A NOTE ON THE BRESLOW ESTIMATOR

BU-1548 -M **January, 2001**

Robert L. Strawderman
Xuelin Huang

Keywords: Ito's formula, Martingale, Nelson-Aalen estimator.

Abstract:
This paper investigates the structure of the bias and mean squared error for a common estimator of the survivor function, namely the Breslow (1972) estimator. In particular, using Ito's change-of-variables formula, new formulas for the bias and mean squared error (hence variance) for this estimator are established. These formulas are subsequently used to investigate a conjecture of Fleming and Harrington (1984). In particular, we show that the mean squared error of the Kaplan-Meier estimator (Kaplan and Meier, 1958) exceeds that of the Breslow estimator whenever the true survival probability is bounded sufficiently far from zero.

Under revision for
Stat & Probability Letters
1548

A Note on the Breslow Estimator

Robert L. Strawderman
Xuelin Huang *

January 3, 2001

Abstract

This paper investigates the structure of the bias and mean squared error for a common estimator of the survivor function, namely the Breslow (1972) estimator. In particular, using Itô's change-of-variables formula, new formulas for the bias and mean squared error (hence variance) for this estimator are established. These formulas are subsequently used to investigate a conjecture of Fleming and Harrington (1984). In particular, we show that the mean squared error of the Kaplan-Meier estimator (Kaplan and Meier, 1958) exceeds that of the Breslow estimator whenever the true survival probalility is bounded sufficiently far from zero.

KEYWORDS: Itô's formula; Martingale; Nelson-Aalen estimator.

*RL Strawderman is Associate Professor, Department of Biometrics, Cornell University, Ithaca NY 14583. Email: rls54@cornell.edu. Xuelin Huang is Ph.D. student, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109-2029. Email: xueling@umich.edu

1 Introduction

In biostatistical studies, interest often centers on the time to an event of interest. The observation of this event, or survival, time may be precluded for some subjects (i.e., censored) by study termination, loss-to-follow-up, etcetera. Suppose there are n independent subjects under study, and we are interested in estimating the survivor function $S(t) = P\{U > t\}$. On each subject, the observed data take the form (X_i, δ_i), where $X_i = \min(U_i, C_i)$ for some potential censoring time C_i and $\delta_i = \boldsymbol{I}(U_i \leq C_i)$. It is assumed that $U_i, i = 1, \ldots, n$, are independent and identically distributed (*i.i.d.*) with absolutely continuous distribution $F = 1 - S$; $C_i, i = 1, \ldots, n$, are *i.i.d.* with survivor function $P\{C > c\} = G(c)$; and, the failure and censoring times are mutually independent.

The most commonly used estimator of $S(t)$ is that originally proposed by Kaplan and Meier (1958). In product integral form, this estimator is written

$$\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{d\bar{N}(s)}{\bar{Y}(s)}\right),$$

where $\bar{N}(t) = \sum_{i=1}^{n} N_i(t)$, $N_i(t) = \boldsymbol{I}(X_i \leq t, \delta_i = 1)$, $\bar{Y}(t) = \sum_{i=1}^{n} Y_i(t)$, and $Y_i(t) = \boldsymbol{I}(X_i \geq t)$. An alternative, and asymptotically equivalent, estimator can be constructed from the Nelson-Aalen estimator of the cumulative hazard function $\Lambda(t)$. Specifically, via the relationship $S(t) = e^{-\Lambda(t)}$, Breslow (1972) proposed

$$\tilde{S}(t) = e^{-\hat{\Lambda}(t)} = \exp\left\{-\int_0^t \frac{d\bar{N}(s)}{\bar{Y}(s)}\right\}$$

as an estimator of $S(t)$. Technically, $\tilde{S}(t)$ is appropriate for continuous U only, and attention in this paper is restricted to this case.

Fleming and Harrington (1984) numerically establish the superiority of the Breslow estimator over the Kaplan-Meier estimator (i.e., in terms of mean squared error) for uncensored data for times t such that $S(t) > 0.2$. They further conjecture, based on the results of a limited simulation study, that for censored data the Breslow estimator continues to have smaller MSE than the Kaplan-Meier estimator whenever the survivor function is sufficiently bounded away from zero. This is interesting because the Kaplan-Meier and Breslow estimators are respectively regarded as being unbiased and biased estimators for $S(t)$. Consequently, the Breslow estimator is less variable than the Kaplan-Meier estimator, presenting an attractive argument for carrying out statistical inference based on the Breslow estimator in the case of small sample sizes. For example, Fleming and Harrington (1984, Table III) shows that the improvement in MSE can be as much as 15% at the median survival time in a classical clinical trials setting (i.e., where the support of the censoring distribution is strictly smaller than that of the failure time distribution).

