J. Statist. Comput. Simul., 2000, Vol. 67, pp. 319–332
Reprints available directly from the publisher
Photocopying permitted by license only

AN ENTROPY-BASED TEST FOR GOODNESS OF FIT OF THE von MISES DISTRIBUTION

ULRIC LUND and S. RAO JAMMALAMADAKA*

Environmental Risk Analysis, Inc. and University of California, Santa Barbara

(Received 10 March 1998; In final form 13 March 2000)

The maximum entropy characterization of the von Mises distribution on the circle is exploited to derive a consistent goodness of fit test for the von Mises distribution. Monte Carlo simulation results are tabulated giving critical values of the test statistic for various sample sizes and values of the concentration parameter. A power analysis is presented for various alternative hypotheses, comparing this entropy statistic to two other competing goodness of fit statistics. The entropy statistic is shown to compare favorably and may be more attractive, especially considering its ease of computation.

Keywords: Circular statistics; Entropy; Goodness of fit; von Mises

1. INTRODUCTION AND REVIEW

For a distribution function F, with density function f on R^1, entropy is measured by the integral

$$H(f) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx.$$

This may also be re-expressed as

$$H(f) = \int_0^1 \log\left\{\frac{d}{dp}F^{-1}(p)\right\}dp,$$

*Corresponding author. Partial support under NSF grant number DMS-9803600 is acknowledged.

for which a consistent estimate is given by Vasicek (1976):

$$H_{mn} = n^{-1} \sum_{i=1}^{n} \log\left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\}. \qquad (1.1)$$

In the above, $x_{(i)}$ indicates the ith order statistic, and $2m$ is the size of the steps used for the spacings $x_{(i+m)} - x_{(i-m)}$, with the restriction that $m < n/2$. Consistency of H_{mn}, as $n \to \infty$, $m \to \infty$ and $m/n \to 0$, is proved in *Theorem 1* of Vasicek (1976).

For density functions defined on R^1, it is well known that for fixed first and second moments, the distribution with the maximum entropy is the normal distribution (Kagan, Linnik and Rao, 1973, p. 410). A test for normality, using this characterization, was first developed by Vasicek (1976). The intuition behind the test is that the maximum entropy for a normal distribution can be calculated analytically, and thus a sample's entropy should be close to this value if the sample comes from a normal distribution. The null hypothesis of normally distributed data is rejected for small sample entropies.

For various sample sizes and spacing widths, Vasicek (1976) tabulates the critical values of the test statistic and provides some power comparisons of the entropy statistic relative to various other goodness of fit tests like the Kolmogorov-Smirnov, Cramer-von Mises, Kuiper V, Watson U^2, Anderson-Darling and Shapiro-Wilk. None of the tests performed uniformly better for all of the alternative hypotheses examined: exponential, gamma, uniform, beta, Cauchy. However, the entropy statistic had the highest power for three of these five alternatives.

2. GOODNESS OF FIT TESTING OF THE VON MISES DISTRIBUTION

A statistic to test for goodness of fit of the von Mises distribution will now be derived. It is based on the von Mises' maximum entropy characterization on the circle, for fixed mean direction and circular variance.

PROPOSITION 1 *Among all probability distributions on the circle, $f(\theta)$, $0 \le \theta < 2\pi$, with fixed mean direction μ_0 and "circular variance"*

$(1-\rho)$, *that is*

$$\int_0^{2\pi} \cos\theta f(\theta)d\theta = \rho\cos\mu_0$$
$$\int_0^{2\pi} \sin\theta f(\theta)d\theta = \rho\sin\mu_0, \tag{2.1}$$

the entropy of $f(\theta)$,

$$H(f) = -\int_0^{2\pi} f(\theta)\log f(\theta)d\theta$$

is maximized when $f(\theta)$ *is the von Mises distribution.*

A proof of this result can be found in Mardia (1972, pp. 65–66), and can also be derived quite easily *via* Theorem 13.2.1 of Kagan, Linnik and Rao (1973, p. 409).

