BIOMETRICS 39, 1077–1081
December 1983

What Price Kaplan–Meier?

Rupert G. Miller, Jr

Division of Biostatistics, Stanford University Medical Center, Stanford, California 94305, U.S.A.

SUMMARY

The asymptotic efficiency of the Kaplan–Meier product-limit estimator, relative to the maximum likelihood estimator of a parametric survival function, is examined under a random-censoring model.

1. Introduction

The Kaplan–Meier (1958) product-limit estimator has become an important tool in the analysis of censored survival data. This is especially true in medical clinical trials in which patients are lost to follow-up and/or the data are analyzed while a number of patients are still surviving.

The product-limit estimator is attractive because it is easy to compute and understand. It has an asymptotic normal distribution with an estimated variance that is easily computed by Greenwood's formula. For the underlying probability structure, no assumptions are required other than the basic one of independence between the survival and censoring variables.

In fact, the product-limit estimator is so seductive that there is a danger of becoming mentally lazy and not considering parametric modelling. Is there a price to be paid for this easy living? This shorter communication examines the asymptotic efficiency of the Kaplan–Meier estimator compared to a parametric estimator of the survival function.

2. Asymptotic Efficiency

Let the survival times $T_1, \ldots, T_n$ be independently identically distributed according to the distribution function $F(t)$, which is assumed to be differentiable for $t > 0$. Similarly, let $C_1, \ldots, C_n$ be independently identically distributed according to $G(t)$, which also is assumed to be differentiable for $t > 0$. In addition, the survival times T_i and the censoring times C_i are assumed to be independent.

The observable random variables are $Y_i = \min\{T_i, C_i\}$ and $\delta_i = I(T_i \leq C_i)$, where $I(\ )$ is the indicator function. The value y_i is the observed length of survival, and δ_i indicates whether the survival time is uncensored or censored. Let $y_{(1)} < \cdots < y_{(n)}$ denote the ordered observed survival times, and let $\delta_{(1)}, \ldots, \delta_{(n)}$ be their corresponding (unordered) indicator values.

The Kaplan–Meier product-limit estimator of $S(t) = 1 - F(t)$ is defined by

$$\hat{S}(t) = \prod_{y_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}. \tag{1}$$

Full discussion of this estimator and its properties can be found in a number of sources (see, for example, Miller, 1981, Ch. 3). Provided that $F(t) < 1$ and $G(t) < 1$, $\hat{S}(t)$ is a consistent

Key words: Kaplan–Meier estimator; Product-limit estimator; Maximum likelihood estimator; Survival function; Efficiency; Random censoring.

estimator of $S(t)$ and is asymptotically normally distributed with variance

$$\mathrm{var}_{\mathrm{asy}}\{\hat{S}(t)\} = \frac{1}{n}\{1 - F(t)\}^2 \int_0^t \frac{f(u)du}{\{1 - F(u)\}^2\{1 - G(u)\}}, \tag{2}$$

where $f(t) = dF(t)/dt$.

If the survival distribution, F, is one of a parametric family of distributions, $\{F_\lambda, \lambda \in \Lambda\}$, then the likelihood function for the sample (y_i, δ_i), $i = 1, \ldots, n$, is

$$L(\lambda) = \prod_{i=1}^{n} \{f_\lambda(y_i)\}^{\delta_i}\{1 - F_\lambda(y_i)\}^{1-\delta_i}, \tag{3}$$

where $f_\lambda(t) = \partial F_\lambda(t)/\partial t$. The maximum likelihood estimator of λ is the value $\hat{\lambda} \in \Lambda$ that maximizes (3), and the associated estimator of $S_\lambda(t) = 1 - F_\lambda(t)$ is $S_{\hat{\lambda}}(t)$. Under standard regularity conditions (see, for example, Rao, 1973, §5f), the estimator $\hat{\lambda}$ is consistent and asymptotically normally distributed with variance $I^{-1}(\lambda)/n$, where

$$I(\lambda) = \int_0^\infty \left\{\frac{\partial \log f_\lambda(u)}{\partial\lambda}\right\}^2 f_\lambda(u)\{1 - G(u)\}\, du + \int_0^\infty \left\{\frac{\partial \log S_\lambda(u)}{\partial\lambda}\right\}^2 S_\lambda(u)g(u)du. \tag{4}$$

The associated estimator, $S_{\hat{\lambda}}(t)$, of the survival probability has an asymptotic normal distribution with variance

$$\mathrm{var}_{\mathrm{asy}}\{S_{\hat{\lambda}}(t)\} = \left(\frac{1}{n}\right) I^{-1}(\lambda) \left\{\frac{\partial S_\lambda(t)}{\partial\lambda}\right\}^2 \tag{5}$$

obtained by the delta method.

