# Discrete random variables

*S520*

*September 6, 2016*

## Discrete random variables

A **discrete random variable** has a countable number of possible outcomes. If the number of possible outcomes is countable, then to characterize these random variables, we can just write down the probability of every possible outcome.

**Probability mass function**

To formalize this idea, we define $f(x)$, the **probability mass function** or **PMF**, as

$$f(x) \equiv P(X = x)$$

for *all* real numbers $x$. To clarify, capital $X$ is a random variable, while little $x$ is some real number. $f(x)$ gives the probability that the random variable $X$ takes the value little $x$. We want this function to give a value for any real number we choose for $x$, even if the resulting probability is zero.

For any PMF, the sum of $f(x)$ over all values of $x$ that have non-zero probability must be 1.

**Example.** I roll a fair six-sided die. Let $X$ be a discrete random variable representing the number facing up. The PMF is:

$$f(x) = \begin{cases} 1/6 & x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise.} \end{cases}$$

With a discrete random variable, once you have the PMF, you can find the probabilities for events made up of multiple outcomes by addition.

Recall that the **cumulative distribution distribution (CDF)** of a random variable $X$, written $F(y)$, is defined as:

$$F(y) \equiv P(X \leq y)$$

for all real numbers $y$. For the fair die roll:

$$F(y) = \begin{cases} 0 & y < 1 \\ 1/6 & 1 \leq y < 2 \\ 2/6 & 2 \leq y < 3 \\ 3/6 & 3 \leq y < 4 \\ 4/6 & 4 \leq y < 5 \\ 5/6 & 5 \leq y < 6 \\ 1 & y \geq 6. \end{cases}$$

## Expected value, variance, and standard deviation

When there are lots of possible outcomes, PMFs and CDFs can get complicated. So we want to be sum up random variables in just a few numbers. Here are the ones we'll use:

The **expected value of a discrete random variable** $X$, denoted $EX$ or $\mu$, is:

$$EX \equiv \mu = \sum_x x \cdot P(X = x) = \sum_x x\, f(x)$$

The **variance of a discrete random variable** $X$, denoted $\mathrm{Var}(X)$, is:

$$\mathrm{Var}(X) = \sum_x (x - \mu)^2 \cdot f(x)$$

The **standard deviation of a discrete random variable** $X$, denoted $\sigma(X)$ or just $\sigma$, is:

$$\sigma \equiv \sqrt{\mathrm{Var}(X)} = \sqrt{\sum_x (x - \mu)^2 \cdot f(x)}$$

We can give intuitive explanations of why these things are interesting, but people usually don't believe them at this stage. But by the end of the course, we'll see that yes, these are good ways to summarize random variables.

**Example.** Let $X$ be a random variable representing the result of a die that has the following property: The probability of rolling a $k$ is proportional to $k$, for $k \in \{1, 2, 3, 4, 5, 6\}$.

Firstly, what's $k$? We know the probabilities have to add up to 1, so

$$1k + 2k + 3k + 4k + 5k + 6k = 1$$

This means $21k = 1$, or $k = 1/21$. We can now write down the PMF in full:

$$f(x) = \begin{cases} 1/21 & x = 1 \\ 2/21 & x = 2 \\ 3/21 & x = 3 \\ 4/21 & x = 4 \\ 5/21 & x = 5 \\ 6/21 & x = 6 \\ 0 & \text{otherwise.} \end{cases}$$

We get the CDF by addition:

$$F(y) = \begin{cases} 0 & y < 1 \\ 1/21 & 1 \le y < 2 \\ 3/21 & 2 \le y < 3 \\ 6/21 & 3 \le y < 4 \\ 10/21 & 4 \le y < 5 \\ 15/21 & 5 \le y < 6 \\ 1 & y \ge 6. \end{cases}$$

To find the expected value of $X$, multiply each value of $X$ by its probability:

$$EX = (1 \times 1/21) + (2 \times 2/21) + (3 \times 3/21) + (4 \times 4/21) + (5 \times 5/21) + (6 \times 6/21) = 91/21$$

Having found $EX$, find the variance from the definition:

$$VarX = 1/21 \times (1 - 91/21)^2 + \cdots + 6/21 \times (6 - 91/21)^2 = 20/9.$$

**Properties of expectation**

- If $X$ always equals a constant $c$, then $EX = c$.

