AN EXTENSION OF THE ARC SINE LAW[1]

By Simeon M. Berman

Columbia University

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of random variables and let

$$S_0 = 0$$
$$S_n = X_1 + \cdots + X_n; \qquad (n \geqq 1)$$

L_n = index at which S_i, $i = 0, 1, \cdots, n$ attains for the first time the value max $(S_0, \cdots, S_n)$;

M_n = index at which S_i, $i = 0, 1, \cdots, n$ attains for the last time the value min $(S_0, \cdots, S_n)$;

N_n = number of $S_0, S_1, \cdots, S_n$ which are positive.

The arc sine law is a general class of limit theorems dealing with the limiting distributions of L_n, M_n and N_n. The first results on the arc sine law were obtained under the assumption of the independence of the $\{X_n\}$ [2], [4], [5]. Later results were obtained using only certain symmetry properties but not necessarily independence [1].

The primary result of this paper is a generalization of E. S. Andersen's arc sine law for sequences of "symmetrically dependent" random variables; this generalization is obtained with the aid of a representation theorem of de Finetti [6, p. 365]. According to that theorem, a sequence of exchangeable (symmetrically dependent) random variables can be represented as conditionally independent random variables. The limiting distributions in the case of exchangeability are obtained from those in the case of independence.

2. Preliminary results.

Lemma 1. *Let $\{X_n\}$ be a sequence of independent random variables with a common distribution function $F(x)$. Then $\lim_{n\to\infty} P\{S_n = 0\}$ is 0 or 1 depending on whether $F(x)$ is nondegenerate or degenerate at 0.*

Proof. The degenerate case is trivial; the nondegenerate case follows from a theorem of Doeblin and Lévy [3] according to which the distribution of the sum of a large number of independent random variables, whose dispersions are uniformly bounded away from zero, has at most small jumps at its points of discontinuity.

Lemma 2. *Let $\{X_n\}$ be a sequence of independent random variables with a common nondegenerate distribution function $F(x)$, which is symmetric, i.e., $1 - F(x) = F(-x)$ at all points of continuity. Let K_n denote any of the random variables L_n, $n - M_n$, or N_n; then*

$$\lim_{n\to\infty} P\{K_n \leqq \alpha n^{\frac{1}{2}}\} = (2/\pi) \sin^{-1} \alpha^{\frac{1}{2}}, \qquad 0 \leqq \alpha \leqq 1. \tag{2.1}$$

Received September 11, 1961; revised January 22, 1962.

[1] Work done under a grant from the National Science Foundation.

PROOF. According to Theorem 3 of [1], (2.1) holds if $\{X_n\}$ is sequence of independent, identically distributed random variables and

$$\lim_{n\to\infty} P\{S_n > 0\} = \tfrac{1}{2}. \tag{2.2}$$

Since $\{X_n\}$ has a symmetric and nondegenerate distribution function, so does $\{S_n\}$; therefore, (2.2) follows from Lemma 1.

3. Extension to exchangeable (symmetrically dependent) sequences. Let $\{X_n\}$ be a sequence of random variables on a probability space $(\Omega\ \mathfrak{A}, P)$ such that the joint d.f. of $(X_1, X_2, \cdots, X_n)$ say $G_n(x_1, \cdots, x_n)$, is a symmetric function for each n. Such random variables were called "symmetrically dependent" by E. S. Andersen [1] and "exchangeable" by Loève [6, p. 364]; the latter's terminology will be used here. According to the fundamental theorem of de Finetti, as formulated by Loève [6, p. 365], there exists a sub-sigma-field $\mathfrak{F}$ of the sigma-field $\mathfrak{A}$ and a conditional d.f. $G_\omega(x)$ such that the $\{X_n\}$ are conditionally independent given $\mathfrak{F}$ with the common conditional d.f. $G_\omega(x)$. More specifically, one may write,

$$G_n(x_1, \cdots, x_n) = \int_\Omega G_\omega(x_1) \cdots G_\omega(x_n)\, dP(\omega) \tag{3.1}$$

where $G_\omega(x)$ is a d.f. for each $\omega \in \Omega$, and an $\mathfrak{F}$-measurable function for each x. In general, for any set $H \in \mathfrak{A}$,

$$P(H) = \int_\Omega P_\omega(H)\, dP(\omega) \tag{3.1$'$}$$

where $P_\omega(H)$ is the conditional probability of H computed under the assumption that the X_n are mutually independent with the common conditional d.f. $G_\omega(x)$.

DEFINITION. The sequence $\{X_n\}$ is S-invariant if every finite dimensional d.f. of the sequence is invariant under any changes in the signs of the $\{X_n\}$.

It is obvious that a sequence of independent random variables with a common d.f. $F(x)$ is S-invariant if and only if $F(x)$ is symmetric, i.e., every one-dimensional d.f. is invariant under changes in signs.

THEOREM 1. *For a sequence of exchangeable random variables, the following three conditions are equivalent.*

(i) *$\{X_n\}$ is S-invariant.*

(ii) *The joint* d.f. *of* (X_1, X_2), $G_2(x_1, x_2)$, *is invariant under changes in the signs of* (X_1, X_2).

(iii) *In the representation* (3.1), *for almost all* ω,

$$G_\omega(x) = 1 - G_\omega(-x), \tag{3.2}$$

for all x, in the continuity set of $G_\omega(x)$.

PROOF. The implication (i)-(ii) requires no proof, while (iii)-(i) follows from (3.1). It remains only to show that (ii)-(iii).

