Session 3: Gibbs Sampling

Bayesian Inference for the Social Sciences
November 3, 2017

Kevin Quinn

The Gibbs Sampling Algorithm

Some Examples
- Sampling from a Bivariate Normal Distribution
- The Linear Regression Model
- A Simple Univariate DLM
- A Hierarchical Beta-Binomial Model
- Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

Suppose we have a density $p(\theta_1, \ldots, \theta_k)$ where θ_i $i = 1, \ldots, k$ can be either univariate or multivariate

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

Suppose we have a density $p(\theta_1, \ldots, \theta_k)$ where θ_i $i = 1, \ldots, k$ can be either univariate or multivariate

We want to take a sample from the joint distribution of $\theta_1, \ldots, \theta_k$ but it is not easy to do this directly

The Gibbs Sampling Algorithm
Some Examples
Sampling from a Bivariate Normal Distribution
The Linear Regression Model
A Simple Univariate DLM
A Hierarchical Beta-Binomial Model
Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

Suppose we have a density $p(\theta_1, \ldots, \theta_k)$ where θ_i $i = 1, \ldots, k$ can be either univariate or multivariate

We want to take a sample from the joint distribution of $\theta_1, \ldots, \theta_k$ but it is not easy to do this directly

Gibbs sampling is another MCMC method that allows us to get a sample approximately from $p(\theta_1, \ldots, \theta_k)$ by iteratively sampling from the full conditional distributions

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

The Gibbs sampling algorithm works as follows

initialize $\theta_2^{(0)}, \ldots, \theta_k^{(0)}$

for (i = 1 to M){

sample $\theta_1^{(i)}$ from $p_1(\theta_1 | \theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)})$

sample $\theta_2^{(i)}$ from $p_2(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)} \ldots, \theta_k^{(i-1)})$

.

.

.

sample $\theta_k^{(i)}$ from $p_k(\theta_k | \theta_1^{(i)}, \ldots, \theta_{k-1}^{(i)})$

store $\theta_1^{(i)}, \ldots, \theta_k^{(i)}$

}

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

The Gibbs sampling algorithm works as follows

initialize $\theta_2^{(0)}, \ldots, \theta_k^{(0)}$

for (i = 1 to M){

sample $\theta_1^{(i)}$ from $p_1(\theta_1|\theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)})$

sample $\theta_2^{(i)}$ from $p_2(\theta_2|\theta_1^{(i)}, \theta_3^{(i-1)} \ldots, \theta_k^{(i-1)})$

.

.

.

sample $\theta_k^{(i)}$ from $p_k(\theta_k|\theta_1^{(i)}, \ldots, \theta_{k-1}^{(i)})$

store $\theta_1^{(i)}, \ldots, \theta_k^{(i)}$

}

Gibbs sampling is a special case of M-H sampling

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

The Gibbs sampling algorithm works as follows

initialize $\theta_2^{(0)}, \ldots, \theta_k^{(0)}$

for (i = 1 to M){

sample $\theta_1^{(i)}$ from $p_1(\theta_1 | \theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)})$

sample $\theta_2^{(i)}$ from $p_2(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)} \ldots, \theta_k^{(i-1)})$

.

.

.

sample $\theta_k^{(i)}$ from $p_k(\theta_k | \theta_1^{(i)}, \ldots, \theta_{k-1}^{(i)})$

store $\theta_1^{(i)}, \ldots, \theta_k^{(i)}$

}

Gibbs sampling is a special case of M-H sampling

Metropolis-Hastings within Gibbs

The Gibbs Sampling Algorithm

The Gibbs sampling algorithm works as follows

initialize $\theta_2^{(0)}, \ldots, \theta_k^{(0)}$

for (i = 1 to M){

  sample $\theta_1^{(i)}$ from $p_1(\theta_1 | \theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)})$

  sample $\theta_2^{(i)}$ from $p_2(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)} \ldots, \theta_k^{(i-1)})$

  .

  .

