*The Annals of Statistics*
1982, Vol. 10, No. 1, 287-296

# ASYMPTOTIC DISTRIBUTIONS OF SLOPE-OF-GREATEST-CONVEX-MINORANT ESTIMATORS

By Sue Leurgans[1]

*University of Wisconsin-Madison, Mathematics Research Center*

Isotonic estimation involves the estimation of a function which is known to be increasing with respect to a specified partial order. For the case of a linear order, a general theorem is given which simplifies and extends the techniques of Prakasa Rao and Brunk. Sufficient conditions for a specified limit distribution to obtain are expressed in terms of a local condition and a global condition. It is shown that the rate of convergence depends on the order of the first non-zero derivative and that this result can obtain even if the function is not monotone over its entire domain. The theorem is applied to give the asymptotic distributions of several estimators.

**1. Introduction.** Suppose for each of $n$ independent variables $X_i$ there is a known $t_i$ such that the distribution of $X_i$ is believed to be determined by and to vary with $t_i$. Let $F_{t_i}$ denote the cumulative distribution function (CDF) of $X_i$, conditional on $t_i$. Let $\theta(\cdot)$ be a specified functional on a subspace of cumulative distribution functions. $\theta$ induces $\mu$, a real-valued function on the space of $t$'s, by $\mu(t) = \theta(F_t)$. $\mu$ is an isotonic function if there is a partial order on the space of $t$'s such that whenever $t$ is "greater than or equal to" $s$, $\mu(t) \geq \mu(s)$. This paper concentrates on the case in which the $t_i$'s are real numbers with the usual ordering and $\mu$ an isotonic function is equivalent to $\mu$ a non-decreasing function. An isotonic (or monotone) estimator of $\mu$ will be an estimator which always has the known monotonicity, but is not restricted to a particular functional form. Use of an isotonic estimator is appropriate if the order relation is certain; that is, if the failure of the observations to exhibit the specified order is an artifact of the randomness of the observations dominating the unknown underlying deterministic increasing function.

The least-squares solution to this problem has been known for some time. Ayer et al. (1955) and van Eeden (1956) describe an estimator $\hat{\mu}_n(t)$, the isotonized mean, which is the monotone function with smallest error sum of squares $\sum_{i=1}^n \{X_i - u(t_i)\}^2$, with $u$ nondecreasing; $\hat{\mu}_n$ adaptively pools observations until the group means are increasing. Barlow et al. (1972, Chapter 1) discuss several algorithms for computation of this estimator and include the fact that, if $s$ is in $(t_{k-1}, t_k]$, $\mu_n(s)$ is the left hand slope at $k$ of the greatest convex minorant of the cumulative sum process of the $X$'s $\{(j, \sum_{i=1}^j X_i),\quad 0 \leq j \leq n\}$; for $j = 0$, the summation is taken to be zero. The asymptotic distribution of this estimator was stated by Brunk (1970). The present paper shows that Brunk's result can be sharpened and extended through the use of a theorem on the distribution of the slopes of greatest convex minorants of processes. This theorem can also be used to extend the results of Prakasa Rao (1969) on estimation of monotone densities, as well as to obtain asymptotic distributions of other estimators. The general theorem is stated in Section 2, applications of the theorem are indicated in Section 3, and the general theorem is proved in Section 4. The final section contains a discussion of the relationship of the results described here to other research.

**2. The general theorem.** The asymptotic distribution of these estimators is of

Received March 1979; revised July 1981.

[1] Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant Nos. MCS77-16974 and MCS78-09525.

*AMS* 1980 *subject classifications.* Primary 60F05, 62E20; secondary 62G05, 62G20.

*Key words and phrases.* Isotonic estimation, asymptotic distribution theory.

interest because the finite sample distributions are especially complicated in all but the very simplest cases. However, to obtain limiting results, it is necessary to specify how the limits are obtained. If, for example, the set of $t$'s to which $X$'s correspond is fixed and hence finite, while the number of $X$'s observed at each $t$ becomes infinite, and if the mean of those $X$'s corresponding to a particular $t$ converges to $\mu(t)$ and $\mu$ takes on a distinct value at each of the $t$'s for which observations are recorded, then these means are asymptotically consistent, asymptotically independent, and asymptotically normal if rescaled in the usual manner; see Parsons (1975) for further discussion of this case.

This paper concentrates on the case in which the number of distinct $t$'s at which observations are made becomes infinite. Exact conditions on the $t$'s appear in the examples below. Meanwhile we assume that for each $n$ we observe $\{(T_{ni}, X_{ni}), i = 1(1)n\}$ where $X_{ni} | T_{ni} \sim F_{T_{ni}}$. We can assume that the observations are indexed so that $T_{ni}$ increases (strictly) with $i$ for every $n$. If the $T$'s are random, $X_{ni}$ is thus the concomitant of the $i$th order statistic of the $T$'s. Inspired by the isotonized mean, we wish to work with estimators of the form $\mu_n(s) = \text{slogcom}(s)\ \{(t, Z_n(t)), t \in \mathscr{T}\}$ where slògcom$(s)\{A\}$ is the left-hand slope at $s$ of the greatest convex minorant of the set of points $A$, $Z_n(t)$ is a random continuous process and $\mathscr{T}$ is an interval containing $s$. Theorem 1 states that if the process $Z_n$ satisfies two conditions, then the asymptotic behavior of $\mu_n(s)$ is known. While the conditions look complicated, they can be described intuitively and verified in practice. Before examining the conditions, note that the proof uses the approximate estimators $\mu_{nc}(s + D_n) = \text{slogcom}(s)\ \{(t, Z_n(t)),\ |t - s| \le 2cn^{-p}\}$, local versions of $\mu_n(s)$.

