

# Probability

Madiba Hudson-Quansah

# Contents

# Chapter 1

## Module 8: Introduction

### 1.1 Introduction To Probability

#### Definition 1.1.1: Probability

A mathematical description of randomness and uncertainty / The likelihood of an event occurring.  
The notation for Probability is  $\mathbf{P}(X)$  where  $X$  is the event.  
Probability is always between  $0 \leq \mathbf{P}(X) \leq 1$  or  $0\% \leq \mathbf{P}(X) \leq 100\%$ .

There are two ways of determining probability:

- Theoretical / Classical - Determined by the nature of the experiment
- Empirical / Observational - Determined by the results of the experiment

### 1.2 Relative Frequency

#### Definition 1.2.1: Relative Frequency

Relative frequency is the number of times an event occurs divided by the total number of trials.

$$\mathbf{P}(X) = \frac{\text{Number of times event occurs}}{\text{Total number of trials}}$$

#### Theorem 1.2.1 The Law of Large Numbers

As the number of trials increases, the relative frequency of an event approaches the theoretical probability of the event.

## Chapter 2

# Module 9: Find the Probability of Events

### 2.1 Sample Spaces and Events

#### Definition 2.1.1: Random Experiment

An experiment whose outcome is determined by chance.

#### Definition 2.1.2: Sample Space

The list of possible outcomes of a random experiment, denoted by  $S$ .

#### Definition 2.1.3: Event

A statement about the nature of the outcome after the experiment has been conducted, denoted by any capital letter except  $S$ .

### 2.2 Equally Likely Outcomes

$$\mathbb{P}(A) = \frac{\text{Number of outcomes in } A}{\text{Number of outcomes in } S}$$

Where  $A$  is an event and  $S$  is the sample space.

### 2.3 Probability Rules

#### 2.3.1 Rule 1: Probability is a Number Between 0 and 1

For any event  $A$ ,  $0 \leq \mathbb{P}(A) \leq 1$ .

#### 2.3.2 Rule 2: Addition Rule

$\mathbb{P}(S) = 1$ , that is the sum of the probabilities of all possible outcomes is 1.

#### 2.3.3 Rule 3: Complement Rule

$\mathbb{P}(A') = 1 - \mathbb{P}(A)$ , that is the probability of the complement of an event is 1 minus the probability the event occurs.

### 2.3.4 Rule 4: Addition Rule for Mutually Exclusive Events

#### Definition 2.3.1: Mutually Exclusive / Disjoint events

Events that cannot happen at the same time.

$$\mathbb{P}(A \text{ or } B) = \mathbb{P}(\text{event } A \text{ occurs or event } B \text{ occurs or both occur})$$

If  $A$  and  $B$  are mutually exclusive, then  $\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B)$

### 2.3.5 Rule 5: Multiplication Rule for Independent Events

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(\text{event } A \text{ occurs and event } B \text{ occurs})$$

#### Definition 2.3.2: Independent Events

Two events  $A$  and  $B$  are said to be independent if the occurrence of one event does not affect the probability of the other event occurring.

#### Definition 2.3.3: Dependent Events

Two events  $A$  and  $B$  are said to be dependent if the occurrence of one event affects the probability of the other event occurring.

$$\text{If } A \text{ and } B \text{ are two independent events, then } \mathbb{P}(A \text{ and } B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

### 2.3.6 Rule 6: General Addition Rule

For any two events  $A$  and  $B$ ,  $\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ and } B)$ . If the events are mutually exclusive, then  $\mathbb{P}(A \text{ and } B) = 0$ , giving us  $\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B)$ , i.e. the addition rule for mutually exclusive events.

