Executive Certificate Program in Machine Learning & Artificial Intelligence

NEENA PANDEY

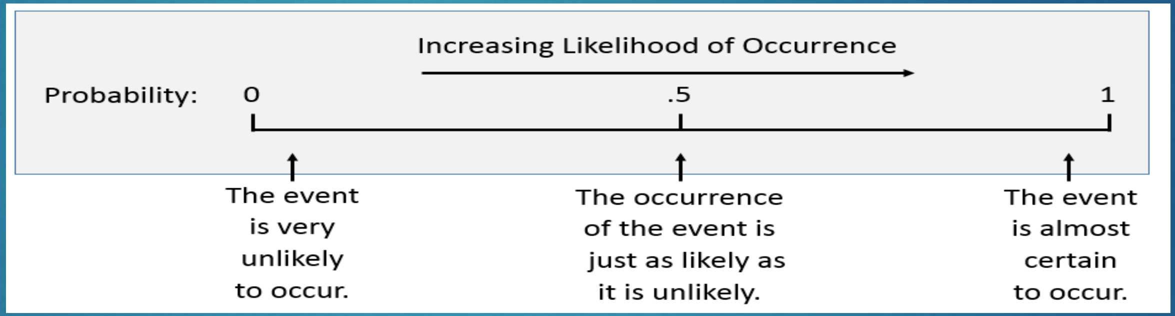
IIM VISAKHAPATNAM

Module 1: Fundamentals of Statistical Learning

- Probability
  - Probability (Basic & Conditional), Bayes' Theorem
  - Probability distributions
- Hypothesis Testing

Probability & Random Expriments

- What is probability?
  - Numerical measure of the likelihood that an event will occur.



- Statistical experiments versus Physical sciences experiments

Business Use Cases

- Managers often base their decisions on an analysis of uncertainties such as the following
  - What are the *chances* that the sales will decrease if we increase prices?
  - What is the *likelihood* a new assembly method will increase productivity?
  - What are the *odds* that a new investment will be profitable?

Assigning Probabilities

Basic Requirements for Assigning Probabilities

1. The probability assigned to each experimental outcome must be between 0 and 1, inclusively.

$$0 \leq P(E_i) \leq 1 \text{ for all } i$$

where E_i is the i th experimental outcome and $P(E_i)$ is its probability

2. The sum of the probabilities for all experimental outcomes must equal 1.

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1$$

where n is the number of experimental outcomes.

Assigning Probabilities

- Classical Method
  - Assigning probabilities based on the assumption of equally likely outcomes; E.g., Rolling a die
- Relative Frequency Method
  - Assigning probabilities based on experimental or historical data; E.g., No. of deliveries on a particular day in a week
- Subjective Method
  - Assigning probability based on judgment.

The best probability estimates often are obtained by combining the estimates from the classical or relative frequency approach with the subjective estimate.

Some more Probability Concepts

- Complement of an Event
- Addition Law – Union of Two Events

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

- Mutually Exclusive events
- Conditional Probability

Conditional Probability

- The probability of an event given that another event has occurred is called a conditional probability.
- The conditional probability of *A* given B has already occurred denoted by $P(A \mid B)$.
- A conditional probability is computed as follows: $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- Examples? Business Use Cases?

Independent events

- Can two events with non-zero probability be both mutually exclusive and independent?
- Mutually exclusive events are always dependent events
- Non-mutually exclusive events may or may not be independent
- Independent events:

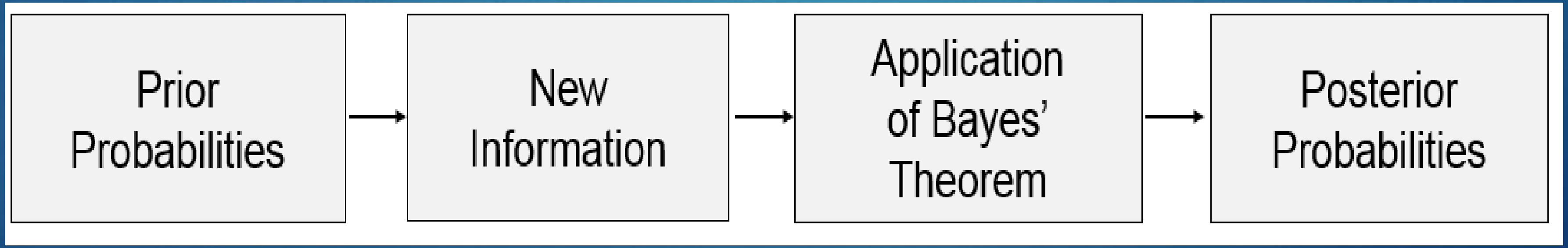
$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

