The Annals of Statistics
1986, Vol. 14, No. 3, 1113–1131

ASYMPTOTICALLY MINIMAX ESTIMATORS FOR DISTRIBUTIONS WITH INCREASING FAILURE RATE[1]

BY JANE-LING WANG

University of California, Davis, and University of Iowa

We construct nonparametric estimators of the distribution function F and its hazard function in the class of all increasing failure rate (IFR) distributions. Denoting the empirical distribution and empirical hazard function by F_n and H_n, respectively, let C_n be the greatest convex minorant of H_n, and G_n the distribution with hazard function C_n. The estimator G_n is itself IFR. We prove that under suitable restrictions on F, and for any fixed λ with $F(\lambda) < 1$, $\sup_{x \le \lambda} n^{1/2}|C_n(x) - H_n(x)|$ and $\sup_{x \le \lambda} n^{1/2}|G_n(x) - F_n(x)|$ both tend to zero in probability. This means that G_n and F_n are asymptotically $n^{1/2}$ equivalent. It follows from Millar (1979) that F_n is asymptotically minimax among the class of all IFR distributions for a large class of loss functions. This property extends to our estimator G_n under some restrictions.

1. Introduction. This paper deals with estimation of a cumulative distribution function and its hazard function. Given a set of observations $X_1, \dots, X_n$ from a common distribution function F, the most standard nonparametric estimator of F is the empirical cumulative distribution function F_n. This estimator F_n of F was proved by Dvoretzky, Kiefer, and Wolfowitz (1956) to be asymptotically minimax among the collection of all (continuous) distribution functions. Therefore, in the absence of additional information about the shape of F (except possibly that F is continuous), the empirical distribution function (or a continuous version of it) is the optimal estimator for the true distribution function F in the asymptotically minimax sense.

However, this does not solve the problem of optimal estimation if one has some information about the shape of the true distribution function. Kiefer and Wolfowitz (1976), motivated by questions arising in reliability theory, reopened the issue and proved that the empirical distribution function is still asymptotically minimax among the class of all concave distribution functions. Note that a distribution function is called concave if it is concave on its interval of support. However F_n, not being concave itself, may be considered inappropriate for some purposes. The problem then is to construct an asymptotically minimax estimator which is concave. It follows immediately from Marshall's lemma that the least concave majorant (LCM) C_n of F_n satisfies

$$\sup_t |C_n(t) - F(t)| \le \sup_t |F_n(t) - F(t)|.$$

Received March 1983; revised November 1985.

[1] This research was supported in part by National Science Foundation Grant MCS 78-25301 and Army Research Office Contract DAAC-29-79-C-0093, and was completed in partial fulfillment of requirements for the Ph.D. degree at the University of California, Berkeley.

AMS 1980 *subject classifications*. Primary 62G05, 62N05; secondary 62G20.

Key words and phrases. Asymptotically minimax estimator, increasing failure rate, greatest convex minorant, (empirical) hazard function.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

This indicates that the LCM C_n of F_n is a better estimator of F than F_n, when one considers loss functions of the Kolmogorov–Smirnov type. It is also the maximum likelihood estimator as mentioned in Grenander (1956). The case of convex distribution functions can be dealt with similarly.

One interesting family of distribution functions occurring in reliability theory is the family of distributions with increasing failure rate (IFR). Millar (1979) proved that F_n is asymptotically minimax among the class of all IFR distribution functions. However, F_n has the drawback that it does not have IFR. The maximum likelihood estimate in this family was found by Grenander (1956). Marshall and Proschan (1965) proved the strong consistency of the MLE and Prakasa Rao (1970) obtained the asymptotic distribution of the MLE of the density function. However, since no estimator of the distribution function in the IFR family has previously been shown to achieve optimality of any kind, it is of interest to look for optimal estimators which have IFR. The fact that F_n is an optimal estimator of F and its hazard function H_n (defined in Section 2) is not convex, leads us to consider modifying H_n into a convex function and then to compare the performance of H_n to its modified version. Denote $\alpha_0(F) = \inf\{x: F(x) > 0\}$ and $\alpha_1(F) = \sup\{x: F(x) < 1\}$. If one tries to measure the goodness of fit of H_n to H by $\sup_{t \le \alpha_1}|H_n(t) - H(t)|$, it is obvious that $\lim_{t \to \alpha_1(F)} H(t) = \infty$, unless F has a jump at α_1, and hence $\sup_{t \le \alpha_1(F)}|H_n(t) - H(t)| = \infty$. Therefore, goodness of fit of an estimator of H should only be measured on intervals bounded away from $\alpha_1(F)$. In this paper for any given number $\lambda < \alpha_1(F)$, we shall consider the problem of estimating H only on $[0, \lambda]$. Motivated by the work of Kiefer and Wolfowitz (1976), we shall modify H_n by its greatest convex minorant (GCM) C_n on $[0, \lambda]$, and under certain restrictions $\sup_{t \le \lambda} n^{1/2}|H_n(t) - C_n(t)|$ will tend to zero in probability.

Let $G_n(t) = 1 - e^{-C_n(t)}$ be the distribution with hazard function C_n. This type of estimator (G_n and C_n) will be used throughout Sections 3 and 4. Several other estimators will also be discussed in Sections 3 and 5.

Section 2 gives some notation and the formal construction of the estimators C_n and G_n. Using a series of lemmas, in Theorem 1 of Section 3 we show that under suitable assumptions, $\sup_{t \le \lambda} n^{1/2}|C_n(t) - H_n(t)| \to 0$ in probability. The proof follows the same pattern as the proof of Theorem 1 in Kiefer and Wolfowitz (1976). Later in Theorem 2 of Section 4, the assumptions of Theorem 1 are relaxed to only that of uniform convexity of the true distribution F. The asymptotic $n^{1/2}$ equivalence of G_n and F_n then follows immediately from the fact that

$$\begin{aligned}\sup_{t \le \lambda}|G_n(t) - F_n(t)| &= \sup_{t \le \lambda}|\exp[-C_n(t)] - \exp[-H_n(t)]| \\ &\le \sup_{t \le \lambda}|C_n(t) - H_n(t)|.\end{aligned}$$

Notice that in the construction of our estimate C_n, one does not need to know $\alpha_0(F)$. (See also Remark 3 of Section 4 for a more detailed discussion.) However, $\alpha_1(F)$ or at least a lower bound for it will have to be known in order to make an arbitrary choice of $\lambda < \alpha_1(F)$. In the absence of such information, it would be much nicer if we just take the GCM of $H_n(x)$ over the entire real line. Let $C_n^*(x)$ be the GCM of H_n on the real line and $G_n^*(x)$ the corresponding distribution function. It will be shown in Section 5 that one can use $C_n^*(x)$ instead of $C_n(x)$ (depending on λ). This makes the estimator practically usable. Section 6 shows the asymptotic minimaxity of the estimators. Section 7 summarizes the result of this paper. Appendixes 1 to 4 provide the proofs of some of the lemmas and theorems.

