Part D: Probabilistic Methods

13. Revision of Discrete Probability

This section is intended to remind you of the definitions and language of discrete probability theory, on the assumption that you have seen most of the ideas before. These notes are based on earlier notes by Dr Barnea and Dr Gerke; of course any errors are my responsibility.

For further background see any basic textbook on probability, for example Sheldon Ross, *A First Course in Probability*, Prentice Hall.

Definition 13.1.

- A *probability measure* p on a finite set Ω assigns a real number p_ω to each $\omega \in \Omega$ so that $0 \leq p_\omega \leq 1$ for each ω and
$$\sum_{\omega \in \Omega} p_\omega = 1.$$
We say that p_ω is the *probability of* ω.
- A *probability space* is a finite set Ω equipped with a probability measure. The elements of a probability space are sometimes called *outcomes*.
- An *event* is a subset of Ω.
- The *probability* of an event $A \subseteq \Omega$, denoted $\mathbf{P}[A]$ is the sum of the probability of the outcomes in A; that is
$$\mathbf{P}[A] = \sum_{\omega \in A} p_\omega.$$

Note that it follows from this definition that $\mathbf{P}[\{\omega\}] = p_\omega$ for each $\omega \in \Omega$. We also have $\mathbf{P}[\varnothing] = 0$ and $\mathbf{P}[\Omega] = 1$.

Example 13.2

(1) To model a throw of a single unbiased die, we take
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
and put $p_\omega = 1/6$ for each outcome $\omega \in \Omega$. The event that we throw an even number is $A = \{2, 4, 6\}$ and as expected, $\mathbf{P}[A] = p_2 + p_4 + p_6 = 1/6 + 1/6 + 1/6 = 1/2$.

(2) To model a throw of a pair of dice we could take
$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$
and give each element of Ω probability $1/36$, so $p_{(i,j)} = 1/36$ for all $(i, j) \in \Omega$. Alternatively, if we know that we only care about the sum of the two dice, we could take $\Omega = \{2, 3, \ldots, 12\}$ with $p_2 = 1/36$, $p_3 = 2/36$, $\ldots$, $p_6 = 5/36$, $p_7 = 6/36$, $p_8 = 5/36$, $\ldots$, $p_{12} = 1/36$.

(3) A suitable probability space for three flips of a coin is

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

where H stands for heads and T for tails, and each outcome has probability $1/8$. To allow for a biased coin we fix $0 \leq q \leq 1$ and instead give an outcome with exactly k heads probability $q^k(1-q)^{3-k}$.

(4) Let $n \in \mathbf{N}$ and let Ω be the set of all permutations of $\{1, 2, \ldots, n\}$. Set $p_\sigma = 1/n!$ for each permutation $\sigma \in \Omega$. This gives a suitable setup for Theorem 2.6. Later we shall use the language of probability theory to give a shorter proof of part (ii) of this theorem.

It will often be helpful to specify events (i.e. subsets of Ω) a little informally. For example, in (3) above we might write $\mathbf{P}$[at least two heads], rather than $\mathbf{P}[\{HHT, HTH, THH, HHH\}]$.

Unions, intersections and complements. Let Ω be a probability space. If A, $B \subseteq \Omega$ then

$$\begin{aligned}\mathbf{P}[A \cup B] = \sum_{\omega \in A \cup B} p_\omega = \sum_{\omega \in A} p_\omega + \sum_{\omega \in B} p_\omega - \sum_{\omega \in A \cap B} p_\omega \\ = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B].\end{aligned}$$

In particular, if A and B are disjoint, i.e. $A \cap B = \varnothing$, then $P[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B]$. The *complement* of an event $A \subseteq \Omega$ is defined to be

$$\bar{A} = \{\omega \in \Omega : \omega \notin A\}.$$

Since

$$1 = \mathbf{P}[\Omega] = \mathbf{P}[A \cup \bar{A}] = \mathbf{P}[A] + \mathbf{P}[\bar{A}]$$

we have $\mathbf{P}[\bar{A}] = 1 - \mathbf{P}[A]$.

Exercise: Restate the Principle of Inclusion and Exclusion (Theorem 5.3) so that it becomes a result about probabilities.

Condition probability and independence.

Definition 13.3. Let Ω be a probability space, and let A, $B \subseteq \Omega$ be events.

- If $\mathbf{P}[B] \neq 0$ then we define the *conditional probability of A given B* by
$$\mathbf{P}[A|B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]}.$$
- The events A, B are said to be *independent* if $\mathbf{P}[A \cap B] = \mathbf{P}[A]\mathbf{P}[B]$.

