TD 1: Linear regression model

Linear Models and Their Generalizations

K. Meziani

Exercise 1 [Proof of Cochran's theorem]

Let Z be a Gaussian random vector in $\mathbb{R}^n$ with $Z \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbb{I}_n)$, where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\sigma > 0$. Let $F_1, \ldots, F_m$ be subspaces of dimension d_i, orthogonal to each other such that $\mathbb{R}^n = F_1 \oplus \cdots \oplus F_m$. For $i = 1, \ldots, m$, let P_{F_i} denote the orthogonal projection matrix onto F_i. Prove that

1. The random vectors $P_{F_1}Z, \ldots, P_{F_m}Z$ have respective distributions $\mathcal{N}(P_{F_1}\boldsymbol{\mu}, \sigma^2 P_{F_1}), \ldots, \mathcal{N}(P_{F_m}\boldsymbol{\mu}, \sigma^2 P_{F_m})$.

2. The random vectors $P_{F_1}Z, \ldots, P_{F_m}Z$ are pairwise independent.

3. The random variables $\frac{\|P_{F_1}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}, \ldots, \frac{\|P_{F_m}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}$ have respective distributions $\chi^2(d_1), \ldots, \chi^2(d_m)$.

4. The random variables $\frac{\|P_{F_1}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}, \ldots, \frac{\|P_{F_m}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}$ are pairwise independent.

Exercise 2. [Proof of Proposition 1. of the chapter 1]

Let X be the *design* matrix of size $n \times (p+1)$. We assume X to be full rank ($rank(X) = p+1$). Let define the following linear model

$$Y = X\beta + \epsilon$$

with $\beta \in \mathbb{R}^{p+1}$. Let

$$\widehat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|^2$$

be the ordinary least square estimator (OLSE).

1. Show that OLSE exists and is unique such that

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

2. **Application for** $p = 1$: Let $(x_1, y_1), \ldots, (x_n, y_n)$ be n pairs of real numbers. Determine the real $\widehat{a}$ and $\widehat{b}$ that minimize RSS $(a, b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2$. Interpret.

Exercise 3.

Let X be a $n \times p$ matrix of rank p. Let $\widehat{Y}$ be the orthogonal project on the space $[X]$ generated by the column vectors of X of a vector Y of $\mathbb{R}^n$. Show that $\sum_{i=1}^{n}(Y_i - \widehat{Y}_i) = 0$ if one of the column vectors of X is the vector $\mathbf{1}_n = (1, \cdots, 1)$. Interpret.

Exercise 4.

We consider the following simple linear regression statistical model: $Y_i = \beta x_i + \varepsilon_i$, for $i = 1, \ldots, n$ where the ε_i are independent, centered, of constant variance. We define two estimators of $\beta \in \mathbb{R}$:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} \quad \text{et} \quad \beta^\star = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}$$

1. What is the logic of construction of these estimators?

2. Show that they are unbiased estimators of β.

3. Compare the variances of these two estimators.

Exercise 5. [An important result]

We consider the Gaussian linear regression model:

$$Y = X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$$

where $\beta \in \mathbb{R}^r$, $Y \in \mathbb{R}^n$ and X matrix of size $n \times r$ of rank r.

1. Recall the matrix closed form of the OLSE and give an unbiased estimator of $\sigma^2 > 0$.

2. Compute the maximum likelihood estimators of β and σ^2.

3. Conclude.

Exercise 6. [Unbiased estimator of σ^2 in the non-Gaussian model]

Consider the following non-Gaussian linear model:

$$Y = X\beta + \epsilon$$

with $\beta \in \mathbb{R}^p$, X of full rank, and the ϵ_i independent, centered and of variance σ^2. We pose:

$$\widehat{\sigma}^2 = \frac{1}{n-p}||Y - X\widehat{\beta}|^2.$$

We note $\mathrm{Tr}(\cdot)$ the trace of a matrix.

1.Show that $(n-p)\widehat{\sigma}^2 = \mathrm{Tr}(\epsilon^\top P_{X^\perp}\epsilon)$

2. Using the fact that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for A and B of respective size $(m \times n)$ and $(n \times m)$, show that

$$(n-p)\mathbb{E}_\beta[\widehat{\sigma}^2] = \sigma^2 \mathrm{Tr}(P_{X^\perp}).$$

3. Deduce that $\mathbb{E}_\beta[\widehat{\sigma}^2] = \sigma^2$.

Exercise 7. [Proof of theorem 4 chapter 4]

Consider the following Gaussian linear model $Y = X\beta + \epsilon$ where $\beta \in \mathbb{R}^r$, X is a full rank matrix of size $n \times r$ $(n > r)$. Let $C \in \mathcal{M}_{q,r}(\mathbb{R})$. We want to test

$$\mathcal{H}_0 : C\beta = \mathbf{0}_q \quad \text{versus} \quad \mathcal{H}_1 : C\beta \neq \mathbf{0}_q$$

We assume that $rg(C) = q \leq r$.Therefore, you will note that $rg(C^\top) = q$ where $C^\top$ is the transpose of C.

