Biological Data Analytics

1. **LTP structure** of the course: 3-1-0

2. **Course Plan:**

Component	Unit	Topics for Coverage
C 1	Unit 1	Review of the basic concepts of Probability (up to Bayes Theorem) and Statistics (Central tendencies and standard deviations)
	Unit 2	Probability Distribution functions: Binomial, Poisson and Normal distributions, Central Limit Theorem and it's applications.
C2	Unit 3	Sampling distribution, Estimation, Interval estimation, Confidence interval, Test of hypotheses, Z-test, t-test, the chi-square test, F-test and ANOVA test.
	Unit 4	Correlation and Regression analyses, Correlation Coefficients, Least square method and curve fittings, Single and multi variable regression.

3. **Text Book:** 'Biostatistics -A Foundation for Analysis in the Health Sciences' by Wayne E. Daniel and Chad L. Gross.

4. **Reference:** 'Fundamental of Biostatistics' by Bernard Rosner.

Syllabus and course structure

Grading Policy

- C1 (2 Assignments + 1 quiz + class interaction+ Review test)
- C2 (2 Assignments+ 1 quiz+ + class interaction+Review test)
- C3 (test or project)

Probability & statistics

- **Probability:** A mathematical framework. Based on laws, theorems lemmas, etc. Well organised. Generally provides a concrete result.

  Eg. Amol has 60% chance of getting A grade in Biostatistics.

- **Statistics:** Application of theory of probability on data. Does not give a concrete result. Messier.

  Eg. The mean infection rate of SARS is between 2.5-3.5 with 95% confidence.

Probability & Statistics Applications

- Determining trends in stock markets
- Weather prediction
- Data Science
- Health Science
- Epidemiology (infectious diseases like COVID-19)
- Many more…

Probability provides the foundation for statistical analyses

Classical Probability : If an experiment (or trial) has **N** mutually exclusive and equally likely outcomes, and if a particular outcome **E** can occur in **m** ways , the probability of the occurrence of **E** is equal to

$$P(E) = \frac{m}{N}$$

The probability of getting a 6 in throwing a die is 1/6.

Relative frequency probability : If some process is repeated a large number of times, N, and if some resulting event with the characteristic **E** occurs **m** times, the relative frequency of occurrence of **E**, **m/N**, will be approximately equal to the probability of **E**.

Out of 300 samples of suspected people 30 are Covid positive. So the probability of Covid positivity is 0.1.

These are examples of Objective way of defining probability.

Probability

A **sample space** (**S**) is the set of all possible outcomes of an experiment. An event in general is a subset of the power set of all distinct outcomes.

Eg: Tossing of two coins. Each toss has two possible outcomes {H} and {T}. The outcomes are {HH}, {TT}, {HT}, {TH}.

Probability law: Assigning a nonnegative number P(A) to any event A (a subset of S)

The sample space of an experiment may consist of a finite or an infinite number of possible outcomes.

Eg: Outcomes in throwing of a dice. Finite no of outcomes.

Decay probability of a radioactive material. Infinite number of outcomes.

A wheel of fortune is continuously calibrated from 0 to 1, so the possible outcomes of an experiment consisting of a single spin are the numbers in the interval **S** = [0,1]. Infinite number of possibilities.

Algebra of Sets

$V \cup T = T \cup V$ *Commutativity*

$V \cap (T \cup U) = (V \cap T) \cup (V \cap U)$ *Distributive law*

$(V^c)^c = V$, $V \cup S = S$

$V \cup (T \cup U) = (V \cup T) \cup (V \cup U)$ *Associative law* $V \cup (T \cap U) = (V \cup T) \cap (V \cup U)$

$V \cap V^c = \emptyset$; $V \cap S = V$.

De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c; \; (A \cap B)^c = A^c \cup B^c$$

Probability Axioms

- (Nonnegativity) $P(A) \geq 0$, for every event A.
- (Additivity) If A and B are two mutually exclusive (disjoint) events, then the probability of their union satisfies

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

- If the sample space has an infinite number of elements and $A_1, A_2, \ldots$ is a sequence of disjoint events, then the probability of their union satisfies

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

- (Normalization) The probability of the entire sample space S is equal to 1, that is,

$$P(S) = 1.$$

Some Properties of Probability Laws

Consider a probability law, and let A, B, and C be events.

(a) If $A \subset B$, then $P(A) \leq P(B)$

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

(c) $P(A \cup B) \leq P(A) + P(B)$

Conditional Probability

The conditional probability of **A** given B is equal to the probability of $A \cap B$ divided by the probability of B, provided the probability of B is not zero.

$$P(A|B) = \frac{P(A \cap B)}{P(B)};\ P(B) \neq 0$$

- Joint probability: $P(A \cap B)$
- For countably finite sets:

$$P(A|B) = \frac{\text{no. of elements in } A \cap B}{\text{no of elements in } B}$$

Example

- In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 10. How likely is it that the first roll was a 6?

- **Solution:** Total no of possibilities = 36

Reduced Sample space {(4,6), (5,5), (6,4)}

Answer should be= 1/3.

Event A={first roll is 6 }

Event B={sum of two rolls is 10}

$P(A \cap B)$=1/36.

P(B)=3/36

Desired probability P(A|B) = $\frac{1/36}{3/36} = 1/3$

Multiplication Rule

If A, B, C are three events then we have-

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

Proof:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \implies P(A \cap B) = P(A \mid B)P(B)$$

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

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

Total Probability Theorem

- If $A_1, A_2, \ldots\ldots, A_n$ are disjoint events and $S = \cup_{i=1}^{n} A_i$ then

  For any event B we have

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

$$= P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)$$

If the sample space is partitioned into a number of scenarios (events) A_i. Then, the probability that an event B occurs is a weighted average of its conditional probability under each scenario.

Subjective probability and Bayes' Theorem

- Sometimes probability of an event is not assigned not by repetitive experiments but it is assigned by prior knowledge or experience. This way one can even assign a probability of a single event. Eg. Probability that cure for cancer will be discovered in 5 years.

- **Bayes' Rule:** If $A_i, i = 1,2,...,n$ are set of disjoint events (non empty) and B is another event. Then the conditional probability that the i^th event A_i occur is given by

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\Sigma_{i=1}^{n} P(A_i)P(B | A_i)}$$

If $A_{i\,s}$ are causes of some effect B. Then Bayes' rule prescribes the probability of the presence of the cause A_i given the effect B has been observed.

- Independent Events: A and B are independent if $P(A \cap B) = P(A)P(B)$