# I.1 Discrete time Markov chains : definition, first properties

September 15, 2020

https://setosa.io/markov/

Copy [[0.1,0.85,0.05,0,0], [0.8,0.15,0,0.05,0], [0,0,0.3,0.4,0.3], [0,0,0.2,0.2,0.6], [0,0,0.1,0.2,0.7]].

then [[0.1,0.88,0.02,0,0], [0.8,0.18,0,0.02,0], [0,0,0.3,0.4,0.3], [0,0,0.2,0.2,0.6], [0.2,0,0.1,0.1,0.6]]

# 2-D Ising

Copy https://www.youtube.com/watch?v=BwyTfMV55RQ

# Epidemics model

https://www.lewuathe.com/covid-19-dynamics-with-sir-model.html (18:00)
https://www.frontiersin.org/articles/10.3389/fpubh.2020.00230/full

# Notation

- $E$ : finite or countable *state space*
- $X$ or $(X_n)_{n\in\mathbb{N}}$ : the *discrete-time chain*, a sequence of $E$ valued random variables
- $(\mathcal{F}_n)_{n\in\mathbb{N}}$ (or sometimes $(\mathcal{F}_n)$ for short) its *natural filtration* : $\mathcal{F}_n := \sigma(X_0, ..., X_n), n \in \mathbb{N}$
- $P$ : the *transition kernel* on $E$, $P : E \times E \to [0, 1]$ such that $\forall x \in E,\ P(x, \cdot)$ is a probability on $E$.

# Discrete-time Markov chain : definition

**Definition**

(a) The sequence of $E$-valued random variables $(X_n, n \in \mathbb{N})$ is a *discrete-time Markov chain* iff
$\forall n \in \mathbb{N}$, $\forall (x_0, ..., x_{n+1}) \in E^{n+2}$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_i = x_i, 0 \leq i \leq n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

# Discrete-time Markov chain : definition

**Definition**

(b) When, for any $x, y \in E$, $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ does not depend on $n$, we denote this quantity by $P(x, y)$, and we say that the chain is time-homogeneous, with transition kernel $P$.

We will always assume time-homogeneity in what follows.

# Discrete-time Markov chain : definition

**Definition**

(c) We say $X$ has initial distribution $\mu$, or that $X$ is started at $\mu$ whenever $\mathbb{P}(X_0 = x_0) = \mu(x_0), x_0 \in E$. We denote by $\mathbb{P}_\mu$ the law of the chain $X$ started at $\mu$. We also use $\mathbb{P}_x$ as shorthand for $\mathbb{P}_{\delta_x}$.

Remark : By (a) and (b) the finite-dimensional laws of $X$ are known (see theorem below) and the existence of $\mathbb{P}_\mu$ thus follows from a simple application of Kolmogorov's extension theorem.

**Definition**

When (a),(b),(c) are satisfied, we simply say $X$ is Markov$(\mu, P)$.

For any $n \in \mathbb{N}$, the law of $(X_0, ...X_n)$ only depends on $\mu, P$ (precisions in the next slide), so does the law of $X$, $\mathbb{P}_\mu$.

# Finite-dimensional distributions of $X$

Let $X$ be Markov$(\mu, P)$.

**Theorem**

*(a) For $n \in \mathbb{N}$, $(x_0, ..., x_n) \in E^{n+1}$. Then*

$$\mathbb{P}_\mu(X_i = x_i, 0 \leq i \leq n) = \mu(x_0) \prod_{i=0}^{n-1} P(x_i, x_{i+1}).$$

By (a) and (b), for $n \geq 1$,

$$\mathbb{P}_\mu(X_n = x_n \mid X_i = x_i, 0 \leq i \leq n-1) = P(x_{n-1}, x_n),$$

which allows to prove the theorem by induction on $n$.

# Finite-dimensional distributions of $X$, Markov$(\mu, P)$

**Theorem**

*(b)* $\mathbb{P}_\mu(X_{n+k} = y \mid X_n = x) = P^k(x, y)$, *where* $P^0(x, y) = \mathbb{1}_{\{x=y\}}$ *and for* $n \geq 0$*,*

$$P^{n+1}(x, y) = \sum_{z \in E} P^n(x, z)P(z, y) = \sum_{z \in E} P(x, z)P^n(z, y).$$

Remark : When $E$ is finite, say with elements $\{x_1, ..., x_N\}$, one can represent $P$ as a $N \times N$ matrix, whose entry at row $i$ and column $j$ is $P_{ij} = P(x_i, x_j)$. Then, $P^n$ defined above simply corresponds to the $n$th power of $P$.

