



Probability

Prof. Pranay Kumar Saha

March 1, 2025





Probability Theory

Probability Theory deals with random phenomena which under certain random experiment yields outcome that have some pattern among themselves.



Radom Variable

- A random variable is a fundamental concept in probability theory, serving as a bridge between abstract random experiments and numerical analysis.
- It is defined as a function that maps each possible outcome of a random experiment to a real number.
- This numerical representation allows for the application of mathematical tools to study:
 - Probabilities
 - Expected values
 - Distributions
- These concepts are essential in statistics and decision-making processes.



Example

- For instance, consider a simple experiment of rolling a six-sided die.
- The outcome could be any number from 1 to 6, each with an equal probability of $\frac{1}{6}$.
- By defining a random variable X as the number shown on the die, we transform the abstract outcome into a numerical value.
- This transformation enables us to calculate:
 - Probabilities
 - Expected values
 - Other statistical measures



Sample Space and Event

- **Sample Space:** Collection of all possible outcome of an Random Event (RE) is called the Sample Space S or Ω .
- **Event:** Any subset of a Sample Space
- **Union of Events**
 - $A \cup B$ represents the occurrence of at least one of A and B .
 - In general, $A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n$ represents the occurrence of at least one of these events.
- **Intersection of Events**
 - $A \cap B$ represents the simultaneous occurrence of both A and B .
 - In general, $A_1 \cap A_2 \cap \dots \cap A_n$ represents the simultaneous occurrence of all these events.
- **Complement of Events:** The complement of an event A , denoted as A^c , represents the non-occurrence of A .



Classical Definition of Probability

The classical definition of probability states that if an experiment has n equally likely outcomes and an event A consists of m favorable outcomes, then the probability of event A occurring is given by:

$$P(A) = \frac{m}{n}$$

where:

- m is the number of favorable outcomes for event A .
- n is the total number of equally likely outcomes in the sample space.



Empirical Definition of Probability

The empirical definition of probability, also known as the relative frequency approach, states that if an experiment is repeated n times and an event A occurs m times, then the probability of A is given by:

$$P(A) = \frac{m}{n}$$

where n is the total number of trials and m is the number of times A occurs. As n increases, $P(A)$ approaches the true probability of the event.



Statistical regularity Definition of Probability

The principle of statistical regularity states that if a large number of trials of a random experiment are conducted under identical conditions, then the relative frequency of occurrence of an event stabilizes around a fixed value, which is considered the probability of that event as $n \rightarrow \infty$.

Mathematically, this can be expressed as:

$$P(A) = \lim_{n \rightarrow \infty} \frac{m}{n}$$

where m is the number of times event A occurs in n trials.



Axiomatic Definition of Probability

It introduced by Kolmogorov, is based on three fundamental axioms:

- **Non-negativity:** For any event A , its probability satisfies:

$$P(A) \geq 0$$

- **Normalization:** The probability of the entire sample space S is 1:

$$P(S) = 1$$

- **Additivity:** For any two mutually exclusive events A and B , the probability of their union is:

$$P(A \cup B) = P(A) + P(B)$$

More generally, for a countable sequence of mutually exclusive events $A_1, A_2, A_3, \dots$:

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



Conditional Probability

- Definition: The probability of an event occurring given that another event has already occurred.
- Notation: $P(A|B)$, read as "the probability of A given B"



Formula for Conditional Probability

The formula for conditional probability is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0$$

This formula calculates the probability of both events A and B occurring, divided by the probability of event B occurring.



Example: Drawing Cards

Consider a standard deck of 52 playing cards. We draw two cards without replacement. Let:

- Event B: The first card drawn is a heart.
- Event A: The second card drawn is a heart.

We want to find $P(A|B)$, the probability that the second card is a heart given that the first card was a heart. Since the first card was a heart, there are 12 hearts left in the remaining 51 cards. Therefore,

$$P(A|B) = \frac{12}{51} \approx 0.235$$



Alternate Method

Alternatively, using the formula: First, find $P(A \cap B)$, the probability that both cards are hearts.

$$P(A \cap B) = P(\text{first is heart}) \times P(\text{second is heart} | \text{first is heart}) = \frac{13}{52} \times \frac{12}{51}$$

Then, $P(B) = P(\text{first is heart}) = \frac{13}{52}$ So,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{13}{52} \times \frac{12}{51}}{\frac{13}{52}} = \frac{12}{51}$$



Law of Total Probability

- A fundamental rule in probability theory.
- Used to compute the total probability of an event by considering all possible scenarios.



