Cardiff University School of Mathematics

MA2500

Foundations of Probability and Statistics

Reading Material

2014-15

Version: 19/12/2014

Contents

Lecture 1 Introduction

1.1 Timetable

In the autumn semester, we will look at the foundations of **probability theory**.

Day	Time	Location	
Mondays	11.10 - 12.00	E/0.15	
Wednesdays	11.10 - 12.00	Tower 0.02	
Fridays	10.00 - 10.50	E/0.15	(even weeks only)

The material is divided into five blocks of four lectures:

Weeks 1–2:	Probability spaces
Weeks 3–4:	Random variables and distributions
Weeks 5–6:	Expectation
Weeks 7–8:	Limit theorems
Weeks 9–10:	Joint distributions

1.2 Flipped classroom

The module will run using a **flipped classroom** approach:

- A traditional lecture is primarily a means of transferring information, from the lecturer to the student. The student is expected to assimilate the information, by reading notes and completing exercises, to gain an understanding of the material, ideally before the next lecture.
- In a flipped classroom, students are expected to acquire the relevant information *before* attending the class. Contact hours are dedicated to discussion and exercises, leading to assimilation and understanding.

Reading material will be provided. This will consist of

(1) definitions,

(2) theorems and proofs,

(3) illustrative examples, and

(4) exercises.

Before each class, you are expected to

(1) learn the definitions,

(2) understand the statements of the theorems,

(3) look at the proofs,

(4) try some of the exercises.

During each class, we will

(1) clarify definitions, theorems and proofs (where necessary),

(2) work on exercises, and

(3) assess each other's work.

Points to note:

- Printed "sample answers" to exercises will **not** be provided.
- Answers will be provided via **screencast**, but strictly on request.

1.3 Assessment

1.3.1 Formative assessment

Answers to exercises can be submitted at any time for assessment and feedback.

1.3.2 Summative assessment

Coursework accounts for 20% of the total marks for the module.

- There will be 10 homework assignments, each worth 2%.
- Hand out: Monday of weeks 2/4/6/8/10 (autumn/spring)
- Hand in: Friday of weeks 2/4/6/8/10 (autumn/spring)

1.3.3 Assessment criteria

Submitted work should be **clear**, **concise** and **correct**.

All work will be assessed according to each of these categories, on a scale of 1 to 5.

Category	1	2	3	4	5
Clear	Obscure	Opaque	Fair	Clear	Crystal
Concise	Long-winded or complete lack of detail	Too long or lack of detail	Fair	Good	Perfect
Correct	Completely incorrect	Major mistakes	Fair	Minor mistakes	Completely correct

Draft Assessment Criteria for MA2500

Formative and summative work should be submitted using the pro-forma shown on the following page.

Lecture 2 Set Theory

To be read in preparation for the **11.00** lecture on **Wed 01 Oct** in **Tower 0.02**.

2.1 Elementary set theory

A set is a collection of distinct *elements*.

- If a is an element of the set A, we denote this by $a \in A$.
- If a is *not* an element of A, we denote this by $a \notin A$.
- The *cardinality* of a set is the number of elements it contains.
- The *empty set* contains no elements, and is denoted by $\emptyset$.

2.1.1 Set relations

Let A, B be sets.

- If $a \in B$ for every $a \in A$, we say that A is a *subset* of B, denoted by $A \subseteq B$.
- If $A \subseteq B$ and $B \subseteq A$, we say that A is *equal* to B, denoted by $A = B$,
- If $A \subseteq B$ and $A \neq B$, we say that A is a *proper subset* of B, denoted by $A \subset B$.

2.1.2 Set operations

Let A, B and Ω be sets, with $A, B \subseteq \Omega$.

- The *union* of A and B is the set $A \cup B = \{a \in \Omega : a \in A \text{ or } a \in B\}$.
- The *intersection* of A and B is the set $A \cap B = \{a \in \Omega : a \in A \text{ and } a \in B\}$.
- The *complement* of A (relative to Ω) is the set $A^c = \{a \in \Omega : a \notin A\}$.

2.1.3 Set algebra

Commutative property:	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Associative property:	$(A \cup B) \cup C = A \cup (B \cup C)$
	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributive property:	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

2.2 Sample space, outcomes and events

Definition 2.1

(1) Any process of observation or measurement whose outcome is uncertain is called a *random experiment*.

(2) A random experiment has a number of possible *outcomes*.

(3) Each time a random experiment is performed, *exactly one* of its outcomes will occur.

(4) The set of all possible outcomes is called the *sample space*, denoted by Ω.

(5) Outcomes are also called *elementary events*, and denoted by $\omega \in \Omega$.

Example 2.2

- $\{1, 2, \ldots, n\}$ is a finite sample space,
- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ is a countably infinite sample space,
- $[0, 1]$ is an uncountable sample space.

Definition 2.3

(1) An *event* A is a subset of the sample space, Ω.

(2) If outcome ω occurs, we say that event A *occurs* if and only if $\omega \in A$.

(3) Two events A and B with $A \cap B = \emptyset$ are called *disjoint* or *mutually exclusive*.

(4) The empty set $\emptyset$ is called the *impossible event*.

(5) The sample space itself is called the *certain event*.

Remark 2.4

- If A occurs and $A \subseteq B$, then B occurs.
- If A occurs and $A \cap B = \emptyset$, then B does not occur.

2.3 Countable unions and intersections

Definition 2.5

Let Ω be any set. The set of all subsets Ω is called its *power set*.

- If Ω is a finite set, its power set is also finite.
- If Ω is a countably infinite set, its power set is uncountable set (Cantor's Theorem).
- If Ω is an uncountable set, its power set is also uncountable.

Definition 2.6

Let $A_1, A_2, \ldots$ be subsets of Ω.

(1) The (countable) *union* of $A_1, A_2, \ldots$ is the set

$$\bigcup_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ for some } A_i\}.$$

(2) The (countable) *intersection* of $A_1, A_2, \ldots$ is the set

$$\bigcap_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ for all } A_i\}.$$

Theorem 2.7 (De Morgan's laws)
For a countable collection of sets $\{A_1, A_2, \ldots\}$,

(1) $\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$,

(2) $\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$.

Proof:

(1) Let $a \in \left(\bigcup_{i=1}^{\infty} A_i\right)^c$. Then $a \notin \bigcup_{i=1}^{\infty} A_i$, and so $a \in A_i^c$ for all A_i.
Hence $\left(\bigcup_{i=1}^{\infty} A_i\right)^c \subseteq \bigcap_{i=1}^{\infty} A_i^c$.
Let $a \in \bigcap_{i=1}^{\infty} A_i^c$. Then $a \notin A_i$ for all A_i, and so $a \notin \bigcup_{i=1}^{\infty} A_i$.
Hence $\bigcap_{i=1}^{\infty} A_i^c \subseteq \left(\bigcup_{i=1}^{\infty} A_i\right)^c$.

(2) Applying part (1) to the collection of sets $\{A_1^c, A_2^c, \ldots\}$,
$\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} \left(A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i$. Taking the complement of both sides,
$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$.

2.4 Collections of sets

Definition 2.8
Let Ω be any set. Any subset of its power set is called a *collection of sets over* Ω.

Let Ω be the sample space of some random experiment. If we are interested whether the events A and B occur, we must also be interested in

- the event $A \cup B$: whether event A occurs *or* event B occurs;
- the event $A \cap B$: whether event A occurs *and* event B occurs;
- the event A^c: whether the event A does *not* occur.

Thus we can not use arbitrary collections of sets over Ω as the basis for investigating random experiments. Instead, we allow only collections which are *closed* under certain set operations.

Definition 2.9
A collection of sets $\mathcal{C}$ over Ω is said to be

(1) *closed under complementation* if $A^c \in \mathcal{C}$ for every $A \in \mathcal{C}$,

(2) *closed under pairwise unions* if $A \cup B \in \mathcal{C}$ for every $A, B \in \mathcal{C}$,

(3) *closed under finite unions* if $\bigcup_{i=1}^{n} A_i \in \mathcal{C}$ for every $A_1, A_2, \ldots A_n \in \mathcal{C}$,

(4) *closed under countable unions* if $\bigcup_{i=1}^{\infty} A_i \in \mathcal{C}$ for every $A_1, A_2, \ldots \in \mathcal{C}$.

Definition 2.10
A collection of sets $\mathcal{F}$ over Ω is called a *field* over Ω if

(1) $\Omega \in \mathcal{F}$,

(2) $\mathcal{F}$ is closed under complementation, and

(3) $\mathcal{F}$ is closed under pairwise unions.

Theorem 2.11 (Properties of fields)
Let $\mathcal{F}$ be a field over Ω. Then

(1) $\emptyset \in \mathcal{F}$,

(2) $\mathcal{F}$ is closed under set differences,

(3) $\mathcal{F}$ is closed under finite unions,

(4) $\mathcal{F}$ is closed under finite intersections.

Proof: See exercises.

Definition 2.12
A collection of sets $\mathcal{F}$ over Ω is called a *σ-field* ("sigma-field") over Ω if

(1) $\Omega \in \mathcal{F}$,

(2) $\mathcal{F}$ is closed under complementation, and

(3) $\mathcal{F}$ is closed under countable unions.

Theorem 2.13 (Properties of σ-fields)
Let $\mathcal{F}$ be a σ-field over Ω. Then

(1) $\emptyset \in \mathcal{F}$,

(2) $\mathcal{F}$ is closed under set differences,

(3) $\mathcal{F}$ is closed under finite unions,

(4) $\mathcal{F}$ is closed under finite intersections,

(5) $\mathcal{F}$ is closed under countable intersections.

Proof: See exercises.

2.5 Borel sets

In many situations of interest, random experiments yield outcomes that are *real numbers*.

Definition 2.14

- The *open interval* (a, b) is the set $\{x \in \mathbb{R} : a < x < b\}$.
- The *closed interval* $[a, b]$ is the set $\{x \in \mathbb{R} : a \leq x \leq b\}$.

Definition 2.15
The *Borel* σ-field over $\mathbb{R}$ is defined to be the smallest σ-field over $\mathbb{R}$ that contains all open intervals.

Remark 2.16

- The Borel σ-field is usually denoted by $\mathcal{B}$, and includes all closed interval, all half-open intervals, all finite sets and all countable sets.

- The elements of $\mathcal{B}$ are called *Borel sets* over $\mathbb{R}$.
- Borel sets can be thought of as the "nice" subsets of $\mathbb{R}$.

Proposition 2.17
The Borel σ-field over $\mathbb{R}$ contains all closed intervals.

Proof: Any closed interval $[a, b]$ can be written as a countable intersection of open intervals:

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n},\ b + \frac{1}{n} \right).$$

Hence $[a, b] \in \mathcal{B}$, because

- for every $n \in \mathbb{N}$, $\left(a - \frac{1}{n}, b + \frac{1}{n} \right) \in \mathcal{B}$, and
- by Theorem 2.13, $\mathcal{B}$ is closed under countable intersections.

2.6 Exercises

Exercise 2.1

1. Let $\mathcal{F}$ be a field over Ω. Show that
   (a) $\emptyset \in \mathcal{F}$,
   (b) $\mathcal{F}$ is closed under set differences,
   (c) $\mathcal{F}$ is closed under pairwise intersections,
   (d) $\mathcal{F}$ is closed under finite unions,
   (e) $\mathcal{F}$ is closed under finite intersections.
2. Let $\mathcal{F}$ be a σ-field over Ω. Show that
   (a) $\mathcal{F}$ is closed under finite unions,
   (b) $\mathcal{F}$ is closed under finite intersections.
   (c) $\mathcal{F}$ is closed under countable intersections.

Exercise 2.2

1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$.
   (a) What is the smallest σ-field containing the event $A = \{1, 2\}$?
   (b) What is the smallest σ-field containing the events $A = \{1, 2\}$, $B = \{3, 4\}$ and $C = \{5, 6\}$?
2. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-fields over Ω.
   (a) Show that $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ is a σ-field over Ω.
   (b) Find a counterexample to show that $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-field over Ω.

Lecture 3 Probability Spaces

To be read in preparation for the **11.00** lecture on **Mon 06 Oct** in **E/0.15**.

3.1 Probability measures

Definition 3.1
Let Ω be a sample space, and let $\mathcal{F}$ be a σ-field over Ω. A *probability measure* on $(\Omega, \mathcal{F})$ is a function

$$\begin{aligned} \mathbb{P}: \quad \mathcal{F} &\to [0,1] \\ A &\mapsto \mathbb{P}(A) \end{aligned}$$

such that $\mathbb{P}(\Omega) = 1$, and for any countable collection of pairwise disjoint events $\{A_1, A_2, \ldots\}$,

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

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.*

Remark 3.2
- The second property is called *countable additivity.*

Remark 3.3
In the more general setting of measure theory:

- The elements of $\mathcal{F}$ are called *measurable sets.*
- The pair $(\Omega, \mathcal{F})$ is called a *measurable space.*
- The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *measure space.*

Example 3.4
A fair six-sided die is rolled once. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for the experiment is given by

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $\mathcal{F} = \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ denotes the power set of Ω,
- $\mathbb{P}(A) = |A|/|\Omega|$ for every $A \in \mathcal{F}$ (where $|A|$ denotes the cardinality of A).

If we are only interested in odd and even numbers, we can instead take

- $\Omega = \{1,2,3,4,5,6\}$,
- $\mathcal{F} = \big\{\emptyset, \{1,3,5\}, \{2,4,6\}, \{1,2,3,4,5,6\}\big\}$
- $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{1,3,5\}) = 1/2$, $\mathbb{P}(\{2,4,6\}) = 1/2$, $\mathbb{P}(\{1,2,3,4,5,6\}) = 1$.

3.2 Null and almost-certain events

Definition 3.5

(1) If $\mathbb{P}(A) = 0$, we say that A is a *null event*.

(2) If $\mathbb{P}(A) = 1$, we say that A occurs *almost surely* (or "*with probability 1*").

Remark 3.6

- A null event is not the same as the impossible event ($\emptyset$).
- An event that occurs almost surely is not the same as the certain event (Ω).

Example 3.7

A dart is thrown at a dartboard.

- The probability that the dart hits a given point of the dartboard is 0.
- The probability that the dart does not hit a given point of the dartboard is 1.

3.3 Properties of probability measures

Theorem 3.8 (Properties of probability measures)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B \in \mathcal{F}$.

(1) Complementarity: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) $\mathbb{P}(\emptyset) = 0$,

(3) Monotonicity: if $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(4) Addition rule: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proof:

(1) Since $A \cup A^c = \Omega$ is a disjoint union and $\mathbb{P}(\Omega) = 1$, it follows by additivity that

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

(2) Since $\emptyset = \Omega^c$ and $\mathbb{P}(\Omega) = 1$, it follows by complemenarity that

$$\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0.$$

(3) Let $A \subseteq B$ and let us write $B = A \cup (B \setminus A)$.
Since A and $B \setminus A$ are disjoint sets, it follows by additivity that

$$\mathbb{P}(B) = \mathbb{P}\big[A \cup (B \setminus A)\big] = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

Hence, because $\mathbb{P}(B \setminus A) \geq 0$, it follows that $\mathbb{P}(B) \geq \mathbb{P}(A)$.

(4) Let us write:

- $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$
- $A = (A \setminus B) + (A \cap B)$
- $B = (B \setminus A) + (A \cap B)$

These are disjoint unions, so by additivity,

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

Hence $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, as required.

3.4 Continuity of probability measures

Theorem 3.9 (Continuity of probability measures)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(1) For an increasing sequence of events $A_1 \subseteq A_2 \subseteq \dots$ in $\mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(2) For a decreasing sequence of events $B_1 \supseteq B_2 \supseteq \dots$ in $\mathcal{F}$,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Proof: To prove the first part, let $A_1 \subseteq A_2 \subseteq \dots$ be an increasing sequence of events, and

$$A = \bigcup_{i=1}^{\infty} A_i.$$

We can write A as a disjoint union

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$$

Since the sets $A_{i+1} \setminus A_i$ are disjoint, by countable additivity we have

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \dots$$

Furthermore, $A_i \subseteq A_{i+1}$ means that $A_{i+1} = (A_{i+1} \setminus A_i) \cup A_i$ is a disjoint union, so

$$\mathbb{P}(A_{i+1} \setminus A_i) = \mathbb{P}(A_{i+1}) - \mathbb{P}(A_i).$$

Hence

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A_1) + \big[\mathbb{P}(A_2) - \mathbb{P}(A_1)\big] + \big[\mathbb{P}(A_3) - \mathbb{P}(A_2)\big] + \dots \\
&= \big[\mathbb{P}(A_1) - \mathbb{P}(A_1)\big] + \big[\mathbb{P}(A_2) - \mathbb{P}(A_2)\big] + \big[\mathbb{P}(A_3) - \mathbb{P}(A_3)\big] + \dots \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

To prove the second part, let $B_1 \supseteq B_2 \supseteq \dots$ be a decreasing sequence of events, and

$$B = \bigcap_{i=1}^{\infty} B_i.$$

Let $A_i = B_i^c$ and $A = B^c$. Then $A_1 \subseteq A_2 \subseteq \dots$ is an increasing sequence, and

$$A = \bigcup_{i=1}^{\infty} A_i.$$

Hence by the first part of the theorem,

$$\begin{aligned}
\mathbb{P}(B) &= 1 - \mathbb{P}(A) \\
&= 1 - \lim_{n\to\infty} \mathbb{P}(A_n) \\
&= \lim_{n\to\infty} (1 - \mathbb{P}(A_n)) \\
&= \lim_{n\to\infty} \mathbb{P}(B_n).
\end{aligned}$$

3.5 Exercises

Exercise 3.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B, C \in \mathcal{F}$. Show that
$$\mathbb{P}(A\cup B\cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A\cap B) - \mathbb{P}(A\cap C) - \mathbb{P}(B\cap C) + \mathbb{P}(A\cap B\cap C)$$
This is called the *inclusion-exclusion principle.*

2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
   (a) Show that $\mathbb{P}(A\cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ for all $A, B \in \mathcal{F}$. This is called *subadditivity.*
   (b) Show that for any sequence $A_1, A_2, \dots$ of events in $\mathcal{F}$,
   $$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$
   This is called *countable subadditivity.*

Exercise 3.2

1. Let A and B be events with probabilities $\mathbb{P}(A) = 3/4$ and $\mathbb{P}(B) = 1/3$.
   (a) Show that $\frac{1}{12} \leq \mathbb{P}(A\cap B) \leq \frac{1}{3}$, and construct examples to show that both extremes are possible.
   (b) Find corresponding bounds for $\mathbb{P}(A\cup B)$.

2. A roulette wheel consists of 37 slots of equal size. The slots are numbered from 0 to 36, with odd-numbered slots coloured red, even-numbered slots coloured black, and the slot labelled 0 coloured green. The wheel is spun in one direction and a ball is rolled in the opposite direction along a track running around the circumference of the wheel. The ball eventually falls on to the wheel and into one of the 37 slots. A player bets on the event that the ball stops in a red slot, and another player bets on the event that the ball stops in a black slot.
   (a) Define a suitable sample space Ω for this random experiment, and identify the events of interest.
   (b) Find the smallest field $\mathcal{F}$ over Ω that contains the events of interest.
   (c) Define a suitable probability measure $(\Omega, \mathcal{F})$ to represent the game.

Exercise 3.3

1. A biased coin has probability p of showing heads. The coin is tossed repeatedly until a head occurs. Describe a suitable probability space for this experiment.

2. A fair coin is tossed repeatedly.
   (a) Show that a head eventually occurs with probability one.
   (b) Show that a sequence of 10 consecutive tails eventually occurs with probability one.
   (c) Show that any finite sequence of heads and tails eventually occurs with probability one.

Lecture 4 Conditional Probability

To be read in preparation for the **11.00** lecture on **Wed 08 Oct** in **Tower 0.02**.

4.1 Conditional probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $B \in \mathcal{F}$.

Definition 4.1
If $\mathbb{P}(B) > 0$, the *conditional probability of A given B* is defined to be

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

4.2 Bayes' theorem

Definition 4.2
A countable collection of sets $\{A_1, A_2, \ldots\}$ is said to form a *partition* of a set B if

(1) $A_i \cap A_j = \emptyset$ for all $i \neq j$, and

(2) $B \subseteq \bigcup_{i=1}^{\infty} A_i$.

Theorem 4.3 (The Law of Total Probability)
If $\{A_1, A_2, \ldots\}$ is a partition of B, then

$$\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Theorem 4.4 (Bayes' Theorem)
If $\{A_1, A_2, \ldots\}$ is a partition of B where $\mathbb{P}(B) > 0$, then

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

4.3 Independence

Definition 4.5
Two events A and B are said to be *independent* if $\mathbb{P}(A|B) = \mathbb{P}(A)$, or equivalently,

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

Definition 4.6
A collection of events $\{A_1, A_2, \ldots\}$ is said to be

(1) *pairwise independent* if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$.

(2) *totally independent* if, for every finite subset $\{B_1, B_2, \ldots, B_m\} \subset \{A_1, A_2, \ldots\}$,

$$\mathbb{P}(B_1 \cap B_2 \cap \ldots \cap B_m) = \mathbb{P}(B_1)\mathbb{P}(B_2)\cdots\mathbb{P}(B_m).$$

This can also be written as $\mathbb{P}\left(\bigcap_{j=1}^m B_j\right) = \prod_{j=1}^m \mathbb{P}(B_j)$.

Remark 4.7
Total independence implies pairwise independence, but not vice versa.

4.4 Conditional probability spaces

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $B \in \mathcal{F}$.

Theorem 4.8
The family of sets $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ is a σ-field over B.

Remark 4.9
$\mathcal{G}$ contains all sets of the form $A \cap B$, where A is some element of $\mathcal{F}$. This means that $A' \in \mathcal{G}$ if and only if there is some $A \in \mathcal{F}$ for which $A' = A \cap B$.

Proof: To show that $\mathcal{G}$ is a σ-field over B, we need to show that

(1) $B \in \mathcal{G}$,

(2) if $A' \in \mathcal{G}$ then $B \setminus A' \in \mathcal{G}$, and

(3) if $A'_1, A'_2, \ldots \in \mathcal{G}$ then $\cup_{i=1}^\infty A'_i \in \mathcal{G}$.

(1) Clearly, $B \in \mathcal{G}$ because there is a set $A \in \mathcal{F}$ for which $B = A \cap B$, namely the set B istelf.

(2) Let $A' \in \mathcal{G}$. Then there exists a set $A \in \mathcal{F}$ for which $A' = A \cap B$.
The complement of A' relative to B can be written as

$$B \setminus A' = B \setminus (A \cap B) = \big[(A \cap B)^c\big] \cap B.$$

- $\mathcal{F}$ is closed under pairwise unions and complementation.
- Since $A, B \in \mathcal{F}$, it thus follows that $(A \cap B)^c \in \mathcal{F}$.
- Hence $B \setminus A'$ can be written as $\big[(A \cap B)^c\big] \cap B$ where $\big[(A \cap B)^c\big] \in \mathcal{F}$
- This shows that $B \setminus A \in \mathcal{G}$.

(3) Let A'_1, A'_2, be elements of $\mathcal{G}$. Then for each A'_i there exists some $A_i \in \mathcal{F}$ such that $A'_i = A_i \cap B$. Using the fact that set intersection is distributive over set union,

$$\cup_i A'_i = \cup_i (A_i \cap B) = \big(\cup_i A_i\big) \cap B.$$

- $\mathcal{F}$ is closed under countable unions.

- Since $A_1, A_2, \ldots \in \mathcal{F}$, it thus follows that $\cup_i A_i \in \mathcal{F}$.
- Hence $\cup_i A'_i$ can be written in the form $\big(\cup_i A_i\big) \cap B$ where $\cup_i A_i \in \mathcal{F}$.
- This shows that $\cup_i A'_i \in \mathcal{G}$.

Thus we have shown that $\mathcal{G}$ is a σ-field over B, as required.

Theorem 4.10
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $B \in \mathcal{F}$, and let $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$.
If $\mathbb{P}(B) > 0$, then

$$\begin{array}{rccc} \mathbb{Q}: & \mathcal{G} & \to & [0,1] \\ & A' & \mapsto & \mathbb{P}(A'|B) \end{array}$$

is a probability measure on $(B, \mathcal{G})$.

Remark 4.11
$(B, \mathcal{G}, \mathbb{Q})$ is called a *conditional probability space.*.

Proof: To show that $\mathbb{Q}$ is a probability measure on $(B, \mathcal{G})$, we need to show that

- $\mathbb{Q}(B) = 1$,
- $\mathbb{Q}\left(\bigcup_i A'_i\right) = \sum_i \mathbb{Q}(A'_i)$ whenever the $A'_i \in \mathcal{G}$ are pairwise disjoint.

First,

$$\mathbb{Q}(B) = \mathbb{P}(B|B) = \frac{\mathbb{P}(B \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

To prove countable additivity, let $A'_1, A'_2, \ldots$ be pairwise disjoint events in $\mathcal{G}$. Then, using the fact that set intersection is distributive over set union,

$$\begin{aligned} \mathbb{Q}(\cup_i A'_i) = \mathbb{P}(\cup_i A'_i \,|\, B) &= \frac{\mathbb{P}\big[(\cup_i A'_i) \cap B\big]}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}\big[\cup_i (A'_i \cap B)\big]}{\mathbb{P}(B)} \\ &= \frac{\sum_i \mathbb{P}(A'_i \cap B)}{\mathbb{P}(B)} \quad \text{because the } A'_i \text{ are disjoint,} \\ &= \sum_i \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\ &= \sum_i \mathbb{Q}(A'_i). \end{aligned}$$

Thus we have shown that $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{G})$, as required.

Remark 4.12
We have shown that $\mathbb{Q}$ is a probability measure on $(B, \mathcal{G})$. Using an almost identical argument, it can be shown that $\mathbb{Q}$ is also a probability measure on $(\Omega, \mathcal{F})$.

- In the probability space $(B, \mathcal{G}, \mathbb{Q})$, outcomes $\omega \notin B$ are excluded from consideration.
- In the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, outcomes $\omega \notin B$ are assigned probability zero.

4.5 Exercises

Exercise 4.1 [Revision]

1. Let Ω be a sample space, and let $A_1, A_2, \ldots$ be a partition of Ω with the property that $\mathbb{P}(A_i) > 0$ for all i.
   (a) Show that $\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.
   (b) Show that $\mathbb{P}(A_i|B) = \dfrac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$.

Exercise 4.2

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and consider the function $\mathbb{Q} : \mathcal{F} \to [0, 1]$ defined by $\mathbb{Q}(A) = \mathbb{P}(A|B)$.
   (a) Show that $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space.
   (b) If $C \in \mathcal{F}$ and $\mathbb{Q}(C) > 0$, show that $\mathbb{Q}(A|C) = \mathbb{P}(A|B \cap C)$.
2. A random number N of dice are rolled. Let A_k be the event that $N = k$, and suppose that $\mathbb{P}(A_k) = 2^{-k}$ for $k \in \{1, 2, \ldots\}$ (and zero otherwise). Let S be the sum of the scores shown on the dice. Find the probability that:
   (a) $N = 2$ given that $S = 4$,
   (b) $S = 4$ given that N is even,
   (c) $N = 2$ given that $S = 4$ and the first die shows 1,
   (d) the largest number shown by any dice is r (where S is unknown).
3. Let $\Omega = \{1, 2, \ldots, p\}$ where p is a prime number. Let $\mathcal{F}$ be the power set of Ω, and let $\mathbb{P} : \mathcal{F} \to [0, 1]$ be the probability measure on $(\Omega, \mathcal{F})$ defined by $\mathbb{P}(A) = |A|/p$, where $|A|$ denotes the cardinality of A. Show that if A and B are independent events, then at least one of A and B is either $\emptyset$ or Ω.

Lecture 5 Random Variables

To be read in preparation for the **11.00** lecture on **Mon 13 Oct** in **E/0.15**.

5.1 Random variables

Random variables are functions that transform abstract sample spaces to the real numbers.

Definition 5.1
Let Ω be the sample space of some random experiment, and let $\mathcal{F}$ be a σ-field of events over Ω. A *random variable* on $(\Omega, \mathcal{F})$ is a function

$$\begin{aligned} X: \quad \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

with the property that $\{\omega : X(\omega) \in B\} \in \mathcal{F}$ for every $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-field over $\mathbb{R}$.

Remark 5.2

- The set $\{\omega : X(\omega) \in B\}$ contains precisely those outcomes that are mapped by X into the set B.
- X is a random variable only if every set of this form is an element of the σ-field $\mathcal{F}$.
- This condition means that, for any Borel set B, the probability that X takes a value in B is well-defined.

Let us define the following notation:

$$\{X \in B\} = \{\omega : X(\omega) \in B\}$$

- The expression $\{X \in B\}$ should not be taken literally: X is a function, while B is a subset of the real numbers.
- Instead, think of $\{X \in B\}$ as the event that X takes a value in B.
- The condition $\{X \in B\} \in \mathcal{F}$ ensures that the probability of this event is well-defined.

We denote the probability of $\{X \in B\}$ by $\mathbb{P}(X \in B)$, by which we mean

$$\mathbb{P}(X \in B) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big)$$

Proposition 5.3
A function $X : \Omega \to \mathbb{R}$ is a random variable if and only if $\{X \leq x\} \in \mathcal{F}$ for every $x \in \mathbb{R}$.

[*Proof omitted.*]

Remark 5.4
To check whether or not a function $X : \Omega \to \mathbb{R}$ is a random variable, by the proposition we need not verify that $\{X \in B\} \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}$. Instead, it is enough to verify only that the sets $\{\omega : X(\omega) \leq x\}$ are included in $\mathcal{F}$ (for every $x \in \mathbb{R}$).

5.2 Indicator variables

The elementary random variable is the *indicator variable* of an event A.

Definition 5.5
The *indicator variable* of an event A is the random variable $I_A : \Omega \to \mathbb{R}$ defined by

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Theorem 5.6
Let A and B be any two events. Then

(1) $I_{A^c} = 1 - I_A$

(2) $I_{A \cap B} = I_A I_B$

(3) $I_{A \cup B} = I_A + I_B - I_{A \cap B}$

Proof: Exercise. Note that for two functions to be equal, they must be equal at every point of their common domain, so for the first part we need to show that $I_{A^c}(\omega) = 1 - I_A(\omega)$ for every $\omega \in \Omega$, and similarly for parts (2) and (3).

5.3 Simple random variables

Definition 5.7
A *simple random variable* is one that takes only finitely many values.

If $X : \Omega \to \mathbb{R}$ is a simple random variable, it can be represented as:

$$X(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega)$$

where

- $\{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$ is the range of X, and
- $\{A_1, A_2, \ldots, A_n\}$ is a partition of the sample space, Ω.

5.4 Probability on $\mathbb{R}$

Definition 5.8
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \mathcal{F})$. The function

$$\begin{aligned} \mathbb{P}_X : \quad \mathcal{B} &\to [0, 1] \\ B &\mapsto \mathbb{P}(X \in B). \end{aligned}$$

is called the *distribution* of X.

Theorem 5.9
$\mathbb{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.

Proof: First we need to show that $\mathbb{P}_X(\mathbb{R}) = 1$:

$$\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}\big(\{\omega : X(\omega) \in \mathbb{R}\}\big) = 1.$$

We also need to show that $\mathbb{P}_X$ is countably additive. If $B_1, B_2, \dots$ is a sequence of pairwise disjoint sets in $\mathcal{B}$, then

$$\begin{aligned}
\mathbb{P}_X\big(\textstyle\bigcup_{i=1}^{\infty} B_i\big) &= \mathbb{P}\big(\{\omega : X(\omega) \in \textstyle\bigcup_{i=1}^{\infty} B_i\}\big) \\
&= \mathbb{P}\big(\textstyle\bigcup_{i=1}^{\infty}\{\omega : X(\omega) \in B_i\}\big) \\
&= \sum_{i=1}^{\infty} \mathbb{P}\big(\{\omega : X(\omega) \in B_i\}\big) \quad \text{because the } B_i \text{ are disjoint,} \\
&= \sum_{i=1}^{\infty} \mathbb{P}_X(B_i),
\end{aligned}$$

which concludes the proof.

Remark 5.10
A random variable X transforms an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into a more tractable probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, where we can apply the methods of *real analysis*.

5.5 Exercises

Exercise 5.1

1. Let Ω be the sample space of some random experiment, and let $\mathcal{F}$ be a σ-field over Ω.
   (a) For any $A \in \mathcal{F}$, show that the function $X : \Omega \to \mathbb{R}$, defined by

   $$X(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$

   is a random variable on $(\Omega, \mathcal{F})$.
   (b) Let $A_1, A_2, \dots, A_n \in \mathcal{F}$ be a partition of Ω and let $a_1, a_2, \dots, a_n \in \mathbb{R}$. Show that the function $X : \Omega \to \mathbb{R}$, defined by

   $$X(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega) \qquad \text{where} \qquad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$

   is a random variable on $(\Omega, \mathcal{F})$.

Lecture 6 Distributions

To be read in preparation for the **11.00** lecture on **Wed 15 Oct** in **Physiology A**.

6.1 Probability on the real line

Let $X : \Omega \to \mathbb{R}$ be a random variable, and recall the probability measure on $(\mathbb{R}, \mathcal{B})$, defined by

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big),$$

where $\mathcal{B}$ is the Borel σ-field over $\mathbb{R}$.

Definition 6.1

(1) The *distribution* of X is the probability measure $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$.

(2) The *cumulative distribution function* (CDF) of X is the function $F(x) = \mathbb{P}(X \leq x)$.

(3) The *survival function* (SF) of X is the function $S(t) = \mathbb{P}(X > t)$.

Remark 6.2

The survival function is also called the *complementary* distribution function. If X represents the *lifetime* of some random system, then $S(t) = \mathbb{P}(X > t)$ is the probability that the system survives beyond time t. In this context, $F(t) = 1 - S(t)$ is called the *lifetime distribution function.*

6.2 Cumulative distribution functions (CDFs)

Proposition 5.3 states that $X : \Omega \to \mathbb{R}$ is a random variable if and only if the sets $\{X \leq x\}$ are *events* over Ω:

$$\{X \leq x\} = \big\{\omega : X(\omega) \leq x\big\} \in \mathcal{F} \quad \text{for all} \quad x \in \mathbb{R}.$$

It can be shown that the probability measure

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big),$$

is uniquely defined by the values it takes on the events $\{X \leq x\}$ for $x \in \mathbb{R}$. Consequently, the distribution of a random variable is uniquely determined by its *cumulative distribution function* (CDF):

Definition 6.3
The *cumulative distribution function* (CDF) of a random variable $X : \Omega \to \mathbb{R}$ is the function

$$\begin{array}{rrcl} F: & \mathbb{R} & \longrightarrow & [0,1] \\ & x & \mapsto & \mathbb{P}(X \leq x). \end{array}$$

Theorem 6.4
Let $F : \mathbb{R} \to [0,1]$ be a CDF. Then there is a unique probability measure $\mathbb{P}_F : \mathcal{B} \to [0,1]$ on the real line with the property that

$$\mathbb{P}_F\big((a,b]\big) = F(b) - F(a)$$

for every such half-open interval $(a,b] \in \mathcal{B}$.

[*Proof omitted.*]

- The triple $(\mathbb{R}, \mathcal{B}, \mathbb{P}_F)$ is sometimes called the *probability space induced by* F.

Remark 6.5
Compare the probability measure $\mathbb{P}_F$ of the interval $(a,b] \subset \mathbb{R}$ to the usual measure of its *length*:

- Length: $\mathbb{L}\big((a,b]\big) = b - a$
- Probability measure: $\mathbb{P}_F\big((a,b]\big) = F(b) - F(a)$.

Thus $\mathbb{P}_F\big((a,b]\big)$ quantifies the "amount of probability" in any given interval $(a,b]$.

6.3 Properties of CDFs

Theorem 6.6
A cumulative distribution function $F : \mathbb{R} \to [0,1]$ has the following properties:

(1) if $x < y$ then $F(x) \leq F(y)$,

(2) $F(x) \to 0$ as $x \to -\infty$,

(3) $F(x) \to 1$ as $x \to +\infty$, and

(4) $F(x+h) \to F(x)$ as $h \downarrow 0$ (right continuity).

Proof:

(1) To show that F is increasing, let $x < y$ and consider the events

$$\begin{aligned} A &= \{X \leq x\} &= \{\omega : X(\omega) \leq x\}, \\ B &= \{X \leq y\} &= \{\omega : X(\omega) \leq y\}. \end{aligned}$$

By construction, $F(x) = \mathbb{P}(A)$ and $F(y) = \mathbb{P}(B)$ so by the monotonicity of probability measures (Theorem 3.8),

$$x < y \Rightarrow A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B) \Rightarrow F(x) \leq F(y).$$

(2) To show that $F(x) \to 0$ as $x \to -\infty$, let

$$B_n = \{X \leq -n\} = \{\omega : X(\omega) \in (-\infty, -n]\} \quad \text{for } n = 1, 2, \ldots$$

so that $F(-n) = \mathbb{P}(X \leq -n) = \mathbb{P}(B_n)$.

The sequence $B_1, B_2, \ldots$ is decreasing ($B_{n+1} \subseteq B_n$), with

$$\bigcap_{n=1}^{\infty} B_n = \emptyset,$$

because for any x, there exists an n such that $x \notin (-\infty, -n]$.

By the continuity of probability measures (Theorem 3.9),

$$\lim_{n\to\infty} F(-n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{n=1}^{n} B_n\right) = \mathbb{P}(\emptyset) = 0,$$

and because $F(x)$ is an increasing function,

$$\lim_{n\to\infty} F(-n) = 0 \quad \Leftrightarrow \quad \lim_{x\to-\infty} F(x) = 0.$$

(3) To show that $F(x) \to 1$ as $x \to \infty$, let

$$A_n = \{X \le n\} = \{\omega : X(\omega) \in (-\infty, n]\} \quad \text{for } n = 1, 2, \ldots,$$

so that $F(n) = \mathbb{P}(X \le n) = \mathbb{P}(A_n)$.

The sequence $A_1, A_2, \ldots$ is increasing ($A_n \subseteq A_{n+1}$), with

$$\bigcup_n A_{n=1}^{\infty} = \Omega,$$

because for any x, there exists an n such that $x \in (-\infty, n]$.

By the continuity of probability measures,

$$\lim_{n\to\infty} F(n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(\Omega) = 1,$$

and because $F(x)$ is an increasing function,

$$\lim_{n\to\infty} F(n) = 1 \quad \Leftrightarrow \quad \lim_{x\to\infty} F(x) = 1.$$

(4) To show that $F(x)$ is right-continuous, let

$$B_n = \left\{\omega : X(\omega) \in \left(-\infty, x + \frac{1}{n}\right]\right\}$$

so that $F(x + 1/n) = \mathbb{P}(X \le x + 1/n) = \mathbb{P}(B_n)$.

The sequence $B_1, B_2, \ldots$ is decreasing ($B_{n+1} \subseteq B_n$), with

$$\bigcap_{n=1}^{\infty} B_n = (-\infty, x],$$

so

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \mathbb{P}\big(\{\omega : X(\omega) \in (-\infty, x]\}\big) = \mathbb{P}\big(\{\omega : X(\omega) \le x\}\big) = F(x).$$

By the continuity of probability measures,

$$F(x) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} F\left(x + \frac{1}{n}\right).$$

which concludes the proof.

Theorem 6.7
Let $F : \mathbb{R} \to [0, 1]$ be a function with properties (i)-(iv) of Theorem 6.6. Then F is a cumulative distribution function.

[*Proof omitted.*]

Remark 6.8
The last two theorems make no explicit reference to random variables:

- many different random variables can have the same distribution function;
- a distribution function can represent many different random variables.

6.4 Discrete distributions and PMFs

The *range* of a random variable $X : \Omega \to \mathbb{R}$ is the set of all possible values it can take:

$$\text{Range}(X) = \{x \in \mathbb{R} : X(\omega) = x \text{ for some } \omega \in \Omega\}.$$

Definition 6.9

- $X : \Omega \to \mathbb{R}$ is called a *discrete random variable* if its range is a countable subset of $\mathbb{R}$.
- A discrete random variable is described by its *probability mass function* (PMF),

$$\begin{array}{rccc} f: & \mathbb{R} & \to & [0,1] \\ & k & \mapsto & \mathbb{P}(X = k), \end{array}$$

  which must have the property that $\sum_k f(k) = 1$.
- A probability mass function defines a *discrete probability measure* on $\mathbb{R}$,

$$\begin{array}{rccc} \mathbb{P}_X: & \mathcal{B} & \to & [0,1] \\ & B & \mapsto & \displaystyle\sum_{k \in B} \mathbb{P}(X = k), \end{array}$$

- The triple $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ is called a *discrete probability space* over $\mathbb{R}$.

6.5 Continuous distributions and PDFs

Definition 6.10

- A cumulative distribution function $F : \mathbb{R} \to [0, 1]$ is said to be *absolutely continuous* if there exists an integrable function $f : \mathbb{R} \to [0, \infty)$ such that

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

- The function $f : \mathbb{R} \to [0, \infty)$ is called the *probability density function* (PDF) of F.
- The triple $(\mathbb{R}, \mathcal{B}, \mathbb{P}_F)$ is called a *continuous probability space* over $\mathbb{R}$.

Definition 6.11
A *continuous random variable* is one whose distribution function is absolutely continuous.

If $X : \Omega \to \mathbb{R}$ is a continuous random variable, then

- $f(x) = F'(x)$ for all $x \in \mathbb{R}$.
- Probabilities correspond to areas under the curve $f(x)$:

$$\mathbb{P}_X\big((a, b]\big) = \mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\, dx.$$

- Note that $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.

Remark 6.12
The continuity of a random variable $X : \Omega \to \mathbb{R}$ refers to the continuity of its distribution function, and *not* to the continuity (or otherwise) of itself as a function on Ω.

