STATS 200A: Homework #5

Professor Yingnian Wu
Assignment: 1-5

Eric Chuu
UID: 604406828

November 1, 2016

Problem 1

Prove the following properties about covariance.

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
3. $\text{Cov}(X, Y) = \text{E}(XY) + \text{E}(X)\text{E}(Y)$
4. $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(Y, Z)$
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
6. $\text{Cor}(X, Y) = \text{Cor}(aX + b, cY + d)$, if $ac > 0$
7. Let $X \sim \text{Unif}[-1, 1]$ and $Y = X^2$. Show that $\text{Cov}(X, Y) = 0$
8. If $(X, Y) \sim f(x, y) = f_X(x)f_Y(y)$. Then $\text{Cov}(X, Y) = 0$

Solution
(1) $\text{Cov}(X, X) = \text{E}[(X - \text{E}[X])(X - \text{E}[X]] = \text{E}[(X - \text{E}[X])^2] = \text{Var}(X)$
(2) Let $\mu_X = \text{E}[X], \mu_Y = \text{E}[Y], \mu_{X+Y} = \text{E}[X] + \text{E}[Y]$. Then, using linearity of expectation, we have

$$\begin{aligned}
\text{Var}(X + Y) &= \text{E}[((X + Y) - \mu_{X+Y})^2] \\
&= \text{E}[((X - \mu_X) + (Y - \mu_y))^2] \\
&= \text{E}[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)] \\
&= \text{E}[(X - \mu_X)^2] + \text{E}[(Y - \mu_Y)^2] + 2\text{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
\end{aligned}$$

(3) Continuing with the variable definitions from (2) and using linearity of expectation, we have

$$\begin{aligned}
\text{Cov}(X, Y) &= \text{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \text{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\
&= \text{E}[XY] - \text{E}[X]\text{E}[Y] - \text{E}[Y]\text{E}[X] + \text{E}[X]\text{E}[Y] \\
&= \text{E}[XY] - \text{E}[X]\text{E}[Y]
\end{aligned}$$

(4)

$$\begin{aligned}
\text{Cov}(aX+b, cY+d) &= \text{E}[(aX+b-\text{E}[aX+b])(cY+d-\text{E}[cY+d])] \\
&= \text{E}[(aX+b-a\text{E}[X]-b)(cY+d-c\text{E}[X]-d)] \\
&= \text{E}[a(X-\text{E}[x])\cdot c(Y-\text{E}[Y])] \\
&= ac\cdot\text{E}[(X-\text{E}[X])(Y-\text{E}[Y])] \\
&= ac\cdot\text{Cov}(X,Y)
\end{aligned}$$

(5)

$$\begin{aligned}
\text{Cov}(X+Y, Z) &= \text{E}[(X+Y=\text{E}[X+Y])(Z-\text{E}[Z])] \\
&= \text{E}[(X-\text{E}[X]+Y-\text{E}[Y])(Z-\text{E}[Z])] \\
&= \text{E}[(X-\text{E}[X])(Z-\text{E}[Z])+(Y-\text{E}[Y])(Z-\text{E}[Z])] \\
&= \text{E}[(X-\text{E}[X])(Z-\text{E}[Z])]+\text{E}[(Y-\text{E}[Y])(Z-\text{E}[Z])] \\
&= \text{Cov}(X,Z)+\text{Cov}(Y,Z)
\end{aligned}$$

(6) If $ac > 0$, then we can use the definition of correlation and part (4) to express the right hand side as

$$\text{Cor}(aX+b, cY+d) = \frac{\text{Cov}(aX+b, cY+d}{\sqrt{\text{Var}(aX+b)}\sqrt{\text{Var}(cY+d)}} = \frac{ac\cdot\text{Cov}(X,Y)}{ac\cdot\sqrt{Var(X)}\sqrt{\text{Var}(Y)}}\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \text{Cor}(X,Y)$$

