Biometrika (1985), **72**, 1, *pp*. 208–210
Printed in Great Britain

Asymptotic points for a test of symmetry about a specified median

By RICHARD A. LOCKHART AND CHRISTOPHER G. McLAREN
Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Summary

We examine the asymptotic distribution and give asymptotic critical values for a statistic suggested by Hill & Rao (1977) for testing the symmetry of a continuous distribution about a specified median.

Some key words: Asymptotic critical value; Median; Test of symmetry.

1. Introduction

Hill & Rao (1977) have discussed a family of statistics, based on Cramér–von Mises statistics, for testing the symmetry of a continuous distribution about a specified median and have suggested one member of the family as a useful test statistic. This statistic, T_n, is calculated as follows. Suppose that $S = \{x_1, \ldots, x_n\}$ is a random sample from an unspecified continuous distribution. We wish to test the null hypothesis that the distribution is symmetric about the origin. Let $z_1, \ldots, z_n$ be the ordered absolute values $|x_j|$, $x_j \in S$ and let N_k and P_k $(k = 1, .., n)$, be the numbers of negative and positive values, respectively, in S such that $|x_j| < z_k$. Let S' be the set of reciprocal values $x_1^{-1}, \ldots, x_n^{-1}$ and define Z'_k, N'_k and P'_k analogously to Z_k, N_k and P_k. Then

$$T_n = \left\{\sum_{k=1}^{n} (N_k - P_k)^2 + \sum_{k=1}^{n} (N'_k - P'_k)^2\right\} \Big/ (2n^2).$$

Large values of T_n constitute evidence against the null hypothesis. Hill & Rao (1977) gave upper tail critical values for $\frac{1}{4}n^2 T_n$ for selected significance levels and for sample sizes 10 to 24. They also gave a representation of the asymptotic null distribution of T_n. In a later paper, Hill & Rao (1981) proposed a different statistic having a known asymptotic null distribution, which can be represented as a sum of weighted chi-squared random variables. They said that the asymptotic distribution of T_n is difficult to work with. However, they reported Monte Carlo studies, which indicate that T_n has good power against alternatives involving a median shift of a symmetric distribution.

In the present paper we show that the asymptotic null distribution of T_n can also be expressed as a sum of weighted chi squared random variables, and calculate asymptotic percentage points. These are very close to those for sample size 24, as suggested by Hill & Rao (1977), and so may be used for $n > 24$. Thus the statistic T_n is available for all n.

2. The asymptotic distribution of T_n

Let $c_j = 1/(2j-1)$ and let c' be the vector $c' = (c_1, c_2, \ldots)$. Also let z_1 and z_2 be independent infinite length random vectors with independent standard normal components. Define a diagonal matrix Q_1 with jth diagonal element c_j^2 and define $Q_2 = 2cc' + Q_1$. Then Hill & Rao (1977) state that the asymptotic distribution of T_n, or any member of the family of statistics from which T_n was selected, is the same as the distribution of D, where

$$\pi^2 D = z_1' Q_1 z_1 + z_2' Q_2 z_2.$$

In view of the normality and independence of z_1 and z_2, the rth cumulant of $\pi^2 D$ (Kendall & Stuart, 1977, p. 382) is

$$\kappa_r = 2^{r-1}(r-1)!\,\mathrm{tr}\,(Q_1^r + Q_2^r).$$

Let $S_p = \Sigma c_j^p = (1-2^{-p})\,\zeta(p)$, where the sum is over $j = 1, \ldots, \infty$, and where $\zeta(p)$ is the zeta function of Riemann (Abramowitz & Stegun, 1965, p. 807), with special values $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$ and $\zeta(8) = \pi^8/9450$. Then by expanding $Q_2^r = (2cc' + Q_1)^r$ and taking the trace for $r = 1, 2, 3$ and 4, we obtain

$$\kappa_1 = 4S_2 = \pi^2/2, \quad \kappa_2 = 8S_2^2 + 12S_4 = \pi^4/4,$$
$$\kappa_3 = 64S_2^3 + 96S_2 S_4 + 64S_6 = 19\pi^6/60, \qquad (1)$$
$$\kappa_4 = 768S_2^4 + 1536S_2^2 S_4 + 768S_2 S_6 + 384S_4^2 + 480S_8 = 529\pi^8/840.$$

If v is an eigenvector of Q_2 corresponding to the eigenvalue λ then the jth equation in the set $Q_2 v = \lambda v$ can be rearranged to give

$$v_j = 2(c'v)\,c_j/(\lambda - c_j^2) \quad (j = 1, 2, \ldots).$$

By forming the product $c'v$ it follows from these equations that any eigenvalue λ of Q_2 must satisfy

$$1 = 2\sum_{j=1}^{\infty} c_j^2/(\lambda - c_j^2). \qquad (2)$$

Hence Q_2 has one eigenvalue larger than c_1^2, and subsequent values between successive terms of the sequence $c_1^2, c_2^2, \ldots$. The terms of this sequence are also the eigenvalues of Q_1.

