MCMC Gibbs Sampler.

Exercises.

Anatoli Iambartsev

IME-USP

[RC] Two-stage Gibbs Sampler.

7.2 The two-stage Gibbs sampler

The *two-stage Gibbs sampler* creates a Markov chain from a joint distribution in the following way. If two random variables X and Y have joint density $f(x, y)$, with corresponding conditional densities $f_{Y|X}$ and $f_{X|Y}$, the two-stage Gibbs sampler generates a Markov chain (X_t, Y_t) according to the following steps:

Algorithm 7 Two-stage Gibbs sampler
Take $X_0 = x_0$
For $t = 1, 2, \ldots$, generate
    1. $Y_t \sim f_{Y|X}(\cdot|x_{t-1})$;
    2. $X_t \sim f_{X|Y}(\cdot|y_t)$.

[RC] Example 7.1. To start with an obvious illustration, consider the bivariate normal model

$$(X, Y) \sim N\left((0,0), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

for which the Gibbs sampler is: given x_t, generate

$$\begin{aligned} Y_{t+1} \mid x_t &\sim N(\rho x_t, 1 - \rho^2), \\ X_{t+1} \mid y_t &\sim N(\rho y_{t+1}, 1 - \rho^2). \end{aligned}$$

[RC] Example 7.1.

Indeed, (from wiki): If N-dimensional $\mathbf{x}$ is partitioned as follows

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N - q) \times 1 \end{bmatrix}$$

and accordingly $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as follows

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N - q) \times 1 \end{bmatrix}$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times q & q \times (N - q) \\ (N - q) \times q & (N - q) \times (N - q) \end{bmatrix}$$

then the distribution of $\mathbf{x}_1$ conditional on $\mathbf{x}_2 = a$ is multivariate normal $(\mathbf{x}_1 \mid \mathbf{x}_2 = a) \sim N(\bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}})$ where

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \boldsymbol{\mu}_2), \overline{\boldsymbol{\Sigma}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

[RC] Example 7.1.

In the bivariate case

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

$\mathbf{x}$ is partitioned into X_1 and X_2, the conditional distribution of X_1 given X_2 is

$$(X_1 \mid X_2 = x_2) \sim N\left(\mu_1 + \rho(x_2 - \mu_2), (1 - \rho^2) \right),$$

Coming back to our example $\mu_1 = \mu_2 = 0$ we obtain

$$\begin{aligned} Y_{t+1} \mid x_t &\sim N(\rho x_t, 1 - \rho^2), \\ X_{t+1} \mid y_t &\sim N(\rho y_{t+1}, 1 - \rho^2). \end{aligned}$$

Then

$$X_{t+1} \mid X_t = x_t \sim N(\rho^2 x_t, 1 - \rho^4)$$

[RC] Example 7.1.

$$X_{t+1} \mid X_t = x_t \sim N(\rho^2 x_t, 1 - \rho^4)$$

Proof:

$$\mathbb{E}(\mathbb{E}(X_{t+1} \mid Y_{t+1}) \mid X_t = x_t)$$
$$= \mathbb{E}(\rho Y_{t+1} \mid X_t = x_t) = \rho \mathbb{E}(Y_{t+1} \mid X_t = x_t) = \rho^2 x_t,$$

$$\mathbb{V}ar(X_{t+1} \mid X_t = x_t) = \mathbb{V}ar(X_{t+1} \mid Y_{t+1}, X_t = x_t)$$
$$= \mathbb{E}(\mathbb{V}ar(X_{t+1} \mid Y_{t+1}) \mid X_t = x_t) + \mathbb{V}ar(\mathbb{E}(X_{t+1} \mid Y_{t+1}) \mid X_t = x_t)$$
$$= \mathbb{E}((1 - \rho^2) \mid X_t = x_t) + \mathbb{V}ar(\rho Y_{t+1} \mid X_t = x_t)$$
$$= 1 - \rho^2 + \rho^2(1 - \rho^2) = 1 - \rho^4$$

$\square$

[RC] Example 7.1.

$$X_{t+1} \mid X_t = x_t \sim N(\rho^2 x_t, 1 - \rho^4)$$

Recursion shows that

$$X_{t+1} \mid X_0 = x_0 \sim N(\rho^{2t} x_0, 1 - \rho^{4t}),$$

which does indeed converge to $N(0, 1)$ as t goes to infinity.

As illustrated by the example, the sequence

$$(X_t, Y_t), t = 1, \ldots, T,$$

produced by a Gibbs sampler converges to the joint distribution f and, as a consequence, both sequences $(X_t)_t$ and $(Y_t)_t$ converge to their respective marginal distributions.

