Convergence of Gibbs Samplers and Output Analysis in a Bayesian Linear Model

Galin Jones

University of Minnesota

Bayesian Linear Model

For $i = 1, \ldots, \mathbf{K}$

$$Y_i | \theta_i, \gamma_i \overset{ind}{\sim} N(\theta_i, \gamma_i^{-1})$$

$$\theta_i | \mu, \lambda_\theta, \lambda_i \overset{ind}{\sim} \mathrm{N}(\mu, \lambda_\theta^{-1} \lambda_i^{-1})$$

$$\mu \sim \mathrm{N}(m_0, s_0^{-1}) \qquad \gamma_i \overset{iid}{\sim} \mathrm{Gamma}(\mathbf{a_3}, b_3)$$

$$\lambda_\theta \sim \mathrm{Gamma}(\mathbf{a_1}, b_1) \qquad \lambda_i \overset{iid}{\sim} \mathrm{Gamma}(a_2, b_2)$$

Posterior density

$$q(\theta, \gamma, \lambda, \mu, \lambda_\theta | y)$$

Bayesian Model

We want to calculate, say,

$$E[\theta_1|y] = \int \theta_1 \, q(\theta, \gamma, \lambda, \mu, \lambda_\theta|y) \, d\theta \, d\gamma \, d\lambda \, d\mu \, d\lambda_\theta$$

and

$$E[\gamma_1|y] = \int \gamma_1 \, q(\theta, \gamma, \lambda, \mu, \lambda_\theta|y) \, d\theta \, d\gamma \, d\lambda \, d\mu \, d\lambda_\theta$$

Gibbs Samplers for Bayesian Model

$$\lambda_\theta|\theta,\mu,\lambda,\gamma \sim \text{Gamma}\left(a_1^*, b_1^*(\lambda,\theta,\mu)\right)$$
$$\lambda_i|\theta,\mu,\lambda_\theta,\gamma \overset{ind}{\sim} \text{Gamma}\left(a_2^*, b_2^*(\lambda_\theta,\theta,\mu)\right)$$
$$\gamma_i|\theta,\mu,\lambda_\theta,\lambda \overset{ind}{\sim} \text{Gamma}\left(a_3^*, b_3^*(\theta)\right)$$
$$(\theta,\mu)|\lambda,\lambda_\theta,\gamma \sim \text{N}_{K+1}(\xi_0, V)$$

Gibbs Samplers:

$$((\theta^{(n)},\mu^{(n)}),\lambda_\theta^{(n)},\lambda^{(n)},\gamma^{(n)}) \to ((\theta^{(n+1)},\mu^{(n+1)}),\lambda_\theta^{(n+1)},\lambda^{(n+1)},\gamma^{(n+1)})$$

Bayesian Model

Simulated data:

```
> Ydata
[1] -1.703497 4.047338
```

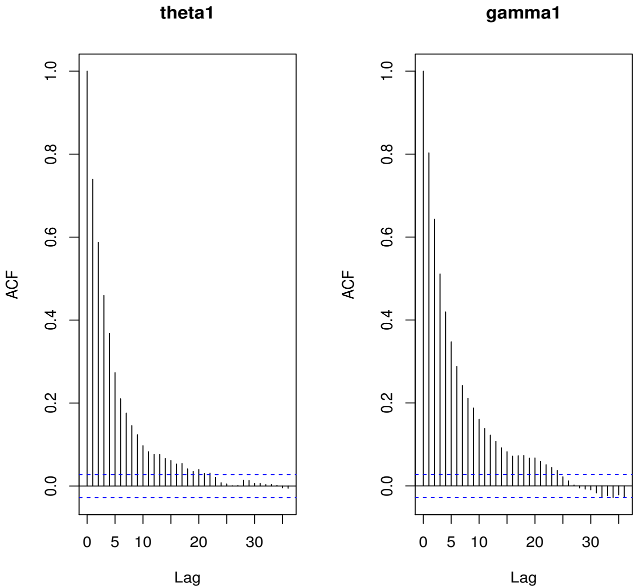
Simulate 5e3 realizations of the random scan Gibbs sampler to obtain:

```
> apply(tg.out, 2, mean)
[1] -1.313975 2.007110
```

