Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R.J.M.M. Does, R. Helmers, C.A.J. Klaassen

Approximating the percentage points of Greenwood's statistic with Cornish-Fisher expansions

Department of Mathematical Statistics Report MS-R8405 April

APPROXIMATING THE PERCENTAGE POINTS OF GREENWOOD'S STATISTIC WITH CORNISH-FISHER EXPANSIONS

R.J.M.M. DOES
Department of Medical Informatics and Statistics, University of Limburg, P.O. Box 616, 6200 MD Maastricht

R. HELMERS
Centre for Mathematics and Computer Science, Amsterdam

C.A.J. KLAASSEN
Department of Mathematics, University of Leyden, P.O. Box 9512, 2300 RA Leyden

It is suggested that approximating the exact percentage points of Greenwood's statistic with Cornish-Fisher expansions is useful for not too small sample sizes.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 62E20, 62E25.
KEY WORDS & PHRASES: Greenwood's statistic, Cornish-Fisher expansions, uniform spacings, goodness-of-fit.
NOTE: This report will be submitted for publication elsewhere.

Report MS-R8405
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. INTRODUCTION

Let G_n denote the sum of the squared lengths of the n+1 spacings obtained when a random sample of size n is drawn from the uniform distribution on (0,1). This statistic was proposed by Greenwood (1946) and it is used in goodness-of-fit problems. However, the exact distribution of G_n under the hypothesis is not known in a manageable form for $n > 3$. Recently there has been some interest in the problem to obtain useful approximations for the percentage points of G_n. Burrows (1979), using recursion and numerical integration, has produced a table of exact percentage points of G_n for n=2 (1) 10. Currie (1981) extended the tabulation of Burrows (1979) up to sample size 20. Hill (1979; cf. also his corrigendum (1981)) fits Johnson curves and lognormal curves and most recently, Stephens (1981) approximated percentage points for the Greenwood statistic for various sample sizes n and values of the level α by fitting Pearson curves to the first four moments.

In this paper we suggest the Cornish-Fisher expansion given in Section 2 to approximate the percentage points for the Greenwood statistic. We present some Monte Carlo results indicating that an approximation of the upper percentage points of G_n by means of Cornish-Fisher expansions will work quite reasonable for $n \geq 12$.

2. CORNISH-FISHER EXPANSIONS

Let $U_1, U_2, \ldots$ be a sequence of independent uniform (0,1) random variables. For n=1,2,... , $U_{1:n} \leq U_{2:n} \leq \ldots \leq U_{n:n}$ denote the ordered $U_1, U_2, \ldots, U_n$. Let $U_{0:n}=0$ and $U_{n+1:n}=1$. Uniform spacings are defined by

$$(2.1) \qquad D_{in} = U_{i:n} - U_{i-1:n}, \qquad \text{for} \quad i=1,2,\ldots,n+1.$$

The statistic

$$(2.2) \qquad G_n = \sum_{i=1}^{n+1} D_{in}^2 ,$$

proposed by Greenwood (1946), is by far the best known statistic based on uniform spacings for testing uniformity. The asymptotic normality and formulae for the exact cumulants were derived by Moran (1947; see also the corrigendum (1981) correcting an error in the formula for the third cumulant). In Does and Helmers (1982) and in Does, Helmers and Klaassen (1984) Edgeworth expansions with approximate cumulants for the sum of a function of uniform spacings with remainder $o(n^{-1})$ were established. For the special case of a quadratic function we can replace these approximate cumulants by their exact counterparts given in Moran (1947, 1981). Related expansions of Cornish-Fisher type for the percentage points of these statistics can be obtained in a standard manner from these results. In this way we arrive at an expansion for the upper percentage points $c_{n\alpha}^*$ of

$$(2.3) \qquad G_n^* = \tfrac{1}{2}n^{-\frac{1}{2}}(n+2)(n+3)^{\frac{1}{2}}(n+4)^{\frac{1}{2}}G_n - n^{-\frac{1}{2}}(n+3)^{\frac{1}{2}}(n+4)^{\frac{1}{2}},$$

which is given by

$$(2.4) \qquad c_{n\alpha}^* = \tilde{c}_{n\alpha} + o(n^{-1}), \qquad \text{as } n \to \infty ,$$

where

$$(2.5) \qquad \tilde{c}_{n\alpha} = u_\alpha + \frac{\kappa_{3n}}{6}(u_\alpha^2-1) + \frac{\kappa_{4n}}{24}(u_\alpha^3-3u_\alpha) - \frac{\kappa_{3n}^2}{36}(2u_\alpha^3-5u_\alpha)$$

with (cf. Moran (1947, 1981))

$$(2.6) \qquad \kappa_{3n} = \frac{(10n-4)(n+3)^{\frac{1}{2}}(n+4)^{\frac{1}{2}}}{n^{\frac{1}{2}}(n+5)(n+6)} ,$$

$$(2.7)\qquad \kappa_{4n} = \frac{(3n^3+303n^2+42n-24)(n+3)(n+4)}{n(n+5)(n+6)(n+7)(n+8)} - 3$$

and $u_\alpha = \Phi^{-1}(1-\alpha)$. Note that G_n^* is G_n exactly standardized.

Although an asymptotic result like (2.4) may bring some comfort to the mathematician such results are obviously useless for applications unless some numerical investigations for finite samples are made. In contrast to this, fitting Johnson or Pearson distributions with the same first four moments (cf. Hill (1979, 1981) and Stephens (1981)) seems to lack such a mathematical foundation.

