

1999

An unbiased estimator for the survival function of censored data

S.N. Chiu

Hong Kong Baptist University, snchiu@hkbu.edu.hk

This document is the authors' final version of the published article.

Link to published article: <http://dx.doi.org/10.1080/03610929908832419>

APA Citation

Chiu, S. (1999). An unbiased estimator for the survival function of censored data. *Communications in Statistics - Theory and Methods*, 28 (9), 2249-2260. <https://doi.org/10.1080/03610929908832419>

This Journal Article is brought to you for free and open access by HKBU Institutional Repository. It has been accepted for inclusion in HKBU Staff Publication by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.

AN UNBIASED ESTIMATOR FOR THE SURVIVAL FUNCTION OF CENSORED DATA

S. N. CHIU

*Department of Mathematics
Hong Kong Baptist University*

Kowloon Tong

Hong Kong

E-mail: snchiu@math.hkbu.edu.hk

Key Words: generalized reduced sample type estimator; Kaplan–Meier;
Nelson–Aalen; random censorship

ABSTRACT

A monotonic, pointwise unbiased and uniformly consistent estimator for the survival function of failure time under the random censorship model is proposed. This estimator is closely related to the Kaplan–Meier, the Nelson–Aalen, and the reduced sample estimator. Large sample properties of the new estimator are discussed.

1. INTRODUCTION

Suppose that individuals in a sample have independent and identically distributed failure times. For each individual there is a period of observation such that observation on the individual ceases at the end of the period (the censoring time) if a failure has not occurred by then. Under the random censorship model, the observation obtained from each subject consists of a censoring indicator and the smaller one of the failure time and the censoring time.

The Kaplan–Meier estimator (Kaplan and Meier, 1958) is usually used to estimate the survival function of the failure time under the random censorship model. The asymptotic optimality of the Kaplan–Meier estimator was first established by Wellner (1982). Gill (1989) has shown that the generalized likelihood approach can be used to obtain its efficiency. The Kaplan–Meier estimator has nonnegative bias, although the bias converges to zero at an exponential rate as $n \rightarrow \infty$ (see e.g. Andersen *et al.*, 1993, p. 257; Fleming and Harrington, 1991, p. 99).

The Nelson–Aalen estimator (Aalen, 1975, 1978; Nelson, 1969, 1972) of the cumulative hazard function of the failure time can be regarded as a maximum likelihood estimator (Johansen, 1978). Similar to the Kaplan–Meier estimator, the Nelson–Aalen estimator has a bias (negative) which converges to zero exponentially as $n \rightarrow \infty$ (see e.g. Andersen *et al.*, 1993, p. 179; Fleming and Harrington, 1991, pp. 92-93).

The Kaplan–Meier and the Nelson–Aalen estimator are closely related and have been studied intensively by using the counting processes and martingale methods. For details see the monographs by Andersen *et al.* (1993), Fleming and Harrington (1991) and Lai and Zheng (1993).

Suppose that all the censoring times are known. Such a situation arises in Type I and Type II censoring, as well as in some applications (see e.g. Cox and

Oakes, 1984, p.5). In such situations the reduced sample estimator (see e.g. Cox and Oakes, 1984, p.11) is a pointwise unbiased estimator of the survival function of the failure time. A disadvantage of using the reduced sample estimator is that this function is not monotonic and so is not a distribution function.

Throughout this paper it is assumed that all of the censoring times are known. Notation will be introduced in Section 2. In Section 3 a monotonic and pointwise unbiased estimator of the survival function of the failure time will be proposed, and the asymptotic distribution and the uniform consistency of it will be established. A comparison of the proposed estimator with the Kaplan–Meier and the reduced sample estimator via Monte-Carlo simulation will be presented in Section 4. In Section 5 the concept of *generalized reduced sample type estimators* will be suggested in order to study the relation between the Nelson–Aalen, the Kaplan–Meier, the reduced sample, and the proposed estimator.

