Sankhyā : The Indian Journal of Statistics
2023, Volume 85-A, Part 1, pp. 1-42

Families of Discrete Circular Distributions with Some Novel Applications

Kanti V. Mardia
University of Leeds, Leeds, UK
University of Oxford, Oxford, UK
Karthik Sriram
Indian Institute of Management Ahmedabad, Ahmedabad, India

Abstract

We give a unified treatment of constructing families of circular discrete distributions. Some of these families are deduced from established distributions such as von Mises and wrapped Cauchy. Some others are derived directly such as a flexible family based on trigonometric sums and the circular location family. Results interrelating these families are discussed. These distributions have been motivated by two examples of discrete circular data: casino roulette spins and smart health acrophase monitoring, and these data are analyzed using our proposed models. We discuss how using continuous circular models for circular discrete data can be misleading.

AMS (2000) *subject classification.* Primary 62H15, 62R30; Secondary 62E10, 62F03.
Keywords and phrases. Conditionalized families, marginalized families, roulette spin data, acrophase data, von Mises distribution, wrapped Cauchy distribution.

1 Introduction

The subject of Directional Statistics has grown tremendously, especially since the 1980's, with advances in "Statistics on Manifolds" leading to new distributions on the hyper-sphere, torus, Stiefel manifold, Grassmann manifold and so on. The progress in this area can be seen through several books

An early version of this work was presented at the Leeds Annual Statistics Research (LASR) Workshop 2019.
Supplementary Information The online version of this article (https://doi.org/10.1007/s13171-022-00298-z) contains supplementary material, which is available to authorized users.

published since then: Fisher et al. (1987), Fisher (1993), Mardia and Jupp (2000), Jammalamadaka and Sengupta (2001), Ley and Verdebout (2017) and Ley and Verdebout (2018). There has been a recent special issue of Sankhya edited by Bharath and Dey (2019). Further, Pewsey and García-Portugués (2021) have given a comprehensive survey of directional statistics and in the discussion to the paper, Mardia (2021) has given a brief history of the subject. However, there is limited development of circular discrete models. There are good choices for continuous models for circular data, but there has been a dearth of models for discrete data. In this paper, we give the first unified treatment of constructing families of circular discrete distributions and present examples of circular data that are observed directly as discrete rather than created by grouping continuous data. The two data that motivated our paper are:

(i) **Roulette wheel data:** A typical European roulette has 37 discrete outcomes, viz. $\{0, 1, 2, \ldots, 36\}$. If the outcome 0 is mapped to 0 radians, then the outcomes get mapped to a regular support of 37 points on the circle given by

$$\left\{ \frac{2\pi r}{37},\ r \in \{0, 1, 2, \ldots, 36\} \right\}. \tag{1.1}$$

In Section 4.1, we consider data sequences obtained from spins of four different European roulette wheels, one from an online roulette simulator and three from two different casino industries. It should be pointed out that Karl Pearson, in the early 1890's, acquired roulette spins data from Monte Carlo to examine the question of whether the roulette wheel was unbiased (see, Plackett, 1983, p.60), and indeed his paper of 1897 has the apt title "The scientific aspect of Monte Carlo roulette" (Pearson, 1897). Few authors have considered this problem but have used linearized methods, beginning with Karl Pearson and subsequently some others, e.g. Ethier (1982), Spencer (2009). Surprisingly, for this important application from the gaming industry, there has been little attention to inference that uses explicitly the circularity of the wheel.

(ii) **Acrophase data:** In non-invasive smart health monitoring, parameters such as Systolic Blood Pressure (SBP) are recorded, by ambulatory devices, at predetermined discrete time points repeated each day. "Acrophase" is defined as the time point at which the maximum SBP reading is recorded on a given day. Typically, acrophase data is extracted from SBP measurements at each half hour during daytime (8 am to 8 pm) and each hour during nighttime (8 pm to 8 am). If we map 8 am to 0 radians and 8 pm to π radians, the acrophase times get mapped to an *irregular support* of 36 points on the circle given by

$$\left\{\frac{2\pi r}{48}, r = 0, 1, 2, \ldots, 24\right\} \bigcup \left\{\frac{2\pi r}{48}, r = 26, 28, \ldots, 44, 46\right\}, \quad (1.2)$$

where the first set in the union corresponds to 25 half-hourly points during daytime and the second set corresponds to the 11 one-hourly points during nighttime.

In any application with discrete data in Linear Statistics, one usually takes into account the discrete nature of the underlying population. Our overall recommendation is the same here: "if one has discrete circular data then one should start with a discrete circular model". Also, the "loss" due to use of a continuous model for circular discrete data can only be assessed after appropriate discrete modeling, which serves as a benchmark. Of course, this issue of discrete versus continuous distributions is a general problem, which is well known and has been dealt successfully in Linear Statistics and we treat this problem here as a model misspecification problem (see, Section 5).

In this paper, we give four methods to construct families of discrete distributions on the circle along with some basic results interrelating the methods. We apply these models to analyze the aforementioned examples of discrete data and also to provide insights based on comparisons among discrete as well as (approximate) continuous models for discrete data. Our methods to construct the probability distributions can be briefly described as follows:

(i) Maximum entropy method: We start with a given set of moment conditions for the discrete distribution on the circle. We then determine the discrete probability distribution with the maximum Shannon entropy among those satisfying the moment constraints.

(ii) Centered wrapping method: We start with a given discrete distribution on the line, and wrap it on the circle to obtain a discrete distribution on the circle.

(iii) Marginalized method: We start with a continuous distribution on the circle, which we refer to as the "parent", and then obtain a discrete distribution on the circle by integrating the probability density function (pdf) on pre-determined arcs on the circle.

(iv) Conditionalized method: We start with a continuous distribution on the circle (parent), and then obtain the discrete distribution on the circle by restricting and normalizing the pdf to a pre-determined lattice on the circle.

In particular, we derive circular discrete distributions from general continuous location families and a family based on trigonometric sums. Key special cases include discrete families deduced from two established continuous distributions, viz. von Mises and wrapped Cauchy. These two distributions are commonly selected for circular data depending on whether the unimodal data has a long tail (wrapped Cauchy) or not (von Mises), which we now describe.

The direction of a unit random vector in two dimensions can be represented by an angle Θ. On the circle, the von Mises distribution for Θ (see, for example, Mardia and Jupp, 2000, p.36) plays the same role as the normal distribution on the line. It belongs to the exponential family with two analogous parameters. Its pdf is given by

$$f_v(\theta|\kappa,\mu) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu)}, \theta \in [0,2\pi), \mu \in [0,2\pi), \kappa \geq 0, \tag{1.3}$$

where μ is the mean direction and κ is the concentration (precision) parameter. The normalization constant $I_0(\kappa)$ is the modified Bessel function of order 0:

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{\kappa^{2r}}{(r!)^2}.$$

For large κ, Θ is approximately normal with mean μ and variance $2/\kappa$, and for $\kappa = 0$, Θ is uniformly distributed on the circle.

Given a distribution on the line, we can wrap it around the circumference of the circle with unit radius. If X is the random variable on the line, the random variable Θ of the wrapped distribution is given by

$$\Theta = (X \bmod 2\pi).$$

A popular example of wrapped distributions is the wrapped Cauchy distribution with its pdf (see, for example, Mardia and Jupp, 2000, p.51)

$$f_c(\theta|\rho,\mu) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \theta \in [0,2\pi), \mu \in [0,2\pi), \rho \in [0,1), \tag{1.4}$$

where μ is the mean direction parameter and ρ is the concentration parameter. It is one of the wrapped distributions whose density has a closed form and is heavy-tailed. When $\rho = 0$ it also reduces to the uniform distribution.

In what follows, Section 2 gives constructions of discrete circular families based on the four methods, along with some examples and results interrelating them. In particular, we deduce families from general continuous location families, and a flexible family based on trigonometric sums. Section 3 deduces some discrete families from established distributions such as von Mises and wrapped Cauchy. We apply some of the models to our discrete data in Section 4. In Section 5, we treat the problem of model misspecification for circular discrete distributions. We conclude the paper with a discussion in Section 6. Some supporting material is given in the supplement, including a table of abbreviations used in the paper.

2 Constructions of Families of Circular Discrete Distributions

In this section, we elaborate on the four different methods to construct families of circular discrete distributions that were mentioned in Section 1. Although our ideas naturally extend to constructing discrete distributions on an irregular support (such as in (1.2)), we will focus here on the regular lattice support (as in (1.1)) which lends itself to some mathematical simplifications.

We denote the set of real numbers by $\mathbb{R}$, non-negative real numbers by $\mathbb{R}^+$, the set of integers by $\mathbb{Z}$, non-negative integers by $\mathbb{Z}^+$ and the cyclic group of integers modulo a given positive integer m by $\mathbb{Z}_m$, i.e.

$$\mathbb{Z}_m = \{0, 1, \ldots, m-1\}. \tag{2.1}$$

The regular circular lattice domain is given by the vertices of a regular polygon on the circle, denoted by $\mathcal{D}_m$, i.e.

$$\mathcal{D}_m = \{2\pi r/m, r \in \mathbb{Z}_m\}. \tag{2.2}$$

We generally use $f(\cdot)$ or $g(\cdot)$ to denote a probability density function (pdf) of a continuous distribution on the line or circle, and $p(\cdot)$ to denote a discrete probability function on $\mathbb{Z}_m$.

2.1. Maximum Entropy Discrete Circular Distributions For a probability function $\{p(r), \quad r \in \mathbb{Z}_m\}$, with $p(r)$ denoting the probability of the point $2\pi r/m \in \mathcal{D}_m$, Shannon's entropy is defined as

$$-\sum_{r=0}^{m-1} p(r) \log p(r). \tag{2.3}$$

Let $t_1, t_2, \ldots, t_q$ be q real valued functions defined on $\mathbb{Z}_m$ and suppose we are interested in discrete distributions $\{p(r), \quad r \in \mathbb{Z}_m\}$ that satisfy a set of pre-selected moment conditions

$$\sum_{r=0}^{m-1} p(r)t_1(r) = a_1, \quad \sum_{r=0}^{m-1} p(r)t_2(r) = a_2, \ldots, \sum_{r=0}^{m-1} p(r)t_q(r) = a_q, \qquad (2.4)$$

with given constants $a_1, a_2, \ldots, a_q$. Then, a useful method to construct discrete distributions is to maximize the entropy among all distributions on $\mathbb{Z}_m$ that satisfy the given conditions. As noted by Kemp (1997), the philosophy behind this construction is that "one should use all the given information and nothing else". The following theorem gives this construction, which follows from Theorem 13.2.1 of Kagan et al. (1973, pp. 408–409) on the line and was adapted in Mardia (1975a) for directional distributions.

Theorem 1 (Maximum Entropy Distributions). *The probability function* $\{p(r), \quad r \in \mathbb{Z}_m\}$ *that maximizes the entropy* (2.3) *subject to the constraints* (2.4) *is of the form*

$$p(r) \quad = \quad \frac{e^{\sum_{j=1}^{q} b_j t_j(r)}}{\sum_{k=0}^{m-1} e^{\sum_{j=1}^{q} b_j t_j(k)}}, \quad r \in \mathbb{Z}_m, \qquad (2.5)$$

provided there exist constants $b_1, b_2, \ldots, b_q$ *satisfying*

$$\frac{\sum_{r=0}^{m-1} t_j(r) e^{\sum_{i=1}^{q} b_i t_i(r)}}{\sum_{k=0}^{m-1} e^{\sum_{i=1}^{q} b_i t_i(k)}} = a_j, \quad j = 1, 2, \ldots, q. \qquad (2.6)$$

In that case, the distribution is unique.

We now give a few examples of maximum entropy discrete distributions.

Example 1. von Mises distribution: Suppose $q = 2$ and $t_1(r) = \cos(2\pi r/m)$ and $t_2(r) = \sin(2\pi r/m)$, a discrete version of the von Mises distribution is of the form

$$p(r) \quad = \quad \frac{e^{\kappa \cos(2\pi r/m - \mu)}}{\sum_{k=0}^{m-1} e^{\kappa \cos(2\pi r/m - \mu)}}, \quad r \in \mathbb{Z}_m, \qquad (2.7)$$

where $\kappa = \sqrt{b_1^2 + b_2^2}$ and $\tan(\mu) = b_2/b_1$.
We note that this also happens to be the conditionalized discrete von Mises distribution that is discussed in more detail later.

