ON THE APPROXIMATION OF A DISTRIBUTION FUNCTION BY AN EMPIRIC DISTRIBUTION[1]

By Jerome Blackman

Syracuse University

1. Summary. Let $x_1, \cdots, x_n$ be independent chance variables with the common distribution function $F(x)$ and the empiric distribution function $F^*(x)$. Let a_n be the value of a which minimizes (1) below. In this paper the asymptotic distribution of $\sqrt{n}\, a_n$ is obtained, subject to certain restrictions on $F(x)$.

2. Introduction. Let $x_1, x_2, \cdots, x_n$ be n independent random variables having the common distribution $F(x)$. Suppose $b_1, b_2, \cdots, b_n$ are the n random variables in ascending order of values and $F^*(x)$ is the empiric distribution function, continuous on the right, with jumps of magnitude $1/n$ at the points $b_1, b_2, \cdots, b_n$. Define the function $H(a)$ by

$$H(a) = \int_{-\infty}^{\infty} |F(x - a) - F^*(x)|^2 \, dF(x - a). \tag{1}$$

Thus $H(a)$ is non-negative for all a and since it is a Borel-measurable function of the random variables $\{x_i\}$, it is also a random variable. The value of a which minimizes $H(a)$ will also be a random variable. If the minimizing value of a is a_n, we shall be concerned with the limiting distribution of a_n as $n \to \infty$. Our main result is the following.

Theorem 2. *If the first three derivatives of $F(x)$ are continuous and bounded, then*

$$\lim_{n\to\infty} P(\sqrt{n}\, a_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu x/\nu} e^{-y^2/2} \, dy,$$

where

$$\nu^2 = \int_0^1 \left| \int_0^y F'(F^{-1}(t)) \, dt - \int_0^y \frac{dx}{x^2} \int_0^x tF'(F^{-1}(t)) \, dt \right|^2 dy,$$

$$\mu = \int_{-\infty}^{\infty} |F'(x)|^2 \, dF(x), \qquad F^{-1}(t) = \sup_x \{x \mid F(x) = t\}.$$

We shall henceforth assume the conditions of Theorem 2 are satisfied. In what follows repeated use will be made of an important result due to Kolmogoroff [2].

Theorem. *Suppose that $F(x)$ is continuous, and define the random variable D_n by $D_n = \text{l.u.b.}_x |F(x) - F^*(x)|$.*

Received June 17, 1954.

[1] Work on this paper was done under sponsorship of the Office of Naval Research. The author would like to thank Professor J. Wolfowitz for many useful suggestions and criticisms in the preparation of this paper.

Then for every fixed $z \geqq 0$, as $n \to \infty$, $P(D_n \leqq zn^{-1/2}) \to L(z)$, *where* $L(z)$ *is the cumulative distribution function which for* $z > 0$ *is given by either of the equivalent relations*

$$L(z) = 1 - 2\sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-2\nu^2 z^2) = \frac{\sqrt{2\pi}}{z} \sum_{\nu=1}^{\infty} \exp\left\{-\frac{(2\nu - 1)^2 \pi^2}{8z^2}\right\}.$$

For $z \leqq 0$ *we have, of course,* $L(z) = 0$.

The Kolmogoroff theorem implies that $F^*(x) \to F(x)$ uniformly in probability and that therefore $a_n \to 0$ in probability. Expanding the right side of (1),

$$\begin{aligned} H(a) &= \int_{-\infty}^{\infty} \{F(x-a) - F^*(x)\}^2 \, dF(x-a) \\ &= \int_{-\infty}^{\infty} F^2(x-a)\, dF(x-a) - 2\int_{-\infty}^{\infty} F(x-a)\, F^*(x)\, dF(x-a) \\ &\qquad + \int_{-\infty}^{\infty} F^{*2}(x)\, dF(x-a) \\ &= \tfrac{1}{3} + \int_{-\infty}^{\infty} F^2(x-a)\, dF^*(x) - \int_{-\infty}^{\infty} F(x-a)\, dF^{*2}(x). \end{aligned} \tag{2}$$

Therefore

$$H'(a) = \left\{-2\int_{-\infty}^{\infty} F(x-a)\, F'(x-a)\, dF^* + \int_{-\infty}^{\infty} F'(x-a)\, dF^{*2}(x)\right\}. \tag{3}$$

