Gibbs Sampler

Componentwise sampling *directly* from the target density $\pi(x)$, $x \in \mathbb{R}^n$.

Define a transition kernel

$$K(x,y) = \prod_{i=1}^{n} \pi(y_i \mid y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_m),$$

and we set

$$r(x) = 0. \quad \text{(move every time)}$$

This transition kernel does not in general satisfy the detailed balance equation,

$$\pi(y)K(y,x) = \pi(x)K(x,y),$$

but it satisfies the balance equation,

$$\int \pi(y)K(y,x)dx = \int \pi(x)K(x,y)dx.$$

Proof in two dimensions

$$\int \pi(y)K(y,x)dx = \pi(y)\int K(y,x)dx.$$

We have

$$K(x,y) = \pi(y_1 \mid x_2)\pi(y_2 \mid y_1),$$

and therefore

$$K(y,x) = \pi(x_1 \mid y_2)\pi(x_2 \mid x_1).$$

Integrate with respect to x:

$$\begin{aligned}
\int K(y,x)dx &= \int \pi(x_1 \mid y_2)\pi(x_2 \mid x_1)dx_1 dx_2 \\
&= \int dx_1 \pi(x_1 \mid y_2) \underbrace{\int \pi(x_2 \mid x_1)dx_2}_{=1} \\
&= \int \pi(x_1 \mid y_2)dx_1 = 1.
\end{aligned}$$

Hence,

$$\int \pi(y)K(y,x)dx = \pi(y).$$

Right hand side:

$$\pi(x)K(x,y) = \pi(x)\pi(y_1 \mid x_2)\pi(y_2 \mid y_1),$$

so

$$\begin{aligned}
\int \pi(x)K(x,y)dx_1 &= \pi(y_1 \mid x_2)\pi(y_2 \mid y_1)\underbrace{\int \pi(x_1,x_2)dx_1}_{=\pi(x_2)} \\
&= \underbrace{\pi(y_1 \mid x_2)\pi(x_2)}_{=\pi(y_1,x_2)}(\pi(y_2 \mid y_1) \\
&= \pi(y_1,x_2)\pi(y_2 \mid y_1).
\end{aligned}$$

Integrating with respect to x_2, we obtain

$$\begin{aligned}
\int \pi(y_1, x_2)\pi(y_2 \mid y_1)dx_2 &= \pi(y_2 \mid y_1) \underbrace{\int \pi(y_1, x_2)dx_2}_{\pi(y_1)} \\
&= \pi(y_2 \mid y_1)\pi(y_1) \\
&= \pi(y_2, y_1) \\
&= \pi(y),
\end{aligned}$$

and the proof is complete.

Algorithm

Componentwise updating:

1. Initialize $x = x^1$ and set $k = 1$.
2. Update $x^k \to x^{k+1}$:
   - Draw x_1^{k+1} from $t \mapsto \pi(t, x_2^k, x_3^k, \cdots, x_n^k)$,
   - Draw x_2^{k+1} from $t \mapsto \pi(x_1^{k+1}, t, x_3^k, \cdots, x_n^k)$,
   - $\vdots$
   - Draw x_n^{k+1} from $t \mapsto \pi(x_1^{k+1} x_2^{k+1}, \cdots, x_{n-1}^{k+1}, t)$.
3. Increase $k \to k + 1$ and repeat from 2. until a desired sample size is reached.

Implementation: an example

Consider the following particular case:

Observation model

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(0, \sigma^2 I).$$

Whitening of the noise: Observe that

$$\mathbf{w} = \frac{1}{\sigma}\mathbf{e} \sim \mathcal{N}(0, I),$$

so the observation equation is equivalent to

$$\mathbf{b}' = A'\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(0, I),$$

where

$$A' = \frac{1}{\sigma}A, \quad \mathbf{b}' = \frac{1}{\sigma}\mathbf{b}.$$

Without loss of generality, we may assume therefore that $\sigma = 1$.

Likelihood density is

$$\pi(\mathbf{b} \mid \mathbf{x}) \propto \exp\left(-\frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2\right).$$

Prior: assume that we have an *a priori* inequality constraint

$$C\mathbf{x} \geq \mathbf{r},$$

the inequality understood componentwise.

