16-August-2025 Data Science Journey Yasir Baig

{ Probability }

Terminolgeies In Probability

- Five most important terminologies in probability
  - Random experiment
  - Trial
  - Outcome
  - Sample Space
  - Event

Random Experiment

A random experiment is an action or process that leads to one of several possible outcomes, and it is not possible to predict the exact outcome in advance

Conditions:

- It has more than one possible outcome
- The exact outcome cann't be predicted inadvance.

Examples

- Rolling a die → outcomes can be 1, 2, 3, 4, 5, 6
- Tossing a coin → outcomes canbe Head or tail
- Titanic dataset example → whether a passanger survived or not.

Trial

A trial refers to a single execution of a random experiment. Each trial produces exactly one outcome

Examples
- Rolling a die once is a trial.
- Tossing a coin once is a trial.
- Checking a survival of one Titanic passenger is trial

Outcome

An outcome is a single possible result of trial

Example:
- Rolling a die → outcome can be "4"
- Tossing a coin → outcome can be "Head"
- Titanic dataset → outcome can be "Passenger survived"

Sample Space

The sample space of a random experiment is the set of all possible outcomes that can occur. Usually, one random experiment hase one sample Space.

Examples:
- Rolling a die → {1, 2, 3, 4, 5, 6}
- Tossing a coin twice → {HH, HT, TH, TT}
- Titanic dataset → {Survived, Not Survived}

Event

An event is a specific subset of outcomes from a sample space. An event can include one or more outcomes.

Example:

Rolling a die:

- Event "Getting an even no" → {2, 4, 6}

Tossing a coin twice:

- Event "Getting exactly one 'Head → {HT, TH}

Titanic dataset:

- Event "Selecting a female passenger who Survived" → Subset of sample space filtered by gender and survival.

Types of Events in Probability

There are 8 type of events in propability.

Simple Event

Also known as elementary event. A simple event is an event that consists of exactly one outcome.

Example:

- When rolling a fair six-sided die, getting a 3 is simple event.

Compound Event

A compound event consists of two or more simple events.

Example:

- When rolling a die, the event "rolling an odd number" is a compound event because it consists of three simple events: rolling a 1, rolling a 3, rolling a 5.

Independent Events

Two events are independent if the occurance of one event does not affect the probability of the occurance of the other event.

Example:

- If you flip a coin and roll a die, the outcomes of the coin flip does not affect the outcome of the die roll.

Dependent Events

Events are dependent if the occurance of one event does affect the probability of the occurance of the other event.

Example:

- If you draw two cards from a deck without replacement, the outcome of the first draw affects the outcomes of the second draw because there are fewer cards left in the deck.

Mutually Exclusive Events

Two events are mutually exclusive (disjoint) if they cannot both occur at the same time.

Example:

- When rolling a die, the events "rolling a 2" and "roll a 4" are mutually exclusive because a single roll cannt result in both a 2 and a 4.

Exhaustive Events

A set of events is exhaustive if at least one of the events must occur when the experiment is performed.

Example:

- When rolling a die, the events " roll an even number " and " roll an old number " are exhaustive because one of them must occur on any roll.

Impossible Event

An impossible event is an event that cannot occur at all, its probability is always zero.

Example:

- Rolling a 7 on a standard six-sided die is an impossible event.

Certain Event

A certain event is an event that will always occur, its probability is always one

Example:

- When rolling a six-sided die, the event "getting a number between 1 and 6" is a certain event

What is Probability?

In simplest terms, the probability is the measure of the linklihood that a particular event will occur. It is a fundamental concept in statistics and is used to make predictions and informed decisions in a wide range of discipline including science, engineering, medicine and social science. Probability is usually expressed as a number b/w 0 & 1 inclusive.

- A probability of 0 means that an event will not happen.
- A probability of 1 means that an event will certainly happened.
- A probability of 0.5 means that an event will happen half the time (or that is likely to happen as not to happen).

Types of Probability

In probability theory, there are two main types of probability

1- Empirical Probability (Experimental Probability)
2- Theoretical Probability (Classical Probability)

Empirical Probability

Empirical probability (also called Experimental Probability) is based on actual observations or experiments.

Formula:

Empirical Probability of Event A = (No. times Event A occurs) / (Total no of trials)

Key Idea:

- Instead of assuming equal likelihood, we rely on real data.
- The more trials we perform, the closer empirical probability gets to the theoretical probability

Example 1: Coin Toss

Suppose we toss a fair coin 100 times.

