

# Lecture 10 (Part 2): Random variables I: definition, events and random variables, independent random variables

Discrete Structures II (Summer 2018)

Rutgers University

Instructor: Abhishek Bhrushundi

References: Relevant parts of chapter 19 of the Math for CS book.

## 1 Introduction to random variables

Recall the steps involved in modelling random experiments: define outcomes in a suitable manner and this defines  $\Omega$  the sample space and define an appropriate probability distribution on  $\Omega$ .  $\Omega$  and  $P$  together are said to define a *probability space*. We denote the probability space by  $(\Omega, P)$ .

Typically, in every random experiment, we are interested in “measuring” some quantity that depends on the outcome of the experiment. Informally, speaking these quantities of interest that are measured during an experiment are called random variables.

For example, consider the experiment of choosing a random person from the population of adults in the age group 18 – 25 in the US (say at the present moment) and then measuring their height. Clearly, height is a quantity that depends on the outcome of the experiment (i.e., it depends on which person was chosen randomly during the experiment). In this case, “height” is a random variable associated with the experiment.

**Definition 1.** Let  $(\Omega, P)$  be a probability space then a random variable  $X$  is a function with domain  $\Omega$  and codomain  $\mathbb{R}$ , i.e.  $X : \Omega \rightarrow \mathbb{R}$ .

Let’s look at some more examples. Consider the experiment of tossing a fair coin 3 times and let  $X$  be the random variable that is equal to the number of heads observed during the 3 tosses. For this experiment, the sample space is nothing but length 3 sequences made from the letters  $H$  and  $T$ . Since we are tossing a fair coin, we define the probability distribution on  $P$  to be the uniform distribution.

$X$  is a function from  $\Omega$  to  $\mathbb{R}$ . Let’s try to see what value  $X$  takes on various outcomes. Recall that

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}.$$

Then,

$$X(TTT) = 0$$

$$X(TTH) = X(THT) = X(HTT) = 1,$$

$$X(THH) = X(HTH) = X(HHT) = 2,$$

$$X(HHH) = 3.$$

For a function  $f : C \rightarrow D$ , recall that the range of  $f$  is all those values in  $D$  which the function  $f$  can take, i.e.

$$\text{Range}(f) = \{d \in D \mid \exists c \in C, f(c) = d\}.$$

For example, consider  $f : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $f(x) = x^2$ , then the codomain of  $f$  as it is defined is  $\mathbb{R}$  but the range consists of all positive real numbers (since the square of a real number can never be negative), i.e.  $\text{Range}(f) = \{x \in \mathbb{R} \mid x > 0\}$ .

The codomain of the random variable  $X$  in the above example is  $\mathbb{R}$  but  $\text{Range}(X) = \{0, 1, 2, 3\}$ .

Here is another example. Suppose you roll a dice, and let  $X$  be the number rolled by the dice. Then  $X$  is a random variable with  $\text{Range}(X) = \{1, 2, \dots, 6\}$ .

Here is one last example which, as we shall see in the next section, is of particular interest. Consider the experiment of tossing a coin 100 times. Let  $Y$  be the random variable that is 1 if all the coin tosses result in the same outcome, and is 0 otherwise. In this case, the range of  $Y$  is just  $\{0, 1\}$ . Let  $E$  denote the event that all the tosses result in the same outcome, then it's almost as if  $Y$  is a “bulb” (i.e.,  $Y$  becomes 1) that turns on at the end of the experiment if event  $E$  happens, and doesn't turn on (i.e.,  $Y = 0$ ) if  $E$  doesn't happen.

## 2 Random variables and events

As we will see, events and random variables have a lot more in common than what one might expect at first. We will show how to use events to define random variables, and then show how random variables naturally define events.

### 2.1 From events to random variables

**Definition 2** (Indicator random variable). *Let  $E$  be an event of a probability space  $(\Omega, P)$ . Define the random variable  $X_E$  as follows: for every outcome  $\omega$  of the sample space  $\Omega$ ,*

$$X_E(\omega) = 1 \text{ if } \omega \text{ is an outcome in } E$$

$$X_E(\omega) = 0 \text{ if } \omega \text{ is NOT an outcome in } E$$

*That is  $X_E$  takes value 1 if the event  $E$  happens, and takes the value 0 if  $E$  doesn't happen.  $X_E$  is called the indicator random variable for the event  $E$  (or associated with the event  $E$ ).*

There are different notations for representing the indicator random variable of an event  $E$ :  $I_E$ ,  $I[E]$ , etc., but I don't mind you using whatever notation you like as long as you explicitly mention that the variable in question is an indicator random variable!

