Mathematical Foundations of Computer Science
Lecture Outline
November 4, 2020

Example (Markov's Inequality). Let X be a non-negative random variable. Then for all $a > 0$, prove that

$$\Pr[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$$

Solution. Intuitively, the claim means that if there is too much of probability mass associated with values above $\mathbf{E}[X]$ then the total contribution of such values to $\mathbf{E}[X]$ would be very large. Formally, the proof is as follows.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x \Pr[X = x] \\
&\geq \sum_{x \geq a} x \Pr[X = x] \\
&\geq a \sum_{x \geq a} \Pr[X = x] \\
&= a \Pr[X \geq a] \\
\therefore \Pr[X \geq a] &\leq \frac{\mathbf{E}[X]}{a}
\end{aligned}$$

Example. Suppose we flip a fair coin n times. Using Markov's inequality bound the the probability of obtaining at least $3n/4$ heads.

Solution. Let X be the random variable denoting the total number of heads in n flips of a fair coin. We know that $\mathbf{E}[X] = n/2$. Applying the above inequality we get

$$\Pr[X \geq 3n/4] \leq \frac{\mathbf{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}$$

Example. Suppose we roll a die. Using Markov's inequality bound the probability of obtaining a number greater than or equal to 7.

Solution. Let X be the random variable denoting the result of the roll of a die. We know that $\mathbf{E}[X] = 3.5$. Using the Markov's inequality we get

$$\Pr[X \geq 7] \leq \frac{\mathbf{E}[X]}{7} \leq \frac{1}{2}$$

As this result shows, Markov's inequality gives a loose bound in some cases.

Variance

We are interested in calculating how much a random variable deviates from its mean. This measure is called *variance*. Formally, for a random variable X we are interested in $\mathbf{E}[X - \mathbf{E}[X]]$. By the linearity of expectation we have

$$\mathbf{E}[X - \mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[X] = 0$$

Note that we have used the fact that $\mathbf{E}[X]$ is a constant and hence $\mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X]$. This is not very informative. While calculating the deviations from the mean we do not want the positive and the negative deviations to cancel out each other. This suggests that we should take the absolute value of $X - \mathbf{E}[X]$. But working with absolute values is messy. It turns out that squaring of $X - \mathbf{E}[X]$ is more useful. This leads to the following definition.

Definition. The *variance* of a random variable X is defined as

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

The *standard deviation* of a random variable X is

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}$$

The standard deviation undoes the squaring in the variance. In doing the calculations it does not matter whether we use variance or the standard deviation as we can easily compute one from the other.

We show as follows that the two forms of variance in the definition are equivalent.

$$\begin{aligned}
\mathbf{E}[(X - \mathbf{E}[X])^2] &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X\mathbf{E}[X]] + \mathbf{E}[X]^2 \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \\
&= \mathbf{E}[X^2] - \mathbf{E}[X]^2
\end{aligned}$$

In step 2 we used the linearity of expectation and the fact that $\mathbf{E}[X]$ is a constant.

Example. Consider three random variables X, Y, Z. Their probability mass distribution is as follows.

$$\Pr[X = x] = \begin{cases} \frac{1}{2}, x = -2 \\ \frac{1}{2}, x = 2 \end{cases}$$

$$\Pr[Y = y] = \begin{cases} 0.001, y = -10 \\ 0.998, y = 0 \\ 0.001, y = 10 \end{cases}$$

$$\Pr[Z = z] = \begin{cases} \frac{1}{3}, z = -5 \\ \frac{1}{3}, z = 0 \\ \frac{1}{3}, z = 5 \end{cases}$$

Which of the above random variables is more "spread out"?

Solution. It is easy to see that $\mathbf{E}[X] = \mathbf{E}[Y] = \mathbf{E}[Z] = 0$.

$$\begin{aligned}
\text{Var}[X] &= \mathbf{E}[X^2] \\
&= 0.5 \cdot (-2)^2 + 0.5 \cdot (2)^2 \\
&= 4 \\
\text{Var}[Y] &= \mathbf{E}[Y^2] \\
&= 0.001 \cdot (-10)^2 + 0.998 \cdot 0^2 + 0.001 \cdot (10)^2 \\
&= 0.2 \\
\text{Var}[Z] &= \mathbf{E}[Z^2] \\
&= (1/3) \cdot (-5)^2 + (1/3) \cdot 0^2 + (1/3) \cdot (5)^2 \\
&= 16.67
\end{aligned}$$

Thus Z is the most spread out and Y is the most concentrated.

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

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

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

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

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

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

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

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 $1 \le i \le n$, 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

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