The main effort of this paper is devoted to the proof of Theorem 1 of Section 3, although it seems to be more appropriate to call Theorem 3 of Section 5 the main theorem of the paper.

2. Construction of the estimators C_n, G_n and some notation. We shall start this section with some definitions.

DEFINITION 1. Let F be a distribution function. The hazard function $H_F(t)$ of $F(t)$ is defined to be $H_F(t) = -\log[1 - F(t)]$. If, in addition, F has a density f, the *failure rate function* $\gamma_F(t)$ of $F(t)$ is defined to be $\gamma_F(t) = f(t)/[1 - F(t)]$ for $F(t) < 1$. Note that $\gamma_F(t)$ is the derivative of the hazard function $H_F(t)$.

DEFINITION 2. A distribution function F is said to have *increasing failure rate* if the support of F is an interval denoted by $[\alpha_0(F), \alpha_1(F)]$, and the hazard function $H_F(t)$ is convex on the support of F.

Marshall and Proschan (1965) proved that an IFR distribution F is absolutely continuous except for the possibility of a jump at $\alpha_1(F)$. Hence the failure rate $\gamma_F(t)$ exists (except possibly at $\alpha_1(F)$) and is a nondecreasing function of t.

For the rest of this paper, F will always be a distribution function with IFR. The shape of F is unknown except that it is known to have IFR. F_n will be its empirical distribution function from a sample of independent and identically distributed observations. The notation $H(t)$ will be used to denote $H_F(t)$, and $H_n(t)$ will be used to denote the empirical hazard function $H_{F_n}(t)$, that is, $H_n(t) = -\log[1 - F_n(t)]$. Also α_0, α_1 will be used instead of $\alpha_0(F)$ and $\alpha_1(F)$. Since $\alpha_0 \geq 0$ for most practical applications, we shall assume $\alpha_0 \geq 0$, although all we need is that $\alpha_0 > -\infty$, which will be implied by the uniform strict convexity assumption (A) in Theorems 1 and 2 on H.

Since we shall estimate F through its hazard function H, for reasons mentioned in Section 1 we shall consider the problem of estimating F and H only on $[0, \lambda]$, for any $\lambda < \alpha_1$. Let C_n be the GCM of H_n on $[0, \lambda]$. This GCM C_n is the supremum of the convex functions that are smaller than or equal to H_n on $[0, \lambda]$. (For references on this subject, see the book by Barlow, Bartholomew, Bremner, and Brunk (1972).) Leurgans (1982) and Groeneboom (1983) provide asymptotic distributions of the slope of the GCM of some processes. Our estimate G_n of F on the restricted range is then the distribution function that has C_n for its hazard function. Explicitly,

$$1 - G_n(t) = \exp\{-C_n(t)\} \quad \text{for } t \text{ in } [0, \lambda].$$

This type of estimator C_n and G_n will be used throughout Sections 3 and 4.

Let us now define some notation and a linear interpolating function for any function on $[\alpha_0, \lambda]$.

Let $\{k_n, n \leq 1\}$ be a sequence of positive integers satisfying $k_n \to \infty$.

Partition the interval $[\alpha_0, \lambda]$ into k_n equal length subintervals $[a_j^n, a_{j+1}^n]$, $j = 0, \dots, k_n - 1$, where

$$a_j^n = \frac{Lj}{k_n} + \alpha_0, \qquad j = 0, \dots, k_n,$$

and

$$L = \lambda - \alpha_0.$$

For any function g on $[\alpha_0, \lambda]$, we define its linear interpolating function $L_n g$ as

$$L_n g(a_j^n) = g(a_j^n) \quad \text{for } j = 0, \dots, k_n$$

and linear on $[a_j^n, a_{j+1}^n]$ for $j = 0, \dots, k_n - 1$.

In particular, $L_n H_n$ is the piecewise linear interpolating process of H_n, and $L_n H$ is the linear interpolating function of H.

Let $\|g\|_\lambda = \sup\{|g(t)|;\ \alpha_0 \leq t \leq \lambda\}$.

For simplicity, let us introduce the following notation: Let $\varepsilon_0 = 1 - F(\lambda)$. Then $\varepsilon_0 > 0$ and is fixed. Let $S(t) = 1 - F(t)$ be the survival functions. Then $S(t) \geq \varepsilon_0$ for all t in $[\alpha_0, \lambda]$. Let $S_n(t) = 1 - F_n(t)$ be the empirical survival function. Since $\lambda < \alpha_1$, the failure rate $\gamma_F(t)$ of F has an upper bound on $[\alpha_0, \lambda]$. Let $M = \|\gamma_F\|_\lambda$ denote this upper bound.

3. Asymptotic $n^{1/2}$ equivalence of C_n and H_n. We shall prove the asymptotic $n^{1/2}$ equivalence of C_n and H_n in this section. Recall that as defined in Section 2, C_n is the GCM of H_n on $[0, \lambda]$. An important assumption in many of our results is that there exist a $c > 0$ such that

$$\text{(A)} \qquad H'(v) - H'(u) \geq c(v - u)$$

for any $u < v$, both in $[\alpha_0, \lambda]$, for which the derivatives exist. This assumption can be described as requiring that H be uniformly convex. Later it will also be necessary to assume there is a $d < \infty$ such that

$$\text{(B)} \qquad H'(v) - H'(u) \leq d(v - u)$$

for any $u < v$, both in $[\alpha_0, \lambda]$, for which the derivatives exist.

The proof follows the same pattern as that of Theorem 1 of Kiefer and Wolfowitz (1976), and will be accomplished in the following five steps.

STEP 1.

LEMMA 1. *For any convex function $C(x)$ on $[\alpha_0, \lambda]$,*

$$\|C_n - C\|_\lambda \leq \|H_n - C\|_\lambda.$$

PROOF. The proof of this lemma is similar to that of Marshall's Lemma B (1970), except that Marshall assumes continuity, which is unnecessary. □

STEP 2. Under some restrictions on F, for suitably chosen $k_n = o(n^{1/3})$, $L_n H_n$ (defined in Section 2) is convex with probability tending to 1. More explicitly, let the event $A_n = \{L_n H_n \text{ is convex on } [\alpha_0, \lambda]\}$.

PROPOSITION 1. *If F has IFR and satisfies assumption* (A), *then for sufficiently large n,*

$$1 - \Pr(A_n) \le 2k_n \exp\left[-\frac{nL^3[\varepsilon_0 c]^2}{96Mk_n^3}\right].$$

PROOF. The proof is given in Appendix 1.

REMARK. The same rate $n^{1/3}$ also appears in Lemma 4.1 of Prakasa Rao (1970), and there is some connection between the two results.

STEP 3. Under A_n, using Lemma 1 with $C(x) = L_n H_n(x)$, we have

$$\begin{aligned}\|C_n - H_n\|_\lambda &\le \|C_n - L_n H_n\|_\lambda + \|L_n H_n - H_n\|_\lambda\\ &\le 2\|L_n H_n - H_n\|_\lambda.\end{aligned}$$

STEP 4.

