# Probability for Statistics - Lecture Notes

## Topic 1 - Probability

### 1.0 Introduction to Probability

**Statistics** concerns collection, analysis and interpretation of data. It is used for making informed decisions in all areas of science, business and government. However, it is often intentionally misused by people by finding ways to interpret the data that are to their own favour.

When the data are resulted from an observational study, there is a possibility that a third variable (called a **confounding variable**) may affect the observed relationship between the two variables of interest. One wants to know how much a confounding variable affects the observed relationship between the two variables; probability theory would be required in answering this question.

**Simpson's Paradox** is that the relationship appears to be in a different direction when the confounding variables are not considered.

**Example 1**

*Death Penalty Verdict by Defendant's Race and Victim's Race*

| Victim's Race | Defendant's Race | Death Penalty | | Percent of Yes |
|---|---|---|---|---|
| | | Yes | No | |
| White | White | 53 | 414 | 11.3% |
| | Black | 11 | 37 | 22.9% |
| Black | White | 0 | 16 | 0% |
| | Black | 4 | 139 | 2.8% |

Finding 1: for each category of victims, the percentage of black defendants receiving death penalty is higher than that for white defendants. However, if the victim's race factor is dropped, one gets the following table:

| Defendant's Race | Death Penalty | | Percent of Yes |
|---|---|---|---|
| | Yes | No | |
| White | 53 | 430 | 11.0% |
| Black | 15 | 176 | 7.9% |

Finding 2: the percentage of black defendants receiving death penalty is lower.

**Probability theory** provides a mathematics foundation for applying statistics where randomness/uncertainties are almost always involved. **Probability** refers to the quantification of randomness/uncertainty (note that probability cannot remove or reduce the randomness/uncertainty).

**Example 2**

A tennis player made 1000 first serves in last year's matches. Among these serves, 750 were successful. Suppose the player's tennis skills stay as the same this year. What is the probability that his next first serve will be successful?

Evidently, the answer is Pr(success) = 750/1000 = 0.75 (75%).

The moral of this example is that the outcome in each trial occurs by chance, but there is a statistical pattern over large number of trials. The probability here is interpreted as the **long-term relative frequency**:

$$\Pr(success) = \lim_{n \to \infty} \frac{n(successes)}{n(trials)}$$

If the probability of a successful first serve is really 0.75, then one would expect the player to succeed around 75% of the first serves in the long run.

Probability has a second interpretation as measuring the uncertainty of knowledge. Probability of Event A conditional on Event B is not the same as the probability of Event B conditional on Event A. In order to rigorously define and systematically study probability, one needs some formal notations and terminology.

### 1.01 Definitions and Properties Related to Probability

Some key definitions of probability include:

> **Experiment**: a process of obtaining an observed result.
> **Trial**: performing an experiment.
> **Outcome**: an observed result of an experiment.
> **Random experiment**: an experiment in which the outcome cannot be predicted with certainty.

A **sample space** (or outcome space), denoted by S, is the set of all possible outcomes of an experiment. Each outcome in S is called a sample point, which is often generically denoted as *w*.

**Example 3**

Let an experiment be the toss of two coins. Suppose we are interested in the observed face of each coin. Then the sample space is:

S = {HH, HT, TH, TT}

This is called a **finite sample space**. To work out the number of heads obtained in the same experiment, the sample space is:

S = {0, 1, 2}

Note that the same experiment can have different sample spaces depending on the objective of the experiment. Now, consider the toss of one coin. If we are interested in the number of tosses required to get a heads, the sample space would be:

S = {1, 2, 3, …}, the set of all natural numbers.

Note that this is called a **countable infinite sample space**.

S is called a **discrete sample space** if S is either finite or countable infinite.

**Example 4**

Observe the lifetime T of a lightbulb. The sample space is:

$$S = \{0 \leq T < \infty\}$$

This is known as a one-dimensional **continuous sample space**.

Note that different experiments can have equivalent sample spaces. For example, experiments of having 'Heads' or 'Tails', 'raining' or 'not raining' and 'male' or 'female' can all have the same sample space S = {yes, no}, where 'yes' and 'no' have different interpretations for each experiment.

A is called an **event** if A is a subset of the outcomes in the sample space S. One says that A has occurred if one performs a random experiment and the outcome of the experiment is in A.