Partition of Sample Space

- The sample space S is divided into mutually exclusive and exhaustive events $B_1, B_2, \dots, B_n$.
- Mutually exclusive: $B_i \cap B_j = \emptyset$ for $i \neq j$
- Exhaustive: $B_1 \cup B_2 \cup \dots \cup B_n = S$



Formula

For any event A , the Law of Total Probability states:

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

This formula sums the probabilities of A occurring given each B_i , weighted by the probability of each B_i .



Example: Machines and Defective Items

A factory has three machines producing items:

- Machine A: Produces 30% of items, 5% defective.
- Machine B: Produces 40% of items, 3% defective.
- Machine C: Produces 30% of items, 4% defective.

We want to find the probability that a randomly selected item is defective, $P(D)$.



Applying the Law

Let B_1, B_2, B_3 be the events that an item is produced by Machine A, B, or C, respectively. Then,

$$P(D) = P(D|B_1) \cdot P(B_1) + P(D|B_2) \cdot P(B_2) + P(D|B_3) \cdot P(B_3)$$

Plugging in the values:

$$P(D) = (0.05)(0.30) + (0.03)(0.40) + (0.04)(0.30)$$



Calculation

Calculate each term:

- Machine A: $(0.05)(0.30) = 0.015$
- Machine B: $(0.03)(0.40) = 0.012$
- Machine C: $(0.04)(0.30) = 0.012$

Sum them up:

$$P(D) = 0.015 + 0.012 + 0.012 = 0.039$$

Thus, the probability of selecting a defective item is 3.9%.



Conclusion

The Law of Total Probability allows us to:

- Break down complex probability problems.
- Compute total probabilities by considering all possible scenarios.
- Apply to real-world situations like quality control.

It's a stepping stone to Bayes' Theorem and other advanced probability concepts.



Bayes' Theorem



Bayes' Theorem

- A fundamental theorem in probability theory.
- Used to update the probability of an event based on new evidence.
- Essential for Bayesian inference and decision-making under uncertainty.



Formula of Bayes' Theorem

For events A and B , Bayes' Theorem states:

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

where:

- $P(A|B)$ is the posterior probability: the probability of A given B .
- $P(B|A)$ is the likelihood: the probability of B given A .
- $P(A)$ is the prior probability of A .
- $P(B)$ is the marginal probability of B .



Example: Medical Test

Consider a medical test for a disease:

- Prevalence of the disease: 1% ($P(D) = 0.01$)
- Test sensitivity (true positive rate): 99% ($P(T+|D) = 0.99$)
- Test specificity (true negative rate): 95% ($P(T-|D^c) = 0.95$)

We want to find the probability that a person has the disease given that they tested positive, $P(D|T+)$.



Applying Bayes' Theorem

To find $P(D|T+)$, use Bayes' Theorem:

$$P(D|T+) = \frac{P(T+|D) \cdot P(D)}{P(T+)}$$

First, we need to calculate $P(T+)$, the total probability of testing positive. Using the Law of Total Probability:

$$P(T+) = P(T+|D) \cdot P(D) + P(T+|D^c) \cdot P(D^c)$$

Note that $P(T+|D^c) = 1 - P(T-|D^c) = 1 - 0.95 = 0.05$



Calculation

Calculate $P(T+)$:

$$P(T+) = (0.99)(0.01) + (0.05)(0.99) = 0.0099 + 0.0495 = 0.0594$$

Now, apply Bayes' Theorem:

$$P(D|T+) = \frac{(0.99)(0.01)}{0.0594} = \frac{0.0099}{0.0594} \approx 0.1667$$

Thus, the probability that a person has the disease given a positive test result is approximately 16.67%.



Independent Event



Independent Events

- A key concept in probability theory.
- Events are independent if the occurrence of one does not affect the probability of the other.



