Bayesian statistics, simulation and software

Module 8: More MCMC: Invariant density, irreducibility, Metropolis-Hastings algorithm

Jesper Møller and Ege Rubak

Department of Mathematical Sciences
Aalborg University

Target density and MCMC

Recall that given a density $\pi(x)$ – which we think of as a target density (e.g. a prior density or a posterior density) from which we want to make simulations – a MCMC algorithm is a way of constructing a Markov chain $X_0, X_1, \ldots$ to produce such simulations (at least approximately).

Today we provide the theory ensuring that this works: we discuss conditions ensuring that π becomes the **limiting density** π, that is, for any event A we have

$$P(X_t \in A) \to \int_A \pi(x)dx \quad \text{as } t \to \infty.$$

NB: Here and in the following we assume π is a probability density function, but everything works as well when it is a probability mass function (then just replace integrals by sums) – or a density for a combination of discrete and continuous random variables...

Invariance

Definition: Invariant density

A Markov chain with transition kernel $P(x, A)$ has **invariant (or stationary or equilibrium) density** $\pi(x)$, if for all events $A \subseteq \Omega$,

$$\int_\Omega \pi(x)P(x, A)dx = \int_A \pi(x)dx.$$

In other words, if at some time t we have that X_t has density π, then X_{t+1} has density π, and hence at any time $s \geq t$ we have that X_s has density π.

It can be shown that if the Markov chain has a limiting density π, then π must be an invariant density of the Markov chain.

Invariance for the Metropolis-Hastings algorithm

Theorem

The Metropolis-Hastings algorithm produces a time homogeneous Markov chain with its target density π as its invariant density.

Irreducible

Definition: Irreducible Markov chain

A Markov chain with invariant density $\pi(x)$ is **irreducible** if for all states $x \in \Omega$ and all events $A \subseteq \Omega$ with $\int_A \pi(x)dx > 0$, there exists a time $n \in \{1, 2, \dots\}$ so that

$$P^n(x, A) > 0.$$

Otherwise it is said to be **reducible**.

Briefly speaking this means that for any feasible event A, no matter at which state x the Markov chain is started, it is possible within a finite time that the Markov chain reaches A.

FACT: If the Markov chain has a limiting density π, then the Markov chain must be irreducible!

Theorem

An irreducible Markov chain has a unique invariant distribution.

FACT: So if a Markov chain has a limiting density π, then it is the unique invariant density!

Metropolis-Hastings algorithm

Theorem

If for all states $x, y \in \Omega$, we have $q(x, y) > 0$ whenever $\pi(y) > 0$, then the MH algorithm produces an irreducible Markov chain and π is its unique invariant density.

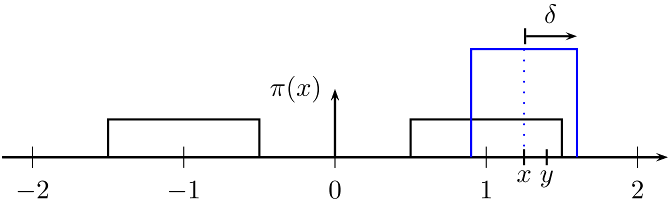
Remark: We return later to what is needed extra in order to ensure convergence of the distribution of X_t towards π. In fact, as we shall soon see, irreducibility is effectively all we need in order to use Monte Carlo estimates!

Remark: It follows that for a Gibbs sampler simulating from a positive density, that is, $\pi(x) > 0$ for all $x = (x_1, \ldots, x_k) \in \Omega = \Omega_1 \times \ldots \times \Omega_k$, we have irreducibility and so π is the unique invariant density.

Example

Consider a target density

$$\pi(x) = \frac{1}{2} \cdot 1\left[|x+1| \le \frac{1}{2}\right] + \frac{1}{2} \cdot 1\left[|x-1| \le \frac{1}{2}\right].$$



Consider a random walk Metropolis algorithm with a uniform proposal density centred at the current value:

$$q(x, y) = \frac{1}{2\delta} 1\left[|x-y| \le \delta\right].$$

Indeed this proposal density is symmetric in x and y: $q(x, y) = q(y, x)$.

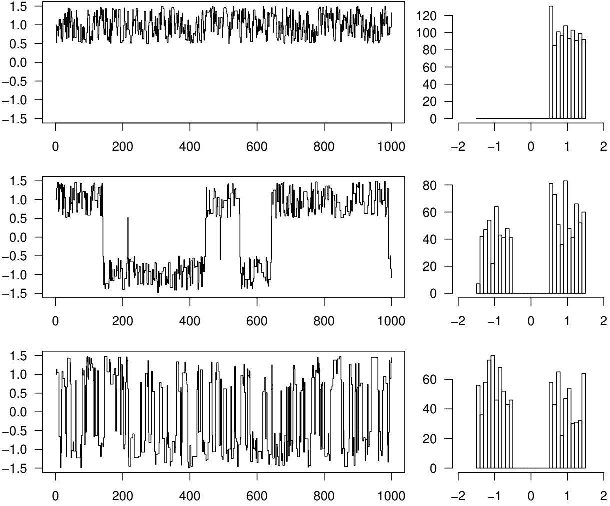
Notice: Irreducible if and only if $\delta > 1$.

