**Math 170B** **Winter 2017**

# Problem Set 2

*Lecturer: Steven Heilman* *Kyle Barron*

## Exercise 1

Let $X, Y$ be random variables with $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[Y^2] < \infty$. Prove the Cauchy-Schwarz inequality:

$$\mathbb{E}[XY] \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}.$$

Then, deduce the following when $X, Y$ both have finite variance:

$$|\text{cov}(X, Y)| \leq \text{var}(X)^{1/2}\text{var}(Y)^{1/2}.$$

(Hint: in the case that $\mathbb{E}[Y^2] > 0$, expand out the product $\mathbb{E}[X - Y\mathbb{E}[XY]/\mathbb{E}[Y^2]]^2$.)

**Proof:** Observe that we can rewrite the above inequality as

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

We can assume that $\mathbb{E}[Y^2] \neq 0$, because if $\mathbb{E}[Y^2] = 0$, then we have $Y = 0$ with probability 1, and $\mathbb{E}[XY] = 0$, so the inequality holds. Then we have

$$\begin{aligned}
0 &\leq \mathbb{E}\left[\left(X - \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}Y\right)^2\right] \\
&= \mathbb{E}\left[X^2 - 2\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}XY + \frac{(\mathbb{E}[XY])^2}{(\mathbb{E}[Y^2])^2}Y^2\right] \\
&= \mathbb{E}[X^2] - 2\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}\mathbb{E}[XY] + \frac{(\mathbb{E}[XY])^2}{(\mathbb{E}[Y^2])^2}\mathbb{E}[Y^2] \\
&= \mathbb{E}[X^2] - \frac{(\mathbb{E}[XY])^2}{\mathbb{E}[Y^2]} \Rightarrow \\
\mathbb{E}[XY]^2 &\leq \mathbb{E}[X^2]\mathbb{E}[Y^2].
\end{aligned}$$

Then for the second part of the problem, let $\tilde{X} = X - \mathbb{E}[X]$ and let $\tilde{Y} = Y - \mathbb{E}[Y]$. Using the Cauchy-Schwarz inequality, we have

$$\text{cov}(X, Y)^2 = \mathbb{E}[\tilde{X}\tilde{Y}]^2 \leq \mathbb{E}[\tilde{X}^2]\mathbb{E}[\tilde{Y}^2] = \text{var}(X)\text{var}(Y).$$

Therefore,

$$|\text{cov}(X, Y)| \leq \text{var}(X)^{\frac{1}{2}}\text{var}(Y)^{\frac{1}{2}}.$$

■

## Exercise 2

Let $X$ be a binomial random variable with parameters $n = 2$ and $p = 1/2$. So $\mathbf{P}(X = 0) = \frac{1}{4}$, $\mathbf{P}(X = 1) = \frac{1}{2}$, and $\mathbf{P}(X = 2) = \frac{1}{4}$. And $X$ satisfies $\mathbb{E}[X] = 1$ and $\mathbb{E}[X^2] = \frac{3}{2}$.

Let $Y$ be a geometric random variable with parameter 1/2. So, for any positive integer $k$, $\mathbf{P}(Y=k)=2^{-k}$. And $Y$ satisfies $\mathbb{E}[Y]=2$ and $\mathbb{E}[Y^2]=6$.

Let $Z$ be a Poisson random variable with parameter 1. So, for any nonnegative integer $k$, $\mathbf{P}(Z=k)=\frac{1}{e}\frac{1}{k!}$. And $Z$ satisfies $\mathbb{E}[Z]=1$ and $\mathbb{E}[Z^2]=2$.

Let $W$ be a discrete random variable such that $\mathbf{P}(W=0)=1/2$ and $\mathbf{P}(W=4)=1/2$, so that $\mathbb{E}[W]=2$ and $\mathbb{E}[W^2]=8$.

Assume that $X$, $Y$, $Z$, and $W$ are all independent. Compute $\mathrm{var}(X+Y+Z+W)$.

