. Replace $\phi_k(\theta)$ by W_k , $k = 0, 1, \dots, m$, we can get a system with m unknown variables. Let $\theta = (\theta_1, \theta_2)$, and $\Phi = (\phi_0, \dots, \phi_m)$. Since the determinant $|\partial\Phi/\partial\theta|$ is usually not zero, the unknown variables can be solved uniquely. Thus, under parametric assumption, the lifetime distribution is usually identifiable even censoring is missing.

3. Nonparametric Estimation of Lifetime Distribution

It is well known that, without parametric assumption, the lifetime distribution in general is not identifiable. However, as we will show, when the censoring distribution is known, the lifetime distribution is estimable or identifiable.

3.1 Estimation with discrete variables

Assume that both the failure time T and the censoring time C are discrete and each may take values $1, 2, \dots$. Let $f(t) = P(T = t)$, $g(t) = P(C = t)$, $S(t) = \sum_{s \geq t} f(s)$ and $K(t) = \sum_{s \geq t} g(s)$. Let $V = T\delta$ and $h(t) = P(V = t)$. Since $h(t) = P(T = t, \delta = 1) = P(T = t, T \leq C) = f(t)K(t)$, $f(t) = P(V = t)/K(t)$. Thus, the estimator for $f(t)$ is,

$$\hat{f}(t) = \frac{d_t}{nK(t)},$$

where d_t is the number of observed failures at time t . It is easy to show that the estimator $\hat{f}(t)$ is unbiased with variance

$$\text{var}\{\hat{f}(t)\} = \frac{f(t)\{1 - f(t)K(t)\}}{\{nK(t)\}}.$$

Remark. This estimator has been derived by Hu, Lawless, and Suzuki (1998) as maximum likelihood estimator and moment estimator. Again, the derivation here is much simpler.

3.2 Estimation with continuous variables

Now assume both T and C are continuous. Denote the density function and the survival function of T as $f(t)$ and $S(t) = P(T > t)$ respectively. Denote the survival function of $V = T\delta$ as $S_V(t) = P(V > t)$. We have, $S_V(t) = P(T > t, \delta = 1) = \int_t^\infty f(s)K(s)ds$. Thus, $dS_V(t) = -f(t)K(t)$, and $f(t) = -dS_V(t)/K(t)$. Thus,

$$S(t) = - \int_t^\infty \frac{dS_V(t)}{K(t)}.$$

Let $H(t)$ be the Heaviside function which is defined as $H(t) = 1$ for $t \geq 0$ and $H(t) = 0$ otherwise. From the observed data, the empirical estimator for $S_V(t)$ is,

$$\hat{S}_V(t) = \frac{1}{n} \sum_{i=1}^n I(V_i > t) = \frac{1}{n} \sum_{i=1}^n \delta_i H(T_i - t).$$

Consequently, an estimator for $S(t)$ is given by

$$\hat{S}(t) = - \int_t^\infty \frac{d\hat{S}_V(s)}{K(s)} = \frac{1}{n} \int_t^\infty \sum_{i=1}^n \delta_i \frac{H'(T_i - s)}{K(s)} ds = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} H(T_i - t). \tag{3.1}$$

Let $\tau = \inf\{t : K(t) = 0\}$. As shown in Appendix B, the estimator is unbiased if $S(\tau) = 0$, and the variance of $\sqrt{n}\hat{S}(t)$ is given by

$$S(t)\{1 - S(t)\} + \int_0^\infty \frac{\lambda^c(u)}{K(u)} \{S(t \vee u) - 2S(t)S(t \vee u) + S^2(t)S(u)\} du,$$

where $\lambda^c(t)$ is the hazard function of censoring time C and \vee is maximum operator defined as $a \vee b = \max(a, b)$.