(7)

$$\begin{aligned}
\text{Cov}(X,Y) = \text{Cov}(X,X^2) &= \text{E}[(X-\text{E}[X])(X^2-\text{E}[X^2])] \\
&= \text{E}\left(X^3 - X\text{E}[X^2] - X^2\text{E}[X] + \text{E}[X]\text{E}[X^2]\right) \\
&= \text{E}[X^3] - \text{E}[X]\text{E}[X^2] - \text{E}[X^2]\text{E}[X] + \text{E}[X]\text{E}[X^2] \\
&= \text{E}[X^3] - \text{E}[X]\text{E}[X^2] \\
&= \int_{-1}^{1}\frac{1}{2}x^3dx - \int_{-1}^{1}\frac{1}{2}xdx\cdot\int_{-1}^{1}\frac{1}{2}x^2dx \\
&= 0 - 0\cdot\frac{1}{3} \\
&= 0
\end{aligned}$$

(8) With Fubini's Theorem, we can perform the following swaps on the integrals

$$\begin{aligned}
\text{Cov}(X,Y) &= \text{E}[XY] - \text{E}[X]\text{E}[Y] \\
&= \int\int xyf(x,y)dxdy - \int xf_X(x)dx\int yf_Y(y)dy \\
&= \int\int xyf_X(x)f_Y(y)dxdy - \int xf_X(x)\int yf_Y(y)dy \\
&= \int xf_X(x)dx\int yf_Y(y)dy - \int xf_X(x)\int yf_Y(y)dy \\
&= 0
\end{aligned}$$

All 8 properties have thus been proven. $\square$

Problem 2

For two random variables X and Y, suppose we want to predict Y by $\alpha + \beta X$ by minimizing $R(\alpha, \beta) = \mathrm{E}[(Y - \alpha\beta X)^2]$. Let $\epsilon = Y - \alpha - \beta X$.

1. Take derivative of R with respect to α, show that $\mathrm{E}[\epsilon] = 0$. Assume that we can exchange the derivative and the expectation.
2. Take the derivative of R with respect to β, show that $\mathrm{Cov}(X, \epsilon) = 0$
3. Based on (1) and (2), solve for the optimal α and β based on $\mu_X = \mathrm{E}[X], \mu_Y = \mathrm{E}[Y], \mathrm{Cov}(X,Y), \sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y)$.
4. Express β in terms of $\rho = \mathrm{Cor}(X, Y)$, as well as σ_x and σ_Y.
5. Prove $\mathrm{Var}(Y) = \mathrm{Var}(\alpha + \beta X) + \mathrm{Var}(\epsilon)$. Let $R^2 = \mathrm{Var}(\alpha + \beta X)/\mathrm{Var}(Y)$. Show that $R^2 = \rho^2$

Solution

(1) Taking the derivative of R with respect to α and setting it equal to 0, we get

$$\frac{\partial R(\alpha, \beta)}{\partial \alpha} = -2\mathrm{E}[Y - \alpha - \beta X] = -2\mathrm{E}[\epsilon] = 0 \implies \mathrm{E}[\epsilon] = 0$$

(2) Taking the derivative of R with respect to β, we get

$$\frac{\partial R(\alpha, \beta)}{\partial \beta} = -2\mathrm{E}[(Y - \alpha - \beta X) \cdot X] = -2\mathrm{E}[\epsilon X] = 0$$

This implies that $\mathrm{E}[\epsilon X] = 0$, and if we then consider the covariance of X and ϵ, then we have

$$\mathrm{Cov}(X, \epsilon) = \mathrm{E}[\epsilon X] - \mathrm{E}[X]\mathrm{E}[\epsilon] = 0 - 0 \cdot \mathrm{E}[X] = 0$$

(3) Using part (1), we see that $\mathrm{E}[Y - \alpha - \beta X] = 0$, so

$$\mathrm{E}[Y] - \alpha - \beta\mathrm{E}[X] = \mu_y - \alpha - \beta\mu_X = 0$$

