Random Variables and Stochastic Process

Dr. G. Omprakash

Assistant Professor, ECE, KLEF

Syllabus

Course Outcome	**Syllabus**
CO1	**THE RANDOM VARIABLE AND MULTIPLE RANDOM VARIABLES**: Introduction, Review of Probability Theory, Definition of a Random Variable, Conditions for a Function to be a Random Variable, Discrete, Continuous and Mixed Random Variables, Distribution and Density functions, Properties, Binomial, Poisson, Uniform, Gaussian, Exponential, Rayleigh, Conditional Distribution, Conditional Density, Properties. Vector Random Variables, Joint Distribution Function, Properties of Joint Distribution, Marginal Distribution Functions, Conditional Distribution and Density, Statistical Independence, Sum of Two Random Variables, Sum of Several Random Variables, Central Limit Theorem: Unequal Distribution, Equal Distributions.
CO2	**OPERATION ON ONE AND MULTIPLE RANDOM VARIABLE-EXPECTATIONS**: Introduction, Expected Value of a Random Variable, Function of a Random Variable, Moments about the Origin, Central Moments, Variance and Skew, Chebychev's Inequality, Characteristic Function, Moment Generating Function, Transformations of a Random Variable: Monotonic Transformations for a Continuous Random Variable, Nonmonotonic Transformations of Continuous Random Variable. OPERATIONS ON MULTIPLE RANDOM VARIABLES: Joint Moments about the Origin, Joint Central Moments, Joint Characteristic Functions, Jointly Gaussian Random Variables: Two Random Variables case, N Random Variables case, Properties, Transformations of Multiple Random Variables, Linear Transformations of Gaussian Random Variables.
CO3	**RANDOM PROCESSES –TEMPORAL CHARACTERISTICS**: The Random Process Concept, Classification of Processes, Deterministic and Non-deterministic Processes, Distribution and Density Functions, Concept of Stationarity and Statistical Independence. First-order stationary Processes, Second-order and Wide-Sense Stationarity, Nth-order and Strict- Sense Stationarity, Time Averages and Ergodicity, Autocorrelation Function and its Properties, Cross-Correlation Function and its Properties, Covariance Functions, Gaussian Random Processes, Poisson Random Process
CO4	**RANDOM PROCESSES -SPECTRAL CHARACTERISTICS**: The Power Density Spectrum: Properties, Relationship between Power Density Spectrum and Autocorrelation Function, The Cross-Power Density Spectrum, Properties. Random Signal Response of Linear Systems: System Response, Convolution, Mean and Mean-squared Value of System Response, Autocorrelation Function of Response, Cross-Correlation Functions of Input and Output.

Textbooks

- Probability, Random Variables & Random Signal Principles, Peyton Z.Peebles, TMH, 4th Edition, 2001.
- Bertsekas, Dimitri, and John N. Tsitsiklis. Introduction to probability. Vol. 1. Athena Scientific, 2008.
- Probability, Random Variables and Stochastic Processes, Athanasios Papoulis and S.Unnikrisha, PHI,4th Edition, 2002.
- Probability and Random Processes with Applications to Signal Processing, Henry Starkand John W.Woods, Pearson Education, 3rd Edition, 2001.
- Probability Theory and Random Processes, P. Ramesh Babu, McGrawHill,2015.
- Schaum's Outline of Probability, Random Variables, and Random Processes, 1997.

Evaluation Plan

EVALUATION PLAN:

Evaluation Type	Evaluation Component	Weightage/ Marks		Assessment Dates	Duration (Hours)	CO1	CO2	CO3	CO4
End Semester Summative Evaluation Total= 40 %	End Semester Exam	Weightage	40		180	10	10	10	10
		Max Marks	100			25	25	25	25
In Semester Formative Evaluation Total= 26 %	Home Assignment and Textbook	Weightage	5		100	1.25	1.25	1.25	1.25
		Max Marks	40			10	10	10	10
	Tutorial	Weightage	15		100	3.75	3.75	3.75	3.75
		Max Marks	100			25	25	25	25
	ALM	Weightage	6		100	1.5	1.5	1.5	1.5
		Max Marks	100			25	25	25	25
In Semester Summative Evaluation Total= 34 %	Semester in Exam-II	Weightage	17		90			8.5	8.5
		Max Marks	50					25	25
	Semester in Exam-I	Weightage	17		90	8.5	8.5		
		Max Marks	50			25	25		

Guidelines

- Office Hours: Saturday 10 AM to 12 Noon (H0-15)
- Mathematics is learnt by practice
  - Don't watch or read Maths!!
- Calculators are must!!
- Maintain a separate notebook
- Tutorial marks are awarded based on your notebook
- No Smartphones during class
- Do not discuss while I am instructing
  - Time will be given to discuss

Aim of the session

Introduction to Course and Probability Theory

Learning Outcomes

Describe the practical applications of Random Variables.

