The Annals of Statistics
1993, Vol. 21, No. 3, 1608–1620

LARGE DEVIATIONS FOR CENSORED DATA

BY I. H. DINWOODIE

Tulane University

Large deviation properties of the Kaplan–Meier estimator are studied and applied to obtain the rate of exponential convergence of the estimator to the underlying survival curve.

1. Introduction. Let $X_1^\circ, X_2^\circ, \dots$ be an i.i.d. sequence of positive random variables with common distribution function F°. One often uses the empirical distribution function F_n° defined by

$$F_n^\circ(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i^\circ \le t\}}$$

to estimate F°. The estimator F_n° is well understood and its fundamental asymptotic properties can be found in Pollard (1984).

Censored data arise in applications when the exact value of each random variable X_i° is not always observed, but one knows a lower bound X_i for X_i° and whether one has observed the true value X_i° or the "censored value" X_i. For example, suppose one is measuring the survival times of patients entering a medical study at various times. Let X_i° represent the survival time of patient i after the patient enters the study, let T represent the time at which the study must end, and let Y_i represent $T - E_i$ where E_i represents the random time in $(0, T)$ at which patient i enters the study. If the study ends at time T, then we observe $(X_i^\circ + E_i) \wedge T$, or equivalently we observe $X_i^\circ \wedge (T - E_i) = X_i^\circ \wedge Y_i$ and we are interested in estimating the distribution of X_i° from these imperfect observations.

The standard estimator for the curve $1 - F^\circ$ in the presence of censored data is the Kaplan–Meier estimator [Kaplan and Meier (1958)], which will be denoted S_n. Gill and Johansen (1990), page 1536, and Gill (1983), page 53, prove uniform convergence of S_n to $1 - F^\circ$ on a fixed interval both a.s. and in probability. Breslow and Crowley (1974), Theorem 5, prove weak convergence on a fixed interval of the process $n^{1/2}(1 - S_n - F^\circ)$ to a Gaussian process. Csörgő and Horváth (1983), Theorem 1, prove a law of the iterated logarithm for a similar estimator which yields a.s. convergence on a fixed interval [under the assumption that $F^x(T) < 1$ (cf. (2.1)]. Thus the asymptotic theory for S_n includes convergence or consistency theorems, a central limit theorem and a law of the iterated logarithm. We present here some large deviation results concerning S_n from Dinwoodie (1990).

Received December 1990; revised May 1992.
AMS 1991 *subject classifications*. Primary 60F10; secondary 62G30.
Key words and phrases. Large deviations, censored data, Kaplan–Meier estimator, survival analysis.

To study large deviations for the estimator is to study the decay rate of the probability that the estimator performs poorly and falls away from the true probability distribution. Large deviation theorems generally establish an asymptotic exponential decay rate, and results for i.i.d. and Markov processes can be stated fairly simply in terms of a variational formula which often has the form of a Young–Fenchel transform [see Varadhan (1984) or Dinwoodie (1993)]. The Kaplan–Meier estimator is a new type of stochastic process with important applications, and so it is interesting both as the source of a new large deviation rate function and also as a practical candidate for efficiency and convergence properties related to large deviations.

We will give the formal definitions and then sketch the results. Let $(\Omega, A, \mathbf{P})$ be a probability space on which is defined a sequence of independent pairs of independent, positive random variables $\{(X_i^\circ, Y_i): i \geq 1\}$. For each $i \geq 1$, let

$$X_i = \min\{X_i^\circ, Y_i\},\quad \delta_i = 1_{\{X_i^\circ \leq Y_i\}}. \tag{1.1}$$

The Kaplan–Meier estimator for the survival curve

$$s_t^\circ = \mathbf{P}\{X_1^\circ > t\} \tag{1.2}$$

is the random variable

$$S_n(t) = \prod_{i=1:\, X_i \leq t}^{n} \left[\frac{n - r_i}{n - r_i + 1}\right]^{\delta_i}, \tag{1.3}$$

where r_i is the rank of $(X_i, 1 - \delta_i)$ in the set $\{(X_k, 1 - \delta_k), 1 \leq k \leq n\}$. Note that r_i depends on n. The ordering is lexicographical, meaning that deaths $(\delta_i = 1)$ come before censored observations $(\delta_i = 0)$ in the case of ties among the set $\{X_i: 1 \leq i \leq n\}$. When no censoring takes place, the Kaplan–Meier estimator reduces to the usual empirical probability distribution based on $\{(X_i^\circ): 1 \leq i \leq n\}$, for which large deviation properties are well known [see Groeneboom, Oosterhoff and Ruymgaart (1979)]. In general, the Kaplan–Meier estimator is a nonlinear function of the new process $\{(X_i, \delta_i)\}$, a process whose law depends on the distributions both of X_i° and Y_i. Many properties of the estimator can be found in the original paper of Kaplan and Meier (1958) or in Miller (1981). The censoring scheme studied here in which the sequence of censoring random variables $\{Y_i: i \geq 1\}$ is an i.i.d. sequence is called "random censoring." An alternative form of censoring occurs when each random variable Y_i is a fixed number c_i. The resulting sequence $\{X_i: i \geq 1\}$ is not i.i.d. and our results do not apply to this censoring scheme. Meier (1975) studies the Kaplan–Meier estimator under fixed censoring.