So if we go back to the example of tossing 100 coins, then the random variable  $X$  in that case was actually an indicator random variable for the event of all 100 tosses resulting in the same outcome, i.e.  $E$  is the event consisting of the all heads outcome and the all tails outcome.

## 2.2 From random variables to events

**Definition 3.** Let  $X$  be a random variable defined on a probability space  $(\Omega, P)$  and let  $c \in \mathbb{R}$  be some real number, then  $[X = c]$  denotes the event that  $X$  takes the value  $c$ , or more formally

$$[X = c] = \{\omega \in \Omega \mid X(\omega) = c\},$$

that is the set of all outcomes on which  $X$  takes value  $c$ .

Again, even though I prefer if you use the notation  $[X = c]$  to denote such events, it's okay if you use other notation as long as you make it clear that the event in question is the one that corresponds to  $X$  taking value  $c$ . Another thing to note here is that  $[X = c]$  is an event, it's a subset of  $\Omega$ , and consists of outcomes for which  $X = c$ .

If we go back to the example of tossing three coins and letting  $X$  be the random variable that counts the number of heads observed, then the event  $[X = 2]$  is give by

$$[X = 2] = \{HHT, HTH, THH\}.$$

Obviously, we are not restricted to using equalities involving random variables to define events. I can also define event  $[X > 1]$ , i.e. the event of  $X$  taking a value larger than 1. In the three coin tosses example,

$$[X > 1] = \{HHT, HTH, THH, HHH\}.$$

## 2.3 Probability and random variables

Having seen this duality between random variables and events, we can now ask questions of the following form: "In the three coin toss example, what is the probability that  $0 < X \leq 2$ ?"

Such questions can easily be answered: if  $X$  is a random variable defined on a probability space  $(\Omega, P)$  then the probability of  $X$  taking a value between  $a$  and  $b$  is basically the probability of the event  $[a \leq X \leq b]$ , i.e.  $P([a \leq X \leq b])$ . We will drop the square brackets in such cases and simply write

$$P(a \leq X \leq b).$$

For example, in the three coin tosses example,  $P(0 < X \leq 2)$  is just  $3/4$  (why?). Here are some more examples.

**Question .** A fair coin is tossed 100 times. Let  $X$  be the number of heads observed. What is  $P(X > 1)$ ?

*Proof.* We are interested in the probability of the event  $[X > 1]$ . Note that the range of  $X$  is  $\{0, 1, \dots, 100\}$ . So the complement of the event  $[X > 1]$  is  $[0 \leq X \leq 1]$ . In fact,  $[0 \leq X \leq 1]$  can be written as the union of two disjoint events:

$$[0 \leq X \leq 1] = [X = 0] \cup [X = 1].$$

(Convince yourself!)  $[X = 0]$  is the event that none of the 100 coin tosses result in heads and thus

$$P(X = 0) = \frac{1}{2^{100}},$$

while  $[X = 1]$  is the event of seeing exactly one heads during the 100 tosses, and so

$$P(X = 1) = \frac{100}{2^{100}}$$

(Can you say why?). Thus, using the sum rule for probability we know that

$$P(0 \leq X \leq 1) = P(X = 0) + P(X = 1) = \frac{101}{2^{100}}.$$

Since  $[X > 1]$  is the complement of  $[0 \leq X \leq 1]$ , we get

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - \frac{101}{2^{100}}.$$

□

**Question .** A fair coin is tossed 100 times. Let  $E$  be the event of seeing an even number of heads, and let  $X$  be the indicator random variable of  $E$ . What is  $P(X = 0)$ ?

*Proof.* Since  $X$  is the indicator random variable of  $E$ ,  $X$  is 1 when  $E$  happens and is 0 if  $E$  doesn't happen. Thus,

$$P(X = 0) = P(E \text{ doesn't happen}) = P(\text{Number of heads observed is odd}).$$

Recall that for coin tosses one possible choice of the sample space is the set of all binary strings of a given length. In this case we are tossing a coin 100 times, so we define the sample space as the set of all binary strings of length 100. Also recall from early lectures that the number of binary strings with an even number of ones is the same as the number of binary strings with an odd number of ones. This means that

$$P(X = 0) = P(\text{Number of heads observed is odd}) = \frac{1}{2}.$$

(It might be a good idea to refresh this if you seemed to have forgotten it)

□

### 3 Independent random variables

Suppose we are sampling a random person from the population of adults in the US in the age group 25 – 30. Let  $X$  be the current annual income of the randomly chosen person, and let  $Y$  be the joint income of their parents when they were growing up. Do we expect  $X$  and  $Y$  to be “correlated”, or do we expect these variables to be able to take values “independent” of each other? We expect them to be correlated! For example, if  $Y \geq \$10,000,000$ , it is highly unlikely that  $X \leq \$40,000$ .