Introduce Probability theory

Motivation

Uncertainty in Everyday Life!!

- **Can we assign a number to an uncertain outcome?**
- Suppose you're waiting at the bus stop. How long will the bus take to arrive?
- What is the time taken to commute to college?
- How many customers will visit the bank today?
- How many calls are received at the customer care?
- How is the weather tomorrow?

Risk Modeling in the Insurance

- An insurance company models expected claims for a health policy.
- Let X= cost of claim from one policyholder.
- Mean=₹10,000
- Variance=₹10,000,000
- **Though mean is small, high variance can ruin the company if a few claims are huge**.

Game Winnings based on coin toss

- Game rule: Toss a fair coin. If heads, you win ₹10, if tails, you lose ₹5.
- Let X be the Winning value
- X= 10 with probability 0.5
- X= -5 with probability 0.5
- Mean=₹2.5; Variance= ₹56.25; Standard Deviation=₹7.5

Game Winnings based on coin toss

- Imagine you're playing this game 10 times:
  - Expected profit is ₹25 overall ($10 \times$ ₹2.5)
- You could easily end up with a loss, or a profit much higher than ₹25. How?

Game Winnings based on coin toss

- Imagine you're playing this game 10 times:
  - Expected profit is ₹25 overall (10 × ₹2.5)
- You could easily end up with a loss, or a profit much higher than ₹25. How?
- In any individual game, the outcome jumps **far** from the average:
  - ₹10 (which is ₹7.5 above the mean)
  - ₹–5 (which is ₹7.5 below the mean)
- **Standard deviation** (₹7.5) is a measure of how far the outcomes are likely to deviate from the average gain — this deviation is the "**risk**"

What happens if standard deviation is just ₹1?

Takeaway

- If you only look at the mean (₹2.5), you might wrongly think:
  - **"I'll always make money!"**
- But variance warns you:
  - **You might also lose money in the short run. Be cautious!**

In finance, insurance, gambling, and everyday decisions, **people care** not just about the average, but **how much volatility or uncertainty there is around that average**.

★ Mean → Planning & Expectation.

✧ Variance → Safety margins, Risk, Quality

Temperature Sensor

Imagine you are using a temperature sensor (say a thermistor or a digital IC like LM35) to measure the room temperature.

- The sensor is connected to a Arduino/Raspberry Pi and gives an analog **output voltage proportional to the temperature**.
- Even if the room temperature is constant, the output voltage fluctuates
- Fluctuations arise due to:
  - Electrical noise (from power lines, ADC quantization, interference)
  - Manufacturing variations
  - Thermal noise inside the sensor's material
  - Small movements or vibrations in the environment

Instead of getting a single constant value, you get random variations around the "true" voltage

Temperature Sensor

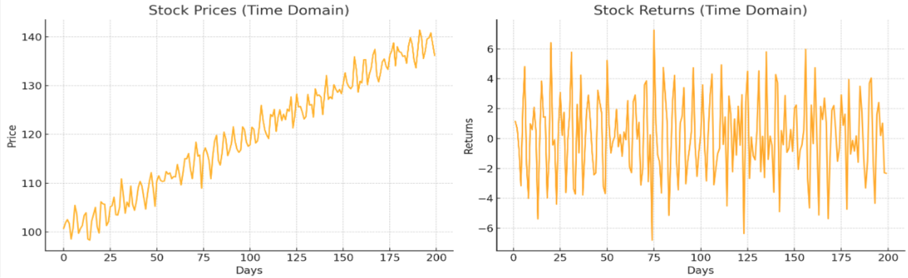
- Collect 100 readings of sensor voltage
- Plot a histogram of these readings
- **The histogram looks like a bell-shaped Gaussian curve**
- Most values cluster near the mean

Spectral Analysis of Stock Prices

Aim:

- Detect seasonal trends
- Understand market volatility
- Identify **dominant periodic patterns**

Not always visible in the time domain!!



Spectral Analysis of Stock Prices

Power Spectral Density (PSD) tells us how the power (or variance) of a signal is distributed across different frequencies.

