MAS 108 — Probability I

Notes 6 — Autumn 2005

Random variables

A *probability space* is a sample space $\mathcal{S}$ together with a probability function P which satisfies Kolmogorov's axioms.

The Holy Roman Empire was, in the words of the historian Voltaire, "neither holy, nor Roman, nor an empire". Similarly, a random variable is neither random nor a variable:

> A *random variable* is a function from a probability space to the real numbers.

Notation We usually use capital letters like X, Y, Z to denote random variables.

Example (Child:part 1) One child is born, equally likely to be a boy of a girl. Then $\mathcal{S} = \{B, G\}$. We can define the random variable X by $X(B) = 1$ and $X(G) = 0$.

Example (One die: part 1) One fair six-sided die is thrown. X is the number showing.

Example (Three coin tosses: part 1) I toss a fair coin three times; I count the number of heads that come up.

Here, the sample space is

$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\};$$

the random variable X is the function which counts the number of Hs in each outcome, so that $X(THH) = 2$, $X(TTT) = 0$, etc.

If x is any real number, the event "$X = x$" is defined to be the event

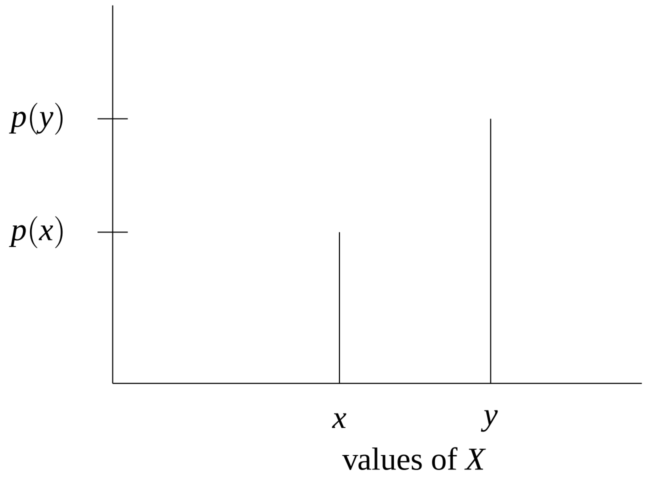
$$\{s \in \mathcal{S} : X(s) = x\},$$

which can also be written as $X^{-1}(x)$. The probability that $X = x$, which is written $P(X = x)$, is just $P(\{s \in \mathcal{S} : X(s) = x\})$. Similarly, other events can be defined in terms of the value(s) taken by X: for example, $P(X \leq y)$ means $P(\{s \in \mathcal{S} : X(s) \leq y\})$.

Notation $P(X = x)$ is written $p(x)$, or $p_X(x)$ if we need to emphasize the name of the random variable. In handwriting, be careful to distinguish between X and x.

The list of values $(x, p(x))$, for those x such that there is an outcome x in $\mathcal{S}$ with $X(s) = x$, is called the *distribution* of X. The function $x \mapsto p(x)$ is called the *probability mass function* of X, often abbreviated to pmf.

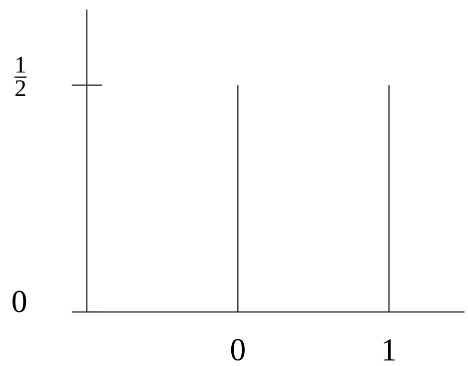
Often we concentrate on the distribution or on the probability mass function and forget the sample space. Sometimes the pmf is given by a table of values, sometimes by a formula. We can also draw the *line graph* of a pmf.



