INFLUENCE FUNCTIONS FOR CENSORED DATA

BY

NANCY MARGARET REID

TECHNICAL REPORT NO. 46
JUNE 1979

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

INFLUENCE FUNCTIONS FOR CENSORED DATA

BY

NANCY MARGARET REID

TECHNICAL REPORT NO. 46

JUNE 1979

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-04

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

1. Introduction

Let X_1^o, X_2^o, ..., X_n^o be independent, identically distributed random variables with distribution function F^o. These are the true lifetimes of the items under observation. Associated with each X_i^o is an independent censoring variable Y_i, and Y_1, Y_2, ..., Y_n are assumed independent, identically distributed with distribution function G. The observations are the n pairs (X_1, δ_1), ..., (X_n, δ_n), where $X_i = \min(X_i^o, Y_i)$ and $\delta_i = 1\{X_i = X_i^o\}$. (Throughout, $1\{A\}$ is the indicator function for the event A.) The distribution function of the X_i's is F and satisfies $1-F(t) = [1-F^o(t)][1-G(t)]$. Two subdistributions 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) $dF^u(t) = [1-G(t)]dF^o(t)$

(ii) $F^u(\infty) = \Pr\{\delta_i = 1\}, \quad F^c(\infty) = \Pr\{\delta_i = 0\}$

(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 c.d.f.'s are indicated by a subscript n. For example,

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

1

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 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

$$\hat{S}^\circ(t) = 1 - \hat{F}^\circ(t) = \begin{cases} \prod_{i: X_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & t \leq X_{(n)} \\ \\ 0 & t > X_{(n)} \ , \end{cases}$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ 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. This is convenient for theoretical reasons, and does not affect asymptotic calculations. However, it is more suitable in practice to leave $\hat{S}^\circ(t)$ undefined for $t > X_{(n)}$ if

$\delta_{(n)}=0$. The function $\hat{S}^\circ(t)$ is constant between uncensored observations. The size of each jump is a function of the sample size n and the censoring pattern between each pair of the ordered uncensored observations.

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, $\{\hat{S}^\circ(t); t \geq 0\}$ 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 Kaplan-Meier estimate was shown by Efron (1967) to be the point of convergence of the iterative scheme defined by

$$n\hat{S}^\circ_{(p)}(t) = \sum_{i=1}^{n} 1\{X_i > t\} + \sum_{i: X_i \leq t} (1-\delta_i) \frac{\hat{S}^\circ_{(p-1)}(t)}{\hat{S}^\circ_{(p-1)}(X_i)},$$

and for this reason he called $\hat{S}^\circ(t)$ a self-consistent estimate of $S^\circ(t)$. This version of the Kaplan-Meier estimate is constructed by the E-M algorithm (Dempster, Laird and Rubin, 1977). The first term $\sum 1\{X_i > t\}$ is the number of observations observed to be larger than t, and the second term is the expected number of observations greater

than t, based on the information in the censored observations and the current estimate of $S°$.

The role of censored and uncensored observations in the construction of $\hat{S}°(t)$ 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 $S°(t)$ can be expressed as a functional of the two subsurvival functions:

$$S°(t) = \exp \int_0^t \frac{dS^u(s)}{(S^u+S^c)(s)} \times \exp \sum_{s \leq t} \ln \frac{S^u(s^+) + S^c(s^+)}{S^u(s^-) + S^c(s^-)} . \qquad (1.1)$$

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

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

Because S_n^u and S_n^c are the maximum likelihood estimates of the subsurvival functions, it follows from the invariance property of the maximum likelihood estimate that the m.l.e. of $S°$, which is the Kaplan-Meier estimate, satisfies

$$\hat{S}°(t) = \exp \int_0^t \frac{dS_n^u(s)}{(S_n^u+S_n^c)(s)} \times \exp \sum_{s \leq t} \ln \frac{S_n^u(s^+) + S_n^c(s^+)}{S_n^u(s^-) + S_n^c(s^-)} . \qquad (1.2)$$

(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).