Example 2. Beran distributions: A more general family than the previous example, a discrete version of the Beran family (Beran, 1979), is obtained by considering constraints on the expected values of $t_j(r) = (\cos(2\pi rj/m),\ \sin(2\pi rj/m))$, which leads to the probability function

$$p(r) \propto e^{\sum_{j=1}^{q}(a_j \cos(2\pi rj/m)+b_k \sin(2\pi rj/m))}, \quad r \in \mathbb{Z}_m. \tag{2.8}$$

We will denote this distribution by $\mathcal{B}_q$, where q is the order of the distribution. So, $\mathcal{B}_1$ is von Mises discrete distribution as in the previous example. $\mathcal{B}_2$ is the discrete generalized von Mises distribution.

We note that this family can be traced back to Maksimov (1967), although his focus for this family is on a characterization for the unknown centering parameter (rather than the concentration parameter), so it is of limited practical importance.

EXAMPLE 3. Geometric distribution: Suppose $q = 1$ and $t_1(r) = r$ the maximum entropy distribution is of the form

$$p(r) \quad = \quad \frac{(1-p)p^r}{1-p^m}, \quad r \in \mathbb{Z}_m, \text{ where } p = e^{b_1}. \tag{2.9}$$

Historically, Mardia (1972, p. 50) proposed the above distribution as a model for roulette outcomes (possibly biased).

The above three examples also arise out of the "conditionalized" construction of discrete circular distributions that we define below in Section 2.3. We note that (2.9) is also the "centered wrapped geometric distribution" discussed below in the next subsection.

2.2. Centered Wrapped Discrete Circular Distributions A natural construction to obtain a circular discrete distribution is to start with a discrete distribution on the line and wrap it on the circle (see, for example, Mardia, 1972, p.50). Let Z be a random variable taking integer values (i.e. in $\mathbb{Z}$) with probability function $p_0(\cdot)$. For a given positive integer m, we define here the wrapped discrete random variable:

$$Z_w = (Z \bmod m) \times (2\pi/m).$$

We note that $Z_w \in \mathcal{D}_m$ and its probability function is given by

$$p_{w0}(r) = P(Z_w = 2\pi r/m) = \sum_{k=-\infty}^{\infty} p_0(r+km), \quad r \in \mathbb{Z}_m. \tag{2.10}$$

It follows that the characteristic function of Z_w is given by

$$\psi_{p,m} = E\left(e^{ipZ_w}\right) = \phi(2\pi p/m), \tag{2.11}$$

where $\phi(\cdot)$ is the characteristic function of Z. In general, these distributions do not have a mean direction or centering parameter, and therefore we construct the "centered wrapped" probability function with a centering parameter t as follows,

$$p(r) = \begin{cases} p_{w0}(r-t+m), & r < t \\ p_{w0}(r-t), & r \geq t, \end{cases} \qquad r, t \in \mathbb{Z}_m. \tag{2.12}$$

Choosing the domain of t as $\mathbb{Z}_m$ ensures probabilities are well defined without changing the domain of the distribution.

Example 4. Centered wrapped Poisson distribution: For the Poisson distribution with mean λ, the wrapped Poisson distribution has the probability function

$$p_{w0}(r) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r+km}}{(r+km)!}, \ r \in \mathbb{Z}_m. \tag{2.13}$$

The centered wrapped probability function with centering parameter t is then given by

$$p(r) = \begin{cases} e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r-t+m+km}}{(r-t+m+km)!}, & r < t \\ e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r-t+km}}{(r-t+km)!} \quad , & r \geq t \end{cases} \qquad r, t \in \mathbb{Z}_m. \tag{2.14}$$

The above expression is a special case of the distributions considered by Mastrantonio et al. (2019).

For practical applications with continuous circular data, it is well known that the selected probability density is continuous at 2π, i.e. the pdf value at 0 is same as its limiting value at 2π. Similarly, along the same lines, a desirable property for a circular discrete probability function $p(\cdot)$ on $\mathbb{Z}_m$ is to have $p(0) = p(m)$. The maximum entropy and the centered wrapping methods do not necessarily ensure this property as is apparent from Examples 3 and 4. However, this property is ensured if we construct discrete distributions by applying the marginalized and conditionalized methods on continuous circular distributions, which we will discuss next.

2.3. Marginalized and Conditionalized Discrete Distributions In this section, we focus on univariate circular constructions based on marginalized and conditionalized methods, whose brief descriptions were given in Section 1.

There has been literature on the marginalized and conditionalized discretization on the line. For example Kemp (1997) and Szabłowski (2001) discuss the conditionalized discrete normal, Inusaha and Kozubowski (2006) discuss the conditionalized discrete Laplace, and Papadatos (2018) derives the characteristic function of the conditionalized discrete Cauchy. The conditionalized approach can also be described as a "plug-in" approach, whereas the marginalized approach is in fact equivalent to the well known probit construction, usually used for univariate and multivariate normal distributions, see for example Joe (2014, p.20). Alzaatreh et al. (2012) and Chakraborty (2015) make references to both of these methods, while developing other methods of constructions. However, there have not been any insights relating these constructions.

Marginalized and conditionalized constructions of circular families of discrete distributions are very recent as proposed in Mardia and Sriram (2020). Besides, there has not been published in-depth analysis of truly discrete circular data. However, particular cases of the conditionalized approach, including von Mises and wrapped Cauchy distributions, have appeared, not only in Mardia and Sriram (2020) but also in Girija et al. (2019) and Imoto et al. (2020). We give a unified treatment of the different methods as a strategy to construct rich classes of discrete distributions on the circle. We derive new results (e.g. Theorems 1 to 4) that offer insights on the inter-relationships between the constructions.

We now define the marginalized and conditionalized discrete families for the circular case. There has been some very recent work on these approaches although not in a comprehensive and unified way, as we describe below. Let Θ be a random variable with pdf $f(\theta)$, $\theta \in [0, 2\pi)$.

DEFINITION 1. The probability function of the marginalized discrete (MD) distribution on the circle is given by

$$p(r) = \int_{\frac{2\pi r}{m}}^{\frac{2\pi(r+1)}{m}} f(\theta)d\theta = F\left(\frac{2\pi(r+1)}{m}\right) - F\left(\frac{2\pi r}{m}\right),\ r \in \mathbb{Z}_m, \quad (2.15)$$

where $F(\cdot)$ is the cumulative distribution function of the pdf $f(\cdot)$.

We note that this is also the probability function of the discrete random variable $\lfloor \frac{m\Theta}{2\pi} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the largest integer less than or equal to the given number.

DEFINITION 2. The probability function of the conditionalized discrete (CD) distribution on the circle is given by

$$\mathrm{p}\,(r) = \frac{f(2\pi r/m)}{\sum_{k=0}^{m-1} f(2\pi k/m)}, r \in \mathbb{Z}_m. \quad (2.16)$$

For simplicity, we will denote both probability functions (2.15) and (2.16) by the same notation, but the choice will be obvious from the context. One question of interest is "can the marginalized and conditionalized methods lead to the same discrete distribution on the circle?" We show that under certain conditions, the two approaches will lead to diffent discrete distributions on the circle except for the trivial uniform case. This has implication when we come to selecting between the two approaches in practice and we give insights based on comparison of the two approaches for some particular cases in Section 5. Theorem 2 below gives one characterization. This theorem is inspired by a similar question for the linear case related to the exponential distribution (see supplement Section A7).

THEOREM 2. *Suppose $f(\cdot)$ is a strictly positive and continuous circular pdf on $[0, 2\pi]$ with $f(\theta) = f(\theta+2\pi)$. Then, the two discretization approaches (i.e. MD and CD) lead to the same discrete distribution, with*

$$\frac{f(a+k\delta)}{\sum_{r=0}^{m-1} f(a+r\delta)} = \frac{\int_{a+k\delta}^{a+(k+1)\delta} f(\theta)d\theta}{\int_a^{a+m\delta} f(\theta)d\theta}, \forall, a, \delta \in [0, 2\pi), \quad \forall\ k \in \mathbb{Z}_m \text{ with } m\delta \leq 2\pi, \tag{2.17}$$

iff f is the uniform density.

PROOF. *Considering (2.17) for $k = 1$ and $k = 0$, and taking their ratio, we get*

$$f(a+\delta) \int_a^{a+\delta} f(\theta)d\theta = f(a) \int_{a+\delta}^{a+2\delta} f(\theta)d\theta. \tag{2.18}$$

Integrating both the left hand side (lhs) and right hand side (rhs) of (2.18) with respect to $\delta \in [0, 2\pi)$, we get

$$\int_0^{2\pi} f(a+\delta) \int_a^{a+\delta} f(\theta)d\theta d\delta = f(a) \int_0^{2\pi} \int_{a+\delta}^{a+2\delta} f(\theta)d\theta d\delta. \tag{2.19}$$

Using continuity of $f(\cdot)$ including $f(\theta+2\pi) = f(\theta)$, we have $\int_a^{a+2\pi} f(\theta)d\theta =$ 1 for any a, and the lhs of (2.19) can be simplified as

$$\int_0^{2\pi} f(a+\delta) \int_a^{a+\delta} f(\theta)d\theta d\delta = \frac{1}{2} \left(\int_a^{a+\delta} f(\theta)d\theta \right)^2 |_{\delta=0}^{2\pi} = 1/2. \tag{2.20}$$

Now, the rhs of (2.19) can be shown to be

$$f(a) \int_0^{2\pi} \int_{a+\delta}^{a+2\delta} f(\theta)d\theta d\delta = f(a)(A - B) = f(a)\pi, \tag{2.21}$$

Equation 2.21 follows because A and B can be simplified as below.

$$\begin{aligned} A &= \int_0^{2\pi}\int_a^{a+2\delta} f(\theta)d\theta d\delta = \delta\int_a^{a+2\delta} f(\theta)d\theta|_0^{2\pi} - \int_0^{2\pi} 2\delta f(a+2\delta)d\delta \\ &= 2\pi\int_a^{a+4\pi} f(\theta)d\theta - \frac{1}{2}\int_0^{4\pi}\delta' f(a+\delta')d\delta' = 4\pi - \left(\pi + \int_0^{2\pi}\delta' f(a+\delta')d\delta'\right), \end{aligned}$$

and

$$B = \int_0^{2\pi}\int_a^{a+\delta} f(\theta)d\theta d\delta = \delta\int_a^{a+\delta} f(\theta)|_0^{2\pi} - \int_0^{2\pi}\delta f(a+\delta)d\delta = 2\pi - \int_0^{2\pi}\delta f(a+\delta)d\delta.$$

Equating the lhs (2.20) and rhs (2.21), we get $f(a) = \frac{1}{2\pi}$. Since a is arbitrary, this means that $f(\cdot)$ must be the uniform pdf on the circle.

It is to be emphasized that the assumption on continuity of the pdf $f(\cdot)$ is crucial in the above theorem. For example, the marginalized and conditionalized methods applied to the wrapped exponential distribution on the circle lead to the same discrete distribution, i.e. geometric distribution. However, the wrapped exponential pdf is not continuous at $\theta = 2\pi$.

A related question is whether the marginalized and conditionalized methods can belong to the same family of distributions. Indeed, it is easy to see that this property will hold for a generalized Cardioid-type family of distributions as given by the following theorem.

THEOREM 3. *Consider the pdf of the parent family defined by*

$$f(\theta) = \frac{1}{2\pi}\sum_{k=1}^{\infty}\eta_k\left(1 + 2\rho_k\cos(\theta - \mu_k)\right), \quad \theta \in [0, 2\pi), \tag{2.22}$$

where $\sum_{k=1}^{\infty}\eta_k = 1$, $\forall\ k$, $\eta_k \geq 0$, $\mu_k \in [0, 2\pi)$, $|\rho_k| < 1/2$. Then the marginalized discrete distribution is also a member of the family of conditionalized discrete distributions.

An interesting connection between the constructions on the circle and line, is given by the following theorem.

THEOREM 4 (Duality). *Consider the following dual approaches to constructing discrete circular distributions supported on $\mathbb{Z}_m$, starting with a real valued random variable X, with a pdf $f(\cdot)$ on $\mathbb{R}$, via either the marginalized or the conditionalized methods of discretization.*

(a) ***Scale, discretize and wrap:*** *Start with the pdf of the scaled random variable $\widetilde{X} = mX/(2\pi)$, obtain the marginalized [or conditionalized]*

discrete probability function on the line and denote the corresponding random variable by $\widetilde{X}_d$. Further, wrap $\tilde{X}_d$, i.e. let $\widetilde{X}_{dw} = (\widetilde{X}_d \mod m)$.