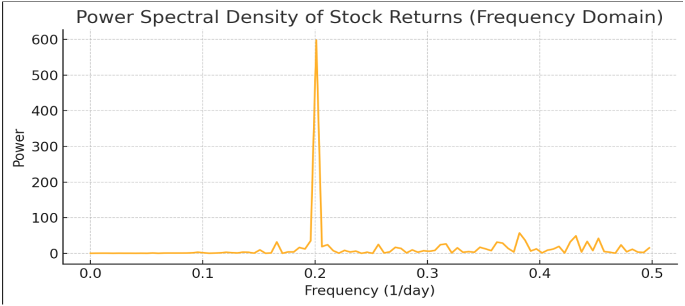
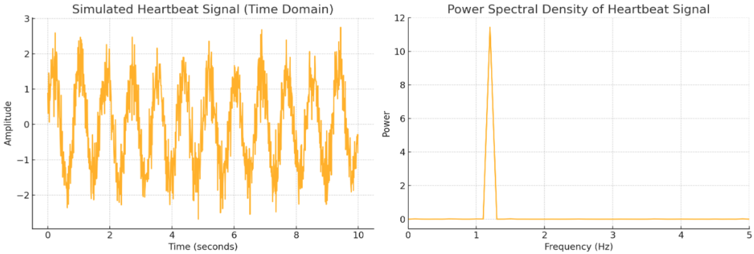


Figure: Clear peak at the frequency (0.2) corresponding to a 5-day cycle

Power Spectral Density of Heartbeat Signal



- Time domain signal (left) resembles a heartbeat, with some added random noise
- A clear peak at around 1.2 Hz (Right)
  - This corresponds to 1.2 beats per second, or 72 beats per minute (BPM).
- PSD analysis helps extract meaningful physiological information
- **Medical devices like ECG monitors, wearable fitness trackers, and ICU systems use spectral analysis**

Videos

- Variable Change
  - `https://www.youtube.com/watch?v=CYyUuIXzGgI`
- Trade Probability
  - `https://://youtube.com/shorts/RuaJiLl4WHc?si=4Uwhn3GfrT-pWRFe`
- Probability Based Mindset for Trading
  - `https://www.youtube.com/watch?v=ldzf9qA3Oks`

Probability Theory

Set Definitions

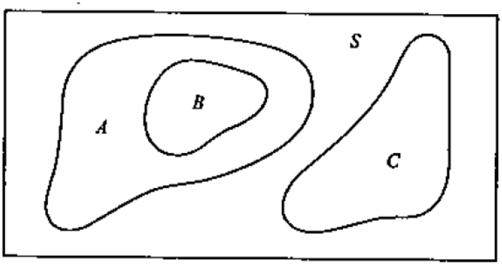
- A ***set*** is a collection of objects.
  - Denoted by a capital letter: "A"
- The objects are called ***elements*** of set.
  - Element is represented by a lower-case letter
- Eg: $A = \{1, 3, 5, 7\}$
  - Set is "A"; 1 is an element of A
- $a \in A \implies a$ is an element of A
- $a \notin A \implies a$ is **not** an element of A

Set Definitions

- $A \subseteq B \implies A$ is known as ***subset*** of B
  - Every element of set A is also an element in another set B
  - A is said to be *contained* in B
- If atleast one element exists in B which is not in A, then A is a ***proper subset*** of B
  - Denoted by: $A \subset B$
- **Null set**: A set is said to be *empty* or *null set* if it has no elements.
  - Denoted by symbol ϕ
- **Disjoint sets**: Two sets A and B are called *disjoint* or ***mutually exclusive*** if they have no common elements

Venn Diagram

Venn Diagram: Sets are represented by closed-plane figures

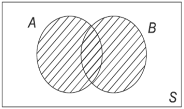


Source: Peyton Z. Peebles,Jr.,2001

Figure: Venn Diagram

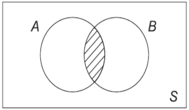
Set Operations: Union and Intersection

- **Union of Sets** A and B is the set of all elements that belong A or B (or both)
  - Denoted by: $A \cup B$



$A \cup B$

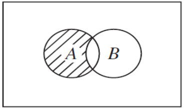
- **Intersection of Sets** A and B is the set of all elements common to both A and B
  - Denoted by: $A \cap B$



$A \cap B$

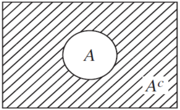
Set Operations: Complement and Difference

- The set containing all elements of A that are not present in B is called the **difference** of A and B
  - Denoted by: $A - B$



$A - B$

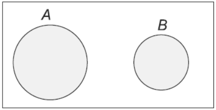
- The **complement** of set A is the set of all elements not in A
  - Denoted by: $\overline{A}$



$\overline{A}$

Disjoint Sets (Mutually Exclusive Sets)

Disjoint Sets: Two sets A and B are said to be disjoint or mutually exclusive if they have no elements in common.



$$A \cap B = \phi$$

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $A \cup B$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $A \cup B$?
    - $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- What is $A \cup C$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $A \cup B$?
  - $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- What is $A \cup C$?
  - $A \cup C = \{1, 3, 4, 5, 6, 7, 8, 12\}$
- What is $B \cup C$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $A \cup B$?
  - $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- What is $A \cup C$?
  - $A \cup C = \{1, 3, 4, 5, 6, 7, 8, 12\}$
- What is $B \cup C$?
  - $B \cup C = \{1, 2, 3, 4, 6, 7, 8, 9, 10, 11\}$
- $A \cap B =?$; $A \cap C =?$; $B \cap C =?$