- Heads appeared 55 times
- Tails appeared 45 times

55 / 100 = 0.55 (or 55%)

Example 2: Marbles in Bag

Bag contains 50 marble in total:

- 20 red, 12 blue, 15 green.

We draw marbles 200 times (with replacement), and record.

- Red marble = 80 times
- Blue marble = 70 times
- Green marble = 50 times

Empirical probability of getting a red marble = 80 / 200 = 0.40 (or 40%)

Theoretical Probability

Theoretical Probability is used when all outcomes in a sample space are equally likely to occur.

Formula:

Theoretical Probability of Events A = (No. of favourable outcomes) / (Total no of outcomes)

Key Idea:

- This is based on mathematical reasoning, not experiment.
- Works well in controlled situations like dice, coin or cards.

Example: A: Tossing a Coin 3 times

- Sample space size = 2^3 = 8 outcomes
- All possibles outcomes: {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}
- Event: Getting exactly 2 heads.
- Favourable outcomes: {HHT, HTH, THH} = 3
- Probability = 3/8 = 0.375 (or 37.5%)

Example B: Rolling 2 dice

- Total possible outcomes = 6 * 6 = 36
- Event: Getting a sum of 7
- Favourable outcomes: {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)} = 6
- Probability = 6/36 = 1/6 ≈ 0.167 (or 16.7%)

Random Variable

What is random variable?

A random variable is NOT like a normal variable that stores values. Instead it is a FUNCTION that maps outcomes of a random process (sample space) to real numbers. This is one of the most confusing concepts at first, but it gets clear with example

In Probability theory:

- A random variable takes an outcome from the

Sample space (input).

- And assigns a real number to it. (output)

Input: An outcomes from a random process (e.g, rolling a die → {1, 2, 3, 4, 5, 6}

Output: A real number assigned to each outcome. (eg, 1 → 1, 2 → 2 etc)

The way outcomes are mapped to numbers depends on how we define the random variable. This choice depends on what aspects of the random process we are studying.

Example: Rolling a fair die

Sample Space(S) = {1, 2, 3, 4, 5, 6}

Random Variable X = "the number that appears on die"

- outcome: Die shows 4 → X = 4
- outcome: Die shows 6 → X = 6

Here, X is a random variable because its value depends on chance.

Types of Random Variables

1) Discrete Random Variable:
   - Take countable value (finite or countably infinite)
   - Example: Rolling a die → {1, 2, 3, 4, 5, 6}
   - Example: Number of students present in a class

2) Contineous Random Variable:
   - Takes value from an interval of real numbers.
   - Example: ~~P~~ Height of Students
   - Example: Time taken to run 100m

What is probability Distribution of RV?

A probability distribution is the set of all possible outcomes of a random variable along with their corresponding probability values

In Simple Words:

- Random Variable → A function that assigns a number to each outcome
- Probability Distribution → Tells us how likely each outcomes is.

Example 1: Tossing a Coin

Random Variable (X):

$X = 1$ if the head comes

$X = 0$ if the tail comes

Sample Space: {Head, Tail}

Random Variable Values: $\{0, 1\}$

Probability Distribution:

$P(X=1) = 0.5$

$P(X=0) = 0.5$

Meaning → The probability of getting Head or Tail is equal (50% – 50%)

Example 2: Rolling a Single Die

Sample Space: $\{1, 2, 3, 4, 5, 6\}$

Random Variable (X): The number that appears on die

Random Variable Values: $\{1, 2, 3, 4, 5, 6\}$

Probability Distribution:

$(P)(X = K) = 1/6$ for each K in $\{1, 2, 3, 4, 5, 6\}$

Meaning → Each face of the die has an equal chances of $1/6$.

Example: 3 Rolling Two Dice

Sample Space: {(1, 1), (1, 2) ..... (6, 6)} Total 36 outcomes

Random Variable (X): The sum of no. on two dice.

