

richardwu.ca

# STAT 333 COURSE NOTES

APPLIED PROBABILITY

STEVE DREKIC • WINTER 2018 • UNIVERSITY OF WATERLOO

Last Revision: March 5, 2018

## Table of Contents

<b>1</b>	<b>January 4, 2018</b>	<b>1</b>
1.1	Example 1.1 solution . . . . .	1
1.2	Example 1.2 solution . . . . .	1
<b>2</b>	<b>January 9, 2018</b>	<b>1</b>
2.1	Example 1.3 solution . . . . .	1
2.2	Example 1.4 solution . . . . .	2
2.3	Example 1.5 solution . . . . .	3
2.4	Exercise 1.3 . . . . .	3
<b>3</b>	<b>January 11, 2018</b>	<b>4</b>
3.1	Theorem 2.1 (conditional variance) . . . . .	4
3.2	Example 2.1 . . . . .	4
3.3	Example 2.2 . . . . .	5
<b>4</b>	<b>Tutorial 1</b>	<b>6</b>
4.1	Exercise 1: MGF of Erlang . . . . .	6
4.2	Exercise 2: MGF of Uniform . . . . .	7
4.3	Exercise 3: Moments from PGF . . . . .	8
4.4	Exercise 4: PGF of Poisson . . . . .	9
<b>5</b>	<b>January 16, 2018</b>	<b>10</b>
5.1	Example 2.3 solution . . . . .	10
5.2	Example 2.4 solution . . . . .	11
5.3	Example 2.5 solution . . . . .	12
<b>6</b>	<b>January 18, 2018</b>	<b>13</b>
6.1	Example 2.6 solution . . . . .	13
6.2	Example 2.7 solution . . . . .	14
6.3	Theorem 2.2 (law of total expectation) . . . . .	15
6.4	Example 2.8 solution . . . . .	16
<b>7</b>	<b>Tutorial 2</b>	<b>16</b>
7.1	Sum of geometric distributions . . . . .	16
7.2	Conditional card drawing . . . . .	17
7.3	Conditional points from interval . . . . .	18

<b>8 January 23, 2018</b>	<b>20</b>
8.1 Example 2.8 solution . . . . .	20
8.2 Theorem 2.3 (variance as expectation of conditionals) . . . . .	21
8.3 Example 2.9 solution . . . . .	22
<b>9 January 25, 2018</b>	<b>23</b>
9.1 Example 2.10 solution ( $P(X < Y)$ ) . . . . .	23
9.2 Example 2.11 solution . . . . .	24
9.3 Example 2.12 solution . . . . .	25
<b>10 Tutorial 3</b>	<b>27</b>
10.1 Mixed conditional distribution . . . . .	27
10.2 Law of total expectations . . . . .	28
10.3 Conditioning on wins and losses . . . . .	29
<b>11 February 1, 2018</b>	<b>30</b>
11.1 Example 3.1 solution . . . . .	30
11.2 Example 3.2 solution . . . . .	32
11.3 Example 3.3 solution . . . . .	32
<b>12 Tutorial 4</b>	<b>32</b>
12.1 Law of total expectations with indicator variables . . . . .	32
12.2 Discrete time Markov chain urn example . . . . .	33
12.3 Discrete time Markov chain weather example . . . . .	34
<b>13 February 6, 2018</b>	<b>35</b>
13.1 Example 3.4 solution . . . . .	35
13.2 Example 3.2 (continued) solution . . . . .	35
13.3 Example 3.5 solution . . . . .	36
13.4 Theorem 3.1 (equivalent states have equivalent periods) . . . . .	36
13.5 Example 3.6 solution . . . . .	37
<b>14 February 8, 2018</b>	<b>37</b>
14.1 Theorem 3.2 (communication and recurrent state $i$ implies recurrent state $j$ ) . . . . .	37
14.2 Theorem 3.3 (communication and recurrent state $i$ implies mutual recurrence among all states) . . . . .	38
14.3 Theorem 3.4 (finite-state DTMCs have at least one recurrent state) . . . . .	38
<b>15 Tutorial 5</b>	<b>39</b>
15.1 Determining diagram, equivalence classes, period, and transience/recurrence of DTMC . . . . .	39
15.2 Discrete time Markov chain consecutive successes example . . . . .	40
15.3 Limiting behaviour of discrete Markov chains . . . . .	41
<b>16 February 13, 2018</b>	<b>42</b>
16.1 Theorem 3.5 (recurrent $i$ and not communicate with $j$ implies $P_{i,j} = 0$ ) . . . . .	42
16.2 Example 3.3 (continued) solution . . . . .	42
16.3 Example 3.7 solution . . . . .	43
16.4 Example 3.8 solution . . . . .	44

<b>17 February 15, 2018</b>	<b>48</b>
17.1 Example 3.9 solution . . . . .	48
17.2 Example 3.10 solution . . . . .	48
<b>18 February 27, 2018</b>	<b>49</b>
18.1 Example 3.11 solution . . . . .	49
18.2 Theorem 3.6 solution . . . . .	49
18.3 Example 3.10 (continued) solution . . . . .	50
18.4 Example 3.12 solution (Gambler's Ruin) . . . . .	51

---

### Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. These notes are my interpretation and transcription of the content covered in lectures. The instructor has not verified or confirmed the accuracy of these notes, and any discrepancies, misunderstandings, typos, etc. as these notes relate to course's content is not the responsibility of the instructor. If you spot any errors or would like to contribute, please contact me directly.

## 1 January 4, 2018

### 1.1 Example 1.1 solution

What is the probability that we roll a number less than 4 given that we know it's odd?

**Solution.** Let  $A = \{1, 2, 3\}$  (less than 4) and  $B = \{1, 3, 5\}$  (odd). We want to find  $P(A | B)$ . Note that  $A \cap B = \{1, 3\}$  and there are six elements in the sample space  $S$  thus

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

### 1.2 Example 1.2 solution

Show that  $BIN(n, p) \sim POI(\lambda)$  when  $\lambda = np$  for  $n$  large and  $p$  small.

**Solution.** Let  $\lambda = np$ . Note that  $p = \frac{\lambda}{n}$   $n > 0$ . From the pmf for  $X \sim BIN(n, p)$

$$\begin{aligned} p(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n(n-1)\dots(n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x} \end{aligned}$$

Recall  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$  so

$$\lim_{n \rightarrow \infty} p(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

## 2 January 9, 2018

### 2.1 Example 1.3 solution

Find the mgf of  $BIN(n, p)$  and use that to find  $E[X]$  and  $Var(X)$ .

**Solution.** Recall the binomial series is

$$(a+b)^m = \sum_{x=0}^m \binom{m}{x} a^x b^{m-x} \quad a, b \in \mathbb{R}, m \in \mathbb{N}$$

Let  $x \sim BIN(n, p)$  and so

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n$$

Taking the mgf  $E[e^{tX}]$

$$\begin{aligned}\Phi_X(t) &= E[e^{tX}] = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x}\end{aligned}$$

from the binomial series we have

$$\Phi_X(t) = (pe^t + 1 - p)^n \quad t \in \mathbb{R}$$

We can take the first and second derivatives for the first and second moment

$$\begin{aligned}\Phi'_X(t) &= n(pe^t + 1 - p)^{n-1} pe^t \\ \Phi''_X(t) &= np[(pe^t + 1 - p)^{n-1} e^t + e^t(n-1)(pe^t + 1 - p)^{n-2} pe^t]\end{aligned}$$

So  $E[X] = \Phi'_X(t) |_{t=0} = np$ .

For the variance, we need the second moment

$$\begin{aligned}E[X^2] &= \Phi''_X(t) |_{t=0} \\ &= np[1 + (n-1)p] \\ &= np + (np)^2 - np^2\end{aligned}$$

So

$$\begin{aligned}Var(X) &= E[X^2] - E[X]^2 \\ &= np + (np)^2 - np^2 - (np)^2 \\ &= np(1-p)\end{aligned}$$

## 2.2 Example 1.4 solution

Show that  $Cov(X, Y) = 0 \not\Rightarrow$  independence.

**Solution.** We show this using a counter example

$p(x, y)$		$y$		$p_X(x)$
		0	1	
$x$	0	0.2	0	0.2
	1	0	0.6	0.6
	2	0.2	0	0.2
$p_Y(y)$		0.4	0.6	1

Note that

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

where

$$\begin{aligned}
 E[XY] &= \sum_{x=0}^2 \sum_{y=0}^1 xyp(x, y) = (1)(1)(0.6) = 0.6 \\
 E[X] &= \sum_{x=0}^2 xp_X(x) = (1)(0.6) + (2)(0.2) = 0.6 + 0.4 = 1 \\
 E[Y] &= \sum_{y=0}^1 yp_Y(y) = (1)(0.6) = 0.6
 \end{aligned}$$

So  $Cov(X, Y) = 0.6 - (1)(0.6) = 0$ . However,  $p(2, 0) = 0.2 \neq p_X(2)p_Y(0) = (0.2)(0.4) = 0.08$ , thus  $X$  and  $Y$  are not independent (they are dependent).

### 2.3 Example 1.5 solution

Given  $X_1, \dots, X_n$  are independent r.v's where  $\Phi_X(t)$  is the mgf of  $X_i$ , show that  $T = \sum_{i=1}^n X_i$  has mgf  $\Phi_T(t) = \prod_{i=1}^n \Phi_{X_i}(t)$ .

**Solution.** We take the definition of the mgf of  $T$

$$\begin{aligned}
 \Phi_T(t) &= E[e^{tT}] \\
 &= E[e^{t(X_1 + \dots + X_n)}] \\
 &= E[e^{tX_1} \cdot \dots \cdot e^{tX_n}] \\
 &= E[e^{tX_1}] \cdot \dots \cdot E[e^{tX_n}] && \text{independence} \\
 &= \prod_{i=1}^n \Phi_{X_i}(t)
 \end{aligned}$$

### 2.4 Exercise 1.3

If  $X_i \sim POI(\lambda_i)$  show that  $T = \sum X_i \sim POI(\sum \lambda_i)$ .

**Solution.** Recall that  $POI(\lambda_i) \sim BIN(n_i, p)$  where  $\lambda_i = n_i p$  and

$$\Phi_{X_i}(t) = (pe^t + 1 - p)^{n_i} \quad \forall t \in \mathbb{R}$$

where  $X_i \sim BIN(n_i, p) \quad i = 1, \dots, m$ .

Therefore

$$\begin{aligned}
 \Phi_T(t) &= \prod_{i=1}^m (pe^t + 1 - p)^{n_i} \\
 &= (pe^t + 1 - p)^{n_1} \cdot \dots \cdot (pe^t + 1 - p)^{n_m} \\
 &= (pe^t + 1 - p)^{\sum n_i} \quad t \in \mathbb{R}
 \end{aligned}$$

By the mgf uniqueness property, we have

$$T = \sum_{i=1}^m X_i \sim BIN\left(\sum_{i=1}^m n_i, p\right)$$

### 3 January 11, 2018

#### 3.1 Theorem 2.1 (conditional variance)

**Theorem 3.1.**

$$\text{Var}(X_1 | X_2 = x_2) = E[X_1^2 | X_2 = x_2] - E[X_1 | X_2 = x_2]^2$$

*Proof.*

$$\begin{aligned} \text{Var}(X_1 | X_2 = x_2) &= E[(X_1 - E[X_1 | X_2 = x_2])^2 | X_2 = x_2] \\ &= E[(X_1^2 - 2E[X_1 | X_2 = x_2]X_1 + E[X_1 | X_2 = x_2]^2) | X_2 = x_2] \\ &= E[X_1^2 | X_2 = x_2] - 2E[X_1 | X_2 = x_2]E[X_1 | X_2 = x_2] + E[X_1 | X_2 = x_2]^2 \\ &= E[X_1^2 | X_2 = x_2] - E[X_1 | X_2 = x_2]^2 \end{aligned}$$

□

#### 3.2 Example 2.1

Suppose that  $X$  and  $Y$  are discrete random variables having joint pmf of the form

$$p(x, y) = \begin{cases} 1/5 & , \text{if } x = 1 \text{ and } y = 0, \\ 2/15 & , \text{if } x = 0 \text{ and } y = 1, \\ 1/15 & , \text{if } x = 1 \text{ and } y = 2, \\ 1/5 & , \text{if } x = 2 \text{ and } y = 0, \\ 2/5 & , \text{if } x = 1 \text{ and } y = 1, \\ 0 & , \text{otherwise.} \end{cases}$$

Find the conditional probability of  $X | (Y = 1)$ . Also calculate  $E[X | Y = 1]$  and  $\text{Var}(X | Y = 1)$ .

**Solution.** Note: for problems of this nature, construct a table.

		$y$			
$p(x, y)$		0	1	2	$p_X(x)$
$x$	0	0	2/15	0	2/15
	1	1/5	2/5	1/15	2/3
	2	1/5	0	0	1/5
$p_Y(y)$		2/5	8/15	1/15	1

Then we have

$$\begin{aligned} p(0 | 1) &= P(X = 0 | Y = 1) = \frac{2/15}{8/15} = \frac{1}{4} \\ p(1 | 1) &= P(X = 1 | Y = 1) = \frac{2/5}{8/15} = \frac{3}{4} \\ p(2 | 1) &= P(X = 2 | Y = 1) = \frac{0}{8/15} = 0 \end{aligned}$$

The conditional pmf of  $X | (Y = 1)$  can be represented as follows

$x$	0	1
$p(x   1)$	1/4	3/4

We observe  $X | (Y = 1) \sim \text{Bern}(3/4)$ . We can take the known  $E[X] = p$  and  $\text{Var}(X)p(1-p)$  for  $X \sim \text{Bern}(p)$ , thus

$$\begin{aligned} E[X | (Y = 1)] &= 3/4 \\ \text{Var}(X | (Y = 1)) &= 3/4(1 - 3/4) = 3/16 \end{aligned}$$

### 3.3 Example 2.2

For  $i = 1, 2$  suppose that  $X_i \sim \text{BIN}(n_i, p)$  where  $X_1, X_2$  are independent (but not identically distributed). Find conditional distribution of  $X_1$  given  $X_1 + X_2 = n$ .

