THE JOINT DISTRIBUTION OF OCCUPATION TOTALS FOR A SIMPLE RANDOM WALK

V. D. BARNETT

(received 14 March 1964, revised 20 May 1964)

1. Summary

A great deal of attention has been given in the literature to the various properties of the simple binomial random walk. Explicit expressions are available for first passage times, absorption probabilities, average duration of the walk up to absorption and other quantities of interest. One aspect of the behaviour of this work which has, however, attracted little attention is the form of the distribution of occupation totals. This paper is devoted to the derivation of an explicit expression for the joint probability generating function of the occupation totals up to absorption, for the binomial random walk in the presence of two absorbing points. The appropriate marginal form of this p.g.f. yields the distribution of the occupation total, and expected occupation total, at any particular lattice point. The limiting forms of these results provide explicit expressions for the corresponding quantities in the case of a binomial random walk having a single absorbing point and, where relevant, in the case of the unrestricted binomial random walk.

A convenient check on these results is provided by the immediate recovery from them of the familiar forms of the absorption probabilities and average duration of the walk.

2. Introduction

Consider a particle which executes a simple random walk on the integers $x = 0, 1, \dots, a$, starting from the point k and with the points 0 and a being absorbing. The walk proceeds according to the following rule. From the current position x, 0 < x < a, the particle moves to x+1 with probability p or to x-1 with probability q = 1-p, and continues in this way until absorption takes place either at 0 or a. We shall not be interested in time-dependent properties of the walk so it is unnecessary to specify the distribution of the time interval between successive steps of the walk. It is well known that, for this walk, ultimate absorption is certain and that the average duration of the walk up to absorption is finite. Consequently it is relevant

to consider the joint distribution of the number of times the particle lands at each of the lattice points up to absorption. Also of interest is the marginal distribution, and its mean, of the number of landings at any particular lattice point, x.

To obtain expressions for these various quantities it is convenient to reinterpret the binomial random walk as a finite Markov chain with state space $\{0, 1, \dots, a\}$, initial state k, and transition matrix

$$(1) Q = \{q_{i,j}\}$$

where

(2)
$$\begin{cases} q_{0,0} = q_{a,a} = 1, \\ q_{i,i+1} = p; \ i = 1, \cdots, a-1, \\ q_{i,i-1} = q; \ i = 1, \cdots, a-1, \\ q_{i,j} = 0; \ j \neq i-1 \text{ or } i+1. \end{cases}$$

Thus we want to obtain an expression for the joint distribution of the numbers of times each of the states is occupied, i.e. for the occupation totals up to the first time the chain is in state 0 or *a*. The marginal forms of this joint distribution then yield immediately the distribution of the occupation total for any particular state, and, *a fortiori*, the expected occupation total for that state, which is seen to have a special relevance in connection with the derivation of the absorption probabilities and the expected duration of the random walk we are considering.

3. The joint p.g.f. of the occupation totals for the walk with two absorbing points

The joint distribution of the occupation totals is most conveniently obtained in the form of its probability generating function. In a paper on "the frequency count of a Markov chain", Good [1] considers a finite Markov chain with general transition matrix Q and initial state distribution $\{q_i\}$; $i = 0, 1, \dots, a$. He shows that the joint p.g.f. of the occupation totals, m_x , is given by

(3)
$$\sum_{\mathbf{m}' \in \geq 0} p(\mathbf{m}) \mathbf{z}^{\mathbf{m}} = (q_0 z_0, \cdots, q_a z_a) [I - QZ]^{-1} \mathbf{e},$$

where

$$Z = \begin{pmatrix} z_0 & & 0 \\ & z_1 & \\ & & \ddots & \\ 0 & & & z_a \end{pmatrix}, \quad e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

This result is easily modified to obtain the joint p.g.f. of the occupation

totals up to absorption when the states 0 and a are absorbing i.e. when $q_{0,x} = \delta_{0,x}$; $q_{a,x} = \delta_{a,x}$. In this case we obtain

(4)
$$\sum_{\mathbf{m}' \in \geq 0} p(\mathbf{m}) \mathbf{z}^{\mathbf{m}} = (q_1 z_1, \cdots, q_{a-1} z_{a-1}) [I - Q^* Z^*]^{-1} f_{\mathbf{n}}$$

where Q^* and Z^* are the truncated forms of the transition matrix Q, and matrix of parameters, Z, obtained in each case by omitting the first and last row and column, and f has the form

