Markov Chain Monte Carlo, Numerical Integration

(See Statistics)

Trevor Gallen

Agenda

- Numerical Integration: MCMC methods
- Estimating Markov Chains
- Estimating latent variables

Numerical Integration: Part II

- Quadrature for one to a few dimensions feasible for well-behaved distributions
- For many-dimensional integrals, we typically use Markov chain Monte Carlo
- There are many different methods
- I discuss a simple one (Gibbs Sampling) and a more complex one (Metropolis-Hastings)
- This is a prominent problem in Bayesian analysis

The Problem

- Say your data is summarized by a two-dimensional problem: height and weight $f(x_1, x_2)$
- You want a population, for $i \in (1, ..., n)$, (x_1^i, x_2^i)
- You have access to the *conditional* marginal distributions. That is, while $f(x_1, x_2)$ is ugly, $f(x_1|x_2)$ and $f(x_2|x_1)$ are easy to sample from.

Gibbs Sampling

1. Start with (x_1^0, x_2^0)
2. Sample $x_1^1 \sim f(x_1|x_2^0)$
3. Sample $x_2^1 \sim f(x_2|x_1^1)$
4. We have a LLN and CLT that states that:

$$\frac{1}{N}\sum_{i=1}^{N} g(x^i) \to \int g(x)f(x)dx$$

Gibbs Sampling: Example

- Assume a distribution:

$$x \sim \mathcal{N}(0, \Sigma) \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

- Then we can get the conditional marginal distributions:

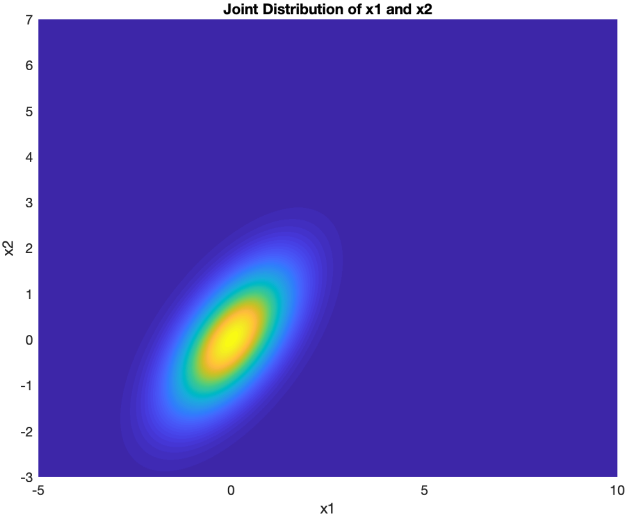
$$x_1 | x_2 \sim \mathcal{N}\left(\rho x_2, (1-\rho)^2\right)$$

$$x_2 | x_1 \sim \mathcal{N}\left(\rho x_1, (1-\rho)^2\right)$$

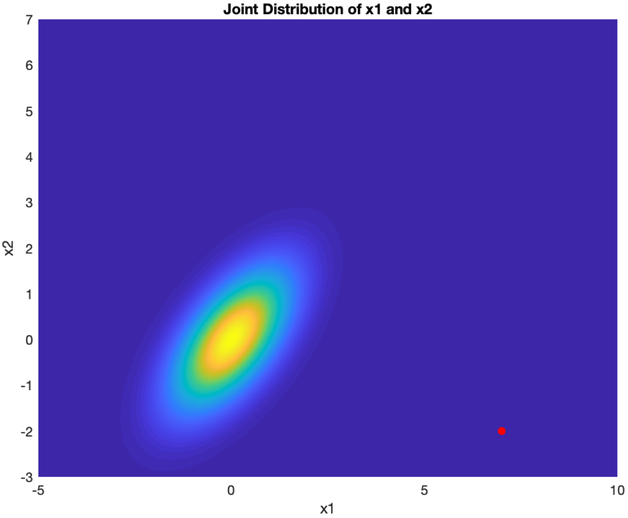
- Iteratively sample from these.

