CHAPTER II

LITERATURE REVIEW

In this chapter, fundamental concepts of mathematics and statistics relevant to Bayesian inference and the analysis of circular data are reviewed. The majority of the concepts on probability and random variables, including Probability Space, Conditional Probability, Independent Events, Classes of Probability Spaces, and Random Variables, are based on the works of Grimmett and Stirzaker (2020), Grimmett and Welsh (2017), Destrempes and Cloutier (2010), and Park, Song, and Yoon (2022).

2.1 Probability Space

The foundation of probability theory lies in the concept of a random experiment (or trial), which is an action whose outcome is uncertain. This concept is subsequently formalized as a mathematical construct called a *probability space*.

2.1.1 Sample Space

In probability theory, a sample space represents the set of all possible outcomes of an experiment. It provides a comprehensive view of the experiment's structure, forming the basis for defining events and assigning probabilities. The sample space is often denoted as Ω, and its nature depends on whether the experiment is discrete or continuous. Understanding the sample space is essential for developing accurate probabilistic models and drawing meaningful inferences.

Definition 2.1. An experiment that can be consistently repeated under controlled conditions, but whose outcome remains uncertain, is referred to as a random experiment.

Definition 2.2. The collection of all possible outcomes of an experiment is called the sample space of the experiment. This sample space, often represented by Ω.

Definition 2.3. An element of the sample space is referred to as a sample point or an elementary outcome.

Definition 2.4. For combined random experiments involving two discrete sample spaces, Ω_1 of size m and Ω_2 of size n, the sample space $\Omega = \Omega_1 \times \Omega_2$, which has a size of mn, is called a discrete combined space.

2.1.2 Event Space

The event space is the collection of all events that can occur in a given experiment, with each event being a subset of the sample space. It represents specific outcomes or sets of outcomes that satisfy certain conditions. The event space allows for the formal definition of events in probabilistic models, enabling probability calculations for each event. Its structure is crucial for understanding the relationships between events and their probabilities.

Definition 2.5. A collection $\mathcal{F}$ of subsets of a sample space Ω is called a *σ-field* if it satisfies the following conditions:

1. $\emptyset \in \mathcal{F}$;
2. If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;
3. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, where A^c is the complement of A.

A σ-field derived from a sample space is called an event space, and each element within this event space is referred to as an event.

Definition 2.6. An event that consists of a singleton set is referred to as an elementary event.

Definition 2.7. Let the sample space be the real line $\mathbb{R}$, the σ-algebra

$$\mathcal{B}(\mathbb{R}) = \sigma(\text{all open intervals})$$

generated by all open intervals (a, b) in $\mathbb{R}$ is known as the Borel σ-field, Borel algebra, or Borel field of $\mathbb{R}$.

2.1.3 Probability Measure

A probability measure provides the formal framework to quantify the likelihood of events within a sample space. This begins with the definition of a measurable space, followed by the introduction of a probability measure function, which assigns probabilities to events. These definitions collectively form a probability space, the essential structure for calculating and analyzing probabilities.

Definition 2.8. The pair $(\Omega, \mathcal{F})$ consisting of a sample space Ω and an event space $\mathcal{F}$ is referred to as a measurable space.

Definition 2.9. A mapping $P : \mathcal{F} \to \mathbb{R}$ is called a probability measure on the measurable space $(\Omega, \mathcal{F})$ if it satisfies the following axioms:

1. $P(A) \geq 0$ for all $A \in \mathcal{F}$.
2. $P(\Omega) = 1$ and $P(\emptyset) = 0$.
3. If a finite number of events $\{B_i\}_{i=1}^{n}$ are mutually exclusive (meaning $B_i \cap B_j = \emptyset$ whenever $i \neq j$), then
$$P\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i).$$
4. If a countable number of events $\{B_i\}_{i=1}^{\infty}$ are mutually exclusive, then
$$P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i).$$

The probability measure P assigns a real number $P(G)$, called the probability of an event G, and it is also denoted by $P\{G\}$ or Pr$\{G\}$. A probability measure is sometimes referred to as a probability function, probability distribution, or simply distribution.

Definition 2.10. The triplet $(\Omega, \mathcal{F}, P)$ consisting of a sample space Ω, an event space $\mathcal{F}$ of Ω, and a probability measure P is referred to as a *probability space*.