Let x be a continuity point of $G_1(x)$. The point (x, x) is then a continuity point of $G_2(x_1, x_2)$. It follows from (ii) that

$$
\begin{aligned}
1 &= P\{X_1 \leqq x\} + P\{X_1 \leqq -x\} \\
&= \int_\Omega [G_\omega(x) + G_\omega(-x)]\, dP(\omega),
\end{aligned}
\tag{3.3}
$$

and that

$$
\begin{aligned}
1 &= P\{X_1 \leqq x, X_2 \leqq x\} + P\{X_1 \leqq x, X_2 > x\} \\
&\quad + P\{X_1 > x, X_2 \leqq x\} + P\{X_1 > x, X_2 > x\} \\
&= P\{X_1 \leqq x, X_2 \leqq x\} + 2P\{X_1 \leqq x, X_2 \leqq -x\} \\
&\quad + P\{X_1 \leqq -x, X_2 \leqq -x\} \\
&= \int_\Omega [G_\omega(x) + G_\omega(-x)]^2\, dP(\omega).
\end{aligned}
\tag{3.4}
$$

It follows from (3.3) and (3.4) that for each x in the continuity set of $G_1(x)$, (3.2) holds for all ω except for those in a set $\Gamma(x)$ of probability zero. Since the continuity set is everywhere dense, some denumerable subset σ is also dense, and for each ω outside $\Gamma = U_{x\varepsilon\sigma}\,\Gamma(x)$, (3.2) holds for all $x \varepsilon \sigma$, and Γ has probability zero.

It was proved in [1] that (2.1) holds if $\{X_n\}$ is exchangeable, S-invariant, and

$$
P\{S_n = 0\} = 0 \qquad n = 1, 2, \cdots. \tag{3.5}
$$

It will now be shown that the sequence δ_n defined by $\delta_n = P\{S_n = 0\}$ always coverges to a limit δ and that (2.1) holds if $\delta = 0$ while $n^{-1}K_n$ has a more general limiting d.f. if $\delta > 0$.

LEMMA 3. *Let $\{X_n\}$ be exchangeable and S-invariant; let*

$$
\Delta = \{\omega\colon G_\omega(\cdot) \text{ is degenerate at } 0\}.
$$

Then, $P(\Delta) \doteq \delta = \lim_{n\to\infty} \delta_n$.

PROOF. According to (3.1′),

$$
\delta_n = \int_\Delta P_\omega\{S_n = 0\}\, dP(\omega) + \int_{\Omega-\Delta} P_\omega\{S_n = 0\}\, dP(\omega).
$$

The integrand in the first integral is equal to 1, while the integrand in the second integral converges boundedly to 0 by Lemma 1.

THEOREM 2. *Let $\{X_n\}$ be exchangeable and S-invariant; then*

$$
\begin{aligned}
\lim_{n\to\infty} P\{n^{-1}K_n \leqq \alpha\} &= 0, && \alpha < 0, \\
&= \delta, && \alpha = 0, \\
&= \delta + (1 - \delta)(2/\pi) \sin^{-1} \alpha^{\frac{1}{2}}, && 0 < \alpha < 1, \\
&= 1, && \alpha \geqq 1.
\end{aligned}
$$

PROOF. According to (3.1′),

$$P\{n^{-1}K_n \leqq \alpha\} = \left(\int_\Delta + \int_{\Omega-\Delta}\right) P_\omega\{n^{-1}K_n \leqq \alpha\}\, dP(\omega).$$

For $\omega \varepsilon \Delta$,

$$\begin{aligned} \lim_{n\to\infty} P_\omega\{n^{-1}K_n \leqq \alpha\} &= 0, & \alpha &< 0 \\ &= 1, & \alpha &\geqq 0; \end{aligned}$$

on the other hand, by Theorem 1 and Lemma 2, for almost all $\omega \notin \Delta$,

$$\begin{aligned} \lim_{n\to\infty} P_\omega\{n^{-1}K_n \leqq \alpha\} &= 0, & \alpha &< 0, \\ &= (2/\pi) \sin^{-1} \alpha^{\frac{1}{2}}, & 0 \leqq \alpha &\leqq 1, \\ &= 1, & \alpha &> 1. \end{aligned}$$

The theorem follows by an application of the bounded convergence theorem.

The result in [1] in a special case of Theorem 2 with $\delta_n = 0$. Theorem 2 also shows that (2.1) is valid even if the weaker relation $\delta_n \rightarrow 0$ holds, while a modified limiting d.f. is obtained in all other cases.

4. Acknowledgments. The author is grateful to Professor Gisiro Maruyama for helpful comments, and to the referee for the references used in the proofs of Lemmas 1 and 2. The author also thanks Professor E. J. Gumbel for obtaining the support necessary for this work.

REFERENCES

[1] ANDERSEN, E. S. (1953) and (1954). On the fluctuations of sums of random variables. *Math. Scand.* **1** 263–285 and **2** 195–223.

[2] DARLING, D. A. (1951). Sums of symmetrical random variables. *Proc. Amer. Math. Soc.* **2** 511–517.

[3] DOEBLIN, W. and LÉVY, P. (1936). Sur les sommes de variables aleatoires independants a dispersions borneés inférieurment. *C. R. Acad. Sci.*, Paris **202** 2027–2028.

[4] ERDOS, P. and KAC, M. (1947). On the number of positive sums of independent random variables. *Bull. Amer. Math. Soc.* **53** 1011–1020.

[5] LÉVY, P. (1939). Sur certains processus stochastique homogènes. *Compositio Math.* **7** 283–339.

[6] LOÈVE, M. (1955). *Probability Theory.* Van Nostrand, New York.