A Reconsideration of the Gibbs' Sampler for Small Area Estimation Models

William R. Bell, U.S. Census Bureau

FCSM 2024

Disclaimer:

The views expressed here are those of the author and not those of the U.S. Census Bureau.

Background: Role of the Gibbs' sampler

- The Gibbs' sampler (Geman and Geman 1984, Gelfand and Smith 1990) has been enormously important to the development of Bayesian inference over the last 40 years.

Background: Role of the Gibbs' sampler

- The Gibbs' sampler (Geman and Geman 1984, Gelfand and Smith 1990) has been enormously important to the development of Bayesian inference over the last 40 years.
    - BUGS and JAGS have also been very important

Background: Role of the Gibbs' sampler

- The Gibbs' sampler (Geman and Geman 1984, Gelfand and Smith 1990) has been enormously important to the development of Bayesian inference over the last 40 years.
  - BUGS and JAGS have also been very important
- Many papers applying it to various statistical models.

Background: Role of the Gibbs' sampler

- The Gibbs' sampler (Geman and Geman 1984, Gelfand and Smith 1990) has been enormously important to the development of Bayesian inference over the last 40 years.
  - BUGS and JAGS have also been very important

- Many papers applying it to various statistical models.

- Natural application to linear mixed models, which are popular in small area estimation (SAE).

Background: Role of the Gibbs' sampler

- The Gibbs' sampler (Geman and Geman 1984, Gelfand and Smith 1990) has been enormously important to the development of Bayesian inference over the last 40 years.
  - BUGS and JAGS have also been very important

- Many papers applying it to various statistical models.

- Natural application to linear mixed models, which are popular in small area estimation (SAE).

- Leading MCMC algorithm; produces dependent simulation sequences.

MCMC: Gibbs' sampler vs alternatives

Question: How good is the Gibbs' sampler compared to alternative algorithms for linear SAE models?

MCMC: Gibbs' sampler vs alternatives

Question: How good is the Gibbs' sampler compared to alternative algorithms for linear SAE models?

Measure of performance of any MCMC algorithm would reflect the amount of dependence in the simulation sequences

- autocorrelation functions (ACFs)
- Monte Carlo variances for estimates of posterior quantities (e.g., for posterior means of parameters)

MCMC: Gibbs' sampler vs alternatives

Question: How good is the Gibbs' sampler compared to alternative algorithms for linear SAE models?

Measure of performance of any MCMC algorithm would reflect the amount of dependence in the simulation sequences

- autocorrelation functions (ACFs)
- Monte Carlo variances for estimates of posterior quantities (e.g., for posterior means of parameters)

We shall examine the performance of the Gibbs' sampler versus alternatives for two examples done using JAGS.

Example 1: Fay-Herriot (1979) model

$$y_i | \theta_i \sim N(\theta_i, v_i) \qquad i = 1, \ldots, m = 51$$

$$\theta_i | \beta, \sigma^2 \sim N(x_i'\beta, \sigma^2)$$

$$p(\beta, \sigma^2) \propto 1$$

- y_i = direct survey estimate of population characteristic θ_i for area i
- v_i = sampling variance of y_i (v_i assumed known)
- x_i = vector of regression variables (3 + intercept) for area i
- β = vector of regression parameters

Example 1: Gibbs' sampler for the FH model

Full conditional distributions for the FH model with $p(\beta, \sigma^2) \propto 1$, $\mathbf{y} = (y_1, \dots, y_m)'$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$:

- $[\beta|\mathbf{y}, \boldsymbol{\theta}, \sigma^2] = [\beta|\boldsymbol{\theta}, \sigma^2] \sim N(\hat{\beta}, V(\hat{\beta}))$

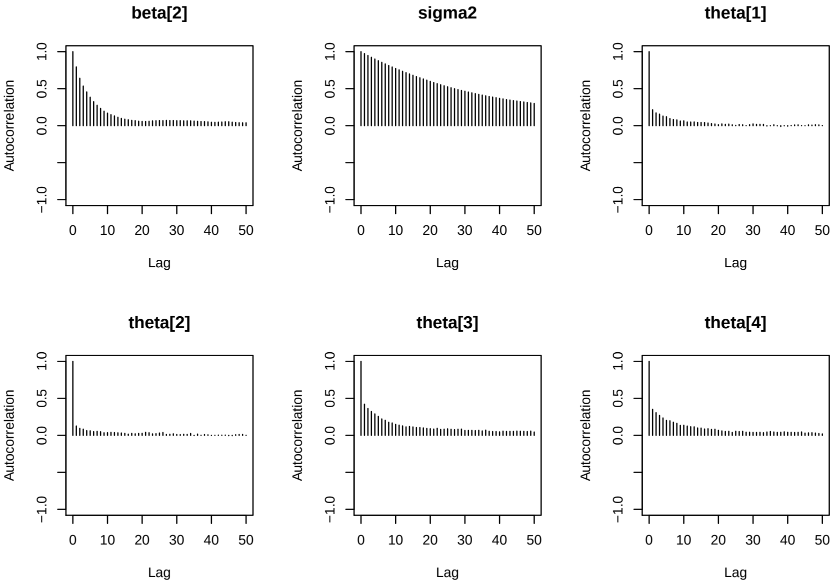
  where $\hat{\beta} = (X'X)^{-1}X'\boldsymbol{\theta},\ V(\hat{\beta}) = \sigma^2(X'X)^{-1} = O(1/m)$.

