

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 $i \in (1, \dots, n)$, (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:

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

GIBBS SAMPLING: EXAMPLE

- ▶ Assume a distribution:

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

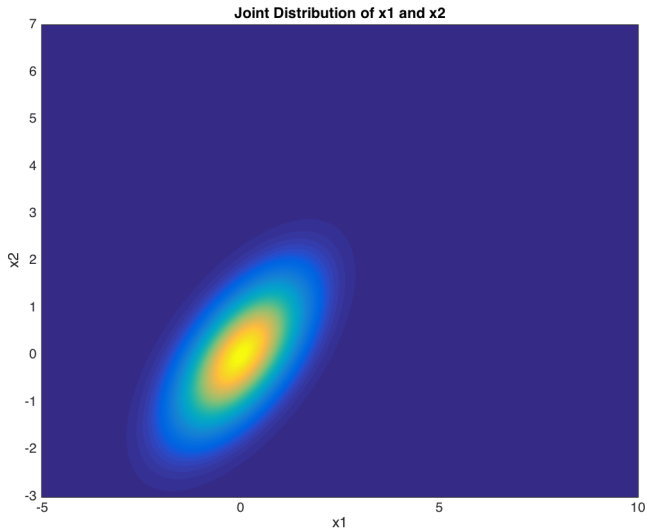
- ▶ Then we can get the conditional marginal distributions:

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

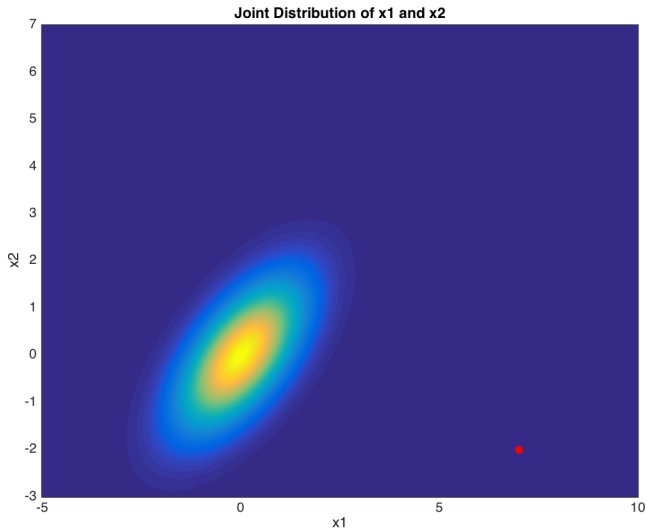
$$x_2 | x_1 \sim \mathcal{N}(\rho x_1, (1 - \rho)^2)$$