  .

  sample $\theta_k^{(i)}$ from $p_k(\theta_k | \theta_1^{(i)}, \ldots, \theta_{k-1}^{(i)})$

  store $\theta_1^{(i)}, \ldots, \theta_k^{(i)}$

}

Gibbs sampling is a special case of M-H sampling

Metropolis-Hastings within Gibbs

Sampling from a Bivariate Normal Distribution

Consider

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma^2_X & \sigma_{X,Y} \\ \sigma_{Y,X} & \sigma^2_Y \end{bmatrix} \right)$$

Let $\rho = \sigma_{X,Y}/\sqrt{\sigma^2_X \sigma^2_Y}$

Sampling from a Bivariate Normal Distribution

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Full conditionals:

$$[X|Y = y] \sim \mathcal{N}(m_1, s_1^2)$$

where

$$m_1 = \mu_X - \mu_Y(\sigma_{X,Y}/\sigma_Y^2) + y(\sigma_{X,Y}/\sigma_Y^2)$$

$$s_1^2 = \sigma_X^2(1 - \rho^2)$$

and

$$[Y|X = x] \sim \mathcal{N}(m_2, s_2^2)$$

where

$$m_2 = \mu_Y - \mu_X(\sigma_{X,Y}/\sigma_X^2) + x(\sigma_{X,Y}/\sigma_X^2)$$

$$s_2^2 = \sigma_Y^2(1 - \rho^2)$$

Sampling from a Bivariate Normal Distribution

The Gibbs sampling algorithm for the bivariate normal distribution is:

initialize $y^{(0)}$

for (i = 1 to M){

  sample $x^{(i)}|y^{(i-1)}$ from $\mathcal{N}(m_1, s_1^2)$

  sample $y^{(i)}|x^{(i)}$ from $\mathcal{N}(m_2, s_2^2)$

  store $x^{(i)}$ and $y^{(i)}$
}

Calculating the Full Conditional Distributions

To calculate the full conditional distributions necessary for Gibbs sampling do the following:

Calculating the Full Conditional Distributions

To calculate the full conditional distributions necessary for Gibbs sampling do the following:

1. Write out the full posterior ignoring constants of proportionality

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Calculating the Full Conditional Distributions

To calculate the full conditional distributions necessary for Gibbs sampling do the following:

1. Write out the full posterior ignoring constants of proportionality
2. Pick a block of parameters (say θ_1) and drop all terms on the rhs of the full posterior that don't depend on θ_1. (The full conditional for θ_1 is proportional to the stuff remaining on the rhs.)

Calculating the Full Conditional Distributions

To calculate the full conditional distributions necessary for Gibbs sampling do the following:

1. Write out the full posterior ignoring constants of proportionality
2. Pick a block of parameters (say θ_1) and drop all terms on the rhs of the full posterior that don't depend on θ_1. (The full conditional for θ_1 is proportional to the stuff remaining on the rhs.)
3. Use your knowledge of distribution theory to figure out what the missing normalizing constant is

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Calculating the Full Conditional Distributions

To calculate the full conditional distributions necessary for Gibbs sampling do the following:

1. Write out the full posterior ignoring constants of proportionality
2. Pick a block of parameters (say θ_1) and drop all terms on the rhs of the full posterior that don't depend on θ_1. (The full conditional for θ_1 is proportional to the stuff remaining on the rhs.)
3. Use your knowledge of distribution theory to figure out what the missing normalizing constant is
4. Repeat steps 2 and 3 for the remaining parameter blocks

The Linear Regression Model

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

The sampling density:

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right].$$

The Gibbs Sampling Algorithm

Some Examples
- Sampling from a Bivariate Normal Distribution
- The Linear Regression Model
- A Simple Univariate DLM
- A Hierarchical Beta-Binomial Model
- Authorship of the Disputed Federalist Papers

The Linear Regression Model

The sampling density:

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^2}\right].$$

The prior:

- Assume β is *a priori* independent of σ^2 so $p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta})p(\sigma^2)$

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{m}, \mathbf{V})$$

$$\sigma^2 \sim \mathcal{IG}(\nu/2, \delta/2)$$

The Linear Regression Model

The posterior (ignoring constants of proportionality):

$$
\begin{aligned}
p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\beta)'(\sigma^2 \mathbf{I}_n)^{-1}(\mathbf{y}-\mathbf{X}\beta)}{2}\right] \\
\times \exp\left[-\frac{(\boldsymbol{\beta}-\mathbf{m})'\mathbf{V}^{-1}(\beta-\mathbf{m})}{2}\right] \\
\times (\sigma^2)^{-(\nu/2+1)} \exp\left[-\frac{\delta/2}{\sigma^2}\right]
\end{aligned}
$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

To find the full conditionals drop terms that don't include the variable of interest.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

To find the full conditionals drop terms that don't include the variable of interest.

