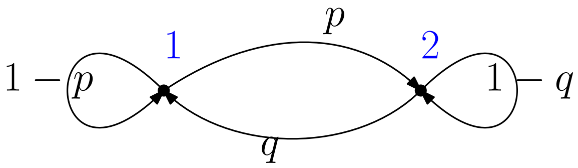
I.2 Classical and important examples of discrete-time chains

September 15, 2020

Two-state chain (exercise 1)



Let $E = \{1, 2\}$ and

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

If $p + q > 0$, a bit of linear algebra leads to

$$P^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}.$$

Two-state chain : special values of parameters

When $p = 0$ or $q = 0$ we'll say the two states do not *communicate*. In that case each state forms a *communication class*. When $p = 0$ δ_1 is invariant, when $q = 0$, δ_2 is invariant.
When both parameters are null, both δ_1 and δ_2 are invariant, hence all probabilities. It is not surprising since the chain remains forever at its starting state. In that case both states (hence both classes) are *recurrent*.
When only one of the parameters is null, the other positive, say $p > 0, q = 0$, state 1 will a.s. be visited finitely many times before the chain gets stuck at 2. Thus 1 is *transient*, while 2 is recurrent.
When $p = q = 1$, the chain can – and does in this case – only visit its starting point at even times, and we'll say that the chain is 2-periodic.

Two-state chain : generic values of parameters

When $pq > 0, (p,q) \neq (1,1)$, the chain is *irreducible, aperiodic, recurrent*, hence *ergodic*. Clearly, then

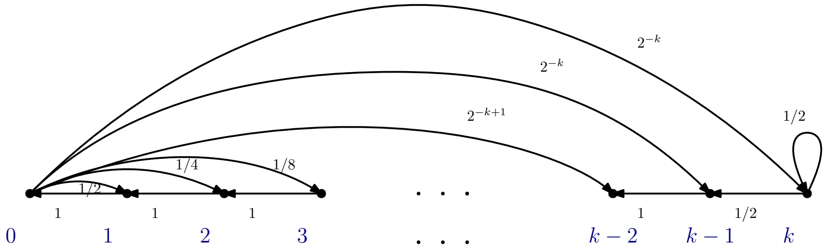
$$P^n \underset{n\to\infty}{\longrightarrow} \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}, \quad \mu P^n \underset{n\to\infty}{\longrightarrow} \left(\frac{q}{p+q}\ \frac{p}{p+q}\right).$$

whatever the initial distribution μ. Thus the law of the chain at time n always converges (exponentially fast) towards

$$\pi := \left(\frac{q}{p+q}\ \frac{p}{p+q}\right).$$

Note that π is not only the *stationary distribution* of the chain, but also its unique invariant probability. It is easily proven directly, as eigenvalue 1 is of multiplicity one for P, hence for P^T also.

Number of consecutive ones at the beginning of a sequence in a moving window of size k (exercise 2)



Number of consecutive ones in a moving window

The above Markov chain is obtained when looking at an infinite sequence of bits (zeroes and ones) on the integer line, by counting the number of consecutive ones from the left in a window of size k which slides by one to the right at each increment of time. We assume further that the window initially reveals bits from 0 to $k-1$, and that bits to the right of the first zero are i.i.d Bernoulli variables with parameter $1/2$.

It is easy that the chain is irreducible, aperiodic and recurrent. We will prove that in such a case there exists a unique invariant probability. It is obvious from the above that the chain reaches stationarity at time k (from time k onwards, the bits which are read are i.i.d Ber$(1/2)$) The (unique) invariant probability of the chain thus corresponds to the distribution of the number of ones in a sequence of k i.i.d Ber$(1/2)$ bits :

$\pi = [1/2\ 1/4\ 1/8\ 1/16\ \cdots\ 2^{-k+1}\ 2^{-k}\ 2^{-k}]$.

Existence, unicity and expression of π can also be checked directly from solving $\pi P = \pi$.

Note that when the chain starts from state $k-1$, it is at 0 at time $k-1$, which is quite far from its stationarity distribution. Hence, in that case, distance from stationarity remains large up to time $k-1$, before it abruptly drops to 0 at time k.
At this point, this is only an informal statement. We need to better define what we mean by "distance from stationarity" to make this statement more precise. See exercise 2 for more details.