- See Gibbs.m

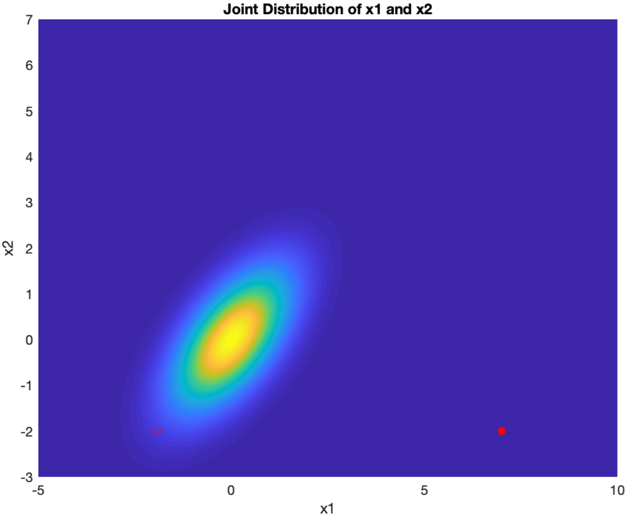
Gibbs Sampling: Example



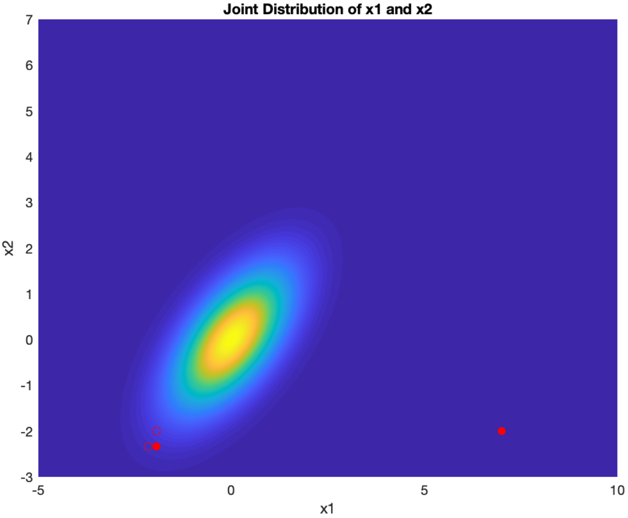
Gibbs Sampling: Example



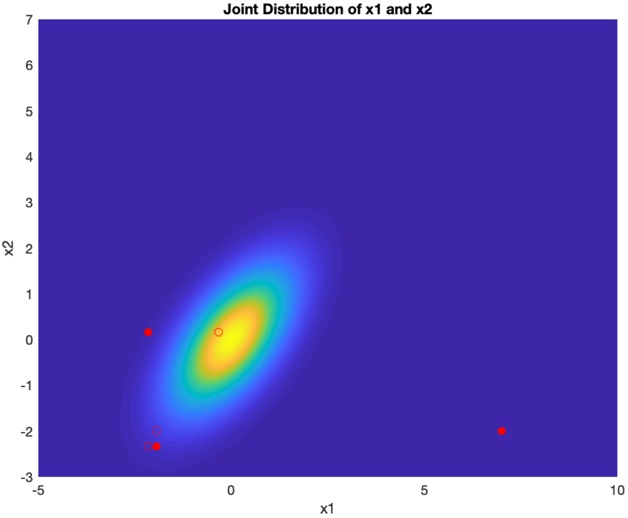
Gibbs Sampling: Example



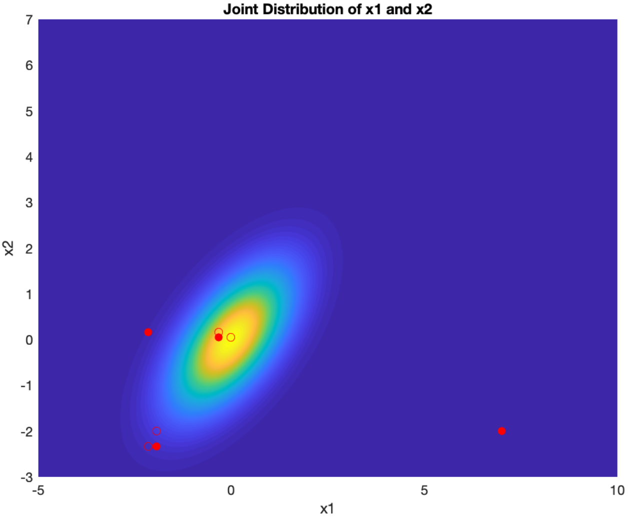
Gibbs Sampling: Example



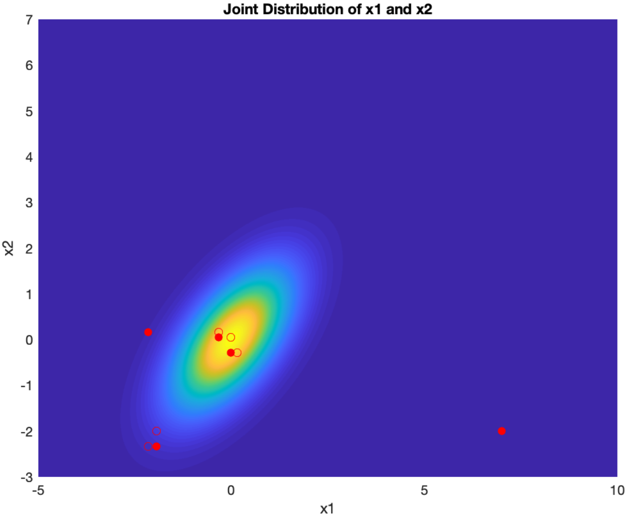
Gibbs Sampling: Example



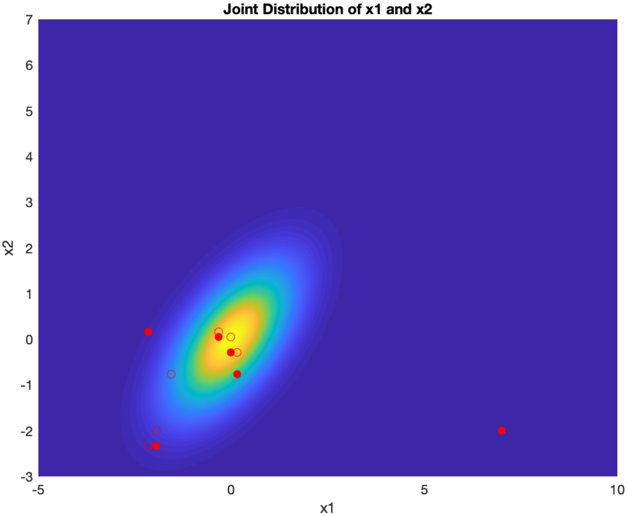
Gibbs Sampling: Example



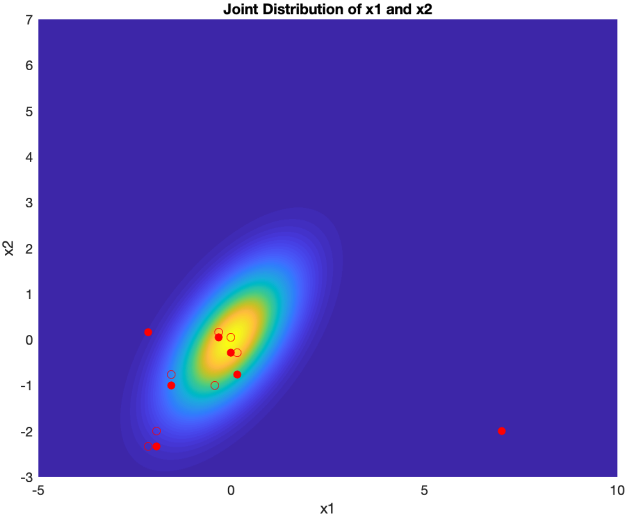
Gibbs Sampling: Example



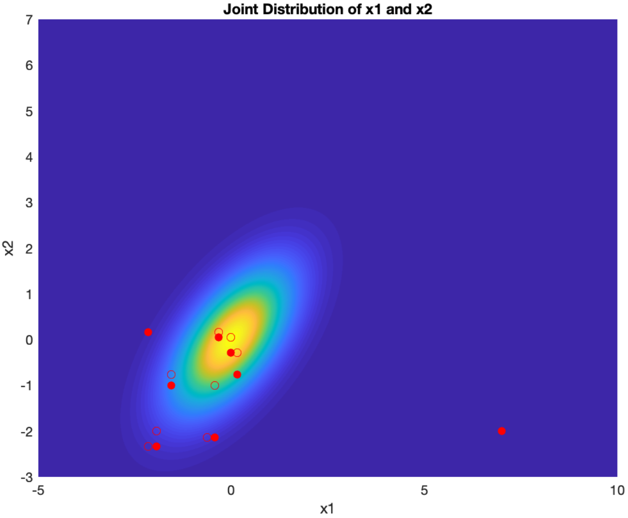
Gibbs Sampling: Example



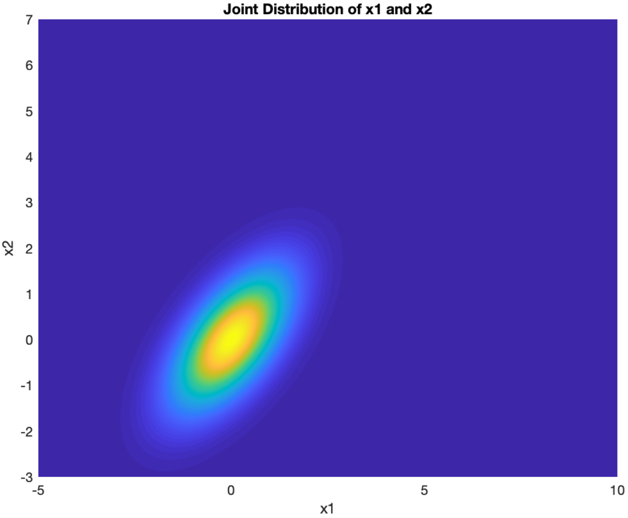
Gibbs Sampling: Example



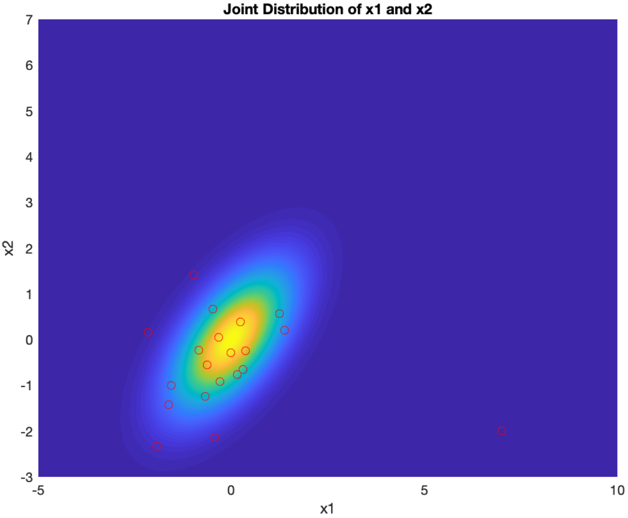
Gibbs Sampling: Example



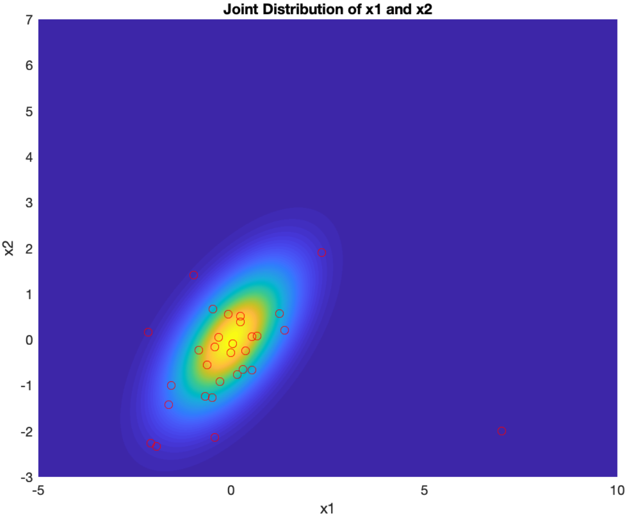
Gibbs Sampling: Example



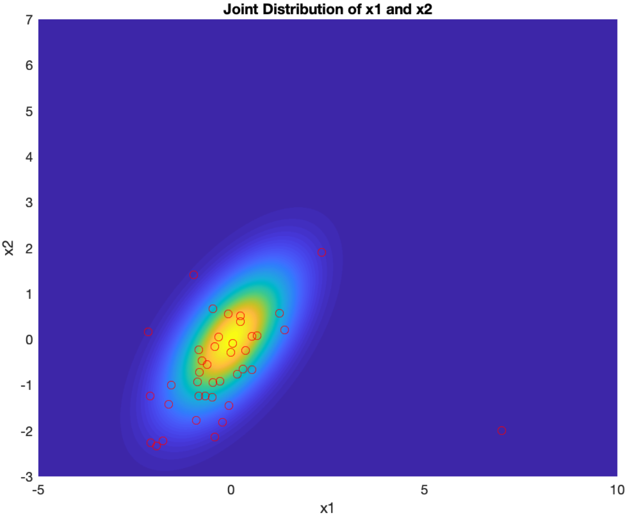
Gibbs Sampling: Example



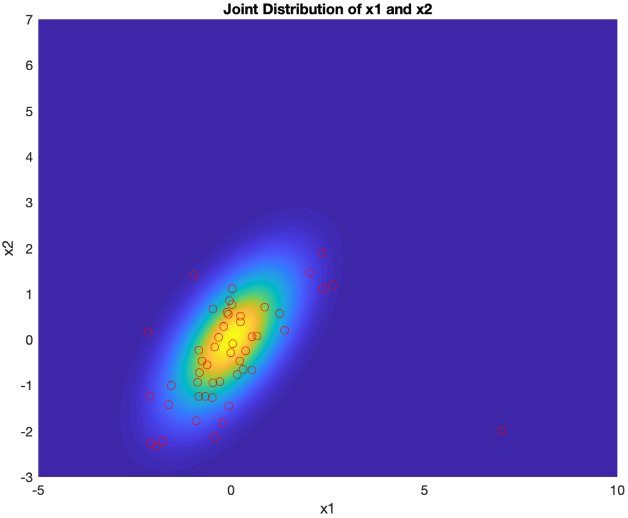
Gibbs Sampling: Example



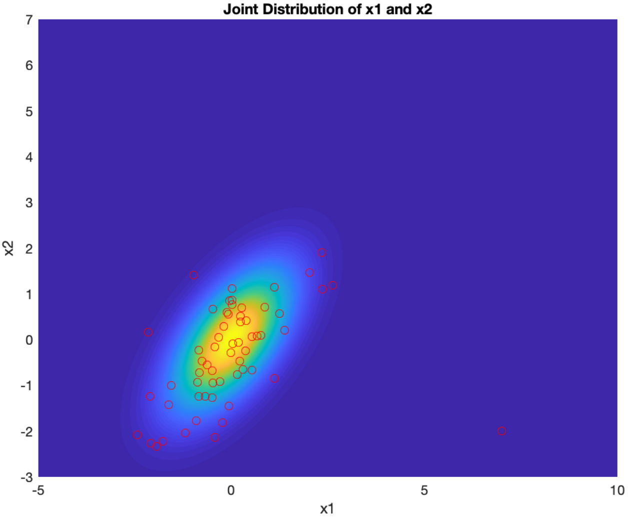
Gibbs Sampling: Example



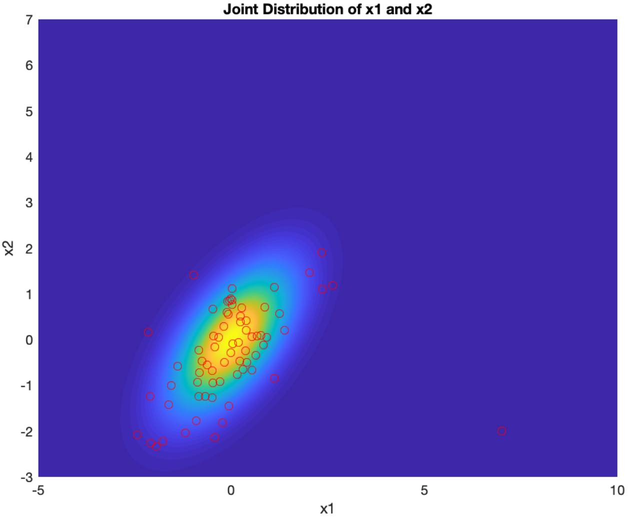
Gibbs Sampling: Example



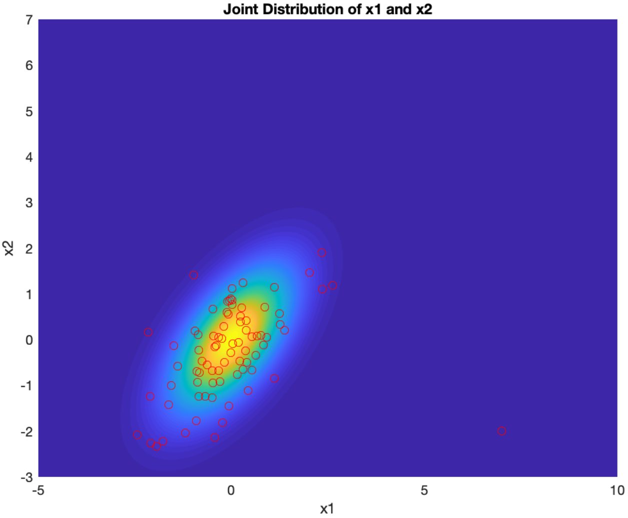
Gibbs Sampling: Example



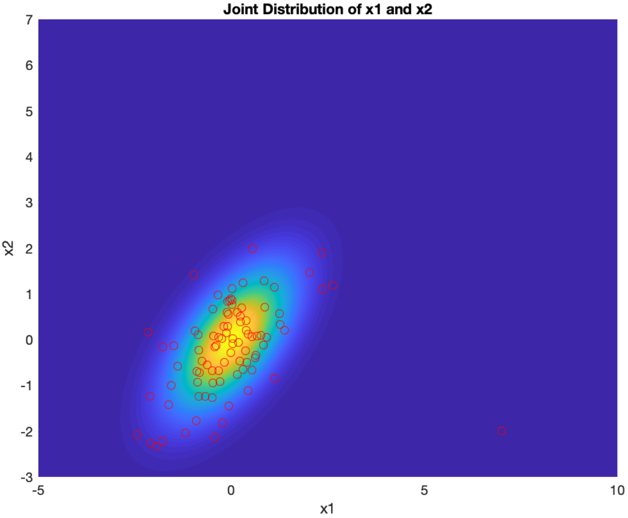
Gibbs Sampling: Example



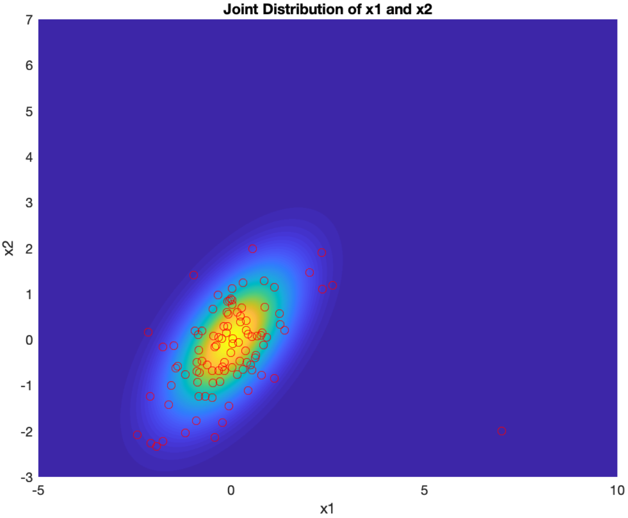
Gibbs Sampling: Example



Gibbs Sampling: Example



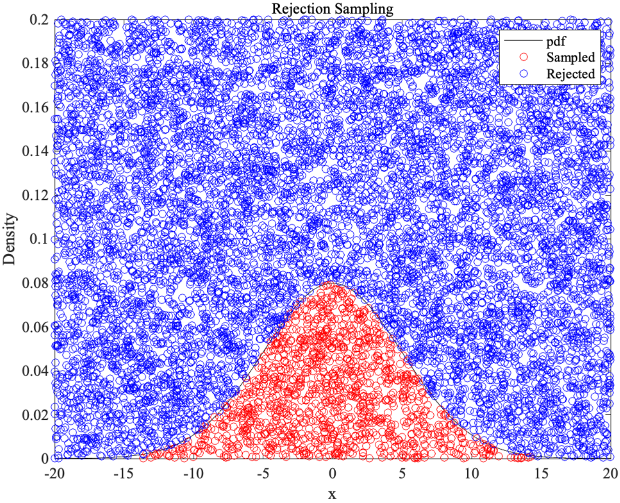
