SMSTC (2007/08)

Probability

www.smstc.ac.uk

Contents

SMSTC (2007/08)
Probability
Lecture 1: Probability spaces

Stan Zachary, Heriot-Watt University[a]

www.smstc.ac.uk

Contents

1.1 Probability models

1.1.1 Introduction

We wish to construct mathematical models to enable us to do calculations with *uncertainty*.

This might concern the result of something yet to happen (next Saturday's weather, the winner of next year's Grand National, the success of a space mission). Alternatively it might concern something which is completely determined, but about which we lack precise knowledge (the winner of yesterday's Grand National if we have not heard the result, the value of a physical constant about which we have limited information).

Specifically, we wish to assign *probabilities* to *events*. An *event* is something which may or may not turn out to be the case (think of it as something on which you could place a bet).

For example, the individual *outcome* of Saturday's weather might be the *maximum temperature*—perhaps along with a lot of other information. We might wish the *probability* of the *event* that the *maximum temperature* will be at least 20°C. This *event* is the collection (set) of *all* those *outcomes* in which the *maximum temperature* is 20°C or more.

Note that *probabilities* reflect our lack of precise knowledge. The probability of a given event will change as our knowledge changes. As Saturday approaches it may become more or less likely that the maximum temperature will be at least 20°C.

[a] s.zachary@hw.ac.uk

1.1.2 Probability spaces

A *probability space* is the mathematical foundation of a probability model. It is usually taken to consist of a triple $(\Omega, \mathcal{F}, \mathbf{P})$:

- Ω is the *sample space*, i.e. the collection of all possible individual *outcomes* ω of the situation under study;
- $\mathcal{F}$ is the *set* of *events* which may be considered; an *event* $E \in \mathcal{F}$ is a set of possible *outcomes* $\omega \in \Omega$, i.e. it is a *subset* of Ω (it follows that $\mathcal{F}$ is a set of subsets of Ω); the *event* E is said to *occur* if the actual outcome ω belongs to E, i.e. $\omega \in E$;
- $\mathbf{P}$ is a *probability measure* which assigns a *probability* $\mathbf{P}(E)$ to every event $E \in \mathcal{F}$.

Example 1.1. *(Roulette wheel.)* We wish to model the result of a single spin of a (single zero) roulette wheel. It is natural to label the possible *outcomes* of this *experiment* as $0, 1, 2, \dots, 36$ and hence to take the *sample space* $\Omega = \{0, 1, 2, \dots, 36\}$. It is also natural to take $\mathcal{F}$ to be the set of *all* subsets of Ω, i.e. any such subset may be considered an *event.* Finally, if the roulette wheel is fair, and hence all outcomes equally likely, it is natural to define the *probability* of any *event* E by $\mathbf{P}(E) = |E|/37$, where $|E|$ is the size of the set E. Some possible events and their probabilities are:

- the event A that the number 36 comes up; we have $A = \{36\}$ and $\mathbf{P}(A) = 1/37$;
- the event A^c that the number 36 does *not* comes up; we have $A^c = \{1, 2, 3, \dots, 35\}$ and $\mathbf{P}(A^c) = 36/37$;
- the event B that an odd number comes up; we have $B = \{1, 3, \dots, 35\}$ and $\mathbf{P}(B) = 18/37$;
- the event C that a number between 1 and 10 comes up; we have $C = \{1, 2, \dots, 10\}$ and $\mathbf{P}(C) = 10/37$;
- the event $B \cap C$ that *both B and C* occur; we have $B \cap C = \{1, 3, 5, 7, 9\}$ and $\mathbf{P}(B \cap C) = 5/37$;
- the event $B \cup C$ that *either B or C* occurs; we have $B \cup C = \{1, 2, \dots, 10, 11, 13, \dots, 35\}$ and $\mathbf{P}(B \cup C) = 23/37$;
- the event $A \cap B$ that *both A and B* occur; we have $A \cap B = \emptyset$ (the empty set) and $\mathbf{P}(\emptyset) = 0$;

Example 1.2. *(Coin tossing.)* A fair coin is tossed n times. It is natural to take the *sample space* Ω to be the set of all possible *outcomes* of this *experiment*, where a single *outcome* is a given *sequence* of heads and tails. It is easy to see (formally by induction *[exercise!]*) that there are 2^n such *outcomes.* It is again also natural to take $\mathcal{F}$ to be the set of all subsets of Ω. Since the coin is fair, we assume that all 2^n *outcomes* are equally likely, so that each is assigned probability 2^{-n}. For each of $k = 0, 1, \dots, n$, let A_k be the event that exactly k of the n tosses turn out to be *heads.* Then, for each such k, we have $\mathbf{P}(A_k) = 2^{-n}\binom{n}{k}$. Note that the *events* A_k, $k = 0, 1, \dots, n$, are *disjoint* (i.e. $A_i \cap A_j = \emptyset$ for all $0 \le i < j \le k$) and together have *union* equal to the entire *sample space* Ω, that is to say that (whatever the outcome of the experiment) exactly one of the events A_k occurs. We say that the events A_k, $k = 0, 1, \dots, n$ form a *partition* of the *sample space* Ω, and we have $\sum_{k=0}^{n} \mathbf{P}(A_k) = 1$.

In both the above examples the *sample space* Ω is *finite.* In this case—and also more generally when Ω is *countable* (see below)—there is never any difficulty in taking the set $\mathcal{F}$ above to be the set of all subsets of Ω, and there is also no difficulty in defining the *probability* $\mathbf{P}(E)$ of any *event* E to be the sum of the probabilities of the individual *outcomes* which comprise that *event.* We require that the probabilities of these individual *outcomes* should be nonnegative and should sum to 1.

In general (e.g. when the *sample space* Ω is *uncountable*) we need to be more careful. First, as we shall see in Section 1.1.4, it is not always possible to consider every subset of Ω as an *event* (although in applications there are never any difficulties) and—at least in principle—$\mathcal{F}$ needs to be explicitly identified. We require that $\mathcal{F}$ should be a *σ-algebra*. *σ-algebras* play a central role in all but the most elementary understandings of probability theory and are discussed in Section 1.1.3 below.

Secondly, when the number of the *outcomes* which define an *event* is *uncountable*, we *cannot* simply define the *probability* of an *event* as a sum of the probabilities of the *outcomes* it contains. Hence we require a more general rule for assigning *probabilities* directly to *events*. This is discussed in Section 1.1.4.

1.1.3 σ-algebras

Recall first that a set is *countable* if it is *finite* or if it may be placed into a *one-one correspondence* with the set of *positive integers*. Examples of *countable* sets are the set $\mathbb{Z}$ of *integers* and the set $\mathbb{Q}$ of *rational numbers*. Examples of *uncountable* sets are the set $\mathbb{R}$ of *real numbers* and any *interval* $[a, b] \subset \mathbb{R}$ where $a < b$. *[Exercise: make sure you understand these examples.]*

We now need to understand the set $\mathcal{F}$ of *events* to which *probabilities* may be assigned in a given model. We require that $\mathcal{F}$ should be a *σ-algebra*, i.e. that it should have the following axiomatic properties

(i) $\mathcal{F}$ is *nonempty*;

(ii) if $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$;

(iii) if the countable collection of events $E_1, E_2, \dots \in \mathcal{F}$, then $\bigcup_{n \geq 1} E_n \in \mathcal{F}$.

Thus a nonempty collection of subsets of Ω is a *σ-algebra* if it is *closed* under the operations of taking *complements* and of taking *countable unions*. This implies that $\mathcal{F}$ is also *closed* under *countable* applications of other standard set-theoretic operations. For example, if $E_1, E_2, \dots \in \mathcal{F}$, then since $\bigcap_{n \geq 1} E_n = \left(\bigcup_{n \geq 1} E_n^c\right)^c$, it follows that $\bigcap_{n \geq 1} E_n \in \mathcal{F}$; thus $\mathcal{F}$ is *closed* under the operation of taking *countable intersections*. It also follows that, in particular, the sets Ω (the entire sample space) and $\emptyset$ (the empty set) *always* belong to the *σ-algebra* $\mathcal{F}$. We remark, for future use, that the *σ-algebra* $\mathcal{F} = (\Omega, \emptyset)$ consisting of just these two sets is referred to as the *trivial σ-algebra*.

At the other extreme, the set of all subsets of Ω is a *σ-algebra*. As remarked above, when the sample space Ω is *countable*, we take this as the *σ-algebra* $\mathcal{F}$ associated with our probability model. Thus every subset of Ω is a legitimate *event*, and we can effectively forget about $\mathcal{F}$. We shall see below that this simple approach is not possible in the case of *uncountable* sample spaces.

It is also necessary to be able to define the *smallest σ-algebra* which contains a given collection of subsets (events) of Ω. This is referred to as the *σ-algebra generated* by the given collection, and is *defined* to be the *intersection* of all those *σ-algebras* each of which contains the given collection of subsets. That this intersection of *σ-algebras* is itself a *σ-algebra* follows since it necessarily has the required closure properties *[exercise!]*.

Example 1.3. For any subset E of Ω (such that $\emptyset \neq E \neq \Omega$) the σ-algebra $\mathcal{G}(E)$ generated by E consists of 4 sets and is given by

$$\mathcal{G}(E) = \{\emptyset, E, E^c, \Omega\}.$$

To see this, observe that this *is* a σ-algebra, and is clearly the smallest σ-algebra which contains E.