Random walks on $\mathbb{Z}^d$ (exercise 3)

Let $(\xi_i)_{i\geq 1}$ i.i.d., taking values in $\mathbb{Z}^d$. Define the *random walk*

$$S_0 = 0, S_n = \sum_{i=1}^{n} \xi_i.$$

Whenever the distribution of ξ only charges the direct neighbours of the origin, we speak of a *nearest-neighbour* random walk, and when $d = 1$ we simply refer to it as *simple* random walk. When the distribution of ξ also charges the origin itself we speak of *lazy* versions of these nearest-neighbour random walks.
The case when $\mathbb{P}(\xi = \mathbf{e}_i) = \mathbb{P}(\xi = -\mathbf{e}_i) = \frac{1}{2d}, i = 1, ..., d$ is refered to as *simple symmetric random walk*.

We assume $\mathbb{P}(\xi = 1) = 1 - \mathbb{P}(\xi = -1) = p \in [0, 1]$. The common lazy version corresponds to $\tilde{P} = (P + Id)/2$.
When $p \in (0, 1)$ the chain is clearly irreducible : for any $n \in \mathbb{N}$, $P^n(0, n) = p^n > 0$, $P^n(0, -n) = (1 - p)^n > 0$. On the other hand, when $p = 0$ or $p = 1$ no two states communicate so each state forms its own class.
The chain is however 2-*periodic* : when starting at 0, it can only visit 0 at even times.

SRW on $\mathbb{Z}$: transience, recurrence

Moreover, there are exactly $\binom{2n}{n}$ trajectories of length $2n$ which start and end up at the origin, so

$$\mathbb{P}_0(S_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n.$$

By Stirling

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim_{n\to\infty} 2^{2n} \frac{1}{\sqrt{\pi n}},$$

so that

$$\mathbb{E}_0[\mathcal{V}_0] = \sum_{n\in\mathbb{N}} \mathbb{P}_0(S_n = 0) = \infty \Leftrightarrow p = 1/2,$$

and it follows from the characterization of recurrence of last section that SRW on $\mathbb{Z}$ is recurrent iff $p = 1/2$.

Simple symmetric RW on $\mathbb{Z}^d$: transience, recurrence

Slightly more delicate combinatorics lead to

$$\mathbb{P}_0(S_{2n} = 0) \sim \frac{c}{n^{d/2}}$$

so that

$$\sum_{n \in \mathbb{N}} \mathbb{P}_0(S_n = 0) = \infty \Leftrightarrow d \leq 2,$$

and it follows that simple symmetric RW on $\mathbb{Z}^d$ is recurrent iff $d \leq 2$.

It can be shown that whatever the value of p there exists no invariant probability for the chain. More precisely, if we try to find all (nonnegative) measures λ such that $\lambda P = \lambda$ we find that for any $n \in \mathbb{Z}$,

$$\begin{aligned} & p\lambda(n-1) + (1-p)\lambda(n+1) = \lambda(n) \\ \Leftrightarrow \quad & p(\lambda(n) - \lambda(n-1)) = (1-p)(\lambda(n+1) - \lambda(n)). \end{aligned}$$

When $p \in \{0, 1/2, 1\}$, it follows that any invariant measure is proportional to the counting measure, that is there exists $\alpha \geq 0$ such that $\lambda(n) = \alpha, n \in \mathbb{Z}$.
When $p \in (0,1) \setminus \{1/2\}$, any invariant measure λ is such that for some $\alpha, \beta \geq 0$, $\lambda(n) = \alpha + \beta \left(\frac{1-p}{p}\right)^n, n \in \mathbb{Z}$.

