Slice samplers

(A very brief introduction)

The basic idea

“To sample from a distribution, simply sample uniformly from the region under the density function and consider only the horizontal coordinates.”

One way to do this is to

- introduce latent (auxiliary) variables,
- then use Gibbs sampling on the area beneath the density.
Suppose we wish to sample from $f(x)$ where $x \in \mathcal{X} \subseteq \mathbb{R}$.

To do this we sample uniformly from the 2-dimensional region under $f(x)$ or $g(x) = cf(x)$. This is implemented as follows:

- introduce a latent variable $y$ with $y|x \sim \mathcal{U}(0, g(x))$;

- this defines a uniform distribution for $x$ and $y$ over the region $\{(x, y) : 0 \leq y \leq g(x)\}$ with density

$$f(x, y) = f(y|x)f(x)$$

$$= \begin{cases} 
\frac{1}{g(x)}f(x) & \text{if } 0 \leq y \leq g(x), \\
0 & \text{otherwise} 
\end{cases}$$

$$= \begin{cases} 
\frac{1}{c} & \text{if } 0 \leq y \leq g(x), \\
0 & \text{otherwise}; 
\end{cases}$$
the conditional distribution for $x \mid y$ has density

$$f(x \mid y) \propto f(x, y)$$

$$\propto \begin{cases} \frac{1}{c} & \text{if } 0 \leq y \leq g(x), \\ 0 & \text{otherwise}, \end{cases}$$

that is,

$$x \mid y \sim \mathcal{U}(S(y)),$$

where $S(y) = \{x : g(x) \geq y\}$.

Here $S(y)$ is the union of intervals that constitute the slice through the density defined by $y$. 
To obtain a sample for $x$, we first sample $(x_i, y_i)$ from $f(x, y)$, then ignore the $y_i$’s.

The structure of this model leads us (naturally) to simulation using Gibbs sampling.

The $i$th iteration of the algorithm is:

- simulate $y_i \sim f(y|x_{i-1}) = U(0, g(x_{i-1}))$,

- simulate $x_i \sim f(x|y_i) = U(S(y_i))$, where $S(y) = \{x : g(x) \geq y\}$.

A key aspect to slice sampling is that only uniform random variates need be simulated.

N.B. Determining the slice $S(y)$ may be tricky!
A simple example

Standard normal slice sampler

Suppose $x \sim \mathcal{N}(0, 1)$, so

$$f(x) \propto g(x) = \exp(-x^2/2),$$

then the slice through the density is

$$S(y) = \left\{ x : -\sqrt{-2 \log(y)} \leq x \leq \sqrt{-2 \log(y)} \right\}.$$

Therefore, the conditional distribution of the latent variable $y$ is

$$y|x \sim \mathcal{U}\left(0, e^{-x^2/2}\right),$$

and the distribution of $x$ conditional on $y$ is

$$x|y \sim \mathcal{U}\left(-\sqrt{-2 \log(y)}, \sqrt{-2 \log(y)}\right).$$

Simulation from these conditional distributions is trivial.
The figure below shows the first 5 iterations of the slice sampler for the standard normal example.

The slice sampler carries out a Gibbs sampler on the area beneath the curve of the density \( g(x) \).
Plot of \((x_i, y_i)\)

Histogram of sampled values

Autocorrelations

Trace plot
Generalisations

More complex distributions

• Determining the slice $S(y)$ can be difficult when $f(x)$ has a complex structure.

• Possible solution: use the product slice sampler.