The technique used here to study S_n is to first establish a contraction, which means that the process S_n is written as the image of a process with known large deviation behavior. The known process will be a sequence of empirical probability measures P_n on $\mathbf{R}_+ \times \{0, 1\}$. The contraction will be denoted $\mathscr{f}$, and hence we write

$$S_n = \mathscr{f} \circ P_n.$$

The probability that the random curve S_n falls in a certain set of curves is the same as the probability that the well-understood process P_n falls in the preimage of this set. The difficulty in carrying out this idea for the Kaplan–Meier estimator is that the contraction $\mathscr{f}$ is not continuous and nonlinear. The results depend on a continuity property of $\mathscr{f}$ only at certain points in its domain (see Lemmas 2.2, 2.3 and 2.4).

Since all results for S_n are proved in the uniform topology on curves on the fixed interval $[0, T]$, we use large deviation results for the measure P_n not in the weak topology but rather the stronger τ-topology proved by Groeneboom, Oosterhoff and Ruymgaart (1979). The τ-topology on measures is very useful because its associated topology on distribution functions is stronger than the topology from the uniform metric. This property is the essence of Lemma 2.2, which together with the very useful Lemma 2.3 of Gill (1981) proves the continuity property of $\mathscr{f}$ at Lemma 2.4. This continuity property is combined with the large deviation results of Groeneboom, Oosterhoff and Ruymgaart (1979) to prove the basic large deviation properties of S_n at Lemma 3.1.

The main result is Theorem 3.1 which is a large deviation theorem for S_n with no restriction on the type of distribution function of either X_1° or Y_1. This theorem intuitively says that the large deviation rate at a certain curve s is the Kullback–Leibler distance from one measure on $\mathbf{R}_+ \times \{0, 1\}$ to another, where the first measure in this distance is a measure transformed from laws for X_1° and Y_1 which would make S_n consistent for s, and the second measure is transformed from the original laws of X_1° and Y_1. Theorem 3.1 specializes to Corollary 3.1 when one is concerned only with small neighborhoods of the true survival curve s°. Corollary 3.2 is the a.s. consistency of S_n and follows from Theorem 3.1 together with the assurance furnished by Lemma 3.2 that the exponential rates of Theorem 3.1 are nontrivial. This consistency result appears to have as few hypotheses as any consistency result in the literature and therefore argues for the use of large deviation methods as a general tool for proving consistency theorems in statistics.

2. $\mathbf{S}_n = \mathscr{f}(\mathbf{P}_n)$. Fix $T > 0$. This section is devoted to studying a representation of the Kaplan–Meier estimator S_n on the interval $[0, T]$ as a function $\mathscr{f}$ of an empirical probability measure P_n defined on $\Xi = \mathbf{R}_+ \times \{0, 1\}$. The map $\mathscr{f}$ will be called a contraction in the large deviation sense of Varadhan (1984). This use of the word "contraction" is unrelated to the more common use of the word in describing a map on a metric space that reduces distances between points. The map $\mathscr{f}$ will be defined on the space of probability measures on $\Xi = \mathbf{R}_+ \times \{0, 1\}$. The technical difficulty is that $\mathscr{f}$ is not continuous at certain probability measures which are supported on $[0, T) \times \{0, 1\}$. The probabilistic analysis succeeds because the empirical law P_n avoids such probability measures anyway when $\mathbf{P}(X_1 \geq T) > 0$.

Let F° denote the distribution function of X_1° and F^y the distribution function of Y_1. With this notation, the distribution function F^x for X_1 is

$$F^x = 1 - (1 - F^\circ)(1 - F^y) = 1 - s^\circ s^y. \tag{2.1}$$

P_n will denote the empirical probability measure on the space Ξ from the observations $\{(X_i, \delta_i): 1 \le i \le n\}$:

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{(X_i, \delta_i) \in A\}}, \tag{2.2}$$

where A is a measurable set in Ξ with the Borel field derived from the product metric. The probability induced on Ξ by (X_1, δ_1) will be called p, and we define

$$X_n^* = \max\{X_i: 1 \le i \le n\}. \tag{2.3}$$

We let $\mathbf{S}$ be the set of bounded, right continuous functions on $\mathbf{R}_+ = [0, \infty)$ with the σ-algebra generated by the open balls or the coordinate maps and we will consider S_n as an element of $\mathbf{S}$. If s_1 and s_2 are two elements of $\mathbf{S}$, then we will use the notation $\|s_1 - s_2\|_T = \sup_{0 \le t \le T} |s_1(t) - s_2(t)|$. We let $\mathbf{M}$ denote the set of probability measures on the metric space Ξ. The topology on $\mathbf{M}$ will be the τ-topology [see Groeneboom, Oosterhoff and Ruymgaart (1979) or Deuschel and Stroock (1989)], which is the one generated by the bounded, measurable functions on Ξ. A function s on $\mathbf{R}_+$ such that $1 - s$ is a probability distribution function will be called a survival curve. If $1 - s$ is a subdistribution function, then s will be called a subsurvival curve.

