STAT 24510 PROBLEM SET 7

YIDUAN DONNA ZHENG

DUE FEBRUARY 23, 2023

Notation. Throughout the problem set, we will be using the following abbreviations consistently:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,\ V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^{(2)}} - (\overline{x})^2 .$$

Problem 1.

Set-up. Consider $y \sim N(X\beta, \sigma^2 I_n$. An unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = \|y - X\hat{\beta}\|^2.$$

1(a).

Question. Find the asymptotic distribution of $\sqrt{n-p}(\hat{\sigma}^2 - \sigma^2)$.

Solution. First, we know that

$$\frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p} = \sum_{i=1}^{n-p} \chi^2_1 \rightsquigarrow N(n-p, 2(n-p)) \tag{$*$}$$

by CLT. The rest follows by moving terms around:

$$\begin{aligned}
(*) &\implies \|y - X\hat{\beta}\|^2 \rightsquigarrow N\big(\sigma^2(n-p), 2(\sigma^2)^2(n-p)\big), \\
&\implies \hat{\sigma}^2 = \frac{\|y - X\hat{\beta}\|^2}{n-p} \rightsquigarrow N\left(\sigma^2, \frac{2(\sigma^2)^2}{n-p}\right), \\
&\implies (\hat{\sigma}^2 - \sigma^2) \rightsquigarrow N\left(0, \frac{2(\sigma^2)^2}{n-p}\right), \\
&\implies \sqrt{n-p}(\hat{\sigma}^2 - \sigma^2) \rightsquigarrow N\left(0, 2(\sigma^2)^2\right).
\end{aligned}$$

1(b).

Question. Find a transform $g(\cdot)$ so that $\sqrt{n-p}(g(\hat{\sigma}^2) - g(\sigma^2)) \rightsquigarrow N(0,1)$.

Solution. Per VST, we want a transformation $g(\cdot)$ such that

$$2(\sigma^2)^2|g'(\sigma^2)|^2 = 1 \implies |g'(\sigma^2)|^2 = \frac{1}{2(\sigma^2)^2} \implies |g'(\sigma^2)| = \frac{1}{\sqrt{2}\sigma^2}.$$

One corresponding solution is

$$g(\sigma^2) = \frac{1}{\sqrt{2}} \log(\sigma^2).$$

1(c).

Question. Using the above asymptotic result to construct an approximate $(1-\alpha)$-confidence interval for σ^2.

Solution. By part (b), we know

$$\sqrt{\frac{n-p}{2}} \cdot \left(\log(\hat{\sigma}^2) - \log(\sigma^2)\right) \rightsquigarrow N(0,1).$$

The corresponding CI is defined by

$$\begin{aligned}
&\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \le \sqrt{\frac{n-p}{2}} \cdot \left(\log(\hat{\sigma}^2) - \log(\sigma^2)\right) \le z_{1-\frac{\alpha}{2}}\right) \\
=&\mathbb{P}\left(\log(\hat{\sigma}^2) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}} \le \log(\sigma^2) \le \log(\hat{\sigma}^2) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}\right) \\
=&\mathbb{P}\left(\log(\hat{\sigma}^2) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}} \le \log(\sigma^2) \le \log(\hat{\sigma}^2) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}\right) \\
=&\mathbb{P}\left(\hat{\sigma}^2 \cdot e^{-z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}} \le \sigma^2 \le \hat{\sigma}^2 \cdot e^{z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}}\right).
\end{aligned}$$

Hence we have the corresponding CI

$$\left[\hat{\sigma}^2 \cdot \exp\left(-z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}\right), \hat{\sigma}^2 \cdot \exp\left(z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2}{n-p}}\right)\right].$$

Problem 2.

Set-up. Consider $y \sim N(X\beta, \sigma^2 I_n)$. Test $H_0 : \beta_S = 0$ against its alternative. The notation S is a subset of $\{0, 1, 2, ..., p-1\}$. For example, if $S = \{1, 2, 3\}$, then the null hypothesis becomes $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, which means we want to test whether the first three covariates are significant or not. We also use the notation s for the size of S and X_S is an $n \times s$ submatrix with columns in S taken from $X \in \mathbb{R}^{n \times p}$.

