Journal of the Korean
Statistical Society
Vol. 26, No. 3, 1997

Mean Lifetime Estimation with Censored Observations

Jinheum Kim[1] and Jeehoon Kim [2]

Abstract

In the simple linear regression model $Y = \alpha_0 + \beta_0 Z + \varepsilon$ under the right censorship of the response variables, the estimation of the mean lifetime $E(Y)$ is an interesting problem. In this paper we propose a method of estimating $E(Y)$ based on the observations modified by the arguments of Buckley and James (1979). It is shown that the proposed estimator is consistent and our proposed procedure in the simple linear regression case can be naturally extended to the multiple linear regression. Finally, we perform simulation studies to compare the proposed estimator with the estimator introduced by Gill (1983).

Key Words : Mean lifetime; Buckley and James estimator; Consistency; Regression; Product-limit estimator; Censoring.

[1]Department of Applied Statistics, University of Suwon, Kyonggido, 445-743, Korea.
[2]Statistical Research Institute, College of Natural Sciences, Seoul National University, Seoul, 151-742, Korea.

1. INTRODUCTION

Let $\{Y_i, i = 1, \ldots, n\}$ be n independent random variables(rv's) satisfying

$$Y_i = \alpha_0 + \beta_0 Z_i + \varepsilon_i \quad (i = 1, \ldots, n), \tag{1.1}$$

where the ε_i are independent and identically distributed(iid) rv's with mean 0 and the Z_i are independent rv's with compact support independent of $\{\varepsilon_i\}$. Suppose that the responses Y_i are not completely observable under the right censorship of the response variables and that the observations are (X_i, Z_i, Δ_i), where $X_i = Y_i \wedge C_i$, $\Delta_i = I_{\{Y_i \le C_i\}}$, and the C_i are independent rv's independent of $\{\varepsilon_i\}$. $\wedge$ denotes the minimum and I_A is the indicator function of the set A.

Let the partial residuals $\eta_i^{\beta_0} = Y_i - \beta_0 Z_i$ be iid with common continuous distribution F_0. Independent of $\{\eta_i^{\beta_0}\}$, let $\xi_i^{\beta_0} = C_i - \beta_0 Z_i$ be also independent rv's with common distribution function G_0. Classically, F_0 is estimated by the product-limit estimator, introduced by Kaplan and Meier (1958). For any real b, define processes N_b and Y_b on $[0, \infty)$ by

$$N_b(t) = \#\{i \le n : \eta_i^b \wedge \xi_i^b \le t,\ \Delta_i = 1\},$$

$$Y_b(t) = \#\{i \le n : \eta_i^b \wedge \xi_i^b \ge t\},$$

where $\#A$ denotes the number of elements of a set A. Then, the product-limit estimator $\hat{F}_b$ calculated from the $\eta_i^b \wedge \xi_i^b$ is given by

$$\hat{F}_b(t) = 1 - \prod_{u \le t} \left\{1 - \frac{\Delta N_b(u)}{Y_b(u)}\right\},$$

where $\Delta N_b(u) = N_b(u) - N_b(u-)$. To overcome problems of definition if the $\max_i(\eta_i^b \wedge \xi_i^b)$ is censored, the product-limit estimator is modified by defining $\hat{F}_b(\max_i(\eta_i^b \wedge \xi_i^b)) = 1$. Define also times $\tau_{F_0} = \sup\{t : F_0(t) < 1\}$ and $\tau_{G_0} = \sup\{t : G_0(t) < 1\}$.

Assuming the Z_i to be nonrandom, Buckley and James (1979) proposed the following method to estimate α_0 and β_0. They started by replacing Y_i by

$$Y_i^* = Y_i \Delta_i + E(Y_i | Y_i > C_i)(1 - \Delta_i), \tag{1.2}$$

by noting the expectation identity $E(Y_i^*) = E(Y_i) = \alpha_0 + \beta_0 Z_i$, and then regressing the Y_i^* instead of the Y_i on the Z_i like the usual least squares estimation to obtain

$$\tilde{\beta} = \left\{\sum_{i=1}^{n} Y_i^*(Z_i - \bar{Z})\right\} / \sum_{i=1}^{n} (Z_i - \bar{Z})^2,$$

$$\tilde{\alpha} = \bar{Y}^* - \tilde{\beta}\bar{Z},$$

where $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$ and $\bar{Y}^* = n^{-1}\sum_{i=1}^n Y_i^*$. However, since $E(Y_i|Y_i > C_i)$ in (1.2) is unknown, they adopted a self-consistency approach and estimated it from the product-limit estimates $\hat{F}_b$. They then replaced $E(Y_i|Y_i > C_i)$ by

$$\hat{E}_b(Y_i|Y_i > C_i) = C_i + \int_{u>\xi_i^b}\{1 - \hat{F}_b(u)\}du/\{1 - \hat{F}_b(\xi_i^b)\}. \qquad (1.3)$$

