The University Of Sheffield.

Graduate Certificate in Statistics

MAS5051: Probability Distributions
(semester one)

Dr Mark Yarrow

2020/21

This course is comprised of forty classe across two semesters consisting of topics outlined in this booklet, six assignments of which five count towards 20% of your grade and a single summer exam. Your exam will be two hours in length, will be marked out of 80 and will be restricted open book. Below details the assignment schedule for this year.

	8.00am release date	Due in at noon
Assignment 0	Wednesday 21 Oct 2019	Wednesday 4 Nov 2019
Assignment 1	Wednesday 18 Nov 2019	Wednesday 2 Dec 2019
Assignment 2	Wednesday 16 Dec 2019	Wednesday 20 Jan 2020
Assignment 3	Wednesday 3 Feb 2020	Wednesday 17 Feb 2020
Assignment 4	Wednesday 3 Mar 2020	Wednesday 17 Mar 2020
Assignment 5	Wednesday 21 Apr 2020	Wednesday 5 May 2020

There are exercises in this booklet after each section which you should complete following the completion of each section. The class sessions contain numerous examples which will be worked through in detail to create a complete course record.

Further support for this module is provided via the MOLE discussion boards. I strongly encourage you to use these discussion boards for any question which you feel others could benefit from knowing the answer to.

You may want to watch the following videos as preliminary material:

1. Functions, (see in particular videos 5 and 6 on binomial expansions and 9 and 10 on exponentials and logarithms.)
2. Differentiation, (also includes some material on series: see video 13.0 for arithmetic and geometric series.)
3. Complex numbers.
4. Vectors.
5. Integration, (including topics such as partial fractions and completing the square.)
6. Matrices, (including sections on matrix multiplication, determinants, inverses, systems of linear equations and eigenvalues and eigenvectors.)

Module leader email address: Mark.Yarrow@Sheffield.ac.uk

Please see below the key corrosponding the the Graduate Certificate timetable. The letter key in the left hand column and the class topics in the right hand column.

Block	**Date**	**Class Topic**
A	14-Oct	1: Sample spaces and events
		2: Defining probabilities
B	21-Oct	3: Conditional probability
		4: Independence
C	28-Oct	5: Introduction to random variables
		6: Discrete random variables
D	4-Nov	7: Expectation (mean)
		8: Variance and standard deviation
E	11-Nov	9: Independence of random variables
		10: Mean and variance of linear combinations
F	18-Nov	11: Bernoulli trials and the Binomial distribution
		12: The Geometric distribution
G	25-Nov	13: Continuous random variables
		14: Mean and variance for continuous random variables
H	2-Dec	15: Moment generating functions
		16: The normal distribution
I	9-Dec	17: Distributions in R
		18: The Poisson distribution
J	16-Dec	19: The Gamma and exponential distributions
		20: The Beta distribution
K	13-Jan	21: Transformations of random variables I
		22: Transformations of random variables II
L	20-Jan	23: χ^2, t and F distributions
		24: Normal approximations to Binomial and Poisson
M	27-Jan	25: The Weak Law of Large Numbers
		26: The Central Limit Theorem
N	3-Feb	27: Conditional probability (continued)
		28: Bayes' Theorem
O	10-Feb	29: Multivariate random variables
		30: The Multinomial distribution
P	17-Feb	31: Multivariate continuous r.v's and conditional distributions
		32: Covariance and correlation
Q	24-Feb	33: Conditional expectation
		34: Multivariate normal distribution
R	3-Mar	35: Transformations of multivariate random variables
		36: Examples and applications of multivariate transformations
S	10-Mar	37: Introduction to Markov Chains I
		38: Introduction to Markov Chains II
T	17-Mar	39: Basic Bayesian inference I
		40: Basic Bayesian inference II

Contents

1 Sample Spaces and Events

Content:

1. Sample spaces.
2. Events.
3. Relationships between events.
4. Set notation for events.

Supplementary reading: Freund §2.1 to 2.3. Note: Freund uses the notation A' for complement. You will also see A^c and $\bar{A}$ used for this.
Supplementary videos: Sample spaces and events.

1.1 Class Content

1.1.1 Motivation

We regularly have to make decisions in day to day life based upon the potential outcomes of that decision. We do this by weighing up the probabilities of the different outcomes.

- A doctor is treating a patient, and is considering whether to prescribe a particular drug. How likely is it that the drug will cure the patient?
- A probation board is considering whether to release a prisoner early on parole. How likely is the prisoner to re-offend?
- A bank is considering whether to approve a loan. How likely is it that the loan will be repaid?
- A government is considering a CO_2 emissions target. What would the effect of a 20% cut in emissions be on global mean temperatures in 20 years' time?

We often use verbal expressions of uncertainty such as "possible", "quite unlikely", "very likely", but these can be inadequate if we want to communicate with each other about uncertainty. Consider the following example.

You are being screened for a particular disease. You are told that the disease is "quite rare", and that the screening test is accurate but not perfect; if you have the disease, the test will "almost certainly" detect it, but if you don't have the disease, there is "a small chance" the test will mistakenly report that you have it anyway. The test result is positive. How certain are you that you really have the disease?

If you have studied probability at GCSE or A-level, you may have seen a definition of probability like this:

> Suppose all the outcomes in an experiment are equally likely. The probability of an event A is defined to be
> $$\mathbb{P}(A) = \frac{\text{number of outcomes in which } A \text{ occurs}}{\text{total number of possible outcomes}}. \tag{1}$$

In this class we will set out some necessary notation and terminology and then try and extend our definition of probability.

Example 1.1. *A deck of 52 playing cards is shuffled thoroughly. What is the probability that the top card is an ace?*

Solution. *After a shuffle that results in a random deck then there are only four options of the 52 in which an Ace would be on the top; the answer is therefore $\frac{4}{52}$ which is simplified to $\frac{1}{13}$.*

Definition 1.2 (Sample Space)**.** *The set of all possible outcomes of an experiment is called the* **Sample Space** *and it is usually denoted by $\mathbb{S}$. Each element of $\mathbb{S}$ is called an* **Event**.

Some examples of sample spaces are:

1. Coin Flip $\mathbb{S} = \{H, T\}$.
2. Dice Roll $\mathbb{S} = \{1, 2, 3, 4, 5, 6\}$.
3. Amount of Rainfall $\mathbb{S} = \{k | k \geq 0, k \in \mathbb{R}\}$.
4. Number of cats and dogs in a shelter $\mathbb{S} = \{(c, d) | (c, d) \in \mathbb{N}^2\}$.

Definition 1.3 (Subset)**.** *A subset C of a set S denoted by $A \subset S$ (alternate notation is $A \subseteq S$) is defined as such if all elements of C are contained within S.*

Definition 1.4 (Event Space)**.** *The event space is a subset of the sample space*

Example 1.5. *For the experiment of rolling two dice simultaniously, write down the sample space and event space for the event "sum of the faces on the two dice is seven".*

Solution.

$$\mathbb{S} = \{(x, y) | x = 1, 2, ..., 6, y = 1, 2, ...6\} \tag{2}$$

and

$$E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \tag{3}$$

or

$$E = \{(z, 7 - z) | z = 1, 2, ..., 6\}. \tag{4}$$

Definition 1.6 (Partition)**.** *A* **partition** *$\{E_1, E_2, ...\}$ of a set $\mathbb{S}$ is a set of subsets of $\mathbb{S}$ such that every element of $\mathbb{S}$ appears in exactly one of the subsets $E_1, E_2, ...$.*

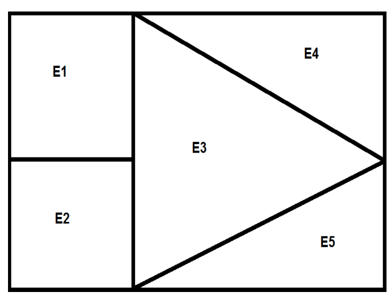

Figure 1: A diagram depicting a partition of the set $\mathbb{S}$.

Example 1.7. *If $\{E_1, E_2, E_3\}$ is a partition of $\mathbb{S}$ such that:*

$$E_1 = \{A, C, V, U\}$$
$$E_2 = \{W, T, H, F, I\}$$
$$E_3 = \{Q, P, L\}$$

explicitly state $\mathbb{S}$.

Solution. $\mathbb{S} = \{A, C, V, U, W, T, H, F, I, Q, P, L\}$.

1.1.2 Defining probabilities

It may surprise you to know that there is no single agreed way to define probability! The following interpretations are all widely used (and there are others).

1.1.2.1 Classical probability

In the case of a finite sample space S, in the **classical** approach we make the *assumption* that each outcome in the sample space is equally likely. If there are k outcomes in the sample space, then the probability of any single outcome is $1/k$. The probability of any event is then calculated as the number of outcomes in which the event occurs, divided by the total number of possible outcomes.

$$\mathbb{P}(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \mathbb{S}}.$$

1.1.2.2 Relative frequency

An alternative formulation which can accommodate events which are not equally likely is to imagine our experiment being repeated many times is to define

$$\mathbb{P}(A) = \frac{\text{number of times } A \text{ occurred}}{\text{number of times experiment is repeated}}.$$

assuming that the proportion, or relative frequency, settles down to a fixed limit as the number of repeats tends to infinity.

1.1.2.3 Subjective probability

The relative frequency approach still cannot cope with all common situations - there are many cases where we cannot repeat experiments even in our imagination. An even more flexible approach is to allow subjective probabilities. Here we allow any individual to specify their own probabilities based on their personal preferences and experiences though we usually insist on some conditions to make them sensible and coherent. This form of probability is often used in the Bayesian approach to statistics to represent 'expert opinion'. Given a sample space $\mathbb{S}$ and two subsets or events A and B, the set operations **union**, **intersection**, **complement** and **difference** all define further events.

1.1.3 Venn Diagrams

The **union** $A \cup B$ corresponds to either A occurring or to B occurring (*or* to both occurring). We can visualise this using a *Venn diagram* (see Figure 2). The rectangle represents the sample space $\mathbb{S}$, the two circles represent the subsets A and B, and the shaded area represents the set $A \cup B$.

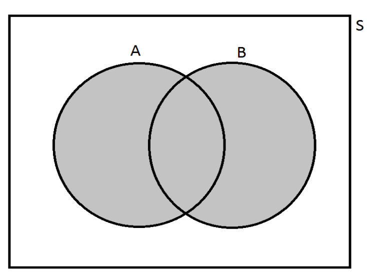

Figure 2: A Venn diagram, with the shaded region showing $A \cup B$ (A, B or both.).

The **intersection** $A \cap B$ corresponds to both A and B occurring.

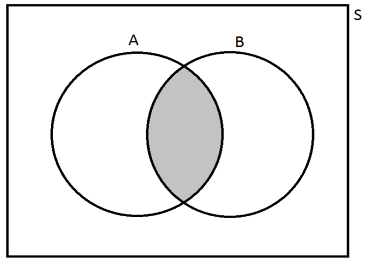

Figure 3: A Venn diagram, with the shaded region showing $A \cap B$ (both A and B).

If there are no elements that are both in A and in B, then $A \cap B = \varnothing$ where $\varnothing$ is the empty set. In this case we say that A and B are **mutually exclusive**.

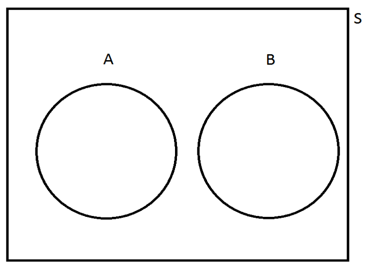

Figure 4: A Venn diagram with mutually exclusive A and B.

Definition 1.8 (Mutual Exclusivity). *Two events A and B are mutually exclusive if there are no common elements in the two event spaces, i.e. $A \cap B = \varnothing$. Here $\varnothing$ is the empty set.*

The **complement** of A, which we will write $\overline{A}$, corresponds to A not occurring. (**Alternative notation**: $\bar{A}$ is also written as A^C or A'.)

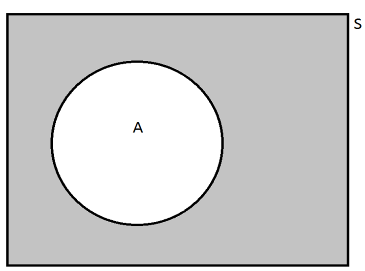

Figure 5: A Venn diagram, with the shaded region showing $\bar{A}$ (all things not in A.)

Note that $\varnothing = \bar{\mathbb{S}}$, and that the event $\varnothing$ is one which cannot happen, as it contains no elements.

The **set difference** $A \setminus B$ corresponds to A occurring but B not.

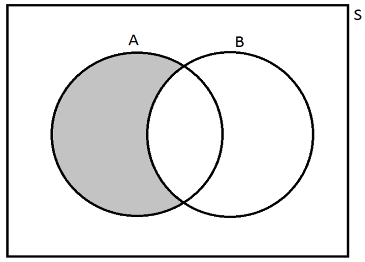

Figure 6: A Venn diagram, with the shaded region showing $A \setminus B$ (all things in A but not in B.)

Note that

$$A \setminus B = A \cap \bar{B}.$$

Example 1.9. *For the sets $A = \{1, 3, 5, 7\}$ and $B = \{-1, 3, 44, 5\}$ write down the following:*

1. $A \cup B$
2. $A \cap B$
3. $A \setminus B$

Solution.

1. $A \cup B = \{1, 3, 5, 7, -1, 44\}$
2. $A \cap B = \{3, 5\}$
3. $A \setminus B = \{1, 7\}$

1.2 Exercises

Exercise 1.10. *In each of the following cases, describe a suitable sample space for the experiment and identify the event indicated as a subset of this sample space. [Do not try to assign probabilities.]*

1. *Experiment: toss a coin three times.*
   Event: the number of heads is even.
2. *Experiment: count the number r of red tomatoes and the number y of yellow tomatoes grown by a gardener.*
   Event: there are more yellow tomatoes than red ones.
3. *Experiment: observe the arrival time of a train, relative to its scheduled arrival time.*
   Event: the train is at least ten minutes late.
4. *Experiment: observe the score in a football match between teams A and B.*
   Event: team A wins the match.
5. *Experiment: measure the quantity of rainfall in a day at each of three weather stations A, B and C.*
   Event: it is wettest at station A.

Exercise 1.11. *Given a sample space $\mathbb{S}$ with $A \subseteq \mathbb{S}$, prove that $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$. Justify each step in your proof carefully.*

Exercise 1.12. *In the Champions League semi-finals, teams are chosen at random to play each other. Two teams out of the four are selected at random for the first semi-final, and the remaining two teams play each other in the second semi-final. Assuming that all random selections are equally likely, if two of the four teams* *are English, what is the probability that they are not drawn to play each other? Show clearly how you have derived your probability.*

Exercise 1.13. *Three dice are rolled, each of different colours to distinguish them from each other. Write down an appropriate sample space for the three dice.*

Exercise 1.14. *The sequence of tosses of a coin until the first heads.*

Exercise 1.15. *Use Venn idagrams to verify the below identities involving three elements $\{A, B, C\} \subset \mathbb{S}$.*

1. $(A \cup B) \cup C = A \cup (B \cup C)$,
2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

2 Defining Probabilities

Content:

1. Axioms of probability.
2. Basic results derived from the axioms.
3. Relationships between events.
4. Examples: equally likely outcomes.

Supplementary reading: Freund §2.4 to 2.5.
Supplementary videos: Defining probabilities.
Related material: Some examples in Ross §2.5 use permutations and combinations, which will be covered later in MAS5050. Freund exercises 2.6, 2.53.

2.1 Class Content

2.1.1 Axioms of probability

Whatever interpretation we put on our probability, we expect it to obey certain rules or **axioms** so that it works mathematically. The core axioms are, For a sample space $\mathbb{S}$, and events A, B such that $A, B \subseteq \mathbb{S}$

(A1) $\mathbb{P}(S) = 1$.

(A2) $0 \leq \mathbb{P}(A) \leq 1$.

(A3) If $A \cap B = \varnothing$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
From these three basic axioms, we can deduce the following

(A4) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

(A5) $\mathbb{P}(\varnothing) = 0$.

(A6) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(A7) If $B \subseteq A$ then $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$ and $\mathbb{P}(B) \leq \mathbb{P}(A)$.

(A8) If $\{E_1, E_2, \ldots, E_n\}$ is a partition of $\mathbb{S}$, then $1 = \mathbb{P}(S) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + ... + \mathbb{P}(E_n)$.

Theorem 2.1. *If A is an event in the discrete sample space $\mathbb{S}$, then $\mathbb{P}(A)$ equals the sum of the probabilities of the individual outcomes comprising A.*

Proof. Let O_i be the ith sequence of outcomes comprising A such that $i \in \{1, 2, \ldots\}$ (can be infinite or not). Therefore

$$A = \cup_{i=1} O_i.$$

Since the individual outcomes, O_i, are mutually exclusive, we can use axiom three to calculate

$$\mathbb{P}(A) = \sum_{i=1} \mathbb{P}(O_i).$$

□

Example 2.2. *When throwing two four sided dice the sample space is given by $\mathbb{S} = \{(x_1, x_2) | x = 1, 2, 3, 4, y = 1, 2, 3, 4\}$ such that x_i is the number on the base of dice $i \in \{1, 2\}$. Let E be the event that $x_1 + x_2 = 4$. Use Theorem 2.1 to show that $\mathbb{P}(E) = \frac{3}{16}$.*

Solution. *The event space $E = \{(1, 3), (2, 2), (3, 1)\}$ and $|\mathbb{S}| = 16$. Therefore using Theorem 2.1*

$$\mathbb{P}(E) = \mathbb{P}((x_1, x_2) = (1, 3)) + \mathbb{P}((x_1, x_2) = (2, 2)) + \mathbb{P}((x_1, x_2) = (3, 1)) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{3}{16}. \quad (5)$$

Example 2.3. *For a given experiment with an infinite number of outcomes O_i for $i \in \mathbb{N}_0$, verify that*

$$\mathbb{P}(O_i) = \left(\frac{1}{2}\right)^i, \text{ for } i = 1, 2, ... \tag{6}$$

is a probability measure.

Solution. *Firstly, it is true that $\mathbb{P}(O_i) \geq 0$. It therefore suffices to show that $\mathbb{P}(\mathbb{S}) = 1$,*

$$\mathbb{P}(\mathbb{S}) = \sum_{i=1}^{\infty} \mathbb{P}(O_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1 \tag{7}$$

using

$$\sum_{i=1}^{\infty} ar^i = \frac{a}{1-r}, \text{ such that } |r| < 1. \tag{8}$$

Here for our example, the first term is $a = \frac{1}{2}$ and the common ratio $r = \frac{1}{2}$.

2.1.2 Counting methods for classical probability

The sample space may be very large, so writing out all the possible elements and counting directly may be laborious. Often, we can apply some standard results for **permutations** and **combinations**, together with a little logic.

Definition 2.4. *Let ${}^nP_r$ denote the number of **permutations**, that is the number of ways of choosing r elements out of n, where the order matters. Then*

$${}^nP_r = \frac{n!}{(n-r)!},$$

here $k! = k \times (k-1) \times (k-2)...2 \times 1$.

Definition 2.5. *Let $\binom{n}{r}$ (also written ${}^nC_r$) denote the number of **combinations**, that is the number of ways of choosing r elements out of n, where the order does not matter. Then*

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Example 2.6. *Consider the National Lottery. On a single ticket, you must choose 6 integers between 1 and 59. A machine also selects 6 integers between 1 and 59, and it is reasonable to assume that each number is equally likely to be chosen. If I buy one ticket:*

1. *What is the probability that I match all 6 numbers?*
2. *What is the probability I match 3 numbers only?*

Solution. *There are $\binom{59}{6}$ outcomes (the order does not matter). In part 1 there is 1 winning outcome, and in part 2 there are $\binom{6}{3}\binom{53}{3}$ outcomes, so the answers are*

$$\frac{1}{\binom{59}{6}} = \frac{1}{45057474} \text{ and } \frac{\binom{6}{3}\binom{53}{3}}{\binom{59}{6}} = 0.0104.$$

We can design 'experiments' in which equally likely outcomes can be assumed. We start with a 'population' of people or items, and then pick a subset of the population, such that each member of the population is equally likely to be selected. This is called (simple) random sampling (some care is needed to make sure we are satisfied that each member really does have the same chance of being selected). This concept is important in Statistics, where we want to make inferences about a population based on what we have observed in a random sample. We can use probability theory to understand how the characteristics of a random sample may differ from the characteristics of a population.

Example 2.7. *A factory has produced 50 items, some of which may be faulty. If an item is tested for a fault, it can no longer be used (testing is 'destructive'). To estimate the proportion of faulty items, a sample of 5 items are selected at random for testing. Suppose out of the 50 items, 10 are faulty.*

1. *What is the probability that none of the 5 items in the sample are faulty?*
2. *What is the probability that the proportion of faulty items in the sample is the same as the proportion of faulty items in the population?*

Solution. *If we select* 5 *items with probability of being faulty* $\frac{1}{5}$. *There are* $\binom{50}{5}$ *possible samples. There are* $\binom{40}{5}$ *possible samples with no faulty items, so the probability is* $\binom{40}{5}/\binom{50}{5}$.

The proportions in the sample and population are the same if 1 item in the sample is faulty; the no. of outcomes which give this is $\binom{40}{4}\binom{10}{1}$, *so the probability is* $\binom{40}{4}\binom{10}{1}/\binom{50}{5}$*=0.4313.*

2.2 Exercises

Exercise 2.8. *If* $A = \{1, -30, 42\}$, $B = \{1, -30\}$ *and* $C = \{-30, 42\}$ *what is*

1. $A \cup B \cup C$
2. $A \cap B \cap C$.

Exercise 2.9. *If* $\mathbb{S} = \{A, B, C\}$, *does* $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.45$ *and* $\mathbb{P}(C) = 0.3$ *represent a valid probability distribution?*

Exercise 2.10. *A five-card poker hand is dealt from a deck of 52 playing cards is said to be a full house if it consists of three of a kind and a pair. If all five-card hands are equally likely, what is the probability of being dealt a full house?*

Exercise 2.11. *If a person visits the fun fair, suppose that the probability that they will have pizza is* 0.44, *eat a burger is* 0.24, *eat a hot dog is* 0.21. *The probability they will have pizza and a burger is* 0.08, *pizza and hot dog is* 0.11, *and the probability they have a burger and a hot dog is* 0.07. *The probability they eat all of pizza, burger and hot dog is* 0.03. *What is the probability that a person visiting the fun fair will eat at least one of these items.*

3 Conditional Probability

Content:

1. Definition of conditional probability.
2. The multiplication rule.

Supplementary reading: Freund §2.6.
Supplementary videos: Conditional probability.

3.1 Class Content

Conditional probability is an important concept that we can use to *modify* a measurement of uncertainty as our information changes.

In previous lectures we have looked at probabilities such as "what is the probability of drawing a king from a pack of playing cards" or "what is the probability of rolling a six on a fair dice". However, you will notice that in the former situation, there is an opportunity to know previous information which may alter the probability of the event occurring. For example, we may have the statement "given we know we have selected a face card, what is the probability we have drawn a king". In the dice example, it wouldn't matter what the previous roll was as each dice roll can be considered independent of previous rolls (more on this later.)

For clarity we will quickly discuss both situations above involving playing cards. For the first example we know that there are 52 cards in a deck and four of these are kings. Therefore the probability of drawing a king is $\frac{4}{52} = \frac{1}{13}$. However, if we know that we have drawn a face card then we are no longer considering the 52 card deck, we are considering the twelve cards which have faces. We know four of these twelve cards are kings. Therefore the probability we draw a king given that we know that the drawn card is a face card is $\frac{4}{12} = \frac{1}{3}$.

Something that should be noted here is that the probability increased ($\frac{1}{3} > \frac{1}{13}$) when we conditioned on knowing more information. This is intuitive, as in life we make more accurate predictions when presented with more information.

Definition 3.1. *We define $\mathbb{P}(E|F)$ to be the conditional probability of E given F, where*

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}. \tag{9}$$

We can interpret this to mean

"If it is known that F has occurred, what is the probability that E has also occurred?"

Note: $\mathbb{P}(E|F)$ is read as "probability of E given F.

If we know that the outcome belongs to the set F, then for E to occur also, the outcome must lie in the intersection $E \cap F$.

Example 3.2. *For a randomly selected individual, suppose the probabilities of the four blood types are $\mathbb{P}(\text{type } O) = 0.45$, $\mathbb{P}(\text{type } A) = 0.4$, $\mathbb{P}(\text{type } B) = 0.1$ and $\mathbb{P}(\text{type } AB) = 0.05$. A test is taken to determine the blood type, but the test is only able to declare that the blood type is either A or B. What is the probability that the blood type is A?*

Solution. *Let E be the event that group is A, and F the event that the group is either A or B. Then $\mathbb{P}(E|F) = \mathbb{P}(E \cap F)/\mathbb{P}(F) = 0.4/(0.4 + 0.1) = 0.8$.*

Definition 3.1 gives us an intuitive way to think about joint probabilities $\mathbb{P}(E \cap F)$, *E and F* occurring. Rearranging (9) we have

$$\mathbb{P}(E \cap F) = \mathbb{P}(F)\mathbb{P}(E|F),$$

clearly also we can swap E and F and write

$$\mathbb{P}(F \cap E) = \mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F|E). \tag{10}$$

This means we can calculate the probability that both E and F occur by considering either

1. the probability that E occurs, and then the probability that F occurs given that E has occurred, or:
2. the probability that F occurs, and then the probability that E occurs given that F has occurred.

The statement given by equation (10) is important and will be used later in the course, particularly in the formulation and discussion of Bayes' Theorem.

Specifying conditional probabilities directly

Equation (9) tells us how to calculate $\mathbb{P}(E|F)$ *if we already know* $\mathbb{P}(E \cap F)$ and $\mathbb{P}(F)$. But in some situations, we may be able to specify $\mathbb{P}(E|F)$ directly, given the information at hand.

