Markov Chain Monte Carlo Methods: Gibbs Sampling

Mihir Arjunwadkar

Centre for Modeling and Simulation
Savitribai Phule Pune University

Basic MCMC: Gibbs sampling for bivariate PDF

Gibbs sampling for a bivariate PDF $f_{XY}(x, y)$

Assume that the two conditional PDFs $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$ are (easily) sampleable.

Start with an arbitrary initial value x_1.

1. Sample $y_1 \sim f_{Y|X}(\cdot|x_1)$ $\Longrightarrow (x_1, y_1)$
2. Sample $x_2 \sim f_{X|Y}(\cdot|y_1)$
3. Sample $y_2 \sim f_{Y|X}(\cdot|x_2)$ $\Longrightarrow (x_2, y_2)$
4. Sample $x_3 \sim f_{X|Y}(\cdot|y_2)$
5. Sample $y_3 \sim f_{Y|X}(\cdot|x_3)$ $\Longrightarrow (x_3, y_3)$
6. ...

Repeat as required.

Basic MCMC: Gibbs sampling for bivariate PDF

Expressing the previous algorithm concisely

1. Start with an arbitrary initial value x_1.
2. (x_{i+1}, y_{i+1}) is obtained from (x_i, y_i) by sampling from the conditionals:

$$
\begin{aligned}
y_{i+1} &\sim f_{Y|X}(\cdot|x_i) \\
x_{i+1} &\sim f_{X|Y}(\cdot|y_{i+1})
\end{aligned}
$$

3. Repeat step 2 as required.

Example: Gibbs sampling for a bivariate normal density

Let

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Let us consider the bivariate normal PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi} |\mathbf{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \right]$$

where $\mathbf{\Sigma}^{-1}$ is the matrix inverse of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}.$$

The determinant $|\mathbf{\Sigma}| = 1 - \rho^2$.

https://www2.stat.duke.edu/courses/Spring12/sta104.1/Lectures/Lec22.pdf

Example: Gibbs sampling for a bivariate normal density

Marginal PDFs

$$
\begin{aligned}
f_X(x) &\equiv N(\mu_X, 1) \\
f_Y(y) &\equiv N(\mu_Y, 1)
\end{aligned}
$$

Example: Gibbs sampling for a bivariate normal density

Conditional PDFs

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{x-\rho y}{2(1-\rho^2)}\right] \equiv N(\rho y, 1-\rho^2) \\
f_{Y|X}(y|x) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{y-\rho x}{2(1-\rho^2)}\right] \equiv N(\rho x, 1-\rho^2)
\end{aligned}
$$

We know how to sample from these univariate normal PDFs ...

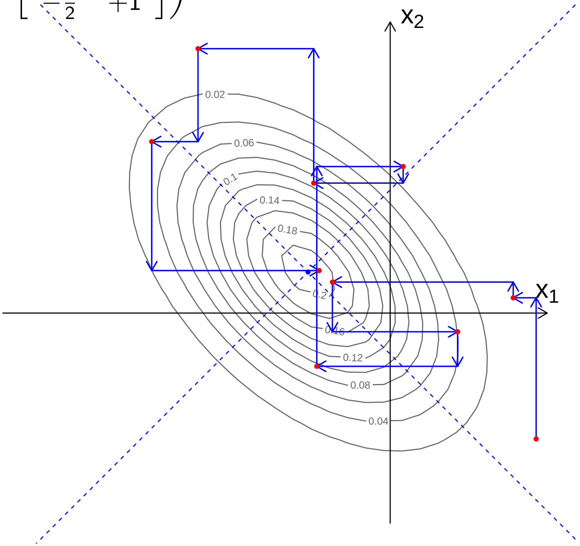
`https://www2.stat.duke.edu/courses/Spring12/sta104.1/Lectures/Lec22.pdf`

Example: Gibbs sampling for bivariate normal

```
rbvnorm.gibbs <- function( n, x1, mu = c( 0, 0 ), rho = 0 )
 {
  # Gibbs sampler for a correlated bivariate normal N( mu, Sigma2 )
  # where
  #  mu = [ mu1 mu2 ]
  # and
  #  covariance matrix Sigma^2 = [ 1   rho ]
  #                               [ rho   1 ]

  stopifnot( abs( rho ) < 1, length( mu ) == 2, is.numeric( mu ) )

  r <- matrix( nrow = n, ncol = 2 )

  sd <- sqrt( 1 - rho^2 )

  r[1,1] <- x1
  r[1,2] <- rnorm( 1, mean = rho * r[1,1], sd = sd )
  for ( i in 2:n )
   {
    r[i,1] <- rnorm( 1, mean = rho * r[i-1,2], sd = sd )
    r[i,2] <- rnorm( 1, mean = rho * r[i,  1], sd = sd )
   }

  cbind( r[,1] + mu[1], r[,2] + mu[2] )
 }
```