- Let $Y = aX + b$. Then $EY = a \cdot EX + b$.

- Let $Y = X_1 + X_2$. Then $EY = E(X_1) + E(X_2)$. It surprises some people that this is always true regardless of whether $X_1$ and $X_2$ are dependent or independent.

**Properties of variance**

- If $X$ always equals a constant $c$, then $\text{Var}(X)$ and $\sigma$ are both 0.

- Let $Y = aX + b$. Then $\text{Var}(Y) = a^2 \cdot \text{Var}(X)$.

- Let $Y = X_1 + X_2$. If $X_1$ and $X_2$ are independent random variables, then $\text{Var}(Y) = \text{Var}(X_1) + \text{Var}(X_2)$. By independent discrete random variables, we mean that

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \, P(X_2 = x_2)$$

for all $x_1, x_2$.

**Example: Bernoulli random variables**

A **Bernoulli(p)** random variable takes the value 1 with probability $p$ and the value 0 with probability $1 - p$. We can think of it as a coin toss with probability $p$ of getting heads, except we count heads as 1 and tails as 0. The PMF and CDF are:

$$f(x) = \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$F(y) = \begin{cases} 0 & y < 0 \\ 1 - p & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

The expected value is $(1 \times p) + (0 \times (1 - p)) = p$ and the variance is

$$\begin{aligned} \text{Var}(X) &= (0 - p)^2(1 - p) + (1 - p)^2 p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= p - p^2 \\ &= p(1 - p). \end{aligned}$$

Now what happens if we add independent Bernoulli random variables together? Let $X_1$, $X_2$, ..., $X_n$ be independent Bernoulli($p$) random variables. From the properties of expected value and variance,

$$
\begin{aligned}
EY &= EX_1 + EX_2 + \cdots + EX_n \\
&= p + p + \cdots + p \\
&= np \\
\mathrm{Var}(Y) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&= p(1-p) + p(1-p) + \cdots + p(1-p) \\
&= np(1-p)
\end{aligned}
$$

Note that the sum of random variables is itself a random variable. In particular, the sum of $n$ independent Bernoulli($p$) random variables gives us a kind of random variable that's so important it needs its own section.

## Binomial random variables

I roll three dice. Let $X$ be a random variable representing the number of dice that show a square number. What's $P(X = 2)$: that is, what's the probability that exactly two of the three dice show a square?

First, we assume the dice are independent, which seems reasonable. By drawing a tree or otherwise, we find:

$$
\begin{aligned}
P(SSN) &= 1/3 \times 1/3 \times 2/3 = 2/27 \\
P(SNS) &= 1/3 \times 2/3 \times 1/3 = 2/27 \\
P(NSS) &= 2/3 \times 1/3 \times 1/3 = 2/27
\end{aligned}
$$

So the probability of getting exactly two squares is $2/27 + 2/27 + 2/27 = 3 \times 2/27 = 6/27$.

Okay, now if I roll ten dice, what's the probability of exactly four squares? This problem is too big to draw the tree, but we can use the same kind of argument:

P(four heads out of 10) = (number of ways of getting 4 heads out of 10) $\times$ (prob. of one particular way)

The number of ways of getting 4 heads out of 10 is the same as the number of ways of choosing 4 out of 10 objects to be "heads" (the rest are "tails" by default.) This is just

$$
C(10, 4) = \frac{10!}{4!6!} = 210
$$

Now we just need to find the probability of way (since all the ways with 4 heads have the same probability.) Pick an easy one:

$$
P(HHHHTTTTTT) = P(H)^4 \cdot P(T)^6 = 0.5^4 \cdot 0.5^6 = \frac{1}{1024}
$$

So the probability of exactly 4 heads out of 10 is $210/1024$ (about 0.2051.)

