Dr. Tim Brereton
Lisa Handl

Winter Term 2014/15

Stochastics III

Problem Sheet 5

Deadline: January 7, 2015 at noon, before the practical

Exercise 1 (2 + 2 + 5 + 3 points)

This exercise is designed to investigate the nature of confidence intervals and estimation in general. In order to do this, we consider a setting where we know the true model (i.e., the correct values of all the model parameters) and are able to generate data from it.

A well-known physical law states that the very interesting physical quantity y depends linearly on two variables: x_1 and x_2. More precisely,

$$y = 10 + 2x_1^2 - 7x_2^2.$$

We want to conduct an "experiment" to reproduce this relationship by measuring y for $x_1, x_2 \in \{0, 1, \ldots, 10\}$ (all possible combinations). We know (for some reason) that the measurement error in our experiment is normally distributed with standard deviation 5 and that the measurement errors of each observation are independent of one another.

a) Set up a design matrix X for the experiment and simulate its output on the computer, i.e., draw the vector $\boldsymbol{y}$ corresponding to X according to the known model.

b) Fit a regression model to the simulated data and plot the data together with the estimated regression function and the true regression function in 3D.

   Hint: Use `plot3d` for the data and `persp3d(..., alpha=0.5, add=T)` for the planes.

c) Repeat the simulation from a) 1000 times. Each time, calculate a confidence interval with confidence level $\alpha = 0.05$ for each β_i (separately). State for each β_i (separately) how frequently the true β_i was *not* contained in the corresponding confidence interval. In how many simulations was *(at least) one* of the true β_i not contained in the corresponding confidence interval?

d) Repeat the simulation from a) 1000 times. Check each time whether the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ is inside the confidence ellipsoid with confidence level $\alpha = 0.05$. How frequently was the true $\boldsymbol{\beta}$ not contained in the confidence ellipsoid?

Please hand in snapshots of your 3D plots as well. You can save the current view of your 3D plot as a PNG by typing `rgl.snapshot(<filename>)`.

Exercise 2 (2 + 3 + 3 points)

Reconsider the dataset *ski.txt* on the course website and the linear models from Exercise 3 a) and b) on Problem Sheet 4.

a) For the model from 3.a): Conduct a test for the hypothesis $H_0 : \beta_2 = \beta_3 = 0$ with significance level $\alpha = 0.05$ and interpret the result.

b) For the model from 3.b): Suppose there is an 11^{th} ski resort with lift capacity $x_0 = 2500$. Calculate a confidence interval for the expected target value $\beta_1 + \beta_2 x_0$ and a prediction interval for $Y_0 = \beta_1 + \beta_2 x_0 + \varepsilon_0$ with confidence level $\alpha = 0.05$.

c) For the model from 3.b): Derive formulas for the bounds of the confidence band for the regression function with confidence level $\alpha = 0.05$. Plot the data, the estimated regression function and the confidence band. Add the confidence interval and the prediction interval from a) to your plot.

Exercise 3 (4 points)

Derive a confidence ellipsoid for linear forms of the parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^\top$, i.e., for $H\boldsymbol{\beta}$, where $H \in \mathbb{R}^{r \times m}$ and $\text{rk}(H) = r \leq m$.

Hint: An n-dimensional ellipsoid is a set

$$E = \left\{ x \in \mathbb{R}^n : (x - c)^\top A (x - c) < b \right\} \subset \mathbb{R}^n,$$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, $c \in \mathbb{R}^n$ is the center of the ellipsoid and $b > 0$.

Merry Christmas and a happy New Year! Enjoy your holidays! :)