Handout 1

STAT5710 Applied Stochastic Processes

The review section is for reference. The rest follows the lecture, so you can fill it in as we go. Nothing here requires a computer.

Review

Law of total probability. If $B_1, B_2, \ldots$ partition the sample space, then

$$P(A) = \sum_i P(A \mid B_i)\, P(B_i)$$

The same identity holds conditional on any event C:

$$P(A \mid C) = \sum_i P(A \mid C \cap B_i)\, P(B_i \mid C)$$

For discrete random variables this reads

$$P(X = x \mid Y = y) = \sum_z P(X = x \mid Y = y,\, Z = z)\, P(Z = z \mid Y = y)$$

Matrix multiplication. If A is $n \times m$ and B is $m \times p$, the (i, j) entry of AB is the inner product of the ith row of A with the jth column of B:

$$(AB)_{ij} = a_i^T b_j = \sum_k a_{ik}\, b_{kj}$$

In particular, for a square matrix P,

$$(P^2)_{ij} = \sum_k p_{ik}\, p_{kj} \qquad \text{and} \qquad P^n = P^{n-1} P$$

Stochastic processes

A **stochastic process** is an indexed collection of random variables $\{X_t : \Omega \to S,\ t \in I\}$ sharing a common codomain S called a “state space”.

Table 1: Both S and I may be discrete or continuous, and the two choices are independent of each other.

Notation	Name	Meaning
X_t	state	the value of the process at t
S	state space	the values the process can take
I	index set	what the process is indexed by; usually time
$\{X_t(\omega)\}$	sample path	one draw of the whole process

Warm up

Consider the board game *Monopoly*, and let X_t denote your position on the board at turn t. Imagine a simplified version of the game where all moves are local (no jail, no chance and community chest).

1. Let Y be the sum of two fair dice, so that one turn moves Y squares forward. The distribution of Y is

$$P(Y = k) = \begin{cases} \frac{k-1}{36}, & 2 \leq k \leq 7 \\ \frac{13-k}{36}, & 7 < k \leq 12 \end{cases}$$

   (a) Verify that $\sum_{k=2}^{12} P(Y = k) = 1$.

   (b) Squares are numbered $1, \ldots, 40$ and wrap around the board, so that you move from square 40 (boardwalk) back to square 1 (go). Explain why

$$P(X_t = x \mid X_{t-1} = s) = P[Y = (x - s) \bmod 40]$$

   and say what this probability is when $(x - s) \bmod 40 = 1$.

Markov chain basics

Definitions

Assume from here on that $\{X_t\}$ has discrete state space and discrete time, with $I = \mathbb{N}$.

2. Fill in the definitions as we go.

   (a) $\{X_t\}$ is a **Markov chain** if, for all $t \in I$ and all $x_0, \dots, x_t \in S$,

   $$P(X_{t+1} = x_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) \ =$$

   (b) In words: the past and the future are ____________________________ given the present.

   (c) A Markov chain is **time-homogeneous** if for every $i, j \in S$ and every $t, s \in I$,

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

   (d) For such a chain we write $p_{ij} =$ __________________________ , the *transition probability* from i to j, and call $P = \{p_{ij}\}$ the __________________________ .

   (e) A matrix P is **stochastic** if _________________________________ for each i.

Examples

3. **SLO weather.** Let X_t be the weather in San Luis Obispo on day t, with $S = \{s, c, r\}$ for sunny, cloudy, and rainy. Suppose $\{X_t\}$ is a time-homogeneous Markov chain with

$$P = \begin{array}{c} \\ s \\ c \\ r \end{array} \begin{array}{c} \begin{array}{ccc} s & c & r \end{array} \\ \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.20 & 0.70 & 0.10 \\ ? & 0.30 & 0.40 \end{bmatrix} \end{array}$$

   (a) What must ? be, and what forces that value?

   (b) If it is cloudy today, what is the probability it is sunny tomorrow?

   (c) The day after tomorrow?

