The Annals of Statistics
1981, Vol 9, No 1, 78–92

INFLUENCE FUNCTIONS FOR CENSORED DATA[1]

BY NANCY REID

University of British Columbia

In this paper influence curves for censored data estimators are considered. The influence curve of the Kaplan-Meier estimate of survival time is calculated, and it is shown that the influence curve provides an alternative derivation of the asymptotic variance of this estimate. A chain rule for influence curves is established and is used to calculate the influence curves of censored data estimators that are functions of the Kaplan-Meier estimate. The robustness properties and asymptotic variances of these estimates follow directly. Some examples of this approach to calculating variances are given. In particular, it is shown how the theory developed for M-estimation and L-estimation can be extended to the censored data case. The necessary differentiability conditions are verified in the appendix.

1. Introduction. Let $X_1^\circ, X_2^\circ, \ldots, X_n^\circ$ be independent, identically distributed random variables with distribution function F°. These are the true lifetimes of the items under observation. Associated with each X_i° is an independent censoring variable Y_i, and $Y_1, Y_2, \ldots, Y_n$ are assumed independent, identically distributed with distribution function G. The observations are the n pairs $(X_1, \delta_1), \ldots, (X_n, \delta_n)$, where $X_i = \min(X_i^\circ, Y_i)$ and $\delta_i = 1\{X_i = X_i^\circ\}$. (Throughout, $1\{A\}$ is the indicator function for the event A.) The distribution function of the X_i's is F and satisfies $1 - F(t) = [1 - F^\circ(t)][1 - G(t)]$. Two subdistribution functions F^u and F^c are defined as

$$F^u(t) = \Pr\{X_i \leq t, \delta_i = 1\},$$

and

$$F^c(t) = \Pr\{X_i \leq t, \delta_i = 0\}.$$

The following relations hold:

(i) $dF^u(t) = [1 - G(t)]\, dF^\circ(t)$
(ii) $F^u(\infty) = \Pr\{\delta_i = 1\}$, $F^c(\infty) = \Pr\{\delta_i = 0\}$
(iii) $F^u(t) + F^c(t) = F(t)$.

Corresponding to each of these distribution functions is an empirical distribution function based on the n observation pairs. These empirical cdf's are indicated by a subscript n. For example,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq t\},$$

and

$$F_n^c(t) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq t, \delta_i = 0\}.$$

Because the random variables X_i are thought of as lifetimes, it will be convenient to have additional notation for the cumulative survival functions corresponding to the cumulative

Received July 1979; revised December 1979.

[1] Research supported by the National Research Council of Canada and the Department of Statistics of Stanford University.

AMS 1970 *subject classifications.* Primary 62G35; secondary 62G05.

Key words and phrases. Influence function, censored data, Kaplan-Meier estimate, robust estimation, differential, M-estimate, L-estimate.

distribution functions. For future reference, $S(t) = 1 - F(t)$ and $S^\circ(t) = 1 - F^\circ(t)$. The subsurvival functions are $S^u(t) = \Pr\{X_i > t, \delta_i = 1\}$ and $S^c(t) = \Pr\{X_i > t, \delta_i = 0\}$. The empirical survival functions are subscripted with an n.

Kaplan and Meier (1958) suggested estimating the conditional probability of failure at time t by the observed proportion of failures at time t, and combining these estimates in the usual manner to obtain an estimate of the underlying survival distribution S°. This gives the Kaplan-Meier estimate, denoted $\hat{S}^\circ$, and defined

$$\begin{aligned} \hat{S}^\circ(t) = 1 - \hat{F}^\circ(t) &= \textstyle\prod_{i:X_{(i)}\le t} \left(\dfrac{n-i}{n-i+1}\right)^{\delta_{(i)}}, \qquad t \le X_{(n)} \\ &= 0, \qquad\qquad\qquad\qquad\qquad t > X_{(n)}, \end{aligned}$$

where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ and $\delta_{(i)} = 1\{X_{(i)} \text{ uncensored}\}$. By convention, uncensored observations are ranked ahead of censored observations with which they are tied.) Note that by defining $\hat{S}^\circ(t)$ to be 0 for $t > X_{(n)}$, we are treating the largest observation as uncensored, whether or not it is.

The properties of the Kaplan-Meier estimate have been studied by Kaplan and Meier (1958), Efron (1967), and Breslow and Crowley (1974). In particular, it is the maximum likelihood estimate, it is strongly consistent, and asymptotically normal. Regarded as a stochastic process, $\{\hat{S}^\circ(t); t \ge 0\}$ converges weakly to a Gaussian process. In the presence of no censoring, $\hat{S}^\circ(t)$ reduces to the usual empirical cumulative survival function. Efron (1967) formulated the random censorship model used here, and Breslow and Crowley (1974) exploited this formulation in proving weak convergence. Meier (1975) established weak convergence of $\hat{S}^\circ(\cdot)$ when the censoring variables are arbitrary unknown constants. The Kaplan-Meier estimate has been extended to the problem of competing risks by Peterson (1975) and Aalen (1976).

The role of censored and uncensored observations in the construction of $\hat{S}^\circ(t)$ can be clarified by a representation of the estimate in terms of the empirical subsurvival functions $S_n^u(t)$ and $S_n^c(t)$. This representation is due to Peterson (1977). He showed that the true survival function $\hat{S}^\circ(t)$ can be expressed as a functional of the two subsurvival functions:

$$S^\circ(t) = \exp \int_0^t \frac{dS^u(s)}{(S^u + S^c)(s)} \times \exp \textstyle\sum_{s\le t} \ln \dfrac{S^u(s^+) + S^c(s^+)}{S^u(s^-) + S^c(s^-)}. \tag{1.1}$$

The region of integration is the union of open intervals of points less than t for which $\hat{S}^u(\cdot)$ is continuous, and the summation is over points s which are points of discontinuity of $\hat{S}^u(\cdot)$. If $\hat{S}^u(\cdot)$ is wholly continuous the second factor is identically 1, and if $\hat{S}^u(\cdot)$ is wholly discrete the first factor is identically 1. The relation (1.1) is a generalization of the relation

$$\hat{S}^\circ(t) = \exp - \int_0^t \frac{dF^\circ(s)}{1 - F^\circ(s)}.$$

Peterson also shows that the Kaplan-Meier estimate satisfies

$$\hat{S}^\circ(t) = \exp \int_0^t \frac{dS_n^u(s)}{(S_n^u + S_n^c)} \times \exp \textstyle\sum_{s\le t} \ln \dfrac{S_n^u(s^+) + S_n^c(s^+)}{S_n^u(s^-) + S_n^c(s^-)}. \tag{1.2}$$

(Even though the first term is 1, we leave it in for later calculations.) This functional form of $\hat{S}^\circ(t)$ will be the form suitable for deriving the influence curve.

The influence curve of a statistic regarded as a functional is the first derivative of the functional evaluated at some point in the space of distribution functions. Differentiation of statistical functionals was originally proposed by von Mises in 1947, and a von Mises statistic is a functional sufficiently regular to have a series expansion in functional derivatives. For a thorough study of von Mises expansions, see Reeds (1976).