In many estimation problems the statistic $\hat{\theta}$ can be expressed as a functional of the empirical cumulative distribution function, $\hat{\theta} = T(F_n)$, and the unknown parameter as the same functional of the true distribution function, $\theta = T(F)$. For example, the sample mean $\hat{\theta} = n^{-1} \sum_{i=1}^{n} X_i$ can be written $\hat{\theta} = T(F_n) = \int x dF_n(x)$; the true mean is $T(F) = \int x dF(x)$. If T is a von Mises functional

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

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

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

5

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

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

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

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

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

$$\frac{d}{d\varepsilon} T(F+\varepsilon(G-F))\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} T((1-\varepsilon)F + \varepsilon\delta_x)\Big|_{\varepsilon=0}$$

$$= \int IC(T,F;y)d(\delta_x - F)(y)$$

$$= IC(T,F;x) - \mu\ .$$

Except for the constant term μ, this derivative measures the effect on the functional T of a small (infinitesimal) change in the weight

6

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. For $T(F) = \int y \, d(G-F)(y)$,
$\frac{d}{d\varepsilon} T(F+\varepsilon(G-F)) \big|_{\varepsilon=0} = \int y \, d(G-F)(y)$, so $IC(T,F;x) = x$. The sample
mean is sensitive to large observations and this is reflected in
the fact that the influence curve is unbounded. The effect on the
sample mean of an additional observation is directly proportional to
the value of the observation. In contrast to this, the influence
curve of the median is a step function. The median is expressed as
a functional by $F(T(F)) = \frac{1}{2}$, and the influence curve is

$$IC(T,F;x) = \begin{cases} - \dfrac{1}{f(F^{-1}(\frac{1}{2}))} \, , & x \leq F^{-1}(\frac{1}{2}) \\[3mm] 0 \, , & \text{otherwise .} \end{cases}$$

A statistical functional with a bounded influence curve is not sensi-
tive to extreme observations, so is robust in this sense. If the
influence curve $IC(T,F;x)$ 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" dis-
tribution; the true underlying distribution is $F_1 = F+\varepsilon H$. 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, sec. 1.6), or for bivariate functionals $T(F_1,F_2)$. The bivariate von Mises expansion is

$$
\begin{aligned}
T(G_1,G_2) = T(F_1,F_2) &+ \int IC_1(T,F_1,F_2;y)d(G_1-F_1)(y) \\
&+ \int IC_2(T,F_1,F_2;y)d(G_2-F_2)(y) + \text{higher order terms} .
\end{aligned} \tag{1.6}
$$

The two influence curves are defined by

$$
\begin{aligned}
\frac{\partial}{\partial \varepsilon} T(F_1+\varepsilon(G_1-F_1), F_2+\delta(G_2-F_2)) \Big|_{\substack{\varepsilon=0 \\ \delta=0}} &= \int IC_1(T,F_1,F_2;y)d(G_1-F_1)(y) \\
\frac{\partial}{\partial \delta} T(F_1+\varepsilon(G_1-F_1), F_2+\delta(G_2-F_2)) \Big|_{\substack{\varepsilon=0 \\ \delta=0}} &= \int IC_2(T,F_1,F_2;y)d(G_2-F_2)(y) .
\end{aligned} \tag{1.7}
$$

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.

8

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

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

and the corresponding functional for the true cumulative hazard function is

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

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.

$$T(S^u + \varepsilon(S_n^u - S^u), \ S^c + \delta(S_n^c - S^c), \ t)$$

$$= -\int_0^t \frac{d[(1-\varepsilon)S^u(s)]}{[S^u + \varepsilon(S_n^u - S^u)](s) + [S^c + \delta(S_n^c - S^c)](s)}$$

$$- \sum_{s \le t} \ell n \frac{[S^u + \varepsilon(S_n^u - S^u)](s^+) + [S^c + \delta(S_n^c - S^c)](s^+)}{[S^u + \varepsilon(S_n^u - S^u)](s^-) + [S^c + \delta(S_n^c - S^c)](s^-)} \, .$$

9

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

$$\frac{\partial}{\partial \varepsilon} T(S^u + \varepsilon(S_n^u - S^u), \; S^c + \delta(S_n^c - S^c), \; t)\bigg|_{(0,0)}$$

$$= \int_0^t \frac{[S^u(s) + S^c(s)](dS^u(s)) + (S_n^u - S^u)(s)dS^u(s)}{[S^u(s) + S^c(s)]^2} - \sum_{s \le t} \left\{ \frac{S^u(s^-) + S^c(s^-)}{S^u(s^+) + S^c(s^+)} \right.$$

$$\left. \cdot \left\{ \frac{(S_n^u - S^u)(s^+)[S^u(s^-) + S^c(s^-)] - (S_n^u - S^u)(s^-)[S^u(s^+) + S^c(s^+)]}{[S^u(s^-) + S^c(s^-)]^2} \right\} \right\}$$

$$= \int_0^t \frac{dS^u(s)}{S^u(s) + S^c(s)} + \int_0^t \frac{(S_n^u - S^u)(s)dS^u(s)}{[S^u(s) + S^c(s)]^2} - \sum_{s \le t} \frac{S_n^u(s^+) - S_n^u(s^-)}{S^u(s) + S^c(s)} \;,$$

where for the last equality we have used the fact that S^u is continuous. A similar calculation gives

