476 Statistics, Spring 2022.
Lecturer: Sergey Nadtochiy.

Lecture 12. LS method for a multidimensional linear regression. (Sections 11.10–11.15)

We study

$$Y = X\beta + \varepsilon,$$

where $X = (1, X^1, \dots, X^k)$ is a (row) random vector with values in $\mathbb{R}^{k+1}$, with the first entry being one, and $\beta = (\beta_0, \dots, \beta_k)^\top \in \mathbb{R}^{k+1}$ is an unknown (column) vector of regression coefficients.

As usual, we always assume that ε is indep. of X, and that ε is mean zero.

We denote observations by $\{X_i, Y_i\}_{i=1}^n$ and always assume that the residuals $\{\varepsilon_i = Y_i - X_i\beta\}$ are i.i.d. We denote by $\tilde{X}$ the matrix whose i-th row is given by X_i, and by $\tilde{Y}$ the (column) vector whose i-th entry is Y_i.

The least-squares (LS) estimator of β, denoted $\hat{\beta}$, is the value of β that attains the minimum in

$$\min_\beta \sum_{i=1}^n (Y_i - \sum_{j=0}^k \beta_j X_i^j)^2 = \min_\beta \|\tilde{Y} - \tilde{X}\beta\|^2 = \min_\beta (\tilde{Y} - \tilde{X}\beta)^\top (\tilde{Y} - \tilde{X}\beta)$$

Thm 1. *Assume that $\{\varepsilon_i\}$ are i.i.d. normal, independent of $\tilde{X}$. Then, the LS estimator $\hat{\beta}$ is a MLE for β.*

Let us compute $\hat{\beta}$. The usual differentiation rules work in the matrix notation:

$$0 = \frac{d}{d\beta}(\tilde{Y} - \tilde{X}\beta)^\top (\tilde{Y} - \tilde{X}\beta) = \frac{d}{d\beta}(-2\beta^\top \tilde{X}^\top \tilde{Y} + \beta^\top \tilde{X}^\top \tilde{X}\beta) = 2\tilde{X}^\top \tilde{X}\beta - 2\tilde{X}^\top \tilde{Y}.$$

If $\tilde{X}^\top \tilde{X}$ is invertible, we have

$$\hat{\beta} = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top \tilde{Y}$$

If not, $\hat{\beta}$ is determined as a solution to

$$\tilde{X}^\top \tilde{X}\hat{\beta} = \tilde{X}^\top \tilde{Y}.$$

Lemma 1. *There always exists a solution to the lin. system of eq-ns for $\hat{\beta}$.*

Ex 1. *Let us recover the familiar expression for $\hat{\beta}$ when $k = 1$. Denote the i-th observation of the explanatory variable by $X_i = (1, Z_i)$. Then,*

$$\tilde{X}^\top \tilde{X} = \begin{pmatrix} n & \sum Z_i \\ \sum Z_i & \sum Z_i^2 \end{pmatrix},$$

$$\tilde{X}^\top \tilde{Y} = \begin{pmatrix} \sum Y_i \\ \sum Z_i Y_i \end{pmatrix}$$

Let us verify that

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{Z}, \quad \hat{\beta}_1 = S_{zy}/S_{zz}$$

solve the desired system of eq-ns:

$$(\tilde{X}^\top \tilde{X})\hat{\beta} = \begin{pmatrix} n\bar{Y} - n\hat{\beta}_1\bar{Z} + \sum Z_i S_{zy}/S_{zz} \\ \sum Z_i(\bar{Y} - \hat{\beta}_1\bar{Z}) + \sum Z_i^2 S_{zy}/S_{zz} \end{pmatrix},$$

which needs to be solved by $\hat{\beta} = (\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{Y}$. *Let us verify that the first entry of the vector in the right hand side of the above equals the first entry of* $\tilde{X}^\top\tilde{Y}$. *We notice that* $\tilde{X}^\top\tilde{Y} = (\sum Y_i, \sum Z_i Y_i)^\top$. *On the other hand,*

$$n\bar{Y} - n\hat{\beta}_1\bar{Z} + \sum Z_i \frac{S_{zy}}{S_{zz}} = n\bar{Y} - n\hat{\beta}_1\bar{Z} + \sum Z_i\hat{\beta}_1 = n\bar{Y} = \sum Y_i.$$

We leave the verification of the fact that the second entries are equal as an exercise.