If T is a von Mises functional,

$$(1.3)\quad T(G) = T(F + G - F) = T(F) + \int \mathrm{IC}(T, F; y)\, d(G - F)(y) + \text{higher order terms.}$$

If the distribution function G is sufficiently close to F, the behavior of $T(G)$ may be described by the behavior of the first two terms in (1.3). This is the basis for the usefulness of influence curves in calculating asymptotic distributions. Substituting F_n for G in (1.3) we have

$$(1.4)\qquad T(F_n) = T(F) + \frac{1}{n} \textstyle\sum_{i=1}^n \mathrm{IC}(T, F; X_i) - \displaystyle\int \mathrm{IC}(T, F; y)\, dF(y) + \text{h.o.t.}$$

The higher order terms are $o_p(1/\sqrt{n})$ because $F_n - F$ is of stochastic order $1/\sqrt{n}$. The random variables $\mathrm{IC}(T, F; X_i)$ are independent and identically distributed with mean $\mu = \int \mathrm{IC}(T, F; y)\, dF(y)$ and variance $\int (\mathrm{IC}(T, F; y) - \mu)^2\, dF(y)$. It follows from the central limit theorem that

$$\sqrt{n}\,(T(F_n) - T(F)) \to_d N(0, \int (\mathrm{IC}(T, F; y) - \mu)^2\, dF(y)).$$

Considering (1.3) as an expansion of $T(F + \epsilon(G - F))$ evaluated at $\epsilon = 1$, about the point $\epsilon = 0$, we see that

$$(1.5)\qquad \frac{d}{d\epsilon} T(F + \epsilon(G - F)) \Big|_{\epsilon=0} = \int \mathrm{IC}(T, F; y)\, d(G - F)(y),$$

which gives us a simple method for calculating influence curves. (Some authors (e.g., Andrews, et al., 1972, Huber 1977), define the influence curve to have mean zero. For our purposes however the above definition is more convenient.)

Hampel (1974) exploited the use of influence curves as a tool in robust estimation. If the distribution function G puts all its mass at the point x, $G(y) = \delta_x(y)$, then

$$\begin{aligned}\frac{d}{d\epsilon} T(F + \epsilon(G - F)) \Big|_{\epsilon=0} &= \frac{d}{d\epsilon} T((1 - \epsilon)F + \epsilon\delta_x) \Big|_{\epsilon=0} \\ &= \int \mathrm{IC}(T, F; y)\, d(\delta_x - F)(y) \\ &= \mathrm{IC}(T, F; x) - \mu.\end{aligned}$$

This derivative measures the effect on the functional T of a small (infinitesimal) change in the weight the distribution function F gives to the point x, that is, the "influence" on the statistic of an additional observation at the point x. The shape of the influence curve provides information about the robustness of a statistic. The sample mean is sensitive to large observations and this is reflected in the fact that the influence curve is unbounded. In contrast to this, the influence curve of the median is a step function. A statistical functional with a bounded influence curve is not sensitive to extreme observations, so is robust in this sense. If the influence curve $\mathrm{IC}(T, F; x)$ is continuous in F, the statistic T is robust to departures from assumptions about the underlying form of F.' Such departures are often modeled as a "contaminated" distribution; the true underlying distribution is $F_1 = F + \epsilon H$. Robust statistics will perform well for such distributions. For a thorough discussion of the uses of the influence curve in robust estimation, the reader is directed to Hampel (1974) and Huber (1977).

The statistic $T(F)$ need not be a functional of only one distribution function. The differential T is merely an operator on the appropriate space of functions, and can be defined for functionals $T(F, u)$, $u \in \mathbb{R}$ (see Reeds, 1976, Section 1.6), or for bivariate functionals $T(F_1, F_2)$. The bivariate von Mises expansion is

$$\begin{aligned}(1.6)\quad T(G_1, G_2) = T(F_1, F_2) &+ \int \mathrm{IC}_1(T, F_1, F_2; y)\, d(G_1 - F_1)(y) \\ &+ \int \mathrm{IC}_2(T, F_1, F_2; y)\, d(G_2 - F_2)(y) + \text{higher order terms.}\end{aligned}$$

The two influence curves are defined by

$$\frac{\partial}{\partial \epsilon} T(F_1 + \epsilon(G_1 - F_1), F_2 + \delta(G_2 - F_2))\Big|_{\epsilon=0;\,\delta=0} = \int \mathrm{IC}_1(T, F_1, F_2; y)\, d(G_1 - F_1)(y)$$

$$\frac{\partial}{\partial \delta} T(F_1 + \epsilon(G_1 - F_1), F_2 + \delta(G_2 - F_2))\Big|_{\epsilon=0;\,\delta=0} = \int \mathrm{IC}_2(T, F_1, F_2; y)\, d(G_2 - f_2)(y).$$

Lambert (1979) has used bivariate functionals for calculating the influence functions of some two-sample estimates.

2. The influence curve of the Kaplan-Meier estimate. The Kaplan-Meier estimate $\hat{S}^\circ(t)$ jumps only at uncensored observations. The size of the jump at each uncensored observation is a function of the number of observations and the pattern of losses occurring before that failure. An additional observation will change all the jump sizes of the estimate. If the new observation is uncensored, an extra jump will be introduced into $\hat{S}^\circ(\cdot)$. However, if the new observation is censored, no such jump will be added. It is this essential difference in the effect of new observations that makes it natural to consider two influence curves, i.e., partial functional derivatives. We have already seen that $\hat{S}^\circ(t)$ can be represented as a bivariate functional of the empirical subsurvival functions of censored and uncensored observations.

To calculate the influence of $\hat{S}^\circ(t)$, we first consider the influence curve of $\hat{\Lambda}^\circ(t)$, defined as $\hat{\Lambda}^\circ(t) = -\ln \hat{S}^\circ(t)$. Writing $\hat{\Lambda}^\circ(t)$ as a bivariate functional, we have

$$\hat{\Lambda}^\circ(t) = T(S_n^u, S_n^c, t) = -\int_0^t \frac{dS_n^u(s)}{(S_n^u + S_n^c)(s)} + \textstyle\sum_{s\le t} -\ln\dfrac{S_n^u(s^+) + S_n^c(s^+)}{S_n^u(s^-) + S_n^c(s^-)}$$

and the corresponding functional for $\Lambda^\circ(t)$

$$T(S^u, S^c, t) = -\int_0^t \frac{dS^u(s)}{(S^u + S^c)} - \textstyle\sum_{s\le t} \ln\dfrac{S^u(s^+) + S^c(s^+)}{S^u(s^-) + S^c(s^-)}.$$

The functions S_n^u *and* S_n^c are discrete, so the first term of $T(S_n^u, S_n^c, \cdot)$ is zero. We assume that the true subsurvival functions are continuous, so the second term of $T(S^u, S^c, \cdot)$ is zero.

For the functions F_1 and G_1 of equation (1.7) we substitute F^u, the true subdistribution function of the uncensored observations, and F_n^u the corresponding empirical subdistribution function. Similarly F_2 and G_2 are replaced by F^c and F_n^c. Because there is a one-to-one relationship between the subdistribution and the corresponding subsurvival functions, we continue to use the more concise notation S^u, S_n^u, S^c, S_n^c.

$$\begin{aligned} &T(S^u + \epsilon(S_n^u - S^u), S^c + \delta(S_n^c - S^c), t) \\ &\quad = -\int_0^t \frac{d[(1-\epsilon)S^u(s)]}{[S^u + \epsilon(S_n^u - S^u)](s) + [S^c + \delta(S_n^c - S^c)](s)} \\ &\qquad - \textstyle\sum_{s\le t} \ln \dfrac{[S^u + \epsilon(S_n^u - S^u)](s^+) + [S^c + \delta(S_n^c - S^c)](s^+)}{[S^u + \epsilon(S_n^u - S^u)](s^-) + [S^c + \delta(S_n^c - S^c)](s^-)}. \end{aligned}$$

Here the integration is over $0 \le s \le t$ because we assume S^u is continuous and the summation is over jump points of S_n^u, which is discrete.