2. NOTATION

Denote by T_1, \dots, T_n the independent failure times of a random sample of size n and by S their common survival function. Suppose that the independent censoring times are C_1, \dots, C_n with common distribution function G , and that the sequence $\{C_i\}$ is independent of the sequence $\{T_i\}$. All T_i 's and C_i 's are random variables in the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let \mathbf{E} denote the expectation in the probability space. Under the random censorship model, the observation obtained from each subject consists of $X_i = \min(T_i, C_i)$ and a censoring indicator $\delta_i = \mathbf{1}(T_i \leq C_i)$ where $\mathbf{1}(\cdot)$ is the indicator function. Denote by H the common distribution function of X_i and let $H_u(t) = \mathbf{P}(X_i \leq t, \delta_i = 1) = - \int_0^t [1 - G(s^-)] dS(s)$. Throughout the paper $0/0$ is defined as 0.

The Kaplan–Meier estimator \widehat{S}_{KM} for the survival function S is given by

$$\widehat{S}_{KM}(t) = \prod_{s \leq t} \left(1 - \frac{\#\{i : X_i = s, \delta_i = 1\}}{\#\{i : X_i \geq s\}} \right) \quad \text{for } t \geq 0,$$

where s in the product ranges over the finite set $R \equiv \{s : s = X_i \text{ for some } i\}$.

The Nelson–Aalen estimator $\widehat{\Lambda}_{NA}$ of the cumulative hazard Λ is:

$$\widehat{\Lambda}_{NA}(t) = \sum_{\{s \in R : s \leq t\}} \frac{\#\{X_i = s, \delta_i = 1\}}{\#\{i : X_i \geq s\}} = \int_0^t \frac{d\widehat{H}_u(s)}{1 - \widehat{H}(s^-)}, \quad \text{for } t \geq 0,$$

where $\widehat{H}_u(s) = \#\{i : X_i \leq s, \delta_i = 1\}/n$ and $\widehat{H}(s) = \#\{i : X_i \leq s\}/n$.

Let C_{\max} denote $\max\{C_i : 1 \leq i \leq n\}$. The reduced sample estimator for S is given by

$$\widehat{S}_{rs}(t) = \frac{\#\{i : X_i > t\}}{\#\{i : C_i \geq t\}} = \frac{1 - \widehat{H}(t)}{1 - \widehat{G}(t^-)} \quad \text{for } t \geq 0,$$

where $\widehat{G}(t) = \#\{i : C_i \leq t\}/n$. It is a pointwise unbiased estimator of S on $[0, C_{\max}]$. The idea behind the reduced sample estimator at time t is simple: Only those subjects with potential censoring times not less than t will be considered. Estimators based only on such subjects will be classified as *reduced sample type* estimators.

3. PROPOSED ESTIMATOR AND ITS LARGE SAMPLE PROPERTIES

In this section the following new estimator \widehat{S} is proposed.

$$\widehat{S}(t) = 1 - \sum_{i=1}^n \frac{\mathbf{1}(X_i \leq t, \delta_i = 1)}{\sum_{j=1}^n \mathbf{1}(X_i \leq C_j, \delta_i = 1)} \quad \text{for } t \geq 0,$$

with the convention that $0/0 = 0$. It is a pointwise unbiased estimator and, unlike the reduced sample estimator, monotonically increasing. Moreover, it is uniform consistent and the errors, after suitable normalization, converge weakly to a Gaussian process as the sample size increases.

Theorem 3.1 *The estimator \widehat{S} is a monotonic and pointwise unbiased estimator of S on $[0, C_{\max}]$.*

Proof. The monotonicity of \widehat{S} is obvious. Its expectation is given by

$$\begin{aligned} \mathbf{E}\widehat{S}(t) &= 1 - \sum_{i=1}^n \mathbf{E} \left[\frac{\mathbf{1}(X_i \leq t, \delta_i = 1)}{\sum_{j=1}^n \mathbf{1}(X_i \leq C_j, \delta_i = 1)} \right] \\ &= 1 - \sum_{i=1}^n \mathbf{E} \left[\frac{\mathbf{1}(T_i \leq t, T_i \leq C_i)}{\sum_{j=1}^n \mathbf{1}(T_i \leq C_j, T_i \leq C_i)} \right] \\ &= 1 - \sum_{i=1}^n \mathbf{E} \left[\frac{\mathbf{1}(T_1 \leq t, T_1 \leq C_i)}{\sum_{j=1}^n \mathbf{1}(T_1 \leq C_j, T_1 \leq C_i)} \right]. \end{aligned}$$

