Cardiff University | School of Mathematics

MA2500

Foundations of Probability and Statistics

Reading Material

2014-15

Version: 10/10/2014

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

- $\{1, 2, \dots, n\}$ is a finite sample space,
- $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ 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 $A_1, A_2, \dots$ 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 $A_1 \subseteq A_2 \subseteq \ldots$ 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 $B_1 \supseteq B_2 \supseteq \ldots$ 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 $A_1 \subseteq A_2 \subseteq \ldots$ be an increasing sequence of events, and

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

We can write A as a disjoint union

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

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

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

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

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

To prove the second part, let $B_1 \supseteq B_2 \supseteq \ldots$ 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 $A_1 \subseteq A_2 \subseteq \ldots$ 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 $A_1, A_2, \ldots$ 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.

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

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

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 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 the set 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, it is not necessary to 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 by:

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

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 $(\Omega, \mathcal{F})$.