PROPOSITION 2. *If $n^{1/2}\|H - L_n H\|_\lambda \to 0$ in probability, then*

$$n^{1/2}\|H_n - L_n H_n\|_\lambda \to 0 \quad \textit{in probability}.$$

PROOF. The proof is given in Appendix 2.

STEP 5. Under some restrictions on F, for suitably chosen k_n such that $n^{1/4}/k_n = o(1)$, we have $n^{1/2}\|H - L_n H\|_\lambda \to 0$ in probability. This follows from Proposition 3.

PROPOSITION 3. *If F has IFR and satisfies assumption* (B) *then* $\|H - L_n H\|_\lambda \le 2dL^2 k_n^{-2}$.

PROOF. The proof is similar to that of Lemma 6 in Kiefer and Wolfowitz (1976), pages 79–80. □

THEOREM 1. *If F has IFR and satisfies assumptions* (A) *and* (B), *then*

$$n^{1/2}\|H_n - C_n\|_\lambda \to 0 \quad \textit{in probability}.$$

PROOF. Let

$$k_n = \left[\frac{L^3\varepsilon_0^2 c^2 n}{300M \log n}\right]^{1/3}.$$

By Proposition 1, for n large enough, $1 - P(A_n) \le n^{-2}$. Since $n^{1/4}/k_n = o(1)$, the results in Steps 3, 4, and 5 now imply the theorem. □

REMARK 1. As mentioned in the beginning of this section, the proof of Theorem 1 follows the same pattern as that of Theorem 1 of Kiefer and Wolfowitz (1976). The essential step is Proposition 1 which is similar to Lemma 4 of their paper. The definition of the linear interpolating function in Kiefer and Wolfowitz (1976) is a little bit different, but the idea is the same. The conditions in our Theorem 1 are weaker than theirs, but their asymptotic $n^{1/2}$ equivalence holds almost surely while ours holds in probability (which is enough for our purposes). A similar weakening of conditions is possible for their results to hold in probability.

REMARK 2. The assumption of existence of the constant d in Theorem 1 can be deleted by using a certain transformation. This will be done in Theorem 2 of the next section. The existence of the constant c will be guaranteed if $\inf_{\alpha_0 \le t \le \lambda} H''(t) > 0$, which essentially means that H is uniformly convex on $[\alpha_0, \lambda]$. This is also assumed in Kiefer and Wolfowitz (1976) as the condition $\beta(F) > 0$ of (3.2).

REMARK 3. Under the assumption of Theorem 1, Steps 4 and 5 imply that for k_n properly chosen (e.g., as in the proof of Theorem 1), $n^{1/2}\|H_n - L_nH_n\|_\lambda \to 0$ in probability. Step 2 then implies that the probability that L_nH_n is convex on $[\alpha_0, \lambda]$ approaches 1. Hence, instead of using C_n (which involves n points), we can use L_nH_n (which only involves k_n points) as the estimator. The advantage of using L_nH_n is that it is much easier to compute than C_n, and by Proposition 1, with probability tending to 1, L_nH_n will be convex. If not, we can then take the GCM of L_nH_n, which is still asymptotically $n^{1/2}$ equivalent to H_n and easier to compute than C_n itself.

It should be noted that although the construction of L_nH_n depends on α_0 (while C_n does not), in the case when α_0 is unknown, one can modify it by the linear interpolating process of H_n on $[X_{(1)}, \lambda]$, where $X_{(1)}$ is the smallest among the X_i's. This will be clear from the proof of Theorem 1.

4. Extension of Theorem 1 by a convex transformation. We shall show in this section that the assumption of existence of the constant d in Theorem 1 can be deleted by use of the convex transformation ϕ^{-1}, where ϕ is defined as in Lemma 2.

LEMMA 2. *If F has IFR and satisfies assumption* (A) *then there is a convex increasing function K defined on some interval $[\alpha_0, \lambda']$ with the following properties*:

(1) *K is twice differentiable with bounded second derivative and* $\inf\{K''(x): \alpha_0 \le x \le \lambda'\} > 0$.
(2) $\phi(x) = H^{-1}(K(x))$ *is a concave function from $[\alpha_0, \lambda']$ to $[\alpha_0, \lambda]$.*

PROOF. The proof will be given in Appendix 3.

THEOREM 2. *If F has IFR and satisfies assumption* (A), *then*

$$n^{1/2}\|H_n - C_n\|_\lambda \to 0 \quad \textit{in probability}.$$

PROOF. By Lemma 2, there is a convex function K on $[\alpha_0, \lambda']$ with properties (1) and (2) in Lemma 2. Let $\phi(x) = H^{-1}(K(x))$ as in Lemma 2. If X is a random variable with distribution function F and hazard function H, and if $Y = \phi^{-1}(X)$, then $P(Y > y) = e^{-H(\phi(Y))} = e^{-K(y)}$.

Hence Y has K as its hazard function and K satisfies all the assumptions in Theorem 1.

Let $X_1, \dots, X_n$ be independent identically distributed samples from F, and let H_n be their empirical hazard function. Let $Y_i = \phi^{-1}(X_i)$, $1 \le i \le n$, and $K_n(y)$ be the corresponding empirical hazard function from Y. Then $H_n(x) = K_n(\phi^{-1}(x))$.

Let D_n be the GCM of K_n on $[0, \lambda']$. Applying Theorem 1, we have

$$n^{1/2}\|D_n - K_n\|_{\lambda'} \to 0 \quad \text{in probability.} \tag{4.1}$$

Let $C^*(x) = D_n(\phi^{-1}(x))$ on $[\alpha_0, \lambda]$. From Lemma 2, C^* is a convex function. Using the fact that $H_n(x) = K_n(\phi^{-1}(x))$ and D_n is a minorant of K_n, we have C^* is a minorant of H_n. Moreover, $C^*(x) - H_n(x) = D_n(\phi^{-1}(x)) - K_n(\phi^{-1}(x))$ for all x. Hence $\|C^* - H_n\|_\lambda \le \|D_n - K_n\|_{\lambda'}$. Since C_n is the GCM of H_n and C^* is a convex minorant of H_n, we have

$$\|C_n - H_n\|_\lambda \le \|C^* - H_n\|_\lambda \le \|D_n - K_n\|_{\lambda'}.$$

From (4.1), the theorem follows. □

REMARK 1. Theorem 2 implies that the only condition needed to guarantee the asymptotic $n^{1/2}$ equivalence of C_n and H_n is the uniform convexity of H (assumption (A)). This is much weaker than the conditions required in Kiefer and Wolfowitz (1976). A similar weakening of conditions is also possible for their other results by using a similar transformation to that in Theorem 2.

