Journal of Statistical
Theory & Methods
1997, Vol. 8, No. 2, pp. 225 ~ 230

A Simple Estimator of Mean Residual Life Function under Random Censoring[1]

Dong-Myung Jeong · Myung-Unn Song · Jae-Kee Song[2]

Abstract

We, in this paper, propose an estimator of mean residual life function by using the residual survival function under random censoring and prove the uniform consistency and weak convergence result of this estimator. Also an example is illustrated by the real data.

Key Words and Phrases: Mean residual life function, Residual survival function, Censored data.

1. Introduction

Let X be a nonnegative random variable with continuous distribution function F and let us define the mean residual life(MRL) function or remaining life expectancy at age x as

$$e(x) = E[X - x|X > x]$$

for $S(x) > 0$, where $S(x) = \Pr[X > x]$ is the survival function of X and $e(x) = 0$ whenever $S(x) = 0$. Note that $e(x)$ is the mean of the remaining lifetime given survival up to time x and is the usual mean if $x = 0$, and uniquely determines the distribution function F via an inversion formula. The MRL function plays very important role in many practical engineering areas and in other applications such as actuarial science and medical research. Hence the estimation problem of MRL function has been investigated by many authors through parametric and nonparametric approaches.

---

[1]This paper was partially supported by the Basic Science Research Institute Programs, Ministry of Education, 1997, Project No. BSRI-97-1403.

[2]Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

Let $T_i, i = 1, 2, \cdots, n$ be independent and identically distributed (i.i.d.) random variables(r.v.'s) with continuous distribution function (d.f.) F and hazard rate function α, and let $C_i, i = 1.2, \cdots, n$ be i.i.d. r.v.'s with d.f. G. Suppose that the two sequences $\{T_i\}_{i=1}^n$ and $\{C_i\}_{i=1}^n$ are independent. We refer to the T_i's as lifetimes and to the C_i's as censoring times. In the random censorship model from the right, the T's may be censored on the right by the C's, so that we only observe the pairs (X_i, δ_i), $i = 1, 2, \cdots, n$, where $X_i = (T_i \wedge C_i)$ and $\delta_i = I(T_i \leq C_i)$. Here and in the sequel, $I(A)$ denotes te indicator function of the event A, $a \wedge b = \min(a, b)$. Thus the X's are i.i.d. r.v's with d.f. H given by $H(t) = 1 - [1 - F(t)][1 - G(t)], 0 \leq t < \infty$.

Now let $\tau = \sup\{t : S(t) > 0\}$. Then under the random censorship model, the MRL function may be written as

$$e(x) = \frac{1}{S(x)} \int_x^\tau S(t)\, dt. \tag{1}$$

for $S(x) > 0$. Yang(1977) proposed the Nelson-Aalen type estimator for (1). Kumazawa(1987) extended the definition of the process based on Yang's(1977) estimator and constructed the Kaplan-Meier type estimator for (1). Jeong *et al.*(1996) introduced the smoothing estimator for (1) and its asymptotic properties.

Let $S_x(t)$ be a conditional probability of surviving age $x+t$ given survival to age x. Then we get

$$S_x(t) = \Pr(T > x + t \,|\, T > x) = \frac{S(x+t)}{S(x)}. \tag{2}$$

Here $S_x(\cdot)$ is called a 'residual survival function' of an individual at age x, and is applied to setting rates and benefits for life insurance.

The residual survival function (2) may be estimated from the only remaining failure times after time x by the product-limit method as follows:

$$\widehat{S}_x(t) = \prod_{(x,x+t]} \left(1 - \frac{dN(u)}{Y(u)}\right), \tag{3}$$

where $N(t) = \Sigma_{i=1}^n I(X_i \leq t, \delta_i = 1)$ is a right-censored counting process and $Y(t) = \Sigma_{i=1}^n I(X_i \geq t)$ is the number at risk just before time t. Note that $\widehat{S}_x$ is a product-limit estimator of S_x and is reduced to the Kaplan-Meier estimator defined by Kaplan and Meier(1958) if $x = 0$ (See, Anderson *et al.*(1993)). Also the asymptotic distribution of (3) is given by an obvious modification of that of the Kaplan-Meier estimator. Therefore $\widehat{S}_x(t)$ is uniform consistent and has weak convergence result, that is, for $x \in [0, \tau)$,

