Lecture 6: Markov Chains

Nicolás Rivera John Sylvester Luca Zanetti Thomas Sauerwald

Lent 2020

UNIVERSITY OF CAMBRIDGE

Outline

Stochastic Processes and Markov Chains

Stopping and Hitting Times

Irreducibility and Stationarity

Stochastic Process

A *Stochastic Process* $X = \{X_t : t \in T\}$ is a collection of random variables indexed by time (often $T = \mathbb{N}$) and in this case $X = (X_i)_{i=0}^{\infty}$.

A vector $\mu = (\mu(i))_{i \in \mathcal{I}}$ is a *Probability Distribution* or *Probability Vector* on $\mathcal{I}$ if $\mu(i) \in [0, 1]$ and

$$\sum_{i \in \mathcal{I}} \mu(i) = 1.$$

Markov Chains

Markov Chain (Discrete Time and State, Time Homogeneous)

We say that $(X_i)_{i=0}^{\infty}$ is a *Markov Chain* on *State Space* $\mathcal{I}$ with *Initial Distribution* μ and *Transition Matrix* P if for any $i \in \mathcal{I}$,

- $\mathbf{P}[\, X_0 = i \,] = \mu(i)$.
- The *Markov Property* holds: for all $t \geq 0$ and any $i_0, \ldots, i_{t+1} \in \mathcal{I}$,

$$\mathbf{P}\Big[\, X_{t+1} = i_{t+1} \Big| X_t = i_t, \ldots, X_0 = i_0 \,\Big] = \mathbf{P}\Big[\, X_{t+1} = i_{t+1} \Big| X_t = i_t \,\Big] := P(i_t, i_{t+1}).$$

From the definition one can deduce that (check!)

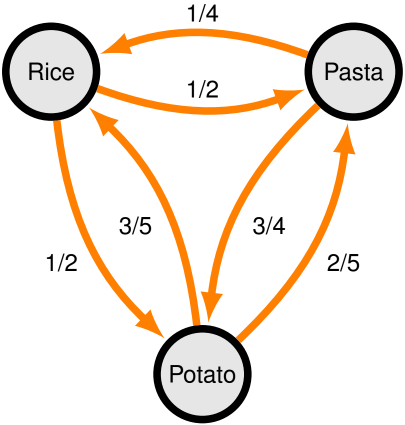
- $\mathbf{P}[\, X_{t+1} = i_{t+1}, X_t = i_t, \ldots, X_0 = i_0 \,] = \mu(i_0) \cdot P(i_0, i_1) \cdots P(i_{t-1}, i_t) \cdot P(i_t, i_{t+1})$
- $\mathbf{P}[\, X_{t+m} = i \,] = \sum_{j \in \mathcal{I}} \mathbf{P}[\, X_{t+m} = i | X_t = j \,] \, \mathbf{P}[\, X_t = j \,]$

If the Markov Chain starts from as single state $i \in \mathcal{I}$ then we use the notation

$$\mathbf{P}_i[X_k = j] := \mathbf{P}[\, X_k = j | X_0 = i \,].$$

What does a Markov Chain Look Like?

Example : the carbohydrate served with lunch in the college cafeteria.



This has transition matrix:

$$
\begin{array}{cc}
 & \begin{array}{ccc} \text{Rice} & \text{Pasta} & \text{Potato} \end{array} \\
P = & \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \\ 3/5 & 2/5 & 0 \end{bmatrix}
\end{array}
\begin{array}{l} \text{Rice} \\ \text{Pasta} \\ \text{Potato} \end{array}
$$

Transition Matrices

The *Transition Matrix* P of a Markov chain (μ, P) on $\mathcal{I} = \{1, \ldots n\}$ is given by

$$P = \begin{pmatrix} P(1,1) & \ldots & P(1,n) \\ \vdots & \ddots & \vdots \\ P(n,1) & \ldots & P(n,n) \end{pmatrix}.$$

- $\rho^t(i)$: probability the chain is in state i at time t.
- $\rho^t = (\rho^t(0), \rho^t(1), \ldots, \rho^t(n))$: *State vector* at time t (Row vector).
- Multiplying ρ^t by P corresponds to advancing the chain one step:

$$\rho^{t+1}(i) = \sum_{j \in \mathcal{I}} \rho^t(j) \cdot P(j,i) \qquad \text{and thus} \qquad \rho^{t+1} = \rho^t \cdot P.$$

- The Markov Property and line above imply that for any $k, t \geq 0$

$$\rho^{t+k} = \rho^t \cdot P^k \qquad \text{and thus} \qquad P^k(i,j) = \mathbf{P}[\, X_k = j | X_0 = i \,].$$

Thus $\rho^t(i) = (\mu P^t)(i)$ and so $\rho^t = \mu P^t = (\mu P^t(1), \mu P^t(2), \ldots, \mu P^t(n))$.

