SF3953: Markov Chains and Processes **Spring 2017**

Lecture 1: Basic Definitions

Lecturer: Jimmy Olsson *February 16*

Goals of this lecture

- To introduce transition kernels and operations on the same.
- To introduce homogeneous Markov chains.

Stochastic processes

Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a probability space, $(\mathsf{X}, \mathcal{X})$ a measurable space, and T a set. We recall the following definitions.

Definition 1.1 (stochastic process)**.** *A family of X-valued random variables indexed by T is called an X-valued* stochastic process indexed by T.

In this course we consider only the cases $T = \mathbb{N}$ and $T = \mathbb{Z}$.

Definition 1.2 (filtration)**.** *A* filtration *of a measurable space $(\Omega, \mathcal{F})$ is an increasing sequence $\{\mathcal{F}_k : k \in \mathbb{N}\}$ of sub-σ-fields of $\mathcal{F}$.*

Definition 1.3 (filtered probability space)**.** *A* filtered probability space $(\Omega, \{\mathcal{F}_k : k \in \mathbb{N}\}, \mathcal{F})$ *is a probability space endowed with a filtration.*

Definition 1.4. *A stochastic process $\{X_k : k \in \mathbb{N}\}$ is said to be* adapted *to the filtration $\{\mathcal{F}_k : k \in \mathbb{N}\}$ if for each $k \in \mathbb{N}$, X_k is $\mathcal{F}_k$-measurable. (Notation: $\{(X_k, \mathcal{F}_k) : k \in \mathbb{N}\}$.)*

Definition 1.5. *The* natural filtration *of a stochastic process $\{X_k : k \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathbb{P}, \mathcal{F})$ is the filtration $\{\mathcal{F}_k^X : k \in \mathbb{N}\}$ defined by*

$$\mathcal{F}_k^X = \sigma(X_j : j \in \mathbb{N}, j \leq k), \quad k \in \mathbb{N}.$$

Kernels

In the following, let $\mathsf{F}(\mathcal{X})$, $\mathsf{F}_+(\mathcal{X})$, and $\mathsf{F}_\mathrm{b}(\mathcal{X})$ denote the sets of measurable functions, non-negative measurable functions and bounded measurable functions on $(\mathsf{X}, \mathcal{X})$, respectively. In addition, $\mathsf{M}_+(\mathcal{X})$ and $\mathsf{M}_1(\mathcal{X})$ denote the sets of measures and probability measures on $(\mathsf{X}, \mathcal{X})$, respectively.

Definition 1.6 (kernel). *Let* $(\mathsf{X}, \mathcal{X})$ *and* $(\mathsf{Y}, \mathcal{Y})$ *be measurable spaces.* A kernel *is a mapping* $K : \mathsf{X} \times \mathcal{Y} \to \bar{\mathbb{R}}_+ = [0, \infty]$ *satisfying the following conditions:*

(i) *for every* $x \in \mathsf{X}$*, the mapping* $\mathcal{Y} \ni A \mapsto K(x, A)$ *is a measure on* $\mathcal{Y}$*,*

(ii) *for every* $A \in \mathcal{Y}$*, the mapping* $\mathsf{X} \ni x \mapsto K(x, A)$ *is a measurable function from* $(\mathsf{X}, \mathcal{X})$ *to* $\bar{\mathbb{R}}_+$*.*[1]

In addition, the kernel K *is said to be*

- bounded *if* $\sup_{x \in \mathsf{X}} K(x, \mathsf{Y}) < \infty$*,*
- Markovian *(or, a* Markov kernel*) if* $K(x, \mathsf{Y}) = 1$ *for all* $x \in \mathsf{X}$*.*

Example 1.7 (kernel on a discrete state space). *Assume that* X *and* Y *are countable sets and denote by* $\wp(\mathsf{Y})$ *the power set of* Y*. In this case, a kernel* K *on* $\mathsf{X} \times \wp(\mathsf{Y})$ *is specified by a (possibly doubly infinite)* transition matrix $\{k(x, y) : (x, y) \in \mathsf{X} \times \mathsf{Y}\}$*. More specifically, define*

$$K : \mathsf{X} \times \wp(\mathsf{Y}) \ni (x, A) \mapsto \sum_{y \in A} k(x, y).$$

Then for each $x \in \mathsf{X}$*, the row* $\{k(x, y) : y \in \mathsf{Y}\}$ *defines a measure on* $\wp(\mathsf{Y})$*. The kernel* K *is Markovian if each row in the matrix sums to one.*

