Introduction to Probability – 2018/19

These notes are a summary of what is lectured. They do not contain all examples or explanations and are NOT a substitute for your own notes. The notes are largely based on previous versions of the module by courtesy of Robert Johnson.

Final exam:

- **The final exam will be OPEN BOOK. All printed, handwritten, photocopied etc. material is permitted. Only the use of electronic communication devices is prohibited.**
- **Calculators are PERMITTED in the final exam (but they will be of no use).**
- **You can obtain a copy of your MARKED EXAM SCRIPT, at about 3 weeks after the final exam. If you wish to obtain a copy of your marked final exam script send me a request by email, by March 29th 2019 at the very latest: w.just@qmul.ac.uk. Use your COLLEGE EMAIL ADDRESS and put your STUDENT ID in the subject field.**

The lecture notes contain the problem sheets as well, at the end of each chapter. Problem sheets contain one homework (labelled A-J, to be handed in at the beginning of the tutorial, and to be collected at the following tutorial) and questions to be discussed in class (labelled 1-30).

Tutorials will start in week 2.

There are no tutorials or lectures in week 7 (reading week).

§0 Introduction

Before starting the module properly we will spend a bit of time thinking about what probability is and where it is used.

Example 0.1 *How many people do you need in a room to have a better than 50% chance that two of them share a birthday?*

This is a simple calculation which we will see how to do later. If you haven't seen this before then try to guess what the answer will be. Most people have rather poor intuition for this kind of question so I would expect there to be a wide range of guesses and for many of them to be far from the true answer.

Example 0.2 *Would you rather be given £5, or toss a coin winning £10 if it comes up heads?*

Would you rather be given £5000, or toss a coin winning £10000 if it comes up heads?

Would you rather be given £1, or toss a coin 10 times winning £1000 if it comes up heads every time?

This example is not entirely a mathematical one and there is no right or wrong answer for each part. However some tools from probability can help to describe the choice we have quantitatively. In each case the average amount we expect to win and the degree of variation in our gain are relevant. Properties of so called random variables can be used to describe this choice. Of course there are lots of extra ingredients which will influence our choice (for instance how useful a particular sum of money is to us, and how much we enjoy the excitement of taking a risk). One attempt to model some of these extra factors mathematically is the idea of utility functions from game theory (this is beyond the module).

Example 0.3 *A suspect's fingerprints match those found on a murder weapon. The chance of a match occurring by chance is around 1 in 50,000. Is the suspect likely to be innocent?*

The third example emphasises how important it is to work out exactly what information we are given and what we want to know. The court would have to consider the probability that the suspect is innocent under the assumption that the fingerprints match. The numerical probability given in the question is the probability that the fingerprints match under the assumption that the suspect is innocent. These are in general completely different quantities. The erroneous assumption that they equal is sometimes called the prosecutor's fallacy. The mathematical tool for considering probabilities given certain assumptions is conditional probability.

These questions all relate to situations where there is some randomness; that is events we cannot predict with certainty. This could be either because they happen by chance (tossing a coin) or they are beyond our knowledge (the innocence or guilt of a suspect). Probability theory is about quantifying randomness.

The question of what probability is does not have an entirely satisfactory answer. We will associate a number to an event which will measure the likelihood of it occurring. But what does this really mean? You can think of it some kind of limiting frequency.

Informal definition of a probability: repeat an experiment (say roll a die) N times. Let A denote an event (say the die shows an even number). Suppose the event comes up m times (among the N repetitions of the experiment). Then the ratio m/N (in the limit of very large values N) denotes the probability of the event A.

We will later give a precise mathematical definition of which roughly speaking defines probability in terms of the properties it should have.

§1 Sample Space and Events

The general setting is we perform an *experiment* and record an *outcome*. Outcomes must be precisely specified, mutually exclusive and cover all possibilities.

Definition 1.1 *The* sample space *is the set of all possible outcomes for the experiment. It is denoted by S (or $\mathcal{S}$ or Ω).*

Definition 1.2 *An* event *is a subset of the sample space. The event* occurs *if the actual* *outcome is an element of the set.*

Example 1.1 *Roll a die (experiment) and write down the number showing (outcome). The sample space is the set containing 1,2,3,4,5,6*

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}\,.$$

The event $A = \{2, 4, 6\}$ *corresponds to the rolling of an even number.*

Example 1.2 *A coin is tossed three times and the sequence of heads/tails is recorded.*

$$\mathcal{S} = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

where, for example, htt means the first toss is a head, the second is a tail, the third is a tail.

If A is the event "exactly one head seen" then

$$A = \{htt, tht, tth\}.$$

Equivalently we could take

$$\mathcal{S}' = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \{\}\}$$

where we record the set of tosses which are head. The outcome $\{1, 3\}$ *means hth.*

$$B = \{\{1, 2\}, \{2\}, \{2, 3\}, \{1, 2, 3\}\}$$

is the event "the second toss is a head".

Example 1.3 *Take an exam repeatedly until you pass.*

$$\mathcal{S} = \{P, FP, FFP, FFFP, \ldots\}$$

where we record the sequence of fails and passes. $\mathcal{S}$ *is infinite.*

Other ways representing the sample space: record the number of attempts

$$\mathcal{S}' = \{1, 2, 3, 4, \ldots\}$$

Record the number of fails

$$\mathcal{S}'' = \{0, 1, 2, 3, \ldots\}$$

Among other things these examples illustrate that the sample space may be finite or infinite, and that sometimes there were several equally good ways to describe outcomes.

Definition 1.3 *An event E is a* simple event *(or* elementary event*) if it consists of a single element of the sample space $\mathcal{S}$.*

Example 1.4 *In example 1.3 the event "pass first time", $E = \{P\}$, is a simple event.*

Basic set theory: We use extensively the terminology and notation of basic set theory. Informally a set is an unordered collection of well- defined distinct objects. For instance $\{b, -2.4, \square\}$ is a set. Strictly speaking $\{2, 3, 3\}$ is not a set (sometimes people use slang and mean by these symbols the two element set $\{2, 3\}$). $\{2, 3, 5, 1\}$ and $\{3, 5, 2, 1\}$ are the same sets as order does not matter. Two sets, A and B are equal, $A = B$,if they contain precisely the same elements. A set can be specified in various ways.

- By listing the objects in it between braces ($\{, \}$) separated by commas, e.g., $\{1, 2, 3, 4\}$.
- (usually for infinite sets) By listing enough elements to determine a pattern, e.g., $\{2, 4, 6, 8, \dots\}$ (the set of positive even integers). A set which can be written as a comma separated list is said to be countable.
- By giving a rule, e.g., $\{x : x \text{ is an even integer}\}$ (read as "the set of all x such that x is an even integer").

If A is a set we write $x \in A$ to mean that the object x is in the set A and say that x is an *element* of A. If x is not an element of A then we write $x \notin A$.

Let A and B be sets.

- $A \cup B$ ("A union B") is the set of elements of A or B (or both)

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

- $A \cap B$ ("A intersection B") is the set of elements of both A and B

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

- $A \setminus B$ ("A take away B") is the set of elements in A but not in B
$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}$$

- $A \triangle B$ ("symmetric difference of A and B") is the set of elements in either A or B but not both
$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

- If all the elements of A are contained in the set B we say that A is a subset of B, $A \subseteq B$.

- If all sets are subsets of some fixed set S then A^c ("the complement of A") is the set of all elements of S which are not elements of A.
$$A^c = S \setminus A$$

- If $A = \{a_1, a_2, \ldots, a_n\}$ denotes a finite set of n elements then $|A| = n$ denotes the *size* (or *cardinality*) of the set (do not confuse the size of a set with the absolute value of a number).

- We say two sets A and B are *disjoint* if they have no element in common, i.e.,
$$A \cap B = \emptyset$$
where $\emptyset = \{\}$ denotes the empty set.

Events: Let A and B denote events (i.e. subsets of the sample space $\mathcal{S}$).

- If A is an event then A^c contains the elements of the sample space which are not contained in A, i.e., A^c is the event that "A does not occur".

- If A and B are events then the event E "A and B both occur" consists of all elements of both A and B, i.e., $E = A \cap B$.

- The event E "at least one of A or B occurs" consists of all elements in A or B, i.e., $E = A \cup B$.

- The event E "A occurs but B does not" consists of all elements in A but not in B, i.e., $E = A \setminus B$.

- The event E "exactly one of A or B occurs" consists of all elements in A or B but not in both, i.e., $E = A \triangle B$.

Example 1.5 *Roll a die, let A be the event that an even number occurs (see example 1.1) and B the event that the outcome is a prime number*

$$B = \{2, 3, 5\}$$

Denote by E_1 the event that "the outcome is an even number or a prime" then

$$E_1 = A \cup B = \{2, 3, 4, 5, 6\}$$

Denote by E_2 the event that "the outcome is either an even number or a prime" then

$$E_2 = A \triangle B = \{3, 4, 5, 6\}$$

Exercise Sheet 1

This sheet is based on the material lectured in week 1. The questions are about sample spaces and events. The sheet will be discussed in tutorials in week 2.

Hand in your homework, problem A, to the tutor at the beginning of the tutorial.

Questions 1,2, and 3 will be discussed in the tutorial.

Problem A: (Homework) You throw an ordinary 6-sided die and write down the number showing. Then you throw it again and write down the number showing.

a) Write down the sample space for this experiment.

b) How many elements does the sample space contain?

Problem 1: Four students, Amanda, Brian, Claire and David, are supposed to be attending a lecture course. In the last lecture of the semester their attendance is recorded.

a) Write down the sample space, explaining your notation carefully.

b) Write down the event "Exactly three of them attend the last lecture" as a set.

c) Write down the event "Amanda attends the last lecture but David does not" as a set.

Problem 2: A race takes place between three horses Adobe, Brandy, and Chopin. It is possible that one or more of them may fall and so fail to complete the race. The finishing horses are recorded in the order in which they finish.

a) Write down the sample space, explaining your notation.

b) Write down the event "The race is won by Adobe" as a set.

c) Write down the event "Brandy falls" as a set.

d) Write down the event "All horses complete the race" as a set.

Problem 3: You toss an ordinary coin repeatedly, recording the outcome of each toss. You do this until you have seen either two Heads or three Tails in total and then you stop.

a) Write down the sample space.

b) Write down the event "you toss the coin exactly four times" as a subset of the sample space.

I perform the same experiment but I do not stop until I have seen either 7 Heads or 8 Tails in total. Let E_i be the event "I toss the coin exactly i times".

c) For which i is it the case that $E_i = \emptyset$?

§2 Properties of Probabilities

We want to assign a numerical value to an event which reflects the chance that it occurs. Probability is a concept/recipe (or in formal terms a function) $\mathbb{P}$ which assigns a (real) number $\mathbb{P}(A)$ to each event A, $0 \leq \mathbb{P}(A) \leq 1$.

The simplest way of doing this is to say that the probability of an event A is the ratio of the number of outcomes in A to the total number of outcomes in S. We sometimes describe this as the situation where all outcomes are equally likely.

Example 2.1 *Roll a die, sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Consider the event A "the number shown is smaller than 3", $A = \{1, 2\}$. Define the probability of A by*

$$\mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|} = \frac{2}{6} = \frac{1}{3}$$

Sometimes this is a reasonable thing to do but often it is not. In the example 2.1 above if the die is biased however this would not be a reasonable notion of probability.

We also run into difficulties if S is infinite. If $S = \mathbb{N}$ (the set of positive integers), see example 1.3 then there is no reasonable way to choose an element of S with all outcomes equally likely. There are however ways to choose a random positive integer in which every outcome has some chance of occurring.

The formal approach is to regard probability as a mathematical construction satisfying certain axioms.

Definition 2.1 (Kolmogorov's Axioms for Probability) *Probability is a function $\mathbb{P}$ which assigns to each event A a real number $\mathbb{P}(A)$ such that:*

a) For every event A we have $\mathbb{P}(A) \geq 0$,

b) $\mathbb{P}(\mathcal{S}) = 1$,

c) If $A_1, A_2, \ldots, A_n$ are n pairwise disjoint events ($A_i \cap A_j = \emptyset$ for all $i \neq j$) then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots + \mathbb{P}(A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Remark:

- The function $\mathbb{P}$ is sometimes called a *probability measure*.
- We say that events satisfying definition 2.1c are *pairwise disjoint* or *mutually exclusive*.
- Notice that definition 2.1c has a version for a countable infinite number of events as well, which we do not cover in this module (e.g. we would need to clarify the notion of an infinite sum first). Further subtleties occur if S is infinite (more particularly if it is not countable).

Example 2.2 *Suppose that the sample space $\mathcal{S}$ is finite. Then the setting*

$$\mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|}$$

gives a probability. This is the case when every outcome in the sample space is equally likely (we say we pick an outcome at random). We need to show that this setting obeys definition 2.1.

a) If $A \subseteq \mathcal{S}$ then $|A| \geq 0$ and

$$\mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|} \geq \frac{0}{|\mathcal{S}|} = 0$$

b)

$$\mathbb{P}(\mathcal{S}) = \frac{|\mathcal{S}|}{|\mathcal{S}|} = 1$$

c) If $A_1, A_2, \ldots A_n$ are pairwise disjoint subsets of $\mathcal{S}$ then

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = |A_1| + |A_2| + \ldots + |A_n|$$

and

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) &= \frac{|A_1| + |A_2| + \ldots + |A_n|}{|\mathcal{S}|} \\
&= \frac{|A_1|}{|\mathcal{S}|} + \frac{|A_2|}{|\mathcal{S}|} + \ldots + \frac{|A_n|}{|\mathcal{S}|} \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots \mathbb{P}(A_n)
\end{aligned}$$

In this situation calculating probabilities becomes counting!

Warning: In general, do not assume that outcomes are equally likely without good reason.

Example 2.3 *Toss a biased coin. The sample space is given by* $\mathcal{S} = \{h, t\}$. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{h\}) = 1/3$, $\mathbb{P}(\{t\}) = 2/3$, *and* $\mathbb{P}(\mathcal{S}) = 1$ *defines a probability measure.*

Starting from the axioms we can deduce various properties. Hopefully, these will agree with our intuition about probability (if they did not then this would suggest that we had not made a good choice of axioms). The proofs of all of these are simple deductions from the axioms.

Proposition 2.1 *If* A *is an event then*

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

The statement makes perfect sense. If $\mathbb{P}(A)$ is the probability of the event A then the probability of the complementary event A^c should be $1-\mathbb{P}(A)$. If we are able to provide a formal proof then we have evidence that definition 2.1 is consistent with real world setups.

Proof: Let A be any event. Set $A_1 = A$ and $A_2 = A^c$. By definition of the complement $A_1 \cap A_2 = \emptyset$ and so we can apply definition 2.1c (with $n = 2$) to get

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

But (again by definition of the complement) $A_1 \cup A_2 = S$ so by definition 2.1b

$$1 = \mathbb{P}(S) = \mathbb{P}(A_1 \cup A_2).$$

Combining both expressions

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

Rearranging this gives the result. $\square$.

We can use the results we have proved to deduce further ones such as the following ones. These are called corollaries as they are an "obvious" consequence of the proposition 2.1.

Corollary 2.1

$$\mathbb{P}(\emptyset) = 0.$$

This statement makes perfect sense as well. The probability of "nothing" is zero.

Proof: By definition of complement $\mathcal{S}^c = \mathcal{S} \setminus \mathcal{S} = \emptyset$. Hence by Proposition 2.1

$$\mathbb{P}(\emptyset) = \mathbb{P}(S^c) = 1 - \mathbb{P}(S)$$

Using definition 2.1b we have

$$\mathbb{P}(\emptyset) = 1 - 1 = 0$$

$\square$.

Corollary 2.2 *If A is an event then $\mathbb{P}(A) \leq 1$.*

Again the statement sounds sensible. Probabilities are always smaller or equal to one.

Proof: By proposition 2.1

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

But A^c is an event, so by definition 2.1a

$$0 \leq \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

and hence

$$\mathbb{P}(A) \leq 1$$

$\square$.

The following statements are less obvious consequences of definition 2.1 and the statements we have shown so far. Thus we call them again propositions.

Proposition 2.2 *If A and B are events and $A \subseteq B$ then*

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

This statement looks sensible as well. If an event B contains all the outcomes of an event A then the former has the higher probability.

Proof: Consider the events $A_1 = A$ and $A_2 = B \setminus A$. Then $A_1 \cap A_2 = \emptyset$ (the two events are pairwise disjoint) and $A_1 \cup A_2 = B$. So by definition 2.1c (with $n = 2$)

$$\mathbb{P}(B) = \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$$

Since $B \setminus A$ is an event definition 2.1a tells us that

$$\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(B \setminus A) \geq 0$$

The statement follows by rearrangement. $\square$.

Proposition 2.3 *If $A = \{a_1, a_2, \ldots, a_n\}$ is a finite event then*

$$\mathbb{P}(A) = \mathbb{P}(\{a_1\}) + \mathbb{P}(\{a_2\}) + \ldots + \mathbb{P}(\{a_n\}) = \sum_{i=1}^{n} \mathbb{P}(\{a_i\}).$$

The statement is quite remarkable. The probability of a (finite) event is the sum of the probabilities of the corresponding simple events. One often writes $\mathbb{P}(a_i)$ for $\mathbb{P}(\{a_i\})$ even though the former one is incorrect.

