UPPSALA UNIVERSITY
Department of Mathematics
© Docent ÖS

Markov Processes, 1MS012
Spring semester 2024

Lecture 1 Markov Processes, 1MS012

Course plan (Lawler's book)

- Introduction, Stochastic processes
- Markov chains (discrete time), Ch. 1, Ch. 2
- Markov chains (continuous time), Ch. 3
- Reversible Markov chains, Ch. 7
- Brownian motion, Ch. 8

1 Repetition

Concepts:

Random variable: Function $X : \Omega \mapsto \mathbb{R}$
(where the set Ω, called the sample space, is a set equipped with a probability measure $P = P(E)$ defined for events $E \subseteq \Omega$. The function X is assumed to be measurable i.e. sets of the form $\{\omega \in \Omega : X(\omega) \leq x\}$ are assumed to be events for any $x \in \mathbb{R}$.)

Distribution function: $F(x) = \underbrace{P(\{\omega \in \Omega : X(\omega) \leq x\})}_{P(X \leq x)}$
gives "full" information about the distribution of X.

X discrete: The **probability function**

$$p(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

gives an alternative description of the probability distribution of X.

X continuous[1]: The **density function** $f(x)$ characterize the distribution via $F(x) = \int_{-\infty}^{x} f(t)dt$.

Expectation: $\mathrm{E}(X)$ measures center of mass
Variance: $\mathrm{Var}(X)$ measures dispersal of the distribution of X.

Common distributions:
Discrete: Bernoulli, Binomial, Poisson, Geometric, Uniform (discrete)
Continuous: Normal, Exponential, Uniform

Recall concepts:
n-dimensional random variables, joint distribution function, conditional distributions, independence.

Example: Suppose (X, Y) is a two-dimensional discrete random variable characterized by the joint probability function $P(X = x, Y = y)$. Note that this implicitly implies that X and Y are discrete random variables defined on the same sample space. If X and Y are independent then

$$\begin{aligned} P(X = i \,|Y = j) \qquad &= \qquad \frac{P(X = i, Y = j)}{P(Y = j)} \\ &\underbrace{=}_{\text{independence}} \quad \frac{P(X = i)P(Y = j)}{P(Y = j)} = P(X = i). \end{aligned}$$

Covariance: $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$
Correlation: $\rho(X, Y) = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$ $\quad (-1 \leq \rho(X, Y) \leq 1)$.

2 Stochastic processes

Definition: Let T be a subset of the real line. A **stochastic process** (or **random process**) is a collection of random variables $\{X_t\}_{t\in T}$ defined on the same sample space Ω.
The set of possible values of these random variables, S, is called the **state space**;
(We sometimes write $X(t)$ instead of X_t.)

[1] More precisely: Absolutely continuous

We call t the time-parameter and will often write $(X_t)_{t\in T} = \{X_t\}_{t\in T}$ in order to stress that it is an ordered set.

We typically stress that we are actually considering a stochastic sequence when T is countable by writing e.g. $(X_n)_{n=0}^{\infty}$ rather than $(X_t)_{t\in T}$ where $T = \{0, 1, 2, ...\}$. (Similarly, if T is uncountable, we prefer writing e.g. $(X_t)_{t\geq 0}$ rather than $(X_t)_{t\in T}$ where $T = \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$.)

Obs: $X_t(\omega)$ is a real number for each fixed $\omega \in \Omega$ and $t \in T$.

The function $t \mapsto X_t(\omega)$ is called the **trajectory** (or realisation) of the random process corresponding to $\omega \in \Omega$.

It is sometimes not natural to call t a time and label states by real numbers:

Example: DNA-strings
$S = (A, G, C, T)$.
$X_n =$"nucleotide at site n".
Identify S with e.g. $S = (1, 2, 3, 4)$ and "time" with "position" in the string.

3 Distributional properties of stochastic processes

A stochastic process, (X_t) can be characterized by its finite-dimensional distributions

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, ..., X_{t_n} \leq x_n),$$

for any choice of $n \geq 1$, $t_1 < t_2 < ... < t_n$ and $x_1, ...x_n \in \mathbb{R}$, i.e. the finite-dimensional distributions specifies the process uniquely (the Kolmogorov extension theorem).

Remark: The finite-dimensional distributions thus give "full" information about all distributions related to the process. The mean function $\mu_X(t) := \mathrm{E}(X_t)$ and the covariance function $\gamma_X(r, s) := \mathrm{Cov}(X_r, X_s)$, $r, s, t \in T$ are "tools" for roughly describing second order properties of the process and are the natural generalisations of the mean and variance as tools for roughly describing the distribution of one single random variable.

It is usually complicated to describe a process by its finite-dimensional dis-

tributions.

Many processes can be described easier.

3.1 Classifications of Stochastic processes

Stochastic processes are classified by:

1. State space (size of S)
   - discrete state space (if S discrete)
   - continuous state space (otherwise)
2. Parameter set (size of T)
   - discrete time (if T discrete)
     (typically $T = \{0, 1, 2, ...\}$, $T = \{1, 2, ...\}$ or $T = \{..-1, 0, 1, 2, ...\}$)
   - Continuous time (T interval)
     (typically $T = [0, \infty)$, $T = (-\infty, \infty)$, or $T = [0, 1]$)
3. Dependence structure of $(X_t)_{t \in T}$

Example: (discrete time, discrete state space)
$(X_n)_{n=1}^{\infty}$ i.i.d. (independent and identically distributed) discrete random variables.

Such random sequences are special cases of "discrete Markov chains" discussed in Lawler, chap. 1 & 2, and in the upcoming lectures:

Definition: A random process $(X_n)_{n=0}^{\infty}$ with countable state space S (w.l.o.g. a set of integers) is called a **Markov chain** if

$$P(X_{n+1} = k \mid X_n = j, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = k \mid X_n = j),$$

holds for any $n \geq 0$ and $j, k, i_{n-1}, ..., i_0 \in S$.

Example: (discrete time, continuous state space)
$(X_n)_{n=1}^{\infty}$ i.i.d. (independent and identically distributed) continuous random variables.

Example: (Continuous time, discrete state space)
Poisson-processes: $T = [0, \infty)$, $S = \{0, 1, 2, ...\}$
(Poisson processes are special cases of "continuous-time Markov chains" discussed in Lawler, chap. 3, and in upcoming lectures.)

Example: (Continuous time, continuous state space)
Brownian motion: $T = [0, \infty)$, $S = (-\infty, \infty)$
(Discussed in Lawler, chap. 8, and in upcoming lectures.)

3.1.1 Random walks

Definition: Let $(X_n)_{n=1}^{\infty}$ be i.i.d. and let $S_n = X_1 + X_2 + ... + X_n$. Then $(S_n)_{n=1}^{\infty}$ is called a **random walk** on the line.

Note that (S_n) is a discrete time stochastic process and $S_{n+1} = S_n + X_{n+1}$.

"Each new value is obtained from the current value+something independent, regardless of the past"

Definition: A **simple random walk** is a random walk $S_n = \sum_{i=1}^{n} X_i$, with $P(X_n = 1) = p$, $P(X_n = -1) = 1 - p$, $n \geq 1$.

Suppose $S_0 = 0$. Let N be the number of returns that the walk ever makes to 0. We may write $N = \sum_{n=1}^{\infty} I_n$, where $I_n = \begin{cases} 1 & \text{if } S_n = 0 \\ 0 & \text{otherwise .} \end{cases}$

Clearly
$E(I_{2n+1}) = P(S_{2n+1} = 0) = 0$, for all n, and

$$E(I_{2n}) = P(S_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n,$$

since it is impossible to return to zero in an odd number of steps, and return to zero in an even number of steps requires half of the steps to be positive, and thus S_{2n} is zero if and only if the Binomially distributed random variable with parameters $2n$ and p counting the number of positive steps is n.

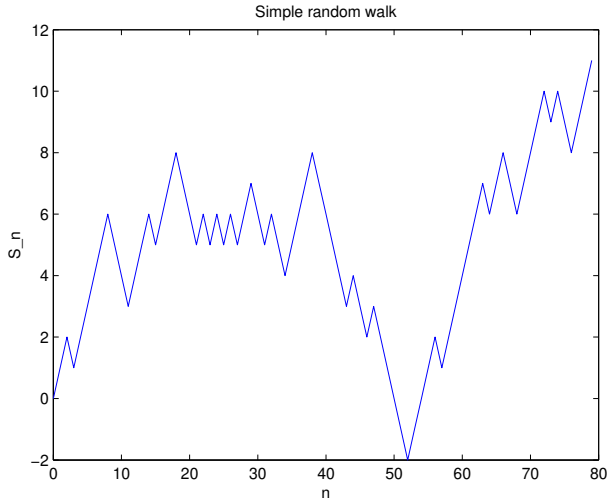


Figure 1: A trajectory of a simple random walk ($p = 1/2$)

Therefore

$$EN = E(\sum_{n=1}^{\infty} I_n) = \sum_{n=1}^{\infty} E(I_n) = \sum_{n=1}^{\infty} E(I_{2n}) = \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n = \sum_{n=1}^{\infty} \frac{2n!}{n!n!} p^n (1-p)^n.$$

By Stirling's formula

$$n! \sim \sqrt{2\pi n} n^n e^{-n},$$

i.e. $n!/(\sqrt{2\pi n} n^n e^{-n}) \to 1$, as $n \to \infty$, and thus

$$\frac{2n!}{n!n!} p^n (1-p)^n \sim \frac{\sqrt{2\pi 2n}(2n)^{2n} e^{-2n}}{(\sqrt{2\pi n} n^n e^{-n})^2} (p(1-p))^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}.$$

Therefore
$EN = \infty$, if $p = 1/2$, (i.e. if $p = 1 - p$)
$EN < \infty$, if $p \neq 1/2$, (i.e. if $p \neq 1 - p$)

Thus
$P(\text{The walk returns to } 0) = 1$ if $p = 1/2$
(since if $P(\text{The walk returns to } 0) < 1$, then N is a geometrically distributed random variable and therefore $EN < \infty$ (a contradiction)).

Similarly $P(\text{The walk returns to } 0) < 1$, if $p \neq 1/2$.

A crucial property for the above argument for the random walk is that it "starts afresh" at each moment of return to zero.
Similarly it can be proved that symmetric simple random walk in d dimensions will return to zero with probability one iff $d \leq 2$:
Kakutani (U.C.L.A. colloquium talk): "A drunk man will find his way home but a drunk bird may get lost forever."

4 Matlab

4.1 Simple random walk

$p = 0.5$;
$n = 79$; % Number of steps
$X = 2. * (\text{rand}(1, n) <= p) - 1$;
% X is a random row-vector of size n with elements +1 and -1.
$S = [0\ \text{cumsum}(X)]$;
% S_n is the cumulative sum of the n first elements of X
plot($[0 : n], S$); % Plot with [0 1 ... n] on x-axis and S_n on y-axis
xlabel('n') % Label of x-axis
ylabel('S_n') % Label of y-axis
title('Simple random walk') % Title of plot

Lecture 2 Markov Processes, 1MS012

5 Markov chains (discrete time)

Definition: A random process $(X_n)_{n=0}^{\infty}$ with countable state space S (w.l.o.g. a set of integers) is called a **Markov chain** if the Markov property

$$P(X_{n+1}=k \mid X_n=j, X_{n-1}=i_{n-1}, ..., X_0=i_0) = \underbrace{P(X_{n+1}=k \mid X_n=j)}_{{}^{n}p_{jk}},$$

holds for any $n \geq 0$ and $j, k, i_{n-1}, ..., i_0 \in S$.

Note that if $(X_n)_{n=0}^{\infty}$ is a Markov chain, then

$$P(\underbrace{X_{n+m}=x_{n+m}, ..., X_{n+1}=x_{n+1}}_{\text{future}}, \underbrace{X_n=j}_{\text{present}}, \underbrace{X_{n-1}=x_{n-1}, ..., X_0=x_0}_{\text{past}})$$

$$= \underbrace{P(\text{future}|\text{present}, \text{past})}_{P(\text{future}|\text{present})} \underbrace{P(\text{present}, \text{past})}_{P(\text{past}|\text{present})P(\text{present})}$$

for any $m > 0$. Thus

$$P(\text{future}, \text{past}|\text{present}) = P(\text{future}|\text{present})P(\text{past}|\text{present})$$

i.e. "future is conditionally independent of past given the present" for a Markov chain.

The Markov chain $(X_n)_{n=0}^{\infty}$ is said to be **time-homogeneous** if

$${}^{n}p_{jk} = \underbrace{p_{jk}}_{\text{transition probabilities}},$$

for all n (i.e. if the transition probabilities don't depend on n).

We will here only study time-homogeneous Markov chains.

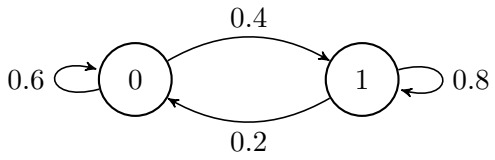
The array $\mathbf{P} = (p_{jk})_{j,k \in S}$, is called the **transition matrix** for the (time-homogeneous) Markov chain.
Note that $\sum_{k \in S} p_{jk} = 1$ for all j, and $p_{jk} \geq 0$, for all j and k, i.e. $\mathbf{P}$ is a stochastic matrix. (A stochastic matrix is a matrix with non-negative entries and row sums equal to one.)

We can illustrate a (time-homogeneous) Markov chain with a **transition diagram**, where vertices represent states, and for any two (not necessarily distinct) states j and k we draw a directed edge labeled by p_{jk} between j and k, if $p_{jk} > 0$.

Example: A Markov chain on $S = \{0, 1\}$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8. \end{pmatrix}$$

has transition diagram



□

Example: Let $(X_n)_{n\geq 0}$ be i.i.d. with $P(X_n = k) = p_k$ for all n and $k \geq 0$. Then (X_n) is a Markov chain with

$$P(X_n = k \mid X_{n-1} = j, X_{n-2} = x_{n-2}, ..., X_0 = x_0) = P(X_n = k) = p_k.$$

Thus (X_n) has transition matrix

$$\mathbf{P} = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

i.e. a transition matrix with identical rows.

Example: Simple random walk: Time-homogeneous Markov chain with $p_{j,j+1} = p$, $p_{j,j-1} = 1 - p$, $p_{jk} = 0$, otherwise.

Example: Paul and Tatiana Ehrenfest (1907)
Molecules of gas in two containers A and B
There is a tiny hole (aperture) between the containers.

At time $n = 0$:
i molecules in A.

$m - i$ molecules in B

At each time point $n = 1, 2, ...$ one of the molecules is chosen uniformly at random and moved to the other container.
Let X_n be the number of molecules in A at time $n \geq 0$.
The sequence $(X_n)_{n=0}^{\infty}$ is a Markov chain with $X_0 = i$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdot & \cdot & \cdot & p_{0m} \\ p_{10} & \cdot & \cdot & \cdot & \cdot & p_{1m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{m0} & p_{m1} & \cdot & \cdot & \cdot & p_{mm} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 1/m & 0 & 1-1/m & 0 & & \cdot\cdot \\ 0 & 2/m & 0 & 1-2/m & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1-1/m & 0 & 1/m \\ \cdot & \cdot & \cdot & 0 & 1 & 0 \end{pmatrix}$$

General question: What happens in the long run?

5.1 Chapman-Kolmogorov equations

If (X_n) is a Markov chain, then

$$P(X_0 = i_0, X_1 = i_1, ..., X_n = i_n) =$$

$$\underbrace{P(X_n = i_n | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1})}_{p_{i_{n-1} i_n}} P(X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}).$$

By applying this recursively, we obtain

$$P(X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = P(X_0 = i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}. \quad (1)$$

The distribution of a (time-homogeneous) Markov chain is thus determined by the transition matrix and the initial distribution (i.e. the distribution of X_0).
How can we calculate the distribution of X_n?
Define $p_{jk}^{(n)} = P(X_{n+m} = k | X_m = j) = P(X_n = k | X_0 = j)$. Since

$$p_{jk}^{(2)} = P(X_2 = k | X_0 = j) = \sum_{i \in S} P(X_2 = k, X_1 = i | X_0 = j)$$

$$\underbrace{=}_{\text{using (1)}} \sum_{i \in S} p_{ji} p_{ik} \leftarrow \text{element in row corresponding to state } j \text{ and column corresponding to state } k \text{ in } \mathbf{P}^2.$$

we see

$$\mathbf{P}^2 = \begin{array}{c} \\ j \end{array} \begin{array}{c} \begin{array}{c} k \end{array} \\ \begin{pmatrix} \cdots & \vdots & \cdots \\ & p_{jk}^{(2)} & \\ & \vdots & \end{pmatrix} \end{array}.$$

More generally, if $p_{jk}^{(n_0)}$ is the element in row corresponding to state j and column corresponding to state k in $\mathbf{P}^{n_0}$ for some n_0 then

$$\begin{aligned} p_{jk}^{(n_0+1)} &= P(X_{n_0+1} = k \mid X_0 = j) \\ &= \sum_{i \in S} \underbrace{P(X_{n_0} = i \mid X_0 = j)}_{p_{ji}^{(n_0)}} \underbrace{P(X_{n_0+1} = k \mid X_{n_0} = i, X_0 = j)}_{p_{ik}} \\ &= \sum_{i \in S} p_{ji}^{(n_0)} p_{ik} \leftarrow \text{element in row corresponding to state } j \text{ and} \\ &\qquad\qquad \text{column corresponding to state } k \text{ in } \mathbf{P}^{n_0+1}. \end{aligned}$$

Therefore, by induction

$$\mathbf{P}^n = \begin{array}{c} \\ j \end{array} \begin{array}{c} \begin{array}{c} k \end{array} \\ \begin{pmatrix} \vdots \\ \cdots \quad p_{jk}^{(n)} \quad \cdots \\ \vdots \end{pmatrix} \end{array}, \qquad \text{for any } n \geq 1.$$

Since $\mathbf{P}^{n+m} = \mathbf{P}^n \mathbf{P}^m$, this implies

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)} \qquad \text{(Chapman-Kolmogorov equations)}$$

We have

$$\underbrace{P(X_n = k)}_{=:\mu_k^n} = \sum_{j \in S} \underbrace{P(X_n = k | X_0 = j)}_{p_{jk}^{(n)}} \underbrace{P(X_0 = j)}_{\mu_j^0}.$$

This means, in vector notation, that the probability distribution $\boldsymbol{\mu}_n = (\mu_k^n : k \in S)$ of X_n, is an $1 \times |S|$-vector satisfying

$$\boldsymbol{\mu}_n = \boldsymbol{\mu}_0 \mathbf{P}^n.$$

Example: Gamblers ruin
Repeated bets at Casino:
Win 1 \$ with probability $p = 18/38$
Lose 1 \$ with probability $1 - p = 20/38$.
Enter the casino with k\$ ($1 \leq k \leq 4$). Leave when fortune is 0\$ or 5\$.
Let X_n be the fortune at "time" n. (At each time-step we bet if we are still in the Casino.) We have $X_0 = k$, $P(X_{n+1} = m + 1 \mid X_n = m) = p,\ m = 1, 2, 3, 4$ $P(X_{n+1} = m - 1 \mid X_n = m) = 1 - p,\ m = 1, 2, 3, 4$, and $P(X_{n+1} = 5 \mid X_n = 5) = P(X_{n+1} = 0 \mid X_n = 0) = 1$.

Thus

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{30} & p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{40} & p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{50} & p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & 0 & 0 \\ 0 & 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

Calculations using Matlab gives

$$\mathbf{P}^n \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.8398 & 0 & 0 & 0 & 0 & 0.1602 \\ 0.6618 & 0 & 0 & 0 & 0 & 0.3382 \\ 0.4640 & 0 & 0 & 0 & 0 & 0.5360 \\ 0.2442 & 0 & 0 & 0 & 0 & 0.7558 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

if n is large.
This indicates that if we for instance enter the casino with 3\$, then the probability that we will leave the casino without money is 0.464. We will verify this theoretically in a later lecture.

