Note 8

Markov processes in discrete time (Markov chains)

1 Random recurrences

Let $(U_n)_{n\geqslant 1}$ be a sequence of i.i.d. uniform random variables taking values in [0, 1] and ϕ_n: $E\times[0,1]\to E$ a given family indexed by integers $n\geqslant 1$ of measurable maps, where $(E,\mathscr{E})$ is a fixed measurable space. Let us fix also a probability distribution ν on $(E,\mathscr{E})$ and consider a family of random variables $(X_n)_{n\geqslant 0}$ defined as follows: X_0 is taken independent of the family $(U_n)_{n\geqslant 1}$ and with law ν and for any $n\geqslant 1$ we let recursively

$$X_n=\phi_n(X_{n-1},U_n).$$

In this way we define a *random recurrence* with initial law ν and transition functions $\phi_n: E\times[0,1]\to E$. Note that, instead of introducing the random variables $(U_n)_{n\geqslant 1}$ and the deterministic functions $(\phi_n)_n$ we could have directly considered the *random* functions $(\Phi_n=\phi_n(\cdot,U_n): E\to E)_n$ which forms an independent family. A random recurrence is the random counterpart of a deterministic (discrete) dynamical system over the space $(E,\mathscr{E})$ where instead of a family of deterministic maps we consider random maps $(\Phi_n)_{n\geqslant 1}$. When each ϕ_n is constant then the process $(X_n)_{n\geqslant 1}$ is easily seen to constitute a family of independent random variables. While if each ϕ_n is (functionally) independent in its second variable, then the sequence $(X_n)_{n\geqslant 0}$ is deterministic conditionally on the value of X_0. In between these two extreme situations: complete independence and complete dependence, lie a family of processes with interesting probabilistic properties.

This setup admits a natural filtration $(\mathscr{F}_n)_{n\geqslant 0}$ which is the filtration generated by the $(U_n)_{n\geqslant 1}$ together with X_0, namely $\mathscr{F}_n=\sigma(X_0,U_1,\ldots,U_n)$ for all $n\geqslant 0$.

Lemma 1. *The random recurrence $(X_n)_{n\geqslant 0}$ satisfies the Markov property wrt. the filtration $(\mathscr{F}_n)_{n\geqslant 0}$, i.e.*

$$\mathbb{E}[f(X_{n+1})|\mathscr{F}_n]=\mathbb{E}[f(X_{n+1})|X_n] \tag{1}$$

for all measureable and bounded $f: E\to\mathbb{R}$.

Proof. (exercise) □

Corollary 2. *The Markov property (1) implies that for all F bounded and measurable wrt. the future σ-algebra $\mathscr{G}_{n+1}=\sigma(X_k: k\geqslant n)$ we also have*

$$\mathbb{E}[F|\mathscr{F}_n]=\mathbb{E}[F|X_n].$$

Proof. The claim follows once we prove it for functions of the form $F=f(X_{n+1},\ldots,X_{n+k})$ for all $k\geqslant 1$ since by the monotone class theorem it will then follow that it holds for any bounded function which is measurable wrt. the σ-algebra generated by $\cup_{k\geqslant 1}\sigma(X_{n+1},\ldots,X_{n+k})$ which is indeed $\mathscr{G}_{n+1}$.

However the Markov property (1) implies that for all $k \geqslant 1$ and all measureable and bounded $f: E^k \to E$ we indeed have

$$\mathbb{E}[f(X_{n+1},\ldots,X_{n+k})|\mathscr{F}_n] = \mathbb{E}[f(X_{n+1},\ldots,X_{n+k})|X_n].$$