In this paper, we establish new formulas for the bias and MSE of the Breslow estimator in Sections 2.3.1 and 2.3.2. These formulas are derived using Itô's stochastic integration formula for semimartingales (Protter, 1990, pp. 71-72). We then show how the formulas for the bias and MSE can be used to facilitate a theoretical comparison between the MSE of the Breslow and Kaplan-Meier estimators of $S(t)$. In particular, we verify in Section 3 the simulation-based conjecture of Fleming and Harrington (1984) that the Breslow estimator has smaller mean squared error than the Kaplan-Meier estimator for times bounded away from the tail of the survivor function.

2 Main results

Exponentially accurate formulas for the bias and MSE of the Breslow estimator shall be established in a stepwise fashion. Some preliminary results are first presented in Section 2.1. Then, Itô's formula for semimartingale processes, specialized to the case of stochastic integrals with respect to counting process martingales, is reviewed in Section 2.2. Finally, using this result, new formulas for the bias and MSE of the Breslow estimator are obtained in Section 2.3.

2.1 Preparations

In addition to the assumptions and counting process notation given earlier, let $J(t) = \boldsymbol{I}(\bar{Y}(t) > 0)$, $\Lambda^*(t) = \int_0^t J(u)\lambda(u)\,du$, $S^*(t) = \exp\{-\Lambda^*(t)\}$, $T = \inf\{s : \bar{Y}(s) = 0\}$, $\pi(t) = P\{X > t\} = S(t)G(t)$, and $\tau = \sup\{t : \pi(t) > 0\}$. These definitions, as well as the explicit assumptions that $t < \tau$ and $\{N_1, \cdots, N_n\}$ is a multivariate counting process (e.g., see Fleming and Harrington, 1991, Def. 2.5.1), shall be used throughout the remainder of this paper.

It is well-known that the Nelson-Aalen estimator $\hat{\Lambda}(t) = \int_0^t J(u)\frac{d\bar{N}(u)}{\bar{Y}(u)}$ is not an unbiased estimator of the cumulative hazard function $\Lambda(\cdot)$. In fact, $E[\hat{\Lambda}(t) - \Lambda^*(t)] = 0$ (Fleming and

Harrington, 1991, Thm. 3.2.1); moreover, the stochastic process $\{M_{\hat{\Lambda}}(t), t \geq 0\}$ with

$$M_{\hat{\Lambda}}(t) = \hat{\Lambda}(t) - \Lambda^*(t)$$

is a local square integrable martingale with respect to the filter $\{\mathcal{F}_t, t \geq 0\}$, where

$$\mathcal{F}_t = \bigvee_{i=1}^{n} \sigma\{N_i(u), \mathbf{I}(X_i \leq u, \Delta_i = 0) : 0 \leq u \leq t\}. \tag{1}$$

The process $M_{\hat{\Lambda}}(\cdot)$ plays a key role in establishing formulas for both the bias and MSE of $\tilde{S}(t)$. This is easy to see by noting that for t such that $S(t) > 0$,

$$\widetilde{S}(t) - S(t) = S(t)\left(\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right) + \mathbf{I}\{T < t\}\tilde{S}(T)\left(1 - \frac{S(t)}{S(T)}\right)$$

cf. Fleming and Harrington (1991, Corollary 3.2.1). The second term on the right hand side is nonnegative and bounded above by $\mathbf{I}\{T < t\}(1 - S(t))$; since $E[\mathbf{I}(T < t)] = (1 - \pi(t))^n$ (e.g., see Fleming and Harrington, 1991, Thm. 3.2.1), the following easily obtain:

$$E[\tilde{S}(t) - S(t)] \;=\; S(t)E\left[\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right] + O((1 - \pi(t))^n) \tag{2}$$

$$MSE(\tilde{S}(t)) \;=\; S^2(t)E\left[\left(\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right)^2\right] + O((1 - \pi(t))^n)\,. \tag{3}$$

We use Itô's change-of-variables formula, reviewed in the next section, to simplify the expectations respectively appearing in (2) and (3).