- $[\sigma^2|\mathbf{y}, \boldsymbol{\theta}, \beta] = [\sigma^2|\boldsymbol{\theta}, \beta] \sim \hat{\sigma}^2 \left(\frac{\chi^2_{m+2}}{m}\right)^{-1}$

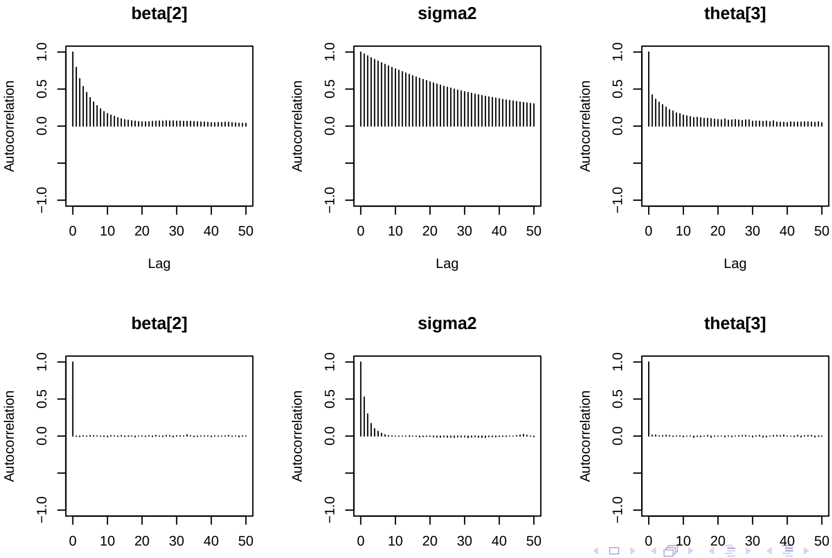
  where $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(\theta_i - x_i'\beta)^2$.

- $[\theta_i|y_i, \beta, \sigma^2] \sim N(\hat{\theta}_i, V(\hat{\theta}_i))$

  where $\hat{\theta}_i = h_i y_i + (1 - h_i)x_i'\beta,\ V(\hat{\theta}_i) = h_i v_i,\ h_i = \frac{\sigma^2}{\sigma^2 + v_i}$.


beta[2]
sigma2
theta[1]
theta[2]
theta[3]
theta[4]
Autocorrelation
Lag
−1.0
0.0
0.5
1.0
0
10
20
30
40
50

Example 1: FH model, ACFs for Gibbs' sampler and for non-hierarchical model specified to JAGS



Example 1: Fay-Herriot model

MCMC variance ratio: $\frac{\text{MC variance of posterior mean for hierarchical specification}}{\text{MC variance of posterior mean for non-hierarchical specification}}$

parameter	β_2	σ^2	θ_1	θ_2	θ_3
variance ratio	10.7	22.4	4.3	3.6	11.6

Example 1: Alternative specifications of the FH model

Hierarchical model specification:

$$y_i|\theta_i \sim N(\theta_i, v_i)$$
$$\theta_i|\beta, \sigma^2 \sim N(x_i'\beta, \sigma^2)$$

Non-hierarchical model specification:

$$y_i = \theta_i + e_i \sim N(x_i'\beta, \sigma^2 + v_i)$$

Here we "integrate out the random effects," which avoids the Gibbs' sampler.

