Lecture 13 (Part 1): Random variables IV: expectation of products of random variables, covariance

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

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

Before we study expectation of products, let’s look at another distribution, called the uniform distribution (this is different from the uniform measure $P$ on a sample space $\Omega$ which assigns probability $\frac{1}{|\Omega|}$ to each outcome, i.e. all outcomes are equally likely).

1 Uniform distribution

A uniform distribution is a random variable $X$ with range $\{a_1, \ldots, a_n\} \subseteq \mathbb{R}$ (defined on a probability space $(\Omega, P)$) with probability mass function $f_X$ defined as follows: for every $a \in \text{Range}(X)$,

$$f_X(a) = P(X = a) = \frac{1}{n}.$$ 

This just means that $X$ is equally likely to take any of the $n$ values in its range. This also means that the expected value of $X$ is

$$E[X] = \sum_{a \in \text{Range}(X)} aP(X = a) = \sum_{i=1}^{n} a_i P(X = a_i) = \sum_{i=1}^{n} \frac{1}{n} a_i = \frac{\sum_{i=1}^{n} a_i}{n}.$$ 

2 Expectation of products of random variables

Suppose $X, Y$ are random variables defined on $(\Omega, P)$. What is $E[XY]$? Turns out that if $X$ and $Y$ are independent, then there is an easy way to compute this:

**Theorem 1.** Let $X$ and $Y$ be independent random variables then $E[XY] = E[X]E[Y]$.

**Proof.** We will use the law of total expectation to rewrite $E[XY]$:

$$E[XY] = \sum_{b \in \text{Range}(Y)} P(Y = b)E[XY | Y = b],$$

which works because the events $[Y = b]$ for $b \in \text{Range}(Y)$ partition $\Omega$. Also, note that given $Y = b$, the random variable $XY$ is just $bX$. So we get,

$$E[XY] = \sum_{b \in \text{Range}(Y)} P(Y = b)E[bX | Y = b] = \sum_{b \in \text{Range}(Y)} P(Y = b)bE[X | Y = b].$$
Here the last equation uses the linearity of conditional expectation. Let’s substitute the expression for \( E[X|Y = b] \) in the above expression:

\[
\mathbb{E}[XY] = \sum_{b \in \text{Range}(Y)} P(Y = b) \mathbb{E}[bX|Y = b] = \sum_{b \in \text{Range}(Y)} P(Y = b) b \left( \sum_{a \in \text{Range}(X)} aP(X = a|Y = b) \right).
\]

Since \( X \) and \( Y \) are independent, we know that \( P(X = a|Y = b) = P(X = a) \), and so

\[
\mathbb{E}[XY] = \sum_{b \in \text{Range}(Y)} P(Y = b) b \left( \sum_{a \in \text{Range}(X)} aP(X = a) \right) = \sum_{b \in \text{Range}(Y)} P(Y = b) b \mathbb{E}[X].
\]

Here the last equation follows from the definition of \( \mathbb{E}[X] \). Since \( \mathbb{E}[X] \) is a constant that does not depend on \( b \) we can pull it out, and so

\[
\mathbb{E}[XY] = \mathbb{E}[X] \left( \sum_{b \in \text{Range}(Y)} bP(Y = b) \right) = \mathbb{E}[X]\mathbb{E}[Y].
\]

More generally, for \( n \) mutually independent random variables, we have the following

**Theorem 2.** If \( X_1, \ldots, X_n \) are mutually independent random variables, then

\[
\mathbb{E}[X_1X_2\ldots X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2] \ldots \mathbb{E}[X_n].
\]

**Question.** Let \( p_1, \ldots, p_k \) be mutually independent random variables such that each \( p_i \) is a uniformly random prime number between 1 and 20 (1 is not a prime). If \( X = p_1p_2\ldots p_k \), what is \( \mathbb{E}[X] \)?

**Proof.** Since all \( n \) random variables are mutually independent, we know that \( \mathbb{E}[X] = \mathbb{E}[p_1] \mathbb{E}[p_2] \ldots \mathbb{E}[p_k] \). It’s easy to see that for each \( i \in \{1, \ldots, k\} \), \( p_i \) is a uniform random variable with range \( \{2, 3, 5, 7, 11, 13, 17, 19\} \), and so

\[
\mathbb{E}[p_i] = \frac{2 + 3 + 5 + 7 + 11 + 13 + 17 + 19}{8} = 9.625,
\]

and so \( \mathbb{E}[X] = (9.625)^k \).

Often we deal with \( n \) variables that are mutually independent and each variable has the same probability distribution. Such variables are called **Independent Identically Distributed (i.i.d.) random variables.**
3 Covariance of random variables

Given two random variables $X, Y$ defined on a probability space, the covariance between them is defined as follows:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

We will first derive an equivalent expression for covariance that will be easier to interpret:

Lemma 3.

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Proof. Let’s write $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. It’s important to recall that $\mu_X$ and $\mu_Y$ are fixed values and not random variables. Then,

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] = \mathbb{E}[XY] - \mu_X \mathbb{E}[Y] - \mu_Y \mathbb{E}[X] + \mu_X \mu_Y.$$

Here the last step follows from linearity of expectation. Putting back $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$, we get

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Clearly, if $X$ and $Y$ are independent, then based on what we saw in the previous section and based on the alternate expression for covariance that we derived, $\text{Cov}(X, Y) = 0$. Is the other way round true? That is, if $\text{Cov}(X, Y) = 0$, then $X$ and $Y$ are independent? This is not true. Consider $X$ with range $\{-1, 0, 1\}$ and probability mass function $f_X$ defined as follows:

$$f_X(-1) = P(X = -1) = \frac{1}{4},$$

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

$$f_X(1) = P(X = 1) = \frac{1}{4},$$

and let $Y = X^2$. Then, $\mathbb{E}[X] = 0$ (why?), and so $\mathbb{E}[X]\mathbb{E}[Y] = 0$. How about $\mathbb{E}[XY]$? Note that

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = (-1)\frac{1}{4} + (0)\frac{1}{2} + (1)\frac{1}{4} = 0.$$

But are $X$ and $Y$ are independent? No, because $Y$ is literally defined as $X^2$ and so, for example, $P(Y = 1 | X = 0) = 0 \neq P(Y = 1)$.

Nevertheless, covariance can be thought of as a measure of how dependent two random variables are on each other. The farther $\text{Cov}(X, Y)$ is from 0, the more “dependent” $X$ and $Y$ are on each other. In fact, it can be shown that if the covariance is “far enough” from 0 (i.e., it is a large enough positive value, or a small enough negative value) then $Y$ can be written as a linear function of $X$, i.e. $Y = aX + b$ for some constants $a$ and $b$ in $\mathbb{R}$. Conversely, if $\text{Cov}(X, Y) = 0$ then all that means is that $Y$ cannot be written as a linear function of $X$, i.e. $Y$ does not have a linear dependence on $X$ (but as we saw, $Y$ can still be dependent on $X$ in a way other than linear, e.g., $Y = X^2$).