Let us proceed to construct the contraction $\not{f}$. Let $\mathbf{SD}$ denote the set of subdistribution functions on $\mathbf{R}_+$ with the uniform metric on $\mathbf{R}_+$ denoted $\| \ \|$. The contraction $\not{f}$ will be a composition of functions from $\mathbf{M}$ to $\mathbf{SD}$ to $\mathbf{S}$, and so $\mathbf{SD}$ can be ignored after some preliminary lemmas. Define the maps $f_k: \mathbf{M} \to \mathbf{SD}$ for $k = 0, 1$ by

$$\begin{aligned} f_0(q)(t) &= q([0,t] \times \{0\}), \qquad 0 \le t < \infty, \\ f_1(q)(t) &= q([0,t] \times \{1\}), \qquad 0 \le t < \infty. \end{aligned} \tag{2.4}$$

We will need a different representation for S_n which will require the following definitions:

$$\begin{aligned} N_n(t) &= n[f_1(P_n)(t)], \\ Y_n(t) &= n[1 - [(f_0 + f_1)(P_n)(t-)]]. \end{aligned} \tag{2.5}$$

$N_n(t)$ can be interpreted as the number of deaths up until and including time t, and $Y_n(t)$ is the number of observations among $\{X_i: 1 \le i \le n\}$ which are greater than or equal to time $t \ge 0$. The following lemma is known but we write out the proof to clarify the ranking procedure in the case of ties in the data. If H is a function such that the left limit $H(s-)$ exists, then we will use the notation

$$\Delta H(s) = H(s) - H(s-).$$

LEMMA 2.1. *The Kaplan–Meier estimator S_n has the representation*

$$S_n(t) = \prod_{s \le t} \left[1 - \frac{\Delta N_n(s)}{Y_n(s)}\right],$$

where $0/0 = 0$.

PROOF. Fix $\omega \in \Omega$ and let $\{s_1 < s_2 < \cdots < s_m\} = \{X_i(\omega) \le t: 1 \le i \le n\}$. Then from (1.3),

$$S_n(t) = \prod_{k=1}^{m} \prod_{X_i = s_k} \left[\frac{n - r_i}{n - r_i + 1} \right]^{\delta_i}.$$

Now the deaths at s_k have ranks $\{n - Y_n(s_k) + 1, \dots, n - Y_n(s_k) + \Delta N_n(s_k)\}$, and so by the telescoping effect of the second product, the last expression is equal to

$$\prod_{k=1}^{m} \left[1 - \frac{\Delta N_n(s_k)}{Y_n(s_k)} \right] = \prod_{s \le t} \left[1 - \frac{\Delta N_n(s)}{Y_n(s)} \right]. \qquad \square$$

Let $\mathbf{G} = \mathbf{SD} \times \mathbf{SD}$. If $G = (G_1, G_0) \in \mathbf{G}$, we define for $t \ge 0$,

$$Y_G(t) = (G_1 + G_0)(\infty) - [(G_1 + G_0)(t-)]. \tag{2.6}$$

We follow Gill (1981) and define $\Phi: \mathbf{G} \to \mathbf{S}$ by

$$\Phi(G_1, G_0)(t) = \prod_{s \le t} \left[1 - \frac{\Delta G_1(s)}{Y_G(s)} \right] \exp\left[-\int_0^t \frac{1}{Y_G(s)} dG_1^c(s) \right], \tag{2.7}$$

where G_1^c is the continuous part of G_1, obtained by subtracting the jumps from G_1. The convention is that $0/0 = 0$, and $\exp(-\infty) = 0$.

From Lemma 2.1, it is immediate that on $[0, T]$,

$$S_n = \Phi(f_1(P_n), f_0(P_n)),$$

since the exponential factor is just 1. This tells us how to define our contraction. Let $\not{f}: \mathbf{M} \to \mathbf{S}$ be defined by

$$\not{f} = \Phi \circ (f_1, f_0). \tag{2.8}$$

We begin with the following technical lemmas about $\not{f}$ to ultimately prove the large deviation results.

LEMMA 2.2. *Both f_0: $\mathbf{M} \to \mathbf{SD}$ and f_1: $\mathbf{M} \to \mathbf{SD}$ are continuous when* $\mathbf{SD}$ *has the uniform metric on* $\mathbf{R}_+$.

PROOF. The proof is essentially the proof of Lemma 2.1 of Groeneboom, Oosterhoff and Ruymgaart (1979), and hence will be omitted. $\square$

The next result is Lemma 2 from Gill (1981).

LEMMA 2.3. *Let $G = (G_1, G_0) \in \mathbf{G}$ be fixed and suppose $Y_G(T) > 0$. Then*

$$\|\Phi(G_1, G_0) - \Phi(H_1, H_0)\|_T \to 0$$

as $\max\{\|G_1 - H_1\|, \|G_0 - H_0\|\} \to 0$.