Example 3.3. *A diagnostic test has been developed for a particular disease. In a trial group of patients known to be carrying the disease, the test successfully detected the disease for 95% of the patients. An individual is selected at random from the population (and so may or may not be carrying the disease). Let D be the event that the individual has the disease, and T be the event that the test declares the individual has the disease. Using the frequency approach to specifying a probability and the information above, which of the following probabilities can we specify?*

- $\mathbb{P}(D)$
- $\mathbb{P}(D \cap T)$
- $\mathbb{P}(T|D)$
- $\mathbb{P}(D|T)$

Solution. $\mathbb{P}(D)$ *is prob. patient has the disease – unknown.*
$\mathbb{P}(D \cap T)$ *is prob. patient has the disease and tests positive – unknown.*
$\mathbb{P}(T|D)$ *is the prob. a patient who has the disease tests positive - estimated at* 0.95 *in the trial.*
$\mathbb{P}(D|T)$ *is the probability a patient who tests positive has the disease – unknown.*

In some cases, the conditional probability will be 'obvious', and you should have the confidence just to write it down!

Example 3.4.
A playing card is drawn at random from a standard deck of 52. Let A be the event that the card is a heart, and B be the event that the card is red. What are $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$?

Solution. $\mathbb{P}(A|B) = 13/26$ *and* $\mathbb{P}(B|A) = 13/13$,

3.2 Exercises

Exercise 3.5. *In the game show "Who wants to be a Millionaire?", a contestant must answer 12 questions correctly to win £1,000,000. Each question is multiple choice, with four possible answers. For each question, explain your reasoning carefully:*

1. *If a contestant guesses each time, picking one of the four answers at random, what is the probability of winning the million pound prize?*
2. *Now suppose the contestant uses the three "lifelines": for one question, two false answers are eliminated; for another question, the audience votes for the correct answer; for a third question, the contestant may phone a friend for advice. Suppose that both the audience and the friend give the correct answer when asked. The contestant uses the answers from the audience and the friend, but otherwise picks an answer at random each time. What is the probability of winning the million pound prize?*

Exercise 3.6. *An online banking website requires its account holders to choose a four digit pin number, and a case-sensitive password made up of digits and letters (the letters can be upper or lower case). Suppose you choose your pin number and a six character password at random (for the password, each character has the same chance of being any digit or upper or lower case letter). When logging in, the website asks you to specify three digits from your pin number, and three characters from your password, all chosen randomly, and will give you three attempts if you make a mistake. If someone tries to log into your account, choosing the three digits and three characters at random, what is the probability that their first attempt is correct, gaining access to your account?*

Exercise 3.7. *In the game show "Deal or No Deal?", a contestant must choose 5 boxes out of 22 to be opened. Each box contains a different amount of money. Once a box has been opened, its contents can no longer be won. For each question, explain your reasoning carefully:*

1. *What is the probability that the contestant chooses the most valuable box?*
2. *What is the probability that the contestant chooses at least one of the top 5 most valuable boxes?*

Exercise 3.8. *Prove that if A, B and C are any three events in a sample space $\mathbb{S}$ such that $\mathbb{P}(A \cap B) \neq 0$, then*

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

Exercise 3.9. *Let us define an experiment as rolling two dice one after another, dice A first followed by dice B. Let us also define the event $E_{a,b}$ as the sum of the faces of the two dice where a is the value of dice A and d the value shown on dice B. Calculate the following:*

1. $\mathbb{P}(E_{a,b} = 7)$,
2. $\mathbb{P}(E_{3,b} = 7)$,
3. $\mathbb{P}(E_{a,b} = 5)$,
4. $\mathbb{P}(E_{3,b} = 5)$.

4 Independence of events

Content:

1. Definition of independence.
2. Independence for more than two events.
3. The law of total probability.

Supplementary reading: Freund §2.7.
Supplementary videos: Independence of events.

4.1 Class Content

In previous classes we have discussed the case where the probability of an event changes when we condition on a different event. Such as the probability we draw a red card (primary event) from a deck given we have chosen a heart (conditioned event). We will discuss the notion of independence. This is where the act of conditioning on an event does not change the probability of the primary event occuring. For example let us say that we roll two dice separately (i.e. they do not bounce off each other). It should be clear that the value observed on the second dice will not be changed given the observed value on the first dice.

Definition 4.1. *Two events E and F are said to be* ***independent*** *if and only if*

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F).$$

Conversely, if this independence relationship is not satisfied, the events are said to be dependent.

We can use the definition of conditional probability to give a more intuitive definition of independence. The events A and B are independent if

$$\mathbb{P}(E|F) = \mathbb{P}(E). \tag{11}$$

We can read this to mean "If E and F are independent, then learning that F has occurred does not change the probability that E will occur (and vice versa)."

Example 4.2. *Suppose two pregnant women are chosen at random, and consider whether each gives birth to a boy or girl (assume there will be no twins, triplets etc.) We can write the sample space as*

$$S = \{(boy,\ boy), (boy,\ girl), (girl,\ boy), (girl,\ girl)\},$$

where, for example, element $(boy,\ girl)$, is the outcome that the first woman gives birth to a boy, and the second woman gives birth to a girl. Define B_i to be the event that the ith woman gives birth to a boy.

1. *With regard to S, what are the events given by B_1, B_2 and $B_1 \cap B_2$?*
2. *Suppose we assume that each outcome in the sample space is equally likely. What are the values of $\mathbb{P}(B_1)$, $\mathbb{P}(B_2)$ and $\mathbb{P}(B_1 \cap B_2)$?*
3. *Are B_1 and B_2 independent?*

Solution. $B_1 = \{(boy, boy), (boy, girl)\}, B_2 = \{(boy, boy), (girl, boy)\}, B_1 \cap B_2 = \{(boy, boy)\}$

There are four outcomes, and B_1, B_2 and $B_1 \cap B_2$ contain 2, 2 and 1 elements respectively, so $\mathbb{P}(B_1) = 1/2, \mathbb{P}(B_2) = 1/2, \mathbb{P}(B_1 \cap B_2) = 1/4$.

$\mathbb{P}(B_1) = 1/2, \mathbb{P}(B_2) = 1/2, \mathbb{P}(B_1 \cap B_2) = 1/4 = \frac{1}{2} \times \frac{1}{2}$, *so they are independent.*

Example 4.3. *Two playing cards are drawn at random from a standard 52 card deck. Let A be the event of at least one ace, and let K be the event of at least one king.*

1. *Assume that each card in the deck has the same chance of being selected. Calculate $\mathbb{P}(A)$, $\mathbb{P}(K)$ and $\mathbb{P}(A \cap K)$.*

2. *Are A and K independent?*

3. *Compare $\mathbb{P}(A)$ with $\mathbb{P}(A|K)$ and comment on the result.*

Solution. *Easier to calculate $\mathbb{P}(\overline{A}) = \frac{48}{52} \times \frac{47}{51}$. So $\mathbb{P}(A) = 1 - \frac{48}{52} \times \frac{47}{51} = 33/221 = 0.149$; $\mathbb{P}(K)$ is the same.*

$\mathbb{P}(A \cap K) = 2 \times \frac{4}{52} \times \frac{4}{51} = 8/663 = 0.0121$

$\mathbb{P}(A)\mathbb{P}(K) = (33/221)^2 \neq 8/663$ *so not independent* $((33/221)^2 = 0.0223)$.

$\mathbb{P}(A|K) = \mathbb{P}(A \cap K)/\mathbb{P}(K) = (8/663)/(33/221) = 8/99 = 0.0808$, *less than* $\mathbb{P}(A)$.

Careful: this last probability is the probability of getting at least one ace given that we get at least one king. This is different from the probability of getting at least one ace given that we know that the first card is a king, which is $4/51 = 0.0784$.

Definition 4.4 (Extension of Definition 4.1)**.** *Events $E_1, E_2, ..., E_k$ are all independent if and only if the probability of the intersections of any pair, triple, ..., subset of these events equals the product of their respective probabilities.*

Example 4.5. *For the elements A, B, C of a sample space $\mathbb{S} = \{A, B, C\}$ are independent if and only if all of the following hold true:*

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B), \\
\mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C), \\
\mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C), \\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).
\end{aligned}$$

The Venn diagram of this is effectively three non overlapping areas denoted by A, B and C.

4.1.1 Calculating joint probabilities: a summary

We have now seen various ways to calculate a joint probability $\mathbb{P}(A \cap B)$. These are as follows.

1. **Assume independence**
   If we think that learning A has occurred will not change our probability of B occurring (and vice versa) then we have
   $$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

   Example 4.6. *If I buy a single National Lottery ticket, the probability I don't win any prize is 53/54. If I buy one ticket every week, what is the probability I win nothing in the first two weeks? What is the probability I win nothing in the first month (four weeks)?*

   Solution. *Because we can assume the event of winning each week is independent of whether you won in a previous week. For 2 weeks $(53/54)^2 = 0.963$; for four weeks $(53/54)^4 = 0.928$.*

2. **Direct calculation using classical probability**
   When using classical probability (assuming the elements of the sample space are equally likely), it may be straightforward to count in the number of outcomes in which both A and B occur, and hence calculate $\mathbb{P}(A \cap B)$ directly. In a sense, we are saying that if we write C as the event $A \cap B$, it is straightforward to calculate $\mathbb{P}(C)$.

   Example 4.7. *One playing card is drawn at random from a standard deck of 52. Let A be the event that the card is a red face card (king, queen or jack). Let B be the event that the card is a king.*

   (a) *Without doing any calculations say whether you expect A&B to be independent?*

   (b) *What is $\mathbb{P}(A \cap B)$?*

   (c) *Confirm your answer to (a).*

Solution. *(a) We wouldn't expect this because whether or not we select a red king as the first card would effect the probability of the second selection. (b) $A \cap B$ (red king) has 2 elements, A has 6 and B has 4, so $\mathbb{P}(A \cap B) = 2/52 = 1/26$. (c) We can confirm they are not independent: $\mathbb{P}(A) = 6/52$ and $\mathbb{P}(B) = 4/52$, so $\mathbb{P}(A)\mathbb{P}(B) = 3/338 < 1/26 = \mathbb{P}(A \cap B)$.*

3. **Using conditional probability**
   As has already been commented, we have

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

If we already know $\mathbb{P}(A)$ and $\mathbb{P}(B|A)$, or $\mathbb{P}(B)$ and $\mathbb{P}(A|B)$, we can calculate $\mathbb{P}(A \cap B)$.

Example 4.8. *(Example 3.3 revisited). Suppose 1% of people in a population have a particular disease. A diagnostic test has a 95% chance of detecting the disease in a person carrying the disease. One person is selected at random. What is the probability that they have the disease and the diagnostic test detects it?*

Solution. *Let T be the event the person tests positive and D be the event they have the disease. Then $\mathbb{P}(D) = 0.01$ and $\mathbb{P}(T|D) = 0.95$, so $\mathbb{P}(D \cap T) = 0.01 \cdot 0.95 = 0.0095$.*

Having three methods may seem confusing! In fact, the conditional probability method can be thought of as *the* way to calculate a joint probability (remembering that conditional probabilities can be specified directly), with the first two methods being short cuts or special cases.

Example 4.9. *Show how the conditional probability method can be used to calculate the joint probability in Example 4.7.*

Solution. *For two weeks let L_i be event we don't win anything in week i. Then $\mathbb{P}(L_2|L_1) = \mathbb{P}(L_2) = 53/54$ by independence, so $\mathbb{P}(L_1 \cap L_2) = \mathbb{P}(L_1)\mathbb{P}(L_2|L_1) = (53/54)^2$.*

4.1.2 Law of total probability

In some situations, calculating a probability of an event is easiest if we first consider some appropriate conditional probabilities, ensuring we cover all possibilities but avoiding any double counting. This leads to the idea of conditioning on a partition

Theorem 4.10. *(**The law of total probability**) Suppose we have a partition $\{E_1, \ldots, E_n\}$ of a sample space S. Then for any event F,*

$$\mathbb{P}(F) = \sum_{i=1}^{n} \mathbb{P}(F \cap E_i),$$

or, equivalently,

$$\mathbb{P}(F) = \sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i).$$

Note: here the Σ symbol means 'summation': for example

$$\sum_{k=1}^{5} k = 1 + 2 + 3 + 4 + 5 = 15.$$

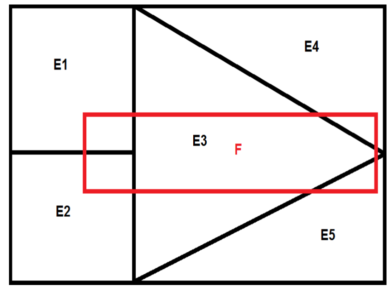

Figure 7: A visual representation of the law of total probability on a sample space with five events.

Remark 4.11. *In Figure 7 you may wish to think about the event F being the probability a person has blue eyes and each of $E_1, E_2, ..., E_5$ being the events that a person has the following hair colours brown, black, blonde, grey and red. The law of probability is simply summing up all of the probability mass associate with having blue eyes. In words this looks like:*
$\mathbb{P}$(Blue eyes)=$\mathbb{P}$(Blue eyes and brown hair)+$\mathbb{P}$(Blue eyes and black hair)+$\mathbb{P}$(Blue eyes and blonde hair) +$\mathbb{P}$(Blue eyes and grey hair)+$\mathbb{P}$(Blue eyes and red hair).

Example 4.12. *Suppose the four teams in this year's Champions League semi-finals are Manchester City, Barcelona, Juventus, and Bayern Munich. Depending on their opponents, you judge Manchester City's probabilities of reaching the final to be*

$$\begin{aligned}
\mathbb{P}(\textit{reach final}|\textit{opponent is Barcelona}) &= 0.2,\\
\mathbb{P}(\textit{reach final}|\textit{opponent is Juventus}) &= 0.4,\\
\mathbb{P}(\textit{reach final}|\textit{opponent is Bayern Munich}) &= 0.5.
\end{aligned}$$

If the semi-final draw has yet to be made (with any two teams having the same probability of being drawn against each other), what is your probability of Manchester City reaching the final?

Solution. *Let F be the probability they reach the final, and D_1, D_2, D_3 be the probabilities they are drawn against Barcelona, Juventus, Bayern respectively. Then* $\mathbb{P}(F) = \mathbb{P}(D_1)\mathbb{P}(F|D_1) + \mathbb{P}(D_2)\mathbb{P}(F|D_2) + \mathbb{P}(D_3)\mathbb{P}(F|D_3) = \frac{1}{3} \cdot 0.2 + \frac{1}{3} \cdot 0.4 + \frac{1}{3} \cdot 0.5 = 11/30 = 0.367$

4.2 Exercises

Exercise 4.13. *In the card game of Blackjack, players attempt to get a total of 21 points in their hand. An ace may be counted as 1 point or 11 points, and face cards (kings, queens, and jacks) are counted as 10 points. For each question, explain your reasoning carefully:*

1. *A player is dealt two cards from a standard deck of 52. What is the probability that these two cards sum to 21 points?*
2. *A player is dealt a 10 and a 6 for his hand, again from a standard deck of 52. His opponent (the dealer) has drawn a 4. If the player draws one more card, what is the probability that the player's total will exceed 21 points? (You should assume that an Ace will be counted as 1 point in this context.)*

Exercise 4.14. *If we roll four dice independent of each other, what is the probability we roll at least one six.*

Exercise 4.15. *If A and B are independent, then prove that A and B^c are also independent (remember B^c is the compliment of B sometimes written as B'.)*

Exercise 4.16. *In a group of 100 men, 55 are clean-shaven, 30 have beards and moustaches, 10 have moustaches only, and 5 have beards only. One man is selected at random. For each question, explain your reasoning carefully:*

1. *What is the probability that the man has a beard?*
2. *If it is known that the selected man has a beard, what is the probability that he has a moustache?*

3. *In this group of 100 men, are the events of having a beard and having a moustache independent for a randomly selected man?*

5 Introduction to Random Variables

Content:

1. Concept of a random variable.

Supplementary reading: Freund §3.1, but distribution functions are later (definition 3.5 and the following material).
Supplementary videos: Introduction to random variables.

5.1 Class Content

We regularly encounter variables in other areas of mathematics, such as in equations like

$$2x + 3y = 6.$$

Here the values x and y are variables. If we consider what is done here, we are representing every pair of values which satisfy this equation using a relationship involving x and y, for example, $(x, y) \in \{(3, 0), (0, 2), (-3, 4), ...\}$.

Informally, in a probabilistic sense, we think of a **random variable** as any uncertain quantity which satisfies the parameters of the proposed setting. For example:

- the number of emergency call-outs received by a fire station in a given week;
- the price of a barrel of oil in one month's time;
- the number of gold medals won by Great Britain at the next summer Olympics.

We cannot say with certainty what any of these quantities are, but probability theory gives us a framework for describing how *likely* different values are. Whereas elements of a sample space may not be numerical, random variables are always numerical quantities, and so, when defining a random variable, we need a rule for getting from the random outcome in the sample space to the value of the random variable.

Definition 5.1. *Given a sample space $\mathbb{S}$, we define a* ***random variable*** *X to be a mapping from $\mathbb{S}$ to a real number $\mathbb{R}$, i.e. we write a random variable as $X(s)$, where $s \in \mathbb{S}$.*

We define the ***range*** *of X to be the set of all possible values of X:*

$$R_X = \{x \in \mathbb{R}; x = X(s) \text{ for some } s \in \mathbb{S}\}.$$

Notice here that we denote random variables as capital letters.

5.2 Exercises

In this section, there is only one exercise question with many separate parts. This is to practice this important step for future problems.

Exercise 5.2. *For each of the following cases, define an appropriate random variable and state the range it can take.*

1. *The amount of times a person blinks each day.*
2. *The distance two pigs are apart from each other in a one meter square pen.*
3. *The roll of a fair six sided dice.*
4. *The temperature outside in Celsius.*
5. *The height of sunflowers in a field.*
6. *Rainfall in the UK in a year.*
7. *Waiting time for a bus.*

8. *Number of darts before the first bullseye.*
9. *The difference in height between a person and the average.*
10. *A dogs distance above sea level.*

All solutions to future problems should begin with a clear definition of your notation, this includes all assigned random variables. There is almost always marks assigned for this step.

6 Discrete Random Variables

Content:

1. Distributions of discrete random variables.
2. The probability mass function.
3. Cumulative distribution function.

Supplementary reading: Freund §3.2 (Note: The function which Freund refers to as the probability distribution is often called the *probability mass function* or simply *probability function*.)
Supplementary videos: Discrete random variables.

6.1 Class Content

In this chapter, we consider **discrete** random variables, in which the number of possible values is either finite or countably infinite. We will also discuss the probability mass function (p.m.f.) which is associated to discrete random variables.

Definition 6.1 (Discrete Random Variable). *A discrete random variable is a random variable which takes specific values and not a value in a given interval.*

Example 6.2. *The following are a list of items followed by a description as to whether they can (or cannot) be represented as a discrete random variable:*

1. *The number of eggs laid by a chicken each year can be modelled as a discrete random variable as it can take integer values, specifically* $\{0, 1, 2, ...\}$.
2. *The height of a randomly chosen member of the public cannot be modelled as a discrete random variable as it can loosely take any value in the interval* $[0, \infty)$.
3. *The number of buses which arrive at a bus stop in an hour can be modelled as a discrete random variable as it can only take the integer values* $\{0, 1, 2, ...\}$.
4. *The value rolled on a fair dice can be modelled as a discrete random variable as it can only take the values* $\{1, 2, 3, 4, 5, 6\}$.
5. *A persons weight cannot be modelled as a discrete random variable as it can (kind of) take any value in the interval* $[0, \infty)$.

All of these cases which cannot be represented as discrete random variables can be modelled as continuous random variables. We will study continuous random variables later in the course. Later in the course we will also discuss cases in which discrete random variables can be approximated to continuous random variables.

6.1.1 Summation notation

Before continuing with the discussion of random variables, we will define the summation notation, and recap some results regarding manipulations of sums. Let X be a random variable taking values $R_X = \{x_1, x_2, \ldots, x_k\}$. For any function g, we define

$$\sum_{x \in R_X} g(x) = \sum_{i=1}^{k} g(x_i) = g(x_1) + g(x_2) + \ldots + g(x_k). \tag{12}$$

For any two constants a and b, we have

$$\begin{aligned} \sum_{x \in R_X} (a + bg(x)) = \sum_{i=1}^{n} (a + bg(x_i)) &= (a + bg(x_1)) + (a + bg(x_2)) + \ldots + (a + bg(x_n)) \\ &= na + b \sum_{i=1}^{n} g(x_i). \end{aligned}$$

Note that

$$\sum_{i=1}^{n} a = na,$$

(avoid a common mistake which is that the sum is not equal to a.)

6.1.2 Probability mass function (p.m.f.)

Definition 6.3. *For a discrete random variable X, we define the* ***probability mass function*** *(p.m.f.) p_X to be*

$$p_X(x) = \mathbb{P}(X = x),$$

where x can be any real number.

In this notation, we use X to represent the random variable, and x to represent a possible value of X. Whereas X refers to a specific random variable, the use of the letter x is arbitrary; we could just as well write $p_X(a) = \mathbb{P}(X = a)$.

A probability mass function must have the following two properties.

1. $p_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. Probability mass functions must sum to 1:

$$\sum_{x \in R_X} p_X(x) = 1.$$

These properties intuitively follow from the axioms of probability covered earlier in the course.

Example 6.4. *For a visual representation of the probability mass function, please see figure 8. Consider an experiment consisting of twenty trials, each with an independent probability p of success. We wish to model the random variable X as the number of successful trials. The p.m.f. given by $P_X(x)$ is the function which outputs the associated probability describing x successful trials. In figure 8 we see the p.m.f. plotted for two values of p. We will consider the simplest case, $p = 0.5$, which can be thought of as a coin flip example.*

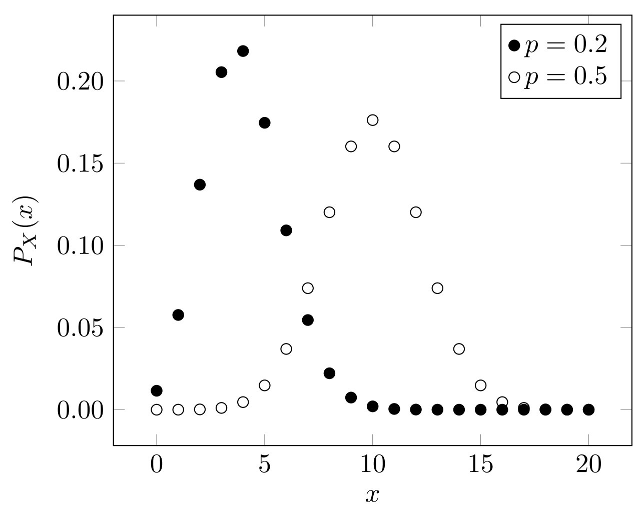

Figure 8: Some plot

For a specific value $x = 10$ in the $p = 0.5$ case, the p.m.f. evaluated at $x = 10$ is $P_X(10) = 0.18$ (approximately). In words this is saying "when $p = 0.5$ there is a probability of 0.18 that our experiment yields ten successes". We can see that different values of p will change the shape of the diagram. In the case when $p = 0.2$, we have that $P_X(10) = 0.01$ (approximately).

Notice the disconnected set of points plotted in figure 8. This is because the p.m.f. does not describe values for $x \notin \{0, 1, 2, ...20\}$.

Example 6.5. *In a simple lottery, two numbers are drawn at random, without replacement, from the numbers 1,2,3,4. You choose two numbers (here the order does not matter). If both 1 and 3 are drawn, you win £10. If either 1 or 3 is drawn (but not both), you win £5. Otherwise, you win nothing. Let X be the amount in pounds that you win. Tabulate $p_X(x)$.*

Solution. *There are six outcomes $\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$ (the order does not matter). We win £10 in one and £5 in four. So*

x	relevant outcomes	$p_X(x)$
0	(2,4)	1/6
5	(1,2),(1,4),(2,3),(3,4)	2/3
10	(1,3)	1/6

6.1.3 Cumulative distribution function

Definition 6.6. *We define the **cumulative distribution function**, abbreviated to c.d.f., F_X of a random variable X (discrete or continuous) to be*

$$F_X(x) = \mathbb{P}(X \leq x),$$

where x can be any real number.

The cumulative distribution function can be written in terms of the probability mass function:

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{a \leq x, a \in R_X} p_X(a). \tag{13}$$

Theorem 6.7. *The values of $F_X(x)$ of the distribution function of a discrete random variable X satisfies:*

1. *$F_X(-\infty) = 0$ and $F_X(\infty) = 1$;*
2. *if $a \leq b$, then $F_X(a) \leq F_X(b)$ for any real numbers a and b.*

Example 6.8. *Suppose England are to play the West Indies in a 3 match cricket test series. Let X be the number of matches won by England. If my probability mass function for X is*

$$p_X(0) = 0.05,\ p_X(1) = 0.2,\ p_X(2) = 0.6,\ p_X(3) = 0.15,$$

tabulate my cumulative distribution function.

Solution.

Range	$F_X(x)$	Relevant Outcome
$x < 0$	0	Win none
$0 \leq x < 1$	0.05	Win 0
$1 \leq x < 2$	0.25	Win 0 or 1
$2 \leq x < 3$	0.85	Win 0 or 1 or 2
$x \geq 3$	1	Win 0 or 1 or 2 or 3

Note that even though X can only take the values 0,1,2 and 3, $F_X(x)$ is defined for all $x \in \mathbb{R}$. Even though $p_X(1.3) = 0$, it would still be correct to state that $F_X(1.3) = \mathbb{P}(X \leq 1.3) = 0.25$.

6.1.4 Quantile functions

The quantile function is related to the inverse of the cumulative distribution function. The function is often used in statistics.

