(24)

Journal of the Indian Statistical Association
Vol. 21. 1983 19—26

COMPARISON OF THE LIMITING EFFICIENCIES OF TWO CHI-SQUARE TYPE TESTS FOR THE CIRCLE

By

J. S. Rao

Department of Mathematics, University of California, Santa Barbara, U. S. A.

and

Y. Yoon

The Wyatt Company, Actuaries and Consultants, Dallas, U. S. A.

ABSTRACT

Two statistics for testing uniformity of circular data, one proposed by the first author and the other by Watson, are compared through their approximate Bahadur efficiencies. The results as to which is superior differ depending on whether one is testing uniformity against unimodal von Mises alternatives or against appropriate multimodal alternaitves.

1. Introduction :

For testing whether a random sample $\theta_1, \cdots, \theta_n$ on the circle comes from a uniform distribution, Ajne (1968) proposed the statistic

$$A_n = \frac{1}{n} \int_0^{2\pi} \left\{ N(\theta) - \frac{n}{2} \right\}^2 d\theta \tag{1·1}$$

where $N(\theta)$ is the number of observations in the semi-circle $[\theta, \theta + \pi)$. The null hypothesis of uniformity is rejected for large values of A_n. Rao (1972*b*) considered the following generalization of A_n which is a variation of the chi-square statistic. Suppose the circular data $\theta_1, \theta_2, \cdots \theta_n$ is grouped into m class intervals of equal width for any fixed positive integer m. Let the i-th class interval be

$$I_i(\alpha) = \left[\alpha + (i-1)\frac{2\pi}{m}, \alpha + i\frac{2\pi}{m} \right), \; i = 1, 2, \cdots, m,$$

and let $n_i(\alpha)$ be the number of observations that fall in $I_i(\alpha)$.

Then

$$\chi^2_n(m) = \frac{m}{2\pi n} \int_0^{2\pi} \sum_{i=1}^{m} \left[n_i(\alpha) - \frac{n}{m} \right]^2 d\alpha. \tag{1·2}$$

It is easily seen that A_n is a special case of $\chi^2_n(m)$ when $m = 2$. Watson (1967) considered a generalization of $(1\cdot1)$ in another direction. For any fixed positive integer m, let

$$S_x = \bigcup_{i=1}^{m} \left[x + \frac{2i-2}{2m} . 2\pi,\; x + \frac{2i-1}{2m} . 2\pi \right)$$

and let $N_m(x)$ be the number of observations falling in this set. Then Watsou (ibid.) suggested the statistic

$$A_n(m) = \frac{2\pi}{n} \int_0^{2\pi} \left\{ N_m(x) - \frac{n}{2} \right\}^2 dx \tag{1·3}$$

for testing uniformity. Clearly $m=1$ in $(1\cdot3)$ corresponds to A_n defined in $(1\cdot1)$. On the other hand, Beran (1969a) considered the following general class of tests of uniformity. Let $f(\theta)$ be any probability density function (p. d. f.) on the circle. Then Beran's class of tests for testing uniformity given by

$$B_n = 2\pi \int_0^{2\pi} \left[\sum_{i=1}^{n} f(\theta + \theta_i) - n\,(2\pi)^{-1} \right]^2 d\theta. \tag{1·4}$$

It is shown in Section 2 that $\chi^2_n(m)$ defined in $(1\cdot2)$ as well as $A_n(m)$ defined in $(1\cdot3)$ are of the form $(1\cdot4)$. Using this idea, conparison is made between tests with respect to their Bahadur efficiency in Section 3.

2. The $\chi^2_n(m)$ test and $A_n(m)$ test as special cases of B_n.

In this section we show that both the statistics $\chi^2_n(m)$ and $A_n(m)$ are of the form $(1\cdot4)$. The circular p. d. f. $f(\cdot)$ coresponding to each of these tests and its Fourier expansion are obtained. An alternte form of B_n (see for instance Mardia 1972, p. 190) which is more useful, is the following. Let the Fourier expansion of $f(\theta)$ be

$$f(\theta) = (2\pi)^{-1} \left\{ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right\}.$$

Then

$$B_n = \sum_{i=1}^{n} \left[\sum_{j=1}^{n} h\,(\theta_i - \theta_j) \right]$$

where

$$h(\theta) = 2\sum_{p=1}^{\infty} \left(\alpha_p^2 + \beta_p^2 \right) \cos p\theta .$$

