Econometrics

Homework 1

Problem 1. Prove the following results:

1. If c is a constant, then $Cov\left(X,c\right)=0$.
2. $Cov\left(X,X\right)=Var\left(X\right).$
3. $Cov\left(X,Y\right)=Cov\left(Y,X\right).$
4. $Cov\left(a_1+b_1X,a_2+b_2Y\right)=b_1b_2Cov\left(X,Y\right),$ where $a_1,a_2,b_1,$ and b_2 are some constants.
5. $Var\left(X+Y\right)=Var\left(X\right)+Var\left(Y\right)+2Cov\left(X,Y\right)$.
6. $Var\left(X-Y\right)=Var\left(X\right)+Var\left(Y\right)-2Cov\left(X,Y\right)$.

Solution. Check any textbook in probability and mathematical statistics.

Problem 2. Let X and Y be two discrete random variables. The set of possible values for X is $\{x_1,\ldots,x_n\}$, and the set of possible values for Y is $\{y_1,\ldots,y_m\}$. The joint PMF is given by

$$p_{ij}^{X,Y}=P\left(X=x_i,Y=y_j\right),\qquad i=1,\ldots n;j=1,\ldots,m.$$

Show that if X and Y are independent then $Cov\left(X,Y\right)=0$.

Solution. Since X and Y are independent, $P\left(X=x_i,Y=y_j\right)=P\left(X=x_i\right)P\left(Y=y_j\right)$ for all i,j. Therefore,

$$\begin{aligned}
E\left[XY\right]=&\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i-E\left[X\right]\right)\left(y_j-E\left[Y\right]\right)P\left(X=x_i,Y=y_j\right)\\
=&\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i-E\left[X\right]\right)\left(y_j-E\left[Y\right]\right)P\left(X=x_i\right)P\left(Y=y_j\right)\\
=&\left(\sum_{i=1}^{n}\left(x_i-E\left[X\right]\right)P\left(X=x_i\right)\right)\left(\sum_{j=1}^{m}\left(y_j-E\left[Y\right]\right)P\left(Y=y_j\right)\right)\\
=&E\left[X\right]E\left[Y\right],
\end{aligned}$$

since

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_j=&a_1\sum_{j=1}^{m}b_j+a_2\sum_{j=1}^{m}b_j+\cdots+a_n\sum_{j=1}^{m}b_j\\
=&\left(a_1+a_2+\cdots+a_n\right)\sum_{j=1}^{m}b_j\\
=&\left(\sum_{i=1}^{n}a_i\right)\left(\sum_{j=1}^{m}b_j\right).
\end{aligned}$$

Problem 3. Let $\{x_i : i = 1, \ldots, n\}$ and $\{y_i : i = 1, \ldots, n\}$ be two sequences. Define the averages

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

1. Show that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

2. Using the result in part (a), show that

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i (x_i - \bar{x}), \text{ and}$$
$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} y_i (x_i - \bar{x}) = \sum_{i=1}^{n} x_i (y_i - \bar{y}).$$

Solution. (a)

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0,$$

because $\sum_{i=1}^{n} x_i = n \cdot \bar{x}$. (b)

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 - \sum_{i=1}^{n} x_i (x_i - \bar{x}) &= \sum_{i=1}^{n} \left[(x_i - \bar{x})^2 - x_i (x_i - \bar{x})\right] \\
&= \sum_{i=1}^{n} \left[\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) - \left(x_i^2 - x_i\bar{x}\right)\right] \\
&= \sum_{i=1}^{n} \left(\bar{x}^2 - x_i\bar{x}\right) \\
&= \bar{x}\sum_{i=1}^{n} (\bar{x} - x_i) \\
&= 0,
\end{aligned}$$

where the last equality follows from $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ proved in (a).

