# Markov Chains, complementary notes

September 19, 2022

We recall that for a sequence $(a_k)_{k\geqslant 0}$, we denote $a_{k:n} = (a_k, a_{k+1}, \dots, a_n)$ and $a_{k:\infty} = (a_k, a_{k+1}, \dots, a_{k+j}, \dots)$. We also recall that $h : \mathsf{X} \to \mathbb{R}$ is called a simple function, if $h = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$, where $A_i \in \mathcal{X}$ and $\alpha_i \in \mathbb{R}$.

## 1 Reminders

### 1.1 Conditional probabilities

Let $(\mathsf{X}, \mathcal{X}, \mathbb{P})$ be probability space. If $A \in \mathcal{X}$, then the following equality holds.

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]\,.$$

Similarly if $\mathcal{F} \subset \mathcal{X}$ is a (sub)-$\sigma$-algebra, then we define the conditional probability (relatively to $\mathcal{F}$) as:

$$\mathbb{P}(\mathbb{1}_A|\mathcal{F}) = \mathbb{E}[\mathbb{1}_A|\mathcal{F}]\,.$$

Recall that, given an integrable random variable $X$ and a (sub)-$\sigma$-algebra $\mathcal{F}$, $\mathbb{E}[X|\mathcal{F}]$ is a $\mathcal{F}$-measurable random variable that satisfies:

$$\forall B \in \mathcal{F}\,, \quad \mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbb{1}_B]\,.$$

### 1.2 Product $\sigma$-algebras

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space. The product $\sigma$-algebra $\mathcal{X} \otimes \mathcal{X}$ is defined as follows:

$$\mathcal{X}^{\otimes 2} = \mathcal{X} \otimes \mathcal{X} := \sigma(\{A \times B : A, B \in \mathcal{X}\})\,.$$

With this definition $(\mathsf{X}^2, \mathcal{X}^{\otimes 2})$ is a measurable space. Similarly, for $k > 0$, we define:

$$\mathcal{X}^{\otimes k} := \sigma(\{A_1 \times A_2 \times \cdots \times A_k : A_1, A_2, \dots, A_k \in \mathcal{X}\})\,.$$

With this definition $(\mathsf{X}^k, \mathcal{X}^{\otimes k})$ is a measurable space.

**Finally**, to define a $\sigma$-algebra on the infinite product $\mathsf{X}^{\mathbb{N}}$, we need the following definition.

**Definition 1.** *We say that $C \in \mathsf{X}^{\mathbb{N}}$ is a cylinder, if there is $k \in \mathbb{N}$ and $A_1, \dots, A_k \in \mathcal{X}$, such that:*

$$C = \prod_{i=1}^n A_i \times \mathsf{X} \times \mathsf{X} \times \cdots \times \mathsf{X} \times \cdots = \prod_{i=1}^n A_i \times \mathsf{X}^{\mathbb{N}}\,.$$

The *cylindrical* $\sigma$-algebra, $\mathcal{X}^{\otimes\mathbb{N}}$, is then defined as:

$$\mathcal{X}^{\otimes\mathbb{N}} = \sigma(\{C : C \text{ is a cylinder}\}) .$$

With this definition $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}})$ is a measurable space.

### 1.3 $\pi$-$\lambda$ theorem

*Reminder: $\pi$-system.* Let $\mathcal{A}$ be a collection of subsets in $\mathsf{X}$. We say that $\mathcal{A}$ is a $\pi$-system, if it is non-empty and is stable by finite intersection. I.e. if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.

*Reminder: $\lambda$-system.* We say that $\mathcal{A}$, a collection of subsets in $\mathsf{X}$, is a $\lambda$-system if the following holds.

1. $\varnothing \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. If $(A_i)_{i\in\mathbb{N}}$ is a collection of disjoint elements in $\mathcal{A}$, then $\bigcup_{i=0}^{\infty} A_i \in \mathcal{A}$.