Example

Consider the sets: $S = \{1 \le integers \le 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $A \cup B$?
  - $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- What is $A \cup C$?
  - $A \cup C = \{1, 3, 4, 5, 6, 7, 8, 12\}$
- What is $B \cup C$?
  - $B \cup C = \{1, 2, 3, 4, 6, 7, 8, 9, 10, 11\}$
- $A \cap B =?$; $A \cap C =?$; $B \cap C =?$
  - $A \cap B = \phi$; $A \cap C = \{1, 3\}$; $B \cap C = \{6, 7, 8\}$

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?
  - $\overline{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$
- What is $\overline{B}$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?
  - $\overline{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$
- What is $\overline{B}$?
  - $\overline{B} = \{1, 3, 4, 5, 12\}$
- What is $\overline{C}$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?
  - $\overline{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$
- What is $\overline{B}$?
  - $\overline{B} = \{1, 3, 4, 5, 12\}$
- What is $\overline{C}$?
  - $\overline{C} = \{2, 5, 9, 10, 11, 12\}$

- What is $A - C$?

Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?
  - $\overline{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$
- What is $\overline{B}$?
  - $\overline{B} = \{1, 3, 4, 5, 12\}$
- What is $\overline{C}$?
  - $\overline{C} = \{2, 5, 9, 10, 11, 12\}$
- What is $A - C$?
  - $A - C = \{5, 12\}$
- What is $B - C$?

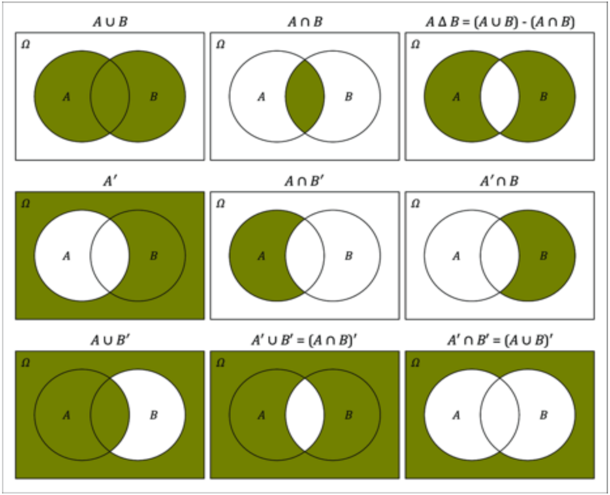
Example

Consider the sets: $S = \{1 \leq integers \leq 12\}$, $A = \{1, 3, 5, 12\}$, $B = \{2, 6, 7, 8, 9, 10, 11\}$, $C = \{1, 3, 4, 6, 7, 8\}$

- What is $\overline{A}$?
  - $\overline{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$
- What is $\overline{B}$?
  - $\overline{B} = \{1, 3, 4, 5, 12\}$
- What is $\overline{C}$?
  - $\overline{C} = \{2, 5, 9, 10, 11, 12\}$
- What is $A - C$?
  - $A - C = \{5, 12\}$
- What is $B - C$?
  - $B - C = \{2, 9, 10, 11\}$

Algebra of Sets

- **Commutative Law**
  - $A \cup B = B \cup A$
  - $A \cap B = B \cap A$
- **Distributive Law**
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- **Associative Law**
  - $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
  - $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$
- **De Morgan's Laws**
  - $(\overline{A \cup B}) = \overline{A} \cap \overline{B}$
  - $(\overline{A \cap B}) = \overline{A} \cup \overline{B}$
- **Duality Principle**: The identity is preserved if we replace
  - $\cup \implies \cap; \cap \implies \cup$
  - $S \implies \phi; \phi \implies S$


$A \cup B$
$A \cap B$
$A \Delta B = (A \cup B) - (A \cap B)$
A'
$A \cap B'$
$A' \cap B$
$A \cup B'$
$A' \cup B' = (A \cap B)'$
$A' \cap B' = (A \cup B)'$
Ω
A
B

Terminology

- **Experiment**: An experiment is a procedure we perform that produces some result.
- **Outcome**: An outcome is a possible result of an experiment.
- **Event**: An event is a certain set of outcomes of an experiment.
  - Interested in characteristic of outcomes
  - It is a subset of a sample space
  - If two events have no common outcomes they are **mutually exclusive**
- **Sample Space**: The set of all possible outcomes in any given experiment is called the sample space (Denoted by symbol S)