Lemma 2.4 below establishes a certain amount of continuity for the map ℓ and is thus the foundation for the large deviation results of Theorem 3.1. If $A \subset \mathbf{M}$, we define

$$
A^{-} = A \cap \{q \in \mathbf{M}: q([0,T) \times \{0,1\}) < 1\},
$$
$$
A^{+} = A \cup \{q \in \mathbf{M}: q([0,T) \times \{0,1\}) = 1\}. \tag{2.9}
$$

LEMMA 2.4. *If U is open in $\mathbf{S}$ with respect to the pseudometric $\| \ \|_T$, then $[\ell^{-1}(U)]^{-}$ is open in $\mathbf{M}$, and if C is closed in $\mathbf{S}$ with respect to $\| \ \|_T$, then $|\ell^{-1}(C)]^{+}$ is closed in $\mathbf{M}$.*

PROOF. Suppose U is open in $\mathbf{S}$ with respect to $\| \ \|_T$ and let $q \in [\ell^{-1}(U)]^{-}$. Then $1 - [(f_0 + f_1)(q)(T-)] = 1 - q\{[0,T) \times \{0,1\}\} > 0$. But since f_0 and f_1 are continuous, there is a neighborhood N_1 of q such that if $r \in N_1$, then $1 - [(f_0 + f_1)(r)(T-)] = 1 - r\{[0,T) \times \{0,1\}\} > 0$. By Lemma 2.3 and the continuity of f_0 and f_1 once again, there is a neighborhood N_2 of q such that if $r \in N_2$, then $\ell(r) = \Phi(f_1, f_0)(r) \in U$. Thus $N_q = N_1 \cap N_2$ is a neighborhood of q contained in $[\ell^{-1}(U)]^{-}$.

Suppose now that C is closed in $\mathbf{S}$. We have seen that $[\ell^{-1}(C^c)]^{-}$ is open in $\mathbf{M}$, and it is immediate that $[\ell^{-1}(C)^{+}]^c = [\ell^{-1}(C^c)]^{-}$, and hence $[\ell^{-1}(C)]^{+}$ is closed in $\mathbf{M}$. □

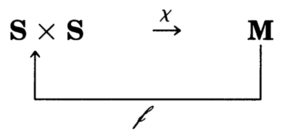
In Section 3 a large deviation rate λ on $\mathbf{S}$ will appear as the infimum of an image of a map χ, which transforms pairs of survival curves (a, σ) to a "censored" law on $\mathbf{R}_{+} \times \{0,1\}$. The map χ gives the law of (X_1, δ_1) when X_1° and Y_1 have survival curves a and σ, respectively. The heuristic description is that the map ℓ undoes the censoring effect of χ.

Lemma 2.5 is essentially Theorem 2.1 of Peterson (1977), although Peterson proved it under the assumption that $f_0(p)$ and $f_1(p)$ do not have jumps at a common point in the interval $[0,T]$. The result is essential for proving the consistency of S_n. If s is a survival curve and σ is a subsurvival curve on $\mathbf{R}_{+}$, we define a new probability measure $\chi(s,\sigma)$ on $\Xi = \mathbf{R}_{+} \times \{0,1\}$ by

$$
\chi(s,\sigma)([0,t] \times \{1\}) = \int 1_{\{0 \le u \le t\}} \sigma(u-)\, d(1-s)(u),
$$

$$
\chi(s,\sigma)([0,t] \times \{0\}) = \int 1_{\{0 \le r \le t\}} s(r)\, d(1-\sigma)(r).
$$

By allowing σ to be a subsurvival curve, we effectively allow a censoring variable with distribution $1 - \sigma$ to take the value ∞ and so the range of χ includes uncensored distributions as well. An imprecise interpretation of ℓ and χ could be the following diagram:



LEMMA 2.5. *Let s be a survival curve and σ a subsurvival curve on $\mathbf{R}_+$. If $s(T-)\sigma(T-) > 0$, then $\mathscr{f}(\chi(s,\sigma)) = s$ on $[0, T]$.*

PROOF. Define the cumulative hazard function G for $F = 1 - s$ by

$$G(t) = \int_{0 \le x \le t} \frac{1}{1 - F(x-)}\, dF(x).$$

Lemma 3.2.1 of Gill (1980) shows that

$$1 - F(t) = \prod_{x \le t} [1 - \Delta G(x)] \exp[-G^c(t)],$$

where G^c is the continuous part of G defined by subtracting the jumps of G. Let $q = \chi(s, \sigma)$. It is enough to show then by the definition of $\mathscr{f}$ at (2.8) and Φ at (2.7) that for $x \in [0, T]$,

$$\Delta G(x) = \frac{\Delta f_1(q)(x)}{1 - f_0(q)(x-) - f_1(q)(x-)}, \tag{2.10}$$

$$G^c(t) = \int_{0 \le x \le t} \frac{1}{1 - f_0(q)(x-) - f_1(q)(x-)}\, df_1(q)^c(x). \tag{2.11}$$