2.2 Conditional Probability

Conditional probability is a fundamental concept and a powerful analytical tool in probability theory. It quantifies the likelihood of an event occurring given the occurrence of another event. This concept is particularly essential when addressing problems with incomplete or partial information, as it often simplifies complex analyses and provides an efficient means of deriving probabilistic results.

Definition 2.11. The probability of an event given that another event has occurred is referred to as conditional probability. Specifically, the conditional probability of event A, given that event B has occurred, is denoted by $\mathbb{P}(A|B)$ and is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

where $\mathbb{P}(B) > 0$.

Theorem 2.1. *For any events A and B such that $0 < \mathbb{P}(B) < 1$,*

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c).$$

More generally, let $B_1, B_2, \ldots, B_n$ be a partition of Ω such that $\mathbb{P}(B_i) > 0$ for all i. The following theorem, called the total probability theorem, is then obtained:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

2.3 Independent Events

The occurrence of an event B can influence the likelihood of another event A, replacing the original probability $P(A)$ with the conditional probability $P(A \mid B)$. If $P(A \mid B) = P(A)$, the events A and B are defined as independent. More formally, independence requires that, for $P(A) > 0$ and $P(B) > 0$, the conditions $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$ are satisfied.

Definition 2.12. If the probability of the intersection of two events A and B, denoted $P(A \cap B)$, is equal to the product of their individual probabilities, i.e.,

$$P(A \cap B) = P(A)P(B),$$

then the events A and B are defined as independent, or mutually independent. Otherwise, the events are dependent.

2.4 Classes of Probability Spaces

The concepts of probability mass functions (PMFs) and probability density functions (PDFs) are introduced here. As discussed by Kim (2010), these functions provide an equivalent representation of the probability measure for describing a probability space. Furthermore, they serve as more practical and efficient tools for performing mathematical operations, particularly differentiation and integration, which are essential in the analysis of probability distributions.

2.4.1 Discrete Probability Spaces

In a discrete probability space, where the sample space Ω is a countable set, it is common to assume $\Omega = \{0, 1, \dots\}$ or $\Omega = \{1, 2, \dots\}$, with the event space $\mathcal{F}$ defined as the power set of Ω, denoted as $\mathcal{F} = 2^{\Omega}$.

Definition 2.13. In a discrete probability space, a function $p(\omega)$ that assigns a real number to each sample point $\omega \in \Omega$ is called a PMF if it satisfies the following conditions:

$$p(\omega) \geq 0, \quad \forall \omega \in \Omega$$

and

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

2.4.2 Continuous Probability Spaces

Let us now consider a continuous probability space with the measurable space $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Here, the sample space Ω is the set of real numbers $\mathbb{R}$, and the event space $\mathcal{F}$ is the Borel sigma-algebra $\mathcal{B}(\mathbb{R})$.

Definition 2.14. Let $(\Omega, \mathcal{B})$ be a measurable space. A real-valued Borel function f is called a probability density function, density function, or simply a density, if it satisfies the following two conditions:

$$f(r) \geq 0, \quad \forall r \in \Omega$$

and

$$\int_{\Omega} f(r)\, dr = 1.$$

2.5 Random Variables

Random phenomena are represented by functions on a probability space, known as *random variables*. The values a random variable assumes are determined by chance, and their likelihoods are captured by a function known as the *distribution function*.

Definition 2.15. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that for each $x \in \mathbb{R}$, the set

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

is an element of the sigma-algebra $\mathcal{F}$. This property is referred to as $\mathcal{F}$*-measurability*.

The distribution of a random variable is a fundamental statistical concept. It describes all the possible values that a random variable can take and the frequency with which each value occurs. Every random variable is associated with a *distribution function*, which is a key tool in statistical analysis.

Definition 2.16. The *distribution function* of a random variable X is defined as the function

$$F : \mathbb{R} \to [0, 1]$$

where $F(x) = \mathbb{P}(X \leq x)$.

The distribution function satisfies the following properties:

Theorem 2.2. *Let F be a distribution function. Then, the following conditions hold:*

1. $\lim_{x\to-\infty} F(x) = 0$,
2. $\lim_{x\to\infty} F(x) = 1$,
3. F *is non-decreasing: if* $x < y$, *then* $F(x) \leq F(y)$,
4. F *is right-continuous: as* $h \to 0^+$, $F(x+h) \to F(x)$.