Definition and Formula

- **Definition:** Two events A and B are independent if:

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

- This means the probability of both events occurring is the product of their individual probabilities.
- If events are not independent, they are dependent.



Example: Coin and Die

Consider the following events:

- Event A : Flipping a head on a fair coin.

$$P(A) = \frac{1}{2}$$

- Event B : Rolling a 3 on a fair six-sided die.

$$P(B) = \frac{1}{6}$$

Since A and B are independent, the probability that both occur is:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$



Conclusion

- Independent events do not influence each other's probabilities.
- For independent events A and B , $P(A \cap B) = P(A) \cdot P(B)$.
- Understanding independence is essential in probability and statistics.



Cumulative Distribution Function (CDF)



Definition of CDF

The Cumulative Distribution Function (CDF) of a random variable X is defined as:

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

It gives the probability that X takes a value less than or equal to x .



Properties of CDF

- Non-decreasing: If $x_1 < x_2$, then $F(x_1) \leq F(x_2)$
- Limits: $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$
- Right-continuous: For any x , $\lim_{h \rightarrow 0^+} F(x + h) = F(x)$



Example: Discrete Random Variable

Consider a fair six-sided die. The random variable X can take values 1, 2, 3, 4, 5, 6, each with probability $\frac{1}{6}$. The CDF $F(x)$ is:

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{6} & \text{if } 1 \leq x < 2 \\ \frac{2}{6} & \text{if } 2 \leq x < 3 \\ \frac{3}{6} & \text{if } 3 \leq x < 4 \\ \frac{4}{6} & \text{if } 4 \leq x < 5 \\ \frac{5}{6} & \text{if } 5 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$



Example: Continuous Random Variable

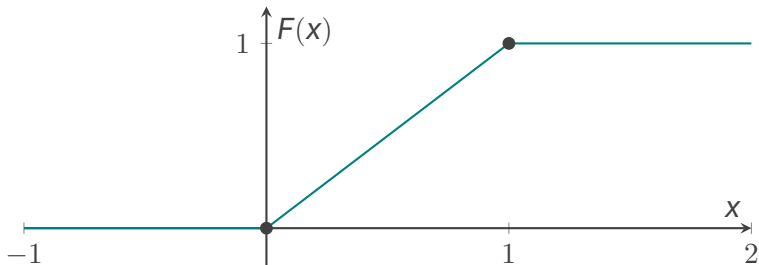
Consider a uniform random variable X on $[0, 1]$. The CDF $F(x)$ is:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Note that $F(x)$ is continuous everywhere.



Graph of CDF (Uniform on $[0,1]$)



This is the CDF of a uniform random variable on $[0, 1]$.



Example: CDF

Consider the function:

$$G(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x < 1 \\ 0.5 & \text{if } x \geq 1 \end{cases}$$

This is **not** a valid CDF because:

- It is not non-decreasing: $G(1^-) = 1$ (limit from left), but $G(1) = 0.5$, so it decreases.
- $\lim_{x \rightarrow \infty} G(x) = 0.5 < 1$, violating the property that a CDF must approach 1.



Types of Random Variable



Types of Random Variables

Random variables are classified into three main types:

- **Discrete:** Take countable values (e.g., integers).
- **Continuous:** Take values in a continuous range (e.g., real numbers).
- **Mixed:** Have both discrete and continuous components.

Each type has a probability distribution and a Cumulative Distribution Function (CDF).



Discrete Random Variable: Die Roll

Example: Let X be the outcome of rolling a fair six-sided die. Possible values are 1, 2, 3, 4, 5, 6, each equally likely.

PMF:

$$P(X = x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

CDF:

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{\lfloor x \rfloor}{6} & \text{if } 1 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$

Observation: The CDF is a step function, jumping by $\frac{1}{6}$ at each integer from 1 to 6.



Continuous Random Variable: Bus Arrival Time

Example: Let X be the time (in minutes) until a bus arrives, equally likely between 0 and 10 minutes.