To see (2.10), observe that $\Delta G(x) = \Delta F(x)/[1 - F(x-)]$. Now

$$\begin{aligned}
\Delta f_1(q)(x) &= q([0,x] \times \{1\}) - q([0,x) \times \{1\}) \\
&= \int 1_{0 \le r \le x} \sigma(r-)\, dF(r) - \int 1_{0 \le r < x} \sigma(r-)\, dF(r) \\
&= \sigma(x-)\, \Delta F(x) \\
&= \sigma(x-) s(x-) \frac{1}{1 - F(x-)}\, \Delta F(x) \\
&= [1 - f_0(q)(x-) - f_1(q)(x-)]\, \Delta G(x),
\end{aligned}$$

which proves (2.10). To prove (2.11),

$$\begin{aligned}
&\int_{0 \le x \le t} \frac{1}{1 - f_0(q)(x-) - f_1(q)(x-)}\, df_1(q)^c(x) \\
&\quad = \int_{0 \le x \le t} \frac{1}{1 - f_0(q)(x-) - f_1(q)(x-)}\, df_1(q)(x) \\
&\qquad - \sum_{x \le t} \frac{\Delta f_1(q)(x)}{1 - f_0(q)(x-) - f_1(q)(x-)} \\
&\quad = \int_{0 \le x \le t} \frac{1}{\sigma(x-) s(x-)} \sigma(x-)\, dF(x) - \sum_{x \le t} \Delta G(x) \\
&\quad = G(t) - \sum_{x \le t} \Delta G(x) = G^c(t).
\end{aligned}$$

□

REMARK. Since the law p of (X_i, δ_i) on $\mathbf{R}_+ \times \{0,1\}$ is $\chi(s^\circ, s^y)$, it follows that $\mathscr{f}(p) = s^\circ$ on $[0,T]$ if $s^\circ(T-)s^y(T-) > 0$ or equivalently if $F^x(T-) < 1$.

Lemma 2.5 says that the map $\chi(\cdot,\cdot)$ parametrizes at least part of the set of measures in $\mathbf{M}$ that are sent to s on $[0,T]$ under $\mathscr{f}$:

$$\chi\colon \{(a,\sigma)\colon a = s \text{ on } [0,T],\ a(T-)\sigma(T-) > 0\}$$
$$\to \{q \in \mathbf{M}\colon \mathscr{f}(q) = s \text{ on } [0,T]\}.$$

The range of χ is identified in the following lemma. Let $q_{|T}$ be the restriction of the measure $q \in \mathbf{M}$ to the σ-algebra generated by open sets in $[0,T] \times \{0,1\}$.

LEMMA 2.6. *If* $s \in \mathbf{S}$, *then*

$$\{\chi(a,\sigma)_{|T}\colon a(T-)\sigma(T-) > 0,\ a = s \textit{ on } [0,T]\}$$
$$= \left\{q_{|T}\colon \mathscr{f}(q) = s \textit{ on } [0,T]\right\}^{-}.$$

PROOF. Let L denote the set on the left and R the set on the right. Let a be a survival curve on $\mathbf{R}_+$ such that $a = s$ on $[0,T]$ and suppose $a(T-)\sigma(T-) > 0$. By Lemma 2.5 it follows that $\mathscr{f}(\chi(a,\sigma)) = s$ on $[0,T]$. Furthermore, if $m = \chi(a,\sigma)$, then using Fubini's theorem,

$$m([T,\infty) \times \{0,1\}) = a(T-)\sigma(T-) > 0$$

and therefore $\chi(a,\sigma) \in \{q \in \mathbf{M}\colon \mathscr{f}(q) = s \text{ on } [0,T]\}^-$. Hence $L \subset R$.

Conversely, let $q \in R$. We will produce a survival curve a and a survival curve σ such that $q = \chi(a,\sigma)$ for sets in $[0,T] \times \{0,1\}$, $a = s$ on $[0,T]$ and $a(T-)\sigma(T-) > 0$. Define a survival curve a on $[0,\infty)$ such that

$$a(t) = \mathscr{f}(q)(t), \qquad t \in [0,T].$$

The definition of $a(t)$ is unimportant on (T,∞). Note that since $\mathscr{f}(q) = a$ on $[0,T]$ and $\mathscr{f}(q)(T-) \geq q([T,\infty) \times \{0,1\}) > 0$, it follows that $a(T-) > 0$. Clearly $a = s$ on $[0,T]$. Define σ on $[0,\infty)$ such that

$$\sigma(t) = \frac{1}{\mathscr{f}(q)(t)} q((t,\infty) \times \{0,1\}), \qquad t \in [0,T],$$

where $\sigma(T) = 0$ in the case where $\mathscr{f}(q)(T) = 0$ and $q((T,\infty) \times \{0,1\}) = 0$. The function σ is nonincreasing on $[0,T]$ by the following argument. Define a probability distribution function H on $\mathbf{R}_+$ by $H(t) = q([0,t] \times \{0,1\})$. Then by representing $1 - H$ through its hazard function as in the proof of Lemma 2.5, we see that

$$\begin{aligned} q((t,\infty) \times \{0,1\}) &= 1 - H(t) \\ &= \prod_{s \leq t} \left[1 - \frac{\Delta H(s)}{1 - H(s-)}\right] \exp\left[-\int_0^t \frac{1}{1 - H(s-)}\, dH^c(s)\right]. \end{aligned}$$

But by letting $H_1(t) = q([0,t] \times \{1\})$ we can write

$$\not{f}(q)(t) = \prod_{s \le t}\left[1 - \frac{\Delta H_1(s)}{1 - H(s-)}\right]\exp\left[-\int_0^t \frac{1}{1-H(s-)}\,dH_1^c(s)\right]$$

and now it is clear that $q((t,\infty) \times \{0,1\})$ decays more quickly than $\not{f}(q)(t)$.

