Lecture 5: Discrete Random Variables

February 11, 2019

1 Examples of random variables.

Roughly speaking, a *random variable* X assigns a number to the outcome of an experiment. Here we'll give a bunch of examples to give you the idea; later we'll give a precise definition.

1.1 Single coin toss

Toss a single coin, and set $X_0 = 1$ if the coin comes up heads, and $X_0 = 0$ if the coin comes up tails.

1.2 Coin tosses

Toss a fair coin 20 times.

1.2.1 Number of heads

Let X_1 denote the number of heads tossed.

1.2.2 Excess of heads over tails.

Let X_2 denote the number of heads minus the number of tails.

1.2.3 Length of longest run

Let X_3 denote the length of the longest run of consecutive heads or tails.

1.3 Wait until heads.

Toss a coin repeatedly until heads comes up. Let X_4 denote the number of tosses.

1.4 Dice

1.4.1 One die

Roll a six-faced die. Let Y_1 denote the number showing.

1.4.2 Two dice

Roll two six-sided dice, and let $Y_{2,1}$ denote the number showing on the first die, and $Y_{2,2}$ the number showing on the second die. Set

$$Y_2 = Y_{2,1} + Y_{2,2},$$

that is, Y_2 denotes the sum of the two dice.

1.5 Darts

Throw 1000 darts at a darboard 12 inches in diameter, with a bulls-eye that is 1 inch in diameter. Let Z_1 denote the number of darts that strike the board. Let Z_2 denote the distance from the center of the board to the closest dart.

The various X's, Y's and Z's above are *random variables.*

2 Events and probabilities; PMF of a random variable.

If X is a random variable and $a \in \mathbf{R}$, then we write $X = a$ for the event consisting of all outcomes for which X has the value a. We write $P(X = a)$ to denote the probability of this event. Similarly, we write things like $X < a$, $P(a < X \leq b)$, *etc.* We illustrate this with the random variables in the examples above.

We have

$$P(X_0 = 1) = P(X_0 = 0) = \frac{1}{2},$$

assuming a fair coin.

We've seen how to calculate things like the probability that in 20 tosses of a fair coin we get exactly 7 heads, so

$$P(X_1 = 7) = \binom{20}{7} \cdot 2^{-20}.$$

We can write $X_2 = 2X_1 - 20$, so $X_1 = \frac{1}{2}(X_2 + 20)$, and thus

$$P(X_2 = -6) = P(X_1 = 7) = \binom{20}{7} 2^{-20},$$

We don't know how to calculate exact probabilities for X_3,, but simulations suggest that the longest run in 20 coin tosses is at least 4 more than 70% of the time, so:

$$P(X_3 < 4) < 0.3.$$

As we've seen, for $i = 1, 2, \ldots,$

$$P(X_4 = i) = 2^{-i}.$$

Assuming fair dice, we have

$$P(Y_{2,1} = i) = \frac{1}{6}$$

for $i \in \{1, \ldots, 6\}$, and

$$P(Y_2 = 6) = \frac{5}{36}, P(Y_2 = 7) = \frac{1}{6}.$$

For the dartboard experiment, assume that the probability of hitting the bull's-eye is $1/100$. Then we can view this as flipping a coin with heads probability $p = 0.01$ one thousand times. This gives

$$P(Z_1 = i) = \binom{1000}{i} \cdot 0.01^i \cdot 0.99^{1000-i}.$$

As we'll see later, there is an accurate, easy-to-compute approximation for this probability.

The random variable Z_2 is a bit of a problem: In all of the preceding examples, our random variable X satisfied $P(X = i) > 0$ on a *discrete* set of values i, and

$P(X_i) = 0$ everywhere else. That's what makes them discrete random variables. However, it would seem that Z_2 could take on any of a continuous range of values; it is a *continuous random variable*, something we will take up later.

