# Markov Chain Monte Carlo

## Gibbs Sampling

Sachin Shanbhag

Department of Scientific Computing
Florida State University,
Tallahassee, FL 32306.

# Contents


- Introduction
  - joint distribution from conditional distributions
  - Hammersley-Clifford Theorem
- Algorithm
- Example: Bivariate Gaussian Distribution
- Relationship between Metropolis-Hastings and Gibbs
  - rejection rate of zero
  - burn-in, thinning, convergence
- Example: Coupled Oscillators


Discovered by brothers Stuart and Donald Geman in 1984.

Named after statistical physicist Gibbs.

# Introduction

Gibbs sampling is an MCMC algorithm for:

- drawing samples from a multivariate joint PDF (jPDF) $\pi(x_1, x_2, ..., x_n)$ by,
- (directly) sampling one random variable at a time from its conditional distribution,

$$\pi(x_i | x_1, x_2, ...x_{i-1}, x_{i+1}, ..., x_n).$$

This is a popular method, when the the full set of conditional PDFs available (*example*: Bayesian analysis, and some physics problems).

It can be thought of as a special case of the Metropolis - Hastings MCMC method, where the proposal move is designed to guarantee acceptance.

# Hammersley-Clifford Theorem

Let's me illustrate a simple form of the theorem that connects joint and conditional PDFs

Consider a simple 2D jPDF $\pi(x, y)$. Show that it can be rewritten in terms of conditional PDFs:

$$\pi(x, y) = \frac{\pi(y|x)}{\int \frac{\pi(y|x)}{\pi(x|y)} dy} \tag{1}$$

See Notes for Proof.

# Philosophical Importance

This provides the philosophical and mathematical basis for Gibbs sampling.

If you have joint PDF $\pi(x, y)$, you can extract marginal and conditional distributions.

The Hammersley-Clifford theorem says "if you know **all** the conditionals, you can (almost) reconstruct the joint PDF!"[1]

[1] subject to compatibility conditions; see Arnold and Press, *J. Am. Stat. Assoc.*, **84**(205), 152-156, **1989**.

# Higher dimensions

You may perhaps find the following argument helpful for any dimension.

If $\mathbf{x} = (x_1, x_2, ..., x_n)$, and $\mathbf{x}_{-i} = (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$, the conditional distribution of $\mathbf{x}_i$ given all the other variables is to the joint distribution:

$$\pi(x_i|\mathbf{x}_{-i}) = \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}_{-i})} \propto \pi(\mathbf{x}),$$

since $\pi(\mathbf{x}_{-i})$ is independent of $x_i$ and may be thought of as a normalization constant.

The individual conditionals are proportional to the joint!

$$\pi(x_i|\mathbf{x}_{-i}) \propto \pi(\mathbf{x}).$$

# Algorithm

Let consider the algorithm for $n = 3$ dimensions.

It can be generalized to arbitrary dimensions easily.

Let $\mathbf{x} = \{x_1, x_2, x_3\}$ represent the state, and $\pi(\mathbf{x})$ be the distribution we seek to sample from.

1. Set $i = 0$; Choose initial state $\mathbf{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, x_3^{(0)}\}$
2. Generate the next sample $\mathbf{x}^{(i+1)}$ by sampling each $x_j$, $(j = 1, 2, 3)$ from its conditional PDF
   - sample $x_1^{(i+1)} \sim \pi(x_1 | x_2^{(i)}, x_3^{(i)})$
   - sample $x_2^{(i+1)} \sim \pi(x_2 | x_1^{(i+1)}, x_3^{(i)})$
   - sample $x_3^{(i+1)} \sim \pi(x_3 | x_1^{(i+1)}, x_2^{(i+1)})$
3. Set $\mathbf{x}^{(i+1)} = \{x_1^{(i+1)}, x_2^{(i+1)}, x_3^{(i+1)}\}$; Set $i = i + 1$.
4. Go to step 2, and repeat.