PDF:

$$f(x) = \begin{cases} \frac{1}{10} & \text{if } 0 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

CDF:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{10} & \text{if } 0 \leq x \leq 10 \\ 1 & \text{if } x > 10 \end{cases}$$

Observation: The CDF increases linearly from 0 to 1 over $[0, 10]$, reflecting equal likelihood.



Mixed Random Variable: Daily Rainfall

Example: Let X be the rainfall amount (in inches) in a day:

- $P(X = 0) = 0.3$ (no rain).
- With probability 0.7, X is equally likely between 0 and 5 inches (rain).

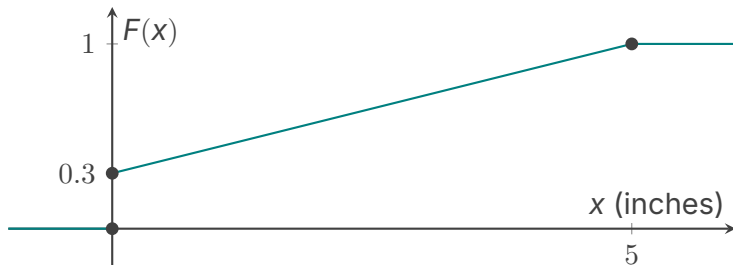
CDF:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.3 & \text{if } x = 0 \\ 0.3 + 0.7 \cdot \frac{x}{5} & \text{if } 0 < x < 5 \\ 1 & \text{if } x \geq 5 \end{cases}$$

Observation: The CDF has a jump of 0.3 at $x = 0$ (discrete part) and increases linearly for $0 < x < 5$ (continuous part).



Graph of Mixed Random Variable CDF (Rainfall)



This shows the CDF with a jump at $x = 0$ and a linear increase from 0 to 5.



Expected Value

Definition

The expected value (mean) of a random variable is the weighted average of all possible values, where the weights are the probabilities.

Discrete Case

$$E[X] = \sum_i x_i \cdot P(X = x_i)$$

Example: For a fair die,

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \dots + 6\left(\frac{1}{6}\right) = 3.5$$

Continuous Case

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

Where $f(x)$ is the PDF of X



Properties of Expected Value

1. **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$ for constants a and b
2. **Constant:** $E[c] = c$ for any constant c
3. **Independence:** If X and Y are independent, $E[X \cdot Y] = E[X] \cdot E[Y]$

Applications

- Financial risk assessment
- Decision theory
- Statistical inference
- Machine learning



Variance and Standard Deviation

Definition

The variance of a random variable measures the spread or dispersion around its expected value.

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

- Standard deviation: $\sigma_X = \sqrt{\text{Var}(X)}$
- Units: Variance has units squared; standard deviation has same units as the original variable
- Always non-negative: $\text{Var}(X) \geq 0$
- $\text{Var}(X) = 0$ if and only if X is a constant



Properties of Variance and Standard Deviation

Key Properties

- $Var(aX + b) = a^2 Var(X)$ for constants a and b
- If X and Y are independent: $Var(X + Y) = Var(X) + Var(Y)$
- In general: $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$



Discrete Probability Distribution



Bernoulli Distribution

Definition

Models a single trial with two possible outcomes: success (1) with probability p or failure (0) with probability $1 - p$.

$$P(X = 1) = p \quad (1)$$

$$P(X = 0) = 1 - p \quad (2)$$

- Mean: $E[X] = p$
- Variance: $Var(X) = p(1 - p)$

Example: Fair Coin Flip

$p = 0.5$, so $E[X] = 0.5$ and $Var(X) = 0.25$



Binomial Distribution

Definition

Models the number of successes in n independent Bernoulli trials, each with probability p .

