**ASSIGNMENT 10**

**Q1.DEFINE THE BAYESIAN INTERPRETATION OF PROBABILITY.**

**ANS.** THE BAYESIAN INTERPRETATION OF PROBABILITY IS A PHILOSOPHICAL AND MATHEMATICAL FRAMEWORK THAT VIEWS PROBABILITY AS A MEASURE OF SUBJECTIVE BELIEF OR DEGREE OF UNCERTAINTY RATHER THAN A FREQUENCY-BASED CONCEPT. IT INCORPORATES PRIOR KNOWLEDGE OR BELIEFS ABOUT AN EVENT OR HYPOTHESIS AND UPDATES THEM BASED ON NEW EVIDENCE USING BAYES' THEOREM.

**Q2.DEFINE PROBABILITY OF A UNION OF TWO EVENTS WITH EQUATION.**

**ANS.** THE PROBABILITY OF THE UNION OF TWO EVENTS, DENOTED AS A AND B, IS THE PROBABILITY THAT AT LEAST ONE OF THE EVENTS A OR B OCCURS. IT CAN BE CALCULATED USING THE EQUATION:

P(A ∪ B) = P(A) + P(B) - P(A ∩ B),

WHERE:

- P(A ∪ B) REPRESENTS THE PROBABILITY OF THE UNION OF EVENTS A AND B.

- P(A) IS THE PROBABILITY OF EVENT A OCCURRING.

- P(B) IS THE PROBABILITY OF EVENT B OCCURRING.

- P(A ∩ B) IS THE PROBABILITY OF THE INTERSECTION OF EVENTS A AND B, I.E., THE PROBABILITY THAT BOTH EVENTS A AND B OCCUR SIMULTANEOUSLY.

**Q3.WHAT IS JOINT PROBABILITY? WHAT IS ITS FORMULA?**

**ANS.** JOINT PROBABILITY REFERS TO THE PROBABILITY OF TWO OR MORE EVENTS OCCURRING SIMULTANEOUSLY. IT MEASURES THE LIKELIHOOD OF THE INTERSECTION OF MULTIPLE EVENTS HAPPENING TOGETHER. THE JOINT PROBABILITY IS DENOTED AS P(A AND B), REPRESENTING THE PROBABILITY OF BOTH EVENT A AND EVENT B OCCURRING.

THE FORMULA FOR CALCULATING THE JOINT PROBABILITY OF TWO EVENTS A AND B DEPENDS ON WHETHER THE EVENTS ARE INDEPENDENT OR DEPENDENT:

1. INDEPENDENT EVENTS:

IF EVENTS A AND B ARE INDEPENDENT, MEANING THAT THE OCCURRENCE OF ONE EVENT DOES NOT AFFECT THE PROBABILITY OF THE OTHER EVENT, THE JOINT PROBABILITY CAN BE CALCULATED USING THE PRODUCT OF THEIR INDIVIDUAL PROBABILITIES:

P(A AND B) = P(A) \* P(B)

2. DEPENDENT EVENTS:

IF EVENTS A AND B ARE DEPENDENT, MEANING THAT THE OCCURRENCE OF ONE EVENT AFFECTS THE PROBABILITY OF THE OTHER EVENT, THE JOINT PROBABILITY CAN BE CALCULATED USING THE CONDITIONAL PROBABILITY:

P(A AND B) = P(A | B) \* P(B)

HERE, P(A | B) REPRESENTS THE CONDITIONAL PROBABILITY OF EVENT A GIVEN THAT EVENT B HAS OCCURRED, AND P(B) IS THE PROBABILITY OF EVENT B OCCURRING.

**Q4.WHAT IS CHAIN RULE OF PROBABILITY?**

**ANS.** THE CHAIN RULE OF PROBABILITY, ALSO KNOWN AS THE MULTIPLICATION RULE, IS A FUNDAMENTAL PRINCIPLE IN PROBABILITY THEORY THAT ALLOWS US TO CALCULATE THE PROBABILITY OF THE INTERSECTION OF MULTIPLE EVENTS. IT PROVIDES A WAY TO DECOMPOSE THE JOINT PROBABILITY OF MULTIPLE EVENTS INTO A PRODUCT OF CONDITIONAL PROBABILITIES.

