STATS 3U03

Sang Woo Park

January 13, 2017

Course Outline

- Textbook: Inroduction to stochastic processes
- Requirement: 5 assignments, 2 tests, and 1 final
- Test 1: Friday, February 10th
- Test 2: Friday, March 17th

Contents

1 Introduction

1.1 Review

Definition 1.1 (Independent random variables)**.** *X and Y are independent iff for any $a, b \in \mathbb{R}$, $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$*

1.2 Stochastic processes

Definition 1.2 (Stochastic process)**.** *Let T be a subset of $[0, +\infty]$. For each $t \in T$, let X_t be a random variable. Then, the collection of $\{X_t : t \in T\}$ is called a stochastic process. Simply put, a stochastic process is just a family of random variables.*

Example 1.2.1. Let $T = \{0\}$. Then, $\{X_0\}$ is a stochastic process.

Example 1.2.2. Let $T = \{1, 2, 3, \dots, m\}$ be a set of finite natural numbers. Then, $\{X_1, X_2, X_3, \dots, m\}$ is a stochastic process.

Example 1.2.3. Let $T = \{0, 1, 2, \dots\}$ be a set of all non-negative integers. Then, $\{X_1, X_2, X_3, \dots\}$ is a stochastic process.

Example 1.2.4. Let $T = [0, +\infty)$ be a set of all non-negative real numbers. Then, $\{X_t : t \geq 0\}$ is a stochastic process.

Definition 1.3 (Time index)**.** *Let T be time index. If $T = \{0, 1, 2, \dots\}$, then the time is discrete. If $T = [0, \infty)$, then time is continuous.*

Definition 1.4 (State Space)**.** *State space, S, is the space space where the random variable takes the values.*

Given a sample space, S, and time index $t \in T$, we can define $X_t(w) \in S$, to describe a stochastic process. Here, $\{X_t : t \in T\}$ describes the dependence relation.

We can further categorize a stochastic process by considering the following two cases: countable and uncountable state space. Time index can also be categorized as follows: discrete and continuous time. Note that each stochastic process must belong to one of the four categories.

Remark. Every stochastic process can be described by the following three factors:

1. Time index
2. State space
3. Dependence relation

Example 1.2.5. Let $S = \{0, 1\}$ and $T = \{0, 1, 2, \dots\}$. Given,

$$X_n = \begin{cases} 1 & \text{with probability of } 1/2 \\ 0 & \text{with probability of } 1/2 \end{cases}$$

$\{X_0, X_1, X_2, \dots\}$ is a stochastic process and is often noted as Bernoulli trials.

2 Markov chains (Discrete time Markov chains)

We will only be dealing with discrete time Markov chains in chapter 1 and 2. In other words, $T = \{0, 1, 2, \dots\}$. It follows that the state space, S, will be at most countable. Finally, Markov describes the dependence relation: $X_0, X_1, X_2, \dots$.

In example 1.2.5, every trial of the Bernoulli trials was independent. On the other hand, in a Markov model, X_{n+1} depends on X_n but not on any past stats, $X_1, X_2, \dots, X_{n-1}$.

2.1 Markov property

Definition 2.1. Markov property *can be expressed as follows:*

$$\begin{aligned} &P\left(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n\right) \\ &= P\left(X_{n+1} = x_{n+1} | X_n = x_n\right) \end{aligned}$$

$P\left(X_{n+1} = y | X_n = x\right)$ is noted as the transition probability and it describes the one step transition from x to y starting at time n. If

$$P\left(X_{n+1} = y | X_n = x\right) = P\left(X_1 = y | X_0 = x\right),$$

then the Markov chain is called to have stationary transition, or homogeneous.