REMARK 2. So far λ has been kept fixed so that the survival function is strictly greater than zero at λ. From the proofs of Propositions 1–3, one can see that it is sometimes not necessary to restrict λ to be fixed. For example, if F has a jump at α_1 and satisfies the assumption of Theorem 2, on $[0, \alpha_1]$ the estimator $C_n(x)$ can be taken to be the GCM of $H_n(x)$ on $[0, \alpha_1]$, and we have $\sup_{0 \le x < \alpha_1} n^{1/2}|H_n(x) - C_n(x)| \to 0$ in probability. Also under the assumption of Theorem 2, the estimator C_n can be taken to be the GCM of H_n on $[0, \lambda_n]$, where $\lambda_n < \alpha_1$ and λ_n tends to α_1 at a suitably slow rate (depending on F). Then we have $n^{1/2}\|H_n - C_n\|_{\lambda_n} \to 0$ in probability.

REMARK 3. We have assumed $\alpha_0 \ge 0$ so far. If $\alpha_0 < 0$ and F satisfies assumption (A), it is clear that $\alpha_0 > -\infty$. Under such circumstances, C_n can be constructed as the GCM of H_n on $[-\infty, \alpha]$ and the result of Theorem 2 still holds. Therefore, in constructing C_n, one needs to know α_1 but not α_0. In the next section, we shall show that even α_1 need not be known.

5. A practical way of constructing the estimators. The construction of C_n in the previous three sections depends on the knowledge of α_1, so as to make an arbitrary choice of λ which is less than α_1. In most real life examples, one does not know α_1. A practical way of constructing the estimator is to take the GCM of $H_n(x)$ over the entire real line. Let us call this GCM $C_n^*(x)$. Let $C_n^\lambda(x)$ denote the GCM of $H_n(x)$ on $[0, \lambda]$, $\lambda < \alpha_1$; that is, C_n^λ is the estimator constructed in Section 2 and used in Sections 3 and 4. The conclusion of Proposition 4 is that for any fixed $a < \lambda$,

$$\Pr\{C_n^\lambda(x) = C_n^*(x) \text{ for all } x \le a\} \to 1.$$

That is, with high probability, the GCM of $H_n(x)$ on the entire support will coincide with the GCM of $H_n(x)$ on $[0, \lambda]$ for x in $[0, a]$.

This suggests that instead of taking our estimator $C_n^\lambda(x)$ as in Sections 2–4, we can simply take the GCM $C_n^*(x)$ of $H_n(x)$ on the entire support and $C_n^*(x)$ will behave just as well as $C_n^\lambda(x)$. In order to formally state Proposition 4, let us introduce the following definition.

DEFINITION 3. A function ϕ on $[a, b]$ is called *strictly convex* if for any $0 < \varepsilon < 1$ and any x, y in $[a, b]$, the following is true:

$$\phi((1-\varepsilon)x + \varepsilon y) < (1-\varepsilon)\phi(x) + \varepsilon\phi(y).$$

NOTE. (1) If F has IFR and satisfies the assumption of Theorem 2, then H is strictly convex.

(2) For a strictly convex function ϕ on $[a, b]$ and any three points $x < y < z$ in $[a, b]$, we have

$$\frac{\phi(y) - \phi(x)}{y - x} < \frac{\phi(z) - \phi(x)}{z - x} < \frac{\phi(z) - \phi(y)}{z - y}.$$

PROPOSITION 4. *If F has IFR and H is strictly convex, then for any $\lambda < \alpha_1$ and any $a < \lambda$,*

$$\Pr\{C_n^\lambda(x) = C_n^*(x) \textit{ for all } x \le a\} \to 1.$$

PROOF. The proof is given in Appendix 4.

Theorem 2 and Proposition 4 now imply the main theorem of the paper.

THEOREM 3. *If F has IFR and there exists a positive constant $c > 0$ such that $H'(v) - H'(u) \ge c(v - u)$ for all $u < v$ in $[\alpha_0, \alpha_1]$ for which the derivative exists, then $n^{1/2}\|H_n - C_n^*\|_\lambda \to 0$ in probability for all $\lambda < \alpha_1$.*

REMARK. Note that the assumption is slightly different from assumption (A).

6. Asymptotic minimaxity of the estimator. Let $\mathscr{F}_\lambda = \{F\colon F$ has IFR with $\alpha_1(F) > \lambda\}$. The asymptotic minimaxity of H_n as an estimator of the true hazard function H restricted on $[-\infty, \lambda]$ among the class $\mathscr{F}_\lambda$ follows along the same line as in Millar (1979). Lemma 1 then implies that this optimality property also extends to the estimator C_n^λ for loss functions which satisfy the assumptions of Millar (1979) and is of the type $l(n^{1/2}\|C_n^\lambda - H\|_\lambda)$ where l is a bounded continuous nondecreasing function. Thus we have

THEOREM 4. *C_n^λ is the asymptotically minimax estimator of the hazard function H among the class $\mathscr{F}_\lambda$.*

As for the estimator C_n^*, Theorem 4 does not apply to C_n^* directly because Lemma 1 fails for C_n^*. Although Theorem 3 implies

$$n^{1/2}\|C_n^* - H_n\|_\lambda \to 0 \quad \text{in probability,} \tag{6.1}$$

the convergence in (6.1) is not uniform over distributions satisfying the assumption of Theorem 3, and there does not seem to be a natural way to restrict the family of distributions in order to obtain uniform convergence of (6.1).

Let G_n^λ, G_n^* be the distribution functions with hazard functions C_n^λ and C_n^*, respectively. It follows from Theorems 1, 2, and 3 that:

COROLLARY 1. (a) *$n^{1/2}\|G_n^\lambda - F_n\|_\lambda \to 0$ in probability under the assumptions of Theorem 2.*

(b) *$n^{1/2}\|G_n^* - F_n\|_\lambda \to 0$ in probability under the assumption of Theorem 3.*

To obtain the asymptotic minimaxity of G_n^* or G_n^λ as an estimator of the true distribution F we encounter the same difficulty for G_n^* as for C_n^*. Therefore we shall only focus on the behavior of G_n^λ. Since uniform convergence is hard to grasp by Theorem 2, we shall restrict ourselves only to distributions which satisfy the assumption of Theorem 1. Checking the proof of Theorem 1, we find that in order to get the uniform convergence of $n^{1/2}\|G_n^\lambda - F_n\|_\lambda \to 0$, it suffices to show that both Proposition 1 and $n^{1/2}\|H - L_nH\|_\lambda \to 0$ hold uniformly in F, which is equivalent to the requirement that both $M[\varepsilon_0 C]^{-2}$ and dL^2 are uniformly bounded from above by some constants. Consider now the restricted family $\mathscr{S}_\lambda$ of IFR distributions which satisfies the assumption of Theorem 1 with $M[\varepsilon_0 C]^{-2}$ and dL^2 both uniformly bounded from above by some constants. We then have $\sup_{F \in \mathscr{S}_\lambda} n^{1/2}\|H_n - C_n^\lambda\|_\lambda \to 0$ in probability, which implies

$$\sup_{F \in \mathscr{S}_\lambda} n^{1/2}\|F_n - G_n^\lambda\|_\lambda \to 0 \quad \text{in probability.} \tag{6.2}$$

Using again the technique of Proposition 6.2 of Millar (1979) it can be checked without difficulty that F_n is still asymptotically minimax among the restricted class $\mathscr{S}_\lambda$. The asymptotic minimaxity of G_n^λ now follows from (6.2) for a loss function which satisfies the assumptions of Millar (1979) and is of the type $l(n^{1/2}\|G_n^\lambda - F\|_\lambda)$, where l is a bounded continuous nondecreasing function.