Full conditional for β:

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

To find the full conditionals drop terms that don't include the variable of interest.

Full conditional for β:

$$p_\beta(\boldsymbol{\beta}|\sigma^2, \mathbf{y}) \propto \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{I}_n)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}\right]$$
$$\times \exp\left[-\frac{(\boldsymbol{\beta}-\mathbf{m})'\mathbf{V}^{-1}(\boldsymbol{\beta}-\mathbf{m})}{2}\right]$$

The Gibbs Sampling Algorithm

Some Examples
- Sampling from a Bivariate Normal Distribution
- The Linear Regression Model
- A Simple Univariate DLM
- A Hierarchical Beta-Binomial Model
- Authorship of the Disputed Federalist Papers

The Linear Regression Model

To find the full conditionals drop terms that don't include the variable of interest.

Full conditional for $\boldsymbol{\beta}$:

$$p_\beta(\boldsymbol{\beta}|\sigma^2, \mathbf{y}) \propto \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{I}_n)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}\right] \times \exp\left[-\frac{(\boldsymbol{\beta}-\mathbf{m})'\mathbf{V}^{-1}(\boldsymbol{\beta}-\mathbf{m})}{2}\right]$$

A bit of algebra reveals

$$p_\beta(\boldsymbol{\beta}|\sigma^2, \mathbf{y}) \propto \exp\left[-\frac{(\boldsymbol{\beta}-\mathbf{m}^*)'\mathbf{V}^{*-1}(\boldsymbol{\beta}-\mathbf{m}^*)}{2}\right]$$

where

$$\mathbf{V}^* = \left(\mathbf{X}'(\sigma^2\mathbf{I})^{-1}\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}$$

$$\mathbf{m}^* = \mathbf{V}^*\left(\mathbf{X}'(\sigma^2\mathbf{I})^{-1}\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right).$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

Full conditional for σ^2:

$$
\begin{aligned}
p_{\sigma^2}(\sigma^2|\boldsymbol{\beta}, \mathbf{y}) &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right] \\
&\quad \times (\sigma^2)^{-(\nu/2+1)} \exp\left[-\frac{\delta/2}{\sigma^2}\right] \\
&\propto (\sigma^2)^{-(n/2+\nu/2+1)} \\
&\quad \times \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+\delta}{2\sigma^2}\right]
\end{aligned}
$$

This is an inverse gamma density with shape parameter $(n+\nu)/2$ and scale parameter $\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+\delta}{2}$.

The intuition here is that ν is acting like an additional number of observations and δ is acting like an additional sum of squared errors.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Linear Regression Model

The Gibbs sampling algorithm for the linear model is

initialize $\sigma^{2^{(0)}}$

for (i = 1 to M){

  sample $\beta^{(i)}$ from $p_\beta(\beta|\sigma^{2^{(i-1)}}, y)$

  sample $\sigma^{2^{(i)}}$ from $p_{\sigma^2}(\sigma^2|\beta^{(i)}, y)$

  store $\beta^{(i)}$ and $\sigma^{2^{(i)}}$
}

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Consider the following model for an observed univariate time-series $\{y_t\}_{t=1}^{T}$:

$$y_t = \theta_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad t = 1, \ldots, T$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Consider the following model for an observed univariate time-series $\{y_t\}_{t=1}^{T}$:

$$y_t = \theta_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad t = 1, \ldots, T$$

with prior beliefs

$$\theta_0 \sim \mathcal{N}(m_0, C_0)$$
$$\theta_t \sim \mathcal{N}(\theta_{t-1}, W) \quad t = 1, \ldots, T$$
$$\sigma^2 \sim \mathcal{I}nv\mathcal{G}amma(\alpha/2, \beta/2)$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Consider the following model for an observed univariate time-series $\{y_t\}_{t=1}^{T}$:

$$y_t = \theta_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad t = 1, \ldots, T$$

with prior beliefs

$$\theta_0 \sim \mathcal{N}(m_0, C_0)$$
$$\theta_t \sim \mathcal{N}(\theta_{t-1}, W) \quad t = 1, \ldots, T$$
$$\sigma^2 \sim \mathcal{I}nv\mathcal{G}amma(\alpha/2, \beta/2)$$