Example:

$$v_j \leq x_j \leq u_j, \quad 1 \leq j \leq n,$$

can be written as

$$\underbrace{\begin{bmatrix} I \\ -I \end{bmatrix}}_{=C} \mathbf{x} \geq \underbrace{\begin{bmatrix} \mathbf{v} \\ -\mathbf{u} \end{bmatrix}}_{=\mathbf{r}}.$$

The prior can be written as

$$\pi_{\text{prior}}(\mathbf{x}) \propto \Theta(C\mathbf{x} - \mathbf{r}),$$

where Θ is the multivariate Heaviside step function.

Observe: the prior may be an improper density, i.e., it may be that the integral is not finite.

Write

$$\pi_{\text{post}}(\mathbf{x}) = \pi(\mathbf{x} \mid \mathbf{b}) \propto \pi(\mathbf{b} \mid \mathbf{x})\pi_{\text{prior}}(\mathbf{x}).$$

Updating of the jth component: we need the *conditional densities*

$$\pi(x_j \mid \underbrace{x_1, x_2, \ldots, x_{j-1}}_{\text{updated}}, \underbrace{x_{j+1}, \ldots, x_n}_{\text{old}}, \mathbf{b})$$

that is, we have to consider the mapping

$$x_j \mapsto \pi_{\text{post}}(\underbrace{x_1, x_2, \ldots, x_{j-1}}_{\text{updated}}, x_j, \underbrace{x_{j+1}, \ldots, x_n}_{\text{old}}),$$

where all the other components except for the jth one are fixed.

Likelihood:

Write

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

Then

$$A\mathbf{x} = \sum_{k=1}^{n} x_k \mathbf{a}_k.$$

Denote

$$A_j = \text{matrix } A \text{ with } j\text{th column eliminated},$$

$$\mathbf{a}_j = j\text{th column of } A,$$

$$\mathbf{x}_j = \text{vector } \mathbf{x} \text{ with the } j\text{th entry eliminated}.$$

We have

$$\begin{aligned} A\mathbf{x} - \mathbf{b} &= x_j \mathbf{a}_j + \sum_{k=1, k \neq j}^{n} x_k \mathbf{a}_k - \mathbf{b} \\ &= x_j \mathbf{a}_j + A_j \mathbf{x}_j - \mathbf{b} \\ &= x_j \mathbf{a}_j - \mathbf{b}_j, \end{aligned}$$

where

$$\mathbf{b}_j = \mathbf{b} - A_j \mathbf{x}_j.$$

Now,

$$\begin{aligned}\|A\mathbf{x}-\mathbf{b}\|^2 &= \|x_j\mathbf{a}_j-\mathbf{b}_j\|^2\\ &= \|\mathbf{a}_j\|^2x_j^2-2x_j\mathbf{a}_j^{\mathrm{T}}\mathbf{b}_j+\|\mathbf{b}_j\|^2\\ &= =\left(\|\mathbf{a}_j\|x_j-\frac{\mathbf{a}_j^{\mathrm{T}}\mathbf{b}_j}{\|\mathbf{a}_j\|}\right)^2+\|\mathbf{b}_j\|^2-\frac{(\mathbf{a}_j^{\mathrm{T}}\mathbf{b}_j)^2}{\|\mathbf{a}_j\|^2}.\end{aligned}$$

Therefore, by denoting

$$t_j=\|\mathbf{a}_j\|x_j,\quad \bar{t}_j=\frac{\mathbf{a}_j^{\mathrm{T}}\mathbf{b}_j}{\|\mathbf{a}_j\|},$$

we have

$$\pi(x_j\mid\mathbf{x}_j,\mathbf{b})\propto\exp\left(-\frac{1}{2}(t_j-\bar{t}_j)^2\right).$$

Prior, i.e., the bounds for t_j, assuming that $\mathbf{x}_j$ is given:

Again, we write

$$C_j = \text{matrix } C \text{ with } j\text{th column eliminated},$$

$$\mathbf{c}_j = j\text{th column of } C,$$

so the bound constraints

$$C\mathbf{x} = x_j\mathbf{c}_j + C_j\mathbf{x}_j \geq \mathbf{r}$$

implies

$$x_j\mathbf{c}_j \geq \mathbf{r} - C_j\mathbf{x}_j,$$

and by scaling with $\|\mathbf{a}_j\|$, we have

$$t_j\mathbf{c}_j \geq \mathbf{q}, \quad \mathbf{q} = \|\mathbf{a}_j\|\big(\mathbf{r} - C_j\mathbf{x}_j\big). \tag{1}$$