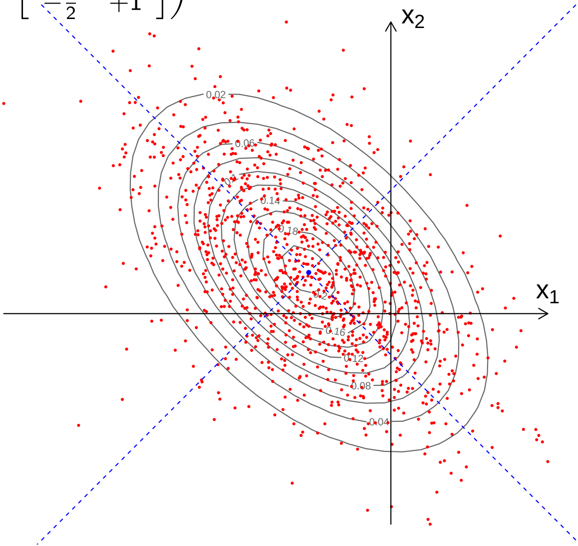
Example: Gibbs sampling for bivariate normal

$N\left(\left(-1,+\frac{1}{2}\right),\begin{bmatrix} +1 & -\frac{1}{2} \\ -\frac{1}{2} & +1 \end{bmatrix}\right)$ :: Gibbs :: Initial trajectory



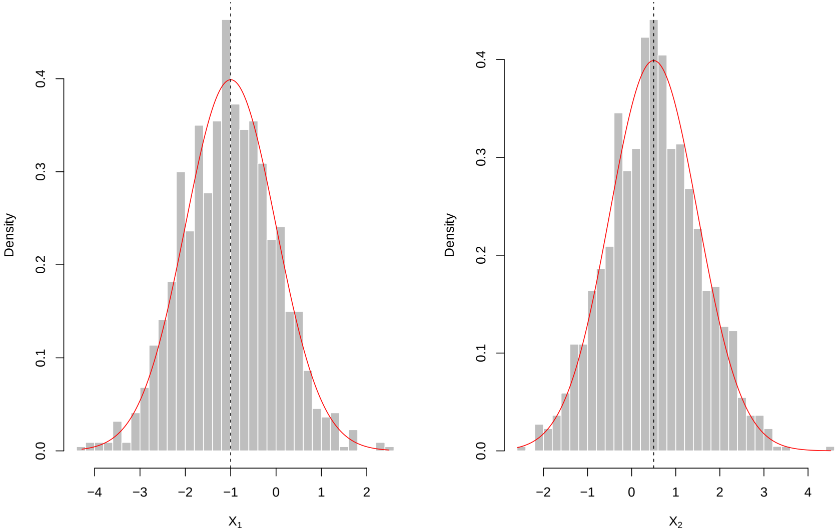
Example: Gibbs sampling for bivariate normal

$N\left(\left(-1,+\frac{1}{2}\right),\begin{bmatrix}+1 & -\frac{1}{2}\\ -\frac{1}{2} & +1\end{bmatrix}\right)$ :: Gibbs :: A larger sample



Example: Gibbs sampling for bivariate normal

$N\left(\left(-1, +\frac{1}{2}\right), \begin{bmatrix} +1 & -\frac{1}{2} \\ -\frac{1}{2} & +1 \end{bmatrix}\right)$ :: Gibbs :: Marginal histograms



Example: Gibbs sampling for bivariate normal

Suppose that the Gibbs sampling trajectory is

$$x_1 \to y_1 \to x_2 \to y_2 \to \ldots \to x_t \to y_t \to \ldots$$

- Time evolution of the joint PDF of (X, Y) under this Markov chain:

  $$N\left(\begin{bmatrix} \mu_X + \rho^{2t} x_1 \\ \mu_Y + \rho^{2t+1} x_1 \end{bmatrix}, \begin{bmatrix} 1 - \rho^{4t} & \rho(1 - \rho^{4t}) \\ \rho(1 - \rho^{4t}) & 1 - \rho^{4t} \end{bmatrix}\right)$$

  for $|\rho| < 1$.

- As $t \to \infty$, this becomes

  $$N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

  which is the target PDF.

More examples

in the handout

on the web

...

Gibbs sampling for $f_{XYZ}(x,y,z)$. Assume that the conditionals

$$f_{Z|XY}(z|x,y), f_{Y|XZ}(y|x,z), f_{X|YZ}(x|y,z)$$

are (easily) sampleable.

