Markov chain Monte Carlo

Geir Storvik

Geilo Winter school 2023

UiO : Universitetet i Oslo

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Outline

Markov chain Monte Carlo

Markov chain Monte Carlo

- Assume now simulating from $\pi(\boldsymbol{x})$ is difficult directly
  - $\pi(\cdot)$ complicated
  - $\boldsymbol{x}$ high-dimensional

Markov chain Monte Carlo

- Assume now simulating from $\pi(\boldsymbol{x})$ is difficult directly
  - $\pi(\cdot)$ complicated
  - $\boldsymbol{x}$ high-dimensional
- Markov chain Monte Carlo:
  - Generates $\{\boldsymbol{x}^{(t)}\}$ sequentially
  - Markov structure: $\boldsymbol{x}^{(t)} \sim P(\cdot|\boldsymbol{x}^{(t-1)})$

Markov chain Monte Carlo

- Assume now simulating from $\pi(\boldsymbol{x})$ is difficult directly
  - $\pi(\cdot)$ complicated
  - $\boldsymbol{x}$ high-dimensional
- Markov chain Monte Carlo:
  - Generates $\{\boldsymbol{x}^{(t)}\}$ sequentially
  - Markov structure: $\boldsymbol{x}^{(t)} \sim P(\cdot|\boldsymbol{x}^{(t-1)})$
- Aim now:
  - The distribution of $\boldsymbol{x}^{(t)}$ converges to $\pi(\cdot)$ as t increases
  - $\hat{\mu}_{MCMC} = N^{-1}\sum_{t=1}^{N} h(\boldsymbol{x}^{(t)})$ converges towards $\mu = E^{\pi}[h(\boldsymbol{x})]$ as t increases
- References:
  - Givens and Hoeting (2012): *Computational statistics*
  - Robert and Casella (1999): *Monte Carlo statistical methods*

Markov chain theory - discrete case

- Assume $\{X^{(t)}\}$ is a Markov chain where $X^{(t)}$ is a discrete random variable

$$\Pr(X^{(t)} = \boldsymbol{x}^* | X^{(t-1)} = x) = P(\boldsymbol{x}^* | \boldsymbol{x}) \qquad \text{transition probabilities}$$

Markov chain theory - discrete case

- Assume $\{X^{(t)}\}$ is a Markov chain where $X^{(t)}$ is a discrete random variable

$$\Pr(X^{(t)} = \boldsymbol{x}^* | X^{(t-1)} = x) = P(\boldsymbol{x}^* | \boldsymbol{x}) \qquad \text{transition probabilities}$$

- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any state infinitely often.
  - aperiodic: Does not go in cycles

Markov chain theory - discrete case

- Assume $\{X^{(t)}\}$ is a Markov chain where $X^{(t)}$ is a discrete random variable

$$\Pr(X^{(t)} = \boldsymbol{x}^*|X^{(t-1)} = x) = P(\boldsymbol{x}^*|\boldsymbol{x}) \qquad \text{transition probabilities}$$

- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any state infinitely often.
  - aperiodic: Does not go in cycles
- Then there exists a unique distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\to\infty} \Pr(X^{(T)} = \boldsymbol{x}^*|X^{(0)} = \boldsymbol{x}) = \pi(\boldsymbol{x}^*)$$

$$\lim_{T\to\infty} \hat{\mu}_{MCMC} = \mu = E^{\pi}[\boldsymbol{h}(X)] \qquad \hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T} h(\boldsymbol{x}^{(t)})$$

Markov chain theory - discrete case

- Assume $\{X^{(t)}\}$ is a Markov chain where $X^{(t)}$ is a discrete random variable

$$\Pr(X^{(t)} = \boldsymbol{x}^* | X^{(t-1)} = x) = P(\boldsymbol{x}^*|\boldsymbol{x}) \qquad \text{transition probabilities}$$

- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any state infinitely often.
  - aperiodic: Does not go in cycles
- Then there exists a unique distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\to\infty} \Pr(X^{(T)} = \boldsymbol{x}^* | X^{(0)} = \boldsymbol{x}) = \pi(\boldsymbol{x}^*)$$

$$\lim_{T\to\infty} \hat{\mu}_{MCMC} = \mu = E^{\pi}[\boldsymbol{h}(X)] \qquad \hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T} h(\boldsymbol{x}^{(t)})$$

