The Annals of Statistics
1974, Vol. 2, No. 5, 1034–1039

CONVERGENCE OF SAMPLE PATHS OF NORMALIZED SUMS OF INDUCED ORDER STATISTICS

BY P. K. BHATTACHARYA

University of Arizona and University of Minnesota

The main result in this paper concerns the limiting behavior of normalized cumulative sums of induced order statistics obtained from n independent two-dimensional random vectors, as n increases indefinitely. By means of a Skorokhod-type embedding of these cumulative sums on Brownian Motion paths, it is shown that under certain conditions the sample paths of these normalized sums converge in a certain sense to a process obtained from the Brownian Motion by a transformation of the time-axis. This yields an invariance principle similar to Donsker's. In particular, the asymptotic distribution of the supremum of the absolute values of these normalized cumulative sums is obtained from a well-known result for the Brownian Motion. Large sample tests of a specified regression function are obtained from these results.

1. Introduction. $(X_1, Y_1), (X_2, Y_2), \cdots$ are independent two-dimensional random vectors each distributed as (X, Y). Let X_{nk} be the kth order statistic obtained from $X_1, \cdots, X_n$. If the marginal distribution of X is continuous, $X_{n1} < \cdots < X_{nn}$ with probability 1 and we can unambiguously define induced order statistics $Y_{n1}, \cdots, Y_{nn}$ as $Y_{nk} = Y_j$ if $X_{nk} = X_j$. Let $m(x)$ denote the conditional expectation and $\sigma^2(x)$ the conditional variance of Y given $X = x$, and let $\psi(t) = \int_{-\infty}^{F^{-1}(t)} \sigma^2(x)\, dF(x), 0 \leqq t \leqq 1$. The main result in this paper concerns the limiting behavior of the sample paths of

$$\{S_{nk} = \sum_{j=1}^{k} (Y_{nj} - m(X_{nj})),\ k = 1, \cdots, n\}\,.$$

By means of a conditional Skorokhod embedding (see Skorokhod (1961), page 163) of $\{S_{nk}\}$ given $X_1, X_2, \cdots$ on Brownian paths (Theorem 1), it is shown that under certain conditions there are processes $\{\xi^{(n)}(t), 0 \leqq t \leqq 1\}$ for each n and a Brownian Motion $\{\xi(t), t \geqq 0\}$ on a common probability space so that $\{\xi^{(n)}(t), 0 \leqq t \leqq 1\}$ has the same distribution as $\{S_{n,[nt]}/(n\psi(1))^{\frac{1}{2}}, 0 \leqq t \leqq 1\}$ and $\sup_{0\leqq t\leqq 1} |\xi^{(n_j)}(t) - \xi(\psi(t)/\psi(1))| \to 0$ a.s. for sufficiently rapidly increasing subsequences $\{n_j\}$ (Theorem 2). This yields an invariance principle similar to Donsker's (1951). In particular, the asymptotic distribution of $\sup_{0\leqq t\leqq 1} |S_{n,[nt]}|/(n\psi(1))^{\frac{1}{2}}$ is the same as the distribution of $\sup_{0\leqq t\leqq 1} |\xi(t)|$. Large sample tests for a specified regression function are obtained from these results.

2. Preliminaries. Let F denote the marginal cdf of X and G_x the conditional cdf of Y given $X = x$. We assume that F and $\{G_x\}$ satisfy the following conditions.

Received February 1972; revised September 1973.

AMS 1970 *subject classifications*. Primary 60F99; Secondary 62E20.

Key words and phrases. Induced order statistics, Skorokhod embedding, invariance principle, test for regression function.

Condition 1. F is continuous.

Condition 2. $\beta(x) = E[\{Y - m(x)\}^4 \mid X = x]$ is bounded above by some constant B on $(-\infty, \infty)$.

Condition 3. $\sigma^2(x) = E[\{Y - m(x)\}^2 \mid X = x]$ is of bounded variation on $(-\infty, \infty)$.

