CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Week 11 Discussion

Expectation and Variance

Graphs

Consider a random graph as in the homework, but slightly different: we still have n nodes, but each edge appears independently with probability p. Let X be the number of isolated nodes (nodes with degree 0). What is $\mathbb{E}(X)$?

Solution: For a vertex v, define the indicator random variable X_v such that that $X_v = 1$ if v is isolated and 0 otherwise. It should be clear that $X = \sum_v X_v$. Observe that $\Pr[X_v = 1] = (1-p)^{n-1}$, since in order for v to be isolated, v cannot be connected to all $n-1$ other vertices. By linearity of expectation:

$$\mathbb{E}(X) = \sum_v \mathbb{E}(X_v) = \sum_v (1-p)^{n-1} = n(1-p)^{n-1}.$$

Balls and Bins again!

Let's go back to another one of the homework questions, regarding balls and bins. In this question, you throw n balls into n bins and you lose \$1 for every ball that was already in the bin. In the homework question, you computed how much you expect to lose. We will now show that with probability at least $\frac{1}{2}$ *no* bin has more than $\sqrt{2n}$ balls.

To do this, let's first consider a simpler question which will help. Suppose that the average number of mistakes made on midterm 2 was 10. Prove that at most $\frac{1}{3}$ of the students made 30 or more mistakes. We can now use this idea to prove the above statement.

Solution: We first solve the simpler question. Let n be the number of students, and m_i be the number of mistakes made by student i. We know that $\frac{1}{n}\sum_{i=1}^{n} m_i = 10$. Assume that more than $\frac{1}{3}$ of the students made 30 or more mistakes - assume without loss of generality that these were the first $\frac{1}{3}$ of the students. Then $m_i \geq 30$ for $i = 1$ to $\frac{n}{3}$. The average number of mistakes made by all students is then more than:

$$\frac{1}{n}\left(\sum_{i=1}^{\frac{n}{3}} 30 + \sum_{i=\frac{n}{3}+1}^{n} m_i\right) \geq \frac{1}{n}\left(\frac{30n}{3} + 0\right) = 10$$

For the inequality, we are using two facts: the first $\frac{1}{3}$ of the students made 30 or more mistakes, and the remaining students made 0 or more mistakes. Now we have a contradiction - the average is > 10, but we previously said that the average is equal to 10.

Let's move on to the original question regarding balls and bins. Let X represent the amount of money you lose. In the homework, you showed that $\mathbb{E}(X) = \frac{n-1}{2}$. Assume the statement above is not true. Then with probability $> \frac{1}{2}$, there is a bin with more than $\sqrt{2n}$ balls. In this case, you pay \$1 for the 2nd ball thrown in this bin, \$2 for the 3rd, etc.. So the total amount you pay is at least:

$$\sum_{i=1}^{\sqrt{2n}} i = \frac{\sqrt{2n}(\sqrt{2n}+1)}{2} = n + \frac{\sqrt{2n}}{2}.$$

For the last equality, we used the summation formula $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Since this occurs with probability more than $\frac{1}{2}$, the expected amount you pay must be at least $\frac{1}{2}(n + \frac{\sqrt{2n}}{2})$. But this is more than $\frac{n-1}{2}$, our previously calculated expectation, so we have reached a contradiction.

Variance

In the homework, you were asked how long it would take you to go to your office given that you were taking a very slow bus to work. Let's change this problem in several ways. First of all, the bus that you are taking is now very, very fast- the bus ride to work takes a negligible amount of time. Also, the bus passes your house once every hour (and it is going towards your office). One last detail: you have a friend who sometimes shows up to take you to work. This friend shows up with probability .5, and when he does, he is at your door exactly when you wake up and drops you off at work in a negligible amount of time.

1. How long will it be in expectation between when you wake up and when you arrive at your office?

   Solution: If you take the bus, you expect it to take between 0 and 59 minutes to get to work (this is how long you expect to wait for the bus, since the time it takes on the bus is negligible). If your friend picks you up, you expect it to take 0 minutes to get to work. Since each of these possibilities occurs with probability .5, you expect it to take $\frac{1}{2} \times 0 + \frac{1}{2} \times \frac{59}{2}$ minutes to get to work. Note that the mean for the bus travel time is different than it was in the homework; this is because we have discretized the time into one minute intervals.

   Let's formalize this intuition. Let X be the random variable which represents how long it takes you to get to work. With probability $\frac{1}{2}$, your friend picks you up, so X is 0. With probability $\frac{1}{2}$, you take the bus. What is X now? In this case, X takes on each value between 0 and 59 with probability $\frac{1}{60}$. Then:

   $$\mathbb{E}(X) = \frac{1}{2} \times 0 + \frac{1}{2} \times \left(\frac{1}{60}\sum_{i=1}^{59} i\right) = \frac{59}{4}.$$

   For the last equality, we used the summation formula $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

2. What is the variance?

   Solution: Now we need to compute $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Let's compute the first term using the same idea as above:

   $$\mathbb{E}(X^2) = \frac{1}{2} \times 0^2 + \frac{1}{2} \times \left(\frac{1}{60}\sum_{i=1}^{59} i^2\right) = \frac{7021}{12}.$$

   For the last equality, we used the summation formula $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. Then the variance is:

   $$\mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{7021}{12} - \left(\frac{59}{4}\right)^2 = \frac{17641}{48}.$$

3. Now let's say your friend becomes more or less reliable- he shows up with probability p. How does the variance change?

Solution: We can replace $\frac{1}{2}$ with $1-p$, since this is now the probability that you take the bus:

$$\mathbb{E}(X^2) = p \times 0^2 + (1-p) \times (\frac{1}{60}\sum_{i=1}^{59} i^2) = \frac{7021(1-p)}{6}.$$

and the variance is:

$$\mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{7021(1-p)}{6} - (\frac{59(1-p)}{2})^2.$$

When $p = 1$, the variance is 0 and the variance increases as p decreases. This makes sense; as your friend becomes less reliable, the variance increases.