Markov Chain Monte Carlo Using the Ratio-of-Uniforms Transformation

Luke Tierney
Department of Statistics & Actuarial Science
University of Iowa

Basic Ratio of Uniforms Method

Introduced by Kinderman and Monahan (1977).

- Suppose f is a (possibly unnormalized) density
- Suppose V, U are uniform on

$$\mathscr{A} = \{(v,u) : 0 < u < \sqrt{f(v/u)}\}$$

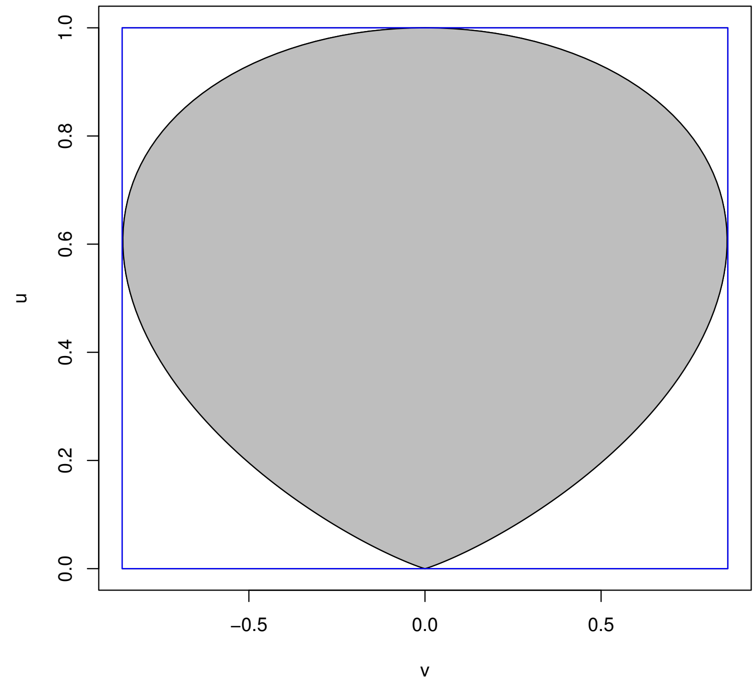
- Then $X = V/U$ has density (proportional to) $f(x)$

Example: Standard Normal Distribution

- Suppose

$$f(x) = \frac{1}{2\pi} e^{-x^2/2}$$

- Then $\mathscr{A}$ is bounded.
- Can use rejection sampling from an enclosing rectangle.