Now

$$\begin{aligned} &\frac{\partial}{\partial \epsilon} T(S^u + \epsilon(S_n^u - S^u), S^c + \delta(S_n^c - S^c), t)\Big|_{(0,0)} \\ &\quad = \int_0^t \frac{dS^u(s)}{S^u(s) + S^c(s)} + \int_0^t \frac{(S_n^u - S^u)(s)\, dS^u(s)}{[S^u(s) + S^c(s)]^2} - \textstyle\sum_{s\le t} \dfrac{S_n^u(s^+) - S_n^u(s^-)}{S^u(s) + S^c(s)}. \end{aligned}$$

(We have used the fact that S^u is continuous.) A similar calculation gives

$$\frac{\partial}{\partial \delta} T(S^u + \epsilon(S_n^u - S^u), S^c + \delta(S_n^c - S^c), t)\,|_{(0,0)} = \int_0^t \frac{(S_n^c - S^c)(s)\, dS^u(s)}{[S^u(s) + S^c(s)]^2}.$$

Writing $-(S_n^u(s) - S^u(s)) = \int_s^\infty d(S_n^u - S^u)(u)$,

$$\int_0^t \frac{[S_n^u(s) - S^u(s)]\, dS^u(s)}{(S^u + S^c)^2(s)} = -\int_0^\infty \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}\, d(S_n^u - S^u)(s),$$

where $s \wedge t = \min(s, t)$. Also

$$-\sum_{s \le t} \frac{S_n^u(s^+) - S_n^u(s^-)}{S^u(s) + S^c(s)} = -\int_0^t \frac{d(S_n^u - S^u)(s)}{S^u(s) + S^c(s)} - \int_0^t \frac{dS^u(s)}{S^u(s) + S^c(s)}.$$

We conclude, using integral representation (1.7), and recalling that $d(S_n^u - S^u)(s) = -d(F_n^u - F^u)(s)$,

(2.1)
$$\mathrm{IC}_1(T, S^u, S^c; s)(t) = \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le t\}}{(S^u + S^c)(s)}$$
$$\mathrm{IC}_2(T, S^u, S^u; s)(t) = \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}.$$

To find the influence curves for the Kaplan-Meier estimate, we write $\hat{S}^\circ(t)$ as the functional $T^\circ(S_n^u, S_n^c, t) = \exp - T(S_n^u, S_n^c, t)$ (because $\hat{S}^\circ(t) = \exp - \hat{\Lambda}^\circ(t)$). Then

(2.2)
$$\mathrm{IC}_1(T^\circ, S^u, S^c; s)(t) = S^\circ(t)\, \mathrm{IC}_1(T, S^u, S^c; s)(t)$$
$$\mathrm{IC}_2(T^\circ, S^u, S^c; s)(t) = S^\circ(t) \mathrm{IC}_2(T, S^u, S^c; s)(t)$$

Certain features of the Kaplan-Meier estimate are reflected in (2.1). The first term of $\mathrm{IC}_1(T, S^u, S^c; s)(t)$ represents the change in the size of each jump in $\hat{\Lambda}^\circ$ when a new observation is added to the sample, and the second term in $\mathrm{IC}_1(T, S^u, S^c; s)(t)$ represents the additional jump introduced when the added observation is uncensored. The influence curve is constant for $s > t$ because a new observation at $s > t$ affects $\hat{\Lambda}^\circ(t)$ only through a change in sample size. The function $\int_0^{s \wedge t} (dS(u)/(S^u + S^c)^2(u))$ is decreasing in s until $s = t$. Note that, for $s < t$

$$\int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le t\}}{(S^u + S^c)(s)}$$
$$= \int_0^s \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1}{(S^u + S^c)(s)} \ge \int_0^s \frac{d(S^u + S^c)(u)}{(S^u + S^c)^2(u)} + \frac{1}{(S^u + S^c)(s)} = 1$$

so $\mathrm{IC}_1(T, S^u, S^c; s)(t)$ is ≥ 1 for $s < t$, and is increasing in s. If both s and t are large, with $s < t$, introducing a new observation at the point s has a large effect on the cumulative hazard at the point t. This is not translated into a large effect on $\hat{S}^\circ(t)$, however, because $S^\circ(t)$ decreases to 0 as $t \to \infty$. Figure (2.1) sketches $\mathrm{IC}_1(T, S^u, S^c; s)(t)$ as functions of s for fixed t, when the underlying distribution F° is exponential. The censoring distribution G is uniform in Figure 2.1.a and exponential in Figure 2.1.b.

In order to calculate the asymptotic variance of $\hat{\Lambda}^\circ(t)$, we substitute expressions (2.1) into the asymptotic expansion (1.6).

$$T(S_n^u, S_n^c, t) - T(S^u, S^c, t) = \int IC_1(T, S^u, S^c; s)(t)\, d(S_n^u - S^u)(s)$$

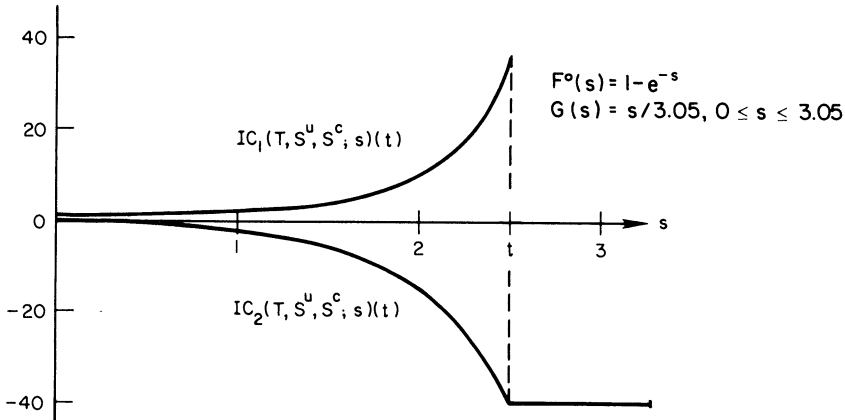

Figure 2.1.a

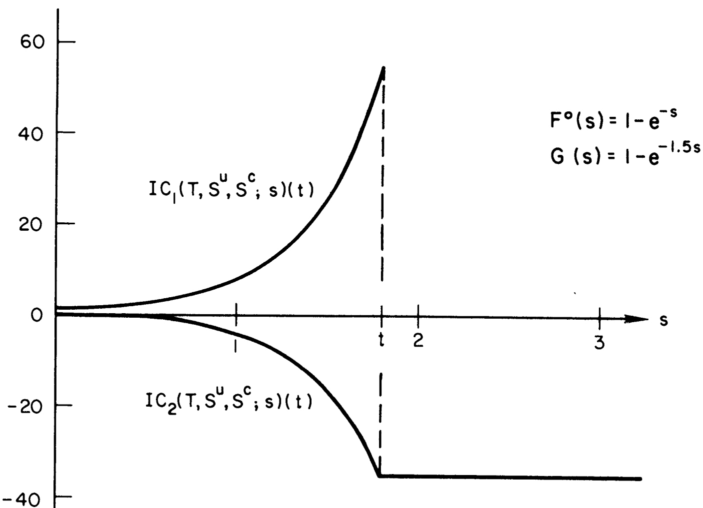

Figure 2.1.b

$$+ \int IC_2(T, S^u, S^c; s)(t)\, d(S_n^c - S^c)(s) + \text{higher order terms}$$

$$= -\frac{1}{n} \sum_{i=1;\,\delta_i=1}^{n} IC_1(T, S^u, S^c; X_i)(t) - \frac{1}{n} \sum_{i=1;\,\delta_i=1}^{n} IC_2(T, S^u, S^c; X_i)(t)$$

$$- \int IC_1(T, S^u, S^c; s)(t)\, dS^u(s) - \int IC_2(T, S^u, S^c; s)(t)\, dS^c(s) + \text{h.o.t.}$$