Example 1.4. Let $\{E_1, E_2, \dots\}$ be a *partition* (finite or countably infinite) of Ω, i.e. a collection of *disjoint* subsets of Ω ($E_i \cap E_j = \emptyset$ whenever $i \neq j$) whose union is the entire sample space Ω. Then the *σ-algebra* $\mathcal{G}(E_1, E_2, \dots)$ generated by this *partition* consists of all countable unions of sub-collections of the events $E_1, E_2, \dots$. To see this, again observe that this *is* a *σ-algebra*, and is clearly the *smallest* which contains all of the events $E_1, E_2, \dots$.

The Borel σ-algebra. Consider the *experiment* of choosing, or the *observation* of, a point randomly distributed on the *real line* $\mathbb{R}$. At a minimum we will wish to be able to consider as an *event* (that the point lies within) any given *interval* within $\mathbb{R}$, where either endpoint of the interval may be *open* or *closed* and where the interval may be *finite*, *half-infinite* (e.g. $(-\infty, a]$ or (a, ∞) for some $a \in \mathbb{R}$) or indeed $\emptyset$ or $\mathbb{R}$ itself. Therefore we need to be able to consider, and to assign *probabilities* to sets within, the *σ-algebra generated* by such *intervals*. This is referred to as the *Borel σ-algebra* on $\mathbb{R}$, and is usually denoted by $\mathcal{B}$. The sets within it are referred to as the *Borel sets*. Notably $\mathcal{B}$ contains all the *intervals* as above, and also all *countable unions* of *intervals*—in particular $\mathcal{B}$ contains all *finite* and *countably infinite* sets. While it is convenient to think of it as the σ-algebra generated by the intervals, it is also generated by smaller collections of sets, e.g. all sets of the form $(-\infty, a]$ for $a \in \mathbb{R}$.

The *Borel σ-algebra* may be similarly defined on any *interval* within $\mathbb{R}$, again as the σ-algebra generated by the intervals within this interval. (We might. for example, wish to consider the experiment of choosing a number at random on the interval $[0, 1]$.)

In another direction the *Borel σ-algebra* may also be defined on the product space $\mathbb{R} \times \mathbb{R} \times \dots$ of finite or countably infinite sequences $x = (x_1, x_2, ...)$ where each $x_i \in \mathbb{R}$. (For example, we might wish to consider a sequence of random numbers.) In this case the *Borel σ-algebra* is that *generated* by the *cylinder sets* of the form $\{x \colon x_i \in B\}$ for all i and for all *Borel sets* B contained in $\mathbb{R}$.

1.1.4 Probability measures

Given a *sample space* Ω and a *σ-algebra* $\mathcal{F}$ of *events* within Ω, a *probability measure* $\mathbf{P}$ is a function on $\mathcal{F}$ which satisfies the following *axioms*.

P1: For any event $E \in \mathcal{F}$,
$$0 \le \mathbf{P}(E) \le 1.$$

P2: For the entire sample space Ω,
$$\mathbf{P}(\Omega) = 1.$$

P3: (**Extended addition rule**) For any countable set of *disjoint* events $(E_j,\ j \in J)$ in $\mathcal{F}$ (i.e. such that $E_i \cap E_j = \emptyset$ for all $i, j \in J$ with $i \neq j$),
$$\mathbf{P}\Big(\bigcup_{j \in J} E_j\Big) = \sum_{j \in J} \mathbf{P}(E_j).$$

Note that (in the usual treatment of probability theory) it is insufficient to require the property P3 to hold for *finite* collections of disjoint events; we need—notably in the theory of stochastic processes—to be able to work with *countable* unions and intersections of events. In particular the *continuity rules* P9 and P10 below follow from P3 as applied to *countable* sets.

We have already remarked that when the sample space Ω is *countable* and the σ-algebra $\mathcal{F}$ is the set of all subsets of Ω, then we may associate probabilities, nonnegative and summing to 1, with the individual *outcomes* in Ω, and define the *probability* of any *event* $E \in \mathcal{F}$ as the sum of probabilities associated with the individual outcomes in E. It is clear that the *probability measure* thus defined does indeed satisfy the axioms P1–P3, and conversely that any *probability measure* $\mathbf{P}$ on $\mathcal{F}$, necessarily satisfying P1–P3, may always be constructed as above.

When the sample space Ω is *uncountable*, it is not in general possible to assign probabilities as above, nor even to assign a probability to every subset of Ω. The following example discuses in detail how to proceed when Ω is the real line.

Example 1.5. *(Choosing a number at random on the real line $\mathbb{R}$.)* Suppose that we wish to choose a number at random on $\mathbb{R}$. As previously discussed it is necessary to define an appropriate *probability measure* $\mathbf{P}$ on $\mathbb{R}$ endowed with (at least) the *Borel σ-algebra $\mathcal{B}$*. Suppose first that we have such a probability measure $\mathbf{P}$. Define the *distribution function* F of $\mathbf{P}$ by

$$F(x) = \mathbf{P}((-\infty, x]), \qquad x \in \mathbb{R}. \tag{1.1}$$

Then F has the following properties:

(i) F is *increasing*;

(ii)
$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1;$$

(iii) F is *right continuous*, i.e.
$$\lim_{x'\downarrow x} F(x') = F(x) \qquad \text{for all } x \in \mathbb{R}.$$

Note that, since F is increasing, for all $x \in \mathbb{R}$, both the *right limit* $\lim_{x'\downarrow x} F(x')$ and *left limit* $\lim_{x'\uparrow x} F(x')$ always exist, but the *left limit* is equal to $F(x)$ if and only if F has no *jump* at x, i.e. F is *continuous* at x.