- ▶ Iteratively sample from these.
- ▶ See Gibbs.m

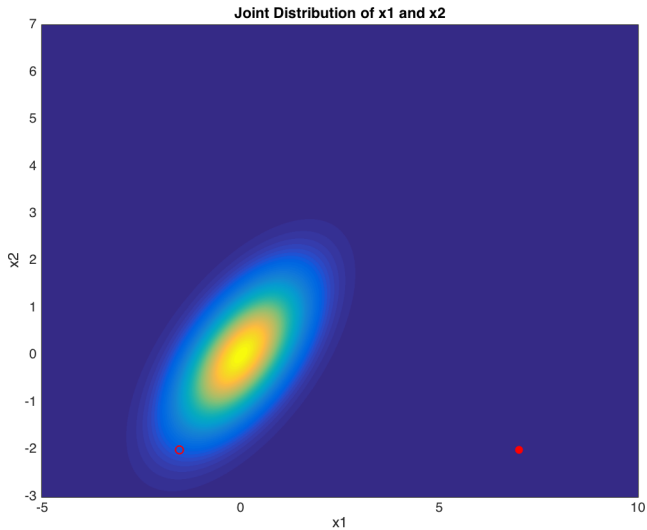
GIBBS SAMPLING: EXAMPLE



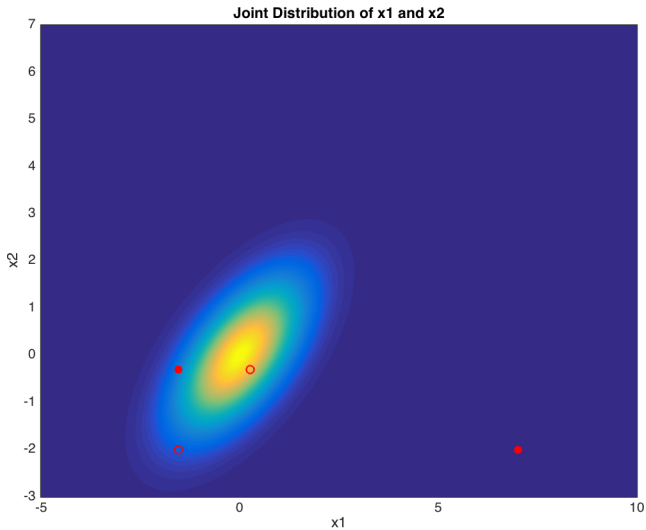
GIBBS SAMPLING: EXAMPLE



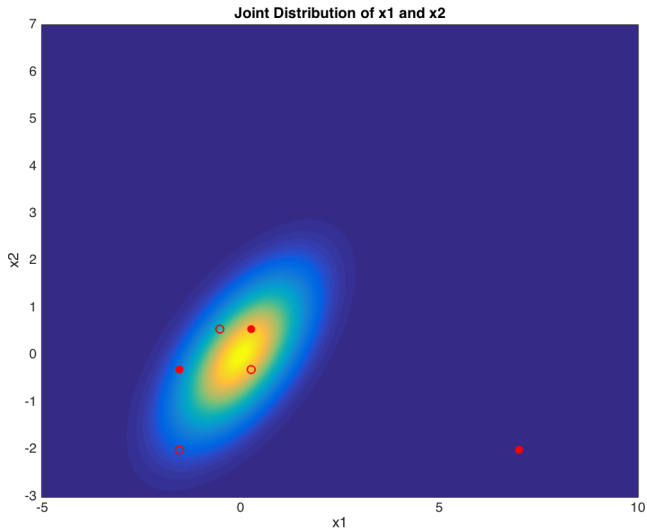
GIBBS SAMPLING: EXAMPLE



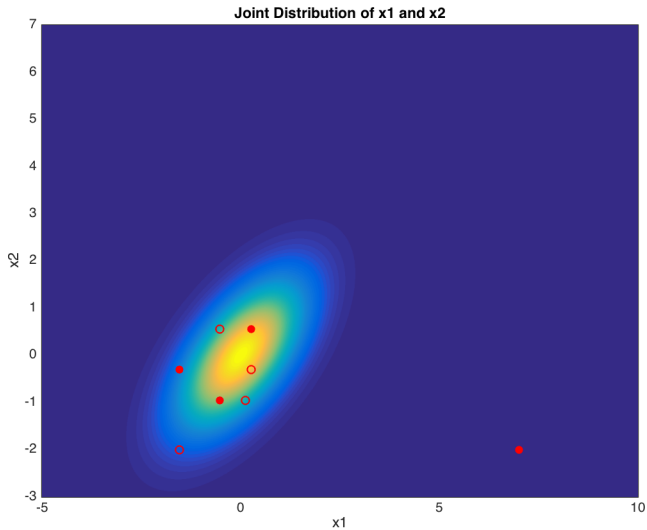
GIBBS SAMPLING: EXAMPLE



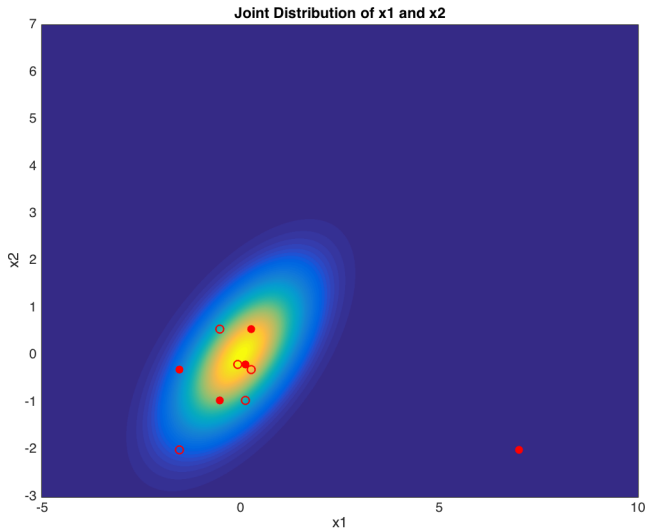
GIBBS SAMPLING: EXAMPLE



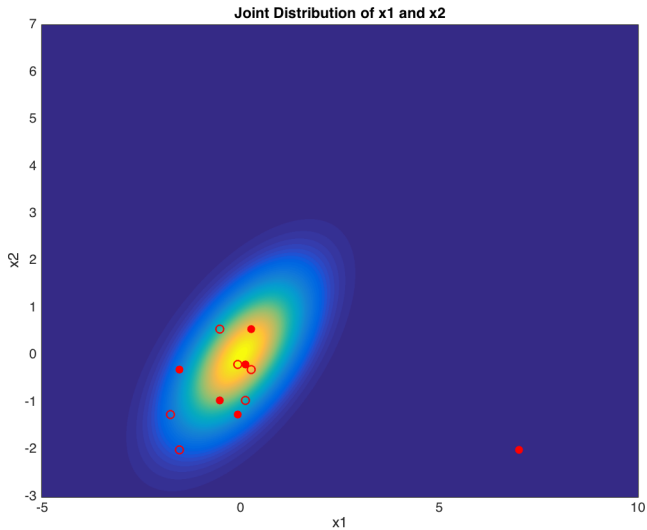
GIBBS SAMPLING: EXAMPLE



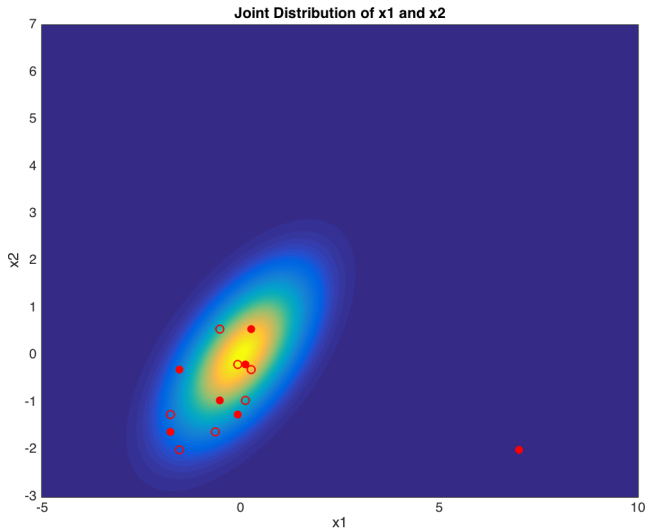
GIBBS SAMPLING: EXAMPLE



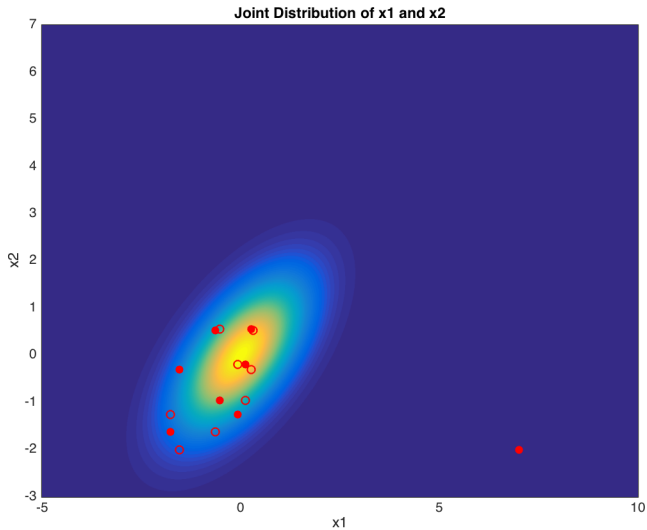
GIBBS SAMPLING: EXAMPLE



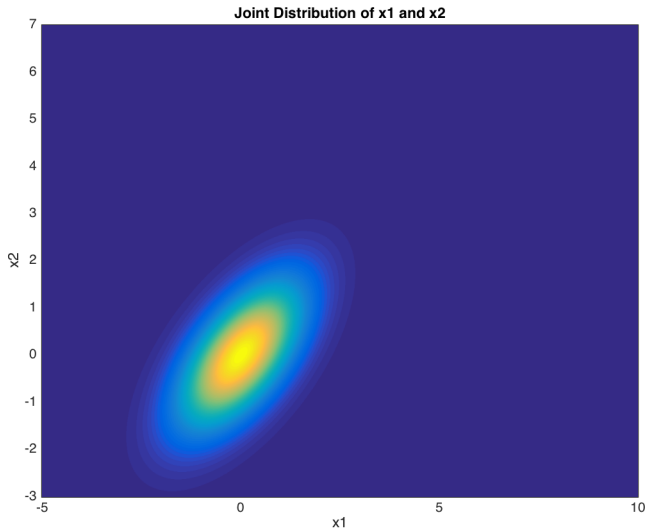
GIBBS SAMPLING: EXAMPLE