This is a simple example of a *Dynamic Linear Model (DLM)*.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Remarks:

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Remarks:

- This model has $(T + 1)$ parameters and T data points. Estimation is possible because of the random walk prior on θ

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Remarks:

- This model has $(T + 1)$ parameters and T data points. Estimation is possible because of the random walk prior on θ
- Without a proper prior on the unobserved θ_0 the joint prior for the θ vector is improper.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Remarks:

- This model has $(T+1)$ parameters and T data points. Estimation is possible because of the random walk prior on θ
- Without a proper prior on the unobserved θ_0 the joint prior for the θ vector is improper.
- This model is a particular instance of a more general class of models called *Markov Random Field (MRF)* models. These models are often used in image processing and spatial statistics.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Remarks:

- This model has $(T+1)$ parameters and T data points. Estimation is possible because of the random walk prior on θ
- Without a proper prior on the unobserved θ_0 the joint prior for the θ vector is improper.
- This model is a particular instance of a more general class of models called *Markov Random Field (MRF)* models. These models are often used in image processing and spatial statistics.
- The general DLM framework can be extended to handle multivariate time-series, covariates in the equation for y, covariates in the equation for θ, non-Gaussian responses, and many other things. In addition, many standard time-series models can be written in the DLM framework.

A SImple Univariate DLM

Remarks:

- This model has $(T+1)$ parameters and T data points. Estimation is possible because of the random walk prior on $\boldsymbol{\theta}$
- Without a proper prior on the unobserved θ_0 the joint prior for the $\boldsymbol{\theta}$ vector is improper.
- This model is a particular instance of a more general class of models called *Markov Random Field (MRF)* models. These models are often used in image processing and spatial statistics.
- The general DLM framework can be extended to handle multivariate time-series, covariates in the equation for y, covariates in the equation for θ, non-Gaussian responses, and many other things. In addition, many standard time-series models can be written in the DLM framework.
- For a thorough treatment of these models see West and Harrison, 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

How do we fit this model?

A SImple Univariate DLM

How do we fit this model?

Note that we can write the full posterior density (up to a constant of proportionality) as:

$$p(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) \propto \left\{\prod_{t=1}^{T} f_{\mathcal{N}}(y_t|\theta_t, \sigma^2)\right\} \left\{\prod_{t=1}^{T} f_{\mathcal{N}}(\theta_t|\theta_{t-1}, W)\right\} \times f_{\mathcal{N}}(\theta_0|m_0, C_0) f_{\mathcal{IG}}(\sigma^2|\alpha/2, \beta/2)$$

where $f_{\mathcal{N}}(\cdot|m, v)$ represents a normal density with mean m and variance v, and $f_{\mathcal{IG}}(\cdot|c, d)$ represents an inverse gamma density with shape c and scale d.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

One approach to model fitting is to use a Gibbs sampler that samples each univariate parameter from its full conditional in each scan of the sampler.

A SImple Univariate DLM

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

One approach to model fitting is to use a Gibbs sampler that samples each univariate parameter from its full conditional in each scan of the sampler.

From the expression for the posterior density above, it follows that the full condition for θ_0 is:

$$p(\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(\theta_1|\theta_0, W) f_{\mathcal{N}}(\theta_0|m_0, C_0)$$

A SImple Univariate DLM

One approach to model fitting is to use a Gibbs sampler that samples each univariate parameter from its full conditional in each scan of the sampler.

From the expression for the posterior density above, it follows that the full condition for θ_0 is:

$$p(\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(\theta_1|\theta_0, W) f_{\mathcal{N}}(\theta_0|m_0, C_0)$$

the full conditional for θ_t, $t \neq 0, T$ is:

$$p(\theta_t|\boldsymbol{\theta}_{\sim t}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(y_t|\theta_t, \sigma^2) f_{\mathcal{N}}(\theta_{t+1}|\theta_t, W) f_{\mathcal{N}}(\theta_t|\theta_{t-1}, W)$$

A SImple Univariate DLM

One approach to model fitting is to use a Gibbs sampler that samples each univariate parameter from its full conditional in each scan of the sampler.

