Stats 318: Lecture # 9

Agenda:

- Motivation
- Basic ideas
- Metropolis algorithm
- Simulations of Gibbs distributions
- Examples

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)
  - Image and signal processing

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)
  - Image and signal processing
  - Combinatorial optimization

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)
  - Image and signal processing
  - Combinatorial optimization
  - Bioinformatics

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)
  - Image and signal processing
  - Combinatorial optimization
  - Bioinformatics
  - ...

Markov chain Monte Carlo (MCMC)

- Better name: simulation by Markov chains
- Widely used
  - Statistical physics
  - Statistics (Bayesian statistics)
  - Image and signal processing
  - Combinatorial optimization
  - Bioinformatics
  - ...
- Originates in physics in early 50's

Problem

- State space $\mathcal{X}$ with prob. dist. π

Problem

- State space $\mathcal{X}$ with prob. dist. π
- Wish to compute

$$\mathbb{E}_\pi f = \sum_x \pi(x) f(x)$$

Problem

- State space $\mathcal{X}$ with prob. dist. π
- Wish to compute

$$\mathbb{E}_\pi f = \sum_x \pi(x) f(x)$$

- How do we do this? e.g. $n = 1000$ & $|\mathcal{X}| = 2^n$

Problem

- State space $\mathcal{X}$ with prob. dist. π
- Wish to compute

$$\mathbb{E}_\pi f = \sum_x \pi(x) f(x)$$

- How do we do this? e.g. $n = 1000$ & $|\mathcal{X}| = 2^n$
- Monte Carlo sampling: Sample $X_1, \ldots, X_n$ i.i.d. π

$$\mathbb{E}_\pi f \approx \frac{1}{n} \sum_i f(X_i)$$

Problem

- State space $\mathcal{X}$ with prob. dist. π
- Wish to compute

$$\mathbb{E}_\pi f = \sum_x \pi(x) f(x)$$

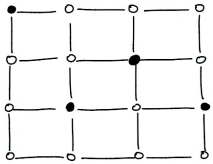
- How do we do this? e.g. $n = 1000$ & $|\mathcal{X}| = 2^n$
- Monte Carlo sampling: Sample $X_1, \ldots, X_n$ i.i.d. π

$$\mathbb{E}_\pi f \approx \frac{1}{n} \sum_i f(X_i)$$

- How do we sample from π?

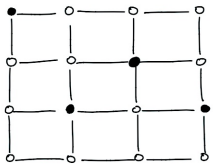
Example: Hard spheres model

- Graph $G = (V, E)$



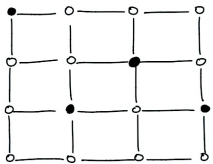
Example: Hard spheres model

- Graph $G = (V, E)$
- Configuration $x \in \{0, 1\}^{|V|}$ s.t. no two adjacent vertices both take on the value 1



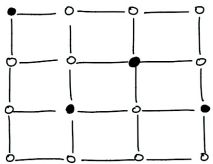
Example: Hard spheres model

- Graph $G = (V, E)$
- Configuration $x \in \{0, 1\}^{|V|}$ s.t. no two adjacent vertices both take on the value 1
- $\mathcal{X}$ is set of feasible configurations $\subset \{0, 1\}^{|V|}$



Example: Hard spheres model

- Graph $G = (V, E)$
- Configuration $x \in \{0, 1\}^{|V|}$ s.t. no two adjacent vertices both take on the value 1
- $\mathcal{X}$ is set of feasible configurations $\subset \{0, 1\}^{|V|}$

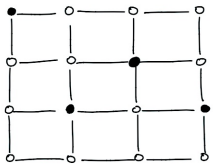


- Uniform dist. on $\mathcal{X}$

$$\pi(x) = \frac{1}{|\mathcal{X}|} \qquad \forall\ x \in \mathcal{X}$$

Example: Hard spheres model

- Graph $G = (V, E)$
- Configuration $x \in \{0, 1\}^{|V|}$ s.t. no two adjacent vertices both take on the value 1
- $\mathcal{X}$ is set of feasible configurations $\subset \{0, 1\}^{|V|}$



