

Bayesian Clustering and Curve Fitting for Time-course Microarray Data

Jun S. LIU, *Department of Statistics, Harvard University, USA*, E-mail: jliu@stat.harvard.edu

Abstract: Gene expression over time is, biologically, a continuous process and can thus be represented by a continuous function, i.e. a curve. Individual genes often share similar expression patterns (functional forms). However, the shape of each function, the number of such functions, and the genes that share similar functional forms are typically unknown. Here we introduce a Bayesian approach that allows a direct discovery of related patterns of gene expression and their underlying functions from data without a priori specification of either cluster number or functional form. This approach accounts for between time-point correlations and accommodates missing data. We apply this method to gene expression data over the life-cycle of *Drosophila melanogaster* and *Caenorhabditis elegans* to discover 17 and 16 unique patterns of gene expression in each species, respectively. New and previously described expression patterns in both species are discovered, the majority of which are biologically meaningful and exhibit statistically significant gene function enrichment.