Discrete Random Variables

Contents

Probability spaces

In whatever we write below, everything is conditioned on our current information I. Because this information is always on the background, we will not be explictily showing it in our notation.

Assume that we are doing an experiment. It doesn't matter what exactly the experiment is. The result of the experiment depends on the values of some physical variables ω which may be unknown to us (epistemic uncertainty) or truly random (aleatory uncertainty). In the language of mathematical probability theory, this ω is called an **event**. The space of all possible ω's, denoted by Ω, is called the **event space**. For today, assume that Ω is a discrete space (otherwise things become a little bit more complicated).

Since, we are uncertain about which ω will appear in nature, we need to assign probabilities over the possible values. Ideally, what we would like to have is some function $\mathbb{P}(A)$ that takes an arbitrary subset A of Ω and tells us how probable it is. That is $\mathbb{P}$ is a function from all subsets of Ω, $\mathcal{F} = \mathcal{P}(\Omega)$, to the real numbers:

$$\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R},$$

There are a few things that this function should satisfy for all A in $\mathcal{F}$

- It should be nonnegative, i.e., $\mathbb{P}(A) \geq 0$.
- One of the ω's must happen, $\mathbb{P}(\Omega) = 1$.
- The obvious rule $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, where $A^c = \Omega \setminus A$ is the complement of A. When these properties are satisfied, we say that $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

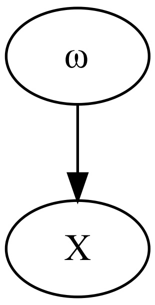
The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Note: If we wanted to show the background information we would be writting $\mathbb{P}[A|I]$.

The mathematical definition of a random variable

Now assume that we are doing a specific experiment that measures something, say an integer. Assume that the physical variables that determine what is the result of the experiment are ω and they take values in a set Ω. We are uncertain about the ω's and we have described this uncertainty using a probability measure $\mathbb{P}$ on some subsets $\mathcal{F}$ of Ω. Call X the result of the experiment. The graph is as follows:

```
from graphviz import Digraph
g = Digraph('omega_X')
g.node('omega', label='<&omega;>')
g.node('X')
g.edge('omega', 'X')
g.render('omega_X', format='png')
g
```



This brings us to the mathematical definition of a random variable:

> A random variable is a function of the event space $X(\omega)$. Maps particular event to some particular quantifiable number

Note: if the event space is not discrete, we need some more restrictions on these functions. You need to take a probability theory course to learn about the technical details.

Now, if $X(\omega)$ takes discrete values, like heads or tails, $0, 1, 2$, etc., then we say that X is a discrete random variable. If $X(\omega)$ takes continuous values, like real numbers, then we say that X is a continuous random variables. Today, we are only going to work with discrete random variables.

Notation:

- We will be using upper case letters to represent random variables, like X, Y, Z.
- We will be using lower case letters to represent the values of random variables, like x, y, z.

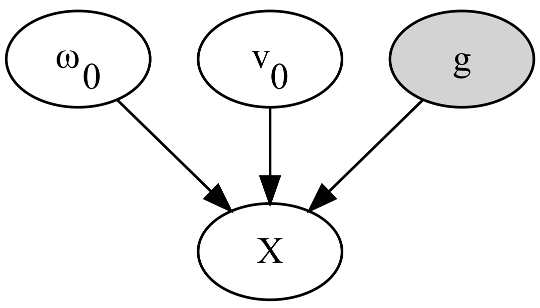
Example: The random variable corresponding to the result of a coin toss (1/2)

Let's consider again the coin tossing example we introduced in the previous lecture. Remember that we denoted with v_0 and ω_0 the initial velocity and angular velocity of the coin. Then, we showed that the variable X representing the coin toss can be predicted exactly, if we knew v_0 and ω_0. Specifically, we derived the following relationship between the result of the coin toss and the initial conditions:

$$X = \begin{cases} T, & \text{if } \frac{2v_0\omega_0}{g} \pmod{2\pi} \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \\ H, & \text{otherwise.} \end{cases}$$

Graphically, this relationship can be represented by:

```
gct = Digraph('coin_toss_g')
gct.node('omega0', label='<&omega;<sub>0</sub>>')
gct.node('v0', label='<v<sub>0</sub>>')
gct.node('g', style='filled')
gct.node('X')
gct.edge('g', 'X')
gct.edge('v0', 'X')
gct.edge('omega0', 'X')
gct.render('coin_toss_g', format='png')
gct
```



Then, we argued that the uncertainty about the value of X is induced by our uncertainty about the values of v_0 and ω_0. It is not that the coin toss is random. It is described in extreme detail by Newton's laws. It is that we do not know what the initial conditions are. So, the state of nature is captured by (v_0, ω_0). Notice that essentially the variable X is a function of (v_0, ω_0). We can write:

$$X = X(v_0, \omega_0).$$

You see that the result of the coin toss X is nothing more but a function of the *true state of nature* (v_0, ω_0). It is just that the value of X is uncertain because the state of nature is uncertain. X is an example of a random variable.

The probability mass function

Take a discrete random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Without loss of generality, assume that X can potentially take infinite $\mathbb{N} = \{1, 2, \dots\}$. Why is this sufficient? The values of X are countably infinite (countable)

- If X takes finite values then we can simply set the probability of values after a given number equal to zero.
- If the values are of another type (e.g., heads and tails) you can just map them to the natural numbers.

