Definition of Markov chain

Lecturer: Partha S. Dey  Date: Aug 29, 2017  Scribe: Lan Wang <lanwang2@illinois.edu>

4.1 Stochastic Matrices

Definition 4.1. (a) A square matrix $P = ((p_{ij}))_{i,j \in I}$ is called a stochastic matrix if

i) $p_{ij} \geq 0 \ \forall \ i, j \in I$ and

ii) $\sum_{j \in I} p_{ij} = 1 \ \forall \ i \in I$.

(b) A square matrix $P$ is called a sub-stochastic matrix if it satisfies (i) and (iii) $\sum_{j=1}^{n} p_{ij} \leq 1 \ \forall \ i$

(c) A square matrix $P$ is called a doubly-stochastic matrix if $P$ and $P^T$ are both stochastic matrices.

Definition 4.2. A directed graph $G = (V,E)$ is a set of vertices $V$ and collection of directed edges $(i,j) \in E \subset V \times V$ from vertex $i$ to vertex $j$. A weighed graph $(V,E,W)$ is a graph $(V,E)$ with the weight of an edge $e = (i,j)$ given by $w_e = w_{ij}$. We will assume $w_{ij} = 0$ iff $(i,j)$ is not an edge in the graph.

Given a Stochastic matrix $P = ((p_{ij}))_{i,j \in I}$ we can define a weighed directed graph with vertex set $I$ and edge weight of $(i,j)$ given by $p_{ij}$. Note that, $\sum_{j \in I} p_{ij} = 1$ for all $i$ implies that the total weight of all the out-edges from the vertex $i$ is 1.

Example 4.3. Given a stochastic matrix

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

then the corresponding edge-weighted directed graph is

![Diagram of a directed graph with weights on edges]
Definition 4.4. (a) A path in a directed graph $G = (V, E)$ is a collection of edges $(i_0, i_1), (i_1, i_2), ..., (i_{n-1}, i_n)$ where $i_0, i_1, ..., i_n \in V$. Moreover, the length of a path is the number of edges in this path and the weight of a path is defined as $\prod_{\text{edges in the path}} \text{edge weight}$.

(b) A cycle is a path with the same starting and ending vertices.

Example 4.5. (a) $\{(2, 3), (3, 2), (2, 3), (3, 2)\}$ is a path of $G$ with length (i.e., number of edges) $= 4$.

(b) $\{(2, 1), (1, 1), (1, 1)\}$ is a path of $G$ with length $= 3$ and weight $= 1/2 \cdot 1 \cdot 1 = 1/2$.

(c) For example 4.3, to reach 1 from 2, there are 3 paths:

$(2, 1), (1, 1)$
$(2, 2), (2, 1)$
$(2, 3), (3, 1)$

The total weight of these paths is

$$(P \cdot P)_{21} = \sum_{k=1}^{n} p_{2k}p_{k1} = p_{21}p_{11} + p_{22}p_{21} + p_{23}p_{31}$$

Exercise 4.1. Show that $(P^k)_{ij} = \text{Total weight of all paths from } i \text{ to } j \text{ of length } k$.

Remark 4.6. $P$ is a stochastic matrix if and only if $P \cdot \hat{1} = \hat{1}$ where $\hat{1} = (1, 1, ..., 1)^T$. Thus, if $P$ is stochastic, then so is $P^k$ for any positive integer $k$.

Theorem 4.7. For a stochastic matrix $P$, the following statements hold:

i) $\text{spec}(P) \subseteq B(0, 1)$

ii) If $p_{ii} > 0$ for all $i$, then $\text{spec}(P) \subseteq B(0, 1) \cup \{1\}$.

Proof. i) For all $i$, $R_i = \sum_{j \neq i} |p_{ij}| = 1 - p_{ii}$. Suppose $z$ is any eigenvalue of $P$. Then by Gershgorin circle theorem, $z \in B(p_{ii}, 1 - p_{ii})$ and thus, $|z| \leq 1$.

ii) Follows from the fact that $B(p_{ii}, 1 - p_{ii}) \subseteq B(0, 1) \cup \{1\}$ if $p_{ii} > 0$.

4.2 Basic Markov chain Monte Carlo (MCMC)

Suppose $X$ is a random variable with values in $I$, we want to estimate $E(X)$.

Theorem 4.8. (Law of large numbers) Let $X_1, X_2, ..., X_n$ be i.i.d with $E(X_1) < \infty$, then $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{n \to \infty} E(X_1)$ in probability and almost everywhere.
Theorem 4.9. (Central limit theorem) If $\text{Var}(X) = \sigma^2 < \infty$, then $\frac{\sqrt{n}}{\sigma} (X_n - \mathbb{E}X) \xrightarrow{n \to \infty} N(0,1)$ in distribution.

Definition 4.10. Given a probability space $(\Omega, \mathcal{F}, P)$, define $\mathbb{N}$ := Time index set and $I$ := Possible values of the random variables. Then a stochastic process indexed by $\mathbb{N}$ is a collection of random variables $(X_i(\omega))_{i \in \mathbb{N}}$ taking values in $I$. Moreover, if $\mathbb{N} = \{0, 1, 2, \ldots\}$, then we get discrete time stochastic process.

Definition 4.11. A discrete-time Markov Chain with initial distribution $\lambda$ and transition matrix $P$ on the state space $I$ is a discrete time stochastic process $\{X_0, X_1, \ldots\}$ indexed by $\mathbb{N} = \{0, 1, \ldots\}$ s.t.

i) $X_0$ has distribution $\lambda$, i.e., $P(X_0 = i) = \lambda_i \forall i \in I$ and

ii) given $X_n = i_n$, $X_{n+1}$ is independent of $X_0, X_1, \ldots, X_{n-1}$ and has distribution $(p_{ij})_{j \in I}$, i.e., $P(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{ij}$.

We will write $(X_i)_{i \geq 0} \sim \text{Markov}(\lambda, P)$. 