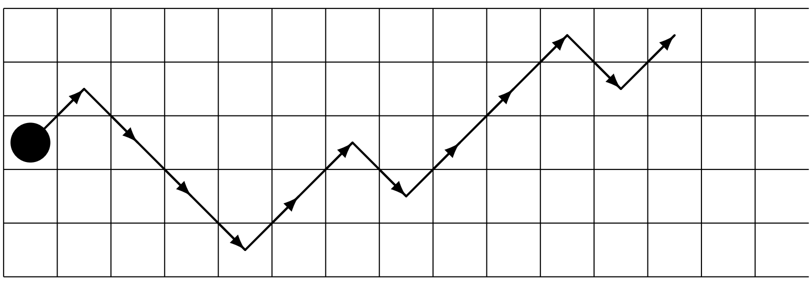
Markov Chains
Compact Lecture Notes and Exercises

September 2009

ACC Coolen
Department of Mathematics
King's College London



part of course 6CCM320A
part of course 6CCM380A

1 Introduction **3**

2 Definitions and properties of stochastic processes **7**
2.1 Stochastic processes 7
2.2 Markov chains 8
2.3 Examples 11

3 Properties of homogeneous finite state space Markov chains **15**
3.1 Simplification of notation & formal solution 15
3.2 Simple properties of the transition matrix and the state probabilities 16
3.3 Classification definitions based on accessibility of states 18
3.4 Graphical representation of Markov chains 19

4 Convergence to a stationary state **21**
4.1 Simple general facts on stationary states 21
4.2 Fundamental theorem for regular Markov chains 23
4.3 Examples 27

5 Some further developments of the theory **31**
5.1 Definition and physical meaning of detailed balance 31
5.2 Definition of characteristic times 32
5.3 Calculation of characteristic times 34

Appendix A Exercises **36**

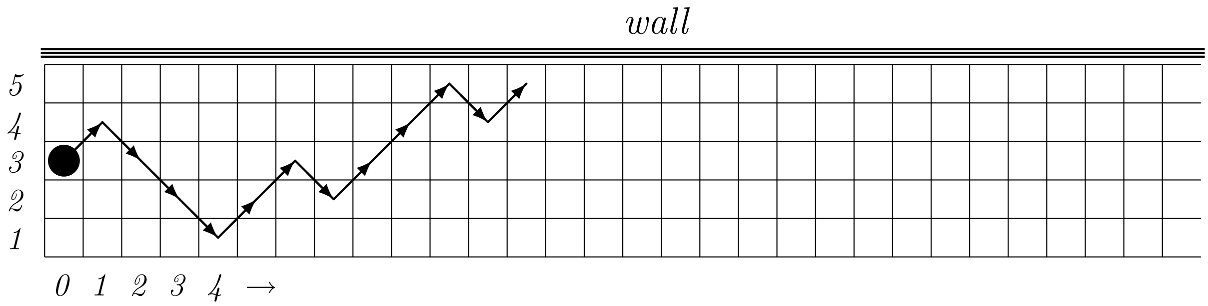
Appendix B Probability in a nutshell **38**

Appendix C Eigenvalues and eigenvectors of stochastic matrices **40**

1. Introduction

- **Random walks**

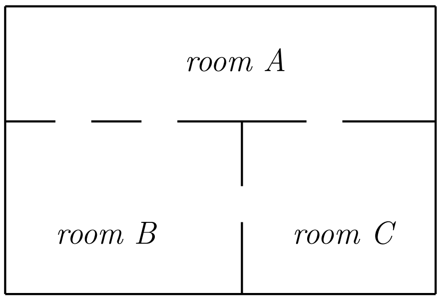
A drunk walks along a pavement of width 5. At each time step he/she moves one position forward, and one position either to the left or to the right with equal probabilities. Except: when in position 5 can only go to 4 (wall), when in position 1 and going to the right the process ends (drunk falls off the pavement).



How far will the walker get on average? What is the probability for the walker to arrive home when his/her home is K positions away?

- **The trained mouse**

A trained mouse lives in the house shown. A bell rings at regular intervals, and the mouse is trained to change rooms each time it rings. When it changes rooms, it is equally likely to pass through any of the doors in the room it is in. Approximately what fraction of its life will it spend in each room?

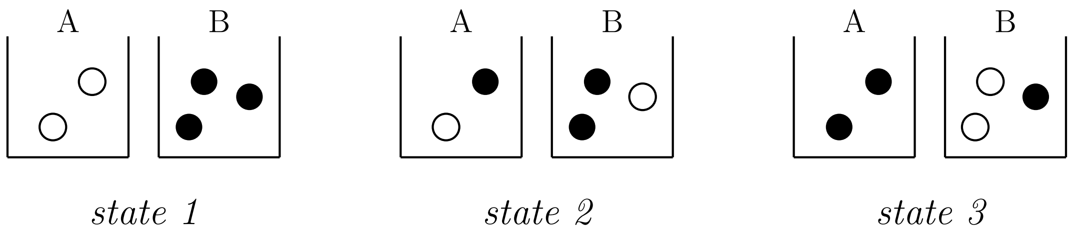


- **The fair casino**

You decide to take part in a roulette game, starting with a capital of C_0 pounds. At each round of the game you gamble £10. You lose this money if the roulette gives an even number, and you double it (so receive £20) if the roulette gives an odd number. Suppose the roulette is fair, i.e. the probabilities of even and odd outcomes are exactly 1/2. What is the probability that you will leave the casino broke?

- **The gambling banker**

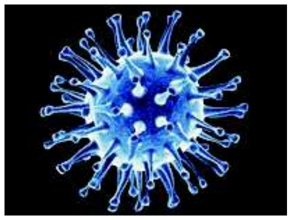
Consider two urns A and B in a casino game. Initially A contains two white balls, and B contains three black balls. The balls are then 'shuffled' repeatedly at discrete time steps according to the following rule: pick at random one ball from each urn, and swap them. The three possible states of the system during this (discrete time and discrete state space) stochastic process are shown below:



A banker decides to gamble on the above process. He enters into the following bet: at each step the bank wins 9M£ if there are two white balls in urn A, but has to pay 1M£ if not. What will happen to the bank?

- **Mutating virus**

  A virus can exist in N different strains. At each generation the virus mutates with probability $\alpha \in (0, 1)$ to another strain which is chosen at random. Very (medically) relevant question: what is the probability that the strain in the n-th generation of the virus is the same as that in the 0-th?

- **Simple population dynamics (Moran model)**

  We imagine a population of N individuals of two types, A and a. 'Birth': at each step we select at random one individual and add a new individual of the same type. 'Death': we then pick at random one individual and remove it. What is the probability to have i individuals of type a at step n? (subtlety: in between birth and death one has $N + 1$ individuals)

- **Google's fortune**

  How did Google displace all the other search engines about ten years ago? (Altavista, Webcrawler, etc). They simply had more efficient algorithms for defining the relevance of a particular web page, given a user's search request. Ranking of webpages generated by Google is defined via a 'random surfer' algorithm (stochastic process, Markov chain!). Some notation:

  $$N = \text{nr of webpages}, \qquad L_i = \text{links away from page } i, \qquad L_i \subseteq \{1, \ldots, N\}$$

  Random surfer goes from any page i to a new page j with probabilities:

  $$\begin{array}{ll} \text{with probability } q: & \text{pick any page from } \{1, \ldots, N\} \text{ at random} \\ \text{with probability } 1 - q: & \text{pick one of the links in } L_i \text{ at random} \end{array}$$

This is equivalent to

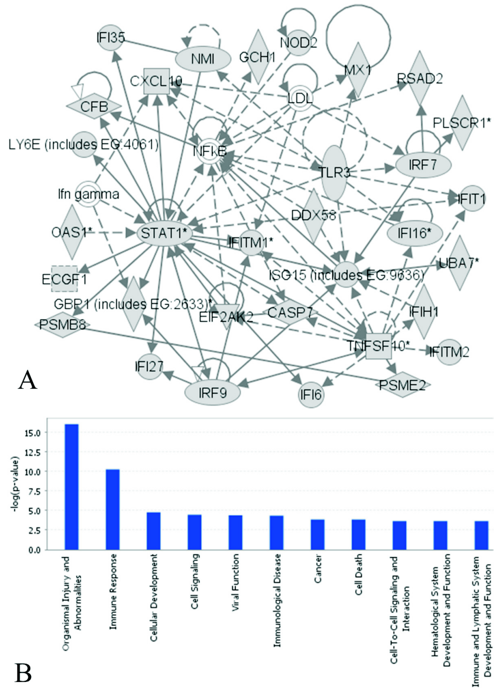
$$j \in L_i: \quad \text{Prob}[i \to j] = \frac{1-q}{|L_i|} + \frac{q}{N}, \qquad j \notin L_i: \quad \text{Prob}[i \to j] = \frac{q}{N}$$

Note: probabilities add up to zero:

$$\begin{aligned}\sum_{j=1}^{N} \text{Prob}[i \to j] &= \sum_{j \in L_i} \text{Prob}[i \to j] + \sum_{j \notin L_i} \text{Prob}[i \to j] \\ &= \sum_{j \in L_i} \Big(\frac{1-q}{|L_i|} + \frac{q}{N}\Big) + \sum_{j \notin L_i} \frac{q}{N} \\ &= |L_i|\Big(\frac{1-q}{|L_i|} + \frac{q}{N}\Big) + \Big(N - |L_i|\Big)\frac{q}{N} \\ &= 1 - q + q \;=\; 1\end{aligned}$$

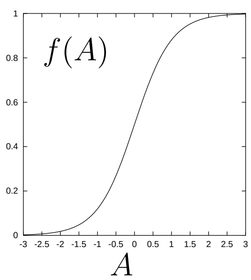
Calculate the fraction f_i of times the site will be visited asymptotically if the above process is iterated for a very large number of iterations. Then f_i will define Google's ranking of the page i. Can we predict f_i? What can one do to increase one's ranking?

- **Gene regulation in cells - cell types & stability**



The genome contains the blueprint of an organism. Each gene in the genome is a code for the production of a specific protein. All cells in an organism contain the same genes, but not all genes are switched on ('expressed'), which allows for different cell types. Let $\theta_i \in \{0,1\}$ indicate whether gene i is switched on. This is controlled by other genes, via a dynamical process of the type

$$\text{Prob}[\theta_i(t{+}1){=}1] = f\Big(\underbrace{\sum_j J^+_{ij}\theta_j(t)}_{\text{activators}} - \underbrace{\sum_j J^-_{ij}\theta_j(t)}_{\text{repressors}}\Big)$$



cell type at time t: $\boldsymbol{\theta}(t) = (\theta_1(t), \ldots, \theta_N(t))$
evolution of cell type:

$$\boldsymbol{\theta}(0) \;\to\; \boldsymbol{\theta}(1) \;\to\; \boldsymbol{\theta}(2) \;\to\; \ldots\;?$$

multiple stationary states $\boldsymbol{\theta}$ of the dynamics?
stable against perturbations (no degenerated 'runaway cells')?
dependence on activation & suppression efficacies $J^{\pm}_{ij}$?

- **Many many other real-world processes ...**

  Dynamical systems with stochastic (partially or fully random) dynamics.
  Some are really fundamentally random, others are 'practically' random.
  E.g.

  – physics: quantum mechanics, solids/liquids/gases at nonzero temperature, diffusion
  – biology: interacting molecules, cell motion, predator-prey models,
  – medicine: epidemiology, gene transmission, population dynamics,
  – commerce: stock markets & exchange rates, insurance risk, derivative pricing,
  – sociology: herding behaviour, traffic, opinion dynamics,
  – computer science: internet traffic, search algorithms,
  – leisure: gambling, betting,

2. Definitions and properties of stochastic processes

We first define stochastic processes generally, and then show how one finds discrete time Markov chains as probably the most intuitively simple class of stochastic processes.

2.1. Stochastic processes

- defn: Stochastic process

  Dynamical system with *stochastic* (i.e. at least partially random) dynamics. At each time $t \in [0, \infty\rangle$ the system is in one state X_t, taken from a set S, the *state space*. One often writes such a process as $X = \{X_t:\ t \in [0, \infty\rangle\}$.

  consequences, conventions

  (i) We can only speak about the probabilities to find the system in certain states at certain times: each X_t is a *random variable*.

  (ii) To define a process fully: specify the probabilities (or probability densities) for the X_t at all t, or give a recipe from which these can be calculated.

  (iii) If time discrete: label time steps by integers $n \geq 0$, write $X = \{X_n:\ n \geq 0\}$.