Definition 6.9. *For $\alpha \in [0, 1]$ the α* ***quantile*** *(or $100 \times \alpha$* ***percentile****) is the smallest value of x such that*

$$F_X(x) \geq \alpha$$

The ***median*** *is the 0.5 quantile.*

Example 6.10. *Why might we be interested in the α quantile? Can you think of an example?*

Solution. *If 1000 people were to take an entrance exam and only the top 10% get into the institute. You might want to know the 0.9 quantile, i.e. what threshold mark did the top 10% exceed.*

6.1.5 Independence of random variables

The idea of independent events extends to random variables. We say that two random variables X and Y are independent of each other if any event defined only using the value of X is independent of any event defined only using the value of Y. More specifically, we can make the following definition for discrete random variables:

Definition 6.11. *Two discrete random variables X and Y are* ***independent*** *if*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),$$

for all x and y, or, equivalently, if

$$\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x).$$

If two random variables are not independent, then we say that they are ***dependent****.*

Example 6.12. *We have a jar with 4 blue balls and 8 red balls, two balls are removed sequentially without replacement. Let X (respectively Y) be the number of blue balls removes at the first (respectively second) draw, so each of X and Y can only be 0 or 1. Are X and Y independent random variables?*

Solution. *No, the probability the second ball is blue will vary depending on whether the first ball was blue or red.*

6.2 Exercises

Exercise 6.13.

1. *Find*

$$S = \sum_{k=1}^{20}(5 + 3k).$$

   Note: You can use the result that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ for any positive integer n.

2. *Find a similar expression for*

$$S = \sum_{k=1}^{n}(a + bk).$$

3. *Use part (b) to show (a) is correct ($n = 20, a = 5, b = 3.$)*

Exercise 6.14. *For the p.m.f. $P_X(0) = \frac{0}{12}$, $P_X(1) = \frac{6}{12}$, $P_X(2) = \frac{2}{12}$, $P_X(3) = \frac{1}{12}$, $P_X(4) = \frac{1}{12}$ and $P_X(5) = \frac{2}{12}$. Find $F_X(4.5)$.*

Exercise 6.15. *Check whether the function given by*

$$g_X(x) = \frac{x+1}{10}, \text{ for } x = 0, 1, ..., 3$$

could serve as the p.m.f. of a discrete random variable.

Exercise 6.16. *Find the distribution function ($F_X(x)$) of the total number of sixes obtained in three rolls of a fair dice.*

7 Expectation

Content:

1. Definition of mean for discrete random variables.
2. Interpretation of mean.

Supplementary reading: Freund §4.1 and 4.2, which also include material about the continuous case.
Supplementary videos: Expectation (mean/average).
Notes: For now, focus on the discrete case in Freund §4.2, in particular the definitions and examples 4.1, 4.3 and 4.5.

The idea of mathematical expectation is effectively the average associated to an experiment. For example, if we conduct the experiment of tossing a coin one thousand times and assign the random variable X as the number of heads tossed. The expected value for this experiment would be the expected number of heads we toss. Unsurprisingly the expected value for this situation is five hundred. Another way to think about the expectation is the value x such that $F_X(x) = 0.5$.

7.1 Class Content

If we knew what the probability your favourite sweet might be present in a mixed bag was and how many sweets were in the bag, an obvious question is how many of your favourite sweets you expect to be in each bag.

Definition 7.1. *The **expectation** $\mathbb{E}(X)$ of a discrete random variable X is defined as*

$$\mathbb{E}(X) = \sum_{x \in R_X} x\mathbb{P}(X = x)$$

*We also refer to the expectation as **expected value** of X or the **mean** of X and write this as:*

$$\mu_X = \mathbb{E}(X).$$

Example 7.2. *Say we were to place a bet of one pound on the probability you select a king from a shuffled deck of cards. If you win you get your original one pound back and are paid five pounds for winning. If you lose the betting shop keeps your one pound bet. Let X be your winnings after a single bet of one pound. What is $\mathbb{E}(X)$?*

Solution. $\mathbb{E}(X) = 5 \cdot \frac{4}{52} + (-1) \cdot \frac{48}{52} = -\frac{28}{52}$. *Therefore you are expected to lose $\frac{28}{52}$ pounds on average after each bet.*

Sometimes it is useful to calculate the expectation of a *function* of X. The following result generalizes the one.

Theorem 7.3. *(The expectation of $g(X)$)*

For any function g of a random variable X, with probability mass function $p_X(x)$,

$$\mathbb{E}(g(X)) = \sum_{x \in R_X} g(x)p_X(x).$$

Example 7.4. *Let X be a random variable with $R_X = \{-1, 0, 1\}$. Define $Y = g(X) = X^2$. Then Y is another random variable, with $R_Y(y) = \{0, 1\}$. What is $\mathbb{E}\{g(X)\}$?*

Solution. $\mathbb{P}(Y = 0) = \mathbb{P}(X = 0)$ *and* $\mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1)$. *So* $\mathbb{E}(g(X)) = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1)$.

Note that it is important to realize that the expected value need not be a value that X can actually take (and generally will not be). The easiest example of this is if we roll a standard six sided dice then the expected value is 3.5 (you can easily check this) which is obviously not a value shown on a standard dice.

7.2 Exercises

Exercise 7.5. *Form a list of 9 digits using the 9 digits in your student registration number and add the digit 5 to the list (For example suppose your registration number is 110258441, so the list of digits is 1,1,0,2,5,8,4,4,1,5). Let X be the number obtained by selecting at random one of the 10 digits from the list, such that each of the 10 digits has the same probability of being selected. Tabulate the probability mass function and cumulative distribution function of X, for integer values of X from 0 to 10 inclusive. Calculate the expectation of X.*

Exercise 7.6. *Above we state that the expected roll of a six sided dice is* 3.5. *Show this.*

Exercise 7.7. *Calculate the expected value associated to the p.m.f.*

$$P_X(x) = \frac{x+1}{10}, \text{ such that } x \in \{0, 1, 2, 3\} \text{ and } P_X(x) = 0 \text{ otherwise.}$$

Exercise 7.8. *Explain why the notion of an expected value would be important for someone working in game design for a casino.*

Exercise 7.9. *Consider the experiment whereby we flip* 10 *fair coins. Let X be the number of heads we toss. Calculate $\mathbb{E}(X)$ and $\mathbb{P}(X = \mathbb{E}(X))$. Discuss what you notice. (Note, when calculating $\mathbb{E}(X)$, make use of the Choose function $\binom{n}{k}$.)*

8 Variance and Standard Deviation

Content:

1. Definition of variance.
2. Calculation of variance.
3. Standard deviation.

Supplementary reading: Freund §4.3.
Supplementary videos: Variance and standard deviation

8.1 Class Content

8.1.1 Moments

Definition 8.1 (Discrete moments about the origin). *The rth moment about the origin of a discrete random variable X is defined as*

$$\mathbb{E}(X^r) = \sum_x x^r P_X(x)$$

for $r = 0, 1, \ldots$.

For completeness we will also define the rth moment about the origin of a continuous random with p.d.f. $f_X(x)$. We will discuss continuous random variables later in the course.

Definition 8.2 (Continuous moment about the origin). *The rth moment about the origin of a continuous random variable X is defined as*

$$\mathbb{E}(X^r) = \int_{-\infty}^{\infty} x^r f_X(x)$$

for $r = 0, 1, \ldots$.

It should be clear that the 1st moment about the origin (i.e. $r = 1$ in Definition 8.1) is the expected 'mean' value of a discrete random variable. We can use Definition 8.2 similarly to calculate the expected value of a continuous random variable.

Definition 8.3. *Specifically we define*

$$\mathbb{E}(X^1) = \mu$$

as the mean (all of the following can be used: average/mean/expected value/1st moment.)

When we start looking at problems which involve multiple variables we will be required to be specific about which variable's mean we are referring to. Therefore, you may want to get into the habit of using μ_X to refer to the mean of X.

Definition 8.4 (Discrete moments about the mean). *The rth moment about the mean of a discrete random variable X is defined as*

$$\mathbb{E}((X - \mu_X)^r) = \sum_x (x - \mu_X)^r P_X(x)$$

for $r = 0, 1, \ldots$.

For completeness we will also define the rth moment about the mean of a continuous random with p.d.f. $f_X(x)$.

Definition 8.5 (Continuous moments about the mean). *The rth moment about the mean of a continuous random variable X is defined as*

$$\mathbb{E}((x - \mu_X)^r) = \int_{-\infty}^{\infty} (x - \mu_X)^r f_X(x)$$

for $r = 0, 1, \ldots$.

8.1.2 Variance

If we can repeat the experiment and observe X lots of times, informally, the expectation of X tells us what we are likely to see 'on average'. It will also be useful to consider how far X might be from its expectation. Consider two random variables X and Y, with the following probability mass functions:

$$p_X(32) = \tfrac{1}{3},\ p_X(36) = \tfrac{1}{3},\ p_X(46) = \tfrac{1}{3},$$
$$p_Y(12) = \tfrac{1}{3},\ p_Y(20) = \tfrac{1}{3},\ p_Y(82) = \tfrac{1}{3}.$$

The expected values of X and Y are given by:

$$\begin{aligned}\mathbb{E}(X) &= 32 \cdot \frac{1}{3} + 36 \cdot \frac{1}{3} + 46 \cdot \frac{1}{3} = 38,\\ \mathbb{E}(Y) &= 12 \cdot \frac{1}{3} + 20 \cdot \frac{1}{3} + 82 \cdot \frac{1}{3} = 38.\end{aligned}$$

You can see both X and Y have the same expected value, but whatever values of X and Y we observe, X will be closer to $\mathbb{E}(X)$ than Y will be to $\mathbb{E}(Y)$.

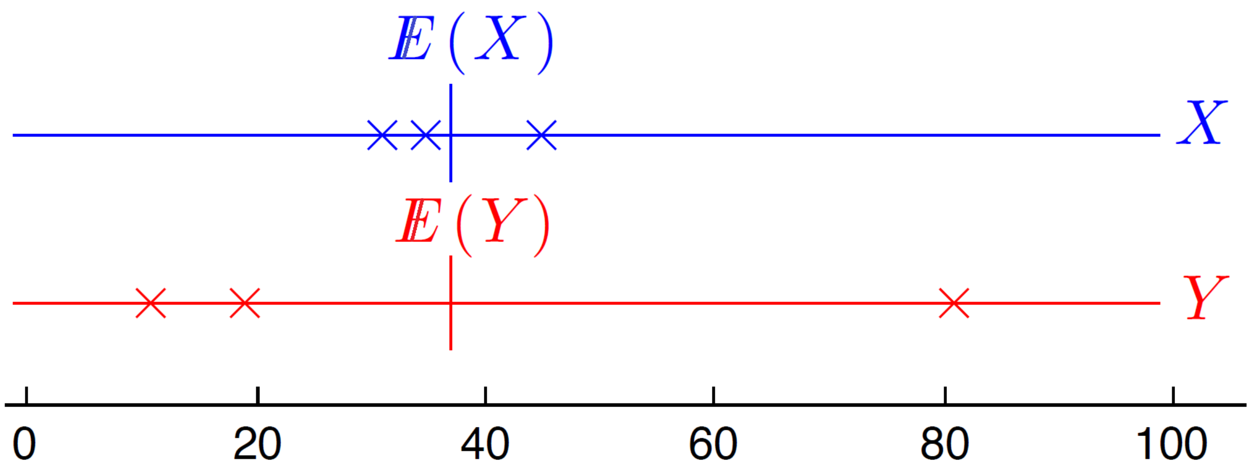

Figure 9: The possible values of X and Y are shown as the crosses (with each value equally likely). Both X and Y have the same mean, but we can see Y must be further away from its mean.

We use the measure of spread known as the **variance** to describe how close a random variable is likely to be to its expected value.

Definition 8.6. *The* ***variance*** *Var(X) of a discrete random variable X is defined as*

$$\mathrm{Var}(X) = \mathbb{E}\left(\{X - \mu_X\}^2\right).$$

We usually denote the variance of the random variable X by σ_X^2:

$$\sigma_X^2 = \mathrm{Var}(X).$$

Example 8.7. *For a random variable X which can take values in $\{1, 3, 5, 7\}$ each with probability $\frac{1}{4}$, calculate Var(X).*

Solution. $\mathbb{E}(X) = \frac{1+3+5+7}{4} = 4$ *so* $Var(X) = \frac{1}{4}\left((1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2\right) = \frac{9+1+1+9}{4} = 5.$

Example 8.8. *What does it mean for a data set to have $\mathbb{E}(X) = k$ and $Var(X) = 0$?*

Solution. *All values in the data set are equal to k (the data all equal k and do not vary).*

The following result is useful for calculating variances:

Theorem 8.9. *(The variance identity) The variance can also be expressed by*

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \tag{14}$$

Note that as long as $\mathrm{Var}(X) > 0$ we can see that

$$\mathbb{E}(X^2) \neq (\mathbb{E}(X))^2.$$

Example 8.10. *Revisit Example 8.7 and use equation (14) to calculate $\mathbb{E}(X^2)$ if we know $Var(X) = 5$?*

Solution. *Using $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ we get that $5 = \mathbb{E}(X^2) - 4^2$ therefore $\mathbb{E}(X^2) = 21$. We can verify this by way of $\mathbb{E}(X^2) = \frac{1}{4}(1^2 + 3^2 + 5^2 + 7^2) = 21$.*

8.1.3 Standard deviation

As the variance is defined as an expected squared difference, the variance will be expressed in units that are the square of the units of X. If we want a measure of spread of data in the same units as X, we take the square root of the variance.

Definition 8.11. *The* ***standard deviation*** *of a random variable X, denoted by σ_X, is the (positive) square root of the variance of X.*

$$\sigma_X = \sqrt{Var(X)}.$$

8.2 Exercises

Exercise 8.12. *For a discrete random variable X with the p.m.f. $P_X(x) = \frac{x+1}{10}$ such that $x = 0, 1, 2, 3$, calculate the first moment about the origin.*

Exercise 8.13. *For a discrete random variable X with the p.m.f. $P_X(x) = \frac{x+1}{10}$ such that $x = 0, 1, 2, 3$, calculate the second moment about the origin.*

Exercise 8.14. *For a discrete random variable X with the p.m.f. $P_X(x) = \frac{x+1}{10}$ such that $x = 0, 1, 2, 3$, calculate the standard deviation.*

Exercise 8.15. *Show that $\mathrm{Var}(X) \geq 0$.*

Exercise 8.16. *Find the mean and standard deviation of the random variable X which takes the values $0, 1 \& 4$ with probabilities $0.25, 0.5$ and 0.25 respectively.*

Exercise 8.17. *Prove equation (14) in Theorem 8.9.*

9 Independence of Random Variables

Content:

1. Definition of independence for random variables.
2. Expectation of a product of independent random variables.

Supplementary reading: Done in section 3.7, after many concepts later in the course.
Supplementary videos: Independence of random variables.

9.1 Class Content

In earlier sections we discussed the independence of events in a sample space. We will now discuss the independence of random variables.

Definition 9.1. *Two random variables X and Y are* independent *if, for any two sets of real numbers A and B,*

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),$$

i.e., the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

For discrete random variables this is equivalent to, for all x and y,

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

The definition can be extended to more than two random variables: a set of random variables $X_1, X_2, \ldots, X_n$ are independent if for sets of real numbers $A_1, A_2, \ldots, A_n$ the events $\{X_i \in A_i\}$ for all i are independent.

Example 9.2. *An eight-sided (octahedral) dice and an ordinary six-sided dice are thrown together. Let X_1 be a random variable representing the score on the eight-sided dice and X_2 be a random variable representing the score on the six-sided dice. Assuming that the dice are fair and that X_1 and X_2 are independent, calculate the probability that both dice show at least 4.*

Solution.

$$\mathbb{P}(X_1 \geq 4 \cap X_2 \geq 4) = \mathbb{P}(X_1 \geq 4)\mathbb{P}(X_2 \geq 4) = \frac{5}{8} \times \frac{1}{2} = \frac{5}{16}.$$

Independence is often assumed in a probabilistic model, as in the above example.

9.1.1 Expectation

An important fact about independent random variables is that the expectation of their product is the product of their expectations:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

We will prove this for discrete random variables X and Y. We can label the set of values X can take by x_i, $i = 1, 2, 3, \ldots$ and the set of values that Y can take by y_j, $j = 1, 2, 3, \ldots$. In the expressions below we assume that X and Y can take infinitely many values; it is easy enough to modify the argument if they can take finitely many values.

We calculate

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) \text{ (by independence)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) \sum_{j=1}^{\infty} y_j \mathbb{P}(Y = y_j) \\
&= \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}
$$

Example 9.3. *Return to the eight-sided and six-sided dice. Assuming that the rolls are independent, calculate the expected product of the numbers on the two dice,* $\mathbb{E}(X_1X_2)$.

Solution: $\mathbb{E}(X_1) = \frac{1}{8}(1 + \cdots + 8) = 9/2$ and $\mathbb{E}(X_2) = \frac{1}{6}(1 + \cdots + 6) = 7/2$. So $\mathbb{E}(X_1X_2) = 63/4(= 15.75)$.

Example 9.4. *Let X and Y be two random variables, both taking values in $i, j \in \{0, 1, 2\}$ where the probabilities $\mathbb{P}(\{X = i\} \cap \{Y = j\})$ are given in the following table:*

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	0.02	0.04	0.14
$X = 1$	0.12	0.16	0.12
$X = 2$	0.36	0.04	0

1. *By considering $\mathbb{P}(X = 0 \cap Y = 0)$, show that X and Y are not independent.*
2. *Calculate $\mathbb{E}(XY)$, and compare it with $\mathbb{E}(X)\mathbb{E}(Y)$.*

Solution. *By adding up the terms in the table, $\mathbb{P}(X = 0) = 0.2$ and $\mathbb{P}(Y = 0) = 0.5$. Under independence we would have $\mathbb{P}(X = 0 \cap Y = 0) = 0.2 \times 0.5 = 0.1$, but in fact we have $\mathbb{P}(X = 0 \cap Y = 0) = 0.02$, so X and Y cannot be independent.*

To calculate $\mathbb{E}(XY)$, we need to work out the values XY can take and their probabilities. Using the table, we have $\mathbb{P}(XY = 0) = 0.68$, $\mathbb{P}(XY = 1) = 0.16$ and $\mathbb{P}(XY = 2) = 0.16$. So $\mathbb{E}(XY) = 0.48$.

We calculate $\mathbb{E}(X) = 1.2$ and $\mathbb{E}(Y) = 0.76$. So $\mathbb{E}(X)\mathbb{E}(Y) = 0.912 \neq \mathbb{E}(XY)$, so the result for independent random variables would not have worked here.

9.2 Exercises

Exercise 9.5. *Use independence of the random variables to calculate the probability that both dice in the above example show a number which is a multiple of 3.*

Exercise 9.6. *Three ordinary fair dice are thrown, and the product of the numbers shown Y is recorded. What is the mean of Y?*

10 Mean and Variance of Linear Combinations

Content:

1. Mean of $aX \pm bY \pm c$ with a, b, c constants.
2. Variance of $aX \pm bY \pm c$ with a, b, c constants.

Supplementary reading: Freund: Theorems 4.2, 4.7 and 4.14 (first part).
Supplementary videos:Mean and variance of linear combinations.

10.1 Class Content

10.1.1 Mean of a linear combination of two random variables

Consider two discrete random variables X and Y (not necessarily independent) with means $\mathbb{E}(X)$ and $\mathbb{E}(Y)$, respectively. An important result concerns the mean of the linear combination $aX + bY$.

Theorem 10.1.
$$\mathbb{E}(aX \pm bY \pm c) = a\mathbb{E}(X) \pm b\mathbb{E}(Y) \pm c.$$

Proof. As X and Y are discrete random variables, we can label the set of values X can take by x_i, $i = 1, 2, 3, \ldots$ and the set of values that Y can take by y_j, $j = 1, 2, 3, \ldots$. In the expressions below we assume that X and Y can take infinitely many values; it is easy enough to modify the argument if they can take finitely many values.

The possible values that $aX \pm bY \pm c$ can take are then of the form $ax_i \pm by_j \pm c$ for some i and j, and we can write the expectation as

$$\mathbb{E}(aX \pm bY \pm c) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(ax_i \pm by_j \pm c)\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\})$$

by the definition of expectation.

Rearranging,

$$\begin{aligned}
\mathbb{E}(aX \pm bY \pm c) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(aX \pm bY \pm c)\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} ax_i\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \pm \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} by_j\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \\
&\pm \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} c\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\})
\end{aligned}$$

The first term in the sum can be rearranged as

$$\sum_{i=1}^{\infty} ax_i \sum_{j=1}^{\infty}\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}),$$

and because

$$\sum_{j=1}^{\infty}\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) = \mathbb{P}(X = x_i)$$

it becomes

$$\sum_{i=1}^{\infty} ax_i\mathbb{P}(X = x_i) = a\mathbb{E}(X).$$

Similarly the second term is $b\mathbb{E}(Y)$. For the final term we can move the c to outside the sum and notice this summation is simply summing over all probabilities. Therefore this final summation equals $c \times 1$. This completes the proof. $\square$

The result also holds for the continuous random variables we will meet in the next few weeks, and can be proved using a version of the above argument with the sums replaced by integrals. It can also be extended to more than two random variables; see the expectation part of Theorem 4.14 in Freund.

Example 10.2. *A shopper visits shop A X times in a week and shop B Y times in a week, where X is a random variable with mean 2 and variance 1 and Y is a random variable with mean 1 and variance 0.25. The amount spent at shop A is always £12 and that at shop B is always £20. Calculate the mean amount spent in a week in the two shops.*

Solution. *The total spent in the two shops is £$(12X + 20Y)$. So the mean total spent is £$(12\mathbb{E}(X) + 20\mathbb{E}(Y))$ =£$(12 \times 2 + 20 \times 1)$ =£44.*

10.1.2 Variance of independent sums

Another important result concerns the variance of a sum of **independent** random variables. This is a special case of the second part of

Theorem 10.3. *If X and Y are independent then*

$$\text{Var}(aX \pm bY \pm c) = a^2 \, \text{Var}(X) + b^2 \, \text{Var}(Y)$$

for any constants a, b, c.

Proof. Writing out the variance of $X + Y$ using the definition,

$$\begin{aligned}
\text{Var}(aX + bY) &= \mathbb{E}\left(((aX + bY) - \mathbb{E}(aX + bY))^2\right) \\
&= \mathbb{E}\big(((aX - \mathbb{E}(aX)) + (bY - \mathbb{E}(bY)))^2\big) \\
&= \mathbb{E}\big((aX - \mathbb{E}(aX))^2\big) + \mathbb{E}\big((bY - \mathbb{E}(bY))^2\big) + 2\mathbb{E}((aX - \mathbb{E}(aX))(bY - \mathbb{E}(bY))) \,.
\end{aligned}$$

The last term here is the *covariance* of X and Y, which we will see more about later in the course. For now, note that if X and Y are independent then so are $aX - \mathbb{E}(aX)$ and $bY - \mathbb{E}(bY)$, so by the results on expectation of a product of independent variables, $E((aX - \mathbb{E}(aX))(bY - \mathbb{E}(bY)))$ can be written as $\mathbb{E}(aX - \mathbb{E}(aX))\mathbb{E}(bY - \mathbb{E}(bY))$, and both terms in the product are zero. So

$$\text{Var}(aX + bY) = \mathbb{E}\big((aX - \mathbb{E}(aX))^2\big) + \mathbb{E}\big((bY - \mathbb{E}(bY))^2\big) = \text{Var}(aX) + \text{Var}(bY).$$

We can show that

$$\begin{aligned}
\text{Var}(\pm aX) &= \mathbb{E}((\pm aX)^2) - \mathbb{E}(\pm aX)^2 \\
&= \left(\sum_{i=1}^{\infty} (\pm aX)^2 P_X(x)\right) - \left(\sum_{i=1}^{\infty} (\pm aX) P_X(x)\right)^2 \\
&= \left(\sum_{i=1}^{\infty} a^2 X^2 P_X(x)\right) - a^2 \left(\sum_{i=1}^{\infty} X P_X(x)\right)^2 \\
&= \left(a^2 \left(\sum_{i=1}^{\infty} X^2 P_X(x)\right) - \left(\sum_{i=1}^{\infty} X P_X(x)\right)^2\right) \\
&= a^2 \, \text{Var}(X).
\end{aligned}$$

We know from our introduction to variance that $\text{Var}(c) = 0$ for any constant c.

Using all of this together, we have a generalised result:

$$\operatorname{Var}(aX \pm bY \pm c) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y)$$

if the variables are independent. □

Example 10.4. *Assume that in Example 10.2 the numbers of visits to the two shops are independent. Calculate the variance of the amount spent.*

Solution. *The amount spent is* $12X + 20Y$. *Now* $\operatorname{Var}(12X + 20Y)$, ***if the variables are independent****, is* $144 \operatorname{Var}(X) + 400 \operatorname{Var}(Y)$. *Hence the* $\operatorname{Var}(12X + 20Y) = 144 + 100 = 244$. *The standard deviation is* $\sqrt{244} = 15.62$ *(properly £15.62).*

Theorem 10.5. *Let* $X_1, X_2, ..., X_n$ *be random variables and* $a_1, a_2, ..., a_n$ *be constants. Also let*

$$Y = \sum_{i=1}^{n} a_i X_i$$

. Then it follows that

$$\mathbb{E}(Y) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i)$$

and

$$\operatorname{Var}(Y) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i) + 2 \sum\sum_{i<j} a_i a_j \operatorname{Cov}(X_i, X_j)$$

where the double summation extends over all values of i *and* j*, from* 1 *to* n*, for which* $i < j$.

The covariance is covered later in the course however for this proof you only need to know that for two random variables W and Z, $\operatorname{Cov}(W, Z) = \mathbb{E}((W - \mathbb{E}(W))(Z - \mathbb{E}(Z)))$. The proof is omitted, you are encouraged to prove this yourself however it can be found in Freund section 4.7.