Note that

$$\frac{\mathbf{1}(T_1 \leq t, T_1 \leq C_i)}{\sum_{j=1}^n \mathbf{1}(T_1 \leq C_j, T_1 \leq C_i)} = \frac{\mathbf{1}(T_1 \leq t, T_1 \leq C_i)}{\sum_{j=1}^n \mathbf{1}(T_1 \leq C_j)}.$$

Thus, for $0 \leq t \leq C_{\max}$,

$$\begin{aligned} \mathbf{E}\widehat{S}(t) &= 1 - \mathbf{E} \left[\mathbf{1}(T_1 \leq t) \frac{\sum_{i=1}^n \mathbf{1}(T_1 \leq C_i)}{\sum_{j=1}^n \mathbf{1}(T_1 \leq C_j)} \right] \\ &= 1 - \mathbf{E} [\mathbf{1}(T_1 \leq t) \mathbf{1}(T_1 \leq C_{\max})] \\ &= 1 - \mathbf{P}\{T_1 \leq t\} \\ &= S(t). \end{aligned}$$

□

Note that the unbiasedness of $\widehat{S}(t)$ does not require that the T_i are independent or that the C_i are independent and identically distributed.

To study the large sample properties, it is more useful to express $\widehat{S}(t)$ as an integral:

$$\widehat{S}(t) = 1 - \sum_{\{s \in \mathcal{R}: s \leq t\}} \frac{\#\{i : X_i = s, \delta_i = 1\}}{\#\{i : C_i \geq s\}} = 1 - \int_0^t \frac{d\widehat{H}_u(s)}{1 - \widehat{G}(s^-)} \quad \text{for } t \geq 0.$$

For $\tau_G < \infty$ and such that $G(\tau_G) < 1$, define $W_n(t) = \sqrt{n} (\widehat{S}(t) - S(t))$, $0 \leq t \leq \tau_G$. The function $\omega \mapsto W_n(\cdot, \omega)$ is a measurable mapping from (Ω, \mathcal{A})

into $(D[0, \tau_G], \mathcal{D})$, where $D[0, \tau_G]$ is the space of functions on $[0, \tau_G]$ that are right-continuous and have left-hand limits, and \mathcal{D} denotes the Borel σ -algebra induced by the Skorokhod topology (see e.g. Billingsley, 1968). The following theorem can be proved by the same argument as in Lai and Zheng (1993, pp. 40-43, Theorem 1).

Theorem 3.2 *Suppose that S and G are continuous. As $n \rightarrow \infty$, W_n converges in distribution to a Gaussian process W in $D[0, \tau_G]$ with $\mathbf{E}[W(t)] = 0$ and*

$$\text{cov}(W(t_1), W(t_2)) = \int_0^{\min(t_1, t_2)} \frac{dH_u(s)}{[1 - G(s)]^2} = - \int_0^{\min(t_1, t_2)} \frac{dS(s)}{1 - G(s)},$$

where $t_1, t_2 \in [0, \tau_G]$.

Let $\text{var}(W(t))$ be denoted by $\sigma^2(t)$. Under the same conditions as Theorem 3.2, the asymptotic variance of $\sqrt{n} \left(\widehat{S}_{KM}(t) - S(t) \right)$ for $0 \leq t \leq \tau_H$ where $\tau_H < \infty$ and such that $H(\tau_H) < 1$ is

$$\sigma_{KM}^2(t) = S(t)^2 \int_0^t \frac{dH_u(s)}{[1 - H(s)]^2} = - \int_0^t \left(\frac{S(t)}{S(s)} \right)^2 \frac{dS(s)}{1 - G(s)}.$$

Thus, if $S(t)/S(s) \approx 1$, then $\sigma^2(t) \approx \sigma_{KM}^2(t)$. That is, $\widehat{S}(t)$ is not only unbiased (which is a desirable property, especially when n is small, a situation which often arises in clinical research) but also quite efficient for small t .

