# Recitation 9

Expectation and Variance

## Expected Value

Intuitively, the expected value is the weighted average of values, kind of a mass center of the probability distribution.

More formally, the *expected value* of a random variable is denoted $\mathbb{E}[X]$ and is defined as

$$\mathbb{E}[X] = \sum_{s \in S} X(s) \Pr(s) = \sum_{r \in X(S)} r \Pr(X = r).$$

We define the *conditional expected values* as follows: Given that event $E$ has occurred, the expectation of random variable $X$ is

$$\mathbb{E}[X \mid E] = \sum_{r \in X(S)} r \Pr(X = r \mid E). \tag{1}$$

Moreover, the *linearity of expectation* can be very useful in calculating expected value: Given that $Z$, $X$, $Y$ are three random variables defined on a sample space $S$ and $a$ and $b$ are two real numbers such that $Z = aX + bY$, we know that $\mathbb{E}[Z] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ must be true.

Let's practice this through a task:

**Task 1**

Tim and Joe are playing a game. They flip a fair coin 3 times. When the coin is heads, Joe pays $1 to Tim; and when the coin is tails, Tim pays $1 to Joe.

a. Let $G_i$ be a random variable representing what Tim gains on the $i$-th round. For instance, $G_3 = -1$ if the coin is tails.

   What is the expected value of $G_i$?

   | |
   |---|
   | Since $\Pr(G_i = 1) = \frac{1}{2}$ and $\Pr(G_i = -1) = \frac{1}{2}$, $\mathbb{E}[G_i] = \frac{1}{2} - \frac{1}{2} = 0$. |

b. Let $G$ be a random variable that represents Tim's *total* gain in this game. What is the expected value of $G$?

$\mathbb{E}[G] = \mathbb{E}[G_1] + \mathbb{E}[G_2] + \mathbb{E}[G_3]$, where $G_i$ is the expected gain from flip $i$. We already know that $\mathbb{E}[G_i] = 0$, so $\mathbb{E}[G] = 0 + 0 + 0 = 0$ as well.

c. What is the expected value of $G$ if the coin is biased and the probability of heads is $p$? in other words, generalize your solution from part b in terms of $p$.

Now, $\mathbb{E}[G_i] = p \cdot 1 + (1 - p) \cdot (-1) = p - 1 + p = 2p - 1$. Thus, $\mathbb{E}[G] = \mathbb{E}[G_1] + \mathbb{E}[G_2] + \mathbb{E}[G_3] = (2p - 1) + (2p - 1) + (2p - 1) = 6p - 3$.

d. Tim and Joe are still using the biased coin from part c. Let $H_1$ be the event that the first coin is heads. What is $\mathbb{E}[G|H_1]$?

$\mathbb{E}[G \mid H_1] = 1 + (2p - 1) + (2p - 1) = 4p - 1$.

$G_1$ is 1 from the first round guaranteed, and the other terms are from part c.

e. Use your answers to calculate $\mathbb{E}[G]$ and $\mathbb{E}[G \mid H_1]$ when $p = 0.7$.

$\mathbb{E}[G] = 6p - 3 = 6 \cdot 0.7 - 3 = 1.2$

$\mathbb{E}[G \mid H_1] = 4p - 1 = 4 \cdot 0.7 - 1 = 1.8$

f. Assume that $p = 0.7$ and let's say we want to change the game to make it "fair." If the flip is tails, then Tim pays a dollar to Joe—how much should Joe pay Tim on Heads so that for any number of flips we know $\mathbb{E}[G] = 0$?

Each $\mathbb{E}[G_i]$ is $0.3 \cdot (-1) + 0.7 \cdot x$, where $x$ is the number of dollars that Joe should pay Tim on Heads. Thus, we have $\mathbb{E}[G] = (0.3 \cdot (-1) + 0.7 \cdot x) \cdot 3 = 0$. Solving for $x$, we have $x = \frac{3}{7}$. So, Joe should pay Tim \$0.43 on heads so that the game is fair.

### Task 2

Tim and Joe are now playing a similar, but different, game. This time they flip a coin **2 times**. Let $X$ be the random variable that is equal to the number of heads and $Y$ the random variable that is equal to the number of tails. At the end of the game, Joe pays Tim $X^2$ dollars. Once again, let $G$ be the random variable for Tim's gain.

a. Calculate $\mathbb{E}[X]^2$ when the probability of heads is $p = 0.7$.

