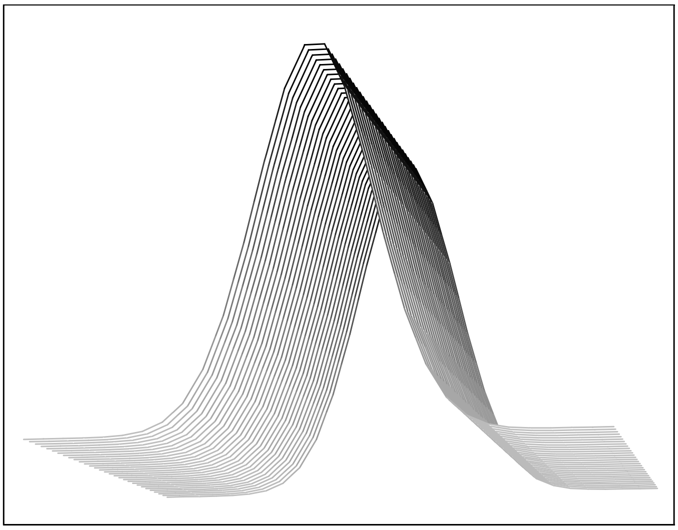
Department of Statistics

STA1503

DISTRIBUTION THEORY I



STUDY GUIDE

Contents

CHAPTER 1

General comments

A word of welcome to the module $STA1503$: *Distribution Theory I*. This is a newly developed module, starting to run for the first time in 2010. *Distribution Theory I* is offered at level 5 of the National Qualifications Framework (NQF), as specified by the Department of Education.

If you are enrolled for a BSc degree with a major in statistics, $STA1503$ *Distribution theory I*, is the third of the three first-level modules. It is also the prerequisite for the second-level module in distribution theory: $STA2603$ *Distribution theory II.* In fact, there are three modules in distribution theory, one at each of the three levels of your degree. The third level module in distribution theory is $STA3703$ *Distribution theory III.* The knowledge given to you in *Distribution theory I* forms the basis for *Distribution theory II* and the combined knowledge in *Distribution theory I* and *Distribution theory II* forms the basis for *Distribution theory III.*

Distribution theory relies heavily on mathematical logic and makes use of different mathematical techniques, e.g. calculus. The prescribed textbook for $STA1503$ also mentions a mathematical prerequisite, namely a thorough knowledge of "first-year college mathematics" and that is why you should either have completed a mathematical course in calculus (first level) or be enrolled for such a course while doing this module.

1.1 The purpose of this module

We would like you to have a clear understanding of the nature of statistics and understand that the objective of statistics is *statistical inference*. Although the emphasis in this module is on the theory behind statistical inference, the relationship between the different topics in distribution theory and practical applications of that knowledge will become clear. Once you have a grip on the knowledge given in this module you will understand how *knowledge* of the distribution of a particular variable can be applied to determine the mean, the variance and several other moments and how the *relationship* between different distributions can be applied to generate data values having specific distributions. As you continue with the different modules in statistics you will see how distribution theory forms the basis for statistical inference, e.g. in testing hypotheses about means, variances, correlation coefficients and regression coefficients.

1.2 Tutorial matter for this module

The tutorial material that you have at your disposal for this module consists of the following:

- the prescribed textbook, *Mathematical Statistics with Applications*
- this study guide as a supplementary tool to the textbook
- three assignments that have to be submitted and contribute towards your year mark
- different other tutorial letters containing e.g. model solutions to the assignments, a trial examination paper and other information about the examination

You have to buy the *prescribed book* from an official UNISA bookseller. We realize that textbooks are expensive articles. However, we believe that this prescribed book is a good reference book that belongs on any statistician's bookshelf and *is also prescribed for other modules*. So you should consider buying this book as an investment in you career and do not sell it once you have completed this module.

The *study guide* is designed to guide you through the relevant sections in the *textbook* in a structured and logical way. This means that the textbook and the study guide should be treated as "inseparable" and that you should use them concurrently as you proceed through the syllabus, section by section. The study guide only will not give you sufficient direction to enable you to pass the examination as its function is to *supplements* the textbook – it is not a *substitution* for the textbook. Maybe it should rather have been called a *textbook guide*!

The three assignments will be included in Tutorial letter 101, which you should have received together with this study guide. Take serious note of the due dates for submission of assignments as submission of an assignment after the due date will most probably result in it not being marked by your lecturer.
There are two lines of tutorial letters, namely a 100-series and a 200-series. Letters with the 100-code give *information* (e.g. Tutorial letter 101, 102...), while letters in the 200-series contain *solutions* (e.g. Tutorial letter $201, 202$, etc.).

1.3 A study programme for STA1503

When you start with this module we assume that you have knowledge of the basic statistics as contained in $STA1501$ and $STA1502$ as well as the mathematics mentioned before. You should aim to read, write and understand mathematical terms and symbols with ease. You may be surprised to see how the different concepts interlink as you do the different modules in your degree. Please realize from the start that you cannot open the tutorial matter for any statistics module only a few days before the examination and expect to pass! The nature of the work is such that you will have to work continuously and refer back to previous knowledge the moment you feel at a loss about a concept. Statistics is a *hands-on* discipline and you should always have paper and pen ready to write down important definitions, theorems or proofs of theorems. Sometimes a mathematical manipulation looks fairly simple when you see it in written form. However, if you close the book and try to derive the same result without peeping, it usually is a different kettle of fish!

It is a good idea to set up a schedule so that you can pace your progress through the tutorial matter. It is difficult to give a general indication of the hours needed as it differs from one individual to the next. You can use the following information as a guideline when you set up your study program, but do not be upset if it takes you longer, as it may even indicate that you are very clever and will reap the benefits of in-depth studies in this and other modules.

- 48 hours on reading and working through the tutorial matter (the textbook and the study guide)
- 36 hours on doing the three assignments
- 36 hours on preparing for the examination

Important facts:

- This is a semester module. A semester is very short and there is no time to waste.
- Assessment (assignments and examination) is in the form of essay-type questions - *not* multiple choice as in $STA1501$ and $STA1502$. In multiple choice assessment only answers are significant and this is often seen as "unfair", so here you have the benefit that calculations, i.e. how you get to the answer, are also evaluated. However, be careful as logic and notation are now very important - a correct answer does not necessarily imply full marks for a question! You will soon realize that both the methods of assessment (multiple choice and essay-type questions) have advantages and disadvantages!
- If you are not able to do the assignments it is a very serious matter as you will most probably not pass the examination.
- You will receive a trial paper, which serves as an illustration of typical questions you can expect. Complete this as a "test" under examination conditions, i.e. put away the study guide and the textbook and complete it within 2 hours!

1.4 Using technology in STA1503

A statistician should never be without at least a scientific pocket calculator with mathematical and statistical functions. In fact, throughout this module we will assume that you have a suitable calculator available for assignment and examination work. Make sure that you have studied the manual of that particular calculator as different operational methods may apply for the different models. Buy it in time and use it often as nervousness during the examination can lead to errors and can cost you marks!

However, in this computer age it is also important that we prepare you for the real-life world of a statistician. Having access to a personal computer is not a prerequisite for $STA1503$. However, in Keller, the prescribed book for $STA1501$ and $STA1502$, there are many examples of calculations using Excel and Minitab software. Try to familiarize yourself with these methods (or use other statistical software). In the case of distribution theory it is very useful to enter different values for the parameters of a specific distribution (e.g. different values of μ and σ for the normal distribution) and see on screen how these changes influence the form and placement of the curves. My experience is that young students are curious and very clever in using technology – approach statistical software in the same way as you do when you play around with your cell phone in an attempt to find out what it is capable of!

1.5 The Textbook

The prescribed book for $STA1503$ is:

Wackerly, Mendenhall & Scheaffer. (2008, Seventh Edition)
Mathematical Statistics with applications
Thomson Books/Cole, International student edition.

From this book the study material for this module is covered in the following chapters:

Chapter 2:	Probability
Chapter 3:	Discrete random variables and their probability distributions (except 3.10)
Chapter 4:	Continuous random variables and their probability distributions (except 4.11)
Chapter 5:	Multivariate probability distributions (only up to 5.7)
Chapter 6:	Functions of random variables (only up to 6.5)

Note:

The same book is prescribed for STA2602 (*Statistical inference*), STA2610 (*Statistical distributions*) and as reference in STA2603 (*Distribution theory II*).

Please note that even though you will not be examined on the contents of Chapter 1, is forms the introduction to the later chapters. I advise you to read through it as reference is made at different places to facts given in Chapter 1.

1.5.1 Reference to the textbook

We have already given you the bibliographic particulars of the book by Wackerly, Mendenhall & Schaeffer, which is the prescribed book for this module. Whenever we refer to this book in the study guide or tutorial letters, we will refer to "**the textbook**" or to "**WM&S".** It will help us if you also do the same in your correspondence with us and in your assignments.
Throughout this guide *a few* of the questions in **WM&S** will be done as examples. Also take note that the answers to all odd questions are given at the back of the book.

1.5.2 Format of the textbook

Before we tackle the work covered in the textbook we will briefly outline the format of the textbook and point out mathematical and other conventions followed in the book. The terminology used by authors of different statistical books varies from book to book. To make it easier for you, we have *not* divided the sections in this guide into *study units* (as in many other modules), but we are using the exact same numbering per topic and subsection as in **WM&S.**

In order to explain the format used in **WM&S,** the table of Contents (right at the beginning) is a good starting point to give you an overview of all the concepts covered. There you will see that each chapter is subdivided into sections. For example, in chapter $\mathbf{2}$, these subsections are numbered "$\mathbf{2}.1$ Introduction", $\mathbf{2}.2$, and so on. Within the sections you will find EXAMPLES, FIGURES, DEFINITIONS, THEOREMS and/or EXERCISES - all numbered in a specific way to assist you. More detail follows.

- Consider EXAMPLE 4.3 : The 4 tells you this is an example in the *fourth* chapter (4 is the digit before the decimal) and 3 tells you that this is the *third* example in that chapter (3 is the digit after the decimal). The beginning and the end of an example are indicated by horizontal lines.

- In the notation for the *figures* in the textbook the same notation is followed.

- In mathematics and statistics the *definition* of a concept is usually the starting point when you solve a problem. In defining a *definition* we read on Wikipedia that " A definition cannot be generated, or used, without the existence of a system, which is organized and structured above a certain level of chaos."

  The notation for definitions is the same as for examples and figures. Furthermore you can spot them very easily as they are place in shaded blocks. You should be able to write down a definition of any concept covered in this module – make them your own, i.e. do not learn a definition by heart as you would a verse in a poem!

- A *theorem* implies serious business in mathematics and statistics! Armed with theorems one can solve many a problem. A theorem is followed by the proof of the theorem. Unless specified otherwise in this study guide, assume that you have to *know all the theorems and their proofs* in the relevant chapters prescribed in this module..

- *Exercises* are placed at the end of most sections right through **WM&S.** The questions are numbered consecutively within the different chapters. In this numbering the digit before the decimal will correspond with the chapter number. The digit after the decimal therefore counts the consecutive questions, regardless of the subsection it is in. So, when you are referred to Exercise 2.38 you know that it is in chapter 2, but you have no idea of the subsection it falls under.
- At the end of each chapter you will find a *Summary* of the knowledge covered in that chapter; *References and Further Readings*, where publications on the topics covered in that chapter are given; lastly, you have more exercises under the heading *Supplementary Exercises*. What more could you want!

1.5.3 Terminology used in the textbook

In statistical terminology there are certain "unwritten rules". You have to know, for example, that we use the uppercase form of an alphabet letter to indicate a *random variable,* while lowercase alphabet letters are used when we refer to the possible values that a random variable can assume within the *sample space* of that a random variable. Even if you have no idea of what I am talking about at this stage, consider the following:
In the expression

$$E(Y) = \sum_y yp(y)$$

it is clear that the alphabet letter Y is a capital letter in $E(Y)$ and a lower case letter (y) in all the other positions. Is this use of the form of the letter a random process, or is the first Y in capital form because of its placement at the beginning of the expression (the same as when you start a sentence with a capital letter)? *No, there is knowledge behind the use of each upper and lower case of a letter*! The rules for the correct notation will be explained in chapter 3. You should then know never to use nonsense notation in assignments or in the examination. Here are examples of such *incorrect notation*:

$$\begin{aligned}
\textbf{NEVER WRITE} \quad E(y) &= \sum_y yp(y) \qquad OR \\
E(Y) &= \sum_y Yp(y) \qquad OR \\
E(Y) &= \sum_y yp(Y) \qquad OR \\
E(Y) &= \sum_Y yp(y) \qquad OR.....
\end{aligned}$$

There are a few *Activities* included in this study guide. These are tasks that you will be doing for your own benefit and you do not have to submit any of them. Activities will be quite easy to find as they are usually at the end of a section and will be placed between horizontal lines. To illustrate I have placed your first activity just here:

Activity 1.1

You have to be *actively* involved (using a pen or computer keyboard, or) when you study statistics, so why do you not construct a nice file and call it **STA1503**? Inside you need different sections - you can decide on the number and what you call them, but there should at least be a section for *formulas*, one for *definitions* and one for *theorems and their proofs.* Please do not think that you will do this *just before* the examination! Then it may be too late because quite a lot of time goes into making proper summaries. It is also an excellent way of familiarizing yourself with notations, etc. In statistics you do not learn by reading - use all your senses!

1. Why do you need to summarize *formulas*?
   Included with the examination paper will be a formula sheet, but remember, you have to *select the correct* formula for a given question. If you used your own list while studying it will be much easier to identify the different formulas, you will not be overwhelmed by the list you receive in the examination. Furthermore, *not each and every* formula will be on the examination list. So, when you receive the list that will be included in the examination paper, you can mark them off on your list. From this it would be wise to make a new list of formulas you have to learn for the examination.

2. Why write down all the *definitions*?
   You can be asked to define a concept and writing down (or typing) a definition is already a first step in learning and understanding it (provided you concentrate when writing/typing it!).

3. Is it not wasting time to write down a *theorem and its proof*?
   No, you learn much better than just "looking" at a proof. You may even think that you understand a proof after looking at it and trying to "recite" it. Remember that lecturers have lots of experience and will be able to see in small errors you make, when proving a theorem, that you learnt that proof by heart. Even if you should succeed in giving us a flawless "poem" in the examination, you are bluffing yourself! Your problems will come in future statistical studies. The knowledge you accumulate while studying can be considered as bricks in the construction of a building. So, would you really use brittle bricks in the building process if it can only lead to a disaster at a later stage?

CHAPTER 2

Probability

As said earlier, the same numbering of the sections and identical headings as in WM&S will be used in this study guide. In many sections the information in the textbook is sufficient, in which case the heading will be given in this guide, but no further notes or additional explanations. *Please take note that this does not mean that you will not be examined on that section* – all the knowledge in chapters $2,\ 3,\ 4\ 5\ and\ 6$ of the textbook are included for examination purposes. We are not re-writing the textbook – the core of knowledge you need sits in the textbook.

Probability, as discussed in WM&S, Chapter $2,$ must be studied as a whole – nothing is omitted for examination purposes. However, seeing that we assume you have some knowledge of probability based on completion of $STA1501$, you have to concentrate on the "new" knowledge and take serious note of the mathematical approach to probability given in WM&S and in this guide.

STUDY	
WM&S:	**Chapter 2**.

2.1 Introduction

You have been introduced to certain probability concepts in the module $STA1501.$ Some of the sections in chapter 2 will therefore seem quite familiar to you and that is fine. The difference is that WM&S gives you the correct *mathematical approach to probability* and to achieve this, you will be introduced to certain basic concepts of set theory. Remember the three approaches in probability? There is the

1. classical approach (popular to help determine probabilities associated with games of chance)
2. relative frequency approach (the long-run relative frequency with which an outcome occurs)
3. subjective approach (based on the degree of belief in the occurrence of an event)

In WM&S the measure for a belief in the occurrence of an event is based on relative frequency.

2.2 Probability and inference

Read this paragraph to understand the need for a theory of probability to provide the foundation for modern statistical inference.

2.3 A review of set notation

You will see that there are differences between the notation of Keller ($STA1501$) and WM&S. Some are simply notation differences, while others form part of set theory notation.

A^c	for "complement of a set " in $STA1501$ now changes to	$\overline{A}$.
A or B	indicating the "union of two sets" is in set theory notation	$A \cup B$.
A and B	indicating the "intersection of two sets " is in set theory notation	$A \cap B$.

Read through section 2.3 of WM&S very carefully and make a summary of the most important facts that you need to remember. The table below is my personal summary of the topics covered in 2.3. of WM&S that I want to remember – not for a moment or a day or ... but as a foundation for current and future studies. Please make your *own summary* – of special interest is the last column, which should contain your personal "reminders".

Description	**Notation**	**Extra information**
The universal set	S	Includes *all* elements
Subset	$A \subset B$	Not the same as $B \subset A$
Empty (null) set	ϕ	ϕ is a *set*, not an element
Union of two sets	$A \cup B$	Two shaded circles
Intersection of sets	$A \cap B$	Circles must overlap
Complement of a set	$\overline{A}$	Careful: $\overline{A} \cap \overline{B} \neq \overline{A \cap B}$
Disjoint (mutually exclusive)	$A \cap B = \phi$	Two circles do not overlap
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Like $2\,(3+4) = 2 \cdot 3 + 2 \cdot 4$
	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	
De Morgan's law (a)	$\overline{A \cap B} = \overline{A} \cup \overline{B}$	
De Morgan's law (b)	$\overline{A \cup B} = \overline{A} \cap \overline{B}$	

EXERCISE 2.1 of WM&S

To find the elements of the subsets you have to refer to the universal set

$$S = \{FF, FM, MF, MM\}$$

$A = \{FF\}$ Remember to use "curly" brackets to list the elements of a set.

$B = \{MM\}$

$C = \{MF, FM, MM\}$ Catchy one! You have to list both MF and FM because the order in which the children are born results in two possibilities. MM is listed because at least one male means that there could have been one or two male children.

$A \cap B = \phi$ because FF and MM have no common elements.

$A \cup B = \{FF, MM\}$ because you put FF and MM together.

$A \cap C = \phi$ because FF is not in C, and no element of C is in A; no common elements.

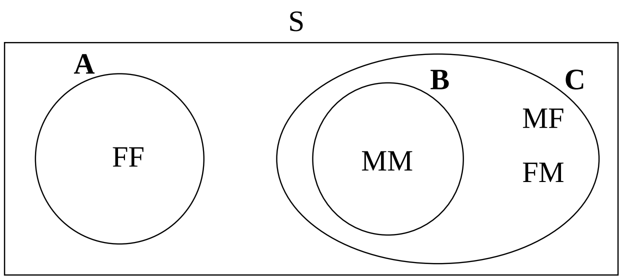
$A \cup C = S$ because you now have all the elements – what was not in C is in A.

$B \cap C = \{MM\}$ because MM occurs as an element in B as well as in C.

$B \cup C = C$ because MM is already an element of C.

$C \cap \overline{B} = \{MF, FM\}$ because MF and FM are the common elements.
Did you have it correct that $\overline{B} = \{FF, MF, FM\}$?

You can draw a Venn diagram of the universal set S and its subsets A, B and C. Note that there are no elements in S that are not in one or more of the sets A, B and C. Furthermore, you can see at a glance that B lies *inside* of C and also that A and C have no overlapping elements – they are disjoint.
The Venn–diagram:



Something else that is interesting *in this example* is that $A \cap C$ and $A \cup C$ are complements, i.e.

$$\overline{A \cap C} = A \cup C = S$$
$$\overline{A \cup C} = A \cap C = \phi$$

2.4 A probabilistic model for an experiment: the discrete case

You have six definitions to learn *and understand*.

In DEFINITION 2.6 the probability of an event A (in the discrete case) is defined, using *three conditions* that must be satisfied in order to assign a probability to an event. The conditions are given in the form of axioms. *The definition of an axiom in Wikipedia is: In traditional logic, an axiom or postulate is a proposition that is not proved or demonstrated but considered to be either self-evident, or subject to necessary decision. Therefore, its truth is taken for granted, and serves as a starting point for deducing and inferring other (theory dependent) truths.*

Axiom 1 A probability is always positive, indicated by ≥ 0.

Axiom 2 Total probabilities over all the events in the sample space is equal to one.

Axiom 3 If none of the events in A overlap, the probability of their union is the same as the summing of the probabilities of the individual events.

EXERCISE 2.15

(a) You are asked to determine $P(E_2)$ and this is done by using axiom 2, namely that the sum of all probabilities must add up to one:

$$\begin{aligned} P(E_1) + P(E_3) + P(E_4) &= 0.01 + 0.09 + 0.81 \\ &= 0.91 \\ P(E_2) &= 1 - 0.91 \\ &= 0.09 \end{aligned}$$

(b) A hit on at least one of the two drillings implies a hit on either the first or on the second drill or on both drills (still *at least* one).

All three these outcomes are disjoint (no overlap), so you *add* the probabilities (axiom 3):

$$\begin{aligned} P(E_1) + P(E_2) + P(E_3) &= 0.01 + 0.09 + 0.09 \\ &= 0.19 \end{aligned}$$

2.5 Calculating the probability of an event: the sample-point method

The *sample-point method* to determine the probability of events defined on a sample space containing a finite (you can count it) or a denumerable (although infinite, you can count them) set of sample points is explained in WM&S through three worked examples.

2.6 Tools for counting sample points

Permutations and combinations are defined in this section and the *first four theorems* of Chapter 2 are given in this section – all four result from combinatorial analysis. Theorems are about the following:

2.1 the $mn-$rule

2.2 calculation of permutations

2.3 partitioning of objects into groups, leading to the multinomial coefficient

2.4 calculation of combinations

EXERCISE 2.49

(a) There are only130 major areas and from these you want to select "pairs" allowing students to have two majors instead of only one. How many *combinations*? Do *not* consider permutations as the order in which the two majors are selected is not important (e.g. Stats and Maths are considered the same selection as Maths and Stats).

$$\begin{aligned}\binom{130}{2} &= \frac{130!}{2!\cdot(130-2)!}\\ &= \frac{(130\cdot 129)}{2\cdot 1}\\ &= 8385\end{aligned}$$

(b) There are 26 letters in the alphabet.