The first condition on $Z_n$ is that the increments of $Z_n$ stay above certain lines over certain regions with sufficiently high probability. These lines depend on $n$ and $c$, although this dependence is suppressed in some of the notation. Therefore weak convergence of the $Z_n$ process will *not* imply Condition 1.

*Condition* 1 *(Hitting Times).*

$$\lim_{c\to\infty} \limsup_{n\to\infty} P\{Z_n(t) - Z_n(s + D_n) < L_i(t), \quad \text{some } t \in I_i\} = 0, \qquad i = 1(1)4,$$

where $L_i(t)$ is a line, $D_n = o_p(n^{-p})$ is a scalar and $I_i$ is an interval,

$$I_1 = (-\infty, s + D_n] \cap \mathscr{T}, \qquad I_2 = [s + D_n + 2cn^{-p}, \infty) \cap \mathscr{T}$$

$$I_3 = [s + D_n, \infty) \cap \mathscr{T}, \qquad I_4 = (-\infty, s + D_n - 2cn^{-p}] \cap \mathscr{T}.$$

The first two lines are

$$L_1(t) = -t(n, c) - (s + D_n - t)\bar{\mu}(n, c), \qquad L_2(t) = -t(n, c) - (s - D_n - t)\bar{\mu}(n, c),$$

where $\bar{\mu}(n, c) = \mu(s) + (2cn^{-p})^{(1-p)/2p}\rho(s), \qquad s_n = s + 2cn^{-p} + D_n,$

$$t(n, c) = \zeta(1 - 2^{(p-1)/2p})2^{(1-p)/2p}\rho(s)n^{-(p+1)/2}c^{(1+p)/2p}, \quad 0 < \zeta < 1$$

for some constants $s$, $\mu(s)$, $p$, $\rho(s)$ and $\zeta$. $L_3$ is obtained by using the formula for $L_1$ with $\bar{\mu}(n, c)$ replaced by $\mu(s) - (2cn^{-p})^{(1-p)/2p}\rho(s)$. $L_4$ is obtained from $L_2$ by making the same substitution and replacing $s + D_n$ by $s - 2cn^{-p}$.

The second condition is that a suitably renormalized local version of $Z_n$ converges to a Wiener process about a convex function. This condition will be used to obtain the limiting behavior of $\mu_{nc}(s)$ for $c$ fixed and $n$ large. Thus the first condition will be used to show that the local behavior of $Z_n$ determines the asymptotic behavior of $\mu_n$.

*Condition* 2 *(Local Weak Convergence).*

$$\left\{t, \frac{n(Z_n(s + 2cn^{-p}t + D_n) - Z_n(s + D_n)) - 2cn^{1-p}t\mu(s)}{(2cn^{1-p})^{1/2}\sigma(s)}\right\}_{|t|\le 1}$$

$$\to_w \quad \left\{t, W(t) + \frac{\rho(s)}{\sigma(s)}(2c)^{1/2p}|t|^{(1+p)/(2p)}\right\}_{|t|\le 1} \qquad \text{as} \quad n \to \infty.$$

where $p$, $s$, $\mu(s)$, $\rho(s)$ and $\sigma(s)$ are constants such that $s$ and $p$ are in (0.1) and $\rho(s)$ and $\sigma(s)$ are positive; $W(t)$ is a two-sided standard Wiener process on $[-1, +1]$, and the convergence is weak convergence on $C[-1, +1]$.

THEOREM 2.1 *If for some constants* $\mu(s)$, $\sigma(s)$, $\rho(s)$ *and* $p$, *the processes* $Z_n$ *satisfy Conditions* 1 *and* 2 *above and* $\mu_n(s) = \text{slogcom}(s + D_n)\ \{(t, Z_n(t)), t \in \mathscr{T}\}$, *then*

$$\frac{n^{(1-p)/2}\{\mu_n(s) - \mu(s)\}}{\{\sigma(s)\}^{1-p}\{\rho(s)\}^{p}} \to_d X^{(p)}, \tag{1}$$

*where*

$$X^{(p)} =_d \text{slogcom}(0)\ \{(t,\ W(t) + |\,t\,|^{(1+p)/2p}),\ |\,t\,| < \infty\}$$

*and* $W(t)$ *is a standard Wiener process on* $\mathbb{R}$ *with* $W(0) = 0$.

We shall see below that the case $p = 1/3$ is most common in applications. In this case, the distribution of $X^{(p)}$ can be described without use of convex minorants. As stated by Prakasa Rao (1969), the distribution of $X^{(1/3)}$ is that of $V/2$, where $V$ is the random value at which $W(t) - t^2$ attains its maximum. Chernoff (1964, Theorem 1, page 37) proves that $V$ has a density of the form $h(x)h(-x)$, where $h$ is a function involving partial derivatives of a particular solution of the heat equation.

**3. Application of the theorem.** This section contains two examples of the application of Theorem 2.1. The first example is an extension of Brunk (1970) to higher order derivatives and random times. The second example discusses isotonized linear combinations of order statistics. Theorem 2.1 can also be used to obtain the results of Prakasa Rao (1969), Prakasa Rao (1970), Barlow and van Zwet (1970) and Barlow et al. (1972).

