Markov chain[1]

- Markov chain
- Stationary distribution
- Ergodicity
- Inefficiency factor and effective sample size
- Centra limit theorem
- Reversible chains

[1] Based on Gamerman and Lopes (2007) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman&Hall/CRC.

A Markov chain is a stochastic process where given the present state, past and future states are independent, i.e.

$$Pr(\theta^{(n+1)} \in A | \theta^{(n)} = x, \theta^{(n-1)} \in A_{n-1}, \ldots, \theta^{(0)} \in A_0)$$

equals

$$Pr(\theta^{(n+1)} \in A | \theta^{(n)} = x)$$

for all sets $A_0, \ldots, A_{n-1}, A \subset S$ and $x \in S$.

When the above equation does not depend on n, the chain is said to be *homogeneous* and a transition function, or kernel $P(x, A)$, can be defined as:

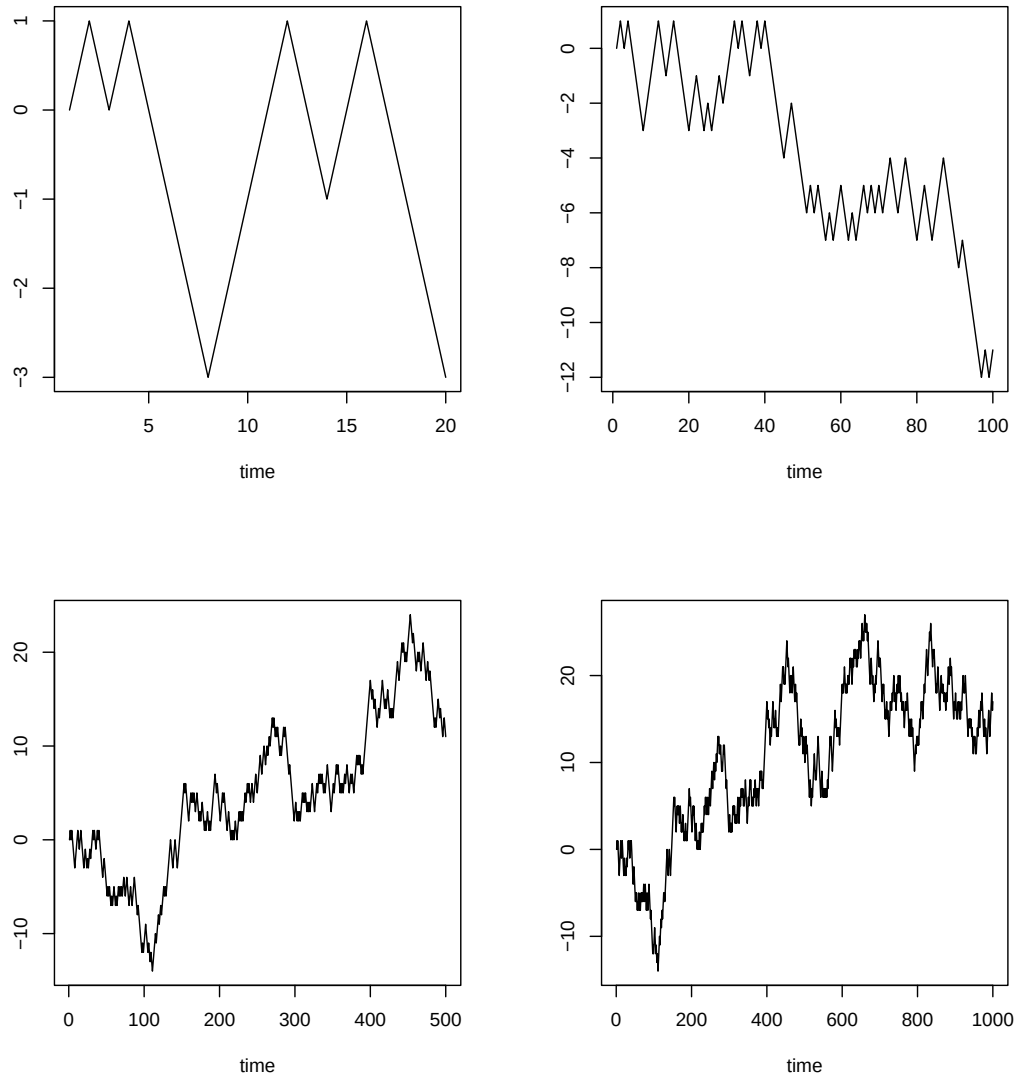
1. for all $x \in S$, $P(x, \cdot)$ is a probability distribution over S;
2. for all $A \subset S$, the function $x \mapsto P(x, A)$ can be evaluated.

