Ch3 Some Special Distributions

Jih-Chang Yu

August 17, 2023

Binomial Distribution

Binomial Distribution

1. Introduction
    - Binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of independent Bernoulli trials.
    - It is widely used in statistics and probability theory to analyze situations with two possible outcomes, such as success/failure or yes/no.
2. Characteristics of Binomial Distribution
    - Fixed number of trials (n): The binomial distribution considers a fixed number of independent trials.
    - Two possible outcomes: Each trial has two possible outcomes, often denoted as success (S) and failure (F).
    - Constant probability of success (p): The probability of success remains constant for each trial.
    - Independence of trials: The outcome of one trial does not affect the outcome of another trial.

Binomial Distribution

1. Probability Mass Function (PMF)
   - The probability mass function of the binomial distribution gives the probability of obtaining a specific number of successes (k) in a fixed number of trials (n).
   - PMF formula: $P(X = k) = C(n, k) \cdot p^k \cdot (1 - p)^{n-k}$, where $C(n, k)$ represents the binomial coefficient.
2. Cumulative Distribution Function (CDF)
   - The cumulative distribution function of the binomial distribution gives the probability of obtaining at most a certain number of successes.
   - CDF formula: $P(X \leq k) = \sum_{i=0}^{k} P(X = i)$
3. Mean and Variance
   - Mean (μ) and Variance (σ^2) formulas for the binomial distribution:

$$\mu = n \cdot p$$

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

Application of Binomial distribution 1: Market Research and Surveys

In market research, we often seek to understand the success rate of specific products or services. For instance, in a market survey questionnaire, respondents are asked whether they have purchased a particular product. The binomial distribution can be utilized to estimate the success rate of purchases within the entire surveyed population.

Application of Binomial distribution 2: Quality Control

In manufacturing, it's essential to monitor the quality of products. For example, in a production batch, we might be interested in whether the products meet certain quality standards. The binomial distribution can help us predict the number of qualified products in large-scale production.

Application of Binomial distribution 3: Biomedical Research

In medical experiments, we often need to study the success rate of certain treatment methods. For example, in drug trials, we want to know if patients exhibit positive responses to a treatment. The binomial distribution can be used to analyze the effectiveness of such treatment methods.

Geometric Distribution

1. Introduction
    - Geometric distribution models the number of trials required to achieve the first success in a sequence of independent Bernoulli trials.
    - It is often used to analyze the waiting time for a specific event to occur, such as the number of flips needed to get the first "heads" in a series of coin tosses.
2. Characteristics of Geometric Distribution
    - Independent trials: Each trial is independent of the others.
    - Two possible outcomes: Each trial has two possible outcomes, often denoted as success (S) and failure (F).
    - Constant probability of success (p): The probability of success remains constant for each trial.
    - Memorylessness property: The probability of achieving the first success is independent of the number of failures that have occurred before.

Geometric Distribution

1. Probability Mass Function (PMF)
    - The probability mass function of the geometric distribution gives the probability of requiring exactly k trials to achieve the first success.
    - PMF formula: $P(X = k) = (1 - p)^{k-1} \cdot p$, where p is the probability of success.
2. Cumulative Distribution Function (CDF)
    - CDF formula: $P(X \leq k) = 1 - (1 - p)^k$
3. Mean and Variance
    - Mean (μ) and Variance (σ^2) formulas for the geometric distribution:

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{1 - p}{p^2}$$

Application of Geometric Distribution 1: Marketing Campaign

Imagine a marketing campaign where an ad is shown to potential customers until the ad gets clicked for the first time. The probability of clicking the ad (p) is relatively low. The geometric distribution helps us model how many times the ad needs to be shown before it gets its first click.

Application of Geometric Distribution 2: Equipment Failure

Consider a scenario where equipment in a factory can fail due to a specific type of malfunction. The probability of such a failure (p) is low. The geometric distribution can be used to estimate how many times the equipment needs to operate before it experiences its first failure.

Application of Geometric Distribution 3: Online Learning

In online learning platforms, a student attempts a quiz multiple times until they answer correctly for the first time. The probability of answering correctly (p) might vary. The geometric distribution helps analyze how many quiz attempts are needed on average before achieving the first correct answer.

