GeeksforGeeks

GATE फर्रें

DATA SCIENCE & ARTIFIACL INTELLIGENCE (DA)

PROBABILITY & STATISTICS

SHORT NOTES

Probability & Statistics GATE DA

1. Counting

1.1 Basic Counting Principles

THE PRODUCT RULE Suppose that a procedure can be broken down into a sequence of two tasks. If there are n_1 ways to do the first task and for each of these ways of doing the first task, there are n_2 ways to do the second task, then there are $n_1 n_2$ ways to do the procedure.

THE SUM RULE If a task can be done either in one of n_1 ways or in one of n_2 ways, where none of the set of n_1 ways is the same as any of the set of n_2 ways, then there are $n_1 + n_2$ ways to do the task.

Principle of inclusion–exclusion

If a task can be done in either n_1 ways or n_2 ways, then the number of ways to do the task is $n_1 + n_2$ minus the number of ways to do the task that are common to the two different ways.
Suppose that A1 and A2 are sets.
$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$.

1.2 Pigeon Hole Principle

THE PIGEONHOLE PRINCIPLE If k is a positive integer and $k + 1$ or more objects are placed into k boxes, then there is at least one box containing two or more of the objects.
THE GENERALIZED PIGEONHOLE PRINCIPLE If N objects are placed into k boxes, then there is at least one box containing at least $\lceil N/k \rceil$ objects.

1.3 Permutation

A permutation is an arrangement in a definite order of a number of objects taken some or all at a time.
The number of r-permutations of a set with n elements is denoted by P(n, r).
If n and r are integers with $0 \le r \le n$, then $P(n,r) = \frac{n!}{(n-r)!}$.

1.4 Combinations

An r-combination of elements of a set is an unordered selection of r elements from the set.
The number of r-combinations of a set with n elements, where n is a nonnegative integer and r is an integer with $0 \le r \le n$, equals
$C(n,r) = \frac{n!}{r!(n-r)!}$

1.5 Binomial Theorem

THE BINOMIAL THEOREM Let x and y be variables, and let n be a nonnegative integer. Then
$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j = \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n$.

Note:
1. Let n and r be nonnegative integers with $r \le n$. Then $C(n,r) = C(n, n-r)$.
2. PASCAL'S IDENTITY Let n and k be positive integers with $n \ge k$. Then
$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$.

1.6 Generalized Permutations and Combinations

Type	Condition	Formula	Explanation
Permutation (Without Repetition)	Order matters, no repetition	$P(n,r) = \frac{n!}{(n-r)!}$	Selecting and arranging r items from n distinct items
Permutation (With Repetition)	Order matters, repetition allowed	n^r	Selecting r items from n options, each can repeat
Combination (Without Repetition)	Order doesn't matter, no repetition	$C(n,r) = \frac{n!}{r!\,(n-r)!}$	Selecting r items from n distinct items, order doesn't matter
Combination (With Repetition)	Order doesn't matter, repetition allowed	$C(n+r-1,r) = \frac{(n+r-1)!}{r!\,(n-1)!}$	Selecting r items from n types, repetitions allowed, order irrelevant

Note:

The number of ways to distribute n distinguishable objects into k distinguishable boxes so that n_i objects are placed into box $i, i = 1,2,\ldots,k$, equals $\frac{n!}{n_1!n_2!\cdots n_k!}$

1.7 Distribution of r balls into n boxes

	n distinguishable boxes		n indistinguishable boxes	
	empty box allowed	no box empty	empty box allowed	no box empty
r disting. balls	n^r	$n!\left\{{r \atop n}\right\}$	$\sum_{i=1}^{n}\left\{{r \atop i}\right\}$	$\left\{{r \atop n}\right\}$
r indisting. balls	$\binom{r+n-1}{r}$	$\binom{r-1}{n-1}$	$\sum_{i=1}^{n}\left\vert{r \atop i}\right\vert$	$\left\vert{r \atop n}\right\vert$

- $\left\vert{r \atop n}\right\vert$: Number of integer partitions of r into n parts (used for indistinguishable balls & boxes).
- $\binom{a}{b}$: Binomial coefficient.
- $S(r,i) = \left\{{r \atop i}\right\} = \frac{1}{i!}\sum_{k=0}^{i}(-1)^k\binom{i}{k}(i-k)^r$

2. Axioms of Probability

2.1 Sample Space and Events

Consider an experiment whose outcome will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Any subset E of the sample space is known as an event