Each a_n must satisfy the equation $H'(a) = 0$. It will be seen below that all solutions of $H'(a) = 0$ which converge in probability to zero as $n \to \infty$ will have the same limiting distribution. Putting $H(a_n) = 0$ we obtain

$$2n^{-1} \sum F(b_i - a_n)F'(b_i - a_n) = n^{-2} \sum \{i^2 - (i-1)^2\}F'(b_i - a_n),$$

where summation is from 1 to n, or

$$\sum\{2nF(b_i - a_n) - 2i + 1\}F'(b_i - a_n) = 0. \tag{4}$$

Since $F(x)$ has a continuous third derivative,

$$\begin{cases} F(b_i - a_n) = F(b_i) - F'(b_i)a_n + \tfrac{1}{2}F''(b_i - \theta a_n)\, a_n^2, & 0 \leqq \theta \leqq 1; \\ F'(b_i - a_n) = F'(b_i) - F''(b_i)a_n + \tfrac{1}{2}F'''(b_i - \psi a_n)a_n^2, & 0 \leqq \psi \leqq 1. \end{cases} \tag{5}$$

Placing (5) in (4) and dividing by n^2 results in

$$\begin{aligned} 0 = {} & \{-2n^{-1}\textstyle\sum F(b_i)F'(b_i) - n^{-2}\sum(-2i + 1)F'(b_i)\} \\ & + \{2n^{-1}\textstyle\sum F(b_i)F''(b_i) + 2n^{-1}\sum F'^2(b_i) + n^{-2}\sum(-2_i + 1)F''(b_i)\}a_n \\ & + T_n(a_n)a_n^2 \end{aligned} \tag{6}$$

where $P\{|T(a_n)| \leqq C\} \geqq 1 - \epsilon$ for some C and for $n > N(\epsilon)$ by the assumption that the derivatives are bounded and that $a_n \to 0$ in probability. Equation (6) is of the form

$$A_n + B_n a_n + T_n(a_n) a_n^2 = 0. \tag{7}$$

Let a_n' be the solution of

$$A_n + B_n a_n' = 0. \tag{8}$$

Subtracting (8) from (7) gives that, with probability $\geqq 1 - \epsilon$ for $n > N(\epsilon)$,

$$|a_n - a_n'| = \frac{|T_n(a_n)|\, a_n^2}{|B_n|} \leqq \frac{C a_n^2}{|B_n|}. \tag{9}$$

It will be shown below that $B_n \to \mu \neq 0$ in probability, so that

$$|a_n - a_n'| \,/\, |a_n| \to 0 \text{ in probability.}$$

It is therefore necessary only to find the limiting distribution of a_n' where

$$\begin{aligned} a_n' = \frac{-A_n}{B_n} &= \frac{n^{-1} \sum [F(b_i) - i/n + 1/2n]\, F'(b_i)}{n^{-1} \sum [F(b_i) - i/n + 1/2n]\, F''(b_i) + n^{-1} \sum F'^2(b_i)} \\ &= \frac{n^{-1} \sum [F(b_i) - i/n] F'(b_i) + \frac{1}{2} n^{-2} \sum F'(b_i)}{n^{-1} \sum [F(b_i) - i/n]\, F''(b_i) + n^{-1} \sum F'^2(b_i) + \frac{1}{2} n^{-2} \sum F''(b_i)}. \end{aligned} \tag{10}$$

3. Lemmas. Three lemmas are useful in the proof of the main result, Theorem 2.

LEMMA 1. *For every* $\epsilon > 0$

$$\lim_{n\to\infty} P(\,|n^{-1/2} F'(b_i)[F(b_i) - i/n] - n^{-1/2} \sum F'(F^{-1}(i/n))[F(b_i) - i/n]| > \epsilon) = 0.$$

PROOF. Since

$$\begin{aligned} &n^{1/2}\{n^{-1} \sum |F'(b_i) - F'(F^{-1}(i/n))| \; |F(b_i) - i/n| \,\} \\ &\qquad \leqq \text{l.u.b.}_i\, |F'(b_i) - F'(F^{-1}(i/n))| \; \text{l.u.b.}_i\, |F(b_i) - i/n|\, n^{1/2}, \end{aligned}$$

we need, by the Kolmogoroff theorem, show only that

$$\text{l.u.b.}_i\, |F'(b_i) - F'(F^{-1}(i/n))| \to 0 \text{ in probability.}$$

Suppose $\epsilon > 0$ and $\eta > 0$ given. Let A be the linear set such that $F'(x) > \eta/4$ for $x \,\varepsilon\, A$ and let A' be the complement of A. Let $D = (b_i\,, F^{-1}(i/n))$ be the open interval with end points b_i and $F^{-1}(i/n)$, and $M = \text{l.u.b.}_x\, |F''(x)|$. By the Kolmogoroff Theorem there is an N such that for $n > N$