This can be obtained by induction. It is true for $k = 1$ and assuming it is true for $k - 1$ then by another monotone class argument we can restrict ourselves to consider functions of the form $f(x_{n+1},\ldots,x_{n+k}) = g(x_{n+1})h(x_{n+2},\ldots,x_{n+k})$ and observe that

$$\mathbb{E}[f(X_{n+1},\ldots,X_{n+k})|\mathscr{F}_n] = \mathbb{E}[g(X_{n+1})\mathbb{E}[h(X_{n+2},\ldots,X_{n+k})|\mathscr{F}_{n+1}]|\mathscr{F}_n] \qquad \text{(conditioning)}$$

$$= \mathbb{E}[g(X_{n+1})\mathbb{E}[h(X_{n+2},\ldots,X_{n+k})|X_{n+1}]|\mathscr{F}_n] \qquad \text{(inductive assumption)}$$

$$= \mathbb{E}[g(X_{n+1})\mathbb{E}[h(X_{n+2},\ldots,X_{n+k})|X_{n+1}]|X_n] \qquad \text{(Markov prop.)}$$

$$= \mathbb{E}[\mathbb{E}[g(X_{n+1})h(X_{n+2},\ldots,X_{n+k})|\mathscr{F}_{n+1}]|X_n] \qquad \text{(conditioning)}$$

$$= \mathbb{E}[g(X_{n+1})h(X_{n+2},\ldots,X_{n+k})|X_n].$$

□

The Markov property then means that for the process $(X_n)_{n\geqslant 0}$ the past σ-algebra $\mathscr{F}_n$ and the future σ-algebra $\mathscr{G}_n$ are independent conditionally on X_n (the present).

For recurrences we can consider the family of probability kernels $P_n: E \times \mathscr{E} \to [0,1]$ given by

$$P_n(x, A) = \mathbb{P}[\phi_n(x, U_1) \in A], \qquad x \in E, A \in \mathscr{E}, n \geqslant 1,$$

which are called *transition kernels*. Then it is not difficult to show that the transition kernel allows to obtain a regular conditional law for X_{n+1} given $\mathscr{F}_n$, namely

$$\mathbb{E}[f(X_{n+1})|\mathscr{F}_n] = \mathbb{E}[f(X_{n+1})|X_n] = (P_{n+1}f)(X_n),$$

where on the r.h.s. we introduce the notation $Tf: E \to E$ to denote the natural action of the probability kernel $T: E \times \mathscr{E} \to [0,1]$ on the space of bounded measureable functions on E:

$$(Tf)(x) := \int_E f(y) T(x, \mathrm{d}y), \qquad x \in E.$$

Note that if μ is a measure on E then we can also define a new measure μT on E by

$$(\mu T)(A) := \mu(T(\cdot, A)) = \int_E T(x, A)\, \mu(\mathrm{d}x), \qquad A \in \mathscr{E}.$$

2 Markov chains

The law of the random recurrence $(X_n)_{n\geqslant 0}$ is completely determined by the family of kernels $(P_n: E \to \pi(E))_{n\geqslant 1}$ and the initial law $\nu \in \Pi(E)$ by the formula

$$\mathbb{E}[f(X_0,\ldots,X_n)] = \int_{E^{n+1}} f(x_0,\ldots,x_n)\, \nu(\mathrm{d}x_0) P_1(x_0, \mathrm{d}x_1) P_2(x_1, \mathrm{d}x_2) \cdots P_n(x_{n-1}, \mathrm{d}x_n). \qquad (2)$$

Remark 3. The integral in the r.h.s. is defined by on product functions of the form $f(x_0, \ldots, x_{n+1}) = f_0(x_0)\cdots f_n(x_n)$ by the formula

$$\nu(f_0 P_1(f_1 P_2(f_2(\cdots P_n(f_n)\cdots))))$$

extended as usual to all measurable functions on $(E^n, \mathscr{E}^{\otimes n})$.

