Problem Set #1 (2^{nd} Half) Solutions

Economics 709B **Fall 2020**

2.1 By applying the law of iterated expectations twice, we get

$$\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[Y|\boldsymbol{X_1},\boldsymbol{X_2},\boldsymbol{X_3}\right]|\boldsymbol{X_1},\boldsymbol{X_2}\right]|\boldsymbol{X_1}\right]=\mathbb{E}\left[\mathbb{E}\left[Y|\boldsymbol{X_1},\boldsymbol{X_2}\right]|\boldsymbol{X_1}\right]=\mathbb{E}\left[Y|\boldsymbol{X_1}\right]$$

2.2 Let a, b be some constants. By the law of iterated expectations,

$$\mathbb{E}\left[YX\right]=\mathbb{E}\left[\mathbb{E}\left[YX|X\right]\right]=\mathbb{E}\left[X\mathbb{E}\left[Y|X\right]\right]=\mathbb{E}\left[X\left(a+bX\right)\right]=a\mathbb{E}\left[X\right]+b\mathbb{E}\left[X^2\right],$$

which is a function of the first and second moments of X.

2.3 If $\mathbb{E}\left[\left|h\left(\boldsymbol{X}\right)e\right|\right]<\infty$, $\mathbb{E}\left[h\left(\boldsymbol{X}\right)e\right]$ exists. By the law of iterated expectations, it follows that $\mathbb{E}\left[h\left(\boldsymbol{X}\right)e\right]=\mathbb{E}\left[\mathbb{E}\left[h\left(\boldsymbol{X}\right)e|\boldsymbol{X}\right]\right]=\mathbb{E}\left[h\left(\boldsymbol{X}\right)\mathbb{E}\left[e|\boldsymbol{X}\right]\right]=0$, since $\mathbb{E}\left[e|\boldsymbol{X}\right]=0$.

2.4 Note that $P\left(Y=0|X=0\right)=\frac{P(Y=0,X=0)}{P(X=0)}=\frac{0.1}{0.1+0.4}=0.2$, $P\left(Y=1|X=0\right)=0.8$, $P\left(Y=0|X=1\right)=0.4$, $P\left(Y=1|X=1\right)=0.6$.

$$\begin{aligned}
\mathbb{E}\left[Y|X=0\right] &= 0\times P\left(Y=0|X=0\right)+1\times P\left(Y=1|X=0\right)=0.8\\
\mathbb{E}\left[Y|X=1\right] &= 0\times P\left(Y=0|X=1\right)+1\times P\left(Y=1|X=1\right)=0.6\\
\mathbb{E}\left[Y^2|X=0\right] &= 0^2\times P\left(Y=0|X=0\right)+1^2\times P\left(Y=1|X=0\right)=0.8\\
\mathbb{E}\left[Y^2|X=1\right] &= 0^2\times P\left(Y=0|X=1\right)+1^2\times P\left(Y=1|X=1\right)=0.6\\
\mathrm{var}\left[Y|X=0\right] &= \mathbb{E}\left[Y^2|X=0\right]-\left(\mathbb{E}\left[Y|X=0\right]\right)^2=0.16\\
\mathrm{var}\left[Y|X=1\right] &= \mathbb{E}\left[Y^2|X=1\right]-\left(\mathbb{E}\left[Y|X=1\right]\right)^2=0.24
\end{aligned}$$

2.5 (c) The mean squared error of a predictor $h\left(\boldsymbol{X}\right)$ for e^2 is

$$\begin{aligned}
\mathbb{E}\left[\left(e^2-h\left(\boldsymbol{X}\right)\right)^2\right] &= \mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)+\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)^2\right]\\
&= \mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)^2+\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)^2+2\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\right]\\
&= \mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)^2\right]+\mathbb{E}\left[\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)^2\right]
\end{aligned}$$

where the third equality holds since $\mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\right]=0$, which we show below. The last expression is minimized at $h\left(\boldsymbol{X}\right)=\sigma^2\left(\boldsymbol{X}\right)$, thus $\sigma^2\left(\boldsymbol{X}\right)$ is the best predictor.

