36-401 Modern Regression Homework #1 Solutions

DUE: September 8, 2017

Problem 1 [20 pts.]

First consider X_1 and X_2 with joint probability function $p(x_1, x_2)$.

$$\begin{aligned}
\mathbb{E}\Big[a_1X_1 + a_2X_2\Big] &= \sum_{x_1}\sum_{x_2}(a_1x_1 + a_2x_2)\cdot p(x_1,x_2) && (1)\\
&= a_1\sum_{x_1}\sum_{x_2}x_1\cdot p(x_1,x_2) + a_2\sum_{x_1}\sum_{x_2}x_2\cdot p(x_1,x_2) && (2)\\
&= a_1\sum_{x_1}x_1\sum_{x_2}p(x_1,x_2) + a_2\sum_{x_2}x_2\sum_{x_1}p(x_1,x_2) && (3)\\
&= a_1\sum_{x_1}x_1\cdot p(x_1) + a_2\sum_{x_2}x_2\cdot p(x_2) && (4)\\
&= a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] && (5)
\end{aligned}$$

Now assume

$$\mathbb{E}\left[\sum_{j=1}^{k}a_jX_j\right] = \sum_{j=1}^{k}a_j\mathbb{E}[X_j]$$

holds for some $k \in \mathbb{Z}^+$, and define

$$Y := \sum_{j=1}^{k}a_jX_j.$$

Then

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{k+1}a_jX_j\right] &= \mathbb{E}\Big[Y + a_{k+1}X_{k+1}\Big]\\
&= \mathbb{E}[Y] + a_{k+1}\mathbb{E}[X_{k+1}] && (6)\\
&= \sum_{j=1}^{k+1}a_j\mathbb{E}[X_j],
\end{aligned}$$

where (6) follows from (1)-(5). Therefore,

$$\mathbb{E}\left[\sum_{j=1}^{m}a_jX_j\right] = \sum_{j=1}^{m}a_j\mathbb{E}[X_j]$$

for any $m \in \mathbb{Z}^+$, by induction.

Notice (5) also implies

$$\mathbb{E}[a_1X + a_2] = a_1\cdot\mathbb{E}[X] + a_2$$

by letting X_2 be a degenerate random variable with $P(X_2 = 1) = 1$.

Problem 2 [20 pts.]

$$\sum_{j=1}^{\infty} P(X \geq j) = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) \tag{7}$$
$$= \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(X = k) \tag{8}$$
$$= \sum_{k=1}^{\infty} k \cdot P(X = k)$$
$$= \mathbb{E}[X]$$

To understand what is happening with the interchange of summations in (8) it may help to write (7) as

$$\begin{aligned}
\sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) = P(X = 1) + P(X = 2) \quad &+ P(X = 3) + P(X = 4) \quad + P(X = 5) + P(X = 6) + \cdots \\
+ P(X = 2) \quad &+ P(X = 3) + P(X = 4) \quad + P(X = 5) + P(X = 6) + \cdots \\
&+ P(X = 3) + P(X = 4) \quad + P(X = 5) + P(X = 6) + \cdots \\
&\qquad\quad\;\, + P(X = 4) \quad + P(X = 5) + P(X = 6) + \cdots \\
&\qquad\qquad\qquad\qquad\quad\;\; + P(X = 5) + P(X = 6) + \cdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad\quad\;\; + P(X = 6) + \cdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \vdots
\end{aligned}$$

(7) sums over this "matrix" by row and (8) sums over it by column.

Alternate approach

One could also sum over all entries of the above matrix with zeros plugged into the lower triangle, i.e.

$$\begin{aligned}
\sum_{j=1}^{\infty} P(X \geq j) &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} P(X = k) \cdot \mathbb{1}_{\{k \geq j\}} \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} P(X = k) \cdot \mathbb{1}_{\{k \geq j\}} \\
&= \sum_{k=1}^{\infty} k \cdot P(X = k) \\
&= \mathbb{E}[X].
\end{aligned}$$

Problem 3 [20 pts.]