Examples

Experiment	Sample Space	Outcome	Event
Throwing a die	$\{1, 2, 3, 4, 5, 6\}$	1 or 2 or 3..	Getting even number $\{2, 4, 6\}$
Tossing two coins	$\{(H, H), (H, T), (T, H), (T, T)\}$	(H, H)	Atleast one Head $\{(H, H), (H, T), (T, H)\}$

Probability

- To each event in the sample space assign a nonnegative number called *probability*
- Probability is a **function** of *event*
- $P(A) \implies$ *probability of event A*
- The assigned probability should satisfy three axioms

Axioms of Probability

The assigned probabilities should satisfy three axioms

- **Nonnegativity**: $P(A) \geq 0$, for every event A
- **Normalization**: The probability of the entire sample space S is equal to 1

$$P(S) = 1$$

  - S is known as the *certain* event
  - Null set ϕ is known as the *impossible event*. ($P(\phi) = 0$)

- **Additivity**: For a sequence of N mutually exclusive events $A_1, A_2, ..., A_N$,

$$P(A_1 \cup A_2 \cup ... \cup A_N) = P(A_1) + P(A_2) + ... + P(A_N)$$

  - Mutually exclusive $\implies$ $A_m \cap A_n = \phi$ if $m \neq n$

Definitions for probability

Three types of definitions for probability.

- Axiomatic definition
- Relative-frequency definition
- Classical definition

Axiomatic definition

Axiomatic Definition

Consider a random experiment with sample space S. For each event, we assign a non-negative number called probability of the event A denoted as $P(A)$. The probability satisfies three axioms

- Axiom 1: $0 \leq P(A) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: For N mutually exclusive events $(A_1, A_2, ..., A_N)$:
  - $P(A_1 \cup A_2 \cup ... \cup A_N) = P(A_1) + P(A_2) + ... + P(A_N)$
  - Mutually exclusive $\implies A_m \cap A_n = \phi$ for all $m \neq n$

Relative-frequency definition

Probability as a relative frequency

Consider a random experiment that is performed n times. If an event occurs $n(A)$ times then the probability of event A; $P(A)$ is defined as

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

Classical Definition

Classical Definition

The probability $P(A)$ of an event A is the ratio of the number of outcomes $n(A)$ of an experiment that are favourable to A to the total number of possible outcomes of the experiment

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

Example

A box contain 80 resistors each having same size and shape. Of the 80 resistors, 18 are 10Ω, 12 are 22Ω, 33 are 27Ω, 17 are 47Ω

$$P(draw\ 10\Omega) = \frac{18}{80}; \qquad P(draw\ 22\Omega) = \frac{12}{80};$$

$$P(draw\ 27\Omega) = \frac{33}{80}; \qquad P(draw\ 47\Omega) = \frac{17}{80};$$

Example

A box contains 25 parts of which 10 are defective. Two parts are being drawn simultaneously in a random manner from the box. The probability of both parts being good is

Soln: Number of ways of drawing 2 parts from 25 parts= $^{25}C_2$.

Number of ways of drawing 2 good parts from the 15 good parts=$^{15}C_2$.

Probability that the both parts are good=$\frac{^{25}C_2}{^{15}C_2} = \frac{7}{20}$

Why Joint Probability

- $P(A \cap B)$=Probability that both A and B occur
- Joint probability **models dependencies**, not just individual likelihoods.
- *"Doctors don't just look at one symptom—they look for patterns that occur together."*
    - Probability a patient has fever and cough

Joint Probability

Joint Probability: The probability $P(A \cap B)$ is called the *joint probability* for two events A and B which intersect in the sample space

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

HR (bpm) \ SBP (mmHg)	**110**	**120**	**130**
60	0.05	0.10	0.05
70	0.10	0.20	0.10
80	0.05	0.10	0.05

Table: Joint probability table showing the distribution of systolic blood pressure (SBP) and heart rate (HR) among patients.

Each cell shows joint probability $P(X = SBP, Y = HR)$

Joint Probability

There are 100 resistors in a box.. Assume each resistor has the same likelihood to be chosen. Define three events: A as draw a 47Ω resistor; B as draw a resistor with 5% tolerance; C as draw a 100Ω resistor.

	Tolerance		
Resistance(Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

Table: Joint frequency distribution of resistance and tolerance of resistors in a box.

Joint Probability

There are 100 resistors in a box.. Assume each resistor has the same likelihood to be chosen. Define three events: A as draw a 47Ω resistor; B as draw a resistor with 5% tolerance; C as draw a 100Ω resistor.

	Tolerance		
Resistance(Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

Table: Joint frequency distribution of resistance and tolerance of resistors in a box.

$$P(A) = P(47\Omega) = \frac{44}{100}$$
$$P(B) = P(5\%) = \frac{62}{100}$$
$$P(C) = P(100\Omega) = \frac{32}{100}$$

Joint Probability

There are 100 resistors in a box.. Assume each resistor has the same likelihood to be chosen. Define three events: A as draw a 47Ω resistor; B as draw a resistor with 5% tolerance; C as draw a 100Ω resistor.