# Example

Let's consider our old friend the bivariate normal.

$$\pi(\mathbf{x}) = \mathcal{N}(\mu, \Sigma),$$

with the mean and covariance matrix,

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix},$$

where the correlation coefficient is $\rho = 0.8$.

Conditional distributions are univariate normal distribution. For $\sigma_1 = \sigma_2 = 1$, the means and standard deviation are given by:

$$\pi(x_1|x_2) = \mathcal{N}(\mu_1 + \rho(x_2 - \mu_2), \sqrt{1-\rho^2}) \tag{2}$$

$$\pi(x_2|x_1) = \mathcal{N}(\mu_2 + \rho(x_1 - \mu_1), \sqrt{1-\rho^2}) \tag{3}$$

# Gibbs Sampling

Let's use Gibbs sampling to sample from the bivariate normal, given these conditional distributions

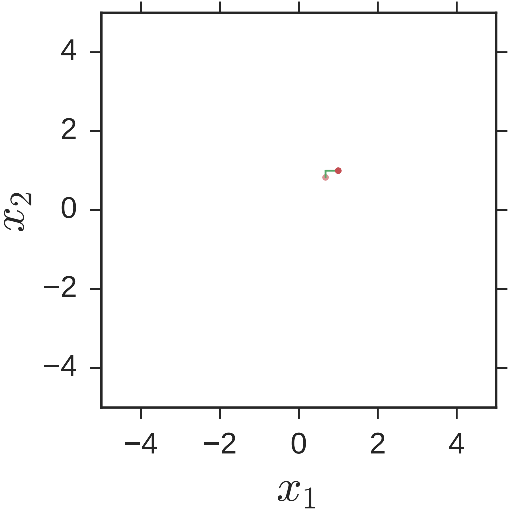
We will need (a) Gibbs update, and (b) driver

```
def gibbsUpdate(x, mu, rho):

    stdv = np.sqrt(1-rho**2)
    newx = np.zeros(x.shape)

    m0      = mu[0] + rho*(x[1]-mu[1])
    newx[0] = np.random.normal(m0, stdv)

    m1      = mu[1] + rho*(newx[0]-mu[0])
    newx[1] = np.random.normal(m1, stdv)

    return newx
```

# Driver

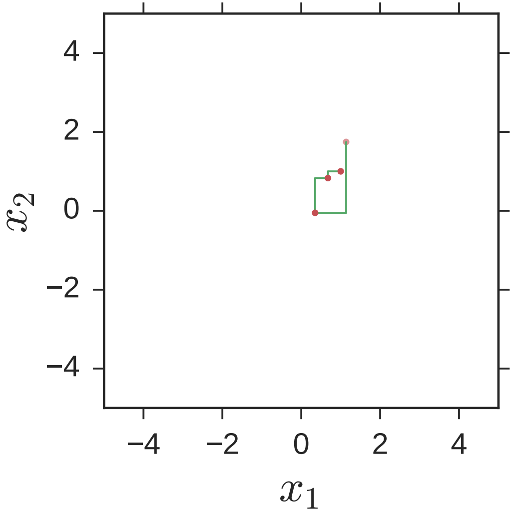
```
def driver(x0, nsteps=5000, mu=np.zeros((2,1)), rho = 0.8):

    # initial state
    x[0,:] = x0.reshape(1,2)

    # main loop
    for i in range(1,nsteps):
        newx = gibbsUpdate(x[i-1], mu, rho)
        x[i,:] = newx.reshape(1,2)

    return x
```