6.6 Exercises

Exercise 6.1

1. Let F and G be CDFs, and let $0 < \lambda < 1$ be a constant. Show that $H = \lambda F + (1 - \lambda)G$ is also a CDF.
2. Let X_1 and X_2 be the numbers observed in two independent rolls of a fair die. Find the PMF of each of the following random variables:
   (a) $Y = 7 - X_1$,
   (b) $U = \max(X_1, X_2)$,
   (c) $V = X_1 - X_2$.
   (d) $W = |X_1 - X_2|$.
3. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^2 & 1 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$
   (a) Find the value of the constant c, and sketch the PDF of X.
   (b) Find the value of $P(X > 3/2)$.
   (c) Find the CDF of X.
4. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^{-d} & \text{for } x > 1, \\ 0 & \text{otherwise.} \end{cases}$
   (a) Find the range of values of d for which $f(x)$ is a probability density function.
   (b) If $f(x)$ is a density function, find the value of c, and the corresponding CDF.
5. Let $f(x) = \dfrac{ce^x}{(1 + e^x)^2}$ be a PDF, where c is a constant. Find the value of c, and the corresponding CDF.
6. Let $X_1, X_2, \ldots$ be independent and identically distributed observations, and let F denote their common CDF. If F is unknown, describe and justify a way of estimating F, based on the observations. [Hint: consider the indicator variables of the events $\{X_j \leq x\}$.]

Lecture 7 Transformations

To be read in preparation for the **11.00** lecture on **Mon 20 Oct** in **E/0.15**.

7.1 Transformations of random variables

Let $\mathcal{B}$ denote the Borel σ-field over $\mathbb{R}$.

Definition 7.1
A function $g : \mathbb{R} \to \mathbb{R}$ is said to be a *measurable function* if $g^{-1}(B) \in \mathcal{B}$ for all $B \in \mathcal{B}$.

Theorem 7.2
Let $X : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \mathcal{F})$, and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then the function $Y = g(X)$, defined by

$$\begin{aligned} Y : \quad \Omega &\to \mathbb{R} \\ \omega &\mapsto g\big[X(\omega)\big], \end{aligned}$$

is also a random variable on $(\Omega, \mathcal{F})$.

Proof: Let $B \in \mathcal{B}$. Then

$$\{Y \in B\} \equiv \{\omega : Y(\omega) \in B\} = \{\omega : g\big[X(\omega)\big] \in B\} = \{\omega : X(\omega) \in g^{-1}(B)\} \equiv \{X \in g^{-1}(B)\}$$

- X is a random variable, so the event $\{X \in B'\} \in \mathcal{F}$ for all $B' \in \mathcal{B}$.
- Since $B \in \mathcal{B}$ and g is a measurable function, it follows that $g^{-1}(B) \in \mathcal{B}$.
- Thus $\{X \in g^{-1}(B)\} \in \mathcal{F}$, because this is true for *every* Borel set.

This holds for every $B \in \mathcal{B}$, so $\{Y \in B\} \in \mathcal{F}$ for all $B \in \mathcal{B}$, as required.

We say that $Y = g(X)$ is a *transformation* of X.

- If we know the distribution of X, how can we deduce the distribution of Y?

In fact, the distribution of $Y = g(X)$ is completely determined by the distribution of X:

$$\begin{aligned}
\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) &= \mathbb{P}\big[\{\omega : Y(\omega) \in B\}\big] \\
&= \mathbb{P}\big[\{\omega : g[X(\omega)] \in B\}\big] \\
&= \mathbb{P}\big[\{\omega : X(\omega) \in g^{-1}(B)\}\big] \\
&= \mathbb{P}\big[X \in g^{-1}(B)\big] \\
&= \mathbb{P}_X\big[g^{-1}(B)\big]
\end{aligned}$$

Remark 7.3

- Theorem 7.2 shows that $g(X) : \Omega \to \mathbb{R}$ is a random variable over $(\Omega, \mathcal{F})$.
- We can also think of $g : \mathbb{R} \to \mathbb{R}$ as a random variable over $(\mathbb{R}, \mathcal{B})$, whose distribution is given by

$$\mathbb{P}(g \in B) = \mathbb{P}\big[g(X) \in B\big] = \mathbb{P}\big[X \in g^{-1}(B)\big] = \mathbb{P}_X\big[g^{-1}(B)\big],$$

  where $\mathbb{P}_X : \mathcal{B} \to [0,1]$ is the distribution of X.
- The distribution of g over $(\mathbb{R}, \mathcal{B})$ is well-defined, because $g^{-1}(B) \in \mathcal{B}$ for all $B \in \mathcal{B}$.

7.2 Support

PMFs and many PDFs are defined to be zero over certain subsets of $\mathbb{R}$. We must ensure that the PMF or PDF of a transformed variable is defined correctly, over appropriate subsets of $\mathbb{R}$.

Definition 7.4

(1) A set $A \subset \mathbb{R}$ is said to be *closed* if it contains all its limit points.

(2) The *support* of an arbitrary function $h : \mathbb{R} \to \mathbb{R}$, denoted by $\text{supp}(h)$, is the smallest closed set for which $h(x) = 0$ for all $x \notin \text{supp}(h)$.

(3) The *support* of a random variable $X : \Omega \to \mathbb{R}$ is defined to be the support of its PMF (discrete case) or PDF (continuous case), denoted by $\text{supp}(f_X)$. This is the smallest closed set that contains the *range* of X.

Remark 7.5

Let X be a random variable and let $g : \mathbb{R} \to \mathbb{R}$. The support of $Y = g(X)$ is the set

$$\text{supp}(f_Y) = \big\{g(x) : x \in \text{supp}(f_X)\big\}.$$

Example 7.6

The PDF of the continuous uniform distribution on $[0,1]$ is

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

- The support of X is $\text{supp}(f_X) = [0,1]$
- For the transformation $g(x) = x^2 + 2x + 3$, the support of $Y = g(X)$ is

$$\text{supp}(f_Y) = \{x^2 + 2x + 3 : x \in [0,1]\} = [3,6].$$

7.3 Transformations of CDFs

Theorem 7.7
Let X be a continuous random variable, and let f_X denote its PDF. Let $g : \mathbb{R} \to \mathbb{R}$ be an injective transformation over $\text{supp}(f_X)$ and let $Y = g(X)$. Finally, let $F_X(x)$ and $F_Y(y)$ respectively denote the CDFs of X and Y.

(1) If g is an increasing function, $F_Y(y) = F_X\big[g^{-1}(y)\big]$.

(2) If g is a decreasing function, $F_Y(y) = 1 - F_X\big[g^{-1}(y)\big]$.

Proof:

(1) If g is increasing, $g(x) \leq y$ implies that $x \leq g^{-1}(y)$ so

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\big[g(X) \leq y\big] = \mathbb{P}\big[X \leq g^{-1}(y)\big] = F_X\big[g^{-1}(y)\big].$$

(2) If g is decreasing, $g(x) \leq y$ implies that $x \geq g^{-1}(y)$ so

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\big[g(X) \leq y\big] &= \mathbb{P}\big[X \geq g^{-1}(y)\big] \\ &= 1 - \mathbb{P}\big[X \leq g^{-1}(y)\big] \quad \text{(because } X \text{ is a continuous r.v.)} \\ &= 1 - F_X\big[g^{-1}(y)\big]. \end{aligned}$$

7.4 Transformations of PMFs and PDFs

Theorem 7.8 (Transformations of PMFs)
Let X be a discrete random variable and let f_X denote its PMF. Let $g : \mathbb{R} \to \mathbb{R}$ be an injective transformation over $\text{supp}(f_X)$, and let $Y = g(X)$. Then the PMF of Y is given by

$$f_Y(y) = f_X\big[g^{-1}(y)\big] \quad \text{for all} \quad y \in \text{supp}(f_Y).$$

Proof:

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}\big[g(X) = y\big] = \mathbb{P}\big[X = g^{-1}(y)\big] = f_X\big[g^{-1}(y)\big].$$

Theorem 7.9 (Transformations of PDFs)
Let X be a continuous random variable and let f_X denote its PDF. Let $g : \mathbb{R} \to \mathbb{R}$ be an injective transformation over $\text{supp}(f_X)$, and let $Y = g(X)$. Then, if the derivative of $g^{-1}(y)$ is continuous and non-zero over $\text{supp}(f_Y)$, the PDF of Y is given by

$$f_Y(y) = f_X\big[g^{-1}(y)\big]\left|\frac{d}{dy}g^{-1}(y)\right| \quad \text{for all} \quad y \in \text{supp}(f_Y).$$

Remark 7.10

- The transformation is equivalent to making a *change of variable* in an integral.
- The scale factor $\left|\frac{d}{dy}g^{-1}(y)\right|$ ensures that $f_Y(y)$ integrates to one.

Proof: For brevity of notation, let $h(y)$ denote the inverse function $g^{-1}(y)$.

(1) If g is increasing, $F_Y(y) = F_X\big[g^{-1}(y)\big]$, so by the chain rule,

$$\begin{aligned} f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X\big[h(y)\big] \\ &= \frac{d}{dh(y)}F_X\big[h(y)\big]\cdot\frac{dh(y)}{dy} \\ &= f_X\big[h(y)\big]\left|\frac{dh(y)}{dy}\right|, \qquad \text{because } \frac{dh(y)}{dy} > 0 \text{ over } \mathrm{supp}(f_Y). \end{aligned}$$

(2) If g decreasing, $F_Y(y) = 1 - F_X\big[g^{-1}(y)\big]$, so by the chain rule,

$$\begin{aligned} f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}\big[1 - F_X[h(y)]\big] \\ &= 0 - \frac{d}{dh(y)}F_X\big[h(y)\big]\cdot\frac{dh(y)}{dy} \\ &= -f_X\big[h(y)\big]\frac{dh(y)}{dy} \\ &= f_X\big[h(y)\big]\left|\frac{dh(y)}{dy}\right|, \qquad \text{because} \frac{dh(y)}{dy} < 0 \text{ over } \mathrm{supp}(f_Y). \end{aligned}$$

7.5 The probability integral transform

Theorem 7.11 (The Probability Integral Transform)
Let X be a continuous random variable, let $F(x)$ denote its CDF, and suppose that the inverse F^{-1} of the CDF exists for all $x \in \mathbb{R}$. Then the random variable $Y = F(X)$ has the continuous uniform distribution on $[0, 1]$.

Proof: Since $F(x) = P(X \leq x)$ is a CDF, we know that $F(x) \in [0, 1]$ for all $x \in \mathbb{R}$. In particular, $P(Y < 0) = 0$ and $P(Y > 1) = 0$. For $y \in [0, 1]$, because the inverse F^{-1} exists we have that

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P\big(F(X) \leq y\big) \\ &= P\big(X \leq F^{-1}(y)\big) \\ &= F\big(F^{-1}(y)\big) \\ &= y, \end{aligned}$$

which is the CDF of the continuous uniform distribution on $[0, 1]$.

Corollary 7.12
Let $F(x)$ be a CDF whose inverse exists for all $x \in \mathbb{R}$, and let $Y \sim \text{Uniform}(0, 1)$. Then F is the CDF of the random variable $X = F^{-1}(Y)$.

- Uniformly distributed pseudo-random numbers in $[0, 1]$ can be generated using sophisticated algorithms.
- Using the probability integral transform, we can convert uniformly distributed pseudo-random samples to pseudo-random samples from other (continuous) distributions:

(1) Generate uniformly distributed pseudo-random numbers $u_1, u_2, \ldots, u_n$ in $[0, 1]$.

(2) Compute $x_i = F^{-1}(u_i)$ for $i = 1, 2, \ldots, n$.

The set $\{x_1, x_2, \ldots, x_n\}$ is a pseudo-random sample from the distribution whose CDF is $F(x)$.

Example 7.13
Given an algorithm that generates uniformly distributed pseudo-random numbers in the range $[0,1]$, show how to generate a pseudo-random sample from the exponential distribution with scale parameter $1/2$.

Solution: The CDF of the exponential distribution with scale parameter $1/2$ (or equivalently, with rate parameter 2) is

$$F(x) = \begin{cases} 1 - e^{-2x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

First we invert F:

$$\begin{aligned} u = 1 - e^{-2x} &\Rightarrow e^{-2x} = 1 - u \\ &\Rightarrow e^{x} = \frac{1}{\sqrt{1-u}} \\ &\Rightarrow x = \log\left(\frac{1}{\sqrt{1-u}}\right) \end{aligned}$$

Having done this, we generate a pseudo-random sample $u_1, u_2, \ldots, u_n$ from the Uniform$(0,1)$ distribution, then compute

$$x_i = \log\left(\frac{1}{\sqrt{1-u}}\right) \quad \text{for each} \quad i = 1, 2, \ldots, n.$$

7.6 Exercises

Exercise 7.1

1. Let X be a discrete random variable, with PMF $f_X(-2) = 1/3$, $f_X(0) = 1/3$, $f_X(2) = 1/3$, and zero otherwise. Find the distribution of $Y = X + 3$.
2. Let $X \sim \text{Binomial}(n, p)$ and define $g(x) = n - x$. Show that $g(X) \sim \text{Binomial}(n, 1-p)$.
3. Let X be a random variable, and let F_X denote its CDF. Find the CDF of $Y = X^2$ in terms of F_X.
4. Let X be a random variable with the following CDF:

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{x^3} & \text{for } x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the CDF of the random variable $Y = 1/X$, and describe how a pseudo-random sample from the distribution of Y can be obtained using an algorithm that generates uniformly distributed pseudo-random numbers in the range $[0,1]$.

Lecture 8 Examples of transformations

To be read in preparation for the **11.00** lecture on **Wed 22 Oct** in **Physiology A**.

8.1 Standard normal CDF $\longrightarrow$ Normal CDF

Example 8.1
Let $Z \sim N(0,1)$. Find the CDF of $X = \mu + \sigma Z$ in terms of the CDF of Z.

Solution: Let F_X and F_Z respectively denote the CDFs of X and Z.

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\mu + \sigma Z \leq x) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right).$$

8.2 Standard normal CDF $\longrightarrow$ Chi-squared CDF

Example 8.2 (The chi-squared distribution)
Let $X \sim N(0,1)$. Find the CDF of $Y = X^2$.

Solution: The PDF of $X \sim N(0,1)$ is given by $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Hence,

$$\mathbb{P}(Y \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\,dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-x^2/2}\,dx.$$

Using a change-of-variable $t = x^2$, we obtain

$$\mathbb{P}(Y \leq y) = \int_0^y \left(\frac{1}{\sqrt{2\pi t}} e^{-t/2}\right) dt.$$

This is the CDF of the *chi-squared* distribution with one degree of freedom.

8.3 Standard uniform CDF $\longrightarrow$ Exponential CDF

Example 8.3
Let $X \sim \text{Uniform}[0,1]$, and let $Y = -\theta \log X$ where $\theta > 0$. Show that $Y \sim \text{Exponential}(\theta)$, where θ is a scale parameter.

Solution: Consider the transformation $g(x) = -\theta \log x$.

- The CDF of X is $F_X(x) = x$ for $x \in [0,1]$ (with $F_X(x) = 0$ for $x < 0$, and $F_X(x) = 1$ for $x > 1$).
- The PDF of X is $f_X(x) = 1$ for $x \in [0,1]$ (and zero otherwise).
- $g(x)$ is a decreasing function (and thus has a unique inverse) over $\text{supp}(f_X) = [0,1]$.
- The support of f_Y is $\text{supp}(f_Y) = \{g(x) : x \in \text{supp}(f_X)\} = [0,\infty)$.
- The inverse transformation is $g^{-1}(y) = e^{-y/\theta}$ over $\text{supp}(f_Y)$.

By Theorem 7.7,

$$F_Y(y) = 1 - F_X\left[g^{-1}(y)\right] = 1 - F_X(e^{-y/\theta}) = 1 - e^{-y/\theta} \quad \text{for} \quad y > 0.$$

This is the CDF of the Exponential(θ) distribution, where θ is a scale parameter.

8.4 Exponential CDF $\longrightarrow$ Pareto CDF

Example 8.4
Let $X \sim \text{Exponential}(\alpha)$ where α is a rate parameter, and let $Y = \theta e^X$, where $\theta > 0$ is a constant. Show that Y has the so-called *Pareto*(θ,α) distribution, whose CDF is given by

$$F_Y(y) = \begin{cases} 1 - \left(\frac{\theta}{y}\right)^\alpha & \text{for } y > \theta \\ 0 & \text{otherwise.} \end{cases}$$

Solution: Consider the transformation $g(x) = \theta e^x$.

- The CDF of X is $F_X(x) = 1 - e^{-\alpha x}$ for $x > 0$ (and zero otherwise).
- $\text{supp}(f_X) = [0,\infty)$.
- $\text{supp}(f_Y) = \{\theta e^x : x \geq 0\} = [\theta,\infty)$.
- $g(x)$ is an increasing function over $\text{supp}(f_X)$; the inverse transformation is $g^{-1}(y) = \log(y/\theta)$.

By Theorem 7.7,

$$F_Y(y) = F_X\left[\log\left(\frac{y}{\theta}\right)\right] = 1 - \exp\left[-\alpha \log\left(\frac{y}{\theta}\right)\right] = \begin{cases} 1 - \left(\dfrac{\theta}{y}\right)^\alpha & \text{for } y > \theta, \\ 0 & \text{otherwise.} \end{cases}$$

as required.

Remark 8.5
Compare the upper-tail probabilities of $X \sim \text{Exponential}(\alpha)$ and $Y \sim \text{Pareto}(\theta,\alpha)$:

$$\mathbb{P}(X > x) = e^{-\alpha x} \qquad \text{and} \qquad \mathbb{P}(Y > y) = \theta^\alpha y^{-\alpha}.$$

In both cases, the rate at which the tail probabilities converge to zero is controlled by the parameter α. However, we can see that $\mathbb{P}(X > x) \to 0$ relatively quickly as $x \to \infty$, the rate of convergence depending "exponentially" on x, while $\mathbb{P}(Y > y) \to 0$ more slowly as $y \to \infty$, with the rate of convergence depending "polynomially" on y. Consequently, the Pareto distribution belongs to the class of *heavy-tailed* distributions.

8.5 Normal PDF $\longrightarrow$ Standard normal PDF

Example 8.6 (The standard normal distribution)
Let $X \sim N(\mu, \sigma^2)$, and define $Z = (X - \mu)/\sigma$. Find the PDF of Z.

Solution: Let $g(X) = \dfrac{X - \mu}{\sigma}$.

- The PDF of X is $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$.
- Because $\sigma > 0$, we see that $g(x)$ is increasing over $\text{supp}(f_X) = (-\infty, \infty)$.
- $g(x)$ has inverse function $g^{-1}(z) = \mu + \sigma z$.
- It is easy to check that $\dfrac{d}{dz} g^{-1}(z) = \sigma$.

Hence,

$$\begin{aligned} f_Z(z) &= f_X\left[g^{-1}(z)\right] \left|\frac{d}{dz} g^{-1}(z)\right| \\ &= \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(\mu + \sigma z) - \mu}{\sigma}\right)^2\right] = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \end{aligned}$$

8.6 Pareto PDF $\longrightarrow$ Standard uniform PDF

Example 8.7 (The Pareto distribution)
The Pareto(θ, α) distribution is a continuous distribution with PDF

$$f_X(x) = \begin{cases} \dfrac{\alpha}{\theta}\left(\dfrac{\theta}{x}\right)^{\alpha+1} & \text{for } x > \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Let $X \sim \text{Pareto}(1, 1)$. Find the PDF of $Y = 1/X$.

Solution: $X \sim \text{Pareto}(1, 1)$ has PDF

$$f_X(x) = \frac{1}{x^2} \quad \text{for } x > 1 \text{ (and zero otherwise).}$$

Let $g(x) = 1/x$.

- $g(x)$ is a monotonic decreasing function over $x > 1$.
- The inverse function is $g^{-1}(y) = 1/y$.
- $\text{supp}(f_Y) = \{g(x) : x \in \text{supp}(f_X)\} = \{1/x : x \in (0, \infty)\} = (0, 1)$.

Hence the PDF of Y is given by

$$\begin{aligned} f_Y(y) = f_X\left[g^{-1}(y)\right] \left|\frac{d}{dy} g^{-1}(y)\right| &= f_X\left(\frac{1}{y}\right) \left|\frac{d}{dy}\left(\frac{1}{y}\right)\right| \\ &= y^2 \left|-\frac{1}{y^2}\right| = \begin{cases} 1 & \text{for } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus Y has the standard uniform distribution.

8.7 Normal PDF $\longrightarrow$ Lognormal PDF

Example 8.8 (The lognormal distribution)
If $X \sim (\mu, \sigma^2)$, then $Y = e^X$ is said to have *lognormal* distribution. Find the PDF of Y.

Solution: Let $g(X) = e^X$.

- Recall that $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$.
- $g(x)$ is an increasing function over $\text{supp}(f_X) = (-\infty, \infty)$.
- $g(x) \in [0, \infty]$ for $x \in \mathbb{R}$ so $\text{supp}(f_Y) = [0, \infty)$.
- $g(x)$ has inverse function $g^{-1}(y) = \log y$, for which $\dfrac{d}{dy}g^{-1}(y) = \dfrac{1}{y}$.

Thus

$$\begin{aligned} f_Y(y) &= f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= f_X(\log y)\left|\frac{1}{y}\right| = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right] \end{aligned}$$

for $y > 0$, and zero otherwise.

8.8 Lomax PDF $\longrightarrow$ Logistic CDF

Example 8.9 (The logistic distribution)
The Lomax(θ, α) distribution[1] is a continuous distribution with PDF

$$f_X(x) = \frac{\alpha}{\theta}\left(1 + \frac{x}{\theta}\right)^{-(\alpha+1)} \quad \text{for } x > 0, \text{ and zero otherwise.}$$

Let $X \sim \text{Lomax}(1, 1)$. Show that the CDF of $Z = \log X$ is given by

$$F_Z(z) = \frac{e^z}{1 + e^z}.$$

This is the CDF of the *standard logistic distribution.*

Solution: Taking $\alpha = 1$ and $\theta = 1$, the PDF of X is

$$f_X(x) = \frac{1}{(1+x)^2}$$

Consider the transformation $g(x) = \log x$.

- $g(x)$ is an increasing function over $\text{supp}(f_X) = (0, \infty)$.
- $\text{supp}(f_Z) = \{g(x) : x \in \text{supp}(f_X)\} = \{\log x : x \in (0, \infty)\} = (-\infty, \infty)$.
- The inverse transformation is $g^{-1}(z) = e^z$, and its first derivative is $\dfrac{d}{dz}g^{-1}(z) = e^z$.

The PDF of $Z = \log X$ is therefore given by

$$f_Z(z) = f_X\left[g^{-1}(z)\right]\left|\frac{d}{dz}g^{-1}(z)\right| = \frac{e^z}{(1+e^z)^2},$$

from which we the required CDF follows by integration.

[1]The Lomax distribution is also known as the *Pareto Type II distribution* and the *shifted Pareto distribution*

8.9 Exercises

Exercise 8.1

1. Let $X \sim \text{Uniform}(-1,1)$. Find the CDF and PDF of X^2.
2. Let X have exponential distribution with rate parameter $\lambda > 0$. The PDF of X is

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the PDFs of $Y = X^2$ and $Z = e^X$.
3. Let $X \sim \text{Pareto}(1,2)$. Find the PDF of $Y = 1/X$.
4. A continuous random variable U has PDF

$$f(u) = \begin{cases} 12u^2(1-u) & \text{for} \quad 0 < u < 1, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the PDF of $V = (1-U)^2$.
5. The continuous random variable U has PDF

$$f_U(u) = \begin{cases} 1+u & -1 < u \leq 0, \\ 1-u & 0 < u \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the PDF of $V = U^2$. (Note that the transformation is not injective over $\text{supp}(f_U)$, so you should first compute the CDF of V, then derive its PDF by differentiation.)
6. Let X have exponential distribution with scale parameter $\theta > 0$. This has PDF

$$f(x) = \begin{cases} \frac{1}{\theta} \exp(-x/\theta) & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the PDF of $Y = X^{1/\gamma}$ where $\gamma > 0$.
7. Suppose that X has the *Beta Type I* distribution, with parameters $\alpha, \beta > 0$. This has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

   where $B(a,b) = \displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ is the so-called *beta function*. Show that the random variable $Y = \dfrac{X}{1-X}$ has the *Beta Type II* distribution, which has PDF

$$f_Y(y) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} \dfrac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Lecture 9 Series and Integrals

To be read in preparation for the **11.00** lecture on **Mon 27 Oct** in **E/0.15**.

9.1 Motivation

Discrete random variables

A discrete random variable can be represented by a sequence of real numbers.

- Let X take values in the set $\mathbb{N} = \{1, 2, 3, \ldots\}$, and let $p_k = \mathbb{P}(X = k)$.
- The PMF of X is the sequence $p_1, p_2, \ldots$.
- The only constraints on the sequence are that its terms p_k are all non-negative, and $\displaystyle\sum_{k=1}^{\infty} p_k = 1$.
- The expectation of X is given by the series $\mathbb{E}(X) = \displaystyle\sum_{k=1}^{\infty} k p_k$.
  - This series does not necessarily converge (to a finite value).
  - It may not even be well-defined.

Continuous random variables

A continuous random variables can be represented by a function $f : \mathbb{R} \to \mathbb{R}$.

- Let X be a continuous random variable, and let $F(x) = \mathbb{P}(X \leq x)$ be its CDF.
- The PDF of X is the function $f(x) = F'(x)$.
- The only constraints on f are that $f(x) \geq 0$ for all $x \in \mathbb{R}$, and $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$.
- The expectation of X is given by the integral $\displaystyle\int_{-\infty}^{\infty} x\, f(x)\, dx$.
  - This integral does not necessarily converge (to a finite value).
  - It may not even be well-defined.

9.2 Series

9.2.1 Convergent sequences

Definition 9.1
An infinite sequence of real numbers $a_1, a_2, \ldots$ is said to *converge* if there exists some $a \in \mathbb{R}$ such that for all $\epsilon > 0$, there exists some $N \in \mathbb{N}$ with

$$|a_n - a| < \epsilon \quad \text{for all} \quad n > N.$$

Remark 9.2
(1) The number a is called the *limit* of the sequence, written as $a = \lim_{n\to\infty} a_n$.

(2) A sequence that is not convergent is said to be *divergent*.

9.2.2 Convergent series

Definition 9.3
Let $a_1, a_2, \ldots$ be a sequence of real numbers. The infinite series $\sum_{n=1}^{\infty} a_n$ is said to be

(1) *convergent* if the sequence of partial sums $\sum_{n=1}^{m} a_n$ converges as $m \to \infty$,

(2) *absolutely convergent* if $\sum_{n=1}^{\infty} |a_n|$ is convergent,

(3) *conditionally convergent* if it is convergent, but is not absolutely convergent,

(4) *divergent* if the sequence of partial sums $\sum_{n=1}^{m} a_n$ diverges as $m \to \infty$.

Example 9.4

- $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ converges.
- $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. (This is the *harmonic series*.)
- $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \log 2$ is converges conditionally. (This is the *alternating harmonic series*.)

The alternating harmonic series is convergent,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \log 2,$$

but not absolutely convergent, because

$$\sum_{n=1}^{\infty} \left| \frac{(-1)^{n+1}}{n} \right| = \sum_{n=1}^{\infty} \frac{1}{n}.$$

9.2.3 Positive and negative parts

If a series $\sum_n a_n$ has both positive and negative terms, we can write it as the difference of two series of non-negative terms:

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-,$$

where

$$\begin{aligned} a_n^+ &= \max\{a_n, 0\} &&= \begin{cases} a_n & \text{if } a_n \geq 0, \\ 0 & \text{otherwise.} \end{cases} \\ a_n^- &= \max\{-a_n, 0\} &&= \begin{cases} -a_n & \text{if } a_n < 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

- $\sum_{n=1}^{\infty} a_n^+$ is called the *positive part* of the series.
- $\sum_{n=1}^{\infty} a_n^-$ is called the *negative part* of the series.

Since

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} a_n^+ + \sum_{n=1}^{\infty} a_n^-,$$

we see that

(1) If $\sum_n a_n$ is absolutely convergent, its positive and negative parts both converge.

(2) If $\sum_n a_n$ is conditionally convergent, its positive and negative parts both diverge.

Consider the alternating harmonic series:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \log 2,$$

- The positive and negative parts of the alternating harmonic series both diverge.
- There is sufficient cancellation between its terms to ensure that the series itself converges.

9.2.4 The Riemann rearrangement theorem

Definition 9.5

(1) A bijection $\phi : \mathbb{N} \to \mathbb{N}$ is called a *permutation* of the labels $\{1, 2, \ldots\}$.

(2) The *rearrangement* of the series $\sum_n a_n$ by the permutation ϕ is the series $\sum_n a_{\phi(n)}$.

Theorem 9.6 (The Riemann rearrangement theorem)
Let $\sum_n a_n$ be a convergent series.

(1) If $\sum_n a_n$ is absolutely convergent, then every rearrangement $\sum_n a_{\phi(n)}$ is absolutely convergent to the same limit.

(2) If $\sum_n a_n$ is conditionally convergent, then for every $a \in \mathbb{R} \cup \{\pm\infty\}$ there exists a permutation $\phi : \mathbb{N} \to \mathbb{N}$ such that $\sum_n a_{\phi(n)} = a$.

[*Proof omitted.*]

- The limit of conditionally convergent series depends on the *order* in which the terms of the series are added together.
- In probability theory, we cannot deal with sums that are conditionally convergent.
- Expectation is only defined if the series $\sum_{i=1}^{\infty} kp_k$ is absolutely convergent.

9.3 Integrals

9.3.1 The Riemann integral

Let $g : [a, b] \to \mathbb{R}$ be a bounded function. A *partition* of $[a, b]$ is a set of intervals

$$\mathcal{P} = \{[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]\}$$

where $a = x_0 < x_1 < x_2 < \ldots < x_n = b$.

The upper and lower *Riemann sums* of g with respect to $\mathcal{P}$ are, respectively,

$$\begin{aligned} U(\mathcal{P},g) &= \sum_{i=1}^{n} M_i\Delta_i \quad \text{where} \quad M_i = \sup\{g(x) : x\in[x_{i-1},x_i]\},\\ L(\mathcal{P},g) &= \sum_{i=1}^{n} m_i\Delta_i \quad \text{where} \quad m_i = \inf\{g(x) : x\in[x_{i-1},x_i]\},\end{aligned}$$

where $\Delta_i = x_i - x_{i-1}$ is the length of the interval $[x_{i-1},x_i]$.

The upper and lower *Riemann integrals* of g on $[a,b]$ are, respectively,

$$\underline{\int_a^b} g(x)\,dx = \sup_{\mathcal{P}} L(\mathcal{P},g) \quad \text{and} \quad \overline{\int_a^b} g(x)\,dx = \inf_{\mathcal{P}} U(\mathcal{P},g).$$

where the supremum and infimum are taken over all possible partitions of $[a,b]$.

If the upper and lower Riemann integrals coincide, we say that g is *Riemann integrable*, in which case their common value is called the *Riemann integral* of g, denoted by

$$\int_a^b g(x)\,dx.$$

To extend the definition to (1) integrals over unbounded intervals, and (2) integrals of unbounded functions, we use limits to define *improper* integrals:

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \lim_{n\to\infty}\int_{-n}^{n} e^{-x^2}\,dx,$$

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = \lim_{\epsilon\to 0}\int_{\epsilon}^{1}\frac{1}{\sqrt{x}}\,dx.$$

9.3.2 Integrable functions

A function $g:\mathbb{R}\to\mathbb{R}$ is said to be *integrable* (in the Riemann sense) if the area between the curve $g(x)$ and the horizontal axis is finite:

$$\int_{-\infty}^{\infty} |g(x)|\,dx < \infty.$$

If a function is not integrable, we say that its integral is *undefined* or *does not exist.*

Let $g:\mathbb{R}\to\mathbb{R}$ be an integrable function. The *integral* of g is the difference between the area above the horizontal axis and the area below the horizontal axis:

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} g^+(x)\,dx - \int_{-\infty}^{\infty} g^-(x)\,dx,$$

where g^+ and g^- are respectively the *positive part* and *negative part* of g:

$$\begin{aligned} g^+(x) &= \max\{g(x),0\} &&= \begin{cases} g(x) & \text{if } g(x)\geq 0,\\ 0 & \text{otherwise.}\end{cases}\\ g^-(x) &= \max\{-g(x),0\} &&= \begin{cases} -g(x) & \text{if } g(x)< 0,\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Note that:

- $g^+(x)\geq 0$ and $g^-(x)\geq 0$ for all $x\in\mathbb{R}$ (i.e. both are non-negative functions).
- $g(x) = g^+(x) - g^-(x)$ for all $x\in\mathbb{R}$.
- $|g(x)| = g^+(x) + g^-(x)$ for all $x\in\mathbb{R}$.

If g is integrable, then $\int g^+(x)\,dx$ and $\int g^-(x)\,dx$ are both finite:

$$\int g^+(x)\,dx + \int g^-(x)\,dx = \int g^+(x) + g^-(x)\,dx = \int |g(x)|\,dx < \infty.$$

If g is not integrable, one or both of $\int g^+(x)\,dx$ and $\int g^-(x)\,dx$ must be infinite:

- if $\int g^+(x)\,dx = \infty$ and $\int g^-(x)\,dx < \infty$, then $\int g(x)\,dx = +\infty$,
- if $\int g^+(x)\,dx < \infty$ and $\int g^-(x)\,dx = \infty$, then $\int g(x)\,dx = -\infty$,
- if $\int g^+(x)\,dx = \infty$ and $\int g^-(x)\,dx = \infty$, we say that $\int g(x)\,dx$ is *undefined*.

Example 9.7
Let $g : \mathbb{R} \to \mathbb{R}$ be the function $g(x) = \begin{cases} \sin x & \text{if } \ 0 \leq x \leq 2\pi \\ 0 & \text{otherwise.} \end{cases}$

- Positive part: $g^+(x) = \begin{cases} \sin x & \text{if } 0 \leq x \leq \pi, \\ 0 & \text{otherwise.} \end{cases}$.
- Negative part: $g^-(x) = \begin{cases} -\sin x & \text{if } \pi \leq x \leq 2\pi, \\ 0 & \text{otherwise.} \end{cases}$.

The area above and below the horizontal axis are respectively

$$\begin{aligned} A^+ &= \int_{-\infty}^{\infty} g^+(x)\,dx = \int_0^{\pi} \sin x\,dx = \big[-\cos x\big]_0^{\pi} = 2, \\ A^- &= \int_{-\infty}^{\infty} g^-(x)\,dx = \int_{\pi}^{2\pi} (-\sin x)\,dx = \big[\cos x\big]_{\pi}^{2\pi} = 2. \end{aligned}$$

Hence the integral of g over $\mathbb{R}$ is

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} g^+(x)\,dx - \int_{-\infty}^{\infty} g^-(x)\,dx = A^+ - A^- = 0.$$

9.3.3 The Riemann-Stieltjes integral

Let $g : [a,b] \to \mathbb{R}$ be a bounded function, let $\mathcal{P}$ be a partition of $[a,b]$ and let $F : \mathbb{R} \to [0,1]$ be a CDF.

The upper and lower *Riemann-Stieltjes sums* of g with respect to $\mathcal{P}$ and F are, respectively,

$$\begin{aligned} U(\mathcal{P}, g, F) &= \sum_{i=1}^{n} M_i \Delta_i \quad \text{where} \quad M_i = \sup\{g(x) : x \in [x_{i-1}, x_i]\}, \\ L(\mathcal{P}, g, F) &= \sum_{i=1}^{n} m_i \Delta_i \quad \text{where} \quad m_i = \inf\{g(x) : x \in [x_{i-1}, x_i]\}, \end{aligned}$$

where $\Delta_i = F(x_i) - F(x_{i-1})$ is the probability measure induced by F of the interval $[x_{i-1}, x_i]$.

The upper and lower *Riemann-Stieltjes integrals* of g on $[a,b]$ are, respectively,

$$\underline{\int_a^b} g(x)\,dF(x) = \sup_{\mathcal{P}} L(\mathcal{P}, g, F) \quad \text{and} \quad \overline{\int_a^b} g(x)\,dF(x) = \inf_{\mathcal{P}} U(\mathcal{P}, g, F).$$

where the supremum and infimum are taken over all possible partitions of $[a,b]$.

- If the upper and lower Riemann-Stieltjes integrals coincide, we say that g is *Riemann-Stieltjes integrable*.

- In this case, their common value is called the *Riemann-Stieltjes integral* of g, denoted by

$$\int_a^b g(x)\,dF(x).$$

Remark 9.8
Let F be the CDF of the *uniform* distribution on $[a,b]$:

$$F(x) = \begin{cases} 0 & x < a, \\ \dfrac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases}$$

In this case, for any interval $[x_{i-1}, x_i] \subseteq [a,b]$ the probability measure induced by F is equal to its length, and the Riemann-Stieltjes integral reduces to the ordinary Riemann integral.

Lecture 10 Expectation

To be read in preparation for the **11.00** lecture on **Wed 29 Oct** in **Physiology A**.

Expectation is to random variables what probability is to events.

- Random events are *sets*, and are studied using *set algebra*.
- Random variables are *functions*, and are studied using *mathematical analysis*.

Elementary probability theory provides the following computational formulae for the expectation of a random variable $X : \Omega \to \mathbb{R}$.

(1) If Ω is a finite sample space with probability mass function $p : \Omega \to \mathbb{R}$, the expectation of X is

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)p(\omega).$$

(2) If X is a discrete random variable with PMF $f(x)$ and range $\{x_1, x_2, \ldots\}$, the expectation of X is

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

(3) if X is a continuous random variable with PDF $f(x)$, the expectation of X is

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

- The convergence of such sums and integrals is not guaranteed under all circumstances.
- If X takes only non-negative values, we can accept that $\mathbb{E}(X) = \infty$.
- If X can take both positive and negative values, we need that $\mathbb{E}(X) < \infty$.

10.1 Indicator variables

Consider the indicator variable of an event A,

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

To be consistent with the probability measure $\mathbb{P}(A)$, the only reasonable definition of expectation for I_A is the following:

Definition 10.1 (Expectation of indicator variables)
The expectation of an indicator variable $I_A : \Omega \to \mathbb{R}$ is defined by

$$\mathbb{E}(I_A) = \mathbb{P}(A)$$

10.2 Simple random variables

Let X be a simple random variable, let $\{x_1, x_2, \ldots, x_n$ denote its range, and let $\{A_1, A_2, \ldots, A_n\}$ be a partition of Ω such that $X(\omega) = x_i$ for all $\omega \in A_i$. Then X can be expressed as a finite linear combination of indicator variables,

$$X(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega) \qquad \text{where} \quad I_{A_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in A_i, \\ 0 & \text{if } \omega \notin A_i. \end{cases}$$

Definition 10.2 (Expectation of simple random variables)
The expectation of a simple random variable $X = \sum_{i=1}^{n} x_i I_{A_i}$ is defined by

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(A_i).$$

Remark 10.3
It can be shown all representations of X as finite linear combinations of inidcator variables yield the same value for $\mathbb{E}(X)$, so the expectation of a simple random variable is well-defined.

Example 10.4
A fair coin is tossed three times. Let $X : \Omega \to \mathbb{R}$ be the random variable

$$X(\omega) = \begin{cases} 1 & \text{if } \ \omega \in A_1 = \{TTT\} \\ 2 & \text{if } \ \omega \in A_2 = \{TTH, THT, HTT\} \\ 3 & \text{if } \ \omega \in A_3 = \{THH, HTH, HHT\} \\ 4 & \text{if } \ \omega \in A_4 = \{HHH\} \end{cases}$$

Compute the expected value of X.

Solution: X is a simple random variable, so

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(A_i) = \left(1 \times \frac{1}{8}\right) + \left(2 \times \frac{3}{8}\right) + \left(3 \times \frac{3}{8}\right) + \left(4 \times \frac{1}{8}\right) = \frac{5}{2}.$$

Definition 10.5
Let X and Y be random variables on Ω.

(1) If $X(\omega) \geq 0$ for all $\omega \in \Omega$, we say that X is *non-negative*. This is denoted by $X \geq 0$.

(2) If $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, we say that X is *dominated* by Y. This is denoted by $X \leq Y$.

Theorem 10.6 (Properties of expectation for simple random variables)
Let $X, Y : \Omega \to \mathbb{R}$ be simple random variables.

(1) **Positivity**. If $X \geq 0$ then $\mathbb{E}(X) \geq 0$.

(2) **Linearity**. For every $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

(3) **Monotonicity**. If $X \leq Y$ then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Proof: Let $\{x_1, x_2, \ldots, x_n\}$ be the range of X, and let $\{y_1, y_2, \ldots, y_m\}$ be the range of Y. Then X and Y can be written as

$$X(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega) \quad \text{and} \quad Y(\omega) = \sum_{j=1}^{m} y_j I_{B_j}(\omega),$$

where $A_i = \{\omega : X(\omega) = x_i\}$ and $B_j = \{\omega : Y(\omega) = y_j\}$

(1) Positivity:

If $X(\omega) \geq 0$ for all ω, we must have that each $x_i \geq 0$. Thus $\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(A_i)$ is a sum of non-negative terms, so $\mathbb{E}(X) \geq 0$.

(2) Linearity:

The composite variable $aX + bY$ is a simple random variable, because it can be written as

$$aX(\omega) + bY(\omega) = \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_j) I_{A_i \cap B_j}(\omega).$$

The expectation of $aX + bY$ is therefore defined to be

$$\begin{aligned}
\mathbb{E}(aX + bY) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_j) \mathbb{P}(A_i \cap B_j) \\
&= a \sum_{i=1}^{n} x_i \sum_{j=1}^{m} \mathbb{P}(A_i \cap B_j) + b \sum_{j=1}^{m} y_j \sum_{i=1}^{n} \mathbb{P}(A_i \cap B_j).
\end{aligned}$$

By additivity:

- $\{A_i \cap B_j\}_{j=1}^{m}$ is a partition of A_i, so $\sum_{j=1}^{m} \mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)$.
- $\{A_i \cap B_j\}_{i=1}^{n}$ is a partition of B_j, so $\sum_{i=1}^{n} \mathbb{P}(A_i \cap B_j) = \mathbb{P}(B_j)$.

