Mathematical Foundations of Computer Science
Lecture Outline
November 8, 2018

Theorem. If $X$ and $Y$ are independent real-valued random variables then

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \text{ and } \mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

The result can be extended to a finite number of random variables.

Note that the converse of the above statement is not true as illustrated by the following example. Let $\Omega = \{a, b, c\}$, with all three outcomes equally likely. Let $X$ and $Y$ be random variables defined as follows: $X(a) = 1, X(b) = 0, X(c) = -1$ and $Y(a) = 0, Y(b) = 1, Y(c) = 0$. Note that $X$ and $Y$ are not independent since

$$\Pr[X = 0 \land Y = 0] = 0, \text{ but } \Pr[X = 0] \cdot \Pr[Y = 0] = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \neq 0.$$ 

Note that for all $\omega \in \Omega$, $X(\omega)Y(\omega) = 0$. Also, $\mathbb{E}[X] = 0$ and $\mathbb{E}[Y] = 1/3$. Thus we have

$$\mathbb{E}[XY] = 0 = \mathbb{E}[X] \mathbb{E}[Y]$$

It is also easy to verify that $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y].$

Example (Chebyshev’s Inequality). Let $X$ be a random variable. Show that for any $a > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}$$

Solution. The inequality that we proved in the earlier homework is called Markov’s inequality. We will use it to prove the above tail bound called Chebyshev’s inequality.

$$\Pr[|X - \mathbb{E}[X]| \geq a] = \Pr[(X - \mathbb{E}[X])^2 \geq a^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} \quad (\text{using Markov’s Inequality}) = \frac{\text{Var}[X]}{a^2}$$

Example. Use Chebyshev’s inequality to bound the probability of obtaining at least $3n/4$ heads in a sequence of $n$ fair coin flips.
Solution. Let $X$ denote the random variable denoting the total number of heads that result in $n$ flips of a fair coin. For $1 \leq i \leq n$, let $X_i$ be a random variable that is 1, if the $i$th flip results in Heads, 0, otherwise. Thus,

$$X = X_1 + X_2 + \cdots + X_n$$

By the linearity of expectation, $\mathbb{E}[X] = n/2$. Since the random variables $X_i$s are independent, we have

$$\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i] = \sum_{i=1}^{n} (1/2 - 1/4) = \frac{n}{4}$$

Using Chebyshev’s inequality, we get

$$\Pr[X \geq 3n/4] = \Pr[X - n/2 \geq n/4]$$

$$= \Pr[X - \mathbb{E}[X] \geq n/4]$$

$$= \frac{1}{2} \cdot \Pr[|X - \mathbb{E}[X]| \geq n/4]$$

$$\leq \frac{1}{2} \cdot \frac{\text{Var}[X]}{n^2/16}$$

$$= \frac{2}{n}$$

Probability Distributions

Tossing a coin is an experiment with exactly two outcomes: heads ("success") with a probability of, say $p$, and tails ("failure") with a probability of $1 - p$. Such an experiment is called a Bernoulli trial. Let $Y$ be a random variable that is 1 if the experiment succeeds and is 0 otherwise. $Y$ is called a Bernoulli or an indicator random variable. For such a variable we have

$$\mathbb{E}[Y] = p \cdot 1 + (1 - p) \cdot 0 = p = \Pr[Y = 1]$$

Thus for a fair coin if we consider heads as "success" then the expected value of the corresponding indicator random variable is 1/2.

A sequence of Bernoulli trials means that the trials are independent and each has a probability $p$ of success. We will study two important distributions that arise from Bernoulli trials: the geometric distribution and the binomial distribution.

The Geometric Distribution

Consider the following question. Suppose we have a biased coin with heads probability $p$ that we flip repeatedly until it lands on heads. What is the distribution of the number of flips? This is an example of a geometric distribution. It arises in situations where we perform a sequence of independent trials until the first success where each trial succeeds with a probability $p$. 
Note that the sample space \( \Omega \) consists of all sequences that end in \( H \) and have exactly one \( H \). That is
\[
\Omega = \{ H, TH, TTH, TTTH, TTTTH, \ldots \}
\]
For any \( \omega \in \Omega \) of length \( i \), \( \Pr[\omega] = (1 - p)^{i-1}p \).