To assess the goodness of fit of a von Mises distribution for a set of data, the sample entropy will be compared to the entropy of a von Mises distribution, which is given by

$$\begin{aligned} H(f) &= -\int_{-\infty}^{\infty} f(\theta)\log f(\theta)d\theta \\ &= -\int_{-\infty}^{\infty} \frac{\exp\{\kappa\cos(\theta-\mu)\}}{2\pi I_0(\kappa)}[-\log(2\pi I_0(\kappa)) + \kappa\cos(\theta-\mu)]d\theta \\ &= \log[2\pi I_0(\kappa)] - \kappa A(\kappa) \\ &= \log\left[\frac{2\pi I_0(\kappa)}{\exp(\kappa A(\kappa))}\right], \end{aligned} \tag{2.2}$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$, and I_0 and I_1 are the modified Bessel functions of the first kind, orders 0 and 1 respectively.

Now, applying *Theorem 1* from Vasicek (1976), under the null hypothesis that a sample of observations $x_1, \ldots, x_n$ comes from a von Mises distribution, the statistic H_{mn} given in (1.1), is a consistent estimate of $H(f)$, defined in (2.2). That is,

$$H_{mn} \xrightarrow{P} \log\left[\frac{2\pi I_0(\kappa)}{\exp(\kappa A(\kappa))}\right],$$

as $n \to \infty$, $m \to \infty$ and $m/n \to 0$. Using the fact that

$$\begin{aligned} H_{mn} &= n^{-1} \sum_{i=1}^{n} \log \left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\} \\ &= \sum_{i=1}^{n} \log \left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\}^{1/n} \\ &= \log \left\{ \frac{n}{2m} \prod_{i=1}^{n} [(x_{(i+m)} - x_{(i-m)})]^{1/n} \right\}, \end{aligned}$$

we use a consistent estimate $\hat{\kappa}$ of κ, to define

$$\begin{aligned} K_{mn} &= \frac{\exp\{H_{mn}\} \exp\{\hat{\kappa} A(\hat{\kappa})\}}{I_0(\hat{\kappa})} \\ &= \frac{n \exp\{\hat{\kappa} A(\hat{\kappa})\}}{2m I_0(\hat{\kappa})} \left[\prod_{i=1}^{n} \left(x_{(i+m)} - x_{(i-m)} \right) \right]^{1/n}. \end{aligned} \tag{2.3}$$

Then, under the null hypothesis, we have

$$\begin{aligned} K_{mn} &\xrightarrow{p} \frac{\exp\{H(f)\} \exp\{\kappa A(\kappa)\}}{I_0(\kappa)} \\ &= 2\pi. \end{aligned}$$

That is,

$$K_{mn} \xrightarrow{p} 2\pi,$$

as $n \to \infty, m \to \infty$ and $m/n \to 0$. Again, samples coming from distributions that are not von Mises, will tend to have *lower* sample entropies, and thus smaller values of K_{mn}. The null hypothesis is therefore rejected for sufficiently small values of K_{mn}. A discussion regarding the definition of spacings on the circle and the order statistics $x_{(i)}$ used in the calculation of (2.3) is presented in Section 2.1.

It is possible to derive the asymptotic distribution of H_{mn} when a simple null hypothesis is used, one in which μ and κ are specified. By taking $u_{(i)} = F_0(x_{(i)})$, and applying H_{mn} to $u_{(1)}, \ldots, u_{(n)}$, one can employ the asymptotic theory for H_{mn} developed by Dudewicz and Van der Muelen (1981) for testing uniformity. However, even for a sample size of 100, Dudewicz *et al.*, concede that the asymptotic theory could not be validated by simulation results. Additionally, as will be remarked later in Section 2.2, one advantage of the entropy statistic presented in this paper, is that the distribution function does not need to be evaluated. This advantage would be lost if the transformation $u_{(i)} = F_0(x_{(i)})$ was required.