Definition 2.2. *Let $\{X_n : n = 0, 1, 2, \dots\}$ be a homogeneous Markov chain. Then,*

$$P_{xy} = P\left(X_1 = y | X_0 = x\right) = P\left(X_{n+1} = y | X_n = x\right),$$

is the one-step transition probability*.*

Definition 2.3. *Following the definition 2.2, we can now define* one-step transition matrix*:*

$$\mathbb{P} = \left(P_{xy}\right)_{x,y \in S}$$

Remark. Given, X_0, $\pi_0(x) = P(X_0 = x)$ is called the *initial distribution.*

Given a Markov chain, we wish to answer the following fundamental questions:

1. Distribution of X_n for any $n \geq 1$.
2. Join distribution of $X_{n_1}, \dots, X_{x_k}$ for any $1 \leq n_1 < n_2 < \cdots < n_k \quad k, \geq 2$.
3. Long time behaviour, i.e.

$$\lim_{n \to \infty} P(X_n = x)$$

Example 2.1.1. We have the following Markov chain: $\{X_n : n = 0, 1, 2, \dots\}$ where $S = \{0, 1\}$. For this model, its initial distribution can be described as follows:

$$\begin{cases} \pi_0(0) = P(X_0 = 0) = a \\ \pi_0(1) = 1 - a \end{cases}$$

Transition probabilities can be written in a similar fashion:

$$\begin{array}{ll} P(X_1 = 1|X_0 = 0) = p, & P(X_1 = 0|X_0 = 0) = 1 - p \\ P(X_1 = 0|X_0 = 1) = q, & P(X_1 = 1|X_0 = 1) = 1 - q \end{array}$$

where $0 \le p, q \le 1$. For this Markov chain, we can consider the following three cases:

Case 1. $p = q = 0$**.**

This case is trivial.

Case 2. $p = q = 1$**.**

This case is also trivial.

Case 3. $0 \le p + q \le 2$**.**

$$\begin{aligned} P(X_{n+1} = 0) &= P(X_{n+1} = 0 \cap X_n = 0) + P(X_{n+1} = 0 \cap X_n = 1) \\ &= P(X_n = 0)P(X_{n+1} = 0|X_n = 0) + P(X_n = 1)P(X_{n+1} = 0|X_n = 1) \\ &= P(X_n = 0)(1 - p) + P(X_n = 1)q \\ &= P(X_n = 0)(1 - p) + (1 - P(X_n = 0))q \\ &= (1 - p - q)P(X_n = 0) + q \end{aligned}$$

We can further expand this as follows:

$$\begin{aligned} P(X_{n+1} = 0) &= (1 - p - q)P(X_n = 0) + q \\ &= (1 - p - q)\big[(1 - p - q)P(X_{n-1} = 0) + q\big] + q \\ &= (1 - p - q)^n P(X_0 = 0) + q \sum_{j=0}^{n-1} (1 - p - q)^j \end{aligned}$$

Note that

$$\sum_{j=0}^{n-1} (1 - p - q)^j \quad = \frac{(1 - p - q)^n - 1}{(1 - p - q) - 1}$$

Therefore, we have

$$
\begin{aligned}
P(X_{n+1} = 0) &= (1-p-q)^n a + q\frac{(1-p-q)^n - 1}{(1-p-q) - 1} \\
&= (1-p-q)^n a - \frac{q}{p+q}\big((1-p-q)^n - 1\big)
\end{aligned}
$$

For this Markov chain, we find that

$$
\lim_{n\to\infty} P(X_n = 0) = \frac{q}{p+q}
$$

2.2 Transition function and initial distribution

Example 2.2.1.

$$
\begin{aligned}
P_{xy} &= P(X_{n+1} = y | X_n = x) \\
&= P(X_1 = y | X_0 = x)
\end{aligned}
$$

Definition 2.4. Transition function, $P(x, y) : S \times X \to [0, 1]$, *satisfies the following conditions:*

1. $p(x, y) \geq 0$.
2. $\sum_{y\in S} p(x, y) = 1$ *for all* $x \in S$.