$$P(\text{l.u.b.}_i\, |F(b_i) - i/n| \geqq \eta^2/16M) < \epsilon. \tag{11}$$

$$\text{l.u.b.}_i\, |\,F(b_i) - i/n| \geqq \left| \int_{b_i}^{F^{-1}(i/n)} F'(x)\, dx \right| \geqq \left| \int_{D \cap A} F'(x)\, dx \right| \geqq \tfrac{1}{4}\eta \,\text{meas}\,[D \cap A].$$

On the set of sample points for which

$$\text{l.u.b.}_i\, |F(b_i) - i/n| < \eta^2/16M$$

we therefore have

$$\text{meas}\,[D \cap A] < \eta/4M \quad \text{for all} \quad i \leqq n.$$

Either meas $(D) < \eta/4M$, in which case $|F'(b_i) - F'(F^{-1}(i/n))| \leqq (\eta/4M)M < \eta$, or there are points γ_i and δ_i for all $i \leqq n$ such that γ_i, $\delta_i \in A'$,

$$|\gamma_i - b_i| < \eta/4M, \qquad |\delta_i - F^{-1}(i/n)| < \eta/4M.$$

Therefore, in the second case

$$\begin{aligned} &|F'(b_i) - F'(F^{-1}(i/n))| \\ &\qquad \leqq |F'(b_i) - F'(\gamma_i)| + |F'(\gamma_i) - F'(\delta_i)| + |F'(\delta_i) - F'(F^{-1}(i/n))| \\ &\qquad \leqq M\,|b_i - \gamma_i| + 2\eta/4 + M\,|\delta_i - F^{-1}(i/n)| \\ &\qquad \leqq \eta/4 + \eta/2 + \eta/4 = \eta. \end{aligned}$$

Therefore this inequality is true in any case except for the set of equations (11). Combining our results we have for all $n > N$

$$P(\text{l.u.b.}_i\,|F'(b_i) - F'(F^{-1}(i/n))| > \eta) < \epsilon,$$

which proves the lemma.

LEMMA 2. *Let* $H_0(x) = F'(F^{-1}(x))$ *for* $0 \leqq x \leqq 1$ *and*

$$H_n(x) = \frac{1}{x}\int_0^x H_{n-1}(t)\,dt. \tag{12}$$

Then both

(a) $\int_0^1 x^{-1} H_n(x)\,dx < \infty$,

(b) $\sum_{k=1}^{\infty} (-1)^{k+1} H_k(x) = \frac{1}{x^2}\int_0^x t\,H_0(t)\,dt$, *uniformly in* x.

PROOF. Let $M = \text{l.u.b.}_x\,|F''(x)|$. Consider the curve $y = F'(x)$ at the point $(x, F'(x))$. If from this point a line of slope M is drawn to the x-axis, the line must be completely on or below the curve $y = F'(x)$ in $(-\infty, x)$, by the mean-value theorem. Since $F(x) = \int_{-\infty}^{x} F'(x)\,dx$, it follows that

$$F(x) \geqq \tfrac{1}{2}F'^2(x)/M. \tag{13}$$

This implies that $\lim_{x\to 0+} H_0(x) = 0$ and by an obvious induction that $\lim_{x\to 0+} H_n(x) = 0$.

Differentiating (12) gives $H_n'(x) = -x^{-1}H_n(x) + x^{-1}H_{n-1}(x)$, or

$$\int_0^y x^{-1} H_n(x)\,dx = \int_0^y x^{-1} H_{n-1}(x)\,dx - H_n(y). \tag{14}$$

Therefore the truth of (a) for $n - 1$ implies it for n and we need show $\int_0^1 x^{-1}H_0(x)\,dx = \int_{-\infty}^{\infty} (F'(x)/F(x))\,dF < \infty$. Define the set E by

$$E = \{x \mid F'(x)/F(x) > F^{-1/2}(x)\}.$$

If E' is the complementary set,

(15) $$\int_{E'} (F'(x)/F(x))\, dF \leqq \int_{E'} F^{-1/2}(x) dF \leqq \int_0^1 x^{-1/2}\, dx = 2.$$