More recently, van Es (1992) considered the asymptotic distribution of a statistic very similar to Vasicek's estimate of entropy (1.1). He shows that for a density f, defined on finite support and satisfying the Lipschitz condition, if $m, n \to \infty$ and $m = o(n^{1/2})$, the statistic defined by

$$V_{mn} = \frac{1}{n-m}\sum_{i=1}^{n-m} \log\left(\frac{n+1}{m}(X_{(i+m)} - X_{(i)})\right) + \sum_{k=2m}^{n} \frac{1}{k} + \log(2m) - \log(n)$$

has the asymptotic distribution given by

$$n^{1/2}(V_{mn} - H(f)) \xrightarrow{D} N(0, \mathrm{var}(\log(f(X)))), \qquad (2.4)$$

where $H(f)$ is the entropy of density f.

Unfortunately, even though it can be shown that the von Mises distribution is Lipschitz, simulations could not confirm the result in (2.4) for the von Mises case. Also, as might be expected, simulations show that for finite n and m the distribution of the test statistic is highly dependent on the choice of m, and there is no guide as to what step size should be used for a fixed and finite n. These two confounding factors make the asymptotic theory inapplicable in practice. For this reason, the asymptotic distribution of the entropy statistic, K_{mn} (2.3), will not be pursued further at this time.

Instead, Monte Carlo methods are used to find critical values for the statistic K_{mn}. For each of various sample sizes and values of κ, 5000 samples were generated from a von Mises distribution. From each sample, K_{mn} was calculated. A condensed table of 95th and 99th percentiles are given in Table I. The choice of step size, m, in the tabulation of critical values was according to the step size that yielded the largest test statistic. Some smaller scale simulations were performed to compare powers using various step sizes, and it was usually found that the step size yielding the largest critical values also

TABLE I Critical values for testing at the 5% (1%) significance level

	Sample size								
κ	$n=20$	$n=25$	$n=30$	$n=35$	$n=40$	$n=45$	$n=50$	$n=75$	$n=100$
0.20	4.11(3.81)	4.39(4.12)	4.57(4.30)	4.69(4.47)	4.82(4.61)	4.92(4.71)	4.99(4.83)	5.27(5.14)	5.43(5.34)
0.40	4.15(3.81)	4.40(4.12)	4.58(4.32)	4.71(4.49)	4.84(4.59)	4.95(4.77)	5.01(4.82)	5.28(5.14)	5.45(5.34)
0.60	4.17(3.83)	4.43(4.16)	4.61(4.37)	4.76(4.53)	4.85(4.65)	4.97(4.75)	5.03(4.86)	5.31(5.16)	5.46(5.36)
0.80	4.16(3.77)	4.44(4.14)	4.62(4.34)	4.78(4.52)	4.89(4.68)	4.99(4.79)	5.08(4.89)	5.33(5.19)	5.49(5.38)
1.00	4.19(3.83)	4.44(4.16)	4.66(4.36)	4.80(4.51)	4.91(4.69)	5.00(4.81)	5.08(4.90)	5.35(5.19)	5.51(5.40)
1.20	4.21(3.82)	4.44(4.13)	4.67(4.42)	4.81(4.54)	4.92(4.72)	5.00(4.76)	5.10(4.88)	5.38(5.24)	5.52(5.41)
1.40	4.23(3.92)	4.46(4.15)	4.66(4.38)	4.82(4.54)	4.93(4.69)	5.04(4.83)	5.11(4.89)	5.39(5.24)	5.54(5.43)
1.60	4.20(3.83)	4.46(4.13)	4.67(4.41)	4.82(4.54)	4.93(4.68)	5.02(4.80)	5.12(4.93)	5.40(5.27)	5.55(5.43)
1.80	4.21(3.88)	4.48(4.13)	4.65(4.40)	4.79(4.54)	4.94(4.69)	5.04(4.84)	5.12(4.91)	5.40(5.24)	5.56(5.43)
2.00	4.20(3.87)	4.44(4.13)	4.67(4.43)	4.81(4.54)	4.94(4.72)	5.02(4.77)	5.13(4.92)	5.41(5.25)	5.57(5.43)
2.20	4.22(3.86)	4.45(4.11)	4.66(4.40)	4.82(4.55)	4.93(4.69)	5.03(4.79)	5.12(4.93)	5.40(5.24)	5.57(5.44)
2.40	4.20(3.83)	4.45(4.14)	4.66(4.41)	4.82(4.58)	4.92(4.67)	5.03(4.82)	5.11(4.86)	5.41(5.25)	5.57(5.44)
2.60	4.20(3.85)	4.48(4.17)	4.65(4.39)	4.82(4.55)	4.93(4.71)	5.03(4.82)	5.12(4.93)	5.40(5.22)	5.56(5.44)
2.80	4.21(3.82)	4.45(4.14)	4.67(4.38)	4.80(4.55)	4.94(4.68)	5.03(4.83)	5.11(4.94)	5.39(5.23)	5.56(5.43)
≥ 3.00	4.21(3.85)	4.46(4.16)	4.66(4.39)	4.81(4.57)	4.92(4.69)	5.02(4.81)	5.10(4.90)	5.38(5.23)	5.54(5.42)
Step Size	$m=3$	$m=4$	$m=4$	$m=4$	$m=5$	$m=5$	$m=5$	$m=7$	$m=8$

