Distribution And Time-Series Modelling Of Ordinal Circular Data

Technical Report No. ASU/2017/1
Dated 15 February, 2017

Jayant Jha

and

Atanu Biswas

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Distribution And Time-Series Modelling Of Ordinal Circular Data

Jayant Jha and Atanu Biswas
Applied Statistics Unit, Indian Statistical Institute
203 B. T. Road, Kolkata - 700 108, India
E-mails: jayantjha@gmail.com and atanu@isical.ac.in

Abstract

In the present paper, we discuss the distribution of a circular random variable having positive probability only on uniformly spaced directions. We used wrapping of geometric distribution to construct two discrete distributions on the circle. The distributional properties and estimation procedures for the underlying model parameters are discussed. Some tests for the distribution are proposed. Simulation studies are done to illustrate the distribution. Then, using the distribution, we carry out the time-series analysis of wind direction data where the sample space also contains the no wind state.

Some keywords and phrases: Discrete Circular Distribution ; Wrapping; Geometric Distribution

1 Introduction

The random variables having the sample space as the angles are called circular or directional random variables. Due to the difference in topological structure, the study of circular random variables is very different from the linear random variables. The usual definitions for different statistics such as mean and variance in the linear setup does not hold for the circular setup. Some distributions have been proposed to analyze the continuous circular random variables. Mardia and Jupp (2000) mentions some distributions for the continuous circular random variables. Wrapping is a very common approach for the distribution modelling in which the probability density of a linear random variable is wrapped to get the probability densities for the circular random variables. If the pdf for a linear random variable X is $g_X(x)$, then the wrapping makes the pdf of a circular random variable Θ as $f_\Theta(\theta) = \sum_{k=-\infty}^{k=\infty} g(\theta + 2\pi k)$. The examples are Wrapped Normal distribution, Wrapped Cauchy distribution etc. Due to the continuity of the probability density function, the wrapping process in the continuous setup maintain the continuity at the boundaries of the circular random variables i.e. $f(\theta) = f(\theta + 2\pi) \forall \theta \in [0, 2\pi)$.

The categorical random variables are of two kinds: nominal and ordinal. The analysis of nominal data means that there is no particular ordering among the categories and hence, such kind of datasets can be analysed by multinomial distributions and this procedure is exactly the same for both the linear and circular random variables. The ordinal random variables have different meanings in the circular setup and the linear setup. For the ordinal random variable in the linear setup, the categories $C_1, C_2, \ldots C_k$ can all be ordered such that we may write $C_1 < C_2 < \ldots < C_k$ for the categories. In the circular setup, this can not be written because if we start from the first category and move over the circle after the last category, again we shall reach the first category. Thus, the neighbouring category of C_1 are C_2 and C_k unlike the linear categories, where C_1 is the farthest from C_k. An example for the circular random variables is the ordering of the 4 directions such that if North(N), East(E) , South(S) and West(W) is the sample space, and if we start from North and move over the circle in clockwise fashion, we move to East, South, West and again North. Due to this kind of circular ordering, there is a need for different kinds of distribution for discrete ordinal circular random variables and discrete ordinal linear random variables.

Using wrapping of linear discrete random variables, the Wrapped Poisson distribution has been considered in Levy(1939) and Mardia (1972). The finite summation form of this pdf involves the number of summation terms equal to the cardinality of sample space and is given in Ball and Blackwell(1992). But even in the finite form, the functional form of the pdf is intractable. Jayakumar and Jacob (2012) mentions a simpler discrete distribution based on Wrapping the Skew Laplace Distribution on Integers. Jacob and Jayakumar (2013) mentions a distribution based on wrapping of the geometric distribution. Both these distributions consider the skewness but a big drawback of the probability mass function is that the 0 angle always has the highest probability mass and thus, we have to define the angle with the largest probability as the angle 0 and then only the analysis can be performed.

The time series analysis of wind direction is very different from the time series analysis of linear random variables because wind direction is measured in angles. Downs and Mardia (2000) contains a comprehensive review of the different techniques applied for the time-series analysis of circular random variables. For continuous circular random variables, such techniques work very well. But, generally the measurement of wind direction contains additional constraints. One of the constraints is that, in many cases, in place of the exact wind direction angle, we observe only one of the finite number of equally spaced directions. For example: North, East, West or South. An additional constraint on the wind direction data is the observation of no wind on many of the days. We shall call such observations as null observations. Thus, in the analysis of the time series of wind directions, we have an ordinal circular data mixed with null observations.

Time series analysis for continuous circular random variable has been mentioned in Kato (2010), which uses some nice distributional properties of Wrapped Cauchy distribution for AR(1) time-series modelling. Our case is different from them in the sense that our dataset contains discrete as well as null observations. Even the concept of null observations used in our case is different from missing observations and censored observations in the sense that a null observation is an actual observation containing information about the occurrence of no wind on a particular day. Thus, we have a sample space which contains uniformly spaced sample points as well as null and on the basis of the observations, our goal of this paper is to predict if there will be a wind in the future time point and if there is wind, then what is the pmf of the wind direction on that time point.

In this paper, our aim is to introduce two alternative symmetric and skewed discrete distribution on the circle with a closed form probability mass function, trigonometric moments and independence of the 0 angle based on the wrapping of the discrete geometric distribution. The model is introduced in Section 2. Section 3 discusses some distributional properties related to the model. In the same section, we mention the tests related to the underlying model parameters. In Section 4, we discuss the time-series modelling based on the Wrapped Geometric distribution with null state in the sample space. Then, we carry out maximum likelihood estimation for underlying model parameters. Later, some tests related to the model are mentioned. Section 5 shows some simulation results. A real data analysis is carried out in the next section as an example. Section 7 concludes.

