Economics 361
Problem Set #3 (Suggested Answers)

Jun Ishii *
Department of Economics
Amherst College

Fall 2022

Question 1: "Moments" of Truth

(a) Let $Z \equiv aX + bY + c$. Show that the variance of Z is equal to $a^2\ \sigma_X^2 + b^2\ \sigma_Y^2 + 2ab\ \sigma_{XY}$

ANS:

$$\begin{aligned} \text{Var}(Z) &= E[\ (\ aX + bY + c - \underbrace{E[aX + bY + c]}_{=a\mu_X + b\mu_Y + c}\ )^2\ ] \\ &= E[\ \ (\ a(X - \mu_X) + b(Y - \mu_Y)\ )^2\ \ ] \\ &= E[\ a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)\ ] \\ &= a^2\ \sigma_X^2 + b^2\ \sigma_Y^2 + 2ab\ \sigma_{XY} \end{aligned}$$

(b) A *proportional* predictor of Y given X is a predictor of the form $\hat{Y}(X) = cX$. Show that the proportional predictor that minimizes the mean squared error (MSE) risk function, $BPP_{MSE}(Y|X)$, is c^*X where $c^* = \frac{E[XY]}{E[X^2]}$

ANS: The proof is similar to the proof for $BLP(Y|X)$. Start with the first order condition, $\frac{d}{dc}R[LF(cX)] = 0$

$$\begin{aligned} \frac{d}{dc}R[LF(cX)] \ = \ \frac{d}{dc}\ E[\ (cX - Y)^2\ ] &= 0 \\ \frac{d}{dc}\ E[\ c^2X^2 - 2cXY + Y^2\ ] &= 0 \\ \frac{d}{dc}\ \left(c^2E[X^2] - 2cE[XY] + E[Y^2]\right) &= 0 \\ 2cE[X^2] - 2E[XY] &= 0 \\ c^* &= \frac{E[XY]}{E[X^2]} \end{aligned}$$

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

For **(c)-(e)**, suppose that X and Y are jointly distributed **bivariate Normal**:

$$f_{XY}(x,y) \quad = \quad \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}$$

where ρ is another parameter of the joint distribution, along with $\mu_X, \mu_Y, \sigma_X, \sigma_Y$.

(c) Show that the marginal distribution of X is Normal with mean μ_X and variance σ_X^2

ANS: The joint distribution of X and Y, given above, can be re-expressed as

$$\begin{aligned}
f_{XY}(x,y) \quad &= \quad \frac{1}{\sqrt{2\pi}\sigma_X} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_X}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right\} \\
&= \quad \underbrace{\frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_X}\right)^2\right\}}_{\text{(A)}} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right\}}_{\text{(B)}}
\end{aligned}$$

The joint distribution can be expressed as the product of (A) the distribution for a Normal(μ_X, σ_X^2) and (B) the distribution for a Normal$(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2))$. The first part of this product does not vary with y. Taking advantage of these two observations

$$f_X(x) \quad = \quad \int_{-\infty}^{+\infty} f_{XY}(x,y)\ dy \quad = \quad \int_{-\infty}^{+\infty} (A)\cdot(B)\ dy \quad = \quad (A)\cdot\underbrace{\int_{-\infty}^{+\infty}(B)\ dy}_{=1} \quad = \quad (A)$$

Note that integrating the pdf for a Normal distribution from $-\infty$ to $+\infty$ yields 1.

(d) Show that $\text{Cov}(X,Y) = \rho\ \sigma_X\ \sigma_Y$. (Alternatively, that the Correlation$(X,Y) = \rho$)

ANS: The marginal distribution for Y can be shown to be Normal with mean μ_Y and σ_Y^2 using similar steps as in **(c)**. Therefore, $E[X] = \mu_X$ and $E[Y] = \mu_Y$. From the expectations handout, $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$. We need only derive $E[XY]$

$$\begin{aligned}
E[XY] \quad &= \quad E_X[\ E_{Y|X}[XY] \qquad \text{by Law of Iterated Expectations} \\
&= \quad E_X[\ X\ E_{Y|X}[\ Y\ ]\ ] \qquad \text{as conditioning on } X \text{ makes it a constant} \\
&= \quad E_X[\ X\ (\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X))\ ] \\
&\qquad\qquad \text{as } f_{Y|X} = \tfrac{f_{XY}}{f_X} = \tfrac{(A)\cdot(B)}{(A)} = (B) \text{ and thus } E_{Y|X}[Y] = E[Y|X] = \mu_Y + \rho\tfrac{\sigma_Y}{\sigma_X}(X-\mu_X) \\
&= \quad E[X]\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(E[X^2] - E[X]\mu_X) \\
&= \quad \mu_X\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\sigma_X^2 \qquad \text{as } E[X^2] - (E[X])^2 = \text{Var}(X) = \sigma_X^2 \\
&= \quad \mu_X\mu_Y + \rho\sigma_Y\sigma_X
\end{aligned}$$

