

430.714

Estimation Theory

Prof. Songhwai Oh
ECE, SNU

MARKOV CHAINS

Lecture outline

Markov-chain foundations

Transitions, return times, irreducibility, and periodicity

Long-run behavior

Stationarity, eigenmodes, and the strong law

Posterior sampling

Metropolis–Hastings and its detailed-balance proof

The Markov property

The next-state distribution depends on the current state.

$$\Pr(X_{n+1} = j \mid X_n = i, \dots, X_0 = i_0) = P_{ij}$$

Time-homogeneous transitions

$$P_{ij} = \Pr(X_{n+1} = j \mid X_n = i)$$

The transition matrix P is the same at every time step.

A transition matrix

$$P_{ij} \geq 0, \quad \sum_j P_{ij} = 1$$

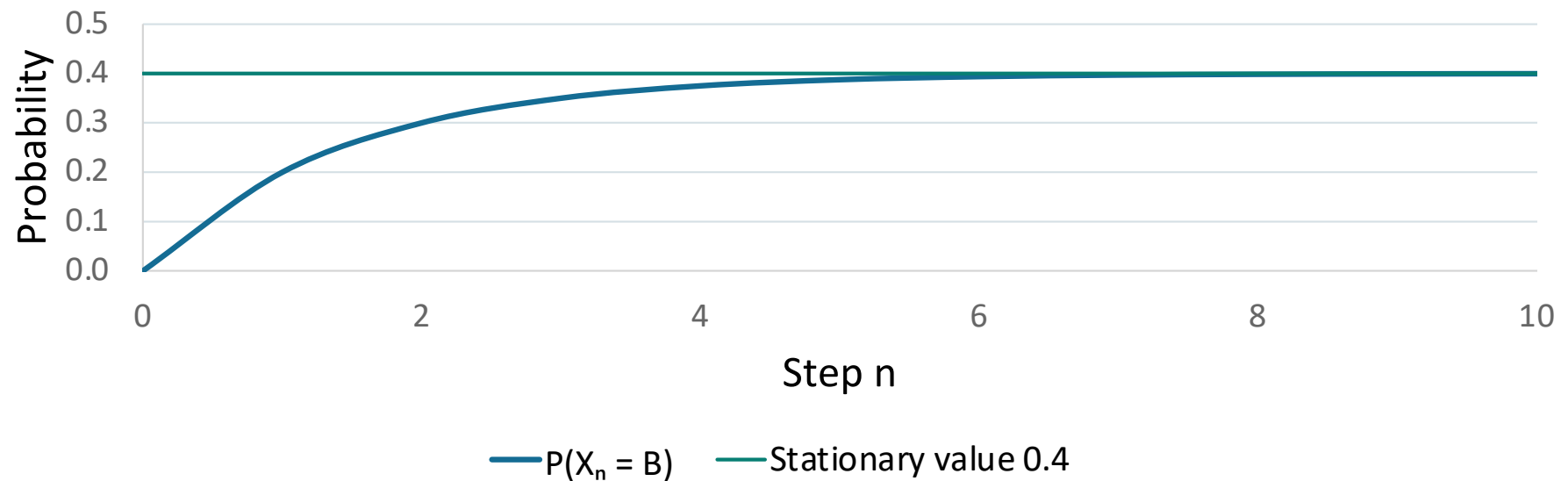
Current / next	A	B
A	0.8	0.2
B	0.3	0.7

A row gives the probabilities of all possible next states.

Evolution of the state distribution

$$\mu_{n+1} = \mu_n P, \quad \mu_n = \mu_0 P^n$$

Example: start in state A.



First return to a state

$$T_y = \inf\{n \geq 1 : X_n = y\}$$

When the chain starts at y , T_y is its first return time.

$$f_y = \Pr(T_y < \infty \mid X_0 = y)$$

A return time may be infinite.

Transient and recurrent states

Transient state

$$f_y < 1$$

There is a positive probability of never returning.

Recurrent state

$$f_y = 1$$

The chain eventually returns with probability one.

Accessibility and irreducibility

$$x \rightarrow y \iff \exists n \geq 0 : (P^n)_{xy} > 0$$

States communicate when each is accessible from the other.

Irreducible chain

$$x \rightarrow y \text{ for every } x, y \in \mathcal{S}$$

Every state belongs to the same communicating class.

Periodicity

$$d(x) = \gcd\{n \geq 1 : (P^n)_{xx} > 0\}$$

Example: deterministic alternation

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad d(A) = d(B) = 2$$

A stationary distribution exists, but the distribution from a fixed starting state keeps alternating.

Aperiodicity

$$d(x) = 1 \quad \text{for every state } x$$

A sufficient condition for each state

$$P_{xx} > 0 \quad \implies \quad d(x) = 1$$

In an irreducible chain, all states share the same period.

A self-loop can break deterministic alternation.

Stationary distribution

$$\pi P = \pi, \quad \pi(x) \geq 0, \quad \sum_x \pi(x) = 1$$

$$\sum_x \pi(x) P_{xy} = \pi(y)$$

Starting with π keeps the marginal distribution unchanged.

π is a normalized left eigenvector with eigenvalue 1.

Example: finding the stationary law

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

$$0.2\pi(A) = 0.3\pi(B), \quad \pi(A) + \pi(B) = 1$$

$$\pi = (0.6, 0.4)$$

Convergence to stationarity

Finite state space

Irreducibility and aperiodicity ensure a unique stationary law.

$$\Pr(X_n = y \mid X_0 = x) \xrightarrow[n \rightarrow \infty]{} \pi(y)$$

The limiting marginal law does not depend on the starting state.

For countable chains, positive recurrence is also required.

Eigenvalues and convergence

$$\lambda_0 = 1, \quad |\lambda_m| \leq 1$$

$$\lambda_* = \max_{1 \leq m \leq d-1} |\lambda_m| \quad \text{Second largest eigenvalue of P}$$

Fact: If P is irreducible and aperiodic, then $\lambda^* < 1$.

Use eigenvalue magnitudes: general eigenvalues may be complex.

Decay of eigenmodes

Assume P is diagonalizable and use left eigenvectors.

$$v_m P = \lambda_m v_m \quad \lambda_0 = 1, \quad |\lambda_m| \leq 1$$

Initial state distribution:

$$\mu_0 = \sum_{m=0}^{d-1} a_m v_m$$

Distribution at time k :

$$\mu_k = \mu_0 P^k = \sum_{m=0}^{d-1} a_m \lambda_m^k v_m$$

Since $\lambda_* < 1$, $\lambda_m^k \rightarrow 0$ as $k \rightarrow \infty$ for all $m > 0$. Hence, $\mu_k \rightarrow a_0 v_0$

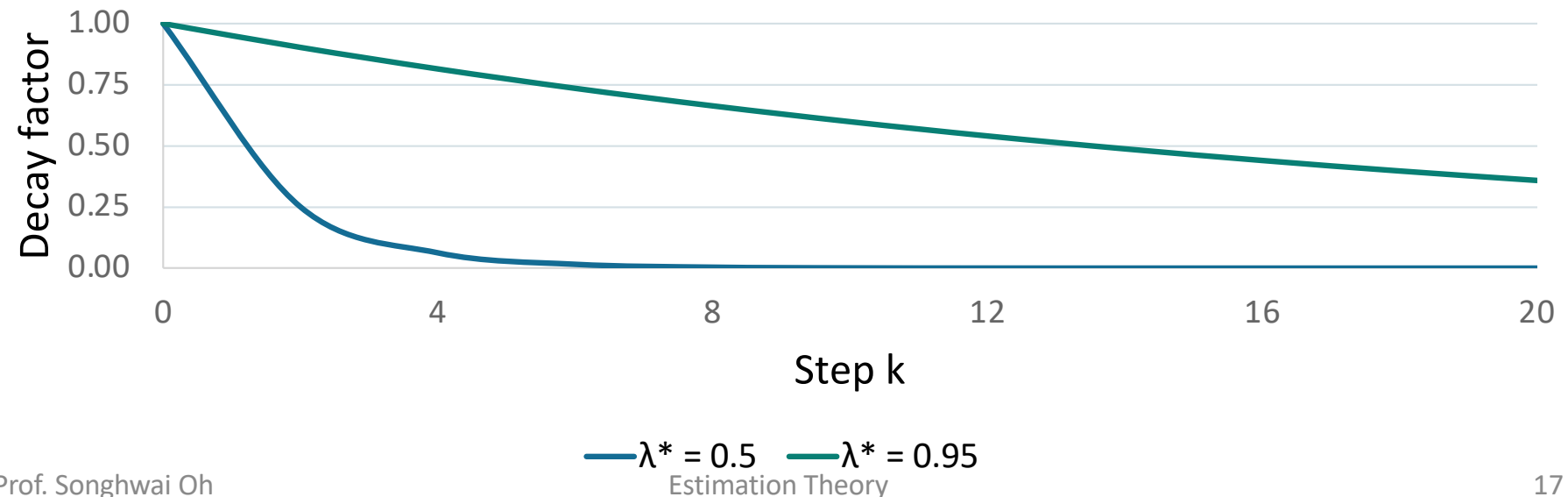
If there is π such that $\pi P = \pi$, then $\pi = a_0 v_0$, where $a_0 = \frac{1}{\sum v_0(i)}$

A bound on the convergence rate

$$\mu_k(x) - \pi(x) = \sum_{m=1}^{d-1} a_m \lambda_m^k v_m$$
$$\left| \mu_k(x) - \pi(x) \right| \leq \left(\sum_{m=1}^{d-1} |a_m v_m(x)| \right) (\lambda_*)^k$$

↑
rate of convergence

Eigenvalues close to unit magnitude give slow decay.



Strong Law for Markov chains

An irreducible positive recurrent chain has stationary law π .

Assume that

$$\sum_x \pi(x) |r(x)| < \infty$$

Then we have

$$\frac{1}{N} \sum_{k=1}^N r(X_k) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \sum_x \pi(x) r(x) = \mathbb{E}_\pi(r(X))$$

Time averages converge even when the samples are dependent.

At stationarity, it is like independent sampling from π .

Detailed balance

$$\pi(x)P_{xy} = \pi(y)P_{yx} \quad \text{for all } x, y$$

At stationarity, each pair has equal flow in both directions.

A sufficient condition for stationarity

Some stationary chains do not satisfy detailed balance.

Detailed balance implies stationarity

$$\begin{aligned}\sum_x \pi(x) P_{xy} &= \sum_x \pi(y) P_{yx} \\ &= \pi(y) \sum_x P_{yx} \\ &= \pi(y)\end{aligned}$$

$$\pi P = \pi$$

Posterior inference

Given a prior and likelihood, infer the latent state from observations.

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}$$

$$p(y) = \int p(y \mid x)p(x) dx$$

The normalization integral may be difficult to compute.

The MCMC construction

$$\pi(x) = p(x \mid y) \propto \tilde{\pi}(x) = p(y \mid x)p(x)$$

Construct transitions that preserve this target distribution.

Simulate the chain from a chosen initial state.

Estimate posterior expectations by averaging along the trajectory.

Metropolis–Hastings acceptance

At current state x , propose a candidate from q .

$$x' \sim q(\cdot \mid x)$$

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x')q(x \mid x')}{\pi(x)q(x' \mid x)} \right\}$$

The reverse-proposal term corrects proposal asymmetry.

Symmetric proposals

$$q(x' | x) = q(x | x')$$

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x')}{\pi(x)} \right\}$$

Example: a candidate with half the target density has acceptance probability 1/2.

The normalization constant cancels

$$\frac{\pi(x')}{\pi(x)} = \frac{\tilde{\pi}(x')/Z}{\tilde{\pi}(x)/Z} = \frac{\tilde{\pi}(x')}{\tilde{\pi}(x)}$$

$$= \frac{p(y | x')p(x')}{p(y | x)p(x)}$$

The acceptance rule does not require $p(y)$.

One Metropolis–Hastings step

1. Propose x' from q conditional on the current state.
2. Compute the Metropolis–Hastings acceptance probability.
3. Draw an independent uniform U .

$$X_{k+1} = \begin{cases} x', & U \leq \alpha(x, x'), \\ x, & U > \alpha(x, x'). \end{cases}$$

Rejections create repeated states in the trajectory.

Estimating a posterior expectation

$$\hat{\mu}_N = \frac{1}{N} \sum_{k=1}^N r(X_{B+k})$$

B is an optional warm-up length. N counts retained steps.

Include the repeated states created by rejected proposals.

$$\hat{\mu}_N \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbb{E}[r(X) \mid y]$$

The transition kernel includes rejections

Discrete state space

$$P(x, x') = q(x' \mid x)\alpha(x, x'), \quad x' \neq x$$

$$P(x, x) = 1 - \sum_{x' \neq x} q(x' \mid x)\alpha(x, x')$$

The continuous kernel also has a probability mass for staying at x .

Why Metropolis–Hastings preserves π

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x')q(x | x')}{\pi(x)q(x' | x)} \right\}$$

$$\begin{aligned}\pi(x)P(x, x') &= \pi(x)q(x' | x)\alpha(x, x') \\ &= \min \{ \pi(x)q(x' | x), \pi(x')q(x | x') \} \\ &= \pi(x')P(x', x)\end{aligned}$$

The minimum is unchanged when x and x' exchange roles.

Detailed balance establishes stationarity of the target.

Worked example: asymmetric proposal

Move	Proposal q	Acceptance α
A to B	0.8	7/12
B to A	0.2	1

$$P = \begin{pmatrix} 8/15 & 7/15 \\ 1/5 & 4/5 \end{pmatrix}, \quad \pi = (0.3, 0.7)$$

$$\pi(A)P_{AB} = \pi(B)P_{BA} = 0.14$$

Conditions for reliable MCMC

Target preservation

Detailed balance establishes the stationary target law.

Reachability and convergence

Check irreducibility and aperiodicity of the constructed chain.

Estimation

Integrable statistics and enough exploration support time averages.