The properties (i)–(iii) follow from the axioms P1–P3, but are most easily deduced once we have used these axioms to establish some further properties of probability measures below.

Conversely, suppose we are given a function F on $\mathbb{R}$ with the above properties (i)–(iii). Then it is an exercise in *measure theory* to show that there *exists* a *unique probability measure* $\mathbf{P}$ on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the *Borel σ-algebra*, such that (1.1) holds. For example, for any *interval* of the form $(a, b]$, we necessarily have, from P3,

$$\begin{aligned}\mathbf{P}((a,b]) &= \mathbf{P}((-\infty,b]) - \mathbf{P}((-\infty,a]) \\ &= F(b) - F(a).\end{aligned}$$

The probability assigned by $\mathbf{P}$ to any *individual point* $x \in \mathbb{R}$ is

$$\begin{aligned}\mathbf{P}(\{x\}) &= \lim_{x'\uparrow x} \mathbf{P}((x', x]) \qquad \text{(by P10 below)} \\ &= F(x) - \lim_{x'\uparrow x} F(x'),\end{aligned}$$

which is the size of the *jump* in the function F at the point x; we have $\mathbf{P}(\{x\}) = 0$ if and only if F is *continuous* at x. Also, by P3, the probability assigned by $\mathbf{P}$ to any *countable union* of *disjoint intervals* is the sum of the probabilities assigned by $\mathbf{P}$ to the individual intervals.

Finally suppose that we are given any *probability measure* $\mathbf{P}$ on the *Borel σ-algebra $\mathcal{B}$* as above. It is both possible and convenient to extend the domain of definition of $\mathbf{P}$ to the enlarged σ-algebra which is defined to be the *smallest* σ-algebra which includes both sets in $\mathcal{B}$ and *also* all subsets of sets with probability measure 0. This latter σ-algebra is the *completion* of the *Borel σ-algebra $\mathcal{B}$* with respect to $\mathbf{P}$.

Example 1.6. *(Choosing a number* uniformly *at random on the interval* $[0, 1]$*.)* Suppose that we wish to choose a number *uniformly* at random from within the interval $[0, 1]$. (This is what the random number generator on a computer is supposed to mimic.) We proceed as in the previous example, replacing the *sample space* $\Omega = \mathbb{R}$ by $\Omega = [0, 1]$. The *distribution function* F on $\Omega = [0, 1]$, given by

$$F(x) = x, \qquad x \in [0, 1],$$

defines the *uniform probability measure* $\mathbf{P}$ on $(\Omega, \mathcal{B})$ (where $\mathcal{B}$ is again the *Borel* σ*-algebra*). In particular, the *probability* assigned by $\mathbf{P}$ to any *interval* in $[0, 1]$ is given by its *length*, and the *probability* assigned by $\mathbf{P}$ to any *single point* in the interval $[0, 1]$ is 0. It can further be shown that it is *not possible* to extend the domain of definition of $\mathbf{P}$ to the set of *all* subsets of $[0, 1]$ while continuing to satisfy the *axioms* P1–P3 above, i.e. for the *uniform probability measure* $\mathbf{P}$, there exist subsets of $[0, 1]$ which *cannot* be considered as *events* (though such sets are very exotic). However, the *Borel* σ*-algebra* $\mathcal{B}$ may be *completed* with respect to the *uniform probability measure* $\mathbf{P}$ as described above to obtain the *Lebesgue* σ*-algebra*.

Further properties of probability measures. From the *axiomatic properties* P1–P3 defined above, we can easily derive the following further *properties*.

P4 (**Complement rule**) For any event E,

$$\mathbf{P}(E^c) = 1 - \mathbf{P}(E).$$

This is one of the most useful properties of probability measures. In complex problems it is often simpler to calculate the probability of the complement of an event of interest. Note also that it follows from P2 and P4 that $\mathbf{P}(\emptyset) = 0$ always.

P5 (**Disjoint intersection rule)** If the events E and F *disjoint* (i.e. $E \cap F = \emptyset$), then

$$\mathbf{P}(E \cap F) = \mathbf{P}(\emptyset) = 0.$$

[Recall the $E \cap F$ is the event that *both* E *and* F occur.]