(a)

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 \cdot p(x) dx \\
&= (a - \mathbb{E}[X])^2 \cdot p(a) + \int_{\mathbb{R}\setminus\{a\}} (x - \mathbb{E}[X])^2 \cdot \underbrace{p(x)}_{= 0 \text{ for all } x \in \mathbb{R}\setminus\{a\}} dx \\
&= (a - \mathbb{E}[X])^2 \cdot 1 \\
&= \left(a - \int_{\mathbb{R}} x \cdot p(x) dx\right)^2 \\
&= (a - a \cdot p(a))^2 \\
&= (a - a)^2 \\
&= 0.
\end{aligned}$$

(b) Here we assume X is discrete and $\text{Var}(X) = 0$, i.e.

$$\begin{aligned}
\text{Var}(X) &= \sum_x (x - \mathbb{E}[X])^2 \cdot p(x) \qquad (9) \\
&= 0.
\end{aligned}$$

Since every term in (9) is nonnegative, the above implies

$$(x - \mathbb{E}[X])^2 \cdot p(x) = 0 \qquad (10)$$

for all x.

For (10) to hold, then any time $p(x) > 0$, we must have $(x - \mathbb{E}[X])^2 = 0$.

Now assume there are two *distinct* values x_1 and x_2 for which $p(x_1) > 0$ and $p(x_2) > 0$. But (10) implies

$$x_1 = x_2 = \mathbb{E}[X],$$

a contradiction. Therefore,

$$P(X = a) = 1,$$

where $a = \mathbb{E}[X]$.

Problem 4 [20 pts.]

$$
\begin{aligned}
\mathrm{Cov}(a + bX, c + dY) &= \mathbb{E}\big[(a + bX - \mathbb{E}[a + bX])(c + dY - \mathbb{E}[c + dY])\big] \\
&= \mathbb{E}\big[(a + bX - a - b \cdot \mathbb{E}[X])(c + dY - c - d \cdot \mathbb{E}[Y])\big] \\
&= \mathbb{E}\big[(bX - b \cdot \mathbb{E}[X])(dY - d \cdot \mathbb{E}[Y])\big] \\
&= \mathbb{E}\big[b \cdot (X - \mathbb{E}[X]) \cdot d \cdot (Y - \mathbb{E}[Y])\big] \\
&= bd \cdot \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] \\
&= bd \cdot \mathrm{Cov}(X, Y),
\end{aligned}
$$

where we have used the linearity of expectation, established in Problem 1.

Problem 5 [20 pts.]

(a)

$$\begin{aligned}\mathbb{E}[Y] &= 5\mathbb{E}[X] + \mathbb{E}[\epsilon] && \text{(by linearity of expectation)}\\ &= 5 \cdot 0 + 0\\ &= 0\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}[Y])^2]\\ &= \mathbb{E}[(5X + \epsilon)^2]\\ &= \mathbb{E}[25X^2 + 10X \cdot \epsilon + \epsilon^2]\\ &= 25 \cdot \mathbb{E}[X^2] + 10 \cdot \mathbb{E}[X \cdot \epsilon] + \mathbb{E}[\epsilon^2]\\ &= 25 \cdot \mathbb{E}[X^2] + 10 \cdot \mathbb{E}[X] \cdot \mathbb{E}[\epsilon] + \mathbb{E}[\epsilon^2] && \text{(by independence)}\\ &= 25 \cdot (\mathrm{Var}(X) + \mathbb{E}[X]^2) + 10 \cdot \mathbb{E}[X] \cdot \mathbb{E}[\epsilon] + \mathrm{Var}(\epsilon) + \mathbb{E}[\epsilon]^2 && \text{(by variance formula)}\\ &= 25 \cdot \left(\frac{1}{3} + 0^2\right) + 10 \cdot 0 \cdot 0 + 1 + 0^2\\ &= \frac{28}{3}\end{aligned}$$