yielded the best power. This phenomenon has also been documented in the papers by Gokhale (1983) and Dudewicz and Van der Muelen (1981).

In practice, κ is generally unknown, in which case we suggest using the maximum likelihood estimate to decide which row to use in the tabulated critical values. Computation of the MLE for κ is detailed in Section 2.1. The simulated power and significance levels in Section 2.2 are obtained *via* the MLE. In particular, it is shown that using an estimate for κ does not adversely affect the significance level of the test.

2.1. Calculation of the Test Statistic

To calculate the statistic

$$K_{mn} = \frac{n\exp\{\hat{\kappa}A(\hat{\kappa})\}}{2mI_0(\hat{\kappa})}\left[\prod_{i=1}^{n}(x_{(i+m)} - x_{(i-m)})\right]^{1/n},$$

a consistent estimate for κ is needed, and the m-step spacings are required.

The maximum likelihood estimate $\hat{\kappa}$ of κ, which is a consistent estimate, is given by

$$A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos(x_i - \hat{\mu}) = \frac{R}{n},$$

where $\hat{\mu}$ is the circular sample mean direction, and R is the length of the resultant of the n unit vectors $(\cos x_i, \sin x_i)$, $i = 1, \ldots, n$. Best and Fisher (1981) provide an approximation of $A^{-1}(x)$:

$$A^{-1}(x) = \begin{cases} 2x + x^3 + (5x^5)/6 & 0 \leq x < 0.53 \\ -0.4 + 1.39x + 0.43/(1-x) & 0.53 \leq x < 0.85 \\ 1/(x^3 - 4x^2 + 3x) & x \geq 0.85, \end{cases}$$

The calculation of $I_0(\hat{\kappa})$ can be carried out by some computer packages, but can also be done using a polynomial approximation given by Abramowitz and Stegun (1970). Letting $t = \kappa/3.75$,

$0 \leq \kappa \leq 3.75$:

$$I_0(\kappa) = 1 + 3.5156229t^2 + 3.0899424t^4 + 1.2067492t^6 + 0.2659732t^8 + 0.0360768t^{10} + 0.0045813t^{12}$$

$\kappa > 3.75$:

$$\kappa^{1/2}e^{-\kappa}I_0(\kappa) = 0.39894228 + 0.0132859t^{-1} + 0.00225319t^{-2} - 0.001157565t^{-3} + 0.00916281t^{-4} - 0.02057706t^{-5} + 0.02635537t^{-6} - 0.01647633t^{-7} + 0.00392377t^{-8}.$$

Circular statistics are ideally invariant with respect to the choice of the zero direction. This will be kept in mind when determining the m-step spacings $x_{(i+m)} - x_{(i-m)}$. There are two ways to define the spacings on the circle. One method is to utilize the circularity of the observations, and to take $x_{(i+m)} = x_{((i+m) \bmod n)} + 2\pi$ for values of $i+m$ larger than n. In the exploration of the sample entropy, H_{mn} (1.1), for von Mises data, this was the first method of spacings used, as it fully exploits the circularity of the data, and also makes H_{mn} invariant with respect to the zero location. Results, however, showed that the statistic behaved irregularly, and did not approximate the population entropy very well, especially for large values of κ.