# Finite-dimensional distributions of $X$, Markov$(\mu, P)$

**Theorem**

*(c)* $\mathbb{P}_\mu(X_n = y) = \sum_{x\in E} \mu(x)P^n(x,y) =: \mu P^n(y)$.

Remark : When $E$ is finite, with elements $\{x_1, ..., x_N\}$, one can represent $\mu$ as a $1 \times N$ line matrix whose $j$th entry is $\mu_j = \mu(x_j)$. Then $\mu P^n$ simply is the usual matrix product.

**Theorem**

*(c) Let* $f : E \to \mathbb{R}$,
$\mathbb{E}_\mu(f(X_n)) = \sum_{x \in E, y \in E} \mu(x) P^n(x,y) f(y) =: \mu P^n f$.

Remark : When $E$ is finite, with elements $\{x_1, ..., x_n\}$, one can represent $f$ as a vector whose $j$th entry is $f_j = f(x_j)$. Then $\mu P^n f$ simply is the usual matrix product.

# Simple Markov property

**Theorem**

*Let $X$ Markov$(\mu, P)$. Conditionnally given $\{X_k = x\}$, the process $(X_{n+k}, n \geq 0)$ is Markov$(\delta_x, P)$, and is independent of $(X_0, ..., X_n)$. Equivalently, for any $A \in \mathcal{F}_k$, $n \in \mathbb{N}$, $(x_k, ..., x_{k+n}) \in E^{n+1}$, we have*

$$\begin{aligned} & \mathbb{P}_\mu(A \cap \{X_{i+k} = x_{i+k}, 0 \leq i \leq n\} \mid X_k = x) \\ = \; & \mathbb{P}_\mu(A) \mathbb{1}_{\{x = x_k\}} \prod_{i=0}^{n-1} P(x_{k+i}, x_{k+i+1}) \\ = \; & \mathbb{P}_\mu(A) \mathbb{P}_x(X_i = x_{k+i}, 0 \leq i \leq n) \end{aligned}$$

# Simple Markov property : a short proof

When, for some $(x_0, ..., x_k) \in E^{k+1}$, $A = \{X_i = x_i, 0 \leq i \leq k\}$, we find

$$\begin{aligned} & \mathbb{P}_\mu(A \cap \{X_{i+k} = x_{i+k}, 0 \leq i \leq n\} \cap \{X_k = x\}) \\ = \quad & \mu(x_0)\mathbb{1}_{\{x_k = x\}} \prod_{i=0}^{n+k-1} P(x_i, x_{i+1}), \end{aligned}$$

and the desired equality easily follows for such an event $A$.
Now, since $E$ is at most countable, the collection of elementary events $\left\{\{X_i = x_i, 0 \leq i \leq k\}, (x_0, ..., x_k) \in E^{k+1}\right\}$ generates $\mathcal{F}_k$ (in fact events in $\mathcal{F}_k$, except the empty set, are unions, at most countable, of elementary events), so we are done.

# Simple Markov property : a somewhat more sophisticated formulation

**Theorem**

*Let $X$ be Markov$(\mu, P)$. Conditionnally given $\mathcal{F}_k$, the process $(X_{n+k}, n \geq 0)$ has law $\mathbb{P}_{X_k}$.*

# Strong Markov property : motivation

The simple Markov property states that if we condition the chain to be at $x$ at time $k$, its trajectory can be simply decomposed into two independent parts : the one up to time $k$, and the trajectory after time $k$, which is again Markov (started at $x$).
It is natural to ask if the property still holds for certain *random* times. A natural condition on such random time is to be able to decide that such random time takes value $n$ by only looking at the trajectory of the process up to time $n$.

# Strong Markov property

**Definition**

$T : \Omega \to \mathbb{N} \cup \{+\infty\}$ is an $(\mathcal{F}_n)_{n\in\mathbb{N}}$ stopping time iff for any $n \in \mathbb{N}$,

$$\{T = n\} \in \mathcal{F}_n.$$

**Theorem**

*Let $X$ Markov$(\mu, P)$, and $T$ be an $(\mathcal{F}_n)$-stopping time. Conditionnally given $\{T < \infty, X_T = x\}$, the process $(X_{T+k}, n \geq 0)$ is Markov$(\delta_x, P)$, and is independent of $(X_0, ..., X_T)$.*

Proof : exercise 10 in the Exercise sheet I.