From the expression for the posterior density above, it follows that the full condition for θ_0 is:

$$p(\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(\theta_1|\theta_0, W) f_{\mathcal{N}}(\theta_0|m_0, C_0)$$

the full conditional for $\theta_t,\ t \neq 0, T$ is:

$$p(\theta_t|\boldsymbol{\theta}_{\sim t}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(y_t|\theta_t, \sigma^2) f_{\mathcal{N}}(\theta_{t+1}|\theta_t, W) f_{\mathcal{N}}(\theta_t|\theta_{t-1}, W)$$

the full conditional for θ_T is:

$$p(\theta_T|\boldsymbol{\theta}_{\sim T}, \sigma^2, \mathbf{y}) \propto f_{\mathcal{N}}(y_T|\theta_T, \sigma^2) f_{\mathcal{N}}(\theta_T|\theta_{T-1}, W)$$

A SImple Univariate DLM

and the full conditional for σ^2 is:

$$p(\sigma^2|\boldsymbol{\theta}, \mathbf{y}) \propto \left\{ \prod_{t=1}^{T} f_{\mathcal{N}}(y_t|\theta_t, \sigma^2) \right\} f_{\mathcal{IG}}(\sigma^2|\alpha/2, \beta/2)$$

The Gibbs Sampling Algorithm

Some Examples
- Sampling from a Bivariate Normal Distribution
- The Linear Regression Model
- A Simple Univariate DLM
- A Hierarchical Beta-Binomial Model
- Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

A bit of algebra reveals that:

$$[\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_0}, V_{\theta_0})$$

where $V_{\theta_0} = (1/C_0 + 1/W)^{-1}$ and $m_{\theta_0} = V_{\theta_0}(m_0/C_0 + \theta_1/W)$

The Gibbs Sampling Algorithm

Some Examples
- Sampling from a Bivariate Normal Distribution
- The Linear Regression Model
- A Simple Univariate DLM
- A Hierarchical Beta-Binomial Model
- Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

A bit of algebra reveals that:

$$[\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_0}, V_{\theta_0})$$

where $V_{\theta_0} = (1/C_0 + 1/W)^{-1}$ and $m_{\theta_0} = V_{\theta_0}(m_0/C_0 + \theta_1/W)$

$$[\theta_t|\boldsymbol{\theta}_{\sim t}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_t}, V_{\theta_t}) \quad t = 1, \ldots, T-1$$

where $V_{\theta_t} = (1/\sigma^2 + 2/W)^{-1}$ and
$m_{\theta_t} = V_{\theta_t}(y_t/\sigma^2 + \theta_{t-1}/W + \theta_{t+1}/W)$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

A bit of algebra reveals that:

$$[\theta_0|\boldsymbol{\theta}_{\sim 0}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_0}, V_{\theta_0})$$

where $V_{\theta_0} = (1/C_0 + 1/W)^{-1}$ and $m_{\theta_0} = V_{\theta_0}(m_0/C_0 + \theta_1/W)$

$$[\theta_t|\boldsymbol{\theta}_{\sim t}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_t}, V_{\theta_t}) \;\; t = 1, \ldots, T-1$$

where $V_{\theta_t} = (1/\sigma^2 + 2/W)^{-1}$ and
$m_{\theta_t} = V_{\theta_t}(y_t/\sigma^2 + \theta_{t-1}/W + \theta_{t+1}/W)$

$$[\theta_T|\boldsymbol{\theta}_{\sim T}, \sigma^2, \mathbf{y}] \sim \mathcal{N}(m_{\theta_T}, V_{\theta_T})$$

where $V_{\theta_T} = (1/\sigma^2 + 1/W)^{-1}$ and $m_{\theta_T} = V_{\theta_T}(y_t/\sigma^2 + \theta_{t-1}/W)$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

and

$$[\sigma^2|\boldsymbol{\theta}, \mathbf{y}] \sim \mathcal{I}nv\mathcal{G}amma\left(\frac{\alpha + T}{2}, \frac{\beta + (\mathbf{y} - \boldsymbol{\theta})'(\mathbf{y} - \boldsymbol{\theta})}{2}\right)$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