The set E consists of a union of open intervals, by the continuity of F and F'. Let $[a, b]$ be one such interval. On $[a, b]$

$$F'(x \geqq F^{1/2}(x), \qquad \int_a^b F'(x)/F^{1/2}(x)\, dx \geqq \int_a^b dx$$

or $2[F^{1/2}(b) - F^{1/2}(a)] \geqq b - a$. This implies that the measure of E is $\leqq 2$. Using the inequality (13)

(16) $$\int_E (F'(x)/F(x))\, dF = \int_E (F'^2(x)/F(x))\, dx \leqq 2M \int_E dx \leqq 4M.$$

This completes the proof of (a). Turning to (b) we note that by (14) $\int_0^y x^{-1} H_n(x)\, dx$ is a decreasing positive function of n for each y and therefore has a limit as $n \to \infty$. Taking this limit in (14) shows that $\lim_{n\to\infty} H_n(y) = 0$ for all y. Writing

$$\int_0^1 x^{-1} H_n(x)\, dx = \int_0^\eta x^{-1} H_n(x)\, dx + \int_\eta^1 x^{-1} H_n(x)\, dx$$
$$\leqq \int_0^\eta x^{-1} H_0(x)\, dx + \int_\eta^1 x^{-1} H_n(x)\, dx$$

we see that η can be chosen so as to make the first of these arbitrarily small, and that the second then goes to zero by bounded convergence. Therefore $\lim_{n\to\infty} \int_0^1 x^{-1} H_n(x)\, dx = 0$. From (14) we obtain

$$\textstyle\sum_1^n H_k(x) = \int_0^x t^{-1} H_0(t)\, dt - \int_0^x t^{-1} H_n(t)\, dt.$$

Since $|\int_0^x t^{-1} H_n(t)\, dt| \leqq \int_0^1 t^{-1} H_n(t)\, dt \to 0$ it follows that $\sum_1^\infty H_k(x)$ converges uniformly to $\int_0^x t^{-1} H_0(t)\, dt$. Let $J_n(x) = \sum_1^n (-1)^{k+1} H_k(x)$. Then $J_n(x) \to J(x)$ uniformly as $n \to \infty$, and

$$\begin{aligned} J'_n &= \textstyle\sum_1^n (-1)^{k+1} H'_k(x) \\ &= x^{-1}[H_0 - H_1 - H_1 + H_2 \cdots (-1)^{n+1} H_{n-1} + (-1)^n H_n] \\ &= x^{-1} H_0 - 2x^{-1} J_{n-1} + x^{-1}(-1)^k H_n. \end{aligned}$$

The right side converges uniformly for $\epsilon \leqq x \leqq 1$ so that for $0 < x \leqq 1$,

$$J' + 2x^{-1} J = x^{-1} H_0(x),\ x^2 J' + 2xJ = xH_0(x),\ d/dx(x^2 J) = xH_0(x),$$

or $J(x) = x^{-2} \int_0^x tH_0(t)\, dt$, proving (b) and thus Lemma 2.

LEMMA 3. *For every* $\epsilon > 0$ *and* $s = 0, 1, 2, \cdots$

$$\lim_{n\to\infty} P\left[\left|\frac{1}{\sqrt{n}}\sum_1^n H_s\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\} - \sqrt{n}\sum_1^n \int_0^{i/n} H_s(x)\,dx\right.\right.$$

$$\left.\left.\left\{\frac{F(b_i) - F(b_i + 1)}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}}\sum_1^n H_{s+1}\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}\right| > \epsilon\right] = 0.$$

PROOF. In what follows we take $F(b_{n+1}) = 1$ whenever it occurs. Letting $G_k = \sum_{i=1}^k H_s(i/n)$ and using the Abel transformation,

$$\frac{1}{\sqrt{n}}\sum_1^n H_s\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}$$

$$= \frac{1}{\sqrt{n}}\sum_1^{n-1} G_i\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}}\,G_n\{F(b_n) - 1\}$$

$$= \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \frac{1}{\sqrt{n}}\,G_n\{F(b_n) - 1\}$$

$$= \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$(16) \qquad + \frac{1}{\sqrt{n}}\,G_n\,\{F(b_n) - 1\}$$

$$= \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{(i+1)/n}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\,\{F(b_i) - F(b_{i+1})\}$$

$$\left\{\frac{1}{F(b_{i+1})} - \frac{n}{i+1}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$- \sqrt{n} \int_0^1 H_s(x)\, dx\, \{F(b_n) - 1\} \left\{1 - \frac{n+1}{n}\right\}$$