- For a two-letter code there are therefore $26 \cdot 26$ possibilities (any letter may be used twice, so the first letter can be selected in 26 ways and after that the second letter can also be selected in 26 ways).
- For a three-letter code there are $26 \cdot 26 \cdot 26$ possibilities (any letter may even be used three times).

  Calculation: $26 \cdot 26 = 676$ two-letter possibilities and

  $26 \cdot 26 \cdot 26 = 17,576$ three-letter possibilities.

  Therefore a total of $676 + 17,576 = 18,252$ possible two and three–letter code combinations can be made.

(c) Either a single major *or* a double major code implies the union of the possibilities and therefore there are $8385 + 130 = 8515$ required.

(d) Yes, because $8515 << 18,252$.

2.7 Conditional probability and the independence of events

The two definitions in this section are extremely important and you will be using them so often that I am sure you will need little effort to recall them! Link the concepts of conditionality and independence in your mind as they "go together".

It may seem that definition 2.10 contains three unrelated conditions for independence, but that is not the case. It is in fact the same rule, just in different forms:

- Read $P(A \mid B)$ as "the probability of event A occurring, given that event B has already happened". If $P(A \mid B) = P(A)$ you can see that event B is "ignored" – it has no influence on the probability that event A occurs. This is then the indication that events A and B are *independent.*
- So, if events A and B are independent then event A can also not have any influence on the occurrence of event B, so $P(B \mid A) = P(B)$.
- Let me use the original definition for conditional probability to show how we get the third condition for independence given in the definition 2.10:

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

Now, if events A and B are *independent*, then

$$P(A \mid B) = P(A)$$

Substitute $P(A)$ for $P(A \mid B)$ in the first equation, then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \text{ becomes } P(A) = \frac{P(A \cap B)}{P(B)}$$

Cross multiply in this fraction and you have

$$P(A \cap B) = P(A) \cdot P(B)$$

EXERCISE 2.77

(a) The definition of event A only refers to education, so it is irrelevant if the person has been convicted or not, so

$$P(A) = 0.10 + 0.30 = 0.40$$

This probability of event A taking place is therefore simply the *total of the first row*.

(b) The definition of event B only refers to conviction, so the education background of the person is irrelevant, so

$$P(B) = 0.10 + 0.27 = 0.37$$

This probability of event B taking place is therefore simply the *total of the first column*.

(c) $P(A \cap B)$ implies that you have to find the overlapping of events A and B. Take the column under *Convicted* and the row *10 years or more* and where the two overlap (where row one and column one cross), you will find the answer.

$$P(A \cap B) = 0.10$$

(d) $P(A \cup B)$ implies that you have to find the events in A as well as those in B without counting the probability where both take place twice. How? There are two ways of reasoning:

- The row one and column one totals, giving the individual probabilities that the events A and B occur, are respectively 0.40 and 0.37. If you simply add these two values, you have counted the probability that both happen simultaneously ($P(A \cap B) = 0.10$) twice. So, simply subtract one of them and you get:

$$P(A \cup B) = 0.40 \ + 0.37 - 0.10 = 0.67$$

- Go to the table and count the probabilities (do not use the totals)

$$P(A \cup B) = 0.10 \ + 0.27 + 0.30 = 0.67$$

(e) This can also be calculated in two ways; both correct, but in some questions the one method can take much longer than the other one.

$$\begin{aligned} \text{Short option} & : \\ P(\overline{A}) &= 1 - P(A) \\ &= 1 - 0.40 \\ &= 0.60 \end{aligned}$$

As alternative you can look for the probabilities where the event of 10 *years or more* does not take place and add them.

$$\begin{aligned} \text{Longer option} & : \\ P(\overline{A}) &= 0.27 + 0.33 \\ &= 0.60 \end{aligned}$$

(f) This question involves finding the event $A \cup B$ first and then the probability of that event, namely $P(A \cup B)$. Once you have this answer you can subtract the answer from one. In this question you have already calculated $P(A \cup B)$ so the answer should come easily:

$$\begin{aligned} P(A \cup B) &= 0.67 \\ P\left(\overline{A \cup B}\right) &= 1 - P(A \cup B) = 1 - 0.67 \\ &= 0.33 \end{aligned}$$

(g) The same reasoning as in **f** applies here:

$$\begin{aligned} P(A \cap B) &= 0.10 \\ P\left(\overline{A \cap B}\right) &= 1 - 0.10 \\ &= 0.90 \end{aligned}$$

(h) Let us assume that you have not yet determined the probability of event B taking place.

Go to the contingency table and use column *Convicted* (for event B) and find where the row 10 *years or more* crosses this column. The probability that corresponds with this is 0.10. You know this is the probability that A and B occur. You know by now that probability is always given as a proper fraction, so we know that 0.10 has to be in the numerator (the top position in the fraction). What is in the denominator (at the bottom)? For a *conditional probability* the denominator is the *total of the row/column of the event given in the condition*. For this question the condition was that event B must have taken place and the total of column one is therefore in the denominator,

$$\begin{aligned} P(A \mid B) &= \frac{\text{numerator}}{\text{denominator}} \\ &= \frac{0.10}{0.37} = 0.270270.. \end{aligned}$$

Of course you can in this case simply fill in the values you calculated earlier into the definition of conditional probability:

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.10}{0.37} \\ &= 0.270270... \end{aligned}$$

(i) The same reasoning as in **h** applies here. From the contingency table you can read off

$$P(B \mid A) = \frac{0.10}{\mathbf{0.40}} = 0.25$$

Note that the denominator is the *row total* for event A, because the condition was that event A must have taken place.

Using the formula and previously calculated probabilities give the same answer:

$$\begin{aligned} P(B \mid A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{0.10}{0.40} \\ &= 0.25 \end{aligned}$$

2.8 Two laws of probability

The three theorems given in this section are very important and significant tools in probability calculations.
Make sure that you know and understand these laws and make the following links in your mind about theorems 2.5 and 2.6:

- **Independence** of events simplifies the **multiplicative** law of probability.
- **Mutual exclusiveness** of events simplifies the **additive** law of probability.

In my personal summary I will simply have:

Independent $P(A \cap B) = P(A) \cdot P(B)$	$\leftrightarrows$	**Multiplicative law** $P(A \cap B) = P(A) \cdot \mathbf{P}(\mathbf{B} \mid \mathbf{A})$
Mutually exclusive $P(A \cup B) = P(A) + P(B)$	$\leftrightarrows$	**Additive law** $P(A \cup B) = P(A) + P(B) - \mathbf{P}(\mathbf{A} \cap \mathbf{B})$

EXERCISE 2.85

$$\begin{aligned}
P\left(A \mid \overline{B}\right) &= \frac{P\left(A \cap \overline{B}\right)}{P\left(\overline{B}\right)} \\
&= \frac{P\left(\overline{B} \mid A\right) \cdot P(A)}{P\left(\overline{B}\right)} \\
&= \frac{[1 - P(B \mid A)] \cdot P(A)}{P\left(\overline{B}\right)} \\
&= \frac{[1 - P(B)] \cdot P(A)}{P\left(\overline{B}\right)} \\
&= \frac{P\left(\overline{B}\right) \cdot P(A)}{P\left(\overline{B}\right)} \\
&= P(A)
\end{aligned}$$

So, A and $\overline{B}$ are independent.

$$\begin{aligned}
P\left(\overline{B} \mid \overline{A}\right) &= \frac{P\left(\overline{B} \cap \overline{A}\right)}{P\left(\overline{A}\right)} \\
&= \frac{P\left(\overline{A} \mid \overline{B}\right) \cdot P\left(\overline{B}\right)}{P\left(\overline{A}\right)} \\
&= \frac{\left[1 - P\left(A \mid \overline{B}\right)\right] \cdot P\left(\overline{B}\right)}{P\left(\overline{A}\right)}
\end{aligned}$$

but A and $\overline{B}$ are independent (from above), so

$$
\begin{aligned}
P\left(\overline{B} \mid \overline{A}\right) &= \frac{[1-P(A)] \cdot P\left(\overline{B}\right)}{P\left(\overline{A}\right)} \\
&= \frac{P\left(\overline{A}\right) \cdot P\left(\overline{B}\right)}{P\left(\overline{A}\right)} \\
&= P\left(\overline{B}\right)
\end{aligned}
$$

So, $\overline{A}$ and $\overline{B}$ are independent.

2.9 Calculating the probability of an event: the event-composition method

This very helpful method is explained in terms of six examples, 2.17 - 2.22, given in this section. Study them all in detail as the principles given in this section form the basis for the next two very important sections.

EXERCISE 2.119

(a) Define the following events:

A : Obtain a sum of 3.

B : Do not obtain a sum of 3 or of 7.

There are $6 \cdot 6 = 36$ possible outcomes for the rolling of the two dice. To get a sum of 3, however, there is only the combination of *one dot* and *two dots* on the two dice.

The probability of getting a total of three on the two dice is

$P(A) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$ (because you can have the $(1+2)$ and the $(2+1)$ combination).

Think of the possible combinations that would give a total of *seven*:

$(1+6)$; $(2+5)$;$(3+4)$ plus, if the order is changed around $(4+3)$;$(5+2)$; $(6+1)$.

In total there are 6 ways to get a total of 7.

$P(\overline{B}) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36}$ giving $P(B) = 1 - \frac{8}{36} = \frac{28}{36}$.

Obtaining a 3 before a sum of 7 can happen in the 1st roll, the 2nd roll, the 3rd roll, etc.

Using the defined events, we can say these rolls will be $A,\ BA,\ BBA,\ BBBA$, etc.

In terms of probabilities we assume that the different rolls are independent, so we have to determine

$$
\begin{aligned}
& P(A) + P(B)P(A) + [P(B)]^2\, P(A) + [P(B)]^3\, P(A) + [P(B)]^4\, P(A) + \ldots \\
= \;& \frac{2}{36} + \frac{28}{36} \cdot \frac{2}{36} + \left(\frac{28}{36}\right)^2 \cdot \frac{2}{36} + \left(\frac{28}{36}\right)^3 \cdot \frac{2}{36} + \ldots
\end{aligned}
$$

Do you recognize this as an infinite geometric series?

From school, the formula for the sum of a geometric series is

$$\begin{aligned} a + ar + ar^2 + ar^3 + ... &= \frac{a}{1-r} \\ \text{In this exercise we have} \quad a &= \frac{2}{36} \quad \text{and} \\ r &= \frac{28}{36} \\ \text{So, back to the worked out series above} \quad &: \\ \frac{2}{36} + \frac{28}{36} \cdot \frac{2}{36} + \left(\frac{28}{36}\right)^2 \cdot \frac{2}{36} + \left(\frac{28}{36}\right)^3 \cdot \frac{2}{36} + .. &= \frac{\frac{2}{36}}{1 - \frac{28}{36}} \\ &= \frac{\frac{2}{36}}{\frac{8}{36}} \\ &= \frac{1}{4} \end{aligned}$$

(b) Now you can try Exercise $2.119\mathbf{b}$ yourself!

Define two sets; one to obtain a sum of 4 and the other *not to* obtain a sum of 4 or 7.

Your answer for this question should be $\frac{1}{3}$ (use the geometric series expansion again).

2.10 The law of total probability and Bayes' rule

Partitioning of the sample space S is defined before the two important theorems are given. Note how a *tree-diagram* is used to solve the problem and how it simplifies a complex question.

EXERCISE 2.135

I will show you to do this using a tree-diagram, but first show you the hard way!
Let $M =$ major airline, $R =$ private airline $C =$ commercial airline $B =$ travel for business.

a.

$$\begin{aligned} P(B) &= P(B \mid M) \cdot P(M) + P(B \mid R) \cdot P(R) + P(B \mid C) \cdot P(C) \\ &= 0.6\,(0.5) + 0.3\,(0.6) + 0.1\,(0.9) \\ &= 0.57 \end{aligned}$$

b.

$$\begin{aligned} P(B \cap R) &= P(B \mid R)P(R) \\ &= 0.3\,(0.6) \\ &= 0.18 \end{aligned}$$

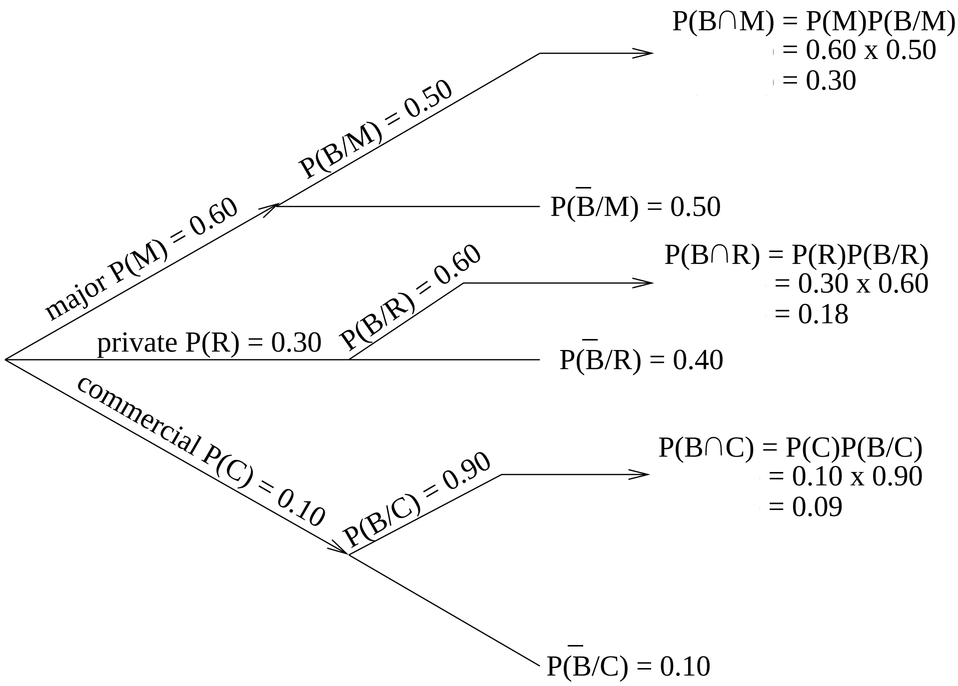
c.

$$P(R \mid B) = \frac{P(B \cap R)}{P(B)} = \frac{0.18}{0.57} = 0.3158$$

d.

$$P(B \mid C) = \frac{P(B \cap C)}{P(C)} = \frac{0.09}{0.10} = 0.90$$

Using a tree diagram:



2.11 Numerical events and random variables

You will understand as we proceed how important it is that you know what is meant by a *random variable*. If you read the definition given in this section and learn it by heart, you will still know nothing. Read the paragraph before the definition is given and make sure you understand what is said there. It is here that you have to take note of the notation for a variable and the values it takes on in the outcome of an experiment as given in examples 2.24 and 2.25.
Use capital letters to indicate the variable to be measured, e.g. Y (as in the explanation).

EXERCISE 2.135

In Exercise 2.112 we were told that each of the three radar sets have a probability of 0.02 to fail in detecting an aeroplane. We are interested in the probability that an aeroplane is detected and the variable of interest is the number of radar sets that detect the aeroplane flying over their area. Denote this variable by the capital letter Y. The possible values that Y can assume in these experiments are only 0 (none detected); 1 (one radar detects it); 2 (two radars detect it); 3 (all three radars detect the aeroplane).
In the proper notation we are going to look at $P(Y = y)$, where$Y$ can take on the values $0, 1, 2$ and 3. Furthermore, let F indicate failure to detect and D indicate detecting the aeroplane.

$$\begin{aligned}
\text{None detected} &: \quad Y = 0 \\
P(Y = 0) &= (0.02)^3 \\
\text{One detection} &: \quad Y = 1 \\
\text{Three possibilities:} & \quad FDF \quad DFF \quad FFD \\
P(Y = 1) &= 3 \cdot (0.02)^2 (0.98) \\
\text{Two detections} &: \quad Y = 2 \\
\text{Three possibilities:} & \quad DDF \quad DFD \quad FDD \\
P(Y = 2) &= 3 \cdot (0.02)(0.98)^2 \\
\text{Three detections} &: \quad Y = 3 \\
P(Y = 3) &= (0.98)^3
\end{aligned}$$

2.12 Random sampling

The topics discussed here will be explained in more detail at a later stage – most probably in other modules. As you continue with your studies, you will see how everything is interlinked – between same-level modules and also modules at other levels. Make sure that you do not allow "gaps" in your knowledge. You will pay dearly for it, either in your studies or in your job as a statistician.

The following concepts are only touched in 2.12 are:

- population
- sampling
  - random sampling
  - simple random sample
  - sampling with replacement
  - sampling without replacement
- design of an experiment

2.13 Summary

Knowing the contents of this chapter implies that you are able to apply

- the concept of a probabilistic model for calculating the probability of an event
- basic concepts in probability theory
- theorems in probability theory
- counting rules used in the sample-point method
- and conceptualize the operations of set algebra and the laws of probability
- the tools and calculating probabilities of numerical events
- probability as the foundation and a tool for statistical inference

---

Activity 2.1

Do you ever, at the end of a day, sit or lie down and think about the things that happened that day: the good stuff, the bad moments, the errors you made, etc.? We should all do this on a daily basis as it can actually improve the quality of our lives (note that I include myself). It is becoming more and more difficult to find time for reflection and in studies it is absolutely essential. Make an attempt to do the following:

1. When you have gone through chapter 2 on probability for the first time, put the textbook and guide away and write down what you can remember about all the stuff you went through. You may be surprised at how little you can actually remember, but that is no reason to lose heart. It only implies that you should go through it all again and then reflect once more. The less you can remember, the more you should be motivated to go back and work through the knowledge again. It is not necessary to write down formulas and definitions, but write in your own words what you can remember. This activity asks you to think "back" – reflect on your knowledge. This way you will know when you have reached the stage that you can feel secure enough to be examined on your knowledge!

2. As you go through the chapter for the second, or third time, do see if you can remember any of the examples discussed in this chapter. Do you realize how an example can "link" knowledge to its application? I am sure you have looked at a theorem and found it impossible to understand and/or remember, but once you have seen it applied in an example, the light comes through.
3. Take your assignment 01 and identify the questions on probability. Then make the link between them and the topics in chapter 2. You can do the same with questions on probability in the trial paper. Make notes on the principles in probability theory that are tested. At first–year level you may find it difficult to find the essential knowledge contained in a chapter or section, but practice will improve this ability.

CHAPTER 3

Discrete random variables and their probability distributions

What is a *random variable*? Does "variable" refer to something changeable, erratic, inconsistent, unpredictable, or...? Why do we add the description "random"? Are we referring to a hit and miss process, is it unplanned, a game of chance, accidental or unpredictable? Of course not! Remember that referring to *random* in a statistical experiment points to the outcome, which is not know ahead of time. Seeing that random variables form the centre of distribution theory you should maybe read paragraph 2.11 again.

In the modules $STA1501$ and $STA1502$ you heard about random variables and were told that they are classified as either *discrete* or *continuous*. Make sure that you know the significant differences between discrete and continuous random variables. In WM&S discrete random variables are discussed in chapter 3 and continuous random variables in chapter 4 to emphasize the difference in their characteristics and resulting analysis.

Recall that this guide alone is not sufficient – you have to study chapter 3 in WM&S in detail and note that assignment and examination questions can be based on any section in this chapter.

STUDY

WM&S: Chapter 3

3.1 Basic definition

Once you really understand why a variable is classified as *discrete*, give attention to the relevance of probability and numerical events as it relates to values of discrete random variables. Read how we get to the concept of a *probability distribution* and understand that we are at this stage referring to *discrete probability distributions* because we are considering variables which are classified as *discrete*.

3.2 The probability distribution for a discrete random variable

In this section the very basic notation, two definitions as well as the characteristics of a discrete probability distribution are given. Give special attention in these basic definitions and characteristics (fundamental for the analysis of random variables) to the difference between discrete and continuous random variables. You will discover as you continue that the probability distribution for a random variable is not an exact presentation, but a *theoretical model.*

EXERCISE 3.13

There is a $\frac{1}{3}$ chance a person has O^+ blood and $\frac{2}{3}$ they do not. Similarly, there is a $\frac{1}{15}$ chance a person has O^- blood and $\frac{14}{15}$ chance they do not. Assuming donors are randomly selected, if $X =$ number of O^+ blood donors and $Y =$ number of O^- blood donors, the probability distributions are

	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$
$p(x)$	$\left(\frac{2}{3}\right)^3 = \frac{8}{27}$	$3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) = \frac{12}{27}$	$3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2 = \frac{6}{27}$	$\left(\frac{1}{3}\right)^3 = \frac{1}{27}$
$p(y)$	$\left(\frac{14}{15}\right)^3 = \frac{2744}{3375}$	$3\left(\frac{14}{15}\right)^2\left(\frac{1}{15}\right) = \frac{588}{3375}$	$3\left(\frac{14}{15}\right)\left(\frac{1}{15}\right)^2 = \frac{42}{3375}$	$\left(\frac{1}{15}\right)^3 = \frac{1}{3375}$

Define $Z = X + Y$ as the number persons with type O blood. The probability that a donor will have type O blood is $\left(\frac{1}{3} + \frac{1}{15} = \frac{6}{15} = \frac{2}{5}\right)$.
The probability distribution of Z can be summarized as

	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$
$p(z)$	$\left(\frac{3}{5}\right)^3 = \frac{27}{125}$	$3\left(\frac{3}{5}\right)^2\left(\frac{2}{5}\right) = \frac{54}{125}$	$3\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)^2 = \frac{36}{125}$	$\left(\frac{2}{5}\right)^3 = \frac{8}{125}$

I want you to consider the characteristics of a discrete random variable and test if they apply in this example above. Do you know what I am referring to? Simply test if

1. all the probabilities lie from 0 to 1, i.e. $0 \leq p(x) \leq 1$ (also for the variables Y and Z separately) and
2. the sum of the probabilities total one, i.e. $\sum_x p(x) = 1$ (also for the variables Y and Z separately)

3.3 The expected value of a random variable or a function of a random variable

Definitions are given for the

1. *expected value* of the random variable Y, which is denoted by $E(Y)$
2. *variance* of the random variable Y, denoted by $V(Y)$, and from previous knowledge
3. *standard deviation* of the random variable Y calculated as the (positive) square root of the variance, thus $\sqrt{V(Y)}$

Five theorems and their proofs are given, namely for the

1. expected value of a real-valued function $g(Y)$ of the discrete random variable Y with probability function $p(y)$
2. expected value of a constant c (e.g. $c = 4$, or $c = 75$, or...)
3. expected value of a real-valued function $g(Y)$ of Y multiplied by some constant c, which then becomes $c \cdot g(Y)$
4. expected value of the sum of a number of (say k) real-valued functions, say $g_1(Y)$, $g_2(Y), g_3(Y), ...g_k(Y)$
5. variance expressed as an expected value

Note that if the probability function $p(y)$ is an accurate characterization of the population frequency distribution, then the expected value is equal to the population mean, i.e.

$$\mathbf{E(Y) = mean}.$$

EXERCISE 3.31

a. The mean of W will be larger than the mean of Y if $E(Y) > 0$. If $E(Y) < 0$, the mean of W will be smaller than $E(Y)$. If $E(Y) = 0$, the mean of W will be equal to $E(Y)$.
b. $E(W) = E(2Y) = 2E(Y) = 2E(X) = 2\mu$ (if we use the notation that $E(X) = \mu$).
c. The variance of W will be larger than σ^2, since the spread of values of W has increased.(If we use the notation that $V(Y) = \sigma^2$.)
d. $V(X) = E\left[X - E(X)^2\right] = E\left[(2Y - E(2Y)^2\right] = E\left[(2Y - 2E(Y))^2\right] = 4E\left[(Y - E(Y))^2\right] = 4\sigma^2$

EXERCISE 3.33

a. $E(aY + b) = E(aY) + E(b) = aE(Y) + b = a\mu + b$
b. $V(aY + b) = E\left[(aY + b - a\mu - b)^2\right] = E\left[(aY - a\mu)^2\right] = a^2E\left[(Y - \mu)^2\right] = a^2\sigma^2$

3.4 The binomial probability distribution

The binomial distribution, based on a binomial experiment, is usually the first discrete probability distribution students are introduced to. Make sure you understand and remember the given properties of a binomial experiment. You should have some knowledge of the binomial distribution even before you start, as the binomial distribution formed part of the study material in $STA1501$. There you learn that a binomial experiment is characterized by four properties:

- There are a fixed number, n, of identical trials.
- There are only two outcomes, referred to a *success* and a *failure.*
- The probability of success on a single trial is denoted as p and the probability of failure by $(1-p)$.
- The trials are independent.

