Computational Statistics
Exam for Spring 2025

Department of Computer and Information Science (IDA), Linköping University

August 29, 2025, 8:00-13:00

Course:	732A89, 732A72 Computational Statistics
Teacher and examiner:	Frank Miller
Allowed aids:	Printed course books, 1 handwritten page (A4, front page, only) with notes
Provided aids:	Helpfiles (lecture slides and some chapters from Givens and Hoeting)
Grades:	A = [36, 40] points, B = [32, 36) points, C = [28, 32) points, D = [24, 28) points, E = [20, 24) points, F = [0, 20) points.
Instructions:	Provide a detailed report that includes plots, conclusions and interpretations. If you are unable to include a plot in your solution file clearly indicate the section of R code that generates it. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in an appendix. In a number of questions, you are asked to do plots. Make sure that they are informative, have correctly labelled axes, informative axes limits and are correctly described. Points may be deducted for poorly done graphs. If you have problems with creating a pdf, you may submit your solutions in text files with unambiguous references to graphics and code that are saved in separate files. There are THREE assignments (with sub–questions) to solve. Provide a separate solution file for each assignment. Include all R code that was used to obtain your answers in your solution files. Make sure it is clear which code section corresponds to which question. If you also need to provide some hand–written derivations, please note the number of the question on each page. Name your solution files as: [your exam account id] [own file description].[format]

Note: If you are not able to solve a part of a question, you can anyway try to show how you would solve the subsequent parts with explaining and providing code examples and you might receive partial points.

1 Optimisation of a one-dimensional function (14 points)

Consider the function

$$g(x) = 2^x/x.$$

It has the derivative $g'(x) = (x\log(2) - 1)2^x/x^2$.

a. Plot the functions $g(x)$ and $g'(x)$ in the interval $[0.5, 3]$ and add a horizontal reference line at 0 to the plot of g'. What do you think the value of x is where the minimum occurs?

b. Write an algorithm based on the secant method and apply it to g' in order to find a local minimum of g in $[0.5, 3]$. Use a convergence criterion based on $|g'(x)|$ to assess the algorithm's convergence.

c. Alternatives to computing the minimum with the secant method are to use the function `optimize` in `R` to find a local optimum of a function within a given interval or to compute the minimum algebraically. Use these two ways to compute the local minimum of $g(x)$ on $[0.5, 3]$ and compare the results with the result obtained with your own secant-code.

2 Bootstrap and permutation test for regression (14 points)

The object `data` in the file `windprod.Rdata` contains the daily average wind speed (in m/s) and the daily production of electricity (in kWh) of one specific wind power plant located in Rättvik for April and May 2025 (61-data points; data source solivind.com). The first value in each row is the wind speed and the second value after a comma is the electricity production.

You are supposed to fit a quadratic regression for electricity produced y_i as dependent variable and wind speed x_i as independent variable, $y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \epsilon_i$, assuming independent errors $\epsilon_i, i = 1, \dots, 61$.

a. Read in the dataset, fit this regression model and estimate the quadratic parameter β_2 together with a 95%-confidence interval using the `R`-function `lm`. Create a plot for produced electricity vs. wind speed and add the estimated regression curve to the plot.

b. Derive a 95%-bootstrap confidence interval for β_2 based on the percentile method. Do not use a bootstrap package for this calculation; program the bootstrap on your own. Use 5000 bootstrap replicates. Plot a histogram with the bootstrap distribution. Check if the confidence interval here and in part a. are similar or not and comment on it.

c. Test now the null hypothesis H_0: "$\beta_2 = 0$ (the relation is linear (a straight line)" versus the alternative H_1: "$\beta_2 \neq 0$" with a permutation test. Generate a null distribution and derive the p-value of the permutation test.

3 Generation of random variables (12 points)

A random variable X has the an exponential distribution with rate $\lambda = 2$. It has the probability density f:

$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

The random variable Z has a Laplace distribution with parameter $\lambda = 2$. For positive values z, the density $g(z)$ of this Laplace distribution has the same shape as the exponential distribution; at a negative value $-z$, the density is the same as for z (symmetric around 0);

$$g(z) = \frac{\lambda}{2} \exp(-\lambda|z|).$$

a. Generate 10000 samples of the random variable X (exponential distribution) using the inverse cummulative distribution function method. While there is a build-in function in R for this distribution, you are supposed to generate random variables based on your own code, **only using the build-in function for the uniform distribution**.

b. Generate 10000 samples of the random variable Z (Laplace distribution) with a method of your choice. While there might be a function in an R-package for this distribution, you are supposed to generate random variables based on your own code, **only using the build-in functions for other distributions in base R**.

c. Use now the sample generated in b. Plot a histogram of the generated random variables following g. Compute a Monte Carlo estimate for $E(Z), Var(Z)$ and for $P(Z < 0.5)$ based on your sample.