36-401/607, HW 1 Solutions

September 7, 2022

NOT FOR SHARING, EVEN AFTER THE CLASS IS OVER

To receive full credit, you must show all steps in your work.

1. Suppose that X and Y are any random variables.

(a) Show that Var[X] $= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. (There should be no need to consider the discrete and continuous cases separately.) **(10 points)**

Solution:

$$
\begin{aligned}
Var[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] &= \mathbb{E}\left[\left(X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\right)\right] && \textbf{[unpacking the square]} \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[2X\mathbb{E}[X]\right] + \mathbb{E}\left[(\mathbb{E}[X])^2\right] && \textbf{[linearity of } \mathbb{E}[\cdot] \textbf{(Eqn A.29a)]} \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}\left[X\mathbb{E}[X]\right] + \mathbb{E}\left[(\mathbb{E}[X])^2\right] && \textbf{[by linearity of } \mathbb{E}[\cdot]\textbf{]} \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 && \textbf{[using } \mathbb{E}[X] = \textbf{constant]} \\
&= \mathbb{E}\left[X^2\right] - 2\left(\mathbb{E}[X]\right)^2 + (\mathbb{E}[X])^2 \\
&= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2
\end{aligned}
$$

(b) Show $Cov[Y, Z] = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]$ **(10 points)**

Solution:

$$
\begin{aligned}
Cov[Y, Z] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])\right] \\
&= \mathbb{E}\left[YZ - \mathbb{E}[Y]Z - Y\mathbb{E}[Z] + \mathbb{E}[Y]\mathbb{E}[Z]\right] \\
&= \mathbb{E}[YZ] - \mathbb{E}\left[\mathbb{E}[Y]Z\right] - \mathbb{E}\left[Y\mathbb{E}[Z]\right] + \mathbb{E}\left[\mathbb{E}[Y]\mathbb{E}[Z]\right] && \textbf{[by linearity of } \mathbb{E}[\cdot]\textbf{]} \\
&= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] - \mathbb{E}[Y]\mathbb{E}[Z] + \mathbb{E}[Y]\mathbb{E}[Z] && \textbf{[pulling constants out of } \mathbb{E}[\cdot]\textbf{]} \\
&= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]
\end{aligned}
$$

(c) Use the definition of the covariance to prove that $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$. **(5 points)**

Solution:

$$\begin{aligned}
\text{Cov}(aX, bY) &= \mathbb{E}[abXY] - \mathbb{E}[aX]\mathbb{E}[bY] \\
&= ab\mathbb{E}[XY] - ab\mathbb{E}[X]\mathbb{E}[Y] \\
&= ab\left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right) \\
&= ab\text{Cov}(X, Y)
\end{aligned}$$

2. Let Y_1 and Y_2 be uncorrelated random variables, i.e., their correlation is zero. Consider $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 - Y_2$.

(a) Find $\text{Cov}(U_1, U_2)$ in terms of the variances of Y_1 and Y_2. **(10 points)**

Solution:
Using properties of the covariance given in lecture,

$$\begin{aligned}
\text{Cov}(U_1, U_2) &= \text{Cov}(Y_1 + Y_2, Y_1 - Y_2) \\
&= \text{Cov}(Y_1, Y_1) - \text{Cov}(Y_1, Y_2) + \text{Cov}(Y_2, Y_1) - \text{Cov}(Y_2, Y_2) \\
&= \text{Var}(Y_1) - \text{Var}(Y_2)
\end{aligned}$$

(b) Find an expression for the coefficient of correlation between U_1 and U_2. **(10 points)**

Solution:
Since Y_1 and Y_2 are uncorrelated, we know that

$$\text{Var}(U_1) = \text{Var}(Y_1 + Y_2) = \text{Var}(Y_1) + \text{Var}(Y_2)$$

and

$$\text{Var}(U_2) = \text{Var}(Y_1 - Y_2) = \text{Var}(Y_1) + \text{Var}(Y_2)$$

Hence,

$$\sqrt{\text{Var}(U_1)\text{Var}(U_2)} = \text{Var}(Y_1) + \text{Var}(Y_2)$$

and

$$\text{Corr}(U_1, U_2) = \frac{\text{Cov}(U_1, U_2)}{\sqrt{\text{Var}(U_1)\text{Var}(U_2)}} = \frac{\text{Var}(Y_1) - \text{Var}(Y_2)}{\text{Var}(Y_1) + \text{Var}(Y_2)}$$

(c) Is it possible that $\text{Cov}(U_1, U_2) = 0$? When does this occur? **(5 points)**

Solution:
Yes, it is possible that U_1 and U_2 would be uncorrelated. It would occur any time that $\text{Var}(Y_1) = \text{Var}(Y_2)$.

