Preface

Here are the course lecture notes for the course MAS108, Probability I, at Queen Mary, University of London, taken by most Mathematics students and some others in the first semester.

The description of the course is as follows:

> This course introduces the basic notions of probability theory and develops them to the stage where one can begin to use probabilistic ideas in statistical inference and modelling, and the study of stochastic processes. Probability axioms. Conditional probability and independence. Discrete random variables and their distributions. Continuous distributions. Joint distributions. Independence. Expectations. Mean, variance, covariance, correlation. Limiting distributions.

The syllabus is as follows:

1. Basic notions of probability. Sample spaces, events, relative frequency, probability axioms.

2. Finite sample spaces. Methods of enumeration. Combinatorial probability.

3. Conditional probability. Theorem of total probability. Bayes theorem.

4. Independence of two events. Mutual independence of n events. Sampling with and without replacement.

5. Random variables. Univariate distributions - discrete, continuous, mixed. Standard distributions - hypergeometric, binomial, geometric, Poisson, uniform, normal, exponential. Probability mass function, density function, distribution function. Probabilities of events in terms of random variables.

6. Transformations of a single random variable. Mean, variance, median, quantiles.

7. Joint distribution of two random variables. Marginal and conditional distributions. Independence.

mobile phone. Then

$$\begin{aligned} P(A) &= 52/104 = 0.5, \\ P(B) &= 87/104 = 0.8365, \\ P(C) &= 83/104 = 0.7981. \end{aligned}$$

Furthermore,

$$\begin{aligned} P(A \cap B) = 45/104 = 0.4327, &\qquad P(A) \cdot P(B) = 0.4183, \\ P(A \cap C) = 39/104 = 0.375, &\qquad P(A) \cdot P(C) = 0.3990, \\ P(B \cap C) = 70/104 = 0.6731, &\qquad P(B) \cap P(C) = 0.6676. \end{aligned}$$

So none of the three pairs is independent, but in a sense B and C 'come closer' than either of the others, as we noted.

In practice, if it is the case that the event A has no effect on the outcome of event B, then A and B are independent. But this does not apply in the other direction. There might be a very definite connection between A and B, but still it could happen that $P(A \cap B) = P(A) \cdot P(B)$, so that A and B are independent. We will see an example shortly.

Example If we toss a coin more than once, or roll a die more than once, then you may assume that different tosses or rolls are independent. More precisely, if we roll a fair six-sided die twice, then the probability of getting 4 on the first throw and 5 on the second is $1/36$, since we assume that all 36 combinations of the two throws are equally likely. But $(1/36) = (1/6) \cdot (1/6)$, and the separate probabilities of getting 4 on the first throw and of getting 5 on the second are both equal to $1/6$. So the two events are independent. This would work just as well for any other combination.

In general, it is always OK to assume that the outcomes of different tosses of a coin, or different throws of a die, are independent. This holds even if the examples are not all equally likely. We will see an example later.

Example I have two red pens, one green pen, and one blue pen. I choose two pens without replacement. Let A be the event that I choose exactly one red pen, and B the event that I choose exactly one green pen.

If the pens are called R_1, R_2, G, B, then

$$\begin{aligned} \mathcal{S} &= \{R_1R_2, R_1G, R_1B, R_2G, R_2B, GB\}, \\ A &= \{R_1G, R_1B, R_2G, R_2B\}, \\ B &= \{R_1G, R_2G, GB\} \end{aligned}$$

Example We can always assume that successive tosses of a coin are mutually independent, even if it is not a fair coin. Suppose that I have a coin which has probability 0.6 of coming down heads. I toss the coin three times. What are the probabilities of getting three heads, two heads, one head, or no heads?

For three heads, since successive tosses are mutually independent, the probability is $(0.6)^3 = 0.216$.

The probability of tails on any toss is $1 - 0.6 = 0.4$. Now the event 'two heads' can occur in three possible ways, as HHT, HTH, or THH. Each outcome has probability $(0.6) \cdot (0.6) \cdot (0.4) = 0.144$. So the probability of two heads is $3 \cdot (0.144) = 0.432$.

Similarly the probability of one head is $3 \cdot (0.6) \cdot (0.4)^2 = 0.288$, and the probability of no heads is $(0.4)^3 = 0.064$.

