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

Lecture 5: Minorisation and Splitting

Lecturer: Jimmy Olsson — *March 16*

Goals of this lecture

- Introduce *small*—"atom-like"—sets.
- Show how a general phi-irreducible Markov chain admitting a small set can be embedded into a larger, atomic *split chain*.

Small sets

In the following, let $(\mathsf{X}, \mathcal{X})$ be a measurable space.

Definition 5.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$, $\nu \in \mathsf{M}_1(\mathcal{X})$, $m \in \mathbb{N}^*$, and $\varepsilon \in (0, 1]$. A set $C \in \mathcal{X}$ is called an* (m, ε, ν)-small set *for P, or simply a* small set, *if $\nu(C) > 0$ and for all $x \in C$ and $A \in \mathcal{X}$,*

$$P^m(x, A) \geq \varepsilon\nu(A).$$

Note that

- if $\varepsilon = 1$, then C is an atom for the kernel P^m.
- for all $x \in \mathsf{X}$, $\{x\}$ is a small—but generally not accessible—set.
- if the state space is countable and P is irreducible (according to the definition in Example 3.4), then every finite set is small.

Proposition 5.2. *Let $C \in \mathcal{X}$ be an accessible (m, ε, ν)-small set for the transition kernel P on $\mathsf{X} \times \mathcal{X}$. Then ν is an irreducibility measure.*

Exercise 5.3. *Prove Proposition 5.2.*

Example 5.4 (the Metropolis-Hastings algorithm (cont.))**.** *Reconsider the Metropolis-Hastings algorithm discussed in Exemple 1.27 and assume that $\mathsf{X} = \mathbb{R}^d$ and that ν is Lebesgue measure on $\mathcal{X} = \mathcal{B}(\mathbb{R}^d)$. Then, if the target and proposal densities h and q are both continuous and positive, then every compact set $C \in \mathcal{X}$ with $\nu(C) > 0$ is small. Indeed,* *let* $\sigma_- = \inf_{(x,x')\in C^2} q(x,x')$*,* $\sigma^+ = \sup_{x\in C} h(x)/\nu h$*, and define for all* $x \in C$ *and* $B \in \mathcal{X}$ *such that* $B \subset C$*,*

$$R_x(B) = \left\{ y \in B : \frac{h(y)q(y,x)}{h(x)q(x,y)} < 1 \right\}.$$

Then for all $x \in C$ *and* B *as above,*

$$\begin{aligned} P(x,B) &\geq \int_B \alpha(x,y)q(x,y)\,\nu(\mathrm{d}y) \\ &\geq \int_{R_x(B)} \frac{h(y)q(y,x)}{h(x)}\,\nu(\mathrm{d}y) + \int_{B\setminus R_x(B)} q(x,y)\,\nu(\mathrm{d}y) \\ &\geq \frac{\sigma_-}{\sigma_+}\int_{R_x(B)} h(y)\,\nu(\mathrm{d}y) + \frac{\sigma_-}{\sigma_+}\int_{B\setminus R_x(B)} h(y)\,\nu(\mathrm{d}y) \\ &\geq \varepsilon\pi_C(B), \end{aligned}$$

where we have set $\varepsilon = (\sigma_-/\sigma_+)\int_C h(y')\,\nu(\mathrm{d}y')$ *and defined the probability measure*

$$\pi_C : \mathcal{X} \ni A \mapsto \frac{\int_{A\cap C} h(y)\,\nu(\mathrm{d}y)}{\int_C h(y')\,\nu(\mathrm{d}y')}.$$

Hence, for all $A \in \mathcal{X}$ *and* $x \in C$*,*

$$P(x,A) \geq P(x, A\cap C) \geq \varepsilon\pi_C(A\cap C) = \varepsilon\pi_C(A),$$

which shows that C *is* $(1,\varepsilon,\pi_C)$*-small for* P*.*

Splitting

Suppose in the following that the chain admits a $(1,\varepsilon,\nu)$-small set $C \in \mathcal{X}$. On the basis of this set, define the *residual kernel*

$$R : \mathsf{X}\times\mathcal{X} \ni (x,A) \mapsto \frac{P(x,A) - \varepsilon\mathbb{1}_C(x)\nu(A)}{1-\varepsilon\mathbb{1}_C(x)}. \tag{5.5}$$

Note that the 1-smallness of C guarantees R to be Markovian.

