A NEW STATISTIC FOR THE ANALYSIS OF CIRCULAR DATA IN GAMMA-RAY ASTRONOMY

R.J. Protheroe
Department of Physics, University of Adelaide
Adelaide, South Australia 5001.

1. INTRODUCTION The analysis of directional data (see e.g., Mardia, 1972) is important in many fields of science. In particular, circular data, where each datum may be represented by a point on the circumference of a circle, often occur in astronomy. For example, the arrival of high energy photons from a pulsar recorded as a function of time may be plotted as a function of the pulsar phase and represented by points on the circumference of a circle if the period of the pulsar is known. In this case one would wish to test for evidence of pulsed emission of high energy photons. This would often be done by binning the data into a linear histogram and using conventional statistical tests such as the χ^2 test, etc., although this procedure has disadvantages for circular data. Alternatively, one may use statistical tests designed specifically for directional data. The latter include the Rayleigh test (see e.g. Mardia, 1972), Kuiper's Kolmogorov-Smirnov type test (Kuiper, 1960), Watson's Cramer-von Mises type test (Watson, 1961) and the Hodges-Ajne's test (Hodges, 1955; Ajne, 1968).

The Rayleigh test and the Hodges-Ajne's test are useful where a sinusoidal component is expected or when the excess events are expected to be concentrated over half the cycle (Ajne, 1968). Unfortunately, however, these tests are not powerful for detecting narrow peaked pulses in the presence of a uniform background. Tests for such a situation are highly desirable for testing, for example, for evidence of ultra-high energy γ-ray emission from neutron star binary X-ray sources where the duration of the γ-ray pulses is observed to be a small fraction of the orbital period (~0.007 in the case of Cygnus X-3 [Samorski and Stamm, 1983], ~0.02 in the case of Vela X-1 [Protheroe et al., 1984] and ~0.05 in the case of LMC X-4 [Protheroe and Clay, 1985]). I have recently proposed a new statistic which meets these requirements (Protheroe, 1985). Here, the test based on this statistic is outlined and its power compared with other tests.

2. PREVIOUS TESTS Leahy et al. (1983) have already shown that the Rayleigh test is not as powerful a test for the presence of narrow pulses as the χ^2 test after binning the data. In binning the data however, two decisions have to be made: (a) where to start binning (i.e. at what phase or direction); (b) how many bins to have (or what is the bin size). The optimum bin size will depend on how wide the pulse is, while ideally the bin origin should be chosen such that the pulse lies in the middle of a bin. This information is however usually not known **a priori** and O'Mongain (1973) has pointed out that re-binning the data would reduce one's confidence in any effect subsequently found. For example, Hillas (1975) has shown that if the bin origin is chosen to maximise the number of events in one bin then the significance of a high count is much less than if the bin origin had been chosen beforehand. An 'educated guess' is

usually made as to how narrow a pulse could be. No guess can be made about the best bin origin however. This is clearly unsatisfactory and a statistic which requires no binning of the data but still provides a powerful test for narrow pulses will be proposed.

3. THE PROPOSED STATISTIC The requirements of the new statistic are: (i) it should be sensitive to the distance between pairs of points on the circle, having a high value if more points are closer together than expected; (ii) it should be sensitive to grouping of several points in the same region; (iii) the sensitivity to a very small distance between an individual pair of points should not be so great that an "accidentally" close pair dominates the statistic. I will define the distance between two observations x_i and x_j as

$$\Delta_{ij} = 0.5 - \left|\{|(x_i - x_j)| - 0.5\}\right|. \qquad (1)$$

Averaging $1/\Delta$ over adjacent pairs would satisfy (i), averaging over all pairs would satisfy (ii) and averaging $(\Delta + 1/n)^{-1}$ over all pairs would satisfy (iii). I have therefore proposed the following statistic:

$$T_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\Delta_{ij} + 1/n)^{-1} \qquad (2)$$

Its null distribution has been computed using Monte Carlo methods for n ranging from 2 to 200. Crictical values for $\alpha = 0.1$, 0.05, 0.01 and 0.005 are given in Table 1 for several values of n. Coefficients of a 4th order polynomial in $\ell n(n)$ are given in Table 2 to facilitate interpolation for other values of n in this range.

The power function (e.g. Hoel, 1971) of the test for $n=50$ and $\alpha=0.05$ has been calculated by the Monte Carlo method for the alternatives to a uniform distribution in the form of a uniform distribution plus a von Mises distribution (e.g. Mardia, 1972).

$$f(x) = (1-b) + b \exp[\kappa \cos 2\pi(x-x_0))]/I_0(\kappa) \qquad (3)$$

where

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{r!^2} (\kappa/2)^{2r}, \qquad (4)$$

κ is the concentration parameter and **b** is the fraction of all events which are not uniformly distributed. The resulting contours of constant power are plotted in Fig 1. The 99% contour for the Rayleigh test has been added to Fig. 1 for comparison. Also shown are 99% contours for tests using Kuiper's V_n and Watson's U^2 statistics and the test based on the statistic Z_2^2 used by Buccheri et al. (1983) in their search for γ-ray emission from radio pulsars. Note that all the tests have very similar powers for broad pulses but that T_n is considerably more powerful than the others for narrow pulses.

