Simulation - Lectures 7 - MCMC: Metropolis Hastings

Lecture version: Sunday 1st March, 2020, 15:36

Robert Davies

Part A Simulation and Statistical Programming

Hilary Term 2020

Metropolis-Hastings Algorithm

- The Metropolis-Hastings (MH) algorithm allows to simulate a Markov Chain with any given stationary distribution.
- We will start with simulation of random variable X on a discrete state space.
- Let $p(x) = \tilde{p}(x)/Z_p$ be the pmf on Ω. We will call p the (pmf of the) target distribution.
- To simplify notations, we assume that $p(x) > 0$ for all $x \in \Omega$
- Choose a 'proposal' transition matrix $q(y|x)$. We will use the notation $Y \sim q(\cdot|x)$ to mean $\Pr(Y = y|X = x) = q(y|x)$.

Metropolis-Hastings Algorithm

Metropolis-Hastings algorithm

1. Either set $X_0 = x_0$, or draw X_0 from some initial distribution
2. For $t = 1, 2, \dots, n-1$:
   2.1 Assume $X_{t-1} = x_{t-1}$.
   2.2 Simulate $Y_t \sim q(\cdot|x_{t-1})$ and $U_t \sim \mathrm{Unif}[0,1]$.
   2.3 If

   $$U_t \leq \alpha(Y_t|x_{t-1})$$

   where

   $$\alpha(y|x) = \min\left\{1, \frac{\tilde{p}(y)q(x|y)}{\tilde{p}(x)q(y|x)}\right\}$$

   set $X_t = Y_t$, otherwise set $X_t = x_{t-1}$.

Metropolis-Hastings Algorithm

- The Metropolis-Hastings algorithm defines a Markov chain with transition matrix P such that, for $x, y \in \Omega$

$$\begin{aligned} P_{x,y} &= \mathbb{P}(X_t = y | X_{t-1} = x) \\ &= q(y|x)\alpha(y|x) + \rho(x)\mathbb{I}(y = x) \end{aligned}$$

  where $\rho(x)$ is the probability of rejection

$$\rho(x) = 1 - \sum_{y \in \Omega} q(y|x)\alpha(y|x).$$

Metropolis-Hastings Algorithm

Theorem

The transition matrix P of the Markov chain generated by the Metropolis-Hastings algorithm is reversible with respect to p and therefore admits p as stationary distribution.

- Proof: We check detailed balance. For $x \neq y$

$$
\begin{aligned}
p(x)P_{x,y} &= p(x)q(y|x)\alpha(y|x) \\
&= p(x)q(y|x)\min\left\{1, \frac{p(y)q(x|y)}{p(x)q(y|x)}\right\} \\
&= \min\left\{p(x)q(y|x), p(y)q(x|y)\right\} \\
&= p(y)q(x|y)\min\left\{\frac{p(x)q(y|x)}{p(y)q(x|y)}, 1\right\} \\
&= p(y)q(x|y)\alpha(x|y) \\
&= p(y)P_{y,x}.
\end{aligned}
$$

Metropolis-Hastings Algorithm

- To run the MH algorithm, we need to specify $X_0 = x_0$ (or $X_0 \sim \lambda$) and a proposal $q(y|x)$.
- We only need to know the target p up to a normalizing constant as α depends only on $p(y)/p(x) = \tilde{p}(y)/\tilde{p}(x)$.
- If the Markov chain simulated by the MH algorithm is irreducible and aperiodic then the ergodic theorem applies.
- Verifying aperiodicity is usually straightforward, since the MCMC algorithm may reject the candidate state y, so $P_{x,x} > 0$ for at least some states $x \in \Omega$.
- In order to check irreducibility we need to check that q can take us anywhere in Ω (so q itself is an irreducible transition matrix), and then that the acceptance step doesn't trap the chain (as might happen if $\alpha(y|x)$ is zero too often).