(5)
$$f = \begin{pmatrix} q_{1,0} + q_{1,a} \\ \vdots \\ q_{a-1,0} + q_{a-1,a} \end{pmatrix}$$

The binomial random walk is simply a particular example of such a chain where Q has the form (2) and

(6)
$$f' = (q, 0, \dots, 0, p).$$

So that, if the starting point is known to be k, the joint p.g.f., $\psi_k(z)$, of the occupation totals up to absorption is given by (4), with $q_i = 0$; $i \neq k$, and where Q and I are given by (2) and (6), respectively. Considering (4) for varying starting point, k, we obtain a system of a-1 equations

(7)
$$\psi(z) = Z^*[I - Q^*Z^*]^{-1}f,$$

so that $\psi_k(z)$ may be obtained from the recurrence relationship

(8)
$$[(Z^*)^{-1}-Q^*]\psi(z) = f.$$

More specifically we must solve

(9)
$$\psi_k(z) = z_a[q\psi_{k-1}(z) + p\psi_{k+1}(z)]; k = 1, 2, \cdots, a-1,$$

subject to the boundary conditions

(10)
$$\psi_0(z) = \psi_a(z) \equiv 1.$$

Solving (9) and (10) is equivalent to inverting $[I-Q^*Z^*]$ in (4) but the more general approach proves more convenient in this case, since we may obtain the solution by an approach similar to the extrapolation method of Kemperman [2], p. 4.

Thus starting with

(11)
$$\psi_1(z) = z_1[q + p\psi_2(z)],$$

the recurrence relationship may be reduced from second order to first order by successively eliminating one of the $\psi_k(z)$ in each equation of the system (9), to obtain finally

[4] The joint distribution of occupation totals for a simple random walk

521

(12)
$$\psi_a(\mathbf{z}) = \frac{1}{p_{z_{a-1}}} \left\{ \left(\frac{1 - H_{a-1}(\mathbf{z})}{1 - H_{a-2}(\mathbf{z})} \right) \psi_{a-1}(\mathbf{z}) - \frac{q^{a-1} z_1 \cdots z_{a-1}}{1 - H_{a-2}(\mathbf{z})} \right\},$$

where $H_x(z)$; $x = 0, 1, \dots, a-1$, is a function of z_1, \dots, z_x satisfying

(13)
$$\begin{cases} H_x(z) = H_{x-1}(z) + pq z_{x-1} z_x [1 - H_{x-2}(z)], \\ H_0 = H_1(z_1) \equiv 0. \end{cases}$$

But the boundary conditions (10) enable $\psi_{a-1}(z)$ to be expressed, by (12), explicitly, in terms of $H_x(z)$, as

(14)
$$\psi_{a-1}(z) = p z_{a-1} \left[\frac{1 - H_{a-2}(z)}{1 - H_{a-1}(z)} \right] + \frac{q^{a-1} z_1 \cdots z_{a-1}}{1 - H_{a-1}(z)}.$$

Reversing the recursive process which lead to (12) we obtain expressions for all the $\psi_k(z)$ in the form

$$\psi_{k}(\mathbf{z}) = p^{a-k} z_{a-1} \cdots z_{k} \left[\frac{1 - H_{k-1}(\mathbf{z})}{1 - H_{a-1}(\mathbf{z})} \right] + q^{k} z_{1} \cdots z_{k} \left[\frac{1 - H'_{a-k-1}(\mathbf{z})}{1 - H'_{a-1}(\mathbf{z})} \right];$$
(15)
$$k = 1, \cdots, a-1,$$

where $H'_x(z)$ is similar to $H_x(z)$ but with the order of the z_i reversed, i.e. $H'_x(z)$ is a function of z_{a-1}, \dots, z_{a-x} satisfying

(16)
$$\begin{cases} H'_{x}(z) = H'_{x-1}(z) + pqz_{a+1-x}z_{a-x}[1-H'_{x-2}(z)], \\ H'_{0} = H'_{1}(z_{a-1}) \equiv 0. \end{cases}$$

