BIOMETRICS 41, 253–259
March 1985

A Comparison of Several Methods of Estimating the Survival Function when There Is Extreme Right Censoring

M. L. Moeschberger[1] and John P. Klein[2]

Departments of Preventive Medicine[1] and Statistics[2], The Ohio State University, 320 West 10th Avenue, Columbus, Ohio 43210, U.S.A.

SUMMARY

When there is extreme censoring on the right, the Kaplan–Meier product-limit estimator is known to be a biased estimator of the survival function. Several modifications of the Kaplan–Meier estimator are examined and compared with respect to bias and mean squared error.

1. Introduction

In human and animal survival studies, as well as in life-testing experiments in the physical sciences, one method of estimating the underlying survival distribution (or the reliability of a piece of equipment) which has received widespread attention is the Kaplan–Meier product-limit estimator (Kaplan and Meier, 1958).

For the situation in which the longest time an individual is in a study (or on test) is not a failure time, but rather a censored observation, it is well known that there are many complex problems associated with any statistical analysis (Lagakos, 1979). In particular, the Kaplan–Meier product-limit estimator is biased on the low side (Gross and Clark, 1975). In the case of many censored observations larger than the largest observed failure time, this bias tends to be worse. Estimated mean survival time and selected percentiles, as well as other quantities dependent on knowledge of the tail of the survival function, will also exhibit such biases.

A practical situation which motivates this study is a large-scale animal experiment conducted at the National Center for Toxicological Research (NCTR), in which mice were fed a particular dose of a carcinogen. The goal of the experiment was to assess the effects of the carcinogen on survival and on age-specific tumor incidence. Toward this end, mice were randomly divided into three groups and followed until death or until a prespecified group censoring time (280, 420, or 560 days) was reached, at which time all those still alive in a given group were sacrificed. Often there were many surviving mice in all three groups at the sacrifice times.

In general, we consider an experiment in which n individuals are under study and censoring is permitted. Let $t_{(1)}, \ldots, t_{(m)}$ denote the m ordered failure times of those m individuals whose failure times are actually observed ($t_{(1)} \leq \cdots \leq t_{(m)}$). The remaining $n - m$ individuals have been censored at various points in time. It will be useful to introduce the notation S_j to denote the number of survivors just prior to time $t_{(j)}$; that is, S_j is the number of individuals still under observation at time $t_{(j)}$, including the one that died at $t_{(j)}$. Then the Kaplan–Meier product-limit estimator (assuming no ties among the $t_{(j)}$) of

Key words: Adjusted Kaplan–Meier survival estimation; Bias of survival function; Life-testing; Right censoring; Survival analysis.

the underlying survival function, $\bar{P}(t) = \Pr(T > t)$, is

$$\hat{\bar{P}}(t) = \begin{cases} 1 & \text{for} \quad t < t_{(1)} \\ \prod_{\ell=1}^{j} (S_\ell - 1)/S_\ell & \text{for} \quad t_{(j)} \leqslant t < t_{(j+1)} \\ 0 & \text{for} \quad t \geqslant t_{(m+1)} \end{cases} \tag{1}$$

for $j = 1, \ldots, m$, where $t_{(m+1)} = t_c$ if the longest time an individual is on study is a censoring time or $t_{(m+1)} = \infty$ if the longest time an individual is on study is a death.

This paper first proposes, in §2, some methods of "completing" the Kaplan–Meier estimator of the survival function by (i) replacing those censored observations that are larger than the last observed failure time by their expected order statistics; (ii) using a Weibull distribution to estimate the tail probability $\bar{P}(t)$, for $t > t_c$; and (iii) employing a method suggested by Brown, Hollander, and Korwar (BHK) (1974). The second purpose is to demonstrate the magnitude of the bias and mean squared error (MSE) of the Kaplan–Meier estimator and to compare all methods of "completing" $\hat{\bar{P}}(t)$ in the context of the aforementioned mouse study, utilizing simulated lifetimes from exponential, Weibull, lognormal, and bathtub-shaped hazard function distributions. These results are presented in §3.

2. Completion of Kaplan–Meier Product-Limit Estimator

2.1 *Expected Order Statistics*

One method of attempting to "complete" $\bar{P}(t)$, $t > t_c$, would be to "estimate" the failure times for those censored observations that are larger than the longest observed lifetime. Let n_c be the number of censored observations larger than $t_{(m)}$. A theorem regarding the conditional distributions of order statistics states that for a random sample of size n from a continuous parent, the conditional distribution of $T_{(u)}$, given $T_{(n-n_c)} = t_{(n-n_c)}$, $u > n - n_c$, is just the distribution of the $(u - n + n_c)$th order statistic in a sample of size n_c drawn from the parent distribution truncated on the left at $t = t_{(n-n_c)}$ (see David, 1981, p. 20).