By the law of iterated expectations and by the definition of conditional variance $\sigma^2\left(\boldsymbol{X}\right)=\mathbb{E}\left[e^2|\boldsymbol{X}\right]$,

$$\begin{aligned}
\mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(e^2-\sigma^2\left(\boldsymbol{X}\right)\right)\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)|\boldsymbol{X}\right]\right]\\
&= \mathbb{E}\left[\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\mathbb{E}\left[e^2-\sigma^2\left(\boldsymbol{X}\right)|\boldsymbol{X}\right]\right]\\
&= \mathbb{E}\left[\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\left(\mathbb{E}\left[e^2|\boldsymbol{X}\right]-\sigma^2\left(\boldsymbol{X}\right)\right)\right]\\
&= \mathbb{E}\left[\left(\sigma^2\left(\boldsymbol{X}\right)-h\left(\boldsymbol{X}\right)\right)\times 0\right]=0.
\end{aligned}$$

2.8 Since $Y|\boldsymbol{X}=\boldsymbol{x} \sim$ Poisson $(\boldsymbol{x}'\boldsymbol{\beta})$, it follows that $\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}] = \text{var}(Y|\boldsymbol{X}=\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$. Therefore, we can justify a linear regression model $Y=\boldsymbol{x}'\boldsymbol{\beta}+e$ with $\mathbb{E}[e|\boldsymbol{x}]=0$, since the conditional expectation function is actually linear.

2.10 True. The mean independence condition $\mathbb{E}[e|X]=0$ implies that $\mathbb{E}[h(X)e]=0$ for any function $h(X)$ of X as long as $\mathbb{E}[|h(X)e|]<\infty$. Then, by the law of iterated expectations $\mathbb{E}\left[X^2e\right]=\mathbb{E}\left[\mathbb{E}\left[X^2e|X\right]\right]=\mathbb{E}\left[X^2\mathbb{E}[e|X]\right]=0$.

2.11 False. Here's a counterexample ... Suppose that X has a symmetric distribution around zero (the odd moments of X are all zero) and is nonzero with positive probability. Suppose $Y=X^2$, and consider a linear projection model $Y=\beta X+e$. Then $\beta=\left(\mathbb{E}[XX']\right)^{-1}\mathbb{E}[XY]=\frac{\mathbb{E}\left[X^3\right]}{\mathbb{E}\left[X^2\right]}=0$, and $e=Y-\beta X=X^2$. Therefore $\mathbb{E}\left[X^2e\right]=\mathbb{E}\left[X^4\right]>0$, since X^4 is always positive.

2.12 False. Mean independence does not imply full independence. Here's a counterexample ...

Consider discrete random variables X and e by

$X\backslash e$	1	0	-1
1	0	$\frac{1}{5}$	0
0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
-1	0	$\frac{1}{5}$	0

then $\mathbb{E}[e|X]=0$ for any $X=1,0,-1$, but $Pr(X=1)\cdot Pr(e=1)=\frac{1}{5}\cdot\frac{1}{5}=\frac{1}{25}\neq 0=Pr(X=1,e=1)$, so X and e are not independent.

2.13 False. See Counterexample in Exercise 2.11.

2.14 False. Mean independence and homoskedasticity do not imply that the random variables X and e are independent. Here's a counterexample ... Consider $Y=Xu$, $\mathbb{E}[u|X]=1$, $\text{var}(u|X)=\sigma^2/X^2$. Consider the CEF error, $e=Y-\mathbb{E}[Y|X]=Xu-X\mathbb{E}[u|X]=X(u-1)$. Even though $\mathbb{E}[e|X]=0$ and $\mathbb{E}\left[e^2|X\right]=\mathbb{E}\left[X^2(u-1)^2|X\right]=X^2\mathbb{E}\left[(u-1)^2|X\right]=X^2\text{var}(u|X)=X^2\left(\sigma^2/X^2\right)=\sigma^2$, e and X are not independent by construction.

