UPPSALA UNIVERSITY
Department of Mathematics
© Docent ÖS

Markov Processes, 1MS012
Spring semester 2024

Lecture 8 Markov Processes, 1MS012

1 Reversible Markov chains

Let (X_n) be a (homogeneous) Markov chain with transition matrix $\mathbf{P} = (p_{jk})$. Let $m > 0$ be an integer and define $Y_n = X_{m-n}$, $0 \leq n \leq m$.

$$Y_0 = X_m, \;\; Y_1 = X_{m-1}, \;\; \ldots, Y_m = X_0.$$

The sequence $(Y_n)_{n=0}^m$ corresponds to running the Markov chain backwards in time.

We check the Markov property:

$$\begin{aligned}
&P(Y_{n+1} = k \mid Y_n = j, Y_{n-1} = y_{n-1}, ..., Y_0 = y_0) \\
&= P(X_{m-n-1} = k \mid X_{m-n} = j, X_{m-n+1} = y_{n-1}, ..., X_m = y_0) \\
&= \frac{P(X_{m-n-1} = k, X_{m-n} = j, X_{m-n+1} = y_{n-1}, ..., X_m = y_0)}{P(X_{m-n} = j, ..., X_m = y_0)} \\
&= \frac{P(X_{m-n+1} = y_{n-1}, ..., X_m = y_0 | X_{m-n} = j, X_{m-n-1} = k) P(X_{m-n} = j, X_{m-n-1} = k)}{P(X_{m-n+1} = y_{n-1}, ..., X_m = y_0 | X_{m-n} = j) P(X_{m-n} = j)} \\
&= \frac{P(X_{m-n} = j, X_{m-n-1} = k)}{P(X_{m-n} = j)} \\
&= \underbrace{P(X_{m-n} = j | X_{m-n-1} = k)}_{p_{kj}} \frac{P(X_{m-n-1} = k)}{P(X_{m-n} = j)} \\
&= p_{kj} \frac{P(Y_{n+1} = k)}{P(Y_n = j)}
\end{aligned}$$

Thus (Y_n) has the Markov property but the transition probabilities may depend on n. If π is a stationary distribution for $\mathbf{P}$ and $X_0 \overset{d}{\sim} \pi$ (i.e. X_0 is π-distributed), then

$$P(Y_{n+1} = k \mid Y_n = j) = p_{kj} \frac{\pi_k}{\pi_j} \qquad \text{(no dependence on } n\text{)}$$

Thus if we run a (homogeneous, stationary) Markov chain, (X_n), backwards in time then we get a (homogeneous stationary) Markov chain.

$$P(X_n = k \mid X_{n-1} = j, X_{n-2} = y_{n-2}, ..., X_0 = y_0) = \underbrace{p_{jk}}_{\text{tr.prob.}}$$

$$P(X_n = k \mid X_{n+1} = j, X_{n+2} = y_{n+2}, ..., X_{n+n_0} = y_{n+n_0}) = \underbrace{\frac{\pi_k p_{kj}}{\pi_j}}_{\text{tr.prob. for reversed chain}}$$

for any integer $n_0 \geq 1$.

Definition: A Markov chain is said to be **reversible** if there exist a probability distribution π such that for any states j and k

$$\underbrace{\pi_k p_{kj}}_{\text{flow out of state } k \text{ to } j} = \underbrace{\pi_j p_{jk}}_{\text{flow in to state } k \text{ from } j} \qquad \text{(detailed balance equations)}$$

Such a distribution is called a **reversible distribution** for the Markov chain.

We thus have

$$P(X_n = k \mid X_{n-1} = j) = P(X_n = k \mid X_{n+1} = j),$$

for a Markov chain starting according to a reversible distribution. Reversible Markov chains thus "looks the same" regardless of whether time runs forwards or backwards.

A reversible distribution is stationary since

$$\pi_j = \pi_j \sum_{k \in S} p_{jk} = \sum_{k \in S} \pi_j p_{jk} = \sum_{k \in S} \pi_k p_{kj}$$

A stationary distribution is not necessarily reversible:

Example: Consider a Markov chain on $S = (1, 2, 3)$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

This Markov chain is irreducible and aperiodic, and the unique stationary distribution is given by $\pi = (\pi_1, \pi_2, \pi_3) = (0.3, 0.4, 0.3)$. The distribution π is not reversible, since e.g.

$$\underbrace{\pi_1}_{=0.3}\underbrace{p_{13}}_{=0} \neq \underbrace{\pi_3}_{=0.3}\underbrace{p_{31}}_{=1/3}.$$

We have earlier seen that stationary distributions are of importence if we are interested in the long rung behaviour of a Markov chain but stationary distributions might be difficult to explicitly find when the state space is large. One approach for finding a stationary distribution is to first make a guess what a stationary distribution might be, based on intuition, and then verify that that the guessed distribution is in fact reversible.

