

# MATH50013 - Probability and Statistics for JMC

Notes by Robert Weingart

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Introduction to Uncertainty . . . . .	4
1.2	Introduction to Statistics . . . . .	4
1.2.1	Population vs. Sample . . . . .	4
1.3	Probability AND Statistics . . . . .	4
1.4	Statistical Modelling . . . . .	4
<b>2</b>	<b>Set Theory Review</b>	<b>4</b>
2.1	Sets, subsets and complements . . . . .	4
2.1.1	Sets . . . . .	4
2.1.2	Membership, subsets, equality, complements, and singletons . . .	4
2.2	Set operations . . . . .	4
2.2.1	Venn diagrams, Unions and Intersections . . . . .	4
2.2.2	Cartesian Products . . . . .	4
2.3	Cardinality . . . . .	4
<b>3</b>	<b>Visual and Numerical Summaries</b>	<b>4</b>
3.1	Visualization . . . . .	4
3.1.1	The histogram . . . . .	4
3.1.2	Empirical CDF . . . . .	4
3.2	Summary Statistics . . . . .	4
3.2.1	Measures of Location . . . . .	4
3.2.2	Measures of Dispersion . . . . .	5
3.2.3	Covariance and Correlation . . . . .	5
3.2.4	Skewness . . . . .	5
3.3	One more visualization: the box-and-whisker plot . . . . .	5
<b>4</b>	<b>Probability</b>	<b>6</b>
4.1	The formal structure . . . . .	6
4.1.1	$\sigma$ -algebras . . . . .	6
4.1.2	Probability measure . . . . .	6
4.2	Interpretations of the probability space . . . . .	6
4.3	Interpretation of the $\sigma$ -algebra . . . . .	6
4.3.1	The sample space ( $S$ ) . . . . .	6
4.3.2	The event space ( $\mathcal{F}$ ) . . . . .	6
4.4	Interpretations of the probability measure ( $P$ ) . . . . .	6
4.4.1	Classical interpretation . . . . .	6

4.4.2	Frequentist interpretation . . . . .	6
4.4.3	Subjective interpretation . . . . .	7
4.5	A few derivations from the axioms . . . . .	7
4.6	Conditional Probability . . . . .	7
4.7	Independent Events . . . . .	7
4.7.1	More Examples . . . . .	7
4.7.2	Conditional Independence . . . . .	7
4.7.3	Joint Events . . . . .	8
4.8	Bayes's Theorem . . . . .	8
4.12	More Examples . . . . .	8
<b>5</b>	<b>Discrete Random Variables</b>	<b>8</b>
5.1	Random Variables . . . . .	8
5.1.1	Cumulative Distribution Function . . . . .	8
5.2	Discrete Random Variables . . . . .	9
5.2.1	Properties of Mass Function $p_X$ . . . . .	9
5.2.2	Discrete Cumulative Distribution Function . . . . .	9
5.2.3	Connection between $F_X$ and $p_X$ . . . . .	9
5.2.4	Properties of Discrete CDF $F_X$ . . . . .	9
5.3	Functions of a discrete random variable . . . . .	10
5.4	Mean and Variance . . . . .	10
5.4.1	Expectation . . . . .	10
5.5.1	Sums of Random Variables . . . . .	10
5.6	Some Important Discrete Random Variables . . . . .	11
5.6.1	Bernoulli Distribution . . . . .	11
5.6.2	Binomial Distribution . . . . .	11
5.6.3	Geometric Distribution . . . . .	11
5.6.4	Poisson Distribution . . . . .	11
5.6.5	Discrete Uniform Distribution . . . . .	11
<b>6</b>	<b>Continuous Random Variables</b>	<b>12</b>
6.0.1	Continuous Cumulative Distribution Function . . . . .	12
6.0.2	Properties of Continuous $F_X$ and $f_X$ . . . . .	12
6.0.3	Transformations . . . . .	12
6.1	Mean, Variance and Quantiles . . . . .	13
6.1.1	Expectation . . . . .	13
6.1.2	Variance . . . . .	13
6.1.3	Quantiles . . . . .	13
6.2	Some Important Continuous Random Variables . . . . .	13
6.2.1	Continuous Uniform Distribution . . . . .	13
6.2.2	Exponential Distribution . . . . .	14
6.2.3	Normal (Gaussian) Distribution . . . . .	14
6.4	Further examples . . . . .	14
<b>7</b>	<b>Joint Random Variables</b>	<b>14</b>
7.0.1	Joint Cumulative Distribution Function . . . . .	14
7.0.2	Properties of Joint CDF $F_{XY}$ . . . . .	14
7.0.3	Joint Probability Mass Functions . . . . .	15
7.0.4	Joint Probability Density Functions . . . . .	15

7.1	Independence, Conditional Probability, Expectation . . . . .	15
7.1.1	Independence and conditional probability . . . . .	15
7.1.2	Expectation . . . . .	16
7.1.3	Conditional Expectation . . . . .	16
7.2	Examples . . . . .	16
7.3	Multivariate Transformations . . . . .	16
7.3.1	Convolutions (sums of random variables) . . . . .	16
7.5.1	General Bivariate Transformations . . . . .	17
<b>8</b>	<b>Estimation</b>	<b>18</b>
8.1	Estimators . . . . .	18
8.1.1	Point estimates . . . . .	18
8.1.2	Bias, Efficiency, Consistency . . . . .	18
8.1.3	Maximum Likelihood Estimation . . . . .	19
8.2	Confidence Intervals . . . . .	19
8.2.1	Normal Distribution with Known Variance . . . . .	19
8.2.2	Normal Distribution with Unknown Variance . . . . .	19
8.2.3	Another way to view the confidence interval: Neyman construction	20
<b>9</b>	<b>Hypothesis Testing</b>	<b>20</b>
9.0.1	Error Rates and Power of a Test . . . . .	20
9.1	Testing for a population mean . . . . .	21
9.1.1	Normal Distribution with Known Variance . . . . .	21
9.1.2	Normal Distribution with Unknown Variance . . . . .	21
9.2	Testing for differences in population means . . . . .	21
9.2.1	Two Sample Problems . . . . .	21
9.2.2	Normal Distributions with Known Variances . . . . .	22
9.2.3	Normal Distributions with Unknown Variances . . . . .	22
9.3	Goodness of Fit . . . . .	22
9.3.1	Count Data and Chi-Square Tests . . . . .	22
9.3.2	Proportions . . . . .	23
9.3.3	Model Checking . . . . .	23
9.3.4	Independence . . . . .	23
9.3.5	The $\chi^2$ distribution and degrees of freedom . . . . .	23
<b>10</b>	<b>Convergence Concepts</b>	<b>23</b>

# 1 Introduction

## 1.1 Introduction to Uncertainty

## 1.2 Introduction to Statistics

### 1.2.1 Population vs. Sample

## 1.3 Probability AND Statistics

## 1.4 Statistical Modelling

# 2 Set Theory Review

## 2.1 Sets, subsets and complements

### 2.1.1 Sets

### 2.1.2 Membership, subsets, equality, complements, and singletons

## 2.2 Set operations

### 2.2.1 Venn diagrams, Unions and Intersections

### 2.2.2 Cartesian Products

## 2.3 Cardinality

# 3 Visual and Numerical Summaries

## 3.1 Visualization

### 3.1.1 The histogram

**Definition.** A **histogram** partitions the range of a sample into **bins** and shows what number of data points in each bin. Rather than frequency, the amount shown can also be relative frequency or density.