As a check, we have

$$0.216 + 0.432 + 0.288 + 0.064 = 1.$$

1.13 Worked examples

Question

(a) You go to the shop to buy a toothbrush. The toothbrushes there are red, blue, green, purple and white. The probability that you buy a red toothbrush is three times the probability that you buy a green one; the probability that you buy a blue one is twice the probability that you buy a green one; the probabilities of buying green, purple, and white are all equal. You are certain to buy exactly one toothbrush. For each colour, find the probability that you buy a toothbrush of that colour.

(b) James and Simon share a flat, so it would be confusing if their toothbrushes were the same colour. On the first day of term they both go to the shop to buy a toothbrush. For each of James and Simon, the probability of buying various colours of toothbrush is as calculated in (a), and their choices are independent. Find the probability that they buy toothbrushes of the same colour.

(c) James and Simon live together for three terms. On the first day of each term they buy new toothbrushes, with probabilities as in (b), independently of what they had bought before. This is the only time that they change their toothbrushes. Find the probablity that James and Simon have differently coloured toothbrushes from each other for all three terms. Is it more likely that they will have differently coloured toothbrushes from each other for all three terms or that they will sometimes have toothbrushes of the same colour?

Solution

(a) Let R,B,G,P,W be the events that you buy a red, blue, green, purple and white toothbrush respectively. Let $x = P(G)$. We are given that

$$P(R) = 3x, \quad P(B) = 2x, \quad P(P) = P(W) = x.$$

Since these outcomes comprise the whole sample space, Corollary 2 gives

$$3x + 2x + x + x + x = 1,$$

so $x = 1/8$. Thus, the probabilities are $3/8$, $1/4$, $1/8$, $1/8$, $1/8$ respectively.

(b) Let RB denote the event 'James buys a red toothbrush and Simon buys a blue toothbrush', etc. By independence (given), we have, for example,

$$P(RR) = (3/8) \cdot (3/8) = 9/64.$$

The event that the toothbrushes have the same colour consists of the five outcomes RR, BB, GG, PP, WW, so its probability is

$$\begin{aligned} & P(RR) + P(BB) + P(GG) + P(PP) + P(WW) \\ = {} & \frac{9}{64} + \frac{1}{16} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} = \frac{1}{4}. \end{aligned}$$

(c) The event 'different coloured toothbrushes in the ith term' has probability $3/4$ (from part (b)), and these events are independent. So the event 'different coloured toothbrushes in all three terms' has probability

$$\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64}.$$

The event 'same coloured toothbrushes in at least one term' is the complement of the above, so has probability $1 - (27/64) = (37)/(64)$. So it is more likely that they will have the same colour in at least one term.

Question There are 24 elephants in a game reserve. The warden tags six of the elephants with small radio transmitters and returns them to the reserve. The next month, he randomly selects five elephants from the reserve. He counts how many of these elephants are tagged. Assume that no elephants leave or enter the reserve, or die or give birth, between the tagging and the selection; and that all outcomes of the selection are equally likely. Find the probability that exactly two of the selected elephants are tagged, giving the answer correct to 3 decimal places.

Solution The experiment consists of picking the five elephants, *not* the original choice of six elephants for tagging. Let $\mathcal{S}$ be the sample space. Then $|\mathcal{S}| = {}^{24}C_5$.

Let A be the event that two of the selected elephants are tagged. This involves choosing two of the six tagged elephants and three of the eighteen untagged ones, so $|A| = {}^{6}C_2 \cdot {}^{18}C_3$. Thus

$$P(A) = \frac{{}^{6}C_2 \cdot {}^{18}C_3}{{}^{24}C_5} = 0.288$$

to 3 d.p.

Note: Should the sample should be ordered or unordered? Since the answer doesn't depend on the order in which the elephants are caught, an unordered sample is preferable. If you want to use an ordered sample, the calculation is

$$P(A) = \frac{{}^{6}P_2 \cdot {}^{18}P_3 \cdot {}^{5}C_2}{{}^{24}P_5} = 0.288,$$

since it is necessary to multiply by the ${}^{5}C_2$ possible patterns of tagged and untagged elephants in a sample of five with two tagged.