Example (Child: part 2) The distribution of X is

x	0	1
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$

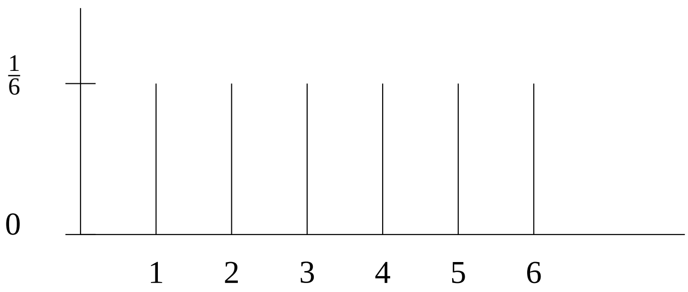
and the line graph is



Example (One die: part 2) The distribution of X is

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

and the line graph is



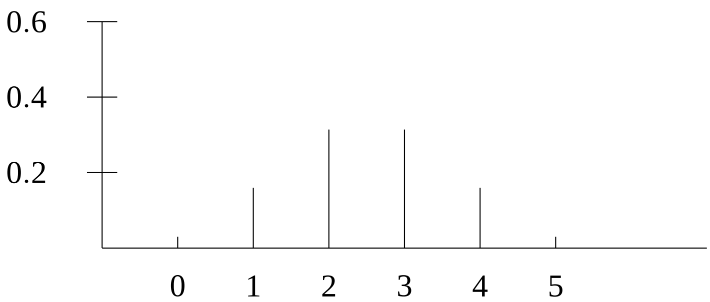
Example A biased coin is tossed five times. Each time it has probability p of coming down heads, independently of all other times. Let X be the number of heads. Then $P(X=m) = {}^5\mathrm{C}_m p^m q^{5-m}$ for $m = 0, \ldots, 5$, where $q = 1-p$. So the pmf is:

m	0	1	2	3	4	5
$p(m)$	q^5	$5pq^4$	$10p^2q^3$	$10p^3q^2$	$5p^4q$	p^5

Different values of p give different distributions. Here are three examples, together with their line graphs.

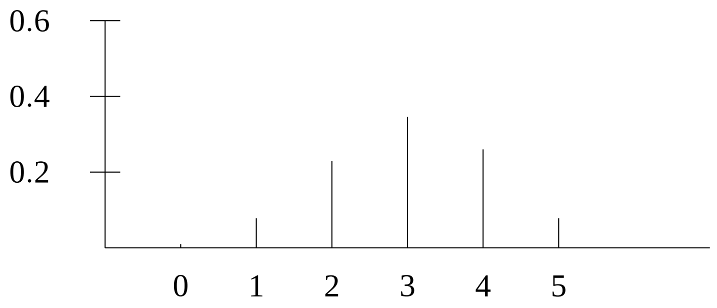
$p = 0.5$

m	0	1	2	3	4	5
$p(m)$	0.031	0.156	0.313	0.313	0.156	0.031



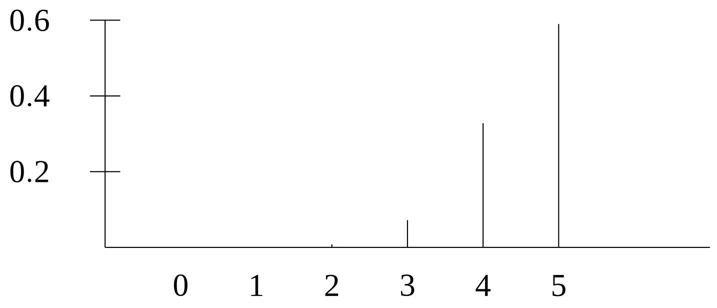
$p = 0.6$

m	0	1	2	3	4	5
$p(m)$	0.010	0.077	0.230	0.346	0.259	0.078



$p = 0.9$

m	0	1	2	3	4	5
$p(m)$	0.000	0.000	0.008	0.073	0.328	0.590



A random variable X is *discrete* if

either (a) $\{X(s) : s \in \mathcal{S}\}$ is finite, that is, X takes only finitely many values,

or (b) $\{X(s) : s \in \mathcal{S}\}$ is infinite but the values X can take are separated by gaps: formally, there is some positive number δ such that if x and y are two different numbers in $\{X(s) : s \in \mathcal{S}\}$ then $|x - y| > \delta$.

For example, X is discrete if it can take only finitely many values (as in all the examples above), or if the values of X are integers.

Note that is X is discrete then $\sum_x p(x) = 1$, where the sum is taken over all values x that X takes.

Example An example where the number of values is infinite is the following: you keep taking an exam until you pass it for the first time; X is the number of times you sit the exam. There is no upper limit on the values of X; for example, you cannot guarantee to pass it even in 100 attempts. So the set of values is $\{1,2,3,\ldots\}$, the set of all positive integers.

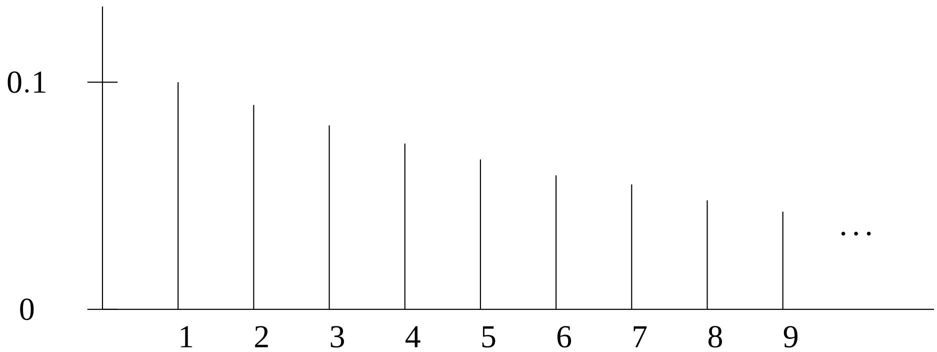
How do we give the probability mass function if the set of values is infinite? Suppose that your probability of passing is p at each attempt, independently of all previous attempts. The event $X = m$ is made up of just the one outcome $FF\ldots FP$ (with $m-1$ Fs); this has probability $q^{m-1}p$ where $q = 1 - p$. So we could just give a formula