**Theorem 1** ($\pi$-$\lambda$ theorem)**.** *If $\mathcal{A}_\pi \subset \mathcal{A}_\lambda$, with $\mathcal{A}_\pi$ (respectively $\mathcal{A}_\lambda$) a $\pi$-system (resp. $\lambda$-system), then $\sigma(\mathcal{A}_\pi) \subset \mathcal{A}_\lambda$.*

**Corollary 1.** *Let $(\mathsf{X}, \mathcal{X})$ be a probability space and $\nu, \mu$ two probability measures on it. If $\mu, \nu$ agree on, $\mathcal{A}$, a $\pi$-system, then, they agree on $\sigma(\mathcal{A})$.*

*Proof.* Denote $\Lambda = \{A \in \mathcal{X} : \mu(A) = \nu(A)\}$. $\Lambda$ is a $\lambda$ system, and $\mathcal{A} \subset \Lambda$. Thus, applying Theorem 1, we obtain that $\sigma(\mathcal{A}) \subset \Lambda$, which completes the proof. □

It is easy to see that for $k \in \mathbb{N}$, the collection of sets $\{A \in \mathcal{X}^{\otimes k} : A = A_1 \times \cdots \times A_k\ ,\ \text{with } A_i \in \mathcal{X}\}$ form a $\pi$-system that generates $\mathcal{X}^{\otimes k}$.

Furthermore, the cylinders form a generating $\pi$-system of $\mathcal{X}^{\otimes\mathbb{N}}$. We will use both of these results in the next section.

## 2 Markov kernels

In this section we provide additional results that complete the proofs of Chapter 1 of the course. The main idea of all of these proofs is to show the given equation (e.g. Markov property) for some simple functions/sets and then apply the $\pi$-$\lambda$ theorem, to prove the given equation for all functions/sets.

### 2.1 Proof of Lemma 1.4.

In the course-notes it was shown that for any $n \in \mathbb{N}$ and measurable, bounded functions $h_0, h_1, \ldots, h_n : \mathsf{X} \to \mathbb{R}$ it holds that:

$$\mathbb{E}\left[\prod_{i=0}^{n} h_i(X_i)\right] = \int_{\mathsf{X}^{n+1}} \prod_{i=0}^{n} h_i(x_i)\nu(\mathrm{d}x_0) \prod_{i=1}^{n} P(x_{i-1}, \mathrm{d}x_i) .$$

In particular, this holds for $h_0, h_1, \ldots, h_n := \mathbb{1}_{A_0}, \mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$, with $A_i \in \mathcal{X}$. Thus, if $A = A_0 \times \cdots \times A_n$, with $A_i \in \mathcal{X}$, then:

$$\begin{aligned}\mathbb{P}(\mathbb{1}_A(X_{0:n})) &= \int_{\mathsf{X}^{n+1}} \nu(\mathrm{d}x_0) \prod_{i=0}^{n} \mathbb{1}_{A_i}(x_n) \prod_{i=1}^{n} P(x_{i-1}, \mathrm{d}x_i) \\ &= \int_{\mathsf{X}^{n+1}} \nu(\mathrm{d}x_0) \mathbb{1}_A(x_{0:n}) \prod_{i=1}^{n} P(x_{i-1}, \mathrm{d}x_i) \,. \end{aligned} \tag{1}$$

Equation (1) shows that the law of $X_{0:n}$ agrees with the probability measure $\nu(\mathrm{d}x_0)\prod_{i=1}^{n} P(x_{i-1}, \mathrm{d}x_i)$ on the events of the form $A = A_0 \times \cdots \times A_n$, with $A_i \in \mathcal{X}$. As explained in Section 1.3 these sets form a $\pi$-system and thus applying Corollary 1 we obtain that both of the laws agree on the whole $\mathcal{X}^{\otimes n+1}$

## 2.2 Proofs for Section 1.3.2 of the course notes

We admitted (see Theorem 1.5 of the course notes), that for every $\nu \in \mathcal{M}_+(\mathsf{X})$ and a Markov kernel $P$, there is a probability measure $\mathbb{P}_\nu$, such that the coordinates processes $(X_k)_{k \geqslant 0}$, where $X_k(w_{0:\infty}) = w_k$, is a Markov chain with initial distribution $\nu$ and a kernel $P$ on $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\nu)$.