WM&S adds one more property by characterizing the *random variable* of interest as *the number of successes observed in the* n *trials*.

If your random variable is such that it can only assume two values, namely 0 and 1, and the probability that the outcome is 1 is equal to p and the probability that the outcome is 0 is equal to $(1-p)$, we call it a *Bernoulli process.* Except for the fist property, this description fits into the list of properties given above describing the binomial random variable. R M Weiers, in his textbook Introduction to Business Statistics, states: "In the binomial distribution, the discrete random variable is the number of success e.g. that occur in n consecutive trials of the Bernoulli process." Understand that by allocating the value *one* to the *number of trials* $(n = 1)$, in a binomial distribution, you can call it a Bernoulli distribution.

Furthermore, a shorthand notation is convenient, namely to indicate the probability of a success *not taking place* as $q = (1-p)$. Please take note of the paragraph just below definition 3.6, stating that *success* not necessarily refers to something "good".

Study the steps in example 3.5, proving that the given data relate to a binomial experiment.
The words "if and only if" are used in definition 3.7. This is well-known mathematical notation. "If and only if" implies that theLHS (left hand side) of an equation is equivalent to the RHS (right hand side) *and vice versa*. "If and only if" means that the

LHS implies RHS and it is at the same time true that RHS implies LHS

We write $LHS \Longrightarrow RHS$ and at the same time $RHS \Longrightarrow LHS$.

The short notation is $LHS \Longleftrightarrow RHS$ and you read $\Longleftrightarrow$ as "if and only if".

Back to definition 3.7:

$p(y) = \binom{n}{y} p^y q^{n-y}$ note that $y = 1, 2, ..., n$

This is a probability expression because $0 \leq p \leq 1$.

Make sure you know the *binomial series expansion*

$$
\begin{aligned}
(p+q) &= \binom{n}{0}p^0q^n + \binom{n}{1}p^1q^{n-1} + \binom{n}{2}p^2q^{n-2} + ... + \binom{n}{n}p^nq^0 \\
&= \frac{n!}{0!\,(n-0)!}\cdot p^0q^n + \frac{n!}{1!\,(n-1)!}\cdot p^1q^{n-1} + \frac{n!}{2!\,(n-2)!}\cdot p^2q^{n-2} + ... + \frac{n!}{n!\,(n-0)!}\cdot p^nq^0
\end{aligned}
$$

Of course you can see that $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$. You should also know that $p^0 = q^0 = 1$

In Theorem 3.7 you are given the values of the two *parameters*, namely the mean and the variance, of this distribution:

$$
\begin{aligned}
\mu &= E(Y) = np \\
&\text{and} \\
\sigma^2 &= V(Y) = npq
\end{aligned}
$$

You should be able to prove these two expressions.

EXERCISE 3.39

Let Y indicate the number of components failing in less than 1000 hours. Then Y has a binomial probability distribution with $n = 4$ and $p = 0.2$.

a. $P(Y=2) = \binom{4}{2}(0.2)^2(0.8)^2 = \frac{4!}{2!2!}\cdot 0.04\cdot 0.64 = 0.1536$

b. The system will operate if $0, 1$ or 2 components fail in less than 1000 hours. So

$$
\begin{aligned}
P(\text{system operates}) &= \binom{4}{0}(0.2)^0(0.8)^4 + \binom{4}{1}(0.2)^1(0.8)^3 + \binom{4}{2}(0.2)^2(0.8)^2 \\
&= \frac{4!}{0!4!}\cdot 1\cdot 0.4096 + \frac{4!}{1!3!}\cdot 0.20\cdot 0.512 + \frac{4!}{2!2!}\cdot 0.04\cdot 0.64 \\
&= 0.4096 + 0.4096 + 0.1536 \\
&= 0.9728\backslash
\end{aligned}
$$

EXERCISE 3.45

a. This is a binomial probability distribution with $n = 3$ and $p = 0.8$

b. The alarm will function if $Y = 1,\ \ 2$ or 3.

$$\begin{aligned} P(Y &\geq 1) = 1 - P(Y=0) \\ &= 1 - \binom{3}{0}(0.2)^3 \\ &= 1 - 0.008 \\ &= 0.992 \end{aligned}$$

EXERCISE 3.55

$$\begin{aligned} E\{Y(Y-1)(Y-2)\} &= \sum_{y=0}^{n} \frac{y(y-1)(y-2)\,n!}{y!\,(n-y)!} p^y (1-p)^{n-y} \\ &= \sum_{y=3}^{n} \frac{n(n-1)(n-2)(n-3)!}{(y-3)!\,(n-3-(y-3))!} p^y (1-p)^{n-y} \\ &= n(n-1)(n-2)\,p^3 \sum_{z=3}^{n-3} p^z (1-p)^{n-3-z} \\ &= n(n-1)(n-2)\,p^3 \cdot 1 \\ &= n(n-1)(n-2)\,p^3 \\ E\{Y(Y-1)(Y-2)\} &= E\left\{Y^3 - 3Y^2 + 2Y\right\} \\ &= E\left(Y^3\right) - 3E\left(Y^2\right) + 2E(Y) \\ E\left(Y^3\right) - 3E\left(Y^2\right) + 2E(Y) &= n(n-1)(n-2)\,p^3 \end{aligned}$$

Therefore $\quad E\left(Y^3\right) = 3E\left(Y^2\right) - 2E(Y) + n(n-1)(n-2)\,p^3$

From theorem 3.7 we have $\quad E\left(Y^2\right) = n(n-1)\,p^2 + np$

and $\quad E(Y) = np$

So,

$$\begin{aligned} E\left(Y^3\right) &= 3\left[n(n-1)\,p^2 + np\right] - 2np + n(n-1)(n-2)\,p^3 \\ &= 3n(n-1)\,p^2 - n(n-1)(n-2)\,p^3 + np \end{aligned}$$

3.5 The geometric probability distribution

Reading through the pages on the geometric probability distribution it is clear that the random variable is also discrete and furthermore, it is "family" of the binomial distribution. A random variable having a geometric probability distribution is defined and the expected value and variance of such a variable is given in the form of a theorem.

EXERCISE 3.73

Place this as a question involving a geometric random variable.

Let Y indicate the number of accounts audited until the first with substantial errors is found.

a. $p = \frac{9}{10}$ or 0.9 and $q = 0.1$
If $p(y) = q^{y-1} \cdot p$ then

$$\begin{aligned} P(Y=3) &= (0.1)^2 \cdot (0.9) \\ &= 0.009 \end{aligned}$$

b.

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) \\ \sum_{y=3}^{\infty} q^{y-1} \cdot p &= 1 - \sum_{y=1}^{2} q^{y-1} \cdot p \\ &= 1 - \left[\left(q^0 \cdot p\right) + \left(q^1 \cdot 0.9\right)\right] \\ &= 1 - [(1 \cdot 0.9) + (0.1 \cdot 0.9)] \\ &= 1 - 0.99 \\ &= 0.01 \end{aligned}$$

3.6 The negative binomial probability distribution

We do not treat this as optional. Therefore Section 3.6 is included for examination purposes.

EXERCISE 3.93

Assuming independent testing and probability of a success (finding a non-defective engine) equal to 0.9 in any test, let Y denote the number of tests on which the third non-defective engine is tested. Assume that Y has a negative binomial distribution with $p = 0.9$. .

a. We are interested in $r=3$ and $y=5$

$$\begin{aligned}\binom{y-1}{r-1}p^r q^{y-r} &= \binom{5-1}{3-1}(0.9)^3(0.1)^{5-3}\\ &= \binom{4}{2}(0.9)^3(0.1)^2\\ &= 6\cdot(0.729)\cdot(0.01)\\ &= 0.04374\end{aligned}$$

b. "On or before the fifth trial" implies $P(Y\leq 5)$ which means $P(Y=5)+P(Y=4)+P(Y=3)$

$$\begin{aligned}\binom{y-1}{r-1}p^r q^{y-r} &= \binom{5-1}{3-1}(0.9)^3(0.1)^{5-3}+\binom{4-1}{3-1}(0.9)^3(0.1)^{4-3}+\binom{3-1}{3-1}(0.9)^3(0.1)^{3-3}\\ &= \binom{4}{2}(0.9)^3(0.1)^2+\binom{3}{2}(0.9)^3(0.1)^1+\binom{2}{2}(0.9)^3(0.1)^0\\ &= 6\cdot(0.729)\cdot(0.01)+3\cdot(0.729)\cdot(0.1)+1\cdot(0.729)\cdot(1)\\ &= 0.04374+0.2187+0.729\\ &= 0.99144\end{aligned}$$

3.7 The hypergeometric probability distribution

Make sure you understand examples 3.1 and 3.5 as they clearly give the characteristics of a hypergeometric probability distribution.

EXERCISE $\mathbf{3.105}$

a. The random variable Y follows a hypergeometric distribution. The probability of being chosen on a trial is dependent on the outcome of previous trials.

b.

$$P(Y \geq 2) = P(Y=2)+P(Y=3)$$

Use: $$\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

with $N = 8;\ n=3;\ r=5;\ y=2,3$

$$\begin{aligned}p(2)+p(3) &= \frac{\binom{5}{2}\binom{3}{1}}{\binom{8}{3}}+\frac{\binom{5}{3}\binom{3}{0}}{\binom{8}{3}}\\ &= \frac{\left(\frac{5!}{2!3!}\right)\left(\frac{3!}{1!2!}\right)}{\left(\frac{8!}{3!5!}\right)}+\frac{\left(\frac{5!}{3!2!}\right)(1)}{\left(\frac{8!}{3!5!}\right)}\\ &= 0.5357+0.1786\\ &= 0.7143\end{aligned}$$

c.

$$
\begin{aligned}
E(Y) &= \frac{3}{\frac{5}{8}} \\
&= 1.875 \\
Variance &= V(Y) = 5 \cdot \frac{5}{8} \cdot \frac{3}{8} \cdot \frac{5}{7} \\
&= 0.5022 \\
SD &= \sqrt{V(Y)} = 0.7087
\end{aligned}
$$

3.8 The Poisson probability distribution

This is a popular probability distribution as Y measures the number of rare events within some time frame (or space, or volume, etc.). The model used in a Poisson distribution is called a *Poisson process*, based on a theory assuming independence of occurrences in non-overlapping intervals. The parameter is λ, which is the *average value* (or mean) of the random variable Y. Of interest is that λ is actually the mean as well as the variance of the distribution. You can use Table 3 in Appendix III to find the values for different Poisson probabilities.

EXERCISE 3.139

Recall that the mean and variance of a Poisson variable is equal to the parameter λ.
With $\lambda = 2$ we have that

$$
\begin{aligned}
E(Y) &= 2 \\
\text{and because} \quad V(Y) &= E\left(Y^2\right) - [E(Y)] \\
\text{we can say that} \quad E\left(Y^2\right) &= V(Y) + [E(Y)]^2 \\
&= 2 + 2^2 \\
&= 6 \\
\text{So} \quad E(X) &= 50 - 2E(Y) - E\left(Y^2\right) \\
&= 50 - 2(2) - 6 \\
&= 40
\end{aligned}
$$

3.9 Moments and moment-generating functions

Definitions $3.12,\ 3.13$ and 3.14 form the basis of the knowledge in this section. This should be part of your "for-ever-knowledge" – make sure that you need not look it up every time you come across a question about moments or moment-generating functions! The two applications in theorem 3.12 are basic knowledge for any statistician, namely that the moments for the random variable can be determined using the moment-generating function and differentiation and/or infinite sum expansions. Make sure that you understand examples $3.23, 3.24$ and 3.25 as they represent typical assignment and examination questions.

If you have forgotten, below are three series expansions (in this section you will need the first one):

$$\begin{aligned} e^{tx} &&&= 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots\ldots.. \\ log\,(1+x) &= log_e\,(1+x) &&= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots\ldots\ldots. \\ log\,(1-x) &= log_e\,(1-x) &&= -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots\ldots\ldots. \end{aligned}$$

EXERCISE 3.145

Using the binomial theorem

$$\begin{aligned} m(t) &= E\left(e^{tY}\right) \\ &= \sum_{y=0}^{n} e^{ty} p^y q^{n-y} \\ &= \sum_{y=0}^{n} \left(pe^t\right)^y q^{n-y} \\ &= \left(pe^t + q\right)^n \end{aligned}$$

EXERCISE 3.155

a.

$$
\begin{aligned}
m(t) &= \frac{1}{6}e^{t} + \frac{2}{6}e^{2t} + \frac{3}{6}e^{3t} \\
\text{Differentiate} &: \\
m^{(1)}(t) &= \frac{1}{6}e^{t}\cdot(1) + \frac{2}{6}e^{2t}\cdot(2) + \frac{3}{6}e^{3t}\cdot(3) \\
&= \frac{1}{6}e^{t} + \frac{4}{6}e^{2t} + \frac{9}{6}e^{3t} \\
m^{(1)}(t) &\mid {}_{t=0} = \frac{1}{6}e^{0} + \frac{4}{6}e^{0} + \frac{9}{6}e^{0} \\
&= \frac{1+4+9}{6} \\
&= \frac{14}{6} \\
&= \frac{7}{3} \\
&= \frac{5}{9}
\end{aligned}
$$

b.

$$
\begin{aligned}
m^{(2)}(t) &= \frac{1}{6}e^{t}\cdot(1) + \frac{4}{6}e^{2t}\cdot(2) + \frac{9}{6}e^{3t}\cdot(3) \\
&= \frac{1}{6}e^{t} + \frac{8}{6}e^{2t} + \frac{27}{6}e^{3t} \\
m^{(2)}(t) &\mid {}_{t=0} = \frac{1}{6}e^{0} + \frac{8}{6}e^{0} + \frac{27}{6}e^{0} \\
&= 6 \\
\sigma^2 &= \mu_2' - \mu^2 \\
&= 6 - \left(\frac{7}{3}\right)^2 \\
&= \frac{5}{9}
\end{aligned}
$$

c. Since $m(t) = E(e^{tY}),\ Y$ can only take on values $1, 2$ and 3 with probabilities $\frac{1}{6},\ \frac{2}{6}$ and $\frac{3}{6}$.

3.10 Probability-generating functions

Any discrete random variable can assume a countable number of values, but these values need not be integers $(..-3,-2,-1,0,1,2,.....)$. A theory has been developed for those variables whose values are part of a counting process. Read the discussion under 3.10 in WM&S on these types of special discrete random variables and some practical examples.
There are two definitions for

(a) a probability–generating function $P(t)=E\left(t^Y\right)$ for the random variable Y

(b) the *k*th factorial moment $\mu_{[k]}=E\left(Y\left(Y-1\right)\left(Y-2\right)\left(Y-3\right).....\left(Y-k+1\right)\right)$ for the random variable $Y.$

The probability–generating function was used in examples 3.26 and 3.27 to determine the mean of a discrete geometric random variable.

EXERCISE 3.165

For the Poisson:

$$\begin{aligned}P(t) &= E(t^Y)\\ &= \sum_{y=0}^{\infty}\frac{\lambda^y e^{-\lambda}t^y}{y!}\\ &= \frac{e^{-\lambda}}{e^{-\lambda t}}\sum_{y=0}^{\infty}\frac{(t\lambda)^y\, e^{-\lambda t}}{y!}\\ &= e^{\lambda(t-1)}\end{aligned}$$

This function $P(t)$ now has to be differentiated with respect to t to find $E(Y),$ and the second derivative will give us the value of $E(Y\left(Y-1\right))$.With these two answers we can calculate the variance. I hope you realized the moment when you started with this exercise that the answers for the expected value and the variance must be λ (in both cases)!

$$\begin{aligned}E(Y) &= \left.\frac{d}{dt}P(t)\right|_{t=1}\\ &= \left.\frac{d}{dt}e^{\lambda(t-1)}\right|_{t=1}\\ &= \left.\lambda e^{\lambda(t-1)}\right|_{t=1}\\ &= \lambda\end{aligned}$$

$$\begin{aligned}
E(Y(Y-1)) &= \left.\frac{d^2}{dt^2}P(t)\right|_{t=1} \\
&\left.\frac{d^2}{dt^2}e^{\lambda(t-1)}\right|_{t=1} \\
&= \left.\lambda^2 e^{\lambda(t-1)}\right|_{t=1} \\
E(Y^2-Y) &= \lambda^2 \\
E\left(Y^2\right)-E(Y) &= \lambda^2 \\
\text{or}\quad E\left(Y^2\right) &= \lambda^2+E(Y) \\
V(Y) &= E\left(Y^2\right)-[E(Y)]^2 \\
&= \left[\lambda^2+E(Y)\right]-[E(Y)]^2 \\
&= \left[\lambda^2+\lambda\right]-[\lambda]^2 \\
&= \lambda
\end{aligned}$$

3.11 Tchebysheff's theorem

This is a very useful theorem for many practical situations, where the distribution of the random variable is often not known. No proof is required, but study the example.

EXERCISE 3.167

(a) The value 6 lies $\frac{11-6}{3} = \frac{5}{3}$ standard deviations below the mean. Similarly, the value 16 lies $\frac{16-11}{3} = \frac{5}{3}$ standard deviation above the mean. By Tchebysheff's theorem, at least $1-\frac{1}{\left(\frac{5}{3}\right)^2} = 64\%$ of the distribution lies in the interval 6 to 16.

(b) By Tchebysheff's theorem, $0.09 = \frac{1}{k^2}$ so $k = \frac{10}{3} = 3.333$. Given that $\sigma^2 = 9$ we have that $\sigma = 3$.

$$\begin{aligned}
k\sigma &= \left(\frac{10}{3}\right)\cdot 3 \\
&= 10 \\
\text{which implies that } C &= 10
\end{aligned}$$

3.12 Summary

Read through the given summary and make sure that a similar summary exists in your mind.

Study objectives of this chapter

After completion of this chapter you should be able to

- use the correct notation for a discrete random variable and its relevant probability distribution
- evaluate the properties of a discrete probability distribution for given scenarios
- define and calculate expected values and variances for random variables and functions of random variables
- prove the expectation theorems
- differentiate between the probability functions for the binomial, geometric, negative binomial, hypergeometric and Poisson random variables
- calculate the mean and variance for the different discrete random variables using their respective discrete probability destitutions
- define and calculate moments about the origin as well as about the mean
- define and calculate the moment-generating function for a discrete random variable
- prove with applications that any of the moments for a variable can be found using differentiation of the moment-generating function and a series expansion
- approximate certain probabilities for *any* probability distribution when only the mean and variance are known by using Tchebysheff's theorem.

Activity 3.1

This activity is once again for your own benefit.

1. Make a summary of the means and variances for the five discrete random variables discussed in this chapter.
2. See if you can determine the moment generating functions for the binomial, geometric, negative binomial and Poisson distributed random variable and then also tabulate them.
3. Compare your summaries for 1 and 2 with Table 3.4 given under **Summary** in 3.12. There is also a summary of Discrete distributions on the last page of WM&S with headings for the specific distribution, its mean, variance and moment generating function. Do not simply copy those summaries as you cannot benefit from doing that! I also recommend that you summarize the properties of the random variables for the different distributions discussed in this section. I placed my summary below so you can compare it with yours.

DISCRETE PROBABILITY DISTRIBUTIONS

Distribution	Characteristics	Probability function
Binomial Y˜bin(n, p)	Y: number of successes n independent & identical trials p : probability of success 2 outcomes: success or failure p remains same Parameters: n and p	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$ $y = 0; 1; ...; n$ and $0 \leq p \leq 1$ finite number of trials
Geometric Y˜geom(p)	Y: number trials up to 1st success Trials: independent & identical p : probability of success 2 outcomes: success or failure p remains same Parameter: p	$p(y) = q^{y-1}p$ $y = 1; 2; 3; ..$ infinite number of trials
Negative binomial Y˜neg bin(r, p)	Y: total number of trials up to rth success Trials: independent & identical p : probability of success 2 outcomes: success or failure Parameters: r and p	$p(y) = \binom{y-1}{r-1} p^r q^{y-r}$ $y = r, , r+1, ...$ infinite number of trials
Hyper-geometric Y˜hyp(r, n, N)	Y: number of elements in sample having characteristic A Trials: dependent (no replacement) 2 characteristics A and B in pop. r : number of A's in population Sample size n large relative to N Parameters: r, n and N	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$ $y = 0, 1, ..$ infinite number of trials
Poisson Y˜Poison(λ)	Y: number of rare events Events: independent Binomial approximation used when $n >>; p <<; \lambda = np < 7$ Parameter: λ	$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$ $y = 0; 1; .,$ $\lambda > 0$ infinite number of trials

CHAPTER 4

Continuous random variables and their probability distributions

STUDY	
WM&S:	**Chapter** 4.