Example: Discrete Distribution on a finite state-space

- Consider a discrete random variable $X \sim p$ on $\Omega = \{1, 2, ..., m\}$ with $\tilde{p}(i) = i$ so $Z_p = \sum_{i=1}^{m} i = \frac{m(m+1)}{2}$.
- One simple proposal distribution is $Y \sim q$ on Ω such that $q(i) = 1/m$.
- Acceptance probability

$$\alpha(y|x) = \min\left\{1, \frac{\tilde{p}(y)q(x|y)}{\tilde{p}(x)q(y|x)}\right\} = \min\left\{1, \frac{y}{x}\right\}$$

- This proposal scheme is clearly irreducible

$$\begin{aligned}\mathbb{P}(X_{t+1} = y|X_t = x) &\geq q(y|x)\alpha(y|x) \\ &= \frac{1}{m}\min(1, y/x) > 0\end{aligned}$$

Example: Discrete Distribution on a finite state-space

- Start from $X_0 = 1$.
- For $t = 1, \ldots, n-1$
  1. Let $Y_t \sim \mathrm{Unif}\{1, 2, ..., m\}$ and $U_t \sim \mathrm{Unif}[0, 1]$
  2. If
  $$U_t \leq \frac{Y_t}{X_{t-1}}$$
  set $X_t = Y_t$, otherwise set $X_t = X_{t-1}$.
- For t large, $X_t \overset{d}{\simeq} X$
- Note: $\alpha(y|x) = \min(1, y/x)$ so $U_t \leq \alpha(y|x) \iff U_t \leq y/x$

Code

```
set.seed(7)
n <- 10000
m <- 20
x <- numeric(n)
x[1] <- 1
for(t in 1:(n - 1)) {
    Yt <- sample(1:m, 1)
    Ut <- runif(1)
    if (Ut <= (Yt / x[t])) {
        x[t + 1] <- Yt
    } else {
        x[t + 1] <- x[t]
    }
}
```

Example: Discrete Distribution on a finite state-space

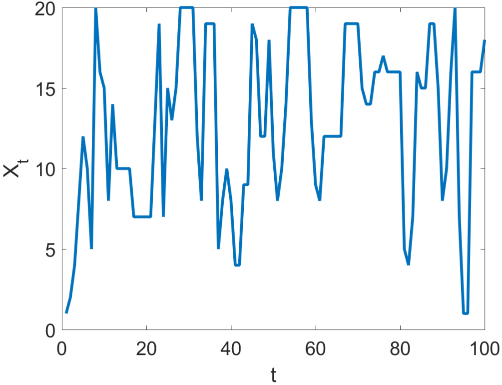


Figure: Realization of the MH Markov chain for $n = 100$ with $m = 20$.

Example: Discrete Distribution on a finite state-space

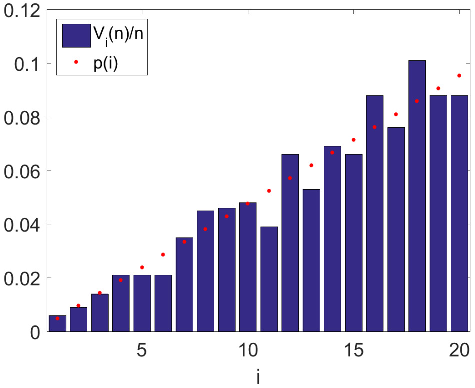


Figure: Average number of visits $V_i(n)/n$ ($n = 1000$) and target pmf $p(i)$

Where $V_i(n) = \sum_{t=0}^{n-1} \mathbb{I}(X_t = i)$

Example: Discrete Distribution on a finite state-space

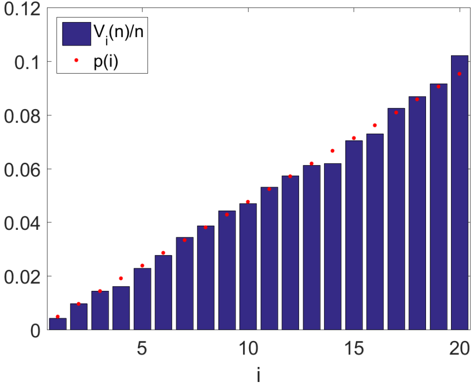


Figure: Average number of visits $V_i(n)/n$ ($n = 10,000$) and target pmf $p(i)$

Example: Poisson Distribution

- We want to simulate $p(x) = e^{-\lambda}\lambda^x/x! \propto \lambda^x/x!$
- For the proposal we use

$$q(y|x) = \begin{cases} \frac{1}{2} & \text{for } y = x \pm 1, x \geq 1 \\ 1 & \text{for } x = 0, y = 1 \\ 0 & \text{otherwise,} \end{cases}$$

  i.e. toss a coin and add or substract 1 to x to obtain y.
- Acceptance probability

$$\alpha(y|x) = \begin{cases} \min\left(1, \frac{\lambda}{x+1}\right) & \text{if } y = x + 1, x \geq 1 \\ \min\left(1, \frac{x}{\lambda}\right) & \text{if } y = x - 1, x \geq 2 \end{cases}$$

  and $\alpha(1|0) = \min(1, \lambda/2)$, $\alpha(0|1) = \min(1, 2/\lambda)$.
- Markov chain is irreducible (check!)