Some Generalizations

- If $f(x) = f_0(x) f_1(x)$ and V, U has density proportional to $f_0(v/u)$ on

$$\mathscr{A} = \{(v,u) : 0 < u < \sqrt{f_1(v/u)}\}$$

  then $X = V/U$ has density proportional to f

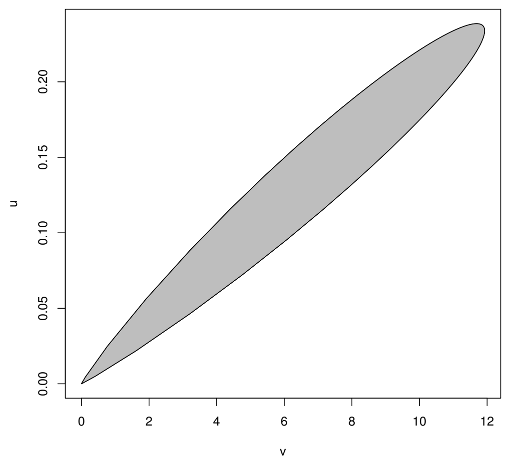
- If V, U are uniform on

$$\mathscr{A} = \{(v,u) : 0 < u < \sqrt{f(v/u + \mu)}\}$$

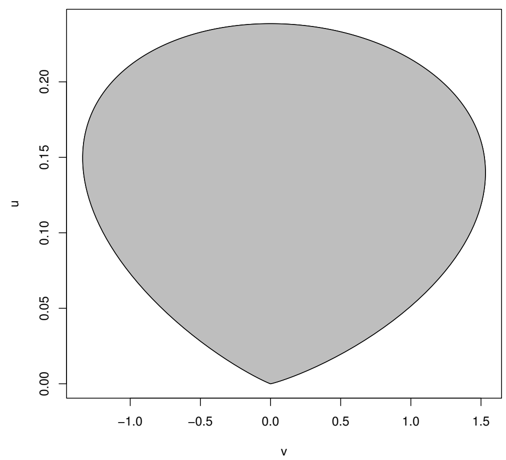
  then $X = V/U + \mu$ has density proportional to f

Some Generalizations (cont.)

- Choosing μ as the mode of f often works well.
- For a Gamma($\alpha = 50$) density $f(x) \propto x^{49} e^{-x}$:



$\mu = 0$



$\mu = \alpha - 1 = 49$

Higher Dimensions

Stefănescu and Văduva (1987); Wakefield, Gelfand and Smith (1991)

If V, U are uniform on

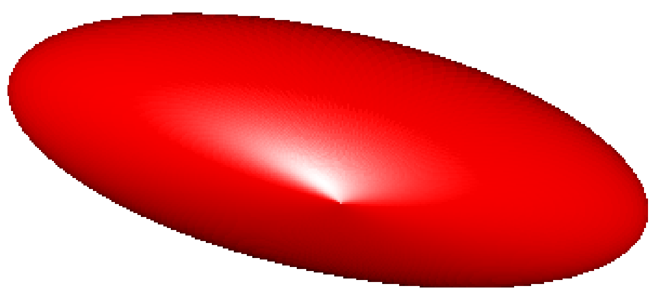
$$\mathscr{A} = \{(v,u) : v \in \mathbb{R}^d, 0 < u < \sqrt[d+1]{f(v/u+\mu)}\}$$

then $X = V/U + \mu$ has density proportional to f

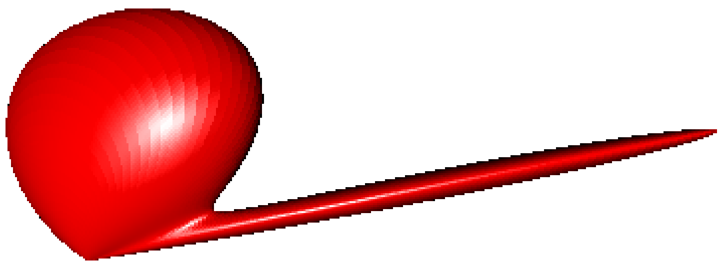
- Can rejection sample from hyper-rectangle
  - Usually not practical in higher dimensions
- Alternative: Sample $\mathscr{A}$ by MCMC

Higher Dimensions (cont.)

Some Regions for $d = 2$:



Bivariate Normal, $\rho = 0.73$



Variance Components

Some Properties of the Region $\mathscr{A}$

- $\mathscr{A}$ is bounded if $f(x)$ and $\|x\|^{d+1} f(x)$ are bounded.
- $\mathscr{A}$ is convex if $f(x)$ is log concave.
- If $f(x) = f_0(x) f_1(x)$ and

$$\mathscr{A} = \{(v,u) : 0 < u < \sqrt[d+1]{f_1(v/u+\mu)}\}$$

  and f_1 is log concave, then
  - $\mathscr{A}$ is convex
  - $\mathscr{A}$ is bounded if and only if $\int f_1(x)dx < \infty$.

Leydold (2001); Hörmann, Leydold, Derflinger (2004)

Auxiliary Variable Approach

Suppose X has density proportional to $f(x)$ on $\mathbb{R}^d$

- Let $U|X=x$ have density $\frac{rd+1}{f(x)}u^{rd}$ for $0<u<\sqrt[rd+1]{f(x)}$ where $r>-1/d$. Then X,U have joint density

$$f_{X,U}(x,u) \propto f(x)u^{rd}/f(x) = u^{rd}$$

for $0<u<\sqrt[rd+1]{f(x)}$.

