Avd. Matematisk statistik

KTH Teknikvetenskap

Sf2955 COMPUTER INTENSIVE METHODS
MCMC
On the Discrete Metropolis-Hastings Algorithm
2009
Timo Koski

1 Introduction

1.1 MCMC

Markov chain Monte Carlo (MCMC) is an important computational tool in Bayesian statistics, since it allows inferences to be drawn from complex posterior distributions, where analytical or numerical integration techniques cannot be applied. There are other important applications in image analysis, optimization, bioinformatics and others.

The idea (in Bayesian statistics) is is to generate a Markov chain via iterative Monte Carlo simulation that has, at least in the asymptotic sense, the desired posterior distribution as its stationary distribution.

Since direct sampling from a posterior distribution may not be possible, the Metropolis-Hastings algorithm starts by generating candidate draws from a so-called proposal distribution. These draws are the corrected so that they behave asymptotically as random observations from the desired stationary or target distribution.

The MC constructed by the algorithm at each stage is thus built in two steps: a **proposal** step and an **acceptance** step. These two steps are associated with the proposal and acceptance distributions, respectively.

1.2 Monte Carlo

Most of the quantities usually computed in applications involving statistical models are integrals (or values derived from integrals), in a generic form

$$E\left[\varphi\left(X\right)\right]=\int\varphi(x)dF(x).$$

A **Monte Carlo** method for computation of $E\left[\varphi\left(X\right)\right]$ requires generation of samples $X_1, X_2, \ldots, X_n, \ldots,$ from the distribution F in order to form

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\left(X_i\right).$$

The law of large numbers (or the ergodic theorem) implies

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\left(X_i\right)\approx E\left[\varphi\left(X\right)\right]$$

for large n. A concise early discussion of Monte Carlo methods is in (Freiberger & Grenander 1971).

One way of describing MCMC is to say that it is a Monte Carlo method for computing $E\left[\varphi\left(X\right)\right]$ applicable to 'difficult' distribution functions F.

1.3 Pseudo-random Numbers

All simulation methods, including MCMC, presuppose the availability of random number generators. There are several different algorithms for generating random numbers, see, e.g., (Freiberger and Grenander 1970, ch. 2). These algorithms are deterministic, and therefore one often talks about pseudo-random numbers. Another possibility someone has talked about is to use certain natural sequences (cosmic radiation, radiactive decay) as generators of random numbers. It is not known to us how the properties of pseudo-random number sequences influence the performance of MCMC.

2 Markov chains

First we recall some notation and facts about Markov chains.

2.1 Notation

- $S = \{0, 1, \ldots, J\}$, $J \leq \infty$, is called a state space. We take here mostly $S \subseteq \mathcal{R} =$ the real numbers.

- $j, i \in S \times S$.

- A *discrete time Markov chain* (MC) $\{X_n\}_{n \geq 0}$ is a random sequence, with values in the discrete state space S, such that

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ = P(X_{n+1} = j | X_n = i) = p_{i|j}. \end{aligned}$$

- $p_{i|j}$ is the one-step transition probability.

- State X_n summarizes the past history needed to predict X_{n+1} for any n.

- $$p_{i|j} \geq 0.$$

- $$\sum_{j=0}^{J} p_{i|j} = 1. \tag{2.1}$$

The transition probability matrix is

$$\mathbf{P} = \left(p_{i|j}\right)_{i=0,j=0}^{J,J}$$

$$\mathbf{P} = \begin{pmatrix} p_{0|0} & p_{0|1} & \cdots & p_{0|J} \\ p_{1|0} & p_{1|1} & \cdots & p_{1|J} \\ \vdots & \vdots & \vdots & \vdots \\ p_{J|0} & p_{J|1} & \cdots & p_{J|J} \end{pmatrix}.$$

If J is finite, $\mathbf{P}$ is a $J+1 \times J+1$ *transition matrix*. If J is $= \infty$, then the notation is

$$\mathbf{P} = \begin{pmatrix} p_{0|0} & p_{0|1} & \cdots & p_{0|j} & \cdots \\ p_{1|0} & p_{1|1} & \cdots & p_{1|j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \dots \\ \vdots & \vdots & \ddots & \ddots & \dots \end{pmatrix}.$$

2.2 A Quick Summary of Some Facts

We shall write

$$\{X_n\}_{n=0}^{\infty} \sim \text{Markov}\left(\mathbf{P}, p_{X_0}\right),$$

where

$$p_{X_0} = (p_0, \dots, p_J)$$

is the *initial distribution*.

$$p_{i|j}(n) = P\left(X_{m+n} = j | X_m = i\right), n \geq 1, i, j \in S$$

are also independent of m. The probabilities $p_{i|j}(n)$ are called the n *-step transition probabilities*. We define

$$p_{i|j}(0) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i. \end{cases}$$

With these elements we define the matrix

$$P(n) = \left(p_{i|j}(n)\right)_{i \in S, j \in S}.$$

The matrix is the n *-step transition probability matrix*. Then

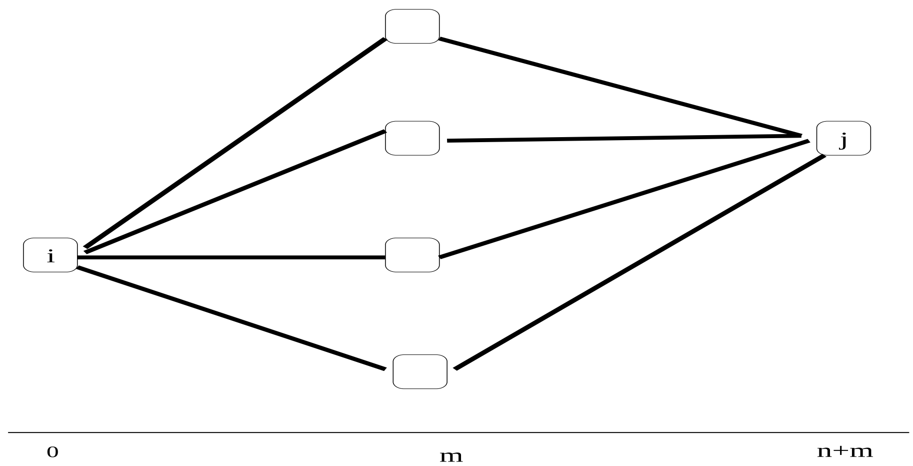
$$P(1) = \mathbf{P}$$

where $\mathbf{P}$ is the one - step transition probability matrix defined first.

