

1 Introduction

The theory of statistics on convex subsets of \mathbb{R}^n has been studied since at least the 7th century, when the Arab scholar Al-Khalil used permutations and combinations to list all possible Arabic words with and without vowels [4]. However, there are several areas relevant to daily life when the methods of linear statistics break down completely. For example, we know that the time halfway between 23:00 and 01:00 hours is 00:00 hours, however taking the average of 23 and 1 leads us to believe that the halfway time is 12:00 hours. We also know that the compass direction between 315 degrees (NW) and 45 degrees (NE) is not 180 degrees (S) but actually 0 deg (N), a difference of 180 degrees! Setting your compass this way would leave you hopelessly lost (ask me how I know). For this reason, over the last century the field of directional or circular statistics has been growing incredibly rapidly, starting with Richard von Mises studying the deviation of atomic weights from integer values and continuing into the present day [18]. At present, the field finds applications in modeling of temporal variations in environmental phenomena such as wind and wave direction, in the modeling of the migration patterns of birds, protein folding, and in any other area that can be assumed to be taking place on S^n , the n -cylinder, or the n -torus. Many great examples of these sorts of processes can be found in the references, in particular in [6], [17], and [18].

Many papers have focused on the modeling of distributions of wind direction (see [1], [6], and [16], among many others, for examples). This is of particular interest to many people in industry and government, who want to know how such variations might affect, for example, the output of a wind power plant. We will continue to follow this trend in this paper. We will investigate the use of these techniques on data from the National Data Buoy Center to model time series of wind direction measured over the ocean.

2 Circular Statistics

3 Distributions on the Circle

3.1 The circular uniform distribution

The easiest generalization from the real line to the circle is the uniform distribution. This is given by the probability density

$$f_X(x) = \frac{1}{2\pi} \tag{3.1}$$

which has the associated distribution function

$$F_X(x) = \frac{x}{2\pi} \tag{3.2}$$

This has all the same properties as the uniform distribution of the real line, so we will not spend time discussing it in detail.

3.2 The von Mises Distribution

There are several different distributions that can be used for modeling wind direction data, however the most common is the von Mises distribution [1]. Its probability density function is given by

$$f_X(x) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)} \quad (3.3)$$

and its cumulative density function is given by [17]

$$F_X(x) = \frac{1}{I_0(\kappa)} \int_0^x e^{\kappa \cos(x-\mu)} dx \quad (3.4)$$

where μ is the angular mean, $\kappa \in [0, \infty)$ is a parameter that determines the spread of the distribution, and $I_0(\kappa)$ is the zeroth order modified Bessel function of the first kind given by

$$I_r(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(r\theta) e^{\kappa \cos(\theta)} d\theta, r = 0, \pm 1, \pm 2, \dots \quad (3.5)$$

We will denote the von Mises distribution with parameters μ and κ by $M(\mu, \kappa)$.

3.2.1 Maximum Likelihood Estimation

We will determine the approximate distribution of the wind direction via maximum likelihood estimation. Consider a set of n data points x_1, x_1, \dots, x_n that are drawn from some distribution $f_X(x; \theta)$ where θ is an unknown parameter or vector of parameters. We can estimate θ by means of maximizing the likelihood function of the distribution. The likelihood function for θ is given by

$$L(\theta) = \prod_{k=1}^n f_X(x_k | \theta) \quad (3.6)$$

and the maximum likelihood estimate is the value $\hat{\theta}$ such that $L(\hat{\theta}) \geq L(\theta)$ for all values of θ and is the solution of the equation $\nabla L(\theta) = 0$, or equivalently $\nabla \log(L(\theta)) = 0$ (called the log-likelihood) which can often reduce the difficulty of the computation [15]. The log-likelihood function for the von Mises distribution given n data points is given by

$$\log(L(\kappa, \mu)) = -n \log(2\pi I_0(\kappa)) + \sum_{k=1}^n e^{\kappa \cos(x_k - \mu)} \quad (3.7)$$

Taking the appropriate gradient and setting it to zero, we can derive the maximum likelihood estimators $\hat{\mu}$ and $\hat{\kappa}$. Let the data points x_1, x_2, \dots, x_n be drawn from the interval $[0, 2\pi)$ according to some distribution and consider the transformation $z_k = e^{ix_k}$. Then the sample mean is given by $\bar{z} = \frac{1}{n} \sum_{k=1}^n z_k$ and the maximum likelihood estimate for μ is given by $\hat{\mu} = \text{Arg}(\bar{z})$. The estimator for κ is given by [3]

$$\bar{R} = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} \quad (3.8)$$

where $\bar{R} = \|\bar{z}\|$. This equation can be solved via your favorite rootfinding algorithm. For the purposes of this paper, the Newton-Raphson method was used.