Question A couple are planning to have a family. They decide to stop having children *either* when they have two boys *or* when they have four children. Suppose that they are successful in their plan.

(a) Write down the sample space.

(b) Assume that, each time that they have a child, the probability that it is a boy is $1/2$, independent of all other times. Find $P(E)$ and $P(F)$ where $E =$ "there are at least two girls", $F =$ "there are more girls than boys".

Solution (a) $\mathcal{S} = \{BB, BGB, GBB, BGGB, GBGB, GGBB, BGGG, GBGG, GGBG, GGGB, GGGG\}$.

(b) $E = \{BGGB, GBGB, GGBB, BGGG, GBGG, GGBG, GGGB, GGGG\}$,
$F = \{BGGG, GBGG, GGBG, GGGB, GGGG\}$.

Now we have $P(BB) = 1/4$, $P(BGB) = 1/8$, $P(BGGB) = 1/16$, and similarly for the other outcomes. So $P(E) = 8/16 = 1/2$, $P(F) = 5/16$.

By the Theorem of Total Probability,

$$\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3) \\ &= (1/3) \times (1/2) + (2/3) \times (1/4) + (2/3) \times (1/4) \\ &= 1/2. \end{aligned}$$

We have reached by a roundabout argument a conclusion which you might think to be obvious. If we have no information about the first pen, then the second pen is equally likely to be any one of the four, and the probability should be $1/2$, just as for the first pen. This argument happens to be correct. But, until your ability to distinguish between correct arguments and plausible-looking false ones is very well developed, you may be safer to stick to the calculation that we did. Beware of obvious-looking arguments in probability! Many clever people have been caught out.

2.5 Bayes' Theorem

There is a very big difference between $P(A \mid B)$ and $P(B \mid A)$.

Suppose that a new test is developed to identify people who are liable to suffer from some genetic disease in later life. Of course, no test is perfect; there will be some carriers of the defective gene who test negative, and some non-carriers who test positive. So, for example, let A be the event 'the patient is a carrier', and B the event 'the test result is positive'.

The scientists who develop the test are concerned with the probabilities that the test result is wrong, that is, with $P(B \mid A')$ and $P(B' \mid A)$. However, a patient who has taken the test has different concerns. If I tested positive, what is the chance that I have the disease? If I tested negative, how sure can I be that I am not a carrier? In other words, $P(A \mid B)$ and $P(A' \mid B')$.

These conditional probabilities are related by *Bayes' Theorem*:

Theorem 2.4 *Let A and B be events with non-zero probability. Then*

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}.$$

The proof is not hard. We have

$$P(A \mid B) \cdot P(B) = P(A \cap B) = P(B \mid A) \cdot P(A),$$

using the definition of conditional probability twice. (Note that we need both A and B to have non-zero probability here.) Now divide this equation by $P(B)$ to get the result.

If $P(A) \neq 0, 1$ and $P(B) \neq 0$, then we can use Corollary 17 to write this as

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid A') \cdot P(A')}.$$

Bayes' Theorem is often stated in this form.

Example Consider the ice-cream salesman from Section 2.3. Given that he sold all his stock of ice-cream, what is the probability that the weather was sunny? (This question might be asked by the warehouse manager who doesn't know what the weather was actually like.) Using the same notation that we used before, A_1 is the event 'it is sunny' and B the event 'the salesman sells all his stock'. We are asked for $P(A_1 \mid B)$. We were given that $P(B \mid A_1) = 0.9$ and that $P(A_1) = 0.3$, and we calculated that $P(B) = 0.59$. So by Bayes' Theorem,

$$P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B)} = \frac{0.9 \times 0.3}{0.59} = 0.46$$

to 2 d.p.

