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 (at most 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 μ_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 μ_0 and transition matrix P, denoted $MC(\mu_0, P)$, if X_0 has distribution μ_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 only $P((X_0, \ldots, X_{n-1}) \in B) > 0$ is considered.

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 $P(X_0 = x_0, X_1 = x_1) = 0 = \mu_0(x_0) p_{x_0 x_1}$.

Induction: Suppose

$$P_j \coloneqq P(X_0 = x_0, \ldots, X_j = x_j) = \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{j-1} x_j}.$$

If this is zero, so is P_{j+1}, and so it is equal to $\mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{j-1} x_j} p_{x_j x_{j+1}}$. If not, then

$$\begin{aligned} P_{j+1} &= P_j P(X_{j+1} = x_{j+1} \mid X_0 = x_0, \ldots, X_j = x_j) \\ &= P_j p_{x_j x_{j+1}} \\ &= \mu_0(x_0) p_{x_0 x_1} \ldots p_{x_{j-1} x_j} p_{x_j x_{j+1}}, \end{aligned}$$

closing the induction. In particular,

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

Now 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 $x_0, \dots, x_N \in S$. Then for any $x_0, \dots, x_{N-1} \in S$,

$$\begin{aligned} P(X_0 = x_0, \dots, X_{N-1} = x_{N-1}) &= \sum_{x_N \in S} P(X_0 = x_0, \dots, X_N = x_N) \\ &= \sum_{x_N \in S} \mu_0(x_0) p_{x_0 x_1} \dots p_{x_{N-2} x_{N-1}} p_{x_{N-1} x_N} \\ &= \mu_0(x_0) p_{x_0 x_1} \dots p_{x_{N-2} x_{N-1}}. \end{aligned}$$

We have by backwards induction that for all $1 \leq i \leq N$,

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

and $P(X_0 = x_0) = \mu_0(x_0)$. This allows us to deduce that

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

by definition of conditional probability. □

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

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

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

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

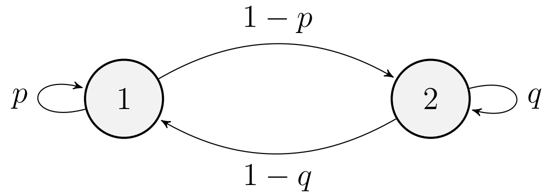
Theorem 1.5 (Weak Law of Large Numbers)**.**

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

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

Examples.

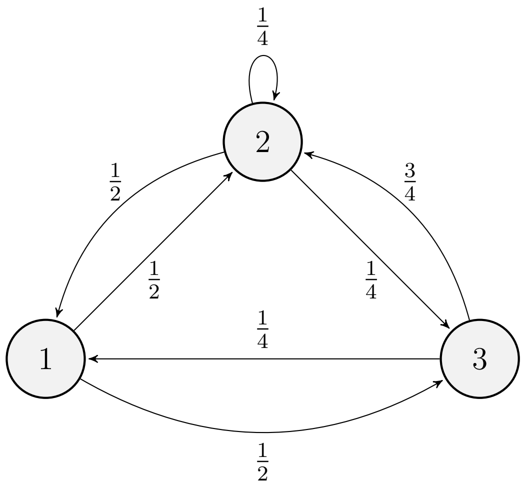
- A two-state Markov chain.



This corresponds to the matrix

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

- A three-state Markov chain.



This has transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 3/4 & 0 \end{pmatrix}.$$

- 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 δ_x is the Krönecker delta.

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

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

  This is notated as

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

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

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

From the central limit theorem,

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

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, μ_0 is viewed as a row vector, and P_{μ_0} is the distribution under the assumption that $X_0 \sim \mu_0$. Also, P_x is under the assumption that $\mu_0 = \delta_x$.

Proof.

$$\begin{aligned} P_{\mu_0}(X_n = y) &= \sum_{\substack{x_j \in S \\ 0 \leq j < n}} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = y) \\ &= \sum_{\substack{x_j \in S \\ 0 \leq j < n}} (\mu_0)_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} y} \\ &= (\mu_0 P^n)_y \end{aligned}$$

□

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 $A \subseteq S^n$,

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

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.

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

□

Lecture 02: Tue 9 Jan '24

Theorem 1.9 (Schur's Lemma)**.**

Lecture 03: Thu 11 Jan '24

Proof. First consider the case where $|S| = 2$. Let $S = \{s_1, s_2\}$. We know that $g_S = as_1 + bs_2$ where $a, b \in \mathbb{Z}$. Now

$$mg_S = (ma - ks_2)s_1 + (mb + ks_1)s_2$$

Choose $k \in \mathbb{N}$ such that $0 \leq ma - ks_2 < s_2$. We can write $mg_S = \tilde{a}s_1 + \tilde{b}s_2$ where $0 \leq \tilde{a} < s_2$.

