Z. Wahrscheinlichkeitstheorie verw. Geb. 22, 33–44 (1972)

Some Variants of Chi-Square for Testing Uniformity on the Circle

J. S. RAO*

1. Introduction

One of the basic problems in the analysis of circularly distributed data is to find whether a given set of observations on the circumference of the unit circle indicate any preferred direction or whether the data can be considered to have come from a uniform distribution on the circumference. We shall assume, throughout this discussion, that the observations are given in terms of angles measured with respect to some suitably chosen origin (or zero direction), taking say, the anticlockwise direction as positive. A "goodness of fit" problem on the circle then is to test whether a random sample $(\alpha_1, \ldots, \alpha_n)$ comes from a population with a completely specified distribution function $F_0(\alpha)$, $0 \leqq \alpha \leqq 2\pi$. If the specified distribution function is continuous, then the points $x_i = F_0(\alpha_i)$ may be considered as observations on the circle of unit circumference, where now, the problem is to test whether the observations $(x_1, \ldots, x_n)$ come from a uniform distribution. Thus a goodness of fit problem on the circle can also be reduced to testing for uniformity on the circle and the two problems are canonically equivalent just as they are on the line.

Broadly speaking, the test procedures available for this purpose on the line may be grouped into three categories viz., (i) the methods based on χ^2 (ii) the methods utilising the empirical distribution functions and (iii) those based on sample spacings, i.e., differences between successive order statistics. However, these methods are not, in general, directly applicable for observations on the circle because of special problems posed by the arbitrary choice of the zero direction. A test statistic should, clearly be independent of this arbitrary origin, in order that it can be meaningfully used with the circular data.

In some cases, modifications of the usual test statistics on the line, so as to make them independent of the choice of origin, have been introduced for use with the circular data. For instance, when employing the methods based on empirical distribution functions, Kuiper [6] and Watson [11] suggested such modifications for the standard Kolmogorov-Smirnov and Cramer-von Mises tests respectively. On the other hand, if the χ^2 methods were to be exploited for testing uniformity, one can make the usual χ^2 test for uniformity invariant under choice of origin, by considering the maximum possible value of χ^2 (for a given number of class intervals) or by taking the average such χ^2. We obtain the asymptotic distribution of the latter in Section 2 by using methods similar to those in Watson

* Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. AFOSR 71-2009.

[12]. Looking at the problem from another angle, if one has a suitable class of parametric alternatives for the observations, one can improve on the usual χ^2 test by concentrating on those alternatives. We show, in Section 3 that a special type of χ^2, suggested by Rao [8], gives a test based on the length of the sample resultant, when testing for uniformity against the class of "close CN alternatives". Finally Rao [10] suggests and studies the third group of tests, based on sample arc lengths, which correspond to the spacings tests on the line. (Ref. Pyke [7] and the references contained therein.)

2. An Average χ^2 on the Circle and Its Asymptotic Distribution

In this section, we consider the problem of testing for uniformity on the basis of a grouped data, by using the χ^2 methods which involve comparing the observed frequencies with those expected. The value of the usual χ^2 statistic depends, in general, on the particular grouping adopted and we therefore suggest an average type of χ^2 and find its asymptotic distribution (a.d.). We also give a computational form for this average χ^2 at the end of this section.

Suppose the circular data consisting of n independent observations $\alpha_1, \ldots, \alpha_n$ is grouped into m class intervals of equal width, the first class starting with a suitably chosen direction α. The number of classes, m, is held fixed throughout our discussion. Let the i^{th} class interval be $I_i(\alpha) = [\alpha + (i-1)\,2\pi/m, \alpha + i\,2\pi/m) = [\alpha^i, \alpha^{i+1})$, say, for $i = 1, \ldots, m$. Suppose $n_i = n_i(\alpha)$ is the number of observations that fall in $I_i(\alpha)$ and let $n = \sum_1^m n_i$ be the total number of observations. For testing uniformity on the basis of this grouped data, the usual χ^2 statistic with equal expected frequencies under the hypothesis is

