ANALYSIS OF MICROARRAY DATA: A MIXED-MODEL FINITE-MIXTURE APPROACH

P.C. Thomson

University of Sydney, and Cooperative Research Centre for Innovative Dairy Products, Camden, Australia

Email: *PeterT@camden.usyd.edu.au*

Frequently, the goal of the analysis of microarray data is to determine the genes that are differentially expressed (DE) as opposed to those that are not differentially expressed (non-DE). Genes are classified as DE if their expression levels differ "significantly" between two (or more) different physiological states. However, this poses a multiple testing problem which may in part be overcome by false discovery rate (FDR) evaluation.

However, an alternative approach is to develop a model for the evaluation of all the gene features simultaneously, and proceed as a model fitting rather than a hypothesis testing process. For this method, a two-stage analysis is performed. (1) All the normalized expression level data are analyzed simultaneously in a large-scale linear mixed model. This model includes random effect terms to describe the both the physical design of the microarrays, as well as the gene effects, or gene-contrast effects. (2) The BLUPs of the gene effects are then fitted to a mixture model of the form $\pi_1 N(0,\sigma_g^2+\sigma_\varepsilon^2)+(1-\pi_1)N(0,\sigma_\varepsilon^2)$, where π_1 is the (prior) probability of a gene being DE, σ_g^2 is the variance of the underlying expression levels of the DE genes, and σ_ε^2 is the residual variance, common to DE and non-DE genes. The mixture model is fitted using the EM algorithm, and in the process, returns posterior probabilities of genes being DE, as well as their estimated effects.

This method will be illustrated by the analysis of a large-scale microarray study of lactation in the tammar wallaby.