If $m_S = \frac{s_1 s_2}{g_S} + 1$, then for all $m \geq m_S$, $s_1 s_2 < mg_S = \tilde{a}s_1 + \tilde{b}s_2 < s_1 s_2 + \tilde{b}s_2$ and so $\tilde{b} > 0$.

Now let $S = \{s_1, s_2, \dots, s_{l+1}\}$ and $F = \{s_1, s_2, \dots, s_l\}$.

Claim: $\tilde{g}_S = \gcd(g_F, s_{l+1})$ is equal to g_S.
Proof of claim: Huh?

If $m \geq m_{g_F, s_{l+1}} + \frac{m_F g_F}{g_S}$, then

$$\begin{aligned}
m g_S - m_F g_F &= \left(m - \frac{m_F g_F}{g_S} \right) g_S \\
&= a g_F + b s_{l+1} \text{ for some } a, b \in \mathbb{Z}_{\geq 0} \\
m g_S &= (a + m_F) g_F + b s_{l+1} \\
&= \sum_{i=1}^{l} a_i s_i + b s_{l+1}
\end{aligned}$$

where all coefficients are non-negative integers. This closes the induction.

□

Definition 1.10 (Extended reals)**.** We define

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

to be the *extended reals*.

Definition 1.11 (Filtration)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A collection $(\mathcal{F}_n)_{n \in \mathbb{N}}$ of σ-algebras is called a *filitration* if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$ for all $n \in \mathbb{N}$.

Definition 1.12 (Natural filtration)**.** Let $(X_n)_{n \geq 0}$ be a sequence of S-values random variables defined on $(\Omega, \mathcal{F}, P)$. For $n \geq 0$, define

$$\begin{aligned}
\mathcal{F}_n &= \left\{ (X_0, X_1, \ldots, X_n)^{-1}(A) \mid A \in S^{n+1} \right\} \\
&= \sigma(X_0, X_1, \ldots, X_n)
\end{aligned}$$

Here, $(X_0, \ldots, X_n)^{-1}(A) = \{\omega \in \Omega \mid (X_0(\omega), \ldots, X_n(\omega)) \in A\}$. This sequence of σ-algebras is called the *natural filtration* of $(X_n)_{n \geq 0}$.

Why is this a σ-algebra? The empty set is in $\mathcal{F}_n$ because $\emptyset \in S^{n+1}$. The complement of any set in $\mathcal{F}_n$ is in $\mathcal{F}_n$ because $(X_0, \ldots, X_n)^{-1}(A^c) = (X_0, \ldots, X_n)^{-1}(A)^c$.

Why is $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$? For any $A \in S^{n+1}$, we have

$$(X_0, \ldots, X_n)^{-1}(A) = (X_0, \ldots, X_n, X_{n+1})^{-1}(A \times S).$$

Definition 1.13 (Stopping time)**.** Suppose $(X_n)_{n\geq 0}$ is a sequence of S-valued random variables on $(\Omega, \mathcal{F}, P)$ with natural filtration $(\mathcal{F}_n)_{n\geq 0}$.

Then $\tau\colon \Omega \to \mathbb{N} \cup \{\infty\}$ is called a *stopping time* with respect to $(\mathcal{F}_n)_{n\geq 0}$ if for all $n \in \mathbb{N}$,

$$\{\tau \leq n\} \in \mathcal{F}_n.$$

This is equivalent to saying that for all $n \in \mathbb{N}$,

$$\mathbf{1}_{\{\tau\leq n\}} = \mathbf{1}_{\{(X_0,\ldots,X_n)\in A\}} \quad \text{for some } A \in S^{n+1}.$$

Consider the simple randow walk $(X_n)_{n\geq 0}$ on $\mathbb{Z}$. Then the event that the hitting time of 10 is at most n is

$$\{T_{10} \leq n\} = \bigcup_{i=1}^{n} \{X_i = 10\}.$$

Examples.

- Let $(X_n)_{n\geq 0}$ be an S-valued stochastic process and let $A \subseteq S$. Let $T_A \coloneqq \inf\{n \geq 1 \mid X_n \in A\}$, where we take $\inf \varnothing$ to be $+\infty$ by convention. Then T_A is a stopping time with respect to the natural filtration associated with $(X_n)_{n\geq 0}$. That is, for all $n \in \mathbb{N}$,

$$\{T_A \leq n\} = \bigcup_{i=1}^{n} \{X_i \in A\} \in \mathcal{F}_n.$$

- SRW(p) started at the origin. Then $L = \sup\{n \geq 1 \mid X_n < 7\}$ is NOT a stopping time.

Now if τ is a stepping time, then $\{\tau = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. This is because

$$\{\tau = n\} = \{\tau \leq n\} \cap \{\tau \leq n-1\}^c$$

where both sets are in $\mathcal{F}_n$.

Proposition 1.14. If τ_1 and τ_2 are stopping times, then so are $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$ and $\tau_1 + \tau_2$.

