Chapter 17

Probability & Statistics

Probability and statistics provide the mathematical foundation for modeling uncertainty, analyzing data, and making quantitative decisions under incomplete information. Probability deals with the theoretical laws governing random phenomena, while statistics concerns the extraction of information from data using probabilistic models. Mathematically, an outcome is said to be random if it cannot be predicted with certainty in advance and is governed only by a probability law.

17.1 Basic Concepts of Probability

17.1.1 Random Experiments and Sample Space

A **random experiment** is a process whose outcome cannot be predicted with certainty. The set of all possible outcomes of a random experiment is called the **sample space** and is denoted by Ω. Any subset $A \subseteq \Omega$ is called an **event**.

17.1.2 Axioms of Probability

A probability measure $P(A)$ satisfies the following axioms:

1. $0 \leq P(A) \leq 1$ for any event A,
2. $P(\Omega) = 1$,
3. For mutually exclusive events A_i,

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)}$$

17.1.3 Conditional Probability and Bayes' Theorem

The **conditional probability** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad P(A|B) \times P(B) = P(B|A) \times P(A)$$

Bayes' theorem:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \qquad \boxed{P(B|A) = \frac{P(A|B)P(B)}{P(A)}}$$

17.2 Random Variables and Probability Distributions

17.2.1 Discrete Random Variables

A **discrete random variable** takes countable values x_i with **probability mass function (PMF)**

$$P(X = x_i) = p(x_i)$$

17.2.2 Continuous Random Variables

A **continuous random variable** has a **probability density function (PDF)** $f(x)$ such that

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

17.2.3 Cumulative Distribution Function (CDF)

The CDF is defined as

$$F(x) = P(X \leq x)$$

17.3 Mathematical Expectation and Moments

17.3.1 Mean and Variance

The **expectation** of a random variable is

$$\boxed{E[X] = \sum_{i=1}^{n} x_i p(x_i)} \quad \text{(discrete)}$$

$$\boxed{E[X] = \int_{-\infty}^{\infty} x f(x)\,dx} \quad \text{(continuous)}$$

The **variance** is

$$\boxed{\mathrm{Var}(X) = E[(X - E[X])^2]}$$

17.3.2 Higher-Order Moments

The **n-th moment** is

$$\boxed{\mu_n = E[X^n]}$$

17.4 Standard Probability Distributions

Many random phenomena encountered in science and engineering can be modeled using a small collection of fundamental probability distributions. These distributions characterize the statistical behavior of discrete and continuous random variables.

17.4.1 Discrete Distributions

17.4.1.1 Bernoulli Distribution

A random variable X is said to follow a **Bernoulli distribution** with parameter p, denoted by $X \sim \text{Bern}(p)$, if it takes the value

$$X = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p \end{cases}$$

The probability mass function (PMF), mean and variance are:

$$\boxed{P(X=x) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}} \quad \boxed{E[X] = p} \quad \boxed{\text{Var}(X) = p(1-p)}$$

This distribution models a **single trial** with two possible outcomes, such as success or failure.

17.4.1.2 Binomial Distribution

A random variable X follows a **Binomial distribution** with parameters n and p, denoted by $X \sim \text{Bin}(n,p)$, if it represents the number of successes in n independent Bernoulli trials. The probability mass function (PMF), mean and variance are:

$$\boxed{P(X=k) = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0,1,2,\ldots,n} \quad \boxed{E[X] = np} \quad \boxed{\text{Var}(X) = np(1-p)}$$

This distribution is widely used in quality control, reliability analysis, and sampling theory.

17.4.1.3 Poisson Distribution

A random variable X follows a **Poisson distribution** with rate parameter $\lambda > 0$, denoted by $X \sim \text{Poisson}(\lambda)$, if it represents the number of events occurring in a fixed interval of time or space, assuming the events occur independently and at a constant average rate. The probability mass function (PMF), mean, and variance are:

$$\boxed{P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2,\ldots} \quad \boxed{E[X] = \lambda} \quad \boxed{\text{Var}(X) = \lambda}$$

The Poisson distribution models the number of events occurring in a fixed interval of time or space, such as radioactive decay or arrival of customers.