Suppose $\exists$ positive functions $f_i$ such that

$$f(x) = c \prod_{i=1}^{k} f_i(x).$$

Then augment the model with latent variables $y = (y_1, y_2, \ldots, y_k)$ where

$$y_i|x \sim \mathcal{U}(0, f_i(x)), \quad \text{independent}$$

so that (again) we have a uniform distribution under the functions $f_i$:

$$f(x, y) = c \prod_{i=1}^{k} \mathbb{I}(y_i \leq f_i(x)).$$

The conditional distribution for \( x|y \) has density

\[
f(x|y) \propto f(x, y) \\
\propto \begin{cases} 
1 & \text{if } 0 \leq y_i \leq f_i(x), \\
\ & \quad i = 1, 2, \ldots, k, \\
0 & \text{otherwise},
\end{cases}
\]

that is,

\[
x|y \sim \mathcal{U}(S(y)),
\]

where

\[
S(y) = \bigcap_{i=1}^{k} S_i(y_i)
\]

and

\[
S_i(y_i) = \{x : f_i(x) \geq y_i\}.
\]

Here \( S(y) \) is the intersection of the slices \( S_i(y_i) \) through the each of the functions \( f_i \) defined by the \( y_i \).
Implementation

The $j$th iteration of the algorithm is:

- **simulate (independently)**

\[ y_i^{(j)} \sim \mathcal{U}(0, f_i(x^{(j-1)})), \ i = 1, 2, \ldots, k \]

- **simulate** $x^{(j)} \sim \mathcal{U}(S(y^{(j)}))$, where

\[
S(y^{(j)}) = \bigcap_{i=1}^{k} \left\{ x : f_i(x) \geq y_i^{(j)} \right\}.
\]

**Multivariate case:** $x \in \mathcal{X} \subseteq \mathbb{R}^p$.

Sample uniformly from the $p + 1$-dimensional region under $g(x)$ using either the simple slice sampler or the product slice sampler ($x \rightarrow x$).
Random sample $x = (x_1, x_2, \ldots, x_n)$ from $f(x|\theta)$, prior density $\pi(\theta)$.

From Bayes’ theorem, the posterior density is

$$\pi(\theta|x) \propto \pi(\theta) \prod_{i=1}^{n} f(x_i|\theta)$$

cf.

$$f(x) \propto \prod_{i=1}^{k} f_i(x)$$

- Conjugate updates
  $\rightarrow$ simple slice sampler.

- Non-conjugate updates
  $\rightarrow$ product slice sampler.
Suppose $X_i | \theta \sim \text{Exp}(\theta)$ and we take a random sample of size $n = 2$, i.e. $x = (x_1, x_2)$.

Our prior for $\theta$ is $\pi(\theta) = \text{const.}$ (i.e. an improper prior). From Bayes’ theorem, the posterior density for $\theta$ is

$$
\pi(\theta|x) \propto \theta e^{-\theta x_1} \times \theta e^{-\theta x_2}
\uparrow \quad \uparrow
f_1(\theta) \quad f_2(\theta)
$$

We can simulate from this posterior density using a product slice sampler with 2 latent variables $\underline{y} = (y_1, y_2)$.

The $j$th iteration of the algorithm is:

- simulate $y_1^{(j)} \sim U\left(0, f_1(\theta^{(j-1)})\right)$
- simulate $y_2^{(j)} \sim U\left(0, f_2(\theta^{(j-1)})\right)$
- simulate $\theta^{(j)} \sim U\left(S(\underline{y}^{(j)})\right)$,

where $S(\underline{y}^{(j)}) = \left\{\theta : f_1(\theta) \geq y_1^{(j)}, f_2(\theta) \geq y_2^{(j)}\right\}$. 

\( y^{(j-1)} \)

\( S_1(y_1^{(j)}) \)

\( S_2(y_2^{(j)}) \)

\( s(y^{(j)}) \)

\( f_1(\theta) \)

\( f_2(\theta) \)
Why use the slice sampler?

✔ Almost automatic method to simulate from new densities – just simulate uniforms.

✔ Applies to most distributions.

✔ Easier to implement than a Gibbs sampler – no need to devise methods to simulate from non-standard distributions.

✔ Can be more efficient than Metropolis-Hastings algorithms – these also need the specification of a proposal.

✗ Determination of $S(\cdot)$ can be tricky.

✗ Some models require lots of latent variables.
For more details, see, for example,

Damien, Wakefield and Walker (1999) Gibbs sampling for non-conjugate and hierarchical models by using auxiliary variables. JRSSB.

