STAT151A Homework 4: Due March 8th

Your name here

Chi squared random variables

Let $s \sim \chi_K^2$. Prove that

- $\mathbb{E}[s] = K$
- $\mathrm{Var}(s) = 2K$ (hint: if $z \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}[z^4] = 3\sigma^4$)
- If $a_n \sim \mathcal{N}(0, \sigma^2)$ IID for $1, \dots, N$, then $\frac{1}{\sigma^2} \sum_{n=1}^{N} a_n^2 \sim \chi_N^2$
- $\frac{1}{K} s \to 1$ as $K \to \infty$
- $\frac{1}{\sqrt{K}}(s - K) \to \mathcal{N}(0, 2)$ as $K \to \infty$
- Let $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ where $a \in \mathbb{R}^K$. Then $\|\boldsymbol{a}\|_2^2 \sim \chi_K^2$
- Let $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ where $a \in \mathbb{R}^K$. Then $\boldsymbol{a}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{a} \sim \chi_K^2$

Predictive variance for different regressors

This question will take the residuals of the training data to be random, and will consider variablity under sampling of the training data. The regressors for both the training data and test data will be taken as fixed.

Let $\boldsymbol{x}_n = (x_{n1}, x_{n2})^\intercal$ be IID normal regressors, with

- $\mathbb{E}[x_{n1}] = \mathbb{E}[x_{n2}] = 0$,
- $\mathrm{Var}(x_{n1}) = \mathrm{Var}(x_{n2}) = 1$, and
- $\mathrm{Cov}(x_{n1}, x_{n2}) = 0.99$.

(Note there is no intercept.)

Assume that $y_n = \beta^\intercal \boldsymbol{x}_n + \varepsilon_n$ for some β, and that the residuals ε_n are IID with mean 0, variance $\sigma^2 = 2$, and are independent of $\boldsymbol{x}_n$.

(a)

Find the limiting distribution of $\sqrt{N}(\hat{\beta} - \beta)$.

(b)

Define the expected prediction error

$$\hat{y}_{\text{new}} - \mathbb{E}\left[y_{\text{new}}\right] := (\hat{\beta} - \beta)^{\mathsf{T}} x_{\text{new}},$$

and approximate the limiting variance $\text{Var}\left(\hat{y}_{\text{new}} - \mathbb{E}\left[y_{\text{new}}\right]\right)$ for the following new regression vectors:

- $x_{\text{new}} = (1, 1)^{\mathsf{T}}$
- $x_{\text{new}} = (1, -1)^{\mathsf{T}}$
- $x_{\text{new}} = (100, 100)^{\mathsf{T}}$
- $x_{\text{new}} = (0, 0)^{\mathsf{T}}$

You may assume that N is large, so that you can apply the CLT to $\sqrt{N}(\hat{\beta} - \beta)$. Even with the CLT approximation your answer will depend on N; just make this dependence explicit.

(c)

Why are some variances in (b) large and some small? Explain each in plain language and intuitive terms.

The sandwich covariance matrix under homoeskedasticity

For this problem, make the following assumptions.

- The regressors are non-random, with $\frac{1}{N} \sum_{n=1}^{N} x_n x_n^{\mathsf{T}} \to \Sigma_X$ for positive definite Σ_X
- The responses are $y_n = \beta^{\mathsf{T}} x_n + \varepsilon_n$ for some unknown β
- The residuals are IID with $\mathbb{E}\left[\varepsilon_n\right] = 0$ and $\text{Var}\left(\varepsilon_n\right) = \sigma^2$ (but not necessarily normal)

Under these assumptions, show that the sandwich covariance matrix and the standard covariance matrix converge to the same quantity. That is, show that

$$\hat{\Sigma}_{sand} = N \left(X^{\mathsf{T}} X\right)^{-1} \left(\sum_{n=1}^{N} x_n x_n^{\mathsf{T}} \hat{\varepsilon}_n^2\right) \left(X^{\mathsf{T}} X\right)^{-1} \to S \quad \text{and} \quad \hat{\Sigma}_h = N \left(X^{\mathsf{T}} X\right)^{-1} \hat{\sigma}^2 \to S$$

for the same S, where $\hat{\sigma}^2 := \frac{1}{N} \sum_{n=1}^{N} \hat{\varepsilon}_n^2$.