- How to find $\pi(\cdot)$ (the stationary distribution): Solve

$$\pi(\boldsymbol{x}^*) = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x})$$

Markov chain theory - discrete case

- Assume $\{X^{(t)}\}$ is a Markov chain where $X^{(t)}$ is a discrete random variable

$$\Pr(X^{(t)} = \boldsymbol{x}^*|X^{(t-1)} = x) = P(\boldsymbol{x}^*|\boldsymbol{x}) \qquad \text{transition probabilities}$$

- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any state infinitely often.
  - aperiodic: Does not go in cycles
- Then there exists a unique distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\rightarrow\infty} \Pr(X^{(T)} = \boldsymbol{x}^*|X^{(0)} = \boldsymbol{x}) = \pi(\boldsymbol{x}^*)$$

$$\lim_{T\rightarrow\infty} \hat{\mu}_{MCMC} = \mu = E^{\pi}[\boldsymbol{h}(X)] \qquad \hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T} h(\boldsymbol{x}^{(t)})$$

- How to find $\pi(\cdot)$ (the stationary distribution): Solve

$$\pi(\boldsymbol{x}^*) = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x})$$

- Our situation: We have $\pi(\boldsymbol{x})$, want to find $P(\boldsymbol{x}^*|\boldsymbol{x})$
  - Note: Many possible $P(\boldsymbol{x}^*|\boldsymbol{x})$!

Markov chain theory - general setting

- Assume $\{\boldsymbol{x}^{(t)}\}$ is a Markov chain where $\boldsymbol{x}^{(t)} \in S$

$$\Pr(\boldsymbol{x}^{(t)} \in A | \boldsymbol{x}^{(t-1)} = \boldsymbol{x}) = P(\boldsymbol{x}, A) = \int_{\boldsymbol{x}^* \in A} P(\boldsymbol{x}^* | \boldsymbol{x}) d\boldsymbol{x}^*$$

giving the transition densities

Markov chain theory - general setting

- Assume $\{\boldsymbol{x}^{(t)}\}$ is a Markov chain where $\boldsymbol{x}^{(t)} \in S$

$$\Pr(\boldsymbol{x}^{(t)} \in A|\boldsymbol{x}^{(t-1)} = \boldsymbol{x}) = P(\boldsymbol{x}, A) = \int_{\boldsymbol{x}^* \in A} P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}^*$$

  giving the transition densities
- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any $A \subset S$ infinitely often.
  - aperiodic: Do not go in cycles

Markov chain theory - general setting

- Assume $\{\boldsymbol{x}^{(t)}\}$ is a Markov chain where $\boldsymbol{x}^{(t)} \in S$

$$\Pr(\boldsymbol{x}^{(t)} \in A | \boldsymbol{x}^{(t-1)} = \boldsymbol{x}) = P(\boldsymbol{x}, A) = \int_{\boldsymbol{x}^* \in A} P(\boldsymbol{x}^* | \boldsymbol{x}) d\boldsymbol{x}^*$$

  giving the transition densities
- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any $A \subset S$ infinitely often.
  - aperiodic: Do not go in cycles
- Then there exists a distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\to\infty} \Pr(\boldsymbol{x}^{(T)} \in A | \boldsymbol{x}^{(0)} = \boldsymbol{x}) = \int_A \pi(\boldsymbol{x}^*) d\boldsymbol{x}^*$$

$$\hat{\mu}_{MCMC} \to \mu$$

Markov chain theory - general setting

- Assume $\{\boldsymbol{x}^{(t)}\}$ is a Markov chain where $\boldsymbol{x}^{(t)} \in S$

$$\Pr(\boldsymbol{x}^{(t)} \in A|\boldsymbol{x}^{(t-1)} = \boldsymbol{x}) = P(\boldsymbol{x}, A) = \int_{\boldsymbol{x}^* \in A} P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}^*$$

  giving the transition densities
- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any $A \subset S$ infinitely often.
  - aperiodic: Do not go in cycles
- Then there exists a distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\rightarrow\infty} \Pr(\boldsymbol{x}^{(T)} \in A|\boldsymbol{x}^{(0)} = \boldsymbol{x}) = \int_A \pi(\boldsymbol{x}^*)d\boldsymbol{x}^*$$