2.2 Probability

Probability refers to the extent of occurrence of events.

$$P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of equally likely outcomes}} = \frac{n(E)}{n(S)}$$

Where:

- P(E): Probability of event E
- n(E): Number of favourable outcomes for event E
- n(S): Total number of outcomes in the sample space

Note: Let A, E, F are the events associated with a random experiment, then

(i) Probability of non occurrence of event A, i.e, P(A') = 1 – P(A)

(ii)If E is a subset of F, then P(E) <= P(F)

(iii) P(E ∪ F) = P(E) + P(F) − P(E ∩ F)

The three axioms of probability

Axiom 1

$$0 \leq P(E) \leq 1$$

Axiom 2

$$P(S) = 1$$

Axiom 3

For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i E_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

We refer to $P(E)$ as the probability of the event E.

2.3 Types of Events

1. Sure (Certain) Event:

An event that is guaranteed to happen.
Formula: $P(E) = 1$

2. Impossible Event:

An event that cannot occur under any circumstances.
Formula: $P(E) = 0$

3. Simple (Elementary) Event:

An event that consists of only a single outcome.

4. Compound (Composite) Event:

An event that consists of two or more outcomes.

5. Mutually Exclusive Events:

Two events are mutually exclusive if they cannot occur at the same time.
Formula: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = 0$

6. Exhaustive Events:

A set of events is exhaustive if at least one of them must occur.
Formula: $P(E_1) + P(E_2) + \cdots + P(E_n) = 1$

7. Independent Events:

Events are independent if the occurrence of one does not affect the occurrence of the other.
Formula: $P(A \cap B) = P(A) \cdot P(B)$

8. Dependent Events:

Events are dependent if the occurrence of one affects the probability of the other.
Formula: $P(A \cap B) = P(A) \cdot P(B \mid A)$

9. Complementary Events:

Two events are complementary if one occurs exactly when the other does not.
Formula: $P(E') = 1 - P(E)$

3. Conditional Probability & Baye's Theorem

3.1 Conditional Probability

The conditional probability of an event A given that another event B has occurred is the probability of A occurring under the condition that B has already occurred.

Formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$

- $P(A \mid B)$: Probability of A given B
- $P(A \cap B)$: Probability that both A and B occur
- $P(B)$: Probability that B occurs

Multiplication Rule of Probability

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

3.2 Law of Total Probability

If $B_1, B_2, \ldots, B_n$ are mutually exclusive and exhaustive events (i.e., one of them must occur), and A is any event, then:

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

- Useful when the probability of A depends on different cases $B_1, B_2, \ldots, B_n$
- Requires: $B_1 \cup B_2 \cup \cdots \cup B_n = S$ and $B_i \cap B_j = \emptyset$ for $i \neq j$

3.2 Baye's Theorem

$P(B \mid A)$: Probability of event B given A occurred

$$P(B \mid A) = \frac{P(B) \cdot P(A \mid B)}{P(A)}$$

3.4 Generalized Baye's theorem

Used to reverse conditional probabilities: Find $P(B_i \mid A)$ when $P(A \mid B_i)$ is known.

$$P(B_i \mid A) = \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j) \cdot P(A \mid B_j)}$$

- $B_1, B_2, \ldots, B_n$: Partition of the sample space
- A : Observed evidence/event
- Denominator is from Law of Total Probability

4. Key measures of central tendency and dispersion

4.1 Mean

The average value of a data set.

Formula:

$$\text{Mean } = \bar{x} = \frac{\sum x_i}{n}$$

Where x_i are the data values and n is the total number of values.

4.2 Median

The middle value when the data is arranged in ascending or descending order.

- If n is odd: median is the middle term
- If n is even: median is the average of the two middle terms (No formula needed; depends on sorting)

4.3 Mode

The value(s) that occur most frequently in the dataset.

- A dataset may be unimodal, bimodal, or multimodal

4.4 Variance

A measure of how much the data values deviate from the mean.

Formula (Population):

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Formula (Sample):

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

4.5 Standard Deviation

The square root of the variance; gives spread in the same units as the data.

Formula:

$$\sigma = \sqrt{\sigma^2} \text{ or } s = \sqrt{s^2}$$

5. Random Variable

5.1 Random Variable - Definition

A random variable is a function that assigns a real number to each outcome of a random experiment.