Example i. random walk

Consider a particle moving independently left and right on the line with successive displacements from its current position governed by a probability function f over the integers and $\theta^{(n)}$ representing its position at instant n, $n \in N$. Initially, $\theta^{(0)}$ is distributed according to some distribution $\pi^{(0)}$. The positions can be related as

$$\theta^{(n)} = \theta^{(n-1)} + w_n = w_1 + w_2 + \ldots + w_n$$

where the w_i are independent random variables with probability function f. So, $\{\theta^{(n)} : n \in N\}$ is a Markov chain in Z.
The position of the chain at instant $t = n$ is described probabilistically by the distribution of $w_1 + \ldots + w_n$.



$Pr\{\theta^{(n)} = \theta^{(n-1)} + i\} = 1/2$, for $i = -1, 1$ and $\theta^{(0)} = 0.0$.

Discrete state spaces

If S is finite with r elements, $S = \{x_1, x_2, \ldots, x_r\}$, a transition matrix P with (i,j)th element given by $P(x_i, x_j)$ can be defined as

$$P = \begin{pmatrix} P(x_1, x_1) & \ldots & P(x_1, x_r) \\ \vdots & & \vdots \\ P(x_r, x_1) & \ldots & P(x_r, x_r) \end{pmatrix}.$$

Transition probabilities from state x to state y over m steps, denoted by $P^m(x,y)$, is given by the probability of a chain moving from state x to state y in exactly m steps. It can be obtained for $m \geq 2$ as

$$\begin{aligned} P^m(x,y) &= Pr(\theta^{(m)} = y | \theta^{(0)} = x) \\ &= \sum_{x_1} \cdots \sum_{x_{m-1}} Pr(y, x_{m-1}, \ldots, x_1 | x) \\ &= \sum_{x_1} \cdots \sum_{x_{m-1}} Pr(y|x_{m-1}) \ldots Pr(x_1|x) \\ &= \sum_{x_1} \cdots \sum_{x_{m-1}} P(x_{m-1}, y) \cdots P(x, x_1) \end{aligned}$$

Chapman-Kolmogorov equations

$$\begin{aligned} P^{n+m}(x,y) &= \sum_z Pr(\theta^{(n+m)} = y | \theta^{(n)} = z, \theta^{(0)} = x) \\ &\times \quad Pr(\theta^{(n)} = z | \theta^{(0)} = x) \\ &= \sum_z P^n(x,z) P^m(z,y) \end{aligned}$$

and (more generally)

$$P^{n+m} = P^n P^m.$$

Marginal distributions

Let

$$\pi^{(n)} = (\pi^{(n)}(x_1), \cdots, \pi^{(n)}(x_r))$$

with the initial distribution of the chain when $n = 0$. Then,

$$\pi^{(n)}(y) = \sum_{x \in S} P^n(x,y) \pi^{(0)}(x)$$

or, in matrix notation,

$$\pi^{(n)} = \pi^{(0)} P^n$$

Example ii. 2-state Markov chain

Consider $\{\theta^{(n)} : n \geq 0\}$, a Markov chain in $S = \{0, 1\}$ with $\pi^{(0)}$ given by

$$\pi^{(0)} = (\pi^{(0)}(0), \pi^{(0)}(1))$$

and transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

It is easy to see that

$$\begin{aligned} Pr(\theta^{(n)} = 0) &= (1-p)Pr(\theta^{(n-1)} = 0) + qPr(\theta^{(n-1)} = 1) \\ &= (1-p-q)^n \pi^{(0)}(0) + q\sum_{k=0}^{n-1}(1-p-q)^k \end{aligned}$$

If $p + q > 0$,

$$Pr(\theta^{(n)} = 0) = \frac{q}{p+q} + (1-p-q)^n \left(\pi^{(0)}(0) - \frac{q}{p+q}\right)$$

If $0 < p + q < 2$ then

$$\lim_{n\to\infty} Pr(\theta^{(n)} = 0) = \frac{q}{p+q} \quad \text{and} \quad \lim_{n\to\infty} Pr(\theta^{(n)} = 1) = \frac{p}{p+q}$$