$$\hat{\mu}_{MCMC} \rightarrow \mu$$

- How to find $p(\cdot)$ (the stationary distribution): Solve

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}$$

Markov chain theory - general setting

- Assume $\{\boldsymbol{x}^{(t)}\}$ is a Markov chain where $\boldsymbol{x}^{(t)} \in S$

$$\Pr(\boldsymbol{x}^{(t)} \in A|\boldsymbol{x}^{(t-1)} = \boldsymbol{x}) = P(\boldsymbol{x}, A) = \int_{\boldsymbol{x}^* \in A} P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}^*$$

  giving the transition densities
- Assume the chain is
  - irreducible: Possible to move from any $\boldsymbol{x}$ to any $\boldsymbol{x}^*$ in a finite number of steps
  - reccurent: The chain will visit any $A \subset S$ infinitely often.
  - aperiodic: Do not go in cycles
- Then there exists a distribution $\pi(\boldsymbol{x})$ such that

$$\lim_{T\to\infty} \Pr(\boldsymbol{x}^{(T)} \in A|\boldsymbol{x}^{(0)} = \boldsymbol{x}) = \int_A \pi(\boldsymbol{x}^*)d\boldsymbol{x}^*$$

$$\hat{\mu}_{MCMC} \to \mu$$

- How to find $p(\cdot)$ (the stationary distribution): Solve

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}$$

- Our situation: We have $\pi(\cdot)$, want to find $P(\boldsymbol{x}^*|\boldsymbol{x})$

Detailed balance

- The task: Find a transition probability/density $P(\boldsymbol{x}^*|\boldsymbol{x})$ satisfying

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) d\boldsymbol{x}$$

Can in general be a difficult criterion to check

Detailed balance

- The task: Find a transition probability/density $P(\boldsymbol{x}^*|\boldsymbol{x})$ satisfying

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) d\boldsymbol{x}$$

  Can in general be a difficult criterion to check
- Sufficient criterion:

$$\pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*) P(\boldsymbol{x}|\boldsymbol{x}^*) \quad \text{Detailed balance}$$

Detailed balance

- The task: Find a transition probability/density $P(\boldsymbol{x}^*|\boldsymbol{x})$ satisfying

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x}$$

  Can in general be a difficult criterion to check
- Sufficient criterion:

$$\pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*)P(\boldsymbol{x}|\boldsymbol{x}^*) \quad \text{Detailed balance}$$

  We then have

$$\begin{aligned}\int_{\boldsymbol{x}} \pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x})d\boldsymbol{x} =& \int_{\boldsymbol{x}} \pi(\boldsymbol{x}^*)P(\boldsymbol{x}|\boldsymbol{x}^*)d\boldsymbol{x}\\ =&\pi(\boldsymbol{x}^*)\int_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{x}^*)d\boldsymbol{x} = \pi(\boldsymbol{x}^*)\end{aligned}$$

  since $P(\boldsymbol{x}|\boldsymbol{x}^*)$ is, for any given $\boldsymbol{x}^*$, a density wrt $\boldsymbol{x}$.

Detailed balance

- The task: Find a transition probability/density $P(\boldsymbol{x}^*|\boldsymbol{x})$ satisfying

$$\pi(\boldsymbol{x}^*) = \int_{\boldsymbol{x}} \pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) d\boldsymbol{x}$$

  Can in general be a difficult criterion to check
- Sufficient criterion:

$$\pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*) P(\boldsymbol{x}|\boldsymbol{x}^*) \quad \text{Detailed balance}$$

  We then have

$$\begin{aligned}\int_{\boldsymbol{x}} \pi(\boldsymbol{x}) P(\boldsymbol{x}^*|\boldsymbol{x}) d\boldsymbol{x} = &\int_{\boldsymbol{x}} \pi(\boldsymbol{x}^*) P(\boldsymbol{x}|\boldsymbol{x}^*) d\boldsymbol{x} \\ =&\pi(\boldsymbol{x}^*) \int_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{x}^*) d\boldsymbol{x} = \pi(\boldsymbol{x}^*)\end{aligned}$$

  since $P(\boldsymbol{x}|\boldsymbol{x}^*)$ is, for any given $\boldsymbol{x}^*$, a density wrt $\boldsymbol{x}$.
- Question: What would be the ideal choice of $P(\boldsymbol{x}^*|\boldsymbol{x})$?
- Question: Why only necessary to check for $\boldsymbol{x}^* \neq \boldsymbol{x}$.
- Question: Why do you think it is called detailed ballance?

