STAT151A Homework 4: Due March 8th

Your name here

1 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 $\boldsymbol{a}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{a} \sim \chi^2_K$

Solutions:

Write $s = \sum_{k=1}^{K} z_k^2$ for $z_k \sim \mathcal{N}\left(0, 1\right)$ IID.

- $\mathbb{E}\left[\sum_{k=1}^{K} z_k^2\right] = \sum_{k=1}^{K} 1 = K$
- $\mathbb{E}\left[\left(\sum_{k=1}^{K}(z_k^2 - 1)\right)^2 = K\mathbb{E}\left[z_k^2 - 1\right]^2\right] = K(\mathbb{E}\left[z_k^4\right] - 2\mathbb{E}\left[z_k^2\right] + 1) = K(3 - 2 + 1) = 2K.$
- This follows from the LLN
- This follows from the CLT
- Direct computation
- Direct computation

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 $\sqrt{N}(\hat{\beta} - \beta)$.

Solution:

By assumption,

$$\frac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{X} \rightarrow \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix} =: \boldsymbol{\Sigma}_{\boldsymbol{X}}.$$

Note that

$$\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} = \frac{1}{1 - 0.99^2} \begin{pmatrix} 1 & -0.99 \\ -0.99 & 1 \end{pmatrix} \approx 50.3 \begin{pmatrix} 1 & -0.99 \\ -0.99 & 1 \end{pmatrix}.$$

Plugging in the definition of $\hat{\beta}$,

$$\hat{\beta} - \beta = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y} - \beta = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}.$$

By the non-IID CLT, $\frac{1}{\sqrt{N}}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon} \rightarrow \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{X}}\right)$.

By the continuous mapping theorem,

$$\sqrt{N}(\hat{\beta} - \beta) \rightarrow \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right) = \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right) \approx \mathcal{N}\left(\boldsymbol{0}, \begin{pmatrix} 100.5 & -99.5 \\ -99.5 & 100.5 \end{pmatrix}.\right).$$

If the students just used the limiting distribution from lecture without proving it, that's fine.

(b)

Define the expected prediction error

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

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

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

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.

Solution:

$\mathrm{Var}\left(\hat{y}_{\mathrm{new}} - \mathbb{E}\left[y_{\mathrm{new}}\right]\right) = x_{\mathrm{new}}^{\intercal} \mathrm{Cov}\left(\hat{\beta}\right) x_{\mathrm{new}}$. Plugging into `R` (or whatever), we get (approximately)

- 2.01
- 400
- 20100
- 0

(c)

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

Solutions:

- x_{n1}, x_{n2} are highly correlated. So $\hat{\beta}$ is well-estimated when both of them change the same way, and poorly estimated when they change independently, since independent changes aren't in the dataset.
- In general, taking $\boldsymbol{x}_n \mapsto c\boldsymbol{x}_n$ increases the variance by 0.
- At $\mathbf{0}$, the prediction is always $\mathbf{0}$, which doesn't depend on $\hat{\beta}$, and so has no variance.

3 The sandwich covariance matrix under homoeskedasticity

For this problem, make the following assumptions.

- The regressors are non-random, with $\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\intercal \rightarrow \boldsymbol{\Sigma_X}$ for positive definite $\boldsymbol{\Sigma_X}$
- The responses are $y_n = \boldsymbol{\beta}^\intercal \boldsymbol{x}_n + \varepsilon_n$ for some unknown $\boldsymbol{\beta}$
- The residuals are IID with $\mathbb{E}\left[\varepsilon_n\right] = 0$ and $\mathrm{Var}\left(\varepsilon_n\right) = \sigma^2$ (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

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

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

Solution:

The key is that by the non-IID LLN,

$$\frac{1}{N}\sum_{n=1}^{N} x_n x_n^\intercal \varepsilon_n^2 \rightarrow \lim_{N\rightarrow\infty}\frac{1}{N}\sum_{n=1}^{N} x_n x_n^\intercal \mathbb{E}\left[\varepsilon_n^2\right] = \sigma^2 \lim_{N\rightarrow\infty}\frac{1}{N}\sum_{n=1}^{N} x_n x_n^\intercal = \sigma^2\boldsymbol{\Sigma_X}.$$

The students do not need to show carefully that $\frac{1}{N}\sum_{n=1}^{N} x_n x_n^\intercal \varepsilon_n^2$ and $\frac{1}{N}\sum_{n=1}^{N} x_n x_n^\intercal \hat{\varepsilon}_n^2$ have the same limit.

Given this, everything cancels and the limits are the same.