Replacing (1.2) by $\hat{Y}_i(b) = Y_i\Delta_i + \hat{E}_b(Y_i|Y_i > C_i)(1 - \Delta_i)$, they proposed to estimate β_0 by iterative solution of the equation

$$b = \left\{\sum_{i=1}^n \hat{Y}_i(b)(Z_i - \bar{Z})\right\} / \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

For the limiting value $\hat{\beta}$ of b, the associated estimate of the intercept α_0 is

$$\hat{\alpha} = \bar{Y}(\hat{\beta}) - \hat{\beta}\bar{Z},$$

where $\bar{Y}(\hat{\beta}) = n^{-1}\sum_{i=1}^n \hat{Y}_i(\hat{\beta})$.

Besides Buckley and James (1979), various estimation methods of β_0 under the above model (1.1) have been proposed in the literature; by Miller (1976) and Koul, Susarla and Van Ryzin (1981). Miller and Halpern (1982), comparing these approaches on the basis of the Stanford heart transplant data, found out that the Buckley-James estimator has better performance than the others as regards bias and estimated variance. Also, Buckley and James (1979) found through simulations that their method gave approximately unbiased estimates for the slope parameter β_0 for a wide range of censoring patterns, some depending on the explanatory variable and others not. For the mathematical justifications of the Buckley-James estimator, James and Smith (1984), Smith (1988) and Ritov (1990) discussed the asymptotic properties including consistency of the Buckley-James estimator.

In this paper we propose a method of estimating the mean lifetime $E(Y)$ based on the observations $\{\hat{Y}_i(\hat{\beta}), i = 1, \ldots, n\}$ modified by the arguments of Buckley and James (1979). In the next section we show that the proposed estimator is consistent. An extension to the multiple linear regression is briefly considered in Section 3. Finally, Section 4 performs simulation studies to investigate how the strength of the regression, and the support and distribution of the censoring variable have an effect on the relative performance of the proposed estimator to the estimator introduced by Gill (1983).

2. SIMPLE LINEAR REGRESSION

Since the Z_i are random in our model (1.1), the arguments of Buckley and James (1979) mentioned in Section 1 must be considered as conditioning on Z_i. That is, $E(Y_i|Y_i > C_i)$ in (1.2) is replaced by $E(Y_i|Y_i > C_i, Z_i)$ and $\hat{E}_b(Y_i|Y_i > C_i)$ in (1.3) by $\hat{E}_b(Y_i|Y_i > C_i, Z_i)$. Noting that $\hat{E}(Y) = \hat{\alpha} + \hat{\beta}\bar{Z}$, we propose the estimator of $E(Y)$ as follows;

$$\hat{E}(Y) = n^{-1}\sum_{i=1}^{n}\hat{Y}_i(\hat{\beta}),$$

which may also be motivated as the empirical estimator of $E(Y)$, since the modified $\hat{Y}_i(\hat{\beta})$ are not subject to the censorship any longer.

In order to prove the consistency of $\hat{E}(Y)$, suppose that under the right censorship, the responses Y_i are observed over the entire interval of support, $\tau_{F_0} \le \tau_{G_0}$, and $\tau_{F_0} < \infty$. In the proofs of the following lemmas and theorems, we assume, without loss of generality, that $\beta_0 = 0$ and Z has support on $[0, 1]$.

Lemma 2.1. Let $\{\beta_n\}$ be any nonrandom sequence converging to β_0.
(i) For any τ such that $F_0(\tau) < 1$,

$$\sup_{t\le\tau}|\hat{F}_{\beta_n}(t) - F_0(t)| \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

(ii)

$$\sup_{t\le\tau_{F_0}}|\hat{F}_{\beta_n}(t) - F_0(t)| \xrightarrow{p} 0, \quad \text{as } n \to \infty,$$

where $\xrightarrow{p}$ denotes convergence in probability.

Proof. (i) Since $F_0(\tau) < 1$, Z has support on $[0, 1]$, and F_0 is continuous, $P(Y - \beta Z < \tau) = E\{F_0(\beta Z + \tau)\} \le F_0(|\beta| + \tau) < 1$ for any β in some neighborhood N_0 of 0. Since $\tau_{F_0} \le \tau_{G_0}$, $P(Y \wedge C - \beta Z < \tau) < 1$ for any β in N_0. Hence, using the Lemma 7.1 (i) and 7.2 (i) of Ritov (1990), (i) holds.