For all $m, n \geq 1$ and $i, j \in S$,

$$p_{i|j}(m+n) = \sum_{k=0}^{J} p_{i|k}(m) \cdot p_{k|j}(n).$$

This is known as the Chapman - Kolmogorov equation. Using a matrix no-



Figur 1: Chapman - Kolmogorov

tation we can write the Chapman - Kolmogorov equation as the following matrix multiplication

$$P(n+m) = P(m) \cdot P(n).$$

Then we have the proposition.

Proposition 2.1

$$P(n) = \mathbf{P}^n.$$

Proof: This is easily proved by induction. The case $n = 1$ follows by our definitions

$$P(1) = \mathbf{P} = \mathbf{P}^1.$$

Assume the claim holds for n, i.e., $P(n) = \mathbf{P}^n$. Then by Chapman-Kolmogorov

$$P(n+1) = \mathbf{P} \cdot P(n),$$

and by induction assumption

$$= \mathbf{P} \cdot P^n = \mathbf{P}^{n+1}$$

as was to be proved. ■

Chapman - Kolmogorov equation can be written as

$$P(n+m) = \mathbf{P}^m \cdot \mathbf{P}^n.$$

Let the initial distribution of X_0 be denoted by $\phi(0)$. In other words,

$$\phi(0) = \left(p_{X_0}(0), \ldots, p_{X_0}(J)\right).$$

Let us denote by

$$\phi(n) = \left(p\left(X_n = 0\right), \ldots, p\left(X_n = J\right)\right)$$

the $1 \times J+1$ vector of the probabilities that the chain visits state j at time n. By marginalization

$$p\left(X_n = j\right) = \sum_{k=1}^{J} p_{k|j} \cdot p\left(X_{n-1} = k\right).$$

This we write using a matrix notation as

$$\phi(n) = \phi(n-1)\mathbf{P}. \tag{2.2}$$

A Markov chain $\{X_n\}_{n=0}^{\infty}$ may be such that the probability $p\left(X_n = j\right)$ is independent of n for all j in the state space. A distribution π a *stationary distribution*, with

$$\pi = \left(\pi_0, \ldots, \pi_J\right),$$

if $p\left(X_0 = j\right) = \pi_j$ for all j implies that $p\left(X_1 = j\right) = \pi_j$ for all j.

Proposition 2.2 *Let* $\{X_n\}_{n=0}^{\infty} \sim \text{Markov}\left(P, \phi(0)\right)$. *Every stationary distibution satisfies the equation*

$$\pi = \pi\mathbf{P}$$

(π is a row vector) with the constraints

$$\sum_{j=1}^{J} \pi_j = 1, \pi_j \geq 0.$$

Proof: Assume first that π is a stationary distribution. Then $\sum_{j=1}^{J} \pi_j = 1$ and $\pi_j \geq 0$ are clear. Since π is stationary , by the definition above we must have $\phi(0) = \pi$ and $\phi(1) = \pi$. But since by (2.2)

$$\phi(n) = \phi(n-1)\mathbf{P},$$

we get that

$$\pi = \pi\mathbf{P}.$$

Assume now that π satisfies $\pi = \pi\mathbf{P}$ and the other constraints. Let $\phi(0) = \pi$. Then

$$\phi(1) = \phi(0)\mathbf{P} = \pi\mathbf{P} = \pi$$

and π is a stationary distribution. ■

Proposition 2.3 *Existence of a stationary distribution: Every MC with a finite state space has* at least one *stationary distribution*

Proof: We give only an outline of the proof. Let p be an arbitrary probability distribution on S. Set

$$p^{(n)} = \frac{1}{n}\left(p + p\mathbf{P} + p\mathbf{P}^2 + \ldots + p\mathbf{P}^{n-1}\right)$$

This is a sequence of probability distributions, i.e. vectors with components with values between zero and one. Thus the well known theorem of Bolzano and Weierstrass shows that we can pick a convergent subsequence $p^{(n_v)}$ which converges componentwise to the vector ϕ. We can show that ϕ is a probability distribution. By our construction we have the recursion relations

$$p^{(n+1)} = \frac{n}{n+1}p^{(n)} + \frac{1}{n+1}p\mathbf{P}^n$$

and

$$p^{(n+1)} = \frac{n}{n+1}p^{(n)}\mathbf{P} + \frac{1}{n+1}p.$$

From the recursion above we get that

$$p^{(n_v+1)} \to \pi$$

and then we get that

$$\pi = \pi\mathbf{P},$$

which proves the claim. ■

Is there convergence to a stationary distribution for **any** $\phi(0)$?

Proposition 2.4 *Let* $\{X_n\}_{n=0}^{\infty} \in \text{Markov}\,(\mathbf{P}, \phi(0))$*. Assume that*

$$\lim_{n\to\infty} \phi(n) = a,$$

where $a = (a_0, \ldots, a_J)$ *is a probability distribution. Then* a *is an stationary distribution.*

Proof: Taking of limits yields

$$a = \lim_{n\to\infty} \phi(n) = \lim_{n\to\infty} \phi(n+1) =$$

$$= \lim_{n\to\infty} (\phi(n)\mathbf{P}) = \left(\lim_{n\to\infty} \phi(n)\right) \mathbf{P} = a\mathbf{P}.$$

■

We need some new definitions.

(a) An MC is **aperiodic**, if there is no state such that return to that state is possible only after t_0, $2t_0$, $3t_0$ ... steps later.

(b) An MC is irreducible means that every state can eventually (observera att 'eventuellt' på svenska betyder ngt annat) be reached from any other state, if not in one step, but then after several steps.

Proposition 2.5 *If a finite MC is aperiodic and irreducible, then for any* $\phi(0)$

$$\lim_{n\to\infty} \phi(n) = \pi,$$

where π *is a unique probability distribution that satisfies*

$$\pi = \pi P.$$

Proof: Omitted. ■

2.3 Time Reversible Markov Chains

$\{X_n\}_{n\geq 0} \sim \text{Markov}\,(\mathbf{P}, \phi(0))$ with the state space S. A probability distribution $\mathbf{a}$ on S is said to be **reversible** for the chain (or for the matrix $\mathbf{P}$), if for all $i, j \in S \times S$ we have

$$a_i p_{i|j} = a_j p_{j|i} \tag{2.3}$$

The MC is said to be **reversible**, if there exists a reversible distribution for it.