The second method is equivalent to Vasicek's (1976) definition of the spacings for linear data, where truncation is used; that is, let

$$x_{(i+m)} = x_{(n)}, \; i+m > n \quad \text{and}$$
$$x_{(i-m)} = x_{(1)}, \; i-m < 1.$$

However, under this definition of spacings, H_{mn} will not be invariant under rotations, because the arc between and $x_{(0)}$ and $x_{(n)}$ is not being utilized, and will vary under different choices of the zero direction. The solution to this problem is to define $x_{(0)}$ and $x_{(n)}$ such that the largest gap between adjacent observations is between $x_{(0)}$ and $x_{(n)}$ (see Fig. 1). Under this definition of the order statistics, the spacings will be invariant under rotations, and hence H_{mn} and K_{mn} will be as well. In terms of sufficiency, disregarding the largest gap can be reconciled with the fact that the length of this omitted gap could be obtained by subtracting the sum of the other arc lengths from 2π.

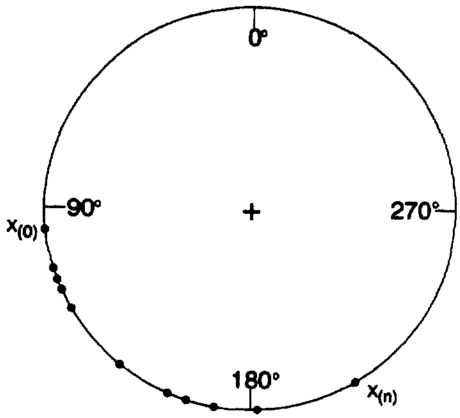

FIGURE 1 Defining the order statistics.

One final remark about the calculation of the test statistic, K_{mn}, is with regards to ties in the data set. As the statistic is a function of the product of the spacings between the observations, any ties among the observations will set the test statistic equal to zero. It is therefore essential that the data set does not contain ties, if the entropy statistic is to be used. Theoretically, as the data is assumed to be from a continuous distribution, ties occur with probability zero. Of course, in practice it is not uncommon to encounter ties due to rounding and discretization of the data. One solution is to employ fixed-variability jittering of the data, as also suggested by Minnotte, Marchette and Wegman (1998) in their work on density mode estimation.

2.2. Power Comparison of the Entropy Statistic with Existing Tests

One of the commonly used statistics for testing goodness of fit of the von Mises distribution is the U^2 statistic. Developed by Watson (1961), it is an invariant adaptation of the Cramer-von Mises statistic:

$$U^2 = n\int_0^{2\pi}\left[F_n(\theta) - F(\theta) - \int_0^{2\pi}\{F_n(\phi) - F(\phi)\}dF(\phi)\right]^2 dF(\theta).$$

Lockhart and Stephens (1985) use the asymptotic distribution of U^2 to tabulate critical values for various cases, including the case when neither the location nor shape parameters of the von Mises distribution are known. They also report that Monte Carlo studies validate their critical values for $n \geq 20$.

The sample version of Watson's statistic is given by

$$U^2 = \sum_{i=1}^{n} \left\{ z_{(i)} - \frac{2i-1}{2n} \right\}^2 - n\left(\bar{z} - \frac{1}{2}\right)^2 + \frac{1}{12n},$$

where $z_i = F(x_i; \hat{\mu}, \hat{\kappa})$. There is no closed form solution for the von Mises distribution function. Numerical methods can be used; otherwise tables are available as well. However, this can be time consuming for relatively large samples.

Bowman (1992) develops a density based, integrated squared error test statistic. Letting

$$e(x) = Ef_n(x),$$

where the expectation is taken under the null hypothesis, he defines the statistic

$$I = \int_{-\infty}^{\infty} [e(x) - f_n(x)]^2 dx.$$

Bowman (1992) proposes using the kernel density estimate f_n, with the smoothing parameter given by Parzen's (1962) formula in the case of the von Mises distribution. Critical values are obtained by Monte Carlo simulations, and are given for various sample sizes and values of κ.

