Gibbs Sampling
Some Theory and Some Practice

Dominic Steinitz

25th June 2014

Outline

Outline

1. Introduction to Markov Chain Methods
   - The Grasshopper's Problems
2. Metropolis
   - Ergodic Theorem
   - Metropolis-Hastings

Outline

Outline

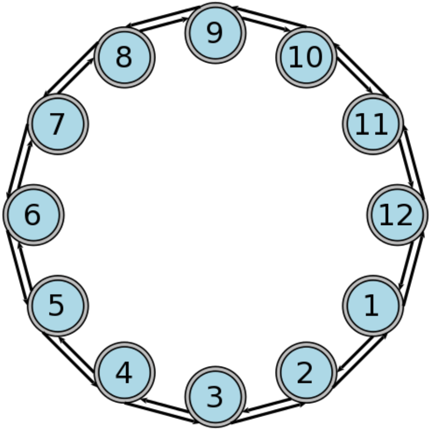
1. Introduction to Markov Chain Methods
   - The Grasshopper's Problems
2. Metropolis
   - Ergodic Theorem
   - Metropolis-Hastings
3. Gibbs Sampling
   - Random Scan
   - Systematic Scan
   - Bivariate Normal
4. Application in Statistics
   - Known Marginal

Outline

1. Introduction to Markov Chain Methods
   - The Grasshopper's Problems
2. Metropolis
   - Ergodic Theorem
   - Metropolis-Hastings
3. Gibbs Sampling
   - Random Scan
   - Systematic Scan
   - Bivariate Normal
4. Application in Statistics
   - Known Marginal
5. The Ising Model

Outline

Visit Each Hour Equally



Suppose Achieve Goal

$$\pi'(n) = p(n \mid n-1)\pi(n-1) + p(n \mid n+1)\pi(n+1)$$

Suppose Achieve Goal

$$\pi'(n) = p(n \mid n-1)\pi(n-1) + p(n \mid n+1)\pi(n+1)$$

$$\pi'(n) = \frac{1}{2}\frac{1}{N} + \frac{1}{2}\frac{1}{N} = \frac{1}{N}$$

Testing Convergence

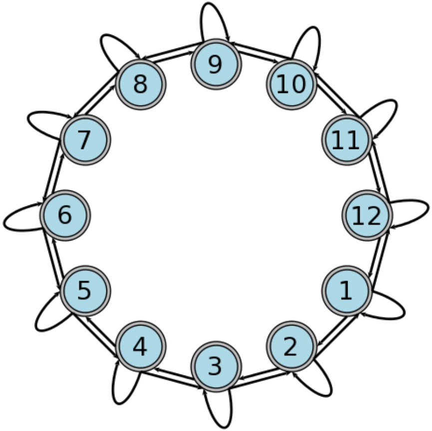
```
ghci> startOnOne
  (1><5)
   [ 1.0, 0.0, 0.0, 0.0, 0.0 ]

ghci> eqProbsMat
  (5><5)
   [ 0.0, 0.5, 0.0, 0.0, 0.5
   , 0.5, 0.0, 0.5, 0.0, 0.0
   , 0.0, 0.5, 0.0, 0.5, 0.0
   , 0.0, 0.0, 0.5, 0.0, 0.5
   , 0.5, 0.0, 0.0, 0.5, 0.0 ]
```

Testing Convergence

```
ghci> take 1 $
      drop 1000 $
      iterate (<> eqProbsMat) startOnOne
  [(1><5)
   [ 0.2, 0.2, 0.2, 0.2, 0.2 ]]
```

Visit Each Hour Proportionally



Suppose Achieve Goal

$$\pi'(n) = p(n \mid n-1)\pi(n-1) + p(n \mid n)\pi(n) + p(n \mid n+1)\pi(n+1)$$

Suppose Achieve Goal

$$\pi'(n) = p(n \mid n-1)\pi(n-1) + p(n \mid n)\pi(n) + p(n \mid n+1)\pi(n+1)$$

$$\begin{aligned}
\pi'(4) &= \frac{1}{2}\pi(3) + \frac{1}{2}\frac{1}{4}\pi(4) + \frac{1}{2}\frac{4}{5}\pi(5) \\
&= \frac{1}{2}\left(\frac{3}{N} + \frac{1}{4}\frac{4}{N} + \frac{4}{5}\frac{5}{N}\right) \\
&= \frac{1}{N}\frac{8}{2} \\
&= \frac{4}{N} \\
&= \pi(4)
\end{aligned}$$