Let us pause a moment to precise this statement. The law μ of the process $(X_n)_{n\geqslant 0}$ is a measure on the infinite product space $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$ where $\mathscr{E}^{\otimes \mathbb{N}}$ is defined as the smalles σ-algebra which makes all the projections $\pi_j: x \in E^{\mathbb{N}} \mapsto x_j \in E$ for $j \geqslant 0$ measurable. What we are assuming above is that this law is determined by its finite dimensional projections. For any $I \subset \mathbb{N}$ let $\pi_I: E^{\mathbb{N}} \to E^I$ be the projection $\pi_I(x) = (x_i)_{i\in I}$. A cylinder set in $\mathscr{E}^{\otimes \mathbb{N}}$ is defined to be of the form $\pi_I^{-1}(A)$ for some $A \in \mathscr{E}^I$ and $I \subset \mathbb{N}$ finite. Cylinder sets generate $\mathscr{E}^{\otimes \mathbb{N}}$ and they are a π-system, therefore by the $\pi - \lambda$ lemma if two measures on $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$ coincide on cylinder sets they are equal.

Note also that existence of the transition kernel is a bit stronger property than the Markov property (1) (this is linked to the existence or not of regular conditional probabilities). In order to abstract the natural properties of random recurrences one defines the class of Markov processes as follows.

Definition 4. *A process $(X_n)_{n\geqslant 0}$ indexed by $\mathbb{N}$ (or $\mathbb{Z}$) is a (discrete time) Markov process on $(E, \mathscr{E})$ wrt. a filtration $(\mathscr{F}_n)_n$ if for all $n \in \mathbb{N}$ (or $\mathbb{Z}$) it satisfies the Markov property (1) with a given family of transition kernels $(P_n)_{n\geqslant 1}$, namely*

$$\mathbb{E}[f(X_{n+1})|\mathscr{F}_n] = (P_{n+1}f)(X_n), \qquad n \geqslant 1.$$

In discrete time Markov processes are usually called Markov chains. We will consider mainly discrete time processes for now, but the notion of Markov process makes sense also in continuous time.

Random recurrences are Markov chains and they are paradigmal examples of such processes. On the other hand, on very general grounds any Markov process can be associated to some random recurrence.

In the following we will abstract from the specific setting of random recurrences and study the general properties of Markov chains.

The space $(E, \mathscr{E})$ is called *state space*. As for random recurrences, the law of a Markov process is completely determined by the initial law ν and the family of kernels $(P_n)_{n\geqslant 1}$ by the formula (2) and a monotone class argument to extend it to the full σ-algebra $\sigma(X_k: k \geqslant 0)$.

3 Canonical realisation and the strong Markov property

We would like now to go in the opposite direction and from an initial law and a family of transition kernels $(P_n)_{n\geqslant 1}$ construct a probability space $(\hat{\Omega}, \hat{\mathscr{F}})$ and a Markov process $(X_n)_{n\geqslant 0}$ on it with such data.

There exists a canonical probability space on which to realise a Markov chain with state space E and it is given by $\hat{\Omega} = E^{\mathbb{N}}$ endowed the product σ-algebra $\mathscr{E}^{\otimes \mathbb{N}}$ which, as we already remarked, is defined as the smallest σ-algebra which makes measurables all the coordinate projections $(\hat{X}_n)_{n\geqslant 0}$ defined naturally as $\hat{X}_n(\omega) = \omega_n$.

On this space we can realise all the Markov chains with initial laws given by δ_x for all $x \in E$. We denote by $\hat{\mathbb{P}}_x$ such probabilities, by $\hat{\mathbb{E}}_x$ the associate expectations. The measure $\hat{\mathbb{P}}_x$ is caracterised by the property that for all measurable and bounded $f: E^{n+1} \to \mathbb{R}$ we have

$$\int_{\hat{\Omega}} f(\omega_0, \ldots, \omega_n) \hat{\mathbb{P}}_x(\mathrm{d}\omega) = \int_{E^{n+1}} f(x_0, \ldots, x_n) \mu_n(\mathrm{d}x) \tag{3}$$

where μ_n is the measure defined by

$$\mu_n(\mathrm{d}x) = \delta_x(\mathrm{d}x_0) P_1(x_0, \mathrm{d}x_1) \cdots P_n(x_{n-1}, \mathrm{d}x_n)$$

as above. The existence of the measure $\hat{\mathbb{P}}_x$ satisfying the property (3) will be for the moment taken as assumption. This is a nontrivial fact which will be further investigated in a later lecture. Once existence is given, uniqueness follows from the $\pi - \lambda$ argument as above.