THE CHAIN RULE STATES THAT THE PROBABILITY OF THE INTERSECTION OF SEVERAL EVENTS CAN BE CALCULATED BY MULTIPLYING THE CONDITIONAL PROBABILITIES OF EACH EVENT GIVEN THE PREVIOUS EVENTS. MATHEMATICALLY, FOR EVENTS A₁, A₂, A₃, ..., Aₙ, THE CHAIN RULE IS EXPRESSED AS:

**P(A₁ ∩ A₂ ∩ A₃ ∩ ... ∩ Aₙ) = P(A₁) \* P(A₂ | A₁) \* P(A₃ | A₁ ∩ A₂) \* ... \* P(Aₙ | A₁ ∩ A₂ ∩ A₃ ∩ ... ∩ Aₙ₋₁)**

IN THIS FORMULA, P(A₁) REPRESENTS THE PROBABILITY OF THE FIRST EVENT A₁, AND EACH SUBSEQUENT TERM P(Aᵢ | A₁ ∩ A₂ ∩ A₃ ∩ ... ∩ Aᵢ₋₁) REPRESENTS THE CONDITIONAL PROBABILITY OF EVENT Aᵢ GIVEN THAT ALL THE PREVIOUS EVENTS HAVE OCCURRED.

**Q5.WHAT IS CONDITIONAL PROBABILITY MEANS? WHAT IS THE FORMULA OF IT?**

**ANS.** CONDITIONAL PROBABILITY IS A MEASURE OF THE PROBABILITY OF AN EVENT OCCURRING GIVEN THAT ANOTHER EVENT HAS ALREADY OCCURRED. IT QUANTIFIES THE LIKELIHOOD OF AN EVENT HAPPENING UNDER A SPECIFIC CONDITION OR CONTEXT.

THE CONDITIONAL PROBABILITY OF AN EVENT A GIVEN AN EVENT B IS DENOTED AS P(A | B), READ AS "THE PROBABILITY OF A GIVEN B." IT REPRESENTS THE PROBABILITY OF EVENT A OCCURRING, ASSUMING THAT EVENT B HAS ALREADY OCCURRED.

THE FORMULA FOR CALCULATING CONDITIONAL PROBABILITY IS:

**P(A | B) = P(A ∩ B) / P(B),**

WHERE:

- P(A | B) REPRESENTS THE CONDITIONAL PROBABILITY OF EVENT A GIVEN EVENT B.

- P(A ∩ B) IS THE JOINT PROBABILITY OF EVENTS A AND B OCCURRING SIMULTANEOUSLY.

- P(B) IS THE PROBABILITY OF EVENT B OCCURRING.

**Q6.WHAT ARE CONTINUOUS RANDOM VARIABLES?**

**ANS.** CONTINUOUS RANDOM VARIABLES ARE VARIABLES THAT CAN TAKE ON ANY VALUE WITHIN A SPECIFIED RANGE OR INTERVAL. UNLIKE DISCRETE RANDOM VARIABLES, WHICH CAN ONLY TAKE ON SPECIFIC, DISTINCT VALUES, CONTINUOUS RANDOM VARIABLES HAVE AN INFINITE NUMBER OF POSSIBLE OUTCOMES WITHIN A CONTINUOUS RANGE.

**Q7.WHAT ARE BERNOULLI DISTRIBUTIONS? WHAT IS THE FORMULA OF IT?**

**ANS.** THE BERNOULLI DISTRIBUTION IS A DISCRETE PROBABILITY DISTRIBUTION THAT MODELS A SINGLE BINARY RANDOM VARIABLE. IT REPRESENTS AN EXPERIMENT WITH TWO POSSIBLE OUTCOMES, OFTEN REFERRED TO AS SUCCESS AND FAILURE, WITH PROBABILITIES OF P AND Q = 1 - P, RESPECTIVELY.

THE PROBABILITY MASS FUNCTION (PMF) OF THE BERNOULLI DISTRIBUTION IS GIVEN BY:

P(X = X) = P^X \* Q^(1-X),

WHERE:

- P(X = X) REPRESENTS THE PROBABILITY THAT THE RANDOM VARIABLE X TAKES THE VALUE X.