Example Consider the clinical test described at the start of this section. Suppose that 1 in 1000 of the population is a carrier of the disease. Suppose also that the probability that a carrier tests negative is 1%, while the probability that a non-carrier tests positive is 5%. (A test achieving these values would be regarded as very successful.) Let A be the event 'the patient is a carrier', and B the event 'the test result is positive'. We are given that $P(A) = 0.001$ (so that $P(A') = 0.999$), and that

$$P(B \mid A) = 0.99, \qquad P(B \mid A') = 0.05.$$

(a) A patient has just had a positive test result. What is the probability that the patient is a carrier? The answer is

$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A')P(A')} \\ &= \frac{0.99 \times 0.001}{(0.99 \times 0.001) + (0.05 \times 0.999)} \\ &= \frac{0.00099}{0.05094} = 0.0194. \end{aligned}$$

(b) A patient has just had a negative test result. What is the probability that the patient is a carrier? The answer is

$$P(A \mid B') = \frac{P(B' \mid A)P(A)}{P(B' \mid A)P(A) + P(B' \mid A')P(A')}$$

$$
\begin{aligned}
&= \frac{1}{4}\cdot\frac{P(S_1\cap S_2)}{P(S_2)}\cdot\frac{P(H_1\cap H_2)}{P(H_2)}\\
&= \frac{1}{4}\cdot P(S_1\mid S_2)\cdot P(H_1\mid H_2)\\
&= \frac{1}{4}\cdot\frac{2}{3}\cdot\frac{2}{51}\\
&= \frac{1}{153}.
\end{aligned}
$$

I thank Eduardo Mendes for pointing out a mistake in my previous solution to this problem.

Question The Land of Nod lies in the monsoon zone, and has just two seasons, Wet and Dry. The Wet season lasts for $1/3$ of the year, and the Dry season for $2/3$ of the year. During the Wet season, the probability that it is raining is $3/4$; during the Dry season, the probability that it is raining is $1/6$.

(a) I visit the capital city, Oneirabad, on a random day of the year. What is the probability that it is raining when I arrive?

(b) I visit Oneirabad on a random day, and it is raining when I arrive. *Given this information*, what is the probability that my visit is during the Wet season?

(c) I visit Oneirabad on a random day, and it is raining when I arrive. *Given this information*, what is the probability that it will be raining when I return to Oneirabad in a year's time?

(You may assume that in a year's time the season will be the same as today but, given the season, whether or not it is raining is independent of today's weather.)

Solution (a) Let W be the event 'it is the wet season', D the event 'it is the dry season', and R the event 'it is raining when I arrive'. We are given that $P(W) = 1/3$, $P(D) = 2/3$, $P(R \mid W) = 3/4$, $P(R \mid D) = 1/6$. By the ToTP,

$$
\begin{aligned}
P(R) &= P(R\mid W)P(W)+P(R\mid D)P(D)\\
&= (3/4)\cdot(1/3)+(1/6)\cdot(2/3) = 13/36.
\end{aligned}
$$

(b) By Bayes' Theorem,

$$P(W\mid R) = \frac{P(R\mid W)P(W)}{P(R)} = \frac{(3/4)\cdot(1/3)}{13/36} = \frac{9}{13}.$$

Binomial random variable $\mathrm{Bin}(n,p)$

Remember that for a Bernoulli random variable, we describe the event $X = 1$ as a 'success'. Now a *binomial random variable* counts the number of successes in n independent trials each associated with a Bernoulli(p) random variable.

For example, suppose that we have a biased coin for which the probability of heads is p. We toss the coin n times and count the number of heads obtained. This number is a $\mathrm{Bin}(n,p)$ random variable.

A $\mathrm{Bin}(n,p)$ random variable X takes the values $0,1,2,\ldots,n$, and the p.m.f. of X is given by

$$P(X = k) = {}^nC_k q^{n-k} p^k$$

for $k = 0,1,2,\ldots,n$, where $q = 1-p$. This is because there are ${}^nC_k$ different ways of obtaining k heads in a sequence of n throws (the number of choices of the k positions in which the heads occur), and the probability of getting k heads and $n-k$ tails in a particular order is $q^{n-k}p^k$.

Note that we have given a formula rather than a table here. For small values we could tabulate the results; for example, for $\mathrm{Bin}(4,p)$:

k	0	1	2	3	4
$P(X=k)$	q^4	$4q^3p$	$6q^2p^2$	$4qp^3$	p^4

Note: when we add up all the probabilities in the table, we get

$$\sum_{k=0}^{n} {}^nC_k q^{n-k} p^k = (q+p)^n = 1$$

as it should be: here we used the *binomial theorem*

$$(x+y)^n = \sum_{k=0}^{n} {}^nC_k x^{n-k} y^k.$$

(This argument explains the name of the binomial random variable!)

