midterm1WIP

Exercise 10. (a) Consider model (3) for $p = 2, X_i = (1, x_i)^\intercal, \theta^* = (\theta_1^*, \theta_2^*)^\intercal$ (similarly to section 1.5). Write explicit expressions for the confidence sets for $\theta^*, \theta_1^*, \theta_2^*$.

To set up explicit expression for the case above, we have:

$$XX^\intercal = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

and $det(XX^\intercal) = n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 = n\sum_{i=1}^n (x_i - \bar{x})^2$, and

$$(XX^\intercal)^{-1} = \frac{n}{det(XX^\intercal)} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

So we have

$$\hat{\theta} = (XX^\intercal)^{-1}XY = \frac{n}{det(XX^\intercal)} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = (\hat{\theta}_1, \hat{\theta}_2)^\intercal = \dots$$

$$\dots = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \bar{y}\sum_i x_i^2 - \bar{x}\sum_i x_i y_i \\ \sum_i x_i y_i - n\bar{y}\bar{x} \end{bmatrix} = (\hat{\theta}_1, \hat{\theta}_2)^\intercal = \hat{\theta}$$

To find a confidence region for θ^*, using a mixture of matrix and summation notation, we use the property:

$$\frac{||(XX^\intercal)^{1/2}(\hat{\theta} - \theta^*)||^2}{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta}_2 x_i)^2} \frac{n-2}{2} \sim F(2, n-2)$$

and denote $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta}_2 x_i)^2}{n-2}$. Where F denotes the F distribution with $df_1 = 2$, and $df_2 = n - 2$.

We can create a confidence interval for θ^*, such that, qF_α denotes the α^{th} quantile for $F(2, n-2)$.

$$P(\frac{||(XX^\intercal)^{1/2}(\hat{\theta} - \theta^*)||^2}{p\hat{\sigma}^2} < qF_{1-\alpha}) = 1 - \alpha = P((\hat{\theta} - \theta^*)^\intercal \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} (\hat{\theta} - \theta^*) < p\hat{\sigma}^2 qF_{1-\alpha})$$

We know that $\frac{(XX^\intercal)^{1/2}(\hat{\theta} - \theta^*)}{\sigma} \sim N(0, I_p)$. We can then set up confidence intervals for θ_1^* and θ_2^*.

For θ_1^*, we can set up a T-statistic by taking the difference of the first parameter estimate and the true estimate and dividing it the corresponding standard error:

$$T_{1(n-2-1)} = \frac{\hat{\theta_1} - \theta_1^*}{\sqrt{\sigma^2[(XX^\intercal)^{-1}]_{11}}} = \frac{\hat{\theta_1} - \theta_1^*}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta}_2 x_i)^2}{n-p} \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

Using T_1 we can set up a % $100(1-\alpha)$ confidence interval for $\hat{\theta_1^*}$ via:

$$\hat{\theta_1^*} \pm T_{1(n-3),\alpha/2} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta}_2 x_i)^2}{n-p} \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

For θ_2^* we have:

$$T_{2(n-3)} = \frac{\hat{\theta_2} - \theta_2^*}{\sqrt{\hat{\sigma^2}[(XX^\intercal)^{-1}]_{22}}} = \frac{\hat{\theta_2} - \theta_2^*}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta_2} x_i)^2}{n-p} \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

With T_2 we can set up a % $100(1-\alpha)$ confidence interval for $\hat{\theta_2^*}$ via:

$$\theta_2^* \pm T_{2(n-3),\alpha/2} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta_2} x_i)^2}{(n-p)\sum_{i=1}^n (x_i - \bar{x})^2}}$$

(b) Find a confidence interval for the expected response $E[Y_i]$ in the model in part (a). The variance of the expected response $var(\hat{Y}) = var(X^\intercal(XX^\intercal)^{-1}XY) = var(X^\intercal(XX^\intercal)^{-1}X(X^\intercal\theta^* + \varepsilon)) = var(X^\intercal(XX^\intercal)^{-1}X\varepsilon) = \sigma^2 X^\intercal(XX^\intercal)^{-1}X$. Using the standard error for $\hat{Y}$, we can set up up the following confidence interval for the expected response for the i^{th} record using a T-statistic:

$$T_{(n-3)} = \frac{\hat{y_i} - y_i}{\sqrt{\hat{\sigma^2} x_i^\intercal (XX^\intercal)^{-1} x_i}} = \frac{\hat{y_i} - y_i}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta_2} x_i)^2}{n-2} x_i^\intercal (XX^\intercal)^{-1} x_i}}$$