# Result: 1 Sample

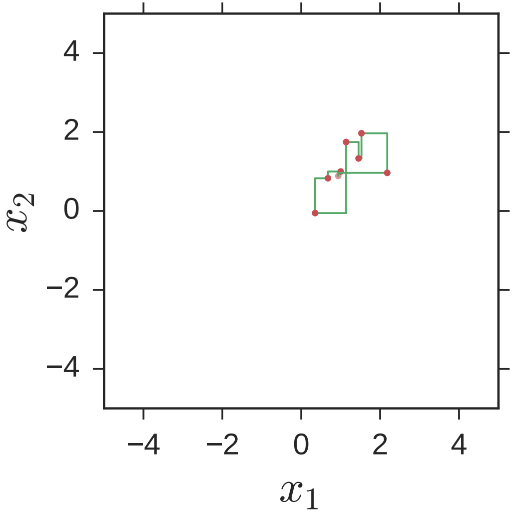



$\mathbf{x}_0 = (1, 1) =$ red dot

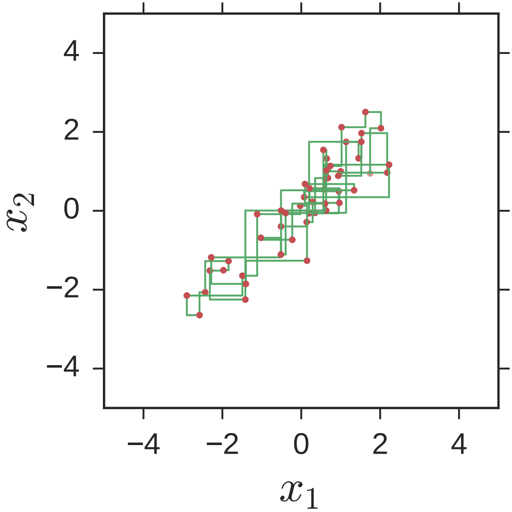
# 3 samples



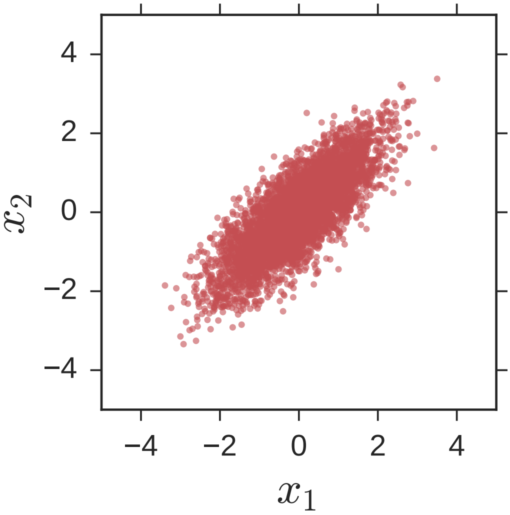
# 8 samples



# 50 samples



# 5000 samples



# Gibbs and MH sampling

Gibbs sampling can be thought of as a particular form of MH sampling.

In MH sampling recall that the acceptance probability is:

$$\text{acc}(o \to n) = \min\left\{1, \frac{\pi_n \pi_{on}}{\pi_o \pi_{no}}\right\},$$

where the second term in the braces is the Hastings ratio.

Consider a simple 2D example, similar to the bivariate Gaussian example above.

Consider a proposal:

$$\mathbf{x}_o = (x_1, x_2) \to \mathbf{x}_n = (x_1, x_2^*),$$

by drawing $x_2^* \sim \pi(x_2^*|x_1)$ from its conditional PDF.

# Gibbs and MH sampling

Since Gibbs sampling works on one component at a time, $x_1$ is the same in both the current and proposed states.

Therefore, the relevant terms in the Hastings ratio,

$$\frac{\pi_n \pi_{on}}{\pi_o \pi_{no}},$$

are:

$$\begin{aligned}
\pi_n &= \pi(x_1, x_2^*) \\
\pi_o &= \pi(x_1, x_2) \\
\pi_{no} &= \pi(x_2^* | x_1) \\
\pi_{on} &= \pi(x_2 | x_1)
\end{aligned}$$

# Gibbs and MH sampling

Thus,

$$\begin{aligned}
\frac{\pi_n \pi_{on}}{\pi_o \pi_{no}} &= \frac{\pi(x_1, x_2^*)\pi(x_2|x_1)}{\pi(x_1, x_2)\pi(x_2^*|x_1)} \\
&= \frac{\pi(x_1, x_2^*)}{\pi(x_2^*|x_1)} \frac{\pi(x_2|x_1)}{\pi(x_1, x_2)} \\
&= \pi(x_1) \frac{1}{\pi(x_1)} \\
&= 1,
\end{aligned}$$

where I have used the relationship between conditional, joint and marginal distributions.