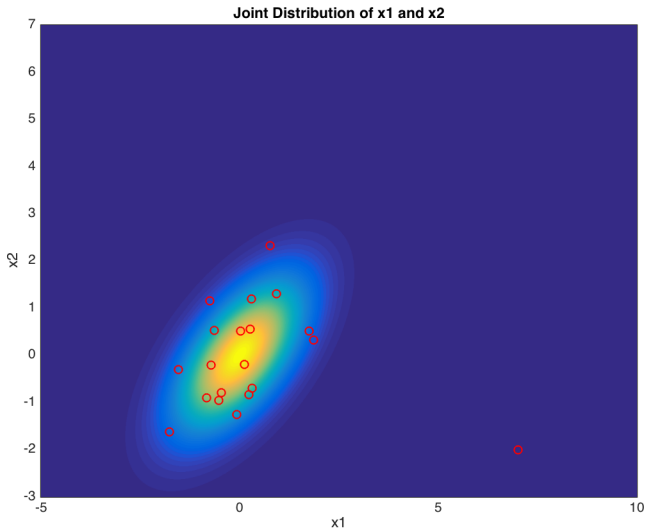
GIBBS SAMPLING: EXAMPLE



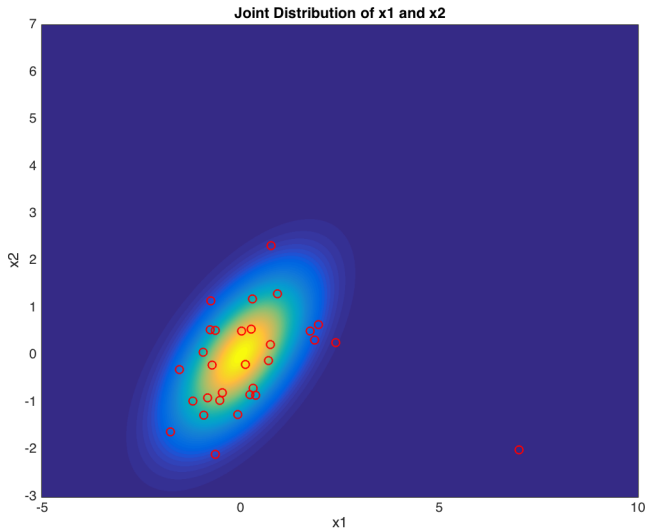
GIBBS SAMPLING: EXAMPLE



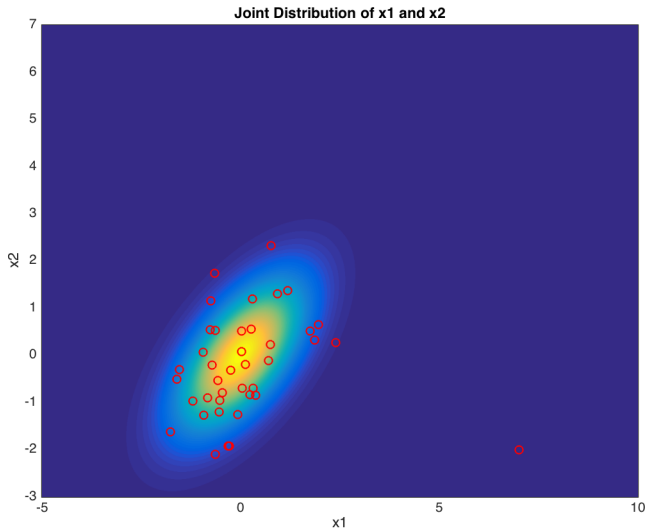
GIBBS SAMPLING: EXAMPLE



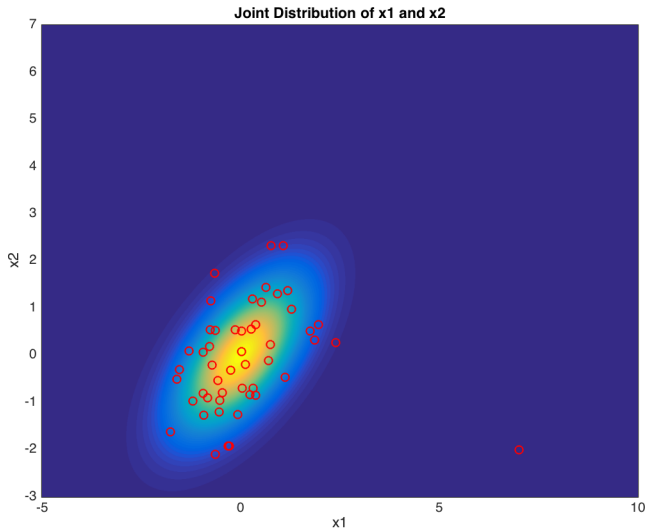
GIBBS SAMPLING: EXAMPLE



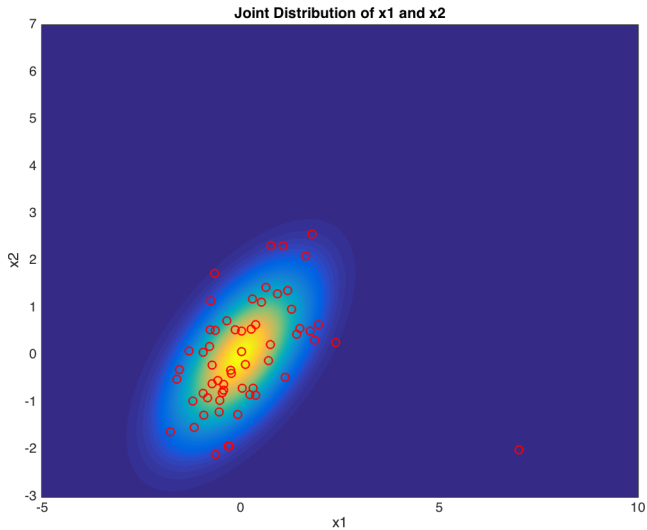
GIBBS SAMPLING: EXAMPLE



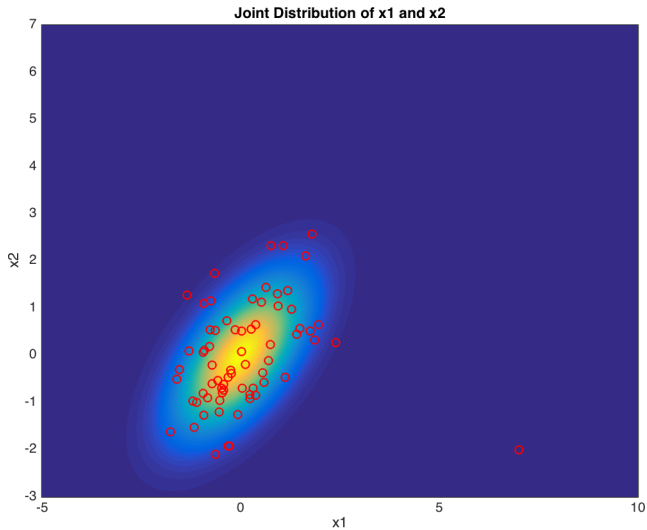
GIBBS SAMPLING: EXAMPLE



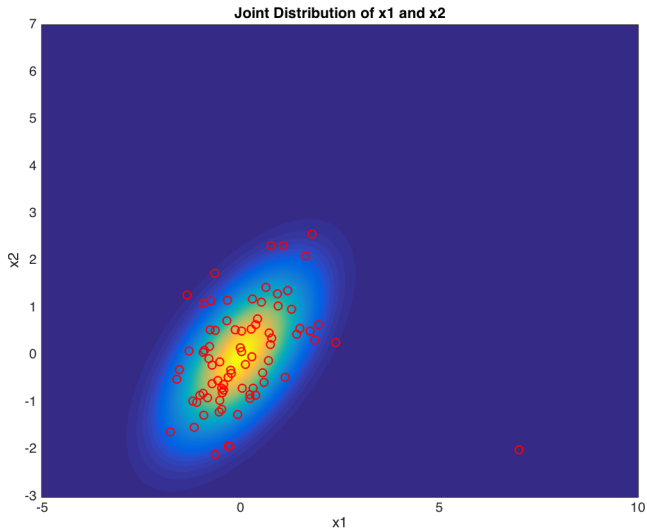
GIBBS SAMPLING: EXAMPLE



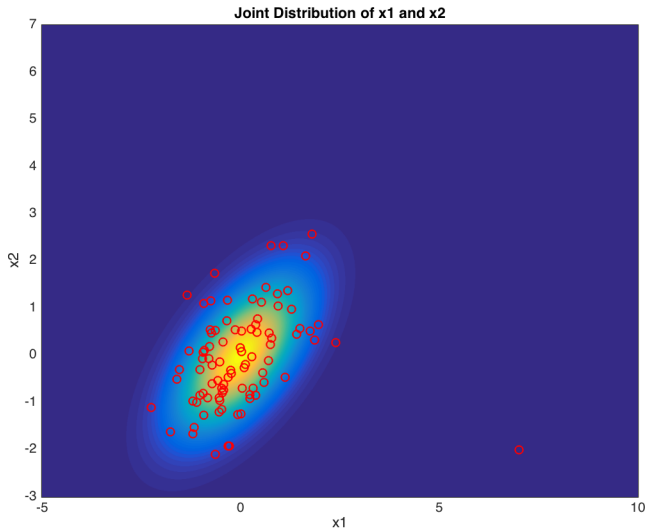
GIBBS SAMPLING: EXAMPLE



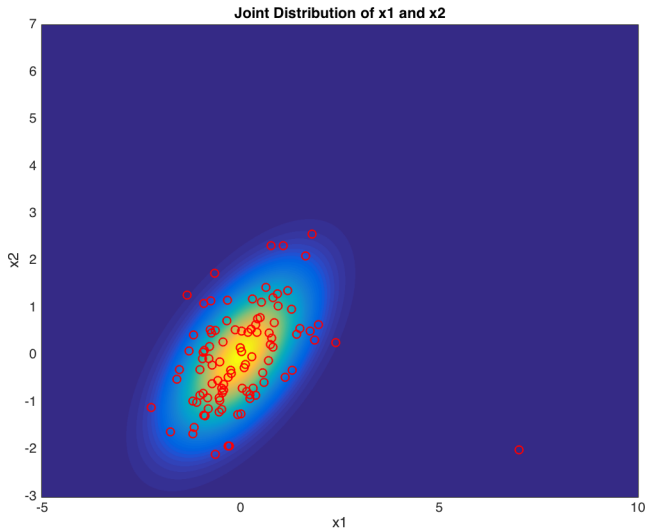
GIBBS SAMPLING: EXAMPLE



GIBBS SAMPLING: EXAMPLE



GIBBS SAMPLING: EXAMPLE