EXAMPLE 1: *The isotonized mean.* Let the functional $\theta(\cdot)$ operate on the space of cumulative distribution functions with finite expectations by assigning to each CDF its expectation. The induced function $\mu$ satisfies $\mu(t) = E(X_t\,|\,t)$. The isotonized mean $\hat{\mu}_n$ alluded to in Section 1 is a natural estimator of $\mu$. If $F_n$ is the empirical distribution function of $\{T_{ni}, 1 \le i \le n\}$ and if

$$\tilde{\mu}_n(u) = \text{slogcom}(u)\{(F_n(T_{nj}), \textstyle\sum_{i=1}^{j} X_{ni}/n), 0 \le j \le n\},$$

then $\hat{\mu}_n$ satisfies $\hat{\mu}_n(r) = \tilde{\mu}_n(F_n(r-) + n^{-1})$. Since $D_n = F_n(r) + n^{-1} - F(r) = O_p(n^{-1/2})$, if $s = F(r)$, then

$$\hat{\mu}_n(r) = \tilde{\mu}_n(s + D_n) = \text{slogcom}(s + D_n)\{(t, Z_n(t)), 0 \le t \le 1\},$$

where $Z_n$ is the random function defined by linear interpolation between points of $\{(j/n, \sum_{i=1}^{j} X_{ni}/n),\quad 0 \le j \le k(n)\}$. Theorem 2.1 will be applied to give the following result:

COROLLARY 3.1. *Assume the following seven conditions are met*:

1. $\{T_{ni}, 1 \le i \le n\}$ *are the order statistics of a sample of size n from a distribution F which possesses a positive derivative* $f(r)$ *in a neighborhood of r.*
2. $E(X_{ni}\,|\,T_{ni}) = \mu(T_{ni})$ *and* $\{X_{ni} - \mu(T_{ni}),\quad 1 \le i \le n\}$ *is a set of mutually independent random variables for each n.*
3. *For every n,* $\{(X_{ni} - \mu(T_{ni})),\quad 1 \le i \le n\}$ *and* $\{T_{ni},\quad 1 \le i \le n\}$ *are independent sets of random variables.*
4. $\text{Var}\ (X_{ni}\,|\,T_{ni}) = \sigma^2 < \infty$ (*and* $\sigma^2 > 0$) *and* $\{(X_{ni} - \mu(T_{ni}))^2,\quad 1 \le i \le n, n \ge 1\}$ *are uniformly integrable.*
5. *For some* $\delta > 0$, $sup_{t \le r-\delta}\, \mu(t) < \mu(r) < \inf_{t \ge (r+\delta)}\, \mu(t)$ *and* $\mu$ *is increasing on* $[r - \delta, r + \delta]$.
6. $\mu$ *has an Nth order derivative at r.*
7. *N is the smallest positive (finite) integer with* $\mu^{(N)}(r) > 0$. *Then*

$$\{f(r)n\}^{N/(2N+1)}\left\{\frac{(N+1)!}{\mu^{(N)}(r)\sigma^{2N}}\right\}^{1/(2N+1)}\{\hat{\mu}_n(r)-\mu(r)\}\to X^{(1/(2N+1))},$$

*where $\hat{\mu}_n(r)$ is the isotonized mean based on $\{(T_{ni}, X_{ni}),\quad 1 \le i \le n\}$ evaluated at r.*

PROOF. The conditions of Theorem 2.1 are checked with $p = (2N+1)^{-1}$, $\sigma(s) = \sigma$, $\rho(s) = \mu^{(N)}(r)/\{f^N(r)(N+1)!\}$, and $Z_n$ the normalized cumulative sum process. Note that $r$ replaces $s$ in the conditions for and statement of Theorem 2.1. Let $\varepsilon(n, c)$ denote $\bar{\mu}(n, c) - (r)$. For notational convenience, we obtain the limiting distribution of $\tilde{\mu}_n(F_n(r-))$. In what follows, $D_n$ denotes $F_n(r-) - F(r)$ and $s = F(r)$.

We sketch the verification of the first Hitting Time Condition; the others are routine variations. In this example, the first Hitting Time Condition reduces to

$$\lim_{c\to\infty}\limsup_{n\to\infty} P\{-\textstyle\sum_{i=k+1}^{n(s+D_n)} X_{ni}/n < -t(n, c) - (s + D_n - k/n)\bar{\mu}(n, c),\quad \text{some } k \le n(s + D_n)\},$$

which involves

$$q(n, c) = P\{\textstyle\sum_{i=k+1}^{n(s+D_n)} (X_{ni} - \mu(T_{ni})) > nt(n, c) + (ns + nD_n - k)\varepsilon(n, c) + -\textstyle\sum_{i=k+1}^{n(s+D_n)} (\mu(T_{ni}) - \mu(r)),\quad \text{some } k \le n(s + D_n)\},$$

the probability that a cumulative sum process crosses a line, where the sequence of cumulative sums depends on $n$ and the line depends on $n$ and on $c$. Since the fourth assumption of the corollary implies that $\mu(T_{ni}) = E(X_{ni} \mid T_{ni}) \le \mu(r)$ for $i \le n(s + D_n)$,

$$q(n, c) \le P\{\textstyle\sum_{i=k}^{n(s+D_n)} (X_{ni} - \mu(T_{ni})) > nt(n, c) + (ns + nD_n - k)\varepsilon(n, c),\quad \text{some } k \le n(s + D_n)\}.$$

This last expression can be written as $P\{S_{n\ell} > nt(n, c) + \ell\varepsilon(n, c),\quad \text{some } 0 \le \ell \le n(s + D_n)\}$, where $S_{n\ell}$ is the $\ell$th cumulative sum of $n$ independent random variables with variance $\sigma^2$. Using the Dubins-Savage inequality (Dubins & Savage, 1965; Dubins & Freedman, 1965) or the Hájek-Rényi Inequality applied to the submartingales $\{S^2_{n\ell},\quad 0 \le \ell \le n\}$ with constants $c_k = \varepsilon(n, c)/\{\sigma^2 + nt(n, c)k\}$ (Chow, Robbins, and Siegmund, 1970, page 25), it can be shown that

$$P\{S_{n\ell} > nt(n, c) + \ell\varepsilon(n, c),\quad \text{some } 0 \le \ell \le n(s + D_n)\} \le (1 + \varepsilon(n, c)nt(n, c)/\sigma^2)^{-1} = (1 + c^{2N+1}2^N(2^N - 1)\rho^2(s)\zeta/\sigma^2)^{-1}.$$

This implies $\lim_{c\to\infty}\limsup_{n\to\infty} q(n, c) = 0$, as desired.