(b) ***Wrap, scale and discretize:*** *Let $X_w = (X \mod 2\pi)$ (i.e. X wrapped on the circle). Now, start with the pdf of the scaled random variable $\widetilde{X}_w = mX_w/(2\pi)$, obtain the marginalized [or conditionalized] discrete probability function on the circle and denote the corresponding random variable by $\widetilde{X}_{wd}$.*

Then, $\widetilde{X}_{dw}$ and $\widetilde{X}_{wd}$ have the same distribution.

PROOF. *For a given random variable Z with continuous support, recall that discretization by marginalized method means taking the random variable $\lfloor Z \rfloor$, and discretization by conditionalized method means taking the random variable Z_d with its probability function defined by $P(Z_d = r) = \frac{f(r)}{\sum_k f(k)}$. First, we will prove the equivalence for the conditionalized discretization method. The probability function of the discrete distribution resulting from process (a) is given by*

$$P(\widetilde{X}_{dw} = r) = \sum_{k\in\mathbb{Z}} P(\widetilde{X}_d = r + km) = \frac{\sum_{k\in\mathbb{Z}} f(2\pi(r + km)/m)}{\sum_{k\in\mathbb{Z}} f(2\pi k/m)}. \tag{2.23}$$

For the process in (b), let us denote the pdf of X_w by f_{x_w}. Then, the probability function of the discrete distribution resulting from process (b) will be

$$P(\widetilde{X}_{wd} = r) = \frac{f_{x_w}(2\pi r/m)}{\sum_{r=0}^{m-1} f_{x_w}(2\pi r/m)} = \frac{\sum_{k\in\mathbb{Z}} f(2\pi r/m + 2\pi k)}{\sum_{k\in\mathbb{Z}} f(2\pi k/m)}. \tag{2.24}$$

Since (2.23) and (2.24) are the same, (a) and (b) yield the same discrete circular distribution. Now, to prove the equivalence under the marginalized method of discretization, we observe that process (a) leads to the probability function given by

$$P(\widetilde{X}_{dw} = r) = \sum_{k\in\mathbb{Z}} P(\widetilde{X}_d = r + km) = \sum_{k\in\mathbb{Z}} \left(F\left(\frac{2\pi(r+1)}{m} + 2\pi k\right) - F\left(\frac{2\pi r}{m} + 2\pi k\right)\right),$$

but process (b) also leads to the same probability function because

$$\begin{aligned} P(\widetilde{X}_{wd} = r) &= P\left(X_w \in \left[\frac{2\pi r}{m}, \frac{2\pi(r+1)}{m}\right]\right) \\ &= \sum_{k\in\mathbb{Z}} \left(F\left(\frac{2\pi(r+1)}{m} + 2\pi k\right) - F\left(\frac{2\pi r}{m} + 2\pi k\right)\right). \end{aligned}$$

2.3.1. General Circular Discrete Location Family. Consider a general circular location family (see, for example, Mardia, 1975b) with probability density function given by

$$f(\theta|\tau,\mu) = g_\tau(\theta-\mu), \quad \theta,\mu \in [0,2\pi), \ \tau \geq 0, \tag{2.25}$$

which we assume to be unimodal with mode at μ. For simplicity, we assume $g_\tau(\theta) = g_\tau(2\pi-\theta)$, $g_\tau(\theta) > 0$ for all $\theta \in [0,2\pi)$ and also that $g_\tau(2\pi) = g_\tau(0)$. Note that the normalizing constant will depend only on τ and not on μ. Here, $\tau \geq 0$ is another parameter in addition to μ, such that $\tau = 0$ corresponds to the case of uniform distribution and the dispersion around the mode decreases as τ increases. For example, $\tau = \kappa$ for the von Mises distribution (1.3), and $\tau = \rho$ for wrapped Cauchy distribution (1.4).

The probability function for the marginalized discrete distribution based on the circular location family (2.25) is given by

$$p(r|m,\tau,\mu) = \int_{2\pi r/m}^{2\pi(r+1)/m} g_\tau(\theta-\mu)d\theta, \quad r \in \mathbb{Z}_m, \ \mu \in [0,2\pi), \tag{2.26}$$

and we will call this distribution the "marginalized discrete circular location family".

Similarly, the probability function of the conditionalized discrete distribution based on the circular location family (2.25) is given by

$$p(r|m,\tau,\mu) = \frac{g_\tau(2\pi r/m-\mu)}{\sum_{r=0}^{m-1} g_\tau(2\pi r/m-\mu)}, \ r \in \mathbb{Z}_m, \mu \in [0,2\pi), \tag{2.27}$$

and we will call this distribution the "conditionalized discrete circular location family".

The characteristic function for the probability function (2.26) or (2.27) is given by

$$\psi_{p,m} = \sum_{r=0}^{m-1} p(r|m,\tau,\mu)e^{ip\frac{2\pi r}{m}}. \tag{2.28}$$

Since $g_\tau(\theta)$ can be expressed in terms of its characteristic function ϕ_p (see Mardia and Jupp, 2000, p.27) as

$$g_\tau(\theta) = \frac{1}{2\pi}\sum_{q=-\infty}^{\infty} \phi_q e^{-iq\theta},$$

it can shown that the characteristic function of the marginalized discrete location family is

$$\psi_{p,m} = \begin{cases} 1, & p = 0 \\ \frac{me^{ip(\mu - \frac{\pi}{m})}\sin(\pi p/m)}{\pi}\sum_{l=-\infty}^{\infty}\frac{\phi_{lm+p}}{lm+p}e^{ilm\mu}, p \in \mathbb{Z}_m\backslash\{0\}, \end{cases} \tag{2.29}$$

and the characteristic function of the conditionalized discrete location family is

$$\psi_{p,m} = e^{ip\mu}\frac{\sum_{l=-\infty}^{\infty}\phi_{lm+p}e^{ilm\mu}}{\sum_{l=-\infty}^{\infty}\phi_{lm}e^{ilm\mu}}, \ p \in \mathbb{Z}_m. \tag{2.30}$$

2.3.2. Circular Discrete Family Based on Trigonometric Sums. In this section, we derive the marginalized and conditionalized discrete distributions starting from the flexible continuous distribution based on trigonometric sums introduced by Fernández-Durán (2004). For a set of complex numbers $\mathbf{c} = \{c_0, c_1, \ldots, c_J\}$ such that

$$\sum_{k=0}^{J}|c_k|^2 = \frac{1}{2\pi}, \tag{2.31}$$

the pdf defined by Fernández-Durán (2004) is

$$f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{J}\left\{a_k\cos(k\theta) + b_k\sin(k\theta)\right\}, \quad \theta \in [0, 2\pi), \tag{2.32}$$

where (a_k, b_k) are such that $a_k - ib_k = 2\sum_{\nu=0}^{J-k}c_{\nu+k}\bar{c}_\nu$. The specific choice of (a_k, b_k) is a necessary and sufficient condition to ensure positivity of the function f. The pdf (2.32) can also be written as

$$f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{J}\rho_k\cos(k\theta - \phi_k), \quad \theta \in [0, 2\pi), \tag{2.33}$$

where

$$\rho_k = \sqrt{a_k^2 + b_k^2} \text{ and } \phi_k = \arctan(b_k/a_k). \tag{2.34}$$

We will refer to the distributions obtained by applying Definitions 1 and 2 on the pdf (2.32) as the "marginalized discrete trigonometric sum" distribution (denoted $MDTS(m, \boldsymbol{c})$) and the "conditionalized discrete trigonometric sum" distribution (denoted $CDTS(m, \boldsymbol{c})$), respectively. It is easy to see from (2.32) and (2.33) that the probability function of $MDTS(m, \boldsymbol{c})$ is given by

$$p(r|m, \boldsymbol{c}) = \frac{1}{m} + \frac{2}{\pi} \sum_{k=1}^{J} \frac{\sin(\pi k/m)\rho_k}{k} \cos\left(\frac{2\pi k(r+1/2)}{m} - \phi_k\right), \ r \in \mathbb{Z}_m, \tag{2.35}$$

and the probability function of $CDTS(m, \boldsymbol{c})$ is given by

$$p(r|m, \boldsymbol{c}) = \frac{1 + 2\sum_{k=1}^{J} \rho_k \cos\left(\frac{2\pi kr}{m} - \phi_k\right)}{m\left(1 + 2 \sum_{\substack{k=1 \\ k=0 (\text{mod m})}}^{J} a_k\right)}, \ r \in \mathbb{Z}_m, \tag{2.36}$$

where ρ_k and ϕ_k are as in (2.34). Note that for $J = 1$, (2.35) and (2.36) give the marginalized and conditionalized discrete cardioid distributions, respectively. We see that the two distributional forms are identical although with different parametrization, which is consistent with Theorem 3.

As in the continuous case, the above constructed discrete families give flexibility in modeling multimodality and skewness in circular discrete data. See also Imoto et al. (2020). In this paper, we will give a particular application in Section 4.1.2.

3. Key Special Discrete Distributions and Their Properties

We now give the marginalized and conditionalized methods for the von Mises (1.3) and the wrapped Cauchy (1.4) as the parent distributions followed by some basic properties including characteristic function, estimation and hypothesis testing. We begin with the definitions of these distributions.

DEFINITION 3. Using (2.15), the probability function for the marginalized discrete von Mises distribution (MDVM) with mean parameter μ and concentration parameter κ, denoted by $MDVM(m, \kappa, \mu)$, is given by

$$p(r|m, \kappa, \mu) = \frac{1}{2\pi I_0(\kappa)} \int_{2\pi r/m}^{2\pi(r+1)/m} e^{\kappa \cos(\theta - \mu)} d\theta, r \in \mathbb{Z}_m, \quad \mu \in [0, 2\pi), \tag{3.1}$$

DEFINITION 4. Using (2.16), the probability function for the conditionalized discrete von Mises (CDVM) distribution with mean parameter μ and concentration parameter κ, denoted by $CDVM(m, \kappa, \mu)$, is given by

$$p(r|m, \kappa, \mu) = \frac{1}{L_0(\kappa, \mu)} e^{\kappa \cos(2\pi r/m - \mu)}, r \in \mathbb{Z}_m, \quad \mu \in [0, 2\pi), \tag{3.2}$$

where the normalizing constant is the reciprocal of the function

$$L_0(\kappa,\mu)=\sum_{r=0}^{m-1} e^{\kappa\cos(2\pi r/m-\mu)}. \tag{3.3}$$

Similarly, we have the following definitions for the wrapped Cauchy case.

DEFINITION 5. The probability function of the marginalized discrete wrapped Cauchy (MDWC) distribution with mean parameter μ and concentration parameter ρ, denoted by $MDWC(m,\rho,\mu)$, is given by

$$\begin{aligned} p(r|m,\rho,\mu) &= \frac{1}{2\pi}\cos^{-1}\left\{\frac{(1+\rho^2)\cos(\frac{2\pi(r+1)}{m}-\mu)-2\rho}{1+\rho^2-2\rho\cos(\frac{2\pi(r+1)}{m}-\mu)}\right\} \\ &\quad -\frac{1}{2\pi}\cos^{-1}\left\{\frac{(1+\rho^2)\cos(\frac{2\pi r}{m}-\mu)-2\rho}{1+\rho^2-2\rho\cos(\frac{2\pi r}{m}-\mu)}\right\} \\ &\quad r\in\mathbb{Z}_m,\quad \mu\in[0,2\pi),\quad \rho\in[0,1). \end{aligned} \tag{3.4}$$

Alternatively, for computational purposes, the above expression can be written as

$$\begin{aligned} p(r|m,\rho,\mu) &= \frac{1}{\pi}\arctan\left(\frac{\frac{1+\rho}{1-\rho}\{\tan(\pi(r+1)/m-\mu/2)-\tan(\pi r/m-\mu/2)\}}{1+\left(\frac{1+\rho}{1-\rho}\right)^2\tan(\pi(r+1)/m-\mu/2)\tan(\pi r/m-\mu/2)}\right), \\ &\quad r\in\mathbb{Z}_m,\quad \mu\in[0,2\pi),\quad \rho\in[0,1). \end{aligned} \tag{3.5}$$

Further, we have

DEFINITION 6. The probability function of the conditionalized discrete wrapped Cauchy distribution with mean parameter μ and concentration parameter ρ, denoted by $CDWC(m,\rho,\mu)$, is given by

$$\begin{aligned} p(r|m,\rho,\mu) &= \frac{1}{D_0(\rho,\mu)}\frac{1}{1-2\rho\cos\left(\frac{2\pi r}{m}-\mu\right)+\rho^2}, \\ &\quad r\in\mathbb{Z}_m,\ \mu\in[0,2\pi),\ \rho\in[0,1). \end{aligned} \tag{3.6}$$

where the normalizing constant is the reciprocal of the function

$$D_0(\rho,\mu)=\frac{m(1-\rho^{2m})}{(1-2\rho^m\cos(m\mu)+\rho^{2m})(1-\rho^2)}. \tag{3.7}$$

The normalizing constant (3.7) is derived in the supplement Section A2, Corollary A1. For simplicity of notation, while writing the marginalized or conditionalized discrete probability functions, we will omit the subscripts such as v or c corresponding to von Mises or Cauchy, but they will be clear from the context and by the explicit mention of κ versus ρ as the concentration parameters. So, we will always denote the discrete probability functions by $p(r|m,\kappa,\mu)$ for von Mises or $p(r|m,\rho,\mu)$ for wrapped Cauchy. We now make some additional observations specific to CDVM and CDWC distributions.