and therefore

$$\text{Cov}(X,Y) \;=\; E[XY]-E[X]E[Y] \;=\; \mu_X\mu_Y+\rho\,\sigma_Y\,\sigma_X-\mu_X\mu_Y \;=\; \rho\,\sigma_Y\,\sigma_X$$

(e) Many empirical economics studies assume that the relevant conditional expectations function, $E[Y|X]$, is linear in X. Critics have commented that this practice makes sense when the joint distribution of (X, Y) is believed to be bivariate Normal but less so for other distributions. Explain.

ANS: We showed earlier that the conditional distribution of Y given X when the two random variables are jointly distributed bivariate Normal is Normal with the following conditional mean

$$E[Y|X] \;=\; \mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$$

The above conditional mean is linear in X. However, when the joint distribution of X and Y is not bivariate Normal, there is no guarantee that $E[Y|X]$ will be linear in X.

For **(f)-(h)**, suppose X and W are independent random variables with

$$E[X] = 0 \ , \ E[X^2] = 1 \ , \ E[X^3] = 0 \ , \ E[W] = 1 \ , \ E[W^2] = 2$$

Let $Y \equiv W + WX^2$. (This problem borrows from Goldberger Exercise 6.7)

(f) Find the $BP_{MSE}(Y|X)$ and $BLP_{MSE}(Y|X)$

ANS: From independence, $E[W|X] = E[W]$ and $E[WX^2|X] = E[W]E[X^2]$

$$\begin{aligned}
BP_{MSE}(Y|X) &= E[Y|X] \quad = \quad E[W + WX^2 \mid X] \quad = \quad E[W|X] + E[W|X] \ X^2 \quad = \quad 1 + X^2 \\
\mathrm{Cov}(X,Y) &= E[XY] - E[X]E[Y] \quad = \quad E[X(W + WX^2] - E[X] \ E[W + WX^2] \\
&= E[X]E[W] + E[W]E[X^3] - E[X]E[W] - E[X]E[W]E[X^2] \\
&= 0 \cdot 1 + 1 \cdot 0 - 0 \cdot 1 - 0 \cdot 1 \cdot 1 \quad = \quad 0 \\
\mathrm{Var}(X,Y) &= E[X^2] - (E[X])^2 \quad = \quad 1 - 0 \quad = \quad 1 \\
E[Y] &= E[W + WX^2] \quad = \quad E[W] + E[W]E[X^2] \quad = \quad 1 + 1 \cdot 1 \quad = \quad 2 \\
\beta^* &= \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \quad = \quad \frac{0}{1} \quad = \quad 0 \\
\alpha^* &= E[Y] - \beta^* E[X] \quad = \quad 2 - 0 \cdot 0 \quad = \quad 2 \\
BLP_{MSE}(Y|X) &= \alpha^* + \beta^* X \quad = \quad 2 + 0 \cdot X \quad = \quad 2
\end{aligned}$$

(g) Change the assumption $E[X^3] = 0$ to $E[X^3] = 1$. Now find the $BP_{MSE}(Y|X)$ and $BLP_{MSE}(Y|X)$

ANS: $BP_{MSE}(Y|X)$ remains the same as its value does not depend on $E[X^3]$. However, $\mathrm{Cov}(X,Y)$ now equals 1. This results in $\beta^* = 1$ and $BLP_{MSE}(Y|X) = 2+X$ (α^* remains the same as $E[X] = 0$)

(h) Which relation remained the same in going from **(f)** to **(g)**? Which changes? Why?
(Do not simply state "because $E[X^3]$ changed ...")

ANS: The change from $E[X^3] = 0$ to $E[X^3] = 1$ leads to $\mathrm{Cov}(X,Y)$ no longer being equal to zero. This in turns shift β^* away from zero to 1, thus changing the $BLP_{MSE}(Y|X)$ from 2 to $2 + X$ But why should $E[X^3]$ have such an impact on β^*?