5.2 Types of Random Variables

- **Discrete Random Variable:**
  Takes countable values (finite or countably infinite)
  Example: Number of heads in 3 coin tosses

- **Continuous Random Variable:**
  Takes uncountably infinite values (within intervals)
  Example: Temperature in a day

5.3 Probability Distribution Function (PDF / PMF)

A. Discrete Random Variable (PMF - Probability Mass Function)

For a discrete random variable X :

$$P(X = x_i) = p_i,$$

where

- x_i : possible value that X can take
- $p_i = P(X = x_i)$: probability that X takes the value x_i

$$\sum_i p_i = 1 \text{ and } 0 \leq p_i \leq 1$$

B. Continuous Random Variable (PDF - Probability Density Function)

For a continuous random variable X :

- $f(x)$: probability density function, such that

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

Conditions:

$$f(x) \geq 0 \text{ for all } x \text{, and } \int_{-\infty}^{\infty} f(x)dx = 1$$

5.4 Cumulative Distribution Function (CDF)

Gives the probability that X takes a value less than or equal to x :

- For discrete:

$$F(x) = \sum_{x_i \leq x} P(X = x_i)$$

- For continuous:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Relationship between PDF and CDF (Continuous Case)

If $F(x)$ is the Cumulative Distribution Function (CDF) of a continuous random variable X, then:

$$f(x) = \frac{d}{dx}F(x)$$

That is, the PDF is the derivative of the CDF.

5.5 Expected Value (Mean)

- For Discrete:

$$E(X) = \sum x_i \cdot P(x_i)$$

- For Continuous:

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

5.6 Variance and Standard Deviation

- **Variance:**

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

- **Standard Deviation:**

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

5.7 Properties of Expectation

Let X and Y be random variables and $a, b \in \mathbb{R}$:

1. **Linearity:**

$$E(aX + b) = aE(X) + b$$

2. **Sum Rule:**

$$E(X + Y) = E(X) + E(Y)$$

3. **If X is constant c :**

$$E(c) = c$$

4. If X and Y are independent random variables, $E(XY) = E(X)E(Y)$ |

5.8 Properties of Variance

1. **Scaling and Shifting:**

$$\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$$

2. **If X and Y are independent:**

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

5.9 Conditional Distribution

(a) Discrete Case - Conditional PMF

For discrete random variables X and Y, the conditional probability mass function of X given $Y = y$ is:

$$P_{X|Y}(x \mid y) = \frac{P(X = x, Y = y)}{P(Y = y)} \quad \text{if } P(Y = y) > 0$$

(b) Continuous Case - Conditional PDF

For continuous random variables X and Y, the conditional probability density function of X given $Y = y$ is:

$$f_{XYY}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0$$

5.10 Conditional Expectation

Conditional expectation gives the expected value of a random variable given that another variable has a specific value.

(a) Discrete Case

If X and Y are discrete random variables, the conditional expectation of X given $Y = y$ is:

$$E[X \mid Y = y] = \sum_{z} x \cdot P(X = x \mid Y = y)$$

(b) Continuous Case

If X and Y are continuous random variables, the conditional expectation of X given $Y = y$ is:

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)dx$$

Here, $f_{XY}(x \mid y)$ is the conditional PDF of X given $Y = y$.

5.11 Law of Total Expectation

(a) Discrete Case

$$E[X] = \sum_{y} E[X \mid Y = y] \cdot P(Y = y)$$

(b) Continuous Case

$$E[X] = \int_{-\infty}^{\infty} E[X \mid Y = y] \cdot f_Y(y)dy$$

6. Types of Random Variable
6.1 Discrete Distributions

1. Discrete Uniform Distribution

- **Notation**: $X \sim U(a, b)$
- **Definition**: Takes integer values from a to b with equal probability.
- **PMF**: $P(X = x) = \frac{1}{b-a+1}$, for $x \in \{a, a+1, \ldots, b\}$
- **CDF**: $F(x) = \frac{\lfloor x \rfloor - a + 1}{b-a+1}$, for $x \geq a$
- **Expectation**: $E[X] = \frac{a+b}{2}$
- **Variance**: $\mathrm{Var}(X) = \frac{(b-a+1)^2-1}{12}$
- **Standard Deviation**: $\sqrt{\mathrm{Var}(X)}$