We define functions $\psi(t)$ and $\psi_n(t)$ on $[0, 1]$ as follows. For any cdf H, let $H^{-1}(t) = \inf\{x : H(x) \geqq t\}$, $0 \leqq t \leqq 1$. Then

$$\psi(t) = \int_{-\infty}^{F^{-1}(t)} \sigma^2(x)\, dF(x) \tag{1}$$

and

$$\psi_n(t) = \int_{-\infty}^{F_n^{-1}(t)} \sigma^2(x)\, dF_n(x) \quad \text{for } 1/n \leqq t \leqq 1\,, \qquad = 0 \quad \text{otherwise,} \tag{2}$$

where F_n is the empirical cdf of $X_1, \cdots, X_n$.

We conclude this section with two lemmas. Lemma 1 gives the conditional distribution of $Y_{n1}, \cdots, Y_{nn}$ given $X_1, \cdots, X_n$ and Lemma 2 establishes the almost sure uniform convergence of $\psi_n(t)$ to $\psi(t)$.

LEMMA 1. *Under Condition* 1, *for every n and almost all $(X_1, \cdots, X_n)$, $Y_{n1}, \cdots, Y_{nn}$ are conditionally independent given $X_1, \cdots, X_n$ with conditional* cdf's $G_{X_{n1}}, \cdots, G_{X_{nn}}$ *respectively.*

PROOF. For any $\mathbf{x}_n = (x_1, \cdots, x_n)$ no two coordinates of which are equal, let $\lambda(k, \mathbf{x}_n) = j$ if x_j is the kth smallest among $x_1, \cdots, x_n$. By Condition 1, $\lambda(k, \mathbf{X}_n)$, $k = 1, \cdots, n$ are defined a.s. and $X_{nk} = X_{\lambda(k,\mathbf{X}_n)}$, $Y_{nk} = Y_{\lambda(k,\mathbf{X}_n)}$. Hence the conditional joint cdf of $Y_{n1}, \cdots, Y_{nn}$ given $X_1, \cdots, X_n$ is the same as the conditional joint cdf of $Y_{\lambda(k,\mathbf{X}_n)}$, $k = 1, \cdots, n$ given $X_{\lambda(k,\mathbf{X}_n)}$, $k = 1, \cdots, n$, which is easily seen to be the product $\prod_{k=1}^n G_{X_{\lambda(k,\mathbf{X}_n)}} = \prod_{k=1}^n G_{X_{nk}}$ due to the independence of Y_i and X_j for every $i \neq j = 1, \cdots, n$.

REMARK. Lemma 1 holds much more generally. In fact if $\lambda(1), \cdots, \lambda(n)$ is any random permutation of $1, \cdots, n$ determined by $X_1, \cdots, X_n$, then $Y_{\lambda(1)}, \cdots, Y_{\lambda(n)}$ are conditionally independent given $X_1, \cdots, X_n$ with conditional cdf's $G_{X_{\lambda(1)}}, \cdots, G_{X_{\lambda(n)}}$. Moreover, for this the condition that F is continuous is not necessary. The only reason this condition is imposed here is to define the induced order statistics in a simple manner and to avoid unnecessary complications.

LEMMA 2. *Under Condition* 3, $\sup_{0 \leq t \leq 1} |\psi_n(t) - \psi(t)| \to 0$ a.s.

PROOF. Since $\sup \sigma^2(x) < \infty$ and $\sigma^2(x)$ is of bounded variation on $(-\infty, \infty)$, the lemma is proved by integration by parts and application of the Glivenko–Cantelli theorem.

3. Convergence of sample paths of $\{S_{nk}\}$. Construct a probability space $(\Omega, \mathscr{F}, P)$ by adjoining an independent Brownian Motion $\xi(t)$ to the probability space of $(X_1, Y_1), (X_2, Y_2), \cdots$ and let $\mathscr{A} \subset \mathscr{F}$ denote the σ-field of $X_1, X_2, \cdots$. We first obtain a conditional Skorokhod representation of $\{S_{nk}, k = 1, \cdots, n\}$ given $\mathscr{A}$. For two stochastic processes we write $\{X(t)\} =_d \{Y(t)\}$ to indicate that the processes have the same distribution.