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:

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

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

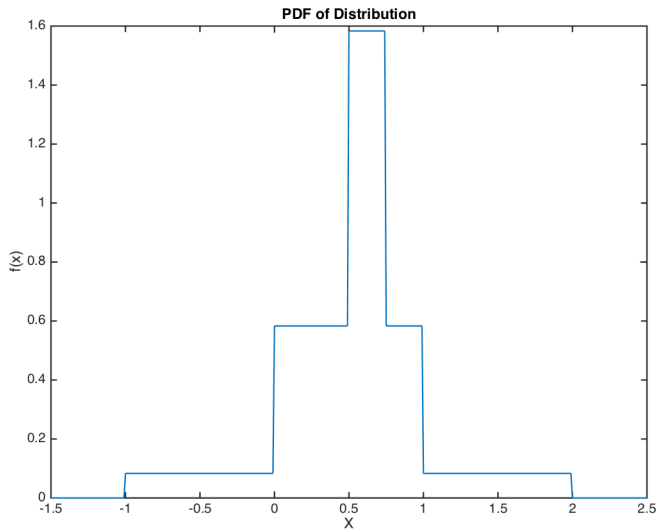
- ▶ Let's say our distribution is one-dimensional:

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

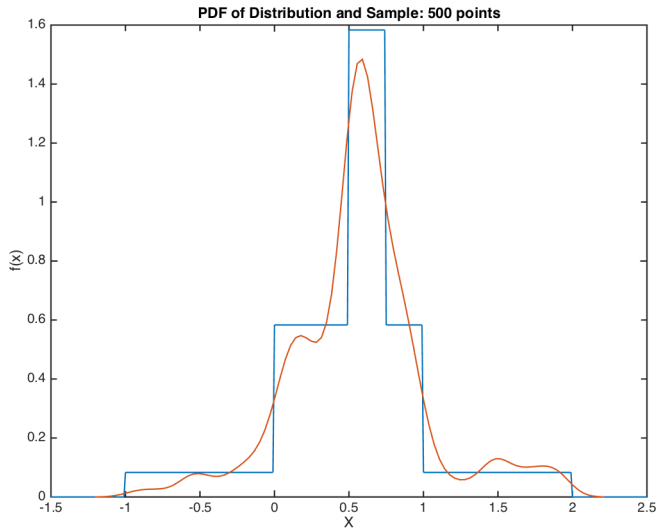
- ▶ Choose sampling distribution centered around current point:

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

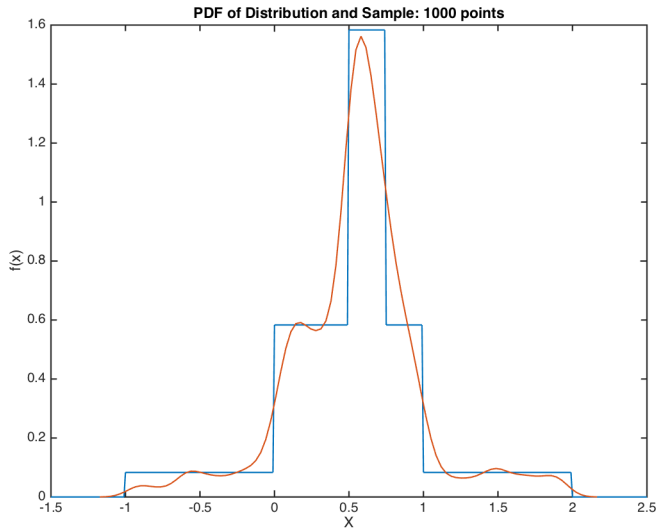
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



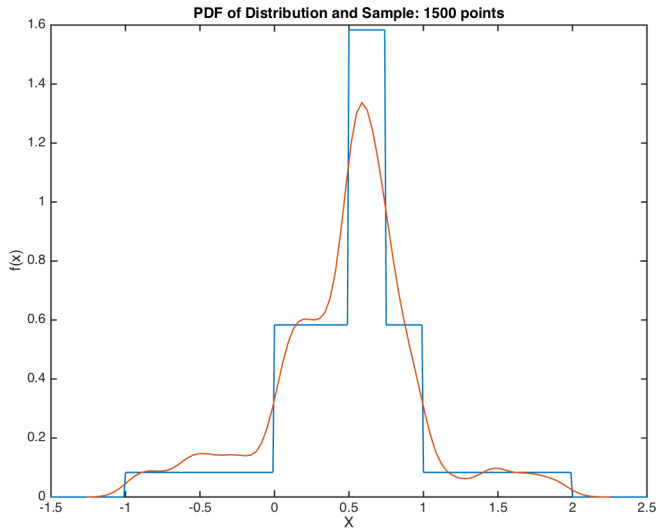
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



