INTERNATIONAL CENTRE
OF EXCELLENCE FOR
EDUCATION IN
MATHEMATICS

Parametric and semi-parametric mixture models
Simon Bartlett, Department of Statistics, Macquarie University

My project over the past six weeks has been on a method for describing Parametric and semi-Parametric Mixture models. These types of models are made up of finite numbers of parametric distributions, i.e. Normal distributions, and non-parametric distributions. I undertook my project in the Statistics department of Macquarie University under the direct supervision of Dr. Jun Ma.

The first part of my project involved researching the EM (Expectation-Maximisation) Algorithm, which was the suggested method for describing these models. The algorithm proceeds iteratively in two steps, E and M. The E-step requires some initial values for the parameters of the models, then it classifies the observed data into the models and calculates the expectation of the log-likelihood. The M-step simply chooses new parameters that maximise the log-likelihood function from the E-step. The steps are then alternated until either sufficient convergence in the log-likelihood or after an appropriate number have been completed. My research involved learning how this algorithm worked and ways to implement it using MATLAB, a numerical computing and programming package similar to C++.

The second part was then to write an implementation of the EM algorithm for the simple case of a mixture of two normal univariate distributions. The program required a mixed sample of data from two distributions and initial estimates of, the ratio of the distributions, their □ and their □. The E-step for the program involved calculating the value of the Probability Density Function (PDF) for each sample value and each set of parameters, and then it classified the data into groups and calculated the log-likelihood for the initial parameters. The M-step for the program involved using the classifications of the data to update the ratio, and then using both of these to update □ and □. This program was then extended to mixtures with a finite number of univariate normal distributions in a natural way.

The third part was then to incorporate into this basic program the multivariate nature more commonly found in normal distributions. This required an overhaul of the coding of the program as the parameters for these distributions changed, □ became a vector and □^{2} became a matrix. However the actual layout as described above did not change significantly. It is worth noting as well that for each program written I also needed to write a sample generating program to test it.

The final part was the addition of other parametric distributions. This required two things. Firstly the distributions themselves needed to be coded into files accessible by the program. Secondly up till this point a method specific to normal distributions had been used for the M-step, on adding different distributions a general maximising method was required. While coding the distributions was easy, finding an easy to code maximising method was where my project ended. One of the ideas which came from this project was that of random (intelligent) parameter estimation, whereby the computer runs the algorithm many times with different parameters removing the need to have to specify the number of models and even their distributions in the mixture.

The experience for me was one that I couldn’t have had if not for this scholarship program. I not only got to work on an important part of my supervisor’s research but was able to go to event organised around meeting like minded students and being able to not only hear about their work but also present my own for the benefit of everyone. I would like to take this opportunity to thank the MQ Stats department and especially my supervisor Jun Ma, the CSIRO for hosting the BDI, and lastly AMSI and the people from ICE-EM for providing me with the opportunity to undertake this research project.

Simon received an ICE-EM Vacation Scholarship in December 2005.
See www.ice-em.org.au/students.html#scholarships05