Example 1.8 (kernel density). *Let* $\lambda \in \mathsf{M}_+(\mathcal{Y})$ *be* σ*-finite and* $k : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}_+$ *a measurable function. Then*

$$K : \mathsf{X} \times \mathcal{Y} \ni (x, A) \mapsto \int_A k(x, y)\, \lambda(\mathrm{d}y)$$

is a kernel. (This follows from the Tonelli-Fubini theorem, which holds for σ*-finite measures.) The kernel* K *is Markovian if* $\int k(x, y)\, \lambda(\mathrm{d}y) = 1$ *for all* $x \in \mathsf{X}$*.*

Let K be a kernel on $\mathsf{X} \times \mathcal{Y}$ and $f \in \mathsf{F}_+(\mathcal{Y})$. Then we define

$$Kf : \mathsf{X} \ni x \mapsto \int f(y)\, K(x, \mathrm{d}y).$$

In addition, we set $Kf = Kf^+ - Kf^-$ for all functions $f \in \mathsf{F}(\mathcal{Y})$ such that $Kf^+$and Kf^- are not both infinity.[2]

Exercise 1.9. *Show that for all* $f \in \mathsf{F}_+(\mathcal{Y})$*,* $Kf \in \mathsf{F}_+(\mathcal{X})$*. (Hint: first, establish the claim for simple functions.) Moreover, show that if* K *is Markovian, then for all* $f \in \mathsf{F}_\mathrm{b}(\mathcal{Y})$*,* $Kf \in \mathsf{F}_\mathrm{b}(\mathcal{X})$*.*

[1] Recall that $\mathcal{B}(\bar{\mathbb{R}}) = \sigma(\{[-\infty, x] : x \in \mathbb{R}\})$. In addition, $\bar{\mathbb{R}}_+$ is furnished with the σ-field $\{B \cap \bar{\mathbb{R}}_+ : B \in \mathcal{B}(\bar{\mathbb{R}})\}$.

[2] Recall that $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$ are both measurable if f is so.

In addition, kernels operates on measures. Let $\mu \in \mathsf{M}_+(\mathcal{X})$ and define the mapping

$$\mu K : \mathcal{Y} \ni A \mapsto \int K(x, A)\, \mu(\mathrm{d}x).$$

Exercise 1.10. *Show that for all $\mu \in \mathsf{M}_+(\mathcal{X})$, $\mu K \in \mathsf{M}_+(\mathcal{Y})$.*

Moreover, we define the *composition* of two kernels. Let K be as above, let $(\mathsf{Z}, \mathcal{Z})$ be a third measurable space, and let L be a kernel on $\mathsf{Y} \times \mathcal{Z}$. Now, define the mapping

$$KL : \mathsf{X} \times \mathcal{Z} \ni (x, A) \mapsto \int L(y, A)\, K(x, \mathrm{d}y).$$

Exercise 1.11. *Show that KL is a kernel on $\mathsf{X} \times \mathcal{Z}$.*

For all $x \in \mathsf{X}$ and $f \in \mathsf{F}_+(\mathcal{Z})$, it holds that $KLf(x) = K(Lf)(x)$. (Prove the claim first for simple functions and then generalise it to general nonnegative measurable functions by using twice the monotone convergence theorem.)

We may define iteratively the nth power of a kernel K on $\mathsf{X} \times \mathcal{X}$ by letting $K^0 : \mathsf{X} \times \mathcal{X} \ni (x, A) \mapsto \delta_x(A)$ and $K^n = K^{n-1}K$ for $n \in \mathbb{N}^*$. For integers $(n, m) \in \mathbb{N}^2$, this implies immediately the *Chapman-Kolmogorov equation*

$$K^{n+m}f = K^{n+m-1}Kf = K^{n+m-2}K(Kf) = K^{n+m-2}K^2 f = \ldots = K^n K^m f$$

for all $f \in \mathsf{F}_{\mathrm{b}}(\mathcal{X})$, implying that $K^{n+m} = K^n K^m$. Finally, we define *tensor products* of kernels. Let K and L be as above and define

$$K \otimes L : \mathsf{X} \times (\mathcal{Y} \otimes \mathcal{Z}) \ni (x, A) \mapsto \int \left(\int \mathbb{1}_A(y, z)\, L(y, \mathrm{d}z) \right) K(x, \mathrm{d}y).$$