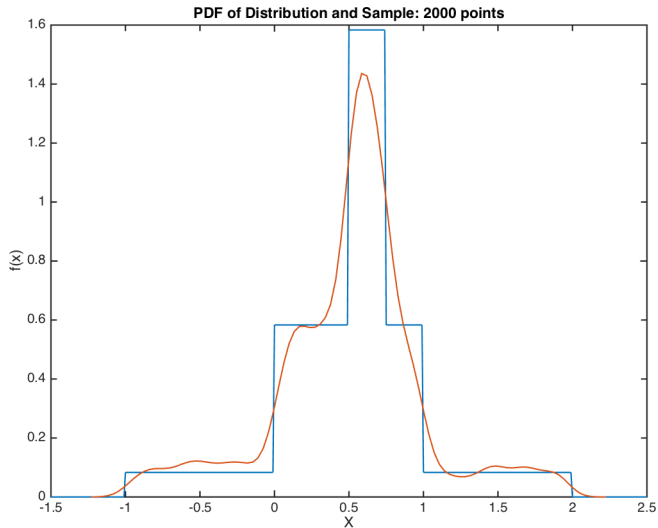
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



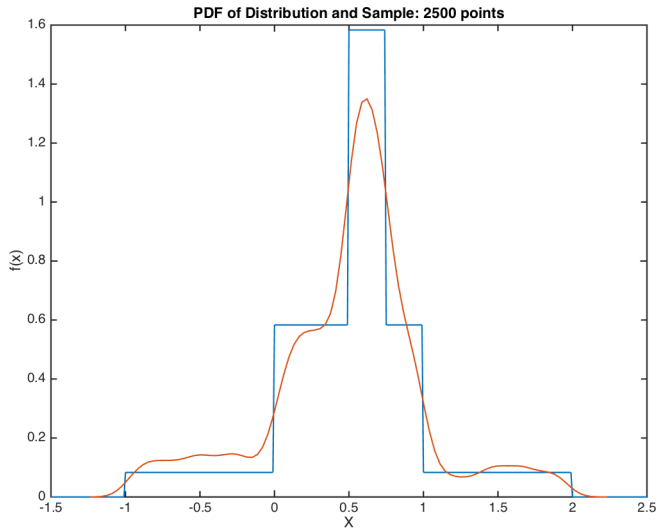
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



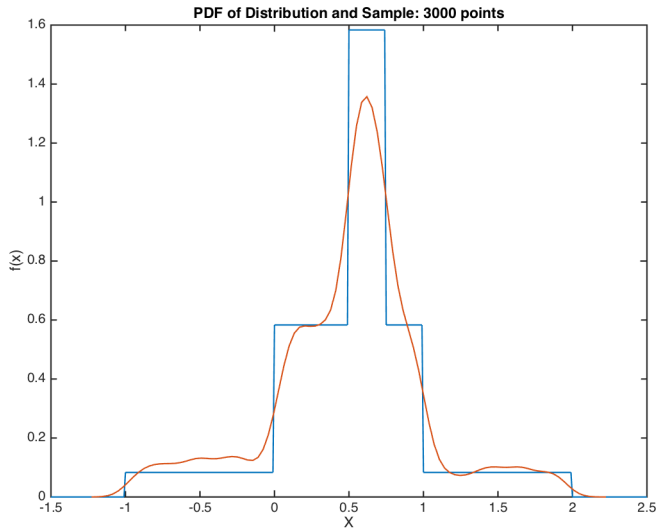
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



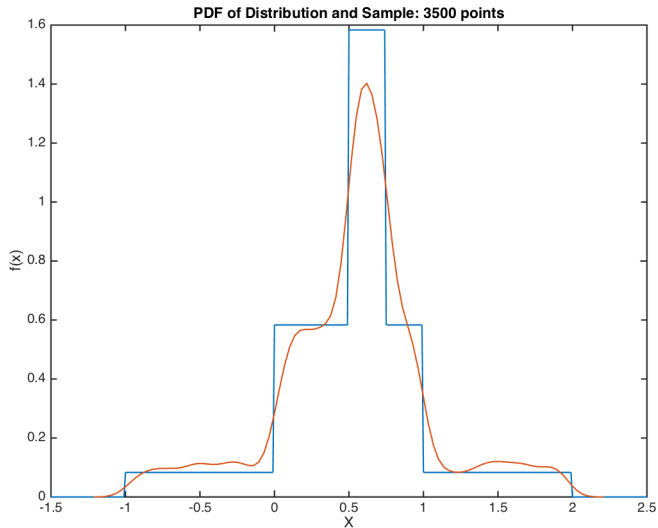
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



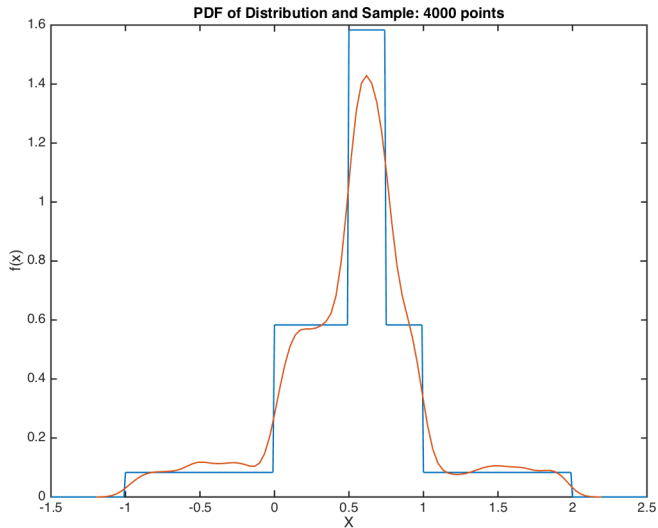
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



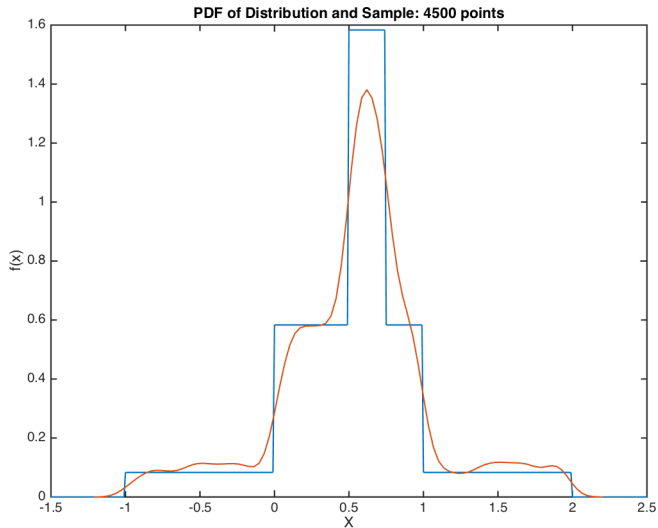
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



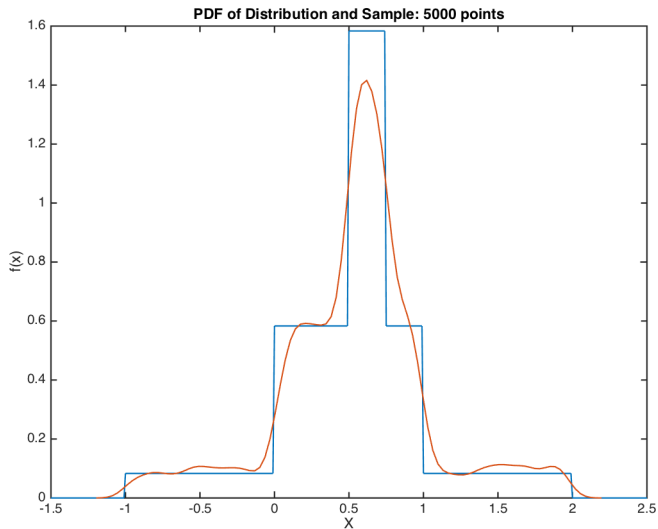
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