$$\begin{aligned} \chi_n^2(\alpha) &= \sum_1^m [n_i(\alpha) - n/m]^2/(n/m) \\ &= n \sum_1^m [p_i(\alpha) - \pi_i^0(\alpha)]^2/\pi_i^0(\alpha) \end{aligned} \tag{2.1}$$

where $p_i(\alpha) = n_i(\alpha)/n$ and $\pi_i^0(\alpha) = 1/m$ denote the observed and hypothetical relative frequencies in the i^{th} class. The statistic given in (2.1) has, for n large, a $\chi^2_{(m-1)}$ distribution under the hypothesis of uniformity. But this statistic clearly depends on the particular grouping adopted or in other words on the starting point α. However, this dependence, it may be remarked, is not peculiar to observations on the circle alone.

The χ^2 statistic (2.1) can be made independent of α (or the particular choice of grouping) by considering, for instance, $\sup_\alpha \chi_n^2(\alpha)$ or $\int_0^{2\pi} \chi_n^2(\alpha)\, d\alpha$. We shall now find the asymptotic distribution (a.d.) of the statistic

$$\begin{aligned} \chi_n^2 &= \frac{1}{2\pi} \int_0^{2\pi} \chi_n^2(\alpha)\, d\alpha \\ &= \frac{m}{2\pi n} \int_0^{2\pi} \sum_{i=1}^m \left[n_i(\alpha) - \frac{n}{m}\right]^2 d\alpha \end{aligned} \tag{2.2}$$

under the hypothesis of uniformity. (Here the subscript n for χ^2 is used to denote the number of observations on which it is based.) The a.d. of χ_n^2 can be evaluated by adopting the following standard method. First, it can be established that the empirical process

$$\left\{\frac{1}{\sqrt{n}}\left(n_i(\alpha)-\frac{n}{m}\right),\ i=1,\ldots,m,\ 0\leqq\alpha<2\pi\right\}$$

converges to a m-variate Gaussian process on $[0, 2\pi)$. Then appealing to the invariance principle, the a.d. of χ_n^2 is the same as that of the limiting statistic χ^2, expressed in terms of the Gaussian process. The Fourier representation of this Gaussian process reduces this χ^2 to an infinite summand of m-variate complex Laplacian variables, whose theory is by now well known (see e.g. Goodman [5]). A more elementary but essentially equivalent approach is given in Watson [12] and we adopt this approach to find the a.d. in our case. Define the indicator random variables

$$\chi_j^i(\alpha)=\begin{cases}1 & \text{if } \alpha^i\leqq\alpha_j<\alpha^{i+1} \text{ i.e., if the } j^{\text{th}} \text{ observation, } \alpha_j\in I_i(\alpha)\\ 0 & \text{otherwise}\end{cases} \tag{2.3}$$

for $j=1,\ldots,n$ and $i=1,\ldots,m$. Then

$$\left(n_i(\alpha)-\frac{n}{m}\right)=\sum_{j=1}^{n}\left(\chi_j^i(\alpha)-\frac{1}{m}\right) \tag{2.4}$$

which, being a periodic function, may be expressed in Fourier series as (the Fourier representation holds for all except a finite number of α where the jumps occur)

$$\left(n_i(\alpha)-\frac{n}{m}\right)=a_{i0}+\sum_{k=1}^{\infty}(a_{ik}\cos k\alpha+b_{ik}\sin k\alpha) \tag{2.5}$$

where

$$a_{i0}=\frac{1}{2\pi}\int_0^{2\pi}\sum_j\left(\chi_j^i(\alpha)-\frac{1}{m}\right)d\alpha=0$$

and for $k\neq 0$

$$\begin{aligned} a_{ik}&=\frac{1}{\pi}\sum_j\int_0^{2\pi}\left(\chi_j^i(\alpha)-\frac{1}{m}\right)\cos k\alpha\, d\alpha\\ &=\frac{1}{\pi}\sum_j\int_{\alpha_j-\frac{2\pi i}{m}}^{\alpha_j-\frac{2\pi(i-1)}{m}}\cos k\alpha\, d\alpha \end{aligned} \tag{2.6}$$

since

$$\chi_j^i(\alpha)=\begin{cases}1 & \text{if } \alpha_j-\dfrac{2\pi}{m}<\alpha^i\leqq\alpha_j \text{ i.e., if } \alpha_j-\dfrac{2\pi i}{m}<\alpha\leqq\alpha_j-\dfrac{2\pi(i-1)}{m}\\ 0 & \text{otherwise.}\end{cases} \tag{2.7}$$