**Solution.** We want to find conditional pmf of  $X | (X_1 + X_2 = n)$ . Let this conditional pmf be denoted by

$$\begin{aligned} p(x_1 | n) &= P(X_1 = x_1 | X_1 + X_2 = n) \\ &= \frac{P(X_1 = x_1, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \end{aligned}$$

Recall:  $X_1 + X_2 \sim \text{BIN}(n_1 + n_2, p)$  so

$$P(X_1 + X_2 = n) = \binom{n_1 + n_2}{n} p^n (1-p)^{n_1 + n_2 - n}$$

Next, consider

$$\begin{aligned} P(X_1 = x_1, X_1 + X_2 = n) &= P(X_1 = x_1, x_1 + X_2 = n) \\ &= P(X_1 = x_1, X_2 = n - x_1) \\ &= P(X_1 = x_1)P(X_2 = n - x_1) && \text{independence} \\ &= \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} \cdot \binom{n_2}{n - x_1} p^{n - x_1} (1-p)^{n_2 - (n - x_1)} \end{aligned}$$

provided that  $0 \leq x_1 \leq n_1$  and

$$\begin{aligned} 0 &\leq n - x_1 \leq n_2 \\ -n_2 &\leq x_1 - n \leq 0 \\ n - n_2 &\leq x_1 \leq n \end{aligned}$$

(from the binomial coefficients). Therefore our domain for  $x_1$  is

$$x_1 = \max\{0, n - n_2\}, \dots, \min\{n_1, n\}$$



Thus we have

$$\begin{aligned}
 p(x_1 | n) &= \frac{P(X_1 = x_1, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \\
 &= \frac{\binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \cdot \binom{n_2}{n-x_1} p^{n-x_1} (1-p)^{n_2-(n-x_1)}}{\binom{n_1+n_2}{n} p^n (1-p)^{n_1+n_2-n}} \\
 &= \frac{\binom{n_1}{x_1} \binom{n_2}{n-x_1}}{\binom{n_1+n_2}{n}}
 \end{aligned}$$

for  $x_1 = \max\{0, n - n_2\}, \dots, \min\{n_1, n\}$ .

Recall: A  $HG(N, r, n)$  (hypergeometric) distribution has pmf

$$\frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \quad x = \max\{0, n - N + r\}, \dots, \min\{n, r\}$$

So this is precisely  $HG(n_1 + n_2, x_1, n)$ .

If you think about it: we are choosing  $x_1$  successes from  $n_1$  trials from the first set  $X_1$  and choosing the remaining  $n - x_1$  successes from  $n_2$  trials from  $X_2$ .

## 4 Tutorial 1

### 4.1 Exercise 1: MGF of Erlang

Find the mgf of  $X \sim \text{Erlang}(\lambda)$  and use it to find  $E[X], \text{Var}(X)$ .

Note that the Erlang's pdf is for  $n \in \mathbb{Z}^+$  and  $\lambda > 0$

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad x > 0$$

**Solution.**

$$\begin{aligned}
 \Phi_X(t) &= E[e^{tX}] = \int_0^\infty e^{tx} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx \\
 &= \int_0^\infty \frac{\lambda^n x^{n-1} e^{-(\lambda-t)x}}{(n-1)!} dx
 \end{aligned}$$

Note that the term in the integral is similar to the pdf of Erlang but for  $\lambda = \lambda - t$ . So we try to fix it so the integral is this pdf of Erlang

$$\begin{aligned}
 \Phi_X(t) &= \int_0^\infty \frac{\lambda^n x^{n-1} e^{-(\lambda-t)x}}{(n-1)!} dx \\
 &= \left(\frac{\lambda}{\lambda-t}\right)^n \int_0^\infty \frac{(\lambda-t)^n x^{n-1} e^{-(\lambda-t)x}}{(n-1)!} dx \\
 &= \left(\frac{\lambda}{\lambda-t}\right)^n \quad t < \lambda
 \end{aligned}$$

since the integral over the positive real line of the pdf of an  $\text{Erlang}(n, \lambda - t)$  is 1 and  $t < \lambda$  must hold so the rate parameter  $\lambda - t$  is positive.

Differentiating,

$$\begin{aligned}\Phi_X^{(1)}(t) &= \frac{d}{dt} \left( \frac{\lambda}{(\lambda - t)^n} \right) \\ &= \frac{n\lambda^n}{(\lambda - t)^{n+1}} \\ \Phi_X^{(2)}(t) &= \frac{d}{dt} \left( \frac{n\lambda^n}{(\lambda - t)^{n+1}} \right) \\ &= \frac{n(n+1)\lambda^n}{(\lambda - t)^{n+2}}\end{aligned}$$

Thus we have

$$\begin{aligned}E[X] &= \Phi_X^{(1)}(0) = \frac{n\lambda^n}{(\lambda - t)^{n+1}} \Big|_{t=0} = \frac{n}{\lambda} \\ E[X^2] &= \Phi_X^{(2)}(0) = \frac{n(n+1)\lambda^n}{(\lambda - t)^{n+2}} \Big|_{t=0} = \frac{n(n+1)}{\lambda^2} \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \frac{n(n+1)}{\lambda^2} - \frac{n}{\lambda} = \frac{n}{\lambda^2}\end{aligned}$$

**Remark 4.1.** To solve any of these mgfs, it is useful to see if one can reduce the integral into a pdf of a known distribution (possibly itself).

## 4.2 Exercise 2: MGF of Uniform

Find the mgf of the uniform distribution on  $(0, 1)$  and find  $E[X]$  and  $\text{Var}(X)$ .

**Solution.** Let  $X \sim U(0, 1)$  so that  $f(x) = 1$   $0 \leq x \leq 1$ . We have

$$\begin{aligned}\Phi_X(t) &= E[e^{tX}] = \int_0^1 e^{tx}(1)dx \\ &= \frac{1}{t} e^{tx} \Big|_{x=0}^{x=1} \\ &= t^{-1}(e^t - 1) \quad t \neq 0\end{aligned}$$

Differentiating

$$\begin{aligned}\Phi_X^{(1)}(t) &= \frac{d}{dt}(t^{-1}(e^t - 1)) \\ &= t^{-1}e^t - t^{-2}(e^t - 1) \\ &= \frac{te^t - e^t + 1}{t^2} \\ \Phi_X^{(2)}(t) &= \frac{d}{dt} \left( \frac{te^t - e^t + 1}{t^2} \right) \\ &= \frac{t^2(te^t + e^t - e^t) - 2t(te^t - e^t + 1)}{t^4} \\ &= \frac{t^2e^t - 2te^t + 2e^t - 2}{t^3}\end{aligned}$$

We may calculate the first two moments by applying **L'Hopital's rule** to calculate the limits

$$\begin{aligned} E[X] &= \Phi_X^{(1)}(t) \Big|_{t=0} = \lim_{t \rightarrow \infty} \frac{te^t - e^t + 1}{t^2} \\ &= \lim_{t \rightarrow \infty} \frac{te^t + e^t - e^t}{2t} \\ &= \lim_{t \rightarrow \infty} \frac{e^t}{2} = \frac{1}{2} \end{aligned}$$

Similarly

$$\begin{aligned} E[X^2] &= \Phi_X^{(2)}(t) \Big|_{t=0} = \lim_{t \rightarrow \infty} \frac{t^2e^t - 2te^t + 2e^t - 2}{t^3} \\ &= \lim_{t \rightarrow \infty} \frac{t^2e^t + 2te^t - 2te^t - 2e^t + 2e^t}{3t^2} \\ &= \lim_{t \rightarrow \infty} \frac{e^t}{3} = \frac{1}{3} \end{aligned}$$

So we have

$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

### 4.3 Exercise 3: Moments from PGF

Suppose  $X$  is a discrete r.v. on  $\mathbb{N}$  with pmf  $p(x)$ . Show how to find the first two moments of  $X$  from its pgf.

**Solution.** By definition, the pgf of  $X$  is  $\Psi_X(z) = E[z^X] = \sum_{x=0}^{\infty} z^x p(x)$ .

If we let  $z = 1$ , then the sum equals 1. However, if we take its derivative with respect to  $z$  just once

$$\Psi_X^{(1)}(z) = \frac{d}{dz} \sum_{x=0}^{\infty} z^x p(x) = \sum_{x=1}^{\infty} x z^{x-1} p(x)$$

Letting  $z = 1$  we can find the first moment

$$\begin{aligned} \Psi_X^{(1)}(1) &= \lim_{z \rightarrow 1} \sum_{x=1}^{\infty} x z^{x-1} p(x) \\ &= \sum_{x=1}^{\infty} x p(x) \\ &= \sum_{x=0}^{\infty} x p(x) && \text{when } x = 0 \text{ the term is 0 anyways} \\ &= E[X] \end{aligned}$$

For the second moment, we consider the second derivative

$$\begin{aligned} \Psi_X^{(2)}(z) &= \frac{d^2}{dz^2} \sum_{x=0}^{\infty} z^x p(x) \\ &= \sum_{x=2}^{\infty} x(x-1) z^{x-2} p(x) \end{aligned}$$

Letting  $z = 1$

$$\begin{aligned}
 \Psi_X^{(2)}(1) &= \lim_{z \rightarrow 1} \sum_{x=2}^{\infty} x(x-1)z^{x-2}p(x) \\
 &= \sum_{x=2}^{\infty} x(x-1)p(x) \\
 &= \sum_{x=0}^{\infty} x(x-1)p(x) \\
 &= E[X(X-1)] \\
 &= E[X^2] - E[X]
 \end{aligned}$$

So we have  $E[X^2] = \Psi_X^{(2)}(1) + \Psi_X^{(1)}(1)$ . To find the variance

$$Var(X) = \Psi_X^{(2)}(1) + \Psi_X^{(1)}(1) - (\Psi_X^{(1)}(1))^2$$

#### 4.4 Exercise 4: PGF of Poisson

Suppose  $X \sim POI(\lambda)$ . Find the pgf of  $X$  and use it to find  $E[X]$  and  $Var(X)$ . The pmf of  $POI(\lambda)$  for  $\lambda > 0$

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

**Solution.**

$$\begin{aligned}
 \Psi_X(z) = E[z^X] &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(z\lambda)^x}{x!} \\
 &= e^{-\lambda} \cdot e^{z\lambda} \\
 &= e^{\lambda(z-1)}
 \end{aligned}$$

where the second equality holds since the summation is the Taylor expansion of  $e^{z\lambda}$ .  
Differentiating

$$\begin{aligned}
 \Psi_X^{(1)}(z) &= \frac{d}{dz} e^{\lambda(z-1)} \\
 &= \lambda e^{\lambda(z-1)} \\
 \Psi_X^{(2)}(z) &= \frac{d}{dz} \lambda e^{\lambda(z-1)} \\
 &= \lambda^2 e^{\lambda(z-1)}
 \end{aligned}$$

The moments are thus

$$\begin{aligned} E[X] &= \Phi_X^{(1)}(1) = \lambda e^{\lambda(1-1)} = \lambda \\ E[X(X-1)] &= \Phi_X^{(2)}(1) = \lambda^2 e^{\lambda(1-1)} = \lambda^2 \\ E[X^2] &= E[X(X-1)] + E[X] = \lambda^2 + \lambda \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

## 5 January 16, 2018

### 5.1 Example 2.3 solution

Let  $X_1, \dots, X_m$  be independent r.v.'s where  $X_i \sim \text{POI}(\lambda_i)$ . Define  $Y = \sum_{i=1}^m X_i$ . Find the conditional distribution  $X_j \mid (Y = n)$ .

**Solution.** We set out to find

$$\begin{aligned} p(x_j \mid n) &= p(X_j = x_j \mid Y = n) = \frac{P(X_j = x_j, Y = n)}{P(Y = n)} \\ &= \frac{P(X_j = x_j, \sum_{i=1}^m X_i = n)}{P(Y = n)} \\ &= \frac{P(X_j = x_j, X_j + \sum_{i=1, i \neq j}^m X_i = n)}{P(Y = n)} \\ &= \frac{P(X_j = x_j, \sum_{i=1, i \neq j}^m X_i = n - x_j)}{P(Y = n)} \\ &= \frac{P(X_j = x_j) P(\sum_{i=1, i \neq j}^m X_i = n - x_j)}{P(Y = n)} \quad \text{independence of } X_i \end{aligned}$$

Remember that if  $X_i \sim \text{POI}(\lambda_i)$ , then

$$Y = \sum_{i=1}^m X_i \sim \text{POI}(\sum_{i=1}^m \lambda_i)$$

which can be derived from mgfs (Exercise 1.3). Therefore

$$\sum_{i=1, i \neq j}^m X_i \sim \text{POI}(\sum_{i=1, i \neq j}^m \lambda_i)$$

Expanding out  $p(x_j \mid n)$  with the pdfs

$$p(x_j \mid n) = \frac{\frac{e^{-\lambda_j} \lambda_j^{x_j}}{x_j!} \cdot \frac{e^{-\sum_{i=1, i \neq j}^m \lambda_i} (\sum_{i=1, i \neq j}^m \lambda_i)^{n-x_j}}{(n-x_j)!}}{\frac{e^{-\sum_{i=1}^m \lambda_i} (\sum_{i=1}^m \lambda_i)^n}{n!}}$$

where  $x_j \geq 0$  and  $n - x_j \geq 0 \Rightarrow 0 \leq x_j \leq n$  (from the factorials).

Cancelling out the  $e^\lambda$  terms and let  $\lambda_Y = \sum_{i=1}^m \lambda_i$

$$\begin{aligned} p(x_j | n) &= \frac{n!}{(n-x_j)!x_j!} \frac{\lambda_j^{x_j}}{\lambda_Y^{x_j}} \frac{(\lambda_Y - \lambda_j)^{n-x_j}}{\lambda_Y^{n-x_j}} \\ &= \binom{n}{x_j} \left(\frac{\lambda_j}{\lambda_Y}\right)^{x_j} \left(1 - \frac{\lambda_j}{\lambda_Y}\right)^{n-x_j} \end{aligned}$$

This is the binomial distribution, so we have

$$X_j | Y = n \sim \text{BIN}\left(n, \frac{\lambda_j}{\lambda_Y}\right)$$

## 5.2 Example 2.4 solution

Suppose  $X \sim \text{POI}(\lambda)$  and  $Y | (X = x) \sim \text{BIN}(x, p)$ . Find the conditional distribution  $X | Y = y$ .

(Note: range of  $y$  depends on  $x$  (that is  $y \leq x$ ). Graphically, we have integral points on and below the  $y = x$  line starting from 0 for both  $x$  and  $y$ ).

**Solution.** We wish to find the conditional pmf given by  $X | Y = y$  or

$$p(x | y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Note that also

$$\begin{aligned} P(Y = y | X = x) &= \frac{P(Y = y, X = x)}{P(X = x)} \\ \Rightarrow P(X = x, Y = y) &= P(X = x)P(Y = y | X = x) \\ &= \frac{e^{-\lambda} \lambda^x}{x!} \cdot \binom{x}{y} p^y (1-p)^{x-y} \end{aligned}$$

for  $x = 0, 1, 2, \dots$  **and**  $y = 0, 1, 2, \dots, x$  (range of  $y$  depends on  $x$ ).

To find the marginal pmf of  $Y$ , we use

$$p_Y(y) = \sum_x p(x, y)$$

To find the support for  $x$ , note that from the graphical region, we realize that  $x = 0, 1, 2, \dots$  **and**  $y = 0, 1, 2, \dots, x$  is equivalent to  $y = 0, 1, 2, \dots$  **and**  $x = y, y+1, y+2, \dots$

So

$$\begin{aligned}
 p_Y(y) &= \sum_{x=y}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \frac{x!}{(x-y)!y!} p^y (1-p)^{x-y} \\
 &= \frac{\lambda^y e^{-\lambda} p^y}{y!} \sum_{x=y}^{\infty} \frac{\lambda^{x-y} (1-p)^{x-y}}{(x-y)!} \\
 &= \frac{e^{-\lambda} (\lambda p)^y}{y!} \sum_{x=y}^{\infty} \frac{[\lambda(1-p)]^{x-y}}{(x-y)!} \\
 &= \frac{e^{-\lambda} (\lambda p)^y}{y!} e^{\lambda(1-p)} \\
 &= \frac{e^{-\lambda p} (\lambda p)^y}{y!} \quad y = 0, 1, 2, \dots
 \end{aligned}$$

Note that  $p_Y(y) \sim \text{POI}(\lambda p)$ .