4. Numerical Simulations

To investigate the behavior of the proposed estimators and the loss of efficiency due to censoring and missing-censoring under parametric assumption, we conduct some simulations with different failure and censoring distributions. First we check the case when both failure time and censoring time are exponential with mean $1/\lambda$ and $1/\phi$ respectively. The estimators $\hat{\lambda}_0$, $\hat{\lambda}_1$, and $\hat{\lambda}_2$ corresponding to the situations of no censoring, random-censoring, and missing-censoring, are given in subsection 2.1. A simulation is carried out with 100 iterations and sample size 50. The results are summarized in Table 1.

Table 1: Parametric estimation when failure time is exponential

λ, ϕ	No censoring $\hat{\lambda}_0, \text{bias}(\hat{\lambda}_0), \text{var}(\hat{\lambda}_0)$	Random-censoring $\hat{\lambda}_1, \text{bias}(\hat{\lambda}_1), \text{var}(\hat{\lambda}_1)$	Missing-censoring $\hat{\lambda}_2, \text{bias}(\hat{\lambda}_2), \text{var}(\hat{\lambda}_2)$
2.0, 1.5	2.042, 0.0421, 0.09737	2.045, 0.0445, 0.17484	2.103, 0.1030, 0.24632
2.0, 2.0	2.042, 0.0421, 0.09737	2.061, 0.0607, 0.19929	2.073, 0.0728, 0.25556
2.0, 2.5	2.042, 0.0421, 0.09737	2.045, 0.0459, 0.22301	2.063, 0.0630, 0.29554

From Table 1 we can see that all these three estimators perform quite well even when censoring is heavy. As we can notice that the estimated values and the theoretical values derived in subsection 2.1 are closely agreeable. For instance, when $\lambda = 2$ and $\phi = 2$, the theoretical asymptotic variance values are: $\text{var}(\hat{\lambda}_0) = \lambda^2/n = 4/50 = 0.08$, $\text{var}(\hat{\lambda}_1) = \lambda(\lambda + \phi)/n = 5/50 = 0.18$, $\text{var}(\hat{\lambda}_2) = \lambda(\lambda + 2\phi)/n = 12/50 = 0.24$. The estimated values are 0.09737, 0.19929, and 0.25556, respectively. Also, the differences between the estimated $\text{var}(\hat{\lambda}_1)$ and $\text{var}(\hat{\lambda}_2)$ are quite small, indicating loss of efficiency due to missing-censoring is small.

Now we check the case when the failure time follows a gamma distribution with shape parameter α and scale parameter λ and the censoring time follows an exponential distribution with mean 0.5. The estimators for α and λ are given in subsection 2.2. A simulation is conducted to assess the performance of these estimators. We use the algorithm proposed by Cheng and Feast (1980) to produce a gamma distribution. In this simulation, we use 100 iterations, and since we find that sample size 50 is too small to get desirable results, we take sample size 500. As mentioned by Hu, Lawless and Suzuki (1998), for missing-censoring data, the sample size usually is large, thus taking sample size 500 is reasonable. Table 2 gives the results.

Table 2: Parametric estimation when failure time is gamma

	No censoring	Random-censoring	Missing-censoring
$\alpha = 3$	2.959, -0.041, 0.04219	3.091, 0.091, 0.36577	3.091, 0.091, 0.36577
$\lambda = 2$	1.973, -0.027, 0.01953	2.172, 0.172, 0.78302	2.148, 0.148, 0.53299
$\alpha = 3$	2.959, -0.041, 0.04220	3.006, 0.006, 0.23242	3.006, 0.006, 0.23242
$\lambda = 3$	2.959, -0.041, 0.04394	3.054, 0.054, 0.62545	3.043, 0.043, 0.51500

The three numbers in each cell are: estimated value, estimated bias, and estimated variance for the parameter under different data settings.

The table shows again that all these three estimators perform quite well. It is interesting to notice that the estimated variance for estimator $\hat{\lambda}_2$ is less than that of the estimator $\hat{\lambda}_1$, this implies that observed censoring times do not provide much information in the estimation of lifetime distribution, and thus the mismeasurement for censoring will not significantly affect the estimation of lifetime distribution. Such conclusion is quite agreeable with the statement made by Meier *et al.* (2003).