Since \widehat{S} is similar to the Nelson–Aalen estimator, it is natural to expect uniform consistency of \widehat{S} over some closed interval. A maximal supermartingale inequality is used to establish this result.

Lemma 3.1 *Suppose that $0/0$ is interpreted as 0 . Let \mathcal{F}_t be the history of all X_i and C_j such that $X_i \leq t$ and $C_j \leq t$ for $0 \leq i, j \leq n$. The process M given by $M(t) = S(t) - \widehat{S}(t)$ for $t \geq 0$ is an \mathcal{F}_t -supermartingale.*

Proof. Let $F = 1 - S$ denote the distribution function of T_i . Clearly $\mathbf{E}|M(t)| \leq 2$ for all $t \geq 0$. Consider $\mathbf{E}[M(t+v)|\mathcal{F}_t]$ for any $v \geq 0$:

$$\begin{aligned}
\mathbf{E}[M(t+v)|\mathcal{F}_t] &= \mathbf{E} \left[\int_0^{t+v} \frac{d\widehat{H}_u(s)}{1 - \widehat{G}(s^-)} - F(t+v) \middle| \mathcal{F}_t \right] \\
&= M(t) - F(t+v) + F(t) + \mathbf{E} \left[\int_t^{t+v} \frac{d\widehat{H}_u(s)}{1 - \widehat{G}(s^-)} \middle| \mathcal{F}_t \right] \\
&= M(t) - F(t+v) + F(t) + F(\min\{t+v, C_{\max}\}) \\
&\quad - F(\min\{t, C_{\max}\}) \\
&\leq M(t)
\end{aligned}$$

□

The uniform consistency (in probability) of \widehat{S} over $[0, \tau_G]$ follows from the theorem below.

Theorem 3.3 *Let $M^* = \sup_{0 \leq t \leq \tau_G} |S(t) - \widehat{S}(t)|$. For each $\varepsilon > 0$,*

$$\mathbf{P}\{M^* \geq \varepsilon\} = O(n^{-1/2}).$$

Proof. Since M is a right-continuous supermartingale, it satisfies the following maximal supermartingale inequality (see e.g. He *et al.*, 1992, p. 53, Theorem 2.42):

$$\begin{aligned}
\varepsilon \mathbf{P}\{M^* \geq \varepsilon\} &\leq \mathbf{E}[M(0)] + 2\mathbf{E}[|M(\tau_G)|] \\
&\leq 2\sqrt{\mathbf{E}[M(\tau_G)^2]} \\
&= O(\sigma(\tau_G)n^{-1/2}).
\end{aligned}$$

□

4. SIMULATION

A comparison of the proposed estimator \widehat{S} with the Kaplan–Meier \widehat{S}_{KM} and the reduced sample estimator \widehat{S}_{rs} for small and large sample sizes via Monte-Carlo simulation is done. Consider that T_i 's follow the gamma distribution with shape parameter 4 and scale parameter 1 and that C_i 's are uniform on $[0, 6]$, $i = 1, \dots, n$. This situation arises when the length of a study is, say, 6 years and n subjects enter the study uniformly during these 6 years. The root mean squared errors of these three estimators obtained from 1000 simulations are given in Figures 4.2.

5. GENERALIZED REDUCED SAMPLE TYPE ESTIMATORS

In this section the concept of the *generalized reduced sample type estimators* will be introduced and the Kaplan–Meier, the Nelson–Aalen, the reduced sample and the proposed estimator all belong in this class of estimators.