However let us stress that existence of $\hat{\mathbb{P}}_x$ is a nontrivial fact and that the measures $(\mu_n)_n$ form a projective systems of measures, i.e. they are such that $\mu_{n+1}(A \times E) = \mu_n(A)$ for all $A \in \mathscr{E}^{\otimes \mathbb{N}}$ and for all $n \geqslant 0$. Morever the measure $\hat{\mathbb{P}}_x$ is such that $\hat{\mathbb{P}}_x(A \times E^{\mathbb{N}}) = \mu_n(A)$ for all $A \in \mathscr{E}^{\otimes \mathbb{N}}$.

Note also that the map $x \mapsto \hat{\mathbb{P}}_x: E \to \Pi(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$ is measurable and therefore one can consider it as a probability kernel from E to $E^{\mathbb{N}}$, obtained as the "lifting" of the family of transition kernels $(P_n)_{n \geqslant 1}$.

The construction of the family of laws $(\hat{\mathbb{P}}_x)_{x \in E}$of laws allows to express the consequences of the Markov property in full generality. On $\hat{\Omega}$ define the *shift* $\theta_n: \hat{\Omega} \to \hat{\Omega}$ as $(\theta_n \omega)_k = \omega_{k+n}$ and $\hat{\mathscr{F}}_n := \sigma(\hat{X}_k: 0 \leqslant k \leqslant n)$. Then we have

$$\hat{\mathbb{E}}_x\big[F \circ \theta_n | \hat{\mathscr{F}}_n\big] = \hat{\mathbb{E}}_{X_n}[F], \tag{4}$$

for any initial state $x \in E$ and any bounded measurable function $F: \hat{\Omega} \to \mathbb{R}$. Note that, on a general probability space is not clear how to define the shift θ_n, this is one of the reason one would like to consider the canonical realisation $(\hat{\Omega}, \mathscr{E}^{\otimes \mathbb{N}}, (\mathbb{P}_x)_{x \in E})$.

Theorem 5. *(Strong Markov property) Let T be a finite stopping time wrt the canonical filtration* $\big(\hat{\mathscr{F}}_n\big)_{n \geqslant 0}$. *Then we have*

$$\hat{\mathbb{E}}_x[F \circ \theta_T | \hat{\mathscr{F}}_T] = \hat{\mathbb{E}}_{X_T}[F].$$

Proof. Take $A \in \hat{\mathscr{F}}_T$ and consider this sequence of equalities

$$\hat{\mathbb{E}}_x[F \circ \theta_T \mathbb{1}_A] = \sum_{k=0}^{\infty} \hat{\mathbb{E}}_x[F \circ \theta_T \mathbb{1}_{A, T=k}] = \sum_{k=0}^{\infty} \hat{\mathbb{E}}_x\big[\hat{\mathbb{E}}_x\big[F \circ \theta_k | \hat{\mathscr{F}}_k\big] \mathbb{1}_{A, T=k}\big]$$

$$= \sum_{k=0}^{\infty} \hat{\mathbb{E}}_x\big[\hat{\mathbb{E}}_{X_k}[F] \mathbb{1}_{A, T=k}\big] = \sum_{k=0}^{\infty} \hat{\mathbb{E}}_x\big[\hat{\mathbb{E}}_{X_T}[F] \mathbb{1}_{A, T=k}\big] = \hat{\mathbb{E}}_x\big[\hat{\mathbb{E}}_{X_T}[F] \mathbb{1}_A\big]$$

valid for all bounded and measurable $F: \hat{\Omega} \to \mathbb{R}$ which allows to conclude the claim. □

