

1. (5 points) Consider the regression model with usual assumptions of the errors

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

- (a) Suppose that the observation  $y_i$  falls directly on the fitted regression line (i.e.,  $y_i = \hat{y}_i$ ). If this case is deleted, would the least squares estimators with the remaining  $n - 1$  cases be changed?
- (b) Suppose that we have fit the straight-line regression model  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ , but the true regression function is

$$\mathbb{E}[y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Is the estimator  $\hat{\beta}_1$  biased?

- (c) Show that  $\hat{\boldsymbol{\beta}}$  and  $\mathbf{e}$  are statistically independent, where  $\hat{\boldsymbol{\beta}}$  is the least squares estimator of  $\boldsymbol{\beta}$  and  $\mathbf{e}$  is the corresponding residuals.

**Solution:** The key relation is

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{1}{1 - h_{ii}} (X'X)^{-1} x_i e_i,$$

where  $h_{ii}$  is the  $i^{\text{th}}$  diagonal element of the hat matrix,  $x_i$  is the  $i^{\text{th}}$  row of the design matrix  $X$ , and  $e_i$  is the residual of the  $i^{\text{th}}$  observation. If the observation fell directly on the fitted regression, then the residual  $e_i = 0$ . Therefore, the least squares estimators remain unchanged.

**Solution:** If the researcher mistakenly omitted a variable that should not have been omitted, then the estimators are biased. We call this the *omitted variable bias*. Recall the estimate for  $\beta_1$  in a simple regression setting:

$$\hat{\beta}_1 = \frac{\text{Cov}(x_1, y)}{\text{Var}(x_1)}.$$

Now plug in the true model instead of  $y$ ,

$$\frac{\text{Cov}(x_1, \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon)}{\text{Var}(x_1)}.$$

By the bilinearity of the covariance operator, it becomes

$$\frac{1}{\text{Var}(x_1)} \{ \text{Cov}(x_1, \beta_0) + \beta_1 \text{Var}(x_1) + \beta_2 \text{Cov}(x_1, x_2) + \text{Cov}(x_1, \epsilon) \}.$$

Covariance with a constant is always zero and regressor variables and the error terms are uncorrelated according to the usual assumptions. Therefore,

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \underbrace{\beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}}_{\text{Bias}}.$$

**Solution:** The least squares estimator of  $\beta$  can be written as follows:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

And for notational convenience, let  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , called the *hat matrix* in linear regression terminology, or *projection matrix* in linear algebra literature. Then

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\beta + \epsilon) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon \\ \mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\beta} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\beta + \epsilon) \\ &= (\mathbf{I} - \mathbf{H})\mathbf{X}\beta - (\mathbf{I} - \mathbf{H})\epsilon. \end{aligned}$$

Since both  $\hat{\beta}$  and  $\mathbf{e}$  are linear combinations of normal variates, they also follow normal distributions. For normal distributions, having no correlation is the same as being independent.

$$\begin{aligned} \text{Cov}[\hat{\beta}, \mathbf{e}] &= \text{Cov}\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon, (\mathbf{H} - \mathbf{I})\epsilon\right] \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \\ &= \mathbf{0} \end{aligned}$$

Therefore, they are independent.

2. (5 points) Consider the regression model with  $k$  regressor variables,  $\mathbf{y} = \mathbf{X}\beta + \epsilon$ , where  $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{V})$ ,  $\mathbf{V} \neq \mathbf{I}$ .

- Obtain the generalized least squares estimator of  $\beta$ .
- Find the expected value and variance of  $\hat{\beta}_{\text{GLS}}$  in the part (a).

**Solution:** When homoskedasticity and uncorrelated observations assumptions don't hold, the ordinary least squares estimators are not appropriate. We take another approach for these situations by transforming the model. Since the covariance matrix is an  $n \times n$  nonsingular & positive definite matrix, there must be a square root matrix of  $\mathbf{V}$  such that  $\mathbf{K}^2 = \mathbf{K}'\mathbf{K} = \mathbf{V}$ . Therefore, if we let  $\mathbf{z} = \mathbf{K}^{-1}\mathbf{y}$ ,  $\mathbf{B} = \mathbf{K}^{-1}\mathbf{X}$ , and  $\mathbf{g} = \mathbf{K}^{-1}\epsilon$ , the model can be recast as

$$\begin{aligned} \mathbf{K}^{-1}\mathbf{y} &= \mathbf{K}^{-1}\mathbf{X}\beta + \mathbf{K}^{-1}\epsilon \\ \mathbf{z} &= \mathbf{B}\beta + \mathbf{g}. \end{aligned}$$

