What is a Random Variable?

Instructor: David Dobor

CIS 2033, Spring 17. February 9, 2017

We carry out a probabilistic experiment. For example, we pick a random student out of a given class. At the end of the experiment we report whether the student was female – that is, we report whether a certain event has occurred. But besides reporting which events occurred, we may also wish to report some other results of the experiment. For example, report the weight of the selected student. Reporting the weight of a student is not quite the same as reporting an event: we are reporting the *numerical value of some quantity associated with the outcome of the experiment*. Such a quantity will be called a *random variable*.

Random variables make the subject of probability much richer and allow us to talk about random numerical quantities and their relations. In the next few lectures we define random variables, talk about ways of describing them, and introduce certain ways of summarizing their properties, namely the expected value, and the variance.

We introduce a fair amount of definitions and notation about the distribution of a random variable. To a large extent, this involves concepts that you're already familiar with but in new notation. We also define the expected value and the variance and a look at some of their properties. We then continue our discussion related to conditioning. We will discuss conditional counterparts of all the concepts that we introduce as well as the concept of independence of random variables.

Random variables can be discrete or continuous. Discrete random variables are conceptually and mathematically much easier. For this reason, in the next two or three lectures, we deal exclusively with discrete random variables aiming to develop a solid understanding.

A WORD OF CAUTION is in order. Unless you make sure that you understand very well every single concept and formula in this part of the course, interpreting the corresponding concepts and formulas will be a real challenge when we move on to continuous random variables. I would also recommend that you pay special attention to notation. Good notation helps you think clearly.

Introduction

In this note we introduce the notion of a random variable. A random variable is, loosely speaking, a numerical quantity whose value is determined by the outcome of a probabilistic experiment. The height of a randomly selected student in your CIS 2033 class is one example.

After giving a general definition, we will focus exclusively on discrete random variables. These are random variables that take values in finite or countably infinite sets. For example, random variables that take integer values are discrete. To any discrete random variable we will associate a probability mass function, which tells us the likelihood of each possible value of the random variable.

Definition of random variables

WE WILL NOW define the notion of a random variable. Very loosely speaking, a random variable is a numerical quantity that takes random values. But what does this mean? We want to be a little more precise and I'm going to introduce the idea through an example.

Suppose that our sample space is a set of students labeled according to their names. For simplicity, let's just label the students as a, b, c, and d.

Figure 1: Our sample space is the set of four students a, b, c and d.

Our probabilistic experiment is to pick a student at random according to some probability law and then record their weight in kilograms. So for example, suppose that the outcome of the experiment was student a, and the weight of that student is 62. Or it could be that the outcome of the experiment is student c, and that c has a weight of 75 kilograms.

The weight of a *particular* student is a number, call it w (small w). But let us think of the *abstract concept of weight*, something that we will denote by W (capital W). Weight is an object whose value is determined once you tell me the outcome of the experiment, that is, once you tell me which student was picked. In this sense, weight is really *a function of the outcome of the experiment*.

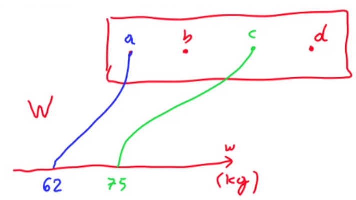

Figure 2: Weight W is a function that associates a particular weight w to each of the students a, b, c and d. Weights of a and b are shown here.

So think of weight as an abstract box that takes as input a student and produces a number, little w, which is the weight of that particular student. Or, more concretely, think of weight with a capital W as a procedure that takes a student, puts him or her on a scale, and reports the result.

In this sense, weight is an object of the same kind as the square root function that's sitting inside your computer. The square root function, perhaps called `sqrt()`, is a subroutine, a piece of code, that takes as input a number, let's say the number 9, and produces another number, in this case the number 3.

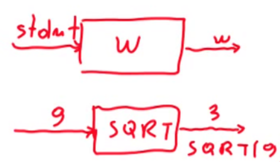

Figure 3: You may think of W in the same way you think of a function such as `sqrt()`. W takes as input a student and outputs his or her weight w.

Notice here the distinction that we will keep emphasizing over and over: *square root of 9, `sqrt(9)`, is a number. It is the number 3. The `sqrt()` is a function.*

Now, let us go back to our probabilistic experiment. Note that a probabilistic experiment such as the one in our example can have several associated random variables. For example, we could have another random variable denoted by H, which is the height of a student recorded in meters.

So if the outcome of the experiment was student a for example, then this random variable H would take a value which is the height of student a, let's say it was 1.7. Or, if the outcome of the experiment was student c, then we would record the height of that student; let's say it turns out to be 1.8. Once again, height H is an abstract object, a function whose value is determined once you tell me the outcome of the experiment.