Metropolis-Hastings algorithms

Metropolis-Hastings algorithms

- $P(\boldsymbol{x}^*|\boldsymbol{x})$ defined through an algorithm:
  1. Sample a candidate value $\boldsymbol{x}^*$ from a proposal distribution $g(\cdot|\boldsymbol{x})$.
  2. Compute the Metropolis-Hastings ratio

$$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}$$

  3. Put

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{x}^* & \text{with probability } \min\{1, R(\boldsymbol{x}, \boldsymbol{x}^*)\} \\ \boldsymbol{x} & \text{otherwise} \end{cases}$$

Metropolis-Hastings algorithms

- $P(\boldsymbol{x}^*|\boldsymbol{x})$ defined through an algorithm:
  1. Sample a candidate value $\boldsymbol{x}^*$ from a proposal distribution $g(\cdot|\boldsymbol{x})$.
  2. Compute the Metropolis-Hastings ratio
  $$R(\boldsymbol{x},\boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}$$
  3. Put
  $$\boldsymbol{Y} = \begin{cases} \boldsymbol{x}^* & \text{with probability } \min\{1, R(\boldsymbol{x},\boldsymbol{x}^*)\} \\ \boldsymbol{x} & \text{otherwise} \end{cases}$$
- For $\boldsymbol{x}^* \neq \boldsymbol{x}$:
$$P(\boldsymbol{x}^*|\boldsymbol{x}) = g(\boldsymbol{x}^*|\boldsymbol{x}) \min\left\{1, \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}\right\}$$
- Note: $P(\boldsymbol{x}|\boldsymbol{x})$ somewhat difficult to evaluate in this case.
- Detailed balance fulfilled?
$$\begin{aligned}\pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x}) =& \pi(\boldsymbol{x})g(\boldsymbol{x}^*|x) \min\left\{1, \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}\right\} \\ =& \min\{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x}), \pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)\} \\ =& \pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*) \min\left\{\frac{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}, 1\right\} = \pi(\boldsymbol{x}^*)P(\boldsymbol{x}|\boldsymbol{x}^*)\end{aligned}$$

Random walk chains

- Popular choice of proposal distribution:

$$\boldsymbol{x}^* = \boldsymbol{x} + \boldsymbol{\varepsilon}$$

- $g(\boldsymbol{x}^*|\boldsymbol{x}) = h(\boldsymbol{x}^* - \boldsymbol{x})$
- Popular choices: Uniform, Gaussian, t-distribution
- Note: If $g(\cdot)$ is symmetric, $g(\boldsymbol{x}^*|\boldsymbol{x}) = g(\boldsymbol{x}|\boldsymbol{x}^*)$ and

$$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{\pi(\boldsymbol{x}^*)}{\pi(\boldsymbol{x})}$$

- Results in the original Metropolis algorithm (Metropolis et al., 1953)
- Extended by Hastings (1970) for non-symmetric $(\cdot)$ to the now-called Metropolis-Hastings algorithm

Independence chains

- Assume $g(\boldsymbol{x}^*|\boldsymbol{x}) = g(\boldsymbol{x}^*)$. Then

$$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x})}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*)} = \frac{\frac{\pi(\boldsymbol{x}^*)}{g(\boldsymbol{x}^*)}}{\frac{\pi(\boldsymbol{x})}{g(\boldsymbol{x})}},$$

  fraction of importance weights!
- Behave very much like importance sampling and SIR
- Difficult to specify $g(\boldsymbol{x})$ for high-dimensional problems
- Theoretical properties easier to evaluate than for random walk versions.

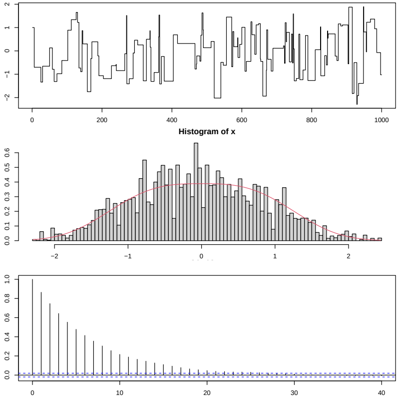
Question: For independence chains, acceptance rates should be as large as possible, for random walk it shou be in the range [0.2,0.7]. Why this difference?