We generalize this using the **binomial formula**. Suppose we flip a coin $n$ times independently, each times with a probability of $p$ of getting heads. Let $X$ be a random variable representing the number of heads. Then

$$
P(X = x) = C(n, x)\, p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}
$$

For large numbers, this is tedious to calculate by hand. Fortunately, there's an R function:

```
dbinom(4, 10, 0.5)
```

## [1] 0.2050781

The inputs into the R function `dbinom()` are the number of heads, the number of coins, and the probability of getting a heads. They have to be in that order.

A **binomial random variable** is a random variable whose possible outcomes are the whole numbers from 0 to $n$, with probabilities given by the binomial formula. Anything that's like a series of coin flips (not necessarily fair: $p$ can be anything from 0 to 1) can be modeled using the binomial. In particular:

- *Sampling with replacement*: The number of individuals in the *sample* with a certain characteristic will follow a binomial distribution, where $p$ is the proportion of individuals in the *population* with the characteristic.

- *Sampling without replacement from a huge population*: Even if the sampling is without replacement, the binomial will be a good approximation whenever the population size is orders of magnitude larger than the sample size.

**Example.** In the U.S., about 55% of adults support same sex marriage. Suppose we take a simple random sample (without replacement) of 1000 U.S. adults. The number in this sample who support same-sex marriage will be a random variable with an (approximate) binomial distribution with $n = 1000$ and $p = 0.55$. Calling this random variable $X$, we write

$$X \sim \text{Binomial}(1000, 0.55).$$

What the R function `dbinom()` is really giving you is this $f(x)$ for the binomial.

```
dbinom(3.5, 6, 0.5)
```

## Warning in dbinom(3.5, 6, 0.5): non-integer x = 3.500000

## [1] 0

This is just telling you that the probability of getting exactly 3.5 heads when you flip 6 coins is zero.

**Example.** I toss a fair coin 50 times. What's the probability of getting 10 heads or less? If we have the PMF, we can just find us as:

$$P(X \leq 10) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) + f(6) + f(7) + f(8) + f(9) + f(10)$$

Adding things up is tiresome, so let's get R to do the work for us. The CDF of a binomial random variable can be evaluated by the R function `pbinom()`:

```
pbinom(10, 50, 0.5)
```

## [1] 1.193067e-05

gives the probability that a binomial random variable with $n = 50$ and $p = 0.5$ takes a value of 10 or less.

**Examples using the binomial PMF and CDF**

A flight with 100 seats accepts 105 reservations On average, 90% show up. What's the prob. exactly 100 people show up?

Assuming independence:

```
dbinom(100, 105, 0.9)
```

```
## [1] 0.02564786
```

What's the prob. at least 1 person doesn't get a seat?

```
# P(Y > 100) = 1 - P(Y <= 100)
1 - pbinom(100, 105, 0.9)
```

```
## [1] 0.01671632
```

Note: Is independence really a good assumption? If you don't think so, you'll need to get better data.

We can also use the binomial in combination with itself or other distributions.