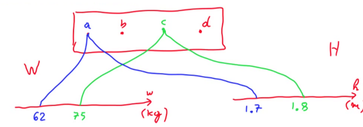

Figure 4: We can associate more than one random variable with the same probabilistic experiment.

Now, given some random variables, we can create new random variables as functions of the original random variables. For example, consider the quantity defined as weight divided by height squared. This quantity is the so-called body mass index,

$$B = \frac{W}{H^2}, \tag{1}$$

and it is also a function on the sample space.

Why is it a function on the sample space? Well, because once an outcome of the experiment is determined, we can compute the body mass index of the selected student: suppose that the outcome of the experiment was student a, then the two numbers, 62 and 1.7, that students weight and height, are also determined. Using those numbers, we can carry out the calculation in formula (1) and find the body mass index of that particular student, which in this case would be 21.5. Similarly, if it happened that student c was selected, then the body mass index would turn out to be some other number. In this case, it would be 23.

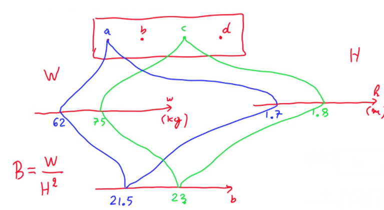

Figure 5: We can associate more than one random variable with the same probabilistic experiment.

Again, we see that the body mass index can be viewed as an abstract concept defined by formula (1). But once an outcome is determined, then the body mass index is also determined. And so the body mass index is really a function of the particular outcome that was selected.

Let us now abstract from the previous discussion. We have seen that random variables are abstract objects that associate a specific value, a particular number, to any particular outcome of a probabilistic experiment. So in that sense, *random variables are functions from the sample space to the real numbers*. They are numerical functions, but as numerical functions they can either take discrete values, for example the integers, or they can take continuous values, let's say on the real line.

For example, if your random variable is the number of heads in 10 consecutive coin tosses, this is a *discrete random variable* that takes values in the set from 0 to 10. If your random variable is a measurement of the time at which something happened, and if your timer has infinite accuracy, then the timer reports a real number and we would have a *continuous random variable*. In this and the next few lectures, we will concentrate on discrete random variables because they are easier to handle. Later on we will move to a discussion of continuous random variables.

Throughout, we want to keep noting this very important distinction that we already brought in the discussion for a particular example, but it needs to be emphasized and re-emphasized. We make a distinction between random variables, which are abstract objects, and the numerical values taken on by the random variables. Random variables are functions on the sample space and they are denoted by uppercase letters; in contrast, we will use lower case letters to indicate numerical values of the random variables. For example, little x is always a real number, as opposed to the random variable X, which is a function.

We emphasize the distinction between random variables, which are abstract objects (such as the notion of weight), and the numerical values taken on by the random variables (such as the weight of a particular person).

Notation: random variable X numerical value x

One point that we made earlier is that we can have several random variables associated with a single probabilistic experiment (think of the weight and height of a randomly selected student). Moreover, we can combine random variables to form new random variables (think of the body mass index). In general, a function of random variables has numerical values that are determined by the numerical values of the original random variables, which are themselves determined by the outcome of the experiment. So a function of random variables is completely determined by the outcome of the experiment and is thus also a random variable.

As an example, we could think of two random variables, X and Y, associated with the same probabilistic experiment, and then define a random variable, let's say $X+Y$. What does that mean? $X+Y$ is a random variable that takes the value little x plus little y when the random variable X takes the value x and Y takes the value y. So X and Y are random variables. $X+Y$ is another random variable. X and Y will take numerical values once the outcome of the experiment has been obtained. And if the numerical values that they take are x and y, then the random variable $X+Y$ will take the numerical value $x+y$.

Meaning of $X+Y$: Random variable $X+Y$ takes value $x+y$ when X takes the value x and Y takes the value y.

Check Your Understanding:

Let X be a random variable associated with some probabilistic experiment, and let x be a number.

(a) Is it always true that $X + x$ is a random variable?

(b) Is it always true that $X - x = 0$?

(a) Think of a concrete example. Let X be the height of a randomly selected student and let $x = 10$. We are dealing with the random variable $X + 10$. It is the random variable that takes the value $a + 10$, whenever the random variable X takes the value a. So yes.

(b) Think of the same concrete example as before. The object $X - 10$, where X is the height of a randomly selected student, has no reason to be equal to 0. (We often use x to denote the realized value of X. But the problem statement never said that the number x considered here had any relation to the realized value of X.) So no.

Probability mass functions

A random variable can take different numerical values depending on the outcome of the experiment. Some outcomes are more likely than others; similarly, some of the possible numerical values of a random variable will be more likely than others.