Testing Convergence

```
ghci> incProbsMat
(5><5)
[ 0.00, 0.500, 0.000, 0.000, 0.5
, 0.25, 0.250, 0.500, 0.000, 0.0
, 0.00, 0.333, 0.167, 0.500, 0.0
, 0.00, 0.000, 0.375, 0.125, 0.5
, 0.10, 0.000, 0.000, 0.400, 0.5 ]

ghci> take 1 $
      drop 1000 $
      iterate (<> incProbsMat) startOnOne
[(1><5)
  [ 6.67e-2, 0.133, 0.199, 0.267, 0.333 ]]
```

Very Simple Markov Chain

$$q(i,j) = \begin{cases} \mathbb{P}(X_{n+1} = j \,|\, X_n = i) = \frac{1}{2} & \text{if } j = i+1 \mod N \\ \mathbb{P}(X_{n+1} = j \,|\, X_n = i) = \frac{1}{2} & \text{if } j = i-1 \mod N \\ \mathbb{P}(X_{n+1} = j \,|\, X_n = i) = 0 & \text{otherwise} \end{cases}$$

Simple Markov Chain

The grasshopper knows that $\pi(i) = i/N$ so it can calculate $\pi(j)/\pi(i) = j/i$ without knowing N.

Simple Markov Chain

The grasshopper knows that $\pi(i) = i/N$ so it can calculate $\pi(j)/\pi(i) = j/i$ without knowing N.

$$p(i,j) = \begin{cases} q(i,j)\left[\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)} \wedge 1\right] & \text{if } j \neq i \\ 1 - \sum_{k:k\neq i} q(i,k)\left[\frac{\pi(k)q(k,i)}{\pi(i)q(i,k)} \wedge 1\right] & \text{if } j = i \end{cases}$$

Simple Markov Chain

$$q(i,j) = \begin{cases} \frac{1}{2}(\frac{j}{i} \wedge 1) & \text{if } j \text{ is 1 step clockwise} \\ \frac{1}{2}(\frac{j}{i} \wedge 1) & \text{if } j \text{ is 1 step anti-clockwise} \\ 1 - \frac{1}{2}(\frac{j^c}{i} \wedge 1) - \frac{1}{2}(\frac{j^a}{i} \wedge 1) & j = i \& j^a, j^c \text{step (anti-)clockwise} \\ 0 & \text{otherwise} \end{cases}$$

Orientation

- Markov chain theory normally interested in when chain has stationary distribution.

Orientation

- Markov chain theory normally interested in when chain has stationary distribution.
- Here we have a distribution and want to create a Markov chain which has it as stationary distribution.

Ergodic Theorem

An irreducible, aperiodic and positive recurrent Markov chain has a unique stationary distribution.

Ergodic Theorem

An irreducible, aperiodic and positive recurrent Markov chain has a unique stationary distribution.

Roughly speaking

- Irreducible means it is possible to get from any state to any other state.

Ergodic Theorem

An irreducible, aperiodic and positive recurrent Markov chain has a unique stationary distribution.

Roughly speaking

- Irreducible means it is possible to get from any state to any other state.
- Aperiodic means that returning to a state having started at that state occurs at irregular times.

Ergodic Theorem

An irreducible, aperiodic and positive recurrent Markov chain has a unique stationary distribution.

Roughly speaking

- Irreducible means it is possible to get from any state to any other state.
- Aperiodic means that returning to a state having started at that state occurs at irregular times.
- Positive recurrent means that the expectation of first time to hit a state is finite (for every state and more pedantically except on sets of null measure).