$$\sqrt{n}\{\widehat{S}_x - S_x\} \;\xrightarrow{d}\; -S_x \cdot Z_x, \tag{4}$$

where Z_x is a mean zero Gaussian process with covariance function

$$Cov\{Z_x(s), Z_x(t)\} = \int_x^{x+(s\wedge t)} \frac{\alpha(u)\, du}{\{1 - F(u)\}\{1 - G(u-)\}}.$$

On the other hand, the MRL function can be represented by using the residual survival function S_x, that is,

$$e(x) = \frac{1}{S(x)} \int_x^\tau S(t)\,dt = \int_0^\tau S_x(t)\,dt. \tag{5}$$

Thus by substituting the estimator $\widehat{S}_x$ for S_x in (5), we propose an estimator of MRL function as

$$\widehat{e}_n(x) = \int_0^{X^*} \widehat{S}_x(t)\,dt, \tag{6}$$

where $X^* = \max(X_1, X_2, \cdots, X_n)$.

2. Main Results

The following two theorems provide the uniform consistency and weak convergence results of the proposed estimator (6).

Theorem 1. Let $\sqrt{n}(\tau - X^*) \xrightarrow{p} 0$. Then as $n \to \infty$,

$$\sup_{x \in [0,\tau)} |\widehat{e}_n(x) - e(x)| \xrightarrow{p} 0.$$

Proof. For a fixed $x \in [0, \tau)$,

$$\begin{aligned} |\widehat{e}_n(x) - e(x)| &= \left| \int_0^{X^*} \widehat{S}_x(t)\,dt - \int_0^\tau S_x(t)\,dt \right| \\ &= \left| \int_0^\tau [\widehat{S}_x(t) - S_x(t)]\,dt - \int_{X^*}^\tau \widehat{S}_x(t)\,dt \right| \\ &\le \int_0^\tau |\widehat{S}_x(t) - S_x(t)|\,dt + \int_{X^*}^\tau \widehat{S}_x(t)\,dt. \end{aligned}$$

Here, the first term of the right-hand side of the inequality converges to zero in probability by the uniform consistency result of $\widehat{S}_x(t)$, Therefore it suffices to show that $\sqrt{n} \int_{X^*}^\tau \widehat{S}_x(t)\,dt \xrightarrow{p} 0$, as $n \to \infty$.

Since

$$\sqrt{n} \int_{X^*}^\tau \widehat{S}_x(t)\,dt = \sqrt{n} \int_{X^*}^\tau [\widehat{S}_x(t) - S_x(t)]\,dt + \sqrt{n} \int_{X^*}^\tau S_x(t)\,dt.$$

By the convergence result of $\widehat{S}_x(t)$ and by the above assumption, this converges to zero in probability. Thus the result follows.

Theorem 2. Suppose $\sqrt{n}(\tau - X^*) \xrightarrow{p} 0$. Then as $n \to \infty$,

$$\sqrt{n}\{\widehat{e}_n - e\} \xrightarrow{d} W_x = -\int_0^\tau S_x(t) Z_x(t)\,dt.$$

Here the limiting distribution $Z_x(\cdot)$ is defined in (4) and the covariance function of $W_x(\cdot)$ is $Cov[W_x(s), W_x(t)] = \sigma_x^2(s \wedge t)$, where

$$\sigma_x^2(t) = \int_x^{x+t} \Big(\int_0^\tau S_x(t)\, dt \Big)^2 \frac{\alpha(u)\, du}{[1 - F(u)][1 - G(u-)]}.$$