Proof: Denote by $A_i = \{a_i\}$, $i = 1, .., n$ the simple events. The events are pairwise disjoint and $A_1 \cup A_2 \cup \ldots \cup A_n = A$. So by definition 2.1c

$$\mathbb{P}(A) = \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots + \mathbb{P}(A_n) = \mathbb{P}(\{a_1\}) + \mathbb{P}(\{a_2\}) + \ldots + \mathbb{P}(\{a_n\})$$

$\square$.

Proposition 2.4 (Inclusion-exclusion for two events) *For any two events A and B we have*

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

The statement is not entirely obvious. For general events the probability of the event "A or B" is normally not the sum of the probabilities of events A and B. Because of some "double counting" one needs to correct by the probability of the event "A and B".

Proof: Consider the three events $E_1 = A \setminus B$, $E_2 = A \cap B$ and $E_3 = B \setminus A$. The events are pairwise disjoint and $E_1 \cup E_2 \cup E_3 = A \cup B$. Hence by definition 2.1c (with $n = 3$)

$$\mathbb{P}(A \cup B) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3)\,.$$

Furthermore $E_1 \cup E_2 = A$ and $E_2 \cup E_3 = B$. Thus definition 2.1c (with $n = 2$) yields

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(E_1) + \mathbb{P}(E_2) \\ \mathbb{P}(B) &= \mathbb{P}(E_2) + \mathbb{P}(E_3)\,. \end{aligned}$$

Since $P(A \cap B) = \mathbb{P}(E_2)$ we finally have

$$\begin{aligned} \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) &= \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_2) \\ &= \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) = \mathbb{P}(A \cup B)\,. \end{aligned}$$

$\square$.

Example 2.4 *Consider a finite sample space $\mathcal{S}$ with probability*

$$\mathbb{P}(E) = \frac{|E|}{|\mathcal{S}|}\,.$$

Example 2.2 tells us that in this case definition 2.1 is fulfilled and in particular proposition 2.4 applies. Let $A \subseteq \mathcal{S}$ and $B \subseteq \mathcal{S}$ denote two events then proposition 2.4 reads

$$\frac{|A \cup B|}{|\mathcal{S}|} = \frac{|A|}{|\mathcal{S}|} + \frac{|B|}{|\mathcal{S}|} - \frac{|A \cap B|}{|\mathcal{S}|}$$

that means

$$|A \cup B| = |A| + |B| - |A \cap B|\,.$$

This last statement is a statement about the sizes of finite sets (which can be proven in other ways as well). Hence probability theory can give results which do not involve any "randomness". This cross fertilisation is essentially what mathematics is about.

Proposition 2.5 (Inclusion-exclusion for three events) *For any three events A, B and C we have*

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

As for two events there exists an "intuitive" argument but that is not a proof.

Proof: Essentially we will apply proposition 2.4 three times. Let $D = A \cup B$ so that $A \cup B \cup C = C \cup D$. Then

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(C \cup D) \\
&= \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C \cap D) \text{ by proposition 2.4} \\
&= \mathbb{P}(C) + \mathbb{P}(A \cup B) - \mathbb{P}(C \cap D) \\
&= \mathbb{P}(C) + \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) - \mathbb{P}(C \cap D) \text{ by proposition 2.4}
\end{aligned}$$

Now $C \cap D = C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$ so that

$$\begin{aligned}
\mathbb{P}(C \cap D) &= \mathbb{P}((C \cap A) \cup (C \cap B)) \\
&= \mathbb{P}(C \cap A) + \mathbb{P}(C \cap B) - \mathbb{P}((C \cap A) \cap (C \cap B)) \text{ by proposition 2.4} \\
&= \mathbb{P}(C \cap A) + \mathbb{P}(C \cap B) - \mathbb{P}(A \cap B \cap C)
\end{aligned}$$

Combining both expressions

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

$\square$.

Remark: There is also an inclusion-exclusion formula for n events. If you like work out this formula and try to prove it (but it won't be easy).

Example 2.5 *Suppose that the probabilities for each of the three events A, B, and C is $1/3$, i.e.*

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{3}.$$

Furthermore assume that the probabilities for each of the events "A and B", "A and C", and "B and C" is $1/10$, i.e.,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{10}.$$

What can be said about the probability of the event that none of A, B, or C occur?

The event we are interested in is the event $(A \cup B \cup C)^c$. *Hence*

$$\begin{aligned}
\mathbb{P}((A \cup B \cup C)^c) &= 1 - \mathbb{P}(A \cup B \cup C) \text{ by proposition 2.1} \\
&= 1 - [\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \\
&\quad \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + P(A \cap B \cap C)] \text{ by proposition 2.5} \\
&= 1 - \left[\frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} + P(A \cap B \cap C)\right] \\
&= \frac{3}{10} - P(A \cap B \cap C)\,.
\end{aligned}$$

Since $A \cap B \cap C \subseteq A \cap B$, *definition 2.1a and proposition 2.2 tell us that*

$$0 \le P(A \cap B \cap C) \le \mathbb{P}(A \cap B) = \frac{1}{10}\,.$$

So

$$\frac{2}{10} \le \mathbb{P}((A \cup B \cup C)^c) \le \frac{3}{10}\,.$$

Note that the bounds are sharp, i.e., there are examples where equality is possible. Consider for instance the sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ *with all outcomes being equally likely. Events* $A = \{1, 2, 3\}$, $B = \{1, 4, 5\}$, $C = \{1, 6, 7\}$ *satisfy the conditions of the example and* $\mathbb{P}((A \cup B \cup C)^c) = 3/10$.

Remark: The various properties of probabilities derived in this paragraph are just based on the basic definition 2.1. Hence few simple axioms can lead to a plethora of results (such structures is essentially what mathematics is about).

Exercise Sheet 2

This sheet is based on the material lectured in week 2. The questions are about set notation and properties of probabilities. The sheet will be discussed in tutorials in week 3.

Hand in your homework, problem B, to the tutor at the beginning of the tutorial, and collect problem A from your tutor to complete your submission.

Questions 4, 5, and 6 will be discussed in the tutorial.

Problem B: (Homework) Let A and B be events with $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/4$, $\mathbb{P}(A \cap B) = 1/10$. Calculate the following probabilities:

a) $\mathbb{P}(B^c)$

b) $\mathbb{P}(A \cup B)$

c) $\mathbb{P}(A \cap B^c)$

d) $\mathbb{P}(A \setminus B)$

Problem 4: Suppose that A, B and C are events.

a) Using the inclusion exclusion principle, or otherwise, show that

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1\,.$$

b) Using the events $C = A \triangle B$ and $D = A \cap B$, or otherwise, show that

$$\mathbb{P}(A \triangle B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)\,.$$

Make sure that each step in your proofs is justified by a definition, axiom or result from lectures (or is a simple manipulation).

Problem 5: One player is chosen at random from a squad of cricketers. Let C be the event "the captain is chosen", B be the event "the player chosen is a batsman", L be the event "the player chosen is left-handed", and I be the event "the player chosen is injured".

a) Express the following in symbols:

i) The event "an injured batsmen is chosen"

ii) The event "the chosen player is a batsmen but is not the captain"

iii) The statement "the captain is a batsman"

iv) The statement "batsmen are fit"

b) Express the following in words:

i) The event I^c

ii) The event $B^c \cup L^c$

iii) The statement $C \subseteq L$

iv) The statement $|I^c| < 11$

c) Suppose that 50% of the squad are batsmen, 25% are lefthanders and 10% are lefthanded batsmen and that the player chosen is equally likely to be any member of the squad. Find the following:

i) the probability that the chosen player is right-handed,

ii) the probability that the chosen player is either a batsman or is left-handed,

iii) the probability that the player is a right-handed batsman.

Problem 6: Let $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and for $A \subseteq \mathcal{S}$ define a probability measure by

$$\mathbb{P}(A) = \frac{1}{12}\left(|A \cap \{1, 2, 3, 4\}| + 2|A \cap \{5, 6, 7, 8\}|\right).$$

a) Verify that this satisfies the axioms for probability.

b) Give an example of a physical situation which you would expect the mathematics above to describe.

§3 Sampling

When all elements of the sample space are equally likely calculating probability often boils down to counting the number of ways of making some selection, see example 2.2. Specifically, we are often interested in finding how many ways there are of choosing r things from an n element set. This is called sampling from an n element set. The answer to this question depends on exactly what we mean by selection: is the order important and is repetition allowed.

a) Ordered Sampling with Replacement

Example 3.1 *How many flags are there consisting of 3 vertical stripes each of which is red, blue, green or white (and if we allow adjacent stripes to be of the same colour)?*

This is an ordered selection of 3 things from a set of 4 things with repetition allowed. There are 4 choices for the first stripe. For each of these there are 4 choices of the second stripe, and for each of these there are 4 choices for the 3rd stripe. In total there are $4 \times 4 \times 4 = 4^3 = 64$ *flags.*

- A set is an unordered collection of different objects. Consider a set $U = \{u_1, u_2, \ldots, u_n\}$ of n elements, $|U| = n$.

- An ordered selection from the set U of r objects which are not necessarily distinct, $(u_1, u_2, \ldots, u_r)$ is called an r-tuple. For instance $(3, -1, 3, 5)$ and $(3, 3, -1, 5)$ are different 4-tuples. A 2-tuple (u_1, u_2) is called a pair (e.g. coordinates in the Cartesian plane).

- If we make an ordered selection of r things from a set U with replacement allowed (that is to say we allow elements to be repeated) then the sample space is the set of all ordered r-tuples of elements of U. That is

$$S = \{(u_1, u_2, \ldots, u_r) : u_i \in U\},$$

- If $|U| = n$ there are n choices for u_1, for each of these there are n choices for u_2, and so on. Hence, we have

$$|S| = |U|^r = n^r.$$

b) Ordered Sampling without Replacement

Example 3.2 *Suppose that in example 3.1 we are not allowed to repeat a colour. There are 4 choices for the first stripe. For each of these there are 3 choices for the second stripe, and for each of these there are 2 choices for the 3rd stripe. In total there are* $4 \times 3 \times 2 = 24$ *flags.*

- Consider a set $U = \{u_1, u_2, \ldots, u_n\}$ of size $|U| = n$.
- If we make an ordered selection of r things from a set U without replacement (that is to say we do not allow elements to be repeated) then the sample space is the set of all ordered r-tuples of distinct elements of U. That is

$$S = \{(u_1, u_2, \ldots, u_r) : u_i \in U \text{ with } u_i \neq u_j \text{ for all } i \neq j\}.$$

- To find the cardinality of this notice that if $|U| = n$ there are n choices for u_1, for each of these choices there are $n-1$ choices for u_2, for each of these there are $n-2$ choices for u_3 and so on. Hence,

$$\begin{aligned} |\mathcal{S}| &= n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) \\ &= \frac{n \times (n-1) \times (n-2\times) \ldots \times (n-r+1) \times (n-r) \times (n-r-1) \times \ldots \times 2 \times 1}{(n-r) \times (n-r-1) \times \ldots \times 2 \times 1} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

where $k! = k \times (k-1) \times \ldots \times 2 \times 1$ ("k factorial") with the convention $0! = 1$.

Example 3.3 *How many permutations of* $(1, 2, \ldots, n)$ *are there? A permutation is an ordered sample of* n *things from the set* $U = \{1, 2, \ldots, n\}$ *of* n *things, i.e., we sample* $r = n$ *objects without replacement. So there are* $n \times (n-1) \times \ldots \times 2 \times 1 = n!/0! = n!$ *permutations.*

c) Unordered Sampling without Replacement

Example 3.4 *How many ways are there to choose 5 players for a penalty shootout from a team of 11 football players (goalkeeper is permitted).?*

If we were to pick the players in shooting order we would make an ordered selection without replacement. So there are $11!/6!$ *ways to do this. Given there are* 5 *players there are* $5!$ *ways to arrange it into shooting order. So* $5! \times$ *number of possibilities* $= 11!/6!$ *and the number of possibilities is* $11!/(6! \times 5!)$.

- Consider a set $U = \{u_1, u_2, \ldots, u_n\}$ of size $|U| = n$.
- If we make an unordered selection of r things from a set U without replacement then we obtain a subset A of U of size r.
- The corresponding sample space is the set of all subsets of r distinct elements of U.

$$\mathcal{S} = \{A \subseteq U : |A| = r\}.$$

- An *ordered* sample is obtained by taking an element of this sample space $\mathcal{S}$ and putting its elements in order. Each element of the sample space can be ordered in $r!$ ways and so (using the formula for ordered selections with repetition, see section 3b, and assuming $|U| = n$) we have that

$$r! \times |\mathcal{S}| = \frac{n!}{(n-r)!},$$

and so

$$|\mathcal{S}| = \frac{n!}{(n-r)!r!}.$$

Remark: This expression is usually written as

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

(read as "n choose r") and is called a *binomial coefficient*. By convention $\binom{n}{r} = 0$ when $r > n$. Binomial coefficients appear for instance in the binomial theorem

$$\begin{aligned}(a+b)^n &= \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n-1}a^1 b^{n-1} + \binom{n}{n}a^0 b^n \\ &= \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.\end{aligned}$$

Try to verify this formula for $n = 2$ or $n = 3$, and if you like prove the expression (e.g. by induction).

d) Summary

To summarise we have shown the following

Theorem 3.1 *The number of ways of selecting (sampling) r things from an n element set is*

a) ordered with replacement/repetition allowed: n^r

b) ordered without replacement/no repetition : $n!/(n-r)!$

c) unordered without replacement/no repetition : $\binom{n}{r}$

Remark: We have not covered/proven the case of unordered sampling with replacement/repetition. This case is rather elaborate to deal with.

Example 3.5 *You have 10 coins, 7 silver ones and 3 copper ones, and you pick 4 coins at random (i.e. all outcomes are equally likely). Let*

- D *be the event that you pick 4 silver coins,*
- E *be the event that you pick 2 silver coins followed by 2 copper coins,*
- F *be the event that you pick 2 silver and 2 cooper coins in any order.*

Find $\mathbb{P}(D)$, $\mathbb{P}(E)$, *and* $\mathbb{P}(F)$.

Since we pick coins at random $\mathbb{P}(A) = |A|/|\mathcal{S}|$. *So wee need to determine the size of the sample space and the size of the event.*

The set U *of objects contains 7 (in principle distinct) silver coins and 3 (distinct) copper coins,* $U = \{c_1, c_2, c_3, s_1, s_2, \ldots, s_6, s_7\}$. *It is important to note that we are going to sample from such a set (and by definition elements of sets are distinct, even though 7 elements have the same property, namely being silver, and 3 have the same property, namely being copper).*

i) *Let us consider the experiment as ordered sampling (that is definitely required for event* E*, for the others we have choices) without replacement. Outcomes are an ordered selection of 4 objects, i.e., a 4-tuple* (u_1, u_2, u_3, u_4) *with* $u_k \in U$*. Since* $n = 10$ *and* $r = 4$ *(in each case we pick 4 coins) the size of the sample space is, according to theorem 3.1b*

$$|\mathcal{S}| = \frac{10!}{6!} = 10 \times 9 \times 8 \times 7\,.$$

Consider event D*. The outcome in* D *is an ordered sample of 4 things from the (sub)set of 7 silver coins (without replacement), for instance* (s_5, s_2, s_6, s_3)*. According to theorem 3.1b*

$$|D| = \frac{7!}{3!} = 7 \times 6 \times 5 \times 4$$

so

$$\mathbb{P}(D) = \frac{|D|}{|\mathcal{S}|} = \frac{7 \times 6 \times 5 \times 4}{10 \times 9 \times 8 \times 7} = frac16\,.$$

Consider event E*. An outcome in* E *is of the form* (s_i, s_j, c_k, c_ℓ)*.*

- *There are 7 choices for* s_i*.*
- *For each of these there are 6 choices for* s_j*.*
- *For each of these there are 3 choices for* c_k*.*
- *For each of these there are 2 choices for* c_ℓ*.*

So

$$|E| = 7 \times 6 \times 3 \times 2$$

and

$$\mathbb{P}(E) = \frac{|E|}{|\mathcal{S}|} = \frac{7 \times 6 \times 3 \times 2}{10 \times 9 \times 8 \times 7} = \frac{1}{20}\,.$$

Consider event F*. There are 6 possible patterns to select 2 silver and 2 copper coins, namely sscc, scsc, sccs, cssc, cscs, and ccss and each occurs in* $7 \times 6 \times 3 \times 2$ *ways (see the previous event). So*

$$|F| = 6 \times 7 \times 6 \times 3 \times 2$$

and

$$\mathbb{P}(F) = \frac{|F|}{|\mathcal{S}|} = \frac{6 \times 7 \times 6 \times 3 \times 2}{10 \times 9 \times 8 \times 7} = \frac{3}{10}\,.$$

ii) Let us consider the experiment as unordered sampling without replacement. That is simpler, but we cannot cover event E as there the order is important. Outcomes are now 4-element subsets $\{u_1, u_2, u_3, u_4\}$ with $u_k \in U$. Since $n = 10$ and $r = 4$ the size of the sample space is, according to theorem 3.1c

$$|\mathcal{S}| = \binom{10}{4} = \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1}\,.$$

Consider event D. Outcomes in D are $\{c_m, c_n, c_k, c_\ell\}$, i.e., 4 element subsets taken from the set of 7 copper coins. We sample $r = 4$ coins from a (sub)set of $n = 7$ coins. So

$$|D| = \binom{7}{4} = \frac{7 \times 6 \times 5 \times 4}{4 \times 3 \times 2 \times 1}$$

and

$$\mathbb{P}(D) = \frac{|D|}{|\mathcal{S}|} = \frac{\binom{7}{4}}{\binom{10}{4}} = \frac{7 \times 6 \times 5 \times 4}{10 \times 9 \times 8 \times 7} = \frac{1}{6}$$

(of course the same result as in i)).