Suppose that each element of Ω has equal probability p. Then

$$\mathbf{P}[A|B] = \frac{|A \cap B|p}{|B|p} = \frac{|A \cap B|}{|B|}$$

is the proportion of elements of B that also lie in A; informally, if we know that the event B has occurred, then the probability that A has also occurred is $\mathbf{P}[A|B]$.

Exercise: Show that if A and B are events in a probability space such that $\mathbf{P}[A]$, $\mathbf{P}[B] \neq 0$, then $\mathbf{P}[A|B] = \mathbf{P}[A]$ if and only if A and B are independent.

Conditional probability can be quite subtle.

Exercise: Let $\Omega = \{HH, HT, TH, TT\}$ be the probability space for two flips of a fair coin, as in Example 13.2(3). Let A be the event that both flips are heads, and let B be the event that at least one flip is a head. Write A and B as subsets of Ω and show that $\mathbf{P}[A|B] = 1/3$.

Example 13.4 (The Monty Hall Problem)**.** On a game show you are offered the choice of three doors. Behind one door is a car, and behind the other two are goats. You pick a door and then the host, *who knows where the car is*, opens another door to reveal a goat. You may then either open your original door, or change to the remaining unopened door. Assuming you want the car, should you change?

Most people find the answer to the Monty Hall problem a little surprising. The Sleeping Beauty Problem, stated below, is even more controversial.

Example 13.5. Beauty is told that if a coin lands heads she will be woken on Monday and Tuesday mornings, but after being woken on Monday she will be given an amnesia inducing drug, so that she will have no memory of what happened that day. If the coin lands tails she will only be woken on Tuesday morning. At no point in the experiment is Beauty told what day it is. Imagine that you are Beauty and are awoken as part of the experiment and asked for your credence that the coin landed heads. What is your answer?

The related statistical issue in the next example is also widely misunderstood.

Example 13.6. Suppose that one in every 1000 people has disease X. There is a new test for X that will always identify the disease in anyone who has it. There is, unfortunately, a tiny probability of $1/250$ that the test will falsely report that a healthy person has the disease. What is the probability that a person who tests positive for X actually has the disease?

Random variables.

Definition 13.7. Let Ω be a probability space. A *random variable* on Ω is a function $X : \Omega \to \mathbf{R}$. If X, $Y : \Omega \to \mathbf{R}$ are random variables then we say that X and Y are *independent* if for all x, $y \in \mathbf{R}$ the events

$$A = \{\omega \in \Omega : X(\omega) = x\} \quad \text{and}$$
$$B = \{\omega \in \Omega : Y(\omega) = y\}$$

are independent.

The following shorthand notation is very useful. If $X : \Omega \to \mathbf{R}$ is a random variable, then '$X = x$' is the event $\{\omega \in \Omega : X(\omega) = x\}$. Similarly '$X \geq x$' is the event $\{\omega \in \Omega : X(\omega) \geq x\}$. We mainly use this shorthand in probabilities, so for instance

$$\mathbf{P}[X = x] = \mathbf{P}\big[\{\omega \in \Omega : X(\omega) = x\}\big].$$

Exercise: Show that X, $Y : \Omega \to \mathbf{R}$ are independent if and only if

$$\mathbf{P}[(X = x) \cap (Y = y)] = \mathbf{P}[X = x]\mathbf{P}[Y = y]$$

for all $x, y \in \mathbf{R}$. (This is just a trivial restatement of the definition.)

Example 13.8. Let $\Omega = \{HH, HT, TH, TT\}$ be the probability space for two flips of a fair coin. Define $X : \Omega \to \mathbf{R}$ to be 1 if the first coin is heads, and zero otherwise. So

$$X(HH) = X(HT) = 1 \quad \text{and} \quad X(TH) = X(TT) = 0.$$

Define $Y : \Omega \to \mathbf{R}$ similarly for the second coin.

(i) The random variables X and Y are independent.

(ii) Let Z be 1 if exactly one flip is heads, and zero otherwise. Then X and Z are independent, and Y and Z are independent.

(iii) There exist $x, y, z \in \{0, 1\}$ such that

$$\mathbf{P}[X = x, Y = y, Z = z] \neq \mathbf{P}[X = x]\mathbf{P}[Y = y]\mathbf{P}[Z = z].$$

This shows that one has to be quite careful when defining independence for a family of random variables. (Except in the Lovász Local Lemma, we will be able to manage with the pairwise independence defined above.)