THEOREM 5. *G_n^λ is the asymptotically minimax estimator of the true distribution function F among the class $\mathscr{S}_\lambda$.*

7. Summary. We have discussed several possible estimators for the hazard function H. A summary of these estimators is given below. All the following results require that H be uniformly convex.

1. If α_1 is known, then for any $\lambda < \alpha_1$, the estimator C_n^λ can be taken to be the GCM of H_{nn} on $[0, \lambda]$ and C_n^λ is asymptotically $n^{1/2}$ equivalent to H_n on $[0, \lambda]$ by Theorem 2.
2. If α_1 is unknown, the estimator C_n^* can be taken to be the GCM of H_n on the entire support of F_n, and C_n^* is asymptotically $n^{1/2}$ equivalent to H_n on $[0, \lambda]$, for any $\lambda < \alpha_1$ by Theorem 3. In particular, if F has a jump at α_1 then C_n^* is asymptotically $n^{1/2}$ equivalent to H_n on $[0, \alpha_1]$ and the corresponding estimator G_n for the distribution function is asymptotically $n^{1/2}$ equivalent to F_n on the whole real line.
3. Under the additional assumption of (B), if we want to save time in computing the estimator for H, then for k_n properly chosen (for example, as in the proof of Theorem 1), one can take L_nH_n to be the estimator if it is convex and otherwise, take the GCM of L_nH_n to be the estimator. In either case, Theorem 1 implies that the adopted estimator will be asymptotically $n^{1/2}$ equivalent to H_n on $[0, \lambda]$ for any $\lambda < \alpha_1$.
4. It is clear that

$$\sup_{\lambda \le x \le \alpha_1} |C_n^*(x) - H_n(x)| = \infty, \quad \text{if } F \text{ is continuous.}$$

Thus, C_n^* will not be close to H_n near the right-hand tail. However, this does not prevent G_n^* from being close to F_n near the right-hand tail. The question is, under what conditions will $\sup_{-\infty < x < \infty} n^{1/2}|G_n^*(x) - F_n(x)| \to 0$ in probability. This remains an open problem.

A final remark is that the technique in this paper can also be applied to distributions with decreasing failure rate by taking the estimator C_n to be the least concave majorant of H_n.

APPENDIX 1

Proof of Proposition 1. Let $x < y < z$ be any three equally spaced points in $[\alpha_0, \lambda]$, such that $y - x = z - y = L/k_n$.
Let $p = S(y)/S(x)$, $q = S(z)/S(x)$; then $0 \le q \le p \le 1$.

LEMMA 3. *Let X be a binomial random variable $B(n, \mu)$ with $\mu \le \frac{1}{2}$. Then for $t \ge 0$,*

$$\Pr\{X - n\mu \ge t\} \le \exp\left\{-\frac{1}{2}\frac{t^2}{n\mu(1-\mu)}\right\},$$

$$\Pr\{X - n\mu \le -t\} \le \exp\left\{-\frac{1}{2}\frac{t^2}{n(1-\mu)+t}\right\}.$$

PROOF. The proof follows from Hoeffding (1963). □

LEMMA 4. *If F has IFR, then $q \geq \frac{1}{2}$ for n sufficiently large.*

PROOF.

$$q = \frac{S(z)}{S(x)} = 1 - \frac{S(x) - S(z)}{S(x)}$$

and

$$\begin{aligned} \frac{S(x) - S(z)}{S(x)} &\leq \frac{|S(x) - S(z)|}{\varepsilon_0} \\ &\leq 2LM(k_n\varepsilon_0)^{-1} \quad (\text{since } f(t) \leq r(t) \leq M) \\ &< \tfrac{1}{2} \quad \text{for large } n. \end{aligned}$$

Hence $q \geq \frac{1}{2}$ for n large enough. □

LEMMA 5. *If F has IFR and satisfies assumption* (A), *then $S^2(y) - S(x)S(z) \geq cL^2k_n^{-2}S(x)S(z)$.*

PROOF.

$$\begin{aligned} H(z) - H(y) &= \int_0^{z-y} H'(y+t)\,dt \\ &= \int_0^{Lk_n^{-1}} H'(y+t)\,dt. \end{aligned}$$

Similarly,

$$H(y) - H(x) = \int_0^{Lk_n^{-1}} H'(x+t)\,dt.$$

Using these and the existence of c,

$$\begin{aligned} \Delta &= [H(z) - H(y)] - [H(y) - H(x)] \\ &= \int_0^{Lk_n^{-1}} [H'(y+t) - H'(x+t)]\,dt \\ &\geq \int_0^{Lk_n^{-1}} c(y-x)\,dt \\ &= cL^2k_n^{-2}. \end{aligned}$$

On the other hand,

$$\Delta = \log\frac{S(y)}{S(z)} - \log\frac{S(x)}{S(y)} = \log\frac{S^2(y)}{S(x)S(z)}.$$

Hence,

$$S^2(y)[S(x)S(z)]^{-1} \geq \exp[cL^2k_n^{-2}].$$

This implies

$$S^2(y) - S(x)S(z) \ge S(x)S(z)\left[\exp\left(cL^2k_n^{-2}\right) - 1\right]$$
$$\ge S(x)S(z)\left(cL^2k_n^{-2}\right). \qquad \square$$

PROOF OF PROPOSITION 1. Since L_nH_n is linear on each of the k_n equal length intervals $[a_j^n, a_{j+1}^n]$, $j = 0, \dots, k_n - 1$,

$$\begin{aligned} A_n &= \bigcap_{j=0}^{k_n-2} \left\{H_n\left(a_{j+1}^n\right) - H_n\left(a_j^n\right) \le H_n\left(a_{j+2}^n\right) - H_n\left(a_{j+1}^n\right)\right\} \\ &= \bigcap_{j=0}^{k_n-2} \left\{S_n\left(a_j^n\right)S_n\left(a_{j+2}^n\right) \le S_n^2\left(a_{j+1}^n\right)\right\} \\ &= \bigcap_{j=0}^{k_n-2} B_{nj}, \end{aligned}$$

where B_{nj} is the event in the above bracket.

Fix some j and let $x = a_j^n$, $y = a_{j+1}^n$, $z = a_{j+2}^n$. Let $B_n(t) = nS_n(t)$ for t in $[\alpha_0, \lambda]$. Then $B_n(t)$ is distributed as a binomial random variable $B(n, S(t))$.

Given $B_n(x) = N$, $B_n(y)$ has binomial distribution $B(N, p)$ where $p = S(y)/S(x)$ as defined above. Given $B_n(x) = N$, $B_n(z)$ has binomial distribution $B(N, q)$ where $q = S(z)/S(x)$.

By Lemma 4, for n large enough, $p \ge q \ge \frac{1}{2}$.