For computational purposes, take t_c as an estimate of the $(n - n_c)$th order statistic. Then find the expected value of the n_c order statistics from the parent distribution truncated on the left at t_c. Since the Weibull distribution with survival function $\bar{P}(t) = \exp(-t^k/\theta)$ has been widely accepted as providing a satisfactory fit for lifetime data, it seems reasonable to employ the results of Weibull distribution theory to complete $\bar{P}(t)$, $t > t_c$. (It should be noted that any distribution which is reasonable for the specific situation may be used.) The expected values of Weibull order statistics up to sample size 40 for location parameter equal to 1 and shape parameter equal to .5 (0.5)4(1)8 may be found in Harter (1969). For larger sample sizes, he states a recurrence relation which may be used.

To compute expected values of the n_c order statistics in question, values for k and θ must be chosen. One approach is to use the maximum likelihood estimators, $\hat{k}$ and $\hat{\theta}$, computed by using all observations to estimate k and θ. A second approach, due to White (1969), uses least squares estimates of k and θ obtained by fitting the model

$$\ln(t_{(j)}) = (1/k) \ln \theta + (1/k) \ln[H(t_{(j)})] \tag{2}$$

to the $t_{(j)}$'s, where $H(t_{(j)})$ is the estimated cumulative hazard rate at $t_{(j)}$ obtained from the Kaplan–Meier estimator. In our Monte Carlo study, we found the maximum likelihood estimators performed better than the least squares estimators in all cases. Consequently, the method of least squares will be dropped from future discussion in this paper.

The survival function for a Weibull random variable, truncated on the left at t_c, is

$$\bar{P}_T(t) = \exp[-(t^k - t_c^k)/\theta], \qquad t > t_c. \tag{3}$$

So, by the theorem on order statistics stated at the beginning of this section, the conditional distribution of $T_{(u)}$, given $T_{(n-n_c)} = t_{(n-n_c)}$ $(u = n - n_c + 1, \ldots, n)$ will be approximated by the $(u - n + n_c)$th order statistic in a sample of n_c drawn from (3). For simplicity, let $j = u - n + n_c$, so that $j = 1, \ldots, n_c$. Now the expected value of the jth order statistic from (3) is

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1} \int_{t_c}^{\infty} t[P_T(t)]^{j-1}[\bar{P}_T(t)]^{n_c-j+1}(kt^{k-1}/\theta)\, dt$$

$$= n_c\binom{n_c - 1}{j - 1} \int_{0}^{\infty} (y^k + t_c^k)^{1/k}[P(y)]^{j-1}[\bar{P}(y)]^{n_c-j+1}(ky^{k-1}/\theta)\, dy \tag{4}$$

where $\bar{P}(y) = \exp(-y^k/\theta)$, $y = (t^k - t_c^k)^{1/k} \geqslant 0$ and $T_{j:n_c}$ is the jth order statistic in a sample of size n_c. Equation (4) can also be written as

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1} \int_{0}^{\infty} (\theta z^k + t_c^k)^{1/k}[P(z)]^{j-1}[\bar{P}(z)]^{n_c-j+1}kz^{k-1}\, dz \tag{5}$$

where $\bar{P}(z) = \exp(-z^k)$, $z = (y/\theta)^{1/k} \geqslant 0$. Now $\mathrm{E}(T_{j:n_c})$ may be crudely estimated by

$$\{\hat{\theta}[\mathrm{E}(Z_{j:n_c})]^{\hat{k}} + t_c^{\hat{k}}\}^{1/\hat{k}} \tag{6}$$

where $\mathrm{E}(Z_{j:n_c})$ is the expected value of the jth order statistic from a sample of size n_c determined from Harter's (1969) tables or recurrence relation, and $\hat{\theta}$ and $\hat{k}$ are maximum likelihood estimators of θ and k, respectively.

These n_c estimated expected order statistics may then be treated as "observed" lifetimes in adjusting (or "completing") the estimated survival function computed in (1). The area under the estimated survival function up to t_c remains unchanged. The area under the extended estimated survival function based on the n_c estimated expected order statistics is then added to the initial area to obtain a more precise estimate of $\bar{P}(t)$ [estimated order statistic (EOS) extension].

