한국통계학회논문집 제3권 2호
1996년 8월 pp. 129-138

Smoothing Mean Residual Life with Censored Data[1]

Dong-Myung Jeong[2], Myung-Unn Song[3] and Jae-Kee Song[4]

Abstract

We propose a smoothing estimator of mean residual life function based on Ghorai and Susarla's (1990) smooth estimator of distribution function under random censorship model and provide the asymptotic properties of this estimator. The Monte Carlo simulation is performed to compare the proposed estimator with the other estimators and an example is also given using the real data.

1. Introduction

Let T 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

$$\begin{aligned} e(t) &= E\{T-t \mid T > t\} \\ &= \frac{1}{S(t)} \int_t^{\infty} S(u)\, du \end{aligned}$$

for $S(t) > 0$, where $S(t) = 1-F(t)$ is the survival function of T and $e(t)=0$ whenever $S(t)=0$. Note that $e(t)$ is the mean of the remaining lifetime given survival up to time t and is the usual mean if $t=0$, and uniquely determines the distribution function F via an inversion formula (See, Hall and Wellner (1981)). The MRL 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.

In the case of the complete data, Yang (1978) showed that the empirical estimator of MRL function is asymptotically unbiased, uniformly strong consistent, and converges weakly to a

1) This paper was supported (in part) by the Basic Science Research Institute Program, Ministry of Education, 1994, Project No. BSRI-94-1403.

2) Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
3) Ph. D. Candidate, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
4) Associate Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

Gaussian Process. Shorack and Wellner (1986) treated estimation of MRL function under an uncensored model. In the present of censoring, the estimation of MRL function has been studied by many authors including Yang (1977), Kumazawa (1987) and Ghorai and Rejtö (1987), and so on.

Let $T_1, T_2, \cdots, T_n$, be independent and identically distributed $(i.i.d.)$ random variables (r.v.'s) with continuous distribution function (d.f.) F, and let $C_1, C_2, \cdots, C_n$, 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 will 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 $\delta_i = I(T_i \le C_i)$. Here and in the sequel, $I(A)$ denotes the indicator function of the event A and $a \wedge b = \min(a, b)$. Thus the observed times X'sare $i.i.d.$ r.v's with d.f. H given by $H(x) = 1 - \{1 - F(x)\}\{1 - G(x)\}$ for $0 \le x < \infty$.

Under random censorship model, the MRL function may be written as

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

for $S(x) > 0$, where $\tau_F = \sup\{x : F(x) < 1\}$. Yang (1977) proposed the Nelson-Aalen type estimator $\hat{e}^{\text{NA}}$ for (1) which is defined by

$$\hat{e}^{\text{NA}}(x) = \frac{1}{\hat{S}^{\text{NA}}(x)} \int_x^{X^*} \hat{S}^{\text{NA}}(t)\,dt ,$$

where $X^* = \max_{1 \le i \le n} X_i$ and $\hat{S}^{\text{NA}} = \exp(-\hat{\Lambda})$, $\hat{\Lambda}$ is the Nelson-Aalen estimator of the cumulative hazard function $\Lambda(x) = -\ln S(x)$ (See, Nelson (1972) and Aalen (1978)). She also proved the uniformly strong consistency and weak convergence results of $\hat{e}^{\text{NA}}$.

Kumazawa (1987) extended the definition of the process based on Yang's (1977) estimator and constructed the Kaplan-Meier type estimator $\hat{e}^{\text{KM}}$ for (1) defined as

$$\hat{e}^{\text{KM}}(x) = \frac{1}{\hat{S}^{\text{KM}}(x)} \int_x^{X^*} \hat{S}^{\text{KM}}(t)\,dt ,$$

where $\hat{S}^{\text{KM}}$ is the Kaplan-Meier estimator (See, Kaplan and Meier (1958)) of S. He also provided under some regularity conditions of the weak convergence of the process over the whole line by using the counting processes and established the asymptotic confidence bands for $\hat{e}^{KM}$. On the other hand, Ghorai and Susarla (1990) proposed a kernel estimator of F based on the kernel estimator of density function and also derived the optimal asymptotic properties.

In this paper, we introduce a smoothing estimator of MRL function based on Ghorai and Susarla's estimator of F under random censorship model and prove uniform consistency and weak convergence results. We also compare the performances of the proposed estimators using Monte Carlo simulation, and illustrate an example using the leukemia data.