Thus,

$$\begin{aligned}
\mathbb{E}(aX + bY) &= a \sum_{i=1}^{n} x_i \mathbb{P}(A_i) + b \sum_{j=1}^{m} y_j \mathbb{P}(B_j) \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y).
\end{aligned}$$

(3) Monotonicity:

If $X \leq Y$ then $Y - X \geq 0$, so

- By positivity, $\mathbb{E}(Y - X) \geq 0$
- By linearity, $\mathbb{E}(Y) - \mathbb{E}(X) \geq 0$, so $\mathbb{E}(X) \leq \mathbb{E}(Y)$ as required.

Example 10.7
Extending example 10.4, let $Y : \Omega \to \mathbb{R}$ be the random variable

$$Y(\omega) = \begin{cases} 2 & \text{if} \quad \omega \in A'_1 = \{TTT, TTH\} \\ 3 & \text{if} \quad \omega \in A'_2 = \{THT, THH\} \\ 4 & \text{if} \quad \omega \in A'_3 = \{HTT, HTH\} \\ 5 & \text{if} \quad \omega \in A'_4 = \{HHT, HHH\} \end{cases}$$

Compute the expected value of (i) Y and (ii) $3X + 2Y$.

Solution: Note that X and Y are both non-negative, and that $X \leq Y$.
Y is a simple random variable, so its expected value is

$$\mathbb{E}(Y) = \sum_{j=1}^{4} b_j \mathbb{P}(B_j) = \left(2 \times \frac{1}{4}\right) + \left(3 \times \frac{1}{4}\right) + \left(4 \times \frac{1}{4}\right) + + \left(5 \times \frac{1}{4}\right) = \frac{14}{4},$$

which verifies that $\mathbb{E}(X) \leq \mathbb{E}(Y)$. By the linearity of expectation for simple random variables,

$$\mathbb{E}(3X + 2Y) = 3\mathbb{E}(X) + 2\mathbb{E}(Y) = \left(3 \times \frac{5}{2}\right) + \left(2 \times \frac{14}{4}\right) = \frac{29}{2}.$$

10.3 Non-negative random variables

Theorem 10.8
For every non-negative random variable $X \geq 0$, there exists an increasing sequence of simple non-negative random variables

$$0 \leq X_1 \leq X_2 \leq \dots$$

with the property that $X_n(\omega) \uparrow X(\omega)$ for each $\omega \in \Omega$ as $n \to \infty$.

[*Proof omitted.*]

Definition 10.9 (Expectation of non-negative random variables)
The *expectation* of a non-negative random variable X is defined to be

$$\mathbb{E}(X) = \lim_{n \to \infty} \mathbb{E}(X_n)$$

where the X_n are simple non-negative random variables with $X_n \uparrow X$ as $n \to \infty$.

Remark 10.10
It can be shown that all approximating sequences yield the same value for $\mathbb{E}(X)$, so the expectation of non-negative random variables is well-defined

Remark 10.11 (Infinite expectation)
The expectation of non-negative random variables can be infinite:

(1) If $\mathbb{E}(X) < \infty$ we say that X has *finite* expectation.

(2) If $\mathbb{E}(X) = \infty$ we say that X has *infinite* expectation.

Theorem 10.12 (Properties of expectation for non-negative random variables)
For non-negative random variables $X, Y \geq 0$,

(1) **Positivity**. If $X \geq 0$ then $\mathbb{E}(X) \geq 0$.

(2) **Linearity**. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for every $a, b \in \mathbb{R}_+$.

(3) **Monotonicity**. If $X \leq Y$ then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

(4) **Continuity**. If $X_n \to X$ as $n \to \infty$, where the X_n are non-negative, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$ as $n \to \infty$.

[*Proof omitted.*]

10.4 Signed random variables

We can extend the definition of expectation to random variables that take both positive and negative values, but only if the random variables are *integrable*:

Definition 10.13 (Integrable random variables)
A random variable X is said to be *integrable* if $\mathbb{E}(|X|) < \infty$.

Definition 10.14 (The positive and negative parts)
The positive and negative parts of a random variable X, denoted by X^+ and X^- respectively, are defined to be

$$\begin{aligned} X^+(\omega) &= \max\{0, X(\omega)\} &= \begin{cases} X(\omega) & \text{if } X(\omega) \geq 0, \\ 0 & \text{if } X(\omega) < 0; \end{cases} \\ X^-(\omega) &= \max\{0, -X(\omega)\} &= \begin{cases} -X(\omega) & \text{if } X(\omega) \leq 0, \\ 0 & \text{if } X(\omega) > 0. \end{cases} \end{aligned}$$

Note that X^+ and X^- are both non-negative random variables, with $X = X^+ - X^-$.

Definition 10.15 (Expectation of signed random variables)
The *expectation* of an integrable random variable $X : \Omega \to \mathbb{R}$ is

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

where X^+ and X^- are respectively the positive part and negative part of X:

Remark 10.16 (Undefined expectation)
Because $|X| = X^+ + X^-$, it follows by the linearity of expectation for non-negative random variables that

$$\mathbb{E}(|X|) = \mathbb{E}(X^+ + X^-) = \mathbb{E}(X^+) + \mathbb{E}(X^-).$$

(1) If X is integrable, then $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are both finite.

(2) If X is not integrable, then one or both of $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ must be infinite:

- if $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) < \infty$, we write $\mathbb{E}(X) = +\infty$;
- if $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) = \infty$, we write $\mathbb{E}(X) = -\infty$;
- if $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) = \infty$, we say that $\mathbb{E}(X)$ *does not exist*.

Theorem 10.17 (Properties of expectation for signed random variables)
Let X and Y be integrable random variables.

(1) **Monotonicity**. If $X \leq Y$, then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

(2) **Linearity**. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for all $a, b \in \mathbb{R}$.

[*Proof omitted.*]

10.5 Exercises

Exercise 10.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $0 \leq X_1 \leq X_2 \leq \ldots$ be an increasing sequence of non-negative random variables over $(\Omega, \mathcal{F})$ such that $X_n(\omega) \uparrow X(\omega)$ ans $n \to \infty$ for all $\omega \in \Omega$. Show that X is a random variable on $(\Omega, \mathcal{F})$.

2. Let X be an integrable random variable. Show that $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

3. Let X and Y be integrable random variables. Show that $aX + bY$ is integrable.

Lecture 11 Computation of Expectation

To be read in preparation for the **11.00** lecture on **Mon 03 Nov** in **E/0.15**.

11.1 Expectation with respect to CDFs

The natural definition of expectation for indicator variables (10.1) was extended to the expectation of simple random variables (10.2), then to non-negative variables (10.3) and finally to signed variables (10.4).

According to definition 10.9, to compute the expectation of a non-negative random variable X, we need to find an increasing sequence $0 \leq X_1 \leq X_2 \leq ...$ of simple random variables for which $X_n \uparrow X$ as $n \to \infty$, then compute the limit of $\mathbb{E}(X_n)$ as $n \to \infty$. This is not feasible in practical applications.

It turns out that the expectation of a random variable can be conventiently expressed as a Riemann-Stieltjes integral with respect to its CDF:

Theorem 11.1
Let $X : \Omega \to \mathbb{R}$ be a non-negative random variable, and let $F : \mathbb{R} \to [0, 1]$ denote its CDF. The expectation of X can be written as

$$\mathbb{E}(X) = \int_0^\infty x \, dF(x).$$

where the right-hand side is the Riemann-Stieltjes integral of x with respect to F.

[*Proof omitted.*]

The following theorem yields a computational formula for expectation, in terms of an ordinary Riemann integral:

Theorem 11.2
If X is non-negative,

$$\mathbb{E}(X) = \int_0^\infty 1 - F(x) \, dx$$

[*Proof omitted.*]

For signed random variables, we first compute $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ using to this formula, and then set $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$ as before. To do this, we must first find the CDFs of X^+ and X^-.

11.2 Discrete distributions

If X is discrete, the Riemann-Stieltjes integral of Theorem 11.1 reduces to a sum:

Theorem 11.3
Let X be a non-negative discrete random variable, let $\{x_1, x_2, \ldots\}$ be its range, and let $f(x)$ denote its PMF. Then

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i f(x_i),$$

provided the sum is absolutely convergent.

[*Proof omitted.*]

Remark 11.4
This expression also holds for signed discrete random variables. Can you prove this?

The following is a special case of Theorem 11.2:

Theorem 11.5
Let X be a discrete non-negative random variable, taking values in the range $\{0, 1, 2, \ldots\}$. Then

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Proof:

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbb{P}(X > k) &= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \mathbb{P}(X = j) \\
&= \sum_{j=1}^{\infty} \mathbb{P}(X = j) + \sum_{j=2}^{\infty} \mathbb{P}(X = j) + \sum_{j=3}^{\infty} \mathbb{P}(X = j) + \ldots \\
&= \mathbb{P}(X = 1) + 2\mathbb{P}(X = 2) + 3\mathbb{P}(X = 3) + \ldots \\
&= \sum_{j=0}^{\infty} jP(X = j) \\
&= \mathbb{E}(X).
\end{aligned}$$

Example 11.6 (Geometric distribution)
Suppose X has the geometric distribution on $\{0, 1, 2, \ldots\}$, with probability-of-success parameter p. Given that the CDF of X is $\mathbb{P}(X \leq k) = 1 - (1-p)^{k+1}$, show that its expected value is equal to $(1-p)/p$.

Solution: Let $q = 1 - p$.

- X is a non-negative random variable.
- The CDF of X is $\mathbb{P}(X \leq k) = 1 - q^{k+1}$, so the complementary CDF is $\mathbb{P}(X > k) = q^{k+1}$.

By Theorem 11.5,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=0}^{\infty} q^{k+1} = q(1 + q + q^2 + q^3 + \ldots) = \frac{q}{1-q} = \frac{1-p}{p}.$$

11.3 Continuous distributions

If X is continuous, the Riemann-Stieltjes integral of Theorem 11.1 reduces to an ordinary Riemann integral:

Theorem 11.7
Let X be a non-negative continuous random variable, and let $f(x)$ denote its PDF. Then

$$\mathbb{E}(X) = \int_0^\infty x f(x)\,dx,$$

provided the integral is absolutely convergent.

[*Proof omitted.*]

Remark 11.8
For signed continuous random variables, $\mathbb{E}(X) = \displaystyle\int_{-\infty}^{\infty} x f(x)\,dx$ provided the integral is absolutely convergent.

Theorem 11.9
Let X be a non-negative continuous random variable, and let F denote its CDF. Then

$$\mathbb{E}(X) = \int_0^\infty 1 - F(x)\,dx$$

[*Proof omitted.*]

Example 11.10 (Rayleigh distribution)
Let X be a continuous random variable having the *Rayleigh* distribution with parameter $\sigma > 0$. This has the following CDF:

$$F(x) = \begin{cases} 1 - e^{-x^2/2\sigma^2} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\mathbb{E}(X) = \sigma\sqrt{\dfrac{\pi}{2}}$.

Solution: Because X is non-negative, by Theorem 11.9 we have

$$\mathbb{E}(X) = \int_0^\infty 1 - F(x)\,dx = \int_0^\infty e^{-x^2/2\sigma^2}\,dx.$$

Changing the variable of integration by setting $t = x/(\sigma\sqrt{2})$, this becomes

$$\mathbb{E}(X) = \sigma\sqrt{2}\int_0^\infty e^{-t^2}\,dt.$$

We recognise the integral as one-half of the famous *Gaussian integral*:

$$\int_{-\infty}^\infty e^{-t^2}\,dt = \sqrt{\pi}.$$

Thus, because e^{-t^2} is an even function,

$$\mathbb{E}(X) = \sigma\sqrt{2}\int_0^\infty e^{-t^2}\,dt. = \frac{\sigma\sqrt{2}}{2}\int_{-\infty}^\infty e^{-t^2}\,dt = \sigma\sqrt{\frac{\pi}{2}},$$

as required.

11.4 Transformed variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X : \Omega \to \mathbb{R}$ be a random variable, let $F : \mathbb{R} \to [0,1]$ be its CDF, and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function.

- By Theorem 7.2, the transformed variable $g(X)$ is a random variable on $(\Omega, \mathcal{F})$.

(1) If $g(X)$ is a non-negative random variable,

$$\mathbb{E}\big[g(X)\big] = \int_0^\infty g(x)\,dF(x)$$

(2) If $g(X)$ is an integrable random variable,

$$\mathbb{E}\big[g(X)\big] = \int_0^\infty g^+(x)\,dF(x) - \int_0^\infty g^-(x)\,dF(x).$$

The latter expression reduces to

$$\mathbb{E}\big[g(X)\big] = \begin{cases} \displaystyle\sum_{i=1}^\infty g^+(x_i)f(x_i) - \sum_{i=1}^\infty g^-(x_i)f(x_i) & \text{when } X \text{ is discrete, and} \\ \displaystyle\int_0^\infty g^+(x)f(x)\,dx - \int_0^\infty g^-(x)f(x)\,dx & \text{when } X \text{ is continuous.} \end{cases}$$

Example 11.11
Let $X \sim \text{Uniform}[-1,1]$ be a continuous random variable. Find $\mathbb{E}(1/X^2)$ and $\mathbb{E}(1/X)$.

Solution: Let $f(x)$ denote the PDF of X:

$$f(x) = \begin{cases} 1/2 & \text{for } \ -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(1) Let $g(x) = 1/x^2$. This is a non-negative function, so

$$\mathbb{E}\left(\frac{1}{X^2}\right) = \int_0^\infty g(x)f(x)\,dx = \frac{1}{2}\int_{-1}^1 \frac{1}{x^2}\,dx = \int_0^1 \frac{1}{x^2}\,dx = \infty.$$

Thus $\mathbb{E}(1/X^2)$ is infinite.

(2) Let $g(x) = 1/x$. This is a signed function, so we must consider its positive and negative parts:

$$g^+(x) = \begin{cases} 1/x & \text{if } \ x \geq 0, \\ 0 & \text{if } \ x < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if } \ x \geq 0, \\ -1/x & \text{if } \ x < 0. \end{cases}$$

The expectation of $g(X) = 1/X$ is therefore given by

$$\begin{aligned} \mathbb{E}\left(\frac{1}{X}\right) &= \int_0^\infty g^+(x)f(x)\,dx - \int_0^\infty g^-(x)f(x)\,dx \\ &= \frac{1}{2}\int_0^1 \frac{1}{x}\,dx - \frac{1}{2}\int_{-1}^0 \frac{-1}{x}\,dx \\ &= \frac{1}{2}\int_0^1 \frac{1}{x}\,dx - \frac{1}{2}\int_0^1 \frac{1}{x}\,dx \\ &= \infty - \infty. \end{aligned}$$

Thus $\mathbb{E}(1/X)$ is undefined.

11.5 Exercises

Exercise 11.1

1. Let X be the score on a fair die, and let $g(x) = 3x - x^2$. Find the expected value and variance of the random variable $Y = g(X)$.

2. A long line of athletes $k = 0, 1, 2, \ldots$ make throws of a javelin to distances $X_0, X_1, X_2, \ldots$ respectively. The distances are independent and identically distributed random variables, and the probability that any two throws are exactly the same distance is equal to zero. Let Y be the index of the first athlete in the sequence who throws further than distance X_0. Show that the expected value of Y is infinite.

3. Consider the following game. A random number X is chosen uniformly from $[0, 1]$, then a sequence $Y_1, Y_2, \ldots$ of random numbers are chosen independently and uniformly from $[0, 1]$. Let Y_n be the first number in the sequence for which $Y_n > X$. When this occurs, the game ends and the player is paid $(n - 1)$ pounds. Show that the expected win is infinite.

4. Let X be a discrete random variable with PMF
$$f(k) = \begin{cases} \dfrac{3}{\pi^2 k^2} & \text{if } \ k \in \{\pm 1, \pm 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$
Show that $\mathbb{E}(X)$ is undefined.

5. Let X be a continuous random variable having the Cauchy distribution, defined by the PDF
$$f(x) = \frac{1}{\pi(1 + x^2)} \qquad x \in \mathbb{R}$$
Show that $\mathbb{E}(X)$ is undefined.

6. A coin is tossed until the first time a head is observed. If this occurs on the nth toss and n is odd, you win $2^n/n$ pounds, but if n is even then you lose $2^n/n$ pounds. Show that the expected win is undefined.

7. Let X be a continuous random variable with uniform density on the interval $[-1, 1]$,
$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1, +1] \\ 0 & \text{otherwise.} \end{cases}$$
Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

8. Let X be a random variable with the following CDF:
$$F(x) = \begin{cases} 0 & \text{for } \quad x \leq 1 \\ 1 - 1/x^2 & \text{for } \quad x \geq 1 \end{cases}$$
Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

9. Let X be a continuous random variable with the following PDF:
$$f(x) = \begin{cases} 1 - |x| & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases}$$
Find the range of integer values $\alpha \in \mathbb{Z}$ for which $\mathbb{E}(X^\alpha)$ exists.

Lecture 12 Concentration Inequalities

To be read in preparation for the **11.00** lecture on **Wed 05 Nov** in **Physiology A**.

12.1 Markov's inequality

If the distribution of a random variable is not known, probabilities can be estimated using the moments of the distribution. A simple upper bound on the tail probability of a non-negative random variable is provided by *Markov's inequality.*

Theorem 12.1 (Markov's inequality)
Let $X \geq 0$ be any non-negative random varible with finite mean. Then for every $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof: Let I_A be the indicator function of event $A = \{\omega : X(\omega) \geq a\}$.

$$I_A(\omega) = \begin{cases} 0 & \text{if } X(\omega) < a, \\ 1 & \text{if } X(\omega) \geq a. \end{cases}$$

- If $\omega \in A$, then $X(\omega) \geq a = aI_A(\omega)$.
- If $\omega \notin A$, then $X(\omega) \geq 0 = aI_A(\omega)$.

In either case, we have $X(\omega) \geq aI_A(\omega)$, so by the monotonicity of expectation (Theorem 10.12),

$$\mathbb{E}(X) \geq a\mathbb{E}(I_A) = a\mathbb{P}(A) \equiv a\mathbb{P}(X \geq a).$$

Hence $\mathbb{P}(X \geq a) \leq \mathbb{E}(X)/a$, as required.

Example 12.2
A fair die is rolled once. Use Markov's inequality to find an upper bound on the probability that we observe a score of at least 5.

Solution: Let X be the number shown on the die. Then X is a non-negative random variable with expectation $\mathbb{E}(X) = 7/2$.

Markov's inequality yields the upper bound

$$\mathbb{P}(X \geq 5) \leq \frac{\mathbb{E}(X)}{5} = \frac{7}{10}.$$

In this example we know that $\mathbb{P}(X \geq 5) = 1/3$, which illustrates that Markov's inequality yields only crude bounds on tail probabilites. Indeed, for the probability $\mathbb{P}(X \geq 3)$ Markov's inequality yields

$$\mathbb{P}(X \geq 3) \leq \frac{\mathbb{E}(X)}{3} = \frac{7}{6}.$$

This tells us nothing useful (because we know that $\mathbb{P}(X \geq 3) \leq 1$).

Theorem 12.3 (Markov's inequality (General form))
Let X be any random varible with finite mean, and let $g : \mathbb{R} \to [0, \infty)$ be a non-negative function. Then for every $a > 0$,

$$\mathbb{P}\big[g(X) \geq a\big] \leq \frac{\mathbb{E}\big[g(X)\big]}{a}.$$

Proof: Let I_A be the indicator function of the event $A = \big\{\omega : g\big[X(\omega)\big] \geq a\big\}$.

$$I_A(\omega) = \begin{cases} 0 & \text{if } g\big[X(\omega)\big] < a, \\ 1 & \text{if } g\big[X(\omega)\big] \geq a. \end{cases}$$

Then $g(X) \geq aI_A$, so by the monotonicity of expectation,

$$\mathbb{E}\big[g(X)\big] \geq a\mathbb{E}(I_A) = a\mathbb{P}(A) = a\mathbb{P}(X \geq a).$$

which concludes the proof.

12.2 Chebyshev's inequality

An upper bound on the absolute deviation of a random variable from its mean is provided by *Chebyshev's inequality.*

Corollary 12.4 (Chebyshev's inequality)
Let X be any random varible with finite mean. Then for all $\epsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Proof: Take $g(X) = \big(X - \mathbb{E}(X)\big)^2$ and $a = \epsilon^2$ in Markov's inequality,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) = \mathbb{P}\Big[\big(X - \mathbb{E}(X)\big)^2 \geq \epsilon^2\Big] \leq \frac{\mathbb{E}\Big[\big(X - \mathbb{E}(X)\big)^2\Big]}{\epsilon^2} = \frac{\mathrm{Var}(X)}{\epsilon^2},$$

as required.

Example 12.5
Suppose that $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = 1$. Find an integer value k such that $\mathbb{P}(|X| \geq k) \leq 0.01$.

Solution: X is not non-negative, so we can not use Markov's inequality here. By Chebyshev's inequality,

$$\mathbb{P}(|X| \geq k) = \mathbb{P}(|X - \mathbb{E}X| \geq k) \leq \frac{\mathrm{Var}(X)}{k^2} = \frac{1}{k^2} \leq \frac{1}{100}$$

so we need $k \geq 10$.

Example 12.6
Let X be a continuous random variable with expected value 3.6 and standard deviation 1.2. Find a lower bound for the probability $\mathbb{P}(1.2 \leq X \leq 6.0)$.

Solution:
The event $\{1.2 \leq X \leq 6.0\}$ can be written as $\{|X - 3.6| \leq 2.4\}$, and by Chebyshev's inequality,

$$\mathbb{P}(|X - 3.6| > 2.4) \leq \frac{\mathrm{Var}(X)}{2.4^2} = \frac{1.2^2}{2.4^2} = \frac{1}{4}.$$

Thus $\mathbb{P}(1.2 \leq X \leq 6.0) \geq 3/4$.

12.3 Bernstein's inequality

Theorem 12.7 (Bernstein's inequality)
Let X be a random variable. Then for all $t > 0$,

$$\mathbb{P}(X > a) \leq e^{-ta}\mathbb{E}(e^{tX}).$$

Proof: For any non-negative random variable Y, Markov's inequality says that

$$\mathbb{P}(Y > y) \leq \frac{\mathbb{E}(Y)}{y}.$$

Let us take $Y = e^{tX}$ and $y = e^{ta}$. Then Y is non-negative, so

$$\mathbb{P}(e^{tX} > e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}} = e^{-ta}\mathbb{E}(e^{tx}).$$

The result then follows by the fact that the exponential function is monotonic increasing, which means that $\mathbb{P}(X > a) = \mathbb{P}(e^{tX} > e^{ta})$.

12.4 Exercises

Exercise 12.1

1. Let $X \sim \text{Uniform}[0,20]$ be a continuous random variable.
   (1) Use Chebyshev's inequality to find an upper bound on the probability $\mathbb{P}(|X-10| \geq z)$.
   (2) Find the range of z for which Chebyshev's inequality gives a non-trivial bound.
   (3) Find the value of z for which $\mathbb{P}(|X-10| \geq z) \leq 3/4$.

2. Let X be a discrete random variable, taking values in the range $\{1,2,\ldots,n\}$, and suppose that $\mathbb{E}(X) = \text{Var}(X) = 1$. Show that $\mathbb{P}(X \geq k+1) \leq k^2$ for any integer k.

3. Let $k \in \mathbb{N}$. Show that Markov's inequality is tight (i.e. cannot be improved) by finding a non-negative random variable X such that
$$\mathbb{P}\big[X \geq k\mathbb{E}(X)\big] = \frac{1}{k}.$$

4. What does the Chebyshev inequality tell us about the probability that the value taken by a random variable deviates from its expected value by six or more standard deviations?

5. Let S_n be the number of successes in n Bernoulli trials with probability p of success on each trial. Use Chebyshev's Inequality to show that, for any $\epsilon > 0$, the upper bound
$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}$$
   is valid for any p.

6. Let $X \sim N(0,1)$.
   (1) Use Chebyshev's Inequality to find upper bounds for the probabilities $\mathbb{P}(|X| \geq 1)$, $\mathbb{P}(|X| \geq 2)$ and $\mathbb{P}(|X| \geq 3)$.
   (2) Use statistical tables to find the area under the standard normal curve over the intervals $[-1,1]$, $[-2,2]$ and $[-3,3]$.
   (3) Compare the bounds computed in part (a) with the exact values found in part (b). How good is the Chebyshev inequality in this case?

7. Let X be a random variable with mean $\mu \neq 0$ and variance σ^2, and define the *relative deviation* of X from its mean by $D = \left|\dfrac{X-\mu}{\mu}\right|$. Show that
$$\mathbb{P}(D \geq a) \leq \left(\frac{\sigma}{\mu a}\right)^2.$$

Lecture 13 Probability Generating Functions

To be read in preparation for the **11.00** lecture on **Mon 10 Nov** in **E/0.15**.

13.1 Generating functions

Generating functions, first introduced by de Moivre in 1730, are power series used to represent sequences of real numbers. It is often easier to work with generating functions than with the original sequences.

Definition 13.1
Let $a = (a_0, a_1, a_2, \ldots)$ be a sequence of real numbers. The *generating function* of the sequence is the function $G_a(t)$, defined for every $t \in \mathbb{R}$ for which the sum converges, by

$$G_a(t) = \sum_{k=0}^{\infty} a_k t^k.$$

The sequence can be reconstructed from $G_a(t)$ by setting

$$a_n = \frac{1}{n!} G_a^{(n)}(0),$$

where $G_a^{(n)}(t)$ is nth derivative of $G_a(t)$. In particular,

$$G(0) = a_0, \quad G'(0) = a_1, \quad G''(0) = a_2, \quad \text{and so on.}$$

Example 13.2
The *convolution* of two sequences $a = (a_0, a_1, a_2, \ldots)$ and $b = (b_0, b_1, b_2, \ldots)$ is another sequence $c = (c_0, c_1, c_2, \ldots)$, whose kth term is

$$c_k = a_0 b_k + a_1 b_{k-1} + \ldots + a_k b_0 = \sum_{i=0}^{k} a_i b_{k-i}.$$

Convolutions can be difficult to handle. However, the generating function of a convolution is just the product of the generating functions of the original sequences:

$$\begin{aligned} G_c(t) = \sum_{k=0}^{\infty} c_k t^k &= \sum_{k=0}^{\infty} \left[\sum_{i=0}^{n} a_i b_{k-i} \right] t^k \\ &= \sum_{i=0}^{\infty} a_i t^i \sum_{k=i}^{\infty} b_{k-i} t^{k-i} = \sum_{i=0}^{\infty} a_i t^i \sum_{j=1}^{\infty} b_j t^j = G_a(t) G_b(t). \end{aligned}$$

A convolution of sequences is replaced by a product of generating functions.

13.1.1 Properties of generating functions*

A generating function $G_a(t)$ is a power series whose coefficients are the terms of the sequence a. All power series have the following properties:

Convergence. There exists a *radius of convergence* $R \geq 0$ such that $G_a(t)$ is absolutely convergent when $|t| < R$, and divergent when $|t| > R$.

Differentiation. $G_a(t)$ may be differentiated or integrated any number of times whenever $|t| < R$.

Uniqueness. If $G_a(t) = G_b(t)$ for all $|t| < R'$, where $0 < R' \leq R$, then $a_n = b_n$ $\forall n$.

Abel's theorem. If $a_k > 0$ for all k, and $G_a(t)$ converges for all $|t| < 1$, then

$$G_a(1) = \lim_{t \uparrow 1} G_a(t) = \lim_{t \uparrow 1} \sum_{k=0}^{\infty} a_k t_k = \sum_{k=0}^{\infty} a_k.$$

13.2 Probability generating functions

Definition 13.3
Let X be a discrete random variable taking values in the range $\{0, 1, 2, \ldots\}$, and let f denote its PMF. The *probability generating function* (PGF) of X is the generating function of its PMF:

$$G(t) = \mathbb{E}(t^X) = \sum_{k=0}^{\infty} f(k) t^k$$

Remark 13.4
- $G(t)$ converges for all $|t| \leq 1$.
- $G(0) = 0$.
- $G(1) = \sum_{k=0}^{\infty} f(k) = 1$.

Example 13.5
The PGFs of some notable discrete distributions on $\{0, 1, 2, \ldots\}$ are computed as follows:

(1) **Constant**: if $\mathbb{P}(X = c) = 1$,

$$G(t) = \sum_{k=0}^{\infty} f(k) t^k = t^c.$$

(2) **Bernoulli**: if $X \sim \text{Bernoulli}(p)$, its PMF is

$$f(k) = \begin{cases} 1-p & \text{if } k = 0, \\ p & \text{if } k = 1, \end{cases}$$

and zero otherwise, so its PGF is

$$G(t) = \sum_{k=0}^{\infty} f(k) t^k = (1-p)t^0 + pt^1 = 1 - p + pt.$$

(3) **Poisson**: if X$\sim$ Poisson(λ), its PMF is

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, 2, \ldots,$$

and zero otherwise, so its PGF is

$$G(t) = \sum_{k=0}^{\infty} f(k) t^k = \sum_{k=0}^{\infty} \left(\frac{\lambda^k e^{-\lambda}}{k!} \right) t^k = e^{-\lambda} \sum_{i=1}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

(4) **Geometric**: if $X \sim \text{Geometric}(p)$, its PMF is

$$f(k) = (1-p)^k p \text{ for } k = 0, 1, 2, \ldots,$$

and zero otherwise, so its PGF is

$$G(t) = \sum_{k=0}^{\infty} f(k)t^k = \sum_{k=0}^{\infty}(1-p)^k p t^k = p\sum_{k=0}^{\infty}\big[(1-p)t\big]^k = \frac{p}{1-(1-p)t} \qquad \text{for all } |t| < \frac{1}{1-p}.$$

Here, we have used the fact that $\displaystyle\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ for $|r| < 1$.

Theorem 13.6
Let X be a random variable, and let $G(t)$ denote its PGF. Then $\mathbb{E}(X) = G'(1)$, and more generally,

$$\mathbb{E}\big[X(X-1)\ldots(X-n+1)\big] = G^{(n)}(1),$$

where $G^{(n)}(1)$ is the nth derivative of $G(t)$ evaluated at $t = 1$.

Proof: Take $t < 1$, and compute the nth derivative of G to obtain

$$\begin{aligned} G^{(n)}(t) &= \sum_{k=0}^{\infty} t^{k-n} k(k-1)\cdots(k-n+1)f(k) \\ &= \mathbb{E}\big[t^{X-n} X(X-1)\cdots(X-n+1)\big] \end{aligned}$$

- If $R > 1$ it follows immediately that $G^{(n)}(1) = \mathbb{E}\big[X(X-1)\cdots(X-n+1)\big]$.
- If $R = 1$, Abel's theorem yields the same conclusion.

Remark 13.7
$\mathbb{E}\big[X(X-1)\ldots(X-n+1)\big]$ is called the nth *factorial moment* of X.

Example 13.8
The variance of X can be written in terms of $G(t)$ as follows:

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

13.2.1 Sums of random variables

Let X and Y be two independent discrete random variables, both taking values in $\{0, 1, 2, \ldots\}$. The PMF of their sum $X + Y$ is given by the convolution of the individual PMFs,

$$\mathbb{P}(X + Y = k) = \sum_{j=0}^{\infty} \mathbb{P}(X = j)\mathbb{P}(Y = k - j).$$

The corresponding PGFs satisfy a more straightforward, multiplicative relationship:

Theorem 13.9
If X and Y are independent, then $G_{X+Y}(t) = G_X(t)G_Y(t)$.

Proof: If X and Y are independent, then t^X and t^Y are also independent, so

$$G_{X+Y}(t) = \mathbb{E}(t^{X+Y}) = \mathbb{E}(t^X t^Y) = \mathbb{E}(t^X)\mathbb{E}(t^Y) = G_X(t)G_Y(t).$$

Corollary 13.10
If $S = X_1 + X_2 + \ldots + X_n$ is a sum of independent random variables taking values in the non-negative integers, its PGF is

$$G_S(t) = G_{X_1}(t)G_{X_2}(t)\cdots G_{X_n}(t)$$

Example 13.11
Show that the PGF of the Binomial(n, p) distribution is $G(t) = (1 - p + pt)^n$.

Solution: Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli variables with parameter p, and let

$$S = X_1 + X_2 + \ldots + X_n.$$

Each X_i has generating function $G(t) = 1 - p + pt$, so by Theorem 13.9,

$$G_S(t) = G_{X_1+X_2+\ldots+X_n}(t) = \big[G(t)\big]^n = (1 - p + pt)^n.$$

Example 13.12
Let $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ) be independent. Show that $X + Y \sim$ Poisson$(\lambda + \mu)$.

Solution: The PGF of X is

$$G_X(t) = \sum_{k=0}^{\infty} f_X(k)t^k = \sum_{k=0}^{\infty} \left(\frac{\lambda^k e^{-\lambda}}{k!}\right) t^k = e^{\lambda(t-1)}.$$

Similarly we have $G_Y(t) = e^{\mu(t-1)}$, so the PGF of $Z = X + Y$ is

$$G_Z(t) = G_X(t)G_Y(t) = e^{(\lambda+\mu)(t-1)}.$$

We recognise this as the PGF of a Poisson$(\lambda + \mu)$ random variable, so $Z \sim$ Poisson$(\lambda + \mu)$.

13.3 Exercises

Exercise 13.1

1. Let $X \sim$ Binomial(m, p) and $Y \sim$ Binomial(n, p). Show that $X + Y \sim$ Binomial$(m + n, p)$,
2. Show that a discrete distribution on the non-negative integers is uniquely determined by its PGF, in the sense that if two such random variables X and Y have PGFs $G_X(t)$ and $G_Y(t)$ respectively, then $G_X(t) = G_Y(t)$ if and only if $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ for all $k = 0, 1, 2, \ldots$.
3. The PGF of a random variable is given by $G(t) = 1/(2 - t)$. What is its PMF?
4. Let $X \sim$ Binomial(n, p). Using the PGF of X, show that
$$\mathbb{E}\left(\frac{1}{1 + X}\right) = \frac{1 - (1 - p)^{n+1}}{(n + 1)p}.$$

Lecture 14 Moment Generating Functions

To be read in preparation for the **11.00** lecture on **Wed 12 Nov** in **Physiology A**.

14.1 Moment generating functions

- PGFs are defined only for discrete random variables taking non-negative integer values.
- MGFs are defined for any random variable.

Definition 14.1
The *moment generating function* (MGF) of a random variable X is a function $M : \mathbb{R} \to [0, \infty]$ given by

$$M(t) = \mathbb{E}(e^{tX}).$$

Remark 14.2
(1) e^{tX} is non-negative, so its expectation is well-defined, and $\mathbb{E}(e^{tX}) \geq 0$.

(2) For a discrete random variable X taking non-negative integer values,

$$M(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\big[(e^t)^X\big] = G(e^t),$$

where G is the PGF of X.

(3) MGFs are related to *Laplace transforms*.

Example 14.3
The MGFs of some notable discrete distributions can be computed as follows:

$$\begin{aligned}
&X \sim \text{Bernoulli}(p): && G(t) = 1 - p + pt && \Rightarrow\ M(t) = 1 - p + pe^t \\
&X \sim \text{Binomial}(n, p): && G(t) = (1 - p + pt)^n && \Rightarrow\ M(t) = (1 - p + pe^t)^n \\
&X \sim \text{Poisson}(\lambda): && G(t) = e^{\lambda(t-1)} && \Rightarrow\ M(t) = e^{\lambda(e^t - 1)}
\end{aligned}$$

MGFs have properties similar to those of PGFs:

Theorem 14.4
(1) If X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

(2) If $Y = a + bX$, then $M_Y(t) = e^{at}M_X(bt)$

Proof:

(1) By independence,

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) = M_X(t)M_Y(t).$$

(2) For $Y = a + bX$,

$$M_Y(t) = \mathbb{E}\big(e^{t(a+bX)}\big) = e^{at}\mathbb{E}\big(e^{btX}\big) = e^{at}M_X(bt).$$

Theorem 14.5
Let $M(t)$ be the MGF of the random variable X. If $M(t)$ converges on an open interval $(-R, R)$ centred at the origin, then

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

Proof: Using the series expansion of e^{tX} and the linearity of expectation,

$$M(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{k!} t^k$$

Corollary 14.6
Let X be a random variable, and let $M(t)$ denote its MGF. Then

$$\mathbb{E}(X^n) = M^{(n)}(0),$$

where $M^{(n)}(0)$ is the nth derivative of $M(t)$ evaluated at $t = 0$. In particular,

$$M(0) = 1, \quad M'(0) = \mathbb{E}(X), \quad M''(0) = \mathbb{E}(X^2), \quad \text{and so on.}$$

Example 14.7 (Exponential distribution)
Let $X \sim \text{Exponential}(\lambda)$ where $\lambda > 0$ is a rate parameter.

(1) Show that the MGF of X is given by $M(t) = \dfrac{\lambda}{\lambda - t}$.

(2) Use $M(t)$ to find the mean and variance of X.

Solution:

(1) The PDF of X is $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ (and zero otherwise), so

$$\begin{aligned} M(t) = \int_0^\infty e^{tx} f(x)\,dx = \lambda \int_0^\infty e^{tx}e^{-\lambda x}\,dx &= \lambda \int_0^\infty e^{-(\lambda - t)x}\,dx \\ &= \frac{-\lambda}{\lambda - t}\Big[e^{-(\lambda-t)x}\Big]_0^\infty = \frac{\lambda}{\lambda - t}. \end{aligned}$$

(2)
- $M'(t) = \dfrac{\lambda}{(\lambda - t)^2}$, which yields $\mu_1 = M'(0) = \dfrac{1}{\lambda}$.
- $M''(t) = \dfrac{2\lambda}{(\lambda - t)^3}$ which yields $\mu_2 = M''(0) = \dfrac{2}{\lambda^2}$.
- Thus $\sigma^2 = \mu_2 - \mu_1^2 = \dfrac{1}{\lambda^2}$.

Example 14.8 (Gamma distribution)
The PDF of the Gamma(k,θ) distribution is given by

$$f(x) = \begin{cases} \dfrac{x^{k-1}e^{-x}}{\Gamma(k)} & x > 0, \\ 0 & \text{othewise.} \end{cases}$$

Show that the MGF of the Gamma(k,θ) distribution is given by

$$M(t) = \frac{1}{(1-\theta t)^k}$$

Solution: Let $X \sim \text{Gamma}(k,1)$. Then

$$\begin{aligned} M(t) = \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} f(x)\,dx &= \frac{1}{\Gamma(k)}\int_0^\infty x^{k-1}e^{-(1-t)x}\,dx \\ &= \frac{1}{\Gamma(k)}\frac{1}{(1-t)^k}\int_0^\infty \big[(1-t)x\big]^{k-1}e^{-(1-t)x}(1-t)\,dx \\ &= \frac{1}{\Gamma(k)}\frac{1}{(1-t)^k}\int_0^\infty y^{k-1}e^{-y}\,dy \\ &= \frac{1}{(1-t)^k}. \end{aligned}$$

The scaled variable $Y = \theta X$ has distribution $Y \sim \text{Gamma}(k,\theta)$, so by the properties of MGFs,

$$M_Y(t) = M_X(\theta t) = \frac{1}{(1-\theta t)^k}.$$

Example 14.9 (Normal distribution)
By first considering the MGF of the $N(0,1)$ distribution, show that the MGF of the $N(\mu,\sigma^2)$ distribution is given by

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Solution: Let $Z \sim N(0,1)$ be a standard normal variable; this has PDF

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \qquad (z \in \mathbb{R}).$$

Thus

$$M(t) = \mathbb{E}(e^{tZ}) = \int_{-\infty}^\infty e^{tz} f(z)\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{tz-\frac{1}{2}z^2}\,dz$$

The exponent $tz - \frac{1}{2}z^2$ can be written as $\frac{1}{2}t^2 - \frac{1}{2}(z-t)^2$, so

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{1}{2}t^2}e^{-\frac{1}{2}(z-t)^2}\,dz \\ &= e^{\frac{1}{2}t^2}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}(z-t)^2}\,dz\right) \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$

Let $X = \mu + \sigma Z$. Then $X \sim N(\mu,\sigma^2)$ and by the properties of MGFs,

$$M_X(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t}e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2},$$

as required.

14.2 Characteristic functions

MGFs are useful, but the expectations that define them may not always be finite. Characteristic functions do not suffer this disadvantage.

Definition 14.10
The *characteristic function* of a random variable X is a function $\phi : \mathbb{R} \to \mathbb{C}$ given by

$$\phi(t) = \mathbb{E}(e^{itX}) \qquad \text{where} \qquad i = \sqrt{-1}.$$

Remark 14.11
- If $M(t)$ is the MGF of X, its characteristic function is given by $\phi(t) = M(it)$.
- $\phi : \mathbb{R} \to \mathbb{C}$ exists for all $t \in \mathbb{R}$.
- Characteristic functions are related to *Fourier transforms*.

Theorem 14.12
(1) If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

(2) If $Y = a + bX$, then $\phi_Y(t) = e^{-iat}\phi_X(bt)$

[*Proof omitted.*]

14.2.1 The inversion theorem

The *inversion theorem* asserts that a random variable is entirely specified by its characteristic function, meaning that X and Y have the same characteristic function if and only if they have the same distribution. We state the inversion theorem only for continuous distributions:

Theorem 14.13 (Fourier inversion theorem)
If X is continuous with density function f and characteristic function ϕ, then

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt$$

at every point x at which f is differentiable.