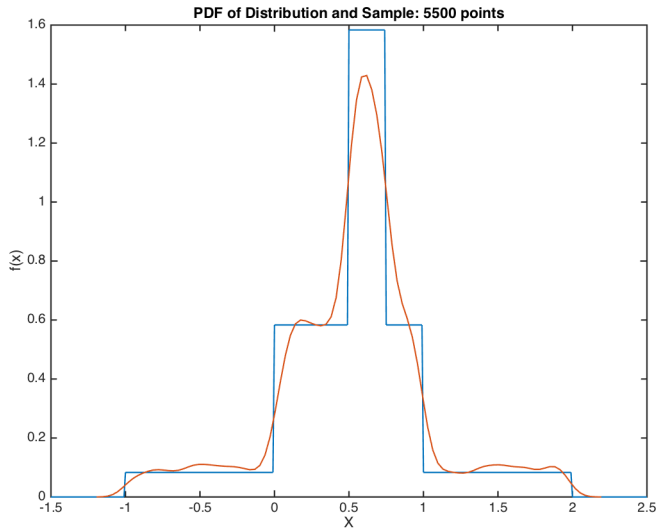
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



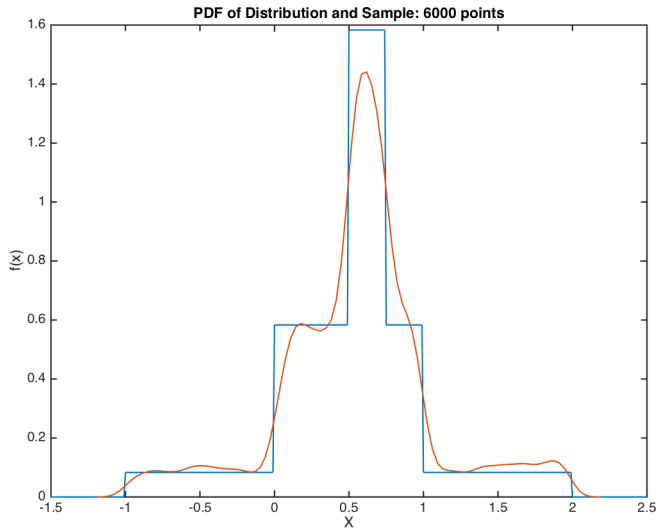
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



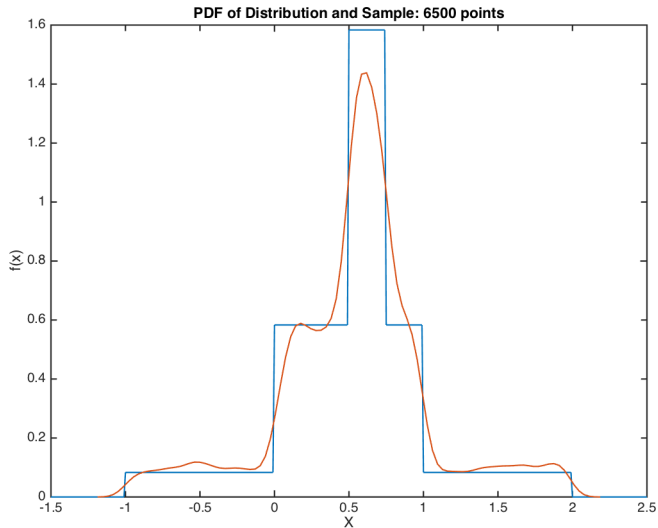
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



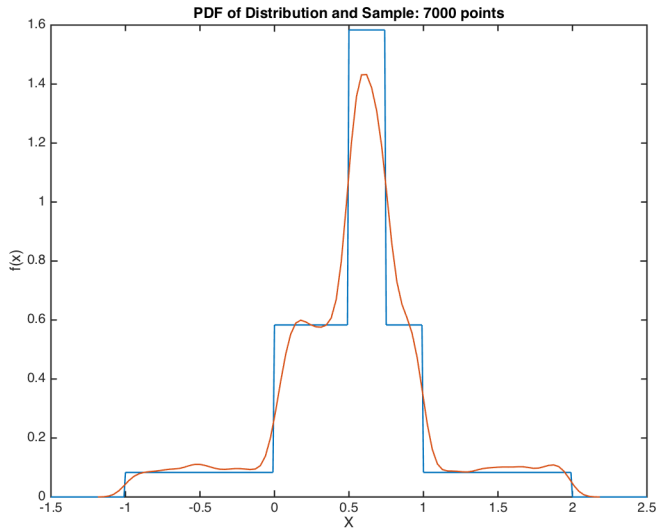
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



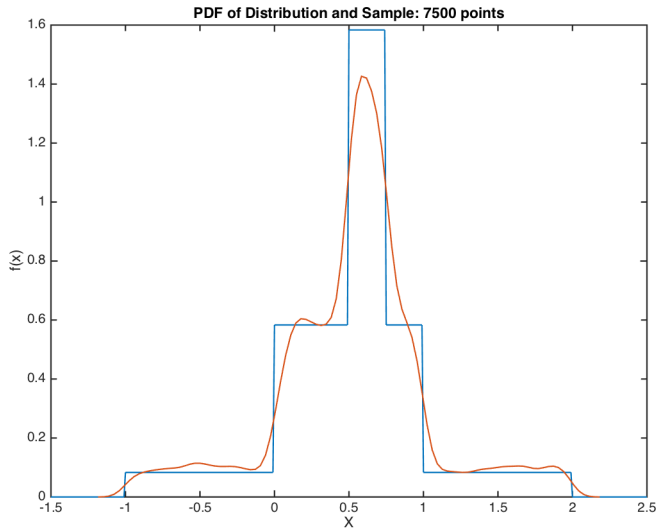
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



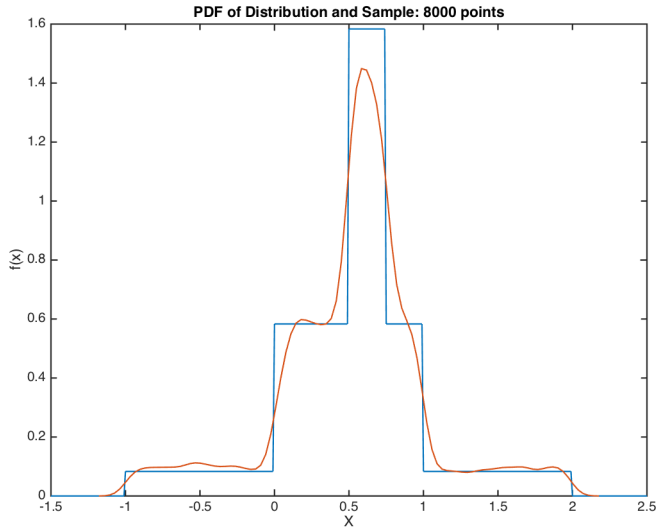
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



