Z. Wahrscheinlichkeitstheorie verw. Geb. 9, 10—15 (1967)

# On Ladder Indices and Random Walk

J. P. Imhof

Received October 18, 1966

## 1. Introduction

Let $X_1, \ldots, X_n$ be exchangeable random variables with continuous, symmetric distribution function. Thus $P[X_i < x_i, i = 1, \ldots, n] = P[e_i X_{\sigma_i} < x_i, i = 1, \ldots, n]$, for any of the $2^n$ sign vectors $e = (e_1, \ldots, e_n)$, $e_i = \pm 1$, and any of the $n!$ permutations $\sigma = (\sigma_1, \ldots, \sigma_n)$ of $(1, \ldots, n)$. The sequence $S_1, \ldots, S_n$ $(S_r = \sum_1^r X_i)$ has the ladder index $I$ if $S_r < S_I$ for $0 \leqq r < I$ $(S_0 = 0)$. Ties among partial sums $S_0, S_1, \ldots, S_n$ occur with probability zero and are disregarded. Let $I_1$ be the smallest ladder index, $I_1, I_2, \ldots, I_k, \ldots$ be the successive ladder indices. We point out that there is complete analogy between ladder indices of $S_1, \ldots, S_n$ and equilibrium values in $2n$ steps of simple symmetric random walk (hereafter called simply "random walk"). A path counting procedure yields for ladder indices a derivation of some results known for random walk. The probability that there is a $k$-th ladder index and that exactly $j$ sums are larger than the $k$-th ladder sum $S_{I_k}$ is found to be

$$(1) \qquad p(n, j, k) = 2^{-2n+k} \binom{2n - 2j - k}{n - j} \binom{2j}{j}, \quad j, k \geqq 0, \quad j + k \leqq n.$$

$(S_0 = 0$ plays the role of zero-th ladder sum, but does not count as one). This provides a transition between the arc sine law of the number of positive sums $(k = 0)$, and the law of the number of ladder indices $(j = 0)$. It is shown that certain results which have no simple explanation in the case of random walk become quite natural in the context of ladder indices.

## 2. Two Equivalent Problems

If an event $E$ relative to $X_1, \ldots, X_n$ has an invariant probability (not depending on the particular continuous distribution function of the symmetric exchangeable random variables $X_1, \ldots, X_n$), a convenient possible way to determine the probability $P[E]$ is to use a path counting procedure. To a set of positive numbers $x_1 < x_2 < \cdots < x_n$ such that

$$(2) \qquad \sum_{i=1}^{n} a_i x_i \neq 0 \quad \text{for all choices } a_i = -1, 0, 1, \text{ with at least one } a_i \neq 0,$$

associate the $2^n n!$ polygonal paths $(x; e, \sigma)$ which join the origin $(0, 0)$ in a Cartesian plane with the successive points $(1, s_1), (2, s_2), \ldots, (n, s_n)$, where $s_r = \sum_{i=1}^{r} e_i x_{\sigma_i}$. $P[E]$ equals the proportion of the paths for which $E$ holds. Conversely, showing that for an event $E$ this proportion does not depend (provided (2) holds) on $x = (x_1, \ldots, x_n)$ establishes that $E$ has an invariant probability. The method has been used for instance by HOBBY and PYKE [5] to derive with relative ease Baxter's generalized arc sine law. A formal justification for this procedure has been given by FRIEDMAN, KATZ and KOOPMANS [4].

Let $E(n, j, k)$ be the event: there are at least $k$ ladder indices for the sequence $S_1, \ldots, S_n$, and $j$ partial sums are larger than the $k$-th ladder sum $S_{I_k}$ ($k = 0, 1, \ldots, n$; $j = 0, 1, \ldots, n-k$). We shall check that it has an invariant probability. The fact can be established directly, with a geometric argument similar to the one used in [5]. While easy to visualize, the map obtained by the shrinking-and-switching procedure is unfortunately clumsy to describe. We use therefore a different argument, based on the following:

**Lemma.** *Let $F_n$ be the event: "$n$ is first ladder index of $S_1, \ldots, S_n$". It has an invariant probability, equal to the probability that in random walk, the first equilibrium occurs at step $2n$.*

*Proof.* The following relation between events is obvious:

$$F_n = [S_i < 0,\ i = 1, \ldots, n-1] - [S_i < 0,\ i = 1, \ldots, n].$$

Set $c_r = 2^{-2r}\binom{2r}{r}$, $r = 0, 1, 2, \ldots$. According to the finite arc sine law,

$$P[S_i < 0,\ i = 1, \ldots, r] = c_r.$$

Therefore,

$$P[F_n] = c_{n-1} - c_n.$$

Comparison with [2, chap. III.4] gives the conclusion.