Poisson Distribution

1. Introduction
    - Poisson distribution is a discrete probability distribution that models the number of events occurring in a fixed interval of time or space.
    - It is commonly used to describe the occurrence of rare events, such as the number of phone calls received in a call center per minute or the number of accidents in a day.
2. Characteristics of Poisson Distribution
    - Average rate (λ): The Poisson distribution is determined by the average rate at which events occur in the given interval.
    - Independence of events: The occurrence of events is assumed to be independent of each other.
    - Constant rate: The rate at which events occur remains constant throughout the interval.
    - Events are rare: The probability of multiple events occurring simultaneously is negligible.

Poisson Distribution (cont.)

4. Probability Mass Function (PMF)
   - The probability mass function of the Poisson distribution gives the probability of observing k events in the interval.
   - PMF formula: $P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$, where λ represents the average rate.
5. Mean and Variance
   - Mean (μ) and Variance (σ^2) formulas for the Poisson distribution:

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

Claim Poisson pmf

Poisson mean

Application of Poisson Distribution 1: Customer Arrivals

Consider a scenario where customers arrive at a service center over time. The Poisson distribution can be used to model the number of customer arrivals within a specific time interval. This is useful for staff scheduling and resource allocation.

Application of Poisson Distribution 2: Defects in Manufacturing

In manufacturing, defects in products can occur randomly. The Poisson distribution can help predict the number of defects in a batch of items, aiding in quality control and process improvement efforts.

Application of Poisson Distribution 3: Call Center Traffic

For call centers, the number of incoming calls in a given time period can be modeled using the Poisson distribution. This information is crucial for staffing levels and optimizing call center operations.

Association of Binomial and Poisson Distributions

The Poisson distribution can be seen as a limiting case of the Binomial distribution when the number of trials (n) is large and the success probability (p) is small.

Binomial Distribution:

- Models the number of successes in a fixed number of trials.
- Discrete probabilities for a finite number of outcomes.

Poisson Distribution:

- Models the number of rare events in a fixed interval.
- Discrete probabilities for an infinite number of outcomes.

When to Use Which

- Use the Binomial distribution when there's a fixed number of trials and each trial is independent.
- Use the Poisson distribution when the number of trials is large, and the success probability is very small, resulting in rare events.

Connection:

- As n becomes large and p becomes small in a Binomial distribution, it approaches the shape of a Poisson distribution.
- The mean of a Poisson distribution (λ) is often related to n and p in a Binomial distribution.

Uniform Distribution

- The uniform distribution is a continuous probability distribution that describes events where all outcomes are equally likely.
- It's often used when no particular outcome is more likely than any other within a certain range.
- The uniform distribution has a constant probability density function (PDF) over a specified interval.
- Probability Density Function (PDF):

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

where a and b are the interval boundaries.

Properties of Uniform Distribution

- Mean: $\mu = \frac{a+b}{2}$ and Variance: $\sigma^2 = \frac{(b-a)^2}{12}$.
- The uniform distribution is symmetric and has a rectangular-shaped PDF.
- The distribution is defined by its interval boundaries a and b.
- The cumulative distribution function (CDF) is a linear function:

$$F(x|a,b) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases}$$

Applications and Examples

- The uniform distribution has various applications in real-world scenarios:
  - Random number generation: Simulating fair dice rolls or lottery drawings.
  - Resource allocation: Distributing resources evenly among possibilities.
  - Statistical sampling: Selecting samples uniformly from a population.
- Example: Suppose we have a random variable representing the time (in minutes) a customer spends in a store. If the time is uniformly distributed between 5 and 15 minutes, find the probability that a customer spends less than 10 minutes in the store.

$$P(X < 10) = F(10|5, 15) = \frac{10 - 5}{15 - 5} = 0.5$$

Conclusion

- The uniform distribution models situations where all outcomes within a specified interval are equally likely.
- It's characterized by its interval boundaries (a and b) and has a constant PDF.
- The mean and variance formulas make it easy to calculate expected values and variability.
- The uniform distribution finds applications in various fields, especially in randomization and sampling.

Applications and Examples

- The uniform distribution has various applications in real-world timing scenarios:
  - **Example 1:** **Website Load Time**
    The time users spend waiting for a website to load can be modeled with a uniform distribution. If the load time is uniformly distributed between 2 to 5 seconds, users are equally likely to experience various load times within that range.
  - **Example 2:** **Bus Arrival Times**
    Modeling the time between bus arrivals at a bus stop can utilize the uniform distribution. If buses arrive uniformly between 10 to 20 minutes, passengers can expect a consistent wait time for each bus.

