Lecture 11: Random variables II: mutual independence, k-wise independence, some obvious facts about random variables, expectation of a random variable, linearity of expectation

Discrete Structures II (Summer 2018)
Rutgers University
Instructor: Abhishek Bhrushundi

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

1 Mutually independent random variables

Definition 1. *Let $X_1, \ldots, X_n$ be random variables defined on a probability space (Ω, P). Then $X_1, \ldots, X_n$ are said to mutually independent if for all $a_1, a_2, \ldots, a_n \in \mathbb{R}$, the events $[X_1 = a_1], [X_2 = a_2], \ldots, [X_n = a_n]$ are mutually independent, or, in other words, for all $a_1, \ldots, a_n \in \mathbb{R}$,*

$$P(X_1 = a_1 \cap X_2 = a_2) = P(X_1 = a_1)P(X_2 = a_2)$$
$$P(X_1 = a_1 \cap X_3 = a_3) = P(X_1 = a_1)P(X_3 = a_3)$$
$$\vdots$$
$$P(X_1 = a_1 \cap X_n = a_n) = P(X_1 = a_1)P(X_n = a_n)$$
$$P(X_2 = a_2 \cap X_3 = a_3) = P(X_2 = a_2)P(X_3 = a_3)$$
$$\vdots$$
$$P(X_{n-1} = a_{n-1} \cap X_n = a_n) = P(X_{n-1} = a_{n-1})P(X_n = a_n)$$
$$P(X_1 = a_1 \cap X_2 = a_2 \cap X_3 = a_3) = P(X_1 = a_1)P(X_2 = a_2)P(X_3 = a_3)$$
$$\vdots$$
$$P(X_{i_1} = a_{i_1} \cap \ldots X_{i_k} = a_{i_k}) = P(X_{i_1} = a_{i_1}) \cdot \ldots \cdot P(X_{i_k} = a_{i_k})$$
$$\vdots$$
$$P(X_1 = a_1 \cap X_2 = a_2 \cap \ldots \cap X_n = a_n) = P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n).$$

As in the case of the definition of independence for two random variables, we can restrict our attention to the ranges of the random variables instead of all real numbers in the definition(convince yourself that the two definitions are equivalent): $X_1, \ldots X_n$ are independent if and only if for all $a_1 \in Range(X_1), a_2 \in Range(X_2), \ldots, a_n \in Range(X_n)$, the event $[X_1 = a_1], [X_2 = a_2], \ldots, [X_n = a_n]$ are mutually independent.

For example, if we roll a dice 100 times, and X_i is the number rolled in the i^{th} roll, then variables $X_1, \ldots, X_{100}$ are mutually independent.

An obvious yet useful consequence is the following:

Fact 2. *If $X_1, \dots, X_n$ are mutually independent random variables defined on a probability space (Ω, P) then for every $a_1 \in Range(X_1), a_2 \in Range(X_2), \dots, a_n \in Range(X_n)$,*

$$P(X_1 = a_1 \cap X_2 = a_2 \cap \dots \cap X_n = a_n) = P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n).$$

1.1 Detour: a warning about the definition of independence

When are events A, B and C of a probability space (Ω, P) mutually independent? If we know that $P(A \cap B \cap C) = P(A)P(B)P(C)$ can we conclude that they are mutually independent?

The anwer is no! It is possible to have three events A, B and C such that $P(A \cap B \cap C) = P(A)P(B)P(C)$ yet they are not mutually independent (See the example given here: `http://www.engr.mun.ca/~ggeorge/MathGaz04.pdf`). Recall that for the three events to be mutually independent, we need

$$P(A \cap B) = P(A)P(B),\ P(A \cap C) = P(A)P(C),\ P(B \cap C) = P(B)P(C),$$

and

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

It is important to keep this in mind!

2 k-wise independence

Ever wonder why we keep referring to more than 2 independent events/random variables as being "mutually independent" and not simply "independent"? This is because there are various kinds of independence possible when we are dealing with three or more events/random variables[1].

