

Introducing a Model to Cluster Patient Survival Times Using the Coxian Phase-Type Distribution

Laura Hill¹ and Adele Marshall¹

¹ Centre of Statistical Science and Operational Research, Queens University Belfast, Northern Ireland

The purpose of this research was to introduce a modelling approach which will represent patient survival using a Coxian phase-type distribution, conditioned on a network of gene expression data. A significant area of research within bioinformatics is gene expression data analysis. DNA microarrays enable us to monitor gene expression values for tens of thousands of genes simultaneously. One of the main areas of statistics has focused on clustering such data. Although there has been development in this area, there is still a substantial amount of research required to further establish sophisticated methods to identify key patterns and characteristics of interest and more specifically link these to patient survival and gene expression data. The Coxian phase-type distribution is a type of Markov model that represents a process as consisting of phases or states which change over time. This research presents a new clustering technique based on the Coxian phase-type distribution, which clusters a group of patients according to their underlying continuous survival times. This clustering technique can be coupled with Bayesian networks, which is a statistical graphical technique used to model inter-relationships between variables. Whilst Bayesian networks have already been considered to represent a network of genes according to their expression values - the combination of Bayesian networks with Coxian phase type distributions extends any previous research carried out in this field. Such a model will not only enable the modelling of relationships between genes of interest, but by including a continuous survival node in the network, can also determine which genes are associated with survival.