In order to model a time series of von Mises random variables, we will also need to understand the distribution of the difference of two von Mises variables. First, we will determine the distribution of a negative von-Mises random variable.

Lemma 3.1. *If $X \in M(\mu, \kappa)$ then $W = -X \in M(-\mu, \kappa)$.*

The proof is as follows: Let $X \in M(\mu, \kappa)$ be a random variable and let $W = -X$. Then $F_W(w) = P(W \leq w) = P(-X \leq w) = 1 - P(X \leq -w) = 1 - F_X(-w)$, where $F_X(x)$ is as given in equation 3.4. By definition, the probability density of W is given by

$$f_W(w) = \frac{d}{dw} \left(1 - \frac{1}{2\pi I_0(\kappa)} \int_0^{-w} e^{\kappa \cos(x-\mu)} dx \right) \quad (3.9)$$

This derivative can be readily evaluated to show that $f_W(w) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(w+\mu)}$ which is the PDF for a $M(-\mu, \kappa)$ random variable.

With the aid of Lemma 3.1, we can determine the distribution of the difference between two von Mises random variables. Let $X \in M(\mu, \kappa)$ and $Y \in M(-\mu, \kappa)$. Since $-Y \stackrel{d}{=} X$ the distribution of the sum $X + Y$ is identical to the distribution of $X - X_1$ where $X_1 \stackrel{d}{=} X$. Letting $Z = X + Y$, we can apply the convolution theorem for sums of random variables and follow the derivation contained in [17] of the distribution of the sum of two von Mises random variables.

$$f_Z(z) = \int_0^{2\pi} f_X(x) f_Y(z-x) dx \quad (3.10)$$

From this, we can see that

$$f_Z(z) = \frac{1}{(2\pi I_0(\kappa))^2} \int_0^{2\pi} e^{\kappa(\cos(x-\mu) + \cos(z-x+\mu))} \quad (3.11)$$

We can apply trigonometric identities to show that $\kappa(\cos(x-\mu) + \cos(z-x+\mu)) = \kappa(\cos(\mu) + \cos(z+\mu)) \cos(x) + \kappa(\sin(\mu) + \sin(z+\mu)) \sin(x)$. Then we can show that the distribution $f_Z(z)$ is given by

$$f_Z(z) = \frac{1}{2\pi I_0(\kappa)^2} I_0 \left(\sqrt{2\kappa^2(1 + \cos(z))} \right) \quad (3.12)$$

This is not a von Mises distribution, however we can still make use of it in the coming sections.

4 Time Series Analysis of Circular Data

One straightforward way to determine the time correlation of the data is by using the formulation of the correlogram given in [7]. They define the sample correlogram ρ as

4.1 Periodograms of circular data

Since we will want to detect periodic behavior in the time series, we will use the formulation for the periodogram of circular data given in [6], which is an estimate of the spectrum of a zero-mean stochastic process with length n . The periodogram of the zero mean process $\{x_k\}$ is given by

$$\hat{g}(p) = \frac{1}{n} \left| \sum_{k=1}^n x_j \exp(i2\pi pk/n) \right|^2 \quad (4.1)$$

We can detect periodicity by the relative size of \hat{g} . When $\hat{g}(p)$ is large there is more energy contained in the component of the process with normalized frequency p/n , and when it is small there is less energy. If x_k is sampled with sampling frequency f_s , the real frequencies are given by $f = \frac{p}{n}f_s$.

Since any sequence of angles can be considered as a sequence of points along the unit circle, we can easily extend this formulation to circular data. Given a sequence of angles $\{x_k\} \in [0, 2\pi)$ the stochastic process $\{z_k = \exp(ix_k)\}$ lies on the unit circle, and we can readily calculate its mean μ using equation CITE EQUATION. Armed with this knowledge, it is readily seen that the periodogram of circular data is given by

$$\hat{g}(p) = \frac{1}{n} \left| \sum_{k=1}^n (z_k - \mu) \exp(i2\pi pk/n) \right|^2 \quad (4.2)$$

4.2 ARMA processes on the circle

$$\rho(k) = \quad (4.3)$$

5 Modeling

Over the years, a variety of models have been introduced for generating data with a proper directional distribution.