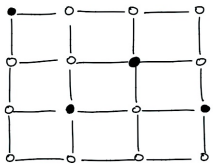
- Uniform dist. on $\mathcal{X}$

$$\pi(x) = \frac{1}{|\mathcal{X}|} \qquad \forall\, x \in \mathcal{X}$$

- Model introduced in statistical physics to capture behavior of a gas: particles have a non-negligible diameter and cannot overlap: $0/1 \leftrightarrow$ empty/occupied location

Example: Hard spheres model

- Graph $G = (V, E)$
- Configuration $x \in \{0,1\}^{|V|}$ s.t. no two adjacent vertices both take on the value 1
- $\mathcal{X}$ is set of feasible configurations $\subset \{0,1\}^{|V|}$



- Uniform dist. on $\mathcal{X}$

$$\pi(x) = \frac{1}{|\mathcal{X}|} \qquad \forall\, x \in \mathcal{X}$$

- Model introduced in statistical physics to capture behavior of a gas: particles have a non-negligible diameter and cannot overlap: $0/1 \leftrightarrow$ empty/occupied location

- Applications in telecommunications

Questions: Average number of ones?

$$f(x) = \sum_{v} x_v \qquad \sum_{x \in \mathcal{X}} \pi(x) f(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f(x) \ ?$$

Questions: Average number of ones?

$$f(x) = \sum_v x_v \qquad \sum_{x \in \mathcal{X}} \pi(x) f(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f(x) \ ?$$

- $|\mathcal{X}|$ exponential in $|V|$

Questions: Average number of ones?

$$f(x) = \sum_v x_v \qquad \sum_{x \in \mathcal{X}} \pi(x) f(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f(x) \ ?$$

- $|\mathcal{X}|$ exponential in $|V|$
- Computation of $|\mathcal{X}|$ is delicate

Questions: Average number of ones?

$$f(x) = \sum_v x_v \qquad \sum_{x \in \mathcal{X}} \pi(x) f(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f(x) \ ?$$

- $|\mathcal{X}|$ exponential in $|V|$
- Computation of $|\mathcal{X}|$ is delicate

Questions: Average number of ones?

$$f(x) = \sum_v x_v \qquad \sum_{x \in \mathcal{X}} \pi(x) f(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f(x) \ ?$$

- $|\mathcal{X}|$ exponential in $|V|$
- Computation of $|\mathcal{X}|$ is delicate

$\rightsquigarrow$ Resort to simulations

Questions: How do we sample from π?

Answers: Suppose we simulate a chain with transition P & unique invariant dist. π

Questions: How do we sample from π?

Answers: Suppose we simulate a chain with transition P & unique invariant dist. π

(1)

$$\frac{1}{T}\sum_{t=1}^{T} f(X_t) \underset{t\to\infty}{\longrightarrow} \quad \mathbb{E}_\pi f \quad \text{[Ergodic theorem]}$$

Questions: How do we sample from π?

Answers: Suppose we simulate a chain with transition P & unique invariant dist. π

(1)

$$\frac{1}{T}\sum_{t=1}^{T} f(X_t) \xrightarrow[t\to\infty]{} \mathbb{E}_\pi f \quad \text{[Ergodic theorem]}$$

(2)

$$X_t \xrightarrow[t\to\infty]{d} \pi \quad \text{[Convergence theorem]}$$

Questions: How do we sample from π?

Answers: Suppose we simulate a chain with transition P & unique invariant dist. π

(1)

$$\frac{1}{T}\sum_{t=1}^{T} f(X_t) \xrightarrow[t\to\infty]{} \mathbb{E}_\pi f \quad \text{[Ergodic theorem]}$$

(2)

$$X_t \xrightarrow[t\to\infty]{d} \pi \quad \text{[Convergence theorem]}$$

Questions: How do we sample from π?