Example — *cont.*

$\delta = 0.5$

$\delta = 1.2$

$\delta = 2.0$



A strong law of large numbers

Theorem: Strong law of large number for Markov chains

Consider an **irreducible** Markov chain with $\pi(x)$ as its invariant density, and a function $h : \Omega \to \mathbf{R}$ so that the **mean** $\mu = \int h(x)\pi(x)dx$ **exists**. For any $m \geq 0$ (the **burn-in**, i.e. the time we start to keep samples), define the **sample mean**

$$\hat{\mu}_n = \frac{1}{n+1} \sum_{t=m}^{m+n} h(X^{(t)}).$$

Then there exists a set $C \subseteq \Omega$ with $\int_C \pi(x)dx = 1$ so that for all $x \in C$

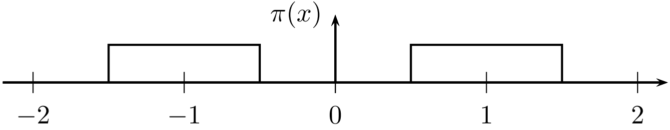
$$P(\hat{\mu}_n \to \mu \text{ as } n \to \infty \,|\, X^{(0)} = x) = 1.$$

The estimator $\hat{\mu}_n$ is a so-called **MCMC estimator** of $\mathbb{E}[h(X)]$.

We say much more about the burn-in later.

Example

- **Setup**: Assume X is distributed as before:

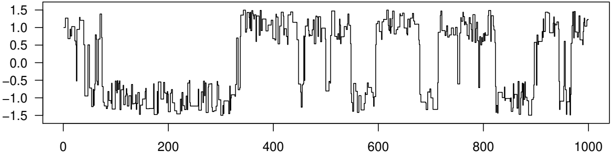


- **Question**: What is the probability $P(X \geq 0)$?
- Notice that $P(X \geq 0) = \mathbb{E}\left[\mathbb{1}[X \geq 0]\right]$ ($= \frac{1}{2}$ of course).
- Accordingly, let $h(x) = \mathbb{1}[x \geq 0]$.
- **Solution**: Generate a realization $x^{(1)}, x^{(2)}, \dots, x^{(1000)}$ of the Markov chain with a proposal density as before so that irreducibility is ensured .
- An MCMC estimate for $P(X \geq 0)$ is then

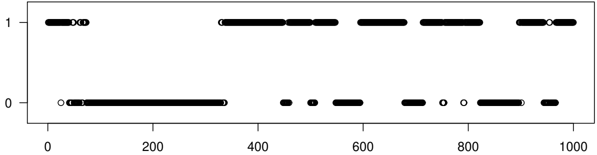
$$\hat{\mu}_{1000} = \frac{1}{1000} \sum_{i=1}^{1000} \mathbb{1}[x^{(i)} \geq 0].$$

Example *cont.*

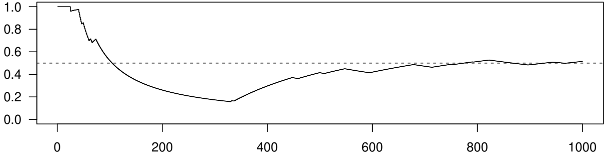
Plot of $x^{(t)}$



Plot of $h(x^{(t)})$



Plot of $\hat{\mu}_t$



Periodicity

Definition: Periodicity and aperiodicity

An irreducible Markov chain is **periodic** if there exists a partition $\Omega = A_0 \cup A_1 \cup A_2 \cup \cdots \cup A_k$ (so $A_i \cap A_j = \emptyset$ whenever $i \neq j$), where $\int_{A_0} \pi(x)dx = 0$, $k \geq 2$ and

- $x \in A_1 \Rightarrow P(x, A_2) = 1$,
- $x \in A_2 \Rightarrow P(x, A_3) = 1$,
- $\vdots$
- $x \in A_k \Rightarrow P(x, A_1) = 1$.

The Markov chain is **aperiodic** if it is not periodic.

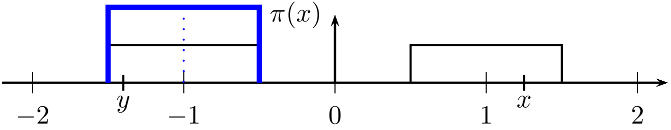
Theorem

If $P(x, \{x\}) > 0$ (that is, $X_{t+1} = X_t$ may happen with a positive probability), then it is an aperiodic Markov chain.