When the kernel P_n does not depends on n we denote it simply by P and we say that the Markov chain has time homogeneous transition probabilities, or that it is an *(time) homogenous* Markov chain. In this case we let $P^{(n)} = P \cdots P$ (n-fold composition of kernels) so that we can write the law of the r.v. X_n as $P^{(n)}(X_0, \cdot)$. We have the Chapman–Kolmogorov equation

$$P^{(n+m)}(X_0, \mathrm{d}z) = \int_E P^{(n)}(X_0, \mathrm{d}y) P^{(m)} f(y, \mathrm{d}z). \tag{5}$$

which states that the family $(P^{(n)})_{n \geqslant 0}$ is a semigroup of probability kernels.

From now on, unless stated otherwise, we will assume that all the Markov chain are time homogeneous.

Example 6. (Random walk on $\mathbb{R}^n$). Let $E = \mathbb{R}^n$ and consider the (homogeneous) Markov chain given by letting $X_0 = x \in \mathbb{R}^n$ and $X_{n+1} = X_n + Z_n$ where $(Z_n)_{n \geqslant 1}$ is a family of i.i.d. random variables on $\mathbb{R}^n$ with given law ρ. Then the kernel P is given by $P(x, \mathrm{d}y) = (\rho * \delta_x)(\mathrm{d}y)$ where $*$ denotes the convolution of measures on $\mathbb{R}^n$. More explicitly

$$Pf(x) = \int_{\mathbb{R}^n} f(x+y) \rho(\mathrm{d}y), \qquad x \in \mathbb{R}^n$$

and $P^{(n)}(x, y) = (\delta_x * \rho^{*n})(\mathrm{d}y)$ for all $n \geqslant 1$.

4 Martingale problems

Martingales are powerful tools also in the study of Markov processes. Let us consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ endowed with a filtration $(\mathscr{F}_n)_{n \geqslant 0}$.

Definition 7. *Let $\mathscr{L}$ be a linear operator defined on the bounded measurable real-valued functions over E. We say that an adapted process $(X_n)_{n \geqslant 0}$ taking values on $(E, \mathscr{E})$ satisfies the martingale problem wrt. $\mathscr{L}$ with initial law ν if for every bounded and measurable $f: E \to \mathbb{R}$ we have that the process $(M_n^f)_{n \geqslant 0}$ defined by*

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} \mathscr{L} f(X_k) \tag{6}$$

is a martingale wrt. the filtration $(\mathscr{F}_n)_{n \geqslant 0}$ and $X_0 \sim \nu$.

Note that this in particular implies that $f(X_{n+1}) = f(X_n) + \Delta M_{n+1}^f + \mathscr{L} f(X_n)$ and therefore

$$\mathbb{E}[f(X_{n+1}) | \mathscr{F}_n] = \mathbb{E}[\Delta M_{n+1}^f + f(X_n) + \mathscr{L} f(X_n) | \mathscr{F}_n] = (f + \mathscr{L} f)(X_n), \qquad \mathbb{P} - a.s.$$

As a consequence of this formula we have that the operator $T = \mathrm{Id} + \mathscr{L}$ satisfies

$$Tf(X_n) = (f + \mathscr{L} f)(X_n) = \mathbb{E}[(f + \mathscr{L} f)(X_n) | X_n] = \mathbb{E}[\mathbb{E}[f(X_{n+1}) | \mathscr{F}_n] | X_n] = \mathbb{E}[f(X_{n+1}) | X_n]. \tag{7}$$