5.1 Markov Processes on the Circle

In [11], Johnson and Wehrle developed an algorithm for generating stationary Markov processes on the circle. Consider three distributions on the circle given by $f_1(\theta)$, $f_2(\phi)$, $g(\cdot)$ and let $F_k(\theta)$ be the distribution function of f_k . The distribution $g(\cdot)$ is called the binding density. We wish to find a joint distribution function $f(\theta, \phi)$ such that the marginal densities are f_1 and f_2 . Johnson and Wehrle showed that

$$f_{\Theta, \Phi}(\theta, \phi) = 2\pi g\{F_{\Theta}(\theta) - F_{\Phi}(\phi)\}f_{\Theta}(\theta)f_{\Phi}(\phi) \quad (5.1)$$

satisfies this condition. This can be seen by substituting $z = F_k(x)$ and integrating to calculate the marginals. They went on to define a Markov process on the unit circle by choosing $f = f_1 = f_2$ and setting the initial density $p(\theta_0) = f(\theta_0)$. We can clearly see from equation 5.1 that the conditional density of each θ_t is

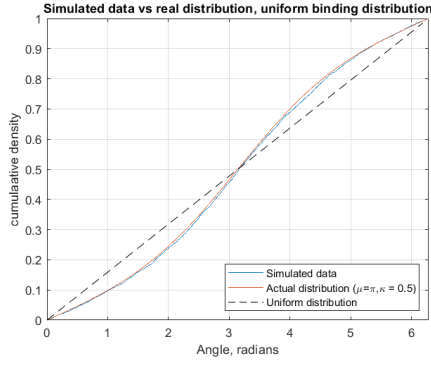
$$p(\theta_t|\theta_{t-1}, \theta_{t-2}, \dots, \theta_0) = p(\theta_t|\theta_{t-1}) = 2\pi g\{F(\theta_t) - F(\theta_{t-1})\}f(\theta_t) \quad (5.2)$$

since each θ_t has the same marginal distribution.

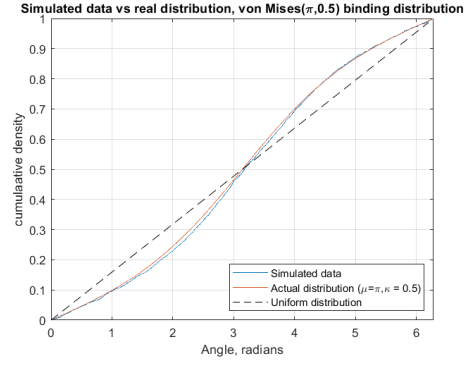
We can use this to apply the Metropolis-Hastings algorithm [19] to generate data that is distributed as f . Let $\{\Theta_t\}$ be a sequence of random variables on the set $I = [0, 2\pi)$ with target distribution $f(\theta)$. First, choose some value $y \in I$ from the transitional density $p(y|\theta_t)$ given in equation 5.2. Now, let

$$\alpha = \min\left\{1, \frac{f(y)p(y|\theta_t)}{f(\theta_t)p(\theta_t|y)}\right\} \quad (5.3)$$

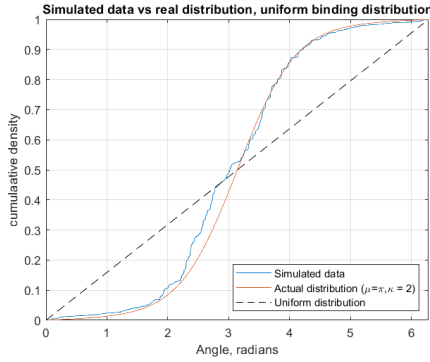
Then, set $\theta_{t+1} = y$ with probability α and $\theta_{t+1} = \theta_t$ with probability $1 - \alpha$, and step forward. The expression in α can be thought of as a sort of likelihood function, where values of y that



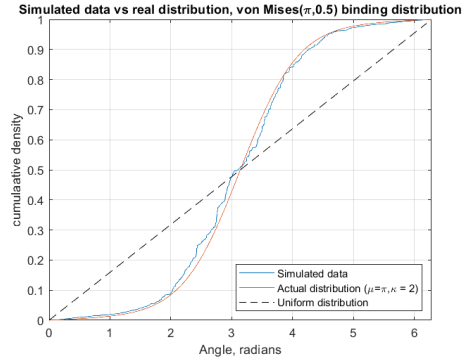
(a) Uniform binding, $\kappa = \frac{1}{2}$



