
**Faculty of Science**

**School of Mathematics and Statistics**

# MATH5905
# Statistical Inference

Lecture Notes

written by Spiridon Penev

Term 1, 2024

- 3.1 Data reduction in statistical inference
- 3.2 Sufficent partition example
- 3.3 Sufficiency principle
- 3.4 Neyman Fisher factorization criterion
- 3.5 Sufficiency examples
- 3.6 Lehmann and Scheffe's method for constructing a minimal sufficient partition
- 3.7 Minimal sufficient examples
- 3.8 One parameter exponential family densities
- 3.9 Generalization to a $k$- parameter exponential family
- 3.10 Ancillary statistic and ancillarity principle
- 3.11 Ancillary examples

- 10.2 Nonparametric bootstrap
- 10.3 Parametric bootstrap
- 10.4 Numerical illustration
- 10.5 Bootstrap estimate of bias
- 10.6 The jackknife estimate of bias.
- 10.7 Relation of bootstrap and jackknife.
- 10.8 Confidence intervals based on the bootstrap

# 1 Elements of probability

# Who am I?

**Professor Spiridon Penev**

Professor of Statistics at the School of Mathematics and Statistics.

About me:

- Born in Bulgaria
- PhD in Statistics from Humboldt University, Berlin

Research interest is Statistical Inference, specifically nonparametric statistics. Other areas of interest:multivariate analysis, specifically latent variable models. Application domains: finance, risk evaluation and risk management.

# A face to the voice!

# Who are you?

What students do we have here today?

# Contact

**Professor Spiridon Penev**

| | |
|---|---|
| E-mail: | s.penev@unsw.edu.au |
| Telephone | 90655376 |
| Office | Anita B. Lawrence Centre 1038 (first floor ) |
| Web-page: | https://research.unsw.edu.au/people/professor-spiridon-ivanov-penev |

**F2F consultations: 10-11am Tuesday; Online consultations: 10-11am Thursday from the moodle page of the course (link via Virtual Classroom)**; individual consultations (if needed): appointments to be made via email **and conducted using Zoom.**

For administrative problems, contact the **Student Services Office** (Mrs Markie Lugton, m.lugton@unsw.edu.au).

# Lectures

I will give four hours of lectures per week except for week 6.

| | | |
|---|---|---|
| **Tuesday** | **18:00 - 20:00** | **Physics Theatre** |
| **Thursday** | **18:00 - 20:00** | **Physics Theatre** |

# Tutorials

Tutorials and computer labs (using RStudio) for this course are flexible and will be held during the lectures. More precise information will be given during lectures.

**Online materials:** Further information, lecture slides, tutorial questions, and other teaching materials will all be provided on Moodle.

A set of tutorial exercises will be available on Moodle. These problems are for you to enhance your mastery of the course. Some of the problems will be done in lectures, but you will learn a lot more if you try to do them before class.

# Software used

We will use the R software during the term. It is one of the most widely used software for Statistical computation and Graphics.

- Install it on your laptop: https://cran.r-project.org
- Next install RStudio, a nice Graphical User Interface to R: http://www.rstudio.com/products/rstudio/download

# Computer laboratories

Computer laboratories (RC-G012 and RC-M020) are open 9-5 Monday-Friday on teaching days. RC-M020 has extended teaching hours (usually 8:30-9pm Monday-Friday, and 9-5 Monday-Friday on non-teaching weeks).

# Course aims

- The aim of the course is to introduce the main ideas and principles behind parametric and non-parametric inference procedures.
- Both frequentist and Bayesian perspectives will be discussed.
- Estimation, confidence set construction and hypothesis testing are discussed within a decision-theoretic framework.
- Both finite sample optimality and asymptotic optimality will be defined and discussed.
- Computationally intensive methods such as bootstrap are discussed and are compared to asymptotic approximations such as Edgeworth expansions and saddlepoint method.
- Students will learn how to determine appropriate inference procedure and to draw inferences using the chosen procedure.