4.1 Introduction

This section starts by highlighting the difference between a discrete and a continuous random variable. If it is *not* possible to identify all the values that a random variable can assume, it cannot be discrete because we *defined* a discrete random variable Y as assuming only a finite number of distinct values. A continuous random variable is defined and the importance is stressed of classification into either discrete or continuous. You know about probability tools such as probability trees, as well as the other definitions of probability used in calculations when the random variable is discrete. What do we have available in terms of probability if the random variable is continuous? What will the differences be in the probability distribution for a discrete and a continuous random variable?

You should know the answers to these questions once we have gone through Chapter 4 in WM&S.

4.2 The probability distribution for a continuous random variable

Definition 4.1 defines a new concept: the *distribution function* of a random variable. Note that the definition does not refer to discrete or continuous – it is applicable to both discrete and continuous random variables. A capital letter F is used to indicate the distribution function (compare this to the lower case f, used in Chapter 3 to denote a probability density function) and $F(y)$ is a *cumulative* function because it is defined as the cumulative probability $P(Y \leq y)$ implying that *any value* from $-\infty$ to y that the variable can assume, is considered.

Read the paragraph under Definition 4.1 carefully, where WM&S states that the distribution function associated with a random variable indicates whether the variable is discrete or continuous. Why did

WM&S not introduce the concept in chapter 3 when the topic of that chapter was *discrete* random variables? I find this very interesting and can see that only introducing you to distribution functions at this stage gave them the opportunity to highlight the differences between the distribution function of a discrete random variable and that of a continuous random variable. The first part of 4.2 therefore introduces you to the distribution function for the *discrete binomial* random variable and only *after that* explanation, you are introduced to the distribution function for a *continuous random variable*.

Theorem 4.1, giving the properties of a distribution function, also applies to both discrete and continuous random variables and you learn that a distribution function is always an increasing function lying between 0 and 1 – regardless of the nature of the random variable it describes.

The next new concept is contained in Definition 4.3: the *probability density function,* indicated by the lower case letter f. The density function describes a continuous random variable in the same way as the probability (or frequency) distribution describes the discrete random variable. *A discrete random variable cannot have a probability density function* – the description only applies to a continuous random variable.

You will now come to realize why you need knowledge of calculus techniques in statistics, because the probability density function can only be determined from the distribution function of a continuous random variable if you *differentiate* the density function. Furthermore, because differentiation and integration are inverse processes, you can integrate the probability density function to get the distribution function.

We say that the derivative of $F(y)$ gives $f(y)$:

$$f(y) = \frac{dF(y)}{dy} = F\,'(y)$$

Alternatively, you integrate $f(y)$ to get $F(y)$:

$$F(y) = \int_{-\infty}^{y} f(t)dt$$

Everything discussed in this chapter is extremely important, so make sure that you understand every part before continuing. There is no shame in repetition of sections – that is how we learn. Study examples 4.2 and 4.3 in detail and take note of their relevance to the properties of a density function (short for probability density function).

In previous modules you learnt about *quantiles*, a concept indicating separation of data into equal-sized groups (examples of quantiles are percentiles, deciles and quartiles). Quantiles are defined in $4.4.$ and the letter p used in this definition gives detail about a *percentile* (ϕ_p can indicate the $100p-$th percentile of Y) and you should remember that a percentile divides the data into 100 equal parts. Recalling that the maximum value of the distribution function is $1,$ you substitute the value 0.5 into the expression for the distribution function to find the median value (which divides the data into two equal parts).

The magic is that the *density function* of a continuous random variable allows us to determine the probability that the variable assumes values within a specified interval. This important method is given in Theorem 4.3.

Note

If the *random variable is continuous*, the difference between the following four expressions

$P(a < Y < b)$, $P(a \leq Y < b)$, $P(a < Y \leq b)$ and $P(a \leq Y \leq b)$

is of no importance.

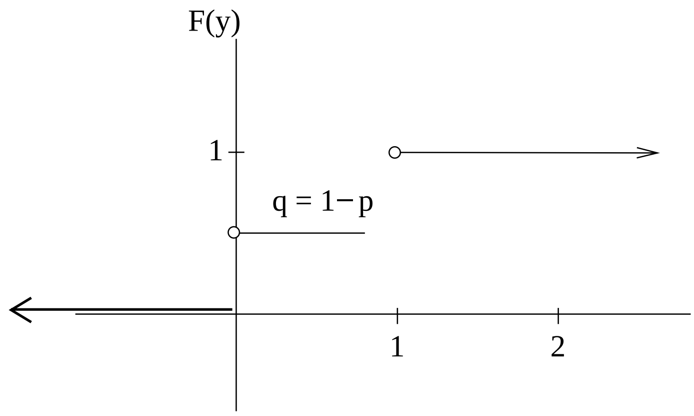
And if the variable is *discrete*? Then it is vital that you distinguish between the four probabilities as they all result in different values of the variable, as can be illustrated with the following example:

$P(2 < Y < 5) = \{3, 4\}$
$P(2 \leq Y < 5) = \{2, 3, 4\}$
$P(2 < Y \leq 5) = \{3, 4, 5\}$
$P(2 \leq Y \leq 5) = \{2, 3, 4, 5\}$

Make sure you understand all the examples in this section and do as many of the exercises as possible.

EXERCISE 4.3

(a) The probability that the variable assumes the value 1 is equal to p and that it takes on the value 0 is equal to $(1-p)$. In a sketch it looks like this:



(b) Test this against the three properties of a distribution function:

$$\begin{aligned} F(-\infty) &= 0 \\ F(\infty) &= 1 \\ F(y_1) &\leq F(y_2) \text{ for any } y_1 \leq y_2 \end{aligned}$$

It is indeed a distribution function.

EXERCISE 4.9

I want you to do this exercise and compare your answer to that given at the back of this textbook (in my textbook the solution is on p 881).

EXERCISE 4.11 (Only **a** to **d**)

Why do I know that this density function represents a continuous random variable? Discrete random variables do not have density functions! Do you remember?

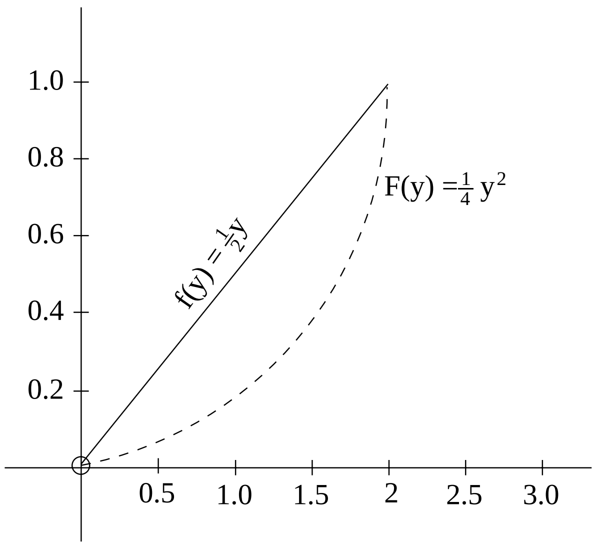
(a) In theorem 4.2 the properties of a density function are given, and to find this constant c, use property 2, namely that $\int_{-\infty}^{\infty} f(y)dy$ must be equal to one. In the calculation below the limits of the integral are changed to 0 to 2 instead of the $-\infty$ to ∞ given in the property. The reason for this is that this given function only exists from 0 to 2.

$$\begin{aligned}
\text{Use } \int_0^2 f(y)dy &= 1 \\
\int_0^2 cydy &= c\int_0^2 ydy \\
c\cdot\frac{1}{2}\int_0^2 (2y)\,dy &= \frac{c}{2}\left[y^2\right]_0^2 \\
&= \frac{c}{2}[4-0] \\
&= 2c \\
\text{Therefore } 2c &= 1 \\
c &= \frac{1}{2}
\end{aligned}$$

(b) Per definition $F(y) = \int_{-\infty}^{y} f(t)dt$. Use this to find the distribution function.

$$\begin{aligned}
F(y) &= \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{y} ctdt \\
&= \int_{0}^{y} \frac{1}{2}tdt = \frac{y}{2}\int_{0}^{y} tdt \\
&= \frac{1}{2}\cdot\frac{1}{2}\int_{0}^{y} 2tdt \\
&= \frac{1}{4}\left[t^2\right]_0^y \\
&= \frac{1}{4}\left[y^2 - 0\right] \\
&= \frac{1}{4}y^2 \text{ for } 0 \le y \le 2
\end{aligned}$$

(c)



(d) Use the distribution function to do this calculation, namely

$$\begin{aligned}
F(y) &= \frac{1}{4}y^2 \\
P(1 \le Y \le 2) &= F(2) - F(1) \\
&= \frac{1}{4}(2)^2 - \frac{1}{4}(1)^2 \\
&= 1 - \frac{1}{4} \\
&= \frac{3}{4} \quad \text{or} \quad 0.75
\end{aligned}$$

4.3 Expected values for continuous random variables

Think back about the knowledge in the previous chapter. Do you remember the concept *expected value* as defined for a discrete random variable in definition 3.4 ?

We now get to a continuation of the parallel concepts for a *continuous random variable.* The actual calculation of the expected value is *quite different* for the two types of variables as you can see in the table below.

Expected value of ...	A **discrete** random variable Y with frequency function $p(y)$	A **continuous** random variable Y with probability density function $f(y)$
The random variable itself:	$E(Y) \quad = \sum_{\text{all } y} yp(y)$	$E(Y) \quad = \int_{-\infty}^{\infty} yf(y)dy$
A function of the variable:	$E[g(Y)] = \sum_{\text{all } y} g(y)p(y)$	$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$
A constant c:	$E(c) \quad = c$	$E(c) \quad = c$
c multiplied by a function	$E(cg(Y) = cE(g(Y)$	$E(cg(Y) = cE(g(Y)$
The sum of more functions of the variable:	$E[g_1(Y) + g_2(Y) + ... + g_k(Y)]$ $= E[g_1(Y)] + E[g_2(Y)] + ...$ $..... + E[g_k(Y)]$	$E[g_1(Y) + g_2(Y) + ... + g_k(Y)]$ $= E[g_1(Y)] + E[g_2(Y)] + ...$ $..... + E[g_k(Y)]$

Just a reminder to be careful:

$E(Y^2) \neq V(Y)$ because
$V(Y) = E(Y^2) - [E(Y)]^2$

EXERCISE 4.25

Here is a small test for you: Did you forget to find the density function before you could calculate the expected value? Do I know it because the question stated that it was the distribution function? Yes, but that was only a confirmation; because a capital letter F is used in $F(Y)$, you should know that it is not the density function $f(Y)$, which is denoted by a small f.

This implies that you first have to determine the density function for this random variable, which is done through differentiation of $F(y)$, i.e.

$$f(y) = \tfrac{d}{dy}F(y)$$

For $y \leq 0:$ $\quad f(y) = \frac{d}{dy}0 = 0$

For $0 < y < 2:$ $\quad f(y) = \frac{d}{dy}\frac{y}{8} = \frac{1}{8}$

For $2 \leq y < 4:$ $\quad f(y) = \frac{d}{dy}\frac{y^2}{16} = \frac{y}{8}$

For $y \geq 4:$ $\quad f(y) = \frac{d}{dy}1 = 0$

The density function is therefore $f(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{8} & 0 < y < 2 \\ \frac{y}{8} & 2 \leq y < 4 \\ 0 & y \geq 4 \end{cases}$

This was testing your knowledge: Did you remember to find the *density function* before you could calculate the *expected value*? Not only did the question state that the given function was the distribution function, but it was also indicated using a small f, which corresponds to the notation for the density function $f(Y)$!

That is why the answer started with differentiation of the given distribution function $F(y)$.

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} yf(y)dy \\
&= \int_{0}^{2} y \cdot \frac{1}{8} dy + \int_{2}^{4} y \cdot \frac{y}{8} dy \\
&= \frac{1}{8}\int_{0}^{2} y \cdot dy + \frac{1}{8}\int_{2}^{4} y^2 dy \\
&= \frac{1}{8}\left[\frac{y^2}{2}\right]_0^2 + \frac{1}{8}\left[\frac{y^3}{3}\right]_2^4 \\
&= \frac{1}{8}\left[\frac{4}{2} - 0\right]_0^2 + \frac{1}{8}\left[\frac{64}{3} - \frac{8}{3}\right]_2^4 \\
&= \frac{1}{4} + \frac{56}{24} \\
&= \frac{6}{24} + \frac{56}{24} \\
&= \frac{62}{24} = \frac{31}{12}
\end{aligned}$$

$$\begin{aligned}
V(Y) &= E(Y^2) - [E(Y)]^2 \\
&= \frac{47}{6} - \left(\frac{31}{12}\right)^2 \\
&= 1.1597 \\
&\approxeq 1.16
\end{aligned}$$

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y)dy \\
&= \int_0^2 y^2 \cdot \frac{1}{8}dy + \int_2^4 y^2 \cdot \frac{y}{8}dy \\
&= \frac{1}{8}\int_0^2 y^2 \cdot dy + \frac{1}{8}\int_2^4 y^3 dy \\
&= \frac{1}{8}\left[\frac{y^3}{3}\right]_0^2 + \frac{1}{8}\left[\frac{y^4}{4}\right]_2^4 \\
&= \frac{1}{8}\left[\frac{8}{3} - 0\right]_0^2 + \frac{1}{8}\left[\frac{256}{4} - \frac{16}{4}\right]_2^4 \\
&= \frac{1}{3} + \frac{15}{2} \\
&= \frac{47}{6}
\end{aligned}
$$

4.4 The uniform probability distribution

Different distributions describing a continuous random variable are discussed in this and the next sections – all in terms of their *density functions (*and not frequency functions), because these random variables are all continuous. The form of a density function is determined by certain values, called the *parameters.* Turn it around and you can say that the character of a distribution is described by the value of its parameters.

I want to take you back to the module $STA1501$, where you were introduced to the word *parameter*. In the first chapter of the prescribed book by Keller you were informed that “a descriptive measure of a population is a parameter” and the corresponding “measure of a sample is called a statistic”. Do you recall this? I hope you made this link because you have to teach your brain to “link” knowledge. What you learnt earlier about a parameter is true and will never change, but now you link the concept *parameter* to a “formula”, called the *density function* which describes a specified distribution. In modules about inference you will learn that the statistician needs the values of population parameters and because it is hardly ever possible to find the population parameters, he/she has to apply sampling procedures and use the values calculated for the resulting *sample statistics* as *estimates* of the true *population parameters*. Notice that, according to the definition of a parameter, the corresponding sample values cannot be called parameters, so they are called *statistics* and these statistics are extremely important as they are used to approximate the population parameters and thus the distribution of the population.

Recalling known facts in a mathematical science is like magic! You should constantly recall previous knowledge and in doing that you strengthen the new information. It is your task to train your brain in this process of embedding new information in between existing knowledge. Once you have started doing this, you will realize that statistics is a very enjoyable subject because you are continuously revising facts and need just a little revision before an examination – *how nice*!

You have to bring the concepts *probability* and *area* together in your mind. Link the fact that a probability can never be larger than *one* to the second property of a density function in Theorem 4.2, namely $\int_{-\infty}^{\infty} f(y)dy = 1$. In this expression the $f(y)$, denotes the density function and is the equation of a graph (call it a curve). For a continuous random variable we deduce that the area under the density function curve represents the probabilities for the variable. The total area under the density curve is therefore equal to *one*. The uniform distribution is very important and yet very easy to use in explanations. Go to the density function given in Figure 4.9, describing a uniform continuous random variable defined over the interval θ_1 to θ_2. This density function is a constant value (because θ_1 to θ_2 are constants) and the graph is not only an ordinary straight line – it is a horizontal line. So, imagine a rectangle formed with base of length $\theta_2 - \theta_1$ and height some constant value on the $f(y)$-axis. What will this axis value be? The area of a rectangle is length times breadth and it has to be equal to *one* – by definition. So, if the distance from $\theta_2 - \theta_1 = 10$ (as in the figure), we know that the value on the $f(y)$-axis has to be $\frac{1}{10}$ (because we have to get a total area of *one* and $\frac{1}{10} \cdot 10 = 1$).

You have to be able to calculate the mean and the variance for each of the distributions discussed in this chapter. See Theorem 4.6.

EXERCISE 4.41

To find the variance you first have to determine the expected value of the variable, $E(Y)$ and also $E\left(Y^2\right)$.

Start with $E(Y)$, using theorem 4:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} yf(y)dy \\ &= \frac{\theta_2 + \theta_1}{2} \end{aligned}$$

$$\text{Now:} \qquad E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)dy$$

$$= \int_{\theta_1}^{\theta_2} y^2 \left(\frac{1}{\theta_2 - \theta_1}\right) dy$$

$$= \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} y^2 dy$$

$$= \frac{1}{\theta_2 - \theta_1} \left[\frac{y^3}{3}\right]_{\theta_1}^{\theta_2}$$

$$= \frac{\theta_2^3 - \theta_1^3}{3(\theta_2 - \theta_1)}$$

$$= \frac{(\theta_2 - \theta_1)\left(\theta_2^2 + \theta_2\theta_1 + \theta_1^2\right)}{3(\theta_2 - \theta_1)}$$

$$= \frac{\left(\theta_1^2 + \theta_1\theta_2 + \theta_2^2\right)}{3}$$

$$\text{So} \qquad V(Y) = E(Y^2) - [E(Y)]^2$$

$$= \frac{\left(\theta_1^2 + \theta_1\theta_2 + \theta_2^2\right)}{3} - \left(\frac{\theta_2 + \theta_1}{2}\right)^2$$

$$= \frac{(\theta_2 - \theta_1)^2}{12}$$

4.5 The normal probability distribution

The density function of the normal distribution function given in Definition 4.8 is rather complex and not easy to remember, and yet it is the continuous probability distribution most widely used! In $STA1501$ and $STA1502$ the properties of the normal distribution, as well as the density function, were given to you, but no further reference was made to the density expression. The two parameters μ and σ are also not new to you. Look carefully at the formula for the normal density function and notice that the only real “unknowns” are μ and σ because all the other symbols are numbers: The e represents the universal constant $e = 2.71828$ and π is also a constant with $\pi = 3.14158...$

Earlier the main interest was in *applications* of the normal distribution and its properties. You could calculate normal probabilities using the normal table and had to know the following:

- The normal distribution is *bell-shaped.*
- The graph is *symmetric* with the mean at the centre.
- The mean μ is a location parameter. In statistical software, do not change σ, but change μ and notice a horizontal movement of the curve.

- The standard deviation σ indicates the spread around the mean. In statistical software, do not change μ, but change σ and notice how the shape of the "bell" changes.
- The distances around the mean can easily be measured in standard deviations.
- All normal distributions can be *standardized* to have a mean 0 and standard deviation 1.
- The *area* under the bell-shaped curve represents *probability.*
- The area from 0 to a value z is the same as the area from 0 to $-z$.
- 0.5 (or 50%) of the area under the curve lies to the right of the mean and the other 0.5 (or 50%) of the area lies to the left of the mean.
- $P(Z > z) = P(Z < -z)$.
- $P(0 < Z < z) =.P(-z < Z < 0)$.

We now move to a more theoretical interest in the normal distribution, its properties, the two parameters and its density function. However, it can still be necessary for you to calculate normal probabilities in certain questions. The manner in which the areas for the normal table in WM&S are listed is different from that given to you in $STA1501$ and $STA150M$. Make sure you can use the table in this prescribed book as this is the one you will receive in the examination.

EXERCISE 4.71

Let $Y =$ the measured resistance of a randomly selected wire.

(a)

$$\begin{aligned} P(0.12 \le Y \le 0.14) &= P\left(\frac{0.12-0.13}{0.005} \le Y \le \frac{0.14-0.13}{0.005}\right) \\ &= P(-2 \le Y \le 2) \\ &= 1 - 2P(Y \ge 2) \\ &= 1 - 2 \cdot 0.0228 \\ &= 0.9544 \end{aligned}$$

(b) Let X be the number of wires that do not meet specifications. Then X is a binomial random variable with $n = 4$ and $p = 0.09954$.

$$\begin{aligned} P(X = 4) &= \binom{4}{4}(0.09954)^4(1-0.09954)^0 \\ &= (0.09954)^4 \\ &= 0.8297 \end{aligned}$$

EXERCISE 4.75

Let Y = volume filled, so that Y is a normal random variable with mean μ and $\sigma = 0.3$ (ounces – the English measuring system).
They require that we find $P(Y > 8) = 0.01$.
For the standard normal distribution $P(Y > z_0) = 0.01$ corresponds to $z_0 = 2.33$.
It must therefore hold that

$$\begin{aligned} 2.33 &= \frac{8-\mu}{0.3} \\ 8-\mu &= 2.33 \cdot 0.3 \\ \mu &= 8 - 0.699 \\ &= 7.301 \end{aligned}$$

4.6 The gamma probability distribution

Populations of continuous random variables with the following specific characteristics:

- non-negativity (only values from 0 to ∞) which implies skewness to the right
- probabilities concentrated near zero
- a gradual drop in probabilities the further you move from the origin

are adequately described by the gamma density function as defined in definition 4.9.

The gamma (Γ) distribution has two parameters:
α, which determines the *shape* of the density function and
β, which is referred to as the *scale parameter*.