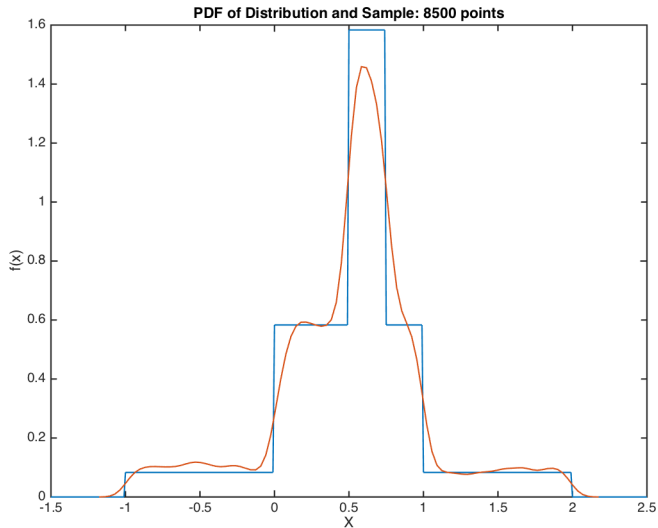
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



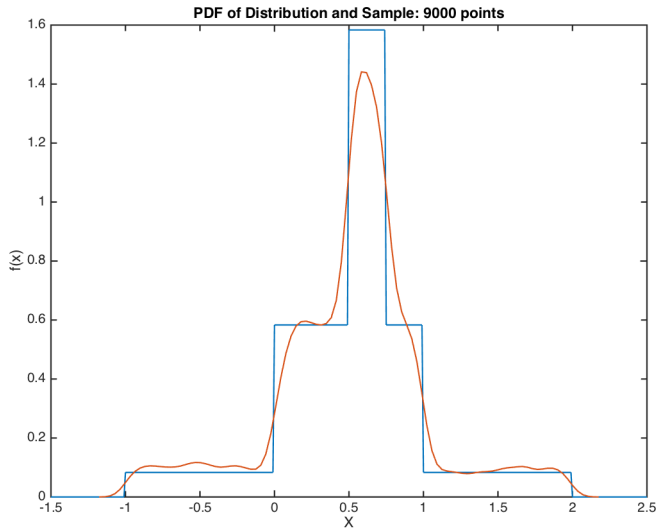
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



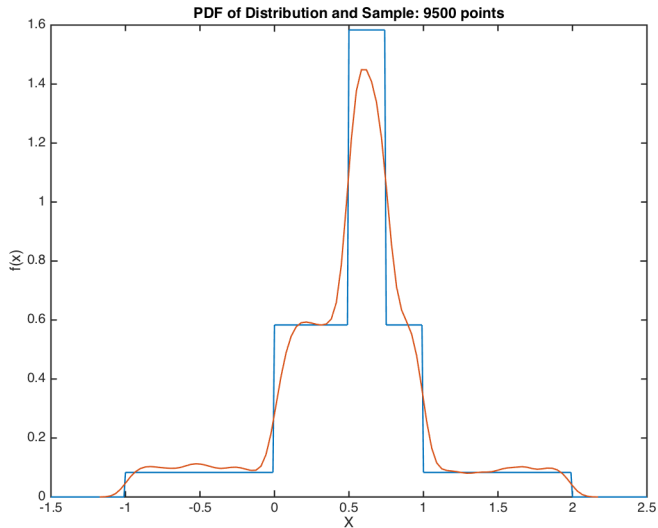
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



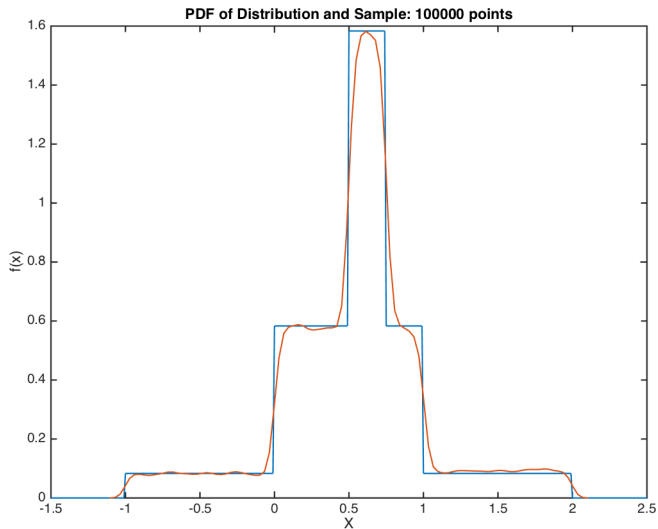
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



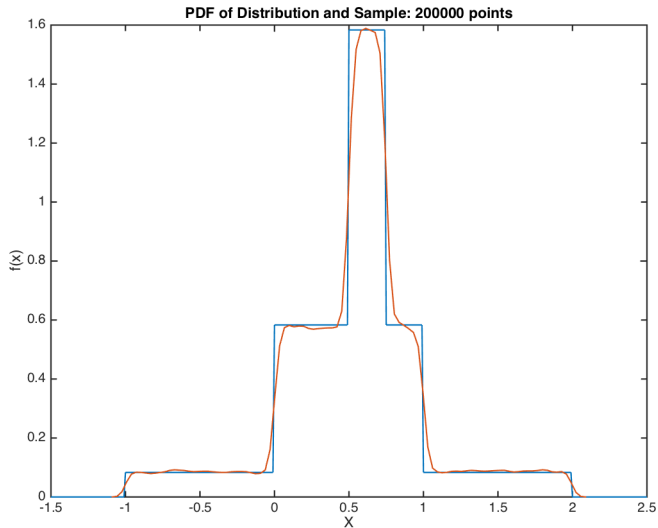
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



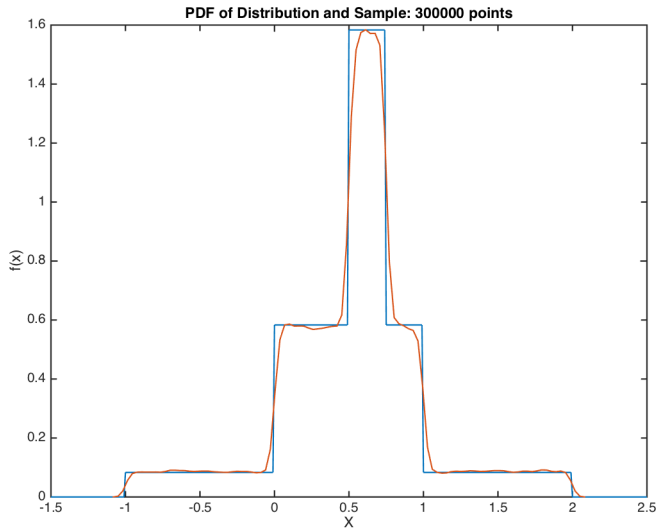
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



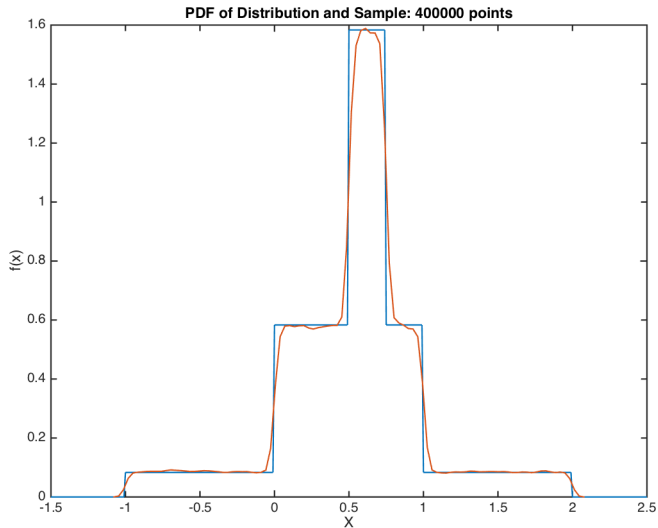
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