Since $q([T,\infty) \times \{0,1\}) > 0$, it follows that $\sigma(T-) > 0$ and therefore $a(T-)\sigma(T-) > 0$. To see that $\chi(a,\sigma)_{|T} = q_{|T}$, observe that for any $t \in [0,T]$ we have $q((t,\infty) \times \{0,1\}) = a(t)\sigma(t)$ and again using Fubini's theorem

$$\begin{aligned} a(t)\sigma(t) &= \int 1_{\{t<u<\infty\}}\sigma(u-)\,d(1-a)(u) + \int 1_{\{t<r<\infty\}}a(r)\,d(1-\sigma)(r) \\ &= \chi(a,\sigma)((t,\infty) \times \{0,1\}). \end{aligned}$$

Thus $\chi(a,\sigma)([0,t] \times \{0,1\}) = q([0,t] \times \{0,1\})$, and on the interval $[0,T]$,

$$f_0(\chi(a,\sigma)) + f_1(\chi(a,\sigma)) = f_0(q) + f_1(q)$$

[cf. (2.4)]. By looking at the definition of $\not{f}$ we see that this implies that $f_1(\chi(a,\sigma)) = f_1(q)$ on $[0,T]$, and hence $f_0(\chi(a,\sigma)) = f_0(q)$ on $[0,T]$. Thus $\chi(a,\sigma) = q$ on sets in $[0,T] \times \{0,1\}$ which proves $R \subset L$. □

3. Large deviation properties of S_n. This section will use the technical results of Section 2 to prove the main results Theorem 3.1 and Corollary 3.1. The notation $K(A,p)$ when $A \subset \mathbf{M}$ will denote the infimum $\inf_{q \in A} K(q,p)$, where $K(q,p)$ is the Kullback–Leibler number

$$K(q,p) = \begin{cases} \int_{\Xi} \log\left[\dfrac{dq}{dp}\right] dq, & \text{if } q \ll p, \\ \infty, & \text{otherwise.} \end{cases}$$

We will also use the conditional Kullback–Leibler number K_T defined by

$$K_T(q,p) = \begin{cases} \int_{\Xi} \log\left[\dfrac{dq}{dp}_{|T}\right] dq, & \text{if } q_{|T} \ll p_{|T}, \\ \infty, & \text{otherwise.} \end{cases}$$

where $dq/dp_{|T}$ denotes the conditional expectation of dq/dp with respect to the σ-algebra generated by open sets in $[0,T] \times \{0,1\}$ and the measure p. Alternatively, $K_T(q,p)$ is the minimum Kullback–Leibler distance over all probability measures which agree with q on sets in $[0,T] \times \{0,1\}$. This second description is a consequence of Jensen's inequality.

When we refer to open and closed sets in $\mathbf{S}$, we do so in this section with respect to the topology from the pseudometric $\| \ \|_T$, since we are interested in studying S_n on the interval $[0,T]$.

LEMMA 3.1. *If U and C are open and closed measurable sets, respectively, in* $\mathbf{S}$, *then*

$$\text{(a)}\qquad \liminf \frac{1}{n} \log \mathbf{P}(S_n \in U,\, X_n^* \geq T) \geq -K\Big(\big[\not{f}^{-1}(U)\big]^-, p\Big),$$

$$\text{(b)}\qquad \limsup \frac{1}{n} \log \mathbf{P}(S_n \in C) \leq -K\Big(\big[\not{f}^{-1}(C)\big]^+, p\Big),$$

where $[\not{f}^{-1}(C)]^+$ *and* $[\not{f}^{-1}(U)]^-$ *are defined at* (2.9).

PROOF. (a) Using the identity $S_n = \not{f}(P_n)$ and the definition at (2.9) of $[\not{f}^{-1}(U)]^-$,

$$\mathbf{P}(S_n \in U,\, X_n^* \geq T) = \mathbf{P}\Big(P_n \in \big[\not{f}^{-1}(U)\big]^-\Big).$$

Then from Lemma 3.1 of Groeneboom, Oosterhoff and Ruymgaart (1979) and Lemma 2.4,

$$\liminf \frac{1}{n} \log \mathbf{P}(S_n \in U,\, X_n^* \geq T) \geq -K\Big(\big[\not{f}^{-1}(U)\big]^-, p\Big).$$

(b) Since $S_n = \not{f}(P_n)$,

$$\mathbf{P}(S_n \in C) = \mathbf{P}\big(P_n \in \not{f}^{-1}(C)\big) \leq \mathbf{P}\Big(P_n \in \big[\not{f}^{-1}(C)\big]^+\Big).$$

Hence from Lemmas 2.4 and 3.1 of Groeneboom, Oosterhoff and Ruymgaart (1979) and Lemma 2.4 above,

$$\limsup \frac{1}{n} \log \mathbf{P}(S_n \in C) \leq -K\Big(\big[\not{f}^{-1}(C)\big]^+, p\Big). \qquad \square$$

Define for each $s \in \mathbf{S}$,

$$\lambda(s) = \inf_{(a,\sigma)} K_T(\chi(a,\sigma), p), \tag{3.1}$$

where the infimum is taken over survival curves a which agree with s on the interval $[0, T]$ and survival curves σ with $a(T-)\sigma(T-) > 0$. If $s(T-) = 0$, then the infimum is taken over the null set and we let $\lambda(s) = \infty$. Let $\Lambda(A) = \inf_{s \in A} \lambda(s)$ when $A \subset \mathbf{S}$.