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

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

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

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

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

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

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

$$\left. \begin{aligned} IC_1(T, S^u, S^c; s)(t) &= \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \leq t\}}{(S^u + S^c)(s)} \\[2em] IC_2(T, S^u, S^u; s)(t) &= \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} \ . \end{aligned} \right\} \tag{2.1}$$

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

$$\left. \begin{aligned} IC_1(T^\circ, S^u, S^c; s)(t) &= S^\circ(t) \ IC_1(T, S^u, S^c; s)(t) \\[2em] IC_2(T^\circ, S^u, S^c; s)(t) &= S^\circ(t) \ IC_2(T, S^u, S^c; s)(t) \ . \end{aligned} \right\} \tag{2.2}$$

Certain features of the Kaplan–Meier estimate are reflected in (2.1). The first term of $IC_1(T, S^u, S^c; s)(t)$ 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 $IC_1(T, S^u, S^c; s)(t)$ 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 $\int_0^{s \wedge t} \frac{dS(u)}{(S^u + S^c)^2(u)}$ is decreasing in s until $s = t$. Note that, for $s < t$

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

$$\geq \int_0^s \frac{d(S^u+S^c)(u)}{(S^u+S^c)^2(u)} + \frac{1}{(S^u+S^c)(s)} = 1 - \frac{1}{(S^u+S^c)(s)} + \frac{1}{(S^u+S^c)(s)} = 1 \, ,$$
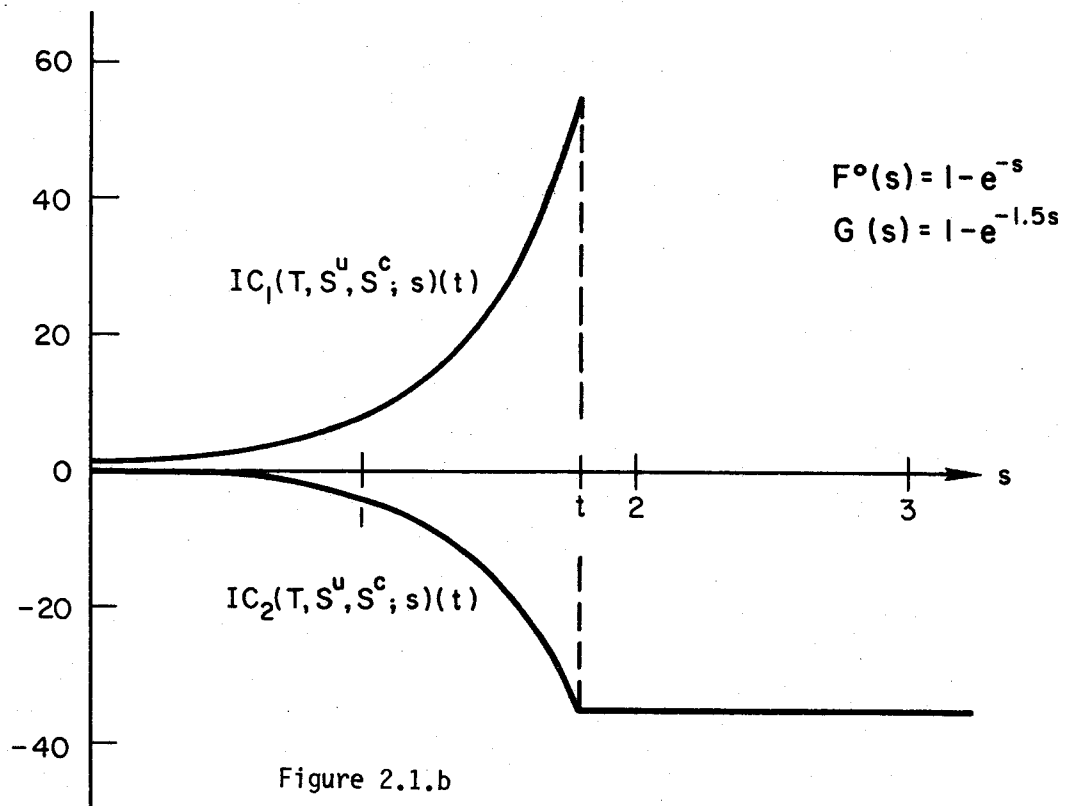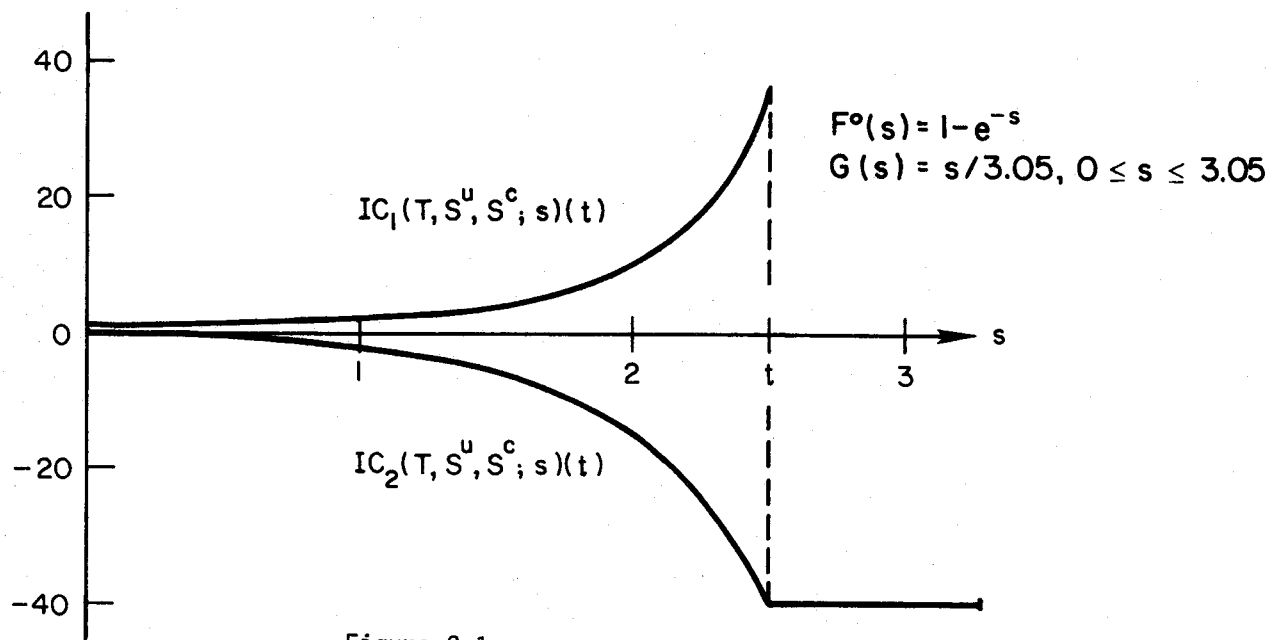
so $IC_1(T,S^u,S^c;s)(t)$ is ≥ 1 for $s<t$, and is increasing in s. If both s and t are large, with $s<t$, introducing a new observation at the point s has a large effect on the cumulative hazard at the point t. This is not translated into a large effect on $\hat{S}^\circ(t)$, however, because $\hat{S}^\circ(t)$ decreases to 0 as $t \to \infty$. Figure (2.1) sketches $IC_i(T,S^u,S^c;s)(t)$ 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.

