Computational Statistics
Exam, Spring 2025

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

May 16, 2025, 8:00-13:00

Course:	732A89, 732A90, 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 Optimization (13 points)

Consider the function

$$g(x, y) = \sin(x) - \frac{1}{5}(x + y)^2 - \frac{1}{25}x^2$$

which has the gradient

$$g'(x, y) = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} \cos(x) - \frac{2}{5}(x + y) - \frac{2}{25}x \\ -\frac{2}{5}(x + y) \end{pmatrix}.$$

It should be maximized for $-5 \le x \le 5$ and $-5 \le y \le 5$.

a. Write an algorithm based on the steepest ascent method using backtracking.

b. Apply the algorithm on the function g and use the two starting points $(-5, 2), (-4, -4)$ for two runs of the algorithm. Plot the iterations done into a two-dimensional plot for (x, y). Based on your runs, what is the global maximum of g?

2 EM algorithm (14 points)

In the lectures, an EM algorithm was presented for a univariate normal mixture model with two components; you can find also an R-text-file `emalg.r` with the code.

a. Use the algorithm from the lecture as start and modify it for the case of three components, i.e. the mixture of three normal distributions. Instead of automatically generating starting values, allow for user specified starting values (ideally as parameters of the `emalg`-function). Change the stopping rule to a criterion which is based on the expectation of the joint log likelihood (function Q in lecture and course book).

b. Use the data `dat` in `mixturedat.Rdata` (same data is in `mixturedat.csv`) and plot a histogram. Based on this histogram, choose reasonable starting values. Explain why you choose them. Determine estimates for the parameters using the function `emalg` with your modifications and your starting values.

c. Check convergence of your solution using convergence plots for all parameters and for Q.

3 Metropolis algorithm (13 points)

The density f of a random variable X is proportional to the product of a density from the normal distribution with mean μ and variance σ^2, $N(\mu, \sigma^2)$, and the function

$$g(x) = \begin{cases} 1, & \text{if } -1 < x < 1, \\ \exp(-|x| + 1), & \text{if } |x| \ge 1. \end{cases}$$

This means that

$$f(x) = c * g(x) * \exp[-(x - \mu)^2/(2\sigma^2)]$$

with a constant c. The parameters μ, σ^2 are given as $\mu = 1.89, \sigma^2 = 1$. It is of interest here to compute the probability $P(x \ge 1.89) = \int_{1.89}^{\infty} f(x)\, dx$.

a. Plot f. You can set the value c to an arbitrary fixed value for this plot.

b. Write an own program using a Metropolis algorithm to generate 10000 draws of X. Compare at least 2 different proposal distributions and decide which of them is better.

c. Plot a histogram of the generated random variables following f. Compute a Monte Carlo estimate for $P(x \ge 1.89)$ and for $E(X)$ based on your generated sample.