Detailed Balance

A Markov chain $p(i,j)$ and a distribution π are said to be in **detailed balance** if

$$\pi(i)p(i,j) = \pi(j)p(j,i)$$

Detailed Balance

A Markov chain $p(i, j)$ and a distribution π are said to be in **detailed balance** if

$$\pi(i)p(i, j) = \pi(j)p(j, i)$$

Summing over i

$$\pi P = \pi$$

So clearly stationary.

Metropolis-Hastings

Let

- π be a probability distribution on the state space Ω with $\pi(i) > 0$ for all i
- (Q, π_0) be an ergodic Markov chain on Ω with transition probabilities $q(i, j) > 0$

The latter condition is slightly stronger than it need be but we will not need fully general conditions.

Metropolis-Hastings

Create a new (ergodic) Markov chain with transition probabilities

$$p_{ij} = \begin{cases} q(i,j)\left[\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)} \wedge 1\right] & \text{if } j \neq i \\ 1 - \sum_{k:k\neq i} q(i,k)\left[\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)} \wedge 1\right] & \text{if } j = i \end{cases}$$

Metropolis-Hastings is Ergodic

$$\begin{aligned}\pi(i)q(i,j)\left[\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)} \wedge 1\right] &= \pi(i)q(i,j) \wedge \pi(j)q(j,i) \\ &= \pi(j)q(j,i)\left[\frac{\pi(i)q(i,j)}{\pi(j)q(j,i)} \wedge 1\right]\end{aligned}$$

Secondly since we have specified that (Q, π_0) is ergodic then clearly (P, π_0) is also ergodic (all the transition probabilities are > 0). So we know the algorithm will converge to the unique distribution we specified to provide estimates of values of interest.

Random Scan

Start with $\pi(\theta_1, \theta_2)$.

$$\begin{cases} \text{Sample } \theta_1^{(i+1)} \sim \pi(\theta_1 \mid \theta_2^{(i)}) & \text{with probability } \frac{1}{2} \\ \text{Sample } \theta_2^{(i+1)} \sim \pi(\theta_2 \mid \theta_1^{(i)}) & \text{with probability } \frac{1}{2} \end{cases}$$

The transition density kernel is then given by

$$\begin{aligned} q(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\theta}^{(i)}) &= \frac{1}{2}\pi(\theta_1^{(i+1)} \mid \theta_2^{(i)})\delta(\theta_2^{(i)}, \theta_2^{(i+1)}) \\ &+ \frac{1}{2}\pi(\theta_2^{(i+1)} \mid \theta_1^{(i)})\delta(\theta_1^{(i)}, \theta_1^{(i+1)}) \end{aligned}$$

where δ is the Dirac delta function.

Detailed Balance

$$\pi(\theta_1, \theta_2)\left[\frac{1}{2}\pi(\theta_1' \mid \theta_2)\delta(\theta_2, \theta_2') + \frac{1}{2}\pi(\theta_2' \mid \theta_1)\delta(\theta_1, \theta_1')\right] =$$

$$\frac{1}{2}\left[\pi(\theta_1, \theta_2)\pi(\theta_1' \mid \theta_2)\delta(\theta_2, \theta_2') + \dots\right] =$$

$$\frac{1}{2}\left[\pi(\theta_1, \theta_2')\pi(\theta_1' \mid \theta_2)\delta(\theta_2, \theta_2') + \dots\right] =$$

$$\frac{1}{2}\left[\pi(\theta_2')\pi(\theta_1 \mid \theta_2')\pi(\theta_1' \mid \theta_2)\delta(\theta_2, \theta_2') + \dots\right] =$$

$$\frac{1}{2}\left[\pi(\theta_1', \theta_2')\pi(\theta_1 \mid \theta_2')\delta(\theta_2', \theta_2) + \dots\right] =$$

$$\pi(\theta_1', \theta_2')\left[\frac{1}{2}\pi(\theta_1 \mid \theta_2')\delta(\theta_2', \theta_2) + \frac{1}{2}\pi(\theta_2 \mid \theta_1')\delta(\theta_1', \theta_1)\right]$$

Detailed Balance

$$\pi(\boldsymbol{\theta})q(\boldsymbol{\theta'},\boldsymbol{\theta}) = \pi(\boldsymbol{\theta'})q(\boldsymbol{\theta},\boldsymbol{\theta'})$$

Hand waving slightly, we can see that this scheme satisfies the premises of the ergodic theorem and so we can conclude that there is a unique stationary distribution and π must be that distribution.