Additionally, the following relationships hold between the distribution function of a random variable X and the probability measure $\mathbb{P}$.

Theorem 2.3. *Let* F *denote the distribution function of* X. *Then*

1. $\mathbb{P}(X > x) = 1 - F(x)$,
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$,
3. $\mathbb{P}(X = x) = F(x) - \lim_{y\to x^-} F(y)$.

Random variables can be divided into two primary types: *discrete* and *continuous*.

Definition 2.17. A random variable X is called *discrete* if it takes values from a countable subset of $\mathbb{R}$, denoted $\{x_1, x_2, \ldots\}$. For a discrete random variable, the *probability mass function* $f : \mathbb{R} \to [0, 1]$ is given by $f(x) = \mathbb{P}(X = x)$.

Definition 2.18. A random variable X is said to be *continuous* if its distribution function is expressible as

$$F(x) = \int_{-\infty}^{x} f(u)\,du \quad \text{for } x \in \mathbb{R},$$

where $f : \mathbb{R} \to [0, \infty)$ is an integrable function, known as the *probability density function* of X.

2.6 Expected Value and Variance

Let $x_1, x_2, \ldots, x_N$ denote the outcomes from N repetitions of an experiment. The arithmetic mean of these outcomes is given by

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Now consider a set of N discrete random variables, each having the same probability mass function f. For any given value x, approximately $Nf(x)$ of the variables X_i, where $i = 1, \ldots, N$, will take the value x. Hence, the average value can be written as

$$m = \frac{1}{N} \sum_x xNf(x) = \sum_x xf(x),$$

where the summation runs over all possible values of X_i. This average is referred to as the *expectation* or *mean value* of the distribution with mass function f.

Definition 2.19. The *expectation* or *mean* of a discrete random variable X with mass function f is defined by

$$\mathbb{E}(X) = \sum_{x:f(x)>0} xf(x), \tag{2.1}$$

provided that the sum converges absolutely.

The expectation, as given in equation (2.1), represents the average of the possible values of X, which can also be expressed as

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x),$$

indicating that each value is weighted by its corresponding probability.

Lemma 2.4. *Let X be a random variable with mass function f, and let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then, the expected value of $g(X)$ is given by*

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x),$$

assuming that the sum converges absolutely.

Definition 2.20. For any positive integer k, the kth *moment* m_k of X is defined as

$$m_k = \mathbb{E}(X^k) = \sum_x x^k\mathbb{P}(X = x),$$

provided the sum converges.

The kth *central moment* σ_k is defined as

$$\sigma_k = \mathbb{E}\left((X - m_1)^k\right),$$

where $m_1 = \mathbb{E}(X)$.

Definition 2.21. Two commonly used moments are:

- $m_1 = \mathbb{E}(X)$, which is the *mean* (or *expectation*) of X, and
- $\sigma_2 = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$, which is the *variance* of X.

These two quantities describe the central tendency and spread of X; that is, m_1 gives the average value of X, while σ_2 quantifies the extent to which X deviates from this average. The mean is often denoted as μ, and the variance is denoted as var(X). The standard deviation is the positive square root of the variance, $\sigma = \sqrt{\text{var}(X)}$.

In addition, higher moments provide further insight into the shape of the distribution. The *skewness* measures the asymmetry of the distribution, while the *kurtosis* describes the thickness of the tails and the peakedness of the distribution relative to a normal distribution.

Definition 2.22. The *skewness* of a random variable X, denoted γ_1, is defined as

$$\gamma_1 = \mathbb{E}\left[\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{var}(X)}}\right)^3\right],$$

provided the expectation exists. This measures the degree of asymmetry around the mean, with positive values indicating a longer right tail and negative values indicating a longer left tail.

Definition 2.23. The *kurtosis* of a random variable X, denoted γ_2, is defined as

$$\gamma_2 = \mathbb{E}\left[\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{var}(X)}}\right)^4\right],$$

provided the expectation exists. This quantifies the concentration of data in the tails and the peakedness relative to a Gaussian distribution, where a value greater than 3 indicates heavier tails or a sharper peak (leptokurtic), and a value less than 3 suggests lighter tails or a flatter peak (platykurtic).