Thus

$$\begin{aligned}
 p(x | y) &= \frac{P(X = x, Y = y)}{P(Y = y)} \\
 &= \frac{\frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{x!}{(x-y)!y!} p^y (1-p)^{x-y}}{\frac{e^{-\lambda p} (\lambda p)^y}{y!}} \\
 &= \frac{e^{-\lambda + \lambda p} [\lambda(1-p)]^{x-y}}{(x-y)!} \\
 &= \frac{e^{-\lambda(1-p)} [\lambda(1-p)]^{x-y}}{(x-y)!} \quad x = y, y+1, y+2, \dots
 \end{aligned}$$

This resembles the POIson distribution with  $\lambda = \lambda(1-p)$  but with a slightly modified domain.

So we see that

$$W | (Y = y) \sim W + y$$

where  $W \sim \text{POI}(\lambda(1-p))$ . This is the **shifted Poisson pmf**  $y$  units to the right (note that  $W$  and  $y$  are random variables).

We can easily find the conditional expectations and variance e.g.

$$E[X | Y = y] = E[W + y] = E[W] + y$$

### 5.3 Example 2.5 solution

Suppose the joint pdf of  $X$  and  $Y$  is

$$f(x, y) = \begin{cases} \frac{12}{5} x(2-x-y) & , 0 < x < 1, 0 < y < 1, \\ 0 & , \text{ elsewhere} \end{cases}$$

Determine the conditional distribution of  $X$  given  $Y = y$  where  $0 < y < 1$ . Also calculate the mean of  $X | (Y = y)$ . (Note: the graphical region is a unit square box where the bottom left corner is at  $0, 0$ : the inside of the box is the support).

**Solution.** Using our theory, we wish to find the conditional pdf of  $X \mid (Y = y)$  given by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

For  $0 < y < 1$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_0^1 \frac{12}{5} x(2 - x - y) dx \\ &= \frac{12}{5} \int_0^1 (2x - x^2 - xy) dx \\ &= \frac{12}{5} \left( x^2 - \frac{x^3}{3} - \frac{x^2 y}{2} \right) \Big|_0^1 \\ &= \frac{12}{5} \left( 1 - \frac{1}{3} - \frac{y}{2} \right) \\ &= \frac{2}{5} (4 - 3y) \end{aligned}$$

So we have

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{\frac{12}{5} x(2 - x - y)}{\frac{2}{5} (4 - 3y)} \\ &= \frac{6x(2 - x - y)}{4 - 3y} \end{aligned}$$

Thus we have

$$\begin{aligned} E[X \mid Y] &= \int_0^1 x \cdot f_{X|Y}(x \mid y) dx \\ &= \frac{5 - 4y}{2(4 - 3y)} \end{aligned}$$

## 6 January 18, 2018

### 6.1 Example 2.6 solution

Suppose the joint pdf of  $X$  and  $Y$  is

$$f(x, y) = \begin{cases} 5e^{-3x-y} & , 0 < 2x < y < \infty, \\ 0 & , \text{otherwise} \end{cases}$$

Find the conditional distribution of  $Y \mid (X = x)$  where  $0 < x < \infty$ .

Note the region of support is a “flag” (upright triangle with downward point) where the slanted part is the line  $y = 2x$ .

**Solution.** We wish to find

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$



For  $0 < x < \infty$

$$\begin{aligned}
 f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\
 &= \int_{2x}^{\infty} 5e^{-3x-y} dy \\
 &= 5e^{-3x} \int_{2x}^{\infty} 5e^{-y} dy \\
 &= 5e^{-3x} (-e^{-y}) \Big|_{2x}^{\infty} \\
 &= 5e^{-3x} e^{-2x} \\
 &= 5e^{-5x}
 \end{aligned}$$

so we have  $f_X(x) \sim \text{Exp}(5)$ .

**Remark 6.1.** The bounds on the integral are in terms of  $y$ : it is dependent on  $x$  in our  $f(x, y)$  definition.

Now

$$\begin{aligned}
 f_{Y|X}(y | x) &= \frac{5e^{-3x-y}}{5e^{-5x}} \\
 &= e^{-y+2x} \quad y > 2x
 \end{aligned}$$

**Note:** recognize the conditional pdf of  $Y | (X = x)$  as that of a shifted exponential distribution ( $2x$  units to the right). Specifically, we have

$$Y | (X = x) \sim W + 2x$$

where  $W \sim \text{Exp}(1)$ . Thus  $E[Y | (X = x)] = E(W) + 2x$  and  $\text{Var}[Y | (X = x)] = \text{Var}(W)$ .

## 6.2 Example 2.7 solution

Suppose  $X \sim U(0, 1)$  and  $Y | (X = x) \sim \text{Bern}(x)$ . Find the conditional distribution  $X | (Y = y)$ .

Note:  $X$  is continuous and  $Y | (X = x)$  is discrete.

**Solution.** We wish to find

$$f_{X|Y}(x | y) = \frac{p(y | x)f_X(x)}{p_Y(y)}$$

From the given information, we have  $f_X(x) = 1$  for  $0 < x < 1$ . Furthermore  $p(y | x) = \text{Bern}(x) = x^y(1-x)^{1-y}$  for  $y = 0, 1$ .

For  $y = 0, 1$  note that  $(\int f(x | y) dx = 1)$

$$\begin{aligned}
 p_Y(y) &= \int_{-\infty}^{\infty} p(y | x)f_X(x) dx \\
 p_Y(y) &= \int_0^1 x^y(1-x)^{1-y} dx
 \end{aligned}$$

To compute this integral, let's check  $p_Y(0)$  and  $p_Y(1)$

$$\begin{aligned} p_Y(0) &= \int_0^1 x^0(1-x)^{1-0} dx \\ &= \int_0^1 1-x dx \\ &= x - \frac{x^2}{2} \Big|_0^1 \\ &= \frac{1}{2} \end{aligned}$$

Similarly, take  $y = 1$  where  $p_Y(1) = \frac{1}{2}$ .

In other words, we have that  $p_Y(y) = \frac{1}{2}$   $y = 0, 1$  so

$$Y \sim \text{Bern}\left(\frac{1}{2}\right)$$

So

$$\begin{aligned} f(x | y) &= \frac{p(y | x)f_X(x)}{p_Y(y)} \\ &= \frac{x^y(1-x)^{1-y} \cdot 1}{\frac{1}{2}} \\ &= 2x^y(1-x)^{1-y} \quad 0 < x < 1 \end{aligned}$$

### 6.3 Theorem 2.2 (law of total expectation)

**Theorem 6.1.** For random variables  $X$  and  $Y$ ,  $E[X] = E[E[X | Y]]$ .

*Proof.* WLOG assume  $X, Y$  are jointly continuous random variables. We note

$$\begin{aligned} E[E[X | Y]] &= \int_{-\infty}^{\infty} E[X | Y = y] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx \right] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} \cdot f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f(x, y) dy \right] dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= E[X] \end{aligned}$$

□

## 6.4 Example 2.8 solution

Suppose  $X \sim GEO(p)$  with pmf  $p_X(x) = (1-p)^{x-1}p$  where  $x = 1, 2, 3, \dots$ . Calculate  $E[X]$  and  $Var(X)$  using the law of total expectation.

**Solution.** Recall  $E[X] = \frac{1}{p}$  and  $Var(X) = \frac{1-p}{p^2}$  where  $X$  models the number of (independent) trials necessary to obtain the first success.

Remember: we could manually solve  $E[X] = \sum_{x=1}^{\infty} (1-p)^{x-1}p$  and similarly  $Var(X) = E[X^2] - E[X]^2$ , or take the derivatives of the mgf  $\Phi_X(t) = E[e^{tX}]$ . This is tedious in general.

## 7 Tutorial 2

### 7.1 Sum of geometric distributions

Let  $X_i$  for  $i = 1, 2, 3$  be independent geometric random variables having the same parameter  $p$ . Determine the value

$$P(X_j = x_j \mid \sum_{i=1}^3 X_i = n)$$

**Solution.** Note that, by construction, the sum of  $k$  independent  $GEO(p)$  random variables is distributed as  $NB(k, p)$ . Recall that

$$\begin{aligned} X_i \sim GEO(p) &\Rightarrow P_{X_i}(x) = (1-p)^{x-1}p & x = 1, 2, 3, \dots \\ Y \sim NB(k, p) &\Rightarrow P_Y(y) = \binom{y-1}{k-1} p^k (1-p)^{y-k} & y = k, k+1, k+2, \dots \end{aligned}$$

Breaking apart the summation

$$\begin{aligned} P(X_j = x_j \mid \sum_{i=1}^3 X_i = n) &= P(X_j = x_j \mid X_j + \sum_{i=1, i \neq j}^3 X_i = n) \\ &= \frac{P(X_j = x_j, X_j + \sum_{i=1, i \neq j}^3 X_i = n)}{P(\sum_{i=1}^3 X_i = n)} \\ &= \frac{P(X_j = x_j, \sum_{i=1, i \neq j}^3 X_i = n - x_j)}{P(\sum_{i=1}^3 X_i = n)} \\ &= \frac{P(X_j = x_j) \cdot P(\sum_{i=1, i \neq j}^3 X_i = n - x_j)}{P(\sum_{i=1}^3 X_i = n)} && X_i\text{'s are independent} \\ &= \frac{(1-p)^{x_j-1}p \cdot \binom{n-x_j-1}{1} p^2 (1-p)^{n-x_j-2}}{\binom{n-1}{2} p^3 (1-p)^{n-3}} && \text{provided that } x_j \geq 1 \text{ and } n - x_j \geq 2 \\ &= \frac{(1-p)^{x_j-1}p \cdot \binom{n-x_j-1}{1} p^2 (1-p)^{n-x_j-2}}{\binom{n-1}{2} p^3 (1-p)^{n-3}} \\ &= \frac{(n-x_j-1)!}{1!(n-x_j-2)!} \cdot \frac{2!(n-3)!}{(n-1)!} \\ &= \frac{2(n-x_j-1)}{(n-1)(n-2)} \quad x_j = 1, 2, \dots, n-2 \end{aligned}$$

Note this is a pmf so we can check

$$\begin{aligned}
 \sum_{x_1}^{n-2} \frac{2(n-x_1)}{(n-1)(n-2)} &= \sum_{x_1}^{n-2} \frac{2(n-1)}{(n-1)(n-2)} - \sum_{x_1}^{n-2} \frac{2x}{(n-1)(n-2)} \\
 &= \frac{2(n-1)(n-2)}{(n-1)(n-2)} - \frac{2}{(n-1)(n-2)} \sum_{x=1}^{n-2} x \\
 &= 2 - \frac{2}{(n-1)(n-2)} \cdot \frac{(n-2)(n-1)}{2} \\
 &= 2 - 1 \\
 &= 1
 \end{aligned}$$

which satisfies the cdf axiom.

## 7.2 Conditional card drawing

Given  $N \in \mathbb{Z}^+$  cards labelled  $1, 2, \dots, N$ , let  $X$  represent the number that is picked. Suppose a second card  $Y$  is picked from  $1, 2, \dots, X$ .

Assuming  $N = 10$ , calculate the expected value of  $X$  given  $Y = 8$ .

**Solution.** Clearly we have that  $P_X(x) = \frac{1}{N}$  where  $x = 1, 2, \dots, N$  and  $P_{Y|X}(y | x) = \frac{1}{x}$  for  $y = 1, 2, \dots, x$ .

To find the conditional distribution of  $X | (Y = y)$  we must identify the joint distribution of  $X, Y$ . It immediately follows that

$$p(x, y) = P(X = x, Y = y) = P_{Y|X}(y | x)P_X(x) = \frac{1}{xN}$$

for  $x = 1, 2, \dots, N$  and  $y = 1, 2, \dots, x$ . or equivalently the range can be re-expressed as

$$y = 1, 2, \dots, N \text{ and } x = y, y + 1, \dots, N$$

**Remark 7.1.** Whenever we want to find the marginal pmf/pdf for a given rv  $Y$ , we generally need to re-map the support such that the support of  $Y$  is independent of the other rv  $X$ .

Note that

$$\begin{aligned}
 P_Y(y) &= \sum_{x=y}^N p(x, y) = \sum_{x=y}^N \frac{1}{xN} \\
 &= \frac{1}{N} \sum_{x=y}^N \frac{1}{x} \quad y = 1, 2, \dots, N
 \end{aligned}$$

Letting  $N = 10$ , we can calculate

$$\begin{aligned}
 E[X | Y = 8] &= \sum_{x=8}^{10} x P_{X|Y}(x | 8) \\
 &= \sum_{x=8}^{10} x \frac{P(x, 8)}{P_Y(8)} \\
 &= \sum_{x=8}^{10} x \frac{\frac{1}{10x}}{\frac{1}{10} \sum_{z=8}^{10} \frac{1}{z}} \\
 &= \sum_{x=8}^{10} x \left( \sum_{z=8}^{10} \frac{1}{z} \right)^{-1} \\
 &= 3 \left( \frac{1}{8} + \frac{1}{9} + \frac{1}{10} \right)^{-1} \\
 &= 3 \left( \frac{242}{720} \right)^{-1} \\
 &= \frac{1080}{121} \approx 8.9256
 \end{aligned}$$

### 7.3 Conditional points from interval

Let us choose a random point from interval  $(0, 1)$  denoted as rv  $X_1$ . We then choose a random point  $X_2$  on the interval  $(0, x_1)$  where  $x_1$  is the realized value of  $X_1$ .

1. Make assumptions about the marginal pdf  $f_1(x_1)$  and conditional pdf  $f_{2|1}(x_2 | x_1)$ .
2. Find the conditional mean  $E[X_1 | X_2 = x_2]$ .
3. Compute  $P(X_1 + X_2 \geq 1)$ .

**Solution.** 1. It makes sense that  $X_1 \sim U(0, 1)$  and  $X_2 | (X_1 = x_1) \sim U(0, x_1)$  so that  $f_1(x_1) = 1$ ,  $0 < x_1 < 1$  and  $f_{2|1}(x_2 | x_1) = \frac{1}{x_1}$  for  $0 < x_2 < x_1 < 1$ .