Note that $IC_i(T^\circ,S^u,S^c;s)(t)$ is a function of s defined at a fixed value of t; it is the influence curve of the point estimate $\hat{S}^\circ(t)$. However, the Kaplan-Meier estimate itself can be regarded as an element in the space of distribution functions, and similarly $IC_i(T^\circ,S^u,S^c;s)(t)$ for fixed s can be regarded as a process in t.

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

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

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

$$= \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i=1}}^n IC_1(T,S^u,S^c;X_i)(t) + \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i=0}}^n IC_2(T,S^u,S^c;X_i)(t)$$

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

12

$F^o(s) = 1 - e^{-s}$
$G(s) = s/3.05,\ 0 \le s \le 3.05$

$IC_1(T, S^u, S^c; s)(t)$

$IC_2(T, S^u, S^c; s)(t)$

Figure 2.1.a

$F^o(s) = 1 - e^{-s}$
$G(s) = 1 - e^{-1.5s}$

$IC_1(T, S^u, S^c; s)(t)$

$IC_2(T, S^u, S^c; s)(t)$

Figure 2.1.b

13

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)

$$\int IC_1 dS^u(s) + \int IC_2 dS^c(s)$$

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

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

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

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

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

Then, disregarding the higher order terms,

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

$$\simeq E\left[\frac{1}{n} \sum_{\substack{i=1 \\ \delta_i = 1}}^n IC_1(T, S^u, S^c; X_i)(t) + \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i = 0}}^n IC_2(T, S^u, S^c; X_i)(t) \right]^2$$

$$= \frac{1}{n} \int IC_1^2(T, S^u, S^c; s)(t)dF^u(s) + \frac{1}{n} \int IC_2^2(T, S^u, S^c; s)(t)dF^c(s) \ .$$

14

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.

