Adv. Appl. Prob. **19**, 505–507 (1987)
Printed in N. Ireland

LETTERS TO THE EDITOR

THE MOMENTS OF THE RANDOM VARIABLE FOR THE NUMBER OF RETURNS OF A SIMPLE RANDOM WALK

ADRIENNE W. KEMP,* *University of St Andrews*

Abstract

The number of returns to the origin for a simple random walk starting and ending at the origin is reconsidered; closed-form expressions in μ are given for μ_3 and μ_4.

RETURNS TO ORIGIN; DIFFERENTIAL EQUATION; GENERALIZED HYPERGEOMETRIC FUNCTION

Let X_k, $k = 1, 2, \cdots$ be independent identically distributed Bernoulli variables with probability generating function $(1+s)/2$, and suppose that $S_j = X_1 + \cdots + X_j$, $j = 1, 2, \cdots, 2n$, with $S_0 = S_{2n} = 0$. Consider the random variable Y given by the number of times that $S_j = 0$ for $j = 0, 1, 2, \cdots, 2n$, i.e. the total number of visits to the origin for a simple random walk starting and finishing at the origin. The support of Y is $2, 3, \cdots, n+1$; let its probability generating function be $G_Y(s) = \sum p_y s^y$.

The variable Y occurs in the random walk treatment of rank order indicators for two-sample tests (Dwass (1967)), and analogously in a distribution-free CUSUM procedure (McGilchrist and Woodyer (1975)). Closed-form expressions for the first two moments were given by Katzenbeisser and Hackl (1986) in an investigation of the usefulness of Y as an alternative to the Kolmogorov–Smirnov two-sample test statistic. The asymptotic behaviour of the moments was examined by Katzenbeisser and Panny (1986).

The purpose of this note is to obtain a differential equation for the moment generating function, and hence a recurrence relation for the moment, by extending the results in Kemp (1968); this enables the corrected moment μ_i to be expressed as a polynomial in μ of degree i. For numerical results highly accurate values of μ are needed, but Katzenbeisser and Hackl's exact formula for μ is dfficult to compute when n is large. A direct derivation of an asymptotic expansion for μ is given, and its numerical accuracy is discussed; this expansion is asymptotically equivalent to that given by Katzenbeisser and Panny, and by Katzenbeisser and Hackl. As $n \to \infty$ the distribution of $Y/(2n^{\frac{1}{2}})$ was shown by Katzenbeisser and Panny to converge to a standard Weibull with parameter 2; calculation of the coefficients of skewness and kurtosis shows that the rate of convergence is very slow.

Received 20 October 1986; revision received 13 November 1986.

* Postal address: Department of Statistics, The North Haugh, University of St Andrews, St Andrews KY16 9SS, Scotland.

Dwass showed that the cumulative probabilities are

$$\sum_{y>k} p_y = 2^k(2n-k)!\,n!/(n-k)!\,(2n)!, \qquad k = 1, 2, \cdots, n. \tag{1}$$

By applying a recursive procedure to this formula Katzenbeisser and Hackl found that

$$\mu = 2^{2n} n!\,n!/(2n));$$

however, when n is large this expression is difficult to calculate accurately. By the Gauss duplication formula for the gamma function,

$$\mu = \pi^{\frac{1}{2}}\Gamma(n+1)/\Gamma(n+\tfrac{1}{2}).$$

Stirling's and Barnes' asymptotic expansions for $\ln\Gamma(n+1)$ and $\ln\Gamma(n+\frac{1}{2})$ then give

$$\begin{aligned} \ln(\mu) &\approx \tfrac{1}{2}\ln(\pi) + (n+\tfrac{1}{2})\ln(n) - n + \tfrac{1}{2}\ln(2\pi) + 1/12n - 1/360n^3 \\ &\quad + 1/1260n^5 - \cdots - n\ln(n) + n - \tfrac{1}{2}\ln(2\pi) + 1/24n \\ &\quad - 7/2880n^3 + 31/40320n^5 - \cdots \\ &\approx \tfrac{1}{2}\ln(n\pi) + 1/8n - 1/192n^3 + 1/640n^5 - \cdots. \end{aligned} \tag{2}$$

This asymptotic expansion is a logarithmic version of that obtained by Katzenbeisser and Panny (1986) (using a result from Panny (1984)), and by Katzenbeisser and Hackl (1986); the logarithmic form is more appropriate for the calculation of the power of μ.

From (2)

$$\mu \doteqdot (n\pi)^{\frac{1}{2}}\exp(1/8n - 1/192n^3). \tag{3}$$

The use of just two terms in the exponent in (3) is sufficient to give two decimal places of accuracy for $n = 1$, three places for $n = 2$, four places for $n = 3$, at least five places for $n > 3$, at least six places for $n > 7$, and at least seven places for $n > 10$.