Thus, from (2.6)

$$a_{ik}=\frac{2\sin\frac{k\pi}{m}}{k\pi}\sum_{j=1}^{n}\cos k\left(\alpha_j-\frac{2\pi(i-1)}{m}\right). \tag{2.8}$$

Similarly b_{ik}, the coefficient of $\sin k\alpha$, can be shown to be

$$b_{ik}=\frac{2\sin\frac{k\pi}{m}}{k\pi}\sum_{j=1}^{n}\sin k\left(\alpha_j-\frac{2\pi(i-1)}{m}\right). \tag{2.9}$$

Now for a fixed k, consider the set of $2m$ coefficients $\{(a_{ik}, b_{ik}), i=1, \ldots, m\}$ given in (2.8) and (2.9). Since the α_j are independently and uniformly distributed on $[0, 2\pi)$, it is easy to check that

$$\begin{aligned} &E(a_{ik})=E(b_{ik})=0\\ &E(a_{ik}\,a_{jk})=E(b_{ik}\,b_{jk})=\frac{2n\sin^2\frac{k\pi}{m}}{k^2\pi^2}\cos(j-i)\frac{2\pi k}{m}\\ &E(a_{ik}\,b_{jk})=\frac{2n\sin^2\frac{k\pi}{m}}{k^2\pi^2}\sin(i-j)\frac{2\pi k}{m}\end{aligned} \tag{2.10}$$

for $i, j=1, \ldots, m$. Further, for every fixed k, when n is large, by the multivariate central limit theorem, the random vector of $2m$ variables

$$\eta'_{nk}=\frac{1}{\sqrt{n}}\{a_{1k}\,b_{1k}\,a_{2k}\,b_{2k}\ldots a_{mk}\,b_{mk}\} \tag{2.11}$$

converges in distribution to a random vector η'_k which has a $2m$ variate normal distribution with means zero and variance covariance matrix

$$\underset{(2m\times 2m)}{\Sigma_k}=E\eta_k\,\eta'_k=\frac{\left(2\sin^2\frac{k\pi}{m}\right)}{\pi^2 k^2}\times\begin{bmatrix} 1 & 0 & \ldots & \cos(m-1)\frac{2\pi k}{m} & -\sin(m-1)\frac{2\pi k}{m}\\ 0 & 1 & \ldots & \sin(m-1)\frac{2\pi k}{m} & \cos(m-1)\frac{2\pi k}{m}\\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots\\ \cos(m-1)\frac{2\pi k}{m} & \sin(m-1)\frac{2\pi k}{m} & \cdots & 1 & 0\\ -\sin(m-1)\frac{2\pi k}{m} & \cos(m-1)\frac{2\pi k}{m} & \cdots & 0 & 1\end{bmatrix} \tag{2.12}$$

with the elements corresponding to those defined in (2.10). Moreover in view of the orthogonality of the Fourier coefficients, for $k \neq k'$, the random vectors η_{nk} and $\eta_{nk'}$ converge to independent normal vectors η_k and $\eta_{k'}$. Now from (2.2) and (2.5),

$$\begin{aligned}\chi_n^2 &= \frac{m}{2\pi n} \sum_{i=1}^{m} \int_0^{2\pi} \left[n_i(\alpha) - \frac{n}{m}\right]^2 d\alpha \\ &= \frac{m}{2\pi n} \sum_{i=1}^{m} \sum_{k=1}^{\infty} (a_{ik}^2 + b_{ik}^2) \cdot \pi \\ &= \frac{m}{2} \sum_{k=1}^{\infty} (\eta'_{nk}\, \eta_{nk}) \\ &= \frac{m}{2} \sum_{k=1}^{\infty} Q_{nk}\end{aligned} \tag{2.13}$$

