General notes

Exam will contain 6 problems. All the problems have equal weights. Midterm duration is 120 minutes. Closed book, one A4 cheatsheet is allowed.

Demo variant «Santa»

1. In a model $y_i = \beta_0 + \beta_1 x_i + u_i$ all Gauss — Markov assumptions are satisfied except $\mathbb{V}\mathrm{ar}(u_i \mid x) = \sigma^2 x_i^4$.

   (a) Derive the most efficient unbiased estimator for β.

   (b) Derive unbiased and consisent estimator for σ^2.

2. Consider the simple regression with L^2-penalty with three observations, $y = (1, 3, 5)$, $x = (1, 2, 2)$. The loss function to be minimized is given by

$$\mathrm{loss}(\hat{\beta}_0, \hat{\beta}_1) = \sum_i (y_i - \hat{y}_i)^2 + \lambda \cdot \hat{\beta}_1^2, \quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

   (a) Find estimator $\hat{\beta}_1$ for any $\lambda \geq 0$.

   (b) Find $\mathbb{E}(\hat{\beta}_1 \mid x)$ and $\mathbb{V}\mathrm{ar}(\hat{\beta}_1 \mid x)$.

   (c) Find the mean squared error $MSE(\hat{\beta}_1)$.

3. The random variables y_1, ..., y_n are independent and identically distributed on $[a; 1]$, where $a \in (0; 1)$ is an unknown parameter. Consider a bootstrap sample y_1^*, ..., y_n^*.

   (a) Find the probabilities $\mathbb{P}(\min y^* > \min y)$, $\mathbb{P}(\min y^* > a)$.

   (b) Describe how to construct naive bootstrap confidence interval for a.

   (c) Find the real coverege probability of a nominal 95% naive bootstrap confidence interval for a.

4. All regressors in X matrix are standardized. There are 3 regressors and 53940 observations. You partially know the singular value decomposition of X:

$$X = UDV^T, \mathrm{diag}(D) = (1.301 \cdot 10^6, 1.022 \cdot 10^4, 42.58), V = \begin{pmatrix} 0.000156 & 0.005953 & -0.999982 \\ 0.007737 & 0.999952 & 0.005954 \\ 0.999970 & -0.007737 & 0.000109 \end{pmatrix}.$$

   (a) Express the first two principal components as the functions of original predictors.

   (b) Find the sample variance of each principal component.

   (c) How much variance does the first principal component explain?

   (d) How much variance do the first two principal components explain?

5. Consider the model $y_i = \beta_0 + \beta_a a_i + \beta_b b_i + u_i$. Let X be the matrix of all regressors including constant, $n = 53940$,

$$(X^T X)^{-1} = \begin{bmatrix} 0.034465493 & -0.000018274 & -0.000557615 \\ -0.000018274 & 0.000082578 & -0.000000771 \\ -0.000557615 & -0.000000771 & 0.000009040 \end{bmatrix}$$

(a) Find the coefficient estimates in the regression $\hat{b}_i = \hat{\gamma}_0 + \hat{\gamma}_1 a_i$, the corresponding R^2_b and VIF_b.

(b) Find the variance of $\hat{\beta}_b$ in the original model assuming $\mathbb{V}\mathrm{ar}(u_i \mid X) = \sigma^2$.

6. Students choose the stochastic calculus book in the library. There are two types of the book: the green one and the red one, the variable g_i is an indicator with $g_i = 1$ when the i-th student wants the green book. There are too many copies of the red book and too few copies of the green book. So the variable a_i is an indicator of the availability: $a_i = 1$ if the green book is available when the i-th student arrives and $a_i = 0$ otherwise. The student will take the book he/she likes if it is available and the red book otherwise.

   The stochastic calculus exam result y_i depends on the book and unobserved student characteristics u_i,

$$y_i = \beta_0 + \beta_1 g_i a_i + u_i.$$

   Here $\mathbb{C}\mathrm{ov}(g_i, u_i) \neq 0$ but $\mathbb{C}\mathrm{ov}(a_i, u_i) = 0$.

   (a) Find plim $\hat{\beta}_1^{\text{ols}}$ from regression $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 g_i a_i$. Is it consistent?

   (b) Find plim $\hat{\beta}_1^{\text{iv}}$ for the same regression with a_i as instrument.