17.4.1.4 Geometric Distribution

A random variable X follows a **Geometric distribution** with parameter p, denoted by $X \sim \text{Geom}(p)$, if it represents the number of trials needed to obtain the first success. The probability mass function (PMF), mean, and variance are:

$$\boxed{P(X=k) = (1-p)^{k-1}p, \quad k = 1,2,3,\ldots} \quad \boxed{E[X] = \frac{1}{p}} \quad \boxed{\text{Var}(X) = \frac{1-p}{p^2}}$$

This distribution exhibits the memoryless property.

17.4.2 Continuous Distributions

17.4.2.1 Uniform Distribution

A random variable X follows a **Uniform distribution** on the interval $[a, b]$, denoted by $X \sim U(a, b)$. The Probability Density Function (PDF), Mean and Variance are given by:

$$\boxed{f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}} \quad \boxed{E[X] = \frac{a+b}{2}} \quad \boxed{\mathrm{Var}(X) = \frac{(b-a)^2}{12}}$$

All values in the interval $[a, b]$ are equally likely.

17.4.2.2 Exponential Distribution

A random variable X follows an **Exponential distribution** with parameter $\lambda > 0$, denoted by $X \sim \mathrm{Exp}(\lambda)$. The Probability Distribution Function, Mean and Variance are:

$$\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}} \quad \boxed{E[X] = \frac{1}{\lambda}} \quad \boxed{\mathrm{Var}(X) = \frac{1}{\lambda^2}}$$

This distribution models waiting times between successive random events and also exhibits the memoryless property.

17.4.2.3 Normal (Gaussian) Distribution

A random variable X follows a **Normal distribution** with mean μ and variance σ^2, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$. The Probability Distribution Function, Mean and Variance are:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \quad \boxed{E[X] = \mu} \quad \boxed{\mathrm{Var}(X) = \sigma^2}$$

It plays a central role in probability theory due to the **Central Limit Theorem**.

17.4.2.4 Gamma Distribution

A random variable X follows a **Gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$, denoted by $X \sim \Gamma(\alpha, \beta)$. The Probability Distribution Function, Mean and Variance are:

$$\boxed{f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}} \quad \boxed{E[X] = \alpha\beta} \quad \boxed{\mathrm{Var}(X) = \alpha\beta^2}$$

The Gamma distribution generalizes both the Exponential distribution and the Erlang distribution and is widely used in queueing theory and reliability analysis.

17.5 Joint Distributions and Independence

17.5.1 Joint Probability Functions

For two continuous random variables X and Y, the joint probability density function (PDF) is denoted by $f_{X,Y}(x, y)$.

17.5.2 Independence

The random variables X and Y are said to be **independent** if their joint PDF factors as

$$\boxed{f_{X,Y}(x,y) = f_X(x) f_Y(y)}$$

17.5.3 Covariance and Correlation

The **covariance** between X and Y is defined as

$$\boxed{\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]}$$

The **correlation coefficient** is defined as

$$\boxed{\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}}$$

17.6 Law of Large Numbers and Central Limit Theorem

17.6.1 Law of Large Numbers

The sample mean converges to the population mean:

$$\boxed{\bar{X}_n \to E[X] \quad \text{as } n \to \infty}$$

17.6.2 Central Limit Theorem

If $X_1, \ldots, X_n$ are independent with mean μ and variance σ^2, then

$$\boxed{\frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1)}$$

$\mathcal{N}(0,1)$ is the standard normal distribution with mean 0 and variance 1.

17.7 Statistical Data Analysis

Statistical data analysis deals with the systematic collection, organization, presentation, and interpretation of data for extracting meaningful information and supporting decision-making under uncertainty.

17.7.1 Organization of Data

Raw data are organized to facilitate analysis.

17.7.1.1 Frequency Tables

A **frequency table** lists data values with corresponding frequencies. If f_i denotes the frequency of x_i, then

$$\sum_{i=1}^{k} f_i = n$$

17.7.1.2 Grouped Data Tables

For large or continuous datasets, observations are grouped into equal-width class intervals with corresponding frequencies.

17.7.1.3 Cumulative Frequency Distributions

The cumulative frequency up to class i is

$$F_i = \sum_{j=1}^{i} f_j$$

and is used to determine medians, quartiles, and percentiles.

