CS 70 Discrete Mathematics and Probability Theory
Fall 2012 Vazirani **Homework 12**

1 Variance

(a) Give a simple example to show that $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ is not necessarily true when X and Y are not independent.

(b) Prove that for any real number c and for any r.v. X, we have that $\mathbf{Var}[cX] = c^2 \mathbf{Var}[X]$.

(c) Give a simple example to show that $\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$ is not necessarily true when X and Y are not independent.

(d) Suppose X is a random variable and $\mathbf{E}[X]$ is finite. Are we guaranteed that $\mathbf{Var}[X]$ is also necessarily finite?

(e) Suppose A, B are events, and $\mathbf{1}_A$, $\mathbf{1}_B$ the indicator variables for these events. Show that $\mathbf{E}[\mathbf{1}_A \cdot \mathbf{1}_B] = \mathbf{Pr}[A \cap B]$.

–Solution–

(a) Toss a fair coin. Let $X = 1$ if heads, $X = 0$ if tails. Let $Y = 0$ if heads, and $Y = 1$ if tails. Note that $\mathbf{E}[XY] = \mathbf{E}[0] = 0$, while $\mathbf{E}[X] = \mathbf{E}[Y] = 1/2$.

(b)

$$
\begin{aligned}
\mathbf{Var}[cX] &= \mathbf{E}\left[(cX - \mathbf{E}[cX])^2\right] \\
&= \mathbf{E}\left[(cX - c\mathbf{E}[X])^2\right] \\
&= \mathbf{E}\left[c^2(X - \mathbf{E}[X])^2\right] \\
&= c^2\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\
&= c^2\mathbf{Var}[X]
\end{aligned}
$$

(c) Define X and Y as in (a). Note that $Y = 1 - X$. Therefore, $\mathbf{Var}[X+Y] = \mathbf{Var}[1] = 0$. $\mathbf{Var}[X] = \mathbf{E}\left[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2\right] = 1/2 - 1/4 = 1/4$. $\mathbf{E}[Y] = 1/4$ by the same reasoning. Therefore, $\mathbf{Var}[X+Y] \neq \mathbf{Var}[X] + \mathbf{Var}[Y]$.

(d) Let $\mu = \mathbf{E}[X]$ be finite. $\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mu^2$. The question becomes, when does $\mathbf{E}\left[X^2\right]$ not converge when $\mathbf{E}[X]$ does?
Consider the following cooked example. Let $X = \sqrt{2}^k$ with probability $\frac{1}{2}^k$ for $k \in 1, 2, \ldots$. Note that $\sum_{i=1}^{\infty}(1/2)^i \to \frac{1}{1-1/2} - 1 = 1$, and therefore describes a probability distribution.

$$
\begin{aligned}
\mathbf{E}[X] &= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\sqrt{2}^i}{2} \\
&\to \frac{1}{1-\sqrt{2}/2} - 1
\end{aligned}
$$

This result is finite, since $\sqrt{2}/2 < 1$. Consider, however, $\mathbf{E}\left[X^2\right]$:

$$
\begin{aligned}
\mathbf{E}\left[X^2\right] &= \lim_{n\to\infty} \sum_{i=1}^{n} 2^i(1/2^i) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} 1 \to \infty
\end{aligned}
$$

Thus, $\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2 \to \infty$.

(e) The random variable $Z = \mathbf{1}_A \mathbf{1}_B = 1$ in the event $A \cap B$ and 0 otherwise. Therefore $\mathbf{E}[Z] = \mathbf{Pr}[A \cap B]$.

2 Random bit strings

Let S be a random bit string of length n. For example $S = 101111001001$ is a bit string of length 12, and we say that it has a *run* of 4 1's starting in position 3.

(a) For a given position j in S, what is the probability that it is a starting point of a run of at least m 1's?

(b) What is the expected number of positions j at which runs of at least m 1's start?

(c) Use Markov's inequality to show that the probability that there exists a run of at least $5\lceil \log n \rceil$ 1's is at most $\frac{1}{n^4}$.

(d) We now consider runs of alternating 1-0 that start with 1 (e.g., 101010101). What is the expected number of places j at which alternating runs (beginning with 1) of at least m bits start?

–Solution–

(a) Let A_j be the event that a run of length m starts at position j:

$$\mathbf{Pr}\left[A_j\right] = \begin{cases} 2^{-m} & j \leq n - m + 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) Let X be the number of positions at which a run of at least m 1's start. By linearity of expectation,

$$\begin{aligned} \mathbf{E}\left[X\right] &= \mathbf{E}\left[\sum_{j=1}^{n} \mathbf{1}_{A_j}\right] \\ &= \sum_{j=1}^{n} \mathbf{E}\left[\mathbf{1}_{A_j}\right] \\ &= \sum_{j=1}^{n} \mathbf{Pr}\left[A_j\right] \\ &= \frac{n - m + 1}{2^m} \end{aligned}$$

(c) In this solution, we use $\log_2 n$. However, note that $\log_b n = \frac{\log_2 n}{\log_2 b}$, for an arbitrary base b.

Using Markov's Inequality,

$$\begin{aligned}
\mathbf{Pr}[X \geq 1] &\leq \mathbf{E}[X] \\
&\leq \frac{n - 5\lceil \log n \rceil}{2^{5\lceil \log n \rceil}} \\
&\leq \frac{n}{2^{5 \log n}} \\
&\leq \frac{n}{n^5} \\
&\leq \frac{1}{n^4}
\end{aligned}$$

(d) The expectation is still $\frac{n-m}{2^m}$.
Let B_j be the event that an alternating run of length m starts at position j.
The alternating run continues until we hit a 1 after a 1 or a 0 after a 0.
As before, this run of length m starting from position j has probability $\mathbf{Pr}[B_j] = 2^{-m}$ if $j \leq n - m + 1$ and 0 otherwise.
Using indicators and linearity of expectation leads us to the result.

