Slice Sampling

Slice Sampling

What is Slice Sampling?

A "black-box" auxiliary variable Markov Chain Monte Carlo (MCMC) method that avoids the need to tune hyperparameters. Introduced by Neal (2003).

The idea of slice sampling.
Suppose we wish to sample from a density for a variable, x, taking values in some subset of R^n . We can do this by sampling uniformly from the (n + 1)-dimensional region that lies under the plot of the density function.

Joint and Marginal Distributions

This idea can be formalized by introducing an auxiliary real variable, y, and defining a joint distribution over x and y that is uniform over the region $U = \{(x, y) : 0 < y < f(x)\}$ below the curve or surface defined by $f(x)$. That is, the joint density for (x, y) is

$$p(x, y) = \frac{1}{Z}\mathbf{1}_{\{0<y<f(x)\}}$$

where $Z = \int f(x)dx$. The marginal density for x is then

$$p(x) = \int_0^{f(x)} \frac{1}{Z} dy = \frac{f(x)}{Z}$$

which is the desired distribution. Thus, if we can sample from the joint distribution $p(x, y)$, we can obtain samples from the marginal distribution $p(x)$.

The Slice Sampling Algorithm

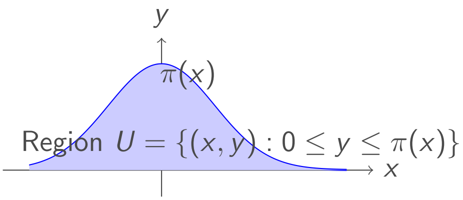
Step 1: Vertical Slice

- Given current position x
- Sample height $y \sim \text{Uniform}(0, \pi(x))$
- Defines horizontal "slice" at height y

Step 2: Horizontal Slice

- Sample new x uniformly from slice $S = \{x : \pi(x) \geq y\}$

The Challenge: In practice, sampling $S = \{x : \pi(x) \geq y\}$ can be difficult!



Key Insight: By alternating between sampling $y|x$ and $x|y$, we create a Markov chain that explores the space under $\pi(x)$ uniformly, with marginal distribution for x being exactly $\pi(x)$

How to sample from S

The Stepping Out Procedure

1. **Create initial interval:**
2. $L = x_t - w \cdot U$, $R = L + w$, where $U \sim \text{Uniform}(0, 1)$
3. **Step out left:**
4. While $\pi(L) \geq y$: $L = L - w$
5. **Step out right:**
6. While $\pi(R) \geq y$: $R = R + w$

The Shrinking Procedure

1. **Sample and shrink:**
2. Loop: $x' \sim \text{Uniform}(L, R)$
3. If $\pi(x') \geq y$: accept $x_{t+1} = x'$
4. Else: shrink $[L, R]$ by setting $L = x'$ or $R = x'$

Adaptive Nature: The algorithm automatically adapts to the local scale of $\pi(x)$. Wide regions are explored with large steps, narrow regions with small steps. Alternative procedures exist for sampling from S e.g., doubling.

Reversibility and Detailed Balance

The geometric intuition: Even though the stepping-out starts from different locations, both procedures ultimately sample uniformly from the same slice S, making the transitions symmetric.

The conditional transition probability $T(x_0 \to x_1|y)$ equals:
$P(\text{sample } y|\text{at } x_0) \times P(\text{sample } x_1|y, \text{starting from } x_0) = [1/\pi(x_0)] \times [1/|S_y|]$
Similarly, the conditional transition probability $T(x_1 \to x_0|y)$ equals:
$[1/\pi(x_1)] \times [1/|S_y|]$
For detailed balance, we need to show that:

$$\pi(x_0) \cdot T(x_0 \to x_1) = \pi(x_1) \cdot T(x_1 \to x_0)$$

The full kernel is obtained by integrating over all possible y:

$$T(x_0 \to x_1) = \int_0^{\min(\pi(x_0), \pi(x_1))} \frac{1}{\pi(x_0)} \cdot \frac{1}{|S_y|} dy$$

Convergence - Continued

Irreducibility

No matter where we start, there's always a positive probability of sampling a very small y value, which creates a very large slice that can connect distant parts of the state space, i.e. the slice is allmost the entire support of $\pi(x)$.

Aperiodicity

$P(x \to x) > 0$ (can stay at current state) $\Rightarrow$ period $= 1$

Ergodic Theorem

Since the chain is detailed balanced, irreducible and aperiodic, it has convergence with stationary distribution $\pi(x)$:

Various topics

- How to choose initial width w?
- Extensions to Multivariate Slice Sampling. Coordinate-wise: Apply one-dimensional slice sampling to each x_i in turn. (Gibbs sampling)
- Elliptical Slice Sampling for Gaussian priors (Murray et al., 2010)
- $\pi(x) \propto N(x|0, \sigma^2)L(x)$
- sample from prior and construct ellipse where the density on the ellipse is proportional to the likelihood