Now we apply the least squares method on the transformed model.

$$\hat{\beta} = \text{argmin}_{\beta} \mathbf{g}'\mathbf{g} = \text{argmin}_{\beta} \epsilon' \mathbf{V}^{-1} \epsilon.$$

Proceeding,

$$\epsilon' V^{-1} \epsilon = (y - X\beta)' V^{-1} (y - X\beta) \quad (1)$$

$$\frac{\partial}{\partial \beta} (y'y - 2\beta' X' V^{-1} y + \beta' X' V^{-1} X \beta) = -2X' V^{-1} y + 2X' V^{-1} X \beta. \quad (2)$$

Setting eqn (2) to zero, we obtain  $\hat{\beta}_{\text{GLS}} = (X' V^{-1} X)^{-1} X' V^{-1} y$ . This is not the only way to achieve the generalized least squares estimators. We can also use the fact that the maximum likelihood estimators are equal to least squares estimators and obtain the same result.

**Solution:** We can plug in the original model for  $y$  in  $\hat{\beta}_{\text{GLS}}$ .

$$\hat{\beta}_{\text{GLS}} = (X' V^{-1} X)^{-1} X' V^{-1} (X\beta + \epsilon).$$

Therefore,

$$\begin{aligned} \mathbb{E} [\hat{\beta}_{\text{GLS}}] &= \mathbb{E} [(X' V^{-1} X)^{-1} X' V^{-1} X \beta] + \mathbb{E} [(X' V^{-1} X)^{-1} X' V^{-1} \epsilon] \\ &= \beta \\ \text{Var} [\hat{\beta}_{\text{GLS}}] &= \text{Var} [(X' V^{-1} X)^{-1} X' V^{-1} X \beta + (X' V^{-1} X)^{-1} X' V^{-1} \epsilon] \\ &= (X' V^{-1} X)^{-1} X' V^{-1} \text{Var} [\epsilon] V^{-1} X (X' V^{-1} X)^{-1} \\ &= (X' V^{-1} X)^{-1} X' V^{-1} \sigma^2 V V^{-1} X (X' V^{-1} X)^{-1} \\ &= \sigma^2 \cancel{(X' V^{-1} X)^{-1} X' V^{-1} X} (X' V^{-1} X)^{-1} \\ &= \sigma^2 (X' V^{-1} X)^{-1}. \end{aligned}$$

3. (5 points) Consider the model

$$y = X\beta + \epsilon,$$

where  $\epsilon$  are i.i.d. normally distributed with mean  $\mathbf{0}$  and variance  $\sigma^2 \mathbf{I}$ . Assume that we have a multicollinearity problem.

- Define the ridge estimator of  $\beta$ .
- Show that the ridge estimator is the solution to the problem

$$\text{Minimize } (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \quad \text{subject to } \beta' \beta \leq c$$

**Solution:** Assume that multicollinearity is present and we need to render  $(X' X)^{-1}$  invertible somehow. We can drop the unbiasedness requirement and obtain a better estimators by adding

*Tikhonov* regularization defined as follows:

$$\hat{\beta}_{\text{RIDGE}} = \operatorname{argmin}_{\beta} (y - X\beta)'(y - X\beta) + \lambda\beta'\beta.$$

By differentiating with respect to  $\beta$  and setting it to zero, we obtain

$$\hat{\beta}_{\text{RIDGE}} = (X'X + \lambda I_n)^{-1} X'y.$$

**Solution:**

4. (4 points) Let  $S$  denote the sample space. Suppose that  $C \subset S$  and  $\mathbb{P}(C) > 0$  and define  $Q(A) = \mathbb{P}(A|C)$ . Show that  $Q$  satisfies the axiom of probability.

**Solution:** First, there are 3 axioms to probability so please add an 's'. Those three are as follows:

- $Q(S) = 1$ . (Unitarity)
- $Q(E) \geq 0$ . (Nonnegativity)
- $Q(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Q(A_i)$  if  $A_i$  are disjoint. (Countable additivity)

Let's prove.