Example: (Model for weather)
Consider the weather during a number of days as a stochastic process, (X_n) with the only possible states $1 : sun$ and $0 : rain$.
Suppose: If it is rain today then the probability that there is rain tomorrow is 0.6, and if it is sun today then the probability that there is sun tomorrow is 0.8 (independent of the weather earlier days). Then (X_n) is a Markov chain on $S = (0, 1)$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8. \end{pmatrix}.$$

Computations in Matlab gives e.g.

$$\mathbf{P}^3 = \begin{pmatrix} 0.376 & 0.624 \\ 0.312 & 0.688 \end{pmatrix}$$

and

$$\mathbf{P}^{15} \approx \begin{pmatrix} 0.3333 & 0.6667 \\ 0.3333 & 0.6667 \end{pmatrix}$$

This shows that the probability that there is rain 3 days from now is 0.376 if it is rain today, and 0.312 otherwise, and the probability that there is rain 15 days from now is approximately 0.3333 regardless of the weather today.

We can actually calculate powers of $\mathbf{P}$ exactly using diagonalization. The matrix $\mathbf{P}$ has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.4$ with (right) eigenvectors $\nu_1 = (1,1)^T$ and $\nu_2 = (-2,1)^T$. being bases for the corresponding eigenspaces. Thus if $\mathcal{P} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}$, then $\mathcal{P}^{-1} = \begin{pmatrix} 1/3 & 2/3 \\ -1/3 & 1/3 \end{pmatrix}$, $\mathcal{P}^{-1}\mathbf{P}\mathcal{P} = D = \begin{pmatrix} 1 & 0 \\ 0 & 0.4 \end{pmatrix}$, and

$$\begin{aligned} \mathbf{P}^n &= \mathcal{P}D^n\mathcal{P}^{-1} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (0.4)^n \end{pmatrix}\begin{pmatrix} 1/3 & 2/3 \\ -1/3 & 1/3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{3}+\frac{2}{3}(0.4)^n & \frac{2}{3}-\frac{2}{3}(0.4)^n \\ \frac{1}{3}-\frac{1}{3}(0.4)^n & \frac{2}{3}+\frac{1}{3}(0.4)^n \end{pmatrix}. \end{aligned}$$

6 Matlab

6.1 The weather model

```
P=[0.6 0.4; 0.2 0.8];
mpower(P,15)
```

6.2 Gamblers ruin

```
m=5;                    % Choose level m for leaving the casino...
n=100;                  % Choose number of time-steps n...
p=18/38;                % Choose probability of winning...
P(1,1)=1;
P(m+1,m+1)=1;
for i=2:m
   P(i,i+1)=p;
   P(i,i-1)=1-p;
end
mpower(P,n)
```

7 Suggested exercises

Basic exercises:
1,2,3,4,5,6,7

Exercises Lawler:
1.1, 1.2, 1.3

Lecture 3 Markov Processes, 1MS012

8 Classification of Markov chains

Definition: State k is accessible from state j, denoted by $j \to k$, if there exists an $m \geq 0$ such that $p_{jk}^{(m)} > 0$.
If $j \to k$ and $k \to j$ then j and k is said to **intercommunicate**, and we write $j \leftrightarrow k$.
Note: A state intercommunicates with itself.

Example: $\star$ Let

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{21} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1 \end{pmatrix}.$$

By drawing a transition diagram we see that $1 \leftrightarrow 2$, $1 \to 3$, $2 \to 3$ and $3 \leftrightarrow 3$.

8.1 Equivalence classes

The state space can be divided into disjoint sets S_i such that all states in any S_i intercommunicates.

Example: $\star$ (continued)

$$\underbrace{(1,2,3\}}_{S} = \underbrace{\{1,2\}}_{S_1} \cup \underbrace{\{3\}}_{S_2}$$

Definition:
(a) A set C is **(stochastically) closed** if $p_{jk} = 0$ for all $j \in C$ and $k \notin C$.
(Intuitively: Cannot leave C if you are there.)
(b) A state j is **absorbing** if $\{j\}$ is closed.
(c) A set C is **irreducible** if $j \to k$, for all $j, k \in C$.

Definition: The Markov chain is said to be irreducible if the state space S is irreducible.

8.2 Recurrence/transience

Let $T_k = \min\{n \geq 1 : X_n = k \,|X_0 = k\}$ be the first return time to state k for a Markov chain starting in state k.

Definition: A state k is said to be
recurrent if $P(T_k < \infty) = 1$
transient if $P(T_k < \infty) < 1$

Definition: A recurrent state with
$E(T_k) = \infty$ is said to be **null-recurrent**
$E(T_k) < \infty$ is said to be **positive-recurrent**

Example: Simple random walk, with $p = 0.5$: All states are recurrent (null).

Example: $\star$ (continued)
States 1 and 2 are transient
State 3 is recurrent (pos.)

Let N be the number of revisits to state k if the Markov chain starts in k. We have:

$$\text{State } k \text{ is recurrent} \Rightarrow E(N) = \infty$$

$$\text{State } k \text{ is transient} \Rightarrow E(N) < \infty,$$

since N is geometrically distributed with parameter $p = 1 - P(T_k < \infty)$ and therefore $E(N) = 1/p$.

Since we can write $N = \sum_{n=1}^{\infty} I_n$, where $I_n = \begin{cases} 1 \text{ if } X_n = k \\ 0 \text{ otherwise} \end{cases}$, if $X_0 = k$, we get

$$E(N) = E(\sum_{n=1}^{\infty} I_n) = \sum_{n=1}^{\infty} E(I_n) = \sum_{n=1}^{\infty} \underbrace{P(X_n = k | X_0 = k)}_{p_{kk}^{(n)}}.$$

We have proved

Theorem: State k is

$$\text{transient} \iff \sum_{n=1}^{\infty} p_{kk}^{(n)} < \infty$$

Thus in particular if k is transient then $p_{kk}^{(n)} \to 0$, as $n \to \infty$.

Since a transient state can only be visited a finite number of times (with probability one) it follows that a Markov chain with finite state space has

at least one recurrent state. This is not necessarily true if the state space is infinite.

Example: The Markov chain on $S = (0, 1, 2, 3....)$ with $p_{jj+1} = 1$ has only transient states.

Corollary: If state k is recurrent and $j \leftrightarrow k$ then j is recurrent
Proof: If $j \leftrightarrow k$ then there exists integers n_1 and n_2 such that $p_{jk}^{(n_1)} > 0$ and $p_{kj}^{(n_2)} > 0$. Since $p_{jj}^{(n_1+n_2+n)} \geq p_{jk}^{(n_1)} p_{kk}^{(n)} p_{kj}^{(n_2)}$ we have

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \geq \sum_{n=1}^{\infty} p_{jj}^{(n_1+n_2+n)} \geq p_{jk}^{(n_1)} p_{kj}^{(n_2)} \underbrace{\sum_{n=1}^{\infty} p_{kk}^{(n)}}_{=\infty} = \infty$$

Conclusions:
Recurrence/transience is a class property.
All states in an irreducible Markov chain with finite state-space are recurrent.

Decomposition theorem: The state space S of a Markov chain can be written uniquely as a union of disjoint classes

$$S = \underbrace{T}_{\text{transient states}} \cup \underbrace{C_1 \cup C_2 \cup ...}_{\text{recurrent states}} ,$$

where T is the set of transient states, and C_i, $i \geq 1$ are closed irreducible sets of recurrent states.

Thus, after rearranging the order of the states, we can express the transition matrix in the form

$$\tilde{\mathbf{P}} = \left(\begin{array}{c|c} \mathbf{C} & \mathbf{0} \\ \hline \mathbf{S} & \mathbf{T} \end{array} \right)$$

where $\mathbf{C}$ is the submatrix of $\mathbf{P}$ which includes only the rows and columns corresponding to the recurrent states, and $\mathbf{T}$ is the submatrix of $\mathbf{P}$ which includes only the rows and columns corresponding to the transient states. The matrix $\mathbf{T}$ is a sub-stochastic matrix, i.e. a matrix with non-negative entries whose row sums are less than or equal to 1.

Example: Let $S = (1, 2, 3, 4)$ and

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{21} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We have $S = T \cup C_1 \cup C_2$, where $T = \{1\}$, $C_1 = \{2, 3\}$, and $C_2 = \{4\}$.
Let $S' = (2, 3, 4, 1)$. The transition matrix

$$\tilde{\mathbf{P}} = \begin{pmatrix} p_{22} & p_{23} & p_{24} & p_{21} \\ p_{32} & p_{33} & p_{34} & p_{31} \\ p_{42} & p_{43} & p_{44} & p_{41} \\ p_{12} & p_{13} & p_{14} & p_{11} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/4 & 0 & 1/4 & 1/2 \end{pmatrix} = \left(\begin{array}{c|c} \mathbf{C} & \mathbf{0} \\ \hline \mathbf{S} & \mathbf{T} \end{array}\right),$$

where $\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{S} = \begin{pmatrix} 1/4 & 0 & 1/4 \end{pmatrix}$, and $\mathbf{T} = \begin{pmatrix} 1/2 \end{pmatrix}$ has the given form.

9 Periodic and aperiodic Markov chains

Definition: Let j be a recurrent state. The **period** $d(j)$ of j is the greatest common divisor of the times at which return to j is possible, i.e.

$$d(j) = \gcd(n \geq 1 : p_{jj}^{(n)} > 0).$$

Definition: A recurrent state j is said to be **aperiodic** if $d(j) = 1$ (otherwise we call j periodic).

Definition: A Markov chain is said to be aperiodic if all states are aperiodic.

Example: Simple random walk is a periodic Markov chain with period 2. since $d(j) = \gcd(2, 4, 6, 8, ...) = 2$ for any j.
The Ehrenfest model is also a Markov chain with period 2.

"Most" Markov chains are aperiodic however...

Theorem: Any irreducible Markov chain with $p_{jj} > 0$ for some j is aperiodic.

Proof: Let k be an arbitrary state. Since the Markov chain is irreducible there exists integers n_0 and m_0 such that $p_{kj}^{(n_0)} > 0$ and $p_{jk}^{(m_0)} > 0$. Since

$$p_{kk}^{(n_0+m_0+m)} \geq p_{kj}^{(n_0)} \underbrace{p_{jj}^{(m)}}_{\geq p_{jj}^{m}} p_{jk}^{(m_0)} > 0,$$

for any $m \geq 0$ it follows that $p_{kk}^{(n)} > 0$ for any $n \geq n_0 + m_0$. Thus $d(k) = 1$ (since any other option would contradict the fact that $d(k)$ is a common

divisor of more than one prime number).

It can be proved that all states have the same period for an irreducible Markov chain, see e.g. Lawler p. 21, so the period is a class property.

Theorem: For any pair of states i and j of an irreducible aperiodic Markov chain, there exist $m_0 = m_0(i,j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq m_0$.

Proof: Since $d(j) = 1$ there exist integers $n_1, ..., n_v$ with greatest common divisor 1 such that $p_{jj}^{(n_k)} > 0$ for all $k = 1, \ldots, v$. Since any sufficiently large n, can be expressed as $n = \sum_{i=1}^{v} c_1 n_i$ for some integers $c_1, ..., c_v$, it follows that $p_{jj}^{(n)} \geq \prod_{k=1}^{v} (p_{jj}^{(n_k)})^{c_k} > 0$. If $p_{ij}^{(m)} > 0$, then $p_{ij}^{(n+m)} \geq p_{ij}^{(m)} p_{jj}^{(n)} > 0$ and thus there exists $m_0 = m_0(i,j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq m_0$.

Definition: A Markov chain is said to be **regular** if for some $n_0 < \infty$

$$p_{ij}^{(n_0)} > 0, \text{ for all } i \text{ and } j.$$

Intuitively regularity means that the Markov chain can take any value at time n_0. Regularity is stronger that irreducibility, since n_0 does not depend on i and j.
It follows from the theorem above, that any irreducible, aperiodic Markov chain with finite state space is regular.

10 Suggested exercises

Extra problems:
A1, A2

Exercises Lawler:
1.7, 1.9 abc

Lecture 4 Markov Processes, 1MS012

11 Hitting times

Definition: The **first passage time** from state j to state k for a Markov chain is

$$T_{jk} = \min(n \geq 1 : X_n = k \mid X_0 = j).$$

Recall:
j is recurrent if

$$P(\underbrace{T_{jj}}_{=:T_j} < \infty) = \sum_{n=1}^{\infty} P(T_{jj} = n) = 1$$

j is transient if

$$\sum_{n=1}^{\infty} P(T_{jj} = n) < 1.$$

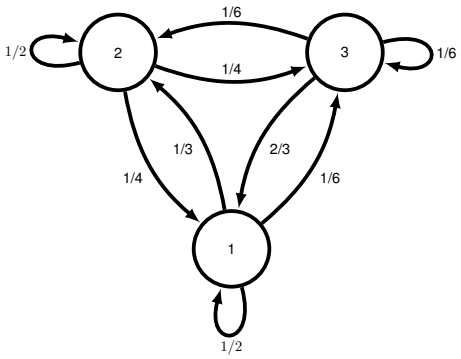
A recurrent state is positive-recurrent if $E(T_j) = \sum_{n=1}^{\infty} nP(T_{jj} = n) < \infty$.

11.1 First-step analysis

The following two examples illustrates a common "trick" in determining properties of passage times by conditioning on the outcomes of some random variable, e.g. on the outcomes of the first step of a Markov chain, in order to obtain a system of equations to solve for the quantities of interest:

Example: Let (X_n) be a Markov chain with state space $S = (1, 2, 3)$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 2/3 & 1/6 & 1/6 \end{pmatrix}.$$



Suppose we want to find, $E(T_{13})$, the expected number of steps needed to reach state 3 from state 1.
By conditioning on the outcomes of the first step, we obtain

$$E(T_{13}) = (E(T_{13})+1)\cdot p_{11}+(E(T_{23})+1)\cdot p_{12}+1\cdot p_{13} = 1+\frac{1}{2}E(T_{13})+\frac{1}{3}E(T_{23})$$

and

$$E(T_{23}) = 1 + \frac{1}{4}E(T_{13}) + \frac{1}{2}E(T_{23}),$$

i.e.

$$\begin{pmatrix} E(T_{13}) \\ E(T_{23}) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \underbrace{\begin{pmatrix} 1/2 & 1/3 \\ 1/4 & 1/2 \end{pmatrix}}_{\mathbf{T}} \begin{pmatrix} E(T_{13}) \\ E(T_{23}) \end{pmatrix},$$

and thus

$$\begin{pmatrix} E(T_{13}) \\ E(T_{23}) \end{pmatrix} = (I-\mathbf{T})^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 1.5 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 4.5 \end{pmatrix}$$

where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Remark: If we consider a Markov chain on $(1,2,3)$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix},$$

i.e. the same Markov chain as in the above example, but where we regard state 3 as an absorbing state, then we calculated the expected time to reach the absorbing state 3 starting at a given transient state, for all transient states. By rearranging the order of the states as $(3,1,2)$ we get the transition matrix

$$\tilde{\mathbf{P}} = \begin{pmatrix} 1 & 0 & 0 \\ 1/6 & 1/2 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

Then $\mathbf{T} = \begin{pmatrix} 1/2 & 1/3 \\ 1/4 & 1/2 \end{pmatrix}$ is the submatrix corresponding to the transient states, and if N_{ij} denotes the number of visits to state j if the Markov chain starts in state i, then

$$\begin{aligned} M &:= I + \mathbf{T} + \mathbf{T}^2 + ... = \begin{pmatrix} \sum_{n=0}^{\infty} p_{11}^{(n)} & \sum_{n=0}^{\infty} p_{12}^{(n)} \\ \sum_{n=0}^{\infty} p_{21}^{(n)} & \sum_{n=0}^{\infty} p_{22}^{(n)} \end{pmatrix} \\ &= \begin{pmatrix} E(N_{11}) & E(N_{12}) \\ E(N_{21}) & E(N_{22}) \end{pmatrix} = (I-\mathbf{T})^{-1} = \begin{pmatrix} 3 & 2 \\ 1.5 & 3 \end{pmatrix}. \end{aligned}$$

The expected number of steps until the chain reaches a recurrent class, here state 3, starting at a given transient state i, is the sum of the expected number of times each transient state is visited i.e. the row-sum of M corresponding to state i;
(Here $E(T_{13}) = E(N_{11}) + E(N_{12}) = 3 + 2 = 5$, and $E(T_{23}) = E(N_{21}) + E(N_{22}) = 1.5 + 3 = 4.5$.)

Example: Gamblers ruin
Repeated bets at Casino:
Win 1 \$ with probability p
Lose 1 \$ with probability $q = 1 - p$.
(E.g roulette: $p = 18/38$, $q = 20/38$)
Enter the casino with k\$ ($1 \leq k \leq N - 1$). Leave when fortune is 0\$ or N\$.
We want to find $w_k = P(\text{Leave casino with } N \text{ \$ })$.
Let X_n be the fortune at "time" n.
(At each time-step we bet in case we are still in the casino.)
The random sequence (X_n) is a Markov chain with $X_0 = k$, and
$P(X_{n+1} = m + 1 \mid X_n = m) = p$, $P(X_{n+1} = m - 1 \mid X_n = m) = q$, $m = 1, 2, ..., N - 1$, and $P(X_{n+1} = N \mid X_n = N) = P(X_{n+1} = 0 \mid X_n = 0) = 1$.
By conditioning on the outcome of the first step, we get

$$\begin{aligned} w_j &= P(\text{reach } N \text{ before } 0 \mid X_0 = j) \\ &= P(\text{reach } N \text{ before } 0 \mid X_1 = j+1, X_0 = j) P(X_1 = j+1 \mid X_0 = j) \\ &\quad + P(\text{reach } N \text{ before } 0 \mid X_1 = j-1, X_0 = j) P(X_1 = j-1 \mid X_0 = j) \\ &= w_{j+1} p + w_{j-1} q, \end{aligned}$$

for any $1 \leq j \leq N - 1$.
Since

$$w_j = \underbrace{(p+q)}_{=1} w_j = w_{j+1} p + w_{j-1} q$$

$$\Leftrightarrow$$

$$w_{j+1} - w_j = \frac{q}{p}(w_j - w_{j-1}),$$

we get

$$w_k - w_{k-1} = (\frac{q}{p})^{k-1}(w_1 - \underbrace{w_0}_{=0}) = (\frac{q}{p})^{k-1} w_1, \; 1 \leq k \leq N,$$

and therefore recursively

$$\begin{aligned} w_k &= (\frac{q}{p})^{k-1} w_1 + w_{k-1} = (\frac{q}{p})^{k-1} w_1 + (\frac{q}{p})^{k-2} w_1 + w_{k-2} \\ &= ... = w_1((\frac{q}{p})^{k-1} + (\frac{q}{p})^{k-2} + ... + (\frac{q}{p})^1 + 1), \; 1 \leq k \leq N. \end{aligned}$$

Thus, if $p = q = 1/2$, then $w_k = kw_1$, and since $w_N = 1$ it follows that

$$w_k = k/N. \tag{2}$$

If $p \neq 1/2$, then

$$w_k = w_1 \frac{1 - (\frac{q}{p})^k}{1 - \frac{q}{p}},$$

and since

$$w_N = 1 \Rightarrow w_1 = \frac{1 - \frac{q}{p}}{1 - (\frac{q}{p})^N},$$

it follows that

$$w_k = \frac{1 - (\frac{q}{p})^k}{1 - (\frac{q}{p})^N}. \tag{3}$$

Remark: In Lecture 2 we performed matlab calculations for the probabilities given in (3) in the special case when $N = 5$, $p = 18/38$ and $q = 20/38$ verifying the formula (3) above that $w_1 \approx 0.1602$, $w_2 \approx 0.3382$, $w_3 \approx 0.5360$, $w_4 \approx 0.7558$, $w_5 = 1$ in that case.
By letting $N \to \infty$ in (3) (and in (2)) we get

$$w_k \to \begin{cases} 1 - (q/p)^k & p > 1/2 \\ 0 & p \leq 1/2 \end{cases}, \qquad \text{as } N \to \infty.$$

Thus if $p > 1/2$, the game is favourable, and there is a positive probability that the gamblers fortune will increase indefinitely.
If $p \leq 1/2$ the gambler will, with probability one, go broke against an infinitely rich adversary.