- X IS THE OUTCOME OF THE RANDOM VARIABLE, WHICH CAN BE EITHER 0 OR 1.

- P IS THE PROBABILITY OF SUCCESS, I.E., THE PROBABILITY THAT X = 1.

- Q = 1 - P IS THE PROBABILITY OF FAILURE, I.E., THE PROBABILITY THAT X = 0.

**Q8.WHAT IS BINOMIAL DISTRIBUTION? WHAT IS THE FORMULA?**

**ANS.** THE BINOMIAL DISTRIBUTION IS A DISCRETE PROBABILITY DISTRIBUTION THAT MODELS THE NUMBER OF SUCCESSES IN A FIXED NUMBER OF INDEPENDENT BERNOULLI TRIALS. IT DESCRIBES THE PROBABILITY OF OBTAINING A SPECIFIC NUMBER OF SUCCESSES (K) IN A GIVEN NUMBER OF TRIALS (N), WHERE EACH TRIAL HAS TWO POSSIBLE OUTCOMES (SUCCESS OR FAILURE) AND THE PROBABILITY OF SUCCESS (P) REMAINS CONSTANT ACROSS ALL TRIALS.

THE PROBABILITY MASS FUNCTION (PMF) OF THE BINOMIAL DISTRIBUTION IS GIVEN BY THE FORMULA:

**P(X = K) = C(N, K) \* P^K \* (1 - P)^(N - K),**

WHERE:

- P(X = K) REPRESENTS THE PROBABILITY THAT THE RANDOM VARIABLE X TAKES THE VALUE K, I.E., THE PROBABILITY OF GETTING K SUCCESSES IN N TRIALS.

- C(N, K) IS THE BINOMIAL COEFFICIENT, ALSO KNOWN AS THE NUMBER OF COMBINATIONS OR "N CHOOSE K," WHICH REPRESENTS THE NUMBER OF WAYS TO CHOOSE K SUCCESSES FROM N TRIALS. IT IS CALCULATED AS C(N, K) = N! / (K! \* (N - K)!), WHERE ! DENOTES THE FACTORIAL FUNCTION.

- P IS THE PROBABILITY OF SUCCESS IN A SINGLE TRIAL.

- (1 - P) IS THE PROBABILITY OF FAILURE IN A SINGLE TRIAL.

- K IS THE NUMBER OF SUCCESSES.

- N IS THE TOTAL NUMBER OF TRIALS.

**Q9.WHAT IS POISSON DISTRIBUTION? WHAT IS THE FORMULA?**

**ANS.** THE POISSON DISTRIBUTION IS A DISCRETE PROBABILITY DISTRIBUTION THAT MODELS THE NUMBER OF EVENTS THAT OCCUR IN A FIXED INTERVAL OF TIME OR SPACE WHEN THE EVENTS ARE RARE AND INDEPENDENT OF EACH OTHER. IT IS OFTEN USED TO DESCRIBE THE OCCURRENCE OF RARE EVENTS SUCH AS ACCIDENTS, PHONE CALLS, OR CUSTOMER ARRIVALS.

THE PROBABILITY MASS FUNCTION (PMF) OF THE POISSON DISTRIBUTION IS GIVEN BY THE FORMULA:

**P(X = K) = (E^(-Λ) \* Λ^K) / K!,**

WHERE:

- P(X = K) REPRESENTS THE PROBABILITY THAT THE RANDOM VARIABLE X TAKES THE VALUE K, I.E., THE PROBABILITY OF K EVENTS OCCURRING.

- E IS THE BASE OF THE NATURAL LOGARITHM (APPROXIMATELY 2.71828).

- Λ (LAMBDA) IS THE AVERAGE RATE OR INTENSITY OF THE EVENTS OCCURRING IN THE GIVEN INTERVAL. IT REPRESENTS THE EXPECTED NUMBER OF EVENTS IN THE INTERVAL.

- K IS THE NUMBER OF EVENTS OCCURRING (CAN BE ANY NON-NEGATIVE INTEGER).

- K! DENOTES THE FACTORIAL OF K.