Systematic Scan

Most references on Gibbs sampling do not describe the random scan but instead something called a systematic scan.
Again for simplicity let us consider a model with two parameters.
In this sampler, we update the parameters in two steps.

$$\text{Sample } \theta_1^{(i+1)} \sim \pi(\theta_1 \mid \theta_2^{(i)})$$
$$\text{Sample } \theta_2^{(i+1)} \sim \pi(\theta_2 \mid \theta_1^{(i+1)})$$

Systematic Scan

- **Not** time-homegeneous!!!

Systematic Scan

- **Not** time-homegeneous!!!
- At each step the transition matrix flips between the two transition matrices given by the individual steps.

Systematic Scan

- **Not** time-homegeneous!!!
- At each step the transition matrix flips between the two transition matrices given by the individual steps.
- Cannot apply the ergodic theorem as it only applies to time-homogeneous processes.

Systematic Scan

- **Not** time-homegeneous!!!
- At each step the transition matrix flips between the two transition matrices given by the individual steps.
- Cannot apply the ergodic theorem as it only applies to time-homogeneous processes.
- ... does not satisfy detailed balance

An Example: The Bivariate Normal

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \Bigg| y \sim N \left[\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right]$$

The conditional distributions are easily calculated to be

$$\theta_1 \,|\, \theta_2, y \sim \mathcal{N}(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$
$$\theta_2 \,|\, \theta_1, y \sim \mathcal{N}(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$

Haskell

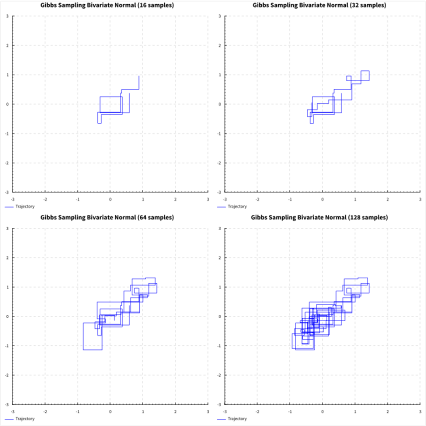
$$
\begin{aligned}
&\mathit{gibbsSampler} ::\\
&\quad \mathit{Double} \to\\
&\quad \mathit{RVarT}\ (\mathit{W.Writer}\ [(\mathit{Double}, \mathit{Double})])\ \mathit{Double}\\
&\mathit{gibbsSampler}\ \theta_2 = \mathbf{do}\\
&\quad \tilde{\theta}_1 \leftarrow \mathit{rvarT}\ (\mathit{Normal}\ (y_1 + \rho * (\theta_2 - y_2))\ (1 - \rho^2))\\
&\quad \mathit{lift}\ \$\ \mathit{W.tell}\ [(\tilde{\theta}_1, \theta_2)]\\
&\quad \tilde{\theta}_2 \leftarrow \mathit{rvarT}\ (\mathit{Normal}\ (y_2 + \rho * (\tilde{\theta}_1 - y_1))\ (1 - \rho^2))\\
&\quad \mathit{lift}\ \$\ \mathit{W.tell}\ [(\tilde{\theta}_1, \tilde{\theta}_2)]\\
&\quad \mathit{return}\ \$\ \tilde{\theta}_2
\end{aligned}
$$

Haskell

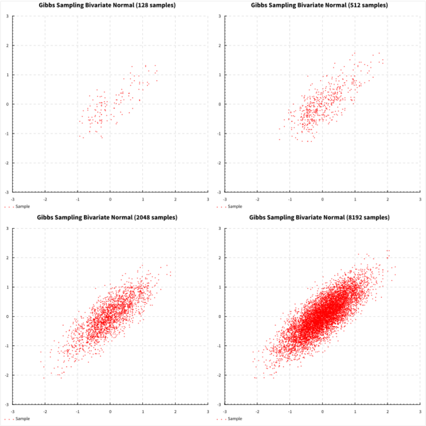
From which we can create an infinite stream of samples.