(ii) Our proof is similar to the Wang's (1987) arguments. For given $\epsilon > 0$, choose $\tau < \tau_{F_0}$ such that $1 > F_0(\tau) > 1 - \epsilon$. Now for $t \in (\tau, \tau_{F_0}]$,

$$\hat{F}_{\beta_n}(\tau) \le \hat{F}_{\beta_n}(t) \le 1,$$

$$1 - \epsilon < F_0(\tau) \le F_0(t) \le 1.$$

Therefore,

$$\sup_{\tau\le t\le\tau_{F_0}}|\hat{F}_{\beta_n}(t) - F_0(t)| \le \max\{\epsilon, 1 - \hat{F}_{\beta_n}(\tau)\}.$$

Combining this, (i) and the arbitrariness of $\epsilon > 0$, (ii) holds.

Lemma 2.2. Let $\{\beta_n\}$ be any nonrandom sequence converging to β_0. For any $\tau < 0$,

$$\int_{-\infty}^{\tau} \hat{F}_{\beta_n}(u)du \xrightarrow{p} \int_{-\infty}^{\tau} F_0(u)du, \quad \text{as } n \to \infty.$$

Proof.

$$\int_{-\infty}^{\tau} \hat{F}_{\beta_n}(u)du = \tau\hat{F}_{\beta_n}(\tau) - \int_{-\infty}^{\tau} ud\hat{F}_{\beta_n}(u) \xrightarrow{p} \tau F_0(\tau) - \int_{-\infty}^{\tau} udF_0(u)$$

by Lemma 7.1 (ii) and 7.2 (ii) of Ritov (1990) and Lemma 2.1.

Theorem 2.3. Let $\{\beta_n\}$ be any nonrandom sequence converging to β_0. Then,

$$\hat{E}(Y) \xrightarrow{p} E(Y), \quad \text{as } n \to \infty.$$

Proof. Note that

$$\hat{Y}_i(\beta_n) = \Delta_i Y_i + (1-\Delta_i)C_i + (1-\Delta_i)\int_{\xi_i^{\beta_n}}^{\infty} \{1 - \hat{F}_{\beta_n}(u)\}du/\{1 - \hat{F}_{\beta_n}(\xi_i^{\beta_n})\}. \tag{2.1}$$

As $n \to \infty$, by the weak law of large numbers,

$$n^{-1}\sum_{i=1}^{n} \Delta_i Y_i \xrightarrow{p} E(\Delta Y),$$

$$n^{-1}\sum_{i=1}^{n} (1-\Delta_i)C_i \xrightarrow{p} E\{(1-\Delta)C\}.$$

For the last term in (2.1),

$$\left| \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i^{\beta_n}}^{\tau_{F_0}+|\beta_n|}\{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(\xi_i^{\beta_n})} - E\left\{(1-\Delta)\frac{\int_{\xi^0}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi^0)}\right\}\right|$$

$$\le I + J + K,$$

where

$$I = \left| \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i^{\beta_n}}^{\tau_{F_0}+|\beta_n|}\{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(\xi_i^{\beta_n})} - \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i^{\beta_n}}^{\tau_{F_0}+|\beta_n|}\{1-F_0(u)\}du}{1-F_0(\xi_i^{\beta_n})}\right|,$$

$$J = \left| \frac{1}{n} \sum_{i=1}^{n} (1-\Delta_i) \frac{\int_{\xi_i^{\beta_n}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi_i^{\beta_n})} - E\left\{ (1-\Delta) \frac{\int_{\xi^{\beta_n}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi^{\beta_n})} \right\} \right|,$$

$$K = \left| E\left\{ (1-\Delta) \frac{\int_{\xi^{\beta_n}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi^{\beta_n})} \right\} - E\left\{ (1-\Delta) \frac{\int_{\xi^{0}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi^{0})} \right\} \right|.$$

In particular,

$$\begin{aligned} I &\le \sup_{t<\tau_{F_0}+|\beta_n|} \left| \frac{\int_t^{\tau_{F_0}+|\beta_n|} \{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(t)} - \frac{\int_t^{\tau_{F_0}+|\beta_n|} \{1-F_0(u)\}du}{1-F_0(t)} \right| \\ &\le \sup_{\tau_{F_0}\le t<\tau_{F_0}+|\beta_n|} \left| \frac{\int_t^{\tau_{F_0}+|\beta_n|} \{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(t)} \right| + \sup_{t<\tau_{F_0}} \left| \frac{\int_{\tau_{F_0}}^{\tau_{F_0}+|\beta_n|} \{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(t)} \right| \\ &\quad + \sup_{t<\tau_{F_0}} \left| \frac{\int_t^{\tau_{F_0}} \{1-\hat{F}_{\beta_n}(u)\}du}{1-\hat{F}_{\beta_n}(t)} - \frac{\int_t^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(t)} \right| \\ &= I1 + I2 + I3. \end{aligned}$$