# Strong Markov property

## Definition

$T : \Omega \to \mathbb{N} \cup \{+\infty\}$ is an $(\mathcal{F}_n)_{n\in\mathbb{N}}$ stopping time iff for any $n \in \mathbb{N}$,

$$\{T = n\} \in \mathcal{F}_n.$$

## Theorem

*Let $X$ Markov$(\mu, P)$, and $T$ be an $(\mathcal{F}_n)$-stopping time. Conditionnally given $\{T < \infty, X_T = x\}$, the process $(X_{T+k}, n \geq 0)$ is Markov$(\delta_x, P)$, and is independent of $(X_0, ..., X_T)$.*

Proof : exercise 10 in the Exercise sheet I.

See also exercise 4.3 for an example of application of the result.

# Strong Markov property : alternate formulation

## Definition

Let $\mathcal{F}_T := \{A : \ \forall n \in \mathbb{N} \ A \cap \{T = n\} \in \mathcal{F}_n\}$

## Theorem

*Let $X$ Markov$(\mu, P)$, and $T$ be an almost surely finte $(\mathcal{F}_n)$-stopping time. Conditionnally given $\mathcal{F}_T$, the process $(X_{T+k}, n \geq 0)$ has distribution $\mathbb{P}_{X_T}$.*

# Strong Markov : application

Let $x \in E$, $X$ Markov $(\delta_x, P)$, and assume that $T = T_x^+ < \infty$ a.s. (we'll say $x$ is *recurrent* when it is the case). Then, by the strong Markov property, $(X_{T+n}, n \geq 0)$ is Markov $(\delta_x, P)$, and is independent of $(X_0, ..., X_T)$.

# Decomposition of a trajectory into excursions from a recurrent state

Let $x$ be recurrent for $X$. Define $T_0 := 0$ and for $i \geq 0$, $T_{i+1} := \inf\{n > T_i : X_n = x\}$, so that $T_i$ is the time of $i$th return of the chain at $x$. Further define, for $i \geq 1$, $\mathfrak{e}_i := (X_{T_{i-1}}, ..., X_{T_i})$ the so-called $i$th *excursion* from state $x$.
By strong Markov property, $(\mathfrak{e}_i, i \geq 1)$ forms a sequence of i.i.d. random variables.
Remark : When $x$ is not recurrent (a.k.a *transient*), a similar decomposition of the trajectory into independent parts holds. However, there are only finitely many parts (a geometric number with parameter $\mathbb{P}_x(T_x^+ = \infty)$), as the chain eventually leaves state $x$ without returning. Beware that these parts are not i.d. More precisely, all finite parts have the same law (that of an excursion conditioned to return at $x$), while the infinite last part has the law of an excursion conditioned not to return.

# Important examples of stopping times

Let $A \subset E$, define $T_A = \inf\{n \geq 0 : X_n \in A\}$ the *entrance time* in $A$, and $T_A^+ = \inf\{n \geq 1 : X_n \in A\}$ the so-called *return time* in $A$. For short, we write $T_x$ ($T_x^+$, resp.) for $T_{\{x\}}$ ($T_{\{x\}}^+$ resp.). Then for any $A \subset E$, $T_A . T_A^+$ are stopping times in the natural fitration of $X$. The proof is left as an easy exercise.
Note however that neither $T_x^+ - 1$ nor $L_A = \sup\{n \geq 0 : X_n \in A\}$ nor $\sup\{n \leq N : X_n \in A\}$ are stopping times.

# Random mapping representation (exercise 9)

**Theorem**

*(a) Let* $X_0 \sim \mu$*, and* $\phi : E \times \Lambda \to E$*. Let* $(Z_n, n \geq 0)$ *be a sequence of i.i.d. random variables, also independent of* $X_0$*, taking values in* $\Lambda$*. Finally set* $X_{n+1} = \phi(X_n, Z_{n+1})$ *for any* $n \in \mathbb{N}$*. Then* $X$ *is a Markov chain started at* $\mu$*.*
*(b) Reciprocally, any (time-homogeneous) Markov chain taking values in* $E$*, started at* $\mu$*, admits such a representation.*

Remark : No particular assumption has to be made on the set $\Lambda$.

