Asymptotic methods for tests of homogeneity for finite mixture models

Michael Stewart
School of Mathematics and Statistics
University of Sydney

PhD Thesis
May 2002

Acknowledgements

I would like to thank some people:

- Professor John Robinson, without whose rich intellect, endless encouragement and supremely agreeable manner this project would never have got off the ground,

- Neville Weber, Malcolm Quine and the rest of the staff in the Statistics group at the University of Sydney, who permitted me to work in an ideal environment to keep the project going and of course

- my wonderful partner Melody Caramins whose inexhaustible love and support got me over the line.

I owe you all everything,

M. S.

Abstract

We present limit theory for tests of homogeneity for finite mixture models. More specifically, we derive the asymptotic distribution of certain random quantities used for testing that a mixture of two distributions is in fact just a single distribution. Our methods apply to cases where the mixture component distributions come from one of a wide class of one-parameter exponential families, both continous and discrete. We consider two random quantities, one related to testing simple hypotheses, the other composite hypotheses. For simple hypotheses we consider the maximum of the standardised score process, which is itself a test statistic. For composite hypotheses we consider the maximum of the efficient score process, which is itself not a statistic (it depends on the unknown true distribution) but is asymptotically equivalent to certain common test statistics in a certain sense. We show that we can approximate both quantities with the maximum of a certain Gaussian process depending on the sample size and the true distribution of the observations, which when suitably normalised has a limiting distribution of the Gumbel extreme value type. Although the limit theory is not practically useful for computing approximate p-values, we use Monte-Carlo simulations to show that another method suggested by the theory, involving using a Studentised version of the maximum-score statistic and simulating a Gaussian process to compute approximate p-values, is remarkably accurate and uses a fraction of the computing resources that a straight Monte-Carlo approximation would.

Contents