Outline

Stochastic Processes and Markov Chains

Stopping and Hitting Times

Irreducibility and Stationarity

Stopping and Hitting Times

A non-negative integer random variable τ is a *Stopping Time* for $(X_i)_{i\geq 0}$ if for every $n \geq 0$ the event $\{\tau = n\}$ depends only on $X_0, \dots, X_n$.

Example - College Carbs Stopping times:

✓ "We had Pasta yesterday"

× "We are having Rice next Thursday"

For two states $x, y \in \mathcal{I}$ we call $h(x, y)$ the *Hitting Time* of y from x:

$$h(x,y) := \mathbf{E}_x[\tau_y] = \mathbf{E}[\,\tau_y | X_0 = x\,] \quad \text{where } \tau_y = \inf\{t \geq 0 : X_t = y\}.$$

For $x \in \mathcal{I}$ the *First Return Time* $\mathbf{E}_x\left[\tau_x^+\right]$ of x is defined

$$\mathbf{E}_x\left[\tau_x^+\right] = \mathbf{E}\left[\,\tau_x^+ | X_0 = x\,\right] \quad \text{where } \tau_x^+ = \inf\{t \geq 1 : X_t = x\}.$$

Comments

- Notice that $h(x, x) = \mathbf{E}_x[\tau_x] = 0$ whereas $\mathbf{E}_x\left[\tau_x^+\right] \geq 1$.
- For any $y \neq x$, $h(x, y) = \mathbf{E}_x\left[\tau_y^+\right]$.
- Hitting times are the solution to the set of linear equations:

$$\mathbf{E}_x\left[\tau_y^+\right] \overset{\text{Markov Prop.}}{=} 1 + \sum_{z \in \mathcal{I}} \mathbf{E}_z[\tau_y] \cdot P(x, z) \qquad \forall x, y \in V.$$

Outline

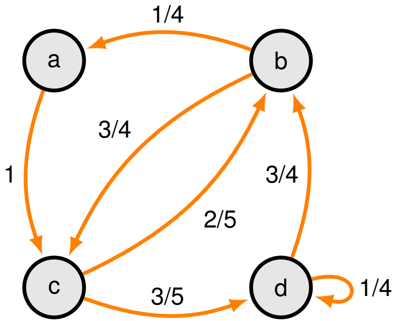
Stochastic Processes and Markov Chains

Stopping and Hitting Times

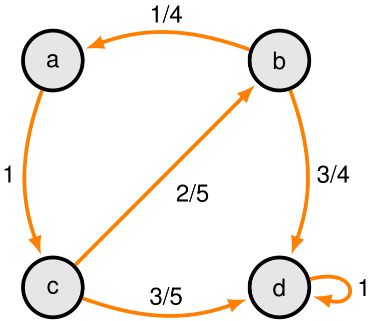
Irreducibility and Stationarity

Irreducible Markov Chains

A Markov chain is *Irreducible* if for every pair of states $(i, j) \in \mathcal{I}^2$ there is an integer $m \geq 0$ such that $P^m(i, j) > 0$.



✓ irreducible



× not-irreducible (thus reducible)

Finite Hitting Theorem

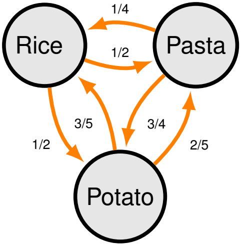
For any states x and y of a finite irreducible Markov chain $\mathbf{E}_x\left[\tau_y^+\right] < \infty$.

Stationary Distribution

A probability distribution $\pi = (\pi(1), \ldots, \pi(n))$ is the *Stationary Distribution* of a Markov chain if $\pi P = \pi$, i.e. π is a left eigenvector with eigenvalue 1.