If $X \sim \mathrm{Bin}(n,p)$, then

$$E(X) = np, \qquad \mathrm{Var}(X) = npq.$$

There are two ways to prove this, an easy way and a harder way. The easy way only works for the binomial, but the harder way is useful for many random variables. However, you can skip it if you wish: I have set it in smaller type for this reason.

Here is the easy method. We have a coin with probability p of coming down heads, and we toss it n times and count the number X of heads. Then X is our $\mathrm{Bin}(n,p)$ random variable. Let X_k be the random variable defined by

$$X_k = \begin{cases} 1 & \text{if we get heads on the } k\text{th toss,} \\ 0 & \text{if we get tails on the } k\text{th toss.} \end{cases}$$

In other words, X_i is the indicator variable of the event 'heads on the kth toss'. Now we have

$$X = X_1 + X_2 + \cdots + X_n$$

(can you see why?), and $X_1, \ldots, X_n$ are *independent* Bernoulli(p) random variables (since they are defined by different tosses of a coin). So, as we saw earlier, $E(X_i) = p$, $\mathrm{Var}(X_i) = pq$. Then, by Theorem 21, since the variables are independent, we have

$$\begin{aligned} E(X) &= p + p + \cdots + p = np, \\ \mathrm{Var}(X) &= pq + pq + \cdots + pq = npq. \end{aligned}$$

The other method uses a gadget called the *probability generating function*. We only use it here for calculating expected values and variances, but if you learn more probability theory you will see other uses for it. Let X be a random variable whose values are non-negative integers. (We don't insist that it takes all possible values; this method is fine for the binomial $\mathrm{Bin}(n,p)$, which takes values between 0 and n. To save space, we write p_k for the probability $P(X = k)$. Now the *probability generating function* of X is the power series

$$G_X(x) = \sum p_k x^k.$$

(The sum is over all values k taken by X.)

We use the notation $[F(x)]_{x=1}$ for the result of substituting $x = 1$ in the series $F(x)$.

Proposition 3.4 *Let $G_X(x)$ be the probability generating function of a random variable X. Then*

(a) $[G_X(x)]_{x=1} = 1$;

(b) $E(X) = \left[\frac{d}{dx} G_X(x)\right]_{x=1}$;

(c) $\mathrm{Var}(X) = \left[\frac{d^2}{dx^2} G_X(x)\right]_{x=1} + E(X) - E(X)^2$.

Part (a) is just the statement that probabilities add up to 1: when we substitute $x = 1$ in the power series for $G_X(x)$ we just get $\sum p_k$.

For part (b), when we differentiate the series term-by-term (you will learn later in Analysis that this is OK), we get

$$\frac{d}{dx} G_X(x) = \sum k p_k x^{k-1}.$$

Now putting $x = 1$ in this series we get

$$\sum k p_k = E(X).$$

For part (c), differentiating twice gives

$$\frac{d^2}{dx^2} G_X(x) = \sum k(k-1) p_k x^{k-2}.$$

Now putting $x = 1$ in this series we get

$$\sum k(k-1) p_k = \sum k^2 p_k - \sum k p_k = E(X^2) - E(X).$$

Adding $E(X)$ and subtracting $E(X)^2$ gives $E(X^2) - E(X)^2$, which by definition is $\mathrm{Var}(X)$.

red balls in the sample. Such an X is called a *hypergeometric* random variable $\mathrm{Hg}(n,M,N)$.

The random variable X can take any of the values $0,1,2,\ldots,n$. Its p.m.f. is given by the formula

$$P(X=k)=\frac{{}^{M}C_k\cdot{}^{N-M}C_{n-k}}{{}^{N}Cn}.$$

For the number of samples of n balls from N is ${}^{N}C_n$; the number of ways of choosing k of the M red balls and $n-k$ of the $N-M$ others is ${}^{M}C_k\cdot{}^{N-M}C_{n-k}$; and all choices are equally likely.

The expected value and variance of a hypergeometric random variable are as follows (we won't go into the proofs):

$$E(X)=n\left(\frac{M}{N}\right),\qquad \mathrm{Var}(X)=n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

You should compare these to the values for a binomial random variable. If we let $p=M/N$ be the proportion of red balls in the hat, then $E(X)=np$, and $\mathrm{Var}(X)$ is equal to npq multiplied by a 'correction factor' $(N-n)/(N-1)$.