Definition 2.24. The *moment generating function* (MGF) of a random variable X, denoted $M_X(t)$, is defined as

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx,$$

for a continuous random variable with density function f, provided the integral exists for some t in a neighborhood of 0. The MGF is a powerful tool that encapsulates all moments of X. Specifically, the kth moment of X can be obtained by taking the kth derivative of $M_X(t)$ with respect to t and evaluating at $t = 0$:

$$m_k = \mathbb{E}(X^k) = \frac{d^k}{dt^k} M_X(t) \bigg|_{t=0},$$

assuming the derivatives exist. Thus, the mean (m_1), variance (via $\mathbb{E}(X^2) - [\mathbb{E}(X)]^2$), skewness, and kurtosis can all be derived from the MGF, linking it directly to the previously defined quantities.

Definition 2.25. The *expectation* of a continuous random variable X with density function f is defined as

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx,$$

provided the integral exists.

Thus, the *mean*, *variance*, *skewness*, and *kurtosis* for a continuous random variable are defined as $\mu = \mathbb{E}(X)$, $\sigma_2 = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$, $\gamma_1 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$, and $\gamma_2 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$, respectively, provided all expectations exist.

2.7 Bayesian Statistics

Bayesian statistics is a branch of statistical theory grounded in the Bayesian interpretation of probability, where probability quantifies the degree of belief or certainty regarding a specific event or outcome. This framework leverages Bayes' theorem to update probabilities as new data are observed, seamlessly integrating prior knowledge with newly acquired information. Such an approach provides enhanced flexibility and adaptability in modeling and inference, particularly in scenarios where data are limited or noisy. By employing Bayes' theorem, parameters of a statistical model or probability distribution can be directly estimated through the assignment of a probability distribution to these parameters, reflecting beliefs about their plausible values. Furthermore, Bayesian statistics establishes a structured methodology for incorporating prior knowledge into the analysis, utilizing concepts such as prior distributions, Bayesian estimators, loss functions, and risk functions to refine inferences and decision-making processes (Lee, 2012).

Theorem 2.5. *Bayes' Theorem serves as a cornerstone for computing the conditional probability of an event based on prior knowledge of related events. It is expressed as*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)},$$

where $\mathbb{P}(A|B)$ denotes the posterior probability, representing the updated likelihood of event A given the occurrence of event B, $\mathbb{P}(B|A)$ is the likelihood, indicating the probability of observing B given that A has occurred, $\mathbb{P}(A)$ is the prior probability of event A, and $\mathbb{P}(B)$ is the marginal probability of event B, acting as a normalizing constant.

2.7.1 Specification of the Prior

In Bayesian analysis, the prior distribution, denoted $\mathbb{P}(A)$ in Bayes' theorem, encapsulates initial beliefs about the parameter of interest before observing the data. Priors can be classified as *informative*, based on existing knowledge or expert opinion, or *uninformative*, designed to exert minimal influence on the posterior, such as a uniform or improper prior (e.g., $h(\theta) = 1$ for $\theta > 0$). The choice of prior is critical, as it shapes the posterior distribution and, consequently, the resulting inferences. For a parameter θ, the prior distribution $p(\theta)$ is specified to reflect the analyst's belief about θ, and when combined with the likelihood $p(x|\theta)$ via Bayes' theorem, it yields the posterior $p(\theta|x) \propto p(x|\theta)p(\theta)$.

2.7.2 Bayesian Estimator

A Bayesian estimator is derived from the posterior distribution to provide a point estimate of the parameter θ. Common estimators include the posterior mean, $\mathbb{E}(\theta|x) = \int \theta p(\theta|x)\, d\theta$, which minimizes the expected squared error, and the posterior median, which is robust to outliers. The choice of estimator often depends on the decision-theoretic framework, where an estimator is selected to optimize a specified criterion, such as minimizing a loss function. The process of obtaining a Bayesian estimator involves the following steps:

1. **Specify the Prior**: Define the prior distribution $p(\theta)$ based on prior knowledge or an uninformative assumption.

2. **Determine the Likelihood**: Model the likelihood $p(x|\theta)$ based on the observed data x and the assumed distribution of the random variable.

3. **Compute the Posterior**: Apply Bayes' theorem to obtain the posterior distribution, $p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta')p(\theta')\,d\theta'}$, where the denominator is the marginal likelihood, ensuring normalization.

4. **Select a Loss Function**: Choose a loss function $L(\theta, \hat{\theta})$ (e.g., squared error loss, $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$) to quantify the penalty of estimation errors.

