**Comparing algorithms based on Likelihood free Markov Chain Monte Carlo methods**

**Amal Agarwal, STAT 540 Project, Fall 2015**

**Introduction:** In the wide class of algorithms dealing with approximate Bayesian computation, Likelihood Free (LF) MCMC methods form a significant component. In the most simplest form, this involves sampling a parameter from the prior, generating dataset from the model given , and accepting as a plausible sample from the posterior depending on the closeness of with actual dataset.[1]

**Problem description:** In this project I will apply the LF-MCMC methods[1] on a simulated dataset from Gaussian mixture model. Subsequently I propose to incorporate the LF-MCMC techniques to a high dimensional matrix normal distribution using Gaussian copula.[2]

**Part I:** Application of LF-MCMC on a simulated dataset: First I will simulate data from a Gaussian mixture model with two modes. Augmented joint posterior LF-MCMC sampler and marginal space LF-MCMC samplers will be used to obtain samples from the posterior. These algorithms will be compared based on the following criteria:

1. Acceptance rates in the MCMC update step: This is governed by the choice of scale parameter in the standard kernel density function used to approximate .[1] Various kernel densities like Uniform, Epanechnikov and Gaussian will be compared based on different
2. Choice of distance: Distance measures[1] like Euclidean, Mahalanobis and Hellinger based on different summary statistics play a crucial role and will be compared based on the length of credible intervals and coverage probabilities of the true parameters.

**Part II:** Application of LF-MCMC to a high dimensional matrix normal distribution using Gaussian copula[2]: Dependence structure of a high dimensional posterior like a matrix normal distribution can be approximated by constructing a Gaussian copula which reduces the application of LF-MCMC methods to low dimensional (bivariate and univariate) posterior distributions.

**Part III:** In addition, for part I, marginal space sampler will be analyzed for different values of S, the number of simulated data points which governs the degree of approximation of the weighting kernel estimate.[1] An error distribution augmented sampler will be implemented and compared against samplers constructed from multiple chain paths based on criteria (a) and (b) in part I. Tempering based on the number of observed data points will be compared against tempering based on [1]

*Model Diagnostics:* If time permits, original model and a test model will be diagnosed for misspecification on the simulated dataset in part I using a variant of error distribution augmented sampler. [1], [3]

**References:**

[1] S. Brooks, A. Gelman, G.L. Jones, X.L. Meng, 2011. Chapter 12, Handbook of Markov Chain Monte Carlo, Chapman and Hall CRC.

[2] J. Li, D.J. Nott, Y. Fan and S.A. Sisson, 2015. Extending approximate Bayesian computation methods to high dimensions via Gaussian copula.

[3] Ratmann, Andrieu, Hinkley, Wiuf and Richardson, 2009. Model criticism based on likelihood free inference, with an application to protein network evolution. Proceedings of the National Academy of Sciences of the USA, 10:10576-10581.