Notice how the values of the two parameters α and β result in *special cases* of the gamma distribution:

1. Let $\alpha = 1$ in the gamma density function and the result is the density function of an *exponential distribution* with $\beta > 0$ as only parameter. Study definition 4.11 as well as the expected value and variance of the exponential distribution given in Theorem 4.10.

$$\begin{aligned} f(y) &= \int_0^\infty \frac{1}{\beta^1 \cdot \Gamma(1)} y^{1-1} e^{-\frac{y}{\beta}} dy \\ &= \int_0^\infty \frac{1}{\beta} e^{-\frac{y}{\beta}} dy \end{aligned}$$

The important characteristic of the exponential distribution is that its mean and variance are both equal to the parameter β.

2. Let $\alpha = \frac{\nu}{2}$ and $\beta = 2$ to find the very useful chi-square $\left(\chi^2\right)$ distribution with ν degrees of freedom. You used the χ^2 distribution and its table in your first-level modules. Read in WM&S how the more accessible chi-square values can be used to find values for the gamma distribution.

EXERCISE 4.81

These are very important deductions that you have to remember as you use them, for example, in the gamma density function.
Here we have to prove the important and often used result that
$\Gamma(1) = 1$ and $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

(a)

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty y^{\alpha-1} e^{-y} dy \\
\Gamma(1) &= \int_0^\infty y^{1-1} e^{-y} dy \\
&= \int_0^\infty 1 \cdot e^{-y} dy \\
&= -e^{-y}\big]_0^\infty \\
&= \left[-e^{-\infty} - \left(-e^{-0}\right) \right] \\
&= [0 - (-1)] \\
&= 1
\end{aligned}$$

(b) You have to use integration by parts to find this integral and it is a rather complex, but interesting, proof.

Consider $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ and use integration by parts:

$\int_0^\infty u dv = uv - \int_0^\infty v du$ and it will imply that you differentiate the $y^{\alpha-1}$ and integrate the e^{-y}

then $u = y^{\alpha-1}$ then $\frac{du}{dy} = (\alpha - 1) y^{(\alpha-1)-1}$ or $du = (\alpha - 1) y^{(\alpha-1)-1} dy$

and if $dv = e^{-y}$ then $v = \int_0^\infty e^{-y} dy = \left[-e^{-y} \mid_0^\infty\right]$

$$\begin{aligned}
\text{So} \qquad (\alpha) &= \int_0^\infty y^{\alpha-1}e^{-y}dy \\
&= \left[-y^{\alpha-1}e^{-y}\,|_0^\infty\right] - \int_0^\infty (\alpha-1)\,y^{(\alpha-1)-1}\left[-e^{-y}\right]dy.
\end{aligned}$$

Now.....$y^{\alpha-1}e^{-y} \Longrightarrow 0$ if $y \Longrightarrow \infty$ (because e^ygrows far faster than any power of y).

$$\begin{aligned}
\text{We get} \quad \Gamma(\alpha) &= 0-(\alpha-1)\int_0^\infty y^{(\alpha-1)-1}[-e^{-y}]dy \\
&= 0+(\alpha-1)\int_0^\infty y^{(\alpha-1)-1}e^{-y}dy \\
&= (\alpha-1)\int_0^\infty y^{(\alpha-1)-1}e^{-y}dy \\
&= (\alpha-1)\int_0^\infty y^{(\alpha-1)-1}e^{-y}dy \\
&= (\alpha-1)\,\Gamma(\alpha-1)
\end{aligned}$$

EXERCISE 4.89

(a) Do you remember that, for the *exponential distribution*, the parameter β represents the mean? This fact will be used in this question.

$$\begin{aligned}
P(Y>200) &= \int_{200}^\infty \frac{1}{100}e^{-\frac{y}{100}}dy \\
&= [-e^{-\frac{y}{100}}]_{200}^\infty \\
&= [-0--e^{-\frac{200}{100}}] \\
&= e^{-2} = 0.13533..
\end{aligned}$$

(b) We need to find the 99th percentile of the distribution of Y:

$$\begin{aligned}
P\left(Y > \phi_{.99}\right) &= \int_{\phi_{.99}}^{\infty} \frac{1}{100} e^{-\frac{y}{100}} dy \\
&= e^{\frac{\phi_{.99}}{100}} \\
\text{This implies that } e^{\frac{\phi_{.99}}{100}} &= 0.01 \\
\phi_{.99} &= -100\ln(.01) \\
&= 460.52
\end{aligned}$$

EXERCISE 4.109

Given is a gamma distribution with $\alpha = 3$ and $\beta = 2$ and also a loss function $L = 30Y + 2Y^2$.

$$\begin{aligned}
E(L) &= E\left(30Y + 2Y^2\right) \\
&= 30E\left(Y\right) + 2E(Y^2) \\
\text{Now use } V(Y) &= E(Y^2) - [E(Y)]^2 . \\
\text{For the gamma} \quad \alpha\beta^2 &= E(Y^2) - [\alpha\beta)]^2 \\
E(Y^2) &= (3)(2)^2 + [(3)(2))]^2 = 48 \\
E(L) &= 30(6) + 2\left(12 + 6^2\right) \\
&= 276 \\
V(L) &= E(L^2) - [E(L)]^2 \\
&= E(900Y^2 + 129Y^3 + 4Y^4) - (276)^2
\end{aligned}$$

$$\begin{aligned}
E(Y^3) &= \int_0^{\infty} \frac{y^3 \cdot y^2}{2^3 \cdot \Gamma(3)} e^{-\frac{y}{20}} dy \\
&= \int_0^{\infty} \frac{y^5}{16} e^{-\frac{y}{20}} dy = 480 \\
E(Y^4) &= \int_0^{\infty} \frac{y^4 \cdot y^2}{2^3 \cdot \Gamma(3)} e^{-\frac{y}{20}} dy \\
&= \int_0^{\infty} \frac{y^6}{16} e^{-\frac{y}{20}} dy = 5760 \\
V(L) &= E(900Y^2 + 129Y^3 + 4Y^4) - (276)^2 \\
&= 900(48) + 129(480) + 4(5760) - 276^2 \\
&= 47\,664
\end{aligned}$$

4.7 The beta probability distribution

When going through chapter 4 and all the continuous distributions, have you noticed how the domain of the random variable differs in the different densities?

Distribution			**Domain**
Uniform			$\theta_1 \leq y \leq \theta_2$
Normal			$-\infty \leq y \leq \infty$
	$\nearrow$	Exponential	
Gamma	$\longrightarrow$	Gamma	$0 \leq y \leq \infty$
	$\searrow$	Chi-square	
Beta			$0 \leq y \leq 1$

You will not be tested on the relation between the incomplete beta function and the binomial probability function. (If you need beta probabilities and quantiles, remember that statistical software is the easiest option.)
The expected value of the beta random variable is given in Theorem 4.11, and in Example 4.11 the beta function is directly integrated as α and β are both integers.

EXERCISE $\mathbf{4.130}$

Try to derive the variance $V(Y)$ of the beta random variable yourself before looking at my solution!

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y)dy \\
&= \int_0^1 y^2 \left[\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\right] dy \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha+1}(1-y)^{\beta-1} dy \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)\Gamma(\beta)}{(\alpha+\beta+1)\Gamma(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha(\alpha+1)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}
$$

$$
\begin{aligned}
V(Y) &= E(Y^2) - [E(Y)]^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \left[\frac{\alpha}{(\alpha+\beta)}\right]^2 \\
&= \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha(\alpha+1)(\alpha+\beta) - \left(\alpha^3+\alpha^2\beta+\alpha^2\right)}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha^3+\alpha^2+\alpha\beta+\alpha^2\beta - \left(\alpha^3+\alpha^2\beta+\alpha^2\right)}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}
\end{aligned}
$$

EXERCISE 4.131

If $\alpha = 1$ and $\beta = 2,$ we have the following density function:

$$
f(y) = \begin{cases} 2(1-y) & 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}
$$

(a)

$$
\begin{aligned}
P(Y < 0.5) &= \int_0^{0.5} 2(1-y)\,dy \\
&= \left[2y - y^2\right]_0^{0.5} \\
&= 1 - 0.25 = 0.75
\end{aligned}
$$

(b)

$$
\begin{aligned}
E(Y) &= \frac{\alpha}{\alpha+\beta} \\
&= \frac{1}{1+2} = \frac{1}{3} \\
V(Y) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{(1)(2)}{(1+2)^2(1+2+1)} \\
&= \frac{2}{9\cdot 4} = \frac{1}{18} \\
\sigma &= \sqrt{\frac{1}{18}} = 0.2357
\end{aligned}
$$

4.8 Some general comments

The few paragraphs under this heading is a general discussion about the selection of the underlying distribution for a given population in inference. Read through it, but you will not be tested on the contents.

4.9 Other expected values

In the discussion on discrete probability distributions you came across the concept of a moment-generating function. Now we are going to investigate the moment-generating function and its calculation if the random variable is continuous. Then moments about the origin and the central moments are explained and you can familiarize yourself with the methods used to find the moment-generating function for a specified distribution by careful inspection of examples $4.12, 4.13$ and 4.14. Carefully work through all three to understand the series expansions. Calculation of such a moment-generating function is a typical examination question.
Theorem 4.12 gives the formula used in the calculation of the moment-generating function if you go beyond the random variable itself and consider some function of the random variable. As we continue, you can see how important it is for you as a statistician to have a solid knowledge of certain mathematical techniques. See the reason for my comment in Example 4.16, where the moment-generating function for a given function of a normally distributed continuous random variable is calculated.
Read the important paragraph about the uses of moment-generating functions (just after Example 4.16).

EXERCISE $\mathbf{4.143}$

From Example 4.14 the moment-generating function (mgf) for the gamma distribution is

$$m(t) = (1 - \beta t)^{-\alpha}$$

Differentiate this expression to find the moments about the origin and using your answers it will be possible, applying a mathematical manipulation, to find the moments around the mean (of which the variance is the second one).

$$\begin{aligned} m\,'(t) &= \frac{d}{dt}(1-\beta t)^{-\alpha} \\ &= -\alpha\,(1-\beta t)^{-\alpha-1}\cdot -\beta \\ &= \alpha\beta\,(1-\beta t)^{-\alpha-1} \end{aligned}$$

$$
\begin{aligned}
m'(t) \quad & | \quad {}_{t=0} = m'(0) = \alpha\beta = E(Y) \\
m''(t) &= \frac{d}{dt}\left[\frac{d}{dt}(1-\beta t)^{-\alpha}\right] \\
&= \frac{d}{dt}\left[\alpha\beta(1-\beta t)^{-\alpha-1}\right] \\
&= \alpha\beta(-\alpha-1)(1-\beta t)^{-\alpha-1-1}\cdot-\beta \\
&= \alpha\beta\left[-(\alpha+1)\right](1-\beta t)^{-\alpha-2}\cdot-\beta \\
&= \alpha\beta^2(\alpha+1)(1-\beta t)^{-\alpha-2} \\
m''(0) &= \alpha\beta^2(\alpha+1) = E(Y^2) \\
V(Y) &= E(Y^2) - [E(Y)]^2 \\
&= \alpha\beta^2(\alpha+1) - (\alpha\beta)^2 \\
&= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 \\
&= \alpha\beta^2
\end{aligned}
$$

EXERCISE 4.145

(a)

$$
\begin{aligned}
E\left(e^{\frac{3T}{2}}\right) &= \int_{-\infty}^{0} e^{\frac{3y}{2}} e^{y}\,dy \\
&= \int_{-\infty}^{0} e^{\frac{5y}{2}}\,dy \\
&= [\frac{2}{5}e^{\frac{5y}{2}}]_{-\infty}^{0} \\
&= \frac{2}{5}
\end{aligned}
$$

(b)

$$
\begin{aligned}
m(t) &= E\left(e^{tY}\right) \\
&= \int_{-\infty}^{0} e^{ty}e^{y}\,dy \\
&= \int_{-\infty}^{0} e^{y(t+1)}dy \\
&= \frac{1}{t+1}e^{y(t+1)}]_{-\infty}^{0} \\
&= \frac{1}{t+1}, t > -1
\end{aligned}
$$

(c) By using the method of mgfs:

$$
\begin{aligned}
E(Y) &= m'(t) = \frac{d}{dt}\left(\frac{1}{t+1}\right) \\
&= \frac{d}{dt}(t+1)^{-1} \\
&= -(t+1)^{-2} \\
m'(0) &= -1 \\
E(Y) &= m''(t) = \frac{d}{dt}\left[\frac{d}{dt}\left(\frac{1}{t+1}\right)\right] \\
&= \frac{d}{dt}\left[-(t+1)^{-2}\right] \\
&= 2(t+1)^{-3} \\
m''(0) &= 2(0+1)^{-3} = 2 \\
V(Y) &= E(Y^2) - [E(Y)]^2 \\
&= 2 - (-1)^2 \\
&= 1
\end{aligned}
$$

4.10 Tchebysheff's theorem

The basic statement in the theorem of Tchebysheff is the same, regardless of the nature of the random variable (discrete or continuous), i.e. Theorem 3.14 and Theorem 4.13 are identical. You will *not* be tested on the proof of Tchebysheff 's theorem in any of the two cases (Chapter 3 and 4). Just make sure that you know what the theorem states and that you can apply it in a question.

EXERCISE 4.147

We require $P(|Y-\mu| \leq 0.1) \geq 0.75 = 1 - \frac{1}{k^2}$.

Thus, $k = 2$.

Using Tchebysheff's inequality, $1 = k\sigma$, or $\sigma = \frac{1}{2}$.

4.11 Expectations of discontinuous functions and mixed probability distributions

This section is *not* included for examination purposes. We will not discuss it at all.

4.12 Summary

Read through the given summary in WM&S and make sure that you have a similar, but personalized summary in your mind.

Study objectives of this chapter

After completion of this chapter you should be able to

- define and characterize the distribution function for a discrete and continuous random variable
- understand the relationship between the distribution function and the density function of a continuous random variable and calculate the distribution function from the density function and vice versa
- describe and determine the expected value, given the distribution of any continuous random variable
- recognize the density functions for the uniform, normal, gamma (including the chi-square and the exponential distribution as special cases) and beta distribution
- define and calculate moments about the origin, as well as moments about the mean, for any continuous distribution using differentiation and/or a series expansion
- associate probabilities with the area under the density function (curve) of a continuous random variable
- use the probability density function of a random variable to find the moment-generating function associated with that random variable
- find the moment generating function of a function of a continuous random variable
- approximate certain probabilities for *any* probability distribution when only the mean and the variance are known by using Tchebysheff's theorem

---

Activity 4.1

This activity is once again for your own benefit.

1. Make a summary of the means and variances for the six continuous distributions discussed in this chapter.
2. See if you can calculate the moment-generating functions for the uniform, normal, gamma (with the chi-square and exponential as special cases) and beta distributed random variable.
3. Compare your summaries for 1 and 2 with Table 2 in Appendix 2 of WM&S (it is also on the second last page of WM&S) to see if you have the correct information. Do not simply copy those

summaries as you cannot benefit from that! Below is my summary where I included some of the characteristics of the different distributions – that is something I always like to add to a summary as it simplifies the process of recognition and comparison of the different functions. So, exactly how your *personal summar*y is going to look is in your own hands!

Name	**Characteristics: continuous densities**	**Density function**
Uniform $Y\tilde{}$unii(θ_1,θ_2)	density function: a constant parameters: $\theta_2$and θ_1 mgf: $m(t)=\dfrac{e^{2\theta_2}-e^{2\theta_2}}{t(\theta_2-\theta_1)}$ mean :$\frac{\theta_2+\theta_1}{2}$; variance:$\frac{(\theta_2-\theta_1)^2}{12}$	$f(y)=\begin{cases}\frac{1}{\theta_2-\theta_1} & \theta_1\leq y\leq\theta_2\\ 0 & \text{elsewhere}\end{cases}$
Normal $Y\tilde{}N(\mu,\sigma^2)$	density fs: bell-shaped around μ parameters: μ and σ μ (location); σ (measures spread) mgf: $m(t)=\exp\left(\mu t+\dfrac{t^2\sigma^2}{2}\right)$ mean : μ; variance: σ^2 $N(\mu,\sigma^2)$ transformable to $N(0,1)$	$f(y)=\left\{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\right. \quad -\infty<y<\infty$ Y defined over positive and negative values
Gamma $Y\tilde{}$gam(α,β)	density fs: long-tail to the right parameters: α and β (both >0) α (describes shape); β (scale) mgf: $m(t)=(1-\beta t)^{-\alpha}$ mean : $\alpha\beta$; variance: $\alpha\beta^2$	$f(y)=\begin{cases}\frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)} & 0\leq y<\infty\\ 0 & \text{elsewhere}\end{cases}$ Y defined only for positive values
Exponential $Y\tilde{}$exp(β)	density function: decreasing parameter: β mgf: $m(t)=(1-\beta t)^{-1}$ mean: β; variance: β^2 special gamma with $\alpha=1$ -	$f(y)=\begin{cases}\frac{1}{\beta}e^{-\frac{y}{\beta}} & 0<y<\infty\\ 0 & \text{elsewhere}\end{cases}$ Y defined only for positive values
Chi-square $Y\tilde{}\chi^2(\nu)$	density fs: positively skewed parameter: ν (degrees of freedom) mgf: $m(t)=(1-2t)^{-\dfrac{\nu}{2}}$ mean: ν; variance: ν^2 special gamma with $\alpha=\frac{\nu}{2}$; $\beta=2$	$f(y)=\begin{matrix}\frac{y^{\frac{\nu}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} & y^2>0\\ 0 & \text{elsewhere}\end{matrix}$ Y defined only for positive values
Beta $Y\tilde{}$beta(α,β)	density fs: different shapes parameters: α and β no mgf in closed form mean: $\dfrac{\alpha}{\alpha+\beta}$; var:$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ also: $Y\tilde{}$beta type I	$f(y)=\begin{matrix}\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & 0\leq y\leq 1\\ 0 & \text{elsewhere}\end{matrix}$ $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ only for $0\leq Y\leq 1$

CHAPTER 5

Multivariate probability distributions

STUDY
WM&S: **Chapter** 5 (from $\mathbf{5.1}$only up to $\mathbf{5.7}$).

Although the heading of this chapter refers to *multivariate* distributions, the emphasis is on *bivariate* distributions. Please concentrate on the bivariate cases and make sure you understand how the move from one to two random variables influences the formulas and calculations. Then, in follow-up modules, it will not be difficult to expand your knowledge to *any number* of variables.

5.1 Introduction

You will very seldom find an experiment in practice involving only one random variable. We therefore have to move to more than one random variable and expand all the knowledge we have so far to two varables (to start with) instead of one only. In $STA1503$ and $STA2603$ we concentrate on *one and two random variables* and when you get to $STA3703$ you will learn about *multivariate probability distributions*. The title of this chapter 5 is a bit misleading as the main interest lies with *bivariate probability distributions*, but that is no problem as you do not have to learn about the multivariate option in this module.

Chapter 5 of WM&S has more sections than those that we are going to discuss in this guide. Sections $\mathbf{5.8}$ to $\mathbf{5.11}$ do not form part of $STA2603$.

5.2 Bivariate (and multivariate) probability distributions

You are given a very understandable explanation of how to progress in your knowledge of one discrete random variable to two discrete random variables, leading to the definition of a joint probability mass function and its characteristics. The definition of a joint distribution function is also just an expansion of the one given in Definition 4.1. Just as in the case of one random variable, you have to know the difference between discrete and continuous random variables and their distribution functions.

Earlier, we defined the *distribution function* for a continuous random variable Y as

$$F(y) = \int_{-\infty}^{y} f(t)dt$$

Now, for two continuous random variables Y_1 and Y_2, we define a *joint distribution function* as

$$\boxed{F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)dt_2dt_1}$$

with the function $f(y_1, y_2)$ the joint probability density function of Y_1 and Y_2.

Recall that we said earlier that statistical knowledge builds up as you continue with your studies. If you did not understand the concepts defined for one random variable, there is no way that you will understand and remember what you get in this section. On the other hand, if you have mastered the concepts for one random variable, what you see now, seems quite similar to what you already know! There are good examples in WM&S to work through. You are shown in Example 5.4 how important the use of mathematical techniques are in solving our statistical problems. Note how the sketches help us to find the region of integration.

As said, we will not give serious attention to increasing the number of random variables up to n in this module, but concentrate on one and two random variables, i.e. up to the bivariate case. In the paragraphs just before exercise 5.1, the formulas for n random variables are given. The beautiful thing about these formulas is that making $n = 1$ results in the single variable formulas, and for $n = 2$ you get the bivariate formulas.

I will show you a number of exercises from WM&S to assist you in understanding the expansion from one variable to two variables.

EXERCISE 5.1

It is very important that you can give the sample space for the given scenario. To be able to do that, make sure that you realize:

- There are only *two* construction jobs and there are *three* firms, A, B and C.
- Then take note of the given information that per definition

  Y_1 denotes the contracts of firm A, so Y_1 can take on the values $0, 1$ and 2

  Y_2 denotes the contracts of firm B, so Y_2 can take on the values $0, 1$ and 2

  Nothing is said about the contracts of firm C!!

(a) The sample space S is therefore all possible allocations of the two construction jobs in terms of Y_1 and Y_2.

  In the table below the joint set indicates:

$(y_1, y_2) =$ (number of allocations to A in the set, number of allocations to B in the set)

S	AA	AB	AC	BA	BB	BC	CA	CB	CC
$(\mathbf{y_1}, \mathbf{y_2})$	$(2,0)$	$(1,1)$	$(1,0)$	$(1,1)$	$(0,2)$	$(0,1)$	$(1,0)$	$(0,1)$	$(0,0)$

Since each sample point is equally likely with probability $\frac{1}{9}$, the joint distribution for Y_1 and Y_2 is given by

$\mathbf{y_2}$ \ $\mathbf{y_1}$	**0**	**1**	**2**
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$
1	$\frac{2}{9}$	$\frac{2}{9}$	0
2	$\frac{1}{9}$	0	0

(b) $F(1,0) = p(0,0) + p(1,0) = \frac{1}{9} + \frac{2}{9} = \frac{3}{9} = \frac{1}{3}$

EXERCISE 5.5

(a) Make sure that you use the limits for Y_1 when you intergrate y_1 (i.e. dy_1) and use the limits for Y_2 when you integrate y_2 (i.e. dy_2).