Example 7.2. Considering the pair of distributions

$$X|\theta \sim \mathcal{B}\text{in}(n, \theta), \quad \theta \sim \mathcal{B}e(a, b),$$

leads to the joint distribution

$$f(x, \theta) = \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{x+a-1} (1-\theta)^{n-x+b-1}.$$

The corresponding conditional distribution of $X|\theta$ is given above, while $\theta|x \sim \mathcal{B}e(x+a, n-x+b)$. The associated Gibbs sampler can be implemented as

```
> Nsim=5000                        #initial values
> n=15
> a=3
> b=7
> X=T=array(0,dim=c(Nsim,1))       #init arrays
> T[1]=rbeta(1,a,b)                #init chains
> X[1]=rbinom(1,n,T[1])
> for (i in 2:Nsim){               #sampling loop
+    X[i]=rbinom(1,n,T[i-1])
+    T[i]=rbeta(1,a+X[i],n-X[i]+b)
+    }
```

[RC] Example 7.3.

Consider the posterior distribution on (θ, σ^2) associated with the joint model

$$
\begin{aligned}
X_i &\sim N(\theta, \sigma^2), \quad i = 1, \ldots, n, \\
\theta &\sim N(\theta_0, \tau^2), \quad \sigma^2 \sim IG(a, b),
\end{aligned}
$$

where $IG(a, b)$ is the inverted gamma distribution (that is, the distribution of the inverse of a gamma variable), with density $b^a (1/x)^{a+1} e^{-b/x} / \Gamma(a)$ and with θ_0, τ^2, a, b specified. Writing $\mathbf{x} = (x_1, \ldots, x_n)$, the posterior distribution on (θ, σ^2) is given by

$$
\begin{aligned}
f(\theta, \sigma^2 \mid \mathbf{x}) \quad \propto \quad & \left[\frac{1}{(\sigma^2)^{n/2}} e^{-\sum_i (x_i - \theta)^2 / (2\sigma^2)} \right] \\
& \times \left[\frac{1}{\tau} e^{-(\theta - \theta_0)^2 / (2\tau^2)} \right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2} \right],
\end{aligned}
$$

from which we can get the full conditionals of θ and σ^2.

[RC] Example 7.3.

We have

$$
\pi(\theta \mid \mathbf{x}, \sigma^2) \quad \propto \quad e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)} e^{-(\theta - \theta_0)^2/(2\tau^2 \sigma^2)},
$$

$$
\pi(\sigma^2 \mid \mathbf{x}, \theta) \quad \propto \quad \left(\frac{1}{\sigma^2}\right)^{(n+2a+3)/2} e^{-\frac{1}{2\sigma^2}\left(\sum_i (x_i - \theta)^2 + (\theta - \theta_0)^2/\tau^2 + 2/b\right)}.
$$

These densities correspond to

$$
\theta \mid \mathbf{x}, \sigma^2 \quad \sim \quad N\left(\frac{\sigma^2}{\sigma^2 + n\tau^2}\theta_0 + \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{x}, \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2}\right)
$$

$$
\sigma^2 \mid \mathbf{x}, \theta \quad \sim \quad IG\left(\frac{n}{2} + a, \frac{1}{2}\sum_i (x_i - \theta)^2 + b\right),
$$

where $\bar{x}$ is the empirical average of the observations, as the full conditional distributions to be used in a Gibbs sampler.

[RC] Example 7.3. A study on metabolism in 15-year-old females yielded the following data, denoted by `x`,

```
> x=c(91,504,557,609,693,727,764,803,857,929,970,1043,
+ 1089,1195,1384,1713)
```

corresponding to their energy intake, measured in megajoules, over a 24 hour period (also available in the dataset `Energy`). Using the normal model above, with θ corresponding to the true mean energy intake, the Gibbs sampler can be implemented as

```
> xbar=mean(x)
> sh1=(n/2)+a
> sigma=theta=rep(0,Nsim)                        #init arrays
> sigma{1}=1/rgamma(1,shape=a,rate=b)        #init chains
> B=sigma2{1}/(sigma2{1}+n*tau2)
> theta{1}=rnorm(1,m=B*theta0+(1-B)*xbar,sd=sqrt(tau2*B))
> for (i in 2:Nsim){
+    B=sigma2[i-1]/(sigma2[i-1]+n*tau2)
+    theta[i]=rnorm(1,m=B*theta0+(1-B)*xbar,sd=sqrt(tau2*B))
+    ra1=(1/2)*(sum((x-theta[i])^2))+b
+    sigma2[i]=1/rgamma(1,shape=sh1,rate=ra1)
+    }
```

