Mathematical Foundations of Computer Science
Lecture Outline
October 31, 2024

Example. In the experiment where we roll one die let X be the random variable denoting the number that appears on the top face. What is $\text{Var}[X]$?

Solution. From the definition of variance, we have

$$\begin{aligned}
\text{Var}[X] &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \\
&= \frac{1}{6}(1+4+9+16+25+36) - \left(\frac{1}{6}(1+2+3+4+5+6)\right)^2 \\
&= \frac{91}{6} - \frac{49}{4} \\
&= \frac{35}{12}
\end{aligned}$$

Example. In the hat-check problem that we did in one of the earlier lectures, what is the variance of the random variable X that denotes the number of people who get their own hat back?

Solution. We can express X as

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

where X_i is the random variable that denotes that is 1 if the ith person receives his/her own hat back and 0 otherwise. We already know from an earlier lecture that $\mathbf{E}[X] = 1$. If $n = 1$ then $\mathbf{E}[X^2] = \mathbf{E}[X_1^2] = \Pr[X_1 = 1] = 1$. In this case, $\text{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = 1 - 1 = 0$, as expected. If $n \geq 2$, $\mathbf{E}[X^2]$ can be calculated as follows.

$$\begin{aligned}
\mathbf{E}[X^2] &= \sum_{i=1}^{n} \mathbf{E}[X_i^2] + 2\sum_{i<j} \mathbf{E}[X_i \cdot X_j] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2] + 2\sum_{i<j} 1 \cdot \Pr[X_i = 1 \cap X_j = 1] \\
&= \sum_{i=1}^{n} \frac{1}{n} + 2\left(\frac{n(n-1)}{2}\right)\left(\frac{1}{n(n-1)}\right) \\
&= n \cdot \frac{1}{n} + 1 \\
&= 2
\end{aligned}$$

$\mathrm{Var}[X]$ is given by

$$\mathrm{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = 2 - 1 = 1$$

Note that like the expectation, the variance is independent of n. This means that it is not likely for many people to get their own hat back even if n is large.

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

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

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

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

Note that $\mathbf{E}[X] = 0$ and $\mathbf{E}[Y] = 1/3$. Also,

$$\mathbf{E}[XY] = \sum_x \sum_y xy \Pr[X = x \cap Y = y] = \sum_x \sum_y xy \Pr[X = x] \Pr[Y = y \mid X = x]$$

Observe that when $x = 0$ or $y = 0$, the summand is clearly 0 and when $x \neq 0$ and $y \neq 0, \Pr[Y = y \mid X = x] = 0$. This is because $\Pr[Y = 1 \mid X = 1] = 0 = \Pr[Y = 1 \mid X = -1]$. Thus we have

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

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

Theorem. If $X_1, X_2, \cdots, X_n$ are random variables that are mutually independent then

$$\mathbf{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbf{E}[X_i]$$

Theorem. If $X_1, X_2, \cdots, X_n$ are random variables that are pairwise independent then

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

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

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

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.

$$\begin{aligned}
\Pr[|X - \mathbf{E}[X]| \geq a] &= \Pr[(X - \mathbf{E}[X])^2 \geq a^2] \\
&\leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} \qquad \text{(using Markov's Inequality)} \\
&= \frac{\mathrm{Var}[X]}{a^2}
\end{aligned}$$

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 ith flip results in Heads, 0, otherwise. Thus,

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

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

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

Using Chebyshev's inequality, we get

$$\begin{aligned}
\Pr[X \geq 3n/4] &= \Pr[X - n/2 \geq n/4] \\
&= \Pr[X - \mathbf{E}[X] \geq n/4] \\
&= \frac{1}{2} \cdot \Pr[|X - \mathbf{E}[X]| \geq n/4] \\
&\leq \frac{1}{2} \cdot \frac{\mathrm{Var}[X]}{n^2/16} \\
&= \frac{2}{n}
\end{aligned}$$

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

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

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 Ω 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:

$$\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{p}{1-p}\sum_{i=1}^{\infty}(1-p)^{i} = \frac{p}{1-p}\cdot\frac{1-p}{1-(1-p)} = 1$$

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.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i=0}^{\infty} i \Pr[X = i] \\
&= \sum_{i=0}^{\infty} i(1-p)^{i-1}p \\
&= \frac{p}{1-p}\sum_{i=0}^{\infty} i(1-p)^{i} \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{(1-(1-p))^2}\right) \qquad \left(\because \sum_{i=0}^{\infty} kx^k = \frac{x}{(1-x)^2},\ \text{for } |x| < 1.\right) \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{p^2}\right) \\
&= \frac{1}{p}
\end{aligned}$$

Another way to compute the expectation is to note that X is a random variable that takes on non-negative integer values. From a theorem proved earlier we know that if X takes on only non-negative integer values then

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$$

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

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-p} \sum_{i=1}^{\infty} (1-p)^{i} = \frac{1}{1-p} \cdot \frac{1-p}{1-(1-p)} = \frac{1}{p}$$