Discrete probability

Discrete random variables and probability distributions

A **discrete random variable** is a variable that can only take on discrete values. For example, if we flip a coin twice, we can only get heads zero times, one time, or two times. We can't get heads 1.5 times, or 0.31 times. The number of heads we can get takes on a discrete set of values: 0, 1, and 2. A **continuous random variable**, on the other hand, can take on any value in a certain interval.

In probability distributions for all random variables, the probabilities of each of the possibilities has to sum to 1, or 100%.

For example, if I flip a coin twice, I can get any of the following outcomes:

$$HH$$

$$HT$$

$$TH$$

$$TT$$

There are four possible outcomes, and one of them where I get 0 heads, so the probability of getting 0 heads is $1/4$. In HT and TH I get 1 heads, so the probability of getting 1 heads is $2/4$. In HH I get 2 heads, so the probability of getting 2 heads is $1/4$.

Now we can tell that this is a valid discrete probability distribution, because

$$\frac{1}{4}+\frac{2}{4}+\frac{1}{4}=1=100\,\%$$

The fact that a valid probability distribution always sums to $100\,\%$ allows us to find missing values in our data. For example, if instead we'd been told that the table below tells us the probability of getting a certain number of heads when we flip a coin twice,

Heads	Probability
0	0.25
1	0.50
2	

we could calculate the missing value by subtracting the known probabilities from 1.00. So we could say that the probability of getting exactly 2 heads is

$$P(2\text{ heads})=1.00-0.25-0.50=0.25$$

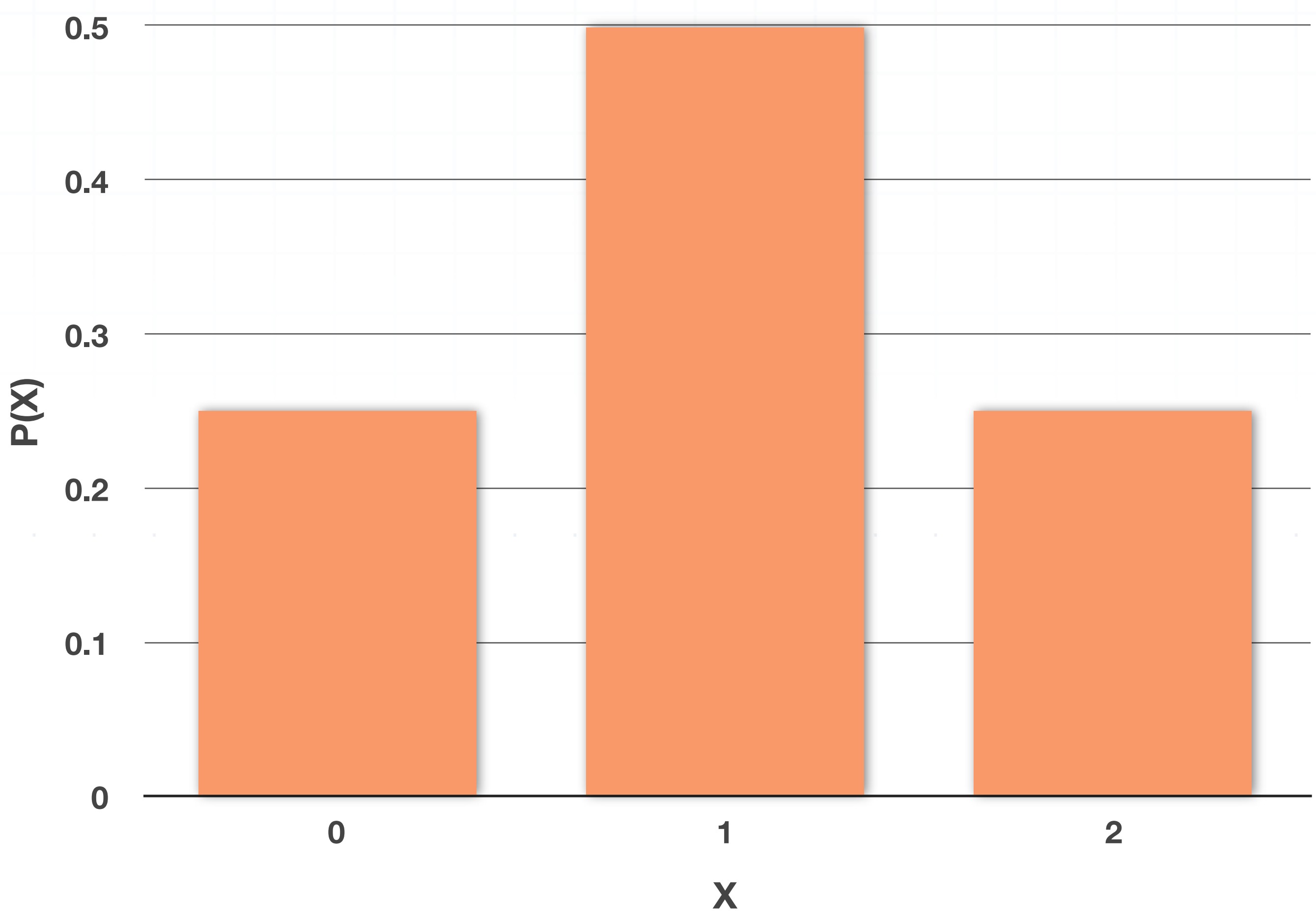
And then we could complete the table.