- defn: Joint state probabilities for process with discrete time and discrete state space

  Processes with discrete time and discrete state space are conceptually the simplest: $X = \{X_n:\ n \geq 0\}$ with $S = \{s_1, s_2, \ldots\}$. From now on we define $\sum_X \equiv \sum_{X \in S}$, unless stated otherwise. Here we can define for any set of time labels $\{n_1, \ldots, n_L\} \subseteq \mathbb{N}^L$:

$$
\begin{aligned}
P(X_{n_1}, X_{n_2}, \ldots, X_{n_L}) = \ & \text{the probability of finding the system} \\
& \text{at the specified times } \{n_1, n_2, \ldots, n_L\} \text{ in the} \\
& \text{states } (X_{n_1}, X_{n_2}, \ldots, X_{n_L}) \in S^L
\end{aligned}
\tag{1}
$$

  consequences, conventions

  (i) The probabilistic interpretation of (1) demands

$$
\text{non–negativity:} \quad P(X_{n_1}, \ldots, X_{n_L}) \geq 0 \quad \forall (X_{n_1}, \ldots, X_{n_L}) \in S^L \tag{2}
$$

$$
\text{normalization:} \quad \sum_{(X_{n_1}, \ldots, X_{n_L})} P(X_{n_1}, \ldots, X_{n_L}) = 1 \tag{3}
$$

$$
\text{marginalization:} \quad P(X_{n_2}, \ldots, X_{n_L}) = \sum_{X_{n_1}} P(X_{n_1}, X_{n_2}, \ldots, X_{n_L}) \tag{4}
$$

(ii) All joint probabilities in (1) can be obtained as marginals of the full probabilities over all times, i.e. of the following *path probabilities* (if T is chosen large enough):

$$P(X_0, X_1, \dots, X_T) = \text{the probability of finding the system at the times } \{0, 1, \dots, T\} \text{ in the states } (X_0, X_1, \dots, X_T) \in S^{T+1} \tag{5}$$

Knowing (5) permits the calculation of any quantity of interest. The process is defined fully by giving S and the probabilities (5) for all T.

(iii) Stochastic dynamical equation: a formula for $P(X_{n+1}|X_0, \dots, X_n)$, the probability of finding a state X_{n+1} given knowledge of the past states, which is defined as

$$P(X_{n+1}|X_0, \dots, X_n) = \frac{P(X_0, \dots, X_n, X_{n+1})}{P(X_0, \dots, X_n)} \tag{6}$$

2.2. Markov chains

Markov chains are discrete state space processes that have the Markov property. Usually they are defined to have also discrete time (but definitions vary slightly in textbooks).

- defn: the Markov property

  A discrete time and discrete state space stochastic process is *Markovian* if and only if the conditional probabilities (6) do not depend on $(X_0, \dots, X_n)$ in full, but only on the most recent state X_n:

$$P(X_{n+1}|X_0, \dots, X_n) = P(X_{n+1}|X_n) \tag{7}$$

  The likelihood of going to any next state at time $n+1$ depends only on the state we find ourselves in at time n. The system is said to have *no memory.*

  consequences, conventions

  (i) For a Markovian chain one has

$$P(X_0, \dots, X_T) = P(X_0) \prod_{n=1}^{T} P(X_n|X_{n-1}) \tag{8}$$

  Proof:

$$\begin{aligned} P(X_0, \dots, X_T) &= P(X_T|X_0, \dots, X_{T-1})P(X_0, \dots, X_{T-1}) \\ &= P(X_T|X_0, \dots, X_{T-1})P(X_{T-1}|X_0, \dots, X_{T-2})P(X_0, \dots, X_{T-2}) \\ &= \quad \vdots \\ &= P(X_T|X_0, \dots, X_{T-1})P(X_{T-1}|X_0, \dots, X_{T-2}) \dots \\ &\qquad \dots P(X_2|X_0, X_1)P(X_1|X_0)P(X_0) \\ Markovian: \quad &= P(X_T|X_{T-1})P(X_{T-1}|X_{T-2}) \dots P(X_2|X_1)P(X_1|X_0)P(X_0) \end{aligned}$$

$$= P(X_0)\prod_{n=1}^{T} P(X_n|X_{n-1}) \qquad []$$

(ii) Let us define the probability $P_n(X)$ to find the system at time $n \geq 0$ in state $X \in S$:

$$P_n(X) = \sum_{X_0} \ldots \sum_{X_n} P(X_0, \ldots, X_n)\delta_{X,X_n} \tag{9}$$

This defines a *time dependent probability measure* on the set S, with the usual properties $\sum_X P_n(X) = 1$ and $P_n(X) \geq 0$ for all $X \in S$ and all n.

(iii) For any two times $m > n \geq 0$ the measures $P_n(X)$ and $P_m(X)$ are related via

$$P_m(X) = \sum_{X'} W_{X,X'}(m,n)P_n(X') \tag{10}$$

with

$$W_{X,X'}(m,n) = \sum_{X_n} \ldots \sum_{X_m} \delta_{X,X_m}\delta_{X',X_n} \prod_{\ell=n+1}^{m} P(X_\ell|X_{\ell-1}) \tag{11}$$

$W_{X,X'}(m,n)$ is the probability to be in state X at time m, given the system was in state X' at time n, i.e. the likelihood to travel from $X' \to X$ in the interval $n \to m$.

Proof:
subtract the two sides of (10), insert the definition of $W_{X,X'}(m,n)$, use (9), and sum over x',

$$\begin{aligned}
\mathrm{LHS} - \mathrm{RHS} &= P_m(X) - \sum_{X'} W_{X,X'}(m,n)P_n(X') \\
&= P_m(X) - \sum_{X'}\sum_{X_n} \ldots \sum_{X_m} \delta_{X,X_m}\delta_{X',X_n}\Big[\prod_{\ell=n+1}^{m} P(X_\ell|X_{\ell-1})\Big]P_n(X') \\
&= \sum_{X_0} \ldots \sum_{X_m} P(X_0, \ldots, X_m)\delta_{X,X_m} \\
&\quad - \sum_{X'}\sum_{X_n} \ldots \sum_{X_m} \delta_{X,X_m}\delta_{X',X_n}\Big[\prod_{\ell=n+1}^{m} P(X_\ell|X_{\ell-1})\Big]\sum_{X'_0} \ldots \sum_{X'_n} P(X'_0, \ldots, X'_n)\delta_{X',X'_n} \\
&= \sum_{X_{n+1}} \ldots \sum_{X_m} \delta_{X,X_m}\Big\{\sum_{X_0} \ldots \sum_{X_n} P(X_0, \ldots, X_m) \\
&\qquad - \sum_{X_n}\Big[\prod_{\ell=n+1}^{m} P(X_\ell|X_{\ell-1})\Big]\sum_{X_0} \ldots \sum_{X_{n-1}} P(X_0, \ldots, X_n)\Big\} \\
&= \sum_{X_1} \ldots \sum_{X_m} \delta_{X,X_m}\Big\{P(X_0, \ldots, X_m) - \Big[\prod_{\ell=n+1}^{m} P(X_\ell|X_{\ell-1})\Big]P(X_0, \ldots, X_n)\Big\} \\
&= 0 \qquad []
\end{aligned}$$

- defn: homogeneous (or stationary) Markov chains

  A Markov chain with transition probabilities that depend only on the length $m-n$ of the separating time interval,

  $$W_{X,X'}(m,n) = W_{X,X'}(m-n), \tag{12}$$

  is called a *homogeneous (or stationary) Markov chain.* Here the absolute time is irrelevant: if we re-set our clocks by a uniform shift $n \to n+K$ for fixed K, then all probabilities to make certain transitions during given time intervals remain the same.

  consequences, conventions

  (i) The transition probabilities in a homogeneous Markov chain obey the *Chapman-Kolmogorov equation*:

  $$\forall X, Y \in S: \quad W_{X,Y}(m) = \sum_{X''} W_{X,X''}(1) W_{X'',Y}(m-1) \tag{13}$$

  The likelihood to go from Y to X in m steps is the sum over all paths that go first in $m-1$ steps to any intermediate state X', followed by one step from X' to X. The Markovian property guarantees that the last step is independent of *how* we got to X'. Stationarity ensures that the likelihood to go in $m-1$ steps to X' is not dependent on *when* various intermediate steps were made.

  Proof:
  Rewrite $P_m(X)$ in two ways, first by choosing $n=0$ in the right-hand side of (10), second by choosing $n=m-1$ in the right-hand side of (10):

  $$\forall X \in S: \quad \sum_{X'} W_{X,X'}(m) P_0(X') = \sum_{X'} W_{X,X'}(1) P_{m-1}(X') \tag{14}$$

  Next we use (10) once more, now to rewrite $P_{m-1}(X')$ by choosing $n=0$:

  $$\forall X \in S: \quad \sum_{X'} W_{X,X'}(m) P_0(X') = \sum_{X'} W_{X,X'}(1) \sum_{X''} W_{X',X''}(m-1) P_0(X'') \tag{15}$$

  Finally we choose $P_0(X) = \delta_{X,Y}$, and demand that the above is true for any $Y \in S$:

  $$\forall X, Y \in S: \quad W_{X,Y}(m) = \sum_{X'} W_{X,X'}(1) W_{X',Y}(m-1)$$ □

- defn: stochastic matrix

  The one-step transition probabilities $W_{XY}(1)$ in a homogeneous Markov chain are from now on interpreted as entries of a matrix $\boldsymbol{W} = \{W_{XY}\}$, the so-called *transition matrix* of the chain, or *stochastic matrix*.

  consequences, conventions:

  (i) In a homogeneous Markov chain one has

  $$P_{n+1}(X) = \sum_Y W_{XY} P_n(Y) \quad \text{for all } n \in \{0, 1, 2, \ldots\} \tag{16}$$

  Proof:
  This follows from setting $m = n+1$ in (10), together with the defn $W_{XY} = W_{XY}(1)$.
  []

  (ii) In a homogeneous Markov chain one has

  $$P(X_0, \ldots, X_T) = P(X_0) \prod_{n=1}^{T} W_{X_n, X_{n-1}} \tag{17}$$

  Proof:
  This follows directly from (8), in combination with our identification of W_{XY} in Markov chains as the probability to go from Y to X in one time step. []

2.3. Examples

Note: the mathematical analysis of stochastic equations can be nontrivial, but most mistakes are in fact made before that, while *translating* a problem into stochastic equations of the type $P(X_n) = \ldots$.

- Some dice rolling examples:

  (i) X_n = number of sixes thrown after n rolls?

  $$\begin{array}{lll} \text{6 at stage } n: & X_n = X_{n-1} + 1, & \text{probability } 1/6 \\ \text{no 6 at stage } n: & X_n = X_{n-1}, & \text{probability } 5/6 \end{array}$$

  So $P(X_n)$ depends only on X_{n-1}, not on earlier values: Markovian!
  If X_{n-1} had been known exactly:

  $$P(X_n|X_{n-1}) = \frac{1}{6}\delta_{X_n, X_{n-1}+1} + \frac{5}{6}\delta_{X_n, X_{n-1}}$$

  If X_{n-1} is not known exactly, average over all possible values of X_{n-1}:

  $$P(X_n) = \sum_{X_{n-1}} \Big[\frac{1}{6}\delta_{X_n, X_{n-1}+1} + \frac{5}{6}\delta_{X_n, X_{n-1}}\Big] P(X_{n-1})$$

  Hence

  $$W_{XY} = \frac{1}{6}\delta_{X, Y+1} + \frac{5}{6}\delta_{X, Y}$$

Simple test: $\sum_X W_{XY} = 1$.

(ii) X_n = largest number thrown after n rolls?

$$\begin{array}{lll} X_n = X_{n-1}: & X_{n-1} \text{ possible throws,} & \text{probability } X_{n-1}/6 \\ X_n > X_{n-1}: & 6 - X_{n-1} \text{ possible throws,} & \text{probability } (6 - X_{n-1})/6 \end{array}$$

So $P(X_n)$ depends only on X_{n-1}, not on earlier values: Markovian!
If X_{n-1} had been known exactly (note: if $X_n > X_{n-1}$ then each of the $6 - X_{n-1}$ possibilities is equally likely):

$$\begin{array}{ll} P(X_n|X_{n-1}) = 0 & \text{for } X_n < X_{n-1} \\ P(X_n|X_{n-1}) = X_{n-1}/6 & \text{for } X_n = X_{n-1} \\ P(X_n|X_{n-1}) = (1 - X_{n-1}/6)/(6 - X_{n-1}) = \dfrac{1}{6} & \text{for } X_n > X_{n-1} \end{array}$$

If X_{n-1} is not known exactly, average over all possible values of X_{n-1}:

$$P(X_n) = \sum_{X_{n-1}} P(X_{n-1}) \begin{cases} 0 & \text{if} \quad X_{n-1} > X_n \\ X_{n-1}/6 & \text{if} \quad X_{n-1} = X_n \\ 1/6 & \text{if} \quad X_{n-1} < X_n \end{cases}$$

Hence

$$W_{XY} = \begin{cases} 0 & \text{if} \quad Y > X \\ Y/6 & \text{if} \quad Y = X \\ 1/6 & \text{if} \quad Y < X \end{cases}$$

Simple test:

$$\sum_X W_{XY} = \sum_{X<Y} 0 + \frac{Y}{6} + \sum_{X>Y} \frac{1}{6} = \frac{Y}{6} + \frac{1}{6}(6 - Y) = 1$$

(iii) X_n = number of iterations since most recent six?

$$\begin{array}{lll} 6 \text{ at stage } n: & X_n = 0, & \text{probability } 1/6 \\ \text{no } 6 \text{ at stage } n: & X_n = X_{n-1} + 1, & \text{probability } 5/6 \end{array}$$

So $P(X_n)$ depends only on X_{n-1}, not on earlier values: Markovian!
If X_{n-1} had been known exactly:

$$P(X_n|X_{n-1}) = \frac{1}{6}\delta_{X_n,0} + \frac{5}{6}\delta_{X_n,X_{n-1}+1}$$

If X_{n-1} is not known exactly, average over all possible values of X_{n-1}:

$$P(X_n) = \sum_{X_{n-1}} \Big[\frac{1}{6}\delta_{X_n,0} + \frac{5}{6}\delta_{X_n,X_{n-1}+1}\Big] P(X_{n-1})$$

Hence

$$W_{XY} = \frac{1}{6}\delta_{X,0} + \frac{5}{6}\delta_{X,Y+1}$$

Simple test: $\sum_X W_{XY} = 1$.