4. **Eugene Onegin.** In 1913 Markov classified the first 20,000 letters of Pushkin's poem as vowels or consonants and tallied pairs of successive letters — the application that gave Markov chains their name. Let $X_t = 1$ if the tth letter is a consonant and 0 if it is a vowel, so $S = \{0, 1\}$ and $I = \{1, \dots, 20000\}$. His estimates were

$$P = \begin{array}{c} \\ v \\ c \end{array} \begin{array}{c} \begin{array}{cc} v & c \end{array} \\ \begin{bmatrix} 0.175 & 0.825 \\ 0.526 & 0.474 \end{bmatrix} \end{array}$$

   (a) Draw the transition graph, labelling each edge with its probability.

   (b) Markov's point was that successive letters are *not* independent. If they were, $P(X_{t+1} = 1 \mid X_t = i)$ would not depend on i. What do these estimates say instead?

5. **Bounded random walk.** A walker on $S = \{0, 1, 2, 3, 4\}$ steps left with probability $1/3$ and right with probability $2/3$, and stops on reaching 0 or 4.

   (a) Write the transition matrix. (What do the rows for states 0 and 4 have to look like?)

$$P = \begin{array}{cc} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0\\ 1\\ 2\\ 3\\ 4 \end{array} & \left[\begin{array}{ccccc} \quad & \quad & \quad & \quad & \quad \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{array}\right] \end{array}$$

   (b) Draw the transitions as a graph, labelling each edge with its probability.

Markov chain or not?

To check the Markov property, verify that conditioning on the whole past gives the same answer as conditioning on the present alone. To check homogeneity, verify that $P(X_{t+1} = j \mid X_t = i)$ does not depend on t.

Example of a non-Markov process

Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. with $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$, and set

$$X_t = \varepsilon_t + \varepsilon_{t-1}, \qquad t = 1, 2, \ldots$$

6. What is the state space of $\{X_t\}$?

7. Compute $P(X_{t+1} = 2 \mid X_t = 0)$. *Hint:* condition on ε_t and use the law of total probability,

$$P(A \mid C) = \sum_i P(A \mid C \cap B_i)\, P(B_i \mid C)$$

8. Now compute $P(X_{t+1} = 2 \mid X_t = 0, X_{t-1} = 2)$. *Hint:* what do $X_{t-1} = 2$ and $X_t = 0$ force ε_{t-1} and ε_t to be? What would ε_{t+1} then have to equal?

9. Compare your answers to 7 and 8. Is $\{X_t\}$ a Markov chain? Explain what goes wrong — which two variables does conditioning on X_t leave coupled?

But wait, look again

10. Now track instead the "coupled" process $Y_t = (\varepsilon_t, \varepsilon_{t-1})$.

(a) List the four states of $\{Y_t\}$.

(b) Given $Y_t = (i_1, i_2)$, the next state is $Y_{t+1} = (\varepsilon_{t+1}, \varepsilon_t)$. Which coordinate of Y_{t+1} is already determined by Y_t, and which is fresh? Use this to argue that $\{Y_t\}$ is a homogeneous Markov chain.

(c) Fill in the transition matrix, writing $-$ for -1 and $+$ for $+1$.

$$P = \begin{array}{c} \\ -- \\ -+ \\ +- \\ ++ \end{array} \begin{array}{c} \begin{array}{cccc} -- & -+ & +- & ++ \end{array} \\ \begin{bmatrix} \quad & \quad & \quad & \quad \\ \quad & \quad & \quad & \quad \\ \quad & \quad & \quad & \quad \\ \quad & \quad & \quad & \quad \end{bmatrix} \end{array}$$

11. Notice that X_t is a function of Y_t: in fact $X_t = \mathbf{1}^T Y_t$. So a process that is *not* Markov turned out to be a function of one that is. This is our first **hidden Markov model**; fill in the terms.

- $\{Y_t\}$ is the ________________________ (unobserved) process
- $\{X_t\}$ are the ________________________ (observed)
- $f(y) = \mathbf{1}^T y$ is the ________________________ function

Multi-step transitions

Two steps

For $n \geq 1$, define the n-step transition probability

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

so that $p_{ij}^{(1)} = p_{ij}$.