$$P(X = m) = q^{m-1}p \qquad \text{for integers } m \geq 1.$$

Alternatively, the following table makes it clear:

m	1	2	3	$\ldots$	n	$\ldots$
$P(X = m)$	p	qp	q^2p	$\ldots$	$q^{n-1}p$	$\ldots$

When $p = 1/10$ the (incomplete) line graph is as follows.



Example (Two dice: part 1) I throw two fair six-sided dice. I am interested in the sum of the two numbers. Here the sample space is

$$\mathcal{S} = \{(i, j) : 1 \leq i, j \leq 6\},$$

and we can define random variables as follows:

$$\begin{aligned} X &= \text{number on first die} \\ Y &= \text{number on second die} \\ Z &= \text{sum of the numbers on the two dice;} \end{aligned}$$

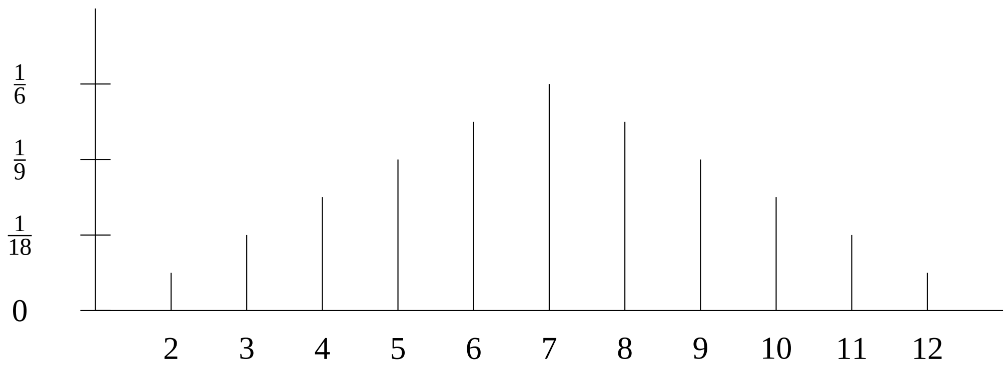
that is, $X(i, j) = i$, $Y(i, j) = j$ and $Z(i, j) = i + j$.

Notice that X and Y are *different* random variables even though they take the same values and their probability mass functions are equal. They are are said to have *the same distribution*. We write $X \sim Y$ in this case.

The target set for Z is the set $\{2, 3, \ldots, 12\}$. Since each outcome has probability $1/36$, the pmf is obtained by counting the number of ways we can achieve each value and dividing by 36. For example, $9 = 6+3 = 5+4 = 4+5 = 3+6$, so $P(Z = 9) = \frac{4}{36}$. We find:

k	2	3	4	5	6	7	8	9	10	11	12
$P(Z = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The line graph follows.



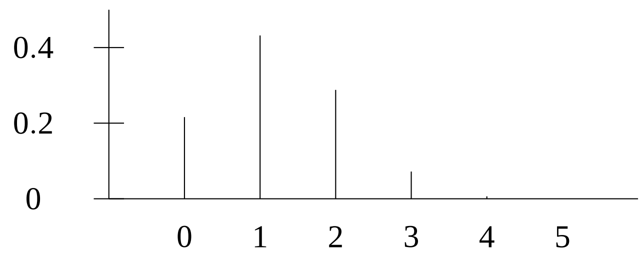
In the first coin-tossing example, if Y is the number of tails recorded during the experiment, then X and Y again have the same distribution, even though their actual values are different (indeed, $Y = 3 - X$).

Example (Sheep: part 1) There are 24 sheep in a field. The farmer shears 6 of them. Later, he comes to the field and randomly samples 5 sheep without replacement. Let X be the number of shorn sheep in his sample. Then

$$P(X = m) = \frac{{}^{6}\mathrm{C}_{m} \times {}^{18}\mathrm{C}_{5-m}}{{}^{24}\mathrm{C}_{5}}$$

for $m = 0, 1, 2, 3, 4, 5$. To 4 decimal places the pmf is as follows.

m	0	1	2	3	4	5
$p(m)$	0.2016	0.4320	0.2880	0.0720	0.0064	0.0001



Expected value

Let X be a discrete random variable which takes the values $a_1, \ldots, a_n$. The *expected value* (also called the *expectation* or *mean*) of X is the number $E(X)$ given by the formula

$$E(X) = \sum_{i=1}^{n} a_i P(X = a_i).$$

That is, we multiply each value of X by the probability that X takes that value, and sum these terms. Often I will write this sum as

$$\sum_x x p(x).$$

I may also write μ_X for $E(X)$, and may abbreviate this to μ if X is clear from the context.