Denote

$$\mathbf{c}_j = \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{Nj} \end{bmatrix} \in \mathbb{R}^N.$$

Make a permutation of the elements c_{ij} of $\mathbf{c}_j$ and $\mathbf{q}$ so that the C_{ij}s are in decreasing order. Assume that the ℓ first elements are positive,

$$c_{1j} \geq \cdots \geq c_{\ell j} > 0,$$

while the entries starting from the $k+1$, $k \geq \ell$ are negative,

$$0 > c_{k+1,j} \geq \cdots \geq c_{Nj}.$$

Writing the inequality (1) component by component and taking the signs into account, we obtain

$$c_{ij}t_j \geq q_i, \text{ or } t_j \geq \frac{q_i}{c_{ij}}, \quad 1 \leq i \leq \ell,$$

and

$$c_{ij}t_j \geq q_i, \text{ or } t_j \leq \frac{q_i}{c_{ij}}, \quad k+1 \leq i \leq N.$$

In addition, one should check that the inequalities corresponding to zero entries are valid, that is,

$$0 \geq r_i, \quad \ell + 1 \leq i \leq k.$$

This is a consistency check, and has no contribution to the sampling strategy.

We therefore have lower and upper bounds for t_j,

$$t_{j,\min} = \max_{1\le i\le \ell}\left(\frac{q_i}{c_{ij}}\right), \quad t_{j,\max} = \min_{k+1\le i\le n}\left(\frac{q_i}{c_{ij}}\right).$$

The conditional probability density of t_j is

$$\pi(t_j \mid \mathbf{x}_j, \mathbf{b}) \propto \exp\left(-\frac{1}{2}(t_j - \bar{t}_j)^2\right), \quad t_{j,\min} \le t_j \le t_{j,\max},$$

and the random draw has to be done from this density.

To do the draws properly, we have to consider three possibilities, each one treated below separately.

1. $\bar{t}_j > t_{j,\max}$. This means that we have to draw from the left tail of the Gaussian distribution. The maximum value of this tail is achieved at $t_j = t_{j,\max}$. We scale the density so that this maximum value equals one:

$$\tilde{\pi}(t_j) = \exp\left(-\frac{1}{2}(t_j - \bar{t}_j)^2 + p\right), \quad p = \frac{1}{2}(t_{j,\max} - \bar{t}_j)^2.$$

   We seek the effective interval where this density is bigger than a prescribed threshold value $\delta > 0$.

Write

$$\tilde{\pi}(t_j) = \delta,$$

take logarithm of both sides to obtain

$$\frac{1}{2}(t_j - \bar{t}_j)^2 - p = \log \frac{1}{\delta},$$

and solve for t_j, bearing in mind that $t_j < \bar{t}_j$,

$$t_j = t_* = \bar{t}_j - \left(2p + 2\log \frac{1}{\delta}\right)^{1/2}.$$

Hence, the effective draw interval is

$$\max(t_{j,\min}, t_*) \leq t_j \leq t_{j,\max}.$$

2. $\bar{t}_j < t_{j,\min}$. This time, we need to draw from the right tail of the distribution. The maximum is attained at $t_j = t_{j,\min}$, and the scaled density now is

$$\tilde{\pi}(t_j) = \exp\left(-\frac{1}{2}(t_j - \bar{t}_j)^2 + p\right), \quad p = \frac{1}{2}(t_{j,\min} - \bar{t}_j)^2.$$

Again, we seek the effective interval, that in this time is

$$t_{j,\min} \leq t_j \leq \min(t_*, t_{j,\max}),$$

where

$$t_* = \bar{t}_j + \left(2p + 2\log\frac{1}{\delta}\right)^{1/2}.$$

3. $t_{j,\min} \leq \bar{t}_j \leq t_{j,\max}$, the maximum thus being within the interval. In this case, the scaled density is directly

$$\tilde{\pi}(t_j) = \exp\left(-\frac{1}{2}(t_j - \bar{t}_j)^2\right),$$

and solving the equation

$$\tilde{\pi}(t_j) = \delta$$

leads to the solutions

$$t_j = t_{*\pm} = \bar{t}_j \pm \left(2\log\frac{1}{\delta}\right)^{1/2},$$

so the active interval for t_j in this case is

$$\max(t_{j,\min}, t_{*,-}) \leq t_j \leq \min(t_{j,\max}, t_{*,+}).$$

Assume now that we have updated the interval to be the effective interval.