Consider event F. An outcome in F has the form $\{s_m, s_n, c_k, c_\ell\} = \{s_m, s_n\} \cup \{c_k, c_\ell\}$. According to theorem 3.1c there are $\binom{7}{2}$ ways to select 2 silver coins from a (sub)set of 7, and $\binom{3}{2}$ ways to select 2 copper coins from a (sub)set of 3. Thus

$$|F| = \binom{7}{2}\binom{3}{2} = \frac{7 \times 6 \times 3 \times 2}{2 \times 1 \times 2 \times 1}$$

and

$$\mathbb{P}(F) = \frac{|F|}{|\mathcal{S}|} = \frac{\binom{7}{2}\binom{3}{2}}{\binom{10}{4}} = \frac{7 \times 6 \times 3 \times 2 \times 4 \times 3 \times 2}{10 \times 9 \times 8 \times 7 \times 2 \times 2} = \frac{3}{10}$$

(of course again the same result).

Example 3.6 (The birthday problem) *Ask 30 people for their birthday (assume nobody is born on 29 February). Let A be the event that there is a repeated birthday, and A^c be the event that there is no repeated birthday.*

Let us consider the experiment as ordered sampling from a set of $n = 365$ days. An outcome in the sample space is a 30-tuple $(u_1, u_2, \ldots, u_{30})$ with u_k denoting the respective birthday. The size of the sample space is, according to theorem 3.1a

$$|\mathcal{S}| = 365^{30}\,.$$

Consider the event A^c*, i.e. a ordered selection of birthdays* $(u_1, u_2, \ldots, u_{30})$ *without repetition. The outcomes of the event* A^c *are ordered samples of size* $r = 30$ *without replacement/repetition (as* A^c *contains non repeated birthdays only) from a set of size* $n = 365$*. According to theorem 3.1a*

$$|A^c| = \frac{365!}{335!} = 365 \times 364 \times 363 \times \ldots \times 337 \times 336\,.$$

Thus

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \text{ by proposition 2.1} \\
&= 1 - \frac{|A^c|}{|\mathcal{S}|} \\
&= 1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \ldots \times \frac{337}{365} \times \frac{336}{365} \\
&= 0.70631624\ldots
\end{aligned}$$

and the probability of a repeated birthday among 30 people is more than 70%.

Example 3.7 (Poker dice) *You roll 5 dice. Find the probability to roll "a pair" (that means outcomes of the type e.g. 33152 etc.).*

We are sampling $r = 5$ *objects from the set* $U = \{1, 2, 3, 4, 5, 6\}$ *of size* $n = 6$*. We sample with replacement/repetition, so consider ordered sampling with replacement.*

Size of the sample space, theorem 3.1a

$$|\mathcal{S}| = |U|^r = 6^5\,.$$

Denote by A the event to "roll a pair". Consider first those outcomes in the event A which have the pair as the first two entries, i.e., outcomes (p, p, r_1, r_2, r_3)*. There are 6 choices for p, for each of those there are 5 choices for* r_1*, for each of those there are 4 choices for* r_2*, for each of those there are 3 choices for* r_3*. Hence there are* $6 \times 5 \times 4 \times 3$ *of such outcomes.*

But A contains as well outcomes where the pair does not come as the first two entries, i.e., outcomes which differ from the pattern $pprrr$*. There are in total "5 choose 2" different patterns* $pprrr$*,* $prprr$*,* $prprr$*,...,* $rrrpp$*, each giving rise to* $6 \times 5 \times 4 \times 3$ *outcomes. Hence*

$$|A| = 10 \times 6 \times 5 \times 4 \times 3\,.$$

Thus

$$\mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|} = \frac{10 \times 6 \times 5 \times 4 \times 3}{6^5} = \frac{25}{54}\,.$$

Remark: It is important when answering questions involving sampling that you read the question carefully and decide what sort of sampling is involved. Specifically, how many things are you selecting, what set are you selecting from, does the order matter, and is repetition allowed or not. Sometimes more than one sort can be used but you must be consistent (see the previous example 3.5).

Exercise Sheet 3

This sheet is based on the material lectured in week 3. The questions are about sampling. The sheet will be discussed in tutorials in week 4.

Hand in your homework, problem C, to the tutor at the beginning of the tutorial, and collect problem B from your tutor to complete your submission.

Questions 7, 8, and 9 will be discussed in the tutorial.

Problem C: (Homework) When I open a bank account I am allocated a 4 digit personal identification number (which may begin with one or more zeros) at random.

a) What is the cardinality of the sample space for this experiment?

b) By computing the cardinality of each of these events find the probability that:

i) Every digit of my number is even.

ii) My number has no repeated digits.

iii) My number is palindromic (reads the same forwards as backwards).

iv) No digit of my number exceeds 7.

v) The largest digit in my number is exactly 7.

vi) The digits in my number are in strictly increasing order.

Problem 7: Each member of a squad of 18 cricketers is either a batsmen or a bowler. The squad comprises 10 batsmen and 8 bowlers. An eccentric coach chooses a team by picking a random set of 11 players from the squad.

a) What is the probability that the team is made up of 6 batsmen and 5 bowlers?

b) What is the probability that the team contains fewer than 3 bowlers?

Problem 8: You are dealt a hand of five cards from a standard well shuffled deck[1].

[1] A deck of playing cards is made up of 52 cards split into 4 suits ($\clubsuit, \diamondsuit, \heartsuit, \spadesuit$) with each suit made up of one card of each of 13 ranks ($2, 3, 4, \ldots, 10$,Jack,Queen,King,Ace).

Consider the following poker hands:

- A flush (all cards of the same suit)
- Four of a kind (four of the same rank and one extra card)
- A full house (three of one rank and two of another rank)

Calculate the probability of getting each hand. In each case simplify the answer as much as you can by hand and then find a numerical value using a calculator.

Problem 9: Let $1 \leq r \leq n$. A subset of $\{1, 2, \ldots, n\}$ of cardinality r is chosen at random.

a) Calculate the probability that 1 is an element of the chosen subset.

b) Without using your answer to part a) calculate the probability that one is not an element of the chosen subset.

c) Deduce that

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}.$$

§4 Conditional Probability

Additional information (a so called "condition") may change the probability of an event.

Example 4.1 *Roll a fair die. Sample space*

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

Consider event A "the number shown is odd"

$$A = \{1, 3, 5\}, \quad \mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|} = \frac{1}{2}$$

and event B "the number shown is smaller than 4"

$$B = \{1, 2, 3\}, \quad \mathbb{P}(B) = \frac{|B|}{|\mathcal{S}|} = \frac{1}{2}$$

Suppose we roll the die and somebody tells us that the number shown is odd (i.e. event A has happened/ event A is given). Now the probability of event B (given event A) is

$$\mathbb{P}(B \text{ given event } A) = \frac{|\{1, 3\}|}{|\{1, 3, 5\}|} = \frac{2}{3}$$

How is this expression linked with $\mathbb{P}(A)$, $\mathbb{P}(B)$, ...?

$$\mathbb{P}(B \text{ given event } A) = \frac{|\{1,3\}|}{|\{1,3,5\}|} = \frac{|\{1,3\}|/|\mathcal{S}|}{|\{1,3,5\}|/|\mathcal{S}|} = \frac{|\{1,3\}|/|\mathcal{S}|}{\mathbb{P}(A)} = \frac{|A \cap B|/|\mathcal{S}|}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

The expression $\mathbb{P}(B \text{ given event } A) = \mathbb{P}(B|A)$ is called conditional probability.

Example 4.2 *You have three pens coloured blue, red, and green. You pick one at random, then you pick another without replacement. Size of the sample space*

$$|\mathcal{S}| = 3 \times 2 = 6$$

Let A be the event that the first pen is red. Let B be the event that the second pen is blue.

$$\begin{aligned}
\mathbb{P}(A) &= \frac{|\{rb, rg\}|}{|\mathcal{S}|} = \frac{1}{3} \\
\mathbb{P}(B) &= \frac{|\{rb, gb\}|}{|\mathcal{S}|} = \frac{1}{3} \\
\mathbb{P}(A \cap B) &= \frac{|\{rb\}|}{|\mathcal{S}|} = \frac{1}{6}
\end{aligned}$$

Clearly $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$

Now assume we know that event A *has happened (i.e. the first pick was a red pen). The probability for the event* B *(with the condition that* A *has happened) is* $1/2$ *(picking at random a blue pen from the remaining two pens) and* $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times 1/2$ *! Here*

$$\frac{1}{2} = \frac{|\{rb\}|}{|\{rb, rg\}|} = \frac{|A \cap B|}{|A|} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

and such a quantity is called a conditional probability (the probability for the event B *conditioned on* A *has occurred).*

Definition 4.1 *If* A *and* B *are events and* $\mathbb{P}(A) \neq 0$ *then the* conditional probability of B given A*, usually denoted by* $\mathbb{P}(B|A)$*, is*

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

Remark:

- The notation $\mathbb{P}(A|B)$ is very unfortunate. $A|B$ is not an event, it has no meaning ! Do not confuse the conditional probability $\mathbb{P}(A|B)$ with $\mathbb{P}(A \setminus B)$, the probability for the event "A and not B".

- Note that the definition does not require that B happens after A. In example 4.2 it would make sense to talk about $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$, the probability that the first pen is red given that the second pen is blue. One way of thinking of this is to imagine that the experiment is performed secretly and the fact that A occurred is revealed to you (without the full outcome being revealed). The conditional probability of B given A is the new probability of B in these circumstances.

- Conditional probability can be used to measure how the occurrence of some event influences the chance of another event occurring.

  If $\mathbb{P}(B|A) < \mathbb{P}(B)$ then A occurring makes B less probable.

  If $\mathbb{P}(B|A) > \mathbb{P}(B)$ then A occurring makes B more probable.

  If $\mathbb{P}(B|A) = \mathbb{P}(B)$ then the event A has no impact on the probability of event B (see section 5).

- In some cases we will use the conditional probability $P(B|A)$ to calculate $\mathbb{P}(A\cap B)$, in some cases we will use $\mathbb{P}(A\cap B)$ to find the conditional probability $\mathbb{P}(B|A)$.

Example 4.3 *Roll a fair die ("pick at random") twice. Consider the events*

- *A: "first roll is a 6"*
- *B: "roll a double"*
- *C: "roll at least one odd number"*

So (if considered as ordered sampling)

$$|\mathcal{S}| = 36$$

and

$$\mathbb{P}(A) = \frac{1}{6}, \qquad \mathbb{P}(B) = \frac{1}{6}, \qquad \mathbb{P}(C) = \frac{3}{4}$$
$$\mathbb{P}(A\cap B) = \frac{1}{36}, \quad \mathbb{P}(B\cap C) = \frac{1}{12}, \quad \mathbb{P}(A\cap C) = \frac{1}{12}$$

Conditional probability of a double given that first roll is a 6

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)} = \frac{1/36}{1/6} = \frac{1}{6}\,.$$

This is somehow obvious.

Conditional probability to roll a 6 first given that one rolls a double

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1}{6}\,.$$

Slightly less obvious.

Conditional probability to roll at least one odd given that one rolls a double

$$\mathbb{P}(C|B) = \frac{\mathbb{P}(C\cap B)}{\mathbb{P}(B)} = \frac{1}{2}\,.$$

Somehow obvious.

Conditional probability to roll a double given that one rolls at least one odd

$$\mathbb{P}(B|C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{1}{9}\,.$$

Probably not obvious at all.

Notice $\mathbb{P}(C|B) < \mathbb{P}(C)$ *so the information that a double is rolled makes it less likely to see an odd number. And* $\mathbb{P}(B|C) < \mathbb{P}(B)$*, so the information that an odd number is seen makes it less likely to have rolled a double.*

But $\mathbb{P}(B|A) = \mathbb{P}(B)$ *and* $\mathbb{P}(A|B) = \mathbb{P}(A)$*. That means rolling a 6 first does not change the chance of seeing a double, and rolling a double does to change the chance of having rolled a 6 first. Events with such a property will be called* independent *and we will discuss the issue in detail in the next section.*

Example 4.4 *Reconsider the event* D *in example 3.5 (picking 4 silver coins from a set of 7 silver and 3 copper coins). We can work out* $\mathbb{P}(D)$ *using conditional probability.*

Let A_i *be the event that the ith coin picked is a silver coin. Obviously* $D = A_1 \cap A_2 \cap A_3 \cap A_4$.

$$\mathbb{P}(A_1) = \frac{7}{10}\,.$$

If A_1 *occurs then the 2nd pick is a coin from 9 of which 6 are silver. So*

$$\mathbb{P}(A_2|A_1) = \frac{6}{9}$$

and by the definition 4.1 of the conditional probability we have

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) = \frac{7}{10} \times \frac{6}{9}\,.$$

If A_1 *and* A_2 *occur then the third pick is from 8 coins 5 of which are silver so*

$$\mathbb{P}(A_3|A_1 \cap A_2) = \frac{5}{8}\,.$$

Again the definition of conditional probability tells us that

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_2)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8}\,.$$

Finally assuming that A_1*,* A_2*, and* A_3 *have occurred the fourth pick is from 7 coins 4 of which are silver so*

$$\mathbb{P}(A_4|A_1 \cap A_2 \cap A_3) = \frac{4}{7}$$

and

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1 \cap A_2 \cap A_3)\mathbb{P}(A_4|A_1 \cap A_2 \cap A_3) = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} \times \frac{4}{7}.$$

The event $A_1 \cap A_2 \cap A_3 \cap A_4$ is the event D in example 3.5 and we have obtained of course the same numerical value for its probability.

We can easily generalise this example for arbitrary events

Theorem 4.1 *Let $E_1, E_2, \ldots, E_n$ be events then*

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_3|E_1 \cap E_2) \times \cdots \times \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

provided that all of the conditional probabilities involved are defined.

One can easily prove the statement by plugging in the definition of conditional probabilities and cancel common factors in the numerator and denominator. But picky pure mathematicians may stick up their noses, so we will use induction

Proof: Consider $n = 2$. Definition 4.1 tells us

$$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)}.$$

Hence the statement of the theorem for $n = 2$ follows

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1).$$

Now assume we have shown that the statement of the theorem holds for $n = k$, i.e.

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_k) = \mathbb{P}(E_1) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_3|E_1 \cap E_2) \times \cdots \times \mathbb{P}(E_k|E_1 \cap E_2 \cap \cdots \cap E_{k-1}) \quad (*)$$

Definition 4.1 tells us

$$\mathbb{P}(E_{k+1}|E_1 \cap E_2 \cap \cdots \cap E_k) = \frac{\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{k+1})}{\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_k)}$$

that means

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{k+1}) = \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_k) \times \mathbb{P}(E_{k+1}|E_1 \cap E_2 \cap \cdots \cap E_k).$$

Using equation (*) we arrive at

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{k+1}) &= \mathbb{P}(E_1) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_3|E_1 \cap E_2) \times \ldots \\ &\quad \cdots \times \mathbb{P}(E_k|E_1 \cap E_2 \cap \cdots \cap E_{k-1}) \times \mathbb{P}(E_{k+1}|E_1 \cap E_2 \cap \cdots \cap E_k)\end{aligned}$$

which means the statement of the theorem holds as well for $n = k+1$ (and by the principle of induction for all $n \geq 2$) $\square$.

Conditional probability provides an alternative approach to questions involving ordered sampling.

Example 4.5 (ordered sampling revisited) *Consider again that we pick r things at random from a set U of size $n = |U|$ in order. Consider a particular (given/fixed) outcome of this experiment $(u_1, u_2, \ldots, u_r)$. Let A denote the event that this particular outcome occurs, $A = \{(u_1, u_2, \ldots, u_r)\}$.*

$$\mathbb{P}(A) = \frac{1}{|\mathcal{S}|}$$

We are going to compute $P(A)$ (i.e. $|\mathcal{S}|$) using conditional probability.

Let E_k denote the event that "the kth pick gives u_k" (with $u_k \in U$). $A = E_1 \cap E_2 \cap \ldots \cap E_r$ is the event to pick a particular ordered sample $(u_1, u_2, \ldots, u_r)$.