Given random variables X, $Y : \Omega \to \mathbf{R}$ we can define new random variables by taking functions such as $X + Y$, aX for $a \in \mathbf{R}$ and XY. We notice that if $z \in \mathbf{R}$ then

$$\{\omega \in \Omega : (X + Y)(\omega) = z\} = \bigcup_{x+y=z} \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}.$$

The events above are disjoint for different x, y, so we get

$$\mathbf{P}[X + Y = z] = \sum_{x+y=z} \mathbf{P}[(X = x) \cap (Y = y)].$$

If X and Y are independent then

$$\mathbf{P}[(X = x) \cap (Y = y)] = \mathbf{P}[X = x]\mathbf{P}[Y = y]$$

and so

$$\mathbf{P}[X + Y = z] = \sum_{x+y=z} \mathbf{P}[X = x]\mathbf{P}[Y = y].$$

(Note that all of these sums have only finitely many non-zero summands, so they are well-defined.)

Exercise: Show similarly that if X, $Y : \Omega \to \mathbf{R}$ are independent random variables then

$$\mathbf{P}[XY = z] = \sum_{xy=z} \mathbf{P}[X = x]\mathbf{P}[Y = y].$$

Expectation and linearity.

Definition 13.9. Let Ω be a probability space with probability measure p. The *expectation* $\mathbf{E}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) p_w.$$

Intuitively, the expectation of X is the average value of X on elements of Ω, if we choose $\omega \in \Omega$ with probability p_ω. Note that

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) p_\omega = \sum_{x \in R} \sum_{\substack{\omega \\ X(\omega)=x}} x p_\omega = \sum_{x \in \mathbf{R}} x \mathbf{P}[X = x].$$

A critical property of expectation is that it is linear. Note that we *do not need to assume independence* in this lemma.

Lemma 13.10. *Let Ω be a probability space. If $X_1, X_2, \ldots, X_k : \Omega \to \mathbf{R}$ are random variables then*

$$\mathbf{E}[a_1 X_1 + a_2 X_2 + \cdots + a_k X_k] = a_1 \mathbf{E}[X_1] + a_2 \mathbf{E}[X_2] + \cdots + a_k \mathbf{E}[X_k]$$

for any $a_1, a_2, \ldots, a_k \in \mathbf{R}$.

Proof. By definition the left-hand side is

$$\begin{aligned} \sum_{\omega \in \Omega} p_\omega \big(a_1 X_1 + \cdots + a_k X_k)(\omega) &= \sum_{\omega \in \Omega} p_\omega \big(a_1 X_1(\omega) + \cdots + a_k X_K(\omega)\big) \\ &= a_1 \sum_{\omega \in \Omega} p_\omega X_1(\omega) + \cdots + a_k \sum_{\omega \in \Omega} X_k(\omega) \end{aligned}$$

which is the right-hand side. □

When X, $Y : \Omega \to \mathbf{R}$ are independent random variables, there is a very useful formula for $\mathbf{E}[XY]$.

Lemma 13.11. *If X, $Y : \Omega \to \mathbf{R}$ are independent random variables then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Exercise: Prove Lemma 13.11 by arguing that

$$\mathbf{E}[XY] = \sum_{z \in \mathbf{R}} z\mathbf{P}[XY = z] = \sum_{z \in R} z \sum_{xy=z} \mathbf{P}[(X = x) \cap (Y = y)]$$

and using independence.

Variance.

Definition 13.12. Let Ω be a probability space. The *variance* $\mathbf{Var}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

$$\mathbf{Var}[X] = \mathbf{E}\big[(X - \mathbf{E}[X])^2\big].$$

The variance measures how much X can be expected to depart from its mean value $\mathbf{E}[X]$. So it is a measure of the 'spread' of X.

It is tempting to define the variance as $\mathbf{E}\big[X - \mathbf{E}[X]\big]$, but by linearity this expectation is $\mathbf{E}[X] - \mathbf{E}[X] = 0$. One might also consider the quantity $\mathbf{E}\big[\,\big|X - \mathbf{E}[X]\big|\,\big]$, but the absolute value turns out to be hard to work with. The definition above works well in practice.

Lemma 13.13. *Let Ω be a probability space.*

(i) *If $X : \Omega \to \mathbf{R}$ is a random variable then*

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

(ii) *If X, $Y : \Omega \to \mathbf{R}$ are independent random variables then*

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Exercise: Show that (ii) can fail if X and Y are not independent. [*Hint: usually a random variable is not independent of itself.*]