2.2 *Weibull Maximum Likelihood Techniques*

A straightforward approach to completing $\hat{\bar{P}}(t)$ is to set

$$\hat{\bar{P}}(t) = \exp(-t^k/\theta) \qquad \text{for} \quad t > t_c. \tag{7}$$

Estimates of k and θ based on all observations can be obtained by either the maximum likelihood (WTAIL) or the least squares method. However, our study found the completion using maximum likelihood estimators was always better in terms of bias and mean squared error.

One suggestion for ostensibly improving this estimator would be to "tie" the estimated tail to the product-limit estimator at t_c. Two methods were attempted to accomplish this goal. First, the likelihood was maximized with respect to k and θ subject to the constraint that $\exp(-t_c^k/\theta) = \hat{\bar{P}}(t_c)$. This method will be referred to as the restricted MLE tail probability estimate (RWTAIL extension). Second, a scale-shift was performed on the tail probability in (7) to tie it to the product-limit estimator. This method led to higher biases and mean squared errors of the survival function and will be dropped from further discussion in this paper.

2.3 *BHK-Type Methods*

The Brown–Hollander–Korwar completion of the product-limit estimator sets

$$\hat{\bar{P}}(t) = \exp(-t/\theta^*) \qquad \text{for} \quad t > t_c \tag{8}$$

where θ^* satisfies $\hat{\bar{P}}(t_c) = \exp(-t_c/\theta^*)$. In the BHK spirit we tried to complete $\bar{P}(t)$ by a Weibull function which used estimates of k and θ, k^* and θ^*, that satisfied the following two relations:

$$\hat{\bar{P}}(t_{(m)}) = \exp(-t_{(m)}^{k^*}/\theta^*)$$

and

$$\hat{\bar{P}}(t_{(m-1)}) = \exp(-t_{(m-1)}^{k^*}/\theta^*).$$

The latter method also led to consistently poor performance and the results will not be presented.

Table 1
Bias/100 (and MSE/100²) for estimating mean survival time for various methods of completion

Distribution	μ	Mean % censored at 560 days	K–M	BHK extension	Estimated order statistic extension	Weibull WTAIL extension	Restricted Weibull RWTAIL extension
Weibull	400	18.7	−2.000[w]	−1.462	−.101[b]	.131	.206
			(4.034)[w]	(2.271)	(1.172)	(1.160)[b]	(1.543)
$k = .5$	500	22.3	−2.802[w]	−2.078	−.176[b]	.208	.299
			(7.886)[w]	(4.498)	(1.922)[b]	(2.344)	(3.292)
	600	25.5	−3.625[w]	−2.704	−.187[b]	.344	.479
			(13.179)[w]	(7.522)	(3.025)[b]	(4.275)	(6.031)
	400	24.6	−.991[w]	−.047	−.046	.016[b]	.0379
			(1.011)[w]	(.215)[b]	(.257)	(.275)	(.343)
$k = 1$	500	32.6	−1.632[w]	−.049	−.047[b]	.073	.116
			(2.696)[w]	(.416)[b]	(.535)	(.508)	(.705)
	600	39.3	−2.359[w]	.022[b]	.034	.140	.214
			(5.592)[w]	(.596)[b]	(.987)	(1.023)	(1.353)
	400	7.5	−.036	.136[w]	−.005	.003[b]	.004
			(.012)[b]	(.053)[w]	(.013)	(.014)	(.014)
$k = 4$	500	34.6	−.314	1.507[w]	−.020	.014[b]	.019
			(.109)	(2.830)[w]	(.036)[b]	(.041)	(.044)
	600	59.9	−.903	5.982[w]	.144	.028[b]	.039
			(.822)	(41.430)[w]	(4.168)	(.147)[b]	(.157)
Lognormal	400	20.6	−.868[w]	−.178[b]	−.544	−.586	−.412
			(.777)[w]	(.179)[b]	(.363)	(.403)	(.267)
$k = 1$	500	29.0	−1.427[w]	−.150[b]	−.865	−.918	−.696
			(2.060)[w]	(.323)[b]	(.855)	(.938)	(.644)
	600	36.9	−2.079[w]	−.022[b]	−1.234	−1.281	−1.038
			(4.345)[w]	(.571)[b]	(1.679)	(1.800)	(1.301)
	400	8.6	−.070	.129[w]	−.047	−.053	−.027[b]
			(.014)[b]	(.056)[w]	(.014)[b]	(.014)[b]	(.014)[b]
$k = 4$	500	29.1	−.330	1.033[w]	−.170	.181	−.135[b]
			(.118)	(1.459)[w]	(.051)	(.055)	(.043)[b]
	600	54.5	−.853	4.430[w]	−.391	−.392	−.356[b]
			(.734)	(23.159)[w]	(.199)	(.199)	(.177)[b]
Bathtub	400	18.6	−1.069	−.185	−.170	1.125[w]	.063[b]
			(1.175)	(.234)[b]	(.260)	(1.745)[w]	(.361)
$p = .1$	500	26.1	−1.722[w]	−.259	−.202	1.523	.046[b]
			(2.996)	(.427)[b]	(.560)	(3.230)[w]	(.608)
	600	32.6	−2.452[w]	−.362	−.310	1.761	.047[b]
			(6.043)[w]	(.727)[b]	(.982)	(4.490)	(1.254)
	400	8.1	−1.786[w]	−1.543	−1.547	−.936	.343[b]
			(3.218)[w]	(2.463)	(2.476)	(1.081)	(.544)[b]
$p = .4$	500	13.3	−2.370[w]	−1.826	−1.814	−.825	.585[b]
			(5.649)[w]	(3.472)	(3.446)	(1.031)[b]	(1.303)
	600	18.7	−3.072[w]	−2.191	−2.175	−.875	.841[b]
			(9.466)[w]	(5.013)	(4.983)	(1.285)[b]	(2.792)

