stat333 Notes

Thaqib M

June 3, 2022

Linear Algebra

Matrix multiplication

If A is a $n \times m$ matrix and B is a $m \times k$ matrix then the matrix AB of dim $n \times k$ is defined by:

$$[AB]_{xy} = \sum_{\text{all}(z)} A_{xz} B_{zy}$$

Inner Product

The inner product (dot product) of 2 vectors $\vec{a}, \vec{b}$ in $\mathbb{R}^n$ is defined as

$$\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = \langle \vec{\mathbf{a}}, \vec{\mathbf{b}} \rangle = \sum_{k=1}^{n} a_k b_k$$

Eigenvalues and Eigenvectors

We can find eigenvalues by solving for the roots of the characteristic polynomial of the matrix $\mathbf{A}$.

$$\det(\mathbf{A} - tI_n) = 0$$

Where I_n is the $n \times n$ identity matrix. Then for each eigenvalue $t = c$ we can solve the system of linear equations

$$(\mathbf{A} - cI_n)\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$$

$\vec{\boldsymbol{x}}$ will be an eigenvector of $\mathbf{A}$.

Assignment Theorems

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f(Y) \cdot \mathbb{E}(X \mid Y)] \tag{hw1q6}$$

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) \tag{hw2q5}$$

1 Week 1

1.1 Definition (Stochastic Process). Let $(X_t)_{t\in T}$ be a collection of random variables this is called a Stochastic Process. T is the *index set*.

1.2 Example (Simple Random Walk on $\mathbb{Z}$)**.** Let $X_i \sim$ iid where $X_i \in \{-1, 1\}$ with

$$P(X_i = 1) = \frac{1}{2}$$
$$P(X_i = -1) = \frac{1}{2}$$

now let

$$S_n = \sum_{i=0}^{n} X_i$$

Then $(S_i)_{k=0}^{\infty}$ is a stochastic process.

1.3 Definition (Transition Probability). Given $(X_s)_{s\leq t}$ we need the probability for X_{t+1}.

$$P(X_{(t+1)} = x_{t+1}|X_1 = x_1, X_2 = x_2, \dots X_t = x_t)$$

1.4 Note. Conditional Probability Properties

$$P(A|B) = \frac{P(AB)}{P(B)}\ P(B) > 0$$
$$P(ABC) = P(A|BC) \cdot P(B|C) \cdot P(C)$$

1.5 Example. Transition Probabilities for SRW on $\mathbb{Z}^d$

$$P\left(\|X_{t+1} - X_t\| \mid (X_s)_{s\leq t}\right) = \frac{1}{2d}$$

1.1 Markov Chains

1.6 Definition (Markov Property)**.** A process has the Markov property if:

$$P(X_{t+1} = x_{t+1} \mid (X_s)_{s \le t}) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

(Next outcome only depends on the previous outcome)

1.7 Note (Markov Chain)**.** A stochastic process that satisfies the Markov property is called a Markov chain.

1.8 Definition (Time Homogeneous Markov Chain)**.** A Markov Chain is called time homogeneous if the following is true

$$P(X_{t+1} = j \mid X_t = i) = P(X_1 = j \mid X_0 = i)$$

1.9 Definition (Stochastic Matrix)**.** A matrix $\mathbf{P}$ is called stochastic if

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \ddots & \end{pmatrix}$$

$$0 \le p_{ij} \le 1$$

$$\sum_{all(j)} p_{i_0 j} = 1 \text{ for fixed } i_0$$

1.10 Definition (Transition Matrix)**.** Let $\mathbf{P}$ be a Stochastic matrix and let p_{ij} = value in $i-$th row and $j-$th column. We define p_{ij} as