2. Note that  $f_{1|2}(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$  and so we need to identify the joint distribution of  $x_1$  and  $x_2$  as well as the marginal distribution of  $X_2$ . We have

$$\begin{aligned}
 f(x_1, x_2) &= f_{2|1}(x_2 | x_1) \cdot f_1(x_1) \\
 &= \frac{1}{x_1} \qquad \qquad \qquad 0 < x_2 < x_1 < 1 \quad 0 < x_1 < 1
 \end{aligned}$$

or equivalently, the region of support can be re-expressed as

$$\begin{aligned}
 0 &< x_2 < 1 \\
 x_2 &< x_1 < 1
 \end{aligned}$$

so the marginal pdf of  $f_2(x_2)$  is

$$\begin{aligned} f_2(x_2) &= \int_{x_1=x_2}^1 p(x_1, x_2) dx_1 \\ &= \int_{x_1=x_2}^1 \frac{1}{x_1} dx_1 \\ &= \ln(x_1) \Big|_{x_1=x_2}^{x_1=1} \\ &= -\ln(x_2) \quad 0 < x_2 < 1 \end{aligned}$$

so the conditional pdf is

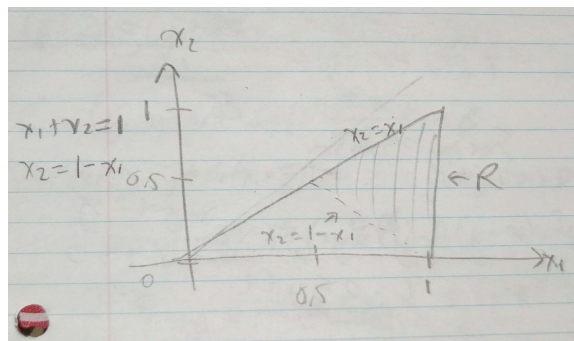
$$\begin{aligned} f_{1|2}(x_1 | x_2) &= \frac{f(x_1, x_2)}{f_2(x_2)} \\ &= \frac{1}{-x_1 \ln(x_2)} \quad 0 < x_2 < x_1 < 1 \end{aligned}$$

Taking the expectation

$$\begin{aligned} E[X_1 | X_2 = x_2] &= \int_{x_1=x_2}^1 x_1 p_{1|2}(x_1, x_2) dx_1 \\ &= \int_{x_1=x_2}^1 x_1 \cdot \frac{1}{-x_1 \ln(x_2)} dx_1 \\ &= \int_{x_1=x_2}^1 \frac{1}{-\ln(x_2)} dx_1 \\ &= \frac{1 - x_2}{-\ln(x_2)} \quad 0 < x_2 < 1 \end{aligned}$$

Exercise: solve for  $\lim_{x_2 \rightarrow 1} E[X_1 | X_2 = x_2]$  (use LHR).

3. The probability that  $X_1 + X_2 \geq 1$  may be calculated by taking the double integral over the region  $R$  of their support where  $X_1 + X_2 \geq 1$  holds. This region may be found as follows:



**Figure 7.1:** The region  $R$  is the support where  $X_1 + X_2 \geq 1$ .

The region  $R$  is equivalent to the bounds  $\frac{1}{2} < x_1 < 1$  and  $1 - x_1 < x_2 < x_1$ .

Integrating  $f(x_1, x_2)$  over  $R$  we obtain

$$\begin{aligned}
 P(X_1 + X_2 \geq 1) &= \int_R \int f(x_1, x_2) dx_2 dx_1 \\
 &= \int_{\frac{1}{2}}^1 \int_{1-x_1}^{x_1} \frac{1}{x_1} dx_2 dx_1 \\
 &= \int_{\frac{1}{2}}^1 \frac{x_2}{x_1} \Big|_{x_2=1-x_1}^{x_2=x_1} dx_1 \\
 &= \int_{\frac{1}{2}}^1 \left(2 - \frac{1}{x_1}\right) dx_1 \\
 &= (2x_1 - \ln(x_1)) \Big|_{x_1=\frac{1}{2}}^{x_1=1} \\
 &= 1 + \ln\left(\frac{1}{2}\right) \\
 &= 1 - \ln(2) \\
 &\approx 0.3068528
 \end{aligned}$$

## 8 January 23, 2018

### 8.1 Example 2.8 solution

Suppose  $X \sim GEO(p)$  with pmf  $p_X(x) = (1-p)^{x-1}p$  for  $x = 1, 2, 3, \dots$ . Calculate  $E[X]$ ,  $Var(X)$  using the law of total expectation.

**Solution.** Recall  $X$  is modelling the number of trials needed to obtain the **1st success**. We want to calculate  $E[X]$  and  $Var(X)$  using the total law of expectation.

Define

$$Y = \begin{cases} 0 & \text{if the 1st trial is a failure} \\ 1 & \text{if the 1st trial is a success} \end{cases}$$

Note that  $Y \sim Bern(p)$  so that  $P_Y(0) = P(Y=0) = 1-p$  and similarly  $P_Y(1) = P(Y=1) = p$ .

Thus by the law of total expectation

$$\begin{aligned}
 E[X] &= E[E[X | Y]] \\
 &= \sum_{y=0}^1 E[X | Y=y] p_Y(y) \\
 &= (1-p)E[X | Y=0] + pE[X | Y=1]
 \end{aligned}$$

Note that

$$X | (Y=1) = 1$$

with probability 1 (one success is equivalent to  $X=1$  for  $GEO(p)$ ), and

$$X | (Y=0) \sim 1 + X$$

(the first one failed, we expect to take  $X$  more trials; same initial problem - recurse. See course notes for formal proof).

Thus we have

$$\begin{aligned}
 E[X] &= (1-p)E[1+X] + p(1) \\
 &= (1-p)(1+E[X]) + p \\
 &= 1 + (1-p)E[X] \\
 &\Rightarrow E[X](1 - (1-p)) = 1 \\
 &\Rightarrow E[X] = \frac{1}{p}
 \end{aligned}$$

as expected.

For  $Var(X)$ , notice that

$$\begin{aligned}
 E[X^2] &= E[E[X^2 | Y]] \\
 &= \sum_{y=0}^1 E[X^2 | Y = y]p_Y(y) \\
 &= (1-p)E[X^2 | Y = 0] + pE[X^2 | Y = 1] \\
 &= (1-p)E[(1+X)^2] + p(1)^2 && \text{from above} \\
 &= (1-p)E[1 + 2X + X^2] + p \\
 &= (1-p)(1 + 2E[X] + E[X^2]) + p \\
 &= 1 + 2(1-p)E[X] + (1-p)E[X^2] \\
 &\Rightarrow E[X^2](1 - (1-p)) = 1 + \frac{2(1-p)}{p} \\
 &\Rightarrow E[X^2] = \frac{1}{p} + \frac{2(1-p)}{p^2}
 \end{aligned}$$

So we have

$$\begin{aligned}
 Var(X) &= E[X^2] - E[X]^2 \\
 &= \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} \\
 &= \frac{p + 2 - 2p - 1}{p^2} \\
 &= \frac{1-p}{p^2}
 \end{aligned}$$

**Remark 8.1.** For law of total expectations, a large part of it is choosing the right random variable to condition on (i.e.  $Y = Bern(p)$  in this example).

## 8.2 Theorem 2.3 (variance as expectation of conditionals)

**Theorem 8.1.** For random variables  $X$  and  $Y$

$$Var(X) = E[Var(X | Y)] + Var(E[X | Y])$$

*Proof.* Recall that

$$Var(X | Y = y) = E[X^2 | Y = y] - E[X | Y = y]^2$$



so more generally we have

$$\text{Var}(X | Y) = E[X^2 | Y] - E[X | Y]^2$$

Taking the expectation of this

$$\begin{aligned} E[\text{Var}(X | Y)] &= E[E[X^2 | Y] - E[X | Y]^2] \\ &= E[E[X^2 | Y]] - E[E[X | Y]^2] \\ &= E[X^2] - E[E[X | Y]^2] \end{aligned} \quad E[A] = E[E[A | B]] \text{ (law of total expectation)}$$

Note that

$$\text{Var}(E[X | Y]) = \text{Var}(v(Y))$$

where  $v(Y) = E[X | Y]$  is a function of  $Y$  (not  $X$ !).

$$\begin{aligned} \text{Var}(v(Y)) &= E[v(Y)^2] - E[v(Y)]^2 \\ &= E[E[X | Y]^2] - E[X]^2 \end{aligned} \quad \text{law of total expectation}$$

Therefore we have

$$\begin{aligned} E[\text{Var}(X | Y)] + \text{Var}(E[X | Y]) &= E[X^2] - E[E[X | Y]^2] + E[E[X | Y]^2] - E[X]^2 \\ &= E[X^2] - E[X]^2 \\ &= \text{Var}(X) \end{aligned}$$

as desired. □

### 8.3 Example 2.9 solution

Suppose  $\{X_i\}_{i=1}^\infty$  is an iid sequence of random variables with common mean  $\mu$  and variance  $\sigma^2$ . Let  $N$  be a discrete, non-negative integer-valued rv that is independent of each  $X_i$ .

Find the mean and variance of  $T = \sum_{i=1}^N X_i$  (referred to as a **random sum**).

**Solution. To find the mean:**

We condition on  $N$  since the value of our  $T$  depends on how many  $X_i$ 's there are which depends on  $N$ . By the law of total expectations

$$E[T] = E[E[T | N]]$$

Note that

$$\begin{aligned} E[T | N = n] &= E\left[\sum_{i=1}^N X_i \mid N = n\right] \\ &= E\left[\sum_{i=1}^n X_i \mid N = n\right] \\ &= \sum_{i=1}^n E[X_i \mid N = n] && \text{due to independence of } X_i \text{ and } N \\ &= \sum_{i=1}^n E[X_i] \\ &= n\mu \end{aligned}$$

So we have  $E[T \mid N] = N\mu$ .

**Remark 8.2.** We needed to first condition on a concrete  $N = n$  in order to unwrap the summation, then revert back to the random variable  $N$ .

Thus we have

$$E[T] = E[E[T \mid N]] = E[N\mu] = \mu E[N]$$

which intuitively makes sense.

**To find the variance:**

We use our previous theorem on variance as expectation of conditionals

$$\text{Var}(T) = E[\text{Var}(T \mid N)] + \text{Var}(E[T \mid N])$$

We know from before that

$$\text{Var}(E[T \mid N]) = \text{Var}(N\mu) = \mu^2 \text{Var}(N)$$

We can break apart the variance as

$$\begin{aligned} \text{Var}(T \mid N = n) &= \text{Var}\left(\sum_{i=1}^N X_i \mid N = n\right) \\ &= \text{Var}\left(\sum_{i=1}^n X_i \mid N = n\right) \\ &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \text{Var}(X_i) && \text{independence of } X_i \\ &= \sigma^2 n \end{aligned}$$

Therefore  $\text{Var}(T \mid N) \text{Var}(T \mid N = n) \Big|_{n=N} = \sigma^2 N$ .

So

$$E[\text{Var}(T \mid N)] = E[\sigma^2 N] = \sigma^2 E[N]$$

and thus

$$\text{Var}(T) = \sigma^2 E[N] + \mu^2 \text{Var}(N)$$

## 9 January 25, 2018

### 9.1 Example 2.10 solution ( $P(X < Y)$ )

Suppose  $X$  and  $Y$  are independent continuous random variables. Find an expression for  $P(X < Y)$ .

**Solution.** Define our event of interest as

$$A = \{X < Y\}$$

Thus we have

$$\begin{aligned}
 P(X < Y) &= P(A) = \int_{-\infty}^{\infty} P(A \mid Y = y) f_Y(y) dy && \text{law of total probability} \\
 &= \int_{-\infty}^{\infty} P(X < Y \mid Y = y) f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} P(X < y \mid Y = y) f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} P(X < y) f_Y(y) dy && X < y \text{ only depends on } X; Y = y \text{ only depends on } Y \\
 &= \int_{-\infty}^{\infty} P(X \leq y) f_Y(y) dy && X \text{ is a continuous rv} \\
 &= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy
 \end{aligned}$$

Suppose that  $X$  and  $Y$  have the same distribution. We expect  $P(X < Y) = \frac{1}{2}$ . Let's verify it with our expression

$$\begin{aligned}
 P(X < Y) &= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} F_Y(y) f_Y(y) dy && X \sim Y
 \end{aligned}$$

Let  $u = F_Y(y)$ , thus  $\frac{du}{dy} = f_Y(y) \iff du = f_Y(y) dy$ . So we have

$$\begin{aligned}
 P(X < Y) &= \int_0^1 u du && \text{domain for a CDF is } [0, 1] \\
 &= \left. \frac{u^2}{2} \right|_0^1 \\
 &= \frac{1}{2}
 \end{aligned}$$

## 9.2 Example 2.11 solution

Suppose  $X \sim \text{Exp}(\lambda_1)$  and  $Y \sim \text{Exp}(\lambda_2)$  are independent exponential rvs. Show that

$$P(X < Y) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

**Solution.** Since  $Y \sim \text{Exp}(\lambda_2)$ , then we have  $f_Y(y) = \lambda_2 e^{-\lambda_2 y}$  for  $y > 0$ . Since  $X \sim \text{Exp}(\lambda_1)$ , we have

$$\begin{aligned}
 F_X(x) &= P(X \leq x) = \int_0^x \lambda_1 e^{-\lambda_1 x} dx \\
 &= -e^{-\lambda_1 x} \Big|_0^x \\
 &= 1 - e^{-\lambda_1 x} \quad x \geq 0
 \end{aligned}$$

From the expression in Example 2.10, we have

$$\begin{aligned}
 P(X < Y) &= \int_0^\infty F_X(y) f_Y(y) dy \\
 &= \int_0^\infty (1 - e^{-\lambda_1 y}) (\lambda_2 e^{-\lambda_2 y}) dy \\
 &= \int_0^\infty \lambda_2 e^{-\lambda y} - \lambda_2 e^{-(\lambda_1 + \lambda_2)y} dy \\
 &= \int_0^\infty \lambda_2 e^{-\lambda y} + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)y} \Big|_0^\infty = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
 &= \frac{\lambda_1}{\lambda_1 + \lambda_2}
 \end{aligned}$$

### 9.3 Example 2.12 solution

Consider an experiment in which independent trials each having probability  $p \in (0, 1)$  are performed until  $k \in \mathbb{Z}^+$  consecutive successes are achieved. Determine the expected number of trials for  $k$  consecutive successes.

**Solution.** Let  $N_k$  be the rv which counts the number of trials needed to obtain  $k$  consecutive successes.

Current goal: we want to find  $E[N_k]$ .

Note: when  $n = 1$ , then we have  $N_1 \sim \text{GEO}(p)$ , and so  $E[N_1] = \frac{1}{p}$ .

For arbitrary  $k \geq 2$ , we will try to find  $E[N_k]$  using the law of total expectations, namely

$$E[N_k] = E[E[N_k | W]]$$

for some  $W$  rv we *choose carefully*.

Suppose we choose  $W$  where (we will later see why this won't work)

$$W = \begin{cases} 0 & \text{if first trial is a failure} \\ 1 & \text{if first trial is a success} \end{cases}$$

So we have

$$\begin{aligned}
 E[N_k] &= \sum_w E[N_k | W = w] P(W = w) \\
 &= P(W = 0) E[N_k | W = 0] + P(W = 1) E[N_k | W = 1] \\
 &= (1 - p) E[N_k | W = 0] + p E[N_k | W = 1]
 \end{aligned}$$

Note that

$$\begin{aligned}
 N_k | (W = 0) &\sim 1 + N_k \\
 N_k | (W = 1) &\sim ?
 \end{aligned}$$

We can't simply have  $N_k | (W = 1) \sim 1 + N_{k-1}$  since  $N_{k-1}$  does not guarantee that the  $k - 1$  consecutive successes are followed immediately after our first  $W = 1$ .