Possible Values of X: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

Probability Distribution:

$P(x = 2) = 1/36 \rightarrow$ only (1, 1)

$P(X = 3) = 2/36 \rightarrow$ (1, 2), (2, 1)

$P(X = 4) = 3/36 \rightarrow$ (1, 3), (2, 2), (3, 1)

$P(X = 5) = 4/36 \rightarrow$ (1, 4), (2, 3), (3, 2), (4, 1)

$P(X = 6) = 5/36 \rightarrow$ (1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

$P(X = 7) = 6/36 \rightarrow$ (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)

$P(X = 8) = 5/36 \rightarrow$ (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)

$P(X = 9) = 4/36 \rightarrow$ (3, 6), (4, 5), (5, 4), (6, 3)

$P(X = 10) = 3/36 \rightarrow$ (4, 6), (5, 5), (6, 4)

$P(X = 11) = 2/36 \rightarrow$ (5, 6), (6, 5)

$P(X = 12) = 1/36 \rightarrow$ (6, 6)

Meaning → The probability of getting sums is not uniform

Example → Getting a 7 is the most likely (6/36, 1/6) while getting 2 or 12 is the least likely (1/36)

Mean of Random Variable

The mean of a random variable, often called the Expected Value (E(X)]), represents the long-run avg. outcome of a random process repeated many times.

In Simple words:

- If you repeat the experiment (like rolling a die) thousands of times, the average result you get will approach the expected value.

Mathematically:

$$E[X] = \Sigma [x * P(x)]$$

Where:

- x = Possible outcome
- $P(X)$ = Probability of that outcomes

Example 1: Rolling a fair Die

possible outcomes (x): {1, 2, 3, 4, 5, 6}

probability of each outcome: 1/6 (since die is fair)

$$E[X] = (1 * 1/6) + (2 * 1/6) + (3 * 1/6) + (4 * 1/6) + (5 * 1/6) + (6 * 1/6)$$
$$= (1 + 2 + 3 + 4 + 5 + 6) / 6$$
$$= 21/6$$
$$= 3.5$$

So, the expected value (mean) of rolling a die = 3.5

Notice that 3.5 is not a possible outcome of the die, but it is the *average* over a very large number of rolls.

```
Import random
import numpy as np
outcomes = []
for i in range (10000):
    outcomes.append (random.randint(1, 6)
print (" Simulated Mean:", np.mean (outcomes))
print(" Theoretical Mean:", 3.5)
```

Variance of Random Variable

The variance of a random variable is a measure of how much the possible values of the variable deffer (spread out) from the mean (expected Value)

In Simple Terms:

- The mean tell us the "average outcome"
- The variance tells us "how far away" the outcomes are from that average, on average

Formula (Discrete Random Variable):

$Var(X) = \Sigma [(x_i - \mu)^2 * P(x_i)]$

where:

x_i = possible values
μ = mean(expected value)
$P(x_i)$ = probability of each outcome

Example:

- Rolling a single die.
- Possible outcomes: $\{1, 2, 3, 4, 5, 6\}$
- Each with equal probability = 1/6
- Mean $\mu = (1+2+3+4+5+6)/6 = 3.5$
- Variance $= \Sigma [(x - 3.2)^2 * (1/6)]$ for x in $\{1, \ldots 6\}$

Python - Code

```
import numpy as np
outcomes = np.array([1, 2, 3, 4, 5, 6])
probabilities = np.array([1/6] * 6)
mean = np.sum(outcomes * probabilities)
variance = np.sum(((outcomes - mean) ** 2) * probabilities)
print("Mean (Expected Value):", mean)
print("Variance:", variance)
```

Venn Diagrams

A venn diagram is a visual way to represent sets & their relationships. In probability, sets represents events &

the diagram helps us understand how events overlap (intersections), combine (unions), or remain separate.

Set Base Theory

1. Universal Set (U):

The set of all possible outcomes of an experiment.

Example: Tossing a die → U = {1, 2, 3, 4, 5, 6}

2. Subset ($A \subseteq U$):

A set that contains some (or all) outcomes from the universal set.

Example: A = {2, 4, 6} (event of rolling an even no.)

3. Complement (A'):

The set of outcomes that are in U but not in A.

Example: A' = {1, 3, 5} (event of rolling an odd no.)

4. Union ($A \cup B$):

All outcomes in either A or B (or both).

Example: A = {2, 4, 6}, B = {1, 2} → $A \cup B$ = {1, 2, 4, 6}

5. Intersection ($A \cap B$):

Outcomes that are in both A & B.

Example: A = {2, 4, 6}, B = {1, 2} → $A \cap B$ = {2}

6. Probability of Universal Set:

P(U) = 1 (since the universal set contains all possible outcome)

Venn Diagram Example (conceptual)

Tossing a die → U = {1, 2, 3, 4, 5, 6}

Let $A = \{2, 4, 6\}$ → Event: rolling even number
Let $B = \{1, 2, 3\}$ → Event: rolling ≤ 3

Union: $A \cup B = \{1, 2, 3, 4, 6\}$
Intersection: $A \cap B = \{2\}$
Complement of A: $A' = \{1, 3, 5\}$

Venn Diagram visually show:
- Circle A overlaps Circle B at $\{2\}$
- All outcomes remain inside universal set U.