2 Symmetric discrete circular distributions

2.1 Wrapped symmetric geometric distribution

For a discrete linear distribution X with support on $x = \frac{2\pi k}{m}$, $m = 0, \pm 1, \ldots, \pm(m-1), \pm m, \pm(m+1), \ldots$, with

$$p_X\left(\frac{2\pi j}{m}\right) = P\left(X = \frac{2\pi j}{m}\right), \tag{2.1}$$

the corresponding wrapped discrete circular distribution Y can be defined on the support $\left\{0, \frac{2\pi}{m}, \ldots, \frac{2\pi(m-1)}{m}\right\}$ as

$$p_Y\left(\frac{2\pi k}{m}\right) = P\left(Y = \frac{2\pi k}{m}\right) = \sum_{j=-\infty}^{\infty} p_X\left(\frac{2\pi}{m}(k+mj)\right), \quad k = 0, 1, \ldots, m-1, \tag{2.2}$$

Specifically we want a distribution of Y such that

$$p_Y\left(\frac{2\pi k}{m}\right) = p_Y\left(\frac{2\pi(m-k)}{m}\right), \quad k = 0, 1, \cdots, m-1.$$

With this objective, we consider $p_X\left(\frac{2\pi j}{m}\right)$ as the convex combination of two random variables X_1 and X_2 such that $p_{X_1}(\cdot)$ is same as (2.1) and

$$p_{X_2}\left(\frac{2\pi k}{m}\right) = p_X\left(2\pi - \frac{2\pi k}{m}\right).$$

Therefore,

$$p_X\left(\frac{2\pi k}{m}\right) = \frac{1}{2}p_{X_1}\left(\frac{2\pi k}{m}\right) + \frac{1}{2}p_{X_2}\left(2\pi - \frac{2\pi k}{m}\right).$$

Writing

$$p_X\left(\frac{2\pi k}{m}\right) = pq^k, \quad k = 0, 1, \ldots,$$

the pmf of Y comes out to be

$$\begin{aligned} p_Y\left(\frac{2\pi k}{m}\right) &= \frac{1}{2}\sum_{j=0}^{\infty} pq^{k+mj} + \frac{1}{2}\sum_{j=0}^{\infty} pq^{m-k+mj} \\ &= \frac{1}{2}\left(\frac{p}{1-q^m}\right)\left(q^k + q^{m-k}\right). \end{aligned} \tag{2.3}$$

It is like wrapping and then taking the simple average of two geometric distributions with same parameter, starting a the point '0', one counter-clockwise and the other clockwise. The basic objective is to find a distribution on the circle so that the probability mass function (pmf) at $\frac{2\pi k}{m}$ is same as the pmf at $\left(2\pi - \frac{2\pi k}{m}\right)$.

To make the pmf independent of the 0 angle and to make the distribution symmetric about the mode, we can modify the pmf as

$$p_Y\left(\frac{2\pi k}{m}\right) = \left(\frac{p}{1-q^m}\right)\frac{1}{(1+q)}\left(q^{\zeta} + q^{m-\zeta}\right). \tag{2.4}$$

where $\zeta = (k-a)(mod)m$ and the parameter space of a is $\{0, 1, \ldots (m-1)\}$. We call the distribution (2.4) a Wrapped Symmetric Geometric (WSG) distribution. Thus, the pmf denoted by (2.4) is symmetric about the direction $\frac{2\pi a}{m}$. As the function $f(x) = q^x + q^{m-x}$ has the maximum at x=0 for any $m > 0$, thus, $\frac{2\pi a}{m}$ is also the mode of the pmf. Thus, (2.4) is a unimodal symmetric discrete circular distribution.

2.2 Asymmetric Wrapped Geometric Distribution

A generalization of the pmf mentioned in 2.4 can be done by taking the parameter value $a \in [0, m)$. This causes asymmetry in the pmf as the mode will again be either $[a]mod(m)$ or $[a+1]mod(m)$ or both, where [.] is the greatest integer less than or equal to a but the new pmf will not remain symmetric about the modal direction unless a is an integer. Thus,

$$p_Y(\frac{2\pi k}{m}) = C(q^{(k-a)mod(m)} + q^{(a-k)mod(m)})$$

where C is proportionality constant such that the sum of the pmf is 1.
C can be calculated in the following way : Denoting the fractional part of a as b, then

$$\sum_{k=0}^{m-1} Pr(Y = \frac{2\pi k}{m}) = 1$$

$$\Rightarrow \frac{1}{C} = \sum_{l=0}^{m-1}(q^{b+l} + q^{m-b-l})$$

. Hence, $C = \frac{1}{(q^b+q^{1-b})}\frac{(1-q)}{(1-q^m)}$.
We shall call this distribution Generalised Wrapped Geometric (GWG) distribution.

3 Parameter Estimation

3.1 Method of Moments Estimation

An equivalent representation of the circular data is in the form of unit complex numbers. If Y is a circular random variable, then $Z = e^{iY}$ is complex valued random variable taking values on the unit circle giving a unique representation for Y. The trignometric moments $\phi_r(Y) = E(Z^r)$, where r is a positive integer can then be calculated for method of moment estimation.

