A FURTHER GENERALIZATION OF THE ARC-SINE LAW

C. C. HEYDE [1]

(Received 15 December 1966)

1. Introduction

Let X_i, $i = 1, 2, 3, \cdots$ be a sequence of independent and identically distributed random variables and write $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$, $n \geqq 1$. Let $I_n(0), I_n(1), \cdots, I_n(n)$ be that unique permutation of $1, 2, \cdots, n$ such that $S_{I_n(0)} \leqq S_{I_n(1)} \leqq \cdots \leqq S_{I_n(n)}$ and such that if $S_j = S_k$ with $j < k$ then $I_n(k) < I_n(j)$. Thus, $I_n(j)$ is an index of the j-th largest partial sum.

In this note, we shall obtain the distribution of the order index $I_n(j)$ in terms of the distribution of the number of positive partial sums in the sequence $0 = S_0, S_1, \cdots, S_n$. Then, under the condition

$$\lim_{n\to\infty} \frac{\Pr(S_1 > 0) + \cdots + \Pr(S_n > 0)}{n} = \alpha, \qquad 0 \leqq \alpha \leqq 1, \tag{1}$$

we shall go on to obtain the limit distribution $\lim_{n\to\infty} \Pr\{I_n([na]) \leqq nx\}$, $0 \leqq a \leqq 1$. This will be seen to constitute a generalization of the limit result of Spitzer [3], Theorem 7.1, on the number of positive partial sums S_k, $0 \leqq k \leqq n$, and proceeds along the lines of an extension of the work of Darling [1]. As with the result of Spitzer, no limit distribution will exist if the condition (1) is not satisfied.

2. Distribution of the order indices

For $n \geqq 0$, take N_n as the number of positive S_k, $0 \leqq k \leqq n$. In addition to the sequence $\{S_k, k = 0, 1, \cdots, n\}$, we introduce for each fixed j the two further sequences

$$\begin{aligned}
S'_0 &= 0, & S''_0 &= 0,\\
S'_1 &= X_j, & S''_1 &= X_{j+1},\\
S'_2 &= X_j + X_{j-1}, & S''_2 &= X_{j+1} + X_{j+2},\\
&\cdots, & &\cdots,\\
S'_j &= X_j + X_{j-1} + \cdots + X_1, & S''_{n-j} &= X_{j+1} + X_{j+2} + \cdots + X_n,
\end{aligned}$$

[1] Research carried out at Aarhus University.

https://doi.org/10.1017/S1446788700005437 Published online by Cambridge University Press

and define random variables N'_j and N''_{n-j} with respect to the S'_i and S''_i, respectively, in the same way as with N_n for the S_i.

Let us look at the event $\{I_n(k) = j\}$. That is, k is one plus the number out of $S_0, S_1, \cdots, S_{j-1}$ that are less than S_j plus the number out of $S_{j+1}, S_{j+2}, \cdots, S_n$ that are less than or equal to S_j. Clearly, the number of $S_0, S_1, \cdots, S_{j-1}$ that are less than S_j is precisely the number of $S'_1, S'_2, \cdots, S'_j$ that are positive or, in other words, N'_j. Furthermore, the number of $S_{j+1}, S_{j+2}, \cdots, S_n$ that are less than or equal to S_j is just the number of $S''_1, S''_2, \cdots, S''_{n-j}$ that are non-positive which is $n-j-1-N''_{n-j}$. We therefore see that the events $\{I_n(k) = j\}$ and $\{N'_j+n-j-N''_{n-j} = k\}$ are the same. Now the primed and double primed random variables are independent as they depend on disjoint subsets of the X_i. Thus,

$$\Pr\{I_n(k) = j\} = \sum_{\nu=\max(0,j+k-n)}^{\min(j,k)} \Pr(N'_j = \nu) \Pr(N''_{n-j} = n-j-k+\nu).$$

Also, the X_i are identically distributed so the prime and double prime can conveniently be dropped at this stage and we obtain the distribution,

$$\Pr\{I_n(k) = j\} = \sum_{\nu=\max(0,j+k-n)}^{\min(j,k)} \Pr(N_j = \nu) \Pr(N_{n-j} = n-j-k+\nu). \tag{2}$$

This result is a generalization of the result of Theorem 1 of Darling [1] which relates to random variables which have continuous and symmetric distributions.