Proposition 2.6 $\{X_n\}_{n\geq 0} \sim$ Markov $(\mathbf{P}, \phi(0))$ with the state space S. Assume that $\mathbf{a}$ on S is reversible for the chain. Then $\mathbf{a}$ is a stationary distribution for the chain.

Proof: By (2.1)

$$a_i = a_i \sum_{j \in S} p_{i|j} = \sum_{j \in S} a_i p_{i|j},$$

and by (2.3)

$$= \sum_{j \in S} a_j p_{j|i},$$

$$\Leftrightarrow$$

$$\mathbf{a} = \mathbf{aP}.$$

Here we regard $\mathbf{a}$ as a $1 \times J$ row vector. ■

3 Basics of MCMC for Discrete One-Dimensional Distributions

3.1 Introduction

The preceding part of these lectures has discussed the problem of finding the stationary distribution for a given transition matrix. Next we are concerned with a kind of converse task, here called the Metropolis problem.

3.2 The Metropolis Problem

Let $\mathbf{f} = \{f_j\}_{j \in S}$ be an arbitrary probability mass function, **target distribution**, on *a discrete subset* S *of* $\mathcal{R}$, i.e.,

- $f_j \geq 0$.
- $\sum_{j=0}^{J} f_j = 1$.

The *Metropolis problem* is to give a Markov chain such that $\mathbf{f}$ is its stationary distribution. We shall next show that it is always possible to solve the stated problem by constructing an appropriate transition probability matrix. In fact there are infinitely many solutions to the stated problem.

3.3 A Solution of the Metropolis Problem

Let $\mathbf{Q} = \left(q_{i|j}\right)_{i=0,j=0}^{J,J}$ be a probability transition matrix on S. Assume that the matrix $\mathbf{Q}$ is **symmetric**, i.e.,

$$q_{i|j} = q_{j|i}, \quad \text{for all } i, j \in S \times S. \tag{3.1}$$

We want to construct an MC $\{X_n\}_{n \geq 0}$ with the state space S and the stationary distribution $\mathbf{f}$. We use the following rules of transition.

Suppose that $X_n = i$. We **propose** a value

$$Y_{n+1} = j$$

drawn from the (row) distribution $\{q_{i|j}\}_{j=0}^{J}$, independently of $X_0, \ldots, X_{n-1}$, Then let us **accept** j with the probability

$$\alpha_{i,j} := \min \left\{ 1, \frac{f_j}{f_i} \right\}, \tag{3.2}$$

whereby acceptance means that the chain moves to j, $X_{n+1} = j$. We **reject** the proposed value j with probability

$$1 - \alpha_{i,j}, \tag{3.3}$$

whereby rejection means that the chain stays at i, $X_{n+1} = i$. The procedure is to be visualized/implemented in terms of an independent random toss of coin with the probability mass function $(1 - \alpha_{i,j}, \alpha_{i,j})$.

We shall now find the transition probabilities $p_{i|j}$ of the chain thus defined. Let $\{\mathbf{ta}\}$ denote the event that the proposed transition is accepted, and let $\{\mathbf{ta}\}^c$ denote the complement. Then, if $i \neq j$,

$$p_{i|j} = P\left(X_{n+1} = j | X_n = i\right) = P\left(Y_{n+1} = j, \mathbf{ta} | X_n = i\right)$$

by definition of proposal and acceptance moves. It follows by the chain rule

$$= P\left(Y_{n+1} = j, |\mathbf{ta}, X_n = i\right) P\left(\mathbf{ta} | X_n = i\right),$$

and since proposal is generated independently of acceptance, given X_n,

$$= P\left(Y_{n+1} = j | X_n = i\right) \alpha_{i,j}.$$

In other words we have obtained for $i \neq j$ that

$$p_{i|j} = q_{i|j} \cdot \min\left\{1, \frac{f_j}{f_i}\right\} = q_{i|j} \cdot \alpha_{i,j}. \tag{3.4}$$

On the diagonal $i = j$ we have

$$p_{i|i} = P\left(X_{n+1} = i|X_n = i\right) = P\left(Y_{n+1} = i, \mathbf{ta}|X_n = i\right) + P\left(Y_{n+1} \neq i, \mathbf{ta}^c|X_n = i\right)$$

and as above

$$= P\left(Y_{n+1} = i|X_n = i\right) P\left(\mathbf{ta}|X_n = i\right) + P\left(Y_{n+1} \neq i|X_n = i\right) P\left(\mathbf{ta}^c|X_n = i\right)$$

$$= P\left(Y_{n+1} = i|X_n = i\right) \alpha_{i,i} + \sum_{j \neq i} P\left(Y_{n+1} = j|X_n = i\right)\left(1 - \alpha_{i,j}\right)$$

$$= q_{i|i}\alpha_{i,i} + \sum_{j \neq i} P\left(Y_{n+1} = j|X_n = i\right)\left(1 - \alpha_{i,j}\right)$$

$$= q_{i|i} + \sum_{j \neq i} q_{i|j}\left(1 - \alpha_{i,j}\right).$$

Thereby we have defined a legitimate transition probability matrix $\mathbf{P}$.

We need to prove that $\mathbf{f}$ is a stationary distribution for Markov chain $\{X_n\}_{n\geq 0} \sim \text{Markov}\left(\mathbf{P}, \phi(0)\right)$, where $\phi(0)$ is an arbitrary initial distribution. We will show that the reversibility condition (2.3) holds w.r.t. $\mathbf{f}$.

For this purpose we consider $i \neq j$ and assume that

$$f_i < f_j.$$

Then due to (3.4), and the assumed symmetry of $\mathbf{Q}$,

$$\begin{aligned} f_i p_{i|j} &= f_i q_{i|j} \min\left\{1, \frac{f_j}{f_i}\right\} \\ &= f_i q_{i|j} \\ &= q_{i|j} \min\left\{1, \frac{f_i}{f_j}\right\} f_j \\ &= q_{j|i}\alpha_{j,i} f_j \\ &= p_{j|i} f_j. \end{aligned} \tag{3.5}$$

In case we have

$$f_j < f_i,$$

we start with

$$f_j p_{j|i} = q_{j|i} \min\left\{1, \frac{f_i}{f_j}\right\} f_j$$

and continue analogously. Hence the reversibility condition (2.3) holds for all i, j, and we have in view of proposition 2.6 solved the Metropolis problem.