Python - Code

```
U = {1, 2, 3, 4, 5, 6}
A = {2, 4, 6}
B = {1, 2, 3}

print(U)
print(A)
print(B)
print(A.union(B))
print(A.intersection(B))
print(U-A,)
```

Contingency Table

A contingency table (also called a cross-tabulation) or (cross tab) is a type of table in probability / Statistics that shows the frequency distribution of variables. It helps us undestand the relationship b/w two categorical variables

Example:

Suppose we roll a die 100 times with 100 students.

we want to analyze whether the outcome is "Even" or "odd". One Variable = "Even / odd" outcome

Another variable = "Greater than 3" or "less than or equal to 3"

Contingency Table Example (100 rolls)

	<=3	>3	Total
Even Numbers	20	30	50
Odd Numbers	25	25	50
Total	45	55	100

This table shows the counts of how many outcomes fall in each category.

Interpretation:

- Total number of outcomes = 100 (since die roll 100 times)
- Even no. (2, 4, 6) came up 50 time, odd no (1, 3, 5) came up 50
- Among even no. 20 were <= 3 (only "2") & 30 were > 3 (4, 6)
- Among odd no. 25 were <= 3 (1, 3) & 25 were > 3 (5)

This helps analyze probability in multiple category once

Formula

P(Event) = Frequency of Event / Total outcomes

Example: P(Even AND > 3) = 30/100 = 0.3

```
Import random
import pandas as pd
rolls = [random.randint(1, 6) for _ in range(100)]
data = {
    "Even/odd": ["Even" if r % 2 == 0 else "odd" for r in rolls],
    "Range": ["<=3" if r <= 3 else ">3" for r in rolls]
}
```

```
df = pd.DataFrame(data)
contingency_table = pd.crosstab(df["Even/odd"], df["Range"],
                    margins = True)
print(contingency_table)
```

Joint Probability

Joint probability refers to the probability of two events happening together. If we have two random variables X and Y, the joint probability is denoted as:

$P(X = x, Y = y)$

This represents the probability that:

X takes a specific value x, AND

Y takes a specific value y simultaneously

Example:

Let's take the titanic dataset as an example

X = Passanger Class (Pclass)

Y = Survival Status (Survived)

The joint probability answers questions like:

- What is the probability that a passanger was in 1st class AND survived?

```
import panda as pd
data = {
    'Survived': [1, 0, 1, 0, 1, 0, 0, 1],
    'Pclass': [1, 1, 2, 3, 3, 2, 1, 3]
}
df = pd.DataFrame(data)
print(pd.crosstab(df['Survived'], df['Pclass']))
```

```
print(pd.crosstab(df['Survived'], df['PLlass'], normalize = 'all',
        margins = True))
print(pd.crosstab(df['Survived'], df['Pclass'], normalize='col'))
```

What are Joint Probability Contribution?

Joint probability contribution refers to the share of probability each (x, Y) combination contributes to the overall probability distribution

For example:

- P(Survived = 1, Pclass = 1) might be 0.20
- P(Surviced = 0, Pclass = 1) might be 0.10

Together, they explain how survival depends on class

Titanic Example

Lets say in Titanic dataset we found

- P(survived = 1, Pclass = 1) = 0.15 (15% of total passangers)
- P(survived = 0, Pclass = 1) = 0.10 (10% of total passangers)

Contribution Interpretation:

- Out of whole population, 15% were "1st class & survived"
- Out of whole population, 10% were "1st class & didn't survive"

These contributions help us understand how each group affects the overall survival probability distribution.

Marginal Probability

Marginal probability is the probability of a single event occuring, regardless, of the outcomes of other random variables.