By theorem 4.1 the probability of this event can be written as

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_r) = \mathbb{P}(E_1) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_3|E_1 \cap E_2) \times \cdots \times \mathbb{P}(E_r|E_1 \cap E_2 \cap \cdots \cap E_{r-1})$$

Note that this expression is valid no matter whether we sample with or without replacement. The latter will come into play when we work out the conditional probabilities.

- *Suppose we sample without replacement and $u_1, u_2, \ldots, u_r$ are distinct. Then*

  $$\mathbb{P}(E_1) = \frac{1}{n}$$

  as we pick the element u_1 at random from the set U of size $|U| = n$.

  $$\mathbb{P}(E_2|E_1) = \frac{1}{n-1}$$

  as we pick the element u_2 at random from the set $U \setminus \{u_1\}$ of size $n-1$. In general

  $$\mathbb{P}(E_i|E_1 \cap E_2 \cap \ldots \cap E_{i-1}) = \frac{1}{n-i+1}$$

as we pick the element u_i *at random from the set* $U\backslash\{u_1, u_2, \ldots, u_{i-1}\}$ *of size* $n-i+1$. *Hence*

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_r) = \frac{1}{n} \times \frac{1}{n-1} \times \ldots \times \frac{1}{n-r+1} = \frac{(n-r)!}{n!}$$

- *Suppose we sample with replacement. Then*

$$\mathbb{P}(E_i | E_1 \cap E_2 \cap \ldots \cap E_{i-1}) = \frac{1}{n}$$

as we pick the element u_i *at random from the set* U *of size* $|U| = n$. *Hence*

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_r) = \left(\frac{1}{n}\right)^r$$

In both cases the answer agrees with $1/|\mathcal{S}|$ *where* $|\mathcal{S}|$ *has been calculated in theorem 3.1b and 3.1a, respectively (recall that our event is a simple event!). The method used in this example is a bit more intuitive and the assumption on what is equally likely is more transparent.*

Exercise Sheet 4

This sheet is based on the material lectured in week 4. The questions are about conditional probability. The sheet will be discussed in tutorials in week 5.

Hand in your homework, problem D, to the tutor at the beginning of the tutorial, and collect problem C from your tutor to complete your submission.

Questions 10, 11, and 12 will be discussed in the tutorial.

Problem D: (Homework) A standard fair die is rolled twice.

- **a)** Find the probability that that sum of the two rolls is at least 9.
- **b)** Find the conditional probability that the first roll is 4 given that the sum of the two rolls is at least 9.
- **c)** Find the conditional probability that the first roll is not 4 given that the sum of the two rolls is at least 9.
- **d)** Find the conditional probability that the first roll is 4 given that the sum of the two rolls is less than 9.
- **e)** Find the conditional probability that the sum of the two rolls is at least 9 given that the first roll is a 4.

Problem 10: When I travel into work each morning I notice whether my train is late and by how much and also whether I am able to get a seat on it. Let A be the event "the train is not late", B be the event "the train is late but by not more than 15 minutes", and C be the event "I am able to get a seat". Suppose that $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/4$, $\mathbb{P}(C) = 1/3$ and $\mathbb{P}(A \cap C) = 1/4$.

- **a)** Find the conditional probability that the train is more that 15 minutes late given that the train is late.
- **b)** Find the conditional probability that I get a seat given that the train is late.

Problem 11: Let A and B be events with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.

a) Prove that $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.

b) What can you say about $\mathbb{P}(A|B^c)$?

c) Prove that if $\mathbb{P}(A|B) > \mathbb{P}(A)$ then $\mathbb{P}(B|A) > \mathbb{P}(B)$.

d) Illustrate parts a) and c) of this question using question D.

Problem 12: Two treatments for a disease are tested on a group of 390 patients. Treatment A is given to 160 patients of whom 100 are men and 60 are women; 20 of these men and 40 of these women recover. Treatment B is given to 230 patients of whom 210 are men and 20 are women; 50 of these men and 15 of these women recover.

a) For which of A and B is there a higher probability that a patient chosen randomly from among those given that treatment recovers? Express this as an inequality between two conditional probabilities.

b) For which of A and B is there a higher probability that a man chosen randomly from among those given that treatment recovers? Express this as an inequality between two conditional probabilities.

c) For which of A and B is there a higher probability that a woman chosen randomly from among those given that treatment recovers? Express this as an inequality between two conditional probabilities.

d) Compare the inequality in part a) with the inequalities in part b) and c). Are you surprised by the result?

§5 Independence

Example 4.3 tells us that the probability $\mathbb{P}(A \cap B)$ being the product $\mathbb{P}(A)\mathbb{P}(B)$ is very special.

Example 5.1 *Roll a fair die twice and consider the events:*

- *A: First roll shows an even number*
- *B: Number shown on the second roll is larger than 4*

Obviously

$$\mathbb{P}(A) = \frac{3 \times 6}{36} = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{6 \times 2}{36} = \frac{1}{3}\,.$$

Furthermore for the event $A \cap B$ ("first roll even and second roll larger than 4")

$$\mathbb{P}(A \cap B) = \frac{3 \times 2}{36} = \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B)\,.$$

Events with such a property are said to be independent. In particular

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

and

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

i.e. the conditional probabilities do not depend on the condition.

Definition 5.1 *We say that the events A and B are (pairwise)* independent *if*

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

Remark: Don't assume independence without good reasons. You may assume that events are independent in the following situations:

i) they are clearly physically unrelated (e.g. depend on different coin tosses),

ii) you calculate their probabilities and find that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ (i.e. to check the definition 5.1),

iii) the question tells you that the events are independent!

Example 5.2 *Reconsider example 4.3, rolling a die twice, with events A: "first roll is a 6", B: "double", and C: "at least one odd number".*

Since $\mathbb{P}(A) = 1/6$, $\mathbb{P}(B) = 1/6$ *and* $P(A \cap B) = 1/36$ *we have*

$$P(A \cap B) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(B)$$

Events A *and* B *are independent.*

Since $\mathbb{P}(A) = 1/6$, $P(C) = 3/4$ *and* $P(A \cap C) = 1/12$ *we have*

$$\mathbb{P}(A \cap C) = \frac{1}{12} \neq \frac{1}{6} \times \frac{3}{4} = \mathbb{P}(A)\mathbb{P}(C)$$

So A *and* C *are not independent.*

Independence is *not* the same as physically unrelated. For example we saw in example 5.2 that if a fair die is rolled twice then the event "first roll is a 6" and the event "both rolls produce the same number" are independent.

As examples 4.3 and 5.2 suggest there is a connection between independence and conditional probability

Theorem 5.1 *Let* A *and* B *be events with* $\mathbb{P}(A) > 0$ *and* $\mathbb{P}(B) > 0$. *The following are equivalent:*

a) A *and* B *are independent,*

b) $\mathbb{P}(A|B) = \mathbb{P}(A)$,

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

This result says roughly that if A and B are independent then telling you that A occurred does not change the probability that B occurred.

Proof: It is sufficient to show that a) implies b), b) implies c), and c) implies a).

a)⇒ b) Suppose A and B are independent, i.e.,

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

Since $\mathbb{P}(B) \neq 0$ we have

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B) \text{ by definition 4.1}$$

b)⇒ c) Suppose that $\mathbb{P}(A|B) = \mathbb{P}(A)$. Then by definition 4.1

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

Since $\mathbb{P}(A) \neq 0$ (and $A \cap B = B \cap A$) it follows

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

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

c)⇒ a) Suppose that $\mathbb{P}(B|A) = \mathbb{P}(B)$. Then

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

which implies $\mathbb{P}(B \cap A) = \mathbb{P}(A)\mathbb{P}(B)$, i.e., A and B are independent.

□.

Example 5.3 *Roll a fair die twice and consider the following events*

- A*: first roll shows odd number*
- B*: second roll shows odd number*
- C*: the sum of both rolls is an odd number*

Obviously $|\mathcal{S}| = 36$ *and* $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$*. Furthermore* $|A \cap B| = 9$ *(as there are three choices each for each roll to be odd) and* $\mathbb{P}(A \cap B) = 9/36$*. So*

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)$$

and A *and* B *are independent.*

$|C| = 18$ (if the sum is odd one roll is odd the other even; there are 9 odd/even pairs and 9 even/odd pairs) and $\mathbb{P}(C) = 18/36 = 1/2$. Furthermore $|A \cap C| = 9$ (there are 9 odd/even pairs) and $\mathbb{P}(A \cap C) = 9/36$. Hence

$$\mathbb{P}(A \cap C) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(C)$$

and the events A and C are independent.

$|B \cap C| = 9$ (there are 9 even/odd pairs) and $\mathbb{P}(B \cap C) = 1/4$- Hence

$$\mathbb{P}(B \cap C) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(B)\mathbb{P}(C)$$

and the events B and C are independent.

The events A, B and C are pairwise independent *(each pair of events is independent).*

But the event $A \cap B \cap C$ is impossible (to be precise: $A \cap B \cap C = \emptyset$; if both outcomes are odd the sum is even) and

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) .$$

For three and more events the notion of independence becomes more sophisticated.

Example 5.4 *Three events A, B and C are called* pairwise independent *if*

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

The three events are called mutually independent *if in addition*

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

The events in example 5.3 are not mutually independent.

It is somehow obvious how to generalise the definition in example 5.4 to 4, 5 and more events. However, the formal definition looks awkward (I am tempted to call such things glorified common sense)

Definition 5.2 *We say that the events $A_1, A_2, \ldots, A_n$ are* mutually independent *(sometimes also written as* each event is independent of all the others*) if for any $2 \leq t \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_t \leq n$ we have*

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_t}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_t}).$$

Example 5.5 *You toss a fair coin three times. Consider the following events*

- *A: The first and the second toss show the same result.*
- *B: The first and the last toss show different results.*
- *C: The first toss shows tail.*

In set notation $A = \{hhh, hht, tth, ttt\}$, $B = \{hht, htt, thh, tth\}$, $C = \{ttt, tth, thh, tht\}$.

Clearly

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{1}{2}.$$

Since $A \cap B = \{hht, tth\}$, $A \cap C = \{thh, tth\}$, $B \cap C = \{thh, tth\}$ we have

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C)$$

and events A, B, C are pairwise independent.

In addition $A \cap B \cap C = \{tth\}$ so that

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

and the events are even mutually independent.

Example 5.6 *Consider two buses, say bus A and bus B, running from station X to station Y along two different routes. Consider the events*

- *A: Bus A is running*
- *B: Bus B is running*

with probabilities $\mathbb{P}(A) = 9/10$ *and* $\mathbb{P}(B) = 4/5$. *Assume the events* A *and* B *to be independent. What is the probability that one can travel from* X *to* Y *by bus?*

The event we are interested in is $A \cup B$ *(Bus* A *or Bus* B *is running). Hence*

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \text{ by proposition 2.4, inclusion-exclusion} \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \text{ by definition 5.1, mutual independence} \\
&= \frac{9}{10} + \frac{4}{5} - \frac{36}{50} = \frac{49}{50}
\end{aligned}$$

Example 5.7 *We have two coins. One is fair and the other has probability* $3/4$ *of coming up heads. We pick a coin at random and toss it twice. Define the events*

- F*: We pick the fair coin.*
- H_1*: The first toss is head.*
- H_2*: The second toss is head.*

Are the events H_1 *and* H_2 *independent?*

$\mathbb{P}(F) = 1/2$ *denotes the probability to pick the fair coin.*

$\mathbb{P}(F^c) = 1/2$ *denotes the probability to pick the biased coin*

$\mathbb{P}(H_1|F) = 1/2$ *denotes the (conditional) probability that the first toss shows a head, assuming we have picked the fair coin.*

$\mathbb{P}(H_1|F^c) = 3/4$ *denotes the (conditional) probability that the first toss shows a head, assuming we have picked the biased coin.*

$\mathbb{P}(H_2|F) = 1/2$ *denotes the (conditional) probability that the second toss shows a head, assuming we have picked the fair coin.*

$\mathbb{P}(H_2|F^c) = 3/4$ *denotes the (conditional) probability that the second toss shows a head, assuming we have picked the biased coin.*

$\mathbb{P}(H_1 \cap H_2|F)$ *denotes the (conditional) probability that the first toss shows head and the second toss shows a head, assuming we have picked the fair coin. Coin tosses of the (fair* *or of the unfair) coin are considered to be independent events (see definition 5.1), i.e., the corresponding probabilities factorise. In the present case where we restrict to the fair coin that means*

$$\mathbb{P}(H_1 \cap H_2|F) = \mathbb{P}(H_1|F)\mathbb{P}(H_2|F) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}\,.$$

Such a property is called conditional independence *with respect to the event F (picking a fair coin).*

$\mathbb{P}(H_1 \cap H_2|F^c)$ *denotes the (conditional) probability that the first toss shows head and the second toss shows a head, assuming we have picked the biased coin. Since subsequent tosses of the same (here: the biased) coin are independent we have*

$$\mathbb{P}(H_1 \cap H_2|F^c) = \mathbb{P}(H_1|F^c)\mathbb{P}(H_2|F^c) = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16}\,.$$

Now compute $\mathbb{P}(H_1)$, $\mathbb{P}(H_2)$ and $\mathbb{P}(H_1 \cap H_2)$ to check for independence.

$$\begin{aligned}\mathbb{P}(H_1) &= \mathbb{P}(H_1 \cap F) + \mathbb{P}(H_1 \cap F^c) = \mathbb{P}(H_1|F)\mathbb{P}(F) + \mathbb{P}(H_1|F^c)\mathbb{P}(F^c) \\ &= \frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} = \frac{5}{8}\end{aligned}$$

Similarly

$$\mathbb{P}(H_2) = \frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} = \frac{5}{8}$$

and

$$\mathbb{P}(H_1 \cap H_2) = \frac{1}{4} \times \frac{1}{2} + \frac{9}{16} \times \frac{1}{2} = \frac{26}{64}\,.$$

Hence

$$\mathbb{P}(H_1 \cap H_2) = \frac{26}{64} \neq \frac{5}{8} \times \frac{5}{8} = \mathbb{P}(H_1)\mathbb{P}(H_2)$$

and the events H_1 and H_2 are not (!) independent.

The concept introduced in the previous example 5.7 can be generalised to arbitrary events.

Definition 5.3 *Two events A and B are said to be* conditionally independent *given an event C if*

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

Exercise Sheet 5

This sheet is based on the material lectured in week 5. The questions are about independent and dependent events. The sheet will be discussed in tutorials in week 6.

Hand in your homework, problem E, to the tutor at the beginning of the tutorial, and collect problem D from your tutor to complete your submission.

Questions 13, 14, and 15 will be discussed in the tutorial.

Problem E: (Homework) You roll a die twice. Let A be the event that the first roll is odd, B the event that you roll at least one 6, and C the event that the sum of the rolls is 7.

a) Are the events A and B independent?

b) Are the events A and B conditionally independent given C?

Problem 13: A positive integer from the set $\{1, 2, 3, \ldots, 36\}$ is chosen at random with all choices equally likely. Let E be the event "the chosen number is even", O be the event "the chosen number is odd", Q be the event "the chosen number is a perfect square", and D_k be the event "the chosen number is a multiple of k". Decide with justification:

a) Are the events E and O independent?

b) Are the events E and Q independent?

c) Are the events O and Q independent?

d) Are the events D_3 and D_4 independent?

e) Are the events D_4 and D_6 independent?

Problem 14: The top card of a thoroughly shuffled deck of playing cards is turned over. Let A be the event "the card is an Ace", R be the event "the card belongs to a red suit

($\diamondsuit$ or $\heartsuit$)", and M be the event "the card belongs to a major suit ($\heartsuit$ or $\spadesuit$)". Show that the events A, R and M are mutually independent.

Problem 15: Prove that if A, B and C are mutually independent events then A and $B \cup C$ are independent.

§6 Total Probability

Example 6.1 *A coin is tossed three times and the sequence of heads/tails is recorded, see example 1.2 for the sample space $\mathcal{S}$. Consider the three events*

- $E_1 = \{htt, hht, hth, htt\}$: *first toss is head*
- $E_2 = \{thh, tht\}$: *first toss is tail and second toss is head*
- $E_3 = \{tth, ttt\}$: *first and second toss are tail*

The three events are pairwise disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$) and $\mathcal{S} = E_1 \cup E_2 \cup E_3$. The three events "split" the sample set into three parts, we call E_1, E_2, E_3 a partition of the sample space.

Definition 6.1 *The events $E_1, E_2, \ldots, E_n$* partition *$\mathcal{S}$ (or* form a partition of *$\mathcal{S}$) if they are pairwise disjoint (i.e. $E_k \cap E_\ell = \emptyset$ if $k \neq \ell$) and $E_1 \cup E_2 \cup \ldots \cup E_n = \mathcal{S}$.*

As we mentioned earlier, conditional probability can be used as an aid to calculating probabilities.