2.2 Itô's Formula for Counting Process Martingales

Itô's formula is a change-of-variables formula useful for continuous functions of martingales and semimartingales that has seen comparatively little use in the survival analysis literature. One important example of the use of Itô's formula can be found in Mykland (1994), where it is used as the basis for deriving versions of the Bartlett identities for martingales. Strawderman and Wells (1997) employ these Bartlett identities in order to simplify cumulant calculations needed for deriving Edgeworth expansions for studentized versions of the Nelson-Aalen and Kaplan-Meier estimators of $\Lambda(t)$ and $S(t)$, respectively.

A formula applicable to general semimartingale processes can be found in Protter (1990, Thm. 32, p. 71). The result below specializes Itô's formula to the case where the stochastic process in question can be written as the stochastic integral $X(t) = \int_0^t H(u)d\bar{M}(u)$, where $H(u)$ is a locally bounded and predictable process and $\bar{M}(t) = \bar{N}(t) - \int_0^t \bar{Y}(u)\lambda(u)du$ is the usual counting process martingale of survival analysis. A proof is given in the appendix. With obvious modification the results below apply more generally to most stochastic integrals arising in survival analysis (e.g., integrals arising in the analysis of the Cox model).

Lemma 1 (Itô's Formula) *Let $X(t) = \int_0^t H(u)d\bar{M}(u)$ where $\bar{M}(t) = \bar{N}(t) - \int_0^t \bar{Y}(u)\lambda(u)du$ is a local square integrable martingale with respect to the filter in (1) and $H(u)$ is a locally bounded predictable process. Let $f(\cdot)$ be a twice-differentiable deterministic function. Then, provided the compensator $\bar{A}(t) = \int_0^t \bar{Y}(u)\lambda(u)du$ is pathwise continuous,*

$$f(X(t)) = \int_0^t \{f(X(s-) + H(s)) - f(X(s-))\} d\bar{M}(s) + \int_0^t D(s)\bar{Y}(s)\lambda(s)ds + f(0) \qquad (4)$$

where $D(s) = f(X(s-) + H(s)) - f(X(s-)) - f'(X(s-))H(s)$.

Remark: The right-hand side of (4) gives the unique Doob-Meyer decomposition of the semi-martingale $f(X(t))$ (Protter, 1990, Thm. 18, p. 107). Since $X(t)$ is a mean zero martingale, these results show that $f(X(t))$ is not a martingale unless $D(s) \equiv 0$, which occurs if $f(x)$ is linear in x. It is evident from (4) that the resulting "bias" is a predictable process whose magnitude depends directly on the degree of departure from linearity. Continuity of the compensator is an important assumption in the above lemma. If $\bar{A}(t)$ has discontinuities, the formulas given in the lemma do not hold, and one must begin with (13) (see the Appendix).

Remark: Lemma 1 is useful for obtaining the bias term for a nonlinear transformation of a martingale. Virtually identical arguments may be used to establish the following companion result for a process $Z(t) = \int_0^t H(s)d\bar{N}(s)$:

$$f(Z(t)) = \int_0^t W(s)d\bar{M}(s) + \int_0^t W(s)\bar{Y}(s)\lambda(s)ds + f(0)$$

for $W(s) = f(Z(s-) + H(s)) - f(Z(s-))$. Hence, Itô's formula also provides an easy method for computing the compensator of a nonlinear, twice-differentiable function of $Z(t)$.

2.3 Formulas for the bias and MSE of $\tilde{S}(t)$

That Itô's formula applies to the problem of computing the bias and MSE of $\tilde{S}(t)$ follows immediately from (2), (3), and the facts that

$$M_{\hat{\Lambda}}(t) \;=\; \hat{\Lambda}(t) - \Lambda^*(t) \;=\; \int_0^t \frac{J(u)}{\bar{Y}(u)} d\bar{M}(u) \tag{5}$$

and $M_{\hat{\Lambda}}(0) = 0$. For convenience, we employ the notation $B(t) = \exp\{-M_{\hat{\Lambda}}(t)\} - 1$, and consider bias and mean squared error in separate subsections.

