Likelihood free MCMC

STAT 540- Project; ABC methods

Amal Agarwal

PENNSTATE

Department of Statistics
Pennsylvania State University

Idea

- Difference in acceptance probability between regular MCMC and Likelihood free MCMC.
- What if likelihood function is analytically unavailable/ computationally expensive to evaluate?
- Likelihood free Rejection sampling algorithm[1]:
  - Sample $\theta' \sim \pi(\theta)$ from the prior.
  - Generate dataset $\underset{\sim}{x}$ from the model $\pi(\underset{\sim}{x}|\theta')$.
  - Accept θ' if $\underset{\sim}{x} \approx \underset{\sim}{y}$.
- $d(\underset{\sim}{x}, \underset{\sim}{y}) \approx d(T(\underset{\sim}{x}), T(\underset{\sim}{y}))$, for some sufficient statistics $T(.)$ of θ.

[1] S.A. Sisson and Y.Fan, Likelihood free MCMC, Chapter 12, Handbook of Markov Chain Monte Carlo, Chapman and Hall CRC

Why it works?

- Consider the following target:

$$\pi_{\text{LF}}(\theta, \underset{\sim}{x} | \underset{\sim}{y}) \propto \pi(\underset{\sim}{y} | \underset{\sim}{x}, \theta)\pi(\underset{\sim}{x} | \theta)\pi(\theta)$$

- Proposal can can factorized as

$$q[(\theta, \underset{\sim}{x}), (\theta', \underset{\sim}{x}')] = q(\theta, \theta')\pi(\underset{\sim}{x}' | \theta')$$

- The acceptance probability in Metropolis hastings algorithm simplifies as:

$$\alpha[(\theta, \underset{\sim}{x}), (\theta', \underset{\sim}{x}')] = \frac{\pi_{\text{LF}}(\theta', \underset{\sim}{x}' | \underset{\sim}{y}) \; q[(\theta', \underset{\sim}{x}'), (\theta, \underset{\sim}{x})]}{\pi_{\text{LF}}(\theta, \underset{\sim}{x} | \underset{\sim}{y}) \; q[(\theta, \underset{\sim}{x}), (\theta', \underset{\sim}{x}')]}$$

$$= \frac{\pi(\underset{\sim}{y} | \underset{\sim}{x}', \theta')\cancel{\pi(\underset{\sim}{x}' | \theta')}\pi(\theta') \; q(\theta', \theta)\cancel{\pi(\underset{\sim}{x} | \theta)}}{\pi(\underset{\sim}{y} | \underset{\sim}{x}, \theta)\cancel{\pi(\underset{\sim}{x} | \theta)}\pi(\theta) \; q(\theta, \theta')\cancel{\pi(\underset{\sim}{x}' | \theta')}}$$

Variants of LF-MCMC Algorithms

- Kernel density estimation for $\pi_\epsilon(\underset{\sim}{y}|\underset{\sim}{x},\theta)$ via sufficient statistic $T(.)$ for θ.

$$\pi_\epsilon(\underset{\sim}{y}|\underset{\sim}{x},\theta) \propto \frac{1}{\epsilon} K\left(\frac{d(T(\underset{\sim}{y}), T(\underset{\sim}{x}))}{\epsilon}\right)$$

- **Error augmented** LF MCMC:

$$\pi_{\text{LF}}(\theta,\underset{\sim}{x},\epsilon|\underset{\sim}{y}) \propto \pi_\epsilon(\underset{\sim}{y}|\underset{\sim}{x},\theta)\pi(\underset{\sim}{x}|\theta)\pi(\theta)\pi(\epsilon)$$

- **Multiple augmented** LF MCMC:

$$\pi_{\text{LF}}(\theta,\underset{\sim}{x}_{1:S},\epsilon|\underset{\sim}{y}) \propto \frac{1}{S}\sum_{s=1}^{S}\pi_\epsilon(\underset{\sim}{y}|\underset{\sim}{x}^s,\theta)\prod_{s=1}^{S}\pi(\underset{\sim}{x}^s|\theta)\pi(\theta)$$

- **Marginal space** LF MCMC:

$$\pi_{\text{LF}}(\theta|\underset{\sim}{y}) \approx \frac{\pi(\theta)}{S}\sum_{s=1}^{S}\pi_\epsilon(\underset{\sim}{y}|\underset{\sim}{x}^s,\theta)$$

LF MCMC for high dimensions

- Curse of dimensionality[2]
  - $\dim(\underset{\sim}{T}) > \dim(\underset{\sim}{\theta})$ for parameter identifiablity.
  - KDE reliable only in low dimensions.
- $C(\underset{\sim}{\theta})$ is the distribution of $(G_1(\theta_1), G_2(\theta_2), ..., G_p(\theta_p))$.
- **Meta Gaussian family** of distributions:

$$g(\underset{\sim}{\theta}) = \frac{1}{|\Lambda|^{1/2}} \exp\left[\frac{1}{2}\eta^T(I - \Lambda^{-1})\eta\right] \prod_{i=1}^{p} g_i(\theta_i)$$

where $\eta_i = \Phi^{-1}[G_i(\theta_i)]$

- **Matrix Normal** distribution: For a $n \times p$ random matrix X, the pdf $\mathcal{MN}_{n,p}(M, U, V)$ is given as:

$$p(X|M, U, V) \propto \frac{\exp\left(-\frac{1}{2}tr(V^{-1}(X - M)^T U^{-1}(X - M))\right)}{|V|^{n/2}|U|^{p/2}}$$

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

Results

Table: LF MCMC for different matrix normal sizes via Gaussian copula

Basic LF MCMC	ESS	ESS/s	Accept rate	Run time
2×2	161	20.05	0.124	48
3×2	159	21.39	0.092	115
3×3	147	17.20	0.061	309
4×3	143	16.05	0.047	592
4×4	137	14.15	0.037	1164

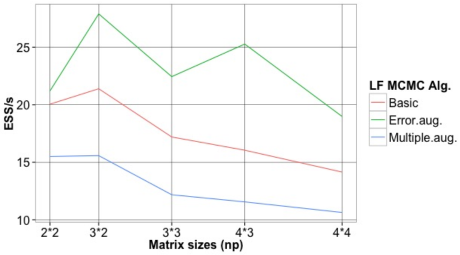

Figure: Comparing ESS/s for different variants of LF MCMC across various matrix normal sizes

Conclusion and Future scope

- Minkowski distance with $p = 3$ and Uniform kernel gave maximum ESS/s for all the variants and matrix sizes.
- Error augmented LF MCMC sampler appears to be performing better in terms of effective sample size per second.
- With increase in matrix sizes, ESS/s decreases for each variant in general. The run times increase exponentially.
- It would be interesting to implement the LF methods via Gaussian copula for higher order tensors following normal distribution.