The following notation will be used, with regard to the $2^n n!$ paths $(x; e, \sigma)$ generated by a set $x_1 < \cdots < x_n$ for which (2) holds:

$t(n, j, k)$ = number of paths of type $(n, j, k)$, i.e. for which $E(n, j, k)$ holds.

$u(n, k)$ = number of paths for which $s_n$ is $k$-th ladder sum, i.e. for wichh $i_k = n$, where $i_k$ designates the path's $k$-th ladder index.

$T(n, j) = t(n, j, 0) = 2^{-n} n! \binom{2n-2j}{n-j}\binom{2j}{j}$, the arc sine frequency.

$U(n, k) = t(n, 0, k)$ = number of paths having exactly $k$ ladder sums.

Dividing each of the above frequencies by $2^n n!$ yields the corresponding probabilities, which we shall denote respectively by:

$$p(n, j, k), \quad q(n, k), \quad P(n, j), \quad Q(n, k).$$

**Theorem.** *$E(n, j, k)$ has an invariant probability, equal* (*in the language of* [2, chap. III]) *to the probability that in $2n$ steps of random walk, the particle returns at least $k$ times to zero and spends $2j$ steps on the positive side after the $k$-th return to zero.*

*Proof.* One can determine $t(n, j, k)$ by reviewing possible lengths $d_h$ ($0 < d_h = i_h - i_{h-1}, h = 1, \ldots, k$) of portions of paths between consecutive ladder indices $i_1, \ldots, i_k$ ($i_0 = 0$), and by requesting that the portion of path of length $n - r$ starting at $s_{i_k}$ have exactly $j$ positive sums ($k \leqq r = d_1 + \cdots + d_k \leqq n - j$).

The components $x_1, \ldots, x_n$ can be assigned to the successive sections of lengths $d_1, \ldots, d_k$ and $n - r$ in $n!\{d_1! \ldots d_k!(n-r)!\}^{-1}$ ways, therefore

$$t(n, j, k) = n! \sum_{r=k}^{n-j} \frac{T(n-r, j)}{(n-r)!} \sum_{d_1 + \cdots + d_k = r} \prod_{h=1}^{k} \frac{u(d_h, 1)}{d_h!}.$$

The known invariance of the arc sine frequencies $T(n, j)$, together with the invariance of the frequencies $u(d, 1)$ which results from the lemma, establishes that $t(n, j, k)$ does not depend on $x_1, \ldots, x_n$, and hence that $E(n, j, k)$ has an invariant probability. The latter is found, dividing by $2^n n!$:

$$p(n, j, k) = \sum_{r=k}^{n-j} P(n-r, j) \sum_{d_1 + \cdots + d_k = r} \prod_{h=1}^{k} q(d_h, 1).$$

According to the lemma, $q(d_h, 1)$ is the probability of first return to zero after $2d_h$ steps of random walk, while $P(n-r, j)$ is known to be the probability that $2j$ out of $2n - 2r$ steps, counted from an equilibrium position, are spent on the positive side. The conclusion follows.

## 3. Derivation of the Probabilities

A recursion formula can be obtained for $t(n, j, k)$ by following the procedure used in [5]. Consider the $n - 1$ components $x_2 < x_3 < \cdots < x_n$, for which (2) holds for sums $\sum\limits_2^n$. Let

$$d = \min \left\{ \left| \sum_2^n a_i x_i \right| : a_i = -1, 0, 1;\ i = 2, \ldots, n; a_i \neq 0 \text{ for at least one } i \right\}.$$

Choose $0 < x_1 < d$. For convenience, set

$$t(n, -1, k) = t(n, j, -1) = t(n-1, n, k) = t(n-1, j, n) = 0.$$

For $j, k \geqq 0$, $j + k \leqq n$, the paths of type $(n, j, k)$ can be obtained from paths of length $n - 1$ constructed with $x_2, \ldots, x_n$, by insertion of $\pm x_1$ in the following ways:

a) In any path of type $(n-1, j-1, k)$, insert $+x_1$ just behind the $k$-th ladder sum $s_{i_k}$, or $\pm x_1$ just behind one of the $j - 1$ sums larger than $s_{i_k}$. This yields $(2j-1)\,t(n-1, j-1, k)$ paths.