Example: Ehrenfest model: $p_{i,i+1} = 1 - i/m$, $i = 0, 1, ..., m-1$, $p_{i,i-1} = i/m$, $i = 1..., m$, for some fixed integer $m \geq 1$.

By recalling the physics behind this model as diffusion of m molecules in two containers a natural guess of the long run distribution of the molecules is that it should be equally likely to find a given molecule in any of the two containers independently of the location of other molecules in the long run. A natural guess is therefore that the number of molecules in the first container has the binomial distribution with parameters m and $p = 1/2$ in the long run.

The $\mathrm{Bin}(m, 1/2)$-binomial distribution, with $\pi_i = \binom{m}{i}(\frac{1}{2})^m$, $i = 0, ..., m$, is reversible since

$$\binom{m}{i}(\frac{1}{2})^m \underbrace{(\frac{m-i}{m})}_{p_{i,i+1}} = \binom{m}{i+1}(\frac{1}{2})^m \underbrace{\frac{i+1}{m}}_{p_{i+1,i}}, \qquad i = 0, 1, ..., m-1.$$

(The fact that π is indeed a distribution (i.e. $\sum \pi_i = 1$) follows by applying the binomial theorem: $2^m = (1+1)^m = \sum_{i=0}^m \binom{m}{i} 1^i 1^{m-i} = \sum_{i=0}^m \binom{m}{i}$.)

1.1 Random walks on graphs

Graph: $G = (V, E)$
$V = \{v_1, .., v_{n_v}\}$ vertices
$E = \{e_1, ..., e_{n_e}\}$ edges

Each edge connects two vertices. Two vertices are called neighbors if they share an edge.

The graph is said to be connected if there is a path from any vertex to any other vertex in the graph.

A random walk on a connected graph is a Markov chain with state space V moving according to the following rules: If $X_n = v_i$, then let X_{n+1} be a neighbor of v_i chosen uniformly at random.

If d_i is the degree of vertex i, i.e. the number of neighbors of v_i then the transition probability of moving from v_i to v_j is $p_{ij} = 1/d_i$ if v_j is a neighbor of v_i, and $p_{ij} = 0$, otherwise.

The detailed balance equation now states

$$\pi_k \underbrace{p_{kj}}_{1/d_k} = \pi_j \underbrace{p_{jk}}_{1/d_j},$$

if j and k are neighbors.

Thus $\pi_k = \pi_1 \frac{d_k}{d_1}$, if the graph is connected, and in that case

$$\sum_{k=1}^{n_v} \pi_k = 1 \Leftrightarrow \pi_1 \frac{\sum_{k=1}^{n_v} d_k}{d_1} = 1 \Leftrightarrow \pi_1 = \frac{d_1}{\sum_{k=1}^{n_v} d_k},$$

which shows that

$$\pi = (\frac{d_1}{\sum_{k=1}^{n_v} d_k}, \frac{d_2}{\sum_{k=1}^{n_v} d_k}, ..., \frac{d_{n_v}}{\sum_{k=1}^{n_v} d_k}),$$

is a reversible distribution for the Markov chain.

2 Markov Chain Monte Carlo (MCMC)

Suppose we wish to pick a random element according to a distribution π.

Idea: Construct a Markov chain, (X_n), having π as its stationary (or even better, reversible) distribution.

By following a trajectory of a convergent Markov chain (X_n), we can estimate the stationary distribution π_k by the proportion of time the chain spends in state k.

We can get approximate samples from π by running the Markov chain for a long time. The random variable X_n will almost be a π-distributed random variable if n is large by the convergence assumption.

Question: How large n do we need to take in order to get a good approximation?
Depends on the (typically unknown) convergence rate. (It can be proved that irreducible aperiodic finite Markov chains converges with an exponential rate.) It is typically hard to give good explicit bounds on the convergence rates.

2.1 Creating Markov chains with a given stationary distribution

Sampling from a given distribution π might be difficult, in particular if π is multidimensional.