- The drunk on the pavement (see section 1). Let $X_n \in \{0,1,\ldots,5\}$ denote the position on the pavement of the drunk, with $X_n = 0$ representing him lying on the street. Let $\sigma_n \in \{-1,1\}$ indicate the (random) direction he takes after step $n-1$ (provided he has a choice at that moment, i.e. provided $X_{n-1} \neq 0, 5$. If σ_n and X_{n-1} had been known exactly:

$$\begin{aligned} X_{n-1} = 0: &\quad X_n = 0 \\ 0 < X_{n-1} < 5: &\quad X_n = X_{n-1} + \sigma_n \\ X_{n-1} = 5: &\quad X_n = 4 \end{aligned}$$

Since $\sigma_n \in \{-1,1\}$ with equal probabilities:

$$P(X_n|X_{n-1}) = \begin{cases} \delta_{X_n,0} & \text{if} \quad X_{n-1} = 0 \\ \frac{1}{2}\delta_{X_n,X_{n-1}+1} + \frac{1}{2}\delta_{X_n,X_{n-1}-1} & \text{if} \quad 0 < X_{n-1} < 5 \\ \delta_{X_n,4} & \text{if} \quad X_{n-1} = 5 \end{cases}$$

If X_{n-1} is not known exactly, average over all possible values of X_{n-1}:

$$W_{XY} = \begin{cases} \delta_{X,0} & \text{if} \quad Y = 0 \\ \frac{1}{2}\delta_{X,Y+1} + \frac{1}{2}\delta_{X,Y-1} & \text{if} \quad 0 < Y < 5 \\ \delta_{X,4} & \text{if} \quad Y = 5 \end{cases}$$

Simple test:

$$\begin{aligned} \sum_X W_{XY} &= \begin{cases} \sum_X \delta_{X,0} & \text{if} \quad Y = 0 \\ \sum_X \left[\frac{1}{2}\delta_{X,Y+1} + \frac{1}{2}\delta_{X,Y-1}\right] & \text{if} \quad 0<Y<5 \\ \sum_X \delta_{X,4} & \text{if} \quad Y = 5 \end{cases} \\ &= \begin{cases} 1 & \text{if} \quad Y = 0 \\ 1 & \text{if} \quad 0<Y<5 \\ 1 & \text{if} \quad Y = 5 \end{cases} \end{aligned}$$

- The Moran model of population dynamics (see section 1). Define X_n as the number of individuals of type a at time n. The number of type A individuals at time n is then $N - X_n$. However, each step involves *two* events: 'birth' (giving $X_{n-1} \to X_n'$), and 'death' (giving $X_n' \to X_n$). The likelihood to pick an individual of a certain type, given the numbers of type a an A are $(X, N-X)$, is:

$$\text{Prob}[a] = \frac{X}{N}, \qquad \text{Prob}[A] = \frac{N-X}{N}$$

First we suppose we know X_{n-1}. After the 'birth' process one has $N+1$ individuals, with X_n' of type a and $N+1-X_n'$ of type A, and

$$P(X_n'|X_{n-1}) = \frac{X_{n-1}}{N}\delta_{X_n',X_{n-1}+1} + \frac{N-X_{n-1}}{N}\delta_{X_n',X_{n-1}}$$

After the 'death' process one has again N individuals, with X_n of type a and $N - X_n$ of type A. If we know X'_n then

$$P(X_n|X'_n) = \frac{X'_n}{N+1}\delta_{X_n,X'_n-1} + \frac{N+1-X'_n}{N+1}\delta_{X_n,X'_n}$$

We can now simply combine the previous results:

$$\begin{aligned}
P(X_n|X_{n-1}) &= \sum_{X'_n} P(X_n|X'_n)P(X'_n|X_{n-1}) \\
&= \sum_{X'_n}\left[\frac{X'_n}{N+1}\delta_{X_n,X'_n-1} + \frac{N+1-X'_n}{N+1}\delta_{X_n,X'_n}\right]\left[\frac{X_{n-1}}{N}\delta_{X'_n,X_{n-1}+1} + \frac{N-X_{n-1}}{N}\delta_{X'_n,X_{n-1}}\right] \\
&= \frac{1}{N(N+1)}\sum_{X'_n}\Big\{X'_nX_{n-1}\delta_{X_n,X'_n-1}\delta_{X'_n,X_{n-1}+1} + X'_n(N-X_{n-1})\delta_{X_n,X'_n-1}\delta_{X'_n,X_{n-1}} \\
&\quad +(N+1-X'_n)X_{n-1}\delta_{X_n,X'_n}\delta_{X'_n,X_{n-1}+1} + (N+1-X'_n)(N-X_{n-1})\delta_{X_n,X'_n}\delta_{X'_n,X_{n-1}}\Big\} \\
&= \frac{(X_{n-1}+1)X_{n-1}}{N(N+1)}\delta_{X_n,X_{n-1}} + \frac{X_{n-1}(N-X_{n-1})}{N(N+1)}\delta_{X_n,X_{n-1}-1} \\
&\quad + \frac{(N-X_{n-1})X_{n-1}}{N(N+1)}\delta_{X_n,X_{n-1}+1} + \frac{(N+1-X_{n-1})(N-X_{n-1})}{N(N+1)}\delta_{X_n,X_{n-1}} \\
&= \frac{(X_{n-1}+1)X_{n-1} + (N+1-X_{n-1})(N-X_{n-1})}{N(N+1)}\delta_{X_n,X_{n-1}} \\
&\quad + \frac{X_{n-1}(N-X_{n-1})}{N(N+1)}\delta_{X_n,X_{n-1}-1} + \frac{(N-X_{n-1})X_{n-1}}{N(N+1)}\delta_{X_n,X_{n-1}+1}
\end{aligned}$$

Hence

$$W_{XY} = \frac{(Y+1)Y+(N+1-Y)(N-Y)}{N(N+1)}\delta_{X,Y} + \frac{Y(N-Y)}{N(N+1)}\delta_{X,Y-1} + \frac{(N-Y)Y}{N(N+1)}\delta_{X,Y+1}$$

Simple test:

$$\sum_X W_{XY} = \frac{(Y+1)Y+(N+1-Y)(N-Y)}{N(N+1)} + \frac{Y(N-Y)}{N(N+1)} + \frac{(N-Y)Y}{N(N+1)} = 1$$

Note also that:

$$Y=0: \quad W_{XY} = \delta_{X,0}, \qquad Y=N: \quad W_{XY} = \delta_{X,N}$$

representing the stationary situations where either the a individuals or the A individuals have died out.

3. Properties of homogeneous finite state space Markov chains

3.1. Simplification of notation & formal solution

Since the state space S is discrete, we can represent/label the states by integer numbers, and write simply $S = \{1, 2, 3, \ldots\}$. Now the X are themselves *integer* random variables. To exploit optimally the simple nature of Markov chains we change our notation:

$$S \to \{1, 2, 3, \ldots\}, \qquad X, Y \to i, j \qquad P_n(X) \to p_i(n), \qquad W_{XY} \to p_{ji} \tag{18}$$

From now on we will limit ourselves for simplicity to Markov chains with *finite state spaces* $S = \{1, \ldots, |S|\}$. This is not essential but removes distracting technical complications.

- defn: homogeneous Markov chains in standard notation

  In our new notation the dynamical eqn (16) of the Markov chain becomes

  $$\forall n \in \mathbb{N},\ \forall i \in S: \quad p_i(n+1) = \sum_j p_{ji}\ p_j(n) \tag{19}$$

  where:

where:	$n \in \mathbb{N}$:	time in the process
	$p_i(n) \geq 0$:	probability that the system is in state $i \in S$ at time n
	$p_{ji} \geq 0$:	probability that, when in state j, it will move to i next

  consequences, conventions

  (i) The probability (17) of the system taking a specific path of states becomes

  $$P(X_0 = i_0, X_1 = i_1, \ldots, X_T = i_T) = \Big(\prod_{n=1}^{T} p_{i_{n-1}, i_n}\Big) p_{i_0}(0) \tag{20}$$

  (ii) Upon denoting the $|S| \times |S|$ transition matrix as $\boldsymbol{P} = \{p_{ij}\}$ and the time-dependent state probabilities as a time-dependent vector $\boldsymbol{p}(n) = (p_1(n), \ldots, p_{|S|}(n))$, the dynamical equation (19) can be interpreted as multiplication from the *right* of a vector by a matrix $\boldsymbol{P}$ (alternatively: from the *left* by $\boldsymbol{P}^\dagger$, where $(\boldsymbol{P}^\dagger)_{ij} = p_{ji}$):

  $$\forall n \in \mathbb{N}: \quad \boldsymbol{p}(n+1) = \boldsymbol{p}(n)\boldsymbol{P} \tag{21}$$

  (iii) the formal solution of eqn (19) is

  $$p_i(n) = \sum_j (\boldsymbol{P}^n)_{ji}\ p_j(0) \tag{22}$$

  proof:
  We iterate formula (21). This gives $\boldsymbol{p}(1) = \boldsymbol{p}(0)\boldsymbol{P}$, $\boldsymbol{p}(2) = \boldsymbol{p}(1)\boldsymbol{P} = \boldsymbol{p}(0)\boldsymbol{P}^2$, $\boldsymbol{p}(3) = \boldsymbol{p}(2)\boldsymbol{P} = \boldsymbol{p}(0)\boldsymbol{P}^3$, etc. Generally: $\boldsymbol{p}(n) = \boldsymbol{p}(0)\boldsymbol{P}^n$. From this one extracts

  $$\forall n \in \mathbb{N},\ \forall i \in S: \quad p_i(n) = (\boldsymbol{p}(0)\boldsymbol{P}^n)_i = \sum_j (\boldsymbol{P}^n)_{ji} p_j(0) \qquad \Box$$

3.2. Simple properties of the transition matrix and the state probabilities

- properties of the transition matrix

  Any transition matrix $\boldsymbol{P}$ must satisfy the conditions below. Conversely, any $|S|\times|S|$ matrix that satisfies these conditions can be interpreted as a Markov chain transition matrix:

  (i) The first is a direct and trivial consequence of the meaning of p_{ij}:
  $$\forall (i,j)\in S^2: \quad p_{ij}\in[0,1] \tag{23}$$
  (ii) normalization:
  $$\forall k\in S: \quad \sum_i p_{ki}=1 \tag{24}$$
  proof:
  This follows from the demand that the state probabilities $p_i(n)$ are to be normalized at all times, in combination with (19) where we choose $p_j(n)=\delta_{jk}$:
  $$1=\sum_i p_i(n+1)=\sum_{ij} p_{ji}p_j(n)=\sum_i p_{ki}$$
  Since this must hold for any choice of $k\in S$, it completes our proof. []

  Note 1: a matrix that satisfies (23,24) is often called 'stochastic matrix'.
  Note 2: instead of (23) we could weaken this first condition to $p_{ij}\geq 0$ for all $(i,j)\in S^2$, since combination (24) will ensure that $p_{ij}\leq 1$ for all $(i,j)\in S^2$.

- conservation of sign and normalization of the state probabilities

  (i) If $\boldsymbol{P}$ is a transition matrix of a Markov chain defined by (19), then
  $$\sum_i p_i(n+1)=\sum_i p_i(n) \tag{25}$$
  proof:
  this follows from (19) and the imposed normalization (24):
  $$\sum_i p_i(n+1)=\sum_i\sum_j p_{ji}p_j(n)=\sum_j\sum_i p_{ji}p_j(n)=\sum_j p_j(n)$$ []

  (ii) If $\boldsymbol{P}$ is a transition matrix of a Markov chain defined by (19), and $p_i(n)\geq 0$ for all $i\in S$, then
  $$p_i(n+1)\geq 0 \quad \text{for all } i\in S \tag{26}$$
  proof:
  this follows from (19) and the non-negatively of all p_{ij} and all $p_j(n)$:
  $$p_i(n+1)=\sum_j p_{ji}p_j(n)\geq\sum_j 0=0$$ []

(iii) If $\boldsymbol{P}$ is a transition matrix of a Markov chain defined by (19), and the $p_i(0)$ represent normalized state probabilities, i.e. $p_i(0) \in [0,1]$ with $\sum_i p_i(0) = 1$, then

$$\forall n \in \mathbb{N}: \quad p_i(n) \in [0,1] \;\text{ for all } i \in S, \qquad \sum_i p_i(n) = 1 \tag{27}$$

proof:
this follows by combining and iterating the previous identities, noting that $p_i(n) \geq 0$ for all $i \in S$ together with $\sum_i p_i(n) = 1$ implies also that $p_i(n) \leq 1$ for all $i \in S$. []

- properties involving powers of the transition matrix

(i) meaning of powers of the transition matrix

$$(\boldsymbol{P}^m)_{ki} \;=\; \text{the probability for the system to move in } m \text{ steps from state } k \text{ to state } i \tag{28}$$

proof:
calculate the stated probability by putting the system at time zero in state k, i.e. $p_j(0) = \delta_{jk}$, and use (22) to find the probability of seeing it n steps later in state i:

$$p_i(n) = \sum_j (\boldsymbol{P}^n)_{ji} \delta_{jk} = (\boldsymbol{P}^n)_{ki}$$

(note: the stationarity of the chain ensures that the result cannot depend on us choosing the moment where the system is state k to be time zero!) []

(ii) If $\boldsymbol{P}$ is a transition matrix, then also $\boldsymbol{P}^\ell$ has the properties of a transition matrix for any $\ell \in \mathbb{N}$: $(\boldsymbol{P}^\ell)_{ik} \geq 0 \;\forall (i,k) \in S^2$ and $\sum_i (\boldsymbol{P}^\ell)_{ki} = 1 \;\forall k \in S$.

proof:
This follows already implicitly from (28), but can also be shown directly by induction. For $m = 0$ one has $\boldsymbol{P}^0 = \mathbb{1}$ (identity matrix), so $(\boldsymbol{P}^0)_{ik} = \delta_{ik}$ and the conditions are obviously met. Next we prove the induction step, assuming $\boldsymbol{P}^\ell$ to be a stochastic‘ matrix and proving the same for $\boldsymbol{P}^{\ell+1}$:

$$(\boldsymbol{P}^{\ell+1})_{ki} \;=\; \sum_j p_{kj} (\boldsymbol{P}^\ell)_{ji} \;\geq\; \sum_j 0 \;=\; 0$$

$$\sum_i (\boldsymbol{P}^{\ell+1})_{ki} = \sum_i \sum_j (\boldsymbol{P}^\ell)_{kj} p_{ji} = \sum_j \Big(\sum_i p_{ji} \Big) (\boldsymbol{P}^\ell)_{kj} = \sum_j (\boldsymbol{P}^\ell)_{kj} = 1$$

Thus also $\boldsymbol{P}^{\ell+1}$ is a stochastic matrix, which completes the proof. []

3.3. Classification definitions based on accessibility of states

- defn: regular Markov chain

$$(\exists n \geq 0): \quad (\boldsymbol{P}^n)_{ij} > 0 \text{ for all } (i,j) \tag{29}$$

Note that if the above holds for some $n \geq 0$, it will hold for all $n' \geq n$ (see exercises). Thus, irrespective of the initial state i, after a finite number of iterations there is a non-zero probability in a regular Markov chain for the system to be in *any* state j.