**Definition.** A geometric random variable \( X \) with parameter \( p \) is given by the following distribution for \( i = 1, 2, \ldots \):
\[
\Pr[X = i] = (1 - p)^{i-1}p
\]
We can verify that the geometric random variable admits a valid probability distribution as follows:
\[
\sum_{i=1}^{\infty} (1 - p)^{i-1}p = p \sum_{i=1}^{\infty} (1 - p)^{i-1} = \frac{p}{1 - p} \sum_{i=1}^{\infty} (1 - p)^{i} = \frac{p}{1 - p} \frac{1 - p}{1 - (1 - p)} = 1
\]
Note that to obtain the second-last term we have used the fact that \( \sum_{i=1}^{\infty} c^i = \frac{c}{1 - c}, |c| < 1 \).

Let’s now calculate the expectation of a geometric random variable, \( X \). We can do this in several ways. One way is to use the definition of expectation.
\[
E[X] = \sum_{i=0}^{\infty} i \Pr[X = i]
\]
\[
= \sum_{i=0}^{\infty} i(1 - p)^{i-1}p
\]
\[
= \frac{p}{1 - p} \sum_{i=0}^{\infty} i(1 - p)^{i}
\]
\[
= \left( \frac{p}{1 - p} \right) \left( \frac{1 - p}{(1 - (1 - p))^2} \right) \quad \therefore \sum_{i=0}^{\infty} kx^k = \frac{x}{(1 - x)^2}, \text{ for } |x| < 1.
\]
\[
= \left( \frac{p}{1 - p} \right) \left( \frac{1 - p}{p^2} \right)
\]
\[
= \frac{1}{p}
\]
Another way to compute the expectation is to note that \( X \) is a random variable that takes on non-negative values. From a theorem proved in last class we know that if \( X \) takes on only non-negative values then
\[
E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]
\]
Using this result we can calculate the expectation of the geometric random variable \( X \). For the geometric random variable \( X \) with parameter \( p \),
\[
\Pr[X \geq i] = \sum_{j=i}^{\infty} (1 - p)^{j-1}p = (1 - p)^{i-1}p \sum_{j=0}^{\infty} (1 - p)^{j} = (1 - p)^{i-1}p \times \frac{1}{1 - (1 - p)} = (1 - p)^{i-1}
\]
Therefore
\[ E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-p} \sum_{i=1}^{\infty} (1-p)^i = \frac{1}{1-p} \cdot \frac{1-p}{1-(1-p)} = \frac{1}{p} \]

**Memoryless Property.** For a geometric random variable \( X \) with parameter \( p \) and for \( n > 0 \),
\[ \Pr[X = n + k \mid X > k] = \Pr[X = n] \]

**Conditional Expectation.** The following is the definition of conditional expectation.
\[ E[Y \mid Z = z] = \sum_y y \Pr[Y = y \mid Z = z], \]
where the summation is over all possible values \( y \) that the random variable \( Y \) can assume.

**Example.** For any random variables \( X \) and \( Y \),
\[ E[X] = \sum_y \Pr[Y = y]E[X \mid Y = y] \]

We can also calculate the expectation of a geometric random variable \( X \) using the memoryless property of the geometric random variable. Let \( Y \) be a random variable that is 0, if the first flip results in tails and that is 1, if the first flip is a heads. Using conditional expectation we have
\[ E[X] = \Pr[Y = 0]E[X \mid Y = 0] + \Pr[Y = 1]E[X \mid Y = 1] \]
\[ = (1-p)E[X] + p \cdot 1 \quad \text{(using the memoryless property)} \]
\[ \therefore pE[X] = 1 \]
\[ E[X] = \frac{1}{p} \]