Probability Functions Figure 1 plots the probability functions of $CDWC(m,\rho,\mu)$ and $CDVM(m,\kappa,\mu)$, for (i) $m = 10$ and (ii) $m = 37$ with $\mu = 2\pi 5/m$ and $\mu = 2\pi 16/m$ respectively, for $\rho = 0.5$ and its mapped κ value. In order to compare the probability functions of $CDWC(m,\rho,\mu)$ and $CDVM(m,\kappa,\mu)$, we need to first map the parameters ρ to κ. We do so by matching their first trigonometric moments given by (3.10) and (3.12) below, i.e. $B(\kappa) = \rho_w$. We note that the CDWC is more spiked and heavy tailed compared to CDVM.

Characteristic Functions We make some observations based on the characteristic functions of $CDVM$ and $CDWC$ distributions.

(a) **CDVM.** For the $CDVM$ distribution (with $\mu = 0$), we can also obtain an alternative simplified form for the characteristic function.

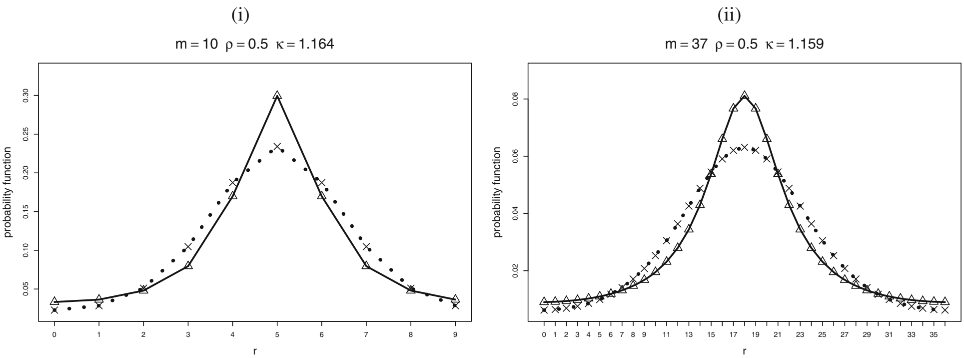

Figure 1: Probability functions of $CDWC(m,\rho,\mu)$(triangles joined by solid line) and $CDVM(m,\kappa,\mu)$ (cross joined by dotted lines) plotted for (i) $m = 10$ and (ii) $m = 37$ with $\mu = 2\pi 5/m$ and $\mu = 2\pi 16/m$ respectively, for $\rho = 0.5$ and its mapped κ value by matching the first trigonometric moment

Let us write

$$L_p(\kappa) = \sum_{r=0}^{m-1} \cos\left(p\frac{2\pi r}{m}\right) e^{\kappa\cos\left(\frac{2\pi r}{m}\right)}. \tag{3.8}$$

So, for $\Theta \sim CDVM(m,\kappa,\mu=0)$, we have

$$\psi_{p,m} = E\left[e^{ip\Theta}\right] \;=\; B_p(\kappa), \text{where } B_p(\kappa) = L_p(\kappa)/L_0(\kappa). \tag{3.9}$$

It then follows by writing $B_1(\kappa) = B(\kappa)$, that

$$B(\kappa) = E\left(\cos\frac{2\pi r}{m}\right) \text{ and } B'(\kappa) = Var\left(\cos\frac{2\pi r}{m}\right). \tag{3.10}$$

$L_p(\kappa)$ is the discrete analogue of the Bessel function $I_p(\kappa)$. It is to be noted that not all of the standard identities of $I_p(\kappa)$ (see Mardia and Jupp (2000), Appendix 1) necessarily hold for its discrete analogue.

(b) **CDWC.** For $\Theta \sim CDWC(m,\rho,\mu=0)$, the pth trigonometric moments ($p \in \mathbb{Z}_m$) are given by

$$\alpha_{p,m} = E\left[\sin\left(p\Theta\right)\right] = 0, \quad \beta_{p,m} = E\left[\cos\left(p\Theta\right)\right] = \frac{\rho^p(1+\rho^{m-2p})}{1+\rho^m}. \tag{3.11}$$

For $p=1$, this leads to the mean resultant length

$$\rho_w = \frac{\rho(1+\rho^{m-2})}{1+\rho^m}. \tag{3.12}$$

In general, $0 \le \rho_w \le 1$ and as $m \to \infty$, $\rho_w \to \rho$. By property of characteristic functions, in general $0 \le \beta_{p,m} \le 1$, and as $m \to \infty$, $\beta_{p,m} \to \rho^p$, as can be seen from (3.11), which is known to be the characteristic function of the wrapped Cauchy distribution as expected. Further, this convergence happens at an exponential rate. To see this, note that

$$\psi_{p,m}(\rho) - \rho^p \;=\; \frac{\rho^p\left(\rho^{m-2p}-\rho^m\right)}{1+\rho^m} = \frac{\rho^{m-p}\left(1-\rho^{2p}\right)}{1+\rho^m}.$$

For any fixed p, it follows that $|\psi_{p,m}(\rho)-\rho^p| \le \rho^{m-p}$, and hence $|\psi_{p,m}(\rho)-\rho^p| = \mathcal{O}(\rho^{m-p})$. In particular, since $\psi_{1,m} = \rho_w$, we have $|\rho_w - \rho| = \mathcal{O}(\rho^{m-1})$.

Estimation The maximum likelihood estimates (mle) of (μ, κ) for $CDVM$ and (μ, ρ) for $CDWC$ can be obtained by iteratively solving the maximum likelihood equations for the two parameters, the details of which are given in Section A5.1 of the supplement. Further, we give asymptotically equivalent estimates to mle which are simpler to compute:

(a). Moment estimates: Characteristic functions for the marginalized discrete and conditionalized discrete location families with cardioid, von Mises and wrapped Cauchy as parent distributions, based on the general formulas (2.29) and (2.30), are given in the supplement (Section A3 Table S2). These can be used to estimate parameters based on matching of trigonometric moments from the data. For $\Theta \sim CDWC(m, \rho, \mu)$, the trigonometric moments have a closed form, given by

$$\begin{aligned} E[\cos(\Theta)] &= A\cos(\mu) + B\cos((m-1)\mu), \\ E[\sin(\Theta)] &= A\sin(\mu) - B\sin((m-1)\mu), \end{aligned}$$

where

$$A = \frac{\rho(1-\rho^{2m-2})}{(1-\rho^{2m})}, B = \frac{\rho^{m-1}(1-\rho^2)}{(1-\rho^{2m})}.$$

For the constrained case when $\mu = 2\pi t/m, t \in \mathbb{Z}_m$, the above equations simplify to

$$E[\cos(\Theta)] = \frac{\rho(1+\rho^{m-2})}{1+\rho^m}\cos(\mu) \text{ and } E[\sin(\Theta)] = \frac{\rho(1+\rho^{m-2})}{1+\rho^m}\sin(\mu).$$

(b). Hybrid estimates: When n is large, plug-in the sample mean direction $(\bar{\theta})$ for μ, and obtain mle for only ρ or κ.

(c) Constrained estimates: When m is large, constrain μ to $\{2\pi t/m, t \in \mathbb{Z}_m\}$. In this case, the normalizing constants (3.3) and (3.7) become free of μ, leading to some simplifications. For $CDVM$, this approach leads to following explicit equations for mle:

$$B(\hat{\kappa}) = \bar{R}\cos\left(\bar{\theta} - \frac{2\pi\hat{t}}{m}\right) \text{ and } \hat{t} = \left[\frac{m\bar{\theta}}{2\pi}\right]_m,$$

where $B(\kappa) = L_1(\kappa)/L_0(\kappa)$, $L_p(\kappa) = \sum_{r=0}^{m-1}\cos\left(p\frac{2\pi r}{m}\right)e^{\kappa\cos\left(\frac{2\pi r}{m}\right)}$, $p \in \mathbb{Z}_m$, and $[x]_m$ denotes the closest integer to x, modulo m.

Asymptotic normality of these estimate follow using the results in Pewsey (2004) for the case (a) and Mardia et al. (2016) for the cases (b) and (c).

Testing of Hypothesis We will give more details in the next section on testing of hypothesis as we apply the methodology as required in the next section but we outline some main points.

The Rayleigh Test is well known to test uniformity under the von Mises distribution, that is to test (see, for example, (Mardia and Jupp, 2000))

$$H_0 : \kappa = 0, \quad H_1 : \kappa > 0,$$

where the mean μ is unknown. It is based on $T_1^2 = 2n\bar{R}^2$, which under H_0, is approximately chi-squared with 2 degrees of freedom. Now, given the data vector $\boldsymbol{w}$ of iid observations under CDVM, the log-likelihood ratio test statistic (T) can be written as

$$T(\boldsymbol{w}, \hat{\kappa}, \hat{\mu}) = -2LL(\boldsymbol{w}|m, \kappa = 0) + 2LL(\boldsymbol{w}|\hat{\kappa}, \hat{\mu}),$$

where $(\hat{\kappa}, \hat{\mu})$ is the mle based on the data vector $\boldsymbol{w}$ and $LL(\cdot)$ is the log-likelihood. If we denote the computed value of T in the data sample by T_d, then

$$\text{p-value} = P\left(T(\boldsymbol{w}, \hat{\kappa}, \hat{\mu}) \geq T_d\right).$$

The test rule is then to reject H_0 for values of the p-value in comparison to a chosen significance level and use bootstrap for the mle and the p-values for the tests. For large m, the T-test = the Rayleigh test. We can extend it easily to the circular location family including CDWC which is easier to use as the normalizing constant is simpler.

4. Examples

We apply some of the models developed in the previous sections to analyze real data on roulette wheel outcomes and smart health monitoring readings on SBP acrophase. In both these situations, data are circular and discrete. Also, in both these situations, data are generated in abundance daily, although they may not usually be accessible in the public domain. Interestingly, the acrophase is an example where the data has an irregular discrete support. Our analysis and findings that are presented below are mainly illustrative of the kind of insights that are possible through the different circular discrete models.

4.1. Roulette Wheel Data: Online Gaming and Casino Spins There has been ongoing search to find a plausible test for testing unbiasedness of a roulette wheel. The problem is now more pressing than ever before with the rise of many online gaming sites, e.g. An Online Gaming Site https://10bestcasinos.co.uk/en-en_d_rl.html. For example, the UK Gambling Commission requires statistical testing to ensure fairness, by an approved third party as per the guidelines provided in their "Testing strategy for compliance with remote gambling and software technical standards" at UK Government Compliance https://www.gamblingcommission.gov.uk/.

Indeed, Pearson (1894, 1897) was captivated by this problem and had acquired his data of $n = 16,563$ roulette spins from the Monte Carlo Casino as recorded in a journal "La Monaco". He constructed three tests, but as the subject of Directional Statistics was still developing, it was usual to ignore the circular aspect of the data. A brief historical insight into his work, along with images of a typical European roulette wheel, are given in the supplement Section A4. Karl Pearson's original sequence of roulette spin data is not available and we work with data from an online roulette simulator as well as some industrial casino data obtained from spins of different European roulette wheels. Note that here the number of outcomes is $m = 37$ (as against an American roulette $m = 38$). In some cases, the data is available as a streaming sequence of outcomes from successive spins of the roulette wheel, and in others we may just have the frequency distribution of the outcomes without knowledge of the sequence. Accordingly, we illustrate different types of analysis.

We look at the following four roulette data:- three different streaming sequence roulette data and one cumulative frequency data.