2.16 The best linear predictor and the conditional mean are different in this exercise, since $m(x)$ is a non-linear function of x.

Since $f(y|x)=\frac{f(x,y)}{\int_0^1 f(x,y)dy}=\frac{(3/2)\left(x^2+y^2\right)}{(3/2)x^2+1/2}1\{0\leq y\leq 1\}$, the conditional mean function $m(x)$ is equal to

$$\mathbb{E}[Y|X=x]=\int_0^1 yf(y|x)\,dy=\frac{6x^2+3}{12x^2+4}. \tag{1}$$

Also,

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[X]=\int_0^1 xf(x)\,dx=\frac{5}{8},\\
\mathbb{E}\left[X^2\right] &= \int_0^1 x^2f(x)\,dx=\frac{7}{15},\\
\mathbb{E}[XY] &= \int_0^1\int_0^1 xyf(x,y)\,dxdy=\frac{3}{8}.
\end{aligned}$$

Therefore, we can compute the coefficients of the best linear predictor α, β as follows:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2} \begin{pmatrix} \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[XY\right] \\ \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] \end{pmatrix} = \begin{pmatrix} \frac{55}{73} \\ -\frac{15}{73} \end{pmatrix}. \qquad (2)$$

Then, by (1) and (2) the best linear predictor $\mathcal{P}\left(Y|X=x\right) = \alpha + \beta x = \frac{55}{73} - \frac{15}{73}x$ and $m\left(x\right) = \frac{6x^2+3}{12x^2+4}$ are different.

4.1 (a) Let $\hat{\mu}_k = \overline{Y^k} = \frac{1}{n}\sum_{i=1}^n Y_i^k$.

(b) $\mathbb{E}\left[\hat{\mu}_k\right] = \mathbb{E}\left[Y_i^k\right] = \mu_k$.

(c) Assume $Y_1, \ldots, Y_n$ are i.i.d. Then, $var(\hat{\mu}_k) = \frac{1}{n}var(Y_i^k) = (\mu_{2k} - \mu_k^2)/n$, which exists if $\mathbb{E}\left[Y_i^{2k}\right] < \infty$.

(d) $\widehat{var}(\hat{\mu}_k) = (\hat{\mu}_{2k} - \hat{\mu}_k^2)/n$, where $\hat{\mu}_{2k} = \frac{1}{n}\sum_{i=1}^n Y_i^{2k}$.

4.2

$$\begin{aligned}
\mathbb{E}\left[(\bar{Y}-\mu)^3\right] &= \frac{1}{n^3}\mathbb{E}\left[\left(\sum_{i=1}^n (Y_i - \mu)\right)^3\right] \\
&= \frac{1}{n^3}\left\{\sum_{i=1}^n \mathbb{E}\left[(Y_i-\mu)^3\right] + 3\sum_{i\neq j}\mathbb{E}\left[(Y_i-\mu)^2(Y_j-\mu)\right] + 6\sum_{i>j>l}\mathbb{E}\left[(Y_i-\mu)(Y_j-\mu)(Y_l-\mu)\right]\right\}
\end{aligned}$$

Assume $Y_1, \ldots, Y_n$ are i.i.d. Then,
$\mathbb{E}\left[(Y_i-\mu)^2(Y_j-\mu)\right] = \mathbb{E}\left[(Y_i-\mu)^2\right]\mathbb{E}\left[(Y_j-\mu)\right] = \mathbb{E}\left[(Y_i-\mu)^2\right]\cdot 0 = 0$.
And, $\mathbb{E}\left[(Y_i-\mu)(Y_j-\mu)(Y_l-\mu)\right] = \mathbb{E}\left[Y_i-\mu\right]\mathbb{E}\left[Y_j-\mu\right]\mathbb{E}\left[Y_l-\mu\right] = 0$.
Then, $\mathbb{E}\left[(\bar{Y}-\mu)^3\right] = \frac{1}{n^2}\mathbb{E}\left[(Y_i-\mu)^3\right]$. The last expression is zero if the skewness of Y_i is zero, $\mathbb{E}\left[(Y_i-\mu)^3\right] = 0$, e.g. Y_i has a symmetric distribution around its mean μ.