Example 3.1 Let

$$\mathbf{f} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{3}\right),$$

and

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & 0 \\ \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{3} \end{pmatrix}.$$

Then the acceptance probabilities are

$$\alpha = \begin{pmatrix} 1 & 1 & 0.6667 & 1 \\ 1 & 1 & 0.6667 & 1 \\ 1 & 1 & 1 & 0 \\ 0.75 & 0.75 & 0 & 1 \end{pmatrix},$$

and the desired transition matrix is

$$\mathbf{P} = \begin{pmatrix} 0.2222 & 0.1667 & 0.1111 & 0.5 \\ 0.1667 & 0.5556 & 0.1111 & 0.1667 \\ 0.1667 & 0.1667 & 0.667 & 0 \\ 0.3750 & 0.125 & 0 & 0.5 \end{pmatrix}.$$

The computations above have been done using the MATLAB^R function in appendix 8.1.

■

4 The One Dimensional Discrete Metropolis Algorithm

The solution of the Metropolis problem, as established above, can be used to simulate a Markov chain $\{X_n\}_{n\geq 0}$ that has the target distribution as the preassigned stationary distribution $\mathbf{f}$. The pertinent simulation algorithm is known as the **Metropolis algorithm**. When $S \subset \mathcal{R}$, this is an algorithm for simulating one dimensional Markov chains.

Definition 4.1 (Metropolis Algorithm) $\mathbf{Q} = \left(q_{i|j}\right)_{i=0,j=0}^{J,J}$ is a symmetric transition probability matrix. Given that $X_n = i$

1. Generate $Y_{n+1} \sim \{q_{i|j}\}_{j=0}^{J}$.
2. Take
$$X_{n+1} = \begin{cases} Y_{n+1} & \text{with probability } \alpha_{i,Y_{n+1}} \\ i & \text{with probability } 1 - \alpha_{i,Y_{n+1}}, \end{cases}$$
where
$$\alpha_{i,j} = \min\left\{1, \frac{f_j}{f_i}\right\}.$$

The distributions $\{q_{i|j}\}_{j=0}^{J}$ are called *proposal* distributions.

■

5 The One Dimensional Discrete Metropolis-Hastings Algorithm

5.1 The Algorithm

In Hastings (1970) the algorithm of Metropolis was generalized by relaxing the requirement that the matrix of proposal distributions $\mathbf{Q}$ be symmetric. The more general simulation algorithm is known as the **Metropolis-Hastings algorithm**.

Definition 5.1 (Metropolis-Hastings Algorithm) $\mathbf{Q} = \left(q_{i|j}\right)_{i=0,j=0}^{J,J}$ is a transition probability matrix. Given that $X_n = i$

1. Generate $Y_{n+1} \sim \{q_{i|j}\}_{j=0}^{J}$.

2. Take

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{with probability } \alpha^H_{i,Y_{n+1}} \\ i & \text{with probability } 1 - \alpha^H_{i,Y_{n+1}}, \end{cases}$$

   where

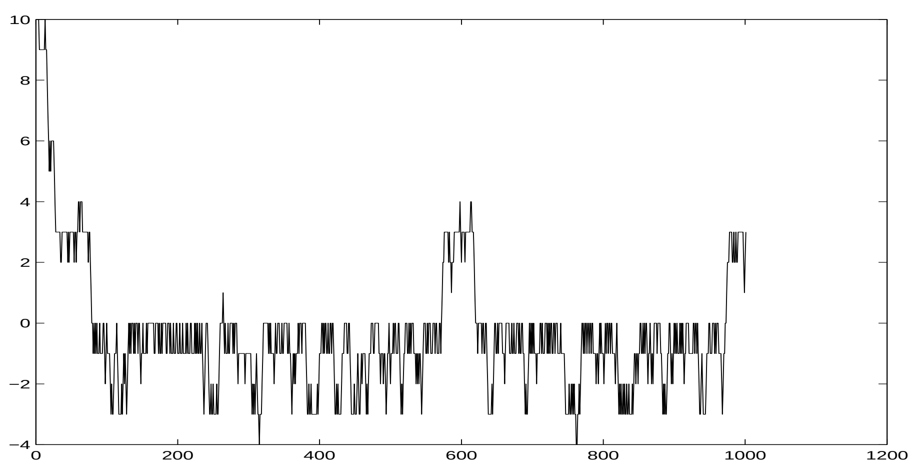
$$\alpha^H_{i,j} = \min\left\{1, \frac{f_j q_{j|i}}{f_i q_{i|j}}\right\}. \tag{5.1}$$

The distributions $\{q_{i|j}\}_{j=0}^{J}$ are called *proposal* distributions.

■

5.2 Burn-In

A relevant point of discussion is the determination of the convergence of the chain, or more precisely, the determination of how close the marginal distribution at iteration n is to the targeted stationary distribution. There are theoretical bounds for the total variation distances between these distributions. More of this will (hopefully) be said later in this course (Häggström 2000, ch. 8, ch. 10).



Figur 2: Burn-in

In pragmatic terms, chain values in the initial stage are far from the stationary distribution and should be discarded. This period is referred to as the **burn-in** period. The phenomenon is illustrated in figure 2. A very simple informal way to do this is to monitor convergence by looking at the simulated path of the chain.

5.3 Examples

The examples in this subsection deal with target distributions on integers. The algorithms are based on taking **Q** as a transition matrix of a suitable *simple random walk*, as defined in Appendix A. The first example, where the target distribution is a Poisson distribution, is of no salient importance as an application of Metropolis-Hastings, since a Poisson distribution can be simulated simpler by other means (e.g, by using exponentially distributed pseudorandom numbers (Gamerman 1997, ch. 1), in view of the theoretical properties of a Poisson process).

Example 5.1 *[Poisson Distribution as the Target Distribution]* Consider **f** as the Poisson distribution with intensity $\lambda > 0$, or,

$$f_i = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots \tag{5.2}$$

Hastings (1970) suggests as **Q** the transition matrix of a reflecting (simple) random walk on the non-negative integers, or

$$q_{0|0} = \frac{1}{2}, q_{0|1} = \frac{1}{2},$$

and

$$q_{i|j} = \begin{cases} \frac{1}{2} & j = i - 1 \\ \frac{1}{2} & j = i + 1 \\ 0 & \text{otherwise.} \end{cases}$$

This gives

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \ldots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \ldots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \ldots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ldots \end{pmatrix}.$$

Hence $\mathbf{Q}$ is in fact symmetric, and the algorithm reduces to that of Metropolis. Then in (5.1),

$$p_{i|j} = q_{i|j} \cdot \alpha^{H}_{i,j} = \begin{cases} \frac{1}{2} \cdot \min\left\{1, \frac{i}{\lambda}\right\} & j = i-1 \\ \frac{1}{2} \cdot \min\left\{1, \frac{\lambda}{i+1}\right\} & j = i+1 \\ 1 - p_{i|i-1} - p_{i|i+1} & j = i \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

For $i = 0$

$$p_{0|j} = \begin{cases} \frac{1}{2} \min(1, \lambda) & j = 1 \\ 1 - \frac{1}{2} \min(1, \lambda) & j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

A chain like the one obtained here is often called a birth and death process (with reflection). Some more facts are found in Appendix A below.