Gibbs Sampling: Example



Rejection Sampling

- Rejection sampling pretty easy way to draw from distributions if you can sample from uniform distribution
- Want to draw from pdf $f(x)$ (x can be vector)
- Draw a uniformly-distributed set of N $\hat{x}$'s and $\hat{y}$'s that span support of f and the domain of f, respectively
- Drop all the samples for which $\hat{y} > f(\hat{x}_i)$
- Feels like dark magic first time you see it (Buffon's needle)

Rejection Sampling-Example



Drop the blues, keep the reds, distributed normally with mean zero standard deviation 5

Metropolis Hastings-Intro

- Likelihoods coming from linearized models, as in macro, are pretty easy.
- Let's say you wanted to do Bayesian estimation on a structural model
- It's pretty hard! Filter, a well-behaved distribution through a maximization problem with kinks, cliffs, and curvature and you quickly have an 'intractable* posterior
- How can we sample from it?
- Metropolis Hastings Algorithm

Metropolis Hastings-Idea

- The core of MH is rejection sampling
- Start out with an x
- Randomly sample new $\hat{x}$
- If $f(\hat{x}) = 0.5f(x)$, want to move there only sometimes (spend more time in more likely regions)

Metropolis Hastings

What if we can only evaluate likelihood at a given point?