Gambler's ruin

For $k \in [|0, N|]$, let $p_k = \mathbb{P}_k(T_N < T_0)$. Obviously $p_0 = 0, p_N = 1$.
For $k \in [|1, N-1|]$, by Markov at time 1, we find that

$$p_k = pp_{k+1} + (1-p)p_{k-1} \Leftrightarrow p(p_{k+1} - p_k) = (1-p)(p_k - p_{k-1}).$$

If $p = 1/2$, it follows that $p_k = k/N, k \in [|0, N|]$.
If $p \in (0,1) \setminus \{1/2\}$, we obtain

$$p_k = \frac{1 - \left(\frac{1-p}{p}\right)^k}{1 - \frac{1-p}{p}}, \; k \in [|0, N|].$$

This computation of exit probabilities from the strip $[|0, N|]$ for SRW is refered to as *gambler's ruin* : $1 - p_k$ indeed corresponds to the probability of ruin for a gambler who enters a game with fortune k, bets 1 at each independent hand with probability p of winning, and leaves the game when he reaches fortune N or has no money left.

Probability of escape for transient SRW

When $p > 1/2$, it follows from the above that
$\mathbb{P}_1(T_0 = +\infty) \geq \lim_{N\to\infty} \mathbb{P}_1(T_N < T_0) = 1 - \frac{1-p}{p} = \frac{2p-1}{p} > 0.$
Therefore, when $p > 1/2$, not only do we know that the walk is transient, but we can precisely compute the probability of escape to infinity from the origin : $2p - 1$.
When $p < 1/2$, by symmetry, we obtain the same result with p replaced with $1 - p$, hence $1 - 2p$.
In both cases, the probability of escape from the origin is $|1 - 2p|$.

Branching process

Let $\{\xi_{n,i}, n \in \mathbb{N}, i \in \mathbb{N}^*\}$ a sequence of i.i.d, $\mathbb{N}$-valued random variables and define

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{n,i}.$$

In an idealized model for the evolution of a population, one may think of $\xi_{n,i}$ as the number of offsprings of the ith individual in the nth generation, so Z_n is the size of the population at the nth generation.

In the case $\mu(1) = 1$ for which the transition kernel simply is Id.

Whatever μ the offspring distribution, when at 0 the chain Z remains stuck there, we say 0 is an *absorbing state*, of course 0 is alone in its communication class.

When $\mu(0) > 0$, any given state leads to 0, so that every other class is opened and therefore transient. Thus $\lim_{n\to\infty} Z_n$ exists a.s. and either takes value $+\infty$ (the population explodes) or 0 (the population goes extinct).

By noticing that $\{\xi_{n,i}, i \geq 1\}$ and Z_n are independent, and writing $G(x) = \sum_{n\geq 0} x^n \mu(n)$ one shows that for any $x \in [0,1]$,

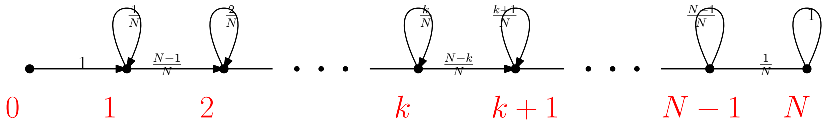
$$G_{n+1}(x) := \mathbb{E}_1[x^{Z_{n+1}}] = G_n(G(x))$$

Now the probability of extinction is

$$\zeta := \lim_{n\to\infty} \mathbb{P}_1(Z_n = 0) = \lim_{n\to\infty} G_n(0),$$

so ζ is the minimal fixed point of G in $[0,1]$.
It is then easy to deduce that $\zeta = 1$ iff $\mu(1) < 1$ and $m := \sum_{i\geq 1} i\mu(i) \leq 1$.

Coupon collector (exercise 4)



Coupon collector

The above chain corresponds to the *coupon collector* model. A total of N coupons are to be collected. The collector receives a coupon choosen independently and uniformly at random at each time increment. If X_n denote the number of distinct coupons that the collector possesses at time n, it is straightforward that X is indeed the Markov chain on $[|0, N|]$ with kernel

$$P(k, k+1) = 1 - P(k, k) = \frac{N-k}{N}, 0 \leq k \leq N-1, \quad P(N, N) = 1.$$

Each state forms its own communication class, all are transient except N which is absorbant. It is obvious that the chain does indeed end up in the state N eventually.