$$
\begin{aligned}
&\mathit{runMCMC} :: [(\mathit{Double}, \mathit{Double})] \\
&\mathit{runMCMC} = \\
&\quad \mathit{drop}\ \mathit{burnIn}\ \$ \\
&\quad \mathit{snd}\ \$ \\
&\quad \mathit{runWriter}\ \$ \\
&\quad \mathit{evalStateT}\ (\mathit{sample}\ (\mathit{iterateM\_}\ \mathit{gibbsSampler}\ 2.5)) \\
&\qquad (\mathit{pureMT}\ 2)
\end{aligned}
$$

Typical Paths



Typical Samples



Prior and Likelihood

$$\pi(\mu, \tau) \propto \frac{1}{\tau} \quad -\infty < \mu < \infty \text{ and } 0 < \tau < \infty$$

$$p(x \mid \mu, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} \right) \exp\left(- \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

Posterior

Re-writing in terms of precision

$$p(x \mid \mu, \tau) \propto \prod_{i=1}^{n} \sqrt{\tau} \exp\left(-\frac{\tau}{2}(x_i - \mu)^2 \right)$$

Thus the posterior is

$$p(\mu, \tau \mid x) \propto \tau^{n/2-1} \exp\left(-\frac{\tau}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

Conditionals

We can re-write the sum in terms of

- Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Conditionals

We can re-write the sum in terms of

- Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$\sum_{i=1}^{n}(x_i - \mu)^2 = (n-1)s^2 + n(\bar{x} - \mu)^2$$

Conditionals

Thus the conditional posterior for μ is

$$p(\mu \mid \tau, x) \propto \exp\left(-\frac{\tau}{2}\left(\nu s^2 + n(\bar{x} - \mu)^2\right)\right)$$
$$\propto \exp\left(-\frac{n\tau}{2}(\mu - \bar{x})^2\right)$$

which we recognise as a normal distribution with mean of $\bar{x}$ and a variance of $(n\tau)^{-1}$.

Conditionals

The conditional posterior for τ is

$$p(\tau \mid \mu, x) \propto \tau^{n/2-1} \exp\left(-\tau \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

which we recognise as a gamma distribution with a shape of $n/2$ and a scale of $\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2$

Marginal

In this particular case, we can calculate the marginal posterior of μ analytically.

$$p(\mu \,|\, x) \propto \left(1 + \frac{n(\mu - \bar{x})^2}{(n-1)s^2}\right)^{-n/2}$$

This is the non-standardized Student's t-distribution $t_{n-1}(\bar{x}, s^2/n)$.

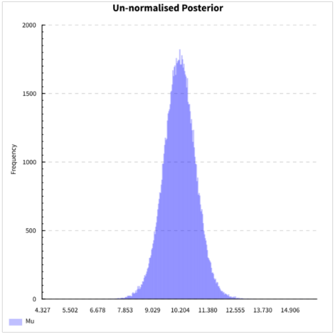
Haskell

$$
\begin{aligned}
&\mathit{gibbsSampler} :: \mathit{MonadRandom}\ m \Rightarrow \\
&\quad \mathit{Double} \to m\ (\mathit{Maybe}\ ((\mathit{Double}, \mathit{Double}), \mathit{Double})) \\
&\mathit{gibbsSampler}\ \tau = \mathbf{do} \\
&\quad \tilde{\mu} \leftarrow \mathit{sample}\ (\mathit{Normal}\ \bar{x}\ (\mathit{recip}\ (\mathit{sqrt}\ (n * \tau)))) \\
&\quad \mathbf{let}\ \mathit{shape} = 0.5 * n \\
&\qquad \mathit{scale} = 0.5 * (\sum x_i^2 + n * \tilde{\mu} \uparrow 2 - 2 * n * \tilde{\mu} * \bar{x}) \\
&\quad \tilde{\tau} \leftarrow \mathit{sample}\ (\mathit{Gamma}\ \mathit{shape}\ (\mathit{recip}\ \mathit{scale})) \\
&\quad \mathit{return}\ \$\ \mathit{Just}\ ((\tilde{\mu}, \tilde{\tau}), \tilde{\tau})
\end{aligned}
$$

