# MA262: Introduction to Stochastic Processes

Naman Mishra

January 2024

## Contents



**Lecture 01:** Thu 04 Jan '24

## 0 The Course

**Texts:**

- *Markov Chains*, J. R. Norris
- *Introduction to Stochastic Processes*, Hoel, Port, Stone
- Karlin and Taylor

**Grading:**

(20%) 2 quizzes

(30%) 1 midterm

(50%) Final

## 1 Discrete time Markov Chains

**Definition 1.1.** Let $S$ be a state set (countable). A matrix $P = (p_{xy}; x, y \in S)$ is called a *stochastic matrix* if $p_{xy} \geq 0$ for all $x, y \in S$ and $\sum_{y \in S} p_{xy} = 1$ for all $x \in S$.

**Definition 1.2.** Let $S$ be a state set, $P = (p_{xy})$ a stochastic matrix, and $\mu_0$ a probability distribution on $S$, *i.e.*, $\mu_0(x) \geq 0$ for all $x \in S$ and $\sum_{x \in S} \mu_0(x) = 1$.

Suppose $X_0, X_1, \ldots$ are random variables defined on the same probability space taking values in $S$. Then $(X_n; n \geq 0)$ is called a Markov chain with initial distribution $\mu_0$ and transition matrix $P$, notated $MC(\mu_0, P)$, if $X_0$ has distribution $\mu_0$ and for all $n \geq 0$,

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = p_{x_n x_{n+1}}$$

whenever $P(X_n = x_n, \ldots, X_0 = x_0) > 0$.

*Notation.* Whenever writing $P(X_n \in A \mid (X_0, \ldots, X_{n-1}) \in B)$, it is understood that $P((X_0, \ldots, X_{n-1}) \in B) > 0$.

**Theorem 1.3.** $(X_n; 0 \leq n \leq N)$ is $MC(\mu_0, P)$ iff

$$P(X_0 = x_0, \ldots, X_N = x_N) = \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{N-1} x_N}$$

for all $x_0, \ldots, x_N \in S$.

*Proof.* Both directions are proven by induction.

Suppose $(X_n; 0 \leq n \leq N)$ is $MC(\mu_0, P)$. $P(X_0 = x_0) = \mu_0(x_0)$. If $P(X_0 = x_0) > 0$, then $P(X_0 = x_0, X_1 = x_1) = \mu_0(x_0) p_{x_0 x_1}$. If $P(X_0 = x_0) = 0$, then $P(X_0 = x_0, X_1 = x_1) \leq P(X_0 = x_0) = 0$, and so $P(X_0 = x_0, X_1 = x_1) = 0 = \mu_0(x_0) p_{x_0 x_1}$.

Suppose

$$P(X_0 = x_0, \ldots, X_n = x_n) = \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n}$$

for all $x_0, \ldots, x_n \in S$. Then

$$\begin{aligned} P(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) &= \sum_{x_n \in S} P(X_0 = x_0, \ldots, X_n = x_n) \\ &= \sum_{x_n \in S} \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n} \\ &= \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n} \end{aligned}$$

Use induction to show that for all $1 \leq i \leq N$,

$$P(X_0 = x_0, \ldots, X_i = x_i) = \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{i-1} x_i}$$

and $P(X_0 = x_0)$. This allows us to deduce that

$$P(X_{i+1} = x_{i+1} \mid X_0 = x_0, \ldots, X_i = x_i) = p_{x_i x_{i+1}}. \qquad \square$$

**Theorem 1.4** (Strong Law of Large Numbers)**.** Suppose $Z_1, Z_2, \ldots$ are iid $\mathbb{R}$-valued random variables and $E[Z_1]$ exists. Then

$$\frac{Z_1 + \cdots + Z_n}{n} \to E[Z_1]$$

as $n \to \infty$, that is,

$$P\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{Z_1(\omega) + \cdots + Z_n(\omega)}{n} = E[Z_1]\right\} = 1.$$

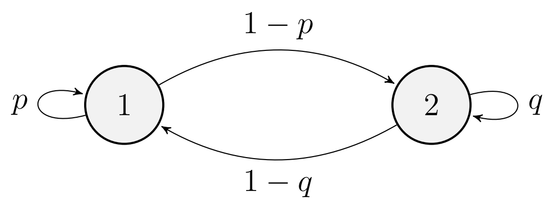
**Theorem 1.5** (Weak Law of Large Numbers)**.**

**Theorem 1.6** (Central Limit Theorem)**.** Suppose $Z_1, Z_2, \ldots$ are iid $\mathbb{R}$-valued random variables and $E[Z_1^2]$ exists. Then

$$\frac{\sqrt{n}}{\sqrt{V(Z_1)}}\left(\frac{Z_1 + \cdots + Z_n}{n} - E[Z_1]\right) \to N(0,1).$$

*Examples.*

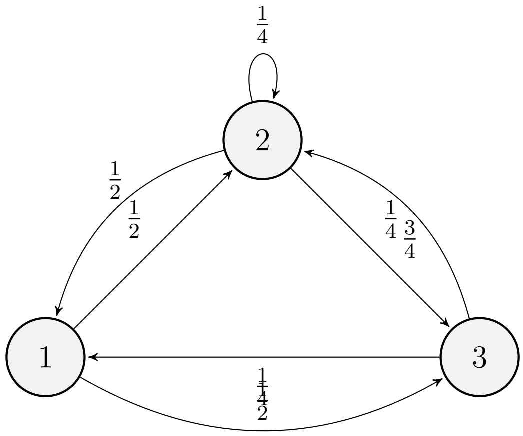
- A two-state Markov chain.