10.2 Exercises

Exercise 10.6. *A discrete random variable* X *has expectation 3 and variance 10. Let* $Y = (X+1)^2$. *Explain what is wrong with the following, and derive the correct expectation of* Y.

$$\begin{aligned}
\mathbb{E}(Y) = E\{(X+1)^2\} &= \mathbb{E}(X^2 + 2X + 1) \\
&= \mathbb{E}(X)^2 + 2\mathbb{E}(X) + 1 \\
&= 3^2 + 2 \times 3 + 1 \\
&= 16.
\end{aligned}$$

Exercise 10.7. *The shops in the previous example increase their prices so that the amount spent in shop A is now £14 and that in shop B £25. Repeat the calculation in the example of the total expected amount spent.*

Exercise 10.8. *Do you think the assumption of independence in Example 10.7 is reasonable?*

Exercise 10.9. *A person's weekly income is a random variable with mean £400 and standard deviation £30, and (independently of their income) they spend an amount which is a random variable with mean £390 and standard deviation £40. Calculate the mean and standard deviation of their net weekly income.*

Exercise 10.10. *Imagine that you buy two scratchcards with probabilities for the different prizes* £0 *with probability 0.689,* £2 *with probability 0.3,* £10 *with probability 0.01 and* £100 *with probability 0.001. Assuming that the prizes from the two scratchcards are independent and they each cost* £1 *to purchase, calculate the expectation and standard deviation of your profit.*

11 Bernoulli Trials and the Binomial Distribution

Content:

1. Bernoulli trials.
2. The binomial distribution.
3. Mean and variance of binomial random variables.

Supplementary reading: Freund §5.3 and 5.4.
Supplementary videos: Bernoulli trials and the Binomial distribution Bernoulli trials and distribution and Binomial distribution.
Related material:

Tasks:

1. Find the mean and variance of a binomial random variable by considering it as the sum of n independent Bernoulli random variables and using the results of Lecture 10.2. (This is Freund exercise 5.6.)
2. Freund exercise 5.40.

Notes:

1. We write $X \sim Bi(n, p)$ if X is a binomial random variable with parameters (n, p), where n is the number of trials and p is the "success" probability.

11.1 Class Content

11.1.1 Bernoulli Distribution

A Bernoulli random variable X can take one of two values: 0 and 1. Examples of 'experiments' that we might describe using a Bernoulli random variable are

- a patient is given a drug, and the drug either 'works': $X = 1$, or does not: $X = 0$;
- a tennis player either wins a match: $X = 1$, or loses: $X = 0$;
- in one year, a house is either burgled: $X = 1$, or not: $X = 0$.

Another way to think of the Bernoulli distribution is the 'coin flip' or 'sucess-fail' distribution. Often a Bernoulli experiment, such as tossing a coin, is referred to as a **Bernoulli trial**.

Definition 11.1. *If a random variable X has a **Bernoulli distribution**, then its probability mass function is*

$$\begin{aligned} P_X(1) &= p, \\ P_X(0) &= 1 - p, \end{aligned}$$

where p is the probability the event happens and $P_X(x) = 0$ otherwise, with $0 \leq p \leq 1$. We write

$$X \sim Bernoulli(p),$$

to mean "X has a Bernoulli distribution with parameter p".

Theorem 11.2. *(Expectation and variance of a Bernoulli random variable)*

For a Bernoulli random variable $X \sim Bernoulli(p)$, we have

$$\begin{aligned} \mathbb{E}(X) &= p, \\ Var(X) &= p(1 - p). \end{aligned}$$

Example 11.3. *Let X be the event we roll a prime number on a six sided dice (i.e. 2,3 or 5). Explain why this follows a Bernoulli distribution, specify its parameters and find $\mathbb{E}(X)$ and $Var(X)$.*

Solution. *Because there are six possible outcomes of which three are prime we have*

$$X \sim Bernoulli\left(\frac{1}{2}\right).$$

This Experiment is distributed according to the bernoulli distribution because we either roll a prime number or not.

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{2}, \\ Var(X) &= \frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{4}. \end{aligned}$$

11.1.2 Binomial Distribution

Consider the following situations:

- 100 patients are given a drug. Each patient either 'responds' to the drug, or does not. X is the number of patients that respond to the drug.
- In a crime survey, 1000 people are selected at random, and asked whether they have been burgled in the last year. X is the number of people who respond 'yes'.
- In a quality control procedure, 20 items are selected at random, and tested to see whether they are faulty. X is the number of faulty items.

In each case we have a fixed number of "trials", each of which can have two possible outcomes often called "success" and "failure", each individual trial here might be described by a Bernoulli random variable. In each of these situations it is reasonable to assume that the trials are independent and that the probability of a "success", which we denote using p, is constant from from one trial to the next. In each case we are interested in counting the total number of successes.

For a general number of trials n the number of possible sequences that contain x successes is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Example 11.4. *If we kick* 10 *balls at a goal, how many different combinations will result in* 5 *goals?*

Solution. *Each kick either results in a goal (G) or not ($\bar{G}$). All sequences of interest are 10-long with* 5 *G's and* 5 *$\bar{G}$'s in any order such as $GGGGG\bar{G}\bar{G}\bar{G}\bar{G}\bar{G}$. The answer is expressed as $\binom{10}{5} = 252$.*

The probability of any individual sequence having x successes in total will be $p^x(1-p)^{n-x}$, since trials are independent.

Definition 11.5. *If a random variable X has a **binomial distribution**, with parameters n (the number of trials) and p (the probability of success in each trial), then the probability mass function of X is given by*

$$P_X(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} = \binom{n}{x}p^x(1-p)^{n-x},$$

for $x = \{0, 1, 2, \ldots, n\}$, and 0 otherwise. We write this

$$X \sim Bin(n, p),$$

to mean "X has a binomial distribution with parameters n (the number of trials) and p (the probability of success in each trial)".

The cumulative distribution function is given by

$$
\begin{aligned}
F_X(x) = \mathbb{P}(X \leq x) &= \sum_{a=0}^{x} p_X(a) \\
&= \sum_{a=0}^{x} \frac{n!}{a!(n-a)!} p^a (1-p)^{n-a}.
\end{aligned}
$$

We cannot simplify this expression, and so calculating the c.d.f. by hand can be tedious. Fortunately, we can do this and other calculations related to the binomial distribution very easily using computer packages such as R and MatLab. How to do this in both MatLab and R are found below.

Theorem 11.6. *(Expectation and variance of a binomial random variable) For $X \sim Bin(n,p)$ we have*

$$
\begin{aligned}
\mathbb{E}(X) &= np \\
Var(X) &= np(1-p).
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x}
\end{aligned}
$$

We will now do a variable change letting $m = n-1$ and $y = x-1$.

$$
\begin{aligned}
\mathbb{E}(X) &= np \sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y} \\
&= np.
\end{aligned}
$$

In order to prove $\mathrm{Var}(X) = np(1-p)$ we will consider

$$
\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} (1-p)^{n-x}
\end{aligned}
$$

This time, we let $y = x-2$ and $m = n-2$.

$$
\mathbb{E}(X(X-1)) = n(n-1)p^2 \sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y}
$$

$$= n(n-1)p^2.$$

We know that

$$\begin{aligned}\text{Var}(X) =&\mathbb{E}(X^2) - \mathbb{E}(X)^2\\ =&\mathbb{E}(X^2) - \mathbb{E}(X) + \mathbb{E}(X) - \mathbb{E}(X)^2\\ =&\mathbb{E}(X^2 - X) + \mathbb{E}(X) - \mathbb{E}(X)^2\\ =&\mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2.\end{aligned}$$

Therefore

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2\\ &= n(n-1)p^2 + np - (np)^2\\ &= np(1-p).\end{aligned}$$

□

11.1.3 The binomial distribution in MatLab

As with most standard distributions, in MatLab, there are commands for calculating the p.m.f., c.d.f., quantile function, and for randomly sampling from the distribution.

- Calculate the p.m.f. : `binopdf(x,n,p)`
  Example: calculate $P_X(2) = \mathbb{P}(X=2)$ when $X \sim Bin(10, 0.3)$.

  ```
  > binopdf(2,10,0.3)
   0.2334744
  ```

  Note: the command is indeed binopdf not binopmf; you will see why this is later in the course.

- Calculate the c.d.f.: `binocdf(x,n,p)`
  Example: calculate $F_X(4) = \mathbb{P}(X \leq 4)$ when $X \sim Bin(30, 0.25)$.

  ```
  > binocdf(4,30,0.25)
   0.0978696
  ```

- To find the α quantile: `binoinv(alpha,n,p)`
  Example: if $X \sim Bin(15, 0.7)$, what is the smallest value of $x \in \{0, 1, \ldots, 15\}$ such that $F_X(x) = \mathbb{P}(X \leq x) \geq 0.4$?

  ```
  > binoinv(0.4,15,0.7)
   10
  ```

  Check:

  ```
  > binocdf(9,15,0.7)
   0.2783786, i.e <0.4
  > binocdf(10,15,0.7)
   0.4845089, i.e. >0.4, so 10 is the smallest $x$ that will do
  ```

- Generate `m` random observations from a binomial distribution: `binornd(n,p,[1,m])`
  Example: generate 10 random observations from the $Bin(100, 0.1)$ distribution.

  ```
   binornd(100,0.1,[1,10])
  ```

  9 11 6 15 10 8 9 8 13 12"

11.1.4 The binomial distribution in R

As with most standard distributions, in R, there are commands for calculating the p.m.f., c.d.f., quantile function, and for randomly sampling from the distribution.

- Calculate the p.m.f. : `dbinom(x,n,p)`
  Example: calculate $P_X(2) = \mathbb{P}(X = 2)$ when $X \sim Bin(10, 0.3)$.

  ```
  > dbinom(2,10,0.3)
   0.2334744
  ```

- Calculate the c.d.f.: `pbinom(x,n,p)`
  Example: calculate $F_X(4) = \mathbb{P}(X \leq 4)$ when $X \sim Bin(30, 0.25)$.

  ```
  > pbinom(4,30,0.25)
   0.0978696
  ```

- To find the α quantile: `qbinom(alpha,n,p)`
  Example: if $X \sim Bin(15, 0.7)$, what is the smallest value of $x \in \{0, 1, \ldots, 15\}$ such that $F_X(x) = \mathbb{P}(X \leq x) \geq 0.4$?

  ```
  > qbinom(0.4,15,0.7)
   10
  ```

  Check:

  ```
  > pbinom(9,15,0.7)
   0.2783786, i.e <0.4
  > pbinom(10,15,0.7)
   0.4845089, i.e. >0.4, so 10 is the smallest $x$ that will do
  ```

- Generate `m` random observations from a binomial distribution: `rbinom(m,n,p)`
  Example: generate 10 random observations from the $Bin(100, 0.1)$ distribution.

  ```
   rbinom(10,100,0.1)
  ```
  9 11 6 15 10 8 9 8 13 12"

Example 11.7. *A company claims that, for a particular product, 8 out of 10 people prefer their brand A over a rival's brand B. You randomly sample 50 people, and ask them whether they prefer brand A to brand B. Let X be the number of people who choose brand A. If the company is right,*

1. *what are the expectation and variance of X?*
2. *what is the probability that $X = 40$?*
3. *what is the probability that $X \leq 30$?*
4. *what is the probability that $X \geq 45$?*

Solution. $X \sim Bin(50, 0.8)$ *assuming their claim and that the sample is genuinely random.*

(1.) So $E(X) = 40$ and $Var(X) = 8$.
(2.) $\mathbb{P}(X = 40) = 0.140$ *("binopdf(40,50,0.8)" in MatLab and "dbinom(40,50,0.8)" in R)*
(3.) $\mathbb{P}(X \leq 30) = 0.000932$ *("binocdf(30,50,0.8)" in MatLab and "pbinom(30,50,0.8) in R")*
(4.) $\mathbb{P}(X \geq 45) = 0.0480$ *("1-binocdf(44,50,0.8)" in MatLab and "1-pbinom(44,50,0.8)" in R)*
Note: Be careful about reversing the inequalities and weak/strict cases. By axioms of probability $\mathbb{P}(X \geq 45) = 1 - \mathbb{P}(X < 45) = 1 - \mathbb{P}(X \leq 44)$ *since X can only take integer values.*

Exam Note 11.8. *For your exam, if asked a question in which you require p.m.f., quantile or c.d.f. values you will be given MatLab or R outputs if it is deemed unreasonable to be asked to calculate these manually. You may be given more than one; you will need to choose carefully and correctly.*

11.2 Exercises

Exercise 11.9. *In a multiple choice test, there are 10 questions, with four answers per question. For each question, only one out of the four answers is correct. If you were to pick one answer at random for each question, calculate the probability of getting exactly 6 out of 10 answers correct. Define any notation that you introduce, and justify your answer*

Exercise 11.10. *A TV chef claims his free-range chickens taste better than battery-farmed chickens. 10 people are given a sample of each to taste, without knowing which is which, and are asked to state which sample they prefer.*

It is suggested that the participants cannot actually taste the difference, and are effectively choosing which sample they prefer at random. If this is true,

1. *what probability distribution would you use to describe the number of people who say they prefer the free-range chicken?*
2. *Using your distribution in (a), calculate the probability that the number of people who say they prefer the free-range chicken is*
    - *not 5;*
    - *no more than 8.*

    Define any notation that you introduce, and justify your answers.

Exercise 11.11. *In a production line, it is estimated that 1 in every 200 items will be faulty. At the quality control stage, each item is visually inspected, and it is believed that there is a 90% chance of detecting a faulty item, and a 5% chance of mistakenly declaring a non-faulty item as being faulty.*

1. *Calculate the probability that a randomly selected item will be declared faulty.*
2. *In a batch of 10 items, what is the probability of two items being declared faulty at the quality control stage? Define any notation that you introduce, and justify your answer.*

Exercise 11.12. *Consider the experiment "A football player kicks 10 footballs at a goal, they score with probability $\frac{1}{3}$". Please give the appropriate distribution and define all notation used carefully.*

Exercise 11.13. *if we roll a six sided dice* 100 *times, what is the expected number of sixes rolled. Define your notation and distribution carefully.*

Exercise 11.14. *Prove directly (as we did for the Binomial distribution above) that for a Bernoulli distributed random variable X, $\mathbb{E}(X) = p$ and $\mathrm{Var}(X) = p(1-p)$.*

12 The Geometric, Negative Binomial and Hypergeometric Distributions

Content:

1. The geometric distribution.
2. Mean and variance.
3. The negative binomial distribution.
4. The hypergeometric distribution.

Supplementary reading: Freund §5.5 and exercises 5.16, 5.17, 5.59. Sometimes the geometric and negative binomial distributions are defined by counting the number of *failures* until the kth success, rather than the number of *trials* needed for k successes. See Freund exercise 5.16, where Y is defined this way.
Supplementary videos: The Geometric distribution, the negative binomial distribution and the hypergeometric distribution.

12.1 Class Content

12.1.1 Geometric distribution

We have previously studied the Binomial distribution as repeated Bernoulli trials. We are sometimes interested in the number of trials in which the first success occurs on the k^{th} trial. For example, the number of successful bowls before the first strike.

Consider the following situations:

- Throwing darts at a dartboard until the bull's eye is hit;
- Buying a national lottery ticket each week until the jackpot is won.

Again, we have a sequence of success/failure Bernoulli trials, but here the number of trials is not fixed in advanced.

Definition 12.1. *If a random variable X has a **geometric distribution**, with parameter p (the probability of a success in any single trial), then the probability mass function of X is given by*

$$P_X(x) = (1-p)^{x-1}p,$$

for $x \in \{1, 2, 3, ...\}$ and 0 otherwise. We write

$$X \sim Geo(p).$$

Notice for the geometric distribution there is no initial number of trials n but only the number of fails until the first success.

Theorem 12.2. *(Expectation and variance of a geometric random variable)*

If $X \sim Geometric(p)$ then

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{p} \\ Var(X) &= \frac{1-p}{p^2}. \end{aligned}$$

Example 12.3. *Suppose we calculate the probability of winning the National Lottery jackpot with a single ticket is $p = \frac{1}{45057474}$. Suppose I buy one ticket per week. Let X be the week number in which I first win the jackpot. What is the expectation of X?*

Solution. *The expectation $E(X) = 45057474$.*

12.1.2 Negative Binomial distribution

If the kth success occurs on the xth trial, there must have been $k-1$ successes in the first $x-1$ trials, the probability of this is $\mathbb{P}(X=k-1)$ which is Binomially distributed with parameters

$$Bin(x-1,p) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k}.$$

As the probability of a success on the xth trial is p, the probability that the xth success occurs on the kth trial is

$$p\binom{x-1}{k-1} p^{k-1}(1-p)^{x-k} = \binom{x-1}{k-1} p^{k}(1-p)^{x-k}.$$

Definition 12.4 (Negative Binomial Distribution)**.** *A Random variable X has a Negative Binomial distribution if and only if it has the p.m.f.*

$$\mathbb{P}(X=x) = \binom{x-1}{k-1} p^{k}(1-p)^{x-k}.$$

Therefore, the number of the trial on which the kth success occurs is a random variable having a Negative Binomial distribution with parameters k and p.

12.1.3 Hypergeometric distribution

Let us consider an experiment whereby we have N items of which M are of one type and $N-M$ are of a different type. The hypergeometric distribution is one in which helps answer questions related to the probability of selecting n items from N items where we would like x of the first type and $n-x$ of the second.

Example 12.5. *We have a bag with 12 red balls and 8 blue balls contained. We select 5 balls at random. What is the probability we selected 2 red balls and 3 blue balls.*

Definition 12.6. *A random variable X has a Hypergeometric distribution with parameters $HyperGeo(n,N,M)$ if and only if its probability mass function is given by*

$$P_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

for $x=0,1,...,n$, $x \le M$ and $n-x \le N-M$.

Solution (Solution to 12.5)**.** *Let X be the number of red balls we select from the bag such that $X \sim HyperGeo(5,20,12)$. The probability we select $x=2$ red balls is*

$$P_X(2) = \frac{\binom{12}{2}\binom{20-12}{5-2}}{\binom{20}{5}} = \frac{77}{323} \approx 0.2384.$$

Theorem 12.7. *The mean and variance of a hypergeometric distributed random variable X are given by*

$$\mathbb{E}(X) = \frac{nM}{N}$$

and

$$\text{Var}(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

Proof is omitted but can be found in Freund.

12.2 Exercises

Exercise 12.8. *Let us say that the probability you hit a bulls eye on a dart board is* 0.1. *What is the probability you score your first bulls eye before the fourth shot.*

Exercise 12.9. *Prove that for a Geometrically distributed random variable* X*, the expected value* $\mathbb{E}(X) = \frac{1}{p}$.

Exercise 12.10. *Let us say that the probability you hit a bulls eye on a dart board is* 0.1. *What is the probability you score your first bulls eye after the third shot.*

Exercise 12.11. *What is the probability that from a bag containing 10 red balls and 5 blue balls that we choose at most one red ball when selecting 3 balls from the bag.*

Exercise 12.12. *In five rounds of snooker Player One wins three times and Player Two wins two times. What is the probability Player One's final win is in before the final match if the probability she wins a game is* 0.6.

13 Continuous Random Variables

Content:

1. Probability density function (p.d.f.)
2. Relationship between p.d.f. and distribution function.
3. Calculating probabilities from the p.d.f.

Supplementary reading:Freund §3.3 and 3.4. Freund exercises 3.17, 3.19, 3.29.
Supplementary videos: Continuous random variables.

13.1 Class Content

In previous sections, we have focussed on discrete random variables, that is random variables which can take particular values in a set. For example a dice roll can take the values $1, 2, 3, 4, 5, 6$ and not other values such as $3.5, 4.99, -1$. In this section we will examine random variables which can take on any value in a particular (continuous) interval. For example the amount of time a person waits for a bus can be modelled as a continuous random variable (loosely) taking any value in the interval $[0, \infty)$. In this section we will discuss many of the continuous analogues to discrete concepts which we have met in earlier sections.

In previous sections we discussed the probability mass function ($P_X(x)$) associated to a discrete random variable (X). We will now discuss the probability density function (p.d.f.) which is the continuous version of the probability mass function.

Definition 13.1. *A **probability density function** (p.d.f. for short) f_X is a function such that both $f_X(x) \geq 0$ for all x, and*

$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1.$$

Note: t here is a dummy variable, so you might see any value here; often we write $\int_{-\infty}^{\infty} f_X(x)\, dx$

Example 13.2. *Find the value k such that the pdf $f_X(x) = kx(x-5)$ is a valid probability density function on $[0, 5]$.*

Solution. *To be valid we must have*

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f_X(x)dx \\
&= \int_{0}^{5} kx(x-5)dx, \text{ since the pdf is zero outside } [0,5]. \\
&= k\left[\frac{x^3}{3} - \frac{5x^2}{2}\right]_0^5 \\
&= k\left(\frac{5^3}{3} - \frac{5^2}{2}\right) \\
&= k \times \frac{-125}{6}
\end{aligned}$$

therefore $k = -\frac{6}{125}$ in order to make the p.d.f. integrate to 1.

Definition 13.3. *A random variable X with p.d.f. f_X has the property that*

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(t)\, dt$$

to describe associated probabilities.

So, whereas a discrete random variable has a probability mass function, a continuous random variable has a probability density function to do this.

We have exactly the same definition of the cumulative distribution function as in the discrete setting:

$$F_X(x) = \mathbb{P}(X \leq x).$$

Essentially $\int$ is the continuous analogue of $\sum$ for a continuous random variable, so this is calculated as an area under a curve instead of a summation:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt,$$

where $f_X(x)$ is the p.d.f. of X.

From the definitions, it follows that

$$\frac{d}{dx} F_X(x) = f_X(x).$$

(essentially differentiating is the continuous analogue of differencing.)

In summary, for a continuous random variable X, we use integrals (area under a curve) to calculate probabilities, and the curve is called the probability density function of X. Note that the condition

$$\int_{-\infty}^{\infty} f_X(t)dt = 1,$$

applies because this integral represents $\mathbb{P}(-\infty < X < \infty)$ which we know must equal 1 because of the axioms of probability. Note also that

$$\mathbb{P}(X = a) = \mathbb{P}(a < X \leq a) = \int_{a}^{a} f_X(x)dx = 0,$$

and that

$$\mathbb{P}(X < a) = \mathbb{P}(X \leq a) - \mathbb{P}(X = a) = \mathbb{P}(X \leq a).$$

Therefore

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = F_X(b) - F_X(a).$$

Example 13.4.
Let X be a random variable with p.d.f. given by $f_X(x) = ke^{-x}$ for $x \in [0, 2]$, and 0 otherwise.

1. *Find the value of k.*
2. *Derive the cumulative distribution function*
3. *Calculate the probability that $X \leq 0.5$.*

Solution.

1. *We must have $\int_0^2 ke^{-x}\, dx = 1$. As $\int_0^2 ke^{-x}\, dx = k(1 - e^{-2})$ we have $k = \frac{1}{1-e^{-2}}$.*
2. $\mathbb{P}(X \leq x) = \int_0^x ke^{-t}\, dt = k[e^{-t}]_0^x = k(1 - e^{-x}) = \frac{1-e^{-x}}{1-e^{-2}}$
3. *Evaluate the previous answer at $x = 0.5$ getting* 0.455.

Exam Note 13.5. *It needs to be clear in your minds that there are many similarities between discrete and continuous random variables. The key differences are that the probability of a single point value for a continuous random variable is 0 and that a spot value $f_X(x)$ of a p.d.f. can exceed 1 so long as it still integrates to 1 over the entire range.*

13.2 Exercises

Exercise 13.6. *Let Z be a random variable with probability density function*

$$f_Z(z) = -\frac{3}{2}z^2 + \frac{3}{2},$$

for $z \in [0, 1]$ and 0 otherwise.

1. *Evaluate the cumulative distribution function of Z.*
2. *Calculate $P(Z > 0.5)$.*

Exercise 13.7. *Find the value k such that the p.d.f. $f_X(x) = kx(x-4)^2$ is a valid probability density function on the interval $x \in [0, 10]$. Leave your answer as an exact fraction in its simplest form.*

Exercise 13.8. *An accident can occur at any point on a 100 mile long road with equal likelihood. Use the probability density function to calculate $\mathbb{P}(x + c \geq X \geq x)$ such that $1 \geq x + c \geq x \geq 0$.*

Exercise 13.9. *If X has the density function*

$$f_X(x) = \begin{cases} ke^{-5x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find k and $\mathbb{P}(1 \leq X \leq 2)$.

Exercise 13.10. *The probability density function for the continuous random variable X is give by*

$$f_X(x) = \begin{cases} \frac{1}{10} & \text{for } 1 \leq x \leq 11 \\ 0 & \text{otherwise.} \end{cases}$$

1. *Draw the graph and verify that the total area under the curve (above the x-axis) is equal to one.*
2. *Calculate $\mathbb{P}(6 > X > 3)$.*

Exercise 13.11. *Show that $f_X(x) = 5x^4$ for $x \in [0, \infty]$ is a valid probability density function. Sketch the graph of $f_X(x)$ and calculate $\mathbb{P}(X > 2)$.*

Exercise 13.12. *Find the distribution function of the random variable X whose probability density is given by*

$$f_X(x) = \begin{cases} 1/3 & \text{for } x \in (0, 1), \\ 1/3 & \text{for } x \in (2, 4), \\ 0 & \text{otherwise.} \end{cases}$$

Also sketch both the density and distribution functions.

14 Mean and Variance for Continuous Random Variables

Content:

1. The mean of a continuous random variable.
2. The mean of a function of a continuous random variable.
3. The variance of a continuous random variable.
4. Uniform random variables.

Supplementary reading: Freund §4.2 and 4.3, continuous case. Freund exercises 4.7, 4.19.
Supplementary videos: Mean and variance for continuous random variables
Related material: The results of class 10.1 on linear combinations apply here too.