[*Proof omitted.*]

14.2.2 The continuity theorem

For a sequence of random variables $X_1, X_2, \ldots$, the *continuity theorem* asserts that if the cumulative distribution functions $F_1, F_2, \ldots$ of the sequence approaches some limiting distribution F, then the charactaristic functions $\phi_1, \phi_2, \ldots$ of the sequence approaches the characteristic function of F.

Definition 14.14
A sequence of distribution functions $F_1, F_2, \ldots$ is said to *converge* to the distribution function F, denoted by $F_n \to F$, if $F_n(x) \to F(x)$ as $n \to \infty$ at each point x at which F is continuous.

Theorem 14.15 (Continuity theorem)
Let $F_1, F_2, \ldots$ and F be distribution functions, and let $\phi_1, \phi_2, \ldots$ and ϕ denote the corresponding characteristic functions.

(1) If $F_n \to F$ then $\phi_n(t) \to \phi(t)$ for all t.

(2) If $\phi_n(t) \to \phi(t)$ then $F_n \to F$ provided $\phi(t)$ exists and is continuous at $t = 0$.

[*Proof omitted.*]

14.3 Exercises

Exercise 14.1

1. Let X be a discrete random variable, taking values in the set $\{-3,-2,-1,0,1,2,3\}$ with uniform probability, and let $M(t)$ denote the MGF of X.
   (1) Show that $M(t) = \frac{1}{7}(e^{-3t} + e^{-2t} + e^{-t} + 1 + e^{t} + e^{2t} + e^{3t})$.
   (2) Use $M(t)$ to compute the mean and variance of X.

2. A continuous random variable X has MGF given by $M(t) = \exp(t^2 + 3t)$. Find the distribution of X.

3. Let X be a discrete random variable with probability mass function

$$\mathbb{P}(X = k) = \begin{cases} q^k p & k = 0,1,2,\dots \\ 0 & \text{otherwise,} \end{cases}$$

   where $0 < p < 1$ and $q = 1 - p$.
   (1) Show that the MGF of X is given by $M(t) = \dfrac{p}{1 - qe^t}$ for $t < -\log q$.
   (2) Find the PGF of X.
   (3) Use the PGF of X to find the PMF of $Y = X + 1$.
   (4) Use $M(t)$ to find the mean and variance of X.

4. Let $M(t)$ denote the MGF of the normal distribution $\mathrm{N}(0,\sigma^2)$. By exanding $M(t)$ as a power series in t, show that the moments μ_k of the $\mathrm{N}(0,\sigma^2)$ distribution are zero if k is odd, and equal to

$$\mu_{2m} = \frac{\sigma^{2m}(2m)!}{2^m m!} \qquad \text{if } k = 2m \text{ is even.}$$

5. Let $X \sim \text{Exponential}(\theta)$ where θ is a scale parameter.
   (1) Show that the MGF of X is $M(t) = \dfrac{1}{1 - \theta t}$.
   (2) By expanding this expression as a power series in t, find the first four non-central moments of X.
   (3) Find the skewness γ_1 and the excess kurtosis γ_2 of X.

6. Let $X_1, X_2, \dots$ be independent and identically distributed random variables, with each $X_i \sim \mathrm{N}(\mu, \sigma^2)$.
   (1) Find the MGF of the random variable $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
   (2) Show that $\bar{X}$ has a normal distribution, and find its mean and variance.

7. Let $X_1 \sim \text{Gamma}(k_1, \theta)$ and $X_2 \sim \text{Gamma}(k_2, \theta)$ be independent random variables. Use the MGFs of X_1 and X_2 to find the distribution of the random variable $Y = X_1 + X_2$.

8. A coin has probability p of showing heads. The coin is tossed repeatedly until exactly k heads occur. Let N be the number of times the coin is tossed. Using the continuity theoram for characteristic functions, show that the distribution of the random variable $X = 2pN$ converges to a gamma distribution as $p \to 0$.

9. Let X and Y be independent and identically distributed random variables, with means equal to 0 and variances equal to 1. Let $\phi(t)$ denote their common characteristic function, and suppose that the random variables $X + Y$ and $X - Y$ are independent. Show that $\phi(2t) = \phi(t)^3\phi(-t)$, and hence deduce that X and Y must be independent standard normal variables.

Lecture 15 The Law of Large Numbers

To be read in preparation for the **11.00** lecture on **Mon 17 Nov** in **E/0.15**.

15.1 Convergence

We define the following notions of convergence for sequences of random variables.

Definition 15.1
Let $X_1, X_2, \dots$ and X be random variables. We say that

(1) $X_n \to X$ *almost surely* if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$,

(2) $X_n \to X$ *in mean square* if $\mathbb{E}(|X_n - X|^2) \to 0$ as $n \to \infty$,

(3) $X_n \to X$ *in mean* if $\mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$,

(4) $X_n \to X$ *in probability*, if for all $\epsilon > 0$, $\mathbb{P}(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$,

(5) $X_n \to X$ *in distribution* if $F_n(x) \to F(x)$ as $n \to \infty$ for every point x at which F is continuous.

Theorem 15.2
(1) Convergence almost surely implies convergence in probability.

(2) Convergence in mean square implies convergence in mean.

(3) Convergence in mean implies convergence in probability.

(4) Convergence in probability implies convergence in distribution.

[*Proof omitted.*]

15.2 The law of large numbers

Theorem 15.3 (The weak law of large numbers)
Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables having finite mean μ, and finite variance. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad \text{in probability as } n \to \infty.$$

Proof:

- By the linearity of expectation, $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \mu$.
- Because the X_i are independent, $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}$.
- Applying Chebyshev's inequality, $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$ for all $\epsilon > 0$.

Because σ^2 is finite, it follows that $\frac{1}{n}\sum_{i=1}^{n}X_i \to \mu$ in probability as $n \to \infty$, as required.

Theorem 15.4 (The law of large numbers: convergence in mean square)
Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having finite mean μ, and finite variance. Then

$$\frac{1}{n}\sum_{i=1}^{n}X_i \to \mu \quad \text{in mean square as } n \to \infty.$$

Proof: Let μ and σ^2 respectively denote the mean and variance of X.

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right)^2\right] = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty.$$

Remark 15.5 (Frequentist probability)
A random experiment is repeated n times under the same conditions. Let A be some random event, and let X_i be the indicator variable of the event that A occurs on the ith trial. Then the sample mean of the X_i is the *relative frequency* of event A over these n repetitions, and by the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}X_i \to \mathbb{P}(A) \quad \text{as } n \to \infty.$$

This shows that the frequentist model, in which probability is defined to be the limit of relative frequency as the number of repetitions increases to infinity, is a reasonable one.

15.3 Bernoulli's law of large numbers*

In the proof of Theorem 15.3, we used Chebyshev's inequality to show that

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \qquad \forall\, \epsilon > 0.$$

We say that the *rate* at which $\bar{X}_n \to \mu$ is of order $O(1/n)$ as $n \to \infty$. In the proof of the following theorem, we use Bernstein's inequality to show that the sample mean of Bernoulli random variables satisfies

$$\mathbb{P}\left(|\bar{X}_n - \mu| > \epsilon\right) \leq e^{-\frac{1}{2}n\epsilon^2} \qquad \forall\, \epsilon > 0.$$

In this case, the rate at which $\bar{X}_n \to \mu$ as $n \to \infty$ is said to be *exponentially fast.*

Theorem 15.6 (Bernoulli's Law of Large Numbers)
Let $X_1, X_2, \ldots$ be independent, with each $X_i \sim \mathrm{Bernoulli}(p)$, and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$ be the sample mean of the first n variables in the sequence. Then for every $\epsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - p| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Proof: Let $\epsilon > 0$ and define $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}\left(\bar{X}_n - p > \epsilon\right) = \mathbb{P}\big[S_n > n(p+\epsilon)\big]$$

Applying Bernstein's inequality (Theorem 12.7) to the random variable S_n with $a = n(p+\epsilon)$,

$$\begin{aligned}
\mathbb{P}\big[S_n > n(p+\epsilon)\big] &\leq e^{-tn(p+\epsilon)}\mathbb{E}(e^{tS_n}) \\
&= e^{-tn(p+\epsilon)}\big[1-p+pe^t\big]^n \\
&= e^{-tn\epsilon}\big[e^{-tp}(1-p+pe^t)\big]^n \\
&= e^{-tn\epsilon}\big[(1-p)e^{-tp}+pe^{t(1-p)}\big]^n
\end{aligned}$$

Using the inequality $e^x \leq x + e^{x^2}$, which holds for all $x \in \mathbb{R}$,

$$\begin{aligned}
(1-p)e^{-tp}+pe^{t(1-p)} &\leq (1-p)\big[-tp+e^{t^2p^2}\big]+p\big[t(1-p)+e^{t^2(1-p)^2}\big] \\
&= (1-p)e^{t^2p^2}+pe^{t^2(1-p)^2} \\
&\leq (1-p)e^{t^2}+pe^{t^2} \\
&= e^{t^2}.
\end{aligned}$$

Hence, for all $t > 0$,

$$\mathbb{P}\left(\bar{X}_n - p > \epsilon\right) = \mathbb{P}\big[S_n > n(p+\epsilon)\big] \leq e^{-tn\epsilon}e^{t^2n} = e^{tn(t-\epsilon)}.$$

- This inequality is valid for all $t > 0$.
- We choose t so that the right-hand side is made as small as possible.
- Because e^x is an increasing function, this corresponds to the minimum value of the exponent.
- We differentiate the exponent $tn(t-\epsilon)$ with respect to t and set this equal to zero.

This yields the value $t = \frac{1}{2}\epsilon$, so

$$\mathbb{P}\left(\bar{X}_n - p > \epsilon\right) = \mathbb{P}\big[S_n > n(p+\epsilon)\big] \leq e^{-\frac{1}{4}n\epsilon^2}.$$

A similar argument shows that

$$\mathbb{P}\left(\bar{X}_n - p < -\epsilon\right) = \mathbb{P}\big[S_n < n(p-\epsilon)\big] \leq e^{-\frac{1}{4}n\epsilon^2},$$

Thus we have

$$\mathbb{P}\left(|\bar{X}_n - p| > \epsilon\right) \leq e^{-\frac{1}{2}n\epsilon^2} \to 0 \quad \text{as } n \to \infty,$$

as required.

15.4 Exercises

Exercise 15.1

1. Let c be a constant, and let $X_1, X_2, \ldots$ be a sequence of random variables with $\mathbb{E}(X_n) = c$ and $\text{Var}(X_n) = 1/\sqrt{n}$ for each n. Show that the sequence converges to c in probability as $n \to \infty$.
2. A fair coin is tossed n times. Does the law of large numbers ensure that the observed number of heads will not deviate from $n/2$ by more than 100 with probability of at least 0.99, provided that n is sufficiently large?

Lecture 16 The Central Limit Theorem

To be read in preparation for the **11.00** lecture on **Wed 19 Nov** in **Physiology A**.

We will need the following result from elementary analysis:

Lemma 16.1
For any constant $c \in \mathbb{R}$,

$$\left(1+\frac{c}{n}\right)^n \to e^c \quad \text{as} \quad n \to \infty.$$

Proof: By the binomial theorem,

$$\begin{aligned}
\left(1+\frac{c}{n}\right)^n &= \sum_{k=1}^{n} \binom{n}{k} \left(\frac{c}{n}\right)^k \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!k!} \left(\frac{c^k}{n^k}\right) \\
&= \sum_{k=1}^{n} \frac{c^k}{k!} \left(\frac{n(n-1)\dots(n-k+1)}{n^k}\right) \\
&= \sum_{k=1}^{n} \frac{c^k}{k!} \left[1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k+1}{n}\right)\right] \\
&\to \sum_{k=1}^{\infty} \frac{c^k}{k!} = e^c \quad \text{as} \quad n \to \infty.
\end{aligned}$$

We will also need the following analogue of Theorem 14.5, which is a consequence of Taylor's theorem for functions of a complex variable. Here, $o(t^k)$ denotes a quantity with the property that $o(t^k)/t^k \to 0$ in the limit as $t \to 0$, and represents an 'error' term that is asymptotically smaller than the other terms of the expression in the limit as $t \to 0$, and which can therefore be neglected when t is sufficiently small. (This is called *Landau notation*.)

Theorem 16.2
If $\mathbb{E}(|X^k|) < \infty$, then

$$\phi(t) = \sum_{j=0}^{k} \frac{\mathbb{E}(X^j)}{j!}(it)^j + o(t^k) \qquad \text{as} \quad t \to 0,$$

[*Proof omitted.*]

16.1 Poisson limit theorem

Theorem 16.3 (The Poisson limit theorem)
If $X_n \sim \text{Binomial}(n, \lambda/n)$ then the distribution of X_n converges to the Poisson(λ) distribution as $n \to \infty$.

Proof: By the continuity theorem for characteristic functions, it is enough to show that the characteristic function of X_n converges to the characteristic function of the Poisson(λ) distribution as $n \to \infty$.

- Binomial(n, p):
$$M(t) = (1 - p + pe^t)^n \quad \Rightarrow \quad \phi(t) = M(it) = (1 - p + pe^{it})^n.$$
- Poisson(λ):
$$M(t) = \exp\left[\lambda(e^t - 1)\right] \quad \Rightarrow \quad \phi(t) = M(it) = \exp\left[\lambda(e^{it} - 1)\right].$$

The characteristic function of $X_n \sim \text{Binomial}(n, \lambda/n)$ is
$$\phi_n(t) = \mathbb{E}(e^{itX_n}) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^{it}\right)^n = \left[1 + \frac{\lambda(e^{it} - 1)}{n}\right]^n$$
By Lemma 16.1,
$$\phi_n(t) \to \exp\left[\lambda(e^{it} - 1)\right] \quad \text{as} \quad n \to \infty.$$
This is the characteristic function of the Poisson(λ) distribution, and the result follows by the continuity theorem for characteristic functions.

16.2 Law of large numbers

Theorem 16.4
Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common mean $\mu < \infty$. Then
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad \text{in distribution as} \quad n \to \infty.$$

- Unlike Theorem 15.3, this result does not require that the X_i have bounded variance.
- Convergence in distribution is however a weaker property than convergence in probability.

Proof: By the continuity theorem for characteristic functions, it is sufficient to show that the characteristic function of $\bar{X}_n$ converges to the characteristic function of the constant μ as $n \to \infty$.

- Let ϕ_X denote the common characteristic function of the X_i.
- Let ϕ_n denote the characteristic function of $\bar{X}_n$.

By the properties of characteristic functions,
$$\begin{aligned}\phi_n(t) &= \phi_{\frac{1}{n}(X_1+X_2+\ldots+X_n)}(t) \\ &= \phi_{(X_1+X_2+\ldots+X_n)}\left(\frac{t}{n}\right) = \left[\phi_X\left(\frac{t}{n}\right)\right]^n\end{aligned}$$
By Theorem 16.2 (with $k = 1$),
$$\phi_X(t) = 1 + it\mu + o(t) \quad \text{as} \quad t \to 0,$$
so by Lemma 16.1
$$\phi_n(t) = \left[\phi_X\left(\frac{t}{n}\right)\right]^n = \left[1 + \frac{it\mu}{n} + o\left(\frac{t}{n}\right)\right]^n \to e^{it\mu} \quad \text{as} \quad n \to \infty.$$
This is the characteristic function of the constant μ, and the result follows by the continuity theorem for characteristic functions.

16.3 Central limit theorem

Let $X_1, X_2, \dots$ be i.i.d. random variables, and consider the partial sums

$$S_n = X_1 + X_2 + \ldots + X_n.$$

By independence, $\mathbb{E}(S_n) = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$.

The central limit theorem says that, *irrespective of the distribution of the X_i*, the distribution of the standardised variables

$$S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges to the standard normal distribution as $n \to \infty$.

Theorem 16.5 (Central limit theorem)
Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed with common mean μ and variance σ^2. If μ and σ^2 are both finite, then the distribution of the normalised sums

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} \qquad \text{where} \qquad S_n = X_1 + \ldots + X_n,$$

converges to the standard normal distribution $\mathrm{N}(0,1)$ as $n \to \infty$.

Proof: Let $Y_i = \dfrac{X_i - \mu}{\sigma}$. Then $\mathbb{E}(Y_i) = 0$ and $\mathrm{Var}(Y_i) = 1$, and

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

- Let $\phi_Y(t)$ denote the common characteristic function of the Y_i.
- Let $\phi_n(t)$ denote the characteristic function of S_n^*.

By Taylor's theorem, if $\mathbb{E}(|Y^k|) < \infty$ we have that

$$\phi(t) = \mathbb{E}(e^{itY}) = \sum_{j=0}^{k} \frac{\mathbb{E}(Y^j)}{j!}(it)^j + o(t^k) \qquad \text{as} \quad t \to 0.$$

Since $\mathbb{E}(Y^2) = \mathrm{Var}(Y) + \mathbb{E}(Y)^2 = 1$ is finite, we can apply this with $k = 2$ to obtain

$$\phi_Y(t) = 1 - \frac{1}{2}t^2 + o(t^2) \quad \text{as} \quad t \to 0$$

By the properties of characteristic functions,

$$\begin{aligned}
\phi_n(t) &= \phi_{\frac{1}{\sqrt{n}}(Y_1+Y_2+\ldots+Y_n)}(t) \\
&= \phi_{Y_1+Y_2+\ldots+Y_n}\left(\frac{t}{\sqrt{n}}\right) \\
&= \left[\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \\
&\to e^{-\frac{1}{2}t^2} \quad \text{as} \quad n \to \infty,
\end{aligned}$$

where the last step follows by Lemma 16.1. This is the characteristic function of the $\mathrm{N}(0,1)$ distribution, and the result follows by the continuity theorem for characteristic functions.

Example 16.6 (Erlang Distribution)
The *Erlang distribution* with parameters $k \in \mathbb{N}$ and $\lambda > 0$ is defined to be the sum of k independent and identically distributed random variables $X_1, X_2, \ldots, X_k$, where each X_i is exponentially distributed with (rate) parameter λ. Show that if $Y \sim \text{Erlang}(k, \lambda)$, then the random variable

$$Z_k = \frac{\lambda Y - k}{\sqrt{k}}$$

has approximately the standard normal distribution when k is large.

Solution: Let $Y \sim \text{Erlang}(k, \lambda)$. Then Y can be written as the sum of k independent and identically distributed random variables X_i:

$$Y = X_1 + X_2 + \ldots + X_k \qquad \text{where} \quad X_i \sim \text{Exponential}(\lambda).$$

Since $X_i \sim \text{Exponential}(\lambda)$ with $\lambda > 0$, we have

$$\mathbb{E}(X_i) = \frac{1}{\lambda} < \infty \quad \text{and} \quad \text{Var}(X_i) = \frac{1}{\lambda^2} < \infty.$$

Furthermore, by independence we have

$$\mathbb{E}(Y) = \sum_{i=1}^{k} \mathbb{E}(X_i) = \frac{k}{\lambda} \qquad \text{and} \qquad \text{Var}(Y) = \sum_{i=1}^{k} \text{Var}(X_i) = \frac{k}{\lambda^2}.$$

Let $Z \sim \text{N}(0,1)$. By the central limit theorem,

$$\frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var}(Y)}} \to Z \quad \text{in distribution as } k \to \infty.$$

i.e.

$$Z_k = \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var}(Y)}} = \frac{Y - k/\lambda}{\sqrt{k/\lambda^2}} = \frac{\lambda Y - k}{\sqrt{k}} \to Z \quad \text{in distribution as } k \to \infty.$$

16.4 Exercises

Exercise 16.1

1. The continuous uniform distribution on (a, b) has the following PDF:

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

   Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

2. The exponential distribution with scale parameter $\theta > 0$ has the following PDF:

$$f(x) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

3. Let $X \sim \text{Binomial}(n_1, p_1)$ and $X_2 \sim \text{Binomial}(n_2, p_2)$ be independent random variables.
   (1) Use the central limit theorem to find the approximate distribution of $Y = X_1 - X_2$ when n_1 and n_2 are both large.

(2) Let $Y_1 = X_1/n_1$ and $Y_2 = X_2/n_2$. Show that $Y_1 - Y_2$ is approximately normally distributed with mean $p_1 - p_2$ and variance $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$ when n_1 and n_2 are both large.

(3) Show that when n_1 and n_2 are both large,

$$\frac{(Y_1 - Y_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \qquad \text{approx.}$$

4. 5% of items produced by a factory production line are defective. Items are packed into boxes of 2000 items. As part of a quality control exercise, a box is chosen at random and found to contain 120 defective items. Use the central limit theorem to estimate the probability of finding at least this number of defective items when the production line is operating properly.

5. Use the central limit theorem to prove the law of large numbers.

6. We perform a sequence of independent Bernoulli trials, each with probability of success p, until a fixed number r of successes is obtained. The total number of failures Y (up to the rth succes) has the *negative binomial* distribution with parameters r and p, so the PMF of Y is

$$\mathbb{P}(Y = k) = \binom{k+r-1}{k}(1-p)^k p^r, \qquad k = 0, 1, 2, \ldots$$

Using the fact that Y can be written as the sum of r independent geometric random variables, show that this distribution can be approximated by a normal distribution when r is large.

Lecture 17 Joint Distributions

To be read in preparation for the **11.00** lecture on **Mon 24 Nov** in **E/0.15**.

17.1 Joint distributions

Definition 17.1
Let $X, Y : \Omega \to \mathbb{R}$ be random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(1) The *joint distribution* of X and Y is the function

$$\begin{array}{rccc} \mathbb{P}_{X,Y}: & \mathcal{B}^2 & \to & [0,1] \\ & (A,B) & \mapsto & \mathbb{P}(X \in A, Y \in B). \end{array}$$

(2) The *joint CDF* of X and Y is

$$\begin{array}{rccc} F_{X,Y}: & \mathbb{R}^2 & \to & [0,1] \\ & (x,y) & \mapsto & \mathbb{P}(X \leq x, Y \leq y). \end{array}$$

(3) The *marginal CDF* of X is the function

$$\begin{array}{rccc} F_X: & \mathbb{R} & \to & [0,1] \\ & x & \mapsto & \mathbb{P}(X \leq x), \end{array}$$

and the marginal CDF of Y is

$$\begin{array}{rccc} F_Y: & \mathbb{R} & \to & [0,1] \\ & y & \mapsto & \mathbb{P}(Y \leq y). \end{array}$$

17.2 Properties of Joint CDFs

Theorem 17.2
Let $F : \mathbb{R}^2 \to [0,1]$ be a joint CDF.

(1) Limiting behaviour:

$$\begin{array}{lll} \lim_{x\to-\infty} F(x,y) = 0, & \lim_{y\to-\infty} F(x,y) = 0, & \lim_{\substack{x\to-\infty \\ y\to-\infty}} F(x,y) = 0, \\ \lim_{x\to+\infty} F(x,y) = F_Y(y), & \lim_{y\to+\infty} F(x,y) = F_X(x), & \lim_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1. \end{array}$$

(2) Monotonicity:

$$F(x,y) \le F(x+u, y+v) \quad \text{for all } u, v \ge 0.$$

(3) Inclusion-exclusion:

$$\mathbb{P}\big(a < X \le b,\ c < Y \le d\big) = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$

(4) Upper continuity:

$$F(x+u, y+v) \longrightarrow F(x,y) \quad \text{as} \quad u \downarrow 0 \text{ and } v \downarrow 0,$$

where $u \downarrow 0$ means that u converges to zero through positive values (a.k.a. "from above").

[*Proof omitted.*]

17.3 Independent random variables

Recall that two events A and B are called *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition 17.3
Two random variables $X, Y : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are said to be *independent* if the events

$$\begin{aligned} \{X \le x\} &\equiv \{\omega : X(\omega) \le x\} \\ \{Y \le y\} &\equiv \{\omega : Y(\omega) \le y\} \end{aligned}$$

are independent for all $x, y \in \mathbb{R}$.

The following lemma is easily proved.

Lemma 17.4
Let X and Y be random variables with joint CDF $F_{X,Y}(x,y)$ and marginal CDFs $F_X(x)$ and $F_Y(y)$ respectively. Then X and Y are independent if and only if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \text{for all} \quad x, y \in \mathbb{R}.$$

17.4 Identically distributed random variables

Definition 17.5
Two random variables $X, Y : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are said to be *identically distributed* if $\mathbb{P}_X = \mathbb{P}_Y$, or equivalently $F_X = F_Y$.

Thus X and Y are identically distributed if and only if

- $\mathbb{P}_X(B) \equiv \mathbb{P}(X \in B) = \mathbb{P}(Y \in B) \equiv \mathbb{P}_Y(B)$ for all $B \in \mathcal{B}$, or equivalently
- $F_X(t) \equiv \mathbb{P}(X \le t) = \mathbb{P}(Y \le t) \equiv F_Y(t)$ for all $t \in \mathbb{R}$.

17.5 Jointly discrete distributions

Definition 17.6

(1) Two random variables X and Y are called *jointly discrete* if the random vector (X, Y) only takes values in a countable subset of $\mathbb{R}^2$.

(2) Two jointly discrete random variables X and Y are described by their *joint PMF*:

$$\begin{aligned} f_{X,Y}: \quad \mathbb{R}^2 &\to [0,1] \\ (x,y) &\mapsto \mathbb{P}(X=x, Y=y). \end{aligned}$$

(3) The *marginal PMF of* X is the function $f_X(x) = \mathbb{P}(X = x)$.

(4) The *marginal PMF of* Y is the function $f_Y(y) = \mathbb{P}(Y = y)$.

Example 17.7

A fair die is rolled once. Let ω denote the outcome, and consider the random variables

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd,} \\ 2 & \text{if } \omega \text{ is even,} \end{cases} \quad \text{and} \quad Y(\omega) = \begin{cases} 1 & \text{if } \omega \leq 3, \\ 2 & \text{if } \omega \geq 4. \end{cases}$$

Find the joint PMF of X and Y.

Solution:

ω	1	2	3	4	5	6
$X(\omega)$	1	2	1	2	1	2
$Y(\omega)$	1	1	1	2	2	2

The joint PMF of X and Y is shown in the following table

	Y=1	Y=2
X=1	1/3	1/6
X=2	1/6	1/3

The marginal PMFs of X and Y are recovered by summing across the rows and columns.

The following lemma is easily proved.

Lemma 17.8

Two jointly discrete random variables X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \text{for all} \quad x, y \in \mathbb{R}.$$

17.6 Jointly continuous distributions

Definition 17.9

(1) Two random variables X and Y are called *jointly continuous* if their joint CDF can be written as

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, du\, dv \qquad x, y \in \mathbb{R}$$

for some integrable function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ called the *joint PDF* of X and Y.

(2) The *marginal PDFs* of X and Y are defined by $f_X(x) = F_X'(x)$ and $f_Y(y) = F_Y'(y)$ respectively, where $F_X(x)$ and $F_Y(y)$ are the marginal CDFs of X and Y respectively.

Example 17.10
A dart is thrown at a circular dartboard of radius ρ. The point at which the dart hits the board determines a distance R from the centre, and an angle Θ with (say) the upward vertical. Assume that the dart does in fact hit the board, and that regions of equal area are equally likely to be hit. Show that R and Θ are jointly continuous random variables.

Solution:

$$\begin{aligned}\mathbb{P}(R \leq r) &= \frac{\pi r^2}{\pi \rho^2} = \frac{r^2}{\rho^2} \qquad \text{for } 0 \leq r \leq \rho. \\ \mathbb{P}(\Theta \leq \theta) &= \frac{\theta}{2\pi} \qquad\qquad \text{for } 0 \leq \theta \leq 2\pi.\end{aligned}$$

Since regions of equal area are equally likely to be hit, the joint distribution function of (R, Θ) satisfies

$$F_{R,\Theta}(r,\theta) = \mathbb{P}(R \leq r, \Theta \leq \theta) = \mathbb{P}(R \leq r)\mathbb{P}(\Theta \leq \theta) = \frac{\theta r^2}{2\pi\rho^2}$$

Thus we have

$$F_{R,\Theta}(r,\theta) = \int_0^r \int_0^\theta f(u,v)\,du\,dv$$

where

$$f(u,v) = \frac{u}{\pi\rho^2}.$$

Hence R and Θ are jointly continuous random variables.

17.6.1 Independence

Lemma 17.11
Two jointly continuous random variables X and Y are independent if and only if

(1) $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$, and

(2) the support of $f_{X,Y}$ is a rectangular region in $\mathbb{R}^2$.

Proof: Since X and Y are jointly continuous,

$$\begin{aligned}F_{X,Y}(x,y) &= \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v)\,du\,dv \\ F_X(x)F_Y(y) &= \left(\int_{-\infty}^x f_X(u)\,du\right)\left(\int_{-\infty}^y f_Y(v)\,dv\right)\end{aligned}$$

- By independence, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, so the two integrals are equal.
- Differentiating both sides with respect to x and y, we get $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Remark 17.12
If the value taken by X affects the range of values taken by Y, then clearly Y dependes on X. Hence if X and Y are independent, we need that $\text{supp}(f_{X,Y})$ can be expressed as the Cartesian product of two sets in $\mathbb{R}$:

$$\text{supp}(f_{X,Y}) = \text{supp}(f_X) \times \text{supp}(f_Y) \qquad \text{where} \quad \text{supp}(f_X), \text{supp}(f_Y) \subseteq \mathbb{R}.$$

For example:

- The unit square is fine: $\text{supp}(f_{X,Y}) = \{(x,y) : 0 \leq x, y \leq 1\} = [0,1] \times [0,1]$.
- The unit disc is not: $\text{supp}(f_{X,Y}) = \{(x,y) : x^2 + y^2 \leq 1\}$.

Example 17.13
Two jointly continuous random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} c(1-x)y & 0 \leq y \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(1) Show that $c = 24$.

(2) Find the marginal PDFs of X and Y.

Solution:

(1)
- For fixed $x \in [0,1]$, we must have $y \in [0,x]$.
- For fixed $y \in [0,1]$, we must have $x \in [y,1]$.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^x c(1-x)y\,dydx = \frac{c}{2}\int_0^1 x^2 - x^3\,dx = \frac{c}{2}\left(\frac{1}{3}-\frac{1}{4}\right) = \frac{c}{24}.$$

By the law of total probability this must equal 1, so $c = 24$.

(2) The marginal density functions are:

$$f_X(x) = 24\int_0^x (1-x)y\,dy = 24(1-x)\left[\frac{y^2}{2}\right]_0^x = \begin{cases} 12x^2(1-x) & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = 24\int_y^1 (1-x)y\,dx = 24y\left[x - \frac{x^2}{2}\right]_y^1 = \begin{cases} 12y(1-y)^2 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Example 17.14
The continuous random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} cxy^2 & 0 < x, y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

(1) Show that $c = 6$.

(2) Find the marginal PDFs of X and Y. Are X and Y independent?

(3) Show that $\mathbb{P}(X + Y \geq 1) = 9/10$.

Solution:

(1) By the law of total probability, we must have that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^1 cxy^2\,dx\,dy = 1.$$

Evaluating the integral,

$$\int_0^1\int_0^1 cxy^2\,dx\,dy = c\int_0^1\left[\frac{x^2y^2}{2}\right]_0^1 dy = c\int_0^1 \frac{y^2}{2}\,dy = c\left[\frac{y^3}{6}\right]_0^1 = \frac{c}{6}$$

Thus we have $c = 6$, so $f_{X,Y}(x,y) = 6xy^2$ over $0 \leq x, y \leq 1$ (and zero otherwise).

(2) The marginal densities are as follows:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy = \int_0^1 6xy^2\,dy = 6x\left[\frac{y^3}{3}\right]_0^1 = \begin{cases} 2x & 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,dx = \int_0^1 6xy^2\,dx = 6y^2\left[\frac{x^2}{2}\right]_0^1 = \begin{cases} 3y^2 & 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

X and Y are independent because the support of $f_{X,Y}(x,y)$ is a rectangular region, and

$$f_{X,Y}(x,y) = 6xy^2 = (2x)(3y^2) = f_X(x)f_Y(y) \quad \text{for all} \quad x,y \in \mathbb{R}.$$

(3) The event $\{X+Y \geq 1\}$ is the same as $\{X \geq 1-Y\}$. For a fixed value $Y=y$,

$$\mathbb{P}(X+y \geq 1) = \mathbb{P}(X \geq 1-y) = \int_{1-y}^1 f_X(x)\,dx = \int_{1-y}^1 2x\,dx = 1-(1-y)^2.$$

Hence,

$$\begin{aligned} P(X+Y \geq 1) &= \int_0^1 \mathbb{P}(X+y \geq 1)f_Y(y)\,dy \\ &= \int_0^1 3y^2\big(1-(1-y)^2\big)\,dy = \int_0^1 6y^3 - 3y^4\,dy = \left[\frac{6y^4}{4} - \frac{3y^5}{5}\right]_0^1 = \frac{9}{10}. \end{aligned}$$

17.7 Exercises

Exercise 17.1

1. Let X be a Bernoulli random variable with parameter p.
   (a) Let $Y = 1-X$. Find the joint PMF of X and Y.
   (b) Let $Y = 1-X$ and $Z = XY$. Find the joint PMF of X and Z.

2. Let X and Y be two independent discrete random variables with the following PMFs:

x	1	2
$f_X(x)$	1/3	2/3

y	-1	0	1
$f_Y(y)$	1/4	1/2	1/4

   (a) Compute the joint PMF of X and Y.
   (b) Compute the joint PMF of the random variables $U = 1/X$ and $V = Y^2$.
   (c) Show that U and V are independent.

3. Two discrete random variables X and Y have the following joint PMF:

$$f_{X,Y}(x,y) = \begin{cases} c|x+y| & \text{for} \quad x,y \in \{-2,-1,0,1,2\}, \\ 0 & \text{otherwise,} \end{cases}$$

   where c is a constant.
   (a) Show that $c = 1/40$.
   (b) Find $\mathbb{P}(X=0, Y=-2)$.
   (c) Find $\mathbb{P}(X=2)$.
   (d) Find $\mathbb{P}(|X-Y| \leq 1)$.

4. Two continuous random variables X and Y have the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2x & \text{if} \quad 0 \leq x, y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the conditional distribution of Y given that $X = x$.
   (b) Find $\mathbb{P}(Y \leq 0.5 | X = 0.5)$ and $\mathbb{P}(Y \leq 0.5 | X = 0.75)$.
   (c) Find the marginal distribution of Y and hence find $\mathbb{P}(Y \leq 0.5)$.

5. Two continuous random variables X and Y have the following joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} c(x^2 + y) & \text{when } -1 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 - x^2, \\ 0 & \text{otherwise.} \end{cases}$$

   where c is a constant.
   (a) Show that $c = 5/4$.
   (b) Find $\mathbb{P}(0 \leq X \leq 0.5)$.
   (c) Find $\mathbb{P}(Y \leq X + 1)$.
   (d) Find $\mathbb{P}(Y = X^2)$.

Lecture 18 Covariance and Correlation

To be read in preparation for the **11.00** lecture on **Wed 26 Nov** in **Tower 0.02**.

18.1 Bivariate Distributions

Definition 18.1
Let $X, Y : \Omega \to \mathbb{R}$ be two random variables defined on the same probability space, let $F_{X,Y}$ denote their joint CDF, and let $g : \mathbb{R}^2 \to \mathbb{R}$ be a (Borel) measureable function on $\mathbb{R}^2$. Then

$$\mathbb{E}\big[g(X,Y)\big] = \iint g(x,y)\,dF(x,y)$$

whenever this integral exists. In particular,

(1) if X and Y are jointly discrete, with joint PMF $f_{X,Y}(x,y)$, then

$$\mathbb{E}\big[g(X,Y)\big] = \sum_{x,y} g(x,y) f_{X,Y}(x,y)$$

whenever this sum exists, and

(2) if X and Y are jointly continuous, with joint PDF $f_{X,Y}(x,y)$, then

$$\mathbb{E}\big[g(X,Y)\big] = \iint g(x,y) f_{X,Y}(x,y)\,dx\,dy$$

whenever this integral exists.

18.2 Covariance

Definition 18.2
The *product moment* of X and Y is defined to be

$$\mathbb{E}(XY) = \iint xy\,dF(x,y)$$

whenever this integral exists. In particular,

(1) if X and Y are jointly discrete, with joint PMF $f_{X,Y}(x,y)$, then

$$\mathbb{E}(XY) = \sum_{x,y} xy\, f_{X,Y}(x,y)$$

whenever this sum is absolutely convergent, and

(2) if X and Y are jointly continuous, with joint PDF $f_{X,Y}(x,y)$, then

$$\mathbb{E}(XY) = \iint xy\, f_{X,Y}(x,y)\, dx\, dy$$

whenever this integral is absolutely convergent.

Definition 18.3

(1) The *covariance* of X and Y is

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

(2) The *correlation coefficient* of X and Y is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

Remark 18.4

- $\mathrm{Cov}(X,Y)$ is the product moment of the *centred* variables $X-\mathbb{E}(X)$ and $Y-\mathbb{E}(Y)$.
- $\rho(X,Y)$ is the product moment of the *standardized* variables $\dfrac{X-\mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}$ and $\dfrac{Y-\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}$.

Remark 18.5 (Variance of sums of random variables)

For any random variables $X_1, X_2, \dots, X_n$,

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^n \sum_{j=i+1}^n \mathrm{Cov}(X_i, X_j).$$

18.3 Correlation

Correlation quantifies the (linear) dependence between random variables.

Definition 18.6

Two random variables X and Y are said to be *correlated* if $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$.

Lemma 18.7

If X and Y are independent, they are uncorrelated.

We prove the lemma only for discrete random variables (the continuous case is similar).

Proof: Since X and Y are independent, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$. Hence

$$\begin{aligned}\mathbb{E}(XY) = \sum_{x,y} xy\, f_{X,Y}(x,y) &= \sum_{x,y} xy\, f_X(x)f_Y(y) \qquad \text{(by independence)}, \\ &= \left(\sum_x x\, f_X(x)\right)\left(\sum_y y\, f_Y(y)\right) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

Theorem 18.8
If X and Y are uncorrelated, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Proof: Let X and Y be uncorrelated. Then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, so

$$\begin{aligned}\text{Var}(X+Y) &= \mathbb{E}\left(\big[X+Y-\mathbb{E}(X+Y)\big]^2\right) \\ &= \mathbb{E}\left(\big[X-\mathbb{E}(X)\big]^2 + 2\big[XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)\big] + \big[Y-\mathbb{E}(Y)\big]^2\right) \\ &= \text{Var}(X) + 2\left[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right] + \text{Var}(Y) \\ &= \text{Var}(X) + \text{Var}(Y).\end{aligned}$$

Example 18.9
Let $Y_1, \ldots, Y_r$ be independent and identically distributed random variables, with each $Y_i \sim \text{Geometric}(p)$. Find the mean and variance of their sum $X = \sum_{i=1}^r Y_i$.

Solution: Since $Y_i \sim \text{Geometric}(p)$, we know that $\mathbb{E}(Y_i) = \dfrac{1-p}{p}$ and $\text{Var}(Y_i) = \dfrac{1-p}{p^2}$. Hence by the linearity of expectation,

$$\mathbb{E}(X) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \ldots + \mathbb{E}(Y_r) = \frac{r(1-p)}{p}$$

and because the Y_i are independent,

$$\text{Var}(X) = \text{Var}(Y_1) + \text{Var}(Y_2) + \ldots + \text{Var}(Y_r) = \frac{r(1-p)}{p^2}$$

Remark: In this case, X has the *negative binomial* distribution, with parameters r and p.

18.4 The Cauchy-Schwarz Inequality

Lemma 18.10
If $X \geq 0$ and $\mathbb{E}(X) = 0$ then $\mathbb{P}(X = 0) = 1$.

[*Proof omitted.*]

Theorem 18.11 (Cauchy-Schwarz inequality)
For any two random variables X and Y,

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

with equality if and only if $\mathbb{P}(Y = aX) = 1$ for some $a \in \mathbb{R}$.

Proof: By Lemma 18.10, X^2 and Y^2 are non-negative random variables, so we can assume that $\mathbb{E}(X^2) > 0$ and $\mathbb{E}(Y^2) > 0$ (otherwise both sides of the inequality are identically zero).
Let $a \in \mathbb{R}$ consider the random variable $Z = aX - Y$.
By the properties of expectation,

$$\begin{aligned}Z^2 \geq 0 \text{ for all } a \in \mathbb{R} &\Rightarrow \mathbb{E}(Z^2) \geq 0 \text{ for all } a \in \mathbb{R} \\ &\Rightarrow \mathbb{E}(a^2X^2 - 2aXY + Y^2) \geq 0 \text{ for all } a \in \mathbb{R} \\ &\Rightarrow a^2\mathbb{E}(X^2) - 2a\mathbb{E}(XY) + \mathbb{E}(Y^2) \geq 0 \text{ for all } a \in \mathbb{R}.\end{aligned}$$

- Let $h(a) = a^2\mathbb{E}(X^2) - 2a\mathbb{E}(XY) + \mathbb{E}(Y^2)$. This is a quadratic expression in a.

Since $h(a) \geq 0$ for all $a \in \mathbb{R}$, the roots of the quadratic equation $h(a) = 0$, given by

$$a = \frac{\mathbb{E}(XY) \pm \sqrt{\mathbb{E}(XY)^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2)}}{\mathbb{E}(X^2)}$$

are either both complex (discriminant is negative) or co-incide (discriminant is zero).

- Hence $\mathbb{E}(XY)^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$, or equivalently, $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$.

Finally, the discriminant is zero if and only if the quadratic has real roots, which occurs if and only if $\mathbb{E}(Z^2) = 0$ for some $a \in \mathbb{R}$, i.e.

$$\mathbb{E}\left((aX - Y)^2\right) = 0 \qquad \text{for some } a \in \mathbb{R}.$$

Hence, because $(aX - Y)^2$ is a non-negative random variable, we have by Lemma 18.10 that

$$\mathbb{E}\big[(aX - Y)^2\big] = 0 \quad \Rightarrow \quad \mathbb{P}\big[(aX - Y)^2 = 0\big] = 1 \quad \Rightarrow \quad \mathbb{P}\big(Y = aX\big) = 1.$$

Corollary 18.12
The correlation coefficient satisfies the inequality

$$|\rho(X, Y)| \leq 1,$$

with equality if and only if $\mathbb{P}(Y = aX + b) = 1$ for some $a \in \mathbb{R}$.