**Q10.DEFINE COVARIANCE.**

**ANS.** COVARIANCE IS A STATISTICAL MEASURE THAT QUANTIFIES THE RELATIONSHIP BETWEEN TWO VARIABLES. IT MEASURES HOW CHANGES IN ONE VARIABLE ARE ASSOCIATED WITH CHANGES IN ANOTHER VARIABLE. IN OTHER WORDS, IT PROVIDES AN INDICATION OF THE LINEAR DEPENDENCY BETWEEN TWO RANDOM VARIABLES.

MATHEMATICALLY, THE COVARIANCE BETWEEN TWO RANDOM VARIABLES X AND Y IS DEFINED AS:

**COV(X, Y) = E[(X - ΜX)(Y - ΜY)],**

WHERE:

- COV(X, Y) REPRESENTS THE COVARIANCE BETWEEN VARIABLES X AND Y.

- E DENOTES THE EXPECTATION OPERATOR.

- X AND Y ARE RANDOM VARIABLES.

- ΜX AND ΜY ARE THE MEANS (OR EXPECTED VALUES) OF X AND Y, RESPECTIVELY.

**Q11.DEFINE CORRELATION**

**ANS.** CORRELATION IS A STATISTICAL MEASURE THAT QUANTIFIES THE STRENGTH AND DIRECTION OF THE LINEAR RELATIONSHIP BETWEEN TWO VARIABLES. IT IS USED TO DETERMINE HOW CLOSELY THE VALUES OF TWO VARIABLES MOVE TOGETHER.

THE CORRELATION COEFFICIENT, DENOTED BY THE SYMBOL "R," REPRESENTS THE CORRELATION BETWEEN TWO VARIABLES X AND Y. IT CAN TAKE VALUES BETWEEN -1 AND 1:

- A CORRELATION COEFFICIENT OF 1 INDICATES A PERFECT POSITIVE LINEAR RELATIONSHIP, MEANING THAT AS ONE VARIABLE INCREASES, THE OTHER VARIABLE INCREASES PROPORTIONALLY.

- A CORRELATION COEFFICIENT OF -1 INDICATES A PERFECT NEGATIVE LINEAR RELATIONSHIP, MEANING THAT AS ONE VARIABLE INCREASES, THE OTHER VARIABLE DECREASES PROPORTIONALLY.

- A CORRELATION COEFFICIENT OF 0 INDICATES NO LINEAR RELATIONSHIP BETWEEN THE VARIABLES. HOWEVER, IT DOES NOT IMPLY THAT THERE IS NO RELATIONSHIP AT ALL BETWEEN THE VARIABLES, AS THEY MIGHT STILL HAVE A NONLINEAR OR NON-MONOTONIC RELATIONSHIP.

THE FORMULA FOR CALCULATING THE CORRELATION COEFFICIENT IS:

**R = COV(X, Y) / (ΣX \* ΣY),**

WHERE:

- COV(X, Y) IS THE COVARIANCE BETWEEN VARIABLES X AND Y.

- ΣX AND ΣY ARE THE STANDARD DEVIATIONS OF VARIABLES X AND Y, RESPECTIVELY.

**Q12.DEFINE SAMPLING WITH REPLACEMENT. GIVE EXAMPLE.**

**ANS.** SAMPLING WITH REPLACEMENT IS A SAMPLING METHOD IN STATISTICS AND PROBABILITY THEORY WHERE EACH SELECTED ITEM FROM A POPULATION IS RETURNED TO THE POPULATION BEFORE THE NEXT SELECTION IS MADE. THIS MEANS THAT AN ITEM CAN BE SELECTED MORE THAN ONCE IN THE SAMPLING PROCESS.

**HERE'S AN EXAMPLE TO ILLUSTRATE SAMPLING WITH REPLACEMENT:**

SUPPOSE WE HAVE A BAG CONTAINING FIVE COLORED BALLS: RED, BLUE, GREEN, YELLOW, AND ORANGE. WE WANT TO RANDOMLY SELECT THREE BALLS FROM THE BAG USING SAMPLING WITH REPLACEMENT.

1. WE RANDOMLY SELECT A BALL FROM THE BAG, LET'S SAY IT'S BLUE. WE NOTE DOWN THE COLOR AND RETURN THE BLUE BALL BACK INTO THE BAG.