We first extend the state space by letting $\check{\mathsf{X}} = \mathsf{X}\times\{0,1\}$ and $\check{\mathcal{X}} = \mathcal{X}\otimes\wp(\{0,1\})$. With each $\mu \in \mathsf{M}_+(\mathcal{X})$ we associate a measure

$$\mu_C : \mathcal{X} \ni A \mapsto \mu(A\cap C)$$

in $\mathsf{M}_+(\mathcal{X})$ as well as a *split measure*

$$\mu^* : \check{\mathcal{X}} \ni \check{A} \mapsto (1-\varepsilon)\int \mathbb{1}_{\check{A}}(x,0)\,\mu_C(\mathrm{d}x) + \int \mathbb{1}_{\check{A}}(x,0)\,\mu_{C^\complement}(\mathrm{d}x) + \varepsilon\int \mathbb{1}_{\check{A}}(x,1)\,\mu_C(\mathrm{d}x) \tag{5.6}$$

in $\mathsf{M}_+(\check{\mathcal{X}})$. Trivially, if $\mu \in \mathsf{M}_1(\mathcal{X})$, then $\mu^* \in \mathsf{M}_1(\check{\mathcal{X}})$. Note that for all $A \in \mathcal{X}$,

$$\mu^*(A \times \{i\}) = \begin{cases} (1-\varepsilon)\mu_C(A) + \mu_{C^\complement}(A) & \text{if } i = 0, \\ \varepsilon \mu_C(A) & \text{if } i = 1, \end{cases}$$

implying especially that $\mu^*(A \times \{0,1\}) = \mu(A)$. Moreover, note that a split measure never assigns any mass to the set $C^\complement \times \{1\}$. Using the definition (5.6), we may split, naturally, also each *kernel* K on $\mathsf{X} \times \mathcal{X}$ according to

$$K^* : \mathsf{X} \times \check{\mathcal{X}} \ni (x, \check{A}) \mapsto [K(x, \cdot)]^*(\check{A}).$$

We now extend P to a Markov kernel $\check{P}$ on $(\check{\mathsf{X}}, \check{\mathcal{X}})$ as follows.

$$\check{P} : \check{\mathsf{X}} \times \check{\mathcal{X}} \ni ((x,i), \check{A}) \mapsto \begin{cases} R^*(x, \check{A}) & \text{if } i = 0, \\ \nu^*(\check{A}) & \text{if } i = 1. \end{cases} \tag{5.7}$$

Note that $\check{\boldsymbol{\alpha}} = \mathsf{X} \times \{1\}$ is an atom for $\check{P}$ with respect to ν^*. For all $\check{\mu} \in \mathsf{M}_1(\check{\mathcal{X}})$, Theorem 2.2 provides a probability measure $\check{\mathbb{P}}_{\check{\mu}}$ on the canonical space $(\check{\mathsf{X}}^{\mathbb{N}}, \check{\mathcal{X}}^{\otimes \mathbb{N}})$ under which the coordinate process, denoted by $\{\check{X}_k : k \in \mathbb{N}\}$, with, for each $k \in \mathbb{N}$, $\check{X}_k = (X_k, D_k)$, and referred to as the *split chain*, is a Markov chain with kernel $\check{P}$ and initial distribution $\check{\mu}$. We denote by $\{\check{\mathcal{F}}_k : k \in \mathbb{N}\}$ the natural filtration of $\{\check{X}_k : k \in \mathbb{N}\}$. In addition, we denote by $\{\check{\mathcal{F}}_k^X : k \in \mathbb{N}\}$ the natural filtration of the marginal process $\{X_k : k \in \mathbb{N}\}$ and set $\check{\mathcal{F}}_\infty^X = \sigma(X_k : k \in \mathbb{N})$.

The following exercise provides conceptual understanding of the dynamics of the split chain.