Metropolis-Hastings algorithm is an algorithm for creating a (reversible) Markov chain with given stationary distribution π. We only need to know π up to a constant i.e. we need to have full information about $\nu = c\pi$ where c is an unknown constant. This generalization is especially relevant when π is the posterior distribution in a Bayesian context. Let $S = \mathbb{Z}$, and let $g(i|j)$, $i, j \in S$ be an arbitrary conditional distribution. In the Metropolis algorithm we suppose that g satisfies the symmetry condition $g(i|j) = g(j|i)$. The function g is referred to as the proposal or jumping distribution. Consider a Markov chain (X_n) generated in the following way. Suppose $X_n = i$. Generate a candidate for X_{n+1} by picking from the distribution $g(j|i)$. Calculate the acceptance ratio $\alpha_{ij} = \nu_j/\nu_i = \pi_j/\pi_i$, which will be used to decide whether to accept or reject the candidate. Generate a uniform random number $u \in [0, 1]$. If $u \leq \alpha_{ij}$, then accept the candidate, i.e. set $X_{n+1} = j$. If $u > \alpha_{ij}$, then reject the candidate, i.e. set $X_{n+1} = i$ instead. (Note that we will always accept the candidate j if $\pi_j \geq \pi_i$.)

The transition probabilities for (X_n) are thus $p_{ij} = g(j|i)\min(1, \frac{\pi_j}{\pi_i})$ and this Markov chain clearly satisfies $\pi_i p_{ij} = \pi_j p_{ji}$ i.e. it is by construction reversible, with reversible distribution π.

In practise we want to choose $g(j|i)$ such that the candidate is easy to simulate. If $|S|$ is finite, then one simple choice of proposal distribution is to choose the candidate uniformly among the other states i.e. let $g(j|i) = 1/(|S| - 1)$ of any $j \neq i$.

We can however more generally often find (not necessarily symmetric) proposal distributions $g(j|i)$ and more general "acceptance probabilities", $a_{ij} = \min(1, \frac{\pi_j g_{i|j}}{\pi_i g_{j|i}})$ such that a Markov chain with transition probabilities $p_{ij} = g(j|i)a_{ij}$, $j \neq i$, is reversible with reversible distribution π with better

convergence rates. The art of finding the "best" proposal distribution and acceptance probabilities is a trade off between computational convenience and quick convergence that will not be discussed here.

The way we apply the Metropolis-Hastings algorithm also depends on the structure of the set where the given probability distribution is supported.

Example: Suppose $S = L^m$, where L is a finite set and m is a fixed positive integer. A probability distribution π on S may be regarded as the joint distribution of m random variables. (One L-valued random variable for each coordinate). Suppose we wish to pick an element of S distributed according to $\pi = (\pi_{\mathbf{j}}, \mathbf{j} \in S)$.

The following procedure defines a Markov chain (X_n) with reversible distribution π:
Let $X_n = (X_n(1), \ldots, X_n(m))$, where $X_n(k)$ denotes the L-valued random variable corresponding to the k:th coordinate of X_n, $1 \leq k \leq m$.

Gibbs sampler:
1. Pick a coordinate $j \in \{1, \ldots, m\}$ uniformly at random.
2. Pick $X_{n+1}(j)$ according to the conditional π-distribution of the value at coordinate j given that all other coordinates takes values according to X_n.
3. Let $X_{n+1}(k) = X_n(k)$ for all $k \neq j$.

This procedure generates a Markov chain with transition probabilities

$$p_{\mathbf{jk}} = \frac{1}{m} \frac{\pi_{\mathbf{k}}}{\sum \pi_{\mathbf{r}}}$$

if **k** differs from **j** in the one fixed coordinate, where the sum is taken over all configurations **r** agreeing with **j** except possibly at the given coordinate. ($p_{\mathbf{jk}} = 0$ if **k** differs from **j** in more than one coordinate, and $p_{\mathbf{jj}} = 1 - \sum_{\mathbf{k} \neq \mathbf{j}} p_{\mathbf{jk}}$.)

By construction

$$\pi_{\mathbf{j}} p_{\mathbf{jk}} = \pi_{\mathbf{k}} p_{\mathbf{kj}},$$

so the constructed Markov chain is reversible with stationary (reversible) distribution π.

This Markov chain is irreducible and aperiodic (at least if $\pi_{\mathbf{j}} > 0$ for all **j**), so the probability distribution of X_n converges to π with an exponential rate and we can thus obtain approximate samples from π by standard MCMC.