# Random mapping representation : proof

(a) First note that for any $k$, $X_k$ is $\sigma(X_0, Z_1, ..., Z_{k-1})$ measurable. In particular, $Z_n$ is independent of $(X_0, ..., X_n)$. Thus

$$\begin{aligned} & \mathbb{P}(X_{n+1} = y \mid X_0 = x_0, ..., X_{n-1} = x_{n-1}, X_n = x) \\ = & \mathbb{P}(\phi(x, Z_n) = y \mid X_0 = x_0, ..., X_{n-1} = x_{n-1}, X_n = x) \\ = & \mathbb{P}(\phi(x, Z_n) = y) \\ = & \int_{\{z \in \Lambda : \phi(x,z) = y\}} d\mathbb{P}_{Z_1}(z) =: P(x, y) \end{aligned}$$

where we have used at the last line that $Z_n$ has the same law as $Z_1$. The above expression does not depend on $n$, as required. Obvisously $P$ defines a transition kernel on $E$, but one can also check directly that
$\sum_{y \in E} \int_{\{z \in \Lambda : \phi(x,z) = y\}} d\mathbb{P}_{Z_1}(z) = \int_\Lambda d\mathbb{P}_{Z_1}(z) = 1.$

(b) We will set $\Lambda = [0,1]$, and $Z_1 \sim \mathrm{Unif}[0,1]$. Even if it means ordering its elements, one can always write $E = \{y_i, 1 \leq i \leq N\}$ when it is finite, or $E = \{y_i, i \in \mathbb{N}\}$ when it is countable. It remains to set

$$\phi(x,z) = y_j \text{ whenever } \sum_{\ell=1}^{j-1} P(x,y_\ell) < z \leq \sum_{\ell=1}^{j} P(x,y_\ell),$$

with the convention $\sum_1^0 = 0$.

Note that the proof of (b) provides an algorithm to simulate a Markov chain with kernel $P$.
It will also be useful for defining free coupling.

# Image of a Markov chain by an application

If $X$ is Markov with state space $E$, and $f : E \to F$ is not injective, beware that the process $(Y_n := f(X_n), n \in \mathbb{N})$ is in general *not* a Markov chain.

Exercise : Find an example for which $Y$ is indeed not Markov.

Challenge : Find a necessary and sufficient condition on $f$ so $Y$ remains Markov.

# Classification of states

**Definition**

We say $x$ leads to $y$ and write $x \to y$ iff $\exists n \in \mathbb{N}$ such that $P^n(x,y) > 0$.
We say $x$ communicates with $y$ and write $x \leftrightarrow y$ iff $x \to y$ and $y \to x$

Recall $T_y := \inf\{n \geq 0 : X_n = y\}$ and introduce $\mathcal{V}_y := \sum_{n \geq 0} \mathbb{1}_{\{X_n = y\}}$. Note that $\{T_y < \infty\} = \{\mathcal{V}_y > 0\}$, and $\mathbb{E}_x[V_y] = \sum_{n \geq 0} P^n(x,y)$.

**Theorem**

*TFAE :*

(i) $x \to y$

(ii) $\mathbb{P}_x[T_y < \infty] > 0$

(iii) $\mathbb{E}_x[\mathcal{V}_y] > 0$.

# Communication classes

It is straightforward that the relation $\to$ is reflexive, transitive, but non symmetric. Now $\leftrightarrow$ is reflexive, transitive and symmetric, hence it is an equivalence relation, and $E$ can be decomposed into *communication* classes (that is, equivalence classes for the relation $\leftrightarrow$).

**Definition**

When all states communicate (i.e. when there is only one communication class) we say that the chain $X$ (or the kernel $P$) is *irreducible*.

**Definition**

A class $C$ is closed iff
$x \in C$ and $x \to y \;\Rightarrow\; y \in C$.

When $C$ is closed, the restriction $\tilde{P}$ of the kernel $P$ to states in $\mathcal{C}$ still is a transition kernel. A Markov chain $\tilde{X}$ defined on $C$ with kernel $\tilde{P}$ is irreducible.

# Recurrence, transience : definition

**Definition**

We say $x$ is *recurrent* iff $\mathbb{P}_x(T_x^+ < \infty) = 1$. Otherwise $x$ is called *transient*.

