Homework 5

To submit via e-mail by h 14:00, July 15

Let X be an $n \times 3$ matrix of rank 3, where the 1st column of X is the intercept term, i.e. $X_{\{1\}} = 1_n$, which is always included in the model. Below the R code to use to generate X:

```
n = 20
x1 = 1:n
set.seed(123)
x2 = rnorm(n)
x3 = rnorm(n,x2, 0.1^2)
X = cbind(x1,x2,x3)
```

Assume a Gaussian model $y \sim N_n(X\beta, \sigma^2 I_n)$ where $\underset{n\times 1}{\mu} = X\beta$ is the mean vector and $\underset{p\times 1}{\beta} = (2,1,1)^\mathsf{T}$ and $\sigma^2 = 1$ are the true model parameters.

1. Compute the prediction error PE_M for the full model $M = \{1,2,3\}$ (including all the predictors) and for the submodel $M = \{1,2\}$ (which does not include the 3rd predictor). For the submodel, calculate the true values of $\underset{2\times 1}{\beta_M} = (X_M^\mathsf{T} X_M)^{-1} X_M^\mathsf{T}\mu$. Comment on the results.

2. Suppose that we want inference for the 2nd predictor but we don't know whether or not include the 3rd, that is, the candidate models are

$$\mathcal{M} = \{\{1,2\},\{1,2,3\}\}$$

Use the model selector $\hat{M}$, which is $\hat{M} = \{1,2,3\}$ if $|\hat{\beta}_{3\cdot\{1,2,3\}}|/\sigma\sqrt{v_{3\cdot\{1,2,3\}}}$ is larger than $z_{1-\alpha/2} = 1.959964$, and $\hat{M} = \{1,2\}$ otherwise, where $v_{i\cdot M}$ is the ith diagonal element of $V_M = (X_M^\mathsf{T} X_M)^{-1}$ and $\alpha = 0.05$. We are interested in the coverage probability of the interval (that ignores the selection)

$$\mathrm{CI}_2 = \hat{\beta}_{2\cdot\hat{M}} \pm z_{1-\alpha/2}\sigma\sqrt{v_{2\cdot\hat{M}}}$$

in two scenarios:

- the target parameter $\beta_{2\cdot\{1,2,3\}}$ is fixed, i.e.

$$\mathrm{P}(\beta_{2\cdot\{1,2,3\}} \in \mathrm{CI}_2)$$

- the target parameter $\beta_{2\cdot\hat{M}}$ is random, i.e.

$$\mathrm{P}(\beta_{2\cdot\hat{M}} \in \mathrm{CI}_2)$$

Write the R code to evaluate via Monte-Carlo simulation the probabilities $\mathrm{P}(\beta_{2\cdot\{1,2,3\}} \in \mathrm{CI}_2)$ and $\mathrm{P}(\beta_{2\cdot\hat{M}} \in \mathrm{CI}_2)$. Comment the results. What is the theoretical coverage probability if there is no selection, i.e. we have always $\hat{M} = \{1,2,3\}$?