Let

$$D(\theta) = \pi - | (\pi - | \theta |) | , \quad 0 \leqslant \theta < 2\pi$$

denote the " circular ditance " between the points θ and the origin. Then $D_{jj'} = D(\theta_j - \theta_{j'})$ gives the circular distance between the jth and j'th observations. In terms of these circular distances, Rao (1972b) showed that the statistic in (1·2) is equal to

$$\chi_n^2 (m) = m - n + \frac{m^2}{2\pi n} \sum_{\substack{\{(j,j'),\, j \neq j'\} \\ D_{jj'} \leqslant 2\pi/m}} \sum (2\pi/m - D_{jj'}) ,$$

which is a computationally useful formula. One can rewrite this as

$$\chi_n^2 (m) = \frac{1}{n} \sum_{j=1}^{n} \sum_{j'=1}^{n} \left\{ \frac{m^2}{2\pi} (2\pi/m - D_{jj'}) \cdot I(D_{jj'}) - 1 \right\}$$

where

$$I(x) = \begin{cases} 1 \text{ if } x \leq 2\pi/m \\ 0 \text{ otherwise.} \end{cases}$$

Thus

$$\chi_n^2 (m) = \frac{1}{n} \sum_{j=1}^{n} \sum_{j'=1}^{n} d(\theta_j - \theta_{j'}) \tag{2·1}$$

with the definition

$$d(\theta) = \frac{m^2}{2\pi} (2\pi/m - D(\theta)) \cdot I(D(\theta)) - 1.$$

Clearly the function $d(\theta)$ is symmetric with respect to zero and periodic with period 2π, and it is easy to check that

$$\int_0^{2\pi} d(\theta)\, d\theta = 0 = \int_0^{2\pi} d(\theta) \sin k\theta\, d\theta,$$

and

$$\int_0^{2\pi} d(\theta) \cos k\theta\, d\theta = \frac{m^2}{\pi k^2} \left(1 - \cos \frac{2k\pi}{m} \right).$$

Hence

$$d(\theta) = \sum_{k=1}^{\infty} \frac{m^2}{\pi^2 k^2} \left(1 - \cos \frac{2k\pi}{m}\right) \cos k\theta$$

$$= \sum_{k=1}^{\infty} \left[\frac{\sin (k\pi/m)}{k\pi/m} \right]^2 \cos k\theta \qquad (2\cdot 2)$$

$$= \sum_{k \neq 0} \left| \alpha_k \right|^2 e^{ik\theta} \quad \text{where } \alpha_k = \left[\frac{\sin (k\pi/m)}{k\pi/m} \right]$$

Now define the p. d. f.

$$f(\theta) = (2\pi)^{-1} \left\{ 1 + 2 \sum_{k=1}^{\infty} \alpha_k \cos k\theta \right\}$$

on the circle and let

$$C_p = \sum_{i=1}^{n} \cos p\theta_i, \quad S_p = \sum_{i=1}^{n} \sin p\theta_i \cdot$$

Then

$$\sum_{i=1}^{n} f(\theta + \theta_i) - \frac{n}{2\pi} = \sum_{p=1}^{\infty} \alpha_p C_p \cos p\theta + \sum_{p=1}^{\infty} (-\alpha_p S_p) \sin p\theta \,.$$

On substituting this in (1.4), and using Parseval's identity,

$$B_n = 2\pi \int_0^{2\pi} \left[\sum_{i=1}^{n} f(\theta + \theta_i) - n(2\pi)^{-1} \right]^2 d\theta$$

$$= \sum_{j=1}^{n} \sum_{j'=1}^{n} 2 \cdot \sum_{k=1}^{\infty} \left(\frac{\sin k\pi/m}{k\pi/m} \right)^2 \cos k(\theta_j - \theta_{j'})$$

$$= n \cdot \chi_n^2(m)$$

from (2·1) and (2·2). Thus $\chi_n^2(m)$ is of the form (1·4).

Beran (1969b) showed that $A_n(m)$ defined in (1·3) is also a special case of B_n by taking $f(x)$ as the symmetric m–modal density

$$f(x) = \begin{cases} \dfrac{p}{\pi} & \text{if } x \in S_a \\[2ex] \dfrac{q}{\pi} & \text{otherwise} \end{cases}$$

where a is an arbitrary location parameter on the circle, $p \neq \frac{1}{2}$ and $p + q = 1$. we also have (ibid.)