Here,

$$I1 \le |\beta_n|, \qquad I2 \le |\beta_n|,$$

and by using Lemma 2.1 and 2.2, and Lemma 1 of James and Smith (1984),

$$I3 \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

Also, $J \xrightarrow{p} 0$ by Chebychev's inequality (with finite variance of C_i), and $K \longrightarrow 0$ by the continuity of F_0 and Lebesgue dominated convergence theorem.

3. MULTIPLE LINEAR REGRESSION

In the multiple linear regression $\underline{Y}_{(n\times1)} = \underline{Z}^T_{(n\times p)}\underline{\beta_0}_{(p\times1)} + \underline{\epsilon}_{(n\times1)}$, the estimator $\hat{\underline{\beta}}$ of $\underline{\beta_0}$ can be obtained by iteratively solving

$$\underline{b} = (\underline{Z}^T\underline{Z})^{-1}\underline{Z}^T\hat{\underline{Y}}(\underline{b}),$$

where $\hat{\underline{Y}}(\underline{b}) = (\hat{Y}_i(\underline{b}))_{(n\times1)}$, $\hat{Y}_i(\underline{b}) = Y_i\Delta_i + \hat{E}_{\underline{b}}(Y_i|Y_i > C_i)(1-\Delta_i)$, and $\hat{E}_{\underline{b}}(Y_i|Y_i > C_i)(1-\Delta_i)$ is the multiple version of (1.3) in Section 1 (James and Smith (1984)). Assuming the $\underline{Z}$ to be random and conditioning on $\underline{Z}$, the proposed procedure of estimating $E(Y)$ in the simple linear regression case can be naturally extended to the multiple linear regression, in which the multiple version estimator of $E(Y)$ is given by

$$\hat{E}(Y) = n^{-1}\sum_{i=1}^{n}\hat{Y}_i(\underline{\hat{\beta}}).$$

Using the same arguments as in the proofs of Lemma 2.1, Lemma 2.2 and Theorem 2.3, we then have;

Theorem 3.1. Let $\{\underline{\beta_n}\}$ be any nonrandom sequence such that $\lim_{n\to\infty}\|\underline{\beta_n} - \underline{\beta_0}\| = 0$, where $\|\cdot\|$ is the norm. Then,

$$\hat{E}(Y) \xrightarrow{p} E(Y), \quad \text{as} \quad n \to \infty.$$

4. SIMULATION RESULTS

To compare the proposed estimator with Gill's estimator such as

$$\int_0^{max_i X_i}\{1 - \hat{F}(u)\}du,$$

where $\hat{F}$ is the product-limit estimator based on $\{(X_i, \Delta_i); i = 1, \ldots, n\}$, some simulation trials were carried out under varying conditions, in which the strength of the regression, and the support and distribution of the censoring variable are included. The simulation reported here had $n = 50$, with Z having a discrete uniform distribution on 50 values, -1.96 (+0.08) 1.96, or uniform(-2,2) and $\beta_0 = 1$; the value of α_0 was varied for the different censoring distributions to achieve about 50% censorship. Four censoring distributions were examined; uniform(0,4) with $\alpha_0 = 2$, corresponding to a censoring distribution with support similar to that of the survival distribution; uniform(-4,8) with $\alpha_0 = 2$, corresponding to a censoring distribution with wider support than that of the survival distribution; uniform(2,2) with $\alpha_0 = 2$, corresponding to Type I censorship; and exponential, with mean 1/3 and $\alpha_0 = 0.4$, corresponding to the heavier censoring at the beginning of the trial. The error term ε was $N(0,\sigma^2)$ with $\sigma = 0.5$ or $\sigma = 2.1$, or $N(0,1)$ truncated at ± 5. There were 1,000 replications for each configuration. Table 1 includes values of bias, mean squared error (mse), and ratio of mse of Gill's estimate to mse of proposed (ratio). All values of bias and mse are the means $(\times 10^4)$ of 1,000 replicates.