Definition 3 (k-wise independent events)**.** *Let $A_1, \dots, A_n$ (assume $n \geq 3$) be events of a probability space (Ω, P), then $A_1, \dots, A_n$ are said to be k-wise independent (for $2 \leq k \leq n$) if any k of them are mutually independent, i.e. if we pick k events from $A_1, \dots, A_n$, say $A_{i_1}, \dots, A_{i_k}$, then $A_{i_1}, \dots, A_{i_k}$ are mutually independent.*

Of course, if $A_1, \dots, A_n$ are n-wise independent then they are basically mutually independent. When $k = 2$, we say $A_1, \dots, A_n$ are *pairwise* independent events.

All this can be extended naturally to random variables:

Definition 4 (k-wise independent random variables)**.** *Let $X_1, \dots, X_n$ (assume $n \geq 3$) be random variables defined on a probability space (Ω, P), then $X_1, \dots, X_n$ are said to be k-wise independent (for $2 \leq k \leq n$) if any k of them are mutually independent, i.e. if we pick k random variables from among $X_1, \dots, X_n$, say $X_{i_1}, \dots, X_{i_k}$, then $X_{i_1}, \dots, X_{i_k}$ are mutually independent.*

When $k = 2$, we say $X_1, \dots, X_n$ are *pairwise* independent random variables. Let's see an example of three random variables that are pairwise independent but not mutually independent.

Suppose we toss a coin three times. Assume that the coin is fair. Let X be 1 if the first coin toss is heads and 0 otherwise, let Y be 1 if the second coin toss is tails and 0 otherwise, and let Z be

[1]For two random variables, there is only one notion of independence.

the indicator random variable for the event of seeing an even number of heads. Then X and Y are independent (why?), X and Z are independent(why?), and Z and Y are independent (why?). This means that any two out of three random variables are independent. However, X, Y, and Z are not mutually independent. To see this, observe that

$$P(X = 1 \cap Y = 1 \cap Z = 1) = 0 \neq P(X = 1)P(Y = 1)P(Z = 1),$$

which contradicts one of the requirements for them to be independent.

3 Some obvious stuff

Here are some obvious facts/definitions that follow from everything we have seen so far about probability, events, and random variables. I will state them without proof and it's up to you to verify them (I recommend doing it!). Let's assume that X is a random variable defined on a probability space (Ω, P).

1. For every $a \in \mathbb{R}$, $0 \leq P(X = a) \leq 1)$.
2. If $a \notin Range(X)$, then $P(X = a) = 0$.
3. $P(X \in \mathbb{R}) = 1$.
4. The sample space Ω is partitioned into disjoint events by the random variable. e.g., if $Range(X) = \{a_1, \ldots, a_k\} \subset \mathbb{R}$ then the events $[X = a_1], \ldots [X = a_k]$ are (i) disjoint and (ii)their union covers the whole of Ω, i.e. $\bigcup_{i=1}^{k}[X = a_i] = \Omega$. In general, $\bigcup_{a \in Range(X)}[X = a] = \Omega$.
5. $\sum_{a \in Range(X)} P(X = a) = 1$
6. If $a, b \in \mathbb{R}$ such that $a \neq b$ then $P(X = a \cup X = b) = P(X = a) + P(X = b)$.
7. If I_1 and I_2 are disjoint subsets of $\mathbb{R}$, then $P(X \in I_1 \cup X \in I_2) = P(X \in I_1) + P(X \in I_2)$. If $I_1 \cap I_2 \neq \emptyset$ then

$$P(X \in I_1 \cup X \in I_2) = P(X \in I_1) + P(X \in I_2) - P(X \in I_1 \cap I_2).$$

8. $P(X \geq a) = 1 - P(X < a)$
9. Let Y be another random variable define on the same probability space. Then if $b \in Range(Y)$ and $a \in Range(X)$, the probability of X taking the value a given that Y takes the value b is

$$P(X = a|Y = b) = \frac{P(X = a \cap Y = b)}{P(Y = b)}.$$

10. X and Y are independent if and only if for every $a \in Range(X)$ and every $b \in Range(Y)$, $P(X = a|Y = b) = P(X = a)$.

