5 Probability and random variables.

5.1 Finite probability spaces. A **finite probability space** is a finite set Ω of possible outcomes to some 'experiment', with probabilities (numbers) $P[\{\omega\}]$ assigned to each element $\omega \in \Omega$, satisfying

$$P[\{\omega\}] \geq 0$$
$$\sum_{\omega \in \Omega} P[\{\omega\}] = 1$$

For example, to represent 3 tosses of a fair coin we could set

$$\Omega = \{HHH, HHT, HTH, HTT \\ THH, THT, TTH, TTT\}$$

and assign a probability of $P[\{\omega\}] = 1/8$ to each element ω of Ω.

One interpretation is that the 'experiment' can be repeated many times, and $P[\{\omega\}]$ is the proportion of times outcome ω will occur in these repeated experiments.

An **event** is a **subset** of Ω and for any event $A \subseteq \Omega$ we write $P[A]$ for the probability of A which is given by summing the probabilities of each outcome in A, i.e.

$$P[A] = \sum_{\omega \in A} P[\{\omega\}].$$

Events can be described either by listing all outcomes in the event, or in words. e.g. in the above example:

$\{HHH, TTT\}$ is the event that all coins land the same way.

$\{HHH, HTH, THH, TTH\}$ is the event that the third coin toss lands on 'heads'.

If A is an event we write A^c for the complement of A, i.e. the event consisting of all elements of Ω that are *not* in A, or in words the event that A does not occur. eg if $A = \{HHH, HTH, THH, TTH\}$ is the event that the third coin toss lands on heads, then the complement A^c is the event $\{HHT, HTT, THT, TTT\}$, i.e. the event that the third coin toss lands on tails.

Given events A and B we say that $A \cap B$ occurs if both A and B occur. For example, if $A = \{HHH, TTT\}$ is the event that all coins land the same way, and $B = \{HHH, HTH, THH, TTH\}$ is the event that the third coin toss lands on 'heads', then $A \cap B = \{HHH\}$, i.e. the event $A \cap B$ is the event that A and B both occur, which is the same as the event that all 3 tosses are heads.

As well as the intersection $A \cap B$, we are sometimes interested in the union $A \cup B$. If A and B are events, then the event $A \cup B$ consists of all elements in either A or in B, or both. For example, if $A = \{HHH, TTT\}$ and $B = \{HHH, HTH, THH, TTH\}$, then

$$A \cup B = \{HHH, HTH, THH, TTH, TTT\}$$

is the event that either the third coin toss lands on heads or the coin tosses are all tails.

We say events A and B are *disjoint* or *incompatible* if $A \cap B$ is the empty set. In our example of 3 coin tosses, the following events are incompatible:

$B = \{HHH, HTH, THH, TTH\}$, the event that the last coin toss is heads.

$D = \{HTT, TTT\}$, the event that the second and third coin tosses are both tails.

It follows from the definition $P[A] = \sum_{\omega \in A} P[\{\omega\}]$ and the assumption that the numbers $P[\{\omega\}], \omega \in \Omega$ are nonnegative and sum to 1, that the following so-called **Axioms of probability** hold:

1. $P[A] \geq 0$ for all events A.
2. $P[\Omega] = 1$
3. If A and B are incompatible then

$$P[A \cup B] = P[A] + P[B]$$

We say events A and B are *independent* if $P[A \cap B] = P[A]P[B]$. In the above example, with $A = \{HHH, TTT\}$ and $B = \{HHH, HTH, THH, TTH\}$, it is the case that $P[A] = 1/4$ and $P[B] = 1/2$ while $P[A \cap B] = 1/8$ so A and B are indeed independent.

The interpretation of independence of A and B is that if we know event A occurs this does not affect the probability of B occurring.

We say events A, B, C are *mutually independent* if any pair of them are independent and also

$$P[A \cap B \cap C] = P[A]P[B]P[C].$$

Likewise for more than three events.

It can be shown that if A and B are independent, then A^c and B are independent. If A, B, C are mutually independent, then A^c, B, C are mutually independent, and so on.

Example. You invest in 3 companies (denoted a, b, c respectively). Companies a, b, c have probability 0.1, 0.2, 0.3 respectively of going bankrupt in the next 5 years (independently of each other). What is the probability that at most 1 of them goes bankrupt in the next 5 years?