It is also called:
- Simple probability
- Unconditional probability

Example
Let X = Passenger Class (Pclass)
Let Y = Survival Status (Survived)

The marginal probability of X = 1 (first class) is the probability that a passengers is in 1st class, regardless of whether they survived
The marginal probability of Y = 1 (survived) is the probability that a passenger survived, regardless of their class.

```
Import pandas as pd
import seaborn as sns
df = sns.load dataset('titanic')
df = df[['class', 'Survived']].dropna()
marginal_survival = df['survived'].value_counts(normalize = True)
print("Marginal Probability of Survival:\n", marginal_survival)

marginal_class = df['class'].value_counts(normalize = True)
print(marginal_class)
```

What is Marginal Distribution?
A margin probability distribution shows probabilities of a single random variable, ignoring (marginalizing over) other.
Example: Distribution of Pclass (ignoring survival)

Conditional Probability

Conditional probability is the measure of probability of an event (A) occuring given that another event (B) has already occured

It is denoted as: $P(A|B)$

$P(A|B) = P(A \cap B) / P(B)$

where:

- $P(A \cap B)$ - Probability that both A and B occur
- $P(B)$ - Probability that event B occurs.

Example 1: Tossing 3 unbiased coins

Question: What is the conditional probability that atleast 2 coins show head, given that atleast 1 coin shows head?

Step 1: Sample space for 3 coins = 8 outcomes

{HHH, HTH, THH, HTT, THT, TTH, TTT}

Step 2: Event B = at least 1 head

B = {HHH, HHT, HTH, THH, HTT, THT, TTH}

$\Rightarrow P(B) = 7/8$

Step 3: Event A = at least 2 heads

A = {HHH, HTH, HHT, THH}

Step 4: $A \cap B$ = outcomes where A and B happen

= {HHH, HHT, HTH, THH} (same as A, since all these contain at least 1 head)

$\Rightarrow P(A \cap B) = 4/8$

Step 5: Conditional Probability

$P(A|B) = P(A \cap B) / P(B)$

$= (4/8) / (7/8)$

$= 4/7$

Final Answer: $P(A|B) = 4/7$

Example 2: Rolling 2 fair dice

Question: What is the conditional probability that the sum is 7, given that the first die shows an old no.?

Step 1: Sample space of 2 dice = 36 outcomes

Step 2: Event B = first die shows odd no.

odd numbers of first die = $\{1, 3, 5\}$

For each choice, second die can be 1–6

$\Rightarrow$ total outcomes in $B = 3 * 6 = 18$

$\Rightarrow P(B) = 18/36 = 1/2$

Step 3: Event A = Sum of dice = 7

possible pairs: (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)

$\Rightarrow$ total outcomes in $A = 6$

$\Rightarrow P(A) = 6/36 = 1/6$

Step 4: $A \cap B$ = cases where sum = 7 AND first die odd

(1, 6), (3, 4), (5, 2)

$\Rightarrow$ total outcomes = 3

$\Rightarrow P(A \cap B) = 3/36 = 1/12$

Step 5: Conditional Probability

$P(A|B) = P(A \cap B) / P(B)$

$= (1/12) / (1/2)$

$= 1/6$

Final Answer: $P(A|B) = 1/6$

Intuition behind Conditional Probability Formula:

- When we say $P(A|B)$, it mean "Probability of A given that B has happened".
- So, we shrink our universe to only those outcomes where B happens.
- The denominator ($P(B)$) tells us the probability of being inside this new universe.
- The numerator ($P(A \cap B)$) tells us how many of these outcomes also satisfy A.
- Dividing them gives us the proportion of cases where A occurs inside the world of B.

Independent Vs Mutually Exclusive Events

Definations:

- Independent events (A, B):
  - Knowing that B happened dës not change the chance of A.
  - Formulas (equivalent when $P(B) > 0$ & $P(A) > 0$):

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

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

Mutually exclusive events (A, B):

- They cannot happen together in a single trial.

Formula: $A \cap B = \phi \Rightarrow P(A \cap B) = 0$

Note: If $P(A) > 0$ and $P(B) > 0$, mutually exclusive events are Not independent (Because independence would require $P(A \cap B) = P(A)\,P(B) > 0$)

- Intuition (Venn):
  - Independence: circles overlap in proportion to their areas.
  - Mutually exclusive: circles don't overlap at all.

Independent: Worked Examples.

Example 1: Coin + Die

- A = "Coin show heads" → $P(A) = 1/2$
- B = "die shows 4" → $P(B) = 1/6$
- Joint: $P(A \cap B) = P(H \ \& \ 4) = (1/2) * (1/6) = 1/12$
- Since $P(A \cap B) = P(A)P(B)$, A & B are independent.

