Solutions to In-Class Problems — Week 13, Mon

Problem 1. Consider the following two gambling games.

Game A: We win \$2 with probability $2/3$ and lose \$1 with probability $1/3$.

Game B: We win \$1002 with probability $2/3$ and lose \$2001 with probability $1/3$.

(a) What is the expected win in each case?

Solution. From the Notes: Let random variables A and B be the payoffs for the two games. For example, A is 2 with probability 2/3 and -1 with probability 1/3. We can compute the expected payoff for each game as follows:

$$\begin{aligned} \mathrm{E}[A] &= 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1, \\ \mathrm{E}[B] &= 1002 \cdot \frac{2}{3} + (-2001) \cdot \frac{1}{3} = 1. \end{aligned}$$

We have the same probability, 2/3, of winning each game and the same expected return for each game. ■

(b) What is the variance in each case?

Solution. The variances of the two games are very different. We can compute the $\mathrm{Var}[A]$ by working "from the inside out" as follows:

$$\begin{aligned} A - \mathrm{E}[A] &= \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ -2 & \text{with probability } \frac{1}{3} \end{cases} \\ (A - \mathrm{E}[A])^2 &= \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ 4 & \text{with probability } \frac{1}{3} \end{cases} \\ \mathrm{E}\left[(A - \mathrm{E}[A])^2\right] &= 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3} \\ \mathrm{Var}[A] &= 2. \end{aligned}$$

Copyright © 2002, Dr. Radhika Nagpal.

Similarly, we have for $\text{Var}[B]$:

$$\begin{aligned}
B - \text{E}[B] &= \begin{cases} 1001 & \text{with probability } \frac{2}{3} \\ -2002 & \text{with probability } \frac{1}{3} \end{cases} \\
(B - \text{E}[B])^2 &= \begin{cases} 1,002,001 & \text{with probability } \frac{2}{3} \\ 4,008,004 & \text{with probability } \frac{1}{3} \end{cases} \\
\text{E}\left[(B - \text{E}[B])^2\right] &= 1,002,001 \cdot \frac{2}{3} + 4,008,004 \cdot \frac{1}{3} \\
\text{Var}[B] &= 2,004,002.
\end{aligned}$$

The variance of Game A is 2 and the variance of Game B is more than two million! Intuitively, this means that the payoff in Game A is usually close to the expected value of $1, but the payoff in Game B can deviate very far from this expected value.

High variance is often associated with high risk. For example, in ten rounds of Game A, we expect to make $10, but could conceivably lose $10 instead. On the other hand, in ten rounds of game B, we also expect to make $10, but could actually lose more than $20,000! ■

Problem 2. Suppose you have learned that the average graduating MIT student's total number of credits is 200.

(a) Knowing only this average, use Markov's inequality to find a best possible upper bound for the fraction of MIT students graduating with at least 235 credits. [1]

Solution. Let X be a random variable with a distribution equal to that of the graduating MIT students' credit count. We are given that $\text{E}[X] = 200$. By Markov's inequality:

$$\Pr\{X \geq 235\} \leq \frac{\text{E}[X]}{235} = \frac{200}{235} \approx 0.85$$

■

(b) Demonstrate that this is a best possible bound by giving a distribution for which this bound holds with equality.

Solution. The bound is attained with equality at the two-point distribution which has non-zero values only at 0 and 235, i.e.

$$\begin{aligned}
\Pr\{X = 235\} &= 200/235 \\
\Pr\{X = 0\} &= 35/235 \\
\Pr\{X = x\} &= 0 \text{ for all other } x
\end{aligned}$$

[1] Ignore the fact that there are practical limits to the amount of time a student can stay at MIT and remain sane; That is, assume that there is no bound on the number of credits a student may earn.

You might wonder how we got to this particular distribution, since the space of all possible probability functions is very large `:-)`. This function could be derived intuitively as follows.

Consider any distribution with the given mean, $\mathrm{E}[X] = 200$. We can 'shift' this distribution around, like sand on a see-saw, subject to the constraint that it remains balanced around the mean. We want to maximize the portion of the distribution that is to the right of the point $x = 235$.

How can we do this? Look at the portion of the distribution to the right of $x = 235$. We need to minimize this portion's contribution to the mean (so we can then maximize its volume). For this, we must move it as close to the mean as possible, i.e. we pile it all up at the point $x = 235$.

Similarly, we need to maximize the contribution of the distribution to the left of the mean. To do this, we move it *away* from the mean, i.e. to the left as far as possible. Since X is non-negative, this means that it all piles up at the point $x = 0$.

Mathematically,

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[X \mid X \geq c] \Pr\{X \geq c\} + \mathrm{E}[X \mid X < c] \Pr\{X < c\} \\ &\geq c \cdot \Pr\{X \geq c\} + 0 \cdot \Pr\{X < c\} \\ &= c \cdot \Pr\{X \geq c\} \end{aligned}$$

Equality holding iff $\mathrm{E}[X \mid X \geq c] = c$ AND $\mathrm{E}[X \mid X < c] = 0$, i.e. $x \geq c \Rightarrow x = c$ and $x < c \Rightarrow x = 0$. ■

(c) Suppose you are now told that no student can graduate with fewer than 170 units. How does this allow you to improve your previous bound? As before, show that this is the best possible bound.