Perhaps we need another  $W$ ,  $W = N_{k-1}$  so we attempt to find

$$E[N_k] = E[E[N_k | N_{k-1}]]$$

Consider

$$E[N_k \mid N_{k-1} = n]$$

conditional on  $N_{k-1} = n$ , defin

$$Y = \begin{cases} 0 & \text{if the } (n+1)\text{th trial is a failure} \\ 1 & \text{if the } (n+1)\text{th trial is a success} \end{cases}$$

Now we have

$$\begin{aligned} E[N_k \mid N_{k-1} = n] &= \sum_y E[N_k \mid N_{k-1} = n, Y = y] P(Y = y \mid N_{k-1} = n) \\ &= P(Y = 0 \mid N_{k-1} = n) E[N_k \mid N_{k-1} = n, Y = 0] \\ &\quad + P(Y = 1 \mid N_{k-1} = n) E[N_k \mid N_{k-1} = n, Y = 1] \\ &= (1-p) E[N_k \mid N_{k-1} = n, Y = 0] + p E[N_k \mid N_{k-1} = n, Y = 1] \quad Y \text{ is independent from } N_{k-1} \end{aligned}$$

Note that

$$\begin{aligned} N_k \mid (N_{k-1} = n \mid Y = 0) &\sim n+1 + N_k \text{ we need to start over again} \\ N_k \mid (N_{k-1} = n \mid Y = 1) &\sim n+1 \text{ with probability } 1 \end{aligned}$$

Therefore

$$\begin{aligned} E[N_k \mid N_{k-1} = n] &= (1-p)(n+1 + E[N_k]) + p(n+1) \\ &= n+1 + (1-p)E[N_k] \end{aligned}$$

which in terms of the rv  $N_{k-1}$

$$E[N_k \mid N_{k-1}] = E[N_k \mid N_{k-1} = n] \Big|_{n=N_{k-1}} = N_{k-1} + 1 + (1-p)E[N_k]$$

Thus from the law of total expectations

$$\begin{aligned} E[N_k] &= E[E[N_k \mid N_{k-1}]] \\ &= E[N_{k-1} + 1 + (1-p)E[N_k]] \\ &= E[N_{k-1}] + 1 + (1-p)E[N_k] \\ \Rightarrow E[N_k] &= \frac{1}{p} + \frac{E[N_{k-1}]}{p} \end{aligned}$$

This is a recurrence relation for  $k = 2, 3, 4, \dots$ . To solve, we check for some  $k$  values to gain some intuition

$$\begin{aligned} k=2 &\Rightarrow E[N_2] = \frac{1}{p} + \frac{E[N_1]}{p} = \frac{1}{p} + \frac{1}{p^2} \\ k=3 &\Rightarrow E[N_3] = \frac{1}{p} + \frac{E[N_2]}{p} = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} \\ &\vdots \end{aligned}$$

$$E[N_k] = \sum_{i=1}^k \frac{1}{p^i} \quad k = 1, 2, 3, \dots$$

by induction

This is the finite geometric series for  $r = \frac{1}{p}$ , thus we have

$$E[N_k] = \frac{\frac{1}{p} - \frac{1}{p^{k+1}}}{1 - \frac{1}{p}}$$

## 10 Tutorial 3

### 10.1 Mixed conditional distribution

Suppose  $X$  is  $Erlang(n, \lambda)$  with pdf

$$f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad x > 0$$

Suppose  $Y \mid (X = x)$  is  $POI(x)$  with pmf

$$p_{Y|X}(y \mid x) = \frac{e^{-x} x^y}{y!} \quad y = 0, 1, 2, \dots$$

Find the condition distribution  $X \mid (Y = y)$ .

**Solution.** The marginal distribution of  $Y$  is characterized by its pmf

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} p_{Y|X}(y \mid x) f_X(x) dx \\ &= \int_{-\infty}^{\infty} p_{Y|X}(y \mid x) f_X(x) dx \\ &= \int_0^{\infty} \frac{e^{-x} x^y}{y!} \cdot \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx \\ &= \frac{\lambda^n}{y!(n-1)!} \int_0^{\infty} x^{n+y-1} e^{-(\lambda+1)x} dx \\ &= \frac{\lambda^n (n+y-1)!}{(\lambda+1)^{n+y} y!(n-1)!} \int_0^{\infty} \frac{(\lambda+1)^{n+y} x^{n+y-1} e^{-(\lambda+1)x}}{(n+y-1)!} dx \\ &= \frac{\lambda^n (n+y-1)!}{(\lambda+1)^{n+y} y!(n-1)!} \quad \text{integral of pdf } Erlang(n+y, \lambda+1) \\ &= \binom{n+y-1}{n-1} \left( \frac{\lambda}{\lambda+1} \right)^n \left( \frac{1}{\lambda+1} \right)^y \quad y = 0, 1, 2, \dots \end{aligned}$$

Note that  $p_Y(y)$  is the Negative Binomial distribution shifted to the left  $n$  units. In other words, it counts the number of “failures” before  $n$  successes, where the probability of success is  $\lambda/(\lambda+1)$ .

The distribution of  $X \mid (Y = y)$  is thus

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{p_{Y|X}(y \mid x) f_X(x)}{p_Y(y)} \\ &= \frac{\frac{e^{-x} x^y}{y!} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}}{\frac{(n+y-1)!}{y!(n-1)!} \frac{\lambda^n}{(\lambda+1)^{n+y}}} \\ &= \frac{(\lambda+1)^{n+y} x^{n+y-1} e^{-(\lambda+1)x}}{(n+y-1)!} \quad x > 0 \end{aligned}$$

Note that  $f_{X|Y}(x | y)$  is exactly the Erlang distribution  $Erlang(n + y, \lambda + 1)$ .

## 10.2 Law of total expectations

1. Let  $\{X_i\}_{i=1}^{\infty}$  an iid sequence of  $EXP(\lambda)$  random variables and let  $N \sim GEO(p)$  be independent of each  $X_i$ . Find  $E[\prod_{i=1}^N X_i]$ .
2. Let  $\{X_i\}_{i=0}^{\infty}$  an iid sequence where  $X_i \sim BIN(10, 1/2^i)$ ,  $i = 0, 1, 2, \dots$ . Also let  $N \sim POI(\lambda)$  be independent of each  $X_i$ . Find  $E[X_N]$ .

**Solution.** 1. We want to first find  $E[\prod_{i=1}^N X_i | N = n]$  (conditioning on  $N = n$ )

$$\begin{aligned}
 E[\prod_{i=1}^N X_i | N = n] &= E[\prod_{i=1}^n X_i | N = n] \\
 &= E[\prod_{i=1}^n X_i] && \text{independence of } X'_i \text{ s and } N \\
 &= \prod_{i=1}^n E[X_i] && \text{independence of } X'_i \text{ s} \\
 &= \prod_{i=1}^n \frac{1}{\lambda} \\
 &= \frac{1}{\lambda^n}
 \end{aligned}$$

Thus by the law of total expectations

$$\begin{aligned}
 E[\prod_{i=1}^N X_i] &= E[E[\prod_{i=1}^N X_i | N = n]] \\
 &= \sum_{n=1}^{\infty} \frac{1}{\lambda^n} (1-p)^{n-1} p \\
 &= \frac{p}{\lambda^n} \sum_{n=1}^{\infty} (1-p)^{n-1} \\
 &= \frac{p}{\lambda} \sum_{n=1}^{\infty} \left(\frac{1-p}{\lambda}\right)^{n-1} \\
 &= \frac{p}{\lambda(1 - \frac{1-p}{\lambda})} \sum_{n=1}^{\infty} \left(\frac{1-p}{\lambda}\right)^{n-1} \left(1 - \frac{1-p}{\lambda}\right) \\
 &= \frac{p}{\lambda(1 - \frac{1-p}{\lambda})} && \text{summation of pmf of } GEO(\frac{1-p}{\lambda}) \\
 &= \frac{p}{\lambda - 1 + p}
 \end{aligned}$$

provided that  $\frac{1-p}{\lambda} < 1$  or  $1-p < \lambda$ .

2. Condition on  $N = n$  we have

$$E[X_N | N = n] = E[X_n | N = n] = E[X_n] = 10 \cdot \frac{1}{2^n} = \frac{10}{2^n}$$

From the law of total expectations

$$\begin{aligned} E[X_N] &= E[E[X_N | N = n]] \\ &= \sum_{n=0}^{\infty} \frac{10}{2^n} \frac{e^{-\lambda} \lambda^n}{n!} \\ &= 10e^{-\lambda/2} \sum_{n=0}^{\infty} \frac{e^{-\lambda/2} \left(\frac{\lambda}{2}\right)^n}{n!} \\ &= 10e^{-\lambda/2} \end{aligned} \quad \text{summation of pmf of } POI(\lambda/2)$$

### 10.3 Conditioning on wins and losses

$A, B, C$  are evenly matched tennis players. Initially  $A$  and  $B$  play a set, and the winner plays  $C$ . The winner of each set continues playing the waiting player until one player wins two sets *in a row*. What is the probability that  $A$  is the overall winner?

**Solution.** Key idea: we condition on wins and losses each time until we can find some sort of recurrent relationship, eliminating trivial cases along the way.

Let  $A$  denote the event that  $A$  is the overall winner, and  $W_i, L_i$  denote that  $A$  wins or loses game  $i$ , respectively. Then we have

$$P(A) = P(W_1)P(A | W_1) + P(L_1)P(A | L_1) = \frac{1}{2}P(A | W_1) + \frac{1}{2}P(A | L_1)$$

We can then continue conditioning on subsequent games and their possible outcomes

$$\begin{aligned} P(A | W_1) &= \frac{1}{2}P(A | W_1, W_2) + \frac{1}{2}P(A | W_1, L_2) \\ &= \frac{1}{2}(1) + \frac{1}{2}[P(A | W_1, L_2, C \text{ wins}) \\ &\quad + \frac{1}{2}P(A | W_1, L_2, C \text{ loses})] \\ &= \frac{1}{2} + \frac{1}{4}P(A | W_1, L_2, C \text{ loses}) \quad \begin{array}{l} P(A | W_1, L_2, C \text{ wins}) = 0 \\ \text{since } C \text{ wins twice in a row} \end{array} \\ &= \frac{1}{2} + \frac{1}{4}[\frac{1}{2}P(A | W_1, L_2, C \text{ loses}, W_4) \\ &\quad + \frac{1}{2}P(A | W_1, L_2, C \text{ loses}, L_4)] \\ &= \frac{1}{2} + \frac{1}{8}P(A | W_1, L_2, C \text{ loses}, W_4) \quad \begin{array}{l} P(A | W_1, L_2, C \text{ loses}, L_4) = 0 \\ \text{since } B \text{ wins twice in a row} \end{array} \\ &= \frac{1}{2} + \frac{1}{8}P(A | W_1) \quad \begin{array}{l} \text{since the probability is the same as} \\ A \text{ winning its second game after a win} \end{array} \end{aligned}$$



Solving this recurrence we get  $P(A | W_1) = \frac{8}{14}$ . Similarly

$$\begin{aligned}
 P(A | L_1) &= \frac{1}{2}P(A | L_1, B \text{ wins}) + \frac{1}{2}P(A | L_1, B \text{ loses}) \\
 &= \frac{1}{2}[\frac{1}{2}P(A | L_1, B \text{ loses}, W_3) + \frac{1}{2}P(A | L_1, B \text{ loses}, L_3)] && P(A | L_1, B \text{ wins}) = 0 \\
 &&& \text{since B wins twice in a row} \\
 &= \frac{1}{4}P(A | L_1, B \text{ loses}, W_3) && P(A | L_1, B \text{ loses}, L_3) = 0 \\
 &&& \text{since C wins twice in a row} \\
 &= \frac{1}{4}P(A | W_1)
 \end{aligned}$$

So  $P(A | L_1) = \frac{2}{14}$ .

Plugging this into our initial equation we get  $P(A) = \frac{5}{14}$ .

## 11 February 1, 2018

### 11.1 Example 3.1 solution

A particle moves along the state  $[0, 1, 2]$  according to a DTMC whose TPM is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

where  $P_{ij}$  is the transition probability  $P(X_n = j | X_{n-1} = i)$ .

Let  $X_n$  denote the position of the particle after the  $n$ th move. Suppose the particle is likely to start in any of the three states.

1. Calculate  $P(X_3 = 1 | X_0 = 0)$ .
2. Calculate  $P(X_4 = 2)$ .
3. Calculate  $P(X_6 = 0, X_4 = 2)$ .

**Solution.** 1. We wish to determine  $P_{0,1}^{(3)}$ . To get this, we proceed to calculate  $P^{(3)} = P^3$ . So we have

$$P^3 = (P^2)P = \begin{bmatrix} 0.54 & 0.26 & 0.2 \\ 0.2 & 0.36 & 0.44 \\ 0.6 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix} \quad (11.1)$$

$$= \begin{bmatrix} 0.478 & 0.264 & 0.258 \\ 0.36 & 0.256 & 0.384 \\ 0.57 & 0.18 & 0.25 \end{bmatrix} \quad (11.2)$$

So  $P(X_3 = 1 | X_0 = 0) = P_{0,1}^{(3)} = 0.264$ .

2. We wish to find  $\alpha_{4,2} = P(X_4 = 2)$ . So

$$\begin{aligned}
 \alpha_4 &= (\alpha_{4,0}, \alpha_{4,1}, \alpha_{4,2}) \\
 &= \alpha_0 P^{(4)} \\
 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) P^3 P \\
 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \begin{bmatrix} 0.478 & 0.264 & 0.258 \\ 0.36 & 0.256 & 0.384 \\ 0.57 & 0.18 & 0.25 \end{bmatrix} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix} \\
 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \begin{bmatrix} 0.4636 & 0.254 & 0.2824 \\ 0.444 & 0.2256 & 0.3304 \\ 0.524 & 0.222 & 0.254 \end{bmatrix} \\
 &= (0.4772, 0.233867, 0.288933)
 \end{aligned}$$

So we have  $P(X_4 = 2) = 0.288933$ .

3. We wish to calculate  $P(X_6 = 0, X_4 = 2)$ , which is

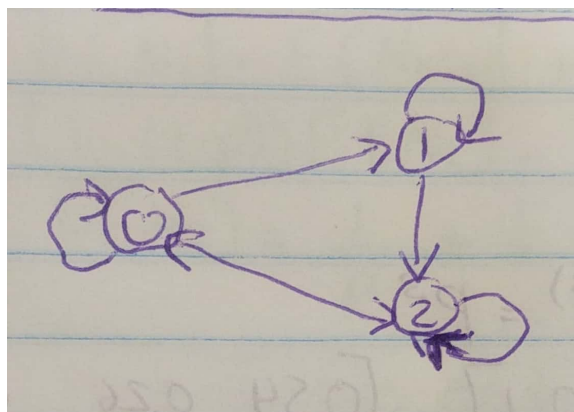
$$\begin{aligned}
 P(X_6 = 0, X_4 = 2) &= P(X_4 = 2)P(X_6 = 0 \mid X_4 = 2) \\
 &= (0.288433)P(X_2 = 0 \mid X_0 = 2) && \text{by stationary assumption} \\
 &= (0.288433)P_{2,0}^{(2)} \\
 &= (0.288433)(0.6) \\
 &= 0.1733598
 \end{aligned}$$

Continued: what are the equivalence classes of the DTMC?