Answers: Suppose we simulate a chain with transition P & unique invariant dist. π

(1)

$$\frac{1}{T}\sum_{t=1}^{T} f(X_t) \xrightarrow[t\to\infty]{} \mathbb{E}_\pi f \quad \text{[Ergodic theorem]}$$

(2)

$$X_t \xrightarrow[t\to\infty]{d} \pi \quad \text{[Convergence theorem]}$$

But how can this be any easier?

Metropolis algorithm

Key idea: Find a trans. matrix P s.t. P is reversible wrt π

Metropolis algorithm

Key idea: Find a trans. matrix P s.t. P is reversible wrt π

- Select transition matrix Q s.t.

$$Q(x, y) > 0 \quad \Leftrightarrow \quad Q(y, x) > 0 \qquad \forall\, x, y \in \mathcal{X}$$

Metropolis algorithm

Key idea: Find a trans. matrix P s.t. P is reversible wrt π

- Select transition matrix Q s.t.

$$Q(x,y) \;>\; 0 \quad \Leftrightarrow \quad Q(y,x) \;>\; 0 \qquad \forall\, x, y \in \mathcal{X}$$

- Select h: $\mathbb{R}_{++} \;\longrightarrow\;]\,0,1\,]$ obeying

$$h(u) = u\; h(1/u) \qquad \text{e.g.} \quad h(u) = u \;\wedge\; 1$$

$$h(u) = \frac{u}{u+1}$$

Metropolis algorithm

Key idea: Find a trans. matrix P s.t. P is reversible wrt π

- Select transition matrix Q s.t.

$$Q(x,y) \;>\; 0 \quad \Leftrightarrow \quad Q(y,x) \;>\; 0 \qquad \forall\, x,y \in \mathcal{X}$$

- Select h: $\mathbb{R}_{++} \;\longrightarrow\;]\,0,1\,]$ obeying

$$h(u) = u\; h(1/u) \qquad \text{e.g.} \quad h(u) = u \;\wedge\; 1$$

$$h(u) = \frac{u}{u+1}$$

- For $x \neq y$

$$R(x,y) = \begin{cases} h\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) & \text{if } Q(x,y) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Metropolis algorithm

Key idea: Find a trans. matrix P s.t. P is reversible wrt π

- Select transition matrix Q s.t.

$$Q(x,y) \;>\; 0 \quad \Leftrightarrow \quad Q(y,x) \;>\; 0 \qquad \forall\, x,y \in \mathcal{X}$$

- Select h: $\mathbb{R}_{++} \;\longrightarrow\;]\,0,1\,]$ obeying

$$h(u) = u\; h(1/u) \qquad \text{e.g.} \quad h(u) = u \;\wedge\; 1$$

$$h(u) = \frac{u}{u+1}$$

- For $x \neq y$

$$R(x,y) = \begin{cases} h\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) & \text{if } Q(x,y) > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Take $P(x,y) = Q(x,y)R(x,y) \qquad x \neq y$

$$P(x,x) = 1 - \sum_{y\,:\,y \neq x} P(x,y)$$

Interpretation

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy
to sample from $Q(x, \cdot)$ for all $x \in X$

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy
to sample from $Q(x, \cdot)$ for all $x \in X$

Algorithm

- Burn in: Initialize X_0

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy to sample from $Q(x, \cdot)$ for all $x \in X$

Algorithm

- Burn in: Initialize X_0
- Loop: for $t = 0, 1, 2, \ldots$

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy
to sample from $Q(x, \cdot)$ for all $x \in X$

Algorithm

- Burn in: Initialize X_0
- Loop: for $t = 0, 1, 2, \ldots$
  - Sample $Y \sim Q(X_t, \cdot)$

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$
2) Accept move wp $R(x, y)$

Select Q so that it is easy
to sample from $Q(x, \cdot)$ for all $x \in X$

Algorithm

- Burn in: Initialize X_0
- Loop: for $t = 0, 1, 2, \ldots$
  - Sample $Y \sim Q(X_t, \cdot)$
  - Sample $U \sim U[0, 1]$