**Solution:** Recall that we can compute variance as $\mathrm{var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2$. Therefore,

$$\begin{aligned}
\mathrm{var}(X) &= \frac{3}{2}-1^2=\frac{1}{2},\\
\mathrm{var}(Y) &= 6-2^2=2,\\
\mathrm{var}(Z) &= 2-1^2=1,\\
\mathrm{var}(W) &= 8-2^2=4.
\end{aligned}$$

Using Corollary 2.25 from the notes, when $X_1,...,X_n$ are independent we have $\mathrm{var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{var}(X_i)$. Therefore, since $X$, $Y$, $Z$, and $W$ are independent we have

$$\mathrm{var}(X+Y+Z+W)=\mathrm{var}(X)+\mathrm{var}(Y)+\mathrm{var}(Z)+\mathrm{var}(W)=\frac{1}{2}+2+1+4=\frac{15}{2}.$$

■

## Exercise 3

Let $X_1,...,X_n$ be random variables with finite variance. Define an $n\times n$ matrix $A$ such that $A_{ij}=\mathrm{cov}(X_i,X_j)$ for any $1\le i,j\le n$. Show that the matrix $A$ is positive semidefinite. That is, show that for any $y=(y_1,...,y_n)\in\mathbb{R}^n$, we have

$$y^TAy=\sum_{i,j=1}^n y_iy_ja_{ij}\ge 0.$$

**Solution:** Let $A$ be an $n\times n$ matrix where $A_{ij}=\mathrm{cov}(X_i,X_j)$ for any $1\le i,j\le n$.

$$\begin{aligned}
y^TAy &= \sum_{i,j=1}^n y_iy_ja_{ij}\\
&= \sum_{i,j=1}^n y_iy_j\mathrm{cov}(X_i,X_j)\\
&= \sum_{j=1}^n y_j\mathrm{cov}\left(\sum_{i=1}^n X_i, X_j\right)\\
&= \mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n y_jX_j\right)\\
&= \mathrm{var}\left(\sum_{i=1}^n X_i\right)\ge 0.
\end{aligned}$$

■

## Exercise 4

(Another Total Expectation Theorem) Using the definition of $\mathbb{E}(X|Y)$, prove the following theorem, which can be considered as a version of a Total Expectation Theorem:

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X).$$

**Proof:** Supposing that $X$ and $Y$ are continuous random variables,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[Y|X=x] f_X(x)dx \\
&= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)} dy \right) f_X(x)dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} y f_Y(y)dy \\
&= \mathbb{E}[Y].
\end{aligned}$$

The proof is identical for discrete random variables by substituting $\int dx$ for $\sum_x$ and $\int dy$ for $\sum_y$. ■

## Exercise 5

If $X$ is a random variable, and if $f(t) := \mathbb{E}[(X-t)^2]$, $t \in \mathbb{R}$, then the function $f : \mathbb{R} \to \mathbb{R}$ is uniquely minimized when $t = \mathbb{E}[X]$. This follows e.g. by writing

$$\begin{aligned}
\mathbb{E}[(X-t)^2] &= \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - t)^2] \\
&= \mathbb{E}[X - \mathbb{E}[X]]^2 + (\mathbb{E}[X] - t)^2 + 2\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - t)] \\
&= \mathbb{E}[X - \mathbb{E}[X]]^2 + (\mathbb{E}[X] - t)^2.
\end{aligned}$$

So, the choice $t = \mathbb{E}[X]$ is the smallest, and it recovers the definition of variance, since $\mathrm{var}(X) = \mathbb{E}[X - \mathbb{E}[X]]^2$.

A similar minimizing property holds for conditional expectation. Let $h : \mathbb{R} \to \mathbb{R}$. Show that the quantity $\mathbb{E}[X - h(Y)]^2$ is minimized among all functions $h$ when $h(Y) = \mathbb{E}[X|Y]$. (Hint: Exercise 4 might be helpful.)

**Proof:**

$$\begin{aligned}
\mathbb{E}[(X - h(Y))^2] &= \mathbb{E}[(X - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - h(Y))^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X|Y])^2] + 2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - h(Y))] + \mathbb{E}[(\mathbb{E}[X|Y] - h(Y))^2].
\end{aligned}$$

It is sufficient to show that $2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - h(Y))] = 0$. Then it is clear that $h(Y) = \mathbb{E}[X|Y]$ minimizes $\mathbb{E}[(X - h(Y))^2]$. From the textbook, we know that $\mathbb{E}[Xh(Y)|Y] = h(Y)\mathbb{E}[X|Y]$. Therefore,

$$\begin{aligned}
2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - h(Y))] &= 2\mathbb{E}\Big[\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - h(Y))]\Big] \\
&= 2\mathbb{E}[(\mathbb{E}[X|Y] - h(Y))]\mathbb{E}[X - \mathbb{E}[X|Y]|Y] \\
&= 2(\mathbb{E}[X|Y] - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - \mathbb{E}[X|Y]) \\
&= 0.
\end{aligned}$$

Therefore $\mathbb{E}[(X - h(Y))^2]$ is minimized when $h(Y) = \mathbb{E}[X|Y]$. ■

## Exercise 6

Toys are stored in small boxes, small boxes are stored in large crates, and large crates comprise a shipment. Let $X_i$ be the number of toys in small box $i \in \{1, 2, ...\}$. Assume that $X_i, X_2, ...$ all have the same CDF. Let $Y_i$ be the number of small boxes in large crate $i \in \{1, 2, ...\}$. Assume that $Y_1, Y_2, ...$ all have the same CDF. Let $Z$ be the number of large crates in the shipment. Assume that $X_1, X_2, ..., Y_1, Y_2, ..., Z$ are all independent, nonnegative integer-valued random variables, each with expected value 10 and variance 16.