For WSG,

$$E(Z^r) = e^{i\gamma a}\frac{(1-q)^2}{(1+q^2-2q\cos\gamma)}$$

where, $\gamma = \frac{2\pi r}{m}$.
Hence, note that $Arg\{E(Z^r)\} = \gamma a$.
Also, $\frac{(1-q)^2}{1+q^2-2qcos(\gamma)}$ is monotonically decreasing in q for any non-zero value of γ for $q \in [0,1]$. Hence, $|E(Z^r)|$ is a monotonically decreasing function in q for any non-zero value of γ. Thus, method of moments will give a unique estimate for a and q and thus, it can be easily used to find the estimates of parameters.
For GWG,

$$E(Z^r) = e^{i\gamma a}\frac{(1-q)(q^b + q^{2-b} - e^{-i\gamma}(q^{1+b} + q^{1-b}))}{(1+q^2-2q\cos\gamma)(q^b + q^{1-b})}$$

where, again $\gamma = \frac{2\pi r}{m}$.
In this case, As $m \to \infty$, $ArgE(Z^r) \to a\gamma$.

3.1.1 Estimation of a by maximum cell frequency for WSG

In case of WSG, the parameter space $\mathcal{A} = \{0,\ldots,m-1\}$ is the parameter space of a which is discrete. Thus, the maximum likelihood estimator will not give a proper confidence interval. Thus, we have to define a new estimator for a.
When X follows a multinomial distribution with probabilities $p_0,\ldots p_{m-1}$, then $\frac{n_k}{n} \overset{a.e}{\to} p_k$ where n_k represents the number of sample observations in the kth state and n is the sample size. Then, if we define the estimator $\delta = argmax_{j=\{0,\ldots m-1\}} n_j$ and let us assume without loss of generality that $p_0 = max\{p_j\}$, then $\frac{n_0}{n} \overset{a.e}{\to} p_0$ and $\frac{n_k}{n} \overset{a.e}{\to} p_k$ for all j. Thus, $\delta = 1 a.e.$. Thus, the estimator δ will correspond to the direction of the mode i.e. a almost surely.
Next, we shall find the exact distribution of δ under the assumption that there is a unique

cell having the maximum number of observations implying δ is unique. Then, by using Corrado(2011), one can get the exact distribution for the maximum frequency in the case of multinomial distribution with known cell frequencies. Thus, we know $P(x) = Pr(Nmax = x)$ where, Nmax is the maximum number of balls in any cell. Now, let the maximum number of balls are in the jth cell. Then,

$$Pr(\delta = j) = \frac{1}{\sum_{k=0}^{m-1} Pr(\delta = k)} \sum_{c=\frac{n}{m}}^{n} P(c) \frac{p_j^c}{(1-p_j)} \sum_{c'=1}^{c-1} P'(c') \tag{3.5}$$

. which means that

$$Pr(\delta = j) = \frac{1}{\sum_{k=0}^{m-1} Pr(\delta = k)} \sum_{c=\frac{n}{m}}^{n} Pr(Nmax = c) \times P(n_j = c, n_s < c/Nmax = c)$$

$$\Rightarrow Pr(\delta = j) = \frac{1}{\sum_{k=0}^{m-1} Pr(\delta = k)} \sum_{c=[\frac{n}{m}]}^{n} Pr(Nmax = c) \sum_{c'=1}^{c-1} Pr(Nmax' = c')$$

where, $Nmax'$ is the maximum among all the cells other than j. Thus, eq.(3.5) gives the required probability. In our setup, we do not know the exact cell frequencies. But, we can find a consistent estimator of q by $\sum\sum_{1\leq<j\leq n} \frac{U_{i,j}}{N}$ which converges in probability to a monotonic function of q. Thus, we approximate the exact cell frequencies by using this estimator of q. Then, as explained in the previous paragraph, we can determine the asymptotic mass function of δ. To estimate the p.m.f of δ, bootstrapping can also be used by taking many samples for the estimated value of q and fixed value of a and getting an empirical estimate of the pmf of $Pr(\delta = j)$ for $j = 0, 1, \ldots (m-1)$.

3.1.2 Estimation of a by mean angle for WSG

Note that the estimation from the maximum cell frequency can only be carried out when there is a unique cell having maximum number of observations. Hence, for real data sets, it is recommended that the sample will be taken till the time the cell with the maximum number of observations is unique. If this is not the case, then the mean angle of the observations should be considered for the estimation.

It has been shown in Section 3.1 that $Arg\{E(Z)\} \to a$. Let, $\phi = Arg(\sum_{i=1}^{n} Z_n)$. Then, if we take the estimator of a, $\hat{a_n}$ to be the angle in the discrete parameter space of a from which ϕ is nearest, then we get an estimator of a. If we define, $d = |\phi - a|$, then the distribution of d does not depend on a, d just measures the distance from the mode whose probability will depend only on q. Thus, approximating the probability by q_n, we can approximate the probability that $Pr|\hat{a}_n - a| < 0.5$ and hence, the probability of true convergence. For the confidence interval of the probability bootstrapping can be used similar to the case of estimation by maximum cell frequency.

3.1.3 Estimation of a by posterior probabilities for WSG

Given the data, one can also estimate the posterior pmf of a in Bayesian setting. Let, Π represent prior probability of a. Then, given the dataset X, the posterior probabilities of a can be found in the following way.

$$P(a = j/X) = \frac{P(X/a = j)P(a = j)}{P(X)}$$

. Thus, the posterior pmf of a can be estimated for a given value of q. For unknown q, we use the estimated value of q to approximate the posterior probabilities.