Thus the joint p.g.f. of the occupation totals up to absorption is given by (15) as a function of the $H_x(z)$ and $H'_x(z)$, and it would seem at first glance that little has been achieved other than to exchange one second order recurrence relationship, (9), for another, (13) (and (16)). However it is a simple matter to evaluate the $H_x(z)$ (and $H'_x(z)$) from (13) (and (16)) successively, starting from the initial values $H_0 = H_1(z_1) \equiv 0$ (and $H'_0 = H'_1(z_{a-1}) \equiv 0$). In this way we get the general result

(17)
$$H_{z}(z_{1}, \cdots, z_{z}) = pq \sum_{i=1}^{z-1} z_{i} z_{i+1} \cdots (pq)^{2} \sum_{j>i+1}^{z-1} \sum_{i=1}^{z-3} z_{i} z_{i+1} z_{j} z_{j+1} \cdots \sum_{i=1}^{z-1} (-pq)^{[x]/2} \sum_{i>k+1}^{z-1} \cdots \sum_{j>i+1}^{z-[x]-3} \sum_{i=1}^{z-[x]-1} z_{i} z_{i+1} z_{j} z_{j+1} \cdots z_{i} z_{i+1} \Big],$$

where

(18)
$$\begin{cases} [x] = x , \text{ if } x \text{ even} \\ = x - 1, \text{ if } x \text{ odd.} \end{cases}$$

An exactly similar expression, with the order of the z_i reversed, is obtained for $H'_x(z)$.

V. D. Barnett

[5]

We might observe that, whilst the general expression (15) for $\psi_k(z)$ appears rather complicated in form, it is, nevertheless, consistent with intuitive ideas about the form of the joint p.g.f. of the occupation totals up to absorption. As indicated by Kemperman [2], p. 48, it is inevitable that the joint distribution of the occupation totals up to absorption must be of the form

(19)
$$p(m) = A_m p^{a-k} (pq)^{(N+k-a)/2} + B_m q^k (pq)^{(N-k)/2}$$

for any walk which leads to absorption on the Nth step; where $(pq)^x$ must be interpreted as zero if x is not integral. The quantities A_m and B_m are the numbers of walks which lead, respectively, to absorption at a on the Nth step and to absorption at 0 on the Nth step.

But (15) generates probabilities p(m) which are of precisely this required form (19), since from (17) it is seen that $[1-H_{k-1}(z)]/[1-H_{a-1}(z)]$ (and $[1-H'_{a-k-1}(z)]/[1-H_{a-1}(z)]$) are power series in pq where for any joint frequency m, such that m'e = N, the relevant power of pq will be $\frac{1}{2}(N-a+k)$ (or $\frac{1}{2}(N-k)$) when N-a+k (or N-k) is even, and zero otherwise.

4. The distribution of the occupation total at a single point, and its mean

Of rather more interest that the joint p.g.f. of the occupation totals over all the points of the walk, is the distribution of the occupation total at a particular point. The p.g.f. of the occupation total at a point, x, may be obtained from the joint p.g.f. (15) simply by putting $z_i = 1$ for $i \neq x$ and $z_x = z$ to give

$$(20) \begin{cases} \psi_{k,x}(z) = p^{a-k} \left\{ \frac{1 - H_{k-1}^{(x)}(z)}{1 - H_{a-1}^{(x)}(z)} \right\} + q^{k} z \left\{ \frac{1 - H_{a-k-1}^{'(x)}(z)}{1 - H_{a-1}^{'(x)}(z)} \right\}, x < k, \\ = z \left[p^{a-k} \left\{ \frac{1 - H_{k-1}^{(k)}(z)}{1 - H_{k-1}^{(k)}(z)} \right\} + q^{k} \left\{ \frac{1 - H_{a-k-1}^{'(k)}(z)}{1 - H_{a-1}^{(k)}(z)} \right\} \right], x = k, \\ = p^{a-k} z \left\{ \frac{1 - H_{k-1}^{(x)}(z)}{1 - H_{a-1}^{'(x)}(z)} \right\} + q^{k} \left\{ \frac{1 - H_{a-k-1}^{'(x)}(z)}{1 - H_{a-k-1}^{'(x)}(z)} \right\}, x > k, \end{cases}$$

where the $H_{u}^{(x)}(z)$ now satisfy the appropriate form of (13), viz.