We should, of course, convince ourselves of the fact that the MC with this transition probability is aperiodic and irreducible. This topic is commented in Appendix A, and in the section Exercises 6.

In practice, if λ is small, this choice of $\mathbf{Q}$ seems to work fairly well and fast to approximate $\mathbf{f}$, since it suffices that the chain visits only the first few integers. If λ is large, then a more general random walk with larger steps will be needed to give a proposal distribution in order that more states will be visited in a reasonably short sample path.

For example, if $\lambda = 0.2$, the target distribution in (5.2) has the values

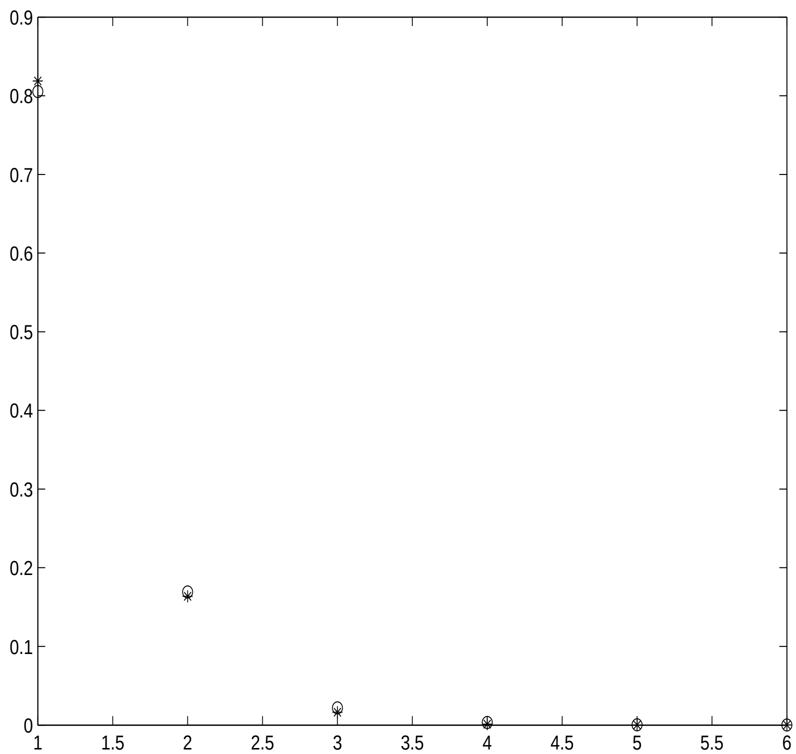
$$f_0 = 0.8187, f_1 = 0.1637, f_2 = 0.0164, f_3 = 0.0011, f_4 = 0.0001, f_5 = 0.0000.$$

$$p_{0|j} = \begin{cases} 0.1 & j = 1 \\ 0.9 & j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, for $i \geq 1$ and $\lambda = 0.2$

$$p_{i|j} = \begin{cases} \frac{1}{2} & j = i-1 \\ 1 - \frac{1}{2} - \frac{0.1}{i+1} & j = i \\ \frac{0.1}{i+1} & j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

In figure 3 one compares for $\lambda = 0.2$ the empirical relative frequencies, as obtained after the 'burn-in' phase of a simulated path with 10 000 samples partly depicted in figure 4, with the target distribution.



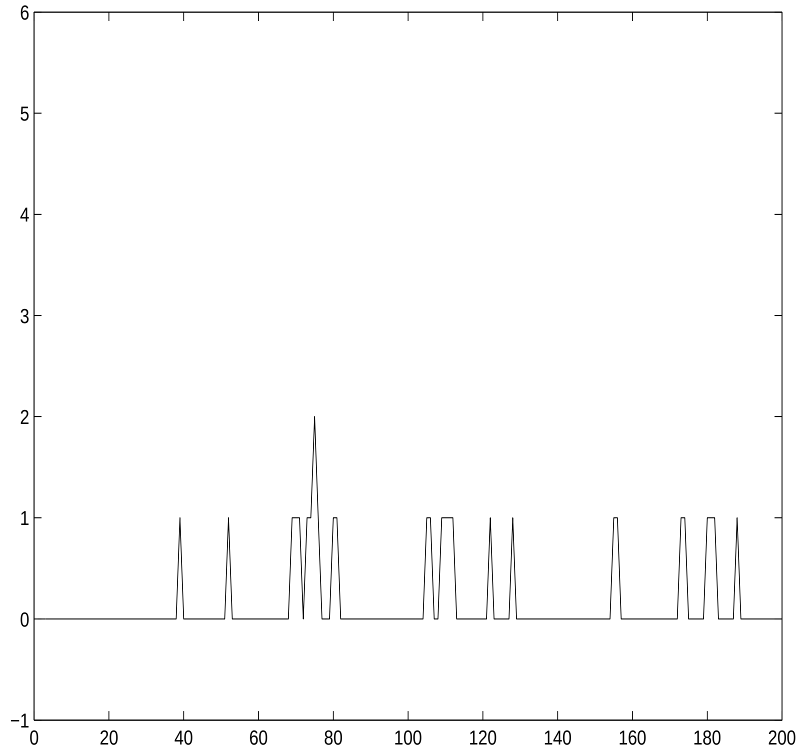
Figur 3: Probability mass function Po(0.2) depicted by $*$ and a simulated empirical relative frequency depicted by o. The scale in x-axis should be shifted by one step to the right.

■

The second example shows a more interesting case, i.e., a more difficult distribution. In this situation the standard technique of using the cumulative distribution function for simulation, see (Häggström 2002, p. 19, Gamerman 1997, ch. 1) is not straightforward, since the normalization constant of the distribution poses a problem.