5. **Derive the Estimator**: Minimize the expected posterior loss, $\mathbb{E}[L(\theta, \hat{\theta})|x] = \int L(\theta, \hat{\theta})p(\theta|x)\,d\theta$, with respect to $\hat{\theta}$. For squared error loss, this yields the posterior mean; for absolute error loss, it yields the posterior median.

This systematic approach underscores the Bayesian paradigm's ability to synthesize prior beliefs and observed data into a coherent estimate, guided by the chosen loss function.

2.7.3 Loss Function and Risk Function

The evaluation of a Bayesian estimator involves the use of a *loss function*, denoted $L(\theta, \hat{\theta})$, which quantifies the cost of estimating the true parameter θ with an estimator $\hat{\theta}$. Common loss functions include squared error loss, $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$, and absolute error loss, $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$. The *risk function*, or expected loss, is defined as

$$R(\theta, \hat{\theta}) = \mathbb{E}[L(\theta, \hat{\theta})] = \int L(\theta, \hat{\theta})p(x|\theta)\,dx,$$

representing the average loss over all possible data x given θ. In Bayesian decision theory, the optimal estimator minimizes the expected posterior risk, $\mathbb{E}[L(\theta, \hat{\theta})|x] = \int L(\theta, \hat{\theta})p(\theta|x)\,d\theta$, integrating the loss over the posterior distribution. This framework links the prior, likelihood, and posterior to decision-making under uncertainty.

Using Bayes' theorem, initial assumptions about the parameter θ are revised based on the evidence provided by the data x. The posterior distribution $p(\theta|x)$ offers a refined estimate of θ, informed by both the prior $p(\theta)$ and the likelihood $p(x|\theta)$, while Bayesian estimators, guided by loss and risk functions, facilitate optimal inference and prediction.

2.8 Wrapped Distributions

Wrapped distributions are frequently used in directional statistics to describe the distribution of circular data. This is because circular data repeats at regular intervals, such as every 360 degrees (or 2π radians), and this property is not captured by traditional probability distributions (Fisher, 1993).

Definition 2.26. If X is a random variable with values on the line and distribution function $F(x)$, the random variable X_W of the wrapped distribution is given by

$$X_W = x \pmod{2\pi}$$

and the distribution function $F_W(\theta)$ of X_W is given by

$$F_W(\theta) = \sum_{k=-\infty}^{\infty} \left[F(\theta + 2\pi k) - F(2\pi k)\right], \qquad 0 < \theta \leq 2\pi.$$

Specifically, if X is a discrete random variable with integer range, the for a fixed $m \in \mathbb{N}$, X_W is defined by

$$X_W = 2\pi x \pmod{2\pi m},$$

which is a random variable with r aligns in the lattice $\left[\frac{2\pi r}{m}, r = 0, 1, \ldots, m-1\right]$ on the circle. The probability mass function $\mathbb{P}_W\left(X_W = \frac{2\pi r}{m}\right)$ of X_W is given by

$$\mathbb{P}_W\left(X_W = \frac{2\pi r}{m}\right) = \sum_{k=-\infty}^{\infty} f(r + km), \qquad r = 0, 1, \ldots, m-1.$$

If X is a continuous random variable, the probability density function $f_W(\theta)$ of X_W is

$$f_W(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k).$$

2.9 Distributions Related Wrapped Data

Distributions related to wrapped data are used to model circular or periodic data, and are important in a variety of fields including meteorology, biology, and engineering.

These distributions treat data as if it were on a circle, with a periodicity of 2π, allowing for more accurate modeling of data that exhibits circular or periodic patterns.

2.9.1 Wrapped Binomial Distribution

A wrapped binomial distribution is the wrapped version of a binomial distribution. It is often used to model circular data that have a binary outcome. The PMF of a wrapped binomial distribution is given by (Girija, Dattatreya Rao, and Srihari, 2014a, 2014b)

$$f_W\left(\theta = \frac{2\pi r}{m}|n, p\right) = \sum_{k=0}^{\lfloor \frac{n-r}{m} \rfloor} \binom{n}{r+km} p^{r+km}(1-p)^{n-(r+km)}, \qquad r = 0, 1, \ldots, m-1,$$

where θ is the observed angle, n is the number of trials, p is the probability of success, and $n \geq m$.