Proof: Apply the Cauchy-Schwarz inequality to $X - \mathbb{E}X$ and $Y - \mathbb{E}Y$:

$$\begin{aligned}\operatorname{Cov}(X, Y)^2 &= \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) \\ &\leq \mathbb{E}\big((X - \mathbb{E}X)^2\big)\mathbb{E}\big((Y - \mathbb{E}Y)^2\big) = \operatorname{Var}(X)\operatorname{Var}(Y),\end{aligned}$$

with equality if and only if there exists $a \in \mathbb{R}$ such that

$$\mathbb{P}\big[Y - \mathbb{E}Y = a(X - \mathbb{E}X)\big] = 1.$$

Hence,

$$|\rho(X, Y)| = \left|\frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}\right| \leq 1$$

with equality if and only if $\mathbb{P}(Y = aX + b) = 1$, where $b = \mathbb{E}Y - a\mathbb{E}X$.

18.5 Exercises

Exercise 18.1

1. Let X and Y be two random variables having the same distribution but which are not necessarily independent. Show that
$$\operatorname{Cov}(X + Y, X - Y) = 0$$
provided that their common distribution has finite mean and variance.

2. Consider a fair six-sided die whose faces show the numbers $-2, 0, 0, 1, 3, 4$. The die is independently rolled four times. Let X be the average of the four numbers that appear, and let Y be the product of these four numbers. Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(Y)$ and $\operatorname{Cov}(X, Y)$.

3. A fair die is rolled twice. Let U denote the number obtained on the first roll, let V denote the number obtained on the second roll, let $X = U + V$ denote their sum and let $Y = U - V$ denote their difference. Compute the mean and variance of X and Y, and compute $\mathbb{E}(XY)$. Check whether X and Y are uncorrelated. Check whether X and Y are independent.

Lecture 19 Conditional Distributions

To be read in preparation for the **11.00** lecture on **Mon 01 Dec** in **E/0.15**.

19.1 Conditional distributions

Let $X, Y : \Omega \to \mathbb{R}$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 19.1

(1) The *conditional distribution* of Y given X is the function

$$\begin{array}{rccl} \mathbb{P}_{Y|X} : & \mathcal{B}^2 & \to & [0,1] \\ & (A,B) & \mapsto & \mathbb{P}(Y \in B \mid X \in A). \end{array}$$

(2) The *conditional CDF* of Y given X is the function

$$\begin{array}{rccl} F_{Y|X} : & \mathbb{R}^2 & \to & [0,1] \\ & (x,y) & \mapsto & \mathbb{P}(Y \leq y \mid X \leq x). \end{array}$$

The following lemma is easily proved.

Lemma 19.2
The conditional CDF of Y given X satisfies

$$F_{Y|X}(x,y) = \frac{F_{X,Y}(x,y)}{F_X(x)},$$

where $F_{X,Y}$ is the joint CDF of X and Y, and F_X is the marginal CDF of X.

19.1.1 Discrete case

Definition 19.3
Let X and Y be jointly discrete random variables, and let x be such that $\mathbb{P}(X = x) > 0$. The *conditional PMF* of Y given $X = x$ is the function

$$\begin{array}{rccl} f_{Y|X} : & \mathbb{R} & \to & [0,1] \\ & y & \mapsto & \mathbb{P}(Y = y \mid X = x). \end{array}$$

The following lemma is easily proved.

Lemma 19.4
The conditional PMF of Y given $X = x$ satisfies

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

Theorem 19.5
If X and Y are jointly discrete random variables, then

$$f_Y(y) = \sum_x f_{Y|X}(y|x) f_X(x).$$

where the sum is taken over the range of X.

Proof:
$$\sum_x f_{Y|X}(y|x) f_X(x) = \sum_x \left(\frac{f_{X,Y}(x,y)}{f_X(x)}\right) f_X(x) = \sum_x f_{X,Y}(x,y) = f_Y(y).$$

19.1.2 Continuous case

Let X and Y be jointly continuous random variables.

- Suppose we observe that X takes the value x.
- Since $\mathbb{P}(X = x) = 0$, we cannot condition on the event $\{X = x\}$.

Definition 19.6
Let X and Y be jointly continuous random variables. The *conditional PDF* of Y given $X = x$ is

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

Theorem 19.7
If X and Y are jointly continuous random variables, then

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,dx.$$

Proof:
$$\int f_{Y|X}(y|x) f_X(x)\,dx = \int \left(\frac{f_{X,Y}(x,y)}{f_X(x)}\right) f_X(x)\,dx = \int f_{X,Y}(x,y)\,dx = f_Y(y).$$

19.2 Conditional expectation

Definition 19.8
(1) The *conditional expectation of* Y *given* $X = x$ is a number,

$$\mathbb{E}(Y|X=x) \;=\; \begin{cases} \displaystyle\sum_y y\, f_{Y|X}(y|x) & \text{(discrete case),} \\ \displaystyle\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy & \text{(continuous case).} \end{cases}$$

(2) The *conditional expectation of Y given X* is a random variable,

$$\begin{aligned} \mathbb{E}(Y|X): \quad \Omega &\to \mathbb{R} \\ \omega &\mapsto \mathbb{E}\big(Y|X = X(\omega)\big). \end{aligned}$$

Remark 19.9
Let $g(x) = \mathbb{E}(Y|X = x)$. The distribution of the random variable $g(X) = \mathbb{E}(Y|X)$ depends only on the distribution of X. Its expectation is given by

$$\mathbb{E}\big[\mathbb{E}(Y|X)\big] = \begin{cases} \displaystyle\sum_x \mathbb{E}(Y|X = x) f_X(x) & \text{(discrete case)}, \\ \displaystyle\int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) f_X(x)\,dx & \text{(continuous case)}. \end{cases}$$

where f_X is the marginal PMF or PDF of X.

19.3 Law of total expectation

Theorem 19.10 (Law of total expectation)
Let X and Y be random variables on the same probability space. Then

$$\mathbb{E}\big[\mathbb{E}(Y|X)\big] = \mathbb{E}(Y).$$

We prove the theorem for continuous random variables (the discrete case follows similarly).

Proof: Let X and Y be jointly continuous random variables.

$$\begin{aligned} \mathbb{E}\big[\mathbb{E}(Y|X)\big] &= \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy\right) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y\, \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy\right) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dy\right) dx \\ &= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\right) dy \\ &= \int_{-\infty}^{\infty} y f_Y(y)\,dy \\ &= \mathbb{E}(Y). \end{aligned}$$

19.4 Law of total variance

Theorem 19.11 (Law of total variance)
Let X and Y be random variables on the same probability space. Then

$$\mathrm{Var}(Y) = \mathbb{E}\big[\mathrm{Var}(Y|X)\big] + \mathrm{Var}\big[\mathbb{E}(Y|X)\big]$$

This is sometimes called the *variance decomposition formula*.

Proof:

- By the law of total expectation,

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\ &= \mathbb{E}\big[\mathbb{E}(Y^2|X)\big] - \mathbb{E}\big[\mathbb{E}(Y|X)\big]^2\end{aligned}$$

- Because $\mathrm{Var}(Y|X) = \mathbb{E}(Y^2|X) - \mathbb{E}(Y|X)^2$,

$$\mathrm{Var}(Y) = \mathbb{E}\big[\mathrm{Var}(Y|X) + \mathbb{E}(Y|X)^2\big] - \mathbb{E}\big[\mathbb{E}(Y|X)\big]^2$$

- Hence, by the linearity of expectation,

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}\big[\mathrm{Var}(Y|X)\big] + \Big(\mathbb{E}\big[\mathbb{E}(Y|X)^2\big] - \mathbb{E}\big[\mathbb{E}(Y|X)\big]^2\Big) \\ &= \mathbb{E}\big[\mathrm{Var}(Y|X)\big] + \mathrm{Var}\big[\mathbb{E}(Y|X)\big].\end{aligned}$$

19.4.1 Variance decomposition

$\mathbb{E}(Y|X)$ can be thought of as a *model* of Y in terms of X.

- $\mathrm{Var}\big[\mathbb{E}(Y|X)\big]$ is the variance of the model. This is called the *explained variance*.
- $Y - \mathbb{E}(Y|X)$ is called the *residual*, representing that part of Y not explained by the model $\mathbb{E}(Y|X)$.
- $\mathrm{Var}(Y|X) = \mathbb{E}\big([Y - \mathbb{E}(Y|X)]^2 \,\big|\, X\big)$ is called the *residual variance* at X.
- $\mathbb{E}\big[\mathrm{Var}(Y|X)\big]$ is the expected residual variance. This is called the *unexplained variance*.

The law of total variance divides the variance into *unexplained* and *explained* components:

$$\mathrm{Var}(Y) = \mathbb{E}\big[\mathrm{Var}(Y|X)\big] + \mathrm{Var}\big[\mathbb{E}(Y|X)\big]$$

or

$$\frac{\mathbb{E}\big[\mathrm{Var}(Y|X)\big]}{\mathrm{Var}(Y)} + \frac{\mathrm{Var}\big[\mathbb{E}(Y|X)\big]}{\mathrm{Var}(Y)} = 1.$$

This idea is important in statistics.

19.4.2 Linear models

Suppose we adopt a *linear model* of Y against X:

$$\mathbb{E}(Y|X) = a + bX.$$

It can be shown that the residual variance is minimised when

$$a = \mathbb{E}(Y) - \left[\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\right]\mathbb{E}(X) \qquad \text{and} \qquad b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

The proportion of the total variance explained by the model is the square of the correlation coefficient:

$$\frac{\mathrm{Var}\big[\mathbb{E}(Y|X)\big]}{\mathrm{Var}(Y)} = \rho(X,Y)^2.$$

This is known as the *coefficient of determination*, usually denoted by R^2, which quantifies the extent to which a linear model $Y = a + bX$ captures the relationship (if any) between X and Y.

19.5 Example

Example 19.12
The jointly continuous random variables X and Y have following joint PDF:

$$f(x,y) = \begin{cases} \frac{21}{4}x^2y & \text{for } x^2 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(1) Find the marginal PDFs of X and Y.

(2) Find the mean and variance of Y.

(3) Find the conditional PDF of X given $Y = y$.

(4) Find the conditional PDF of Y given $X = x$.

(5) Are X and Y independent?

(6) Verify that $\mathbb{E}(Y) = \mathbb{E}\big[\mathbb{E}(Y|X)\big]$.

Solution: The support of the joint PDF $f(x,y)$ is the set $\{(x,y) : x^2 < y < 1\}$.

- This is the region between the vertical lines $x = -1$ and $x = +1$, bounded above by the horizontal line $y = 1$ and below by parabola $y = x^2$.

In particular,

- For fixed $x \in [-1,1]$, we must have $y \in [x^2, 1]$.
- For fixed $y \in [0,1]$, we must have $x \in [-\sqrt{y}, +\sqrt{y}]$.

(1) The marginal distributions are computed as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2y\,dy = \frac{21}{4}x^2\left[\frac{y^2}{2}\right]_{x^2}^{1} = \begin{cases} \frac{21}{8}x^2(1-x^4) & -1 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2y\,dx = \frac{21}{4}y\left[\frac{x^3}{3}\right]_{-\sqrt{y}}^{\sqrt{y}} = \begin{cases} \frac{7}{2}y^{5/2} & 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(2) The expected value and variance of Y are computed as follows:

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^1 y\left(\frac{7y^{5/2}}{2}\right)dy = \frac{7}{9}, \\ \mathbb{E}(Y^2) &= \int_0^1 y^2\left(\frac{7y^{5/2}}{2}\right)dy = \frac{7}{11}, \\ \mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 &= \frac{7}{11} - \frac{49}{81} = \frac{28}{891}. \end{aligned}$$

(3) The conditional PDF of X given that $Y = y$ is

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{(21/4)x^2y}{(7/2)y^{5/2}} = \begin{cases} \dfrac{3}{2}x^2y^{-3/2} & -\sqrt{y} \le x \le \sqrt{y}, \\ 0 & \text{otherwise.} \end{cases}$$

(4) The conditional PDF of Y given that $X = x$ is

$$f_{Y|X}(y\,|\,x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{(21/4)x^2y}{(21/8)x^2(1-x^4)} = \begin{cases} \dfrac{2y}{1-x^4} & x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(5) X and Y are not independent because the conditional PDF of Y given $X = x$ depends on x.

(6) The conditional expected value of Y given that $X = x$ is

$$\begin{aligned} g(x) = \mathbb{E}(Y|X=x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y\,|\,x)\,dy \\ &= \int_{x^2}^{1} y\left(\frac{2y}{1-x^4}\right)dy = \frac{2}{1-x^4}\left[\frac{y^3}{3}\right]_{x^2}^{1} = \frac{2(1-x^6)}{3(1-x^4)} \end{aligned}$$

Hence the conditional expectation of Y given X is the random variable

$$\mathbb{E}(Y|X) = \frac{2(1-X^6)}{3(1-X^4)}$$

and its expected value is

$$\begin{aligned} \mathbb{E}\big[\mathbb{E}(Y|X)\big] &= \int_{-\infty}^{\infty} \mathbb{E}(Y|X=x) f_X(x)\,dx \\ &= \frac{2}{3}\int_{-1}^{1}\left(\frac{1-x^6}{1-x^4}\right)\left(\frac{21}{8}x^2(1-x^4)\right)dx \\ &= \frac{7}{4}\int_{-1}^{1} x^2(1-x^6)\,dx = \frac{7}{9}. \end{aligned}$$

Thus $\mathbb{E}\big[\mathbb{E}(Y|X)\big] = \mathbb{E}(Y)$ as required.

19.6 Exercises

Exercise 19.1

1. A fair coin is tossed three times. Let I_j be the indicator variable of the event that a head occurs on the jth toss. Compute the conditional expectation $E(Y|X)$ and verify the identity $E(E(Y|X)) = E(Y)$ in each of the following cases:
   (1) $X = \max\{I_1, I_2, I_3\}$ and $Y = \min\{I_1, I_2, I_3\}$,
   (2) $X = I_1 + I_2$ and $Y = I_2 + I_3$.

2. Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} c(x+y) & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   (1) Show that $c = 1$.
   (2) Compute the conditional expectation $\mathbb{E}(Y|X)$.
   (3) Verify the identity $\mathbb{E}\big(\mathbb{E}(Y|X)\big) = \mathbb{E}(Y)$.

3. Let the joint density of random variables X and Y be

$$f(x,y) = \begin{cases} cxy & \text{for} \quad 0 \leq x, y \leq 1 \text{ where } \; x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   (1) Compute the normalization constant c.
   (2) Compute the conditional expectation $\mathbb{E}(Y|X)$.
   (3) Verify the identity $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$.

Lecture 20 The Bivariate Normal Distribution

To be read in preparation for the **11.00** lecture on **Wed 03 Dec** in **Physiology A**.

20.1 Bivariate transformations

Definition 20.1
Let $h : \mathbb{R}^2 \to \mathbb{R}^2$ and let $(u, v) = h(x, y)$. The *Jacobian determinant* of the transformation h is the determinant of its 2×2 matrix of partial derivatives:

$$J = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

Theorem 20.2
Let U and V be jointly continuous random variables, let $f_{U,V}$ be their joint PDF, let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be an injective transform over the support of $f_{U,V}$ and let $(X, Y) = g(U, V)$. Then the joint PDF of X and Y is given by

$$f_{X,Y}(x, y) = |J| f_{U,V}\left[g^{-1}(x, y)\right] \quad \text{where} \quad J = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} \quad \text{with} \quad (u, v) = g^{-1}(x, y).$$

Remark 20.3
The absolute value $|J|$ is a scale factor, which ensures that $f_{X,Y}(x, y)$ integrates to one.

Example 20.4
Let U and V be continuous random variables, and let $X = U + V$ and $Y = U - V$.

(1) Find the joint PDF of X and Y in terms of the joint PDF of U and V.

(2) If $U, V \sim \text{Exponential}(1)$ are independent, find the joint PDF of X and Y.

Solution:

(1) • The transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $g(u, v) = (u + v, u - v)$.

- To compute the inverse transformation $g^{-1}:\mathbb{R}^2\to\mathbb{R}^2$, we solve the equations

$$x = u+v \quad \text{and} \quad y = u-v.$$

- This yields $u=\frac{1}{2}(x+y)$ and $v=\frac{1}{2}(x-y)$.
- Thus the inverse transformation is

$$(u,v) = g^{-1}(x,y) = \left[\frac{1}{2}(x+y), \frac{1}{2}(x-y)\right].$$

The Jacobian determinant is given by

$$J = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}.$$

Hence the joint PDF of X and Y is

$$\begin{aligned} f_{X,Y}(x,y) &= |J| f_{U,V}(u,v) \\ &= \left|-\frac{1}{2}\right| f_{U,V}\left[\frac{1}{2}(x+y), \frac{1}{2}(x-y)\right] \\ &= \frac{1}{2} f_{U,V}\left[\frac{1}{2}(x+y), \frac{1}{2}(x-y)\right]. \end{aligned}$$

(2) Let U and V be independent with $U,V\sim \text{Exponential}(1)$.

By independence, the joint PDF of U and V is

$$f_{U,V}(u,v) = \begin{cases} e^{-(u+v)} & u,v>0 \\ 0 & \text{otherwise.} \end{cases}$$

To compute the support of $f_{X,Y}$, since $u>0$ and $v>0$ we have $x>0$, so

- $\min(y) = \min(u-v) = -x$ (which occurs when $u=0$ and $v=x$), and
- $\max(y) = \max(u-v) = x$ (which occurs when $u=x$ and $v=0$).

Thus, substituting for $u+v = \frac{1}{2}(x+y)+\frac{1}{2}(x-y) = x$, we obtain

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}e^{-x} & \text{for } x>0 \text{ and } -x<y<x, \\ 0 & \text{otherwise.} \end{cases}$$

20.2 The bivariate normal distribution

Theorem 20.5
if $X\sim N(\mu_1,\sigma_1^2)$ and $Y\sim N(\mu_2,\sigma_2^2)$ are independent, then

$$X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2).$$

[*Proof omitted.*]

Corollary 20.6
If $U,V\sim N(0,1)$ are independent, then $aU+bV\sim N(0,a^2+b^2)$ for all $a,b\in\mathbb{R}$.

Definition 20.7
A pair of random variables U and V have the *standard bivariate normal distribution* if their joint PDF $f : \mathbb{R}^2 \to [0, \infty)$ can be written as

$$f_{U,V}(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\big(u^2 - 2\rho uv + v^2\big)\right)$$

where ρ is a constant satisfying $-1 < \rho < 1$.

Definition 20.8
A pair of random variables X and Y are said to have *bivariate normal distribution* with means μ_1 and μ_2, variances σ_1^2 and σ_2^2 and correlation ρ, if their joint PDF can be written as

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

The following lemma can be used to derive many properties of the bivariate normal distribution.

Lemma 20.9
Let $U, V \sim N(0, 1)$ be independent, let $\rho \in (-1, +1)$. Then the random variables

$$\begin{aligned} X &= \mu_1 + \sigma_1 U, \\ Y &= \mu_2 + \sigma_2\big(\rho U + \sqrt{1-\rho^2}V\big) \end{aligned}$$

have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ.

Proof: To find the joint PDF of X and Y, let $g(u, v)$ denote the transformation:

$$g(u, v) = \big[\mu_1 + \sigma_1 u, \mu_2 + \sigma_2(\rho u + \sqrt{1-\rho^2}v)\big].$$

The inverse transformation is

$$g^{-1}(x, y) = \left(\frac{x-\mu_1}{\sigma_1}, \frac{1}{\sqrt{1-\rho^2}}\left[\left(\frac{y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right]\right)$$

The joint PDF of X and Y is $f_{X,Y}(x, y) = |J| f_{U,V}(u, v)$, where J is the Jacobian determinant of the inverse transformation:

$$J = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{\sigma_1} & 0 \\ \dfrac{1}{\rho\sigma_1} & \dfrac{1}{\sigma_2\sqrt{1-\rho^2}} \end{vmatrix} = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

Because U and V are independent,

$$f_{U,V}(u, v) = f_U(u) f_V(v) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) \qquad u, v \in \mathbb{R}.$$

and since

$$\begin{aligned} u^2 + v^2 &= \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{1-\rho^2}\left[\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] \\ &= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] \end{aligned}$$

it follows that

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_1\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

as required.

The following theorem shows that if X and Y have bivariate normal distribution, then any linear combination of X and Y is normally distributed.

Theorem 20.10
Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Then

$$aX + bY \sim N\big(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\sigma_1\sigma_2\rho + b^2\sigma_2^2\big)$$

Proof: Let $Z = aX + bY$, let U and V be independent standard normal random variables, and let

$$\begin{aligned} X' &= \mu_1 + \sigma_1 U \\ Y' &= \mu_2 + \sigma_2\big(\rho U + \sqrt{1-\rho^2}V\big) \end{aligned}$$

By Lemma 20.9, X and Y have the same joint distribution as X' and Y', so $Z = aX + bY$ has the same distribution as

$$Z' = aX' + bY' = (a\mu_1 + b\mu_2) + (a\sigma_1 + b\sigma_2\rho)U + b\sigma_2\sqrt{1-\rho^2}V$$

Because $U, V \sim N(0,1)$ are independent, it follows by Corollary 20.6 that

$$Z' \sim N\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\sigma_1\sigma_2\rho + b^2\sigma_2^2\right),$$

so $Z = aX + bY$ has normal distribution, as required.

20.3 Properties of the bivariate normal distribution

Theorem 20.11
Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Then

(1) $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$,

(2) ρ is the correlation coefficient of X and Y, and

(3) X and Y are independent if and only if $\rho = 0$.

Proof: Let $U, V \sim N(0,1)$ and define

$$\begin{aligned} X &= \mu_1 + \sigma_1 U \\ Y &= \mu_2 + \sigma_2\big(\rho U + \sqrt{1-\rho^2}V\big) \end{aligned}$$

(1) In the proof of Theorem 20.10:
- taking $a = 1$ and $b = 0$ yields $X \sim N(\mu_1, \sigma_1^2)$, and
- taking $a = 0$ and $b = 1$ yields $Y \sim N(\mu_2, \sigma_2^2)$.

(2) Using the fact that $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$ for all $a, b, c, d \in \mathbb{R}$,

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}\big[\mu_1 + \sigma_1 U, \mu_2 + \sigma_2(\rho U + \sqrt{1-\rho^2}V)\big] \\ &= \sigma_1\sigma_2\text{Cov}(U, \rho U + \sqrt{1-\rho^2}V) \\ &= \sigma_1\sigma_2\big[\rho\mathbb{E}(U^2) + \sqrt{1-\rho^2}\mathbb{E}(UV)\big] \\ &= \sigma_1\sigma_2\rho. \end{aligned}$$

Thus $\rho = \dfrac{\text{Cov}(X)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$ as required.

(3) If X and Y are independent, they are uncorrelated. If X and Y are uncorrelated then $\rho = 0$, so

the joint PDF of X and Y satisfies

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)\times\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right) \\ &= f_X(x)f_Y(y). \end{aligned}$$

Because this holds for all $x, y \in \mathbb{R}$, it follows that X and Y are independent.

20.4 Conditional distributions

Theorem 20.12
Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Then the conditional distribution of Y given $X = x$ is also normal, with conditional mean and variance given by

$$\mathbb{E}(Y|X=x) = \mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x-\mu_1),$$

$$\mathrm{Var}(Y|X=x) = \sigma_2^2(1-\rho^2),$$

and the conditional mean and variance of Y given X is

$$\mathbb{E}(Y|X) = \mathbb{E}(Y) + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\big[X - \mathbb{E}(X)\big],$$

$$\mathrm{Var}(Y|X) = \mathrm{Var}(Y)(1-\rho^2).$$

Proof: Let $U, V \sim N(0,1)$ be independent, and define the random variables

$$\begin{aligned} X &= \mu_1 + \sigma_1 U, \\ Y &= \mu_2 + \sigma_2\big[\rho U + \sqrt{1-\rho^2}V\big] \\ &= \mu_2 + \sigma_2\left[\rho\left(\frac{X-\mu_1}{\sigma_1}\right) + \sqrt{1-\rho^2}V\right]. \end{aligned}$$

If X is fixed at x, then Y is a linear transformation of V, so the conditional distribution of Y given that $X = x$ is a normal distribution. Furthermore, since $\mathbb{E}(V) = 0$ and $\mathrm{Var}(V) = 1$ we have

$$\mathbb{E}(Y|X=x) = \mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x-\mu_1)$$

$$\mathrm{Var}(Y|X=x) = \sigma_2^2(1-\rho^2)$$

as required.

20.5 Exercises

Exercise 20.1

1. Let X and Y have standard bivariate normal distribution, with joint PDF given by
$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$
where ρ is a constant satisfying $-1 < \rho < 1$.
   (a) Check that $f(x,y)$ is indeed a joint PDF, by verifying that $f(x,y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$.
   (b) Check that $\text{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy = \rho$.
   (c) Show that if X and Y are uncorrelated, then they are independent.
2. Let X and Y have standard bivariate normal distribution. Find the conditional distribution of Y given $X = x$, and hence show that $\mathbb{E}(Y|X) = \rho X$.
3. Let X and Y have standard bivariate normal distribution. Show that X and $Z = \dfrac{Y - \rho X}{\sqrt{1-\rho^2}}$ are independent standard normal random variables.
4. Let X and Y have standard bivariate normal distribution, and let $Z = \max\{X, Y\}$. Show that $\mathbb{E}(Z) = \sqrt{(1-\rho)/\pi}$ and $\mathbb{E}(Z^2) = 1$.
5. Let $U, V \sim N(0,1)$. Show that the random variables $X = U + V$ and $Y = U - V$ are independent.
6. Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Show that the conditional distribution of Y given $X = x$ is
$$N\left(\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x - \mu_1), \sigma_2^2(1-\rho^2)\right).$$
7. (a) Let X and Y be jointly continuous random variables, and let $f_{X,Y}$ be their joint PDF. Show that the PDF of the random variable $X + Y$ can be written as
$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_{X,Y}(x, t-x)\,dx = \int_{-\infty}^{\infty} f_{X,Y}(t-y, y)\,dy.$$
   (b) Hence, or otherwise, show that if $U, V \sim N(0,1)$ are independent, then $U + V \sim N(0,2)$. (This is a special case of Theorem 20.5.)

Exercises

Exercise 2.1

1. Let $\mathcal{F}$ be a field over Ω. Show that
   (a) $\emptyset \in \mathcal{F}$,
   (b) $\mathcal{F}$ is closed under set differences,
   (c) $\mathcal{F}$ is closed under pairwise intersections,
   (d) $\mathcal{F}$ is closed under finite unions,
   (e) $\mathcal{F}$ is closed under finite intersections.
2. Let $\mathcal{F}$ be a σ-field over Ω. Show that
   (a) $\mathcal{F}$ is closed under finite unions,
   (b) $\mathcal{F}$ is closed under finite intersections.
   (c) $\mathcal{F}$ is closed under countable intersections.

Exercise 2.2

1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$.
   (a) What is the smallest σ-field containing the event $A = \{1, 2\}$?
   (b) What is the smallest σ-field containing the events $A = \{1, 2\}$, $B = \{3, 4\}$ and $C = \{5, 6\}$?
2. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-fields over Ω.
   (a) Show that $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ is a σ-field over Ω.
   (b) Find a counterexample to show that $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-field over Ω.

Exercise 3.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B, C \in \mathcal{F}$. Show that

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

   This is called the *inclusion-exclusion principle.*
2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
   (a) Show that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ for all $A, B \in \mathcal{F}$. This is called *subadditivity.*
   (b) Show that for any sequence $A_1, A_2, \ldots$ of events in $\mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

   This is called *countable subadditivity.*

Exercise 3.2

1. Let A and B be events with probabilities $\mathbb{P}(A) = 3/4$ and $\mathbb{P}(B) = 1/3$.
   (a) Show that $\frac{1}{12} \leq \mathbb{P}(A \cap B) \leq \frac{1}{3}$, and construct examples to show that both extremes are possible.
   (b) Find corresponding bounds for $\mathbb{P}(A \cup B)$.
2. A roulette wheel consists of 37 slots of equal size. The slots are numbered from 0 to 36, with odd-numbered slots coloured red, even-numbered slots coloured black, and the slot labelled 0 coloured green. The wheel is spun in one direction and a ball is rolled in the opposite direction along a track running around the circumference of the wheel. The ball eventually falls on to the wheel and into one of the 37 slots. A player bets on the event that the ball stops in a red slot, and another player bets on the event that the ball stops in a black slot.

(a) Define a suitable sample space Ω for this random experiment, and identify the events of interest.

(b) Find the smallest field $\mathcal{F}$ over Ω that contains the events of interest.

(c) Define a suitable probability measure $(\Omega, \mathcal{F})$ to represent the game.

Exercise 3.3

1. A biased coin has probability p of showing heads. The coin is tossed repeatedly until a head occurs. Describe a suitable probability space for this experiment.
2. A fair coin is tossed repeatedly.
   (a) Show that a head eventually occurs with probability one.
   (b) Show that a sequence of 10 consecutive tails eventually occurs with probability one.
   (c) Show that any finite sequence of heads and tails eventually occurs with probability one.

Exercise 4.1 [Revision]

1. Let Ω be a sample space, and let $A_1, A_2, \ldots$ be a partition of Ω with the property that $\mathbb{P}(A_i) > 0$ for all i.
   (a) Show that $\mathbb{P}(B) = \displaystyle\sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.
   (b) Show that $\mathbb{P}(A_i|B) = \dfrac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$.

Exercise 4.2

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and consider the function $\mathbb{Q} : \mathcal{F} \to [0, 1]$ defined by $\mathbb{Q}(A) = \mathbb{P}(A|B)$.
   (a) Show that $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space.
   (b) If $C \in \mathcal{F}$ and $\mathbb{Q}(C) > 0$, show that $\mathbb{Q}(A|C) = \mathbb{P}(A|B \cap C)$.
2. A random number N of dice are rolled. Let A_k be the event that $N = k$, and suppose that $\mathbb{P}(A_k) = 2^{-k}$ for $k \in \{1, 2, \ldots\}$ (and zero otherwise). Let S be the sum of the scores shown on the dice. Find the probability that:
   (a) $N = 2$ given that $S = 4$,
   (b) $S = 4$ given that N is even,
   (c) $N = 2$ given that $S = 4$ and the first die shows 1,
   (d) the largest number shown by any dice is r (where S is unknown).
3. Let $\Omega = \{1, 2, \ldots, p\}$ where p is a prime number. Let $\mathcal{F}$ be the power set of Ω, and let $\mathbb{P} : \mathcal{F} \to [0, 1]$ be the probability measure on $(\Omega, \mathcal{F})$ defined by $\mathbb{P}(A) = |A|/p$, where $|A|$ denotes the cardinality of A. Show that if A and B are independent events, then at least one of A and B is either $\emptyset$ or Ω.

Exercise 5.1

1. Let Ω be the sample space of some random experiment, and let $\mathcal{F}$ be a σ-field over Ω.
   (a) For any $A \in \mathcal{F}$, show that the function $X : \Omega \to \mathbb{R}$, defined by
   $$X(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$
   is a random variable on $(\Omega, \mathcal{F})$.
   (b) Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be a partition of Ω and let $a_1, a_2, \ldots, a_n \in \mathbb{R}$. Show that the function $X : \Omega \to \mathbb{R}$, defined by
   $$X(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega) \qquad \text{where} \qquad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$
   is a random variable on $(\Omega, \mathcal{F})$.

Exercise 6.1

1. Let F and G be CDFs, and let $0 < \lambda < 1$ be a constant. Show that $H = \lambda F + (1-\lambda)G$ is also a CDF.

2. Let X_1 and X_2 be the numbers observed in two independent rolls of a fair die. Find the PMF of each of the following random variables:
   (a) $Y = 7 - X_1$,
   (b) $U = \max(X_1, X_2)$,
   (c) $V = X_1 - X_2$.
   (d) $W = |X_1 - X_2|$.

3. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^2 & 1 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$
   (a) Find the value of the constant c, and sketch the PDF of X.
   (b) Find the value of $P(X > 3/2)$.
   (c) Find the CDF of X.

4. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^{-d} & \text{for } x > 1, \\ 0 & \text{otherwise.} \end{cases}$
   (a) Find the range of values of d for which $f(x)$ is a probability density function.
   (b) If $f(x)$ is a density function, find the value of c, and the corresponding CDF.

5. Let $f(x) = \dfrac{ce^x}{(1+e^x)^2}$ be a PDF, where c is a constant. Find the value of c, and the corresponding CDF.

6. Let $X_1, X_2, \ldots$ be independent and identically distributed observations, and let F denote their common CDF. If F is unknown, describe and justify a way of estimating F, based on the observations. [Hint: consider the indicator variables of the events $\{X_j \leq x\}$.]

Exercise 7.1

1. Let X be a discrete random variable, with PMF $f_X(-2) = 1/3$, $f_X(0) = 1/3$, $f_X(2) = 1/3$, and zero otherwise. Find the distribution of $Y = X + 3$.

2. Let $X \sim \text{Binomial}(n, p)$ and define $g(x) = n - x$. Show that $g(X) \sim \text{Binomial}(n, 1-p)$.

3. Let X be a random variable, and let F_X denote its CDF. Find the CDF of $Y = X^2$ in terms of F_X.

4. Let X be a random variable with the following CDF:

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{x^3} & \text{for } x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the CDF of the random variable $Y = 1/X$, and describe how a pseudo-random sample from the distribution of Y can be obtained using an algorithm that generates uniformly distributed pseudo-random numbers in the range $[0, 1]$.

Exercise 8.1

1. Let $X \sim \text{Uniform}(-1, 1)$. Find the CDF and PDF of X^2.

2. Let X have exponential distribution with rate parameter $\lambda > 0$. The PDF of X is

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the PDFs of $Y = X^2$ and $Z = e^X$.

3. Let $X \sim \text{Pareto}(1, 2)$. Find the PDF of $Y = 1/X$.

4. A continuous random variable U has PDF

$$f(u) = \begin{cases} 12u^2(1-u) & \text{for} \quad 0 < u < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = (1-U)^2$.

5. The continuous random variable U has PDF

$$f_U(u) = \begin{cases} 1+u & -1 < u \leq 0, \\ 1-u & 0 < u \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = U^2$. (Note that the transformation is not injective over $\text{supp}(f_U)$, so you should first compute the CDF of V, then derive its PDF by differentiation.)

6. Let X have exponential distribution with scale parameter $\theta > 0$. This has PDF

$$f(x) = \begin{cases} \frac{1}{\theta}\exp(-x/\theta) & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $Y = X^{1/\gamma}$ where $\gamma > 0$.

7. Suppose that X has the *Beta Type I* distribution, with parameters $\alpha, \beta > 0$. This has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $B(a,b) = \displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ is the so-called *beta function*. Show that the random variable $Y = \dfrac{X}{1-X}$ has the *Beta Type II* distribution, which has PDF

$$f_Y(y) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} \dfrac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 10.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $0 \leq X_1 \leq X_2 \leq \ldots$ be an increasing sequence of non-negative random variables over $(\Omega, \mathcal{F})$ such that $X_n(\omega) \uparrow X(\omega)$ ans $n \to \infty$ for all $\omega \in \Omega$. Show that X is a random variable on $(\Omega, \mathcal{F})$.

2. Let X be an integrable random variable. Show that $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

3. Let X and Y be integrable random variables. Show that $aX + bY$ is integrable.

Exercise 11.1

1. Let X be the score on a fair die, and let $g(x) = 3x - x^2$. Find the expected value and variance of the random variable $Y = g(X)$.

2. A long line of athletes $k = 0, 1, 2, \ldots$ make throws of a javelin to distances $X_0, X_1, X_2, \ldots$ respectively. The distances are independent and identically distributed random variables, and the probability that any two throws are exactly the same distance is equal to zero. Let Y be the index of the first athlete in the sequence who throws further than distance X_0. Show that the expected value of Y is infinite.

3. Consider the following game. A random number X is chosen uniformly from $[0, 1]$, then a sequence $Y_1, Y_2, \ldots$ of random numbers are chosen independently and uniformly from $[0, 1]$. Let Y_n be the first number in the sequence for which $Y_n > X$. When this occurs, the game ends and the player is paid $(n-1)$ pounds. Show that the expected win is infinite.

4. Let X be a discrete random variable with PMF

$$f(k) = \begin{cases} \dfrac{3}{\pi^2 k^2} & \text{if } \ k \in \{\pm 1, \pm 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\mathbb{E}(X)$ is undefined.

5. Let X be a continuous random variable having the Cauchy distribution, defined by the PDF

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

Show that $\mathbb{E}(X)$ is undefined.

6. A coin is tossed until the first time a head is observed. If this occurs on the nth toss and n is odd, you win $2^n/n$ pounds, but if n is even then you lose $2^n/n$ pounds. Show that the expected win is undefined.

7. Let X be a continuous random variable with uniform density on the interval $[-1,1]$,

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1,+1] \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

8. Let X be a random variable with the following CDF:

$$F(x) = \begin{cases} 0 & \text{for } \ x \leq 1 \\ 1 - 1/x^2 & \text{for } \ x \geq 1 \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

9. Let X be a continuous random variable with the following PDF:

$$f(x) = \begin{cases} 1 - |x| & \text{if } x \in [-1,1] \\ 0 & \text{otherwise.} \end{cases}$$

Find the range of integer values $\alpha \in \mathbb{Z}$ for which $\mathbb{E}(X^\alpha)$ exists.

Exercise 12.1

1. Let $X \sim \text{Uniform}[0,20]$ be a continuous random variable.
   (1) Use Chebyshev's inequality to find an upper bound on the probability $\mathbb{P}(|X-10| \geq z)$.
   (2) Find the range of z for which Chebyshev's inequality gives a non-trivial bound.
   (3) Find the value of z for which $\mathbb{P}(|X-10| \geq z) \leq 3/4$.

2. Let X be a discrete random variable, taking values in the range $\{1, 2, \ldots, n\}$, and suppose that $\mathbb{E}(X) = \text{Var}(X) = 1$. Show that $\mathbb{P}(X \geq k+1) \leq k^2$ for any integer k.

3. Let $k \in \mathbb{N}$. Show that Markov's inequality is tight (i.e. cannot be improved) by finding a non-negative random variable X such that
$$\mathbb{P}\big[X \geq k\mathbb{E}(X)\big] = \frac{1}{k}.$$

4. What does the Chebyshev inequality tell us about the probability that the value taken by a random variable deviates from its expected value by six or more standard deviations?

5. Let S_n be the number of successes in n Bernoulli trials with probability p of success on each trial. Use Chebyshev's Inequality to show that, for any $\epsilon > 0$, the upper bound

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}$$

is valid for any p.

6. Let $X \sim N(0,1)$.
   (1) Use Chebyshev's Inequality to find upper bounds for the probabilities $\mathbb{P}(|X| \geq 1)$, $\mathbb{P}(|X| \geq 2)$ and $\mathbb{P}(|X| \geq 3)$.
   (2) Use statistical tables to find the area under the standard normal curve over the intervals $[-1,1]$, $[-2,2]$ and $[-3,3]$.
   (3) Compare the bounds computed in part (a) with the exact values found in part (b). How good is the Chebyshev inequality in this case?

7. Let X be a random variable with mean $\mu \neq 0$ and variance σ^2, and define the *relative deviation* of X from its mean by $D = \left|\frac{X-\mu}{\mu}\right|$. Show that
$$\mathbb{P}(D \geq a) \leq \left(\frac{\sigma}{\mu a}\right)^2.$$

Exercise 13.1

1. Let $X \sim \text{Binomial}(m,p)$ and $Y \sim \text{Binomial}(n,p)$. Show that $X+Y \sim \text{Binomial}(m+n,p)$,

2. Show that a discrete distribution on the non-negative integers is uniquely determined by its PGF, in the sense that if two such random variables X and Y have PGFs $G_X(t)$ and $G_Y(t)$ respectively, then $G_X(t) = G_Y(t)$ if and only if $\mathbb{P}(X=k) = \mathbb{P}(Y=k)$ for all $k = 0,1,2,\ldots$.

3. The PGF of a random variable is given by $G(t) = 1/(2-t)$. What is its PMF?

4. Let $X \sim \text{Binomial}(n,p)$. Using the PGF of X, show that
$$\mathbb{E}\left(\frac{1}{1+X}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

Exercise 14.1

1. Let X be a discrete random variable, taking values in the set $\{-3,-2,-1,0,1,2,3\}$ with uniform probability, and let $M(t)$ denote the MGF of X.
   (1) Show that $M(t) = \frac{1}{7}(e^{-3t}+e^{-2t}+e^{-t}+1+e^{t}+e^{2t}+e^{3t})$.
   (2) Use $M(t)$ to compute the mean and variance of X.

2. A continuous random variable X has MGF given by $M(t) = \exp(t^2+3t)$. Find the distribution of X.

3. Let X be a discrete random variable with probability mass function
$$\mathbb{P}(X=k) = \begin{cases} q^k p & k = 0,1,2,\ldots \\ 0 & \text{otherwise,} \end{cases}$$
   where $0 < p < 1$ and $q = 1-p$.
   (1) Show that the MGF of X is given by $M(t) = \dfrac{p}{1-qe^t}$ for $t < -\log q$.
   (2) Find the PGF of X.
   (3) Use the PGF of X to find the PMF of $Y = X+1$.
   (4) Use $M(t)$ to find the mean and variance of X.