4 Expectation of random variables

Consider the set of all students in the summer 2018 CS 206 class. Let this be our sample space Ω. Furthermore, let us assume that we picking a student at random so that every student is equally likely to be picked, and thus we use the uniform distribution on Ω. Let X the following random variable defined on Ω: for every student $\omega \in \Omega$, $X(\omega)$ is the midterm score of the student ω. Then, the average midterm score of the class is

$$\frac{\sum_{\omega\in\Omega} X(\omega)}{|\Omega|} = \sum_{\omega\in\Omega} X(\omega)\frac{1}{|\Omega|} = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}).$$

(Can you see why all these equalities are true?). The expression $\sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$ is called the *expected value* of the random variable X, and is denoted by $\mathbb{E}[X]$, i.e.

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}).$$

The definition is the same even if we were dealing with a random variable X defined on some arbitrary probability space (Ω, P) where P is not necessarily the uniform distribution on Ω, and is some arbitrary probability distribution on Ω. Even in this general case, one can think of the $\mathbb{E}[X]$ as the "average value" of the random variable X.

The expected value of a random variable provides crucial information about the behavior of the random variable, i.e. what values does the random variable typically take and with what probability. Let's see some examples.

Question . Suppos we toss two fair coins and let X be the number of heads observed. What is $\mathbb{E}[X]$?

Proof. There are four outcomes, i.e. $\Omega = \{HH, HT, TH, TT\}$ and we have the uniform distribution on Ω. Furthermore, the random variable X is defined as follows:

$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0.$$

Then, using the above formula for $\mathbb{E}[X]$ we have

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}) = X(HH)P(\{HH\})+X(HT)P(\{HT\})+X(TH)P(\{TH\})+X(TT)P(\{TT\})$$

$$= 2\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 0\cdot\frac{1}{4} = 1.$$

$\square$

4.1 An alternate definition for expectation

Let us go back to the setup with Ω being the set of all students, a student being randomly chosen, and X being the midterm score of the randomly picked student. Then we saw that the average of the class was basically $\mathbb{E}[X]$. Here is another way to calculate the average of the class. Let's assume that the distinct scores scored by the students on the midterm are $a_1, \ldots, a_k$, i.e. the range of X is $\{a_1, \ldots, a_k\}$. Let's collect all the students who scored a_1 points. Note that this is basically the event/set $[X = a_1]$. Similarly, we can collect all the students who scored a_2 and the set of those students is basically $[X = a_2]$, and we can repeat the same for $a_3, \ldots, a_k$. Thus, we have partitioned the students by their scores into k groups (obviously, the groups are disjoint). Another way to compute the average is as follows (convince yourself):

$$\frac{\sum_{i=1}^{k}(a_i \cdot |[X = a_i]|)}{|\Omega|} = \sum_{i=1}^{k} a_i \frac{|[X = a_i]|}{|\Omega|} = \sum_{i=1}^{k} a_i P(X = a_i) = \sum_{a \in Range(X)} aP(X = a).$$

Thus, in this case,

$$\mathbb{E}[X] = \sum_{a \in Range(X)} aP(X = a).$$

Turns out that one can prove that this is true for all random variables defined on arbitrary probability spaces (Ω, P), and so this is a general formula for the expectation.

Let's see an example.

Question . Suppose we toss two dice, and let X be the sum of the numbers rolled by the two dice. Find $\mathbb{E}[X]$.

Proof. We will use the second definition for expectation for this problem. To do so, first notice that the range of X is $\{2, \ldots, 12\}$, and we need to find $P(X = 2), P(X = 3), \ldots, P(X = 12)$ and then use those values in the above formula. It's not hard to check that (I encourage you to verify this):

$$P(X = 2) = P(X = 12) = \frac{1}{36},$$

$$P(X = 3) = P(X = 11) = \frac{2}{36},$$

$$P(X = 4) = P(X = 10) = \frac{3}{36},$$

$$P(X = 5) = P(X = 9) = \frac{4}{36},$$

$$P(X = 6) = P(X = 8) = \frac{5}{36},$$

$$P(X = 7) = \frac{6}{36}.$$

Thus, using the alternate definition of expectation, we have

$$\mathbb{E}[X] = \sum_{a \in Range(X)} aP(X = a) = \sum_{i=2}^{12} i \cdot P(X = i)$$

$$= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \ldots + 7 \cdot \frac{6}{36} + \ldots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7.$$

$\square$

If you thought the solution to the last problem was a little cumbersome, how about you try this:

Question . Suppose we toss 100 dice, and let X be the sum of the numbers rolled by the 100 dice. Find $\mathbb{E}[X]$.