Theorem 6.1 (Theorem of total probability) *If $E_1, E_2, \ldots, E_n$ partition $\mathcal{S}$ and $\mathbb{P}(E_k) > 0$ for all k then for any event A we have*

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|E_1)\mathbb{P}(E_1) + \mathbb{P}(A|E_2)\mathbb{P}(E_2) + \ldots + \mathbb{P}(A|E_n)\mathbb{P}(E_n) \\ &= \sum_{k=1}^{n} \mathbb{P}(A|E_k)\mathbb{P}(E_k).\end{aligned}$$

Remark: This theorem is often used to calculate the (total) probability of A, $\mathbb{P}(A)$, if we know the conditional probabilities $\mathbb{P}(A|E_i)$ (i.e. probabilities under certain constraints) and the so called marginal probabilities $\mathbb{P}(E_i)$. The technique is called *conditioning*.

Proof: Let $A_i = A \cap E_i$. These events are pairwise disjoint (since $A_i \subset E_i$ and the sets E_j are pairwise disjoint) and

$$A_1 \cup A_2 \cup \ldots \cup A_n = A \cap (E_1 \cup E_2 \cup \ldots \cup E_n) = A \cap \mathcal{S} = A\,.$$

So by definition 2.1c

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots + \mathbb{P}(A_n)\,.$$

Since $\mathbb{P}(E_i) > 0$ we have (for any $1 \le i \le n$)

$$\mathbb{P}(A_i) = \mathbb{P}(A \cap E_i) = \frac{\mathbb{P}(A \cap E_i)}{\mathbb{P}(E_i)}\mathbb{P}(E_i) = \mathbb{P}(A|E_i)\mathbb{P}(E_i)$$

and the last two equations yield the statement of the theorem. $\square$.

Example 6.2 *The probability that an icecream seller sells all his stock depends on the weather*

weather	*sunny*	*cloudy*	*rain*
probability to sell all stock	9/10	3/5	3/10

Tomorrow the weather will be sunny with probability $1/2$*, cloudy with probability* $1/3$ *and raining with probability* $1/6$*. Find the probability of the event* A *that they sells all their stock.*

Let E_s *be the event that the weather is sunny,* E_c *the event that the weather is cloudy, and* E_r *that the weather is rainy. These events are pairwise disjoint. Let us further assume that this covers all the weather conditions (i.e.* E_s, E_c, E_r *is a partition). The table tells us conditional probabilities. By theorem 6.1*

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A|E_s)\mathbb{P}(E_s) + \mathbb{P}(A|E_c)\mathbb{P}(E_c) + \mathbb{P}(A|E_r)\mathbb{P}(E_r) \\
&= \frac{9}{10} \times \frac{1}{2} + \frac{3}{5} \times \frac{1}{3} + \frac{3}{10} \times \frac{1}{6} \\
&= \frac{14}{20} = \frac{7}{10}\,.
\end{aligned}$$

There exists as well an analogue of theorem 6.1 for conditional probabilities.

Theorem 6.2 *If* $E_1, E_2, \ldots, E_n$ *partition* $\mathcal{S}$*, and* A *and* B *are events with* $\mathbb{P}(B \cap E_i) > 0$ *for all* i *then*

$$\begin{aligned}
\mathbb{P}(A|B) &= \mathbb{P}(A|B \cap E_1)\mathbb{P}(E_1|B) + \mathbb{P}(A|B \cap E_2)\mathbb{P}(E_2|B) + \ldots + \mathbb{P}(A|B \cap E_n)\mathbb{P}(E_n|B) \\
&= \sum_{i=1}^{n} \mathbb{P}(A|B \cap E_i)\mathbb{P}(E_i|B).
\end{aligned}$$

Proof: Using (see definition 4.1)

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

and applying theorem 6.1 to $\mathbb{P}(A \cap B)$ we have

$$\mathbb{P}(A|B) = \frac{1}{\mathbb{P}(B)} \left[\mathbb{P}(A \cap B|E_1)\mathbb{P}(E_1) + \mathbb{P}(A \cap B|E_2)\mathbb{P}(E_2) + \ldots + \mathbb{P}(A \cap B|E_n)\mathbb{P}(E_n)\right] \quad (*)$$

Now for any term of the sum (for any $1 \leq i \leq n$)

$$\begin{aligned}
\frac{1}{\mathbb{P}(B)}\mathbb{P}(A \cap B|E_i)\mathbb{P}(E_i) &= \frac{1}{\mathbb{P}(B)} \frac{\mathbb{P}(A \cap B \cap E_i)}{P(E_i)}\mathbb{P}(E_i) \text{ by definition 4.1} \\
&= \frac{1}{P(B)}\mathbb{P}(A \cap B \cap E_i)\frac{\mathbb{P}(B \cap E_i)}{\mathbb{P}(B \cap E_i)} \text{ using } \mathbb{P}(B \cap E_i) > 0 \\
&= \frac{\mathbb{P}(A \cap B \cap E_i)}{\mathbb{P}(B \cap E_i)} \frac{\mathbb{P}(B \cap E_i)}{\mathbb{P}(B)} \\
&= \mathbb{P}(A|B \cap E_i)\mathbb{P}(E_i|B) \text{ by definition 4.1}
\end{aligned}$$

Hence equation (*) gives the statement of the theorem. $\square$.

Example 6.3 *We have two coins. One is fair and the other has probability $3/4$ of coming up heads. We pick a coin at random and toss it (see example 5.7).*

a) What is the probability to get a head ?

b) Suppose we get a head. What is the probability that the coin is fair ?

c) Suppose we get a head. What is the probability that a second toss of the same coin gives head again ?

Use the notation of example 5.7 for the events

- *F: We pick the fair coin.*
- *H_1: The first toss is head.*
- *H_2: The second toss is head.*

a) Given the fair coin the probability for head is $1/2$ that means (in probability notation)

$$\mathbb{P}(H_1|F) = \frac{1}{2}.$$

F^c is the event to pick the biased coin which has probability $3/4$ to show head, that means

$$\mathbb{P}(H_1|F^c) = \frac{3}{4}.$$

The events F and F are pairwise disjoint (by definition) and $\mathcal{S} = F \cup F^c$, i.e. the events are a partition. Theorem 6.1 tells us that

$$\mathbb{P}(H_1) = \mathbb{P}(H_1|F)\mathbb{P}(F) + \mathbb{P}(H_1|F^c)\mathbb{P}(F^c) = \frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} = \frac{5}{8}$$

since we pick coins at random ($\mathbb{P}(F) = \mathbb{P}(F^c) = 1/2$)

b) We are looking for $\mathbb{P}(F|H_1)$. By definition 4.1

$$\mathbb{P}(F|H_1) = \frac{\mathbb{P}(F \cap H_1)}{\mathbb{P}(H_1)} = \frac{\mathbb{P}(H_1|F)\mathbb{P}(F)}{\mathbb{P}(H_1)} = \frac{1/2 \times 1/2}{5/8} = \frac{2}{5}$$

In particular $\mathbb{P}(F|H_1) \neq \mathbb{P}(H_1|F)$.

c) We are looking for $\mathbb{P}(H_2|H_1)$. Using theorem 6.2 and the partition F, F^c we have

$$\mathbb{P}(H_2|H_1) = \mathbb{P}(H_2|H_1 \cap F)\mathbb{P}(F|H_1) + \mathbb{P}(H_2|H_1 \cap F^c)\mathbb{P}(F^c|H_1)$$

We have (part b)

$$\mathbb{P}(F|H_1) = \frac{2}{5}$$

and (see problem 11a, sheet 4)

$$\mathbb{P}(F^c|H_1) = 1 - \mathbb{P}(F|H_1) = 1 - \frac{2}{5}.$$

Using conditional independence of coin tosses we have (see example 5.7)

$$\begin{aligned}
\mathbb{P}(H_2|H_1 \cap F) &= \frac{P(H_2 \cap H_1 \cap F)}{\mathbb{P}(H_1 \cap F)} = \frac{\mathbb{P}(H_2 \cap H_1|F)\mathbb{P}(F)}{\mathbb{P}(H_1|F)\mathbb{P}(F)} \\
&= \frac{\mathbb{P}(H_2 \cap H_1|F)}{\mathbb{P}(H_1|F)} = \frac{\mathbb{P}(H_2|F)\mathbb{P}(H_1|F)}{\mathbb{P}(H_1|F)} = \mathbb{P}(H_2|F) = \frac{1}{2}
\end{aligned}$$

and similarly

$$\mathbb{P}(H_2|H_1 \cap F^c) = \mathbb{P}(H_2|F^c) = \frac{3}{4}.$$

Hence

$$\mathbb{P}(H_2|H_1) = \frac{1}{2} \times \frac{2}{5} + \frac{3}{4} \times \frac{3}{5} = \frac{13}{20}.$$

As we have seen $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ are very different things (see e.g. example 6.3). The following theorem relates these two conditional probabilities .

Theorem 6.3 (Bayes' theorem) *If A and B are events with $\mathbb{P}(A), \mathbb{P}(B) > 0$ then*

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

Proof: By definition 4.1

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

$\square$.

Example 6.4 *There is a disease which 0.1% of the population suffers from. A test for the disease has probability 99% of giving a positive result for someone with the disease and a 0.5% chance of showing that a healthy person has the disease (a so called false positive). What is the probability that a person testing positive does have the disease?*

Let D be the event that a person tested has the disease, and P be the event that the test is positive.

We know that

$$\begin{aligned}
\mathbb{P}(D) &= \frac{1}{1000} \text{ assuming tested people are picked at random} \\
\mathbb{P}(P|D) &= \frac{99}{100} \\
\mathbb{P}(P|D^c) &= \frac{5}{1000} \text{ since } D^c \text{ is the event the person tested is healthy}
\end{aligned}$$

We are looking for $\mathbb{P}(D|P)$.

Bayes' theorem, theorem 6.3, tells us that

$$\mathbb{P}(D|P) = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P)} = \frac{99/100 \times 1/1000}{\mathbb{P}(P)}$$

Since D, D^c is a partition theorem 6.1 tells us that

$$\begin{aligned}
\mathbb{P}(P) &= \mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^c)\mathbb{P}(D^c) \\
&= \mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^c)(1 - \mathbb{P}(D)) \\
&= \frac{99}{100} \times \frac{1}{1000} + \frac{5}{1000} \times \frac{999}{1000} = \frac{5985}{1000000}
\end{aligned}$$

Hence

$$\mathbb{P}(D|P) = \frac{990}{5985} = \frac{22}{133} = 0.16541\ldots$$

Thus assuming the test is positive there is only a 17% chance that the person is infected. That means at about 83% of positive tests are false positives. Does it mean the test is useless, or is there anything one can do about this ?

Remark: The prosecutors fallacy, example 0.3, has similar features.

Exercise Sheet 6

This sheet is based on the material lectured in week 6. The questions are about the law of total probability and Bayes' theorem. The sheet will be discussed in tutorials in week 8.

Hand in your homework, problem F, to the tutor at the beginning of the tutorial, and collect problem E from your tutor to complete your submission.

Questions 16, 17, and 18 will be discussed in the tutorial.

Problem F: (Homework) Two important members of a cricket team are injured, and each has probability $1/3$ of recovering before the match. The recoveries of the two players are independent of each other. If both are able to play then the team has probability $3/4$ of winning the match, if only one of them plays then the probability of winning is $1/2$ and if neither play the probability of winning is $1/16$. What is the probability that the match is won?

Problem 16: Which of the following partition $\mathcal{S}$ when A and B are arbitrary events. In each case give a brief reason.

a) the four events A, A^c, B, B^c,

b) the two events $A, B \setminus A$,

c) the four events $A \setminus B, B \setminus A, A \cap B, (A \cup B)^c$,

d) the three events $A \cap B, A \triangle B, A^c \cap B^c$,

e) the three events $A, B, (A \cup B)^c$.

Problem 17: Mimi and Rodolfo are looking for a key in the dark. Suppose that the key may be under the table, behind the bookshelf or in the corridor and has a $1/3$ chance of being in each of these places. Mimi searches under the table; if the key is there she has a $3/5$ chance of finding it. Rodolfo searches behind the bookshelf; if the key is there he has a $1/5$ chance of finding it.

a) Calculate the probability that the key is found.

b) Suppose that the key is found. Calculate the conditional probability that it is found by Rodolfo.

c) Suppose that the key is not found. Calculate the conditional probability that it is in the corridor.

Problem 18: A lawyer for the prosecution in a murder trial observes that a suspect's fingerprints match those found at the crime scene and that there is a 1 in 50,000 chance of such a match occurring by chance. They argue that this means there is a 1 in 50,000 chance of the suspect being innocent.

a) Why is such an argument flawed?

b) Suppose that the city has a population of 1 million and the murderer is assumed to be one of these. If there is no evidence against the suspect apart from the fingerprint match then it is reasonable to regard the suspect as a randomly chosen citizen. Under this assumption what is the probability that they are innocent?

c) How does the argument change if one knows that the suspect is among the 100 acquaintances of the victim.

§7 Random Variables

Example 7.1 *We roll two fair dice and record the sum. The outcome of each experiment is a pair* (k, j) *with* k *and* j *being an integer from the set* $\{1, 2, 3, 4, 5, 6\}$ *and the sample space is the set of such pairs*

$$\mathcal{S} = \{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\} = \{(j, k) : j, k \in \{1, 2, 3, 4, 5, 6\}\}$$

Recording the sum means that for an outcome (j, k) *we record the value* $k+j$*. This recipe* $(j, k) \mapsto k+j$ *is a function* X *taking inputs from the sample space* $\mathcal{S}$ *and giving an integer. Such a function is called a* random variable *. For instance*

$$X((2, 5)) = 7, \quad X((3, 3)) = 6$$

Note that the function X *takes as input a pair (an element from the sample space !) and not just two numbers.*

Definition 7.1 *A* random variable *is a function from* $\mathcal{S}$ *to* $\mathbb{R}$*.*

Remark: If $\mathcal{S}$ is uncountable then this definition is not quite correct. It turns out that some functions are too complicated to regard as random variables (just as some sets are too complicated to regard as events). This subtlety is well beyond the scope of this module and will not concern us at all.

Example 7.2 *Toss a fair coin three times (see example 1.2 for the sample space* $\mathcal{S}$*). Let* X *denote the number of heads seen.* X *is a random variable which takes values* $0, 1, 2, 3$*. For instance*

$$X(hhh) = 3, \quad X(hth) = 2$$

Let Y *be the number of tails seen.* Y *is another random variable for the same experiment. E.g.* $Y(htt) = 2$*.* $Z = \max\{X, Y\}$ *is another random variable, e.g.,*

$$Z(hht) = \max\{X(hht), Y(hht)\} = \max\{2, 1\} = 2$$

Let X be a random variable which takes values $x_1, x_2, \ldots$. We will denote by $X = x_k$ the event (i.e. a collection of outcomes, a subset of the sample space) such that all outcomes

ω of the event (all elements of the subset) give the value $X(\omega) = x_k$, i.e. $X = x_k$ denotes the set

$$\{\omega \in \mathcal{S} : X(\omega) = x_k\}$$

Example 7.3 *Consider the experiment and the random variable of example 7.1. The event $X = 5$ (two dice showing 5 in total) is*

$$\{(1,4),(2,3),(3,2),(4,1)\}\,.$$

Similarly $X \leq x_k$ denotes the event

$$\{\omega \in \mathcal{S} : X(\omega) \leq x_k\}\,.$$

Example 7.4 *Consider the setup of example 7.1. Then $X \leq 3$ denotes the event*

$$\{(1,1),(1,2),(2,1)\}\,.$$

If Y is another random variable $X \leq Y + 3$ denotes the event

$$\{\omega \in \mathcal{S} : X(\omega) \leq Y(\omega) + 3\}$$

and so on.

Example 7.5 *Consider the experiment and the random variable of example 7.2. $X > Y$ denotes the event that the number of heads seen is larger than the number of tails, i.e.,*

$$\{hhh, hht, hth, thh\}\,.$$

Assuming that the coin is fair the probability of this event is given by

$$\mathbb{P}(X > Y) = \frac{1}{2}$$

Remark: Don't confuse events and random variables. If X is a random variable then $\mathbb{P}(X = 2)$ makes sense as $X = 2$ is an event (subset of the sample space). But $\mathbb{P}(X)$ is meaningless as X is not an event (X is a function, not a set).