Normal Distribution

1. Introduction
    - Normal distribution, also known as Gaussian distribution, is a continuous probability distribution that is symmetric and bell-shaped.
    - It is widely used in statistics and probability theory due to its mathematical tractability and its prevalence in nature.
    - Many real-world phenomena, such as heights, weights, and IQ scores, can be modeled using the normal distribution.
2. Characteristics of Normal Distribution
    - Symmetry: The distribution is symmetric around its mean.
    - Bell-shaped curve: The probability density function (PDF) forms a bell-shaped curve.
    - Mean and variance: The mean (μ) represents the center of the distribution, and the variance (σ^2) controls the spread.
    - Standard deviation: The standard deviation (σ) is the square root of the variance and measures the average distance from the mean.

Normal Distribution (cont.)

1. Standard Normal Distribution
    - The standard normal distribution is a specific case of the normal distribution with a mean of 0 and a standard deviation of 1.
    - It is often denoted by Z and is used for standardizing values to calculate probabilities.
    - The standard normal distribution has a well-known cumulative distribution function (CDF) called the standard normal table or Z-table.

pdf of Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In this formula, x is the random variable, μ is the mean of the distribution, and σ is the standard deviation.

Properties of Normal Distribution

Gamma distribution

- The gamma distribution is a continuous probability distribution that is commonly used to model the wait times between events.
- It is a generalization of the exponential distribution and includes the exponential distribution as a special case.
- The gamma distribution has two shape parameters: α (shape) and β (scale).
- Probability Density Function (PDF):

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $\Gamma(\alpha)$ is the gamma function.

Probability Density Function

- The probability density function (PDF) of the gamma distribution is defined as:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

- Mean: $\mu = \frac{\alpha}{\beta}$ and Variance: $\sigma^2 = \frac{\alpha}{\beta^2}$.
- The gamma distribution is right-skewed for $\alpha < 1$ and becomes more symmetric as α increases.
- As α approaches infinity, the gamma distribution approaches a normal distribution.

Applications and Examples

- The gamma distribution has various applications in different fields:
  - Reliability engineering: Modeling lifetimes of electronic components.
  - Finance: Modeling waiting times between stock price changes.
  - Queuing theory: Modeling inter-arrival times of customers.
- Example: Consider a manufacturing process where the time between equipment failures follows a gamma distribution with $\alpha = 2$ and $\beta = 0.5$.
- Probability of failure within 5 hours:

$$P(X \leq 5) = \int_0^5 f(x|2, 0.5)\, dx$$

Applications and Examples

- The gamma distribution has various applications in real-world scenarios:
  - **Example 1:** **Service Time Modeling**
    In a customer service setting, the time a customer service representative spends assisting a customer can be modeled using a gamma distribution. If the average service time is 10 minutes and the shape parameter is 2, it can capture the variability in service times.
  - **Example 2:** **Machine Repair Time**
    In manufacturing, the time it takes to repair a machine after it fails can be modeled with a gamma distribution. If the shape parameter is low, it indicates a high probability of quick repairs, while a higher shape parameter captures longer repair times.

Exponential Distribution

The Exponential Distribution is a fundamental probability distribution with wide applications in various fields.

- Memoryless property
- Describes time between events in a Poisson process

Definition and Properties

Definition: The Exponential Distribution is defined by its probability density function (PDF):

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

where λ is the rate parameter.

- Probability of event in interval $[x, x + \Delta x]$ is $\lambda e^{-\lambda x} \Delta x$ as Δx approaches 0.

Properties:

- Memoryless: Future events are independent of past events.
- Mean: $\mu = \frac{1}{\lambda}$
- Variance: $\sigma^2 = \frac{1}{\lambda^2}$

Definition and Properties

Definition: The Exponential Distribution is defined by its probability density function (PDF):

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

where λ is the rate parameter.

- Probability of event in interval $[x, x + \Delta x]$ is $\lambda e^{-\lambda x} \Delta x$ as Δx approaches 0.

Properties:

- Memoryless: Future events are independent of past events.
- Mean: $\mu = \frac{1}{\lambda}$
- Variance: $\sigma^2 = \frac{1}{\lambda^2}$

Cumulative Distribution Function (CDF)

The Cumulative Distribution Function (CDF) of the Exponential Distribution is:

$$F(x) = 1 - e^{-\lambda x}$$

- Represents the probability that a random variable X is less than or equal to x.