This gives us enough information to implement the first sampling algorithm for this model:

initialize $\boldsymbol{\theta}^{(0)}$ and ${\sigma^2}^{(0)}$

for (i = 1 to M){

for (t = 0 to T){

sample $\theta_t^{(i)}$ from $\mathcal{N}(m_{\theta_t}, V_{\theta_t})$

}

sample ${\sigma^2}^{(i)}$ from $\mathcal{I}nv\mathcal{G}amma\left(\frac{\alpha+T}{2}, \frac{\beta+(\mathbf{y}-\boldsymbol{\theta}^{(i)})'(\mathbf{y}-\boldsymbol{\theta}^{(i)})}{2}\right)$

store $\boldsymbol{\theta}^{(i)}$ and ${\sigma^2}^{(i)}$

}

A SImple Univariate DLM

Another approach, called the forward-filtering-backward-sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994), is more complicated but allows one to sample θ in one block.

A SImple Univariate DLM

Another approach, called the forward-filtering-backward-sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994), is more complicated but allows one to sample $\boldsymbol{\theta}$ in one block.

Let $D_t \equiv \{y_t, D_{t-1}\}$ denote the information available up to time t where by convention we set $D_0 = \{W, m_0, C_0, \alpha, \beta\}$.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Another approach, called the forward-filtering-backward-sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994), is more complicated but allows one to sample $\boldsymbol{\theta}$ in one block.

Let $D_t \equiv \{y_t, D_{t-1}\}$ denote the information available up to time t where by convention we set $D_0 = \{W, m_0, C_0, \alpha, \beta\}$.

We can exploit the structure of the model to factor the full conditional density for $\boldsymbol{\theta}$ as:

$$p(\boldsymbol{\theta}|\sigma^2, D_T) = p(\theta_T|\sigma^2, D_T) \times p(\theta_{T-1}|\theta_T, \sigma^2, D_{T-1}) \times$$
$$p(\theta_{T-2}|\theta_{T-1}, \sigma^2, D_{T-2}) \times \cdots \times p(\theta_1|\theta_2, \sigma^2, D_1) \times p(\theta_0|\theta_1, \sigma^2, D_0)$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Another approach, called the forward-filtering-backward-sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994), is more complicated but allows one to sample $\boldsymbol{\theta}$ in one block.

Let $D_t \equiv \{y_t, D_{t-1}\}$ denote the information available up to time t where by convention we set $D_0 = \{W, m_0, C_0, \alpha, \beta\}$.

We can exploit the structure of the model to factor the full conditional density for $\boldsymbol{\theta}$ as:

$$p(\boldsymbol{\theta}|\sigma^2, D_T) = p(\theta_T|\sigma^2, D_T) \times p(\theta_{T-1}|\theta_T, \sigma^2, D_{T-1}) \times p(\theta_{T-2}|\theta_{T-1}, \sigma^2, D_{T-2}) \times \cdots \times p(\theta_1|\theta_2, \sigma^2, D_1) \times p(\theta_0|\theta_1, \sigma^2, D_0)$$

The forward-filtering-backward-sampling algorithm works by using the Kalman filter to calculate the quantities necessary to construct the densities on the rhs of the equation above and to then sample the components of $\boldsymbol{\theta}$ backward in time.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm
Some Examples
Sampling from a Bivariate Normal Distribution
The Linear Regression Model
A Simple Univariate DLM
A Hierarchical Beta-Binomial Model
Authorship of the Disputed Federalist Papers

A SImple Univariate DLM

Forward-Filtering-Backward-Sampling Algorithm:

initialize ${\sigma^2}^{(0)}$
for (i = 1 to M){
  for (t = 1 to T){
    $a_t \leftarrow m_{t-1};$ $\quad R_t \leftarrow C_{t-1} + W$
    $f_t \leftarrow a_t;$ $\quad Q_t \leftarrow R_t + \sigma^2$
    $e_t \leftarrow y_t - f_t;$ $\quad A_t \leftarrow R_t / Q_t$
    $m_t \leftarrow a_t + A_t e_t;$ $\quad C_t \leftarrow R_t - A_t Q_t A_t$
  }
  sample $\theta_T^{(i)}$ from $\mathcal{N}(m_T, C_T)$
  for (t = (T-1) to 1){
    $B \leftarrow C_t / R_{t+1}$
    $h \leftarrow m_t + B(\theta_{t+1} - a_{t+1})$
    $H \leftarrow C_t - BR_{t+1}B$
    sample $\theta_t^{(i)}$ from $\mathcal{N}(h, H)$
  }
  sample ${\sigma^2}^{(i)}$ from $\mathcal{I}nv\mathcal{G}amma\left(\frac{\alpha+T}{2}, \frac{\beta+(\mathbf{y}-\boldsymbol{\theta}^{(i)})'(\mathbf{y}-\boldsymbol{\theta}^{(i)})}{2}\right)$
  store $\boldsymbol{\theta}^{(i)}$ and ${\sigma^2}^{(i)}$
}

A SImple Univariate DLM

Forward-Filtering-Backward-Sampling Algorithm:

initialize $\sigma^{2^{(0)}}$
for (i = 1 to M){
  for (t = 1 to T){
    $a_t \leftarrow m_{t-1}$; $\quad R_t \leftarrow C_{t-1} + W$
    $f_t \leftarrow a_t$; $\quad Q_t \leftarrow R_t + \sigma^2$
    $e_t \leftarrow y_t - f_t$; $\quad A_t \leftarrow R_t / Q_t$
    $m_t \leftarrow a_t + A_t e_t$; $\quad C_t \leftarrow R_t - A_t Q_t A_t$
  }
  sample $\theta_T^{(i)}$ from $\mathcal{N}(m_T, C_T)$
  for (t = (T-1) to 1){
    $B \leftarrow C_t / R_{t+1}$
    $h \leftarrow m_t + B(\theta_{t+1} - a_{t+1})$
    $H \leftarrow C_t - B R_{t+1} B$
    sample $\theta_t^{(i)}$ from $\mathcal{N}(h, H)$
  }
  sample $\sigma^{2^{(i)}}$ from $\mathcal{I}nv\mathcal{G}amma\left(\frac{\alpha+T}{2}, \frac{\beta+(\mathbf{y}-\boldsymbol{\theta}^{(i)})'(\mathbf{y}-\boldsymbol{\theta}^{(i)})}{2}\right)$
  store $\boldsymbol{\theta}^{(i)}$ and $\sigma^{2^{(i)}}$
}

While this algorithm is much more complicated, at each scan it produces a draw of $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\sigma^2, \mathbf{y})$ rather than T draws from the univariate full conditionals.

A SImple Univariate DLM

Trump approval ratings

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A Hierarchical Beta-Binomial Model

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Consider the following data:

- i indexes geographic units (census blocks)
- m_i is the number of survey respondents in geographic unit i
- Y_i is the number of "yes" responses to a particular yes/no question
- θ_i is the fraction of "yes" responders among all residents of geographic unit i who are in the sampling frame
- We believe the geographic-unit-specific "yes" response fractions vary quite a bit across the population of geographic units.

How should we model these data?

A Hierarchical Beta-Binomial Model

Consider the following model that we might use:

$$
\begin{aligned}
Y_i &\overset{ind.}{\sim} \mathcal{B}inomial(m_i, \theta_i), \qquad i = 1, \ldots, n \\
\theta_i &\overset{ind.}{\sim} \mathcal{B}eta(\alpha, \beta), \qquad i = 1, \ldots, n \\
\alpha &\sim \mathcal{E}xponential(a) \\
\beta &\sim \mathcal{E}xponential(b)
\end{aligned}
$$

where it is assumed that α and β are a priori independent.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A Hierarchical Beta-Binomial Model

In what follows, I'll use the following shorthand:

$$f_{bin}(y|m,\theta) = \binom{m}{y}\theta^{y}(1-\theta)^{m-y}$$

$$f_{beta}(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$f_{exp}(\alpha|a) = \begin{cases} a\exp(-a\alpha) & \text{if } \alpha \geq 0 \\ 0 & \text{if } \alpha < 0 \end{cases}$$

$$f_{bb}(y|m,\alpha,\beta) = \binom{m}{y}\frac{B(y+\alpha,\ m-y+\beta)}{B(\alpha,\ \beta)}$$

where Γ is the gamma function and B is the beta function. $\mathbf{y}$ denotes the n-vector of the realized values of Y_is. Also, we'll let $\boldsymbol{\theta}$ denote the n-vector of θ_i values.