In particular, if the numbers M and $N-M$ of red and non-red balls in the hat are both very large compared to the size n of the sample, then the difference between sampling with and without replacement is very small, and indeed the 'correction factor' is close to 1. So we can say that $\mathrm{Hg}(n,M,N)$ is approximately $\mathrm{Bin}(n,M/N)$ if n is small compared to M and $N-M$.

Consider our example of choosing two pens from four, where two pens are red, one green, and one blue. The number X of red pens is a $\mathrm{Hg}(2,2,4)$ random variable. We calculated earlier that $P(X=0)=1/6$, $P(X=1)=2/3$ and $P(X=2)=1/6$. From this we find by direct calculation that $E(X)=1$ and $\mathrm{Var}(X)=1/3$. These agree with the formulae above.

Geometric random variable $\mathrm{Geom}(p)$

The geometric random variable is like the binomial but with a different stopping rule. We have again a coin whose probability of heads is p. Now, instead of tossing it a fixed number of times and counting the heads, we toss it until it comes down heads for the first time, and count the number of times we have tossed the coin. Thus, the values of the variable are the positive integers $1,,2,3,\ldots$ (In theory we might never get a head and toss the coin infinitely often, but if $p>0$ this possibility is 'infinitely unlikely', i.e. has probability zero, as we will see.) We always assume that $0<p<1$.

More generally, the number of independent Bernoulli trials required until the first success is obtained is a geometric random variable.

The p.m.f of a Geom(p) random variable is given by

$$P(X=k) = q^{k-1}p,$$

where $q = 1-p$. For the event $X = k$ means that we get tails on the first $k-1$ tosses and heads on the kth, and this event has probability $q^{k-1}p$, since 'tails' has probability q and different tosses are independent.

Let's add up these probabilities:

$$\sum_{k=1}^{\infty} q^{k-1}p = p + qp + q^2p + \cdots = \frac{p}{1-q} = 1,$$

since the series is a geometric progression with first term p and common ratio q, where $q < 1$. (Just as the binomial theorem shows that probabilities sum to 1 for a binomial random variable, and gives its name to the random variable, so the geometric progression does for the geometric random variable.)

We calculate the expected value and the variance using the probability generating function. If $X \sim \text{Geom}(p)$, the result will be that

$$E(X) = 1/p, \qquad \text{Var}(X) = q/p^2.$$

We have

$$G_X(x) = \sum_{k=1}^{\infty} q^{k-1}px^k = \frac{px}{1-qx},$$

again by summing a geometric progression. Differentiating, we get

$$\frac{\mathrm{d}}{\mathrm{d}x}G_X(x) = \frac{(1-qx)p + pxq}{(1-qx)^2} = \frac{p}{(1-qx)^2}.$$

Putting $x = 1$, we obtain

$$E(X) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Differentiating again gives $2pq/(1-qx)^3$, so

$$\text{Var}(X) = \frac{2pq}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

For example, if we toss a fair coin until heads is obtained, the expected number of tosses until the first head is 2 (so the expected number of tails is 1); and the variance of this number is also 2.

for $0 \le y \le 2$; of course $F_Y(y) = 0$ for $y < 0$ and $F_Y(y) = 1$ for $y > 2$. (Note that $Y \le y$ if and only if $X \le y^2$, since $Y = \sqrt{X}$.)

(c) We have

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} y/2 & \text{if } 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

The argument in (b) is the key. If we know Y as a function of X, say $Y = g(X)$, where g is an increasing function, then the event $Y \le y$ is the same as the event $X \le h(Y)$, where h is the *inverse function* of g. This means that $y = g(x)$ if and only if $x = h(y)$. (In our example, $g(x) = \sqrt{x}$, and so $h(y) = y^2$.) Thus

$$F_Y(y) = F_X(h(y)),$$

and so, by the Chain Rule,

$$f_Y(y) = f_X(h(y))h'(y),$$

where h' is the derivative of h. (This is because $f_X(x)$ is the derivative of $F_X(x)$ with respect to its argument x, and the Chain Rule says that if $x = h(y)$ we must multiply by $h'(y)$ to find the derivative with respect to y.)