12. Take $n = 2$. Condition on the state at the intermediate time $t - 1$ and apply the conditional law of total probability from the review:

$$p_{ij}^{(2)} \;=\; P(X_t = j \mid X_{t-2} = i) \;=\; \sum_k P(\underline{\hspace{3cm}} \mid X_{t-1} = k,\, X_{t-2} = i)\; P(\underline{\hspace{3cm}} \mid X_{t-2} = i)$$

    (a) Fill in the two blanks.

    (b) The first factor conditions on both X_{t-1} and X_{t-2}. Apply the Markov property. Which one-step transition probability is this?

    (c) The second factor is also a one-step transition probability. Which one?

    (d) Put the pieces together:

$$p_{ij}^{(2)} \;=\; \sum_k \underline{\hspace{2.5cm}} \cdot \underline{\hspace{2.5cm}}$$

13. Compare 12(d) with the expression for $(P^2)_{ij}$ in the review. What have you just shown?

An example

Questions 14–15 use the chain on $S = \{1, 2, 3\}$ with

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

14. List every path of length two from state 1 to state 3, then use them to compute $p_{13}^{(2)}$ directly.

15. Now compute P^2 and confirm that its $(1, 3)$ entry matches your answer to 14.

$$P^2 = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix} \end{array}$$

The general case

Nothing above was special to $n = 2$. Suppose $p_{ij}^{(n)} = (P^n)_{ij}$ holds for some n (the induction hypothesis) and condition on the state at time $t-1$ again — this time the first leg takes n steps and the second takes one.

16. Complete the inductive step.

$$p_{ij}^{(n+1)} = P(X_t = j \mid X_{t-n-1} = i) = \sum_k \underbrace{P(X_{t-1} = k \mid X_{t-n-1} = i)}_{A} \underbrace{P(X_t = j \mid X_{t-1} = k, X_{t-n-1} = i)}_{B}$$

(a) Write A and B as transition probabilities. Say which simplification uses the induction hypothesis and which uses the Markov property.

$$A = \underline{\hspace{2cm}}$$

$$B = \underline{\hspace{2cm}}$$

(b) Recognize the sum as a matrix product to finish:

$$p_{ij}^{(n+1)} = \sum_k \underline{\hspace{2cm}} \cdot \underline{\hspace{2cm}} = (\underline{\hspace{3cm}})_{ij} = (P^{n+1})_{ij}$$

(c) In words, what does this say about getting from i to j in $n+1$ steps?

Chapman-Kolmogorov equations

Question 16 split the $n+1$ steps at the *last* one, but nothing forces that choice. Since $P^n = P^m P^{n-m}$ for any $0 < m < n$,

$$(P^n)_{ij} = \sum_k (P^m)_{ik} \left(P^{n-m}\right)_{kj}$$

Written out, this says

$$P(X_t = j \mid X_{t-n} = i) \;=\; \sum_k \underbrace{P(X_{t-n+m} = k \mid X_{t-n} = i)}_{i \to k \text{ in } m \text{ steps}} \; \underbrace{P(X_t = j \mid X_{t-n+m} = k)}_{k \to j \text{ in } n-m \text{ steps}}$$

These are the **Chapman-Kolmogorov equations**: to get from i to j in n steps, add up the probabilities of passing through *each* state k at a fixed intermediate time.

17. Which choice of m recovers question 16? Which recovers question 12?

18. **A chain with more than one path.** In question 14 only one two-step path ran from 1 to 3. Now take

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

   (a) Find both two-step paths from 1 to 3 and give the probability of each.

   (b) Add them to get $p_{13}^{(2)}$, then check that the same number falls out of Chapman–Kolmogorov with $m = 1$.

The Markov property

Every conditioning so far has been on the *immediately* preceding state. But the defining characteristic of Markov chains holds more generally when conditioning on the *most recent* state.