# Assessment Details

| Task | Due | Weight | Duration |
|---|---|---|---|
| Assignment 1 | End of week 4 | 10% | 2 weeks |
| Assignment 2 | End of week 9 | 10% | 2 weeks |
| Mid-term test | Week 7 (26/03) | 20% | 135 min |
| Final examination | May | 60% | 2 hours* |

(*) Assignments and Mid-term test will be submitted online via the Assignment tool of Moodle. The Mid-term test will be a time-released online assignment. The final will be a pen and paper invigilated exam for all students. Details-TBA

In all assessments marks will be awarded for correct working and appropriate explanations, NOT just for the final answer.

Consult the course outline on Moodle for more details.

# Warning

Late assignments will *not* be accepted! The current deadline structures already accommodate the possibility of unexpected circumstances that may lead students to require additional days for submission. Consequently, the School of Mathematics and Statistics has decided to universally opt out of the Short Extension provision for all its courses, having pre-emptively integrated flexibility into our assessment deadlines.

# Skills to be developed

- Learn how statistical inference arises from the first principles of probability theory;
- Learn the fundamental principles of inference: sufficiency, likelihood, ancillarity, and equivariance;
- Learn the concepts of finite-sample and asymptotic efficiency of an inference procedure;
- Master the parametric and non-parametric delta method, asymptotic normality, Edgeworth expansions and saddlepoint method;
- Be able to estimate key population parameters of interest, to test hypotheses about them and to construct confidence regions;
- Be able to use in practice the parametric, nonparametric, Bayes and robust inference;
- Learn how to use the computer package **R** to generate output for the most common inference procedures and for computer-intensive calculations such as bootstrapping and robust estimation.

# My philosophy for MATH5905

- Lecture notes provide a brief reference source for this course.
- At this stage, these are skeleton lecture notes only. Throughout the course other materials and textbooks will be used for deeper understanding.
- New ideas and skills are first introduced and demonstrated in lectures, then students develop these skills by applying them to specific tasks in tutorials and assessments.
- Computing skills will be used to some extent but this is not a course in computing; the computing part is mainly used to illustrate the theory/methodology.

# Any questions?

Please feel free to interrupt me at any time!

# 1.2 Probability

These lecture notes were originally written and developed by Professor Spiridon Penev.

Standard univariate distributions like binomial, Poisson, normal, Cauchy, logistic, exponential are assumed to be known and are summarised in the Table of Common Distributions on pages 621–626 of **CB**. A copy can be found on Moodle.

The revision mainly the follows the sections of the CB reference.

## 1.2.1 Events and probabilities

An experiment that includes randomness can be modelled with probabilities. An event $A$, for example: It will be raining tomorrow is assigned a probability, $P(A)$, which is a number between $0$ and $1$. Here, the certain event has probability $1$, while the impossible event has probability $0$.

In the simplest probabilistic model, there is a finite number $m$ of possibilities (often called outcomes) and each of them has the same probability $1/m$.

Furthermore, a collection of $k$ outcomes, where $k$ is less than or equal to $m$, is called an event $A$ and its probability is calculated as $k/m$. That is:

$$P(A) = \frac{\text{the number of outcomes in A}}{\text{the total number of outcomes}} = \frac{k}{m}.$$

## Example 1.1

Suppose there are $n$ people in a Zoom meeting.

i) Find the probability that at least two people have the same birthday.

ii) Calculate the probability for $n = 22$.

iii) Calculate the probability for $n = 23$.

**Solution:**

(i) Let $A_n$ be the event that at least two people have the same birthday. The number of outcomes **not** in the event $A_n$ is

$$k = 365 \times 364 \times \cdots \times (365 - n + 1).$$

The total number of possible outcomes in the sample space of all birthday combinations is

$$m = 365 \times 365 \times \ldots 365 = 365^n.$$

The probability that all birthdays are distinct is

$$P(A_n^c) = \frac{k}{m} = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

Hence the probability that two or more people have the same birthday in the Zoom meeting is

$$P(A_n) = 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

ii) For $n = 22$ this probability is

```
n <- 22
1 - prod(365:(365 - n + 1))/365^n
```

```
#> [1] 0.4756953
```

that is $P(A_{22}) = 0.48$.

iii) For $n = 23$ this probability is

```
n <- 23
1 - prod(365:(365 - n + 1))/365^n
```

```
#> [1] 0.5072972
```

that is $P(A_{23}) = 0.51$.