Applying this formula in our example we have

$$f_Y(y) = \frac{1}{4} \cdot 2y = \frac{y}{2}$$

for $0 \le y \le 2$, since the p.d.f. of X is $f_X(x) = 1/4$ for $0 \le x \le 4$.

Here is a formal statement of the result.

Theorem 4.5 *Let X be a continuous random variable. Let g be a real function which is either strictly increasing or strictly decreasing on the support of X, and which is differentiable there. Let $Y = g(X)$. Then*

(a) the support of Y is the image of the support of X under g;

(b) the p.d.f. of Y is given by $f_Y(y) = f_X(h(y))|h'(y)|$, where h is the inverse function of g.

For example, here is the proof of Proposition 3.6: *if $X \sim N(\mu, \sigma^2)$ and $Y = (X - \mu)/\sigma$, then $Y \sim N(0,1)$.* Recall that

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

Note the 2 in the line labelled "by the Chain Rule". If you blindly applied the formula of Theorem 4.5, using $h(y) = \sqrt{y}$, you would not get this 2; it arises from the fact that, since $Y = X^2$, each value of Y corresponds to two values of X (one positive, one negative), and each value gives the same contribution, by the symmetry of the p.d.f. of X.

4.5 Worked examples

Question Two numbers X and Y are chosen independently from the uniform distribution on the unit interval $[0,1]$. Let Z be the maximum of the two numbers. Find the p.d.f. of Z, and hence find its expected value, variance and median.

Solution The c.d.f.s of X and Y are identical, that is,

$$F_X(x) = F_Y(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 < x < 1, \\ 1 & \text{if } x > 1. \end{cases}$$

(The variable can be called x in both cases; its name doesn't matter.)

The key to the argument is to notice that

$$Z = \max(X,Y) \leq x \quad \text{if and only if } X \leq x \text{ and } Y \leq x.$$

(For, if both X and Y are smaller than a given value x, then so is their maximum; but if at least one of them is greater than x, then again so is their maximum.) For $0 \leq x \leq 1$, we have $P(X \leq x) = P(Y \leq x) = x$; by independence,

$$P(X \leq x \text{ and } Y \leq x) = x \cdot x = x^2.$$

Thus $P(Z \leq x) = x^2$. Of course this probability is 0 if $x < 0$ and is 1 if $x > 1$. So the c.d.f. of Z is

$$F_Z(x) = \begin{cases} 0 & \text{if } x < 0, \\ x^2 & \text{if } 0 < x < 1, \\ 1 & \text{if } x > 1. \end{cases}$$

The median of Z is the value of m such that $F_Z(m) = 1/2$, that is $m^2 = 1/2$, or $m = 1/\sqrt{2}$.

We obtain the p.d.f. of Z by differentiating:

$$f_Z(x) = \begin{cases} 2x & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then we can find $E(Z)$ and $\mathrm{Var}(Z)$ in the usual way:

$$E(Z) = \int_0^1 2x^2 \mathrm{d}x = \frac{2}{3}, \qquad \mathrm{Var}(Z) = \int_0^1 2x^3 \mathrm{d}x - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Appendix A

Mathematical notation

The Greek alphabet

Mathematicians use the Greek alphabet for an extra supply of symbols. Some, like π, have standard meanings. You don't need to learn this; keep it for reference. Apologies to Greek students: you may not recognise this, but it is the Greek alphabet that mathematicians use!
Pairs that are often confused are zeta and xi, or nu and upsilon, which look alike; and chi and xi, or epsilon and upsilon, which sound alike.

Name	Capital	Lowercase
alpha	A	α
beta	B	β
gamma	Γ	γ
delta	Δ	δ
epsilon	E	ε
zeta	Z	ζ
eta	H	η
theta	Θ	θ
iota	I	ι
kappa	K	κ
lambda	Λ	λ
mu	M	μ
nu	N	ν
xi	Ξ	ξ
omicron	O	o
pi	Π	π
rho	P	ρ
sigma	Σ	σ
tau	T	τ
upsilon	Υ	υ
phi	Φ	ϕ
chi	X	χ
psi	Ψ	ψ
omega	Ω	ω