Compute the expected value and variance of the number of toys in the shipment.

**Solution:**

Let $T_1$ be the number of toys in the large crate. Let $y \in \mathbb{N}_+$. Given that $Y_1 = y$, we know that $T_1 = X_1 + ... + X_y$. So using independence, we have

$$\mathbb{E}[T_1|Y_1 = y] = \mathbb{E}[X_1 + X_2 + ... + X_y|Y_1 = y] = y\mathbb{E}[X_1] = 10y.$$

So by the definition of conditional expectation, $\mathbb{E}[T_1|Y_1] = 10Y_1$. We know from Exercise 4 that

$$\mathbb{E}[T_1] = \mathbb{E}[\mathbb{E}[T_1|Y_1]] = \mathbb{E}[10Y_1] = 10\mathbb{E}[Y_1] = 100.$$

For the variance, from the definition of conditional variance and Corollary 2.25, we observe that

$$\begin{aligned}
\text{var}(T_1|Y_1 = y) &= \mathbb{E}[(T_1 - 10y)^2|Y_1 = y] \\
&= \mathbb{E}[(X_1 + X_2 + ... + X_y - 10y)^2|Y_1 = y] \\
&= \mathbb{E}[(X_1 + X_2 + ... + X_y - \mathbb{E}[X_1 + ... + X_y])^2] \\
&= \text{var}(X_1 + X_2 + ... + X_y) \\
&= y\text{var}(X_1) \\
&= 16y.
\end{aligned}$$

So, $\text{var}(T_1|Y_1) = 16Y_1$, and by Proposition 2.34 we have

$$\begin{aligned}
\text{var}(T_1) &= \mathbb{E}[\text{var}(T_1|Y_1)] + \text{var}(\mathbb{E}[T_1|Y_1]) \\
&= \mathbb{E}[16Y_1] + \text{var}(10Y_1) \\
&= 16\mathbb{E}[Y_1] + 100\text{var}(Y_1) \\
&= 160 + 1600 = 1760.
\end{aligned}$$

Now let $W$ be the number of toys in the entire shipment. Let $z \in \mathbb{N}_+$. Given that $Z = z$, we know that $W = T_1 + T_2 + ... + T_z$. So using independence we have

$$\mathbb{E}[W|Z = z] = \mathbb{E}[T_1 + T_2 + ... + T_z|Z = z] = z\mathbb{E}[T_1] = 100z.$$

So by the definition of conditional expectation, $\mathbb{E}[W|Z] = 100Z$. By Exercise 4,

$$\mathbb{E}[W] = \mathbb{E}[\mathbb{E}[W|Z]] = \mathbb{E}[100Z] = 100\mathbb{E}[Z] = 1000.$$

From the definition of conditional variance and Corollary 2.25, we observe that

$$\begin{aligned}
\text{var}(W|Z = z) &= \mathbb{E}[(W - 100z)^2|Z = z] \\
&= \mathbb{E}[(T_1 + T_2 + ... + T_z - 100z)^2|Z = z] \\
&= \mathbb{E}[(T_1 + T_2 + ... + T_z - \mathbb{E}[T_1 + ... + T_z])^2] \\
&= \text{var}(T_1 + T_2 + ... + T_z) \\
&= z\text{var}(T_1) \\
&= 1760z.
\end{aligned}$$

So $\mathrm{var}(W|Z) = 1760Z$, and by Proposition 2.34 we have

$$\begin{aligned}\mathrm{var}(W) &= \mathbb{E}[\mathrm{var}(W|Z)] + \mathrm{var}(\mathbb{E}[W|Z]) \\ &= \mathbb{E}[1760Z] + \mathrm{var}(100Z) \\ &= 1760\mathbb{E}[Z] + 10000\mathrm{var}(Z) \\ &= 1760 \cdot 10 + 10000 \cdot 16 \\ &= 177600.\end{aligned}$$

So the expected number of toys in the shipment is 1000 with a variance of 177600. ■

## Exercise 7

Let $0 < p < 1$. Suppose you have a biased coin which has a probability $p$ of landing heads, and probability $1 - p$ of landing tails, each time it is flipped. Also, suppose you have a fair six-sided die (so each face of the cube has a distinct label from the set $\{1, 2, 3, 4, 5, 6\}$, and each time you roll the die, any face of the cube is rolled with equal probability.)