The (general) Markov property. If $\{X_t\}$ is a time-homogeneous Markov chain, then for any $m \geq 1$,

$$P(X_{t+1} = j \mid X_{t-m} = i,\ X_{t-m-1} = x_{t-m-1},\ \dots) \ = \ P(X_{t+1} = j \mid X_{t-m} = i) \ = \ p_{ij}^{(m+1)}$$

In words: conditioning on the state $m+1$ steps back *together with the entire history before it* is the same as conditioning on that one state alone.

19. Verify the case $m = 1$ by conditioning on X_t and applying the law of total probability. Fill in the blanks below

$$P(X_{t+1} = j \mid X_{t-1} = i, \dots) \ = \ \sum_k \underbrace{P(X_t = k \mid X_{t-1} = i, \dots)}_{A} \ \underbrace{P(X_{t+1} = j \mid X_t = k,\ X_{t-1} = i, \dots)}_{B}$$

$$A \ = \ \underline{\qquad\qquad\qquad} \qquad\qquad\qquad B \ = \ \underline{\qquad\qquad\qquad}$$

$$= \ \underline{\qquad\qquad\qquad}$$

$$= \ p_{ij}^{(\underline{\quad})}$$

The general case follows by induction on m, just as in question 16. The argument is left as an exercise.

Marginal distributions

20. To get the marginal distribution of X_t we also need an initial distribution α on X_0, where $\alpha_i = P(X_0 = i)$. Complete the computation:

$$P(X_t = j) = \sum_i P(X_t = j \mid X_0 = i)\, P(X_0 = i) = \sum_i \underline{\hspace{2cm}} \cdot \underline{\hspace{2cm}} = (\underline{\hspace{2.5cm}})_j$$

Note that this still depends on t. But consider the output below, which shows P^{11} for the same chain.

```r
P <- matrix(c(1/2, 1/2,   0,
              1/2,   0, 1/2,
                0, 1/2, 1/2), nrow = 3, byrow = TRUE)

Pn <- P
for (i in 1:10) Pn <- Pn %*% P
round(Pn, 4)
```

```
       [,1]   [,2]   [,3]
[1,] 0.3335 0.3335 0.3330
[2,] 0.3335 0.3330 0.3335
[3,] 0.3330 0.3335 0.3335
```

21. What do you notice about the rows of P^{11}? What does that suggest about $P(X_t = j)$ as $t \to \infty$, and how much do you think the answer depends on the initial distribution α?

Joint distributions

Everything so far has described one time point at a time. But we now have enough information to characterize the joint distribution of the chain at arbitrarily many observation times.

22. For any times $t_1 < t_2 < \cdots < t_n$,

$$P(X_{t_1} = i_1,\, X_{t_2} = i_2,\, \ldots,\, X_{t_n} = i_n) \;=\; \left(\alpha P^{t_1}\right)_{i_1} \, p_{i_1 i_2}^{(t_2-t_1)} \, p_{i_2 i_3}^{(t_3-t_2)} \cdots p_{i_{n-1} i_n}^{(t_n-t_{n-1})}$$

(a) Explain where each factor comes from, and why the first one has a different form than the rest.

(b) Now specialize to consecutive times $0, 1, \ldots, t$, so that every gap is a single step:

$$P(X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_t = i_t) \;=\; \underline{\qquad\quad} \prod_{k=1}^{t} \underline{\qquad\qquad\quad}$$

(c) A finite, time-homogeneous Markov chain is therefore completely specified by just two objects. What are they?

Remark. The joint distribution above is a finite-dimensional object, because the chain is observed at a finite collection of times. However, we can consider the sequence $X_0, X_1, X_2, \ldots$ and observe that the sample paths $\{X_t(\omega)\}_{t\in\mathbb{N}}$ are points in an infinite-dimensional space. Remarkably, the finite-dimensional distributions suffice to construct a probability measure on the path space, *i.e.*, an infinite-dimensional probability distribution. In practical terms, this allows us to ask questions about limits as $t \to \infty$.