P6 (**Inclusion-exclusion rule**) For *any* two events E and F,

$$\mathbf{P}(E \cup F) = \mathbf{P}(E) + \mathbf{P}(F) - \mathbf{P}(E \cap F).$$

This is proved, for example, by noting that, from P3, we have the two relations

$$\mathbf{P}(E \cup F) = \mathbf{P}(E) + \mathbf{P}(F \setminus E)$$
$$\mathbf{P}(F) = \mathbf{P}(E \cap F) + \mathbf{P}(F \setminus E).$$

and eliminating $\mathbf{P}(F \setminus E)$ (recall the $F \setminus E$ is the event that F *but not* E occurs).

Example 1.7. Outside work on an oil platform cannot proceed if it is either too wet or too windy. The probability of the event A that work is cancelled because it is too wet is 0.3, and the probability of the event B that work is cancelled because it is too windy is 0.2. The probability of the event $A \cap B$ that it is simultaneously both too wet and too windy is 0.1. Find the probability that work can proceed.

The event that it is either too wet or too windy is $A \cup B$ and we have, by P6,

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) = 0.3 + 0.2 - 0.1 = 0.4.$$

The event C that work can proceed is the complement of the event $A \cup B$ and hence, by P4, the probability that work can proceed is given by

$$\mathbf{P}(C) = 1 - \mathbf{P}(A \cup B) = 1 - 0.4 = 0.6.$$

P7 (**Subset rule**) If $E \subseteq F$, then

$$\mathbf{P}(E) \leq \mathbf{P}(F).$$

P8 (**Boole's inequality**) For *any* countable collection of events $E_1, E_2, \dots$ (finite or infinite),

$$\mathbf{P}\left(\bigcup_{\text{all } n} E_n\right) \leq \sum_{\text{all } n} \mathbf{P}(E_n).$$

(Compare this with the property P3 for *disjoint* sets.)

P9 (**Continuity rule for increasing sequences**) Let $(E_n,\, n \geq 1)$ be an *increasing* sequence of events, i.e.

$$E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots.$$

Then

$$\mathbf{P}\left(\bigcup_{n\geq 1} E_n\right) = \lim_{n\to\infty} \mathbf{P}(E_n).$$

To see this, note that the event $\bigcup_{n\geq 1} E_n$ is also the union of the sequence of *disjoint* events $(F_n,\, n \geq 1)$ where $F_1 = E_1$ and $F_n = E_n \setminus \bigcup_{i=1}^{n-1} E_i$. Hence, from P3,

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{n\geq 1} E_n\right) &= \mathbf{P}\left(\bigcup_{n\geq 1} F_n\right) \\
&= \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{m=1}^{n} F_m\right) \\
&= \lim_{n\to\infty} \mathbf{P}(E_n).
\end{aligned}$$

As a special case, note that when $\bigcup_{n\geq 1} E_n = \Omega$, then $\lim_{n\to\infty} \mathbf{P}(E_n) = \mathbf{P}(\Omega) = 1$.

Also as a more general corollary we have that, for *any* countable sequence of events $(E_n,\, n \geq 1)$,

$$\mathbf{P}\left(\bigcup_{n\geq 1} E_n\right) = \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{i=1}^{n} E_i\right).$$

P10 (**Continuity rule for decreasing sequences**) Let $(E_n,\, n \geq 1)$ be an *decreasing* sequence of events, i.e.

$$E_1 \supseteq E_2 \supseteq E_3 \supseteq \dots.$$

Then

$$\mathbf{P}\left(\bigcap_{n\geq 1} E_n\right) = \lim_{n\to\infty} \mathbf{P}(E_n).$$

To see this, note that the sequence of complements $(E_n^c,\, n \geq 1)$ is *increasing*, and so, from P9,

$$\begin{aligned}
\mathbf{P}\left(\bigcap_{n\geq 1} E_n\right) &= 1 - \mathbf{P}\left(\left(\bigcap_{n\geq 1} E_n\right)^c\right) \\
&= 1 - \mathbf{P}\left(\bigcup_{n\geq 1} E_n^c\right) \\
&= 1 - \lim_{n\to\infty} \mathbf{P}(E_n^c) \\
&= \lim_{n\to\infty} \mathbf{P}(E_n).
\end{aligned}$$

As a special case, note that when $\bigcap_{n\geq 1} E_n = \emptyset$, then $\lim_{n\to\infty} \mathbf{P}(E_n) = \mathbf{P}(\emptyset) = 0$.