This is not enough to say something more about T. In particular we would like to conclude that there exists a unique transition kernel P such that $Pf = Tf$ for all bounded functions $f: E \to \mathbb{R}$. A sufficient condition for this is that the martingale problem is not only solved by a single measure $\mathbb{P}$ but that there exists on $(\Omega, \mathscr{F})$ a family of measure $(\mathbb{P}_x)_{x \in E}$ such that under $\mathbb{P}_x$ the process $(X_n)_n$ solves the martingale problem for $\mathscr{L}$ with initial law δ_x. In this case we have for all $x \in E$ that eq. (7) is satisfied when $n = 0$ $\mathbb{P}_x$ almost surely and this means that

$$Tf(x) = Tf(X_0)(\omega) = \mathbb{E}_x[f(X_1) | X_0](\omega) = \mathbb{E}_x[f(X_1)], \qquad \text{for } \mathbb{P}_x\text{-almost all } \omega \in \Omega,$$

since one has $\mathbb{P}_x(X_0 = x) = 1$ and so the conditional probability can be simplified. And from this we can conclude that $Tf(x) = \mathbb{E}_x[f(X_1)]$, that is T is indeed given by a proability kernel P which is nothing else that the law of X_1 under the measure $\mathbb{P}_x$. So in this case every solution of a martingale problem is a Markov process with transition kernel $P = \mathbb{I} + \mathscr{L}$.

Conversely if $(X_n)_{n\geqslant 0}$ is a Markov process with kernel P then it follows easily that it also a solution of the martingale problem with generator $\mathscr{L} = P - \mathbb{I}$, indeed

$$\mathbb{E}[\Delta M^f_{n+1}|\mathscr{F}_n] = \mathbb{E}[f(X_{n+1}) - f(X_n) - \mathscr{L}f(X_n)|\mathscr{F}_n] = \mathbb{E}[f(X_{n+1})|\mathscr{F}_n] - (Pf)(X_n) = 0$$

and therefore $(M^f_n)_{n\geqslant 0}$ is a martingale for all f.

Since both the maringale problem and the Markov property can be formulated taking into account properties of a given proces $(X_n)_{n\geqslant 0}$ one can simply consider the canonical process $(X_n)_{n\geqslant 0}$ on the space space $(\hat{\Omega}, \mathscr{F}) = (E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$ and say that a measure $\mathbb{Q}$ on $(\hat{\Omega}, \mathscr{F})$ is a Markov process (resp. a solution of the martingale problem) if the canonical process is a Markov process. (resp. a solution to the martingale problem). We can then formulate the following result

Theorem 8. *A family of laws $(\mathbb{P}_x)_{x\in E}$ for which $\mathbb{P}_x(X_0 = x) = 1$ for all $x \in E$ is a (homogeneous) Markov process with transition kernel P iff it solves the martingale problem for the generator $\mathscr{L} = P - \mathbb{I}$.*

An immediate consequence of this identification is that the solution of the martingale problem is unique in law, namely two families $(\mathbb{P}_x)_{x\in E}$ and $(\tilde{\mathbb{P}}_x)_{x\in E}$ which solve the martingale problem for $\mathscr{L}$ must be equal.

We call $\mathscr{L} = P - \mathbb{I}$ the *generator* of the Markov chain.

Note that eq. (6) gives Doob's decomposition of the process $(f(X_n))_{n\geqslant 0}$ as the sum of a martingales and a previsible process (and an initial value).

Special classes of functions are those for which $(f(X_n))_{n\geqslant 0}$ is a martingale itself, or a supermartingale or a submartingales from every possible starting point X_0. Respectively they are given by the condition $\mathscr{L}f(x) = 0$, $\mathscr{L}f(x) \leqslant 0$ or $\mathscr{L}f(x) \geqslant 0$ for all $x \in E$. A function for which $\mathscr{L}f = 0$ is called *harmonic* (for $\mathscr{L}$), similarly the functions for which $\mathscr{L}f \leqslant 0$ are *superharmonic* (since the corresponds to supermartingales) and those for which $\mathscr{L}f \geqslant 0$ are *subharmonic*.