where $Q_{nk} = \eta'_{nk}\, \eta_{nk}$ is a quadratic form in η_{nk}. However, since η_{nk} converges in law to η_k as $n \to \infty$, the a.d. of Q_{nk} is that of $Q_k = \eta'_k\, \eta_k$ where η_k is the random vector with means zero and covariance matrix Σ_k given in (2.12). Because of this and the independence of the quadratic forms Q_{nk} and $Q_{nk'}$ for $k \neq k'$, the a.d. of

$$S_{nN} = \sum_{k=1}^{N} Q_{nk} \tag{2.14}$$

is the same as that of

$$S_N = \sum_{k=1}^{N} Q_k \tag{2.15}$$

for any finite N, i.e.,

$$S_{nN} \xrightarrow{L} S_N \tag{2.16}$$

where $\xrightarrow{L}$ denotes convergence in law. If $S_{n\infty}$ and S_∞ stand for the corresponding infinite summands of the quadratic forms, we show below that

$$S_{n\infty} \xrightarrow{L} S_\infty \tag{2.17}$$

by arguments similar to those in Beran [3].

In $F_X(.)$ denotes the distribution function of the subscripted random variable X, then for any arbitrary continuity point x of $F_{S_\infty}(x)$, we have

$$|F_{S_{n\infty}}(x) - F_{S_\infty}(x)| \leq |F_{S_{n\infty}}(x) - F_{S_{nN}}(x)| + |F_{S_{nN}}(x) - F_{S_N}(x)| + |F_{S_N}(x) - F_{S_\infty}(x)|. \tag{2.18}$$

But since

$$E(Q_{nk}) = 4m \frac{\sin^2\left(\dfrac{k\pi}{m}\right)}{k^2 \pi^2} = E(Q_k),$$

$$E|S_{n\infty} - S_{nN}| = \sum_{k=N+1}^{\infty} E(Q_{nk})$$

and

$$E|S_\infty - S_N| = \sum_{k=N+1}^{\infty} E(Q_k)$$

are the tails of convergent series. Therefore by Markov's inequality

$$(S_{n\infty} - S_{nN}) \xrightarrow{p} 0,$$
$$(S_\infty - S_N) \xrightarrow{p} 0$$

uniformly in n as $N \to \infty$. Hence for any $\varepsilon > 0$, there exists an N independent of n such that

$$\begin{aligned} |F_{S_{n\infty}}(x) - F_{S_{nN}}(x)| &< \varepsilon/3 \\ |F_{S_\infty}(x) - F_{S_N}(x)| &< \varepsilon/3. \end{aligned} \tag{2.19}$$

Now for this choice of N, we can get a n_0 such that for all $n > n_0$,

$$|F_{S_{nN}}(x) - F_{S_N}(x)| < \varepsilon/3 \tag{2.20}$$

in view of (2.16). For such an n, (2.18), (2.19) and (2.20) imply

$$|F_{S_{n\infty}}(x) - F_{S_\infty}(x)| < \varepsilon. \tag{2.21}$$

Further since the distribution function of S_∞ is continuous, by Polya's theorem, this convergence in (2.21) is uniform in x. Thus the distribution of

$$\chi_n^2 = \frac{m}{2} \sum_{k=1}^{\infty} Q_{nk}$$

converges not only weakly but uniformly to that of

$$\frac{m}{2} \sum_{k=1}^{\infty} Q_k \quad \text{as } n \to \infty.$$

Now the distribution of $Q_k = \eta_k' \eta_k$ is not difficult to obtain but notice that $\Sigma_k = ((0))$ if k is a multiple of m and for $k \neq 0$ (mod. m), it is only of rank 2 as can be seen from the fact that the third order principal minors of Σ_k vanish. Therefore η_k can be reduced to a two-dimensional random variable Y_k by means of a transformation

$$\underset{(2m\times 1)}{\eta_k} = \underset{(2m\times 2)}{B_k} \; \underset{(2\times 1)}{Y_k} \tag{2.22}$$