1. We start with some (multi-dimensional) value x^i and a proposal distribution $g(x|x^i)$

2. Grab a new sample from our proposal distribution:

$$x' \sim g(x|x^i)$$

3. Calculate acceptance probability:

$$pr(x^i, x') = \min\left\{1, \frac{f(x')}{f(x^i)} \frac{g(x^i|x')}{g(x', x^i)}\right\}$$

4. Accept the new value with probability $pr(x^i, x')$, otherwise, stay there.

5. This again converges in distribution to the true distribution.

A convenient proposal density

- If our proposal density is symmetric

$$g(x^i|x') = g(x', x^i)$$

- This is called random-walk Metropolis-Hastings
- Our acceptance probability is easy:

$$pr(x^i, x') = \min\left\{1, \frac{f(x')}{f(x^i)}\right\}$$

- Acceptance probability can be derived from wanting our Markov chain's stationary distribution to be equal to $f(x)$ conditional on switching probability

Metropolis-Hastings: Example

- Let's say our distribution is one-dimensional:

$$x \sim 0.5U(0,1) + 0.25U(-1,2) + 0.25U(0.5,0.75)$$

- Choose sampling distribution centered around current point:

$$g(x'|x) \sim \mathcal{N}(x, 0.1)$$

Sampling PDF

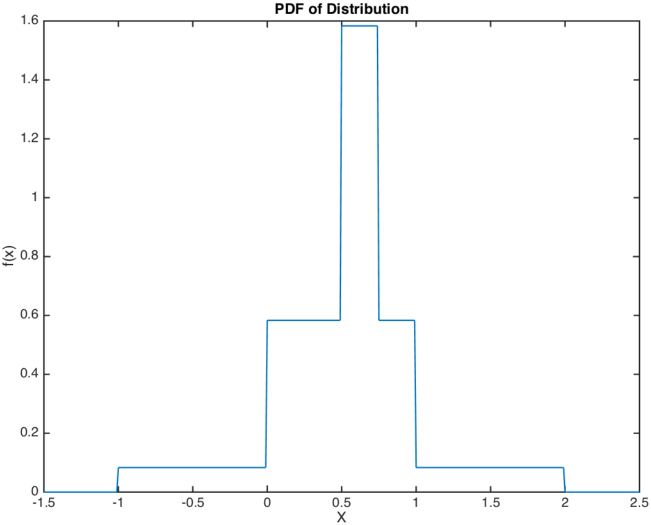
- Let's say our distribution is one-dimensional:

$$x \sim 0.5U(0,1) + 0.25U(-1,2) + 0.25U(0.5,0.75)$$

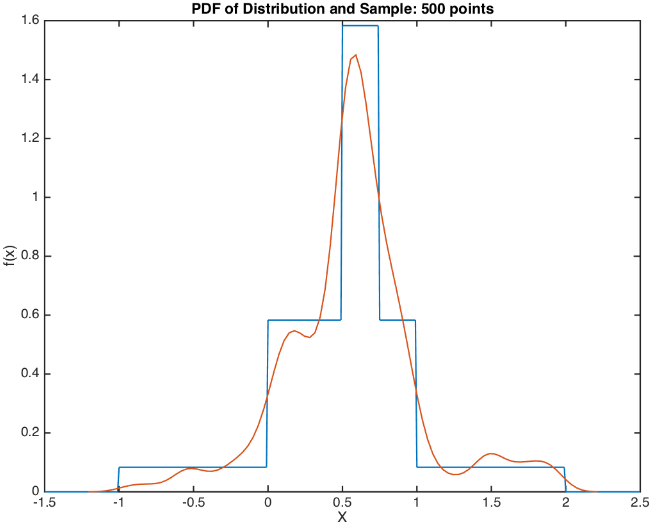
- Choose sampling distribution centered around current point:

$$g(x'|x) \sim \mathcal{N}(x, 0.1)$$

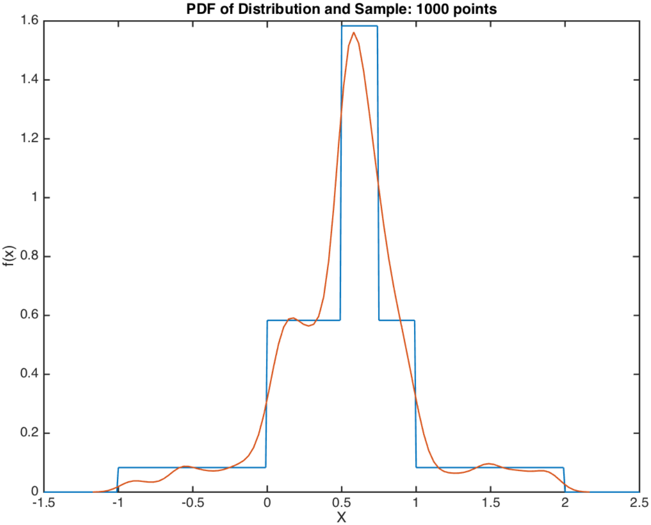
Metropolist-Hastings Random Walk: Example



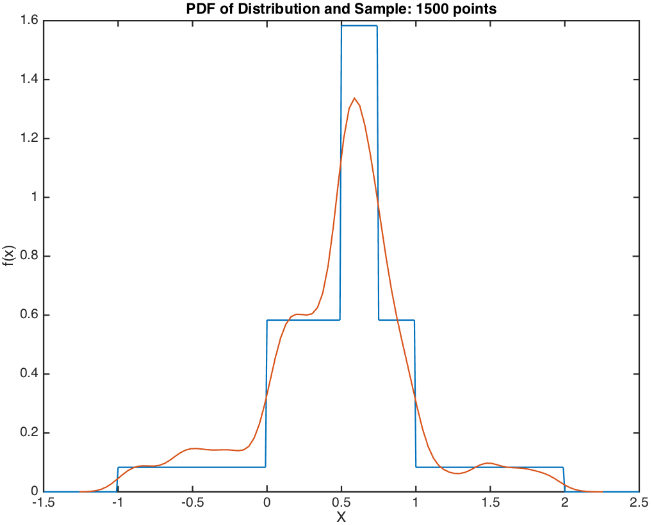
Metropolist-Hastings Random Walk: Example



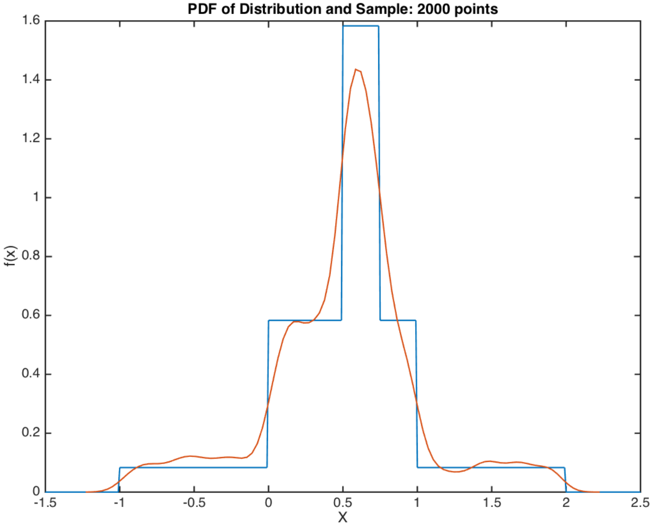
Metropolist-Hastings Random Walk: Example