You get the same answer for $\int_0^{\frac{1}{3}}\int_0^{\frac{1}{2}} 3y_1 dy_1 dy_2$ and $\int_0^{\frac{1}{2}}\int_0^{\frac{1}{3}} 3y_1 dy_2 dy_1$ because in both cases the above comment applies.

$$\begin{aligned}
P(Y_1 &\leq \frac{1}{2}, Y_2 \leq \frac{1}{3}) \\
&= \int_0^{\frac{1}{3}}\int_0^{\frac{1}{2}} 3y_1 dy_1 dy_2 \\
&= \int_0^{\frac{1}{3}} \left(\frac{3}{2} y_1^2 \Big|_0^{\frac{1}{2}} \right) dy_2 \\
&= \frac{3}{2}\int_0^{\frac{1}{3}} \left(\left(\frac{1}{2}\right)^2 - (0)^2 \right) dy_2 \\
&= \frac{3}{2}\cdot\frac{1}{4}\int_0^{\frac{1}{3}} dy_2 \\
&= \frac{3}{8}(y_1)\,|_0^{\frac{1}{3}} \\
&= \frac{3}{8}\cdot\frac{1}{3} \\
&= \frac{1}{8} = 0.125
\end{aligned}$$

By now I am sure that you realize the importance of your mathematical knowledge of double integrals.

A reminder about the order of integration: Find the inner integral with its correct limits and see that in your mind's eye inside brackets. Then start with the information inside the brackets.

$$\int_0^{\frac{1}{3}} \left(\int_0^{\frac{1}{2}} 3y_1 dy_1 \right) dy_2$$

In this example the bracket implies integrating y_1 and you have the limits for y_1 in the definite integral, namely 0 up to $\frac{1}{2}$ (inside that brackets).

Once you have calculated that inner integral, you are left with one integration, namely for y_2. Notice that the limits in the remaining integal are those for y_2 namely 0 to $\frac{1}{3}$.

$$\int_0^{\frac{1}{3}} (- - - - - -)\, dy_2$$

If the $dy_1 dy_2$ were given the other way round, namely $dy_2 dy_1$,you would have first integrated over dy_2 and used the integral with limits 0 to $\frac{1}{3}$.Once this was completed, you would integrate over dy_1 with limits 0 up to $\frac{1}{2}$. You should get the same answer!

(b)

$$\begin{aligned}
P(Y_2 &\leq \frac{Y_1}{2}) \\
&= \int_0^1 \int_0^{\frac{y_1}{2}} 3y_1 dy_2 dy_1 \\
&= \int_0^1 3y_1 \left[(y_2) \mid_0^{\frac{1}{2}} \right] dy_1 \\
&= \int_0^1 3y_1 \left[\left(\frac{y_1}{2}\right) - (0) \right] dy_1 \\
&= \int_0^1 \frac{3}{2} y_1^2 dy_1 \\
&= \frac{3}{2} \int_0^1 y_1^2 dy_1 \\
&= \frac{3}{2} \left(\frac{y_1^3}{3} \right) \Big|_0^1 \\
&= \frac{3}{2} \cdot \left(\frac{1}{3} - 0 \right) \\
&= \frac{1}{2} = 0.500
\end{aligned}$$

EXERCISE 5.7

(a)

$$\begin{aligned}
P(Y_1 &< 1, Y_2 > 5) = \int_0^1 \int_5^\infty e^{-(y_1+y_2)} dy_2 dy_1 \\
&= \int_0^1 \int_5^\infty e^{-y_1} e^{-y_2} dy_2 dy_1 \\
&= \int_0^1 e^{-y_1} \left[\int_5^\infty e^{-y_2} dy_2 \right] dy_1 \\
&= \int_0^1 e^{-y_1} \left(-e^{-y_2} \Big|_5^\infty \right) dy_1 \\
&= \int_0^1 e^{-y_1} \left(-0 + e^{-5} \right) dy_1 \\
&= e^{-5} \int_0^1 e^{-y_1} dy_1 \\
&= e^{-5} \left(-e^{-y_1)} \Big|_0^1 \right) \\
&= e^{-5} \left(-e^{-1} + e^0 \right) \\
&= e^{-5} \left(1 - e^{-1} \right) \\
&= 0.00426
\end{aligned}$$

(b)

$$
\begin{aligned}
P(Y_1 + Y_2 &< 3) = P(Y_1 < 3 - Y_2) \\
&= \int_0^3 \left[\int_5^{3-y_2} e^{-(y_1+y_2)} dy_1 \right] dy_2 \\
&= \int_0^3 \left[\int_5^{3-y_2} e^{-(y_1+y_2)} dy_1 \right] dy_2 \\
&= \int_0^3 e^{-y_2} \left[\int_0^{3-y_2} e^{-y_1} dy_1 \right] dy_2 \\
&= \int_0^3 e^{-y_2} \left(-e^{-y_1} \; |_0^{3-y_2} \right) dy_2 \\
&= \int_0^3 e^{-y_2} \left(-e^{y_2-3} + e^0 \right) dy_2 \\
&= \int_0^3 e^{-y_2} \left(1 - e^{-3} e^{+y_2} \right) dy_2 \\
&= \int_0^3 \left[e^{-y_2} - e^{-3} \right] dy_2 \\
&= \left[-e^{-y_2} - e^{-3} y_2 \right] |_0^3 \\
&= \left[\left(-e^{-3} - 3e^{-3} \right) - (-1 - 0) \right] \\
&= \left(-4e^{-3} + 1 \right)
\end{aligned}
$$

EXERCISE 5.9

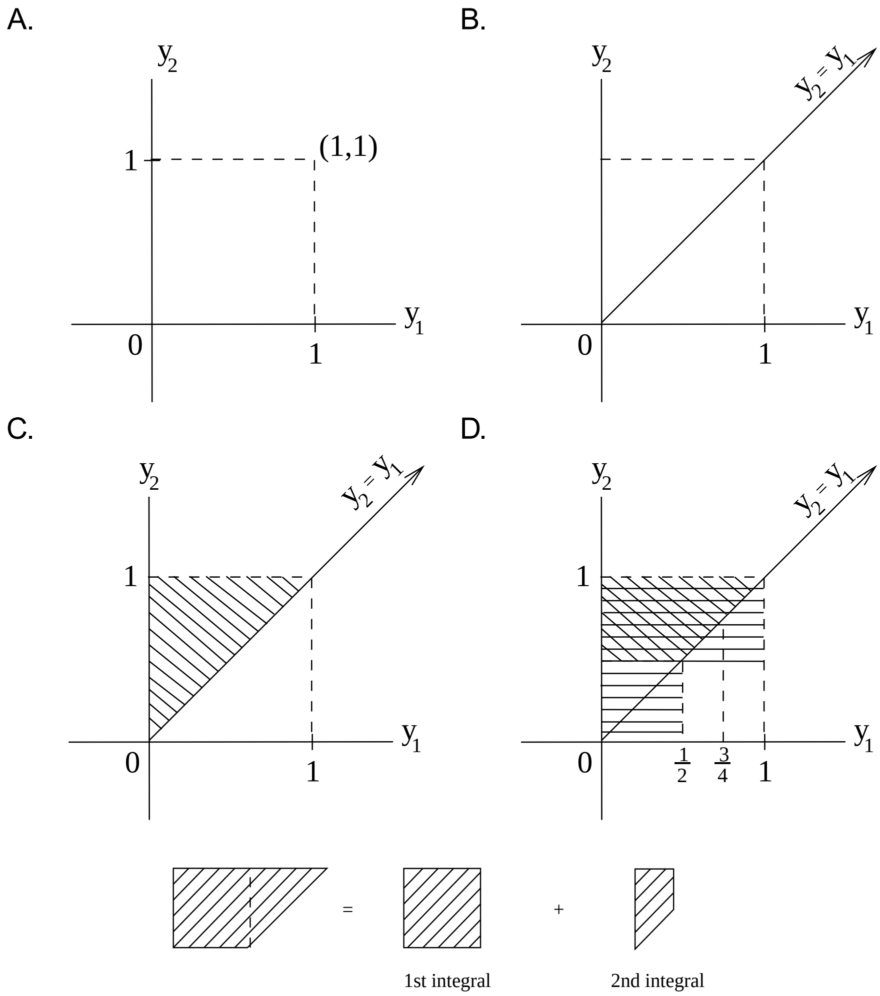
(a)

$$\begin{aligned}
\int_0^1 \int_0^{y_2} k\,(1-y_2)\,dy_1 dy_2 &= \int_0^1 \left[\int_0^{y_2} k\,(1-y_2)\,dy_1\right] dy_2 \\
&= \int_0^1 \left[(ky_1 - ky_2y_1)\ |_0^{y_2}\right] dy_2 \\
&= \int_0^1 \left[\left(ky_2 - ky_2^2\right) - 0\right] dy_2 \\
&= \int_0^1 \left(ky_2 - ky_2^2\right) dy_2 \\
&= \left(\frac{ky_2^2}{2} - \frac{ky_2^3}{3}\right)\Big|_0^1 \\
&= \left(\frac{k\,(1)^2}{2} - \frac{k\,(1)^3}{3}\right) - 0 \\
&= \frac{3k-2k}{6} \\
&= \frac{k}{6}
\end{aligned}$$

For a density function the value of the integral must be equal to 1, so

$$\begin{aligned}
\frac{k}{6} &= 1 \\
k &= 6
\end{aligned}$$

(b) Note the limits for the variables in $f(y_1, y_2) = 6(1 - y_2)$, namely $0 \le y_1 \le y_2 \le 1$. This implies that the probability requested must be found in two parts and can be illustrated using a few graphs. (I am not considering $f(y_1, y_2)$ at all – I am simply looking at y_1 and y_2 as you did in school with straight line graphs.)



- Graph A shows you the regions for y_1 and y_2 ; both may only extend from 0 to 1.
- In graph B the straight line $y_2 = y_1$ is included (think of it as $y = x$ in school).
- In graph C the straight line area above the graph $y_2 = y_1$ is shaded. That is to satisfy the given condition that, apart from both variables lying from 0 to 1, it is necessary that $y_1 \le y_2$.
- In graph D the two sections $y_1 \le \frac{3}{4}$ and $y_2 > \frac{1}{2}$ have been shaded. Find their overlapping section, but make sure you are still considering the region given in graph C.

The little sketches below the four graphs A to D explain why the integration must be done in two parts. The section for integration is the section in sketch D that was shaded twice and that is the part we break up into two sections for integration. The need for the split lies in the different regions of integration.

- The first part is integrated over the region ($\frac{1}{2}$ to1) for *both* y_1 and y_2.
- The second integration is over ($\frac{1}{2}$ to $\frac{3}{4}$) for y_1 and from the line (i.e. $y_2 = y_1$) to 1 for y_2.

I hope you understand this now!!

$$P(Y_1 \leq \frac{3}{4}, Y_2 \geq \frac{1}{2}) = \int_{\frac{1}{2}}^{1}\int_{\frac{1}{2}}^{1} 6\,(1-y_2)\,dy_1 dy_2 + \int_{\frac{1}{2}}^{\frac{3}{4}}\int_{y_1}^{1} 6\,(1-y_2)\,dy_2 dy_1$$

$$= \int_{\frac{1}{2}}^{1}\left[\int_{\frac{1}{2}}^{1} 6\,(1-y_2)\,dy_1\right] dy_2 + \int_{\frac{1}{2}}^{\frac{3}{4}}\left[\int_{y_1}^{1} 6\,(1-y_2)\,dy_2\right] dy_1$$

Now you can determine these integrations yourself!!

$$= \frac{24}{64} + \frac{7}{64}$$

$$= \frac{31}{64}$$

EXERCISE 5.11

I want you to recall your previous school knowledge and not imagine that everything we teach you in this module is new and difficult! Look at the given sketch and you will see it is a triangle. Do you remember the formula for the area of a triangle? Ignore the fact that this triangle is placed in the Cartesian plane and follow my reasoning below:

Area triangle = $\frac{1}{2}\cdot$ base $\cdot$ perperdicular height.

In this triangle the base is 2 units long and the perpendicular height is 1 unit, so

Area triangle = $\frac{1}{2} \cdot\ 2 \cdot\ 1 = 1.$

This answer of 1 is the same as the value of the area under any density curve – remember?

Given that Y_1 and Y_2 are *uniformly distributed,* implies an area in the form of a *rectangle*. We do not have the detail to sketch this density, but we are given a triangle indicating restrictions on the range of the random variables Y_1 and Y_2.

(a) You have to determine

$$P(Y_1 \leq \frac{3}{4}, Y_2 \leq \frac{3}{4})$$

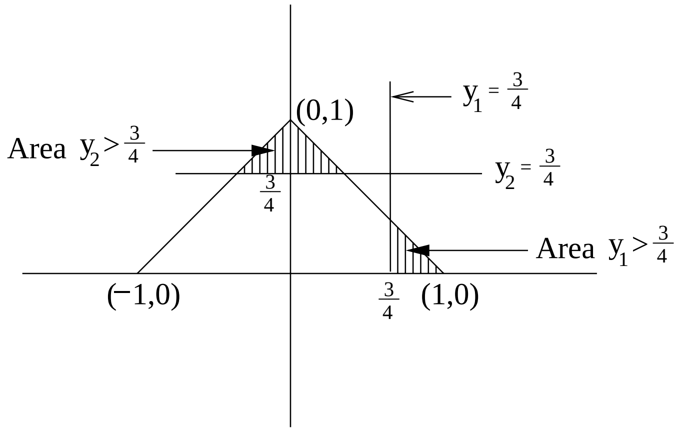
In the sketch below I added two straight lines:

- A horizontal line with equation $y_2 = \frac{3}{4}$. The area *under* this line, but still in the triangle will indicate $y_2 < \frac{3}{4}$ and the area *above* this line is where $y_2 > \frac{3}{4}$.
  I have shaded the area $y_2 > \frac{3}{4}$, but only where it overlaps with the triangle.
- A vertical line with equation $y_1 = \frac{3}{4}$. The area *to the left of* this line, but still in the triangle,

will indicate $y_1 < \frac{3}{4}$ and the area *to the right of* this line is where $y_1 > \frac{3}{4}$.

I have shaded the area $y_1 > \frac{3}{4}$, but only where it overlaps with the triangle.

The area corresponding to the question is the *area of the triangle that is not shaded in my sketch.*



In practice it is easier to determine the two small shaded areas in the triangle than to find the area of the non-shaded area. We have a neat way to do just that and find the answer in an easier way because we have calculated that the total area of the triangle is one (note that we are just calculating the areas of the two shaded triangles using school knowledge about the area of a triangle):

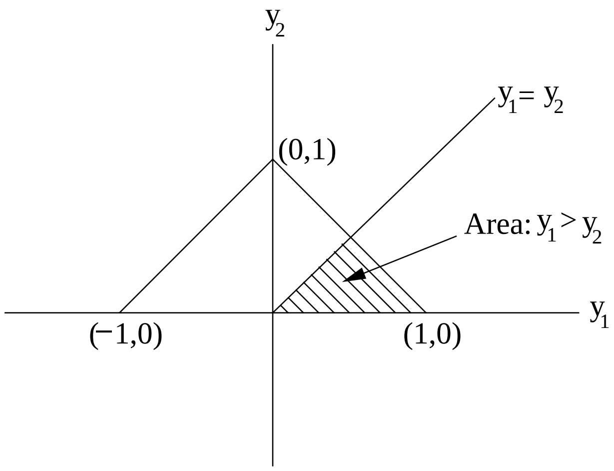
$$\begin{aligned}
P(Y_1 &\leq \frac{3}{4}, Y_2 \leq \frac{3}{4}) = 1 - P(Y_1 \geq \frac{3}{4}) - P(Y_2 \geq \frac{3}{4}) \\
&= 1 - \text{half of } \left[\left(\frac{1}{4} + \frac{1}{4}\right) \cdot \left(\frac{1}{4}\right)\right] - \text{half of } \left[\left(\frac{1}{4}\right) \cdot \left(\frac{1}{4}\right)\right] \\
&= 1 - \left[\left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) \cdot \left(\frac{1}{4}\right)\right] - \left[\left(\frac{1}{2}\right) \cdot \left(\frac{1}{4}\right) \cdot \left(\frac{1}{4}\right)\right] \\
&= 1 - \frac{1}{16} - \frac{1}{32} \\
&= \frac{29}{32}
\end{aligned}$$

(b)

$$P(Y_1 - Y_2 \geq 0) = P(Y_1 \geq Y_2)$$

I am once again giving you a sketch and in this sketch, I have added the graph of $y_1 = y_2$.

I shaded the area *below* the line and this represents the region where $y_1 \geq y_2$, restricted within the given triangle.



Can you see that the shaded area is *one quarter* of the total area of the triangle? Therefore

$$P(Y_1 - Y_2 \geq 0) = P(Y_1 \geq Y_2) = \frac{1}{4}.$$

EXERCISE 5.15

$$\begin{aligned}
P(Y_1 &< 2, Y_2 > 1) = \int_0^1 \int_1^{y_1} e^{-y_1} dy_2 dy_1 \quad \text{remember } Y_1 \geq Y_2 \\
&= \int_0^1 \int_{y_2}^2 e^{-y_1} dy_1 dy_2 \quad \text{if } y_2 \text{ goes from } 1 \text{ to } 2;\ y_1 \text{ goes from } y_2 \text{ to} 2 \\
&= \int_0^1 \left(\int_{y_2}^2 e^{-y_1} dy_1 \right) dy_2 \\
&= \int_0^1 \left[-e^{-y_1} \right] |_{y_2}^2 dy_2 \\
&= \int_0^1 \left[-e^{-2} - -e^{-y_2} \right] dy_2 \\
&= \int_0^1 \left[-e^{-2} + e^{-y_2} \right] dy_2 \\
&= \left[-y_2 e^{-2} + \left(-e^{-y_2} \right) \right] |_1^2 \\
&= \left[-2e^{-2} + \left(-e^{-2} \right) \right] - \left[-1e^{-2} + \left(-e^{-1} \right) \right] \\
&= \left[-2e^{-2} - e^{-2} \right] - \left[-e^{-2} - e^{-1} \right] \\
&= -2e^{-2} - e^{-2} + e^{-2} + e^{-1} \\
&= e^{-1} - 2e^{-2}
\end{aligned}$$

5.3 Marginal and conditional probability distributions

Now we get to the *marginal probability function*, where you eliminate one of the two random variables. This is applicable for two discrete and continuous random variables.

Discrete variable: The marginal probability function of y_1 : sum over the values of y_2.

Notation: $p_1(y_1)$

Discrete variable: The marginal probability function of y_2 : sum out over the values of y_1.

Notation: $p_2(y_2)$

Continuous variable: The marginal density function of y_1 : integrate over the values of y_2.

Notation: $f_1(y_1)$

Continuous variable: The marginal density function of y_2 : integrate over the values of y_1.

Notation: $f_2(y_2)$

EXERCISE 5.19

When the joint distribution of two discrete random variables y_1 and y_2 are displayed like this, the marginal probability functions are simply the row and column totals. I am sure you recognized this question and can link it to 5.1, which I discussed in the previous section.
Here we once again have the joint distribution of Y_1 and Y_2:

		$\mathbf{y_1}$			
		0	**1**	**2**	**Row totals**
$\mathbf{y_2}$	**0**	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}+\frac{2}{9}+\frac{1}{9}=\frac{4}{9}$
	1	$\frac{2}{9}$	$\frac{2}{9}$	0	$\frac{2}{9}+\frac{2}{9}=\frac{4}{9}$
	2	$\frac{1}{9}$	0	0	$\frac{1}{9}+0+0=\frac{1}{9}$
Column totals:		$\frac{1}{9}+\frac{2}{9}+\frac{1}{9}=\frac{4}{9}$	$\frac{2}{9}+\frac{2}{9}=\frac{4}{9}$	$\frac{1}{9}+0+0=\frac{1}{9}$	

(a) The marginal distribution of Y_1 (column totals) can be given in a table:

y_1	0	1	2
$p_1(y_1)$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

(b) For the binomial distribution we have

$$\begin{aligned} p(y) &= \binom{n}{y} p^y (1-p)^{n-y} \\ f_1(0) &= \binom{2}{0}\left(\frac{1}{3}\right)^0 (\frac{2}{3})^2 = \frac{4}{9} \\ f_1(1) &= \binom{2}{1}\left(\frac{1}{3}\right)^1 (\frac{2}{3})^1 = \frac{4}{9} \\ f_1(2) &= \binom{2}{2}\left(\frac{1}{3}\right)^2 (\frac{2}{3})^0 = \frac{1}{9} \end{aligned}$$

Obviously, as expected, there is no contradiction in the results!

EXERCISE 5.23

(a)

$$\begin{aligned} f_2(y_2) &= \int_{y_2}^{1} 3y_1 dy_1 \\ &= 3\left(\frac{y_1^2}{2}\right)\Big|_{y_2}^{1} \\ &= 3\left[\frac{1}{2} - \frac{y_2^2}{2}\right] \\ &= \frac{3}{2} - \frac{3}{2}y_2^2 \qquad\qquad 0 \leq y_2 \leq 1 \end{aligned}$$

Note that it is extremely important when you give the equation for any distribution that you specify the relevant limits for the variable concerned. I know lecturers who penalize seriously if students do not give the limits. (I hope you realize that I am referring to the $0 \leq y_2 \leq 1$!!!)

(b) You need to consider $f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$ and give the limits of y_2. Because the denominator in any fraction may never be zero, you know that $f_2(y_2) \neq 0$ or more specifically $\frac{3}{2} - \frac{3}{2}y_2^2 \neq 0$. None of the values of y_2 from 0 to 1 will make this expression equal to zero, so y_2 can be ≥ 0. Furthermore it was given that $y_2 \leq y_1 \leq 1$.The answer to the question is therefore:

$f(y_1 \mid y_2)$ is defined over $y_2 \leq y_1 \leq 1$, with the constant $y_2 \geq 0$.