where B_k is such that $B_k B_k' = \Sigma_k$, B_k is of rank 2 and Y_k is distributed as $N(0, I_2)$ (see e.g. Rao [9], p. 440), I_2 standing for an identity matrix of order 2. Because of the fact that Y_k is distributed as $N(0, I_2)$, the characteristic function (cf.) of

$$Q_k = \eta_k' \eta_k = Y_k'(B_k' B_k) Y_k$$

is given by

$$\phi_k(t) = \det^{-\frac{1}{2}} |I_2 - 2it B_k' B_k|. \tag{2.23}$$

In our case, Σ_k given in (2.12), can be written as $B_k B_k'$ where

$$B_k' = \frac{\sqrt{2}\sin\dfrac{k\pi}{m}}{k\pi} \times \begin{bmatrix} 1 & 0 & \cos\dfrac{2\pi k}{m} & -\sin\dfrac{2\pi k}{m} \ldots \cos(m-1)\dfrac{2\pi k}{m} & -\sin(m-1)\dfrac{2\pi k}{m} \\ 0 & 1 & \sin\dfrac{2\pi k}{m} & \cos\dfrac{2\pi k}{m} \ldots \sin(m-1)\dfrac{2\pi k}{m} & \cos(m-1)\dfrac{2\pi k}{m} \end{bmatrix}. \tag{2.24}$$

For this B_k, since

$$B_k' B_k = \frac{2m\sin^2\dfrac{k\pi}{m}}{k^2\pi^2} I_2, \tag{2.25}$$

the cf. of Q_k, from (2.23), is

$$\begin{aligned}\phi_k(t) &= \det^{-\frac{1}{2}}\left| I_2 - 4itm\frac{\sin^2\dfrac{k\pi}{m}}{k^2\pi^2} I_2 \right| \\ &= \left(1 - \frac{4m\sin^2\dfrac{k\pi}{m}}{k^2\pi^2} it\right)^{-1}.\end{aligned} \tag{2.26}$$

Now from the independence of η_k and $\eta_{k'}$ for $k \neq k'$, the asymptotic cf. of χ_n^2 can be written down using (2.13) and (2.26) as

$$\begin{aligned}\phi(t) &= \prod_{k=1}^{\infty}\left(1 - \frac{2m^2\sin^2\dfrac{k\pi}{m}}{k^2\pi^2} it\right)^{-1} \\ &= \prod_{k=1}^{\infty}(1 - it\lambda_k)^{-1}\end{aligned} \tag{2.27}$$

with

$$\lambda_k = 2\sin^2\frac{k\pi}{m} \Big/ \left(\frac{k\pi}{m}\right)^2. \tag{2.28}$$

If k is a multiple of m, the corresponding λ_k is zero so that it does not contribute anything to the cf. of χ^2. We mentioned this earlier by saying that for such a k, $\Sigma_k = ((0))$ so that the contribution of the corresponding quadratic form Q_k is zero. From (2.27), the asymptotic distribution of the χ^2 statistic can be formally written down. If $f(x)$ denotes the density function we have by the inversion formula

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\left[\prod_k (1 - it\lambda_k)^{-1}\right] dt. \tag{2.29}$$

Since the λ_k are positive and distinct there are only simple poles for the integrand and they occur in the lower half of the complex plane at $t_k = \frac{1}{i\lambda_k}$. The integral (2.29) can then be evaluated by closing a contour in the lower half plane and $f(x)$ is given by i times the sum of the residues at the poles. Thus

$$f(x) = \sum_k c_k \exp[-x/\lambda_k] \quad \text{for } x \geqq 0 \tag{2.30}$$

where

$$c_k = [\lambda_k \prod_{k' \neq k} (1 - \lambda_{k'}/\lambda_k)]^{-1}$$

and zero otherwise. From (2.27) one can write down the μ^{th} cumulant, κ_μ of the distribution and we have

$$\kappa_\mu = (\mu-1)! \sum_{j=1}^{\infty} \lambda_j^\mu = 2^\mu (\mu-1)! \frac{m^{2\mu}}{\pi^{2\mu}} \sum_{j=1}^{\infty} \sin^{2\mu}\left(\frac{j\pi}{m}\right) \Big/ j^{2\mu}.$$

For getting the percentage points for this distribution, some numerical approximations may have to be found or use may be made of the first four cumulants to find a suitable Pearson curve approximation.