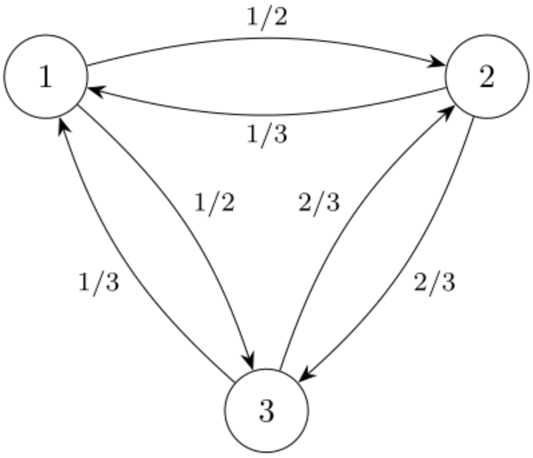
$$p_{ij} = P(X_t = j \mid X_{t-1} = i)$$

(probability of going from state i to state j in the chain).
This is called the transition matrix for $(X_t)_{t \in T}$.

1.11 Example. Transition Matrix Consider this transition matrix The transition matrix for this Markov Chain is

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix} \end{array}$$

this can be visualized as:



1.1.1 Multistep Transition Probabilities

1.12 Definition.

$$[P(n, n+m)]_{xy} = P(X_{n+m} = y \mid X_n = x)$$

1.13 Theorem. Multistep Transition Probability Matrix Let $(X_t)_{t\in T}$ be a stochastic process satisfying the Markov property and be *time homogeneous* and let $\mathbf{P}$ be the transition matrix.

$$[P(n, n+m)]_{xy} = \mathbf{P}^m_{xy}$$

1.14 Lemma.

$$[P(n, m+1+n)]_{xy} = \sum_{\text{all}(z)} [P(n, m+n)]_{xz} P_{zy}$$

Proof. To go from state $x \to y$ we must add up all probabilities of going to an intermediate state $\mathbf{z}$, $x \to \mathbf{z} \to y$ we add possibilities of $\mathbf{z}$.

$[P(n, m+1+n)]_{xy} = P(X_{m+1+n} = y \mid X_n = x)$

$= \sum_{\text{all}(z)} P(X_{m+1+n} = y, X_{n+m} = z \mid X_n = x)$ Marginal probability function (stat240)

$= \sum_{\text{all}(z)} P(X_{m+1+n} = y \mid X_{n+m} = z, X_n = x) P(X_{n+m} = z \mid X_n = x)$ conditional probability

Since X_t satisfies the Markov property we get

$$= \sum_{\text{all}(z)} P(X_{m+1+n} = y \mid X_{n+m} = z) P(X_{n+m} = z \mid X_n = x)$$

By definition we have $P(X_{m+1+n} = y \mid X_{n+m} = z) = P_z y$ and $P(X_{n+m} = z \mid X_n = x) = [P(n, n+m)]_{xz}$. □

Using Lemma 1.14 we can prove the Theorem 1.13.
Since 1.14's result is the definition of matrix multiplication we get

$$[P(n, m+1+n)]_{xy} = [P(n, m+n)P]_{xy}$$

by induction on m with base case $P(n, n+1) = P$ we get

$$[P(n, m+1+n)]_{xy} = \mathbf{P}^m$$

Since RHS does not depend on n we can write $P(n, n+m) = P(m)$ and time homogeneity applies for any m number of steps.

$$P(X_{n+m} = y \mid X_n = x) = P(X_m = y \mid X_0 = x)$$

2 Week 2

2.1 Initial Data

Let $(X_n)_{n \in I}$ be a time homogeneous Markov chain. We denote these by $0, 1, 2, \ldots |I| - 1$. We represent the state space as:

$$\{i_1, i_2, \ldots, i_{|I|}\} = \mathcal{X}$$

Let $\mathbf{P}$ be the transition matrix for this Markov chain.