College carbs example:

$$\underset{\pi}{\left(\frac{4}{13}, \frac{4}{13}, \frac{5}{13}\right)} \cdot \underset{P}{\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \\ 3/5 & 2/5 & 0 \end{pmatrix}} = \underset{\pi}{\left(\frac{4}{13}, \frac{4}{13}, \frac{5}{13}\right)}$$



A Markov chain reaches *Equilibrium* if $\rho^t = \pi$ for some t. If equilibrium is reached it *Persists*: If $\rho^t = \pi$ then $\rho^{t+k} = \pi$ for all $k \geq 0$ since

$$\rho^{t+1} = \rho^t P = \pi P = \pi = \rho^t.$$

Existence and Uniqueness of a Positive Stationary Distribution

Let P be finite, irreducible M.C., then there exists a unique probability distribution π on $\mathcal{I}$ such that $\pi = \pi P$ and $\pi(x) = 1/\mathbf{E}_x\left[\tau_x^+\right] > 0$, $\forall x \in \mathcal{I}$.

Proof: [Existence] Fix $z \in \mathcal{I}$ and define $\mu(y) = \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_t = y, \tau_z^+ > t\right]$, this is the expected number of visits to y before returning to z. For any state y, we have $0 < \mu(y) \leq \mathbf{E}_z\left[\tau_z^+\right] < \infty$ since P is irreducible. To show $\mu P = \mu$

$$\mu P(y) = \sum_{x \in \mathcal{I}} \mu(x) \cdot P(x,y) = \sum_{x \in \mathcal{I}} \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_t = x, \tau_z^+ > t\right] \cdot P(x,y)$$

$$= \sum_{x \in \mathcal{I}} \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_t = x, X_{t+1} = y, \tau_z^+ > t\right]$$

$$= \sum_{t=0}^{\infty} \sum_{x \in \mathcal{I}} \mathbf{P}_z\left[X_t = x, X_{t+1} = y, \tau_z^+ > t\right] = \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_{t+1} = y, \tau_z^+ > t\right]$$

$$= \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_{t+1} = y, \tau_z^+ > t+1\right] + \mathbf{P}_z\left[X_{t+1} = y, \tau_z^+ = t+1\right]$$

$$= \mu(y) - \mathbf{P}_z\left[X_0 \overset{(a)}{=} y, \tau_z^+ > 0\right] + \sum_{t=0}^{\infty} \mathbf{P}_z\left[X_{t+1} \overset{(b)}{=} y, \tau_z^+ = t+1\right] = \mu(y).$$

Where (a) and (b) are 1 if $y = z$ and 0 otherwise so cancel. Divide μ though by $\sum_{x \in \mathcal{I}} \mu(x) < \infty$ to turn it into a probability distribution π. $\square$

Existence and Uniqueness of a Positive Stationary Distribution

Let P be finite, irreducible M.C., then there exists a unique probability distribution π on $\mathcal{I}$ such that $\pi = \pi P$ and $\pi(x) = 1/\mathbf{E}_x\left[\tau_x^+\right] > 0$, $\forall x \in \mathcal{I}$.

Proof: [Uniqueness] Assume P has a stationary distribution μ and let $\mathbf{P}[X_0 = x] = \mu(x)$. We shall show μ is uniquely determined

$$
\begin{aligned}
\mu(x) \cdot \mathbf{E}_x\left[\tau_x^+\right] &\stackrel{\text{Hw1}}{=} \mathbf{P}[X_0 = x] \cdot \sum_{t \geq 1} \mathbf{P}\left[\tau_x^+ \geq t \mid X_0 = x\right] \\
&= \sum_{t \geq 1} \mathbf{P}\left[\tau_x^+ \geq t, X_0 = x\right] \\
&= \mathbf{P}[X_0 = x] + \sum_{t \geq 2} \mathbf{P}[X_1 \neq x, \ldots, X_{t-1} \neq x] - \mathbf{P}[X_0 \neq x, \ldots, X_{t-1} \neq x] \\
&\stackrel{(a)}{=} \mathbf{P}[X_0 = x] + \sum_{t \geq 2} \mathbf{P}[X_0 \neq x, \ldots, X_{t-2} \neq x] - \mathbf{P}[X_0 \neq x, \ldots, X_{t-1} \neq x] \\
&\stackrel{(b)}{=} \mathbf{P}[X_0 = x] + \mathbf{P}[X_0 \neq x] - \lim_{t \to \infty} \mathbf{P}[X_0 \neq x, \ldots, X_{t-1} \neq x] \stackrel{(c)}{=} 1.
\end{aligned}
$$

A sum S is Telescoping if

$$S = \sum_{i=0}^{n-1} a_i - a_{i+1} = a_0 - a_n.$$

Equality (a) follows as μ is stationary, equality (b) since the sum is telescoping and (c) by Markov's inequality and the Finite Hitting Theorem. $\square$