1. For unitarity, we will use the definition.

$$Q(S) = \mathbb{P}(S|C) = \frac{\mathbb{P}(S, C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = 1.$$

2. For nonnegativity, if  $E$  is an empty set, by definition,

$$Q(E) = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(C)} = 0.$$

If  $E$  is a nonempty set, by definition,

$$Q(E) = \frac{\mathbb{P}(E, C)}{\mathbb{P}(C)} = \begin{cases} 0, & E \cap C = \emptyset \\ > 0, & E \cap C \neq \emptyset \end{cases}.$$

3. For countable additivity, by definition, we should consider the set relation

$$\left( \bigcup_{i=1}^{\infty} A_i \right) \cap C = \bigcup_{i=1}^{\infty} (A_i \cap C).$$

If  $A_i$  are disjoint, then  $A_i \cap C$  should also be disjoint. Therefore,

$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\mathbb{P}(\bigcup_{i=1}^{\infty} (A_i \cap C))}{\mathbb{P}(C)} = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap C)}{\mathbb{P}(C)} = \sum_{i=1}^{\infty} Q(A_i).$$

5. (3 points) Construct an example where  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  but  $X_n + Y_n$  does not converge in distribution to  $X + Y$ .

**Solution:** Let

$$X_n = \begin{cases} Z, & n \text{ is even} \\ -Z, & n \text{ is odd} \end{cases}, \quad Y_n = \begin{cases} -Z, & n \text{ is even} \\ Z, & n \text{ is odd} \end{cases},$$

where  $Z$  is a continuous random variable symmetric about zero. Then both  $X_n, Y_n$  will converge in distribution to  $Z$  but  $X_n + Y_n$  will converge in probability to zero.

6. (6 points) Suppose that we observe  $X_1, \dots, X_n$  from  $\mathcal{N}(\mu, \sigma^2)$ . Find the uniformly minimum variance unbiased estimator (UMVUE) of  $\eta = \mu/\sigma^2$ .

**Solution:** By Lehmann-Scheffé theorem, if  $\eta$  is estimable, then there is a unique unbiased estimator of  $\eta$  that is of the form  $h(T)$  where  $T$  is the sufficient & complete statistic for  $P \in \mathcal{P}$ . Since all  $X_i$  are normal, jointly sufficient & complete statistics are  $(\bar{X}, S^2)$ . We only need to find an unbiased estimator  $\hat{\eta}$  that is a function  $h(\bar{X}, S^2)$ . Since

$$\bar{X} \xrightarrow{P} \mu, \quad S \xrightarrow{P} \sigma,$$

we shall start with  $\bar{X}/S$  and compute  $\mathbb{E}[\bar{X}/S]$ . By Basu's theorem,  $\bar{X}$  and  $S$  are independent; therefore  $\mathbb{E}[\bar{X}/S] = \mathbb{E}[\bar{X}] \mathbb{E}[1/S]$ .

$$\mathbb{E}\left[\frac{\bar{X}}{S}\right] = \mu \mathbb{E}\left[\frac{1}{S}\right]$$

Since  $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ ,

$$S^2 \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right).$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{S}\right] &= \int_0^\infty y^{-1/2} \left(\frac{n-1}{2\sigma^2}\right)^{(n-1)/2} \left(\Gamma\left(\frac{n-1}{2}\right)\right)^{-1} y^{(n-1)/2-1} e^{-(n-1)y/(2\sigma^2)} dy \\ &= \Gamma\left(\frac{n-2}{2}\right) \left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^{-1} \left\{\frac{n-1}{2}\right\}^{1/2} \frac{1}{\sigma} \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\frac{\bar{X}}{S}\right] = \Gamma\left(\frac{n-2}{2}\right) \left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^{-1} \left\{\frac{n-1}{2}\right\}^{1/2} \frac{\mu}{\sigma}.$$

To make this unbiased, we should multiply it by a constant:

$$\hat{\eta}_{\text{UMVUE}} = C \frac{\bar{X}}{S}, \quad C = \Gamma\left(\frac{n-1}{2}\right) \left\{\Gamma\left(\frac{n-2}{2}\right)\right\}^{-1} \left\{\frac{n-1}{2}\right\}^{-1/2}.$$

7. (4 points) Let  $X \sim \text{Exp}(1)$ , and define  $Y$  to be the integer part of  $X + 1$ , that is

$$Y = i + 1 \text{ if and only if } i \leq X < i + 1, \quad i = 0, 1, 2, \dots$$

Find the conditional distribution of  $X - 4$  given  $Y \leq 5$ .