With this statistic a % $100(1-\alpha)$ confidence interval for y_i can be created:

$$y_i \pm T_{n-3,\alpha/2} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\theta_1} - \hat{\theta_2} x_i)^2}{n-2} x_i^\intercal \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} x_i}$$

Section 1.1

Exercise 3. Consider the linear regression model from exercise 1. Suppose, that the target of estimation is $h^\intercal\theta$ for some determinate non-zero vector $h \in R^p$. Find expression for the LSE of $h^\intercal\theta$. Is this estimate optimal in sense of Gauss-Markov theorem, i.e. does it have the smallest variance among all linear unbiased estimators?

—Start with this —By Gauss Markov, we know that a BLUE estimator has $Var(\theta_{OLS}) = \sigma^2(XX^\intercal)^{-1})$. However in the case of heterscedastic noise, we have $Var(\theta) = (XX^\intercal)^{-1}XDX^\intercal(XX^\intercal)^{-1}$, which must be greater than $\sigma^2(XX^\intercal)^-1)$. An so, in this case, our estimator is not BLUE. Study the same issue for the target $\eta = H^\intercal\theta$, where $H \in R^{q \times p}$ is some non-zero matrix with $q \le p$.

Section 1.3

Exercise 6. Let $L1, L2$ be some subspaces in R^n, and $L2 \subseteq L1 \subseteq R^n$. Let $PL1, PL2$ denote orthogonal projections on these subspaces. Prove the following properties:
(a) $PL2 - PL1$ is an orthogonal projection,
(b) $|PL2| \le |PL1|$ $\forall x \in R^n$,
(c) $PL2 \cdot PL1 = PL2$

Section 2.1

Exercise 7. (a) Using the notation from section 2.1, consider $X \sim N(\mu, I_n)$ for some $\mu \in R^n$. Find $E(Q(X))$ and $Var(Q(X))$

For $Q(X) = \sum_i \sum_j a_{ij} X_i X_j = X^\intercal AX, X \sim N(\mu, I_n)$, we have, using the property of trace operator:

$$E(Q(X)) = tr(E(Q(X)) = E(tr(Q(X)) = E(tr(X^\intercal AX)) = E(tr(AXX^\intercal)) = tr(AE(XX^\intercal))$$

Since $E(XX^\intercal) = I_n + \mu\mu^\intercal$, we have,

$$tr(AE(XX^\intercal)) = tr(A(I_n + \mu\mu^\intercal)) = trA + tr(A\mu\mu^\intercal) = trA + \mu^\intercal A\mu$$

$Var(Q(X)) =$

(b) Generalize the results from part (a) to the case $X \sim N(\mu, \Sigma)$ for some positive-definite covariance matrix $\Sigma \in R^{n\times n}$. For $X \sim N(\mu, \Sigma)$ we have,

$$E(Q(X)) = tr(AE(XX^\intercal)) = tr(A(\Sigma + \mu\mu^\intercal)) = tr(A\Sigma) + tr(A\mu\mu^\intercal) = tr(A\Sigma) + \mu^\intercal A\mu$$

$Var(Q(X)) =$

Section 2.2

Exercise 9. In the Gaussian linear regression model 3, consider the target of estimation $\eta = H^\intercal\theta^*$, where $H \in R^{q\times p}$ is some non-zero matrix with $q \leq p$. Find an analogue of the quadratic form $S2$ (from (4)) for the new target η^*, and prove for the new quadratic form statements similar to (e) from Theorem 2.1, and Corollary 2.1.2.

Exercise 11. Find an elliptical confidence set for the expected response $E[Y]$ in model (3).

Exercise 12. Construct simultaneous confidence intervals (e.g., as in Corollary 2.2.1) for the expected responses $E[Y_1], ..., E[Y_n]$ in model (3).