## Exercise 1.1 (revision)

Suppose there are $n$ people in a Zoom meeting.

i) Find the probability that at least one person has the same birthday as you.

ii) Find the value of $n$, that is the number of people needed in the Zoom meeting, so that the probability that at least one person has the same birthday as you approaches $\frac{1}{2}$.

# 1.2.2 Conditional probability and independence

We want to define conditional probabilities $P(A|B)$, which are calculated by updating the probability $P(A)$ of a particular event under the additional information that a second event $B$ has occurred.

Such conditional probabilities can be calculated as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

where $P(A \cap B)$ is the probability that both $A$ and $B$ occur.

Additionally, when events $A$ and $B$ are independent then

$$P(A \cap B) = P(A)P(B) \quad \text{and} \quad P(A|B) = P(A).$$

## Example 1.2

Four cards are dealt from the top of a well-shuffled deck of 52 playing cards. Find the probability that all four cards are aces.

**Solution:**

We can calculate this probability by the methods of the previous section. The number of distinct groups of four cards is

$$\binom{52}{4} = 270725$$

Only one of these groups consists of the four aces and every group is equally likely, so the probability of being dealt all four aces is

$$\frac{1}{270725}.$$

Another way to calculate this probability is to first consider the probability that the first card is an ace, which is $4/52$. Given that the first card is an ace, the probability that the second card is an ace is $3/51$. Continuing this argument, we obtain:

$$\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{270725}$$

## Example 1.3

Let us now see how the conditional probabilities change given that some aces have already been drawn. Four cards will again be dealt from a well-shuffled deck, and we now calculate:

$$P(\text{4 aces in 4 cards} \,|\, i \text{ aces in } i \text{ cards}),\ i = 1, 2, 3$$

The event "4 aces in 4 cards" is a subset of the event "$i$ aces in $i$ cards". Hence, from the definition of conditional probability we have:

$$\begin{aligned} & P(\text{4 aces in 4 cards} \,|\, i \text{ aces in } i \text{ cards}) \\ &= \frac{P(\{\text{4 aces in 4 cards }\} \cap \{i \text{ aces in } i \text{ cards}\})}{P(i \text{ aces in } i \text{ cards})} \\ &= \frac{P(\text{4 aces in 4 cards })}{P(i \text{ aces in } i \text{ cards})} \end{aligned}$$

The numerator has already been calculated and the denominator can be calculated using a similar argument.

$$P(i \text{ aces in } i \text{ cards}) = \frac{k}{m} = \frac{\binom{4}{i}}{\binom{52}{i}}$$

Hence, the conditional probability is

$$\begin{aligned} P(4 \text{ aces in } 4 \text{ cards } | i \text{ aces in } i \text{ cards}) &= \frac{\binom{52}{i}}{\binom{52}{4}\binom{4}{i}} \\ &= \frac{(4-i)!48!}{(52-i)!} \\ &= \frac{1}{\binom{52-i}{4-i}} \end{aligned}$$

For $i = 1, 2$ and $3$ the conditional probabilities are $0.00005$, $0.00082$ and $0.02041$.

```
i <- 1
1/choose(52 - i, 4 - i)
```

```
#> [1] 4.801921e-05
```

```
i <- 2
1/choose(52 - i, 4 - i)
```

```
#> [1] 0.0008163265
```

```
i <- 3
1/choose(52 - i, 4 - i)
```

```
#> [1] 0.02040816
```

# 1.2.3 Bayes' Theorem

Consider the following two equations that arise from the definition of the conditional probability:

$$P(A \cap B) = P(A|B)P(B) \quad \text{and} \quad P(A \cap B) = P(B|A)P(A)$$

Equating and rearranging these two formulas gives **Bayes' theorem**:

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

Additionally, for a general partition $A_1, A_2, \ldots, A_n$ of the sample space $S$ with $P(A_i) > 0$ for all $i = 1, \ldots, n$, we have

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

for each $j = 1, \ldots, n$ which follows from the law of total probability

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

## Example 1.4

When a coded telegraph message is sent, there are sometimes errors in transmission. In particular, Morse code uses "dots" and "dashes", which are known to occur in the proportion of $3:4$.

This means that for any given symbol:

$$P(\text{dot sent}) = \frac{3}{7} \quad \text{and} \quad P(\text{dash sent}) = \frac{4}{7}.$$

Suppose there is interference on the transmission line, with probability $\frac{1}{8}$ a dot is mistakenly received as a dash, and vice versa. If we receive a dot, what is the probability that a dot was actually transmitted?