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

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

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

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

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

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

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).

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

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

$$+ 2 \int_0^t (S^u + S^c)(s) \int_0^s \frac{dS^u(u)}{(S^u + S^c)^2 (u)} \frac{dS^u(s)}{(S^u + S^c)^2 (s)}$$

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

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 calcu-

lation because the jackknife is a finite sample approximation to the

influence function. An approximation to expression (1.5), with $F = F_n$,

the empirical c.d.f. based on a sample $X_1, X_2, \ldots, X_n, \varepsilon = \frac{1}{n-1}$,

and $G = \delta_{X_i}$, a point mass at the ith sample point, is

$$\frac{T(F + \varepsilon(G-F)) - T(F)}{\varepsilon} = (n-1)(\hat{\theta} - \hat{\theta}_{-i}) = \hat{\theta} - \tilde{\theta}_i$$

$$\triangleq \widehat{IC}(\hat{\theta}; X_i) \ ,$$

where $\hat{\theta} = T(F_n)$ is the parameter estimate, $\hat{\theta}_{-i}$ is the estimate of

θ recomputed with the ith sample point deleted, and $\tilde{\theta}_i$ is the ith

pseudo-value. If $\widehat{IC}(T(F_n); y)$ converges to $IC(T, F; y)$, 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 $\widehat{IC}(T(F_n);y)$ is also related to Tukey's sensitivity curve, discussed, for example, in Huber (1977).

It is not possible to conclude that $\hat{\Lambda}°(t)$ 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 <u>Frechet differentiable</u> 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 $\frac{1}{\sqrt{n}}$. These regularity conditions are verified in the Appendix.

3. <u>Functions of the Kaplan-Meier Estimate</u>

Parameters of the underlying distribution function $F°$ can be estimated using $\hat{F}°$. For example, Kaplan and Meier (1958) suggested estimating the mean of $F°$ by $\hat{\mu}^{KM} = \int x d\hat{F}°(x)$. The influence curves for the functionals $T°(S^u, S^c; t) = 1 - \hat{F}°(t)$ 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 $\hat{\mu}_{KM}$, and other estimators based on $\hat{F}°(t)$.

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

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

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

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

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

$$dV\Big|_G \circ \lambda = \int IC(T_1, T_2(G); t)(dT_2\Big|_G \circ \lambda) \{dt\}$$

$$= \int IC(T_1, T_2(G); t) \int IC(T_2, G; s)(\{dt\})\lambda\{ds\}$$

$$= \iint IC(T_1, T_2(G); t) \ IC(T_2, G; s)(\{dt\})\lambda\{ds\}$$

$$\overset{\Delta}{=} \int IC(V, G; s)\lambda\{ds\} \ .$$

Hence

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

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

$$IC_1(V, F_1, F_2; s) = \int IC(T_1, T_2(F_1, F_2); t) \ IC_1(T_2, F_1, F_2; s)(\{dt\})$$

$$IC_2(V, F_1, F_2; s) = \int IC(T_1, T_2(F_1, F_2); t) \ IC_2(T_2, F_1, F_2; s)(\{dt\}) \ .$$

(3.1)

Integrating (3.1) by parts we obtain

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

which is valid as long as $IC_i(T_2,F_1,F_2;s)(t) \times IC(T_1,T_2;t)$ 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 $\hat{S}^\circ(t) = T^\circ(S_n^u, S_n^c, t)$ and $V(S_n^u, S_n^c) = T_1(\hat{S}^\circ) = T_1(T^\circ(S^u,S^c,\cdot))$. Let $g(t) = \frac{d}{dt} IC(T_1,T_2;t)$. Then

$$IC_1(V,S^u,S^c;s) = \int_0^\infty S^\circ(t)\left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u+S^c)^2(u)} + \frac{1\{s\leq t\}}{(S^u+S^c)(s)}\right)g(t)dt$$

$$(3.3)$$

$$IC_2(V,S^u,S^c;s) = \int_0^\infty S^\circ(t)\left(\int_0^{s\wedge t} \frac{dS^u(u)}{(S^u+S^c)^2(s)}\right)g(t)dt .$$

From the asymptotic expansion (1.6),

$$n * \text{A. Var. } V = \int_0^\infty IC_1^2(V,S^u,S^c;s)dF^u(s) + \int_0^\infty IC_2^2(V,S^u,S^c;s)dF^c(s)$$

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

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

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

19

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

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

$$+ \int_0^\infty \left\{ \int_0^\infty S^\circ(t) \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} \, g(t) dt \right\}^2 \, d(F^u + F^c)(s) \ . \qquad (3.4)$$

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

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

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

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

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

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

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:

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

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: $\hat{\mu}^{KM} = \int_0^\infty \hat{S}^\circ(t) dt$.