- defn: existence of paths

$$i \to j: \quad \exists n \geq 0 \text{ such that } (\boldsymbol{P}^n)_{ij} > 0 \tag{30}$$

$$i \not\to j: \quad \nexists n \geq 0 \text{ such that } (\boldsymbol{P}^n)_{ij} > 0 \tag{31}$$

In the first case it is possible to get to state j after starting in state i. In the second case one can *never* get to j starting from i, irrespective of the number of steps executed.

- defn: communicating states

$$i \leftrightarrow j: \quad \exists n, m \geq 0 \text{ such that } (\boldsymbol{P}^n)_{ij} > 0 \text{ and } (\boldsymbol{P}^m)_{ji} > 0 \tag{32}$$

Given sufficient time, we can always get from i to j and also from j to i.

- defn: closed set

$$\text{any set } C \subseteq S \text{ of states such that } (\forall i \in C,\ j \notin C): \ i \not\to j \tag{33}$$

So no state inside C can ever reach any state outside C via transitions allowed by the Markov chain, irrespective of the number of iterations. Put differently, $(\boldsymbol{P}^n)_{ij} = 0$ for all $n \geq 0$ is $i \in C$ and $j \notin C$.

- defn: absorbing state

A state which constitutes a *closed set* with just one element. So if i is an absorbing state, one cannot leave this state ever via transitions of the Markov chain.
Note: if i is absorbing, then $p_{ij} = 0$ for all $j \neq i$. Since also $\sum_j p_{ij} = 1$, we conclude that $p_{ii} = 1$ and $p_{ij} = 0$ for all $j \neq i$:

$$i \in S \text{ is absorbing} \qquad \text{if and only if} \qquad p_{ij} = \delta_{ij} \tag{34}$$

- defn: irreducible set of states

This is any set $C \subseteq S$ of states such that:

$$\forall i, j \in C: \ i \leftrightarrow j \tag{35}$$

All states in an irreducible set are connected to each other, in that one can go from any state in C to any other state in C in a finite number of steps.

- defn: ergodic (or 'irreducible') Markov chain

  A Markov chain with the property that the complete set of states S is itself irreducible. Equivalently, one can go from any state in S to any other state in S in a finite number of steps.

  Ergodicity appears to be very similar to regularity (see above); let us clarify the relation between the two:

  (i) All *regular* Markov chains are also *ergodic*.

  proof:
  This follows from the definition of regular Markov chains: $(\exists n \geq 0): \ (\boldsymbol{P}^n)_{ij} > 0$ for all (i,j). It follows that one can indeed go from any state i to any state j in n steps of the dynamics. Hence the chain is ergodic. []

  (ii) The converse is not true: not all ergodic Markov chains are regular.

  proof:
  We only need to give one example of an ergodic chain that is not regular. The following will do:

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

  Clearly $\boldsymbol{P}$ is a stochastic matrix (although in the limit where there is at most randomness in the initial conditions, not in the dynamics), and $\boldsymbol{P}^2 = \mathbb{1}$. So one has $\boldsymbol{P}^n = \boldsymbol{P}$ for n odd and $\boldsymbol{P}^n = \mathbb{1}$ for n even. We can then solve the dynamics using (22) and write

  $$n \text{ even}: \quad p_1(n) = p_1(0), \quad p_2(n) = p_2(0)$$
  $$n \text{ odd}: \quad p_1(n) = p_2(0), \quad p_2(n) = p_1(0)$$

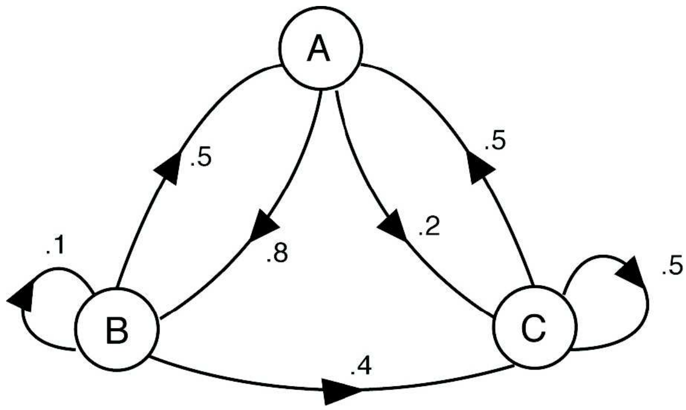
  However, as there is no n for which $\boldsymbol{P}^n$ has all entries nonzero, this chain is not regular. (see exercises for other examples). []

3.4. Graphical representation of Markov chains

Graphical representation: appropriate and possible only if $|S|$ is small!

nodes: all states $i \in S$ of the Markov chain
arrows: all allowed one-step transitions

arrow from i to j
if and only if $p_{ij} > 0$



Markov graph of transiton probabilites
between states A, B and C

translation of concepts in terms of network:

(i) $j \to i$: there is a path from j to i, following arrows
$j \not\to i$: there is no path from j to i, following arrows

(ii) communicating states $i \leftrightarrow j$:
there is path from j to i, and also a path from i to j, following arrows

(iii) ergodic Markov chain:
there is a path from *any* node j to *any* node i, following arrows

(iv) closed set: subset of nodes from which one cannot escape following arrows

(v) absorbing state: node with no outgoing arrows ('sink')

4. Convergence to a stationary state

4.1. Simple general facts on stationary states

- defn: stationary state of a Markov chain

  A stationary state of a Markov chain defined by the equation $p_i(n+1) = \sum_j p_{ji} p_j(n)$ is a vector $\boldsymbol{p} = (p_1, \ldots, p_{|S|})$ such that

$$\forall i \in S: \quad p_i = \sum_j p_{ji} p_j \qquad \text{and} \qquad \forall i \in S: \quad p_i \geq 0, \quad \sum_j p_j = 1 \tag{36}$$

  Thus $\boldsymbol{p}$ is a left eigenvector of the stochastic matrix, with eigenvalue $\lambda = 1$ and with non-negative entries, and represents a time-independent solution of the Markov chain.

- chains for which $\lim_{n\to\infty} \boldsymbol{P}^n$ exists

  If a transition matrix has the property that $\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{Q}$, then this has many useful consequences:

  (i) The matrix $\boldsymbol{Q} = \lim_{n\to\infty} \boldsymbol{P}^n$ is a stochastic matrix.

  proof:
  This follows trivially from the fact that $\boldsymbol{P}^n$ is a stochastic matrix for any $n \geq 0$. []

  (ii) The solution (22) of the Markov chain will also converge:

$$\lim_{n\to\infty} p_i(n) = \sum_j Q_{ji}\; p_j(0) \tag{37}$$

  proof: trivial []

  (iii) Each such limiting vector $\boldsymbol{p} = \lim_{n\to\infty} \boldsymbol{p}(n)$, which $\boldsymbol{p}(n) = (p_1(n), \ldots, p_{|S|}(n))$, is a stationary state of the Markov chain. In such a solution the probability to find any state will not change with time.

  proof:
  Since $\boldsymbol{Q}$ is a stochastic matrix, it follows from $p_i = \sum_j Q_{ji} p_j(0)$, in combination with the fact (established earlier) that stochastic matrices map normalized probabilities onto normalized probabilities, that the components of $\boldsymbol{p}$ obey $p_i \geq 0$ and $\sum_i p_i = 1$. What remains is to show that $\boldsymbol{p}$ is a left eigenvector of $\boldsymbol{P}$:

$$\begin{aligned}
\sum_j p_{ji} p_j &= \sum_{jk} p_{ji} Q_{kj} p_k(0) = \sum_k \Big(\sum_j Q_{kj} p_{ji}\Big) p_k(0) \\
&= \sum_k (\boldsymbol{QP})_{ki} p_k(0) = \lim_{n\to\infty} \sum_k (\boldsymbol{P}^{n+1})_{ki} p_k(0) \\
&= \lim_{n\to\infty} \sum_k (\boldsymbol{P}^n)_{ki} p_k(0) = \sum_k Q_{ki} p_k(0) = p_i
\end{aligned}$$

  []

(iv) The stationary solution to which the Markov chain evolves is *unique* (independent of the choice made for the $p_i(0)$) if and only if Q_{ij} is independent of i for all k, i.e. all rows of the matrix $\boldsymbol{Q}$ are identical.

proof:
Suppose we choose our initial state to be $k \in S$, so $p_i(0) = \delta_{ik}$. This would lead to the stationary solution $p_j = \sum_i Q_{ij} p_i(0) = Q_{kj}$ for all $j \in S$. It follows that the stationary probabilities are independent of the choice made for k if and only if Q_{kj} is independent of k for all j. []

- <u>convergence of time averages</u>

In many practical problems one is not necessarily interested in stationary states as defined above, but rather in the asymptotic averages over time of state probabilities, viz. in $\lim_{M\to\infty} M^{-1}\sum_{n\leq M} p_i(n) = \lim_{M\to\infty} M^{-1}\sum_{n\leq M}\sum_j (\boldsymbol{P}^n)_{ji} p_j(0)$. Hence one would like to know whether the following limit exists, and find it:

$$\overline{\boldsymbol{Q}} = \lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \boldsymbol{P}^n \tag{38}$$

We note: if $\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{Q}$, one will recover $\overline{\boldsymbol{Q}} = \boldsymbol{Q}$ (since the individual $\boldsymbol{P}^n$ are all bounded). In other words, if the first limit $\boldsymbol{Q}$ exists then also the second one $\overline{\boldsymbol{Q}}$ will exist, and the two will be identical. There will, however, be Markov chains for which $\boldsymbol{Q} = \lim_{n\to\infty} \boldsymbol{P}^n$ does not exist, yet $\overline{\boldsymbol{Q}}$ does.

To see what could happen, let us return to the earlier example of an ergodic chain that is not regular,

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

One has $\boldsymbol{P}^n = \boldsymbol{P}$ for n odd and $\boldsymbol{P}^n = \mathbf{1}$ for n even, so $\lim_{n\to\infty} \boldsymbol{P}^n$ does not exist. However, for this Markov chain the limit $\overline{\boldsymbol{Q}}$ *does* exist:

$$\begin{aligned}
\overline{\boldsymbol{Q}} &= \lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \boldsymbol{P}^n \\
&= \lim_{M\to\infty} \frac{1}{M}\left\{\sum_{n=1}^{\frac{1}{2}(M+1)} \boldsymbol{P}^{2n-1} + \sum_{n=0}^{\frac{1}{2}M} \boldsymbol{P}^{2n}\right\} \;=\; \lim_{M\to\infty} \frac{1}{M}\left\{\sum_{n=1}^{\frac{1}{2}(M+1)} \boldsymbol{P} + \sum_{n=0}^{\frac{1}{2}M} \mathbf{1}\right\} \\
&= \frac{1}{2}\boldsymbol{P} + \frac{1}{2}\mathbf{1}
\end{aligned}$$

4.2. Fundamental theorem for regular Markov chains

This theorem states:

> If $\boldsymbol{P}$ is the stochastic matrix of a regular Markov chain, then $\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{Q}$ exists, and all rows of the matrix $\boldsymbol{Q}$ are identical. Hence *the Markov chain will always converge to a unique stationary state, independent of the initial conditions.*

proof:
as a result of the importance of this property, one can find several different proofs in literature. Here follows one version which is rather intuitive, in that it focuses on how the differences between the stationary state probabilities and the actual time-dependent state probabilities become smaller and smaller as the process progresses.

- Existence of a stationary state

  First we prove that there exists a left eigenvector of $\boldsymbol{P} = \{p_{ij}\}$ with eigenvalue $\lambda = 1$, then we prove that there exists such an eigenvector with non-negative entries.

