COMPSCI 230: Discrete Mathematics for Computer Science — April 22, 2019

Lecture 25

Lecturer: Debmalya Panigrahi — *Scribe: Erin Taylor*

1 Overview

In this lecture, we introduce random variables. We give examples of common random variables, as well as prove properties about them.

2 Random Variables

Definition 1. *A* random variable *X on a probability space is a total function whose domain is the sample space.*

The codomain of a random variable could be anything, but most often it is the set of real numbers. The range could be much smaller, however, even just $\{0,1\}$ as we will see later. Note that random variables aren't variables, they're functions. Random variables also do not have probabilities, as they are not events.

Example 1: A *Bernoulli random variable*, also called an indicator random variable, has range $\{0,1\}$. Recall that any collection of outcomes in the sample space is an event. Consider the event defined by the set of all outcomes that map to 1. The random variable is said to indicate this event.

Next, we will define the expectation of a random variable. Intuitively, the expectation is similar to an average of the value of random variable over t, and this is exactly the case if all outcomes are equally likely, but in general, not all outcomes may be equally likely which requires the following definition.

2.1 Expectation

Definition 2. *The* expectation *of a random variable X defined on sample space S is*

$$\mathrm{E}[X] = \sum_{\omega \in S} X(\omega)\Pr(\omega).$$

Corollary 1. *For any random variable X,*

$$\mathrm{E}[X] = \sum_{x} x \Pr(X = x).$$

Proof.

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{\omega \in S} X(\omega) \Pr(\omega) \\
&= \sum_{x} \sum_{\omega \in [X=x]} X(\omega) \Pr(\omega) \\
&= \sum_{x} \sum_{\omega \in [X=x]} x \Pr(\omega) && \text{by definition of event } [X = x] \\
&= \sum_{x} x \sum_{\omega \in [X=x]} \Pr(\omega) \\
&= \sum_{x} x \Pr(X = x) && \text{by definition of } \Pr(X = x).
\end{aligned}$$

□

Note that the expectation of a constant c is just the constant: $\mathrm{E}[c] = c$. We will find the expectation of a few random variables next.

Example 2: Consider the Bernoulli random variable X.

$$X = \begin{cases} 1 & \text{with probability (w/p)} \quad p \\ 0 & \text{w/p} \quad 1 - p \end{cases}$$

Then, we can compute the expectation of X:

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

So, the expectation of an indicator random variable is equal to the probability of the event that it indicates.

Example 3: Suppose we were tossing a (possibly biased) coin. The coin lands heads with probability p and tails with probability $(1 - p)$. We continue tossing the coin until we see a heads. How many tosses do we expect to make? We can define the number of coin tosses with a random variable, known as a *geometric random variable*.

If the first toss is a heads, the random variable has value 1. If the first time we see a heads is in the second toss, then the random variable has value 2, and so on. In summary:

$$X = \begin{cases} 1 & \text{w/p} \;\; p \\ 2 & \text{w/p} \;\; (1-p)p \\ 3 & \text{w/p} \;\; (1-p)^2 p \\ & \vdots \\ k & \text{w/p} \;\; (1-p)^{k-1} p \end{cases}$$

Notice that there is no maximum value of this random variable, we could continue tossing the coin forever (though this would likely occur with small probability). Let's compute the expectation of X, which is an arithmetic-geometric sum:

$$
\begin{array}{r}
\mathrm{E}[X] = S = 1 \cdot p + 2(1-p)p + 3(1-p)^2 p + \dots \\
\underline{(1-p)S = (1-p)p + 2(1-p)^2 p + 3(1-p)^3 p + \dots}
\end{array}
$$

$$
\begin{aligned}
\Rightarrow pS &= (1-p)p + (1-p)^2 p + (1-p)^3 + \dots \\
\Rightarrow S &= 1 + (1-p) + (1-p)^2 + (1-p)^3 + \dots \\
&= \frac{1}{1-(1-p)} = \frac{1}{p}
\end{aligned}
$$

For a fair coin, this implies that we would expect to toss the coin two times before we see a heads.

Theorem 2. *For random variables X and Y:*

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

This is known as the linearity of expectation, and is a useful property for calculating expectations.

Proof of Theorem 2.

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{\omega \in S} (X(\omega) + Y(\omega)) \Pr(\omega) \\
&= \sum_{\omega \in S} X(\omega) \Pr(\omega) + Y(\omega) \Pr(\omega) \\
&= \sum_{\omega \in S} X(\omega) \Pr(\omega) + \sum_{\omega \in S} Y(\omega) \Pr(\omega) \\
&= \mathrm{E}[X] + \mathrm{E}[Y]
\end{aligned}
$$

$\square$

Next, we state another simple property of expectations.

Lemma 3. *For random variable X and constant $a \in \mathbb{R}$:*

$$\mathrm{E}[aX] = a \cdot \mathrm{E}[X].$$

Using the above properties of expectation, we can derive the following corollary.

Corollary 4 (Corollary of Theorem 2). *For random variables $X_1, \dots, X_n$ and constants $a_1, \dots, a_n \in \mathbb{R}$, we have:*

$$\mathrm{E}\Big[\sum_{i=1}^{n} a_i X_i\Big] = \sum_{i=1}^{n} a_i \cdot \mathrm{E}[X_i].$$

Let's try to apply these properties.