If the parametric family of distributions is characterized by two parameters $\{F_{\lambda,\alpha}, \lambda \in \Lambda, \alpha \in A\}$, then (5) is replaced by

$$\mathrm{var}_{\mathrm{asy}}\{S_{\hat{\lambda},\hat{\alpha}}(t)\} = \left(\frac{1}{n}\right)\boldsymbol{\partial}\mathbf{S}^T\mathbf{I}^{-1}(\lambda, \alpha)\boldsymbol{\partial}\mathbf{S}, \tag{6}$$

where $\boldsymbol{\partial}\mathbf{S}$ is the vector of partial derivatives of $S_{\lambda,\alpha}(t)$ with respect to λ and α, and $\mathbf{I}(\lambda, \alpha)$ is the Fisher information matrix defined appropriately for observations subject to censoring.

The asymptotic efficiency, $E(t; \lambda)$, of the Kaplan–Meier estimator relative to the maximum likelihood estimator of the survival probability $S_\lambda(t)$ is the ratio of (5) over (2). Similarly, for two parameters the efficiency, $E(t; \lambda, \alpha)$, is the ratio of (6) over (2).

In the case in which there is no censoring, i.e. $G(t) \equiv 0$, the product-limit estimator is $1 - F_n(t)$, where $F_n(t)$ is the ordinary sample distribution function. For this case, (2) reduces to a simple binomial variance, and the Fisher information is the usual expectation of squared or cross-product partial derivatives. Surprisingly, no efficiency comparison of the sample distribution function with the maximum likelihood estimators appears to have been reported in the literature.

There is no simple relationship between the cases with and without censoring: censoring can either increase or decrease the efficiency of the Kaplan–Meier estimator, depending on how the censoring distribution, G, interacts with the survival distribution, F.

3. Specific Cases

For an exponential survival distribution, $S_\lambda(t) = \exp(-\lambda t)$, with exponential censoring distribution $1 - G(t) = \exp(-\mu t)$, the reciprocal of $E(t; \lambda)$ is

$$E^{-1}(t; \lambda) = \frac{1}{(1 + \rho)^2(\lambda t)^2}[\exp\{(1 + \rho)\lambda t\} - 1], \tag{7}$$

Table 1
Asymptotic efficiencies of the K–M estimator relative to the MLE for the exponential distribution, with exponential censoring

CP	ρ	SF: / λt:	.50 / .69	.25 / 1.39	.10 / 2.30
.50	1.0		.64	.51	.21
.25	.33		.56	.64	.46
0	0		.48	.64	.59

$SF = \exp(-\lambda t)$; $CP = \rho/(1 + \rho)$.

where $\rho = \mu/\lambda$. The function (7) approaches $+\infty$ as $t \rightarrow 0$ or $t \rightarrow +\infty$, and it has a minimum value of 1.54 near $(1 + \rho)\lambda t = 1.59$. Thus, the Kaplan–Meier estimator has a peak efficiency of $1/1.54 = .65$ relative to the maximum likelihood estimator, and it has zero efficiency at the opposite ends of the time scale. Also, as the censoring becomes heavier, that is, as $\rho \rightarrow +\infty$, the Kaplan–Meier efficiency drops to zero.

Table 1 presents some asymptotic efficiencies at different surviving fractions $SF = \mathrm{pr}(T > t) = \exp(-\lambda t)$ for several censoring proportions $CP = \mathrm{pr}(C < T) = \rho/(1 + \rho)$. The case of no censoring arises when $\rho = 0$. Similar efficiencies are obtained for an exponential distribution with uniform censoring on $(0, c)$.

For a Weibull distribution, $S_\lambda(t) = \exp\{-(\lambda t)^\alpha\}$, with known index α subject to exponential censoring, the reciprocal of $E(t; \lambda)$ is given by

$$E^{-1}(t; \lambda) = \frac{\alpha}{(\lambda t)^{2\alpha}}(1 - CP)\int_0^{\lambda t} u^{\alpha-1}\exp(\rho u + u^\alpha)du, \tag{8}$$

where $\rho = \mu/\lambda$ and the censoring proportion is

$$CP = \rho\int_0^\infty \exp(-\rho u - u^\alpha)\, du. \tag{9}$$

Table 2 presents the asymptotic efficiencies of the Kaplan–Meier estimator for the same combinations of censoring proportions and surviving fractions as in Table 1 but when the Weibull index is $\alpha = 2$. The rows for no censoring are identical in the two tables. The reason for this is that without censoring the Weibull distribution is just an exponential distribution with a transformed time scale. Analogous expressions and similar efficiencies can be obtained for the Weibull distribution with uniform censoring and also for the scale parameter in the gamma distribution.