A Hierarchical Beta-Binomial Model

We can write the posterior for α, β, and $\boldsymbol{\theta}$ as:

$$p(\alpha, \beta, \boldsymbol{\theta}|\mathbf{y}) \propto \left[f_{exp}(\alpha|a) f_{exp}(\beta|b) \prod_{i=1}^{n} f_{beta}(\theta_i|\alpha, \beta) \right] \left\{ \prod_{i=1}^{n} f_{bin}(y_i|m_i, \theta_i) \right\}$$

The terms in the square brackets are our prior for $(\alpha, \beta, \boldsymbol{\theta})$. The piece in the curved brackets is the likelihood given $\boldsymbol{\theta}$.

Note that y_i is assumed to be conditionally independent of (α, β) given θ_i. Thus, after conditioning on $\boldsymbol{\theta}$, α and β don't show up in the likelihood.

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A Hierarchical Beta-Binomial Model

We can also write the posterior for α and β marginalized over $\boldsymbol{\theta}$ as:

$$\begin{aligned} p(\alpha, \beta|\mathbf{y}) &\propto f_{exp}(\alpha|a) f_{exp}(\beta|b) \prod_{i=1}^{n} \int_0^1 f_{beta}(\theta_i|\alpha,\beta) f_{bin}(y_i|m_i,\theta_i) d\theta_i \\ &\propto f_{exp}(\alpha|a) f_{exp}(\beta|b) \prod_{i=1}^{n} f_{bb}(y_i|m_i,\alpha,\beta) \end{aligned} \tag{1}$$

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A Hierarchical Beta-Binomial Model: Model-Fitting Strategy 1

The idea here is to sample (α, β) from $p(\alpha, \beta|\mathbf{y})$ and to then sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\mathbf{y}, \alpha, \beta)$ where the values of α and β we condition on are those drawn from the first step.

In essence, this is just a variation on the method of composition. $p(\boldsymbol{\theta}|\mathbf{y}, \alpha, \beta)$ is easy to sample from—each $[\theta_i|\mathbf{y}, \alpha, \beta] \overset{ind.}{\sim} \mathcal{B}eta(\alpha + y_i,\ \beta + m_i - y_i)$.

Sampling (α, β) from $p(\alpha, \beta|\mathbf{y})$ is more complicated. To do that we'll use a random walk Metropolis algorithm.

A Hierarchical Beta-Binomial Model: Model-Fitting Strategy 2

A second approach does not analytically marginalize over $\boldsymbol{\theta}$.

Here we iteratively sample from $[\alpha, \beta | \mathbf{y}, \boldsymbol{\theta}]$ and $[\boldsymbol{\theta} | \mathbf{y}, \alpha, \beta]$.

If we could directly sample from these conditional distributions this would be an example of Gibbs sampling.

Directly sampling from $[\alpha, \beta | \mathbf{y}, \boldsymbol{\theta}]$ is not possible so we need to use a Metropolis step for this piece. This gives rise to what is sometimes called a Metropolis within Gibbs setup.

Before proceeding, note that

$$p(\alpha, \beta | \mathbf{y}, \boldsymbol{\theta}) \propto f_{exp}(\alpha | a) f_{exp}(\beta | b) \prod_{i=1}^{n} f_{beta}(\theta_i | \alpha, \beta) \tag{2}$$

(the likelihood for $\mathbf{y}$ drops out because $\mathbf{y}$ and (α, β) are conditionally independent given $\boldsymbol{\theta}$.)

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

A Hierarchical Beta-Binomial Model

Which of these model-fitting strategies do you think will work best? Why?

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Authorship of the Disputed Federalist Papers

The Gibbs Sampling Algorithm

Some Examples

Sampling from a Bivariate Normal Distribution

The Linear Regression Model

A Simple Univariate DLM

A Hierarchical Beta-Binomial Model

Authorship of the Disputed Federalist Papers

Another approach to this authorship problem is to fit a two-component multinomial mixture model to these data with the constraint that the essays with known authorship are in particular clusters with certainty.

`R` example