We want to prove the following link between $\mathbb{P}_\nu$ and $\mathbb{P}_x$ (recall that $\mathbb{P}_x$ is the notation of $\mathbb{P}_{\delta_x}$, where $\delta_x$ is the dirac probability measure at $x$)

$$\mathbb{P}_\nu(A) = \int_{\mathsf{X}} \nu(\mathrm{d}x) \mathbb{P}_x(A) \,. \tag{2}$$

Notice that for Equation (2) to make sense we also need to prove that the mapping $x \mapsto \mathbb{P}_x(A)$ is $\mathcal{B}([0,1])/\mathcal{X}$ measurable.

We have the following proposition.

**Proposition 1.** *For all $A \in \mathcal{X}^{\otimes \mathbb{N}}$, the following holds:*

1. *The mapping $x \mapsto \mathbb{P}_x(A)$ is $\mathcal{B}([0,1])/\mathcal{X}$ measurable.*
2. $\mathbb{P}_\nu(A) = \int_{\mathsf{X}} \nu(\mathrm{d}x)\mathbb{P}_x(A)$.

*Proof.* The proof will be done in two steps.
*1st step. The collection of sets that satisfy both points is a $\lambda$-system.*
*2nd step. Both points are true for cylinders.*

Since the set of cylinders is a $\pi$-system that generates $\mathcal{X}^{\otimes \mathbb{N}}$ (see Section 1.3), applying the $\pi$-$\lambda$ theorem will prove the proposition.

The proof of the first step immediately follows from the definition of a $\lambda$-system.

To prove the second step, we will prove by induction a slightly stronger result:

$$\begin{gathered}\forall n \geqslant 0 \text{ and } h_0, \ldots, h_n : \mathsf{X} \mapsto \mathbb{R} \text{ bounded and measurable } : \\ x \mapsto \mathbb{E}_x\left[\prod_{i=0}^{n} h_i(X_i)\right] \text{ is measurable} \\ \mathbb{E}_\nu\left[\prod_{i=0}^{n} h_i(X_0)\right] = \int_{\mathsf{X}} \nu(\mathrm{d}x) \mathbb{E}_x\left[\prod_{i=0}^{n} h_i(X_i)\right] \,. \end{gathered} \tag{3}$$

Indeed, if we prove (3), then choosing $h_i = \mathbb{1}_{A_i}$ this will imply the second step.

For $n = 0$, the property is immediate, since $\mathbb{E}_x[h_i(X_0)] = h_i(x)$. Therefore, assume that the property is true for $n \geqslant 0$, we have that:

$$\begin{aligned}\mathbb{E}_x\left[\prod_{i=0}^{n+1} h_i(X_i)\right] &= h_0(x)\int_{\mathsf{X}}^{n+1} P(x, \mathrm{d}x_1)\prod_{i=1}^{n} h_i(x_i)P(x_i, \mathrm{d}x_{i+1})h(x_{n+1}) \\ &= h_0(x)\int_{\mathsf{X}}^{n+1} P(x, \mathrm{d}x_1)\prod_{i=1}^{n} h_i(x_i)\prod_{j=1}^{n-1} P(x_j, \mathrm{d}x_{j+1})Ph(x_n) \\ &= \mathbb{E}_x\left[Ph(X_n)\prod_{i=0}^{n} h_i(X_i)\right],\end{aligned}$$

where the second equality is obtained by integrating through $x_{n+1}$. Thus the last equality allows us to apply the induction and we obtain that $x \mapsto \mathbb{E}_x\left[\prod_{i=0}^{n+1} h_i(X_i)\right]$ is measurable.

Furthermore,

$$\begin{aligned}\int_{\mathsf{X}} \nu(\mathrm{d}x)\mathbb{E}_x\left[\prod_{i=0}^{n+1} h_i(X_i)\right] &= \int_{\mathsf{X}} \nu(\mathrm{d}x)\mathbb{E}_x\left[Ph(X_n)\prod_{i=0}^{n} h_i(X_i)\right] \\ &= \mathbb{E}_\nu\left[Ph(X_n)\prod_{i=0}^{n} h_i(X_i)\right] \\ &= \int_{\mathsf{X}^{n+1}} h_i(x_0)\nu(\mathrm{d}x_0)\prod_{i=1}^{n} h_i(x_i)P(x_{i-1}, \mathrm{d}x_i)Ph_{n+1}(x_n) \\ &= \int_{\mathsf{X}^{n+2}} h_i(x_0)\nu(\mathrm{d}x_0)\prod_{i=1}^{n} h_i(x_i)P(x_{i-1}, \mathrm{d}x_i)h(x_{n+1})P(x_n, \mathrm{d}x_{n+1}) \\ &= \int_{\mathsf{X}^{n+2}} h_i(x_0)\nu(\mathrm{d}x_0)\prod_{i=1}^{n+1} h_i(x_i) \\ &= \mathbb{E}_\nu\left[\prod_{i=0}^{n+1} h_i(X_i)\right].\end{aligned}$$

where the second equality comes from the induction property.