(b) Von Mises binding, $\kappa = \frac{1}{2}$



(c) Uniform binding, $\kappa = 2$



(d) Von Mises binding, $\kappa = 2$

Figure 5.1: Simulated data vs target distribution for various distributions and binding densities, $n = 4000$

are more "likely" to have come from $f(\theta)$ than θ_t get accepted automatically and others only get accepted with some probability $\alpha < 1$ depending on their similarity to the expected distribution. Results for $\theta \in M(\mu = \pi, \kappa = \{\frac{1}{2}, 2\})$ with $n = 4000$ and both a uniform and $M(\pi, \frac{1}{2})$ binding density are shown in figure 5.1. Based purely on these results it appears that for larger values of κ the distribution of $\{\Theta_t\}$ converges to the target distribution more slowly.

This is a nice result, however in order to simulate realistic data our model must also have realistic autocorrelation. As we can see in figure 5.2, the simulated data has almost no autocorrelation, but data from NDBC buoys is highly autocorrelated. We will discuss ways to remedy this in the next section by introducing circular autoregressive models.

5.2 Circular Autoregressive Models

remove?

5.2.1 Wrapped and Linked Processes

There are several ways to generate autocorrelated circular data in the literature. One of the more common approaches is the wrapped autoregressive (WAR(p)) process, by which a Gaussian process $\{x_t\}$ on the real line is converted to a circular time series by setting

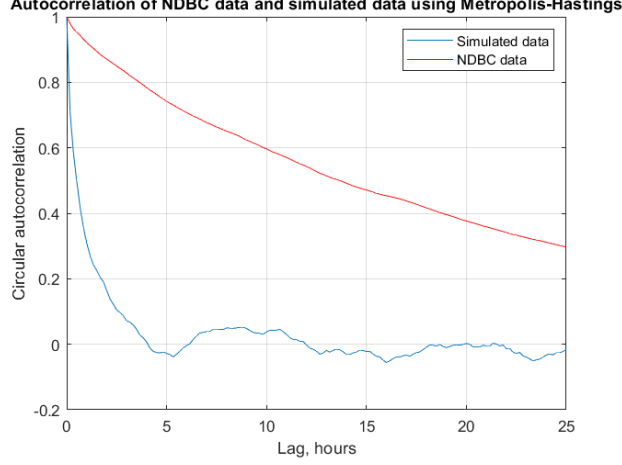


Figure 5.2: Autocorrelation of data generating using Metropolis-Hastings vs real data

$$y_t = x_t \pmod{2\pi} \quad (5.4)$$

These models can be seen as a missing data problem: The unwrapped observations can be written as $x_t = y_t + 2\pi k_t$ where k_t is an integer [9]. Estimation can be done via expectation maximization, which can be very difficult for any somewhat complex model. There are Markov Chain Monte Carlo methods to do this fitting proposed in [5].

A second commonly employed model is the linked or circular ARMA(p,q) process (LARMA(p,q) or CARMA(p,q)) [9]. This method associates a linear time series model x to a circular time series y with mean μ by means of a link function $g(t)$ via

$$x = g^{-1}(y - \mu) \quad (5.5)$$

where $g(t)$ is an increasing function on \mathbb{R} with range 2π such that $g(0) - g(-\infty) = \pi$. The most common link functions are probit transformations given by $g(x) = (2\Phi(x) - 1)\pi + \mu$, where $\Phi(x)$ is the normal CDF, or functions of the form $g(x) = 2 \arctan(x) + \mu$. If the aim is to have x be nearly Gaussian, the probit transformation is preferred. In these models the inverse of x is called an inverse autoregressive process (IAR(p)) in which the conditional mean of a circular distribution is given by [9]

$$\mu_{t|t-1} = \mu + g\{\phi_1 g^{-1}(y_{t-1} - \mu) + \dots + \phi_p g^{-1}(y_{t-p} - \mu)\} \quad (5.6)$$

When $p = 1$ and $\phi_1 \approx 1$ we can rewrite the IAR(1) model as

$$\mu_{t|t-1} \approx \mu + \phi_1(y_{t-1} - \mu) \quad (5.7)$$

This shows us that when $\phi_1 = 1$, $\mu_{t|t-1} = y_{t-1}$ so our process is a random walk [9]. However, there is one significant issue with this approach: it ignores that $y = 0$ is the same point as $y = 2\pi$, and it makes it impossible to jump directly from, for example, $y = \epsilon$ to $y = 2\pi - \delta$ in spite of the fact that those values are actually close to each other. The next section describes one attempt to resolve this problem.