Let $N$ be the number of coin flips you need to do until the first head appears. Now, roll the fair die $N$ times. Let $S$ be the sum of the results of the $N$ rolls of the die. Compute $\mathbb{E}[S]$ and $\mathrm{var}(S)$.

**Solution:**

Let $N$ be the number of coin flips that it takes to make the first head appear. Let $S$ be the sum of the results of the $N$ rolls of the die. Then $N$ is a geometric random variable with parameter $p$. Therefore we have

$$\mathbb{E}[N] = \frac{1}{p} \text{ and } \mathrm{var}(N) = \frac{1-p}{p^2}.$$

Let $D_i$ be the random variable corresponding to each roll of the dice. Note that $\mathbb{E}[D_i] = 3.5$ and that $\mathrm{var}(D_i) = \frac{17.5}{6}$. Let $n \in \mathbb{N}_+$, so that $S = D_1 + D_2 + ... + D_n$ when $N = n$. Then

$$\mathbb{E}[S|N = n] = \mathbb{E}[D_1 + ... + D_n|N = n] = n\mathbb{E}[D_1] = 3.5n.$$

By the definition of conditional expectation, we have $\mathbb{E}[S|N] = 3.5N$. By Exercise 4 we have

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|N]] = \mathbb{E}[3.5N] = 3.5\mathbb{E}[N] = \frac{3.5}{p}.$$

From the definition of conditional variance and Corollary 2.25, we have

$$\begin{aligned}\mathrm{var}(S|N = n) &= \mathbb{E}[(S - \mathbb{E}[S|N])^2|N = n] \\ &= \mathbb{E}[(D_1 + D_2 + ... + D_n - 3.5N)^2|N = n] \\ &= \mathrm{var}(D_1 + ... + D_n) \\ &= n\mathrm{var}(D_1) \\ &= n\frac{17.5}{6}.\end{aligned}$$

So then $\text{var}(S|N) = N\frac{17.5}{6}$. Then by proposition 2.34,

$$\begin{aligned}\text{var}(S) &= \mathbb{E}[\text{var}(S|N)] + \text{var}(\mathbb{E}[S|N]) \\ &= \mathbb{E}[N\frac{17.5}{6}] + \text{var}(3.5N) \\ &= \frac{17.5}{6}\mathbb{E}[N] + 3.5^2 \cdot \text{var}(N) \\ &= \frac{17.5}{6} \cdot \frac{1}{p} + \frac{49}{4} \cdot \frac{1-p}{p^2} \\ &= \frac{147 - 112p}{12p^2}.\end{aligned}$$

So we have $\mathbb{E}[S] = 3.5\frac{1-p}{p}$ and $\text{var}(S) = \frac{147-112p}{12p^2}$. ∎

## Exercise 8

Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function. Assume that $f$ is convex. That is, $f''(x) \geq 0$, or equivalently, the graph of $f$ lies above any of its tangent lines. That is, for any $x, y \in \mathbb{R}$,

$$f(x) \geq f(y) + f'(y)(x - y).$$

Let $X$ be a discrete random variable. By setting $y = \mathbb{E}[X]$, prove **Jensen's inequality**:

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

In particular, choosing $f(x) = x^2$, we have $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$.

**Proof:** Let $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable and convex. Therefore, for all $x, y \in \mathbb{R}$,

$$f(x) \geq f(y) + f'(y)(x - y).$$

Let $x = X$ and let $y = \mathbb{E}[X]$. Then

$$f(X) \geq f(\mathbb{E}[X]) + f'(\mathbb{E}[X])(X - \mathbb{E}[X]).$$

Notice however, that $\mathbb{E}[X]$ is a constant, so its derivative is zero. Now taking the expectation of both sides, we get

$$\begin{aligned}\mathbb{E}[f(X)] &\geq \mathbb{E}[f(\mathbb{E}[X]) + f'(\mathbb{E}[X])(X - \mathbb{E}[X])] \\ &= f(\mathbb{E}[X]) + \mathbb{E}[f'(\mathbb{E}[X])(X - \mathbb{E}[X])] \\ &= f(\mathbb{E}[X]).\end{aligned}$$

Therefore $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ for a convex function $f$. ∎