Roulette data 1 Our first data is of size $n = 1000$, a sequence of outcomes from successive spins of an online European roulette simulator available at (Online Simulator: last accessed 9-sept-2020)

Roulette data 2 This data has outcomes from successive spins of a real European roulette recorded in a casino in Slovenia, $n = 8299$.

Roulette data 3 This data has outcomes from successive spins from the same casino as roulette data 2, but from a different roulette wheel, $n = 8106$.

Roulette data 4 This data is only of cumulative frequency of roulette wheel spins from an industrial consulting project at University of Leeds, $n = 3094$.

Rows (i)–(iv) of Table 1 give their respective frequency distributions. We will deal with sequential roulette Data 1, Data 2 and Data 3 in the next subsection (Section 4.1.1) and the cumulative roulette Data 4 in the subsequent subsection (Section 4.1.2).

4.1.1. Analysis of the Streaming Sequence Roulette Data 1–3. We consider here roulette Data 1, Data 2 and Data 3. The main challenge is to

Table 1: Frequency distributions for the roulette Data 1, 2, 3 and 4. The entries of angular positions (r) in the first row correspond to angles $2\pi r/m$ on the circle, the second row shows the corresponding label on the roulette wheel and the other rows show the frequencies of outcomes

	r	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	label	0	26	3	35	12	28	7	29	18	22	9	31	14	20	1	33	16	24	5
(i)	Data 1	45	19	34	20	29	25	25	19	34	16	29	25	32	31	32	18	29	32	28
(ii)	Data 2	242	228	231	213	241	211	202	221	208	216	230	199	232	230	220	208	196	215	203
(iii)	Data 3	230	202	184	231	200	221	169	187	224	212	232	221	255	236	225	190	189	199	214
(iv)	Data 4	76	78	72	85	74	101	93	71	67	70	67	103	104	102	93	82	84	73	65

	r	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	Total
	label	10	23	8	30	11	36	13	27	6	34	17	25	2	21	4	19	15	32	
(i)	Data 1	27	25	31	38	33	32	26	11	21	24	27	21	29	33	31	22	23	24	1000
(ii)	Data 2	244	260	213	197	219	240	244	242	227	221	239	233	236	227	229	221	218	243	8299
(iii)	Data 3	196	203	206	207	213	247	222	199	201	233	250	254	274	239	220	256	218	247	8106
(iv)	Data 4	90	80	73	101	75	86	79	89	87	75	91	85	94	106	89	78	86	70	3094

detect a possible bias in a roulette based on a streaming sequence of spin outcomes. We note here that since the roulette Data 1, 2 and 3 are available as a time series, we carried out a test for serial independence, in the lines of Watson and Beran (1967), but adapting to discrete data (more details are in Mardia and Sriram (2020)). The test indicates that overall there is no dependence. Our analyses and findings for these data are as follows. The main challenge is to detect a possible bias in a roulette based on a streaming sequence of spin outcomes.

Analysis 1 First, we carry out testing unbiased (of the wheels) using only cumulative frequencies for the roulette Data 1, 2, and 3 given in Table 1, which is equivalent to testing for uniformity, i.e. $H_0 : \tau = 0$ (unbiased wheel) vs. $H_1 : \tau \neq 0$ (biased wheel). For this purpose, we use a log-likelihood ratio test statistic (denoted T), which is computed as the difference in the log-likelihoods at the maximum likelihood estimates (mle) and at the null hypothesis $\tau = 0$. Recall for the continuous circular location scale family, $\tau = 0$ corresponds to the uniform distribution on the circle, so a specific value of μ is not required for H_0. We carry out this analysis using the CDWC model. Table 2 shows the mle, test statistic and the p-value for each the roulette Data 1,2 and 3. Comparing the p-value with a 5% significance level, we conclude that the evidence for bias does not exist for roulette data 1, is weak for roulette Data 2 and strong for roulette Data 3. The estimated mode for data 3 is $\hat{\mu} = 5.34$ approximately corresponding to the angular

Table 2: Based on CDWC model, results of Analysis 1 for roulette data 1-3

	Roulette	n	$\hat{\rho}$	SE($\hat{\rho}$)	$\hat{t} = \lfloor \frac{m\hat{\mu}}{2\pi} \rfloor$	$\hat{\theta} = \hat{\mu}$	$SE(\hat{\theta})$	T	$SE(T)$	p-value
(i)	data 1	1000	0.019	0.017	17	2.976	1.236	0.688	3.010	0.711
(ii)	data 2	8299	0.020	0.007	30	5.106	0.414	6.676	5.420	0.046
(iii)	data 3	8106	0.030	0.008	31	5.340	0.279	14.333	7.924	0.000

position $r \approx 31$. See Supplement Section A5 Figure S2 for the circular histogram of data 3. It turns out these conclusions happen to be the same if we had used the CDVM model, or if we had used some alternative tests known in the context of continuous data (see Supplement Section A5 for details).

Analysis 2 Analysis 1 does not use the streaming nature of the outcomes, that is we have used only the cumulative frequency not the full time series. Here, we delve further into the sequence of outcomes. Let us denote the sequence of angular positions of the roulette outcomes by $\{w_i : i \in \{1, 2, \ldots, n\}\}$. The mapping of angular positions to labels on the roulette wheel is given in Table 1. Our goal is to estimate a change-point in the data. The model with a changepoint at $i = K$ can be constructed as follows:

$$w_i \sim \begin{cases} p(\cdot|m, \tau_1, \mu_1) & \text{if } i \le K \\ p(\cdot|m, \tau_2, \mu_2) & \text{if } i > K, \end{cases} \tag{4.1}$$

where $p(\cdot|m, \tau, \mu)$ is the probability function of the discrete circular distribution as in (2.26) and (2.27). Of particular interest is to detect a change from uniformity, i.e. $\tau_1 = 0$, where μ_1 can be arbitrary but we take it as 0 without loss of generality. The likelihood for the data can then be written as

$$L(\boldsymbol{w}|(\tau_1 = 0, \tau_2), (\mu_1 = 0, \mu_2), K) = \prod_{i=1}^{K} p(w_i|m, \tau_1 = 0, \mu_1 = 0) \times \prod_{i=K+1}^{n} p(w_i|m, \tau_2, \mu_2). \tag{4.2}$$

We can apply a Bayesian approach with standard Markov Chain Monte Carlo methods for estimation, using non-informative flat priors on the unknown parameters, viz τ_2, μ_2 and K. For our data, we carry out the change-point analysis using the $CDWC(m, \rho, \mu)$ model, first based on the full data sequence and then based on partial data sequences. Specifically, we use a Gibbs sampling procedure to obtain the posterior distributions, by taking the support of $(\rho_2,\ \mu_2,\ K)$ to be a suitably fine grid of values, viz. $\rho_2 \in \{0, 0.001, 0.002, \ldots, 0.999\}$, $\mu_2 \in \{0, 0.001 \times 2\pi, 0.002 \times 2\pi, \ldots, 0.999 \times 2\pi\}$

and $K \in \{1, 2, \ldots, n\}$. The posterior distribution summaries obtained for each of the full sequences of roulette data 1,2, 3 are shown in Table 3. For any given roulette data sequence, we conclude that there is evidence for a changepoint if the 95% highest probability density (hpd) credible interval for ρ_2 is removed from 0. Accordingly, we conclude that there is no evidence for a change from uniformity in data sequences 1 and 2. For roulette data 1 and 2, we also see that the 95% hpd interval for K spans a very large range between 1 and n, as one would expect if the distribution of K is close to a discrete uniform on $\{1, 2, \ldots, n\}$. This further supports the absence of a changepoint in roulette data 1 and 2.

However, there is evidence for a changepoint in roulette data 3 since the 95% hpd interval for ρ_2 is clearly removed from 0. We can also see that the estimated posterior mode for the changepoint in roulette data 3 is $K = 1226$. Therefore, subsequent to the changepoint, the distribution of outcomes in roulette data 3 has changed from uniform distribution to one that has a single mode (at $\mu_2 = 5.305$ with angular position $r \approx 31$). Such a bias might be resulting from a slight "tilt" in the roulette wheel downwards to favour such a mode. In Section 4.1.2, we also look at a different type of bias, possibly resulting from "wobble" of a roulette wheel.

Testing Streaming Consistency Further, to see how early the changepoint would have been detected, we treat the spins as streaming data by increasing the number of spins sequentially by 500. So, we apply the same changepoint detection procedure on partial data sequences of roulette data 3, i.e. spins $1 : u$ for $u \in \{500, 1000, 1500, \ldots, 8000, 8106\}$. Figure 2 shows the plots of the 95% hpd intervals, posterior mean for ρ_2 and mode for K for different choices of the upper bound. We can see that starting from an upper bound of 4000 onwards, the 95% credible interval for ρ_2 appears to be removed from 0, and for 5500 onwards it is even more clearly removed from 0. Correspondingly the posterior mode for K is settled somewhere between 1000 and 2000. Also, the 95% hpd for K appears to be stabilize for the data range 1:5500 and after. So, while we start detecting the change weakly based on the first 4500 outcomes, the evidence becomes stronger as we start including outcomes 1:5500 and after.

We now comment on how our work is basically different from some other work on circular change point detection. Pewsey and García-Portugués (2021) have given a survey of changepoint detection with continuous angular data, but our method is for discrete circular data. There has been

Table 3: Posterior distribution summaries from the $CDWC$ changepoint model, for parameters (μ_2, ρ_2, K) based on full data ranges (i.e., $1:n$) for roulette data 1,2,3. The distribution before changepoint is assumed to be uniform, i.e. $\rho_1 = 0, \mu_1 = 0$

	ρ_2			μ_2		K				
Roulette	Mean	sd	95% hpd interval	Mean	sd	n	Mode	sd	95% hpd interval	Changepoint
data 1	0.071	0.113	[0 , 0.288]	2.747	1.822	1000	998	305.34	[91, 1000]	No
data 2	0.031	0.060	[0 , 0.086]	5.338	0.949	8299	8294	2828.61	[244 , 8298]	No
data 3	0.035	0.010	[0.015 , 0.053]	5.305	0.270	8106	1226	723.13	[3, 2612]	Yes, at 1226

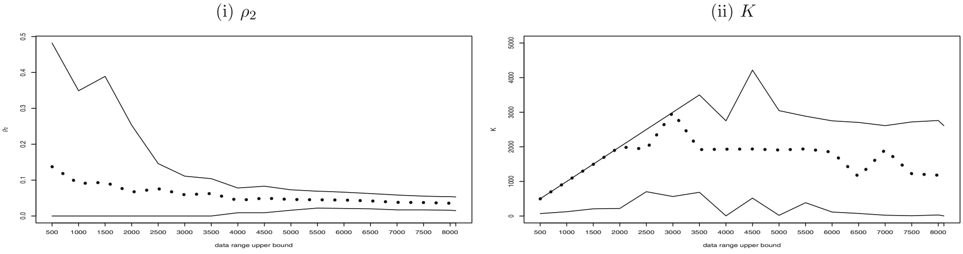

Figure 2: Results of changepoint analysis based on different partial data ranges for Data 3. plot (i) shows posterior summaries for ρ_2 and (ii) for K plotted against data range upper bound. In each plot, solid lines mark the 95% credible intervals. The dotted line in plot (i) shows the posterior mean for ρ_2 and in plot (ii) shows the posterior mode for K

other work in this area motivated by control charts, e.g. Lombard and Maxwell (2012), Laha and Gupta (2011) and Rao and Girija (2020). Compared to these approaches, our strategy is different as it is model-based with unknown parameters, whereas the method in Lombard and Maxwell (2012) is nonparametric. Further, control charts, e.g. Laha and Gupta (2011) and Rao and Girija (2020), are not suitable for testing uniformity for the following reasons. Broadly speaking, in their method, given a distribution, n samples of size n_1 are generated. For each sample, based on the n_1 observations, the circular mean is computed. Then two quantities are found after sorting the n values, viz. (a) CCR (clockwise control ray) by eliminating the first α percent values and (b) ACR (anti-clockwise control ray) by eliminating the last α percent values. Such an approach may not be suitable for checking deviation from uniformity because the ACR and CCR will be wide and symmetric around π and so it will not have any observations that fall in the rejection region if the data is actually coming from a distribution concentrated around π. Although their work is also based on some model assumptions, it requires a priori fixing of parameter values to determine the ACR and CCR.