Defining μ to be $\lambda^{-\frac{1}{2}}$, we can write (2) as

$$1 = \pi\mu[2\pi\mu\,\Sigma\,1/(\pi^2 j^2 - \pi^2\mu^2) - \tfrac{1}{2}\pi\mu\,\Sigma\,1/\{\pi^2 j^2 - \pi^2(\tfrac{1}{2}\mu)^2\}],$$

where the sums are over $j = 1, \ldots, \infty$. Because (Abramowitz & Stegun, 1965, p. 75)

$$1/z - \cot z = 2z\,\Sigma\,1/(\pi^2 k^2 - z^2) \quad (z \neq \pm k\pi;\ k = 0, 1, \ldots),$$

where the sum is over $k = 1, \ldots, \infty$, and

$$2\cot z = \cot(\tfrac{1}{2}z) - \tan(\tfrac{1}{2}z),$$

it follows that μ must satisfy

$$2/(\pi\mu) = \tan(\tfrac{1}{2}\pi\mu).$$

This can be solved numerically to give eigenvalues of Q_2.

If $\lambda_1, \lambda_2, \ldots$ are the decreasing sequence of eigenvalues of Q_1 and Q_2 together, and χ_j^2 $(j = 1, 2, \ldots)$ are independent chi-squared random variables with one degree of freedom then, in distribution,

$$\pi^2 D = \sum_{j=1}^{\infty} \lambda_j \chi_j^2. \qquad (3)$$

3. Critical values of the asymptotic distribution

We approximate $\pi^2 D$ by using the first 50 terms of the series (3). Let

$$\pi^2 D^* = \sum_{j=1}^{50} \lambda_j \chi_j^2.$$

Then $\pi^2 D^*$ differs in mean and variance from $\pi^2 D$ by $2{\cdot}02002 \times 10^{-2}$ and $5{\cdot}5 \times 10^{-6}$ respectively.

The distribution of $\pi^2 D^*$ is evaluated at specific points by numerical inversion of its characteristic function following the method of Imhof (1961). Critical values of $\pi^2 D^*$ are found by searching for points giving appropriate values of the distribution.

The critical values of D given in Table 1 are obtained by adding $2{\cdot}02002 \times 10^{-2}$ to the points of $\pi^2 D^*$ to correct the difference in mean and then scaling by $1/\pi^2$.

The cumulants, (1), were also used to fit a Pearson curve to the distribution of $\pi^2 D$; the percentage points agreed well with those in Table 1. Furthermore, following a suggestion of a referee who observed that $\kappa_2 = \kappa_1^2$, as for the exponential distribution, we suggest that interpolation in the table be made by fitting a linear relation between the percentage points and the logarithm of the significance level.

Table 1. *Upper tail critical points for* T_n

Sample size	Significance level 50%	25%	15%	10%	5%	2·5%	1%	0·5%
20				1·175	1·525	1·930	2·535	
21				1·145	1·526	1·984	2·397	
22				1·147	1·576	1·890	2·461	
23				1·181	1·510	1·949	2·516	
24				1·137	1·550	1·953	2·422	
∞	0·337	0·647	0·899	1·111	1·489	1·886	2·428	2·847

The critical values for $\frac{1}{4}n^2 T_n$ for $n = 20$ to 24 given by Hill & Rao (1977) have been rescaled by $4/n^2$ and are also given in Table 1. The exact distribution of T_n is discrete, but it can be seen that the asymptotic approximation is excellent for $n = 20$ to 24; and it can safely be assumed that the approximation will work for all $n > 24$ with good accuracy for practical use.

Acknowledgements

The authors thank Professor Michael Stephens for suggesting this problem, and Dr John Paine for helpful suggestions concerning the computations.

References

Abramowitz, M. & Stegun, I. A. (Eds). (1965). *Handbook of Mathematical Functions*. New York: Dover.

Hill, D. L. & Rao, P. V. (1977). Tests of symmetry based on Cramér–von Mises statistics. *Biometrika* **64**, 489–94.

Hill, D. L. & Rao, P. V. (1981). Tests of symmetry based on Watson statistics. *Comm. Statist.* A **10**, 1111–25.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419–26.

Kendall, M. G. & Stuart, A. (1977). *The Advanced Theory of Statistics*, **1**, 3rd ed. London: Griffin.

[*Received June* 1984. *Revised July* 1984]