(For this question, I will introduce some of my own notations, and I will add them as the problem progresses.)

2(a).

Question. The LSE under H_0 is given by

$$\hat{\beta}_{H_0} = (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T y.$$

Find the distribution of $\hat{\beta}_{H_0}$ under H_0.

Solution. First, a new notation: let C be the 'choice matrix' such that $X_{S^c} = XC$, given S. It can be quickly verified that C has dimension $p \times (p-s)$, with an explicit form of

$$C = \begin{pmatrix} e_{j_1} & e_{j_2} & \cdots & e_{j_{(p-s)}} \end{pmatrix},$$

where e_{j_k}'s are the standard basis and $\{j_k\} = S^c$. Then we can write

$$\hat{\beta}_{H_0} = (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T y = (C^T X^T X C)^{-1} C^T X^T y.$$

This is a linear transformation of y and thus follows a normal distribution. It remains to calculate the expectation and the covariance matrix. First, note that under the null, we are essentially claiming

$$Y \sim N(XCC^T\beta, \sigma^2 I_n) \sim N(X_{S^c} \underbrace{C^T\beta}_{\beta_{S^c}}, \sigma^2 I_n).$$

Then

$$\hat{\beta}_{H_0} = (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T (X_{S^c}\beta_{S^c} + \sigma^2 Z).$$

To calculate the expectation, the first term is fixed and the second term vanishes (because $Z \sim N(0, I_n)$), so

$$\mathbb{E}[\hat{\beta}_{H_0}] = (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T X_{S^c} \beta_{S^c} = \beta_{S^c}.$$

For the covariance matrix, the first term vanishes, so

$$\begin{aligned}
\mathbb{V}[\hat{\beta}_{H_0}] &= (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T \cdot \sigma^2 I_n \cdot X_{S^c} (X_{S^c}^T X_{S^c})^{-1} \\
&= \sigma^2 (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T X_{S^c} (X_{S^c}^T X_{S^c})^{-1} \\
&= \sigma^2 (X_{S^c}^T X_{S^c})^{-1}.
\end{aligned}$$

Hence

$$\hat{\beta}_{H_0} \sim N\big(\beta_{S^c},\ \sigma^2 (X_{S^c}^T X_{S^c})^{-1}\big).$$

2(b).

Question. The fits under H_0 and H_1 are

$$\hat{y}_{H_0} = X_{S^c} \hat{\beta}_{H_0} = X_{S^c} (X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T y$$

and

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$$

with $\hat{\beta}$ being the LSE of the full model. Prove that

$$\|y - \hat{y}_{H_0}\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - \hat{y}_{H_0}\|^2.$$

Solution. One way to proceed is to show the cross-term is 0. However, I will just use a property already proven in class. Recall we have shown the Pythagorean theorem for LSE of the full model: specifically,

$$\|y - Xb\|^2 = \|y - X\hat{\beta}\|^2 + \|X\hat{\beta} - Xb\|^2,$$

for any $b \in \mathbb{R}^p$. Thus, to show the above equality is true, we just need to find the b that satisfies $\hat{y}_{H_0} = Xb$. But then

$$\hat{y}_{H_0} = X_{S^c}\hat{\beta}_{H_0} = XC\hat{\beta}_{H_0},$$

and we can set $b = C\hat{\beta}_{H_0}$. We can also verify the dimensions $p \times (p-s) \cdot (p-s) \times 1 = p \times 1$, and this b is a valid choice.