Solution. We can now apply Markov's inequality to the nonnegative variable $Y = X - 170$, with expectation $\mathrm{E}[Y] = \mathrm{E}[X - 170] = \mathrm{E}[X] - 170 = 30$. So,

$$\Pr\{X \geq 235\} = \Pr\{X - 170 \geq 235 - 170\} = \Pr\{Y \geq 65\}$$

Therefore:

$$\begin{aligned} \Pr\{X \geq 235\} &= \Pr\{Y \geq 65\} \\ &\leq \frac{\mathrm{E}[Y]}{64} \\ &\leq \frac{30}{65} \approx 0.46 \end{aligned}$$

As above, we achieve an optimum (equality in the bound) when our distribution consists of two spikes: one at $(x - 170) = c - 170$, i.e. $x = 235$, and one at $(x - 170) = 0$, i.e. $x = 170$.

$$\begin{aligned} \Pr\{X = 235\} &= (200 - 170)/(235 - 170) = 30/65 \\ \Pr\{X = 170\} &= 35/65 \\ \Pr\{X = x\} &= 0 \text{ for all other } x \end{aligned}$$

■

(d) Now suppose you *further* learn that the standard deviation of the total credits per graduating student is 7. Give a best possible bound on the fraction of students who can graduate with at least 235 credits.

Solution. Use the Chebyshev inequality to bound the probability. The variance of X is the square of the standard deviation, or 49. The variance of Y is the same as that of X, by the linearity of variance. That is, $\text{Var}[Y] = \text{Var}[X - 170] = \text{Var}[X] - \text{Var}[170] = 49 - 0$. (The variance of a constant is 0).

$$
\begin{aligned}
\Pr\{X \geq 235\} &= \Pr\{Y \geq 65\} \\
&= \Pr\{Y - \text{E}[Y] \geq 65 - \text{E}[Y]\} \\
&= \Pr\{Y - 30 \geq 35\} \\
&\leq \Pr\{|Y - 30| \geq 35\} \\
&\leq \frac{\text{Var}[Y]}{35^2} \\
&\leq \frac{49}{1225} = \frac{1}{25}
\end{aligned}
$$

This is a much better bound than before!

■

Problem 3. In this problem we will derive Chebyshev's Theorem from a corollary of Markov's Theorem.

(a) Explain why the following corollary of Markov's theorem holds:

Corollary. *For any random variable R, any positive integer k, and any $x > 0$,*

$$\Pr\{|R| \geq x\} \leq \frac{\text{E}\left[|R|^k\right]}{x^k}.$$

Solution. This can be seen by letting the random variable in Markov's theorem be $|R|^k$, which is nonnegative for any random variable R and positive integer k. Notice as well that since $|R| \geq x$ iff $|R|^k \geq x^k$, the probabilities of the two events are the same. Combining these facts we get,

$$\Pr\{|R| \geq x\} = \Pr\left\{|R|^k \geq x^k\right\} \leq \frac{\text{E}\left[|R|^k\right]}{x^k}.$$

Note: Even though Markov's theorem applies only to nonnegative random variables, this corollary applies to *all* random variables. This implies that Chebyshev's theorem, derived in the next part, also applies to all random variables. ■

(b) Use the above corollary to prove the following:

Theorem (Chebyshev). *Let R be a random variable, and let x be a positive real number. Then*

$$\Pr\{|R - \mathrm{E}[R]| \geq x\} \leq \frac{\mathrm{Var}[R]}{x^2}.$$

(Hint: Consider the case where $k = 2$).

Solution. The special case of this corollary when $k = 2$ can be applied to bound the random variable, $|R - \mathrm{E}[R]|$, that measures R's deviation from its mean. Namely

$$\Pr\{|R - \mathrm{E}[R]| \geq x\} = \Pr\left\{(R - \mathrm{E}[R])^2 \geq x^2\right\} \leq \frac{\mathrm{E}\left[(R - \mathrm{E}[R])^2\right]}{x^2},$$

where the inequality follows from the corollary applied to the random variable, $|R - \mathrm{E}[R]|$. So we can bound the probability that the random variable R deviates from its mean by more than x by an expression decreasing as $1/x^2$ multiplied by the constant $\mathrm{E}\left[(R - \mathrm{E}[R])^2\right]$. This constant is the *variance of* R. Hence,

$$\Pr\{|R - \mathrm{E}[R]| \geq x\} \leq \frac{\mathrm{Var}[R]}{x^2}.$$

■

Problem 4. Prove that the following two formulas for calculating variance are equivalent:

$$\mathrm{Var}[R] ::= \mathrm{E}\left[(R - \mathrm{E}[R])^2\right].$$
$$\mathrm{Var}[R] = \mathrm{E}\left[R^2\right] - \mathrm{E}^2[R],$$

Solution. From the Notes: Remember that $\mathrm{E}\left[R^2\right]$ is generally not equal to $\mathrm{E}^2[R]$. The expected value of a product is the product of the expected values only for independent variables, and R is not independent of itself unless it is constant.

Proof. Write $\mu = \mathrm{E}[R]$. Then

$$\begin{aligned}
\mathrm{Var}[R] &= \mathrm{E}\left[(R - \mu)^2\right] \\
&= \mathrm{E}\left[R^2 - 2R \cdot \mu + \mu^2\right] \\
&= \mathrm{E}\left[R^2\right] - \mathrm{E}[2R \cdot \mu] + \mathrm{E}\left[\mu^2\right] \\
&= \mathrm{E}\left[R^2\right] - 2\,\mathrm{E}[R] \cdot \mu + \mu^2 \\
&= \mathrm{E}\left[R^2\right] - 2\mathrm{E}^2[R] + \mathrm{E}^2[R] \quad \text{(definition of } \mu\text{)} \\
&= \mathrm{E}\left[R^2\right] - \mathrm{E}^2[R].
\end{aligned}$$

The first step uses the definition of variance. In the second step, we multiply out the squared term. The third step uses linearity of expectation. There are two transformations on the fourth line. In the second term, we pull the constant 2μ out of the expectation. In the third term, we use the fact that the expectation of a constant, namely μ^2, is that constant. The final step is simplification. □

■