Interpretation

1) Propose a state y sampled from $Q(x, \cdot)$

2) Accept move wp $R(x, y)$

Select Q so that it is easy to sample from $Q(x, \cdot)$ for all $x \in X$

Algorithm

- Burn in: Initialize X_0
- Loop: for $t = 0, 1, 2, \ldots$
  - Sample $Y \sim Q(X_t, \cdot)$
  - Sample $U \sim U[0, 1]$
  - If $U < R(X_t, Y)$, $X_{t+1} = Y$ [Accept]

    Otherwise $X_{t+1} = X_t$ [Stay put]

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Assume $Q(x,y) > 0$ (otherwise equality is trivial)

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Assume $Q(x,y) > 0$ (otherwise equality is trivial)

$$\pi(x)P(x,y) = \pi(x)Q(x,y)R(x,y) = \pi(x)Q(x,y)h\left[\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right]$$

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Assume $Q(x,y) > 0$ (otherwise equality is trivial)

$$\pi(x)P(x,y) = \pi(x)Q(x,y)R(x,y) = \pi(x)Q(x,y)h\left[\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right]$$

$$= \pi(y)Q(y,x)h\left[\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right]$$

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Assume $Q(x,y) > 0$ (otherwise equality is trivial)

$$\begin{aligned}\pi(x)P(x,y) = \pi(x)Q(x,y)R(x,y) &= \pi(x)Q(x,y)h\left[\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right] \\ &= \pi(y)Q(y,x)h\left[\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right] \\ &= \pi(y)Q(y,x)R(y,x)\end{aligned}$$

Convergence

Proposition

$\pi(x) > 0 \quad \forall\, x \in \mathcal{X}$

(i) P *reversible for* π

(ii) P *irreducible if* Q *irreducible*

(iii) $h(u) < 1 \Longrightarrow P$ *aperiodic*

Proof: (i) To show $\pi(x)P(x,y) = \pi(y)P(y,x)$

Assume $Q(x,y) > 0$ (otherwise equality is trivial)

$$\begin{aligned}
\pi(x)P(x,y) = \pi(x)Q(x,y)R(x,y) &= \pi(x)Q(x,y)h\left[\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right] \\
&= \pi(y)Q(y,x)h\left[\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right] \\
&= \pi(y)Q(y,x)R(y,x) \\
&= \pi(y)P(y,x)
\end{aligned}$$

(ii) $Q^t(x,y) > 0 \Longrightarrow P^t(x,y) > 0$

(ii) $Q^t(x,y) > 0 \Longrightarrow P^t(x,y) > 0$

Why? $P^t(x,y) \geq Q^t(x,y) \times \underbrace{\text{Prob. accept all moves}}_{>0} > 0$

(ii) $Q^t(x,y) > 0 \Longrightarrow P^t(x,y) > 0$

$$\text{Why?} \quad P^t(x,y) \geq Q^t(x,y) \times \underbrace{\text{Prob. accept all moves}}_{>0} > 0$$

(iii) Do it!

Choice of h

- Assume Q is symmetric: $Q(x, y) = Q(y, x)$

Choice of h

- Assume Q is symmetric: $Q(x, y) = Q(y, x)$
- Accept wp $h(\pi(y)/\pi(x))$

Choice of h

- Assume Q is symmetric: $Q(x, y) = Q(y, x)$
- Accept wp $h(\pi(y)/\pi(x))$
- Reversibility

$$\pi(x)\cancel{Q(x,y)}h(\pi(y)/\pi(x)) = \pi(y)\cancel{Q(y,x)}h(\pi(x)/\pi(y))$$
$$\Updownarrow$$
$$h(u) = u\ h(1/u)$$

Choice of h

- Assume Q is symmetric: $Q(x, y) = Q(y, x)$
- Accept wp $h(\pi(y)/\pi(x))$
- Reversibility

$$\pi(x)\cancel{Q(x,y)}h(\pi(y)/\pi(x)) = \pi(y)\cancel{Q(y,x)}h(\pi(x)/\pi(y))$$
$$\Updownarrow$$
$$h(u) = u\ h(1/u)$$