	Tolerance		
Resistance(Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

Table: Joint frequency distribution of resistance and tolerance of resistors in a box.

$P(A) = P(47\Omega) = \frac{44}{100}$
$P(B) = P(5\%) = \frac{62}{100}$
$P(C) = P(100\Omega) = \frac{32}{100}$
Joint Probabilities are :
$P(A \cap B) = P(47\Omega \cap 5\%) = \frac{28}{100}$
$P(A \cap C) = P(47\Omega \cap 100\Omega) = 0$
$P(B \cap C) = P(5\% \cap 100\Omega) = \frac{24}{100}$

Joint Probability

"In a school of 100 students: 60 like music, 50 like dance, and 30 like both. What's the probability that a randomly chosen student likes both music and dance?"

Why Conditional Probability

We update our beliefs based on new information.

Conditional Events:

- "What is the chance a person has diabetes **given** they are obese?"
- "What is the chance it rains today **given** it's cloudy?"

Would you guess differently if you didn't know the person was obese?

- **Conditional probability is the foundation of Bayes' Theorem**
  - Used in Machine Learning, weather prediction, Medical diagnosis, spam filtering.
- Conditional probability is how we reason in real life — when we say '**if**'

Conditional Probability

Conditional Probability: Given some event B with nonzero probability ($P(B) > 0$). The *conditional probability* of an event A, given B is

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)}; \text{ where } P(B) \neq 0$$

Rewriting the above equation

$$P(A \cap B) = P(A \setminus B) P(B) \tag{1}$$

- Reflects the fact that probability of an event A may depend on second event B
- If A and B are mutually exclusive $\implies A \cap B = \phi \implies P(A \setminus B) = 0$

Problem: Conditional Probability

	Tolerance		
Resistance(Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

Table: Joint frequency distribution of resistance and tolerance of resistors in a box.

$P(A) = P(47\Omega) = \frac{44}{100}$
$P(B) = P(5\%) = \frac{62}{100}$
$P(C) = P(100\Omega) = \frac{32}{100}$
Joint Probabilities are :
$P(A \cap B) = P(47\Omega \cap 5\%) = \frac{28}{100}$
$P(A \cap C) = P(47\Omega \cap 100\Omega) = 0$
$P(B \cap C) = P(5\% \cap 100\Omega) = \frac{24}{100}$
What is $P(A \setminus B)$? $P(A \setminus C)$? $P(B \setminus C)$?

Problem: Conditional Probability

	Tolerance		
Resistance(Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

Table: Joint frequency distribution of resistance and tolerance of resistors in a box.

$P(A) = P(47\Omega) = \frac{44}{100}$
$P(B) = P(5\%) = \frac{62}{100}$
$P(C) = P(100\Omega) = \frac{32}{100}$
Joint Probabilities are :
$P(A \cap B) = P(47\Omega \cap 5\%) = \frac{28}{100}$
$P(A \cap C) = P(47\Omega \cap 100\Omega) = 0$
$P(B \cap C) = P(5\% \cap 100\Omega) = \frac{24}{100}$
What is $P(A \setminus B)$? $P(A \setminus C)$? $P(B \setminus C)$?

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} = \frac{28}{62}; P(A \setminus C) = \frac{P(A \cap C)}{P(C)} = 0; P(B \setminus C) = \frac{P(B \cap C)}{P(C)} = \frac{24}{32};$$

Problems

- We toss a fair coin three successive times. Find the conditional probability $P(A \setminus B)$ when A and B are the events A={more heads than tails come up} and B={ 1st toss is a head}. (Bertsekas Eg 1.6)

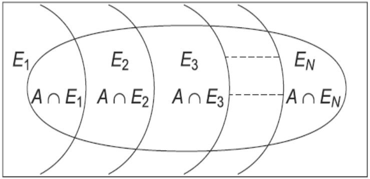
Why Total Probability

There is an **uncertainty** due to **multiple causes**

- *Imagine your internet is down in the morning before a deadline. It could be your router, your ISP, or your laptop. What's the chance you'll be able to fix it in 10 minutes?*
- *There is a delay in the package delivery. The delay could be due to weather, logistics, or traffic. Each contributes to the total risk.*

The Law of Total Probability lets you think systematically, even when life is uncertain

Total Probability



Given N mutually exclusive events $E_1, E_2, ..., E_N$.
Mutually exclusive $\implies$ $E_m \cap E_n = \phi$, where $m \neq n \in \{1, 2, ..., N\}$ and

$$\bigcup_{n=1}^{N} E_n = S$$

The total probability of event A is given by

$$P(A) = \sum_{n=1}^{N} P(A \setminus E_n) P(E_n) \tag{2}$$

Total Probability (How)