Definition 2.5. *Given a transition function,* $p(x, y)$, *a transition matrix is defined as follows:*

$$
\mathbb{P} = (p(x, y))_{x,y\in S}
$$

Example 2.2.2.

$$
\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
$$

Example 2.2.3.

$$
\begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 5/8 \\ 0 & 1/4 & 3/4 \end{pmatrix}
$$

Definition 2.6. Initial distribution *is a probability mass function (pmf) that is defined as follows*

$$
\pi_0(x) = p(X_0 = x).
$$

Note that it must satisfy the following conditions:

1. $\pi_0(x) \geq 0$
2. $\sum_{x\in S} \pi_0(x) = 1$

Theorem 2.1. *Let $\{x_N : n = 0, 1, 2, \dots\}$ be a Markov chain with initial distribution $\pi_0(x)$, and one-step transition matrix $\mathbb{P} = (p(x, y))_{x,y \in S}$. Then, the distribution of X_n is*

$$\begin{aligned} P(X_n = x_n) &= \sum_{x_0 \in S} \sum_{x_1 \in S} \cdots \sum_{x_{n-1} \in S} \pi_0(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \\ &= \pi_0 \underbrace{\mathbb{P}\mathbb{P} \dots \mathbb{P}}_{n} \end{aligned}$$

Proof. For any $n \geq 1, \quad x_n \in S$

$$\begin{aligned} P(X_n = x_n) &= P(X_n = x_n, x_0 \in S, X_1 \in S, \dots, X_{n-1} \in S) \\ &= \sum_{x_0 \in S} \sum_{x_1 \in S} \cdots \sum_{x_{n-1} \in S} P(X_n = x_n, X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned}$$

Note that

$$\begin{aligned} &P(X_n = x_n, X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= P(X_0 = x_0) P(X_1 = x_1 | X_0 = x_0) P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \\ &\cdots P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned}$$

Using the Markov property, it is evident that the equation above is equivalent to $P(X_0 = x_0) P(X_1 = x_1 | X_0 = x_0) \cdots P(X_n = x_n | X_{n-1} = x_{n-1})$. $\square$

Example 2.2.4. Simple random walk is a Markov chain:

$$\begin{aligned} X_0 &= 0 \\ X_1 &= \begin{cases} 1 & p \\ -1 & q \end{cases} \\ X_2 &= \begin{cases} X_1 + 1 & p \\ X_1 - 1 & q \end{cases} \\ X_n &= \begin{cases} X_{n-1} + 1 & p \\ X_{n-1} - 1 & q \end{cases} \end{aligned}$$

where $S = \{0, \pm 1, \pm 2, \dots\}$.

Example 2.2.5 (Ehrenfest chain)**.** Suppose that we have a box and a inivisble bar that divides the box into region I and II. d balls are placed in a box. Initially, n balls are distributed in region I and $d - n$ balls are distributed in region II. You pick a ball at random. If it's from region I, you put it back in region II. If it's from region II, you put it back in region I.

First, note that this Markov chain has a state space of $S = \{0, 1, 2, \dots, d\}$. We observe that

$$P(0, y) = \begin{cases} 0 & y > 1 \\ 1 & y = 1 \end{cases}$$

$$P(1, y) = \begin{cases} 0 & y \neq 0, 2 \\ \frac{1}{d} & y = 0 \\ 1 - \frac{1}{d} & y = 2 \end{cases}$$

In general, we have

$$P(x, y) = \begin{cases} 0 & y \neq x \pm 1 \\ 1 - \frac{x}{d} & y = x + 1 \\ \frac{x}{d} & y = x - 1 \end{cases}$$

Combining these results, we have the following transition matrix:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & \\ \frac{1}{d} & 0 & 1 - \frac{1}{d} & & \\ & & & \ddots & \\ & & & & \end{pmatrix}_{(d+1)\times(d+1)}$$

Example 2.2.6 (Birth-Death Markov chain)**.** At each time step, one person can die and a new person can be born:

$$X_{n+1} = \begin{cases} p_x & y = X_n + 1 \\ q_x & y = X_n - 1 \\ r_x & y = X_n \\ 0 & \text{else} \end{cases}$$

Example 2.2.7 (Queuing chain)**.** At each time step, one customer is served and new customers arrive:

$$X_{n+1} = \begin{cases} y_{n+1} & \text{if } X_n = 0 \\ X_n - 1 + y_{n+1} & \text{if } X_n \geq 1 \end{cases}$$