Example 2: Two dice (classic non-trivial independence)

- Roll two fair dice (ordered pair)
- A = "Sum is 7" → favourable: (1,6), (2,5), (3,4), (4,3), (5,2) (6,1) → $6/36 = 1/6$
- B = "first die is odd" → first die $\in \{1, 3, 5\}$, second die any → $18/36 = 1/2$
- Check: $P(A)P(B) = (1/6)*(1/2) = 1/2 = P(A \cap B)$
- A and B are independent.

Example 3: Cards with replacement (independent draws)

- A = "first draw is an Ace" → 4/52
- B = "Second draw is an Ace (after replacing the first) → still 4/52
- $P(A \cap B) = (4/52) * (4/52) = P(A)P(B)$ = independent.

Dependent (for contrast): Worked Example

Card without replacement (depedent draws)

- A = "first draw is an Ace" → 4/52

- B = "Second draw is an Ace"
- $P(B \mid A) = 3/51$ (one Ace removed)
- $P(B \mid A^c) = 4/51$ (no Ace removed)
- Since $P(B \mid A) \neq P(B \mid A^c)$, the second draw depends on the first
- Also $P(A \cap B) = (4/52) * (3/51) \neq P(A)P(B)$

Mutually Exclusive: Worked Examples

Example 1: Single Coin toss

- A = "Head", B = "Tails"
- $A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$
- With $P(A) = P(B) = 1/2$, independence world require $P(A \cap B) = 1/4$, but its 0.
  - A and B are mutually exclusive not independent

Bayes Theorem

Bayes Theorem is a rule in probability that helps us update the probability of an event (hypothesis) when we get new evidance.

Formula: $P(A \mid B) = [P(A) * P(B \mid A)] / P(B)$

where

$P(A)$: Prior probability of events A (before seeing evidence)

$P(B \mid A)$: likelihood (probability of observing B given A is true)

$P(B)$: Marginal probability of B (total probability of evidence)

$P(A \mid B)$: Posterior probability (updated probability of A after seeing B)

Derivation Intuition

From conditional probability

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

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

Rearranging

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

Substitute into first equation:

$$P(A \mid B) = [P(A) * P(B \mid A)] / P(B)$$

Usage in Machine Learning

- Naive Bayes Classifier
- Medical diagnosis
- Fraud detection
- Bayesian inference

Example Problem

A medical test detects a disease

- P(Disease) = 0.01 (1% people have disease)
- P(Positive | Disease) = 0.99 (99% sensitivety)
- P(Posibve | No Disease) = 0.05 (5% false positive rate)

Question: What is the probability that a person has the disease given that their test result is positive?

Using Bayes:

P(Disease | Positive) = [P(Disease) * P(Posibve| Disease)] / P(Positive)

where P(Posibve) = P(Disease) * P(Positive | Disease) + P(No Disease) * P(Positive | No Disease)

Discrete Probability Distribution

A discrete probability distribution gives the probabilities of outcomes for a discrete random variable (a variable that takes only countable values like 0, 1, 2, ...). It tells us how likely each outcome is

Example:

- Number of heads in coin toss
- Number of emails in an hour
- Rolling a die

Conditions:

1) Probability must be between 0 & 1 → $0 \leq P(X=x) \leq 1$
2) Total probability = 1 → $\sum P(X=x) = 1$

Example – Tossing 2 coins:

Sample Space = {HH, HT, TH, TT}

Let X = number of Tails

X(no. of tails)	Outcomes	$P(X=x)$
0	{HH}	1/4
1	{HT, TH}	2/4
2	{T, T}	1/4

Total probability = $1/4 + 2/4 + 1/4 = 1$

Important Formulas:

PMF (Probability Mess Function)

$f(x) = P(X=x)$

Example: Tossing 2 coins → $P(X=1) = 2/4$

CDF (cumulative Distribution Function)

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

Example: Rolling a 4-sided die

$P(1 < X \leq 3) = F(3) - F(1) = 0.7 - 0.2 = 0.5$

Mean (Expected Value)

$E[X] = \sum (x \cdot P(X=x))$

Variance

$Var[X] = \sum ((x-\mu)^2 \cdot P(X=x))$

Steps to Find DPD

- Write the sample space
- Define the random Variable X
- Assign probabilities for value of X
- Make a probability distribution table.

Types of DPD

In discrete probability distributions, the random variable takes distinct values (like, 0, 1, 2, ...)