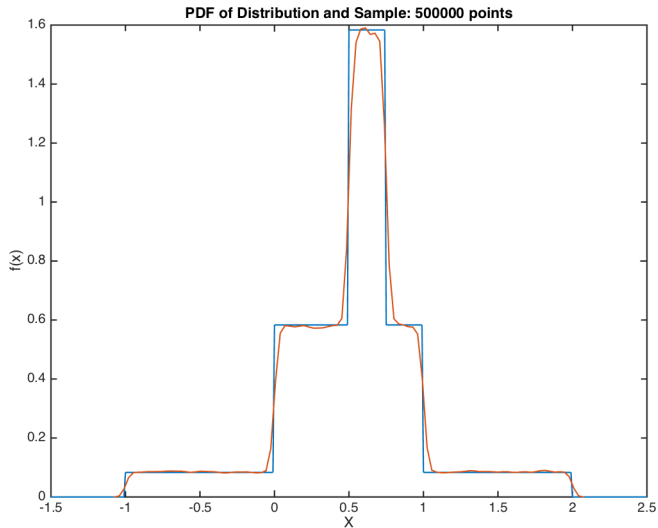
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



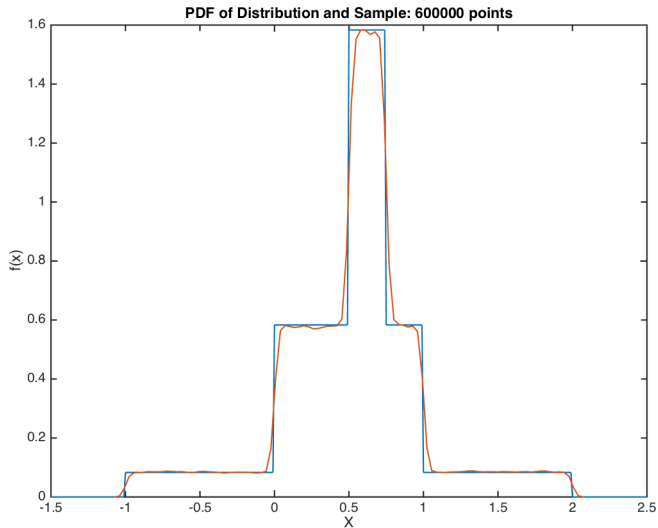
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



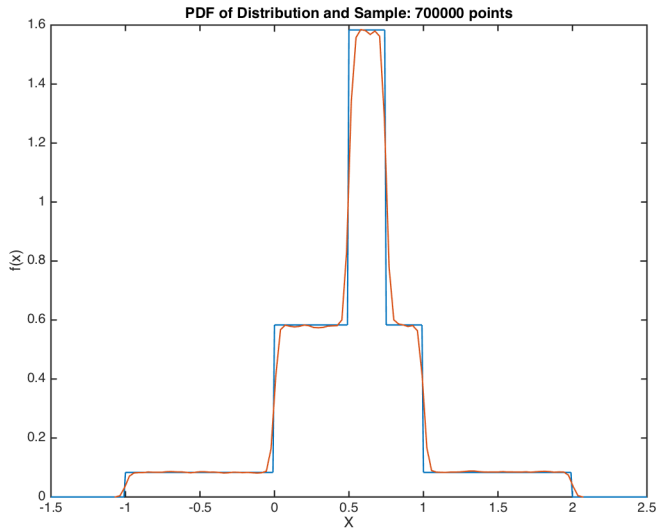
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



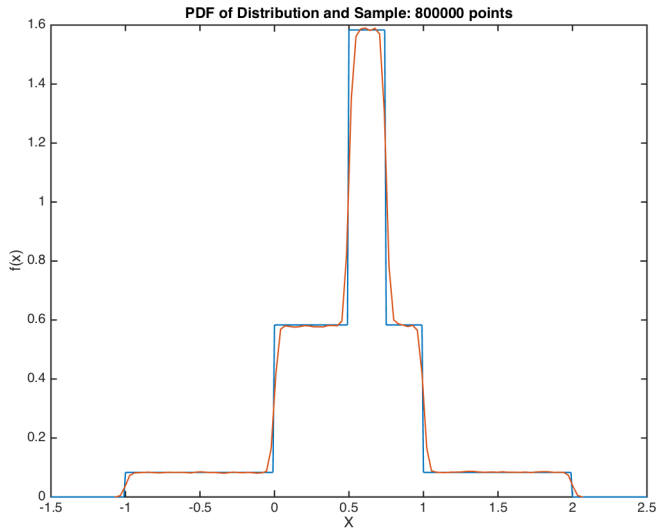
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



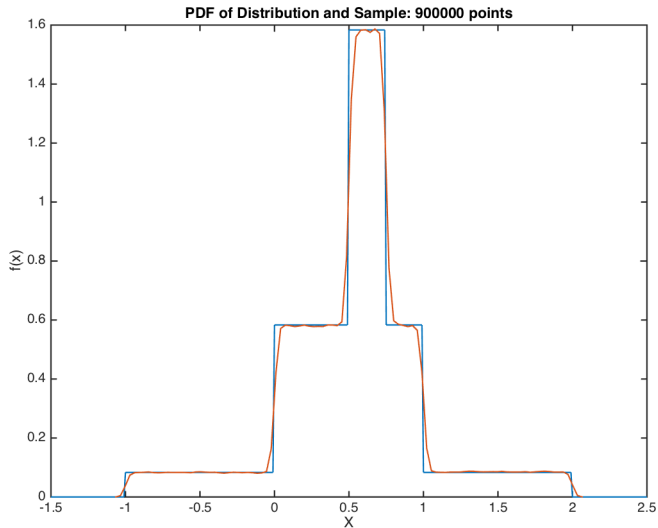
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



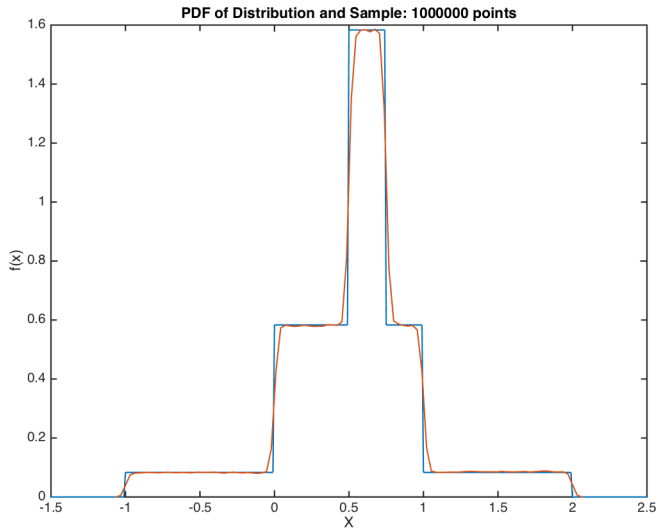
METROPOLIST-HASTINGS RANDOM WALK: EXAMPLE



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

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!)