Solving for the optimal α, we get $\alpha = \beta\mu_X + \mu_y$, where the value of β can be optimally chosen by solving for β given in part (2). Since $\mathrm{Cov}(\epsilon, X) = 0$, we can expand and solve for β:

$$\begin{aligned} 0 &= \mathrm{Cov}(\epsilon, X) = \mathrm{Cov}(Y - \alpha - \beta X, X) = \mathrm{Cov}(Y, X) - \beta\mathrm{Cov}(X, X) \\ &\Rightarrow \mathrm{Cov}(X, Y) = \beta\mathrm{Cov}(X, X) \\ &\Rightarrow \beta = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Cov}(X, X)} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X^2} \end{aligned}$$

(4) Since $\rho = \mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/\sigma_X\sigma_Y$, then $\mathrm{Cov}(X, Y) = \rho\sigma_X\sigma_Y$, and we can express β as found in the previous part in terms of ρ, σ_X, σ_Y as follows

$$\beta = \frac{\mathrm{Cov}(X, Y)}{\sigma_X^2} = \frac{\rho\sigma_X\sigma_Y}{\sigma_X^2} = \rho\frac{\sigma_Y}{\sigma_X}$$

(5) We use property (2) and (5) from problem 1 to show the following:

$$\begin{aligned} \mathrm{Var}(Y) = \mathrm{Var}(\alpha + \beta X + \epsilon) &= \mathrm{Var}(\alpha + \beta X) + \mathrm{Var}(\epsilon) + 2\mathrm{Cov}(\alpha + \beta X, \epsilon) \\ &= \mathrm{Var}(\alpha + \beta X) + \mathrm{Var}(\epsilon) + 2 \cdot \beta^2\mathrm{Cov}(X, \epsilon) \\ &= \mathrm{Var}(\alpha + \beta X) + \mathrm{Var}(\epsilon) + 2 \cdot 0 \\ &= \mathrm{Var}(\alpha + \beta X) + \mathrm{Var}(\epsilon) \end{aligned}$$

Then we use the value of β as calculated in part (4) to show that

$$R^2 = \frac{\mathrm{Var}(\alpha + \beta X)}{\mathrm{Var}(Y)} = \frac{\beta^2\mathrm{Var}(X)}{\mathrm{Var}(Y)} = \rho^2\frac{\mathrm{Var}(Y)}{\mathrm{Var}(X)} \cdot \frac{\mathrm{Var}(X)}{\mathrm{Var}(Y)} = 1$$

$\square$

Problem 3

Define the terms and prove the following:

1. $\text{E}[\text{E}[Y|X]] = \text{E}[Y]$
2. $\text{Var}(Y) = \text{Var}(\text{E}(Y|X)) + \text{E}[\text{Var}(Y|X)]$
3. $\text{Cov}(X,Y) = \text{E}[\text{Cov}(X,Y|Z)] + \text{Cov}(\text{E}[X|Z].\text{E}[Y|Z]])$

Solution
Let X, Y be random variables, where $Y \sim f_{Y|X}(y|x)$, and we define

$$h(X) := \text{E}[Y|X=x] = \int y f_{Y|X}(y|x) dy \tag{1}$$

(1) Using the definition of expectation, conditional probability, and transformation defined in (1) above, we can calculate $\text{E}[\text{E}[Y|X]]$,

$$\begin{aligned}
\text{E}[\text{E}[Y|X]] = \text{E}[h(X)] &= \int h(x) f_X(x) dx \\
&= \int \left[\int y f_{Y|X}(y|x) dy \right] f_X(x) dx \\
&= \int \int y f(x,y) dx dy \\
&= \text{E}[Y]
\end{aligned}$$

