STAT151A Homework 4: Due March 8th

Your name here

Chi squared random variables

Let $s \sim \chi^2_K$. Prove that

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

Predictive variance for different regressors

This question will take the training data to be random, and will consider variablity under sampling of the training data.

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

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

(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

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

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

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

(Your answers 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

Assume homeskedastic errors; that is, that ε_n is independent of $\boldsymbol{x}_n$, with $\mathbb{E}\left[\varepsilon_n | \boldsymbol{x}_n\right] = 0$ and $\mathbb{E}\left[\varepsilon_n | \boldsymbol{x}_n\right] = \sigma^2$ for all n.

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

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

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