  (i) If $\boldsymbol{P}$ is a stochastic matrix, it has at least one left eigenvector with eigenvalue $\lambda = 1$.

  proof:
  left eigenvectors of $\boldsymbol{P}$ are right eigenvectors of $\boldsymbol{P}^\dagger$, where $(\boldsymbol{P}^\dagger)_{ij} = p_{ji}$. Since $\det(\boldsymbol{P} - \lambda\mathbf{1}) = \det(\boldsymbol{P} - \lambda\mathbf{1})^\dagger = \det(\boldsymbol{P}^\dagger - \lambda\mathbf{1})$, the left- and right eigenvalue polynomials of any matrix $\boldsymbol{P}$ are identical. Since $\boldsymbol{P}$ has a *right* eigenvector with eigenvalue 1 (namely $(1,\ldots,1)$, due to the normalization $\sum_j p_{ij} = 1$ for all $i \in S$), we know that there exists at least one *left* eigenvector ϕ with eigenvalue one: $\sum_i \phi_i p_{ij} = \phi_j$ for all $j \in S$. []

  (ii) If $\boldsymbol{P}$ is the stochastic matrix of a regular Markov chain, then each left eigenvector with eigenvalue $\lambda = 1$ has strictly positive components.

  Let $\boldsymbol{p}$ be a left eigenvector of $\boldsymbol{P}$. We define $S^+ = \{i \in S |\ p_i > 0\}$. Let $n > 0$ be such that $(\boldsymbol{P}^n)_{ij} > 0$ for all $i, j \in S$ (this n exists since the chain is regular). We write the left eigenvalue equation for $\boldsymbol{P}^n$, which is satisfied by $\boldsymbol{p}$, and we sum over all $j \in S^+$ (using $\sum_j (\boldsymbol{P}^n)_{ij} = 1$ for all i, due to $\boldsymbol{P}^n$ being also a stochastic matrix):

$$\sum_{j\in S^+} p_j = \sum_{j\in S^+}\sum_i (\boldsymbol{P}^n)_{ij} p_i = \sum_{j\in S^+}\Big[\sum_{i\in S^+}(\boldsymbol{P}^n)_{ij}p_i + \sum_{i\notin S^+}(\boldsymbol{P}^n)_{ij}p_i\Big]$$

$$\sum_{j\in S^+} p_j - \sum_{i,j\in S^+}(\boldsymbol{P}^n)_{ij}p_i = \sum_{j\in S^+}\sum_{i\notin S^+}(\boldsymbol{P}^n)_{ij}p_i$$

$$\sum_{i\in S^+} p_i\Big[1 - \sum_{j\in S^+}(\boldsymbol{P}^n)_{ij}\Big] = \sum_{j\in S^+}\sum_{i\notin S^+}(\boldsymbol{P}^n)_{ij}p_i$$

$$\sum_{i\in S^+} p_i\Big[\sum_{j\notin S^+}(\boldsymbol{P}^n)_{ij}\Big] = \sum_{i\notin S^+} p_i\Big[\sum_{j\in S^+}(\boldsymbol{P}^n)_{ij}\Big]$$

The left-hand side is a sum of non-negative terms and the right-hand side is a sum of non-positive terms; hence *all terms on both sides must be zero*:

$$\text{LHS}: \quad (\forall i \in S^+): \; p_i = 0 \;\; \text{or} \;\; \sum_{j \notin S^+} (\boldsymbol{P}^n)_{ij} = 0$$

$$\text{RHS}: \quad (\forall i \notin S^+): \; p_i = 0 \;\; \text{or} \;\; \sum_{j \in S^+} (\boldsymbol{P}^n)_{ij} = 0$$

Since we also know that $(\boldsymbol{P}^n)_{ij} > 0$ for all $i, j \in S$, and that $p_i > 0$ for all $i \in S^+$:

$$\text{LHS}: \quad S^+ = \emptyset \;\; \text{or} \;\; S^+ = S$$

$$\text{RHS}: \quad (\forall i \notin S^+): \; p_i = 0 \;\; \text{or} \;\; S^+ = \emptyset$$

This leaves two possibilities: either $S^+ = S$ (i.e. all components p_i positive), or $S^+ = \emptyset$ (i.e. all components p_i negative or zero). In the former case we have proved our claim already; in the latter case we can construct a new eigenvector via $p_i \to -p_i$ for all $i \in S$, which will then have non-negative components only. What remains is to show that none of these can be zero. If a p_i were to be zero then the eigenvalue equation would give $\sum_j (\boldsymbol{P}^n)_{ji} p_j = 0$, from which it would follow (regular chain!) that $\sum_j p_j = 0$; thus *all* components must be zero since $p_j \geq 0$ for all j. This is impossible since $\boldsymbol{p}$ is an eigenvector. This completes the proof. []

- Convergence to the stationary state

Having established the existence of a stationary state, the second part of the proof of the fundamental theorem consists in showing that the Markov chain must have a stationary state as a limit, whatever the initial conditions $p_i(0)$, and that there can only be one such stationary state.

(i) If $\boldsymbol{P}$ is the stochastic matrix of a regular Markov chain, and $\boldsymbol{p} = (p_1, \ldots, p_{|S|})$ is a stationary state of the chain, then $\lim_{n\to\infty} p_k(n) = p_k$ for all initial conditions.

proof:
Let m be such that $(\boldsymbol{P}^n)_{ij} > 0$ for all $n \geq m$ and all $i, j \in S$. Define $z = \min_{ij} (\boldsymbol{P}^m)_{ij}$, so $z > 0$, and choose $n \geq m$. Define the sets $S^+(n) = \{k \in S |\; p_k(n) > p_k\}$, $S^-(n) = \{k \in S |\; p_k(n) < p_k\}$, and $S^0(n) = \{k \in S |\; p_k(n) = p_k\}$, as well as the sums $U^\pm(n) = \sum_{k \in S^\pm(n)} [p_k(n) - p_k]$. By construction, $U^+(n) \geq 0$ and $U^-(n) \leq 0$ for all n. We inspect how the $U^\pm(n)$ evolve as n increases:

$$U^\pm(n+1) - U^\pm(n) = \sum_{k \in S^\pm(n+1)} [p_k(n+1) - p_k] - \sum_{k \in S^\pm(n)} [p_k(n) - p_k]$$

$$= \sum_{k \in S^\pm(n+1)} \sum_{\ell} [p_\ell(n) - p_\ell] p_{\ell k} - \sum_{\ell \in S^\pm(n)} [p_k(n) - p_k]$$

$$= \sum_{\ell \in S^\pm(n)} [p_\ell(n) - p_\ell] \Big[\sum_{k \in S^\pm(n+1)} p_{\ell k} - 1 \Big] + \sum_{\ell \notin S^\pm(n)} [p_\ell(n) - p_\ell] \Big[\sum_{k \in S^\pm(n+1)} p_{\ell k} \Big]$$

$$= -\sum_{\ell\in S^{\pm}(n)} [p_\ell(n)-p_\ell]\Big[\sum_{k\notin S^{\pm}(n+1)} p_{\ell k}\Big] + \sum_{\ell\notin S^{\pm}(n)} [p_\ell(n)-p_\ell]\Big[\sum_{k\in S^{\pm}(n+1)} p_{\ell k}\Big]$$

So we find

$$\begin{aligned}
U^+(n{+}1)-U^+(n) = & -\sum_{\ell\in S^+(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\notin S^+(n+1)} p_{\ell k}\Big] \\
& + \sum_{\ell\notin S^+(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\in S^+(n+1)} p_{\ell k}\Big] \leq 0 \\
U^-(n{+}1)-U^-(n) = & -\sum_{\ell\in S^-(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\notin S^-(n+1)} p_{\ell k}\Big] \\
& + \sum_{\ell\notin S^-(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\in S^-(n+1)} p_{\ell k}\Big] \geq 0
\end{aligned}$$

Thus $U^+(n)$ is a non-increasing function of n, and $U^-(n)$ a non-decreasing. Next we inspect what happens at time intervals of m steps. This means repeating the above steps with $S^{\pm}(n+1)$ replaced by $S^{\pm}(n+m)$, and $p_{\ell k}$ replaced by $(\boldsymbol{P}^m)_{\ell k}$:

$$\begin{aligned}
U^+(n{+}m)-U^+(n) = & -\sum_{\ell\in S^+(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\notin S^+(n+m)} (\boldsymbol{P}^m)_{\ell k}\Big] \\
& + \sum_{\ell\notin S^+(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\in S^+(n+m)} (\boldsymbol{P}^m)_{\ell k}\Big] \\
\leq & -z\sum_{\ell\in S^+(n)} |p_\ell(n)-p_\ell|\Big[|S|-|S^+(n+m)|\Big] \\
& -z\sum_{\ell\notin S^+(n)} |p_\ell(n)-p_\ell||S^+(n+m)| \\
U^-(n{+}m)-U^-(n) = & -\sum_{\ell\in S^-(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\notin S^-(n+m)} (\boldsymbol{P}^m)_{\ell k}\Big] \\
& + \sum_{\ell\notin S^-(n)} \Big[p_\ell(n)-p_\ell\Big]\Big[\sum_{k\in S^-(n+m)} (\boldsymbol{P}^m)_{\ell k}\Big] \\
\geq & \, z\sum_{\ell\in S^-(n)} |p_\ell(n)-p_\ell|\Big[|S|-|S^-(n+m)|\Big] \\
& +z\sum_{\ell\notin S^-(n)} |p_\ell(n)-p_\ell||S^-(n+m)|
\end{aligned}$$

Since $U^+(n)$ is bounded from below, it is a Lyapunov function for the dynamics, and must tend to a limit $\lim_{n\to\infty} U^+(n) \geq 0$. Similarly, $U^-(n)$ is nondecreasing and bounded from above, and must tend to a limit $\lim_{n\to\infty} U^-(n) \leq 0$ If these limits have been reached we must have equality in the above inequalities, giving

$$\sum_{\ell\in S^+(n)} |p_\ell(n)-p_\ell|\Big[|S|-|S^+(n+m)|\Big] = \sum_{\ell\notin S^+(n)} |p_\ell(n)-p_\ell||S^+(n+m)| = 0$$

$$\sum_{\ell\in S^-(n)} |p_\ell(n)-p_\ell|\Big[|S|-|S^-(n+m)|\Big] = \sum_{\ell\notin S^-(n)} |p_\ell(n)-p_\ell||S^-(n+m)| = 0$$

or

$$\left\{S^+(n)=\emptyset \;\text{ or }\; S^+(n+m)=S\right\} \text{ and } \left\{S^-(n)=\emptyset \;\text{ or }\; S^+(n+m)=\emptyset\right\}$$
$$\left\{S^-(n)=\emptyset \;\text{ or }\; S^-(n+m)=S\right\} \text{ and } \left\{S^+(n)=\emptyset \;\text{ or }\; S^-(n+m)=\emptyset\right\}$$

Using the incompatibility of the combination $S^+(n+m)=S$ and $S^+(n+m)=\emptyset$ (with the same for $S^-(n+m)$), we see that always at least one of the two sets $S^{\pm}(n)$ must be empty. Finally we prove that in fact both must be empty. For this we use the identity $0=\sum_k[p_k(n)-p_k]=\sum_{k\in S^+(n)}|p_k(n)-p_k|-\sum_{k\in S^-(n)}|p_k(n)-p_k|$, which can never be satisfied if $S^+(n)=\emptyset$ and $S^-(n)\neq\emptyset$ or vice versa. Hence we are left with $S^+(n)=S^-(n)=\emptyset$, so $p_k(n)=p_k=0$ for all k. []

(ii) A regular Markov chain has exactly one stationary state.

proof:
We already know there exists at least one stationary state. Let us now assume there are two distinct stationary states $\boldsymbol{p}=(p_1,\ldots,p_{|S|})$ and $\boldsymbol{q}=(q_1,\ldots,q_{|S|})$. We choose $\boldsymbol{p}$ as the stationary state in the sense of the previous result, and take $\boldsymbol{q}$ as the initial state (so $p_i(0)=q_i$ for all i). The previous result then states:

$$\text{for all } k: \qquad \lim_{n\to\infty}\sum_i q_i(\boldsymbol{P}^n)_{ik}=p_k \quad \text{giving} \quad q_k=p_k$$

A contradiction with the assumption $\boldsymbol{p}\neq\boldsymbol{q}$, which completes the proof. []

- If $\boldsymbol{P}$ is the stochastic matrix of a regular Markov chain, then $\lim_{n\to\infty}\boldsymbol{P}^n=\boldsymbol{Q}$ exists, and all rows of the matrix $\boldsymbol{Q}$ are identical.

proof:
We just combine the previous results. We have already shown that there is one unique stationary state $\boldsymbol{p}=(p_1,\ldots,p_{|S|})$, and that $\lim_{n\to\infty}\sum_i p_i(0)(\boldsymbol{P}^n)_{ij}=p_j$ for all $j\in S$ and all initial conditions $\{p_i(0)\}$. We now choose as our initial conditions $p_i(0)=\delta_{ik}$, for which we then find

$$\lim_{n\to\infty}(\boldsymbol{P}^n)_{kj}=p_j$$

Since this is true for all choices of $k\in S$, we have shown

$$\lim_{n\to\infty}\boldsymbol{P}^n=\boldsymbol{Q}, \qquad Q_{kj}=p_j \;\text{ for all k}$$

Thus the limit exists, and all rows of $\boldsymbol{Q}$ are identical. []

4.3. Examples

Let us inspect some new and earlier examples and apply what we have learned in the previous pages. The first example in fact covers all Markov chains with two states.