2.3.1 Bias

Letting $f(x) = e^{-x}$ and $X(t) = M_{\hat{\Lambda}}(t)$, we see that $f'(x) = -e^{-x}$, $f(X(0)) = 1$ and $H(s) = J(s)/\bar{Y}(s)$. As an immediate consequence of Lemma 1, we obtain the representation

$$\begin{aligned}\exp\{-M_{\hat{\Lambda}}(t)\} - 1 &= \int_0^t \exp(-M_{\hat{\Lambda}}(u-)) \left(\exp\left\{-\frac{J(u)}{\bar{Y}(u)}\right\} - 1\right) d\bar{M}(u) \\ &+ \int_0^t \exp(-M_{\hat{\Lambda}}(u-)) \left(\exp\left\{-\frac{J(u)}{\bar{Y}(u)}\right\} - 1 + \frac{J(u)}{\bar{Y}(u)}\right) \bar{Y}(u)\lambda(u)du. \end{aligned} \tag{6}$$

From (2) it suffices to compute the expectation of (6). Notice that

$$\exp\{-\frac{J(u)}{\bar{Y}(u)}\} - 1 + \frac{J(u)}{\bar{Y}(u)} \;=\; J(u)\left(\exp\{-\frac{1}{\bar{Y}(u)}\} - 1 + \frac{1}{\bar{Y}(u)}\right);$$

thus, using (5) and the fact that $J(u)/S^*(u) = J(u)/S(u)$,

$$E\left[B(t)\right] \;=\; E\left[\int_0^t \frac{\tilde{S}(u-)}{S(u)} g(\bar{Y}(u))J(u)\lambda(u)du\right], \tag{7}$$

where $g(u) = u\exp(-1/u) - u + 1$ for $u > 0$ and $g(0) = 1$. Combining (2) and (7),

$$E\left[\tilde{S}(t) - S(t)\right] \;=\; S(t)E\left[\int_0^t \frac{\tilde{S}(u-)}{S(u)} g(\bar{Y}(u))J(u)\lambda(u)du\right] + O((1-\pi(t))^n). \tag{8}$$

This formula for the bias of the Breslow estimator appears to be new. Since $g(u)$ is a positive decreasing function, (8) also reflects the fact that the bias of $\tilde{S}(t)$ is known to be positive. This follows directly from the facts that $\tilde{S}(t) \geq \hat{S}(t)$ (e.g., Breslow and Crowley, 1974, Lemma 1) and that $\hat{S}(t)$ is positively biased (Fleming and Harrington, 1991, Lemma 3.2.1).

2.3.2 MSE

By (3), the computation of $MSE(\tilde{S}(t))$ requires $E[B^2(t)] = E\left[\left(\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right)^2\right]$. Applying Lemma 1 with $f(x) = (e^{-x} - 1)^2$ and taking expectations of both sides leads to

$$E\left[B^2(t)\right] = E\left[\int_0^t \left(f(M_{\hat{\Lambda}}(u-) + J(u)/\bar{Y}(u)) - f(M_{\hat{\Lambda}}(u-)) - f'(M_{\hat{\Lambda}}(u-))\frac{J(u)}{\bar{Y}(u)}\right) \bar{Y}(u)\lambda(u)du\right],$$

where $f'(x) = 2e^{-x} - 2e^{-2x}$. Easy algebra shows that, for $x, h \in \mathbb{R}$,

$$f(x+h) - f(x) - f'(x)h = e^{-2x}(e^{-2h} - 1 + 2h) - 2e^{-x}(e^{-h} + h - 1),$$

from which we obtain

$$\begin{aligned} E\left[B^2(t)\right] &= E\left[\int_0^t \exp\{-2M_{\hat{\Lambda}}(u-)\}\left(\exp\{-2\frac{J(u)}{\bar{Y}(u)}\} - 1 + 2\frac{J(u)}{\bar{Y}(u)}\right)\bar{Y}(u)\lambda(u)du\right] \\ &\quad - 2\,E\left[\int_0^t \exp\{M_{\hat{\Lambda}}(u-)\}\left(\exp\{-\frac{J(u)}{\bar{Y}(u)}\} - 1 + \frac{J(u)}{\bar{Y}(u)}\right)\bar{Y}(u)\lambda(u)du\right]. \end{aligned}$$

Each of the above expectations can be now be simplified exactly as was done in the previous section. By (3), the MSE of the Breslow estimator is then obtained:

$$\begin{aligned} \mathrm{MSE}(\tilde{S}(t)) &= S^2(t)E\left[\int_0^t \left(\frac{\tilde{S}(u-)}{S(u)}\right)^2 h(\bar{Y}(u))J(u)\lambda(u)du\right] \\ &- 2\,S^2(t)E\left[\int_0^t \frac{\tilde{S}(u-)}{S(u)}g(\bar{Y}(u))J(u)\lambda(u)du\right] + O((1-\pi(t))^n), \qquad (9) \end{aligned}$$

where $h(u) = u\exp(-2/u) - u + 2$ for $u > 0$, $h(0) = 2$, and $g(u)$ is as defined earlier.