2. Bernoulli Distribution

- **Notation**: $X \sim \mathrm{Bern}(p)$
- **Definition**: Random variable with two outcomes: success (1) and failure (0).
- **PMF**: $P(X = x) = p^x(1-p)^{1-x}$, for $x \in \{0,1\}$
- **CDF**: Step function at 0 and 1
- **Expectation**: $E[X] = p$
- **Variance**: $\mathrm{Var}(X) = p(1-p)$
- **Standard Deviation**: $\sqrt{p(1-p)}$

3. Geometric Distribution

- **Notation**: $X \sim \mathrm{Geo}(p)$
- **Definition**: Number of trials until first success.
- **PMF**: $P(X = x) = (1-p)^{x-1}p$, for $x = 1,2,3, \ldots$
- **CDF**: $F(x) = 1 - (1-p)^x$
- **Expectation**: $E[X] = \frac{1}{p}$
- **Variance**: $\mathrm{Var}(X) = \frac{1-p}{p^2}$
- **Standard Deviation**: $\sqrt{\frac{1-p}{p^2}}$

4. Binomial Distribution

- **Notation**: $X \sim \mathrm{Bin}(n, p)$
- **Definition**: Number of successes in n independent Bernoulli trials.
- **PMF**: $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, for $k = 0,1, \ldots, n$
- **CDF**: Sum of PMFs up to k
- **Expectation**: $E[X] = np$
- **Variance**: $\mathrm{Var}(X) = np(1-p)$
- **Standard Deviation**: $\sqrt{np(1-p)}$

5. Poisson Distribution

- **Notation**: $X \sim$ Poisson(λ)
- **Definition**: Models number of events in a fixed interval with average rate λ.
- **PMF**: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, for $k = 0,1,2, \ldots$
- **CDF**: Sum of PMFs up to k
- **Expectation**: $E[X] = \lambda$
- **Variance**: $\mathrm{Var}(X) = \lambda$
- **Standard Deviation** $\sqrt{\lambda}$

6.2 Continuous Distributions

1. Continuous Uniform Distribution

- **Notation**: $X \sim U(a, b)$
- **Definition**: Uniform density over interval $[a, b]$
- **PDF**: $f(x) = \frac{1}{b-a}$, for $a \leq x \leq b$
- **CDF**: $F(x) = \frac{x-a}{b-a}$, for $a \leq x \leq b$
- **Expectation**: $E[X] = \frac{a+b}{2}$
- **Variance**: $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
- **Standard Deviation**: $\frac{b-a}{\sqrt{12}}$

2. Exponential Distribution

- **Notation**: $X \sim \text{Exp}(\lambda)$
- **Definition**: Time between events in a Poisson process.
- **PDF**: $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$
- **CDF**: $F(x) = 1 - e^{-\lambda x}$
- **Expectation**: $E[X] = \frac{1}{\lambda}$
- **Variance**: $\text{Var}(X) = \frac{1}{\lambda^2}$
- **Standard Deviation**: $\frac{1}{\lambda}$

3. Normal Distribution

- **Notation**: $X \sim N(\mu, \sigma^2)$
- **Definition**: Bell-shaped symmetric distribution.
- **PDF**:

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

- **CDF**: No closed form; denoted by $\Phi(x)$
- **Expectation**: $E[X] = \mu$
- **Variance**: $\text{Var}(X) = \sigma^2$
- **Standard Deviation**: σ

4. Standard Normal Distribution

- **Notation**: $Z \sim N(0,1)$
- **Definition**: Normal distribution with mean 0 and variance 1.
- **PDF**:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- **CDF**: Denoted by $\Phi(z)$
- **Expectation**: $E[Z] = 0$
- **Variance**: $\text{Var}(Z) = 1$
- **Standard Deviation**: 1

6.3 Moment Generating Function (MGF)

- The Moment Generating Function of a random variable $\mathbf{X}$ is defined as:
  $$M_X(t) = \mathbb{E}[e^{tX}]$$
  Where t is a real number such that the expectation exists.

Why MGF?

- Helps in finding moments (mean, variance, etc.) of a distribution.
- Uniquely determines the distribution (if it exists in an open interval around 0).
- Simplifies calculations for the sum of independent random variables.

