Abstract

Markov chain Monte Carlo (MCMC) simulation is the central tool for Bayesian posterior sampling. However, the method suffers from the high computational cost of likelihood evaluation when dealing with large datasets, which has become increasingly common in today's Big Data context. This thesis investigates two recently proposed algorithms to speed up MCMC sampling, namely the Random Walk Metropolis (RWM) subsampling design and the embarrassingly parallel MCMC algorithm. For RWM subsampling algorithm, the analytic work focuses on investigating the effect of the bias-knob ε on the approximation performance. The experiments show that a higher ε corresponds to a higher bias together with an over-estimated variance in the approximation distribution, given a fixed number of samples. For the embarrassingly parallel MCMC method, the experiments demonstrate the good performance of the Gaussian approximation technique for the kernel density. In addition, based on the latter algorithm, this thesis attempts another parallel MCMC setting, together with two additional approximation techniques, i.e. quantile approximation and marginal Gaussian approximation. In particular, experiments are conducted to demonstrate the sensitivity of the quantile approximation technique to the number of subsets in the parallel MCMC setting.