3 Expansions for the MSE of $\tilde{S}(t)$ and $\hat{S}(t)$

In the absence of censoring, Fleming and Harrington (1984) show that Breslow's estimator has a smaller MSE than the Kaplan-Meier estimator whenever the true survival probability $S(t)$ exceeds 1/5. They also did a simulation study to compare the MSEs of $\tilde{S}(t)$ and $\hat{S}(t)$ under random censoring, and conjectured based on their results that a similar result could be established in the presence of censoring. We now explore whether this is indeed the case. For this we turn to asymptotics, and to keep things manageable, consider differences in MSE up to terms of order $o(n^{-2})$. We only sketch the arguments here, leaving the details to the reader. The calculations involve tedious but straightforward expansions.

3.1 An expansion for MSE($\tilde{S}(t)$)

The MSE of $\tilde{S}(t)$ is given in (9). Setting $h(x,s) = \exp(-2x)h(1/s) - 2\exp(-x)g(1/s)$, expanding this function using a bivariate Taylor series, and substituting $x = M_{\hat{\Lambda}}(u-)$ and $s = [\bar{Y}(u)]^{-1}$, it can be shown that

$$\frac{\text{MSE}(\tilde{S}(t))}{S^2(t)} = E\left[\int_0^t \left(\frac{1}{\bar{Y}(u)} - \frac{1}{\bar{Y}^2(u)} - 3\frac{M_{\hat{\Lambda}}(u-)}{\bar{Y}(u)} + \frac{7}{2}\frac{M_{\hat{\Lambda}}^2(u-)}{\bar{Y}(u)}\right) J(u)\lambda(u)du\right] + o(n^{-2}).$$

For $t < \tau$,

$$E\left[\int_0^t \left(\frac{1}{\bar{Y}(u)} - \frac{1}{\bar{Y}^2(u)}\right) J(u)\lambda(u)du\right] = \left(\frac{1}{n} - \frac{1}{n^2}\right)\int_0^t \frac{\lambda(u)}{\pi(u)}du + o(n^{-2}).$$

Using the asymptotic U-statistic representation of $M_{\hat{\Lambda}}(t)$ devised by Lai and Wang (1993, pp. 521-524) and employing calculations like those found in Strawderman and Wells (1997, Appendix C), one may prove that

$$\begin{aligned} E\left[\int_0^t \frac{M_{\hat{\Lambda}}(u-)}{\bar{Y}(u)} J(u)\lambda(u)du\right] &= E\left[M_{\hat{\Lambda}}(t)\int_0^t \frac{\lambda(u)J(u)}{\bar{Y}(u)}du\right] \\ &= -\frac{1}{n^2}E\left[M_{\hat{\Lambda}}(t)\int_0^t \bar{Y}(u)\frac{\lambda(u)}{\pi^2(u)}du\right] + o(n^{-2}) \\ &= \frac{1}{2n^2}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 + o(n^{-2}). \end{aligned}$$

A considerably more tedious argument shows that

$$E\left[\int_0^t \frac{M_{\hat{\Lambda}}^2(u-)}{\bar{Y}(u)} J(u)\lambda(u)du\right] = \frac{1}{2n^2}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 + o(n^{-2}).$$

Putting these calculations together, we obtain the following expansion for MSE($\tilde{S}(t)$):

$$\text{MSE}(\tilde{S}(t)) = \frac{S^2(t)}{n}\left[\int_0^t \frac{\lambda(u)}{\pi(u)}du + \frac{R_B(t)}{n}\right] + o(n^{-2}), \tag{10}$$

where

$$R_B(t) = \frac{1}{4}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 - \int_0^t \frac{\lambda(u)}{\pi(u)}du.$$

REMARK: The formulas given for the bias and MSE of $\widetilde{S}(t)$ further imply that

$$\text{Var}(\widetilde{S}(t)) = S^2(t)\left(E\left[\left(\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right)^2\right] - E\left[\exp\{-M_{\hat{\Lambda}}(t)\} - 1\right]^2\right) + O((1-\pi(t))^n).$$

Using the above expansion results, we find that

$$\operatorname{Var}(\widetilde{S}(t)) = \frac{S^2(t)}{n}\left(1 - \frac{1}{n}\right)\int_0^t \frac{\lambda(u)}{\pi(u)}du + o(n^{-2}),$$

suggesting that the usual Delta Method approximation may provide a better-than-expected method for approximating $\operatorname{Var}(\widetilde{S}(t))$. Simulation results (not shown) confirm that the usual Delta Method approximation provides an excellent approximation to $\operatorname{Var}(\widetilde{S}(t))$. These results also show that it is a poor approximation to $\operatorname{MSE}(\tilde{S}(t))$ when $\bar{Y}(t)$ is small.