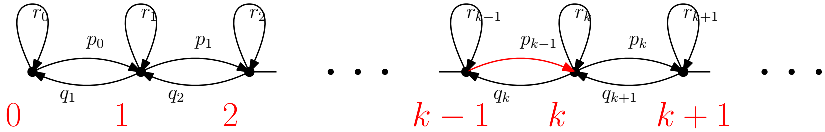
Coupon collector : time of collection

By the strong Markov property at T_k,
$T_{k+1} - T_k := G_k \sim \mathrm{Geom}(\frac{N-k}{N})$, and is independent of T_k, so under $\mathbb{P}_0$, the time of collection T_N is given by $\sum_{k=0}^{N-1} G_k$.
One has $\mathbb{E}[T_N] = \sum_{k=0}^{N-1} \frac{N}{N-k} \sim N\log(N)$, while

$$\mathrm{Var}[T_N] = \sum_{k=0}^{N-1} \frac{k}{N} \frac{N^2}{(N-k)^2} \sim \frac{N^2\pi^2}{6}$$

Thus $\frac{\mathrm{Var}[T_N]}{(\mathbb{E}[T_N])^2} \to 0$ and by Markov's inequality, it follows that $\frac{T_N}{N\log(N)} \to 1$ as $N \to \infty$, in probability.

Birth and death chains (exercise 5)



The class of 0 is $\{0, ..., N_1\}$ where $N_1 := \inf\{n \in \mathbb{N} : p_n q_{n+1} = 0\}$. In particular the chain is irreducible iff $p_i > 0 \forall i \in \mathbb{N}, q_i > 0 \forall i \in \mathbb{N}^*$. We discuss the transience or recurrence of the chain in that case later on.
Furthermore, for any $k \geq 1$, if $N_k < \infty$ the class of $N_k + 1$ is given by $\{N_k + 1, ..., N_{k+1}\}$ where $N_{k+1} := \inf\{n \geq N_k + 1 : p_n q_{n+1} = 0\}$.
When $N_1 < \infty$, we have either $p_{N_1} = 0$, in which case the finite class of 0 is clearly recurrent, or $p_{N_1} > 0, q_{N_1+1} = 0$ in which case the finite class of 0 is opened, hence transient (and the chain eventually will leave that class forever). Similar reasoning applies for further finite classes : either they are opened and transient, or closed and recurrent. If there exists one last infinite class, its transience or recurrence will be discussed later on.

Irreductible birth and death chains : transience vs recurrence

As we did in the gambler's ruin computations,let $h_k = \mathbb{P}_k(T_N < T_0), k \in [|0, N|]$. Of course $h_0 = 0, h_N = 1$, and if $k \in [|1, ..., N-1|]$, by Markov at time 1, we find that

$$\begin{aligned} & h_k = r_k h_k + p_k h_{k+1} + q_k h_{k-1} \\ \Leftrightarrow \quad & p_k(h_{k+1} - h_k) = q_k(h_k - h_{k-1}). \end{aligned}$$

It follows that, in the irreductible case,

$$h_k = h_1 \left(\sum_{\ell=0}^{k-1} \prod_{i=1}^{\ell} \frac{q_i}{p_i} \right),$$

and thus the chain is transient iff $\sum_{\ell=0}^{\infty} \prod_{i=1}^{\ell} \frac{q_i}{p_i} < \infty$.

Irreductible birth and death chains : invariant measures

In this case, $\lambda P = \lambda$ reads $\lambda_0 = r_0\lambda_0 + q_1\lambda_1$, and

$$\lambda_k = r_k\lambda_k + p_{k-1}\lambda_{k-1} + q_{k+1}\lambda_{k+1},\ \forall k \geq 1$$

Solutions to $\lambda P = \lambda$, if they exist, are clearly unique up to a multiplicative constant. Note that the simpler *detailed balance equations*

$$p_k\lambda_k = q_{k+1}\lambda_{k+1},\ \forall k \geq 0$$

indeed imply $\lambda P = \lambda$, and are easier to solve, we find :

$$\lambda_k = \lambda_0 \prod_{i=0}^{k-1} \frac{p_i}{q_{i+1}}.$$

When λ solves detailed balance, we say the chain is *reversible* with respect to λ. Finally observe that there exists a (unique) invariant probability iff $\sum_{k\geq 0} \prod_{i=0}^{k-1} \frac{p_i}{q_{i+1}} < \infty$.

