

Hidden Markov Models For DNA Sequence Segmentation : Simulation and Evaluation

Darfiana Nur¹, Kerrie L. Mengersen² and Yan-Xia Lin³

¹*School of Mathematical and Physical Sciences, University of Newcastle, Australia*

²*School of Mathematical Sciences, Queensland University of Technology, Australia*

³*School of Mathematics and Applied Statistics, University of Wollongong , Australia*

This paper focuses on the simulation, estimation and sensitivity of a hidden Markov model describing homogeneous segments of DNA sequences. We consider three main aims : (i) to present a simulation experiment of hidden Markov models (HMM) for DNA sequence segmentation; (ii) to evaluate the accuracy and precision of estimates from a Bayesian analysis of this model under various parameter configurations and (iii) to assess the sensitivity of the estimates to the specification of the prior.

Overall, we find that (i) simulation is simple, fast and can be undertaken in R; (ii) in general the Bayesian analysis provided satisfactory estimates in terms of both precision and accuracy and (iii) the estimates are sensitive to the prior on the transition probabilities between segments, but less sensitive to the prior on the transition probabilities between nucleotides within segments.

We suggest that such a simulation study should be undertaken for all complex statistical models in order to better understand the properties of the model and the resultant estimates.