The third and fourth terms of this expansion are the bivariate analogue of the term $\mu = \int IC(T, F; x)\, dF(x)$ in (1.4). We first show that for each fixed t, the sum of these two terms is zero. From (2.1)

$$\int IC_1\, dS^u(s) + \int IC_2\, dS^c(s) = \int_0^\infty \left(\int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} \right) d(S^u + S^c)(s) + \int_0^t \frac{dS^u(s)}{(S^u + S^c)(s)}$$

$$= \int_0^t \int_0^\infty [d(S^u + S^c)(s)] \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \int_0^t \frac{dS^u(s)}{(S^u + S^c)(s)}$$

$$= \int_0^t \frac{-dS^u(u)}{(S^u + S^c)(u)} + \int_0^t \frac{dS^u(s)}{(S^u + S^c)(s)} = 0.$$

Then, disregarding the higher order terms,

$E\{T(S_n^u, S_n^c, t) - T(S^u, S^c, t)\}^2$

$$\simeq E\left[\frac{1}{n}\sum_{i=1;\delta_i=1}^n \mathrm{IC}_1(T, S^u, S^c; X_i)(t) + \frac{1}{n}\sum_{i=1;\delta_i=1}^n \mathrm{IC}_2(T, S^u, S^c; X_i)(t)\right]^2$$

$$= \frac{1}{n} IC_1^2(T, S^u, S^c; s)(t)\, dF^u(s) + \frac{1}{n}\int \mathrm{IC}_2^2(T, S^u, S^c; s)(t)\, dF^c(s).$$

The limit of $E(T(S_n^u, S_n^c, t) - T(S^u, S^c, t))^2$ as $n \to \infty$ gives the asymptotic variance of the estimate.

$n \times$ A. Var $T(S_n^u, S_n^c, t) = n \times$ A. Var $\hat{\Lambda}^\circ(t)$

$$= \int_0^\infty \left\{\frac{1\{s \le t\}}{(S^u + S^c)^2(s)} + \left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}\right)^2 + 2\frac{1\{s \le t\}}{(S^u + S^c)(s)}\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}\right\} dF^u(s)$$

$$+ \int_0^\infty \left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}\right)^2 dF^c(s)$$

$$= \int_0^t \frac{dF^u(s)}{(S^u + S^c)^2(s)} + \int_0^\infty \left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)}\right)^2 d(F^u + F^c)(s)$$

$$+ 2\int_0^\infty \frac{1\{s \le t\}}{(S^u + S^c)(s)}\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} dF^u(s).$$

The third term of this expression equals the negative of the second term. This is established by integrating the second term by parts (see Miller, 1975). Recalling from Section 1 that $dF^u(s) = [1 - G(s)]\, dF^\circ(s)$ and $S^u(s) + S^c(s) = 1 - F(s) = [1 - G(s)][1 - F^\circ(s)]$, we see that the asymptotic variance of $\hat{\Lambda}^\circ(t)$ is

$$\frac{1}{n}\int_0^t \frac{dF^\circ(s)}{[1 - F(s)][1 - F^\circ(s)]},$$

which is equation (7.11) of Breslow and Crowley (1974).

The above calculation appears in Miller (1975) in connection with the use of the jackknife for the Kaplan-Meier estimate. The jackknife and the influence curve approaches lead to the same asymptotic calculation because the jackknife is a finite sample approximation to the influence function. This approximation is defined by $\mathrm{IC}(\hat{\theta}; X_i) = \hat{\theta} - \tilde{\theta}_i$, where $\hat{\theta} = T(F_n)$ is the parameter estimate and $\tilde{\theta}_i$ is the ith pseudovalue. If $\mathrm{IC}(T(F_n); y)$ converges to $\mathrm{IC}(T, F; y)$, then the jackknife estimate of variance converges to the asymptotic variance given by the influence function approach. For a fuller discussion of this point, the reader is referred to Miller (1978) or Miller (1974). The function $\mathrm{IC}(T(F_n); y)$ is also related to Tukey's sensitivity curve, discussed, for example, in Huber (1977).

It is not possible to conclude that $\hat{\Lambda}^\circ(t)$ is asymptotically normal with the above variance without examining the regularity conditions under which "higher order terms" in expression (1.6) converge to zero as $n \to \infty$. The first condition is that the functional $T(S^u, S^c, \cdot)$ be Frechet differentiable in each argument S^u and S^c. In order to check Frechet differentiability it is necessary to specify a norm on the space of subdistribution functions. The second condition is that S_n^u and S_n^c converge to S^u and S^c in this norm, at stochastic rate $1/\sqrt{n}$. These regularity conditions are verified in the Appendix.

3. Functions of the Kaplan-Meier estimate. Parameters of the underlying distribution function F° can be estimated using $\hat{F}^\circ$. For example, Kaplan and Meier (1958) suggested estimating the mean of F° by $\hat{\mu}^{\mathrm{KM}} = \int x \, d\hat{F}^\circ(x)$. The influence curves for the functionals $T^\circ(S^u, S^c; t) = 1 - \hat{F}^\circ(t)$ and $T_1(G) = \int x \, dG(x)$ are known. In this section we establish a chain rule for influence curves, enabling us to find the influence curve and asymptotic variance of $\hat{\mu}_{\mathrm{KM}}$, and other estimators based on $\hat{F}^\circ(t)$.

The chain rule follows directly from the chain rule for differential operators. Let $\mathscr{B}$ be the space of bounded positive measures. We suppose that there are two functionals, $T_1\colon \mathscr{B} \to \mathbb{R}$ and $T_2\colon \mathscr{B} \to \mathscr{B}$. Then $V = T_1(T_2)$ is a functional from $\mathscr{B}$ to $\mathbb{R}$. The differential of T_1, denoted dT_1, maps $\mathscr{B}$ to $\mathbb{R}$ and is defined by

$$dT_1|_{F} \circ \mu = \int \mathrm{IC}(T_1, F; t)\mu\{dt\}.$$

Here $|_F$ means evaluated at the point F. Similarly dT_2 maps $\mathscr{B}$ to $\mathscr{B}$ and is defined by

$$(dT_2|_G \circ \nu)(t) = \int \mathrm{IC}(T_2, G; s)(t)\nu\{ds\}.$$

By the chain rule, $dV|_G = dT_1|_{T_2(G)} \circ dT_2|_G$. Substituting, we find

$$\begin{aligned} dV|_G &= \int \mathrm{IC}(T_1, T_2(G); t)(dT_2|_G \circ \lambda)\{dt\} \\ &= \int \mathrm{IC}(T_1, T_2(G); t) \int \mathrm{IC}(T_2, G; s)(\{dt\})\lambda\{ds\} \\ &= \int\int \mathrm{IC}(T_1, T_2(G); t)\mathrm{IC}(T_2, G; s)(\{dt\})\lambda\{ds\} \\ &\triangleq \int \mathrm{IC}(V, G; s)\lambda\{ds\}. \end{aligned}$$