- Let $V=XU^r$. Then V,U have joint density

$$f_{V,U}(v,u) \propto u^{rd}|J(v,u)| = u^{rd}u^{-rd} = 1$$

for $0<u<\sqrt[rd+1]{f(v/u^r)}$.

Auxiliary Variable Approach (cont.)

Generalized RU sampler: $X = V/U^r$ with V,U uniform on

$$\mathscr{A}_r = \{(u,v) : 0 < u < \sqrt[rd+1]{f(v/u^r)}\}$$

- $r = 1$: standard RU sampler
- $r = 0$: slice sampler

A Simple Gibbs Sampler Approach

Alternate sampling $V|U=u$ and $U|V=v$

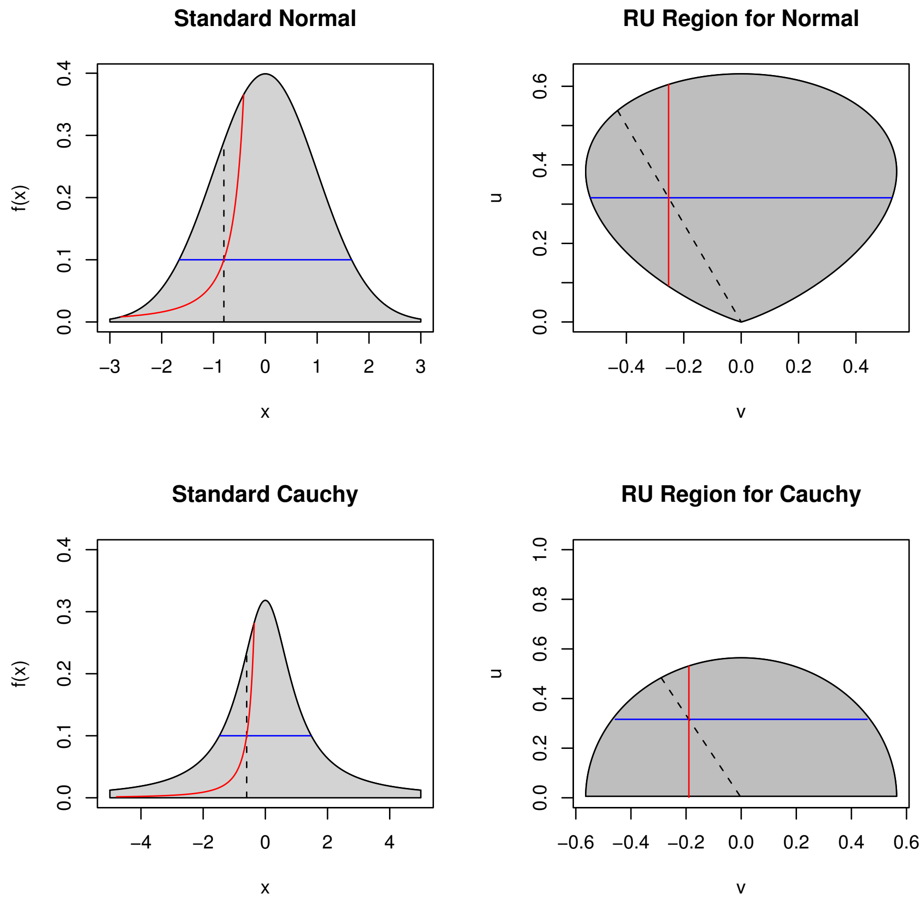
- Sampling $V|U=u$ is equivalent to

$$\begin{aligned} \text{sample} \quad & X|U=u \\ \text{set} \quad & V = Xu^{rd} \end{aligned}$$

- Sampling $V|U=u$ or $X|U=u$
  - may not be practically feasible for large d
  - may be theoretically useful to consider

Call this the simple RU sampler.