b) In any path of type $(n-1, j, k-1)$, insert $+x_1$ just behind one of the ladder sums $s_0 = 0, s_{i_1}, \ldots, s_{i_{k-1}}$. This yields $k\,t(n-1, j, k-1)$ paths.

c) In any path of type $(n-1, j, k)$, insert $-x_1$ just behind one of $s_0, s_{i_1}, \ldots, s_{i_k}$, or $\pm x_1$ just behind one of the $n - 1 - j - k$ sums which are neither ladder sums, nor larger than $s_{i_k}$. This provides the last term in:

$$\begin{aligned} t(n, j, k) = {} & (2j-1)\,t(n-1, j-1, k) + k\,t(n-1, j, k-1) + \\ & + (2n - 2j - k - 1)\,t(n-1, j, k), \quad j, k \geqq 0,\ j + k \leqq n. \end{aligned} \tag{3}$$

For $k = 0$ and the boundary values $T(1, 0) = T(1, 1) = 1$, the solution is given by the arc sine frequencies $T(n, j)$. For $j = 0$, one knows from random walk theory [3] that the solution is necessarily given by

$$U(n,k) = 2^{-n+k} n! \binom{2n-k}{n}, \quad k = 0, 1, \ldots, n,$$

corresponding to the probabilities

$$Q(n,k) = 2^{-2n+k} \binom{2n-k}{n}. \tag{4}$$

This is easy to obtain also by induction, from the values $U(1,0) = U(1,1) = 1$, and $U(n,0) = (2n-1)!! = 2^{-n} n! \binom{2n}{n}$. As is known [2], the most likely numbers of ladder indices are, for all $n > 0$, 0 and 1. More precisely,

$$Q(n,0) = Q(n,1) > Q(n,2) > \cdots > Q(n,n) = 2^{-n}.$$

Rather than solve (3) directly, it is easier to proceed via $u(n,k)$. This again is known from random walk [3], but a derivation based on counting paths is instructive. One has, for $k = 1, \ldots, n$:

$$2n\, U(n-1, k-1) - U(n, k-1) = u(n,k) - u(n,k-1). \tag{5}$$

In fact, consider for each choice of $n-1$ of the numbers $x_1, \ldots, x_n$, the $U(n-1, k-1)$ paths having $k-1$ ladder sums. Complete them to length $n$ by adding $\pm$ the missing $x_i$ in last position. This yields $2n\, U(n-1, k-1)$ paths, namely the $u(n,k)$ paths having $s_n$ as $k$-th ladder sum, plus all those having $k-1$ ladder sums, $s_n$ not being one of them. They number $U(n,k-1) - u(n,k-1)$, hence (5).

Let us write (3) for $j = 0$. This gives, with $t(n,0,k) = U(n,k)$ and with $k$ replaced by $k-1$:

$$U(n,k-1) = (k-1)\, U(n-1,k-2) + (2n-k)\, U(n-1,k-1).$$

Comparison with (5) now yields, for $k = 1, \ldots, n$,

$$u(n,k) - u(n,k-1) = k\, U(n-1,k-1) - (k-1)\, U(n-1,k-2).$$

Summation from 1 to $k$ leaves simply,

$$u(n,k) = k\, U(n-1,k-1). \tag{6}$$

The probability that $n$ is $k$-th ladder index is therefore

$$q(n,k) = 2^{-2n+k} \frac{k}{2n-k} \binom{2n-k}{n}, \quad k = 1, \ldots, n.$$

Notice the index of the maximum partial sum is $n$ if and only if $n$ is a ladder index, so that one has:

$$P(n,0) = \sum_{k=1}^{n} q(n,k) = \frac{1}{2n} \sum_{k=1}^{n} k\, Q(n-1,k-1).$$

From this, an easy calculation gives the expected number of ladder sums among $S_1, \ldots, S_n$ to be $(2n+2)\, P(n+1, 0) - 1$. Furthermore, the known relation $\sum_{1}^{n} k^{-1} q(n,k) = (2n)^{-1}$ reduces to $\sum_{0}^{n-1} Q(n-1,k) = 1$.