1. Show that if $Z \sim \mathcal{N}_q(\mathbf{0}_q, \Sigma)$ then $Z^\top \Sigma^{-1} Z \sim \chi_q^2$.

2. Show that $C(X^\top X)^{-1} C^\top$ is a symmetric and invertible matrix.

3. Recall the ordinary least squares expression $\widehat{\beta}$.

4. What is the law of $\widehat{\beta}$?

5. Deduce the law of $C\widehat{\beta}$ under the hypothesis $\mathcal{H}_0$.

6. Deduce that, under $\mathcal{H}_0$,

$$R = \frac{(C\widehat{\beta})^\top (C(X^\top X)^{-1}C^\top)^{-1}(C\widehat{\beta})}{\sigma^2} \sim \chi_q^2.$$

7. Conclude that, under $\mathcal{H}_0$,

$$F = \frac{\widehat{\beta}^\top C^\top (C(X^\top X)^{-1}C^\top)^{-1}C\widehat{\beta}}{q\widehat{\sigma}^2}$$

is distributed according to a Fisher distribution with $(q, n - r)$ degrees of freedom. *Each step of the reasoning must be carefully justified.*

8. Justify and construct a test of $\mathcal{H}_0$ against $\mathcal{H}_1$ of level α.

Exercise 8. [MCQ]

We have observations $(x_i, y_i) \in \mathbb{R}^2$, $\forall i = 1, \ldots, n$. We consider the following classical Gaussian linear model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \forall i = 1, \ldots, n$$

where $(\beta_0, \beta_1) \in \mathbb{R}^2$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ are *i.i.d.*

Let $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Assume X is a full rank mlatrix and note $\widehat{\beta}_0$ and $\widehat{\beta}_1$ the least squares estimators of β_0 and β_1.

For each of the following questions, give the answer.

> a) Yes   b) No   c) not always

1. Are the variables Y_i independent and identically distributed?

> a) Yes   b) No   c) not always

2. Does the regression line calculated on the observations pass through the mean point $(\overline{x}, \overline{y})$?

> a) Yes   b) No   c) Only if I am lucky

3. Is it possible to find estimators of β_0 and β_1 with smaller variance than the ordinary least squares estimators ?

> a) Yes   b) No   c) Maybe.

4. Are $\widehat{\beta}_0$ and $\widehat{\beta}_1$ independent?

> a) Yes   b) No   c) It depends on the matrix X

5. If the coefficient of determination R^2 calculated on the observations is equal to 1, are the points $(x_i, y_i)_{i=1,\ldots,n}$ aligned?

> a) Yes   b) No   c) Not necessarily

6. Are $\widehat{Y}$ and $Y - \widehat{Y}$ independent?

> a) Yes   b) No   c) It depends on the matrix X

7. Are $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$ and $Y - \widehat{Y}$ independent?

> a) Yes   b) No   c) It depends on the matrix X

8. Is the maximum likelihood estimator of σ^2 unbiased ?

> a) Yes   b) No   c) We don't know

3

Exercise 9. [This exercise will be solved without the tools of linear algebra]

Let $(x_1, y_1), (x_n, y_n)$ be n pairs of real numbers. We suppose that y_i are the realization of Y_i whose law is given by the following equation:

$$Y_i = a + bx_i + \varepsilon_i, \quad \varepsilon_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$$

1. Determine $\widehat{A}$ and $\widehat{B}$ the maximum likelihood estimators of a and b. Interpret the estimators.

2. Show that these estimators are unbiased.

3. Calculate the variance of the estimators $\mathbb{Var}_\beta(\widehat{A})$ and $\mathbb{Var}_\beta(\widehat{B})$. How do these variances vary as a function of σ^2 and the experimental design $x_1, \ldots, x_n$

4. Compute the covariance of $\widehat{A}$ and $\widehat{B}$. Comment.

5. Let $\widehat{Y}_i = \widehat{A} + \widehat{B}x_i$ and $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$. Show that $\sum_{i=1}^{n} \widehat{\varepsilon}_i = 0$.

6. Show that $\frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{n-2}$ is an unbiased estimator of σ^2.

7. Let x_{n+1} be another value. We define $\widehat{Y}_{n+1} = \widehat{A} + \widehat{B}x_{n+1}$. Compute the variance of this prediction.

8. Furthermore, let $Y_{n+1} = A + Bx_{n+1} + \varepsilon_{n+1}$. Calculate the variance of $\widehat{\varepsilon}_{n+1} = Y_{n+1} - \widehat{Y}_{n+1}$. Compare it to the variance of ε_i (for $i = 1, \ldots, n$).

9. *Gauss-Markov Theorem*:

   a. Show that $\widehat{B}$ is written as a linear combination of the observations (we will explain the weights).

   b. Consider $\widetilde{B} = \sum_{i=1}^{n} \lambda_i Y_i$ another unbiased estimator of B, written as a linear combination of Y_i. Show that $\sum_{i=1}^{n} \lambda_i = 0$ and $\sum_{i=1}^{n} \lambda_i x_i = 0$.

   c. Deduce that $\mathbb{Var}_\beta(\widetilde{B}) \geq \mathbb{Var}_\beta(\widehat{B})$