2. Main Results

With random censored data, Ghorai and Susarla (1990) proposed a kernel estimator of d.f. F by smoothing the Kaplan-Meier estimator, which is defined by

$$\hat{F}(x) = \int \frac{1}{h_n} k\left(\frac{x-y}{h_n}\right) \hat{F}^{KM}(y)dy ,$$

where h_n is a bandwidth or a smoothing parameter and $k(\cdot)$ is a kernel function. The following two lemmas due to Ghorai and Susarla are needed to develop the properties of the our estimator.

Lemma 2.1. Suppose that the kernel k is a bounded probability density function which has finite support and a symmetric about zero. Let, for some positive integer $m \geq 1$,

(i) $\lim_{n\to\infty} h_n^{m+1}(n/\log\log n)^{1/2} = 0,$

(ii) either $\sup_x | f^{(m)}(x) | < \infty$ or $\int | f^{(m)}(x) | dx < \infty.$

Then as $n\to\infty$,

$$\sup_{0\leq x<\tau_F} | \hat{F}(x)-F(x) | \xrightarrow{p} 0 .$$

Lemma 2.2. If $\sqrt{n}\, h_n^m \to 0$ as $n\to\infty$ and $| f^{(m)} |$ is integrable. Then as $n\to\infty$,

$$\sqrt{n}\ \{ \hat{F} - F \} \xrightarrow{d} Z , \tag{2}$$

where Z is a zero mean Gaussian process with covariance function

$$Cov\{Z(s), Z(t)\} = S(s)S(t)\int_0^{s\wedge t} \frac{dF(u)}{S(u)^2(1-G(u))^2} .$$

By substituting the smoothing estimator $\hat{F}(x)$ for $F(x)$ in MRL function, we propose a smoothing estimator $\hat{e}(x)$ given by

$$\hat{e}(x) = \frac{1}{\hat{S}(x)} \int_x^{X^*} \hat{S}(u)du , \tag{3}$$

where $\hat{S}(x) = 1 - \hat{F}(x)$.

Now from the Lemmas 2.1 and 2.2, we obtain the following main asymptotic results of (3).

Theorem 2.1. Suppose $\sqrt{n}\int_{X^*}^{\tau_F} S(x)dx \xrightarrow{p} 0$. Then as $n \to \infty$,

$$\sup_{0 \le x < \tau_F} | \hat{e}(x) - e(x) | \xrightarrow{p} 0 .$$

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

$$\begin{aligned} | \hat{e}(x) - e(x) | &= \left| \frac{1}{\hat{S}(x)} \int_x^{X^*} \hat{S}(t)dt - \frac{1}{S(x)} \int_x^{\tau_F} S(t)dt \right| \\ &= | \hat{S}(x)S(x) |^{-1} \left| S(x)\int_x^{\tau_F} \{ \hat{S}(t)-S(t)\}dt \right. \\ &\quad \left. - \{ \hat{S}(x)-S(x)\} \int_x^{\tau_F} S(t)dt - S(x)\int_{X^*}^{\tau_F} \hat{S}(t)dt \right| \\ &\le | \hat{S}(x)S(x) |^{-1} \Big(S(x)\int_x^{\tau_F} | \hat{S}(t)-S(t) | dt \\ &\quad + | \hat{S}(x) - S(x) | \int_x^{\tau_F} S(t)\,dt + S(x)\int_{X^*}^{\tau_F} \hat{S}(t)\,dt \Big) . \end{aligned}$$

By combining the consistency result of $\hat{S}$ with partial integration, the first and second terms of the right-hand side of the inequality converge to zero in probability. On the other hand, the main part of third term is rewritten as

$$\sqrt{n}\int_{X^*}^{\tau_F} \hat{S}(t)\,dt = \sqrt{n}\int_{X^*}^{\tau_F} \{\hat{S}(t) - S(t)\}\,dt - \sqrt{n}\int_{X^*}^{\tau_F} S(t)\,dt .$$

From Lemma 2.1 and the above assumption of this theorem, the third term converge to zero in probability. Thus the result follows. □

Theorem 2.2. Suppose $\sqrt{n}\int_{X^*}^{\tau_F} S(x)dx \xrightarrow{p} 0$. Then as $n \to \infty$,

$$\sqrt{n}\{\hat{e} - e\} \xrightarrow{d} W,$$

where W is a mean zero Gaussion process and is given by

$$W(x) = \{S(x)\}^{-2}\left[S(x)\int_t^{\tau_F} Z(t)dt - Z(x)\int_t^{\tau_F} S(t)dt \right].$$