4.1.2. Analysis of the Cumulative Frequency Data 4. Here, we consider Data 4 where only a cumulative frequency roulette outcomes is available, i.e. instead of a streaming sequence we only have the frequency distribution of outcomes. As part of an industrial consulting project at University of Leeds, Baines (1990) recommended a protocol for certifying a casino roulette wheel as "unbiased" using five different statistical tests, two based on chi-square statistic and the other three based on variations of the Rayleigh test (see

Supplement Section A5.4 for detailed description of these tests which were given by K. V. Mardia in a personal correspondence). As per the protocol, if all five statistical tests resulted in the acceptance of the Null Hypothesis (H_0) (i.e. no evidence of bias), then the wheel would be certified as having passed the randomness test. However, if any of the five statistical tests resulted in the acceptance of the Alternative Hypothesis (H_1)(i.e. evidence of bias), a second series of at least 370 spins would be collected and the five statistical tests repeated on the new data as well as the combined data from the two series. The choice of the level for the tests would be contextual, i.e. chosen between 5% and 0.1% depending on the acceptable risk of wrongly concluding that the wheel is biased. The cost implications also influence the number of runs to be considered for such tests.

Based on an analysis of different sub-series of roulette spins as well as the combined series, Al Baines and Kanti Mardia reported evidence to suggest that the roulette had a quadri-modal bias with possible modes at the roulette slot positions 12,20,30 and 21, which correspond to angular positions $r \in \{4, 13, 22, 32\}$. However, his methodology is not based on a statistical model, so it does not systematically identify the nature of the specific alternative in H_1, in particular the positions of modes and their likelihood.

To allow some flexibility to capture the multimodal aspect of this data and to achieve the highest log pseudo marginal likelihood (see e.g. Carlin and Louis, 2009, p. 215), we chose $J = 4$ to fit the CDTS model as in (2.36) to the combined series of roulette outcomes used in Baines (1990). We refer to this as 'Data 4' and give its frequency distribution in row (iv) of Table 1. We find it convenient to estimate the model using a Bayesian approach with a standard MCMC approach based on a random walk Metropolis-Hastings algorithm. Figure 3 shows the fitted model along with the frequency distribution of the data. The fitted model suggests modes at angular positions $r \in \{4, 13, 23, 31\}$, which is somewhat consistent with Al Baines' conclusion of 4 equi-spaced modes, although our finding suggests that the successive modes may not exactly be equi-spaced. As per our model, the angle subtended between successive modes (in degrees) are 87.6, 97.3, 77.8 and 97.3 respectively. The roulette slots corresponding to these angular positions are 12, 20, 11, 2, with estimated probabilities 0.0279, 0.0319, 0.0289 and 0.0305 respectively. So, our finding goes a bit beyond Baines (1990) to suggest that the roulette wheel possibly has an asymmetric wobble (i.e. modes that are not exactly equi-spaced and with unequal probabilities).

4.2. Acrophase Data: Ambulatory BP Monitoring Systolic blood pressure (SBP) has a circadian rhythm. To monitor it, patients wear ambulatory devices that regularly measure and record blood pressure. In non-invasive

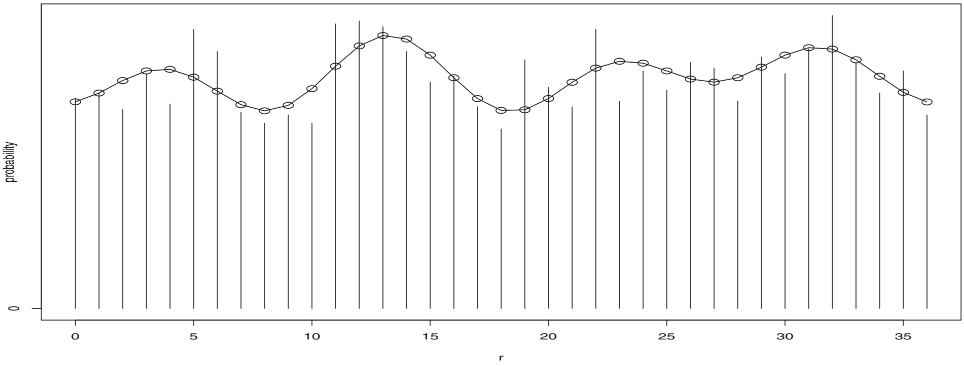

Figure 3: CDTS model fit to the combined series of roulette outcomes in Baines (1990)

smart health monitoring, these readings are typically recorded at predetermined, possibly irregular, discrete time points during daytime and nighttime. Our objective is to analyze "acrophase", the time at which maximum SBP is attained in a day. Monitoring the acrophase can provide an automated early warning of a possible medical condition before it becomes clinically obvious. Typically, the readings taken by industrial monitoring devices are more frequent during daytime (e.g. each half hour) than at nighttime (e.g. each hour). For more details, see Acrophase Times https://www.londoncardiovascularclinic.co.uk/cardiology-info/investigation/24-hr-ambulatory-blood-pressure.

Therefore, the resulting acrophase data is circular, discrete and supported on an irregular lattice on the circle.

We use acrophase data on SBP based on readings taken for an individual from 3-31-1998 to 7-7-2000, collected by the Halberg Chronobiology Center (University of Minnesota). We note that for a few days, there were multiple time points where maximum SBP was achieved. In such cases, we retained all such time points, thus resulting in a total of 880 data points. As mentioned in the introduction, typically, acrophase data is extracted from SBP measurements at each half hour during daytime (8 am to 8 pm) and each hour during nighttime (8 pm to 8 am). If we map 8 am to 0 radians and 8 pm to π radians, the acrophase times get mapped to an irregular support of $m = 36$ points on the circle given by

$$\left\{\frac{2\pi r}{48}, r = 0, 1, 2, \ldots, 24\right\} \bigcup \left\{\frac{2\pi r}{48}, r = 26, 28, \ldots, 44, 46\right\}. \tag{4.3}$$

Note that unlike the case of regular support where the points in the support are expressed in terms of $\mathbb{Z}_m$, here the points in the irregular support are expressed as $2\pi r/48$ to accommodate the half-hour and one-hour time points. Table 4 shows the frequency distribution of this acrophase data. Our initial interest is to estimate the centering and concentration parameters for this data. While the centering parameter will be indicative of the most likely timing of acrophase, the concentration parameter will indicate the extent of variability around that timing. We adapt CD distributions to an irregular support as follows and prefix such distributions by ICD, that is, for example, ICDVM stands for the CDVM with irregular support. Let $\mathcal{S} = \{\theta_0, \theta_2, \ldots, \theta_{m-1}\}$ denote the irregular circular lattice support (4.3). The conditionalized discrete probability function for any given parent pdf $f(\cdot)$ on the irregular support $\mathcal{S}$, ICD, is given by

$$p(\theta_l) = \frac{f(\theta_l)}{\sum_{k=0}^{m-1} f(\theta_k)}, \quad \theta_l \in \mathcal{S}. \tag{4.4}$$

Table 5 shows the estimated centering (μ) and concentration (κ) for the ICDVM model. We found it computationally convenient to estimate the parameters using a Bayesian approach by assuming non-informative priors on the parameters, specifically we take uniform priors for $\mu \in [0, 2\pi)$. Since the concentration parameter in CDVM is κ, a mapping is done to an equivalent value of ρ based on matching of first moment between von Mises and wrapped Cauchy distributions. To get a prior for κ, we first map it to ρ and use a uniform prior on $\rho \in [0, 1)$. From Table 5, we note that the estimates are

$$\mu = 2.462 \ (\approx 17:30 \ hrs), \kappa = 1.114, \ \rho = 0.48$$

so there is moderate concentration. Further, we note that the approximate 99% confidence interval for μ is $17:30$ hours ± 33 minutes, slightly different from the mode of 19:00 hrs observed in the data in Table 4. This can be the effect of skewness in the data as seen in its histogram given in Fig. 4, which has led us to explore the analysis with the parent as a skew circular distribution since one of our objectives is a construction that allow flexible families of plausible discrete circular distributions.

We have selected the circular distribution of Kato and Jones (2015), which has four parameters that control the first four trigonometric moments, leading to unimodal symmetrical as well as skew distributions as particular cases. This family also has an analytically tractable normalizing constant and its pdf is given by

$$g_{KJ}(\theta) = \frac{1}{2\pi}\left(1 + 2\gamma \frac{\cos(\theta - \mu) - \rho\cos\lambda}{1 + \rho^2 - 2\rho\cos(\theta - \mu - \lambda)}\right), \ -\pi < \theta \leq \pi, \tag{4.5}$$

Table 4: Frequency distribution of the acrophase data. Also shown is r corresponding to each time point, which maps to the circle by $2\pi r/48$

Time	08:00	08:30	09:00	09:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30
r	0	1	2	3	4	5	6	7	8	9	10	11
freq	16	13	8	21	7	7	16	15	19	27	27	28
Time	14:00	14:30	15:00	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00	19:30
r	12	13	14	15	16	17	18	19	20	21	22	23
freq	31	27	27	30	24	34	46	58	54	76	79	50
Time	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00
r	24	26	28	30	32	34	36	38	40	42	44	46
freq	39	33	16	7	5	1	2	6	4	5	3	19

Table 5: Parameter estimates for ICDVM on irregular support for the acrophase data

Parameter	Estimate	se	95% interval
μ	2.462 ($\approx 17:30$ hrs)	0.049	[2.367, 2.554]
κ	1.114	0.060	[0.994, 1.230]
ρ	0.485	0.020	[0.444, 0.522]

where the parameters are constrained by

$$0 \leq \rho < 1, 0 \leq \gamma \leq (1+\rho)/2, -\pi \leq \mu, \lambda \leq \pi, \text{ and } \rho\gamma\cos\lambda \geq (\rho^2 + 2\gamma - 1)/2. \tag{4.6}$$

The conditionalized discrete distribution from (4.5), is given by the probability function

$$p\left(r|m,\rho,\mu,\gamma,\lambda\right) = \frac{1}{D^\star}\left(1 + 2\gamma\frac{\cos(\frac{2\pi r}{m} - \mu) - \rho\cos\lambda}{1+\rho^2 - 2\rho\cos(\frac{2\pi r}{m} - \mu - \lambda)}\right), r \in \mathbb{Z}_m, \tag{4.7}$$

where

$$D^\star = m\left(1 + 2\gamma\rho^{m-1}\frac{\cos(m(\mu+\lambda) - \lambda) - \rho^m\cos\lambda}{1+\rho^{2m} - 2\rho^m\cos(m(\mu+\lambda))}\right), \tag{4.8}$$

with the same constraints on parameters as in (4.6). The normalizing constant $D^\star$ is derived in the supplement (corollary A2). We will call this family, the conditionalized discrete Kato-Jones family, or briefly as CDKJ family. The CDWC is obtained as a special case when $\lambda = 0$ and $\gamma = \rho$. Note that the constraints ensure that the probability function in (4.5) is positive and hence also for the discretized version (4.7). We can now obtain from (4.4), the probability function for ICDKJ.

Adapting the method of moments approach in Kato and Jones (2015), we can obtain the estimates to use in the probability function of ICDKJ. The moment estimates are

$$\hat{\mu} = 2.248, \hat{\lambda} = 0.816, \hat{\rho} = 0.495, \hat{\gamma} = 0.584.$$

Figure 4 shows the histogram for the acrophase data along with the fitted ICDKJ probability function with these estimates. It can be seen that the mode of the fitted ICDKJ distribution is 18:30 hours approximately, which is roughly what is seen in the histogram. Also, visually, the fit captures the skew behaviour in the data adequately.