4.3 μ and $\mathbb{E}\left[X_iX_i'\right]$ are the population means of Y_i and the matrix X_iX_i', and $\bar{Y}$ and $\frac{1}{n}\sum_{i=1}^n X_iX_i'$ are the sample analogs of the unknown population means. As sample size n increases, the sample mean $\frac{1}{n}\sum_{i=1}^n X_iX_i'$ converges in probability to $\mathbb{E}\left[X_iX_i'\right]$ by WLLN under $\mathbb{E}\left[\|X_iX_i'\|\right] < \infty$.

4.4 False. Let $X = (X_1, X_2, \ldots, X_n)'$, $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_n)'$, $X^2 = \left(X_1^2, \ldots, X_n^2\right)'$. For OLS residual vector $\hat{e}$, X and $\hat{e}$ are orthogonal, i.e. $X'\hat{e} = 0$. In general, this does not imply $X^2\hat{e} \neq 0$ unless X and X^2 are linearly dependent, i.e., X is a constant vector $(1, \ldots, 1)'$.

Counter example: $n = 3$, $\{(Y_i, X_i)\} = \{(7,2), (4,3), (8,4)\}$. $\hat{\beta} = \frac{\sum X_iY_i}{\sum X_i^2} = 2$, $\hat{e}_1 = 3$, $\hat{e}_2 = -2$, $\hat{e}_3 = 0$. Here $\sum_{i=1}^3 X_i\hat{e}_i = 0$, but $\sum_{i=1}^n X_i^2\hat{e}_i = -6 \neq 0$.

4.5 Under assumption $\mathbb{E}\left[\boldsymbol{e}|\boldsymbol{X}\right] = 0$, we have

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}|\boldsymbol{X}\right] &= \mathbb{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}|\boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left[\boldsymbol{Y}|\boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left[\boldsymbol{X}\beta + \boldsymbol{e}|\boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\mathbb{E}\left[\boldsymbol{X}\beta|\boldsymbol{X}\right] + \mathbb{E}\left[\boldsymbol{e}|\boldsymbol{X}\right]\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\beta \\
&= \beta.
\end{aligned}$$

Let $var(\boldsymbol{e}|\boldsymbol{X}) = E(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}) = \Omega$. Then

$$\begin{aligned} var\left[\hat{\beta}|\boldsymbol{X}\right] &= \mathbb{E}\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|\boldsymbol{X}\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}|\boldsymbol{X}\right] \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}\right]\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\Omega\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}. \end{aligned}$$

4.6 As noted in class, $\tilde{\beta}$ is linear and unbiased if $\tilde{\beta} = A'Y$, where A is an $n \times k$ function of X such that $A'X = I_k$. The variance of $\tilde{\beta}$ is

$$\text{var}\left(\tilde{\beta}|X\right) = A'\text{var}\left(Y|X\right)A = A'\Omega A,$$

Set $C = A - \Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}$. Note that $X'C = 0$. Then

$$\begin{aligned} \text{var}\left(\tilde{\beta}|X\right) - \left(X'\Omega^{-1}X\right)^{-1} &= A'\Omega A - \left(X'\Omega^{-1}X\right)^{-1} \\ &= \left(C + \Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}\right)'\Omega\left(C + \Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}\right) - \left(X'\Omega^{-1}X\right)^{-1} \\ &= C'\Omega C + C'\Omega\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} + \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\Omega C \\ &+ \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} - \left(X'\Omega^{-1}X\right)^{-1} \\ &= C'\Omega C = \left(\Omega^{1/2}C\right)'\left(\Omega^{1/2}C\right), \end{aligned}$$

The last term is positive semi-definite.