Should we stop sampling?

```
> ess(tg.out)
[1] 606.5829 482.4933
```

Bayesian Model



Bayesian Model

After 5e3:

```
> apply(tg.out, 2, mean)
[1] -1.313975 2.007110

> ess(tg.out)
[1] 606.5829 482.4933
```

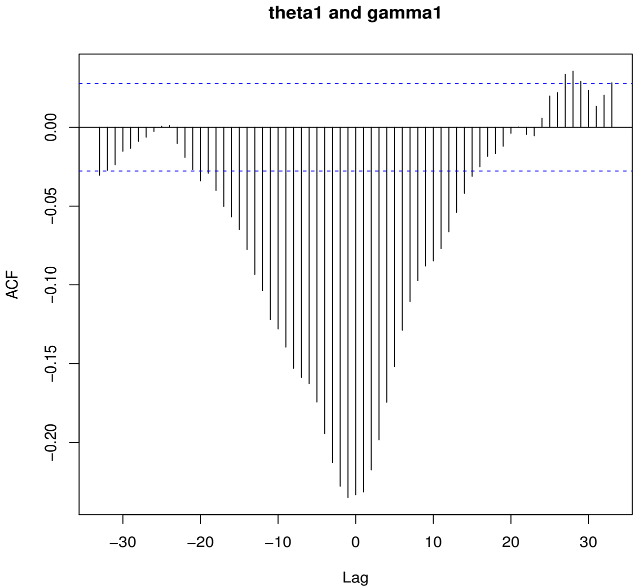
After 1.5e4:

```
> apply(tg.out1, 2, mean)
[1] -1.374471 2.018791

> ess(tg.out1)
[1] 2014.954 1475.851
```

We could keep going, but should we stop?

Bayesian Model



Multivariate Output Analysis

If $g : \mathcal{X} \to \mathbb{R}^p$, set

$$\eta = E_F g(X) = \int_{\mathcal{X}} g(x) F(dx) \ .$$

<u>SLLN</u>

$$\eta_n = \frac{1}{n} \sum_{i=0}^{n-1} g(X_i) \stackrel{a.s}{\to} E_F g(X) = \eta \quad n \to \infty$$

<u>CLT</u>

$$\sqrt{n}(\eta_n - \eta) \stackrel{d}{\to} \mathrm{N}_p(0, \Sigma) \quad n \to \infty$$

Multivariate Output Analysis

CLT

$$\sqrt{n}(\eta_n - \eta) \stackrel{d}{\to} \mathrm{N}_p(0, \Sigma) \quad n \to \infty$$

$$\Sigma = \Lambda + \sum_{k=1}^{\infty} \left[\mathrm{Cov}_F[g(X_1), g(X_{1+k})] + \mathrm{Cov}_F[g(X_1), g(X_{1+k})]^T \right]$$

where $\mathrm{Var}_F[g(X)] = \Lambda$ and if the Markov chain is reversible:

$$\Sigma = \Lambda + 2 \sum_{k=1}^{\infty} [\mathrm{Cov}_F[g(X_1), g(X_{1+k})]]$$

Multivariate Output Analysis

Estimating Σ:

Initial sequence estimators (Dai and Jones (2017), *J. Multivariate Analysis*)

Spectral variance estimators (Vats, Flegal, and Jones (2017), *Bernoulli*)

Batch means (Vats, Flegal, and Jones (2017), *Submitted*)

If the Markov chain is geometrically ergodic and $\|E_F g(X)\|^{2+\delta} < \infty$ for some $\delta > 0$, then spectral variance and batch means estimators are strongly consistent for Σ.

Output Analysis

If Σ_n estimates Σ, then a $100(1-\alpha)\%$ confidence region is

$$C(n) = \{\eta \in \mathbb{R}^p \,:\, n(\eta_n - \eta)^T \Sigma_n^{-1} (\eta_n - \eta) \leq F_*(\alpha)\}$$

If the Markov chain is geometrically ergodic and $\|E_F g(X)\|^{2+\delta} < \infty$ for some $\delta > 0$, then $C(n)$ is asymptotically valid in the sense that it will have coverage probability $1 - \alpha$.

