stat333 Notes

Thaqib M

May 15, 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

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

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:

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

(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

$$P(X_{t+1} = x_t \mid X_t = x_t) = P(X_1 = x_1 \mid X_0 = x_0)$$

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

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

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

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

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

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

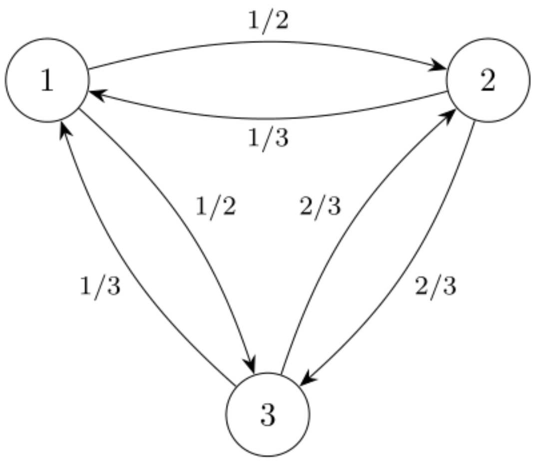
(probability of going from state i to state j in the chain).

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 $P(X_{n+m} = z \mid X_n = x) = [P(n, n+m)]_{xz}$. $\square$

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 $\lim\limits_{n\to\infty} \vec{\mu}P^n = \vec{\pi}$ exists then we can write

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

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

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

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

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 $\mathbf{P}$

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

Then $\boldsymbol{\vec{\pi}}$ 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

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

Proof. Sketch.

Let $\vec{\nu} = \left[\begin{smallmatrix} 1 & 1 & \cdots & 1 \end{smallmatrix}\right]$ be a vector

$$[\vec{\nu}\mathbf{P}]_k = \sum_j \vec{\nu}_j \mathbf{P}_{kj} = \sum_j \mathbf{P}_{kj} = 1 = \vec{\nu}_k$$

So $\vec{\nu}$ is a measure on $\mathbf{P}$. Then scaling $\vec{\nu}$ gives us a stationary distribution with $\vec{\boldsymbol{\pi}}_k = \frac{1}{N}$. $\square$