$$- \sqrt{n} \int_0^1 H_s(x)\, dx \left\{(F(b_n) - 1) \frac{n+1}{n} + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}} \{F(b_n) - 1\}\, G_n$$

$$= \sum_1^7 A_i .$$

We first must show that A_2, A_4, A_5, A_6, and A_7 converge to zero in probability. That this is so for A_5, A_6, and A_7 is a simple consequence of the fact that since $F(b_n)$ represents the maximum of n uniformly distributed independent random variables, the distribution function of $1 - F(b_n)$ is $1 - (1 - x)^n$ for $0 \leqq x \leqq 1$, from which it follows that $n^{1-\eta}[1 - F(b_n)]$ converges to zero in probability for $\eta > 0$. Of course, $n^{-1}G_n$ is bounded.

Using the Abel transformation

$$A_2 = \sqrt{n} \sum_1^{n-2} \left\{ \int_{i/n}^{(i+1)/n} H_s(x)\, dx - \frac{1}{n} H_s\left(\frac{i+1}{n}\right)\right\} \left\{F(b_1) - F(b_{i+1}) + \frac{i+1}{n} - \frac{1}{n}\right\}$$

$$+ \sqrt{n} \left\{\frac{1}{n} G_{n-1} - \int_0^{(n-1)/n} H_s(x)\, dx\right\} \left\{F(b_1) - F(b_n) + \frac{n-1}{n}\right\} \tag{17}$$

$$|A_2| \leqq n \frac{\sqrt{n}}{n} \operatorname*{l.u.b.}_i \max_{\frac{i}{n} \leqq x \leqq \frac{i+1}{n}} \left| H_s(x) - H_s\left(\frac{i+1}{n}\right)\right| \operatorname*{l.u.b.}_i \left|F(b_i) - \frac{i}{n}\right|$$

$$+ O(\sqrt{n} \left\{ |1 - F(b_n)| + \frac{1}{n} + F(b_1) \right\}.$$

The last term clearly goes to zero in probability since $n^{1-\eta}F(b_1)$ converges to zero in probability for $\eta > 0$. (The distribution function of $F(b_1)$ is $1 - (1 - x)^n$ for $0 \leqq x \leqq 1$.) By the Kolmogoroff theorem, given $\epsilon > 0$, we may choose A and N_1 such that for $n > N_1$

$$P(\text{l.u.b.}_i\, |F(b_i) - i/n|\, \sqrt{n} > A) < \epsilon.$$

Since $H_s(x)$ is uniformly continuous, we may choose N_2 so that for $n > N_2$

$$\operatorname*{l.u.b.}_i \max_{\frac{i}{n} \leqq x \leqq \frac{i+1}{n}} \left| H_s(x) - H_s\left(\frac{i+1}{n}\right)\right| \leqq \frac{\epsilon}{A}.$$

For $n > \max(N_1, N_2)$ the probability is less than ϵ that the first term on the right of (17) exceeds ϵ.

$$|A_4| \leqq \sqrt{n} \operatorname*{l.u.b.}_i \left|F(b_i) - \frac{i}{n}\right|^2 \sum_1^n \left\{\int_0^{i/n} H_s(x)\, dx\right\} \{F(b_{i+1}) - F(b_i)\} \frac{1}{F(b_{i+1})} \frac{n}{i+1}$$

$$\leqq \sqrt{n} \operatorname*{l.u.b.}_i \left|F(b_i) - \frac{i}{n}\right|^2 \max_x H_s(x) \sum_1^n \frac{F(b_{i+1}) - F(b_i)}{F(b_{i+1})}$$

$$\leqq C\sqrt{n}\,\underset{i}{\text{l.u.b.}}\left|F(b_i) - \frac{i}{n}\right|^2 \int_{F(b_1)}^{1} \frac{dx}{x}$$

$$= C\sqrt{n}\,\underset{i}{\text{l.u.b.}}\left|F(b_i) - \frac{i}{n}\right|^2 \quad |\ln F(b_1)|.$$

By the usual argument $|A_4| \to 0$ in probability. We now return to A_3.

$$A_3 = -\frac{\sqrt{n}}{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$= -\frac{1}{\sqrt{n}}\sum_1^n H_{s+1}\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}$$

$$(18) \qquad + \frac{1}{\sqrt{n}}\sum_1^{n-1} \frac{n}{i+1}\int_{i/n}^{(i+1)/n} H_s(x)\,dx\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \frac{1}{\sqrt{n}}\frac{n}{n+1}\int_0^1 H_s(x)\,dx\cdot\frac{1}{n} + \frac{1}{\sqrt{n}}\,n\int_0^{1/n} H_s(x)\,dx\left\{F(b_1) - \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$= \sum_1^5 B_i\,.$$