(21)
$$\begin{cases} H_0 = H_1 \equiv 0, \\ H_{y}^{(x)} = H_{y-1}^{(x)} + pq[1 - H_{y-2}^{(x)}], y < x, \\ H_{x}^{(x)}(z) = H_{x-1}^{(x)} + pqz[1 - H_{x-2}^{(x)}], \\ H_{x+1}^{(x)}(z) = H_{x}^{(x)}(z) + pqz[1 - H_{x-1}^{(x)}], \\ H_{y}^{(x)}(z) = H_{y-1}^{(x)}(z) + pq[1 - H_{y-2}^{(x)}(z)], y > x + 1. \end{cases}$$

and the $H'_{\mathbf{y}}^{(x)}(z)$ satisfy the corresponding appropriate form of (16).

It is easy to confirm that the solution of (21) is

(22)
$$\begin{cases} H_{y}^{(x)}(z) = H_{y}, & y < x, \\ = 1 - z(1 - H_{y}) - (1 - z)(1 - H_{x-1})(1 - H_{y-x}), & y \ge x, \end{cases}$$

where H_y , independent of z, is the solution of the recurrence relationship

(23)
$$\begin{cases} H_{y} = H_{y-1} + pq(1 - H_{y-2}), \\ H_{0} = H_{1} \equiv 0. \end{cases}$$

Similarly

(24)
$$\begin{cases} H'_{\boldsymbol{y}}^{(\boldsymbol{x})}(z) = H_{\boldsymbol{y}}, \quad \boldsymbol{y} < \boldsymbol{a} - \boldsymbol{x}, \\ = 1 - z(1 - H_{\boldsymbol{y}}) - (1 - z)(1 - H_{\boldsymbol{a} - \boldsymbol{x} - 1})(1 - H_{\boldsymbol{y} + \boldsymbol{x} - \boldsymbol{a}}), \quad \boldsymbol{y} \geq \boldsymbol{a} - \boldsymbol{x}. \end{cases}$$

But the explicit form of H_y is immediately obtained, by the standard method of solution of the second order linear recurrence relationship, as

(25)
$$\begin{cases} H_{y} = 1 - \frac{p^{y+1} - q^{y+1}}{p-q}, & p \neq q, \\ = 1 - \frac{y+1}{2^{y}}, & p = q, \end{cases}$$

thus yielding an explicit expression for $\psi_{k,x}(z)$ from (20). We have 1) $\phi \neq q$,

(26)
$$\begin{cases} \psi_{k,x}(z) = \frac{A + (B - A)z}{C + (B - C)z}, & x < k, \\ = \frac{Bz}{C + (B - C)z}, & x = k, \\ = \frac{D + (B - D)z}{C + (B - C)z}, & x > k, \end{cases}$$

where

(27)
$$\begin{cases} A = p^{a-k}(p^{x}-q^{x})(p^{k-x}-q^{k-x}), \\ B = (p^{a}-q^{a})(p-q), \\ C = (p^{x}-q^{x})(p^{a-x}-q^{a-x}), \\ D = q^{k}(p^{x-k}-q^{x-k})(p^{a-x}-q^{a-x}), \end{cases}$$

or

It may be observed from (26) and (28) that the distribution of the occupation total at the starting point, k, is geometric; and modified geometric at any other lattice point $x \neq k$. Proof of the fact that the distribution is geometric or modified geometric does not require the detailed calculations above. In fact this follows immediately by simple probabilistic arguments involving first passage probabilities from k to x and x to x. The explicit form of the distribution of the occupation total at x is seen, by this approach, to depend on the two quantities $\mu_{k,x}$ and $\mu_{x,x}$ (or $\mu_{k,k}$ alone, if x = k), where $\mu_{k,x}$ is the expected occupation total at x for a walk starting from k. However, evaluation of $\mu_{k,x}$ and $\mu_{x,x}$ to specify the explicit form of the distribution again involves calculations substantially the same as those above, since we must again solve a second order recurrence relationship similar to (9) to obtain $\mu_{k,x}$ and $\mu_{x,x}$.