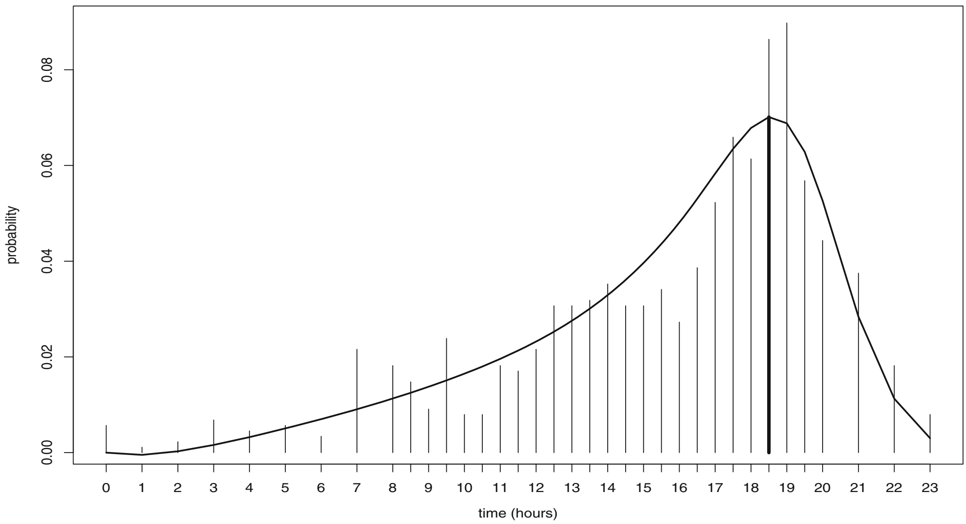

Figure 4: Histogram for the acrophase data along with the fitted ICDKJ probability function in solid line

5 Model Misspecification

In this section, we present some insights into the effect of model misspecification, specifically where a true model is a given discrete model but we apply statistical methodology developed for continuous circular distributions. As we have mentioned in Section 1, in any application with discrete data in Linear Statistics, one usually takes into account the discrete nature of the underlying population so the same principle applies here for Directional Statistics but due to lack of developments of Discrete Circular Statistics, Statistical Methodology for continuous circular statistics has been used in the literature and it is of some interest to learn where this use could lead to conflicting inference. The simulations experiments (where necessary) are given to provide some preliminary insight only and presenting elaborate simulation studies is far beyond the scope and the core of this methodology paper.

We will study three situations:

Case 1. The effect on the basic summary statistics.

Case 2. Behaviour of the maximum likelihood estimates. Case 2a Irregular case; Case 2b Regular case.

Case 3. Power of the test of uniformity.

We give an overview of these three studies and our selection of n, m, κ. Case 1 is generic and a population based comparison so n is not relevant; m covers a wide range of values, $\mu = 0$, and a fixed moderate concentration $\rho = 0.5, \kappa = 1.159$ (guided by the acrophase data with $\rho = 0.485, \kappa = 1.114$).

Case 2 has two parts Case 2a and Case 2b. Case 2a mimics the acrophase data so the support is irregular as it is for that Example, where the data is generated with $n = 880$, $\mu = 0.785$ and $\rho = 0.6$ ($\kappa = 1.516$) from the ICDVM discrete distribution so the ground truth is known. Note that for simplicity, we have taken here $\mu = \pi/4 = 0.785$ as the mean parameter can be arbitrary, and the concentration parameter $\kappa = 1.516$ computed for the acrophase data as $\bar{R} = 0.58$, so $\rho = 0.6$ which maps approximately to $\kappa = 1.516$.

Case 2b is for regular cases, $n = 1000$, $m = 10, 20$ and varying values of the concentration parameter $\kappa = 1, 2.5, 10$ so n is nearly same as for the acrophase data and its concentration parameter is covered in this range but m is kept lower than the acrophase data of $m = 36$ as the case is here regular the effect would be more evident at the lower values of m. The choice of the distributions are given below in each case.

Case 3 is motivated by the roulette data sets where n is larger and the concentration parameter is low so we have taken $\kappa = 0.05$ ($\rho = 0.03$), for each of different choices of $n = 1000, 10000$ and $m = 10, 37$.

Case 1. The Effect on the Basic Summary Statistics We carry out a study, where we sample from the marginalized and conditionalized wrapped Cauchy distributions with the parameters $\mu = 0, \rho = 0.5, \kappa = 1.159$ and varying values of m. We study the first two trigonometric moments, since these are the basic moments to capture the behavior of the distribution (through $E\cos(\theta)$) and its first order departure (through $E\cos(2\theta)$). Table 6 shows these trigonometric moments computed for $MDWC(m, \rho, \mu)$ and $CDWC(m, \rho, \mu)$ for $\rho = 0.5$, $\mu = 0$, for different values of m. As a reference, for wrapped Cauchy, $E\cos(\theta) = 0.5$ and $E\cos(2\theta) = 0.25$. Thus discretization matters for MDWC if m is less than 20, i.e. bin with angle 18 degrees. Surprisingly, the effect is a bit smaller for CDWC. Incidentally, the two moments for CDWC are obtained using the exact results given by (3.11) and these moments for MDWC were computed using simulated data of size 200000. Our study gives a specific flavor of the effect rather than the very broad Sheppard's correction (Mardia, 1972), which is too generic. Our conclusion is that discretization matters for $m \leq 20$.

It should be noted that this study has broader implications as these moments indirectly would influence inference problems associated with the

Table 6: Trigonometric moments computed for $MDWC(m, \rho, \mu)$ and $CDWC(m, \rho, \mu)$ for $\rho = 0.5$, $\mu = 0$, for different values of m. As reference, for WC, $\rho = E\cos(\theta) = 0.5$ and $E\cos(2\theta) = 0.25$

	MDWC⋆		CDWC	
m	$E\cos(\theta)$	$E\cos(2\theta)$	$E\cos(\theta)$	$E\cos(2\theta)$
3	0.159	0.159	0.667	0.667
5	0.368	0.038	0.545	0.364
10	0.466	0.190	0.501	0.254
15	0.485	0.221	0.500	0.250
20	0.493	0.232	0.500	0.250
30	0.495	0.242	0.500	0.250
50	0.497	0.248	0.500	0.250
100	0.503	0.247	0.500	0.250
500	0.499	0.248	0.500	0.250

⋆ For MDWC the moments were computed using simulated data of size 200000

von Mises and the wrapped Cauchy distributions, and specific cases are examined below in Case 2 and Case 3

Case 2. Behaviour of the Maximum Likelihood Estimates

Case 2a The irregular case.

We first assess the effect of using VM vs ICDVM via a sampled drawn from ICDVM which mimics the acrophase data, that is the irregular support, and the same data size ($n = 880$) and call it "simulated acrophase data" where we take the true value of ICDVM to be $\mu = 0.785$, $\kappa = 1.516$, and $\rho = 0.6$. Table 7 shows the estimation of centering parameter (μ) and concentration parameter (κ) for this simulated acrophase data. Parts (a) and (b) of Table 7 do not consider the discrete nature of the data. Part (a) uses sample statistics along with bootstrap (1000 resamples) standard errors, namely the circular mean of the data for μ and the mean resultant length $\bar{R}$ for ρ. Part (b) estimates the parameters by assuming a continuous model, namely von Mises. Both (a) and (b) are unable to closely estimate the true μ and κ (ρ) parameters unlike the discrete ICDVM model shown in part (c). Especially, the 95% interval for κ in parts (a) and (b) do not capture the true value. Thus the inappropriateness of applying techniques that are otherwise meant for continuous data become even more apparent when we are dealing with discrete circular data on an irregular support. In summary, this experiment clearly illustrates that using methods meant for continuous data on discrete data can be misleading.

Table 7: Estimated center and concentration parameters for data sampled from ICDVM with the same irregular support as acrophase data. The true values of parameters are $\mu = 0.785$, $\rho = 0.6$, $\kappa = 1.516$

Parameter	Estimate	se	95% interval
(a)Sample statistics with bootstrap standard errors			
μ	0.912	0.045	[0.822, 1.008]
ρ	0.493	0.020	[0.457, 0.532]
κ	1.139	0.061	[1.031, 1.265]
(b) von Mises model			
μ	0.910	0.046	[0.822, 1.001]
ρ	0.491	0.020	[0.451, 0.529]
κ	1.131	0.060	[1.014, 1.253]
(c) ICDVM model on irregular support			
μ	0.812	0.033	[0.747, 0.877]
ρ	0.608	0.015	[0.578, 0.637]
κ	1.553	0.063	[1.429, 1.676]

Case 2b The regular case.

We use CDVM (CDWC) as the true model so samples are drawn from this distribution and VM (WC) as the misspecified model and calculate the mle for $m = 10, 20$ with different values of the concentration parameter using the VM model. We find that the mle using the misspecified model leads to biased estimates especially when there is a high concentration parameter in the true model.

The details are as follows. We simulate 1000 datasets, each of size $n = 1000$ for each choice of $m \in \{10, 20\}$ and $\kappa \in \{1, 2.5, 10\}$ for $CDVM(m, \kappa, \mu = 0)$, and $\rho \in \{0.5, 0.6, 0.8\}$ for $CDWC(m, \rho, \mu = 0)$. We compute the mle from the conditionalized discrete model and compare it with that obtained from the continuous model. Table 8 shows the comparison of bias, standard deviation(sd) and mean squared error (mse)calculated based on the 1000 simulated datasets under different scenarios of m and the concentration parameter (κ or ρ). Part (a) of the table compares CDVM with VM, and part (b) compares CDWC with WC. We used the “CircStats” library in R to compute the mle under von Mises and wrapped Cauchy models.

In part (a) of Table 8 we can see that for $m = 10$, the bias, sd and mse for CDVM are comparable with that of VM for $\kappa = 1, 2.5$, but much smaller than that of VM for $\kappa = 10$ (i.e larger concentration parameter). However, as m increases to 20, the differences between $CDVM$ and VM decrease.

Table 8: Bias, SD and MSE for the mle for the concentration parameter (κ in (a), ρ in (b)) computed based on 1000 datasets, each of size $n = 1000$ for each choice of $m \in \{10, 20\}$ and $\kappa \in \{1, 2.5, 10\}$ for $CDVM(m, \kappa, \mu = 0)$, and $\rho \in \{0.5, 0.6, 0.8\}$ for $CDWC(m, \rho, \mu = 0)$

(a) MLE : CDVM vs. VM

		CDVM			VM		
m	κ	Bias	SD	MSE	Bias	SD	MSE
10	1	0.004	0.055	0.055	0.002	0.053	0.053
10	2.5	0.002	0.092	0.092	−0.003	0.092	0.092
10	10	0.008	0.385	0.385	1.513	0.690	1.663
20	1	0.000	0.052	0.052	−0.002	0.051	0.051
20	2.5	0.005	0.095	0.095	−0.000	0.094	0.094
20	10	0.035	0.432	0.433	0.042	0.433	0.435

(b) MLE : CDWC vs. WC

		CDWC			WC		
m	ρ	Bias	SD	MSE	Bias	SD	MSE
10	0.5	0.001	0.016	0.016	0.013	0.019	0.023
10	0.6	−0.000	0.013	0.013	0.058	0.024	0.063
10	0.8	−0.000	0.007	0.007	$\star$0.200	0.000	0.200
20	0.5	−0.001	0.017	0.017	−0.000	0.017	0.017
20	0.6	0.000	0.014	0.014	0.001	0.014	0.014
20	0.8	−0.000	0.007	0.007	0.063	0.020	0.066

$\star$ mle, on using the program mle.wrappedcauchy() in R converged to $\rho = 1$ for this case

In part (b) of Table 8, the differences are more pronounced. For $m = 10$, we see that the bias, sd and mse for CDWC are comparable with that of WC for $\rho = 0.5$, but bias, sd and mse are much less than that of WC for $\rho = 0.6$ and 0.8. With increasing m to 20, the differences decrease at $\rho = .6$ decrease but remain at $\rho = 0.8$ which indicates that a larger m would be required for the two to match. So, using the continuous model for discrete data with possibly high concentration and moderate m can lead to highly biased and inaccurate estimation of parameters. Further, how large m needs to be for a continuous approximation to work depends on the concentration parameter. For example, in problems such as mixture estimation, where the underlying nature of concentration parameters are *apriori* unknown, it is more appropriate to work with the discrete distribution, as inferences from (the approximate) continuous model can be misleading.

Case 3. Power of the Test of Uniformity Here we will concentrate on testing uniformity on highly dispersed data, as we have for our roulette data (Section 4.1.1). The test based on a discrete distribution, CDVM or CDWC, and the Rayleigh test based on a continuous distribution, will lead to similar conclusions (see the supplement, Section 5).