To show the results more explicitly, however, let us actually verify that the cross term is 0. First, denote

$$H = X(X^TX)^{-1}X^T,\ H_c = X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T.$$

Then

$$HH_c = X(X^TX)^{-1}X^TXC(C^TX^TXC)^{-1}C^TX^T = XC(C^TX^TXC)^{-1}C^TX^T = H_c. \quad (2.1)$$

$$H_cH = XC(C^TX^TXC)^{-1}C^TX^TX(X^TX)^{-1}X^T = XC(C^TX^TXC)^{-1}C^TX^T = H_c. \quad (2.2)$$

Therefore,

$$(I_n - H)(H - H_c) = H - HH - H_c + HH_c = H - H - H_c + H_c = 0_n,$$

and the cross term must be 0:

$$(y - \hat{y})^T(\hat{y} - \hat{y}_{H_0}) = ((I_n - H)y)^T(H - H_c)y = y^T(I_n - H)(H - H_c)y = 0.$$

2(c).

Question. Are $\|y - \hat{y}\|^2$ and $\|\hat{y} - \hat{y}_{H_0}\|^2$ independent? Why?

Solution. They are independent. We start by observing that

$$\|y - \hat{y}\|^2 = \|(I_n - H)y\|^2,\ \|\hat{y} - \hat{y}_{H_0}\|^2 = \|(H - H_c)y\|^2.$$

Both are linear transformations of y, so they are jointly Gaussian, and it remains to verify the parts $(I_n - H)y$ and $(H - H_c)y$ have covariance 0. One way to see this is from the Pythagorean equation from part (b). The two components have to be uncorrelated (which happens if and only if they are orthogonal), otherwise the equality should not hold due to the existence of a non-negligible cross-term. More concretely, by part (b),

$$\mathrm{Cov}((I_n - H)y, (H - H_c)y) = (I_n - H)\mathbb{V}[y](H - H_c) = \sigma^2(I_n - H)(H - H_c) = 0_n.$$

2(d).

Lemma 2.1. *Both H_c and $H - H_c$ are projection matrices.*

Proof. First, H_c is symmetric by structure, and it is idempotent since

$H_c H_c = XC(C^T X^T XC)^{-1} C^T X^T XC(C^T X^T XC)^{-1} C^T X^T = XC(C^T X^T XC)^{-1} C^T X^T = H_c.$

Next, the symmetry of $H - H_c$ follows directly from the symmetry of each component, so it remains to verify idempotence. By (2.1) and (2.2),

$$(H - H_c)(H - H_c) = HH - H_c H - HH_c + H_c H_c = H - H_c - H_c + H_c = H - H_c.$$

$\square$

Question. Find the distribution of $\|y - \hat{y}_{H_0}\|^2/\sigma^2$ and $\|\hat{y} - \hat{y}_{H_0}\|^2/\sigma^2$ under H_0.

Solution. First, under the null,

$$\begin{aligned}
y - \hat{y}_{H_0} &= (I_n - H_c)y = (I_n - H_c)(X_{S^c}\beta_{S^c} + \sigma Z) \\
&= (X_{S^c} - X_{S^c}(X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T X_{S^c})\beta_{S^c} + \sigma(I_n - H_c)Z \\
&= (X_{S^c} - X_{S^c})\beta_{S^c} + \sigma(I_n - H_c)Z \\
&= \sigma^2(I_n - H_c)Z.
\end{aligned}$$

Therefore,

$$\frac{\|y - \hat{y}_{H_0}\|^2}{\sigma^2} = \frac{\sigma^2\|(I_n - H_c)Z\|^2}{\sigma^2} = \|(I_n - H_c)Z\|^2 \sim \chi^2_{\text{rank}(I_n - H_c)}$$

by lemma from class. But then, since both are projection matrices

$$\text{rank}(I_n - H_c) = \text{Tr}(I_n - H_c) = \text{Tr}(I_n) - \text{Tr}(H_c) = n - (p - s) = n - p + s.$$