Example

- Assume $p(x) \propto \exp(-|x|^3/3)$
- Proposal distribution $N(x, \sigma_p^2)$
- `Example_MH_cubic.R`
- $N = 10\,000$ samples, varying values of σ_p^2

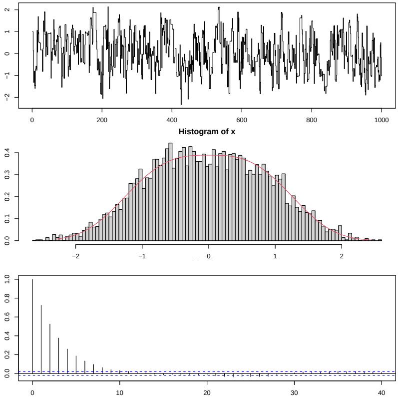
Example - $\sigma_p = 10$

Acceptance rate = 0.13



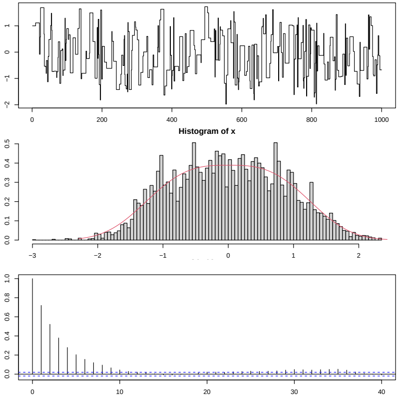
Example - $\sigma_p = 1$

Acceptance rate = 0.72



Example - $\sigma_p = 4$

Acceptance rate = 0.26



M-H and unknown constant

- Assume now $\pi(\boldsymbol{x}) = c \cdot q(\boldsymbol{x})$ with c unknown.

$$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{c \cdot q(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{c \cdot q(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{q(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{q(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})}$$

- Do not depend on c!

M-H and multivariate settings

- $\boldsymbol{x} = (X_1, ..., X_p)$
- Typical in this case: Only change one or a few components at a time.
  1. Choose index j (randomly)
  2. Sample $X_j^* \sim g_j(\cdot|\boldsymbol{x})$, put $X_k^* = X_k$ for $k \neq j$
  3. Compute

$$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{\pi(x_j^*|\boldsymbol{x}_{-j})g_j(x_j|\boldsymbol{x}^*)}{\pi(x_j|\boldsymbol{x}_{-j})g_j(x_j^*|\boldsymbol{x})}$$

  4. Put

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{x}^* & \text{with probability } \min\{1, R(\boldsymbol{x}, \boldsymbol{x}^*)\} \\ \boldsymbol{x} & \text{otherwise} \end{cases}$$

M-H and multivariate settings

- $\boldsymbol{x} = (X_1, ..., X_p)$
- Typical in this case: Only change one or a few components at a time.
  1. Choose index j (randomly)
  2. Sample $X_j^* \sim g_j(\cdot|\boldsymbol{x})$, put $X_k^* = X_k$ for $k \neq j$
  3. Compute
  $$R(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\pi(\boldsymbol{x}^*)g(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})g(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{\pi(x_j^*|\boldsymbol{x}_{-j})g_j(x_j|\boldsymbol{x}^*)}{\pi(x_j|\boldsymbol{x}_{-j})g_j(x_j^*|\boldsymbol{x})}$$
  4. Put
  $$\boldsymbol{Y} = \begin{cases} \boldsymbol{x}^* & \text{with probability } \min\{1, R(\boldsymbol{x}, \boldsymbol{x}^*)\} \\ \boldsymbol{x} & \text{otherwise} \end{cases}$$
- Can show that this version also satisfies detailed balance
- Can even go through indexes systematic
  - Should then consider the whole loop through all components as one iteration

Example

- Assume $\pi(\boldsymbol{x}) \propto \exp(-||\boldsymbol{x}||^3/3) = \exp(-[||\boldsymbol{x}||^2]^{3/2}/3) =$
- Algorithm given current value $\boldsymbol{x}^t$:

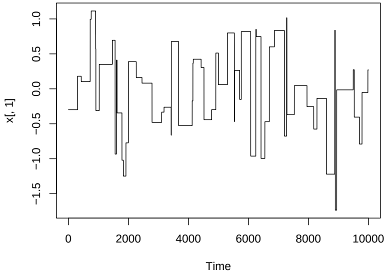
Sample $j \in \{1, ..., p\}$
Put

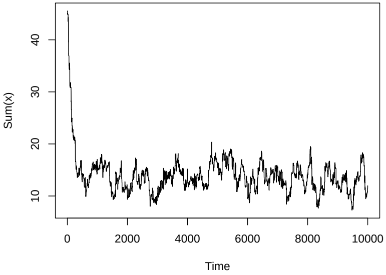
$$x_k^* = \begin{cases} x_k^t & \text{if } k \neq j \\ N(x_k^t, \sigma_k^2) & \text{if } k = j \end{cases}$$

Calculate $R(\boldsymbol{x}^t, \boldsymbol{x}^*) = \pi(\boldsymbol{x}^*)/\pi(\boldsymbol{x}^t))$
Generate $u \sim \text{Uniform}[0, 1]$
if $u < R(\boldsymbol{x}, \boldsymbol{x}^*)$ **then**
    Put $\boldsymbol{x}^{t+1} = \boldsymbol{x}^*$
else
    Put $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t$
end if

`Example_MH_cubic_multivariate.R`

Results multivariate cubic





Acceptance rate: 0.30

Gibbs sampling

Gibbs sampling

- Assume $\boldsymbol{x} = (X_1, \ldots, X_p)$
- Aim: Simulate $\boldsymbol{x} \sim \pi(\boldsymbol{x})$

Gibbs sampling

- Assume $\boldsymbol{x} = (X_1, \ldots, X_p)$
- Aim: Simulate $\boldsymbol{x} \sim \pi(\boldsymbol{x})$
- Gibbs sampling:
  1. Select starting values $\boldsymbol{x}^{(0)}$ and set $t = 0$
  2. Generate, in turn

$$
\begin{aligned}
X_1^{(t+1)} &\sim p(x_1 | x_2^{(t)}, x_3^{(t)}, \ldots, x_p^{(t)}) \\
X_2^{(t+1)} &\sim p(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)}) \\
&\vdots \\
X_{p-1}^{(t+1)} &\sim p(x_{p-1} | x_1^{(t+1)}, \ldots, x_{p-2}^{(t+1)}, x_p^{(t)}) \\
X_p^{(t+1)} &\sim p(x_p | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})
\end{aligned}
$$

Gibbs sampling

- Assume $\boldsymbol{x} = (X_1, \ldots, X_p)$
- Aim: Simulate $\boldsymbol{x} \sim \pi(\boldsymbol{x})$
- Gibbs sampling:
  1. Select starting values $\boldsymbol{x}^{(0)}$ and set $t = 0$
  2. Generate, in turn

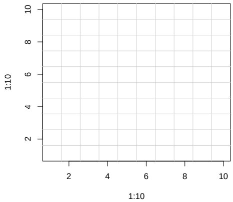
$$
\begin{aligned}
X_1^{(t+1)} &\sim p(x_1 | x_2^{(t)}, x_3^{(t)}, \ldots, x_p^{(t)}) \\
X_2^{(t+1)} &\sim p(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)}) \\
&\vdots \\
X_{p-1}^{(t+1)} &\sim p(x_{p-1} | x_1^{(t+1)}, \ldots, x_{p-2}^{(t+1)}, x_p^{(t)}) \\
X_p^{(t+1)} &\sim p(x_p | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})
\end{aligned}
$$

  3. Increment t and go to step 2.
- Completion of step 2 is called a cycle

Example - Ising model

- Ising model: $x_i \in \{-1, 1\}$, i lattice

$$p(\boldsymbol{x}) = \frac{1}{Z} e^{\beta \sum_{i \sim j} I(x_i = x_j)}$$

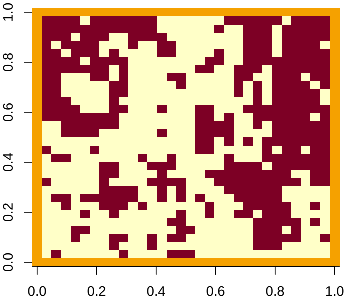


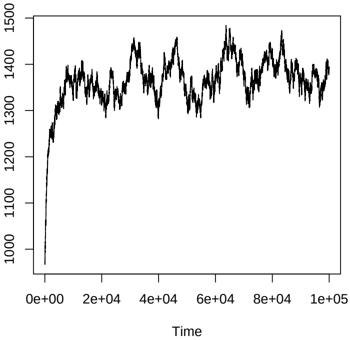
- Conditional distributions:

$$\Pr(x_i = 1 | \boldsymbol{x}_{-i}) = \frac{e^{-\beta \sum_{j:j \sim i} I(x_j = 1)}}{e^{-\beta \sum_{j:j \sim i} I(x_j = 1)} + e^{-\beta \sum_{j:j \sim i} I(x_j = -1)}}$$