(2) First recall that conditional variance is defined as

$$\text{Var}(Y|X=x) = \text{E}[(Y - h(X))^2 | X = x] \tag{2}$$

Define $\epsilon := Y - h(X)$. Consider

$$\begin{aligned}
\text{E}[\epsilon|X] = \text{E}[Y - h(X)|X] &= \text{E}[Y - \text{E}[Y|X]|X] && (3) \\
&= \text{E}[Y|X] - \text{E}[Y|X] && (4) \\
&= 0 && (5)
\end{aligned}$$

where the equality in (5) follows from the result proven in part (1). Then for any function g defined on X, we can again use the result from part (1) to calculate the following,

$$\text{E}[\epsilon g(X)] = \text{E}[\text{E}[\epsilon g(X)|X]] = \text{E}[g(X)\text{E}[\epsilon|X]] = \text{E}[g(X) \cdot 0] = 0 \tag{6}$$

Using the conclusions from (6) and(7), we then have

$$\text{Cov}(\epsilon, h(X)) = \text{E}[\epsilon h(X)] - \text{E}[\epsilon]\text{E}[h(X)] = 0 - 0 \cdot \text{E}[h(X)] = 0 \tag{7}$$

Then the variance of Y can be expressed as

$$\begin{aligned}
\text{Var}(Y) = \text{Var}(h(X) + \epsilon) &= \text{Var}(h(X)) + \text{Var}(\epsilon) + 2\text{Cov}(h(X), \epsilon) && (8) \\
&= \text{Var}(h(X)) + \text{Var}(\epsilon) && (9) \\
&= \text{Var}(\text{E}[Y|X]) + \text{E}[(\epsilon - \text{E}[\epsilon])^2] && (10) \\
&= \text{Var}(\text{E}[Y|X]) + \text{E}[\epsilon^2] && (11) \\
&= \text{Var}(\text{E}[Y|X]) + \text{E}[(Y - h(X))^2] && (12) \\
&= \text{Var}(\text{E}[Y|X]) + \text{E}[\text{E}[(Y - h(X))^2|X]] && (13) \\
&= \text{Var}(\text{E}[Y|X]) + \text{E}[\text{Var}(Y|X)] && (14)
\end{aligned}$$

where equality (10) follows from (8), equality (12) follows from (6), equality (14) follows from the result of part (1), and equality (15) follows from the definition of conditional variance.

(3) Let X, Y, Z be random variables, and define $h(Z) = \mathrm{E}[X|Z]$ and $g(Z) = \mathrm{E}[Y|Z]$. Recall the definition of conditional covariance

$$\mathrm{Cov}(X, Y|Z) = \mathrm{E}[(X - \mathrm{E}[X|Z])(Y - \mathrm{E}[Y|Z])|Z] \tag{15}$$

In addition, we define $\epsilon := X - h(Z)$ and $\delta = Y - g(Z)$. Consider the following intermediate calculations:

$$\begin{aligned}
\mathrm{E}[\epsilon] = \mathrm{E}[X - h(Z)] &= \mathrm{E}[\mathrm{E}[(X - h(Z)|Z)]] && (16)\\
&= \mathrm{E}[\mathrm{E}[X|Z] - \mathrm{E}[h(Z)]] && (17)\\
&= \mathrm{E}[\mathrm{E}[X|Z] - \mathrm{E}[X|Z]] && (18)\\
&= \mathrm{E}[0] && (19)\\
&= 0 && (20)
\end{aligned}$$

By a similar calculation, $\mathrm{E}[\delta] = 0$. For any function g defined on Z,

$$\begin{aligned}
\mathrm{E}[\epsilon g(Z)] &= \mathrm{E}[\mathrm{E}[\epsilon g(Z)|Z]] && (21)\\
&= \mathrm{E}[g(Z)\mathrm{E}[\epsilon|Z]] && (22)\\
&= \mathrm{E}[g(Z) \cdot 0] && (23)\\
&= 0 && (24)
\end{aligned}$$

