

MATH 680 - ASSIGNMENT 1

To be handed in not later than 10pm, 20th February 2015.

Please submit your solutions as pdf via myCourses.

This project focuses on the paper:

Acceleration of the EM algorithm by using quasi-Newton methods. M. Jamshidian and R. I. Jennrich, *Journal of the Royal Statistical Society, Series B*, **59**:569–587, 1997.

In this paper the authors study amendments to the EM algorithm aimed at improving speed of convergence. For example, for a two component mixture of Poisson distributions model (section 5.1, pp. 576–8), the authors study three example data sets:

Ex 1. A simulated example using the following data

Value j	0	1	2	3	4	5	6
Count c_j	552	703	454	180	84	23	4

Ex 2. A data set from Hasselblad (1969)

Estimation of finite mixtures of distributions from the exponential family. V. Hasselblad, *Journal of the American Statistical Association*, **64**, 1459–1471, 1969.

Value j	0	1	2	3	4	5	6	7	8	9
Count c_j	162	267	271	185	111	61	27	8	3	1

Ex 3. A data set from Thisted (1988, p. 188)

Elements of Statistical Computing: Numerical Computation. R. A. Thisted, 1988.

Value j	0	1	2	3	4	5	6
Count c_j	3062	587	284	103	33	4	2

- (a) Write out in full the details of the EM update steps for the two component Poisson mixture case.
5 Marks
- (b) For the three data sets above, implement the ordinary EM algorithm to verify (or refute) the claims made by the authors concerning its slow convergence. Use the same starting values as in the paper.
10 Marks
- (c) For at least one of the proposed amended EM methods described in the paper, verify (or refute) the authors claim of an improvement in convergence properties.
10 Marks
- (d) Perform a direct optimization of the missing data likelihood based on the Poisson mixture mass function for the observed data alone, and compare convergence properties with those of the EM approaches. You may use `optim` to do this.
5 Marks

Please hand in no more than 5 pages of text and tables, plus your annotated computation code in a separate file. The code should produce reproducible results.