The Gibbs sampler as described above is a particular case of Metropolis-Hastings algorithm with acceptance probabilities being one.

3 Suggested exercises

Extra problems:
b2

Exercises Lawler:
7.1, 7.3

Lecture 9 Markov Processes, 1MS012

4 Markov Chain Monte Carlo (MCMC)

Recall:
Suppose we wish to pick a random element according to a distribution π.

We can get *approximate* samples from π by constructing a convergent Markov chain, (X_n), with stationary distribution π. If n is large then X_n will be almost π-distributed.

Question: How large n do we need to take in order to get a good approximation?
Depends on the (typically unknown) convergence rate. (It can be proved that irreducible aperiodic finite Markov chains converges with an exponential rate.) It is typically hard to give good explicit bounds on the convergence rates.

The following method can be used to obtain *perfect* samples from the stationary distribution of π:

4.1 The Propp-Wilson method

This is a method to generate (exact) samples from the stationary distribution π of a convergent Markov chain.
Advantages:
• No need to know bounds on the convergence rates for the Markov chain.
• The method gives non-biased samples from π.

In order to be able to describe how this method works, we recall some basic facts about Markov chains:

Theorem: Any Markov chain can be generated by random (i.i.d.) iterations of functions. In particular, if (X_n) is a Markov chain with transition matrix $\mathbf{P} = (p_{ij})$ and countable state space S then there exists a function $f : S \times [0,1] \to S$, such that if $I \sim U(0,1)$ is a random variable uniformly

distributed on the unit interval, then $P(f(i,I)=j)=p_{ij}$.

We may regard $f(i,I)$ as a random function if we write $f_I(i)=f(i,I)$ and interpret I as the random choice of function index.

Thus if $(I_n)_{n=1}^{\infty}$ is a sequence of independent random variables uniformly distributed on the unit interval, then it follows that

$$P(f_{I_n}\circ f_{I_{n-1}}\circ\cdots\circ f_{I_1}(i)=j)=p_{ij}^{(n)},$$

where $p_{ij}^{(n)}$ denotes the element on row i and column j in the matrix $\mathbf{P}^n$.

If we define

$$X_n(i)=f_{I_n}\circ f_{I_{n-1}}\circ\cdots\circ f_{I_1}(i),\quad n\geq 1,\quad X_0(i)=i$$

then $(X_n(i))$ is a Markov chain starting at the state $X_0(i)=i$.

Since the sequence (I_n) is independent and identically distributed (i.i.d.) it also follows that

$$P(f_{I_1}\circ f_{I_2}\circ\cdots\circ f_{I_n}(i)=j)=p_{ij}^{(n)},$$

but the sequence of random variables $(\tilde{X}_n(i))$ defined by

$$\tilde{X}_n(i):=f_{I_1}\circ f_{I_2}\circ\cdots\circ f_{I_n}(i),\quad n\geq 1,\quad \tilde{X}_0(i)=i$$

does not form a Markov chain.

Theorem: Let $T=\inf\{n : \tilde{X}_n(i) \text{ does not depend on } i\}$. Suppose $P(T<\infty)=1$. Then the common value $Y=\tilde{X}_T$ is a π-distributed random variable.

To see this, note that if $n\geq T$, then

$$\tilde{X}_n(i)=\tilde{X}_T(f_{I_{T+1}}\circ\cdots\circ f_{I_n}(i))=\tilde{X}_T,$$

and thus

$$P(X_n(i)=j)=P(\tilde{X}_n(i)=j)\to P(\tilde{X}_T(i)=j)=P(Y=j)=\pi_j,$$

as $n\to\infty$.

We can thus simulate from the stationary distribution π without actually calculating what it is by simulating I_1, I_2,..,I_n until

$$\tilde{X}_n(i):=f_{I_1}\circ f_{I_2}\circ\cdots\circ f_{I_n}(i)$$

does not depend on i. The simulated common value of $\tilde{X}_n(i)$ is then a realisation of a π-distributed random variable.

This method is called the Propp-Wilson perfect sampling method.

It is numerically costly to check if

$$\tilde{X}_n(i) := f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i)$$

does not depend on i if S is large.