Hence

$$\mathrm{IC}(V, G; s) = \int \mathrm{IC}(T_1, T_2(G); t)\mathrm{IC}(T_2, G; s)(\{dt\}).$$

The extension to bivariate functionals is again straightforward. If $T_2\colon \mathscr{B} \times \mathscr{B} \to \mathscr{B}$ then $V = T_1(T_2)$ maps $\mathscr{B} \times \mathscr{B}$ to $\mathbb{R}$ and has two influence functions:

$$\begin{aligned} \mathrm{IC}_1(V, F_1, F_2; s) &= \int \mathrm{IC}(T_1, T_2(F_1, F_2); t)\mathrm{IC}_1(T_2, F_1, F_2; s)(\{dt\}) \\ \mathrm{IC}_2(V, F_1, F_2; s) &= \int \mathrm{IC}(T_1, T_2(F_1, F_2); t)\mathrm{IC}_2(T_2, F_1, F_2; s)(\{dt\}). \end{aligned} \tag{3.1}$$

Integrating (3.1) by parts we obtain

$$\mathrm{IC}_i(V, F_1, F_2; s) = \int \mathrm{IC}_i(T_2, F_1, F_2; s)(t)\left[\frac{d}{dt}\mathrm{IC}(T_1, T_2; t)\right] dt, \qquad i = 1, 2, \tag{3.2}$$

which is valid as long as $\mathrm{IC}_i(T_2, F_1, F_2; s)(t) \times \mathrm{IC}(T_1, T_2; t)$ vanishes at $t = -\infty$ and $t = \infty$.

Using the relations (3.2) we can calculate the asymptotic variance of any functional of the Kaplan-Meier estimate as long as the influence curve of that functional is known. In this case $\hat{S}^\circ(t) = T^\circ(S^u_n, S^c_n, t)$ and $V(S^u_n, S^c_n) = T_1(\hat{S}^\circ) = T_1(T^\circ(S^u_n, S^c_n, \cdot))$. Let $g(t) = (d/dt)$ $\mathrm{IC}(T_1, T_2; t)$. From (2.2) we see that $\mathrm{IC}_i(T^\circ, S^u, S^c; S)(t)$ is zero at both endpoints $t = 0$ and $t = \infty$. Then

$$\mathrm{IC}_1(V, S^u, S^c; s) = \int_0^\infty S^\circ(t)\left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le t\}}{(S^u + S^c)(s)}\right) g(t)\, dt \tag{3.3}$$

$$\mathrm{IC}_2(V, S^u, S^c; s) = \int_0^\infty S^\circ(t)\left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(s)}\right) g(t)\, dt.$$

From the asymptotic expansion (1.6),

$$n \times \text{A. Var. } V = \int_0^\infty \mathrm{IC}_1^2(V, S^u, S^c; s)\, dF^u(s) + \int_0^\infty \mathrm{IC}_2^2(V, S^u, S^c; s)\, dF^c(s)$$

$$= \int_0^\infty \left\{\int_0^\infty S^\circ(t)\left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le t\}}{(S^u + S^c)(s)}\right) g(t)\, dt\right\}^2 dF^u(s)$$

$$+ \int_0^\infty \left\{\int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} g(t)\, dt\right\}^2 dF^c(s)$$

$$\tag{3.4} = \int_0^\infty \left\{\int_0^\infty S^\circ(t) \frac{1\{s \le t\}}{(S^u + S^c)(s)} g(t)\, dt\right\}^2 dF^u(s)$$

$$+ 2\int_0^\infty \left\{\int_0^\infty S^\circ(t) \frac{1\{s \le t\}}{(S^u + S^c)(s)} g(t)\, dt\right\}$$

$$\cdot \left\{\int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} g(t)\, dt\right\} dF^u(s)$$

$$+ \int_0^\infty \left\{\int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} g(t)\, dt\right\}^2 d(F^u + F^c)(s).$$

We write the third term of this expression as $\int x dy$ where

$$x = \left(\int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u}{(S^u + S^c)^2} g(t)\, dt\right)^2$$

and

$$y = -(S^u + S^c)(s),$$

and integrate by parts. The third term of expression (3.4) is

$$-\left(\int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} g(t)\, dt\right)^2 (S^u + S^c)(s)\Big|_{s=0}^{s=\infty}$$

$$+ 2\int_0^\infty (S^u + S^c)(s) \int_0^\infty S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} g(t)\, dt$$

$$\cdot \int_0^\infty S^\circ(t) \frac{1\{s \le t\}}{(S^u + S^c)^2(s)} g(t)\, dt\, dS^u(s)$$

$$= 2\int_0^\infty \left\{\int_0^\infty (S^\circ(t) \int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(s)} g(t)\, dt\right\}$$

$$\cdot \left\{\int_0^\infty S^\circ(t) \frac{1\{s \le t\}}{(S^u + S^c)(s)} g(t)\, dt\right\} dS^u(s).$$

This is precisely the negative of the second term, and we are left with a relatively simple expression for the asymptotic variance of any functional of the Kaplan-Meier estimate:

$$\tag{3.5} \text{A. Var. } V(\hat{S}^\circ) = \frac{1}{n}\int_0^\infty \frac{1}{[1 - F(s)]^2}\left\{\int_s^\infty S^\circ(t)\, g(t)\, dt\right\}^2 dF^u(s).$$

The influence function approach can be used to prove the asymptotic normality of V. The functional V will be Frechet differentiable when T_1 and T° are Frechet differentiable. (In fact, to prove asymptotic normality it is sufficient that V be compactly differentiable, a slightly weaker requirement.) Differentiability of T° is discussed in the Appendix. The functionals T_1 in all the examples that follow are known to be differentiable. (See, for example, Reeds, 1976, Chapters 5 and 6.)

EXAMPLE 1. The Kaplan-Meier mean: $\hat{\mu}^{\text{KM}} = \int_0^T \hat{S}^\circ(t)\, dt$. The appropriate functional is $T_1(F) = \int^T t dF(t)$, with influence curve $\text{IC}(T_1, F; t) = t\, 1\{t \le T\}$. Writing $V(S_n^u, S_n^c) = T_1(T^\circ(S_n^u, S_n^c, \cdot))$, it follows that

$$\text{IC}_1(V, S^u, S^c; s) = \int_0^T S^\circ(t) \left\{ \frac{1\{s \le t\}}{(S^u + S^c)(s)} + \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} \right\} dt;$$

the influence curve of the mean is the mean of the influence curve. From (3.5)

$$\text{A. Var}\, \hat{\mu}^{\text{KM}} = \frac{1}{n} \int_0^T \frac{1}{[1 - F(s)]^2} \left(\int_s^T S^\circ(t)\, dt \right)^2 dF^u(s),$$

which (except for the upper limit T) is the result given by Breslow and Crowley (1974, equation 8.2). The upper limit T must satisfy $T < \infty$ and $S(T) > 0$. The condition $T < \infty$ is required for the function $(d/dt)\text{IC}(T_1, T^\circ; t)$ of (3.2) to be integrable and the condition $S(T) > 0$ is required in the proof of differentiability of the functional T°. Susarla and Van Ryzin (1979) have generalized the result to $\hat{\mu}_n = \int_0^{M_n} \hat{S}^\circ(t)\, dt$, where $M_n \uparrow \infty$ as $n \to \infty$.