The PMF of a random variable X (also the *distribution* of X) is the function

$$P_X : \mathbf{R} \to [0, 1],$$

defined by

$$P_X(a) = P(X = a).$$

As we said, the PMF has a nonzero value only on a discrete subset of **R**, and is zero elsewhere. This means that it is possible to form the infinite sum of all the values of the PMF. We have

$$\sum_{\{a : P_X(a) > 0\}} P_X(a) = 1.$$

It is as though we redefined the underlying sample space of the experiment to be the set of possible values X could take on—in this case, the PMF of the random variable is the same as the PMF as we defined it earlier. Often we define random variables without specifying the underlying experiment, simply by giving the PMF.

Example. The PMF of the random variable Y_2 above (sum of two dice) has PMF graphed in Figure 1. The PMF of X_4 is graphed in Figure 2. The bounding curves in the figure are there just to show more clearly the shape of PMF, but are not part of the graph itself.

3 Formal definition; operations on random variables

Here is the formal definition: A random variable on a probability space (S, P) is just a function

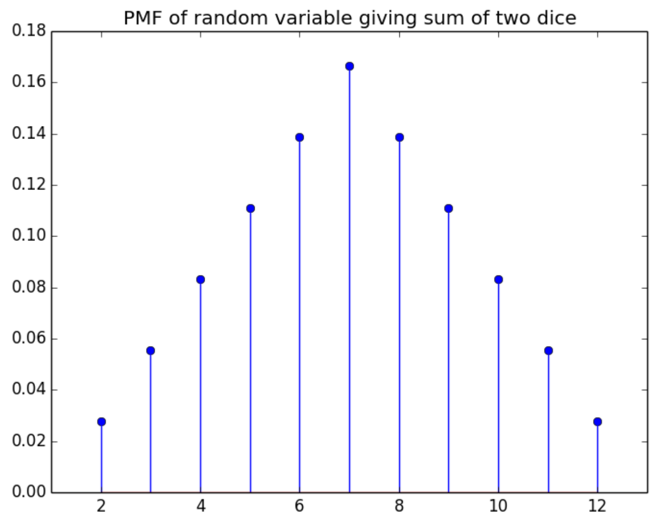
$$X : S \to \mathbf{R}.$$

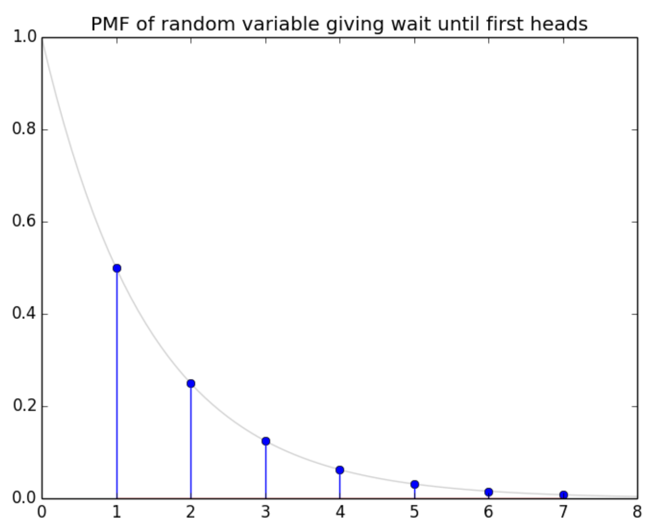
Then $P(X = a)$ is just a special shorthand for the probability of the event

$$\{s \in S : X(s) = a\}.$$

For example

$$X_1(\text{HHTHHTHHTTHHHTHHTTTH}) = 12.$$


PMF of random variable giving sum of two dice
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
2
4
6
8
10
12


PMF of random variable giving wait until first heads
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8

In practice you hardly ever use this functional notation for random variables.

By thinking of random variables this way, you can make sense of what it means to add or subtract (or even multiply) two different random variables on the same sample space, or multiply random variables by constants: In our examples, we saw $X_2 = 2X_1 - 20$ for the coin example, and $Y_2 = Y_{2,1} + Y_{2,2}$. Note that in this example, $Y_2, X_{2,1}$ and $X_{2,2}$ are all defined on the same sample space

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

which is the set of outcomes of the roll of two dice.

It's a little harder to say what the underlying sample space for the variables Z_1 and Z_2 is. We can think of it as the set of sequences $(p_1, \ldots, p_{1000})$ of 1000 points in the interior of the dartboard. This is a continuous sample space, although Z_1 is a discrete random variable defined on this space.

4 Independent random variables

Two random variables X_1, X_2 on the same probability space are *independent* if for all $a, b \in \mathbf{R}$,

$$\{s \in S : X_1(s) = a\}, \{s \in S : X_2(s) = b\}$$

are independent events. Recall this means that for all a and b,

$$P((X_1 = a) \wedge (X_2 = b)) = P(X_1 = a) \cdot P(X_2 = b).$$

This implies that for any *sets* A and B of values,

$$\{s \in S : X_1(s) \in A\}, \{s \in S : X_2(s) \in B\}$$

are also independent events.

Independence provides a formal justification for some of the probability calculations we made earlier. For example, we would naturally assume that the random variables that we denoted above by $Y_{2,1}$ and $Y_{2,2}$, representing the individual outcomes of each of the two dice, are independent. This then allows us to compute the PMF of the sum $Y_2 = Y_{2,1} + Y_{2,2}$. For example, the even $Y_2 = 8$ is the disjoint union of the events

$$(Y_{2,1} = 2) \wedge (Y_{2,2} = 6)$$

$$(Y_{2,1} = 3) \wedge (Y_{2,2} = 5)$$

$$(Y_{2,1} = 4) \wedge (Y_{2,2} = 4)$$
$$(Y_{2,1} = 5) \wedge (Y_{2,2} = 3)$$
$$(Y_{2,1} = 6) \wedge (Y_{2,2} = 2).$$

Independence implies that each of these events has probability $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. So $P_{Y_2}(8) = \frac{5}{36}$. This is an important general principle: If we know the distributions of two random variables, and we know that they are independent, then we can compute the distribution of their sum (and, likewise, their product, or any other operation performed on them).

Note that Y_2 and $Y_{2,2}$ are not independent–of course they aren't, because the sum of the two dice really does depend on what shows up on the second die! But let's verify this formally, using the definition. One counterexample suffices for this:

$$P(Y_2 = 12) = \frac{1}{36}.$$
$$P(Y_{2,2} = 5) = \frac{1}{6},$$

but

$$P((Y_2 = 12) \wedge (Y_{2,2} = 5)) = 0 \neq \frac{1}{36} \cdot \frac{1}{6}.$$

We can extend this to talk about a collection of more than two *mutually independent* random variables: $X_1, \ldots, X_k$ are mutually independent if for all $a_1, \ldots, a_k \in \mathbf{R}$,

$$P\big(\bigwedge_{j=1}^{k} (X_j = a_j)\big) = \prod_{j=1}^{k} P(X_j = a_j)$$

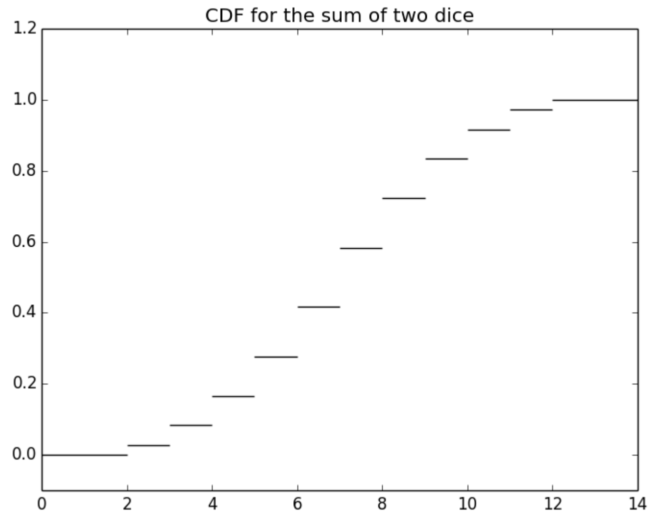
5 Cumulative distribution function

The *cumulative distribution* function (CDF) of a random variable X, denote F_X, is a function

$$F_X : \mathbf{R} \to [0,1]$$

defined by

$$F_X(a) = P(X \leq a).$$



For discrete random variables, we can compute this as

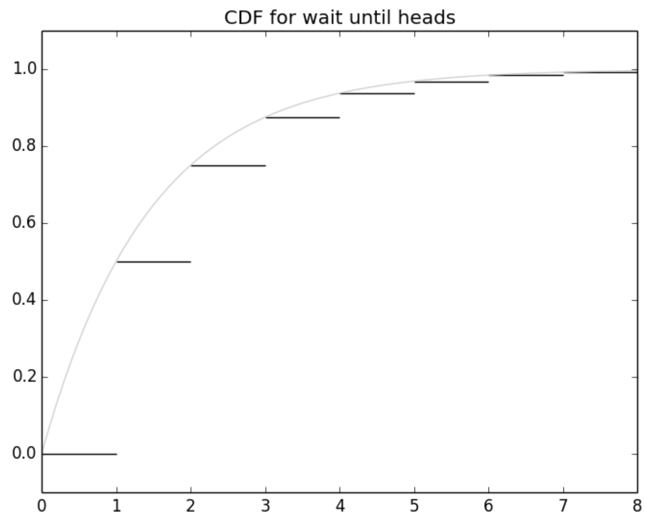
$$F_X(a) = \sum_{b \leq a} P_X(b),$$

since the terms on the right-hand side are all zero except on a discrete set of values. The figures below show the CDF of the two random variables Y_2 and X_4 whose PMFs were plotted above. Note that for discrete random variables, the CDF is a step function. This will change when we get to continuous random variables. (By the way, the horizontal line segments on these figures should appear with an open dot at the right-hand endpoints! For instance, in the first figure, $F_X(2) = 1/36$, not 0. I just got lazy....)

6 Using the CDF to generate random values according to a given distribution.

The CDF of a random variable is always monotone non-decreasing and rises from 0 to 1. As the second example indicates, it does not have to reach 1 (or 0) but must approach these values as limits as the argument approaches $\pm\infty$.

Suppose you want to generate random numbers that are not distributed uniformly, but that obey the distribution of some random variable. For example, we

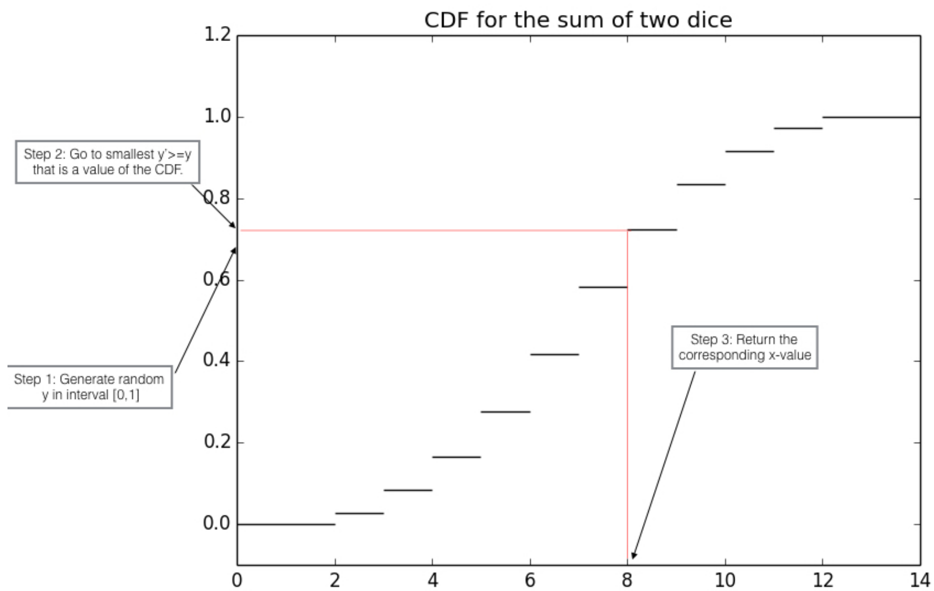


might want to generate a bunch of numbers in the range 2,...,12 that are distributed like the sum of two dice. Using the built-in `randint` function will let us generate integers in this range, but they will be uniformly distributed: 2 will occur approximately as often as 7. But we would actually like 2 to occur with frequency approximately $P_X(2) = \frac{1}{36}$, where X is the random variable that represents the sum of two successive rolls of a single die.

Of course we could do this easily using `randint`: Call it twice to generate two random integers uniformly distributed in the range 1,...,6 and add them. But we will describe a different method, applicable to any random variable, which essentially involves inverting the cumulative distribution function of the random variable.

The figure below illustrates the general procedure, using the sum of two dice as the example.

First generate a random value y uniformly in the interval $[0, 1]$. (This is actually a continuous distribution, which we will discuss formally later—but just think about the output of `random`.) In the picture, y is about 0.68. The idea is to find the corresponding x-value–*i.e.,* to solve $y = F_X(x)$. Now the CDF F_X is not one-to-one or onto, so this equation has no solution. Instead, we look for the



smallest value $y' \geq y$ that is in the range of the CDF. In this case

$$F_X(7) = P(X \leq 7) = \frac{21}{36} = 0.583, F_X(8) = P(X \leq 8) = \frac{26}{36} = 0.722,$$

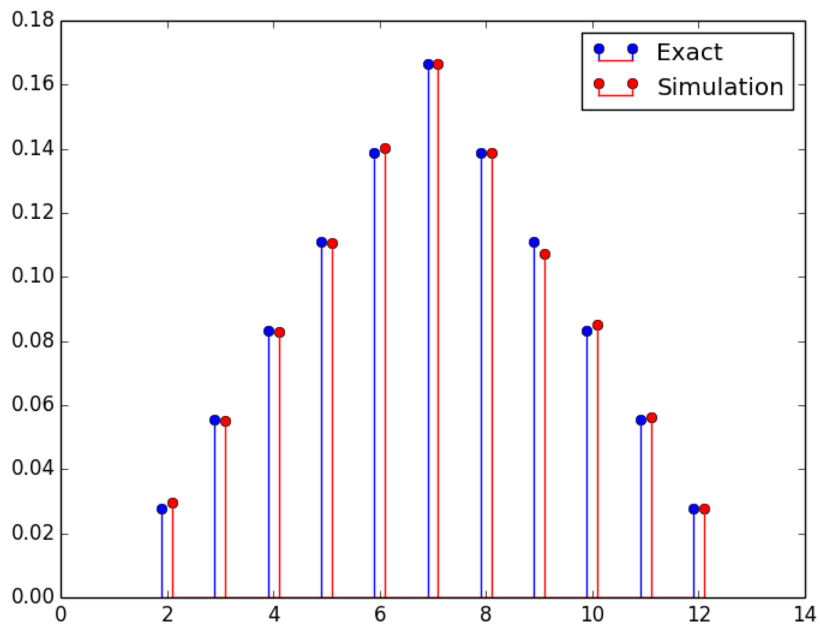
so we have $y' = \frac{26}{36}$, and the procedure returns the value 8.

What is the probability of generating 8 by this process? We would have to have, as above,

$$P(X \leq 7) < y \leq P(X \leq 8),$$

so the probability is $P(X \leq 8) - P(X \leq 7) = P(X = 8)$, which is what we want. Likewise, any value a of the random variable will be generated with probability $P(X = a)$.

There is an implementation posted on the website. The figure below shows the results: 20,000 values in the range 2 to 12 were generated using this algorithm, and their relative frequencies (in red) were plotted next to the exact values of the probability mass function. As you can see, there is very close agreement.


0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
12
14
Exact
Simulation