2.1 Definition (Distribution Row Vector)**.**

$$\mu_j = P(X_0 = i_j)$$

Then the row vector $\vec{\mu}$ of $\dim = 1 \times |I|$ is defined as

$$\vec{\mu} = \left[\mu_1, \mu_2, \ldots, \mu_{|I|}\right]$$

$\vec{\mu}$ is called the distribution of X_0 denoted by $X_0 \sim m\vec{u}$.
The distribution vector for X_n is denoted by $\mu(n)$.

2.2 Theorem. Distribution of X_n The distribution row vector of X_n for a time homogeneous Markov chain is given by μP^n

Proof. Sketch.

$$P(X_n = i_k) = \sum_{j=1}^{|I|} P(X_n = i_k \mid X_0 = i_j) P(X_0 = i_j) = \sum \vec{\mu_j} P_{jk} = [\vec{\mu} P]_k$$

Implies $X_n \sim \vec{\mu} P^n$ □

2.2 Conditional Expectation

Given $f : \mathcal{X} \to \mathbb{R}$ what is the expected value of $f(X_m)$ given an initial distribution?
The function f on a finite state space $\mathcal{X}$ is equivalent to a vector $\vec{f} \in \mathbb{R}^{|\mathcal{X}|}$

$$\vec{f} = \begin{pmatrix} f(1) \\ f(2) \\ \dots \\ f(n) \end{pmatrix}$$

The conditional expectation for $f(X_m)$ given $X_0 \sim \vec{\mu}$ is denoted by

$$\mathbb{E}(f(X_m) \mid X_0 \sim \vec{\mu})$$

By definition of conditional expectation we get

$$\begin{aligned}
\mathbb{E}(f(X_m) \mid X_0 \sim \vec{\mu}) &= \sum_{k=1}^{|\mathcal{X}|} f(i_k) P(X_m = i_k \mid X_0 \sim \vec{\mu}) \\
&= \sum_{\text{all}(k)} f(i_k) \left[\vec{\boldsymbol{\mu}} \mathbf{P}^m\right]_k \\
&= \sum_{\text{all}(k)} \vec{\boldsymbol{f}}_k \left[\vec{\boldsymbol{\mu}} \mathbf{P}^m\right]_k \\
&= \langle \vec{\boldsymbol{\mu}} \mathbf{P}^m, \vec{\boldsymbol{f}} \rangle
\end{aligned}$$

2.3 Stationary Distribution

Suppose $X_0 \sim \vec{\mu}$ then the distribution for $X_n \sim \vec{\mu}(n)$ then what is the limit of $\vec{\mu}(n)$ as $n \to \infty$. Suppose the limit $\lim_{n\to\infty} \vec{\mu}P^n = \vec{\pi}$ exists then we can write

$$\vec{\boldsymbol{\pi}} = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}\mathbf{P}^n = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}\mathbf{P}^{n-1}P = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}(n-1)\mathbf{P} = \vec{\boldsymbol{\pi}}\mathbf{P}$$

So $\vec{\pi}$ is an `left eigenvector of` $\mathbf{P}$ with `eigenvalue 1`.

2.3 Definition (Stationary Distribution)**.** A probability vector $\vec{\boldsymbol{\pi}}$ is the Stationary Distribution for the stochastic matrix $\mathbf{P}$ if

$$\sum_k \vec{\boldsymbol{\pi}}_k = 1$$
$$\vec{\boldsymbol{\pi}}\mathbf{P} = \vec{\boldsymbol{\pi}}$$

2.4 Definition (Stationary Measure)**.** A measure $\vec{\nu}$ on $\mathcal{X}$ $\left(\vec{\nu} \in \mathbb{R}^{|\mathcal{X}|}\right)$ if

$$\vec{\nu}_i \geq 0$$
$$\sum \vec{\nu}_i > 0$$
$$\vec{\nu}\mathbf{P} = \vec{\nu}$$

2.5 Proposition (Stationary Distribution from Measure)**.** If $|X| < \infty$ and $\vec{\nu}$ is a stationary measure on $\mathbf{P}$

$$\vec{\boldsymbol{\pi}} = \frac{1}{\sum_i \vec{\nu}_i}\vec{\nu}$$

Then $\vec{\boldsymbol{\pi}}$ is a stationary distribution by definition.