4. Let $M(t)$ denote the MGF of the normal distribution $\mathrm{N}(0,\sigma^2)$. By exanding $M(t)$ as a power series in t, show that the moments μ_k of the $\mathrm{N}(0,\sigma^2)$ distribution are zero if k is odd, and equal to
$$\mu_{2m} = \frac{\sigma^{2m}(2m)!}{2^m m!} \qquad \text{if } k = 2m \text{ is even.}$$

5. Let $X \sim \text{Exponential}(\theta)$ where θ is a scale parameter.

(1) Show that the MGF of X is $M(t) = \dfrac{1}{1-\theta t}$.

(2) By expanding this expression as a power series in t, find the first four non-central moments of X.

(3) Find the skewness γ_1 and the excess kurtosis γ_2 of X.

6. Let $X_1, X_2, \dots$ be independent and identically distributed random variables, with each $X_i \sim \mathrm{N}(\mu, \sigma^2)$.

   (1) Find the MGF of the random variable $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

   (2) Show that $\bar{X}$ has a normal distribution, and find its mean and variance.

7. Let $X_1 \sim \mathrm{Gamma}(k_1, \theta)$ and $X_2 \sim \mathrm{Gamma}(k_2, \theta)$ be independent random variables. Use the MGFs of X_1 and X_2 to find the distribution of the random variable $Y = X_1 + X_2$.

8. A coin has probability p of showing heads. The coin is tossed repeatedly until exactly k heads occur. Let N be the number of times the coin is tossed. Using the continuity theoram for characteristic functions, show that the distribution of the random variable $X = 2pN$ converges to a gamma distribution as $p \to 0$.

9. Let X and Y be independent and identically distributed random variables, with means equal to 0 and variances equal to 1. Let $\phi(t)$ denote their common characteristic function, and suppose that the random variables $X + Y$ and $X - Y$ are independent. Show that $\phi(2t) = \phi(t)^3\phi(-t)$, and hence deduce that X and Y must be independent standard normal variables.

Exercise 15.1

1. Let c be a constant, and let $X_1, X_2, \dots$ be a sequence of random variables with $\mathbb{E}(X_n) = c$ and $\mathrm{Var}(X_n) = 1/\sqrt{n}$ for each n. Show that the sequence converges to c in probability as $n \to \infty$.

2. A fair coin is tossed n times. Does the law of large numbers ensure that the observed number of heads will not deviate from $n/2$ by more than 100 with probability of at least 0.99, provided that n is sufficiently large?

Exercise 16.1

1. The continuous uniform distribution on (a, b) has the following PDF:

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

   Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

2. The exponential distribution with scale parameter $\theta > 0$ has the following PDF:

$$f(x) = \begin{cases} \dfrac{1}{\theta}e^{-x/\theta} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

3. Let $X \sim \mathrm{Binomial}(n_1, p_1)$ and $X_2 \sim \mathrm{Binomial}(n_2, p_2)$ be independent random variables.

   (1) Use the central limit theorem to find the approximate distribution of $Y = X_1 - X_2$ when n_1 and n_2 are both large.

   (2) Let $Y_1 = X_1/n_1$ and $Y_2 = X_2/n_2$. Show that $Y_1 - Y_2$ is approximately normally distributed with mean $p_1 - p_2$ and variance $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$ when n_1 and n_2 are both large.

   (3) Show that when n_1 and n_2 are both large,

$$\frac{(Y_1 - Y_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \qquad \text{approx.}$$

4. 5% of items produced by a factory production line are defective. Items are packed into boxes of 2000 items. As part of a quality control exercise, a box is chosen at random and found to contain 120 defective items. Use the central limit theorem to estimate the probability of finding at least this number of defective items when the production line is operating properly.

5. Use the central limit theorem to prove the law of large numbers.

6. We perform a sequence of independent Bernoulli trials, each with probability of success p, until a fixed number r of successes is obtained. The total number of failures Y (up to the rth succes) has the *negative binomial* distribution with parameters r and p, so the PMF of Y is

$$\mathbb{P}(Y = k) = \binom{k+r-1}{k}(1-p)^k p^r, \qquad k = 0, 1, 2, \ldots$$

Using the fact that Y can be written as the sum of r independent geometric random variables, show that this distribution can be approximated by a normal distribution when r is large.

Exercise 17.1

1. Let X be a Bernoulli random variable with parameter p.
   (a) Let $Y = 1 - X$. Find the joint PMF of X and Y.
   (b) Let $Y = 1 - X$ and $Z = XY$. Find the joint PMF of X and Z.

2. Let X and Y be two independent discrete random variables with the following PMFs:

x	1	2
$f_X(x)$	1/3	2/3

y	-1	0	1
$f_Y(y)$	1/4	1/2	1/4

   (a) Compute the joint PMF of X and Y.
   (b) Compute the joint PMF of the random variables $U = 1/X$ and $V = Y^2$.
   (c) Show that U and V are independent.

3. Two discrete random variables X and Y have the following joint PMF:

$$f_{X,Y}(x, y) = \begin{cases} c|x+y| & \text{for} \quad x, y \in \{-2, -1, 0, 1, 2\}, \\ 0 & \text{otherwise,} \end{cases}$$

   where c is a constant.
   (a) Show that $c = 1/40$.
   (b) Find $\mathbb{P}(X = 0, Y = -2)$.
   (c) Find $\mathbb{P}(X = 2)$.
   (d) Find $\mathbb{P}(|X - Y| \leq 1)$.

4. Two continuous random variables X and Y have the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2x & \text{if} \quad 0 \leq x, y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the conditional distribution of Y given that $X = x$.
   (b) Find $\mathbb{P}(Y \leq 0.5 | X = 0.5)$ and $\mathbb{P}(Y \leq 0.5 | X = 0.75)$.
   (c) Find the marginal distribution of Y and hence find $\mathbb{P}(Y \leq 0.5)$.

5. Two continuous random variables X and Y have the following joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} c(x^2 + y) & \text{when } -1 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 - x^2, \\ 0 & \text{otherwise.} \end{cases}$$

   where c is a constant.

(a) Show that $c = 5/4$.
(b) Find $\mathbb{P}(0 \leq X \leq 0.5)$.
(c) Find $\mathbb{P}(Y \leq X + 1)$.
(d) Find $\mathbb{P}(Y = X^2)$.

Exercise 18.1

1. Let X and Y be two random variables having the same distribution but which are not necessarily independent. Show that
$$\mathrm{Cov}(X + Y, X - Y) = 0$$
provided that their common distribution has finite mean and variance.

2. Consider a fair six-sided die whose faces show the numbers $-2, 0, 0, 1, 3, 4$. The die is independently rolled four times. Let X be the average of the four numbers that appear, and let Y be the product of these four numbers. Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(Y)$ and $\mathrm{Cov}(X, Y)$.

3. A fair die is rolled twice. Let U denote the number obtained on the first roll, let V denote the number obtained on the second roll, let $X = U + V$ denote their sum and let $Y = U - V$ denote their difference. Compute the mean and variance of X and Y, and compute $\mathbb{E}(XY)$. Check whether X and Y are uncorrelated. Check whether X and Y are independent.

Exercise 19.1

1. A fair coin is tossed three times. Let I_j be the indicator variable of the event that a head occurs on the jth toss. Compute the conditional expectation $E(Y|X)$ and verify the identity $E(E(Y|X)) = E(Y)$ in each of the following cases:
   (1) $X = \max\{I_1, I_2, I_3\}$ and $Y = \min\{I_1, I_2, I_3\}$,
   (2) $X = I_1 + I_2$ and $Y = I_2 + I_3$.

2. Let X and Y be continuous random variables with joint density function
$$f(x, y) = \begin{cases} c(x + y) & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
   (1) Show that $c = 1$.
   (2) Compute the conditional expectation $\mathbb{E}(Y|X)$.
   (3) Verify the identity $\mathbb{E}\big(\mathbb{E}(Y|X)\big) = \mathbb{E}(Y)$.

3. Let the joint density of random variables X and Y be
$$f(x, y) = \begin{cases} cxy & \text{for} \quad 0 \leq x, y \leq 1 \text{ where } \ x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$
   (1) Compute the normalization constant c.
   (2) Compute the conditional expectation $\mathbb{E}(Y|X)$.
   (3) Verify the identity $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$.

Exercise 20.1

1. Let X and Y have standard bivariate normal distribution, with joint PDF given by
$$f(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)}(x^2 - 2\rho xy + y^2)\right)$$
where ρ is a constant satisfying $-1 < \rho < 1$.
   (a) Check that $f(x, y)$ is indeed a joint PDF, by verifying that $f(x, y) \geq 0$ and $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy = 1$.

(b) Check that $\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy = \rho$.

(c) Show that if X and Y are uncorrelated, then they are independent.

2. Let X and Y have standard bivariate normal distribution. Find the conditional distribution of Y given $X = x$, and hence show that $\mathbb{E}(Y|X) = \rho X$.

3. Let X and Y have standard bivariate normal distribution. Show that X and $Z = \dfrac{Y-\rho X}{\sqrt{1-\rho^2}}$ are independent standard normal random variables.

4. Let X and Y have standard bivariate normal distribution, and let $Z = \max\{X,Y\}$. Show that $\mathbb{E}(Z) = \sqrt{(1-\rho)/\pi}$ and $\mathbb{E}(Z^2) = 1$.

5. Let $U, V \sim N(0,1)$. Show that the random variables $X = U + V$ and $Y = U - V$ are independent.

6. Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Show that the conditional distribution of Y given $X = x$ is

$$N\left(\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x-\mu_1), \sigma_2^2(1-\rho^2)\right).$$

7. (a) Let X and Y be jointly continuous random variables, and let $f_{X,Y}$ be their joint PDF. Show that the PDF of the random variable $X + Y$ can be written as

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_{X,Y}(x, t-x)\,dx = \int_{-\infty}^{\infty} f_{X,Y}(t-y, y)\,dy.$$

(b) Hence, or otherwise, show that if $U, V \sim N(0,1)$ are independent, then $U + V \sim N(0,2)$. (This is a special case of Theorem 20.5.)

Answers to Exercises

Exercise 2.1

1. Let $\mathcal{F}$ be a field over Ω. Show that
   (a) $\emptyset \in \mathcal{F}$,

   Answer: $\mathcal{F}$ is closed under complementation, and $\emptyset = \Omega^c$ where $\Omega \in \mathcal{F}$, so $\emptyset = \Omega^c$.

   (b) $\mathcal{F}$ is closed under set differences,

   Answer: Let $A, B \in \mathcal{F}$. Then $A \setminus B = A \cap B^c = (A^c \cup B)^c$ (De Morgan's laws). Hence $A \setminus B \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and pairwise unions.

   (c) $\mathcal{F}$ is closed under pairwise intersections,

   Answer: Let $A, B \in \mathcal{F}$. Then $A \cap B = (A^c \cup B^c)^c$ (De Morgan's laws). Hence $A \cap B \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and pairwise unions.

   (d) $\mathcal{F}$ is closed under finite unions,

   Answer: Proof by induction. Suppose that $\mathcal{F}$ is closed under unions of n sets (where $n \geq 2$). Let $A_1, A_2, \ldots, A_{n+1} \in \mathcal{F}$. By the inductive hypothesis, $\cup_{i=1}^{n} \in \mathcal{F}$, so $\cup_{i=1}^{n+1} A_i = \left[\cup_{i=1}^{n} A_i\right] \cup A_{n+1} \in \mathcal{F}$ because $\mathcal{F}$ is closed under pairwise unions.

   (e) $\mathcal{F}$ is closed under finite intersections.

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Then $\cap_{i=1}^{n} A_i = \left[\cup_{i=1}^{n} A_i^c\right]^c$ (De Morgan's laws). Hence $\cap_{i=1}^{n} A_i \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and finite unions.

2. Let $\mathcal{F}$ be a σ-field over Ω. Show that
   (a) $\mathcal{F}$ is closed under finite unions,

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Since $\mathcal{F}$ is closed under countable unions and $\emptyset \in \mathcal{F}$,
   $$\cup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \ldots \in \mathcal{F}.$$

   (b) $\mathcal{F}$ is closed under finite intersections.

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complementation and finite unions,
   $$\cap_{i=1}^{n} A_i = A_1 \cap \ldots \cap A_n = (A_1^c \cup \ldots \cup A_n^c)^c \in \mathcal{F}.$$

   (c) $\mathcal{F}$ is closed under countable intersections.

Answer: Let $A_1, A_2, \ldots \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complementation and countable unions,

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}.$$

Exercise 2.2

1. Let $\Omega = \{1,2,3,4,5,6\}$.
   (a) What is the smallest σ-field containing the event $A = \{1,2\}$?

Answer: A σ-field must contain $\emptyset$ and Ω, and be closed under complementation and countable unions.
The smallest σ-field containing $A = \{1,2\}$ is therefore

$$\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4,5,6\}, \Omega\}$$

   (b) What is the smallest σ-field containing the events $A = \{1,2\}$, $B = \{3,4\}$ and $C = \{5,6\}$?

Answer:
$$\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\}, \{3,4,5,6\}, \Omega\}$$

2. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-fields over Ω.
   (a) Show that $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ is a σ-field over Ω.

Answer: $\mathcal{H}$ is a σ-field because:
- $\emptyset \in \mathcal{F}$ and $\emptyset \in \mathcal{G}$ so $\emptyset \in \mathcal{H}$;
- if A belongs to both $\mathcal{F}$ and $\mathcal{G}$, then A^c belongs to both $\mathcal{F}$ and $\mathcal{G}$, so $\mathcal{H}$ is closed under complementation;
- if $A_1, A_2, \ldots$ all belong to both $\mathcal{F}$ and $\mathcal{G}$, then their union also lies in both $\mathcal{F}$ and $\mathcal{G}$, so $\mathcal{H}$ is closed under countable unions.

   (b) Find a counterexample to show that $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-field over Ω.

Answer: Let $\Omega = \{a,b,c\}$, $\mathcal{G} = \big\{\emptyset, \{a\}, \{b,c\}, \Omega\big\}$ and $\mathcal{G} = \big\{\emptyset, \{a,b\}, \{c\}, \Omega\big\}$. Then

$$\mathcal{H} = \mathcal{F} \cup \mathcal{G} = \big\{\emptyset, \{a\}, \{c\}, \{a,b\}, \{b,c\}, \Omega\big\}.$$

Hence $\{a\} \in \mathcal{H}$ and $\{c\} \in \mathcal{H}$, but $\{a,c\} \notin \mathcal{H}$ so $\mathcal{H}$ is not a σ-field.

Exercise 3.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B, C \in \mathcal{F}$. Show that

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

This is called the *inclusion-exclusion principle.*

Answer: Set union is an associative operator: $A \cup B \cup C = (A \cup B) \cup C$, so by the addition rule,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}\big((A \cup B) \cup C\big) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}\big((A \cup B) \cap C\big).\end{aligned}$$

Set intersection is distributive over set union: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, so by the addition rule,

$$\begin{aligned}\mathbb{P}\big((A \cup B) \cap C\big) &= \mathbb{P}\big((A \cap C) \cup (B \cap C)\big) \\ &= \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}\big((A \cap C) \cap (B \cap C)\big) \\ &= \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C).\end{aligned}$$

2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
   (a) Show that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ for all $A, B \in \mathcal{F}$. This is called *subadditivity*.

   Answer: TODO

   (b) Show that for any sequence $A_1, A_2, \ldots$ of events in $\mathcal{F}$,
   $$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$
   This is called *countable subadditivity*.

   Answer: TODO

Exercise 3.2

1. Let A and B be events with probabilities $\mathbb{P}(A) = 3/4$ and $\mathbb{P}(B) = 1/3$.
   (a) Show that $\frac{1}{12} \leq \mathbb{P}(A \cap B) \leq \frac{1}{3}$, and construct examples to show that both extremes are possible.

   Answer:
   - Lower bound: $\mathbb{P}(A \cup B) \leq 1$ so $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1 = \frac{1}{12}$.
   - Upper bound: $A \cap B \subseteq A$ and $A \cap B \subseteq B$, so $\mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\} = \frac{1}{3}$.

   Example: let $\Omega = \{1, 2, \ldots, 12\}$ with each outcome equally likely, and let $A = \{1, 2, \ldots, 9\}$.
   - Let $B = \{9, 10, 11, 12\}$. Then $\mathbb{P}(A \cap B) = \mathbb{P}(\{9\}) = \frac{1}{12}$.
   - Let $B = \{1, 2, 3, 4\}$. Then $\mathbb{P}(A \cap B) = \mathbb{P}(\{1, 2, 3, 4\}) = \frac{1}{3}$.

   (b) Find corresponding bounds for $\mathbb{P}(A \cup B)$.

   Answer:
   - Upper bound: $\mathbb{P}(A \cup B) \leq \min\{\mathbb{P}(A) + \mathbb{P}(B), 1\} = 1$.
   - Lower bound: $\mathbb{P}(A \cup B) \geq \max\{\mathbb{P}(A), \mathbb{P}(B)\} = 3/4$.

   These bounds are attained in the above example.

2. A roulette wheel consists of 37 slots of equal size. The slots are numbered from 0 to 36, with odd-numbered slots coloured red, even-numbered slots coloured black, and the slot labelled 0 coloured green. The wheel is spun in one direction and a ball is rolled in the opposite direction along a track running around the circumference of the wheel. The ball eventually falls on to the wheel and into one of the 37 slots. A player bets on the event that the ball stops in a red slot, and another player bets on the event that the ball stops in a black slot.
   (a) Define a suitable sample space Ω for this random experiment, and identify the events of interest.

Answer: A suitable sample space for the experiment is $\Omega = \{0, 1, 2, \ldots, 36\}$. The events of interest are $G = \{0\}$, $R = \{1, 3, \ldots, 35\}$ and $B = \{2, 4, \ldots, 36\}$.

(b) Find the smallest field $\mathcal{F}$ over Ω that contains the events of interest.

Answer: The smallest field of sets containing the events G, R and B is

$$\mathcal{F} = \{\emptyset, G, R, B, G \cup R, G \cup B, R \cup B, \Omega\}.$$

$\mathcal{F}$ is indeed a field of sets, because
- $\Omega \in \mathcal{F}$,
- $\mathcal{F}$ is closed under complementation,
  - $\emptyset^c = \Omega \in \mathcal{F}$ and $\Omega^c = \emptyset \in \mathcal{F}$,
  - $G^c = R \cup B \in \mathcal{F}$, $R^c = B \cup G \in \mathcal{F}$ and $B^c = R \cup G \in \mathcal{F}$,
  - $(G \cup R)^c = B \in \mathcal{F}$, $(G \cup B)^c = R \in \mathcal{F}$ and $(R \cup B)^c = G \in \mathcal{F}$
- $\mathcal{F}$ is closed under pairwise unions, for example
  - $R \cup \emptyset = R \in \mathcal{F}$ and $R \cup \Omega = \Omega \in \mathcal{F}$,
  - $R \cup B \in \mathcal{F}$ and $R \cup G \in \mathcal{F}$,
  - $R \cup (R \cup B) = R \cup B \in \mathcal{F}$,
  - $R \cup (R \cup G) = R \cup G \in \mathcal{F}$,
  - $R \cup (B \cup G) = \Omega \in \mathcal{F}$.

  and so on.

(c) Define a suitable probability measure $(\Omega, \mathcal{F})$ to represent the game.

Answer: A suitable probability measure over $(\Omega, \mathcal{F})$ is given by

$$\begin{aligned}
&\mathbb{P}(\emptyset) = 0, \\
&\mathbb{P}(R) = 18/37,\ \mathbb{P}(B) = 18/37,\ \mathbb{P}(G) = 1/37, \\
&\mathbb{P}(R \cup B) = 36/37,\ \mathbb{P}(R \cup G) = 19/37,\ \mathbb{P}(B \cup G) = 19/37, \\
&\mathbb{P}(\Omega) = 1.
\end{aligned}$$

This is indeed a probability measure, because
- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}$ is additive over $\mathcal{F}$; for example,
  - $\frac{36}{37} = \mathbb{P}(R \cup B) = \mathbb{P}(R) + \mathbb{P}(B) = \frac{18}{37} + \frac{18}{37} = \frac{36}{37}$,
  - $\frac{19}{37} = \mathbb{P}(R \cup G) = \mathbb{P}(R) + \mathbb{P}(G) = \frac{18}{37} + \frac{1}{37} = \frac{19}{37}$,
  - $\frac{19}{37} = \mathbb{P}(B \cup G) = \mathbb{P}(B) + \mathbb{P}(G) = \frac{18}{37} + \frac{1}{37} = \frac{19}{37}$,

  and so on.

Exercise 3.3

1. A biased coin has probability p of showing heads. The coin is tossed repeatedly until a head occurs. Describe a suitable probability space for this experiment.

Answer: The sample space is the set of all finite sequences of tails followed by a head, together with the infinite sequence of tails:

$$\Omega = \{T^n H : n \geq 0\} \cup \{T^\infty\}.$$

The σ-field can be taken to be the power set of Ω, and the probability measure can be defined on the elementary events by

$$\mathbb{P}(T^n H) = (1-p)^n p,$$
$$\mathbb{P}(T^\infty) = \lim_{n\to\infty}(1-p)^n = 0 \text{ if } p \neq 0.$$

2. A fair coin is tossed repeatedly.
   (a) Show that a head eventually occurs with probability one.

Answer: Let A_n be the event that no heads occur in the first n tosses, and let A be the event that no heads occur at all. Then $A_1, A_2, \ldots$ is a decreasing sequence ($A_{n+1} \subset A_n$), with $A = \cap_{i=1}^\infty A_n$. Hence by the continuity property of probability measures,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n) = \lim_{n\to\infty}\left(\frac{1}{2}\right)^n = 0,$$

or alternatively,

$$\mathbb{P}(\text{no heads}) = \lim_{n\to\infty}\mathbb{P}(\text{no heads in the first } n \text{ tosses}) = \lim_{n\to\infty}\left(\frac{1}{2}\right)^n = 0.$$

Thus we are certain of eventually observing a head.

   (b) Show that a sequence of 10 consecutive tails eventually occurs with probability one.

Answer: Let us think of the first $10n$ tosses as disjoint groups of consecutive outcomes, each group of length 10. The probability any one of the n groups consists of 10 consecutive tails is 2^{-10}, independently of the other groups. The event that one of the groups consists of 10 consecutive tails is a subset of the event that a sequence of 10 consecutive tails appears anywhere in the first $10n$ tosses. Hence, using the continuity of probability measures,

$$\begin{aligned}
\mathbb{P}(10T \text{ eventually appears}) &= \lim_{n\to\infty}\mathbb{P}(10T \text{ occurs somewhere in the first } 10n \text{ tosses}) \\
&\geq \lim_{n\to\infty}\mathbb{P}(10T \text{ occurs as one of the first } n \text{ groups of } 10) \\
&= 1 - \lim_{n\to\infty}\mathbb{P}(10T \text{ does not occur as one of the first } n \text{ groups of } 10) \\
&= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{2^{10}}\right)^n = 1.
\end{aligned}$$

Thus we are certain of eventually observing sequence of 10 consecutive tails.

   (c) Show that any finite sequence of heads and tails eventually occurs with probability one.

Answer: Let s be a fixed sequence of length k. As in the previous part, we think of the first kn tosses as n distinct groups of length k. The event that the one of these groups is exactly equal to s is a subset of the event that first kn tosses contains at least one instance of s. Hence

$$\begin{aligned}
\mathbb{P}(s \text{ eventually appears}) &= \lim_{n\to\infty}\mathbb{P}(s \text{ occurs somewhere in the first } kn \text{ tosses}) \\
&\geq \lim_{n\to\infty}\mathbb{P}(s \text{ occurs as one of the first } n \text{ groups of } k) \\
&= 1 - \lim_{n\to\infty}\mathbb{P}(s \text{ does not occur as one of the first } n \text{ groups of } k) \\
&= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{2^k}\right)^n = 1.
\end{aligned}$$

Thus we are certain of eventually observing the sequence s.

- In an infinite sequence of coin tosses, anything that can happen, does happen!

Exercise 4.1 [Revision]

1. Let Ω be a sample space, and let $A_1, A_2, \ldots$ be a partition of Ω with the property that $\mathbb{P}(A_i) > 0$ for all i.

   (a) Show that $\mathbb{P}(B) = \displaystyle\sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.

   Answer: Bookwork: this is the *partition theorem*, also known as the *law of total probability.*

   (b) Show that $\mathbb{P}(A_i|B) = \dfrac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$.

   Answer: Bookwork: this is *Bayes' formula.*

Exercise 4.2

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and consider the function $\mathbb{Q} : \mathcal{F} \to [0, 1]$ defined by $\mathbb{Q}(A) = \mathbb{P}(A|B)$.

   (a) Show that $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space.

   Answer:
   - $\mathbb{Q}(\Omega) = \mathbb{P}(\Omega|B) = 1$.
   - Let $\{A_i\}_{i=1}^{\infty}$ be a countable collection of pairwise disjoint events in $\mathcal{F}$.
     Since $\mathcal{F}$ is a σ-field, $\{A_i \cap B\}_{i=1}^{\infty}$ is also a countable collection of pairwise disjoint events in $\mathcal{F}$. Hence

   $$\mathbb{Q}(\cup_i A_i) = \frac{\mathbb{P}\big[(\cup_i A_i) \cap B\big]}{\mathbb{P}(B)} = \frac{\mathbb{P}\big[\cup_i (A_i \cap B)\big]}{\mathbb{P}(B)} = \frac{\sum_i \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_i \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_i \mathbb{Q}(A_i).$$

   (b) If $C \in \mathcal{F}$ and $\mathbb{Q}(C) > 0$, show that $\mathbb{Q}(A|C) = \mathbb{P}(A|B \cap C)$.

   Answer: Since $\mathbb{Q}$ is a probability measure,

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

   This shows that the order in which we impose the conditions B and C does not matter.

2. A random number N of dice are rolled. Let A_k be the event that $N = k$, and suppose that $\mathbb{P}(A_k) = 2^{-k}$ for $k \in \{1, 2, \ldots\}$ (and zero otherwise). Let S be the sum of the scores shown on the dice. Find the probability that:

   (a) $N = 2$ given that $S = 4$,

   Answer:

   $$\begin{aligned}\mathbb{P}(N = 2|S = 4) &= \frac{\mathbb{P}(\{N = 2\} \cap \{S = 4\})}{\mathbb{P}(\{S = 4\})} \\ &= \frac{\mathbb{P}(S = 4|N = 2)\mathbb{P}(N = 2)}{\sum_{k=1}^{n} \mathbb{P}(S = 4|N = k)\mathbb{P}(N = k)} \\ &= \frac{1/12 \times 1/4}{(1/6 \times 1/2) + (1/12 \times 1/4) + (3/6^3 \times 1/8) + (1/6^4 \times 1/16)} \\ &= \end{aligned}$$

(b) $S = 4$ given that N is even,

Answer:

$$\begin{aligned}\mathbb{P}(S=4|N \text{ even}) &= \frac{\mathbb{P}(S=4|N=2)\times(1/4)+\mathbb{P}(S=4|N=4)\times(1/16)}{\mathbb{P}(N \text{ even})} \\ &= \frac{(1/12\times 1/4)+(1/1296\times 1/16)}{1/4+1/16+1/64+\cdots} \\ &= \end{aligned}$$

(c) $N = 2$ given that $S = 4$ and the first die shows 1,

Answer: Let D be the score on the first die.

$$\begin{aligned}\mathbb{P}(N=2|S=2,D=1) &= \frac{\mathbb{P}(N=2,S=4,D=1}{\mathbb{P}(S=4,D=1)} \\ &= \frac{1/6\times 1/6\times 1/4}{(1/6\times 1/6\times 1/4)+(1/6\times 2/36\times 1/8)+(1/6^4\times 1/16)} \\ &= \end{aligned}$$

(d) the largest number shown by any dice is r (where S is unknown).

Answer: Let M be the maximum number shown on the dice. For $r \in \{1,2,3,4,5,6\}$,

$$\begin{aligned}\mathbb{P}(M\leq r) &= \sum_{k=1}^{\infty}\mathbb{P}(M\leq r|N=k)\frac{1}{2^k} \\ &= \sum_{k=1}^{\infty}\left(\frac{r}{6}\right)^k\frac{1}{2^k} \\ &= \frac{r}{12}\left(1-\frac{r}{12}\right)^{-1} \\ &= \frac{r}{12-r}.\end{aligned}$$

3. Let $\Omega = \{1,2,\dots,p\}$ where p is a prime number. Let $\mathcal{F}$ be the power set of Ω, and let $\mathbb{P}:\mathcal{F}\to[0,1]$ be the probability measure on $(\Omega,\mathcal{F})$ defined by $\mathbb{P}(A)=|A|/p$, where $|A|$ denotes the cardinality of A. Show that if A and B are independent events, then at least one of A and B is either $\emptyset$ or Ω.

Answer: Let A and B be independent events with $|A| = a$, $|B| = b$ and $|A\cap B| = c$.

- By independence, $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- This means that $(a/p)(b/p) = (c/p)$ and therefore $ab = pc$.
- If $ab \neq 0$, then p divides ab.
- Since p is prime, either p divides a, or p divides b (by the fundamental theorem of arithmetic).
- Hence $a = p$ or $b = p$ (or both).
- Thus follows that $A = \Omega$ or $B = \Omega$ (or both).

Exercise 5.1

1. Let Ω be the sample space of some random experiment, and let $\mathcal{F}$ be a σ-field over Ω.
   (a) For any $A \in \mathcal{F}$, show that the function $X:\Omega\to\mathbb{R}$, defined by

$$X(\omega) = \begin{cases}1 & \text{if } \omega\in A,\\ 0 & \text{if } \omega\notin A,\end{cases}$$

is a random variable on $(\Omega, \mathcal{F})$.

Answer: For any $B \in \mathcal{B}$,
- if $1 \in B$, then $\{\omega : X(\omega) \in B\} = A$, which is contained in $\mathcal{F}$;
- if $1 \notin B$, then $\{\omega : X(\omega) \in B\} = \emptyset$, which is also contained in $\mathcal{F}$.

(b) Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be a partition of Ω and let $a_1, a_2, \ldots, a_n \in \mathbb{R}$. Show that the function $X : \Omega \to \mathbb{R}$, defined by

$$X(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega) \qquad \text{where} \qquad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$

is a random variable on $(\Omega, \mathcal{F})$.

Answer: For any $B \in \mathcal{B}$,

$$\{\omega : X(\omega) \in B\} = \cup\{A_i : a_i \in B\} \in \mathcal{F},$$

because $\mathcal{F}$ is closed under finite unions.

Exercise 6.1

1. Let F and G be CDFs, and let $0 < \lambda < 1$ be a constant. Show that $H = \lambda F + (1-\lambda)G$ is also a CDF.

Answer: Let $H(x) = \lambda F(x) + (1-\lambda)G(x)$. It is easy to show that H has the following properties:
- if $x < y$ then $H(x) \leq H(y)$,
- $H(x) \to 0$ as $x \to -\infty$,
- $H(x) \to 1$ as $x \to +\infty$, and
- $H(x + \epsilon) \to H(x)$ as $\epsilon \downarrow 0$.

Thus H is a distribution function.

2. Let X_1 and X_2 be the numbers observed in two independent rolls of a fair die. Find the PMF of each of the following random variables:

(a) $Y = 7 - X_1$,

Answer: $P(Y = k) = 1/6$ for $k = 1, \ldots, 6$.

(b) $U = \max(X_1, X_2)$,

Answer: Let $U = \max\{X_1, X_2\}$. Then since $\{X_1 \leq k\}$ and $\{X_2 \leq k\}$ are independent events,

$$\begin{aligned} P(U \leq k) &= P(X_1 \leq k \text{ and } X_2 \leq k) \\ &= P(X_1 \leq k)P(X_2 \leq k) \\ &= (k/6) \cdot (k/6) = k^2/36 \end{aligned}$$

Thus,

$$P(U = k) = P(U \leq k) - P(U \leq k-1) = \frac{(k^2 - (k-1)^2)}{36} = \frac{(2k-1)}{36}$$

(c) $V = X_1 - X_2$.

Answer: The values of $V = X_1 - X_2$ at each point of the sample space $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$ are

		j=1	2	3	4	5	6
i	1	0	1	2	3	4	5
	2	−1	0	1	2	3	4
	3	−2	−1	0	1	2	3
	4	−3	−2	−1	0	1	2
	5	−4	−3	−2	−1	0	1
	6	−5	−4	−3	−2	−1	0

The required probabilities are obtained by counting the number of outcomes that give the same value of $V = X_1 - X_2$:

v	−5	−4	−3	−2	−1	0	1	2	3	4	5
$P(V=v)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

(d) $W = |X_1 - X_2|$.

Answer:

w	0	1	2	3	4	5
$P(W=w)$	6/36	10/36	8/36	6/36	4/36	2/36

3. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^2 & 1 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$

(a) Find the value of the constant c, and sketch the PDF of X.

Answer: The PDF must integrate to 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_1^2 cx^2\,dx = \left[\frac{cx^3}{3}\right]_1^2 = \frac{7c}{3} = 1$$

so $c = 3/7$. (The sketch is a quadratic curve between $x = 1$ and $x = 2$.)

(b) Find the value of $P(X > 3/2)$.

Answer:

$$P(X > 3/2) = \int_{3/2}^2 \frac{3x^2}{7}\,dx = \left[\frac{x^3}{7}\right]_{3/2}^2 = \frac{37}{56}$$

(c) Find the CDF of X.

Answer: For $1 \leq x \leq 2$,

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_1^x \frac{3x^2}{7}\,dx = \left[\frac{x^3}{7}\right]_1^x = \frac{x^3 - 1}{7}$$

so the CDF of X is

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{7}(x^3 - 1) & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

4. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^{-d} & \text{for } x > 1, \\ 0 & \text{otherwise.} \end{cases}$

   (a) Find the range of values of d for which $f(x)$ is a probability density function.

   > **Answer:** The function $f(x) = cx^{-d}$ is only integrable if $d > 1$, in which case
   > $$\int_{-\infty}^{\infty} f(x)\,dx = \int_{1}^{\infty} \frac{c}{x^d}\,dx = \left[\frac{-c}{(d-1)x^{d-1}}\right]_1^{\infty} = \frac{c}{d-1}$$

   (b) If $f(x)$ is a density function, find the value of c, and the corresponding CDF.

   > **Answer:** If $f(x)$ is a probability density function, we require that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, so we must have that $c = d-1$. The corresponding distribution function is
   > $$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{1}^{\infty} \frac{d-1}{u^d}\,du = \left[\frac{-1}{x^{d-1}}\right]_1^{x} = 1 - \frac{1}{x^{d-1}}$$
   > for $x > 1$, and zero otherwise.

5. Let $f(x) = \dfrac{ce^x}{(1+e^x)^2}$ be a PDF, where c is a constant. Find the value of c, and the corresponding CDF.

   > **Answer:** By inspection, $f(x) = F'(x)$ where $F(x) = \frac{ce^x}{1+e^x}$. Writing this as $F(x) = \frac{c}{e^{-x}+1}$ it is easy to see that $F(x) \to c$ as $x \to \infty$, so we must have that $c = 1$.

6. Let $X_1, X_2, \dots$ be independent and identically distributed observations, and let F denote their common CDF. If F is unknown, describe and justify a way of estimating F, based on the observations. [Hint: consider the indicator variables of the events $\{X_j \leq x\}$.]

   > **Answer:** Let X be a random variable with same CDF, and let $I_j(x)$ the indicator variable of the event $\{X_j \leq x\}$. Then
   > $$\mathbb{P}(X \leq x) \approx \frac{1}{n}\sum_{j=1}^{n} I_j(x).$$
   > The RHS yields the proportion of observations that are at most equal to x.

Exercise 7.1

1. Let X be a discrete random variable, with PMF $f_X(-2) = 1/3$, $f_X(0) = 1/3$, $f_X(2) = 1/3$, and zero otherwise. Find the distribution of $Y = X + 3$.

   > **Answer:** The function $g(x) = x + 3$ is injective, with $g^{-1}(y) = y - 3$, so
   > $$f_Y(1) = 1/3, \quad f_Y(3) = 1/3, \quad f_Y(5) = 1/3.$$
   > Note that $\text{supp}(f_X) = \{-2, 0, 2\}$, and $\text{supp}(f_Y) = \{g(x) : x \in \text{supp}(f_X)\} = \{1, 3, 5\}$.

2. Let $X \sim \text{Binomial}(n, p)$ and define $g(x) = n - x$. Show that $g(X) \sim \text{Binomial}(n, 1-p)$.

Answer: $g(x) = n - x$ is a decreasing function on $[0, n]$: its (unique) inverse is $g^{-1}(y) = n - y$. By Theoem 7.8, the PMF of $Y = g(X)$ is

$$f_Y(y) = f_X\left[g^{-1}(y)\right] = f_X(n-y) = \binom{n}{n-y} p^{n-y}(1-p)^{n-(n-y)} = \binom{n}{y}(1-p)^y(1-(1-p))^{n-y},$$

which is the PMF of the Binomial$(n, 1-p)$ distribution.

3. Let X be a random variable, and let F_X denote its CDF. Find the CDF of $Y = X^2$ in terms of F_X.

Answer:

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \sqrt{y}) - \mathbb{P}(X < -\sqrt{y}) \\ &= \begin{cases} F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

4. Let X be a random variable with the following CDF:

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{x^3} & \text{for } x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the CDF of the random variable $Y = 1/X$, and describe how a pseudo-random sample from the distribution of Y can be obtained using an algorithm that generates uniformly distributed pseudo-random numbers in the range $[0, 1]$.

Answer: Let $g(x) = 1/x$ denote the transformation.

- $\text{supp}(f_X) = [1, \infty] \Rightarrow \text{supp}(f_Y) = [0, 1]$.
- The inverse transformation: $g^{-1}(y) = 1/y$.

Because $g(x)$ is a decreasing function over $\text{supp}(f_X)$,

$$F_Y(y) = 1 - F_X\left[g^{-1}(y)\right] = 1 - F_X\left(\frac{1}{y}\right) = \begin{cases} 0 & y < 0 \\ y^3 & 0 \leq y \leq 1 \\ 1 & y > 1. \end{cases}$$

To find a pseudo-random sample from the distribution of Y, we use the fact that $F_Y(Y) \sim \text{Uniform}(0, 1)$. Let $u = F_Y(y)$. Then

$$y = F_Y^{-1}(u) = u^{1/3}.$$

The required sample is obtained by generating a pseudo-random sample $u_1, u_2, \ldots, u_n$ from the Uniform$(0, 1)$ distribution, then computing

$$y_i = u_i^{1/3} \quad \text{for } i = 1, 2, \ldots, n.$$

Exercise 8.1

1. Let $X \sim \text{Uniform}(-1, 1)$. Find the CDF and PDF of X^2.

Answer: The PDF of X is

$$f_X(x) = \begin{cases} 1/2 & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

For $x \in [-1, 1]$,

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-1}^{x} \frac{1}{2}\,dt = \left[\frac{t}{2}\right]_{-1}^{x} = \frac{1}{2}(x+1).$$

The CDF of X is:

$$F(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}(x+1) & -1 \leq x \leq 1, \\ 1 & x > 1. \end{cases}$$

Let $Y = X^2$. For $0 \leq y \leq 1$ we have

$$\begin{aligned} \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \sqrt{y}) - \mathbb{P}(X \leq -\sqrt{y}) \\ &= \sqrt{y}. \end{aligned}$$

Hence the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ \sqrt{y} & 0 \leq y \leq 1, \\ 1 & y > 1. \end{cases}$$

and the PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{2}y^{-1/2} & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

2. Let X have exponential distribution with rate parameter $\lambda > 0$. The PDF of X is

$$f(x) = \begin{cases} \lambda\exp(-\lambda x) & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDFs of $Y = X^2$ and $Z = e^X$.

Answer:

(1) The transformation $g(x) = x^2$ is monotonic increasing over $[0, \infty)$; its inverse function is

$$g^{-1}(y) = \sqrt{y}, \quad \text{which has first derivative} \quad \frac{d}{dy}g^{-1}(y) = \frac{1}{2\sqrt{y}}.$$

Since $\text{supp}(f_X) = [0, \infty)$ it follows immediately that $\text{supp}(f_Y) = [0, \infty)$.
For $y > 0$,

$$f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right| = \lambda\exp(\lambda\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| = \frac{\lambda}{2\sqrt{y}}\exp(-\lambda\sqrt{y}).$$

Hence the PDF of $Y = X^2$ is given by

$$f_Y(y) = \begin{cases} \dfrac{\lambda}{2\sqrt{y}}\exp(-\lambda\sqrt{y}) & y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(2) The transformation $g(x) = e^x$ is a monotonic increasing function over $[0, \infty)$; its inverse function is

$$g^{-1}(z) = \log y \quad \text{and} \quad \frac{d}{dy}g^{-1}(z) = \frac{1}{z}.$$

Since $\text{supp}(f_X) = [0, \infty)$ it follows immediately that $\text{supp}(f_Z) = [1, \infty)$.
For $z \geq 1$,

$$f_Z(z) = f_X\left[g^{-1}(z)\right] \left|\frac{d}{dz} g^{-1}(z)\right| = \lambda \exp(-\lambda \log z) \left|\frac{1}{z}\right| = \lambda z^{-(\lambda+1)}.$$

Hence the PDF of $Z = e^X$ is given by

$$f_Z(z) = \begin{cases} \lambda z^{-(\lambda+1)} & z \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

3. Let $X \sim \text{Pareto}(1, 2)$. Find the PDF of $Y = 1/X$.

Answer: $X \sim \text{Pareto}(1, 2)$ has PDF

$$f_X(x) = \begin{cases} \dfrac{2}{x^3} & x > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $g(x) = 1/x$.

- $g(x)$ is monotonic decreasing over $x > 1$; the inverse transformation is $g^{-1}(y) = 1/y$.
- $\text{supp}(f_Y) = \{x^{-1} : x > 1\} = (0, 1)$.

Hence the PDF of Y is given by

$$f_Y(y) = f_X\left[g^{-1}(y)\right] \left|\frac{d}{dy} g^{-1}(y)\right| = f_X\left(\frac{1}{y}\right) \left|-\frac{1}{y^2}\right| = \begin{cases} 2y & \text{for } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

4. A continuous random variable U has PDF

$$f(u) = \begin{cases} 12u^2(1-u) & \text{for} \quad 0 < u < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = (1-U)^2$.