3.2 Tests

3.2.1 One sample test for WSG when a is known or unknown

Let $Y_1, \dots Y_n$ are i.i.d WSG(a,q). let, p_i denote the probability of ith state and if we define $U_{i,j} = cos(Y_i - Y_j)$, then $U_{i,j}$ and $U_{k,l}$ are independent for $i \neq j \neq k \neq l$. Given a state i, let, $j = i + \alpha * mod(m)$ denote the state such that $\phi_{i,j} = \alpha$ represents the angle between the states i and j. Now, for any a,

$$P\{U_{i,j} = \cos(\alpha)\}/P\{U_{i,k} = cos(\beta)\} = \sum_{r=0}^{m-1} p_r p_{r+\alpha*mod(m)} = P(U_{i,j} = \alpha)$$

. Thus, $U_{i,j}$ and $U_{i,k}$ are pairwise independent and hence, we can find the asymptotic distribution of the pairwise summation of $U_{i,j}$'s. Also, we can find $E(U_{i,j})$ and $Var(U_{i,j}$ for WSG(a,q). Hence, by Central Limit Theorem,

$$\frac{\sum\sum_{1\leq<j\leq n} \frac{U_{i,j}}{N} - E(U_{i,j})}{Var(U_{i,j})} \sim N(0,1) \tag{3.6}$$

where, $N = \binom{n}{2}$. Thus, an asymptotic test for $q = q_0$ for WSG can be performed on this basis. Let, we want to test $H_0 : q = q_0$ vs $H_1 : q \neq q_0$. Then, we can find out the confidence interval (a, b) for the test at some pre-defined level of significance under H_0 and hence, test the hypothesis.

3.2.2 Two-sample test for WSG when a is known or unknown

Using (3.6), we can define asymptotic test for $H_0 : q_X = q_Y$ vs $H_1 : q_X \neq q_Y$, where X and Y follow $WSG(q_X, a_X)$ and $WSG(q_Y, a_Y)$ respectively. Denoting $T(X) = \frac{\sum\sum_{1\leq<j\leq n} \frac{U_{i,j}(X)}{N} - E(U_{i,j}(X))}{Var(U_{i,j}(X))}$ and $T(Y) = \frac{\sum\sum_{1\leq<j\leq n} \frac{U_{i,j}(Y)}{N} - E(U_{i,j}(Y))}{Var(U_{i,j}(Y))}$, we can get under H_0, $T(X) - T(Y) \sim N(0, 2)$ and thus, we can find the confidence interval for the test.

3.2.3 Likelihood ratio tests for GWG

For GWG, the likelihood ratio tests based on MLE can be done in all the general cases as the parameter space for a and q both are continuous. We will reiterate the results in that case. Let L_0 is the maximum likelihood under H_0 and L_1 is the maximum likelihood under $H_0 \cup H_1$, then $\Lambda = -2\log(L_0/L_1) \sim \chi^2_\nu$, where ν is the difference between the number of extra parameters to be estimated.

4 Time-Series Model

Let, $\{Y_t\}$ be the time series where, the sample space of Y_t is $\Omega = \{0, 1, \ldots m-1\}$ for all t. Then, using WSG, we can define the conditional probability model for Y_t/Y_{t-1} as the following:

$$Y_t/Y_{t-1} \sim WSG(Y_{t-1}, q)$$

Then, this formulation ensures that the wind direction at the next time-point follows WSG with mean direction the same as the observed direction on the time t-1 and q denotes the variability. The higher the value of q, the more is the variability of Y_t. This is a Markov-Chain where the distribution of $Y_t/Y_{t-1}, \ldots$ is the same as the distribution of Y_t/Y_{t-1}.

But, generally in the wind direction observations the sample space is different. In the wind direction dataset, we have some observations when there is no wind. Thus, we shall try to incorporate these conditions in our model. For this purpose, we shall use a mixture distribution which gives some probability to the null and the other probabilities are divided into the probability of observing other directions. Moreover, we incorporate the consecutive number of time-points when there is no wind in our model.

Let, N denote the state of no wind. Then, the sample space of Y_t is $\Omega \cup N$. Thus, the model can be formulated in the following way:

When $X_{t-1} \in \Omega$,

$$X_t/X_{t-1} \sim p\mathbb{1}_{X_t=N} + (1-p)\mathbb{1}_{X_t\in\Omega} WSG(X_{t-1}, q) \tag{4.7}$$

When $X_{t-1} = N$,

$$X_t/X_{t-1} \sim p(h)\mathbb{1}_{X_t=N} + (1-p(h))\mathbb{1}_{X_t\in\Omega} WSG(X_{t-h}, q(h)) \tag{4.8}$$

where, $h = min\{b : X_{t-b} \in \Omega\}$ and $p(h), q(h) \in [0, 1]$ are non-decreasing functions of h and $\mathbb{1}$ denotes the indicator variable.

Thus, the pmf in the state X_t depends only on the previous state when there is a wind in the last state but when the last state is a state wih no wind, then it depends on h and the wind direction observed in the last day when there was wind in case. Now, the interpretation of $p(h)$ and $q(h)$ is such that if the last state is the null state (having no wind), then $p(h)$ denotes the probability of null in the next time point. Also, by making it a non-decreasing function of h, the probability of the next occurrence of null-state increases with the number of previous consecutive days of no wind observation. Thus, $p(h)$ satisfies the following conditions:

1. $0 \leq p(h) \leq 1$ for all $h = 0, 1, \ldots$.
2. $p(1) = p$.
3. $p(h)$ is a non-decreasing function of h.
4. $\lim_{h\to\infty} p(h) = 1$.

Similarly, $q(h)$ is also dependent on h and it increases with h meaning that the variability increases as the number of days for last observed wind day increases. Thus, $q(h)$ satisfies the following conditions:

1. $0 \leq q(h) \leq 1$ for all $h = 1, \ldots$.
2. $q(1) = q$.
3. $q(h)$ is a non-decreasing function of h.

4. $\lim_{h\to\infty} q(h) = 1$.