  This corresponds to the matrix

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

- A three-state Markov chain.




- Simple random walk on $\mathbb{Z}$. Staring from 0, at each step, move right with probability $p$ and left with probability $q = 1 - p$. $P(X_{n+1} = x + 1 \mid X_n = x) = p$ and $P(X_{n+1} = x - 1 \mid X_n = x) = q$. All other probabilities are 0.

  Such a simple random walk is called symmetric if $p = q = \frac{1}{2}$. A special case is where $\mu_0 = \delta_x$ for some $x \in \mathbb{Z}$, where $\delta_x$ is the Krönecker delta.

  Aside: Suppose $Z_1, \dots, Z_k$ are random variables taking values in a state set $S$ defined on a probability space $(\Omega, \mathcal{F}, P)$. and $\tilde{Z}_1, \dots, \tilde{Z}_k$ are rvs taking values in a state set $S$ defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. Then $(Z_1, \dots, Z_k)$ and $(\tilde{Z}_1, \dots, \tilde{Z}_k)$ are said to be identically distributed if

$$P(Z_1 = x_1, \dots, Z_k = x_k) = P(\tilde{Z}_1 = x_1, \dots, \tilde{Z}_k = x_k).$$

  This is notated as

$$(Z_1, \dots, Z_k) \stackrel{d}{=} (\tilde{Z}_1, \dots, \tilde{Z}_k).$$

  Suppose that $Y_1, Y_2, \dots$ are iid $\begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$. We have that $(X_n; n \geq 0) \stackrel{d}{=} (\sum_{j=1}^{n} Y_j; n \geq 0)$. Then from the weak law of large numbers,

$$\frac{X_n}{n} \to E[Y_1] = 2p - 1.$$

  From the central limit theorem,

$$\frac{X_n - n(p-q)}{\sqrt{n}\sqrt{1-(p-q)^2}} \to N(0, 1).$$

  On a graph, a simple symmetric random walk is a random walk on a graph where each

$$p_{xy} = \begin{cases} \frac{1}{deg(x)} & \text{if } y \sim x, \\ 0 & \text{otherwise.} \end{cases}$$

On $\mathbb{Z}^2$, a simple random walk is given by $p_N$, $p_E$, $p_S$, $p_W$, where $p_N + p_E + p_S + p_W = 1$. At each step, move up with probability $p_N$, right with probability $p_E$, down with probability $p_S$, and left with probability $p_W$.

- Consider a shooting game with 4 modes: $N$ (normal), $D$ (distance), $W$ (windy) and $DW$ (distance and windy). The game changes mode randomly to a mode different from the current mode with direcred graph $K_4$ with some edge weights.

**Theorem 1.7.** If $(X_n; n \geq 0)$ is a DTMC with transition matrix $P$, then

$$P_{\mu_0}(X_n = y) = (\mu_0 P^n)(y).$$

In particular, $P_x(X_n = y) = (P^n)_{x,y} = p_{xy}^{(n)}$.

Here, $\mu_0$ is viewed as a row vector, and $P_{\mu_0}$ is the distribution under the assumption that $X_0 \sim \mu_0$.

*Proof.*

$$\begin{aligned} P_{\mu_0}(X_n = y) &= \sum_{x_0,\dots,x_{n-1} \in S} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = y) \\ &= \sum_{x_0,\dots,x_{n-1} \in S} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) p_{x_{n-1}y} \\ &= (\mu_0 P^{n-1})_y \end{aligned}$$

□

**Theorem 1.8.** Let $(X_n; n \geq 0)$ be $MC(\mu_0, P)$. Then for any $n \geq 0$, $l \geq 1$, $x_n, \dots, x_{n+l} \in S$ and $A \subseteq S^n$,

$$\begin{aligned} &P_{\mu_0}(X_i = x_i, n < i \leq n + l \mid X_n = x_n, (X_0, \dots, X_{n-1}) \in A) \\ &\qquad = P_{x_n}(X_1 = x_{n+1}, \dots, X_l = x_{n+l}) \end{aligned}$$

In other words, conditioning on $X_n = x_n$ and $(X_0, \dots, X_{n-1}) \in A$, the process $(X_n, X_{n+1}, \dots)$ is $MC(\delta_{x_n}, P)$.

*Proof.*

$$
\begin{aligned}
&P(X_{n+l} = x_{n+l}, \ldots, X_n = x_n, (X_0, \ldots, X_{n-1}) \in A) \\
&\qquad = p_{x_n x_{n+1}} \ldots p_{x_{n+l-1} x_{n+l}} \sum_{(x_0, \ldots, x_{n-1}) \in A} \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n} \\
&\qquad = P_{\mu_0}(X_n = x_n, (X_0, \ldots, X_{n-1}) \in A) p_{x_n x_{n+1}} \ldots p_{x_{n+l-1} x_{n+l}}
\end{aligned}
$$

□