Lecture 3:

Initial Data and Conditional Expectation

Last Time

Thm 1. Let $(X_t)_{t\in I}$ be a time homogeneous Markov chain, then the m-step transition probability

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

is the m-th power of the transition matrix P, evaluated at the row x and column y, i.e., $[P^m]_{xy}$.

Today

1°. Remark 1.

$$[P^m]_{xy} = \mathbb{P}(X_{n+m}=y \mid X_n=x)$$
$$= \mathbb{P}(X_m=y \mid X_0=x)$$

does not depend on n. This implies time homogeneity applies not only to $\mathbb{P}(X_{n+1}=y \mid X_n=x)=\mathbb{P}(X_1=y \mid X_0=x)$, but also $\mathbb{P}(X_{n+m}=y \mid X_n=x)=\mathbb{P}(X_m=y \mid X_0=x)$.

Remark 2. Also, Theorem 1 implies

$$\boxed{P(n, n+m) = P^m, \quad \forall n, m \in \mathbb{N}.}$$

The left hand side (LHS) does not depend on n as well. Denote it by $P(m)$. That is,

$$\boxed{P(m) = P^m, \quad \forall m \in \mathbb{N}.}$$

Remark 3. Chapman – Kolmogorov equation

$$\boxed{[P(m+n)]_{xy} = \sum_{z \in \mathcal{X}} [P(m)]_{xz} [P(n)]_{zy}, \quad \forall m, n \in \mathbb{N}, \forall x, y \in \mathcal{X}.}$$

<u>Pf.</u> From Theorem 1,

$$\begin{aligned} LHS &= [P^{m+n}]_{xy} \\ &= [P^m \cdot P^n]_{xy} \\ &= \sum_{z \in \mathcal{X}} [P^m]_{xz} \cdot [P^n]_{zy} \\ &= \sum_{z \in \mathcal{X}} [P(m)]_{xz} \cdot [P(n)]_{zy} = RHS. \end{aligned}$$ $\blacksquare$

Recall

If $A \in \mathcal{M}_{m \times n}$, $B \in \mathcal{M}_{n \times p}$, then $A \cdot B \in \mathcal{M}_{m \times p}$.

Moreover,

$$\begin{aligned} [AB]_{ij} &= \sum_{k=1}^{n} A_{ik} \cdot B_{kj} \\ &= \langle \overrightarrow{A_{i,\cdot}}, \overrightarrow{B_{\cdot,j}} \rangle. \end{aligned}$$

Prepared by Kelvin Shuangjian ZHANG (ks3zhang@waterloo.ca)

2°. Ex 1. (Gambler's Ruin).

X_t — Gambler's Wallet at time t.

N : Once $X_t = N$, exit game.

0 : Once $X_t = 0$, exit game.

Flip coins. $\begin{cases} \text{Head} & \text{with prob. } 0.4 \\ \text{Tail} & \text{with prob. } 0.6 \end{cases}$

If Head, $X_t = X_{t-1} + 1$, $\forall\ 0 < X_{t-1} < N$.
Tail, $X_t = X_{t-1} - 1$,

$$\mathbb{P}(X_t = y \mid X_{t-1} = x) = \begin{cases} 0.4, & y = x+1; \\ 0.6, & y = x-1; \\ 0, & \text{otherwise}. \end{cases} \quad \forall 0<x<N.$$

$$\mathbb{P}(X_t = y \mid X_{t-1} = x) = \begin{cases} 1, & y = x; \\ 0, & \text{otherwise}. \end{cases} \quad x=0,N.$$

Q: Is it a Markov chain? Yes.
$\mathbb{P}(X_t = x_t \mid (X_s)_{s\le t-1} = (x_s)_{s\le t-1}) = \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1})$

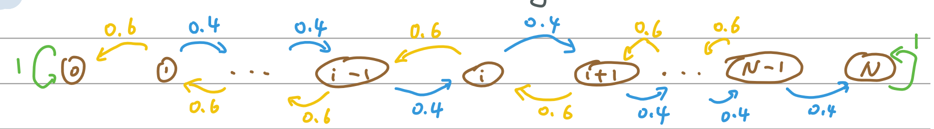
Q: Is it time homogeneous? Yes
$\mathbb{P}(X_{t+1} = y \mid X_t = x) = \mathbb{P}(X_1 = y \mid X_0 = x)$.