2.9.2 Wrapped Geometric Distribution

The wrapped geometric distribution still has the same concept with the classical geometric distribution. It is used to model circular data that have a binary outcome and exhibit clustering or periodicity. The PMF of a wrapped geometric distribution is given by (Jacob, and Jayakumar, 2013)

$$f_W\left(\theta = \frac{2\pi r}{m}|p\right) = \sum_{k=0}^{\infty}(1-p)^{r+km}p, \qquad r = 0, 1, \ldots, m-1,$$

where θ is the observed angle, and p is the probability of success.

2.9.3 Wrapped Poisson Distribution

The wrapped Poisson distribution is a circular distribution obtained by wrapping a Poisson distribution. Circular data that have a countable outcome can be modeled by this distribution. Its PMF is given by (Mardia, and Jupp, 2001)

$$f_W\left(\theta = \frac{2\pi r}{m}|\lambda\right) = e^{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r+km}}{(r+km)!} \qquad r = 0, 1, \ldots, m-1,$$

where θ is the observed angle, and λ is the rate parameter.

2.9.4 Wrapped Normal Distribution

The other well-known wrapped distribution is the wrapped normal distribution. Circular data, such as angles or directions, is often modeled by this distribution. Its PDF is given by (Mardia, and Jupp, 2001)

$$f_W(\theta|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(2k\pi+\theta-\mu)^2}{2\sigma^2}\right\},$$

where θ is the observed angle, μ is the mean angle, and σ is the standard deviation.

2.9.5 Wrapped Cauchy Distribution

For the circular data that exhibit heavy-tailed behavior, the wrapped Cauchy distribution is often used to model. It has a PDF given by (Jammalamadaka, and Gupta, 2014)

$$f_W(\theta|\mu,\gamma) = \frac{1}{2\pi}\frac{1-\gamma^2}{1+\gamma^2-2\gamma\cos(\theta-\mu)},$$

where θ is the observed angle, μ is the mean angle, and γ is the concentration parameter.

2.9.6 Wrapped Exponential Distribution

The wrapped exponential distribution is a probability distribution commonly used to model cyclic or periodic phenomena. It finds applications in various fields, including the analysis of animal migration patterns, seasonal cycle modeling of waves in oceanography, and diagnosing daily cycles of pill intake in the human body.

This distribution is derived by mapping the exponential distribution onto a circular or periodic domain. To recall, the distribution of a random variable X is exponential if its PDF is given by:

$$f(x \mid \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where X is a continuous random variable and $\lambda > 0$ is the rate parameter, representing the inverse of the average time between events.

Hence, the distribution of a random variable X follows a wrapped exponential distribution if its PDF is given by:

$$\begin{aligned} f_\Theta(\theta \mid \lambda) &= \sum_{k=-\infty}^{\infty} f_X(\theta + 2\pi k \mid \lambda), \theta \in [0, 2\pi) \\ &= \sum_{k=0}^{\infty} \lambda e^{-\lambda(\theta+2\pi k)} \\ &= \lambda e^{-\lambda\theta} \sum_{k=0}^{\infty} e^{(-2\pi\lambda)k} \\ &= \frac{\lambda e^{-\lambda\theta}}{1 - e^{-2\pi\lambda}}, \end{aligned}$$

where Θ is a wrapped random variable and λ is a parameter. The PDF of the wrapped exponential distribution is a decreasing function. Consequently, as the angle increases, the likelihood of observing a larger angle diminishes.

2.9.7 Hurwitz Zeta Function

The Hurwitz Zeta function can be seen as an extension of the Riemann Zeta function. It is named after the Austrian mathematician Adolf Hurwitz, who introduced it in the late 19th century. Its definition is as follows

$$\zeta(s, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^s},$$

where

- a is a real number, $a \neq 0, -1, -2, \ldots$, and
- s is a complex number with real part greater than 1 (Re$(s) > 1$).

However, the Hurwitz Zeta function can be also represented in the integral form as

$$\zeta(s, a) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{x^{s-1} e^{-ax}}{1 - e^{-x}} dx.$$

2.10 Related Research

Ravindran and Ghosh (2011) discusses the use of Bayesian methodology for analyzing circular data. They propose using wrapped distributions, which are probability distributions that are defined on a circle rather than a line, to model circular data. The context shows how the wrapped Cauchy distribution and the wrapped normal distribution can be used to model circular data. The performance of these models to traditional methods such as circular-linear regression and circular kernel density estimation was compared. Through a simulation study and the analysis of real data on bird migration patterns, the authors provide evidence for the effectiveness of their Bayesian methodology using wrapped distributions for the analysis of circular data. The article also establishes that this approach offers a flexible and efficient means of analyzing circular data, and is capable of producing more accurate results when compared to traditional methods.

