Computational Statistics
Advanced Course, 7.5 HEC, Spring 2022
Home Assignment 2 until February 3

Mahmood Ul Hassan and Frank Miller, Department of Statistics

January 27, 2022

You are supposed to solve these problems individually without any cooperation. Record your answers clearly. If you cite material from other sources, or use intellectual ideas or code from others, point out this clearly with stating the source. Please include your `R` computer code and relevant output in your solution. Send an email to `mahmood.ul-hassan@stat.su.se` with your solutions until the above deadline with a file name like `CompStat2022HA2_[Your name].pdf` and extra file(s) with similar name with your computer code.

1 Three-dimensional minimization

For $x, y, z \in [-5, 5]$, consider the three dimensional function

$$f(x,y,z) = \frac{1}{2}(x^4 + y^4 + z^4) - 8(x^2 + y^2 + z^2) + \frac{5}{2}(x + y + z).$$

The aim is to search for the position (x^*, y^*, z^*) of the global minimum of f inside of the given range for x, y, z (you do not need to consider the boundary when one coordinate is ± 5, here).

a. Visualize the function f by drawing contour plots for at least three specific z-values.

b. Search (x^*, y^*, z^*) using the steepest ascent algorithm (here it is rather a "steepest descent" algorithm). Use several different starting value(s) in order to find a global minimum.

c. Visualize the function f by drawing a contour plot specifically for $z = z^*$ which you computed as z-coordinate of the optimal solution.

2 EM algorithm for mixture of three normal components

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

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. Important: Use this provided algorithm to start with and generalize it; do not write a completely new code.

b. Modify the stopping criterion such that the stopping does not depend on a scaling of data. This means that if data in another unit is analysed, e.g. numbers are multiplied with or divided by 1000, the results after stopping should be the same.

c. Use the dataset `threepops.csv` in Athena which contains $n = 230$ observations. Create first a histogram of the data. Fit then a normal mixture model with three components using your program in a. Which estimates do you get for the model parameters?

d. Provide plots of current estimates for the model parameters versus iteration-number.

3 EM-algorithm for bivariate normal mixture

In this problem, you are supposed to generalize the EM-algorithm `emalg.r` to bivariate normal mixtures.

a. Use the algorithm from the lecture as start and modify it for the case of **two-dimensional observations** which come from the mixture of two bivariate normal distributions. Allow for user specified starting values. Add the computation of the expected log-likelihood Q to the algorithm.

b. Use the dataset `bivardat.csv` in Athena which contains $n = 1000$ observations. Create a two-dimensional point plot of the data. Choose starting values of all model-parameters and discuss why you have chosen these values. You might test different possibilities before you decide upon the starting values.

c. Fit the bivariate normal mixture model to the data using your program from b. Check the convergence of all model-parameters and of the expected log-likelihood.

4 Gauss-Hermite integration

Nodes 5-9 and the weights for the 9-point Gauss-Hermite integration are given in the left of the tables below. The nodes are symmetric around 0 which means that nodes 1-4 are the negative values of nodes 6-9 with the same weight (e.g. $x_4 = -x_6 = -0.723551$, $A_4 = A_6 = 0.432652$).

a. Compute the integrals $\int_{-\infty}^{\infty} f(x)\, dx$ of the following density functions $f(x)$ using this 9-point Gauss-Hermite integration. Theoretically, we know that the exact integral is 1 and the result shows you how good the 9-point Gauss-Hermite performs in this situation.

   (i) $f(x) =$ the density of the t-distribution with 3 degrees of freedom, and with 10 degrees of freedom,

   (ii) $f(x) = \frac{1}{2}e^{-|x|}$ (double-exponential distribution),

   (iii) $f(x) =$ the density of the normal distribution $N(0, 1)$.

   Comment on these results.

b. Take one of the cases in a. where you think that the result from the calculation was least satisfactory. Compute the integral with the 18-point Gauss-Hermite integration. The nodes and weights can be based on the right of the following tables. Comment on the result.

9-point Gauss-Hermite		
i	Nodes x_i	Weights A_i
5	0	0.720235
6	0.723551	0.432652
7	1.46855	0.088475
8	2.26658	0.00494362
9	3.19099	0.000039607

18-point Gauss-Hermite		
i	Nodes x_i	Weights A_i
10	0.258268	0.483496
11	0.77668	0.284807
12	1.30092	0.0973017
13	1.83553	0.01864
14	2.3863	0.00188852
15	2.96138	9.1811e-5
16	3.57377	1.8107e-6
17	4.24812	1.0467e-8
18	5.04836	7.8282e-12