### 3.1.2 Empirical CDF

**Definition.** The **indicator function** is defined as  $I(\text{false}) := 0$  and  $I(\text{true}) = 1$ .

**Definition.** The **empirical cumulative distribution function** of a sample is

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$$

## 3.2 Summary Statistics

### 3.2.1 Measures of Location

**Definition.** The **arithmetic mean** is  $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ .

**Definition.** The **geometric mean** is  $x_G := (\prod_{i=1}^n x_i)^{\frac{1}{n}}$ .

**Definition.** The **harmonic mean** is  $x_H := n \left( \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$

**Definition.** The  **$i$ th order statistic**, written  $x_{(i)}$ , is the  $i$ th smallest value of the sample. For non-integer values of the form  $i + \alpha$  with  $\alpha \in (0, 1)$ , we define

$$x_{(i+\alpha)} := (1 - \alpha)x_{(i)} + \alpha x_{(i+1)}$$

**Definition.** The **median** is  $x_{(\frac{n+1}{2})}$ .

**Definition.** The **mode** is the most frequently occurring value. If there are multiple then the sample is **multimodal**.

### 3.2.2 Measures of Dispersion

**Definition.** The **mean square** or **sample variance** is

$$s_x^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

**Definition.** The **root mean square** or **sample standard deviation** is

$$s_x := \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

**Definition.** The **range** is  $x_{(n)} - x_{(1)}$ .

**Definition.** The **first quartile** is  $x_{(\frac{1}{4}(n+1))}$ . The **third quartile** is  $x_{(\frac{3}{4}(n+1))}$ . The **interquartile range** is the difference between the third and first quartiles.

### 3.2.3 Covariance and Correlation

**Definition.** For a sample where each data point is an  $(x_i, y_i)$  pair, the **covariance** is

$$s_{xy} := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x} \bar{y}$$

.

**Definition.** For a sample as above, the **correlation** is

$$r_{xy} := \frac{s_{xy}}{s_x s_y}$$

### 3.2.4 Skewness

**Definition.** The **skewness** is  $\frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^3$ .

## 3.3 One more visualization: the box-and-whisker plot

**Definition.** A **box-and-whisker plot** shows the median, first and third quartiles, points within  $\frac{3}{2} \times IQR$  of the quartiles, and any outliers.

## 4 Probability

### 4.1 The formal structure

#### 4.1.1 $\sigma$ -algebras

**Definition 4.1.1.** A  $\sigma$ -algebra associated with  $S$  is a set  $\mathcal{F}$  of subsets of  $S$  where  $S \in \mathcal{F}$ ,  $\mathcal{F}$  is closed under complements with respect to  $S$ , and  $\mathcal{F}$  is closed under countable unions.

**Proposition.**  $\emptyset \in \mathcal{F}$ .  $\mathcal{F}$  is also closed under countable intersections.

#### 4.1.2 Probability measure

**Definition 4.1.2.** A **probability measure** is a function  $P : \mathcal{F} \rightarrow \mathbb{R}$  where  $P(E) \geq 0$  for any  $E$ ,  $P(S) = 1$ , and for countably many disjoint sets  $E_i$ ,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

. A triple  $(S, \mathcal{F}, P)$  as previously defined is a **probability space**.

### 4.2 Interpretations of the probability space

### 4.3 Interpretation of the $\sigma$ -algebra

#### 4.3.1 The sample space ( $S$ )

**Definition.** The **sample space**  $S$  is the set of all possible outcomes of an experiment.

#### 4.3.2 The event space ( $\mathcal{F}$ )

**Definition.** An **event** is a subset  $E \subset S$ .  $\mathcal{F}$  is the set of all possible events being considered (which may not include all possible combinations of outcomes).

**Definition.**  $E_1$  and  $E_2$  are **mutually exclusive** iff  $E_1 \cap E_2 = \emptyset$  i.e. they cannot both happen at once.

### 4.4 Interpretations of the probability measure ( $P$ )

#### 4.4.1 Classical interpretation

**Definition.** In the **classical interpretation**,  $S$  consists of finitely many equally likely **elementary events** and  $P(E) = \frac{|E|}{|S|}$ . For an infinite  $S$ , this can still be applied by replacing cardinality above with a different measure.

#### 4.4.2 Frequentist interpretation

**Definition.** In the **frequentist interpretation**, when an experiment is repeated infinitely many times, the proportion of trials in which  $E$  occurs approaches  $P(E)$ .

### 4.4.3 Subjective interpretation

**Definition.** In the **subjective interpretation**,  $P(E)$  is the degree of belief a person has that  $E$  occurs.

## 4.5 A few derivations from the axioms

**Proposition.** For  $E, F \in \mathcal{F}$ ,

- $P(\emptyset) = 0$
- $P(E) \leq 1$
- $P(\overline{E}) = 1 - P(E)$
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- $P(E \cap \overline{F}) = P(E) - P(E \cap F)$
- $E \subseteq F \implies P(E) \leq P(F)$

## 4.6 Conditional Probability

**Definition 4.6.1.** For  $P(F) > 0$  the **conditional probability of  $E$  given  $F$**  is

$$P(E \mid F) := \frac{P(E \cap F)}{P(F)}$$

**Proposition.** For  $P(F) > 0$ ,

- For any  $E \in \mathcal{F}$ ,  $P(E \mid F) \geq 0$
- $P(F \mid F) = 1$
- For  $E_1, \dots, E_n \in \mathcal{F}$  pairwise disjoint,  $P(\bigcup_{i=1}^n E_i \mid F) = \sum_{i=1}^n P(E_i \mid F)$

## 4.7 Independent Events

**Definition 4.7.1.**  $E, F \in \mathcal{F}$  are **independent** iff  $P(E \cap F) = P(E)P(F)$ .  $E_1, \dots, E_n$  are **independent** iff for any subset  $E_{i_1}, \dots, E_{i_l}$  we have  $P\left(\bigcap_{j=1}^l E_{i_j}\right) = \prod_{j=1}^l P(E_{i_j})$ .