Let A, B, C be the events that Company a, Company b, Company c respectively go bankrupt. So $P[A] = 0.1$, $P[B] = 0.2$, $P[C] = 0.3$, and A, B, C are assumed mutually independent. The probability that none of them goes bankrupt is

$$\begin{aligned} P[A^c \cap B^c \cap C^c] &= P[A^c]P[B^c]P[C^c] \\ &= (1 - 0.1)(1 - 0.2)(1 - 0.3) = 0.504 \end{aligned}$$

The probability that only A goes bankrupt is

$$\begin{aligned} P[A \cap B^c \cap C^c] &= P[A]P[B^c]P[C^c] \\ &= 0.1(1 - 0.2)(1 - 0.3) = 0.056 \end{aligned}$$

and similarly

$$\begin{aligned} P[A^c \cap B \cap C^c] &= P[A^c]P[B]P[C^c] \\ &= (1 - 0.1)(0.2)(1 - 0.3) = 0.126 \end{aligned}$$

$$\begin{aligned} P[A^c \cap B^c \cap C] &= P[A^c]P[B^c]P[C] \\ &= (1 - 0.1)(1 - 0.2)(0.3) = 0.216 \end{aligned}$$

so the answer is

$$\begin{aligned} &0.504 + 0.056 + 0.126 + 0.216 \\ &= 0.902 \end{aligned}$$

5.2 Bernoulli trials

The discrete probability space that will interest us the most is the *Bernoulli trials* model, which generalizes the coin-tossing example to an arbitrary number of coin tosses, and allows for the possibility that the coin is biased.

In general terms consider a sequence of n 'trials', each of which can end in 'success' or 'failure'. Let p be a parameter with $(0 < p < 1)$. In the Bernoulli trials model, it is assumed that each trial has probability p of success and that different trials are mutually independent. In the case $n = 3$, the sample space is similar to the earlier example:

$$\begin{aligned}\Omega = \{&SSS, SSF, SFS, SFF,\\ &FSS, FSF, FFS, FFF\}\end{aligned}$$

but now, given p and setting $q = 1 - p$, the probabilities of these outcomes are

$$\begin{aligned} P[\{SSS\}] &= p^3 \\ P[\{SSF\}] = P[\{SFS\}] = P[\{FSS\}] &= p^2 q \\ P[\{SFF\}] = P[\{FSF\}] = P[\{FFS\}] &= pq^2 \\ P[\{FFF\}] &= q^3 \end{aligned}$$

It can then be shown that if A_i denotes the event that the ith trial is a success, so that e.g.

$$A_2 = \{SSS, SSF, FSS, FSF\}$$

then

$$P[A_i] = p, \quad i = 1, 2, 3$$

and events A_1, A_2 and A_3 are mutually independent.

Similarly, if $n = 5$ then a typical element of the sample space is $SSFFS$ which has probability p^3q^2. Again, if A_i is the event that the ith trial is successful, it has probability p and events A_1, A_2, A_3, A_4, A_5 are mutually independent.

In general, for n Bernoulli trials with success probability p, the elements of the probability space are sequences of the form $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with each x_i equal to either S or F. An outcome $\mathbf{x} = (x_1, \ldots, x_n)$ with k of the x_i's equal to S and the other $n - k$ of them equal to F has probability $p^k(1-p)^{n-k}$.

Going back to the case $n = 3$, let E_2 denote the event that precisely two of the three trials are successful. Thus

$$E_2 = \{SSF, SFS, FSS\}$$

and each element of E has probability p^2q (with $q = 1 - p$). Hence

$$P[E_2] = \sum_{\omega \in E_2} P[\{\omega\}] = 3p^2q$$

so for example if $n = 3, p = 3/4$ then

$$P[E_2] = 3\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right) = \frac{27}{64}$$

So for three tosses of coin that is biased $75 : 25$ in favour of heads, the probability that precisely 2 of the tosses land heads is $27/64$.

Similarly, if E_1 is the event that just one of the three trials is successful, then

$$P[E_1] = P[\{SFF, FSF, FFS\}] = 3pq^2$$

while if E_0 is the event that no trial is a success and E_3 is the event that all three trials are successful, then

$$P[E_0] = P[\{FFF\}] = q^3$$
$$P[E_3] = P[\{SSS\}] = p^3$$

To take another example, consider 5 Bernoulli trials, and let E_3 denote the event that precisely 3 of them are successful. We can list the outcomes in E_3 as

$$\{SSSFF, SSFSF, SSFFS, SFSSF, SFSFS,$$
$$SFFSS, FSSSF, FSSFS, FSFSS, FFSSS\}$$

a total of 10 outcomes which each have probability p^3q^2. Hence for 5 trials,

$$P[E_3] = 10p^3q^2$$

In general, for n trials the probability of having precisely k successes is

$$P[E_k] = \binom{n}{k} p^k q^{n-k}$$

where $\binom{n}{k}$ denotes number of distinct sequences of k S's and $(n-k)$ F's. It turns out that $\binom{n}{k}$ is given by the formula

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

where we set $n! = 1 \times 2 \times \cdots \times n$ for $n \geq 1$, and we set $0! = 1$. For example,

$$\binom{5}{3} = \frac{5!}{3!2!}$$
$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1}$$
$$= \frac{20}{2} = 10.$$

Also, it can be shown that $\binom{n}{k}$ is the entry in the $(k+1)$st entry in the $(n+1)$st row of Pascal's triangle, where each entry is obtained by summing the two entries directly above it.

Example. Suppose each day the stock market either goes up (with probability 0.6) or down (with probability 0.4), and different days are assumed independent. What is the probabilty that the stock market goes up on precisely 2 days out of the next 4?