THEOREM 1. *If Condition* 1 *holds and if* $\beta(x) = E[\{Y - m(x)\}^4 \mid X = x]$ *exist for all* x, *then for every* n, *there exist stopping times* $T_{n1}, \cdots, T_{nn}$ *of the Brownian Motion* $\{\xi(t), t \geqq 0\}$ *such that*

(a) $(S_{n1}, \cdots, S_{nn}) =_d (\xi(T_{n1}), \cdots, \xi(T_{n1} + \cdots + T_{nn}))$.
(b) $T_{n1}, \cdots, T_{nn}$ *are conditionally independent given* $\mathscr{A}$ a.s.
(c) $E[T_{nk} \mid \mathscr{A}] = \sigma^2(X_{nk})$ a.s.
(d) $E[T_{nk}^2 \mid \mathscr{A}] \leqq C\beta(X_{nk})$ a.s., *where* C *is a constant.*

PROOF. Argue conditionally given $\mathscr{A}$ in $(\Omega, \mathscr{F}, P)$. Then by Lemma 1, the random variables $Y_{nk} - m(X_{nk})$, $k = 1, \cdots, n$ are mutually independent with mean 0, variances $\sigma^2(X_{nk})$ and fourth moments $\beta(X_{nk})$ for almost all sample points. In the conditional argument the theorem thus becomes the same as the well-known theorem of Skorokhod (1961, page 163).

By means of the above embedding theorem we now study the convergence of normalized cumulative sums of induced order statistics. The following is the main theorem of this section.

THEOREM 2. *Under Conditions* 1–3, *there exist processes* $\{\xi^{(n)}(t), 0 \leqq t \leqq 1\}$ *and a Brownian Motion* $\{\xi(t), t \geqq 0\}$ *on a common probability space such that*

(a) *for each* n,

$$\{\xi^{(n)}(t), 0 \leqq t \leqq 1\} =_d \{S_{n,[nt]}/(n\psi(1))^{\frac{1}{2}}, 0 \leqq t \leqq 1\},$$

(b) *for any sufficiently rapidly increasing subsequence* $\{n_j\}$

$$\sup_{0 \leqq t \leqq 1} |\xi^{(n_j)}(t) - \xi(\psi(t)/\psi(1))| \to 0 \quad \text{a.s.},$$

where $\psi(t)$ *is as defined in* (1).

PROOF. We shall prove the theorem in the context of the probability space $(\Omega, \mathscr{F}, P)$. For each n, construct random stopping times $T_{n1}, \cdots, T_{nn}$ of $\xi(t)$ as in Theorem 1. Then for each n,

$$\{S_{n,[nt]}/(n\psi(1))^{\frac{1}{2}}, 0 \leqq t \leqq 1\} =_d \left\{\frac{1}{(n\psi(1))^{\frac{1}{2}}} \xi(T_{n1} + \cdots + T_{n,[nt]}), 0 \leqq t \leqq 1\right\}$$
$$=_d \left\{\xi\left(\frac{T_{n1} + \cdots + T_{n,[nt]}}{n\psi(1)}\right), 0 \leqq t \leqq 1\right\}.$$

Thus the processes

$$\left\{\xi^{(n)}(t) = \xi\left(\frac{T_{n1} + \cdots + T_{n,[nt]}}{n\psi(1)}\right), 0 \leqq t \leqq 1\right\}$$

satisfy (a). We shall now show that these processes also satisfy (b). Arguing as in the proof of Theorem 13.8 of Breiman (1968) and using Lemma 2, it will suffice to show that $\sup_{0\le t\le 1}|n^{-1}\sum_{k=1}^{[nt]} T_{nk} - \psi_n(t)| \to_p 0$ as $n\to\infty$, where $\psi_n(t)$ is as defined in (2).

Now for any $\varepsilon > 0$ and $n > 1/\varepsilon$,

$$\begin{aligned}\sup_{0\le t\le 1}&|n^{-1}\sum_{k=1}^{[nt]} T_{nk} - \psi_n(t)| \\ &\le \sup_{1/n\le t\le 1}(t/[nt])|\sum_{k=1}^{[nt]}\{T_{nk}-\sigma^2(X_{nk})\}| \\ &\le \varepsilon \sup_{1\le k\le[\varepsilon n]}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| \\ &\quad + \sup_{[\varepsilon n]\le k\le n}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}|\,. \end{aligned}\tag{3}$$