**Example.** 100 people each toss 10 fair coins What's the probability the majority get more heads than tails?

```
p = 1 - pbinom(5, 10, 0.5)
1 - pbinom(50, 100, p)
```

```
## [1] 0.004577976
```

**Example.** Let $X$ be a Binomial$(n = 3, p = 0.6)$ random variable.

- *What's the PMF?* By plugging into the binomial formula, we get:

$$
f(x) = \begin{cases}
0.064 & x = 0 \\
0.288 & x = 1 \\
0.432 & x = 2 \\
0.216 & x = 3 \\
0 & \text{otherwise.}
\end{cases}
$$

- *What's the CDF?* Remembering that the CDF is the "less than or equal to" probability, we find its values by adding up values of the PMF.

$$
F(y) = \begin{cases}
0 & y < 0 \\
0.064 & 0 \le y < 1 \\
0.064 + 0.288 = 0.352 & 1 \le y < 2 \\
0.064 + 0.288 + 0.432 = 0.784 & 2 \le y < 3 \\
0.064 + 0.288 + 0.432 + 0.216 = 1 & y \ge 3
\end{cases}
$$

- *What's the expected value?* All we do is multiply each of the possible outcomes by its probability (NOT by the CDF), then take the sum:

$$EX = (0 \times 0.064) + (1 \times 0.288) + (2 \times 0.432) + (3 \times 0.216) = 1.8$$

- *What's the variance?* For each possible outcome $x$, find $(x - \mu)^2 \cdot f(x)$, where $\mu$ is 1.8.

$$x = 0 : (0 - 1.8)^2 \times 0.064 = .20736$$
$$x = 1 : (1 - 1.8)^2 \times 0.288 = .18432$$
$$x = 2 : (2 - 1.8)^2 \times 0.432 = .01728$$
$$x = 3 : (3 - 1.8)^2 \times 0.216 = .31104$$

Add these terms to get the variance: $.20736 + \cdots + .31104 = 0.72$.

- *What's the standard deviation?* It's the square root of the variance, which is about 0.849.

But wait – there's an easier way. If $X_1$ is the number of heads in one $p$-coin toss, and $X_2$ is the number of heads in a second independent $p$-coin toss, and so on, then let $Y$ be the number of heads in $n$ tosses of $p$-coins. Then $Y$ is the sum of $n$ independent Bernoulli random variables. We've already studied this! This means:

- The expected value of a Binomial$(n, p)$ is $np$
- The variance of a Binomial$(n, p)$ is $np(1 - p)$
- The standard deviation of a Binomial$(n, p)$ is $\sqrt{np(1 - p)}$

These formulae save us having to calculate expected value and standard deviation from the definitions every time we see a binomial.

**Binomial combos**

Sometimes we have to do some work to get the right value of $p$ to use in the binomial – by drawing a tree, or using another distribution, or using the binomial itself.

**Example.** (From Takehome 1, Fall 2015.) *I give a ten question true/false statistics test to a class of ten chimpanzees. The chimpanzees, who do not know any statistics, randomly guess true or false, independently for each question and independently of each other. A score of at least 8 out of 10 is required to pass the test. What is the probability that at least one of the chimpanzees passes the test?*

Consider Chimp Number One. What's the probability it passes?

```
prob.pass = 1 - pbinom(7, 10, 0.5)
print(prob.pass)
```

```
## [1] 0.0546875
```

Each of the ten chimps has the same probability of passing. So the number of chimps that pass is a binomial random variable $Y$ with $n = 10$ and $p$ given by `prob.pass` above. The probability at least one chimp passes is $P(Y \geq 1) = 1 - P(Y \leq 0)$:

```
1 - pbinom(0, 10, prob.pass)
```

```
## [1] 0.4301586
```

So there's a 43% chance at least one chimp passes.

If we had 100 chimps, the probability of at least one chimp passing would be:

```
1 - pbinom(0, 100, prob.pass)
```

```
## [1] 0.9963897
```

That is, with 100 chimps, it's almost certain at least one chimp would pass based on random guessing.

The didactic points here are:

1. Things with low probability can happen.
2. Any one thing with a low probability is unlikely to happen.
3. If you consider lots of low probability events, it can be quite likely that one of them will happen.

This all seems very obvious, but failing to consider this – not just by lottery players but also by people with PhDs who should know better – is arguably the major problem with statistical inference in the real world. We'll return to this subject in chapter 12.