Recall our discussion about moments. Specifically, the third moment represents skewness (asymmetry) of the distribution. $E[X^3] = 0$ and $E[X] = 0$, combined, imply that the distribution is symmetric around zero; errors associated with negative values of X are weighed similarly as errors associated with positive values of X. But a linear function of X with a non-zero slope will necessarily lead to different errors for positive values of X than for associated negative values of X (for a given Y value). Minimizing the sum of squared errors implies that we want the error associated with $X = +c$ to be the same as the error associated with $X = -c$ (as $X = +c$ and $X = -c$ are equally weighted). This indicates that BLP_{MSE} should be a constant (zero slope). But when $E[X^3] = 1$, the distribution is skewed positively, positive values of X are more likely than the corresponding negative values. In which case, we want BLP to be positively sloped. Hence, BLP_{MSE} is positively sloped when $E[X^3] = 1$.

Question 2: Curved Roof

These questions are modified versions of Exercises 4.1 and 5.1 in Goldberger's *A Course in Econometrics* textbook.

Consider the following joint pdf for continuous random variables X and Y

$$f(x,y) \;=\; \begin{cases} \frac{3}{11}\left(x^2+y\right) & \text{for } 0 \le x \le 2 \text{ and } 0 \le y \le 1 \\ 0 & \text{for all other values of } x \text{ and } y \end{cases}$$

(a) Show that the above joint pdf does not violate $P(S) = 1$. i.e. probability over possible joint realizations "sum up" to 1.

ANS: Just do the integration

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\;dx\;dy &= \int_0^1\int_0^2 \frac{3}{11}\left(x^2+y\right)\;dx\;dy \\ &= \int_0^1 \frac{3}{11}\left(\frac{1}{3}x^3+xy\right)\Big|_0^2\;dy \;=\; \int_0^1 \frac{3}{11}\left(\frac{8}{3}+2y-0\right)\;dy \\ &= \frac{3}{11}\left(\frac{8}{3}y+y^2\right)\Big|_0^1 \;=\; \frac{3}{11}\left(\frac{8}{3}+1\right) \;=\; 1 \end{aligned}$$

(b) Derive the marginal pdf of X and of Y

ANS:

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} f(x,y)\;dy \;=\; \int_0^1 \frac{3}{11}\left(x^2+y\right)\;dy \\ &= \frac{3}{11}\left(x^2y+\frac{1}{2}y^2\right)\Big|_0^1 \;=\; \frac{3}{11}\left(x^2+\frac{1}{2}\right) \quad \text{for } 0 \le x \le 2 \text{ and } 0 \text{ otherwise} \\ f(y) &= \int_{-\infty}^{\infty} f(x,y)\;dx \;=\; \int_0^2 \frac{3}{11}\left(x^2+y\right)\;dx \\ &= \frac{3}{11}\left(\frac{1}{3}x^3+xy\right)\Big|_0^2 \;=\; \frac{3}{11}\left(\frac{8}{3}+2y\right) \quad \text{for } 0 \le y \le 1 \text{ and } 0 \text{ otherwise} \end{aligned}$$

(c) Derive the conditional pdf of Y given X for $0 \le x \le 2$.

ANS:

$$f(y|x) \;=\; \frac{f(x,y)}{f(x)} \;=\; \begin{cases} \frac{\frac{3}{11}\left(x^2+y\right)}{\frac{3}{11}\left(x^2+\frac{1}{2}\right)} = \frac{x^2+y}{x^2+\frac{1}{2}} & \text{for } 0 \le y \le 1 \\ 0 & \text{for all other } y \end{cases}$$

(d) Calculate the following moments: $E[X], E[Y], E[X^2], E[Y^2], E[XY]$.