14.1 Class Content

Similarly as to in the discrete case, we can calculate the moments of continuous random variables. In this section we will look at the first moment $\mathbb{E}(X)$ and second moment $\mathbb{E}(X^2)$ as a tool to calculate $\mathrm{Var}(X)$. We will also look at the expected value of functions of random variables $g(X)$.

Definition 14.1. *For a continuous random variable X, the expectation of $g(X)$, for some function g defined on R_X, is defined as*

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

Setting $g(X) = X$ gives

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

and we again use the notation

$$\mathbb{E}(X) = \mu_X,$$

with μ_X referred to as the mean of X.

Variance has the same definition as before, but it is now calculated using integration

$$\begin{aligned} \mathrm{Var}(X) &= E\{(X - \mu_X)^2\} \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx. \end{aligned}$$

Note: for a continuous random variable X, the identity

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

still holds, so, as with discrete random variables, we can calculate $\mathrm{Var}(X)$ by calculating $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$.

All the properties of expectation and variance that we had in the discrete setting still hold. For all random variables:

$$\mathbb{E}(aX \pm bY \pm c) = a\mathbb{E}(X) \pm b\mathbb{E}(Y) \pm c,$$

and, if X and Y are independent,

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X)\mathbb{E}(Y), \\ \mathrm{Var}(aX \pm bY \pm c) &= a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \end{aligned}$$

and if X and Y are not independent then

$$\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \pm 2ab\mathrm{Cov}(X,Y)$$

Example 14.2. *Calculate the expectation and variance of the random variable defined in Example 13.4.*

Solution. $\mathbb{E}(X) = k\int_0^2 xe^{-x}\ dx$. *Integrating by parts we see that* $\mathbb{E}(X) = k([-xe^{-x}]_0^2 + \int_0^2 e^{-x}dx = k(-2e^{-2} + 1 - e^{-2})$ *and substituting for* k *gives* $\frac{1-3e^{-2}}{1-e^{-2}} = 0.687$.

$\mathbb{E}(X^2) = k\int_0^2 x^2e^{-x}\ dx = k([-x^2e^{-x}]_0^2 + [-2xe^{-x}]_0^2 + [-2e^{-x}]_0^2) = k(2 - 10e^{-2}) = \frac{2(1-5e^{-2})}{1-e^{-2}} = 0.748$. *(For this you were required to use integration by parts twice.)*

So $Var(X) = 0.748 - 0.687^2 = 0.276$.

14.2 Exercises

Exercise 14.3. *Let* Z *be a random variable with probability density function*

$$f_Z(z) = -\frac{3}{2}z^2 + \frac{3}{2},$$

for $z \in [0,1]$ *and 0 otherwise. Find the expectation and standard deviation of* Z.

Exercise 14.4. *Find the expected value of the following p.d.f.*

$$f_X(x) = \begin{cases} \frac{x+1}{8} & \textit{for } 4 \geq x \geq 2 \\ 0 & \textit{otherwise} \end{cases}$$

Exercise 14.5. *For a random variable* X *with p.d.f.* $f_X(x) = \frac{1}{2}\sin(x)$ *for* $x \in [0,\pi]$ *and* 0 *elsewhere. Calculate* $\mathbb{E}(X)$.

15 Moment Generating Functions

Content:

1. Moment Generating Functions.

Supplementary reading: Freund §4.5.
Supplementary videos: Moment generating functions.

Up until now you have come across first and second order moments, namely the first moment $\mathbb{E}(X)$ defined as the mean of the random variable X and the second moment about the mean $\mathbb{E}\{(X-\mu_X)^2\}$ which we usually call the variance $\mathrm{Var}(X)$. Covariance and correlation which we discuss later in the course are also second order moments, but of joint distributions.

Both series extend in the obvious way

$$n^{th} \text{ moment} \equiv \mathbb{E}(X^n)$$

and

$$n^{th} \text{ moment about the mean} \equiv \mathbb{E}(X-\mu_X)^n.$$

These moments capture increasingly complex aspects of the random variable's distribution:

- first order $\equiv$ location
- second order $\equiv$ spread
- third order $\equiv$ skewness
- fourth order $\equiv$ peakedness

however orders one and two are by far the most common.

15.1 Class Content

Let X be a random variable. Two important associated numerical quantities are the expected values $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$, from which we can compute the variance, using

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

More generally we might try to find the n**th moment** $\mathbb{E}(X^n)$ for $n \in \mathbb{N}$. For example $\mathbb{E}(X^3)$ and $\mathbb{E}(X^4)$ give information about the shape of the distribution and are used to calculate quantities called the **skewness** and **kurtosis**. We can try to calculate moments directly, but there is also a useful shortcut.

Definition 15.1. *The* ***moment generating function*** *(or mgf) is defined for all $t \in \mathbb{R}$ by:*

$$M_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \sum_{x\in R_X} e^{tx} P_X(x) & \textit{if } X \textit{ is discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \textit{if } X \textit{ is continuous.} \end{cases} \tag{15}$$

We note that

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}(X^k) \tag{16}$$

by way of the Maclaurin series expansion.

Theorem 15.2.

$$\frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} = \mathbb{E}(X^n).$$

Proof.

$$\begin{aligned}
\frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} &= \frac{d^n}{dt^n} \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}(X^k)\bigg|_{t=0} \\
&= \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{k!} \left(\frac{d^n}{dt^n} t^k\right)\bigg|_{t=0} \\
&= \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{k!} \left(k(k-1)\dots(k-(n-1))t^{k-n}\right)\bigg|_{t=0} \\
&= \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{(k-n)!} t^{k-n}\bigg|_{t=0}.
\end{aligned}$$

At $t = 0$, $t^{k-n} = 0$ for all k except $k = n$ because $0^0 = 1$. The reasoning for this is as follows: For any value x such that $-1 < x < 1$ can write Y as the following convergent sum

$$\begin{aligned}
Y &= x^0 + x^1 + x^2 + \dots \\
xY &= x^1 + x^2 + x^3 + \dots \\
Y - xY &= x^0 + x^1 - x^1 + x^2 - x^2 + x^3 - x^3 + \dots = x^0 = 1 \\
Y &= \frac{1}{1-x}.
\end{aligned}$$

Therefore we have

$$\frac{1}{1-x} = x^0 + x^1 + x^2 + \dots$$

As this is defined for $-1 < x < 1$ we can set $x = 0$ to conclude

$$\frac{1}{1-0} = 0^0 + 0^1 + 0^2 + \dots = 0^0 + 0 + 0 + \dots = 0^0.$$

Therefore the mgf calculation becomes

$$\begin{aligned}
\frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} &= \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{(k-n)!} t^{k-n}\bigg|_{t=0} \\
&= \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{(k-n)!} 0^{k-n} \\
&= \frac{\mathbb{E}(X^n)}{(n-n)!} 0^0 \\
&= \mathbb{E}(X^n).
\end{aligned}$$

□

As a demonstration of Theorem 15.2 we differentiate (15) to get

$$\frac{d}{dt} M_X(t) = \mathbb{E}(X e^{tX}),$$

and so

$$\frac{d}{dt} M_X(t)\bigg|_{t=0} = \mathbb{E}(X).$$

Differentiating (15) twice gives

$$\frac{d^2}{dt^2} M_X(t) = \mathbb{E}(X^2 e^{tX}),$$

and hence

$$\frac{d^2}{dt^2} M_X(t)\bigg|_{t=0} = \mathbb{E}(X^2).$$

Theorem 15.3. *1. If X and Y are independent, then for all $t \in \mathbb{R}$*

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

2. If a and b are constants, then

(a)

$$M_{X+a}(t) = \mathbb{E}(e^{(X+a)t}) = e^{at}M_X(t).$$

(b)

$$M_{bX}(t) = \mathbb{E}(e^{bXt}) = M_X(bt).$$

(c)

$$M_{\frac{X+a}{b}}(t) = \mathbb{E}\left(e^{\left(\frac{X+a}{b}\right)t}\right) = e^{\frac{a}{b}t}M_X\left(\frac{t}{b}\right).$$

It follows (by using mathematical induction) that if $S(n) = X_1 + X_2 + \cdots + X_n$ is a sum of i.i.d random variables having common mgf M_X then

$$M_{S(n)}(t) = M_X(t)^n. \tag{17}$$

Example 15.4. *For a random variable X with p.d.f. $f_X(x) = \lambda e^{-x\lambda}$ for $x \geq 0$ such that $\lambda \in \mathbb{R}$ and $f_X(x) = 0$ otherwise. Show that*

$$\mathbb{E}(X^n) = \frac{n!}{\lambda^n}.$$

Solution.

$$\begin{aligned}
M_X(t) &= \int_0^\infty e^{xt} f_X(x)dx \\
&= \int_0^\infty e^{xt}\lambda e^{-x\lambda}dx \\
&= \lambda \int_0^\infty e^{-x(\lambda-t)}dx \\
&= \lambda \left[\frac{1}{-(\lambda-t)}e^{-x(\lambda-t)}\right]_0^\infty = \frac{\lambda}{\lambda-t}, \text{ if } t-\lambda<0.
\end{aligned}$$

We can now find the closed form for the nth moment if $t - \lambda < 0$.

$$\begin{aligned}
\mathbb{E}(X^n) &= \left.\frac{d^n}{dt^n}M_X(t)\right|_{t=0} \\
&= \left.\frac{d^n}{dt^n}\frac{\lambda}{\lambda-t}\right|_{t=0} \\
&= \left.\frac{\lambda n!}{(\lambda-t)^{n+1}}\right|_{t=0}, \text{ by noticing a pattern when repeatedly differentiating} \\
&= \frac{n!}{\lambda^n}
\end{aligned}$$

as required.

15.2 Exercises

Exercise 15.5. *Let $X \sim Bin(n, p)$. Starting from the probability mass function, find the moment generating function of X and use this technique to confirm that X is the sum of n i.i.d. random variables each of which has a $Bernoulli(p)$ distribution.*

Exercise 15.6. *If X is a random variable (discrete or continuous), what can you say about $M_X(0)$?*

Exercise 15.7. *Find the moment generating function for the r.v. with p.m.f. $P_X(x) = 2\left(\frac{1}{3}\right)^x$ for x=1,2,... and zero otherwise. Verify that $\mathbb{E}(X) = \frac{3}{2}$.*

16 The Normal Distribution

Content:

1. The normal distribution.
2. Sum of independent normal random variables.
3. Standardisation.
4. Calculating probabilities using the standard normal.

Supplementary reading: Freund §6.5, up to Example 6.4. Freund exercise 6.71. Document on sums of independent normal random variables.
Supplementary videos: (The normal distribution).

Notes:

1. The notation for a normal random variable X with mean μ and variance σ^2 is $X \sim N(\mu, \sigma^2)$. Note that it is the *variance* that appears here, not the standard deviation σ. If Y is a normal random variable with mean 12 and standard deviation 4, we write $Y \sim N(12, 16)$. Often, when a problem includes more than one variable we use subscripts to represent the different variables. For example, for X we have $\mu_X = \mathbb{E}(X)$ and $\sigma_Y = \mathrm{Var}(Y)$.
2. To calculate $\mathbb{P}(X \leq \mathtt{x})$ where $X \sim N(0,1)$ in R we can use the following command: `pnorm(x)`
3. The above command can be easily generalised to calculate $\mathbb{P}(X \leq \mathtt{x})$ where $X \sim N(\mathtt{m}, \mathtt{s}^2)$ in R as follows: `pnorm(x, mean = m, sd = s)`
4. To find the value of x such that $\mathbb{P}(X \leq x) = \mathtt{p}$ where $X \sim N(0,1)$ in R we can use the following command: `qnorm(p)`
5. The above command can also be easily generalised to find the value of x such that $\mathbb{P}(X \leq x) = \mathtt{p}$ where $X \sim N(\mathtt{m}, \mathtt{s}^2)$ in R as follows: `qnorm(p, mean = m, sd = s)`
6. To generate a random sample of n observations from the $N(0,1)$ distribution in R we can use the following command: `rnorm(n)`
7. If we want to generate a random sample of n observations from the $N(\mathtt{m}, \mathtt{s}^2)$ distribution, then we must also specify the mean and standard deviation of the distribution in R as follows: `rnorm(n, mean = m, sd = s)`

16.1 Class Content

16.1.1 The standard normal distribution: $N(0,1)$

The normal distribution is very important distribution in both probability and statistics. Before studying it, we first introduce the **Gaussian integral**:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \tag{18}$$

This is used to establish the correct value of the normalizing constant for the normal distribution (i.e. the value of k to ensure the distribution integrates to 1.)

We first define the **standard** normal distribution, before considering the more general case.

Definition 16.1. *If a random variable Z has a **standard normal distribution**, then its probability density function is given by*

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

We write

$$Z \sim N(0,1),$$

to mean "Z has the standard normal distribution."

Theorem 16.2. *(Expectation and variance of a standard normal random variable)*
If $Z \sim N(0,1)$, then

$$\begin{aligned} \mathbb{E}(Z) &= 0 \\ \mathrm{Var}(Z) &= 1. \end{aligned}$$

16.1.2 The cumulative distribution function of $N(0,1)$

The c.d.f. is

$$F_Z(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

However, we can't evaluate this integral analytically, and have to use numerical methods. There are various statistical tables available that give the value of $F_Z(z)$ for different z, but these have now largely been superseded by modern computing packages, and we will see how to calculate $F_Z(z)$ using MatLab and R in Section 16.1.5. The c.d.f. is plotted below (right plot).

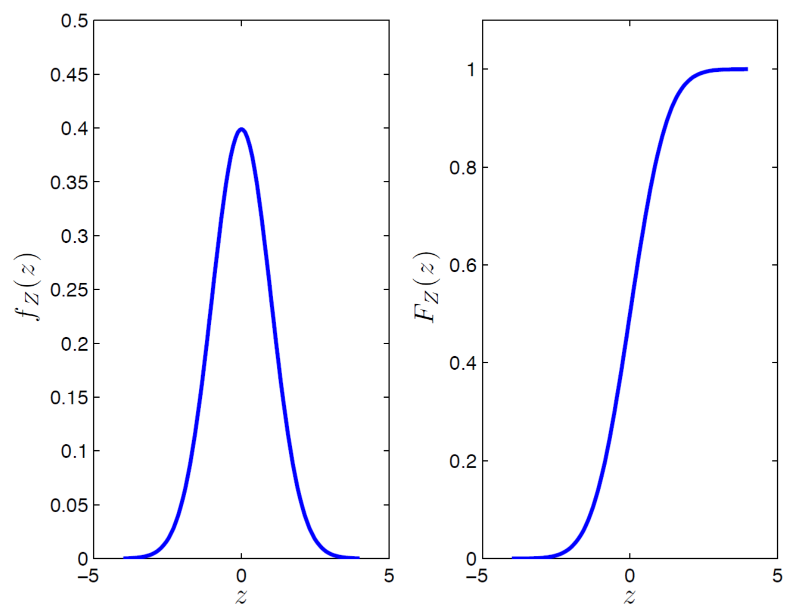

Figure 10: The p.d.f. (left plot) and c.d.f. (right plot) of a $N(0,1)$.

The notation Φ (Phi) is commonly used to represent the c.d.f., and ϕ (phi) to represent the p.d.f.:

$$\begin{aligned} \Phi(z) &= \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \\ \phi(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \end{aligned}$$

Then

$$\frac{d}{dz}\Phi(z) = \phi(z).$$

Theorem 16.3. *(Relationship between $\Phi(z)$ and $\Phi(-z)$.)*

$$\Phi(-z) = 1 - \Phi(z).$$

This can be seen in Figure 11 below.

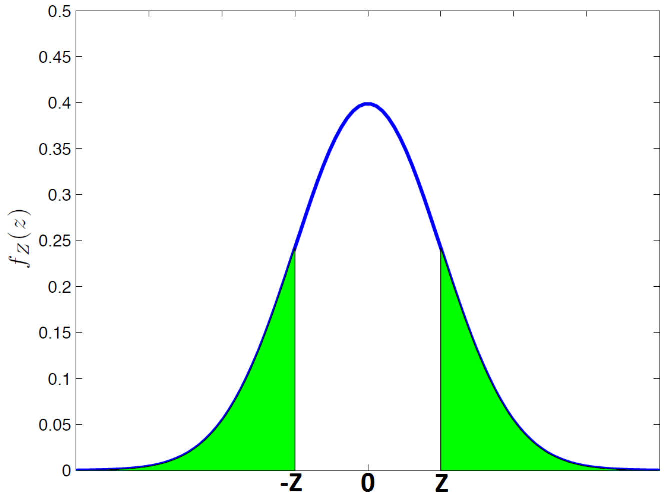

Figure 11: As f_Z is a symmetric about $z = 0$, the area under the curve between $-\infty$ and $-z$ is the same as the area under the curve between z and ∞.

We denote the quantile function by Φ^{-1}. If we want z such that $\mathbb{P}(Z \leq z) = \alpha$, then we write

$$\begin{aligned} \Phi(z) &= \alpha \\ z &= \Phi^{-1}(\alpha). \end{aligned}$$

These quantities are used extensively in statistics.

16.1.3 The general normal distribution $N(\mu, \sigma^2)$

The standard normal distribution is one example of the family of normal distributions. Its mean is 0 and variance is 1, but, in general, normal random variables can have any values for the mean and variance (though variances cannot be negative, of course).

Definition 16.4. *If a random variable X has a **normal distribution** with mean μ and variance σ^2, then its probability density function is given by*

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

for $x \in \mathbb{R}$ and $\sigma > 0$. Normal distributions are used very widely in many situations, for example:

- many physical characteristics of humans and other animals, for example the distribution of heights in a particular group, can be well represented with a normal distribution;
- scientists often assume that 'measurement errors' are normally distributed;
- normal distributions are commonly used in finance to model changes in stock prices.

Indeed this is why the distribution has the name 'normal'! You may also see it referred to as the Gaussian distribution after the famous Mathematician Gauss.

Theorem 16.5. *(Definition of a general normal random variable via transformation of a standard normal random variable)*

Let $Z \sim N(0, 1)$, *and define*

$$X = \mu + \sigma Z.$$

Then $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$ *and*

$$X \sim N(\mu, \sigma^2),$$

It's worth noting that we can invert this relationship between a standard normal random variable Z and a 'general' normal random variable X.

- given $Z \sim N(0, 1)$, we can obtain $X \sim N(\mu, \sigma^2)$ by transforming Z:

$$X = \mu + \sigma Z.$$

- given $X \sim N(\mu, \sigma^2)$, we can obtain $Z \sim N(0, 1)$ by transforming X:

$$Z = \frac{X - \mu}{\sigma},$$

  and we refer to transforming X to get a standard $N(0, 1)$ random variable as **standardising** X.

Traditionally, we would calculate the c.d.f. of X via standardising and using the $\Phi(z)$ function:

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

where $\Phi(z)$ is given in statistical tables for various values of z. As discussed before, statistical tables have become largely obsolete given computer packages such as MatLab, although the technique of standardising is still computationally useful.

By plotting the density function, we can see the effect of changing the value of μ and σ^2 and so interpret these parameters more easily. If we change μ whilst leaving σ^2 unchanged (Figure 12, top plot), the p.d.f. 'shifts' along the x-axis, but the shape of the p.d.f. is unchanged.

The variance parameter σ^2 determines how 'spread out' the p.d.f. is. If we increase σ^2, whilst leaving μ unchanged (Figure 12, bottom plot), the peak of the p.d.f. is in the same place, but we get a flatter curve. This is to be expected, remembering that random variables with larger variances are more likely to be further away from their expectations.

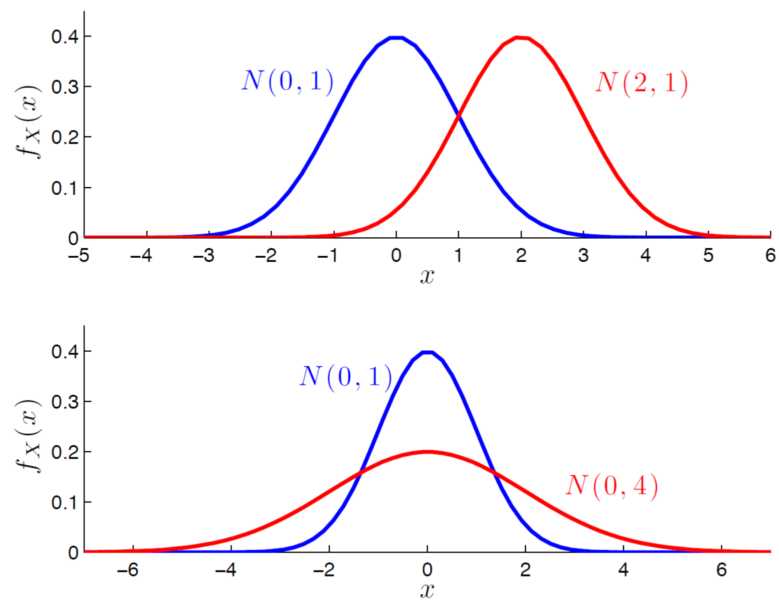

Figure 12: Top plot: pdfs for the $N(0,1)$ and $N(2,1)$ distributions. Bottom plot: pdfs for the $N(0,1)$ and $N(0,4)$ distribution.

16.1.4 The normal distribution in MatLab

MatLab will calculate the p.d.f., c.d.f. and quantile functions, and will also generate normal random variables. **Note that in MatLab, we specify the standard deviation rather than the variance.**

- Calculate the pdf: `normpdf(x,mu,sigma)`
  Example: calculate $f_X(2)$ when $X \sim N(1,4)$.

  ```
  normpdf(2,1,2)
  0.1760327
  ```

- Calculate the cdf: `normcdf(x,mu,sigma)`
  Example: calculate $F_X(-1) = \mathbb{P}(X \leq -1)$ when $X \sim N(1,4)$.

  ```
  normcdf(-1,1,2)
  0.1586553
  ```

- Invert the c.d.f. to find the α quantile: `norminv(alpha,mu,sigma)`
  Example: if $Z \sim N(0,1)$, what value of z satisfies the equation $F_Z(z) = \mathbb{P}(Z \leq z) = 0.95$?

  ```
  norminv(0.95,0,1)
  1.644854
  ```

  Check:
  ```
  normcdf(1.644854,0,1)
  0.95
  ```

16.1.5 The normal distribution in R

R will calculate the p.d.f., c.d.f. and quantile functions, and will also generate normal random variables. **Note that in R, we specify the standard deviation rather than the variance.**

- Calculate the pdf: `dnorm(x,mu,sigma)`
  Example: calculate $f_X(2)$ when $X \sim N(1, 2^2)$.

  ```
  dnorm(2,1,2)
  0.1760327
  ```

- Calculate the cdf: `pnorm(x,mu,sigma)`
  Example: calculate $F_X(-1) = \mathbb{P}(X \leq -1)$ when $X \sim N(1, 4)$.

  ```
  pnorm(-1,1,2)
  0.1586553
  ```

- Invert the c.d.f. to find the α quantile: `qnorm(alpha,mu,sigma)`
  Example: if $Z \sim N(0, 1)$, what value of z satisfies the equation $F_Z(z) = \mathbb{P}(Z \leq z) = 0.95$?

  ```
  qnorm(0.95,0,1)
  1.644854
  ```

  Check:
  ```
  pnorm(1.644854,0,1)
  0.95
  ```

Example 16.6. *(from Ross, 2010).*
An expert witness in a paternity suit testifies that the length, in days, of human gestation is approximately normally distributed, with mean 270 days and standard deviation 10 days. The defendant has proved that he was out of the country during a period between 290 days before the birth of the child and 240 days before the birth of the child, so if he is the father, the gestation period must have either exceeded 290 days, or been shorter than 240 days. How likely is this?

Solution. $X \sim N(270, 100)$*, and we want* $1 - \mathbb{P}(240 \leq X \leq 290)$*. In MatLab* `1-(normcdf(290,270,10)-normcdf(240,270,10))` *gives* 0.0241.

16.1.6 The two-σ rule

For a standard normal random variable Z,

$$\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95.$$

Since $\mathbb{E}(Z) = 0$ and $\text{Var}(Z) = 1$, the probability of Z being within two standard deviations of its mean value is approximately 0.95 (ie $\mathbb{P}(-2 \leq Z \leq 2) = 0.9545$ to 4 d.p.). We illustrate this in Figure 13.

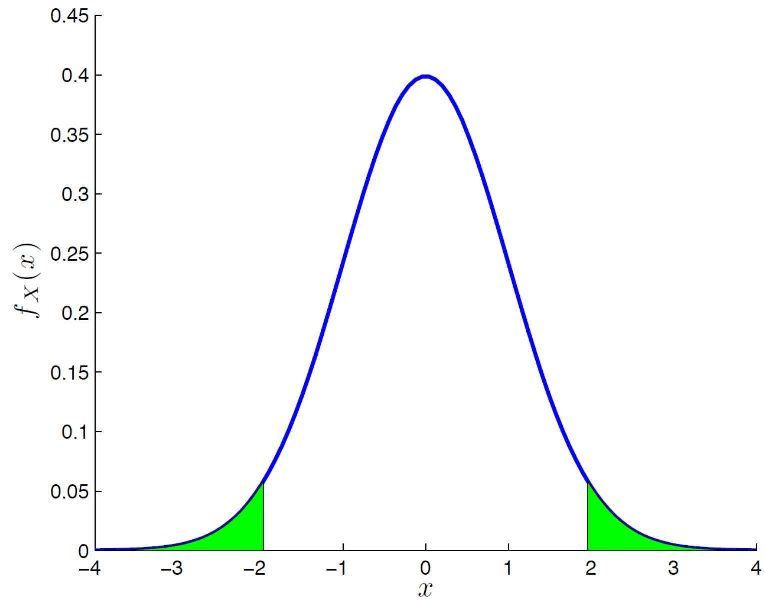

Figure 13: The p.d.f. of a $N(0,1)$ random variable (mean 0, standard deviation 1). There is a 95% probability that a normal random variable will lie within 1.96 standard deviations of its mean.

