Spectral Gap of Markov Chains

Pedro Cardoso

Instituto Superior Técnico - LisMath Seminar

November 14, 2018

Summary

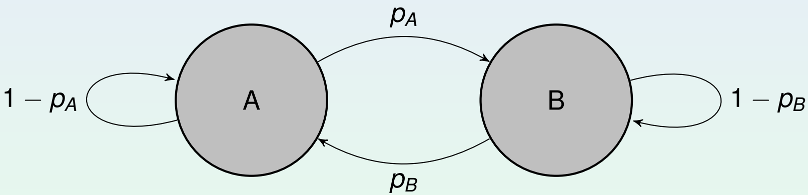
A concrete example

Consider bedrooms A and B. Regardless of where did you spend the last night, every morning you toss a coin which is always kept in the bedroom where you wake up. If the coin lands heads up, you will remain sleeping in the same place. If not, you will change your bedroom.

A concrete example

Denote the probabilities for the coins kept in bedrooms A and B to lands tails up by p_A and p_B, respectively. Then the probabilities for the coins kept in bedrooms A and B to lands heads up are $1 - p_A$ and $1 - p_B$. Since you do not want to be stuck in the same bed forever, we assume $p_A, p_B > 0$. This setting is a example of a Markov chain, which state space is the set $S = \{A, B\}$.

Graphic representation



Transition matrix

The standard representation of Markov chains with a finite state space is the transition matrix, whose entries are the transition probabilities between states. Denoting this matrix by P, we can write

$$P = \begin{bmatrix} 1 - p_A & p_A \\ p_B & 1 - p_B \end{bmatrix}.$$

Remark: the probability of you going to wake up in a specific bedroom tomorrow (**future**) **does not depend on** whether room you woke up yesterday (**past**), **given the** bedroom you are waking up today (**present**).

State of the chain

Denote the bedroom where you sleep in day k by X_k, with $k \in \{0,\ 1,\ \ldots\}$. Also, X_k is the k-th state of the chain. The possible states are the elements of S.

Then $\mathbb{P}(X_{k+1} = A | X_k = A) = 1 - p_A$, $\mathbb{P}(X_{k+1} = B | X_k = A) = p_A$, $\mathbb{P}(X_{k+1} = A | X_k = B) = p_B$ and $\mathbb{P}(X_{k+1} = B | X_k = B) = 1 - p_B$.

Probability distribution of X_k

We will denote the probability distribution of X_k by the vector μ_k, i.e.,

$$\mu_k = [\mu_k(A) \quad \mu_k(B)] = [\mathbb{P}(X_k = A) \quad \mathbb{P}(X_k = B)].$$

$$\begin{aligned}\mathbb{P}(X_{k+1} = A) &= \mathbb{P}(X_k = A)\mathbb{P}(X_{k+1} = A|X_k = A)\\ &+ \mathbb{P}(X_k = B)\mathbb{P}(X_{k+1} = A|X_k = B),\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_{k+1} = B) &= \mathbb{P}(X_k = A)\mathbb{P}(X_{k+1} = B|X_k = A)\\ &+ \mathbb{P}(X_k = B)\mathbb{P}(X_{k+1} = B|X_k = B).\end{aligned}$$

The equalities above are equivalent to the matrix equality $\mu_{k+1} = \mu_k \cdot P$, which illustrates why is very convenient to write the chain as a matrix . Therefore, $\mu_1 = \mu_0 \cdot P$; more generally, an inductive argument leads to $\mu_k = \mu_0 \cdot P^k, \forall k \geq 0$.

Stationary distribution

There is a special probability distribution:

$$\pi = \begin{bmatrix} \frac{p_B}{p_A + p_B} & \frac{p_A}{p_A + p_B} \end{bmatrix}$$

such that $\pi = \pi \cdot P$; we say π is a **stationary distribution** of the Markov chain P. If $\mu_0 = \pi$, then $\mu_1 = \mu_0 \cdot P = \pi \cdot P = \pi$; more generally, an inductive argument leads to $\mu_k = \pi, \forall k \geq 0$. In the same way, if $\mu_m = \pi$ for some m, then $\mu_k = \pi, \forall k \geq m$. Because of this, we say that if $\mu_m = \pi$ for some m, then the chain will have achieved the equilibrium.

Reversibility

Denote $\frac{p_B}{p_A+p_B}$ and $\frac{p_A}{p_A+p_B}$ by $\pi(A)$ and $\pi(B)$, respectively. Note that

$$\pi(A)\mathbb{P}(X_k+1 = B|X_k = A) \ = \ \frac{p_A p_B}{p_A + p_B} \ = \ \pi(B)\mathbb{P}(X_k+1 = A|X_k = B).$$

Then, we have

$$\pi(x)\mathbb{P}(X_k + 1 = y|X_k = x) \ = \ \pi(y)\mathbb{P}(X_k + 1 = x|X_k = y), \forall x, y \in S. \tag{1}$$

Our Markov chain satisfies (1), which is known as **reversibility** with respect to π.

Total variation distance

We are specially interested in evaluating how close the Markov chain is to the equilibrium in the day k. In order to answer this question, we will make use of a distance between μ_k and π.

Definition

The **total variation distance** *between two probability distributions α and β in S is defined by*

$$\|\alpha - \beta\|_{TV} = \max_{E \subset S} |\alpha(E) - \beta(E)|.$$

Remark: $|\alpha(S) - \beta(S)| = |1 - 1| = 0$ and $|\alpha(\emptyset) - \beta(\emptyset)| = |0 - 0| = 0$.

Evaluating the total variation distance

Denote $\mu_k - \pi$ by Δ_k. Then

$$
\begin{aligned}
&\Delta_{k+1}(A) = \mathbb{P}(X_{k+1} = A) - \pi(A) \\
=&\mathbb{P}(X_{K+1} = A | X_k = A)\mathbb{P}(X_k = A) \\
+&\mathbb{P}(X_{K+1} = A | X_k = B)\mathbb{P}(X_k = B) - \pi(A) \\
=&(1 - p_A)(\mu_k(A)) + p_B(1 - \mu_k(A)) - \pi(A) \\
=&(1 - p_A - p_B)(\mu_k(A)) + (p_A + p_B)\pi(A) - \pi(A) \\
=&(1 - p_A - p_B)(\mu_k(A) - \pi(A)) \\
=&(1 - p_A - p_B)\Delta_k(A).
\end{aligned}
$$

Evaluating the total variation distance

An inductive argument leads to $\Delta_k(A) = (1 - p_A - p_B)^k \Delta_0(A), \forall k \geq 0$.
Since $\mu_k(A) - \pi(A) = -\big(\mu_k(B) - \pi(B)\big)$,

$$\begin{aligned}\|\mu_k - \pi\|_{TV} &= |\mu_k(A) - \pi(A)| = |\Delta_k(A)| \\ =&|(1 - p_A - p_B)^k \Delta_0(A)| = |(1 - p_A - p_B)|^k \|\mu_0 - \pi\|_{TV}, \forall k \geq 0\end{aligned}$$

Since $0 < p_A < 1, 0 < p_B < 1$, we get $0 < p_A + p_B < 2$ and $|(1 - p_A - p_B)| < 1$. Therefore, we conclude that the distance between the chain and the equilibrium decays exponentially with rate $|(1 - p_A - p_B)|$.
Remark: the eigenvalues of P are $\beta_1 = 1$ and $\beta_2 = 1 - p_A - p_B$.

Markov chain on a general finite set S

Hereafter, S is a finite set with $n \geq 2$ elements. A sequence of states $(X_0, X_1, \dots)$ is a **Markov chain with state space S and transition matrix P** if $\forall x, y \in S, \forall k \geq 1$, and all events $H_{k-1} = \cap_{j=0}^{k-1} [X_j = x_j]$ satisfying $\mathbb{P}(H_{k-1} \cap [X_k = x]) > 0$, we have

$$\mathbb{P}(X_{k+1} = y | H_{k-1} \cap [X_k = x]) = \mathbb{P}(X_{k+1} = y | X_k = x) = P(x, y). \quad (2)$$

Equation (2) is often called the **Markov property**. Intuitively, given the present, the future is independent of the past.

Distribution of X_k

The x-th row of P is the distribution $P(x, \cdot)$; it has non-negative real entries such that

$$\sum_{y \in S} P(x, y) \;=\; 1\,, \quad \forall\, x \in S\,.$$

We will denote the distribution of X_k by the row vector μ_k:

$$\mu_k(x) = \mathbb{P}(X_k = x), \forall x \in S.$$

Matrix notation

The knowledge of P suffices to describe all the probability transitions and we will identify a Markov chain with its transition matrix.
In matrix notation, we have

$$\mu_{k+1} = \mu_k \cdot P, \forall k \geq 0,$$

which leads to

Proposition

$$\mu_k = \mu_0 \cdot P^k, \forall k \geq 0.$$

Irreducibility of a Markov chain

Definition

A chain P is called **irreducible** *if for any two states $x, y \in S$, there is a positive integer $k(x, y)$ such that $P^{k(x,y)}(x, y) > 0$.*

Remark: the identity matrix I_n of order $n > 1$ is not irreducible: if $n = 2$, we have

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Stationary distribution of a Markov chain

Definition

A **stationary distribution** π *of a Markov chain P is a probability distribution satisfying*

$$\pi = \pi \cdot P.$$

Then, the following result holds:

Proposition

Let P be a Markov chain with stationary distribution π*. Then* $\pi = \pi \cdot P^k, \forall k \geq 0.$

Positivity of the stationary distribution

Proposition

Let P be an irreducible Markov chain with stationary distribution π. Then $\pi(x) > 0, \forall x \in S$.

Remark: if P is the identity matrix I_n of order $n > 1$, every distribution μ is stationary, since $\mu \cdot I_n = \mu$.

Existence of a stationary distribution

Proposition

Let P be the transition matrix of a Markov chain on a finite state space S. Then there is at least one stationary distribution π for the chain P.

Uniqueness of the stationary distribution

Proposition

Let P be an irreducible Markov chain. Then P has only one stationary distribution π.

Remark: if P is the identity matrix I_n of order $n > 1$, every distribution μ is stationary, since $\mu \cdot I_n = \mu$.

Reversibility of a Markov chain

Given a Markov chain P and a probability distribution π on S, the **detailed balance equations** are

$$\pi(x)P(x,y) = \pi(y)P(y,x), \quad \forall x, y \in S. \tag{3}$$

Definition

If there is a probability distribution π which satisfies (3)*, we say P is* **reversible** *(with respect to π).*

Proposition

If P is a Markov chain which is reversible with respect to π, then π is a stationary distribution of P.

Intuition of the reversibility

Proposition

Let P be a Markov chain, reversible with respect to π. If the initial distribution is π, given a finite sequence of states $(y_0, y_1, \ldots, y_{k-1}, y_k)$, we have

$$\begin{aligned}&\mathbb{P}(X_0 = y_0, X_1 = y_1, \ldots, X_{k-1} = y_{k-1}, X_k = y_k)\\ =&\mathbb{P}(X_0 = y_k, X_1 = y_{k-1}, \ldots, X_{k-1} = y_1, X_k = y_0).\end{aligned}$$

Reversibility $\times$ irreducibility

The following Markov chain is irreducible, but it is not reversible.

Example (Biased random walk on the $n-$ cycle, $n \geq 3$)

We describe the transitions by a function $P : \mathbb{Z}_n \times \mathbb{Z}_n \to [0,1]$ *such as*

$$P(x,y) \;=\; \begin{cases} p \neq \frac{1}{2}\,, & \textit{if} \quad y - x = 1 (\textit{mod } n), \\ q = 1 - p\,, & \textit{if} \quad y - x = -1 (\textit{mod } n), \\ 0\,, & \textit{otherwise}. \end{cases}$$

Remark: the identity matrix I_n of order $n > 1$ is reversible with respect to any probability distribution μ, but it is not irreducible.

Properties of the eigenvalues of a Markov chain

Proposition

If β is a eigenvalue of a finite state Markov chain, then $|\beta| \leq 1$.

Proposition

Let P be an irreducible Markov chain. Then the geometric multiplicity of the eigenvalue $\beta_1 = 1$ is 1.

Remark: if the transition matrix is the identity matrix I_n of order $n > 1$, then $I_n \cdot v = v, \forall v$, i.e., the geometric multiplicity of $\beta_1 = 1$ is $n > 1$.

The inner product $< \cdot, \cdot >_\pi$

Hereafter, we identify a function $f : S \mapsto \mathbb{R}$ with a column vector $\in \mathbb{R}^n$. We define $\ell^2_\pi(S)$ as the vector space of the column vectors $\in \mathbb{R}^n$ with the inner product with respect to the probability distribution π, given by

$$< f, g >_\pi = \sum_{x \in S} f(x)g(x)\pi(x), \ \forall f, g : S \mapsto \mathbb{R}.$$

Example: If $f(x) = 1, \forall x \in S$, then

$$< f, f >_\pi = \sum_{x \in S} f(x)f(x)\pi(x) = \sum_{x \in S} 1 \cdot 1 \cdot \pi(x) \ = 1.$$

P may be self-adjoint

If we identify P with the linear operator $P : \ell^2_\pi(S) \to \ell^2_\pi(S)$, we get

Proposition

Let P be a transition matrix of a Markov chain. Then

$$< Pf, g >_\pi = < f, Pg >_\pi, \forall f, g \in \ell^2_\pi(S),$$

if, and only if, P is reversible with respect to π.

Spectral Theorem for Markov chains

Proposition

Hereafter, let P be an irreducible, reversible Markov chain with stationary distribution π. We have

a) *The linear operator $P : \ell^2_\pi(S) \to \ell^2_\pi(S)$ is diagonalizable.*

b) *Denote the eigenvalues of the matrix P by β_j, $1 \leq j \leq n$. Then they may be written in descending order, such that*

$$1 = \beta_1 > \beta_2 \geq \ldots \geq \beta_n \geq -1.$$

Probability distribution: discrete time k

Notice that we have

$$P^k(x,y) \;=\; \mathbb{P}([X_0 = x] \cap [X_k = y] | [X_0 = x]).$$

Enumerating the elements of S, the x-th row of P^k is the probability distribution

$$P^k(x,\cdot) = \mathbb{P}([X_0 = x] \cap [X_k = \cdot] | [X_0 = x]),$$

which is denoted by P^k_x.

Spectral gap of discrete-time Markov chains

Proposition

Let $\beta_* = \max\{|\beta_n|, \beta_2\}$, $\pi_* = \min_{x\in S}\{\pi(x)\} > 0$*. Then,*

$$2\|P_x^k - \pi\|_{TV} \;\leq\; {\pi_*}^{-1/2}\beta_*^k,\ \forall x \in S, \forall k \geq 0.$$

We define the spectral gap of the Markov chain P in this setting as $\gamma_* = 1 - \beta_*$. Note that if $\beta_n > -1$, then $\beta^* < 1$.

Discrete-time setting

In the discrete-time setting, change of states occur only at $k = 1, 2, 3, \ldots, etc$. In this case, if T_j is the difference between the $(j+1) - th$ and the $j - th$ transition instants, $T_j = 1, \forall j$.

Continuous-time setting

Hereafter, we will consider a continuous-time setting. Let $T_1, T_2, \dots$ be independent identically distributed exponential random variables of unit rate. That is, each T_j takes values in $[0, \infty)$ and has distribution function

$$\mathbb{P}(T_j \le t) = \begin{cases} 1 - e^{-t}, & \text{if } t \ge 0, \\ 0, & \text{if } t < 0. \end{cases}$$

Intuitively, now the difference between two transition instants may be any positive number.

Continuous-time setting

Let $(\Phi_k)_{k=0}^{\infty}$ be a discrete-time Markov chain with transition matrix P, independent of the random variables $(T_k)_{k=1}^{\infty}$. Let $C_0 = 0$ and $C_k := \sum_{j=1}^{k} T_j$, for $k \geq 1$. Define $X_t := \Phi_k$, if $C_k \leq t < C_{k+1}$. Then $(X_t)_{t \geq 0}$ is a continuous-time Markov chain with transition matrix P.

Probability distribution: continuous time t

We define H^t by

$$H^t(x,y) \ := \ \mathbb{P}([X_0 = x] \cap [X_t = y] | [X_0 = x]).$$

Enumerating the elements of S, the x-th row of H^t is the probability distribution

$$H^t(x,\cdot) = \mathbb{P}([X_0 = x] \cap [X_t = \cdot] | [X_0 = x]),$$

which is denoted by H^t_x.

Spectral gap of continuous-time Markov chains

Proposition

Let $\pi_* = \min_{x \in X}\{\pi(x)\} > 0$. *Then,*

$$2\|H_x^t - \pi\|_{TV} \ \leq \ \pi_*^{-1/2} e^{-(1-\beta_2)t}, \ \forall x \in S, \forall t \geq 0.$$

We define the spectral gap of the Markov chain P in this setting as $\gamma = 1 - \beta_2 > 0$.

Dirichlet forms for the chain P

Given a column $f \in \mathbb{R}^n$, we can evaluate the following Dirichlet forms:

Definition

Let $f \in \mathbb{R}^n$. We define the first Dirichlet form $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ as

$$\mathcal{E}(f) \ := < (I - P)f, f >_\pi = \sum_{x \in S} \big((I - P)f\big)(x) f(x) \pi(x).$$

Definition

Let $f \in \mathbb{R}^n$. We define the second Dirichlet form $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ as

$$\mathcal{F}(f) \ := < (I + P)f, f >_\pi = \sum_{x \in S} \big((I + P)f\big)(x) f(x) \pi(x).$$

Auxiliar Markov chain

Let $\overline{P}$ be an auxiliar irreducible Markov chain in S with known eigenvalues which is reversible with respect to $\overline{\pi}$. Denote the eigenvalues of the matrix $\overline{P}$ by $\overline{\beta_j}$, $1 \leq j \leq n$. Then they may be written in descending order, such that

$$1 = \overline{\beta}_1 > \overline{\beta}_2 \geq \ldots \geq \overline{\beta}_n \geq -1.$$

Let $\overline{\mathcal{E}}$ and $\overline{\mathcal{F}}$ be the first and second Dirichlet forms with respect to $\overline{P}$, respectively.

Dirichlet forms for the chain $\overline{P}$

Given a column $f \in \mathbb{R}^n$, we can evaluate the following Dirichlet forms:

Definition

Let $f \in \mathbb{R}^n$. We define the first Dirichlet form $\overline{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}$ as

$$\overline{\mathcal{E}}(f) \ := < (I - \overline{P})f, f >_{\overline{\pi}} = \sum_{x \in S} \big((I - \overline{P})f\big)(x) f(x) \overline{\pi}(x).$$

Definition

Let $f \in \mathbb{R}^n$. We define the second Dirichlet form $\overline{\mathcal{F}} : \mathbb{R}^n \to \mathbb{R}$ as

$$\overline{\mathcal{F}}(f) \ := < (I + \overline{P})f, f >_{\overline{\pi}} = \sum_{x \in S} \big((I + \overline{P})f\big)(x) f(x) \overline{\pi}(x).$$

Comparison between two Markov chains

Proposition

In the notations above, the following assertions hold:

a) *If there is a positive constant A such that $\overline{\mathcal{E}} \leq A\mathcal{E}$ and a positive constant B such that $\overline{\pi} \geq B\pi$, then*

$$\beta_j \leq 1 - \frac{B}{A}(1 - \overline{\beta}_j),\ 2 \leq j \leq n.$$

b) *If there is a positive constant A^* such that $\overline{\mathcal{F}} \leq A^*\mathcal{F}$ and a positive constant B such that $\overline{\pi} \geq B\pi$, then*

$$\beta_j \geq -1 + \frac{B}{A^*}(1 + \overline{\beta}_j),\ 2 \leq j \leq n.$$

Example: complete graph with loops

Proposition

Let $\overline{P} \equiv 1/n$, which means each entry of $\overline{P}$ is equal to $1/n$. Then the n eigenvalues of $\overline{P}$ are

$$\overline{\beta_1} = 1;\ \overline{\beta_2} = \overline{\beta_3} = \ldots = \overline{\beta}_n = 0.$$

Example: complete graph without loops

Proposition

Let $\overline{P}$ be the matrix of the simple random walk in a complete graph (without loops), that is, each entry of the principal diagonal of $\overline{P}$ is null and the rest of the entries are equal to $\frac{1}{n-1}$. Then, the n eigenvalues of $\overline{P}$ are

$$\overline{\beta_1} = 1; \ \overline{\beta_2} = \overline{\beta_3} = \ldots = \overline{\beta}_n = -\frac{1}{n-1}.$$

Example: simple random walk on the $n-cycle$

We describe the transitions by a function $\overline{P} : \mathbb{Z}_n \times \mathbb{Z}_n \to [0,1]$ such as

$$P(x,y) = \begin{cases} \frac{1}{2}, & \text{if} \quad y - x = 1(mod\ n), \\ \frac{1}{2}, & \text{if} \quad y - x = -1(mod\ n), \\ 0, & \text{otherwise.} \end{cases}$$

Proposition

The n eigenvalues of the simple random walk on the n-cycle are $\cos(\frac{2\pi j}{n})$, *where* $0 \leq j \leq n-1$.

Spectral gap $\times$ the first Dirichlet form

Proposition

The spectral gap $\gamma = 1 - \beta_2 > 0$ *satisfies*

$$\gamma = \min_{\substack{f \in \mathbb{R}^n \\ Var_\pi(f) \neq 0}} \frac{\mathcal{E}(f)}{Var_\pi(f)},$$

where

$$Var_{\pi(f)} = \sum_{x \in S} (f(x))^2 \pi(x) - \Big(\sum_{x \in S} f(x)\pi(x) \Big)^2, \forall f \in \mathbb{R}^n.$$

Bibliography

- Persi Diaconis and Laurent Saloff-Coste. *Probability: theory and examples.*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, fourth edition, 2010.
- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, second edition, 2013.

THANK YOU!