$\mathbb{E}[X] = 0.7 + 0.7 = 1.4$. So, $\mathbb{E}[X]^2 = 1.4^2 = 1.96$

b. Calculate $\mathbb{E}[G]$ when $p = 0.7$.

$\mathbb{E}[G] = \mathbb{E}[X^2] = (0.7)^2(4) + \binom{2}{1}(0.3)(0.7)(1) = 2.38$

c. Does $\mathbb{E}[G] = \mathbb{E}[X]^2$?

No!

In general for any random variable with nonzero variance we have $\mathbb{E}[X^2] > \mathbb{E}[X])^2$.

d. Find an example that shows that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ does not hold where $X$ and $Y$ are not independent variables.

Hint: Try $X$ and $Y$ as defined above.

We know that $X$ and $Y$, as defined in this problem, are not independent variables. We know that $\mathbb{E}[X]\mathbb{E}[Y] = 1 \cdot 1 = 1$. We can also calculate $\mathbb{E}[XY]$ to be $(\frac{1}{4})(2)(0) + (\frac{1}{2})(1)(1) = (\frac{1}{4})(0)(2) = \frac{1}{2}$.

Thus, this is an example of where $X$ and $Y$ are dependent variables where $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$.

e. Prove that if $X$ and $Y$ are two independent random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Assume $X$ and $Y$ are independent. So for any $x$ and $y$ in their range we have $\Pr(X = x \wedge Y = y) = \Pr(X = x)\Pr(Y = y)$. We now use the definition of expected value.

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x \in X(S)} \sum_{y \in Y(S)} xy \Pr(X = x \wedge Y = y) && (2) \\
&= \sum_{x \in X(S)} \sum_{y \in Y(S)} xy \Pr(X = x)\Pr(Y = y) && (3) \\
&= \left( \sum_{x \in X(S)} x \Pr(X = x) \right) \left( \sum_{y \in Y(S)} y \Pr(Y = y) \right) && (4) \\
&= \mathbb{E}[X]\mathbb{E}[Y] && (5)
\end{aligned}$$

## Bayes' Rule

Bayes' Rule can be summarized as

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

where $A$ and $B$ are events and $\Pr(B) \neq 0$.

**Task 3**

Assume Brown's CS department has an evaluation system for CS courses based on student evaluations. In any class, the students can fill the evaluation form and give a score of $0, 1$, or $2$ to the course. Let $X$ be the random variable of this score. The students of CS0220 either like the course with probability $3/4$ (Event $L$) or they do not like the course with probability $1/4$ (Event $\neg L$).

Assume that the conditional probability distribution of $X$ given $L$ is

$$\Pr(X = 0 \mid L) = 1/8$$

$$\Pr(X = 1 \mid L) = 1/4$$

$$\Pr(X = 2 \mid L) = 5/8$$

and given that they do not like the course ($\neg L$) it is

$$\Pr(X = 0 \mid \neg L) = 9/10$$

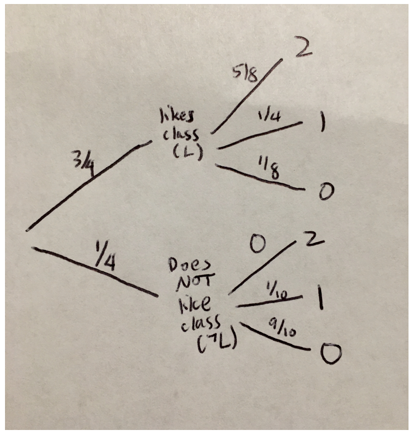
$$\Pr(X = 1 \mid \neg L) = 1/10$$

$$\Pr(X = 2 \mid \neg L) = 0.$$

a. If a student has given score of 0 to CS0220, what is the probability that they do not like the course?

$\Pr(\neg L \mid X = 0) = \frac{\Pr(X=0|\neg L)\Pr(\neg L)}{\Pr(X=0)} = \frac{\frac{9}{10}\cdot\frac{1}{4}}{\frac{1}{4}\cdot\frac{9}{10}+\frac{3}{4}\cdot\frac{1}{8}} = \frac{12}{17}$. The formula for conditional probability $\Pr(\neg L \mid X = 0) = \frac{\Pr(\neg L \cap X=0)}{\Pr(X=0)}$ gives the same answer.