We restrict ourselves to discrete random variables, and we will describe these relative likelihoods in terms of the so-called *probability mass function*, or PMF for short, which gives the probability of the different possible numerical values. The PMF is also sometimes called the *probability law* or the *probability distribution* of a discrete random variable.

Let us illustrate the idea in terms of a simple example. We have a probabilistic experiment with four possible outcomes. We also have a probability law on the sample space. To keep things simple, we assume that all four outcomes in our sample space are equally likely.

We introduce a random variable that associates a number with each possible outcome as shown in figure 7. The random variable X can take one of three possible values, namely 3, 4, or 5. Let us focus on one of those numbers – let's say the number 5.

We can think of the event that $X = 5$. Which event is this? This is the event that the outcome of the experiment led to the random variable taking a value of 5. We can identify this event in the sample space: it consists of two elements, namely a and b.

More formally, the event that we're talking about is the set of all outcomes for which the value – the numerical value of our random variable, which is a function of the outcome – happens to be equal to 5. In this example it happens to be a set consisting of two elements. That is, $X = 5$ is an event:

$$\text{Event } (X = 5) = \{\omega \in \Omega : X(w) = 5\} = \{a, b\}$$

and this event has a probability assosiated with it. We will denote that probability as follows:

$$p_X(5) : \text{ denotes the probability of the event } X = 5.$$

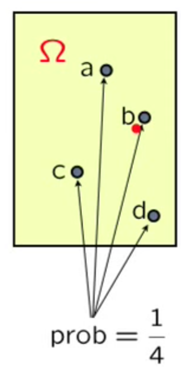

Figure 6: Sample space for an experiment with 4 equally likely outcomes.

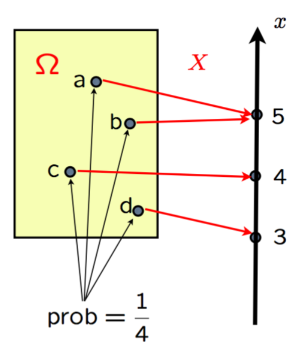

Figure 7: Defining a random variable on the sapmple space.

In our case this probability is equal to 1/2, because we have two outcomes, each one has probability 1/4 and these two probabilities add up to 1/2.

More generally, we will be using the follwoing notation to denote the probability of the event that the random variable X takes on a particular value x:

$$p_X(x) = P(X = x)$$

This is just a piece of notation, not a new concept. We're dealing with a probability, and we indicate it using this particular notation.

More formally, the probability that we're dealing with is the probability – the total probability – of all outcomes for which the numerical value of our random variable is this particular number, x:

$$\boxed{p_X(x) = \mathbf{P}(X = x) \;= \mathbf{P}\Big(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\}\Big)}$$

A few things to notice. We use a subscript, X, in $p_X(\cdot)$ to indicate which random variable we're talking about. This will be useful if we have several random variables involved. For example, if we have another random variable on the same sample space, Y, then it would have its own probability mass function which would be denoted with this particular notation:

$$p_Y(y).$$

The argument of the PMF, which is the little x in $p_X(x)$, ranges over the possible values of the random variable (capital) X. So here we're really dealing with a function – a function that we could denote just by p_X. Moreover, we can *plot this function*. Let's do that.

In our particular example, the interesting values of x are 3, 4, and 5. The associated probabilities are as follows:

- the value of 5 – this is the event that the outcome was either a or b – is obtained with probability 1/2,
- the value of 4 – this is the event that the outcome is c – has probability 1/4,
- the value of 3 – this is the event that the outcome is d – is also obtained with probability 1/4.

So the probability mass function is a function of an argument x. And for any x, it specifies the probability that the random variable takes on that particular value x.

Figure 8: **Probability Mass Function** of the random variable X. The random variable X is also shown in figures 7 and 9.

A few more things to notice. The probability mass function is always non-negative, since we're talking about probabilities and probabilities are always non-negative:

$$p_X(x) \geq 0.$$

In addition, since the total probability of all outcomes is equal to 1, the probabilities of X taking these different values should also add to 1:

$$\sum_x p_X(x) = 1.$$

In terms of our picture, the event that $X = 3$, the event circled in red in figure 8, the event that $X = 4$, the event circled in green in figure 8, and the event that $X = 5$, the event circled in blue in figure 8, are disjoint, and together they cover the entire sample space. So their probabilities should add to 1. And the probabilities of these events are the probabilities of the different values of the random variable X. So the probabilities of these different values should also add to 1:

$$p_X(3) + p_X(4) + p_X(5) = 1.$$

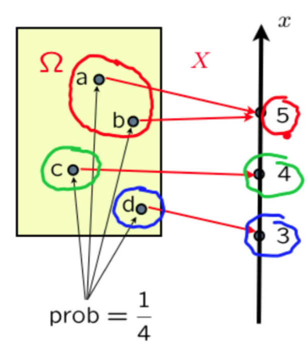

Figure 9: Probabilities of the disjoint events comprising the sample space - the red, blue and green events - add up to 1: 1/2 + 1/4 + 1/4 = 1.