The expected occupation totals may be obtained by inspection from (26), (27) and (28). Thus we see that

(29)
$$\begin{cases} \mu_{k\,x} = \frac{(p^{x} - q^{x})(p^{a-k} - q^{a-k})q^{k-x}}{(p^{a} - q^{a})(p - q)}, & x \leq k, \\ = \frac{(p^{k} - q^{k})(p^{a-x} - q^{a-x})p^{x-k}}{(p^{a} - q^{a})(p - q)}, & x \geq k, \end{cases}$$

or

2) $\phi = q$.

1) $\phi \neq a$.

(30)
$$\begin{cases} \mu_{k,x} = 2x(1-k/a), & x \leq k, \\ = 2k(1-x/a), & x \geq k. \end{cases}$$

An interesting check of these results is provided be deriving from (29) and (30) the absorption probabilities, $q_k(0)$ and $q_k(a)$, and the expected duration of the walk D_k , the forms of which are well known. Obviously $q_k(0) = q\mu_{k,1}$; $q_k(a) = p\mu_{k,a-1}$ and $D_k = \sum_{x=1}^{a-1} q_{k,x}$, and substituting (29) or (30) for $\mu_{k,x}$ we obtain expressions which agree with those given, for example, by Feller [3], p. 313 et seq.

524

A reference in the literature which is relevant to this problem is included in the work of Kemeny and Snell [4] on "finite Markov chains". In particular they discuss (Chapter 3) the significance of the truncated transition matrix Q^* in the study of absorbing Markov chains, and derive the explicit form of $\mu_{k,x}$ for the walk discussed in this paper, in the particular case a = 4. For this single situation their result agrees with the general expression, (29), for $\mu_{k,x}$ given above.

5. The walk having a single absorbing point at the origin

The analogous results for a walk with a single absorbing point at the origin, may be immediately obtained as the limiting form of those given in the previous section. In particular, letting $a \to \infty$ in (26)-(30), we obtain the appropriate expressions for $\psi_{k,x}(z)$ and $\mu_{k,x}$ in this case.

Thus if $p \neq q$, $\psi_{k,x}(z)$ is again given by (26), but the form of A, B, C and D now depends on the relative magnitudes of p and q, so that

a)
$$p > q$$
,

(31)
$$\begin{pmatrix} A = [1 - (q/p)^{k-x}][1 - (q/p)^{x}] \\ B = p - q, \\ C = [1 - (q/p)^{x}], \\ D = (q/p)^{x}[(q/p)^{k-x} - 1], \end{cases}$$

b) p < q,

(32)
$$\begin{cases} A = 0, \\ B = q - \phi, \\ C = [1 - (\phi/q)^{x}], \\ D = [1 - (\phi/q)^{x-k}]. \end{cases}$$

When p = q, we have

(33)
$$\begin{cases} \psi_{k,x}(z) = \frac{z}{2x + (1 - 2x)z}, & x \leq k, \\ = \frac{2(x - k) + [1 - 2(x - k)]z}{2x + (1 - 2x)z}, & x > k. \end{cases}$$

Similarly we have

10 1 - 0

(34)
$$\begin{pmatrix} \mu_{k,x} = \frac{(q/p)^{k} [(p/q)^{x} - 1]}{(p-q)}, & x \leq k, \\ = \frac{1 - (q/p)^{k}}{(p-q)}, & x \geq k, \end{pmatrix}$$

b) $\phi < q$,

(35)
$$\begin{cases} \mu_{k,x} = \frac{(p/q)^{x} - 1}{(p-q)}, & x \leq k, \\ = \frac{(p/q)^{x} (1 - (q/p)^{k})}{(p-q)}, & x \geq k \end{cases}$$

and, when p = q,

(36)
$$\begin{pmatrix} \mu_{k,x} = 2x, & x \leq k, \\ = 2k, & x \geq k. \end{pmatrix}$$

Some rather interesting qualitative observations may be drawn from these results for the walk with a single absorbing point at the origin.