Note that when $x$ is recurrent, by strong Markov at $T_1, T_2, \ldots$ the successive return times at $x$, it must be that $\mathcal{V}_x = +\infty$ a.s. under $\mathbb{P}_x$.

**Theorem**

*If $x$ is recurrent and $x \to y$ then $x \leftrightarrow y$ and $y$ is recurrent. Hence recurrence is a class property.*

# Recurrence, transience : class properties

Proof : By applying the strong Markov at successive returns at $x$, under $\mathbb{P}_x$, $\mathcal{V}_x \sim \mathrm{Geom}(\mathbb{P}(T_x^+ = \infty))$ (where a parameter 0 geometric variable simply takes value $+\infty$ a.s.). Hence TFAE :

(i) $x$ is transient.

(ii) $\mathbb{E}_x[\mathcal{V}_x] = \sum_{n\in\mathbb{N}} P^n(x,x) < \infty$.

Now consider $x$ recurrent and $y$ such that $x \to y$. For some $k \in \mathbb{N}$ we have $P^k(x,y) =: p > 0$. If $y \not\to x$, we would have $\mathbb{P}_x^+(T_x = +\infty) \geq p > 0$, a contradiction. Thus $\exists \ell \in \mathbb{N} : P^{ell}(y,x) > 0$. It remains to see that

$$\begin{aligned}
\sum_{i\in\mathbb{N}} P^i(y,y) &\geq \sum_{n\geq 0} P^{\ell+n+k}(y,y) \\
&\geq \sum_{n\geq 0} P^{\ell}(y,x)P^n(x,x)P^k(x,y) = +\infty
\end{aligned}$$

using that $\sum_{n\geq 0} P^n(x,x) = +\infty$. By equivalence of (i) and (ii) above we conclude that $y$ is recurrent.

# Recurrence and transience : immediate properties

1. If $x$ recurrent and $x \leftrightarrow y$, then $\mathbb{P}_x(T_y < \infty) = \mathbb{P}_y(T_x < \infty) = 1$, and $\mathcal{V}_y = +\infty$ a.s. under $\mathbb{P}_x$.
2. An opened class is always transient.
3. A finite closed class is always recurrent
4. If $\mathbb{E} < \infty$ there is at least one finite, closed, recurrent class.

We leave the proofs as exercises.
Remark : An infinite closed class can be either recurrent or transient (think, e.g., of the SRW example)

# Invariant distributions

## Definition

The measure $\pi$ on $E$ is *invariant* for $X$ (or $P$) iff $\pi P = \pi$.

If $\pi$ is an invariant *probability* and $X_0 \sim \pi$ then $X_n \sim \pi$ for all $n \geq 0$.
If $\pi_1, \pi_2$ are invariant *probabilities*, and $\alpha \in [0, 1]$ then $\alpha\pi_1 + (1 - \alpha)\pi_2$ is also invariant. Thus the set of all invariant probabilities of $X$ is *convex*.

# Stationary distributions are invariant

**Theorem**

*If, for some $\mu$ probability on $E$, $\mu P^n \to \pi$ as $n \to \infty$ then $\pi$ is an invariant probability.*

Proof : By Fatou

$$\begin{aligned} \pi P(y) &= \left(\lim_{n\to\infty} \mu P^n\right) P \\ &= \sum_{x\in E} \left(\lim_{n\to\infty} \mu P^n(x)\right) P(x,y) \\ &\leq \lim_{n\to\infty} \mu P^{n+1}(y) = \pi(y) \end{aligned}$$

Since both sides are probabilities, we must have equality and we conclude.

Remark : A slight adaptation of the reasoning, hence the result, remain true even if we only assume convergence along a subsequence.

# Reversibility

## Definition

We say $X$ is *reversible* w.r.t $\lambda$ iff

$$\lambda(x)P(x,y) = \lambda(y)P(y,x) \ \forall (x,y) \in E^2.$$

The above equations are called *detailed balance equations*.

## Theorem

*If $\lambda$ solves detailed balance, then it is invariant.*

Proof : If $\lambda$ solves detailed balance, then

$$\lambda P(y) = \sum_{x \in E} \lambda(x)P(x,y) = \sum_{x \in E} \lambda(y)P(y,x) = \lambda(y).$$