Note: Drawing out a diagram like this may be helpful for this problem:




b. Use the definition of conditional expected value (Equation 1) and find $\mathbb{E}[X \mid \neg L]$

$\mathbb{E}[X \mid \neg L] = \sum r \Pr(X = r \mid \neg L) = 2 \cdot 0 + 1 \cdot \frac{1}{10} + 0 \cdot \frac{9}{10} = \frac{1}{10}$.

c. *Optional:* Find $\mathbb{E}[X]$.

$$\begin{aligned}\mathbb{E}[X] =& 2\Pr(X=2) + 1\Pr(X=1) + 0\Pr(X=0)\\ =& 2\left(\Pr(X=2 \mid L)\Pr(L) + \Pr(X=2 \mid \neg L)\Pr(\neg L)\right)\\ &+ \Pr(X=1 \mid L)\Pr(L) + (\Pr(X=1 \mid \neg L)\Pr(\neg L)\\ &= 2 \cdot \frac{3}{4} \cdot \frac{5}{8} + 1 \cdot \frac{3}{4} \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} \cdot \frac{1}{10} = \frac{23}{20} = 1.15.\end{aligned}$$

Alternative solution: $\mathbb{E}[X] = \Pr(L)\mathbb{E}[X \mid L] + \Pr(\neg L)\mathbb{E}[X \mid \neg L] = \frac{3}{4} \cdot (\frac{5}{8} \cdot 2 + \frac{1}{4} \cdot 1 + \frac{1}{8} \cdot 0) + \frac{1}{4} \cdot (0 \cdot 2 + \frac{1}{10} \cdot 1 + \frac{9}{10} \cdot 0) = \frac{23}{20} = 1.15$

## Checkpoint #1 — Call over a TA!

## Variation from the mean

Sometimes measuring the mean (expectation) of a random variable doesn't give us enough information: it can be helpful to know how much we expect the variable to *stray* from its average.

*Markov's inequality* gives a generally coarse estimate of the probability that a random variable takes a value much larger than its mean.

**Theorem (Markov).** If $R$ is a nonnegative random variable, then for all $x > 0$,

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}.$$

Expressed differently:

**Corollary.** If $R$ is a nonnegative random variable, then for all $c \geq 1$,

$$\Pr[R \geq c \cdot \mathbb{E}[R]] \leq \frac{1}{c}.$$

That is: the probability of $R$ being more than $c$ times its mean is at most $1/c$.

A related notion is that of *variance*:

**Definition.** The *variance* $\mathrm{Var}[R]$ of a random variable $R$ is defined to be $\mathbb{E}[(R - \mathbb{E}[R])^2]$.

Unpacking this from the inside out: $R - \mathbb{E}[R]$ is a random variable measuring the distance between $R$ and its mean at each outcome. Averaging the square of this gives us a sense of, overall, how far $R$ tends to be from its mean.

There is an equivalent way to state this:

**Lemma.** For any random variable $R$,

$$\mathrm{Var}[R] = \mathbb{E}[R^2] - (\mathbb{E}[R])^2.$$

### Task 4

Prove the above lemma!

For notational convenience, we will define $\mu := \mathbb{E}[R]$.

$$\begin{aligned}
\mathrm{Var}[R] &= \mathbb{E}[(R - \mu)^2] \\
&= \mathbb{E}[R^2 - 2\mu R + \mu^2] \\
&= \mathbb{E}[R^2] - 2\mu\mathbb{E}[R] + \mu^2
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[R^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[R^2] - \mu^2 \\
&= \mathbb{E}[R^2] - (\mathbb{E}[R])^2
\end{aligned}
$$

Note that the third line follows by linearity of expectation, with $\mathbb{E}[\mu^2] = \mu^2$ since $\mu$ is a constant.

This leads us to state Chebyshev's theorem, an application of Markov's inequality:

**Theorem (Chebyshev).** Let $R$ be a random variable and $x \in \mathbb{R}^+$. Then

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

**Task 5**

Suppose you flip a fair coin 100 times. The coin flips are all mutually independent.

a. What is the expected number of heads?