2. WE AGAIN RANDOMLY SELECT A BALL FROM THE BAG, AND THIS TIME WE GET RED. WE NOTE DOWN THE COLOR AND RETURN THE RED BALL BACK INTO THE BAG.

3. ONCE MORE, WE RANDOMLY SELECT A BALL FROM THE BAG, AND LET'S SAY WE GET BLUE AGAIN. WE NOTE DOWN THE COLOR AND RETURN THE BLUE BALL BACK INTO THE BAG.

**Q13.WHAT IS SAMPLING WITHOUT REPLACEMENT? GIVE EXAMPLE.**

**ANS.** SAMPLING WITHOUT REPLACEMENT IS A SAMPLING METHOD IN STATISTICS AND PROBABILITY THEORY WHERE EACH SELECTED ITEM FROM A POPULATION IS NOT RETURNED TO THE POPULATION BEFORE THE NEXT SELECTION IS MADE. THIS MEANS THAT ONCE AN ITEM IS SELECTED, IT IS REMOVED FROM THE POPULATION AND CANNOT BE SELECTED AGAIN IN SUBSEQUENT DRAWS.

**HERE'S AN EXAMPLE TO ILLUSTRATE SAMPLING WITHOUT REPLACEMENT:**

SUPPOSE WE HAVE A DECK OF PLAYING CARDS CONSISTING OF 52 CARDS, INCLUDING FOUR SUITS: HEARTS, DIAMONDS, CLUBS, AND SPADES. WE WANT TO RANDOMLY SELECT THREE CARDS FROM THE DECK USING SAMPLING WITHOUT REPLACEMENT.

1. WE SHUFFLE THE DECK TO ENSURE RANDOMNESS.

2. WE DRAW THE FIRST CARD FROM THE DECK, LET'S SAY IT'S THE ACE OF HEARTS. WE NOTE DOWN THE CARD AND REMOVE IT FROM THE DECK.

3. WE DRAW THE SECOND CARD FROM THE REMAINING DECK, AND LET'S SAY IT'S THE KING OF SPADES. WE NOTE DOWN THE CARD AND REMOVE IT FROM THE DECK.

4. FINALLY, WE DRAW THE THIRD CARD FROM THE REMAINING DECK, AND LET'S SAY IT'S THE 7 OF CLUBS. WE NOTE DOWN THE CARD AND REMOVE IT FROM THE DECK.

**Q14.WHAT IS HYPOTHESIS? GIVE EXAMPLE.**

**ANS.** IN STATISTICS AND RESEARCH METHODOLOGY, A HYPOTHESIS IS A STATEMENT OR PROPOSITION THAT SUGGESTS A POSSIBLE EXPLANATION OR RELATIONSHIP BETWEEN VARIABLES. IT IS A FUNDAMENTAL COMPONENT OF SCIENTIFIC INVESTIGATION AND SERVES AS A STARTING POINT FOR EMPIRICAL TESTING AND DATA ANALYSIS.

A HYPOTHESIS TYPICALLY CONSISTS OF TWO PARTS: THE NULL HYPOTHESIS (H0) AND THE ALTERNATIVE HYPOTHESIS (HA). THE NULL HYPOTHESIS ASSUMES THAT THERE IS NO SIGNIFICANT RELATIONSHIP OR EFFECT BETWEEN VARIABLES, WHILE THE ALTERNATIVE HYPOTHESIS SUGGESTS THAT THERE IS A SIGNIFICANT RELATIONSHIP OR EFFECT.

**HERE'S AN EXAMPLE TO ILLUSTRATE A HYPOTHESIS:**

LET'S SAY A RESEARCHER WANTS TO INVESTIGATE WHETHER A NEW DRUG IS EFFECTIVE IN REDUCING THE SYMPTOMS OF A CERTAIN MEDICAL CONDITION. THEY COULD FORMULATE THE FOLLOWING HYPOTHESIS:

**NULL HYPOTHESIS (H0):** THE NEW DRUG HAS NO EFFECT ON REDUCING THE SYMPTOMS OF THE MEDICAL CONDITION.

**ALTERNATIVE HYPOTHESIS (HA):** THE NEW DRUG IS EFFECTIVE IN REDUCING THE SYMPTOMS OF THE MEDICAL CONDITION.