On the other hand, if  $Y$  was the month they were born in (we can represent the months using numbers from 1 to 12) then we don't expect  $X$  and  $Y$  to be correlated. We expect the probability of  $X$  taking a certain value to remain unchanged even if  $Y$  was fixed to some value. We can use this to formalize what it means for two random variables to be independent.

**Definition 4.** Let  $(\Omega, P)$  be a probability space and let  $X$  and  $Y$  be random variables defined on it. We say that  $X$  and  $Y$  are independent if, for every  $a \in \mathbb{R}, b \in \mathbb{R}$  the events  $[X = a]$  and  $[Y = b]$  are independent, or in other words, for every  $a \in \mathbb{R}, b \in \mathbb{R}$ ,

$$P(X = a \cap Y = b) = P(X = a)P(Y = b).$$

Note that if some number  $c$  is not the range of  $X$  then  $P(X = c) = 0$ , and similarly if some number  $d$  is not in the range of  $Y$  then  $P(Y = d) = 0$ , and so it doesn't make sense to consider values that are not in the range of  $X$  or  $Y$  in the above definition (because for such values both the LHS and RHS of the equation in the above definition are trivially zero). We can instead say that  $X$  and  $Y$  are independent if and only if for every  $a \in \text{Range}(X)$  and every  $b \in \text{Range}(Y)$ ,  $P(X = a \cap Y = b) = P(X = a)P(Y = b)$  (Convince yourself that this is indeed true).

Let's look at some examples. Suppose we consider the tossing 100 coins experiment again, and let  $X$  be the number of heads we see, and let  $Y$  be the indicator random variable for the event  $E$  that all tosses result in the same outcome (all tosses result in heads or all tosses result in tails), then clearly  $X$  and  $Y$  are not independent:  $P(X = 35 \cap Y = 1) = 0$  because if  $Y = 1$  then  $X$  must be 0 or 100, whereas  $P(X = 35) > 0$  and  $P(Y = 1) > 0$  and so  $P(Y = 1)P(X = 35) > 0$ .

Consider the experiment of rolling two dice, one blue and one red, and let  $X$  be the random variable that is the number rolled by the red dice and let  $Y$  be the number rolled by the blue dice, then  $X$  and  $Y$  are independent. To see this, note that the range of both  $X$  and  $Y$  is  $\{1, \dots, 6\}$ , and for any  $a$  such that  $a < 1$  or  $a > 6$   $P(X = a) = P(Y = a) = 0$ , so only need to prove that for every  $a, b \in \{1, \dots, 6\}$ ,

$$P(X = a \cap Y = b) = P(X = a)P(Y = b).$$

To see this is true, let  $a$  and  $b$  be arbitrary numbers in  $\{1, \dots, 6\}$ . We know that  $P(X = a) = P(Y = b) = 1/6$  (make sure you know why this is true). On the other hand,  $P(X = a \cap Y = b) = \frac{1}{36}$  since there is only one outcome where the red dice rolls the number  $a$  and the blue dice rolls the number  $b$ . Thus,  $P(X = a \cap Y = b) = P(X = a)P(Y = b)$ .

**Question .** Consider the experiment of tossing a coin two times, and let  $X$  be 1 if the 1<sup>st</sup> coin toss is heads and 0 otherwise, and let  $Y$  be the indicator random variable for the event that we observe an even number of heads. Show that  $X$  and  $Y$  are independent.

*Proof.* First notice that the range of  $X$  and  $Y$  is just  $\{0, 1\}$ , and so we need to show that for every  $a, b \in \{0, 1\}$ ,  $P(X = a \cap Y = b) = P(X = a)P(Y = b)$ . Here I will only show you the proof for  $a = 0$  and  $b = 1$ . You can easily complete the proof for the other cases.

We know that  $P(X = 0) = 1/2$  and  $P(Y = 1) = 1/2$ . What about  $P(X = 0 \cap Y = 1)$ ?  $[X = 0] \cap [Y = 1]$  is the event consisting of all outcomes where the first coin is tails and there are an even number of heads. The only possible outcome of this form is  $TT$  (why?), and thus  $P(X = 0 \cap Y = 1) = 1/4$ . This show that  $P(X = 0 \cap Y = 1) = P(X = 0)P(Y = 1)$ .  $\square$