For a path to be of type $(n, j, k)$, the $k$-th ladder index must take one of the values $r = k, \ldots, n - j$. The initial portion of $r$ segments must correspond to "$k$-th ladder index is $r$", the terminal portion of $n - r$ segments to "$j$ positive partial sums". Taking into account the $\binom{n}{r}$ choices of $r$ initial $x_i$'s, this gives

$$t(n, j, k) = \sum_{r=k}^{n-j} \binom{n}{r} u(r, k)\, T(n - r, j),$$

which becomes in terms of probabilities:

$$p(n, j, k) = \sum_{r=k}^{n-j} q(r, k)\, P(n - r, j).$$

Noticing that $P(n - r, j) = \binom{2j}{j} P(n - r, 0)$, one obtains

$$p(n, j, k) = \binom{2j}{j} \sum_{r=k}^{n-j} q(r, k)\, P(n - j - r, 0)$$
$$= \binom{2j}{j} p(n - j, 0, k) = \binom{2j}{j} Q(n - j, k).$$

This establishes (1).

Let us also point out that there are two ways in which to evaluate the probability of obtaining at least $k$ ladder indices: they correspond to the two members in the equality

$$\sum_{j=0}^{n-k} p(n, j, k) = \sum_{j=0}^{n-k} Q(n, k + j).$$

## 4. Some Remarks

If $L_n$ is the index of the maximum partial sum and if we call ladder* indices the successive indices $L_n = I_1^* < I_2^* < \cdots < n$ for which $S_{I^*j} > S_i$, $i = I_j^* + 1, \ldots, n$, the probability of having $k$ ladder indices and $m$ ladder* indices $(0 < k + m \leqq n)$ is invariant. In fact, if $P^*(n, k, m)$ is said probability, the map

$$(x_1, \ldots, x_{L_n}, x_{L_{n+1}}, \ldots, x_n) \to (x_1, \ldots, x_{L_n}, -x_n, \ldots, -x_{L_n+1})$$

shows that

$$P^*(n, k, m) = q(n, k + m).$$

The pair $(k, m)$ takes one of the possible values corresponding to $k, m = 0, 1, \ldots, n$, $0 < k + m \leqq n$, thus

$$\sum_{r=1}^{n} (r + 1)\, q(n, r) = 1.$$

On the other hand, if down-ladder indices of $S_1, \ldots, S_n$ are defined to be ladder indices of $-S_1, \ldots, -S_n$, the joint distribution of ladder and down-ladder indices is not invariant.

It is obvious, for reasons of symmetry, that $P$[$n$ is first ladder index] $= P$[there is no ladder index]. In the language of random walk, this yields the wellknown equality $P$[first return to zero is at step $2n$] $= P$[no return to zero in the first $2n$ steps].

Consider now the $2^n n!$ paths $(x; e, \sigma)$ determined by a vector $x$ for which (2) holds. To each particular path, corresponding to $e_1 x_{\sigma_1}, \ldots, e_n x_{\sigma_n}$, make correspond its reverse, determined by $-e_n x_{\sigma_n}, \ldots, -e_1 x_{\sigma_1}$. For $0 < m \leqq n$, let $C_{m,n}$ be the set of all paths for which one at least of the indices $m, m+1, \ldots, n$ is a ladder index. One realizes at once that a path does not belong to $C_{m,n}$, if and only if its reverse belongs to $C_{n-m+1,n}$. In other words, if $A_{m,n}$ is the event: "one at least of the indices $m, m+1, \ldots, n$ is a ladder index for the sequence $S_1, \ldots, S_n$",

$$P[A_{m,n}] + P[A_{n-m+1,n}] = 1 .$$

In the language of random walk, this gives (when $n$ is replaced by $m+n-1$) the formula established by BLACKWELL, DEUEL and FREEDMAN [1].

I thank an Editor for suggesting improvements of the manuscript.

## References

1. BLACKWELL, D., P. DEUEL, and D. FREEDMAN: The last return to equilibrium in a coin-tossing game. Ann. math. Statistics **35**, 1344 (1964).
2. FELLER, W.: An introduction to probability theory and its applications, Vol. 1, 2nd ed. New York: Wiley 1957.
3. — The number of zeros and changes of sign in a symmetric random walk. Enseignement math., II. Sér. **3**, 229–235 (1957).
4. FRIEDMAN, N., M. KATZ, and L. H. KOOPMANS: On tests of symmetry for continuous distributions. Technical Report No. 119. Department of Math., Univ. of New Mexico 1966.
5. HOBBY, CH., and R. PYKE: Combinatorial results in fluctuation theory. Ann. math. Statistics **34**, 1233–1242 (1963).

Institut de Mathématiques
16, Blvd. d'Yvoy
CH 1211 Genève 4