To complete the proof of Lemma 3 it is sufficient to show that $|B_i| \to 0$ in probability for $i = 2, 3, 4, 5$. The simplest arguments suffice for $i = 2, 3, 4$. In order to treat B_5 note the identity

$$2\sum (a_i - a_{i+1})a_{i+1} = a_1^2 - a_{K+1}^2 - \sum (a_i - a_{i+1})^2,$$

where sums are from 1 to K. If we set $a_i = F(b_i) - i/n$, we obtain

$$|2\sum \{F(b_i) - F(b_{i+1}) + 1/n\} \quad \{F(b_i) - i/n\}|$$
$$\leqq |F(b_1) - 1/n|^2 + |F(b_{K+1}) - (K+1)/n|^2 + \sum \{F(b_i) - F(b_{i+1}) + 1/n\}^2$$
$$\leqq 2\,\text{l.u.b.}_i\,|F(b_i) - i/n|^2 + \sum \{F(b_i) - F(b_{i+1})\}^2 + 3/n.$$

From the joint distribution of the quantities $\{F(b_i) - F(b_{i+1})\}$ given in [3] it is simple to show that

$$P\{F(b_{i+1}) - F(b_i) \geqq h\} = [1-h]^n, \qquad i = 1, 2, \cdots, n.$$

Therefore

$$P\{\underset{1<i<n}{\text{l.u.b.}}\,|F(b_{i+1}) - F(b_i)| < h\} = 1 - P(\textstyle\bigcup_1^n |F(b_{i+1}) - F(b_i)| \geqq h)$$

$$\geqq 1 - \sum_1^n [1-h]^n = 1 - n[1-h]^n.$$

From this it follows that if $-1 < \alpha < 0$

$$\lim_{n\to\infty} P\{ \operatorname{l.u.b.}_{1\leqq i\leqq n} |F(b_{i+1}) - F(b_i)| < n^{\alpha}\} = 1,$$

so that, choosing $\alpha = -7/8$, there exists an N such that, for $n > N$, with probability greater than $1 - \epsilon$

$$|\sum \{F(b_i) - F(b_{i+1}) + 1/n\} \quad \{F(b_{i+1}) - i/n\}| \leqq \operatorname{l.u.b.}_i |F(b_i) - i/n|^2 + n^{-3/4} + 3/2n.$$

Now applying the Abel transformation to B_5,

$$|B_5| \leqq \sqrt{n}\left[\operatorname{l.u.b.}_{i} \left|F(b_i) - \frac{i}{n}\right|^2 + \frac{1}{n^{3/4}} + \frac{1}{n}\right] \tag{19}$$
$$\cdot \sum_1^{n-1} \left|\frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx - \frac{n}{i+2}\int_0^{(i+1)/n} H_s(x)\,dx\right|$$
$$+ \sqrt{n}\left[\operatorname{l.u.b.}_{i} \left|F(b_i) - \frac{i}{n}\right|^2 + \frac{1}{n^{3/4}} + \frac{3}{2n}\right]\frac{n}{n+1}\int_0^1 H_s(x)\,dx.$$

The second term on the right clearly converges to zero in probability. We observe that

$$x^{-1}\int_0^x H_s(t)\,dt = \int_0^x t^{-1}[H_s(t) - H_{s+1}(t)]\,dt$$

and that by Lemma 2 the right side converges absolutely. Therefore

$$\sum_1^{n-1}\left|\frac{n}{i}\int_0^{i/n} H_s(x)\,dx - \frac{n}{i+1}\int_0^{(i+1)/n} H_s(x)\,dx\right|$$
$$= \sum_1^{n-1}\left|\int_0^{i/n}\frac{1}{x}[H_s - H_{s+1}]\,dx - \int_0^{(i+1)/n}\frac{1}{x}[H_s - H_{s+1}]\,dx\right|$$
$$= \sum_1^{n-1}\left|\int_{i/n}^{(i+1)/n}\frac{1}{x}[H_s - H_{s+1}]\,dx\right| \leqq \int_0^1 \frac{1}{x}\{|H_s| + |H_{s+1}|\}\,dx = C_1.$$