Thus, the statement (3) is proved, which completes the proof of this proposition.

## 2.3 Markov property

For $k \in \mathbb{N}$, let $X_k : \mathsf{X}^{\mathbb{N}} \to \mathsf{X}$, be defined as $X_k(w_{0:\infty}) = w_k$. For $\nu$ a probability measure on $(\mathsf{X}, \mathcal{X})$ and a kernel $P$ on $(\mathsf{X}, \mathcal{X})$, let $\mathbb{P}_\nu$ be the probability measure on $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}})$ such that $(X_k)$ is a Markov chain with initial distribution $\nu$ and a kernel $P$.

The Markov's property states that for a bounded or non-negative, measurable function $h : \mathsf{X}^{\mathbb{N}} \to \mathbb{R}$, and for $k \in \mathbb{N}$,

$$\mathbb{E}_\nu[h(X_{k:\infty})|\sigma(X_{0:k})] = \mathbb{E}_{X_k}[h(X_{0:\infty})], \quad \mathbb{P}_\nu \text{ almost surely.} \tag{4}$$

We stress that both of the terms of the equality are $\mathcal{F}_k$-measurable random variables, where given $w \in \mathsf{X}^{\mathbb{N}}$, the right-hand term becomes $\mathbb{E}_{X_k(w)}[h(X_{0:\infty})]$: the expectation of $h(X_{0:\infty})$ if the initial distribution of $(X_k)$ was a dirac in $X_k(w)$.

The proof of the Markov's property will be done in four steps.

*First step.* The Markov's property is valid if $h(X_{0:\infty}) = h_0(X_0)$, where $h_0 : \mathsf{X} \to \mathbb{R}$ is measurable and bounded or non-negative.

*Second step.* The Markov's property is valid if $h(X_{0:\infty}) = \prod_{i=0}^{n} h_i(X_i)$, where each $h_i$ : $\mathsf{X} \to \mathbb{R}$ is measurable and bounded or non-negative.

*Third step.* Notice that the set of cylinders is a $\pi$-system and the set of $A \in \mathcal{X}^{\otimes \mathbb{N}}$ such that the Markov's property is valid for $\mathbb{1}_A$ is a $\lambda$-system. Thus, applying the $\pi$-$\lambda$ theorem and the *second step*, we obtain that the Markov's property is valid for $h = \mathbb{1}_A$, with $A \in \mathcal{X}^{\otimes \mathbb{N}}$.

*Fourth step.* Finally, any measurable, bounded or non-negative $h : \mathsf{X}^{\mathbb{N}} \to \mathbb{R}$ can written as an increasing limit of simple functions. Applying the monotone convergence theorem, combined with the *third step*, completes the proof for arbitrary $h$.

**Lemma 1.** *Let $h : \mathsf{X} \to \mathbb{R}$ be measurable, for $k \in \mathbb{N}$, it holds that*

$$\mathbb{E}_\nu[h(X_{k+1}) | \sigma(X_{0:k})] = \mathbb{E}_{X_k}[h(X_1)] = \int_{\mathsf{X}} h(y) P(X_k, \mathrm{d}y) = Ph(X_k)\,. \tag{5}$$

*Proof.* Notice that Equation (4) immediatel holds for $h = \mathbb{1}_A$ by the definition of a Markov chain. Furthermore, any bounded or non-negative measurable function $h$ can be written as a simple function. In other words, there is a sequence of real numbers $(\alpha_k)$ and a sequence of measurable sets $(A_k)$ such that:

$$\sum_{i=0}^{n} \alpha_i \mathbb{1}_{A_i} \uparrow_i h\,.$$

Thus, Equation (5) follows from the monotone convergence theorem. □

**Lemma 2.** *For any $k, j \in \mathbb{N}$ and any bounded or non-negative measurable functions $h_0, \ldots, h_j$ : $\mathsf{X} \to \mathbb{R}$, it holds that:*

$$\mathbb{E}\left[\prod_{i=0}^{j} h_i(X_{k+i}) | \mathcal{F}_k\right] = h_0(X_k) \int_{\mathsf{X}^j} P(X_k, \mathrm{d}x_1) \prod_{i=1}^{j} h_i(x_j) \prod_{i=1}^{j} P(x_{i-1}, \mathrm{d}x_i) = \mathbb{E}_{X_k}\left[\prod_{i=0}^{j} h_i(X_i)\right],$$