One can prove that $K \otimes L$ is a kernel on $\mathsf{X} \times (\mathcal{Y} \otimes \mathcal{Z})$. (Indeed, let $\mathcal{H}$ be the set of bounded functions f such that $f(y, \cdot) \in \mathsf{F}(\mathcal{Z})$ for all $y \in \mathsf{Y}$ and $\int f^{\pm}(\cdot, z)\, L(\cdot, \mathrm{d}z) \in \mathsf{F}(\mathcal{Y})$, and prove, using the functional monotone class theorem, that $\mathsf{F}_{\mathrm{b}}(\mathcal{Y} \otimes \mathcal{Z}) \subset \mathcal{H}$. Then, prove that $K \otimes L(x, \cdot) \in \mathsf{M}_+(\mathcal{Y} \otimes \mathcal{Z})$ for all $x \in \mathsf{X}$ by using twice the monotone convergence theorem.)

Exercise 1.12. *Show that*

- *if K and L are both bounded, so is $K \otimes L$.*
- *if K and L are both Markovian, so is $K \otimes L$.*

Also for the tensor product we define the nth power; more specifically, set $K^{\otimes 1} = K$ and define iteratively $K^{\otimes n} = K \otimes K^{\otimes (n-1)}$ for $n \in \mathbb{N}^* \setminus \{1\}$, i.e., $K^{\otimes n}$ is a kernel on $\mathsf{X} \times \mathcal{X}^{\otimes n}$. Carrying through the recursion yields the alternative expression

$$K^{\otimes n} : \mathsf{X} \times \mathcal{X}^{\otimes n} \ni (x_0, A) \mapsto \int \cdots \int \mathbb{1}_A(x_1^n) \prod_{k=0}^{n-1} K(x_k, \mathrm{d}x_{k+1}),$$

where we used the vector notation $x_m^n = (x_m, \dots, x_n)$ for $(m,n) \in \mathbb{Z}^2$ with $m \leq n$.

We also define the tensor product of a *measure* $\mu \in \mathsf{M}_+(\mathcal{X})$ and a kernel K on $\mathsf{X} \times \mathcal{Y}$ as the *measure*

$$\mu \otimes K : \mathcal{X} \otimes \mathcal{Y} \ni A \mapsto \int \left(\int \mathbb{1}_A(x,y)\, K(x, \mathrm{d}y) \right) \mu(\mathrm{d}x).$$

Homogeneous Markov chains

Definition 1.13 (homogeneous Markov chain)**.** *Let* $(\mathsf{X}, \mathcal{X})$ *be a measurable space and* P *a Markov kernel on* $\mathsf{X} \times \mathcal{X}$*. Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_k : k \in \mathbb{N}\}, \mathbb{P})$ *be a filtered probability space. An adapted stochastic process* $\{(X_k, \mathcal{F}_k) : k \in \mathbb{N}\}$ *is called a* homogeneous Markov chain with kernel P *if for all* $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$[3] *and* $k \in \mathbb{N}$*,*

$$\mathbb{E}\left[h(X_{k+1}) \mid \mathcal{F}_k\right] = Ph(X_k), \quad \mathbb{P}\text{-}a.s.$$

The distribution of X_0 *is called the* initial distribution*.*

A homogeneous Markov chain is always a homogeneous Markov chain with respect to its natural filtration (why?). Thus, we will always assume in the following that $\{\mathcal{F}_k : k \in \mathbb{N}\}$ is the natural filtration of $\{X_k : k \in \mathbb{N}\}$.

Theorem 1.14. *Let* P *be a Markov kernel on* $\mathsf{X} \times \mathcal{X}$ *and* μ *a probability measure on* $(\mathsf{X}, \mathcal{X})$*. An* X*-valued stochastic process* $\{X_k : k \in \mathbb{N}\}$ *is a homogeneous Markov chain with kernel* P *and initial distribution* μ *if and only if for all* $k \in \mathbb{N}$*, the distribution* X_0^k *is* $\mu \otimes P^{\otimes k}$*.*

Proof. We assume that $\{X_k : k \in \mathbb{N}\}$ is a homogeneous Markov chain with kernel P. Fix $k \in \mathbb{N}$ and let $\mathcal{H}_k$ be the vector space of functions $h \in \mathsf{F}_\mathrm{b}(\mathcal{X}^{\otimes(k+1)})$ such that

$$\mathbb{E}\left[h(X_0^k)\right] = \mu \otimes P^{\otimes k} h. \tag{1.15}$$

We show that $\mathsf{F}_\mathrm{b}(\mathcal{X}^{\otimes(k+1)}) \subset \mathcal{H}_k$ using Theorem 1.29. We proceed by induction and assume that $\mathsf{F}_\mathrm{b}(\mathcal{X}^{\otimes k}) \subset \mathcal{H}_{k-1}$. Let $\mathcal{C} = \{A_0 \times \cdots \times A_k : A_j \in \mathcal{X}, 0 \leq j \leq k\}$. Note that $\mathcal{C}$ is closed under finite intersections. To prove that $\mathbb{1}_A \in \mathcal{H}_k$ for all $A \in \mathcal{C}$, write, using

[3] Note that by monotone convergence, the same property will hold also for all $h \in \mathsf{F}_+(\mathcal{X}^{\otimes(n+1)})$.