5.3 Score-driven Models

The next class of models is based on the use of a score function to dynamically update an equation for the conditional mean at each time step. This section draws heavily from [9], which is a thorough discussion of score-driven models. The score function is defined as the gradient of the log-likelihood function, and so for the $M(\mu, \kappa)$ distribution it is given by taking the gradient of equation 3.7:

$$s(\mu, \kappa) = \nabla \log L_M(\mu, \kappa) = \begin{bmatrix} \frac{\partial \log L}{\partial \mu} \\ \frac{\partial \log L}{\partial \kappa} \end{bmatrix} = \begin{bmatrix} \kappa \sin(x - \mu) \\ \cos(x - \mu) - R \end{bmatrix} \quad (5.8)$$

where R is the mean resultant length.

The score function is used as follows. Let u_t be the score of an observation taken at time t and let $\phi < 1$ be a parameter. A score-driven update for the conditional mean $\mu_{t|t-1}$ is given by the recurrence relation

$$\mu_{t|t-1} = (1 - \phi)\mu + \phi\mu_{t-1|t-2} + au_{t-1} \quad (5.9)$$

where a is a parameter that determines the influence of the score on the updating and μ is the unconditional mean of the random variable. When $a = 0$ this expression has a fixed point at $\mu_t = \mu$. The effect of the score function is such that if the score of step $t - 1$ is large, the next step will be large and if it is small, the next step will be smaller.

This can be used to define the conditional distribution of the variable x_t on the real line given by $x_t = \mu_{t|t-1} + \epsilon_t$, where ϵ_t has zero mean. In the event that an observation falls outside of the support, it can be wrapped to the variable y_t so that it falls in the support. if we have $\epsilon_t \in M(0, \kappa)$, the score u_t is given by

$$u_t = \sin(y_t - \mu_{t|t-1}) \quad (5.10)$$

If, instead of using the conditional mean, we use the unconditional mean μ , we can derive the score-driven circular autoregressive model (SCAR) with order p as

$$\mu_{t|t-1} = \mu + \phi_1 \sin(y_{t-1} - \mu) + \dots + \phi_p \sin(y_{t-p} - \mu) \quad (5.11)$$

5.4 Stochastic Differential Equation Models

Our last (and most successful) potential model candidate comes from the field of stochastic differential equations. A basic stochastic differential equation is the Ornstein-Uhlenbeck process given by

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \quad (5.12)$$

where dW_t is a Wiener process, that is, a continuous time stochastic process with independent, normally distributed increments with mean 0 and variance equal to the time step so that $W_{t+\Delta t} - W_t \in N(0, \Delta t)$. This process is a modification of the continuous random walk, and can be thought of as a continuous analogue to the AR(1) process [20].

One of the first continuous time processes on the circle was proposed in [13]. This process, known as the von Mises process, can be thought of as a continuous time analogue to the Ornstein-Uhlenbeck process [8]. It was defined as the solution to the familiar looking stochastic differential equation

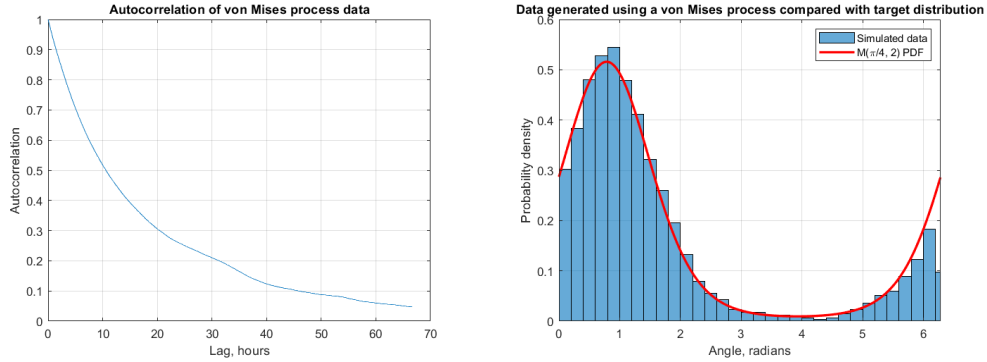
$$d\Theta_t = \alpha \sin(\mu - \Theta_t)dt + \sigma dW_t \quad (5.13)$$