Questions

- Question 1
  - P(A) = 0.30; p(B) = 0.40; A and B are mutually exclusive;
  - p(A Intersection B) = ?
  - P(A | B)= ?

Bayes' Theorem

- Probability Analysis begins with 'prior probabilities'
- Then, additional information is obtained – from a sample, product test, report
- Using this information, 'posterior probabilities' are calculated – using Bayes' Theorem – A means to revise prior probabilities



Bayes' Theorem: Applications

- Use case
  - Impact of marketing approaches

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

- Bayes' theorem is applicable when the events for which we want to compute posterior probabilities are mutually exclusive and collectively exhaustive

Bayes' Theorem: Applications

- Question 1
  - Two suppliers (A1 & A2)
  - 65% supplies from A1 & 35% from A2
  - $P(G|A1) = 0.98$ & $p(G|A2) = 0.95$
  - Question is: A bad part is detected, what is the probability that it came from A1 and what is the probability that it came from A2?
- Solution
  - If a part is chosen at random, $p(A1) = 0.65$; $p(A2) = 0.35$
  - $P(A1|B) = p(A1)*p(B|A1)/(p(A1)*p(B|A1) + p(A2)*p(B|A2))$

Random Variables & Probability Distribution

- Random variables and probability distributions are models for populations of data
- Random Variable – numerical description of the outcome of an experiment
- Discrete & Continuous Random variables
- Discrete Random variables
  - X = the number of people arriving in a bank branch in a day
- Continuous Random variables
  - X = time between customer arrivals in a minute, at a bank
  - X = percentage of project complete after six months

Probability Distributions - Discrete

- Probability Distribution
  - How are probabilities distributed over the values of the random variable
  - Prob Distribution of number obtained on one roll of a die

Number obtained	x	1	2	3	4	5	6
Probability of x	f(x)	1/6	1/6	1/6	1/6	1/6	1/6

  - Prob distribution of the no of automobiles sold during a day at a dealer

x	0	1	2	3	4	5
f(x)	0.18	0.39	0.24	0.14	0.04	0.01

Probability Distributions - Discrete

- Discrete Uniform Probability Distribution
- Discrete Uniform Probability Function $f(x) = 1/n$

where: n = the number of values the random variable may assume

The values of the random variable are equally likely.

- Binomial Probability Distribution
  - 'n' identical independent trials, two possible outcomes in each trial
  - The probability of both events does not change from trial to trial
  - Trials are independent
  - Example: Showing an advertisement to 100 families, each either purchases a policy or not; Families are selected randomly (independent trials)

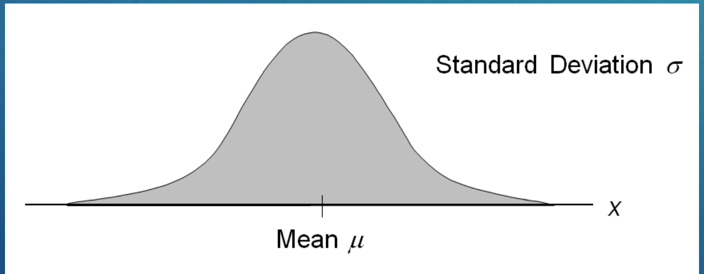
Probability Distributions - Continuous

- Continuous Uniform Probability Distribution
  - Probability is proportional to the interval's length
  - a, b: smallest, largest value the variable can assume
  - E.g., Flight time of an airplane traveling from Delhi to Mumbai
  - E.g., Battery life of iPad Mini is uniformly distributed between 8.5 and 12 hours – what's the probability that the battery life will be more than 11 hours?