2.6 Definition (Bi-stochastic Matrix)**.** A stochastic matrix is Bi-stochastic if

$$\sum_{\text{all}(i)} P_{ij_0} = 1 \qquad\qquad \text{for fixed } j_0$$

Sum of all rows = 1 and sum of all columns = 1.

2.7 Proposition (Stationary Distribution for Bi-stochastic Matrices)**.** If $\mathbf{P}$ is a **Bi-stochastic** transition matrix for Markov chain with finite state space $\mathcal{X}$ with $|\mathcal{X}| = N$ then the stationary distribution is given by

$$\vec{\boldsymbol{\pi}} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

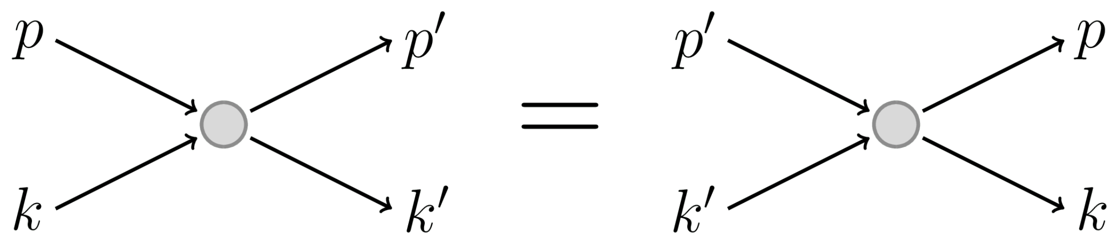
2.4 Detail Balance Condition

2.8 Definition (Detail Balance Condition)**.** $\vec{\pi}$ has the detail balance condition if:

$$\vec{\pi}_x \mathbf{P}_{xy} = \vec{\pi}_y \mathbf{P}_{yx}$$

2.9 Note. Detail balance condition means $\mathbb{P}(X_1 = x, X_0 = y) = \mathbb{P}(X_1 = y, X_0 = x)$.

2.10 Theorem (Detail Balance and Stationary Distribution)**.** If $X_0 \sim \vec{\pi}$ and $\vec{\pi}$ satisfies the detail balance condition then $X_n \sim \vec{\pi}$ for all $n \geq 1$



3 Week 3

3.1 Communicating States

3.1 Definition (communicating states). A state x communicates with y if $\exists n \geq 1$ such that

$$[\mathbf{P}^n]_{xy} > 0$$

denoted by $x \to y$.

3.2 Note. $\mathbb{P}(A \mid X_{n-1} = x) = \mathbb{P}_x(A)$ and $\mathbb{E}(\cdot \mid X_n = x) = \mathbb{E}_x(\cdot)$

3.3 Definition (Time of the first return / first hitting time).

$$\tau_x = \min\{n \mid X_n = x\}$$

$$\rho_{xy} = \mathbb{P}_x(\tau_y < \infty)$$

$\rho_{xy} = \mathbb{P}(X_n$ returns to y given it starts at $x)$.

3.4 Note.

$$1 - \rho_{xy} = \mathbb{P}_x(\tau_y = \infty)$$

3.5 Lemma (Communicating states and return probability). $x \to y \iff \rho_{xy} > 0$.

3.6 Lemma (Transitivity). $x \to y$ and $y \to z \Rightarrow x \to z$

3.7 Definition (Time of $k-th$ return).

$$\tau_x^k = \min\{n > \tau_x^{k-1} \mid X_n = x\}$$

where $\tau_x^1 = \tau_x$.