Definition 1.13 and the induction hypothesis,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_A(X_0^k)\right] &= \mathbb{E}\left[\prod_{i=0}^{k}\mathbb{1}_{A_i}(X_i)\right] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1}\mathbb{1}_{A_i}(X_i)\mathbb{E}\left[\mathbb{1}_{A_k}(X_k)\mid\mathcal{F}_{k-1}\right]\right] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1}\mathbb{1}_{A_i}(X_i)P\mathbb{1}_{A_k}(X_{k-1})\right] && (1.16) \\
&= \mu\otimes P^{\otimes(k-1)}\left(\prod_{i=0}^{k-1}\mathbb{1}_{A_i}\times P\mathbb{1}_{A_k}\right) && (1.17) \\
&= \mu\otimes P^{\otimes k}\left(\prod_{i=0}^{k}\mathbb{1}_{A_i}\right). && (1.18)
\end{aligned}$$

In addition, let $\{h_n\}_{n\in\mathbb{N}^*}$ be a sequence of increasing functions in $\mathcal{H}_k$ and let $h=\lim_{n\to\infty}h_n$. Then by using twice the monotone convergence theorem we conclude that $h\in\mathcal{H}_k$. (Indeed, proceed like

$$\mathbb{E}\left[h(X_0^k)\right]=\lim_{n\to\infty}\mathbb{E}\left[h_n(X_0^k)\right]=\lim_{n\to\infty}\mu\otimes P^{\otimes k}h_n=\mu\otimes P^{\otimes k}h,$$

where we used, in the last step, that $\mu\otimes P^{\otimes k}$ is a measure.) As the induction hypothesis is trivially true for $k=0$, necessity follows by Theorem 1.29.

Conversely, assume that the identity (1.15) holds for all $k\in\mathbb{N}$ and $h\in\mathsf{F}_\mathrm{b}(\mathcal{X}^{\otimes(k+1)})$. Pick $k\in\mathbb{N}$ arbitrarily and show that for all $h\in\mathsf{F}_\mathrm{b}(\mathcal{X})$ and $\mathcal{F}_{k-1}$-measurable (with $\mathcal{F}_{k-1}=\sigma(X_j:j\le k-1)$) bounded Y,

$$\mathbb{E}\left[Y(h(X_k)-Ph(X_{k-1}))\right]=0\Leftrightarrow\mathbb{E}\left[Yh(X_k)\right]=\mathbb{E}\left[YPh(X_{k-1})\right], \tag{1.19}$$

(implying that $\mathbb{E}[h(X_k)\mid\mathcal{F}_{k-1}]=Ph(X_{k-1})$, $\mathbb{P}$-a.s.).

Exercise 1.20. *Establish* (1.19).

□

Invariant measures and stationarity

Definition 1.21 (invariant/sub-invariant measure)**.** *Let P be a Markov kernel on $\mathsf{X}\times\mathcal{X}$. A non-zero σ-finite measure $\mu\in\mathsf{M}_+(\mathcal{X})$ is said to be invariant (or sub-invariant) with respect to P if $\mu P=\mu$ (or $\mu P\le\mu$).*

Note that we assume an invariant measure to be σ-finite. (Consider the case $\mathsf{X} = \mathbb{R}$ and $P(x, A) = \delta_{x+1}(A)$ for all $A \in \mathcal{X} = \mathcal{B}(\mathbb{R})$. If μ is the counting measure, then $\mu P = \mu$ for all $A \in \mathcal{X}$. However, μ is not σ-finite.)

Recall that a stochastic process is $\{X_k : k \in \mathbb{N}\}$ on $(\Omega, \mathcal{X}, \mathbb{P})$ is *stationary* if for all $(k, p) \in \mathbb{N}^2$, the law of X_k^{k+p} does not depend on k.