17.7.2 Descriptive Statistics

Descriptive statistics summarize large datasets using numerical measures.

17.7.2.1 Measures of Central Tendency

Measures of central tendency describe the typical or representative value of a dataset.

- **Arithmetic Mean:** The mean is the average value of the data and is defined by

  $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  where $x_1, x_2, \ldots, x_n$ are the observed values.
- **Median:** The median is the central value of the dataset after arranging the observations in ascending or descending order.
  - If n is **odd**, the median is

    $$\text{Median} = x_{\frac{n+1}{2}}$$

  - If n is **even**, the median is

    $$\text{Median} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

- **Mode:** The mode is the value that occurs with the highest frequency in the dataset.
  - A dataset may be **unimodal**, **bimodal**, or **multimodal**.
  - If all values occur with equal frequency, the dataset has **no mode**.

17.7.3 Measures of Dispersion

- **Range:** $\boxed{x_{\text{max}} - x_{\text{min}}}$
- **Variance:** $\boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- **Standard Deviation:** $\boxed{\sigma = \sqrt{\sigma^2}}$

17.7.4 Skewness and Kurtosis

Skewness tells whether a distribution is symmetric or tilted left/right.

$$\gamma_1 = \frac{1}{n\sigma^3} \sum_{i=1}^{n} (x_i - \bar{x})^3$$

Kurtosis : tells how peaked the distribution is and how heavy its tails are.

$$\gamma_2 = \frac{1}{n\sigma^4} \sum_{i=1}^{n} (x_i - \bar{x})^4 - 3$$

17.7.5 Correlation Analysis

The **Pearson correlation coefficient** is

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad -1 \le r \le 1$$

17.7.6 Regression Analysis

The **simple linear regression model** is

$$y = a + bx$$

where a and b are obtained by the least squares method.

17.8 Estimation Theory

Estimation theory deals with the use of sample data to infer the numerical values of unknown population parameters. Since complete population information is rarely available, statistical inference relies on carefully constructed estimators and confidence intervals to quantify uncertainty in parameter estimation.

17.8.1 Point Estimation

In point estimation, an unknown population parameter is estimated using a single numerical value computed from a *sample*. Since the entire population is usually unavailable, all estimation is necessarily based on sample data.

A statistic $\hat{\theta}$ calculated from a sample is called a **point estimator** of the population parameter θ. The numerical value obtained after computation is called the **point estimate** .

Common point estimators include:

$$\hat{\mu} = \bar{x}, \qquad \hat{\sigma}^2 = s^2$$

where $\bar{x}$ is the sample mean and s^2 is the sample variance.

Because point estimates are based on samples, they are subject to sampling variability and do not exactly equal the true population parameters.

17.8.1.1 Properties of Good Estimators

A point estimator is considered **good** if it satisfies the following important properties:

- **Unbiasedness:** An estimator $\hat{\theta}$ is said to be unbiased if its expected value equals the true population parameter:

  $$E[\hat{\theta}] = \theta$$

  This means that, on average, the estimator neither overestimates nor underestimates the parameter.

- **Consistency:** An estimator is consistent if it becomes closer to the true parameter value as the sample size increases:

  $$\hat{\theta} \to \theta \quad \text{as } n \to \infty$$

  Thus, large samples improve the accuracy of the estimate.

- **Efficiency (Minimum Variance):** Among all unbiased estimators of a parameter, the estimator with the smallest variance is called the most efficient, since it shows the least random fluctuation from sample to sample.

- **Sufficiency:** A statistic is sufficient if it contains all the information in the sample relevant to the estimation of the parameter, so that no additional sample data can improve the estimate.

17.8.2 Interval Estimation

Point estimation provides only a single numerical value as an estimate of an unknown population parameter. However, since all estimates are based on sample data, there is always **sampling uncertainty. Interval estimation** addresses this limitation by specifying a range of values within which the true population parameter is expected to lie with a known level of confidence.

