II. Discrete Random Variables

Discrete Random variables

Until now, we have discussed probability almost entire in the context of *events*, in which there are only two possibilities: either they happen or they don't, with some associated probabilities. This is limiting in two ways. First, we often want to talk about numerical measurements associated with these events (i.e., the probability of getting 2 heads in 3 coin tosses, regardless of the exact sequence). Second, we often want to think simultaneously about the probabilities associated with all possible values of these measurements (i.e., the three probabilities associated with getting 1, 2 or 3 heads in 3 coin tosses). In short, we want a mathematical framework that lets us quantitatively consider a variable that can take many values but is random in some way.

Definition: A **random variable** X is a *mapping* from the sample space Ω to the real line[1]:

$$X : \Omega \to \mathbb{R},$$

i.e., X assigns a real number to every possible outcome in the sample space.

Examples.

1. In an experiment involving drawing M&Ms from a bag, the number of candies you have to draw before repeating a color.
2. In a complex system, the number of days until a part failure, the number of parts that failed today, the number of customers affected by a failure, etc.

Random variables give us a succinct way to talk about numerical outcomes; before, we talked about the event that there were two

[1] As you might expect, we can also define random variables that are complex-valued, vector-valued, matrix-valued, etc.

heads, the event that there were three heads, etc. Now we can simply talk about "the number of random heads".

The examples above are discrete random variables because the values they take are from a discrete set. Continuous random variables also abound: the current temperature in the room, the amount of time until a light bulb burns out, etc. Since the discrete case is easier, we will start by developing some essential tools here and then generalize.

A word on notation

Random variables are just that — random; all of the interesting discussion about them occurs before we know their value. We will use *capital letters* to denote random variables, e.g., X for the number of heads in five flips.

We will use *lowercase letters* to denote particular outcomes a random variable might take. For example, we might ask about

$$\mathrm{P}\left(\{X = k\}\right) \quad \text{for } k = 2, 3, 4.$$

Where this will really start to get important (and confusing) is when we talk about multiple random variables, in which case it is easier (in the long run) to let x denote the particular value that X might take, y the value that Y might take, etc. This will seem a little unnatural at first, but just remember: X (or Y or Z ...) is a random variable, while x (or y or z) is just an old-fashioned regular variable.

Probability mass functions (pmfs) for discrete random variables

A random variable is completely described by the probabilities of the values it can take. These are encapsulated in the **probability mass function** for X, which we denote as $p_X(k)$. For discrete random variables, the definition is straightforward:

$$p_X(k) = \mathrm{P}\left(X = k\right).$$

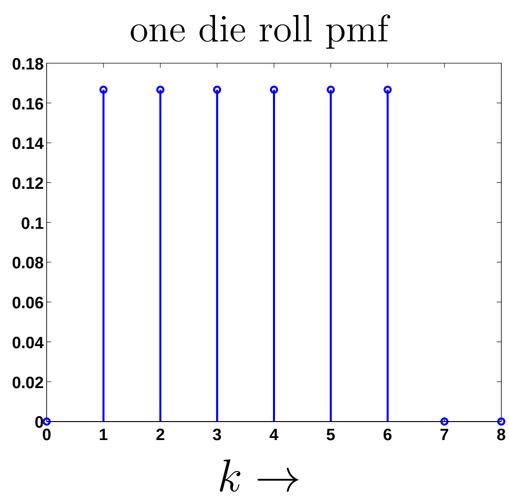
(Notice that we have adopted the notation $\mathrm{P}\left(X = k\right)$ over the strictly-more-correct-but-also-more-clunky $\mathrm{P}\left(\{X = k\}\right)$.)

Examples.