The probability mass function of the random variable X, denoted by $f_X(x)$, gives the probability of X taking the value x. Mathematically, it is defined by:

$$f_X(x) := \mathbb{P}(X = x) = \mathbb{P}\left(\{\omega : X(\omega) = x\}\right).$$

Notice that we are just gathering in a set all the states of nature ω that give an experiment with value x, $X = k$, and then we find probability of that set.

If you are 100% sure about which random variable you are talking about, feel free to use the much simpler notation:

$$p(x) \equiv p(X = x) \equiv f_X(x) = \mathbb{P}\left(\{\omega : X(\omega) = x\}\right).$$

This is the notation we will employ from this point on. We will only use the strict mathematical notation when we have no choice.

Note: If we wanted to show the background information we would be writing $p(x|I)$.

Properties of the probability mass function

There are some standard properties of the probability mass function that is worth memorizing:

- The probability mass function is nonnegative:

$$p(x) \geq 0,$$

for all x in $\mathbb{N}$.

- The probability mass function is normalized:

$$\sum_{x=0}^{\infty} p(x) = 1.$$

This is a direct consequence of the fact that X must take a value.

- Take any set of possible values of X, A. The probability of X taking values in A is:

$$p(X \in A) = \sum_{x \in A} p(x).$$

Example: The random variable corresponding to the result of a coin toss (2/2)

Let's write down the probability mass function of the coin toss random variable X. Without loss of generality, we can map heads to the number 0 and tails to the number 1. We need to specify the probability of one of these events, as the probability of the other one is trivially defined. For a fair coin we have:

$$p(X = 0) = \text{probability of heads} = \frac{1}{2}.$$

From this, because of the normalization constraint:

$$p(X = 0) + p(X = 1) = 1,$$

we get that:

$$p(X = 1) = \frac{1}{2}.$$

This is an example of a special random variable taking two discrete values 0 and 1, which we call the Bernoulli random variable. We will see it in an example later on.

Functions of discrete random variables

Consider a random variable X taking values in $\mathbb{N}$ with probability mass function $p(x)$. Now, consider a function $g(x)$. We can now define a new random variable:

$$Y = g(X).$$

The this random variable takes values in:

$$g(\mathbb{N}) := \{g(x) : x \in \mathbb{N}\}.$$ g(x) such that x is a natural number

It has its own probability mass function (pmf) which we can define using the pmf of X:

$$p(y) = p(Y = y) = p(X \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} p(x),$$

where $g^{-1}(y)$ is the set of x's that map to y through g, i.e., x is in the pre-image of the mapping to the particular y

$$g^{-1}(y) := \{x \in \mathbb{N} : g(x) = y\}.$$

This is formal definition of the uncertainty propagation problem. The correspondence is that X represents the parameters of a physical model, and $Y = g(X)$ is the uncertain result of the physical model.

Expectation of random variables

The expectation of a random variable is defined to be:

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x p(x).$$

You can think of the expectation as the value of the random variable that one should "expect" to get. However, take this interpretation with a grain of salt because it may be a value that the random variable has a zero probability of getting…

Example: Expectation of a coin toss

The expectation of the coin toss random variable is:

$$\mathbb{E}[X] = 0 \cdot p(X = 0) + 1 \cdot p(X = 1) = 0.5.$$

Of course, this is not a value that the random variable can get.

Properties of the expectation

Here are some properties of the expectation. The proof of some of these properties will be given as homework.

- Take any constant c. Then we have:

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c.$$

- For any λ real number, we also have:

$$\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X].$$

- Take two random variables X and Y. Then we have:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

- Now consider any function $g(x)$. We can now define the expectation of $g(X)$ as the expectation of the random variable $Y = g(X)$. It is quite easy to show that:

$$\mathbb{E}[g(X)] = \sum_{x=0}^{\infty} g(x) p(x).$$

- Assume that $g(x)$ is a convex function, then:

$$g\left(\mathbb{E}[X]\right) \le \mathbb{E}[g(X)].$$

This is known as Jensen's inequality. The convex transformation of a mean is less than or equal to the mean applied after convex transformation.

Variance of random variables

The variance of a random variable X is defined as the expectation of the square deviation from its expectation, i.e.:

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

You can think of the variance as the spread of the random variable around its expectation. However, do not take this too literally for discrete random variables.

Example: Variance of a coin toss

Let's calculate the variance of the coin toss. We need:

$$\mathbb{E}\left[X^2\right] = 0^2 \cdot p(X=0) + 1^2 \cdot p(X=1) = 0.5.$$ g(X) = X^2

So, using the formula above we get:

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 = 0.5 - (0.5)^2 = 0.5 - 0.25 = 0.25.$$ By FOIL-ing

Properties of the variance

Here are some properties of the variance.

- It holds that:

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2.$$

- For any constant c, we have:

$$\mathbb{V}[X + c] = \mathbb{V}[X].$$

- For any constant c, we have:

$$\mathbb{V}[cX] = c^2\mathbb{V}[X].$$

By Ilias Bilionis (ibilion[at]purdue.edu)