Definition 7.2 *The* probability mass function *(pmf) of a random variable X is the function which given x_k has output $\mathbb{P}(X = x_k)$*

$$x_k \mapsto \mathbb{P}(X = x_k)\,.$$

Example 7.6 *Consider example 7.1. Since outputs are picked at random and $|\mathcal{S}| = 36$*

$$\begin{aligned}
\mathbb{P}(X=2) &= \frac{|\{(1,1)\}|}{36} = \frac{1}{36} \\
\mathbb{P}(X=3) &= \frac{|\{(1,2),(2,1)\}|}{36} = \frac{2}{36} = \frac{1}{18} \\
\mathbb{P}(X=4) &= \frac{|\{(1,3),(2,2),(3,1)\}|}{36} = \frac{3}{36} = \frac{1}{12} \\
&\vdots \\
\mathbb{P}(X=11) &= \frac{|\{(5,6),(6,5)\}|}{36} = \frac{2}{36} \\
\mathbb{P}(X=12) &= \frac{|\{(6,6)\}|}{36} = \frac{1}{36}
\end{aligned}$$

and we can summarise these results in a table (the mass probability function)

x_k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = x_k)$	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

Example 7.7 *Toss a fair coin until you see head. Denote the sample space by $\mathcal{S} = \{h, th, tth, ttth, \ldots\}$ (see as well example 1.3). Let T be the number of tosses. T is a random variable, e.g. $T(tth) = 3$, with values in $\mathbb{N}$. If $n \in \mathbb{N}$ then $P(T = n) = 1/2^n$, using the probability for a fair coin and independence of subsequent tosses (e.g. $T = 4$ is the simple event $\{ttth\}$ and $\mathbb{P}(\{ttth\}) = (1/2)^4 = 1/2^4$). Probability mass function*

n	1	2	3	4	...
$\mathbb{P}(T = n)$	1/2	1/4	1/8	1/16	...

Definition 7.3 *A random variable X is* discrete *if the set of values that X takes is either finite or countably infinite.*

All our examples so far are discrete random variables, e.g., in example 7.6 with a finite set of values, or in example 7.7 with a countable set of values.

Example 7.8 *Probabilities in the mass distribution function add up to one.*

- *In example 7.6*

$$\begin{aligned}\sum_{k=1}^{12}\mathbb{P}(X=k) &= \mathbb{P}(X=1)+\mathbb{P}(X=2)+\ldots+\mathbb{P}(X=12)\\ &= \frac{1}{36}+\frac{1}{18}+\frac{1}{12}+\frac{1}{9}+\frac{5}{36}+\frac{1}{6}+\frac{5}{36}+\frac{1}{9}+\frac{1}{12}+\frac{1}{18}+\frac{1}{36}=1\,.\end{aligned}$$

- *In example 7.7*

$$\begin{aligned}\sum_{k=1}^{\infty}\mathbb{P}(T=k) &= \mathbb{P}(T=1)+\mathbb{P}(T=2)+\mathbb{P}(T=3)+\ldots\\ &= \frac{1}{2}+\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\ldots\\ &= \frac{1}{2}\left(1++\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\ldots\right)\\ &= \frac{1/2}{1-1/2}=1\end{aligned}$$

where we have used the formula for the geometric series (see problem 21b, sheet 7)

$$\sum_{k=0}^{\infty} z^k = 1+z+z^2+z^3+\ldots=\frac{1}{1-z},\quad |z|<1$$

with $z=1/2$*.*

Such a property holds for any discrete random variable.

Lemma 7.1 *If* X *is a discrete random variable with values* $x_1, x_2, \ldots$ *then*

$$\sum_k \mathbb{P}(X=x_k)=\mathbb{P}(X=x_1)+\mathbb{P}(X=x_2)+\ldots=1$$

where the sum is over all values X *takes (either finite or infinite).*

Proof: Suppose that X takes the values x_k. Let A_k be the event $X=x_k$. The events are pairwise disjoint (for $i\neq j$ we have $A_i\cap A_j=\emptyset$; otherwise $A_i\cap A_j$ would contain an element, say ω and by definition $X(\omega)=x_i$ and $X(\omega)=x_j$, violating the assumption

$i \neq j$). Furthermore $A_1 \cup A_2 \cup \ldots = \mathcal{S}$ (for any $\omega \in \mathcal{S}$ $X(\omega)$ takes a value, say $X(\omega) = x_i$, so $\omega \in A_i$). By definition 2.1c

$$\begin{aligned}
\sum_k \mathbb{P}(X = x_k) &= \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \ldots \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots = \mathbb{P}(A_1 \cup A_2 \cup \ldots) = \mathbb{P}(\mathcal{S}) = 1
\end{aligned}$$

$\square$.

The values the random variable X takes will be normally denoted by x_k. When we write a $\sum_k$ then the summation is over all values that the random variable takes, either finite of infinite.

Exercise Sheet 7

This sheet is based on the material lectured in week 8. The questions are about random variables. The sheet will be discussed in tutorials in week 9.

Hand in your homework, problem G, to the tutor at the beginning of the tutorial, and collect problem F from your tutor to complete your submission.

Questions 19, 20, and 21 will be discussed in the tutorial.

Problem G: (Homework) A coin which has probability p of coming up heads is tossed three times. Let X be the number of heads observed.

a) State the range of the function X, i.e., list the values which the random variables take.

b) Compute the probability mass function of the random variable X, i.e., compute the probabilities $\mathbb{P}(X = x_k)$ where x_k is any of the values in the range of the function X.

c) Confirm the statement in lemma 7.1, i.e., show that

$$\sum_k \mathbb{P}(X = x_k) = 1$$

for the special random variable considered in this problem.

Problem 19: A random variable X has the following probability mass function:

k	-2	-1	0	1	2
$P(X = k)$	$1/10$	$2/5$	$1/4$	$1/5$	$1/20$

a) Calculate the probability of each of the following events:

i) $X = 2$; ii) $X = 3$; iii) $X \leq 1$; iv) $X^2 < 2$.

b) Let Y be a new random variable defined by $Y = X^2 + 4$.

i) What values does Y take?

ii) Find the probability mass function of Y.

Problem 20: A bag contains 6 red marbles and 2 blue marbles. I choose 5 at random without replacement. Let B be the number of blue marbles I end up with and R be the number of red marbles I end up with. Find the probability mass function of B. Without doing any more calculations write down the probability mass function of R.

Problem 21:

a) Suppose $z \neq 1$ is a real number and n a positive integer. Compute the so called geometric sum

$$S_n = 1 + z + z^2 + z^3 + \ldots + z^{n-1} = \sum_{k=0}^{n-1} z^k .$$

Hint: For instance take the difference of S_n and zS_n and show that this difference just contains two terms, namely $1 - z^n$. Hence conclude that

$$S_n = \frac{1 - z^n}{1 - z} .$$

b) Assume that $|z| < 1$. By taking the limit $n \to \infty$ derive the expression for the so called geometric series

$$1 + z + z^2 + z^3 + \ldots = \sum_{k=0}^{\infty} z^k = \frac{1}{1 - z} .$$

§8 Expectation and Variance

Suppose you toss 10 fair coins and you count the number of heads. Let X be the random variable which gives the number of heads. How many heads do you see "on average". Intuition tells us that this value is at about 5, and that a "typical" outcome differs from this value little (say by 1 or 2 at most).

Expectation and variance of the random variable X quantifies these concepts. The expectation of X is the value we are expected to see in most outcomes and the variance of X (to be precise: the square root of the variance) gives us a range by which typical experiments may differ from the expected value.

Definition 8.1 *If X is a discrete random variable which takes values x_k then the* expectation *of X (or the expected value of X), called $E(X)$, is defined by*

$$\begin{aligned} E(X) &= \sum_k x_k \mathbb{P}(X = x_k) \\ &= x_1 \mathbb{P}(X = x_1) + x_2 \mathbb{P}(X = x_2) + x_3 \mathbb{P}(X = x_3) + \ldots \end{aligned}$$

The variance *of X, called $Var(X)$, is defined by*

$$\begin{aligned} Var(X) &= \sum_k [x_k - E(X)]^2 \mathbb{P}(X = x_k) \\ &= [x_1 - E(X)]^2 \mathbb{P}(X = x_1) + [x_2 - E(X)]^2 \mathbb{P}(X = x_2) + [x_3 - E(X)]^2 \mathbb{P}(X = x_3) + \ldots \end{aligned}$$

Remark: The idea is that the expectation is the average value X takes. The variance measures how sharply concentrated X is about E(X), with a small variance meaning sharply concentrated and a large variance meaning spread out.

Example 8.1 *Calculate the expectation and the variance for the random variable in example 7.6 (roll two dice and record the sum).*

- *Expectation*

$$\begin{aligned} E(X) &= 2 \times \mathbb{P}(X = 2) + 3 \times \mathbb{P}(X = 3) + \ldots + 12 \times \mathbb{P}(X = 12) \\ &= 2 \times \frac{1}{36} + 3 \times \frac{1}{18} + 4 \times \frac{1}{12} + 5 \times \frac{1}{9} + 6 \times \frac{5}{36} + 7 \times \frac{1}{6} \\ &\quad + 8 \times \frac{5}{36} + 9 \times \frac{1}{9} + 10 \times \frac{1}{12} + 11 \times \frac{1}{18} + 12 \times \frac{1}{36} \\ &= 7 \end{aligned}$$

- *Variance*

$$\begin{aligned} Var(X) &= (2-7)^2\mathbb{P}(X=1) + (3-7)^2\mathbb{P}(X=3) + \ldots + (12-7)^2\mathbb{P}(X=12) \\ &= 5^2 \times \frac{2}{36} + 4^2 \times \frac{2}{18} + 3^2 \times \frac{2}{12} + 2^2 \times \frac{2}{9} + 1^2 \times \frac{10}{36} \\ &= \frac{35}{6} \end{aligned}$$

Example 8.2 *Calculate the expectation of the random variable in example 7.7 (toss a coin until you see head)*

$$\begin{aligned} E(T) &= 1\mathbb{P}(T=1) + 2\mathbb{P}(T=2) + 3\mathbb{P}(T=3) + 4\mathbb{P}(T=4) + \ldots \\ &= \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + \ldots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots \\ &\quad + \frac{1}{4} + 2 \times \frac{1}{8} + 3 \times \frac{1}{16} + \ldots \\ &= 1 + \frac{1}{2}E(T) \end{aligned}$$

Hence $E(T) = 2$.

Remark: If f is a function and $f(X)$ denotes a new random variable then the expectation of $f(X)$ is given by

$$\mathrm{E}(f(X)) = \sum_k f(x_k)\mathbb{P}(X=x_k) = f(x_1)\mathbb{P}(X=x_1) + f(x_2)\mathbb{P}(X=x_2) + \ldots$$

Proposition 8.1 *If X is a discrete random variable then*

$$\begin{aligned} Var(X) &= \sum_k (x_k)^2\mathbb{P}(X=x_k) - [E(X)]^2 \\ &= \big((x_1)^2\mathbb{P}(X=x_1) + (x_2)^2\mathbb{P}(X=x_2) + \ldots\big) - [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Proof: Using $[x_k - \mathrm{E}(X)]^2 = (x_k)^2 - 2\mathrm{E}(X)x_k + [\mathrm{E}(X)]^2$ definition 8.1 tells us

$$\mathrm{Var}(X) = \sum_k (x_k)^2\mathbb{P}(X=x_k) - 2\mathrm{E}(X)\sum x_k\mathbb{P}(X=x_k) + [\mathrm{E}(X)]^2\sum_k \mathbb{P}(X=x_k)$$

If we use definition 8.1 for the second term, and lemma 7.1 for the third term, we arrive at

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_k (x_k)^2 \mathbb{P}(X = x_k) - 2\mathrm{E}(X)\mathrm{E}(X) + [\mathrm{E}(X)]^2 \times 1 \\
&= \sum_k (x_k)^2 \mathbb{P}(X = x_k) - [\mathrm{E}(X)]^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 .
\end{aligned}$$

$\square$.

Proposition 8.2 (Properties of expectation) *Let X be a discrete random variable and $c \in \mathbb{R}$ be a constant.*

i) $E(c) = c$

ii) $E(X + c) = E(X) + c$

iii) $E(cX) = cE(X)$

iv) If $m \leq X(\omega) \leq M$ for all $\omega \in \mathcal{S}$ then

$$m \leq E(X) \leq M.$$

Proof:

i) Denote by $X = c$ the constant random variable (which takes the single value $x_1 = c$), so that $\mathbb{P}(X = c) = 1$. Then by definition 8.1

$$\mathrm{E}(c) = c\mathbb{P}(X = c) = c$$

ii)

$$\begin{aligned}
\mathrm{E}(X + c) &= \sum_k (x_k + c)\mathbb{P}(X = x_k) \\
&= \sum_k x_k \mathbb{P}(X = x_k) + \sum_k c\mathbb{P}(X = x_k) \\
&= \mathrm{E}(X) + c \times 1 \text{ using definition 8.1 and lemma 7.1}
\end{aligned}$$

iii)

$$\begin{aligned}\mathrm{E}(cX) &= \sum_k c x_k \mathbb{P}(X = x_k) \\ &= c\sum_k x_k \mathbb{P}(X = x_k) \\ &= c\mathrm{E}(X)\end{aligned}$$

iv) Since every value that X takes is lower or equal to M, $x_k \leq M$, we have that $x_k\mathbb{P}(X = x_k) \leq M\mathbb{P}(X = x_k)$ (recall probabilities are not negative !) and

$$\mathrm{E}(X) = \sum_k x_k \mathbb{P}(X = x_k) \leq \sum_k M\mathbb{P}(X = x_k) = M\sum_k \mathbb{P}(X = x_k) = M.$$

Similarly, since every value that X takes is greater or equal to m, $x_k \geq m$, we have that

$$\mathrm{E}(X) = \sum_k x_k \mathbb{P}(X = x_k) \geq m\sum_k \mathbb{P}(X = x_k) = m.$$

$\square$.

Proposition 8.3 (Properties of variance) *Let X be a discrete random variable and $c \in \mathbb{R}$ be a constant.*

i) $\mathit{Var}(X) \geq 0$

ii) $\mathit{Var}(c) = 0$

iii) $\mathit{Var}(cX) = c^2\,\mathit{Var}(X)$

iv) $\mathit{Var}(X + c) = \mathit{Var}(X)$

Proof: For most of these we can either use the definition 8.1 of variance or proposition 8.1. However one or other of these approaches may be easier.

i)

$$\mathrm{Var}(X) = \sum_k (x_k - \mathrm{E}(X))^2\mathbb{P}(X = x_k) \geq 0$$

since the square of any real number is non-negative and probabilities are all non-negative.

ii) Since $\mathrm{E}(c) = c$, see proposition 8.2

$$\mathrm{Var}(c) = (c - c)^2 \times 1 = 0.$$

iii) By proposition 8.1

$$\begin{aligned}
\mathrm{Var}(cX) &= \mathrm{E}((cX)^2) - (\mathrm{E}(cX))^2 \\
&= c^2\mathrm{E}(X^2) - (c\mathrm{E}(X))^2 \quad \text{(By proposition 8.2iii)} \\
&= c^2(\mathrm{E}(X^2) - (\mathrm{E}(X))^2) \\
&= c^2\mathrm{Var}(X).
\end{aligned}$$

iv) By proposition 8.1

$$\begin{aligned}
\mathrm{Var}(X + c) &= \mathrm{E}((X + c)^2) - (\mathrm{E}(X + c))^2 \\
&= \mathrm{E}(X^2 + 2cX + c^2) - (\mathrm{E}(X) + c)^2 \\
&= \mathrm{E}(X^2) + 2c\mathrm{E}(X) + c^2 - ((\mathrm{E}(X))^2 + 2c\mathrm{E}(X) + c^2) \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \\
&= \mathrm{Var}(X)
\end{aligned}$$

$\square$.

Example 8.3 *Consider the setup in example 7.1 (roll two dice). If $X = k + j$ denotes the sum of the roll, and $Y = (k + j)/2$ the mean value of the rolls then example 8.1 tells us*

$$E(Y) = E(X/2) = \frac{E(X)}{2} = \frac{7}{2}$$

and

$$Var(Y) = Var(X/2) = \frac{Var(X)}{4} = \frac{35}{24}$$

Exercise Sheet 8

This sheet is based on the material lectured in week 9. The questions are about expectation and variance. The sheet will be discussed in tutorials in week 10.

Hand in your homework, problem H, to the tutor at the beginning of the tutorial, and collect problem G from your tutor to complete your submission.

Questions 22, 23, and 24 will be discussed in the tutorial.

Problem H: (Homework) Let X be a discrete random variable with $\mathrm{E}(X) = 5$, $\mathrm{Var}(X) = 2/3$. Find the following:

i) $\mathrm{E}(3X)$; ii) $\mathrm{Var}(3X)$; iii) $\mathrm{E}(4-3X)$; iv) $\mathrm{Var}(4-3X)$; v) $\mathrm{E}(4-3X^2)$.

Problem 22: A coin which has probability p of coming up heads is tossed three times. Let X be the number of heads observed (see problem G, sheet 7).

a) Compute the expectation of X.

b) Compute the variance of X.

Problem 23: Suppose $|z| < 1$. Using the geometric series, or otherwise, compute the following expressions:

a)

$$\sum_{k=1}^{\infty} k z^{k-1} = 1 + 2z + 3z^2 + 4z^3 + \dots .$$

b)

$$\sum_{k=1}^{\infty} k^2 z^{k-1} = 1 + 4z + 9z^2 + 16z^3 + \dots .$$

Problem 24: Let A be a random variable taking values in the set $\{0, 1, 2, 3, \dots, n\}$.

a) Show that

$$\mathrm{E}(A) = \sum_{i=1}^{n} \mathbb{P}(A \geq i).$$

b) Deduce that if $\mathrm{E}(A) < 1$ then A takes the value 0 with positive probability.

c) Use part a) to prove that for all $1 \leq t \leq n$ we have

$$\mathbb{P}(A \geq t) \leq \frac{\mathrm{E}(A)}{t}.$$

§9 Special Discrete Random Variables

Certain probability mass functions occur so often that it is convenient to give them special names. In this section we study a few of these.

a) Bernoulli distribution

Bernoulli(p) trial: An experiment with the two outcomes called "success" and "failure" with $\mathbb{P}(\{\text{success}\}) = p$ is called a *Bernoulli*(p) *trial*.

