CS 70 Discrete Mathematics and Probability Theory
Summer 2021

DIS 07A

1 Probabilistic Bounds

A random variable X has variance $\text{Var}(X) = 9$ and expectation $\mathbb{E}[X] = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

(a) $\mathbb{E}[X^2] = 13$.

(b) $\mathbb{P}[X = 2] > 0$.

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

(d) $\mathbb{P}[X \leq 1] \leq 8/9$.

(e) $\mathbb{P}[X \geq 6] \leq 9/16$.

Solution:

(a) TRUE. Since $9 = \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2^2$, we have $\mathbb{E}[X^2] = 9 + 4 = 13$.

(b) FALSE. It is not necessary for a random variable to be able to take on its mean as a value. Construct a random variable X that satisfies the conditions in the question but does not take on the value 2. A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 1/2$, and $a \neq b$. The expectation must be 2, so we have $a/2 + b/2 = 2$. The variance is 9, so $\mathbb{E}[X^2] = 13$ (from Part (a)) and $a^2/2 + b^2/2 = 13$. Solving for a and b, we get $\mathbb{P}[X = -1] = \mathbb{P}[X = 5] = 1/2$ as a counterexample.

(c) FALSE. The median of a random variable is not necessarily the mean, unless it is symmetric. Construct a random variable X that satisfies the conditions in the question but does not have an equal chance of being less than or greater than 2. A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = p, \mathbb{P}[X = b] = 1 - p$. Here, we use the same approach as part (b) except with a generic p, since we want $p \neq 1/2$. The expectation must be 2, so we have $pa + (1 - p)b = 2$. The variance is 9, so $\mathbb{E}[X^2] = 13$ and $pa^2 + (1 - p)b^2 = 13$. Solving for a and b, we find the relation $b = 2 \pm 3/\sqrt{x}$, where $x = (1 - p)/p$. Then, we can find an example by plugging in values for x so that $a, b \leq 10$ and $p \neq 1/2$. One such counterexample is $\mathbb{P}[X = -7] = 1/10, \mathbb{P}[X = 3] = 9/10$.

(d) TRUE. Let $Y = 10 - X$. Since X is never exceeds 10, Y is a non-negative random variable. By Markov's inequality,

$$\mathbb{P}[10 - X \geq a] = \mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a} = \frac{\mathbb{E}[10 - X]}{a} = \frac{8}{a}.$$

Setting $a = 9$, we get $\mathbb{P}[X \le 1] = \mathbb{P}[10 - X \ge 9] \le 8/9$.

(e) TRUE. Chebyshev's inequality says $\mathbb{P}[|X - \mathbb{E}[X]| \ge a] \le \operatorname{Var}(X)/a^2$. If we set $a = 4$, we have

$$\mathbb{P}[|X-2| \ge 4] \le \frac{9}{16}.$$

Now we observe that $\mathbb{P}[X \ge 6] \le \mathbb{P}[|X-2| \ge 4]$, because the event $X \ge 6$ is a subset of the event $|X-2| \ge 4$.

2 Easy As

A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course next semester to give yourself a well-deserved break after working hard in CS 70. At the first lecture, the professor announces that grades will depend only on two homework assignments. Homework 1 will consist of three questions, each worth 10 points, and Homework 2 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, in the spirit of the class, the GSIs are very lazy, and to save time the grading will be done as follows. For each student's Homework 1, the GSIs will choose an integer randomly from a distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. They'll mark each of the three questions with that score. To grade Homework 2, they'll again choose a random number from the same distribution, independently 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? Use Chebyshev's inequality to conclude that you have less than a 5% chance of getting an A when the grades are randomly chosen this way.

Solution:

Let X be the total number of points you receive in the class. Then $X = X_1 + X_2$ where X_1 is the number points received on Homework 1 and X_2 is the number of points received on Homework 2. Your Homework 1 score is generated as $X_1 = 3Y_1$, where the r.v. Y_1 represents the integer that the GSI chose when grading it. Similarly, $X_2 = 4Y_2$. The problem statement tells us that Y_1 and Y_2 are independent, both with mean 5 and variance 1, so $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = 5$ and $\operatorname{Var}(Y_1) = \operatorname{Var}(Y_2) = 1$. Thus,

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3\,\mathbb{E}[Y_1] + 4\,\mathbb{E}[Y_2] = 35,$$
$$\operatorname{Var}(X) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) = 9\operatorname{Var}(Y_1) + 16\operatorname{Var}(Y_2) = 25.$$

Using Chebyshev's Inequality, we get

$$\mathbb{P}[X \ge 60] \le \mathbb{P}[|X-35| \ge 25] \le \frac{\operatorname{Var}(X)}{25^2} = \frac{1}{25}.$$

Unfortunately, any student will have at most a 4% chance of getting an A.

Note that although we calculated a bound for $\mathbb{P}[|X-35| \geq 25]$, which is the probability that you will get 60 or above or 10 or below, we cannot divide by 2 to refine our bound unless the distribution is symmetric about its mean. In this case, the distribution is not symmetric.