5. Discussion

This paper considers the estimation of the lifetime distribution with missing-censoring data by using a kind of empirical estimation method, rather than the traditional maximum likelihood method. The estimators and their biases and variances are derived in a simple and uniform way. Theoretical analysis and simulation studies show that the loss of information due to missing-censoring is negligible if the model is correctly specified. Compared with the traditional maximum likelihood method, the approach advocated in this paper is much simpler and much easier to implement. As we can see, it would be quite hard to tackle the parameter estimation for gamma distribution by using maximum likelihood method. The method used in the present paper might be also applicable to other censored data.

Appendix A

To obtain the bias and the variance of the estimator $\hat{\lambda}_1$ in (2.2), denote $W_0 = \sum_{i=1}^n \delta_i$, $W_1 = \sum_{i=1}^n X_i$, and,

$$\frac{W_0 - E(W_0)}{E(W_0)} = \delta W_0, \quad \frac{W_1 - E(W_1)}{E(W_1)} = \delta W_1.$$

We have,

$$\begin{aligned}\hat{\lambda}_1 - \lambda &= \frac{W_0}{W_1} - \lambda = \frac{E(W_0)(1 + \delta W_0)}{E(W_1)(1 + \delta W_1)} - \lambda = \lambda \left(\frac{1 + \delta W_0}{1 + \delta W_1} \right) - \lambda \\ &= \lambda \{ \delta W_0 - \delta W_1 - \delta W_0 \delta W_1 + (\delta W_1)^2 + \text{terms of order 3 or more} \}.\end{aligned}$$

Ignoring the terms of order 3 or more, we have,

$$\begin{aligned}\text{bias}(\hat{\lambda}_1) &= E(\hat{\lambda}_1 - \lambda) = -\lambda E(\delta W_0 \delta W_1) + \lambda E(\delta W_1)^2 \\ &= -\lambda \frac{\text{cov}(W_0, W_1)}{E(W_0)E(W_1)} + \frac{\lambda}{E^2(W_1)} \text{var}(W_1).\end{aligned}$$

When both failure time and censoring time are exponential, the data fit the propositional hazard model. As shown in Chen *et al.* (1982), δ and X are independent. Thus, $\text{cov}(W_0, W_1) = \text{cov}(\sum \delta_i, \sum X_i) = n \text{cov}(\delta, X) = 0$. Since X follows exponential distribution with mean $1/(\lambda + \phi)$, we know $E(W_1) = nE(X) = n(\lambda + \phi)^{-1}$, and $\text{var}(W_1) = n \text{var}(X) = n(\lambda + \phi)^{-2}$. Hence, $\text{bias}(\hat{\lambda}_1) = \lambda/n$. Similarly,

$$\begin{aligned}\text{var}(\hat{\lambda}_1) &\approx E(\hat{\lambda}_1 - \lambda)^2 \approx \lambda^2 E\{(\delta W_0)^2 + (\delta W_1)^2 - 2\delta W_0 \delta W_1\} \\ &= \lambda^2 \frac{\text{var}(W_0)}{E^2(W_0)} + \lambda^2 \frac{\text{var}(W_1)}{E^2(W_1)} - 2\lambda^2 \frac{\text{cov}(W_0, W_1)}{E(W_0)E(W_1)} \\ &= \lambda^2 \frac{n \frac{\lambda \phi}{(\lambda + \phi)^2}}{n^2 \left(\frac{\lambda}{\lambda + \phi}\right)^2} + \lambda^2 \frac{n \frac{1}{(\lambda + \phi)^2}}{n^2 \left(\frac{1}{\lambda + \phi}\right)^2} - 0 \\ &= \frac{\lambda \phi}{n} + \frac{\lambda^2}{n} = \frac{\lambda(\lambda + \phi)}{n}.\end{aligned}$$