For the Weibull distribution with unknown scale and index parameters (λ, α) and

Table 2
Asymptotic efficiencies of the K–M estimator relative to the MLE for the Weibull distribution with known index $\alpha = 2$, with exponential censoring

CP	ρ	SF: / λt:	.50 / .83	.25 / 1.18	.10 / 1.52
.50	.87		.57	.58	.40
.25	.34		.52	.63	.52
0	0		.48	.64	.59

$SF = \exp(-(\lambda t)^2\}$; CP is given by (9).

Table 3
Asymptotic efficiencies of the K–M estimator relative to the MLE for the Weibull distribution with unknown index $\alpha = 2$, with exponential censoring

CP	ρ	SF: / λt:	.50 / .83	.25 / 1.18	.10 / 1.52
.50	.87		.60	.62	.56
.25	.34		.63	.63	.62
0	0		.66	.64	.65

$SF = \exp\{-(\lambda t)^{\alpha}\}$; CP is given by (9).

exponential censoring, the Fisher information matrix simplifies to

$$\mathbf{I}(\lambda, \alpha) = \int_0^{\infty} \alpha u^{\alpha-1}\exp(-\rho u - u^{\alpha})\mathbf{A}(u; \lambda, \alpha)\mathbf{A}^T(u; \lambda, \alpha)du, \tag{10}$$

where

$$\mathbf{A}^T(u; \lambda, \alpha) = \left[\frac{\alpha}{\lambda}, \frac{1}{\alpha}\{1 + \log(u^{\alpha})\}\right]. \tag{11}$$

Table 3 gives the asymptotic efficiencies, $E(t; \lambda, \alpha)$, for the same combinations of censoring proportions and surviving fractions as in Table 2 but when the Weibull distribution has an unknown index $\alpha = 2$. The censoring proportions are still given by (9). The efficiencies in Table 3 are higher than those in Table 2 because a second parameter now has to be estimated by the maximum likelihood method.

4. Discussion

In the specific cases considered in §3 the asymptotic efficiencies of the Kaplan–Meier estimator are low, especially for high censoring proportions or for surviving fractions that are close to one or zero. Low efficiencies for high surviving fractions, i.e. when t is near zero, are not a cause for worry because the variances of both the Kaplan–Meier estimator and the maximum likelihood estimator are small and the surviving fraction is accurately estimated. However, for low surviving fractions, i.e. when t is large, both variances are large and any drop in efficiency represents a real loss of accuracy. Parametric modelling should be considered as a means of increasing the precision in the estimation of small tail probabilities.

The reader should realize that the preceding comparisons are loaded against the Kaplan–Meier estimator. The efficiencies have been computed on the 'home turf' of the maximum likelihood estimator, that is, under the assumption that the parametric family of distributions has been correctly selected. But in practice where is the oracle that informs the statistician of the correct choice? With an incorrectly selected family, the maximum likelihood estimator is asymptotically biased. In the limit, as $n \to \infty$, the mean square error for the Kaplan–Meier estimator will tend to zero, unlike that for the maximum likelihood estimator.

For large samples, for example where $n > 100$, with only moderate censoring, for example $CP < .25$, it should be reasonably clear whether or not a particular family of distributions fits the data well. If it does, then parametric modelling produces better estimates. Still, in the main body of the data the product-limit estimator may be sufficiently accurate for practical purposes. If interest is focused on estimation for the extreme high quantiles, for example where $\mathrm{pr}(T > t) < .01$, then the Kaplan–Meier estimator is usually worthless, but very large samples, for example where $n > 500$, are required to check the validity of the parametric extrapolation.

For small- and moderate-sized samples the question is whether to stick with a nonparametric approach or to try parametric modelling. Studies are under way to investigate quantitatively the performance of the maximum likelihood estimator when the underlying model is incorrect. When completed, these studies should provide some guidelines on when it is beneficial to try parametric modelling and when it is not.

Acknowledgement

This research was supported by NIH Grant No. R01 GM21215.

Résumé

L'efficacité asymptotique de l'estimateur de Kaplan–Meier relatif à l'estimateur du maximum de vraisemblance d'une fonction paramétrique de survie est examinée sous un modèle de censure aléatoire.

References

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

Miller, R. G., Jr (1981). *Survival Analysis.* New York: Wiley.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

Received January 1982; *revised September* 1982