The Rate Parameter λ

- λ is a positive number representing the average rate of event occurrences.
- It quantifies how frequently events happen per unit time (or distance).
- In other words, λ indicates the average number of events that occur in a given timeframe.

Applications

Reliability Analysis:

- Modeling time until failure in systems with constant hazard rates.

Queuing Theory:

- Analyzing waiting times between arrivals in service systems.

Financial Modeling:

- Modeling time between extreme events in finance.

Conclusion

The Exponential Distribution is a versatile tool with wide-ranging applications due to its memoryless property.

- Foundational concept in probability theory and statistics.
- Memoryless property: Significant in various fields.

As researchers explore more complex distributions, the Exponential Distribution remains a crucial building block for statistical analysis and inference.

Example 1: Reliability Analysis

Scenario: Manufacturing Company and Component Failure

- Analyzing time until failure for electronic component.
- Goal: Understand average operational lifetime.

Solution:

- Average time until failure: 500 hours ($\lambda = \frac{1}{500}$)
- Mean (μ) = 500 hours
- Variance (σ^2) = 500^2 hours^2

Introduction to the Chi-Square Distribution

- Probability distribution used in statistics.
- Widely applied in hypothesis testing and confidence interval estimation.

Definition

Definition of the Chi-Square Distribution

- A continuous probability distribution.
- Parameter: Degrees of Freedom (df).
- Probability density function (PDF):

$$f(x; df) = \frac{1}{2^{df/2}\Gamma(df/2)} \cdot x^{(df/2)-1} \cdot e^{-x/2}$$

Properties

Properties of the Chi-Square Distribution

- Non-negative values: $x \geq 0$
- Shape depends on degrees of freedom.
- Mean: $E(X) = df$
- Variance: $Var(X) = 2df$

Applications

Applications of the Chi-Square Distribution

- Hypothesis testing: Compare observed data with expected values.
- Goodness-of-fit tests: Assess if observed data fits a theoretical distribution.
- Confidence interval estimation: Determine the interval where a parameter is likely to fall.

Example

Example: Goodness-of-Fit Test

- Scenario: Testing whether observed data follows a specific distribution.
- Steps: Calculate chi-square statistic and compare with critical value.
- Decision: If statistic exceeds critical value, reject null hypothesis.

Conclusion

- Chi-Square Distribution is versatile in statistics.
- Key role in hypothesis testing, confidence intervals, and goodness-of-fit tests.
- Understanding its properties and applications is essential for data analysis.

Sum of Independent Random Variables

- Consider the sum of independent random variables from various distributions.
- Explore how properties of the sum differ based on the underlying distributions.

Sum of Independent Bernoulli Random Variables

- **Distribution**: Bernoulli random variables with parameter p.
- **Properties**:
  - Sum of n independent Bernoulli RVs follows a Binomial distribution.
  - Parameters: n (number of trials), p (probability of success).
  - Mean: np, Variance: $np(1 - p)$.

Sum of Independent Poisson Random Variables

- **Distribution**: Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$.
- **Properties**:
  - Sum of n independent Poisson RVs follows a Poisson distribution.
  - Parameter: $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.
  - Mean: λ, Variance: λ.

Sum of Independent Exponential Random Variables

- **Distribution**: Exponential random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$.
- **Properties**:
  - Sum of n independent Exponential RVs follows a Gamma distribution.
  - Parameters: n (shape), $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ (rate).
  - Mean: $\frac{n}{\lambda}$, Variance: $\frac{n}{\lambda^2}$.

Sum of Independent Chi-Square Random Variables

- **Distribution**: Chi-square random variables with degrees of freedom $k_1, k_2, \ldots, k_n$.
- **Properties**:
  - Sum of n independent Chi-square RVs follows a Chi-square distribution.
  - Degrees of freedom: $k = k_1 + k_2 + \ldots + k_n$.
  - Mean: k, Variance: $2k$.

Sum of Independent Normal Random Variables

- **Distribution**: Normal random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.
- **Properties**:
    - Sum of n independent Normal RVs follows a Normal distribution.
    - Mean: $\mu = \mu_1 + \mu_2 + \ldots + \mu_n$.
    - Variance: $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$.

Conclusion

- Summing independent random variables from different distributions yields various new distributions.
- Properties of the sum depend on the properties of the individual distributions.
- Understanding these properties is crucial in probability theory and statistics.