Also as a more general corollary we have that, for *any* countable sequence of events $(E_n,\, n \geq 1)$,

$$\mathbf{P}\Big(\bigcap_{n\geq 1} E_n\Big) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcap_{i=1}^{n} E_i\Big).$$

Example 1.8. Suppose that, as in Example 1.2, a fair coin is tossed n times. Let A_0 be the event that no heads are obtained and let B be the event that at least 1 head is obtained. Suppose we require $\mathbf{P}(B)$. We have

$$\mathbf{P}(B) = 1 - \mathbf{P}(B^c) = 1 - \mathbf{P}(A_0) = 1 - 2^{-n}.$$

Suppose now that the coin is tossed infinitely often. We take Ω to be the set of all possible infinite sequences of heads and tails (this is the product space $\{H,T\} \times \{H,T\} \times \dots$) and the corresponding σ-algebra $\mathcal{F}$ to be that generated by the *cylinder sets* as in the discussion of the *Borel σ-algebra* in Section 1.1.3, i.e. to be the smallest σ-algebra which contains all events of the form "the ith toss is a head". Suppose that we require the probability of the event B that at least one head is ever obtained. For each $n \geq 1$ define also B_n to be the event that at least one head is obtained in the first n tosses. Then $B_1 \subset B_2 \subset \dots$ and so it follows from P9 that

$$\mathbf{P}(B) = \mathbf{P}\Big(\bigcup_{n\geq 1} B_n\Big) = \lim_{n\to\infty} \mathbf{P}(B_n) = \lim_{n\to\infty} (1 - 2^{-n}) = 1.$$

1.1.5 Random variables

Example 1.9. Suppose that we are again interested in the number of heads obtained in n tosses of the fair coin. It is convenient to denote this by X. The *observed value* $X(\omega)$ of the *random variable* X depends on the *outcome* ω of the *experiment*. Thus, formally, the *random variable* X is a *function* defined on the *sample space* Ω. For example, if the coin is tossed 3 times we have

outcome ω:	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(\omega)$:	3	2	2	1	2	1	1	0

For any $k = 0, 1, \dots, n$ the set $\{\omega\colon X(\omega) = k\}$ is just the *event* that $X = k$, i.e. the event A_k defined in Example 1.2 that precisely k heads are obtained.

Since the coin is fair and all outcomes (sequences of heads and tails) of the n tosses are considered equally likely, we have

$$\mathbf{P}(X = k) = \mathbf{P}(A_k) = 2^{-n}\binom{n}{k}, \qquad k = 0, 1, \dots, n.$$

This information (the probabilities associated with the different values, or ranges of values, that X may take) is referred to as the *distribution* of the random variable X. It may be given in many different forms. In particular, the function f_X given by $f_X(k) = \mathbf{P}(X = k)$ for $k = 0, 1, \dots, n$ is referred to as the *probability function* (or, sometimes, *probability mass function*) of the random variable X. The function F_X given by $F_X(k) = \mathbf{P}(X \leq k)$ for $k = 0, 1, \dots, n$, is referred to as the *distribution function* (or, occasionally, *cumulative distribution function*) of the random variable X. Clearly either may be computed from the other, and so both convey the same information.

Now define also the *random variable* Y to be the number of *tails* obtained in the n tosses of the fair coin. We have $Y = n - X$, so clearly X and Y are different. However, for all k, we have $\mathbf{P}(X = k) = \mathbf{P}(Y = k) = 2^{-n}\binom{n}{k}$, so two *different random variables* may have the *same distribution*.

We now consider the general definition of a (real-valued) *random variable* X on a *probability space* $(\Omega, \mathcal{F}, \mathbf{P})$. Clearly X has to be a *function* from Ω to $\mathbb{R}$ (or to some subset of $\mathbb{R}$). For any such function X and any subset $B \subseteq \mathbb{R}$, define $X^{-1}(B) = \{\omega\colon X(\omega) \in B\}$, i.e. the set of outcomes ω for which $X(\omega)$ belongs to the set B. For X to be a *random variable*, we clearly require

$$X^{-1}(B) \in \mathcal{F}, \tag{1.2}$$

i.e. $X^{-1}(B)$ to be an *event* (to which we can assign a *probability*) for all sets B of interest, e.g. $B = (-\infty, x]$ or $B = \{x\}$ for any $x \in \mathbb{R}$. It is a straightforward result that the collection of sets B for which (1.2) holds is itself a σ-algebra on $\mathbb{R}$, so we can most neatly express our requirement as that (1.2) should hold for all B belonging to the *Borel σ-algebra* $\mathcal{B}$ on $\mathbb{R}$. Formally we say that the function X from Ω to $\mathbb{R}$ should be *measurable*—more precisely that it should be $(\mathcal{F}, \mathcal{B})$-*measurable*. (However, note that, also from the above result, to show X is *measurable* it will be sufficient to *verify* (1.2) for *Borel* sets B which *generate* $\mathcal{B}$, e.g. all sets of the form $B = (-\infty, x]$, $x \in \mathbb{R}$.) We therefore make the following definition.

Definition 1.1. A (real-valued) *random variable* X on $(\Omega, \mathcal{F}, \mathbf{P})$ is a *measurable* function from Ω to $\mathbb{R}$.

[Exercise: prove the result in the above paragraph: use observations such as $\{\omega\colon X(\omega) \in B^c\} = \{\omega\colon X(\omega) \in B\}^c$*.]*
Random variables are further discussed in Lecture 3.