So

$$\frac{\|y - \hat{y}_{H_0}\|^2}{\sigma^2} \sim \chi^2_{n-p+s}.$$

Similarly, we have

$$\begin{aligned}
\hat{y} - \hat{y}_{H_0} &= (H - H_c)y = (H - H_c)(X_{S^c}\beta_{S^c} + \sigma Z) \\
&= [(X(X^T X)^{-1} X^T) - (X_{S^c} - X_{S^c}(X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T)](X_{S^c}\beta_{S^c} + \sigma^2 Z) \\
&= (X_{S^c} - X_{S^c})\beta_{S^c} + \sigma^2(X(X^T X)^{-1} X^T) - (X_{S^c} - X_{S^c}(X_{S^c}^T X_{S^c})^{-1} X_{S^c}^T)]Z \\
&= \sigma(H - H_c)Z.
\end{aligned}$$

It follows that

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2}{\sigma^2} = \frac{\sigma^2\|(H - H_c)Z\|^2}{\sigma^2} = \|(H - H_c)Z\|^2 \sim \chi^2_{\text{rank}(H - H_c)}.$$

Since

$$\text{rank}(H - H_c) = \text{Tr}(H - H_c) = \text{Tr}(H) - \text{Tr}(H_c) = p - (p - s) = s.$$

We conclude

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2}{\sigma^2} \sim \chi^2_s.$$

2(e).

Question. Construct an F-test for the testing problem.

Solution. Per results derived in class, we know

$$\|y - \hat{y}\|^2 \sim \chi^2_{n-p}.$$

Then by part (c), (d) and a theorem covered in class,

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2/(\sigma^2 s)}{\|y - \hat{y}\|^2/(\sigma^2(n-p))} = \frac{\|\hat{y} - \hat{y}_{H_0}\|^2/s}{\|y - \hat{y}\|^2/(n-p)} \sim F_{s,n-p}.$$

Thus, given a confidence level α, we can just reject when

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2/s}{\|y - \hat{y}\|^2/(n-p)} > F_{s,n-p,1-\alpha}.$$

2(f).

Question. Show that if $S = \{1, 2, ..., p-1\}$, we get the results that we have learned in the class.

Solution. If $S = \{1, 2, ..., p-1\}$, $S^c = \{0\}$, and then

$$X_{S^c} = XC = \mathbb{1}_n,\ \beta_{S^c} = C^T\beta = \beta_0,$$

which means

$$(X_{S^c}^T X_{S^c})^{-1} = \frac{1}{n},$$

and

$$\hat{y}_{H_0} = \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^T y = \overline{y}.$$

This allows us to recover the ANOVA decomposition (part b):

$$\left\|y - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^T y\right\|^2 = \|y - \hat{y}\|^2 + \left\|\hat{y} - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^T y\right\|^2.$$

We also recover

$$\frac{\left\|y - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^T y\right\|^2}{\sigma^2} \sim \chi^2_{n-p+(p-1)} = \chi^2_{n-1},$$

$$\frac{\left\|\hat{y} - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^T y\right\|^2}{\sigma^2} \sim \chi^2_s = \chi^2_{p-1}.$$

Hence

$$\frac{\|y - \hat{y}_{H_0}\|^2/p-1}{\|\hat{y} - \hat{y}_{H_0}\|^2/n-1} > F_{p-1,n-1,1-\alpha}.$$

Problem 3.

Set-up. Let $y_i \overset{\text{iid}}{\sim} N(\beta_0, \sigma^2)$ for $i = 1, ..., n$.

3(a).

Question. We fit the mean model and get

$$\hat{\beta}_0 = \underset{\beta_0}{\text{argmin}} \sum_{i=1}^{n} (y_i - \beta_0)^2.$$

Find the MSE $\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2]$.

Solution. If we take derivative w.r.t. β_0, we can recover $\hat{\beta}_0 = \overline{y}$, as discussed in class. We have, by linearity of expectations,

$$\mathbb{E}[\hat{\beta}_0] = \mathbb{E}[\overline{y}] = \frac{\sum_{i=1}^{n} \mathbb{E}[y_i]}{n} = \frac{n\beta_0}{n} = \beta_0,$$

so the estimator is unbiased. For variance,

$$\mathbb{V}[\hat{\beta}_0] = \mathbb{V}[\overline{y}] = \frac{\sigma^2}{n}.$$

Hence the MSE is

$$\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2] = \mathbb{V}[\hat{\beta}_0] + \mathbb{E}^2[\hat{\beta}_0 - \beta_0] = \frac{\sigma^2}{n}.$$

3(b).