1. Start with arbitrary initial values x_1, y_1.
2. $(x_{i+1}, y_{i+1}, z_{i+1})$ is obtained from (x_i, y_i, z_i) by sampling from the conditionals:

$$\begin{aligned} z_{i+1} &\sim f_{Z|XY}(\cdot|x_i, y_i) \\ y_{i+1} &\sim f_{Y|XZ}(\cdot|x_i, z_{i+1}) \\ x_{i+1} &\sim f_{X|YZ}(\cdot|y_{i+1}, z_{i+1}) \end{aligned}$$

3. Repeat step 2 as required.

Generalizing Gibbs for a multivariate PDF

Suppose

$f(x_1, \ldots, x_k)$: Target multivariate PDF

$f_i(x_i | x_{-i})$: conditional PDF of x_i given everything else (x_{-i})

New notation $x_{-i} \equiv$ "all variables except the ith"

Gibbs for a multivariate PDF: Sequential scan

1. Start with arbitrary initial values $x_1, \ldots, x_k$.
2. For $i = 1, \ldots, k$:
   Sample $x_i \sim f_i(\cdot | x_{-i})$
3. Repeat step 2 as required.

Gibbs for a multivariate PDF: Random scan

1. Start with arbitrary initial values $x_1, \ldots, x_k$.
2. For $i \in$ a random permutation of $1, \ldots, k$:
   Sample $x_i \sim f_i(\cdot|x_{-i})$
3. Repeat step 2 as required.

Gibbs for a multivariate PDF

- The basic variant discussed changes one variable at a time.
- Block versions, where the conditionals are expressed over groups of variables instead of single variables, are also possible.

Why Gibbs works

- Gibbs algorithm sets up a Markov chain, because the next state depends only on the current state.
- Gibbs can be shown to be a special case of Metropolis-Hastings with acceptance probability $= 1$.

Gibbs as Metropolis-Hastings with P(accept)=1

- $f(x_1, \ldots, x_k)$: Target multivariate PDF
- $f_i(x_i|x_{-i})$: conditional PDF of x_i given everything else :: x_{-i}
- Suppose the current state of the Markov chain is $(x_1, \ldots, x_k)$.
- The Gibbs sampler is now ready to change the ith variable x_i.
- Think of Gibbs as Metropolis-Hastings with proposal PDF

$$q(\cdot|x_i, x_{-i}) = f_i(\cdot|x_{-i}).$$

- Generate a candidate from this proposal PDF,

$$x_i' \sim f_i(\cdot|x_{-i}).$$

Gibbs as Metropolis-Hastings with P(accept)=1

Recall the Metropolis-Hastings form of the acceptance probability

$$a(\text{proposed}|\text{current}) = \min\left\{1, \frac{f(\text{proposed})}{f(\text{current})} \times \frac{q(\text{current}|\text{proposed})}{q(\text{proposed}|\text{current})}\right\}$$

In our Gibbs sampling setting,

- Current state of the chain: (x_i, x_{-i})
- Proposed state of the chain: (x_i', x_{-i})

Gibbs as Metropolis-Hastings with P(accept)=1

For convenience, let us write $f(x_1, \ldots, x_k)$ as $f(x_i, x_{-i})$.

M-H acceptance probability for Gibbs sampling is
$a(x_i', x_{-i} | x_i, x_{-i}) = \min(1, \alpha)$, where

$$
\begin{aligned}
\alpha &= \frac{f(x_i', x_{-i})}{f(x_i, x_{-i})} \times \frac{q(x_i, x_{-i} | x_i', x_{-i})}{q(x_i', x_{-i} | x_i, x_{-i})} \\
&= \frac{f(x_i', x_{-i})}{f(x_i, x_{-i})} \times \frac{f_i(x_i | x_{-i})}{f_i(x_i' | x_{-i})} \\
&= \frac{f_i(x_i' | x_{-i}) f(x_{-i})}{f_i(x_i | x_{-i}) f(x_{-i})} \times \frac{f_i(x_i | x_{-i})}{f_i(x_i' | x_{-i})} \\
&= 1.
\end{aligned}
$$

That is, $a(x_i', x_{-i} | x_i, x_{-i}) = 1$. This is true for any i.

https://ermongroup.github.io/cs323-notes/probabilistic/gibbs/
http://theanalysisofdata.com/notes/metropolis.pdf

Convergence, caveats, and burn-in/equilibration considerations applicable to M-H are also applicable to Gibbs – except for tunable parameters in the proposal PDF of M-H.

Further reading

- A classic

  Explaining the Gibbs Sampler. Casella & George, The American Statistician, Vol. 46, No. 3 (Aug., 1992), pp. 167-174. http://www.jstor.org/stable/2685208

- A perspective on comparison between Gibbs and M-H:

  https://stats.stackexchange.com/questions/104815/gibbs-sampling-versus-general-mh-mcmc

- Advanced texts with rigorous treatment of the subject
  - Monte Carlo Statistical Methods. Robert & Casella, Springer 2004
  - Advanced Markov Chain Monte Carlo Methods. Liang, Liu, & Carroll, Wiley 2010