Contents

1 Introduction to Probability and Statistics

This course deals with data, generally numbers, that may arise in experiments or in observations and whose values cannot be predicted with certainty in advance (**non-deterministic**). Source of uncertainty lies in the variation of conditions or incomplete knowledge. Some can be controlled and modeled. But there is always some residual variation that is left over, no matter how good the model.

Statistics: body of scientific knowledge that deals with the collection, description, analysis, and interpretation of observations or experimental results.

Probability: Body of scientific knowledge that provides the language and tools for carrying out statistics

Population: Set of individuals or objects being studied (real or hypothetical)

Experiment $\mathcal{E}$: Mechanism for selecting a set/multi set of a population. Usual assumption is that each member of the population has the same chance begin selected. Examples: randomize patients to treat, determine population decline, coin tosses

Random Variable X: Value attached to each member of a population.

Probability Distribution P: Description of all values X can take, including multiplicities. Essentially says how often values are repeated. Important because some are more prevalent than others.

Parameter θ: Characteristic of a population (or equivalently, distribution) such as an average

Sample Space S: Set of all possible outcomes of an experiment. Comprised of sample points {s}. Discrete or continuous. Finite or infinite.

Event E: A subset of a sample space

Random Sample: The subset of the population obtained via experiment

Data: Values of the random variables in the random sample. Denoted by $\{ x_1, x_2, x_3, ..., x_n\}$

Example 2.3: A communication system sends messages back and forth. The messages can wither be late or on time. There are three messages analyzed. Describe the sample space (how many outcomes are possible).

Ans. The first one can be either on time or late, the second can also be on time or late, and the third can be on time or late. Each messages can take on two different states. Thus we multiply together all the possible outcomes for each, and we get:

$$2 * 2 * 2 = 2^3 = 8$$

Example 2.4: A car company produces a car which can be customized with the following: 2 types of gear shits, 2 types of sunroofs, 3 types of stereo, and 4 types of colors. How many sample points are in S?

Ans. We use the rule of multiplying through the different options we have.

$$|S| = 2 * 2 * 3 * 4 = 48$$

Notation:

Union: $E_1 \cup E_2$ — **Intersect**: $E_1 \cap E_2$ — **Complement**: E' — **Empty Set** : $\emptyset$

DeMorgan's Laws: $(A \cup B)' = A' \cap B'$ — **Distributive Laws:** $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$