## Chapter 3

# Module 10: Conditional Probability and Independence

### Definition 3.0.1: Conditional Probability

The probability an event occurs as a result of another event. I.e. Probability of event  $B$ , given event  $A$  is,

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(A)}$$

### 3.1 Independence

When two events are independent, the probability of one event occurring does not affect the probability of the other event, i.e.

$$\mathbb{P}(B | A) = \mathbb{P}(B)$$

$$\mathbb{P}(A | B) = \mathbb{P}(A)$$

$$\mathbb{P}(B | A) = \mathbb{P}(B | A')$$

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

### 3.2 The General Multiplication Rule

For any two dependent events  $A$  and  $B$

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(A) \times \mathbb{P}(B | A)$$

### 3.3 Probability Trees

#### Definition 3.3.1: Probability Tree

A diagram that shows the sample space of a random experiment and the probability of each outcome.

#### 3.3.1 Bayes' Theorem

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A) \times \mathbb{P}(B | A)}{\mathbb{P}(A) \times \mathbb{P}(B | A) + \mathbb{P}(A') \times \mathbb{P}(B | A')}$$

# Chapter 4

## Module 11: Random Variables

### Definition 4.0.1: Random Variable

Assigns a unique numerical value to the outcome of a random experiment.

### Definition 4.0.2: Discrete Random Variable

A random variable that can take on a finite number of values. Discrete random variables are usually counts.

### Definition 4.0.3: Continuous Random Variable

A random variable that can take on an infinite number of values. Continuous random variables are usually measurements.

## 4.1 Discrete Random Variables

### 4.1.1 Notation

For a given event  $X$ , the probability of  $X$  is denoted by  $\mathbb{P}(X)$ . For a given value  $x$ , the probability of  $X$  is denoted by  $\mathbb{P}(X = x)$ , i.e. the probability that  $X$  takes on the value  $x$ .

### 4.1.2 Probability Distribution

#### Definition 4.1.1: Probability Distribution

The list of all possible values of a random variable and their corresponding probabilities.

Any probability distribution must satisfy the following two conditions:

- $0 \leq \mathbb{P}(X = x) \leq 1$  - The probability of any value of  $X$  is between 0 and 1.
- $\sum_x \mathbb{P}(X = x) = 1$  - The sum of the probabilities of all possible values of  $X$  is 1.

### 4.1.3 Key Words

- At least / No less than -  $x \geq$
- At most / No more than -  $x \leq$
- Less than / fewer than -  $x <$
- More than / greater than -  $x >$

- Exactly -  $x =$

#### 4.1.4 Mean and Variance of a Discrete Random Variable

##### 4.1.4.1 Mean

###### Definition 4.1.2: Mean / Expected value of a Discrete Random Variable

The average value of a random variable, denoted by  $\mu$ .

For a given random variable  $X$ , the mean is given by

$$\mu_X = \sum_{i=1}^n x_i p_i$$

Where  $x_i$  is the value of  $X$  and  $p_i$  is the probability of  $X$  taking on the value  $x_i$ .

##### 4.1.4.1.1 Applications of the Mean

- The mean of a random variable is the long-term average value of the random variable.
- The mean of a random variable is the centre of the probability distribution of the random variable.

##### 4.1.4.2 Variance

###### Definition 4.1.3: Variance

The average of the squared differences between each value of a random variable and the mean of the random variable, denoted by  $\sigma^2$ .

For a given random variable  $X$ , the variance is given by

$$\sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 p_i$$

And standard deviation is given by

$$\sigma_X = \sqrt{\sigma_X^2}$$

Where  $x_i$  is the value of  $X$  and  $p_i$  is the probability of  $X$  taking on the value  $x_i$ .

##### 4.1.4.3 Rules for Mean and Variance of Random Discrete Variables

###### 4.1.4.3.1 Adding or Subtracting a Constant to a Random Variable

If  $Y = X + c$ , then  $\mu_Y = \mu_X + c$ ,  $\sigma_Y^2 = \sigma_X^2$  and  $\sigma_Y = \sigma_X$ .

If  $Y = X - c$ , then  $\mu_Y = \mu_X - c$ ,  $\sigma_Y^2 = \sigma_X^2$  and  $\sigma_Y = \sigma_X$ .