(b)

$$\begin{aligned}\mathbb{E}[Y^2] &= \mathrm{Var}(Y) + \mathbb{E}[Y]^2\\ &= \frac{28}{3} + 0^2\\ &= \frac{28}{3}\end{aligned}$$

(c)

$$\begin{aligned}\mathbb{E}[Y \mid X = x] &= \mathbb{E}[5X + \epsilon \mid X = x]\\ &= \mathbb{E}[5X \mid X = x] + \mathbb{E}[\epsilon \mid X = x] && \text{(by linearity of conditional expectation)}\\ &= 5x + \mathbb{E}[\epsilon \mid X = x]\\ &= 5x + \mathbb{E}[\epsilon] && \text{(by independence of } X \text{ and } \epsilon\text{)}\\ &= 5x\end{aligned}$$

(d)

$$\begin{aligned}\mathbb{E}[Y^3] &= \mathbb{E}[(5X + \epsilon)^3]\\ &= \mathbb{E}[125X^3] + \mathbb{E}[75X^2 \cdot \epsilon] + \mathbb{E}[15X \cdot \epsilon^2] + \mathbb{E}[\epsilon^3]\\ &= 125\mathbb{E}[X^3] + 75\mathbb{E}[X^2 \cdot \epsilon] + 15\mathbb{E}[X \cdot \epsilon^2] + \mathbb{E}[\epsilon^3]\\ &= 125 \int_{-1}^{1} x^3 \cdot \frac{1}{2} dx + 75\mathbb{E}[X^2 \cdot \epsilon] + 15\mathbb{E}[X \cdot \epsilon^2] + \mathrm{Skew}[Z]\\ &= 75\mathbb{E}[X^2 \cdot \epsilon] + 15\mathbb{E}[X \cdot \epsilon^2]\\ &= 75\mathbb{E}[X^2] \cdot \mathbb{E}[\epsilon] + 15\mathbb{E}[X] \cdot \mathbb{E}[\epsilon^2] && \text{(by independence; see A1)}\\ &= 0\end{aligned}$$

(e)

$$\begin{aligned}
\mathrm{Cov}(\epsilon, \epsilon^2) &= \mathbb{E}\big[(\epsilon - \mathbb{E}[\epsilon])(\epsilon^2 - \mathbb{E}[\epsilon^2])\big] \\
&= \mathbb{E}[\epsilon \cdot (\epsilon^2 - 1)] \\
&= \mathbb{E}[\epsilon^3] - \mathbb{E}[\epsilon] \\
&= \mathrm{Skew}[Z] \\
&= 0.
\end{aligned}$$

Although $\mathrm{Cov}(\epsilon, \epsilon^2) = 0$, this *does not* imply that they are independent! For example, consider that

$$\begin{aligned}
P(\epsilon \le 1, \epsilon^2 \le 1) &= P(\epsilon \le 1) \\
&\neq P(\epsilon \le 1) \cdot P(\epsilon^2 \le 1).
\end{aligned}$$

Hence, ϵ and ϵ^2 are not independent.

Appendix

(A1) Let X and Y be independent random variables. Then, by definition, for any (x, y)

$$P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y). \tag{11}$$

Now consider X^2 and Y. We have

$$\begin{aligned}
P(X^2 \le x, Y \le y) &= P(-\sqrt{x} \le X \le \sqrt{x}, Y \le y) \\
&= P(X \le \sqrt{x}, Y \le y) - P(X \le -\sqrt{x}, Y \le y) \\
&= P(X \le \sqrt{x}) \cdot P(Y \le y) - P(X \le -\sqrt{x}) \cdot P(Y \le y) \quad \text{(by independence of } X \text{ and } Y\text{)} \\
&= \big(P(X \le \sqrt{x}) - P(X \le -\sqrt{x})\big) \cdot P(Y \le y) \\
&= P(X^2 \le x) \cdot P(Y \le y).
\end{aligned}$$