Table 1. Simulations Comparing Estimates of Mean Lifetime Based on 1,000 Replications in Each Configuration

Covariate Dist.	Error Dist.	Censoring Dist.	Gill Bias	Gill Mse	Proposed Bias	Proposed Mse	Ratio
Discrete Uniform	$N(0,0.5^2)$	$U(0,4)$	-434	481	23	377	1.276
		$U(-4,8)$	44	547	31	336	1.628
		$U(2,2)$	-5328	2937	-60	429	6.846
		$E(1/3)$	-2308	928	6	478	1.941
	Tr. $N(0,1)$ at ± 5[a]	$U(0,4)$	-1035	643	-116	620	1.037
		$U(-4,8)$	-111	830	-72	682	1.217
		$U(2,2)$	-6202	4002	-122	715	5.597
		$E(1/3)$	-3152	1505	-91	716	2.102
	$N(0,2.1^2)$	$U(0,4)$	-3194	1973	-1180	1624	1.215
		$U(-4,8)$	-324	1887	-157	1833	1.029
		$U(2,2)$	-9528	9481	-2321	2030	4.670
		$E(1/3)$	-6173	4708	-2245	1939	2.428
$U(-2,2)$	$N(0,0.5^2)$	$U(0,4)$	-374	487	84	421	1.157
		$U(-4,8)$	-128	568	-60	387	1.468
		$U(2,2)$	-5264	2876	138	454	6.335
		$E(1/3)$	-2319	938	20	447	2.098
	Tr. $N(0,1)$ at ± 5[a]	$U(0,4)$	-937	628	-24	635	0.989
		$U(-4,8)$	-12	820	-35	617	1.329
		$U(2,2)$	-6230	4031	55	789	5.109
		$E(1/3)$	-3118	1486	-176	684	2.173
	$N(0,2.1^2)$	$U(0,4)$	-3294	2104	-1162	1821	1.155
		$U(-4,8)$	-425	1877	-280	1823	1.030
		$U(2,2)$	-9582	9550	-2360	1919	4.977
		$E(1/3)$	-6040	4588	-2108	1916	2.395

[a] It refers to the standard normal distribution truncated at ± 5.

Conclusions from the results in Table 1 are as follows.

(i) Our proposed estimator is always preferred; both its bias and mse are consistently smaller than those of Gill's estimator.

(ii) These two methods are most effective under the strongest regression model, $N(0, 0.5^2)$ with $R^2 = 0.85$. This is probably due to the asymmetric effect of the largest σ^2 on censoring; a large positive ε results in a large Y that is likely to be censored. In contrast, a large negative ε results in a value of Y likely to be uncensored. Thus, the weakest regression, $N(0, 2.1^2)$ with $R^2 = 0.25$, has observed survival times more likely to be small to true value.

(iii) Two estimators considered are, in general, less effective when the support of the censoring distribution is narrower. This trend is remarkably found out for Gill's estimator; when the censoring values cover a narrow range relative to that of the survival times, it is probably that high values of Y, likely to correspond to extreme values of Z, will be censored. Thus, under a constant censoring distribution and early censoring, the estimators are based on data with the downwardly censored values at the higher points.

In conclusion, the proposed estimator has substantially better performance than Gill's estimator with no concern for modifications of censored observations. However, our proposed estimator, being smaller than Gill's estimator, is also negatively biased when the regression is weak or when the support of censoring distribution is narrow relative to that of survival distribution. Thus, the proposed method may be effective in the situations where the assumed regression model is appropiate and the censoring values cover wider range than that of survival times, this being difficult to check in practice.

ACKNOWLEDGEMENTS

The authors greatly appreciate two referees for their helpful comments.

REFERENCES

(1) Buckley, J. and James, I. R. (1979). Linear Regression with Censored Data, *Biometrika*, **66**, 429-436.

(2) Gill, R. D. (1983). Large Sample Behaviour of the Product-Limit Estimator on the Whole Line, *The Annals of Statistics*, **11**, 49-58.

(3) James, I. R. and Smith, P. J. (1984). Consistency Results for Linear Regression with Censored Data, *The Annals of Statistics*, **12**, 590-600.

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

(5) Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression Analysis with Randomly Right-Censored Data, *The Annals of Statistics*, **9**, 1276-1288.

(6) Miller, R. G. (1976). Least Squares Regression with Censored Data, *Biometrika*, **63**, 449-464.

(7) Miller, R. G. and Halpern, J. (1982). Regression with Censored Data, *Biometrika*, **69**, 521-531.

(8) Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data, *The Annals of Statistics*, **18**, 303-328.

(9) Smith, P. J. (1988). Asymptotic Properties of Linear Regession Estimators under a Fixed Censorship Model, *Australian Journal of Statistics*, **30**, 52-66.

(10) Wang, J-G. (1987). A Note on the Uniform Consistency of the Kaplan-Meier Estiamtor, *The Annals of Statistics*, **15**, 1313-1316.