Exercise 5.8. *Let $\check{x} = (x, d) \in \check{\mathsf{X}}$ and check that*

– *if $x \notin C$ and $d = 0$, then $\check{\mathbb{P}}_{\check{x}}(X_1 \in \cdot) = P(x, \cdot)$.*

– *if $x \in C$ and $d = 0$, then $\check{\mathbb{P}}_{\check{x}}(X_1 \in \cdot) = R(x, \cdot)$.*[1]

– *if $x \in C$ and $d = 1$, then $\check{\mathbb{P}}_{\check{x}}(X_1 \in \cdot) = \nu$.*

– *$\check{\mathbb{P}}_{\check{x}}$-a.s.,*

$$\check{\mathbb{P}}_{\check{x}}\left(D_k = i \mid \check{\mathcal{F}}_k^X\right) = \begin{cases} (1-\varepsilon)\mathbb{1}_C(X_k) + \mathbb{1}_{C^\complement}(X_k) & \text{if } i = 0, \\ \varepsilon \mathbb{1}_C(X_k) & \text{if } i = 1. \end{cases}$$

Exercise 5.9. *Show that for all $\mu \in \mathsf{M}_1(\mathcal{X})$,*

$$\mu^* \check{P} = (\mu P)^*.$$

[1] Note that for all $x \notin C$, $P(x, \cdot) = R(x, \cdot)$, and the first two cases could hence be summarised using only the residual kernel. Nevertheless, we have chosen to write explicitly P in the first case for clarity.

The following is the main result of this lecture.

Theorem 5.10. *Let P be a phi-irreducible transition kernel on $\mathsf{X} \times \mathcal{X}$, let $C \in \mathcal{X}$ be an accessible $(1, \varepsilon, \nu)$-small set for P, and let $\mu \in \mathsf{M}_1(\mathcal{X})$. Then for all $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$ and all $k \in \mathbb{N}$,*

$$\check{\mathbb{E}}_{\mu^*}\left[h(X_{k+1}) \mid \check{\mathcal{F}}_k^X\right] = Ph(X_k), \quad \check{\mathbb{P}}_{\mu^*}\text{-a.s.} \tag{5.11}$$

In addition, $\check{\mathbb{P}}_{\mu^*}(X_0 \in \cdot) = \mu$.

Consequently, under $\check{\mathbb{P}}_{\mu^*}$, the marginal process $\{X_k : k \in \mathbb{N}\}$ is a Markov chain with respect to its natural filtration with kernel P and initial distribution μ. Now, let $\mathbb{P}_\mu$ be the unique law on $(\mathsf{X}^\mathbb{N}, \mathcal{X}^{\otimes \mathbb{N}})$ under which the coordinate process, which we, by abuse of notation, denote $\{X_k : k \in \mathbb{N}\}$ as well, is a Markov chain with kernel P and initial distribution μ. Let $\{\mathcal{F}_k : k \in \mathbb{N}\}$ denote its natural filtration and set $\mathcal{F}_\infty = \sigma(X_k : k \in \mathbb{N})$.

Since an $\check{\mathcal{F}}_\infty^X$-measurable random variable Y depends only on the X-component of each path, we may identify the same with an $\mathcal{F}_\infty$-measurable random variable denoted, by abuse of notation, by the same symbol, Y. Theorem 5.10 implies the following result, which will be instrumental in the coming developments.

Exercise 5.12. *Show that for all bounded $\check{\mathcal{F}}_\infty^X$-measurable Y,*

$$\check{\mathbb{E}}_{\mu^*}[Y] = \mathbb{E}_\mu[Y]. \tag{5.13}$$

Proof of Theorem 5.10. Pick $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$. We show, using the functional monotone class theorem, that for all bounded $\check{\mathcal{F}}_k^X$-measurable Z,

$$\check{\mathbb{E}}_{\mu^*}[Zh(X_{k+1})] = \check{\mathbb{E}}_{\mu^*}[ZPh(X_k)]. \tag{5.14}$$