The expected value is a kind of 'generalised average': if each of the values is equally likely, so that each has probability $1/n$, then $E(X) = (a_1 + \cdots + a_n)/n$, which is just the average of the values.

There is an interpretation of the expected value in terms of mechanics. If we put a mass p_i on the axis at position a_i for $i = 1,\ldots,n$, where $p_i = P(X = a_i)$, then the centre of mass of all these masses is at the point $E(X)$. In other words, if we make the line graph out of metal (and do not include the vertical axis) then the graph will balance at the point $E(X)$ on the horizontal axis.

If the random variable X takes infinitely many values, say a_1, a_2, a_3, ..., then we define the expected value of X to be the infinite sum

$$E(X) = \sum_{i=1}^{\infty} a_i P(X = a_i).$$

Of course, now we have to worry about whether this means anything, that is, whether this infinite series is convergent. This is a question which is discussed at great length in analysis. We won't worry about it too much. Usually, discrete random variables will only have finitely many values; in the few examples we consider where there are infinitely many values, the series will usually be a geometric series or something similar, which we know how to sum. In the proofs below, we assume that the number of values is finite.

Example (Child: part 3)

$$E(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}.$$

Example (One die: part 3)

$$E(X) = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3.5.$$

Example (Sheep: part 2)

$$\begin{aligned} E(X) &= 0 \times 0.2016 + 1 \times 0.4320 + 2 \times 0.2880 + 3 \times 0.0720 + 4 \times 0.0064 + 5 \times 0.0001 \\ &= 1.2501 \end{aligned}$$

to 4 decimal places. We shall see later that it should be exactly $5/4$: what we have here is affected by rounding error.

Example (Two dice: part 2)

$$\begin{aligned} E(Z) &= 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} \\ &= \frac{1}{36}(2+6+12+20+30+42+40+36+30+22+12) \\ &= 7. \end{aligned}$$

Variance

While $E(X)$ gives the centre of gravity of the distribution, the spread of the disributiion is measured by the average value of $(X-\mu_X)^2$.

The *variance* of X is the number $\mathrm{Var}(X)$ given by

$$\mathrm{Var}(X)=\sum_x (x-\mu_X)^2 p(x),$$

where $\mu_X = E(X)$. Sometimes it is written as σ_X^2, or just as σ^2 if X is clear from the context.

Example (Sheep: part 3)

m	0	1	2	3	4	5
$p(m)$	0.2016	0.4320	0.2880	0.0720	0.0064	0.0001
$m-1.25$	-1.25	-0.25	0.75	1.75	2.75	3.75

$$\mathrm{Var}(X)=(-1.25)^2\times 0.2016+\cdots+(3.75)^2\times 0.0001=0.7743.$$

(Again, there is rounding error, because the exact value should be $285/(16\times 23)$, which is 0.7745 to 4 decimal places.)

Theorem 3 If $E(X)=\mu$ then $\mathrm{Var}(X)=\sum_x x^2p(x)-\mu^2$.

Proof

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_x (x-\mu)^2 p(x)\\
&= \sum_x \left(x^2-2\mu x+\mu^2\right) p(x)\\
&= \sum_x \left[x^2p(x)-2\mu x p(x)+\mu^2 p(x)\right]\\
&= \sum_x x^2p(x)-\sum_x 2\mu x p(x)+\sum_x \mu^2 p(x)\\
&= \sum_x x^2p(x)-2\mu\sum_x xp(x)+\mu^2\sum_x p(x)\\
&= \sum_x x^2p(x)-2\mu^2+\mu^2 \qquad \text{because } \textstyle\sum_x xp(x)=\mu \text{ and } \sum_x p(x)=1\\
&= \sum_x x^2p(x)-\mu^2. \quad \blacksquare
\end{aligned}$$

So now we have two methods of calculating variance. The first is affected by rounding errors in the calculation of μ, and also by rounding errors in the calculation of the values $p(x)$. The second method is more badly affected by rounding errors in the $p(x)$, and it can cause computer or calculator overflow if X takes very large values; however, the calculations are usually simpler.

Two further properties of expected value and variance can be used as a check on your calculations.

- The expected value of X always lies between the smallest and largest values of X. (Can you prove this?)

- The variance of X is never negative. (We will prove this in a little while.)