Now

$$\sum_1^{n-1}\left|\frac{n}{i+1}\int_0^{i/n} H_s\,dx - \frac{n}{i+2}\int_0^{(i+1)/n} H_s\,dx\right|$$
$$\leqq \sum_1^{n-1}\left|\frac{n}{i+1}\int_0^{(i+1)/n} H_s\,dx - \frac{n}{i+2}\int_0^{(i+2)/n} H_s\,dx\right|$$
$$+ \sum_1^{n-1}\frac{n}{i+1}\int_{i/n}^{(i+1)/n} H_s\,dx + \sum_1^{n-1}\frac{n}{i+2}\int_{(i+1)/n}^{(i+2)/n} H_s\,dx$$
$$\leqq C_1 + 2\max_{0\leqq x\leqq 1} H_s(x)\sum_1^{n-1}\frac{1}{i} \leqq C_1 + C_2 \ln n.$$

With this estimate the first term in (19) converges to zero in probability and Lemma 3 is completely proved.

4. Proof of main result. In statistics the random variable $F(b_{i+1}) - F(b_i)$ are known as coverages and the random variables

$$\{[F(b_{i+1}) - F(b_i)]/F(b_{i+1})\}, \qquad i = 0, 1, 2, \cdots, n,$$

are independent (as usual $F(b_0) = 0$ and $F(b_{n+1}) = 1$) with density $i(1-u)^{i-1}\,du$. From this one calculates easily that the independent random variables

$$X_{i,n} = \left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\}$$

have mean 0, variance $n^{-2}[i/(2+i)]$, and $E(\,|X_{i,n}|^3\,) \leqq 16n^{-3}$. In what follows it will be necessary to apply the Central Limit Theorem to sums of the form $S_n = \sum_{i=1}^n a_{i,n}X_{i,n}$. Although the Liapunoff version of the Central Limit Theorem is usually stated for sums of the form $(\sum \sigma_i^2)^{-1/2}\sum Y_i$, in this slightly more general form the proof [1] goes through with no change at all. It will be easy to show that the Liapunoff condition,

$$\lim_{n\to\infty} \frac{\sum |a_{i,n}|^3\, E(|X_{i,n}|^3)}{\sum |a_{i,n}|^2\, E(X_{i,n}^2)]^{3/2}} = 0, \tag{20}$$

holds. Let

$$Y_{k,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n H_k(i/n)\left[F(b_i) - \frac{i}{n}\right], \qquad W_{k,n} = \sqrt{n}\sum_{i=1}^n \int_0^{i/n} H_k(x)\,dx\,X_{i,n},$$

$$V_n = \sqrt{n}\sum_{i=1}^n\left\{\int_0^{i/n} H_0(x)\,dx - \int_0^{i/n}\frac{dx}{x^2}\int_0^x tH_0(t)\,dt\right\}X_{i,n}.$$

Theorem 1 shows that we can replace $n^{-1/2}\sum_1^n F'(b_i)[F(b_i) - i/n]$ by V_n which is the sum of independent random variables.

THEOREM 1. *For every* $\epsilon > 0$

$$\lim_{n\to\infty} P\left[\left|n^{-1/2}\sum_1^n F'(b_i)[F(b_i) - i/n] - V_n\right| > \epsilon\right] = 0.$$

PROOF. By Lemma 1 it is sufficient to show that

$$\lim_{n\to\infty} P[|Y_{0,n} - V_n| > \epsilon] = 0. \tag{21}$$

This probability is

$$\begin{aligned} P[|Y_{0,n} - V_n)| > \epsilon] &= P\Bigg[\Bigg|\sum_{k=0}^M (-1)^k\{Y_{k,n} - W_{k,n} + Y_{k+1,n}\} + \sum_{k=0}^M (-1)^k W_{k,n} - V_n \\ &\qquad\qquad + (-1)^{M+1}Y_{M+1,n}\Bigg| > \epsilon\Bigg] \qquad (22) \\ &\leqq \sum_{k=0}^M P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon/3M] \\ &\qquad + P\left[\left|\sum_{k=0}^M (-1)^k W_{k,n} - V_n\right| < \epsilon/3\right] + P[|Y_{M+1,n}| > \epsilon/3]. \end{aligned}$$