The weak convergence condition involves the convergence of the process defined by linear interpolation between the points of

$$\left(\frac{\ell}{k(n)}, \textstyle\sum_{i=1}^{\ell} \dfrac{\{X_{n,Q(n)+i} - \mu(T_{n,Q(n)+i})\}}{\sigma\{k(n)\}^{1/2}} + \textstyle\sum_{i=1}^{\ell} \dfrac{\{\mu(T_{n,Q(n)+i}) - \mu(r)\}}{\sigma\{k(n)\}^{1/2}}\right)_{0\le|\ell|\le k(n)}, \tag{3}$$

where $k(n) = 2cn^{1-p}$ and $Q(n) = nF_n(r)$. By assumption 3, the process defined by the first sum is independent of that defined by the second sum. Assumption 4 ensures that the Lindeberg condition holds for $\{X_{n,Q(n)+i} - \mu((T_{n,Q(n)+i})/n),\quad 1 \le i \le k(n)\}$ and thus that the first process in (3) converges weakly to Brownian motion (Billingsley, 1968, page 77, pr. 2).

It remains to show that the second term in (3) converges weakly to the nonrandom process $\mu^{(N)}(r)(2c)^{(2N+1)/2}|t|^{N+1}/\{f(r)^{N+1}(N+1)!\sigma\}$. The following lemma will be useful:

LEMMA 3.1. *If $V_n(t) = F_n^{-1}(F_n(r) + tk(n)/n)$ and if assumption 1 of Corollary 3.1 holds, then*

$$V_n(t) - V_n(0) \sim 2cn^{-1/(2N+1)}t/f(r).$$

PROOF. Since assumption 1 implies that $n^{-1/2}(F_n^{-1} - F^{-1})$ converges weakly to a Gaussian process in a neighborhood of $r$, $F_n^{-1}(t) = F^{-1}(t) + n^{-1/2}B_n(t)$, where $B_n(t)$ converges weakly to a Gaussian process, and

$$\begin{aligned} V_n(t) - V_n(0) &= F^{-1}(F_n(r) + tk(n)/n) + n^{-1/2}B_n(F_n(r) + tk(n)/n) \\ &\quad - F^{-1}(F_n(r)) - n^{-1/2}B_n(F_n(r)) \\ &= F^{-1}(F_n(r) + tk(n)/n) - F^{-1}(F_n(r)) + n^{-1/2}O(\sup B_n(x)). \end{aligned}$$

Since with probability 1,

$$F_n(r) = F(r) + O((\ln n/n)^{1/2}) = F(r) + o(k(n)/n)$$

(cf. Csaki, 1968), expansion of $F^{-1}$ in a Taylor series at $F(r)$ shows that

$$V_n(t) - V_n(0) = tk(n)/\{nf(F^{-1}(F(r)))\} + O_p(n^{-1/2}) + o(k(n)/n).$$

The lemma follows.

The second term in (3) can be written as the sum of a Lebesgue-Stieltjes integral with respect to $F$ and an integral with respect to $(F_n - F)$:

$$\frac{n\sigma^{-1}}{\{k(n)\}^{1/2}}\int_{V_n(0)}^{V_n(t)} \{\mu(x) - \mu(r)\}\, dF(x) + \frac{n\sigma^{-1}}{\{k(n)\}^{1/2}}\int_{V_n(0)}^{V_n(t)} \{\mu(x) - \mu(r)\}\, d(F_n - F)(x).$$

Assumptions 5 and 6 imply that the first integral is, to first order,

$$\begin{aligned} &n\{k(n)\}^{-1/2}\{V_n(t) - r\}^{N+1}\sigma^{-1}\mu^{(N)}(r)f(r)/(N+1)! \\ &\qquad = \mu^{(N)}(r)[n^{1/(2N+1)}\{V_n(t) - V_n(0)\}]^{N+1}f(r)/\{(N+1)!\sigma\}, \end{aligned}$$

which converges to $\mu^{(N)}(r)\{2ct/f(r)\}^{N+1}f(r)/\{(N+1)!\sigma\}$ by Lemma 3.1. Since $N$ is odd, this is the process desired. The second integral will be shown to be negligible. The differentiability conditions imply that the second integral is bounded above by

$$\begin{aligned} &n\,|\mu(V_n(t)) - \mu(V_n(0))|\,\{|F_n(V_n(t)) - F(V_n(t))| - |F_n(V_n(0)) - F(V_n(0))|\}\{k(n)\}^{-1/2} \\ &\qquad \le 2k(n)^{-1/2}[\mu^{(N)}(r)\{V_n(t) - V_n(0)\}^N + o((V_n(t) - V_n(0))^N)]\sup_x|F_n(x) - F(x)|. \end{aligned}$$

This bound is therefore of order $[n/\{k(n)\ln n\}]^{1/2}(V_n(1) - r)^N$, which Lemma 3.1 implies is of order

$$\{k(n)/n\}^N[n/\{k(n)\ln n\}]^{1/2} = n^{1/2-(N+1)/(2N+1)}(\ln n)^{-1/2} = o(1).$$

This completes the verification of the weak convergence condition.

EXAMPLE 2. *Smoothly weighted linear combinations of order statistics; equally spaced observations.* Let $J$ be a smooth (see below) weight function defined on [0, 1] with $\int J(u)\,du = 1$. $\theta(F)$ is defined to be the solution of $\int J(u)Q(u - \theta(F))\,du = 0$, where $Q = F^{-1}$, for all continuous $F$ such that the integral is well-defined. For all $F$ members of a specific translation family, $\theta(F)$ is a percentile of $F$. Which percentile $\theta(F)$ gives depends on the weight function and on the shape of $F$. For example, if $J$ is symmetric about ½ and the distribution determined by $F$ is symmetric, $\theta(F)$ is the median of $F$. The weight function $J$ can be used to construct the following process from which a slogcom estimator will be obtained.