**Solution.** Remember we have the TPM

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

To answer questions of this nature, it is useful to draw a **statement transition diagram**



We see that all states communicate with each other (there is some path from state  $i$  to  $j$  and vice versa). There is only one equivalence class, namely  $\{0, 1, 2\}$ . This is an **irreducible DTMC**.

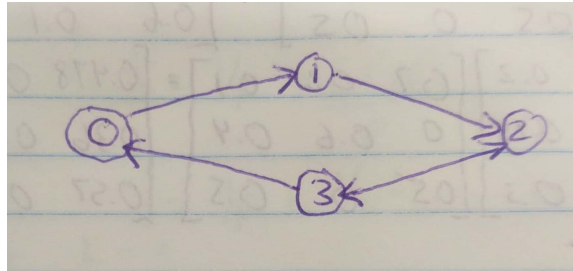
### 11.2 Example 3.2 solution

Consider a DTMC with TPM

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

What are the equivalence classes of this DTMC?

**Solution.** Using a state diagram we have



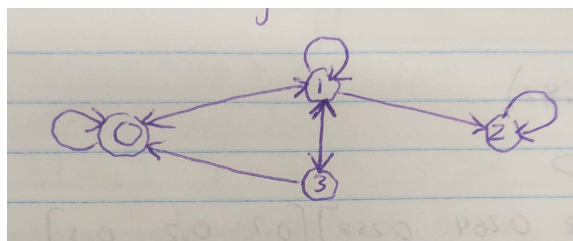
From the diagram, there is only one equivalence class  $\{0, 1, 2, 3\}$ . This DTMC is irreducible.

### 11.3 Example 3.3 solution

Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}$$

What are the equivalence classes of this DTMC?



**Solution.** From the state diagram there are two equivalence classes:  $\{2\}$  and  $\{0, 1, 3\}$ . Thus this DTMC is not irreducible.

## 12 Tutorial 4

### 12.1 Law of total expectations with indicator variables

Suppose the number of people who get on the ground floor of an elevator follows  $POI(\lambda)$ . If there are  $m$  floors above the ground floor and if each person is equally likely to get off at each of the  $m$  floors, independent of where the others get off, calculate the expected number of stops the elevator will make before discharging all passengers.

**Solution.** Let  $X_i$  denote the whether or not someone gets off at floor  $i$ , that is

$$X_i = \begin{cases} 0 & \text{no one gets off at floor } i \\ 1 & \text{someone gets off at floor } i \end{cases}$$

It is easier to think of the case where no one gets off at a floor. That is for  $N$  people

$$P(X_i) = (1 - \frac{1}{m})^N$$

Since  $X_i$  is bernoulli, we have

$$E[X_i | N = n] = 1 - (1 - \frac{1}{m})^n$$

Let  $X = X_1 + \dots + X_m$  denote the total number of stops for the elevator. Thus we have

$$\begin{aligned} E[X] &= E[E[X | N = n]] = \sum_{n=0}^{\infty} E[\sum_{i=1}^m X_i | N = n] p_N(n) \\ &= \sum_{n=0}^{\infty} m(1 - (1 - \frac{1}{m})^n) \cdot \frac{e^{-\lambda} \lambda^n}{n!} \\ &= m \left[ \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} - \sum_{n=0}^{\infty} \frac{e^{-\lambda} ((1 - \frac{1}{m})\lambda)^n}{n!} \right] \\ &= m \left[ 1 - e^{-\frac{\lambda}{m}} \sum_{n=0}^{\infty} \frac{e^{-(1-\frac{1}{m})\lambda} ((1 - \frac{1}{m})\lambda)^n}{n!} \right] \\ &= m(1 - e^{-\frac{\lambda}{m}}) \end{aligned}$$

where the second last and third last equality follows from the summation of Poisson distributions.

## 12.2 Discrete time Markov chain urn example

Three white and three black balls are distributed in two urns in such a way that each contains three balls. We say that the system is in state  $i$ ,  $i = 0, 1, 2, 3$ , if the first urn contains  $i$  white balls. At each step, we draw one ball from each urn and place the ball drawn from the first urn into the second, and conversely with the ball from the second urn. Let  $X_n$  denote the state of the system after the  $n$ th step. Explain why  $\{X_n, n = 0, 1, 2, \dots\}$  is a Markov chain and calculate its transition probability matrix.

**Solution.** Since the determination of the number of white balls in the first urn in the  $n$ th step is only dependent on the same information in the  $(n-1)$ th step, the process  $\{X_n, n = 0, 1, 2, \dots\}$  satisfies the Markov property assumption, and hence is a Markov chain.

Let Urn 1 contain  $i$  white balls and  $3-i$  black balls (and Urn 2 has  $3-i$  white balls and  $i$  black balls). There are 9 possible ways to choose a ball from each so we do a case analysis on what color ball is chosen from each

Urn 1	Urn 2	$X_n$	$X_{n+1}$	
W	B	$i$	$i-1$	with probability $i^2/9$
W	W	$i$	$i$	with probability $i(3-i)/9$
B	B	$i$	$i$	with probability $(3-i)i/9$
B	W	$i$	$i+1$	with probability $(3-i)^2/9$

So for  $i = 0, 1, 2, 3$  we have

$$P_{i,i-1} = \frac{i^2}{9} \quad P_{i,i} = \frac{2i(3-i)}{9} \quad P_{i,i+1} = \frac{(3-i)^2}{9}$$

thus we have the TPM

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

### 12.3 Discrete time Markov chain weather example

On a given day, the weather is either clear (C), overcast (O), or raining (R). If the weather is clear today, then it will be C, O, or R tomorrow with respective probabilities 0.6, 0.3, 0.1. If the weather is overcast today, then it will be C, O, or R tomorrow with probabilities 0.2, 0.5, 0.3. If the weather is raining today, then it will be C, O, or R tomorrow with probabilities 0.4, 0.2, 0.4. Construct the one-step transition probability matrix and use it to find  $\alpha_1$ ,  $P^{(2)}$ , and the probability that it rains on both the first and third days. Assume that the initial probability row vector is given by  $\alpha_0 = (0.5, 0.3, 0.2)$ .

**Solution.** Let state 0, 1, 2 be the states for clear, overcast, and raining. Then we can easily construct the TPM from the given probabilities of transitioning to the other states given an initial state

$$P = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$$

To find  $\alpha_1$ , we simply take  $\alpha_0 P$  or

$$\alpha_1 = (0.5, 0.3, 0.2) \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.2 & 0.4 \end{bmatrix} = (0.44, 0.34, 0.22)$$

To find  $P^{(2)}$ , we take  $PP$  or

$$\begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.2 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.46 & 0.35 & 0.19 \\ 0.34 & 0.37 & 0.29 \\ 0.44 & 0.30 & 0.26 \end{bmatrix}$$

We know the probability it rains on the first day  $\alpha_{1,2}$  or 0.22.

**Remark 12.1.** We can't just take  $\alpha_{3,2}$  as the probability it rains on the third day then apply the multiplication rule since we have some prior that the first day also rains. So we must use conditional probability here.

Thus we have

$$\begin{aligned} P(X_3 = 2, X_1 = 2) &= P(X_3 = 2 \mid X_1 = 2)P(X_1 = 2) \\ &= P(X_2 = 2 \mid X_0 = 2)P(X_1 = 2) && \text{stationary assumption} \\ &= P_{2,2}^{(2)}\alpha_{1,2} = (0.26)(0.22) = 0.0572 \end{aligned}$$

Where  $P(X_2 = 2 \mid X_0 = 2) = P_{2,2}^{(2)}$  since we already know day 0 has state 2.

## 13 February 6, 2018

### 13.1 Example 3.4 solution

Consider the DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

which has equivalence classes  $\{0, 3\}$ ,  $\{1\}$ , and  $\{2\}$ . Determine the period of each state.

**Solution.** Consider the state 0. There are many paths which we can go from state 0 to 0 in  $n$  steps, but one obvious one is simply going from 0 to 0  $n$  times which has probability  $(P_{0,0})^n$ . Therefore

$$P_{0,0}^{(n)} \geq (P_{0,0})^n = (1/3)^n > 0 \quad \forall n \in \mathbb{Z}^+$$

Thus  $d(0) = \gcd\{n \in \mathbb{Z}^+ \mid P_{0,0}^{(n)} > 0\} = \gcd\{1, 2, 3, \dots\} = 1$  (there is a way to get from 0 to 0 in any  $n \in \mathbb{Z}^+$  steps, so we take the gcd of  $\mathbb{Z}^+$  which is 1).

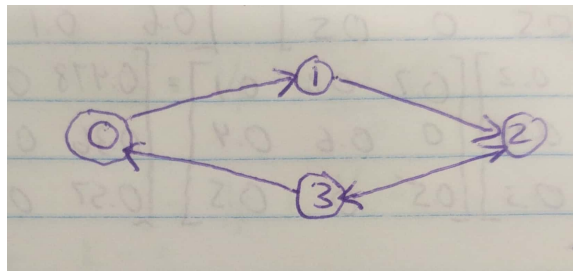
In fact, since every term on the main diagonal of  $P$  is positive, the same argument holds for every state. Thus  $d(1) = d(2) = d(3) = 1$ .

### 13.2 Example 3.2 (continued) solution

Recall for the DTMC with TPM

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

there is 1 equivalence class  $\{0, 1, 2, 3\}$  with state diagram



(note the bi-direction between 2 and 3). Determine the period for each state.

**Solution.** We see that in one of the loops  $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 0$  we have

$$P_{0,0}^{(n)} > 0 \text{ for } n = 4, 8, 12, 16, \dots$$

Also we have (for the cycle  $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 3 \rightarrow 0$ )

$$P_{0,0}^{(n)} > 0 \text{ for } n = 6, 10, 14, 18, \dots$$

Thus  $d(0) = \gcd\{4, 6, 8, 10, 12\} = 2$ .

Following a similar line of logic, we find

$$\begin{aligned} d(1) &= \gcd\{4, 6, 8, 10, 12, \dots\} = 2 \\ d(2) &= \gcd\{2, 4, 6, 8, 10, 12, \dots\} = 2 \\ d(3) &= \gcd\{2, 4, 6, 8, 10, 12, \dots\} = 2 \end{aligned}$$

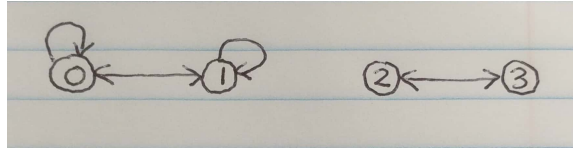
### 13.3 Example 3.5 solution

Consider the DTMC with TPM

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

Find the equivalence class of this DTMC and determine the period of each state.

**Solution.** To determine the equivalence class we draw the state diagram



Clearly the equivalence classes are  $\{0, 1\}$  and  $\{2, 3\}$ .

As in Example 3.4, the main diagonal terms for rows 0 and 1 are positive (i.e.  $P_{0,0}, P_{1,1} > 0$ ) and so  $d(0) = d(1) = 1$ . For states 2 and 3 the DTMC will continually alternate (with probability 1) between each other at every step (i.e.  $2 \rightarrow 3 \rightarrow 2 \rightarrow 3 \rightarrow \dots$ ). Thus it is clear that

$$\begin{aligned} d(2) &= \gcd\{n \in \mathbb{Z}^+ \mid P_{2,2}^{(n)} > 0\} = \gcd\{2, 4, 6, 8, \dots\} = 2 \\ d(3) &= \gcd\{n \in \mathbb{Z}^+ \mid P_{3,3}^{(n)} > 0\} = \gcd\{2, 4, 6, 8, \dots\} = 2 \end{aligned}$$

### 13.4 Theorem 3.1 (equivalent states have equivalent periods)

**Theorem 13.1.** If  $i \leftrightarrow j$  (they communicate), then  $d(i) = d(j)$ .

*Proof.* Assume  $i \neq j$ . Since  $i \leftrightarrow j$ , we know by definition that  $P_{i,j}^{(n)} > 0$  for some  $n \in \mathbb{Z}^+$  and  $P_{j,i}^{(m)} > 0$  for some  $m \in \mathbb{Z}^+$ . Moreover since state  $i$  is accessible from state  $j$  and state  $j$  is accessible from state  $i$ ,  $\exists s \in \mathbb{Z}^+$  such that  $P_{j,j}^{(s)} > 0$ .

Clearly we have that

$$P_{i,i}^{(n+m)} \geq P_{i,j}^{(n)} \cdot P_{j,i}^{(m)} > 0$$

(paths that take  $n$  steps to  $i$  to  $j$  then  $m$  steps to  $j$  to  $i$  is one such possible path from  $i$  to  $i$  in  $n + m$  steps. There could be more  $n + m$  paths hence the  $\geq$ . This also follows from the Chapman-Kolmogorov equations.)

In addition,

$$P_{i,i}^{(n+s+m)} \geq P_{i,j}^{(n)} \cdot P_{j,j}^{(s)} \cdot P_{j,i}^{(m)} > 0$$

So we have paths with  $n + m$  and  $n + s + m$  steps, thus  $d(i)$  divides both  $n + m$  and  $n + s + m$ . Therefore it follows that  $d(i)$  divides their difference, namely  $(n + s + m) - (n + m) = s$ . Since this holds true for *any*  $s$  which satisfies  $P_{j,j}^{(s)} > 0$ , then it must be the case that  $d(i)$  divides  $d(j)$ .

Using the same line of logic, it is straightforward to show  $d(j)$  divides  $d(i)$ .

Putting these two arguments together, we deduce that  $d(i) = d(j)$ . □

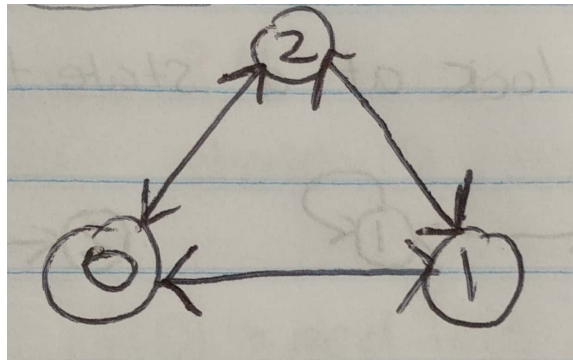
### 13.5 Example 3.6 solution

Consider the DTMC with TPM

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

Find the equivalence classes and determine the period of the states.

**Solution.** The state transition diagram looks like



Clearly the DTMC is irreducible (i.e. there is just one class  $\{0, 1, 2\}$ ).