Gibbs Samplers for Bayesian Model

The deterministically updated Gibbs sampler and the random scan Gibbs sampler are geometrically ergodic if $2a_1 + K - 2 > 0$ and $a_3 > 1$. (Johnson and Jones (2015) *J. Multivariate Analysis*)

Some conditional Metropolis-Hastings samplers are geometrically ergodic. (Jones, Roberts, and Rosenthal (2014) *Adv. Applied Prob.*)

Bayesian Model

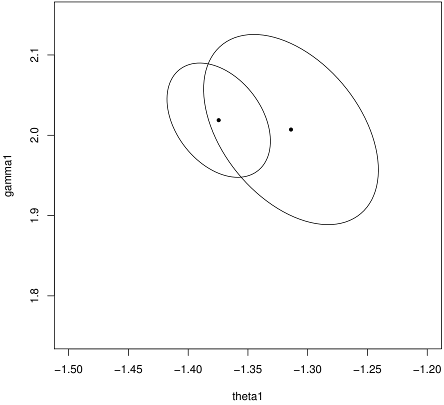


Figure: Confidence ellipses and estimates after 5e3 and 1.5e4 iterations.

Relative-Volume Stopping Rules

If $C(n)$ is the $100(1-\alpha)\%$ confidence region, then

$$\text{Volume}(C(n)) = k_n|\Sigma_n|^{1/2} \to 0 \quad n \to \infty$$

Terminate the first time that the volume of the confidence region is less than an ϵth fraction of the posterior standard deviation.

More formally, stop the first time after n^* that

$$\text{Volume}(C(n))^{1/p} + n^{-1} \leq \epsilon|\Lambda_n|^{1/2p}$$

Effective Sample Size

Define

$$\text{ESS} = n \left[\frac{|\Lambda|}{|\Sigma|} \right]^{1/p}$$

when $p = 1$ this reduces to the familiar

$$\text{ESS} = \frac{n}{1 + \sum_{i=1}^{\infty} \text{Corr}_F(g(X_1), g(X_{1+k}))}$$

We estimate ESS with

$$\text{ESS}_n = n \left[\frac{|\Lambda_n|}{|\Sigma_n|} \right]^{1/p}$$

Bayesian Model

n	ESS	ESS_1	ESS_2
5e3	582.9	606.5	482.5
1.5e4	1742.6	2015	1475.9

ESS as stopping rule

Stopping the first time after n^* that

$$\text{Volume}(C(n))^{1/p} + n^{-1} \leq \epsilon |\Lambda_n|^{1/2p}$$

is asymptotically equivalent to stopping when

$$\text{ESS}_n \geq \frac{2^{2/p}\pi}{(p\Gamma(p/2))^{2/p}} \frac{\chi^2_{1-\alpha,p}}{\epsilon^2}$$

Bayesian Model

To achieve a Monte Carlo error that is at most 10% of the posterior standard deviation with 90% confidence when $p = 2$ we need

$$\text{ESS}_n \geq 1447$$

```
> dim(tg.out1)
[1] 15000 2
> multiESS(tg.out1)
1742.674
```

```
> apply(tg.out1, 2, mean)
[1]-1.374471 2.018791
```

Bayesian Model

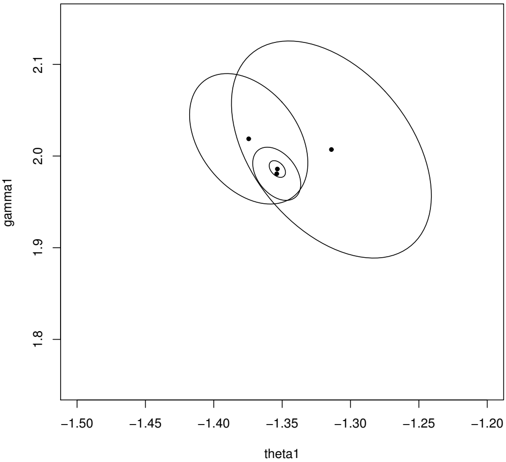


Figure: Ellipses and estimates after 5e3, 1.5e4, 1e5, and 1e6 iterations.

Discussion

The multivariate nature of MCMC estimation has largely been ignored.

Effective sample size can be used to assess the simulation in a principled manner.

Convergence rate of the Markov chain is *key*.

All of the output analysis methods in this talk are in the `mcmcse` R package available on CRAN.