Recall that the test of uniformity for a discrete location family ((2.27) in the paper) is formulated as $H_0 : \tau = 0$ vs. $H_1 : \tau \neq 0$, where $\tau = \kappa$ for CDVM and $\tau = \rho$ for CDWC. This can be tested based on the discrete distribution using the test statistic T as discussed in Section A5. An alternative approach is to use the Rayleigh test based on test statistic $\bar{R}$. Table 9 shows the power computed for a 5% test based on either statistic at a specific alternative, viz. $(\kappa = 0.05, \mu = 0)$ for CDVM and $(\rho = 0.03, \mu = 0)$ for CDWC, for each of different choices of data-size $n \in \{1000, 10000\}$ and $m \in \{10, 37\}$. The 5% critical values and the power for T are computed based 10000 simulated datasets under the null and alternative hypothesis respectively, from the conditionalized discrete distribution (CDVM in (a) and CDWC in (b) with a given value of m) of the given data-size n. The 5% critical values and the power for $\bar{R}$ are computed based 10000 simulated datasets under the null and

Table 9: shows the power computed for a 5% test based on the statistics T and $\bar{R}$ at a specific alternative, viz. $(\kappa = 0.05, \mu = 0)$ for CDVM and $(\rho = 0.03, \mu = 0)$ for CDWC, for each of different choices of data-size $n \in \{1000, 10000\}$ and $m \in \{10, 37\}$. The 5% critical values and the power for T are computed based 10000 simulated datasets under the null and alternative hypothesis respectively

(a) CDVM: $H_0 : \kappa = 0 \quad H_1 : \kappa \neq 0$

Power for 5% test at $\kappa = 0.05, \mu = 0$

n	m	T	Rayleigh ($\bar{R}$)
1000	10	0.1445	0.1606
1000	37	0.1637	0.1668
10000	10	0.8976	0.8967
10000	37	0.8997	0.8992

(b) CDWC: $H_0 : \rho = 0 \quad H_1 : \rho \neq 0$

Power for 5% test at $\rho = 0.03, \mu = 0$

n	m	T	Rayleigh ($\bar{R}$)
1000	10	0.207	0.212
1000	37	0.214	0.209
10000	10	0.974	0.976
10000	37	0.976	0.976

alternative hypothesis respectively, from the continuous distribution (VM in (a) and WC in (b)) of the given data-size n. For both CDVM and CDWC, we find that the power of the test based on T is comparable to that based on $\bar{R}$ (i..e Rayleigh).

We now give some general remarks

REMARK 1. A key takeaway from these comparison studies is that discreteness of the data cannot be ignored in general and such data may need to be modeled using discrete circular distributions. At least, these discrete distributions provide a bench mark to assess any loss incurred in using continuous distribution.

REMARK 2. MD distributions are more relevant for the rounded circular data, whereas for any naturally circular discrete data, CD distributions are more appropriate.

REMARK 3. For $\lim m \to \infty$, the CD and MD distributions tend to their parent distribution as these are the Riemannian sums.

We have dealt here with the model misspecification problem between continuous and discrete distributions as this is the main focus of the paper; at the same time we have used both CDVM and CDWC in our comparison study wherever relevant. Incidentally, Mardia and Sriram (2020) show comparison among various different discrete circular models. For example, a comparison study of choices among conditionalized discrete distributions is based on divergence measures such as Kullback-Leibler, L_1 and L_2. It is found that CDVM and CDWC can be very different from each other but the CDVM and conditionalized discrete wrapped normal are very close to each other. These results are similar to the well known results for their respective continuous parent circular distribution: the von Mises, wrapped normal and wrapped Cauchy distributions.

6 Discussion

We have proposed flexible families of circular discrete distributions encompassing well established continuous circular distributions, such as von Mises and wrapped Cauchy. Our analysis of model misspecification (Section 5) highlights the importance of using discrete circular models for discrete data. We have selected the marginalized and conditionalized approaches for our analysis, but other constructions such as the centered wrapped families can be explored further. Also one can further explore the Beran family, in particular $\mathcal{B}_2$ and $\mathcal{B}_3$ given in Section 2.1. We have derived some insightful theorems interrelating the different methods of constructions. In particular, we have given an interesting characterization that under some regularity conditions, marginalized and conditionalized discrete distributions will be the same if and only if the parent circular distribution is uniform (leading to the discrete uniform). The marginalized and conditionalized families of distributions have a significant potential for further development beyond Directional Statistics. For example, we have also proved a characterization on the line that under some conditions, these two approaches lead to the same discrete distribution if and only if the parent is the exponential distribution (see supplement Section A7).

We note that some properties of the parent distributions such as unimodality and symmetry are inherited by the circular MD and CD. Also, maximum entropy characterization of the von Mises distribution is inherited by CDVM. However not everything carries over, namely, the normalizing constant for the CDVM depends on both the parameters, the Rayleigh test is no longer the likelihood ratio test for CDVM (supplement Section A5.1), and for CDWC the trigonometric moments are not as simple.

It is worth noting that Karl Pearson recognized that the roulette wheel data goes beyond coin tossing data experiments and raises some difficult inference problems in assessing unbiasedness (see Supplement A4 for more details). Perhaps due to the unavailability of adequate discrete circular models , it did not have much impact at the time. Also, bias in a roulette wheel can be due to a "tilt" or can due to "wobble" as we have seen in our examples. It could be said that wherever there is a wheel, there is inherently a natural circular discrete data, for example, a wheel used in some TV shows-Wheel of Fortune and other shows, in Industries (bicycle wheel, umbrella, and so on).

The field is full of new challenges in statistical methodology, for example, the marginalized and conditionalized approaches are amenable to extensions. We have outlined some extensions in Mardia and Sriram (2020), namely, alternative approaches to allow for an irregular lattice support, and extensions to higher manifolds such as the torus. However, it turns out that regular discretization on the sphere is not straightforward (see supplement, Section A6), and there can be multiple ways of constructing conditionalized discrete distributions.

Our overall recommendation from this paper is that "If you have circular discrete data, you should start with a discrete model".

7 Supplementary material

Online supporting material for this paper includes a supplement pdf file that gives details as referenced in the main paper.

Acknowledgments. We wish to thank Arthur Pewsey for his help on some R queries related to his book Pewsey et al. (2013), to Shogo Kato for confirming a query, and to Peter Green, John Kent, Arnab Laha, Florian Pfaff and John Wootton for helpful discussions. We are deeply grateful to Mihael Perman for kindly providing us with the real data (roulette data 2 and 3) related to a casino in Slovenia, and Germaine Cornelissen-Guillaume for the real data on acrophase. We appreciate Neil Spencer for making us aware of some other relevant work on roulette and are most grateful to Colin Goodall for carefully reviewing this current manuscript. The first author would also like to thank the Leverhulme Trust for the Emeritus Fellowship. We are grateful to the referee for the helpful comments.

Compliance with Ethical Standards. This is to confirm that there is no conflict of interest.

References

ALZAATREH, A., LEE, C. and FAMOYE, F. (2012). On the discrete analogues of continuous distributions. *Stat. Methodol.* **9**, 589–603.

BAINES, A. (1990). Testing a roulette for operational randomness/bias (link). University of Leeds Industrial Services Ltd. Report C3034.

BERAN, R. (1979). Exponential models for directional data. *Ann. Stat.* **7**, 1162–1178.

BHARATH, K. and DEY, D. (2019). Statistics on non-Euclidean Spaces and Manifolds [Special issue]. Sankhyā A **81**.

CARLIN, B. P. and LOUIS, T. A. (2009). *Bayesian methods for data analysis, 3rd edn.* CRC Press, Taylor and Francis Group, Boca Raton.

CHAKRABORTY, S. (2015). Generating discrete analogues of continuous probability distributions-a survey of methods and constructions. *J. Stat. Distrib. Appl.* **2**, 1–30.

ETHIER, S. N. (1982). Testing for favorable numbers on a roulette wheel. *J. Am. Stat. Assoc.* **77**, 660–665.

FERNÁNDEZ-DURÁN, J. J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics* **60**, 499–503.

FISHER, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.

FISHER, N. I., LEWIS, T. and EMBLETON, B. (1987). *Statistical analysis of spherical data*. Cambridge University Press, Cambridge.

GIRIJA, S. V. S., SRIHARI, G. V. L. N. and SRINIVAS, R. (2019). On discrete wrapped cauchy model. *Math. Theory Model.* **9**, 15–23.

IMOTO, T., SHIEH, G. S. and SHIMIZU, K. (2020). Discrete circular distributions with applications to shared orthologs of paired circular genomes. *Comput. Model. Eng. Sci.* **123**, 1131–1149.

INUSAHA, S. and KOZUBOWSKI, T. J. (2006). A discrete analogue of the Laplace distribution. *J. Stat. Plan. Inference* **136**, 1090–1102.

JAMMALAMADAKA, R. and SENGUPTA, A. (2001). *Topics in circular statistics*. Chapman and Hall/CRC.

JOE, H. (2014). *Monographs on applied statistics and probability*. CRC Press, Taylor Francis Group, Chapman and Hall.

KAGAN, A. M., LINNIK, Y. V. and RAO, C. R. (1973). *Characterization problems in mathematical statistics*. Wiley, New York.

KATO, S. and JONES, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika* **102**, 181–190.

KEMP, A. W. (1997). Characterizations of a discrete normal distribution. *J. Stat. Plan. Inference* **63**, 223–229.

LAHA, A. K. and GUPTA, A. (2011). Statistical quality control of directional data. In *International Conference on Advanced Data Analysis Business analytics and Intelligence, Ahmedabad, India. 8–9 January 2011 (link)*.

LEY, C. and VERDEBOUT, T. (2017). *Modern directional statistics*. Chapman and Hall/CRC.

LEY, C. and VERDEBOUT, T. (2018). *Applied directional statistics, modern methods and case studies*. Chapman and Hall/CRC.

LOMBARD, F. and MAXWELL, R. K. (2012). A cusum procedure to detect deviations from uniformity in angular data. *J. Appl. Stat.* **39**, 1871–1880.

MAKSIMOV, V. M. (1967). Necessary and sufficient statistics for the family of shifts of probability distributions on continious bicompact groups (english-tranlation-link(last-checked-22-03-2021)). *Teor. Veroyatnost. i Primenen.* **12**, 307–321.

MARDIA, K. V. (1972). *Statistics of directional data*. Academic Press, London.

MARDIA, K. V. (1975a). Statistics of directional data (with discussion). *J. R. Stat. Soc. B* **37**, 349–393.

MARDIA, K. V. (1975b). Characterizations of directional distributions. In *Statistical Distributions in Scientific Work* (G. P., Patil, S. Kotz and J. K. Ord, eds), vol. 3, pp. 365–385. Reidel, Dordrecht (55, 171, 262).

MARDIA, K. V. (2021). Comments on: Recent advances in directional statistics. *Test* **30**, 59–63.

MARDIA, K. V. and JUPP, P. E. (2000). *Directional statistics*. Wiley.

MARDIA, K. V., KENT, J. T. and LAHA, A. K. (2016). Score matching estimators for directional distributions. arXiv:1604.08470.

MARDIA, K. V. and SRIRAM, K. (2020). Families of discrete circular distributions with some novel applications. Arxiv (version 1) arXiv:2009.05437.

MASTRANTONIO, G., JONA LASINIO, G., MARUOTTI, A. and CALISE, G. (2019). Invariance properties and statistical inference for circular data. *Stat. Sin.* **29**, 67–80.

PAPADATOS, N. (2018). The characteristic function of the discrete Cauchy distribution. arXiv:1809.09443.

PEARSON, K. (1894). Science and monte carlo. *Fortn. Rev., New Series* **55**, 183–193.

PEARSON, K. (1897). The scientific aspect of Monte Carlo roulette. In *The Chances of Death and Other Studies in Evolution*, Vol 1, Chapter II, pp. 42–62. Revised version of Pearson (1894). London: Edward Arnold.

PEWSEY, A. (2004). The large-sample joint distribution of key circular statistics. *Metrika* **60**, 25–32.

PEWSEY, A. and GARCÍA-PORTUGUÉS, E. (2021). Recent advances in directional statistics. *Test* **30**, 1–58.

PEWSEY, A., NEUÄUSER, M. and RUXTON, G. D. (2013). *Circular statistics in R.* Oxford University Press, Oxford.

PLACKETT, R. L. (1983). Karl pearson and the chi-squared test. *Int. Stat. Rev.* **51**, 59–72.

RAO, D. A. V. and GIRIJA, S. V. S. (2020). *Angular statistics.* CRC Press, Taylor and Francis Group, Boca Raton.

SPENCER, N. H. (2009). Overcoming the multiple-testing problem when testing randomness. *J. R. Stat. Soc. (Ser. C)* **58**, 543–553.

SZABŁOWSKI, P. J. (2001). Discrete normal distribution and its relationship with Jacobi theta functions. *Stat. Probab. Lett.* **52**, 289–299.

WATSON, G.S. and BERAN, R.J. (1967). Testing a sequence of unit vectors for serial correlation. *J. Geophys. Res.* **72**, 5655–5659.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kanti V. Mardia
University of Leeds, Leeds, UK
University of Oxford, Oxford, UK
E-mail: k.v.mardia@leeds.ac.uk

Karthik Sriram
Indian Institute of Management
Ahmedabad, Ahmedabad, India

Paper received: 28 April 2022; accepted 26 September 2022.