From which we can create an infinite stream of samples.

$$
\begin{aligned}
&\mathit{gibbsSamples} :: [(\mathit{Double}, \mathit{Double})] \\
&\mathit{gibbsSamples} = \mathit{evalState}\ (\mathit{unfoldrM}\ \mathit{gibbsSampler}\ \tau_0) \\
&\quad (\mathit{pureMT}\ 1)
\end{aligned}
$$

Posterior for μ

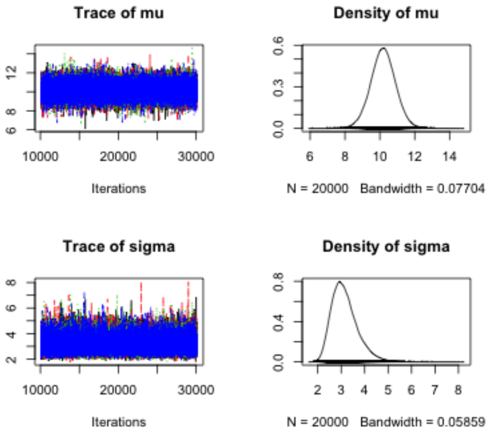


JAGS

JAGS is a mature, DSL for building Bayesian statistical models using Gibbs(?) sampling.

```
model {
for (i in 1:N) {
x[i] ~ dnorm(mu, tau)
}
mu ~ dnorm(0, 1.0E-6)
tau <- pow(sigma, -2)
sigma ~ dunif(0, 1000)
}
```

Output for JAGS

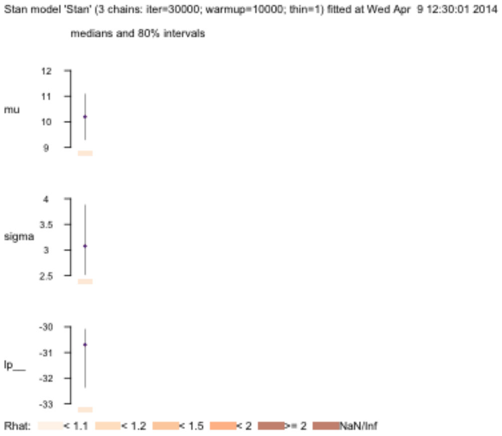


STAN

STAN: DSL similar to JAGS but newer, uses HMC, allows variable re-assignment, cannot really be described as declarative.

```
data {
  int<lower=0> N;
  real x[N];
}
parameters {
  real mu;
  real<lower=0,upper=1000> sigma;
}
model{
  x       ~ normal(mu, sigma);
  mu      ~ normal(0, 1000);
}
```

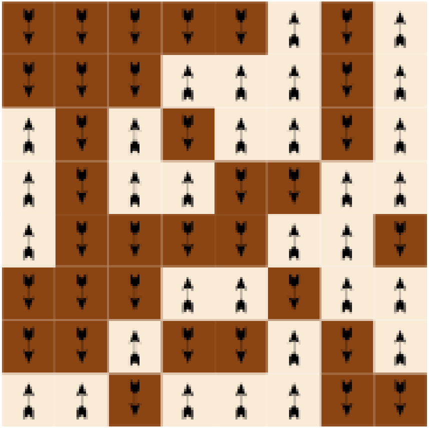
Output for STAN



Curie Temperature

https://www.youtube.com/watch?v=YzwGzJm41_o

Ising in a Nutshell



Boltzmann Distribution

To calculate the total magnetization, pick random configurations according to the Boltzmann distribution.

$$\mathbb{P}(\sigma) = \frac{\exp(-E(\sigma)/k_B T)}{Z(T)}$$

T temperature, k_B Boltzmann's constant, E energy.

$$E(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

$Z(T)$ normalizing constant, J a constant.

$$Z(T) = \sum_{\sigma} \exp(-E(\sigma)/k_B T)$$

Uniform Sampling

But what about the normalizing constant Z? Even for a modest grid size say 10×10, the number of states that needs to be summed over is extremely large $2^{10\times 10}$.