Obviously, following the same strategy as before might take too long. It turns out that expectation of random variables has a powerful property that makes such problems almost trivial to solve!

4.2 Linearity of expectation

Theorem 5 (Linearity of expectation)**.** *Let X and Y be random variables defined on a probability space (Ω, P). Then,*

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

Notice that we do not require any conditions on X and Y, and most surprisingly, then don't need to be independent!

Before we prove the linearity of expectation, let's go back to the problem of rolling 100 dice. Let X_1 be the number rolled by the first dice, X_2 the number rolled by the second dice, ..., X_{100} be the number rolled by the 100^{th} dice. Then X, the sum of the numbers rolled by all the dice, is simply $X_1 + X_2 + \ldots + X_n$. Using linearity of expectation, we get

$$\mathbb{E}[X] = \sum_{i=1}^{100} \mathbb{E}[X_i].$$

What is $\mathbb{E}[X_i]$ for some i? This is easy to calculate (convince yourself):

$$\mathbb{E}[X_i] = \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5,$$

and thus $\mathbb{E}[X] = 350$. Observe how much simpler this solution is compared to the approach we used earlier.

We will look at another nontrivial application of the linearity of expectation.

Question . Suppose there are 100 rabbits $R_1, \ldots, R_{100}$ and 100 rabbit holes $H_1, \ldots, H_{100}$ such that, to begin with, R_i is in H_i. All the 100 rabbits come out of their holes and are hopping around till they see an eagle swooping down. This makes them go helter-skelter and then run back into the holes so that (i) each rabbit is equally likely to go into any of the 100 holes and (ii) no two rabbits go into the same hole. If X is the number of rabbits that go back into their own hole, what is $\mathbb{E}[X]$?

Proof. Remember that we dealt with this problem when we studided inclusion-exclusion. If you go back and look at the notes from that lecture, you will realize that simply computing the expectation using the formula $\mathbb{E}[X] = \sum_{i=1}^{100} i \cdot P(X = i)$ will be really cumbersome, since even just computing $P(X = 0)$ needs the use of inclusion-exclusion and turns out to be not so easy (imagine doing it for $P(X = 1), P(X = 2), \ldots, P(X = 100)$). It can certainly be done but there is a much much simpler way as we shall see now.

Let X_i be the indicator random variable for the event of R_i going back into its own hole H_i. Then, the total number of rabbits that go back into their own hole is simply $X = X_1 + X_2 + \ldots + X_n$, and thus, using the linearity of expectation, we have that

$$\mathbb{E}[X] = \sum_{i=1}^{100} \mathbb{E}[X_i].$$

Let's compute $\mathbb{E}[X_i]$. We know that the range of X_i is just $\{0, 1\}$, and that

$$P(X = 1) = P(R_i \text{ goes back into } H_i).$$

Also, $P(X = 0) = 1 - P(X = 1)$ (why?). Then, using the definition of expectation of X_i we have

$$\mathbb{E}[X_i] = 1 \cdot P(R_i \text{ goes back into } H_i) + 0 \cdot (1 - P(R_i \text{ goes back into } H_i)) = P(R_i \text{ goes back into } H_i)$$

All we need now is to compute $P(R_i$ goes back into $H_i)$. But this is super-easy! Since we are told that each rabbit is equally likely to go into any of the 100 holes[2], we have that

$$P(R_i \text{ goes back into } H_i) = \frac{1}{100}.$$

(Convince yourself!) This means that

$$\mathbb{E}[X] = \sum_{i=1}^{100} \frac{1}{100} = 1.$$

Thus, on an average, only one rabbit goes into its own hole (Can you see intuitively as to why the expectation is so low!?). □

We used a very simple (yet very useful) fact in the above solution:

Fact 6. *Let X be an indicator random variable of an event E of a probability space (Ω, P). Then*

$$\mathbb{E}[X] = P(E).$$

Proof.

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

□

[2]Another way to see this is to represent all the outcomes using permutations of 1 to 100, using the uniform distribution on the sample space, and then realizing that the event of R_i going back to H_i is the set of all permutations where i appears in position i, and there are 99! such permutations, and thus the probability of R_i going to H_i is $\frac{99!}{100!} = \frac{1}{100}$.