*where $\mathcal{F}_k = \sigma(X_{0:k})$*

*Proof.* Fixing $k \in \mathbb{N}$, we will prove this lemma by induction on $j \in \mathbb{N}$. For $j = 0$, the result is immediate. Thus, assume that the result holds for some $j \geqslant 0$. Write down:

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{j+1} h_i(X_{k+i}) | \mathcal{F}_k\right] &= \mathbb{E}\left[\mathbb{E}[h_{j+1}(X_{k+j+1}) | \mathcal{F}_{k+j}] \prod_{i=0}^{j} h_i(X_{k+i})\right] \\ &= \mathbb{E}\left[Ph_{j+1}(X_{k+j}) \prod_{i=0}^{j} h_i(X_{k+i})\right],\end{aligned}$$

where the last equality comes from Lemma 1. Notice that in the last term, the expression under the expectation can be rewritten as $\prod_{i=0}^{j} \tilde{h}_i(X_{k+i})$, where $\tilde{h}_j = h_j P h_{j+1}$ and $\tilde{h}_i = h_i$ for $i < j$. Thus, we can apply the induction property and obtain:

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{j+1} h_i(X_{k+i}) | \mathcal{F}_k\right] &= \mathbb{E}_{X_k}\left[\prod_{i=0}^{j} \tilde{h}_i(X_i)\right] \\ &= \mathbb{E}_{X_k}\left[\prod_{i=0}^{j+1} h_i(X_i)\right]\end{aligned}$$

□

Finally, the sets $A \in \mathcal{X}^{\otimes \mathbb{N}}$ of the form $A = A_0 \times A_1 \times \ldots A_n \times \mathsf{X}^{\mathbb{N}}$ (the cylinders) form a $\pi$-system that generate $\mathcal{X}^{\otimes \mathbb{N}}$. By Lemma 2, the Markov's property is satisfied for the functions $h = \mathbb{1}_A$, where $A$ is a cylinder. Furthermore, the set of $A$ such that the Markov's property is satisfied is a $\lambda$-system. Thus, applying the $\pi$-$\lambda$ theorem, we have that the Markov's property is satisfied for $\mathbb{1}_A$, with $A$ any element of $\mathcal{X}^{\otimes \mathbb{N}}$. Thus, it is also satisfied for simple functions [1]. Finally, since any measurable function $h : \mathcal{X}^{\otimes \mathbb{N}} \to \mathbb{R}$ can be written as an increasing limit of simple functions, applying the monotone convergence theorem we obtain the Markov's property for all measurable $h : \mathsf{X}^{\mathbb{N}} \to \mathbb{R}$.

$$\begin{aligned}\mathbb{P}(X_{0:n} \in A) = \mathbb{E}[\mathbb{1}_A(X_{0:n})] = \mathbb{E}\left[\prod_{i=0}^{n} \mathbb{1}_{A_i}(X_i)\right] &= \mathbb{E}\left[\mathbb{E}[\mathbb{1}_{A_n}(X_n)|\mathscr{F}_{n-1}] \prod_{i=0}^{n-1} \mathbb{1}_{A_i}(X_i)\right] \\ &= \mathbb{E}\left[P(X_{n-1}, A_n) \prod_{i=0}^{n-1} \mathbb{1}_{A_i}(X_i)\right]\end{aligned}$$

and we apply the induction.

Thus the equality of laws hold for any cylinder, which are a $\pi$-system. If two measures agree on a $\pi$-system, then they agree on the whole space. □

[1] $h$ is a simple function if it can be written as a linear combination of indicators: $\sum_{i=0}^{k} \alpha_k \mathbb{1}_{A_k}$, where $\alpha_i$ is some real number.