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

$$\begin{aligned}\sqrt{n}\{\hat{e}(x) - e(x)\} &= \sqrt{n}\left(\frac{1}{\hat{S}(x)}\int_x^{X^*}\hat{S}(t)dt - \frac{1}{S(x)}\int_x^{\tau_F} S(t)dt\right) \\ &= \{\hat{S}(x)S(x)\}^{-1}\left(S(x)\int_x^{\tau_F}\sqrt{n}\{\hat{S}(t) - S(t)\}dt\right. \\ &\quad \left. - \sqrt{n}\{\hat{S}(x) - S(x)\}\int_x^{\tau_F} S(t)dt - S(x)\sqrt{n}\int_{X^*}^{\tau_F}\hat{S}(t)dt\right).\end{aligned}$$

Now let $D[0, \tau_F)$ be the space of functions on the interval $[0, \tau_F)$ that are right continuous and have left-hand limits. Let **d** be the Skorohod metric on $D[0, \tau_F)$, and let us define a map $H: D[0, \tau_F) \to D[0, \tau_F)$ by having

$$H(Z)(x) = S(x)\int_x^{\tau_F} Z(t)dt - Z(x)\int_x^{\tau_F} S(t)dt$$

for $Z \in D[0, \tau_F)$, where the limiting distribution Z is defined in (2). Then H is a continuous map with respect to **d**. Thus by the Lemma 2.2, continuity theorem in Billingsley(1968) and the above assumption of this theorem, the result follows. □

Remark. The covariance function of the limit distribution $W(\cdot)$ defined in Theorem 2.2 is given by, for $0 \le s \le t < \tau_F$,

$$\begin{aligned}Cov\{W(s), W(t)\} = \{S(s)S(t)\}^{-2}\Big[& S(s)S(t)E\left(\int_s^{\tau_F}\int_t^{\tau_F} Z(u)Z(v)du\,dv\right) \\ & + E\{Z(s)Z(t)\}\int_s^{\tau_F} S(v)dv\int_t^{\tau_F} S(u)du \\ & - S(s)\int_t^{\tau_F} S(u)du\,E\left(Z(t)\int_s^{\tau_F} Z(v)dv\right) \\ & - S(t)\int_s^{\tau_F} S(v)dv\,E\left(Z(s)\int_t^{\tau_F} Z(u)du\right)\Big].\end{aligned}$$

3. Simulation Studies

In this section, a Monte Carlo simulation studies were carried out to compare the performances of the proposed estimator with the Kaplan-Meier type estimator and the Nelson-Aalen type estimator in terms of bias and estimated mean squared error(MSE).

The simulation scheme is designed with various sample of size ($n = 10, 20$ and 30) and the lifetime distribution (increasing, decreasing and constant failure rate). In each simulation, failure times with weibull distributions were generated. These values were then subject to be censored to the right by independent and exponentially distributed random variate with hazard rate of 0.067, 0.429 and 0.866. Here the values of hazard rates were calculated to make censoring rate to be 10%, 30% and 50%, respectively. This simulation procedure is repeated 500 times in order to get estimates of bias and MSE of the three type estimators. To construct the smoothing estimator of MRL function, we use the Epanechinikov kernel

$$k(x) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}x^2\right) & \text{if } |x| < \sqrt{5}, \\ 0 & \text{otherwise} . \end{cases}$$

Since the optimum choice of the bandwidth depends on the unknown density and its derivatives, the bandwidth is optimally selected at each time point, with which the MSE of the smoothing estimator is minimized.

Simulation results are tabulated in Tables 1-3. In tables, the values of estimates, biases and MSE's of the three type estimators are given at time points corresponding to quantiles of $0.1, 0.3, 0.5, 0.7, 0.9$.

In our simulation studies, we may see the following results : (1) In general, the estimators look like to be under-estimated because of truncation beyond the largest observed value X^* in calculating of the MRL function estimator. (2) In each quantile point, the MSE's of proposed estimator are decreased as the sample size increases and censoring rate decreases. In particular, the proposed estimator gets much larger MSE's at upper tail points. (3) For almost all cases, the MSE's of the proposed estimator are smaller than those of the other estimators. Hence the smoothing estimator is slightly better than the Kaplan-Meier type estimator and the Nelson-Aalen type estimator in terms of MSE.