###### 4.1.4.3.2 Multiplying a Random Variable by a Constant $> 1$

If  $Y = cX$ ,  $c > 1$ , then  $\mu_Y = c\mu_X$ ,  $\sigma_Y^2 = c^2\sigma_X^2$  and  $\sigma_Y = c\sigma_X$

###### 4.1.4.3.3 Multiplying a Random Variable by a Constant $< 1$

If  $Y = cX$ ,  $c < 1$ , then  $\mu_Y = c\mu_X$ ,  $\sigma_Y^2 = c^2\sigma_X^2$  and  $\sigma_Y = c\sigma_X$



#### 4.1.4.3.4 Linear Transformation of a Random Variable

If  $Y = a + bX$ , then  $\mu_Y = a + b\mu_X$ ,  $\sigma_Y^2 = b^2\sigma_X^2$  and  $\sigma_Y = |b|\sigma_X$

#### 4.1.4.3.5 Sum of Two Random Variables

If  $Z = X + Y$ , then  $\mu_Z = \mu_X + \mu_Y$ ,  $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$  and  $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$ . Only if  $X$  and  $Y$  are independent.

### 4.1.5 Binomial Random Variables

#### Definition 4.1.4: Binomial Random Variable

A random variable that counts the number of successes in a fixed number of independent trials, denoted by  $X \sim \text{Bin}(n, p)$ . Where  $n$  is the number of trials and  $p$  is the probability of success.

#### Definition 4.1.5: Binomial Experiment

Random experiments that satisfy the following conditions:

- A fixed number of trials, denoted by  $n$ .
- Each trial is independent of the others.
- There are only two possible outcomes for each trial, success or failure.
- There is a constant probability of success, denoted by  $p$ , for each trial, which can be expressed as the complement of the probability of failure,  $q = 1 - p$ .

#### Note:-

The number ( $X$ ) of success in a sample of size  $n$  taken without replacement from a population with proportion  $p$  of successes is approximately binomial with  $n$  and  $p$  as long as the sample size is at most 10% of the population size ( $N$ ). I.e.

$$n \leq 0.1N$$

Or

$$N \geq 10n$$

To calculate the probability of a binomial random variable, we use the formula

$$\mathbf{P}(X = x) = \binom{n}{x} p^x q^{n-x}, \text{ where } x = 0, 1, 2, \dots, n$$

Where  $n$  is the number of trials,  $x$  is the number of successes,  $p$  is the probability of success and  $q$  is the probability of failure.

If  $X$  is Binomial with parameters  $n$  and  $p$ , then

$$\mu_X = np$$

And

$$\begin{aligned} \sigma_X^2 &= np(1-p) \\ \sigma_X &= \sqrt{np(1-p)} \end{aligned}$$

## 4.2 Continuous Random Variables

### 4.2.1 Probability Distribution

For a continuous random variable  $X$ , the probability distribution is given by the *probability density function*, whose properties are

- $f(x) \geq 0$  for all  $x$ .
- $\int_{-\infty}^{\infty} f(x) dx = 1$
- The probability that  $X$  takes on a value between  $a$  and  $b$  is given by

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$$

**Note:-**

- The probability that a continuous random variable takes on a specific value is always 0.
- The strictness of the inequality does not matter, i.e.  $\mathbb{P}(X \geq a) = \mathbb{P}(X > a)$

### 4.2.2 Normal Random Variables

**Definition 4.2.1: Normal Random Variable**

A random variable that has a bell-shaped probability distribution, denoted by  $X \sim N(\mu, \sigma^2)$ . Where  $\mu$  is the mean and  $\sigma^2$  is the variance.

For a normally distributed random variable  $X$ :

- There is a 68% chance that  $X$  takes on a value within one standard deviation of the mean, i.e.  $0.68 = \mathbb{P}(\mu - \sigma < X < \mu + \sigma)$
- There is a 95% chance that  $X$  takes on a value within two standard deviations of the mean, i.e.  $0.95 = \mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma)$
- There is a 99.7% chance that  $X$  takes on a value within three standard deviations of the mean, i.e.  $0.997 = \mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma)$

#### 4.2.2.1 Finding Probabilities for Normal Random Variables

##### 4.2.2.1.1 Standardizing Values

**Definition 4.2.2: z-score**

The number of standard deviations a value is from the mean of a normal random variable, denoted by  $z$ .