The generator of a Markov chain satisfies the following property:

Theorem 9. *(Maximum principle) Let $D \in \mathscr{E}$ and $T_{D^c} = \inf\{n \geqslant 0 : X_n \in D^c\}$. Assume that T_{D^c} is finite almost surely, then if $f : E \to \mathbb{R}$ is a bounded function which is subharmonic for $\mathscr{L}$ in D (i.e. $\mathscr{L}f(x) \geqslant 0$ for all $x \in D$) we have*

$$\sup_{x\in D} f(x) \leqslant \sup_{x\in D^c} f(x).$$

Proof. Consider the submartingale $(f(X_n))_{n\geqslant 0}$ under the measure $\mathbb{P}_x$. By the submartingale property, the optional stopping theorem and dominated convergence, that

$$f(x) = \mathbb{E}_x[f(X_0)] \leqslant \mathbb{E}_x[f(X_{n\wedge T_{D^c}})] \xrightarrow[n\to\infty]{} \mathbb{E}_x[f(X_{T_{D^c}})] \leqslant \sup_{z\in D^c} f(z),$$

for all $x \in D$. □

5 Computations of probabilities related to a Markov chain

For many reasons we are interested to compute certain probabilities related to a Markov chain, for example: the probability to reach some given set before reaching another set, the expected time to exit some given region, the average of certain additive functionals of the trajectory of the chain. In these situations the Markov property provides a drastric simplication of the orginal problem by relating it to the solution of certain linear equations.

Let us give some examples.

Let $(X_n)_{n\geqslant 0}$ the canonical realisation of a Markov chain on $(E,\mathscr{E})$ and let $A,B\in\mathscr{E}$ with $A\cap B=\emptyset$. Assume we want to compute $\mathbb{P}_x(X \text{ reaches } A \text{ before reaching } B)$. We can let $T_A=\inf\{n\geqslant 0: X_n\in A\}$ and similarly for T_B. Then $\mathbb{P}_x(X \text{ reaches } A \text{ before reaching } B)=\mathbb{P}_x(T_A<T_B)$. Now by definition if $x\in A$ then $u(x)=1$ while if $x\in B$ then $u(x)=0$. If $x\notin A\cup B$ we can reason as follows: the chain has do at least one step to reach A or B and we have $\mathbb{1}_{T_A<T_B}=\mathbb{1}_{T_A<T_B}\circ\theta_1$ on $X_0\notin A\cup B$ (think to why). In this case we can condition on the value of this first step and by the Markov property we have

$$\mathbb{P}_x(T_A<T_B)=\mathbb{E}_x[\mathbb{1}_{T_A<T_B}\circ\theta_1]=\mathbb{E}_x[\mathbb{E}_x[\mathbb{1}_{T_A<T_B}\circ\theta_1|\mathscr{F}_1]]=\mathbb{E}_x[\mathbb{E}_{X_1}[\mathbb{1}_{T_A<T_B}]]$$

(pay attention to the meaning of this last expression). If we define $u(x)=\mathbb{P}_x(T_A<T_B)$ we have established that $u(x)=\mathbb{E}_x[u(X_1)]=(Pu)(x)$ for $x\in(A\cup B)^c$. Using the language of the generator:

$$\begin{cases}(\mathscr{L}u)(x) &= 0, & x\in(A\cup B)^c,\\ u(x) &= 1, & x\in A,\\ u(x) &= 0, & x\in B.\end{cases} \tag{8}$$