**Solution:**

From the information provided we have that

$$P(\text{dot sent}) = \frac{3}{7} \quad \text{and} \quad P(\text{dash sent}) = \frac{4}{7}.$$

Using Bayes' theorem, we can write:

$$P(\text{dot sent} \mid \text{dot received}) = P(\text{dot received} \mid \text{dot sent}) \frac{P(\text{dot sent})}{P(\text{dot received})}.$$

Additionally, we can write

$$\begin{aligned} P(\text{dot received}) &= P(\text{dot received} \cap \text{dot sent}) + P(\text{dot received} \cap \text{dash sent}) \\ &= P(\text{dot received} \mid \text{dot sent})P(\text{dot sent}) + \\ &\qquad P(\text{dot received} \mid \text{dash sent})P(\text{dash sent}) \\ &= \frac{7}{8} \times \frac{3}{7} + \frac{1}{8} \times \frac{4}{7} \\ &= \frac{25}{56} \end{aligned}$$

Combining these results, we have that the probability of correctly receiving a dot is

$$P(\text{dot sent} \mid \text{dot received}) = \frac{(7/8) \times (3/7)}{25/56} = \frac{21}{25}.$$

## Exercise 1.2 (revision)

Suppose that $5\%$ of men and $0.25\%$ of women are colour-blind. A person is chosen at random and that person is colour-blind. Find the probability that the person is male. Assume males and females to be in equal numbers.

## Exercise 1.3 (at lecture)

Two litters of a particular rodent species have been born, one with two brown-haired and one grey-haired (litter 1), and the other with three brown-haired and two grey-haired (litter 2). We select a litter at random and then select an offspring at random from the selected litter.

i) Find the probability that the animal chosen is brown-haired.

ii) Given that a brown-haired offspring was selected, find the probability that the sampling was from litter 1.

# 1.3.1 Random variables

In many experiments, it is easier to deal with a summary variable, a so-called random variable, than with the original probability structure.

A random variable $X$ is defined as a function from a sample space $S$ into the set of real numbers:

$$X : S \to \mathbb{R}$$

For example, consider an experiment where two dice are thrown, we can define the random variable $X$ as a sum of the numbers rolled.

# 1.3.2 Probability mass function (pmf) for discrete random variables

Let us now consider real-valued realisations $x$ of a discrete random variable $X$. The probability mass function (pmf) of a discrete random variable $X$

$$f(x) = P(X = x),$$

describes the distribution of $X$ by assigning probabilities for the events $\{X = x\}$.

## 1.3.3 Cumulative distribution function (cdf)

A cumulative distribution function (cdf) of a random variable $X$ is defined by:

$$F(x) = P(X \leq x), \text{ for all } x$$

### Theorem 1.1

The function $F(x)$ is a cdf if and only if the following three conditions hold:

i) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

ii) $F(x)$ is a nondecreasing function of $x$

iii) $F(x)$ is right-continuous, that is, for every number $x_0$,

$$\lim_{x \downarrow x_0} F(x) = F(x_0).$$

This theorem is useful to determine whether a function is a valid cdf.

## Example 1.5

Consider the function $F(x) = 1 - (1-p)^x$ where $x = 1, 2, \ldots$ and $0 < p < 1$. Show that the conditions in the above theorem are satisfied.

**Solution:**

First, since $F(x) = 0$ for all $x < 0$,

$$\lim_{x \to -\infty} F(x) = 0$$

and

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} 1 - (1-p)^x = 1$$

where $x$ goes through only integer values when this limit is taken. To verify property (ii) we note that the sum: $F(x) = \sum_{i=1}^{x} (1-p)^{i-1} p$ contains more positive terms as $x$ increases.

# 1.3.4 Probability density function (pdf) Theorem

The probability density function (pdf) $f(x)$ of a continuous random variable $X$ is a function that satisfies

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{for all } x$$

## Theorem 1.2

A function $f(x)$ is a pdf (or pmf) of a random variable $X$ if and only if the following two conditions hold:

i) $f(x) \geq 0$ for all $x$

ii) $\sum_x f(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f(x)dx = 1$ (pdf)

# 1.3.5 Transformations

We now focus on transformations of random variables, which consider a function of a random variable $X$ with a known cdf $F(x)$. We will often be able to gain complete knowledge about the distribution of the transformed variable, or in other cases, will be able to acquire some understanding of the average behaviour of this transformed random variable.