12 Suggested exercises

Basic exercises:
11, 14

Extra problems:
a2, A2, A4, B1

Exercises Lawler:
1.11, 1.17

UPPSALA UNIVERSITY
Department of Mathematics
© Docent ÖS

Markov Processes, 1MS012
Spring semester 2024

Lecture 5 Markov Processes, 1MS012

1 Markov chains in the long run

Example: Let

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

From computer computations we see that

$$\mathbf{P}^n = \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix} \approx \begin{pmatrix} 0.33 & 0.67 \\ 0.33 & 0.67 \end{pmatrix}$$

if n is large. It seems to be a limiting matrix with identical rows.

Thus

$$\underbrace{(P(X_n = 0), P(X_n = 1))}_{(\pi_0^n, \pi_1^n)} = (\pi_0^0, \pi_1^0)\mathbf{P}^n \approx (0.33, 0.67),$$

if n is large, and the initial distribution seems to be "forgotten".

What can be said about the limit vector (distribution?) (π_0, π_1) if it exists?

1.1 Markov chains (finite state-space)

Suppose (X_n) is a Markov chain with finite state-space $S = (1, 2, ..., n_0)$, $n_0 < \infty$, and

$$\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k,$$

for all j and k in S.

Then

(a) $\pi = (\pi_1, \ldots, \pi_{n_0})$ is a probability distribution since

$$1 = \sum_k p_{jk}^{(n)} \Rightarrow 1 = \lim_{n\to\infty} \sum_k p_{jk}^{(n)} \underbrace{=}_{\text{since } S \text{ is finite}} \sum_k \lim_{n\to\infty} p_{jk}^{(n)} = \sum_k \pi_k$$

(b) π is a solution to the equation $\pi = \pi\mathbf{P}$, i.e. $\pi_k = \sum_j \pi_j p_{jk}$ for all $k \in S$, since

$$\begin{aligned} \pi_k &= \lim_{n\to\infty} p_{ik}^{(n)} \underbrace{=}_{\text{Chapm.Kolm.}} \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n-1)} p_{jk} \\ &\underbrace{=}_{\text{since } S \text{ is finite}} \sum_{j\in S} \underbrace{\lim_{n\to\infty} p_{ij}^{(n-1)}}_{\pi_j} p_{jk} \end{aligned} \tag{1}$$

A probability distribution satisfying $\pi = \pi\mathbf{P}$ is called a **stationary distribution** for the Markov chain since if the Markov chain starts according to π, i.e. $P(X_0 = k) = \pi_k$ for all k, then X_n is π-distributed for each n since

$$\underbrace{\boldsymbol{\mu}_n}_{\text{dist. of } X_n} = \underbrace{\boldsymbol{\mu}_0}_{\text{dist. of } X_0} \mathbf{P}^n = \pi\mathbf{P}^n = \underbrace{\pi\mathbf{P}}_{\pi}\mathbf{P}^{n-1} = \ldots = \pi.$$

More generally it follows that $(X_n)_{n=0}^{\infty}$ is a stationary sequence i.e. $(X_0, X_1, ..., X_k)$ has the same distribution as $(X_n, X_{n+1}, ..., X_{n+k})$, for any fixed n and k.

If X_n is π-distributed where π is stationary, then we sometimes say that the chain is in equilibrium.

(c) In the long run the proportion of time spent in state j is $\pi_j = \frac{1}{\mathrm{E}(T_j)}$. This can be seen by first observing that it is true if j is transient. If j is recurrent and $T_{jj}^{(n)}$ denotes the time for the nth revisit to j starting at j then by the law of large numbers

$$\frac{T_{jj}^{(n)}}{n} = \frac{T_{jj}^{(1)} + (T_{jj}^{(2)} - T_{jj}^{(1)}) + ... + (T_{jj}^{(n)} - T_{jj}^{(n-1)})}{n} \to \mathrm{E}(T_j),$$

as $n \to \infty$ (with probability one). Therefore it follows that

$$\frac{\text{number of revisits in } j \text{ up to time } T_{jj}^{(n)}}{T_{jj}^{(n)}} = \frac{n}{T_{jj}^{(n)}} \to \frac{1}{\mathrm{E}(T_j)}, \tag{2}$$

as $n \to \infty$ (with probability one).

Let $n(N)$ denote the number of revisits to state j made at time N. If we define $I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{otherwise} \end{cases}$, then,

$$\frac{n(N)}{T_{jj}^{(n(N)+1)}} \leq \frac{\sum_{n=1}^{N} I_n}{N} \leq \frac{n(N)}{T_{jj}^{(n(N))}},$$

and therefore from (2) and a "sandwich"-argument, it follows that

$$\frac{\sum_{n=1}^{N} I_n}{N} = \frac{|\{n \leq N : X_n = j\}|}{N} \to \frac{1}{\mathrm{E}(T_j)},$$

as $N \to \infty$ (with probability one).

Now

$$\mathrm{E}(\frac{\sum_{n=1}^{N} I_n}{N}) = \frac{1}{N} \sum_{n=1}^{N} \underbrace{\mathrm{E}(I_n)}_{P(X_n=j)} \to \pi_j,$$

as $N \to \infty$. Therefore $\pi_j = \frac{1}{\mathrm{E}(T_j)}$.

Example: (continued) Let (X_n) be a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}.$$

Note that (X_n) is irreducible. We want to find all stationary distributions, i.e. solutions to $\pi\mathbf{P} = \pi$, where $\pi = (\pi_1, \pi_2)$, $\pi_1 + \pi_2 = 1$, $\pi_1 \geq 0$, $\pi_2 \geq 0$:

$$\pi\mathbf{P} = \pi \Leftrightarrow (\pi_1, \pi_2) = (0.6\pi_1 + 0.2\pi_2, 0.4\pi_1 + 0.8\pi_2).$$

We thus need to solve

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 & = \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 & = \pi_2 \\ \pi_1 + \pi_2 & = 1 \\ \pi_1 \geq 0, \ \pi_2 \geq 0. & \end{cases}$$

This system has unique solution $\pi_1 = 1/3$, $\pi_2 = 2/3$, and thus $\pi = (1/3, 2/3)$ is the unique stationary distribution for the Markov chain.

We have earlier seen that $\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k$, for all j and k in S, in this example, and therefore, by property c) above, it follows that $\mathrm{E}(T_1) = 3$ and $\mathrm{E}(T_2) = 3/2$.

Example: Let

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.2 & 0.8 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix $\mathbf{P}$ is a transition matrix for a reducible (i.e. not irreducible) Markov chain.

If $\pi\mathbf{P} = \pi$ for a distribution π i.e. a vector $\pi = (\pi_1, \pi_2, \pi_3)$ with $\sum_{i=1}^{3} \pi_i = 1$, and $\pi_i \geq 0$, $i = 1, 2, 3$, then

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 & = \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 & = \pi_2 \\ \pi_3 & = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 & = 1 \\ \pi_1 \geq 0,\ \pi_2 \geq 0,\ \pi_3 \geq 0 & \end{cases}$$

This system has many solutions. Each solution is of the form

$$(\pi_1, \pi_2, \pi_3) = t(1/3, 2/3, 0) + (1-t)(0, 0, 1),$$

for some fixed $0 \leq t \leq 1$.

Thus there are many stationary distributions in this case. We cannot hope that $\lim_{n\to\infty} p_{jk}^{(n)}$ exists and is independent of j in this case since if we start in $\{1, 2\}$, then we will stay there, and similarly if we start in $\{3\}$.

The limiting behavior of $p_{jk}^{(n)}$ here thus depends on *both* j and k.

If $X_0 \in \{1, 2\}$, then we may regard this as the state space and thus $p_{j1}^{(n)} \to 1/3$, $p_{j2}^{(n)} \to 2/3$ as $n \to \infty$, for any $j \in \{1, 2\}$. Trivially, $p_{33}^{(n)} \to 1$ and $p_{3j}^{(n)} \to 0$ as $n \to \infty$, for any $j \in \{1, 2\}$, and therefore

$$\mathbf{P}^n \to \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

as $n \to \infty$.

Recall: Decomposition theorem:

$$S = T \cup C_1 \cup C_2 \cup ...,$$

where T are the set of transient states and C_i, $i \geq 1$ are closed irreducible sets of recurrent states.

If S is finite and (X_n) starts in a transient state then it will end up in some set of recurrent states in the long run, but it is random what set it will be.

1.2 Markov chains (countably infinite state space)

If S is infinite and $\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k$ for all j and k then it can happen that

$$1 = \sum_{k\in S} p_{jk}^{(n)} = \lim_{n\to\infty} \sum_{k\in S} p_{jk}^{(n)} > \sum_{k\in S} \lim_{n\to\infty} p_{jk}^{(n)} = \sum_{k\in S} \pi_k.$$

Example: Simple random walk: $\lim_{n\to\infty} p_{jk}^{(n)} = 0$ for all j and k. No stationary distribution can exist.
Recall: All states are null-recurrent for a simple random walk (with $p = 1/2$), i.e. $\mathrm{E}(T_k) = \infty$, for all k.

Theorem: If

$$\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k,$$

and $\sum_k \pi_k = 1$, then π is a stationary distribution.

Proof: In general, taking limits as in (1) gives $\pi_k \geq \sum_{j\in S} \pi_j p_{jk}$, for any $k \in S$, but if the inequality is strict for some k, then

$$1 = \sum_{k\in S} \pi_k > \sum_{k\in S} \sum_{j\in S} \pi_j p_{jk} = \sum_{j\in S} \pi_j = 1,$$

which is a contradiction. Thus $\pi_k = \sum_{j\in S} \pi_j p_{jk}$, for any $k \in S$, i.e. π is a stationary distribution.

Theorem: If (X_n) is an irreducible Markov chain then

All states are positive-recurrent $\Leftrightarrow$ A stationary distribution exists

If a stationary distribution π exists then it is unique and given by $\pi_k = \frac{1}{\mathrm{E}(T_k)}$.

Proof: ($\Rightarrow$)
Suppose k is a positive recurrent state. Let $\nu_i(k)$ be the expected number of visits of the Markov chain to state i between two successive visits to state k. That is, if $X_0 = k$, we define

$$\nu_i(k) = \mathrm{E}(\sum_{n=1}^{\infty} I(X_n = i, T_k \geq n)) = \sum_{n=1}^{\infty} P(X_n = i, T_k \geq n | X_0 = k),$$

where I denotes the indicator function. It is sufficient to check that $\nu = (\nu_i(k) : i \in S)$ satisfies $\nu P = \nu$, since by definition $\sum_i \nu_i(k) = \mathrm{E}(T_k) < \infty$, and we can therefore normalize ν and obtain a stationary probability vector, by dividing each component with $\mathrm{E}(T_k)$.

We have

$$\begin{aligned}
\nu_j(k) &= \sum_{n=1}^{\infty} P(X_n = j, T_k \geq n | X_0 = k) \\
&= \sum_{n=1}^{\infty} \sum_{i \in S} P(X_n = j, X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{n=1}^{\infty} \sum_{i \in S} \underbrace{P(X_n = j | X_{n-1} = i, T_k \geq n, X_0 = k)}_{p_{ij}} P(X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{i \in S} \sum_{n=1}^{\infty} p_{ij} P(X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{i \in S} p_{ij} \underbrace{\sum_{m=0}^{\infty} P(X_m = i, T_k > m | X_0 = k)}_{\nu_i(k)} = \sum_{i \in S} \nu_i(k) p_{ij}.
\end{aligned}$$

Thus if $\pi_i = \frac{\nu_i(k)}{\mathrm{E}(T_k)}$, then $\pi = (\pi_i : i \in S)$ is a stationary distribution.

Proof: ($\Leftarrow$)
Suppose $\pi = (\pi_k : k \in S)$ is a stationary distribution.

We first observe that all states then have to be recurrent, since if all states were transient, then $p_{jk}^{(n)} \to 0$, as $n \to \infty$, for all j and k, and then $\pi_k = \sum_{j \in S} \pi_j p_{jk}^{(n)} \to 0$, as $n \to \infty$, for all j and k which contradicts the assumption that π is a probability distribution.

Next observe that $\pi_k > 0$ for all k. We can see this by noting that if $\pi_k = 0$, for some $k \in S$, then by stationarity $0 = \pi_k = \sum_{j \in S} \pi_j p_{jk}^{(n)} \geq \pi_j p_{jk}^{(n)}$, for all j and n. Since we may choose n such that $p_{jk}^{(n)} > 0$ (irreducibility), it follows that $\pi_j = 0$ for all j which contradicts the assumption that π is a probability distribution.

It remains to prove that $\mathrm{E}(T_k)\pi_k = 1$, for all $k \in S$.

Let (X_n) be a stationary Markov chain starting according to π, i.e. $P(X_0 = k) = \pi_k$, for all $k \in S$. Then

$$
\begin{aligned}
\mathrm{E}(T_k)\pi_k &= \sum_{n=1}^{\infty} P(T_k \geq n | X_0 = k) P(X_0 = k) \\
&= \sum_{n=1}^{\infty} P(T_k \geq n, X_0 = k) \\
&= P(X_0 = k) + \sum_{n=2}^{\infty} P(T_k \geq n, X_0 = k) \\
&= P(X_0 = k) + \sum_{n=2}^{\infty} P(X_0 = k, X_m \neq k, 1 \leq m \leq n-1) \\
&= P(X_0 = k) \\
&\quad + \sum_{n=2}^{\infty} (P(X_m \neq k, 1 \leq m \leq n-1) - P(X_m \neq k, 0 \leq m \leq n-1)) \\
&= P(X_0 = k) \\
&\quad + \sum_{n=2}^{\infty} (P(X_m \neq k, 0 \leq m \leq n-2) - P(X_m \neq k, 0 \leq m \leq n-1)) \\
&\qquad\qquad \text{(by stationarity)} \\
&= P(X_0 = k) \\
&\quad + P(X_0 \neq k) + \lim_{n \to \infty} P(X_m \neq k, 0 \leq m \leq n-1) \\
&\qquad\qquad \text{(sum of telescoping serie)} \\
&= 1 \\
&\qquad\qquad \text{(using recurrence of } k\text{)}
\end{aligned}
$$

Thus $\mathrm{E}(T_k) = \frac{1}{\pi_k} < \infty$, for any $k \in S$, and the stationary distribution π is therefore unique. □

It follows from the proof above that positive recurrence is a class property i.e. if $\mathrm{E}(T_i) < \infty$ and $i \leftrightarrow j$ then $\mathrm{E}(T_j) < \infty$. Therefore in particular if S is

finite then irreducible Markov chains have only positively recurrent states.

Example: Let (X_n) be a Markov chain on $S = (0, 1, 2, \ldots)$ with $p_{i,i+1} = 1/2 = p_{i,0}$, $i \geq 0$. Then $\pi = (\pi_0, \pi_1, \pi_2, \ldots) = (1/2, 1/4, 1/8, \ldots)$ is the unique stationary distribution for (X_n). Thus if T_k denotes the first return time to state k then $E(T_k) = 1/\pi_k = 2^{k+1}$. □

The following theorem generalizes the law of large numbers in the discrete setting:

Theorem: (Law of large numbers for discrete Markov chains)
Let (X_n) be an irreducible Markov chain with state space $S = \{0, 1, 2 \ldots\}$ and suppose a stationary distribution π exists. Then π is unique and for any bounded function $f : S \to \mathbb{R}$,

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{n} \to \sum_k f(k)\pi_k$$

with probability one as $n \to \infty$.

Remark: The law of large numbers for discrete Markov chains is an example of an ergodic theorem where the "time average" $\frac{\sum_{k=0}^{n-1} f(X_k)}{n}$ of f converges to the "space average" $\sum_k f(k)\pi_k$ of f. □

Proof: The uniqueness of π, where $\pi_k = \frac{1}{\mathrm{E}(T_k)}$, $k \in S$, follows from the theorem above. By the same argument as in proving property (c) above we see that the long run the proportion of time spent in state k is π_k, for any $k \in S$, i.e. if $n_k(n)$ denotes the number of visits to state k before time n, then $\frac{n_k(n)}{n} \to \pi_k$ as $n \to \infty$, with probability one for any $k \in S$. Note that if F is a finite set then $\sum_{k \notin F} \frac{n_k(n)}{n} = 1 - \sum_{k \in F} \frac{n_k(n)}{n} = \sum_{k \notin F} \pi_k - \sum_{k \in F} (\frac{n_k(n)}{n} - \pi_k)$,

so

$$\begin{aligned}
\left|\frac{\sum_{k=0}^{n-1} f(X_k)}{n} - \sum_k f(k)\pi_k\right| &= \left|\sum_{k\in S}(\frac{n_k(n)}{n} - \pi_k)f(k)\right| \\
&\leq \sup_{k\in S}|f(k)|\sum_{k\in S}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&= \sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + \sup_{k\in S}|f(k)|\sum_{k\notin F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\leq \sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + \sup_{k\in S}|f(k)|\sum_{k\notin F}\left(\frac{n_k(n)}{n} + \pi_k\right) \\
&\leq 2\sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + 2\sup_{k\in S}|f(k)|\sum_{k\notin F}\pi_k,
\end{aligned}$$

and by chosing F large enough we can make the right hand side as small as we want by chosing n sufficiently large.

□

2 Matlab

2.1 Stationary distribution

```
P=[0.6 0.4; 0.2 0.8]
pi=null(transpose(P-eye(2)))/sum(null(transpose(P-eye(2))))
% πP = π ⇔ π(P − I) = 0 ⇔ (P − I)^t π^t = 0^t.
% Thus π^t need to be in the null-space of (P − I)^t in order for π to be
% a solution.
% A stationary distribution is a normalized such solution.

P=[0.6 0.4 0; 0.2 0.8 0; 0 0 1]
v=null(transpose(P-eye(3)))                % Gives a matrix with 2 column-
                                           % vectors spanning the null-space
```

```
pi1=v(:,1)/sum(v(:,1))
pi2=v(:,2)/sum(v(:,2))          % normalized-null-space-spanning-vectors

% Both examples above can be treated with:
v=null(transpose(P-eye(length(P))))
pi=v.*(ones(length(P),1)*(1./sum(v)))
% (In the second example the matrix pi will have column-vectors
% pi1= (0.3333 0.6667 0) and pi2= (0 0 1).)
```

3 Suggested exercises

Basic exercises:
8–10, 12–16

Extra problems:
a1

Exercises Lawler:
1.9 de, 1.18

Lecture 6 Markov Processes, 1MS012

4 Markov chain convergence theorem

Convergence theorem: Let (X_n) be an aperiodic irreducible Markov chain with transition probability matrix $\mathbf{P}$ and stationary distribution π. Then

$$p_{jk}^{(n)} \to \pi_k = \frac{1}{\mathrm{E}(T_k)} \quad \text{as } n \to \infty \tag{3}$$

for all j and k.

Thus here: $S = T \cup C_1 \cup C_2 \cup \ldots = C_1$ (i.e. no transient states and only one closed irreducible set of recurrent states)

Note that (3) does not mean that any particular trajectory converge.

Proof: We will prove the convergence theorem using a probabilistic technique known as "coupling".

Suppose (X_n) starts at $X_0 = j$.

Let (Y_n) be another Markov chain with transition probability matrix $\mathbf{P}$ starting according to the stationary distribution (i.e. with $P(Y_0 = k) = \pi_k$ for any k) and suppose (Y_n) is independent of (X_n). (Note that Y_n has distribution π for all n.)

Let $T = \min(n \geq 1 : X_n = Y_n)$.

Let

$$Z_n = \begin{cases} X_n & \text{if } T > n, \\ Y_n & \text{if } T \leq n. \end{cases}$$

Clearly X_n and Z_n have the same distribution for any fixed n.

Thus

$$P(X_n = k) = P(Z_n = k) = \underbrace{P(Z_n = k, T \leq n)}_{P(Y_n = k, T \leq n)} + \underbrace{P(Z_n = k, T > n)}_{P(X_n = k, T > n)}$$

and

$$P(Y_n = k) = \underbrace{P(Y_n = k, T \leq n)}_{P(Y_n = k, T \leq n)} + \underbrace{P(Y_n = k, T > n)}_{P(Y_n = k, T > n)}.$$

This implies that

$$\underbrace{|P(X_n = k) - P(Y_n = k)|}_{|p_{jk}^{(n)} - \pi_k|} \leq P(T > n) \to 0, \tag{4}$$

as $n \to \infty$, since $((X_n, Y_n))$ is an irreducible[1] Markov chain with stationary distribution $\pi(j,k) = \pi_j \pi_k$ and therefore positively recurrent and in particular, all states of the form (i,i), $i \in S$ (i.e. the "diagonal") will eventually be reached. We can interpret T as the time to hit the "diagonal".