EXAMPLE 2. The Kaplan-Meier median: $\hat{m}^{\text{KM}} = \inf\{m: \hat{F}^\circ(m) \ge 1/2\}$. Let m be the median of the underlying distribution F° which we suppose has a density f°. The appropriate functional $T_1(G)$ is defined by $G(T_1(G)) = 1/2$. The influence curve is an indicator function and

$$g(t) = \frac{\delta(t - m)}{f^\circ(m)},$$

where $\delta(x)$ is the Dirac delta function. Writing $V(S_n^u, S_n^c) = \hat{m}^{\text{KM}}$,

$$\begin{aligned}\text{IC}_1(V, S^u, S^c; s) &= \frac{S^\circ(m)}{f^\circ(m)} \left\{ \int_0^{s \wedge m} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le m\}}{(S^u + S^c)(s)} \right\} \\ &= \frac{1}{2f^\circ(m)} \text{IC}_1(T^\circ, S^u, S^c; s)(m)\end{aligned}$$

and

$$\begin{aligned}\text{A. Var}\, \hat{m}^{\text{KM}} &= \frac{1}{n} \int_0^\infty \frac{1}{[1 - F(s)]^2} \left\{ \int_s^\infty \frac{S^\circ(t)\delta(t - m)}{f^\circ(m)} dt \right\}^2 dF^u(s) \\ &= \frac{1}{n} \int_0^m \frac{1}{[1 - F(s)]^2} \frac{[S^\circ(m)]^2}{[f^\circ(m)]^2} dF^u(s) \\ &= \frac{1}{n} \frac{1}{4[f^\circ(m)]^2} \int_0^m \frac{dF^u(s)}{[1 - F(s)]^2}.\end{aligned}$$

This formula was derived by a different method in Sander (1975a).

EXAMPLE 3. Kaplan-Meier L-estimators. An L-estimate is a linear combination of order statistics $\hat{\theta} = (1/n) \sum_{i=1}^n a_{in} X_{(i)}$, where $a_{ni} = J(i/n + 1)$ for some bounded differentiable function $J(\cdot)$ on [0, 1] satisfying $J(0) = 0$. The estimate $\hat{\theta}$ can be written as a functional of the empirical cdf as $T_1(F_n)\ \int x J(F_n(x))\, dF_n(x)$. The corresponding Kaplan-Meier L-estimate is defined as $T_1(\hat{F}^\circ)$. The influence function for this functional is

$$\mathrm{IC}(T_1, F; t) = \int_0^t J(F(s))\, ds,$$

so

$$g(t) = J(F(t)).$$

The functional T_1 has been shown to be Frechet differentiable by Boos (1979). The conditions Boos imposes on J are weaker than those given above.

Defining $V(S_n^u, S_n^c) = T_1(T^\circ(S_n^u, S_n^c))$, we have

$$\mathrm{IC}_1(V, S^u, S^c; s) = \int_0^T S^\circ(t)J(F^\circ(t))\left\{\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1[s \le t]}{(S^u + S^c)(s)}\right\} dt$$

$$\mathrm{A.\ Var}\ V = \frac{1}{n}\int_0^\infty \frac{1}{[1 - F(s)]^2}\left\{\int_s^T S^\circ(t)J(F^\circ(t))\, dt\right\}^2 dF^u(s)$$

$$(3.6) \qquad = \frac{1}{n}\int_0^\infty \frac{1}{[1 - F(s)]^2}\int_0^T\int_0^T S^\circ(t)J(F^\circ(t))S^\circ(u)J(F^\circ(u)) \cdot 1\{s \le t\}\, 1\{s \le u\}\, dt\, du\, dF^u(s)$$

$$= \frac{1}{n}\int_0^T\int_0^T S^\circ(t)J(F^\circ(t))S^\circ(u)J(F^\circ(u))\int_0^{u\wedge t} \frac{dF^u(s)}{[1 - F(s)]^2}\, dt\, du.$$

As in Example 1, $T < \infty$ and $S(T) > 0$. Expression (3.6) was derived by a different method in Sander (1975b), with the same restriction on T.

A special example of an L-estimate is the p-trimmed mean. Let $\hat{\mu}^{\mathrm{KM}}(p) = 1/(1 - 2p) \int_p^{1-p} \hat{F}^{\circ -1}(t)\, dt$. This is $T_1(\hat{F}^\circ)$ with $J(t) = 1/(1 - 2p) \times 1\{p \le t \le 1 - p\}$. Here $2p$ is a number between 0 and 1 representing the proportion of extreme observations excluded from the calculation of the mean. The influence curve for the trimmed mean is a combination of those of the median and mean, and reflects the fact that new observations in the middle of the range of the data affect the statistic in proportion to their value, but extreme observations have a bounded effect on the statistic.

$$\begin{aligned} \mathrm{IC}(T_1, F; t) &= (1 - 2p)^{-1}F^{-1}(p), & t &< F^{-1}(p) \\ &= (1 - 2p)^{-1}t, & F^{-1}(p) \le t &< F^{-1}(1 - p) \\ &= (1 - 2p)^{-1}F^{-1}(1 - p), & t &> F^{-1}(1 - p). \end{aligned}$$

$$\mathrm{A.\ Var}\ \hat{\mu}^{\mathrm{KM}}(p) = \frac{1}{n}\frac{1}{(1 - 2p)^2}\int_x^y\int_x^y S^\circ(t)S^\circ(u)\int_0^{t\wedge u} \frac{dF^u(s)}{[1 - F(s)]^2}\, dt\, du,$$

where

$$x = F^{\circ -1}(p), \qquad y = F^{\circ -1}(1 - p).$$

EXAMPLE 4. Kaplan-Meier M-estimates. An M-estimate is defined as the solution to

$$\frac{1}{n}\textstyle\sum_{i=1}^n \psi(X_i, \hat{\theta},) = 0,$$

for some given function $\psi(\cdot)$. If $\psi(t, \theta) = \partial/\partial\theta \log f(t, \theta)$, where f is the density of the random variables $X_1, \cdots, X_n$, then $\hat{\theta}$ is the maximum likelihood estimate. The appropriate functional form is $\int \psi(t, T_1(F))\, dF(t) = 0$. The influence curve of T_1 is

$$\mathrm{IC}(T_1, F; t) = \frac{\psi(t, T_1(F))}{\int \psi'(t, T_1(F))\, dF(t)},$$

where $\psi'(t, \theta) = (\partial/\partial\theta)\psi(t, \theta)$. The influence curve is directly proportional to the function ψ that defines the estimate. Thus M-estimates with influence curves of a desired form are easy to define. Huber (1964) proposed M-estimates as a generalization of maximum likelihood estimates, with desirable robustness properties. Two important examples are Huber's M-estimate,

$$\begin{aligned} \psi(t) &= -k \qquad t < -k \\ &= t \qquad -k \le t < k \\ &= k \qquad k < t, \end{aligned}$$

and Tukey's biweight

$$\begin{aligned} \psi(t) &= t(1 - t^2)^2, \qquad |t| \le 1 \\ &= 0 \qquad |t| > 1. \end{aligned}$$

These ψ-functions were suggested for the problem of locating the center of a symmetric distribution, in which case the defining equation becomes $\int \psi(x - T(F))\, dF(x) = 0$. In applying M-estimators to survival data, it will usually be appropriate to transform the observations (possibly by taking logarithms) in order to symmetrize the underlying distribution. In addition, in practice it is usually necessary to estimate the scale parameter of the underlying distribution, but this will not be considered here.