Slice and RU Sampler Comparison



Other MCMC Approaches

Updating V and U separately:

- Updating V:
  - by any method that leaves $V|U = u$ or $X|U = u$ invariant
  - one or more coordinates at a time, or all at once
  - Neal (2003) describes many approaches
- Updating U:
  - for log convex f can use adaptive rejection
  - again, methods from Neal (2003) can be used

Other MCMC Approaches (cont.)

Marginal updating of X:

- update to X_n by any method that is invariant for f
- generate $U_0|X = x_n$ as in the auxiliary variable approach
- compute $V = X_n U_0$
- choose U_1 by a method that leaves $U|V = v$ invariant
- compute $X_{n+1} = V/U_1 = X_n U_0/U_1$

This is a random rescaling of X_n that leaves f invariant

Other MCMC Approaches (cont.)

Joint sampling of V,U.

- Hit and run (random direction) methods
  - uniform conditionals for log concave f
  - other univariate uniform samplers when $\mathscr{A}$ is not convex
- Random walk Metropolis-Hastings methods
  - For symmetric, e.g. Gaussian, proposal acceptance probability is zero or one.
- Shape/scaling of $\mathscr{A}$ is important.

Other MCMC Approaches (cont.)

Suppose g is an approximation to f with RU region $\mathscr{B}$, and $q(y|x)$ is a symmetric proposal kernel. Let

$$\widetilde{q}(y|x) = \lambda \frac{1}{|\mathscr{B}|} 1_{\mathscr{B}}(y) + (1-\lambda) q(y|x)$$

Then the Metropolis-Hastings acceptance probabilities are

$$\alpha(x,y) = \begin{cases} 1 & \text{if } x \in \mathscr{A} \cap \mathscr{B} \text{ or } y \in \mathscr{A} \cap \mathscr{B}^c \\ \left(1 + \frac{\lambda}{1-\lambda} \frac{1}{|\mathscr{B}| q(y|x)}\right)^{-1} & \text{if } x \in \mathscr{A} \cap \mathscr{B}^c \text{ and } y \in \mathscr{A} \cap \mathscr{B} \\ 0 & \text{otherwise} \end{cases}$$

Theoretical Properties

Uniform ergodicity: $\pi(x) \propto f(x)$, $\pi P = \pi$ and

$$\|P^n(x, dy) - \pi(dy)\|_{\text{TV}} \leq M\rho^n$$

for some $M < \infty$ and $\rho < 1$

- If f is log concave with mode at the origin, then the simple RU Gibbs sampler is uniformly ergodic.
- If f is log concave, then the hit-and-run sampler is uniformly ergodic.
- If f and $\|x\|^{rd+1} f(x)$ are bounded then Gaussian MH samplers are uniformly ergodic.

A Simple Example

Suppose $f(x)$ is a density on $\mathbb{R}^d$ with

$$f(x) \propto \left(\frac{2}{1+\|x\|^2}\right)^{d+1}$$

The RU region $\mathscr{A}$ is the sphere

$$\mathscr{A} = \{(v,u) : \|v\|^2 + (u-1)^2 \leq 1\}$$

Simple RU sampler, run of 10,000, $d = 10, 20, 50, 100, 200$:

- absolute lag one autocorrelations are around 0.03

A Simple Example (cont.)

- Gaussian random walk Metropolis-Hastings:
  - optimal proposal SD seems to be around $1.5/d^{0.9}$
  - absolute lag d autocorrelations are around 0.5
- Hit and run:
  - absolute lag d autocorrelations are around 0.35
- Coordinate-wise Gaussian random walk M-H:
  - optimal proposal SD seems to be around $2.4/\sqrt{d}$
  - absolute lag one autocorrelations are around 0.6
- Coordinate-wise Gibbs:
  - absolute lag one autocorrelations are around 0.35