**Proposition.**  $E$  and  $F$  are independent  $\implies E$  and  $\overline{F}$  are independent.

**Proposition.**  $E$  and  $F$  are independent  $\iff P(E \mid F) = P(E)$ .

### 4.7.1 More Examples

### 4.7.2 Conditional Independence

**Definition.** For  $E_1, E_2, F \in \mathcal{F}$ ,  $E_1$  and  $E_2$  are **conditionally independent given  $F$**  iff  $P(E_1 \cap E_2 \cap F) = P(E_1 \mid F)P(E_2 \mid F)$ .

### 4.7.3 Joint Events

**Definition.** When combining multiple independent experiments, a **probability table** can be used to show the probabilities of all elementary events (i.e. combinations of an elementary event in each experiment).

## 4.8 Bayes's Theorem

**Theorem 4.9.** (*Bayes's*) For  $E, F \in \mathcal{F}$  with  $P(E) > 0$  and  $P(F) > 0$ ,

$$P(E | F) = \frac{P(F | E)P(E)}{P(F)}$$

**Theorem 4.10.** (*The Law of Total Probability*) For a partition  $E_1, \dots$  of  $S$ , and any  $F \in \mathcal{F}$ ,  $P(F) = \sum_i P(F | E_i)P(E_i)$ .

**Theorem 4.11.** (*Bayes's applied to a partition*) For a partition  $E_1, \dots$  of  $S$  with  $P(E_i) > 0$  for all  $i$  and  $F \in \mathcal{F}$  with  $P(F) > 0$ ,

$$P(E_i | F) = \frac{P(F | E_i)P(E_i)}{\sum_j P(F | E_j)P(E_j)}$$

### 4.12 More Examples

## 5 Discrete Random Variables

### 5.1 Random Variables

**Definition 5.1.1.** A **random variable** is a measurable mapping  $X : S \rightarrow \mathbb{R}$  where  $\forall x \in \mathbb{R}, \{s \in S : X(s) \leq x\} \in \mathcal{F}$ .

**Definition 5.1.2.** The **range** of  $X$  is  $\mathbb{X}$ , the image of  $S$  under  $X$ .

**Definition.** The **probability distribution** of  $X$  is

$$P_X(X \in B) := P(\{s \in S : X(s) \in B\})$$

where  $B \subseteq \mathbb{R}$ .

**Notation.** For brevity we write  $\{X \in B\} := \{s \in S : X(s) \in B\}$  (TODO: doesn't this make  $P$  and  $P_X$  interchangeable?) and  $\{a < X \leq b\} := \{X \in (a, b]\}$  etc.

#### 5.1.1 Cumulative Distribution Function

**Definition 5.1.3.** The **cumulative distribution function** of  $X$  is  $F_X : \mathbb{R} \rightarrow [0, 1]$  where  $F_X(x) = P_X(X \leq x)$ .

**Definition.** A function  $f$  is **right-continuous** iff for any decreasing sequence  $x_i \rightarrow x$  we have  $f(x_i) \rightarrow f(x)$ .

**Proposition.** A CDF is right-continuous.

**Proposition.**  $F_X$  is a CDF iff all the following hold:



- $F_X$  is right-continuous
- $F_X(\mathbb{R}) \subseteq [0, 1]$
- $F_X$  is monotonically increasing
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow \infty} F_X(x) = 1$

## 5.2 Discrete Random Variables

**Definition 5.2.1.** A random variable is **discrete** iff its range is finite or countably infinite.

**Definition 5.2.2.** For a DRV  $X$ , the **probability mass function**  $p_X : \mathbb{R} \rightarrow [0, 1]$  is  $p_X(x) = P_X(X = x)$  for  $x \in \mathbb{X}$  and  $p_X(x) = 0$  for  $x \notin \mathbb{X}$ .

**Definition.** The **support** of  $X$  is  $\{x \in \mathbb{R} : p_X(x) > 0\}$ . Usually this is  $\mathbb{X}$ .

### 5.2.1 Properties of Mass Function $p_X$

**Proposition.** An arbitrary function  $p_X$  can be a PMF for  $X$  iff  $\forall x \in \mathbb{X}, p_X(x) \geq 0$  and  $\sum_{x \in \mathbb{X}} p_X(x) = 1$ .

### 5.2.2 Discrete Cumulative Distribution Function

**Definition.** The **cumulative distribution function** of a DRV  $X$  is  $F_X(x) = P(X \leq x)$  (TODO: is this not what it always is?).

### 5.2.3 Connection between $F_X$ and $p_X$

**Proposition.** For  $\mathbb{X} = \{x_1, \dots\}$  with the  $x_i \leq x_{i+1}$  for all  $i$ ,

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$$

Equivalently,

$$\forall i \geq 1, p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

### 5.2.4 Properties of Discrete CDF $F_X$

**Proposition.** We have

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow \infty} F_X(x) = 1$
- $\lim_{h \rightarrow 0^+} F_X(x + h) = F_X(x)$
- $a < b \implies F_X(a) \leq F_X(b)$
- For  $a < b$ ,  $P(a < X \leq b) = F_X(b) - F_X(a)$

### 5.3 Functions of a discrete random variable

**Proposition.** For a DRV  $X$  and  $g : \mathbb{X} \rightarrow \mathbb{R}$ ,  $Y = g(X)$  is also a DRV. We have

$$p_Y(y) = \sum_{x \in \mathbb{X}: g(x)=y} p_X(x)$$

### 5.4 Mean and Variance

**Notation.** All the functions defined in this section are of type  $\mathbf{RV} \rightarrow \mathbb{R}$ .

#### 5.4.1 Expectation

**Definition 5.4.1.** The **expected value** or **mean** of a DRV  $X$  is

$$E_X(X) := \sum_{x \in \mathbb{X}} xp_X(x)$$

It is often abbreviated to  $E(X)$ . For the case  $E_Y(X)$  with  $Y \neq X$ , see below.