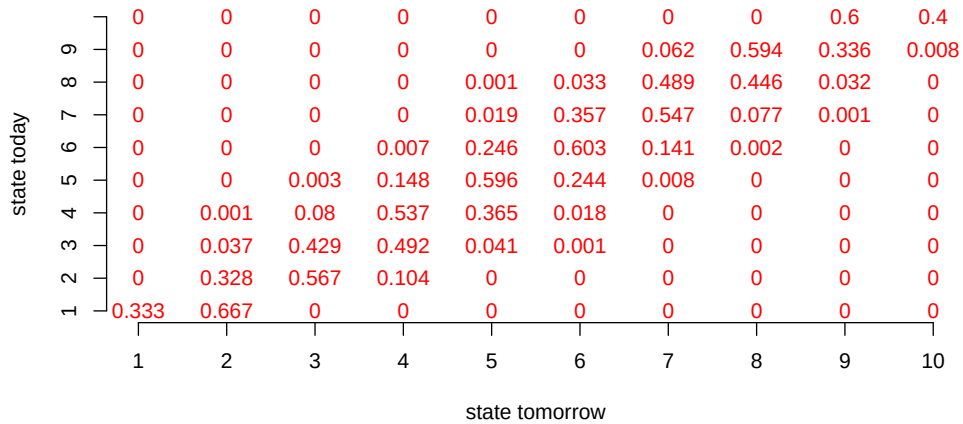
Stationary distributions

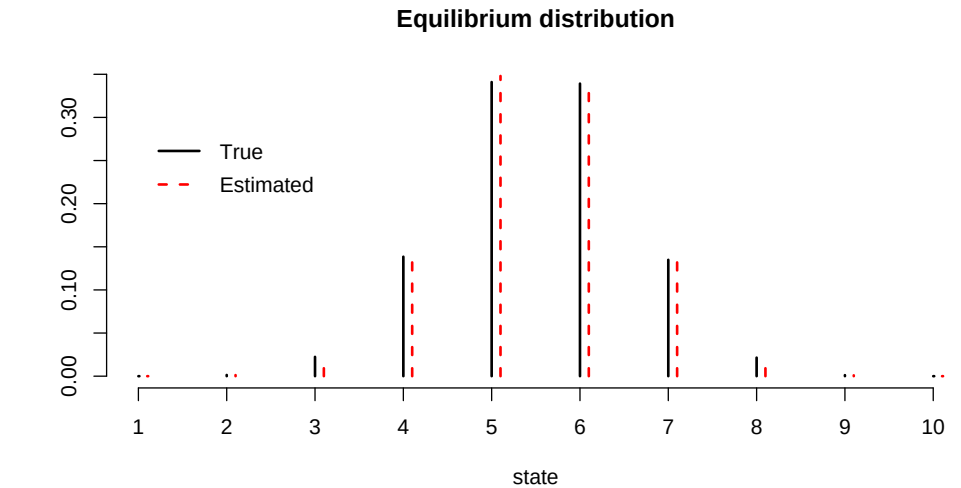
A fundamental problem for Markov chains is the study of the asymptotic behavior of the chain as the number of iterations $n \longrightarrow \infty$.
A key concept is that of a *stationary distribution* π. A distribution π is said to be a stationary distribution of a chain with transition probabilities $P(x, y)$ if

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y), \quad \forall y \in S$$

or in matrix notation as $\pi P = \pi$.
If the marginal distribution at any given step n is π then the next step distribution is $\pi P = \pi$.
Once the chain reaches a stage where π is its distribution, all subsequent distributions are π.
π is also known as the *equilibrium distribution*.

Example iii. 10-state Markov chain





Top panel: transition probabilities, $Pr(x|y)$

Bottom panel: equilibrium distribution, $\pi(x)$.

Ergodicity

A chain is said to be geometrically ergodic if $\exists \lambda \in [0, 1)$, and a real, integrable function $M(x)$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\lambda^n$$

for all $x \in S$. If $M(x) = M$, then the ergodicity is uniform.

Uniform ergodicity $\Rightarrow$ geometric ergodicity $\Rightarrow$ ergodicity.

The smallest λ satisfying the above condition is called the rate of convergence.

A very large value of $M(x)$ may slow down convergence considerably.

Ergodic theorem

Once ergodicity of the chain is established, important limiting theorems can be stated. The first and most important one is the ergodic theorem.