Note that if $X$ is a random variable with a cdf $F(x)$, then any function of $X$, such that $Y = g(X)$, is also a random variable.

We introduce a subscript in the notation of a cdf and pdf to distinguish between two different random variables $X$ and $Y$.

## Theorem 1.3

Let X be a random variable with cdf function $F_X(.)$ and density $f_X(.)$. Let $Y = g(X)$ and $F_Y(.)$ be the cdf of $Y$. Put

$$S_X = \{x : f_X(x) > 0\} \quad \text{and} \quad S_Y = \{y : y = g(x) \text{ for some } x \in S_X.\}$$

i) If $g$ is increasing on $S_X$ then $F_Y(y) = F_X(g^{-1}(y))$ for $y \in S_y$.

ii) If $g$ is decreasing on $S_X$ and $X$ is continuous random variable then $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in S_y$.

**Proof:** at lecture.

## Example 1.6

Let $X$ be a uniformly distributed random variable $X \sim U(0,1)$ with the density function $f_X(x) = 1$ if $0 < x < 1$ and $0$ otherwise. Find the density of the transformed variable

$$Y = g(X) = -\log(X).$$

**Solution:**

The cdf is $F_X(x) = x$ when $0 \leq x \leq 1$. We now make a transformation:

$$Y = g(X) = -\log(X)$$

We can easily verify that $g(x)$ is a decreasing function of $x$ and since $X$ ranges from $0$ to $1$, $Y = -\log(X)$ ranges from $0$ to $\infty$, that is $S_Y = (0, \infty)$.

For $y > 0$, $y = -\log(x)$ implies $x = e^{-y}$, hence $g^{-1}(y) = e^{-y}$.

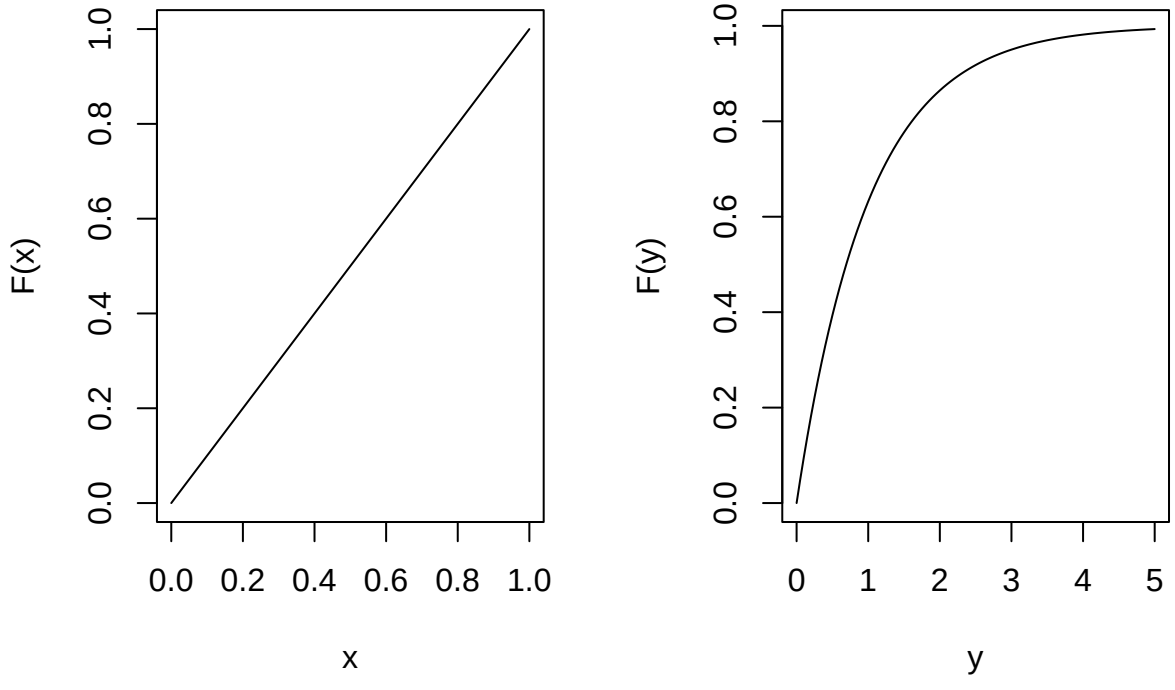
Hence, using the above theorem, we have for $y > 0$

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

Additionally, $F_Y(y) = 0$ for $y \leq 0$. We recognise $F_Y(y)$ as the cdf of the standard exponential distribution.