**Theorem 5.5.** For a **function of interest**  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have

$$E_X(g(X)) = \sum_{x \in \mathbb{X}} g(x)p_X(x)$$

*This is the only situation where we can have  $E_X(Y)$  with  $X \neq Y$ .*

**Proposition.**  $E$  is linear.

**Definition 5.5.1.** For a DRV  $X$ , the **variance** of  $X$  is

$$\text{Var}_X(X) := E_X((X - E_X(X))^2) = E(X^2) - E(X)^2$$

**Proposition.** For  $a, b \in \mathbb{R}$ ,  $\text{Var}(aX + b) = a^2 \text{Var}(X)$

**Definition 5.5.2.** For a DRV  $X$ , the **standard deviation** of  $X$  is

$$\text{sd}(X) := \sqrt{\text{Var}_X(X)}$$

**Definition 5.5.3.** For a DRV  $X$ , the **skewness** of  $X$  is

$$\gamma_1 := \frac{E_X((X - E_X(X))^3)}{\text{sd}_X(X)^3}$$

#### 5.5.1 Sums of Random Variables

**Proposition.** For  $X_1, \dots, X_n$  (possibly with different distributions, not necessarily independent) with sum  $S_n$ , we have

$$E(S_n) = \sum_{i=1}^n E(X_i)$$

and

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i)$$

**Proposition.** For  $X_1, \dots, X_n$  independent with sum  $S_n$ , we have

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i)$$

and

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i)$$

**Proposition.** For  $X_1, \dots, X_n$  independent and identically distributed with sum  $S_n$ ,  $E(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2$ , we have

$$E\left(\frac{S_n}{n}\right) = \mu$$

and  $\text{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$

## 5.6 Some Important Discrete Random Variables

$X$	$\mathbb{X}$	$p_X(x)$	$E(X)$	$\text{Var}(X)$	$\gamma_1$
$X \sim \text{Bernoulli}(p)$	$\{0, 1\}$	$p^x(1-p)^{1-x}$	$p$	$p(1-p)$	$\frac{1-2p}{\sqrt{p(1-p)}}^*$
$X \sim \text{Binomial}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{x} p^x(1-p)^{n-x}$	$np$	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$
$X \sim \text{Geometric}(p)$	$\{1, 2, \dots\}$	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$
$X \sim \text{Poisson}(\lambda)$	$\{0, 1, \dots\}$	$\frac{e^{-\lambda} \lambda^x}{x!}$	$\lambda$	$\lambda$	$\frac{1}{\sqrt{\lambda}}$
$X \sim \text{U}(\{1, \dots, n\})$	$\{1, \dots, n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	0

\*: The skewness of the Bernoulli distribution is not given in the official notes.

### 5.6.1 Bernoulli Distribution

$X \sim \text{Bernoulli}(p)$  chooses between 1 and 0 where  $P(X = 1) = p$ .

### 5.6.2 Binomial Distribution

$X \sim \text{Binomial}(n, p)$  is the total number of successes after  $n$  Bernoulli trials with probability  $p$ .

### 5.6.3 Geometric Distribution

$X \sim \text{Geometric}(p)$  is the number of Bernoulli trials with probability  $p$  it will take to have the first success.

### 5.6.4 Poisson Distribution

$X \sim \text{Poisson}(\lambda)$  is the number of occurrences of an event that occurs at a rate of  $\lambda$ .

### 5.6.5 Discrete Uniform Distribution

$X \sim \text{U}(\{1, \dots, n\})$  is a random value out of  $\{1, \dots, n\}$ .

## 6 Continuous Random Variables

**Definition 6.0.1.** A random variable  $X$  is absolutely **continuous** iff there exists a measurable non-negative function  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  (the **probability density function**) where

$$\forall B \subseteq \mathbb{R}, P(X \in B) = \int_{x \in B} f_X(x) dx$$

### 6.0.1 Continuous Cumulative Distribution Function

**Definition 6.0.2.** The **cumulative distribution function** of a CRV  $X$  is  $F_X(x) = P(X \leq x)$  (as for any RV).

**Proposition.** For a CRV  $X$ ,  $F_X(x) = \int_{-\infty}^x f_X(x') dx'$

### 6.0.2 Properties of Continuous $F_X$ and $f_X$

**Proposition.** For a CRV  $X$ ,

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow \infty} F_X(x) = 1$
- If  $F_X$  is differentiable at  $x$  then  $f_X(x) = F'_X(x)$
- $\forall a \in \mathbb{R}, P(X = a) = 0$
- For  $a < b$ ,  $P(a < X \leq b) = F_X(b) - F_X(a)$
- $f_X(X)$  is not a probability, so we do not require  $f_X(x) \leq 1$
- $X$  is uniquely defined by  $f_X$

**Proposition.** An arbitrary function  $f_X$  is a PDF for a CRV iff  $\forall x \in \mathbb{R}, f_X(x) \geq 0$  and  $\int_{-\infty}^{\infty} f_X(x) dx = 1$  ( $f_X$  is **normalised**).

### 6.0.3 Transformations

**Proposition.** For  $Y = g(X)$  with  $g$  strictly monotonically increasing, we have

$$F_Y(y) = F_X(g^{-1}(y))$$

and

$$f_Y(y) = f_X(g^{-1}(y)) g^{-1'}(y)$$

**Proposition.** For  $Y = g(X)$  with  $g$  strictly monotonically decreasing, we have

$$F_Y(y) = 1 - F_X(g^{-1}(y))$$

and

$$f_Y(y) = -f_X(g^{-1}(y)) g^{-1'}(y)$$

## 6.1 Mean, Variance and Quantiles

### 6.1.1 Expectation

**Definition 6.1.1.** The **mean** or **expectation** of a CRV  $X$  is

$$E(X) := \int_{-\infty}^{\infty} x f_X(x) dx$$

**Definition.** For any measurable **function of interest**  $g : \mathbb{R} \rightarrow \mathbb{R}$  we have

$$E(g(X)) := \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

**Proposition.**  $E$  is linear.

### 6.1.2 Variance

**Definition 6.1.2.** The **variance** of a CRV  $X$  is

$$\text{Var}_X(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

**Proposition.** For  $a, b \in \mathbb{R}$ ,  $\text{Var}(aX + b) = a^2 \text{Var}(X)$

### 6.1.3 Quantiles

**Definition 6.1.3.** For  $\alpha \in [0, 1]$ , we  **$\alpha$ -quantile** of a CRV  $X$  is

$$Q_X(\alpha) := F_X^{-1}(\alpha)$$

so that  $P(X \leq Q_X(\alpha)) = \alpha$ .