How to find moments using MGF:

- The $\mathbf{n}$-th moment about the origin is obtained by differentiating the MGF $\mathbf{n}$ times:
  $$\mu'_n = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}$$

- Example:
- Mean: $\mu = M'_X(0)$
- Variance: $\sigma^2 = M''_X(0) - (M'_X(0))^2$

Discrete Distributions:

1. **Uniform (Discrete)** $U(a,b)$
$$M_X(t) = \frac{1}{b-a+1} \sum_{x=a}^{b} e^{tx}$$
2. **Bernoulli(p)**
$$M_X(t) = (1-p) + pe^t$$
3. **Geometric(p)** (Number of trials until first success, starting from 1)
$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}, \text{ for } t < -\ln(1-p)$$
4. **Binomial(n, p)**
$$M_X(t) = [(1-p) + pe^t]^n$$

Negative Binomial(r, p) (Number of trials until r successes)

$$M_X(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r, t < -\ln(1-p)$$

5. **Poisson** (λ)
$$M_X(t) = \exp\big(\lambda(e^t - 1)\big)$$

Continuous Distributions:

1. **Uniform (Continuous)** $U(a,b)$
$$M_X(t)=\frac{e^{tb}-e^{ta}}{t(b-a)}, t\neq 0$$
2. **Exponential** (λ)
$$M_X(t)=\frac{\lambda}{\lambda-t}, t<\lambda$$
3. **Normal** (μ,σ^2)
$$M_X(t)=\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right)$$
4. **Standard Normal** $\mathcal{N}(0,1)$
$$M_X(t)=\exp\left(\frac{1}{2}t^2\right)$$

7. Joint Probabilty

7.1 Joint Probability Distributions

A Joint Probability Distribution describes the probability behavior of two or more random variables together.
There are two types:

Discrete Joint Distribution

- Defined for discrete **random variables** X and Y.
- Represented using a **Joint PMF (Probability Mass Function):**
$$p_{X,Y}(x,y)=P(X=x,Y=y)$$

Continuous Joint Distribution

- Defined for continuous random variables X and Y.
- Represented using a Joint PDF (Probability Density Function):
$$f_{X,Y}(x,y)$$

7.2 Joint PMF (Discrete Case)

- Let X and Y be discrete random variables.
- The Joint PMF:
$$p_{X,Y}(x,y)=P(X=x,Y=y)$$
- Must satisfy:
$$\sum_x\sum_y p_{X,Y}(x,y)=1$$
- To Find Probabilities:
- For exact values:
$$P(X=a,Y=b)=p_{X,Y}(a,b)$$
- For events over sets:
$$P(X\in A,Y\in B)=\sum_{x\in A}\sum_{y\in B}p_{X,Y}(x,y)$$

7.3 Joint PDF (Continuous Case)

- For continuous variables X and Y, the Joint PDF:
$$f_{X,Y}(x,y)$$
- Must satisfy:
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)dxdy=1$$

To Find Probabilities:

- For exact values:

$P(X=a,Y=b)=0$ (always zero in continuous)

- For intervals or regions:
$$P((X,Y)\in R)=\iint_R f_{X,Y}(x,y)dxdy$$

7.4 Marginal Distributions

- **Discrete**:
$$p_X(x)=\sum_y p_{X,Y}(x,y), p_Y(y)=\sum_x p_{X,Y}(x,y)$$
- **Continuous**:
$$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)dy, f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)dx$$

7.5 Conditional Distributions

- **Discrete**:
$$p_{X|Y}(x\mid y)=\frac{p_{X,Y}(x,y)}{p_Y(y)}\text{ (if }p_Y(y)>0)$$
- **Continuous**:
$$f_{X|Y}(x\mid y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}\text{ (if }f_Y(y)>0)$$

7.6 Independence

- X and Y are independent iff:
$p_{X,Y}(x,y)=p_X(x)\cdot p_Y(y)$ (discrete)
$f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)$ (continuous)

7.7 Expectation and Variance

- $E[X]=\sum_x x\cdot p_X(x), E[Y]=\sum_y y\cdot p_Y(y)$ (or integral form for continuous)

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

7.8 Expectation of functions

- $E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y)$ (discrete)
- $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)dxdy$ (continuous)

7.9 Conditional Expectation

- $E[X \mid Y = y] = \sum_x x \cdot p_{X|Y}(x \mid y)$ (discrete)
- $E[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)dx$ (continuous)

7.10 Conditional Variance

- $\text{Var}(X \mid Y = y) = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2$

7.11 Properties of Conditional Expectation

Let X, Y, Z be random variables, $a, b \in R$, and $g: R \to R$. Assuming all the following expectations exist, we have