Remark: If $(X_n)_{n=0}^{\infty}$ is irreducible and *periodic* with period d then we can use the convergence theorem for the aperiodic Markov chains $(X_{nd+k})_{n=0}^{\infty}$, for any fixed $k = 0, 1, ..., d-1$. Note however, that irreducibility may be lost:

Example: The Markov chain (X_n) on $S = (1,2)$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

is irreducible and periodic with period 2, and has unique stationary distribution $\pi = (1/2, 1/2)$. The Markov chain (X_{2n}) has transition matrix $\mathbf{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and is not irreducible. Any probability distribution on S is stationary for (X_{2n}).

5 Branching processes

Branching processes are processes used in modeling how sizes of a population grows in generations. Let (X_n) be a stochastic sequence, where we regard X_n as the size of a population at generation n. Consider the following simple model:

Assumption: Suppose each individual in generation n, gives birth to a family of offspring of random size in generation $n+1$, where the offspring sizes

[1]Irreducibility here means that $(i_0, i_1) \to (j_0, j_1)$ for any states i_0, i_1, j_0, and j_1. We know by irreducibility of the individual chains that there exist numbers k and l such that $p_{i_0 j_0}^{(k)} > 0$, and $p_{i_1 j_1}^{(l)} > 0$. By aperiodicity we may choose $l = k$ (See Lecture 3) which means that $(i_0, i_1) \to (j_0, j_1)$. Note that we have exponentially quick convergence rate in (4) in case S is finite, since by irreducibility of $((X_n, Y_n))$ we see that there exists a $p > 0$ and a positive integer n_0 such that $P(T > nn_0) \leq (1-p)^n$.

are independent and identically distributed non-negative integer valued random variables. (We also tacitly assume that an individual only lives for one generation so an individual dies immediately after its family of offspring has been created.)

Let p_m denote the probability that an individual has m offspring.

Then (X_n) is a Markov chain and

$$P(X_{n+1} = j | X_n = k) = P(Y_{n,1} + \ldots + Y_{n,k} = j),$$

where $Y_{n,1}, \ldots, Y_{n,k}$ are independent and identically distributed random variables with $P(Y_{n,j} = m) = p_m$, for all n, j and $m \geq 0$.

Let $D_n = P(X_n = 0 | X_0 = 1)$ denote the extinction probability by the n:th generation for a line starting with one individual, and let

$$D = \lim_{n \to \infty} D_n$$

denote the ultimate extinction probability of the line. (The limit exists since (D_n) is a non-decreasing sequence bounded from above.)

If there are j offspring in the first generation, then to die out by the nth generation, each of these lines must die out in $n-1$ generations. Since they proceed independently, this probability is D_{n-1}^j.

Thus

$$D_n = p_0 + p_1 D_{n-1} + p_2 D_{n-1}^2 + p_3 D_{n-1}^3 + \ldots = G(D_{n-1}), \tag{5}$$

and thus, by taking limits, $D = G(D)$, where

$$G(s) = \sum_{k=0}^{\infty} p_k s^k = p_0 + p_1 s + p_2 s^2 + \ldots$$

denotes the probability generating function of the offspring distribution.

Note that if $p_0 > 0$ and $p_0 + p_1 < 1$, then $G(s) > 0$ for $s \geq 0$ and

$$G'(s) = p_1 + 2p_2 s + 3p_3 s^2 + \ldots \geq 0,$$

and

$$G''(s) = 2p_2 + 6p_3 s + \ldots > 0,$$

so G is increasing and convex. By definition $G(1) = 1$ and $G'(1) = \mu$ where μ denotes the expected number of offspring of an individual.

Thus all positive solutions to $G(x) = x$ must be larger than or equal to one if and only if $\mu \leq 1$, and consequently $D = 1$ if $\mu \leq 1$.

If $\mu > 1$ then there exist a solution to $G(x_0) = x_0$ with $x_0 < 1$. We claim that D is the smallest solution to $G(x) = x$, i.e. $x_0 = D$. In order to prove the claim note first that $D_0 = 0$, and thus we see from (5), and the non-decreasing property of G on $[0, 1]$, that $D_n = G^n(0) \leq G^n(x_0) = x_0$. The claim now follows by taking limits.

Let us call the offspring distribution non-degenerate if $p_1 < 1$. We have proved the following theorem:

Theorem: Suppose the offspring distribution is non-degenerate with probability generating function G. If $D = \lim_{n\to\infty} P(X_n = 0 \mid X_0 = 1)$ denotes the ultimate extinction probability, then

$$D = 1 \quad \Leftrightarrow \quad \mu \leq 1,$$

and if $\mu > 1$, then D is the unique solution to the equation $D = G(D)$, with $0 < D < 1$.

Thus the population will eventually die out if the expected number of offspring of an individual is less than or equal to one, and there is a positive probability that the population does not die out otherwise.

Remark: In the degenerated case when $p_1 = 1$, then obviously $D = 0$. More generally $D > 0$ if and only if $p_0 > 0$.

Remark: Equation (5) can be generalized, and we can express the distribution of $(X_n|X_0 = 1)$ in terms of G; Let $G_{X_n}(s) = \mathrm{E}(s^{X_n}|X_0 = 1)$ denote the probability generating function of X_n, given $X_0 = 1$. Arguing in a similar way as in (5) we get

$$\begin{aligned} G_{X_n}(s) &= \mathrm{E}(s^{X_n}|X_0 = 1) = \sum_{k=0}^{\infty} p_k \mathrm{E}(s^{X_n}|X_1 = k) = \sum_{k=0}^{\infty} p_k (G_{X_{n-1}}(s))^k \\ &= G(G_{X_{n-1}}(s)) = \ldots = G^n(s), \end{aligned}$$

where we, in the last step, used that $G_{X_1}(s) = G(s)$. Thus in particular we get

$$\mathrm{E}(X_n) = G'_{X_n}(1) = (G^n)'(1) = G'(\underbrace{G^{n-1}(1)}_{1}) \underbrace{(G^{n-1})'(1)}_{\mathrm{E}(X_{n-1})} = \mu \mathrm{E}(X_{n-1}) = \ldots = \mu^n.$$

6 Appendix

The Markov chain convergence theorem can also be proved using non-probabilistic techniques and in particular, if the state space is finite, as a corollary of the Perron-Frobenius theorem of linear algebra;

6.1 The Perron–Frobenius theorem

Theorem: Let A be a matrix with non-negative entries such that A is primitive, i.e. A^n has only positive entries for some n. Then

(a) There is an eigenvalue $r > 0$ of A with algebraic multiplicity 1 such that $|\lambda| < r$ for any other eigenvalue λ of A.

(b) There exists a unique positive *probability vector* π satisfying $\pi A = r\pi$, and a unique positive vector v such that $Av^T = rv^T$ with $\pi v^T = 1$.

(c) $\lim_{k\to\infty} A^k/r^k = v^T\pi$.

Remark: If we extend the definition of irreducibility and period for transition matrices to general non-negative matrices in the natural way, then it can be proved that a non-negative matrix A is irreducible and aperiodic if and only if A is primitive.

If $\mathbf{P}$ is a transition matrix then the row sums are equal to one and thus $v^T = \mathbf{1}^T$ (the column vector containing only ones) is an eigenvector corresponding to the eigenvalue $r = 1$. As a corollary of the Perron-Frobenius theorem we thus get:

Corollary: Let $\mathbf{P}$ be the transition matrix of an irreducible aperiodic Markov chain with finite state space. Then

(a) $r = 1$ is an eigenvalue of $\mathbf{P}$ with algebraic multiplicity 1, and $|\lambda| < 1$ for any other eigenvalue λ of $\mathbf{P}$.

(b) There exists a unique positive *probability vector* π satisfying $\pi\mathbf{P} = \pi$.

(c) $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{1}^T\pi$.

We will not prove the core part ((a), and (b)) of the Perron-Frobenius theorem here, but an outline of proof is given in Lawler, exercise 1.20.

All real matrices cannot be diagonalized, but if A is an irreducible and aperiodic non-negative matrix, then there exist an invertible matrix $\mathcal{P}$ such that $\mathcal{P}^{-1}A\mathcal{P} = J$, where the first row vector of $\mathcal{P}^{-1}$ is π, the first column vector of $\mathcal{P}$ is v^T, and where J is a block diagonal matrix with

$$\lim_{k\to\infty} J^k/r^k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Thus

$$\lim_{k\to\infty} A^k/r^k = \lim_{k\to\infty} (\mathcal{P}J^k\mathcal{P}^{-1})/r^k = \mathcal{P}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}\mathcal{P}^{-1} = v^T\pi,$$

explaining the main idea why (a) and (b) implies (c).

7 Suggested exercises

Extra problems:
a3

Exercises Lawler:
1.12, 2.2, 2.3, 2.8

Lecture 7 Markov Processes, 1MS012

8 Markov chains generated by random iterations of functions

Example: Simple random walk on $S = \mathbb{Z}$ can be generated by random iterations of functions where the function $w_1(x) = x + 1$ is chosen with probability p and the function $w_2(x) = x - 1$ is chosen with probability $1-p$. More precisely, if (I_n) is a sequence of independent random variables with $P(I_n = 1) = p$ and $P(I_n = 2) = 1-p$ and we define $X_{n+1} = f_{I_{n+1}}(X_n)$, i.e.

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

then (X_n) is a simple random walk.

In general, random independent iterations of functions from a set S to S generates a Markov chain with state space S.

Example: $S = (0, 1, 2)$. Random iterations with $w_1(x) = 0$ chosen with probability p_1, $w_2(x) = x$ chosen with probability p_2, and $w_3(x) = 2$ chosen with probability p_3, where $p_1 + p_2 + p_3 = 1$, $p_i \geq 0$, generates a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_1 + p_2 & 0 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & 0 & p_2 + p_3 \end{pmatrix}.$$

Conversely, we can also generate any (time-homogeneous) Markov chain by i.i.d. iterations of functions. We use this property when we want to simulate a trajectory of a Markov chain.

Example: Suppose $S = (0, 1, 2)$, and

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

Most computers can generate sequences of numbers $(U_n)_{n\geq 1}$ that more or less well behaves like a sequence of independent $U(0,1)$-distributed random variables.

At time-step n we use U_{n+1} to perform the transitions.

If $0 < U_{n+1} < 1/2$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 0$,
If $1/2 \leq U_{n+1} < 3/4$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 1$,
If $3/4 \leq U_{n+1} < 1$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 2$,

If $0 < U_{n+1} < 1/4$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 0$,
If $1/4 \leq U_{n+1} < 3/4$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 1$,
If $3/4 \leq U_{n+1} < 1$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 2$,

If $0 < U_{n+1} < 1/4$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 0$,
If $1/4 \leq U_{n+1} < 1/2$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 1$,
If $1/2 \leq U_{n+1} < 1$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 2$.

By construction $X_{n+1} = f_{U_{n+1}}(X_n)$, i.e.

$$X_n = (f_{U_n} \circ f_{U_{n-1}} \circ \cdots \circ f_{U_1})(X_0)$$

where (U_n) is a sequence of independent random variables uniformly distributed on $(0,1)$, and $p_{jk} = P(f_{U_n}(j) = k)$, for all $j, k \in S$.

Note that the functions $f_u, u \in (0,1)$ are only of 4 different types in this case. Let

$$g = \begin{pmatrix} 0 & 1 & 2 \\ i & j & k \end{pmatrix},$$

denote a function $g : \{0,1,2\} \to \{0,1,2\}$ with $g(0) = i$, $g(1) = j$ and $g(2) = k$.

If

$$w_1 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

i.e. $w_1(x) = 0$,

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

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

$$w_4 = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 2 & 2 \end{pmatrix},$$

i.e. $w_4(x) = 2$, then random (i.i.d.) iterations with these functions (chosen with equal probabilities) generates a Markov chain with transition matrix $\mathbf{P}$.

The above examples illustrates and can in a straightforward manner be generalized to a proof of the following theorem:

Theorem: (X_n) is a (time-homogeneous) Markov chain $\Leftrightarrow$ (X_n) can be generated by random (i.i.d.) iterations of functions

Remark: If the state space of (X_n) is uncountable, then we may regard the statement of this theorem as the definition of a time-homogeneous Markov chain.

8.1 Fractals

Fractals are sets with self-similarity properties. Many fractals can be obtained by a recursive construction or by random iterations with affine maps.

Example: Sierpinski triangle: $S =$ a triangle in $\mathbb{R}^2$
Pick three points with coordinates A, B, C in the plane building the vertices of a triangle. Consider a Markov chain starting at $X_0 = A$ where, in each step, if $X_n = x$ we let

$$\begin{aligned} X_{n+1} &= f_A(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } A \\ &= x + \frac{1}{2}(A - x) = \frac{x}{2} + \frac{A}{2}, \end{aligned}$$

with probability $1/3$,

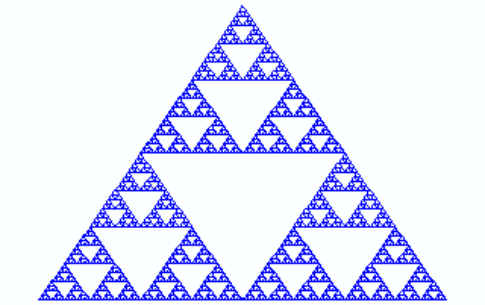
$$\begin{aligned} X_{n+1} &= f_B(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } B \\ &= x + \frac{1}{2}(B - x) = \frac{x}{2} + \frac{B}{2}, \end{aligned}$$

with probability $1/3$,

$$\begin{aligned} X_{n+1} &= f_C(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } C \\ &= x + \frac{1}{2}(C - x) = \frac{x}{2} + \frac{C}{2}, \end{aligned}$$

with probability $1/3$.

If (I_n) is a sequence of independent random variables uniformly distributed on (A, B, C), then $X_{n+1} = f_{I_{n+1}}(X_n)$. The only reachable points are points on the *Sierpinski triangle*. A random trajectory will, with probability one, "draw" the Sierpinski triangle.



The Sierpinski triangle

Example: Cantor set: $S = [0, 1]$
The Markov chain (X_n) generated by independent random iterations with the functions $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$ chosen with equal probabilities (starting at e.g. $X_0 = 0$) generates points on the middle-third *Cantor set*. A random trajectory will, with probability one, "draw" the Cantor set.

The first 5 sets approximating the limiting Cantor set. If (X_n) is a Markov chain on $S = [0, 1]$ generated by i.i.d. iterations with the maps $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$, then the value of X_n can be found in the $n : th$ approximating set of the Cantor set.

Example: Barnsley's fern: $S = \mathbb{R}^2$
The Markov chain (X_n) generated by independent random iterations with the functions

$$f_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$f_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix},$$

$$f_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.2 & -0.26 \\ 0.23 & 0.22 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix},$$

$$f_4\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0.44 \end{pmatrix},$$

(starting at e.g. the origin) generates points on *Barnsley's fern*. A random trajectory will, with probability one, "draw" Barnsley's fern. Barnsley suggested to choose f_1 with probability 0.01, f_2 with probability 0.85, f_3 with probability 0.07, and f_4 with probability 0.07, in order to make this algorithm for drawing Barnsley's fern quick.

Barnsley's fern

9 Matlab

9.1 Simulating a Markov chain with a given transition matrix

```
n=79                                    % Number of steps to be simulated
P=[1/2 1/4 1/4; 1/4 1/2 1/4; 1/4 1/4 1/2];
startstate=1;                           % Choose initial state
MCsim(n,P,startstate)


% In the textfile MCsim.m ...
function MCsim(n,P,startstate)
i=0;
j=startstate;
x=[j];
```

```
while i< n
   j=f(j,rand,P);
   x=[x j];
   i=i+1;
end;
x
end


function y=f(x,u,P)
r=[0 cumsum(P(x,:))];
for j=1:length(P)
   if ((u>r(j)) & (u<=r(j+1)));
      y=j;
   end
end
end
```

9.2 Simulating a trajectory of a Markov chain drawing the Sierpinski triangle

```
n=50000;              % Number of steps to be simulated
A1=[1/2 0; 0 1/2];
A2=[1/2 0; 0 1/2];
A3=[1/2 0; 0 1/2];
b1=[0;0];
b2=[1/2;0];
b3=[1/4;sqrt(3)/4];
p1=1/3;
p2=1/3;
p3=1/3;
startstate=[0;0];
x=startstate;
trajectory=[x];       % We store the trajectory in the
                      % matrix ``trajectory''
for i = 1:n
  u=rand;
  if u<p1
     x=A1*x+b1;
  elseif u<p1+p2
     x=A2*x+b2;
  else
     x=A3*x+b3;
  end
trajectory=[trajectory x];
end
plot(trajectory(1,:),trajectory(2,:),'.')
                      % Draws the Sierpinski triangle
title('The Sierpinski triangle')
axis off;   %  To avoid printing the coordinate axes
```

9.3 Simulating a trajectory of a Markov chain drawing Barnsley's fern

```
n=50000;   % Number of iterations
A1=[0 0;0 0.16];
A2=[0.85 0.04;-0.04 0.85];
A3=[0.2 -0.26;0.23 0.22];
A4=[-0.15 0.28;0.26 0.24];
b1=[0;0];
b2=[0;1.6];
b3=[0;1.6];
b4=[0;0.44];
p1=0.01;
p2=0.85;
p3=0.07;
p4=0.07;
startstate=[0;0];
x=startstate;
trajectory=[x];
for i = 1:n
  u=rand;
  if u<p1
     x=A1*x+b1;
  elseif u<p1+p2
     x=A2*x+b2;
  elseif u<p1+p2+p3
     x=A3*x+b3;
  else
     x=A4*x+b4;
  end
trajectory=[trajectory x];
end
plot(trajectory(1,:),trajectory(2,:),'g.')
                    % Draws a green fern
title('Barnsleys fern')
axis off;
```

10 Suggested exercises

Basic exercises:
17

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

UPPSALA UNIVERSITY
Department of Mathematics

Markov Processes, 1MS012
Spring semester 2024

Lecture 10 Markov Processes, 1MS012

1 Markov processes (continuous time)

Definition: A random process $(X_t)_{t\geq 0}$ with countable state space (w.l.o.g. a set of integers) is a (discrete) **Markov process** if, for any n and time-points $t_1 < ... < t_{n+1}$ and states $j_1, ..., j_{n+1}$

$$P(X_{t_{n+1}} = j_{n+1} \mid X_{t_n} = j_n, X_{t_{n-1}} = j_{n-1}, ..., X_{t_1} = j_1)$$
$$= P(X_{t_{n+1}} = j_{n+1} \mid X_{t_n} = j_n).$$

Thus the "future" is conditionally independent of the "past" given the present.

Time-homogeneous Markov process:

$$p_{ij}(t) = P(X_t = j \mid X_0 = i) = P(X_{t+s} = j \mid X_s = i),$$

for all $t, s \geq 0$.

Observe: $p_{ii}(0) = 1$ and $p_{ij}(0) = 0$ for all $i \neq j$, so if

$$\mathbf{P}(t) = \begin{pmatrix} \ddots & \vdots & \cdot \\ \dots & p_{ij}(t) & \dots \\ \cdot & \vdots & \ddots \end{pmatrix},$$

then $\mathbf{P}(0) = I$ (the identity matrix). The matrices $\mathbf{P}(t)$ are called the **matrices of transition probabilities**. These matrices are stochastic matrices i.e. they have non-negative entries and row-sums $= 1$.

Example: Poisson process $(N_t)_{t\geq 0}$, with intensity parameter $\lambda > 0$: Stochastic process on $S = \{0, 1, 2, \dots\}$ starting at $N_0 = 0$ with

$$P(N_{t_{n+1}} = j_{n+1} \mid N_{t_n} = j_n, N_{t_{n-1}} = j_{n-1}, ..., N_{t_1} = j_1)$$

$$\underbrace{:=}_{\text{independent increments}} \qquad P(N_{t_{n+1}} - N_{t_n} = j_{n+1} - j_n)$$

$$\underbrace{:=}_{N_t - N_s \sim Po(\lambda(t-s)),\, 0 \le s < t} \qquad e^{-\lambda(t_{n+1}-t_n)} \frac{(\lambda(t_{n+1}-t_n))^{j_{n+1}-j_n}}{(j_{n+1}-j_n)!},$$

if $j_k \le j_{k+1}$, for all $k = 1, 2, ..., n$.
($P(N_{t_{n+1}} = j_{n+1} \mid N_{t_n} = j_n, N_{t_{n-1}} = j_{n-1}, ..., N_{t_1} = j_1) = 0$, otherwise.)
(With $N_t - N_s \sim Po(\lambda(t-s)$ we thus mean that the increment $N_t - N_s$ has the Poisson distribution with intensity parameter $\lambda(t-s)$.)

