

Lecture-23: DTMC: Invariant Distribution

1 Invariant Distribution

Let $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}^+}$ be a time-homogeneous Markov chain with transition probability matrix $P : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$.

Definition 1.1. For any countable set \mathcal{X} , we define the set of probability mass functions over \mathcal{X} as

$$\mathcal{M}(\mathcal{X}) \triangleq \left\{ \nu \in [0, 1]^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \nu_x = 1 \right\}.$$

Definition 1.2. A probability distribution $\pi : \mathcal{X} \in \mathcal{M}(\mathcal{X})$ is said to be **stationary distribution** or invariant distribution for the Markov chain X if it satisfies the **global balance equation** $\pi = \pi P$.

Definition 1.3. When the initial distribution of a Markov chain is $\nu \in \mathcal{M}(\mathcal{X})$, then the condition probability is denoted by $P_\nu : \mathcal{F} \rightarrow [0, 1]$ defined by

$$P_\nu(A) \triangleq \sum_{x \in \mathcal{X}} \nu(x) P_x(A) \text{ for all events } A \in \mathcal{F}.$$

Definition 1.4. For a Markov chain $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}^+}$, we denote the distribution of random variable $X_n : \Omega \rightarrow \mathcal{X}$ by $\nu_n \in \mathcal{M}(\mathcal{X})$ for all $n \in \mathbb{Z}_+$. That is, $\nu_n(x) \triangleq P_{\nu_0} \{X_n = x\}$ for all $x \in \mathcal{X}$.

Remark 1. We observe that $\nu_n(x) = \sum_{z \in \mathcal{X}} \nu_0(z) (P^n)_{zx}$ for all $x \in \mathcal{X}$.

Remark 2. Facts about the invariant distribution π .

- i. The global balance equation $\pi = \pi P$ is a matrix equation, that is we have a collection of $|\mathcal{X}|$ equations $\pi_y = \sum_{x \in \mathcal{X}} \pi_x p_{xy}$ for each $y \in \mathcal{X}$.
- ii. Balance equation across cuts is $\pi_y(1 - p_{yy}) = \pi_y \sum_{x \neq y} p_{yx} = \sum_{x \neq y} \pi_x p_{xy}$.
- iii. The invariant distribution π is left eigenvector of stochastic matrix P with the largest eigenvalue 1. The all ones vector is the right eigenvector of this stochastic matrix P for the eigenvalue 1.
- iv. From the Chapman-Kolmogorov equation for initial probability vector π , we have $\pi = \pi P^n$ for $n \in \mathbb{N}$. That is, if $\nu_0 = \pi$, then $\nu_n = \pi$ for all $n \in \mathbb{Z}_+$.
- v. Resulting process with initial distribution π is stationary, and hence have shift-invariant finite dimensional distributions. For example, for any $k, n \in \mathbb{Z}_+$ and $x_0, \dots, x_n \in \mathcal{X}$, we have

$$P_\pi \{X_0 = x_0, \dots, X_n = x_n\} = P_\pi \{X_k = x_0, \dots, X_{k+n} = x_n\} = \pi_{x_0} p_{x_0 x_1} \dots p_{x_{n-1} x_n}.$$

- vi. If the Markov chain is irreducible, with $\pi_x > 0$ for some $x \in \mathcal{X}$. Then for any $y \in \mathcal{X}$, we have $p_{xy}^{(m)} > 0$ for some $m \in \mathbb{N}$. Hence, $\pi_y \geq \pi_x p_{xy}^{(m)} > 0$. That is, the entire invariant vector is positive.
- vii. Any scaled version of π satisfies the global balance equation. Therefore, for any invariant vector $\alpha \in \mathcal{X}^{\mathbb{R}^+}$ of transition matrix P , the sum $\|\alpha\|_1 = \sum_{x \in \mathcal{X}} \alpha_x$ must be finite for positive recurrent Markov chains, to normalize such vectors and get a unique invariant measure $\pi = \frac{\alpha}{\|\alpha\|_1}$.

Theorem 1.5. *An irreducible Markov chain with transition probability matrix P is positive recurrent iff there exists a unique invariant probability measure $\pi \in \mathcal{M}(\mathcal{X})$ that satisfies global balance equation $\pi = \pi P$ and $\pi_x = \frac{1}{\mu_{xx}} > 0$ for all $x \in \mathcal{X}$.*

Proof. We will first show that positive recurrence implies the existence of invariant distribution, and then its converse.

Implication: Let $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}_+}$ be a positive recurrent Markov chain with initial state $X_0 = x$. Recall that $N_y(n) = \sum_{k=1}^n \mathbb{1}_{\{X_k=y\}}$ denotes the number of visits to state $y \in \mathcal{X}$ in the first n steps of the Markov chain. It follows that $\sum_{y \in \mathcal{X}} N_y(n) = n$ for each $n \in \mathbb{N}$. Let $H_x \triangleq \tau_x^{(1)}$ be the first recurrence time to state x , then we have $N_x(H_x) = 1$ and $\sum_{y \in \mathcal{X}} N_y(H_x) = H_x$.

Existence: We denote $v_y \triangleq \mathbb{E}_x[N_y(H_x)]$ for each $y \in \mathcal{X}$. We observe that $v_y \geq 0$ for each state $y \in \mathcal{X}$, in particular $v_x = 1$, and $\sum_{y \in \mathcal{X}} v_y = \mathbb{E}_x H_x = \mu_{xx} < \infty$ since X is positive recurrent. We will show that the vector $v = (v_x : x \in \mathcal{X})$ satisfies the global balance equations $v = vP$, and since v is summable, $\pi = \frac{v}{\sum_{x \in \mathcal{X}} v_x}$ is an invariant distribution for the Markov chain X . To see that the vector v satisfies the global balance equations, we observe from the monotone convergence theorem

$$v_y = \mathbb{E}_x N_y(H_x) = \mathbb{E}_x \sum_{n \in \mathbb{N}} \mathbb{1}_{\{X_n=y, n \leq H_x\}} = \sum_{n \in \mathbb{N}} P_x \{X_n = y, n \leq H_x\}.$$

Let $\lambda_{xy}^{(n)} \triangleq P_x \{X_n = y, n \leq H_x\}$. Observe that $\lambda_{xy}^{(1)} = p_{xy}$ for each $y \in \mathcal{X}$. For $n \geq 2$, we have

$$\begin{aligned} \lambda_{xy}^{(n)} &= \sum_{z \neq x} P_x \{X_n = y, X_{n-1} = z, n \leq H_x\} \\ &= \sum_{z \neq x} P(\{X_n = y\} \mid \{X_{n-1} = z, n \leq H_x, X_0 = x\}) P_x \{X_{n-1} = z, n \leq H_x\} \\ &= \sum_{z \neq x} P(\{X_n = y\} \mid \{X_{n-1} = z\}) P_x \{X_{n-1} = z, n-1 \leq H_x\} = \sum_{z \neq x} \lambda_{xz}^{(n-1)} p_{zy}. \end{aligned}$$

From the definition of $\lambda_{xy}^{(n)}$, we have $v_y = \sum_{n \in \mathbb{N}} \lambda_{xy}^{(n)}$ for each $y \in \mathcal{X}$. Therefore,

$$v_y = p_{xy} + \sum_{n \geq 2} \sum_{z \neq x} \lambda_{xz}^{(n-1)} p_{zy} = v_x p_{xy} + \sum_{z \neq x} v_z p_{zy} = \sum_{x \in \mathcal{X}} v_x p_{xy}.$$

Hence, $\pi = \frac{v}{\sum_{x \in \mathcal{X}} v_x}$ is an invariant measure of the transition matrix P , and $\pi_x = \frac{v_x}{\sum_{y \in \mathcal{X}} v_y} = \frac{1}{\mu_{xx}} > 0$.

Uniqueness: Next, we show that this is a unique invariant measure independent of the initial state x , and hence $\pi_y = \frac{1}{\mu_{yy}} > 0$ for all $y \in \mathcal{X}$. For uniqueness, we observe from the Chapman-Kolmogorov equations and invariance of π that $\pi = \frac{1}{n} \pi (P + P^2 + \dots + P^n)$. Hence,

$$\pi_y = \sum_{x \in \mathcal{X}} \pi_x \frac{1}{n} \sum_{k=1}^n p_{xy}^{(k)}, \quad y \in \mathcal{X}.$$

Taking limit $n \rightarrow \infty$ on both sides, and exchanging limit and summation on right hand side using bounded convergence theorem for summable series π , we get for all $y \in \mathcal{X}$

$$\pi_y = \frac{1}{\mu_{yy}} \sum_{x \in \mathcal{X}} \pi_x = \frac{1}{\mu_{yy}} > 0.$$

Converse: Let π be the positive invariant distribution of Markov chain X . Then, if the Markov chain was transient or null recurrent, we would have $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{xy}^{(k)} = 0$. Since π is an invariant vector, we get $\pi = \pi P^k$ for each $k \in \mathbb{N}$ and hence $\pi = \pi \frac{1}{n} \sum_{k=1}^n P^k$. Taking limit on both sides, we have $\pi = 0$, yielding a contradiction for its positivity.

□

Corollary 1.6. *An irreducible Markov chain on a finite state space \mathcal{X} has a unique positive stationary distribution π .*

Definition 1.7. An irreducible, aperiodic, positive recurrent Markov chain is called **ergodic**.

Remark 3. Additional remarks about the stationary distribution π .

i. For a Markov chain with multiple positive recurrent communicating classes $\mathcal{C}_1, \dots, \mathcal{C}_m$, one can find the positive equilibrium distribution for each class, and extend it to the entire state space \mathcal{X} denoting it by π_k for class $k \in [m]$. It is easy to check that any convex combination $\pi = \sum_{k=1}^m \alpha_k \pi_k$ satisfies the global balance equation $\pi = \pi P$, where $\alpha_k \geq 0$ for each $k \in [m]$ and $\sum_{k=1}^m \alpha_k = 1$. Hence, a Markov chain with multiple positive recurrent classes have a convex set of invariant probability measures, with the individual invariant distribution π_k for each positive recurrent class $k \in [m]$ being the extreme points.

ii. Let $\mu(0) = e_x$, that is let the initial state of the positive recurrent Markov chain be $X_0 = x$. Then, we know that

$$\pi_y = \frac{1}{\mu_{yy}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{xy}^{(k)} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_x N_y(n).$$

That is, π_y is limiting average of number of visits to state $y \in \mathcal{X}$.

iii. If a positive recurrent Markov chain is aperiodic, then limiting probability of being in a state y is its invariant probability, that is $\pi_y = \lim_{n \rightarrow \infty} p_{xy}^{(n)}$.

Theorem 1.8. *For an ergodic Markov chain X with invariant distribution π , and n th step distribution $\mu(n)$, we have $\lim_{n \rightarrow \infty} \mu(n) = \pi$ in the total variation distance.*

Proof. Consider independent time homogeneous Markov chains $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}^+}$ and $Y : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}^+}$ each with transition matrix P . The initial state of Markov chain X is assumed to be $X_0 = x$, whereas the Markov chain Y is assumed to have an initial distribution π . It follows that Y is a stationary process, while X is not. In particular,

$$\mu_y(n) = P_x \{X_n = y\} = p_{xy}^{(n)}, \quad P_\pi \{Y_n = y\} = \pi_y.$$

Let $\tau = \inf\{n \in \mathbb{Z}_+ : X_n = Y_n\}$ be the first time that two Markov chains meet, called the **coupling time**.

Finiteness: First, we show that the coupling time is almost surely finite. To this end, we define a new Markov chain on state space $\mathcal{X} \times \mathcal{X}$ with transition probability matrix Q such that $q((x, w), (y, z)) = p_{xy} p_{wz}$ for each pair of states $(x, w), (y, z) \in \mathcal{X} \times \mathcal{X}$. The n -step transition probabilities for this couples Markov chain are given by

$$q^{(n)}((x, w), (y, z)) \triangleq p_{xy}^{(n)} p_{wz}^{(n)}.$$

Ergodicity: Since the Markov chain X with transition probability matrix P is irreducible and aperiodic, for each $x, y, w, z \in \mathcal{X}$ there exists an $n_0 \in \mathbb{Z}_+$ such that $q^{(n)}((x, w), (y, z)) = p_{xy}^{(n)} p_{wz}^{(n)} > 0$ for all $n \geq n_0$ from a previous Lemma on aperiodicity. Hence, the irreducibility and aperiodicity of this new **product** Markov chain follows.

Invariant: It is easy to check that $\theta(x, w) = \pi_x \pi_w$ is the invariant distribution for this product Markov chain, since $\theta(x, w) > 0$ for each $(x, w) \in \mathcal{X} \times \mathcal{X}$, $\sum_{x, w \in \mathcal{X}} \theta(x, w) = 1$, and for each $(y, z) \in \mathcal{X} \times \mathcal{X}$, we have

$$\sum_{x, w \in \mathcal{X}} \theta(x, w) q((x, w), (y, z)) = \sum_{x \in \mathcal{X}} \pi_x p_{xy} \sum_{w \in \mathcal{X}} \pi_w p_{wz} = \pi_y \pi_z = \theta(y, z).$$

Recurrence: This implies that the product Markov chain is positive recurrent, and each state $(x, x) \in \mathcal{X} \times \mathcal{X}$ is reachable with unit probability from any initial state $(y, w) \in \mathcal{X} \times \mathcal{X}$.

In particular, the coupling time is almost surely finite.

Coupled process: Second, we show that from the coupling time onwards, the evolution of two Markov chains is identical in distribution. That is, for each $y \in \mathcal{X}$ and $n \in \mathbb{Z}_+$,

$$P_{X_\tau} \{X_n = y, n \geq \tau\} = P_{Y_\tau} \{Y_n = y, n \geq \tau\}.$$

This follows from the strong Markov property for the joint process where τ is stopping time for the joint process $((X_n, Y_n) : n \in \mathbb{Z}_+)$ such that $X_\tau = Y_\tau$, and both marginals have the identical transition matrix.

Limit: For any $y \in \mathcal{X}$, we can write the difference as

$$\left| p_{xy}^{(n)} - \pi_y \right| = \left| P_x \{X_n = y, n < \tau\} - P_\pi \{Y_n = y, n < \tau\} \right| \leq 2P_{\delta_x, \pi}(\tau > n).$$

Since the coupling time is almost surely finite for each initial state $x, y \in \mathcal{X}$, we have $\sum_{n \in \mathbb{N}} P_{\delta_x, \pi} \{\tau = n\} = 1$ and the tail-sum $P_{\delta_x, \pi} \{\tau > n\}$ goes to zero as n grows large, and the result follows.

□