Conductance model (exercise 6)

Consider $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a locally finite graph, $\mathcal{V}$ being the (at most countable) set of vertices, $\mathcal{E}$ the set of unoriented edges (at most a finite number from each vertex), that is pairs of vertices. For more generality, we authorize self-loops, that is edges of the form $\{x, x\}$ for some $x \in \mathcal{V}$. Let

$$c : \begin{cases} \mathcal{E} \to \mathbb{R}_+^* \\ e \to c(e) \end{cases}$$

a conductance function, $c(e)$ or $c(x, y)$ being the conductance of edge $e = \{x, y\}$.
For a vertex x denote $c(x) = \sum_{e : e = \{x,y\}} c(e)$ the sum of all conductances of edges from x. For short we'll talk of $c(x)$ as the conductance at x. Finally introduce

$$P(x, y) := \begin{cases} \frac{c(x,y)}{c(x)} & \text{if } \{x, y\} \in \mathcal{E} \\ 0 & \text{if } \{x, y\} \notin \mathcal{E} \end{cases}$$

The chain on $\mathcal{V}$ with kernel P is called a conductance model (on $\mathcal{G}$, with conductance function c).

In other words, when at vertex x at time n, the chain chooses an edge incident from x proportionally to its conductance, and goes at time $n+1$ to the other extremity of that edge (possibly x itself if the choosen edge is a self-loop).
The chain is irreducible iff $\mathcal{G}$ is connected (otherwise communication classes are the connected components of $\mathcal{G}$). Since the graph is unoriented, all classes are closed, so all finite classes are recurrent. Infinite classes may be transient or recurrent, we'll come back to this later on.
It is straightforward that the chain is reversible with respect to λ such that for some $\alpha > 0$, $\lambda(x) = \alpha c(x), x \in \mathcal{V}$.
As we will establish later on, there exists an invariant probability (and the chain is positive recurrent) iff $c_{\mathcal{G}} := \sum_{x \in \mathcal{V}} c(x) < \infty$, and in that case it is given by $\pi(x) = \frac{c(x)}{c(\mathcal{G})}, x \in \mathcal{V}$.

Theorem

Any conductance model on a connected graph is an irreductible reversible chain.
Conversely, any irreductible, reversible chain is a conductance model.

Proof : We have gone over the direct implication in the previous slide.
Conversely, consider an irreducible reversible Markov chain.
Conductances at each vertex are set equal to the invariant measure, say λ, with respect to which the chain is reversible. The conductance along a given edge, say $\{x, y\}$, is then given by $c(x, y) := \lambda(x)P(x, y) = \lambda(y)P(y, x)$.
Note that we could have choosen to multiply all conductances with the same positive factor without changing the chain.

Conductance model : SRW on a graph

Whenever all edges are equiped with the same conductance, e.g. 1, we get SRW on the graph $\mathcal{G}$, where at each step, the chain goes to one nearest neighbour in the graph choosen uniformly at random. In that case the invariant measure are the uniform measures, so there exists an invariant probability iff $\mathcal{V}$ is finite. Important examples include

- SRW on the complete graph with N vertices,
- simple (symmetric) RW on $\mathbb{Z}^d$ (edges are between nearest neighbours on the usual lattice),
- simple (symmetric) random walk on the d-dimensional discrete torus $\left(\frac{\mathbb{Z}}{N\mathbb{Z}}\right)^d$
- SRW on the hypercube $\{0,1\}^d = \left(\frac{\mathbb{Z}}{2\mathbb{Z}}\right)^d$,
- and many others : SRW on the triangular lattice, SRW on a random d-regular graph, shuffling by random transpositions, etc...