Here's a summary of these properties in our new notation:

- **Properties:** $p_X(x) \geq 0$ $\quad \sum_x p_X(x) = 1$

Example

Let us now go through an example to illustrate the general method for calculating the PMF of a discrete random variable. We will revisit our familiar example involving two rolls of the four-sided die and let X be the result of the first roll and Y be the result of the second roll.

Notice that we're using uppercase letters. And this is because X and Y are random variables.

In order to do any probability calculations, we also need the probability law. To keep things simple, let us assume that every possible outcome has the same probability. Since we have 16 outcomes, each one gets assigned the probability of 1/16.

We will concentrate on a particular random variable, which we call Z, defined to be the sum of the random variables X and Y. So if X and Y both happen to be 1, then Z will take the value of 2. If X is 2 and Y is 1 our random variable Z will take the value of 3, and so on.

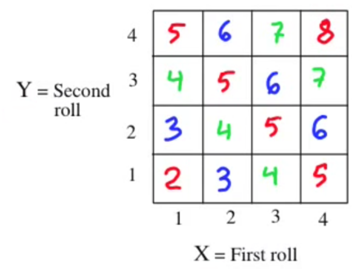

Figure 10: The familiar tetrahedral dice. We are interested in the random variable Z, which we define as the sum of the two faces shown. Thus $Z = X + Y$. We filled in the squares with the values z taken on by the random variable Z.

What we want to do now is to calculate the PMF of this random variable Z. What does it mean to calculate the PMF? We need to find its value for all choices of z, that is for all possible values in the range of our random variable Z.

The way we're going to do it is to consider each possible value of z, one at a time, and for any particular value find out what are the outcomes – the elements of the sample space – for which our random variable Z takes on the specific value, and add the probabilities of those outcomes:

- repeat for all z:
  - collect all possible outcomes for which Z is equal to z
  - add their probabilities

To illustrate this process, let us calculate the value of the PMF when z is 2. This is by definition the probability that our random variable Z takes the value of 2; that is, we want $P(Z = 2) = p_Z(2)$. Now, the event $Z = 2$ can only happen when $X = 1$ and $Y = 1$ which corresponds to only one element of the sample space – which is the lower left square in figure 9 – which has probability 1/16.

We can continue the same way for other values of z. For example, the value of PMF at z equal to 3 is the probability that $Z = 3$. This is an event that can happen in two ways – it corresponds to two outcomes marked with the blue "3"s in fugure 9 – and so it has probability 2/16.

Continuing similarly, the probability that $Z = 4$ is equal to 3/16. We can continue this way and calculate the remaining entries of our PMF.

After you are done, you end up with a plot shown in figure 11. We will find it very convenient to plot PMFs of random variables in the rest of the course.

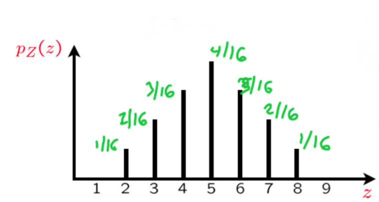

Figure 11: PMF of the random variable Z.

Check Your Understanding: PMF calculation

As in the example just shown, consider two rolls of a 4-sided die, with all 16 outcomes equally likely. As before, let X be the result of the first roll and Y be the result of the second roll. Define $W = XY$. Find the numerical values of $p_W(4)$ and $p_W(5)$.

- The event $W = 4$ may occur in three different ways: $(1,4)$, $(2,2)$, $(4,1)$. Since all 16 outcomes of the two rolls are equally likely, $p_W(4) = P(W = 4) = 3/16$.
- The event $W = 5$ cannot happen, and so $p_W(5) = P(W = 5) = 0$.

Check Your Understanding: Random variables versus numbers

Let X be a random variable that takes integer values, with PMF $p_X(x)$. Let Y be another integer-valued random variable and let y be a number.

(a) Is $p_X(y)$ a random variable or a number?

(b) Is $p_X(Y)$ a random variable or a number?

(a) Recall that $p_X(\cdot)$ is a function that maps real numbers to real numbers. So, when we give it a numerical argument, y, we obtain a number.

(b) In this case, we are dealing with a function, the function being $p_X(\cdot)$, of a random variable Y. And a function of a random variable is a random variable. Intuitively, the "random" value of $p_X(Y)$ is generated as follows: we observe the realized value y of the random variable Y, and then look up the numerical value $p_X(y)$.