Proof. For a fixed $x \in [0, \tau)$,

$$\begin{aligned} \sqrt{n}\{\widehat{e}_n(x) - e(x)\} &= \sqrt{n}\Big(\int_0^{X^*} \widehat{S}_x(t)\, dt - \int_0^\tau S_x(t)\, dt \Big) \\ &= \int_0^\tau \sqrt{n}\{\widehat{S}_x(t) - S_x(t)\}\, dt - \sqrt{n} \int_{X^*}^\tau \widehat{S}_x(t)\, dt. \end{aligned}$$

The first term of the right-hand side of the second equality converges weakly to a Gaussian process by the convergence result of $\widehat{S}_x$ and the second term converges to zero in probability by the above assumption of this theorem. Thus the result follows by the Slutsky's theorem.

Remark 1. Under the conditions of Theorem 2, we have, for the $t < \tau$, as $n \to \infty$,

$$\widehat{\sigma}_x^2(t) \xrightarrow{p} \sigma_x^2(t)$$

where

$$\widehat{\sigma}_x^2(t) = n \int_x^{x+t} \Big(\int_0^\tau \widehat{S}_x(t)\, dt \Big)^2 \frac{J(u)}{[y(u)]^2}\, dN(u).$$

This means that $\widehat{\sigma}_x^2$ is a consistent estimator of the variance function of the limiting process W_x in Theorem 2.

Remark 2. The residual survival function estimator is estimated by using only the remaining failure times after time x while the Kaplan-Meier estimator is estimated using the all observed failure times. This explains the difference between the two. However, the asymptotic distribution of the residual survival function estimator is analogous to that of the Kaplan-Meier estimator. Furthermore, in practical calculation, the estimates of (6) are the same as the Kaplan-Meier type estimates proposed by Kumazawa(1987). Thus if we want to estimate the remaining life after certain time then the proposed estimator (6) is simpler and more useful than the Kaplan-Meier type estimator because that (6) is estimated by using the partial data.

3. An Example

As an example, let us consider the well-known acute myelogenous leukemia (AML) maintenance study. A clinical trial to evaluate the efficacy of maintenance

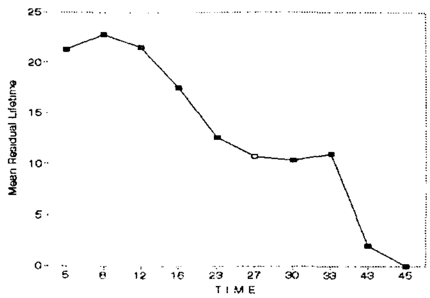

Figure 1: Estimates of $\widehat{e}_n(x)$ for AML data

chemotherapy for AML was conducted by Embury *et al.*(1977) at Stanford University. After reaching a state of remission through treatment by chemotherapy, the patients who entered the study were randomized into two groups. The first group received maintenance chemotherapy; the second or control group did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse, that is, increased the length of remission. For a preliminary analysis during the course of the trial the data which consist of length of complete remission, in weeks, were obtained. In this example, we use only nonmaintained group data, which are censored 1 of the 12 observations. The data are as follows : 5, 5, 8, 8, 12, 16*, 23, 27, 30, 33, 43, 45, where '$*$' denotes a censored.

Figure 1 displays the values of estimates of MRL function at all observed time points for nonmaintained group data. In this figure, one can see that the mean residual lifetime decreases as the length of remission increases.

References

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

2. Embury, S. H., Elias, L., Heller, P. H., Hood, C.E., Greenberg, P. L., and Schrier, S. L. (1977). Remission maintenance therapy in acute myelogenous leukemia, *Western Journal of Medicine*, **126**, 267-272.

3. Jeong, D. M., Song, M. U., and Song, J. K. (1996). Smoothing mean residual life with censored data, *The Korean Communications in Statistics*, **3**, 129-138.

4. Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457-481.

5. Kumazawa, Y. (1987). A note an estimator of life expectancy with random censorship, *Biometrika*, **74**, 655-658.

6. Yang, G. L. (1977). Life expectancy under random censorship, *Stochastic processes and their applications*, **6**, 33-39.