For the Kaplan-Meier estimate, $V(S_n^u, S_n^c) = T_1(\hat{F}^\circ)$ is defined implicitly by

$$\int \psi(t - T_1(\hat{F}^\circ))\, d\hat{F}^\circ(t) = 0.$$

From (3.3) and (3.5)

$$\mathrm{IC}_1(V, S^u, S^c; s) = \int_0^T S^\circ(t) \frac{1}{\alpha} \psi'(t - T(F^\circ)) \left\{ \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \le t\}}{(S^u + S^c)(s)} \right\} dt,$$

$$\text{A. Var } V = \frac{1}{n} \int_0^T \frac{1}{[1 - F(s)]^2} \left\{ \int_s^T \frac{1}{\alpha} S^\circ(t) \psi'(t - T(F^\circ))\, dt \right\}^2 dF^u(s),$$

where

$$\alpha = \int \psi'(t - T(F^\circ))\, dF^\circ(t).$$

APPENDIX.

In this section we give sufficient conditions for the asymptotic expansions of Section 2 to be valid.

A von Mises expansion of a functional T is analogous to a Taylor expansion of a function with the successive derivatives of f replaced by the successive differentials of T. In order to expand T in such a series it is necessary to check that T is k times differentiable, where k is determined by the nature of the problem. For expansions like (1.3), $k = 1$; higher values of k could be used for Edgeworth expansion results.

Let B_1, B_2 be normed topological vector spaces, and let T: $B_1 \to B_2$ be a given functional. Given $F \in B_1$, let dT_F be a linear transformation from B_1 to B_2. Define R: $B_1 \to B_2$ by $T(F + H) = T(F) + dT_F \circ H + R(F + H)$.

DEFINITION A.1. The functional T is Frechet differentiable at F iff

$$\| R(F + H) \|_{B_2} / \| H \|_{B_1} \to 0 \text{ as } \| H \|_{B_1} \to 0.$$

In order to verify definition A.1 it is necessary to specify a norm on B_1 and B_2. If $\| F_n - F \|_{B_1}$ is $O_p(1/\sqrt{n})$ then the limit law of $T(F_n)$ may be deduced from the limit law of the first two terms of its von Mises expansion, $T(F) + dT_F(F_n - F)$. Thus (1.4) can be expressed as

$$T(F_n) = T(F) + \int \mathrm{IC}(T, F; x)\, d(F_n - F)(x) + o_p(\| F_n - F \|_{B_1})$$

and we conclude

$$\sqrt{n}\,(T(F_n) - T(F)) \to_d N(0, \int (\mathrm{IC} - E_F\mathrm{IC})^2\, dF(x)).$$

In what follows we will show that the functional $T(S^u, S^c)$ of Section 2 is Frechet differentiable with respect to the sup-norm. It is well known that $\| F_n - F \|_\infty = O_p(1/\sqrt{n})$ (see, e.g., Billingsley (1968, Section 16)) and hence $\| S_n^u - S^u \|_\infty = O_p(1/\sqrt{n})$ by definition of S_n^u. From this the asymptotic normality of $T(S_n^u, S_n^c)$ follows directly. In addition it is easy to get a law of the iterated logarithm for $T(S_n^u, S_n^c)$ by using the fact that $\sqrt{n}\, \| S_n^u - S^u \|_\infty = O((\log \log n)^{1/2})$ w.p. 1. Our approach is in the spirit of Boos and Serfling (1979). For many statistical functionals that are not Frechet differentiable, the approach of Reeds (1976) via compact differentiability can be used to obtain limit theorems of the same type.

Let $\mathscr{B}$ be the space of subsurvival functions, i.e., the space of decreasing left continuous functions: $\mathbb{R}^+ \to [0, \alpha]$, where $\alpha \le 1$. Let $\| W_1 - S_1 \|_\infty = \sup_{0 \le x < \infty} | W_1(x) - S_1(x) |$.

THEOREM A.1. *The function $T(S_1, S_2)$ defined by* (A.1) *is Frechet differentiable at $S_1 \in \mathscr{B}$, $S_2 \in \mathscr{B}$ with respect to $\| \cdot \|_\infty$ in each argument S_1 and S_2 for each fixed $t < \infty$ satisfying $S_1(t) > 0$, $S_2(t) > 0$. The differential $dT_{S_1,S_2}(W_1 - S_1, S_2)$ is given by* (A.2).

PROOF. Let

$$C_F = \{\text{continuity intervals of a survival function } F\}$$

$$D_F = \{\text{jump points of } F\} \tag{A.1}$$

$$T(S_1, S_2) = \int_{C_{S_1[0,t]}} \frac{dS_1(s)}{(S_1 + S_2)(s)} + \sum_{D_{S_1[0,t)}} \ln \frac{S_1(s^+) + S_2(s^+)}{S_1(s^-) + S_2(s^-)}.$$

We first consider differentiability with respect to the first argument, S_1

$$T(W_1, S_2) = \int_{C_{W_1[0,t]}} \frac{dW_1(s)}{(W_1 + S_2)(s)} + \sum_{D_{W_1[0,t)}} \ln \frac{W_1(s^+) + S_2(s^+)}{W_1(s^-) + S_2(s^-)}$$

$$\begin{aligned} dT_{S_1,S_2}\circ(W_1 - S_1, S_2) = &\int_{C_{S_1}\cap C_{W_1}\cap[0,t]} \frac{d(W_1 - S_1)(s)}{(S_1 + S_2)(s)} \\ &- \int_{C_{S_1}\cap C_{W_1}\cap[0,t]} \frac{(W_1 - S_1)(s)\,dS_1(s)}{(S_1 + S_2)^2(s)} \\ &+ \sum_{D_{S_1}\cup D_{W_1}\cup[0,t)} \left\{ \frac{(W_1 - S_1)(s^+)}{(S_1 + S_2)(s^+)} - \frac{(W_1 - S_1)(s^-)}{(S_1 + S_2)(s^-)} \right\}. \end{aligned} \tag{A.2}$$

(For convenience, we have assumed that W_1 and S_1 do not jump exactly at t.) Expression (A.1) is a valid representation of the cumulative hazard only when S_1 and S_2 do not jump at the same point (Peterson, 1977), so $S_2(s^+) = S_2(s^-)$ in all the expressions above.

First, assume S_1 is continuous and W_1 is a (left-continuous) step function. The expressions simplify to

$$T(S_1, S_2) = \int_0^t \frac{dS_1(s)}{(S_1 + S_2)(s)}$$

$$T(W_1, S_2) = \sum_{D_{W_1[0,t)}} \ln \frac{W_1(s^+) + S_2(s)}{W_1(s^-) + S_2(s)}$$

$$\begin{aligned} dT_{S_1,S_2}\circ(W_1 - S_1, S_2) = &- \int_0^t \frac{dS_1(s)}{(S_1 + S_2)(s)} - \int_0^t \frac{(W_1 - S_1)(s)\,dS_1(s)}{(S_1 + S_2)^2(s)} \\ &+ \sum_{D_{W_1[0,t)}} \frac{W_1(s^+) - W_1(s^-)}{(S_1 + S_2)(s)}. \end{aligned}$$

(This last expression is derived in Section 2.)