Following these conditions, we define

$$p(h) = \frac{p\alpha(h-1)+p}{p\alpha(h-1)+1}$$

and

$$q(h) = \frac{q\alpha(h-1)+q}{q\alpha(h-1)+1}$$

where, $\alpha \in [0,\infty)$. Note that when $\alpha = 0$, then $p(h) = p$, $q(h) = q$ for all h and then, the time series process becomes a homogeneous Markov chain and hence, the usual techniques for Markov Chain can be applied. But, the general model has 3 parameters $\alpha, p, q \in [0,\infty)\times[0,1]^2$. Note that, in our paper we have taken only α for easy of description of the model as well as computation. In most practical applications, this is not true and thus, we may use α_1, $\alpha_2 \in [0,\infty)^2$ in place of α for $p(h)$ and $q(h)$ respectively.

4.1 Discussion On Stationary Properties

In Discrete-state and discrete-time homogeneous Markov chain, let P denotes the transition probability matrix, then the stationary distribution of the states π can be found out by the equation πP=π. In the time-series model, if there is no null state, then the time series is a homogeneous Markov chain and the stationary distribution of the states can be found out in a similar manner. Moreover, as P is symmetric with respect to different states, the stationary probabilities of the states will be the same and the stationary probability of the states are $\Pi_i = \frac{1}{m}$ for all $i \in \{0,2,\ldots,m-1\}$ where m is the number of states. If we include the null state also and $\alpha = 0$, then also we can find the stationary probabilities directly. In fact, for any value of α , due to symmetricity of the transition probability matrix with respect to the non-null states, $\Pi_i = \frac{1-\Pi}{m}$ where Π is the stationary probability of the null state. Moreover, whenever $\alpha > 0$, $p(h)$ is a monotonically increasing function of h, thus, $\Pi_n > \xi_n$, where ξ_n is the limiting probability of the null-state when $\alpha = 0$ and Π_n is the limiting probability of the null-state when $\alpha > 0$. Thus, the lower bound on Π_n can be found by calculating the stationary distribution of the homogeneous markov chain with $\alpha = 0$. Also, it should be noted here that the stationary distribution remains the same for all values of α when $p(h) = p$ is a constant but q(h) is a monotonically increasing function of h. This is because in this case, the probabilities in transition probability matrix are symmetrically redistributed among the non-null states while the probability for reaching the null-state remains the same. The stationary probability for the null-state when p(h) is a constant can be found by the 2×2 transition probability matrix with $p_{ii} = p$ and $p_{ij} = 1 - p$.

Given the model and an initial state X_t, to compute the distribution at X_{t+h}, a different matrix at each state has to be calculated. This requires huge computation but if we assume a maximum value for h and call it h_{max}, then a lower and upper bound on the probabilities of each state can be very easily calculated by two homogeneous Markov chains with lower bound for p(h) as p and upper bound as $\frac{p\alpha(h_{max}-1)+p}{p\alpha(h_{max}-1)+1}$ and lower bound for q(h) as q and upper bound as $\frac{q\alpha(h_{max}-1)+q}{q\alpha(h_{max}-1)+1}$,and thus, by considering these two Markov chains, the bounds on probabilities can be directly computed.

4.2 Estimation of parameters

The estimation of the underlying model parameters can be performed using Maximum Likelihood Estimation procedure. For estimation, we have to consider an additional assumption in the model that at the first time point there is a wind direction. This assumption is necessary because we do not know the distribution of X_1/X_0 when X_0 is the null state because we do not know h. Thus, assuming that $X_0 \in \Omega$ and $X_0 = x_0$ is known, the pmf of $(X_1, \ldots X_n)$ can be written as:

$$L = f_{X_1, \ldots X_n}(x_1, \ldots x_n) = \Pi_{t=1}^{n} f(X_t = x_t / X_{t-1} = x_{t-1}; h)$$

Thus, the MLE's of the parameters can be found out using

$$\frac{dlogL}{d\theta} = 0$$

where, $\theta = (q, \alpha, p)$

If the MLE's do not take values on the boundaries of the parameter set and the MLE's are $\hat{\theta}_n$ then,

$$\sqrt{n}(\hat{\theta}_n - \theta) \sim N(0, \Sigma)$$

where Σ is the inverse of the Fisher Information matrix. Moreover,

$$\frac{d^2 logL}{dpdq} = 0$$

and thus, asymptotically $\hat{p}$ is independent of $\hat{q}$.

4.3 Testing of homogeneity

It has been mentioned before that when $\alpha = 0$, then the the time-series is a homogeneous Markov-chain and thus, the analysis becomes much easier as the transition probability matrix remains a constant. Thus, denoting

$H_0 : \alpha = 0$ and H_1: not H_0

and denoting by L_0, L_1, the maximum likelihood under null and alternate hypothesis respectively. Now, under H_0, a parameter of interest is on the boundary and hence, under H_0, $\Lambda = -2log\frac{L_0}{L_1}$ follows an equal mixture of 0 and χ^2_1 as mentioned in Self and Liang (1987).

5 Simulation

We carried out the simulation studies for the two distribution models and the time-series model. For both the WSG and GWG distributions, we considered 3 different values of q and a and 2 different values of number of categories (m) at two different sample sizes (n). The simulation results based on 10000 simulations for the WSG distribution are given in Table 1 and the simulation results based on 10000 simulations for the WSG distribution are given in Table 2. For the WSG distribution, we estimated the modal direction by the maximum cell frequency (δ_{max}) as well as the mean (δ_{mean}) as described in the previous section. We also estimated the mean and standard error (given in parentheses) over all simulations of $h(q) = E_q\{cos(\theta_i - \theta_j)\}$ which is a monotonically decreasing function of q. In all the tables, the true value of h(q) is also reported alongside the value of q. We have taken a=0 due to no loss in generality for all the simulations for WSG. Table 1 shows that as q decreases, the estimates become better and hence, for highly concentrated datasets and large sample sizes, we get very good estimates of the modal direction by both mode and the mean. At higher values of q, the estimates are not good. This is due to the fact as q gets larger, the difference in the probabilities of the cells decreases and the distribution becomes more uniform.