Thus, Gibbs sampling is a special case of MH sampling, where the proposal is always accepted.

# Implications

Due to its relationship with MH, Gibbs sampling inherits all the "features" of MH sampling, including:

- burn-in
- correlation in samples
- block-averaging
- convergence diagnostics

The good news is that, we can use the same tools we developed for MH MCMC.

Let us reconsider an extra example we have seen before.

# Example: Sampling from a Uniform Disc

Suppose we want to generate samples $(x, y)$ distributed uniformly on a 2D disc of radius 1.

The joint PDF,

$$\pi(x, y) = \frac{1}{\pi}, \quad x^2 + y^2 \leq 1.$$

We have already seen this problem before. Let us discuss the following methods:

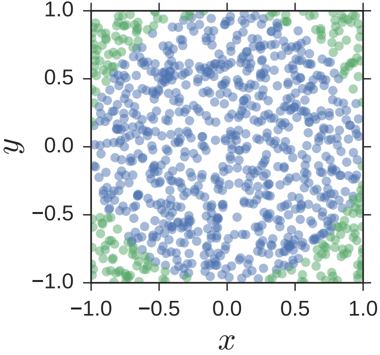
- accept-reject
- transformation method
- Metropolis MCMC
- Gibbs sampling

# Accept-Reject

Draw samples from independent 1D uniform distributions,

$$x \sim U[-1, 1]$$
$$y \sim U[-1, 1]$$

and retain only those that satisfy $x^2 + y^2 \leq 1$.




Disadvantages:

- points are rejected
- rejection worse in higher dimensions $(n)$
- Rejected fraction:

$$1 - \frac{\pi^{n/2}}{2^n \Gamma(\frac{n}{2} + 1)}$$

# Transformation Method

Map $(r,\, \theta)$ and $(x, y)$. Sample without rejection from

$$g(r, \theta) = R(r)T(\theta)$$

where,

$$R(r) = 2r, \quad 0 < r < 1$$
$$T(\theta) = \frac{1}{2\pi}, \quad 0 < \theta \leq 2\pi$$

Direct sample from $R(r)$ by setting $r \sim \sqrt{u}$, where $u \sim U[0, 1]$.

Disadvantage:

Not easy to construct simple maps in higher dimensions.

# MCMC

It is possible to write a Metropolis sampler to sample from the 2D joint distribution.

$$x_n \sim U[x_c - \delta, x_c + \delta]$$
$$y_n \sim U[y_c - \delta, y_c + \delta]$$

with a suitable $\delta$.

Since probability density is flat, most proposals are accepted.

Moves to points outside the disc are rejected.

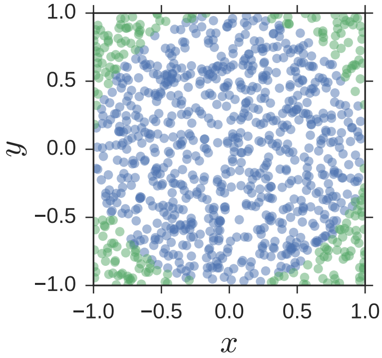
This property will help for higher dimensions, since we won't veer away too far from the "disc" or hypersphere.

The big disadvantage is that samples are now correlated!

# Gibbs Sampler

For this problem, it is easy to write down the conditional distributions:

$$\pi(x|y) = U[-\sqrt{1-y^2}, \sqrt{1-y^2}]$$
$$\pi(y|x) = U[-\sqrt{1-x^2}, \sqrt{1-x^2}]$$




Advantages:

- no rejection
- conditional easy to sample from
- extends to higher dimensions