The ergodic average of a real-valued function $t(\theta)$ *is the average* $\bar{t}_n = (1/n)\sum_{i=1}^{n} t(\theta^{(i)})$. *If the chain is ergodic and* $E_\pi[t(\theta)] < \infty$ *for the unique limiting distribution* π *then*

$$\bar{t}_n \overset{a.s.}{\to} E_\pi[t(\theta)] \text{ as } n \to \infty$$

which is a Markov chain equivalent of the law of large numbers.

It states that averages of chain values also provide strongly consistent estimates of parameters of the limiting distribution π despite their dependence.

Inefficiency factor

Define the autocovariance of lag k of the chain $t^{(n)} = t(\theta^{(n)})$ as $\gamma_k = Cov_\pi(t^{(n)}, t^{(n+k)})$, the variance of $t^{(n)}$ as $\sigma^2 = \gamma_0$, the autocorrelation of lag k as $\rho_k = \gamma_k/\sigma^2$ and $\tau_n^2/n = Var_\pi(\bar{t}_n)$.
It can be shown that

$$\tau_n^2 = \sigma^2 \left(1 + 2\sum_{k=1}^{n-1} \frac{n-k}{n}\rho_k\right) \to \tau^2 = \sigma^2 \left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)$$

as $n \to \infty$.

The term between parentheses in the above equation can be called *inefficiency factor* or *integrated autocorrelation time* because it measures how far $t^{(n)}$s are from being a random sample and how much $Var_\pi(\bar{t}_n)$ increases because of that.

Effective sample size

The inefficiency factor can be used to derive the *effective sample size*

$$n_{\text{eff}} = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k}$$

which can be thought of as the size of a random sample with the same variance since

$$Var_{\pi}(\bar{t}_n) = \sigma^2 / n_{\text{eff}}.$$

It is important to distinguish between

$$\sigma^2 = Var_{\pi}[t(\theta)] \quad \text{and} \quad \tau^2$$

the variance of $t(\theta)$ under the limiting distribution π and the limiting sampling variance of $\sqrt{n}\,\bar{t}$, respectively.

Note that under independent sampling they are both given by σ^2. They are both variability measures but the first one is a characteristic of the limiting distribution π whereas the second is the uncertainty of the averaging procedure.

Central limit theorem

If a chain is uniformly (geometrically) ergodic and $t^2(\theta)$ ($t^{2+\epsilon}(\theta)$) is integrable with respect to π (for some $\epsilon > 0$) then

$$\frac{\bar{t}_n - E_\pi[t(\theta)]}{\tau/\sqrt{n}} \xrightarrow{d} N(0,1),$$

as $n \to \infty$.

Just as the ergodic theorem provides theoretical support for the use of ergodic averages as estimates, the above equation provides support for evaluation of approximate confidence intervals.

Reversible chains

Let $(\theta^{(n)})_{n\geq 0}$ be an homogeneous Markov chain with transition probabilities $P(x,y)$ and stationary distribution π.

Assume that one wishes to study the sequence of states $\theta^{(n)}, \theta^{(n-1)}, \ldots$ in reversed order. It can be shown that this sequence is a Markov chain with transition probabilities are

$$\begin{aligned} P_n^*(x,y) &= Pr(\theta^{(n)} = y \mid \theta^{(n+1)} = x) \\ &= \frac{Pr(\theta^{(n+1)} = x \mid \theta^{(n)} = y) Pr(\theta^{(n)} = y)}{Pr(\theta^{(n+1)} = x)} \\ &= \frac{\pi^{(n)}(y) P(y,x)}{\pi^{(n+1)}(x)} \end{aligned}$$

and in general the chain is not homogeneous.

If $n \to \infty$ or alternatively, $\pi^{(0)} = \pi$, then

$$P_n^*(x, y) = P^*(x, y) = \pi(y)P(y, x)/\pi(x)$$

and the chain becomes homogeneous.

If $P^*(x, y) = P(x, y)$ for all x and $y \in S$, the Markov chain is said to be *reversible*. The reversibility condition is usually written as

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

for all $x, y \in S$.

It can be interpreted as saying that the rate at which the system moves from x to y when in equilibrium, $\pi(x)P(x, y)$, is the same as the rate at which it moves from y to x, $\pi(y)P(y, x)$.

For that reason, the above equation is sometimes referred to as the *detailed balance equation*; *balance* because it equates the rates of moves through states and *detailed* because it does it for every possible pair of states.