For the GWG distribution, we took three different values of q and for each value of q we took three different values of a at two sample sizes for two different values of m.

Table 1: Estimates of the parameters (with standard errors in parentheses) when the true distribution is WSG, m=5 and a=0 for q=0.25, 0.5, 0.75 and n=70, 200.

q=0.25;h(q)=0.3575			
$n = 200$	$\delta_{max} = (10000, 0, 0, 0, 0)$	$\delta_{mean} = (10000, 0, 0, 0, 0)$	$\hat{h(q)} = 0.3783(0.0005)$
$n = 70$	$\delta_{max} = (10000, 0, 0, 0, 0)$	$\delta_{mean} = (10000, 0, 0, 0, 0)$	$\hat{h(q)} = 0.3701(0.0008)$
q=0.5;h(q)=0.0699			
$n = 200$	$\delta_{max} = (9975, 15, 0, 0, 10)$	$\delta_{mean} = (9992, 3, 0, 0, 5)$	$\hat{h(q)} = 0.0704(0.0003)$
$n = 70$	$\delta_{max} = (9371, 287, 9, 13, 330)$	$\delta_{mean} = (9412, 295, 5, 1, 287)$	$\hat{h(q)} = 0.0557(0.0004)$
q=0.75;h(q)=0.0032			
$n = 200$	$\delta_{max} = (5300, 1551, 688, 735, 1726)$	$\delta_{mean} = (5248, 1929, 411, 474, 1938)$	$\hat{h(q)} = -0.0017(0.0008)$
$n = 70$	$\delta_{max} = (3699, 1758, 1107, 1276, 2160)$	$\delta_{mean} = (3799, 2162, 914, 973, 2147)$	$\hat{h(q)} = -0.0013(0.0002)$

Table 2: Estimates of the parameters (with standard errors in parentheses) when the true distribution is WSG, m=5 and a=0 for q=0.25, 0.5, 0.75 and n=70, 200.

q=0.25;h(q)=0.6296			
$n = 200$	$\delta_{max} = (10000, 0, 0, 0, 0, 0, 0, 0)$	$\delta_{mean} = (10000, 0, 0, 0, 0, 0, 0, 0)$	$\hat{h(q)} = 0.6241(0.0004)$
$n = 70$	$\delta_{max} = (10000, 0, 0, 0, 0, 0, 0, 0)$	$\delta_{mean} = (10000, 0, 0, 0, 0, 0, 0, 0)$	$\hat{h(q)} = 0.6137(0.0007)$
q=0.5;h(q)=0.2080			
$n = 200$	$\delta_{max} = (9991, 3, 0, 0, 0, 0, 0, 6)$	$\delta_{mean} = (10000, 0, 0, 0, 0, 0, 0, 0)$	$\hat{h(q)} = 0.2063(0.0004)$
$n = 70$	$\delta_{max} = (9491, 255, 1, 0, 0, 0, 1, 252)$	$\delta_{mean} = (9758, 118, 0, 0, 0, 0, 0, 120)$	$\hat{h(q)} = 0.1977(0.0007)$
q=0.75;h(q)=0.0156			
$n = 200$	$\delta_{max} = (6283, 1429, 260, 85, 57, 83, 277, 1526)$	$\delta_{mean} = (6557, 1610, 95, 18, 4, 17, 89, 1606)$	$\hat{h(q)} = 0.0104(0.0001)$
$n = 70$	$\delta_{max} = (3935, 1539, 700, 366, 288, 415, 770, 1987)$	$\delta_{mean} = (4342, 2074, 467, 159, 117, 157, 470, 2198)$	$\hat{h(q)} = 0.0002(0.0002)$

Table 3: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=5, q=0.25 and n=70, 200 for different values of a.

n=200	a=0	$\hat{q} = 0.2407(0.0002)$	$\hat{a*} = 0.0001(0.0031)$
	a=0.25	$\hat{q} = 0.2503(0.0003)$	$\hat{a*} = 0.3100(0.0036)$
	a=0.5	$\hat{q} = 0.2497(0.0003)$	$\hat{a*} = 0.6289(0.0023)$
n=70	a=0	$\hat{q} = 0.2338(0.0004)$	$\hat{a*} = 0.0000(0.0082)$
	a=0.25	$\hat{q} = 0.2509(0.0006)$	$\hat{a*} = 0.2971(0.0147)$
	a=0.5	$\hat{q} = 0.2490(0.0005)$	$\hat{a*} = 0.6261(0.0074)$

Table 4: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=5, q=0.5 and n=70, 200 for different values of a.

n=200	a=0	$\hat{q} = 0.4858(0.0004)$	$\hat{a*} = -0.0003(0.0122)$
	a=0.25	$\hat{q} = 0.5013(0.0005)$	$\hat{a*} = 0.2965(0.0256)$
	a=0.5	$\hat{q} = 0.5034(0.0005)$	$\hat{a*} = 0.6196(0.0243)$
n=70	a=0	$\hat{q} = 0.4776(0.0007)$	$\hat{a*} = 0.0012(0.0346)$
	a=0.25	$\hat{q} = 0.4991(0.0009)$	$\hat{a*} = 0.2923(0.0811)$
	a=0.5	$\hat{q} = 0.5080(0.0009)$	$\hat{a*} = 0.6015(0.1046)$

We estimated the mean and standard error (in parentheses) for q and circular mean and dispersion for $a* = \frac{2\pi a}{m}$. The results for the simulation are given in Table 3-8.
For the time series simulation, we chose 2 different sample sizes at a fixed value of p, q and 2 different values of α near the data analysis example. The results are shown in Table 9.