## 6.2 Some Important Continuous Random Variables

$X$	$\mathbb{X}$	$f_X(x)$	$F_X(x)$	$E(X)$	$\text{Var}(X)$
$X \sim \text{U}(a, b)$	$(a, b)$	$\begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$X \sim \text{Exp}(\lambda)$	$[0, \infty)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$X \sim \text{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$	$\mu$	$\sigma^2$

### 6.2.1 Continuous Uniform Distribution

$X \sim \text{U}(a, b)$  or  $X \sim \text{Uniform}(a, b)$  is uniformly distributed on the interval  $(a, b)$  and 0 elsewhere.

**Definition.** The **standard uniform** is  $\text{Uniform}(0, 1)$ .

**Proposition.**  $X \sim \text{Uniform}(0, 1) \implies (a + (b - a)X) \sim \text{Uniform}(a, b)$ .

### 6.2.2 Exponential Distribution

$X \sim \text{Exp}(\lambda)$  is the time until an event occurring at rate  $\lambda$  occurs.

**Proposition.**  $X \sim \text{Exp}(\lambda)$  exhibits the **Lack of Memory Property**:

$$\forall x, t > 0, P(X > t + x \mid X > t) = P(X > x)$$

**Proposition.** If the number of events occurring in an interval of size  $x$  is  $N_x \sim \text{Poisson}(\lambda x)$  then the separation between two events is  $X \sim \text{Exp}(\lambda)$ .

### 6.2.3 Normal (Gaussian) Distribution

$X \sim N(\mu, \sigma^2)$  has no obvious interpretation.

**Definition.**  $X \sim N(0, 1)$  is the **standard normal distribution** or **unit normal distribution**. It has the PDF

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

and the CDF

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

**Proposition.**  $X \sim N(0, 1) \implies (\sigma X + \mu) \sim N(\mu, \sigma^2)$

**Theorem 6.3.** (Central Limit Theorem) For  $X_1, \dots, X_n$  independent and identically distributed with mean  $\mu$  and variance  $\sigma^2$ ,

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$

## 6.4 Further examples

## 7 Joint Random Variables

**Definition 7.0.1.** For RVs  $X$  and  $Y$  with the same sample space, the **joint probability distribution** is  $P_{XY}(B_X, B_Y) := P(X^{-1}(B_X) \cap Y^{-1}(B_Y))$  where  $B_X, B_Y \subseteq \mathbb{R}$ .

### 7.0.1 Joint Cumulative Distribution Function

**Definition 7.0.2.** The **joint cumulative distribution function** is

$$F_{xy}(x, y) := P_{XY}(X \leq x, Y \leq y).$$

**Proposition.**  $F_X(x) = F_{XY}(x, \infty)$  and  $F_Y(y) = F_{XY}(\infty, y)$ .

### 7.0.2 Properties of Joint CDF $F_{XY}$

**Proposition.** An arbitrary function  $F_{XY}$  is a valid joint CDF iff the following hold:

- $\forall x, y \in \mathbb{R}, F_{XY}(x, y) \in [0, 1]$
- $\forall x_1, x_2, y \in \mathbb{R}, x_1 < x_2 \implies F_{XY}(x_1, y) \leq F_{XY}(x_2, y)$
- $\forall x, y_1, y_2 \in \mathbb{R}, y_1 < y_2 \implies F_{XY}(x, y_1) \leq F_{XY}(x, y_2)$
- $\forall x, y \in \mathbb{R}, F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = 0$
- $F_{XY}(\infty, \infty) = 1$

### 7.0.3 Joint Probability Mass Functions

**Definition 7.0.3.** For DRV's  $X, Y$ , the **joint probability mass function** is  $p_{XY}(x, y) := P_{XY}(X = x, Y = y)$ .

**Proposition.**  $p_X(x) = \sum_{y \in \mathbb{Y}} p_{XY}(x, y)$  and  $p_Y(y) = \sum_{x \in \mathbb{X}} p_{XY}(x, y)$

**Proposition.** An arbitrary function  $p_{XY}$  is a valid joint PMF iff  $\forall x, y \in \mathbb{R}, p_{XY}(x, y) \in [0, 1]$  and  $\sum_{y \in \mathbb{Y}} \sum_{x \in \mathbb{X}} p_{XY}(x, y) = 1$ .

### 7.0.4 Joint Probability Density Functions

**Definition.** CRVs  $X$  and  $Y$  are **jointly continuous** iff  $\exists f_{XY} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  where

$$\forall B_{XY} \subseteq \mathbb{R} \times \mathbb{R}, P_{XY}(B_{XY}) = \int_{(x,y) \in B_{XY}} f_{XY}(x, y) dx dy$$

Then  $f_{XY}$  is the **joint probability density function** of  $X$  and  $Y$ .

**Proposition.** For jointly continuous CRVs, we have

$$F_{XY}(x, y) = \int_{t=-\infty}^y \int_{s=-\infty}^x f_{XY}(s, t) ds dt$$

**Definition 7.0.4.** (Not actually a definition) The joint PDF is

$$f_{XY} = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$

**Proposition.**  $f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x, y) dy$  and  $f_Y(y) = \int_{x=-\infty}^{\infty} f_{XY}(x, y) dx$

**Proposition.** An arbitrary function  $f_{XY}$  is a valid joint PDF iff  $\forall x, y \in \mathbb{R}, f_{XY}(x, y) \geq 0$  and  $\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$ .

## 7.1 Independence, Conditional Probability, Expectation

### 7.1.1 Independence and conditional probability

**Definition.** RVs  $X$  and  $Y$  are **independent** iff  $\forall B_X, B_Y \subseteq \mathbb{R}, P_{XY}(B_X, B_Y) = P_X(B_X)P_Y(B_Y)$ .

**Definition 7.1.1.** CRVs  $X$  and  $Y$  are **independent** iff  $\forall x, y \in \mathbb{R}, f_{XY}(x, y) = f_X(x)f_Y(y)$ .

**Definition 7.1.2.** For RVs  $X$  and  $Y$ , the **conditional probability distribution** is

$$P_{Y|X}(B_Y | B_X) := \frac{P_{XY}(B_X, B_Y)}{P_X(B_X)}$$

**Proposition.**  $X$  and  $Y$  are independent  $\iff \forall B_X, B_Y \subseteq \mathbb{R}, P_{Y|X}(B_Y | B_X) = P_Y(B_Y)$ .