Example: Poisson Distribution

- Set $X_0 = 1$.
- For $t = 1, \ldots, n-1$
  1. If $X_{t-1} = 0$, set $Y_t = 1$
  2. Otherwise, simulate $V_t \sim \mathrm{Unif}[0,1]$
     - 2.1 If $V_t \leq \frac{1}{2}$, set $Y_t = X_{t-1} + 1$.
     - 2.2 Otherwise set $Y_t = X_{t-1} - 1$.
  3. Simulate $U_t \sim \mathrm{Unif}[0,1]$.
  4. If $U_t \leq \alpha(Y_t|X_{t-1})$, set $X_t = Y_t$, otherwise set $X_t = X_{t-1}$.

Example: Poisson distribution

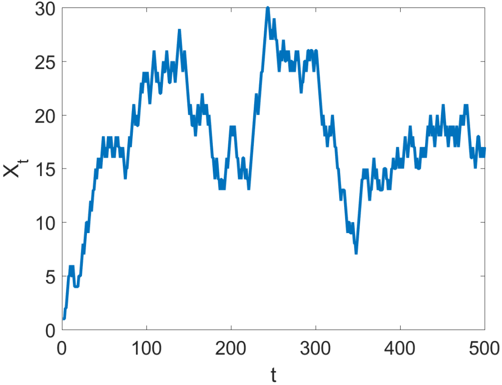


Figure: Realization of the MH Markov chain for $n = 500$ with $\lambda = 20$.

Example: Poisson distribution

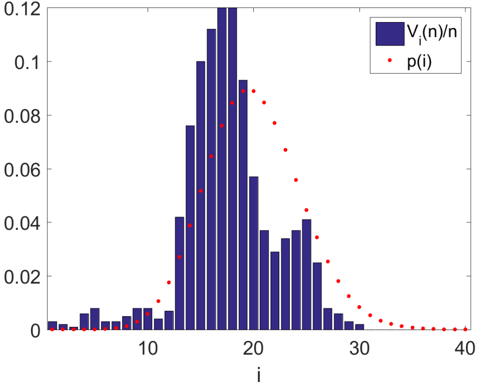


Figure: Average number of visits $V_i(n)/n$ ($n = 1000$) and target pmf $p(i)$

Example: Poisson distribution

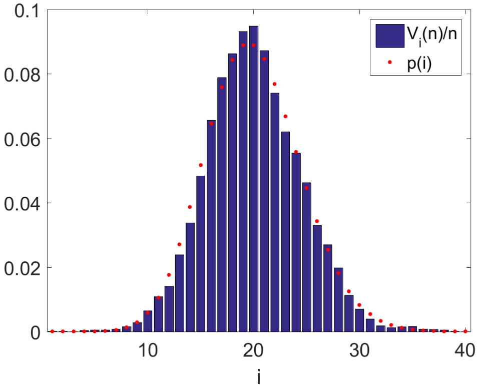


Figure: Average number of visits $V_i(n)/n$ ($n = 10,000$) and target pmf $p(i)$

Example: Image

- Consider a $m_1 \times m_2$ image, where $I(i,j) \in \{0, 1, \ldots, 256\}$ is the gray level of pixel $(i,j) \in \Omega = \{0, \ldots, m_1 - 1\} \times \{0, \ldots, m_2 - 1\}$
- Consider a discrete random variable taking values in Ω
- Unnormalized pdf

$$\tilde{p}((i,j)) = I(i,j)$$

- Proposal transition probabilities

$$q((y_1, y_2)|(x_1, x_2)) = q(y_1|x_1)q(y_2|x_2)$$

with

$$q(y_1|x_1) = \begin{cases} 1/3 & \text{if } y_1 = x_1 \pm 1 \text{ or } y_1 = x_1, \quad \mod m_1 \\ 0 & \text{otherwise} \end{cases}$$

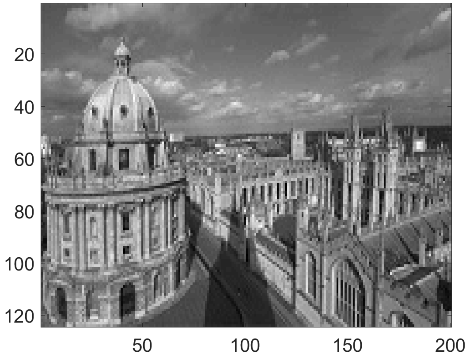
and similarly for $q(y_2|x_2)$.