Q: What's the transition matrix?

Prepared by Kelvin Shuangjian ZHANG (ks3zhang@waterloo.ca)

$$P = \begin{array}{c} \\ 0 \\ \vdots \\ i \\ \vdots \\ N \end{array}\begin{array}{c} \begin{array}{cccccccc} 0 & 1 & \cdots & i-1 & i & i+1 & \cdots & N \end{array} \\ \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0.6 & 0 & 0.4 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \end{array}$$

Q: What's the visualizing chain?



Remark 4. For the case $N = 4$,

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array},$$

$$P^2 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0.24 & 0 & 0.16 & 0 \\ 0.36 & 0 & 0.48 & 0 & 0.16 \\ 0 & 0.36 & 0 & 0.24 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array},$$

$$\lim_{n\to\infty} P^n = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{57}{65} & 0 & 0 & 0 & \frac{8}{65} \\ \frac{45}{65} & 0 & 0 & 0 & \frac{20}{65} \\ \frac{27}{65} & 0 & 0 & 0 & \frac{38}{65} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$P^3, P^4, \ldots,$

Prepared by Kelvin Shuangjian ZHANG (ks3zhang@waterloo.ca)

3° Initial Data & Row vectors

Let X_0 be the initial random variable of the time homogeneous Markov chain $(X_n)_{n\in I}$. Denote by $0, 1, 2, \cdots, |I|-1$ the elements of I. Denote by $i_1, i_2, i_3, \cdots, i_{|\mathcal{X}|}$ the elemts of the state space $\mathcal{X}$. Let P be the transition matrix such that $P_{jk} = \mathbb{P}(X_1 = i_k \mid X_0 = i_j)$ the transition probability from the state i_j to i_k. Denote by

$\mu_j = \mathbb{P}(X_0 = i_j)$ and the row vector

$\vec{\mu} = (\mu_1, \mu_2, \cdots \mu_{|\mathcal{X}|})$ represents the Law of X_0.

$$\left.\begin{array}{l}\vec{\mu} \in M_{1\times|\mathcal{X}|}\\ P \in M_{|\mathcal{X}|\times|\mathcal{X}|}\end{array}\right\} \Rightarrow \vec{\mu} P \in M_{1\times|\mathcal{X}|}$$

Q: What is $\mathbb{P}(X_1 = i_k)$?

A:
$$\begin{aligned}\mathbb{P}(X_1 = i_k) &= \sum_{j=1}^{|\mathcal{X}|} \mathbb{P}(X_1 = i_k, X_0 = i_j)\\ &= \sum_{j=1}^{|\mathcal{X}|} \mathbb{P}(X_1 = i_k \mid X_0 = i_j)\cdot \mathbb{P}(X_0 = i_j)\\ &= \sum_{j=1}^{|\mathcal{X}|} P_{jk}\cdot \vec{\mu}_j = [\vec{\mu} P]_k .\end{aligned}$$

Multiplication Rule

$$\begin{aligned}&\mathbb{P}(A_1, A_2)\\ &= \mathbb{P}(A_1 \mid A_2)\cdot \mathbb{P}(A_2)\\ &= \mathbb{P}(A_2 \mid A_1)\cdot \mathbb{P}(A_1).\end{aligned}$$

Q: What is $\mathbb{P}(X_n = i_k)$?

A: $\mathbb{P}(X_n = i_k)$

$$= \sum_{j=1}^{|\mathcal{X}|} \mathbb{P}(X_n = i_k, X_0 = i_j)$$

$$= \sum_{j=1}^{|\mathcal{X}|} \mathbb{P}(X_n = i_k | X_0 = i_j) \cdot \mathbb{P}(X_0 = i_j)$$

$$= \sum_{j=1}^{|\mathcal{X}|} [P^n]_{jk} \cdot \vec{\mu}_j$$

$$= [\vec{\mu} P^n]_k .$$

Prop 1. In sum, once $\vec{\mu}$ and P are given, we know the distribution of X_n, $\forall n \in I$.

$$\mathbb{P}(X_n = i_k) = [\vec{\mu} P^n]_k .$$

Remark 5. There exists a one-to-one correspondence between $\{$ Probability distributions on $\mathcal{X}$ $\}$ and

$$\left\{ \nu \in \mathbb{R}^{|\mathcal{X}|} \,\middle|\, \sum_{j=1}^{|\mathcal{X}|} \nu_j = 1, \ \nu_j \geq 0, \forall j = 1, 2, \cdots, |\mathcal{X}| \right\} =: \mathcal{M}_1(\mathcal{X}).$$

Here the elements in $\mathcal{M}_1(\mathcal{X})$ are called probability vectors.

e.g. $\vec{\mu} = (1, 0, 0, 0) = \delta_1$ is called a "Dirac mass".

4°. Conditional Expectations.

$\mathbb{P}(X_m = i_k \mid X_0 = i_j)$ is a conditional probability.

Let $f: \mathcal{X} \to \mathbb{R}$ be a real-valued function on $\mathcal{X}$.

Note: There is a one-to-one correspondence between $\{$functions $f: \mathcal{X} \to \mathbb{R}\}$ and $\{$vectors $\vec{f} \in \mathbb{R}^{|\mathcal{X}|}\}$.

Q: What is the expectation of $f(X_m)$ given $X_0 = i_j$?

A:

$$\begin{aligned}
&\mathbb{E}[f(X_m) \mid X_0 = i_j] \\
&= \sum_{k=1}^{|\mathcal{X}|} f(X_m = i_k) \cdot \mathbb{P}(X_m = i_k \mid X_0 = i_j) \\
&= \sum_{k=1}^{|\mathcal{X}|} f(i_k) \cdot [P^m]_{jk} \\
&= \sum_{k=1}^{|\mathcal{X}|} \vec{f}_k \cdot [P^m]_{jk} \\
&= [P^m \vec{f}]_j .
\end{aligned}$$

e.g. let $|\mathcal{X}| = 2$,
In the case, $f: \mathcal{X} \to \mathbb{R}$ corresponds to a vector
$$\vec{f} = \begin{bmatrix} f(i_1) \\ f(i_2) \end{bmatrix} \in \mathbb{R}^2 .$$

Notation.
$$\vec{f} = \begin{bmatrix} f(i_1) \\ f(i_2) \\ \vdots \\ f(i_{|\mathcal{X}|}) \end{bmatrix} .$$

5°. Expectation with initial data

Let $\vec{\mu}$ be the distribution of X_0, P be the transition matrix, and $f: \mathcal{X} \longrightarrow \mathbb{R}$ be represented by a vector $\vec{f}$ such that

$\vec{f}_k = f(i_k), \ \forall k = 1, 2, \cdots, |\mathcal{X}|$.

Q: What is the expectation of $f(X_m)$?

A:
$$
\begin{aligned}
& \mathbb{E}[f(X_m) \mid X_0 \sim \vec{\mu}] \\
&= \sum_{k=1}^{|\mathcal{X}|} f(X_m = i_k) \cdot \mathbb{P}(X_m = i_k \mid X_0 \sim \vec{\mu}) \\
&= \sum_{k=1}^{|\mathcal{X}|} f(i_k) \cdot [\vec{\mu} P^m]_k \\
&= \sum_{k=1}^{|\mathcal{X}|} \vec{f}_k \cdot [\vec{\mu} P^m]_k \\
&= \vec{\mu} \, P^m \, \vec{f}
\end{aligned}
$$

$1 \times |\mathcal{X}|$ "row vector" = law/distribution of X_0

$|\mathcal{X}| \times 1$ "column vector" = function on $\mathcal{X}$.

This is the end of this lecture!

Prepared by Kelvin Shuangjian ZHANG (ks3zhang@waterloo.ca)