- For an experiment involving the roll of one fair six-sided die, let X be the number of "pips" facing upwards at the end of the roll (i.e., the numerical value of the result of the roll). Then

  $$p_X(k) = \begin{cases} 1/6 & k = 1, 2, \ldots, 6 \\ 0 & \text{otherwise.} \end{cases}$$

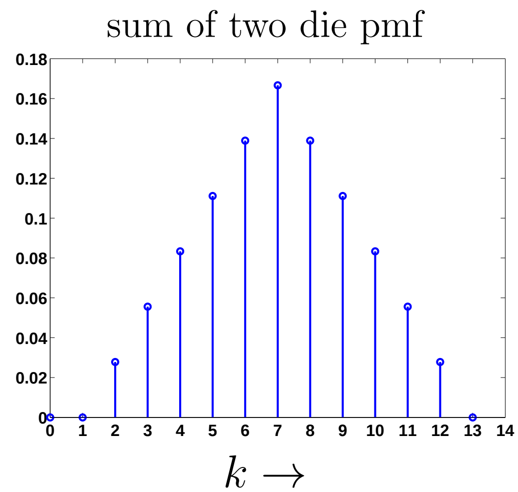
  This is called a *discrete uniform random variable*, and the pmf is illustrated below:



- For an experiment involving the roll of two fair six sided dice, let X be the sum of the values of the two rolls. Then

$$p_X(k) = \begin{cases} \frac{k-1}{36} & 2 \leq k \leq 7 \\ \frac{13-k}{36} & 8 \leq k \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

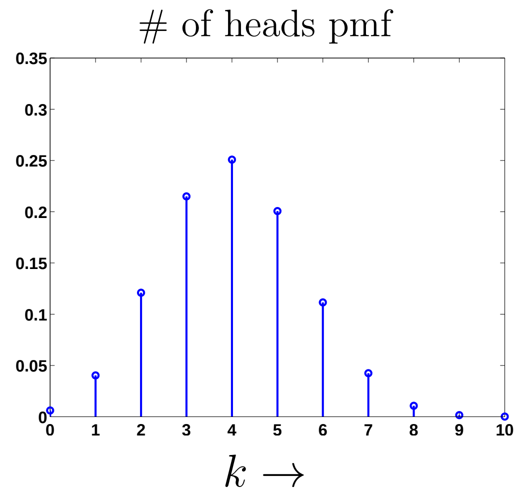
This pmf is illustrated below:



- An unfair coin is flipped 10 times, the probability that is lands on heads on any one flip is 0.4, and the flips are independent of one another. Let X be the total number of heads. Then

$$p_X(k) = \binom{10}{k}(0.4)^k(0.6)^{10-k} \quad \text{for } k = 0, 1, 2, \ldots, 10$$

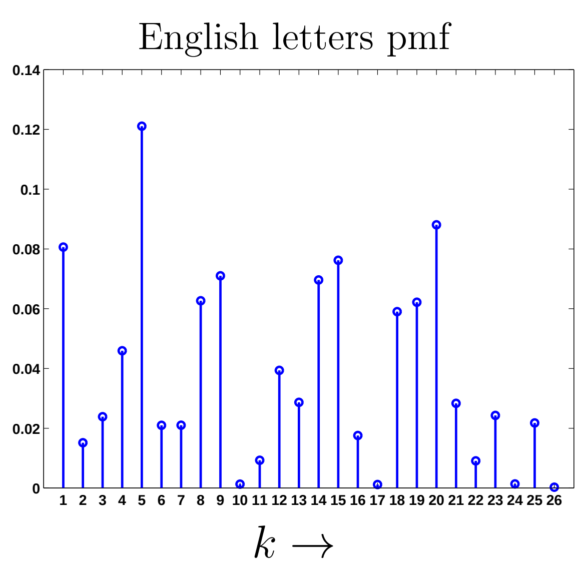
This pmf is illustrated below:



- You open a book written in English to a random page and place your finger down at a random location. Let

$$
\begin{aligned}
X &= 1, \quad \text{if the closest letter is an “a” (upper- or lower-case)} \\
X &= 2, \quad \text{if the closest letter is a “b”} \\
&\vdots \\
X &= 26, \quad \text{if the closest letter is a “z”.}
\end{aligned}
$$

The corresponding pmf is illustrated below:



Properties for pmfs

Every pmf obeys two properties which follow immediately from the Kolmogorov axioms:

1. Positivity:
$$p_X(k) \geq 0 \quad \text{for all } k$$

2. Normalization:
$$\sum_k p_X(k) = 1$$

When we combine these we also see that $p_X(k) \leq 1$ for all k.

Examples of important discrete pmfs

Bernoulli random variables.