(c) To determine

$$f(y_2 \mid y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

first find $f_1(y_1)$ – the marginal distribution of y_1.

$$\begin{aligned} f_1(y_1) &= \int_{0}^{y_1} 3y_1 dy_2 \\ &= 3y_1 \int_{0}^{y_1} dy_2 \\ &= 3y_1(y_2)\mid_0^{y_1} \\ &= 3y_1(y_1 - 0) \\ &= 3y_1^2 \end{aligned}$$

Now it is possible to find the conditional distribution

$$\begin{aligned} f(y_2 \mid y_1) &= \frac{f(y_1, y_2)}{f_1(y_1)} \\ &= \frac{3y_1}{3y_1^2} \\ &= \frac{1}{y_1} \qquad\qquad 0 \le y_2 \le y_1 \end{aligned}$$

Conditioned on $Y_1 = y_1$ we see that y_2 has a uniform distribution on the interval $(0, y_1)$. Recall that the calculation of a probability from θ_1 to θ_2 for a continuous random variable, said to have a uniform distribution, is given by $\frac{\theta_2 - \theta_1}{\theta_1}$. In this case

$$\begin{aligned} P\left(Y_2 > \frac{1}{2} \mid Y_1 = \frac{3}{4}\right) &= \frac{\frac{3}{4} - \frac{1}{2}}{\frac{3}{4}} \\ &= \frac{1}{3} \end{aligned}$$

You could also have reasoned as follows:

$$\begin{aligned} P\left(Y_2 > \frac{1}{2} \mid Y_1 = \frac{3}{4}\right) &= \int_{-\infty}^{1} f(y_2 \mid y_1)\, dy_2 \\ &= \int_{-\infty}^{1} \frac{1}{y_1} dy_2 \\ \text{But} \quad y_1 &= \frac{3}{4} \text{ and } y_2 \le y_1, \text{ so this becomes} \\ P\left(Y_2 > \frac{1}{2} \mid Y_1 = \frac{3}{4}\right) &= \frac{1}{y_1} \int_{\frac{1}{2}}^{\frac{3}{4}} dy_2 \\ &= \frac{1}{y_1} \; y_2 \Big|_{\frac{1}{2}}^{\frac{3}{4}} \\ &= \frac{1}{y_1}\left(\frac{3}{4} - \frac{1}{2}\right) \\ &= \frac{1}{\frac{3}{4}}\left(\frac{3}{4} - \frac{1}{2}\right) \quad \text{because } y_1 = \frac{3}{4} \\ &= \frac{1}{3} \end{aligned}$$

EXERCISE 5.25

(a)

$$\begin{aligned}
f_1(y_1) &= \int_0^{\infty} e^{-(y_1+y_2)} dy_2 \\
&= e^{-y_1} \int_0^{\infty} e^{-y_2} dy_2 \\
&= e^{-y_1} \left[\left(-e^{-y_2} \right) |_0^{\infty} \right] \\
&= e^{-y_1} \left[-0 + e^{-0} \right] \\
&= e^{-y_1} \qquad y_1 > 0
\end{aligned}$$

$$\begin{aligned}
f_2(y_2) &= \int_0^{\infty} e^{-(y_1+y_2)} dy_1 \\
&= e^{-y_2} \int_0^{\infty} e^{-y_1} dy_1 \\
&= e^{-y_2} \left[\left(-e^{-y_1} \right) |_0^{\infty} \right] \\
&= e^{-y_2} \left[-0 + e^{-0} \right] \\
&= e^{-y_2} \qquad y_2 > 0
\end{aligned}$$

Both the marginal distributions have exponential density functions with $\beta = 1$.

(b)

$$\begin{aligned}
P(1 < Y_1 < 2.5) &= \int_1^{2.5} e^{-y_1} dy_1 \\
&= \left[\left(-e^{-y_1} \right) |_1^{2.5} \right] \\
&= e^{-1} - e^{-2.5} = 0.2858
\end{aligned}$$

And in the same way you will find

$$P(1 < Y_2 < 2.5) = e^{-1} - e^{-2.5} = 0.2858$$

(c) As said before, the denominator $f_2(y_2)$ in the fraction for the conditional density function $f(y_1 \mid y_2) = \frac{f(y_1,y_2)}{f_2(y_2)}$ may not assume the value 0.

Seeing that e^{-y_2} cannot become zero there is no further restriction than the given $y_2 > 0$

(d)

$$\begin{aligned} f(y_1 \mid y_2) &= \frac{f(y_1, y_2)}{f_2(y_2)} \\ &= \frac{e^{-(y_1+y_2)}}{e^{-y_2}} \\ &= e^{-y_1}, \qquad y_1 > 0 \\ \text{So} \quad f(y_1 \mid y_2) &= f_1(y_1) \end{aligned}$$

(e)

$$\begin{aligned} f(y_2 \mid y_1) &= \frac{f(y_1, y_2)}{f_1(y_1)} \\ &= \frac{e^{-(y_1+y_2)}}{e^{-y_1}} \\ &= e^{-y_2}, \qquad y_2 > 0 \\ \text{So} \quad f(y_2 \mid y_1) &= f_2(y_2) \end{aligned}$$

(f) The answers are the same.

(g) The probabilities will be the same.

EXERCISE 5.27

(a)

$$\begin{aligned} f_1(y_1) &= \int_{y_1}^{1} 6(1-y_2)\,dy_2 \\ &= 6\int_{y_1}^{1} dy_2 - 6\int_{y_1}^{1} y_2 dy_2 \\ &= 6\left[(y_2)\,\Big|_{y_1}^{1}\right] - 6\left[\left(\frac{y_2^2}{2}\right)\Bigg|_{y_1}^{1}\right] \\ &= 6\left[(y_2)\,\Big|_{y_1}^{1}\right] - 3\left[\left(y_2^2\right)\,\Big|_{y_1}^{1}\right] \\ &= 6\left[(1-y_1)\right] - 3\left[\left(1-y_1^2\right)\right] \\ &= 6 - 6y_1 - 3 + 3y_1^2 \\ &= 3\left(y_1^2 - 2y_1 + 1\right) \\ &= 3(y_1-1)^2 \qquad 0 \le y_1 \le 1 \end{aligned}$$

$$\begin{aligned} f_2(y_2) &= \int_0^{y_2} 6(1-y_2)\,dy_1 \\ &= 6(1-y_2)\int_0^{y_2} dy_1 \\ &= 6(1-y_2)\left[(y_1)\,|_0^{y_2}\right] \\ &= 6(1-y_2)(y_2-0) \\ &= 6y_2(1-y_2) \qquad 0 \le y_2 \le 1 \end{aligned}$$

(b) Use the standard form of the conditional density function of y_2 given y_1.

$$f(y_2 \mid y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

Note:

Treat the two integrals in the numerator and in the denominator of the conditional density separately in your calculation. It is quite lengthy, but not so complex!

Use the given limits for y_1 and y_2 and keep in mind that $y_1 \le y_2$.

$$\begin{aligned} P(Y_2 &\le \frac{1}{2} \mid Y_1 \le \frac{3}{4}) = \frac{\int_0^{\frac{1}{2}}\int_0^{y_2} 6(1-y_2)\,dy_1 dy_2}{\int_0^{\frac{3}{4}} 3(y_1-1)^2\,dy_1} \\ &= \frac{\int_0^{\frac{1}{2}}\left[\int_0^{y_2} 6(1-y_2)\,dy_1\right] dy_2}{3\int_0^{\frac{3}{4}} (y_1-1)^2\,dy_1} \\ &= \frac{\int_0^{\frac{1}{2}} \left[6y_1(1-y_2)\,|_0^{y_2}\right] dy_2}{3\int_0^{\frac{3}{4}} \left(y_1^2 - 2y_1 + 1\right) dy_1} \\ &= \frac{\int_0^{\frac{1}{2}} \left[\left(6y_2 - 6y_2^2\right) - 6(0 - 0y_2)\right] dy_2}{3\left[\left(\frac{y_1^3}{3} - y_1^2 + y_1\right)|_0^{\frac{3}{4}}\right]} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\int_0^{\frac{1}{2}} \left[6\left(y_2 - y_2^2\right)\right] dy_2}{3\left[\frac{1}{3}\left(\frac{27}{64}\right) - \frac{9}{16} + \frac{3}{4}\right]} \\
&= \frac{\left[6\left(\frac{y_2^2}{2}\right) - 6\left(\frac{y_2^3}{3}\right) |_0^{\frac{1}{2}}\right]}{\frac{27}{64} - \frac{27}{16} + \frac{9}{4}} \\
&= \frac{\left[3y_2^2 - 2\left(y_2^3\right) |_0^{\frac{1}{2}}\right]}{\frac{27-108+144}{64}} \\
&= \frac{\left[3\left(\frac{1}{4}\right) - 2\left(\frac{1}{8}\right) - \left(3\left(0\right) - 2\left(0\right)\right)\right]}{\frac{63}{64}} \\
&= \frac{\left[\frac{3}{4} - \frac{1}{4}\right]}{\frac{63}{64}} \\
&= \frac{\left(\frac{1}{2}\right)}{\frac{63}{64}} \\
&= \frac{1}{2} \cdot \frac{64}{63} \\
&= \frac{32}{63}
\end{aligned}
$$

(c)

$$
\begin{aligned}
f\left(y_1 \mid y_2\right) &= \frac{f\left(y_1, y_2\right)}{f_2\left(y_2\right)} \\
&= \frac{6\left(1 - y_2\right)}{6y_2\left(1 - y_2\right)} \\
&= \frac{1}{y_2} \qquad 0 \le y_1 \le y_2 \le 1
\end{aligned}
$$

(d)

$$
\begin{aligned}
f\left(y_2 \mid y_1\right) &= \frac{f\left(y_1, y_2\right)}{f_1\left(y_1\right)} \\
&= \frac{6\left(1 - y_2\right)}{3\left(1 - y_1\right)^2} \\
&= \frac{2\left(1 - y_2\right)}{\left(1 - y_1\right)^2} \qquad 0 \le y_1 \le y_2 \le 1
\end{aligned}
$$

(e) From the answer in (d):

$$
\begin{aligned}
f\left(y_2 \mid y_1\right) &= \frac{6\left(1-y_2\right)}{3\left(1-y_1\right)^2} \\
f\left(y_2 \mid \frac{1}{2}\right) &= \frac{6\left(1-y_2\right)}{3\left(1-\frac{1}{2}\right)^2} \\
&= \frac{6\left(1-y_2\right)}{\frac{3}{4}} \\
&= 8\left(1-y_2\right) \qquad \frac{1}{2} \leq y_2 \leq 1
\end{aligned}
$$

Therefore

$$
\begin{aligned}
P(Y_2 &\geq \frac{3}{4} \mid Y_1 = \frac{1}{2}) = \int_{\frac{3}{4}}^{1} 8\left(1-y_2\right) dy_2 \\
&= 8\int_{\frac{3}{4}}^{1} \left(1-y_2\right) dy_2 \\
&= 8\left[\left(y_2 - \frac{y_2^2}{2}\right)\Big|_{\frac{3}{4}}^{1}\right] \\
&= 8\left[\left(1-\frac{1^2}{2}\right)\left(\frac{3}{4}-\frac{\left(\frac{3}{4}\right)^2}{2}\right)\right] \\
&= 8\left[\frac{1}{2}-\left(\frac{3}{4}-\frac{\left(\frac{9}{16}\right)}{2}\right)\right] \\
&= 8\left[\frac{16}{32}-\left(\frac{24}{32}-\frac{9}{32}\right)\right] \\
&= \frac{8}{1}\cdot\frac{1}{32} \\
&= \frac{1}{4}
\end{aligned}
$$

5.4 Independent random variables

Definition 5.8, and Theorem 5.4 give the conditions for two random variables Y_1 and Y_2 to be independent.

In 5.8 the definition applies to the distribution functions of the two variables – whether discrete or continuous.

Theorem 5.4 has two sections:

- giving the condition in terms of probability functions for discrete random variables
- giving the condition in terms of density functions for continuous random variables

Read through theorem 5.5 (not important for this module)

Work through the examples. Many of these are continuations of previous discussions, now requesting you to test if the two variables are independent or not.

EXERCISE 5.45

No independence. Take for example

$$P(Y_1 = 2, Y_2 = 2) = 0 \neq P(Y_1 = 2) \cdot P(Y_2 = 2) = \frac{1}{9} \cdot \frac{1}{9}$$

EXERCISE 5.51

(a) Note that

$$\begin{aligned} f_1(y_1) &= \int_0^\infty e^{-(y_1+y_2)} dy_2 \\ &= e^{-y_1} \qquad y_1 > 0 \end{aligned}$$

and

$$\begin{aligned} f_2(y_2) &= \int_0^\infty e^{-(y_1+y_2)} dy_1 \\ &= e^{-y_2} \qquad y_2 > 0 \end{aligned}$$

It is easy to see that $f(y_1, y_2) = f_1(y_1) \cdot f_2(y_2)$ indicating that Y_1 and Y_2 are independent.

(b) Yes, the conditional probabilities are the same as the marginal probabilities.

EXERCISE 5.53

The ranges of y_1 and y_2 depend on each other, so Y_1 and Y_2 cannot be independent. (Easy to remember!)

5.5 Expected value of a function of random variables

Concentrate on the bivariate case as illustrated in the three examples.I am sure that you realize by now that if you do not have the necessary integration skills, you cannot do the statistical computation and interpretation.

5.6 Special theorems

Expanding the three results in Theorem 4.5 will give you the result in Theorems 5.6 to 5.8. Thinking in terms of independence as defined in section 5.4, brings us to Theorem 5.9.

EXERCISE 5.75

This is now where you must start using the knowledge of the different distributions and recognize density and probability functions.

(a) In Exercise 5.25 you calculated the marginal distributions for this given density function, namely

$$f_1(y_1) = \int_0^\infty e^{-(y_1+y_2)} dy_2 \quad = e^{-y_1} \qquad y_1 > 0$$

and
$$f_2(y_2) = \int_0^\infty e^{-(y_1+y_2)} dy_1 = e^{-y_2} \qquad y_2 > 0.$$

These are exponential functions with $\beta = 1$. Do you remember this?

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-\frac{y}{\beta}} & 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$$ which becomes $$f(y) = \begin{cases} e^{-y} & 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$$ if $\beta = 1$.

That brings us to the knowledge of the exponential function and its parameters: In theorem 4.10 you learnt that, for the exponential distribution

$$\mu = E(Y) = \beta \text{ and } \sigma^2 = V(Y) = \beta^2.$$

If $\beta = 1$ this becomes

$$\mu = E(Y) = 1 \text{ and } \sigma^2 = V(Y) = 1^2 = 1.$$

Continue with pre-knowledge about this specific density function and recall that we deduced in Exercise 5.51 that Y_1 and Y_2 are independent.

So, now when we get to the actual question in Exercise 5.75, we apply this knowledge as follows:

$$\begin{aligned} E(Y_1+Y_2) &= E(Y_1)+E(Y_2) \\ &= 1+1 \\ &= 2 \\ V(Y_1+Y_2) &= V(Y_1)+V(Y_2) \\ &= 1+1 \\ &= 2 \end{aligned}$$

(b)

$$\begin{aligned} P(Y_1-Y_2>3) &= P(Y_1>3+Y_2) \\ &= \int_0^\infty \int_{3+y_2}^\infty e^{-(y_1+y_2)} dy_1 dy_2 \\ &= \int_0^\infty \left[\int_{3+y_2}^\infty e^{-(y_1+y_2)} dy_1 \right] dy_2 \\ &= \int_0^\infty e^{-y_2} \left[\int_{3+y_2}^\infty e^{-y_1} dy_1 \right] dy_2 \\ &= \int_0^\infty e^{-y_2} \left[-e^{-y_1} \mid_{3+y_2}^\infty \right] dy_2 \\ &= \int_0^\infty e^{-y_2} \left[-e^{-\infty} - -e^{-(3+y_2)} \right] dy_2 \\ &= \int_0^\infty e^{-y_2} \left[0 + e^{-3-y_2} \right] dy_2 \\ &= \int_0^\infty e^{-y_2} e^{-3-y_2} dy_2 \\ &= e^{-3} \int_0^\infty e^{-2y_2} dy_2 \\ &= e^{-3} \left[-\frac{1}{2} e^{-2y_2} \mid_0^\infty \right] \\ &= e^{-3} \left[-\frac{1}{2} \left(e^{-\infty} - e^{-0} \right) \right] \\ &= e^{-3} \left[-\frac{1}{2} (0-1) \right] \\ &= \frac{1}{2} e^{-3} = 0.0249 \end{aligned}$$

(c)

$$\begin{aligned}
P(Y_1 - Y_2 < -3) &= P(Y_2 > 3 + Y_1) \\
&= \int_0^{\infty} \int_{3+y_1}^{\infty} e^{-(y_1+y_2)} dy2dy_1 \\
&= \int_0^{\infty} \left[\int_{3+y_1}^{\infty} e^{-(y_1+y_2)} dy_2 \right] dy_1 \\
&= \int_0^{\infty} e^{-y_1} \left[\int_{3+y_1}^{\infty} e^{-y_2} dy_2 \right] dy_1 \\
&= \int_0^{\infty} e^{-y_1} \left[-e^{-y_1} \mid_{3+y_1}^{\infty} \right] dy_1 \\
&= \int_0^{\infty} e^{-y_1} \left[-e^{-\infty} - -e^{-(3+y_1)} \right] dy_1 \\
&= \int_0^{\infty} e^{-y_1} \left[0 + e^{-3-y_1} \right] dy_1 \\
&= \int_0^{\infty} e^{-y_1} e^{-3-y_1} dy_1 \\
&= e^{-3} \int_0^{\infty} e^{-2y_1} dy_1 \\
&= e^{-3} \left[-\frac{1}{2} e^{-2y_1} \mid_0^{\infty} \right] \\
&= e^{-3} \left[-\frac{1}{2} \left(e^{-\infty} - e^{-0} \right) \right] \\
&= e^{-3} \left[-\frac{1}{2} (0 - 1) \right] \\
&= \frac{1}{2} e^{-3} = 0.0249
\end{aligned}$$

(d)

$$\begin{aligned}
E(Y_1 - Y_2) &= E(Y_1) - E(Y_2) \\
&= 1 - 1 \\
&= 0 \\
V(Y_1 - Y_2) &= V(Y_1) + V(-Y_2) \\
&= V(Y_1) + V(Y_2) \\
&= 1 + 1 \\
&= 2
\end{aligned}$$

(e) They are the same; both are equal to 2.

EXERCISE 5.77

(a) You have determined the marginal densities in Exercise 5.27 and can use those answers to continue from:

$$\begin{aligned} f_1(y_1) &= 3\left(y_1-1\right)^2 \qquad 0 \le y_1 \le 1 \\ f_2(y_2) &= 6y_2\left(1-y_2\right) \qquad 0 \le y_2 \le 1 \\ E\left(Y_1\right) &= \int_0^1 y_1 \cdot 3\left(y_1-1\right)^2 dy_1 \\ &= 3\int_0^1 y_1\left(y_1-1\right)^2 dy_1 \\ &= 3\int_0^1 y_1\left(y_1^2-2y_1+1\right) dy_1 \\ &= 3\int_0^1 \left(y_1^3-2y_1^2+y_1\right) dy_1 \\ &= 3\left[\left(\frac{y_1^4}{4}-\frac{2y_1^3}{3}+\frac{y_1^2}{2}\right)\Bigg|_0^1\right] \\ &= 3\left[\left(\frac{1}{4}-\frac{2}{3}+\frac{1}{2}\right)\right] \\ &= 3\left(\frac{3-8+6}{12}\right) \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned}
E(Y_2) &= \int_0^1 y_2 \cdot 6y_2(1-y_2)\,dy_2 \\
&= 6\int_0^1 y_2^2(1-y_2)\,dy_2 \\
&= 6\int_0^1 \left(y_2^2 - y_2^3\right)dy_2 \\
&= 6\left[\left(\frac{y_2^3}{3} - \frac{y_2^4}{4}\right)\Big|_0^1\right] \\
&= 6\left[\left(\frac{1}{3} - \frac{1}{4}\right)\right] \\
&= 6\left(\frac{1}{12}\right) \\
&= \frac{1}{2}
\end{aligned}$$

(b)

$$\begin{aligned}
E\left(Y_1^2\right) &= \int_0^1 y_1^2 \cdot 3(y_1-1)^2\,dy_1 \\
&= 3\int_0^1 y_1^2(y_1-1)^2\,dy_1 \\
&= 3\int_0^1 y_1^2\left(y_1^2 - 2y_1 + 1\right)dy_1 \\
&= 3\int_0^1 \left(y_1^4 - 2y_1^4 + y_1^2\right)dy_1 \\
&= 3\left[\left(\frac{y_1^5}{5} - \frac{2y_1^4}{4} + \frac{y_1^3}{3}\right)\Big|_0^1\right] \\
&= 3\left[\left(\frac{1}{5} - \frac{1}{2} + \frac{1}{3}\right)\right] \\
&= 3\left(\frac{6-15+10}{30}\right) \\
&= \frac{1}{10} \\
V(Y_1) &= E\left(Y_1^2\right) - \left[E(Y_1)\right]^2 \\
&= \frac{1}{10} - \left(\frac{1}{4}\right)^2 \\
&= \frac{3}{80}
\end{aligned}$$

$$
\begin{aligned}
E(Y_2) &= \int_0^1 y_2^2 \cdot 6y_2(1-y_2)\,dy_2 \\
&= 6\int_0^1 y_2^3(1-y_2)\,dy_2 \\
&= 6\int_0^1 \left(y_2^3 - y_2^4\right)dy_2 \\
&= 6\left[\left(\frac{y_2^4}{4}-\frac{y_2^5}{5}\right)\Big|_0^1\right] \\
&= 6\left[\left(\frac{1}{4}-\frac{1}{5}\right)\right] \\
&= 6\left(\frac{1}{20}\right) \\
&= \frac{3}{10} \\
V(Y_2) &= E\left(Y_2^2\right) - \left[E(Y_2)\right]^2 \\
&= \frac{3}{10} - \left(\frac{1}{2}\right)^2 \\
&= \frac{1}{20}
\end{aligned}
$$

(c)

$$
\begin{aligned}
E(Y_1 - 3Y_2) &= E(Y_1) - 3E(Y_2) \\
&= \frac{1}{4} - 3\left(\frac{1}{2}\right) \\
&= -\frac{5}{4}
\end{aligned}
$$

EXERCISE 5.79

Refer back to question 5.11, as the same information is given.
In the process of finding the expected value of Y_1Y_2,please note that it is necessary to break the given shaded triangle up into two parts: the one part lies to the left of the y_2-axis and the other half lies to the right of the y_2-axis.The reason for this split is that the limits of integration for these two halves are different. In the section to the left of the y_2-axis the border of the triangle is formed by a straight line with equation $y_2 = y_1 + 1$ (a line with gradient *positive one*, cutting the y_2-axis at $y_2 = 1$), while the right-hand side of the triangle has (as border) the line $y_2 = -y_2 + 1$ (a line with gradient negative one, cutting the y_2-axis at $y_2 = 1$).

$$
\begin{aligned}
E(Y_1Y_2) &= \int_{-1}^{0}\int_{0}^{1+y_1} y_1y_2dy_2dy_1 + \int_{0}^{1}\int_{0}^{1-y_1} y_1y_2dy_2dy_1 \\
&= \int_{-1}^{0}\left[\int_{0}^{1+y_1} y_1y_2dy_2\right]dy_1 + \int_{0}^{1}\left[\int_{0}^{1-y_1} y_1y_2dy_2\right]dy_1 \\
&= \int_{-1}^{0} y_1\left[\int_{0}^{1+y_1} y_2dy_2\right]dy_1 + \int_{0}^{1} y_1\left[\int_{0}^{1-y_1} y_2dy_2\right]dy_1 \\
&= \int_{-1}^{0} y_1\left[\left.\frac{y_2^2}{2}\right|_0^{1+y_1}\right]dy_1 + \int_{0}^{1} y_1\left[\left.\frac{y_2^2}{2}\right|_0^{1-y_1}\right]dy_1 \\
&= \int_{-1}^{0} y_1\left[\frac{(1+y_1)^2}{2}\right]dy_1 + \int_{0}^{1} y_1\left[\frac{(1-y_1)^2}{2}\right]dy_1 \\
&= \frac{1}{2}\int_{-1}^{0}\left(y_1+2y_1^2+y_1^3\right)dy_1 + \frac{1}{2}\int_{0}^{1}\left(y_1-2y_1^2+y_1^3\right)dy_1 \\
&= \frac{1}{2}\left(\left.\frac{y_1^2}{2}+\frac{2}{3}y_1^3+\frac{y_1^4}{4}\right|_{-1}^{0}\right) + \frac{1}{2}\left(\left.\frac{y_1^2}{2}+\frac{2}{3}y_1^3+\frac{y_1^4}{4}\right|_{0}^{1}\right) \\
&= \frac{1}{2}\left[0-\left(\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right)\right] + \frac{1}{2}\left[\left(\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right)-0\right] \\
&= \frac{1}{2}\left[-\frac{1}{2}+\frac{2}{3}-\frac{1}{4}\right] + \frac{1}{2}\left[\left(\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right)\right] \\
&= 0
\end{aligned}
$$

5.7 The covariance of two random variables

Now we get back to something you should be familiar with, namely *covariance* and *correlation coefficients*. They measure the dependence between two variables, so if there is no dependence, the values of these measures should be zero. Read the very clear explanation of the significance of these concepts and make sure you understand how this can be illustrated in a figure.