Let $U = Np - B_n(y)$ and $V = B_n(z) - Nq$. Let B_{nj}^c be the complement of B_{nj}. For n large enough, consider the conditional probability

$$\begin{aligned} P\left\{B_{nj}^c | B_n(x) = n\right\} &= \Pr\left\{S_n^2(y) < S_n(x)S_n(z) | B_n(x) = N\right\} \\ &= \Pr\left\{B_n^2(y) < B_n(x)B_n(z) | B_n(x) = N\right\} \\ &= \Pr\left\{(Np - U)^2 < N(Nq + V) | B_n(x) = N\right\} \\ &\le \Pr\left\{(Np)^2 - 2UNp < N^2q + NV | B_n(x) = N\right\} \\ &= \Pr\left\{N\left(p^2 - q\right) < (V + 2Up) | B_n(x) = N\right\} \\ &\le \Pr\left\{N\left(p^2 - q\right) < V + 2U | B_n(x) = N\right\}. \end{aligned} \tag{A1.1}$$

Let $t = \frac{1}{3}N(p^2 - q)$. Since F has IFR,

$$\Delta = [H(z) - H(y)] - [H(y) - H(x)] \ge 0.$$

But

$$\Delta = \log \frac{S^2(y)}{S(x)S(z)};$$

hence $S^2(y) \ge S(x)S(z)$ and $p^2 \ge q$. This implies $t \ge 0$. Also $p^2 - q \le 1 - q$; hence $0 \le t \le \frac{1}{3}N(1 - q)$.

Putting t into our calculation (A1.1), we have

$$\begin{aligned}
&\Pr\{3t < V + 2U | B_n(x) = N\} \\
&\quad \le \Pr\{V > t \text{ or } U > t | B_n(x) = N\} \\
&\quad \le \exp\left\{-\frac{1}{2}\frac{t^2}{Nq(1-q)}\right\} + \exp\left\{-\frac{1}{2}\frac{t^2}{N(1-p)+t}\right\} \quad \text{(by Lemma 3)} \\
&\quad \le \exp\left\{-\frac{1}{2}\frac{t^2}{N(1-q)}\right\} + \exp\left\{-\frac{1}{2}\frac{t^2}{N(1-q)+\frac{1}{3}N(1-q)}\right\} \\
&\quad \le 2\exp\left\{-\frac{1}{2}\frac{t^2}{\frac{4}{3}N(1-q)}\right\} \\
&\quad = 2\exp\left\{-\frac{N(p^2-q)^2}{24(1-q)}\right\}.
\end{aligned}$$

Let $a = (p^2 - q)^2/(24(1-q))$. Then $0 \le a \le 1/24$. We have proved so far that, for n large enough,

$$\Pr\{B_{nj}^c | B_n(x) = N\} \le 2\exp\{-aN\}. \tag{A1.2}$$

Now

$$\begin{aligned}
\Pr\{B_{nj}^c\} &= E\{\Pr[S_n^2(y) < S_n(x)S_n(z) | B_n(x)]\} \\
&\le E\{2\exp[-aB_n(x)]\} \quad \text{(from (A1.2))} \\
&= 2[1 - S(x) + S(x)e^{-a}]^n \quad \text{(since } B_n(x) \text{ is distributed as } B(n, S(x))) \\
&= 2[1 - S(x)(1 - e^{-a})]^n \\
&\le 2\exp[-nS(x)(1 - e^{-a})] \quad \left(\text{because } \left(1 - \frac{\xi}{n}\right)^n \le e^{\xi}, \text{ for } 0 \le \xi \le n\right) \qquad \text{(A1.3)} \\
&\le 2\exp\left[-nS(x)\left(a - \frac{a^2}{2}\right)\right] \\
&\le 2\exp\left[-\frac{a}{2}nS(x)\right] \quad \text{(since } a < 1).
\end{aligned}$$

Consider

$$\begin{aligned}
aS(x) = \frac{(p^2 - q)^2 S(x)}{24(1-q)} &= \frac{[S^2(y) - S(x)S(z)]^2}{24[S(x) - S(z)]S^2(x)} \\
&\ge \frac{[L^2 c S(x) S(z) k_n^{-2}]^2}{24[S(x) - S(z)]S^2(x)} \quad \text{(by Lemma 5)} \\
&\ge \frac{[\varepsilon_0 L^2 c]^2}{24 k_n^4 [F(z) - F(x)]} \quad \text{(since } S(x) \ge \varepsilon_0).
\end{aligned}$$

Now

$$F(z) - F(x) = \int_x^z f(t)\,dt \le 2LMk_n^{-1}.$$

Hence

$$aS(x) \ge \frac{L^3[\varepsilon_0 c]^2}{48Mk_n^3}. \tag{A1.4}$$

From (A1.3) and (A1.4)

$$\Pr\{B_{nj}^c\} \le 2\exp\left\{-\frac{nL^3[\varepsilon_0 c]^2}{96Mk_n^3}\right\}.$$

Hence

$$1 - \Pr(A_n) = \Pr\left(\bigcup_{j=0}^{k_n-2} B_{nj}^c\right)$$
$$\le 2k_n \exp\left\{-\frac{nL^3[\varepsilon_0 c]^2}{96Mk_n^3}\right\}. \qquad \square$$

APPENDIX 2

Proof of Proposition 2.

LEMMA 6. *Let $X_n(t) = n^{1/2}[H_n(t) - H(t)]$. Given $\varepsilon > 0$, there exist $\delta > 0$ and an integer N_1 such that for all $n \ge N_1$,*

$$\Pr[\sup\{|X_n(t) - X_n(s)|;\ \textit{for all } |t-s| < \delta \textit{ and } t, s, \textit{ in } [\alpha_0, \lambda]\} > \varepsilon] < \varepsilon.$$

PROOF. It can be checked by standard procedures that the process $X_n(t)$ converges weakly to a Gaussian process $Z(t)$ with continuous paths in $D[\alpha_0, \lambda]$, where $EZ(t) = 0$, $\operatorname{Var} Z(t) = F(t)[1 - F(t)]^{-1}$, and $\operatorname{Cov}(Z(s), Z(t)) = F(s)[1 - F(s)]^{-1}$ if $s < t$. Since $Z(t)$ has continuous paths on $D[\alpha_0, \lambda]$ with probability 1, the result follows from the tightness conditions on Skorokhod topology. $\square$

PROOF OF PROPOSITION 2. Given $\varepsilon > 0$, since $k_n \to \infty$ there exists N_2 such that for all $n \ge N_2$, $Lk_n^{-1} < \delta$ for the δ in Lemma 6.