Note that

$$P_{0,0}^{(1)} = 0$$

$$P_{0,0}^{(2)} \geq P_{0,1}P_{1,0} = \left(\frac{1}{2}\right)^2 = \frac{1}{4} > 0$$

$$P_{0,0}^{(3)} \geq P_{0,1}P_{1,2}P_{2,0} = \left(\frac{1}{2}\right)^3 = \frac{1}{8} > 0$$

Clearly  $d(0) = \gcd\{2, 3, \dots\} = 1$ .

By the above theorem, we have  $d(0) = d(1) = d(2) = 1$ .

## 14 February 8, 2018

### 14.1 Theorem 3.2 (communication and recurrent state $i$ implies recurrent state $j$ )

**Theorem 14.1.** If  $i \leftrightarrow j$  (communicate) and state  $i$  is recurrent, then state  $j$  is recurrent.

*Proof.* Since  $j \leftrightarrow i$ ,  $\exists m, n \in \mathbb{N}$  such that

$$P_{i,j}^{(m)} > 0$$

$$P_{j,i}^{(n)} > 0$$

Also since state  $i$  is recurrent, then we have

$$\sum_{l=1}^{\infty} P_{i,i}^{(l)} = \infty$$



Suppose that  $s \in \mathbb{Z}^+$ . Note that

$$P_{j,j}^{(n+s+m)} \geq P_{j,i}^{(m)} \cdot P_{i,i}^{(s)} \cdot P_{i,j}^{(n)}$$

Now to show that  $j$  is recurrent, we show that the following series diverges

$$\begin{aligned} \sum_{k=1}^{\infty} P_{j,j}^{(k)} &\geq \sum_{k=n+m+1}^{\infty} P_{j,j}^{(k)} \\ &= \sum_{s=1}^{\infty} P_{j,j}^{(n+s+m)} \\ &\geq \sum_{s=1}^{\infty} P_{j,i}^{(m)} \cdot P_{i,i}^{(s)} \cdot P_{i,j}^{(n)} \\ &= P_{j,i}^{(m)} \cdot P_{i,j}^{(n)} \sum_{s=1}^{\infty} P_{i,i}^{(s)} \\ &= \infty \end{aligned}$$

since  $P_{j,i}^{(m)}, P_{i,j}^{(n)} > 0$  and the series diverges by our premise. Therefore  $j$  is recurrent.  $\square$

**Remark 14.1.** A by-product of the above theorem is that if  $i \leftrightarrow j$  and state  $i$  is transient, then state  $j$  is transient.

## 14.2 Theorem 3.3 (communication and recurrent state $i$ implies mutual recurrence among all states)

**Theorem 14.2.** If  $i \leftrightarrow j$  and state  $i$  is recurrent, then

$$f_{i,j} = P(\text{DTMC ever makes a future visit to state } j \mid X_0 = i) = 1$$

*Proof.* Clearly the result is true if  $i = j$ . Therefore suppose that  $i \neq j$ . Since  $i \leftrightarrow j$ , the fact that state  $i$  is recurrent implies that state  $j$  is recurrent by the previous theorem and  $f_{j,j} = 1$ .

To prove  $f_{i,j} = 1$ , suppose that  $f_{i,j} < 1$  and try to get a contradiction.

Since  $i \leftarrow j$ ,  $\exists n \in \mathbb{Z}^+$  such that  $P_{j,i}^{(n)} > 0$  i.e. each time the DTMC visits state  $j$ , there is the possibility of being in state  $i$   $n$  time units later with probability  $P_{j,i}^{(n)} > 0$ .

If we are assuming that  $f_{i,j} < 1$ , then this implies that the probability of returning to state  $j$  after visiting  $i$  in the future is not guaranteed (as  $1 - f_{i,j} > 0$ ). Therefore

$$\begin{aligned} 1 - f_{j,j} &= P(\text{DTMC never makes a future visit to state } j \mid X_0 = j) \\ &= P_{j,i}^{(n)} \cdot (1 - f_{i,j}) \\ &> 0 \qquad \text{both } > 0 \end{aligned}$$

This implies that  $1 - f_{j,j} > 0$  or  $f_{j,j} < 1$ , which is a contradiction. Therefore  $f_{i,j} = 1$ .  $\square$

## 14.3 Theorem 3.4 (finite-state DTMCs have at least one recurrent state)

**Theorem 14.3.** A finite-state DTMC has at least one recurrent state.

*Proof.* Equivalently, we want to show that not all states can be transient.

Suppose that  $\{0, 1, 2, \dots, N\}$  represents the states of the DTMC where  $N < \infty$  (finite).

To prove that not all states can be transient, we suppose they are all transient and try to get a contradiction.

Now for each  $i = 0, 1, \dots, N$ , if state  $i$  is assumed to be transient, we know that after a *finite* amount of time (denoted by  $T_i$ ), state  $i$  will never be visited again. As a result, after a finite amount of time  $T = \{T_0, T_1, \dots, T_n\}$  has gone by *none of the states will be visited again*.

However, the DTMC **must be in some state** after time  $T$  but we have exhausted all states from the DTMC to be in. This is a contradiction thus not all states can be transient in a finite state DTMC.  $\square$

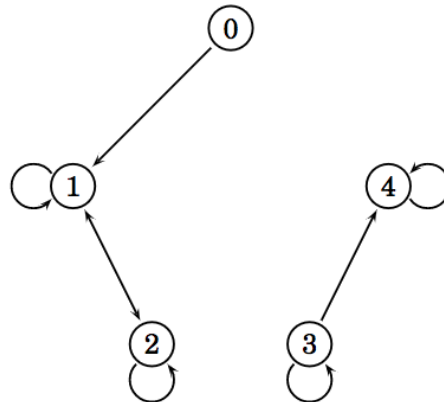
## 15 Tutorial 5

### 15.1 Determining diagram, equivalence classes, period, and transience/recurrence of DTMC

For the following Markov chain, draw its state transition diagram, determine its equivalence classes, and the periods of states within each class, and determine whether they are transient or recurrent

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

**Solution.** The state diagram for the DTMC is



From the diagram we see the only closed loop path containing more than one state is  $1 \rightarrow 2 \rightarrow 1$ , thus we have the equivalence classes  $\{0\}$ ,  $\{1, 2\}$ ,  $\{3\}$ ,  $\{4\}$ .

Clearly  $0 \rightarrow 1$  with a probability of 1 so 0 is transient and its period is  $\infty$ .

Notice for states 1, 2, 3, 4 the main diagonal entries  $P_{i,i} > 0$  so their periods are 1.

For transience/recurrence, note that

$$\begin{aligned} \sum_{n=1}^{\infty} P_{0,0}^{(n)} &= \sum_{n=1}^{\infty} 0 &< \infty \Rightarrow \{0\} \text{ is transient} \\ \sum_{n=1}^{\infty} P_{3,3}^{(n)} &= \sum_{n=1}^{\infty} \left(\frac{3}{4}\right)^n = \frac{\frac{3}{4}}{1 - \frac{3}{4}} = 3 &< \infty \Rightarrow \{3\} \text{ is transient} \\ \sum_{n=1}^{\infty} P_{4,4}^{(n)} &= \sum_{n=1}^{\infty} 1 &= \infty \Rightarrow \{4\} \text{ is recurrent} \end{aligned}$$

For class  $\{1, 2\}$ , consider the values of  $f_{1,1}^{(n)}$

$$\begin{aligned} f_{1,1}^{(1)} &= P(X_1 = 1 \mid X_0 = 1) &&= P_{1,1} = \frac{2}{3} \\ f_{1,1}^{(2)} &= P(X_2 = 1, X_1 \neq 1 \mid X_0 = 1) = P(X_2 = 1, X_1 = 2 \mid X_0 = 1) &&= P_{1,2}P_{2,1} = \frac{1}{3} \cdot \frac{1}{5} \\ f_{1,1}^{(3)} &= P(X_3 = 1, X_2 \neq 1, X_1 \neq 1 \mid X_0 = 1) \\ &= P(X_3 = 1, X_2 = 2, X_1 = 2 \mid X_0 = 1) &&= P_{1,2}P_{2,2}P_{2,1} = \frac{1}{3} \cdot \frac{4}{5} \cdot \frac{1}{5} \\ f_{1,1}^{(4)} &= P(X_4 = 1, X_3 = 2, X_2 = 2, X_1 = 2 \mid X_0 = 1) &&= P_{1,2}(P_{2,2})^2P_{2,1} = \frac{1}{3} \cdot \left(\frac{4}{5}\right)^2 \cdot \frac{1}{5} \\ f_{1,1}^{(n)} &= P(X_n = 1, X_{n-1} = 2, \dots, X_2 = 2, X_1 = 2 \mid X_0 = 1) &&= P_{1,2}(P_{2,2})^{n-2}P_{2,1} = \frac{1}{3} \cdot \left(\frac{4}{5}\right)^{n-2} \cdot \frac{1}{5} \end{aligned}$$

So we have

$$\begin{aligned} f_{1,1} &= \sum_{n=1}^{\infty} f_{1,1}^{(n)} \\ &= \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{5} \sum_{n=2}^{\infty} \left(\frac{4}{5}\right)^{n-2} \\ &= \frac{2}{3} + \frac{1}{15} \sum_{n=0}^{\infty} \left(\frac{4}{5}\right)^n \\ &= \frac{2}{3} + \frac{1}{15} \frac{1}{1 - \frac{4}{5}} \\ &= \frac{2}{3} + \frac{1}{3} \\ &= 1 \end{aligned}$$

which by definition implies class  $\{1, 2\}$  is recurrent.

## 15.2 Discrete time Markov chain consecutive successes example

Let  $k \in \mathbb{Z}^+$  and consider an experiment in which independent trials, each having success probability  $p \in (0, 1)$ , are conducted indefinitely. We say that the system is in state  $i$  at time  $n$  if after the  $n$ th trial we have observed  $i$  consecutive successes occurring at times  $n - i + 1, n - i + 2, \dots, n - 1, n$  where  $i = 1, 2, \dots, k - 1$ . In addition, we

say that the system is in state  $k$  at time  $n$  if we have ever observed at least  $k$  consecutive successes occur at least once by time  $n$ . Finally, if we have not observed  $k$  consecutive successes by time  $n$ , and the  $n$ th trial was a failure, then we say that the system is in state 0. Letting  $X_n$  denote the state of the system after the  $n$ th trial, calculate its transition probability matrix.

**Solution.** Note at any state  $k = 1, 2, \dots, k-1$  we always have a probability of  $1-p$  of failing and going back to state 0.

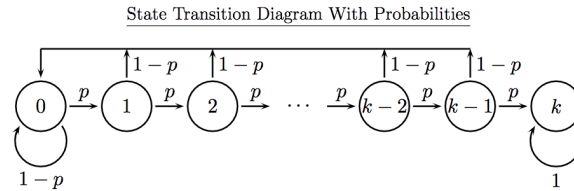
To transition from state  $k$  to state  $k+1$  successes, we must get a success on the next trial with probability  $p$ , unless it's already in state  $k$ .

Finally, once we reach state  $k$  (we have seen  $k$  consecutive successes), any successes or failures will not change the system's state at  $k$ , so  $P_{k,k} = 1$ .

Thus we have

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & \dots & k-2 & k-1 & k \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ k-2 \\ k-1 \\ k \end{matrix} & \begin{bmatrix} 1-p & p & 0 & \dots & 0 & 0 & 0 \\ 1-p & 0 & p & \dots & 0 & 0 & 0 \\ 1-p & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & & & & & \\ 1-p & 0 & 0 & \dots & 0 & p & 0 \\ 1-p & 0 & 0 & \dots & 0 & 0 & p \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

The corresponding state diagram of this DTMC is



### 15.3 Limiting behaviour of discrete Markov chains

For a DTMC  $\{X_n, n \in \mathbb{N}\}$  suppose we are given initial probability row vector  $\alpha_0 = (1, 0)$  and TPM

$$P = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix} \end{matrix}$$

where  $0 < p < 1$ . Find  $P^{(n)}$  and  $\alpha_n$ .

**Solution.** We know the  $n$ -step TPM is simply the one-step TPM multiplied by itself  $n$  times or

$$P^{(n)} = \prod_{i=1}^n \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix}$$

Note that

$$\begin{aligned}
 P_{0,0}^{(n)} &= \sum_{k=0}^1 P_{0,k}^{(n-1)} P_{k,0} = P_{0,0}^{(n-1)} P_{0,0} + P_{0,1}^{(n-1)} P_{1,0} = P_{0,0}^{(n-1)} (1-p) & P_{1,0} &= 0 \\
 &= P_{0,0}^{(n-2)} (1-p)^2 = \dots = P_{0,0}^{(1)} (1-p)^{n-1} = (1-p)^n \\
 P_{1,0}^{(n)} &= \sum_{k=0}^1 P_{1,k}^{(n-1)} P_{k,0} = P_{1,0}^{(n-1)} P_{0,0} + P_{1,1}^{(n-1)} P_{1,0} = P_{1,0}^{(n-1)} (1-p) & P_{1,0} &= 0 \\
 &= P_{1,0}^{(n-2)} (1-p)^2 = \dots = P_{1,0}^{(1)} (1-p)^{n-1} = 0(1-p)^{n-1} = 0 \\
 P_{1,1}^{(n)} &= \sum_{k=0}^1 P_{1,k}^{(n-1)} P_{k,1} = P_{1,0}^{(n-1)} P_{0,1} + P_{1,1}^{(n-1)} P_{1,1} = P_{1,1}^{(n-1)} 1 = \dots = P_{1,1}^{(1)} = 1
 \end{aligned}$$

for all  $n \in \mathbb{N}$ . Using the fact that each row sum of a TPM must equal 1 we have

$$P_{0,1}^{(n)} = 1 - P_{0,0}^{(n)} = 1 - (1-p)^n \quad n \in \mathbb{N}$$

So we have

$$P^{(n)} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} (1-p)^n & 1 - (1-p)^n \\ 0 & 1 \end{bmatrix} \end{matrix} \quad n \in \mathbb{N}$$

and by definition

$$\alpha_n = \alpha_0 P^{(n)} = (1, 0) \begin{bmatrix} (1-p)^n & 1 - (1-p)^n \\ 0 & 1 \end{bmatrix} = ((1-p)^n, 1 - (1-p)^n)$$

## 16 February 13, 2018

### 16.1 Theorem 3.5 (recurrent $i$ and not communicate with $j$ implies $P_{i,j} = 0$ )

**Theorem 16.1.** If state  $i$  is recurrent and state  $i$  does not communicate with state  $j$ , then  $P_{i,j} = 0$ .

*Proof.* Let us assume that  $P_{i,j} > 0$  (one-step transition probability) and try to get a contradiction.

Then  $P_{j,i}^{(n)} = 0 \forall n \in \mathbb{Z}^+$  otherwise states  $i, j$  communicate.

However, the DTMC, starting in state  $i$ , would have a positive probability of at least  $P_{i,j}$  of never returning to state  $i$ . This contradicts the recurrence of state  $i$ , therefore we must have  $P_{i,j} = 0$ .  $\square$

### 16.2 Example 3.3 (continued) solution

Recall our earlier DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}$$

Determine whether each state is transient or recurrent.