Table 1. Biases and MSE's of $\hat{e}^{KM}$, $\hat{e}^{NA}$ and $\hat{e}$ under decreasing failure rate model

(F : Weib (1.0, 0.5), G : Exp (0.067), Censoring rate : 10%)

pt	TYPE	TRUE	N = 10 EST	BIAS	MSE	N = 20 EST	BIAS	MSE	N = 30 EST	BIAS	MSE
	KM		1.883	-.320	1.4904	1.955	-.256	1.0735	2.016	-.195	.7149
0.1	NA	2.211	2.173	-.038	1.8957	2.188	-.023	1.3166	2.215	.004	.8350
	SM		1.859	-.352	1.4434	1.924	-.287	1.0438	1.988	-.222	.7126
	KM		2.329	-.384	2.3937	2.387	-.326	1.7198	2.485	-.228	1.1943
0.3	NA	2.713	2.662	-.051	3.0074	2.666	-.047	2.0800	2.728	.015	1.4020
	SM		2.484	-.229	1.9604	2.576	-.137	1.4006	2.678	-.035	.9827
	KM		2.809	-.577	4.0611	2.922	-.464	2.9417	3.031	-.355	2.0452
0.5	NA	3.386	3.177	-.210	4.9597	3.262	-.125	3.4750	3.336	-.050	2.3459
	SM		3.192	-.194	2.6648	3.099	-.287	1.7779	3.245	-.142	1.1795
	KM		3.314	-1.094	9.8606	3.606	-.802	6.9005	3.840	-.568	5.0466
0.7	NA	4.408	3.618	-.790	10.9162	4.007	-.401	7.7613	4.234	-.174	5.6280
	SM		3.219	-1.118	5.6213	3.629	-.779	3.2038	3.966	-.442	1.7881
	KM		2.149	-4.456	35.0769	3.307	-3.298	28.3389	4.194	-2.412	23.5819
0.9	NA	6.605	2.201	-4.404	35.0576	3.469	-3.135	28.6828	4.450	-2.155	23.9420
	SM		1.758	-4.847	29.9938	2.831	-3.773	22.4474	3.709	-2.896	15.4063

Table 2. Biases and MSE's of $\hat{e}^{KM}$, $\hat{e}^{NA}$ and $\hat{e}$ under constant failure rate model

(F : Weib (1.0, 1.0), G : Exp (0.429), Censoring rate : 30%)

pt	TYPE	TRUE	N = 10 EST	BIAS	MSE	N = 20 EST	BIAS	MSE	N = 30 EST	BIAS	MSE
	KM		.918	-.082	.1364	.935	-.065	.0824	.966	-.034	.0508
0.1	NA	1.000	.992	-.008	.1489	.994	-.007	.0909	1.017	.017	.0566
	SM		1.005	.005	.1210	1.002	.001	.0746	1.002	.002	.0490
	KM		.912	-.088	.1960	.913	-.084	.1132	.967	-.034	.0811
0.3	NA	1.000	.987	-.013	.2116	.982	-.018	.1247	1.027	.027	.0918
	SM		.974	-.026	.1383	1.007	.007	.0840	1.027	.027	.0585
	KM		.849	-.151	.2942	.878	-.122	.1692	.941	-.059	.1192
0.5	NA	1.000	.917	-.083	.3104	.949	-.051	.1834	1.010	.010	.1308
	SM		.866	-.134	.1884	.961	-.039	.1058	.980	-.020	.0708
	KM		.682	-.318	.4678	.782	-.218	.3069	.915	-.085	.2408
0.7	NA	1.000	.714	-.286	.4807	.842	-.158	.3227	.986	-.014	.2570
	SM		.631	-.368	.3734	.830	-.170	.2145	.870	-.130	.1170
	KM		.203	-.797	.8482	.339	-.661	.7574	.526	-.474	.6361
0.9	NA	1.000	.206	-.764	.8495	.348	-.652	.7611	.538	-.462	.6354
	SM		.172	-.828	.8145	.286	-.714	.6920	.435	-.585	.5188

Table 3. Biases and MSE's of $\hat{e}^{KM}$, $\hat{e}^{NA}$ and $\hat{e}$ under increasing failure rate model

(F : Weib (1.0, 2.0), G : Exp (0.866), Censoring rate : 50%)