- Consider two-state Markov chains with state space $S = \{1, 2\}$. Let $p_i(n)$ denote the probability to find the process in state i at time n, where $i \in \{1, 2\}$ and $n \in \mathbb{N}$. Let $\{p_{ij}\}$ represent the transition matrix of the process.

  What is the most general stochastic matrix for this set-up? We need a real-valued matrix with entries $p_{ij} \in [0, 1]$ (condition 1), such that $\sum_j p_{ij} = 1$ for all i (condition 2).
  First row: $p_{11} + p_{12} = 1$ so $p_{12} = 1 - p_{11}$ with $p_{11} \in [0, 1]$
  Second row: $p_{21} + p_{22} = 1$ so $p_{22} = 1 - p_{21}$ with $p_{21} \in [0, 1]$
  Call $p_{11} = \alpha \in [0, 1]$ and $p_{21} = \beta \in [0, 1]$ and we have

  $$\boldsymbol{P} = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix} \qquad \alpha, \beta \in [0, 1]$$

  For what values of (α, β) does this process have absorbing states? A state i is absorbing iff $p_{ij} = \delta_{ij}$ for $j \in \{1, 2\}$. Let us check the two candidates:
  state 1 absorbing iff $\alpha = 1$ (giving $p_{11} = 1,\ p_{12} = 0$)
  state 2 absorbing iff $\beta = 0$ (giving $p_{21} = 0,\ p_{22} = 1$)
  So our process has absorbing states only for $\alpha = 1$ and for $\beta = 0$.

  As soon as $\alpha \neq 1$ and $\beta \neq 0$ we see in a simple diagram that our process is ergodic. If in addition $\alpha > 0$ and $\beta < 1$, we have $p_{ij} > 0$ for all (i, j), so it is even regular. Next let us calculate the stationary solution(s). These are left-eigenvectors of $\boldsymbol{P}$ with eigenvalue $\lambda = 1$, that have non-negative components only and $p_1 + p_2 = 1$. So we have to solve

  $$p_1 = p_1 p_{11} + p_2 p_{21} : \quad (1 - \alpha)p_1 = \beta(1 - p_1)$$
  $$p_2 = p_1 p_{12} + p_2 p_{22} : \quad (1 - \alpha)p_1 = \beta(1 - p_1)$$

  Result:

  $$p_1 = \frac{\beta}{1 - \alpha + \beta}, \qquad p_2 = \frac{1 - \alpha}{1 - \alpha + \beta},$$

  Let us next calculate the general solution $p_i(n) = \sum_j p_j(0)(\boldsymbol{P}^n)_{ji}$ of the Markov chain, for the case $\beta = 1 - \alpha$. This requires finding expressions for $\boldsymbol{P}^n$. Note that now

  $$\boldsymbol{P} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix} \qquad \alpha \in [0, 1]$$

  $\boldsymbol{P}$ is symmetric, so we must have two orthogonal eigenvectors, and left- and right eigenvectors are identical. Inspect the eigenvalue polynomial:

  $$\det \begin{pmatrix} \alpha - \lambda & 1-\alpha \\ 1 - \alpha & \alpha - \lambda \end{pmatrix} = 0 \quad \Rightarrow \quad (\lambda - \alpha)^2 - (1 - \alpha)^2 = 0 \quad \Rightarrow \quad \lambda = \alpha \pm (1 - \alpha)$$

There are two eigenvalues: $\lambda = 1$ and $\lambda = 2\alpha - 1$. The corresponding eigenvectors are

$$\lambda_1 = 1: \quad \begin{cases} \alpha x_1 + (1-\alpha)x_2 = x_1 \\ (1-\alpha)x_1 + \alpha x_2 = x_2 \end{cases} \quad \Rightarrow \quad x_1 = x_2 \quad \Rightarrow \quad (x_1, x_2) = (1, 1)$$

$$\lambda_2 = 2\alpha - 1: \quad \begin{cases} \alpha x_1 + (1-\alpha)x_2 = (2\alpha-1)x_1 \\ (1-\alpha)x_1 + \alpha x_2 = (2\alpha-1)x_2 \end{cases} \quad \Rightarrow \quad x_1 = -x_2 \quad \Rightarrow \quad (x_1, x_2) = (1, -1)$$

Normalize the two eigenvectors: $\hat{e}^{(1)} = (1,1)/\sqrt{2}$, $\hat{e}^{(2)} = (1,-1)/\sqrt{2}$.
We can now expand $\boldsymbol{P}$ in eigenvectors: $p_{ij} = \sum_{\ell=1}^{2} \lambda_\ell e_i^{(\ell)} e_j^{(\ell)}$. More generally, since eigenvectors of $\boldsymbol{P}$ with eigenvalue λ are also eigenvectors of $\boldsymbol{P}^n$ with eigenvalue λ^n:

$$\begin{aligned} (\boldsymbol{P}^n)_{ij} &= \sum_{\ell=1}^{2} \lambda_\ell^n e_i^{(\ell)} e_j^{(\ell)} = e_i^{(1)} e_j^{(1)} + (2\alpha-1)^n e_i^{(2)} e_j^{(2)} \\ &= \frac{1}{2} + (2\alpha-1)^n e_i^{(2)} e_j^{(2)} \end{aligned}$$

Here this gives

$$\begin{aligned} p_i(n) &= \sum_j p_j(0) \Big(\frac{1}{2} + (2\alpha-1)^n e_i^{(2)} e_j^{(2)} \Big) \\ &= \frac{1}{2}[p_1(0) + p_2(0)] + (2\alpha-1)^n \frac{e_i^{(2)}}{\sqrt{2}} [p_1(0) - p_2(0)] \end{aligned}$$

or

$$\begin{aligned} p_1(n) &= \frac{1}{2} + \frac{1}{2}(2\alpha-1)^n [p_1(0) - p_2(0)] \\ p_2(n) &= \frac{1}{2} - \frac{1}{2}(2\alpha-1)^n [p_1(0) - p_2(0)] \end{aligned}$$

Let us inspect what happens for large n. The limits $\lim_{n\to\infty} p_i(n)$ exists if and only if $|2\alpha - 1| < 1$, i.e. $0 < \alpha < 1$. If the latter condition holds, then $\lim_{n\to\infty} p_1(n) = \lim_{n\to\infty} p_2(n) = 1/2$.

So-called *periodic states* i are defined by the following property: starting from state i, $p_i(n) > 0$ only if n is a multiple of some integer $\lambda_i \geq 2$. For $\alpha \in (0,1)$ this clearly doesn't happen (stationary state); this leaves $\alpha \in \{0, 1\}$.
$\alpha = 1$: $p_1(n) = \frac{1}{2} + \frac{1}{2}[p_1(0) - p_2(0)]$, $p_2(n) = \frac{1}{2} - \frac{1}{2}[p_1(0) - p_2(0)]$
both are independent of time, so never periodic states.
$\alpha = 0$: $p_1(n) = \frac{1}{2} + \frac{1}{2}(-1)^n[p_1(0) - p_2(0)]$, $p_2(n) = \frac{1}{2} - \frac{1}{2}(-1)^n[p_1(0) - p_2(0)]$
both are oscillating. To get zero values repeatedly one needs $p_1(0) - p_2(0) = \pm 1$, which happens for $p_1(0) \in \{0, 1\}$.
for $p_1(0) = 0$ we get $p_1(n) = \frac{1}{2} - \frac{1}{2}(-1)^n$, $p_2(n) = \frac{1}{2} + \frac{1}{2}(-1)^n$
for $p_1(0) = 1$ we get $p_1(n) = \frac{1}{2} + \frac{1}{2}(-1)^n$, $p_2(n) = \frac{1}{2} - \frac{1}{2}(-1)^n$
conclusion: system has periodic states for $\alpha = 0$.

- X_n = number of sixes thrown after n dice rolls?

$$W_{XY} = \frac{1}{6}\delta_{X,Y+1} + \frac{5}{6}\delta_{X,Y}$$

Translated into our standard conventions, we have $S = \{0, 1, 2, 3, \ldots\}$, $(X, Y) \to (j, i) \in S^2$, and hence

$$p_{ij} = \frac{1}{6}\delta_{i,j-1} + \frac{5}{6}\delta_{ij}$$

This is not a finite state space process, the set S is not bounded. Although we can still do several things, the proofs of the previous subsection cannot be used. One easily convinces oneself that $(\boldsymbol{P}^n)_{ij}$ will be of the form

$$(\boldsymbol{P}^n)_{ij} = \sum_{\ell=0}^{n} a_{n,\ell}\delta_{i,j-\ell}$$

with non-negative coefficients $a_{n,\ell}$ that must obey $\sum_{\ell=0}^{n} a_{n,\ell} = 1$ for any $n \geq 0$. One then finds a simple iteration for these coefficient by inspecting $(\boldsymbol{P}^{n+1})$:

$$\begin{aligned}
(\boldsymbol{P}^{n+1})_{ij} &= \sum_k (\boldsymbol{P}^n)_{ik} p_{kj} = \sum_{\ell=0}^{n}\sum_k a_{n,\ell}\delta_{i,k-\ell}\Big[\frac{1}{6}\delta_{k,j-1} + \frac{5}{6}\delta_{kj}\Big] \\
&= \frac{1}{6}\sum_{\ell=0}^{n} a_{n,\ell}\delta_{i,j-1-\ell} + \frac{5}{6}\sum_{\ell=0}^{n} a_{n,\ell}\delta_{i,j-\ell} \\
&= \frac{1}{6}\sum_{\ell=1}^{n+1} a_{n,\ell-1}\delta_{i,j-\ell} + \frac{5}{6}\sum_{\ell=0}^{n} a_{n,\ell}\delta_{i,j-\ell} \\
&= \frac{5}{6}a_{n,0}\delta_{ij} + \sum_{\ell=1}^{n}\Big[\frac{1}{6}a_{n,\ell-1} + \frac{5}{6}a_{n,\ell}\Big]\delta_{i,j-\ell} + \frac{1}{6}a_{n,n}\delta_{i,j-n-1}
\end{aligned}$$

Hence:

$$a_{n+1,0} = \frac{5}{6}a_{n,0}, \qquad a_{n+1,n+1} = \frac{1}{6}a_{n,n}, \qquad 0<\ell<n+1: \quad a_{n+1,\ell} = \frac{1}{6}a_{n,\ell-1} + \frac{5}{6}a_{n,\ell}$$

The first two are calculated directly, starting from $a_{1,0} = \frac{5}{6}$ and $a_{1,1} = \frac{1}{6}$, giving $a_{n,0} = (\frac{5}{6})^n$ and $a_{n,n} = (\frac{1}{6})^n$. The others are obtained by iteration. It is clear from these expressions that there cannot be a stationary state. The long time average transition probabilities, if they exist, would be $\overline{Q}_{ij} = \phi(j-i)$, where

$$\phi(k < 1) = 0, \qquad \phi(k \geq 1) = \lim_{m\to\infty}\frac{1}{m}\sum_{n=k}^{m} a_{n,k}$$

Using $a_{n,k} \in [0, 1]$ and $\sum_j \overline{Q}_{ij} = 1$ one can show quite easily that $\phi(k) = 0$ for all k. So although the overall probabilities to find *any* state continue to add up to one, the continuing growth of the state space (this process is diffusive in nature) is such that the individual probabilities $\overline{Q}_{ij}$ vanish asymptotically.

- X_n = largest number thrown after n dice rolls?

$$W_{XY} = \begin{cases} 0 & \text{if} \quad Y > X \\ Y/6 & \text{if} \quad Y = X \\ 1/6 & \text{if} \quad Y < X \end{cases}$$

Translated into our standard conventions, we have $S = \{1, \ldots, 6\}$, $(X, Y) \to (j, i) \in S^2$, and hence

$$p_{ij} = \begin{cases} 0 & \text{if} \quad i > j \\ i/6 & \text{if} \quad i = j \\ 1/6 & \text{if} \quad i < j \end{cases}$$

We immediately notice that the state $i = 6$ is absorbing, since $p_{66} = 1$. Once we are at $i = 6$ we can never escape. Thus the chain cannot be regular, and the absorbing state is a stationary state. We could calculate $\boldsymbol{P}^n$ but that would be messy and hard work. We suspect that the system will always end up in the absorbing state $i = 6$, so let us inspect what happens to $(\boldsymbol{P}^n)_{i6}$:

$$\begin{aligned} (\boldsymbol{P}^{n+1})_{i6} - (\boldsymbol{P}^n)_{i6} &= \sum_k (\boldsymbol{P}^n)_{ik} p_{k6} = (\boldsymbol{P}^n)_{i6} + \frac{1}{6} \sum_{k<6} (\boldsymbol{P}^n)_{ik} - (\boldsymbol{P}^n)_{i6} \\ &= (\boldsymbol{P}^n)_{i6} + \frac{1}{6} \Big[1 - (\boldsymbol{P}^n)_{i6}\Big] - (\boldsymbol{P}^n)_{i6} \\ &= \frac{1}{6} \Big[1 - (\boldsymbol{P}^n)_{i6}\Big] \geq 0 \end{aligned}$$

We conclude from this that each $(\boldsymbol{P}^n)_{i6}$ will increase monotonically with n, and converge to the value one. Thus, the system will always ultimately end up in the absorbing state $i = 6$, irrespective of the initial conditions.

