FINDING TRANSCRIPTION FACTOR BINDING SITE : MARKOV CHAIN MONTE CARLO CONVERGENCE

<u>D. Nur</u>^{\dagger}

University of Newcastle, Callaghan, Australia

[†]E-mail: *Darfiana.Nur@newcastle.edu.au*

Transcription regulation is controlled by coordinated binding of one or more transcription factors (TF) in the promoter regions of genes. The detection and location of transcription factor binding sites (TFBS) in DNA is very crucial in understanding the mechanics of gene expression. TFBS in many species tend to occur as clusters, also known as Cis-Regulatory Modules (CRM), for which the number of sites and distance between the sites vary in a module. The computational identification of TFBS using the analysis of DNA sequence data, such as Markov Chain Monte Carlo (MCMC) methods, has emerged in the last decade as a major new technology for the elucidation of transcription regulatory networks. MCMC methods have revolutionised Bayesian statistics, enabling evaluation of complex distributions and thus facilitating careful modelling in a very wide range of disciplines. An important consideration in the implementation of these methods, however, is whether the chain converges in some formal sense to the target distribution and, if so, how quickly. The aim of this talk is to address the convergence diagnostics of TFBS identification and CRM through Gibbs sampler by applying diagnostic tools to the output produced by running their samplers.