Stockholm
University

# Estimating convergence of Markov chain Monte Carlo simulations

Kristoffer Sahlin[*]

December 2011

## Abstract

An important research topic within Markov chain Monte Carlo (MCMC) methods is the estimation of convergence of a simulation. The simulation is divided in to two parts, pre- and post-convergence, where the pre-convergence part known as burn-in is discarded and the post-convergence part is used for inference. Recently, MCMC methods have become a popular way of analyzing phylogenetic models. As more and larger phylogenetic data sets are analyzed, there is a need for automated procedures estimating both convergence and sufficient run length of a simulation. Since MCMC methods are used in a variety of different research fields there are several different methods for evaluating the output of a run. In this thesis, we construct a diagnostic for estimating the burn-in of the chain. We then evaluate this diagnostic together with well known convergence diagnostics used in other fields on simulations performed on three different phylogenetic data sets. We also propose an algorithmic procedure for verifying convergence and sufficient run length of an MCMC-simulation.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:ksahlin@kth.se . Supervisor: Sebastian Höhna.