where theta0, tau2, a, and b are specified values. The posterior means of θ and σ^2 are 872.402 and 136,229.2, giving as an estimate of σ 369.092. Histograms of the posterior distributions of $\log(\theta)$ and $\log(\sigma)$ are given in Figure.
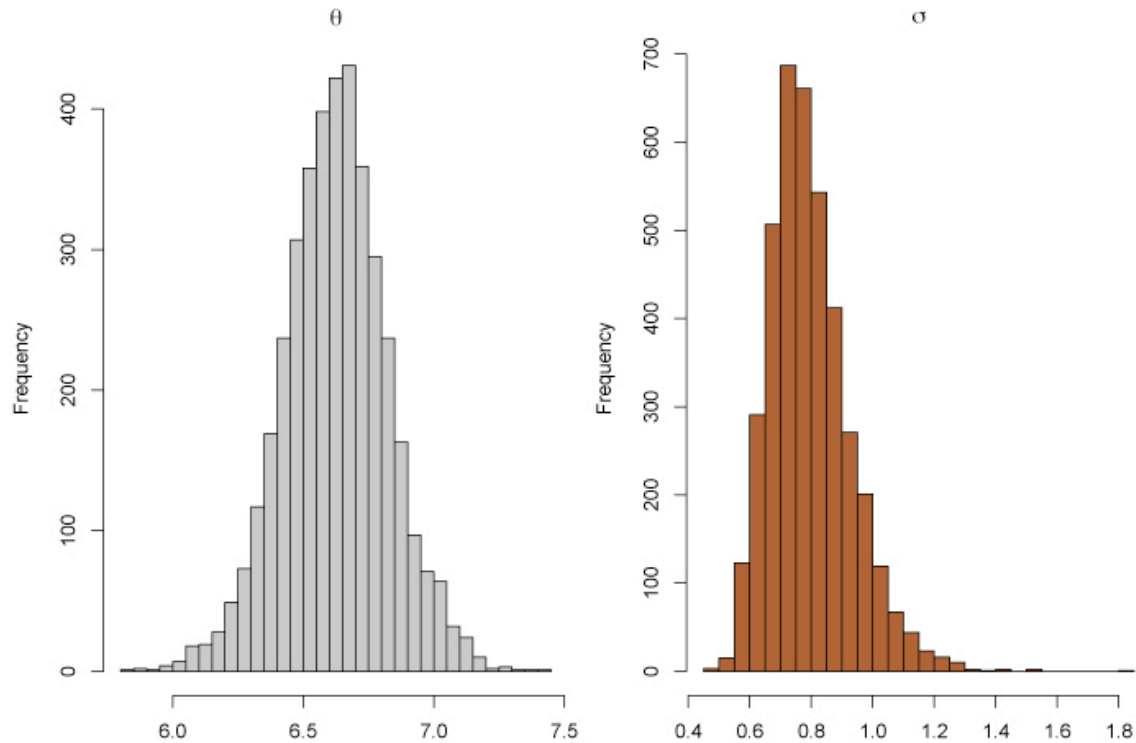
Fig. 7.2. Histograms of marginal posterior distributions of the log-mean and log-standard deviation from the Gibbs sampler of Example 7.3 based on 5000 iterations, with $a = b = 3$, $\tau_2 = 10$ and $\theta_0 = 5$. The 90% interval for $\log(\theta)$ is $(6.299, 6.960)$ and for $\log(\sigma)$ it is $(0.614, 1.029)$.

[RC] Example 7.3.

Writing $\mathbf{x} = (x_1, \ldots, x_n)$, the posterior distribution on (θ, σ^2) is given by

$$f(\theta, \sigma^2 \mid \mathbf{x}) \quad \propto \quad \left[\frac{1}{(\sigma^2)^{n/2}} e^{-\sum_i (x_i - \theta)^2 / (2\sigma^2)} \right]$$

$$\times \left[\frac{1}{\tau} e^{-(\theta - \theta_0)^2 / (2\tau^2)} \right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2} \right],$$

To find a full conditional (that is, the conditional distribution of one parameter conditional on all others), we merely need to pick out all of the terms in the joint distribution that involve that parameter. For example

$$f(\theta \mid \sigma^2, \mathbf{x}) \quad \propto \quad \left[\frac{1}{(\sigma^2)^{n/2}} e^{-\sum_i (x_i - \theta)^2 / (2\sigma^2)} \right] \times \left[\frac{1}{\tau} e^{-(\theta - \theta_0)^2 / (2\tau^2)} \right]$$

$$f(\sigma^2 \mid \theta, \mathbf{x}) \quad \propto \quad \left[\frac{1}{(\sigma^2)^{n/2}} e^{-\sum_i (x_i - \theta)^2 / (2\sigma^2)} \right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2} \right],$$

7.3 The multistage Gibbs sampler

There is a natural extension from the two-stage Gibbs sampler to the general multistage Gibbs sampler. Suppose that, for some $p > 1$, the random variable $\mathbf{X} \in \mathcal{X}$ can be written as $\mathbf{X} = (X_1, \ldots, X_p)$, where the X_i's are either unidimensional or multidimensional components. Moreover, suppose that we can simulate from the corresponding conditional densities $f_1, \ldots, f_p$, that is, we can simulate

$$X_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p \sim f_i(x_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$$

for $i = 1, 2, \ldots, p$. The associated *Gibbs sampling* algorithm (or *Gibbs sampler*) is given by the following transition from $X^{(t)}$ to $X^{(t+1)}$:

