4. Heteroscedastic Regression

For the next three problems, consider the following setting.

We measure characteristics of n individuals, sampled randomly from a population. Let (X_i, y_i) be the observed data of the ith individual, where $y_i \in \mathbb{R}$ is the dependent variable and $X_i \in \mathbb{R}^p$ is the vector of p **deterministic** explanatory variables. Our goal is to estimate the coefficients of $\beta = (\beta_1, \ldots, \beta_p)^T$ in the linear regression:

$$y_i = X_i^T \beta + \epsilon_i, \qquad i = 1, \ldots, n$$

We will consider the case where the model is potentially **heteroscedastic** (i.e. the error terms ϵ_i are **not** i.i.d.).

More specifically, assume that the vector $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ is a n-dimensional Gaussian with mean 0 and known **nonsingular** covariance matrix Σ. Denote by $\mathbb{X}$ the matrix in $\mathbb{R}^{n \times p}$ whose rows are $\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T$ and by $\mathbf{Y}$ the vector with coordinates $y_1, \ldots, y_n$.

Instead of the usual Least Squares Estimator, **instead consider the estimator $\hat{\beta}$ that minimizes, over all $\beta \in \mathbb{R}^p$,**

$$(\mathbf{Y} - \mathbb{X}\beta)^T \Sigma^{-1} (\mathbf{Y} - \mathbb{X}\beta).$$

(a) A Generalized Estimator

1/1 point (graded)

Let I_n be the $n \times n$ identity matrix. If $\Sigma = \sigma^2 I_n$ (i.e. homoscedastic ε) for some $\sigma^2 > 0$, then which one of the following statement about $\hat{\beta}$ **must** be true? **Make no assumptions about the rank of $\mathbb{X}$.**

- ( ) $\hat{\beta}$ has positive entries.
- (•) $\hat{\beta}$ is the least squares estimator.
- ( ) $\hat{\beta}$ is the unique minimizer of the specified loss.
- ( ) The components of $\hat{\beta}$ are independent.

✔

Submit

You have used 1 of 3 attempts

(b) The Maximum Likelihood Estimator

1/1 point (graded)

In this exercise, we will prove that $\hat{\beta}$ is equal to the Maximum Likelihood Estimator, even for general Σ. Recall the form of the n-dimensional Gaussian density from Lecture 10.

Let Σ be an arbitrary $n \times n$ matrix. The maximum likelihood estimator β_{MLE} is the value of β **maximizes** the log-likelihood function $\ell(\beta) = \ln L(\mathbb{X}, \mathbb{Y}; \beta)$.

Write down the function ℓ, in terms of $\mathbb{X}$, $\mathbf{Y}$, β, Σ, and n.

(Type **X** for $\mathbb{X}$, **Y** for $\mathbf{Y}$, **Sigma** for Σ. Type **trans(X)** for the transpose $\mathbb{X}^T$, **det(X)** for the determinant $\det \mathbb{X}$, and **X^(-1)** for the inverse $\mathbb{X}^{-1}$, of a matrix $\mathbb{X}$.)

$\ell(\beta) = \ln L(\mathbb{X}, \mathbb{Y}; \beta) =$ -(n/2)*ln(2*pi)-ln(det(Sigma))/2-1/2*(trans(Y-X*beta))*Sigma^(-1)*(Y-X*beta)

STANDARD NOTATION

Submit You have used 4 of 4 attempts

Correct (1/1 point)

(c)

2/2 points (graded)

Assume that the rank of $\mathbb{X}$ is p. (As a consequence, for any nonsingular $n \times n$ matrix Q, the $p \times p$ matrix product $X^T Q X$ is also nonsingular.)

Which of the following is a correct formula for $\hat{\beta}$ in terms of $\mathbb{X}$, $\mathbb{Y}$ and Σ?

- (selected) $(\mathbb{X}^T \Sigma^{-1} \mathbb{X})^{-1} \mathbb{X}^T \Sigma^{-1} \mathbb{Y}$
- $(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbb{Y}$
- $(\mathbb{X}^T \Sigma^{-1} \mathbb{X})^{-1} \mathbb{Y}^T \Sigma^{-1} \mathbb{X}$
- $\mathbb{X}^{-1} \mathbb{Y}$

Using the result from the above, which of the following correctly characterizes the distribution of $\hat{\beta}$?

- $\mathcal{N}(0, \Sigma)$
- $\mathcal{N}\left(0, \left(\mathbb{X}^T \Sigma^{-1} \mathbb{X}\right)^{-1}\right)$
- $\mathcal{N}\left(0, \left(\mathbb{X}^T \mathbb{X}\right)^{-1}\right)$
- $\mathcal{N}\left(\beta, \left(\mathbb{X}^T \Sigma^{-1} \mathbb{X}\right)^{-1}\right)$
- $\mathcal{N}\left(\beta, \left(\mathbb{X}^T \mathbb{X}\right)^{-1}\right)$

✔

Submit You have used 2 of 3 attempts

Discussion

Topic: Unit 6 Linear Regression:Homework 10 Linear regression / 4. Heteroscedastic Regression

Hide Discussion

Add a Post

All Posts

(b) The Maximum Likelihood Estimator incomplete log-likelihood?

question posted about 6 hours ago by **sandipan_dey**

For MLE, we have log-likelihood function of the form $l(X, Y; \beta, p) = f_1(p) + f_2(\Sigma) + f_3(X, Y, \beta, \Sigma)$. But grader does not accept $f_1(p)$ and when I entered $f_2(\Sigma) + f_3(X, Y, \beta, \Sigma)$ it got rejected. Should we get rid of the $f_2(\Sigma)$ part too? Also, as per the initial definition we have $Y = X^T\beta + \epsilon$, for MLE we are asked to minimize $(Y - X\beta)^T \Sigma^{-1} (Y - X\beta)$ (notice it's $X\beta$ and NOT $X^T\beta$ here), should not it be $X^T\beta$ instead in the objective function?

This post is visible to everyone.

Add a Response

1 response

Erocha (Community TA)
about 6 hours ago

There is no p in this question. Note that the input is $\mathbb{X}$ (that you type as **X**), so it is $\mathbb{X}\beta$ and there is no transpose. Please, make sure you understood the text on the top.

@Erocha -- DELETED by SD -- here it's $\mathcal{N}_n\left(.\right)$ and not $\mathcal{N}_p\left(.\right)$ - it's in the lecture already, my bad, thank you for helping.

thank you.

posted about 6 hours ago by **sandipan_dey**

Add a comment

Preview

Submit

Showing all responses

Add a response:

Preview

Submit