A **confidence interval** for a population parameter θ is written as

$$P(\theta_1 \le \theta \le \theta_2) = 1 - \alpha$$

where:

- $1 - \alpha$ is the **confidence level**
- α is the **level of significance**

This statement means that, over many repeated samples, a proportion $(1 - \alpha)$ of the constructed intervals will contain the true parameter value.

17.8.2.1 Confidence Interval for the Population Mean (Known Variance)

If the population variance σ^2 is known and the population is normally distributed (or the sample size is sufficiently large), the confidence interval for the population mean μ is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where:

- ▷ $\bar{x}$ is the sample mean
- ▷ n is the sample size
- ▷ σ is the population standard deviation
- ▷ $z_{\alpha/2}$ is the **critical value** from the standard normal distribution

The **critical value** $z_{\alpha/2}$ represents the number of standard deviations from the mean required to achieve the chosen confidence level $1-\alpha$. It determines the width of the confidence interval.

17.8.2.2 Confidence Interval for the Population Mean (Unknown Variance)

If the population variance is unknown and the sample size is small, the confidence interval for the population mean μ is based on Student's t-distribution:

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

where:

- ▷ $\bar{x}$ is the sample mean
- ▷ n is the sample size
- ▷ s is the sample standard deviation
- ▷ $t_{\alpha/2, n-1}$ is the **critical value** from the t-distribution with $n-1$ degrees of freedom

The t-distribution accounts for the additional uncertainty introduced by estimating the population variance from the sample.

17.8.3 Confidence Interval for the Population Variance

If the population is normally distributed, the confidence interval for the population variance σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

where:

- ▷ s^2 is the sample variance
- ▷ n is the sample size
- ▷ $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are the **critical values** from the chi-square distribution with $n-1$ degrees of freedom

This interval accounts for the sampling uncertainty in estimating the population variance.

17.8.4 Sample Size Determination (Known Population Standard Deviation)

If the population standard deviation σ is known, the required sample size for estimating the population mean with maximum allowable error E and confidence level $1-\alpha$ is

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$$

where:

- n is the required sample size
- σ is the population standard deviation
- E is the maximum allowable error (margin of error)
- $z_{\alpha/2}$ is the critical value from the standard normal distribution

This ensures that the margin of error does not exceed E.

17.8.5 Sample Size Determination (Unknown Population Standard Deviation)

If the population standard deviation is unknown, an estimate s obtained from a pilot sample is used:

$$n = \left(\frac{z_{\alpha/2}\, s}{E}\right)^2$$

where:

- n is the required sample size
- s is the estimated standard deviation
- E is the maximum allowable error
- $z_{\alpha/2}$ is the critical value from the standard normal distribution

This approximation is widely used in practice when prior knowledge of population variability is unavailable.

17.9 Statistical Hypothesis Testing

Statistical hypothesis testing is a formal method for drawing conclusions about population parameters using sample data. It provides a probabilistic framework for testing scientific claims in the presence of uncertainty.

17.9.1 Null and Alternative Hypotheses

A **statistical hypothesis** is a statement about a population parameter. Two competing hypotheses are formulated:

- **Null Hypothesis** (H_0): The default assumption that no effect, difference, or change exists.
- **Alternative Hypothesis** (H_1): The competing claim that contradicts H_0.

Common forms of hypotheses are:

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0 \quad \text{(two-tailed test)}$$

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0 \quad \text{(right-tailed test)}$$

$$H_0 : \mu \geq \mu_0, \qquad H_1 : \mu < \mu_0 \quad \text{(left-tailed test)}$$

17.9.2 Level of Significance and Errors

The **level of significance** α is the probability of rejecting the null hypothesis when it is actually true. Typical values are

$$\alpha = 0.10, \quad 0.05, \quad 0.01$$

Two types of decision errors may occur:

- **Type I Error** : Rejecting H_0 when H_0 is true (probability = α)
- **Type II Error** : Failing to reject H_0 when H_0 is false (probability = β)

The **power of a test** is defined as

$$\text{Power} = 1 - \beta$$

It represents the probability of correctly rejecting a false null hypothesis.