The statistic χ_n^2 given in (2.2) in the integral form can now be expressed as a finite sum and given the following computational form

$$\begin{aligned}
\chi_n^2 &= \frac{m}{2\pi n} \int_0^{2\pi} \sum_{i=1}^{m} \left[n_i(\alpha) - \frac{n}{m}\right]^2 d\alpha \\
&= \frac{m}{2\pi n} \sum_i \int_0^{2\pi} \left\{ [\sum_j \chi_j^i(\alpha)]^2 - 2\frac{n}{m} [\sum_j \chi_j^i(\alpha)] + \frac{n^2}{m^2} \right\} d\alpha \\
&= \frac{m}{2\pi n} \sum_i \int_0^{2\pi} \left\{ \left(1 - \frac{2n}{m}\right) [\sum_j \chi_j^i(\alpha)] + [\sum_{j \neq j'} \chi_j^i(\alpha) \chi_{j'}^i(\alpha)] + \frac{n^2}{m^2} \right\} d\alpha
\end{aligned} \tag{2.31}$$

since $[\chi_j^i(\alpha)]^2 \equiv \chi_j^i(\alpha)$. Now from (2.7)

$$\chi_j^i(\alpha)\, \chi_{j'}^i(\alpha) = \begin{cases} 1 & \text{if either } \alpha_{j'} - \frac{2\pi i}{m} \leqq \alpha < \alpha_j - \frac{2\pi(i-1)}{m} \\ & \quad \text{or } \alpha_j - \frac{2\pi i}{m} \leqq \alpha < \alpha_{j'} - \frac{2\pi(i-1)}{m} \\ 0 & \text{otherwise.} \end{cases} \tag{2.32}$$

Using (2.7) and (2.32) in (2.31),

$$\chi_n^2 = \frac{m}{2\pi n} \sum_i \left\{ \left(1 - \frac{2n}{m}\right) \left[\sum_j \frac{2\pi}{m}\right] + \sum_j \sum_{j'} \left(\frac{2\pi}{m} - D_{jj'}\right) + \frac{2\pi n^2}{m^2} \right\} \\ \left\{(j,j'): j \neq j',\ D_{jj'} \leqq \frac{2\pi}{m}\right\} \tag{2.33}$$

where $D_{jj'}$ denotes the "circular" distance (i.e., smaller of the two distances) between the observations α_j and $\alpha_{j'}$ on the circumference. Simplifying (2.33) further

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

$$= m - n + \frac{m^2}{2\pi n} \sum_{\substack{j \neq j' \\ \left\{(j,j'):D_{jj'} \leqq \frac{2\pi}{m}\right\}}}\sum \left(\frac{2\pi}{m} - D_{jj'}\right). \tag{2.34}$$

Thus the invariant version (2.2) of the usual χ^2 statistic, depends on the sample arc lengths between the observations. These arc lengths play a crucial role in the case of the circle as any invariant test statistic has to be a function of these values. A detailed discussion about these arc lengths can be found in Rao [10]. The statistic χ_n^2, in this form (2.34), looks like a "U-statistic" (see e.g. Fraser [4]) which it is not in the strictest sense. The usual central limit theorem for the U-statistics thus fails in our case. Finally it may be remarked that the Ajne's statistic A_n discussed in Ajne [1] and Watson [12] is a special case of our χ^2 statistic when the number of class intervals into which the circumference is divided is only two.