$$f(x) = 1/(b - a) \quad \text{for } a \leq x \leq b$$
$$= 0 \quad \text{elsewhere}$$

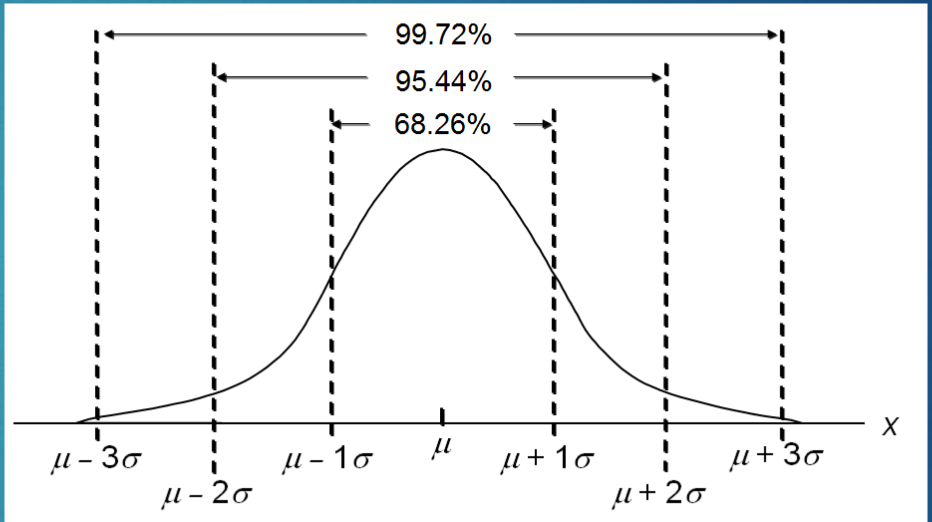
- Normal Probability Distribution

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



Normal Probability Distribution

- **Empirical Rule**
- 68.26% of values of a normal random variable are within ± 1 standard deviation of its mean.
- 95.44% of values of a normal random variable are within ± 2 standard deviations of its mean.
- 99.72% of values of a normal random variable are within ± 3 standard deviations of its mean.



Hypothesis Testing

- Performed to determine if a statement about the value of a population parameter should or should not be rejected
  - Null Hypothesis: H_0: tentative assumption about a population parameter
  - Alternate Hypothesis: H_1: Opposite of Null Hypothesis
- Developing Null & Alternate Hypothesis is very crucial
- Type 1 Error
  - Rejecting Null Hypothesis when it is true
  - The probability of Type 1 error – Level of significance
- Type II Error
  - Accepting Null Hypothesis when it is false

Type 1 & Type 11 Errors

	Population Condition	
Conclusion	H_0 True	H_0 False
Accept H_0	Correct Conclusion	Type II Error
Reject H_0	Type 1 Error	Correct Conclusion

The Importance of p-value

- P-value: Support provided by the sample for the Null hypothesis
- If p-value is less than or equal to the level of significance (α Type 1 error), the value of the test statistic is in the rejection region
  - Reject H_0(β=0), if the p-value <= α
- Guidelines for interpreting p-values
  - Less than 0.01: Overwhelming evidence to conclude H_1 is true.
  - Between 0.01 & 0.05: Strong evidence to conclude H_1 is true.
  - Between .05 and .10: Weak evidence to conclude H_1 is true.
  - Greater than .10: Insufficient evidence to conclude H_1 is true

Example of Linear Regression (for p-value)

- We first check the distributions of the data collected (both dependent and independent variables)
- We check assumptions for setting a linear regression model
- On running linear regression, you get a model – coefficients for independent variable
- Then you need to check the significance of those coefficients – and hence we need the p-value
- Compare the p-value with the level of significance you have decided
  - A low p-value indicates that coefficient is likely not equal to zero (i.e., Reject H0)
  - A high p-value means we cannot conclude that the independent variable affects the dependent variable