3. Suppose we generate a random variable Y in the following way. First we flip a fair coin. If the coin is heads, take Y to have a Unif$(0,1)$ distribution. If the coin is tails, take Y to have a Unif$(3,4)$ distribution.

(a) Find the mean of Y. **(10 points)**

Solution:

Let $X \sim$ Bernoulli$(p = 0.5)$, where $X = 1$ if heads, and $X = 0$ if tails. After we observe $X = 1$, we draw $Y|X = 1 \sim$ Unif$(0,1)$, and after we observe $X = 0$, we draw $Y|X = 0 \sim$ Unif$(3,4)$. From the definition of conditional expectation:

$$\begin{aligned} \mathbb{E}[Y|X=1] &= 1/2 \\ \mathbb{E}[Y|X=0] &= 7/2 \end{aligned}$$

Using the "law of total expectation", it follows that:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y|X=1]P(X=1) + \mathbb{E}[Y|X=0]P(X=0) \\ &= (7/2)(1/2) + (1/2)(1/2) = 2 \end{aligned}$$

(b) Find the standard deviation of Y. **(10 points)**

Solution:

From the definition of conditional variance:

$$\mathrm{Var}[Y|X=1] = \mathrm{Var}[Y|X=0] = 1/12$$

Now using the "law of total variance", it follows that:

$$\begin{aligned} \mathrm{Var}[Y] &= \mathbb{E}[\mathrm{Var}[Y|X]] + \mathrm{Var}[\mathbb{E}[Y|X]] \\ &= \mathbb{E}[1/12] + \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[\mathbb{E}[Y|X]]^2 \\ &= 1/12 + (7/2)^2(1/2) + (1/2)^2(1/2) - 2^2 = 7/3 \end{aligned}$$

Thus, the standard deviation is $\sqrt{7/3}$.

4. Let $r(x)$ be a function of x and let $s(y)$ be a function of y. Show that

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

Also, show that $\mathbb{E}(r(X)|X) = r(X)$. (For convenience, you can just consider the continuous case.) **(10 points)**

Solution:

Following the hint, we show that the statements are true for $X = x$, where x is an arbitrary value/realization of X, and then just use the definition of conditional expectations:

$$\begin{aligned} \mathbb{E}[r(X)s(Y)|X = x] &= \int r(x)s(y)f(y|x)dy \\ &= r(x)\int s(y)f(y|x)dy \\ &= r(x)\mathbb{E}[s(Y)|X = x] \end{aligned}$$

If $s(Y) = 1$ then we have $\mathbb{E}[r(X)|X = x] = r(x)\mathbb{E}[1|X = x] = r(x)$.

---

5. Read the following news story from the Washington Post (Childhood exposure to second-hand smoke linked to lung cancer decades later, August 17, 2018) and answer the questions below. `https://wapo.st/2wI3Swb`

(a) What is the response or outcome variable being studied? **(2 points)**

Solution:

Chronic obstructive pulmonary disease, or COPD, in adulthood. (Okay if they just say lung disease)

(b) What is the predictor variable discussed in the news story? **(3 points)**

Solution:

Exposure to second-hand smoke during childhood.

(c) Look at the statistical analysis here:
`https://www.ajpmonline.org/article/S0749-3797(18)31876-2/fulltext`
It's okay if you don't understand the details of the modeling – you're not expected to. What other variables did they include in the model? **(8 points)**

Solution:

gender
race
education
BMI
physical activity
time spent sitting
alcohol use
healthy eating diet score
marital status
(Okay to say age. The paper says they stratified on age, which is one way of controlling

for age effects. If you say prevalent lung or vascular disease, that's wrong–the paper says they did not adjust for those, as they are on the causal pathway)

(d) What was the study design? **(2 points)**

Solution:

Observational

(e) Did the researchers have a causal question in mind? **(5 points)**

Solution:

Yes, it's clear from the article that they care about a causal question. They want to understand if exposure to second hand smoke in childhood causes disease in adults. This is why they control for potential confounders – and also why they did a large, multi-year longitudinal (that is, expensive) study.