Simplifying properties:
• constant maps:
If f_{I_k} is a constant map for some k then $\tilde{X}_n(i) := f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i)$ does not depend on i for all $n \geq k$.
• monotone maps: If $f_u(i) = f(i, u)$, are monotone maps in i (i.e. increasing or decreasing) for any fixed u, then $\tilde{X}_n(i)$ will be monotone, and, thus if $S = (1, 2, ..., n_0)$, and $\tilde{X}_k(1) = \tilde{X}_k(n_0)$, for some k, then $\tilde{X}_n(i)$ does not depend on i for any $n \geq k$.
(Sandwiching technique)

Example: (of a Markov chain that can be generated by random iterations where constant maps are chosen with positive probability)
Suppose $S = (0, 1, 2)$, and

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

We have earlier seen that a Markov chain with transition matrix $\mathbf{P}$ can be obtained by independent iterations with $w_1(x) = 0$,

$$w_2 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix},$$

$$w_3 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 2 \end{pmatrix},$$

and $w_4(x) = 2$, chosen with equal probabilities in each iteration step.

Thus we can construct a function $f : S \times [0, 1] \to S$, such that if (U_n) is a sequence of independent random variables, uniformly distributed on $[0, 1]$, then $p_{jk} = P(f_{U_n}(j) = k)$, for all $j, k \in S$, and the probability that $f_{U_n}(j)$ is a constant map (does not depend on j) is 0.5.

Let $\bar{T} = \inf\{n : f_{U_n}(i) \text{ does not depend on } i\}$. Then $\bar{T}$ is a geometrically distributed random variable with parameter 0.5, and $Y = \tilde{X}_{\bar{T}}$ is a π-distributed random variable.

Remark: In general, if

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdot & p_{1n} \\ p_{21} & \cdot & \cdot & p_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ p_{n1} & p_{n2} & \cdot & p_{nn} \end{pmatrix},$$

is a transition probability matrix for a Markov chain with a column containing only strictly positive elements, then we can similarly as above find a finite set of functions, with at least one of the functions being constant, such that random iterations with these functions will generate the Markov chain. The perfect sampling algorithm will therefore require (at most) a geometrically distributed number of uniformly distributed random variables.

Example: (of a Markov chain that can be generated by random iterations with monotone maps)
Let $S = (1, 2, ..., k)$ and $p_{i,i+1} = 1/2$, $i = 1, 2, ..., k-1$, $p_{i,i-1} = 1/2$, $i = 2, ..., k$, and $p_{11} = p_{kk} = 1/2$. It is easy to see that the uniform distribution with $\pi_i = 1/k$, $i \in S$, is a reversible (and thus stationary) distribution.

Let $f_s(x) = \begin{cases} 1 & \text{if } x = 1 \\ x - 1 & \text{otherwise} \end{cases}$, for $0 \leq s < 1/2$,

and

$$f_s(x) = \begin{cases} x + 1 & \text{if } 1 \leq x \leq k-1 \\ k & \text{if } x = k \end{cases}, \quad \text{for } 1/2 \leq s < 1.$$

All maps f_s are increasing;
$f_s(x) \leq f_s(y)$ if $x \leq y$ for any $s \in [0, 1]$.

It follows that $\tilde{X}_n$ is an increasing function.

Thus if $\tilde{X}_n(1) = \tilde{X}_n(k)$, then $\tilde{X}_n(i) = \tilde{X}_n(1)$, for any $i \in S$, i.e. $\tilde{X}_n$ is constant. (Since $\tilde{X}_n(1) \leq \tilde{X}_n(2) \leq ... \leq \tilde{X}_n(k)$.)

The above example may be regarded as a "toy"-example of how we can simulate from the stationary distribution of Markov chains that can be generated by random iterations with monotone maps. The following example is technically more involved but illustrates the same idea when simulating

from a complicated probability distribution (the Boltzmann distribution).

Example: The Ising model

Let $G = (V, E)$ be a graph. The Ising model is a way of picking a random element of the set $S = (-1, 1)^V$, i.e. of randomly assigning the values $+1$ or -1 to the vertices of G.

Physical interpretation:
Vertices: atoms in a ferromagnetical material. Each vertex then has an associated small magnet and $+1$ and -1 are two possible orientations, spins, of the magnet.

For a configuration $\xi \in \{-1, 1\}^V$ define the energy

$$H(\xi) = -\sum_{<x,y>\in E} \xi(x)\xi(y),$$

where $\xi(x)$ and $\xi(y)$ are the spins of vertices x and y, $< x, y >$ denotes an edge between x and y, and the sum is taken over all possible edges.

Thus each edge adds 1 to the energy if its endpoints have opposite spins.

