STAT151A Quiz 4 (Mar 12th)

Please write your full name and email address:

The OLS estimator is given by $\hat{\beta} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$.

You have 20 minutes for this quiz.

There are three parts, (a), (b), and (c), each weighted equally..

This quiz will use some facts about the standard normal distribution. If $z \sim \mathcal{N}(0, 1)$, then:

- $\mathbb{E}[z] = 0$
- $\mathbb{E}[z^2] = 1$
- $\mathbb{E}[z^3] = 0$
- $\mathbb{E}[z^4] = 3$

Recall that the OLS estimator of $y_n \sim \beta^\intercal \boldsymbol{x}_n$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$.

(a)

For this quiz, we will take $a_n \sim \mathcal{N}(0, 1)$ IID. Take $b_n = a_n^2$. Note that the pairs (a_n, b_n) are IID, but a_n and b_n are not independent.

Let $\boldsymbol{x}_n = (a_n, b_n)^\intercal$. Let $\boldsymbol{\beta} = (\beta_a, \beta_b)^\intercal$. Let $y_n = \boldsymbol{\beta}^\intercal \boldsymbol{x}_n + \varepsilon_n$, where ε_n are IID with $\mathbb{E}[\varepsilon_n] = 0$ and $\mathrm{Var}(\varepsilon_n) < \infty$.

Note that a_n, b_n, β_a, and β_b are all scalars, and both $\boldsymbol{x}_n$ and $\boldsymbol{\beta}$ are 2–vectors.

Let $\hat{\boldsymbol{\beta}}$ denote the OLS estimator of $y_n \sim \boldsymbol{\beta}^\intercal \boldsymbol{x}_n$, that is, of y_n regressed on both a_n and b_n. **Note that the regression for $\hat{\boldsymbol{\beta}}$ does not include a constant.**

Let $\hat{\alpha}$ denote the OLS estimator of $y_n \sim \alpha a_n$, that is, of y_n regressed on a_n alone. **Note that the regression for $\hat{\gamma}$ does not include a constant, and does not include b_n.**

Prove that, as $N \to \infty$,

- $\hat{\alpha} \to \beta_a$ and
- $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta} = \begin{pmatrix} \beta_a \\ \beta_b \end{pmatrix}$.

(b)

For part (b), let $\hat{y}_{\text{new}} = \beta_a a_{\text{new}}$, and assume that $\beta_b \neq 0$. **Note that $\hat{y}_{\text{new}}$ is formed with β_a, not $\hat{\alpha}$.** Prove that

$$\mathbb{E}\left[y_{\text{new}} - \hat{y}_{\text{new}}\right] \neq 0,$$

where the expectation is taken over a_{new}, b_{new}, and ε_{new}. That is, when you exclude b_n from the regression, the predictions evaluated at the limit β_a are biased.

(c)

For part (c), let $\hat{y}'_{\text{new}} = \boldsymbol{\beta}^\intercal \boldsymbol{x}_{\text{new}}$. **Note that $\hat{y}'_{\text{new}}$ is formed with $\boldsymbol{\beta}$, not $\hat{\boldsymbol{\beta}}$.** Prove that

$$\mathbb{E}\left[y'_{\text{new}} - \hat{y}_{\text{new}}\right] = 0,$$

where the expectation is taken over a_{new}, b_{new}, and ε_{new}. That is, when you include b_n from the regression, the predictions evalutead at the limit $\boldsymbol{\beta}$ are unbiased.