**Definition 7.1.3.** For CRVs  $X$  and  $Y$ , the **conditional probability density function** is

$$f_{Y|X}(y | x) := \frac{f_{XY}(x, y)}{f_X(x)}$$

**Proposition.**  $X$  and  $Y$  are independent  $\iff \forall x, y \in \mathbb{R}, f_{Y|X}(y | x) = f_Y(y)$ .

### 7.1.2 Expectation

**Definition 7.1.4.** For DRV's  $X$  and  $Y$ :

$$E_{XY}(g(X, Y)) := \sum_{y \in \mathbb{Y}} \sum_{x \in \mathbb{X}} g(x, y) p_{XY}(x, y)$$

**Definition 7.1.5.** For CRV's  $X$  and  $Y$ :

$$E_{XY}(g(X, Y)) := \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

**Proposition.** Both versions of  $E$  are linear.

**Proposition.**  $E_{XY}(g_1(X) + g_2(Y)) = E_X(g_1(X)) + E_Y(g_2(Y))$ . If  $X$  and  $Y$  are independent then  $E_{XY}(g_1(X)g_2(Y)) = E_X(g_1(X))E_Y(g_2(Y))$ .

### 7.1.3 Conditional Expectation

**Definition 7.1.6.** The **conditional expectation** of  $Y$  given  $X = x$  is

$$E_{Y|X}(Y | X = x) := \sum_{y \in \mathbb{Y}} yp(y | x)$$

or

$$E_{Y|X}(Y | X = x) := \int_{y=-\infty}^{\infty} yf(y | x)dy$$

**Definition.** The **covariance** of  $X$  and  $Y$  is

$$\sigma_{XY} = \text{Cov}(X, Y) := E_{XY}((X - E_X(X))(Y - E_Y(Y)))$$

**Definition 7.1.7.** The **correlation** of  $X$  and  $Y$  is

$$\rho_{XY} = \text{Cor}(X, Y) := \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

**Proposition.**  $X$  and  $Y$  are independent  $\implies \sigma_{XY} = \rho_{XY} = 0$ .

## 7.2 Examples

## 7.3 Multivariate Transformations

### 7.3.1 Convolutions (sums of random variables)

**Theorem 7.4.** (Convolution Theorem) For independent RV's  $X$  and  $Y$  and  $Z = X + Y$ ,

$$p_Z(z) = \sum_{x \in \mathbb{X}} p_X(x)p_Y(z - x)$$

or

$$p_Z(z) = \int_{\mathbb{R}} f_X(x)f_Y(z - x)dx$$

**Theorem 7.5.** If  $X \sim N(\mu_X, \sigma_X^2)$  and  $Y \sim N(\mu_Y, \sigma_Y^2)$  are independent then  $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ .



### 7.5.1 General Bivariate Transformations

**Proposition.** For DRVs  $X$  and  $Y$  with  $U = g_1(X, Y)$  and  $V = g_2(X, Y)$ ,

$$p_{UV}(u, v) = \sum_{(x, y) \in A} p_{XY}(x, y)$$

where

$$A := \{(x, y) : (g_1(x, y), g_2(x, y)) = (u, v)\}$$

**Proposition.** For CRVs  $X$  and  $Y$  with  $U = g_1(X, Y)$  and  $V = g_2(X, Y)$ , and given  $u := g_1(x, y)$  and  $v := g_2(x, y)$ ,

$$f_{UV}(u, v) = f_{XY}(x, y) \left| \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right|$$

where

$$A := \{(x, y) : (g_1(x, y), g_2(x, y)) = (u, v)\}$$

**Definition.** The **Gamma function** is  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ , defined for  $\alpha \in (0, \infty)$ .

**Proposition.** We have:

- $\forall \alpha > 1, \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\Gamma(1) = 1$
- $\forall n \in \mathbb{N}, \Gamma(n) = (n - 1)!$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

**Definition.** The **Gamma distribution**  $X \sim \text{Gamma}(\alpha, \beta)$  with  $\alpha, \beta > 0$  has the following properties:

- $f_X(x) := \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
- $\mathbb{X} = (0, \infty)$
- $E(X) = \frac{\alpha}{\beta}$
- $\text{Var}(X) = \frac{\alpha}{\beta^2}$

**Definition.** The **Beta function** is  $B(\alpha, \beta) := \int_0^1 x^{\alpha-1} (1 - x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ .

**Definition.** The **Beta distribution**  $X \sim \text{Beta}(\alpha, \beta)$  has PDF  $f_X(x) := \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1 - x)^{\beta-1}$  and  $\mathbb{X} = (0, 1)$

**Theorem 7.6.** If  $X \sim \text{Gamma}(\lambda, \beta)$  and  $Y \sim \text{Gamma}(\xi, \beta)$  are independent then  $X + Y \sim \text{Gamma}(\lambda + \xi, \beta)$

## 8 Estimation

**Notation.** In this section we consider random variables which are known to have a distribution depending on an unknown parameter (so that  $X \sim \text{DIST}(\theta)$  where  $\text{DIST}$  is some distribution).  $\Theta$  is the set of all possible values of  $\theta$ . For properties of  $X$  which depend only on the distribution (essentially all of them), we use the notation  $\mid \theta$  to indicate this dependence. For instance, we write  $P_{X|\theta}(x \mid \theta)$  to mean whatever  $P(X)$  would be if the missing parameter of the distribution were  $\theta$ . Note that this is entirely unrelated to all previous uses of the symbol  $\mid$  in this document.

### 8.1 Estimators

**Notation.** Throughout this section, we consider a set of  $n$  independent and identically distributed random variables  $\underline{X} = (X_1, \dots, X_n)$ .

**Definition 8.1.1.** A **statistic** is a random variable  $T$  which depends on  $\underline{X}$ . The corresponding lowercase letter  $t : \mathbb{R}^n \rightarrow \mathbb{R}$  is used to represent a realised value of  $T$ .

**Definition.** An **estimator** is a statistic used to compute unknown parameters  $\theta$  of the distribution of  $\underline{X}$ . Its realised values are called **estimates**.

#### 8.1.1 Point estimates

**Definition.** A **point estimate** is an estimator which estimates a single unknown parameter  $\theta$ . The official notes call this an estimate even though, according to the previous definition, it is an estimator rather than an estimate. The distribution of the point estimate,  $P_{T|\theta}$ , will depend on the same unknown parameter  $\theta$ .

#### 8.1.2 Bias, Efficiency, Consistency

**Definition.** The **bias** of an estimator  $T$  for a parameter  $\theta$  is

$$\text{bias}(T, \theta) := E(T - \theta \mid \theta) = E(T \mid \theta) - \theta$$

**Definition.**  $T$  is **unbiased**  $\iff \forall \theta \in \Theta, \text{bias}(T, \theta) = 0$ .

**Proposition.** *For any distribution, the mean of a sample is an unbiased estimator for the mean of the distribution.*

**Definition.** The **bias-corrected sample variance** of  $\underline{X}$  is

$$S_{n-1}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

This is an unbiased estimator for the variance of any distribution.