**Example 5**

There are 4 white marbles and 2 black marbles in a bag. Now take out two marbles at random.

The sample space is S = {WB, WW, BB, BW}. Each of the 4 sample points constitutes an event, i.e. $A_1$ = {(W,W,)}, $A_2$ = {(W,B)}, $A_3$ = {(B,W)}, $A_4$ = {(B,B)} are all events. The following are also events:

1. $E_1$ = {the first draw is a black} = {(B,B), (B,W)}
2. $E_2$ = {the second draw is a black} = {(B,B), (W,B)}
3. $E_3$ = {exactly one draw is a black} = {(B,W), (W,B)}

The differences between $E_1$, $E_2$, $E_3$ and $A_1$, $A_2$, $A_3$, $A_4$ is that the former ones are decomposable (contains more than one), while the latter ones are elementary events.

An event is called an **elementary event** if it contains exactly one outcome of the experiment. There are two special events: the empty set ø is called the **null event** (impossible event). It cannot happen at any time. The sample space S itself is called the **certain event** which always happens.

For operations of events:

1. Different events can be defined in one sample space.
2. An objective of the probability theory is to enable computing the probabilities of complicated events from that of simple events.
3. For this reason, one needs to study the relations among different events.

Operations of events are the same as those of sets:

1. A B , where A is a **sub-event** of B, or B contains A, implying that outcome *w* is in B if *w* is in A.
2. **Intersection**; A B or AB represents the event 'A and B'., so AB has occurred if both A and B have occurred.
3. **Union**; A U B represents the event 'A or B'. So A U B has occurred if either A or B has occurred.
4. **Complement**; A' represents the event 'not A'.

Some definitions include:

1. Events $A$ and $B$ are **mutually exclusive** if $A \cap B = \varnothing$.
2. Events $A_1, A_2, A_3, \cdots$ are **mutually exclusive** if they are pair-wisely mutually exclusive, i.e. $A_i \cap A_j = \varnothing$ for any $i \neq j$.
3. Events $A_1, A_2, \cdots A_k$ are **collectively exhaustive** if $A_1 \cup A_2 \cup \cdots \cup A_k = S$.
4. $A_1, A_2, \cdots A_k$ are **mutually exclusive and exhaustive** if $A_1 \cup A_2 \cup \cdots \cup A_k = S$ and $A_i \cap A_j = \varnothing$ for any $i \neq j$

The priority of the operations are: (1) complement, (2) intersection, (3) union or difference.

The properties of the event operations are:

1. **Commutative laws.** $A \cup B = B \cup A,\ A \cap B = B \cap A$.
2. **Associative laws.** $A \cup (B \cup C) = (A \cup B) \cup C,\ A \cap (B \cap C) = (A \cap B) \cap C$.
3. **Distributive laws.** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. **De Morgan's laws.** $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$.

   **Proof:** If $\omega \in (A \cup B)'$, $\Rightarrow \omega \notin A \cup B$. $\Rightarrow \omega \notin A$ and $\omega \notin B$.

   $\Rightarrow \omega \in A'$ and $\omega \in B'$. $\Rightarrow \omega \in A' \cap B'$. $\Rightarrow (A \cup B)' \subset A' \cap B'$.

   Similarly, we can show that $A' \cap B' \subset (A \cup B)'$.

   Therefore, $(A \cup B)' = A' \cap B'$.

   Similarly one can show that $(A \cap B)' = A' \cup B'$.

### 1.02 Classic Probability Models

Given a random experiment and the associated sample space S, the primary objective of probability modelling is to assign to each event A a real number *P(A),* called the probability of A. Sometimes, this can be determined from the long-term relative frequency on A or can be determined by the degrees of belief about A. Generally, there is no universal method for determining the probability value P(A) which must be dependent on the nature of the experiment involved.

However, prior to focusing on specifying P(A), one can deduce an abstract mathematical definition for probability by following the properties of the relative frequencies of events. With that definition and the associated properties of probability, one can be able to find the probability for any general event once some background information about the experiment is provided.

Probability is a real valued set function P that assigns to each event A in the sample space S a number P(A), called the probability of the event A, such that the following properties below are satisfied.