Some Common types are:

- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution
- Geometric Distribution

1- Bernoulli Distribution

A random experiment with only Two possible outcomes

- Success (1) with probability P
- Failure (0) with probability 1 - P

Probability Mass Function (PMF):

- $P(X = x) = P$ if $x = 1$

  $1 - P$ if $x = 0$

Example:

- Toss a Coin Once
- Left success = Head → P = 0.5
- Failure = Tail → 1 - P = 0.5
- So, $P(X = 1) = 0.5$, $P(X = 0) = 0.5$

2- Binomial Distribution

Repeated Bernoulli trials (n times).

- Counts the no. of successes in 'n' independent trials.
- Each trial has probability of sucess = p, failure = 1 - p

Probability Mass Funtion (PMF):

- P(X = x) = (n choose x) * p^x * (1-p)^(n-x)
- where (n choose x) = n! / [x! * (n-x)!]

Example:

- Toss a coin 3 times (n = 3, p = 0.5)
- Finding probability of getting exactly 2 Heads.
- P(X = 2) = (3 choose 2) * (0.5)^2 * (0.5)^1

  = 3 * 0.25 * 0.5

  = 0.375

3- Poisson Distribution

Describe the probability of a given number of events happening in a fixed interval of time or space.

- Controlled by parameter λ (mean no. of events)

Probability Mass Funtion (PMF)

- P(X = x) = (λ^x * e^(-λ)) / x!

Example:

- A call center receives on average 5 calls per min (λ=5)
- Find probability of receiving exactly 3 calls in min
- P(X = 3) = (5^3 * e^(-5)) / 3!

  = (125 * e^(-5)) / 6

  = 0.1404

4- Geometric Distribution

Models the no. of trials until the first success occurs.

Probability of success = p, probability of failure = (1-p)

Probability Mass Funtion (PMF)

- P(X = x) = (1-p)^(x-1) * p

- where X = number of trial in which first success happens.

Example:

- Toss a coin until first Head appears ($p = 0.5$)
- Find probability that the first Head appears on the 3rd toss.
- $P(X = 3) = (1 - 0.5)\text{^}(3 - 1) * 0.5$
  $= (0.5)\text{^}2 * 0.5$
  $= 0.125$

Contineous Probability Distribution (CPD)

CPDs describe probabilities for contineous random variables (e.g, height, time, prices). Values can take Any number in a range.

Key Functions

1- PDF (Probability Density Function)

- Probability "density" at a point.
- Area under curve = probability

2- CDF (Cumulative Distribution Function)

- $P(X \leq x)$, probability variable is less than/equal x
- CDF = $\int$ PDF, PDF = derivative of CDF

Normal (Gaussian) Distribution

- Bell-Shaped, symmetric around mean μ
- Defined by mean (μ) and std. deviation (σ)

Formula:

- $f(x) = 1/(\sigma\sqrt{(2\pi)}) * \exp(-(x-\mu)\text{^}2/(2\sigma\text{^}2))$

Rule of thumb:

- 68% data within $\mu \pm \sigma$

- 95% data within $\mu \pm 2\sigma$
- 99.7% data within $\mu \pm 3\sigma$

Uniform Distribution

- All values b/w [a, b] equally likely

Formula

- $f(x) = 1/(b-a)$, for $a \leq x \leq b$

Mean = $(a+b)/2$

Var = $(b-a)^2/12$

Exponential Distribution

- Time b/w events in a poisson process. Defined by rate λ (events per unit time)

Formula

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

Mean = $1/\lambda$

Var = $1/\lambda^2$

Chi-Squared Distribution

- Special case of Gamma distribution. Used in hypothesis testing (χ^2 testing). Defined by degrees of freedom (k)

Mean = k

Var = 2k

Beta Distribution

- Models probabilities (values strictly in [0, 1]). Defined by shape parameters α and β

Formula:

- $f(x) \propto x^{(\alpha-1)} * (1-x)^{(\beta-1)}$, for $0 \leq x \leq 1$

Mean = $\alpha/(\alpha+\beta)$

Var = $(\alpha\beta) / [(\alpha+\beta)^2 * (\alpha+\beta+1)]$
Uses: Bayesian inference (priors / posteriors for probabilities)

Gamma Distribution
- Generalization of exponential distribution. Defined by shape parameter k and rate λ.

Formula
- $f(x) = (\lambda^k / \Gamma(k)) * x^{(k-1)} * e^{(-\lambda x)}$, for $x \geq 0$

Mean = k / λ
Var = k / λ^2
Uses: Modeling waiting times, Bayesian statistics

Sampling Distribution
A sampling distribution is the probability distribution of a statistic (like mean or proportion) obtained from repeated random samples of a population.