The Ising model at temperature T means that we pick a configuration $\xi \in (-1, 1)^V$ according to the Boltzmann distribution

$$\pi(\xi) = \pi_T(\xi) = \frac{e^{-\frac{H(\xi)}{T}}}{\sum_{\eta\in\{-1,1\}^V} e^{-\frac{H(\eta)}{T}}}.$$

The Boltzmann distribution is of interest in physics and chemistry. It has many applications e.g. in describing gas molecule speeds. (The factor $e^{-\frac{H(\xi)}{T}}$ is called the Boltzmann-factor.)

The model thus favors configurations with low energy.

Extreme cases:
$T \approx \infty$:
Uniform distribution:
Each configuration has the same probability i.e. each vertex is assigned +1 or -1 with equal probability independently.
$T \approx 0$:

The two configurations with all vertices labeled +1 and all vertices labeled -1 are chosen with equal probability.

Recall how we can construct a Markov chain, (X_n), having a given probability distribution π as its stationary (reversible) distribution (provided the distribution is supported on a set of configurations (i.e. labellings of vertices by a finite set of numbers):

Gibbs sampler:
1. Pick a vertex $v \in V$ uniformly at random
2. Pick $X_{n+1}(v)$ according to the conditional π-distribution of the value at v given that all other vertices takes values according to X_n.
3. Let $X_{n+1}(w) = X_n(w)$ for all vertices $w \in V$ except v.

For the Ising model step 2 states:
If $X_n(w) = \xi(w)$ for all $w \neq v$ then let $X_{n+1}(v) = 1$ with probability

$$\frac{\pi(\xi^+)}{\pi(\xi^+) + \pi(\xi^-)},$$

where $\xi^+(w) := \xi^-(w) := \xi(w)$, for all $w \neq v$, and $\xi^+(v) := 1$ and $\xi^-(v) = -1$. This can be rewritten in a more explicit form as

$$\frac{\pi(\xi^+)}{\pi(\xi^+) + \pi(\xi^-)} = \frac{\pi(\xi^+)/\pi(\xi^-)}{\pi(\xi^+)/\pi(\xi^-) + 1},$$

where

$$\frac{\pi(\xi^+)}{\pi(\xi^-)} = \frac{e^{-\frac{H(\xi^+)}{T}}}{e^{-\frac{H(\xi^-)}{T}}} = e^{\frac{H(\xi^-) - H(\xi^+)}{T}} = e^{\frac{2}{T}(N_\xi^+ - N_\xi^-)},$$

where
N_ξ^+ =number of neighbors of v with $+1$ in ξ, and N_ξ^- =number of neighbors of v with -1 in ξ. (We thus add +2 for each neighbor of v labeled by +1 and subtract 2 for each neighbor of v labeled by -1.)

Thus the transition probabilities can be expressed purely in terms of computationally convenient "local" quantities.

Define the partial order on S by $\xi \preceq \eta$ if $\xi(w) \leq \eta(w)$ for all $w \in V$.

For any fixed vertex $v \in V$ and $s \in (0, 1)$, let

$$f_{(s,v)}(\xi)(v) = \begin{cases} 1 & \text{if } s < \frac{e^{\frac{2}{T}(N_\xi^+ - N_\xi^-)}}{e^{\frac{2}{T}(N_\xi^+ - N_\xi^-)} + 1} \\ -1 & \text{otherwise,} \end{cases}$$

and $f_{(s,v)}(\xi)(w) = \xi(w)$, if $w \neq v$.

By construction if $\xi \preceq \eta$ then $f_{s,v}(\xi) \preceq f_{s,v}(\eta)$.

The stationary distribution for the Markov chain obtained by random iterates of these functions is π_T.

If (U_n) is a sequence of independent random variables uniformly distributed on the unit interval and (V_n) is a sequence of independent random variables uniformly distributed on V also independent of (U_n), then if

$$f_{(U_1,V_1)} \circ \cdots \circ f_{(U_n,V_n)}(\xi_1) = f_{(U_1,V_1)} \circ \cdots \circ f_{(U_n,V_n)}(\xi_{-1}),$$

where $\xi_1(w) = 1$, for any $w \in V$ and $\xi_{-1}(w) = -1$, for any $w \in V$, then it follows that $f_{(U_1,V_1)} \circ \cdots \circ f_{(U_n,V_n)}$ is a constant map and the common value is a perfect draw from π_T.

5 Suggested exercises

Extra problems:
b1