Question. We take the covariates $x_1, ..., x_n$, fit a regression model, and get

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0}{\text{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Find the MSE $\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2]$.

Solution. The results we derived for univariate linear regression still holds regardless of the fact that $\beta_1 = 0$, so

$$\mathbb{E}[\hat{\beta}_0] = \beta_0,\ \mathbb{V}[\hat{\beta}_0] = \frac{\sigma^2}{n}\left(1 + \frac{\overline{x}^2}{V}\right),$$

where

$$V = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} > 0.$$

Hence

$$\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2] = \mathbb{V}[\hat{\beta}_0] + \mathbb{E}^2[\hat{\beta}_0 - \beta_0] = \frac{\sigma^2}{n}\left(1 + \frac{\overline{x}^2}{V}\right).$$

3(c).

Question. Since the data is generated by the simpler model $N(\beta_0, \sigma^2)$, both approaches - the simpler model in (a) or the more complicated model in (b) - are correct. However, they give different results in terms of estimation accuracy (variance). Discuss.

Solution. From parts (a) and (b), we can see that both estimators are unbiased. However, unless $\overline{x} = 0$, we have $\overline{x}^2/V > 0$, and the more complicated model in (b) would return larger variance (MSE) than the simple model in (a). Hence overfitting decreases estimator accuracy.

Problem 4.

Set-up. Let $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ independently for $i = 1, ..., n$.

Question. We fit the mean model and get

$$\hat{\beta}_0 = \underset{\beta_0}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0)^2.$$

Find the MSE $\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2]$.

Solution. The minimizing process is still the same as Problem 3, so the minimizer remains $\hat{\beta}_0 = \overline{y}$. However, we now have

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_0] &= \frac{\sum_{i=1}^{n} \mathbb{E}[y_i]}{n} = \beta_0 + \beta_1 \frac{\sum_{i=1}^{n} x_i}{n} = \beta_0 + \beta_1 \overline{x}, \\
\mathbb{V}[\hat{\beta}_0] &= \mathbb{V}\left[\frac{\sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}[y_i] = \frac{\sigma^2}{n}
\end{aligned}$$

by independence. Thus the estimator is now biased, and the MSE is

$$\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2] = \mathbb{V}[\hat{\beta}_0] + \mathbb{E}^2[\hat{\beta}_0 - \beta_0] = \frac{\sigma^2}{n} + \beta_1^2 \overline{x}^2.$$

4(b).

Question. We fit the regression model and get

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Find the MSE $\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2]$.

Solution. This is still the univariate results we derived in class, and the answers remain the same as part (b) of Problem 3:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_0] &= \beta_0 \\
\mathbb{V}[\hat{\beta}_0] &= \frac{\sigma^2}{n} \left(1 + \frac{\overline{x}^2}{V}\right), \\
\mathbb{E}[(\hat{\beta}_0 - \beta_0)^2] &= \frac{\sigma^2}{n} \left(1 + \frac{\overline{x}^2}{V}\right).
\end{aligned}$$

4(c).

Question. When the data is generated by the more complicated model, using the simpler one in (a) to fit the data may cause problems. Discuss.

Solution. The major problem here is that now the estimator for the simple model is biased by a magnitude of $\beta_1 \overline{x}$. This greatly decreases the quality of prediction. We can also compare the variance

$$\frac{\sigma^2}{n} + \beta_1^2 \overline{x}^2 \text{ versus } \frac{\sigma^2}{n} + \frac{\sigma^2 \overline{x}^2}{nV}.$$

Thus, if $\overline{x} > 0$ and

$$\beta_1^2 > \frac{\sigma^2}{nV},$$

then the variance of the simpler model is also larger than that of the more complicated model, which contrasts with the discussion in Problem 3(c). However, the primary concern lies with the expectation (first-order versus second-order concerns).

(I collaborated with Chad Schmerling and Justin Jung.)