The estimator \widehat{S} and the Nelson–Aalen estimator $\widehat{\Lambda}$ are closely related. The differential of \widehat{S} ,

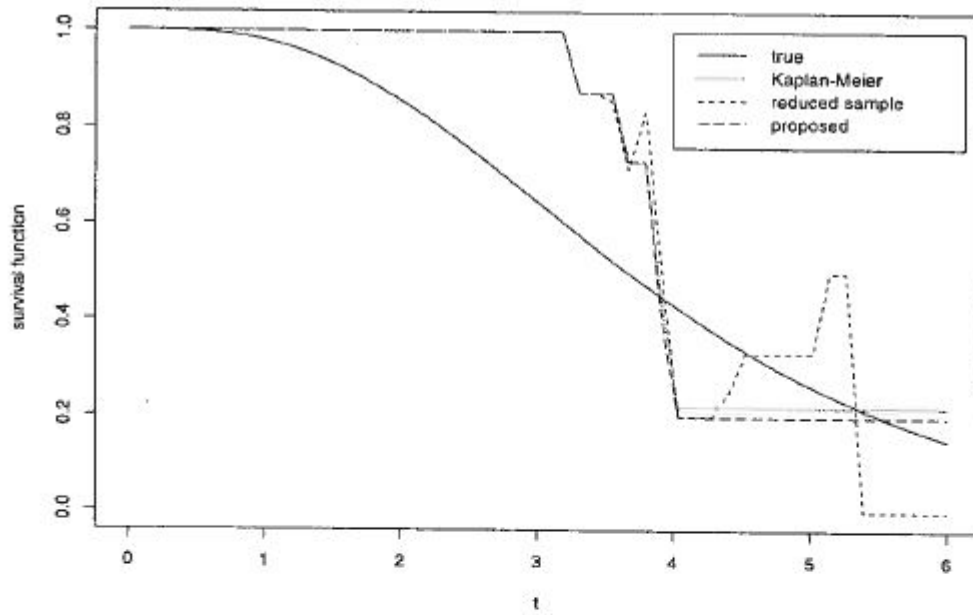
$$d\widehat{S}(s) = -\frac{d\widehat{H}_u(s)}{1 - \widehat{G}(s^-)} = -\frac{d\#\{i : X_i \leq s, \delta_i = 1\}}{\#\{i : C_i \geq s\}}, \quad s > 0,$$

is in fact a reduced sample type estimator of $dS(s) = -dH_u(s)/[1 - G(s^-)]$, since only those subjects with potential censoring times not less than s will be considered. The differential of $\widehat{\Lambda}_{NA}$,

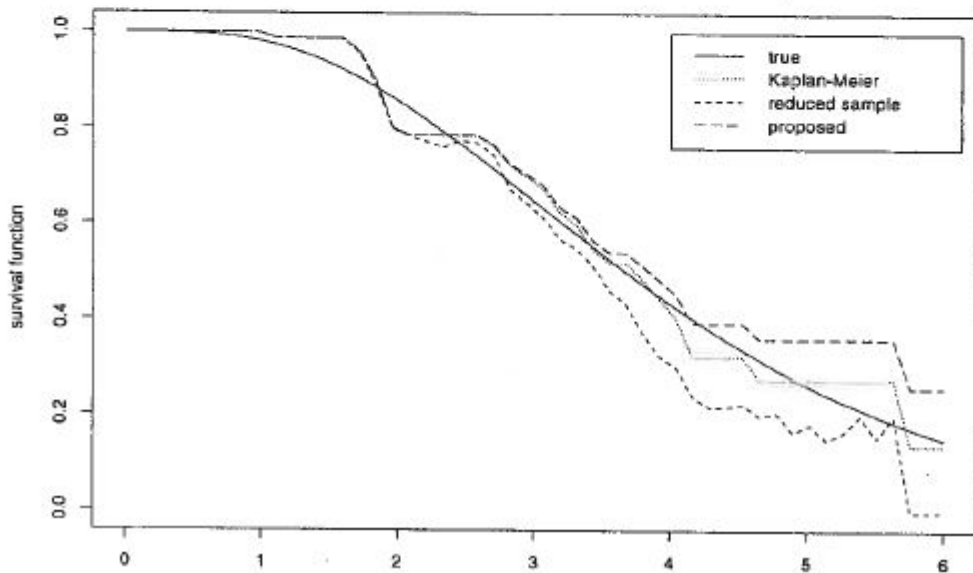
$$d\widehat{\Lambda}_{NA}(s) = \frac{d\widehat{H}_u(s)}{1 - \widehat{H}(s^-)} = \frac{d\widehat{H}_u(s)}{1 - \widehat{G}(s^-)} \times \frac{1 - \widehat{G}(s^-)}{1 - \widehat{H}(s^-)} = -\frac{d\widehat{S}(s)}{\widehat{S}_{rs}(s^-)}, \quad s > 0,$$