Example 5.2 *[A Difficult Target Distribution]* Let the state space be $S = \ldots, -2, -1, 0, 1, 2, \ldots,$. The target distribution is

$$f_i = C \cdot \left(i - \frac{1}{2}\right)^4 e^{-3|i|} \cos^2(i), \quad i \in S. \tag{5.4}$$



Figur 4: A (portion of a) simulated path of the MC with Po(0.2) as the target distribution.

Here

$$C = \frac{1}{\sum_{i=-\infty}^{\infty} \left(i - \frac{1}{2}\right)^4 e^{-3|i|} \cos^2(i)}, \tag{5.5}$$

which seems to have no explicit expression. We need to construct the proposal distributions. Again, we take as $\mathbf{Q}$ the transition matrix of a simple random walk, see Appendix A, by

$$q_{i|j} = \begin{cases} \frac{1}{2} & j = i - 1 \\ \frac{1}{2} & j = i + 1 \\ 0 & \text{otherwise.} \end{cases}$$

for all $i, j \in S \times S$. Then we get for $j = i - 1$ or $j = i + 1$.

$$
\begin{aligned}
\alpha_{i,j} &= \min\left\{1, \frac{f_j q_{j|i}}{f_i q_{i|j}}\right\} \\
&= \min\left\{1, \frac{\frac{1}{2}C \cdot \left(j - \frac{1}{2}\right)^4 e^{-3|j|}\cos^2(j)}{\frac{1}{2}C \cdot \left(i - \frac{1}{2}\right)^4 e^{-3|i|}\cos^2(i)}\right\} \\
&= \min\left\{1, \frac{\left(j - \frac{1}{2}\right)^4 e^{-3|j|}\cos^2(j)}{\left(i - \frac{1}{2}\right)^4 e^{-3|i|}\cos^2(i)}\right\}.
\end{aligned}
\tag{5.6}
$$

Note that C has cancelled out, so that $\alpha_{i,j}$ does not depend on C. The expression for $\alpha_{i,j}$ in (5.6) can be readily computed using Matlab^R or any other numerical software.

In figure 5 there is shown the histogram, as obtained from the simulated path with 10 000 samples partly depicted in figure 6 simulated by Metropolis-Hastings with the target distribution in (5.4).

■

6 Exercises

1. Prove that the Metropolis-Hastings algorithm defines a Markov chain with the desired target distribution as the stationary distribution.

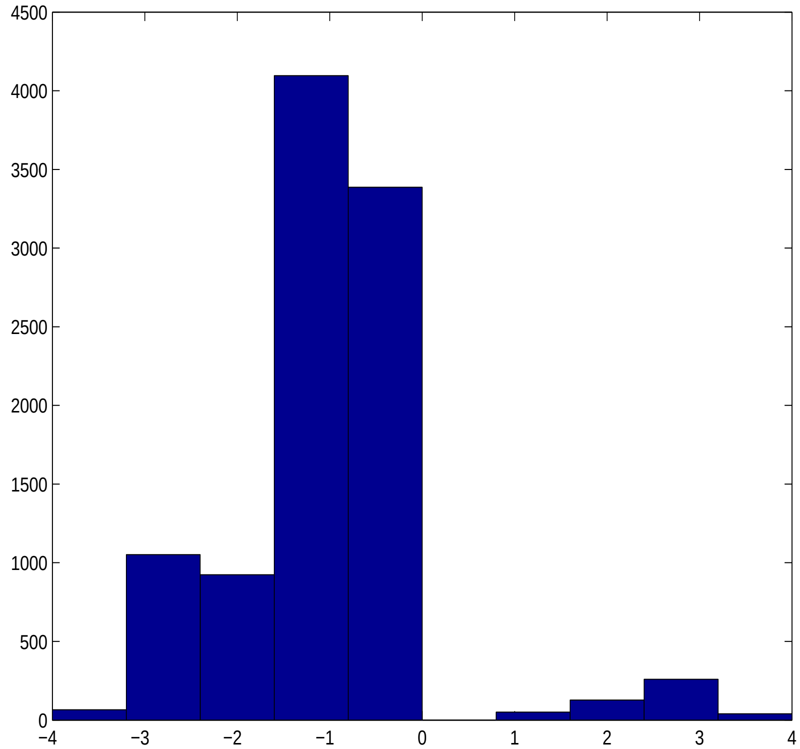
2. In *Barker's algorithm* for MCMC one takes the acceptance probability as

$$\alpha^{\mathrm{B}}_{i,j} = \frac{f_j}{f_i + f_j}.$$

   Show that a Metropolis-Hastings algorithm with this acceptance probability generates an MC with the target distribution as stationary distribution under a certain condition (which ?) on $\mathbf{Q}$.

3. Give and implement, e.g., with Matlab$^{\mathrm{R}}$ functions, c.f. Appendix B below, a Metropolis-Hastings algorithm for simulating the target distribution

$$\mathbf{f} = (1/3, 1/5, 2/15, 1/3)$$



Figur 5: The histogram for the MC with (5.4) as the target distribution.

Show a histogram of the distribution from your simulations. Take into account the burn-in period, by discarding a suitable portion of samples used in computing your histogram.

4. **(a)** Explain why a birth and death chain with reflection (Appendix A) is aperiodic and irreducible.

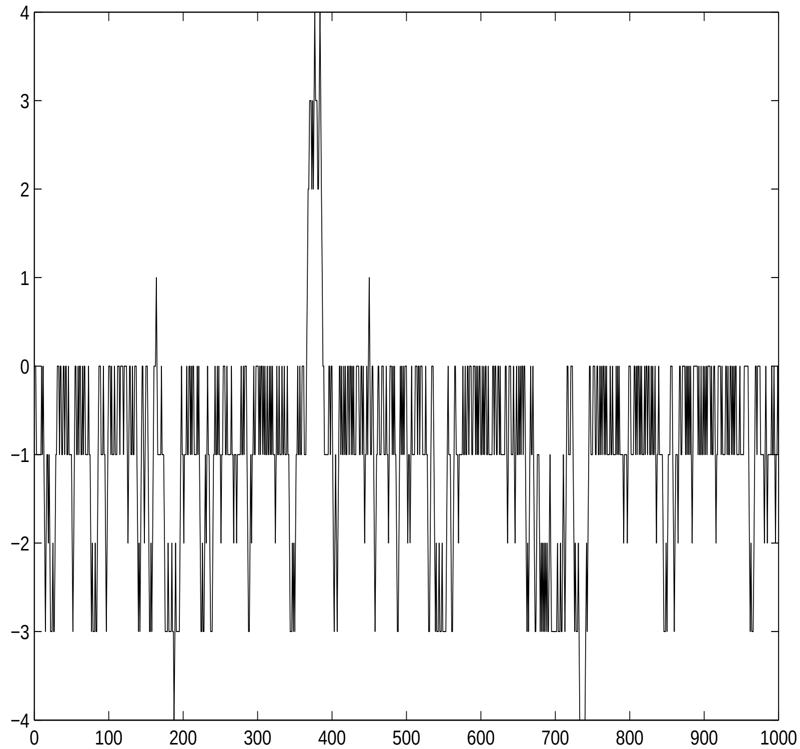
   (b) Show that the quantities π_i defined in (A.7), i.e.,