$$
\begin{aligned}
E[X] &= \int_0^2 x\ f(x)\ dx &=& \int_0^2 \frac{3}{11}\left(x^3+\frac{1}{2}x\right)\ dx &=& \frac{3}{11}\left(\frac{1}{4}x^4+\frac{1}{4}x^2\right)\Big|_0^2 &=& \frac{15}{11} \\
E[Y] &= \int_0^1 y\ f(y)\ dy &=& \int_0^1 \frac{3}{11}\left(\frac{8}{3}y+2y^2\right)\ dy &=& \frac{3}{11}\left(\frac{4}{3}y^2+\frac{2}{3}y^3\right)\Big|_0^1 &=& \frac{6}{11} \\
E[X^2] &= \int_0^2 x^2\ f(x)\ dx &=& \int_0^2 \frac{3}{11}\left(x^4+\frac{1}{2}x^2\right)\ dx &=& \frac{3}{11}\left(\frac{1}{5}x^5+\frac{1}{6}x^3\right)\Big|_0^2 &=& \frac{116}{55} \\
E[Y^2] &= \int_0^1 y^2\ f(y)\ dy &=& \int_0^1 \frac{3}{11}\left(\frac{8}{3}y^2+2y^3\right)\ dy &=& \frac{3}{11}\left(\frac{8}{9}y^3+\frac{1}{2}y^4\right)\Big|_0^1 &=& \frac{25}{66} \\
E[XY] &= \int_0^1\int_0^2 xy\ f(x,y)\ dx\ dy &=& \int_0^1\int_0^2 \frac{3}{11}\left(x^3y+xy^2\right)\ dx\ dy \\
&= \int_0^1 \frac{3}{11}\left(\frac{1}{4}x^4y+\frac{1}{2}x^2y^2\right)\Big|_0^2\ dy &=& \frac{3}{11}\left(2y^2+\frac{2}{3}y^3\right)\Big|_0^1 &=& \frac{8}{11}
\end{aligned}
$$

(e) Find the best predictor of Y given X under MSE: $BP_{MSE}(Y|X)$

ANS:

$$
BP_{MSE}(Y|X) = E[Y|X] = \int_0^1 y\ f(y|x)\ dy = \int_0^1 \frac{x^2y+y^2}{x^2+\frac{1}{2}}\ dy = \frac{\frac{1}{2}x^2+\frac{1}{3}}{x^2+\frac{1}{2}}
$$

(f) Find the best linear predictor of Y given X under MSE: $BLP_{MSE}(Y|X)$

ANS:

$$
\begin{aligned}
BLP_{MSE}(Y|X) &= a^*+b^*X = \left(E[Y]-\frac{\sigma_{XY}}{\sigma_X^2}E[X]\right)+\frac{\sigma_{XY}}{\sigma_X^2}X \\
&= \frac{6}{11}-\frac{\frac{8}{11}-(\frac{6}{11}\frac{15}{11})}{\frac{116}{55}-\frac{15}{11}\frac{15}{11}}\left(\frac{15}{11}-X\right) \\
&\approx 0.636-0.066\ X
\end{aligned}
$$

(g) Figure 5.2 (p.55) of the Goldberger text shows $BLP_{MSE}(Y|X)$ closer to $BP_{MSE}(Y|X)$ at high, rather than low, values of x. Explain why.

> **HINT:** Goldberger suggests that you may "see" the answer in Figure 4.5 (p.42), which graphs the marginal pdf of X, $f(x)$

ANS: There are several possible explanations. The one that Goldberger hints at centers on the view of the BLP as a linear approximation to the BP (see Chapter 5.5). As Figure 4.5 shows, much of the probability mass is in the higher values of x. Therefore, it is not surprising to find the BLP fitting the BP more closely for the higher values of x at the expense of the lower values; the probability weights assigned to the loss at the higher valued x are higher – so the errors of the higher valued x contribute more to the mean loss.

Question 3: Patent Race, Part II

(a) For each of the four possible actions for the incumbent,(file now, file in $t+1$, file in $t+2$, file in $t+3$), calculate the expected value of the profit the incumbent would earn, in terms of θ_E. **HINT**: the relevant random variable here is the innovation success/failure of the entrant

ANS: Consider the possible payoffs for each of the four actions

- **File Now**: \$500 million with certainty
- **File t+1**:
  - Entrant fails at $t+1$ (\$600 million) with probability $(1-\theta_E)$
  - Entrant succeeds at $t+1$ (\$0) with probability θ_E
- **File t+2**:
  - Entrant fails at $t+1$ and $t+2$ (\$700 million) with probability $(1-\theta_E)^2$
  - Entrant succeeds at $t+2$ (\$100 million) with probability $(1-\theta_E)\theta_E$
  - Entrant succeeds at $t+1$ (\$0) with probability θ_E
- **File t+3**:
  - Entrant fails at $t+1$ through $t+3$ (\$800 million) with probability $(1-\theta_E)^3$
  - Entrant succeeds at $t+3$ (\$200 million) with probability $(1-\theta_E)^2\theta_E$
  - Entrant succeeds at $t+2$ (\$100 million) with probability $(1-\theta_E)\theta_E$
  - Entrant succeeds at $t+1$ (\$0) with probability θ_E

Therefore

$$\begin{aligned}
E[\text{File Now}] &= \$500 \text{ million} \\
E[\text{File } t+1] &= \$600(1-\theta_E) \text{ million} \\
E[\text{File } t+2] &= \$(700(1-\theta_E)^2 + 100(1-\theta_E)\theta_E) \text{ million} \\
E[\text{File } t+3] &= \$(800(1-\theta_E)^3 + 200(1-\theta_E)^2\theta_E + 100(1-\theta_E)\theta_E) \text{ million}
\end{aligned}$$

(b) Suppose that the incumbent must commit to one of the four possible actions now (t). What is the largest value of θ_E for which the incumbent, seeking to maximize its "expected profit," would choose to delay filing the patent (i.e. **not** file now)?

