RVs and their probability distributions

In these notes, I will use the following notation:
The probability distribution (function) on a sample space will be denoted by $P(A)$ - upper case P. Here A represents either an outcome or an event.
The probability distribution (function) (pdf) for a discrete (finite or infinite) random variable will be denoted by $p(X = x)$ or $p_X(x)$ or just $p(x)$ if there is only one random variable under discussion - note, lower-case p! Here x represents a possible value of the random variable.
The probability density function for a continuous random variable will be denoted by f(x).

Motivation

Why use the sample space $S = \{HH, HT, TH, TT\}$ for the experiment "tossing two balanced coins"? What is wrong with using the sample space $S = \{\text{two heads, one head, no heads}\}$?

Another example: why use the sample space consisting of 36 ordered pairs for the experiment "rolling two balanced dice" when all we care about is the sum of the numbers showing on the two dice? Here the sample space is so large and unwieldy, and if we declared instead that the sample space is $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ (the possible total number of spots showing), it might be more useful.

On the other hand, there is more than one way to throw a total of 7, but only one way to throw a 2. So I probably would not think that these are equally likely events.

If we assume that the two tosses or the two rolls are independent, that means that the probability of getting heads (for instance) on the second toss does not depend on whether or not heads shows on the first toss. How could one coin be influenced by what another coin does? Or equivalently, how can a coin be influenced by whether or not it landed heads up last time I tossed it? If the tossing procedure is anything like reasonable, we can easily believe that the tosses must be independent. And the same goes for the dice.

I can use that independence to compute what the probabilities of the four elements of S given in the first sample space, from knowing only that the coin is balanced, by reasoning as follows: the probability of getting heads on the first toss (or the first coin) is 1/2. The probability of getting heads on the second toss, given that the first one was heads, is still 1/2, because of independence. Hence the probability of getting heads on the first toss and on the second toss is the product 1/2 times 1/2 or 1/4. Same for all three other outcomes (work it out!) and so, assuming only that the coin is balanced, the sample space which has four events in it is naturally an equally likely outcomes sample space. The other one is not, if we then view its outcomes as events in the first sample space.

The same argument works for the experiment "rolling two balanced dice". The sample space consisting of 36 ordered pairs is naturally an equally likely outcomes sample space, and the other is not. This is the main reason for preferring such large sample spaces.

But then we want to talk about the sum of the numbers on the two dice, or about the number of heads that show! So this is the reason for defining a concept to represent such things. That is a random variable.

Random Variables: A random variable is a variable whose value depends on the outcome of a random process (experiment). The values of a random variable in this course will be real numbers. The random variable acts like a function: it assigns a number to each outcome in S.

Random variables are traditionally designated by using capital letters, starting with X.

Finite Random Variables

A **finite random variable** is one which has only a finite number of possible values.

Probability Distribution Function (pdf)

A probability distribution (function) for a finite random variable X is a function $p(X = k)$ which assigns a probability to each possible value k of X. (We will sometimes shorthand this notation by writing just $p(k)$ for $p(X = k)$.)

If we already have a probability distribution on the underlying sample space, the pdf for the random variable is defined using that probability distribution. For each possible value k of the RV X, we find all of the outcomes which give that value of the RV and add their probabilities together.

Example of a probability distribution function for an RV:

Experiment "toss two balanced coins".
$S = \{HH, HT, TH, TT\}$ and each outcome has probability $\frac{1}{4}$ as argued in the Motivation section.
Define a RV on S as follows:

$$X = \text{the number of heads that show}$$

Then the possible values of X are 0, 1, and 2.

We define the pdf for X by looking at each value of the random variable as if it described an event and computing the probability of that event as usual, by adding the probabilities of the individual outcomes:
$p(X = 0) = P(TT) = \frac{1}{4}$
$p(X = 1) = P(HT) + P(TH) = \frac{1}{2}$
$p(X = 2) = P(HH) = \frac{1}{4}$

So, putting this all together (and using shorthand) the pdf for X is given by
$p(0) = \frac{1}{4}$
$p(1) = \frac{1}{2}$
$p(2) = \frac{1}{4}$

Experiment: Roll two coins which are unbalanced, so that $P(HH) = \frac{1}{2}$, $P(HT) = \frac{1}{4}$, $P(TH) = \frac{1}{8}$, $P(TT) = \frac{1}{8}$. Now for $X =$ the number of heads that show, we have:

$p(X = 0) = P(TT) = \frac{1}{8}$
$p(X = 1) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$
$p(X = 2) = P(HH) = \frac{1}{2}$

Properties of the pdf of a random variable:
The pdf of a finite or discrete RV has similar properties to those of a pdf on a sample space:

- $0 \leq p(x) \leq 1$ for all possible values x
- $\sum p(x) = 1$, where the sum is taken over all possible values of X
- The probability that $X = x_1$ or $X = x_2$ is $p(x_1) + p(x_2)$, etc. (The various values of the RV represent mutually exclusive events and so the probability of their unions are found by adding.)