Jacob and Jayakumar (2013) propose a new distribution called the wrapped geometric (WG) distribution for modeling circular data that takes values in the set of positive integers. The authors derive the probability mass function and cumulative distribution function of the WG and compare it to other circular distributions, such as the von Mises and wrapped Cauchy distributions. They demonstrate the usefulness of the WG through simulation studies and by applying it to real data sets related to animal movement and wind direction. The authors conclude that the WG is a useful tool for modeling circular count data that takes values in the set of positive integers, particularly in cases where other circular count distributions may not be appropriate.

Umehara, Okada, and Naruse (2018) propose a method for estimating time series of angles that are wrapped or unwrapped using sequential Bayesian filters. The authors consider both circular data, such as wind direction, and linear data, such as stock prices, and develop a framework that can handle both types of data. The authors also propose a method for estimating hyperparameters, such as the observation noise variance and the transition noise variance, using the maximum a posteriori estimator. They demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to wind direction and stock prices. The authors conclude that their method is a powerful tool for estimating time series of angles that are wrapped or unwrapped, and it can be used in a wide range of applications, including weather forecasting and financial analysis.

Srihari, Girija, and Rao (2018a) introduce a new probability distribution called the discrete wrapped exponential (DWE) distribution for modeling circular data that is discrete in nature. The authors derive the probability mass function and cumulative distribution function of the DWE and compare it to other circular distributions, such as the wrapped Cauchy and von Mises distributions. They demonstrate the usefulness of the DWE through simulation studies and by applying it to real data sets related to wind direction and animal movement. The authors conclude that the DWE is a useful tool for modeling discrete circular data, particularly in cases where other circular distributions may not be appropriate. They also discuss some characteristics of the DWE, such as its moments and order statistics.

Srihari, Girija, and Rao (2018b) propose a new distribution called the wrapped negative binomial (WNB) distribution for modeling circular count data. The authors derive the probability mass function and cumulative distribution function of the WNB and compare it to other circular count distributions, such as the wrapped Poisson and the wrapped binomial. They demonstrate the usefulness of the WNB through simulation studies and by applying it to real data sets related to bird migration and crime patterns. The authors conclude that the WNB is a flexible and useful tool for modeling circular count data, particularly in cases where other circular count distributions may not be appropriate.

Malay, Xiaolong, Ashis, and Ruoyang (2019) propose a method for selecting non-subjective prior distributions for Bayesian analysis of wrapped Cauchy distributions. The authors demonstrate that traditional prior distributions, such as the Jeffreys prior, can be inappropriate for wrapped Cauchy distributions, and they propose a new class of priors based on the confluent hypergeometric function. The authors demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to animal movement and wind direction. The authors conclude that their method provides a useful tool for Bayesian analysis of wrapped Cauchy distributions, and it can be used in a wide range of applications, including ecology and meteorology.

Ignacio, Concha, and Pedro (2019) propose a method for classification of circular data using wrapped Cauchy distributions within a Bayesian framework. The authors demonstrate that traditional classifiers, such as linear discriminant analysis, are not appropriate for circular data, and they introduce the use of wrapped distributions to model circular data. They develop a Bayesian framework that uses the wrapped Cauchy distribution to model circular data and demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to wind direction and animal movement. The authors conclude that their method is a powerful tool for classification of circular data, and it can be used in a wide range of applications, including meteorology and animal behavior analysis.

Girija, Srihari, and Srinivas (2019) introduces a new probability distribution called the discrete wrapped Cauchy distribution (DWCD) for modeling circular data that is discrete in nature. The authors derive the probability density function and cumulative distribution function of the DWCD and compare it to other circular distributions, such as the wrapped Cauchy and von Mises distributions. They demonstrate the utility of the DWCD through simulation studies and by applying it to real data sets related to wind direction and animal movement. The authors conclude that the DWCD is a useful tool for modeling discrete circular data, particularly in cases where other circular distributions may not be appropriate.