$$\pi_i = Cv_i, i \geq 1, \pi_0 = C$$

   constitute in fact the stationary distribution for a birth and death chain with reflection at origin.

5. Let the target distribution be

$$f_i = C \cdot e^{-j^4}, j = \ldots, -2, -1, 0, 1, 2, \ldots, .$$



Figur 6: A (portion of a) simulated path of the MC with (5.4) as the target distribution.

(a) Give the Metropolis-Hastings algorithm for this target distribution, when the proposal distribution is given by a simple random walk (Appendix A) with $p = \frac{1}{4}$.

(b) Write a code, e.g., with Matlab^{R} functions, for the pertinent algorithm to simulate samples from this distribution, and show the histograms of the distribution from your simulations. Take into account the burn-in period, by discarding a suitable portion of samples used in computing your histogram.

6. Consider the Metropolis-Hastings algorithm in example 5.1 with the $\mathrm{Po}(\lambda)$ as the target distribution.

(a) Show that the chain defined by the algorithm is recurrent using the criterion in (A.5).

(b) Check that the stationary distribution defined by means of (A.6) is in fact the Poisson distribution Po(λ), like it should.

7. Write a code for, e.g., with Matlab^{R} functions, for the Metropolis-Hastings algorithm in example 5.1 with the Po (λ) as the target distribution, when

   (a) $\lambda = 0.1$.

   (b) $\lambda = 3.2$.

   Show empirical histograms of your simulations. Take into account the burn-in period by discarding a suitable portion of samples used in computing your histograms.

7 Appendix A: Random Walks & Birth and Death Chains

7.1 Simple Random Walk

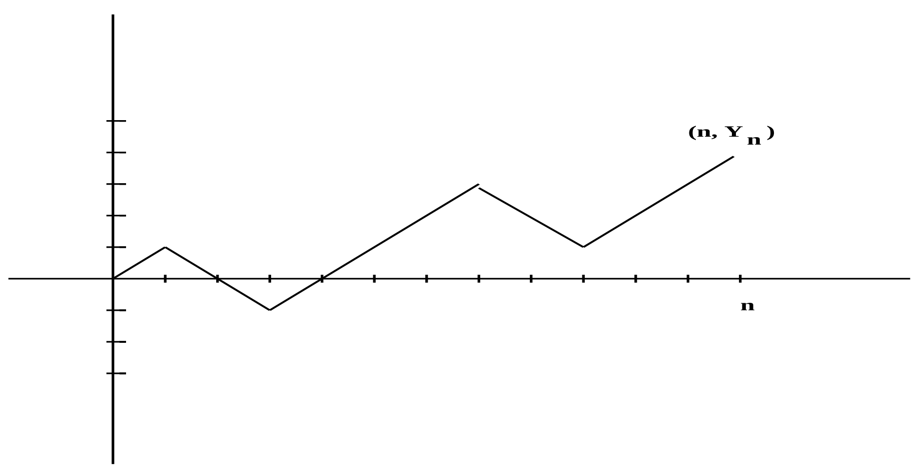
Let for $\{Z_i\}_{i\geq 1}$ be a sequence of I.I.D. random variables with values in $\{-1,1\}$ and with

$$Z_i = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } q = 1-p. \end{cases} \tag{A.1}$$

We take $Y_0 = h$, h is an integer. Let

$$Y_{n+1} = h + \sum_{i=1}^{n+1} Z_i = Y_n + Z_{n+1}. \tag{A.2}$$

This is called a *simple random walk*. If $p = q = 1/2$, we refer to the *classical simple random walk*. This is thought of as a random motion of a particle that inhabits one of the integer points of the real line. The graphic representation of an outcome of the sequence $\{(n, Y_n)|n =, 1, \ldots,\}$, is called a *path* of the particle, see figure 7, where $h = 0$.



Figur 7: A path of a simple random walk

Lemma 7.1 The simple random walk is spatially homogeneous, that is

$$P(Y_n = j \mid Y_0 = h) = P(Y_n = j + b \mid Y_0 = h + b).$$

Proof:

$$P(Y_n = j \mid Y_0 = h) = P\left(\sum_{i=1}^{n} Z_i = j - h\right).$$

On the other hand

$$P(Y_n = j + b \mid Y_0 = h + b) = P\left(\sum_{i=1}^{n} Z_i = j - h\right).$$

■

Lemma 7.2 The simple random walk is temporally homogeneous, that is

$$P(Y_n = j \mid Y_0 = h) = P(Y_{n+m} = j \mid Y_m = h).$$

Proof:

$$P(Y_n = j \mid Y_0 = h) = P\left(\sum_{i=1}^{n} Z_i = j - h\right)$$

and since Z_i are I.I.D.,

$$= P\left(\sum_{i=m+1}^{m+n} Z_i = j - h\right) = P(Y_{n+m} = j \mid Y_m = h).$$

■

Lemma 7.3 The simple random walk has the Markov property, that is,

$$P(Y_{n+1} = j \mid Y_0, Y_1, \ldots, Y_n) = P(Y_{n+1} = j \mid Y_n).$$

Proof:

$$P(Y_{n+1} = j \mid Y_0, Y_1, \ldots, Y_n = a) =$$

$$= P(Z_{n+1} = j - a) = P(Z_{n+1} = j - a \mid Y_n = a) = P(Y_{n+1} = j \mid Y_n = a)$$

■

Hence $\{Y_n\}_{n\geq 0} \sim \text{Markov}(\mathbf{Q}, \delta_h)$ with the integers as state space, where δ_h

is the unit mass at h, and the transition probability matrix, $\mathbf{Q}$, of the simple random walk has the entries

$$q_{i|j} = \begin{cases} p & \text{if } j = i+1 \\ q = 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