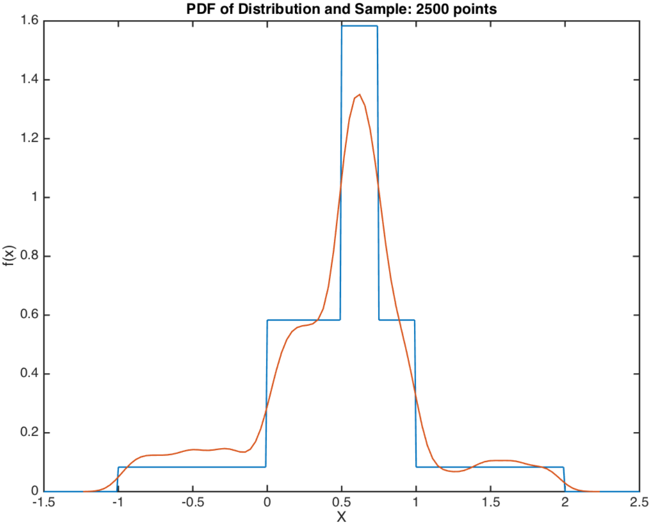
Metropolist-Hastings Random Walk: Example



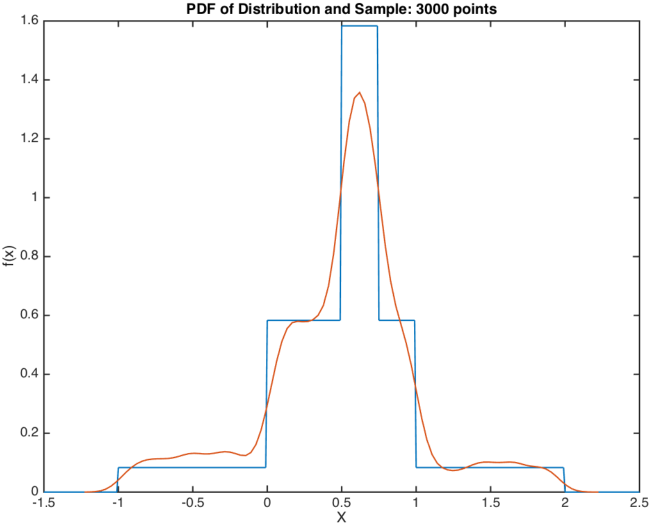
Metropolist-Hastings Random Walk: Example



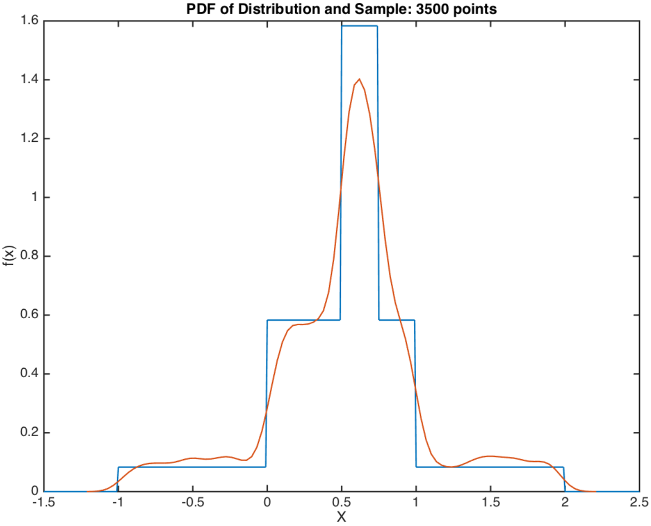
Metropolist-Hastings Random Walk: Example



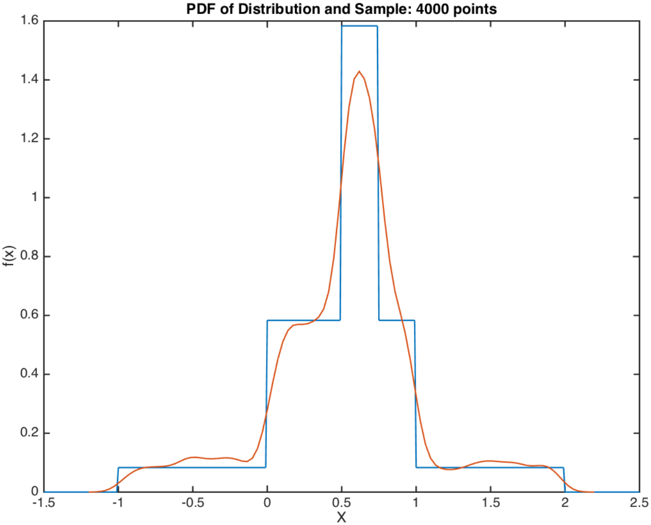
Metropolist-Hastings Random Walk: Example



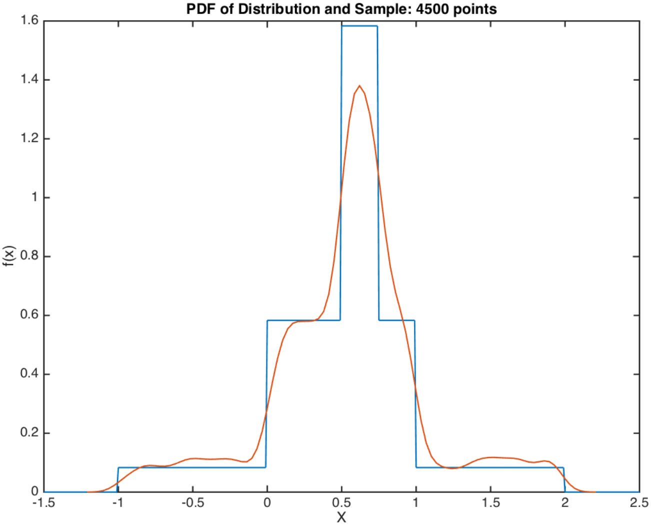
Metropolist-Hastings Random Walk: Example



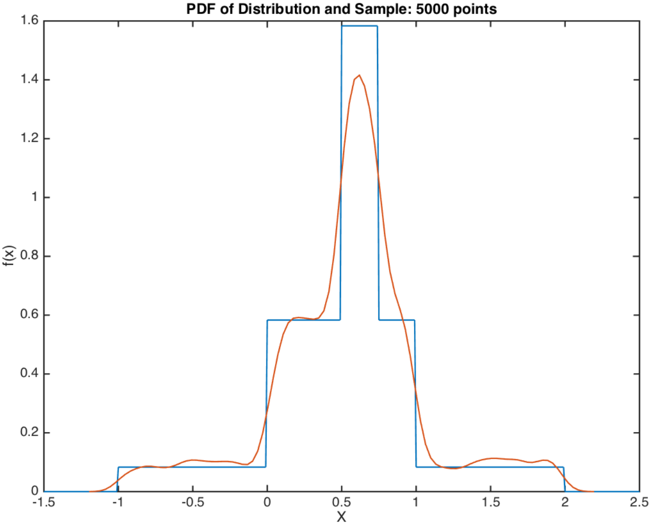
Metropolist-Hastings Random Walk: Example



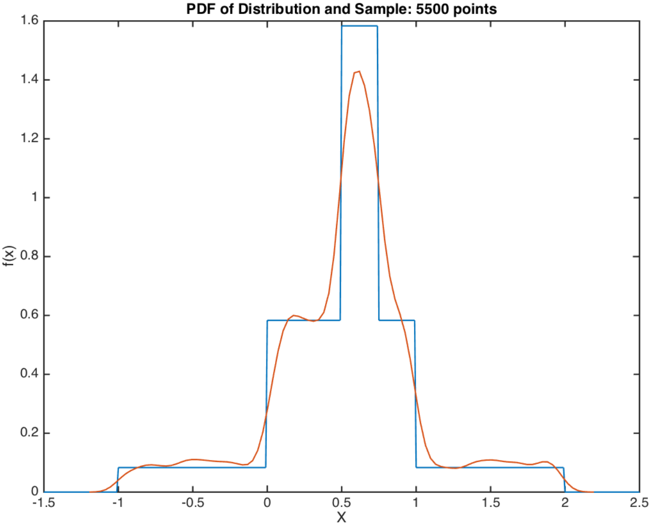
Metropolist-Hastings Random Walk: Example