Using the well-known result of Sparre-Andersen that

$$\Pr(N_n = k) = \Pr(N_k = k) \Pr(N_{n-k} = 0), \; 0 \leqq k \leqq n,$$

we have

$$\begin{aligned}\Pr(N_j = \nu) \Pr(N_{n-j} = n-j-k+\nu) \\ &= \Pr(N_\nu = \nu) \Pr(N_{j-\nu} = 0) \Pr(N_{n-j-k+\nu} = n-j-k+\nu) \Pr(N_{k-\nu} = 0) \\ &= \Pr(N_k = \nu) \Pr(N_{n-k} = n-j-k+\nu),\end{aligned}$$

so that from (2),

$$\Pr\{I_n(k) = j\} = \Pr\{I_n(j) = k\}. \tag{3}$$

We have therefore established the following theorem.

THEOREM 1. *The random variable $I_n(j)$ has the same distribution as the random variable $N'_j+n-j-N''_{n-j}$, the primed and double primed random variables being independent.*

https://doi.org/10.1017/S1446788700005437 Published online by Cambridge University Press

3. Limit theorem

We shall establish the following theorem.

THEOREM 2. *Suppose the random variables X_i are such that the condition* (1) *is satisfied. Then, for* $0 \leqq a \leqq 1$,

$$\lim_{n\to\infty} \Pr \{n^{-1}I_n([na]) \leqq x\} = G_{a,\alpha}(x),$$

where

(4)

$$G_{a,\alpha}(x) = \left(\frac{\sin \pi\alpha}{\pi}\right)^2 \int_0^x \left\{\int_{v=\max(0,u+a-1)}^{\min(u,a)} \frac{dv}{(u-v)^\alpha (1-a-u+v)^{1-\alpha} v^{1-\alpha}(a-v)^\alpha}\right\} du$$

$$(0 \leqq x \leqq 1,\ 0 < a < 1,\ 0 < \alpha < 1),$$

$$G_{0,\alpha}(x) = 1-F_\alpha(1-x), \qquad G_{1,\alpha}(x) = F_\alpha(x),$$

$$G_{a,0}(x) = \begin{cases} 0 & (x < 1-a), \\ 1 & (x \geqq 1-a), \end{cases} \qquad G_{a,1}(x) = \begin{cases} 0 & (x < a), \\ 1 & (x \geqq a), \end{cases}$$

and $F_\alpha(x)$ *is given in the relations* (5). *If the condition* (1) *is not satisfied then* $\Pr \{n^{-1} I_n([na]) \leqq x\}$ *does not tend to a limit as* $n \to \infty$.

PROOF. From Theorem 1 we see that $n^{-1} I_n([na])$ has the same distribution as $n^{-1} N'_{[na]}+1-n^{-1}[na]-n^{-1}N''_{n-[na]}$, the primed and double primed terms being independent. Further, the results of Spitzer [3], Theorem 7.1, tell us that as $n \to \infty$, $n^{-1}N_n$ converges in law to a random variable with distribution function F_α given by

$$F_0(x) = \begin{cases} 0 & (x < 0), \\ 1 & (x \geqq 0), \end{cases}$$

$$(5) \qquad F_\alpha(x) = \frac{\sin \pi\alpha}{\pi} \int_0^x u^{\alpha-1}(1-u)^{-\alpha} du \qquad (0 \leqq x \leqq 1, 0 < \alpha < 1),$$

$$F_1(x) = \begin{cases} 0 & (x < 1), \\ 1 & (x \geqq 1). \end{cases}$$

It is therefore clear that as $n \to \infty$, $n^{-1} I_n([na])$ will converge in law to a random variable with the same distribution as $aY_1+(1-a)(1-Y_2)$, where Y_1 and Y_2 are independent and each has distribution function F_α. It remains only to examine the particular cases.