**Definition.** Given two unbiased estimators for the same parameter,  $\hat{\Theta}$  and  $\hat{\Psi}$ ,  $\hat{\Theta}$  is **more efficient** than  $\hat{\Psi}$  iff

$$\left( \forall \theta \in \Theta, \text{Var}(\hat{\Theta} \mid \theta) \leq \text{Var}(\hat{\Psi} \mid \theta) \right) \wedge \left( \exists \theta \in \Theta : \text{Var}(\hat{\Theta} \mid \theta) < \text{Var}(\hat{\Psi} \mid \theta) \right)$$

$\hat{\Theta}$  is **efficient** iff it is more efficient than all other estimators.

**Definition.**  $\hat{\Theta}$  is **consistent** iff it converges in probability to  $\theta$ , that is to say

$$\forall \theta \in \Theta, \forall \varepsilon > 0, \lim_{n \rightarrow \infty} P_{\hat{\Theta}|\theta} \left( \left| \left( \hat{\Theta} \mid \theta \right) - \theta \right| > \varepsilon \right) = 0$$

**Proposition.**  $\hat{\Theta}$  is unbiased  $\implies \hat{\Theta}$  is consistent.

### 8.1.3 Maximum Likelihood Estimation

**Definition.** The **likelihood function** is

$$L(\theta \mid \underline{x}) := \prod_{i=1}^n p_{X|\theta}(x_i)$$

or

$$L(\theta \mid \underline{x}) := \prod_{i=1}^n f_{X|\theta}(x_i)$$

where  $\underline{x} = (x_1, \dots, x_n)$  is a sample of  $\underline{X}$ . Note that this is yet another different usage of  $\mid$ .

**Definition.** The **maximum likelihood estimate** is  $\hat{\theta}_{MLE} := \operatorname{argmax}_{\theta \in \Theta} L(\theta \mid \underline{x})$ .

**Definition.** The **log-likelihood function** is  $\ell(\theta \mid \underline{x}) := \log L(\theta \mid \underline{x})$

**Definition.** The **maximum likelihood estimator** is defined like the maximum likelihood estimate and uses the same notation, but uses the RVs  $\underline{X}$  instead of a specific sample  $\underline{x}$ .

## 8.2 Confidence Intervals

### 8.2.1 Normal Distribution with Known Variance

**Definition.** The  $(1 - \alpha)$  **confidence interval** for the mean  $\mu$  given a known variance  $\sigma^2$  is

$$\left[ \bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

where  $z_\alpha$  is the  $\alpha$ -quantile of  $N(0, 1)$ . Then a sample of size  $n$  with this distribution should have  $\bar{x}$  within this range  $1 - \alpha$  of the time.

### 8.2.2 Normal Distribution with Unknown Variance

**Proposition.** If  $\mu$  and  $\sigma^2$  are both unknown then

$$\frac{\bar{X} - \mu}{S_{n-1}/\mu} \sim \text{Student}(n-1)$$

where

$$S_{n-1} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Then the  $(1 - \alpha)$  confidence level for  $\mu$  is

$$\left[ \bar{x} - t_{n-1, 1-\frac{\alpha}{2}} \frac{s_{n-1}}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\frac{\alpha}{2}} \frac{s_{n-1}}{\sqrt{n}} \right]$$

where  $t_{\nu, \alpha}$  is the  $\alpha$ -quantile of  $\text{Student}(\nu)$ .

### 8.2.3 Another way to view the confidence interval: Neyman construction

**Definition.** The **Neyman construction** is a graph with values of the estimator along the horizontal axis and values of the parameter along the vertical axis. For each value of the parameter, indicate a belt of values in which the estimator is expected to lie for that value. Draw a vertical line at the observed estimate. Then the range of parameter values whose belts intersect this line is the confidence interval.

## 9 Hypothesis Testing

**Definition.** To test an unknown parameter  $\theta \in \Theta$ , partition  $\Theta$  into  $\Theta_0$  and  $\Theta_1$ . The **null hypothesis** is  $H_0 : \theta \in \Theta_0$  (usually chosen to represent the absence of a finding - no correlation, the change has no effect on the data etc). The **alternative hypothesis** is  $H_1 : \theta \in \Theta_1$ . From now on, we use the  $|$  notation from the last chapter with  $H_0$  and  $H_1$  instead of individual values of  $\theta$ .

**Definition.** To perform a hypothesis test, choose a **rejection region**  $R \subseteq \mathbb{R}$  such that  $P(T \in R \mid H_0) = \alpha$  is low. This  $\alpha$  is the **significance level** of the test. Given an observed value  $t(\underline{x})$ , reject the null hypothesis iff  $t \in R$ . The **p-value** of a test is the significance level that lies on the boundary between rejecting and not rejecting the null hypothesis.

### 9.0.1 Error Rates and Power of a Test

**Definition.** **Type I error** occurs when  $H_0$  is rejected, but is actually true. The probability of this is  $\alpha$  by definition. **Type II error** occurs when  $H_0$  is not rejected, but actually  $H_1$  is true. Its probability is  $\beta := P(T \notin R \mid \theta \in \Theta_1)$ .

**Definition 9.0.1.** The **power** of a hypothesis test is  $1 - \beta = P(T \in R \mid \theta \in \Theta_1)$ .

**Definition.** A **simple hypothesis** or **point hypothesis** has the form  $\theta = \theta_0$ . A **composite hypothesis** has the form  $\theta < \theta_0$  or  $\theta > \theta_0$ . A **two-sided test** tests a simple hypothesis against its negation. A **one-sided test** tests a composite hypothesis against its negation (usually the case  $\theta = \theta_0$  is included in  $H_0$ ).

Dist(s)	$H_0$	Test Stat	$R$
$N(\mu, \sigma^2)$	$\mu = \mu_0$	$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\{z \mid  z  > z_{1-\frac{\alpha}{2}}\}$
$N(\mu, \sigma^2)^*$	$\mu = \mu_0$	$\frac{\bar{X} - \mu}{S_{n-1}/\mu}$	$\{t \mid  t  > t_{n-1, 1-\frac{\alpha}{2}}\}$
$N(\mu_X, \sigma_X^2),$ $N(\mu_Y, \sigma_Y^2)$	$\mu_X = \mu_Y$	$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}}$	$\{z \mid  z  > z_{1-\frac{\alpha}{2}}\}$
$N(\mu_X, \sigma^2),$ $N(\mu_Y, \sigma^2)^*$	$\mu_X = \mu_Y$	$\frac{\bar{X} - \bar{Y}}{S_{n_1+n_2-2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$\{t \mid  t  > t_{n_1+n_2-2, 1-\frac{\alpha}{2}}\}$
$\theta \in \mathbb{R}^m,  \mathbb{X}  = k$	$\theta = \theta_0$	$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}^\dagger$	$\{x^2 \mid x^2 > \chi_{k-m-1, 1-\alpha}^2\}$
$ \mathbb{X}  = k,$ $ \mathbb{Y}  = \ell$	$X$ and $Y$ independent	$\sum_{j=1}^\ell \sum_{i=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}^{2\dagger}$	$\{x^2 \mid x^2 > \chi_{(k-1)(\ell-1), 1-\alpha}^2\}$

★:  $\sigma^2$  is not known.