- The drunk on the pavement (see section 1):

$$W_{XY} = \begin{cases} \delta_{X,0} & \text{if} \quad Y = 0 \\ \frac{1}{2}\delta_{X,Y+1} + \frac{1}{2}\delta_{X,Y-1} & \text{if} \quad 0 < Y < 5 \\ \delta_{X,4} & \text{if} \quad Y = 5 \end{cases}$$

Here $S = \{0, \ldots, 5\}$, and

$$p_{ij} = \begin{cases} \delta_{0,j} & \text{if} \quad i = 0 \\ \frac{1}{2}\delta_{i,j-1} + \frac{1}{2}\delta_{i,j+1} & \text{if} \quad 0 < i < 5 \\ \delta_{4,j} & \text{if} \quad i = 5 \end{cases}$$

Again we have an absorbing state: $p_{00} = 1$. Once we are at $i = 0$ (i.e. on the street) we stay there. The chain cannot be regular, and the absorbing state is a stationary state. Let us inspect what happens to the likelihood of being in the absorbing state:

$$(\boldsymbol{P}^{n+1})_{i0} - (\boldsymbol{P}^n)_{i0} = \sum_k (\boldsymbol{P}^n)_{ik} p_{k0} = (\boldsymbol{P}^n)_{i0} + \frac{1}{2} \sum_{k=1}^{4} (\boldsymbol{P}^n)_{ik} \delta_{k,1} - (\boldsymbol{P}^n)_{i0} = \frac{1}{2} (\boldsymbol{P}^n)_{i1} \geq 0$$

A bit of further work shows that also this system will always end in the absorbing state.

5. Some further developments of the theory

5.1. Definition and physical meaning of detailed balance

- defn: probability current in Markov chains

  The net probability current $J_{i\to j}$ from any state $i \in S$ to any state $j \in S$, in the stationary state of a Markov chain characterized by the stochastic matrix $\boldsymbol{P} = \{p_{ij}\}$ and state space S, is defined as

$$J_{i\to j} = p_i p_{ij} - p_j p_{ji} \tag{39}$$

  where $\boldsymbol{p} = \{p_i\}$ defines the stationary state of the chain.

  consequences, conventions

  (i) The current is by definition anti-symmetric under permutation of i and j:

$$J_{j\to i} = p_j p_{ji} - p_i p_{ij} = -\Big[p_i p_{ij} - p_j p_{ji}\Big] = -J_{i\to j}$$

  (ii) Imagine that the chain represents a random walk of a particle, in a stationary state. Since p_i is the probability to find the particle in state i, and p_{ij} the likelihood that it subsequently moves from i to j, $p_i p_{ij}$ is the probability that we *observe the particle moving from i to j*. With multiple particles it would be proportional to the *number* of observed moves from i to j. Thus $J_{i\to j}$ represents the net balance of observed transitions between i and j in the stationary state; hence the term 'current'. If $J_{i\to j} > 0$ there are more transitions $i \to j$ than $j \to i$; if $J_{i\to j} < 0$ there are more transitions $j \to i$ than $i \to j$.

  (iii) Conservation of probability implies that the sum over all currents is always zero:

$$\begin{aligned}\sum_{ij} J_{i\to j} &= \sum_{ij}\Big[p_i p_{ij} - p_j p_{ji}\Big] = \sum_i p_i\Big(\sum_j p_{ij}\Big) - \sum_j p_j\Big(\sum_i p_{ji}\Big)\\ &= \sum_i p_i - \sum_j p_j = 1 - 1 = 0\end{aligned}$$

- defn: detailed balance

  A regular Markov chain characterized by the stochastic matrix $\boldsymbol{P} = \{p_{ij}\}$ and state space S is said to obey detailed balance if

$$p_j p_{ji} = p_i p_{ij} \qquad \text{for all} \ \ i, j \in S \tag{40}$$

  where $\boldsymbol{p} = \{p_i\}$ defines the stationary state of the chain. Markov chains with detailed balance are sometimes called 'reversible'.

  consequences, conventions

(i) We see that detailed balance represents the special case where *all* individual currents in the stationary state are zero: $J_{i\to j} = p_i p_{ij} - p_j p_{ji} = 0$ for all $i, j \in S$.

(ii) Detailed balance is a stronger condition than stationarity. All Markov chains with detailed balance have by definition a unique stationary state, but not all regular Markov chains with stationary states obey detailed balance.

(iii) From (40) one can derive stationarity by summing over j:
$$\forall i \in S: \quad \sum_j p_j p_{ji} = \sum_j p_i p_{ij} = p_i \sum_j p_{ij} = p_i$$

(iv) Markov chains used to model closed physical many-particle systems with noise are usually of the detailed balance type, as a result of the invariance of Newton's laws of motion under time reversal $t \to -t$.

So far we studied situations where we are given a Markov chain and want to know its properties. However, sometimes we are faced with the inverse problem: given a state space S and stationary probabilities $\boldsymbol{p} = (p_1, \ldots, p_{|S|})$, find a simple stochastic matrix $\boldsymbol{P}$ for which the associated Markov chain will give $\lim_{n\to\infty} \boldsymbol{p}(n) = \boldsymbol{p}$. Simple: one that is easy to implement in computer programs. The detailed balance condition allows for this to be achieved in a simple way, leading to so-called

5.2. Definition of characteristic times

- defn: statistics of first passage times (FPT)

$$f_{ij}(n) = \text{probability that, starting from } i, \text{ first visit to } j \text{ occurs at time } n \tag{41}$$

By definition, $f_{ij}(n) \in [0,1]$.
Can be expressed in terms of path probabilities (20):

$$f_{ij}(n) = \overbrace{\sum_{i_1,\ldots,i_{n-1}}}^{\text{sum over all paths}} \overbrace{\Big(\prod_{m=1}^{n-1}(1-\delta_{i_m,i})\Big)}^{\text{not visiting } i \text{ earlier}} \overbrace{p_{i,i_1}\Big(\prod_{m=2}^{n-1} p_{i_{m-1},i_m}\Big)p_{i_m,j}}^{\text{path probability}} \tag{42}$$

Consequences, conventions:

(i) Probability of *ever* arriving at j upon starting from i:
$$f_{ij} = \sum_{n>0} f_{ij}(n) \tag{43}$$

(ii) Probability of *never* arriving at j upon starting from i:

$$1 - f_{ij} = 1 - \sum_{n>0} f_{ij}(n) \tag{44}$$

(iii) mean recurrence time μ_i:

$$\mu_i = \sum_{n>0} n f_{ii}(n) \tag{45}$$

note: $\mu_i \geq f_{ii}$

(iv) recurrent state i: $f_{ii} = 1$ (system will *always* return to i at some point in time)
null recurrent state: $\mu_i = \infty$ (returns to i, but average time taken diverges)
positive recurrent state: $\mu_i < \infty$ (returns to i, average time taken finite)

(v) transient state i: $f_{ii} < 1$ (nonzero chance that system will never return to i at any point in time)

(vi) periodic/aperiodic states:

Consider the set $N_i = \{n \in \mathbb{N}^+ |\ f_{ii}(n) > 0\}$ (all times at which it is possible for the system to be in state i, given it started in state i). Define $\lambda_i \in \mathbb{N}^+$ as the largest common divisor of the elements in N_i, i.e. λ_i is the smallest integer such that all elements in N_i can be written as integer multiples of λ_i.
A state i is periodic if $\lambda_i > 1$, aperiodic if $\lambda_i = 1$.

Markov chains with periodic states are apparently systems subject to persistent oscillation, whereby certain states can be observed only on specific times that occur periodically, and not on intermediate ones.

(vii) relation between $f_{ij}(n)$ and the probabilities $p_{ij}(n)$, for $n > 0$:

$$p_{ij}(n) = \sum_{r=1}^{n} p_{jj}(n-r) f_{ij}(r) \tag{46}$$

proof:
consider the probability $p_{ij}(n)$ to go in n steps from state i to j. There are n possibilities for when the *first* occurrence of state j occurs after time zero: after $r = 1$ iterations, after $r = 2$ iterations, etc, until we have i showing up first after n iterations. If the first occurrence is after more than n steps, then there is no contribution to $p_{ij}(n)$. Hence

$$\begin{aligned} p_{ij}(n) &= \sum_{r=1}^{n} f_{ij}(r).\text{Prob}[X_n = j | X_r = j] \\ &= \sum_{r=1}^{n} p_{jj}(n-r) f_{ij}(r) \qquad [] \end{aligned}$$

note: not easy to use this eqn to calculate the $f_{ij}(n)$ in practice.

5.3. Calculation of characteristic times

- defn: generating functions for state transitions and first passage times

  Let $p_{ij}(n) = (\boldsymbol{P}^n)_{ij}$ (itself a stochastic matrix),
  for $s \in \mathbb{R}$, $|s| < 1$:
$$P_{ij}(s) = \sum_{n\geq 0} s^n p_{ij}(n), \qquad F_{ij}(s) = \sum_{n\geq 0} s^n f_{ij}(n) \tag{47}$$
  with the definition extension $f_{ij}(0) = 0$.
  Clearly $|P_{ij}(s)|, |F_{ij}(s)| \leq \sum_{n\geq 0} s^n = 1/(1-s)$
  (uniformly convergent since $|s| < 1$)

  Consequences, conventions

  (i) simple algebraic relation between generating functions, for arbitrary (i,j):
$$P_{ij}(s) - \delta_{ij} = P_{jj}(s)F_{ij}(s) \tag{48}$$
  proof:
  Multiply (46) by s^n and sum over $n \geq 0$. Use $p_{ij}(0) = (\boldsymbol{P}^0)_{ij} = \delta_{ij}$:
$$\begin{aligned}
P_{ij}(s) &= \delta_{ij} + \sum_{n>0} s^n \sum_{r=1}^{n} p_{jj}(n-r) f_{ij}(r) \\
&= \delta_{ij} + \sum_{n>0}\sum_{r=1}^{n} \Big(s^{n-r}p_{jj}(n-r)\Big)\Big(s^r f_{ij}(r)\Big) \quad \text{define } m = n-r \\
&= \delta_{ij} + \sum_{m\geq 0}\sum_{n>0}\sum_{r=1}^{n} \delta_{r+m,n}\Big(s^m p_{jj}(m)\Big)\Big(s^r f_{ij}(r)\Big) \\
&= \delta_{ij} + \sum_{m\geq 0}\sum_{n>0}\sum_{r>0} \delta_{r+m,n}\Big(s^m p_{jj}(m)\Big)\Big(s^r f_{ij}(r)\Big) \\
&= \delta_{ij} + \sum_{m\geq 0}\sum_{r>0} \Big(s^m p_{jj}(m)\Big)\Big(s^r f_{ij}(r)\Big) \\
&= \delta_{ij} + P_{jj}(s)F_{ij}(s)
\end{aligned}$$
  []

  (ii) corollary:
$$i \neq j: \quad P_{ij}(s) = P_{jj}(s)F_{ij}(s), \qquad \text{so } F_{ij}(s) = P_{ij}(s)/P_{jj}(s) \tag{49}$$
$$i = j: \quad P_{ii}(s) = 1/[1-F_{ii}(s)], \qquad \text{so } F_{ii}(s) = 1 - 1/P_{ii}(s) \tag{50}$$

- Possible ways to use $F_{ij}(s)$: differentiation
$$\lim_{s\uparrow 1} \frac{d}{ds}F_{ij}(s) = \lim_{s\uparrow 1} \sum_{n>0} ns^{n-1} f_{ij}(n) = \sum_{n>0} nf_{ij}(n) \tag{51}$$
$$\lim_{s\uparrow 1} \frac{d^2}{ds^2}F_{ij}(s) = \lim_{s\uparrow 1} \sum_{n>1} n(n-1)s^{n-2} f_{ij}(n) = \sum_{n>0} n(n-1)f_{ij}(n) \tag{52}$$

average n_{ij} and variance Δ_{ij} of first passage times for $i \to j$:

$$n_{ij} = \sum_{n>0} n f_{ij}(n) = \lim_{s\uparrow 1} \frac{d}{ds} F_{ij}(s) \tag{53}$$

$$\Delta_{ij}^2 = \sum_{n>0} n^2 f_{ij}(n) - \Big[\sum_{n>0} n f_{ij}(n)\Big]^2 = \lim_{s\uparrow 1} \frac{d^2}{ds^2} F_{ij}(s) + n_{ij}[1 - n_{ij}] \tag{54}$$

Appendix A. Exercises

(i) Show that the 'trained mouse' example in section 1 defines a discrete time and discrete state space homogeneous Markov chain. Calculate the stochastic matrix W_{XY} as defined by eqn (16) for this chain.

(ii) Show that the 'fair casino' example in section 1 defines a discrete time and discrete state space homogeneous Markov chain. Calculate the stochastic matrix W_{XY} as defined by eqn (16) for this chain.

(iii) Show that the 'gambling banker' example in section 1 defines a discrete time and discrete state space homogeneous Markov chain. Calculate the stochastic matrix W_{XY} as defined by eqn (16) for this chain.

(iv) Show that the 'mutating virus' example in section 1 defines a discrete time and discrete state space homogeneous Markov chain. Calculate the stochastic matrix W_{XY} as defined by eqn (16) for this chain.

(v) Let $\boldsymbol{P} = \{p_{ij}\}$ be an $N \times N$ stochastic matrix. Prove the following statement. If for some $m \in \mathbb{N}$ one has $(\boldsymbol{P}^m)_{ij} > 0$ for all $i, j \in \{1, \ldots, N\}$, then for all $n \geq m$ it is true that $(\boldsymbol{P}^n)_{ij} > 0$ for all $i, j \in \{1, \ldots, N\}$.