If the mean drift, i.e. in this case p-q, is zero or towards the origin, the distribution of the occupation total is geometric, rather than modified geometric, at all points on the same side of the starting point as the absorbing point i.e. for all $x \leq k$. It is possible to give a simple intuitive justification of this result. It is well-known, see e.g. Lindley [5], that ultimate absorption is certain for any one-dimensional random walk with a single absorbing barrier, where the drift is zero or directed towards the absorbing barrier. Thus in the present situation any point $x \leq k$ is attained with probability 1 whereafter subsequent arrivals at x depend only on the probability of return to x, $f_{x,x}$, for a walk starting at x. Consequently the distribution of the occupation total $N_{k,x}$ at x is given by

(37)
$$p_{k,x}(n) \equiv \Pr\{N_{k,x} = n\} = (1 - f_{x,x})f_{x,x}^{n-1}; \text{ all } k,$$

which is geometric, as required. Obviously this distribution should be independent of the starting point of the walk, k, and this is seen to be the case by the forms of (26) and (32), and (33).

Another interesting effect is revealed by the expressions (34) and (35) for $\mu_{k,x}$; namely that, for drift away from the absorbing point, i.e. $p \ge q$, the expected occupation totals at points on the opposite side of the starting point to the absorbing point are constant and depend only on the starting point, k. This effect is not associated with a constant form of the distribution of the occupation totals for $x \ge k$, except in the case p = q. Again simple intuitive arguments confirm this observed effect.

As in the previous section, the forms of the results (31)-(36) are conveniently checked by recovery of the probability of ultimate absorption, q(0), and the expected duration of the walk, D_k (when p < q, so that this is finite), for which explicit expressions are available in the literature. (See e.g. Feller [3] p. 318.)

526

[9]

6. The unrestricted walk

Finally we may complete the discussion for the binomial walk by obtaining explicit expressions for the distribution of the occupation totals, and the expected occupation totals, for the unrestricted walk. These appear as the appropriate limiting forms of the expressions obtained for the walk with a single absorbing point at the origin.



Figure 1. Expected occupation totals for the binomial random walk.

V. D. Barnett

Assuming, without loss of generality, that the unrestricted walk starts at the origin we obtain, by putting x = y+k and letting k tend to infinity,

(38)
$$\begin{cases} \psi_{\boldsymbol{y}}(z) = 1 - \left\{\frac{1 - |p-q|}{2q}\right\}^{\boldsymbol{y}} \frac{1 - z}{1 - (1 - |p-q|)z}, \quad \boldsymbol{y} \ge 0, \\ = 1 - \left\{\frac{2p}{1 + |p-q|}\right\}^{\boldsymbol{y}} \frac{1 - z}{1 - (1 + |p-q|)z}, \quad \boldsymbol{y} \le 0, \end{cases}$$

and

(39)
$$\begin{cases} \mu_{\mathbf{y}} = \frac{1}{|p-q|} \left\{ \frac{1-|p-q|}{2q} \right\}^{\mathbf{y}}, \quad \mathbf{y} \ge 0 \\ = \frac{1}{|p-q|} \left\{ \frac{2p}{1+|p-q|} \right\}^{\mathbf{y}}, \quad \mathbf{y} \le 0. \end{cases}$$

Obviously, when p = q return to any point is certain so that the distribution of the occupation total at that point is improper and the expected occupation total infinite. We see, in this case, from (38) and (39), that not only is the expected occupation total constant at any point in the direction of drift, but its distribution is geometric and independent of the particular point concerned. Again it is possible to justify this observation by simple intuitive probabilitistic arguments.

In conclusion it is perhaps of interest to summarise the various qualitative properties of the $\mu_{k,x}$ which are discussed above, for the different cases of 2, 1 or 0 absorbing points. This is best done diagrammatically. Figure 1 presents the typical forms of the $\mu_{k,x}$ in the different situations. For convenience the $\mu_{k,x}$ are presented as continuous curves rather than discrete points.

References

- Good, I. J. (1961), The frequency count of a finite Markov chain and the transition to continuous time, Ann. Math. Stat. 32, 41-49.
- [2] Kemperman (1961), The Passage Problem for a Stationary Markov Chain, Univ. of Chicago Press.
- [3] Feller, W. (1960), An Introduction to Probability Theory and its Applications, I. 2nd Edition. Wiley, New York.
- [4] Kemeny, J. G. and Snell, J. L. (1960), Finite Markov Chains, Van Nostrand, Princeton, N.J., U.S.A.
- [5] Lindley, D. V. (1952), The theory of queues with a single server, Proc. Camb. Phil. Soc. 48, 277-289.

University of Birmingham.

528