Uniform Sampling

But what about the normalizing constant Z? Even for a modest grid size say 10×10, the number of states that needs to be summed over is extremely large $2^{10\times 10}$.

Idea: could draw R random samples $(\sigma^{(i)})_{0\leq i<R}$ uniformly then use

$$Z_R \triangleq \sum_{0}^{R-1} \exp(-\beta\sigma_i)$$

Uniform Sampling

Now can estimate e.g. the magnetization

$$\langle M \rangle = \sum_{\sigma} M(\sigma) \frac{\exp(-\beta E(\sigma))}{Z(T)}$$

by

$$\widehat{\langle M \rangle} = \sum_{i=0}^{R-1} M(\sigma) \frac{exp(-\beta E(\sigma(i)))}{Z_R}$$

Problem

Statistical physics tells us that systems with large numbers of particles will occupy a small portion of the state space with any significant probability.

Problem

Statistical physics tells us that systems with large numbers of particles will occupy a small portion of the state space with any significant probability.
High dimensional distribution concentrated on small region of the state space: typical set T volume is given by $T \approx 2^H$ where H is the (Shannon) entropy.

$$H = -\sum_{\sigma} P(\sigma)\log_2(P(\sigma))$$

Problem

Statistical physics tells us that systems with large numbers of particles will occupy a small portion of the state space with any significant probability.
High dimensional distribution concentrated on small region of the state space: typical set T volume is given by $T \approx 2^H$ where H is the (Shannon) entropy.

$$H = -\sum_{\sigma} P(\sigma) \log_2(P(\sigma))$$

Actual value of the (mean) magnetization will determined by the values that M takes on T.

Problem

Uniform sampling will only give a good estimate if we make R large enough that we hit T at least a small number of times.

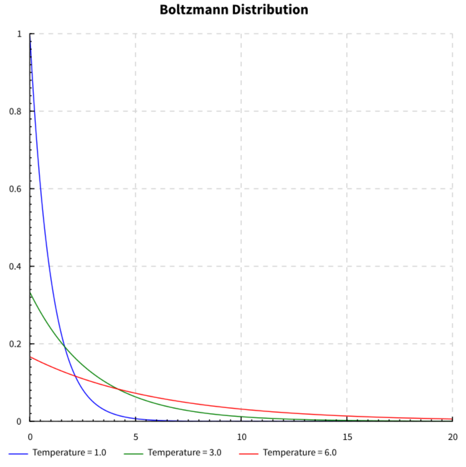
Problem

Uniform sampling will only give a good estimate if we make R large enough that we hit T at least a small number of times. The total size of the state space is 2^N and $T \approx 2^H$, so there is a probability of $2^H/2^N$ of hitting T.

Problem

Uniform sampling will only give a good estimate if we make R large enough that we hit T at least a small number of times.
The total size of the state space is 2^N and $T \approx 2^H$, so there is a probability of $2^H/2^N$ of hitting T.
Thus we need roughly 2^{N-H} samples to hit T.

Boltzmann Distribution



High Temperature

At high temperatures, the Boltzmann distribution flattens out so roughly all of the states have an equal likelihood of being occupied. We can calculate the (Shannon) entropy for this.

$$H \approx \sum_{\sigma} \frac{1}{2^N} \log_2 2^N = N$$

Low Temperature

- At "low" temperatures e.g. the temperature at which the phase transition occurs, $T = 2.269$, the entropy is approximately $0.13N$.

Low Temperature

- At "low" temperatures e.g. the temperature at which the phase transition occurs, $T = 2.269$, the entropy is approximately $0.13N$.
- So uniform sampling would require $\sim 2^{(N-N)}$ samples at high temperatures but $\sim 2^{(N-0.13N)} \approx 2^{N/2}$ at temperatures of interest. Even for our modest 10×10 grid this is $2^{50} \approx 10^{17}$ samples!