3.2 An expansion for $\operatorname{MSE}(\hat{S}(t))$

We now require a similar expansion for $\operatorname{MSE}(\hat{S}(t))$. From Fleming and Harrington (1991, Lemma 3.2.1 and Eqn. 2.13),

$$\operatorname{MSE}(\hat{S}(t)) = S^2(t) E\left[\int_0^t \left(\frac{\hat{S}(u-)}{S^*(u-)}\right)^2 \frac{J(u)}{\bar{Y}(u)}\lambda(u)du\right] + O((1-\pi(t))^n).$$

Defining

$$M_{\hat{S}}(t) = \frac{\hat{S}(t)}{S^*(t)} - 1,$$

the following identity is obtained:

$$\frac{\operatorname{MSE}(\hat{S}(t))}{S^2(t)} = E\left[\int_0^t \left(M_{\hat{S}}^2(u-) + 2M_{\hat{S}}(u-) + 1\right)\frac{J(u)}{\bar{Y}(u)}\lambda(u)du\right] + O((1-\pi(t))^n).$$

Similarly to before, it can be established that

$$E\left[\int_0^t \frac{J(u)}{\bar{Y}(u)}\lambda(u)du\right] = \frac{1}{n}\int_0^t \frac{\lambda(u)}{\pi(u)}du + \frac{1}{n^2}\left(\int_0^t \frac{\lambda(u)}{\pi^2(u)}du - \int_0^t \frac{\lambda(u)}{\pi(u)}du\right) + o(n^{-2}).$$

Using the asymptotic U-statistic representation of $M_{\hat{S}}(t)$ devised by Gross and Lai (1996, pp. 513-515) and employing calculations like those found in Strawderman and Wells (1997, Appendix C), one may also prove that

$$E\left[\int_0^t M_{\hat{S}}(u-)\frac{J(u)}{\bar{Y}(u)}\lambda(u)du\right] = -\frac{1}{2n^2}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 + o(n^{-2})$$

and that

$$E\left[\int_0^t M_{\hat{S}}^2(u-)\frac{J(u)}{\bar{Y}(u)}\lambda(u)du\right] = \frac{1}{2n^2}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 + o(n^{-2}).$$

Consequently, the expansion analogous to (10) for the Kaplan-Meier estimator is

$$\mathrm{MSE}(\hat{S}(t)) = \frac{S^2(t)}{n}\left[\int_0^t \frac{\lambda(u)}{\pi(u)}du + \frac{R_{KM}(t)}{n}\right] + o(n^{-2}) \tag{11}$$

where

$$R_{KM}(t) = \int_0^t \frac{\lambda(u)}{\pi^2(u)}du - \int_0^t \frac{\lambda(u)}{\pi(u)}du - \frac{1}{2}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2.$$

3.3 On $\mathrm{MSE}(\tilde{S}(t)) - \mathrm{MSE}(\hat{S}(t))$

The results of the previous two sections now imply

$$\begin{aligned}\mathrm{MSE}(\tilde{S}(t)) - \mathrm{MSE}(\hat{S}(t)) &= n^{-2}S^2(t)\left(R_B(t) - R_{KM}(t)\right) + o(n^{-2}) \\ &= n^{-2}S^2(t)\left[\frac{3}{4}\left(\int_0^t \frac{\lambda(u)}{\pi(u)}du\right)^2 - \int_0^t \frac{\lambda(u)}{\pi^2(u)}du\right] + o(n^{-2}). \quad (12)\end{aligned}$$

This expansion is valid provided $t < \tau$ (equivalently, $\pi(t) > 0$), a condition that causes no difficulty in the case of uncensored data. However, for censored data, this expansion does not hold for values of t such that $\pi(t) = 0$. Such is the case, for example, when $G(t) = 0$ and $S(t) > 0$, a situation that commonly occurs in clinical trial applications. We consider the implications of (12) separately for uncensored and censored data separately in the next two sections for t such that $\pi(t) > 0$.