We now apply Theorem 1(b), (c), (d) the Hájek–Rényi (1955) inequality, and use Condition 2 to get

$$P[\sup_{1\le k\le[\varepsilon n]}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > x\,|\,\mathscr{A}] \le CBx^{-2}\sum_{k=1}^{[\varepsilon n]}k^{-2} \quad \text{a.s.}$$

and therefore,

$$P[\sup_{1\le k\le[\varepsilon n]}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > x] \le CBx^{-2}\sum_{k=1}^{[\varepsilon n]}k^{-2}\,,\tag{4}$$

and

$$P[\sup_{[\varepsilon n]\le k\le n}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > x\,|\,\mathscr{A}] \le CBx^{-2}n\{[\varepsilon n]\}^{-2} \quad \text{a.s.}$$

and therefore,

$$P[\sup_{[\varepsilon n]\le k\le n}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > x] \le CBx^{-2}n\{[\varepsilon n]\}^{-2}\,.\tag{5}$$

From (3), (4) and (5) we have for any $\delta > 0$ and $\varepsilon > 0$,

$$\begin{aligned}\limsup_{n\to\infty} &P[\sup_{0\le t\le 1}|n^{-1}\sum_{k=1}^{[nt]} T_{nk} - \psi_n(t)| > \delta] \\ &\le \limsup_{n\to\infty} P[\varepsilon \sup_{1\le k\le[\varepsilon n]}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > \delta/2] \\ &\quad + \limsup_{n\to\infty} P[\sup_{[\varepsilon n]\le k\le n}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > \delta/2] \\ &= \limsup_{n\to\infty} P[\sup_{1\le k\le[\varepsilon n]}|k^{-1}\sum_{j=1}^{k}\{T_{nj}-\sigma^2(X_{nj})\}| > \delta/2\varepsilon] + 0 \\ &\le 4CB\varepsilon^2\delta^{-2}\sum_{k=1}^{\infty}k^{-2}\,,\end{aligned}$$

which goes to 0 for any given $\delta > 0$ by allowing ε to tend to zero. This concludes the proof.

REMARK. Theorem 2 implies weak convergence in the uniform topology (see e.g. Breiman (1968), Theorem 13.12). In particular, the asymptotic distribution of $\sup_{0\le t\le 1}|S_{n,[nt]}|(n\psi(1))^{\frac{1}{2}}$ is the same as the distribution of $\sup_{0\le t\le 1}|\xi(\psi(t)/\psi(1))| = \sup_{0\le t\le 1}|\xi(t)|$, the last equality being a consequence of the fact that by Condition 1, $\psi(t)/\psi(1)$ increases continuously from 0 to 1 as t increases from 0 to 1.

4. Testing a specified regression function. Using the results of the last section, we can construct tests for a specified regression function. We want to test the null hypothesis that the regression function $m(x)$ of Y on X in a bivariate distribution is equal to a specified function $m_0(x)$. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent samples from this distribution. We then compute the order statistics $X_{n1}, \cdots, X_{nn}$ of the X-observations and the induced order statistics $Y_{n1}, \cdots, Y_{nn}$

of the Y observations, and let

$$S_{nk} = \sum_{j=1}^{k} \{Y_{nj} - m_0(X_{nj})\} .$$

Then under the null hypothesis, in view of the Remark following Theorem 2,

$$(6) \qquad P[\max_{k=1,\cdots,n} |S_{nk}|/(n\psi(1))^{\frac{1}{2}} \leqq \lambda] = \sum_{k=-\infty}^{\infty} (-1)^k \int_{(2k-1)\lambda}^{(2k+1)\lambda} (2\pi)^{-\frac{1}{2}} e^{-x^2/2}\, dx .$$

However, $\psi(1) = \int_{-\infty}^{\infty} \sigma^2(x)\, dF(x)$ is unknown, but

$$\hat{\psi}_n(1) = n^{-1} \sum_{j=1}^{n} \{Y_{nj} - m_0(X_{nj})\}^2 = n^{-1} \sum_{j=1}^{n} \{Y_j - m_0(X_j)\}^2$$

is a consistent estimator of $\psi(1)$ and (6) holds with $\psi(1)$ replaced by $\hat{\psi}_n(1)$. We can now use the large sample level α test:

Test 1. Reject the null hypothesis if and only if

$$\max_{k=1,\cdots,n} |S_{nk}|/(n\hat{\psi}_n(1))^{\frac{1}{2}} \geqq \lambda_\alpha$$

where $\sum_{k=-\infty}^{\infty} (-1)^k \int_{(2k-1)\lambda_\alpha}^{(2k+1)\lambda_\alpha} (2\pi)^{-\frac{1}{2}} e^{-x^2/2}\, dx = 1 - \alpha$.

