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 distribution for a variable, x, taking values in some subset of R^n , whose density is proportional to some function $f(x)$. We can do this by sampling uniformly from the (n + 1)-dimensional region that lies under the plot of $f(x)$.

Introduction

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

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

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 $p(x)$.

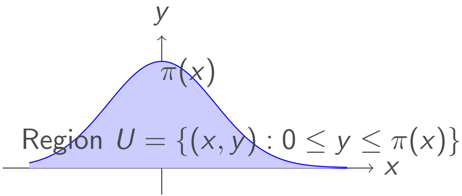
Intuition Behind Slice Sampling

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
- $S = \{x : \pi(x) \geq y\}$
- Gives new sample from $\pi(x)$



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 $\pi(x)$

The Slice Sampling Algorithm

Basic Algorithm

Given: Current state x_t, target distribution $\pi(x)$

Algorithm

1. **Sample auxiliary variable:** Draw $y \sim \text{Uniform}(0, \pi(x_t))$
2. **Find the slice:** Identify $S = \{x : \pi(x) \geq y\}$
3. **Sample from the slice:** Draw $x_{t+1} \sim \text{Uniform}(S)$

The Challenge: Finding and Sampling from S

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

Key Idea: Leave Distribution Invariant

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.

Why w is Not a Critical Tuning Parameter

The Width Parameter w: Efficiency vs. Correctness

Yes, w IS technically a tuning parameter, BUT...

Traditional MCMC (e.g., RW-Metropolis)

- **Poor tuning $\rightarrow$ Poor mixing**
- σ too small $\rightarrow$ Tiny steps, stuck
- σ too large $\rightarrow$ High rejection
- Can take exponentially long
- Affects correctness in finite time

Slice Sampling with w

- **Poor w $\rightarrow$ More computation**
- w too small $\rightarrow$ Many step-outs
- w too large $\rightarrow$ More shrinking
- Always finds correct slice
- Only affects efficiency

Why w is Not a Critical Tuning Parameter

Why w is Robust: The Self-Correcting Mechanism

Case 1: $w \ll$ typical slice width

- Initial $[L, R]$ doesn't cover slice
- Stepping-out expands it
- ✓ Still finds correct slice!

Case 2: $w \gg$ typical slice width

- Initial $[L, R]$ too wide
- Shrinking contracts it
- ✓ Still samples correctly!

Mathematical Guarantee

For ANY $w > 0$:

- The stationary distribution is ALWAYS $\pi(x)$
- Detailed balance is satisfied
- Convergence is guaranteed

Key Distinction

Why Slice Sampling Converges

Formal Convergence Properties

1. Detailed Balance

Let $T(x'|x)$ be the transition kernel. We need: $\pi(x) \cdot T(x'|x) = \pi(x') \cdot T(x|x')$

Proof sketch:

- Given x, sample $y \sim \text{Uniform}(0, \pi(x))$
- Probability density of moving from x to x':

$$T(x'|x) = \int_0^{\min(\pi(x),\pi(x'))} \frac{1}{\pi(x)} \cdot \frac{1}{|S_y|} dy$$

  where $|S_y|$ is the length of slice $\{z : \pi(z) \geq y\}$
- This is symmetric: $T(x'|x) = T(x|x') \Rightarrow$ detailed balance holds

Convergence - Continued

2. Irreducibility

For any x, x' where $\pi(x) > 0$ and $\pi(x') > 0$:

$$P(x \to x') \geq \int_0^{\min(\pi(x),\pi(x'))} \frac{1}{\pi(x)} \cdot P(x' \text{ sampled from } S_y)dy > 0$$

3. Aperiodicity

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

Ergodic Theorem

Since the chain is irreducible, aperiodic, with stationary distribution $\pi(x)$:

$$\lim_{n\to\infty} \|P(X_n \in \cdot | X_0 = x_0) - \pi(\cdot)\|_{TV} = 0$$

Convergence: Additional Mathematical Details

Detailed Balance - Complete Argument

Consider augmented state space (x, y) with invariant distribution:

$$\pi^*(x, y) = \frac{1}{Z} \cdot \mathbf{1}\{0 \leq y \leq \pi(x)\}$$

where $Z = \int \pi(x)dx$ is the normalization constant.

The Gibbs Sampler View

Slice sampling is a Gibbs sampler on the augmented space:

- **Step 1:** Sample $y|x \sim \text{Uniform}(0, \pi(x))$
- **Step 2:** Sample $x|y \sim \text{Uniform}(\{x : \pi(x) \geq y\})$

Each conditional distribution is correct:

$$p(y|x) = \frac{1}{\pi(x)} \cdot \mathbf{1}\{0 \leq y \leq \pi(x)\} \qquad (1)$$