(vi) Consider the following matrix:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$

Show that $\boldsymbol{P}$ is a stochastic matrix. Prove that $\boldsymbol{P}$ is ergodic but not regular.

(vii) Let i be an absorbing state of a Markov chain defined by the stochastic matrix $\boldsymbol{P}$, i.e. $p_{ij} = \delta_{ij}$ for all $j \in S$. Prove that the chain cannot be regular. Hint: prove first by induction with respect to $n \geq 0$ that $(\boldsymbol{P}^n)_{ij} = \delta_{ij}$ for all $j \in S$ and all $n \geq 0$, where i denotes the absorbing state.

(viii) Let i be an absorbing state of a Markov chain defined by the stochastic matrix $\boldsymbol{P}$, i.e. $p_{ij} = \delta_{ij}$ for all $j \in S$. Prove that the state $\boldsymbol{p}$ defined by $p_j = \delta_{ij}$ is a stationary state of the process.

(ix) Consider the stochastic matrix $\boldsymbol{P}$ of the 'trained mouse' example in section 1 (which you calculated in an earlier exercise). Determine whether or not this chain is ergodic. Determine whether or not this chain is regular. Determine whether or not this chain has absorbing states. Give the graphical representation of the chain.

(x) Consider the stochastic matrix $\boldsymbol{P}$ of the 'fair casino' example in section 1 (which you calculated in an earlier exercise). Determine whether or not this chain is ergodic. Determine whether or not this chain is regular. Determine whether or not this chain has absorbing states.

(xi) Consider the stochastic matrix $\boldsymbol{P}$ of the 'gambling banker' example in section 1 (which you calculated in an earlier exercise). Calculate its eigenvalues $\{\lambda_i\}$ and verify that $|\lambda_i| \leq 1$ for each i. Determine whether or not this chain is ergodic. Determine whether or not this chain is regular. Determine whether or not this chain has absorbing states. Give the graphical representation of the chain.

(xii) Consider the stochastic matrix $\boldsymbol{P}$ of the 'mutating virus' example in section 1 (which you calculated in an earlier exercise). Calculate its eigenvalues $\{\lambda_i\}$ and verify that $|\lambda_i| \leq 1$ for each i. Determine whether or not this chain is ergodic. Determine whether or not this chain is regular. Determine whether or not this chain has absorbing states. Give the graphical representation of the chain.

(xiii) Consider the stochastic matrix $\boldsymbol{P}$ of the 'trained mouse' example in section 1 (which you analysed in earlier exercises). Find out whether the chain has stationary states. If so, calculate them and determine whether the stationary state is unique. If not, does the time average limit $\overline{\boldsymbol{Q}} = \lim_{m\to\infty} m^{-1}\sum_{n\leq m}(\boldsymbol{P}^n)$ exist? Use your results to answer the question(s) put forward for this example in section 1.

(xiv) Consider the stochastic matrix $\boldsymbol{P}$ of the 'fair casino' example in section 1 (which you analysed in earlier exercises). Find out whether the chain has stationary states. If so, calculate them and determine whether the stationary state is unique. If not, does the time average limit $\overline{\boldsymbol{Q}} = \lim_{m\to\infty} m^{-1}\sum_{n\leq m}(\boldsymbol{P}^n)$ exist? Use your results to answer the question(s) put forward for this example in section 1.

(xv) Consider the stochastic matrix $\boldsymbol{P}$ of the 'gambling banker' example in section 1 (which you analysed in earlier exercises). Find out whether the chain has stationary states. If so, calculate them and determine whether the stationary state is unique. If not, does the time average limit $\overline{\boldsymbol{Q}} = \lim_{m\to\infty} m^{-1}\sum_{n\leq m}(\boldsymbol{P}^n)$ exist? Use your results to answer the question(s) put forward for this example in section 1.

(xvi) Consider the stochastic matrix $\boldsymbol{P}$ of the 'mutating virus' example in section 1 (which you analysed in earlier exercises). Find out whether the chain has stationary states. If so, calculate them and determine whether the stationary state is unique. If not, does the time average limit $\overline{\boldsymbol{Q}} = \lim_{m\to\infty} m^{-1}\sum_{n\leq m}(\boldsymbol{P}^n)$ exist? Use your results to answer the question(s) put forward for this example in section 1.

Appendix B. Probability in a nutshell

Definitions & Conventions. We define 'events' $\boldsymbol{x}$ as n-dimensional vectors, drawn from some event set $A \subseteq \mathrm{I\!R}^n$. We associate with each event a real-valued and non-negative probability measure $p(\boldsymbol{x}) \geq 0$. If A is discrete and countable, each component x_i of $\boldsymbol{x}$ can only assume values from a discrete set A_i so $A \subseteq A_1 \otimes A_2 \otimes \ldots \otimes A_n$. We drop explicit mentioning of sets where possible; e.g. $\sum_{x_i}$ will mean $\sum_{x_i \in A_i}$, and $\sum_{\boldsymbol{x}}$ will mean $\sum_{\boldsymbol{x} \in A}$, etc. No problems arise as long as the arguments of $p(\ldots)$ are symbols; only once we evaluate probabilities for explicit values of the arguments we need to indicate to which components of $\boldsymbol{x}$ such values are assigned. The probabilities are normalized according to $\sum_{\boldsymbol{x}} p(\boldsymbol{x}) = 1$.

Interpretation of Probability. Imagine a system which generates events $\boldsymbol{x} \in A$ sequentially, giving the infinite series $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$. We choose an arbitrary one-to-one index mapping $\pi : \{1, 2, \ldots\} \to \{1, 2, \ldots\}$, and one particular event $\boldsymbol{x} \in A$ (in that order), and calculate for the first M sequence elements $\{\boldsymbol{x}_{\pi(1)}, \ldots, \boldsymbol{x}_{\pi(M)}\}$ the frequency $f_M(\boldsymbol{x})$ with which $\boldsymbol{x}$ occurred:

$$f_M(\boldsymbol{x}) = \frac{1}{M} \sum_{m=1}^{M} \delta_{\boldsymbol{x}, \boldsymbol{x}_{\pi(m)}} \qquad \begin{cases} \delta_{\boldsymbol{x},\boldsymbol{y}} = 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} \\ \delta_{\boldsymbol{x},\boldsymbol{y}} = 0 & \text{if } \boldsymbol{x} \neq \boldsymbol{y} \end{cases}$$

We define *random events* as those generated by a system as above with the property that for each one-to-one index map π, for each event $\boldsymbol{x} \in A$ the frequency of occurrence $f_M(\boldsymbol{x})$ tends to a limit as $M \to \infty$. This limit is then defined as the 'probability' associated with $\boldsymbol{x}$:

$$\forall \boldsymbol{x} \in A: \qquad p(\boldsymbol{x}) = \lim_{M \to \infty} f_M(\boldsymbol{x})$$

Since $f_M(\boldsymbol{x}) \geq 0$ for each $\boldsymbol{x}$, and since $\sum_{\boldsymbol{x}} f_M(\boldsymbol{x}) = 1$ for any M, it follows that $p(\boldsymbol{x}) \geq 0$ and that $\sum_{\boldsymbol{x}} p(\boldsymbol{x}) = 1$ (as it should).

Marginal & Conditional Probabilities, Statistical Independence. The so-called 'marginal probablities' are obtained upon summing over individual components of $\boldsymbol{x} = (x_1, \ldots, x_n)$:

$$p(x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_n) = \sum_{x_\ell} p(x_1, \ldots, x_n) \tag{B.1}$$

In particular we obtain (upon repeating this procedure):

$$p(x_i) = \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} p(x_1, \ldots, x_n) \tag{B.2}$$

Marginal probabilities are normalized. This follows upon combining their definition with the basic normalization $\sum_{\boldsymbol{x}} p(\boldsymbol{x}) = 1$, e.g.

$$\sum_{x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_n} p(x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_n) = 1, \qquad \sum_{x_i} p(x_i) = 1$$

For any two disjunct subsets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_\ell\}$ of the index set $\{1, \ldots, n\}$ (with necessarily $k + \ell \leq n$) we next define the 'conditional probability'

$$p(x_{i_1}, \ldots, x_{i_k} | x_{j_1}, \ldots, x_{j_\ell}) = \frac{p(x_{i_1}, \ldots, x_{i_k}, x_{j_1}, \ldots, x_{j_\ell})}{p(x_{j_1}, \ldots, x_{j_\ell})} \tag{B.3}$$

(B.3) gives the probability that the k components $\{i_1, \ldots, i_k\}$ of $\boldsymbol{x}$ take the values $\{x_{i_1}, \ldots, x_{i_k}\}$, given the knowledge that the ℓ components $\{j_1, \ldots, j_\ell\}$ take the values $\{x_{j_1}, \ldots, x_{j_\ell}\}$.

The concept of statistical independence now follows naturally. Loosely speaking: statistical independence means that conditioning in the sense defined above does not affect any of the marginal probabilities. Thus the n events $\{x_1, \ldots, x_n\}$ are said to be statistically independent if for *any* two disjunct subsets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_\ell\}$ of $\{1, \ldots, n\}$ we have

$$p(x_{i_1}, \ldots, x_{i_k} | x_{j_1}, \ldots, x_{j_\ell}) = p(x_{i_1}, \ldots, x_{i_k})$$

This can be shown to be equivalent to saying

$$\{x_1, \ldots, x_n\} \text{ are independent : } \quad \begin{array}{l} p(x_{i_1}, \ldots, x_{i_k}) = p(x_{i_1})p(x_{i_2}) \ldots p(x_{i_k}) \\ \text{for every subset } \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\} \end{array} \tag{B.4}$$

Appendix C. Eigenvalues and eigenvectors of stochastic matrices

For $\boldsymbol{P}^m$ to remain well-defined for $m \to \infty$, it is vital that the eigenvalues are sufficiently small. Let us inspect the right- and left-eigenvector equations $\sum_j p_{ij} x_j = \lambda x_i$ and $\sum_j p_{ji} y_j = \lambda y_i$, where $\boldsymbol{P}$ is a stochastic matrix, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{|S|}$ and $\lambda \in \mathbb{C}$ (since $\boldsymbol{P}$ need not be symmetric, eigenvalues and eigenvectors need not be real-valued). Much can be extracted from the two defining properties $p_{ik} \geq 0\ \forall (i,k) \in S^2$ and $\sum_i p_{ik} = 1\ \forall k \in S$ alone. We will write complex conjugation of $z \in \mathbb{C}$ as $\overline{z}$. Note that eigenvectors $(\boldsymbol{x}, \boldsymbol{y})$ of $\boldsymbol{P}$ need not be probabilities in the sense of the $\boldsymbol{p}(n)$, as they could have negative or complex entries.

- The spectra of left- and right- eigenvalues of $\boldsymbol{P}$ are identical.

  proof:
  This follows from the fact that a right eigenvector of $\boldsymbol{P}$ is automatically a left eigenvector of $\boldsymbol{P}^\dagger$, where $(\boldsymbol{P}^\dagger)_{ij} = p_{ji}$, so

  $$\begin{aligned} &\text{right eigenv polynomial}: && \det[\boldsymbol{P} - \lambda \mathbf{1}] = 0 \\ &\text{left eigenv polynomial}: && \det[\boldsymbol{P}^\dagger - \lambda \mathbf{1}] = 0 \quad \Rightarrow \quad \det[(\boldsymbol{P} - \lambda \mathbf{1})^\dagger] = 0 \end{aligned}$$

  the proof then follows from the general property $\det \boldsymbol{A} = \det \boldsymbol{A}^\dagger$. []

- All eigenvalues λ of stochastic matrices $\boldsymbol{P}$ obey

  $$\boldsymbol{y}\boldsymbol{P} = \lambda \boldsymbol{y},\ \boldsymbol{y} \neq \mathbf{0}: \quad |\lambda| \leq 1 \tag{C.1}$$

  proof:
  we start from the left eigenvalue equation $\lambda y_i = \sum_j p_{ji} y_j$, take absolute values of both sides, sum over $i \in S$, and use the triangular inequality:

  $$\begin{aligned} \sum_i |\lambda y_i| = \sum_i |\sum_j p_{ji} y_j| \ \Rightarrow\ |\lambda| \sum_i |y_i| &= \sum_i |\sum_j p_{ji} y_j| \\ &\leq \sum_i \sum_j |p_{ji} y_j| = \sum_i \sum_j p_{ji} |y_j| = \sum_j |y_j| \end{aligned}$$

  Since $\boldsymbol{y} \neq \mathbf{0}$ we know that $\sum_i |y_i| \neq 0$ and hence $|\lambda| \leq 1$. []

- Left eigenvectors belonging to eigenvalues $\lambda \neq 1$ of stochastic matrices $\boldsymbol{P}$ obey

  $$\boldsymbol{y}\boldsymbol{P} = \lambda \boldsymbol{y},\ \boldsymbol{y} \neq \mathbf{0}: \quad \text{if } \lambda \neq 1 \text{ then } \sum_i y_i = 0 \tag{C.2}$$

  proof:
  we simply sum both sides of the eigenvalue equation over $i \in S$ and use the normalization of the columns of $\boldsymbol{P}$:

  $$\lambda y_i = \sum_j p_{ji} y_j \ \Rightarrow\ \lambda \sum_i y_i = \sum_j y_j$$

  Clearly, either $\sum_i y_i = 0$ or $\lambda = 1$. This proves our claim. []