Let $\{X_i, 1 \le i \le k\}_{(j)}$ denote the $j$th order statistic of the set $X_1, \cdots, X_k$ and take $T_{ni} = i/n$. Then for $s$ in (0, 1) fixed, define

$$Z_n(s + \ell/n) = \textstyle\sum_{j=1}^{\ell} J(j/(\ell+1))\{X_{n\langle ns\rangle+i}, 1 \le i \le \ell\}_{(j)}/n,\ \ell \text{ positive},$$

$$Z_n(s + \ell/n) = -\textstyle\sum_{j=1}^{-\ell} J(j/(-\ell+1))\{X_{n\langle ns\rangle-i+1}, 1 \le i \le -\ell\}_{(j)}/n,\ \ell \text{ negative},$$

and

$$Z_n(s) = 0,$$

where $\langle ns \rangle$ is the least integer greater than or equal to $ns$. The process $Z_n$ can be thought of as the cumulative sum process of Corollary 3.1 pinned at $s$ (that is, $Z_n(t) - Z_n(s)$) with each sum of random variables replaced by the $J$-weighted sum of the order statistics of the same set of random variables. Extend $Z_n$ to a continuous process on [0, 1] by linear interpolation and define $\mu_n(s) = \text{slogcom}(s)\{(t, Z_n(t)), 0 \le t \le 1\}$. For any finite set of integers $A$, define $N(A)$ to be the number of elements of $A$. Let

$$J_n(A) = \sum_{j \in A} J(j/(N(A) + 1))\{X_{n,i} i \in A\}_{(j)}/n.$$

Then $\mu_n(s)$ can be written as $[N(L^*) \max J_n(L) + N(U^*) \min J_n(U)]/N(L^* \cup U^*)$, where the maximum is taken over the sets $L$ of the form $\{i \le j \le \langle ns \rangle\}$ for some $i$, $L^*$ is the largest such set for which the maximum is attained, the minimum is over sets $U$ of the form $\{\langle ns \rangle + 1 \le j \le k\}$, and $U^*$ is the largest such set for which the minimum is obtained. Note that $L^*$ and $U^*$ are disjoint.

COROLLARY 3.2. *If the following six assumptions are met, then*

$$n^{N/(2N+1)} \left\{ \frac{(N+1)!}{\sigma^{2N} \mu^{(N)}(s)} \right\}^{1/(2N+1)} \{\mu_n(s) - \mu(s)\} \to_d X^{((2N+1)^{-1})} \quad \text{as} \quad n \to \infty.$$

1. $X_{ni} - \mu(i/n)$ *are independent, identically distributed random variables with cumulative distribution function* $F$.
2. $\int J(u)Q(u)\, du = 0$, $\int J(u)\, du = 1$, *and* $\mu$ *is non-decreasing on* (0, 1].
3. $J$ *is continuously differentiable nonnegative function whose derivative* $J'$ *satisfies a Hölder condition for some* $\gamma$, $1/2 < \gamma \le 1$, *that is*, $|J'(u) - J'(v)| \le K|u - v|^\gamma$ *for finite* $K$. *The support of* $J$ *is a compact subset of* (0, 1).
4. $\sigma^2 = \int\int J(F(x))J(F(y))F(\min(x, y))\{1 - F(\max(x, y))\}\, dx\, dy > 0$.
5. $\mu$ *has an Nth order derivative at* $s$, $0 < s < 1$, *where* $N$ *is the smallest finite integer with* $\mu^{(N)}(s) > 0$.
6. $F$ *is absolutely continuous, with strictly positive density* $f$ *such that* $f$ *converges to zero at infinity and* $f'$ *is bounded.*

The first two conditions describe the model and assert that the weight function $J$ is appropriate. The third condition, which includes a requirement that $J$ trim, is used to verify the Local Weak Convergence. The nonnegativity of $J$ will be used in the proof of the Hitting Time Condition. The fourth condition is more a definition than a condition, since the third condition ensures the integral is finite. The fifth condition describes the local behavior of $\mu$ at $s$. The sixth condition is a regularity condition used to obtain the Cornish-Fisher expansion needed to compute the drift component of the local weak convergence.

Corollary 3.2 shows that if the $X_{ni}$ are all members of the translation family generated by the CDF $F$, the relative efficiency of two different isotonized linear combinations of order statistics with weight functions $J_1$ and $J_2$ is determined by the ratio $\sigma(J_1, F)/\sigma(J_2, F)$, where $\sigma(J, F) = \int\int J(u)J(v)\{\min(u, v) - uv\}\, dQ(u)\, dQ(v)$. Corollary 3.1 shows that if $F$ has finite variance, the same formula gives the efficiency of an isotonized linear combination of order statistics relative to the isotonized mean, although the weight function of the mean does not satisfy the conditions of Corollary 3.2. This same ratio is the asymptotic relative efficiency of two linear combinations of order statistics for estimating the location parameter of independent, identically distributed random variables whose distribution is a member of the location family generated by $F$. Therefore, all the comparisons known for a simple location problem carry over to isotonic estimation. In particular, if $F$ does not have a variance, Corollary 3.2 applies, and the isotonized version of any linear combination of order statistics which trims will converge in the familiar manner. However, Corollary 3.1 does not apply. This extreme case shows that the isotonized mean is sensitive to wild observations and isotonized trimmed linear combinations of order statistics are more robust to heavy tails.