3. Testing Uniformity against a Specific Class of Parametric Alternatives

In this section we consider the problem of testing uniformity from a different point of view. Suppose a specific class of parametric distributions can be considered as plausible alternatives to uniformity. Such a class can often be specified at least approximately, in practice. Then we can improve on the χ^2 given in (2.1) by concentrating on these alternatives and using a χ^2 statistic suggested by Rao [8] where one compares the estimated cell frequencies under the alternatives with those under the hypothesis. Suppose it is desired to test for uniformity, against the class of unimodal symmetric densities given by

$$g(\alpha|\rho,\gamma) = \frac{1}{2\pi} + \rho\cos(\alpha-\gamma), \qquad 0 \leqq \alpha < 2\pi \tag{3.1}$$

where $0 \leqq \gamma < 2\pi$ and $\rho \geqq 0$ denote the location and concentration parameters respectively. It may be observed that circular normal densities close to the hypothesis of uniformity (i.e., with a low value for the concentration parameter) can be put in this form. Under these alternatives (3.1), the i^{th} cell has probability

$$\begin{aligned}\pi_i(\alpha) &= \int_{\alpha+(i-1)2\pi/m}^{\alpha+i2\pi/m}\left[\frac{1}{2\pi} + \rho\cos(\beta-\gamma)\right]d\beta \\ &= \frac{1}{m} + \rho\cos\gamma\left\{2\sin\frac{\pi}{m}\cos\left(\alpha + \frac{(2i-1)\pi}{m}\right)\right\} \\ &\quad + \rho\sin\gamma\left\{2\sin\frac{\pi}{m}\sin\left(\alpha + \frac{(2i-1)\pi}{m}\right)\right\} \\ &= \frac{1}{m}\{1 + \xi x_i + \eta y_i\}\end{aligned} \tag{3.2}$$

where

$$\xi=\rho\cos\gamma, \quad \eta=\rho\sin\gamma$$

$$x_i=2m\sin\frac{\pi}{m}\cos\left(\alpha+\frac{(2i-1)\pi}{m}\right),$$
$$y_i=2m\sin\frac{\pi}{m}\sin\left(\alpha+\frac{(2i-1)\pi}{m}\right). \tag{3.3}$$

Now in order that the required asymptotic distribution holds good for the statistic given in (3.9), these π_i's should be estimated by any method of estimation that is "efficient" in the sense of Rao i.e., satisfying the assumption (3) of Rao [8]. The validity of this assumption can be verified easily for the estimates $\hat{\xi}$ and $\hat{\eta}$ obtained by minimising the quantity

$$L=\sum_1^m (p_i-\pi_i)^2/\pi_i^0=\frac{1}{m}\sum_1^m (m p_i-1-\xi x_i-\eta y_i)^2. \tag{3.4}$$

On using the trigonometric relations

$$\sum_1^m \cos\frac{(2i-1)\pi}{m}=\sum_1^m \sin\frac{(2i-1)\pi}{m}=\sum_1^m \cos\frac{(2i-1)\pi}{m}\sin\frac{(2i-1)\pi}{m}=0$$
$$\sum_1^m \cos^2\frac{(2i-1)\pi}{m}=\sum_1^m \sin^2\frac{(2i-1)\pi}{m}=m/2, \tag{3.5}$$

the estimating equations simplify and we get

$$\hat{\xi}=m\sum_1^m p_i x_i\Big/\sum_1^m x_i^2=\bar{c}/m\sin\frac{\pi}{m}$$
$$\hat{\eta}=m\sum_1^m p_i y_i\Big/\sum_1^m y_i^2=\bar{s}/m\sin\frac{\pi}{m} \tag{3.6}$$

where

$$n\bar{c}=\sum_1^m n_i\cos\left(\alpha+\frac{(2i-1)\pi}{m}\right) \quad\text{and}\quad n\bar{s}=\sum_1^m n_i\sin\left(\alpha+\frac{(2i-1)\pi}{m}\right) \tag{3.7}$$

are nothing but the components of the vector resultant based on the grouped data. Thus from (3.2) and (3.6),

$$\hat{\pi}_i=\frac{1}{m}(1+\hat{\xi}x_i+\hat{\eta}y_i)$$
$$=\frac{1}{m}+(2/m)\left\{\bar{c}\cos\left(\alpha+\frac{(2i-1)\pi}{m}\right)+\bar{s}\sin\left(\alpha+\frac{(2i-1)\pi}{m}\right)\right\}. \tag{3.8}$$