Now consider the bias and variance of estimator $\hat{\lambda}_2$. Let $W_0 = \sum_{i=1}^n \delta_i$, $W_1 = \sum_{i=1}^n T_i \delta_i$, and denote,

$$\frac{W_0 - E(W_0)}{E(W_0)} = \delta W_0, \quad \frac{W_1 - E(W_1)}{E(W_1)} = \delta W_1.$$

So,

$$\begin{aligned}\hat{\lambda}_2 - \lambda &= \frac{W_0^2}{nW_1} - \lambda = \frac{E^2(W_0)(1 + \delta W_0)^2}{nE(W_1)(1 + \delta W_1)} - \lambda = \lambda \left\{ \frac{(1 + \delta W_0)^2}{1 + \delta W_1} - 1 \right\} \\ &= \lambda \{ 2\delta W_0 - \delta W_1 + (\delta W_0)^2 + (\delta W_1)^2 - 2\delta W_0 \delta W_1 + o(|\delta W_0| + |\delta W_1|)^2 \}.\end{aligned}$$

Thus,

$$\begin{aligned}\text{bias}(\hat{\lambda}_2) &\approx \lambda E(\delta W_0)^2 + \lambda E(\delta W_1)^2 - 2\lambda E(\delta W_0 \delta W_1) \\ &= \lambda \frac{\text{var}(W_0)}{E^2(W_0)} + \lambda \frac{\text{var}(W_1)}{E^2(W_1)} - 2\lambda \frac{\text{cov}(W_0, W_1)}{E(W_0)E(W_1)}.\end{aligned}$$

And,

$$\text{var}(\hat{\lambda}_2) \approx 4\lambda^2 E(\delta W_0)^2 + \lambda^2 E(\delta W_1)^2 - 4\lambda^2 E(\delta W_0 \delta W_1).$$

By some calculation, we find, $E(W_0) = n\lambda/(\lambda + \phi)$, $\text{var}(W_0) = n\lambda\phi/(\lambda + \phi)^2$, $E(W_1) = n\lambda/(\lambda + \phi)^2$, $\text{var}(W_1) = n\lambda(\lambda + 2\phi)/(\lambda + \phi)^4$, $\text{cov}(W_0, W_1) = n\lambda\phi/(\lambda + \phi)^3$. Plugging in these values in above bias and variance expressions, we obtain,

$$\text{bias}(\hat{\lambda}_1) = \frac{\lambda + \phi}{n},$$

and

$$\text{var}(\hat{\lambda}_1) = \frac{\lambda(\lambda + 2\phi)}{n}.$$

Appendix B

To study the estimator $\hat{S}(t)$ in (3.1), we use martingale arguments since the parametric method in Appendix A is not suitable for nonparametric estimation.

For the mean of $\hat{S}(t)$, we have

$$\begin{aligned} \hat{S}(t) - S(t) &= 1 - \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} \{I(T_i \geq t) - S(t)\} \\ &= \hat{S}(\infty) + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} \{I(T_i \geq t) - S(t)\}. \end{aligned}$$

Then, $E\{\hat{S}(\infty)\} = 1 - E\{I(T < C)/K(T)\} = 1 - \int_0^\tau f(t)dt = S(\tau)$, and

$$\begin{aligned} &E\left[\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} \{I(T_i \geq t) - S(t)\}\right] \\ &= E\left[\frac{I(T \leq C)}{K(T)} \{I(T \geq t) - S(t)\}\right] \\ &= S(t) - S(\tau) - S(t)\{1 - S(\tau)\} \\ &= -S(\tau)\{1 - S(t)\}. \end{aligned}$$

Thus, $E\{\hat{S}(t) - S(t)\} = S(\tau)S(t)$. So, $\hat{S}(t)$ is unbiased when $S(\tau) = 0$.