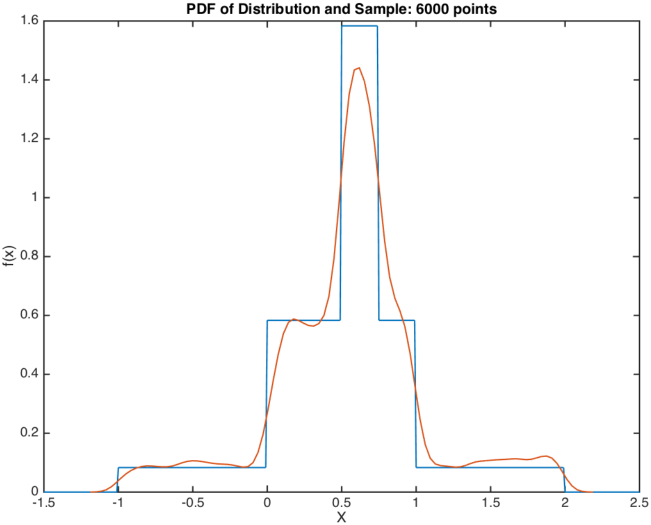
Metropolist-Hastings Random Walk: Example



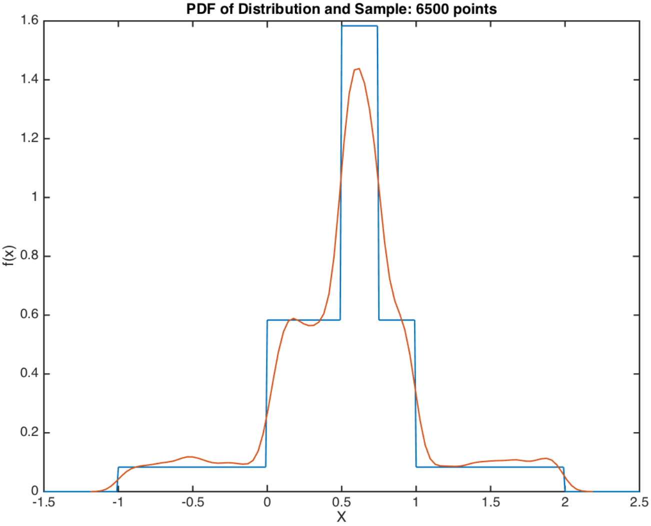
Metropolist-Hastings Random Walk: Example



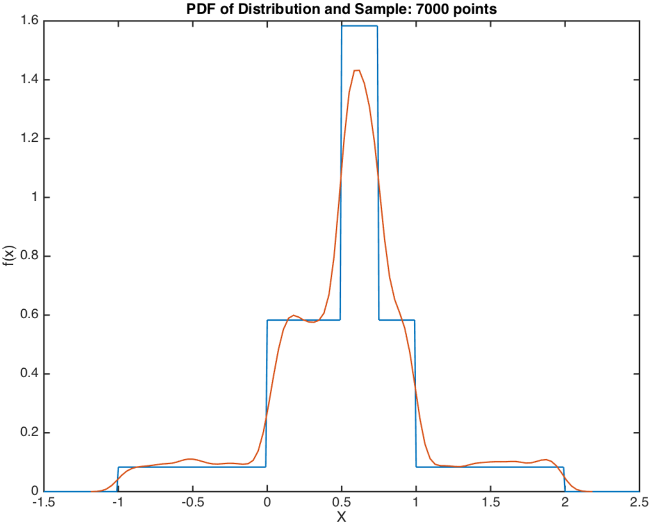
Metropolist-Hastings Random Walk: Example



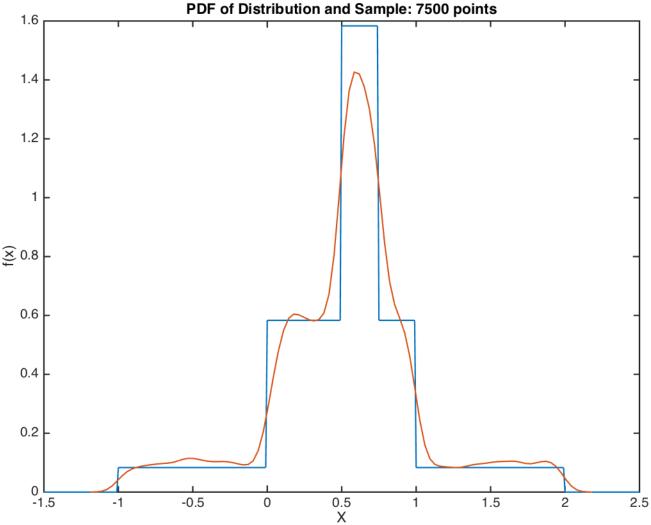
Metropolist-Hastings Random Walk: Example



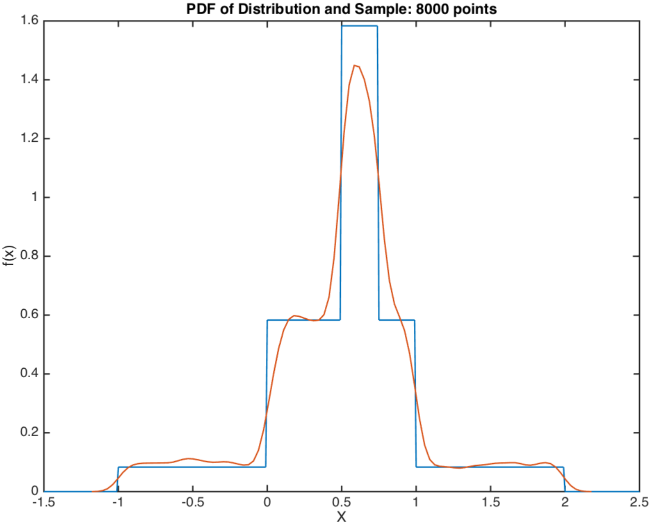
Metropolist-Hastings Random Walk: Example



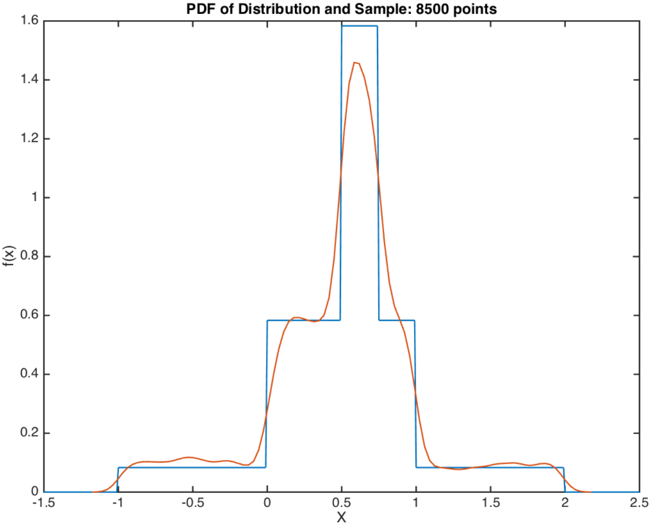
Metropolist-Hastings Random Walk: Example



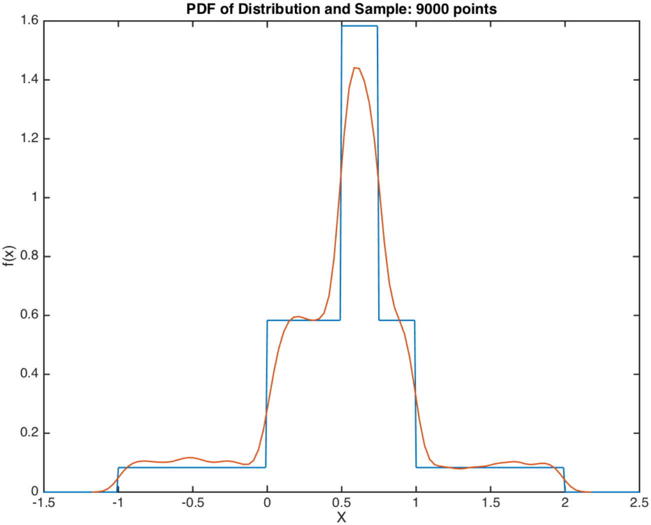
Metropolist-Hastings Random Walk: Example



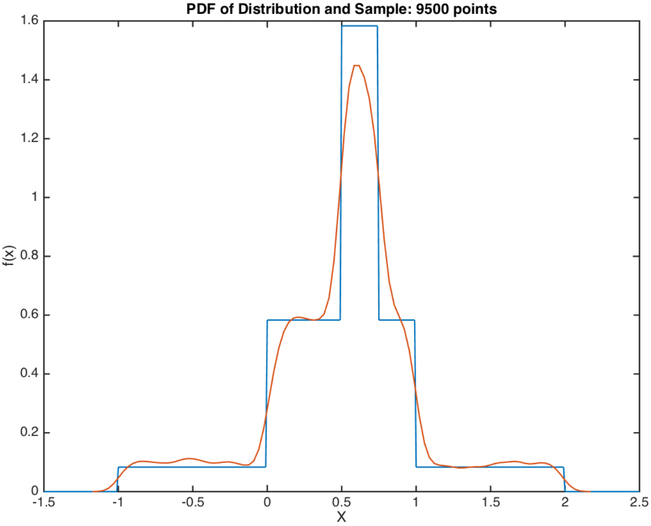
Metropolist-Hastings Random Walk: Example



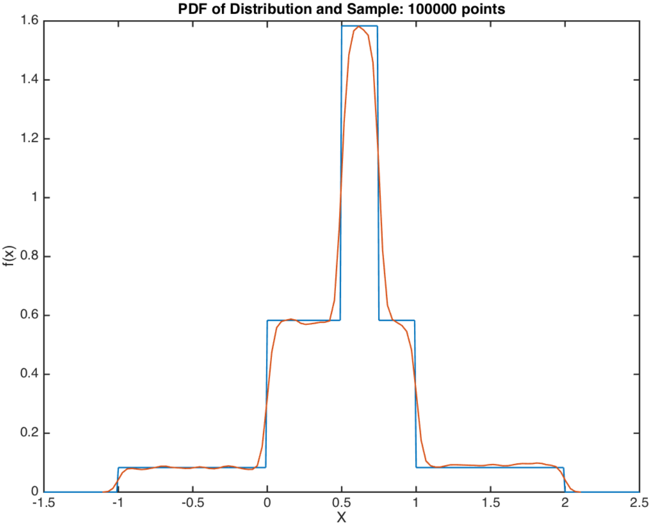
Metropolist-Hastings Random Walk: Example



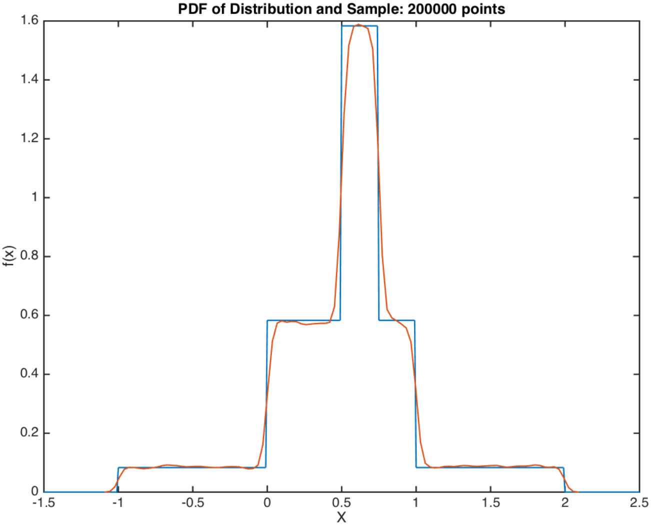
Metropolist-Hastings Random Walk: Example



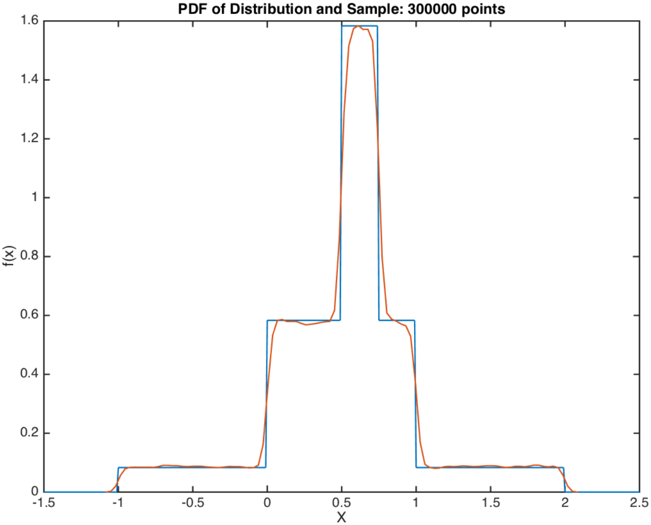
Metropolist-Hastings Random Walk: Example



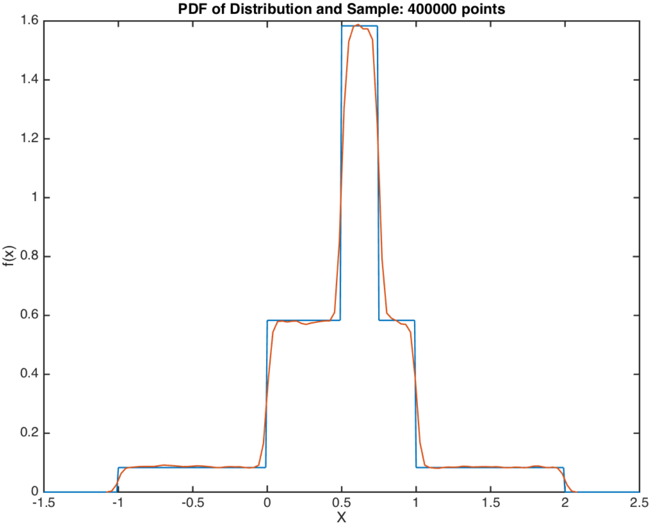
Metropolist-Hastings Random Walk: Example



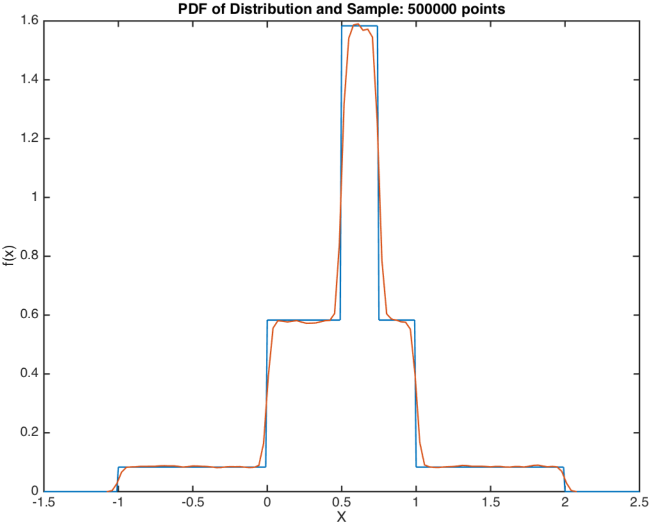
Metropolist-Hastings Random Walk: Example



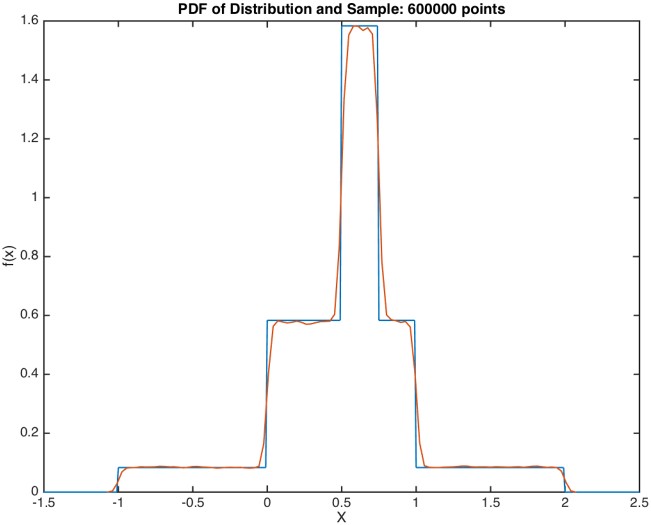
Metropolist-Hastings Random Walk: Example



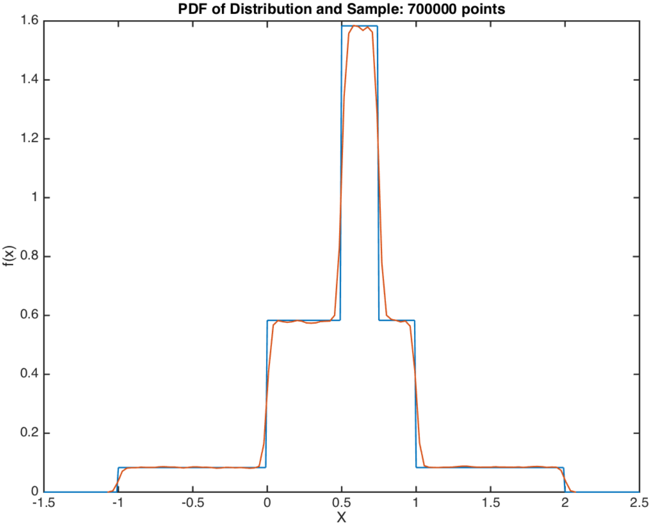
Metropolist-Hastings Random Walk: Example