We have seen in Section 1 that the appropriate functional is $T_1(F) = \int t \, dF(t)$, with influence curve $IC(T_1, F; t) = t$. Writing $V(S_n^u, S_n^c) = T_1(T^\circ(S_n^u, S_n^c, \cdot))$, it follows that

$$IC_1(V, S^u, S^c; s) = \int_0^\infty S^\circ(t) \left\{ \frac{1\{s \leq t\}}{(S^u + S^c)(s)} + \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} \right\} dt \, ;$$

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

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

which is the result given by Breslow and Crowley (1974, equation 8.2).

21

It may be more relevant in practice to calculate the restricted mean (Meier, 1975), $\hat{\mu}_b^{KM} = \int_0^b \hat{S}^\circ(t)dt$, particularly when the largest observation of a sample is censored. The asymptotic variance of $\hat{\mu}_b^{KM}$, from (2.5), is

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

Example 2. The Kaplan-Meier median $\hat{m}^{KM} = \inf\{m: F^\circ(m) \geq \frac{1}{2}\}$.

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 $G(T_1(G)) = \frac{1}{2}$. The influence curve is an indicator function (see Section 1), so

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

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

$$
IC_1(V, S^u, S^c; s) = \frac{S^\circ(m)}{f^\circ(m)} \left\{ \int_0^{s \wedge m} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \leq m\}}{(S^u + S^c)(s)} \right\}
$$

$$
= \frac{1}{2f^\circ(m)} IC_1(T^\circ, S^u, S^c; s)(m)
$$

and

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

$$
= \frac{1}{n} \int_0^m \frac{1}{[1-F(s)]^2} \frac{[S^\circ(m)]^2}{[f^\circ(m)]^2} dF^u(s)
$$

$$
= \frac{1}{n} \frac{1}{4[f^\circ(m)]^2} \int_0^m \frac{dF^u(s)}{[1-F(s)]^2} \ .
$$

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

<u>Example 3.</u> Kaplan-Meier L-estimators.

An L-estimate is a linear combination of order statistics $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} a_{in} X_{(i)}$, where $a_{ni} = J(\frac{i}{n+1})$ 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 c.d.f. as $T_1(F_n) = \int x\, J(x) dF_n(x)$. The corresponding Kaplan-Meier L-estimate is defined as $T_1(\hat{F}^\circ)$. The influence function for this functional is

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

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

$$IC_1(V, S^u, S^c; s) = \int_0^\infty S^\circ(t) J(S^\circ(t)) \left\{ \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \leq t\}}{(S^u + S^c)(s)} \right\} dt$$

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

$= \frac{1}{n} \int_0^\infty \frac{1}{[1-F(s)]^2} \int_0^\infty \int_0^\infty S^\circ(t) J(F^\circ(t)) S^\circ(u) J(S^\circ(u))$

$\cdot\, 1\{s \leq t\}\, 1\{s \leq u\} dt\, du\, dF^u(s)$

$= \frac{1}{n} \int_0^\infty \int_0^\infty S^\circ(t) J(S^\circ(t)) S^\circ(u) J(S^\circ(u)) \int_0^{u \wedge t} \frac{1}{[1-F(s)]^2} dt\, du \, . \quad (3.6)$

23

Expression (3.6) was derived by a different method in Sander (1975b).

A special example of an L-estimate is the p-trimmed mean. Let $\hat{\mu}^{KM}(p) = \frac{1}{1-2p} \int_p^{1-p} t \, d\hat{F}^\circ(t)$. This is $T_1(\hat{F}^\circ)$ with $J(t) = \frac{1}{1-2p} \times 1\{p \le t \le 1-p\}$. 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.

$$IC(\hat{\mu}^{KM}(p), F; t) = \begin{cases} (1-2p)^{-1} F^{-1}(p) , & t < F^{-1}(p) \\ (1-2p)^{-1} t , & F^{-1}(p) \le t < F^{-1}(1-p) \\ (1-2p)^{-1} F^{-1}(1-p) , & t > F^{-1}(1-p) . \end{cases}$$

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

where

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

Example 4. Kaplan–Meier M-estimates.