When X_1° and Y_1 are discrete random variables there is an explicit formula for λ [Dinwoodie (1990)].

THEOREM 3.1. *If U and C are open and closed measurable sets, respectively, in* $\mathbf{S}$, *then*

$$\text{(a)}\qquad \liminf \frac{1}{n} \log \mathbf{P}(S_n \in U,\, X_n^* \geq T) \geq -\Lambda(U),$$

$$\text{(b)}\qquad \limsup \frac{1}{n} \log \mathbf{P}(S_n \in C) \leq -\min\left\{\Lambda(C), \log\left[\frac{1}{F^x(T-)}\right]\right\}.$$

PROOF. First observe by Lemma 2.6 that if s is a survival curve and a and σ are survival curves with $a(T-)\sigma(T-) > 0$ and $a = s$ on $[0, T]$, then

$$\begin{aligned}\lambda(s) &= \inf_{(a,\sigma)} K_T(\chi(a,\sigma), p)\\ &= K_T\Big(\bigcup_{(a,\sigma)} \{\chi(a,\sigma)\}, p\Big)\\ &= K_T\big(\{q \in \mathbf{M}: \mathscr{f}(q) = s \text{ on } [0,T]\}^-, p\big)\\ &= K\big(\{q \in \mathbf{M}: \mathscr{f}(q) = s \text{ on } [0,T]\}^-, p\big).\end{aligned}$$

Now to prove (a), we see from the above that

$$\begin{aligned}\Lambda(U) &= \inf_{s\in U} K\big(\{q \in \mathbf{M}: \mathscr{f}(q) = s \text{ on } [0,T]\}^-, p\big)\\ &= K\Big(\bigcup_{s\in U} \{q \in \mathbf{M}: \mathscr{f}(q) = s \text{ on } [0,T]\}^-, p\Big)\\ &= K\big([\mathscr{f}^{-1}(U)]^-, p\big)\end{aligned}$$

and thus (a) follows from part (a) of Lemma 3.1.

To prove (b), we argue as above that $\Lambda(C) = K([\mathscr{f}^{-1}(C)]^-, p)$. We also have the inequality

$$\min\left\{K\big([\mathscr{f}^{-1}(C)]^-, p\big), \log\frac{1}{F^x(T-)}\right\} \le K\big([\mathscr{f}^{-1}(C)]^+, p\big),$$

since if $q \in [\mathscr{f}^{-1}(C)]^+$ and $q([0,T) \times \{0,1\}) = 1$, then

$$K(q,p) \ge \log[1/F^x(T-)]$$

by Jensen's inequality, whereas if $q \in [\mathscr{f}^{-1}(C)]^+$ and $q([0,T) \times \{0,1\}) < 1$, then $q \in [\mathscr{f}^{-1}(C)]^-$. Together we have

$$\min\left\{\Lambda(C), \log\left[\frac{1}{F^x(T-)}\right]\right\} \le K\big([\mathscr{f}^{-1}(C)]^+, p\big),$$

and so (b) follows from part (b) of Lemma 3.1. □

COROLLARY 3.1. *Assume $F^x(T-) < 1$. For $\varepsilon > 0$ suitably small,*

(a) $$\liminf \frac{1}{n} \log \mathbf{P}(\|S_n - s^\circ\|_T < \varepsilon) \ge -\Lambda(\{s: \|s - s^\circ\|_T < \varepsilon\}),$$

(b) $$\limsup \frac{1}{n} \log \mathbf{P}(\|S_n - s^\circ\|_T \ge \varepsilon) \le -\Lambda(\{s: \|s - s^\circ\|_T \ge \varepsilon\}).$$

PROOF. Let $U = \{s: \|s - s^\circ\|_T < \varepsilon\}$. Then U is open, measurable, and hence (a) follows immediately from part (a) of Theorem 3.1.

Now consider (b). If $s^\circ(T) = 1$, then $S_n(t) = 1$ a.s. for $t \in [0, T]$ and thus $\mathbf{P}(\|S_n - s^\circ\|_T \ge \varepsilon) = 0$ and (b) is true in this case.