The sum in the second term can be evaluated as

$$\sum_{k=0}^{M} (-1)^k W_{k,n} - V_n$$

$$= \sqrt{n} \sum_{i=1}^{n} \left\{ \int_0^{i/n} \sum_{k=0}^{M} (-1)^k H_k(x)\, dx - \int_0^{i/n} H_0(x)\, dx + \int_0^{i/n} \frac{dx}{x^2} \int_0^x tH_0(t)\, dt \right\} X_{i,n}$$

$$= \sqrt{n} \left\{ \sum_{i=1}^{n} \int_0^{i/n} \left[\sum_{k=1}^{M} (-1)^k H_k(x)\, dx + \frac{1}{x^2} \int_0^x tH_0(t)\, dt \right] dx \right\} X_{i,n}.$$

By Lemma 2 the integrals appearing in the sum may be made uniformly less than $\sqrt{\epsilon^3/27}$ for all $M \geqq M_1$. The variance of $\sum (-1)^k W_{k,n} - V_n$ is then $\leqq \epsilon^3/27$ for all n.

By the Tchebycheff inequality

$$P\left[\left| \sum_{k=0}^{M} (-1)^k W_{k,n} - V_n \right| > \frac{\epsilon}{3} \right] \leqq \frac{\epsilon}{3}, \quad \text{all } n, \text{ all } M \geqq M_1; \tag{23}$$

$$|Y_{M+1,n}| \leqq \sqrt{n}\ \text{l.u.b.}_i\ |F(b_i) - i/n|\ \max_x H_{M+1}(x). \tag{24}$$

Since $\lim_{k\to\infty} H_k(x) = 0$ uniformly, we can, by the Kolmogoroff theorem, find an M_2 and N_1 such that for $M \geqq M_2$ and $n > N_1$

$$P(|Y_{M+1,n}| > \epsilon/3) \leqq \epsilon/3. \tag{25}$$

Now fix an M, say $M = \max(M_1, M_2)$. Lemma 3 states that

$$\lim_{n\to\infty} P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon] = 0$$

for all k and all $\epsilon > 0$. Therefore for $N > N_2$

$$\sum P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon/3M] \leqq \epsilon/3. \tag{26}$$

Combining (22), (23), (25), and (26) we obtain $P[|Y_{0,n} - V_n| > \epsilon] \leqq \epsilon$ for all $n > \max(N_1, N_2)$, which proves Theorem 1.

It is now easy to prove Theorem 2, our main result, as stated at the outset. The variance of V_n, which is the sum of independent random variables, is asymptotically for large n

$$\frac{1}{n} \sum_{i=1}^{n} \left[\int_0^{i/n} H_0(t)\, dt - \int_0^{i/n} \frac{dx}{x^2} \int_0^x t\, H_0(t)\, dt \right]^2$$

$$\to \int_0^1 \left| \int_0^y H_0(t)\, dt - \int_0^y \frac{dx}{x^2} \int_0^x tH_0(t)\, dt \right|^2 dy = \nu^2.$$

It is easily verified that the Liapunoff condition (20) is satisfied, so that by the Central Limit Theorem

$$\lim_{n\to\infty} P(V_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\nu} e^{-t^2/2}\, dt.$$

By Theorem 1,

$$(27) \qquad \lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}} \sum_1^n F'(b_i)\left[F(b_i) - \frac{i}{n}\right] < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\nu} e^{-t^2/2}\, dt.$$

Returning to (10),

$$\lim_{n\to\infty} \frac{1}{2n^{3/2}} \sum_1^n F'(b_i) = \lim_{n\to\infty} \frac{1}{2n^2} \sum_1^n F''(b_i) = 0; \frac{1}{n} \sum_1^n \left[F(b_i) - \frac{i}{n}\right] F''(b_i) \to 0$$

in probability.

By the weak law of large numbers, $n^{-1}\sum_1^n F'(b_i)$ converges in probability to $\int_{-\infty}^{\infty} F'^2(x)\, dF(x) = \mu$. Combining these with (9), (10), and (27) gives our final result,

$$\lim_{n\to\infty} P(\sqrt{n}\, a_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\mu/\nu)x} e^{-t^2/2}\, dt.$$

The constants ν and μ cannot be zero. This is obvious for μ. If ν were zero, it would be necessary to have $\int_0^y H_0(t)\, dt = \int_0^y x^{-2}\, dx \int_0^x tH_0(t)\, dt$. Differentiating this twice leads to the equation $H_0(t) = c/t$, which is impossible.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 215–216.

[2] A. Kolmogoroff, "Confidence limits for an unknown distribution function," *Ann. Math. Stat.*, Vol. 12, (1941), pp. 461–463.

[3] S. Malmquist, "On a property of order statistics from a rectangular distribution," *Skand. Aktuarietids*, Vol. 33 (1950), pp. 214–222.