Probabilities and statistics

Lecturer: Prof. Dr. Sara van de Geer
Prof. Dr. Martin Larsson

Serie 11

May 18th, 2015

Q1. Gauss-Markov Theorem We want to study linear regression models. We do m experiments with explanatory variables $(x_i)_{i=1}^m \subseteq \mathbb{R}^n$ and with a scalar dependent variable $(y_i)_{i=1}^n \subseteq \mathbb{R}$. We suppose that for all i, the underlying model is given by

$$y_i = \beta \cdot x_i + \epsilon_i \quad \beta \in \mathbb{R}^n \tag{1}$$

where (ϵ_i) is a i.i.d sequence such that $\mathbb{E}(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. We want to estimate β. We say that $\hat{\beta}$ is an unbiased estimator of β if

$$\mathbb{E}\left(\hat{\beta}\right) = \beta.$$

Additionally we say that $\hat{\beta}$ is linear if there exists a matrix, D, only depending on X such that $\hat{\beta} = DY$ .We will also say that a matrix $A \lesssim B$ if $B - A$ is a positive semidefinite matrix.

(a) Show that (1) is equivalent to

$$Y = X\beta + \epsilon, \tag{2}$$

where $Y = \begin{pmatrix} y_1 \\ \vdots \\ y^t \end{pmatrix}$, $X = \begin{pmatrix} x_1^t \\ \vdots \\ x_m^t \end{pmatrix}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}$.

(b) Show that the normal linear regression model (example 3.1 of the Skript) is a linear unbiased estimator. We will call its associated matrix K.

(c) Compute the covariance matrix of $\bar{\beta}$, the estimator of the normal linear regression model. **Hint:** Remember that if $Z \in \mathbb{R}^n$ is a random variable and C is a matrix then $V(CZ) = CZC^t$, where $V(\cdot)$ is the covariance matrix.

(d) Show that if $\hat{\beta} = (K + C)Y$ is an unbiased estimator, then $CX = 0$.

(e) Show that the covariance matrix of $\hat{\beta}$ is such that

$$V(\hat{\beta}) \gtrsim V(\bar{\beta}).$$

Q2. In a lake we want to estimate the amount of fishes in a lake (we assume there is only one type of fish on the lake). For this we mark 5 fishes and we let them mix with the others, when they are well mixed we fish 11, and we realize that there are 3 marked and 8 non-marked. What is the maximum-likelihood estimator for the amount of fishes?.

Q3. Let $(X_i)_{i=1}^{2n+1}$ a sequence of i.i.d normal random variables with mean μ and variance σ unknown. We take two different estimators for μ:

$$T_{2n+1}^{(1)} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} X_i,$$

$$T_{2n+1}^{(2)} = X_{(n+1)},$$

where $X_{(1)} < X_{(2)} < ... < X_{(2n+1)}$ are the ordered results.

(a) With the help of the Central Limit Theorem find sequences $c_n^{(1)}$ and $c_n^{(2)}$ so that

$$\mathbb{P}\left(|T_{2n+1}^{(i)} - \mu| \leq c_n^{(i)}\right) \to 0.95.$$

(b) Find $q \in \mathbb{R}^+$ so that

$$\frac{c_{nq}^2}{c_n^1} \to 1,$$

how can we interpret, in words, q?.

Q4. A gas station estimates that it takes at least α minutes for a change of oil. The actual time varies from costumer to costumer. However, one can assume that this time will be well represented by an exponential random variable. The random variable X, therefore, possess the following density function

$$f(t) = e^{\alpha - t}\mathbf{1}_{\{t \geq \alpha\}},$$

i.e. $X = \alpha + Z$ where $Z \sim Exp(1)$. The following values were recorded from 10 clients randomly selected (the time is in minutes):

$$4.2, 3.1, 3.6, 4.5, 5.1, 7.6, 4.4, 3.5, 3.8, 4.3.$$

Estimate the parameter α using the estimator of maximum likelihood.