A reminder: For two random variables Y_1 and Y_2:

Name	Notation	Expression	Comments
	$(Y_1-\mu_1)(Y_2-\mu_2)$		Measure of linear dependence
Covariance of Y_1 and Y_2	$Cov(Y_1,Y_2) \;=$	$E[(Y_1-\mu_1)(Y_2-\mu_2)]$	$Cov(Y_1,Y_2)=E(Y_1Y_2)-\mu_1\mu_2$ $Cov(Y_1,Y_2)=0$ if independent
Correlation coefficient	$\rho \;=$	$\dfrac{Cov(Y_1,Y_2)}{\sigma_1\sigma_2}$	$-1\leq\rho\leq 1$

EXERCISE 5.95

Refer back to Exercises 5.11 and 5.79, as this question is a follow-up on those two.

(a)

$$\begin{aligned} E(Y_1) &= \int_{-1}^{1} y_1 dy_1 \\ &= \left.\frac{y_1^2}{2}\right|_{-1}^{1} \\ &= \frac{1^2}{2} - \frac{(-1)^2}{2} \\ &= \frac{1}{2} - \frac{1}{2} \\ &= 0 \end{aligned}$$

We have proved in Exercise 5.79 that $E(Y_1Y_2) = 0$, so we can now say that

$$\begin{aligned} Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\ &= 0 - 0 \cdot E(Y_2) \\ &= 0 \end{aligned}$$

(b) If the variables have to be independent, it should be true that (e.g.)

$$P\left(Y_1 \leq \frac{1}{4}; Y_2 \leq \frac{1}{4}\right) \text{ must be equal to } P\left(Y_1 \leq \frac{1}{4}\right) \cdot P\left(Y_2 \leq \frac{1}{4}\right).$$

Just looking at the sketch you can see that this is not true, so the two variables are dependent.

(c) If $Cov(Y_1, Y_2) = 0,$ then $\rho = \dfrac{Cov(Y_1, Y_2)}{\sigma_1\sigma_2} = 0.$

(d) Do you remember what is meant by "if and only if"? It implies that a statement is "reversable". But *independence of variables and covariance* is not an "if and only if" statement!
Although independence of variables implies that the covariance will be zero, the reverse statement is not true.
If $Cov(Y_1, Y_2) = 0,$ it does not imply that the variables have to be independent.

Do you understand?
Suppose that only male players are allowed to play soccer, then we can say:
Being a soccer player implies ($\Longrightarrow$) the person is male.
This statement cannot be reversed, because

being a male does not necessarily imply that the person is a soccer player.

In short we can say:

soccer player	$\Longrightarrow$	male
male	$\not\Rightarrow$	a soccer player

The rest of this chapter does not form part of the contents of this module.

CHAPTER 6

Functions of random variables

STUDY
WM&S: **Chapter 6** (up to 6.5).

6.1 Introduction

Read through this to find your way through what you have studied so far as well as the road ahead. Take note of the assumption explained in the last paragraph, which states that for the rest of the work discussed in the book you have to assume that the populations are much larger than the samples taken from them. The result of this assumption is that the random variables can be considered as independent of one another.

For discrete random variables independence implies

$$p(y_1, y_2, y_3, ...y_n,) = p(y_1) \cdot p(y_2) \cdot p(y_3) \cdot p(y_4).... \cdot p(y_n)$$

For continuous random variables independence implies

$$f(y_1, y_2, y_3, ...y_n,) = f(y_1) \cdot f(y_2) \cdot f(y_3) \cdot f(y_4)... \cdot f(y_n)$$

6.2 Finding the probability distribution of a function of random variables

In this section you are introduced to three methods for finding the joint probability distribution of different functions of random variables. The fourth method using Jacobians will not be discussed in this module as it is covered at third level in Distribution Theory III. Read this paragraph 6.2, but accept that you will not really understand it until *after* you have studied the three methods in detail as given in sections 6.3 to 6.5. During the revision process, when you come back to this paragraph, you will really understand the short summaries.

When you get to the module $STA2603$ (Distribution Theory II) and later $STA3703$ (Distribution Theory III) these functions of random variables (also called *composite* functions) will reappear, but at a higher level of complexity. Make sure that you master this introduction to the very important topic of composite functions and transformations. The examples and exercises in sections 6.3 to 6.5 of WM&S include single variable densities as well as joint density functions. In this module you will concentrate on the single variable density function and different composite forms of that variable. The same topics, but for joint density functions of two random variables, will be covered in $STA2603$, and in $STA3703$ you will learn about multivariate transformations using Jacobians. So, make sure that you achieve the necessary basic knowledge of these different methods.

6.3 The method of distribution functions

Do you remember the definition of a distribution function?

$$F_Y(y) = P(Y \leq y) \text{ for } -\infty < y < \infty$$

We only consider continuous random variables.

Example 6.1 is typical of the questions you can expect in an examination or assignment. You must also study Example 6.2 in depth, even though it uses a joint density function because the joint function is very elementary and contains only one of the two variables. Go through these two examples in detail while you are reading my notes to guide you through the different steps. I have also given you a full solution of Exercise 6.1 to ensure that you really understand the method of distribution functions.

Example 6.1

In this first example only *one* random variable Y, with its density function, is given. Finding the density function for the new variable, $U = 3Y - 1$, is therefore not difficult to follow, but you should understand that you move from the *density* function of the variable Y to the *density* function of the variable U via the *distribution* function of $\mathbf{U}$ (in this example using the relationship $U = 3Y - 1$).

$$F_U(u) = P(U \leq u)$$

Now let $U = 3Y - 1$

therefore $F_U(u) = P(3Y - 1 \leq u)$

Make Y the subject of the inequality :

$$F_U(u) = P(Y \leq \frac{u+1}{3})$$

As we continue with examples you will realize how important the definition region of the composite variable is, and finding the new limits is not always easy. For this reason you have to read and re-read the paragraph in Example 6.1 where determining the limits for U is discussed. Once you are sure about that, continue with the distribution function:

$$F_Y(y) = \int_{-\infty}^{y} f(t)dt$$

$$F_U(u) = \int_{-\infty}^{\frac{u+1}{3}} f(y)dy$$

You can follow the rest in the book. Note the presentation of the distribution function over the different regions. Hopefully you remember all the properties of a distribution function. If you have forgotten, go back to 4.2 and refresh your memory!

Example 6.2

Here you are given a *joint density function* of the two variables Y_1 and Y_2 and you have to find the density function of the new variable U with $U = Y_1 + Y_2$. Now you will understand why I stressed the importance of the definition region of the new variable. Figure 6.1 shows you exactly where the joint function is positive. Make sure you understand, and whenever you get a question on this section do not hesitate to make such a sketch! You are supposed to have the skill to make these sketches based on your knowledge of graphs in the Cartesian plane at school level.

It would not be impossible to find the area of the dark shaded region, but because of its positioning you would have to break it up into two sections. So, to simplify matters, the distribution function of U is rewritten in terms of the complement $[1 - P(U \leq u)]$. The lighter shaded area is then integrated and the answer subtracted from one.

If you have difficulty in understanding the limits of integration, remember what we said earlier and look at this:

$$\int_{u}^{1} \left(\int_{0}^{y_1 - u} 3y_1y_2\mathbf{dy_2} \right) dy_1$$

The inner integral is integrated to y_2 (see the $\mathbf{dy_2}$), and the given the limits in the integral, namely 0 to $(y_1 - u)$, are therefore the limits of y_2. This implies that we integrate from 0 to $(y_1 - u)$. To see this, look at the lightly shaded triangle and move vertically inside it (up from the y_1-axis). Your movement is from $y_2 = 0$ (the equation of the y_1-axis) upwards and in every vertical move you are blocked by the line with equation $y_1 - y_2 = u$ (rewritten with y_2 as subject it is $y_2 = y_1 - u$). Do you see now that y_2 moves from 0 to $(y_1 - u)$?

Now move your attention to $\mathbf{dy_1}$ which indicates the horizontal movement from u to 1 (according to the graph). Why from u? The intercept of the line $y_2 = y_1 - u$ on the y_1-axis is found by letting $y_2 = 0$:

$$\begin{aligned} y_2 &= y_1 - u \\ 0 &= y_1 - u \\ y_1 &= u \end{aligned}$$

The integration as well as the listing of the distribution function and the density function is given in WM&S. Do not look at it and think that you understand it – do it yourself! The illustration of the two functions in Figure 6.2 is also very interesting (see how the distribution function satisfies the properties) and the derivation of the expected value is a good example of testing your pre-knowledge.

EXERCISE 6.1

(a) $U_1 = 2Y - 1$

Determine the limits for u :

$$\begin{aligned} 0 &\leq y \leq 1 \Longrightarrow 0 \leq \frac{u+1}{2} \leq 1 \\ &\Longrightarrow 0 \leq u+1 \leq 2 \\ &\Longrightarrow -1 \leq u \leq 1 \end{aligned}$$

The region in which we work is determined by the inequality $2y - 1 \leq u$ or $y \leq \frac{u+1}{2}$. Because of the $\leq$-sign, we shade (in our minds) *below* the graph of $y = \frac{u+1}{2}$ (see the limits of the integral).

The distribution function of $U_1 = 2Y - 1$ is

$$\begin{aligned} F_{U_1}(u) &= P(U_1 \leq u) \\ &= P(2Y - 1 \leq u) \\ &= P\left(Y \leq \frac{u+1}{2}\right) \\ &= \int_0^{\frac{u+1}{2}} 2(1-y)\,dy \\ &= \left(2y - y^2\right)\Big|_0^{\frac{u+1}{2}} \\ &= 2(\frac{u+1}{2}) - (\frac{u+1}{2})^2 \\ &= \frac{2(2u+2) - \left(u^2 + 2u + 1\right)}{4} \\ &= \frac{-u^2 + 2u + 3}{4} \end{aligned}$$

The distribution function is formally defined as

$$F_{U_1}(u) = \begin{cases} 0 & u \leq -1 \\ \dfrac{-u^2 + 2u + 3}{4} & -1 < u < 1 \\ 1 & u > 1 \end{cases}$$

$$\begin{aligned} f_{U_1}(u) &= \frac{d}{du} F_{U_1}(u) \\ &= \frac{-2u+2}{4} \\ &= \frac{1-u}{2} \end{aligned}$$

$$f_{U_1}(u) = \begin{cases} \dfrac{1-u}{2} & -1 < u < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(b) $U_2 = 1 - 2Y$

Determine the limits for u :

$$\begin{aligned} 0 &\leq y \leq 1 \Longrightarrow 0 \leq \frac{1-u}{2} \leq 1 \\ &\Longrightarrow 0 \leq 1-u \leq 2 \\ &\Longrightarrow -1 \leq u \leq 1 \end{aligned}$$

The region in which we work is determined by the inequality $1 - 2y \leq u$ or $y \geq \frac{1-u}{2}$. Because of the $\geq$-sign, we shade (in our minds) *above* the graph of $y = \frac{1-u}{2}$ (see the limits of the integral).

The distribution function of $U_2 = 1 - 2Y$ is

$$\begin{aligned} F_{U_2}(u) &= P(U_2 \leq u) \\ &= P(1 - 2Y \leq u) \\ &= P\left(Y \leq \frac{1-u}{2}\right) \\ &= \int_{\frac{1-u}{2}}^{1} 2(1-y)\,dy \\ &= \left(2y - y^2\right)\Big|_{\frac{1-u}{2}}^{1} \\ &= \left[2(1) - 1^2\right] - \left[2(\frac{1-u}{2}) - (\frac{1-u}{2})^2\right] \\ & 1 - (1-u) + (\frac{1-u}{2})^2 \\ &= \frac{4u + \left(u^2 + 2u + 1\right)}{4} \\ &= \left(\frac{1+u}{2}\right)^2 \end{aligned}$$

The distribution function is formally defined as

$$F_{U_2}(u) = \begin{cases} 0 & u \leq -1 \\ \left(\dfrac{1+u}{2}\right)^2 & -1 < u < 1 \\ 1 & u > 1 \end{cases}$$

$$\begin{aligned} f_{U_2}(u) &= \frac{d}{du} F_{U_2}(u) \\ &= \left(\frac{1+u}{2}\right)^2 \\ &= 2\left(\frac{1+u}{2}\right) \cdot \frac{1}{2} \\ &= \frac{1+u}{2} \end{aligned}$$

$$f_{U_2}(u) = \begin{cases} \dfrac{1+u}{2} & -1 < u < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(c) $U_3 = Y^2$

Determine the limits for u :

$$\begin{aligned} 0 &\leq y \leq 1 \Longrightarrow 0 \leq \sqrt{u} \leq 1 \\ &\Longrightarrow 0^2 \leq u \leq 1^2 \\ &\Longrightarrow 0 \leq u \leq 1 \end{aligned}$$

The region in which we work is determined by the inequality $y^2 \leq u$ or $y \leq \sqrt{u}$. Because of the $\leq$-sign, we shade (in our minds) *below* the graph of $y = \frac{u+1}{2}$ (see the limits of the integral).

The distribution function of $U_3 = Y^2$ is

$$\begin{aligned} F_{U_3}(u) &= P(U_3 \leq u) \\ &= P\left(Y^2 \leq u\right) \\ &= P\left(Y \leq \sqrt{u}\right) \\ &= \int_0^{\sqrt{u}} 2(1-y)\,dy \\ &= \left(2y - y^2\right)\Big|_0^{\sqrt{u}} \\ &= \left[2(\sqrt{u}) - \left(\sqrt{u}\right)^2\right] \\ &= 2\sqrt{u} - u \end{aligned}$$

The distribution function is formally defined as

$$F_{U_3}(u) = \begin{cases} 0 & u \leq 0 \\ 2\sqrt{u} - u & 0 < u < 1 \\ 1 & u > 1 \end{cases}$$

$$\begin{aligned} f_{U_3}(u) &= \frac{d}{du} F_{U_3}(u) \\ &= \frac{d}{du}\left(2\sqrt{u} - u\right) \\ &= \frac{1}{\sqrt{u}} - 1 \text{ or } \frac{1-\sqrt{u}}{\sqrt{u}} \end{aligned}$$

$$f_{U_3}(u) = \begin{cases} \frac{1-\sqrt{u}}{\sqrt{u}} & 0 < u < 1 \\ 0 & \text{elsewhere} \end{cases}$$

6.4 The method of transformations

If the composite function U is either increasing or decreasing, this method of transformations can be applied. In this section you are also told that the function $h(y)$ need not be increasing or decreasing for *all* values of y, but only for the values of y such that $f_Y(y)$ is positive. The method is described in detail; an easy-to-apply formula is deduced and you should study examples $6.6,\ 6.7$ and 6.8.

$$f_U(u) = f_Y\left[h^{-1}(u)\right]\frac{dh^{-1}(u)}{du}$$

Shorter notation :

$$f_U(u) = f_Y\left[h^{-1}(u)\right]\left|\frac{dh^{-1}}{du}\right|$$

Once you understand this method, it always seems much easier than the method of distribution functions. The result is that this method can be treated as "cake recipe", but always remember how it was deduced.

The steps are as folows:

- Solve for y in terms of u, i.e. find $h^{-1}(u)$.
- Substitute this answer into $f_Y(y)$.
- Differentiate $h^{-1}(u)$ to u, i.e. find $\frac{dh^{-1}(u)}{du}$.
- Find the absolute value of $\frac{dh^{-1}(u)}{du}$.
- Substitute the answers into the formula $f_U(u) = f_Y\left[h^{-1}(u)\right]\left|\frac{dh^{-1}}{du}\right|$.

Please try the examples yourself before looking at the solutions. Also keep in mind that you can be forced in any question on the topic *functions of random variables* to use a specified method (distribution function method or method of transformations) and if you apply any other method, you will not get the marks for that question!

EXERCISE 6.23

The density function as well as the functions of random variables in this exercise are identical to those in Exercise 6.1, but now we are going to use the transformation method.

(a) Note that you find the region for the new variable U in the same way as in the distribution function method.

For the information in this question we have determined that $-1 \leq u \leq 1$.

$$\begin{aligned}
&\text{If} \quad U = 2Y - 1 \quad \text{then} \quad & Y &= \tfrac{U+1}{2} \\
&\tfrac{dy}{du} = \tfrac{1}{2} \quad & \text{and } f_U(u) &= f_Y\left[h^{-1}(u)\right]\left|\frac{dh^{-1}}{du}\right| \\
& & &= 2\left(1 - \tfrac{u+1}{2}\right) \cdot \left|\tfrac{1}{2}\right| \\
& & &= \left(1 - \tfrac{u+1}{2}\right) \\
& & &= \tfrac{1-u}{2}
\end{aligned}$$

(b) For the information in this question we have determined that $-1 \leq u \leq 1$.

$$\begin{aligned}
&\text{If} \quad U = 1 - 2Y \quad \text{then} \quad & Y &= \tfrac{1-U}{2} \\
&\tfrac{dy}{du} = -\tfrac{1}{2} \quad & \text{and } f_U(u) &= f_Y\left[h^{-1}(u)\right]\left|\frac{dh^{-1}}{du}\right| \\
& & &= 2\left(1 - \tfrac{1-u}{2}\right) \cdot \left|-\tfrac{1}{2}\right| \\
& & &= \left(1 - \tfrac{1-u}{2}\right) \\
& & &= \tfrac{1+u}{2}
\end{aligned}$$

(c) For the information in this question we have determined that $0 \leq u \leq 1$

$$\begin{aligned}
&\text{If} \quad U = Y^2 \quad \text{then} \quad & Y &= \sqrt{U} \\
&\tfrac{dy}{du} = -\tfrac{1}{2\sqrt{u}} \quad & \text{and } f_U(u) &= f_Y\left[h^{-1}(u)\right]\left|\frac{dh^{-1}}{du}\right| \\
& & &= 2\left(1 - \sqrt{u}\right) \cdot \left|\tfrac{1}{2\sqrt{u}}\right| \\
& & &= \left(\tfrac{1-\sqrt{u}}{\sqrt{u}}\right)
\end{aligned}$$

6.5 The method of moment-generating functions

You may not quite appreciate this method at this stage, as you have not yet "played around" enough with different distributions and may not be able to recognize a moment-generating function as belonging to a specific distribution. We expect you to understand the principles involved in this method, but will not yet expect you to recognize moment-generating functions.

It is very important that you know the statement given in Theorem 6.1. It states that a moment-generating function uniquely defines a probability distribution of a random variable. Although you do not have to prove this theorem, you have to know that unique characteristic of a moment-generating function.

Go through Examples 6.10 and 6.11. The result of 6.10 seems quite familiar because you have used it in previous modules. Do you recognize it? In past modules you often needed to determine a probability for a random variable Y, having any general normal distribution. Without knowing why, you *standardized* the variable Y, using its mean μ and standard deviation σ to find a new variable $Z = \frac{Y-\mu}{\sigma}$, then having a standard normal distribution. That allowed you to use the standard normal or $N(0,1)$ table. In this example you can now see why that was the right thing to do!

Example 6.11 is a little more complicated, but the result is well known to statisticians. The deduction shows how you can prove, using a moment-generating function, that the square of a random variable

Z which is normally distributed with mean 0 and variance 1, namely Z^2, will have a chi-square distribution with one degree of freedom. The reasoning in this example is also based on recognition of the moment-generating function of the new variable $U = Y^2$. This result is also widely used and you are definitely going to hear more about it in other distribution theory modules!

The rest of this chapter does not form part of the contents of this module.