Bernoulli(p) distribution: The random variable X, with values $X(\text{success}) = 1$ and $X(\text{failure}) = 0$, has the probability mass function

n	0	1
$P(X = n)$	$1 - p$	p

which we call the Bernoulli(p) distribution. We write $X \sim \text{Bernoulli}(p)$.

Expectation and variance:

$$\begin{aligned} \mathrm{E}(X) &= 0 \times (1-p) + 1 \times p = p \\ \mathrm{Var}(X) &= (0-p)^2 \times (1-p) + (1-p)^2 p = p(1-p)(p + (1-p)) = p(1-p) \end{aligned}$$

Remark: If a (biased) coin with probability p coming up "head" is tossed, and if X is the random variable $X(\text{head}) = 1$ and $X(\text{tail}) = 0$, then the pmf of X is given by the Bernoulli(p) distribution. We use the shorthand notation $X \sim \text{Bernoulli}(p)$.

b) Binomial distribution

Perform n independent Bernoulli trials (i.e. if E_i denotes the event "ith trial is a success" then $E_1, E_2, \ldots, E_n$ are mutually independent events). Let X be the number of successes. To determine the pmf of X we need to evaluate the probabilities $\mathbb{P}(X = k)$ for $k = 0, 1, 2, \ldots, n$:

- First consider the special outcome that the first k trials are "success" and the remaining $n - k$ trials are "failure". Using mutual independence the probability of this simple event is $p^k(1-p)^{n-k}$.

- The event $X = k$ contains $\binom{n}{k}$ different outcomes with k "successes" and $n-k$ "failures" nd each outcome occurs with the same probability $p^k(1-p)^{n-k}$.

Hence we obtain the pmf

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

We say X has the Binomial(n,p) (Bin(n,p)) distribution, $X \sim \text{Bin}(n,p)$.

Remark: Some useful identities:

- Binomial theorem
$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

- $$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \times (n-1)!}{k \times (k-1)!(n-k)!} = \frac{n}{k}\binom{n-1}{k-1}$$

Expectation: Using the just mentioned identities

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{n-k} = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = np\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell}(1-p)^{n-1-\ell} \\
&= np(p+(1-p))^{n-1} = np\,.
\end{aligned}$$

Variance: Apply the aforementioned identity twice, i.e.,

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} = \frac{n}{k}\frac{n-1}{k-1}\binom{n-2}{k-2}\,.$$

Now

$$\begin{aligned}
\mathrm{E}(X^2) - \mathrm{E}(X) &= \mathrm{E}(X(X-1)) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^k(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^{\ell}(1-p)^{n-2-\ell} \\
&= n(n-1)p^2(p+(1-p))^{n-2} = n(n-1)p^2\,.
\end{aligned}$$

So

$$E(X^2) = n(n-1)p^2 + E(X) = n(n-1)p^2 + np$$

and

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p)\,.$$

Remark: If a (biased) coin with probability p coming up "head" is tossed n times, and if X denotes the random variable counting the numbers of heads then $X \sim \mathrm{Bin}(n, p)$.

Example 9.1 *A bag contains N balls M of which are red. You choose n balls with replacement. Each random pick will result in a red ball with probability $\frac{M}{N}$ and the outcome of each pick is independent of all the others (i.e. we perform n independent Bernoulli trials with $p = M/N$).*

Let R denote (the random variable counting) the number of red balls. Then $R \sim Bin(n, M/N)$ and

$$E(R) = n\frac{M}{N}, \quad Var(R) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\,.$$

c) Geometric distribution

Suppose we make an unlimited number of independent Bernoulli(p) trials and let T be the number of trials up to and including the first success, see example 7.7. $T = k$ consists of a single event with $k-1$ "failures", each of which with probability $1-p$, and one success with probability p. Hence

$$P(T = k) = (1-p)^{k-1}p$$

We say T has the *geometric distribution*, $T \sim \mathrm{Geom}(p)$.

Remark: Some useful identities

- (see problem 23a), sheet 8)

$$\sum_{k=1}^{\infty} kz^{k-1} = 1 + 2z + 3z^2 + 4z^3 + \ldots = \frac{1}{(1-z)^2} \quad (*)$$

- (see problem 23b), sheet 8)

$$\sum_{k=1}^{\infty} k^2 z^{k-1} = 1 + 4z + 9z^2 + 16z^3 + \ldots = \frac{2}{(1-z)^3} - \frac{1}{(1-z)^2} \qquad (**)$$

Expectation:

$$\begin{aligned}
\mathrm{E}(T) &= \sum_{k=1}^{\infty} kP(T=k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= p\left(1 + 2(1-p) + 3(1-p)^2 + 4(1-p)^3 + \ldots\right) \\
&= p\frac{1}{[1-(1-p)]^2} \text{ using (*)} \\
&= \frac{1}{p}
\end{aligned}$$

Variance: Recall $\mathrm{Var}(T) = \mathrm{E}(T^2) - [\mathrm{E}(T)]^2$

$$\begin{aligned}
\mathrm{E}(T^2) &= \sum_{k=1}^{\infty} k^2 P(T=k) = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p \\
&= p\left(1 + 4(1-p) + 9(1-p)^2 + 16(1-p)^3 + \ldots\right) \\
&= p\left(\frac{2}{[1-(1-p)]^3} - \frac{1}{[1-(1-p)]^2}\right) \text{ using (**)} \\
&= \frac{2}{p^2} - \frac{1}{p}
\end{aligned}$$

Therefore

$$\mathrm{Var}(T) = \mathrm{E}(T^2) - [\mathrm{E}(T)]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Example 9.2 *A biased coin has probability p to show head. We toss the biased coin until we see head for the first time. Let T denote the random variable counting the number of coin tosses. $T \sim Geom(p)$ and $\mathbb{P}(T = k)$ is the probability that we perform exactly k coin tosses.*

$\mathbb{P}(T > k)$ is the probability that we perform more than k coin tosses (i.e. that the first k tosses are all tail). Obviously (recall $T > k$ and $T = \ell$ are events/sets)

$$(T > k) = (T = k+1) \cup (T = k+2) \cup (T = k+3) \cup \ldots$$

and since all the simple events are disjoint

$$\begin{aligned}\mathbb{P}(T > k) &= \mathbb{P}(T = k+1) + \mathbb{P}(T = k+2) + \mathbb{P}(T = k+3) + \dots \\ &= (1-p)^k p + (1-p)^{k+1} p + (1-p)^{k+2} p + \dots \\ &= (1-p)^k p \left(1 + (1-p) + (1-p)^2 + (1-p)^3 + \dots\right) \\ &= (1-p)^k p \frac{1}{1-(1-p)} = (1-p)^k .\end{aligned}$$

The function $k \mapsto \mathbb{P}(T > k)$ is called cumulative distribution function.

d) Poisson distribution

Consider a random variable X which takes values $k = 0, 1, 2, 3, \dots$. Denote by $\lambda > 0$ a fixed positive real number. Define the pmf of the random variable X by

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} .$$

We say that X has the Poisson(λ) distribution, in short $X \sim$ Poisson(λ).

Remark:

- There is no simple way to illustrate the Poisson(λ) distribution. For instance, if X counts the number of emissions in radioactive decay in a fixed time span then $X \sim$ Poisson(λ).
- If $0 < \lambda \leq 1$ then the pmf $\mathbb{P}(X = k)$ is monotonic decreasing. If $\lambda > 1$ then the pmf has a maximum at a finite value $k > 0$.

Remark: Some useful identity: the Taylor series of the exponential function

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots$$

Using the Taylor series of the exponential function we have

$$\sum_{k=0}^{\infty} \mathbb{P}(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

see lemma 7.1.

Expectation:

$$\mathrm{E}(X) = \sum_{k=0}^{\infty} k\mathbb{P}(X=k) = \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda \times \lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^\ell}{\ell!} = \lambda$$

Variance: To compute the variance consider again

$$\begin{aligned}\mathrm{E}(X^2) - \mathrm{E}(X) &= \mathrm{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1)\mathbb{P}(X=k) = \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} \\ &= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^2 \times \lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^\ell}{\ell!} = \lambda^2\end{aligned}$$

Hence $\mathrm{E}(X^2) = \lambda^2 + \lambda$ and

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

e) Summary

	range	pmf $\mathbb{P}(X=k)$	$\mathrm{E}(X)$	$\mathrm{Var}(X)$
$X \sim \mathrm{Bin}(n,p)$	$X = 0, 1, \ldots, n$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
$X \sim \mathrm{Geom}(p)$	$X = 1, 2, 3, \ldots$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$X \sim \mathrm{Poisson}(\lambda)$	$X = 0, 1, 2, \ldots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ

Exercise Sheet 9

This sheet is based on the material lectured in week 10. The questions are about special discrete random variables. The sheet will be discussed in tutorials in week 11.

Hand in your homework, problem I, to the tutor at the beginning of the tutorial, and collect problem H from your tutor to complete your submission.

Questions 25, 26, and 27 will be discussed in the tutorial.

Problem I: (Homework) Suppose that $A \sim \text{Poisson}(3)$, $B \sim \text{Geom}(1/3)$, and that $C \sim \text{Bin}(4, 1/6)$ are random variables. Find the following probabilities. (You should leave your answers involving powers of e but do simplify all factorials and other powers.)

i) $\mathbb{P}(A = 2)$; ii) $\mathbb{P}(A > 2)$; iii) $\mathbb{P}(B = 3)$; iv) $\mathbb{P}(B \leq 3)$; v) $\mathbb{P}(C = 2)$.

Problem 25: An electrical component is installed on a certain day and is inspected on each subsequent day. Let G be the number of days that the component lasts for before breaking.

a) Suppose that $G \sim \text{Geom}(p)$. Show that for any $k, \ell \geq 0$

$$\mathbb{P}(G > k + \ell | G > k) = \mathbb{P}(G > \ell).$$

b) Say in words what the conclusion of part a) means for the lifetime of the component. Why do you think this is sometimes called the "memoryless property" of the geometric distribution?

c) Show that if a random variable H (which takes positive integer values) satisfies $\mathbb{P}(H > k + \ell | H > k) = \mathbb{P}(H > \ell)$ for all $k, \ell \geq 0$ then H must have a geometric distribution.

Problem 26: Let X be the number of fish caught by a fisherman in one afternoon. Suppose that X is distributed $\text{Poisson}(\lambda)$. Each fish has probability p of being a salmon independently of all other fish caught. Let Y be the number of salmon caught.

a) Suppose that the fisherman catches m fish. What is the probability that k of them are salmon?

b) Show that:

$$\mathbb{P}(Y = k) = \sum_{m \geq k} \mathbb{P}(Y = k|X = m)\mathbb{P}(X = m).$$

c) Hence find the probability mass function of Y. What is the name of the distribution of Y?

Problem 27: A fair coin is tossed four times. Let N be the number of instances of a head followed by another head in the sequence of tosses.

a) Professor Habakuk Tibatong from the Oxbridge Institute of Random Knowledge proposes the following solution: There are three possible ways in which we could have a head followed by another head (at the first and second, the second and third, or third and fourth toss). We have a probability $1/2 \times 1/2 = 1/4$ of getting a head followed by another head at each of these positions. Hence N is the number of successes in three Bernoulli$(1/4)$ trials and so $N \sim \text{Bin}(3, 1/4)$.

Explain what is wrong with this argument. Justify your reason carefully.

b) Determine the correct probability mass function of N.

c) Calculate the expectation and variance of N.

§10 Several Random Variables

Example 10.1 *A bag contains* 3 *red balls,* 2 *blue balls, and* 2 *green balls. We pick* 3 *balls at random. Let R denote the number of red balls, and B the number of blue balls. R and B are two random variables defined on the same sample space.*

In this section we will consider random variables defined on the same sample space.

Definition 10.1 *Denote by X and Y two discrete random variable defined on the same sample space. Assume X and Y take values $(x_1, x_2, \ldots)$ and $(y_1, y_2, \ldots)$, respectively. The function*

$$(x_k, y_\ell) \mapsto \mathbb{P}(X = x_k \cap Y = y_\ell)$$

is called the joint probability mass function *of X and Y.*

Remark: Often one writes $\mathbb{P}(X = x_k, Y = y_\ell)$ instead of $\mathbb{P}(X = x_k \cap Y = y_\ell)$.

Example 10.2 *Consider the experiment of example 10.1. Use unordered sampling without replacement, section 3c, to work out probabilities.*

$|\mathcal{S}| = \binom{7}{3} = 35$*. To work out, e.g., $\mathbb{P}(R = 1, B = 1)$ recall that we sample a red ball from a subset of* 3*, a blue ball from a subset of* 2 *and a green ball from a subset of* 2 *giving* $\binom{3}{1}\binom{2}{1}\binom{2}{1}$ *possibilities. Hence $\mathbb{P}(R = 1, B = 1) = 12/35$. The other probabilities can be worked out in the same way and are summarised in a table.*

B \ R	0	1	2	3
0	0	3/35	6/35	1/35
1	2/35	12/35	6/35	0
2	2/35	3/35	0	0

The joint pmf satisfies (see as well lemma 7.1)

$$\sum_k \sum_\ell \mathbb{P}(X = x_k, Y = y_\ell) = 1\,.$$

The joint pmf and the pmf for X and Y, the so called *marginal distributions*, are related by

$$\mathbb{P}(X = x_k) = \sum_{\ell} \mathbb{P}(X = x_k, Y = y_\ell), \quad \mathbb{P}(Y = y_\ell) = \sum_{k} \mathbb{P}(X = x_k, Y = y_\ell)\,.$$

Example 10.3 *Consider the joint pmf of example 10.2.*

Clearly the entries in the table add up to one.

Marginal distributions: column sum

k	0	1	2	3
$\mathbb{P}(R = k)$	$4/35$	$18/35$	$12/35$	$1/35$

row sum

ℓ	0	1	2
$\mathbb{P}(B = \ell)$	$10/35$	$20/35$	$5/35$

Expectations:

$$E(R) = \sum_{k=0}^{3} k\mathbb{P}(R = k) = \frac{0 \times 4 + 1 \times 18 + 2 \times 12 + 3 \times 1}{35} = \frac{9}{7}$$

$$E(B) = \sum_{\ell=0}^{2} \ell\mathbb{P}(B = \ell) = \frac{0 \times 10 + 1 \times 20 + 2 \times 5}{35} = \frac{6}{7}$$

Expectations are defined and can be computed in the obvious way (see definition 8.1).

Definition 10.2 *If $g(X, Y)$ is a (real valued) function of the two discrete random variables X and Y then*

$$E(g(X, Y)) = \sum_{k} \sum_{\ell} g(x_k, y_\ell)\mathbb{P}(X = x_k, Y = y_\ell)$$

defines the expectation of $g(X, Y)$.

Example 10.4 *Consider the experiment in example 10.1. $R + B$ is the number of red and blue balls, $R \times B$ is the product of the number of red and blue balls. Using the joint probability mass function of example 10.2 we have*

$$E(R+B) = 0 \times 0 + 1 \times \left(\frac{3}{35} + \frac{2}{35}\right) + 2 \times \left(\frac{6}{35} + \frac{12}{35} + \frac{2}{35}\right) + 3 \times \left(\frac{1}{35} + \frac{6}{35} + \frac{3}{35}\right) = \frac{15}{7}\,.$$

Note that $E(R+B) = E(R) + E(B)$. *Furthermore*

$$E(R \times B) = 1 \times \frac{12}{35} + 2 \times \left(\frac{6}{35} + \frac{3}{35}\right) = \frac{6}{7}\,.$$

Note that $E(R \times B) \neq E(R) \times E(B)$.

Theorem 10.1 *If* X *and* Y *are discrete random variables then*

$$E(X+Y) = E(X) + E(Y)\,.$$

Proof:

$$\begin{aligned}
\mathrm{E}(X+Y) &= \sum_k \sum_\ell (x_k + y_\ell)\mathbb{P}(X = x_k, Y = y_\ell) \\
&= \sum_k \sum_\ell x_k \mathbb{P}(X = x_k, Y = y_\ell) + \sum_k \sum_\ell y_\ell \mathbb{P}(X = x_k, Y = y_\ell) \\
&= \sum_k x_k \left(\sum_\ell \mathbb{P}(X = x_k, Y = y_\ell)\right) + \sum_\ell y_\ell \left(\sum_k \mathbb{P}(X = x_k, Y = y_\ell)\right) \\
&= \sum_k x_k \mathbb{P}(X = x_k) + \sum_\ell y_\ell \mathbb{P}(Y = y_\ell) = \mathrm{E}(X) + \mathrm{E}(Y)
\end{aligned}$$

$\square$.