These are the simplest random variables of all. They only take two values, 1 or 0. The pmf is defined by a single parameter p, the probability that $X = 1$:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0. \end{cases}$$

Bernoulli random variables are useful for things like modeling coin flips, bits, yes/no decisions, wins/loss, make/miss, Republican/Democrat, etc.

See also `http://bit.ly/UDt7Wc` (wikipedia).

Binomial random variables.

We consider a fixed number n of independent Bernoulli random variables (with parameter p), then set

$$X = \text{the number of the } n \text{ trials that had value } 1$$

As we have seen already

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

While it is not immediately obvious, it is possible to check that $p_X(k)$ is a valid pmf in the sense that $\sum_{k=0}^{n} p_X(k) = 1$. (This is a consequence of something called the "Binomial Theorem".)

Binomial random variables are useful for modeling the number of "successes" (or failures) over a series of independent trials.

- What is the probability that there is an error in k of the n bits I transmit?
- What is the probability that LeBron James makes k out of the next n free throws he shoots?
- What is the probability that k out of the next n visitors to my web site click on a certain link?

See also `http://bit.ly/UDtkIW` (wikipedia).

Geometric random variables.

A *geometric* random variable models the number of Bernoulli trials it will take to have our first success. We consider a (possibly infinitely long) series of Bernoulli trials with parameter p, and let

$$X = \text{the number of trials until I see the first 1.}$$

Then the pmf for X is

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots.$$

(This is simply the probability that we get $k-1$ zeros in a row and then a one.) This pmf satisfies the required normalization property since

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} (1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1.$$

Geometric random variables are good models for discrete "waiting" processes:

- How many times will I flip this coin before I see a "heads"?
- How many pages can I print out before the printer jams?
- How many attempts until LeBron James hits his next three-pointer?

See also http://bit.ly/TAjUQm (wikipedia).

Poisson random variables.

Poisson random variables are useful for modeling the number of events that occurred over a certain stretch of time:

- How many packets will be routed to this server in the next minute?
- How many cars will pass under the 5th street bridge between 2:00p and 2:07p this afternoon?
- How many photons will hit this detector in the next 5 ms?

The pmf for a Poisson random variable X is given by

$$p_X(k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

where $\lambda \geq 0$ is an *intensity parameter* (i.e., the larger λ, the more events we can expect to happen in a given interval).

It is clear that $p_X(k) \geq 0$; we can check the normalization property by recalling the *Taylor series* expansion for e^λ:

$$\begin{aligned} \mathrm{e}^\lambda &= 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}. \end{aligned}$$

Thus

$$\sum_{k=0}^{\infty} p_X(k) = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1.$$

See also http://bit.ly/UDtLTF (wikipedia).

Functions of a random variable

Random variables are just that: variables. They can be manipulated using algebraic rules just like standard variables. In particular, if you plug a random variable into a function, the output will be another random variable.

Here is a simple example. Suppose that X is a random variable modeling the Atlanta rainfall each day measured in inches (rounded to the nearest inch). I might be more interested in the rainfall in centimeters, which I can compute as

$$Y = 2.54\, X.$$

Y is another discrete random variable, and we can compute its pmf from the pmf of X.

In general, if $Y = g(X)$ is a function of a random variable X, we can compute the pmf of Y using

$$p_Y(y) = \sum_{\{k|g(k)=y\}} p_X(k).$$

(Now that we are starting to talk about multiple random variables, we will start using notation like $p_Y(y)$ and $p_X(x)$. Remember that X and Y are random variables, but x and y stand for the particular values X and Y might take. This might seem a little confusing now, but it is much easier in the long run since it helps you keep track of which value goes with which random variable. This will all get a bit clearer as we see more examples.)

Example: Suppose that X is a discrete random variable with pmf

$$p_X(k) = \begin{cases} 1/6, & \text{when } k \text{ is an integer with } -3 \le k \le 2 \\ 0, & \text{otherwise} \end{cases}.$$

1. Sketch the pmf of X.

2. Let $Y = |X|$. Calculate and sketch the pmf of Y.

3. Let $Y = X^2$. Calculate and sketch the pmf of Y.