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n \text{ where } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Mean: $E[X] = np$
- Variance: $Var(X) = np(1 - p)$

Example: 10 Coin Flips

$n = 10, p = 0.5$, so $E[X] = 5$ and $Var(X) = 2.5$



Binomial Distribution: Properties and Applications

Key Properties

- Sum of Bernoulli RVs:
 $X = \sum_{i=1}^n X_i$
- Reproductive property:
 $Bin(n_1, p) + Bin(n_2, p) = Bin(n_1 + n_2, p)$
- Normal approximation when n is large: $X \approx N(np, np(1 - p))$
- Poisson approximation when n is large and p is small

Applications

- Quality control: Defective items in batches
- Medicine: Success/failure of treatments
- Finance: Up/down price movements
- Survey sampling: Properties in finite populations



Geometric Distribution

Definition

Models the number of Bernoulli trials needed to get the first success.

$$P(X = k) = (1 - p)^{k-1}p \quad \text{for } k = 1, 2, 3, \dots$$

- Mean: $E[X] = \frac{1}{p}$
- Variance: $Var(X) = \frac{1-p}{p^2}$

Example: Number of rolls until a 6 appears on a die

$$p = \frac{1}{6}, \text{ so } E[X] = 6 \text{ and } Var(X) = 30$$



Poisson Distribution

Definition

Models the number of events occurring in a fixed interval when events happen at a constant average rate λ .

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, 2, \dots$$

- Mean: $E[X] = \lambda$
- Variance: $Var(X) = \lambda$



Poisson Distribution: Properties and Applications

Key Properties

- Limit of Binomial:
 $Bin(n, p) \approx Pois(np)$ when n is large, p is small
- For Poisson process: Number of events in time t is $Pois(\lambda t)$
- Normal approximation when λ is large: $X \approx N(\lambda, \lambda)$
- Independence of counts in disjoint intervals

Applications

- Call center arrivals
- Radioactive decay events
- Network traffic modeling
- Insurance claims
- Rare events: Accidents, natural disasters
- Mutations in DNA sequences



Continuous Probability Distributions



Normal Distribution

Standard Normal Distribution

When $\mu = 0$ and $\sigma = 1$, we get the standard normal distribution $Z \sim N(0, 1)$

- Any normal RV can be standardized: $Z = \frac{X - \mu}{\sigma}$
- CDF: $\Phi(z) = P(Z \leq z)$ (no closed form, values tabulated)



Normal Distribution: Properties and Applications

Key Properties

- 68-95-99.7 rule:
 - 68% of data within $\mu \pm \sigma$
 - 95% of data within $\mu \pm 2\sigma$
 - 99.7% of data within $\mu \pm 3\sigma$
- Reproductive property: If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- Linear transformations: If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$



Normal Distribution: Applications

Scientific Applications

- Measurement errors in physical sciences
- Heights, weights in populations
- IQ scores and standardized tests
- Approximation for binomial when n is large
- Foundation for parametric statistical tests

Financial Applications

- Stock price movements (log-normal)
- Black-Scholes option pricing model
- Value at Risk (VaR) calculations
- Portfolio theory and optimization
- Credit scoring models



Central Limit Theorem

Central Limit Theorem Connection

The normal distribution's importance is largely due to the CLT, which states that sums of independent random variables tend toward a normal distribution regardless of their original distributions.



Exponential Distribution

Definition

Models the time between events in a Poisson process with rate parameter λ .

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

- Mean: $E[X] = \frac{1}{\lambda}$
- Variance: $Var(X) = \frac{1}{\lambda^2}$
- Median: $\frac{\ln(2)}{\lambda}$
- Mode: 0
- CDF: $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$



Exponential Distribution: Properties and Applications

Key Properties

- Memoryless property: $P(X > s + t | X > s) = P(X > t)$
- Relationship to Poisson process: If events occur according to a Poisson process with rate λ , then the time between consecutive events follows $\text{Exp}(\lambda)$
- Minimum of independent exponentials:
 $\min(X_1, X_2, \dots, X_n) \sim \text{Exp}(\lambda_1 + \lambda_2 + \dots + \lambda_n)$
- Relationship to other distributions:
 - Special case of Gamma distribution with shape parameter $\alpha = 1$
 - $X^2/2 \sim \text{Exp}(\lambda/2)$ if $X \sim \chi^2(2)$



Uniform Distribution

Definition

Every value in interval $[a, b]$ has equal probability.

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

Applications

- Mean: $E[X] = \frac{a+b}{2}$
- Variance: $Var(X) = \frac{(b-a)^2}{12}$
- Median: $\frac{a+b}{2}$
- Random number generation
- Rounding errors in computing
- Prior distributions in Bayesian analysis



Probability

Thank You for Listening!