$$T(W_1, S_2) - T(S_1, S_2) - dT_{S_1,S_2}\circ(W_1 - S_1, S_2)$$

$$(*) \quad = \sum_{S\in D_{W_1[0,t)}}\left\{\frac{W_1(s^+) + S_2(s)}{W_1(s^-) + S_2(s)} - 1\right\} + \int_0^t \frac{(W_1 - S_1)(s)\, dS_1(s)}{(S_1 + S_2)^2(s)} - \int_0^t \frac{dW_1(s)}{(S_1 + S_2)(s)}$$

$$+ o(\| W_1 - S_1 \|_\infty)$$

where we have used the fact that

$$\left|\frac{W_1(s^+) - W_1(s^-)}{W_1(s^-) + S_2(s)}\right| \le \frac{2}{|S_2(t)|} \| W_1 - S_1 \|_\infty .$$

The first term simplifies to $\int_0^t (dW_1(s)/(W_1 + S_2)(s))$, so

$$(*) \quad = \int_0^t \frac{(W_1 - S_1)(s)\, dS_1(s)}{(S_1 + S_2)^2(s)} + \int_0^t \left(\frac{1}{W_1(s) + S_2(s)} - \frac{1}{S_1(s) + S_2(s)}\right) dW_1(s)$$

$$+ o(\| W_1 - S_1 \|_\infty)$$

$$= \int_0^t \frac{(W_1 - S_1)(s)\, dS_1(s)}{(S_1 + S_2)^2(s)} - \int_0^t \frac{(W_1 - S_1)(s)\, dW_1(s)}{(W_1 + S_2)(s)(S_1 + S_2)(s)} + o(\| W_1 - S_1 \|_\infty).$$

It remains to show that the first two terms of this expression are $o(\| W_1 - S_1 \|_\infty)$.

$$\left| \int_0^t \frac{(W_1 - S_1)(s)\, dS_1(s)}{(S_1 + S_2)^2(s)} - \int_0^t \frac{(W_1 - S_1)(s)\, dW_1(s)}{(W_1 + S_2)(s)(S_1 + S_2)(s)} \right|$$

$$= \left| \int_0^t \frac{[(W_1 + S_2)(s)\, dS_1(s) - (S_1 + S_2)(s)\, d(W_1 - S_1)(s)}{(S_1 + S_2)^2(s)(W_1 + S_2)(s)} \right|$$

$$\le \left| \int_0^t \frac{(W_1 - S_1)(s)[W_1(s)\, dS_1(s) - S_1(s)\, dW_1(s)]}{(S_1 + S_2)^2(s)(W_1 + S_2)(s)} \right|$$

$$+ \left| \int_0^t \frac{(W_1 - S_1)(s) S_2(s)\, d(S_1 - W_1)(s)}{(S_1 + S_2)^2(s)(W_1 + S_2)(s)} \right| .$$

To show that the first term is $o(\| W_1 - S_1 \|_\infty)$, we use the following:

$$\text{(a)} \quad \left| \int_0^t (W_1 - S_1)(s) W_1(s)\, dS_1(s) - \int_0^t (W_1 - S_1)(s) S_1(s)\, dS_1(s) \right|$$

$$= \left| \int_0^t (W_1 - S_1)^2(s)\, dS_1(s) \right| \le \| W_1 - S_1 \|_\infty \| W_1 - S_1 \|_\infty | S_1(t) - S_1(0) |$$

By a similar argument we have:

$$\text{(b)} \quad \left| \int_0^t (W_1 - S_1)(s) S_1(s)\, dW_1(s) - \int_0^t (W_1 - S_1)(s) S_1(s)\, dS_1(s) \right| \le \| W_1 - S_1 \|_\infty^2 S_1(0).$$

To show that the second term is $o(\| W_1 - S_1 \|_\infty)$ we apply equation (b). This concludes the proof of the theorem for the first argument, S_1, and the case S_1 continuous, W_1 discrete. Differentiability with respect to S_2 is straightforward and will be omitted.

To show (*) is $o(\| S_1 - W_1 \|_\infty)$ when S_1, W_1 are arbitrary members of $\mathscr{B}$ (composed of discrete and continuous parts) involves essentially the same argument, with some additional analysis to show that the (S_1) measure of sets of the $\{D_{W_1} \cap C_{S_1}\}$ goes to zero as $\| S_1 - W_1 \|_\infty \to 0$. The calculations are somewhat more tedious than the above and will not be presented here. In fact, as Boos and Serfling (1979) emphasize, for the asymptotic distribution results of Section 2 it is sufficient to establish that (*) is $O_p(\| W_1 - S_1 \|)$ for the special case $W_1 = S_{1n}$, the empirical survival function associated with S_1.

It is also possible to show Frechet differentiability by establishing that the Gateaux differential $dT_{S_1,S_2}(H_1, S_2)$ is continuous in S_1, in the sup norm, for all H_1. (This is Proposition A.2.3 of Reeds, 1976.) Such a verification involves calculations essentially similar to those presented above.

In the proof of Theorem A.2.1, bounding the expression (*) is simplified by the fact that we are integrating over the finite interval [0, t]. The condition that $S_2(t)$ be strictly positive is natural in the context of the censored data problem. In this case S_2 represents the subsurvival function for the censored observations. It makes sense to require that the support of the censoring distribution extend beyond the point at which the survival function is being estimated. For a fuller discussion of this point, see Breslow and Crowley (1974).

Acknowledgment. This paper is based on the author's Ph.D. dissertation at Stanford University under the supervision of Professor Rupert Miller. I would like to thank Professor Miller for his guidance and many valuable suggestions. I am grateful to André Cabannes for helpful discussions, especially about the material in the appendix. I would also like to thank the referee for careful reading of the manuscript.

REFERENCES

AALEN, ODD (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.

ANDREWS, D. F., BICKEL, P. J., et al. (1972). *Robust Estimates of Location.* Princeton Univ. Press.

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959.

BOOS, D. D. and SERFLING, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. *Ann. Statist.* **8** 618–624.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under censorship. *Ann. Statist.* **2** 437–453.

EFRON, B. (1967). The two sample problem with censored data. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** 831–853. Univ. California Press, Berkeley.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1977). *Robust Statistics.* SIAM Philadelphia.

KAPLAN, E. L., and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

LAMBERT, D. (1979). Influence functions for testing. Technical Report No. 152, Depart. Statist., Carnegie-Mellon Univ., Pittsburgh.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics.* 67–87. (J. Gani, ed.) Applied Probability Trust, Sheffield.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15.

MILLER, R. G. (1975). Jackknifing censored data. Technical Report No. 14, Depart. Statist., Stanford Univ.

MILLER, R. G. (1978). The jackknife: survey and applications. Proceedings of the twenty-third conference on the design of experiments in army research development and testing. ARO Report 78-2.

MISES, R., v (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

PETERSON, A. V. (1975). Nonparametric estimation in the competing risks problem. Technical Report No. 13, Depart. Statist., Stanford Univ.

PETERSON, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical sub-survival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

REEDS, J. A. (1976). On the definition of Von Mises functionals. Ph.D. thesis, Depart. Statist., Harvard Univ. Cambridge.

SANDER, J. M. (1975a). The weak convergence of quantiles of the product-limit estimator. Technical Report No. 5, Depart. Statist., Stanford Univ.

SANDER, J. M. (1975b). Asymptotic normality of linear combinations of functions of order statistics with censored data. Technical Report No. 8, Department of Stat., Stanford Univ.

SUSARLA, V. and VAN RYZIN, J. (1980). Large sample-theory for an estimator of mean survival time from censored samples. *Ann. Statist.* **8** 1001–1016.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C.
CANADA V6T 1W5