4. CONCLUSION I have proposed a new statistic for the analysis of circular data. The test based on this statistic has recently been used in a search for ultra-high energy γ-ray from neutron star binary X-ray sources (Protheroe and Clay, 1985). The statistic is designed

n	α = .10	.05	.01	.005
2	1.82	1.90	1.98	1.99
3	2.24	2.40	2.68	2.78
4	2.60	2.78	3.15	3.28
5	2.91	3.09	3.49	3.65
6	3.17	3.36	3.77	3.94
7	3.40	3.59	4.01	4.18
8	3.61	3.80	4.21	4.38
9	3.80	3.98	4.39	4.56
10	3.97	4.15	4.55	4.72
12	4.27	4.44	4.82	5.00
14	4.53	4.69	5.06	5.23
16	4.75	4.92	5.27	5.44
18	4.96	5.11	5.45	5.62
20	5.14	5.29	5.62	5.78
30	5.86	5.99	6.28	6.42
40	6.38	6.50	6.76	6.89
50	6.79	6.90	7.14	7.26
60	7.12	7.23	7.46	7.57
70	7.41	7.51	7.73	7.83
80	7.66	7.75	7.96	8.05
90	7.88	7.97	8.16	8.25
100	8.07	8.16	8.35	8.43
120	8.42	8.50	8.66	8.74
130	8.57	8.64	8.80	8.88
140	8.71	8.78	8.92	9.00
160	8.96	9.03	9.15	9.22
180	9.19	9.25	9.34	9.41
200	9.39	9.44	9.51	9.58

Table 1. Critical values $T_{n,\alpha}$ for the proposed test of uniformity with the test statistic T_n (equation 2). $Pr(T_n > T_{n,\alpha}) = \alpha$. (Reproduced from Protheroe, 1985).

α	0.1	0.05	0.01	0.005
A_0	1.368	1.214	0.5523	0.3170
A_1	0.3467	0.8400	2.3416	2.807
A_2	0.4862	0.2436	−0.4819	−0.6597
A_3	−0.07407	−0.02658	0.1155	0.1418
A_4	0.004515	0.001132	−0.009005	−0.01038

Table 2. Polynomial Coefficients for interpolating $T_{n,\alpha}$. The 4th order polynomial $T_{n,\alpha} = \sum_{i=0}^{4} A_i[\ell n(n)]^i$ is valid for $2 < n < 200$. . (Reproduced from Protheroe, 1985).

specifically for situations where one requires a test of uniformity which is powerful against alternatives in which a small fraction of the observations are grouped in a small range of directions, or phases.

ACKNOWLEDGEMENTS This work has been supported in part by provision of a Queen Elizabeth II Fellowship and by a grant from the Australian Research Grants Scheme.

REFERENCES

Ajne, B., 1968: Biometrika, **55**, 343-354.
Buccheri, R. et al., 1983: Astron. Astrophysics, **128**, 245-251.
Hillas, A.M., 1975: Proc. 14th Int. Cosmic Ray Conf. (Munich), **9** 3439-3443.
Hodges, J.L., 1955: Ann. Math. Statist., **26**, 523-527.
Hoel, P.G. 1971: "Intoduction to Mathematical Statistics" 4th Edition (New York: J. Wiley and Sons).
Kuiper, N.H. 1960: Ned. Akad. Wet. Proc., **A63**, 38-47.
Leahy, D.A. et al., 1983: Astrophys. J., **272**, 256-258.
Lloyd-Evans, J. et al., 1983: Nature, **305**, 784-786.
Mardia, K.V. 1972: "Statistics of Directional Data" (London: Academic Press).
O'Mongain, E. 1973: Nature, **241**, 376-379.
Protheroe, R.J., et al. 1984: Ap.J. (Lett.), 280, L47-L50.
Protheroe, R.J. 1985: Astron. Expr. (submitted).
Protheroe, R.J. and Clay, R.W. 1985: Nature, in the press; see also these proceedings, paper OG 2.6-10.
Samorski, M. and Stamm, W. 1983a: Proc. 18th Int. Cosmic Ray Conf. (Bangalore), **1**, 135-138.
Watson, G.S. 1961: Biometrika, **48**, 109-114.

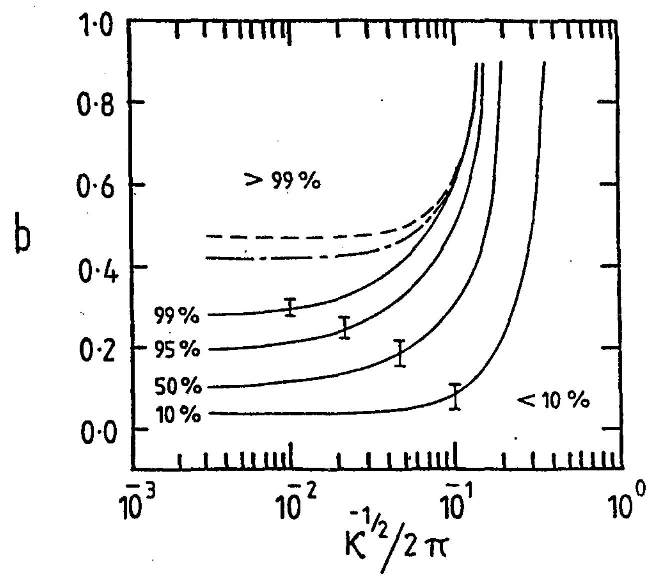

Figure 1. Contour map of the power function $P(b, \kappa)$ for $\alpha = 0.05$ for the test based on the statistic T_n. 99% contours for the Rayleigh test (dashed line) and Kuiper's test (chain line) are also given for comparison. (The 99% contours for the Z_2^2 test and for Watson's test are indistinguishable from that for Kuiper's test).