Further Work and Open Questions

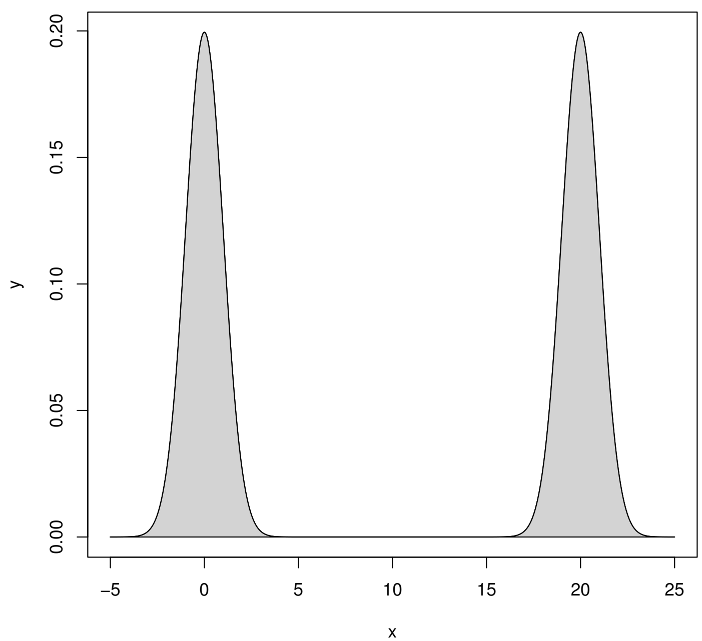
- Computational issues
  - numerical considerations
  - computational efficiency
  - efficient single coordinate updates
- Transformations and parameterization
  - location scale choices for f
  - scaling of U
  - nonlinear transformations, e.g. logs
  - opportunities for adaptive methods

Further Work and Open Questions (cont.)

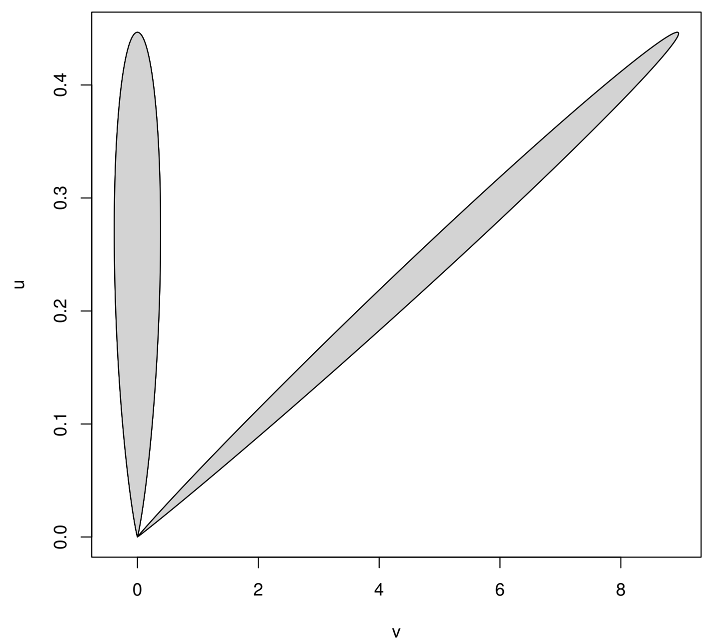
- Use with marginals and conditionals
  - sample joint conditionals for blocking
  - sample marginals if available (variance components)
- Multi-modality issues
  - $\mathscr{A}$ is connected at the origin
  - allows transitions between modes

A Multi-Modal Target Density

Two well-separated modes:



Mixture of Normals



RU Region

Further Work and Open Questions (cont.)

- Diagnostic uses
  - maybe useful for discovering additional modes
  - unbounded behavior may indicate improper f
- Theoretical considerations
  - explicit transition kernels and convergence rates
  - perfect sampling opportunities
  - hybrids to make other samplers uniformly ergodic

Conclusions

- The value of the ratio of uniforms transformation for random variate generation is well known
- It also has interesting uses in MCMC construction.
- Its simplicity leads to interesting theoretical properties.
- With good scaling there is promise for good practical results.
- More work and experience is needed to fully understand the practical and theoretical value.