For the von Mises case, Bowman (1992) defines the integrated squared error statistic as

$$I = \frac{1}{n} M(0, A^{-1}(A(h)^2) + \frac{2}{n(n-1)} \sum_{i<j} M(x_i - x_j, A^{-1}[A(h)^2])$$
$$- \frac{2}{n} \sum_{i} M(x_i - \hat{\mu}, A^{-1}[A(h)^2 A(\hat{\kappa})]) + M(0, A^{-1}[A(h)^2 A(\hat{\kappa})^2]).$$

with

$$h = A^{-1}[\exp\{(4/3)^{2/5} n^{-2/5} \log A(\hat{\kappa})\}],$$

and M is the von Mises density function. Clearly, the computation of K_{mn} is far simpler than that of I, as I requires numerous evaluations of A and A^{-1}.

Computationally, the entropy statistic, K_{mn}, is much less cumbersome than either of the two statistics above, and practitioners of statistics would find K_{mn} an attractive alternative to U^2 and I.

However, this computational edge over U^2 and I would not be meaningful if the entropy statistic did not compete in terms of power. Power comparisons between the three statistics follow. For samples of size 25, 50, 75 and 100, Bowman (1992) performs Monte Carlo power comparisons between U^2 and I. The alternative hypotheses considered are mixtures of the von Mises distribution:

1. Bimodal
   $(1/2)VM(\pi, 5) + (1/2)VM((3\pi/2), 5)$
2. Skewed
   $(2/3)VM(\pi, 3) + (1/3)VM(0.62\pi, 3)$
3. Long-tailed
   $(1/3)VM(\pi, 8) + (2/3)VM(\pi, 0.1)$
4. Half von Mises
   $\begin{cases} 2VM(\pi, 2) & \pi \le \theta \le 2\pi \\ 0 & 0 < \theta < \pi. \end{cases}$

With 1000 iterations used for each case, Bowman's (1992) results are reproduced in Table II. He shows that the integrated squared error statistic outperforms U^2 for the bimodal and skewed alternatives. For the long-tailed distribution, U^2 is far superior, while the two statistics are fairly equal for the half von Mises. Simulated power of the entropy

TABLE II Power comparison between U^2, I and K_{mn}. Powers for I and U^2 are from Bowman (1992)

	Sample size											
	$n=25$			$n=50$			$n=75$			$n=100$		
Dist.	I	U^2	$K_{4,25}$	I	U^2	$K_{5,50}$	I	U^2	$K_{7,75}$	I	U^2	$K_{8,100}$
1	0.74	0.58	0.63	0.97	0.91	0.90	1.00	0.99	0.97	1.00	1.00	0.99
2	0.14	0.10	0.14	0.22	0.16	0.23	0.27	0.24	0.28	0.47	0.29	0.32
3	0.14	0.38	0.16	0.24	0.56	0.35	0.40	0.75	0.49	0.51	0.87	0.69
4	0.43	0.45	0.74	0.82	0.81	0.98	0.97	0.93	1.00	1.00	0.99	1.00

statistic, K_{mn}, for the same alternatives used by Bowman (1992), is also given in Table II. 5000 iterations have been used in this simulation. Under the bimodal alternative, K_{mn} beats U^2 for small samples, $n=25$, and its power is only slightly lower than that of U^2 and I for the other sample sizes. For the skewed alternative, K_{mn} is uniformly better than U^2, and again quite close to I. U^2 has a clear advantage under the long-tailed alternative, however, K_{mn} has uniformly better power than I. Under the half von Mises alternative, it is interesting to note that although I is based on kernel methods, and thus smooths over the sharp edge of the half von Mises distribution, it still outperformed U^2. However, K_{mn} outperformed both I and U^2, most decisively for the smaller sample sizes.