The invariance principle also applies to the asymptotic distribution of

$$\{n\psi(1)\}^{-\frac{1}{2}} \int_0^1 S_{n,[nt]}\, dt.$$

Thus under the null hypothesis, $\{n\psi(1)\}^{-\frac{1}{2}} \int_0^1 S_{n,[nt]}\, dt$ converges in distribution to $\int_0^1 \xi(\psi(t)/\psi(1))\, dt$ where $\xi(t)$ is a Brownian Motion. It is easily seen that $\int_0^1 \xi(\psi(t)/\psi(1))\, dt$ is a normal random variable with mean 0 and variance $\{\psi(1)\}^{-1} \int_0^1 \int_0^1 \psi(\min(s, t))\, ds\, dt$. Hence under the null hypothesis,

$$\int_0^1 S_{n,[nt]}\, dt/[n \int_0^1 \int_0^1 \psi(\min(s, t))\, ds\, dt]^{\frac{1}{2}}$$

is asymptotically normally distributed with mean 0 and variance 1. The function $\psi(t)$ can be estimated from the sample by

$$\hat{\psi}_n(t) = n^{-1} \sum_{k=1}^{[nt]} \{Y_{nk} - m_0(X_{nk})\}^2 .$$

To see that $\hat{\psi}_n(t)$ is a uniformly consistent estimate of $\psi(t)$, note that

$$\sup_{0\leqq t\leqq 1} |\hat{\psi}_n(t) - \psi(t)| \leqq \sup_{0\leqq t\leqq 1} |\hat{\psi}_n(t) - \psi_n(t)| + \sup_{0\leqq t\leqq 1} |\psi_n(t) - \psi(t)|$$

where $\psi_n(t)$ is as defined in (2). By Lemma 2, $\sup_{0\leqq t\leqq 1} |\psi_n(t) - \psi(t)| \to 0$ a.s., and it can be shown in a way analogous to the proof of Theorem 2, that $\sup_{0\leqq t\leqq 1} |\hat{\psi}_n(t) - \psi_n(t)| \to_p 0$. Hence,

$$\int_0^1 \int_0^1 \hat{\psi}_n(\min(s, t))\, ds\, dt \to_p \int_0^1 \int_0^1 \psi(\min(s, t))\, ds\, dt ,$$

and consequently,

$$W_n = \int_0^1 S_{n,[nt]}\, dt/[n \int_0^1 \int_0^1 \hat{\psi}_n(\min(s,t))\, ds\, dt]^{\frac{1}{2}}$$

is also asymptotically normally distributed with mean 0 and variance 1 under the null hypothesis. We can now use the following large sample level α tests:

Test 2a. Reject the null hypothesis if and only if

$$W_n \geqq \Phi^{-1}(1 - \alpha) ,$$

or

Test 2b. Reject the null hypothesis if and only if

$$W_n \leqq \Phi^{-1}(\alpha)\ ,$$

where Φ is the cdf of a normal random variable with mean 0 and variance 1.

By a little algebraic simplification, we have

$$W_n = \sum_{j=1}^{n-1} (n - R_{nj})\{Y_j - m_0(X_j)\}/[\sum_{j=1}^{n-1} (n^2 - R_{nj}^2)\{Y_j - m_0(X_j)\}^2]^{\frac{1}{2}}\ ,$$

where R_{nj} is the rank of X_j among $X_1, \cdots, X_n$. In this form, W_n is computed easily.

Test 1 would guard against all possible alternatives, whereas Tests 2a and 2b would guard against alternatives $m(x) > m_0(x)$ and $m(x) < m_0(x)$ respectively.

REFERENCES

[1] Breiman, Leo (1968). *Probability*. Addison-Wesley, Reading.

[2] Donsker, M. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* No. 6.

[3] Hájek, J. and Rényi, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.* **6** 281–283.

[4] Skorokhod, A. V. (1961). *Studies in the Theory of Random Processes*. Kiev Univ. Press. (English translation (1965). Addison-Wesley, Reading. The page number referred to in the text is from the English translation.)

Department of Mathematics
University of Arizona
Tucson, Arizona 85721