An M-estimate is defined as the solution to

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

for some given function $\psi(\cdot)$. If $\psi(x,\theta) = \frac{\partial}{\partial\theta} \log f(x,\theta)$, where f is the density of the random variables $X_1, \ldots, X_n$, then

24

$\hat{\theta}$ is the maximum likelihood estimate. The appropriate functional

form is $\int \psi(x, T_1(F)) dF(x) = 0$. The influence curve of T_1 is

$$IC(T_1, F; x) = \frac{\psi(x, T_1(F))}{\int \psi'(x, T_1(F)) dF(x)} ,$$

where $\psi'(x, \theta) = \frac{\partial}{\partial \theta} \psi(x, \theta)$. The influence curve is directly pro-

portional 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,

$$\psi(x) = \begin{cases} -k & x < -k \\ x & -k \leq x < k \\ k & k < x , \end{cases}$$

and Tukey's biweight

$$\psi(x) = \begin{cases} x(1-x^2)^2, & |x| \leq 1 \\ 0 & |x| > 1 . \end{cases}$$

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)

$$IC_1(V, S^u, S^c; s) = \int_0^\infty S^\circ(t) \frac{1}{\alpha} \psi'(t - T(F^\circ)) \left\{ \int_0^{s \wedge t} \frac{dS^u(u)}{(S^u + S^c)^2(u)} + \frac{1\{s \leq t\}}{(S^u + S^c)(s)} \right\} dt ,$$

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

where

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

4. 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 f, 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 $R: B_1 \to B_2$ and $Q: R \times B_1 \to B_2$ by

$$T(F+H) = T(F) + dT_F \circ H + R(F+H)$$

$$Q(t,H) = \begin{cases} R(F+tH)/t & t \neq 0 \\ 0 & t=0 \ . \end{cases}$$

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

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

A different, but equivalent, definition of Frechet differentiability is given in Reeds (1976, Chapter 2).

Definition A.2.   (Reeds)   The functional   T   is Frechet differen-
tiable at   F   iff

$\forall$ bounded sets $B \subseteq B_1$

$\|Q(t,H)\|_{B_2} \to 0$ as $t \to 0$ uniformly in $H \in B$.

This definition is convenient for clarifying the relationship between
different kinds of differentiability. If the phrase "$\forall$ bounded sets
$B \subseteq B_1$" is replaced by "$\forall$ compact sets $B \subseteq B_1$", the functional   T
is compactly differentiable. Definition A.2 is valid for nonnormed
topological vector spaces.

In the above definitions, it is necessary that   F   and   F+H   be
elements of B_1. Most statistical applications use for B_1 the space
of cumulative distribution functions generated by probability measures.
In this case   H   itself cannot be a cumulative distribution function,
but is usually written as   G-F   where   G, F   are distribution functions.
This emphasizes the underlying idea that   H   is a small perturbation
of   F. Reeds does not restrict his definitions to the space of func-
tions of total mass   1.

In checking definition A.1, the candidate for dT_F is the Gateaux
differential of   T.

Definition A.3. The functional   T   is Gateaux differentiable at
F   iff

$$\lim_{t \to 0} \frac{T(F+tH) - T(F)}{t}$$

exists. The limit is called the Gateaux differential, and is written
$dT_F \circ H$. There is no confusion with this label, because a function

28

which is Frechet differentiable is also Gateaux differentiable, and the differentials are the same. Compact differentiability is weaker than Frechet and stronger than Gateaux differentiability.

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 $O_p\left(\frac{1}{\sqrt{n}}\right)$, 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, $T(F) + dT_F \circ (F_n - F)$. Thus (1.4) can be expressed as

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

and we conclude

$$\sqrt{n}\ (T(F_n) - T(F)) \xrightarrow{d} N\left(0,\ \int (IC - E_F IC)^2\ dF(x)\right)\ .$$

The representation $dT_F \circ (G-F) = \int IC(T,F;x) d(G-F)(x)$ holds for most statistical functionals; when it does not T is not a von Mises functional (by definition), so can't have a von Mises expansion. If G and F have total mass 1, $IC(T,F;x)$ as defined is only unique up to translation by constants, so is often defined to have mean zero. In the case of the Kaplan-Meier estimates, F and G are subdistribution functions so this convention is not necessary.

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 $\|F_n - F\|_\infty = O_p\left(\frac{1}{\sqrt{n}}\right)$ (see, e.g., Billingsley (1968, Section 16)) and hence $\|S_n^u - S^u\|_\infty = O_p\left(\frac{1}{\sqrt{n}}\right)$ 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 via compact differentiability can be used to obtain limit theorems of the same type.