Algorithm 8 The Multistage Gibbs Sampler

At iteration $t = 1, 2, \ldots,$, given $\mathbf{x}^{(t)} = (x_1^{(t)}, \ldots, x_p^{(t)})$, generate

    1. $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \ldots, x_p^{(t)})$;

    2. $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})$;

    $\vdots$

    p. $X_p^{(t+1)} \sim f_p(x_p | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})$.

[RC] Example 7.5. A hierarchical specification for the normal model is the one-way random effects model. There are different ways to parameterize this model, but a possibility is as follows

$$
\begin{aligned}
X_{ij} &\sim N(\theta_i, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \\
\theta_i &\sim N(\mu, \tau^2), \quad i = 1, \ldots, k, \\
\mu &\sim N(\mu_0, \tau_\mu^2), \\
\sigma^2 &\sim IC(a_1, b_1), \quad \tau^2 \sim IC(a_2, b_2), \quad \sigma_\mu^2 \sim IC(a_3, b_3).
\end{aligned}
$$

[RC] Example 7.5.

Now, if we proceed as before and write down the joint distribution from this hierarchy, we can derive the set of full conditionals

$$\theta_i \quad \sim \quad N\left(\frac{\sigma^2}{\sigma^2 + n_i\tau^2}\mu + \frac{n_i\tau^2}{\sigma^2 + n_i\tau^2}\bar{X}_i, \frac{\sigma^2\tau^2}{\sigma^2 + n_i\tau^2}\right), \quad i = 1, \ldots, k,$$

$$\mu \quad \sim \quad N\left(\frac{\tau^2}{\tau^2 + k\sigma_\mu^2}\mu_0 + \frac{k\sigma_\mu^2}{\tau^2 + k\sigma_\mu^2}\bar{\theta}, \frac{\sigma_\mu^2\tau^2}{\tau^2 + k\sigma_\mu^2}\right),$$

$$\sigma_\mu^2 \quad \sim \quad IG\left(\frac{n}{2} + a_1, \frac{1}{2}\sum_{ij}(X_{ij} - \theta_i)^2 + b_1\right),$$

$$\tau_\mu^2 \quad \sim \quad IG\left(\frac{k}{2} + a_2, \frac{1}{2}\sum_{i}(\theta_i - \mu)^2 + b_2\right),$$

$$\sigma_\mu^2 \quad \sim \quad IG\left(\frac{1}{2} + a_3, \frac{1}{2}(\mu - \mu_0)^2 + b_3\right),$$

where $n = \sum_i n_i$ and $\bar{\theta} = \sum_i n_i\theta_i/n$.

[CG] Estimate the density.

"Gibbs sampling can be used to estimate the density itself by averaging the final conditional densities from each Gibbs sequence. The values $X_k = x_4$ yield a realization of $X_1, \ldots, X_m \sim f(x)$, the values $Y_k = y_k$ yield a realization of $Y_1, \ldots, Y_m \sim f(y)$. Moreover, the average of the conditional densities $f(xIY_k = y_k)$ will be a close approximation to $f(x)$, and we can estimate $f(x)$ with

$$\widehat{f}(x) = \frac{1}{m} \sum_{i=1}^{m} f(x \mid y_i), \qquad (1)$$

where $y_1, \ldots, y_m$ is the sequence of realized values of final Y observations from each Gibbs sequence. The theory behind the calculation in (1) is that the expected value of the conditional density is

$$\mathbb{E}[f(x \mid Y)] = \int f(x \mid y) f(y) dy = f(x),$$

a calculation mimicked by (1), since $y_1, \ldots, y_m$ approximate a sample from $f(y)$."

Exercises.

1.   [RC] Example 7.10, Exercises 7.11.

2. Construct Gibbs sampler for Ising model em $\Lambda = [0, N]^d \subset \mathbb{Z}^d$. Calculate explicitly conditional probabilities for Gibbs sampler.

3. Simulate the trajectory of the Gibbs sequence

$$(X_0, Y_0), (X_1, Y_1), \ldots, (X_m, Y_m), \ldots$$

with $f(x \mid y) = ye^{-yx}, x \in (0, \infty)$, $f(y \mid x) = xe^{-xy}, y \in (0, \infty)$. What can you say about recurrence of this sequence?

References.

[CG] Casella G. and George E.I. *Explaining the Gibbs Sampler.*

[RC] Cristian P. Robert and George Casella. *Introducing Monte Carlo Methods with R.* Series "Use R!". Springer