Thus (N_t) is a Markov process with

$$p_{ij}(t) = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, & j \ge i \\ 0, & j < i \end{cases}$$

Chapman-Kolmogorov equations:

$$\underbrace{P(X_{t+s} = k \mid X_0 = i)}_{p_{ik}(s+t)} = \sum_{j \in S} \underbrace{P(X_s = j \mid X_0 = i)}_{p_{ij}(s)} \underbrace{P(X_{t+s} = k \mid X_s = j, X_0 = i)}_{p_{jk}(t)}$$

If we express the Chapman-Kolmogorov equations in matrix form we get:

$$\underbrace{\begin{pmatrix} \ddots & \vdots & . \\ \dots & p_{ik}(s+t) & \dots \\ . & \vdots & \ddots \end{pmatrix}}_{\mathbf{P}(s+t)} = \underbrace{\begin{pmatrix} \ddots & \vdots & . \\ p_{i1}(s) & p_{i2}(s) \dots & p_{i.}(s) \\ . & \vdots & \ddots \end{pmatrix}}_{\mathbf{P}(s)} \underbrace{\begin{pmatrix} \ddots & p_{1k}(t) & . \\ \dots & p_{2k}(t) & \dots \\ . & p_{.k}(t) & \ddots \end{pmatrix}}_{\mathbf{P}(t)}$$

Note:
For Markov chains (with discrete time): If we know the transition matrix $\mathbf{P}$, then we can calculate $\mathbf{P}^n$, for all n.

For Markov processes (with continuous time): If we know the matrix $\mathbf{P}(t)$, for some t, then we cannot calculate $\mathbf{P}(t)$ for all t. (only for $\mathbf{P}(t), \mathbf{P}(2t)$,$\mathbf{P}(3t), ...$)

If we let $t \to 0$ and note that

$$\frac{p_{ij}(t) - p_{ij}(0)}{t} = \begin{cases} \frac{p_{ij}(t)}{t} & i \ne j \\ \frac{p_{ij}(t)-1}{t} & i = j, \end{cases}$$

we get

$$p_{ij}(t) \approx \begin{cases} tp'_{ij}(0) & i \neq j \\ 1 + tp'_{ij}(0) & i = j, \end{cases}$$

for small t. (provided these derivatives exist)

It therefore seems plausible that the matrix of derivatives $(p'_{ij}(t))$ at time $t = 0$ for a Markov process would play a similar role as the transition matrix $\mathbf{P}$ for a Markov chain in characterizing the process.

Remark: It can be proved that the derivatives $(p'_{ij}(t))$ exist for Markov processes with right-continuous trajectories satisfying the property that only a finite number of jumps can be made in each finite time-interval.

Definition: Let (X_t) be a (discrete, time-homogeneous) Markov process. Suppose there exists $q_{ij} \geq 0$ for $j \neq i$ and $q_{ii} \leq 0$ such that

$$P(X_t = j \mid X_0 = i) = q_{ij}t + o(t),$$

and

$$P(X_t \neq i \mid X_0 = i) = q_i t + o(t),$$

where $q_i = \sum_{j \neq i} q_{ij} =: -q_{ii}$, and $o(t)$, small ordo of t, is a function satisfying $\lim_{t \to 0} o(t)/t = 0$. The numbers q_{ij} are called the **transition intensities** from state i to state j and the array

$$\mathbf{Q} = \begin{pmatrix} \ddots & \vdots & . \\ \dots & q_{ij} & \dots \\ . & \vdots & \ddots \end{pmatrix}_{i,j \in S}$$

is called the **intensity matrix, or the infinitesimal generator** of the Markov process.

Remark: Note that

$$\sum_j q_{ij} = \sum_{j \neq i} q_{ij} + q_{ii} = -q_{ii} + q_{ii} = 0,$$

for any state i, so row sums in $\mathbf{Q}$ are zero by definition.

Remark: Note that

$$P(X_t = i \mid X_0 = i) = 1 - q_i t + o(t),$$

for any state i, since $P(X_t = i \mid X_0 = i) + P(X_t \neq i \mid X_0 = i) = 1$.

Remark: Note that if S is finite, then $\mathbf{P} := \frac{1}{\max q_i}\mathbf{Q} + I$ is a stochastic matrix. This makes it possible to translate some properties for discrete time Markov chains to the continuous time case, see e.g. Lawler, exercise 3.4.

Theorem: Let $(X_t)_{t\geq 0}$ be a (discrete, time-homogeneous) Markov process with intensity matrix $\mathbf{Q}$. If T_1 denotes the time for the first jump of the process, then:

(i) $P(T_1 \leq t \mid X_0 = i) = 1 - e^{-q_i t}$ i.e. if the process starts in state i then T_1 is exponentially distributed with expectation $1/q_i$.

(ii) $P(X_{T_1} = j \mid X_0 = i) = \frac{q_{ij}}{q_i}$, for $j \neq i$ i.e. if the process starts in state i then the probability that the next state is j is q_{ij}/q_i.

(iii) T_1 and X_{T_1} are conditionally independent given that $X_0 = i$.

Sketch of proof: By the Markov property and homogeneity of the chain it follows that T_1 has the "lack of memory" property (see the appendix below), and therefore T_1 must be exponentially distributed. The remaining part of the proof follows from properties of exponentially distributed random variables stated in the appendix.

Construction of a Markov process, $(X_t)_{t\geq 0}$, from an intensity matrix $\mathbf{Q}$:

1. Choose a starting point Z_0.

2. If $Z_0 = j$ then the process stays in state j a random time interval $U_0 \sim \text{Exp}(q_j)$. (U_0 is called the first holding time.) (We will typically only consider cases when $0 < q_j < \infty$. If $q_j = 0$ then the process will never leave state j, and if $q_j = \infty$ then the process leaves j immediately.)

3. Next choose a new state, Z_1, according to the matrix $R = (r_{ij})$, with $r_{ij} = q_{ij}/q_i$, $i \neq j$, and $r_{ii} = 0$.

4. Repeat inductively steps 2 and 3. If $Z_j = k$ let $U_j \sim \text{Exp}(q_k)$ (be holding times independent of $U_0, ..., U_{j-1}$)

If $J_n = \sum_{i=0}^{n-1} U_i$ and $J_n \to \infty$ then we can define X_t for any $t \geq 0$ by $X_t = Z_m$ if $J_m \leq t < J_{m+1}$.

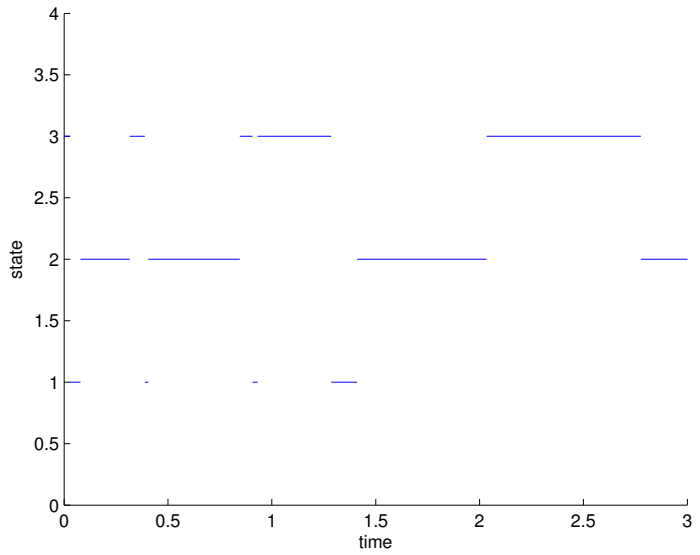


Figure 1: Trajectory of a Markov process with state space $S = (1, 2, 3)$

Theorem: The process (X_t) defined as above is a (homogeneous) Markov process.

Remark: The Markov chain (Z_n) with transition matrix R is called the jump-chain corresponding to (X_t). Note that the definition of R above presumes that $q_i > 0$. To avoid ambiguity we define $r_{ij} = 0$, for $i \neq j$ and $r_{ii} = 1$ in case $q_i = 0$.

Example: Construction of the Poisson process:
$q_j = \lambda$ for all j (for some parameter $\lambda > 0$)
$r_{i,i+1} = 1$ for any $i \geq 0$.

2 Appendix: Some properties of exponentially distributed random variables

The exponential distribution may be viewed as a continuous counterpart to the geometric distribution.

We say that a random variable X is exponentially distributed with parameter $\lambda > 0$ (notation $X \sim \text{Exp}(\lambda)$) if its density function is given by $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$, $f_X(x) = 0$, otherwise.

Distribution function: $F_X(x) = 1 - e^{-\lambda x}$, $x \geq 0$.

If $X \sim \text{Exp}(\lambda)$ then $E(X) = \int_0^\infty P(X > x)dx = \int_0^\infty e^{-\lambda x} = \frac{1}{\lambda}$.

Exponentially distributed random variables have many interesting properties:

Theorem: (Lack of memory property.) Let $X \sim \text{Exp}(\lambda)$. Then

$$P(X > x + y \mid X > y) = P(X > x), \text{ for all } x, y > 0.$$

The exponential distribution is the only non-negative continuous distribution having the lack of memory property.

If $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$ are independent then

$$\begin{aligned} P(X_1 < X_2) &= \int_{\mathbb{R}} P(X_1 < X_2 \mid X_1 = s) f_{X_1}(s) dt \\ &= \int_0^\infty \underbrace{P(X_2 > s)}_{e^{-\lambda_2 s}} \underbrace{f_{X_1}(s)}_{\lambda_1 e^{-\lambda_1 s}} ds \\ &= \lambda_1 \int_0^\infty e^{-(\lambda_1+\lambda_2)s} ds = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \end{aligned} \tag{1}$$

and if $T = \min(X_1, X_2)$ then $T \sim \text{Exp}(\lambda_1 + \lambda_2)$ since

$$\begin{aligned} P(T > t) &= P(X_1 > t, X_2 > t) \\ &\underbrace{=}_{\textit{by indep.}} P(X_1 > t)P(X_2 > t) = e^{-(\lambda_1+\lambda_2)t}. \end{aligned}$$

The events $\{\min(X_1, X_2) > t\}$ and $\{X_1 < X_2\}$ are also independent, since a slight generalization of (1) gives $P(t < X_1 < X_2) = \lambda_1 e^{-(\lambda_1+\lambda_2)t}/(\lambda_1 + \lambda_2)$.

More generally:

Theorem: Let $X_1 \sim \text{Exp}(\lambda_1), X_2 \sim \text{Exp}(\lambda_2), ...$ be independent exponentially distributed random variables with $0 < \sum_i \lambda_i < \infty$. Let $X = \inf_k X_k$.

Then this infimum is attained at a unique random value K of k (with probability 1). Moreover X and K are independent and

$$X \sim \mathrm{Exp}(\sum_i \lambda_i),$$

and

$$P(K = k) = \frac{\lambda_k}{\sum_i \lambda_i}.$$

3 Matlab

3.1 Program for simulating a Markov processes with finite state space and given generator Q (Figure 1)

```
Q=[-12 10 2; 1 -2 1; 2 2 -4];                    % Choose intensity matrix
Tmax=3;                  % The process will be simulated up to time Tmax
z(1) = 3;          % Choose initial state in 1,2,...,m (in case Q has size m × m)
                                % (We store Z_n in z(n+1) and J_n in J(n+1).)


for i=1:length(Q)                                % corresponding jump-chain
   for j=1:length(Q)
     if i==j
       R(i,j)=0;
     else
       R(i,j)=-Q(i,j)/Q(i,i);
     end
   end
end
J(1) = 0;                                                % start time at t=0
i = 1;
while J(i) < Tmax,
   J(i+1) = J(i) + log(rand)/Q(z(i),z(i));                % Next jump-time
   u = rand;                          % Next state is chosen according to R
   j = 1;
   s = R(z(i),1);
   while ((u > s) & (j < length(Q) )),
     j=j+1;
     s=s+R(z(i),j);
   end
   z(i+1) = j;
   i=i+1;
end
cla;
hold on                             % A trajectory plot consists of many...
for i=1:(length(J)-1),
   plot([J(i) J(i+1)], [z(i) z(i)]);                    % ... line-segments
end
axis([0 Tmax 0 (length(Q)+1)]);
xlabel('time');
ylabel('state');
```

4 Suggested exercises

Basic exercises:
18, 20

Exercises Lawler:
3.1, 3.3

Lecture 11 Markov Processes, 1MS012

5 Forward & Backward equations

Recall:
A (well behaved) Markov process gives rise to an intensity matrix or generator $\mathbf{Q}$. Conversely given a matrix $\mathbf{Q} = (q_{jk})$ where q_{jk}, for all $j, k \in S$ satisfy
(a) $q_{jk} \geq 0$, $j \neq k$.
(b) $-\infty < q_{jj} \leq 0$.
(c) $\sum_{k \in S} q_{jk} = 0,\ j \in S$,
we can construct a Markov process (at least if S is finite), with holding times, exponentially distributed with parameters $-q_{jj} = q_j < \infty$, and jump chain also specified by the entries in $\mathbf{Q}$.

The matrix $\mathbf{Q}$ gives all information about the Markov process in the sense that for small time steps h

$$\begin{aligned} p_{ij}(h) &= P(X_{t+h} = j \mid X_t = i) = q_{ij}h + o(h),\ i \neq j \\ p_{jj}(h) &= P(X_{t+h} = j \mid X_t = j) = 1 + q_{jj}h + o(h) = 1 - q_j h + o(h) \end{aligned}$$

Markovprocesses in discrete time:
$P(X_n = k)$ is "easy" to find if we know the distribution of X_0 and the transition matrix $\mathbf{P}$.

Markovprocesses continuous time:
Not obvious how to use $\mathbf{Q}$ to find $P(X_t = k)$. We can derive equations for $p_{ij}(t) = P(X_t = j | X_0 = i)$ that hopefully can be solved:

Forward equations: (S finite)
We study the derivatives of $p_{ij}(t)$.

$$\begin{aligned} \lim_{h\to 0} \frac{p_{ij}(t+h) - p_{ij}(t)}{h} &\underbrace{=}_{CK-equations} \lim_{h\to 0} \frac{\sum_{k\in S} p_{ik}(t)p_{kj}(h) - p_{ij}(t)}{h} \\ &\underbrace{=}_{S\ finite} \sum_{k\in S\setminus\{j\}} p_{ik}(t) \underbrace{\lim_{h\to 0} \frac{p_{kj}(h)}{h}}_{q_{kj}} + p_{ij}(t) \underbrace{\lim_{h\to 0} \frac{p_{jj}(h) - 1}{h}}_{q_{jj}} \\ &= \sum_{k\in S} p_{ik}(t) q_{kj}. \end{aligned}$$

The equations

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) q_{kj}$$

are called **Kolmogorov's forward equations**.

Forward equations expressed in matrix form: $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$

$$\underbrace{\begin{pmatrix} \ddots & \vdots & . \\ \dots & p'_{ij}(t) & \dots \\ . & \vdots & \ddots \end{pmatrix}}_{\mathbf{P}'(t)} = \underbrace{\begin{pmatrix} \ddots & \vdots & . \\ p_{i1}(t) & p_{i2}(t) \dots & p_{i|S|}(t) \\ . & \vdots & \ddots \end{pmatrix}}_{\mathbf{P}(t)} \underbrace{\begin{pmatrix} \ddots & q_{1j} & . \\ \dots & q_{2j} & \dots \\ . & q_{|S|j} & \ddots \end{pmatrix}}_{\mathbf{Q}}$$

(Actually we just proved that the right-derivative exists but a similar arguments shows that also the left derivative exists.)

Example: A machine works an exponential time with mean $\frac{1}{\mu}$. It takes an exponential time with mean $\frac{1}{\lambda}$ to repair it. If the machine works at time 0, what is the probability that it will not work at time t?
We model the state of the machine with a Markov process. Let (X_t) be a Markov process with state space $S = (0, 1)$ where X_t represents the state of the process at time t: Let
$X_t = 1$ mean that the machine works at time t, and
$X_t = 0$ mean that it doesn't work at time t.
When X_t enters state 1 it stays there an $\mathrm{Exp}(\mu)$-distributed time and then goes to 0 and stays there an $\mathrm{Exp}(\lambda)$-distributed time and then goes to state 1 and so on.
We have $q_0 = \lambda = -q_{00}$, and $q_1 = \mu = -q_{11}$. Thus we get the intensity matrix

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$

(since row sums of Q are zero).
We want to find $p_{10}(t) = P(X_t = 0 | X_0 = 1)$.
The forward equations gives:

$$p'_{10}(t) = -\lambda p_{10}(t) + \mu p_{11}(t) = -\lambda p_{10}(t) + \mu(1 - p_{10}(t))$$

$$= \mu - (\lambda + \mu) p_{10}(t),$$

with the boundary condition $p_{10}(0) = 0$.
The equation

$$p'_{10}(t) + (\lambda + \mu) p_{10}(t) = \mu,$$

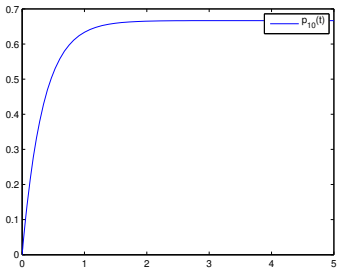


Figure 2: (X_t) with parameters $\lambda = 1$, and $\mu = 2$.
$\lim_{t\to\infty} p_{10}(t) = \mu/(\lambda+\mu) = 2/3$.

can be solved by the standard trick of equivalently writing it as

$$\underbrace{p'_{10}(t)e^{(\lambda+\mu)t} + (\lambda+\mu)e^{(\lambda+\mu)t}p_{10}(t)}_{\frac{d}{dt}(p_{10}(t)e^{(\lambda+\mu)t})} = \mu e^{(\lambda+\mu)t}.$$

By integrating both sides we get

$$p_{10}(t)e^{(\lambda+\mu)t} = \frac{\mu}{\lambda+\mu}e^{(\lambda+\mu)t} + c,$$

where c is a constant. The boundary condition $p_{10}(0) = 0 \Rightarrow c = \frac{-\mu}{\lambda+\mu}$.
Thus

$$p_{10}(t) = \frac{\mu}{\lambda+\mu}(1 - e^{-(\lambda+\mu)t}).$$

5.1 Kolmogorov's backward equations

If S is infinite then the Kolmogorov forward equations may fail to hold.
To avoid such problems we derive another set of equations.
By conditioning on the state of the process at time h we get

$$\begin{aligned}
\lim_{h\to 0}\frac{p_{ij}(t+h)-p_{ij}(t)}{h} &\underbrace{=}_{CK-equations} \lim_{h\to 0}\frac{\sum_{k\in S}p_{ik}(h)p_{kj}(t)-p_{ij}(t)}{h} \\
&\underbrace{=}_{Can\ be\ justified} \sum_{k\in S\setminus\{i\}}\underbrace{\lim_{h\to 0}\frac{p_{ik}(h)}{h}}_{q_{ik}}p_{kj}(t) + p_{ij}(t)\underbrace{\lim_{h\to 0}\frac{p_{ii}(h)-1}{h}}_{q_{ii}} \\
&= \sum_{k\in S} q_{ik}p_{kj}(t).
\end{aligned}$$

We have thus derived **Kolmogorov's backward equations**

$$p'_{jk}(t) = \sum_{r\in S} q_{jr}p_{rk}(t)$$

expressed in matrix form as:

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t).$$

Thus if the forward equation also holds then

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q}.$$

Theorem: If $(X_t)_{t\geq 0}$ is a Markov process with finite state space,then

$$\mathbf{P}(t) = e^{t\mathbf{Q}} := \sum_{k=0}^{\infty} \mathbf{Q}^k t^k/k! = \mathbf{I} + t\mathbf{Q} + \frac{t^2\mathbf{Q}^2}{2!} + \frac{t^3\mathbf{Q}^3}{3!} + ...,$$

is the unique solution of the forward and backward equations, with initial condition $\mathbf{P}(0) = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

This explicitly explains why $\mathbf{Q}$ is called the "generator" of the process.

