STATISTICAL MODELLING

Contents

RANDOM PHENOMENA AND PROBABILITY

Random Phenomena

- Deterministic phenomenon: Phenomenon whose outcome can be predicted with a very high degree of confidence
  - Example: Age of a person (using date of birth stated in Aadhaar card)
- Stochastic phenomenon: Phenomenon which can have many possible outcomes for same experimental conditions. Outcome can be predicted with limited confidence
  - Example: Outcome of a coin toss

Characterizing random phenomena

- Sources of error in observed outcomes
  - Lack of knowledge of generating process (model error)
  - Errors in sensors used for observing outcomes (measurement error)
- Types of random phenomena
  - Discrete: Outcomes are finite
    - Coin toss : {H, T}
    - Throw of a dice : {1, 2, 3, 4, 5, 6}
  - Continuous: Infinite number of outcomes
    - Body temperature measurement in deg F

Sample space, events (discrete phenomena)

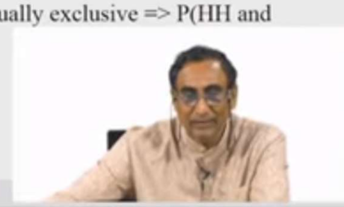
- Sample space
  - Set of all possible outcomes of a random phenomenon
    - Coin Toss : S = {H, T}
    - Two coin tosses: S = {HH, HT, TH, TT}
- Event
  - Subset of the sample space
    - Occurrence of a head in first toss of a two coin toss experiment A = {HH, HT}
    - Outcomes of a sample space are elementary events

Probability Measure

- Probability measure is a function that assigns a real value to every outcome of a random phenomena which satisfies following axioms
  - $0 \leq P(A) \leq 1$ (Probabilities are non-negative and less than 1 for any event A)
  - $P(S) = 1$ (one of the outcomes should occur)
  - For two mutually exclusive events A and B
    - $P(A \cup B) = P(A) + P(B)$
- Interpretation of probability as a frequency :
  - Conduct an experiment (coin toss) N times. If N_A is number of times outcome A occurs then $P(A) = N_A/N$

Exclusive and Independent Events

- Independent events
  - Two events are independent if occurrence of one has no influence on occurrence of other
    - Formally A and B are independent events if and only if $P(A \cap B) = P(A) \times P(B)$
    - In a two coin toss experiment, the occurrence of head in second toss can be assumed to be independent of occurrence of head or tail in first toss, then P(HH) = P(H in first toss) × P(H in second toss) = 0.5 × 0.5 = 0.25
- Mutually exclusive events
  - Two events are mutually exclusive if occurrence of one implies other event does not occur
    - In a two coin toss experiment, events {HH} and {HT} are mutually exclusive => P(HH and HT) = P(HH) + P(HT) = 0.25 + 0.25 = 0.5



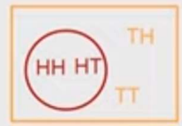
Some rules of probability

- Following important probability rules can be proved using Venn diagrams

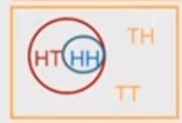
S = ▭ A = ◯ B = ◯

All outcomes are equally likely

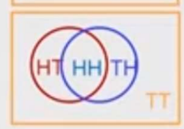
If A^c is the complement of event A^c,
$P(A^C) = P(S) - P(A) = 1 - P(A) = 0.5$



If $B \subseteq A$, $P(B) \leq P(A)$; $0.25 < 0.5$



$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.5 + 0.5 - 0.5*0.5 = 0.75$



Conditional Probability

- If two events A and B are not independent, then information available about the outcome of event A can influence the predictability of event B
- Conditional probability
  - $P(B \mid A) = P(A \cap B)/P(A)$ if $P(A) > 0$
  - $P(A \mid B)P(B) = P(B \mid A)P(A)$ - Bayes formula
  - $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$
- Example: two (fair) coin toss experiment
  - Event A : First toss is head = {HT, HH}
  - Event B : Two successive heads = {HH}
  - $Pr(B) = 0.25$ (no information)
  - Given event A has occurred $Pr(B|A) = 0.5 = 0.25/0.5 = P(A \cap B)/P(A)$

Example

In a manufacturing process 1000 parts are produced of which 50 are defective. We randomly take a part from the day's production

- Outcomes : {A=Defective part B = Non-defective part}
- P(A) = 50/1000, P(B) = 950/1000

- Suppose we draw a second part without replacing the first part
  - Outcomes : {C = Defective part D = Non-defective part}
  - Pr(C) = 50/1000 (no information about outcome of first draw)
  - P(C | A) = 49/999 (given information that first draw is defective)
  - Pr(C | B) = 50/999 (given information that first draw is non-defective)
  - P(C) = 49/999*50/1000 + 50/999*950/1000 = 50/1000
  - $P(A \mid C) = P(A \cap C)/P(C) = P(C \mid A)P(A)/P(C) = 49/999$