Now suppose that $s^\circ(T) < 1$. Since $F^x(T-) < 1$, it follows that $\log[1/F^x(T-)] > 0$. Let $C_\varepsilon = \{s\colon \|s - s^\circ\|_T \geq \varepsilon\}$. We will show that $\lim_{\varepsilon\to 0}\Lambda(C_\varepsilon) = 0$, which will imply

$$\Lambda(C_\varepsilon) = \min\left\{\Lambda(C_\varepsilon), \log\left[\frac{1}{F^x(T-)}\right]\right\}$$

for $\varepsilon > 0$ sufficiently small, and thus (b) will follow from part (b) of Theorem 3.1. Let $\eta > 0$ and let p_η be a probability measure on $\mathbf{R}_+ \times \{0,1\}$ such that $K(p_\eta, p) < \eta$ and such that $\mathscr{S}(p)$ is not identically s° on $[0, T]$. Such a measure can be constructed by making the derivative $dp_\eta/dp \leq 1$ on $[0,T] \times \{0\}$, $dp_\eta/dp \geq 1$ on $[0,T]\times\{1\}$ and $dp_\eta/dp > 1$ on a part of $[0,T]\times\{1\}$ which has positive probability under p. For positive ε sufficiently small that $\varepsilon < \|\mathscr{S}(p_\eta) - s^\circ\|_T$, we have

$$\Lambda(C_\varepsilon) \leq \lambda\big(\mathscr{S}(p_\eta)\big) \leq K(p_\eta, p) < \eta.$$

This shows that $\lim_{\varepsilon\to 0}\Lambda(C_\varepsilon) = 0$ and completes the proof. □

The following lemma shows that part (b) of Corollary 3.1 gives a nontrivial exponential convergence rate of S_n to s°.

LEMMA 3.2. *If $F^x(T-) < 1$, then $\Lambda\{s \in \mathbf{S}\colon \|s - s^\circ\|_T \geq \varepsilon\} > 0$ for every $\varepsilon > 0$.*

PROOF. Let $C = \{s \in \mathbf{S}\colon \|s - s^\circ\|_T \geq \varepsilon\}$. It is enough to show that

$$K\Big(\big[\mathscr{S}^{-1}(C)\big]^+, p\Big) > 0, \tag{3.2}$$

since $\Lambda(C) = K([\mathscr{S}^{-1}(C)]^-, p) \geq K([\mathscr{S}^{-1}(C)]^+, p)$ by Lemma 2.6 and the definition of λ. To show (3.2), it is enough to show that $p \notin [\mathscr{S}^{-1}(C)]^+$, because $K(q,p) = 0$ if and only if $q = p$, and $K([\mathscr{S}^{-1}(C)]^+, p)$ is attained for some element in $[\mathscr{S}^{-1}(C)]^+$ [Groeneboom, Oosterhoff and Ruymgaart (1979), Lemmas 2.2 and 2.3]. Since $F^x(T-) < 1$, it follows that $p([T,\infty)\times\{0,1\}) > 0$, and hence

$$p \in \{q \in \mathbf{M}\colon q([0,T)\times\{0,1\}) < 1\}.$$

Thus it remains to show that $p \notin \mathscr{S}^{-1}(C)$. But this is immediate since $\mathscr{S}(p) = s^\circ \notin C$. □

REMARK. Lemma 3.2 completes the main idea of the paper, which can be roughly summarized as follows. Since $S_n = \mathscr{S}(P_n)$ and $\mathscr{S}$ is reasonably continuous, S_n is distance ε from s° only if P_n is outside a certain neighborhood of p. But the probability of this happening decays exponentially, as we know from the large deviation results of Groeneboom, Oosterhoff and Ruymgaart (1979).

We conclude with the following corollary of Theorem 3.1 and Lemma 3.2. For the Kaplan–Meier estimator set to zero strictly beyond X_n^*, the result is given in Theorem 2.1 of Földes, Rejtő and Winter (1980).

COROLLARY 3.2. *Assume* $F^x(T-) < 1$. *For each* $\varepsilon > 0$, *there exists* $c > 0$ *such that* $\mathbf{P}(\|S_n - s^\circ\|_T \geq \varepsilon) \leq e^{-nc}$ *for* n *sufficiently large. Hence* S_n *converges a.s. to* s° *uniformly on* $[0, T]$.

Acknowledgments. The work presented here is part of a dissertation completed under the supervision of Professor S. L. Zabell. I would like to thank the referees for many valuable comments.

REFERENCES

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

CSÖRGŐ, S. and HORVÁTH, L. (1983). The rate of strong uniform consistency for the product limit estimator. *Z. Wahrsch. Verw. Gebiete* **62** 411–426.

DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations*. Academic, San Diego.

DINWOODIE, I. H. (1990). Large deviations for censored data. Ph.D. thesis, Northwestern Univ.

DINWOODIE. I. H. (1993). Identifying a large deviation rate function. *Ann. Probab.* **21** 216–231.

FÖLDES, A., REJTŐ, L. and WINTER, B. B. (1980). Strong consistency properties of nonparametric estimators for randomly censored data. *Period. Math. Hungar.* **11** 233–250.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math. Centre Tracts* **124**. Math. Centrum, Amsterdam.

GILL, R. D. (1981). Testing with replacement and the product limit estimator. *Ann. Statist.* **9** 853–860.

GILL, R. D. (1983). Large sample behaviour of the product limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586.

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. In *Perspectives in Probability and Statistics* (J. Gani, ed.) 67–87. Applied Probability Trust, Sheffield.

MILLER, R. G. (1981). *Survival Analysis*. Wiley, New York.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

VARADHAN, S. R. S. (1984). *Large Deviations with Applications*. SIAM, Philadelphia.

TULANE UNIVERSITY
DEPARTMENT OF MATHEMATICS
NEW ORLEANS, LOUISIANA 70118