Answer:

- The transformation $g(u) = (1-u)^2$ is monotonic decreasing over $[0, 1]$.
- The inverse transformation is $g^{-1}(v) = 1 - v^{1/2}$, for which $\dfrac{d}{dv} g^{-1}(v) = -\dfrac{1}{2v^{1/2}}$.
- Since $\text{supp}(f_U) = (0, 1)$ it follows that $\text{supp}(f_V) = (0, 1)$.

Hence for $0 < v < 1$ the PDF of V is

$$\begin{aligned} f_V(v) &= f_U\left[g^{-1}(v)\right] \left|\frac{d}{dv} g^{-1}(v)\right| \\ &= 12(1 - v^{1/2})^2 v^{1/2} \left|-\frac{1}{2v^{1/2}}\right| \\ &= 6(1 - v^{1/2})^2, \end{aligned}$$

and zero otherwise.

5. The continuous random variable U has PDF

$$f_U(u) = \begin{cases} 1 + u & -1 < u \leq 0, \\ 1 - u & 0 < u \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = U^2$. (Note that the transformation is not injective over $\text{supp}(f_U)$, so you should first compute the CDF of V, then derive its PDF by differentiation.)

Answer: Let $g(u) = u^2$. This is not injective over $\text{supp}(f_U) = (-1, 1)$, and does not therefore have a unique inverse over this interval. Instead we will compute the CDF of V, then obtain the PDF by differentiation.
For $0 < v < 1$,

$$\begin{aligned} F_V(v) = P(V \leq v) &= P(U^2 \leq v) \\ &= P(-\sqrt{v} \leq U \leq \sqrt{v}) \\ &= \int_{-\sqrt{v}}^{+\sqrt{v}} f_U(u)\, du \\ &= \int_{-\sqrt{v}}^{0} (1+u)\, du + \int_{0}^{+\sqrt{v}} (1-u)\, du \\ &= \left[u + \frac{u^2}{2}\right]_{-\sqrt{v}}^{0} + \left[u - \frac{u^2}{2}\right]_{0}^{\sqrt{v}} \\ &= \sqrt{v} - \frac{v}{2} + \sqrt{v} - \frac{v}{2} \\ &= 2\sqrt{v} - v. \end{aligned}$$

The CDF is therefore

$$F_V(u) = \begin{cases} 0 & v \leq 0, \\ 2\sqrt{v} - v & 0 < v < 1, \\ 1 & v \geq 1. \end{cases}$$

The PDF is then found by differentiation with respect to v:

$$f_V(u) = \begin{cases} v^{-1/2} - 1 & \text{for } 0 \leq v < 1, \\ 0 & \text{otherwise.} \end{cases}$$

6. Let X have exponential distribution with scale parameter $\theta > 0$. This has PDF

$$f(x) = \begin{cases} \frac{1}{\theta}\exp(-x/\theta) & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $Y = X^{1/\gamma}$ where $\gamma > 0$.

Answer: Let $g(x) = x^{1/\gamma}$.

- g a monotonic increasing function over $\text{supp}(f_X) = \{x : x > 0\}$, so its inverse exists:
- The inverse transformation is $g^{-1}(y) = y^\gamma$, for which $\dfrac{d}{dy}g^{-1}(y) = \gamma y^{\gamma-1}$.
- $\text{supp}(f_X) = \{x : x > 0\}$ means that $\text{supp}(f_Y) = \{y : y > 0\}$.

Since $f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\dfrac{d}{dy}g^{-1}(y)\right|$, we obtain

$$f_Y(y) = \begin{cases} (\gamma/\theta)y^{\gamma-1}\exp(-y^\gamma/\theta) & \text{for } \ y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is called the *Weibull* distribution (with scale parameter θ and shape parameter γ).

7. Suppose that X has the *Beta Type I* distribution, with parameters $\alpha, \beta > 0$. This has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ is the so-called *beta function*. Show that the random variable $Y = \dfrac{X}{1-X}$ has the *Beta Type II* distribution, which has PDF

$$f_Y(y) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} \dfrac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Answer: Let $g(x) = x/(1-x)$

- $g(x)$ is monotonic increasing on $\text{supp}(f_X) = [0,1]$.
- The inverse transformation is $g^{-1}(y) = \dfrac{y}{1+y}$, which has derivative $\dfrac{d}{dy}g^{-1}(y) = \dfrac{1}{(1+y)^2}$.
- Since $\text{supp}(f_X) = [0,1]$, we see that $\text{supp}(f_Y) = [0,\infty)$.

Thus for $y > 0$, the PDF of Y is

$$\begin{aligned} f_Y(y) &= f_X\left[g^{-1}(y)\right] \left|\frac{d}{dy}g^{-1}(y)\right| \\ &= \frac{1}{B(\alpha,\beta)}\left(\frac{y}{1+y}\right)^{\alpha-1}\left(\frac{1}{1+y}\right)^{\beta-1}\left|\frac{1}{(1+y)^2}\right| \\ &= \frac{1}{B(\alpha,\beta)}\frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}}, \end{aligned}$$

and zero otherwise.

Exercise 10.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $0 \leq X_1 \leq X_2 \leq \ldots$ be an increasing sequence of non-negative random variables over $(\Omega, \mathcal{F})$ such that $X_n(\omega) \uparrow X(\omega)$ ans $n \to \infty$ for all $\omega \in \Omega$. Show that X is a random variable on $(\Omega, \mathcal{F})$.

Answer: Let $x \in \mathbb{R}$. Since the X_n are random variables, we have (by definition) that $\{X_n \leq x\} \in \mathcal{F}$ for every $n \in \mathbb{N}$. Since $\mathcal{F}$ is closed under countable intersections,

$$\{X \leq x\} = \bigcap_{n=1}^{\infty}\{X_n \leq x\} \in \mathcal{F}$$

so X is a random variable.

2. Let X be an integrable random variable. Show that $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Answer: Since $|X| = X^+ + X^-$, by the triangle inequality

$$|\mathbb{E}(X)| = |\mathbb{E}(X^+) - \mathbb{E}(X^-)| \leq \mathbb{E}(X^+) + \mathbb{E}(X^-) = \mathbb{E}(|X|),$$

3. If $X \leq Y$ then $X^+ \leq Y^+$ and $X^- \geq Y^-$ so
$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \leq \mathbb{E}(Y^+) - \mathbb{E}(Y^-) = \mathbb{E}(Y),$$

4. Let X and Y be integrable random variables. Show that $aX + bY$ is integrable.

Answer: To show that $aX + bY$ is integrable, first we have by the triangle inequality that
$$|aX + bY| \leq |a||X| + |b||Y|.$$
By the linearity and monotonicity of expectation for non-negative random variables,
$$\mathbb{E}(|aX + bY|) \leq |a|\mathbb{E}(|X|) + |b|\mathbb{E}(|Y|)$$
and since $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y|) < \infty$, it follows that $\mathbb{E}(|aX + bY|) < \infty$, so $aX + bY$ is integrable.

Exercise 11.1

1. Let X be the score on a fair die, and let $g(x) = 3x - x^2$. Find the expected value and variance of the random variable $Y = g(X)$.

Answer: The expectation of $Y = 3X - X^2$ is determined by the distribution of X,
$$\begin{aligned}\mathbb{E}(Y) &= \sum_{x=1}^{6} y(x)f(x) = \sum_{x=1}^{6}(3x - x^2) \times \frac{1}{6} \\ &= \frac{1}{6}\left(3\sum_{x=1}^{6} x - \sum_{x=1}^{6} x^2\right) = \frac{-14}{3}\end{aligned}$$
and
$$\begin{aligned}\mathbb{E}(Y^2) &= \sum_{x=1}^{6} y^2(x)f(x) = \sum_{x=1}^{6}(3x - x^2)^2 \times \frac{1}{6} \\ &= \frac{1}{6}\left(9\sum_{x=1}^{6} x^2 - 6\sum_{x=1}^{6} x^3 + \sum_{x=1}^{6} x^4\right) = \frac{448}{6}\end{aligned}$$
Hence
$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{448}{6} - \frac{196}{9} = \frac{476}{9}$$

2. A long line of athletes $k = 0, 1, 2, \ldots$ make throws of a javelin to distances $X_0, X_1, X_2, \ldots$ respectively. The distances are independent and identically distributed random variables, and the probability that any two throws are exactly the same distance is equal to zero. Let Y be the index of the first athlete in the sequence who throws further than distance X_0. Show that the expected value of Y is infinite.

Answer: Y is a discrete random variable, taking values in the set $\{1, 2, \ldots\}$.

- The event $\{Y > k\}$ means that out of the first $k + 1$ throws, the initial throw was the furthest.

Because the distances $X_0, X_1, \ldots, X_k$ are identically distributed, it follows that
$$\mathbb{P}(Y > k) = \frac{1}{k+1}.$$

Thus,

$$\mathbb{P}(Y = k) = \mathbb{P}(Y > k - 1) - \mathbb{P}(Y > k) = \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}$$

so

$$\mathbb{E}(Y) = \sum_{n=0}^{\infty} k\mathbb{P}(Y = k) = \sum_{n=0}^{\infty} \frac{1}{k+1} = \sum_{n=1}^{\infty} \frac{1}{k} = \infty.$$

3. Consider the following game. A random number X is chosen uniformly from $[0, 1]$, then a sequence $Y_1, Y_2, \ldots$ of random numbers are chosen independently and uniformly from $[0, 1]$. Let Y_n be the first number in the sequence for which $Y_n > X$. When this occurs, the game ends and the player is paid $(n - 1)$ pounds. Show that the expected win is infinite.

Answer: Let Z be the amount won.

$$\begin{aligned}
\mathbb{P}(Z = k|X = x) &= \mathbb{P}(Y_1 \leq x, Y_2 \leq x, \ldots, Y_k \leq x, Y_{k+1} > x) \\
&= \mathbb{P}(Y_1 \leq x)\mathbb{P}(Y_2 \leq x) \ldots \mathbb{P}(Y_k \leq x)\mathbb{P}(Y_{k+1} > x) \qquad \text{(by independence)} \\
&= x^k(1 - x)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}(Z = k) &= \int_0^1 x^k(1 - x)\, dx \\
&= \left[\frac{1}{k+1}x^{k+1} - \frac{1}{k+2}x^{k+2}\right]_0^1 \\
&= \frac{1}{k+1} - \frac{1}{k+2} \\
&= \frac{1}{(k+1)(k+2)}
\end{aligned}$$

Thus,

$$\mathbb{E}(Z) = \sum_{k=0}^{\infty} k\left(\frac{1}{(k+1)(k+2)}\right) = \infty.$$

4. Let X be a discrete random variable with PMF

$$f(k) = \begin{cases} \dfrac{3}{\pi^2 k^2} & \text{if } k \in \{\pm 1, \pm 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\mathbb{E}(X)$ is undefined.

Answer: Let $X = X^+ - X^-$ where

$$X^+ = \max\{X, 0\} = \begin{cases} X & \text{if } X \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$X^- = \max\{-X, 0\} = \begin{cases} -X & \text{if } X < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned}\mathbb{E}(X^+) &= \sum_{k=1}^{\infty} k\left(\frac{3}{\pi^2 k^2}\right) &&= \frac{3}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty \\ \mathbb{E}(X^-) &= \sum_{k=-\infty}^{-1}(-k)\left(\frac{3}{\pi^2 k^2}\right) &&= \frac{3}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty\end{aligned}$$

so $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$ is undefined.

5. Let X be a continuous random variable having the Cauchy distribution, defined by the PDF

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

Show that $\mathbb{E}(X)$ is undefined.

Answer: The expectation of X is

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_+) - \mathbb{E}(X_-) \\ &= \int_0^{\infty} x f(x)\,dx - \int_{-\infty}^{0}(-x)f(x)\,dx \\ &= \int_0^{\infty}\frac{x}{\pi(1+x^2)}\,dx - \int_0^{\infty}\frac{x}{\pi(1+x^2)}\,dx\end{aligned}$$

If $x > 1$ then $x^2 > 1$ and therefore $2x^2 > 1 + x^2$, so

$$\frac{x}{1+x^2} > \frac{1}{2x} \qquad \text{for all } x > 1$$

Consequently,

$$\int_0^{\infty}\frac{x}{1+x^2}\,dx > \int_1^{\infty}\frac{x}{1+x^2}\,dx > \frac{1}{2}\int_1^{\infty}\frac{1}{x}\,dx = \infty$$

Thus X is not integrable:

$$\mathbb{E}(|X|) = \mathbb{E}(X_+) + \mathbb{E}(X_-) = 2\int_0^{\infty}\frac{x}{\pi(1+x^2)}\,dx = \infty$$

and $\mathbb{E}(X)$ is not defined.

6. A coin is tossed until the first time a head is observed. If this occurs on the nth toss and n is odd, you win $2^n/n$ pounds, but if n is even then you lose $2^n/n$ pounds. Show that the expected win is undefined.

Answer: Let X represent the amount won. $\mathbb{P}(\text{First head occurs on } n\text{th toss}) = 1/2^n$, so

$$\begin{aligned}\mathbb{E}(X) &= \sum_{n=1}^{\infty}\left(\frac{(-1)^{n+1}2^n}{n} \times \frac{1}{2^n}\right) \\ &= \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n} \\ &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots\end{aligned}$$

This is the alternating harmonic series, which is not absolutely convergent. Hence the expected win is undefinned.

Remark. It is known that the alternating harmonic series is convergent:

$$\sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \log 2$$

However, the series is not absolutely convergent, because

$$\sum_{n=0}^{\infty} \left| \frac{(-1)^{n+1}}{n} \right| = \sum_{n=0}^{\infty} \frac{1}{n} = \infty$$

The Riemann rearrangement theorem says that if a series is convergent but not absolutely convergent, then its limit depends on the order in which its terms are added. For example

$$\begin{aligned}
\log 2 &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \ldots \\
&= 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} + \ldots \\
&= \left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) + \ldots \\
&= 1 - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \ldots \\
&= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots\right) \\
&= \frac{1}{2}\log 2
\end{aligned}$$

which is absurd, since $\log 2 \neq 0$. The expectation $\mathbb{E}(X) = \sum_x g(x)f(x)$ of a discrete random variable cannot be sensibly defined unless the series $\sum_x g(x)f(x)$ is absolutely convergent.

7. Let X be a continuous random variable with uniform density on the interval $[-1, 1]$,

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1, +1] \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

Answer: Let F be the CDF of X, let $g : \mathbb{R} \to \mathbb{R}$, and recall the following:

- If $g(X)$ is non-negative random variable, its expectation with respect to F is

$$\mathbb{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

(For non-negative random variables, we can accept that its expectation is infinite.)

- If $g(X)$ is a signed random variable, its expectation with respect to F is only defined if

$$\int_{-\infty}^{\infty} |g(x)|f(x)\,dx < \infty.$$

If this condition holds, the expectation is given by

$$\mathbb{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx$$

where $g^+(x)$ and $g^-(x)$ are respectively is the positive and negative parts of $g(x)$:

$$g^+(x) = \begin{cases} g(x) & \text{if} \quad g(x) \geq 0, \\ 0 & \text{if} \quad g(x) < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad g(x) \geq 0, \\ -g(x) & \text{if} \quad g(x) < 0. \end{cases}$$

(1) $g(x) = x$. In this case, $g(x)$ is a signed function. Since

$$|g(x)| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0, \end{cases}$$

we see that the expectation exists:

$$\int_{-\infty}^{\infty} |g(x)|f(x)\,dx = \frac{1}{2}\int_{-1}^{0}(-x)\,dx + \frac{1}{2}\int_{0}^{1} x\,dx = \int_{0}^{1} x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} < \infty.$$

The positive and negative parts of g are

$$g^+(x) = \begin{cases} x & \text{if} \quad x \geq 0, \\ 0 & \text{if} \quad x < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad x \geq 0, \\ -x & \text{if} \quad x < 0. \end{cases}$$

Thus we have

$$\begin{aligned} \mathbb{E}(X) = \mathbb{E}\big[g(X)\big] &= \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx \\ &= \frac{1}{2}\int_0^1 x\,dx - \frac{1}{2}\int_{-1}^{0}(-x)\,dx \\ &= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 - \frac{1}{2}\left[\frac{-x^2}{2}\right]_{-1}^0 \\ &= \left(\frac{1}{4} - 0\right) - \left(0 + \frac{1}{4}\right) = 0. \end{aligned}$$

Note that, if we regard an integral as the "area between a curve and the x-axis", the positive part gives the area above the x-axis (which has a positive sign), and the negative part gives the area below the x-axis (which has a negative sign): the integral is zero because these two areas are of equal magnitude.

(2) $g(x) = x^2$. In this case, $g(x)$ is a non-negative function, so

$$\mathbb{E}(X^2) = \mathbb{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x)f(x)\,dx = \frac{1}{2}\int_{-1}^{1}\frac{x^2}{2}\,dx = \int_0^1 x^2\,dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

(3) $g(x) = x^3$. In this case, $g(x)$ is a signed function. Since

$$|g(x)| = \begin{cases} x^3 & \text{if } x \geq 0, \\ -x^3 & \text{if } x < 0, \end{cases}$$

we see that its expectation exists:

$$\int_{-\infty}^{\infty} |g(x)|f(x)\,dx = \frac{1}{2}\int_{-1}^{0}(-x^3)\,dx + \frac{1}{2}\int_{0}^{1} x^3\,dx = \int_{0}^{1} x^3\,dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4} < \infty.$$

The positive and negative parts of g are

$$g^+(x) = \begin{cases} x^3 & \text{if} \quad x^3 \geq 0, \\ 0 & \text{if} \quad x^3 < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad x^3 \geq 0, \\ -x^3 & \text{if} \quad x^3 < 0. \end{cases}$$

Since $x^3 \geq 0$ if and only if $x \geq 0$, these can be written as:

$$g^+(x) = \begin{cases} x^3 & \text{if} \quad x \geq 0, \\ 0 & \text{if} \quad x < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad x \geq 0, \\ -x^3 & \text{if} \quad x < 0. \end{cases}$$

Thus we have

$$\begin{aligned} \mathbb{E}(X^3) = \mathbb{E}\big[g(X)\big] &= \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx \\ &= \frac{1}{2}\int_0^1 x^3\,dx - \frac{1}{2}\int_{-1}^{0}(-x^3)\,dx \\ &= \frac{1}{2}\left[\frac{x^4}{4}\right]_0^1 - \frac{1}{2}\left[\frac{-x^4}{4}\right]_{-1}^0 \\ &= \left(\frac{1}{8} - 0\right) - \left(0 + \frac{1}{8}\right) = 0. \end{aligned}$$

(4) $g(x) = 1/x$. In this case, $g(x)$ is a signed function. Since

$$|g(x)| = \begin{cases} 1/x & \text{if } x \geq 0, \\ -1/x & \text{if } x < 0, \end{cases}$$

we see that its expectation does *not* exist:

$$\begin{aligned} \int_{-\infty}^{\infty} |g(x)| f(x)\,dx &= \frac{1}{2}\int_{-1}^{0} \frac{-1}{x}\,dx + \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx \\ &= \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx + \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx \\ &= \int_{0}^{1} \frac{1}{x}\,dx \\ &= \infty. \end{aligned}$$

Another way of seeing that the expectation is undefined is to consider the positive and negative parts of g:

$$g^{+}(x) = \begin{cases} 1/x & \text{if} \quad 1/x \geq 0, \\ 0 & \text{if} \quad 1/x < 0, \end{cases} \qquad \text{and} \qquad g^{-}(x) = \begin{cases} 0 & \text{if} \quad 1/x \geq 0, \\ -1/x & \text{if} \quad 1/x < 0. \end{cases}$$

Since $1/x \geq 0$ if and only if $x \geq 0$, these can be written as:

$$g^{+}(x) = \begin{cases} 1/x & \text{if} \quad x \geq 0, \\ 0 & \text{if} \quad x < 0, \end{cases} \qquad \text{and} \qquad g^{-}(x) = \begin{cases} 0 & \text{if} \quad x \geq 0, \\ -1/x & \text{if} \quad x < 0. \end{cases}$$

Thus we have

$$\begin{aligned} \mathbb{E}\left(\frac{1}{X}\right) = \mathbb{E}\big[g(X)\big] &= \int_{-\infty}^{\infty} g^{+}(x) f(x)\,dx - \int_{-\infty}^{\infty} g^{-}(x) f(x)\,dx \\ &= \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx - \frac{1}{2}\int_{-1}^{0} \frac{-1}{x}\,dx \\ &= \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx - \frac{1}{2}\int_{0}^{1} \frac{1}{x}\,dx \\ &= \infty - \infty. \end{aligned}$$

so $\mathbb{E}(1/X)$ is undefined.

(5) $g(x) = 1/x^2$. In this case, $g(x)$ is a non-negative function, so

$$\mathbb{E}\left(\frac{1}{X^2}\right) = \mathbb{E}_F(g) = \int_{-\infty}^{\infty} g(x) f(x)\,dx = \frac{1}{2}\int_{-1}^{1} \frac{1}{x^2}\,dx = \int_{0}^{1} \frac{1}{x^2}\,dx = \infty.$$

so $\mathbb{E}(1/X^2)$ is infinite (which is acceptable because $1/X^2$ is non-negative).

8. Let X be a random variable with the following CDF:

$$F(x) = \begin{cases} 0 & \text{for} \quad x \leq 1 \\ 1 - 1/x^2 & \text{for} \quad x \geq 1 \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

Answer:

$$f(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \geq 1, \\ 0 & \text{otherwisee.} \end{cases}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \int_1^\infty x\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^2}\,dx = 2\left[-\frac{1}{x}\right]_1^\infty = 2 \\
\mathbb{E}(X^2) &= \int_1^\infty x^2\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x}\,dx = \infty \\
\mathbb{E}\left(\frac{1}{X}\right) &= \int_1^\infty \frac{1}{x}\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^4}\,dx = 2\left[-\frac{1}{3x^3}\right]_1^\infty = \frac{2}{3} \\
\mathbb{E}\left(\frac{1}{X^2}\right) &= \int_1^\infty \frac{1}{x^2}\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^5}\,dx = 2\left[-\frac{1}{4x^4}\right]_1^\infty = \frac{1}{2}
\end{aligned}
$$

9. Let X be a continuous random variable with the following PDF:

$$
f(x) = \begin{cases} 1-|x| & \text{if } x \in [-1,1] \\ 0 & \text{otherwise.} \end{cases}
$$

Find the range of integer values $\alpha \in \mathbb{Z}$ for which $\mathbb{E}(X^\alpha)$ exists.

Answer: For $\alpha > 0$,

$$
\mathbb{E}(X^\alpha) = \int_{-1}^0 x^\alpha(1+x)\,dx + \int_0^1 x^\alpha(1-x)\,dx < \infty
$$

Let $\alpha < 0$. If α is even then X^α is non-negative, so

$$
\mathbb{E}(X^\alpha) = \mathbb{E}\big((X^+)^\alpha\big) = +\infty
$$

If α is odd,

$$
\mathbb{E}(X^\alpha) = \mathbb{E}\big((X^+)^\alpha\big) - \mathbb{E}\big((X^-)^\alpha\big) = \infty - \infty
$$

so in this case the moment $\mathbb{E}(X^\alpha)$ does not exist.

Exercise 12.1

1. Let $X \sim \text{Uniform}[0,20]$ be a continuous random variable.
   (1) Use Chebyshev's inequality to find an upper bound on the probability $\mathbb{P}(|X-10| \geq z)$.
   (2) Find the range of z for which Chebyshev's inequality gives a non-trivial bound.
   (3) Find the value of z for which $\mathbb{P}(|X-10| \geq z) \leq 3/4$.

Answer:

(1) By Chebyshev's inequality, $\mathbb{P}(|X-10| \geq z) \leq \dfrac{\text{Var}(X)}{z^2} = \dfrac{100}{3z^2}$.

(2) For a non trivial bound, we need that $\mathbb{P}(|X-10| \geq z) \leq \dfrac{100}{3z^2} < 1$ and hence $z^2 > \dfrac{100}{3}$.
We reject the case $z = -10/\sqrt{3}$ because $\mathbb{P}(|X-10| > -10/\sqrt{3}) = 1$.
Thus we conclude that $z > 10/\sqrt{3}$.

(3) This time we need that $\mathbb{P}(|X-10| \geq z) \leq \dfrac{100}{3z^2} < \dfrac{3}{4}$ and hence $z^2 > \dfrac{400}{9}$.
As before, we reject the case $z = -20/3$ because $\mathbb{P}(|X-10| > -20/3) = 1$.
Thus we conclude that $z > 20/3$.

2. Let X be a discrete random variable, taking values in the range $\{1, 2, \ldots, n\}$, and suppose that $\mathbb{E}(X) = \text{Var}(X) = 1$. Show that $\mathbb{P}(X \geq k + 1) \leq k^2$ for any integer k.

> **Answer:** Using the fact that $X - 1 \geq 0$,
>
> $$\mathbb{P}(X \geq k + 1) = \mathbb{P}(X - 1 \geq k) = \mathbb{P}(|X - 1| \geq k).$$
>
> By Chebyshev's inequality, with $\mathbb{E}(X) = 0$ and $\text{Var}(X) = 1$,
>
> $$\mathbb{P}(|X - 1| \geq k) \leq \frac{\text{Var}(X)}{k^2} = \frac{1}{k^2}$$

3. Let $k \in \mathbb{N}$. Show that Markov's inequality is tight (i.e. cannot be improved) by finding a non-negative random variable X such that

$$\mathbb{P}\big[X \geq k\mathbb{E}(X)\big] = \frac{1}{k}.$$

> **Answer:** Let X be a random variable taking values in the set $\{0, k\}$, such that $\mathbb{P}(X = k) = 1/k$ and $\mathbb{P}(X = 0) = 1 - 1/k$. Then $\mathbb{E}(X) = 1$ and $\mathbb{P}\big(X \geq k\mathbb{E}(X)\big) = \mathbb{P}\big(X \geq k\big) = \mathbb{P}\big(X = k\big) = 1/k$ as required.

4. What does the Chebyshev inequality tell us about the probability that the value taken by a random variable deviates from its expected value by six or more standard deviations?

> **Answer:** For any random variable X with finite variance σ^2,
>
> $$\mathbb{P}(|X - \mu| \geq 6\sigma) \leq \frac{\sigma^2}{(6\sigma)^2} = \frac{1}{36}.$$

5. Let S_n be the number of successes in n Bernoulli trials with probability p of success on each trial. Use Chebyshev's Inequality to show that, for any $\epsilon > 0$, the upper bound

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}$$

is valid for any p.

> **Answer:** For the Binomial(n, p) distribution, Chebyshev's inequality yields
>
> $$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}$$
>
> The result then follows by the fact that for any p,
>
> $$p(1-p) = \frac{1}{4} - \left(\frac{1}{4} - p + p^2\right) = \frac{1}{4} - \left(\frac{1}{2} - p\right)^2 \leq \frac{1}{4}$$

6. Let $X \sim N(0, 1)$.
   (1) Use Chebyshev's Inequality to find upper bounds for the probabilities $\mathbb{P}(|X| \geq 1)$, $\mathbb{P}(|X| \geq 2)$ and $\mathbb{P}(|X| \geq 3)$.
   (2) Use statistical tables to find the area under the standard normal curve over the intervals $[-1, 1]$, $[-2, 2]$ and $[-3, 3]$.
   (3) Compare the bounds computed in part (a) with the exact values found in part (b). How good is the Chebyshev inequality in this case?

Answer:

(1) $\mathbb{P}(|X| \geq 1) \leq 1$, $\mathbb{P}(|X| \geq 2) \leq 1/4$ and $\mathbb{P}(|X| \geq 3) \leq 1/9$.

(2) From tables, $\mathbb{P}(|X| \geq 1) = 0.3173$, $\mathbb{P}(|X| \geq 2) = 0.0455$ and $\mathbb{P}(|X| \geq 3) = 0.0027$.

(3) Chebyshev's inequality provides only crude bounds on the tail probabilities of the standard normal distribution.

7. Let X be a random variable with mean $\mu \neq 0$ and variance σ^2, and define the *relative deviation* of X from its mean by $D = \left|\frac{X-\mu}{\mu}\right|$. Show that

$$\mathbb{P}(D \geq a) \leq \left(\frac{\sigma}{\mu a}\right)^2.$$

Answer: By Chebyshev's inequality,

$$\mathbb{P}(D \geq a) = \mathbb{P}\left(\left|\frac{X-\mu}{\mu}\right| \geq a\right) = \mathbb{P}\big(|X-\mu| \geq |\mu|a\big) \leq \frac{\sigma^2}{\mu^2 a^2}$$

Exercise 13.1

1. Let $X \sim \text{Binomial}(m, p)$ and $Y \sim \text{Binomial}(n, p)$. Show that $X + Y \sim \text{Binomial}(m + n, p)$,

Answer: The PGFs of X and Y are

$$G_X(t) = (1 - p + pt)^m \quad \text{and} \quad G_Y(t) = (1 - p + pt)^n$$

Using the properties of PGFs,

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (1 - p + pt)^m(1 - p + pt)^n = (1 - p + pt)^{m+n},$$

which we recognise as the PGF of the $\text{Binomial}(m + n, p)$ distribution.

2. Show that a discrete distribution on the non-negative integers is uniquely determined by its PGF, in the sense that if two such random variables X and Y have PGFs $G_X(t)$ and $G_Y(t)$ respectively, then $G_X(t) = G_Y(t)$ if and only if $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ for all $k = 0, 1, 2, \ldots$.

Answer: The PGFs of X and Y are

$$G_X(t) = \sum_{k=1}^{\infty} \mathbb{P}(X = k)t^k \quad \text{and} \quad G_Y(t) = \sum_{k=1}^{\infty} \mathbb{P}(Y = k)t^k$$

Clearly, if $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ for all $k = 0, 1, 2, \ldots$, then $G_X(t) = G_Y(t)$. Conversely, $G_X(1) = 1$ implies that its power series expansion (about the origin) is unique, and likewise for G_Y. Thus if $G_X = G_Y$, their power series must have identical coefficients, so $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ for all $k = 0, 1, 2, \ldots$, as required

3. The PGF of a random variable is given by $G(t) = 1/(2 - t)$. What is its PMF?

Answer: To find the PMF, we need to express $G(t)$ as a power series:

$$G_X(t) = \frac{1}{2-t} = \frac{1}{2}\left(1-\frac{s}{2}\right)^{-1} = \frac{1}{2}\left(1+\frac{t}{2}+\left(\frac{t}{2}\right)^2+\dots\right) - \sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^{k+1} t^k$$

Thus the PMF of X

$$\mathbb{P}(X=k) = \left(\frac{1}{2}\right)^{k+1} \quad \text{for } k=0,1,2,\dots.$$

4. Let $X \sim \text{Binomial}(n,p)$. Using the PGF of X, show that

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

Answer: Let $G(t)$ be the PGF of X. Then $G(t) = \mathbb{E}(t^X) = (q+pt)^n$ where $q = 1-p$. Now

$$\int_0^1 t^x\,dt = \left[\frac{t^{1+x}}{1+x}\right]_0^1 = \frac{1}{1+x},$$

so

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \mathbb{E}\left(\int_0^1 t^X\,dt\right) = \int_0^1 \mathbb{E}(t^X)\,dt = \int_0^1 (q+pt)^n\,dt = \frac{1-q^{n+1}}{(n+1)p}$$

Exercise 14.1

1. Let X be a discrete random variable, taking values in the set $\{-3,-2,-1,0,1,2,3\}$ with uniform probability, and let $M(t)$ denote the MGF of X.
   (1) Show that $M(t) = \frac{1}{7}(e^{-3t}+e^{-2t}+e^{-t}+1+e^t+e^{2t}+e^{3t})$.
   (2) Use $M(t)$ to compute the mean and variance of X.

Answer:

(1) $M(t) = \mathbb{E}(e^{Xt}) = \sum_k e^{tk}p(k) = \frac{1}{7}(e^{-3t}+e^{-2t}+e^{-t}+1+e^t+e^{2t}+e^{3t})$

(2) Using the power series expansion $e^x = \displaystyle\sum_{k=0}^{\infty}\frac{x^k}{k!}$, the MGF of X can be written as

$$\begin{aligned} M(t) = \frac{1}{7}\Bigg[&\left(1-3t+\frac{9t^2}{2}+\dots\right)+\left(1-2t+\frac{4t^2}{2}+\dots\right)+\left(1-t+\frac{t^2}{2}+\dots\right) \\ &+1+\left(1+t+\frac{t^2}{2}+\dots\right)+\left(1+2t+\frac{4t^2}{2}+\dots\right)+\left(1+3t+\frac{9t^2}{2}+\dots\right)\Bigg] \\ = 1+\frac{4t^2}{2}+\dots \end{aligned}$$

- The mean is the coefficient of t, so $\mu = 0$.
- The second non-central moment is the coefficient of $\frac{1}{2}t^2$, so $\mu_2' = 4$.
- The variance is therefore $\sigma^2 = \mu_2' - \mu^2 = 4$.

A more direct method is to use the fact that $\mu = M_X'(0)$ and $\mu_2' = M_X''(0)$. Since

$$M'(t) = \frac{1}{7}(-3e^{-3t}-2e^{-2t}-e^{-t}+e^t+2e^{2t}+3e^{3t}),$$
$$M''(t) = \frac{1}{7}(9e^{-3t}+4e^{-2t}+e^{-t}+e^t+4e^{2t}+9e^{3t})$$

it follows that

$$M'(0) = \frac{1}{7}(-3-2-2+1+2+3) = 0,$$
$$M''(0) = \frac{1}{7}(9+4+1+1+4+9) = 4.$$

2. A continuous random variable X has MGF given by $M(t) = \exp(t^2 + 3t)$. Find the distribution of X.

Answer:
$$M_X(t) = \exp(t^2+3t) = \exp(3t + \frac{1}{2}2t^2)).$$
The MGF is of the form $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$, which is the MGF of a normal distribution. The mean μ is 3 and the variance σ^2 is 2, so $X \sim N(3,2)$. Note that under reasonable conditions, a random variable can be uniquely identified by its MGF: only normally distributed random variables have MGFs of the form $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

3. Let X be a discrete random variable with probability mass function

$$\mathbb{P}(X=k) = \begin{cases} q^k p & k = 0,1,2,\dots \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < p < 1$ and $q = 1-p$.

(1) Show that the MGF of X is given by $M(t) = \dfrac{p}{1-qe^t}$ for $t < -\log q$.
(2) Find the PGF of X.
(3) Use the PGF of X to find the PMF of $Y = X+1$.
(4) Use $M(t)$ to find the mean and variance of X.

Answer:

(1) The MGF of X is

$$M(t) = \mathbb{E}(e^{Xt}) = \sum_{k=0}^{\infty} e^{kt}q^k p = p\sum_{k=0}^{\infty}\left[qe^t\right]^k = \frac{p}{1-qe^t}$$

provided that $|qe^t| < 1$, or equivalently that $t < -\log q$.

(2) Since $M(t) = G(e^t)$, the PGF is

$$G(t) = M(\log t) = \frac{p}{1-qt}.$$

Note that $G(0) = 0$ and $G(1) = 1$.

(3) To derive the PMF of $Y = X+1$, we have

$$G_Y(t) = \mathbb{E}(t^Y) = \mathbb{E}(t^{X+1}) = \mathbb{E}(t^X.t) = t\mathbb{E}(t^X) = tG_(t)$$

so by part (i),

$$G_Y(t) = \frac{pt}{1-qt} = pt(1 + qt + q^2t^2 + \dots$$

Comparing the coefficients in this expression with those of $G_Y(t)$ expressed as a power series,

$$G_Y(t) = \sum_{k=0}^{\infty}\mathbb{P}(Y=k)t^k = \mathbb{P}(Y=0) + \mathbb{P}(Y=1)t + \mathbb{P}(Y=2)t^2 + \dots$$

we see that

$$\begin{aligned}
\mathbb{P}(Y=0) &= \text{constant term} &&= 0 \\
\mathbb{P}(Y=1) &= \text{coefficient of } t &&= p \\
\mathbb{P}(Y=2) &= \text{coefficient of } t^2 &&= p(1-p) \\
\mathbb{P}(Y=3) &= \text{coefficient of } t^3 &&= p(1-p)^2 \text{ etc.}
\end{aligned}$$

As we might expect, $\mathbb{P}(Y=k) = \mathbb{P}(X=k-1)$ for $k=0,1,\ldots$.

(4)

$$M'(t) = \frac{d}{dt}\left(\frac{p}{1-qe^t}\right) = \frac{pqe^t}{(1-qe^t)^2}$$

so

$$\mu = M'(0) = \frac{pq}{(1-q)^2} = \frac{q}{p} = \frac{1-p}{p}.$$

Similarly,

$$M''_X(t) = \frac{d}{dt}\left(\frac{pqe^t}{(1-qe^t)^2}\right) = \frac{pqe^t(1+qe^t)}{(1-qe^t)^3}$$

so

$$\mu'_2 = M''_X(0) = \frac{q(1+q)}{p^2} = \frac{(1-p)(2-p)}{p}$$

and hence

$$\mathrm{Var}(X) = \mu'_2 - \mu^2 = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

4. Let $M(t)$ denote the MGF of the normal distribution $\mathrm{N}(0,\sigma^2)$. By exanding $M(t)$ as a power series in t, show that the moments μ_k of the $\mathrm{N}(0,\sigma^2)$ distribution are zero if k is odd, and equal to

$$\mu_{2m} = \frac{\sigma^{2m}(2m)!}{2^m m!} \qquad \text{if } k=2m \text{ is even.}$$

Answer: The MGF of the $N(\mu,\sigma^2)$ distribution is $\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$, so

$$M(t) = \exp\left(\frac{\sigma^2 t^2}{2}\right).$$

Expand this as a power series in t:

$$M(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{\sigma^4 t^4}{4\times 2!} + \frac{\sigma^6 t^6}{8\times 3!} + \ldots + \frac{\sigma^{2m} t^{2m}}{2^m \times m} + \ldots$$

The kth moment μ_k is the coefficient of $t^k/k!$ in this expansion. In particular, the skewness $\gamma_1 = \mu_3/\sigma^3$ is zero, and the excess kurtosis is

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\sigma^4 4!}{4\times 2!}\cdot\frac{1}{\sigma^4} - 3 = 0.$$

5. Let $X \sim \text{Exponential}(\theta)$ where θ is a scale parameter.

(1) Show that the MGF of X is $M(t) = \dfrac{1}{1-\theta t}$.

(2) By expanding this expression as a power series in t, find the first four non-central moments of X.

(3) Find the skewness γ_1 and the excess kurtosis γ_2 of X.

Answer:

(1)

$$\begin{aligned}
M(t) = \mathbb{E}(e^{Xt}) &= \int e^{xt}\frac{1}{\theta}e^{-x/\theta}\,dx \\
&= \frac{1}{\theta}\int_0^\infty e^{-x(1-\theta t)/\theta}\,dx \\
&= -\frac{1}{1-\theta t}\int_0^\infty \frac{d}{dx}e^{-x(1-\theta t)/\theta}\,dx \\
&= -\frac{1}{1-\theta t}\left[e^{-x(1-\theta t)/\theta}\right]_0^\infty \\
&= \frac{1}{1-\theta t}.
\end{aligned}$$

(2) Using the binomial expansion,

$$M(t) = (1-\theta t)^{-1} = 1 + \theta t + \theta^2 t^2 + \theta^3 t^3 + \theta^4 t^4 + \dots$$

The non-central moment μ_k is the coefficient of $\frac{t^k}{k!}$ in this power series expansion, so $\mu = \theta$, $\mu_2 = 2\theta^2$, $\mu_3 = 6\theta^3$ and $\mu_4 = 24\theta^4$.

(3) The skewness and excess kurtosis are

$$\begin{aligned}
\gamma_1 &= \frac{\mu_3}{\sigma^3} = \frac{2\theta^3}{\theta^3} = 2, \quad \text{and} \\
\gamma_2 &= \frac{\mu_4}{\sigma^4} - 3 = \frac{9\theta^4}{\theta^4} - 3 = 6,
\end{aligned}$$

respectively.

6. Let $X_1, X_2, \dots$ be independent and identically distributed random variables, with each $X_i \sim \mathrm{N}(\mu, \sigma^2)$.

(1) Find the MGF of the random variable $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$.

(2) Show that $\bar{X}$ has a normal distribution, and find its mean and variance.

Answer: Let $M(t)$ denote the common MGF of the random variables X_i:

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Moment generating functions have the following properties:

- $M_{X+Y}(t) = M_X(t)M_Y(t)$
- If $Y = a + bX$ then $M_Y(t) = e^{at}M_X(bt)$.

In view of these, the MGF of $\bar{X}$ is

$$\begin{aligned}
M_{\bar{X}}(t) &= M_{\frac{1}{n}(X_1+X_2+\dots+X_n)}(t) \\
&= M_{X_1+X_2+\dots+X_n}\left(\frac{t}{n}\right) \\
&= \left[M\left(\frac{t}{n}\right)\right]^n = \exp\left[\frac{\mu t}{n} + \frac{\sigma^2 t^2}{2n^2}\right]^n \\
&= \exp\left[\mu t + \frac{1}{2}\left(\frac{\sigma^2}{n}\right)t^2\right]
\end{aligned}$$

which is the MGF of a normal distribution with mean μ and variance σ^2/n.

7. Let $X_1 \sim \text{Gamma}(k_1, \theta)$ and $X_2 \sim \text{Gamma}(k_2, \theta)$ be independent random variables. Use the MGFs of X_1 and X_2 to find the distribution of the random variable $Y = X_1 + X_2$.