(i) $E[a \mid Y] = a$
(ii) $\text{E[aX + bZ} \mid \text{Y]} = \text{aE[X} \mid \text{Y]} + \text{bE[Z} \mid \text{Y]}$
(iii) $E[X \mid Y] \geq 0$ if $X \geq 0$.
(iv) $\text{E[X} \mid \text{Y]} = \text{E[X]}$ if X and Y are independent.
(v) $\text{E[E[X} \mid \text{Y]]} = \text{E[X]}$
(vi) $\text{E[Xg(Y)} \mid \text{Y]} = \text{g(Y)E[X} \mid \text{Y]}$. In particular, $\text{E[g(Y)} \mid \text{Y]} = \text{g(Y)}$.
(vii) $\text{E[X} \mid \text{Y, g(Y)]} = \text{E[X} \mid \text{Y]}$ (viii) $\text{E[E[X} \mid \text{Y, Z]} \mid \text{Y]} = \text{E[X} \mid \text{Y]}$

7.12 Law of Total Expectation

$$E[X] = \sum_y E[X \mid Y = y] \cdot P(Y = y) \quad \text{(discrete)}$$

$$E[X] = \int E[X \mid Y = y] \cdot f_Y(y)dy \quad \text{(continuous)}$$

7.13 Law of Total Variance

$$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y])$$

8. Covariance & Correlation

8.1 Covariance

- Covariance measures the linear relationship between two random variables X and Y.

Definition:

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

Alternate Formula:

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Note:

1. $\text{cov}(X,Y)$ will be positive if large values of X tend to occur with large values of Y, and small values of X tend to occur with small values of Y.

   For example, if X is height and Y is weight of a randomly selected person, we would expect cov(X, Y) to be positive.

2. $\text{cov}(X,Y)$ will be negative if large values of X tend to occur with small values of Y, and small values of X tend to occur with large values of Y.

   For example, if X is age of a randomly selected person, and Y is heart rate, we would expect X and Y to be negatively correlated (older people have slower heart rates).

3. If X and Y are independent, then there is no pattern between large values of X and large values of Y so $\text{cov}(X,Y) = 0$. However, $\text{cov}(X,Y) = 0$ does NOT imply that X and Y are independent, unless X and Y are Normally distributed.

8.2 Properties of Covariance

(Properties of Covariance) Let X, Y, Z be random variables, and let c be a constant.
Then:

1. **Covariance-Variance Relationship**: $\text{Var}[X] = \text{Cov}[X,X]$
2. **Pulling Out Constants**:
   $$\text{Cov}[cX,Y] = c \cdot \text{Cov}[X,Y]$$
   $$\text{Cov}[X,cY] = c \cdot \text{Cov}[X,Y]$$
3. **Distributive Property**:
   $$\text{Cov}[X+Y,Z] = \text{Cov}[X,Z] + \text{Cov}[Y,Z]$$
   $$\text{Cov}[X,Y+Z] = \text{Cov}[X,Y] + \text{Cov}[X,Z]$$
4. **Symmetry**: $\text{Cov}[X,Y] = \text{Cov}[Y,X]$
5. **Constants cannot covary**: $\text{Cov}[X,c] = 0$.

8.3 Correlation

- Correlation Coefficient (denoted ρ or $\operatorname{Corr}(X, Y)$) measures the strength and direction of a linear relationship between two variables.

Formula:

$$\rho_{X,Y} = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Where:

- $\sigma_X = \sqrt{\operatorname{Var}(X)}$
- $\sigma_Y = \sqrt{\operatorname{Var}(Y)}$
- $\rho \in [-1,1]$
- $\rho = 1$: Perfect positive linear correlation
- $\rho = -1$: Perfect negative linear correlation
- $\rho = 0$: No linear correlation

8.4 Properties of Correlation

1. $-1 \leq \rho_{X,Y} \leq 1$
2. $\rho_{X,Y} = \rho_{Y,X}$
3. $\rho_{X,Y} = 0$ implies no linear relationship
4. Correlation is scale-invariant:
   $\operatorname{Corr}(aX + b, Y) = \operatorname{sign}(a) \cdot \operatorname{Corr}(X, Y)$
5. If X and Y are independent, then $\rho = 0$ (but $\rho = 0$ does not imply independence)

9. Statistics

9.1 Markov's Inequality

For any non-negative random variable X and $a > 0$:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Useful when only the mean is known.

9.2 Chebyshev's Inequality

For any random variable X with finite mean μ and variance σ^2, and for any $k > 0$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Works for any distribution, not just normal.