is the ratio of reduced sample type estimators of $-dS(s)$ and $S(s^-)$, and $d\Lambda(s) = -dS(s)/S(s^-)$. Thus, $d\widehat{\Lambda}_{NA}(s)$ is of the reduced sample type, although it is not unbiased.

Since $S(t)$ is equal to $\prod_{(0,t]}(1 - d\Lambda(s))$, where \prod denotes the product integral, the Kaplan–Meier estimator $\widehat{S}_{KM}(t) = \prod_{(0,t]}(1 - d\widehat{\Lambda}_{NA}(s))$ builds



$n = 10$



$n = 100$

Figure 4.1: The true survival function, and the estimates obtained from the Kaplan-Meier, the reduced sample and the proposed estimator, based on a sample of size n .

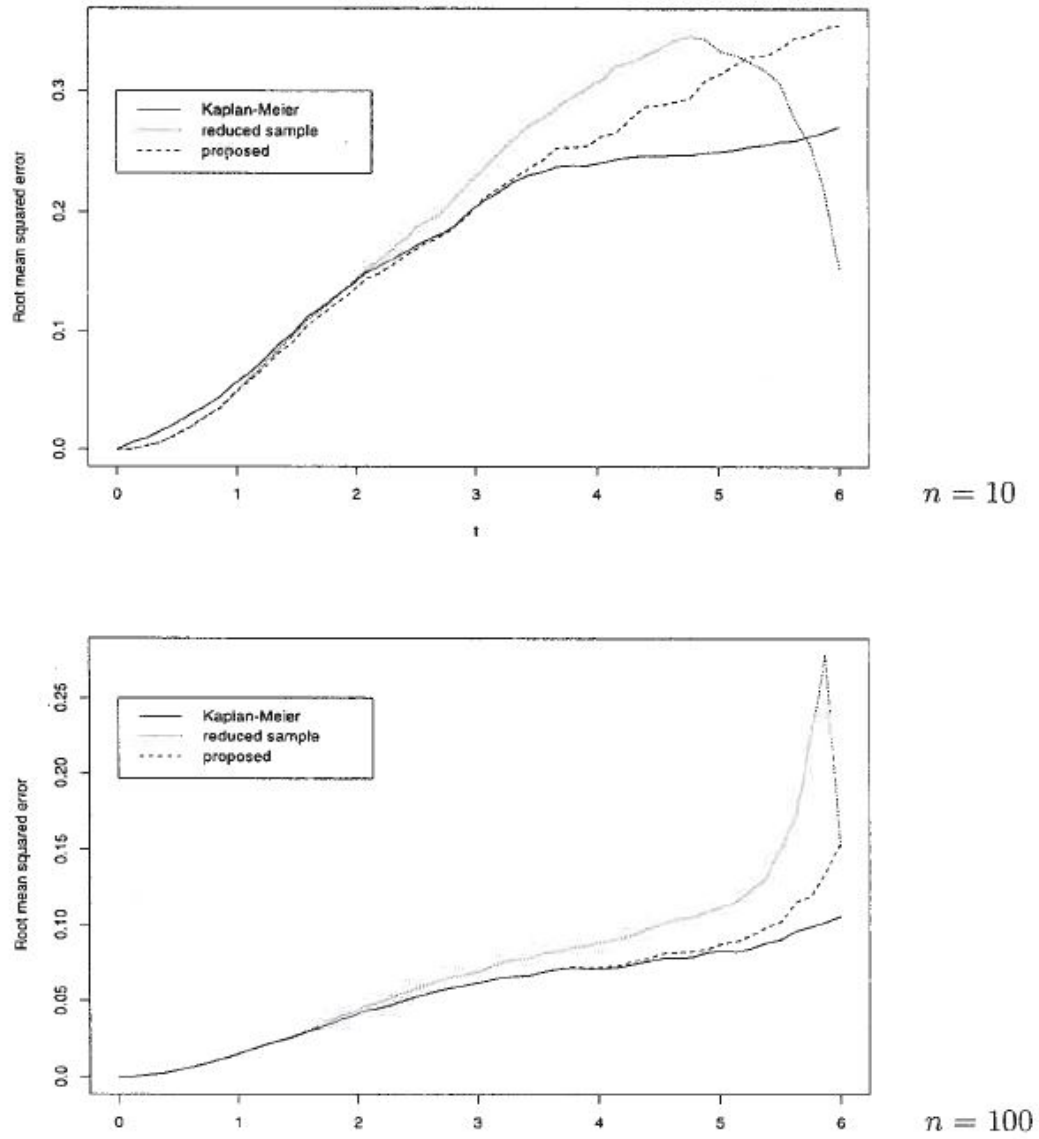


Figure 4.2: Root mean squared errors of the Kaplan-Meier, the reduced sample and the proposed estimator obtained from 1000 samples of size n .

ordered sample type estimators of the differential $d\Lambda(s)$ for $0 < s \leq t$. It motivates the following definition.

Definition 5.1 *Let f be a real-valued function defined on \mathbb{R} . For $t \in \mathbb{R}$, $\widehat{f}(t)$ is a generalized reduced sample type estimator of $f(t)$ if there exists a countable or uncountable collection of functions $\{f_i\}$ defined on \mathbb{R} such that $\widehat{f}(t)$ depends solely on reduced sample type estimators of $\{f_i(s)\}$ for $s \leq t$.*

Thus, the Kaplan–Meier estimator is a generalized reduced sample type estimator. The differential $d\widehat{S}(s)$ is of the reduced sample type. Consequently, the estimator $\widehat{S}(t) = \int_0^t d\widehat{S}(s)$ is also a generalized reduced sample estimator, and it is closely related to the Kaplan–Meier estimator because

$$\widehat{S}_{KM}(t) = \boldsymbol{\pi}_{(0,t]} \left(1, \frac{d\widehat{S}(s)}{\widehat{S}_{rs}(s^-)} \right).$$