3. MONTE CARLO

In Table 3.1 we give the results of a small Monte Carlo experiment into the accuracy of the Cornish-Fisher approximation $\tilde{c}_{n\alpha}$ to the upper percentage points $c_{n\alpha}^*$ of G_n^* (cf. (2.3) - (2.5)). We compare this approximation (referred to as CF) to the one introduced by Stephens (1981) which is based on Pearson curves (referred to as PS).

Let $\alpha(CF)$ denote the probability that G_n^* exceeds $\tilde{c}_{n\alpha}$. Similarly, let $\alpha(PS)$ denote the probability that G_n^* exceeds the approximation for $c_{n\alpha}^*$, determined by fitting Pearson curves in Stephens (1981). To obtain a rough idea of the performance of these approximations we estimated (with the aid of a Monte Carlo simulation based on 40,000 samples) the 'errors' $\Delta(CF) = \alpha - \alpha(CF)$ and $\Delta(PS) = \alpha - \alpha(PS)$ for some values of α and n; the standard deviations σ_α of these pseudo-estimators are given in the last row of the table.

Thus for n=12 and α=.05 the deviation $\Delta(CF)$ is estimated as .004±.002, which means that the true level of significance lies in between .044 and .048 when testing at the nominal level .05.

TABLE 3.1

sample size	$\Delta(CF).10^2$				$\Delta(PS).10^2$			
	$1-\alpha$				$1-\alpha$			
n	.90	.95	.975	.99	.90	.95	.975	.99
12	-.2	.4	.4	.3	.2	.1	.0	-.0
14	-.1	.5	.5	.3	.3	.1	.0	-.0
16	-.3	.5	.5	.3	.2	.1	.0	-.1
18	-.4	.4	.5	.4	.1	-.0	-.1	.0
20	-.2	.6	.5	.4	.3	.2	-.0	.0
25	-.2	.6	.6	.3	.2	.3	.1	-.0
30	-.5	.4	.5	.3	-.0	-.0	-.0	-.0
40	-.4	.2	.5	.3	-.1	-.2	-.0	-.0
50	-.4	.3	.4	.3	-.1	-.1	-.0	-.1
60	-.4	.3	.6	.4	-.2	-.0	.1	.1
80	-.2	.2	.3	.2	-.1	-.1	-.0	-.0
100	.0	.4	.4	.2	.1	.1	.1	.1
200	.1	.3	.2	.2	.1	.1	.1	.1
500	.2	.1	.1	.1	.1	.1	.1	.1
$\sigma_\alpha.10^2$	.2	.1	.1	.0	.2	.1	.1	.0

4. CONCLUSIONS

Our numerical results indicate that Cornish-Fisher approximation behaves quite satisfactory for the cases that were investigated, although its performance is obviously inferior to that of Stephens' Pearson curves approximation.

It should be noted that the approximate percentage points $\tilde{c}_{n\alpha}$ (cf. (2.5)) can be computed for any value of α and n. On the other hand, to apply the Pearson curves method of Stephens (1981) one has either to rely on interpolation for values α and n different from those occurring in Stephens (1981) or to extend his table.

Table 3.1 suggests that the Cornish-Fisher approximation is a conservative one for $\alpha \leq .05$ and $n \geq 12$. Only upper percentage points of G_n^* are considered. The reason being that the Cornish-Fisher approximation for estimating lower percentage points of G_n^* turns out to behave rather disappointing for not very large (say $n < 100$) sample sizes. Fortunately, however, mainly upper percentage points are of interest in applications, e.g. in testing for uniformity.

ACKNOWLEDGEMENT. The authors are grateful to Mr. M.P.E. Janssen for assistance in computing Table 3.1.

REFERENCES.

[1] BURROWS, P.M. (1979), *Selected percentage points of Greenwood's statistic*, J.R. Statist. Soc. A, 142, 256-258.

[2] CURRIE, I.D. (1981), *Further percentage points of Greenwood's statistic*, J. Roy. Statist. Soc. A, 144, 360-363.

[3] DOES, R.J.M.M. and R. HELMERS (1982), *Edgeworth expansions for functions of uniform spacings*, in: Coll. Math. Soc. J. Bolyai, Vol. 32: Nonparametric Statistical Inference, B.V. Gnedenko, M.L. Puri and I. Vincze (eds), North-Holland, Amsterdam, 203-212.

[4] DOES, R.J.M.M., R. HELMERS and C.A.J. KLAASSEN (1984), *On the Edgeworth expansion for the sum of a function of uniform spacings*, Centre for Mathematics and Computer Science Report MS-R8404, Amsterdam.

[5] GREENWOOD, M. (1946), *The statistical study of infectious diseases*, J. Roy. Statist. Soc. A, 109, 85-109.

[6] HILL, I.D. (1979), *Approximating the distribution of Greenwood's statistic with Johnson distributions*, J. Roy. Statist. Soc. A, 142, 378-380, Corrigendum, J. Roy. Statist. Soc. A, 144 (1981), 388.

[7] MORAN, P.A.P. (1947), *The random division of an interval*, J. Roy. Statist. Soc. B, 9, 92-98, Corrigendum, J. Roy. Statist. Soc. A, 144 (1981), 388.

[8] STEPHENS, M.A. (1981), *Further percentage points for Greenwood's statistic*, J. Roy. Statist. Soc. A, 144, 364-366.