In summary, the $-\log$ transformed uniform $[0, 1]$ random variable is standard exponentially distributed.

```
par(mfrow = c(1,2))
curve(punif(x), 0, 1, ylab = "F(x)")
curve(pexp(x), 0, 5, xlab = "y", ylab = "F(y)")
```



# 1.3.6 Density transformation formula

## Theorem 1.4

Let $X$ have a pdf $f_X(x)$ and let $Y = g(X)$, where $g$ is a monotone function. Let $S_X$ and $S_Y$ be as above and suppose that $f_X(x)$ is continuous on $S_X$ and that $g^{-1}(y)$ has a continuous derivative on $S_Y$. Then the pdf of $Y$ is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & \text{for } y \in S_Y \\ 0 & \text{otherwise.} \end{cases}$$

**Proof:** Follows from Theorem 1.3 by taking the derivative of the cdf $F_Y(y)$ with respect to $y$ to obtain the pdf $f_Y(y)$. □

## Example 1.7

Let $f_X(x)$ be the gamma pdf

$$f_X(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, \quad 0 < x < \infty$$

where $\beta$ is a positive constant and $n$ is a positive integer. Find the pdf of transformed variable $g(X) = 1/X$.

**Solution:**

We note first that $S_Y = S_X = (0, \infty)$. Furthermore, if $y = g(x)$, then $g^{-1}(y) = 1/y$ and $\frac{d}{dy} g^{-1}(y) = -1/y^2$. Applying the density transformation formula for $y \in (0, \infty)$ we obtain:

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y}\right)^{n-1} e^{-1/(\beta y)} \frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y}\right)^{n+1} e^{-1/(\beta y)} \end{aligned}$$

which is a special case of an inverse gamma pdf.

# 1.3.7 Probability integral transform

**Theorem 1.5 (Probability integral transform)**

Let $X$ be a continuous random variable with a cdf $F_X(.)$. The the random variable $Y = F_X(X)$ is uniformly distributed on $[0, 1]$.

**Proof:**

By applying Theorem 1.3 and noting that $F_x$ is a continuous and monotone increasing transformation, we have that

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

and $S_Y = \{0 \leq y \leq 1\}$. Hence $Y$ is uniformly distributed on $[0, 1]$. □

This fact is, in particular, useful in random number generation from a given distribution. If it is required to generate an observation $X$ from a population with cdf $F_X(x)$, we need only to generate a uniform random number $U$, between $0$ and $1$ and solve for $x$ in the equation $F_X(x) = u$.

Clearly there are often more computationally efficient methods for random number generation. However, this method is still useful because of its general applicability.

# 1.4.1 Expected values

The expected value of a distribution can be understood as a measure of the centre of a distribution, which is obtained by weighting the values of the random variable according to the probability distribution.

The formal definition states, that the expected value or mean of a random variable $g(X)$, denoted by $\mathbb{E}(g(X))$ is:

$$\mathbb{E}(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in S_x} g(x) f_X(x) = \sum_{x \in S_x} g(x) P(X = x) & \text{if } X \text{ is discrete} \end{cases}$$

given that the interval or sum exists.

## Example 1.8

Suppose $X$ has an exponential $\lambda$ distribution, with density

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad 0 \le x < \infty, \quad \lambda > 0.$$

Determine the expected value of $X$.

**Solution:**

The $\mathbb{E}(X)$ is given by

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\infty \frac{1}{\lambda} x e^{-x/\lambda} dx \\
&= -x e^{-x/\lambda} \Big|_0^\infty + \int_0^\infty e^{-x/\lambda} dx \qquad \text{(integration by parts)} \\
&= \int_0^\infty e^{-x/\lambda} dx = \lambda.
\end{aligned}$$

## 1.4.2 Moments

Moments of a distribution are an important class of expectations. For each integer $n$, the $n$th moment of $X$ is

$$\mu_n' = \mathbb{E}(X^n).$$

The $n$th central moment of $X$, $\mu_n$, is defined by

$$\mu_n = \mathbb{E}((X-\mu)^n),$$

where $\mu = \mu_1' = \mathbb{E}(X)$.