If we now consider any normal random variable $X \sim N(\mu, \sigma^2)$, the probability that it will lie within a distance of two standard deviations from its mean is also approximately 0.95. This is straightforward to verify by the usual transformation argument:

$$\begin{aligned}
\mathbb{P}\left(|X-\mu| \leq 1.96\sigma\right) &= \mathbb{P}(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma) \\
&= \mathbb{P}\left(-1.96 \leq \frac{X-\mu}{\sigma} \leq 1.96\right) \\
&= \mathbb{P}(-1.96 \leq Z \leq 1.96) \\
&= 0.95,
\end{aligned}$$

(with $\mathbb{P}\left(|X-\mu| \leq 2\sigma\right) = 0.9545$ to 4 d.p.).

In Statistics, there is a convention of using 0.05 as a threshold for a 'small' probability (more of this later), though the choice of 0.05 is arbitrary. However, the two-σ rule is an easy to remember fact about normal random variables, and can be a useful yardstick in various situations. This reasoning underlies the use of "± 2 standard errors" in many engineering applications, e.g. error bars on graphs.

16.1.7 Sums of independent Normals

From Lecture 10, we know that if two random variables X and Y are independent, then we have

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

and

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

If we know that X and Y both have Normal distributions, then we have the further fact that the distribution of the sum remains Normal, so if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ and X and Y are independent then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. As usual, we can generalise this to more than two random variables:

Theorem: If $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ for each i, then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

For those who are interested in how to prove this, note the following fact about moment generating functions:

If X and Y are independent with moment generating functions $M_X(t)$ and $M_Y(t)$ then $X+Y$ has moment generating function given by the product, $M_{X+Y}(t) = M_X(t)M_Y(t)$. This is because $\mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY})$ (by properties of the exponential function) and so is $\mathbb{E}(e^{tX})\mathbb{E}(e^{tY})$ (by independence). Again, this fact extends to more than two random variables.

It is now straightforward to show, using the moment generating function for the Normal as shown in Theorem 6.6 of Freund, that if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ and X and Y are independent, then $M_{X+Y}(t)$ is the same as a moment generating function of $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$; that this is enough to conclude the result follows from further facts about moment generating functions which appear in §6.6 of Freund (lecture 24). Again, the method generalises to where we have more than two random variables.

16.2 Exercises

Exercise 16.7. *Let $X \sim N(0.5, 0.25)$. Using the output from an MatLab session below, calculate*

1. $\mathbb{P}(X > 1.5)$;
2. $\mathbb{P}(0 < X < 1.5)$.

Note that two of the following four MatLab commands are not relevant!

```
normcdf(2,0,1)
0.9772499
normcdf(1.5,0,1)
0.9331928
normcdf(0,0,1)
0.5
normcdf(1,0,1)
0.8413447
```

Exercise 16.8. *Find the moment generating function the normally distributed random variable $X \sim N(\mu\sigma^2)$.*

17 Distributions in R

Content:

1. Using R for calculations involving standard probability distributions.

Supplementary videos: Distributions in R.

Tasks: In this section we will focus on solving the following problem using R. The weight of men in a particular population can be approximated by a Normal distribution with mean 75kg and standard deviation 4kg.

1. What is the probability that a randomly selected male weighs more than 80kg?
2. What is the shortest interval that contains 50% of the weights in the population?

17.1 Class Content

17.1.1 R commands

For each standard distribution, R offers a family of commands of the form:

- `dnorm` which evaluates the density of the Normal distribution with given mean and standard deviation.
- `pnorm` which evaluates the distribution function of the Normal distribution with given mean and standard deviation.
- `qnorm` which evaluates the quantile function of the Normal distribution with given mean and standard deviation.
- `rnorm` which offers random generation from the Normal distribution with given mean and standard deviation.

but with '`norm`' replaced by the appropriate abbreviation for the distribution in question and the mean and standard deviation replaced by whatever parameters are relevant for that distribution (you can view the possibilities through, e.g., `help(dnorm)`).

For the final function you must declare the sample size required and it is also possible to set the seed for the random number generator, with, e.g., `set.seed(1234)`. This is mainly used so that you can recreate the data set again later, if necessary.

Once you have decided on the relevant command, it is issued with the relevant parameters in a standard order (again check with the help command, though this can be varied if you are careful to specify what order you are in fact following). Some examples based on the standard Normal will illustrate the idea:

- `rnorm(20,0,1)` generates a sample of 20 random values from a standard Normal.
- `pnorm(0.5,0,1)` returns the value of the standard Normal cumulative distribution function at $z = 0.5$.
- `dnorm(0.5,0,1)` returns the value of the standard Normal density function at $z = 0.5$.
- `qnorm(0.5,0,1)` returns z such that $P(Z \leq z) = 0.5$ for the standard Normal.

Note that for the Normal you must give the standard deviation and NOT the variance.

Once you are familiar with these commands, it is possible to embed them in more complex instructions to R. For example: `hist(rnorm(5000,0,1))` will draw the histogram of 5000 randomly generated standard Normal variates.

One needs to understand that R functions can operate on whole vectors of data, which can be set up through the simple device of the R concatenation function $c(\ldots,\ldots)$. This creates a vector of numbers from the list you type (separated by commas). Many R functions will take such vectors as input and return appropriate vectors as output. Now we are in a position to issue instructions such as `dnorm(c(-2,-1,0,1,2),0,1)`.This produces the output

```
[1] 0.05399097 0.24197072 0.39894228 0.24197072 0.05399097
```

which is (ignoring the label [1]), the standard Normal density at each of the 5 points.

Note that these functions will simplify hand probability calculations somewhat. For example, it is not necessary to standardise general Normals to the $N(0,1)$ form to calculate probabilities, or to use symmetry to evaluate quantiles for probabilities less than 0.5. However, one would still need to recall that probabilities sum to one to evaluate $\mathbb{P}(X > k)$ from the cdf $\mathbb{P}(X \leq k)$ which is available.

17.2 Exercises

Exercise 17.1. *Solve the following problems using R. The weight of men in a particular population can be approximated by a Normal distribution with mean 75kg and standard deviation 4kg.*

1. *What is the probability that a randomly selected male weighs more than 80kg?*
2. *What is the shortest interval that contains 50% of the weights in the population?*

18 The Poisson Distribution

Content:

1. The Poisson distribution.
2. Mean and variance.
3. Uses of the Poisson distribution.
4. Poisson approximation to the binomial.
5. Sum of independent Poisson variables.

Supplementary reading: Freund §5.7. Freund exercises 5.33, 5.77.
Supplementary videos: The Poisson distribution.

18.1 Class Content

18.1.1 Poisson

The Poisson distribution is used to represent more general **count** data: the number of times an event occurs in a finite interval in time or space. We use it particularly when individual events can be regarded as 'rare'. Some situations that we might model using a Poisson distribution are as follows.

- The number of arrivals at an Accident & Emergency ward in one night.
- The number of burglaries in a city in a year.
- The number of goals scored by a team in a football match.
- The number of leaks in 1km section of water pipe.

Definition 18.1. *If a random variable X has a* ***Poisson distribution****, with parameter $\lambda > 0$, then its probability mass function is given by*

$$p_X(x) = \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for $x \in \{0, 1, 2, ...\}$ and 0 otherwise. We write

$$X \sim Po(\lambda),$$

to mean "X has a Poisson distribution with rate parameter λ".

The Poisson distribution has a single parameter λ, known as the rate parameter. Shortly, we will show that $\mathbb{E}(X) = \lambda$, so you can interpret λ as the expected number of times the event will occur in a specified time period.

Theorem 18.2. *(Poisson random variable: expectation and variance) If $X \sim Poisson(\lambda)$ then*

$$\mathbb{E}(X) = \lambda = Var(X).$$

Theorem 18.3. *The cumulative distribution function of the Poisson distribution with rate λ is given by*

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leq x) &= \sum_{a=0}^{x} p_X(a) \\ &= \sum_{a=0}^{x} \frac{e^{-\lambda}\lambda^a}{a!}. \end{aligned}$$

As with the binomial distribution, this is tedious to calculate by hand, but easy to calculate using computer packages.

18.1.2 The Poisson distribution in MatLab

- Calculate the p.m.f.: `poisspdf(x,lambda)`
  Example: calculate $p_X(4) = \mathbb{P}(X = 4)$ when $X \sim Po(2)$.
  `poisspdf(4,2)`
  `0.09022352`

- Calculate the c.d.f.: `poisscdf(x,lambda)`
  Example: calculate $F_X(10) = \mathbb{P}(X \leq 10)$ when $X \sim Po(6.5)$.
  `poisscdf(10,6.5)`
  `0.9331612`

- To find the α quantile: `poissinv(alpha,lambda)`
  Example: if $X \sim Po(4.2)$, what is the smallest value of $x \in \{0, 1, \ldots, 15\}$ such that $F_X(x) = \mathbb{P}(X \leq x) \geq 0.8$?
  `poissinv(0.8,4.2)`
  `6`

- To generate 10 random observations from the $Po(20)$ distribution.
  `poissrnd(20,1,10)`
  `15 29 20 30 23 19 20 12 32 20`
  Here the 20 represents the rate parameter, the 1 and 10 represent the number of random observations we want; here we want a matrix with dimensions 1 and 10 of observations (i.e. one set (row) of 10 observations.)

18.1.3 The Poisson distribution in R

- Calculate the p.m.f.: `dpois(x,lambda)`
  Example: calculate $p_X(4) = \mathbb{P}(X = 4)$ when $X \sim Po(2)$.
  `dpois(4,2)`
  `0.09022352`

- Calculate the c.d.f.: `ppois(x,lambda)`
  Example: calculate $F_X(10) = \mathbb{P}(X \leq 10)$ when $X \sim Po(6.5)$.
  `ppois(10,6.5)`
  `0.9331612`

- To find the α quantile: `qpois(alpha,lambda)`
  Example: if $X \sim Po(4.2)$, what is the smallest value of $x \in \{0, 1, \ldots, 15\}$ such that $F_X(x) = \mathbb{P}(X \leq x) \geq 0.8$?
  `qpois(0.8,4.2)`
  `6`

- To generate 10 random observations from the $Po(20)$ distribution.
  `rpois(10,20)`
  `23 20 26 16 20 17 30 22 22 25`
  Here the 20 represents the rate parameter and 10 represents the number of random observations we want.

Example 18.4. *Suppose X, the number of accidents at a road junction in one year has a Poisson distribution with rate parameter 5.*

1. *What are the expectation and variance of X?*
2. *What is the probability that $X = 0$?*

3. *What is the probability that $X \leq 5$?*
4. *What is the probability that $X \geq 1$?*

Solution. $X \sim Poisson(5)$, $\mathbb{E}(X) = Var(X) = 5$, $\mathbb{P}(X = 0) = \frac{e^{-5} \times 5^0}{0!} = 0.00674$, $\mathbb{P}(X \leq 5) = 0.616$ *using* `poisscdf(5,5)` , $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X \leq 0) = 0.9933$ *using* `1-poisscdf(0,5)`. *Note: By definition* $0! = 1$. *Again, take care with weak/strict inequalities.*

18.1.4 Binomial approximation to the Poisson distribution

Consider a random variable X with follows a Binomial distribution with parameters $B\left(n, \frac{\lambda}{n}\right)$ and p.m.f.

$$\begin{aligned}
p_X(x) =& \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
=& \frac{n(n-1)...(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
=& \frac{\frac{n}{n}\left(1 - \frac{1}{n}\right) ... \left(1 - \frac{x-1}{n}\right)}{x!} \lambda^x \left(\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right)^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-x}
\end{aligned}$$

We see that for fixed x and λ,

$$\lim_{n \to \infty} \frac{n}{n} \left(1 - \frac{1}{n}\right) ... \left(1 - \frac{x-1}{n}\right) \to 0$$

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} \to 1$$

$$\lim_{n \to \infty} \left(\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right)^{-\lambda} \to e$$

the limiting distibution becomes

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, ...$$

18.2 Exercises

Exercise 18.5. *Defects occur along an undersea cable in accordance with a Poisson process of rate* $\lambda = 0.1$ *per mile.*

1. *What is the probability that no defects occur in the first two miles of cable ?*
2. *What is the probability that there is one defect in the first two miles and three in the first eight ?*

Exercise 18.6. *Sales of a household article are assumed to occur as a Poisson process with rate 5 per month. How long must the dealer wait to be at least 99% sure of selling at least 150?*

Exercise 18.7. *According to data collected by the British Geological Survey, earthquakes of magnitude between 3 and 3.9 on the Richter scale occur in the UK on average (mean) 3 times per year. Suppose we choose to model the number of such earthquakes occurring next year with a Poisson distribution. Suggest a suitable rate parameter for the Poisson distribution. Using this parameter value,*

1. *what is the probability that there will be precisely 3 such earthquakes next year?*
2. *What is the probability that there will be at least 2 such earthquakes next year?*

Exercise 18.8. *Let* $X \sim Poisson(5)$. *If it is known that* $4 \leq X \leq 6$, *what is the probability that* $X = 5$? *Justify your answer.*

Exercise 18.9. *Consider the observation that an average of 5 buses arrive at a bus stop in a 30 minute time frame. Let* X *denote the number of buses that arrives at the stop.*

1. *Write down the correct distribution to model the above situation; please define your notation carefully.*
2. *Find the probability we get exactly 4 buses in half an hour.*
3. *The probability more than 2 buses arrive at the bus stop in a given thirty minute window.*

Exercise 18.10. *Prove that the moment generating function of a Poisson distributed random variable X is given by*

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Exercise 18.11. *Prove that the expectation of a Poisson distributed random variable X is given by*

$$\mathbb{E}(X) = \lambda.$$

19 The Gamma, Uniform and Exponential Distributions

Content:

1. The Uniform distribution.
2. The Exponential distribution.
3. The Gamma distribution.
4. Memorylessness.
5. Examples of the exponential distribution in applications.

Supplementary reading: Freund §6.2-6.3, up to Example 6.1. Freund exercise 6.16
Supplementary video: Uniform, The exponential distribution, and Gamma.

Notes:

1. There are different parametrisations for both the Gamma and Exponential distributions. For the Exponential, a common parametrisation is to say that $X \sim Exp(\lambda)$ if the p.d.f. is $f(x) = \lambda e^{-\lambda x}$ for $x > 0$, which corresponds to Freund's notation with $\theta = 1/\lambda$.
2. Similarly for the Gamma, an alternative parametrisation is to write the p.d.f. (for $x > 0$) as

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},$$

with two parameters α (the same as Freund's α) and λ (equal to Freund's $1/\beta$).

19.1 Class Content

19.1.1 The uniform distribution

The uniform distribution is used to describe a random variable that is constrained to lie in some interval $[a, b]$, but has the same probability of lying in any subinterval of a fixed width. The uniform distribution is an important concept in probability theory.

Definition 19.1. *If a random variable X has a **uniform distribution** over the interval $[a, b]$, then its probability density function is given by*

$$f_X(x) = \frac{1}{b-a},$$

for $x \in [a, b]$, and 0 otherwise. We write

$$X \sim U[a, b],$$

to mean "X has a uniform distribution over the interval $[a, b]$."

Theorem 19.2. *(Cumulative distribution function of a uniform random variable) If $X \sim U[a, b]$, then for $x \in [a, b]$*

$$F_X(x) = \frac{x-a}{b-a}.$$

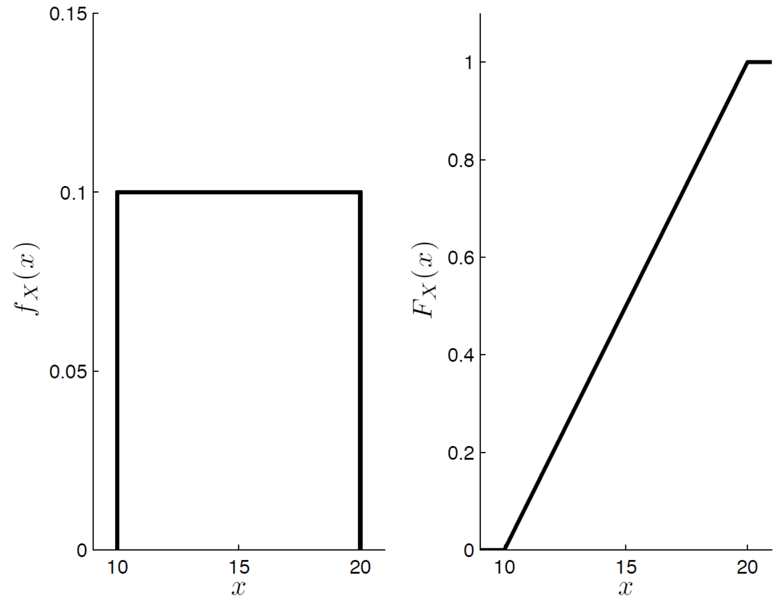

Figure 14: The p.d.f. and c.d.f. of a Uniform distributed random variable $X \sim U[10, 20]$.

Theorem 19.3. *(Expectation and variance of a uniform random variable) If $X \sim U[a,b]$, then*

$$\begin{aligned} \mathbb{E}(X) &= \mu_X = \frac{a+b}{2}, \\ Var(X) &= \sigma_X^2 = \frac{(b-a)^2}{12}. \end{aligned}$$

Example 19.4. *Let $X \sim U[-1,1]$. Calculate $\mathbb{E}(X)$, $Var(X)$ and $\mathbb{P}(X \leq -0.5|X \leq 0)$.*

Solution. *By theorem,* $\mathbb{E}(X) = 0$, $Var(X) = 1/3$. $\mathbb{P}(X \leq -0.5|X \leq 0)$ $= \frac{\mathbb{P}((X \leq -0.5) \cap (X \leq 0))}{\mathbb{P}(X \leq 0)} = \frac{\mathbb{P}(X \leq -0.5)}{\mathbb{P}(X \leq 0)} = (1/4)/(1/2) = 1/2$.

Exam Note 19.5. *Steps when presented with any and all probability distribution questions:*

1. *If the question is presented as a block of text, extract out the useful information.*
2. *Decide if the problem is discrete or continuous. This can usually be done by asking yourself if the available options the random variable can take has a finite number (such as rolling a rice or picking a card) or if it can be any real number in an interval (such as rainfall being anywhere in $[0, \infty)$.)*
3. *Decided on an appropriate model chosen.*
4. *Write your distribution out with its parameters in; for example $X \sim Bin(n = 10, p = 0.35)$.*
5. *Remember when calculating the c.d.f. from the p.m.f. of a discrete distribution you are summing and to calculate the c.d.f. from the p.d.f. of a continuous function you integrate.*
6. *Use of axioms of probability or determine unknown constants in the p.m.f. or p.d.f.*
7. *Finally, after answering what the question is asking step back and ask yourself if the answer you obtained makes logical sense. Obviously if you have a probability $\notin [0,1]$, then alarms should sound. But also if you calculate rainfall in mm as -0.3, the expected value of a dice roll as 6.2 or your expected waiting time for the next bus is 140 hours you should be having a quick look through your calculations to check for errors.*

19.1.2 The exponential distribution

The exponential distribution is used to represent the "time to an event". Examples of 'experiments' that we might describe using an exponential random variable are

- a patient with heart disease is given a drug, and we observe the time until the patient's next heart attack;
- a new car is bought and we observe how many miles the car is driven before it has its first breakdown.

Definition 19.6. *If a random variable X has an **exponential distribution**, with rate parameter $\lambda > 0$, then its probability density function is given by*

$$f_X(x) = \lambda e^{-\lambda x},$$

for $x \geq 0$, and 0 otherwise. We write

$$X \sim Exp(\lambda),$$

to mean "X has an exponential distribution with rate parameter λ".

Theorem 19.7. *Cumulative distribution function of an exponential random variable*

If $X \sim Exp(\lambda)$, then

$$F_X(x) = 1 - e^{-\lambda x}.$$

We can see that $\lim_{x\to\infty} F_X(x) = 1$ (so that "$F_X(\infty) = 1$"), so that, as required of a p.d.f., the density integrates to 1. We plot both the p.d.f. and c.d.f. of an $Exp(2)$ random variable in Figure 15

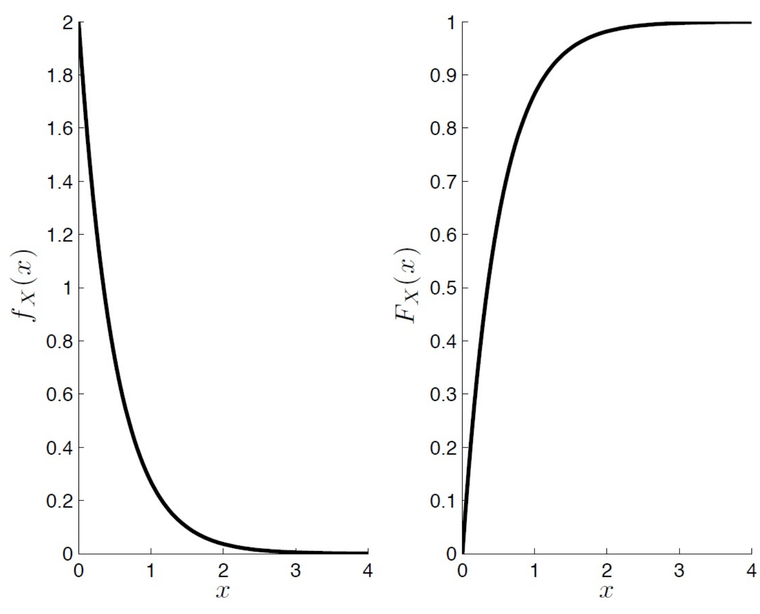

Figure 15: The p.d.f. (left plot) and c.d.f. (right plot) of an exponential random variable X with rate parameter $\lambda = 2$.

Theorem 19.8. *(Expectation and variance of an exponential random variable)*

If $X \sim Exp(\lambda)$, then

$$\mathbb{E}(X) = \frac{1}{\lambda},$$

$$Var(X) = \frac{1}{\lambda^2}.$$

Theorem 19.9. *(The 'lack of memory' or 'memoryless' property of an exponential random variable)*
If $X \sim Exp(\lambda)$, then

$$\mathbb{P}(X > x + a | X > a) = \mathbb{P}(X > x).$$

In other words, exponential random variables have the interesting property that they 'forget' how 'old' they are. If the lifetime of some object has an exponential distribution, and the object survives from time 0 to time a, it will 'carry on' as if it was starting at time 0. There is a generalization of the exponential distribution (the gamma distribution) which allows for aging and so is a more useful model in many practical situations, but we will not study it here.

Example 19.10. *A computer is left running continuously until it first develops a fault. The time until the fault, X, is to be modelled with an exponential distribution. The expected time until the first fault is 100 days.*

1. *If $X \sim Exp(\lambda)$, identify the value of λ. What is the standard deviation of X?*
2. *What is the probability that is the computer develops a fault within the first 100 days?*
3. *If the computer is still working after 100 days, what is the probability that it will still be working after 150 days?*

Solution.
1. *If there is one fault in 100 days, then $\lambda = \frac{1}{100}$. The standard deviation is $\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = 100$.*
2. $\int_0^{100} 0.01 exp(-0.01x)dx = 1 - exp(-1) \approx 0.632$.
3. *Using memoryless property* $\mathbb{P}(X \geq 150 | X \geq 100) = \mathbb{P}(X \geq 50) = 1 - (1 - exp(-50/100)) = exp(-0.5) \approx 0.606$.

19.1.3 The Gamma distribution

The Gamma function is a very flexible function which can be used as a model to fit many real life situations.

Definition 19.11. *We define the Gamma function as*

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz \tag{19}$$

for $\alpha > 0$.

Definition 19.12. *We define the p.d.f. of a Gamma distributed random variable X as*

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

for $\alpha > 0$ and $\lambda > 0$. Here the constant $\frac{\lambda^\alpha}{\Gamma(\alpha)}$ is used as a scaling factor to ensure that $\int_0^\infty f_X(x)dx = 1$.

Some useful facts about the Gamma distribution include:

1. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
2. $\Gamma(\alpha) = (\alpha - 1)!$ (if $\alpha \in \mathbb{N}$).
3. $\Gamma(1) = 1$.

You should prove all of these to exercise your understanding.