[b] Best estimation method.
[w] Worst estimation method.

3. A Comparison of the Various Methods

A simulation study of data such as that collected at NCTR was performed. Three groups of 48 lifetimes were simulated with all testing stopping at 280, 420, and 560 days, respectively, for the three groups. Distributions with mean survival times of 400, 500, and 600 days were used. The generated lifetimes greater than or equal to the sacrifice time for each particular group were considered as censored. The remaining set of observed lifetimes, along with the number censored at the three sacrifice times, constituted a single sample. For each of the distributions studied, 1000 such samples were generated. Weibull distributions with shape parameters .5, decreasing failure rate, 1, constant failure rate, and 4,

Table 2

Bias/100 (*and MSE*/100^2) *for estimating* 90*th percentile for various methods of completion*

Distribution	μ	K–M	BHK extension	Estimated order statistic extension	Weibull WTAIL extension	Restricted Weibull RWTAIL extension
Weibull	400	−5.017[w]	−2.858	1.691	.234[b]	.458
		(25.185)[w]	(9.358)	(16.424)	(7.524)[b]	(10.812)
$k = .5$	500	−7.655[w]	−4.620	1.897	.418[b]	.642
		(58.604)[w]	(22.711)	(24.276)	(14.319)[b]	(21.442)
	600	−10.306[w]	−6.390	2.213	.734[b]	1.064
		(106.21)[w]	(42.449)	(36.895)	(25.419)[b]	(37.911)
	400	−3.610[w]	.064[b]	.248	.084	.067
		(13.035)[w]	(1.892)[b]	(2.423)	(1.980)	(2.945)
$k = 1$	500	−5.913[w]	.096[b]	.289	.121	.306
		(34.963)[w]	(2.995)[b]	(4.681)	(4.361)	(5.903)
	600	−8.216[w]	.244[b]	.610	.418	.550
		(67.459)[w]	(4.198)[b]	(9.247)	(8.331)	(10.792)
	400	−.045	.098[w]	−.007[b]	−.037	−.011
		(.038)[b]	(.236)[w]	(.060)	(.047)	(.063)
$k = 4$	500	−1.195	5.324[w]	−.031	−.026	.024[b]
		(1.429)	(33.091)[w]	(.146)	(.141)[b]	(.177)
	600	−2.554	17.913[w]	.120	.090	.068[b]
		(6.524)	(355.02)[w]	(.794)	(.676)	(.641)[b]
Lognormal	400	−2.628[w]	−.044[b]	−1.263	−1.758	−.967
		(6.908)[w]	(1.526)[b]	(1.979)	(3.407)	(1.673)
$k = 1$	500	−4.680[w]	.213[b]	−2.354	−2.718	−1.908
		(21.902)[w]	(2.708)[b]	(6.153)	(7.909)	(4.751)
	600	−6.736[w]	.759[b]	−3.507	−3.766	−2.980
		(45.373)[w]	(4.764)[b]	(13.123)	(14.981)	(10.257)
	400	−.085	.161	−.038	−.162[w]	−.024[b]
		(.060)[b]	(.409)[w]	(.081)	(.065)	(.093)
$k = 4$	500	−1.251	3.722[w]	−.584	−.657	−.484[b]
		(1.566)	(17.654)[w]	(.403)	(.495)	(.318)[b]
	600	−2.621	13.695[w]	−1.214	−1.236	−1.158[b]
		(6.872)	(210.30)[w]	(1.616)	(1.662)	(1.498)[b]
Bathtub	400	−3.629[w]	−.177	.053[b]	−.104	.105
		(13.167)[w]	(1.717)[b]	(2.052)	(2.058)	(3.190)
$p = .1$	500	−6.068[w]	−.457	−.071	−.208	.004[b]
		(36.826)[w]	(2.955)[b]	(4.702)	(3.619)	(5.245)
	600	−7.997[w]	−.318	.043	−.244	−.014[b]
		(63.954)[w]	(4.330)[b]	(7.786)	(7.608)	(9.923)
	400	−.347	.143[b]	.276	1.154[w]	.981
		(.273)[b]	(.844)	(1.078)	(3.877)	(4.747)[w]
$p = .4$	500	−1.425	.521[b]	.764	1.699	1.718[w]
		(2.035)	(1.540)[b]	(2.067)	(8.574)	(10.714)[w]
	600	−3.554[w]	−.137	.132[b]	2.304	2.450
		(12.628)[w]	(1.804)[b]	(2.352)	(17.530)	(22.456)