Proof. We have

$$
\begin{aligned}
\{\tau_1 \wedge \tau_2 \le n\} &= \{\tau_1 \le n\} \cup \{\tau_2 \le n\} \\
\{\tau_1 \vee \tau_2 \le n\} &= \{\tau_1 \le n\} \cap \{\tau_2 \le n\} \\
\{\tau_1 + \tau_2 \le n\} &= \bigcup_{i=0}^{n} \{\tau_1 \le i\} \cap \{\tau_2 \le n-i\}
\end{aligned}
$$

□

Problem 1.1. Give an example of two stopping times τ_1 and τ_2 such that $\Pr(\tau_1 \le \tau_2) = 1$ but $\tau_2 - \tau_1$ is not a stopping time.

Solution. Consider the SSRW(p) started at the origin, with

$$
\begin{aligned}
\tau_1 &= \inf\{n \ge 1 \mid X_n = 10\} \\
\tau_2 &= \inf\{n \ge \tau_1 \mid X_n = 0\}.
\end{aligned}
$$

Theorem 1.15 (Strong Markov property)**.** Let $(X_n)_{n\ge 0}$ be in $MC(\mu_0, p)$, and let τ be a stopping time. Let $A = [0, \infty)$. Then

$$
\begin{aligned}
\Pr_{\mu_0}(X_{\tau+1} = x_1, \dots, X_{\tau+n} = x_n \mid \tau \in A, X_\tau = x) \\
= \Pr_x(X_1 = x_1, \dots, X_n = x_n)
\end{aligned}
$$

Proof. The SMP is equivalent to

$$
E_{\mu_0}[f((X_{\tau+j})_{j\ge 0}) \mid \tau \in A, X_\tau = x] = E_x\left[f\left((X_j)_{j\ge 0}\right)\right]
$$

for any bounded function $f\colon S^\infty \to \mathbb{R}$. Now

$$
\begin{aligned}
&P_{\mu_0}(X_{\tau+1} = x_1, \dots, X_{\tau+n} = x_n, \tau \in A, X_\tau = x) \\
&\quad = \sum_{m\in A} P_{\mu_0}(X_{m+1} = x_1, \dots, X_{m+n} = x_n, \tau = m, X_m = x) \\
&\quad = \sum_{m\in A} P_{\mu_0}(\tau = m, X_m = x) P_x(X_1 = x_1, \dots, X_n = x_n) \\
&\quad = P_x(X_1 = x_1, \dots, X_n = x_n) P_{\mu_0}(\tau \in A, X_\tau = x)
\end{aligned}
$$

□

Definition 1.16. Suppose X takes values in $\mathbb{N} \cup \{\infty\}$, and let $p_k := \Pr(X = k)$, $k \in \mathbb{N} \cup \{\infty\}$. Then the *probability generating function* of X is defined as

$$\begin{aligned} G_X(s) &= p_0 + p_1 s + p_2 s^2 + \dots, \quad s \in (-1, 1) \\ &= E[s^X] \end{aligned}$$

where we take $s^\infty = 0$ for $|s| < 1$.

Remarks. The left limit $G_X(1^-) = \lim_{s \uparrow 1} G_X(s) = 1 - p_\infty$.

If $p_\infty > 0$, then $EX = \infty$. Otherwise, $EX = \sum_{k=0}^{\infty} k p_k = G'_X(1^-)$.

Let X, Y be two random variables taking values in $\mathbb{N} \cup \infty$ and $G_X(s) = G_Y(s) \forall\, s \in (-1, 1)$, then $X \overset{d}{=} Y$.

Problem 1.2. Suppose $(X_n)_{n \geq 0}$ is an SRW(p) on $\mathbb{Z}$ started at the origin. Find $\mathcal{G} = G_{T_{-1}}$.

Solution.

$$\begin{aligned} G(s) &= E_0[s^{T_{-1}}] \\ &= pE_0[s^{T_{-1}} \mid X_1 = 1] + qE_0[s^{T_{-1}} \mid X_1 = -1] \\ &= pE_1[s^{T_{-1}}] + qs \\ &= pE_1[s^{1+T_{-1}}] + qs \\ &= psE_0[s^{T_{-2}}] + qs \end{aligned}$$

Since $s^\infty = 0$ by our convention, we have

$$\begin{aligned} E_0[s^{T_{-2}}] = E_0[s^{T_{-2}} \mathbf{1}_{T_{-1} < \infty}] & \qquad = \sum_m E_0[s^{T_{-2}} \mathbf{1}_{T_{-1} = m}] \\ &= \sum_m \Pr_0(T_{-1} = m) E_{-1}[s^{m + T_{-2}}] \\ &= \sum_m \Pr_0(T_{-1} = m) s^m E_0[s^{T_{-1}}] \\ &= G(s)^2 \end{aligned}$$

Thus

$$\begin{aligned} G(s) &= psG(s)^2 + qs \\ G(s) &= \frac{1 \pm \sqrt{1 - 4pqs}}{2ps} \end{aligned}$$

Claim: $G(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}$ for all $s \in (-1, 1) \setminus \{0\}$.