Theorem 19.13. *The nth moment about the origin of the Gamma distribution is given by*

$$\mathbb{E}(x^n) = \frac{\Gamma(n + \alpha)}{\lambda^n \Gamma(\alpha)}.$$

Proof.

$$\begin{aligned}
\mathbb{E}(x^n) &= \int_0^\infty x^n \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{n+\alpha-1} e^{-\lambda x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{n+\alpha-1}{\lambda} \int_0^\infty x^{n+\alpha-2} e^{-\lambda x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{n+\alpha-1}{\lambda} \frac{n+\alpha-2}{\lambda} \ldots \frac{\alpha-3}{\lambda} \frac{\alpha-2}{\lambda} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx \\
&= \frac{n+\alpha-1}{\lambda} \frac{n+\alpha-2}{\lambda} \ldots \frac{\alpha-3}{\lambda} \frac{\alpha-2}{\lambda} \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\
&= \frac{n+\alpha-1}{\lambda} \frac{n+\alpha-2}{\lambda} \ldots \frac{\alpha-3}{\lambda} \frac{\alpha-2}{\lambda} (1) \\
&= \frac{\Gamma(n+\alpha)}{\lambda^n \Gamma(\alpha)}.
\end{aligned}$$

□

Theorem 19.14. *The mean and variance of the Gamma distribution are given by*

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}$$

and

$$\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Proof. The proof of these follows directly from Theorem 19.13. □

19.2 Exercises

Exercise 19.15. *Radioactive decay can be modelled using an exponential distribution. It is estimated that the half life of carbon-14 (used for radiocarbon dating) is 5730 years. This means that the probability of one carbon-14 atom decaying into nitrogen-14 within 5730 years is estimated to be 0.5. Find the expected time taken for one carbon-14 atom to decay into nitrogen-14. Define any notation that you introduce, and justify your answer.*

Exercise 19.16. *A cyclist leaves a bicycle chained to some railings, and returns five hours later to find that the bike has been stolen. Define T to be the time in which the bike was stolen, counting in hours from when the bike was left by the owner. Assuming that T has a uniform distribution, calculate*

1. *the mean of T and $\mathbb{E}(T^2)$;*
2. *the probability that T lies between 3 and 4 four hours;*
3. *the 95th percentile of the distribution of T;*
4. *the probability that $T = 2$.*

Exercise 19.17. *A patient with gastroesophageal reflux disease is treated with a new drug to relieve pain from heartburn. Following treatment, the time until the patient next experiences the symptoms is recorded. The doctor treating the patient thinks there is a 50% chance that the patient will stay symptom-free for at least 30 days. For each question, define any notation that you introduce, and justify your answer:*

1. *If the time until recurrence of symptoms is to be modelled using an exponential distribution, find the rate parameter of this distribution based on the doctor's judgement.*
2. *What is the probability the patient will remain symptom-free for at least 60 days?*

3. *If, after 30 days, the patient has remained symptom-free, what is the probability the patient will be symptom-free for at least another 30 days? (Note that the answer differs from that to part (b).)*

Exercise 19.18. *In question 19.17, a doctor now treats* 50 *patients with the drug. Define* T_i *as the time* T_i *until patient* i *next experiences the symptoms. Assuming the times are independent and identically distributed, each with an exponential distribution, find the expectation and variance of*

$$\bar{T} = \frac{1}{50}\sum_{i=1}^{50} T_i.$$

Exercise 19.19. *Show that, with the correct parametrisation fixing* $\lambda > 0$*, the gamma distribution is a generalisation of the exponential distribution.*

20 The Beta Distribution

Content:

1. The Beta distribution.

Supplementary reading: Freund §6.4. Freund exercise 6.27.
Supplementary videos: The beta distribution. Related material:

20.1 Class Content

This short class introduces another flexible probability distribution called the beta distribution.

Definition 20.1. *A random variable X has a beta distribution and is referred to as a beta random variable if and only if its probability density function is given by*

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where both $\alpha > 0$ and $\beta > 0$.

We know that for a probability density function to be valid then

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = 1$$

which is rearranged to make

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{22}$$

which is an important relationship which will be used when calculating moments.

20.2 Exercises

Exercise 20.2. *Show that the nth moment of a beta distributed random variable X is given by*

$$\mathbb{E}(X^n) = \frac{\Gamma(\alpha+\beta)\Gamma(n+\alpha)}{\Gamma(\alpha)\Gamma(n+\alpha+\beta)}.$$

Exercise 20.3. *Prove that a beta distributed random variable X has Variance*

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

21 Solutions To Class Questions

21.1 Solutions To Sample Spaces and Events Questions

Solution (1.10). *1. The sample space is* $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$*. The given event is* $\{HHT, HTH, THH, TTT\}$*.* **Note:** *The method of listing elements is appropriate here because the sample space is small. It is less and less appropriate the bigger the sample space.*

2. *A suitable sample space can be written*

$$\{(0,0), (0,1), (1,0), (0,2), (1,1), (2,0), (0,3), (1,2), (2,1), (3,0), \ldots\}.$$

Note: You cannot list all the elements here because the sample space is infinite. You can use a list style specification by choosing a sensible patterned ordering so it is obvious how it continues. The event is $\{(r, y) \in \mathbb{S} : y > r\}$*.*

3. *S:* $\mathbb{R}$ *(real numbers)*
(Train could be early, but we don't know how early.)

$$\textit{Event: } [10, \infty) = \{x : x \in \mathbb{R},\ x \geq 10\}$$

4.

$$\mathbb{S}\textit{: } \{(x, y) : x, y \in \mathbb{N}_0\}$$

(x denotes A's no. of goals, y denotes B's)

$$\textit{Event: } \{(x, y) : x, y \in \mathbb{N}_0,\ x > y\}$$

5.

$$\mathbb{S}\textit{: } \big\{(x, y, z) : x, y, z \in \mathbb{R}^+\big\}$$

$$\textit{Event: } \big\{(x, y, z) : x, y, z \in \mathbb{R}^+,\ x > y,\ x > z\big\}$$

(x denotes A's rainfall, y denotes B's rainfall, z denotes C's rainfall.)

Solution (1.11). *We have*

$$A \cup \bar{A} = \mathbb{S},$$

and so

$$\mathbb{P}(A \cup \bar{A}) = \mathbb{P}(\mathbb{S}) = 1, \textit{ by axiom 1.}$$

Now, $A \cap \bar{A} = \varnothing$*, and so*

$$\mathbb{P}(A \cup \bar{A}) = \mathbb{P}(A) + \mathbb{P}(\bar{A}), \textit{ by axiom 3.}$$

Hence

$$\begin{aligned} \mathbb{P}(A) + \mathbb{P}(\bar{A}) &= 1, \\ \Rightarrow \mathbb{P}(\bar{A}) &= 1 - \mathbb{P}(A). \end{aligned}$$

Solution (1.12). *Denote the four teams by A, B, C and D, with A and B the two English teams. Team A has 3 possible opponents, and once A's opponent has been decided, the remaining fixture will also be decided. The possible draws are*

A v B and C v D,
A v C and B v D,
A v D and B v C.

If each draw is equally likely, then the probability that A is not drawn to play against B is 2/3, so the probability that the two English teams avoid each other in the semi-finals is 2/3.

Solution (1.13)**.** *For three six sided dice denoted by R for red, B for blue and Y for yellow. The sample same is given by*

$$\mathbb{S} = \{(R, B, Y)|R = 1, 2, ..., 6, B = 1, 2, ..., 6, Y = 1, 2, ..., 6\}. \tag{23}$$

Solution (1.14)**.**

$$\mathbb{S} = \{T^k H|k \in \mathbb{N} \cup \{0\}\}. \tag{24}$$

Solution (1.15)**.** *Diagrams of solutions ommitted. Verify statements using drawings.*

21.2 Solutions To Defining Probabilities Questions

Solution (2.8). *1.* $\{1, -30, 42\}$.

2. $\{-30\}$.

Solution (2.9). *No, as* $0 \leq \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1.05 \nleq 1$, *as axiom 2 does not hold.*

Solution (2.10). *We start by denoting the event A as the event the hand which is dealt is a full house. We will begin by breaking this down into parts by noticing that there are $\binom{4}{3}$ ways to pick the three of a kind and $\binom{4}{2}$ ways to pick the two of a kind. Let us say without loss of generality (without compromising the answer) that we select the member of the triple first. There are thirteen options to select this card (one of A,K,...,3,2,1), this leaves twelve options to select the member of the pair. We multiply all of these components together to get the number of options for different full houses, we will denote this by n:*

$$n = 13 \times \binom{4}{3} \times 12 \times \binom{4}{2}.$$

We also know that the total number of five card hands from a 52 card desk, denoted by N is:

$$N = \binom{52}{5}.$$

Because all of these N options are equally likely, we have that the probability of the event A is given by

$$\mathbb{P}(A) = \frac{n}{N} = \frac{13 \times \binom{4}{3} \times 12 \times \binom{4}{2}}{\binom{52}{5}} = \frac{6}{4165}.$$

Solution (2.11). *Let P, B and H denote the event they eat a pizza, burger and hot dog respectively. We therefore have that*

$$\begin{aligned}\mathbb{P}(P \cup B \cup H) &= \mathbb{P}(P) + \mathbb{P}(B) + \mathbb{P}(H) - \mathbb{P}(P \cap B) - \mathbb{P}(P \cap H) - \mathbb{P}(B \cap H) + \mathbb{P}(P \cap B \cap H) \\ &= 0.44 + 0.24 + 0.21 - 0.08 - 0.11 - 0.07 + 0.03 = 0.66.\end{aligned}$$

21.3 Solutions To Conditional Probability Questions

Solution (3.5). *1. You may have given the following answer: the probability of any single guess being correct is* $1/4$*, so the probability of all 12 guesses being correct is* $(1/4)^{12}$*. This is correct, but how do we justify it? Consider the sample space* S *of sequences of 12 answers. If we label the four possible answers for any single question as A,B,C,D, then one sequence of answers to the 12 questions, one element of* S*, could be written as* $BAACB\ldots$*. There must be* 4^{12} *elements of* S*, with only one resulting in the million pound prize, so if the contest picks an answer at random each time, any element of* S *must have the same chance of being selected, so the probability of winning the prize is* $\frac{1}{4^{12}}$*.*

2. Now the contestant only has to guess correctly 10 times. Eliminating two false answers on one question, the number of possible sequences of answers is 2×4^9*, so the probability of winning is* $\frac{1}{2 \times 4^9}$*.*

Solution (3.6). *Define* N *to be the number of possible entries (the choice of three digits and three characters) to the website, so for example, one entry could be 1,0,4,a,H,7. For each digit, there are 10 possibilities, and for each password character, there are* $26 + 26 + 10 = 62$ *possibilities. For three digits and three characters, we have* $N = 10^3 \times 62^3 = 238328000$ *possibilities. There are* N *possible entries, and one of them will work, so the probability is* $1/N = \frac{1}{238328000}$*.*

Solution (3.7). *1. There are* $\binom{22}{5} = 26334$ *choices of 5 boxes from the 22. If you have chosen the most valuable box, the remaining four chosen boxes can be any from the remaining 21. The number of choices that include the most valuable box is therefore* $\binom{21}{4} = 5985$*, so the probability of choosing the most valuable box is* $\frac{5985}{26334} = \frac{5}{22}$*.*

2. There are $\binom{17}{5} = 6188$ *choices that contain none of the top 5 most valuable boxes, so all of the remaining* $26334 - 6188 = 20146$ *choices must contain at least one of the top 5 most valuable boxes, so the probability this occurs is* $\frac{20146}{26334} = \frac{1439}{1881} = 0.765$*.*

Note: Here, as in many cases it is useful to look at complimentary events - because the compliment may be unique while there are many ways of the original occurring. For example here we could have 1st and 4th or 2nd , 3rd and 5th or just 5th most valuable boxes. Whereas it is sumple to identify 'none of the five most valuable'.

Solution (3.8). *Proof.* We will begin by writing $A \cap B \cap C$ as $D \cap C$ where $D = A \cap B$. We then apply equation (10) twice as follows

$$\begin{aligned}\mathbb{P}(D \cap C) &= [\mathbb{P}(D)]\mathbb{P}(C|D) \\ &= [\mathbb{P}(A \cap B)]\mathbb{P}(C|A \cap B) \\ &= \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B).\end{aligned}$$

□

Solution (3.9). *The solutions are:*

1. $\mathbb{P}(E_{a,b} = 7) = \mathbb{P}(E_{1,6}) + \mathbb{P}(E_{2,5}) + \mathbb{P}(E_{3,4}) + \mathbb{P}(E_{4,3}) + \mathbb{P}(E_{5,2}) + \mathbb{P}(E_{6,1}) = 6 \times \frac{1}{36} = \frac{1}{6}$.
2. $\mathbb{P}(E_{3,b} = 7) = \mathbb{P}(E_{3,4}|A = 3) = \frac{\mathbb{P}(E_{3,4} \cap A=3)}{\mathbb{P}(A=3)} = \frac{1/36}{1/6} = \frac{1}{6}$.
3. $\mathbb{P}(E_{a,b} = 5) = \mathbb{P}(E_{1,4}) + \mathbb{P}(E_{2,3}) + \mathbb{P}(E_{3,2}) + \mathbb{P}(E_{4,1}) = 4 \times \frac{1}{36} = \frac{1}{9}$.
4. $\mathbb{P}(E_{3,b} = 5) = \mathbb{P}(E_{3,2}|A = 3) = \frac{\mathbb{P}(E_{3,2} \cap A=3)}{\mathbb{P}(A=3)} = \frac{1/36}{1/6} = \frac{1}{6}$.

You should take some time to think about why the solutions to 1 and 2 are the same but the answer to 3 and 4 are different. You should also deduce why answers 2 and 4 are the same but 1 and 3 are different.

21.4 Solutions To Independence Questions

Solution (4.13). *1. The total number of possible hands is* $\binom{52}{2}$*=1326. For the two cards to sum to 21, one card must be an ace, and the other must be a 10 or face card. There are four aces, and sixteen cards worth 10, so there are* $4 \times 16 = 64$ *possible hands worth 21. Hence the probability the two cards sum to 21 is* $\frac{64}{1326} = 0.0483$ *(to 3 s.f.).*

2. *There are 49 cards to choose from. The player must draw a 6 or higher, and (noting that one 10 and 6 have already been drawn), there are 30 such cards worth 6 or higher. The probability he will exceed 21 is therefore* $\frac{30}{49} = 0.6122$.

Solution (4.14). *Let* E_i *be the number of sixes we roll, here* $i = 0, 1, 2, 3, 4$. *We wish to calculate* $\mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4)$. *However, this is not the most intuitive way of approaching this problem. Instead we should notice that*

$$\mathbb{P}(E_0) + \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = 1$$

and deduce that

$$1 - \mathbb{P}(E_0) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4).$$

Therefore we only need to calculate $\mathbb{P}(E_0)$ *to solve this problem. We know, because of dice independence, the probability we do not roll a six is* $\frac{5}{6}$ *on each dice which implies that* $\left(\frac{5}{6}\right)^4$ *is the probability we do not roll a six on any of the four dice. Therefore*

$$\mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = 1 - \mathbb{P}(E_0) = 1 - \left(\frac{5}{6}\right)^4 = \frac{671}{1296}.$$

Solution (4.15). *Since we can write* $A = (A \cap B) \cup (A \cap B^c)$ *(draw a Venn diagram to convince yourself of this) and we know that for any two events* E *and* F *it holds that* $\mathbb{P}(E \cup E) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(F \cap E)$.

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}((A \cap B) \cup (A \cap B^c)) \\
&= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) - \mathbb{P}((A \cap B) \cap (A \cap B^c)), \text{ using equation above involving } E \text{ and } F, \\
&= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) - 0, \text{ as } B \text{ and } B^c \text{ do not intersect}, \\
&= \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A \cap B^c) - 0, \text{ as } A \text{ and } B \text{ are independent.}
\end{aligned}$$

Therefore by rearranging we have

$$\begin{aligned}
\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\mathbb{P}(B^c).
\end{aligned}$$

Hence A *and* B^c *are independent.*

Solution (4.16). *For a randomly selected man in the group of 100 let* B *be the event that the selected man has a beard, and* M *be the event that the selected man has a moustache. We have*

1. $P(B) = 35/100$, *as 35 men out of 100 in the group have beards.*
2. *We now consider*

$$P(M|B) = \frac{P(M \cap B)}{P(B)}.$$

   Now $P(M \cap B) = 30/100$, *as 30 men have both beards and moustaches, so* $P(M|B) = 30/35$.

3. *For the events of having a beard and having a moustache to be independent, we must have*

$$P(M|B) = P(M),$$

   but $P(M) = 40/100 \neq 30/35$, *as 40 men have moustaches, so the events are not independent. (Alternatively, you can show that* $P(M \cap B) \neq P(M)P(B)$*).*

21.5 Solutions To Introduction to Random Variables Questions

Solution (5.2). *The solutions given below are not the only way in which you can write the solution. In particular the labels given to the random variables are not unique (sometimes $X, Y, Z...$ and sometimes more intuitive letters.)*

1. *The amount of times a person blinks each day: Let X be the number of times a person blinks each day, here X can take any non negative integer value ($\mathbb{N} \cup \{0\}$).*
2. *The distance two pigs are apart from each other in a one meter square pen: Let P be the distance between the two pigs, here $P \in (0, \sqrt{2})$. [note the curly brackets here represent 'does not include' as the pigs cannot have distance 0 or $\sqrt{2}$ exactly.]*
3. *The roll of a fair six sided dice: Let D be the number rolled on a fair six sided dice, here $D \in \{1, 2, 3, 4, 5, 6\}$.*
4. *The temperature outside in Celsius: Let C be the temperature outside in Celsius, here $C \in \mathbb{R}$. [The Real numbers are all numbers which exist on the number line. We are being a little 'hand-wavey' here with the upper and lower bounds on temperature.]*
5. *The height of sunflowers in a field: Let F be the height of a sunflower, here $F \geq 0$.*
6. *Rainfall in the UK in a year: Let R be the amount of rainfall in the UK each year, here $R \geq 0$.*
7. *Waiting time for a bus: Let W be the waiting time for a bus, here $W \geq 0$.*
8. *Number of darts before the first bullseye: Let D be the number of failed shots before the first bullseye, $D \geq 0$ and $D \in \mathbb{Z}$.*
9. *The difference in height between a person and the average: Let H be the difference in heigh between a person and the average, $H \in \mathbb{R}$. [Again, we are a little vague with the bounds on height.]*
10. *A dogs distance above sea level: Let Z be the dogs distance above sea level, here $Z \in \mathbb{R}$. [Again, we are a little vague with the bounds on height.]*

21.6 Solutions To Discrete Random Variables Questions

Solution (6.13). *1.*

$$\begin{aligned}
S &= \sum_{k=1}^{20} 5 + 3k \\
&= \sum_{k=1}^{20} 5 + 3\sum_{k=1}^{20} k \\
&= 100 + 3 \times \frac{20(1+20)}{2} \\
&= 730
\end{aligned}$$

2.

$$\begin{aligned}
S &= \sum_{k=1}^{n} (a + bk) \\
&= \sum_{k=1}^{n} a + b\sum_{k=1}^{n} k \\
&= na + b\frac{n(1+n)}{2} \\
&= \frac{n}{2}\left(b(1+n) + 2a\right)
\end{aligned}$$

3.

$$S = \frac{n}{2}\left(b(1+n)+2a\right) = \frac{20}{2}\left(3(1+20)+2\times 5\right) = 730.$$

Solution (6.14).

$$F_X(4.5) = F_X(4) = P(X = \{0,1,2,3,4\}) = p_X(0)+p_X(1)+p_X(2)+p_X(3)+p_X(4) = \frac{5}{6}.$$

Solution (6.15). *Yes, because* $g_X(x) \in [0,1]$ *for all* $x \in \{0,1,2,3\}$ *and* $\sum_{k=0}^{3} g_X(k) = 1$.

Solution (6.16). *Given that*

$$\begin{aligned}
P(X=0) &= \frac{5}{6}\times\frac{5}{6}\times\frac{5}{6} = \left(\frac{5}{6}\right)^3 \\
P(X=1) &= \frac{1}{6}\times\frac{5}{6}\times\frac{5}{6} + \frac{5}{6}\times\frac{1}{6}\times\frac{5}{6} + \frac{5}{6}\times\frac{5}{6}\times\frac{1}{6} = 3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 \\
P(X=2) &= \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} + \frac{1}{6}\times\frac{5}{6}\times\frac{1}{6} + \frac{5}{6}\times\frac{1}{6}\times\frac{1}{6} = 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) \\
P(X=3) &= \frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = \left(\frac{1}{6}\right)^3.
\end{aligned}$$

Hence the distribution function is given by

$$F_X(x) = \begin{cases} 0, & \text{for } x \in (-\infty, 0), \\ \left(\frac{5}{6}\right)^3, & \text{for } x \in [0,1), \\ \left(\frac{5}{6}\right)^3 + 3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2, & \text{for } x \in [1,2), \\ \left(\frac{5}{6}\right)^3 + 3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right), & \text{for } x \in [2,3), \\ 1, & \text{for } x \in [3,\infty). \end{cases} \tag{25}$$

21.7 Solutions To Expectation Questions

Solution (7.5). *Suppose your registration number is 110258441, so the list of digits is 1,1,0,2,5,8,4,4,1,5. Note that this solution is given int he context of the question, your answer will be different due to your unique registration number. Then $p_X(x)$ is the number of times the digits occurs in the list, divided by 10, and*

$$F_X(x) = \sum_{n=0}^{x} p_X(n).$$

We have

x	$p_X(x)$	$F_X(x)$
0	0.1	0.1
1	0.3	0.4
2	0.1	0.5
3	0.0	0.5
4	0.2	0.7
5	0.2	0.9
6	0.0	0.9
7	0.0	0.9
8	0.1	1.0
9	0.0	1.0
10	0.0	1.0

To calculate the expected value of X, we first obtain

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=0}^{9} x p_X(x)\\ &= 0 \times 0.1 + 1 \times 0.3 + 2 \times 0.1 + 4 \times 0.2 + 5 \times 0.2 + 8 \times 0.1\\ &= 3.1.\end{aligned}$$

Solution (7.6). *We know that the sample space for a fair dice roll is $\{1, 2, ...6\}$ with each of the options equally likely to show (i.e. $P_X(x) = \frac{1}{6}$ for all $x \in \{1, 2, ..., 6\}$). Therefore if X is the expected value shown on the dice, then*

$$\mathbb{E}(X) = \sum_{x=1}^{6} x \times \frac{1}{6} = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6} = 3.5.$$

Solution (7.7).

$$\mathbb{E}(X) = \sum_{x=0}^{3} x P_X(x) = \sum_{x=0}^{3} x \frac{x+1}{10} = 0 \times \frac{0+1}{10} + 1 \times \frac{1+1}{10} + 2 \times \frac{2+1}{10} + 3 \times \frac{3+1}{10} = 2.$$

Solution (7.8). *If you are working for a casino in games design, your aim is to ensure the house turns a profit off each game. Therefore you would like the expected gain for each player to be negative. For example, if 100 people play your game at £1 each, you would hope that the expected payout (customer win) to be less than £100.*

Solution (7.9). *We will first devise the p.m.f. for this experiment. We have $n = 10$ tosses each with probability $p = 0.5$, we also know that there are $\binom{10}{x}$ ways of having x successes. Therefore we have the p.m.f. is*

$$P_X(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}.$$

We know that

$$
\begin{aligned}
\mathbb{E}(X) = &\sum_{x=0}^{10} x\binom{10}{x}(0.5)^x(0.5)^{10-x} = 0.5^{10}\sum_{x=0}^{10} x\binom{10}{x} \\
=&0.5^{10}\left(0\binom{10}{0}+1\binom{10}{1}+2\binom{10}{2}+\cdots+8\binom{10}{8}+9\binom{10}{9}+10\binom{10}{10}\right) \\
=&0.5^{10}(0\times 1+1\times 10+2\times 45+\cdots+8\times 45+9\times 10+10\times 1) \\
=&0.5^{10}\,(5120) = 5.
\end{aligned}
$$

We will now calculate

$$
\begin{aligned}
\mathbb{P}(X=\mathbb{E}(X)) =&\mathbb{P}(X=5) \\
=&\binom{10}{5}(0.5)^5(0.5)^{10-5} \\
=&252\times\frac{1}{2^{10}}\approx 0.24609375.
\end{aligned}
$$

What we see here is that though the expected number of heads is five, there is still only probability 0.25 *that we toss five heads. This notion becomes more noticeable when the number of tosses* n *in the experiment increases. For example, if we toss* 100 *coins, the expected value is* 50 *but the probability of tossing* 50 *heads is approximately* 0.0796*.*

We will examine a more efficient way of solving problems like this in later topics.

21.8 Solutions To The Standard Deviation and Variance Questions

Solution (8.12)**.** $\mathbb{E}(X) = \sum_{x=0}^{3} x \frac{x+1}{10} = 0 \times \frac{1}{10} + 1 \times \frac{2}{10} + 2 \times \frac{3}{10} + 3 \times \frac{4}{10} = 2.$

Solution (8.13)**.** $\mathbb{E}(X^2) = \sum_{x=0}^{3} x^2 \frac{x+1}{10} = 0^2 \times \frac{1}{10} + 1^2 \times \frac{2}{10} + 2^2 \times \frac{3}{10} + 3^2 \times \frac{4}{10} = 5.$

Solution (8.14)**.** $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 5 - 2^2 = 1.$

Solution (8.15)**.** *We know that* $\mathrm{Var}(X) = \mathbb{E}\left((X - \mu_X)^2\right)$*. It is clear that* $(X - \mu_X)^2 \geq 0$ *and* $P_X(x) \geq 0$ *by the axioms of probability. Therefore each term of*

$$\mathbb{E}((X - \mu_X)^2) = \sum_x (x - \mu_X)^2 P_X(x)$$

is non negative, therefore $\mathrm{Var}(X) \geq 0$*.*

Solution (8.16)**.**

1. $\mathbb{E}(X) = 0 \times 0.25 + 1 \times 0.5 + 4 \times 0.25 = 1.5$
2. $\mathbb{E}(X^2) = 0^2 \times 0.25 + 1^2 \times 0.5 + 4^2 \times 0.25 = 4.5$, $\mathrm{Var}(X) = E(X^2) - E(X)^2 = 4.5 - 1.5^2 = 2.25$, $\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{9/4} = 3/2.$

Solution (8.17)**.** *Proof omitted, see Freund or start a discussion board thread.*

21.9 Solutions To Independence of Random Variables Questions

Solution (9.5). *This is*

$$\mathbb{P}(X_1 \in \{3,6\} \cap X_2 \in \{3,6\}) = \mathbb{P}(X_1 \in \{3,6\})\mathbb{P}(X_2 \in \{3,6\}) = \frac{2}{8} \times \frac{2}{6} = \frac{1}{12}.$$

Solution (9.6). *Let X_1, X_2, X_3 be the scores on the three dice. As they are independent*

$$\mathbb{E}(Y) = \mathbb{E}(X_1 X_2 X_3) = \mathbb{E}(X_1)\mathbb{E}(X_2)\mathbb{E}(X_3) = \left(\frac{7}{2}\right)^3 = \frac{343}{8} = 42.875.$$

21.10 Solutions To Mean and Variance of Linear Combinations Questions

Solution (10.6)**.** *In line 2, it is not true that* $\mathbb{E}(X^2) = (\mathbb{E}(X))^2$*, since* $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$*, and we know that* $Var(X) \neq 0$*. The correct derivation is*

$$\begin{aligned}
\mathbb{E}(Y) = E\{(X+1)^2\} &= \mathbb{E}(X^2 + 2X + 1) \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(X) + 1 \\
&= Var(X) + \mathbb{E}(X)^2 + 2\mathbb{E}(X) + 1 \\
&= 10 + 3^2 + 2 \times 3 + 1 \\
&= 26.
\end{aligned}$$

Solution (10.7)**.** *The total spent in the two shops is £*$(14X + 25Y)$*. So the mean total spent is £*$(14\mathbb{E}(X) + 25\mathbb{E}(Y))$ *=£*$(14 \times 2 + 25 \times 1)$ *=£53.*

Solution (10.8)**.** *It depends on the details. For example, if the shops are next door a high number of visits to shop A may make a high number of visits to B more likely as well, in which case* X *and* Y *would not be independent. Alternatively, if the shops are alternatives for the same goods the shopper may be less likely to visit B as well as A in the same week, so a high number of visits to A would make a high number of visits to B less likely; again* X *and* Y *would not be independent. In other circumstances, though,* X *and* Y *could be independent.*

Solution (10.9)**.** *Let* X *be the income and* Y *be the amount spent. We want* $\mathbb{E}(X - Y)$ *and* $\mathrm{Var}(X - Y)$*. We have* $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 10$ *(pounds). Because of the assumption of independence,* $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 30^2 + 40^2 = 2500$*, so the standard deviation is £50.*

Solution (10.10)**.** *Let* X_1 *be your profit from the first scratchcard, and* X_2 *be your profit from the second, both measured in pounds. Your overall profit is then* $X_1 + X_2$*. As* X_1 *and* X_2 *are independent,* $\mathbb{E}(X_1) = \mathbb{E}(X_2) = -0.2$ *and* $Var(X_1) = Var(X_2) = 11.56$*. Hence* $\mathbb{E}(X_1 + X_2) = -0.4$*; as we assumed the prizes to be independent we know* $Var(X_1 + X_2) = 11.56 + 11.56 = 23.12$*. Hence the standard deviation of* $X_1 + X_2$ *is* $\sqrt{23.12} = 4.81$*.*

If we define Y *to be your prize, with range* $R_Y = \{0, 2, 10, 100\}$*, you will find that* $\mathbb{E}(Y) = 0.8$*, and* $Var(Y) = 11.56$*. Can you explain this?*

21.11 Solutions To Bernoulli Trials and the Binomial Distribution Questions

Solution (11.9). *Let X be the number of questions that you get correct. If you pick one answer at random each time, then we have $X \sim Bin(10, 0.25)$, and so*

$$\mathbb{P}(X = 6) = \binom{10}{6}(0.25)^6(0.75)^4 = 0.016 \text{ (to 3 d.p.)}$$

Solution (11.10).