$$A_n(m) = \frac{n\pi}{2} - \frac{2}{n} \sum_{j<k} \sum D(\hat{x}_j, \hat{x}_k)$$

where

$$\hat{x}_i = mx_i \pmod{2\pi}, \quad i = 1, \cdots, n,$$

and $D(x, y)$ is the circular distance between x and y. Again taking,

$$d^*(\theta) = \begin{cases} \frac{\pi}{2} - m\theta, & \text{if } 0 \leq \theta < \frac{\pi}{m} \\ \frac{\pi}{2} + m\left(\theta - \frac{2\pi}{m}\right) & \text{if } \pi/m \leq \theta \leq 2\pi/m \end{cases} \left(\bmod \frac{2\pi}{m}\right),$$

we have

$$A_n(m) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d^*(\theta_i - \theta_j). \tag{2·3}$$

It can be checked that

$$\int_0^{2\pi} d^*(\theta)\, d\theta = 0 = \int_0^{2\pi} d^*(\theta) \sin k\theta\, d\theta, \quad \text{and}$$

$$\int_0^{2\pi} d^*(\theta) \cos k\theta\, d\theta = \begin{cases} \frac{4m^2}{k^2} = \frac{4}{(2i-1)^2} & \text{if } k = (2i-1)m \\ & i = 1, 2, 3, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Hence
$$d^*(\theta) = 2 \sum_{i=1}^{\infty} \frac{2}{\pi(2i-1)^2} \cos(2i-1)m\theta$$

$$= \sum_{k \neq 0} \left|c_k\right|^2 e^{ik\theta} \tag{2·4}$$

where

$$\left|c_k\right|^2 = \begin{cases} \frac{2}{\pi(2i-1)^2}, & k = (2i-1)m \\ & i = 0, \pm 1, \pm 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

3. Approximate Bahadur Efficiencies of $\chi_n^2(m)$ and $A_n(m)$.

We first briefly set forth the concept of Bahadur approximate slope (BAS) and Bahadur approximate efficiency (BAE) that would be used in this section. Let $\{T_n\}$ be a sequence of real-valued statistics based on a random sample of size n. Bahadur (1960) defined $\{T_n\}$ to be a standard sequence for testing the hypothesis $H_0 : \kappa = \kappa_0$ if the following three conditions are satisfied.

(i) There exists a continuous probability distribution functin F such that $\lim_{n\to\infty} P_{\kappa_0}(T_n < x) = F(x)$.

(ii) There exists a constant $a, 0 < a < \infty$, such that

$$\log(1 - F(x)) = -\frac{ax^2}{2}(1 + o(1)) \quad \text{where } o(1) \to 0 \text{ as } x \to \infty.$$

(iii) There exists a real-valued function $b(\kappa)$ for any κ in the alternative, with $0 < b(\kappa) < \infty$, such that

$$\lim_{n\to\infty} P_{\kappa}(|T_n/\sqrt{n} - b(\kappa)| > x) = 0 \quad \text{for every } x > 0$$

Bahadur (1960) defines BAS of the sequence $\{T_n\}$ to be $c(\kappa) = ab^2(\kappa)$ and BAE of a standard sequence $\left\{T_n^{(1)}\right\}$ relative to another standard sequence $\left\{T_n^{(2)}\right\}$ to be

$$F_{12}(\kappa) = \frac{c_1(\kappa)}{c_2(\kappa)}.$$

In what follows, κ will denote the concentration parameter for von Mises alternatives (see (3·2) and (3·5)) and the hypothesis of uniformity corresponds to $H_0 : \kappa = 0$. Since both $A_n(m)$ and $\chi_n^2(m)$ can be expressed in the form (4·1) we can use the result (see Rao (1972a) Lemma 2.3; also Beran (1969a)) that the BSA of B_n under an alternative with distribution function $G(x)$ is given by

$$S(G) = \sum_{k\neq 0} |c_k|^2 \cdot |d_k|^2 \Big/ \max_k |c_k|^2 \qquad (3\cdot 1)$$

where $\{c_k\}$ are the Fourier coefficients of $f(\cdot)$ in B_n and $\{d_k\}$ the Fourier coefficients of G. Since $A_n(m)$ is based on dividing the whole circumference into $2m$ intervals, a valid comparison of $A_n(m)$ should be made against $\chi_n^2(2m)$. This is done first against unimodal von Mises alternatives with p. d. f.

$$g_1(\theta) = (2\pi I_0(\kappa))^{-1} e^{\kappa\cos\theta}, \quad 0 \leqslant \theta < 2\pi. \qquad (3\cdot 2)$$