By a similar calculation, we also have that for any function g defined on Z, $\mathrm{E}[\delta g(Z)] = 0$. We can then calculate the covariance of ϵ and δ,

$$\begin{aligned}
\mathrm{Cov}(\epsilon, \delta) &= \mathrm{Cov}(X - h(Z), Y - g(Z)) && (25)\\
&= \mathrm{E}[(X - h(Z) - \mathrm{E}[X] + \mathrm{E}[h(Z)])(Y - g(Z) - \mathrm{E}[Y] + \mathrm{E}[g(Z)])] && (26)\\
&= \mathrm{E}[(X - h(Z))(Y - g(Z))] && (27)\\
&= \mathrm{E}[\mathrm{E}[(X - h(Z))(Y - g(Z))|Z]] && (28)\\
&= \mathrm{E}[\mathrm{Cov}(X, Y|Z)] && (29)
\end{aligned}$$

note that the equality in (29) follows from Adam Law, as proved in part (1). Then, if we consider the the conclusions in lines (21), (25), and (30), we can express the covariance of X, Y as

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(h(Z) + \epsilon, g(Z) + \delta) && (30)\\
&= \mathrm{Cov}(h(Z), g(Z) + \delta) + \mathrm{Cov}(\epsilon, g(Z) + \delta) && (31)\\
&= \mathrm{Cov}(h(Z), g(Z)) + \mathrm{Cov}(h(Z), \delta) + \mathrm{Cov}(\epsilon, g(Z)) + \mathrm{Cov}(\epsilon, \delta) && (32)\\
&= \mathrm{Cov}(h(Z), g(Z)) + \mathrm{Cov}(\epsilon, \delta) && (33)\\
&= \mathrm{Cov}(\mathrm{E}[X|Z], \mathrm{E}[Y|Z]) + \mathrm{E}[\mathrm{Cov}(X, Y|Z)] && (34)
\end{aligned}$$

All three properties have thus been proven. $\square$

Problem 4

Let $Z \sim N(\mu, \sigma^2)$. Let $X = Z + \epsilon_1$ and $Y = Z + \epsilon_2$, where ϵ_1, ϵ_2 follow $N(0, \tau^2)$ independently, and both are independent of Z. Calculate $\mathrm{E}[X], \mathrm{Var}[X]$, and $\mathrm{Cov}(X, Y)$ using the formulas in Problem 1.

Solution
Since Z is normally distributed with mean μ, and ϵ_1 is normally distributed with mean 0, then we can use linearity of expectation to find the expectation of X

$$\mathrm{E}[X] = \mathrm{E}[Z + \epsilon_1] = \mathrm{E}[Z] + \mathrm{E}[\epsilon_1] = \mu + 0 = \mu$$

Using property (2) from Problem 1, we can find the variance of X,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(Z + \epsilon_1) \\
&= \mathrm{Var}(Z) + \mathrm{Var}(\epsilon_1) + 2\mathrm{Cov}(Z, \epsilon_1) \\
&= \sigma + \tau + 0 \\
&= \sigma + \tau
\end{aligned}$$

where $\mathrm{Cov}(Z, \epsilon_1) = 0$ because ϵ_1 is independent of Z. Then we can apply property (6) twice to find the covariance of X, Y,

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(Z + \epsilon_1, Z + \epsilon_2) \\
&= \mathrm{Cov}(Z, Z + \epsilon_2) + \mathrm{Cov}(\epsilon_1, Z + \epsilon_2) \\
&= \mathrm{Cov}(Z, Z) + \mathrm{Cov}(Z, \epsilon_2) + \mathrm{Cov}(\epsilon_1, Z) + \mathrm{Cov}(\epsilon_1, \epsilon_2) \\
&= \mathrm{Var}(Z) + 0 + 0 + 0 \\
&= \mathrm{Var}(Z)
\end{aligned}$$

since Z is independent of both ϵ_1, ϵ_2, and ϵ_1 is independent of ϵ_2. $\square$

Problem 5

For a random person, let X be a whether he or she smokes or not, let Z be the background variable, and let Y be the health of this person. Is $\mathrm{E}[Y|X=1]-\mathrm{E}[Y|X=0]$ equal to $\mathrm{E}[\mathrm{E}[Y|X=0,Z]]$, where the outer expectation is with respect to the marginal distribution of Z? When are the two equal to each other?

Solution