9.3 Central Limit Theorem (CLT) for Sample Mean

The Central Limit Theorem, one of the cornerstone results in probability and statistics, states that if $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables, each with mean μ and variance σ^2, then as the sample size n becomes large, the distribution of the sample mean $\bar{X} = \frac{1}{n}\sum X_i$ tends toward a normal distribution, regardless of the original distribution of X :

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

This approximation improves with larger n; it is widely used for constructing confidence intervals and hypothesis testing, even when the data are not normally distributed.

9.4 Central Limit Theorem for Sum

An extension of the CLT applies not just to the sample mean but also to the sum of i.i.d. random variables; that is, the sum $S_n = X_1 + X_2 + \cdots + X_n$ is also approximately normally distributed when n is large, with mean and variance scaled accordingly:

$$S_n \sim N(n\mu, n\sigma^2)$$

9.5 Point Estimate

- A single value used to estimate a population parameter.
- Computed from sample data.
- Common point estimates:
- $\bar{X}$ estimates μ
- $\hat{p} = \frac{x}{n}$ estimates population proportion p
- No indication of uncertainty - only a single guess.

9.6 Interval Estimate

- Provides a range of plausible values for a population parameter.
- More informative than a point estimate.
- General form:
  Point Estimate $\pm$ Margin of Error
- Margin of error depends on confidence level and sample variability.

9.7 Confidence Interval

A confidence interval is a type of interval estimate that gives a range of values, computed from the sample data, which is likely to contain the true population parameter with a specified level of confidence

(typically 90%, 95%, or 99%); for estimating a population mean when the population standard deviation σ is known and the sample size n is sufficiently large, the confidence interval is given by:

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

Here, $z_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to the desired confidence level; the interval is random because it is derived from sample data, and over repeated samples, a certain proportion (e.g., 95%) of such intervals will contain the true parameter.

9.8 Properties of Confidence Interval

Confidence intervals possess several important properties:

1. A higher confidence level results in a wider interval, reflecting greater certainty at the cost of precision.
2. As the sample size increases, the interval becomes narrower, indicating improved estimate accuracy.
3. The width of the confidence interval depends on the variability of the estimator and the underlying distribution;
4. The confidence level refers to the long-run success rate of the procedure: in repeated sampling, the true parameter would fall within the constructed interval the specified proportion of the time.

9.9 T-Distribution

- Used when:
- Population standard deviation σ is unknown
- Sample size is small (typically $n<30$)
- Definition:

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

where S is the sample standard deviation.

- Degrees of freedom $= n-1$
- Properties:
- Symmetric and bell-shaped like normal
- Heavier tails (more uncertainty due to estimating σ)
- As $n \rightarrow \infty, t$-distribution $\rightarrow$ standard normal $N(0,1)$

10. Hypothesis Testing: Z-Test
10.1 Introduction

- Z-test is a statistical method used to test hypotheses about population means or proportions when the population variance is known and the sample size is large (typically $n \geq 30$).
- Test statistic follows Standard Normal Distribution (Z-distribution).

10.2 One-Tailed Z-Test

- Used when alternative hypothesis $(\mathrm{H_a})$ suggests a directional change.

Left-tailed test:

- $\mathrm{H}_0: \mu=\mu_0$
- $H_3: \mu<\mu_0$
- Reject H_0 if $Z<-Z\_\alpha$

Right-tailed test

- $H_0: \mu=\mu_0$
- $H_8: \mu>\mu_0$
- Reject H_0 if $\mathrm{Z}>\mathrm{Z}\_\alpha$

10.3 Two-Tailed Z-Test

- Used when $H_a: \mu \neq \mu_0$, i.e., deviation in either direction.
- Reject H_0 if $|\mathrm{Z}|>\mathrm{Z}\_(\alpha/2)$

10.4 Z-Test for Proportion

- Testing population proportion:
- Let $\hat{\mathrm{p}}=\mathrm{x}/\mathrm{n}$ (sample proportion), p (population proportion)
- Test statistic:

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

- Used for both one-tailed and two-tailed depending on the alternative hypothesis.