where again W_t is a Wiener process, $\alpha > 0$ is a drift coefficient, μ is the mean direction, and σ is a diffusion coefficient. This process has a unique stationary distribution given by the $M(\mu, \frac{2\alpha}{\sigma^2})$ distribution. By assuming unit time steps, this can be easily reformulated into a stochastic difference equation as

$$\theta_{t+1} - \theta_t = \alpha \sin(\mu - \theta_t) + \sigma \epsilon_t \quad (5.14)$$

where $\epsilon \in N(0, 1)$. If we know our stationary distribution, we can develop a model using equation 5.14 to generate autocorrelated data with the desired distribution. This formulation also allows us to leverage regression techniques to determine the value of α .

An example of the data generated by this model is shown in figure 5.3. The data shown consists of 50000 samples generated with a target stationary distribution given by the $M(\pi/4, 2)$ distribution and with $\alpha = 0.01$. The data is strongly autocorrelated with a $1/e$ decorrelation time of approximately 15 hours assuming a sample rate of 6 samples per hour (which aligns with the NDBC data).



(a) Sample autocorrelation of data generated from a von Mises process (b) Simulated data vs stationary distribution

Figure 5.3: Autocorrelation of simulated data and comparison with stationary distribution

6 Results

7 Conclusions

References

- [1] Al Yammahi, Aishah & Prashanth R. Marpu & Taha B. M. J. Ouarda (2021), "Modeling directional distributions of wind data in the United Arab Emirates at different elevations", Arabian Journal of Geosciences
- [2] Artes, Rinaldo & Clelia M. C. Tolo (2009) An Autoregressive Model for Time Series of Circular Data, Communications in Statistics - Theory and Methods, 39:1, 186-194, DOI: 10.1080/03610920802650338
- [3] Borradaile, G. J. (2003). Statistics of earth science data : their distribution in time, space, and orientation. Springer. ISBN 978-3-662-05223-5

- [4] Broemeling, Lyle D. (1 November 2011). "An Account of Early Statistical Inference in Arab Cryptology". *The American Statistician*. 65 (4): 255–257. doi:10.1198/tas.2011.10191
- [5] Coles, S. G. (1998). Inference for circular distributions and processes. *Statistical Computing*, 5
- [6] Craig, Peter S (1988), "Time Series Analysis for Directional Data", Phd Thesis
- [7] Fisher, N.I and A.J. Lee (1994), "Time Series Analysis of Circular Data", *Journal of the Royal Statistical Society. Series B*
- [8] García-Portugués, Eduardo, Michael Sørensen, Kanti V. Mardia, and Thomas Hamelryck (2019), "Langevin diffusions on the torus: estimation and applications", *Statistics and Computing*
- [9] Harvey, Andrew; Hurn,Stan, and Stephen Thiele (2019) "Modeling Directional (Circular) Time Series", <https://doi.org/10.17863/CAM.43915>
- [10] Holzmann H, Munk A, Suster M, Zucchini W (2006) Hidden Markov models for circular and linear–circular time series. *Environ Ecol Stat* 13:325–347
- [11] Johnson RA, Wehrly TE (1979) Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* 66:255–2566
- [12] Kato,Shogo (2010); "A Markov Process for Circular Data" , *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*
- [13] Kent, J. (1975). Discussion of paper by K. V. Mardia. *J. Roy. Statist. Soc. Ser. B*, 37(3):377–378
- [14] Kuiper, N. H. (1960) "Tests concerning random points on a circle". *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*. 63: 38–47.
- [15] Larsen, Richard J and Morris L. Marx, "An Introduction to Mathematical Statistics and Its Applications, 6th Edition", Pearson Education, 2018
- [16] Mahrt, Larry (2011) "Surface Wind Direction Variability", *Journal of Applied Meteorology and Climatology*,
- [17] Mardia, Kantil; Jupp, Peter E. (1999). *Directional Statistics*. Wiley. ISBN 978-0-471-95333-3.
- [18] Mardia, K. V. (1975), "Statistics of Directional Data", Read before the Royal Statistical Society at a meeting organized by the Research Section on Wednesday, March 19th, 1975
- [19] Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculation by fast computing machines, *J. Chem. Phys.*, 21, 1087-1092
- [20] Øksendal, Bernt (YEAR), *Stochastic Differential Equations: An Introduction with Applications*, Springer-Verlag Heidelberg New York