3 Chebyshev Inequality

A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course so that you can devote all of your time to CS70. At the first lecture, the professor announces that grades will depend only a midterm and a final. The midterm will consist of three questions, each worth 10 points, and the final will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points.

However, speaking with the professor in office hours you hear some very disturbing news. He tells you that to save time he will be grading as follows. For each student's midterm, he'll choose a number randomly from a distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. He'll mark each of the three questions with that score. To grade the final, he'll again choose a random number from the same distribution, independent of the first number, and will mark all four questions with that score.

If you take the class, what will the mean and variance of your total class score be? Can you conclude that you have less than a 5% chance of getting an A? Why?

–Solution–

Let X be the score given to a problem on the midterm, and Y be the score given to a problem on the final. Let $Z = 3X + 4Y$ be the total score at the end of the class.
$\mathbf{E}[Z] = 3\mathbf{E}[X] + 4\mathbf{E}[Y] = 35$
$\mathbf{Var}[Z] = 3^2\mathbf{Var}[X] + 4^2\mathbf{Var}[Y] = 25$
The probability of getting an A is:

$$\begin{aligned}
\mathbf{Pr}[Z \geq 60] &= \mathbf{Pr}[(Z-35) \geq 25] \\
&\leq \mathbf{Pr}[|Z-35| \geq 25] \\
&\leq \frac{\mathbf{Var}[Z]}{25^2} \text{ by Chebyshev} \\
&\leq \frac{1}{25}
\end{aligned}$$

Thus, the probability of getting an A is less than 4%.

4 Poisson Distribution

A textbook has on average one misprint per page. What is the chance that you see exactly 4 misprints on page 1? What is the chance that you see exactly 4 misprints on some page in the textbook of 250 pages?

–Solution–

Model the misprints on a page as a poisson random variable with parameter $\lambda = 1$. Each page is treated independently.

$$\begin{aligned} \mathbf{Pr}[X = k] &= \frac{\lambda^k e^{-\lambda}}{k!} \\ \mathbf{Pr}[X = 4] &= \frac{1}{4! \cdot e} \end{aligned}$$

The probability of a single page not having exactly 4 errors is:

$$1 - \frac{1}{4! \cdot e}$$

Treating each page independently, the probability of at least one page having exactly 4 errors is:

$$1 - \left(1 - \frac{1}{4! \cdot e}\right)^{250} \approx 0.979$$

5 Geometric Distribution

James Bond is imprisoned in a cell from which there are three possible ways to escape: an air-conditioning duct, a sewer pipe and the door (which is unlocked). The air-conditioning duct leads him on a one-hour trip whereupon he falls through a trap door onto his head, much to the amusement of his captors. The sewer pipe is similar but takes two hours to traverse. Each fall produces temporary amnesia and he is returned to the cell immediately after each fall. Assume that he always immediately chooses one of the three exits from the cell with probability $\frac{1}{3}$. On the average, how long does it take before he realizes that the door is unlocked and escapes?

Hint: If you are doing complicated calculations you're taking the wrong approach.

–Solution–

Let A_k be the event where Bond failed in his first $k-1$ attempts and selected the AC duct or sewer in the k-th attempt, and let B_k be the event that Bond failed in his first $k-1$ attempts and selected the sewer in the k-attempt. The time the k-th attempt takes is $\mathbf{1}_{A_k} + \mathbf{1}_{B_k}$, because when Bond selects the sewer both indicators are 1 and if Bond selects the AC duct then only the first indicator is 1. Therefore,

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}\left[\sum_{k=1}^{\infty}(\mathbf{1}_{A_k} + \mathbf{1}_{B_k})\right] \\
&= \sum_{k=1}^{\infty}(\mathbf{E}[\mathbf{1}_{A_k}] + \mathbf{E}[\mathbf{1}_{B_k}]) \\
&= \sum_{k=1}^{\infty}(\mathbf{Pr}[A_k] + \mathbf{Pr}[B_k]) \\
&= \sum_{k=1}^{\infty}\left(\frac{1}{3}\cdot\left(\frac{2}{3}\right)^{k-1} + \frac{2}{3}\cdot\left(\frac{2}{3}\right)^{k-1}\right) \\
&= \sum_{k=1}^{\infty}\left(\frac{2}{3}\right)^{k-1} \\
&= \frac{1}{1-\frac{2}{3}} \\
&= 3
\end{aligned}$$

6 Extra Credit

Consider the following game: The dealer shuffles a regular deck of 52 cards and successively turns over one card at a time. After any card, you are allowed to say "stop". If the next card is red, you win, and if it is black, you lose. You must say stop at some point during the game.

What is your optimal strategy (i.e. when should you say "stop") and what is your probability of winning under this strategy?

–Solution–

Consider some arbitrary magical strategy that tells you when to say "stop." The probability that the next card is red is the same as the probability that the last card is red - all the permutations of the remaining cards are equally likely. Regardless of your strategy, whatever your odds are for the next card being red are exactly the same as the odds of the last card being red. Since your strategy is then equivalent to waiting until the very last card, all strategies have, on average, an equal chance at winning, which is 1/2.
Therefore, all strategies are equally good and will win with probability 1/2.