Consider now the higher moments. From (1)

$$P_y = 2^{y-1}(2n-y)!\,(y-1)n!/(n-y+1)!\,(2n)!, \qquad y = 2, \cdots, n+1,$$

and

$$G_Y(s) = s^2\,{}_2F_1[2, 1-n; 2-2n; 2s]/(2n-1);$$

this shows that the distribution of Y belongs to the wide class of discrete distributions studied by Kemp (1968).

Corollary (a) of Theorem 9 in Kemp (1968) states that the probability generating function

$$H(s) = \frac{s^c\,{}_pF_q[a_1, \cdots, a_p; b_1, \cdots, b_q; \lambda s]}{{}_pF_q[a_1, \cdots, a_p; b_1, \cdots, b_q; \lambda]}$$

satisfies the differential equation

$$(\theta - c)(\theta - c + b_1 - 1)\cdots(\theta - c + b_q - 1)H(s) = \lambda s(\theta - c + a_1)\cdots(\theta - c + a_p)H(s),$$

where θ is the differential operator sd/ds.

Now the moment generating function about a constant m is $\exp(-mt)K(e^t)$ where $K(s)$ is the probability generating function, and

$$\exp(-mt)[\theta^j K(s)]_{s=e^t} = (D+m)^j\{\exp(-mt)K(e^t)\}$$

where D is the differential operator d/dt; so the moment generating function

$\exp(-mt)H(e^t)$ satisfies

$$(D+m-c)(D+m-c+b_1-1)\cdots(D+m-c+b_q-1)\{\exp(-mt)H(e^t)\}$$
$$=\lambda e^t(D+m-c+a_1)\cdots(D+m-c+a_p)\{\exp(-mt)H(e^t)\}.$$

Identifying the coefficient of t^i in this equation gives a recurrence relation for the uncorrected moments when $m=0$, and one for the corrected moments when $m=\mu$.

Because $G_Y(s)$ has the form $H(s)$

$$(D+\mu-2)(D+\mu-2n-1)\{\exp(-\mu t)G_Y(e^t)\}$$
$$=2e^t(D+\mu)(D+\mu-n-1)\{\exp(-\mu t)G_Y(e^t)\}$$

and

$$\mu_{i+2}=-(2\mu+1)\mu_{i+1}+(4n+2-\mu-\mu^2)\mu_i$$
$$-\sum_{j=1}^{i}\binom{i}{j}[2\mu_{i+2-j}+2(2\mu-n-1)\mu_{i+1-j}+2\mu(\mu-n-1)\mu_{j-1}].$$

Putting $i=0$ gives Katzenbeisser and Hackl's result

$$\mu_2=2(2n+1)-\mu-\mu^2.$$

When $i=1$ and $i=2$ we get

$$\mu_3=2(n+1-\mu)\mu-(3+2\mu)\mu_2$$
$$=-6(2n+1)-\mu(6n-1)+3\mu^2+2\mu^3$$

and

$$\mu_4=2(n+1-\mu)\mu+(8n+4-9\mu-\mu^2)\mu_2-(5+2\mu)\mu_3$$
$$=2(2n+1)(8n+19)+\mu(12n-13)-16\mu^2-6\mu^3-3\mu^4$$

respectively.

The highest power of μ in the recurrence formula for μ_{i+2} is given by $\mu\mu_{i+1}$ and $\mu^2\mu_i$. So, by induction, μ_i is expressible as a polynomial in μ of degree i, for all finite i. Asymptotic approximations for the moments are obtainable by the use of expression (3) for μ.

Katzenbeisser and Panny note that, as $n\to\infty$, $Y/(2n^{\frac{1}{2}})$ tends to a standard Weibull distribution with parameter 2. However, the rate of convergence appears to be too slow for this result to be of much practical use, either for the Katzenbeisser and Hackl two-sample test or for the corresponding non-parametric CUSUM procedure. When $n=10$, 100, 1000, 10000, ∞, the coefficients of skewness and kurtosis $(\mu_3/\mu_2^{3/2},\ \mu_4/\mu_2^2)$ take the values (0·168, 2·388), (0·498, 2·947), (0·590, 3·149), (0·618, 3·214) and (0·631, 3·245).

References

DWASS, M. (1967) Simple random walk and rank order statistics. *Ann. Math. Statist.* **38,** 1042–1053.

KATZENBEISSER, W. AND HACKL, P. (1986) An alternative to the Kolmogorov–Smirnov two-sample test. *Commun. Statist. Theory Methods* **15,** 1163–1177.

KATZENBEISSER, W. AND PANNY, W. (1986) A note on the higher moments of the random variable T associated with the number of returns of a simple random walk. *Adv. Appl. Prob.* **18,** 279–282.

KEMP, A. W. (1968) A wide class of discrete distributions and the associated differential equations. *Sankhyā* A **30,** 401–410.

MCGILCHRIST, C. A. AND WOODYER, K. D. (1975) Note on a distribution-free CUSUM technique. *Technometrics* **17,** 321–325.

PANNY, W. (1984) *The maximal deviation of lattice paths. Mathematical Systems in Economics* Vol. 91. Verlagsgruppe Athenäum/Hain/Hanstein, Königstein/Ts.