Applying theorem 10.1 repeatedly, that means, $\mathrm{E}(X+Y+Z) = \mathrm{E}(X+Y) + \mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y) + \mathrm{E}(Z)$ and using the properties of expectations, proposition 8.2iii), we arrive at

Corollary 10.1 (Linearity of expectation) *If* $X_1, X_2, \ldots, X_n$ *are discrete random variables and* $c_1, c_2, \ldots, c_n$ *real valued constants then*

$$E(c_1 X_1 + c_2 X_2 + \ldots + c_n X_n) = c_1 E(X_1) + c_2 E(X_2) + \ldots + c_n E(X_n)\,.$$

Example 10.5 *Consider the sample space of a Bernoulli*(p) *trial and define* n *random variables* $X_1, X_2, \ldots, X_n$ *such that* $X_i(\mathit{success}) = 1$ *and* $X_i(\mathit{failure}) = 0$*, see section 9a. Hence* $E(X_1) = E(X_2) = \ldots E(X_n) = p$*. Consider the random variable* $X = X_1 + X_2 + \ldots + X_n$*. Then by corollary 10.1 (see as well section 9b)*

$$E(X) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) = np\,.$$

Definition 10.3 *Two random variables are* independent *if for all x_k and y_ℓ we have*

$$\mathbb{P}(X = x_k, Y = y_\ell) = \mathbb{P}(X = x_k)\mathbb{P}(Y = y_\ell)\,.$$

Equivalently, the random variables X and Y are independent if for all x_k, y_ℓ, the events $X = x_k$ and $Y = y_\ell$ are independent.

The random variables $X_1, X_2, \ldots, X_n$ are independent *if for all $x_1, x_2, \ldots, x_n$ we have*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\ldots\mathbb{P}(X_n = x_n).$$

Example 10.6 *The random variables R and B in the experiment of example 10.1 are not independent.*

For instance, using the probabilities in example 10.2 and example 10.3

$$\frac{12}{35} = \mathbb{P}(R = 1, B = 1) \neq \mathbb{P}(R = 1)\mathbb{P}(B = 1) = \frac{18}{35} \times \frac{20}{35}\,.$$

Remark: Some useful identity

$$(a_1 + a_2)(b_1 + b_2) = a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_2$$

i.e.

$$\left(\sum_{k=1}^{2} a_k\right)\left(\sum_{\ell=1}^{2} b_\ell\right) = \sum_{k,\ell=1}^{2} a_k b_\ell$$

The relation holds for general sums

$$\left(\sum_{k} a_k\right)\left(\sum_{\ell} b_\ell\right) = \sum_{k,\ell} a_k b_\ell$$

Theorem 10.2 *If X and Y are independent discrete random variables then:*

i) $E(XY) = E(X)E(Y)$,

ii) $Var(X + Y) = Var(X) + Var(Y)$.

Proof:

i)

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_k \sum_\ell x_k y_\ell \mathbb{P}(X = x_k, Y = y_\ell) \\
&= \sum_k \sum_\ell x_k y_\ell \mathbb{P}(X = x_k)\mathbb{P}(Y = y_\ell) \text{ by independence} \\
&= \left(\sum_k x_k \mathbb{P}(X = x_k)\right)\left(\sum_\ell y_\ell \mathbb{P}(Y = y_\ell)\right) \\
&= \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}$$

ii)

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}((X+Y)^2) - [\mathrm{E}(X+Y)]^2 \\
&= \mathrm{E}(X^2 + 2XY + Y^2) - [\mathrm{E}(X) + \mathrm{E}(Y)]^2 \\
&= \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}(Y^2) - [\mathrm{E}(X)]^2 - 2\mathrm{E}(X)\mathrm{E}(Y) - [\mathrm{E}(Y)]^2 \\
&= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 + \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 + 2\left(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\right) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) \text{ by part i)}
\end{aligned}$$

$\square$.

Using theorem 10.2 repeatedly, and the properties of the variance, proposition 8.3iii) we arrive at

Corollary 10.2 *If $X_1, X_2, \ldots, X_n$ are independent discrete random variables and $c_1, c_2, \ldots, c_n$ real valued constants then*

$$Var(c_1X_1 + c_2X_2 + \ldots + c_nX_n) = c_1^2\, Var(X_1) + c_2^2\, Var(X_2) + \ldots + c_n^2\, Var(X_n)\,.$$

Remark: Note that corollary 10.2 for the variance requires the random variables to be independent, while the linearity of the expectation value, corollary 10.1, holds for dependent random variables as well.

Example 10.7 *Consider the experiment of example 10.5 and assume the Bernoulli trials to be independent, i.e., $X_1, X_2, \ldots, X_n$ are independent random variables. Since $Var(X_i) = p(1-p)$, see section 9a, corollary 10.2 tells us that (see section 9b as well)*

$$Var(X) = Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n) = np(1-p)$$

Exercise Sheet 10

This sheet is based on the material lectured in week 11. The questions are about problems with several random variables. The sheet will be discussed in tutorials in week 12.

Hand in your homework, problem J, to the tutor at the beginning of the tutorial, and collect problem I from your tutor to complete your submission.

Questions 28, 29, and 30 will be discussed in the tutorial.

Problem J: (Homework) Suppose that X, Y, Z are random variables with $X \sim \mathrm{Bin}(7, 1/6)$, $Y \sim \mathrm{Geom}(1/2)$, $Z \sim \mathrm{Poisson}(6)$. Suppose further that X and Y are independent but that X and Z are not independent. Which of the following can be determined from this information? Find the value of those which can be determined.

a) $\mathrm{E}(X+Y)$

b) $\mathrm{E}(X+Z)$

c) $\mathrm{E}(X+2Y+3Z)$

d) $\mathrm{E}(X^2+Y^2+Z^2)$

e) $\mathrm{Var}(X+Y)$

f) $\mathrm{Var}(X+Z)$

g) $\mathrm{Var}(X+2Y+3Z)$.

Problem 28: Two fair standard dice are rolled. Let A be the number of 1s seen and B be the number of 2s seen in the outcome. Find the joint distribution of A and B and the two marginal distributions. Are A and B independent random variables.

Problem 29: Let X and Y be discrete random variables with probability mass functions given by

x_k	0	1
$\mathbb{P}(X = x_k)$	$1/2$	$1/2$

and

y_ℓ	0	1	2
$\mathbb{P}(Y = y_\ell)$	1/3	1/3	1/3

.

Furthermore assume that

$$\mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 1, Y = 2) = 0\,.$$

Find the joint probability mass function of X and Y.

Problem 30: You roll 5 fair dice. Let X denote the sum of the numbers shown. Compute the expectation and the variance of X.

§11 Covariance and Conditional Expectation

Let us introduce some quantity which tell us by "how much" two random variables depend on each other.

Definition 11.1 *If X and Y are discrete random variables the* covariance of X and Y *is defined to be*

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

The correlation coefficient of X and Y *is defined to be*

$$corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Remark:

- If X and Y are independent, definition 10.3, then according to theorem 10.2 we have $\mathrm{Cor}(X,Y) = \mathrm{corr}(X,Y) = 0$.
- If X and Y have the tendency to be either both positive or both being negative (to be precise: $X - \mathrm{E}(X)$ and $Y - \mathrm{E}(Y)$ both being positive/negative) then the covariance and the correlation coefficient will be positive. If there is a tendency for X being positive and Y being negative, or vice versa, then the covariance and the correlation coefficient will be negative.
- The correlation coefficient obeys $-1 \leq \mathrm{corr}(X,Y) \leq 1$ (see problem 32c, sheet 11).

Example 11.1 *Consider the experiment of example 10.1. Examples 10.3 and 10.4 tell us that*

$$Cor(R,B) = E(RB) - E(R)E(B) = \frac{6}{7} - \frac{6}{7} \times \frac{9}{7} = -\frac{12}{49}$$

The covariance is negative, that makes sense. Having more red balls (on average) makes it likely to have fewer blue balls (on average).

Proposition 11.1 (Properties of covariance and correlation coefficient) *If X and Y are random variables then:*

i) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

ii) If a, b, c, d *are real valued constants then*

$$Cov(aX + b, cY + d) = ac\,Cov(X,Y).$$

iii) If a, b, c, d *are real valued constants and* $a, c > 0$ *then*

$$corr(aX + b, cY + d) = corr(X,Y).$$

For the proof see problem 32, sheet 11.

Example 11.2 *Consider the experiment of example 10.1 with joint probability mass function as in example 10.2. What is the probability to pick k red balls, assuming that we pick one blue ball?*

The corresponding conditional probability reads

$$\mathbb{P}(R = k|B = 1) = \frac{\mathbb{P}(R = k, B = 1)}{\mathbb{P}(B = 1)}$$

The conditional probabilities can be obtained by dividing the values in the second row in the table of example 10.2 by the relevant marginal probability

k	0	1	2	3
$\mathbb{P}(R = k\|B = 1)$	$2/20$	$12/20$	$6/20$	0

The result is a conditional probability mass function. Note that rows and columns in the table of the joint pmf, example 10.2, do not give probability mass functions as the numbers do not add up to one.

We can work out as well the expectation of red balls, given that we have one blue ball

$$E(R|B = 1) = 0 \times \frac{2}{20} + 1 \times \frac{12}{20} + 2 \times \frac{6}{20} + 3 \times 0 = \frac{24}{20} = \frac{6}{5}\,.$$

$E(R) > E(R|B = 1)$ *(see example 10.3) which makes perfect sense as having a blue ball means having fewer red balls on average.*

Definition 11.2 *Let X be a discrete random variable and A is an event with $\mathbb{P}(A) > 0$. The conditional probability*

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

defines the probability mass function of X given A. The corresponding expectation

$$E(X|A) = \sum_k x_k \mathbb{P}(X = x_k|A)$$

is called the conditional expectation.

Remark: One may consider $X|A$ ("random variable X given an event A") as a random variable with pmf being given by definition 11.2

Example 11.3 *A standard fair die is rolled twice. Let X be the number showing on the first roll and A be the event "at least one odd number is rolled". Find the distribution of $X|A$.*

We have that $\mathbb{P}(A) = 3/4$.

Consider a fixed value k. k may be even or odd.

If k is odd then the event $(X = k) \cap A$ ("first roll k/odd and at least one odd number") is equal to $X = k$ ("first roll k/odd"). Hence

$$\mathbb{P}(X = k|A) = \frac{\mathbb{P}(X = k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(X = k)}{\mathbb{P}(A)} = \frac{1/6}{3/4} = \frac{2}{9}.$$

If k is even then the event $X = k \cap A$ ("first roll k/even and at least one odd number") is equal to the event "first roll k/even and second roll odd") so that (using independence of rolls) $P(X = k \cap A) = \mathbb{P}(X = k) \times 1/2 = 1/6 \times 1/2$ and

$$\mathbb{P}(X = k|A) = \frac{\mathbb{P}(X = k \cap A)}{\mathbb{P}(A)} = \frac{1/6 \times 1/2}{3/4} = \frac{1}{9}.$$

It follows that the conditional pmf of $X|A$ is

k	1	2	3	4	5	6
$\mathbb{P}(X = k\|A)$	2/9	1/9	2/9	1/9	2/9	1/9

In addition

$$E(X|A) = \frac{2 + 2 + 6 + 4 + 10 + 6}{9} = \frac{10}{3}.$$

The following statement can be considered to be a special case of theorem 6.1

Theorem 11.1 *Denote by $A_1, A_2, \ldots, A_n$ a partition of $\mathcal{S}$ (that is the sets are pairwise disjoint and their union is $\mathcal{S}$) with $\mathbb{P}(A_i) > 0$ for all i,. If X is a discreet random variable then*

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

Proof: Apply the Theorem of Total Probability, theorem 6.1, for the event $X = x_k$ with the partition $A_1, \ldots, A_n$

$$\mathbb{P}(X = x_k) = \sum_{i=1}^{n} \mathbb{P}(X = x_k|A_i)\mathbb{P}(A_i)\,.$$

By definition

$$\mathrm{E}(X) = \sum_k x_k \mathbb{P}(X = x_k).$$

so

$$\begin{aligned}
\mathrm{E}(X) &= \sum_k x_k \sum_{i=1}^{n} \mathbb{P}(X = x_k|A_i)\mathbb{P}(A_i) = \sum_{i=1}^{n}\sum_k x_k \mathbb{P}(X = x_k|A_i)\mathbb{P}(A_i) \\
&= \sum_{i=1}^{n} \mathrm{E}(X|A_i)\mathbb{P}(A_i)\,.
\end{aligned}$$

$\square$.

Example 11.4 *We toss a fair coin. If the coin comes up heads we roll a standard 6-sided die. If it comes up tails we roll a standard 4-sided die (with sides numbered 1, 2, 3, 4). Let X be the number we roll. Find its expectation.*

Let H be the event "coin shows head". Clearly H, H^c is a partition and $\mathbb{P}(H) = \mathbb{P}(H^c) = 1/2$ (as the coin is fair).

$X|H$ is the (random variable) number shown by the die given the coins shows head, i.e., the number shown by the 6 sided die. Clearly its expectation reads

$$E(X|H) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}\,.$$

$X|H^c$ is the (random variable) number shown by the die given the coin shows tail, i.e., the number shown by the 4 sided die. Clearly

$$E(X|H^c) = \frac{1+2+3+4}{4} = \frac{5}{2}\,.$$

Theorem 11.1 with the partition H, H^c gives

$$E(X) = \mathbb{P}(X|H)\mathbb{P}(H) + \mathbb{P}(X|H^c)\mathbb{P}(H^c) = \frac{7}{2} \times \frac{1}{2} + \frac{5}{2} \times \frac{1}{2} = 3$$

Example 11.5 *(simple random walk) Consider an infinite one dimensional lattice where sites are labelled by integers. Consider a walker on this lattice and let X_t denote the position of the walker at discrete time t where $t = 0, 1, 2, 3, \ldots$.*

At time $t = 0$ the walker is at site 0, i.e., $X_0 = 0$. At each time step the walker moves with probability $1/2$ to either the site left or right of their current position, i.e., if $X_{t-1} = \ell$ then $X_t = \ell + 1$ with probability $1/2$ and $X_t = \ell - 1$ with probability $1/2$. Written in terms of conditional probabilities $\mathbb{P}(X_t = \ell + 1|X_{t-1} = \ell) = 1/2$, $\mathbb{P}(X_t = \ell - 1|X_{t-1} = \ell) = 1/2$, and $\mathbb{P}(X_t = k|X_{t-1} = \ell) = 0$ if $k \neq \ell \pm 1$.

Compute the expectation of X_t.

Using theorem 11.1 with the partition $A_\ell = (X_{t-1} = \ell)$ we have

$$E(X_t) = \sum_{\ell=-\infty}^{\infty} E(X_t|X_{t-1} = \ell)\mathbb{P}(X_{t-1} = \ell)\,.$$

Definition 11.2 tells us

$$E(X_t|X_{t-1} = \ell) = \sum_{k=-\infty}^{\infty} k\mathbb{P}(X_t = k|X_{t-1} = \ell) = \frac{1}{2}(\ell + 1) + \frac{1}{2}(\ell - 1) = \ell\,.$$

Hence

$$E(X_t) = \sum_{\ell=-\infty}^{\infty} \ell\mathbb{P}(X_{t-1} = \ell) = E(X_{t-1})$$

Therefore $E(X_t) = E(X_0) = 0$ (i.e. on average the walker does not move).

Exercise Sheet 11

This sheet is based on the material lectured in week 12. The questions are about covariance and conditional expectation. The sheet is for self study.

There is no homework. Collect problem J during my office hour in the first week of the second term to complete your submission.

Questions 31, 32, and 33 are for self study.

Problem 31: A fair die is rolled. Let N be the number showing on the die. Now a fair coin is tossed N times. Let X be the number of heads seen. Find the conditional distribution of X given $N = n$ and hence find the expectation of X.

Problem 32: Let X and Y denote two random variables.

a) Using the proof of theorem 10.2ii, or otherwise, show that

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)\,.$$

b) If a, b, c, d are real valued constants show that

$$\mathrm{Cov}(aX+b, cY+d) = ac\mathrm{Cov}(X,Y)\,.$$

c) Assume in addition that $a > 0$ and $c > 0$. Using the result of part b) and properties of the variance, or otherwise, show that

$$\mathrm{corr}(aX+b, cY+d) = \mathrm{corr}(X,Y)\,.$$

Problem 33: Consider the simple random walk on an infinite lattice (see example 11.5).

a) Show that

$$\mathrm{E}((X_t)^2 | X_{t-1} = \ell) = \frac{1}{2}(\ell+1)^2 + \frac{1}{2}(\ell-1)^2 = \ell^2 + 1\,.$$

b) Using theorem 11.1 show that

$$\mathrm{E}((X_t)^2) = \mathrm{E}((X_{t-1})^2) + 1\,.$$

c) Using the result of part b) show that

$$\mathrm{Var}(X_t) = \mathrm{E}((X_t)^2) = t\,.$$