Theorem 1.22. *A Markov chain $\{(X_k, \mathcal{F}_k) : k \in \mathbb{N}\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_k : k \in \mathbb{N}\}, \mathbb{P})$ with kernel P on $(\mathsf{X}, \mathcal{X})$ is a stationary process if and only if its initial distribution is invariant with respect to P.*

Exercise 1.23. *Prove Theorem 1.22.*

Definition 1.24 (reversibility)**.** *Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. A σ-finite measure ξ on $\mathcal{X}$ is said to be* reversible with respect to P *if the measure $\xi \otimes P$ on $\mathcal{X}^{\otimes 2}$ is symmetric, i.e. for all $h \in \mathsf{F}_\mathrm{b}(\mathcal{X}^{\otimes 2})$,*

$$\iint h(x, x')\, \xi(\mathrm{d}x)\, P(x, \mathrm{d}x') = \iint h(x', x)\, \xi(\mathrm{d}x)\, P(x, \mathrm{d}x'). \tag{1.25}$$

Note that if $\{X_k : k \in \mathbb{N}\}$ is a Markov chain with kernel P and initial distribution ξ, then the reversibility condition (1.25) means that $\mathbb{E}_\xi[h(X_0, X_1)] = \mathbb{E}_\xi[h(X_1, X_0)]$.

Exercise 1.26. *Show that if ξ is reversible with respect to P, then ξ is invariant with respect to P.*

Example 1.27 (the Metropolis-Hastings algorithm)**.** Markov chain Monte Carlo *(MCMC) is a general method for simulating from distributions known up to a constant of proportionality. Let ν be a (σ-finite) measure on some state space $(\mathsf{X}, \mathcal{X})$ and let $h \in \mathsf{F}_+(\mathcal{X})$ such that $0 < \int h(x)\, \nu(\mathrm{d}x) < \infty$. Assume for simplicity that h is positive and define the distribution*

$$\pi : \mathcal{X} \ni A \mapsto \frac{\int_A h(x)\, \nu(\mathrm{d}x)}{\int h(x)\, \nu(\mathrm{d}x)}.$$

The Metropolis-Hastings *(MH)* algorithm *generates a Markov chain $\{X_k : k \in \mathbb{N}\}$ with invariant measure π as follows. Let $Q : \mathsf{X} \times \mathcal{X} \to [0, 1]$ be a* proposal kernel *with positive kernel density $q \in \mathsf{F}_+(\mathcal{X}^2)$ with respect to ν, i.e., for all $(x, A) \in \mathsf{X} \times \mathcal{X}$, $Q(x, A) = \int_A q(x, y)\, \nu(\mathrm{d}y)$. Given X_k, a candidate X_{k+1}^* is sampled from $Q(X_k, \cdot)$. With probability $\alpha(X_k, X_{k+1}^*)$, where*

$$\alpha : \mathsf{X}^2 \ni (x, y) = 1 \wedge \frac{h(y) q(y, x)}{h(x) q(x, y)},$$

this proposal is accepted and the chain moves to $X_{k+1} = X_{k+1}^*$; otherwise, the candidate is rejected and the chain remains at $X_{k+1} = X_k$. Consequently, $\{X_k : k \in \mathbb{N}\}$ is a Markov chain with kernel

$$P : \mathsf{X} \times \mathcal{X} \ni (x, A) \mapsto \int_A \alpha(x, y) q(x, y)\, \nu(\mathrm{d}y) + \rho(x) \delta_x(A),$$

with

$$\rho : \mathsf{X} \ni x \mapsto \int \{1 - \alpha(x, y)\} q(x, y) \, \nu(\mathrm{d}y)$$

being the probability of rejection.

Exercise 1.28. *Show that the target π is reversible with respect to the MH kernel P.*

A The functional monotone class theorem

Theorem 1.29. *Let $\mathcal{H}$ be a vector space of bounded functions on Ω and $\mathcal{C}$ a class of subsets of Ω stable by finite intersection. Assume that $\mathcal{H}$ satisfies*

(i) $\mathbb{1}_\Omega \in \mathcal{H}$ *and for all* $A \in \mathcal{C}$, $\mathbb{1}_A \in \mathcal{H}$.

(ii) If $\{f_n, n \in \mathbb{N}\}$ *is an increasing sequence of functions of* $\mathcal{H}$ *such that* $\sup_{n \in \mathbb{N}} f_n = \lim_{n \to \infty} f_n = f$ *is bounded, then* $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all the bounded $\sigma(\mathcal{C})$-measurable functions.

References

[1] J. Jacod and P. Protter. *Probability Essentials.* Springer, 2000.