10.5 Z-Test for Mean (Two Populations)

- Used when comparing means of two independent populations.
- Let sample means be $\dot{x}_1, \dot{x}_2$; population variances ${\sigma_1}^2, {\sigma_2}^2$ known; sample sizes n_1, n_2.
- Test statistic:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- One-tailed: $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$
- Two-tailed: $\mu_1 \neq \mu_2$

10.6 Z-Test for Proportion (Two Populations)

- Testing equality of proportions $\mathbf{p}_1$ and $\mathbf{p}_2$.
- Combined proportion:

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, q = 1 - p$$

- Test statistic:

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

11. Hypothesis Testing: T-Test

11.1 Introduction

- A T-Test is used to test hypotheses when the population standard deviation is unknown and the sample size is small (typically $\mathrm{n} < 30$).
- The test statistic follows the Student's t-distribution with appropriate degrees of freedom (df).
- Types of t -tests:
- One-sample t-test
- Two-sample t-test (Independent samples)
- Paired t-test

11.2 One-Sample T-Test

- Used to compare the sample mean $\bar{x}$ to a known value μ (population mean)
- Null Hypothesis (H_0): $\mu = \mu_0$
- Test Statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where:

- $\bar{x}$: sample mean
- s : sample standard deviation
- n : sample size
- $\mathrm{df} = \mathrm{n} - 1$

11.3 Two-Sample T-Test (Independent Samples)

- Used to compare the means of two independent populations.
- Assumes equal variances (can be extended to unequal variances too).
- Null Hypothesis (H_0): $\mu_1 = \mu_2$
- Test Statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where:

- $\bar{x}_1, \bar{x}_2$: sample means
- n_1, n_2 : sample sizes
- s_p : pooled standard deviation
- Pooled Variance (s^2) :

$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

- $\mathrm{df} = \mathrm{n}_1 + \mathrm{n}_2 - 2$

11.4 Paired T-Test (Dependent Samples)

- Used when the samples are paired or related (e.g., before-after, same subject in different conditions).
- Focuses on the difference in paired observations.
- Let $d_i = x_i - y_i$, and $\bar{d} = \frac{\sum d_i}{n}$
- Null Hypothesis (H_0): $\mu\_\mathrm{d} = 0$ (no difference)
- Test Statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- Standard deviation of differences:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n - 1}}$$

- $df = n - 1$

12. Chi-Square Test

The Chi-Square (X^2) Test is a non-parametric statistical test used to examine the association between categorical variables or to test the goodness of fit of observed data with expected data.

12.1 Types of Chi-Square Tests

1. Chi-Square Goodness of Fit Test
2. Chi-Square Test for Independence (Association)

Assumptions of Chi-Square Test

- Observations must be independent.
- Categories must be mutually exclusive.
- Data must be in frequency form (not percentages).
- Expected frequency in each cell should be ≥ 5.

12.2 Chi-Square Goodness of Fit Test

- **Purpose**: To test if the observed categorical data match an expected distribution.
- **Hypotheses**:
- H_0 : The data follow the expected distribution.
- H_1 : The data do not follow the expected distribution.
- **Test Statistic**:

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

where:

- O_i : Observed frequency for category i
- E_i : Expected frequency for category i
- Degrees of Freedom:

$$df = k - 1$$

  where k is the number of categories.

- **Decision Rule:**
  - Reject H_0 if $\chi^2_{\text{calculated}} > \chi^2_{\text{critical}}$ (from the χ^2 table at given α and df).
  - Fail to reject H_0 if $\chi^2_{\text{calculated}} \leq \chi^2_{\text{critical}}$.

STAR MENTOR CS/DA

KHALEEL SIR
ALGORITHM & OS
29 YEARS OF TEACHING EXPERIENCE

SATISH SIR
DISCRETE MATHEMATICS
BE in IT from MUMBAI UNIVERSITY

VIJAY SIR
DBMS & COA
M. TECH FROM NIT
14+ YEARS EXPERIENCE

SAKSHI MA'AM
ENGINEERING MATHEMATICS
IIT ROORKEE ALUMNUS

AVINASH SIR
APTITUDE
10+ YEARS OF TEACHING EXPERIENCE

CHANDAN SIR
DIGITAL LOGIC
GATE AIR 23 & 26 / EX-ISRO

MALLESHAM SIR
M.TECH FROM IIT BOMBAY
AIR – 114, 119, 210 in GATE
(CRACKED GATE 8 TIMES)
14+ YEARS EXPERIENCE

PARTH SIR
DA
IIIT BANGALORE ALUMNUS
FORMER ASSISTANT PROFESSOR

SHAILENDER SIR
C PROGRAMMING & DATA STRUCTURE
M.TECH in Computer Science
15+ YEARS EXPERIENCE

AJAY SIR
PH.D. IN COMPUTER SCIENCE
12+ YEARS EXPERIENCE