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}(x_i - \bar{x})\, y_i - \sum_{i=1}^{n}(x_i - \bar{x})\, \bar{y} \\
&= \sum_{i=1}^{n}(x_i - \bar{x})\, y_i - \bar{y}\sum_{i=1}^{n}(x_i - \bar{x}) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})\, y_i,
\end{aligned}$$

where the last equality follows from $\sum_{i=1}^{n}(x_i-\bar{x})=0$ proved in (a). The proof of

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(y_i-\bar{y})\,x_i$$

is similar.

Problem 4. The distribution of income X within countries is often modeled as a Pareto distribution with parameter $A>0$ and $\gamma>0$, with CDF

$$F_X(x)=\begin{cases}0 & \text{if } x\le A \text{ and}\\ 1-\left(\frac{x}{A}\right)^{-\gamma}, & \text{if } x>A.\end{cases}$$

Assume $\gamma>1$ and calculate $E[X]$.

Solution. The PDF is $f_X(x)=0$ if $x\le 0$ and $f_X(x)=\gamma x^{-\gamma-1}A^{\gamma}$ if $x>A$. Now, note that

$$\int_{-\infty}^{z} x f_X(x)\,\mathrm{d}x=\int_{A}^{z} x f_X(x)\,\mathrm{d}x=\frac{\gamma}{\gamma-1}\left(A-A^{\gamma}z^{-\gamma+1}\right)$$

converges if $\gamma>1$ and diverges if $\gamma<1$, as $z\to\infty$. So if $\gamma>1$, then $E[X]=\frac{\gamma A}{\gamma-1}$.

Problem 5. Show

$$E[(Y-E[Y|X])(E[Y|X]-E[Y])]=0.$$

Solution. By LIE,

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

Problem 6. Show that if $E[Y|X]=E[Y]$, then $Cov(X,Y)=0$.

Solution. By LIE, $E[XY]=E[E[XY|X]]=E[XE[Y|X]]=E[X\cdot E[Y]]=E[X]E[Y]$.

Problem 7. The conditional PDF of Y given X is

$$f_{Y|X}(y|x)=\frac{y+x}{1/2+x},$$

for $0<y<1$. Find $\mathrm{E}[Y|X=x]$.

Solution.

$$\mathrm{E}[Y|X=x]=\int_0^1\frac{y(y+x)}{1/2+x}\mathrm{d}y=\frac{2+3x}{3+6x}.$$

Problem 8. A Monet expert is given a painting purported to be a lost Monet. He is asked to assess the chances that it is genuine and has the following information:
• In general, only 1% of the "found" paintings he receives turn out to be genuine, an event well call G
• "Found" paintings have a different frequency of use of certain pigments than genuine Monets do: (i) cadmium yellow Y appears in 20% "found" paintings, but only 10% genuine ones (ii) raw umber U appears in 80% of "found" paintings, but only 40% of genuine ones (iii) burnt sienna S appears in 40% of "found" paintings, but 60% of genuine paintings
• This particular painting uses burnt sienna, but not cadmium yellow or raw umber.
What is the probability that this particular painting is genuine? Do we have to make any additional assumptions to answer the question?

Solution. This problem has the following structure: the problem seems to tell us what colors ("data" $\mathrm{S}\cap Y^c\cap U^c$) are how likely to appear given that the painting is genuine ("state of the world" G), i.e. $P(B|A)$. But we actually want to know how likely the painting is genuine given the colors that were used in it, i.e $P(A|B)$. So we are trying to switch the order of conditioning, so we'll try to use Bayes' Theorem. Let's first compile the information contained in the problem:

$$\begin{aligned} P(Y) &= 0.2 \\ P(Y|G) &= 0.1 \\ P(U) &= 0.8 \\ P(U|G) &= 0.4 \\ P(S) &= 0.4 \\ P(S) &= 0.4 \\ P(S|G) &= 0.6 \\ P(G) &= 0.01. \end{aligned}$$