Given N mutually exclusive events $E_1, E_2, ..., E_N$.
Mutually exclusive $\implies E_m \cap E_n = \phi$, where $m \neq n \in \{1, 2, ..., N\}$.
The sample space $S = E_1 \cup E_2 \cup ... \cup E_N$.
We can write

$$\begin{aligned} A &= (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup ... \cup (A \cap E_N) \\ P(A) &= P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + ... + P(A \cap E_N) \\ &= P(E_1)P(A \setminus E_1) + P(E_2)P(A \setminus E_2) + ... + P(E_N)P(A \setminus E_N) \\ &= \sum_{n=1}^{N} P(A \setminus E_n)P(E_n) \end{aligned}$$

$$P(A) = \sum_{n=1}^{N} P(A \setminus E_n)P(E_n)$$

Total Probability: Problem

You have a car that sometimes doesn't start in the morning. There are only three possible causes: A_1: Battery is dead (probability = 0.3); A_2 : Fuel tank is empty (probability = 0.2); A_3: Starter motor issue (probability = 0.5). Let us say the Probability the car won't start given each cause:

- $P(No\ Start \setminus A_1) = 0.9$
- $P(No\ Start \setminus A_2) = 0.8$
- $P(No\ Start \setminus A_3) = 0.4$

Question: What is the probability that the car won't start tomorrow?

$$\begin{aligned} P(No\ Start) &= \sum_{i=1}^{3} P(No\ Start \setminus A_i) P(A_i) \\ &= (0.9)(0.3) + (0.8)(0.2) + (0.4)(0.5) = 0.63 \end{aligned}$$

Why Bayes' Theorem

Revise beliefs with new evidence

- Bayes' Theorem helps us update probabilities when new information becomes available.
- It's used in AI, medicine, forensics, finance, and robotics
  - Naive Bayes is used in classification
- Medical Diagnosis: Update disease probability after a test result
- Weather Forecasting: If it's cloudy, what's the chance it'll rain?
  - Bayes helps you revise your forecast as new conditions (e.g., humidity, temperature) arrive.

"Bayes' Theorem is how smart systems — and people — get smarter over time."

Bayes' Theorem

Consider and event 'A' of the sample space 'S'. Let $E_1, E_2, ..., E_N$ be N mutually exclusive and exhaustive events associated with the sample space. Then for any event A we have

$$P(E_n \setminus A) = \frac{P(E_n)P(A \setminus E_n)}{P(A)} = \frac{P(E_n)P(A \setminus E_n)}{\sum_{n=1}^{N} P(E_n)P(A \setminus E_n)}$$

Proof: From Conditional probability Eqn(1), we have

$$P(E_n \setminus A) = \frac{P(A \cap E_n)}{P(A)}$$

Bayes' Theorem

Proof: From conditional probability

$$P(E_n \setminus A) = \frac{P(A \cap E_n)}{P(A)}; \text{ where } P(A) \neq 0$$

Similarly

$$P(A \setminus E_n) = \frac{P(A \cap E_n)}{P(E_n)}; \text{ where } P(E_n) \neq 0$$

Equate $P(A \cap E_n)$ in the above two equations. We have

$$P(E_n \setminus A) = \frac{P(A \setminus E_n)P(E_n)}{P(A)}$$

Using law of total probability (eqn(2))

$$P(E_n \setminus A) = \frac{P(A \setminus E_n)P(E_n)}{\sum_{n=1}^{N} P(E_n)P(A \setminus E_n)}$$

Independent Events

Two events A and B are *statistically independent* if the probability of occurrence of one event is not affected by the occurrence of the other event.

$$P(A \setminus B) = P(A) \qquad or \qquad P(B \setminus A) = P(B) \tag{3}$$

From conditional probability

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} \tag{4}$$

Substitute eqn(3) in the above equation. We get

$$P(A) = \frac{P(A \cap B)}{P(B)} \implies P(A \cap B) = P(A)P(B); \tag{5}$$

Two events A and B are independent if

$$P(A \cap B) = P(A)P(B) \tag{6}$$

Multiple Independent Events

Three events A_1, A_2 and A_3 are independent if they are independent as pair and independent as a triple. They must satisfy the following equations

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned}$$

Multiplication Theorem on Probability

If $A_1, A_2, \cdots, A_n$ be n events associated with an experiment then

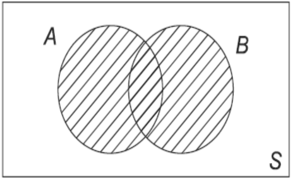
$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2 \setminus A_1)P(A_3 \setminus A_1 \cap A_2) \cdots P(A_n \setminus \cap_{i=1}^{n-1} A_i)$$

How:?