3.3.1 Uncensored data

Fleming and Harrington (1984, Table 1) show that Breslow's estimator has a smaller MSE than the Kaplan-Meier estimator whenever the true survival probability $S(t)$ exceeds 1/5. In fact, the point at which the MSEs become equal depends on n, as can be seen in the respective formulas for the exact MSEs given there. For example, with $n = 10$ this difference is negative (i.e., the Breslow estimator beats the Kaplan-Meier estimator) if $S(t) > 0.183$; for $n = 20$, $MSE(\tilde{S}(t)) - MSE(\hat{S}(t)) < 0$ provided $S(t) > 0.186$. Plots of the difference in MSE for increasing n show that the sequence of differences remains negative for $S(t) > 0.2$.

This fact is reflected in (12). In particular, for uncensored data, it is easily shown that for $t < \tau$ and integers $k \geq 1$,

$$\int_0^t \frac{\lambda(u)}{\pi^k(u)}du = \int_0^t \frac{\lambda(u)}{S^k(u)}du = \frac{1}{k}\left(\frac{1}{S^k(t)} - 1\right).$$

Substituting the above formula for $k = 1, 2$ into (12) and simplifying, we obtain

$$\mathrm{MSE}(\tilde{S}(t)) - \mathrm{MSE}(\hat{S}(t)) = \frac{(S(t)-1)(5S(t)-1)}{4n^2} + o(n^{-2}).$$

Evidently, this difference is equal to zero at $S(t) = 1/5$, negative for $S(t) > 1/5$, and positive for $S(t) < 1/5$. This asymptotic analysis further justifies conclusions drawn by Fleming and Harrington (1984).

3.3.2 Censored data

Fleming and Harrington (1984, §3) assert that the Breslow estimator has smaller MSE than the Kaplan-Meier estimator when $S(t) \geq 1/5$ in the presence of censoring. These conclusions are based on Monte Carlo simulation results that assume $S(t) = 1 - t$ (i.e., failure times are distributed as $U(0,1)$) and $G(t) = 1 - t/a$ (i.e., censoring times are distributed as $U(0,a)$) for $a = 1/2, 1$, and 2. The formula (12) clearly shows that a single cutpoint cannot suffice in general since the difference in MSE depends on the censoring distribution as well.

It is very difficult to establish a sharp bound on (12) in the case where the censoring distribution is allowed to be arbitrary. However, a crude bound can be easily established using Cauchy's Inequality. Notice that

$$\left(\int_0^t \frac{\lambda(u)}{\pi(u)}\right)^2 \leq \int_0^t \frac{\lambda(u)}{\pi^2(u)}\,du \int_0^t \lambda(u)\,du\,.$$

Thus,

$$\mathrm{MSE}(\tilde{S}(t)) - \mathrm{MSE}(\hat{S}(t)) \leq \frac{1}{n^2}\left(\frac{3}{4}\Lambda(t) - 1\right)\int_0^t \frac{\lambda(u)}{\pi^2(u)}\,du\,.$$

Hence, for $\Lambda(t) < \frac{4}{3}$ or equivalently $S(t) > e^{-\frac{4}{3}} \approx 0.264$, we have $\mathrm{MSE}(\tilde{S}(t)) - \mathrm{MSE}(\hat{S}(t)) < 0$ (to terms of $o(n^{-2})$) *provided* $G(\Lambda^{-1}(4/3)) > 0$. That is, the Breslow estimator generally has smaller MSE than the Kaplan-Meier estimator when $S(t) > e^{-\frac{4}{3}}$ and $G(t^*) > 0$ for $t^* = \Lambda^{-1}(4/3)$.

References

Andersen, P.K., Borgan, O., Gill, R., and Keiding, N. (1993), *Statistical Models Based on Counting Processes,* New York: Springer-Verlag.

Breslow, N.E. (1972), "Discussion of Professor Cox's paper," *Journal of the Royal Statistical Society, Series B*, 34, 216-217.

Breslow, N.E. and Crowley, J. (1974), "A large sample study of the life table and product-limit estimates under random censorship," *Annals of Statistics*, 2, 437-443.

Fleming, T.R., and Harrington, D.P. (1984), "Nonparametric estimation of the survival distribution in censored data," *Communications in Statistics - Theory and Methods*, 13, 2469-2486.