1.2 Classical Probability

1.2.1 Introduction

As we have already seen, there are many probability models (e.g. coin tossing, urn models) in which the sample space Ω contains some *finite* number n of possible *outcomes* (*sample points*) each of which is considered equally likely and each of which is therefore assigned probability $1/n$. The σ-algebra $\mathcal{F}$ is taken to be the set of all subsets of Ω and the probability $\mathbf{P}(E)$ of any *event* $E \in \mathcal{F}$ is then given by

$$\mathbf{P}(E) = \frac{|E|}{n}, \tag{1.3}$$

where $|E|$ denotes the total number of *outcomes* comprising the *event* E.
Thus many problems in probability theory involve *counting*. This is not as easy as it seems.

1.2.2 Fundamental counting results

Suppose that we have n objects (let them be labelled $1, 2, \ldots, n$), and that we wish to choose k of them. In how many ways can we do this? The answer depends on

- whether the k objects are to be chosen
  - *without replacement*, i.e. each of the n objects may be chosen at most once (so that, if $n = 8$ and $k = 4$, the choice $(1, 3, 1, 2)$ is *not* allowed)
  - *with replacement*, each of the n objects may be chosen as often as we like (the choice $(1, 3, 1, 2)$ *is* allowed);
- whether the k objects are considered as
  - an *ordered* collection, i.e. different orderings of the same k objects are counted as distinct (so, with $n = 8$ and $k = 4$, the choices $(1, 2, 4, 6)$ and $(6, 2, 1, 4)$ are distinct)

– an *unordered* collection, i.e. different orderings of the same k objects are counted as being the same.

We have the following results.

- *Without replacement, ordered.* For this we require $k \leq n$. Then the k objects may be obtained in
$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$
different ways. To see this, note that there are n ways to choose the first, $n-1$ ways to choose the second, etc, and that the total number of ordered choices is the *product* of these. *[Why?]*
- *Without replacement, unordered.* For this we again require $k \leq n$. Since the k objects to be chosen are required to be distinct, each unordered set of them may be ordered in $k!$ different ways. Hence the total number of unordered choices is the total number of ordered choices divided by $k!$, and, from the previous result, this is the *binomial coefficient*
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
- *With replacement, ordered.* (Here there are no restrictions on n and k.) Arguing as for the first result above, we find that the k objects may be obtained in n^k different ways.
- *With replacement, unordered.* (Here again there are no restrictions on n and k.) This time the argument is slightly more tricky. Consider any given unordered collection of size k of objects, where each of these has one of the labels $1, 2, \dots, n$. For example, for $n = 10$, $k = 7$, we might have the choice (with numbers arranged in ascending order) given by $(1, 2, 2, 4, 6, 6, 6)$. Represent this choice instead as
$$\circ \mid \circ\, \circ \mid\mid \circ \mid\mid \circ\, \circ\, \circ \mid\mid\mid\mid$$
where (in general) the k circles represent the k objects and the $n-1$ vertical bars separate the assigned labels $1, 2, \dots, n$. Then the number of distinct choices is the number of distinct patterns as above, and this is the number of ways in which, out of $n+k-1$ symbols, k may be chosen as circles to correspond to the k objects (the remaining $n-1$ symbols being the vertical bars). Hence, finally, he number of ways in which the k objects may be chosen is the binomial coefficient
$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}.$$

Remark 1.1. Another interpretation of the last result: the answer can also be seen as the number of different ways in which k indistinguishable objects, e.g. balls, can be placed in a total of n bins (where some bins may be empty).

1.2.3 Use of multinomial coefficients

Suppose that, in a population of distinct n objects, each is to be assigned to one of r numbered groups, in such a way that there are k_i objects in the ith group, $1 \leq i \leq r$, where (necessarily) $\sum_{i=1}^{r} k_i = n$. Then the number of ways in which this can be done is given by the *multinomial coefficient*
$$\frac{n!}{k_1! \dots k_r!}. \tag{1.4}$$

To see this, note that the number of possible choices of objects for assignment to the first group is the binomial coefficient $\binom{n}{k_1}$; given any such choice, the number of possible further choices for assignment to the second group is the binomial coefficient $\binom{n-k_1}{k_2}$, and so on; once the choice of objects for the $(r-1)$st group have been made, the remaining objects automatically go in the rth group. Hence the total number of assignments of objects to groups is

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{k_{r-1}+k_r}{k_{r-1}},$$

and this is easily seen to evaluate to the expression in (1.4).

In the special case $r = 2$ the expression in (1.4) reduces to the binomial coefficient $\binom{n}{k_1}$ (in this case note again that once the k_1 objects for the first group have been chosen, the remaining objects necessarily go in the second group).

