UPPSALA UNIVERSITY
Department of Mathematics

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

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

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 distributions.

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.

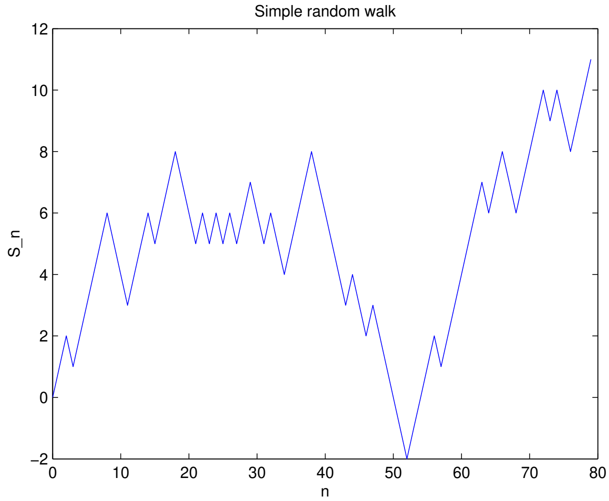


Figure 1: A trajectory of a simple random walk ($p = 1/2$)

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

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

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

for any $m > 0$. Thus

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

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.

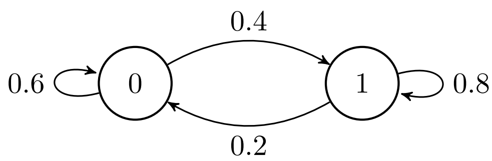
The array $\mathbf{P} = (p_{jk})_{j,k \in S}$, is called the **transition matrix** for the (time-homogeneous) Markov chain.
Note that $\sum_{k \in S} p_{jk} = 1$ for all j, and $p_{jk} \geq 0$, for all j and k, i.e. $\mathbf{P}$ is a stochastic matrix. (A stochastic matrix is a matrix with non-negative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



$\square$

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

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

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

$$\begin{aligned} p_{jk}^{(2)} &= P(X_2 = k | X_0 = j) = \sum_{i \in S} P(X_2 = k, X_1 = i | X_0 = j) \\ &\underbrace{=}_{\text{using (1)}} \sum_{i \in S} p_{ji} p_{ik} \leftarrow \text{element in row corresponding to state } j \text{ and} \\ &\qquad\qquad \text{column corresponding to state } k \text{ in } \mathbf{P}^2. \end{aligned}$$

we see

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

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

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

Therefore, by induction

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

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

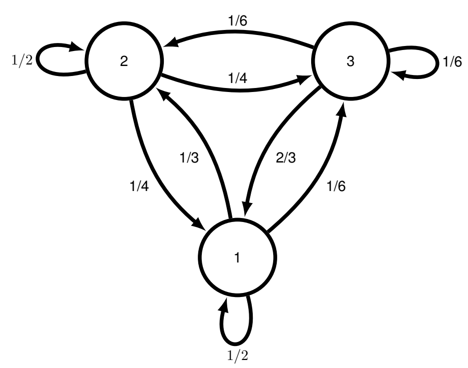
j is transient if
$$\sum_{n=1}^{\infty} P(T_{jj} = n) < 1.$$

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 $w_k = P$(Leave casino with N \$).
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

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

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