**Solution:** This is a very sly way of stating  $Y = \lceil X \rceil$ . Then the condition  $Y \leq 5$  is equivalent to  $X \leq 5$ . Therefore, we should start with the c.d.f. of  $Z = X - 4 | Y \leq 5$ .

$$\begin{aligned} F_Z(z) &= \frac{\mathbb{P}(X - 4 \leq z, X \leq 5)}{\mathbb{P}(X \leq 5)} \\ &= \frac{\mathbb{P}(X \leq 4 + z, X \leq 5)}{\mathbb{P}(X \leq 5)} \end{aligned}$$

This now divides into

$$\frac{\mathbb{P}(X \leq 4 + z, X \leq 5)}{\mathbb{P}(X \leq 5)} = \begin{cases} 1, & \text{if } k \geq 1 \\ \frac{\mathbb{P}(X \leq 4 + z)}{\mathbb{P}(X \leq 5)}, & \text{if } -4 < k < 1 \end{cases}.$$

Since we all know that the c.d.f. of an exponential distribution with parameter 1 is  $F_X(x) = 1 - e^{-x}$ ,

$$\frac{\mathbb{P}(X \leq 4 + z)}{\mathbb{P}(X \leq 5)} = \frac{1 - e^{-(4+z)}}{1 - e^{-5}}.$$

To get the p.d.f., differentiate it with respect to  $z$ , which yields

$$f_Z(z) = \frac{e^{-(4+z)}}{1 - e^{-5}}, \quad -4 < z < 1.$$

8. (6 points) Let  $Y \sim \text{Poisson}(\lambda)$  and  $X|Y = y \sim \text{Bin}(y, p)$ .

- (a) Determine the distribution of  $X$ .  
(b) Find the conditional pdf of  $Y$  given  $X = x$ .

**Solution:** Marginalizing out  $Y$ ,

$$\begin{aligned} f_X(x) &= \sum_{y=x}^{\infty} f_{X|Y}(x|y) f_Y(y) \\ &= \sum_{y=x}^{\infty} \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{\{\lambda(1-p)\}^{y-x}}{(y-x)!} \\ &= \frac{e^{-\lambda p} (\lambda p)^x}{x!}. \end{aligned}$$

Therefore,  $X \sim \text{Poisson}(\lambda p)$ .

**Solution:**

$$\begin{aligned}
 f_{Y|X}(y|x) &= \frac{e^{-\lambda} \lambda^y}{y!} \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} \frac{x!}{e^{-\lambda p} (\lambda p)^x} \\
 &= \frac{e^{-\lambda} \lambda^y p^x (1-p)^{y-x} e^{\lambda p} (\lambda p)^{-x}}{(y-x)!}
 \end{aligned}$$

9. (6 points) Suppose that we observe  $X$  having a distribution  $P \in \{P_0, P_1\}$ , and that we wish to test the hypothesis  $H_0 : P = P_0$  versus  $H_1 : P = P_1$ , based on an observation of  $X$ . Let  $P_0$  and  $P_1$  be given by

$x$	0	1	2	3
$P_0$	0.03	0.05	0.12	0.80
$P_1$	0.15	0.10	0.30	0.45

- (a) Find non-randomized most powerful test(s) at level  $\alpha = 0.05$ .  
 (b) Find randomized most powerful test(s) at level  $\alpha = 0.05$ .

**Solution:****Solution:**

10. (6 points) Let  $X_n$  be a random variable with  $\mathbb{P}(X_n = i/n) = 1/n, 1 \leq i \leq n$ . Find the limit distribution of  $X_n$  as  $n \rightarrow \infty$ .

**Solution:** To get the limit distribution we should start with the c.d.f.

$$F_{X_n}(x) = \frac{\lfloor nx \rfloor}{n}.$$

By constructing consecutive inequalities,  $nx-1 \leq \lfloor nx \rfloor \leq nx$ , we can get the limit  $\lim_{n \rightarrow \infty} F_{X_n}$ .

$$\lim_{n \rightarrow \infty} \frac{nx-1}{n} \leq \lim_{n \rightarrow \infty} \frac{\lfloor nx \rfloor}{n} \leq \lim_{n \rightarrow \infty} \frac{nx}{n}.$$

Therefore,  $F_{X_n} \xrightarrow{n \rightarrow \infty} F_X$  where  $F_X = x$ . This is the c.d.f. of a continuous uniform distribution on  $(0, 1)$ .