3.2 Recurrent and Transient States

3.8 Definition (Recurrent and Transient States). A state $x \in \mathcal{X}$ is called **recurrent** if

$$\rho_{xx} = 1$$

and **transient** if

$$\rho_{xx} < 1$$

3.9 Theorem (Escaping path). If $x \to y$ and $\rho_{xy} < 1$ then x is transient.

3.10 Theorem (Corollary of Escaping Path theorem). If $x \to y$ and x is recurrent then $\rho_{xy} = 1$.

3.3 Strong Markov Property

3.11 Definition (Stopping Time)**.** T is a stopping time if the occurrence (or non occurrence) of the event $\{T = n\}$ can be determined by $\{X_0, \ldots, X_n\}$.

3.12 Theorem (Strong Markov Property)**.** Suppose T is a stopping time. Given $T = n$ and $X_T = y$ the random variables $\{X_{T+k}\}_{k=0}^{\infty}$ behave like a Markov chain starting from initial state y. That is

$$\mathbb{P}(X_{T+1} = z \mid X_T = y, T = n) = \mathbb{P}(X_1 = z \mid X_0 = y) = \mathbf{P}_{yz}$$

3.13 Lemma ($k-$th return time and the strong Markov property)**.** Let τ_y^k be the $k-$th return time to y. Then the strong Markov property implies

$$\mathbb{P}_x(\tau_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1} \textbf{ or } \mathbb{P}_y(\tau_y^k < \infty) = \rho_{yy}^k \qquad \forall k \geq 1$$

3.14 Note. From the above **lemma** if we have $\rho_{yy} = 1$ (y is recurrent) then the chain returns to y for infinitely many k and it continually recurs in the Markov chain.
Otherwise if $\rho_{yy} < 1$ (y is transient) then $\rho_{yy}^k \to 0$ as $k \to \infty$ so after sometime y is never visited in the chain.

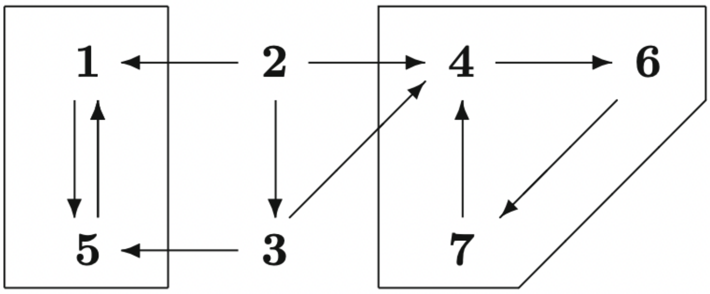
4 Week 4

4.1 Classification of States

4.1 Definition (Closed)**.** A set A is **closed** if it is impossible to get out. Formally $c \in A$ and $y \notin A$ then $P_{xy} = 0$

4.2 Definition (irreducible)**.** A set B is irreducible if every state is reachable from another in k steps or every state communicates with with all other states. Formally

$$x, y \in B \Rightarrow x \to y$$



4.3 Lemma (Commutating recurrent states)**.** If x is recurrent and $x \to y$ then y is recurrent

4.4 Lemma (Existence of recurrent states in finite closed sets)**.** If A is finite and closed then $\exists x \in A$ such that x is recurrent.

4.5 Theorem (Closed and irreducible sets are recurrent)**.** If $C \subseteq \mathcal{X}$ is **finite**, closed and irreducible then all $x \in C$ are recurrent.

4.6 Theorem (Decomposition Theorem)**.** If $\mathcal{X}$ is finite then

$$\mathcal{X} = T \cup R_1 \cup R_2 \cup \cdots \cup R_k$$

where T is the set of transient states and R_i for $1 \le i \le k$ are are closed irreducible sets of recurrent states.