The proof of Corollary 3.2 can be described as showing that linear functions of order statistics behave almost like sums of independent random variables. In Leurgans (1978), it was shown that $Z_n$ has the same distribution as the sum of eight terms, six of which converge to zero in $C[0, 1]$. The remaining terms are $\sum_{j=1}^{tk(n)} Z_{nj} + C_n(t)$, where the $Z_{nj}$ are independent random variables and $C_n(t)$ is a nonrandom process. The assumptions of Corollary 3.2 can be used to show that

$$C_n(t) \sim k(n)t\mu(s) + \sum_{j=1}^{tk(n)} \{\mu(s + j/n) - \mu(s)\}/\{tk(n)\}.$$

The weak convergence condition then follows as in Example 2.1.

The representation of the order statistic process alluded to above can also be used to verify the Hitting Time Condition. The verification of the first condition will be sketched.

The first step is to reduce the first condition to

$$\lim_{c\to\infty}\limsup_{n\to\infty} P\{-\sum_{j=1}^{\ell} J(j/(\ell + 1))\{X_{n,\langle ns\rangle -i+1},\quad 1 \le i \le \ell\}_{(j)} < nL_1(s - \ell/n),\quad 0 < \ell < ns\}$$

$$= \lim_{c\to\infty}\limsup_{n\to\infty} P\{\sum_{j=1}^{\ell} J(j/\ell + 1)\{X_{n,\langle ns\rangle -i+1},\quad 1 \le i \le \ell\}_{(j)} > nt(n, c) + \ell(\mu(s) + \varepsilon(n, c)),\ 0 < \ell < ns\}.$$

If $X'_{n,i} := X_{n,i} - \mu(i/n) + \mu(s)$, then $X'_{n,i} \ge X_{n,i}$. Because a location family was assumed, the $X'_{n,i}$ will be iid. Since $J$ is assumed nonnegative, each probability above is less than or equal to

$$P[\sum_{j=1}^{\ell} J(j/(\ell + 1))\{X'_{n,i},\quad 1 \le i \le \ell\}_{(j)} > nt(n, c) + \ell\{\mu(s) + \varepsilon(n, c)\},\ \text{some } \ell,\ 0 < \ell < ns].$$

The representation of a linear function of order statistic process as a cumulative sum process plus seven other processes can now be used to determine the behavior of the limiting process. Since the $X'_{n,i}$ have identical distributions, $C_n(t) = nt\mu(s)$ and two of the six remainder processes are identically zero. If $Y_{in}$ is the $i$th remainder process, it suffices to show that

$$\lim_{c\to\infty}\limsup_{n\to\infty} P\{\sup_{0\le t\le s} | Y_{in}(t) | > nt(n, c)/5\} = 0$$

and

$$\lim_{c\to\infty}\limsup_{n\to\infty} P\{\sum_{i=1}^{\ell} Z_{ni} \ge nt(n, c)/5 + \varepsilon(n, c)\ell/5,\ \text{some } 0 < \ell \le ns\} = 0.$$

The statements for the remainder terms are verified in Leurgans (1978, Section V.3). The cumulative sum term follows as in the preceding example.

Corollary 3.2 can be extended in the manner of Corollary 3.1 to $T_{ni}$ which are either other tractable deterministic sequences or order statistics from a suitable distribution.

**4. Proof of Theorem 2.1.** Let $Y_n$ denote the right-hand side of (1) and $X_{nc}$ the same expression with $\mu_{nc}$ replacing $\mu_n$. The theorem will be established in the following steps:

1. For all $c$, $X_{nc} \to_d X_c$ as $n \to \infty$. ($X_c$ will be defined below.)
2. $X_c \to_d X^{(p)}$ as $c \to \infty$.
3. $\lim_{c\to\infty}\liminf_{n\to\infty} P\{\mu_{nc}(s) = \mu_n(s)\} = \lim_{c\to\infty}\liminf_{n\to\infty} P\{X_{nc} = Y_n\} = 1$.

By Theorem 4.2 of Billingsley (1968, page 25), $Y_n \to_d X^{(p)}$, which implies Theorem 2.1.

*Step* 1. Since adding a line to a function increases the slope of the function's convex minorant by the slope of the line

$$\mu_{nc}(s) - \mu(s)$$
$$= \text{slogcom}(s + D_n)\{(t, Z_n(t) - Z_n(s + D_n) - (t - s - D_n)\mu(s)),\ | t - (s + D_n) | \le 2cn^{-p}\}.$$

Translating the process so that the time scale of the function whose convex minorant is being obtained is $[-1, +1]$ and rescaling,

$$\frac{(2cn^{1-p})^{1/2}}{\sigma(s)}\{\mu_{nc}(s) - \mu(s)\}$$

$$= \text{slogcom}(0)\left\{\left(t, \frac{n(Z_n(s + D_n + 2cn^{-p}t) - Z_n(s + D_n)) - 2cn^{1-p}t\mu(s)}{(2cn^{1-p})^{1/2}\sigma(s)}\right), |t| \le 1\right\}, \tag{4}$$

the local weak convergence condition implies that the expression above converges as $n \to \infty$ to

$$\text{slogcom}(0)\{(\text{t}, \text{W(t)} + (2\text{c})^{1/(2p)}\rho(s)/\sigma(s)\,|t|^{(1+p)/2p}), |t| \le 1\},$$

which can be shown (using scale properties of the Wiener process) to equal $\{2cK(s)\}^{1/2}X_c$, where

$$X_c = \text{slogcom}(0)\{(t, W(t) + |t|^{(1+p)/2p}), |t| \le 2cK(s)\},$$

$$K(s) = \{\rho(s)/\sigma(s)\}^{2p}.$$

Dividing (4) by $\{2cK(s)\}^{1/2}$, we see that for fixed $c$, $X_{nc}$ converges in distribution to $X_c$.