Mapping representation for the lazy SRW on the hypercube (exercise 9)

Assume $(\zeta_i)_{i\geq 1}$ are i.i.d according to the uniform distribution on $\{1, .., d\}$, and $(\xi_i)_{i\geq 1}$ are i.i.d, according to $\mathrm{Ber}(1/2)$, independent of $(\zeta_i)_{i\geq 1}$.
Now consider $\Lambda = \{1, ..., d\} \times \{0, 1\}$, so that $(\zeta_i, \xi_i) \in \Lambda$ and

$$\phi : \begin{cases} \Lambda \times \{0,1\}^d & \to \{0,1\}^d \\ ((\zeta, \xi), x) & \to y = \phi(x) \end{cases},$$

such that $y_i = x_i$ if $i \neq \zeta$, and $y_\zeta = \xi$.
Then $X_{n+1} = \phi((\zeta_n, \xi_n), X_n)$ defines a lazy random walk on the d-dimensional hypercube.
Note that the time at which every coordinate has been updated is exactly the collection time in a coupon collector.

Random walks on groups (exercise 7)

Consider a group $(G, \bullet)$, and a probability μ on G. The chain X with kernel $P(g, h \bullet g) = \mu(h), g \in G, h \in G$ is a random walk on G. This includes many examples : simple random walk on $\mathbb{Z}^d$, on the discrete torus, on the hypercube, ...
This is the right setting for modelling card shuffling : when $G = \mathfrak{t}_n$ is the group of permutations, the configuration of a deck of n cards can be given by $g \in G$ and one can think of X as the successive steps in the shuffling of the deck. A particular method of shuffling simply corresponds to a particular choice of μ. We will, however, only look at idealized, specific choices of μ for which computations remain trackable.

Let X be a random walk on the group G associated with measure μ, as above.

Theorem

- *X is irreducible iff μ generates G, that is, iff* $\mathcal{H} := \{h_1 \bullet h_2 \cdots \bullet h_n : n \in \mathbb{N}, \mu(h_i) > 0 \; \forall i \in [|1, n|]\} = G$.
- *Uniform measures on G are always invariant for X*
- *There exists an invariant probability iff G is finite. In that case, it is unique and it is the uniform one.*
- *If $\mu(h) = \mu(h^{-1})$ for any $h \in G$ then X is reversible. This is a necessary and sufficient condition for reversibility with respect to the uniform measure. This is however, a sufficient but not necessary condition for reversibility in general.*

Proof : see exercise 7

Stock management

Let $(\xi_i, i \geq 1)$ a sequence of i.i.d integer valued variables, think of ξ_n as the demand at time n. A very simple model for stock management is to assume that stock is renewed at (maximum) level $b \in \mathbb{N}^*$ each time it passes below the threshold $a \in \mathbb{N}^*$. We assume obviously that $a < b$. Then

$$X_{n+1} = \begin{cases} (X_n - \xi_n)^+ & \text{if } X_n > a \\ (b - \xi_n)^+ \text{ if } X_n \leq a \end{cases}$$

gives the level of stock at time $n+1$ as a simple function of the level at time n, X_n and the demand at that time, ξ_n. The chain always takes values in $[|0, b|]$. It is a bit cumbersome to treat the general case, though we can say that when $\mathbb{P}(\xi_1 = 0) \in (0,1)$ the class of b is always recurrent, and if there are any other they must be transient. There exists a unique invariant probability, which allows to compute quantities of interest as, e.g. the asymptotic proportion of time demand can not be fully satisfied.

Interacting particle systems : see Liggett

This is a very large framework, in which the state space is typically $S^{\mathcal{V}}$, where S is a (typically finite) space of *states*, and $\mathcal{V}$ are the *interacting particles* (or agents), corresponding to the vertices in a locally finite graph. The state of $x \in \mathcal{V}$ at time $n+1$, say $\xi_{n+1}(x)$, will depend on its own at time n, $\xi_n(x)$, that of its neighbours $\{\xi_n(y), y \sim x\}$, and some added independent randomness.
Most of these models are more practical to study in their continuous-time setting (more on that later on).
Mean-field versions correspond to the case when the underlying graph is the complete graph.
These include colouring of graphs, zero-range process, contact process, Glauber dynamics for the Ising model, exclusion processes, voter model, models for rumour/epidemics (SIR/SIS), etc... They would require a full course on the subject.