ANS: Set each of the expected payoffs from filing after t equal to \$500 million and solve for θ_E. This provides the value of θ_E such that θ_E value for which the firm is indifferent between filing now and filing at that delayed time. You should be able to show that the expected payoff monotonically declines with θ_E (take derivative or use intuition – growing θ_E reduces the odds of successful delay). So values of θ_E below those threshold values lead to the firm filing now.

$$\begin{aligned} E[\text{File t+1}] &= \$500 \text{ million} \implies \theta_E^* = \frac{1}{6} \\ E[\text{File t+2}] &= \$500 \text{ million} \implies \theta_E^* = \frac{1}{6} \qquad \text{Note: } \theta_E \in [0,1] \\ E[\text{File t+3}] &= \$500 \text{ million} \implies \theta_E^* = \frac{1}{6} \end{aligned}$$

You could have solved for these threshold values individually (the last one requiring you to solve a trinomial ... yuck) **or** you could have intuited the solution.

Note that each period you delay, you are trading off the risk of losing \$500 million (with probability θ_E) with the opportunity of winning an additional \$100 million (with probability $1 - \theta_E$). So, the firm is indifferent toward delaying an additional period when $500\theta_E = 100(1 - \theta_E)$ ... $\theta_E^* = \frac{1}{6}$.[1]

The firm is indifferent toward delay at $\theta_E = \frac{1}{6}$. For greater values, the firm prefers to file now. For lesser values, the firm prefers to delay.

[1]This is the same "marginal" intuition that drives much of economics: indifference occurs where marginal benefit equals marginal cost (MB = MC)

Question 4: A Past Quiz Problem (Modified)

Consider the discrete random variables X and Y defined as follows

$$X = \begin{cases} +1 & \text{with probability } \frac{1}{3} \\ 0 & \text{with probability } \frac{1}{3} \\ -1 & \text{with probability } \frac{1}{3} \end{cases} \quad \text{and} \quad Y \equiv X^2$$

(a) Derive the (marginal) distribution of Y

ANS:

Note that $(X = +1)$ necessarily implies $(Y = +1)$, similarly $(X = 0)$ necessarily implies $(Y = 0)$ and $(X = -1)$ necessarily implies $(Y = 1)$. Therefore, Y can only realize one of two values, $y \in \{0, 1\}$ with any positive probability.

Moreover

$$\begin{aligned} P_Y(Y = 0) &= P_X(X = 0) = f_X(0) = \frac{1}{3} \\ P_Y(Y = 1) &= P_X(X = -1) + P_X(X = +1) = f_X(-1) + f_X(+1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Therefore $f(y) = \begin{cases} \frac{2}{3} & \text{if } y = 1 \\ \frac{1}{3} & \text{if } y = 0 \\ 0 & \text{otherwise} \end{cases}$

(b) Show that (i) $\mathrm{Cov}(X, Y) = 0$ but (ii) X and Y are *not* statistically independent of each other. **Hint:** for the latter claim, look for a counter-example ...

ANS:

Let us start with (ii). If $X = +1$, $Y = +1$ necessarily. So $f(Y = 1|X = 1) = 1$. But $f_Y(1) = \frac{2}{3}$. Therefore $f(Y = 1|X = 1) \neq f_Y(1)$ and (X, Y) are not statistically independent of each other. There are other counter-examples as well.