Why its matters:
- Helps in inferential statistics (estimating population from sample)
- Forms the basis for confidence intervals & hypothesis tests.

Key Terms:
- Population: whole group being studied
- Sample: subset of population
- Parameter: Value from population (e.g, μ, σ)
- Statistic: Value from sample (e.g $\bar{x}$, s)
- Standard Error: (std. deviation of sampling distribution
- CLT (central limit Theorem): Sample means approach a normal distribution as n ↑

Types of Sampling Distributions

1- Chi-Square Test (χ^2 Test)

The chi-square test is a statistical method used to check if there is a relationship (associated) between two categorical variables It compares observed frequencies (actual data) with expected frequencies (if no relationship existed)

Example:

- Variable 1 = Favorite Color
- Variable 2 = Favourite Ice Cream
- Ho (Null Hypothesis): No relationship
- H_1 (Alternative Hypothesis): There is a relationship

Formula: $\chi^2 = \Sigma (O_i - E_i)^2 / E_i$

where O_i = observed frequency, E_i = Expected frequency

Steps of Chi-Square Test

- Define Hypothesis
  - H_0: Variables are independent
  - H_1: Variables are dependent
- Collect Data
  - Example (contingency Table: Ice Cream Vs Gender)

	chocolate	Vanilla	Strawberry
Male	20	15	10
Female	25	20	30

- Calculate Expected Frequency
  - E_i = (Row Total x Column Total) / Grand Total

    Example: Male - chocolate = $(45 \times 45)/120 = 16.875$
- Apply Formala:

  $\chi^2 = \Sigma (O_i - E_i)^2 / E_i$

- Degrees of freedom
  - $df = (rows - 1) \times (columns - 1)$
- Compare χ^2 value with Chi-Square Table
  - If $\chi^2 >$ critical value → reject H_0
  - Otherwise → Fail to reject H_0
- Interpret Results
  - $p < 0.05$ → significant relationship exists.
  - $p > 0.05$ → No significant relationship

Types of Chi-square Tests

1- Test of Independence:
- checks if two variables are related.
- Examples: Gender Vs Ice cream Preference

2- Goodness-of-Fit Test:
- checks if data fits a specific distribution
- Example: Is a die fair?

2- F-Test In Statistics

- The F-test checks whether the variances of two populations (or sample) are equal. It is based on the F-distribution

F- Statistics Formula
- For samples
  - $F = (S1^2 / S2^2)$

  (where $S1^2 \geq S2^2$, i.e, always divide larger variance by smaller)

$df1 = n1 - 1$ → numerator degrees of freedom

$df2 = n2 - 1$ → denominator degrees of freedom

- Hypothesis
  1- Right tailed (most common)
    - $H_0: \sigma_1^2 = \sigma_2^2$ (variances equal)
    - $H_1: \sigma_1^2 > \sigma_2^2$ (variance 1 greater)
  2- Left-tailed
    - $H_0: \sigma_1^2 = \sigma_2^2$
    - $H_1: \sigma_1^2 < \sigma_2^2$
  3- Two-tailed
    - $H_0: \sigma_1^2 = \sigma_2^2$
    - $H_1: \sigma_1^2 \neq \sigma_2^2$

3- T- Distribution

- A probability distribution used when
  - Sample Size is small ($n \leq 30$)
  - Population standard deviation (σ) in unknown

Similar to normal distribution but has heavier tails
Discovered by W.S. Gosset (pen name: "Student")

- Formula for t-Score
  - $t = (\bar{x} - \mu) / (S / \sqrt{n})$

where

$\bar{x}$ = sample mean

μ = population mean

s = Sample standard deviation

n = Sample Size

Central Limit Theorem

CLT states: For large sample size ($n > 30$), the distribution of sample means becomes ~ Normal, regardless of the population distribution (normal, skewed)

Means of sampling distribution = population mean (μ)
Standard deviation = $\sigma/\sqrt{n}$ (called standard error)
Formula:

$$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$$
$$z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$$

Example

- population mean = 100, σ = 20, sample size n = 100
- Standard Error = $\sigma/\sqrt{n} = 20/\sqrt{100} = 2$
- Distribution of sample mean: N(100, 2)

Steps to Solve CLT Problems

- Identify mean (μ), std dev (σ), sample size (n)
- Compute standard error = $\sigma/\sqrt{n}$
- Find Z-score = $(\bar{x} - \mu)/(\sigma/\sqrt{n})$
- Use Z-table to find probability