Denote the BAS of $\chi^2_n(2m)$ and $A_n(m)$ against the distribution function G by S_1 (G) and S_2(G) respectively. The Fourier coefficinets of $g_1(\theta)$ are given by (see for instance Mardia (1972) p. 62) $d_l(\kappa) = d_l(\kappa)/I_0(\kappa)$ where $I_n(\kappa)$ denotes the Bessel function of the first kind of purely imaginary argument and has the expansion

$$I_n(\kappa) = \frac{(\kappa/2)^n}{\Gamma(n+1)}\left\{1 + \frac{(\kappa/2)^2}{1\cdot(n+1)} + \frac{(\kappa/2)^4}{1\cdot 2\cdot(n+1)\ (n+2)} + \cdots\right\} \quad (3\cdot 3)$$

From (2·2). (2·4) and (3·1) it follows that

$$S_1(G_1) = \sum_{l\neq 0}\left(\frac{\sin\ l\pi/2m}{l\ \sin\ \pi/2m}\right)^2 \left(\ I_l(\kappa)/I_0(\kappa)\ \right)^2$$

$$S_2(G_1) = \sum_{i=-\infty}^{\infty} \frac{1}{(2i-1)^2}\left(\ I_{(2i-1)m}\ (\kappa)/I_0(\kappa)\ \right)^2. \quad (3\cdot 4)$$

To get an idea about local efficiency (i. e., for alternatives close to the null hypothesis) one may use the expansion (3·3) and let $\kappa \to 0$ in the above expressions in (3·4). We see that the limiting BAE of $A_n(m)$ relative to $\chi^2_n(2m)$ is zero unless $m = 1$, in which case these two tests are of course, the same. This is not surprising in view of the fact that when there are more than 2 class intervals (i. e., $m > 1$) the $\chi^2_n(2m)$ test takes into account the frequencies in all these class intervals simultaneously whereas the $A_n(m)$ test lumps such frequencies into two groups, those belonging to the set S_π and those that do not. For this reason, one would expect the $\chi^2_n(m)$ test to asymptotically more efficient than the $A_n(m)$ test for testing uniformity against any unimodal alternatives.

On the other hand, if we take the m-modal von Mises density

$$g_2(\theta) = (\ 2\pi I_0(\kappa)\ )^{-1}\ e^{\kappa\ \cos\ m\theta},\ 0 \leqslant \theta < 2\pi. \quad (3\cdot 5)$$

as the alternative which has the Fourier coefficients

$$d_l = \begin{cases} \dfrac{I_p(\kappa)}{I_0(\kappa)} & \text{if} \quad l = pm \\ 0 & \text{otherwise,} \end{cases}$$

then we have

$$S_1(G_2) = \sum_{i=-\infty}^{\infty}\left(\ (2i-1)\ m \sin\ \frac{\pi}{2m}\right)^{-2}(\ I_{(2i-1)m}\ (\kappa)/I_0(\kappa)\ )^2$$

and

$$S_2(G_2) = \sum_{i=-\infty}^{\infty} (2i-1)^{-2} \left(I_{(2i-1)m}(\kappa)/I_0(\kappa)\right)^2 = S_2(G_1)$$

Therefore the BAE of $\chi^2_n(2m)$ relative to $A_n(m)$ is $\left(m \sin \frac{\pi}{2m} \right)^{-2}$. This relative efficiency is one when $m = 1$, but $A_n(m)$ is somewhat more efficient than $\chi^2_n(2m)$ for $m > 1$. In fact, for large m, this value tends to $\frac{4}{\pi^2}$. Thus in testing uniformity against m-modal von Mises alternatives given in (3·5), Watson's statistic $A_n(m)$ has better asymptotic efficiency.

References

Ajne, B. (1968) : A simple test for uniformity of a circular distribution, *Biometrika* **55,** 343–354.

Bahadur, R. R. (1960) : Stochastic comparison of tests, *Ann. Math. Statist,* **31,** 276–295.

Beran, R. J. (1969a) : Asymptotic theory of a class of tests for uniformity of a circular distribution. *Ann. Math. Statist.* **40,** 1196–206.

Beran, R. J. (1969b) : The derivation of nonparametric two sample tests for uniformity of a circular distribution, *Biometrika* **56,** 561–570.

Mardia, K. V. (1972) : *Statistics of Directional Data,* Academic Press.

Rao, J. S. (1972a) : Bahadur Efficiencies of some tests for uniformity on the circle, *Ann. Math Statist.* **43,** 468–479.

Rao, J. S. (1972b) : Some variants of chi-square for testing uniformity on the circle, *Z. Waherscheinlichkeits-theore Verw, Geb.* **22,** 33–44.

Watson, G. S. (1967) : Some Problems in the Statistics of Directions, *Bull. Int. Statis. Inst.* **42,** 374–385.