Now (i)

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E[XY] - E[X]\ E[Y] = E[X^2 \cdot X] - E[X]\ E[X^2] = E[X^3] - E[X]\ E[X^2] \\ &= \underbrace{\left(\sum_{x=-1}^{+1} x^3 \cdot \frac{1}{3} \right)}_{=0} - \underbrace{\left(\sum_{x=-1}^{+1} x \cdot \frac{1}{3} \right)}_{=0} \cdot \underbrace{\left(\sum_{x=-1}^{+1} x^2 \cdot \frac{1}{3} \right)}_{=\frac{2}{3}} \\ &= 0 \end{aligned}$$

(c) Solve for (i) $BP_{MSE}(Y|X)$ and (ii) $BLP_{MSE}(Y|X)$

ANS: Start with (i) $BP_{MSE}(Y|X) = E[Y|X]$

$$E[Y|X] \;=\; E[X^2|X] \;=\; X^2 \;=\; Y$$

as conditioning on X effectively fixes the value of X (makes it no longer random)

Yes, the $BP_{MSE}(Y|X)$ is a **perfect** predictor for this example !!!

Now (ii) $BLP_{MSE}(Y|X) = a^* + b^*X$

$$\begin{aligned}
b^* &= \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \;=\; 0 \;\text{ as } \text{Cov}(X,Y) = 0 \text{ as shown in } \textbf{(b)} \\
a^* &= E[Y] - b^*E[X] \;=\; E[Y] \;\text{ as } b^* = 0 \\
&= \sum_{y=0}^{+1} y \cdot f(y) \;=\; 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \;=\; \frac{2}{3}
\end{aligned}$$

So $BLP_{MSE}(Y|X) = E[Y] = \frac{2}{3}$

(d) Often, the $BP_{MSE}(Y|X)$ and $BLP_{MSE}(Y|X)$ share similar intuition as to how X is informative about Y. But for the (X,Y) given for this problem, this does not hold true. For one of the two predictors (under MSE), X is incredible informative about Y but for the other X is entirely uninformative about Y. Explain which is which and why.

ANS: As shown above, $BP_{MSE}(Y|X) = Y$ and is, thus, a **perfect** predictor. But $BLP_{MSE}(Y|X) = E[Y]$ and is, thus, an **uninformative** predictor in that it does not use any information about the realization of X. This is because $BLP_{MSE}(Y|X)$ relies heavily on the **second moment** between X and Y but this second moment is uninformative as $\text{Cov}(X,Y) = 0$ for this example. This contrasts $BP_{MSE}(Y|X)$ which relies on the **conditional distribution** of Y given X, not just that second moment.

"Food for Thought": Getting Ready to Regress

Let (X, Y) be two random variables with some well defined joint distribution. Consider a third random variable defined as $Z \equiv \gamma_0 + \gamma_1 X + Y$. Also ...

- You are given a **random** sample of N draws of (Z, X): $\{Z_i, X_i\}_{i=1}^N$
- You do **not** know the values of the constants (γ_0, γ_1) or the associated Y draws: $\{Y_i\}_{i=1}^N$

(a) Suppose you were given the values of the associated Y draws. How would you "estimate" the unknown parameter values (γ_0, γ_1)? How good of an estimate would they be?

ANS: No need to estimate. Just solve the system of (linear) equations! Let (X_i, Y_i, Z_i) and (X_j, Y_j, Z_j) be two observations from the sample where at least one of the random variables has a different value. Then solve the following two (linearly independent) equations for (γ_0, γ_1)

$$\begin{aligned} Z_i &= \gamma_0 + \gamma_1 X_i + Y_i \\ Z_j &= \gamma_0 + \gamma_1 X_j + Y_j \end{aligned}$$

(b) Derive the BP and BLP of Z given X under the MSE criterion as a function of $\{\gamma_0, \gamma_1, X\}$ and any relevant moments derived from the joint distribution of $\{X, Y, Z\}$ (and from other distributions that can be derived from that joint distribution).

ANS: From notes, $BP_{MSE}(Z|X) = E[Z|X]$

$$E[Z|X] \quad = \quad E[\gamma_0 + \gamma_1 X + Y \mid X] \quad = \quad \gamma_0 + \gamma_1 X + E[Y|X]$$

From notes, $BLP_{MSE}(Z|X) = \alpha^* + \beta^* X$ where $\alpha^* = E[Z] - \beta^* E[X]$ and $\beta^* = \frac{Cov[Z,X]}{Var[X]}$

Note that $E[X], E[Z], Cov[Z, X], Var[X]$ are all moments that can be derived from the joint distribution.

(c) How does your answer to **(b)** change if you were also told that X and Y were distributed independently of each other and that $E[Y] = 0$?

ANS: As X and Y are distributed independently of each other, $E[Y|X] = E[Y] = 0$. Therefore $E[Z|X] = \gamma_0 + \gamma_1 X$. Note that the $BP_{MSE}[Z|X]$ is now linear in X ...