Fleming, T.R., and Harrington, D.P. (1991), *Counting Processes and Survival Analysis*, NY: John Wiley and Sons.

Gross, S. and Lai, T.L. (1996), "Bootstrap methods for truncated and censored data," *Statistica Sinica*, 6, 509-530.

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

Lai, T.L., and Wang, J.Q. (1993), "Edgeworth expansions for symmetric statistics with applications to bootstrap methods," *Statistica Sinica*, 3, 517-542.

Mykland, P. (1994), "Bartlett identities for martingales," *Annals of Statistics*, 22, 21-38.

Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach.* Springer: New York.

Strawderman, R.L. and Wells, M.T. (1997). "Accurate bootstrap confidence limits for the cumulative hazard and survivor functions under random censoring," *Journal of the American Statistical Association*, 92, 1356-1374.

Appendix: Proof of Lemma 1

We first give a reasonably general form of Itô's formula. In what follows, all processes are defined on a common probability space equipped with a right-continuous filtration. Let X be an adapted process with right-continuous paths having left-hand limits (cadlag) and suppose further that for $t \geq 0$, $X(t) = X(0) + M(t) + A(t)$ where $X(0)$ is finite and $\mathcal{F}_0-$measurable, $M(t)$ is a local martingale with $M(0) = 0$, and $A(t)$ is an adapted cadlag predictable process with $A(0) = 0$ and nondecreasing paths of finite variation on every bounded interval. Define $\Delta X(s) = X(s) - X(s-)$, the optional variation process $[X](s) = X^2(s) - 2\int_0^s X(u-)dX(u)$ with $[X](0) = X^2(0)$, and its continuous part $[X]^c(s) = [X](s) - \sum_{0 \leq s \leq t}(\Delta X(s))^2$.

As defined, X is a semimartingale (Protter, 1990, p. 88). Suppose further that $[X]^c \equiv 0$, or that $[X](s) = \sum_{0 \leq s \leq t}(\Delta X(s))^2$; then, $X(s)$ is a *quadratic pure jump* semimartingale (Protter, 1990, p. 63). Finally, let $f(\cdot)$ be a twice-differentiable function. Then, $f(X)$ is itself a semimartingale, and by the general form of Itô's formula (Protter, 1990, Thm. 32, p. 71),

$$f(X(t)) - f(X(0)) = \int_0^t f'(X(s-))dX(s) + \sum_{0<s\leq t} \{f(X(s)) - T(s,X)\}, \tag{13}$$

where $T(s,X) = f(X(s-)) + f'(X(s-))\Delta X(s)$.

Now, let $X(t) = \int_0^t H(u)d\bar{M}(u)$ where $\bar{M}(t) = \bar{N}(t) - \bar{A}(t)$ is a local square integrable martingale with respect to the filter in (1) and $H(u)$ is a locally bounded predictable process. Since $[X](t) = \int_0^t H^2(s)d\bar{N}(s)$ (see, for example, Andersen *et al.*, 1993, pp. 68-78), $X(t)$ is quadratic pure jump martingale. Moreover, $dX(s) = H(s)d\bar{M}(s)$, and since the compensator $\bar{A}(t) = \int_0^t \bar{Y}(s)\lambda(s)ds$ is continuous, $\Delta X(s) = H(s)\Delta\bar{M}(s) = H(s)\Delta\bar{N}(s)$. Thus, (13) becomes

$$\begin{aligned} f(X(t)) - f(X(0)) = &\int_0^t f'(X(s-))H(s)d\bar{M}(s) \\ &+ \sum_{0<s\leq t} \left\{f(X(s)) - f(X(s-)) - f'(X(s-))H(s)\Delta\bar{N}(s)\right\}. \end{aligned}$$

Since $X(s) = X(s-) + H(s)\Delta\bar{N}(s)$ and $\Delta\bar{N}(s) = 0$ or 1, it follows that $f(X(s)) - f(X(s-)) = [f(X(s-) + H(s)) - f(X(s-))]\,\Delta\bar{N}(s)$ and hence that

$$\sum_{0<s\leq t} \left\{f(X(s)) - f(X(s-)) - f'(X(s-))H(s)\Delta\bar{N}(s)\right\} = \int_0^t D(s)d\bar{N}(s),$$

where $D(s)$ is given in the statement of the theorem. The theorem now follows by making the substitution $d\bar{N}(s) = d\bar{M}(s) + \bar{Y}(s)\lambda(s)ds$ and rearranging terms.