1 Properties of $\hat{\beta}$

Thm 2. *Assume* $\mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top$ *is well defined. Then,*

$$\mathbb{E}\hat{\beta} = \beta$$

Proof:

$$\mathbb{E}\hat{\beta} = \mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top(\tilde{X}\beta + \varepsilon) = \beta + \mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\mathbb{E}\varepsilon = \beta.$$

■

To make a more accurate inference about β (i.e., to construct confidence intervals and to test hypotheses), we need to know further distributional characteristics of $\hat{\beta}$. One of such characteristics is the covariance matrix of $\hat{\beta}$. Recall that, for any random vector ξ with values in $\mathbb{R}^m$, its covariance matrix Σ is given by

$$\Sigma_{ij} = \mathbb{E}(\xi^i - \mathbb{E}\xi^i)(\xi^j - \mathbb{E}\xi^j), \quad i,j = 1,\dots,m.$$

Thm 3. *Assume* $\mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top$ *and* $\mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}$ *are well defined. Then, the covariance matrix of* $\hat{\beta}$ *is*

$$cov(\hat{\beta}) = \sigma^2\mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}$$

Proof:

By conditioning on $\tilde{X}$,

$$\mathbb{E}(\hat{\beta}-\beta)(\hat{\beta}-\beta)^\top = \mathbb{E}\left[\mathbb{E}\left((\hat{\beta}-\beta)(\hat{\beta}-\beta)^\top \,|\, \tilde{X}\right)\right] = \mathbb{E}\left[\mathbb{E}\left((\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\varepsilon\varepsilon^\top\tilde{X}(\tilde{X}^\top\tilde{X})^{-1} \,|\, \tilde{X}\right)\right]$$

$$= \mathbb{E}\left((\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\mathbb{E}(\varepsilon\varepsilon^\top \,|\, \tilde{X})\tilde{X}(\tilde{X}^\top\tilde{X})^{-1}\right) = \mathbb{E}\left((\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\mathbb{E}(\varepsilon\varepsilon^\top)\tilde{X}(\tilde{X}^\top\tilde{X})^{-1}\right) = \sigma^2\mathbb{E}(\tilde{X}^\top\tilde{X})^{-1}.$$

■

Ex 2. *For* $k = 1$, $X_i = (1, Z_i)$,

$$(\tilde{X}^\top\tilde{X})^{-1} = \begin{pmatrix} \frac{\sum Z_i^2}{nS_{zz}} & -\frac{\bar{Z}}{S_{zz}} \\ -\frac{\bar{Z}}{S_{zz}} & \frac{1}{S_{zz}} \end{pmatrix},$$

and we recover the known expressions for $V(\hat{\beta}_0)$ *and* $V(\hat{\beta}_1)$ *for deterministic* $\{Z_i\}$.

Thm 4. *Assume that* $\{X_i\}_{i=1}^n$ *are i.i.d. with well defined and invertible covariance matrix. Then* $\hat{\beta}$ *is consistent.*

Thm 5. *If* $\tilde{X}$ *is det-c and* $\{\varepsilon_i\}$ *are i.i.d. normal, then* $\hat{\beta}$ *is a normal vector.*

2 Estimating σ^2

Recall that, for $k = 1$, $X_i = (1, Z_i)$, we had

$$\sigma^2 \approx \tilde{S}^2 = \frac{1}{n-2} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Z_i)^2$$

We extend this definition to general $k \geq 1$.