17.9.3 Test Statistics and Decision Rules

A **test statistic** is a numerical quantity computed from sample data whose sampling distribution is known when H_0 is true. Decisions are made using either:

- The **critical region** method
- The **p-value** method

Decision rules:

- Reject H_0 if the test statistic falls in the critical region
- Reject H_0 if p-value $\leq \alpha$

17.9.4 Common Statistical Tests

17.9.4.1 z-Test

Used for testing a population mean when the population variance is known and the sample size is large:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}, \qquad z \sim \mathcal{N}(0, 1) \text{ under } H_0$$

17.9.4.2 t-Test

Used when the population variance is unknown and the sample size is small:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad t \sim t_{n-1} \text{ under } H_0$$

17.9.4.3 Chi-Square (χ^2) Test

Used for:

- testing population variance

- goodness-of-fit tests
- tests of independence in contingency tables

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

17.9.4.4 *F*-Test

Used to compare two population variances and in analysis of variance (ANOVA):

$$F = \frac{s_1^2}{s_2^2}$$

17.9.5 Procedure for Hypothesis Testing

The standard steps are:

1. Formulate H_0 and H_1
2. Specify the level of significance α
3. Select the appropriate test statistic
4. Determine the critical region or compute the p-value
5. Make the statistical decision and interpret the result

17.9.6 Interpretation of Results

Rejecting H_0 does not prove that H_1 is absolutely true, and failing to reject H_0 does not prove that H_0 is true. All conclusions from hypothesis testing are **probabilistic** and subject to sampling uncertainty.

17.10 Regression and Correlation Analysis

Regression and correlation analysis quantify the strength, direction, and functional form of relationships between variables.

17.10.1 Simple Linear Regression

The **simple linear regression model** is

$$y = a + bx$$

where a is the intercept and b is the regression coefficient.

17.10.2 Least Squares Method

The least squares estimates of a and b satisfy the normal equations:

$$\sum y = na + b \sum x$$
$$\sum xy = a \sum x + b \sum x^2$$

17.11 Stochastic Processes (Introduction)

A **stochastic process** is a family or collection of random variables indexed by a parameter, usually time. If $X(t)$ denotes the value of a random variable at time t, then a stochastic process is denoted by

$$\{X(t),\ t \in T\}$$

where T is the index set, typically representing time. If T is discrete (e.g., $t = 0, 1, 2, \dots$), the process is called a **discrete-time stochastic process**. If T is continuous (e.g., $t \geq 0$), it is called a **continuous-time stochastic process**.

Unlike ordinary random variables, which describe a single random outcome, stochastic processes model the **evolution of randomness over time**. They are widely used to represent systems that evolve in an uncertain manner, such as communication networks, stock prices, queuing systems, radioactive decay, population growth, and traffic flow.

17.11.1 Markov Processes

A **Markov process** is a stochastic process that satisfies the **Markov property**, which states that the future evolution of the process depends only on its present state and not on its past history. Mathematically,

$$P(X_{n+1} = x \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots) = P(X_{n+1} = x \mid X_n = x_n)$$

Markov processes are widely used in queueing theory, reliability analysis, inventory control, and machine learning.

17.11.2 Poisson Processes

A **Poisson process** is a continuous-time stochastic process that models the occurrence of random events over time. It is characterized by the following properties:

- The number of events in disjoint time intervals are independent.
- The probability of one event occurring in a short interval Δt is approximately $\lambda \Delta t$.
- The probability of more than one event occurring in a short interval is negligible.

If $N(t)$ denotes the number of events up to time t, then

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Poisson processes are used to model arrivals of customers, system failures, network traffic, and radioactive emissions.

17.11.3 Random Walks

A **random walk** is a stochastic process in which a variable evolves by taking successive random steps. A simple one-dimensional random walk is defined by

$$X_n = X_{n-1} + Y_n$$

where Y_n takes values $+1$ or -1 with equal probability. Random walks form the foundation for modeling diffusion, Brownian motion, stock market fluctuations, and search algorithms.

Importance of Stochastic Processes

Stochastic processes provide the mathematical framework for analyzing systems that evolve under uncertainty. They are fundamental in:

- Communication and computer networks
- Financial modeling and option pricing
- Control systems and signal processing
- Queueing systems and operations research
- Physics, biology, and economics

Thus, stochastic processes extend probability theory from single random experiments to entire random evolutions over time.