- Simulations: $\beta = 0.7$, $ss = \sum_{i \sim j} I(x_i = x_j)$
- R-script: `Ising.R`

Example - Ising model





Properties of Gibbs sampler (random scan)

- Gibbs sampling (random scan):
  1. Select starting values $\boldsymbol{x}^{(0)}$ and set $t = 0$
  2. Sample $j \sim \text{Uniform}\{1, \ldots, p\}$
  3. Sample $X_j^{(t+1)} \sim p(x_j | \boldsymbol{x}_{-j}^{(t)})$
  4. Put $X_k^{(t+1)} = X_k^{(t)}$ for $k \neq j$

Properties of Gibbs sampler (random scan)

- Gibbs sampling (random scan):
  1. Select starting values $\boldsymbol{x}^{(0)}$ and set $t = 0$
  2. Sample $j \sim \text{Uniform}\{1, \ldots, p\}$
  3. Sample $X_j^{(t+1)} \sim p(x_j | \boldsymbol{x}_{-j}^{(t)})$
  4. Put $X_k^{(t+1)} = X_k^{(t)}$ for $k \neq j$
- The chain $\{\boldsymbol{x}^{(t)}\}$ is Markov
- Detailed balance:
  - Consider $\boldsymbol{x}, \boldsymbol{x}^*$ where $x_j \neq x_j^*$ while $x_k = x_k^*$ for $k \neq j$

Properties of Gibbs sampler (random scan)

- Gibbs sampling (random scan):
  1. Select starting values $\boldsymbol{x}^{(0)}$ and set $t = 0$
  2. Sample $j \sim \text{Uniform}\{1, \ldots, p\}$
  3. Sample $X_j^{(t+1)} \sim p(x_j|\boldsymbol{x}_{-j}^{(t)})$
  4. Put $X_k^{(t+1)} = X_k^{(t)}$ for $k \neq j$
- The chain $\{\boldsymbol{x}^{(t)}\}$ is Markov
- Detailed balance:
  - Consider $\boldsymbol{x}, \boldsymbol{x}^*$ where $x_j \neq x_j^*$ while $x_k = x_k^*$ for $k \neq j$

$$
\begin{aligned}
\pi(\boldsymbol{x})P(\boldsymbol{x}^*|\boldsymbol{x}) =& \pi(\boldsymbol{x})p^{-1}p(x_j^*|\boldsymbol{x}_{-j}) \\
=& \pi(\boldsymbol{x}_{-j})p(x_j|\boldsymbol{x}_{-j})p^{-1}p(x_j^*|\boldsymbol{x}_{-j}) \\
=& \pi(\boldsymbol{x}_{-j}^*)p(x_j|\boldsymbol{x}_{-j}^*)p^{-1}p(x_j^*|\boldsymbol{x}_{-j}^*) \\
=& \pi(\boldsymbol{x}^*)p^{-1}p(x_j|\boldsymbol{x}_{-j}^*) \\
=& \pi(\boldsymbol{x}^*)P(\boldsymbol{x}|\boldsymbol{x}^*)
\end{aligned}
$$

Tuning the Gibbs sampler

- Random or deterministic scan?
  - Deterministic scan most common (?)
  - When high correlation, random scan can be more efficient

Tuning the Gibbs sampler

- Random or deterministic scan?
  - Deterministic scan most common (?)
  - When high correlation, random scan can be more efficient
- Blocking:
  - When dividing $\boldsymbol{x} = (X_1, ..., X_p)$, each X_j can be vectors
  - Making each X_j as large as possible will typically improve convergence
  - Especially beneficial when high correlation between single components