For this purpose, let $\mathcal{H}$ be the vector space of bounded $\check{\mathcal{F}}_k^X$-measurable Z such that (5.14) holds true and let $\mathcal{C} = \{\cap_{j=0}^k X_j^{-1}(B_j) : B_j \in \mathcal{X}, 0 \le j \le k\}$. We pick arbitrarily $A \in \mathcal{C}$ and show that $\mathbb{1}_A = \prod_{j=0}^k \mathbb{1}_{B_j}(X_j)$ belongs to $\mathcal{H}$. Write, using the Markov property and Exercise 5.8,

$$\begin{aligned}
\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^k \mathbb{1}_{B_j}(X_j)h(X_{k+1})\right] &= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^k \mathbb{1}_{B_j}(X_j)\check{\mathbb{E}}_{\check{X}_k}[h(X_1)]\right] \\
&= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^k \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C\times\{0\}}(\check{X}_k)Rh(X_k)\right] && (5.15) \\
&+ \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^k \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C\times\{1\}}(\check{X}_k)\nu h\right] && (5.16) \\
&+ \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^k \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C^\complement\times\{0\}}(\check{X}_k)Ph(X_k)\right]. && (5.17)
\end{aligned}$$

For the first term (5.15), write, using again the Markov property,

$$
\begin{aligned}
&\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C\times\{0\}}(\check{X}_k)Rh(X_k)\right] \\
&= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k-1} \mathbb{1}_{B_j}(X_j)\check{\mathbb{E}}_{\mu^*}\left[\mathbb{1}_{C\cap B_k}(X_k)\mathbb{1}_{\{0\}}(D_k)Rh(X_k) \mid \check{\mathcal{F}}_{k-1}\right]\right] \\
&= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k-1} \mathbb{1}_{B_j}(X_j)\check{\mathbb{E}}_{\check{X}_{k-1}}\left[\mathbb{1}_{C\cap B_k}(X_1)\mathbb{1}_{\{0\}}(D_1)Rh(X_1)\right]\right] \\
&= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k-1} \mathbb{1}_{B_j}(X_j)\check{\mathbb{E}}_{\check{X}_{k-1}}\left[\mathbb{1}_{C\cap B_k}(X_1)\check{\mathbb{P}}_{\check{X}_{k-1}}\left(D_1=0 \mid X_1\right)Rh(X_1)\right]\right]. \qquad (5.18)
\end{aligned}
$$

Now, by Exercise 5.8, for all $\check{x} \in \check{\mathsf{X}}$,

$$
\mathbb{1}_{C\cap B_k}(X_1)\check{\mathbb{P}}_{\check{x}}\left(D_1=0 \mid X_1\right) = \mathbb{1}_{C\cap B_k}(X_1)(1-\varepsilon), \quad \check{\mathbb{P}}_{\check{x}}\text{-a.s.},
$$

and applying the previous identity and the Markov property to (5.18) yields

$$
\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C\times\{0\}}(\check{X}_k)Rh(X_k)\right] = (1-\varepsilon)\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C}(X_k)Rh(X_k)\right]. \qquad (5.19)
$$

The terms (5.16) and (5.17) are treated similarly, yielding

$$
\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C\times\{1\}}(\check{X}_k)\nu h\right] = \varepsilon\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C}(X_k)\nu h\right], \qquad (5.20)
$$

$$
\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C^\complement\times\{0\}}(\check{X}_k)Ph(X_k)\right] = \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C^\complement}(X_k)Ph(X_k)\right]. \qquad (5.21)
$$

Finally, combining (5.19)–(5.21) provides, using the definition of the residual kernel,

$$
\begin{aligned}
\check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)h(X_{k+1})\right] &= \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C}(X_k)\{(1-\varepsilon)Rh(X_k)+\varepsilon\nu h\}\right] \\
&\quad + \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)\mathbb{1}_{C^\complement}(X_k)Ph(X_k)\right] = \check{\mathbb{E}}_{\mu^*}\left[\prod_{j=0}^{k} \mathbb{1}_{B_j}(X_j)Ph(X_k)\right],
\end{aligned}
$$

showing that $\mathbb{1}_A \in \mathcal{H}$.

