Part D: Probabilistic methods

16. 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.

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

Definition 16.1.

- A *probability measure* p on a finite set Ω assigns a real number p_ω to each $\omega \in \Omega$, so that $0 \leq p_\omega \leq 1$ 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 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 probabilities of the elements of A; that is, $\mathbf{P}[A] = \sum_{\omega \in A} p_\omega$.

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

Example 16.2.

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

(2) For 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 $1/36$. Alternatively, if we only care about the sum, we could take $\Omega = \{2, 3, \ldots, 12\}$ with probabilities $p_2 = 1/36$, $p_3 = 2/36$, $\ldots$, $p_6 = 5/36$, $p_7 = 6/36$, $p_8 = 5/36$, $\ldots$, $p_{12} = 1/36$.

(3) Suppose we repeatedly flip a coin. For three flips, a suitable probability space would be

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

where H stands for heads, and T for tails. To allow for a biased coin, we fix $0 \leq q \leq 1$ and give each sequence with exactly k H's probability $q^k(1-q)^{3-k}$.

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

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

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

$$\mathbf{P}[A \cup B] = \sum_{\omega \in A \cup B} p_\omega = \sum_{\omega \in A} p_\omega + \sum_{\omega \in B} p_\omega = \mathbf{P}[A] + \mathbf{P}[B].$$

More generally, for any $A, B \subseteq \Omega$ we have

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

The *complement* of an event A is

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

Since

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

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

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

Conditional probability and independence.

Definition 16.3. Let Ω be a probability space and let $A, B \subseteq \Omega$ be events.

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

The justification for the definition of $\mathbf{P}[A|B]$ is that it defines the probability that A occurs if we already know that B occurred. Note that if $\mathbf{P}[B] \neq 0$, then A and B are independent events if and only if $\mathbf{P}[A|B] = \mathbf{P}[A]$. In other words, the probability of A does not change if we know that B occurred.

Conditional probability can be quite subtle.

Exercise: Show that if $\mathbf{P}[A]$, $\mathbf{P}[B] \neq 0$, then $\mathbf{P}[A|B] = \mathbf{P}[A]$ if and only if $\mathbf{P}[B|A] = \mathbf{P}[B]$. Do you find this intuitively reasonable?

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

Example 16.4 (The Monty Hall Problem)**.** On a game show you are offered the choice of three doors. Behind one of the doors is a desirable prize (conventionally, a car) and behind the others, something valueless (conventionally, goats). You pick a door and then the host, *who knows where the prize is*, opens another door to reveal a goat. You may then either open your original door, or change to the other unopened door. Should you change?

Many people find the answer to the Monty Hall Problem a little surprising. The Sleeping Beauty Problem, in which the sleeper is woken either once or twice according to the toss of a coin, and after each waking is asked for her credence that the coin landed heads, is still controversial.

Random variables.

Definition 16.5. Let Ω be a probability space. A *random variable* 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 will be very useful. If $X : \Omega \to \mathbf{R}$ is a random variable then we write $X = x$ for $\{\omega \in \Omega : X(\omega) = x\}$. We mainly use this shorthand in probabilities: thus

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

Example 16.6. 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 $X(HH) = X(HT) = 1$ and $X(TH) = X(TT) = 0$. Define Y similarly for the second coin.

(i) The random variables X and Y are independent.
(ii) Let Z be 1 if exactly one flip results in heads, and 0 otherwise. Then X and Z are independent, and Y and Z are independent.
(iii) There exist $x, y, z \in \mathbf{R}$ 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 of a family of random variables. (Except in the Lovász Local Lemma we will be able to manage with the pairwise independence defined above.)

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 $x, y \in \mathbf{R}$.

We can also define new random variables by looking at functions like $X + Y$, aX for $a \in \mathbf{R}$ and XY. The constant functions $c : \Omega \to \mathbf{R}$ defined by $c(\omega) = c$ for all $\omega \in \Omega$ will also be useful. We notice that

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

The above events are disjoint for different x and 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 we have

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

Note that in the sums above we sum over all $x, y \in \mathbf{R}$. However, there are only a finite number of x's and y's such that

$$\{\omega \in \Omega : X(\omega)=x, Y(\omega)=y\} \neq \varnothing.$$

Therefore the non-zero contributions come from finitely many terms, and so the sums make sense.

Exercise: Show similarly that

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

and that if X and Y are independent, then

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

Expectation and linearity.

Definition 16.7. Let Ω be a probability space with probability p. The *expectation* $\mathbf{E}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

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

Intuitively the expectation represents the average value of X on elements of Ω, if we pick many elements independently, choosing $\omega \in \Omega$ with probability p_ω. Note that

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

A critical property of expectation is that it is linear. This is easily proved, but can be very useful in the problems we shall deal with. Note that we do *not assume independence in this lemma.*

Lemma 16.8. *Let Ω be a probability space. If $X_1, X_2, \ldots, X_n : \Omega \to \mathbf{R}$ are random variables then*

$$\mathbf{E}[a_1X + a_2X_2 + \cdots + a_nX_n] = a_1\mathbf{E}[X_1] + a_2\mathbf{E}[X_2] + \cdots + a_n\mathbf{E}[X_n]$$

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

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

Lemma 16.9. *If $X, Y : \Omega \to \mathbf{R}$ are independent random variables then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Exercise: prove the lemma by arguing that

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

and then using independence to factorise the right-hand side.

Variance.

Definition 16.10. Let Ω be a probability space. The *variance* $\mathbf{Var}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

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

The variance measures the expected difference of X from its average value. 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 expression is always zero. One might also try $\mathbf{E}\big[|X - \mathbf{E}[X]|\big]$, but the problem is that the absolute value is hard to work with. The definition above works well in practice.

Lemma 16.11. *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.]