In particular, the variance of a random variable $X$ is its second central moment,

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

Hence, the variance is a measure of the degree of spread of a distribution around its mean.

## Exercise 1.4 (revision)

Let $\mu_n$ denote the $n$th central moment of a random variable $X$. Two quantities of interest, in addition to mean and variance, are

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}} \quad \text{and} \quad \alpha_4 = \frac{\mu_4}{\mu_2^2}$$

The value of $\alpha_3$ is called the skewness and $\alpha_4$ the kurtosis. The skewness measures the lack of symmetry in the pdf. The kurtosis measures the peakedness or flatness of the pdf.

i) Show that if a pdf is symmetric about a point $a$, then $\alpha_3 = 0$. Hint: Show that $\mu_3 = 0$ for a general density function $f(x)$.

ii) Calculate $\alpha_4$ for the following:

$$f(x) = \frac{1}{2}, \quad -1 < x < 1.$$

# 1.4.3 Covariance and correlation

Covariance and correlation are given by the following formulas:

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))] \\ &= \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

$$\begin{aligned} \rho &= Cor(X,Y) \\ &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \end{aligned}$$

# 1.5.1 Random vector

Now we will generalise the concepts of cumulative distribution function, probability mass function and density function for univariate random variables to allow multivariate modelling. Finally, we also explain how the independence of random variables relates to the construction of appropriate models.

Let us now consider a random vector

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \in R^p$$

where $p \geq 2$ and there are $p$ different components, each of which is a random variable with a cumulative distribution function $F_{X_i}(x_i), i = 1, 2, \ldots, p$. Each of the functions $F_{X_i}(x_i)$ is called a marginal distribution.

# 1.5.2 Joint cumulative distribution Function

The joint cdf of the random vector $\boldsymbol{X}$ is:

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{x}) &= P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p) \\ &= F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_p) \end{aligned}$$

## 1.5.3 Joint probability mass/density function

In case of a discrete vector $\boldsymbol{X}$ the probability mass function is defined as

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p)$$

If a density $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_p)$ exists such that

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f_{\boldsymbol{X}}(\boldsymbol{t}) dt_1 \ldots dt_p$$

then $\boldsymbol{X}$ is a continuous random vector with a joint density function of $p$ arguments $f_{\boldsymbol{X}}(\boldsymbol{x})$. In this case the following holds:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^p F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \partial x_2 \ldots \partial x_p}$$

If $\boldsymbol{X}$ has $p$ independent components then

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_p}(x_p)$$

holds and, equivalently, also

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{X_1}(x_1)P_{X_2}(x_2)\ldots P_{X_p}(x_p)$$

and

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_p}(x_p)$$

holds.

## Exercise 1.5 (revision)

A pdf is defined by

$$f(x,y) = \begin{cases} C(x+2y) & \text{if} \quad 0 < y < 1 \quad \text{and} \quad 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of C.

# 1.5.4 Marginal distributions

The previous slides defined what a joint distribution of a random vector $\mathbf{X}$ is. This section will explain how to obtain marginal distribution or conditional distribution for some components of the random vector $\mathbf{X}$ when the joint distribution of $\mathbf{X}$ is known.

The marginal cdf of the first $k < p$ components of the vector $\mathbf{X}$ is defined in a natural way as follows:

$$\begin{aligned} &P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k) \\ &\quad = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k, X_{k+1} \leq \infty, \ldots, X_p \leq \infty) \\ &\quad = F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_k, \infty, \infty, \ldots, \infty) \end{aligned}$$

The marginal density of the first $k$ components can be obtained by partial differentiation:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_p) dx_{k+1} \ldots dx_p$$

## Exercise 1.6 (revision)

Consider the pdf:

$$f(x,y) = \begin{cases} C(x+2y) & \text{if} \quad 0 < y < 1 \quad \text{and} \quad 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal distribution of $X$.