ACKNOWLEDGEMENTS

I thank M. C. Ngai for his help in the simulation and the referee for helpful comments and suggestions.

BIBLIOGRAPHY

- Aalen, O. O. (1975). Statistical inference for a family of counting processes. PhD thesis, University of California, Berkeley.
- Aalen, O. O. (1978). “Nonparametric inference for a family of counting processes,” *Ann. Statist.*, **6**, 701-726.
- Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

- Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.
- Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman & Hall, London.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, Wiley, New York.
- Gill, R. D. (1989). "Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1)," *Scand. J. Statist.*, **16**, 97-128.
- He, S. W., Wang, J. G. and Yan, J. A. (1992). *Semimartingale Theory and Stochastic Calculus*, CRC Press, Boca Raton.
- Johansen, S. (1978). "The product limit estimator as maximum likelihood estimator," *Scand. J. Statist.*, **5**, 195-199.
- Kaplan, E. L. and Meier, P. (1958). "Non-parametric estimation from incomplete observations," *J. Amer. Statist. Assoc.*, **53**, 457-481, 562-563.
- Lai, T. L. and Zheng, Z. K. (1993). *Survival Analysis*, Zhejiang Publishing House of Science and Technology, Zhejiang (in Chinese).
- Nelson, W. (1969). "Hazard plotting for incomplete failure data," *J. Qual. Technol.*, **1**, 27-52.
- Nelson, W. (1972). "Theory and applications of hazard plotting for censored failure data," *Technometrics*, **14**, 945-965.
- Wellner, J. A. (1982). "Asymptotic optimality of the product limit estimator," *Ann. Statist.*, **10**, 595-602.