We introduce a new notation, $x^+ = x \vee 0$, which is essentially $\max(x, 0)$. Using this notation, we can rewrite the Markov chain as follows

$$X_{n+1} = (X_n - 1)^+ + y_n$$

Example 2.2.8 (Branching Markov chain)**.** If $X_0 = 0$, then $X_n = 0$ for all $n \geq 1$. We call 0 an *absorbing state.*

Suppose $X_0 \geq 1$. An individual, i, will produce y_i number of offsprings at each generation. Then, we will have

$$X_1 = y_1^{(1)} + \cdots + y_{X_0}^{(1)}$$

Each individual in generation 1 will also produce offsprings. Then,

$$X_2 = y_1^{(2)} + \cdots + y_{X_0}^{(2)}$$

We wish to understand how the population will evolve over time. To do so, we can look at the expected value. It's clear that the population will grow if $E[y] > 1$. On the other hand, if $E[y] < 1$, the population will eventually die out.

Example 2.2.9 (Wright-Fisher Markov chain)**.** For this Markov chain, we start by make the following assumptions:

1. The population size is fixed.
2. No generation overlap.

Within the population, there are N number of individuals of two types: I and II. Let X_0 be number of type I individuals at time 0. Each individual in generation 1 pick its parent from generation 0 at random. This process is equivalent to repeating Bernoulli trials N times (also equivalent to binomial).

Therefore, we have

$$\begin{aligned}
X_1 &\sim \text{Bin}(N, \frac{X_0}{N}) \\
X_2 &\sim \text{Bin}(N, \frac{X_1}{N}) \\
&\vdots \\
X_{n+1} &\sim \text{Bin}(N, \frac{X_n}{N})
\end{aligned}$$

2.3 Joint distribution

Given a Markov chain with π_0 and $\mathbb{P}$, how do we find (1) the distribution of X_n and (2) the joint distribution of X_n and X_m where $n < m$?

From the previous section, recall that $\pi_n = \pi_0 \underbrace{\mathbb{P}\mathbb{P}\ldots\mathbb{P}}_{n}$.

Example 2.3.1. Consider the following transition matrix:

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

Supose that $\pi_0 = (1, 0, 0, 0)$. Then, we have

$$\begin{aligned}
\pi_1 &= (1, 0, 0, 0)\mathbb{P} = (1/2, 1/2, 0, 0) \\
\pi_2 &= (1, 0, 0, 0)\mathbb{P}\mathbb{P} = (1/4, 1/2, 1/4, 0)
\end{aligned}$$

Eventually, all states will converge to the absorbing state and stay there.

To find the join distribution, we first note that

$$\begin{aligned}
P(X_n = x, X_m = x_m) &= P(X_n = x_n)P(X_m = x_m|X_n = x_n) \\
&= P(X_m = x_m)P(X_{m-n} = x_m|X_0 = x_n)
\end{aligned}$$

Definition 2.7. *For any interger m, m-step transition matrix is given by*

$$P^m(x, y) = P(X_m = y|X_0 = x).$$

When $m = 0$, we have

$$p^0(x, y) = \begin{cases} 1 & y = x \\ 0 & y \neq x \end{cases}$$

We can decompose m-step transition matrix as follows:

$$\begin{aligned}
P^m(x, y) &= P(X_m = y|X_0 = x) \\
&= P(X_m = y, x_1 \in S, \dots, X_{m-1} \in S|X_0 = x) \\
&= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{m-1} \in S} P(x_0, x_1) \dots P(x_{n-1}, y)
\end{aligned}$$

Then, we have

$$\begin{aligned}
P(X_m = y) &= P(X_m = y, X_m \in S) \\
&= \sum_{x_0 \in S} P(X_m = y, X_0 = x_0) \\
&= \sum_{x_0 \in S} P(X_0 = x_0)P^m(X_0, y) \\
&= \sum_{x_0 \in S} \pi_0(x_0)P^m(X_0, y)
\end{aligned}$$