1. *Let X be the number of people who say they prefer the free-range chicken. Then, under the assumption each person is equally likely to choose either variety, we may suppose that $X \sim Bin(10, 0.5)$.*
2. (a) *We require*

$$\begin{aligned}\mathbb{P}(X \neq 5) &= 1 - P(X = 5)\\ &= 1 - \binom{10}{5}(0.5)^5(1-0.5)^{10-5}\\ &= 0.754 \text{ (to 3 d.p.)}.\end{aligned}$$

(b) *We require*

$$\begin{aligned}\mathbb{P}(X \leq 8) &= 1 - P(X = 9) - P(X = 10)\\ &= 1 - \binom{10}{9}(0.5)^9(1-0.5)^{10-9} - \binom{10}{10}(0.5)^{10}(1-0.5)^{10-10}\\ &= 0.989 \text{ (to 3 d.p.)}.\end{aligned}$$

Solution (11.11).

1. *Let F be the event that the item is faulty, and D be the event that the item is declared to be faulty. The probabilities we have been given in the question are*

$$\begin{aligned}\mathbb{P}(F) &= \frac{1}{200} = 0.005\\ \mathbb{P}(D|F) &= 0.9\\ \mathbb{P}(D|\bar{F}) &= 0.05\end{aligned}$$

Using the law of total probability,

$$\mathbb{P}(D) = \mathbb{P}(D|F)\mathbb{P}(F) + \mathbb{P}(D|\bar{F})\mathbb{P}(\bar{F}) = 0.9 \times 0.005 + 0.05 \times 0.995 = 0.05425.$$

2. *Let X be the number of items declared faulty in a batch of 10. Assuming independence between items, $X \sim Bin(10, 0.05425)$, so*

$$\mathbb{P}(X = 2) = \binom{10}{2} \times 0.05425^2 \times (1 - 0.05425)^8 = 0.085 \text{ (to 3 d.p.)}.$$

Solution (11.12). *For $n = 10$ balls, each with probability $p = \frac{1}{3}$ of successfully scoring. Let X be the number of goals: $X \sim Bin(10, \frac{1}{3})$.*

Solution (11.13). *Let X be the number of sixes rolled on one dice. We see that X follows a Binomial distribution with parameters $X \sim B(100, 1/6)$. Therefore $\mathbb{E}(X) = np = 100/6$ and $\mathrm{Var}(X) = 100(1/6)(5/6) = 500/36 = 125/9$.*

Solution (11.14). *Proof ommitted, start a discussion board thread if you are stuck on this problem.*

21.12 Solutions To The Geometric, Negative Binomial and Hypergeometric Distributions Questions

Solution (12.8). $P_X(1) + P_X(2) + P_X(3) = p + p(1-p)^2 + p(1-p)^3 = 0.1 + 0.1(0.9) + 0.1(0.9)^2 = 0.271.$

Solution (12.9). *We are trying to find an expression for*

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} xp(1-p)^{x-1}.$$

Let us first consider

$$S(p) = \sum_{x=1}^{\infty} (1-p)^x.$$

We can calculate $S(p)$ *exactly as it is a geometric series with common ratio satisfying* $|1-p| < 1$*. Therefore* $S(p) = \frac{1}{p} - 1$*. We can see a relationship between* $S(p)$ *and the original summation whereby*

$$\mathbb{E}(X) = -p\frac{d}{dp}S(p) = -p\frac{d}{dp}\left(\sum_{x=1}^{\infty}(1-p)^x\right) = -p\frac{d}{dp}\left(\frac{1}{p} - 1\right) = \frac{1}{p}$$

as required.

Solution (12.10). $1 - (P_X(1) + P_X(2) + P_X(3)) = 1 - (0.1 + 0.1(0.9) + 0.1(0.9)^2) = 0.729.$

Solution (12.11). *Let* X *be the number of red balls we select from the bag such that* $X \sim HyperGeo(3, 15, 10)$*.*

$$\mathbb{P}(X \le 1) = P_X(0) + P_X(1) = \frac{\binom{10}{0}\binom{5}{3}}{\binom{15}{3}} + \frac{\binom{9}{1}\binom{4}{2}}{\binom{15}{3}} = \frac{2}{91} + \frac{20}{91} = \frac{22}{91}.$$

Solution (12.12). *This is a negative binomial problem which can be solved either by exhaustion (without too much effort) or as follows:*

$$1 - \mathbb{P}(X = 5) = 1 - \binom{4}{2}0.6^3(0.4)^2 = 0.79264.$$

21.13 Solutions To Continuous Random Variables Questions

Solution (13.6). *1. Given the density function $f_Z(z) = -\frac{3}{2}z^2 + \frac{3}{2}$ for $z \in [0,1]$ and 0 otherwise, we have*

$$\begin{aligned}\mathbb{P}(Z \leq z) &= \int_0^z f_Z(t)dt = \int_0^z \left(-\frac{3}{2}t^2 + \frac{3}{2}\right)dt \\ &= \left[-\frac{t^3}{2} + \frac{3}{2}t\right]_0^z = \frac{3}{2}z - \frac{z^3}{2}.\end{aligned}$$

More formally, we have

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{3}{2}z - \frac{z^3}{2} & 0 \leq z \leq 1 \\ 1 & z > 1. \end{cases}$$

2.

$$\mathbb{P}(Z > 0.5) = 1 - F_Z(0.5) = 0.3125.$$

Solution (13.7). *We must solve*

$$1 = \int_{-\infty}^{\infty} kx(x-4)^2 dx$$

where we can write as

$$k\int_0^{10} \left(x^3 - 8x^2 + 16x\right)dx = k\left[\frac{x^4}{4} - \frac{8x^3}{3} + 8x^2\right]_0^{10}.$$

When we input our limits in we get

$$1 = \frac{k \times 1900}{3}$$

which we can rearrange to gain

$$k = \frac{3}{1900}$$

.

Solution (13.8).

$$\mathbb{P}(x + c \geq X \geq x) = \int_x^{x+c} f_X(t)dt = \int_x^{x+c} \frac{1}{100}dt = \left[\frac{t}{100}\right]_x^{x+c} = \frac{c}{100}.$$

This makes sense, if the accident can happen at any point in the road uniformly, then any interval of length c will have probability $\frac{c}{100}$ of having an accident in.

Solution (13.9). *We start by finding k*

$$1 = \int_0^{\infty} ke^{-5t}dt = \left(\frac{-1}{5}ke^{-5t}\right)_0^{\infty} = \frac{-1}{5}k(0-1) = \frac{k}{5}.$$

Thus $k = 5$. We will now use this to calculate

$$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 5e^{-5t}dt = [-e^{-5t}]_1^2 = -e^{-10} + e^{-5}.$$

Solution (13.10). *The distribution function is given by*

$$F_X(x) = \begin{cases} 0 & \textit{for } x \in (-\infty, 0] \\ x/3 & \textit{for } x \in (0,1], \\ 1/3 & \textit{for } x \in (1,2], \\ x/3 - 1/3 & \textit{for } x \in (2,4], \\ 1 & \textit{for } x \in (4,\infty) \end{cases}$$

21.14 Solutions To Mean and Variance for Continuous Random Variables Questions

Solution (14.3). *To find the expectation,*

$$\begin{aligned}\mathbb{E}(Z) &= \int_0^1 z f_Z(z)dz = \int_0^1 \left(-\frac{3}{2}z^3 + \frac{3}{2}z\right)dz \\ &= \left[-\frac{3}{8}z^4 + \frac{3}{4}z^2\right]_0^1 = 0.375.\end{aligned}$$

and to get the standard deviation we first find

$$\begin{aligned}\mathbb{E}(Z^2) &= \int_0^1 z^2 f_Z(z)dz = \int_0^1 \left(-\frac{3}{2}z^4 + \frac{3}{2}z^2\right)dz \\ &= \left[-\frac{3}{10}z^5 + \frac{3}{6}z^3\right]_0^1 = 0.2.\end{aligned}$$

Therefore

$$sd(Z) = \sqrt{VarZ} = \sqrt{\mathbb{E}(Z^2) - \mathbb{E}(Z)^2} = 0.244 \text{ to 3 d.p.}$$

Solution (14.4). *The density function is given by*

$$F_X(x) = \begin{cases} 0 & \text{for } x \in (-\infty, 2] \\ \int_{-\infty}^{\infty} \frac{t+1}{8}dt = \int_2^x \frac{t}{8} + \frac{1}{8}dt = \left[\frac{t^2}{16} + \frac{t}{8}\right]_2^x = \frac{1}{16}(x+4)(x-2) & \text{for } x \in (2, 4] \\ 1 & \text{for } x \in (4, \infty) \end{cases}$$

Solution (14.5).

$$\mathbb{E}(X) = \int_0^{\pi} \frac{t}{2}\sin(t)dt = \left[-\frac{t}{2}\cos(t)\right]_0^{\pi} - \int_0^{\pi} \frac{-\cos(t)}{2}dt = \left[\frac{\pi}{2} + 0\right] + \left[\frac{\sin(t)}{2}\right]_0^{\pi} = \frac{\pi}{2}.$$

21.15 Solutions To Moment Generating Functions Questions

Solution (15.5). *If $X \sim Bin(n,p)$,*

$$M_X(t) = \sum_{r=0}^{n} e^{tr}\binom{n}{r}p^r(1-p)^{n-r} = \sum_{r=0}^{n}\binom{n}{r}(e^t p)^r(1-p)^{n-r} = [pe^t + 1 - p]^n,$$

by the binomial theorem.

On the other hand, if $S_n = Y_1 + Y_2 + \cdots + Y_n$ is the sum of n i.i.d. Bernoulli(p) random variables, then using Theorem 15.3 and the fact that if $Y \sim Bernoulli(p)$ then $M_Y(t) = pe^t + 1 - p$ (as shown in the notes), then

$$M_{S(n)}(t) = [pe^t + 1 - p]^n.$$

Since the mgf determines the distribution uniquely, we conclude that $S(n) \sim Bin(n,p)$.

Solution (15.6). *If X is discrete, then $M_X(t) = \sum_{x\in R_X} e^{tx} p_X(x)$ and so*

$$M_X(0) = \sum_{x\in R_X} p_X(x) = 1.$$

If X is continuous and has pdf f, then $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$, and so $M_X(0) = \int_{-\infty}^{\infty} f(x)dx = 1$.

Solution (15.7).

$$\mathbb{E}(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} 2\left(\frac{1}{3}\right)^x = 2\sum_{x=1}^{\infty}\left(\frac{e^t}{3}\right)^x = 2\frac{\frac{e^t}{3}}{1-\frac{e^t}{3}} = \frac{2}{3e^{-t}-1}.$$

This series approximation holds when $\left|\frac{e^t}{3}\right| < 1$ which is valid as we are interested the case when $t = 0$.

$$\mathbb{E}(X) = \frac{d}{dt}\left.\frac{2}{3e^{-t}-1}\right|_{t=0} = \left.\frac{6e^{-t}}{(3e^{-t}-1)^2}\right|_{t=0} = \frac{3}{2}.$$

21.16 Solutions To The Normal Distribution Questions

Solution (16.7). *1.* 0.0227501,

2. 0.1359052.

Solution (16.8). *By definition*

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[-2xt\sigma^2 + (x-\mu)^2\right]\right\} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[(x-(\mu+t\sigma^2))^2 - 2\mu t\sigma^2 - t^2\sigma^4\right]\right\} dx \\
&= e^{\mu t + \frac{1}{2}t^2\sigma^2} \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-(\mu+t\sigma^2))^2}{\sigma}\right]\right\} dx\right) = e^{\mu t + \frac{1}{2}t^2\sigma^2}
\end{aligned}$$

as required. This is because

$$1 = \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-(\mu+t\sigma^2))^2}{\sigma}\right]\right\} dx\right)$$

as this is the p.d.f. of a normal random variable with parameters $N(\mu + t\sigma^2, \sigma)$*.*

21.17 Solutions To Distributions in R Questions

Solution (Solution to Exercise 17.1). *1. Firstly identify the required probability (a sketch might help). Starting at -∞ we want to know how much probability has accumulated by the time we get to 80kg and then to subtract this value from one.*

$$1 - \Phi\left(\frac{80-75}{4}\right) = 1 - \Phi\left(\frac{5}{4}\right) = 1 - \Phi(1.25)$$

Using `pnorm(1.25,0,1)` we find $\Phi(1.25) = 0.8944$.

So the probability that a randomly selected male weighs more than 80kg is

$$1 - 0.8944 = 0.1056.$$

2. First, sketch the density and identify the range we are looking for. A little thought shows that the shortest interval will be obtained by having the interval symmetrically placed in the centre of the distribution. Thus we want to know the weights (in kg) (say x_1 and x_2) that provide lower and upper bounds on the interval containing 50% of the weights. By symmetry this means that each tail must contain 25%. Thus,

$$\Phi\left(\frac{x_1-75}{4}\right) = 0.25 \textit{ and } \Phi\left(\frac{x_2-75}{4}\right) = 0.75$$

We can now use the `qnorm` function in the form `qnorm(0.75,0,1)` to find that the 0.75 quantile (upper quartile) is 0.6745, i.e., $\Phi(0.6745) = 0.75$, and thus

$$\begin{aligned} \frac{x_2-75}{4} &= 0.6745 \\ x_2 &= (0.6745 \times 4) + 75 = 77.698kg. \end{aligned}$$

Then, by symmetry,

$$x_1 = 75 - (77.698 - 75) = 72.302kg.$$

or we can use `qnorm(0.25,0,1)` to identify that -0.6745 is the lower quartile and work as above.

As a result, we can state that 50% of the men in the population in question weigh between approximately 72.3kg and 77.7kg.

21.18 Solutions To The Poisson Distribution Questions

Solution (18.5). *Solution. We are used to t measuring time but there is no reason why it has to. Here it measures length - $N(t)$ is the number of defects found in t miles.*

1. $t = 2, \lambda = 0.1$,
$$\mathbb{P}(N(2) = 0) = e^{-0.2} = 0.8187.$$

2.
$$\begin{aligned}
\mathbb{P}(N(2) = 1, N(8) = 3) &= \mathbb{P}(N(2) = 1, N(8) - N(2) = 2) \\
&= \mathbb{P}(N(2) = 1)\mathbb{P}(N(8) - N(2) = 2) \quad \textit{by indept. increments} \\
&= \mathbb{P}(N(2) = 1)\mathbb{P}(N(6) = 2) \quad \textit{by stationary increments} \\
&= e^{-0.2}(0.2) \times e^{-0.6}\frac{(0.6)^2}{2} = 0.0162.
\end{aligned}$$

Solution (18.6). *We require a length of time t months satisfying*
$$P(N(t) \geq 150) \geq 0.99$$
For this to be satisfied, $N(t)$ will need to have a large expected value, and so a normal approximation should be good.
$$N(t) \sim Po(5t) \overset{\text{approx.}}{\sim} \mathcal{N}(5t, 5t).$$
So we require
$$\begin{aligned}
1 - \Phi\left(\frac{149.5 - 5t}{\sqrt{5t}}\right) &\geq 0.99 \\
\frac{149.5 - 5t}{\sqrt{5t}} &\leq \Phi^{-1}(0.01) \\
&= -2.3263 \\
5t - 5.2018\sqrt{t} - 149.5 &\geq 0.
\end{aligned}$$
The corresponding equation is quadratic in $\sqrt{t}$, with roots
$$\sqrt{t} = \frac{5.2018 + \sqrt{(5.2018)^2 + 4 \times 5 \times 149.5}}{2 \times 5}.$$
The positive root is
$$\sqrt{t} = 6.0130.$$
Hence we require
$$t \geq (6.0130)^2 = 36.16.$$

Solution (18.7). *Let X be the number of UK earthquakes next year that are of magnitude between 3 and 3.9 on the Richter scale. If we suppose that $X \sim Poisson(\lambda)$, and we observe, on average, 3 such earthquakes per year, then as $\mathbb{E}(X) = \lambda$, a suitable choice is $\lambda = 3$. Then*

1.
$$\mathbb{P}(X = 3) = \frac{e^{-3}3^3}{3!} = 0.224 \textit{ to 3 d.p.}$$

2.
$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - \frac{e^{-3}3^0}{0!} - \frac{e^{-3}3^1}{1!} = 0.801 \textit{ to 3 d.p.}$$

Solution (18.8)**.** *Given $X \sim Poisson(5)$ we require*

$$\begin{aligned}
\mathbb{P}(X=5|4 \leq X \leq 6) &= \frac{\mathbb{P}(X=5 \text{ and } 4 \leq X \leq 6)}{\mathbb{P}(4 \leq X \leq 6)} \\
&= \frac{\mathbb{P}(X=5)}{\mathbb{P}(4 \leq X \leq 6)} \\
&= \frac{\mathbb{P}(X=5)}{\mathbb{P}(X=4)+\mathbb{P}(X=5)+\mathbb{P}(X=6)} \\
&= \frac{\frac{e^{-5}5^5}{5!}}{\frac{e^{-5}5^4}{4!}+\frac{e^{-5}5^5}{5!}+\frac{e^{-5}5^6}{6!}} \\
&= 0.353.
\end{aligned}$$

Solution (18.9)**.**

1. *Let X be the number of buses that arrive at the bus stop in half an hour: $X \sim Poisson(5)$.*
2. $\mathbb{P}(X=4) = \frac{e^{-5}5^4}{4!} = \frac{625}{24e^5} \approx 0.1755$.
3. $\mathbb{P}(X>2) = 1-\mathbb{P}(X \leq 2) = 1-\mathbb{P}(X=0)-\mathbb{P}(X=1)-\mathbb{P}(X=2) = 1-\frac{e^{-5}5^0}{0!}-\frac{e^{-5}5^1}{1!}-\frac{e^{-5}5^2}{2!} = 1-\frac{37}{2e^5} \approx 0.8753$.

Solution (18.10)**.** *Proof omitted, begin a discussion board thread for further class discussion.*

Solution (18.11)**.**

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \lambda e^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^{x-1}}{x!}
\end{aligned}$$

let $y = x-1$

$$\begin{aligned}
\mathbb{E}(X) &= \lambda e^{-\lambda}\sum_{y=0}^{\infty}(y+1)\frac{\lambda^y}{(y+1)!} \\
&= \lambda e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda.
\end{aligned}$$

21.19 Solutions To The Gamma, Uniform and Exponential Distributions Questions

Solution (19.15). *Let T be the time taken for one atom to decay. If $T \sim Exponential(\lambda)$ and we estimate that $\mathbb{P}(T \leq 5730) = 0.5$, then*

$$\begin{aligned} F_T(5730) &= 1 - \exp(-\lambda 5730) = 0.5 \\ \Rightarrow \lambda &= \frac{\ln 0.5}{-5730}, \end{aligned}$$

and so

$$\mathbb{E}(T) = \frac{1}{\lambda} = \frac{5730}{\ln 2}.$$

Hence the expected time for one carbon-14 atom to decay is 8267 years (to the nearest year).

Solution (19.16). *We have $T \sim U[0, 5]$.*

1. $\mathbb{E}(T) = (0 + 5)/2 = 2.5$, *and*

$$\mathbb{E}(T^2) = Var(T) + \{\mathbb{E}(T)\}^2 = \frac{5^2}{12} + \frac{25}{4} = \frac{25}{3}.$$

2. $F_T(t) = t/5$, *so*

$$\mathbb{P}(3 < T < 4) = F_T(4) - F_T(3) = 1/5.$$

3. *We require the 95th percentile $t_{0.95}$ such that $F_T(t_{0.95}) = 0.95$, where $F_T(t) = t/5$. Therefore the 95th percentile is $0.95 \times 5 = 4.75$.*

4. $\mathbb{P}(T = 2) = 0$, *because T is a continuous random variable.*

Solution (19.17). *Let T be the time (in days) until the patient next experiences symptoms. We suppose that $T \sim Exp(\lambda)$.*

1. *The doctor believes that $\mathbb{P}(T \geq 30) = 0.5$, so*

$$\begin{aligned} 0.5 = \mathbb{P}(T \geq 30) &= \int_{30}^{\infty} f_T(t)dt = \int_{30}^{\infty} \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_{30}^{\infty} = e^{-\lambda 30} \\ \Rightarrow \lambda &= 0.023 \text{ to 3 d.p.} \end{aligned}$$

2. *We require*

$$\begin{aligned} \mathbb{P}(T \geq 60) &= \int_{60}^{\infty} f_T(t)dt = \int_{60}^{\infty} \lambda e^{-\lambda t} dt \\ &= \left[-e^{-\lambda t}\right]_{60}^{\infty} = e^{-60\lambda} = 0.25, \end{aligned}$$

Writing $\lambda = -(\ln 0.5)/30$ will avoid rounding error here: $\exp\{-60 \times (-\ln 0.5)/30\} = \exp(\ln 0.25)$.

3. *By the lack of memory property of the exponential distribution*

$$\mathbb{P}(T \geq 30 + 30 | T \geq 30) = \mathbb{P}(T \geq 30) = 0.5.$$

Solution (19.18). *We have $T_i \sim Exp(\lambda)$, and, based on the doctor's judgement,*

As above $\lambda = 0.023$. Now, as $T_1, \ldots, T_{50}$ are i.i.d., using the rules for expectation and variance we have

$$\mathbb{E}(\bar{T}) = \frac{1}{50}\sum_{i=1}^{50} \mathbb{E}(T_i) = \frac{1}{50} \times 50 \times \frac{1}{\lambda} = -30/(\ln 0.5) \simeq 43.28,$$

and

$$Var(\bar{T}) = \frac{1}{50^2}\sum_{i=1}^{50} Var(T_i) = \frac{1}{50^2} \times 50 \times \frac{1}{\lambda^2} = 18/(\ln 0.5)^2 \simeq 37.46.$$

Solution (19.19). *We know that the p.d.f. of a gamma distribution is*

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

for $\alpha > 0$ *and* $\lambda > 0$. *If we set* $alpha = 1$ *we attain the p.d.f. of the exponential distribution as*

$$f_X(x) = \begin{cases} \frac{\lambda}{\Gamma(1)} x^{1-1} e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

for $\lambda > 0$.

21.20 Solutions To The Beta Distribution Questions

Solution (20.2)**.**

$$\begin{aligned}\mathbb{E}(X^n) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^n x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}dx\end{aligned}$$

We know from equation (22) *that*

$$\int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(n+\alpha)\Gamma(\beta)}{\Gamma(n+\alpha+\beta)}.$$

Therefore

$$\begin{aligned}\mathbb{E}(X^n) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n+\alpha)\Gamma(\beta)}{\Gamma(n+\alpha+\beta)} \\ &= \frac{\Gamma(\alpha+\beta)\Gamma(n+\alpha)}{\Gamma(\alpha)\Gamma(n+\alpha+\beta)}.\end{aligned}$$

Solution (20.3)**.** *Using the result from Exercise 20.2 we can calculate*

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(\alpha+\beta+2)} - \left(\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)}\right)^2 \\ &= \frac{\Gamma(\alpha+\beta)\alpha(\alpha+1)\Gamma(\alpha)}{\Gamma(\alpha)(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta)} - \left(\frac{\Gamma(\alpha+\beta)\alpha\Gamma(\alpha)}{\Gamma(\alpha)(\alpha+\beta)\Gamma(\alpha+\beta)}\right)^2 \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\end{aligned}$$