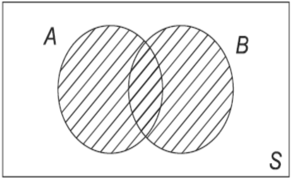
Let us write

$$P(\cap_{i=1}^{n} A_i) = P(A_1) \underbrace{\frac{P(A_1 \cap A_2)}{P(A_1)}}_{P(A_2 \setminus A_1)} \underbrace{\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}}_{P(A_3 \setminus A_1 \cap A_2)} \cdots \underbrace{\frac{P(\cap_{i=1}^{n} A_i)}{P(\cap_{i=1}^{n-1} A_i)}}_{P(A_n \setminus \cap_{i=1}^{n-1} A_i)}$$

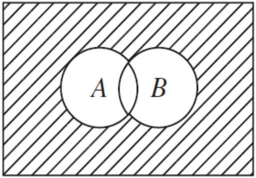
Set Operations



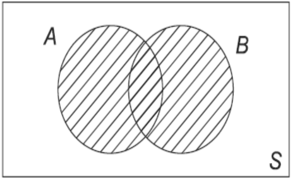
Set Operations



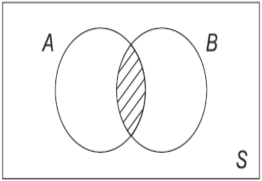
$A \cup B$

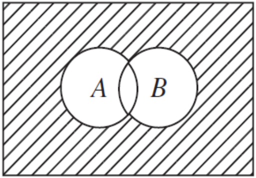


Set Operations



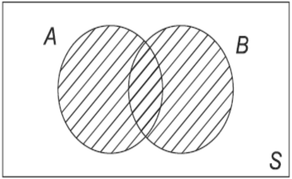
$A \cup B$



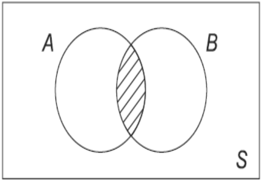


$\overline{A \cup B}$

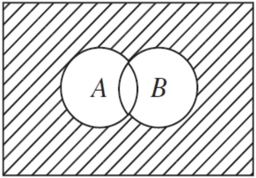
Set Operations



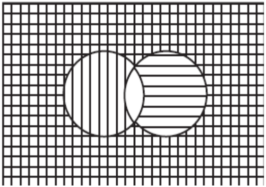
$A \cup B$



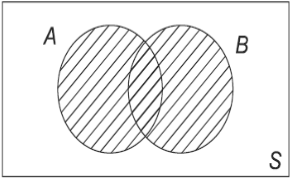
$A \cap B$



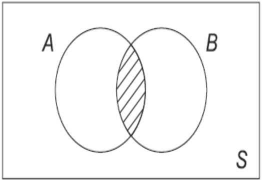
$\overline{A \cup B}$



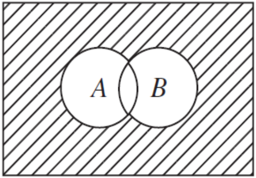
Set Operations



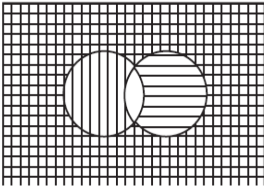
$A \cup B$



$A \cap B$

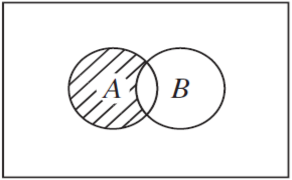


$\overline{A \cup B}$

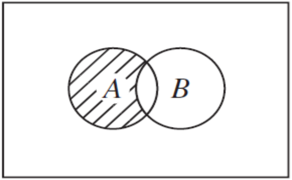


$\overline{A} \cup \overline{B}$

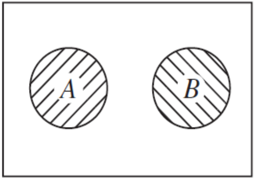
Set Operations

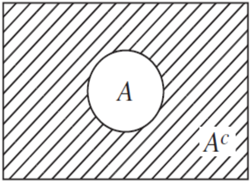


Set Operations



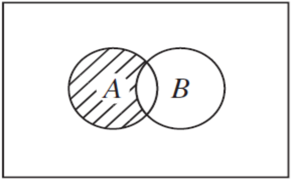
$A - B$



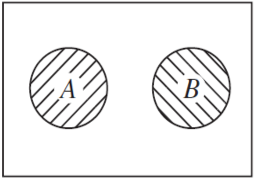


$\overline{A}$

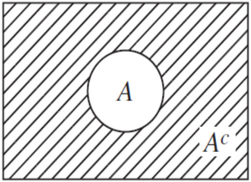
Set Operations



$A - B$



$A \cap B = \phi$



$\overline{A}$

Acknowledge various sources for the images.
Thankyou