For the asymptotic variance of $\sqrt{n}\hat{S}(t)$, let $\lambda^c(t)$ denote the hazard function for the censoring distribution. As in Zhao and Tsiatis (1997), define a filtration $F(u)$ as the σ -algebras generated by

$$\sigma\{I(T_i \leq u, T_i \leq C_i), I(C_i \leq u), K(s), 0 \leq s < \infty, i = 1, \dots, n\}.$$

The corresponding martingale process $M_i^c(u)$ can be expressed as

$$M_i^c(u) = N_i^c(u) - \int_0^u \lambda^c(t) Y_i(t) dt,$$

where

$$N_i^c(u) = I(C_i \leq u, C < T_i), \quad Y_i(u) = I(T_i \geq u, C_i \geq u).$$

By the equality from Robin and Rotnitzky (1992, p313),

$$\frac{\delta_i}{K(T_i)} = 1 - \int_0^\infty \frac{dM_i^c(u)}{K(u)}.$$

Since $\hat{S}(\infty)$ is a constant, we can assume $\hat{S}(\infty) = 0$. Thus,

$$\begin{aligned} n^{\frac{1}{2}} \{\hat{S}(t) - S(t)\} &= n^{-\frac{1}{2}} \sum_{i=1}^n \frac{\delta_i}{K(T_i)} \{I(T_i \geq t) - S(t)\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \{I(T_i \geq t) - S(t)\} - n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\infty \frac{dM_i^c(u)}{K(u)} \{I(T_i \geq t) - S(t)\}. \end{aligned}$$

From the results for empirical estimator, the asymptotic variance from the first term goes to $S(t)\{1 - S(t)\}$. From the martingale theory, the second one converges to,

$$\begin{aligned} E \int_0^\infty \{I(T_i \geq t) - S(t)\}^2 \frac{I(T_i \geq u) I(C_i \geq u) \lambda^c(u)}{K^2(u)} du \\ = \int_0^\infty \frac{\lambda^c(u)}{K(u)} \{S(t \vee u) - 2S(t)S(t \vee u) + S^2(t)S(u)\} du. \end{aligned}$$

As claimed in Zhao and Tsiatis (1997), these two terms are uncorrelated. Hence, the asymptotic variance of $\sqrt{n}\hat{S}(t)$ is

$$S(t)\{1 - S(t)\} + \int_0^\infty \frac{\lambda^c(u)}{K(u)} \{S(t \vee u) - 2S(t)S(t \vee u) + S^2(t)S(u)\} du.$$

References

- Chen, Y.Y., Hollander, M. and Langberg, N. A. (1982). Small-sample results for Kaplan-Meier estimator. *Journal of the American Statistical Association* **77**, 141-144.
- Cheng, R. C. and Feast, G. M. (1980). Gamma variate generators with increased shape parameter range. *Communications of the ACM* **23**(7), 389-394.

- Hu, X. J., Lawless, J. F. and Suzuki, K. (1998). Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics* **40**, 3-13.
- Kalbfleisch, J. D., Lawless, J. F. and Robinson, J.A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics* **33**, 273-285.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.
- Meier, A. S., Richardson, B. A. and Hughes, J. (2003). Discrete proportional hazards models for mismeasured outcomes. *Biometrics* **59**, 947-954.
- Phillips, J. F. and Sweeting, T. J. (1996). Estimation for censored exponential data when the censoring times are subject to error. *Journal of the Royal Statistical Society. Series B* **58**, 775-783.
- Singh, B. (2002). Inferences concerning exponential distributions in the presence of randomly right censored data with missing censored values. *Lifetime Data Analysis* **8**, 69-88.
- Suzuki, K. (1985). Estimation of lifetime parameters from incomplete field data. *Technometrics* **27**, 263-271.
- Zhao, H. and Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84**, 339-348.

Received December 16, 2009; accepted May 18, 2010.

Jiantian Wang
Department of Mathematics
Kean University
Union, NJ, 07083, USA
jwang@kean.edu