If S is finite and $\mathbf{Q}$ can be diagonalized, with $\mathcal{P}^{-1}\mathbf{Q}\mathcal{P} = \mathbf{D}$, for some invertible matrix $\mathcal{P}$, where $\mathbf{D}$ is a diagonal matrix with diagonal entries d_i, then

$$\mathbf{P}(t) = \mathcal{P}e^{t\mathbf{D}}\mathcal{P}^{-1},$$

where $e^{t\mathbf{D}}$ is the diagonal matrix with diagonal entries $e^{d_i t}$.

If S infinite then it is possible to prove that if there exists a solution, $\mathbf{P}(t)$ to the forward and backward equations with $\sum_{j\in S} p_{ij}(t) = 1$, for all t and $i \in S$, then this is the unique solution.
In general it is hard to explicitly solve the forward and backward equations if S has many states and $\mathbf{Q}$ has no particular structure. Typically the forward and backward equations have to be solved by numerical methods.

6 Matlab

6.1 Program for Figure 2

```
lambda=1;      % parameter values
mu=2;          %
x0=0;          % starting point
[t,y]=ode45(@hogerled,[0,5],x0,[],lambda,mu);
               % Numerically solves ODEs of the form y'(t)=f(t,y)
               % with given starting values, on a specified interval
               % with specified parameters.
               % The function f(t,y) is specified in the m-file
               % hogerled.m below.
plot(t,y)
legend('p_{10}(t)')


% In the file hogerled.m
function fty=hogerled(t,y,lambda,mu);
fty=mu-(lambda+mu)*y;
```

7 Suggested exercises

Basic exercises:
19

Extra problems:
c2

Exercises Lawler:
3.5

Lecture 12 Markov Processes, 1MS012

8 Class structure

Definition: State k is **accessible** from state j ($j \to k$) if $p_{jk}(t) > 0$ for some $t \geq 0$.
k and j are intercommunicating ($j \leftrightarrow k$) if $j \to k$ and $k \to j$

Theorem: Let (X_t) be a discrete Markov process with intensity matrix $\mathbf{Q}$. Let j and k be two (distinct) states. Then

$$j \to k$$
$$\Leftrightarrow$$

there exists an admissible path from j to k with respect to the matrix $\mathbf{Q}$ i.e. if either $q_{jk} > 0$ or there exists states $j_1, ..., j_n$ such that $q_{jj_1} > 0$, $q_{j_1j_2} > 0$,..., $q_{j_nk} > 0$.

$$\Leftrightarrow$$
$$p_{jk}(t) > 0, \quad \text{for } \textit{all } t > 0,$$

The last equivalence can be seen since holding times of arbitrary small length have positive probability.

Conclusion:
Don't need to bother about aperiodicity for Markov processes in continuous time.

Closed sets and irreducible sets/processes are defined in analogy with the discrete time case.

The definition of recurrent and transient states has to be presented with care since all states are re-visited an infinite amount of time during the first holding time;

Definition: We say that j is recurrent if j is recurrent for the jump-chain,

otherwise j is transient.

A state j is recurrent if, starting at j the process visits j at arbitrary large time-points (with probability 1). A state j is transient otherwise.

Let T_{jj} be the first time the process returns to j after the first jump. Thus if $X_0 = j$ then $T_{jj} = \inf(t > J_1 : X_t = j)$, where J_1 is the time for the first jump. If j is a recurrent state, and $ET_{jj} = \infty$, then j is said to be null-recurrent. Otherwise we call j non-null, (or positive) recurrent.

If $q_j = 0$ then we may define j to be non-null (positive) recurrent.

Theorem: j is recurrent iff $\int_0^\infty p_{jj}(t)dt = \infty$.

Similarly as for Markov chains in discrete time it can also be proved that
• if i and j are two states with $i \leftrightarrow j$ then i is positive recurrent iff j is positive recurrent.
• All states are positive recurrent for a finite irreducible Markov process.

8.1 Stationary distributions and the long run

Suppose $S = (1, 2, ..., n)$ is finite and

$$p_{jk}(t) \to \pi_k \qquad \text{as } t \to \infty, \tag{2}$$

for all j and k, i.e.

$$\mathbf{P}(t) = \begin{pmatrix} p_{11}(t) & \cdot & \cdot & p_{1n}(t) \\ p_{21}(t) & \cdot & \cdot & p_{2n}(t) \\ \cdot & \cdot & \cdot & \cdot \\ p_{n1}(t) & \cdot & \cdot & p_{nn}(t) \end{pmatrix} \to \begin{pmatrix} \pi_1 & \pi_2 & \cdot & \pi_n \\ \pi_1 & \pi_2 & \cdot & \pi_n \\ \pi_1 & \pi_2 & \cdot & \pi_n \\ \pi_1 & \pi_2 & \cdot & \pi_n \end{pmatrix}$$

Then, $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$, has the following properties:

1. If $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_n(t)) := (P(X_t = 1), \ldots, P(X_t = n))$ is the $1 \times n$ vector representing the probability distribution of X_t then

   $$\boldsymbol{\mu}(t) = \boldsymbol{\mu}(0)\mathbf{P}(t) \to \pi,$$

   as $t \to \infty$. This can be seen from (2) since

   $$\mu_k(t) = P(X_t = k) = \sum_{j \in S} P(X_t = k \mid X_0 = j)P(X_0 = j) = \sum_{j \in S} p_{jk}(t)\mu_j(0)$$

   $$\to \sum_{j \in S} \pi_k \mu_j(0) = \pi_k, \qquad \text{as } t \to \infty.$$

2.
$$\pi = \pi \mathbf{P}(s),$$
for all s, i.e. π is a stationary (or equilibrium) distribution.

This can be seen by letting $t \to \infty$ in the Chapman-Kolmogorov identity since
$$\underbrace{\boldsymbol{\mu}(0)\mathbf{P}(t+s)}_{\to \pi} = \underbrace{\boldsymbol{\mu}(0)\mathbf{P}(t)}_{\to \pi}\mathbf{P}(s)$$
(Thus if $X_0 \overset{d}{\sim} \pi$ then (X_t) forms a stationary process.)

3.
$$\pi \mathbf{Q} = \mathbf{0}$$
This can be seen since $\lim_{t\to\infty} p_{jk}(t) = \pi_k \Rightarrow \lim_{t\to\infty} p'_{jk}(t) = 0$ and since by the forward equations, $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$,
$$\underbrace{\pi \mathbf{P}'(t)}_{\to \mathbf{0}} = \underbrace{\pi \mathbf{P}(t)}_{\pi}\mathbf{Q}.$$

4. The distribution
$$\nu = \frac{1}{\sum \pi_k q_k}(\pi_1 q_1, ..., \pi_n q_n),$$
is a stationary distribution for the jump chain, i.e. $\nu \mathbf{R} = \nu$, where
$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdot & \cdot & r_{1n} \\ r_{21} & \cdot & \cdot & r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ r_{n1} & \cdot & \cdot & r_{nn} \end{pmatrix}$$
is defined by $r_{ij} = q_{ij}/q_i$ if $i \neq j$, and $r_{ii} = 0$ for any i. (Recall the implicit assumption that $q_i > 0$ for all i, so r_{ij} is well defined.)

Check:
Let $\nu_j = \frac{\pi_j q_j}{\sum \pi_k q_k}$. The distribution $\nu = (\nu_1, ..., \nu_n)$ is stationary, since
$$\nu_j = \frac{\pi_j q_j}{\sum \pi_k q_k} = \frac{-\pi_j q_{jj}}{\sum \pi_k q_k} = \frac{\sum_{i\neq j} \pi_i q_{ij}}{\sum \pi_k q_k} = \sum_{i\neq j} \frac{\pi_i q_i}{(\sum \pi_k q_k)} \frac{q_{ij}}{q_i} = \sum_i \nu_i r_{ij},$$
for each j, where in the third equality we used that $\pi \mathbf{Q} = \mathbf{0}$.

5. The stationary distribution π satisfies
$$\pi_k = \frac{1}{q_k E(T_{kk})},$$

i.e. π_k is the proportion of time spent in state k.
The proof of this uses the renewal theorem.
(Interpretation: On average it takes a time of length $E(T_{kk})$ between arrivals to the state k. During this time we spend on average a time of length $1/q_k$ in the state k and the rest of the time in other states.)

Theorem: If (X_t) is an irreducible (discrete, time-homogeneous) Markov process such that all states are positive recurrent then

$$p_{jk}(t) \to \pi_k = \frac{1}{q_k E(T_{kk})},$$

as $t \to \infty$ for all j and k, and $\pi = (\pi_j)_{j \in S}$, is uniquely stationary for the process, and the unique solution to the equation $\pi \mathbf{Q} = \mathbf{0}$, with $\sum_{i \in S} \pi_i = 1$, and $\pi_i \geq 0$, for any i.

How can we check if an irreducible Markov process is positive recurrent?

Theorem: Let (X_t) be an irreducible Markov process with intensity matrix $\mathbf{Q}$ such that there is a unique solution to the forward and backward equations. Then

$$(X_t) \text{ is positive recurrent}$$
$$\Leftrightarrow$$
there exists a solution to $\pi \mathbf{Q} = \mathbf{0}$, with $\sum_{i \in S} \pi_i = 1$, and $\pi_i \geq 0$, for any i.

Remark: If S is finite then $P(t) = e^{\mathbf{Q}t}$ is the unique solution to the forward and backward equations.

Example: Let $(X_t)_{t \geq 0}$ be a Markov process on $S = (0, 1, 2)$ with intensity matrix

$$\mathbf{Q} = \begin{pmatrix} q_{00} & q_{01} & q_{02} \\ q_{10} & q_{11} & q_{12} \\ q_{20} & q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 2 & -3 & 1 \\ 0 & 2 & -2 \end{pmatrix}.$$

Suppose we wish to find the limit

$$\lim_{t \to \infty} P(X_t = 2 \mid X_0 = 0).$$

Since the Markovprocess (X_t) is irreducible with a finite state space it follows from the convergence theorem that

$$\lim_{t \to \infty} P(X_t = 2 \mid X_0 = 0) = \pi_2,$$

where $\pi = (\pi_0, \pi_1, \pi_2)$ is the unique probability vector solving the equation

$$\pi \mathbf{Q} = \mathbf{0}.$$

This system has solution $\pi = (4/7, 2/7, 1/7)$.
Thus

$$\lim_{t\to\infty} P(X_t = 2 \mid X_0 = 0) = 1/7.$$

Example: (continued)
Consider the same Markov process as in the previous example.
Suppose, starting at state 0, we wish to find the expected time it takes to reach state 2.
Let $T_{i2} = \inf(t > J_1 : X_t = 2 \mid X_0 = i)$ be the first passage time from state i to state 2, $i = 0, 1, 2$.
In order to find $E(T_{02})$ we will use the idea of "first step analysis" but since time is continuous the first "interesting time-step" is naturally the time when the first jump occurs. More precisely

$$E(T_{02}) = \sum_{k=0}^{2} E(T_{02} \mid X_{J_1} = k, X_0 = 0) \underbrace{P(X_{J_1} = k \mid X_0 = 0)}_{r_{0k}},$$

where $r_{ik} = -q_{ik}/q_{ii}$, $i \neq j$ are the transition probabilities of the jump chain, i.e.

$$\mathbf{R} = \begin{pmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 1 & 0 \end{pmatrix}.$$

Since T_{02} is the time it takes to leave state 0 plus the time it takes to reach state 2 from the new state reached after the first jump from 0, it follows from the Markov property that

$$E(T_{02} \mid X_{J_1} = k, X_0 = 0) = E(J_1 \mid X_0 = 0) + E(T_{k2}),\ k = 0, 1,$$

$E(T_{02}|X_{J_1} = 2, X_0 = 0) = E(J_1|X_0 = 0)$.
From the intensity matrix we see that $J_1|(X_0 = 0) \sim \text{Exp}(q_0)$ and $J_1|(X_0 = 1) \sim \text{Exp}(q_1)$, where $q_0 = -q_{00} = 1$, and $q_1 = -q_{11} = 3$. Thus $E(J_1|X_0 = 0) = 1$ and $E(J_1|X_0 = 1) = 1/3$.
Therefore

$$E(T_{02}) = \underbrace{E(J_1 \mid X_0 = 0)}_{=1} + E(T_{12}) r_{01} = 1 + E(T_{12}).$$

Similarly

$$E(T_{12}) = (1/q_1) + r_{10} E(T_{02}) = 1/3 + (2/3) \cdot E(T_{02}).$$

Thus $E(T_{02}) = 1 + E(T_{12}) = 1 + 1/3 + (2/3)E(T_{02})$, i.e. $E(T_{02}) = 4$.

9 Matlab

9.1 Code for solving $\pi\mathbf{Q} = \mathbf{0} \Leftrightarrow \mathbf{Q}'\pi' = \mathbf{0}'$ for probability vectors π.

```
Q=[-1 1 0; 2 -3 1; 0 2 -2]

pi= null(transpose(Q))/ sum(null(transpose(Q)))
```

10 Suggested exercises

Basic exercises:
21–22, 27

Extra problems:
b3, B3, B4

Exercises Lawler:
3.8

Lecture 13 Markov Processes, 1MS012

11 Birth-processes

Definition: A birth-process on $S = (0, 1, 2, \ldots)$ is a Markov process with generator

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdot & \cdot \\ 0 & -\lambda_1 & \lambda_1 & 0 & \cdot & \cdot \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

The λ_j:s are called birthrates.
A birth-process is thus a Markovprocess with (trivial) jump chain $r_{j,j+1} = 1$ and holding times λ_j for all states j.

Birth-processes can be used as models for the size of a growing population, where new individuals are born with intensity λ_i if the size of the population is i. The Poisson process corresponds to the special case when $\lambda_i = \lambda$ i.e. the case when the birthrates do not depend on the population size.

11.1 Transition probabilities for a birth-process

Forward equations $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$:

$$p'_{ij}(t) = \lambda_{j-1}p_{i,j-1}(t) - \lambda_j p_{i,j}(t), \quad j \geq 1.$$

Since $p_{ij}(t) = 0$, if $j < i$, we get
(I):

$$p'_{ii}(t) = -\lambda_i p_{ii}(t)$$

(II):

$$p'_{ij}(t) = \lambda_{j-1}p_{i,j-1}(t) - \lambda_j p_{ij}(t), \; j > i$$

Since $p_{ii}(0) = 1$, (I) has solution $p_{ii}(t) = e^{-\lambda_i t}$, for all $i \geq 0$.
We now solve (II):

$$\underbrace{e^{\lambda_j t}(p'_{ij}(t) + \lambda_j p_{ij}(t))}_{\frac{d}{dt}(e^{\lambda_j t}p_{ij}(t))} = (\lambda_{j-1}p_{i,j-1}(t))e^{\lambda_j t}$$

By integrating both sides and using $p_{ij}(0) = 0$, for $j > i$, we finally get the recursive equations

$$p_{ij}(t) = e^{-\lambda_j t}\lambda_{j-1}\int_0^t p_{i,j-1}(s)e^{\lambda_j s}ds \tag{3}$$

for $j > i$.

Remark: It is an exercise to explicitly solve the recursive equations (3) in the special case $\lambda_i = \lambda$ (corresponding to the Poisson process), to obtain the solution

$$p_{ij}(t) = e^{-\lambda t}(\lambda t)^{j-i}/(j-i)!, \quad j \geq i,$$

in that case.

11.2 Honest and dishonest processes

Let J_n be the time for the n:th jump. Let $J_\infty = \lim_{n\to\infty} J_n$.

If $J_\infty < \infty$ then the process makes an infinite number of jumps just before time J_∞. The process is then explosive and $\mathbf{Q}$ does not define the process for all times in this case. This occurs if λ_i grows sufficiently fast.

Since $U_i := (J_{i+1} - J_i) \sim \text{Exp}(\lambda_i)$ and thus

$$EJ_\infty = E(\sum_{i=0}^{\infty} U_i) = \sum_{i=0}^{\infty} E(U_i) = \sum_{i=0}^{\infty} 1/\lambda_i,$$

we have:

Theorem: (Feller-Lundberg)

$$\sum_{i=0}^{\infty} 1/\lambda_i < \infty \Rightarrow P(J_\infty < \infty) = 1 \qquad \text{(explosion always)}$$

It is also possible to prove

$$\sum_{i=0}^{\infty} 1/\lambda_i = \infty \Rightarrow P(J_\infty = \infty) = 1 \qquad \text{(no explosion)}$$

Example: Poisson process $\lambda_i = \lambda$, does not explode since $\sum_{i=0}^{\infty} 1/\lambda = \infty$.

Definition: A Markov process $(X_t)_{t\geq 0}$ is said to be

honest if $P(J_\infty = \infty) = 1$
dishonest if $P(J_\infty = \infty) < 1$

We thus have a simple criteria for telling if a birth-process is honest or not. For more general Markov processes it is much harder:

Theorem: The Markov process $(X_t)_{t\geq 0}$ with generator $\mathbf{Q}$ is honest if one of the following conditions hold.

- $\sup_j q_j < \infty$
- The state space S is finite
- The process starts in a recurrent state.

Note that these are only sufficient conditions for a Markov process to be honest.

Theorem: If $(X_t)_{t\geq 0}$ is irreducible with generator $\mathbf{Q}$ then:

$$\text{all states are positive recurrent} \quad \Leftrightarrow \quad \begin{array}{l}(X_t)_{t\geq 0} \text{ is honest and } \pi\mathbf{Q} = \mathbf{0} \\ \text{for some probability distribution } \pi.\end{array}$$

Obs:
Birth-processes are not irreducible, and stationary distributions do not exist for such processes.

12 Birth-death processes

Definition: A birth death process is a Markov process on $S = (0, 1, 2, ...)$ with generator

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdot & \cdot \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdot & \cdot \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Thus $i \mapsto i+1$ (birth) and $i \mapsto i-1$ (death) are the only possible transitions.

$\lambda_0, \lambda_1, \ldots$ birthrates/arrival rates
$\mu_1, \mu_2, \ldots$ deathrates/departure rates

(the latter notation if regarded as the number of people in a system)

The process will be irreducible if $\lambda_n > 0$, and $\mu_n > 0$ for all n.

In order to find stationary distributions we solve $\pi \mathbf{Q} = \mathbf{0}$.

$$-\lambda_0\pi_0 + \mu_1\pi_1 = 0 \quad \text{(I)}$$
$$\lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} = 0, \qquad n \geq 1 \quad \text{(II)}$$

Thus

$$\mu_{n+1}\pi_{n+1} - \lambda_n\pi_n \underbrace{=}_{\text{(II)}} \mu_n\pi_n - \lambda_{n-1}\pi_{n-1} \underbrace{=}_{\text{(II)}} \ldots = \mu_1\pi_1 - \lambda_0\pi_0 \underbrace{=}_{\text{(I)}} 0, \; n \geq 0,$$

i.e.

$$\mu_{n+1}\pi_{n+1} = \lambda_n\pi_n, \qquad n \geq 0, \qquad \text{(detailed balance equations)}$$

so

$$\pi_n = \frac{\lambda_{n-1}}{\mu_n}\pi_{n-1} = \ldots = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}\pi_0, \qquad n \geq 1.$$

If π is a probability distribution then

$$\sum_{n=0}^{\infty}\pi_n = 1 \Leftrightarrow (1 + \sum_{n=1}^{\infty}\frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1})\pi_0 = 1$$

This partly proves

Theorem:
Let $(X_t)_{t\geq 0}$ be a birth-death process with birth intensities $\lambda_i > 0$, $i \geq 0$, and death intensities $\mu_i > 0$, $i \geq 1$.
If

$$\sum_{k=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k} < \infty$$

and

$$\sum_{k=1}^{\infty}\frac{\mu_1\mu_2\cdots\mu_k}{\lambda_1\lambda_2\cdots\lambda_k} = \infty \quad (4)$$

then a unique stationary distribution π, exists and

$$p_{ij}(t) = P(X_t = j | X_0 = i) \to \pi_j,$$

as $t \to \infty$, where

$$\pi_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \pi_0, \qquad \text{and} \qquad \pi_0 = (1 + \sum_{k=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k})^{-1}.$$

Remark: It can be proved that (4) is a condition for recurrence of the state 0.