It remains to show that for all increasing sequences $\{Z_n : n \in \mathbb{N}^*\} \subset \mathcal{H}$, also $Z = \lim_{n\to\infty} Z_n$ belongs to $\mathcal{H}$. However, this follows by the monotone convergence theorem, since

$$\check{\mathbb{E}}_{\mu^*}\left[Zh(X_{k+1})\right] = \lim_{n\to\infty} \check{\mathbb{E}}_{\mu^*}\left[Z_n h(X_{k+1})\right] = \lim_{n\to\infty} \check{\mathbb{E}}_{\mu^*}\left[Z_n Ph(X_k)\right] = \check{\mathbb{E}}_{\mu^*}\left[ZPh(X_k)\right].$$

This completes the proof of (5.11).

Finally, since for all $A \in \mathcal{X}$,

$$\check{\mathbb{P}}_{\mu^*}(X_0 \in A) = \mu^*(A \times \{0,1\}) = \mu(A),$$

the last statement follows □

Theorem 5.22. *Under the assumptions of Theorem 5.10, $\check{\boldsymbol{\alpha}} = \mathsf{X} \times \{1\}$ is an accessible atom and ν^* is an irreducibility measure for the split kernel $\check{P}$. More generally, if $B \in \mathcal{X}$ is accessible for P, then $B \times \{0,1\}$ is accessible for the split kernel.*

Proof. We show that $\check{\boldsymbol{\alpha}}$ is accessible. For this purpose, let $x \in \mathsf{X}$ and write, using Exercise 5.8,

$$\begin{aligned}\check{\mathbb{P}}_{(x,1)}(\sigma_{\check{\boldsymbol{\alpha}}} < \infty) \geq \check{\mathbb{P}}_{(x,1)}(X_1 \in C, D_1 = 1)& \\ = \check{\mathbb{E}}_{(x,1)}\left[\check{\mathbb{P}}_{(x,1)}(D_1 = 1 \mid X_1)\mathbb{1}_C(X_1)\right] &= \varepsilon\nu(C) > 0.\end{aligned}$$

In addition, using σ-additivity and Exercise 5.8,

$$\begin{aligned}\check{\mathbb{P}}_{(x,0)}\left(\sigma_{\check{\boldsymbol{\alpha}}} < \infty\right) &= \sum_{\ell=1}^{\infty} \check{\mathbb{P}}_{(x,0)}\left(\sigma_{\check{\boldsymbol{\alpha}}} = \ell\right) \geq \sum_{\ell=1}^{\infty} \check{\mathbb{P}}_{(x,0)}\left(\sigma_{C\times\{0,1\}} = \ell, D_\ell = 1\right) \\ &= \sum_{\ell=1}^{\infty} \check{\mathbb{E}}_{(x,0)}\left[\check{\mathbb{P}}_{(x,0)}\left(D_\ell = 1 \mid \mathcal{F}_\ell^X\right)\mathbb{1}_{\{\sigma_{C\times\{0,1\}} = \ell\}}\right] = \varepsilon \sum_{\ell=1}^{\infty} \check{\mathbb{P}}_{(x,0)}\left(\sigma_{C\times\{0,1\}} = \ell\right) \\ &= \varepsilon \check{\mathbb{P}}_{(x,0)}\left(\sigma_{C\times\{0,1\}} < \infty\right).\end{aligned}$$

It is enough to consider $x \notin C$, in which case $(\delta_x)^* = \delta_{(x,0)}$. Now, using Exercise 5.12,

$$\check{\mathbb{P}}_{(x,0)}(\sigma_{\check{\boldsymbol{\alpha}}} < \infty) \geq \varepsilon \mathbb{P}_x\left(\sigma_C < \infty\right) > 0,$$

since C is accessible. It follows that $\check{\boldsymbol{\alpha}}$ is accessible and by Theorem 3.14, ν^* is an irreducibility measure for $\check{P}$.

By the previous, Theorem 3.7, and Proposition 5.2, for all $\eta \in (0,1)$, νK_η and $\nu^* \check{K}_\eta = (\nu K_\eta)^*$ (the latter identity follows by monotone convergence) are maximal irreducibility measures for P and $\check{P}$, respectively. Thus, if $B \in \mathcal{X}$ is accessible for P, then

$$\nu^* \check{K}_\eta(B \times \{0,1\}) = (\nu K_\eta)^*(B \times \{0,1\}) = \nu K_\eta(B) > 0,$$

which shows that $B \times \{0,1\}$ is accessible for $\check{P}$. □