Examples:

1. For example, for our previously defined random variable $X=$ the number of heads that show in tossing two balanced coins, we can compute the probability that X is 1 or 2:

   $P(X = 1 \; or \; X = 2) = p(1) + p(2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

2. Experiment "rolling two balanced dice" This has a sample space consisting of 36 ordered pairs describing the two numbers which show. This is the equally likely sample space for this experiment under the classical approach. Each outcome has probability $\frac{1}{36}$. It is useful to represent these outcomes as points in the first quadrant of the Cartesian plane. All of the points lie in a 6-by-6 square.

   We can define the random variable $X=$ the sum of the two numbers that show. You may wish to label each point of the square with its sum. The possible values of X are the whole numbers 2 through 12 inclusive. Note that some sums appear more than once.

   Then we can compute the pdf for X (details omitted, but it is strongly recommended that you write out the sample space and compute them for yourself):
   $p(2) = \frac{1}{36}$
   $p(3) = \frac{2}{36}$ (because there are 2 ways to get a sum of 3)
   $p(4) = \frac{3}{36}$ (because there are 3 ways to get a sum of 4, etc.)
   $p(5) = \frac{4}{36}$
   $p(6) = \frac{5}{36}$
   $p(7) = \frac{6}{36}$
   $p(8) = \frac{5}{36}$
   $p(9) = \frac{4}{36}$
   $p(10) = \frac{3}{36}$
   $p(11) = \frac{2}{36}$
   $p(12) = \frac{1}{36}$

   We can use this pdf to find other probabilities.

   For example, the probability that the sum will be either 2 or 7 is
   $P(X = 2 \; or \; X = 7) = p(2) + p(7) = \frac{1}{36} + \frac{6}{36} = \frac{7}{36}$.

   Also, the probability that the sum is at least 10 is
   $P(X \geq 10) = p(10) + p(11) + p(12) = \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$

Infinite Discrete Random Variables: an Example

An **(infinite) discrete random variable** is one which has a countably infinite number of possible values. In other words, the possible values can be out into a one-to-one correspondence with the natural numbers.

The definitions and properties given above for the finite case go through without change in case the RV has possible values in an infinite discrete set.

We have so far seen exactly one example of an infinite, discrete sample space. Recall that for the experiment "toss a balanced coin until heads shows for the first time" has an infinite sample space:

$$S = \{H, TH, TTH, TTTH, TTTTH, ...\}$$

This is discrete because it can be put into one-to-one correspondence with the natural numbers. (Informally, you can count the elements, even though the counting never ends.)

Following the argument given in the Motivation section, we should arrive at the following probability distribution for this sample space:

$P(H) = \frac{1}{2}$

$P(TH) = (\frac{1}{2})(\frac{1}{2}) = (\frac{1}{2})^2$ (the probability of getting T on the first throw times the probability of getting heads on the second throw)

$P(TTH) = (\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = (\frac{1}{2})^3$ (the probability of getting T on the first throw times the probability of getting T on the second throw times the probability of getting H on the third throw)

etc. We are not computing the powers because the object here (since the sample space is infinite) is to get a general formula for the pdf.

The general formula can be seen from the process above:
$P(\underbrace{T...T}_{n\ tails} H) = \left(\frac{1}{2}\right)^{n+1}$

Verify that this satisfies the requirements to be a pdf. We need to check that all the probabilities are between 0 and 1 (which is obviously true) and that the sum of all the probabilities for all elements of the sample space is 1, which is not so obvious.

The sum of all the probabilities is

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1}$$

which is a geometric series $\sum_{n=0}^{\infty} ar^n$ with $a = \frac{1}{2}$ and $r = \frac{1}{2}$. The series converges when $|r| < 1$. The sum of this series is then $\frac{a}{1-r}$, so the sum of all these probabilities is

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

Now we will define a RV on this experiment.:

$$X = \text{ the number of tails that show}$$

The possible values of X are all the whole numbers 0, 1, 2, ..., so X is a discrete RV. (Note: it is possible to define a finite RV on an infinite sample space. The number of possible values of the RV may be smaller than the number of outcomes in the sample space: it just cannot be larger.)

Then from the pdf we just derived, we can find a pdf for X:

$p(0) = P(H) = \frac{1}{2}$
$p(1) = P(TH) = (\frac{1}{2})^2$
$p(2) = P(TTH) = (\frac{1}{2})^3$
$p(3) = P(TTTH) = (\frac{1}{2})^4$
etc.

As before, we see that

$p(n) = (\frac{1}{2})^{n+1}$

and the sum of the probabilities over all possible values of x is 1.

This example is important in the historical development of the theory of probability and we will be seeing it again with a different random variable. It is known as the Saint Petersburg Paradox.

Exercise: Work out the case of tossing an unbalanced coin until heads shows for the first time. Let's say that the probability of heads on one toss is $\frac{3}{5}$ and the probability of tails on one toss is then $\frac{2}{5}$. Find the pdf for the sample space (as a general formula) and then find the pdf for the RV X= the number of tails that show, and show that the sum of all the probabilities is 1. (It is important in doing this that you recreate the thinking of the example we just worked, and not just try to put numbers in the right places.)