Bayes theorem tells us that

$$P(G|\mathrm{S}\cap Y^c\cap U^c) = \frac{P(\mathrm{S}\cap Y^c\cap U^c|G)\,P(G)}{P(\mathrm{S}\cap Y^c\cap U^c)}.$$

However, know only marginal probability of each color, but would need joint probabilities (both conditional on G and unconditional). Therefore we have to make an additional assumption at this point, and the simplest way to attack this is to assume that the use of pigments is independent across the three colors, both unconditionally and conditional on G, i.e.

$$P(\mathrm{S}\cap Y^c\cap U^c|G) = P(S|G)\,P(Y^c|G)\,P(U^c|G) = 0.6\times 0.9\times 0.6$$

and

$$P(\mathrm{S}\cap Y^c\cap U^c) = P(S)\,P(Y^c)\,P(U^c) = 0.4\times 0.8\times 0.2.$$

Using Bayes' theorem we get therefore that under the independence assumption

$$P(G|\mathrm{S}\cap Y^c\cap U^c) = \frac{0.6\times 0.9\times 0.6\times 0.01}{0.4\times 0.8\times 0.2} = \frac{81}{1600}.$$

To see how much this assumption mattered, can invent a different dependence structure among the different types of pigments: suppose that for genuine Monets, every painting using sienna S also uses umbra U for sure. Then, by the definition of conditional probabilities

$$P(\mathrm{S} \cap Y^c \cap U^c | G) \leq P(S \cap U^c | G) = P(U^c | S \cap G) P(S|G) = 0,$$

so that for a true Monet, it is impossible to find sienna S, but not umbra, therefore we'd know for sure that the painting in questions can't be a Monet (note that since our painting had this combination, it has to be possible for "found" paintings in general). So, summing up, this problem did not give us enough information to answer the question.

Problem 9. For any given two random variables X and Y, we define

$$\mathrm{Var}[Y \mid X] = \mathrm{E}\left[(Y - \mathrm{E}[Y \mid X])^2 \mid X\right].$$

Suppose that $\mathrm{E}[Y \mid X] = 1/4$ and $\mathrm{E}[Y^2 \mid X] = 1/8$. Then show that for any function g, $\mathrm{Var}[Y \mid g(X)] = 1/16$. Use the following facts: for any function g, $\mathrm{E}[\mathrm{E}[Y \mid g(X)] \mid X] = \mathrm{E}[Y \mid g(X)]$ and $\mathrm{E}[\mathrm{E}[Y \mid X] \mid g(X)] = \mathrm{E}[Y \mid g(X)]$.

Solution. By using the fact that $\mathrm{E}[Y \mid g(X)] = \mathrm{E}[\mathrm{E}[Y \mid X] \mid g(X)]$,

$$\begin{aligned}
\mathrm{Var}[Y \mid g(X)] &= \mathrm{E}\left[(Y - \mathrm{E}[Y \mid g(X)])^2 \mid g(X)\right] \\
&= \mathrm{E}\left[(Y - \mathrm{E}[\mathrm{E}[Y \mid X] \mid g(X)])^2 \mid g(X)\right] \\
&= \mathrm{E}\left[\left(Y - \frac{1}{4}\right)^2 \mid g(X)\right] \\
&= \mathrm{E}\left[Y^2 \mid g(X)\right] - \frac{1}{2}\mathrm{E}[Y \mid g(X)] + \mathrm{E}\left[\frac{1}{16} \mid g(X)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[Y^2 \mid X\right] \mid g(X)\right] - \frac{1}{2}\mathrm{E}[\mathrm{E}[Y \mid X] \mid g(X)] + \frac{1}{16} \\
&= \mathrm{E}\left[\frac{1}{8} \mid g(X)\right] - \frac{1}{2}\mathrm{E}\left[\frac{1}{4} \mid g(X)\right] + \frac{1}{16} \\
&= \frac{1}{8} - \frac{1}{8} + \frac{1}{16} \\
&= \frac{1}{16}.
\end{aligned}$$