Heads	Probability
0	0.25
1	0.50
2	**0.25**

Keep in mind that we often use capital X to represent a discrete random variable. Which means that, for the example of flipping the coin twice, we could call the number of heads X, and the probability of getting a certain number of heads $P(X)$. And we could give the probability distribution table as

X	P(X)
0	0.25
1	0.50
2	0.25

Or we could take the same information and graph the distribution this way:



Expected value

Once we have a probability distribution for a discrete random variable, X, we can calculate the **expected value** $E(X)$, which is the mean of X. The expected value is often referred to as the "long term average." When we run an experiment over and over and over again, this is the mean we'd *expect* to find. To find this value for a discrete random variable, we have to "weight" each value.

For example, if we want to find the expected value for the number of heads when we flip a coin two times, we'll multiply each value of X by the corresponding value of $P(X)$, and then add them all together.

X	P(X)
0	0.25
1	0.50
2	0.25

So the expected value is

$$E(X) = \mu_X = 0(0.25) + 1(0.50) + 2(0.25)$$

$$\mu_X = 0 + 0.50 + 0.50$$

$$\mu_X = 1$$

Therefore, on average, we'll expect to get 1 heads when we flip a coin two times. We can then extrapolate this to guess that, for example, we should get 50 heads when we flip a coin 100 times.

Variance and standard deviation

We can also find the variance and standard deviation for discrete random variables. To find the variance, we'll take the difference between X and the mean, μ_X, square that difference, and then multiply the result by the probability of X, called $P(X)$. We'll do that for each value of X, and then add all those results together to get the **variance**, σ_X^2.

$$\sigma_X^2 = \sum_{i=1}^{n} (X_i - \mu)^2 P(X_i)$$

Let's find the variance when X is the number of heads we get when we flip a coin two times, remembering that we already found $E(X) = 1$ for this probability distribution.

$$\sigma_X^2 = (0 - 1)^2(0.25) + (1 - 1)^2(0.50) + (2 - 1)^2(0.25)$$

$$\sigma_X^2 = (-1)^2(0.25) + (0)^2(0.50) + (1)^2(0.25)$$

$$\sigma_X^2 = 1(0.25) + 0(0.50) + 1(0.25)$$

$$\sigma_X^2 = 0.25 + 0.25$$

$$\sigma_X^2 = 0.50$$

We can also find the standard deviation of X, σ_X, which is just the square root of the variance.

$$\sigma_X = \sqrt{0.50}$$

$$\sigma_X \approx 0.71$$

Let's do an an example where we find mean, variance, and standard deviation of a discrete random variable.

Example

We're playing a game of chance in which a computer randomly chooses four numbers from 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, with replacement. We pay $\$3$ to play the game. If we pick the same four numbers as the computer, we win $\$10{,}000$ and get our $\$3$ back, so we profit $\$10{,}000$. If we fail to match all four of the computer's numbers, we lose our $\$3$, and our profit is $-\$3$. What is our expected profit in the long run if we play this game over and over again?

Let X be the profit on each play. X is a discrete random variable whose possible values are only -3 and $10{,}000$.

Since the computer picks from 10 numbers, the probability that we choose one of the computer's numbers is $1/10$, and the probability of choosing all four numbers correctly is

$$\left(\frac{1}{10}\right)\left(\frac{1}{10}\right)\left(\frac{1}{10}\right)\left(\frac{1}{10}\right)=\frac{1}{10{,}000}=0.0001$$

Therefore, the probability we win is 0.0001 and the probability we lose is $1-0.0001=0.9999$.

X	P(X)	XP(X)
-3	0.9999	-2.9997
10,000	0.0001	1

Now we can find the expected value of playing the game long term.

$$E(X) = \mu_X = -2.9997 + 1 = -1.9997$$

So if we play this game over and over again, we expect to lose a little less than $\$2$ each time we play. To think about it a different way, here's our expected profit based on number of games played.

If we play 100 games, we expect approximately
$100(-\$1.9997) = -\$199.97$ in loss

If we play $1{,}000$ games, we expect approximately
$1{,}000(-\$1.9997) = -\$1{,}999.70$ in loss

If we play $10{,}000$ games, we expect approximately
$10{,}000(-\$1.9997) = -\$19{,}997.00$ in loss

The variance is

$$\sigma_X^2 = (-3 - (-1.9997))^2(0.9999) + (10{,}000 - (-1.9997))^2(0.0001)$$

$$\sigma_X^2 = (-1.0003)^2(0.9999) + (10{,}001.9997)^2(0.0001)$$

$$\sigma_X^2 = 1.0005 + 10{,}003.9998$$

$$\sigma_X^2 \approx 10{,}005$$

so the standard deviation is

$$\sigma_X \approx 100.02$$
