Mathematical Foundations of Computer Science
Lecture Outline
October 24, 2024

Example. Suppose we throw n balls into n bins with the probability of a ball landing in each of the n bins being equal. What is the expected number of empty bins?

Solution. First Approach: The following approach was discussed in class. Let X be the random variable denoting the number of empty bins. For $0 \le i < n$, let X_i be a random variable that is i if exactly i bins are empty and 0, otherwise. We have

$$X = \sum_{i=1}^{n-1} X_i$$

By the linearity of expectation, we have

$$\mathbf{E}[X] = \sum_{i=1}^{n-1} \mathbf{E}[X_i] = \sum_{i=1}^{n-1} \mathbf{E}[X_i] = \sum_{i=1}^{n-1} i \Pr[X_i = i] = \sum_{i=1}^{n-1} i \Pr[X = i]$$

The last equality follows because exactly one of the X_is will be non-zero and if $X_i \neq 0$ then $X = X_i$. Note that we have not made any progress as we are back to using the original definition of expectation to solve the problem.

Second Approach: Let X be the random variable denoting the number of empty bins. Let X_i be a random variable that is 1 if the ith bin is empty and is 0 otherwise. Clearly

$$X = \sum_{i=1}^{n} X_i$$

By linearity of expectation, we have

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\
&= \sum_{i=1}^{n} \Pr[X_i = 1] \\
&= \sum_{i=1}^{n} \left(\frac{n-1}{n}\right)^n \\
&= \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^n
\end{aligned}$$

As $n \to \infty$, $(1 - \frac{1}{n})^n \to \frac{1}{e}$. Hence, for large enough values of n we have

$$\mathbf{E}[X] = \frac{n}{e}$$

Example. The following pseudo-code computes the minimum of n distinct numbers that are stored in an array A. What is the expected number of times that the variable *min* is assigned a value if the array A is a random permutation of the n elements.

```
FindMin(A, n):
min ← A[1]
for i ← 2 to n do
    if (A[i] < min) then
        min ← A[i]
return min
```

Solution. Let X be the random variable denoting the number of times that *min* is assigned a value. We want to calculate $\mathbf{E}[X]$. Let X_i be the random variable that is 1 if *min* is assigned $A[i]$ and 0 otherwise. Clearly,

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Using the linearity of expectation we get

$$\begin{aligned} \mathbf{E}[X] &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\ &= \sum_{i=1}^{n} \Pr[X_i = 1] \qquad (1) \end{aligned}$$

Note that $\Pr[X_i = 1]$ is the probability that $A[i]$ contains the smallest element among the elements $A[1], A[2], \ldots, A[i]$. Since the smallest of these elements is equally likely to be in any of the first i locations, we have $\Pr[X_i = 1] = \frac{1}{i}$. Thus equation (1) becomes

$$\mathbf{E}[X] = \sum_{i=1}^{n} \frac{1}{i} = H(n) \approx \ln n + c$$

where c is a constant less than 1.

Example. Suppose there are k people in a room and n days in a year. On average how many pairs of people share the same birthday?

Solution. Let X be the random variable denoting the number of pairs of people sharing the same birthday. For any two people i and j, let X_{ij} be an indicator random variable that is 1 if i and j have the same birthday and is 0 otherwise. Clearly $X = \sum_{i,j} X_{ij}$. Using the linearity of expectation we get

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i,j} \mathbf{E}[X_{ij}] \\
&= \sum_{i,j} \Pr[X_{ij} = 1] \\
&= \sum_{i,j} \frac{1}{n} \\
&= \frac{\binom{k}{2}}{n} \\
&= \frac{k(k-1)}{2n}
\end{aligned}$$

Assuming $n = 365$, the smallest value of k for which the RHS is at least 1 is 28.

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

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

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