To standardize a normal random variable  $X$ , we must find its z-score, given by

$$z = \frac{x - \mu}{\sigma}$$

##### 4.2.2.1.2 Finding Probabilities with the z-score

**Definition 4.2.3: Normal Table**

A table that shows the probability that a standard normal random variable takes on a value less than a given z-score.

Using the z-score we can find the probability that a normal random variable takes on a value less than a given value  $x$ , by tracing the z-score to the normal table.

$$\mathbb{P}(X < x) = \mathbb{P}(Z < z)$$

On a standard normal table z-score are written to two decimal places as row headers and for additional precision the column headers are the first two decimal places of the z-score.

### 4.2.3 Uniform Distribution

#### Definition 4.2.4: Uniform Distribution

,denoted by  $X \sim U(a, b)$

For a random variable  $X$ , if is uniformly distributed over the interval  $a$  and  $b$  then its *probability distribution density function* is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

The mean, variance, and standard deviation of a uniformly distributed random variable is given by

$$\begin{aligned} \mu_X &= \frac{a+b}{2} \\ \sigma_X^2 &= \frac{(b-a)^2}{12} \\ \sigma_X &= \sqrt{\frac{(b-a)^2}{12}} \end{aligned}$$

# Chapter 5

## Module 12: Sampling Distributions

### 5.1 Parameters vs. Statistics

#### Definition 5.1.1: Parameter

A numerical value that describes a characteristic of a population, denoted by a Greek letter, e.g.  $\mu, \sigma^2$ .

#### Definition 5.1.2: Statistic

A numerical value that describes a characteristic of a sample, denoted by a Roman letter, e.g.  $\bar{x}, s^2$ .

#### Definition 5.1.3: Proportion

A statistic that estimates the proportion of a population or sample that has a certain characteristic, denoted by  $p$  for a population and  $\hat{p}$  for a sample.

#### Definition 5.1.4: Sampling Variability

The variability of a statistic from one sample to another.

### 5.2 Behaviour of Sample Proportion $\hat{p}$

#### 5.2.1 Centre

The mean of the sample proportion is the same as the population proportion, i.e.

$$\mu_{\hat{p}} = p$$

As it is reasonable to expect all the sample proportions in repeated samples to average out to the underlying population.

#### 5.2.2 Spread

The sample size has an effect on the spread of the distribution of the sample proportion, i.e. the **larger the sample size, the less spread out the distribution** of the sample proportion and **more spread for smaller sample sizes**. We can describe the spread of the distribution of the sample proportion more precisely by finding the actual standard deviation of the sample proportion. i.e.

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Where  $p$  is the population proportion and  $n$  is the sample size.

### 5.2.3 Shape

The shape of the distribution of the sample proportion is approximately normal if the sample size is large enough. I.e. if

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

#### Definition 5.2.1: Sampling of Distribution of $\hat{p}$

The distribution of the values of the sample proportions  $\hat{p}$  in repeated samples.

## 5.3 Behaviour of Sample Mean $\bar{X}$

### 5.3.1 Centre

The mean of the sample mean is the same as the population mean, i.e.

$$\mu_{\bar{X}} = \mu$$

### 5.3.2 Spread

The sample size has an effect on the spread of the distribution of the sample mean, i.e. the **larger the sample size, the less spread out the distribution** of the sample mean and **more spread for smaller sample sizes**. We can describe the spread of the distribution of the sample mean more precisely by finding the actual standard deviation of the sample mean. i.e.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

### 5.3.3 Shape

The shape of the distribution of the sample mean is approximately normal if the sample size is large enough. I.e. if

$$n \geq 30$$

#### Definition 5.3.1: Sampling of Distribution of $\bar{X}$

The distribution of the values of the sample mean  $\bar{X}$  in repeated samples.