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

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

Proof: Let $C_F = \{$continuity intervals of a survival function $F\}$

$\qquad\qquad D_F = \{$jump points of $F\}$

$$T(S_1, S_2) = \int_{C_{S_1}[0,t]} \frac{dS_1(s)}{(S_1+S_2)(s)} + \sum_{D_{S_1}[0,t)} \ln \frac{S_1(s^+)+S_2(s^+)}{S_1(s^-)+S_2(s^-)} \cdot \quad (A.1)$$

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

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

30

$$dT_{S_1,S_2} \circ (W_1-S_1, S_2) = \int_{C_{S_1} \cap C_{W_1} \cap [0,t]} \frac{d(W_1-S_1)(s)}{(S_1+S_2)(s)}$$

$$- \int_{C_{S_1} \cap C_{W_1} \cap [0,t]} \frac{(W_1-S_1)(s)dS_1(s)}{(S_1+S_2)^2(s)} \qquad (A.2)$$

$$+ \sum_{D_{S_1} \cup D_{W_1} \cap [0,t)} \left\{ \frac{(W_1-S_1)(s^+)}{(S_1+S_2)(s^+)} - \frac{(W_1-S_1)(s^-)}{(S_1+S_2)(s^-)} \right\} \quad .$$

(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

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

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

$$dT_{S_1,S_2} \circ (W_1-S_1, S_2) = - \int_0^t \frac{dS_1(s)}{(S_1+S_2)(s)} - \int_0^t \frac{(W_1-S_1)(s)dS_1(s)}{(S_1+S_2)^2(s)}$$

$$+ \sum_{D_{W_1}[0,t)} \frac{W_1(s^+)-W_1(s^-)}{(S_1+S_2)(s)} \quad .$$

(This last expression is derived in Section 2.)

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

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

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

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

where the last line follows from the fact that

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

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

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

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

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

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

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

32

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

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

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

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

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

(a): $\left| \int_0^t (W_1-S_1)(s)W_1(s)dS_1(s) - \int_0^t (W_1-S_1)(s)S_1(s)dS_1(s) \right|$

$$= \left| \int_0^t (W_1-S_1)^2(s)dS_1(s) \right|$$

$$\leq \sup_{0 \leq s \leq t}{}^2 |(W_1-S_1)(s)| \left| \int_0^t dS_1(s) \right|$$

$$\leq \|W_1-S_1\|_\infty \|W_1-S_1\|_\infty \underbrace{|S_1(t)-S_1(0)|}_{\in [0,1]}$$

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

(b): $\left| \int_0^t (W_1-S_1)(s)S_1(s)dW_1(s) - \int_0^t (W_1-S_1)(s)S_1(s)dS_1(s) \right|$

$$= \left| \int_0^t (W_1-S_1)(s)S_1(s)d(W_1-S_1)(s) \right|$$

$$\leq S_1(0) \left| \int_0^t (W_1-S_1)(s)d(W_1-S_1)(s) \right|$$

$$\leq \underbrace{S_1(0)}_{\in [0,\alpha]} \|W_1-S_1\|_\infty \|W_1-S_1\|_\infty$$

$$= o(\|W_1-S_1\|_\infty) .$$

To show that the second term is $o(\|W_1-S_1\|_\infty)$, 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 $\mathcal{B}$ (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 type $\{D_{W_1} \cap C_{S_1}\}$ 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.

34

It is also possible to show Frechet differentiability by estab-
lishing 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. (See, for
example, the discussion following Theorem 5 of Breslow and Crowley
(1974).) It is probably possible to relax this condition on $S_2(t)$
by using a result like Proposition 11.4.18 of Royden (1968, p. 233).

5. 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 University Press: Princeton.

Boos, D. D. (1979). A differential for L-statistics. Ann. Statist. 7, No. 5.

Boos, D. D., and Serfling, R. J. (1980). On differentials, asymptotic normality and almost sure behavior of statistical functions. Ann. Statist. 8, No. 3.

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.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. R. Statist. Soc. B 39, 1-38.

Efron, B. (1967). The two sample problem with censored data. Proc. 5th Berkeley Symp. 4, 831-853. University of 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.

Meier, P. (1975). Estimation of a distribution function from incomplete observations. Perspectives in Probability and Statistics, J. Gani, editor. 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, Department of Statistics, Stanford University: Stanford.

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, Department of Statis-
    tics, Stanford University: Stanford.

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, Department of Statistics, Harvard University:
    Cambridge.

Royden, H. L. (1968). _Real Analysis._ MacMillan: New York.

Sander, J. M. (1975a). The weak convergence of quantiles of the
    product-limit estimator. Technical Report No. 5, Department
    of Statistics, Stanford University: Stanford.

Sander, J. M. (1975b). Asymptotic normality of linear combinations
    of functions of order statistics with censored data. Technical
    Report No. 8, Department of Statistics, Stanford University:
    Stanford.

37