Example 2: Bivariate functional measurement error model

For $i = 1, \ldots, m = 51$, let $y_i = (y_{1i}, y_{2i})'$ and $\theta_i = (\theta_{1i}, \theta_{2i})'$. Then

$$y_i|\theta_i \sim N(\theta_i, \mathbf{V}_i), \quad \mathbf{V}_i = \begin{bmatrix} v_{i11} & 0 \\ 0 & v_{i22} \end{bmatrix}$$

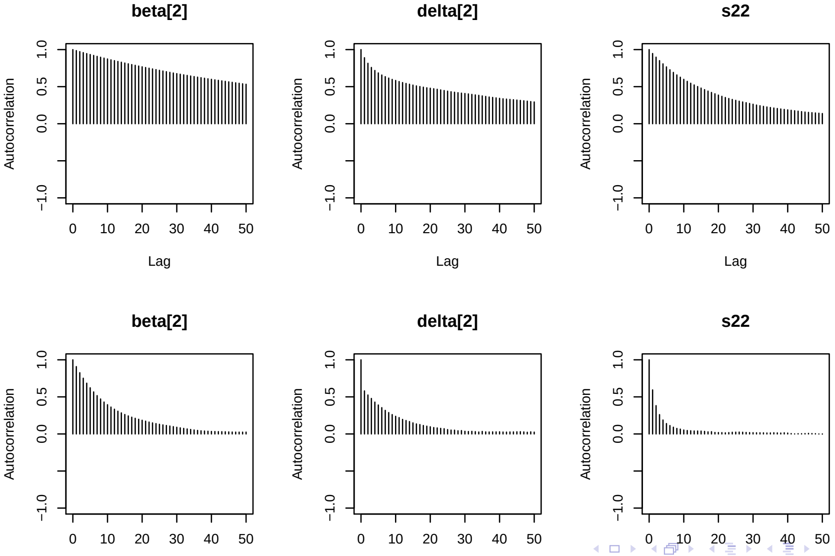
$$\begin{aligned} \theta_{i1}|x_i, \beta_1, \delta_1 &\sim N\left(\beta_1 x_i + z_{i1}'\delta_1, \sigma_{11}\right) \\ \theta_{i2}|x_i, \beta_2, \delta_2 &\sim N\left(\beta_2 x_i + z_{i2}'\delta_2, \sigma_{22}\right) \\ \mathrm{cov}(\theta_{i1}, \theta_{i2}|x_i, \beta, \delta) &= \sigma_{12} \qquad \rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}} \end{aligned}$$

$$X_i|x_i \sim N(x_i, C_i) \quad C_i \text{ assumed known}$$

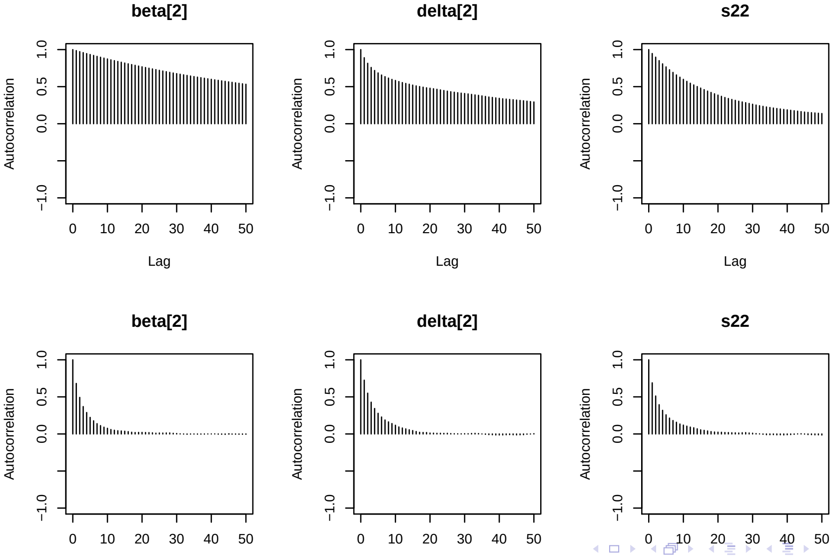
Note that

$$E(\theta_{ij}|X_i, \beta, \delta) = \beta_j X_i + z_{ij}'\delta_j \qquad \mathrm{var}(\theta_{ij}|X_i, \beta, \delta) = \beta_j^2 C_i + \sigma_{jj}.$$

Example 2: Bivariate functional measurement error model ACFs - Gibbs' sampler vs non-hierarchical spec to JAGS



Example 2: Bivariate functional measurement error model ACFs - Gibbs' sampler vs independence chain



Example 2: Bivariate functional measurement error model

MCMC variance ratio h/nh: $\frac{\text{MC variance of posterior mean for hierarchical spec}}{\text{MC variance of posterior mean for non-hierarchical spec}}$

MCMC variance ratio h/Ind: $\frac{\text{MC variance of posterior mean for hierarchical spec}}{\text{MC variance of posterior mean for independence chain}}$

parameter	β_2	δ_2	σ_{22}	ρ
variance ratio h/nh	6.7	5.7	7.0	9.8
variance ratio h/Ind	18.2	9.3	4.1	

Why the poor performance of the Gibbs' sampler?