With $n = 4$ and $p = 0.6$ the answer is

$$P[E_2] = \binom{4}{2}(0.6)^2(0.4)^2 = 6(0.36)(0.16)$$
$$= 0.3456$$

5.3 Probability distributions

When a sample space Ω of a probability space is a set of **numbers**, the random outcome is called a **random quantity**. For example if you roll a die the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. In such cases the random outcome is often denoted by a capital letter such as X, and for each possible outcome x (i.e. each $x \in \Omega$) we write $P[X = x]$ for $P[\{x\}]$. Also we write $P[X \leq x]$ for $P[\{y \in \Omega : y \leq x\}$. More generally for any set of numbers A we write $P[X \in A]$ for $P[A]$.

In the above example, assuming the die is fair, if X denotes the score on the die we have

$$P[X = 4] = 1/6,$$
$$P[X \leq 2] = 1/3,$$
$$P[X \in \{1, 3, 5\}] = 1/2$$

and so on.

If the set of possible values for X is finite, we say X is a **discrete** random quantity. The **probability distribution** of a discrete random quantity X is given by listing the set of all possible values x for X and the associated probabilities $P[X = x]$, often in a table, for example

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

x	100	0	−20
$P(X = x)$	0.2	0.7	0.1

Given a sample space (not necessarily a set of numbers), we can also construct a random quantity by specifying a rule for assigning a numerical 'score' to each element of a sample space (i.e., a **function** from the sample space Ω to the set of (real) numbers). Such a rule is called a **random variable** (often also denoted by letters such as X). The resulting random quantity would then also be denoted by the same letter!

For example, consider 2 coin tosses (so $\Omega = \{HH, HT, TH, TT\}$) and define a random variable X which counts the number of heads (we tend to use letters such as X for random variables, too!). That is,

$$X(HH) = 2, X(TT) = 0,$$
$$X(HT) = 1, X(TH) = 1.$$

Assuming a fair coin, the resulting random quantity (also denoted X) has probability distribution given by its set of possible values $\{0, 1, 2\}$ and corresponding probabilities $P[X = 0] = 1/4$, $P[X = 1] = 1/2$, and $P[X = 2] = 1/4$.

To take another example, go back to the earlier example where X is the score on a'fair die'. We could construct a new random quantity Y by taking (for example) $Y = (X - 3)^2$. What is the probability distribution of Y?

Answer: The possible values for Y are
$\{0, 1, 4, 9\}$, and

$$P[Y = 0] = P[Y = 9] = 1/6$$
$$P[Y = 1] = P[Y = 4] = 1/3$$

Note that these probabilities are nonnegative and add to 1. This should always be the case for any random quantity.

For our purposes, the most important discrete probability distributions are the family of **Binomial** distributions. Given a nonnegative integer n and a number p between 0 and 1, a Binomial random quantity with parameters n and p is obtained by counting the total number of successes in n Bernoulli trials with probability p of success.

For example consider 3 Bernoulli trials, so

$$\begin{aligned}\Omega = \{&SSS, SSF, SFS, SFF,\\ &FSS, FSF, FFS, FFF\}\end{aligned}$$

and consider the random variable X which counts the number of successes, so that

$$X(SSS) = 3, X(SSF) = 2, X(SFF) = 1 \quad \text{etc.}$$

The corresponding random quantity is simply the 'number of successes in 3 bernoulli trials' with parameter p, and has a Binomial distribution with

$$\begin{aligned}P[X=0] = (1-p)^3,\ \ P[X=1] = 3p(1-p)^2,\\ P[X=2] = 3p^2(1-p),\ \ P[X=3] = p^3.\end{aligned}$$

In the particular case $p = 1/4$, this comes to

$$\begin{aligned}P[X=0] = 27/64, \quad P[X=1] = 27/64,\\ P[X=2] = 9/64, \quad P[X=3] = 1/64\end{aligned}$$

Note that these probabilities add to 1, which is as it should be.

Using the formula from earlier, we obtain that if X is Binomial with parameters n and p then

$$P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad (k = 0, 1, 2, \ldots, n)$$

Example. Suppose each day the stock market either goes up (with probability 0.6) or down (with probability 0.4), and different days are assumed independent. What is the probabilty that the stock market goes up on at least 2 days out of the next 4?

Let X be the number of days in the next 4 days when the stock market goes up. Then X is binomial with parameters $n = 4$ and $p = 0.6$, so

$$P[X \geq 2] = P[X=2] + P[X=3] + P[X=4]$$

We worked out earlier that $P[X=2] = 0.3456$. Also

$$\begin{aligned}P[X=3] &= \binom{4}{3}(0.6)^3(0.4)^1\\ &= 4(.216)0.4 = 0.3456\\ P[X=4] &= \binom{4}{4}(0.6)^4(0.4)^0\\ &= 1(0.1296)\times 1 = 0.1296\end{aligned}$$

so the answer is

$$0.3456 + 0.3456 + 0.1296 = 0.8208$$