Remark: If $\lambda_{n_0} = 0$, for some n_0, then we regard $S = (0, \ldots, n_0)$ as the state space and if the process is irreducible (regarded as a process on S), then

$$p_{ij}(t) = P(X_t = j | X_0 = i) \to \pi_j,$$

as $t \to \infty$, where

$$\pi_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \pi_0, \qquad \text{and} \qquad \pi_0 = (1 + \sum_{k=1}^{n_0} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k})^{-1}.$$

12.1 Population models

Let $X_t =$ number of individuals in a population at time t.
Suppose $X_t = i$, at some fixed time t. Suppose $B \sim \text{Exp}(\lambda_i)$ is the time until next birth, and $D \sim \text{Exp}(\mu_i)$ is the time until the next death, and B and D are independent. If $D < B$ then the next jump will be $i \mapsto i - 1$. By properties of the exponential distribution if follows that $P(D < B) = \frac{\mu_i}{\lambda_i + \mu_i}$.
If J_1 is the time for the next jump, then $J_1 = \min(B, D) \sim \text{Exp}(\lambda_i + \mu_i)$.
It follows from the memoryless property of the exponential distribution that conditional on $X_{J_1} = j$, the process starts afresh from j at time J_1.
Thus $(X_t)_{t \geq 0}$ is a birth-death process with birthrates λ_i, and deathrates μ_i.

Example: Each individual produces new individuals with rate λ and dies with rate μ.
This is a birth-death process with $\lambda_n = n\lambda$, $\mu_n = n\mu$.

12.2 Some common queueing models

$X_t =$ number of people on line for service.
Independent interarrival times $\sim \text{Exp}(\lambda)$, $\lambda > 0$, and service times $\sim \text{Exp}(\mu)$, $\mu > 0$.

Example: $M/M/1$ queue
One server.
This is a birth-death process with $\lambda_n = \lambda$, $\mu_n = \mu$.

(The two M:s refer to the fact that both inter-arrival and service times are exponential and thus memoryless, and hence arrival and service times are Markovian.)
If $\lambda < \mu$ then $\pi_k = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^k$, is the unique stationary distribution.
(Shifted geometric distribution with parameter $1 - \frac{\lambda}{\mu}$.)

Example: $M/M/k$ queue
k servers.
This is a birth-death process with $\lambda_n = \lambda$, $\mu_n = \begin{cases} n\mu & \text{if } n \leq k \\ k\mu & \text{if } n > k \end{cases}$.
If $\lambda < k\mu$ then there is a unique stationary distribution.

Example: $M/M/\infty$ queue
infinitely many servers.
This is a birth-death process with $\lambda_n = \lambda$, $\mu_n = n\mu$.
The distribution $\pi_k = e^{-\lambda/\mu}\frac{(\lambda/\mu)^k}{k!}$, is the unique stationary distribution.
(Poisson distribution with parameter λ/μ.)

13 Suggested exercises

Basic exercises:
23–26, 28

Extra problems:
c1, B2

Exercises Lawler:
3.12

UPPSALA UNIVERSITY
Department of Mathematics
© Docent ÖS

Markov Processes, 1MS012
Spring semester 2024

Lecture 14 Markov Processes, 1MS012

1 Some applications

1.1 Card shuffling

$X_n =$ the order of a deck of cards after shuffling n times. State space S the set of permutations of $1, 2, ..., N$ where N is the number of cards.
$|S| = N!$

Example: Shuffling dynamics: take top card and put it at a random position. (X_n) is clearly irreducible and aperiodic since if $x = (x_1, ..., x_N)$ and $y = (y_1, ..., y_N)$ are 2 permutations, then it is possible to go from x to y in N shuffles, and it is also possible to go from a state to itself if the upper most card is chosen.

Observe that all cards under the original bottom card are equally likely to be in any order. The deck is thus in a completely random order when the original bottom card is first reinserted.
Let T be the time when this happens.
$T = T_1 + ... + T_N$ where T_i is the time it takes for the original bottom card to move from position i from below one step upwards for $i = 1, ..., N-1$, and $T_N = 1$. Since T_i is geometrically distributed with parameter i/N it follows that $E(T_i) = N/i$. Thus

$$E(T) = \sum_{i=1}^{N} E(T_i) = N \sum_{i=1}^{N} \frac{1}{i} \sim N \log N.$$

1.2 Google PageRank: Decide the order search results are displayed.

In 1998 Larry Page and Sergey Brin at Stanford university founded Google Inc., the company behind the Google search engine.

Let W be the set of webpages that can be reached by Google. W is stored in a database of size $n = |W|$.

Each of the n pages are regularly ranked with respect to popularity measured by the degree of links from other pages.

To each query submitted to the search engine, Google displays those pages that match the query in order of page ranking.

The basic idea behind the Google search engine is roughly the following: Consider an imaginary internet surfer randomly clicking on links. We rank webpages according to the probability distribution on W describing the location of the surfer after a huge number of clicks.

Basic Markov model:

The set W can be regarded as the set of nodes in a directed graph where we put a directed edge between two nodes (web-pages) v_1 and v_2 if there is a hyperlink from v_1 to v_2.

Consider a Markov chain, (X_n), with state space W obtained by the following dynamics; If $X_n = v \in W$, then, with probability $p = 0.85$ we pick one of the outgoing links of v at random with equal probabilities, and with probability $1-p$ we move to a randomly chosen page (among the n pages). (If there are no outgoing links of v, i.e. if v is a **dangling node** then one remedy to make this a well defined Markov chain is to then always choose a random page.)

This Markov chain has finite state space and is irreducible and aperiodic and thus there exists a unique stationary distribution.

Pages are ranked according to the stationary probabilities for this Markov chain.

Other choices of parameter values p are of course also possible. The natural

choice of $p = 1$ (corresponding to the rough idea described above) leads to many problems, since the long run behavior of the Markov chain will then depend on the initial state. A small value of p gives quick convergence, but makes it hard to distinguish differences in rankings.

Some major advantages/disadvantages of the method:
Advantage: Rather hard to manipulate ranking of a single website.
(A way to manipulate it is e.g. to build *link farms* i.e. a big set of pages where every page have links to every other page in the set.)
Disadvantage: Favors older webpages.

1.3 Calculating integrals by using a coin

Suppose we want to calculate

$$\int_a^b f(x)dx,$$

for some continuous function f where a and b are real numbers.

Our tool: A coin.

Assume w.l.o.g. that $a = 0$ and $b = 1$. Consider random iterations with $w_0(x) = x/2$, and $w_1(x) = x/2 + 1/2$ chosen with probability $1/2$ each, i .e. let

$$X_{n+1} = w_{I_{n+1}}(X_n),$$

where (I_n) is i.i.d. with $P(I_n = 0) = P(I_n = 1) = 1/2$. Let $X_0 = x_0$ be an arbitrary point in $[0, 1]$.

(X_n) is a Markov chain.

We claim that the average of f along a trajectory, $\frac{\sum_{k=0}^{N-1} f(X_k)}{N}$, converges to $\int_0^1 f(x)dx$, as $N \to \infty$, (with probability one).

In order to motivate why the claim holds, divide the interval according to the base 2-expansion; Let

$$\Delta_{i_1,\dots,i_k} = (i_1 2^{-1} + i_2 2^{-2} + \dots + i_k 2^{-k}, i_1 2^{-1} + i_2 2^{-2} + \dots + (i_k + 1)2^{-k}).$$

If for any fixed k, $x \in \Delta_{i_1,\ldots,i_k}$, then $w_0(x) \in \Delta_{0i_1,\ldots,i_{k-1}}$, and $w_1(x) \in \Delta_{1i_1,\ldots,i_{k-1}}$.

Thus if Y_n is the interval where X_n belongs then (Y_n) is an irreducible aperiodic Markov chain with state space $S = \{\Delta_{i_1,\ldots,i_k}; i_j \in \{0,1\}, 1 \leq j \leq k\}$. By symmetry Y_n will be uniformly distributed on S as $n \to \infty$.

We have

$$\frac{\sum_{n=0}^{N-1} f_k(Y_n)}{N} \to \sum f_k(\Delta_{i_1,\ldots,i_k})|\Delta_{i_1,\ldots,i_k}| = \int_0^1 f_k(x)dx$$

as $N \to \infty$ with probability one, if f_k is a constant function on each interval $\Delta_{i_1,\ldots,i_k}$.

Thus also

$$\frac{\sum_{n=0}^{N-1} f_k(X_n)}{N} \to \sum f_k(\Delta_{i_1,\ldots,i_k})|\Delta_{i_1,\ldots,i_k}| = \int_0^1 f_k(x)dx.$$

By letting $k \to \infty$ (if $f_k \to f$), we get

$$\frac{\sum_{n=0}^{N-1} f(X_n)}{N} \to \int_0^1 f(x)dx,$$

as $N \to \infty$ with probability one.

1.4 Yahtzee

Yahtzee is a popular dice game. The object of the game is to score the most points by rolling five dice to make certain combinations. On each turn, a player gets up to three rolls of the dice. He or she can save any dice that are wanted to complete a combination and then re-roll the other dice. Suppose a player wants as many ”sixes” as possible. We can model the number of ”sixes” after n rolls with a Markov chain.

Let X_n denote the number of sixes the player has after n rolls. We are interested in finding the distribution of X_3 given that $X_0 = 0$. If we have j sixes after n rolls then the number of new sixes after $n+1$ rolls will be binomially distributed with parameters $5 - j$ and $1/6$. Thus (X_n) is a Markov chain with state space $S = (0, 1, 2, 3, 4, 5)$ with $(X_{n+1} - j|X_n = j) \sim \text{Bin}(5-j, 1/6)$, i.e.

$$P(X_{n+1} = i + j|X_n = j) = \binom{5-j}{i} (\frac{1}{6})^i (\frac{5}{6})^{5-i-j}, \quad 0 \leq i \leq 5 - j,$$

so

$$p_{jk} = P(X_{n+1} = k | X_n = j) = \binom{5-j}{k-j} (\frac{1}{6})^{k-j} (\frac{5}{6})^{5-k}, \quad j \leq k \leq 5,$$

i.e. (X_n) has transition matrix

$$\mathbf{P} = \begin{pmatrix} \binom{5}{0}(\frac{1}{6})^0(\frac{5}{6})^5 & \binom{5}{1}(\frac{1}{6})^1(\frac{5}{6})^4 & \binom{5}{2}(\frac{1}{6})^2(\frac{5}{6})^3 & \binom{5}{3}(\frac{1}{6})^3(\frac{5}{6})^2 & \binom{5}{4}(\frac{1}{6})^4(\frac{5}{6})^1 & \binom{5}{5}(\frac{1}{6})^5(\frac{5}{6})^0) \\ 0 & \binom{4}{0}(\frac{1}{6})^0(\frac{5}{6})^4 & \binom{4}{1}(\frac{1}{6})^1(\frac{5}{6})^3 & \binom{4}{2}(\frac{1}{6})^2(\frac{5}{6})^2 & \binom{4}{3}(\frac{1}{6})^3(\frac{5}{6})^1 & \binom{4}{4}(\frac{1}{6})^4(\frac{5}{6}^0) \\ 0 & 0 & \binom{3}{0}(\frac{1}{6})^0(\frac{5}{6})^3 & \binom{3}{1}(\frac{1}{6})^1(\frac{5}{6})^2 & \binom{3}{2}(\frac{1}{6})^2(\frac{5}{6})^1 & \binom{3}{3}(\frac{1}{6})^3(\frac{5}{6})^0 \\ 0 & 0 & 0 & \binom{2}{0}(\frac{1}{6})^0(\frac{5}{6})^2 & \binom{2}{1}(\frac{1}{6})^1(\frac{5}{6})^1 & \binom{2}{2}(\frac{1}{6})^2(\frac{5}{6})^0 \\ 0 & 0 & 0 & 0 & \binom{1}{0}(\frac{1}{6})^0(\frac{5}{6})^1 & \binom{1}{1}(\frac{1}{6})^1(\frac{5}{6})^0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\approx \begin{pmatrix} 0.4019 & 0.4019 & 0.1608 & 0.0322 & 0.0032 & 0.0001 \\ 0 & 0.4823 & 0.3858 & 0.1157 & 0.0154 & 0.0008 \\ 0 & 0 & 0.5787 & 0.3472 & 0.0694 & 0.0046 \\ 0 & 0 & 0 & 0.6944 & 0.2778 & 0.0278 \\ 0 & 0 & 0 & 0 & 0.8333 & 0.1667 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus

$$\mathbf{P}^3 \approx \begin{pmatrix} 0.0649 & 0.2363 & 0.3440 & 0.2504 & 0.0912 & 0.0133 \\ 0 & 0.1122 & 0.3266 & 0.3566 & 0.1731 & 0.0315 \\ 0 & 0 & 0.1938 & 0.4233 & 0.3081 & 0.0748 \\ 0 & 0 & 0 & 0.3349 & 0.4876 & 0.1775 \\ 0 & 0 & 0 & 0 & 0.5787 & 0.4213 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

so in particular

$$P(X_3 = k | X_0 = 0) \approx \begin{cases} 0.0649, & k = 0 \\ 0.2363, & k = 1 \\ 0.3440, & k = 2 \\ 0.2504, & k = 3 \\ 0.0912 & k = 4 \\ 0.0133 & k = 5 \end{cases}.$$

(This can also be seen without involving Markov chains in the following way: The probability of not getting a six on a given dice in 3 turns is $(5/6)^3$ independently of the other dice. Thus $(X_3|X_0 = 0) \sim \text{Bin}(5, 1 - (5/6)^3) = \text{Bin}(5, \frac{91}{216})$.)

UPPSALA UNIVERSITY
Department of Mathematics

Markov Processes, 1MS012
Spring semester 2024

Lecture 15 Markov Processes, 1MS012

1 Brownian motion

1.1 History

Robert Brown (1827), english botanist: describe the motion of pollen particles moving in a gas or in a container of water.

Albert Einstein (1905): mathematical description using laws of physics as bombardment of molecules in the surrounding medium.
(Independent of Brown's work.)

Norbert Wiener (1923): Formulated Brown's observations with mathematical rigor.

Remark:
Louis Bachelier (1900) proposed a "Brownian motion model" for the movement of prices in the French bond market. Although ignored by academics for many decades it is now regarded as the first step in the mathematical theory of stock markets.

1.2 Construction of Brownian motion from a simple random walk

We can model the phenomena observed by Brown with simple random walk in 3 dimensions where the motions in each coordinate are independent symmetric simple random walks.

At each Δt time unit take a step of size Δx either to the left or to the right.

If X_t is the position at time t then

$$X_{n\Delta t} = S_n \Delta x,$$

where $S_n = \sum_{j=1}^{n} I_j$ is a simple random walk with (I_n) i.i.d. with $P(I_n = 1) = P(I_n = -1) = 0.5$.

Since

$$E(X_{n\Delta t}) = E(S_n \Delta x) = \Delta x E(\sum_{j=1}^{n} I_j) = \Delta x \sum_{j=1}^{n} \underbrace{E(I_j)}_{=0} = 0$$

and

$$Var(X_{n\Delta t}) = Var(S_n \Delta x) = (\Delta x)^2 Var(S_n) = (\Delta x)^2 Var(\sum_{j=1}^{n} I_j)$$

$$\underbrace{=}_{by\ independence} (\Delta x)^2 \sum_{j=1}^{n} \underbrace{Var(I_j)}_{=1} = n(\Delta x)^2,$$

and

$$X_t = (\Delta x) S_{\lfloor \frac{t}{\Delta t} \rfloor},$$

where $\lfloor t/\Delta t \rfloor$ is the largest integer less than or equal to $t/\Delta t$, we have $EX_t = 0$, $Var(X_t) = (\Delta x)^2 \lfloor \frac{t}{\Delta t} \rfloor$.

X_t is approximately normally distributed with these parameters, if Δt is small, since a sum of many i.i.d. random variables is approximately normally distributed by the central limit theorem.

Let Δx and Δt tend to 0 in a way resulting in a non-trivial limiting process:

$$(\Delta x)^2 = \sigma^2 \Delta t \Rightarrow Var(X_t) \to \sigma^2 t,$$

for any constant $\sigma^2 > 0$.

Then the limiting process is a called a Brownian motion.

Definition: A random process $(X_t)_{t \geq 0}$ is called a Brownian motion or Wiener process with parameter $0 < \sigma^2 < \infty$ if

(a) $X_0 = 0$

(b) (X_t) has
Independent increments: The increments $X_{t_1}, X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, ..., X_{t_n} - X_{t_{n-1}}$ are independent random variables for any choice of time-points $0 < t_1 < t_2 < ... < t_n$.

Stationary increments: The distribution of $X_{t+h} - X_t$ does not depend on t for any $h > 0$

(c) $X_{t+h} - X_t \sim N(0, \underbrace{\sigma^2 h}_{\text{(Variance)}})$, for all $h > 0$.

(d) (X_t) has continuous trajectories

Definition: A Brownian motion with $\sigma^2 = 1$ is called a **standard Brownian motion**.

The above arguments in deriving the Brownian motion as a limiting process uses a special case of the invariance principle/functional central limit theorem:
Suppose $\Delta t = 1/n$, and $\Delta x = 1/\sqrt{n}$ above and let

$$B_t^{(n)} = (\Delta x) S_{\lfloor \frac{t}{\Delta t} \rfloor} = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}.$$

The functional CLT states that $(B_t^{(n)})_{t \geq 0}$ converges (in distribution) to a standard Brownian motion, $(B_t)_{t \geq 0}$, as $n \to \infty$. (The central limit theorem is a special case of the the functional CLT obtained by letting $t = 1$.)

Example: Suppose (S_k) is a simple symmetric random walk with $(S_0 = 0)$. We have earlier calculated (Gamblers ruin, Lecture 4)

$$P(S_k \text{ reaches } -i < 0 \text{ before } j > 0) = \frac{j}{i+j},$$

if i and j are positive integers. Thus for large n

$$P(B_t^{(n)} \text{ reaches } -a < 0 \text{ before } b > 0) \approx \frac{b\sqrt{n}}{a\sqrt{n} + b\sqrt{n}} = \frac{b}{a+b},$$

for any real $a, b > 0$.

By the invariance principle it follows that

$$P(B_t \text{ reaches } -a < 0 \text{ before } b > 0) = \frac{b}{a+b}.$$

2 Further properties of Brownian motion

1. Recurrence/transience like a simple random walk:
Recall: Symmetric simple random walk in 1 or 2 dimensions is recurrent but symmetric simple random walk in 3 dimensions is transient.
Shizuo Kakutani stated this: "A drunk man will find his way home but a drunk bird may get lost forever."
This property is inherited by the Brownian motion where recurrence here means that any line segment (or ball) is re-visited with probability one in one or two dimensions, but balls need not be revisited in 3d.

2. "erratic" trajectories:
It can be proved that trajectories of a Brownian motion are nowhere differentiable.

3. A Brownian motion (X_t) can be standardized:

$$B_t = \frac{X_t}{\sigma}$$

is a standard Brownian motion.

Since a Brownian motion (X_t) has independent increments, and

$$X_{t_1} \sim N(0, \sigma^2 t_1),\ X_{t_{k+1}} - X_{t_k} \sim N(0, \sigma^2(t_{k+1} - t_k)),\ k = 1, ..., n-1,$$

it follows that the finite dimensional density functions of (X_t) has the form

$$f_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_1, ..., x_n) = f_{X_{t_1}}(x_1) f_{X_{t_2} - X_{t_1}}(x_2 - x_1) \cdots f_{X_{t_n} - X_{t_{n-1}}}(x_n - x_{n-1}),$$

where for $Z \sim N(0, \sigma^2)$, we use the notation $f_Z(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ for its density function.