Tuning the Gibbs sampler

- Random or deterministic scan?
  - Deterministic scan most common (?)
  - When high correlation, random scan can be more efficient
- Blocking:
  - When dividing $\boldsymbol{x} = (X_1, \ldots, X_p)$, each X_j can be vectors
  - Making each X_j as large as possible will typically improve convergence
  - Especially beneficial when high correlation between single components
- Hybrid Gibbs sampling
  - If $p(x_j|\boldsymbol{x}_{-j})$ is difficult to sample from, use an Metropolis-Hastings step for this component
  - Example ($p = 5$)
    1. Sample $X_1^{(t+1)} \sim p(x_1|\boldsymbol{x}_{-1}^{(t)})$
    2. Sample $(X_2^{(t+1)}, X_3^{(t+1)})$ through an M-H step
    3. Sample $X_4^{(t+1)}$ through another M.H step
    4. Sample $X_5^{(t+1)} \sim p(x_5|\boldsymbol{x}_{-5}^{(t+1)})$

Lemmings data

- Model

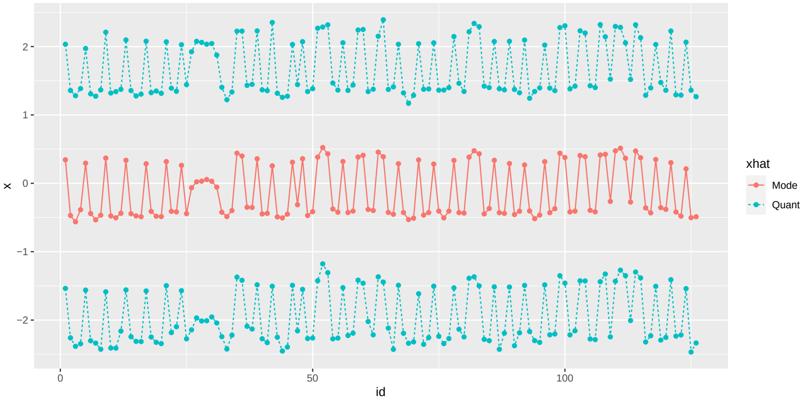
$$\begin{aligned} \boldsymbol{y}_t &\sim \text{Binom}\left(1, \frac{\exp(\boldsymbol{x}_t)}{1+\exp(\boldsymbol{x}_t)}\right) \\ \boldsymbol{x}_t &= a\boldsymbol{x}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \\ a &\sim \text{Uniform}[0, 1] \end{aligned}$$

- Simulation of a given $\boldsymbol{x}$:

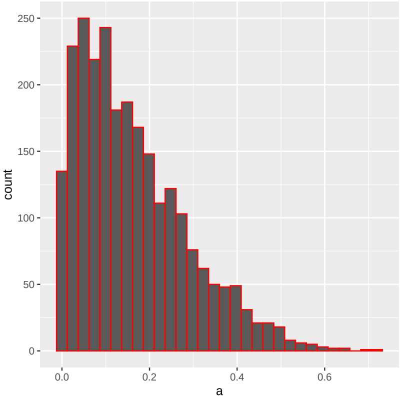
$$a|\boldsymbol{x} \sim N(\frac{\sum_{i=2}^{T} x_i x_{i-1}}{\sum_{i=2}^{T} x_{i-1^2}} I(0 < a < 1)$$

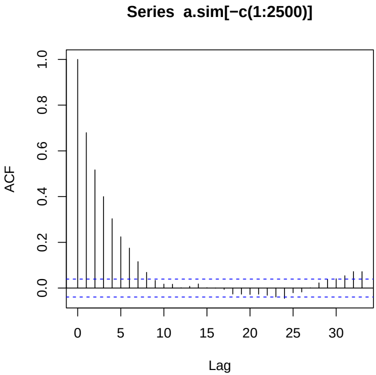
- Simulaton of x_i:
  1. Simulate $x_i^*|\boldsymbol{x}_{-i} = N\left(\frac{a(x_{i-1}+x_{i+1})}{1+a2}, \sigma\frac{1-a^2}{1-a^4}\right)$
- Calculate $r = \frac{p(y_i|x_i^*)}{p(y_i|x_i)}$
- Accept x_i^* with probability $\min\{1, r\}$

Lemmings - results



Lemmings - results





G. H. Givens and J. A. Hoeting. *Computational statistics*, volume 710. John Wiley & Sons, 2012.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

C. P. Robert and G. Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.