Coelho-Barros, Achcar, Martinez, Davarzani, and Grabsch (2019) propose a Bayesian approach for analyzing data using the segmented Weibull distribution. The authors derive the probability density function and cumulative distribution function of the segmented Weibull distribution and develop a Bayesian framework for estimating the parameters of the distribution. They propose a non-informative prior distribution for the parameters and demonstrate how to use Markov chain Monte Carlo methods to estimate the posterior distribution of the parameters. The authors demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to cancer survival times and wind speed. The authors conclude that their approach provides a useful tool for modeling data with changing hazard rates, and it can be used in a wide range of applications, including engineering and health sciences.

Yu and González (2020) propose a Bayesian approach for estimating fault and volcano surface ground deformation models from wrapped satellite interferometric phase data. The authors use the wrapped Cauchy distribution to model the phase data and develop a Bayesian inversion algorithm to estimate the surface ground deformation models. The authors demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to the 2010 Chile earthquake and the 2011 eruption of the Puyehue-Cordón Caulle volcano in Chile. The authors conclude that their method is a powerful tool for estimating surface ground deformation models from wrapped satellite interferometric phase data, and it can be used in a wide range of applications, including earthquake and volcanic activity monitoring.

Bailey and Codling (2020) discusses the emergence of the wrapped Cauchy distribution in mixed directional data, which is data that contains a mixture of directional and non-directional components. The authors demonstrate that when directional data is mixed with non-directional data, the resulting distribution often exhibits characteristics of the wrapped Cauchy distribution. They develop a statistical framework for modeling mixed directional data using the wrapped Cauchy distribution and demonstrate its effectiveness through simulation studies and by applying it to real data sets related to animal movement and ocean currents. The authors conclude that their approach provides a useful tool for analyzing mixed directional data, and it can be used in a wide range of applications, including ecology and oceanography.

Okhli and Nooghabi (2020) propose a Bayesian approach for modeling positive-valued insurance claim data with outliers using the contaminated exponential distribution. The authors demonstrate that traditional distributions, such as the exponential and gamma distributions, can be inadequate for modeling insurance claim data with outliers, and they propose the use of the contaminated exponential distribution to model such data. They develop a Bayesian framework for estimating the parameters of the contaminated exponential distribution and demonstrate its effectiveness through simulation studies and by applying it to real insurance claim data. The authors conclude that this approach provides a useful tool for modeling insurance claim data with outliers and can be used in a wide range of applications in the insurance industry.

Chaudhary (2020) propose a Bayesian approach for analyzing data using the two-parameter exponentiated log-logistic (TELL) distribution. The author derives the probability density function, cumulative distribution function, and hazard function of the TELL distribution and discusses its various characteristics, such as its moments, skewness, and kurtosis. The author then develops a Bayesian framework for estimating the parameters of the TELL distribution, including the shape parameter, scale parameter, and exponentiation parameter, and demonstrates the effectiveness of their approach through simulation studies and by applying it to real data sets related to income and survival times. The author concludes that their approach provides a useful tool for modeling data with heavy tails and can be applied in a wide range of applications, including economics and health sciences.

Ghaderinezhad, Ley, and Loperfido (2020) proposes a Bayesian approach for analyzing data using skew-symmetric distributions, which are distributions that exhibit skewness in both positive and negative directions. The authors develop a general framework for modeling skew-symmetric distributions and demonstrate its usefulness by applying it to several specific distributions, including the skew normal, skew Student's t, and skew Laplace distributions. They discuss how to select appropriate prior distributions for the parameters of skew-symmetric distributions and develop a Bayesian inference algorithm for estimating the parameters. The authors demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to finance and economics. The authors conclude that their approach provides a powerful tool for analyzing data with skewness in both positive and negative directions, and it can be used in a wide range of applications.

Aslam and Feroze (2020) propose a Bayesian approach for analyzing data using a two-component mixture of Weibull distributions. The authors develop a likelihood function for the mixture distribution and propose an approximation technique based on the Laplace approximation and the saddlepoint approximation to estimate the posterior distribution of the parameters. They also develop a Bayesian model selection criterion based on the integrated completed likelihood to determine the number of components in the mixture distribution. The authors demonstrate the effectiveness of their approach through simulation studies and by applying it to real data sets related to survival times and wind speed. The authors conclude that their approach provides a powerful tool for analyzing data using mixture models, and it can be used in a wide range of applications, including engineering and environmental sciences.