Therefore u is harmonic in $(A\cup B)^c$ with prescribed values in $A\cup B$. In many situation this family of equations allows to completely detemine the function u among all the positive functions. In general however there could be multiple solutions to the system (8) which do not give our initial probability u. Assume $v\geqslant 0$ is another such solutions, then by definition the process $(v(X_n))_n$ is a martingale until it stays away from $A\cup B$, namely the stopped process $(v(X_n^{T_A\wedge T_B}))_n$ is a martingale. Therefore by the martingale property, for all $n\geqslant 0$,

$$v(x)=\mathbb{E}_x[v(X_0^{T_A\wedge T_B})]=\mathbb{E}_x[v(X_n^{T_A\wedge T_B})]$$

$$=\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A<T_B}]+\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A=T_B=\infty}]+\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A>T_B}]$$

$$\geqslant\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A<T_B}]+\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A>T_B}].$$

Letting $n\to\infty$ this quantity converges to

$$\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A<T_B}]+\mathbb{E}_x[v(X_n^{T_A\wedge T_B})\mathbb{1}_{T_A>T_B}]\to\mathbb{E}_x[v(X_{T_A})\mathbb{1}_{T_A<T_B}]+\mathbb{E}_x[v(X_{T_B})\mathbb{1}_{T_A>T_B}]$$

$$=\mathbb{E}_x[\mathbb{1}_{T_A<T_B}]=u(x)$$

Therefore we have proved that $v(x)\geqslant u(x)$ for any positive solution to (8) and moreover we can also show that if $T_A\wedge T_B<+\infty$ a.s. then $v(x)=u(x)$ namely, there exisys only one positive solution to the above system.

Remark 10. This line of reasoning draws a strong link between the asymptotic behaviour of certain Markov processes and the solutions of a wide class of linear equations. In continuous time the generator of continuous Markov processes corresponds to second order differential equation and this link become a powerful tool to study elliptic and parabolic equations, especially when the coefficients or the domain are not very regular.

Probabilities are not the only quantities amenable to this kind of approach. Let us consider for example the average time $u(x) = \mathbb{E}_x[T_A]$ to reach a given set $A \in \mathscr{E}$ starting from the point $x \in A$. In this case if $x \in A^c$ we have $T_A = T_A \circ \theta_1 + 1$ and the Markov property gives

$$u(x) = \mathbb{E}_x[T_A \circ \theta_1] = \mathbb{E}_x[1 + \mathbb{E}_{X_1}[T_A]] = 1 + \mathbb{E}_x[u(x)] = 1 + (Pu)(x)$$

while if $x \in A$ then $u(x) = 0$. Therefore the equation for u now reads

$$\begin{cases} (\mathscr{L}u)(x) = -1, & x \in A^c, \\ u(x) = 0, & x \in A, \end{cases} \tag{9}$$

and again $\mathbb{E}_x[T_A]$ is the smallest positive solution to this equation, indeed, if v is another solution the process $M_n = v(X_n^{T_A}) + (n \wedge T_A)$ is a martingale and then

$$v(x) = \mathbb{E}_x[M_0] = \mathbb{E}_x[M_n] = \mathbb{E}_x[v(X_n^{T_A})] + \mathbb{E}_x[(n \wedge T_A)] \geqslant \mathbb{E}_x[n \wedge T_A]$$

and by monotone convergence, as $n \to \infty$ we have $v(x) \geqslant \mathbb{E}_x[T_A] = u(x)$.

Note that we have $T_A = \sum_{k=0}^{T_A - 1} 1$. Therefore the same reasoning can be applied to more general quantities of the form

$$u(x) = \mathbb{E}_x\left[\sum_{k=0}^{T_A - 1} g(X_k) + q(X_{T_A}) \mathbb{1}_{T_A < \infty}\right]$$

for given (positive) functions $g, q: E \to \mathbb{R}_+$ and set $A \in \mathscr{E}$. In this case the equation satisfied by u is of the general form

$$\begin{cases} (\mathscr{L}u)(x) = -g(x), & x \in A^c, \\ u(x) = q(x), & x \in A. \end{cases}$$

And is not difficult to prove that for any solution v of this system

$$M_n = v(X_n) + \sum_{k=0}^{T_A - 1} g(X_k)$$

is a positive martingale and that $v \geqslant u$.