[b] Best estimation method.
[w] Worst estimation method.

increasing failure rate, were used. Lognormal distributions, failure rate changes from increasing to decreasing, with first two moments comparable to the above Weibull distributions with $k = 1$ and $k = 4$, were also used. Finally, a bathtub hazard model of Glaser (1980), failure rate changes from decreasing to increasing, was used. This distribution is a mixture of an exponential of parameter λ with probability $1 - p$ and a gamma with parameter λ and index 3 with probability p. Mixing parameters of $p = .1$ and $p = .4$ were used.

The bias and MSE for the estimation of the tail probabilities, i.e., the completed portion of the product-limit estimator, were calculated for each hypothesized distribution and for each competing method of completion. Since these results were extremely similar to those found in estimating mean survival time, $\hat{\mu} = \int_0^\infty \hat{\bar{P}}(t)\, dt$, we show only the bias and MSE of each competing estimator of μ in Table 1. This also allows us to demonstrate the magnitude of the bias and MSE of the product-limit estimator of μ. The bias and MSE for estimating the 90th percentile are also presented for the various estimation methods in Table 2. As one would expect, the Kaplan–Meier (K–M) estimator performs considerably more poorly than the other estimation schemes. The BHK extension does very well if the underlying distribution is exponential or lognormal with first two moments compatible with the exponential. BHK does reasonably well for the bathtub-shaped hazard model, but it performs very poorly for the Weibull with increasing failure rate and for the lognormal with first two moments compatible with the Weibull.

The remaining three extensions (EOS, WTAIL, and RWTAIL) appear to be somewhat comparable. Each of them is best under certain circumstances although many times the biases and MSEs are so close to one another that they are essentially equivalent. Only the EOS extension has the desirable property of never being worst. It usually is competitive with the method that is best. Ordering the extensions from the standpoint of simplicity, from simplest to most complex, we have BHK, WTAIL, RWTAIL, and EOS.

In summary, the Kaplan–Meier estimator should probably be extended in the presence of extreme right censoring. The choice of extension depends on one's knowledge of the distribution of lifetimes under consideration and the extent of computer facilities available. If the data follow an exponential-type distribution or if no computer facilities are present, the BHK method is the extension of choice due to its simplicity. If the data exhibit a nonconstant failure rate and computer facilities are available, then the RWTAIL or EOS extensions seem to be advisable.

Acknowledgement

This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-82-0307.

Résumé

On sait que l'estimateur de Kaplan–Meier est un estimateur biaisé de la fonction de survie quand le pourcentage d'observations censurées est très élevé. Plusieurs modifications de l'estimateur de Kaplan–Meier sont examinées et comparées du point de vue de leurs biais et écarts moyens quadratiques.

References

Brown, B. W., Jr., Hollander, M., and Korwar, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling (eds), 291–302. Philadelphia: Society for Industrial and Applied Mathematics.

David, H. A. (1981). *Order Statistics*. New York: Wiley.

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association* **75**, 667–672.

Gross, A. J. and Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

Harter, H. L. (1969). *Order Statistics and Their Use in Testing and Estimation. Vol.* 2. Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force.

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

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics* **35,** 139–156.

White, J. S. (1969). The moments of log-Weibull order statistics. *Technometrics* **11,** 373–386.

Received June 1980; *revised May* 1984.