Example: Simulation of an image

- Average number of visits to each pixel (i, j):
  $V_{(i,j)}(n) = \frac{1}{n}\sum_{t=0}^{n-1} \mathbb{I}(X_t = (i, j))$

Example: Simulation of an image

- Target pmf



Metropolis-Hastings algorithm on $\mathbb{R}^d$

- The Metropolis-Hastings algorithm generalizes to continuous state-space where $\Omega \subseteq \mathbb{R}^d$ with
  1. p is a pdf on Ω
  2. $q(\cdot|x)$ is a pdf on Ω for any $x \in \Omega$
- The Metropolis-Hastings algorithm thus defines a Markov chain on $\Omega \subseteq \mathbb{R}^d$
- Precise definition of Markov chains on $\mathbb{R}^d$ is beyond the scope of this course. We will just state the most important results without proof. Assume for simplicity that $p(x) > 0$ for all $x \in \Omega$

Metropolis-Hastings algorithm on $\mathbb{R}^d$

- The Markov chain $X_0, X_1, \dots$ on $\Omega \subseteq \mathbb{R}^d$ is irreducible if for any $x \in \Omega$ and $A \subset \Omega$, there is n such that

$$\mathbb{P}(X_n \in A | X_0 = x) > 0$$

Theorem

If the Metropolis-Hastings chain is irreducible, then for any function ϕ such that $\mathbb{E}_p[|\phi(X)|] < \infty$, the MH estimator is strongly consistent

$$\widehat{\theta}_n^{MH} = \frac{1}{n} \sum_{t=0}^{n-1} \phi(X_t) \to \theta \quad \textit{almost surely as } n \to \infty$$

Example: Gaussian distribution

- Let $X \sim N(\mu, \sigma^2)$ with

$$p(x) = (2\pi\sigma^2)^{-1/2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- MH algorithm with target pdf p and proposal transition pdf

$$q(y|x) = \begin{cases} 1 & \text{for } y \in [x - 1/2, x + 1/2] \\ 0 & \text{otherwise} \end{cases}$$

- Acceptance probability

$$\alpha(y|x) = \min\left(1, \frac{p(y)q(x|y)}{p(x)q(y|x)}\right) = \min\left(1, e^{-\frac{(y-\mu)^2}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^2}}\right)$$

Example: Gaussian distribution

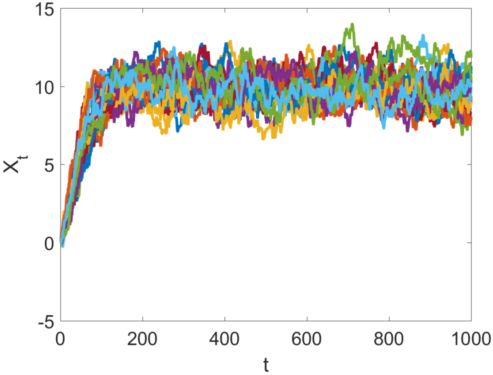


Figure: Realizations from the MH Markov chain $(X_0, \ldots, X_{1000})$ with $X_0 = 0$.

Example: Gaussian distribution

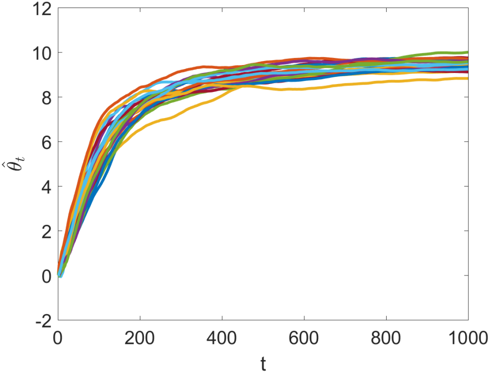


Figure: MH estimates $\widehat{\theta}_t = \frac{1}{t}\sum_{i=0}^{t-1} X_i$ of $\theta = \mathbb{E}_p[X]$ for different realizations of the Markov chain.

Recap

- Given an proposal transition matrix $q(y|x)$ that is irreducible over Ω, MCMC:Metropolis-Hastings defines an algorithm that produces a Markov chain that admits p as a stationary distribution
- **Next time:** Properties of MCMC, tuning parameters, and Gibbs sampling