4.7 Definition (Number of visits)**.** $N(y)$ is the number of visits to y after initial time.

4.8 Lemma (Expected number of visits)**.**

$$\mathbb{E}_x[N(y)] = \begin{cases} 0 & \rho_{xy} = 0 \\ \frac{\rho_{xy}}{1-\rho_{yy}} & \rho_{xy} > 0 \end{cases}$$

4.9 Lemma (Expected number of visits II)**.**

$$\mathbb{E}_x[N(y)] = \sum_{n=1}^{\infty} [\mathbf{P}^n]_{xy}$$

4.10 Theorem (Equivalent condition for recurrence)**.** y is recurrent if and only if

$$\sum_{n=1}^{\infty} [\mathbf{P}^n]_{yy} = \mathbb{E}_y[N(y)] = \infty$$

4.2 Existence of Stationary measure

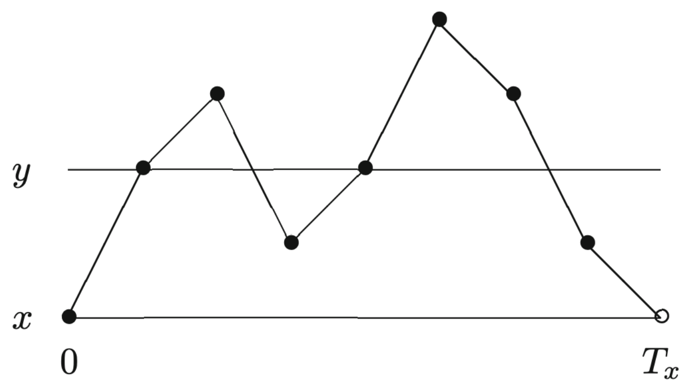
4.11 Theorem (Existence of Stationary measure)**.** Suppose $\mathcal{X}$ is irreducible and recurrent there exists a stationary measure $\vec{\mu}$ with

$$0 < \mu_y < \infty \qquad y \in \mathcal{X}$$

Let $x \in \mathcal{X}$ be recurrent by Existence of recurrent states in finite closed sets. We define $\vec{\mu}^x$ as

$$\mu_y^x = \mathbb{E}_x[\# \text{ of visits to } y \text{ before } x] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, \tau_x > n)$$

$\vec{\mu}^x$ is a stationary measure for $\mathbf{P}$.



5 Week 5

5.1 Definition (Ergodicity)**.** If $\vec{\pi}$ is a stationary distribution given $X_0 \sim \vec{\mu}$ if we have

$$\vec{\mu}\mathbf{P}^n \to \vec{\pi}$$

then $\mathbf{P}$ has Ergodicity.

5.2 Theorem (Ergodicity equivalent definition)**.**

$$\vec{\mu}\mathbf{P}^n \to \pi \iff [P^n]_{xy} \to \vec{\pi}_y \qquad \forall x \in \mathcal{X}$$

5.3 Remark. If y is transient then $\mathbb{E}_X[N(y)] < \infty$. Then

$$\mathbb{E}_X[N(y)] = \sum_{n=1}^{\infty} [\mathbf{P}^n]_{xy}$$

Then $[\mathbf{P}^n]_{xy} \to 0$. Meaning $\vec{\pi}_y = 0$, so all transient states have Ergodicity.

5.4 Definition (Periodicity)**.** A state x has a period d_x if

$$I_x = \{n \geq 0 \mid [P^n]_{xx} > 0\} \qquad \mathrm{d}_x = \gcd I_x$$

x is aperiodic if $\mathrm{d}_x = 1$ and periodic if $\mathrm{d}_x > 1$.

5.5 Definition (Class property)**.** A property $\mathcal{K}$ of a state x is called a **class property** if x has $\mathcal{K}$, $x \to y$ and $y \to x$ then y has $\mathcal{K}$.

5.6 Lemma. Periodicity is a class property If $x \to y$ and $y \to x$ then $\mathrm{d}_x = \mathrm{d}_y$.

5.7 Lemma. I_x is closed under addition

$$a, b \in I_x \Rightarrow (a + b) \in I_x$$

5.8 Lemma. If x is aperiodic then $\exists n_0$ such that for all $n \geq n_0$ $n \in I_x$.

5.9 Lemma. If $P_{xx} > 0$ then x is aperiodic.