Therefore, we have

$$(P(X_m = x_m))_{x_m \in S} = \pi_0 \mathbb{P}^m$$

Definition 2.8 (Hitting time)**.** *Given $A \subset S$, hitting time T_A is defined as follows:*

$$T_a = \min\{n \geq 1 : X_n \in A\}$$

If $A = \{x\}$, then we have $T_x = T_x$. Note that

- $T_A \geq 1$
- *If $x_n \notin A$ for all $n \geq 1$. we have $T_A = +\infty$*

Now, we wish to understand the distribution of T_y given thatn $X_0 = x$. First, note that we have

$$P_x(T_y) = 1 = P(T_y = 1|X_0 = x) = P(x, y)$$

Similarly, we have

$$\begin{aligned} P_x(T_y = 2) &= P_x(x_1 \neq y, x_2 = y) \\ &= \sum_{w \neq y} P(x, w)P(w, y) \end{aligned}$$

Generally, we have

$$\begin{aligned} P_x(T_y = n+1) &= P_x(x_1 \neq y, \ldots, x_n \neq y, x_{n+1} = y) \\ &= \frac{P(X_0 = x, X_1 \neq y \ldots X_n \neq y, X_{n+1} = y)}{P(X_0 = x)} \\ &= \frac{P(X_0 = x, X_1 \neq y)}{P(X_0 = x)} \frac{P(X_0 = x, X_1 \neq y \ldots X_n \neq y, X_{n+1} = y)}{P(X = x, X_1 \neq y)} \\ &= \sum_{x_1 \neq y} P(x, x_1)P_{x_1}(T_y = n) \end{aligned}$$

Note that the last result follows from the Markov property.

Lemma 2.1.

$$P^m(x, y) = \sum_{k=1}^{m} P_x(T_y = k)P^{m-k}(y, y)$$

Proof.

$$\begin{aligned} P^m(x, y) &= P(X_m = y|X_0 = x) \\ &= P(X_m = y, T_y \leq m|x_0 = x) \\ &= \sum_{k=1}^{m} P(X_m = y, T_y = k|X_0 = x) \\ &= \sum_{k=1}^{m} \frac{P(X_0 = x, T_y = k, X_m = y)}{P(X_0 = x)} \\ &= \sum_{k=1}^{m} \frac{P(X_0 = x, T_y = k)}{P(X_0 = x)} \frac{P(X_0 = x, T_y = k, X_m = y)}{P(X_0 = x, T_y = k)} \\ &= \sum_{k=1}^{m} P_x(T_y = k)P(X_m = y|X_0 = x, x \neq y, x_k = y) \\ &= \sum_{k=1}^{m} P_x(T_y = k)P(X_m = y|x_k = y) \end{aligned}$$

□

Before we define recurrent and transient states, we introduce the following notation:

$$P_{xy} = P_x(T_y \leq \infty) = \sum_k P_x(T_y = k).$$

Definition 2.9 (Recurrent and Transient states)**.** *A state x is called* recurrent *if $P_{xx} = 1$. Ohterwise, it is called* transient.

We introduce more notations:

- $I_x(y) = \begin{cases} 1 & y = x \\ 0 & \text{else} \end{cases}$ (indicator function of x).
- $N(y) = \sum_{n=1}^{\infty} I_y(X_n)$

Theorem 2.2.

1. $P_x(N(y) \geq m) = P_{xy} P_{yy}^{m-1}$
2. $P_x(N(y) = m) = P_{xy} P_{yy}^{m-1}(1 - P_{yy})$
3. $P_x(N(y) = 0) = 1 - P_{xy}$

Proof. First, assume that theorem 1 is true. Then, we have

$$\begin{aligned} P_x(N(y) = m) &= P_x(N(y) \geq m) - P_x(N(y) \geq m+1) \\ &= P_{xy} P_{yy}^{m-1} - P_{xy} P_{yy}^{m} \\ &= P_{xy} P_{yy}^{m-1}(1 - P_{yy}) \end{aligned}$$

Now, we want to prove theorem 3:

$$\begin{aligned} P_x(N(y) = 0) &= 1 - P_x(N(y) \geq 1) \\ &= 1 - P_{xy} \end{aligned}$$

□