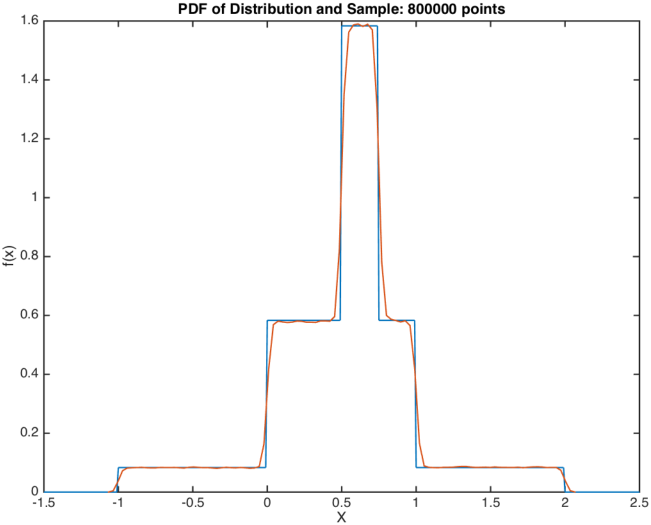
Metropolist-Hastings Random Walk: Example



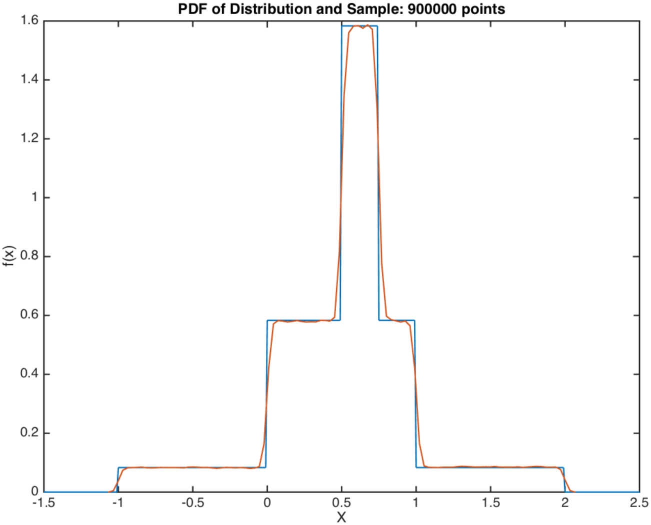
Metropolist-Hastings Random Walk: Example



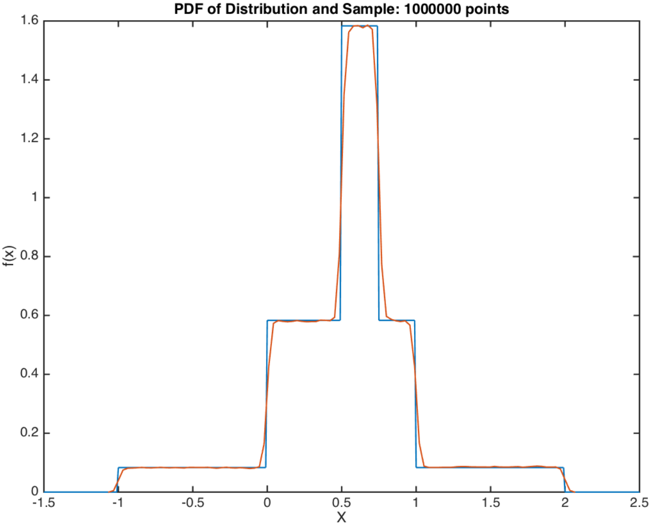
Metropolist-Hastings Random Walk: Example



Metropolist-Hastings Random Walk: Example



Metropolist-Hastings Random Walk: Example



Sampling PDF

- Let's say our distribution is one-dimensional:

$$x \sim 0.5U(0,1) + 0.25U(-1,2) + 0.25U(0.5,0.75)$$

- Choose sampling distribution centered around current point:

$$g(x'|x) \sim \mathcal{N}(x, 0.1)$$

Why learn MH & Numerical Integration?

- Many-dimensional problems
- Bayesian estimation
  - Write down model
  - Write down distribution of parameters $f(\theta)$
  - Simulate many models to get model distribution of data $f(x|\theta)$
  - Update your beliefs: $f(\theta|x) \propto f(\theta)f(x|\theta)$
- Typically need to draw from posterior distribution without an analytical calculation
- Use M-H

Metropolis-Hasting Example-Motivation

- Consider a human capital accumulation problem (as in Homework 1, 2 in 2023)
- Households choose how much time to study vs. work, no asset accumulation
- VFI problem, but want to estimate depreciation rate and ability level
- Bayesian Structural Estimation!

Model/Problem

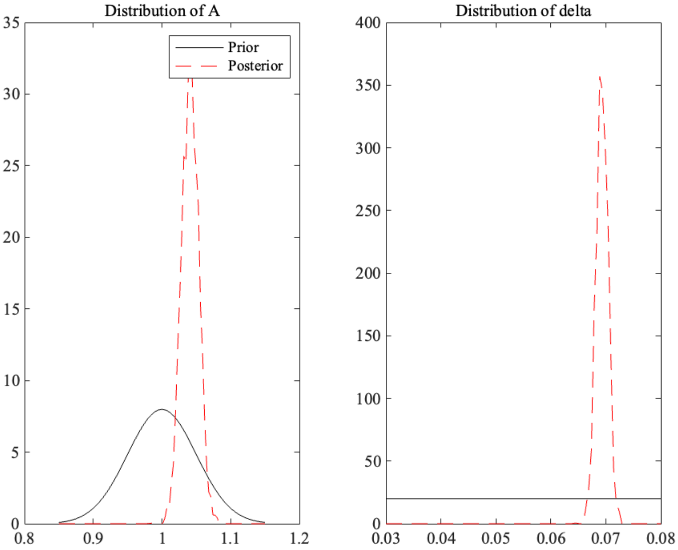
$$V(h) = \max_i \{log((1-i)h) + \beta V((1-\delta)h + Ai^{\gamma})\}$$

- Where A, δ are to be estimated from decision data in the problem (γ is known).
- For convenience (real structural problem would have more natural probabilities), $i^{obs} = i^{data} + \epsilon$ where $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$
- For convenience, I solve this for a grid of δ and A
- This gives me pr(data—parameter)
- Metropolis-Hastings lets me combine model based probability p(parameter) to get p(parameter—data) without having to integrate across

Model

- Assume prior distribution of $\{A, \delta\}$ (here independently distributed $f(A) \sim \mathcal{N}(1, 0.05)$ $f(\delta) \sim \mathcal{U}(0.03, 0.08)$
- Use transition distribution $\Delta\theta \sim \mathcal{N}(0, 0.01)$
- Have probability of choice conditional on parameter $g(i|\theta)$ from structural model
- Have starting point for MCMC $\theta_0 \sim \mathcal{N}(1, 0.05)$
  - Draw new $\theta_1 = \theta_0 + \Delta\theta$
  - Compare: $r = \frac{g(i|\theta_1)f(\theta_1)}{g(i|\theta_0)f(\theta_0)}$
  - Either accept or reject based on MH algo
  - Repeat many times
  - Now have sampling from posterior distribution
- See Main.m, VFIGridSolver.m, and MCMC.m

Model


Distribution of A
Prior
Posterior
35
30
25
20
15
10
5
0
0.8
0.9
1
1.1
1.2
Distribution of delta
400
350
300
250
200
150
100
50
0
0.03
0.04
0.05
0.06
0.07
0.08

Aside: Multiple Hypothesis Testing and Maximum F-Statistics

- Data is tortured. When you see...
  - ..an experiment with multiple test groups or with without very strong theoretical justification, be skeptical!
  - ...a regression that could have been run differently, or with many potential controls, weighting options, and unit-of-observation choices, be skeptical!
- Not all bad: t-statistics might just be heuristics...when I see 0.01 in a regression where I could imagine 10 other setups, I know the real p-value is around 0.1.
- But we might want to take statistics seriously, or data-mine honestly
- We can simulate the distribution of the maximum F-statistic, or the maximum t-statistic.
- See MonteCarlo.do and similar exercises (Note: Stata!)