Table 5: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=5, q=0.75 and n=70, 200,1000 for different values of a.

n=200	a=0	$\hat{q} = 0.7046(0.0008)$	$\hat{a*} = -0.0021(0.2017)$
	a=0.25	$\hat{q} = 0.7140(0.0008)$	$\hat{a*} = 0.1461(0.2705)$
	a=0.5	$\hat{q} = 0.7186(0.0008)$	$\hat{a*} = 0.2961(0.3000)$
n=70	a=0	$\hat{q} = 0.6514(0.0009)$	$\hat{a*} = -0.0030(0.3667)$
	a=0.25	$\hat{q} = 0.6588(0.0010)$	$\hat{a*} = 0.1179(0.4215)$
	a=0.5	$\hat{q} = 0.6592(0.0010)$	$\hat{a*} = 0.2585(0.4624)$

Table 6: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=8, q=0.25 and n=70, 200 for different values of a.

n=200	a=0	$\hat{q} = 0.2433(0.0003)$	$\hat{a*} = 0.0003(0.0010)$
	a=0.25	$\hat{q} = 0.2520(0.0003)$	$\hat{a*} = 0.1858(0.0020)$
	a=0.5	$\hat{q} = 0.2494(0.0003)$	$\hat{a*} = 0.3942(0.0008)$
n=70	a=0	$\hat{q} = 0.2373(0.0004)$	$\hat{a*} = -0.0004(0.0025)$
	a=0.25	$\hat{q} = 0.2556(0.0005)$	$\hat{a*} = 0.1797(0.0026)$
	a=0.5	$\hat{q} = 0.2488(0.0005)$	$\hat{a*} = 0.3920(0.0031)$

Table 7: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=8, q=0.5 and n=70, 200 for different values of a.

n=200	a=0	$\hat{q} = 0.4929(0.0003)$	$\hat{a*} = -0.0005(0.0029)$
	a=0.25	$\hat{q} = 0.4996(0.0003)$	$\hat{a*} = 0.1959(0.0040)$
	a=0.5	$\hat{q} = 0.5000(0.0003)$	$\hat{a*} = 0.3906(0.0041)$
n=70	a=0	$\hat{q} = 0.4909(0.0005)$	$\hat{a*} = 0.0079(0.0046)$
	a=0.25	$\hat{q} = 0.4975(0.0005)$	$\hat{a*} = 0.2139(0.0098)$
	a=0.5	$\hat{q} = 0.500(0.0005)$	$\hat{a*} = 0.3970(0.0131)$

Table 8: Estimates of the parameters (with standard errors in parentheses) when the true distribution is GWG, m=5, q=0.75 and n=70, 200,1000 for different values of a.

n=200	a=0	$\hat{q} = 0.7411(0.0004)$	$\hat{a*} = 0.0578(0.0310)$
	a=0.25	$\hat{q} = 0.7435(0.0005)$	$\hat{a*} = 0.2928(0.0472)$
	a=0.5	$\hat{q} = 0.7459(0.0005)$	$\hat{a*} = 0.4451(0.0571)$
n=70	a=0	$\hat{q} = 0.7272(0.0007)$	$\hat{a*} = -0.0078(0.2732)$
	a=0.25	$\hat{q} = 0.7300(0.0007)$	$\hat{a*} = 0.3878(0.1043)$
	a=0.5	$\hat{q} = 0.7291(0.0007)$	$\hat{a*} = 0.5175(0.1171)$

Table 9: Estimates of the parameters (with standard errors in parentheses) for time series at different values of n and α with m=8

Parameters	n=200	n=50
$p = 0.4$	0.4106 (0.0007)	0.4216 (0.0012)
$q = 0.25$	0.2521 (0.0006)	0.2639 (0.0012)
$\alpha = 1.5$	1.7410 (0.0104)	1.8619 (0.0161)
$p = 0.4$	0.4048 (0.0006)	0.4114 (0.0010)
$q = 0.25$	0.2517 (0.0005)	0.2587 (0.0009)
$\alpha = 1$	1.0800 (0.0054)	1.2815 (0.0123)

Table 10: Cataract surgery dataset

Category	1	2	3	4	5
n_j	44	15	4	0	7

6 Data Analysis

The Data analysis for two sets of data: first is the cataract surgery data mentioned in Biswas et.al.(2016) and other from the wind direction data observed over two years in Kolkata taken from www.weatheronline.in was used to illustrate the distribution models mentioned in the previous sections.

6.1 Cataract surgery data

In Biswas et.al.(2016), they had the dataset for axis of astigmatism for 70 eyes of 70 patients. In the paper, they first converted the data to make a new unimodal dataset and then attempted to find the distribution of the data. We have taken the unimodal data in this paper and then, converted it into a dataset with 5 categories. If we denote the dataset by D and the number of observations in jth category by n_j, then n_j is the the number of observations between $(\frac{2\pi(j-1)}{5} - \frac{\pi}{5}, \frac{2\pi(j-1)}{5} + \frac{\pi}{5}]$. Thus, the number of observations in the categories are given in Table 10

We estimated the parameters by assuming both WSG and GWG distribution. The parameters for WSG were estimated first by estimating the mean and variance of h(q) by taking all pairs of $\cos(\theta_i - \theta_j)$ where, $i \neq j$. Then, using this estimated value of h(q), an estimate of q is found out. Then, using this estimate of q, the distribution of δ_{max} is estimated. The parameters of GWG were estimated using MLE. The parameters estimate for GWG are $q = 0.1425(0.0017)$ and $a* = 0.2435(0.0065)$. The point estimates for WSG are $h(q) = 0.4599(q = 0.1035), \delta_{max} = \delta_{mean} = 1$ (the first category). The estimate of standard deviation of h(q) is 0.0216.The bootstrap confidence interval for a at this estimated value of h(q) gives the estimated probability of finding the correct a at this value of q to be greater than 0.999.