*Step* 2. The only difference in the definitions of $X_c$ and $X^{(p)}$ is that in $X_c$ the set of points is restricted to $|t| \le 2cK(s)$. Therefore to show $X_c$ converges in distribution to $X^{(p)}$, it is necessary to show that large values of $t$ do not affect the convex minorant of $W(t) + |t|^{(1+p)/2p}$. Since $p < 1$ implies the exponent of $|t|$ is greater than one, the proof of this step follows from $W(t)/t \to_{\text{a.s.}} 0$ $(t \to \infty)$, as is pointed out by Wright (1981). For an explicit proof in the case $p = 1/3$, see Prakasa Rao (1969, Lemma 6.2, page 34).

Figure 1 displays a realization of the process $Z_n$ and the lines $\ell 1$, connecting $(s + D_n - cn^{-p}, Z_n(s + D_n - cn^{-p}))$ and $(s + D_n + cn^{-p}, Z_n(s + D_n + cn^{-p}))$; $\ell 2$, which has slope $\mu(s) - (2cn^{-p})^{(1-p)/2p}\rho(s)$ and intersects the graph of $Z_n$ at $s + D_n - cn^{-p}$; and $\ell 3$, which has slope $\bar{\mu}(n, c)$ and intersects $Z_n$ at $s + D_n + cn^{-p}$. If the region $R$ lying above all three lines is convex, as in the figure, no points of $Z_n$ in this region can affect slogcom$(s + D_n)\{(t, Z_n(t)), t \in \mathbb{R}\}$ and $\mu_{nc}(s)$ will equal $\mu_n(s)$ if $Z_n(t) > \ell 2(t)$ and $Z_n(t) > \ell 3(t)$ for $|t - s - D_n| > 2cn^{-p}$. The region $R$ will be convex if $Z_n(s + D_n - cn^{-p}) = \ell 1(s + D_n - cn^{-p}) > \ell 3(s +$

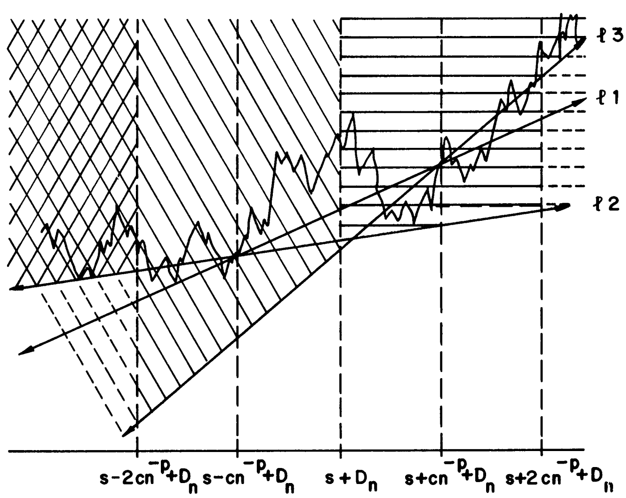


FIG. 1 Realization of the process $Z_n$ and lines $\ell_1$, $\ell_2$, $\ell_3$.

$D_n - cn^{-p}$) and $Z_n(s + D_n + cn^{-p}) > \ell 2(s + D_n + cn^{-p})$, which necessarily occurs if $Z_n(t) > \ell 3(t)$ for $|t - (s + D_n - cn^{-p})| < cn^{-p}$ and $Z_n(t) > \ell 2(t)$ for $|t - (s + D_n + cn^{-p})| < cn^{-p}$. Therefore $\mu_{nc}(s) = \mu(s)$ if $Z_n(t) > \ell 3(t)$ for $|t - (s + D_n + cn^{-p})| > cn^{-p}$ and $Z_n(t) > \ell 2(t)$ for $|t - (s + D_n - cn^{-p})| > cn^{-p}$. Therefore it suffices to show that the conditions imply that the probability $Z_n$ lies above two lines, for each of two separate intervals of $t$, is one in the appropriate limit. We shall show that the Hitting Time Conditions with $i = 1$ and the Local Weak Convergence Condition imply that $Z_n$ lies above $\ell 3$ for $t \le s + D_n$ with appropriately high probability. The other three Hitting Time Conditions are used in the same manner, and then the Bonferroni Inequality can be used to complete the proof.

Thus it remains to show that $\lim_{c\to\infty}\liminf_{n\to\infty} p(n, c) = 1$, where after rearranging we have $p(n, c) = P\{Z_n(s + D_n) - \ell 3(s + D_n) \ge \ell 3(t) - Z_n(t) + Z_n(s + D_n) - \ell 3(s + D_n), t \le s + D_n\}$. The probability that $Z_n(s + D_n) - \ell 3(s + D_n)$ exceeds $\ell 3(t) - Z_n(t) + Z_n(s + D_n) - \ell 3(s + D_n)$ is greater than the probability that $Z_n(s + D_n) - \ell 3(s + D_n)$ is greater than a fixed constant $t(n, c)$ and that this fixed constant is greater than $\ell 3(t) - Z_n(t) + Z_n(s + D_n) - \ell 3(s + D_n)$. Applying the Bonferroni Inequality to the intersection of the above two events, and recalling the definition of $L_1(t)$ in the Hitting Time Condition, it is easy to show that

$$p(n, c) \ge P\{Z_n(s + D_n) - \ell 3(s + D_n) \ge t(n, c)\} - P\{Z_n(t) - Z_n(s + D_n) < L_1(t), \text{ some } t \le s + D_n\}. \tag{5}$$

The Hitting Time Condition therefore implies the $\lim_{c\to\infty}\liminf_{n\to\infty}$ of the last term (minus sign included) is zero. Using the Local Weak Convergence Condition with $t = -\frac{1}{2}$ it can be shown (Leurgans, 1978, Chapter 3, Section 3) that the $\liminf_{n\to\infty}$ of the first term is $1 - \Phi((\zeta - 1)\lambda c^{(2p)^{-1}}\sqrt{2})$, where $0 < \zeta < 1$ (from the definition of $t(n, c)$); $\lambda = (2^{(1-p)/2p} - 1)\rho(s)/(\sigma(s)\sqrt{2})$ is positive because $p < 1$; and $\Phi$ is the cumulative distribution function of the standard normal distribution. Therefore the $\lim_{c\to\infty}\liminf_{n\to\infty}$ of the first term in (5) is 1, and the proof of the theorem is complete.