3 Tightest Bounds

A random variable X has expectation $\mathbb{E}[X] = 12$ and variance $\text{Var}[X] = 4$.

(a) For $I \sim \text{Bernoulli}(p)$ and $Y = aI + b$, come up with values for a, b, and p such that $\mathbb{E}[Y] = 12$ and $\text{Var}[Y] = 4$. What are the possible values of your Y?

(b) Find the tightest bounds you can for $\mathbb{P}[4 \leq X \leq 20]$.

(c) Find the tightest bounds you can for $\mathbb{P}[9 \leq X \leq 20]$.

(d) Find the tightest bounds you can for $\mathbb{P}[X \geq 16]$.

(e) Find the tightest bounds you can for $\mathbb{P}[X^2 \geq 225]$.

Solution: For parts (b) through (e), draw a distribution, label it with $\mathbb{E}[X] = 12$, and shade the regions for which we're bounding the probability. Then compare it to the regions that we can bound using Chebyshev's.

(a) To satisfy the given conditions, we must have

$$\mathbb{E}[aI+b] = ap+b = 12 \qquad \text{Var}[aI+b] = a^2p(1-p) = 4.$$

Solving for a and b, we get that

$$a = \frac{2}{\sqrt{p(1-p)}} \qquad b = 12 - ap.$$

We can plug in any value of p. Some easy options are $p = \frac{1}{2}$, which yields $4I + 10$ and $p = \frac{1}{5}$, which yields $5I + 11$. The possible values in each case are $\{10, 14\}$ and $\{11, 16\}$, respectively. For the rest of the problem, we will use $Y = 4I + 10$ for $I \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

(b) This is a symmetric interval about $\mathbb{E}[X]$:

$$\begin{aligned}\mathbb{P}[4 \leq X \leq 20] &= \mathbb{P}[-8 \leq X - 12 \leq 8] \\ &= \mathbb{P}[|X-12| \leq 8] \\ &= 1 - \mathbb{P}[|X-12| \geq 8].\end{aligned}$$

Apply Chebyshev's inequality:

$$\begin{aligned}
\mathbb{P}[|X-12| \geq 8] &\leq \frac{4}{8^2} \\
-\mathbb{P}[|X-12| \geq 8] &\geq -\frac{4}{8^2} \\
1-\mathbb{P}[|X-12| \geq 8] &\geq 1-\frac{4}{8^2} \\
\mathbb{P}[4 \leq X \leq 20] &\geq 1-\frac{4}{8^2}.
\end{aligned}$$

The tightest upper bound is 1, since if $X = Y$, then $\mathbb{P}[4 \leq X \leq 20] = 1$.

Our tightest bounds are $1-\frac{4}{8^2} \leq \mathbb{P}[4 \leq X \leq 20] \leq 1$.

(c) This region is not a symmetric interval about $\mathbb{E}[X]$, but it is bounded below by one:

$$\begin{aligned}
\mathbb{P}[9 \leq X \leq 20] &\geq \mathbb{P}[9 \leq X \leq 15] \\
&= \mathbb{P}[-3 \leq X-12 \leq 3] \\
&= 1-\mathbb{P}[|X-12| \geq 3].
\end{aligned}$$

Applying Chebyshev's yields

$$\mathbb{P}[9 \leq X \leq 20] \geq 1-\frac{4}{3^2}.$$

The tightest upper bound is again 1, which happens if $X = Y$.

Our tightest bounds are $1-\frac{4}{3^2} \leq \mathbb{P}[9 \leq X \leq 20] \leq 1$.

(d) We cannot use Markov's since X is not necessarily nonnegative, so we can only use Chebyshev.

$$\begin{aligned}
\mathbb{P}[X \geq 16] &\leq \mathbb{P}[X \leq 8 \cup X \geq 16] \\
&= \mathbb{P}[|X-12| \geq 4] \\
&\leq \frac{4}{4^2}
\end{aligned}$$

The tightest lower bound is 0, which happens if $X = Y$.

Our tightest bounds are $0 \leq \mathbb{P}[X \geq 16] \leq \frac{4}{4^2}$.

(e) Note that X^2 is a nonnegative random variable, so check Markov's. First, $E[X^2] = \text{Var}[X] + \mathbb{E}[X]^2 = 4+144 = 148$. By Markov,

$$\mathbb{P}[X^2 \geq 225] \leq \frac{148}{225}.$$

We compare with Chebyshev:

$$\begin{aligned}
\mathbb{P}[X^2 \geq 225] &= \mathbb{P}[X \leq -15 \cup X \geq 15] \\
&\leq \mathbb{P}[X \leq 9 \cup X \geq 15] \\
&= \mathbb{P}[|X-12| \geq 3] \\
&\leq \frac{4}{3^2}.
\end{aligned}$$

The tightest upper bound is $\frac{4}{3^2}$ by Chebyshev. The tightest lower bound is 0, which happens if $X = Y$.

Our tightest bounds are $0 \leq \mathbb{P}[X^2 \geq 225] \leq \frac{4}{3^2}$.