Let $X$ be the random variable denoting the number of heads.

$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{100}] = \underbrace{(1 \cdot \frac{1}{2}) + (1 \cdot \frac{1}{2}) + \cdots (1 \cdot \frac{1}{2})}_{\text{100 times}} = 100 \cdot \frac{1}{2} = 50$

b. What upper bound on the probability that the number of heads is at least 70 can we derive using Markov's inequality?

$\Pr[X \geq 70] \leq \frac{\mathbb{E}[X]}{70} = \frac{50}{70}$

Using Markov's inequality, we can derive a corresponding upper bound of $\frac{5}{7} \approx 0.714$.

c. What is the variance of the number of heads? The following theorem may help:

**Theorem.** Let $X$ and $Y$ be independent random variables. Then

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

(*Note*: This does **not** hold as a general property of variance!)

Given that all coin flips are mutually independent, we can say

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots \mathrm{Var}[X_{100}].$$

For $X_i$ such that $1 \leq i \leq 100$,

$$\mathrm{Var}[X_i] = \mathbb{E}[(X_i)^2] - (\mathbb{E}[X_i])^2.$$

Calculating the terms independently, we get

$$\mathbb{E}[(X_i)^2] = (1)^2 \cdot \frac{1}{2} + (0)^2 \cdot \frac{1}{2} = \frac{1}{2}$$

$$(\mathbb{E}[X_i])^2 = (\frac{1}{2})^2 = \frac{1}{4}$$

So

$$\mathrm{Var}[X_i] = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Since this generally holds for the $i^{\text{th}}$ flip, we can conclude

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots \mathrm{Var}[X_{100}] = \underbrace{\frac{1}{4} + \frac{1}{4} + \cdots \frac{1}{4}}_{\text{100 times}} = 25.$$

d. What upper bound does Chebyshev's Theorem give us on the probability that the number of heads is either less than 30 or greater than 70?

$\Pr[X > 70 \cup X < 30] = \Pr[|X - \mathbb{E}[X]| \geq 20]$

By Chebyshev's Theorem, $\Pr[|X - \mathbb{E}[X]| \geq 20] \leq \frac{25}{20^2} = \frac{1}{16}$.

Thus, we are given an upper bound of $\frac{1}{16} = 0.0625$.

**Task 6**

A herd of dinosaurs is stricken by an outbreak of cold dino disease. The disease lowers a dinosaur's body temperature from normal levels, and a dino will die if its temperature goes below 90 degrees F. The disease epidemic is so intense that it lowered the average temperature of the herd to 85 degrees. Body temperatures as low as 70 degrees, but no lower, were actually found in the herd.

a. Use Markov's inequality to prove that at most 3/4 of the dinos could survive.

*Hint*: This isn't necessarily a direct result from applying Markov's inequality to the expected value itself. Think about how the bound can be expressed differently – perhaps in relation to the minimum temperature.

Let random variable $T$ denote a dinosaur's body temperature in degrees F.

$\mathbb{E}[T]$ is the average temperature of the herd, so $\mathbb{E}[T] = 85$

Applying Markov's inequality to $T$, we get

$$\Pr[T \geq 90] \leq \frac{\mathbb{E}[T]}{90} = \frac{85}{90}$$

.

But $\frac{85}{90} = \frac{17}{18} > \frac{3}{4}$, so this is not tight enough of a bound.

Instead, apply Markov's inequality to $T - 70$:

$$\Pr[T \geq 90] = \Pr[T - 70 \geq 20] \leq \frac{\mathbb{E}[T-70]}{20} = \frac{85-70}{20} = \frac{15}{20} = \frac{3}{4}.$$

b. Suppose there are 400 dinos in the herd. Show that the bound from part a is the best possible by giving an example set of temperatures for the dinos so that the average herd temperature is 85 and 3/4 of the dinos will have a high enough temperature to survive.

Let 300 dinosaurs have a temperature of 90 degrees and the remaining 100 have a temperature of 70 degrees. The average temperature of the herd is $\frac{300 \cdot 90 + 100 \cdot 70}{400} = \frac{34000}{400} = 85$ degrees.

## Checkoff — Call over a TA!