- Maximum acceptance rate

$$h(u) \leq u \quad \forall\ u \leq 1$$
$$h(u) \leq 1 \quad \forall\ u \geq 1$$

Choice of h

- Assume Q is symmetric: $Q(x, y) = Q(y, x)$
- Accept wp $h(\pi(y)/\pi(x))$
- Reversibility

$$\pi(x)\cancel{Q(x,y)}h(\pi(y)/\pi(x)) = \pi(y)\cancel{Q(y,x)}h(\pi(x)/\pi(y))$$
$$\Updownarrow$$
$$h(u) = u\ h(1/u)$$

- Maximum acceptance rate

  $h(u) \leq u \quad \forall\ u \leq 1$

  $h(u) \leq 1 \quad \forall\ u \geq 1$
- Maximum: $h(u) = \min(u, 1)$

Sampling for the model of hard spheres

1. Initialize X_0 (with all entries zero, say)
2. For $t = 1, 2, \ldots$
   - Pick a vertex v at random
   - If all neighbors are empty, toss a coin and put value of fair bit at v
   - Otherwise, do nothing

Sampling for the model of hard spheres

1. Initialize X_0 (with all entries zero, say)
2. For $t = 1, 2, \ldots$
   - Pick a vertex v at random
   - If all neighbors are empty, toss a coin and put value of fair bit at v
   - Otherwise, do nothing

The chain converges to π. Why?

Equivalent formulation (Metropolis form)

1. Initialize X_0 (with all entries zero, say)
2. For $t = 1, 2, \ldots$
   - Pick a vertex v at random and flip the bit
   - Accept move with probability

$$\begin{cases} 1/2 & \text{if } y \text{ is feasible} \\ 0 & \text{otherwise} \end{cases}$$

Equivalent formulation (Metropolis form)

1. Initialize X_0 (with all entries zero, say)
2. For $t = 1, 2, \ldots$
   - Pick a vertex v at random and flip the bit
   - Accept move with probability

$$\begin{cases} 1/2 & \text{if } y \text{ is feasible} \\ 0 & \text{otherwise} \end{cases}$$

This is a Metropolis algorithm with $h(u) = u/(1+u)$ and

$$Q(x,y) = \begin{cases} \frac{1}{|V|} & \text{if } x \sim y \text{ differ in one site} \\ 0 & \text{otherwise} \end{cases}$$

$$Q(x,y) = Q(y,x) \qquad \Longrightarrow \qquad R(x,y) = h(\pi(y)/\pi(x)) = \frac{1}{2} \times \mathbb{1}(\pi(y) > 0)$$

Equivalence

- x feasible

Equivalence

- x feasible
- $x \sim y$ and $x_v \neq y_v$

Equivalence

- x feasible
- $x \sim y$ and $x_v \neq y_v$
- $\alpha(x, v) = 1$ if neighbors of v are empty, 0 otherwise

Equivalence

- x feasible
- $x \sim y$ and $x_v \neq y_v$
- $\alpha(x, v) = 1$ if neighbors of v are empty, 0 otherwise
- First description

$$P(x, y) = \alpha(x, v) \times \frac{1}{n} \times \frac{1}{2}$$

Equivalence

- x feasible
- $x \sim y$ and $x_v \neq y_v$
- $\alpha(x, v) = 1$ if neighbors of v are empty, 0 otherwise
- First description

$$P(x, y) = \alpha(x, v) \times \frac{1}{n} \times \frac{1}{2}$$

- Second description

$$T(x, y) = Q(x, y)R(x, y) = \frac{1}{n} \times \frac{1}{2} \times \mathbb{1}(\pi(y) > 0)$$

Equivalence

- x feasible
- $x \sim y$ and $x_v \neq y_v$
- $\alpha(x, v) = 1$ if neighbors of v are empty, 0 otherwise
- First description

$$P(x, y) = \alpha(x, v) \times \frac{1}{n} \times \frac{1}{2}$$

- Second description

$$T(x, y) = Q(x, y)R(x, y) = \frac{1}{n} \times \frac{1}{2} \times \mathbb{1}(\pi(y) > 0)$$