Def 1.

$$\tilde{S}^2 := \frac{1}{n-(k+1)} \|\tilde{Y} - \tilde{X}\hat{\beta}\|^2$$

Thm 6. *Assume $(\tilde{X}^\top \tilde{X})^{-1}$ is well defined. Then, $\tilde{S}^2$ is an unbiased est-r of σ^2.*

Thm 7. *If $(\tilde{X}^\top \tilde{X})^{-1}$ is well defined and $\{\varepsilon_i\}$ are i.i.d. normal, then*

$$\frac{n-(k+1)}{\sigma^2} \tilde{S}^2 \sim \chi^2(n-(k+1)),$$

and $\tilde{S}^2$ is indep. of each $\hat{\beta}_i$.

3 Hypotheses tests and confidence intervals for linear functions of $\hat{\beta}$

Recall that the predicted value of Y, given $x = (1, x^1, \ldots, x^k)$, is

$$\mathbb{E}(Y|X = x) = x\beta,$$

which is a linear combination of the entries of β. Hence, the problem of statistical inference about the predicted value of Y (and about β), as before, reduces to making inference about linear combinations of the form

$$\theta := a^\top \beta, \quad a \in \mathbb{R}^{k+1}.$$

Naturally, we estimate θ via

$$\hat{\theta} := a^\top \hat{\beta}.$$

Thm 8. *$\hat{\theta}$ is always unbiased (assuming that $\{\varepsilon_i\}$ are i.i.d. and independent of $\tilde{X}$), and it is consistent whenever $\hat{\beta}$ is consistent.*

Thm 9. $V(\hat{\theta}) = V(a^\top \hat{\beta}) = \sigma^2 (a^\top \mathbb{E}(\tilde{X}^\top \tilde{X})^{-1} a)$.

Corollary 1. *Assuming normal residuals,*

$$U := \frac{\hat{\theta} - \theta}{\sigma \sqrt{a^\top (\tilde{X}^\top \tilde{X})^{-1} a}} \sim N(0, 1),$$

$$V := \frac{\hat{\theta} - \theta}{\tilde{S} \sqrt{a^\top (\tilde{X}^\top \tilde{X})^{-1} a}} \sim T(n-(k+1))$$

Using U and V, we can test hypotheses on θ (or on β_j) and construct confidence intervals for θ.

Ex 3. *Consider a polynomial model*

$$Y = \beta_0 + \beta_1 Z + \beta_2 Z^2 + \varepsilon,$$

and the sample

$$\{z_i\} : -2, -1, 0, 1, 2$$

$$\{y_i\} : 0, 0, 1, 1, 3$$

Assume normal ε*.*

Q 1. *Does the data present sufficient evidence of the presence of nonlinearity in the relation* $x \mapsto y$*, at* 0.05 *level?*

We need to test

$$H_0 = \{\beta_2 = 0\}, \quad H_a = \{\beta_2 \neq 0\}.$$

To this end, we compute

$$\hat{\beta} \approx \begin{pmatrix} 0.571 \\ 0.7 \\ 0.214 \end{pmatrix},$$

$$\tilde{s}^2 := \frac{1}{n-(k+1)} \|\tilde{y} - \tilde{x}\hat{\beta}\|^2 = \frac{1}{n-3} \sum_{i=1}^{n} (y_i - 0.571 - z_i 0.7 - z_i^2 0.214)^2 \approx 0.232, \quad \tilde{s} \approx 0.48.$$

$$v := \frac{\hat{\beta}_2 - 0}{\tilde{s}\sqrt{(0,0,1)(\tilde{x}^\top \tilde{x})^{-1}(0,0,1)^\top}} \approx \frac{0.214}{0.48\sqrt{1/14}} \approx 1.67,$$

$$t_{0.025} \approx 4.303.$$

Thus, we accept H_0*.*

The p-value is

$$p = \mathbb{P}(T(2) > 1.67) \approx 0.23686.$$

Q 2. *Construct a* 0.95*-conf. int-l for* β_2*.*

Using V *as a pivot, we obtain the confidence interval:*

$$\hat{\beta}_2 \pm t_{0.025}\tilde{s}\sqrt{(0,0,1)(\tilde{x}^\top \tilde{x})^{-1}(0,0,1)^\top} \approx (0.214 \pm 0.552).$$