In addition to comparing K_{mn} to the results of Bowman (1992), the entropy statistic was compared to U^2 for two other alternatives, which closely resemble the von Mises distribution: the cardoid and triangular distributions. The results are given in Table III. The entropy statistic clearly has higher power for both alternatives and all sample sizes. Bowman's (1992) I statistic was not included in this comparison since it is not commonly used.

Bowman (1992) comments that an entropy statistic for testing goodness of fit of the von Mises distribution changed too quickly as a function of κ, creating difficulty in the control of the level of the test. Table IV shows the simulated level of K_{mn}, generating data from a von

TABLE III Power comparison between U^2 and K_{mn}, based on 5000 iterations

	Sample size							
	$n=25$		$n=50$		$n=75$		$n=100$	
Distribution	$K_{4,25}$	U^2	$K_{5,50}$	U^2	$K_{7,75}$	U^2	$K_{8,100}$	U^2
1. Cardoid	0.11	0.08	0.15	0.11	0.21	0.14	0.24	0.18
2. Triangular	0.09	0.06	0.12	0.08	0.16	0.08	0.22	0.10

TABLE IV Simulated levels of K_{mn}, based on 5000 iterations ($\alpha=0.05$)

	Sample size			
κ	$n=25$	$n=50$	$n=75$	$n=100$
1	0.0496	0.0524	0.0472	0.0498
3	0.0526	0.0560	0.0546	0.0472
5	0.0532	0.0518	0.0468	0.0538

Mises distribution with various sample sizes and values for κ. The control of type I error does not seem to be a problem in our work, as the simulated levels are all quite close to the desired 0.05.

2.3. Concluding Remarks

The maximum entropy characterization of the von Mises distribution, as proposed in Section 2, is over the class of all probability distributions on the circle with fixed mean direction and circular variance. Since the entropy statistic, K_{mn}, is based on this characterization, the statistic should not be used to distinguish between the von Mises distribution and a distribution outside of this class of distributions. In particular, the uniform distribution on $[0, 2\pi)$ has an undefined mean direction, and thus the entropy statistic should not be used when the alternative hypothesis is uniformity. Instead, to test between the uniform and von Mises distributions, the Rayleigh test is recommended (see Mardia, 1972, p. 133 and Fisher, 1993, p. 82).

Acknowledgements

The authors wish to thank a referee for several helpful comments improving the focus of the paper.

References

Abramowitz, M. and Stegun, I. A. (1970) *Handbook of Mathematical Functions*, New York: National Bureau of Standards.

Best, D. J. and Fisher, N. I. (1981) The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters, *Communications in Statistics – Simulations and Computations*, **B10**(5), 493 – 502.

Bowman, A. W. (1992) "Density based tests for goodness-of-fit", *Journal of Statistical Computation and Simulation*, **40**, 1 – 13.

Dudewicz, E. J. and Van der Meulen, E. C. (1981) "Entropy-based tests of uniformity", *Journal of the American Statistical Association*, **76**, 967 – 974.

Fisher, N. I. (1993) *Statistical Analysis of Circular Data*, Cambridge: Cambridge University Press.

Gokhale, D. V. (1983) "On entropy-based goodness-of-fit tests", *Computational Statistics and Data Analysis*, **1**, 157 – 165.

Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973) Characterization Problems in Mathematical Statistics, New York: John Wiley and Sons.

Lockhart, R. A. and Stephens, M. A. (1985) "Tests of fit for the von Mises distribution", *Biometrika*, **72**, 647 – 652.

Mardia, K. V. (1972) *Statistics of Directional Data*, London: Academic Press.
Minnotte, M. C., Marchette, D. J. and Wegman, E. J. (1998) "The Bumpy Road to the Mode Forest", *Journal of Computational and Graphical Statistics*, **7**, 239–251.
Van Es, B. (1992) "Estimating Functionals Related to a Density by a Class of Statistics Based on Spacings", *Scandinavian Journal of Statistics*, **19**, 61–72.
Vasicek, O. (1976) "A test of normality based on sample entropy", *Journal of the Royal Statistical Society Series B*, **38**, 54–59.
Watson, G. S. (1961) "Goodness-of-fit tests on a circle", *Biometrika*, **48**, 109–114.