- Equivalence

$$\mathbb{1}(\pi(x) > 0)\alpha(x, v) = \mathbb{1}(\pi(x) > 0)\mathbb{1}(\pi(y) > 0)\alpha(x, v) = \mathbb{1}(\pi(x) > 0)\mathbb{1}(\pi(y) > 0)$$

Variation

$h(u) = \min(u, 1)$

1. Initialize X_0 (with all entries zero, say)
2. For $t = 1, 2, \ldots$
   - Pick a vertex v at random and flip the bit
   - Accept move with probability

$$\begin{cases} 1 & \text{if } y \text{ is feasible} \\ 0 & \text{otherwise} \end{cases}$$

Numerical example

Hard spheres example (click link)

Simulations of Gibbs distributions

- Energy distribution: $E : \mathcal{X} \to \mathbb{R}$
- β: real-valued parameter—inverse temperature

Simulations of Gibbs distributions

- Energy distribution: $E : \mathcal{X} \to \mathbb{R}$
- β: real-valued parameter—inverse temperature

Gibbs distribution

$$\pi(x) = \frac{1}{Z} e^{-\beta E(x)}$$

Partition function

$$Z = \sum_{x \in S} e^{-\beta E(x)}$$

Simulations of Gibbs distributions

- Energy distribution: $E : \mathcal{X} \to \mathbb{R}$
- β: real-valued parameter—inverse temperature

Gibbs distribution

$$\pi(x) = \frac{1}{Z} e^{-\beta E(x)}$$

Partition function

$$Z = \sum_{x \in S} e^{-\beta E(x)}$$

Gibbs distribution characterizes a system in thermodynamical equilibrium at temperature

$$T = 1/\beta$$

Ising Model

Simplified model of interacting spins

- Spins located at vertices of a graph $G = (V, E)$
- Each spin x_v can take on two values: ± 1
- State space $\mathcal{V} = \{-1, 1\}^{|V|}$

Energy of a spin configuration

$$E(I) = -\frac{1}{2} \sum_{v \neq v'} J_{v,v'} x_v x_{v'} - \sum_v B_v x_v$$

- J is symmetric
- $J_{v,v'} = 0$ if v and v' are not adjacent
- $J_{v,v'} > 0$ if $v \sim v'$ (*ferromagnetic* interaction): spins tend to be aligned
- $J_{v,v'} < 0$ if $v \sim v'$ (*anti-ferromagnetic* interaction): spins tend to be misaligned
- B_v: strength of (electro)magnetic field at v

Problem

$$\pi(x) = \frac{1}{Z} e^{-\beta E(x)} \qquad E(x) = -\sum_{e=(v,v')\in E} J_e x_v x_{v'} - \sum_v B_v x_v$$

(edges counted only once)

Simulate a typical spin configuration

This is where a Metropolis algorithm is enormously useful

A Metropolis chain

- Q: pick a sign at random and flip sign
- $h(u) = \min(u, 1)$

A Metropolis chain

- Q: pick a sign at random and flip sign
- $h(u) = \min(u, 1)$

Transition matrix $P(x, y)$

1. Pick a site and flip the sign (new state is y)
2. Accept move with probability $\min(\pi(y)/\pi(x), 1)$

A Metropolis chain

- Q: pick a sign at random and flip sign
- $h(u) = \min(u, 1)$

Transition matrix $P(x, y)$

1. Pick a site and flip the sign (new state is y)
2. Accept move with probability $\min(\pi(y)/\pi(x), 1)$

Probability of accepting move is

$$\frac{e^{-\beta E(y)}}{e^{-\beta E(x)}} \wedge 1 = e^{-\beta(E(y)-E(x))} \wedge 1$$

Numerical example

Ising model example (click link)

Very important remark

To run the Metropolis algorithm we do not need to know the partition function or normalizing constant Z since all calculations involve the ratio

$$\frac{\pi(y)}{\pi(x)}$$

Normalizing constant cancels out!

Very important remark

To run the Metropolis algorithm we do not need to know the partition function or normalizing constant Z since all calculations involve the ratio

$$\frac{\pi(y)}{\pi(x)}$$

Normalizing constant cancels out!

Enormously important: how would we know Z?