†:  $O_i = (\text{freq}(\underline{X}, x_i), E_i = np_X(x_i))$

‡:  $O_{ij} = \text{freq}((\underline{X}, \underline{Y}), (x_i, y_j)), E_{ij} = \text{freq}(\underline{X}, x_i) \text{freq}(\underline{Y}, y_j)$

## 9.1 Testing for a population mean

### 9.1.1 Normal Distribution with Known Variance

**Proposition.** For  $n$  variables i.i.d with distribution  $N(\mu, \sigma^2)$  with  $\mu$  unknown and  $\sigma^2$  known and a hypothesis  $H_0 : \mu = \mu_0$ , use the test statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \Phi$$

and the rejection region  $R = (-\infty, -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, \infty) = \{z \mid |z| > z_{1-\frac{\alpha}{2}}\}$  has significance level  $\alpha$ . The  $p$ -value is  $2(1 - \Phi(|z|))$ .

### 9.1.2 Normal Distribution with Unknown Variance

**Proposition.** If  $\sigma^2$  is also unknown then we use

$$T = \frac{\bar{X} - \mu}{S_{n-1}/\mu} \sim \text{Student}(n-1)$$

and the rejection region  $R = (-\infty, -t_{n-1, 1-\frac{\alpha}{2}}) \cup (t_{n-1, 1-\frac{\alpha}{2}}, \infty) = \{t \mid |t| > t_{n-1, 1-\frac{\alpha}{2}}\}$ .

**Proposition.**  $\chi^2(k)$  has mean  $k$  and variance  $2k$ .

## 9.2 Testing for differences in population means

### 9.2.1 Two Sample Problems

**Notation.** In this section we have a sample  $\underline{X} = (X_1, \dots, X_{n_1})$  i. i. d with distribution  $P_X$  and another sample  $\underline{Y} = (Y_1, \dots, Y_{n_2})$  i. i. d with distribution  $P_Y$  with  $\mu_X$  and  $\mu_Y$  unknown and the samples independent from each other. In the special case where  $\underline{X}$  and  $\underline{Y}$  are **paired**,  $n_1 = n_2$  and each pair  $(X_i, Y_i)$  is possibly dependent.

### 9.2.2 Normal Distributions with Known Variances

**Proposition.** If  $P_X = N(\mu_X, \sigma_X^2)$  with  $\mu_X$  unknown and  $P_Y = N(\mu_Y, \sigma_Y^2)$  with  $\mu_Y$  unknown, we can test  $H_0 : \mu_X = \mu_Y$  using the test statistic

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim \Phi$$

and the rejection region  $R = \{z \mid |z| > z_{1-\frac{\alpha}{2}}\}$ .

### 9.2.3 Normal Distributions with Unknown Variances

**Definition.** If  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$  unknown, the **bias-corrected pooled sample variance** is

$$\frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2}$$

This is an unbiased estimator for  $\sigma^2$ .

**Proposition.** In the situation above, test  $H_0 : \mu_X = \mu_Y$  using

$$T = \frac{\bar{X} - \bar{Y}}{S_{n_1+n_2-2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{Student}(n_1 + n_2 - 2)$$

and  $R = \{t \mid |t| > t_{n_1+n_2-2, 1-\frac{\alpha}{2}}\}$

## 9.3 Goodness of Fit

### 9.3.1 Count Data and Chi-Square Tests

**Notation.** In this section, we have  $\underline{X}$  with range  $\{x_1, \dots, x_k\}$  and pmf  $p_X(j) = P(X = x_j \mid \theta)$  where  $\theta \in \mathbb{R}^m$  (there are  $m$  unknown parameters which we express as an  $m$ -vector). We write  $p_j \equiv p_X(j)$  for brevity. Consider the observed frequencies  $\underline{O} = (O_1, \dots, O_k)$  with  $O_j = |\{i \in \{1, \dots, n\} \mid X_i = x_j\}|$ . We will not be referring to  $\underline{X}$  directly after this, so just ignore the fact that  $x_i$  can we can't express extracting values from a sample  $\underline{x}$  right now because the notation is the same as the values in the range. For a hypothesis  $H_0 : \theta = \theta_0$ , we can compute the expected probabilities  $\{p_1, \dots, p_k\}$  and thus the expected frequencies  $\underline{E} = (E_1, \dots, E_k)$  with  $E_j = np_j$ .

**Proposition.** In this situation, test  $H_0 : \theta = \theta^n$  using the **chi-squared statistic**

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi_{k-m-1}^2$$

and  $R = \{x^2 \mid x^2 > \chi_{k-m-1, 1-\alpha}^2\}$

### 9.3.2 Proportions

### 9.3.3 Model Checking

### 9.3.4 Independence

**Proposition.** For two DRVs  $X$  and  $Y$  with ranges  $\{x_1, \dots, x_k\}$  and  $\{y_1, \dots, y_\ell\}$ , test  $H_0 : X$  and  $Y$  are independent by creating a **contingency table**:

	$y_1$	$y_2$	$\dots$	$y_\ell$	
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1\ell}$	$n_{1\bullet}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2\ell}$	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\dots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{k\ell}$	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$\dots$	$n_{\bullet \ell}$	$n$

where the  $n_{ij}$  are the observed frequencies and the numbers at the end of each row and column are the sums of those frequencies. Let the expected frequency at each  $i, j$  be  $\frac{n_{i\bullet}n_{\bullet j}}{n}$  and perform a chi-squared test on these observations and expectations. Use the rejection region  $\{x^2 \mid x^2 > \chi^2_{(k-1)(\ell-1), 1-\alpha}\}$ .

### 9.3.5 The $\chi^2$ distribution and degrees of freedom

**Proposition.** The parameter of the  $\chi^2$  distribution represents the size of the sample minus the number of ways in which the expectations depend on the sample.

## 10 Convergence Concepts