Let S_nH be the piecewise shifted H. That is,

$$S_nH(t) = H(t) + [H_n(a_j^n) - H(a_j^n)] \quad \text{for } a_j^n \le t < a_{j+1}^n.$$

Then

$$S_nH(a_j^n) = H_n(a_j^n) \quad \text{for } j = 0, \dots, k_n.$$

Consider

$$n^{1/2}(H_n - S_nH)(t) = n^{1/2}\{[H_n(t) - H(t)] - [H_n(a_j^n) - H(a_j^n)]\}$$
$$= X_n(t) - X_n(a_j^n) \quad \text{for } a_j^n \le t < a_{j+1}^n.$$

Then

$$n^{1/2}\|H_n - S_nH\|_\lambda = \sup_{0 \le j \le k_n - 1} \sup_{a_j^n \le t < a_{j+1}^n} |X_n(t) - X_n(a_j^n)|$$
$$\le \sup_{|t-s|<\delta} |X_n(t) - X_n(s)| \quad \text{if } n > N_2.$$

Lemma 6 now implies

$$n^{1/2}\|H_n - S_nH\|_\lambda \to 0 \quad \text{in probability.} \tag{A2.1}$$

Recall the definition of L_n, the linear interpolation in Section 2:

$$L_nH_n(a_j^n) = H_n(a_j^n) = S_nH(a_j^n) = L_n(S_nH(a_j^n)), \qquad j = 0, \dots, k_n.$$

Since L_n is the piecewise linear interpolation process,

$$L_nH_n(t) = L_n(S_nH(t)) \quad \text{for all } t.$$

We have

$$(H_n - L_nH_n)(t) = (H_n - S_nH)(t) + (S_nH - L_nS_nH)(t)$$
$$+ (L_nS_nH - L_nH_n)(t)$$
$$= (H_n - S_nH)(t) + (S_nH - L_nS_nH)(t)$$
$$= (H_n - S_nH)(t) + (H - L_nH)(t),$$

since S_nH is the piecewise shifted H.

If $n^{1/2}\|H - L_nH\|_\lambda \to 0$ in probability, by (A2.1) this will imply $n^{1/2}\|H_n - L_nH_n\|_\lambda \to 0$ in probability. □

APPENDIX 3

Proof of Lemma 2.

PROOF. From the definition of IFR distributions, H^{-1} exists provided one defines $H^{-1}(0) = \alpha_0$. Thus $H^{-1}(y)$ is concave on $[0, H(\lambda)]$ and differentiable except for countably many points. Let N_0 be the exceptional set where $H^{-1}(y)$ is not differentiable. Let J be the complement of N_0, that is, $J = [0, H(\lambda)] - N_0$ and f be the derivative of H^{-1} on J. Since H' is bounded on $H^{-1}(J)$, we have $f(x) > \delta$ for some $\delta > 0$ and all x in J, and $f(u) = [H'(H^{-1}(u))]^{-1} >$

$[H'(H^{-1}(v))]^{-1} = f(v)$ for any $u < v$ on J. Also,

$$\begin{aligned} f(u) - f(v) &= [H'(H^{-1}(u))]^{-1} - [H'(H^{-1}(v))]^{-1} \\ &= \frac{H'(H^{-1}(v)) - H'(H^{-1}(u))}{H'(H^{-1}(u))H'(H^{-1}(v))} \\ &\geq C[H^{-1}(v) - H^{-1}(u)]\, f(u)f(v) \\ &\geq C[(v-u)f(v)]\, f(u)f(v). \end{aligned}$$

The last step is due to the fact that f is decreasing on J. Therefore,

$$\text{(A3.1)} \qquad \frac{f(u) - f(v)}{(v-u)f(v)} \geq cf(u)f(v) \geq c\delta^2$$

for any pair $u < v$ in J.

We are now looking for a decreasing linear function $g(u) = b - au$ with $g(u) > 0$ on $[0, H(\lambda)]$ so that the ratio $f(u)/g(u)$ will be decreasing on J. This can be done in the following way.

Given $b > 0$, there exists $a > 0$ such that $b - aH(\lambda) > 0$ and $0 < a < c\delta^2[b - aH(\lambda)]$. From (A3.1)

$$a < \frac{f(u) - f(v)}{(v-u)f(v)}[b - aH(\lambda)] \leq \frac{f(u) - f(v)}{(v-u)f(v)}[b - av]$$

for all $u < v$ in J. Hence

$$[f(u) - f(v)][b - av] > af(v)(v-u),$$

and this implies

$$\frac{f(u)}{b - au} > \frac{f(v)}{b - av} \quad \text{for } u < v \text{ in } J.$$

Hence, we have found g such that $f(u)/g(u)$ is decreasing on J and $g(u) - g(v) = a(v-u)$. Let $G(y) = \alpha_0 + \int_0^y g(u)\,du$. Then G is concave and $G'' = -a$ on $[0, H(\lambda)]$. Let $K(x) = G^{-1}(x)$ on $[\alpha_0, \lambda']$, where $[\alpha_0, \lambda']$ is the range of G on $[0, H(\lambda)]$. Then K is convex and K'' exists on $[\alpha_0(F), \lambda']$. Moreover

$$K''(x) = [G^{-1}(x)]'' = -\frac{G''(G^{-1}(x))}{[G'(G^{-1}(x))]^3} = a[g(G^{-1}(x))]^{-3}.$$

This means $K''(x) \geq ab^{-3} > 0$ for all x in $[\alpha_0, \lambda']$ and K'' has a finite upper bound because g is bounded away from zero. Part (1) has thus been proved.

Let $\phi(x) = H^{-1}(K(x))$. From the way K is defined, ϕ is a function from $[\alpha_0, \lambda']$ to $[\alpha_0, \lambda]$. Note that $\phi'(x)$ will exist if $K'(x)$ exists, and at any point x where $K'(x)$ exists,

$$\phi'(x) = \frac{K'(x)}{H'[H^{-1}(K(x))]} = \frac{f(K(x))}{g(K(x))}.$$

Hence, $\phi'(x)$ exists and is a decreasing function on $[\alpha_0, \lambda']$ except for countably many points. This proves the concavity of ϕ and hence part (2). $\square$

APPENDIX 4

Proof of Proposition 4.

PROOF. The proof will be separated into two parts.

(i) This is the case where F has a jump at α_1 with $1 - F(\alpha_1 -) > 0$. In the proof of Lemma 6 in Appendix 2, we showed that $\sqrt{n}(H_n(x) - H(x))$ converges in distribution (in the Prohorov sense) to a Gaussian process $Z(x)$. Let $\| \ \|_{\alpha_1}$ denote the supremum norm$_\alpha$ of a function on $[\alpha_0, \alpha_1)$. By Donsker's invariance principle (1952), $n^{1/2}\|H_n - H\|_{\alpha_1}$ is bounded in probability. For any $\varepsilon > 0$, there exists N large enough such that $\mathrm{pr}\{\|H_n - H\|_{\alpha_1} < \varepsilon\} > 1 - \varepsilon$ for all $n \geq N$. Let E_n be the event that $\|H_n - H\|_{\alpha_1} < \varepsilon$. Then we have $P(E_n) > 1 - \varepsilon$ for $n \geq N$. Let b be a fixed point between a and λ. Let m be the right-hand derivative of $C_n^\lambda(x)$ at $x = a$. Under E_n, $H_n(x)$ lies entirely within the band $\{H(x) - \varepsilon, H(x) + \varepsilon\}$. Hence

$$m < \frac{[H(b) + \varepsilon] - [H(a) - \varepsilon]}{b - a} = \frac{H(b) - H(a) + 2\varepsilon}{b - a}.$$