Full conditional distributions for the FH model with $p(\beta, \sigma^2) \propto 1$, $\mathbf{y} = (y_1, \ldots, y_m)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$:

- $[\beta|\mathbf{y}, \boldsymbol{\theta}, \sigma^2] = [\beta|\boldsymbol{\theta}, \sigma^2] \sim N(\hat{\beta}, V(\hat{\beta}))$

  where $\hat{\beta} = (X'X)^{-1}X'\boldsymbol{\theta},\ V(\hat{\beta}) = \sigma^2(X'X)^{-1} = O(1/m)$.

- $[\sigma^2|\mathbf{y}, \boldsymbol{\theta}, \beta] = [\sigma^2|\boldsymbol{\theta}, \beta] \sim \hat{\sigma}^2 \left(\frac{\chi^2_{m+2}}{m}\right)^{-1}$

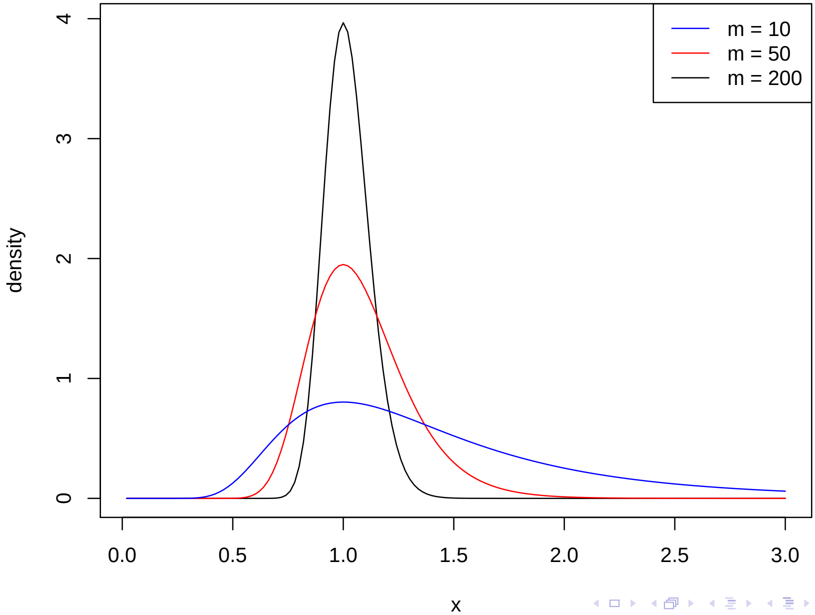
  where $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(\theta_i - x_i'\beta)^2$.

- $[\theta_i|y_i, \beta, \sigma^2] \sim N(\hat{\theta}_i, V(\hat{\theta}_i))$

  where $\hat{\theta}_i = h_i y_i + (1 - h_i)x_i'\beta,\ V(\hat{\theta}_i) = h_i v_i,\ h_i = \frac{\sigma^2}{\sigma^2 + v_i}$.

"Thinning" of the MCMC chains does not solve the dependence problem.

Densities of m/χ^2_{m+2}



Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.

Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.
- Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models, *Biometrics*, **56**, 768-774.

Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.

- Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models, *Biometrics*, **56**, 768-774.
  - accept/reject algorithm with a proposal density that approximates the posterior of the variance components; proposal remains fixed throughout simulations (independence chain)

Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.

- Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models, *Biometrics*, **56**, 768-774.

  - accept/reject algorithm with a proposal density that approximates the posterior of the variance components; proposal remains fixed throughout simulations (independence chain)
  - adapted to multivariate measurement error model by Arima, Bell, Datta, Franco, and Liseo (2017)

Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.

- Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models, *Biometrics*, **56**, 768-774.
    - accept/reject algorithm with a proposal density that approximates the posterior of the variance components; proposal remains fixed throughout simulations (independence chain)
    - adapted to multivariate measurement error model by Arima, Bell, Datta, Franco, and Liseo (2017)

- STAN

Alternative algorithms to the Gibbs' sampler for mixed linear models

- Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, **9**, 17-26.

- Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models, *Biometrics*, **56**, 768-774.
    - accept/reject algorithm with a proposal density that approximates the posterior of the variance components; proposal remains fixed throughout simulations (independence chain)
    - adapted to multivariate measurement error model by Arima, Bell, Datta, Franco, and Liseo (2017)

- STAN

- SAS

Conclusions

- Though convenient, the numerical performance of the Gibbs' sampler for mixed linear models, such as models used in small area estimation, is quite poor.
- More efficient algorithms are readily available, including by simply programming mixed linear models in non-hierarchical form in BUGS or JAGS.