**5. Discussion.** Example 1 is a generalization of Brunk's Theorem 5.2. It should be remarked that Brunk's condition that "the observations satisfy Lindeberg's condition" can mislead the unwary: from the proof of Example 1 we see that the observations must satisfy local Lindeberg conditions, which are unrelated to a global Lindeberg Condition. Wright's paper also generalizes Brunk's Theorem, and is the only paper known to the author with results for $N > 1$. Wright does not require that $N$ be an integer and allows a different variance structure, but otherwise his results correspond to Example 1.

Robertson and Wright (1975) discuss monotone estimators of the form max min $J_n(L \cup U)$, with $L$ and $U$ defined as in Example 3.2. Unlike $\mu_n(s)$, $\bar{\mu}_n(s)$ is always a monotone function of $s$. Robertson and Wright give conditions under which $\bar{\mu}_n(s)$ is consistent for $\mu(s)$, but their methods do not give a rate of convergence. Corollary 3.2 gives such rates for the slogcom estimators $\mu_n$, and suggests that $\bar{\mu}_n$ has the same asymptotic behavior, even though $\mu_n$ and $\bar{\mu}_n$ are identical only in the case of Example 3.1. Isotonized percentiles of the $\bar{\mu}_n$ type are also discussed by Casady and Cryer (1976).

Recall that the isotonized mean at $s(\hat{\mu}_n(s))$ is the mean of the $x_{ni}$'s over an adaptively chosen neighborhood of $s$. Theorem 5.8 of Barlow, et al. (1972) and Theorem 3.2 of Davis (1972) point out that for each $s$, if slightly wider deterministic windows centered at $s$ are used, the resulting estimators converge more rapidly. However, this result appears to be the same sort of superefficiency result obtained in Example 3.1 for $N > 1$. In the case of Barlow et al. (1972), $s$ must be at the center of every window. In Example 3.1, $s$ must be exactly a point at which $\mu'(s) = 0$, but some other derivative is positive. If one is interested in estimation of an entire function, both kinds of $s$ are isolated. Also, the deterministic window estimators need not give monotone estimators of $\mu(s)$.

The fact that $\mu_n$ can be consistent in some cases even when $\mu$ is not monotone is reminiscent of Theorem 3.4 of Barlow, et al. (1972), which states that in the normal case with equal known variances, likelihood ratio tests that group means exhibit a specified partial order against the null hypothesis that the means are all equal is an unbiased test of some alternatives which do not have the specified partial order against the same null hypothesis. The application to estimation does not appear to have been noted previously.

**6. Acknowledgments.** This paper represents an extension of parts of the author's Ph.D. thesis prepared in the Stanford Statistics Department, under the guidance of Thomas W. Sager. The dissertation was commenced while the author was a NIH Trainee and completed on an NSF Graduate Fellowship. David Siegmund suggested the use of the Dubins-Freedman and Hájek-Rényi Inequalities.

The paper has benefitted by the detailed comments of the referee, whose suggestions led to a much shorter proof of part of Example 3.1.

## REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T., and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMMER, J. M., and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

BARLOW, R. E. and VAN ZWET, W. R. (1970). Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: Strong consistency. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 159–173, Cambridge.

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

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading.

BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.). 177–195, Cambridge.

CASADY, R. J., and CRYER, J. D. (1976). Monotone percentile regression. *Ann. Statist.* **4** 532–541.

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

CHOW, Y. S., ROBBINS, H., and SIEGMUND, D. (1971). *Great Expectations.* Houghton Mifflin, Boston.

CSAKI, E. (1968). An iterated logarithm law for semi-martingales and its application to the empirical distribution function. *Studia Sci. Math. Hungar.* **3** 287–292.

DAVIS, H. T. (1972). Nonparametric quantal response estimation procedures, Ph.D. thesis, Univ. of Chicago.

DUBINS, L. E. and FREEDMAN, D. A. (1965). A sharper form of the Borel-Cantelli Lemma and the Strong Law. *Ann. Math. Statist.* **36** 800–807.

DUBINS, L. E. and SAVAGE, L. J. (1965). A Tchebycheff-like inequality for stochastic processes. *Proc. Nat. Acad. Sci.* **53** 274–275.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications.* **II** 2nd ed., Wiley, New York.

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

LEURGANS, S. (1978). Asymptotic distribution theory in generalized isotonic regression. Ph.D. Dissertation, Dept. of Statist., Stanford Univ.

PARSON, L. V. (1975). Distribution theory of isotonic estimators. Ph.D. thesis, Univ. of Iowa.

PRAKASA, RAO, B. L. S. (1966). Asymptotic distributions in some non-regular statistical problems. Tech. Report 9, Dept. of Statist. and Probab., Michigan State Univ.

PRAKASA, RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā* Ser. A **31** 23–26.

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

ROBERTSON, T. and WRIGHT, F. T. (1975). Consistency in generalized isotonic regression. *Ann. Statist.* **3** 350–362.

VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Nederl. Akad. Wetensch Proc. Ser. A* **59**/*Indag. Math.* **18** 444–455.

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448.

DEPARTMENT OF STATISTICS
COLLEGE OF LETTERS AND SCIENCE
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WI 53706