Note that, according to the statement of the problem, the order or numbering of the groups is important, but not the order of the objects within each group. For example, in the case $n = 4$, $k_1 = k_2 = 2$ (and numbering the objects $1, 2, 3, 4$) the 6 possible assignments are

$$(12)(34) \quad (34)(12) \quad (13)(24) \quad (24)(13) \quad (14)(23) \quad (23)(14).$$

or equivalently, in a dual notation,

$$1122 \quad 2211 \quad 1212 \quad 2121 \quad 1221 \quad 2112 \tag{1.5}$$

where, for example, 1122 means that the first and second objects are assigned to group 1, while the third and fourth are assigned to group 2.

An alternative physical formulation of the same mathematical problem is the following. Suppose that, of n balls, k_i have the colour i for $1 \le i \le r$, where (again necessarily) $\sum_{i=1}^{r} k_i = n$. In how many different ways (distinct sequences of colours) can these balls be arranged? It is easy to see that the answer is again given by (1.4) (e.g. consider again the above example, and see that (1.5) now represents the various possible arrangements).

In the original formulation of the problem, and in the special case $k_1 = \cdots = k_r$, we might also wish to consider the problem in which the order or numbering of the groups is not important. (Thus, in the above example, the assignments (12)(34) and (34)(12) would be considered the same.) In this case the number of possible assignments of objects to groups is given by dividing the expression in (1.4) by $r!$.

Example 1.10. A total of 12 people are to be divided into 3 groups of 4 each (so as to take part in some tournament). The order of the groups is unimportant. According to the above the number of ways in which this can be accomplished is

$$\frac{12!}{3!(4!)^3} = 5775$$

(a result which may easily be checked by direct counting).

1.3 Exercises

1–1. A physical device only works if the ambient temperature lies within the range $[T_1, T_2]$ (where $T_1 < T_2$). Suppose that the probability of the event A that the temperature is below T_2 is 0.6, and that the probability of the event B that the temperature is above T_1 is 0.75. Show that the probability that the device works is 0.35

1–2. A fair die (one for which each face is likely to show) is given two *independent* throws, so that all 36 outcomes are equally likely. Find the distribution of the random variable N which is defined to be the total of the numbers obtained on the two throws. Show that the most likely value for this total is 7.

1–3. (Chung.) Six mountain climbers decide to divide into three groups for the final assault on the peak. The groups will be of sizes 1, 2, 3 respectively, and all orders of assault by the three groups are considered. Show that the that the total number of possible groupings and assault orders is 360. In the case where instead each of the three groups is to be of size 2, show that the above total reduces to 90. Understand why the same formula is not applicable in both cases.

1–4. An urn contains 5 red and 3 green balls. Three balls are chosen at random, in succession, and without replacement.

(a) Show that the probability that all three balls drawn are red is 5/28.

(b) Show that the probability that the third ball drawn is red is 5/8. (Make sure you give a quick derivation of this result.)

Show also that in the case where instead the balls are drawn with replacement, the above probabilities become 125/512 and 5/8 respectively.

1–5. *Hypergeometric distribution.* Suppose that n objects are partitioned into a objects of Type 1, say, and $n - a$ objects of Type 2. A random choice is made of b objects, all $\binom{n}{b}$ choices being equally likely. Let the random variable X denote the number of those objects chosen which are of Type 1. Show that

$$\mathbf{P}(X = k) = \frac{\binom{a}{k}\binom{n-a}{b-k}}{\binom{n}{b}},$$

where, necessarily, $\max(0, a + b - n) \leq k \leq \min(a, b)$.

1–6. (Chung.) A pack of cards is shuffled and the cards are then dealt one at a time. Show that the probability that the 4 aces occur consecutively is $24/(52 \times 51 \times 50)$.

1–7. *Bose Einstein statistics.* Suppose that k indistinguishable objects are to be placed in n numbered boxes. Show that the number of ways in which this can be done is

$$\binom{n+k-1}{k}.$$

(Understand that this is just the same as the "With replacement, ordered" problem considered in Section 1.2.2.)

1–8. *Birthday problem.* Suppose there are n days in the year (on Earth we have $n = 365$, but it may be different elsewhere). Show that, in a gathering of k people, the probability $p_{n,k}$ that at least two share the same birthday is given by

$$p_{n,k} = 1 - \frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-k+1}{n}.$$

For $n = 365$ show that the smallest value of k such that $p_{n.k} > 0.5$ is given by $k = 23$.

1–9. *Matching problem.* A collection of n letters fall out of their envelopes and are replaced at random. Let the random variable N be the number of letters which are replaced in their correct envelopes.

(a) In the case $n = 4$, find the distribution of the random variable N (i.e. find $\mathbf{P}(N = k)$ for $k = 0, 1, \dots, 4$).

(b) For each of $n = 1, 2, \dots, 6$, find the probability $\mathbf{P}(N = 0)$ that no letters are replaced in their correct envelopes.

(c) Show that, as $n \to \infty$, the above probability converges to e^{-1} (*difficult*).

References