Low Temperature

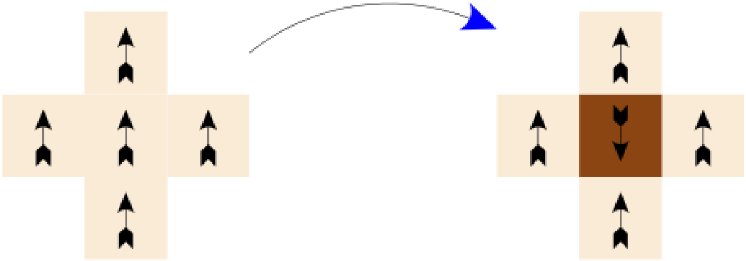
- At "low" temperatures e.g. the temperature at which the phase transition occurs, $T = 2.269$, the entropy is approximately $0.13N$.
- So uniform sampling would require $\sim 2^{(N-N)}$ samples at high temperatures but $\sim 2^{(N-0.13N)} \approx 2^{N/2}$ at temperatures of interest. Even for our modest 10×10 grid this is $2^{50} \approx 10^{17}$ samples!
- Enter Metropolis and his team: construct a Markov chain with a limiting distribution of the distribution required — does not require the evaluation of the partition function — samples high density areas with high probability (although theoretical results substantiating this latter point seem to be hard to come by).

Gibbs

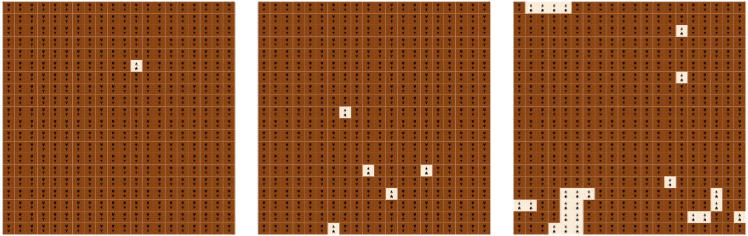
Use random scan and

$$p(\sigma_i = +1 \mid \sigma_{-i}) = \frac{\exp\left(J/k_b T \sum_{\langle i,j \rangle} \sigma_j\right)}{\exp\left(J/k_b T \sum_{\langle i,j \rangle} \sigma_j\right) + \exp\left(-J/k_b T \sum_{\langle i,j \rangle} \sigma_j\right)}$$

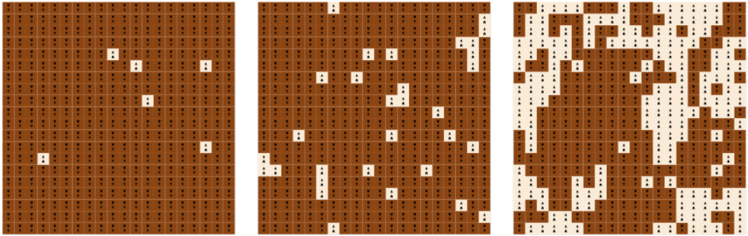
An Possible Flip



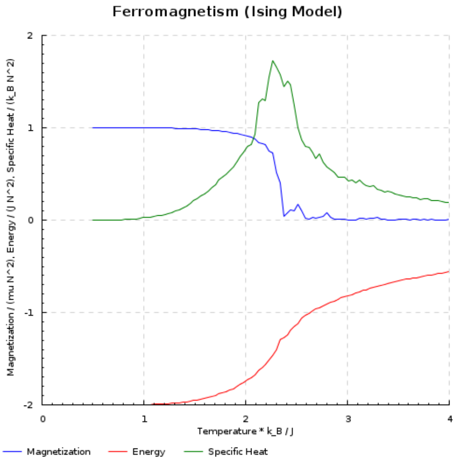
Evolution at $T = 2$, steps $= 100, 1000, 10{,}000$



Evolution at $T = 3$, steps $= 100, 1000, 10{,}000$



The Phase Transition Revealed



Acknowledgements

- http://education.mrsec.wisc.edu/463.htm
- http://www.inference.phy.cam.ac.uk/itila/book.html
- http://idontgetoutmuch.wordpress.com