On the other hand, for any $y \geq \lambda$,

$$\begin{aligned} \frac{H_n(y) - H_n(a)}{y - a} &> \frac{H(y) - H(a) - 2\varepsilon}{y - a} \\ &\geq \frac{H(y) - H(a)}{y - a} - \frac{2\varepsilon}{\lambda - a} \\ &> \frac{H(\lambda) - H(a)}{\lambda - a} - \frac{2\varepsilon}{\lambda - a}. \end{aligned}$$

If ε is small enough, by strict convexity of H we have

$$[H(\lambda) - H(a)]/(\lambda - a) - 2\varepsilon/(\lambda - a) > [H(b) - H(a) + 2\varepsilon]\big/(b - a).$$

Hence, $[H_n(y) - H_n(a)]/(y - a) > m$ for any $y \geq \lambda$. This means that for $y \geq \lambda$, $H_n(y)$ lies above the line passing through $(a, H_n(a))$ with slope m. Since $C_n^\lambda(a) \leq H_n(a)$, $H_n(y)$ also lies above the line passing through $(a, C_n^\lambda(a))$ with slope m. Thus for all $y \geq \lambda$, $H_n(y)$ does not affect $C_n^\lambda(a)$, hence does not affect $C_n^\lambda(x)$ for all $x \leq a$. We have shown that under E_n, $C_n^\lambda(x) = C_n^*(x)$ for all $x \leq a$. Thus, $\mathrm{pr}\{C(x) = C_n^\lambda(x)$ for C_n^* all $x \leq a\} \to 1$.

(ii) This is the case where F is continuous. If X distributes as F, then $F(X)$ distributes as $U(0,1)$, the uniform distribution on $[0,1]$. Also its survival function $S(X) = 1 - F(X)$ distributes as $U(0,1)$. From Karlin (1972), page 250, for $\alpha \geq 1$, $\Pr\{S_n(x) \leq \alpha S(x)$ for all $x\} = 1 - 1/\alpha$ for all n. This implies $\Pr\{H_n(x) \geq -\log\alpha + H(x)$ for all $x\} = 1 - 1/\alpha$ for all n. For any $\varepsilon > 0$, let $b = \log\alpha = -\log\varepsilon$. We have

$$\Pr\{H_n(x) \geq H(x) - b\} = 1 - \varepsilon. \tag{A4.1}$$

As in (i), $n^{1/2}\|H_n - H\|_\lambda$ is bounded in probability. Hence, given any $\delta > 0$, there exists N_1 large enough such that

$$\Pr\{\|H_n - H\| < \delta\} > 1 - \varepsilon \quad \text{for } n \geq N_1.$$

Let m be the right-hand derivative of $C_n^\lambda(x)$ at $x = a$. We want to show that with high probability, for x close to α_1 and n large enough, $H_n(x)$ is above the line passing through $(a, C_n^\lambda(a))$ with slope m. Hence, what happens in the right-hand tail does not affect $C_n^\lambda(x)$ for $x \leq a$. To achieve this goal, let E_n be the event that $\|H_n - H\|_\lambda < \delta$. Under E_n, H_n lies within the band $\{H(x) - \delta, H(x) + \delta\}$. Hence, $m < [H(\lambda) - H(a) + 2\delta]/(\lambda - a)$ and $H_n(a) < H(a) + \delta$. From (A4.1), it is sufficient to show that $H(x) - b \geq [H(a) + \delta] + [H(\lambda) - H(a) + 2\delta](\lambda - a)^{-1}(x - a)$ for x close to α_1, or equivalently, $[H(x) - H(a)]/(x - a) \geq (b + \delta)/(x - a) + [H(\lambda) - H(a) + 2\delta]/(\lambda - a)$ for x close to α_1.

If $\alpha_1 < \infty$, the ratio $[H(x) - H(a)]/(x - a)$ increases to ∞ as x tends to $\alpha_1(F)$. Hence there exists x_0 such that for $x \geq x_0$, $[H(x) - H(a)]/(x - a) \geq (b + \delta)/(x - a) + [H(\lambda) - H(a) + 2\delta]/(\lambda - a)$.

If $\alpha_1 = \infty$, this means $(b + \delta)/(x - a) \to 0$ as $x \to \alpha_1$. Since $[H(x) - H(a)]/(x - a) > [H(\lambda) - H(a)]/(\lambda - a)$ for $x > \lambda$, with δ small enough, there exists x_0, such that for $x \geq x_0$, $[H(x) - H(a)]/(x - a) \geq (b + \delta)/(x - a) + [H(\lambda) - H(a) + 2\delta]/(\lambda - a)$.

So far, we have shown that given $\varepsilon > 0$, there exist $x_0(\varepsilon)$ and $N_1 > 0$ such that for $n \geq N_1$

$$\text{(A4.2)} \quad \Pr\left\{\text{for all } y \geq x_0,\ H_n(y) \text{ does not affect } C_n^\lambda(x) \text{ for all } x \leq a\right\} \geq 1 - 2\varepsilon.$$

For this x_0, $n^{1/2}\|H_n - H\|_{x_0}$ is bounded in probability. Using the same argument as in (i), one can show that there exists $N_2 > 0$ such that for $n \geq N_2$,

$$\text{(A4.3)} \quad \Pr\left\{C_n^\lambda(x) = C_n^{x_0}(x) \text{ for all } x \leq a\right\} \geq 1 - \varepsilon.$$

Combining (A4.2) and (A4.3) we have

$$\Pr\left\{C_n^\lambda(x) = C_n^*(x) \text{ for all } x \leq a\right\} \geq 1 - 3\varepsilon$$

for n large enough. This proves the result for (ii). □

Acknowledgments. This paper is dedicated to the memory of the late Professor Jack Kiefer who first introduced and guided me to research in this area. I am grateful to Professor Lucien Le Cam for many inspiring discussions during the course of this work. The referee and the Associate Editor have made many helpful comments and suggestions.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

GRENANDER, U. (1956). On the theory of mortality measurement. Part II. *Scand. Actuar. J.* **39** 125–153.

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KARLIN, S. (1972). *A First Course in Stochastic Processes.* Academic, New York.

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. verw. Gebiete* **34** 73–85.

LEURGANS, S. (1982). Asymptotic distributions of slope of greatest-convex-minorant estimators. *Ann. Statist.* **10** 287–296.

MARSHALL, A. W. (1970). Discussion of Barlow and van Zwet's papers. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 175–176. Cambridge Univ. Press.

MARSHALL, A. W. and PROSCHAN, F. (1965). Maximum likelihood estimation for distributions with monotone failure rates. *Ann. Math. Statist.* **36** 69–77.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. verw. Gebiete* **48** 233–252.

PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **41** 507–519.

WANG, J.-L. (1982). Asymptotically minimax estimators for distributions with increasing failure rate. Ph.D. dissertation. Univ. California, Berkeley.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616