World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:14, No:12, 2020

An Infinite Mixture Model for Modelling Stutter Ratio in Forensic Data Analysis

Authors : M. A. C. S. Sampath Fernando, James M. Curran, Renate Meyer

Abstract : Forensic DNA analysis has received much attention over the last three decades, due to its incredible usefulness in human identification. The statistical interpretation of DNA evidence is recognised as one of the most mature fields in forensic science. Peak heights in an Electropherogram (EPG) are approximately proportional to the amount of template DNA in the original sample being tested. A stutter is a minor peak in an EPG, which is not masking as an allele of a potential contributor, and considered as an artefact that is presumed to be arisen due to miscopying or slippage during the PCR. Stutter peaks are mostly analysed in terms of stutter ratio that is calculated relative to the corresponding parent allele height. Analysis of mixture profiles has always been problematic in evidence interpretation, especially with the presence of PCR artefacts like stutters. Unlike binary and semi-continuous models; continuous models assign a probability (as a continuous weight) for each possible genotype combination, and significantly enhances the use of continuous peak height information resulting in more efficient reliable interpretations. Therefore, the presence of a sound methodology to distinguish between stutters and real alleles is essential for the accuracy of the interpretation. Sensibly, any such method has to be able to focus on modelling stutter peaks. Bayesian nonparametric methods provide increased flexibility in applied statistical modelling. Mixture models are frequently employed as fundamental data analysis tools in clustering and classification of data and assume unidentified heterogeneous sources for data. In model-based clustering, each unknown source is reflected by a cluster, and the clusters are modelled using parametric models. Specifying the number of components in finite mixture models, however, is practically difficult even though the calculations are relatively simple. Infinite mixture models, in contrast, do not require the user to specify the number of components. Instead, a Dirichlet process, which is an infinite-dimensional generalization of the Dirichlet distribution, is used to deal with the problem of a number of components. Chinese restaurant process (CRP), Stick-breaking process and Pólya urn scheme are frequently used as Dirichlet priors in Bayesian mixture models. In this study, we illustrate an infinite mixture of simple linear regression models for modelling stutter ratio and introduce some modifications to overcome weaknesses associated with CRP.

Keywords : Chinese restaurant process, Dirichlet prior, infinite mixture model, PCR stutter
Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development
Conference Location : Chicago, United States
Conference Dates : December 12-13, 2020