Answer: The $MGFs$ of X_1 and X_2 are:

$$M_{X_1}(t) = \frac{1}{(1-\theta t)^{k_1}} \quad \text{and} \quad M_{X_2}(t) = \frac{1}{(1-\theta t)^{k_2}}.$$

Since X_1 and X_2 are independent,

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = \frac{1}{(1-\theta t)^{k_1+k_2}}$$

This is the MGF of a $\text{Gamma}(k_1 + k_2, \theta)$ random variable.

8. A coin has probability p of showing heads. The coin is tossed repeatedly until exactly k heads occur. Let N be the number of times the coin is tossed. Using the continuity theoram for characteristic functions, show that the distribution of the random variable $X = 2pN$ converges to a gamma distribution as $p \to 0$.

Answer: Let $N = Z_1 + Z_2 + \ldots + Z_k$ with $Z_i \sim \text{Geometric}(p)$ independent. Let $\phi_Z(t)$ denote the common characteristic function of the Z_i:

$$\phi_Z(t) = \mathbb{E}(e^{itZ}) = \frac{pe^{it}}{1-qe^{it}} \qquad \text{where } Z \sim \text{Geometric}(p) \text{ and } q = 1-p.$$

By the properties of characteristic functions, the characteristic function of N is

$$\phi_N(t) = \big[\phi_Z(t)\big]^k = \left[\frac{pe^{it}}{1-qe^{it}}\right]^k$$

and the characteristic function of $X = 2pN$ is

$$\phi_X(t) = \phi_N(2pt) = \left[\frac{pe^{2pit}}{1-(1-p)e^{2pit}}\right] = \left[\frac{1+2pit+o(1)}{1-2it+o(1)}\right] \to \frac{1}{1-2it} \qquad \text{as} \quad p \to 0,$$

where $o(1)$ respresnts a quantity that tends to zero as $p \to 0$. This is the characteristic function of the $\text{Gamma}(k, \frac{1}{2})$ distribution. The result then follows by the continuity theorem for characteristic functions.

9. Let X and Y be independent and identically distributed random variables, with means equal to 0 and variances equal to 1. Let $\phi(t)$ denote their common characteristic function, and suppose that the random variables $X+Y$ and $X-Y$ are independent. Show that $\phi(2t) = \phi(t)^3\phi(-t)$, and hence deduce that X and Y must be independent standard normal variables.

Answer: Let $U = X+Y$ and $V = X-Y$. Since U and V are independent, we have $\phi_{U+V}(t) = \phi_U\phi_V$, or equivalently $\phi_{2X} = \phi_{X+Y}\phi_{X-Y}$. Thus, since $\phi_{2X}(t) = \phi(2t)$,$\phi_{X+Y} = \phi(t)^2$ and $\phi X - Y(t) = \phi(t)\phi(-t)$, it follows that

$$\phi(2t) = \phi(t)^3\phi(-t)$$

To show that $X, Y \sim \text{N}(0,1)$, it is sufficient to show that $\phi(t) = e^{-\frac{1}{2}t^2}$ (by the inversion theorem). It can be shown that characteristic functions are symmetric: $\phi(t) = \phi(-t)$ for all t.
Thus we have $\phi(2t) = \phi(t)^4$, and hence

$$\phi(t) = \phi\left(\frac{1}{2}t\right)^4 = \phi\left(\frac{1}{4}t\right)^{16} = \ldots = \phi\left(\frac{1}{2^n}t\right)^{2^{2^n}} \qquad \text{for } n = 0, 1, 2, \ldots.$$

Hence,

$$\phi(t) = \left[1 - \frac{1}{2}\left(\frac{t}{2^n}\right)^2 + \dots\right]^{2^{2^n}} \to e^{-\frac{1}{2}t^2} \qquad \text{as } n \to \infty.$$

Exercise 15.1

1. Let c be a constant, and let $X_1, X_2, \dots$ be a sequence of random variables with $\mathbb{E}(X_n) = c$ and $\text{Var}(X_n) = 1/\sqrt{n}$ for each n. Show that the sequence converges to c in probability as $n \to \infty$.

Answer: Let $\epsilon > 0$. By Chebyshev's inequality,

$$\mathbb{P}\big(|X_n - c| \geq \epsilon\big) \leq \frac{\text{Var}(X_n)}{\epsilon^2} = \frac{1}{\epsilon^2\sqrt{n}}$$

for all $n \in \mathbb{N}$. Thus $\lim_{n\to\infty} \mathbb{P}\big(|X_n - c| \geq \epsilon\big) = 0$, so the sequence converges to c in probability.

2. A fair coin is tossed n times. Does the law of large numbers ensure that the observed number of heads will not deviate from $n/2$ by more than 100 with probability of at least 0.99, provided that n is sufficiently large?

Answer: Yes, because the indicator variable has finite mean and variance.

Exercise 16.1

1. The continuous uniform distribution on (a, b) has the following PDF:

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

Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

Answer: The mean is

$$\mu = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2},$$

and the second moment is

$$\mu_2 = \int_a^b \frac{x^2}{b-a}\,dx = \frac{a^2+ab+b^2}{3},$$

so the variance is

$$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(b-a)^2}{12}$$

By the central limit theorem, if X is a random variable with mean μ and variance σ^2, the distribution of the sample mean $\bar{X}$ of a random sample of n independent observations is approximately $N(\mu, \frac{\sigma^2}{n})$, the approximation being better for larger n. In this case, the approximate distribution of $\bar{X}$ is $N\left(\frac{a+b}{2}, \frac{(b-a)^2}{12n}\right)$.

2. The exponential distribution with scale parameter $\theta > 0$ has the following PDF:

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

Answer:

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\theta}\int_0^\infty x e^{-x/\theta}\,dx = \theta, \\ \mathbb{E}(X^2) &= \frac{1}{\theta}\int_0^\infty x^2 e^{-x/\theta}\,dx = 2\theta^2 \\ \mathrm{Var}(Y) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \theta^2. \end{aligned}$$

By the CLT, the distribution of $\bar{X}$ is approximately $\mathrm{N}(\mu, \frac{\sigma^2}{n})$, the approximation being better for larger n. In this case, the approximate distribution of $\bar{X}$ is $\mathrm{N}\big(\theta, \theta^2/n\big)$.

3. Let $X \sim \text{Binomial}(n_1, p_1)$ and $X_2 \sim \text{Binomial}(n_2, p_2)$ be independent random variables.

(1) Use the central limit theorem to find the approximate distribution of $Y = X_1 - X_2$ when n_1 and n_2 are both large.

(2) Let $Y_1 = X_1/n_1$ and $Y_2 = X_2/n_2$. Show that $Y_1 - Y_2$ is approximately normally distributed with mean $p_1 - p_2$ and variance $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$ when n_1 and n_2 are both large.

(3) Show that when n_1 and n_2 are both large,

$$\frac{(Y_1 - Y_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \qquad \text{approx.}$$

Answer:

(1) The mean and variance of X_1 are respectively $n_1 p_1$ and $n_1 p_1 q_1$ where $q_1 = 1 - p_1$. The mean and variance of X_2 are respectively $n_2 p_2$ and $n_2 p_2 q_2$ where $q_2 = 1 - p_2$. Since $Y = X_1 - X_2$ is a linear combination of random variables,

$$\mathbb{E}(Y) = \mathbb{E}(X_1) - \mathbb{E}(X_2) = n_1 p_1 - n_2 p_2$$

and since X_1 and X_2 are independent,

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = n_1 p_1 q_1 + n_2 p_2 q_2.$$

Because both X_1 and X_2 are the sums of Bernoulli random variables, the CLT applies, so the approximate distribution of Y is

$$Y \sim N(n_1 p_1 - n_2 p_2, n_1 p_1 q_1 + n_2 p_2 q_2)$$

(2) For large n_1 and n_2, by the CLT the distribution of X_1 is approximately $N(n_1 p_1, n_1 p_1(1-p_1))$ for n_1 large, and the distribution of X_2 is approximately $N(n_2 p_2, n_2 p_2(1-p_2))$ for n_2 large. Thus the distributions of Y_1 and Y_2 are approximately $N\left(p_1, \frac{p_1 q_1}{n_1}\right)$ and $N\left(p_2, \frac{p_2 q_2}{n_2}\right)$ respectively, and the distribution of $Y_1 - Y_2$ is therefore approximately $N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$ for large n_1 and n_2.

(3) The usual standardization for the normal distribution (subtract the mean and divide by the standard deviation) yields the result. This is used in devising approximate tests and confidence intervals for the difference of proportions.

4. 5% of items produced by a factory production line are defective. Items are packed into boxes of 2000 items. As part of a quality control exercise, a box is chosen at random and found to contain 120 defective items. Use the central limit theorem to estimate the probability of finding at least this number of defective items when the production line is operating properly.

Answer: Let X be the number of defective items in a box. Then $X \sim \text{Binomial}(n, p)$ with $n = 2000$ and $p = 0.05$. Since n is large, X has approximately normal distribution with mean equal to $np(1-p) = 100$, and variance equal to $npq = 95$. The standardized variable $Z = (X - 100)/\sqrt{95}$ has therefore approximately the standard normal distribution $N(0,1)$. Thus

$$\mathbb{P}(X \geq 120) = \mathbb{P}\left(Z \geq \frac{120 - 100}{\sqrt{95}}\right) = \mathbb{P}(Z \geq 2.0520) \approx 0.0202$$

where the probability $\mathbb{P}(Z \geq 2.0520) \approx 0.0202$ can be obtained from statistical tables.

5. Use the central limit theorem to prove the law of large numbers.

Answer: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, and define $S_n = \sum_{i=1}^n X_i$. To prove the (weak) law of large numbers, we need to show that

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \qquad \text{as} \quad n \to \infty$$

Now,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \geq \frac{n\epsilon}{\sigma\sqrt{n}}\right) = \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \geq \frac{\sqrt{n}\epsilon}{\sigma}\right)$$

By the central limit theorem, $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ is approximately distributed according to $N(0,1)$, so this probability is approximated by the area under the standard normal curve between $\dfrac{\sqrt{n}\epsilon}{\sigma}$ and infinity, which approaches zero as $n \to \infty$.

6. We perform a sequence of independent Bernoulli trials, each with probability of success p, until a fixed number r of successes is obtained. The total number of failures Y (up to the rth succes) has the *negative binomial* distribution with parameters r and p, so the PMF of Y is

$$\mathbb{P}(Y = k) = \binom{k + r - 1}{k}(1-p)^k p^r, \qquad k = 0, 1, 2, \ldots$$

Using the fact that Y can be written as the sum of r independent geometric random variables, show that this distribution can be approximated by a normal distribution when r is large.

Answer: If $Y \sim \text{NB}(r, p)$, we can write

$$Y = X_1 + X_2 + \ldots + X_r \qquad \text{where} \quad X_i \sim \text{Geometric}(p).$$

Let $X \sim \text{Geometric}(p)$. Since $\text{Var}(X) < \infty$, it follows by the central limit theorem that

$$\frac{Y - r\mathbb{E}(X)}{\sqrt{r\text{Var}(X)}} \to \text{N}(0,1) \quad \text{in distribution as } r \to \infty.$$

In fact, since $\mathbb{E}(X) = (1-p)/p$ and $\text{Var}(X) = (1-p)/p^2$, we see that Y can be approximated by the $\text{N}\left(\dfrac{r(1-p)}{p}, \dfrac{r(1-p)}{p^2}\right)$ distribution as $r \to \infty$.

Exercise 17.1

1. Let X be a Bernoulli random variable with parameter p.
   (a) Let $Y = 1 - X$. Find the joint PMF of X and Y.

Answer:

$$f_{X,Y}(x,y) = \begin{cases} p & \text{if } (x,y) = (1,0) \\ 1-p & \text{if } (x,y) = (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

   (b) Let $Y = 1 - X$ and $Z = XY$. Find the joint PMF of X and Z.

Answer:

$$f_{X,Z}(x,z) = \begin{cases} 1-p & \text{if } (x,z) = (0,0) \\ p & \text{if } (x,z) = (1,0) \\ 0 & \text{otherwise.} \end{cases}$$

2. Let X and Y be two independent discrete random variables with the following PMFs:

x	1	2
$f_X(x)$	1/3	2/3

y	-1	0	1
$f_Y(y)$	1/4	1/2	1/4

   (a) Compute the joint PMF of X and Y.

Answer: Since X and Y are independent we have that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, which yields the following joint distribution.

		y			
		-1	0	1	$f_X(x)$
x	1	1/12	1/6	1/12	1/3
	2	1/6	1/3	1/6	2/3
	$f_Y(y)$	1/4	1/2	1/4	1

   (b) Compute the joint PMF of the random variables $U = 1/X$ and $V = Y^2$.

Answer: Let $f_{U,V}$ denote the joint PMF of U and V. Clearly, U takes the values 0.5 and 1, while V takes the values 0 and 1. We compute (for example)

$$f_{U,V}(0.5,1) = f_{X,Y}(2,-1) + f_{X,Y}(2,1) = 1/6 + 1/6 = 1/3.$$

Thus we obtain joint PMF shown in the following table:

		v		
		0	1	$f_U(u)$
u	1/2	1/3	1/3	2/3
	2	1/6	1/6	1/3
	$f_V(v)$	1/2	1/2	1

   (c) Show that U and V are independent.

Answer: The marginal PMFs $f_U(u)$ and $f_V(v)$ are computed by summing the rows and columns of the joint PMF table. Thus we see that U and V are independent because $f_{U,V}(u,v) = f_U(u)f_V(v)$ for every pair of values (u,v).

3. Two discrete random variables X and Y have the following joint PMF:

$$f_{X,Y}(x,y) = \begin{cases} c|x+y| & \text{for} \quad x,y \in \{-2,-1,0,1,2\}, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant.

(a) Show that $c = 1/40$.

Answer: First we tabulate the values of $|x+y|$:

		y				
		-2	-1	0	1	2
	-2	4	3	2	1	0
	-1	3	2	1	0	1
x	0	2	1	0	1	2
	1	1	0	1	2	3
	2	0	1	2	3	4

Because the probabilities must sum to 1, it follows that $c = 1/40$.

(b) Find $\mathbb{P}(X = 0, Y = -2)$.

Answer: The joint PMF of X and Y, along with their marginal distributions, are shown in the following table,

		y					
		-2	-1	0	1	2	
	-2	4/40	3/40	2/40	1/40	0	10/40
	-1	3/40	2/40	1/40	0	1/40	7/40
x	0	2/40	1/40	0	1/40	2/40	6/40
	1	1/40	0	1/40	2/40	3/40	7/40
	2	0	1/40	2/40	3/40	4/40	10/40
		10/40	7/40	6/40	7/40	4/40	

Hence

$$\mathbb{P}(X = 0, Y = -2) = f_{X,Y}(0,-2) = 2/40 = 1/20.$$

(c) Find $\mathbb{P}(X = 2)$.

Answer: $\mathbb{P}(X = 2) = f_X(2) = 10/40 = 1/4$

(d) Find $\mathbb{P}(|X - Y| \leq 1)$.

Answer:

$$\begin{aligned}\mathbb{P}(|X-Y|\leq 1) &= \mathbb{P}(-1\leq X-Y\leq 1)\\ &= \mathbb{P}(X-Y=-1,0 \text{ or } 1)\\ &= \mathbb{P}\big((X=Y-1)\cup(X=Y)\cup(X=Y+1)\big)\\ &= 8/40+12/40+8/40\\ &= 7/10.\end{aligned}$$

4. Two continuous random variables X and Y have the joint PDF

$$f_{X,Y}(x,y)=\begin{cases}2x & \text{if} \quad 0\leq x,y\leq 1\\ 0 & \text{otherwise.}\end{cases}$$

(a) Find the conditional distribution of Y given that $X=x$.

Answer: The marginal PDF of X is

$$f_X(x)=\int_y f(x,y)\,dy=\int_0^1 2x\,dy=\big[2xy\big]_0^1=2x \quad \text{for } 0\leq x\leq 1.$$

The conditional PDF of Y given $X=x$ is defined as

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{2x}{2x}=1 \quad \text{for } 0\leq y\leq 1$$

Thus the conditional PDF of Y given $X=x$ is $f_{Y|X}(y|x)=1$ for $0\leq y\leq 1$, and because this is independent of x, the random variables X and Y are independent.
The marginal PDF of Y is

$$f_Y(y)=\int_0^1 2x\,dx=\big[x^2\big]_0^1=1 \quad \text{for } 0\leq y\leq 1,$$

and the conditional PDF of X given $Y=y$ is

$$f_{X|Y}(x|y)=2x \quad \text{for } 0\leq x\leq 1$$

(b) Find $\mathbb{P}(Y\leq 0.5|X=0.5)$ and $\mathbb{P}(Y\leq 0.5|X=0.75)$.

Answer: Because X and Y are independent ,

$$\mathbb{P}(Y\leq 0.5|X=0.5)=\mathbb{P}(Y\leq 0.5|X=0.75)=\int_0^{0.5} 1\,dy=0.5$$

In general, when X and Y are not independent, the probability that $Y\leq 0.5$ would depend on the value taken by X.

(c) Find the marginal distribution of Y and hence find $\mathbb{P}(Y\leq 0.5)$.

Answer:

$$\mathbb{P}(Y\leq 0.5)=\int_0^{0.5} 1\,dy=\big[y\big]_0^{0.5}=0.5$$

5. Two continuous random variables X and Y have the following joint PDF:

$$f_{X,Y}(x,y)=\begin{cases}c(x^2+y) & \text{when } -1\leq x\leq 1 \text{ and } 0\leq y\leq 1-x^2,\\ 0 & \text{otherwise.}\end{cases}$$

where c is a constant.

(a) Show that $c = 5/4$.

Answer: The joint PDF is non-zero over the region between the curve $y = 1 - x^2$ and the x-axis.

- For $x \in [-1, 1]$ fixed, we integrate y over the range $0 \leq y \leq 1 - x^2$.
- For $y \in [0, 1]$ fixed, we integrate x over the range $-\sqrt{1-y} \leq x \leq +\sqrt{1-y}$.

To find c, the joint PDF must integrate to 1.

$$\begin{aligned}
\int\int f(x,y)\,dx\,dy = 1 &\Rightarrow \int_{x=-1}^{1}\int_{y=0}^{1-x^2} c(x^2+y)\,dx\,dy = 1 \\
&\Rightarrow \int_{-1}^{1}\left[c\left(x^2y + \frac{y^2}{2}\right)\right]_0^{1-x^2} dx = 1 \\
&\Rightarrow \int_{-1}^{1} c\left(x^2(1-x^2) + \frac{(1-x^2)^2}{2}\right) dx = 1 \\
&\Rightarrow \int_{-1}^{1} c\left(x^2 - x^4 + \frac{1}{2} - x^2 + \frac{x^4}{2}\right) dx = 1 \\
&\Rightarrow \int_{-1}^{1} \frac{c}{2}(1-x^4)\,dx = 1 \\
&\Rightarrow \left[\frac{c}{2}\left(x - \frac{x^5}{5}\right)\right]_{-1}^{1} = 1 \\
&\Rightarrow \frac{c}{2}\left(1 - \frac{1}{5} + 1 - \frac{1}{5}\right) = 1
\end{aligned}$$

Thus $c = 5/4$.

(b) Find $\mathbb{P}(0 \leq X \leq 0.5)$.

Answer: To calculate $\mathbb{P}(0 \leq X \leq 1/2)$, we first compute the marginal PDF of X.

$$\begin{aligned}
f_X(x) &= \int_0^{1-x^2} \frac{5}{4}(x^2+y)\,dy \\
&= \left[\frac{5}{4}\left(x^2y + \frac{y^2}{2}\right)\right]_0^{1-x^2} \\
&= \frac{5}{8}(1-x^4) \quad \text{for } -1 \leq x \leq 1.
\end{aligned}$$

Then we integrate this over x in the range $[0, 1/2]$:

$$\begin{aligned}
\mathbb{P}(0 \leq X \leq 1/2) &= \int_0^{1/2} \frac{5}{8}(1-x^4)\,dx \\
&= \left[\frac{5}{8}\left(x - \frac{x^5}{5}\right)\right]_0^{1/2} = \frac{79}{256}
\end{aligned}$$

(c) Find $\mathbb{P}(Y \leq X + 1)$.

Answer: $Y \leq X + 1$ is satisfied for pairs (X, Y) in the region between the curves $y = 1 + x$ and the x-axis when $-1 \leq x \leq 0$, and between the curve $y = 1 - x^2$ and the x-axis when $0 \leq x \leq 1$.

- For fixed $x \in [-1, 0]$, we must integrate y over the range $0 \leq y \leq 1 + x$.
- For fixed $x \in [0, +1]$, we must integrate y over the range $0 \leq y \leq 1 - x^2$.

It helps to draw a plot of the curves $y = 1 + x$ and $y = 1 - x^2$

$$\begin{aligned}
\mathbb{P}(Y \leq X+1) &= \int_{-1}^{0}\int_{0}^{1+x} f(x,y)\,dy\,dx + \int_{0}^{1}\int_{0}^{1-x^2} f(x,y)\,dy\,dx \\
&= \int_{-1}^{0}\int_{0}^{1+x} \frac{5}{4}(x^2+y)\,dy\,dx + \int_{0}^{1}\int_{0}^{1-x^2} \frac{5}{4}(x^2+y)\,dy\,dx \\
&= \int_{-1}^{0}\left[\frac{5}{4}\left(x^2y+\frac{y^2}{2}\right)\right]_0^{1+x} dx + \int_{0}^{1}\left[\frac{5}{4}\left(x^2y+\frac{y^2}{2}\right)\right]_0^{1-x^2} dx \\
&= \int_{-1}^{0}\frac{5}{4}\left(x^2(1+x)+\frac{(1+x)^2}{2}\right) dx + \int_{0}^{1}\frac{5}{4}\left(x^2(1-x^2)+\frac{(1-x^2)^2}{2}\right) dx \\
&= \int_{-1}^{0}\frac{5}{4}\left(x^2+x^3+\frac{1}{2}+x+\frac{x^2}{2}\right) dx + \int_{0}^{1}\frac{5}{4}\left(x^2-x^4+\frac{1}{2}-x^2+\frac{x^4}{2}\right) dx \\
&= \left[\frac{5}{4}\left(\frac{x}{2}+\frac{x^2}{2}+\frac{x^3}{2}+\frac{x^4}{4}\right)\right]_{-1}^{0} + \left[\frac{5}{4}\left(\frac{x}{2}-\frac{x^5}{10}\right)\right]_0^1 \\
&= \frac{13}{16}
\end{aligned}$$

(d) Find $\mathbb{P}(Y = X^2)$.

Answer: $\mathbb{P}(Y = X^2) = 0$. This is because the region in the (x, y) plane over which we integrate is a curve, $y = x^2$ (which has zero area). If we integrate first with respect to y, the range of integration is x^2 to x^2, giving zero for the definite integral. This is analagous to the calculation of $P(X = k)$ for a continuous univariate random variable (in which case the answer is again 0).

Exercise 18.1

1. Let X and Y be two random variables having the same distribution but which are not necessarily independent. Show that
$$\mathrm{Cov}(X+Y, X-Y) = 0$$
provided that their common distribution has finite mean and variance.

Answer: Perhaps the simplest method is the following: let $U = X + Y$ and $V = X - Y$. Then

$$\begin{aligned}
\mathrm{Cov}(X+Y, X-Y) &= \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) \\
&= \mathbb{E}\big[(X+Y)(X-Y)\big] - \mathbb{E}(X+Y)\mathbb{E}(X-Y) \\
&= \mathbb{E}(X^2-Y^2) - \big[\mathbb{E}(X)+\mathbb{E}(Y)\big]\big[\mathbb{E}(X)-\mathbb{E}(Y)\big] \qquad \text{(by the linearity of expectation)} \\
&= \mathbb{E}(X^2) - \mathbb{E}(Y^2) - \mathbb{E}(X)^2 + \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(Y)^2 \quad \text{(by linearity again)} \\
&= \big[\mathbb{E}(X^2) - \mathbb{E}(X)^2\big] - \big[\mathbb{E}(Y^2) - \mathbb{E}(Y)^2\big] \\
&= \mathrm{Var}(X) - \mathrm{Var}(Y).
\end{aligned}$$

Since X and Y have the same distribution, their variances must be equal, so $\mathrm{Cov}(X+Y, X-Y) = 0$.

2. Consider a fair six-sided die whose faces show the numbers $-2, 0, 0, 1, 3, 4$. The die is independently rolled four times. Let X be the average of the four numbers that appear, and let Y be the product of these four numbers. Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(Y)$ and $\mathrm{Cov}(X, Y)$.

Answer: Let X_1, X_2, X_3, X_4 be independent discrete random variables on the set $\{-2, 0, 0, 1, 3, 4\}$. Each X_i is identically distributed according to the following PMF:

k	-2	0	1	3	4
$\mathbb{P}(X_i = k)$	1/6	1/3	1/6	1/6	1/6

Hence for $i = 1, 2, 3, 4$,

$$\begin{aligned}
\mathbb{E}(X_i) &= \frac{1}{6}(-2+0+0+1+3+4) = 1, \\
\mathbb{E}(X_i^2) &= \frac{1}{6}(4+0+0+1+9+16) = \frac{30}{6} = 5, \\
\mathrm{Var}(X_i) &= \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = 4.
\end{aligned}$$

Let $X = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$. Then

$$\mathbb{E}(X) = \frac{1}{4}\big(\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_4)\big) = 1$$

By independence,

$$\mathrm{Var}(X) = \frac{1}{16}\big(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_4)\big) = 1$$

so $\mathbb{E}(X^2) = \mathrm{Var}(X) + \mathbb{E}(X)^2 = 2$.
and $Y = X_1X_2X_3X_4$. By independence,

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(X_1X_2X_3X_4) \\
&= \mathbb{E}(X_1)\mathbb{E}(X_2)\mathbb{E}(X_3)\mathbb{E}(X_4) \\
&= 1
\end{aligned}$$

and because the X_i are identically distributed,

$$\begin{aligned}
\mathbb{E}(XY) &= \frac{1}{4}\mathbb{E}\big(\mathbb{E}(X_1 + X_2 + X_3 + X_4)X_1X_2X_3X_4\big) \\
&= \mathbb{E}(X_1^2)\mathbb{E}(X_2)\mathbb{E}(X_3)\mathbb{E}(X_4) \\
&= \mathbb{E}(X_1^2) = 5
\end{aligned}$$

so $\mathrm{Cov}(XY) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 4$.

3. A fair die is rolled twice. Let U denote the number obtained on the first roll, let V denote the number obtained on the second roll, let $X = U + V$ denote their sum and let $Y = U - V$ denote their difference. Compute the mean and variance of X and Y, and compute $\mathbb{E}(XY)$. Check whether X and Y are uncorrelated. Check whether X and Y are independent.

Answer: Let $U, V \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$ be independent (and identically distributed) random variables, and define $X = U + V$ and $Y = U - V$.

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(U) + \mathbb{E}(V) = 7 \\
\mathbb{E}(Y) &= \mathbb{E}(U) - \mathbb{E}(V) = 0
\end{aligned}$$

By independence,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(U) + \mathrm{Var}(V) = 35/6 \\
\mathrm{Var}(Y) &= \mathrm{Var}(U) + \mathrm{Var}(V) = 35/6
\end{aligned}$$

Because U and V are identically distributed, and

$$XY = (U+V)(U-V) = U^2 - V^2$$

it follows that

$$\mathbb{E}(XY) = \mathbb{E}(U^2) - \mathbb{E}(V^2) = 0$$

X and Y are uncorrelated, since

$$\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$$

However X and Y are not independent, because (for example)

$$\mathbb{P}(Y=0) \neq \mathbb{P}(Y=0|X=12) = 1$$

Exercise 19.1

1. A fair coin is tossed three times. Let I_j be the indicator variable of the event that a head occurs on the jth toss. Compute the conditional expectation $E(Y|X)$ and verify the identity $E(E(Y|X)) = E(Y)$ in each of the following cases:
   (1) $X = \max\{I_1, I_2, I_3\}$ and $Y = \min\{I_1, I_2, I_3\}$,
   (2) $X = I_1 + I_2$ and $Y = I_2 + I_3$.

Answer:

				Part (i)		Part (ii)	
Outcome	I_1	I_2	I_3	$\max\{I_1,I_2,I_3\}$	$\min\{I_1,I_2,I_3\}$	I_1+I_2	I_2+I_3
HHH	1	1	1	1	1	2	2
HHT	1	1	0	0	1	2	1
HTH	1	0	1	0	1	1	1
THH	0	1	1	0	1	1	2
HTT	1	0	0	0	1	1	0
THT	0	1	0	0	1	1	1
TTH	0	0	1	0	1	0	1
TTT	0	0	0	0	0	0	0

(1) $X = \max\{I_1, I_2, I_3\}$ and $Y = \min\{I_1, I_2, I_3\}$.
The joint and marginal distributions of X and Y are:

	$Y=0$	$Y=1$	f_X
$X=0$	1/8	0	1/8
$X=1$	6/8	1/8	7/8
f_Y	7/8	1/8	

The conditional distributions of Y given $X = x$ are:

	y	
	0	1
$P(Y=y\|X=0)$	1	0
$P(Y=y\|X=1)$	6/7	1/7

The conditional expectations of Y given $X = x$ are:

- $\mathbb{E}(Y|X=0) = (0 \times 1) + (1 \times 0) = 0$,
- $\mathbb{E}(Y|X=1) = \big(0 \times \frac{6}{7}\big) + \big(1 \times \frac{1}{7}\big) = 1/7$.

To verify the law of total expectation:

$$\mathbb{E}(Y) = \left(0 \times \frac{7}{8}\right) + \left(1 \times \frac{1}{8}\right) = \frac{1}{8},$$

and

$$\begin{aligned}\mathbb{E}\big(\mathbb{E}(Y|X)\big) &= \mathbb{E}(Y|X=0)P(X=0) + \mathbb{E}(Y|X=1)P(X=1) \\ &= \left(0 \times \frac{1}{8}\right) + \left(\frac{1}{7} \times \frac{7}{8}\right) = \frac{1}{8} \\ &= \mathbb{E}(Y).\end{aligned}$$

(2) $X = I_1 + I_2$ and $Y = I_2 + I_3$.
The joint and marginal distributions of X and Y are:

	$Y=0$	$Y=1$	$Y=2$	f_X
$X=0$	$1/8$	$1/8$	0	$2/8$
$X=1$	$1/8$	$2/8$	$1/8$	$4/8$
$X=2$	0	$1/8$	$1/8$	$2/8$
f_Y	$2/8$	$4/8$	$2/8$	

The conditional distributions of Y given $X = x$ are:

	y		
	0	1	2
$P(Y=y\|X=0)$	$1/2$	$1/2$	0
$P(Y=y\|X=1)$	$1/4$	$1/2$	$1/4$
$P(Y=y\|X=2)$	0	$1/2$	$1/2$

The conditional expectations Y given $X = x$ are:

- $\mathbb{E}(Y|X=0) = (0 \times 1/2) + (1 \times 1/2) + (2 \times 0) = 1/2$,
- $\mathbb{E}(Y|X=1) = (0 \times 1/4) + (1 \times 1/2) + (2 \times 1/4) = 1$,
- $\mathbb{E}(Y|X=2) = (0 \times 1/2) + (1 \times 0) + (2 \times 1/2) = 3/2$.

To verify the law of total expectation:

$$\mathbb{E}(Y) = \left(0 \times \frac{2}{8}\right) + \left(1 \times \frac{4}{8}\right) + \left(2 \times \frac{2}{8}\right) = \frac{1}{2},$$

and

$$\begin{aligned}\mathbb{E}\big(\mathbb{E}(Y|X)\big) &= \mathbb{E}(Y|X=0)P(X=0) + \mathbb{E}(Y|X=1)P(X=1) + \mathbb{E}(Y|X=2)P(X=2) \\ &= \left(\frac{1}{2} \times \frac{1}{4}\right) + \left(1 \times \frac{1}{2}\right) + \left(\frac{3}{2} \times \frac{1}{4}\right) = \frac{1}{8} + \frac{1}{2} + \frac{3}{8} = \frac{1}{2} \\ &= \mathbb{E}(Y).\end{aligned}$$

2. Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} c(x+y) & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(1) Show that $c = 1$.
(2) Compute the conditional expectation $\mathbb{E}(Y|X)$.

(3) Verify the identity $\mathbb{E}\big(\mathbb{E}(Y|X)\big) = \mathbb{E}(Y)$.

Answer:

(1) By the law of total probability,

$$\int_0^1 \int_0^1 (x+y)\,dx\,dy = 1 \quad \text{so} \quad c = 1$$

(2) The marginal densities are

$$f_X(x) = \int_0^1 (x+y)\,dy = (x+1/2) \qquad 0 \leq x \leq 1$$
$$f_Y(x) = \int_0^1 (x+y)\,dy = (y+1/2) \qquad 0 \leq y \leq 1$$

The conditional expectation of Y given $X = x$ is

$$\begin{aligned}
\mathbb{E}(Y|X=x) &= \int_0^1 y\left(\frac{f_{X,Y}(x,y)}{f_X(x)}\right)dy \\
&= \int_0^1 y\left(\frac{x+y}{x+1/2}\right)dy \\
&= \left[\frac{3xy^2+2y^3}{6x+3}\right]_0^1 \\
&= \frac{3x+2}{6x+3}
\end{aligned}$$

so $\mathbb{E}(Y|X) = \dfrac{3X+2}{6X+3}$.

(3) Check:

$$\begin{aligned}
\mathbb{E}(Y) &= \int_0^1 y f_Y(y)\,dy = \int_0^1 y\left(y+\frac{1}{2}\right)dy = \frac{7}{12} \\
\mathbb{E}\big(\mathbb{E}(Y|X)\big) &= \int_0^1 \mathbb{E}(Y|X=x) f_X(x)\,dx \\
&= \int_0^1 \left(\frac{3x+2}{6x+3}\right)\left(x+\frac{1}{2}\right)dx \\
&= \int_0^1 \left(\frac{x}{2}+\frac{1}{3}\right)dx = \frac{7}{12}
\end{aligned}$$

3. Let the joint density of random variables X and Y be

$$f(x,y) = \begin{cases} cxy & \text{for} \quad 0 \leq x,y \leq 1 \text{ where } \ x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(1) Compute the normalization constant c.
(2) Compute the conditional expectation $\mathbb{E}(Y|X)$.
(3) Verify the identity $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$.

Answer:

(1) The marginal PDF of X is

$$f_X(x) = c\int_0^{1-x} xy\,dy = \frac{cx(1-x)^2}{2}$$

To find c,

$$\int_0^1 f_X(x)\,dx = 1, \qquad \text{so} \quad c = 24.$$

Thus

$$\begin{aligned} f_X(x) &= 12x(1-x)^2 \qquad 0 \le x \le 1 \\ f_Y(y) &= 12y(1-y)^2 \qquad 0 \le y \le 1 \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(X) &= 12\int_0^1 x(1-x)^2\,dx = 2/5 \\ \mathbb{E}(Y) &= 12\int_0^1 y(1-y)^2\,dy = 2/5 \end{aligned}$$

(2) To compute $\mathbb{E}(Y \mid X)$,

$$\begin{aligned} \mathbb{E}(Y \mid X = x) &= \int_0^1 y\left(\frac{f_{X,Y}(x,y)}{f_X(x)}\right)dy \\ &= \int_0^{1-x} y\left(\frac{24xy}{12x(1-x)^2}\right)dy \\ &= \frac{24x}{12x(1-x)^2}\int_0^{1-x} y^2\,dy \\ &= \frac{24x}{12x(1-x)^2}\left[\frac{(1-x)^3}{3}\right] \\ &= \frac{2}{3}(1-x) \end{aligned}$$

so $\mathbb{E}(Y \mid X) = 2(1-X)/3$.

(3) Check:

$$\begin{aligned} \mathbb{E}\big(\mathbb{E}(Y \mid X)\big) &= \int_0^1 \mathbb{E}(Y \mid X = x) f_X(x)\,dx \\ &= \frac{2}{3}\int_0^1 (1-x) f_X(x)\,dx \\ &= \frac{2}{3}\big(1 - \mathbb{E}(X)\big) = \frac{2}{3}\left(1 - \frac{2}{5}\right) = \frac{2}{5} \\ &= \mathbb{E}(Y) \end{aligned}$$

Exercise 20.1

1. Let X and Y have standard bivariate normal distribution, with joint PDF given by

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

where ρ is a constant satisfying $-1 < \rho < 1$.

(a) Check that $f(x,y)$ is indeed a joint PDF, by verifying that $f(x,y) \ge 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$.

Answer: TODO

(b) Check that $\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy = \rho$.

Answer: TODO

(c) Show that if X and Y are uncorrelated, then they are independent.

Answer: TODO

2. Let X and Y have standard bivariate normal distribution. Find the conditional distribution of Y given $X = x$, and hence show that $\mathbb{E}(Y|X) = \rho X$.

Answer: The conditional distribution of Y given $X = x$ is $N(\rho x, 1-\rho^2)$.

3. Let X and Y have standard bivariate normal distribution. Show that X and $Z = \dfrac{Y-\rho X}{\sqrt{1-\rho^2}}$ are independent standard normal random variables.

Answer: TODO

4. Let X and Y have standard bivariate normal distribution, and let $Z = \max\{X,Y\}$. Show that $\mathbb{E}(Z) = \sqrt{(1-\rho)/\pi}$ and $\mathbb{E}(Z^2) = 1$.

Answer: TODO

5. Let $U, V \sim N(0,1)$. Show that the random variables $X = U+V$ and $Y = U-V$ are independent.

Answer:

- The transformation is $g(u,v) = (u+v, u-v)$.
- To compute the inverse transformation, consider $x = u+v$ and $y = u-v$.
- Solving these, we obtain $u = \frac{1}{2}(x+y)$ and $v = \frac{1}{2}(x-y)$.
- Thus the inverse transformation is $(u,v) = g^{-1}(x,y) = \left(\frac{1}{2}(x+y), \frac{1}{2}(x-y)\right)$

The Jacobian determinant of $g^{-1}(x,y)$ is

$$J = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} [r]\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

The joint PDF of U and V is

$$f(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right)$$

Now,

$$\begin{aligned} u^2+v^2-2\rho uv &= \left(\frac{1}{2}(x+y)\right)^2 + \left(\frac{1}{2}(x-y)\right)^2 - 2\rho\left(\frac{1}{2}(x+y)\right)\left(\frac{1}{2}(x-y)\right) \\ &= \frac{1}{2}x^2(1-\rho) + \frac{1}{2}y^2(1+\rho) \end{aligned}$$

The joint PDF of X and Y is therefore

$$\begin{aligned} f(x,y) &= \frac{1}{2}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{x^2(1-\rho)}{4(1-\rho^2)} - \frac{y^2(1+\rho)}{4(1-\rho^2)}\right) \\ &= \frac{1}{\sqrt{4\pi(1+\rho)}}\exp\left(-\frac{x^2}{4(1+\rho)}\right) \times \frac{1}{\sqrt{4\pi(1-\rho)}}\exp\left(-\frac{y^2}{4(1-\rho)}\right) \end{aligned}$$

This is the product of the PDF of a $N\big(0,2(1+\rho)\big)$ variable and the PDF of a $N\big(0,2(1-\rho)\big)$ variable. Thus X and Y are independent.

6. Let X and Y have bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Show that the conditional distribution of Y given $X = x$ is

$$N\left(\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x-\mu_1), \sigma_2^2(1-\rho^2)\right).$$

Answer: TODO

7. (a) Let X and Y be jointly continuous random variables, and let $f_{X,Y}$ be their joint PDF. Show that the PDF of the random variable $X+Y$ can be written as

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_{X,Y}(x, t-x)\,dx = \int_{-\infty}^{\infty} f_{X,Y}(t-y, y)\,dy.$$

Answer: Let $A = \{(x,y) : x+y \leq z\} \subset \mathbb{R}^2$. Then

$$\mathbb{P}(X+Y\leq z) = \iint_A f(x,y)\,dxdy = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{z-x} f_{X,Y}(x,y)\,dy\,dx$$

We change the variable of integration (in the inner integral), making the substitution $y = t - x$:

$$\begin{aligned} F_{X+Y}(z) = \mathbb{P}(X+Y\leq z) &= \int_{x=-\infty}^{\infty}\int_{t=-\infty}^{z} f_{X,Y}(x,t-x)\,dt\,dx \\ &= \int_{t=-\infty}^{z}\int_{x=-\infty}^{\infty} f_{X,Y}(x,t-x)\,dx\,dt \end{aligned}$$

where the final equality follows by reversing the order of integration. Thus the PDF of $X+Y$ is

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_{X,Y}(x,t-x)\,dx \qquad \text{as required.}$$

(b) Hence, or otherwise, show that if $U, V \sim N(0,1)$ are independent, then $U+V \sim N(0,2)$. (This is a special case of Theorem 20.5.)

Answer: By part (a), if two random variables X and Y are independent, the PDF of $X+Y$ is the *convolution* of the marginal PDFs:

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)\,dx = \int_{-\infty}^{\infty} f_X(t-y)f_Y(y)\,dy.$$

U and V are independent, so their joint PDF is

$$f(u,v) = f_U(u)f_V(v) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}(u^2+v^2)\right) \qquad u,v\in\mathbb{R}.$$

Let $W = U + V$. Then because U and V are independent,

$$\begin{aligned}
f_W(w) &= \int_{-\infty}^{\infty} f_U(u) f_V(w-u)\,du \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(u^2+(w-u)^2\right)\right]\,du \\
&= \frac{1}{2\pi}e^{-\frac{1}{4}w^2}\int_{-\infty}^{\infty} \exp\left[-\left(u-\frac{w}{2}\right)^2\right]\,du
\end{aligned}$$

We change the variable of integration, by making the substitution $t = \sqrt{2}\left(u - \dfrac{w}{2}\right)$:

$$f_W(w) = \frac{1}{2\sqrt{\pi}}e^{-\frac{1}{4}w^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}w^2}\,dv = \frac{1}{2\sqrt{\pi}}e^{-\frac{w^2}{4}},$$

which is the PDF of the $N(0,2)$ distribution