Example 4: Suppose we toss a coin n times. How many times do we expect to see heads? We can define a random variable to represent the number of times we see heads. Suppose the probability of getting heads on each coin toss was p. Then, at each toss we could define a random variable that takes value 1 when the toss is heads. The sum of these variables is exactly how many heads we see in n tosses. This leads to the following random variable.

A *binomial random variable* is the sum of n identical and independent Bernoulli random variables. In our example, the i^{th} Bernoulli variable is 1 if the i^{th} toss is a heads, i.e.

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ toss is heads (w/p } p) \\ 0 & \text{if the } i^{th} \text{ toss is tails (w/p } 1-p) \end{cases}$$

The number of heads we see, X, is exactly the sum of the X_is. Consider our random variable X. It could take on values 0 (when we see no heads) to n (when we see all heads). What's the probability of seeing k heads? If we fix the k tosses on which we see the heads, this happens with probability $p^k(1-p)^k$. Then, there are $\binom{n}{k}$ ways to pick these k tosses. Thus:

$$X = \begin{cases} 0 & \text{w/p } (1-p)^n p^0 \binom{n}{0} \\ 1 & \text{w/p } (1-p)^{n-1} p^1 \binom{n}{1} \\ & \vdots \\ k & \text{w/p } (1-p)^{n-k} p^k \binom{n}{k} \\ & \vdots \\ n & \text{w/p } (1-p)^0 p^n \binom{n}{n} \end{cases}$$

Let's find the expectation of X. Working the above probabilities seems unwieldy, but recall X is the sum of n Bernoulli random variables: $X = X_1 + \cdots + X_n$.

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[X_1 + \cdots + X_n] \\ &= \mathrm{E}[X_1] + \mathrm{E}_[X_2] + \cdots + \mathrm{E}[X_n] && \text{by linearity of expectation} \\ &= np && X_i \text{ is a Bernoulli random variable with } \mathrm{E}[X_i] = p. \end{aligned}$$

2.2 Independence

Definition 3. *Random variables X and Y are independent iff*

$$\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y).$$

Lemma 5. *If X, Y are independent random variables, then*

$$\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y].$$

When the above equation holds, we say that X and Y are uncorrelated. While independent variables are uncorrelated, the converse is not true. In other words, there are uncorrelated variables that are not independent. Can you find such a pair of random variables?

Proof of Lemma 5.

$$\begin{aligned} \mathrm{E}[XY] &= \sum_x \sum_y xy \Pr(X = x, Y = y) \\ &= \sum_x \sum_y xy \Pr(X = x) \Pr(Y = y) && \text{by independence of } X \text{ and } Y \\ &= \left(\sum_x x \Pr(X = x) \right) \cdot \left(\sum_y y \Pr(Y = y) \right) \\ &= \mathrm{E}[X] \cdot \mathrm{E}[Y]. \end{aligned}$$

$\square$

2.3 Variance

Example 5: Consider the following 3 random variables.

$$X_1 = 1 \text{ w/p } 1$$
$$X_2 = \begin{cases} 0 & \text{w/p } 1/2 \\ 2 & \text{w/p } 1/2 \end{cases}$$
$$X_3 = \begin{cases} -10^6 + 1 & \text{w/p } 1/2 \\ 10^6 + 1 & \text{w/p } 1/2 \end{cases}$$

These three random variables take on very different values, but they have the same expectation. We will consider a measure of how likely a random variable is to be away from its expectation, and this will give us one way to distinguish between the above variables.

Naively, we might think we could capture this information with the quantity $\mathrm{E}[X - \mathrm{E}[X]]$. Why does this not work?

$$\mathrm{E}[X - \mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[X] = 0,$$

for all random variables X. We could instead consider $\mathrm{E}[|X - \mathrm{E}[X]|]$, but absolute values can be quite difficult to work with algebraically. Instead, we use the following definition.

Definition 4. *The* variance *of a random variable X, denoted* $\mathrm{Var}(X)$, *is* $\mathrm{E}[(X - \mathrm{E}[X])^2]$.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X - \mathrm{E}[X])^2] \\
&= \mathrm{E}[X^2 - 2X\mathrm{E}[X] + \mathrm{E}^2[X]] && \mathrm{E}^2[X] \text{ denotes } (\mathrm{E}[X])^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[2X\mathrm{E}[X]] + \mathrm{E}[\mathrm{E}^2[X]] && \text{by linearity of expectation} \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}^2[X] && \text{since } \mathrm{E}[X] \text{ is a constant} \\
&= \mathrm{E}[X^2] - \mathrm{E}^2[X].
\end{aligned}$$

Example 6: Let X be a Bernoulli variable, i.e.,

$$X = \begin{cases} 1 & \text{w/p } p \\ 0 & \text{w/p } 1 - p \end{cases}.$$

Let's calculate the variance of X. Since X only takes on the values 0 and 1, $X^2 = X$. Recall that $\mathrm{E}[X] = p$. Calculating the variance,

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}^2[X] = p - p^2 = p(1 - p).$$

3 Summary

In this lecture, we introduced random variables. We saw examples such as Bernoulli, Geometric, and Binomial random variables. We also defined expectation, independence, and variance and saw examples of each.