Example

Consider again the target density

$$\pi(x) = \frac{1}{2} \cdot 1 \left[|x+1| \leq \frac{1}{2} \right] + \frac{1}{2} \cdot 1 \left[|x-1| \leq \frac{1}{2} \right].$$



Consider the proposal density

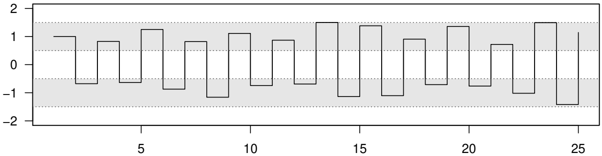
$$q(x,y) = \mathbb{1} \left[\left| y + \mathsf{sign}(x) \right| \leq \frac{1}{2} \right].$$

Accordingly:

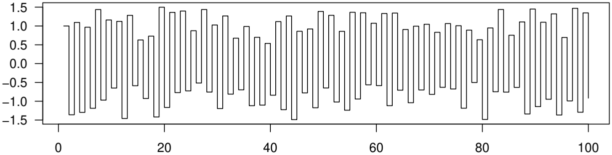
- If $x > 0$, then $y \sim$ Unif$([-1.5, -0.5])$ and $a(x,y) = 1$.
- If $x < 0$, then $y \sim$ Unif$([0.5, 1.5])$ and $a(x,y) = 1$.

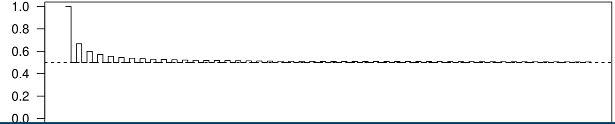
Periodicity: Example *cont.*

Thus the Markov chain is irreducible but periodic (switching between the intervals $[-1.5, -0.5]$ and $[0.5, 1.5]$):



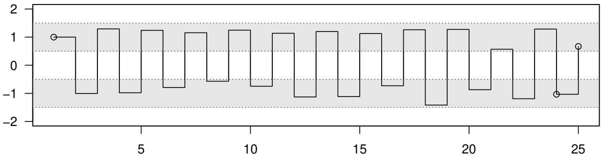
Law of large numbers still "works" (since the Markov chain is irreducible):





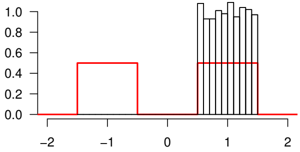
Periodicity: Example *cont.*

Since the Markov chain is periodic

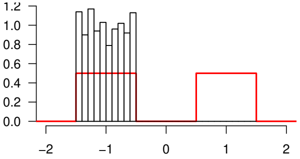


it does *not* not converge towards a limiting distribution:

1000 replicates of $x^{(24)}$ when $x^{(0)} = 1$



1000 replicates of $x^{(23)}$ when $x^{(0)} = 1$



Markov chain convergence theorem

Theorem: Markov chain convergence theorem

For an irreducible and aperiodic Markov chain with invariant density $\pi(x)$, there exists $C \subseteq \Omega$, so that $\int_C \pi(x)dx = 1$ and for all $x \in C$ and $A \subseteq \Omega$ we have

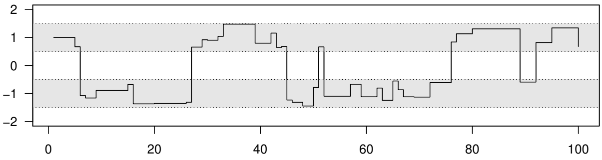
$$P(X_t \in A | X_0 = x) \to \int_A \pi(x)dx \quad \text{as } t \to \infty.$$

In other words, no matter where the chain starts (except in the "π-nullset" C), as the time t goes along, the distribution of X_t converges towards the target distribution with density π.

If the Markov chain is Harris recurrent (this technical concept is not defined in this course), then $C = \Omega$ (so no worries about if we started outside C).

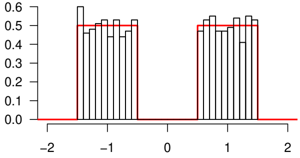
Convergence: Example

Consider the irreducible and aperiodic chain from earlier ($\delta = 2$):

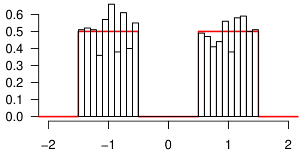


Clearly, this Markov chain *does* converge:

1000 replicates of X_{24} when $X_0 = 1$



1000 replicates of X_{23} when $X_0 = 1$



Convergence: Example *cont.*

When $X_0 = 1$, from the top left to the bottom right, 1000 replicates of $X_0, \ldots, X_{11}$, respectively: