Markov chains: theory and applications

Parimal Parag

Nov 22, 2021

Introduction

Probability space

Probability space $(\Omega, \mathcal{F}, P)$

- Sample space is an abstract set Ω
- Collection of subsets of sample space $\mathcal{F}$ that contain Ω, are closed under countable unions and complements
- Probability function $P : \mathcal{F} \to [0, 1]$

Probability space $(\Omega, \mathcal{F}, P)$

- Let $\Omega = \{H, T\}^{\mathbb{N}}$
- $\mathcal{F} = \sigma(E_1, \dots, E_n, \dots)$ where $E_n \triangleq \{\omega \in \Omega : \omega_n = H\}$
- Probability defined as $P(\cap_{i \in S} E_i \cap_{i \in T} E_i^c) \triangleq p^{|S|}(1 - p)^{|T|}$ for all finite disjoint subsets $S, T \subseteq \mathbb{N}$

Discrete stochastic processes

Discrete stochastic processes X

- Discrete time index set $\mathbb{N}$
- Countable state space $\mathcal{X}$
- A mapping $X : \Omega \to \mathcal{X}^{\mathbb{N}}$ such that $\cap_{k=1}^{n}\{X_k = x_k\} \in \mathcal{F}$ for all $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathcal{X}$
- In particular, $(X_1, \dots, X_n)$ is a random vector for all $n \in \mathbb{N}$

Independent and identically distributed processes $Z : \Omega \to \{0,1\}^{\mathbb{N}}$

- For $(\Omega, \mathcal{F}, P)$ of the previous example, let $Z_n = \mathbb{1}_{\{\omega_n = H\}}$
- For any $z_1, \dots, z_n \in \{0,1\}$, let $S \triangleq \{i \in [n] : z_i = 1\}$
- The probability of joint event

$$P(\cap_{k=1}^{n}\{Z_k = z_k\}) = \prod_{k=1}^{n} P(\{Z_k = z_k\}) = p^{|S|}(1-p)^{n-|S|}$$

Markov property

History of the process X until time n

- Collection of events of the form $H_n = \cap_{k=1}^{n}\{X_k = x_k\}$

Discrete time Markov chain X

Conditioned on the present, future is independent of the past

$$P(\{X_{n+1} = y\} \mid \{X_n = x\} \cap H_{n-1}) = P(\{X_{n+1} = y\} \mid \{X_n = x\})$$

Random walk S

- Let $Z : \Omega \to \{0,1\}^{\mathbb{N}}$ be an independent step size sequence
- The random walk $S : \Omega \to \mathbb{N}^{\mathbb{N}}$ defined as $S_n \triangleq \sum_{k=1}^{n} Z_k$ is Markov

$$\begin{aligned} P(\{S_{n+1} = y\} \mid \{S_n = x\} \cap H_{n-1}) &= P(\{S_{n+1} = y\} \mid \{S_n = x\}) \\ &= P\{Z_{n+1} = y - x\} \end{aligned}$$

Transition matrix and homogeneity

Transition matrix $P(n+1) : \mathcal{X} \times \mathcal{X} \to [0,1]$

- (x, y)th element is $P_{xy}(n+1) \triangleq P(\{X_{n+1} = y\} \mid \{X_n = x\})$
- xth row is $(P_{xy}(n+1) : y \in \mathcal{X})$ the conditional probability distribution of X_{n+1} given the event $\{X_n = x\}$

Homogeneity

When the transition matrix $P(n) = P$ for all times n

Random walk S

When the underlying step-size sequencer is *i.i.d.* , then the random walk S is homogeneous

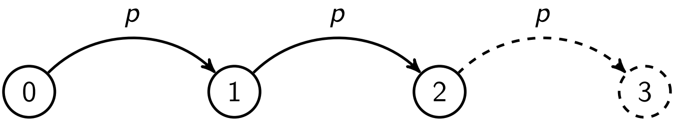
$$P_{xy}(n+1) = P\{Z_{n+1} = y - x\} = P\{Z_1 = y - x\} = P_{xy}(1)$$

Representation

Transition matrix P as weighted transition graph $(\mathcal{X}, E, w)$

- Set of nodes is the state space $\mathcal{X}$
- Set of directed edges $E = \{(x, y) \in \mathcal{X} \times \mathcal{X} : p_{xy} > 0\}$
- Weight of edge $(x, y) \in E$ is $w_{xy} = P_{xy}$

Random walk on non-negative integers



Random representation

For each Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, there exists a function f and independent sequence $Z : \Omega \to \mathcal{X}^{\mathbb{N}}$, such that $X_{n+1} = f(X_n, Z_{n+1})$

Chapman Kolmogorov Equations

n-step transition probability matrix $P^{(n)}$

- (x, y)th element is $P_{xy}^{(n)} \triangleq P(\{X_{m+n} = y\} \mid \{X_m = x\})$
- One can show that $P^{(n+m)} = P^{(n)}P^{(m)}$ to conclude that $P^{(n)} = P^n$
- If $\nu(n)$ denotes the distribution of X_n, then

$$\nu(n) = \nu(0)P^n$$

Three questions

- What is the limiting distribution $\lim_{n\to\infty} \nu(n)$?
- What is the average time spent in a state x?
- What is the mean return time to state x?

Stopping Time

Stopping time $\tau : \Omega \to \mathbb{N}$

- Value of stopping time being equal to n completely determined by process until time n

$$\{\tau = n\} \in \sigma(X_1, \ldots, X_n)$$

- τ is finite almost surely
- First hitting time to a state x

$$\tau_x \triangleq \inf\{n \in \mathbb{N} : X_n = x\}$$

- τ_x is a stopping time if finite almost surely

Strong Markov Property

Strong Markov Property

Markov property holds at stopping times

$$P(\{X_{\tau+1} = y\} \mid \{X_\tau = x\}) = P_{xy}$$

Regeneration for Markov chain X

- If τ_x is finite almost surely
- $(X_{\tau_x}, X_{\tau_x+1}, \ldots,)$ distributed identically to X for $X_0 = x$
- X_{τ_x+n} independent of $(X_0, \ldots, X_{\tau_x-1})$

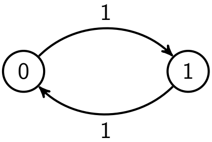
Invariant distribution

Periodic Markov chain

Periodicity

- For a state x, we can define the set of time steps for possible return
$$A(x) \triangleq \{n \in \mathbb{N} : P_{xx}^{(n)} > 0\}$$
- Periodicity of state x: $d(x) = \gcd A(x)$
- Aperiodic when $d(x) = 1$

Two state Markov chain with $X_0 = 0$



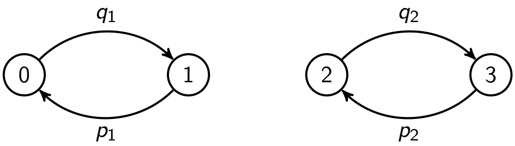
- $\nu(n) = \begin{bmatrix} 1 & 0 \end{bmatrix}$ when n even and $\nu(n) = \begin{bmatrix} 0 & 1 \end{bmatrix}$ when n odd
- No limit for $\nu(n)$

Reducible Markov chain

Irreducibility

- A state y is reachable from state x if $P_{xy}^{(n)} > 0$ for some n
- If all states are reachable from one another, then P is irreducible

Reducible Markov chain



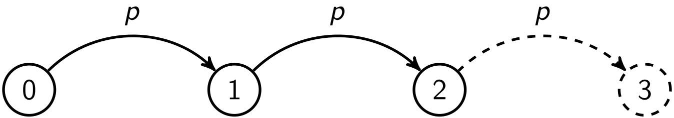
- $\nu(n)$ different when $X_0 \in \{0, 1\}$ and when $X_0 \in \{2, 3\}$
- No limit for $\nu(n)$

Transience and Recurrence

State x for a Markov chain X

- Transient if $X_0 = x$ and there is a finite probability of no return to state x
- Recurrent if $X_0 = x$ and the recurrence time τ_x is finite almost surely
  - Positive recurrent if mean recurrence time $\mathbb{E}[\tau_x \mid \{X_0 = x\}]$ is finite
  - null recurrent otherwise

Transient Markov chain



- $P\{S_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}$
- S_n/n converges to $\mathbb{E}Z_1$

Limiting distribution

Ergodic Markov chain

Irreducible, aperiodic, and positive recurrent

Number of visits to a state x in n time steps

- $N_x(n) \triangleq \sum_{k=1}^{n} \mathbb{1}_{\{X_k=x\}}$
- Mean number of visits $\sum_{k=1}^{n} \nu_x(k)$
- $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \nu_x(k) = \lim_{n\to\infty} \nu_x(n)$
- $\lim_{n\to\infty} \nu_x(n) = \lim_{n\to\infty}[\nu(0)P^n]_x = \lim_{n\to\infty} P^n_{yx}$

Average number of visits to a state x

- Let $\tau_x^{(k)}$ be the kth recurrence time to state x
- $n \geqslant \sum_{i=1}^{N_x(n)} \tau_x^{(i)}$
- $\lim_{n\to\infty} \frac{1}{n} N_x(n) = \frac{1}{\mathbb{E}[\tau_x|\{X_0=x\}]}$ almost surely

Invariant distribution

Ergodic Markov chain X with transition matrix P

- Unique probability vector π such that $\pi = \pi P$
- Left eigenvector with maximum eigenvalue 1
- If $\nu(0) = \pi$, then $\nu(n) = \pi P^n = \pi$ for all n
- If $\nu(0) = \pi$, then $\frac{1}{n}\sum_{k=1}^{n} \nu(k) = \pi$
- For a positive recurrent Markov chain

$$\pi_x = \frac{1}{\mathbb{E}[\tau_x \mid \{X_0 = x\}]}$$

- Showing that $(\mathbb{E}_x N_y(\tau_x) : y \in \mathcal{X})$ is a positive eigenvector for P with eigenvalue 1

Ergodic Markov chain X

$$\pi_x = \frac{1}{\mathbb{E}_x[\tau_x]} = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_x[N_x(n)] = \lim_{n\to\infty} \nu_x(n)$$

Computing invariant distribution

Cut method

- Global balance equation $\pi = \pi P$
- Let $A \subseteq \mathcal{X}$, then $(A, \mathcal{X} \setminus A)$ is a partition of state space
- Probability flux balances across cuts

$$\sum_{y \notin A} \sum_{x \in A} \pi_x P_{xy} = \sum_{x \in A} \sum_{y \notin A} \pi_y P_{yx}$$

Transform method

- When the state space $\mathcal{X} = \mathbb{Z}_+$ and P_{xy} is tridiagonal and homogeneous
- Discrete transform $\Pi(z) \triangleq \sum_{x \in \mathbb{Z}_+} \pi_x z^x$
- $\sum_{x \in \mathbb{Z}_+} \sum_{y \in \mathbb{Z}_+} \pi_y P_{yx} z^x = \sum_{x \in \mathbb{Z}_+} \pi_{x-1} P_{x-1,x} z^x + \sum_{x \in \mathbb{Z}_+} \pi_x P_{x,x} z^x + \sum_{x \in \mathbb{Z}_+} \pi_{x+1} P_{x+1,x} z^x$

Application

Page Rank

Internet as a network of webpages

- Let the set of webpages be denoted by $\mathcal{X}$
- If there exists a hyperlink from page x to page y, then $(x, y) \in E$

Ordering webpages

- Rank of page x denoted by r_x
- $r_x \propto d_{\mathrm{in}}(x) = \sum_{y \in \mathcal{X}} \mathbb{1}_{\{(y,x) \in E\}}$ number of webpages referring to it
- Agnostic to referral by important webpages

Importance of webpages

Importance shared by a page to all its referred pages

$$r_x = \sum_{(y,x)\in E} \frac{r_y}{d_{\mathrm{out}}(y)}$$

- Ranking vector $(r_x : x \in \mathcal{X})$ is an invariant vector for transition probability matrix H where $H_{yx} = \frac{1}{d_{\mathrm{out}}(y)} \mathbb{1}_{\{(y,x)\in E\}}$
- Ranking is proportional to invariant distribution of a Markov chain with transition matrix H

Computing invariant distribution

- Markov random walk under transition matrix H with initial distribution $\nu(0)$
- $\nu(n) = \nu(n-1)H = \nu(0)H^n$
- Stop if $\|\nu(n) - \nu(n-1)\| < \epsilon$

Regularizing webpage graphs

Networked pages

- Transition matrix H may not be irreducible
- Teleportation probability $(1-\alpha)$
- Personalized preference probability distribution v for webpages
- Modified transition probability

$$G_{xy} = \alpha H_{xy} + (1-\alpha) v_y$$

- Modified transition matrix $G = \alpha H + (1-\alpha)\mathbf{1} v^T$