Table 11: Wind direction dataset

Direction	N	NE	E	SE	S	SW	W	NW
n_j	30	16	22	24	52	28	9	19

Table 12: Estimated pmf of a

Direction	p_j
N	3.049×10^{-13}
NE	7.970×10^{-13}
E	4.113×10^{-13}
SE	1.040×10^{-4}
S	0.999
SW	4.664×10^{-6}
W	8.554×10^{-13}
NW	2.474×10^{-14}

6.2 Wind direction data

Wind direction for two consecutive years 2014 and 2015 were taken over Dumdum in Kolkata. The wind direction data was given in 8 categories. The data given was in percentage terms with the magnitude of percentage expressed as integers. There were some days when there was no wind and hence, those days were excluded from the data. We assumed the given wind direction data to be composed of 200 days, 100 days for 2014 and the same for 2015. Then, the sum was done over the two year for each wind direction to form the ordinal circular dataset. The dataset is shown in 11. Similar to the cataract surgery dataset, the estimates were found for this dataset. For GWG, the maximum likelihood estimates are $q = 0.6792(1.3561 \times 10^{-3}$ and $a* = 3.9999(2.8972 \times 10^{-4})$. For WSG, the point estimates are $h(q) = 0.0207(q = 0.6748)$, $\delta_{max} = S$ and $\delta_{mean} = SE$. The estimate of standard deviation of h(q) is 0.0073. As the value of q is very high, in place of maximum cell frequency and mean angle, we estimated the pmf of a by posterior probabilities as mentioned in 3.1.3 by taking the prior pmf to be uniform. The estimated posterior probabilities are given in Table 12.

Table 13: Estimates of parameters of time series

Parameters	MLE under $\alpha = 0$	MLE under the full model
p	0.6200 (0.0047)	0.4298 (0.0095)
q	0.4434 (0.0077)	0.2270 (0.0065)
α		1.3979 (0.8792)

6.3 Wind Direction Time-series Example

Wind direction for one week starting from 20th December, 2016 were taken at Baranagar in Kolkata. The wind direction was noted for every three hours. There were 8 uniformly spaced wind directions and no wind in the sample space of random variable associated with each time point. We carry out the time series analysis under the model mentioned in Section 4. Maximum likelihood estimation was carried out to estimate the underlying model parameters. First we tested the hypothesis that $H_0 : \alpha = 0$ vs $H_1 : \alpha > 0$ for homogeneity. At 5% level, the hypothesis was rejected with p-value 0.0005. Then, we estimated the model parameters. The MLEs under $\alpha = 0$ and $\alpha > 0$ along with the estimated variance of the MLE's are mentioned in Table 13.

7 Conclusion

The paper is an attempt to solve the issues regarding the ordinal circular random variables. We gave two models based on the wrapping of the geometric distribution for modelling. The distribution theory developed can be extended by appropriately wrapping other discrete distributions also. Further, we developed the time-series analysis based on one of the distributions. Moreover, we incorporated a different kind of sample space containing null-state which depended on a different random variable i.e. wind speed. By defining different functions for p and q, this can also be extended further. We also found in our data analysis that even when we fix $\alpha = 1$, then also the results were much better than the homogeneous case. If α is fixed and known, then there are two parameters to estimate and the MLE's of the parameters are asymptotically independent and the likelihood function is more tractable. But, in that case, the testing could not be performed. If we use α_1, α_2 as mentioned at the end of Section 4, then asymptotically (p, α_1) and (q, α_2) are independent. The model can also be used when the time points are not equidistant. The theory developed can also be used when predicting other datasets such as the time series analysis of the time of maximum rainfall in a day given that there are some days where there is no rain. Other such examples could also be found.

References

Ball F. and Blackwell, P. (1992). A Finite Form For The Wrapped Poisson Distribution. *Advances in Applied Probability* 24, 221-222.

Corrado, Charles J.(2011). The exact distribution of the maximum, minimum and the range of Multinomial/Dirichlet and Multivariate Hypergeometric frequencies. *Statistics and Computing* 21(3), 249-259.

Jacob, S. and Jayakumar, K.(2013). Wrapped Geometric Distribution: A new Probability Model for Circular Data. *Journal of Statistical Theory and Applications* 12(4), 348-355.

Jayakumar, K. and Jacob, S. (2012). Wrapped Skew Laplace Distribution on Integers: A New Probability Model for Circular Data *Open Journal of Statistics*

Levy, P. (1939). L'addition des variables aleatoires defines sur une circonference. *Bull. Soc. Math.France* 67, 1-41.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics.* John Wiley and Sons, Chichester.

Biswas, A., Jha, J. and Dutta, S. (2015). Modelling of Circular Random variables with a spike at zero. *Statistics and Probability Letters* 109, 194-201.

Kopylev,L. (2012).Constrained Parameters in Applications: Review of Issues and Approaches *ISRN Biomathematics* 2012.

Kato,S.(2010).A Markov process for circular data *Journal of the Royal Statistical Society, Series B* 2010. 72, 655-672