3 Suggested exercises

Basic exercises:
29–32

Extra problems:
c3

Exercises Lawler:
8.4abc, 8.9

Lecture 16 Markov Processes, 1MS012

4 Transformations of Brownian Motion

Many transformations of Brownian motion generates another Brownian motion:

Theorem: Suppose $(B_t)_{t\geq 0}$ is a standard Brownian motion. Then the transformed processes

1. $B_1(t) = cB_{t/c^2}$, for fixed $c > 0$
2. $B_2(t) = tB_{1/t}$ for $t > 0$, $B_2(t) = 0$ for $t = 0$
3. $B_3(t) = B_{t+h} - B_h$, for fixed $h > 0$.

are all standard Brownian motions.

Roughly all these results says that Brownian motion is stochastically self-similar in various ways. Such properties are useful when studying properties of trajectories of Brownian motion.

Proofs: We need to check the defining properties of Brownian motion for each of the transformed processes.

1. $$B_1(t) = cB_{t/c^2}$$

   (a) $B_1(0) = cB_{0/c^2} = cB_0 = 0$.

   (b)–(c) If $t_1 < t_2 < .. < t_n$, then $t_1/c^2 < t_2/c^2 \leq t_3/c^2 < ... < t_n/c^2$, and the corresponding increments B_{t_1/c^2}, and $B_{t_n/c^2} - B_{t_{k-1}/c^2}$, $k = 2, ..., n$ are thus independent. Multiples of each of these increments by c are also independent i.e. $B_1(t_1)$, and $B_1(t_k) - B_1(t_{k-1})$, $k = 2, ..., n$ are independent, so $B_1(t)$ has independent increments.

   The increments

   $$B_1(t + h) - B_1(t) = cB_{(t+h)/c^2} - cB_{t/c^2} = c(B_{(t+h)/c^2} - B_{t/c^2})$$

are clearly normally distributed as a constant multiple of a normally distributed random variable. Since the increment $B_{(t+h)/c^2} - B_{t/c^2}$ has mean zero, then

$$B_1(t+h) - B_1(t) = c(B_{(t+h)/c^2} - B_{t/c^2}))$$

must have mean zero. Since the mean is zero it follows that the variance is

$$\begin{aligned} E((B_1(t+h) - B_1(t))^2) &= E((cB_{(t+h)/c^2} - cB_{t/c^2})^2) \\ &= c^2 E((B_{(t+h)/c^2} - B_{t/c^2})^2) \\ &= c^2((t+h)/c^2 - t/c^2) = h, \end{aligned}$$

so $B_1(t+h) - B_1(t) \sim N(0,h)$. Thus B_1 has both stationary, and normally distributed increments with $\sigma^2 = 1$.

(d) B_1 has continuous trajectories, since compositions of continuous functions are continuous.

2.

$$B_2(t) = tB_{1/t}$$

(a) By definition, $B_2(0) = 0$.

(b)–(c)

$$B_2(t) - B_2(s) = tB_{1/t} - sB_{1/s} = (t-s)B_{1/t} - s(B_{1/s} - B_{1/t}),\ s < t$$

is the difference of independent normally distributed random variables each with mean 0, so the difference is normally distributed with mean 0. We need to check that the normal random variable has the correct variance:

$$\begin{aligned} E((B_2(t+h) - B_2(t))^2) &= E((B_2(t) - B_2(t+h))^2) \\ &= E((tB_{1/t} - (t+h)B_{1/(t+h)})^2) \\ &= E(\Big(tB_{1/t} - tB_{1/(t+h)} + tB_{1/(t+h)} \\ &\quad -(t+h)B_{1/(t+h)} + hB_0\Big)^2) \\ &= t^2 E((B_{1/t} - B_{1/(t+h)})^2) \\ &\quad +h^2 E((B_{1/(t+h)} - B_0)^2) \\ &= t^2(1/t - 1/(t+h)) \\ &\quad +h^2/(t+h) \\ &= h \end{aligned}$$

(In the 4:th equality we used independence of the increments $B_{1/t} - B_{1/(t+h)}$ and $B_{1/(t+h)} - B_0$. Thus $B_2(t)$ has stationary and normally distributed increments with the correct variance.

To prove that B_2 has independent increments is hard and beyond the scope of this course. One can rely on the fact that a Gaussian process with mean 0 and covariance function $\min(s,t)$ is a standard Brownian motion, and thus prove it indirectly.
Note that

$$\begin{aligned} Cov(B_s, B_t) &= E(B_s B_t) - E(B_s)E(B_t) = E(B_s B_t) \\ &= E(B_s(B_t - B_s) + B_s^2) \\ &= E(B_s)E(B_t - B_s) + E(B_s^2) = E(B_s^2) = s, \end{aligned}$$

if $s < t$, so

$$Cov(B_2(s), B_2(t)) = st \min(1/s, 1/t) = \min(s,t),$$

is a direct consequence of the same property for standard Brownian motion.

(d) The argument that $\lim_{t\to 0} B_2(t) = 0$ is equivalent to showing that $\lim_{t\to\infty} B_t/t = 0$. To show this is beyond the scope of this course. The translation property in the third statement of this theorem proves continuity at every value of t.

3.

$$B_3(t) = B_{t+h} - B_h$$

(a)

$$B_3(0) = B_{0+h} - B_h = 0.$$

(b)–(c) The increment

$$B_3(t+h) - B_3(t) = (B_{t+2h} - B_h) - (B_{t+h} - B_h) = B_{t+2h} - B_{t+h}$$

is by definition normally distributed with mean 0 and variance h.

If $t_1 < t_2 < ... < t_n$ then the increments

$$B_3(t_k) - B_3(t_{k-1}) = B_{t_k+h} - B_{t_{k-1}+h}$$

are independent by the property of independent increments of B_t.

(d) B_3 is continuous since it is a difference of continuous functions.

5 The Markov property

Consider a stochastic process (X_t) where X_t are continuous random variables.

Note that $P(X_t = a) = 0$, for all $a \in \mathbb{R}$, so expressions like $P(X_{t+h} = j \mid X_t = i)$ do not make sense.

We want to express the Markov property in a way that is useful in all cases. The following definition can then be used;

Definition: We say that a random process $(X_t)_{t\geq 0}$ is a Markov process if

$$P(X_{t_n} \leq x \mid X_{t_1}, ..., X_{t_{n-1}}) = P(X_{t_n} \leq x \mid X_{t_{n-1}})$$

for all $x \in \mathbb{R}$ and $t_1 < t_2 < ... < t_{n-1} < t_n$.

The transition density

$$f(t, y|s, x) = \frac{\partial}{\partial y} P(X_t \leq y \mid X_s = x),$$

is an alternative way to characterize Markov processes with continuous time and (uncountable) state space $\mathbb{R}$. The reason for writing this function of 4 variables like this is to stress that it is to be interpreted as the conditional density that $X_t = y$ given that $X_s = x$.

The Chapman-Kolmogorov equations states that

$$f(t, y|s, x) = \int_{-\infty}^{\infty} f(u, z|s, x) f(t, y|u, z) dz, \qquad s < u < t$$

Example: Consider a standard Brownian motion, (B_t). Suppose we know that $B_s = x$ where $s \geq 0$ and $x \in \mathbb{R}$. Conditional on this $B_t \sim N(x, t - s)$, for $t \geq s$, i.e.

$$\underbrace{\frac{\partial}{\partial y} P(B_t \leq y \mid B_s = x)}_{f(t,y|s,x)} = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-s)}}, \qquad -\infty < y < \infty.$$

From the Chapman-Kolmogorov equations (following the calculations of Einstein), it is possible to derive the equations,

$$\frac{\partial f}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial y^2} \qquad \text{(forward equation)}$$

$$\frac{\partial f}{\partial s} = -\frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2} \quad \text{(backward equation)},$$

for the transition density of a Brownian motion with variance parameter σ^2.

Note the connection between partial differential equations (PDEs) and random walks.

6 Diffusion processes

Diffusion processes is a class of processes generalizing Brownian motion.

Definition: Suppose (X_t) is a Markov process with continuous sample paths and suppose there exist functions $a(t,x)$, $b(t,x)$ such that

$$P(|X_{t+h} - X_t| > \epsilon \mid X_t) = o(h), \text{ for all } \epsilon > 0$$

$$E(X_{t+h} - X_t \mid X_t) = \underbrace{a(t, X_t)h}_{\text{instantaneous mean}} + o(h)$$

$$E((X_{t+h} - X_t)^2 \mid X_t) = \underbrace{b(t, X_t)h}_{\text{instantaneous variance}} + o(h)$$

Then (X_t) is called a diffusion process.

The above defines a stochastic differential equation.

$$dX_t = a(t, X_t)dt + \sqrt{b(t, X_t)}dB_t$$

where dB_t corresponds to infinitesimal increments of a standard Brownian motion. (In order to solve such equations we need theory from stochastic calculus not covered by this course.)

Under some additional technical conditions it is possible to derive the following equations specifying the transition densities.

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial y}(a(t,y)f) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(t,y)f) \quad \text{(forward equation)}$$

$$\frac{\partial f}{\partial s} = -a(s,x)\frac{\partial f}{\partial x} - \frac{1}{2}b(s,x)\frac{\partial^2 f}{\partial x^2} \quad \text{(backward equation)}$$

The diffusion process is time homogeneous if $a(t,x) = a(x)$ and $b(t,x) = b(x)$.

Brownian motion:
$a(t,x) = 0$, $b(t,x) = \sigma^2$.

The coefficients for a Brownian motion thus do not depend on x and t. This reflects the fact that Brownian motion is homogeneous in space and time by the independent and stationary increments properties.

Brownian motion with drift: $a(t,x) = m$, $b(t,x) = \sigma^2$.

$$X_t = mt + \sigma B_t$$

Compare:
Markov chain (discrete time countable state space): "generated" by a transition matrix $\mathbf{P} = (p_{ij})$.
Markov process (continuous time, countable state space): No smallest time unit. Generator: matrix of derivatives in 0, $\mathbf{Q} = \mathbf{P}'(0) = (p'_{ij}(0))$.
The transition probabilities are given as the solutions to the forward and backward equations $\mathbf{P}'(t) = \mathbf{QP}(t) = \mathbf{P}(t)\mathbf{Q}$.
Diffusion processes (continuous time, (uncountable) state space $\mathbb{R}$): The transition densities are given as the solutions to the forward and backward equations. The functions $a(t,x)$ and $b(t,x)$ play the role of $\mathbf{Q}$.

7 Geometric Brownian motion

$$X_t = e^{Y_t},$$

where Y_t is a Brownian motion with drift.

Used in modeling stock market prices. (e.g. in the Black-Scholes theory)
S_t asset price at time t (e.g. stock)
(The assumption that stock prices are log-normally distributed is doubtful. Real world data indicates some heavy-tailed distribution.)

Interpretation

$$dS_t = rS_t dt + \sigma S_t dB_t$$

$$\underbrace{\frac{dS_t}{S_t}}_{return} = \underbrace{rdt}_{mean\ return} + \underbrace{\underbrace{\sigma}_{volatility} dB_t}_{Gaussian\ random\ disturbance}$$

By using stochastic calculus (not a part of this course) it can be proved that the solution is given by

$$S_t = S_0 e^{(r-\sigma^2/2)t+\sigma B_t}.$$

8 Suggested exercises

Exercises Lawler:
8.7

Lecture 17 Markov Processes, 1MS012

9 The reflection principle and the distribution of $\max_{0\le u\le t} B_u$ and $|B_t|$.

Let $(B_t)_{t\ge 0}$ be a standard Brownian motion. Fix a state $x > 0$. Let τ_x be the first time at which B_t attains the value x.

Define

$$B_t^\star = \begin{cases} B_t & t \le \tau_x \\ x - (B_t - x) & t \ge \tau_x \end{cases} \qquad \text{reflection of } B_t.$$

If $B_t > x$, then by continuity of trajectories of $(B_t)_{t\ge 0}$, it follows that $\tau_x < t$, i.e. B_u (and $B_u^\star$) attain the value x somewhere in the interval $0 \le u < t$.

By symmetry,

$$\begin{aligned} P(B_t > x \mid \tau_x < t) &= P(B_t^\star > x \mid \tau_x < t) = P(B_t < x \mid \tau_x < t) \\ &= 1 - \underbrace{P(B_t \ge x \mid \tau_x < t)}_{P(B_t > x \mid \tau_x < t)} = 1/2. \end{aligned}$$

Therefore

$$P(B_t > x) = P(B_t > x,\ \tau_x < t) = \underbrace{P(B_t > x \mid \tau_x < t)}_{1/2} \quad \underbrace{P(\tau_x < t)}_{P(\max_{0\le u\le t} B_u > x)},$$

and thus

$$P(\max_{0\le u\le t} B_u > x) = 2P(B_t > x) = 2(1 - P(B_t \le x)) = 2(1 - \Phi(x/\sqrt{t})),$$

where Φ denotes the distribution-function of a $N(0,1)$-distributed random variable, i.e, $\phi(x) = \frac{d}{dx}\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Since

$$P(|B_t| > x) = P(B_t > x) + P(B_t < -x) = 2P(B_t > x)$$

it follows that Brownian motion reflected at the origin $|B_t|$ has the same distribution as $\max_{0\le u\le t} B_u$ for any fixed t. The common density is given by

$$f(x) = \frac{d}{dx}(1 - (2(1 - \Phi(x/\sqrt{t})))) = \frac{d}{dx}(2\Phi(x/\sqrt{t}) - 1)$$

$$= 2\phi(x/\sqrt{t})(1/\sqrt{t}) = \sqrt{\frac{2}{\pi t}}e^{-\frac{x^2}{2t}}, \qquad x \geq 0.$$

These processes are continuous time Markov process whose sample paths are continuous. The mean and variance are given by

$$E(\max_{0\leq u\leq t} B_u) = E|B_t| = \sqrt{\frac{2}{\pi t}} \underbrace{\int_0^\infty xe^{-\frac{x^2}{2t}}dx}_{[te^{-\frac{x^2}{2t}}]_{x=0}^{x=\infty}} = \sqrt{\frac{2t}{\pi}},$$

and

$$Var(\max_{0\leq u\leq t} B_u) = Var|B_t| = E(B_t^2) - 2t/\pi = Var(B_t) - 2t/\pi = t - 2t/\pi = t(1-2/\pi).$$

10 Zeros of Brownian motion

Theorem: Let $(B_u)_{u\geq 0}$ be a standard Brownian motion and $0 \leq s < t$ be two fixed time points. Then

$$P(B_u \text{ has at least one zero in the interval}(s,t)) = 1 - \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}}.$$

Proof: Suppose $0 < s < t$, and let E be the event that B_u has at least one zero in the interval (s,t), i.e. that there exists an u with $s < u < t$ such that $B_u = 0$. Let τ_x be the first time the Brownian motion takes the value x. By the reflection principle

$$P(\tau_x \leq t) \underbrace{=}_{\text{if } x>0} P(\max_{0\leq u\leq t} B_u \geq x) = 2(1 - \Phi(\frac{x}{\sqrt{t}})), \quad t > 0,$$

where Φ denotes the distribution function of a standard normal random variable thus satisfying $\frac{d\Phi(x)}{dx} = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Thus

$$\begin{aligned} P(E \mid B_s = \omega) &= P(\tau_{-\omega} < t-s) \underbrace{=}_{sym.} P(\tau_\omega < t-s) \\ &= \int_0^{t-s} \frac{d(P(\tau_\omega \leq y))}{dy}dy \\ &= \int_0^{t-s} 2\phi(\frac{\omega}{\sqrt{y}})\frac{\omega}{2y^{3/2}}dy = \int_0^{t-s} \frac{\omega e^{-\omega^2/(2y)}}{\sqrt{2\pi y^3}}dy, \quad \omega > 0. \end{aligned}$$

Thus

$$\begin{aligned}
P(E) &= \int_{-\infty}^{\infty} P(E \mid B_s = \omega) P(B_s = d\omega) \\
&= 2\int_{0}^{\infty} P(E \mid B_s = \omega) P(B_s = d\omega) \\
&= 2\int_0^{\infty}\int_0^{t-s} \frac{\omega e^{-\omega^2/(2y)}}{\sqrt{2\pi y^3}} \frac{1}{\sqrt{2\pi s}} e^{-\omega^2/(2s)} dy d\omega \\
&= \frac{1}{\sqrt{s}\pi}\int_0^{t-s} y^{-3/2} \underbrace{\int_0^{\infty} \omega e^{\frac{-\omega^2}{2}(\frac{1}{y}+\frac{1}{s})} d\omega}_{\frac{ys}{y+s}} dy \\
&= \frac{\sqrt{s}}{\pi}\int_0^{t-s} \frac{1}{(y+s)\sqrt{y}} dy \\
&\quad \text{Substitute } v = \sqrt{y/s}\text{, i.e. } y = v^2 s,\ dy = 2vsdv \\
&= \frac{1}{\pi}\int_0^{\sqrt{(t-s)/s}} \frac{1}{(v^2 s + s)v} 2vsdv = \frac{2}{\pi}\underbrace{\int_0^{\sqrt{(t-s)/s}} \frac{1}{1+v^2} dv}_{\arctan(\sqrt{(t-s)/s})} \\
&= \frac{2}{\pi}\arccos(\sqrt{\frac{s}{t}}) = \frac{2}{\pi}(\frac{\pi}{2} - \arcsin(\sqrt{\frac{s}{t}})) = 1 - \frac{2}{\pi}\arcsin(\sqrt{\frac{s}{t}}).
\end{aligned}$$

Corollary: Let $V = \sup\{u < t : B_u = 0\}$, be the time for the last zero before time $t > 0$. Then

$$\begin{aligned}
P(V > s) &= P(B_u \text{ has at least one zeros in } (s,t)) \\
&= 1 - \frac{2}{\pi}\arcsin(\sqrt{\frac{s}{t}}), \quad 0 < s < t.
\end{aligned}$$

Thus V has distribution function

$$F_V(s) = P(V \le s) = \frac{2}{\pi}\arcsin(\sqrt{\frac{s}{t}}), \quad 0 < s < t,$$

and hence density function

$$f_V(s) = \frac{d}{ds}F_V(s) = \frac{1}{\pi}\frac{1}{\sqrt{s(t-s)}}, \quad 0 < s < t.$$

Note that $f_V(s)$ is symmetric around $s = t/2$. Note also that $P(V > 0) = 1$, i.e. for any $t > 0$ the Brownian motion has a zero in the interval $(0, t)$ with probability one. It therefore follows that B_u has infinitely many zeros in $[0, t]$ for any $t > 0$.

11 Brownian bridge

Brownian bridge: B_t, $0 \leq t \leq 1$, conditional on the the event that $B_1 = 0$.

Let $(Z_t)_{t=0}^1$ be a Brownian bridge. We want to find the density function of Z_s for $0 < s < 1$.

More generally we find the conditional density of B_s given that $B_t = b$, $s < t$. (Thus Brownian bridge corresponds to $b = 0$ and $t = 1$.)

Conditional density:

$$\begin{aligned} f_{B_s|B_t=b}(x) &= \frac{f_{B_s,B_t}(x,b)}{f_{B_t}(b)} \underbrace{=}_{\text{indep. incr.}} \frac{f_{B_s}(x) f_{B_t-B_s}(b-x)}{f_{B_t}(b)} \\ &= \frac{\frac{1}{\sqrt{2\pi s}} e^{-\frac{x^2}{2s}} \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(b-x)^2}{2(t-s)}}}{\frac{1}{\sqrt{2\pi t}} e^{-\frac{b^2}{2t}}} \\ &= \frac{1}{\sqrt{2\pi \frac{s}{t}(t-s)}} e^{-\frac{x^2}{2s} - \frac{(b-x)^2}{2(t-s)} + \frac{b^2}{2t}} \\ &= \frac{1}{\sqrt{2\pi \frac{s}{t}(t-s)}} e^{-\frac{t}{2s(t-s)}(x - \frac{bs}{t})^2} \end{aligned}$$

Thus

$$B_s|B_t = b \ \sim N(\frac{bs}{t}, \frac{s(t-s)}{t}).$$

Therefore, in particular

$$E(B_s|B_t) = \frac{s}{t} B_t,$$

$$Var(B_s|B_t) = \frac{s}{t}(t-s),$$

and

$$Z_t \sim N(0, t(1-t)), \quad 0 \leq t \leq 1.$$

Proposition: If B_t is a standard Brownian motion, then

$$Z_t = B_t - tB_1, \ 0 \leq t \leq 1,$$

is a Brownian bridge.

12 Suggested exercises

Exercises Lawler:
8.4def, 8.11