7.2 Birth-Death Chains With Reflection at Origin

Let us consider a Markov chain $\{X_n\}_{n\geq 0}$ with the non-negative integers as the state space, and with the transition matrix $\mathbf{P}$ with the arrays for $i \geq 1$, c.f., figure 8,

$$p_{i|j} = \begin{cases} q_i & j = i-1 \\ p_i & j = i+1 \\ r_i & j = i \\ 0 & \text{otherwise.} \end{cases} \tag{A.3}$$

where $r_i = 1 - q_i - p_i$, $q_i > 0$, $p_i > 0$. By *reflection at origin* we mean the folowing boundary condition at $i = 0$

$$p_{0|j} = \begin{cases} p_0 & j = 1 \\ r_0 = 1 - p_0 & j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

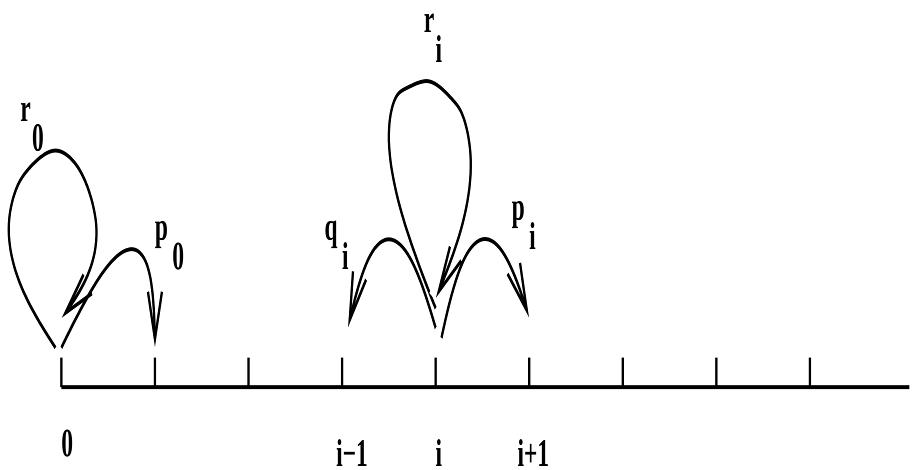
This is a rudimentary model for statistics of population sizes, where during each period only one new individual, or none, is born or dies. The reflection at origin means that the population does not die out, but obtains a new founding member, as it were. We call this a birth and death chain with reflection, with the acronym BDC(r).

It is not difficult to argue that a BDC is irreducible and aperiodic, this is left as an exercise (section 6). Let us set

$$\gamma_0 = 1, \gamma_i = \frac{\prod_{l=1}^{i} q_l}{\prod_{l=1}^{i} p_l}, i \geq 1. \tag{A.4}$$

It has been shown that the BDC(r) is recurrent, if and only if

$$\sum_{i=1}^{\infty} \gamma_i = +\infty. \tag{A.5}$$



Figur 8: Birth-Death Chain Transitions

The tools given in these lectures are not well-suited for proving this assertion. Let us also set

$$\upsilon_0 = 1, \upsilon_i = \frac{\prod_{l=0}^{i-1} p_l}{\prod_{l=1}^{i} q_l}, i \geq 1. \tag{A.6}$$

In the same vein it can be shown that the stationary distribution of the BDC(r) is given by

$$\pi_i = C\upsilon_i, \tag{A.7}$$

where C is i the normalization constant. An exercise in section 6 checks this w.r.t. example 5.1.

8 Appendix B: Some Relevant Matlab Functions

8.1 A Matlab^R function for P

The next Matlab^R function (Englund 2000) delivers a matrix alpha with the probabilities of acceptance, and the corresponding transition probability matrix P.

```
function [P,alpha]=pmatris(f,Q)
```

```
s=size(f);
n=s(2);
P=zeros(n,n);
for i=1:n,
for j=1:n,
alpha(i,j)=min(f(j)*Q(j,i)/max(f(i)*Q(i,j),eps),1);
if j~=i,
P(i,j)=alpha(i,j)*Q(i,j);
end,
end,
end,
for i=1:n,
su=sum(P');
P(i,i)=1-su(i);
end;
```

8.2 A Matlab^R function for the discrete Metropolis-Hastings algorithm

This is a Matlab^R function that implements the Metropolis-Hastings algorithm for target distributions on finite subsets of positive integers. It uses as subroutines the function `pmatris` above and the update function `generate` implemented using (Häggström (2002) p. 20), but not shown here.

```
function x=metropolis(f,Q,n,start);

  [P,alpha]=pmatris(f,Q);
x=start;
last=start;
for i=1:n
y=generate(1,Q(last,:));
if rand(1)< alpha(last,y)
x=[x y];
last=y;
else,
x=[x last];
end,
end,
```

8.3 A Matlab^R function for the discrete Metropolis-Hastings algorithm in example 5.2

First we give a function for the acceptance probabilities.

```
function alpha=alphaexempel(last,y)
a=(y-(1/2))^(4)* exp(-3*abs(y))*(cos(y)^(2));
b=(last-(1/2))^(4)* exp(-3*abs(last))*(cos(last)^(2));
alpha=min(1, a/b);
```

In the next function the variable p represents the probability of a simple random walk taking the step up. Hence Z_i in (A.1) are implemented as `z= -sign(u-p)` , where u are pseudorandom numbers simulating samples of $U(0, 1)$.

```
function x=metropolisexempel(n,start,p);
x=start;
last=start;
for i=1:n
u=rand(1);
z= -sign(u-p)
y= last +z;
if rand(1)< alphaexempel(last,y)
x=[x y];
last=y;
else,
x=[x last];
end,
end,
```

9 References and further reading:

1 Journal articles and technical reports:

- W.K.Hastings (1970): Monte Carlo sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57, pp. 97−109.

2 Books:

- G. Englund (2000): *Datorintensiva metoder i matematisk statistik.* Institutionen för matematik, KTH, Stockholm.
- W.Freiberger & U. Grenander (1971): *A Short Course in Computational Probability and Statistics.* Springer Verlag. New York, Heidelberg, Berlin.

- D. Gamerman (1997): *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC. Boca Raton, London, New York, Washington.
- O. Häggström (2002): *Finite Markov Chains and Algorithmic Applications.* London Mathematical Society, Student Texts, Cambridge University Press, Cambridge.