			N = 10			N = 20			N = 30		
pt	TYPE	TRUE	EST	BIAS	MSE	EST	BIAS	MSE	EST	BIAS	MSE
	KM		.598	-.038	.0391	.615	-.021	.0213	.625	-.012	.0127
0.1	NA	.636	.634	-.002	.0401	.644	.008	.0227	.651	.015	.0139
	SM		.640	.003	.0307	.647	.010	.0174	.646	.009	.0122
	KM		.473	-.032	.0498	.477	-.027	.0240	.494	-.010	.0172
0.3	NA	.504	.502	-.002	.0515	.507	.003	.0260	.523	.019	.0192
	SM		.498	-.015	.0336	.517	.012	.0185	.524	.020	.0134
	KM		.352	-.072	.0608	.382	-.042	.0325	.405	-.018	.0215
0.5	NA	.424	.372	-.052	.0621	.409	-.014	.0346	.436	.012	.0237
	SM		.341	-.082	.0444	.391	-.032	.0231	.439	.015	.0155
	KM		.215	-.141	.0702	.271	-.086	.0498	.322	-.034	.0357
0.7	NA	.357	.228	-.135	.0712	.287	-.070	.0514	.345	-.011	.0376
	SM		.194	-.163	.0613	.258	-.098	.0380	.325	-.032	.0235
	KM		.049	-.233	.0708	.086	-.197	.0649	.131	-.152	.0568
0.9	NA	.282	.050	-.233	.0710	.088	-.195	.0652	.134	-.149	.0570
	SM		.043	-.240	.0688	.072	-.210	.0602	.118	-.165	.0490

4. An Illustration

As an example, let us consider the well-known acute myelogenous leukemia (AML) data of Embury et al. (1977), which consist of length of remission, in weeks, of maintained group and nonmaintained group. The first group received maintenance chemotherapy; the second group did not. In this example, we use only nonmaintained group data, which are censored 1 of the 12 observation. 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 Kaplan-Meier estimator and Ghorai and Susarla's kernel estimator with $h_n = 3$ of the survival function. In this Figure, the kernel estimates are well smoothed, and are larger than the Kaplan-Meier estimates at each time points. Figure 2 presents the three estimators, the Kaplna-Meier type estimator, the Nelson-Aalen type estimator, and the smoothing estimator, of MRL function. From this we may see that the smoothing estimates are slightly larger than the other estimates for all time points.