If $0 < a < 1$, $0 < \alpha < 1$, aY_1 has density $\pi^{-1} \sin \pi a\, x^{\alpha-1}(a-x)^{-\alpha}$, $0 \leqq x \leqq a$, while $(1-a)(1-Y_2)$ has density $\pi^{-1} \sin \pi a\, x^{-\alpha}(1-a-x)^{\alpha-1}$, $0 \leqq x \leqq 1-a$. The density of the random variable $aY_1+(1-a)(1-Y_2)$ is therefore

$$\left(\frac{\sin \pi\alpha}{\pi}\right)^2 \int_{y=\max(0,x+a-1)}^{\min(x,a)} \frac{dy}{(x-y)^\alpha (1-a-x+y)^{1-\alpha} y^{1-\alpha}(a-y)^\alpha} \qquad (0 \leqq x \leqq 1),$$

as required.

The other cases can be read off immediately using relations (5). If $0 \leqq a \leqq 1$, $\alpha = 0$, then $n^{-1} I_n([na])$ converges in law to $1-a$ as $n \to \infty$, and hence converges in probability to $1-a$. Similarly, if $0 \leqq a \leqq 1$, $\alpha = 1$, $n^{-1} I_n([na])$ converges in probability to a as $n \to \infty$. On the other hand, we see that $\lim_{n\to\infty} \Pr\{n^{-1} I_n([na]) \leqq x\}$ is $1-F_\alpha(1-x)$, if $a = 0$ or $F_\alpha(x)$ if $a = 1$.

Finally, if the condition (1) is not satisfied then the relation

$$\frac{\Pr(S_1 > 0) + \cdots + \Pr(S_n > 0)}{n} = E\,\frac{N_n}{n}$$

shows us that $n^{-1} N_n$ cannot converge in distribution and so neither can $n^{-1} N'_{[na]} + 1 - n^{-1}[na] - n^{-1} N''_{n-[na]}$ or, in other words, $n^{-1} I_n([na])$. This completes the proof of the theorem.

Theorem 2 of Darling [1] is the particular case of our Theorem 2 where the X_i are restricted to have a continuous and symmetric distribution. The case $0 < \alpha < 1$ of our theorem could have been established along parallel lines to the proof of Theorem 2 of [1] by making use of Theorem 2 of Heyde [2] in which it is shown that there must exist a function of slow variation L such that

$$k^{1-\alpha}(n-k)^\alpha \frac{L(n-k)}{L(k)} \Pr(N_n = k) = \frac{\sin \pi\alpha}{\pi} + o(k, n)$$

where $o(k, n)$ tends to zero uniformly in k and n as $\min(k, n-k) \to \infty$.

In order to read off the results of Spitzer's generalization [3], Theorem 7.1, of the arc-sine law from our Theorem 2 we note that

$$\Pr(N_n = k) = \Pr\{I_n(n-k) = 0\} = \Pr\{I_n(0) = n-k\},$$

so that

$$\begin{aligned} \Pr(N_n \leqq nx) &= \Pr(N_n \leqq [nx]) \\ &= \Pr(n-[nx] \leqq I_n(0) \leqq n) \\ &\to 1 - G_{0,\alpha}(1-x) = F_\alpha(x), \text{ as } n \to \infty. \end{aligned}$$

References

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

[2] C. C. Heyde, 'Some local limit results in fluctuation theory', *J. Austral. Math. Soc.* 7 (1967), 455–464.

[3] F. L. Spitzer, 'A combinatorial lemma and its application to probability theory', *Trans. Amer. Math. Soc.* 82 (1956), 232–339.

University of Sheffield