Let $E_{0i}=n\pi_i^0$ and $E_{1i}=n\hat{\pi}_i$, denote respectively the estimated frequencies in the i^{th} cell, under the restrictions imposed by the hypothesis, and when there are no restrictions on the parameters. Then Rao [8] suggests the statistic

$$T=\sum_{1}^{m}(E_{1i}-E_{0i})^2/E_{0i}=n\sum_{1}^{m}(\hat{\pi}_i-\pi_i^0)^2/\pi_i^0 \tag{3.9}$$

to test the hypothesis that the data follows a uniform distribution given that the admissible set of probabilities $\pi_i=\pi_i(\rho,\gamma)$ have the representation given in (3.2). Using the relations (3.5), the statistic (3.9) reduces to

$$\begin{aligned} T&=n\,m\sum_{1}^{m}\left\{(2/m)\left[\bar{c}\cos\left(\alpha+\frac{(2i-1)\pi}{m}\right)+\bar{s}\sin\left(\alpha+\frac{(2i-1)\pi}{m}\right)\right]\right\}^2 \\ &=2n(\bar{c}^2+\bar{s}^2)=2R^2/n \end{aligned} \tag{3.10}$$

where R^2 is the squared length of the resultant based on the given grouped data. The value of this R^2, computed on the basis of the grouped data apparently depends on the particular grouping adopted. But as is well known the length of the resultant based on the ungrouped or raw data is independent of the choice of the zero direction so that the statistic (3.10), is invariant if the effect of grouping is ignored. However, since the grouping correction needed for R^2 turns out to be quite negligible even when we have only about ten to twelve class intervals (see e.g. Batschelet [2]), the statistic is almost as good as an invariant one if m is not too small.

Now, under the density (3.1), the cell probabilities π_i have a parametric representation in terms of two independent parameters ρ and γ or equivalently in terms of $\xi=\rho\cos\gamma$ and $\eta=\rho\sin\gamma$. On the other hand, the hypothesis of uniformity is equivalent to the simple hypothesis that the parameter point is the origin i.e., $\xi=0, \eta=0$. Hence the statistic T, given in (3.10), has asymptotically a χ^2 distribution with 2 degrees of freedom (ref. Rao [8]). It is interesting that the special type of χ^2 test (3.9) reduces to an analogue of the classical Rayleigh's test for grouped data on the circle.

I am grateful to Professor C. R. Rao of the Indian Statistical Institute for his encouragement and advice.

References

1. Ajne, B.: A simple test for uniformity of a circular distribution. Biometrika **55**, 343–354 (1968).
2. Batschelet, E.: Statistical methods for the analysis of problems in animal orientation and certain biological rhythms. AIBS Monograph (1965).
3. Beran, R.J.: The power of some tests for uniformity of a circular distribution. Tech. Report No. 78, Johns Hopkins University (1968).
4. Fraser, D.A.S.: Nonparametric methods in statistics. New York: John Wiley 1957.
5. Goodman, N.R.: Statistical analysis based on a certain multivariate complex Gaussian distribution. (An introduction.) Ann. math. Statistics **34**, 152–177 (1963).
6. Kuiper, N.H.: Tests concerning random points on a circle. Nederl. Akad. Wet. Proc., Ser. A **63**, 38–47 (1960).
7. Pyke, R.: Spacings. J. roy. statist. Soc., Ser. B **27**, 395–449 (1965).

8. Rao, C. R.: A study of large sample test criteria through properties of efficient estimates. Sankhyā, Ser. A **23**, 25–40 (1961).
9. – Linear statistical inference and its applications. New York: John Wiley 1965.
10. Rao, J. S.: Some contributions to the analysis of circular data. Unpublished Ph. D. thesis, Indian Statistical Institute, Calcutta (1969).
11. Watson, G. S.: Goodness of fit tests on a circle. Biometrika **48**, 109–114 (1961).
12. – Another test for uniformity of a circular distribution. Biometrika **54**, 675–677 (1967).

J. S. Rao
Department of Mathematics
Indiana University
Bloomington, Indiana 47401
USA

(Received July 28, 1970/September 20, 1971)