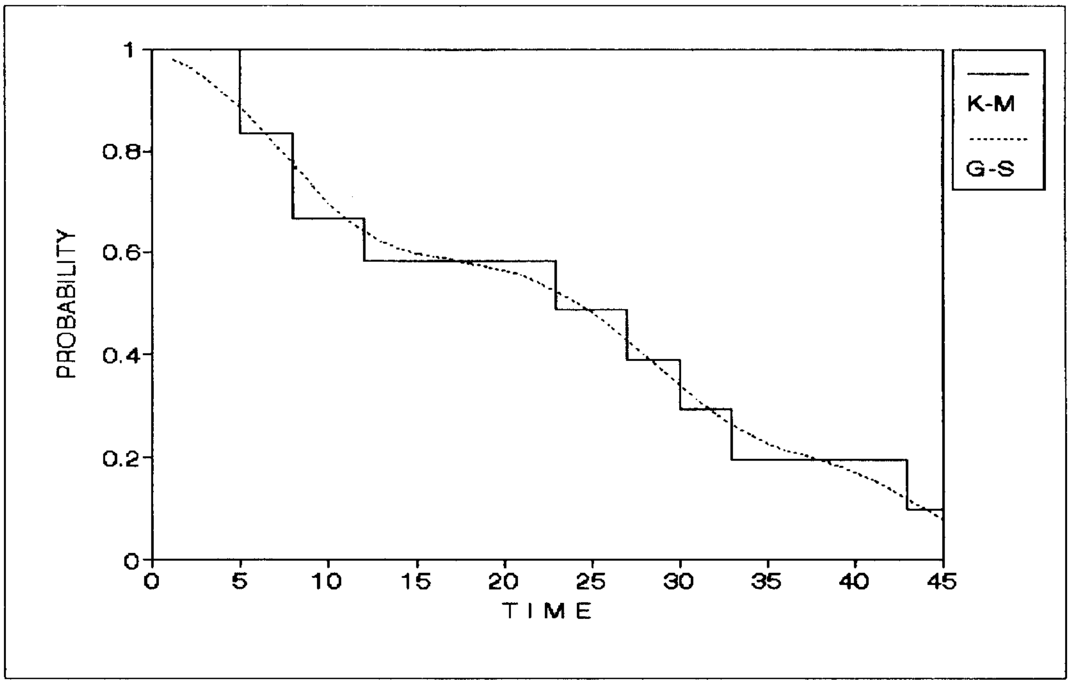

Figure 1. Estimates of $\hat{S}^{\text{KM}}$ and $\hat{S}$ for AML data

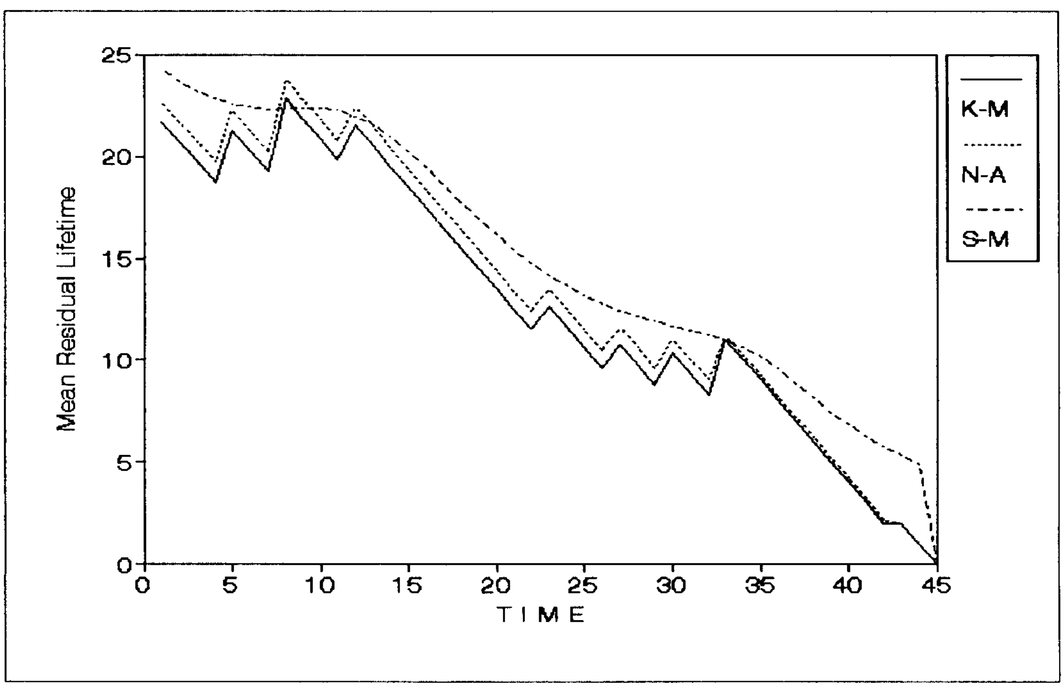

Figure 2. Estimates of $\hat{e}^{\text{KM}}$, $\hat{e}^{\text{NA}}$ and $\hat{e}$ for AML data

References

[1] Aalen, O. (1978). Nonparametric inference for a family of counting processes, *Annals of Statistics*, Vol. 6, 701-726.

[2] Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley & and Sons, New York.

[3] 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*, Vol. 126, 267-272.

[4] Ghorai, J.K. and Rejtö, L. (1987). Estimation of mean residual life with censored data under proportional hazard model, *Communication in Statistics, Theory of Methods*, Vol. 16(7), 2097-2114.

[5] Ghorai, J.K. and Susarla, V. (1990). Kernel estimation of a smooth distribution function based on censored data, *Metrika*, Vol. 37, 71-86.

[6] Hall, W.J. and Wellner, J.A. (1981). Mean residual life, *In Statistics and Related Topics*, Ed, M. Csörgö, D.A. Dawson, J.N.K. Rao and Md.E. Saleh, 169-184, New York, North-Holland.

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

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

[9] Nelson, W.B. (1972). Theory and applications of hazards plotting for censored failure data, *Technometrics*, Vol. 14, 945-996.

[10] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York.

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

[12] Yang, G.L. (1978). Estimation of Biometric function, *Annals of Statistics*, Vol. 6, 112-116.