## 1.5.5 Conditional distributions

The conditional density $X$ when $X_{r+1} = x_{r+1}, \ldots, X_p = x_p$ is defined by

$$f_{(X_1,\ldots,X_r|X_{r+1},\ldots,X_p)}(x_1,\ldots,x_r|x_{r+1},\ldots,x_p) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{X_{r+1},\ldots,X_p}(x_{r+1},\ldots,x_p)}$$

The above conditional density is interpreted as the joint density of $X_1, \ldots, X_r$ when $X_{r+1} = x_{r+1}, \ldots, X_p = x_p$ and is only defined when $f_{X_{r+1},\ldots,X_p}(x_{r+1},\ldots,x_p) \neq 0$.

We note that, in the case of mutual independence the $p$ components, all conditional distributions do not depend on the conditions and it holds:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} F_{X_i}(x_i) \quad \text{and} \quad f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} f_{X_i}(x_i).$$

# 1.5.6 Moments

Given the density $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the random vector $\boldsymbol{X}$ the joint moments of order $s_1, s_2 \ldots, s_p$ :

$$E(X_1^{s_1} \ldots X_p^{s_p}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{s_1} \ldots x_p^{s_p} f_{\boldsymbol{X}}(x_1, \ldots, x_p) dx_1 \ldots dx_p$$

# 1.5.7 Density transformation formula

Assume that the $p$ existing random variables $X_1, X_2, \ldots, X_p$ with given density $f_{\boldsymbol{X}}(\boldsymbol{x})$ have been transformed by a smooth (i.e. differentiable) one-to-one transformation into $p$ new random variables $Y_1, Y_2, \ldots, Y_p$, i.e. a new random vector $\boldsymbol{Y} \in \mathbb{R}^p$ has been created by calculating:

$$Y_i = y_i(X_1, X_2, \ldots, X_p), \quad i = 1, 2, \ldots, p$$

The question is how to calculate the density $g_{\boldsymbol{Y}}(\boldsymbol{y})$ of $\boldsymbol{Y}$ by knowing the transformation functions $y_i(X_1, X_2, \ldots, X_p), i = 1, 2, \ldots, p$ and the density $f_{\mathbf{X}}(\mathbf{x})$ of the original random vector.

Since the transformation of the $X$ into $Y$ is assumed to be one-to-one, its inverse transformation $X_i = x_i(Y_1, Y_2, \ldots, Y_p), i = 1, 2, \ldots, p$ also exists and then the following density transformation formula applies:

$$g_{\boldsymbol{Y}}(y_1, \ldots, y_p) = f_{\boldsymbol{X}}\Big(x_1(y_1, \ldots, y_p), \ldots, x_p(y_1, \ldots, y_p)\Big)\Big|J(y_1, \ldots, y_p)\Big|$$

where $J(y_1, \ldots, y_p)$ is the Jacobian of the transformation:

$$J(y_1, \ldots, y_p) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_p} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}$$

## Exercise 1.7 (exercise)

Given the pdf

$$f(x,y) = \begin{cases} C(x+2y) & \text{if} \quad 0 < y < 1 \quad \text{and} \quad 0 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Find the pdf of the random variable

$$Z = \frac{9}{(X+1)^2}.$$

Notice that this is a one-dimensional density transformation.

## Exercise 1.8 (at lecture)

Let $X$ and $Y$ be independent, standard normal random variables. Consider the transformation $U = X + Y$ and $V = X - Y$. Find the joint density of $U$ and $V$.

# 1.5.8 Multivariate normal distribution

We will only need the **non-degenerated** multivariate normal.

Consider the term

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

in the univariate case of

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$$

which can generalize to:

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}).$$

Here $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{X} \in \mathbb{R}^p$ is the expected value of the random vector $\boldsymbol{X} \in \mathbb{R}^p$ and the matrix

$$\Sigma = \mathbb{E}(X-\mu)(X-\mu)' = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \in \mathcal{M}_{p,p}$$

is the covariance matrix (assumed to be positive definite).

- On the diagonals are the variances of the $p$ random variables sometimes we simply denote $\sigma_{ii}$ by $\sigma_i^2$;
- $\sigma_{ij}, i \neq j$ are the covariances.

The final result:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{\frac{1}{2}}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}, -\infty < x_i < \infty, i = 1, 2 \dots, p$$