**Solution.** Note that  $\sum_{n=1}^{\infty} P_{2,2}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty$  which implies state 2 is recurrent.

Looking at the possible transitions that can take place among states 0, 1, and 3 we strongly suspect state 1 to be transient (since there is a positive probability of never returning to state 1 if a transition to state 2 occurs). To show this formally assume state 1 is recurrent and try to get a contradiction.

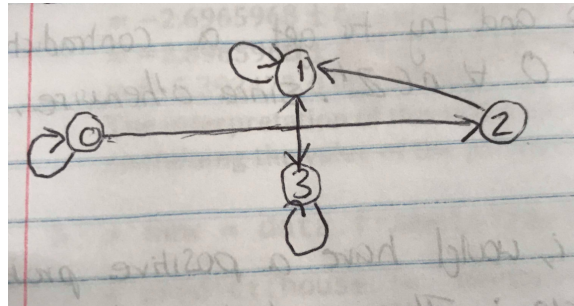
So if state 1 is recurrent, note that state 1 does not communicate with state 2. By theorem 3.5, we have that  $P_{1,2} = 0$ , but in fact  $P_{1,2} = \frac{1}{8} > 0$ . This is a contradiction, so state 1 must be transient and so  $\{0, 1, 3\}$  is transient.

### 16.3 Example 3.7 solution

Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 1 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} \end{bmatrix}$$

Determine whether each state is transient or recurrent.



**Solution.** The equivalence classes are  $\{0\}$ ,  $\{2\}$ ,  $\{1, 3\}$ .

Note that (since 0 is in its own equivalence classes)

$$\sum_{n=1}^{\infty} P_{0,0}^{(n)} = \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^n = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3} < \infty$$

so state 0 is transient.

Furthermore (since state 2 is in its own equivalence class)

$$\sum_{n=1}^{\infty} P_{2,2}^{(n)} = \sum_{n=1}^{\infty} 0 = 0 < \infty$$

so state 2 is also transient.

On the other hand, concerning the class  $\{1, 3\}$  we observe

$$f_{1,1}^{(1)} = P_{1,1} = \frac{1}{3}$$

AND

$$f_{1,1}^{(n)} = \sum_{n=2}^{\infty} (2/3)(3/5)^{n-2}(2/5) \quad n \geq 2$$

Now

$$\begin{aligned}
 f_{1,1} &= \sum_{n=1}^{\infty} f_{1,1}^{(n)} = \frac{1}{3} \sum_{n=2}^{\infty} \left(\frac{2}{3}\right) \left(\frac{3}{5}\right)^{n-2} \left(\frac{2}{5}\right) \\
 &= \frac{1}{3} + \left(\frac{2}{3}\right) \left(\frac{2}{5}\right) \sum_{n=2}^{\infty} \left(\frac{3}{5}\right)^{n-2} \\
 &= \frac{1}{3} + \left(\frac{2}{3}\right) \left(\frac{2}{5}\right) \cdot \frac{1}{1 - \frac{3}{5}} \\
 &= 1
 \end{aligned}$$

By definition since  $f_{1,1} = 1$  state 1 is recurrent and so class  $\{1, 3\}$  is recurrent.

(Or we could have concluded from Theorem 3.4 that because  $\{0\}$  and  $\{2\}$  are both transient,  $\{1, 3\}$  must be recurrent).

### 16.4 Example 3.8 solution

Consider a DTMC  $\{X_n, n \in \mathbb{N}\}$  whose state space is all integers i.e.  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ . Suppose the TPM for the DTMC satisfies

$$P_{i,i-1} = 1 - p \text{ and } P_{i,i+1} = p \quad \forall i \in \mathbb{Z} \text{ and } 0 < p < 1$$

In other words from any state, we can either jump up or down one state. This is often referred to as the *Random Walk*. Characterize the behaviour of this DTMC in terms of its equivalence classes, periodicity, and transience/recurrence.

**Solution.** First of all since  $0 < p < 1$  all states clearly communicate with each other. This implies that  $\{X_n, n \in \mathbb{N}\}$  is an irreducible DTMC.

Hence we can determine its periodicity (and likewise its transience/recurrence) by analyzing any state we wish. Let us select state 0. Starting from state 0, note that state 0 cannot possibly be visited in an odd number of transitions since we are guaranteed to have the number of up(down) jumps exceed the number of down(up) jumps (??? why). Thus  $P_{0,0}^{(1)} = P_{0,0}^{(3)} = P_{0,0}^{(5)} = \dots = 0$  or equivalent

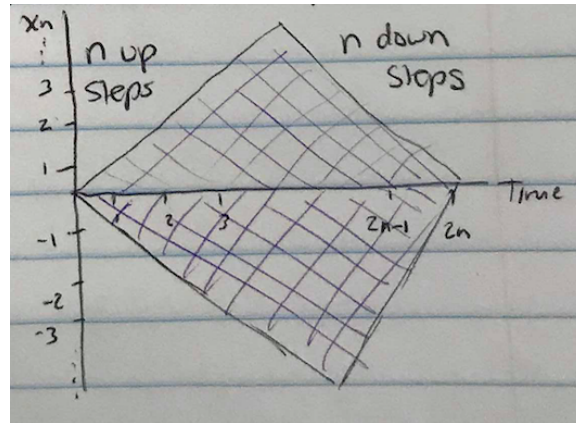
$$P_{0,0}^{(2n-1)} = 0 \quad \forall n \in \mathbb{Z}^+$$

However since it is clearly possible to return to state 0 in an even number of transitions, it immediately follows that  $P_{0,0}^{(2n)} > 0 \quad \forall n \in \mathbb{Z}^+$ .

Hence  $d(0) = \gcd\{n \in \mathbb{Z}^+ \mid P_{0,0}^{(n)} > 0\} = \gcd\{2, 4, 6, \dots\} = 2$ .

Finally to determine whether state 0 is transient or recurrent, let us consider

$$\sum_{n=1}^{\infty} P_{0,0}^{(n)} = \sum_{n=1}^{\infty} P_{0,0}^{(2n)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n$$



That is we need to take  $n$  up steps with probability  $p$  and  $n$  down steps with probability  $1 - p$ . Note that  $\binom{2n}{n}$  accounts for the number of ways to arrange this.

Recall: ratio test for series.

Suppose that  $\sum_{n=1}^{\infty} a_n$  is a series of positive terms and

$$L = \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n}$$

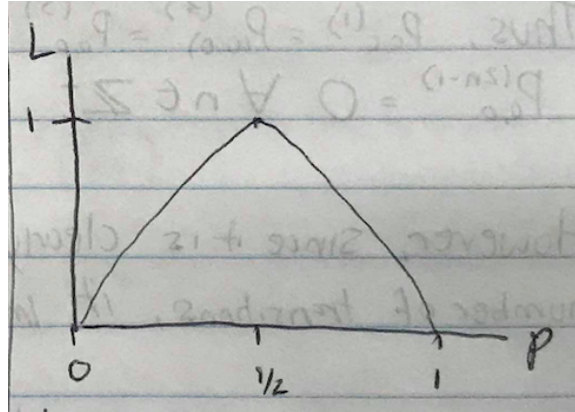
1. If  $L < 1$  then the series converges.
2. If  $L > 1$  then the series diverges.
3. If  $L = 1$  then the test is inconclusive.

In our case we obtain

$$\begin{aligned}
 L &= \lim_{n \rightarrow \infty} \frac{\frac{(2(n+1))!}{(n+1)!(n+1)!} \cdot p^{n+1}(1-p)^{n+1}}{\frac{2n!}{n!n!} \cdot p^n(1-p)^n} \\
 &= \lim_{n \rightarrow \infty} \frac{(2n+2)!}{(n+1)!(n+1)!} \cdot \frac{n!n!}{2n!} \cdot p(1-p) \\
 &= \lim_{n \rightarrow \infty} \frac{(2n+2)(2n+1)}{(n+1)(n+1)} \cdot p(1-p) \\
 &= \lim_{n \rightarrow \infty} \frac{4n^2 + 6n + 2}{n^2 + 2n + 1} \cdot p(1-p) \\
 &= 4p(1-p)
 \end{aligned}$$

A plot of  $L = 4p(1-p)$  reveals the following shape





Note that if  $p \neq \frac{1}{2}$  then  $L < 1$  which means state 0 is transient (and so is DTMC).

However, if  $p = \frac{1}{2}$  then  $L = 1$  and the ratio test is inconclusive. We need another approach to handle  $p = \frac{1}{2}$  (this approach can handle both  $p = \frac{1}{2}$  and  $p \neq \frac{1}{2}$ ).

Recall that  $f_{i,j} = P(\text{DTMC ever makes a future visit to state } j \mid X_0 = i)$ .

Letting  $q = 1 - p$  and conditioning on the state of the DTMC at time 1 we have

$$\begin{aligned}
 f_{0,0} &= P(\text{DTMC ever makes a future visit to state 0} \mid X_0 = 0) \\
 &= P(X_1 = -1 \mid X_0 = 0) \cdot P(\text{DTMC ever makes a future visit to state 0} \mid X_1 = -1, X_0 = 0) \\
 &\quad + P(X_1 = 1 \mid X_0 = 0) \cdot P(\text{DTMC ever makes a future visit to state 0} \mid X_1 = 1, X_0 = 0) \\
 &\quad \text{by Markov property (only most recent state matters) and stationarity (we can treat time 1 as time 0)} \\
 &= q \cdot f_{-1,0} + p \cdot f_{1,0}
 \end{aligned}$$

Using precisely the same logic it follows that

$$f_{1,0} = q \cdot 1 + p \cdot f_{2,0} \tag{16.1}$$

Consider  $f_{2,0}$ : in order to visit state 0 from state 2, we first have to visit state 1 again, which happens with probability  $f_{2,1}$ . This is **equivalent to**  $f_{1,0}$ , so  $f_{2,1} = f_{1,0}$  (probability of ever making a visit to the state one step down).

Given you make a visit back to state 1, then we must make a visit back to state 0, which happens to with probability  $f_{1,0}$ .

Putting this together we have

$$f_{2,0} = f_{2,1} \cdot f_{1,0} = f_{1,0}^2$$

Thus from equation 16.1 we have

$$f_{1,0} = q + p \cdot f_{1,0}^2$$

which is a quadratic in the form

$$pf_{1,0}^2 - f_{1,0} + q = 0$$

Applying the quadratic formula we get

$$\begin{aligned}
 f_{1,0} &= \frac{1 \pm \sqrt{1 - 4pq}}{2p} \\
 &= \frac{1 \pm \sqrt{(p+q)^2 - 4pq}}{2p} \\
 &= \frac{1 \pm \sqrt{p^2 - 2pq + q^2}}{2p} \\
 &= \frac{1 \pm |p - q|}{2p}
 \end{aligned}$$

There can only be **one unique solution** for  $f_{1,0}$  which means that one of

$$\begin{aligned}
 r_1 &= \frac{1 + |p - q|}{2p}; \text{ or} \\
 r_2 &= \frac{1 - |p - q|}{2p}
 \end{aligned}$$

must be inadmissible (intuitively: we see that if  $p < 0.5$  then  $r_1$  will be bigger than 1 which is not possible for probabilities so  $r_1$  does not work for all  $p$ ).

To determine which it is, suppose that  $q > p$ . Then  $|p - q| = -(p - q)$  and the 2 roots become

$$\begin{aligned}
 r_1 &= \frac{1 - (p - q)}{2p} = \frac{q + q}{2p} = \frac{q}{p} > 1 \\
 r_2 &= \frac{1 + (p - q)}{2p} = \frac{p + p}{2p} = 1
 \end{aligned}$$

Thus we must have

$$f_{1,0} = \frac{1 - |p - q|}{2p}$$

Note: using the exact same approach

$$f_{-1,0} = \frac{1 - |p - q|}{2q}$$

With knowledge of  $f_{1,0}$  and  $f_{-1,0}$  we find that

$$\begin{aligned}
 f_{0,0} &= qf_{-1,0} + pf_{1,0} \\
 &= q\left(\frac{1 - |p - q|}{2q}\right) + p\left(\frac{1 - |p - q|}{2p}\right) \\
 &= 1 - |p - q|
 \end{aligned}$$

Suppose  $p > q$  i.e.  $1 - q > q$  or  $2q < 1$ :

$$\begin{aligned}
 f_{0,0} &= 1 - (p - q) \\
 &= 1 - p + q \\
 &= 2q < 1
 \end{aligned}$$

since  $f_{0,0} < 1$  we have that state 0 is transient.

Similarly for  $p < q$  we get  $f_{0,0} = 2p < 1$  so state 0 is transient.

When  $p = q$ , then we have  $f_{0,0} = 1$  so state 0 is recurrent under  $p = q = \frac{1}{2}$ .

## 17 February 15, 2018

### 17.1 Example 3.9 solution

Consider the DTMC with TPM

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

Determine if  $\lim_{n \rightarrow \infty} P^{(n)}$  exists (i.e. the limiting behavior of the DTMC).

**Solution.** There are 2 equivalence classes, namely  $\{0, 2\}$  and  $\{1\}$ .

Each class is recurrent with periods 2 and 1, respectively.

For  $n \in \mathbb{Z}^+$ , note that

$$P^{(2n)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which follows by taking  $P^{(2)}$  and multiplying by itself arbitrarily many times. Also

$$P^{(2n-1)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

which follows by taking  $P^{(1)}$  and multiply it arbitrarily many times by  $P^{(2)}$  (identity).

As such,  $\lim_{n \rightarrow \infty} P^{(n)}$  does not exist since  $P^{(n)}$  alternates between those two matrices..

However, while  $\lim_{n \rightarrow \infty} P_{0,0}^{(n)}$  and  $\lim_{n \rightarrow \infty} P_{0,2}^{(n)}$  (and many other) do not exist, note that some limits do exist such as

$$\lim_{n \rightarrow \infty} P_{0,1}^{(n)} = 0$$

and

$$\lim_{n \rightarrow \infty} P_{1,1}^{(n)} = 1$$

### 17.2 Example 3.10 solution

Consider the DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Determine if  $\lim_{n \rightarrow \infty} P^{(n)}$  exists.

**Solution.** There is only one equivalence class and so the DTMC is irreducible. Also it is straightforward to show that the DTMC is aperiodic and recurrent.

It can be shown that

$$\lim_{n \rightarrow \infty} P^{(n)} = \begin{bmatrix} \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \end{bmatrix}$$

(we can find this by using some software and repeatedly applying matrix exponentiation; we will later demonstrate a way to solve this for certain DTMCs (see [Example 3.10 \(continued\) solution](#)).

Note that this matrix has **identical rows**. This implies that  $P_{i,j}^{(n)}$  converges to a value as  $n \rightarrow \infty$  which is the same for all initial states  $i$ . In other words, there is a limiting probability the DTMC will be in state  $j$  as  $n \rightarrow \infty$  and this probability is independent of the initial states.

## 18 February 27, 2018

### 18.1 Example 3.11 solution

Consider the DTMC with TPM

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ 0 & 0 & 1 \end{bmatrix}$$

Examine  $\lim_{n \rightarrow \infty} P^{(n)}$  and explain why the limiting probability of being in a state depends on the initial state of this DTMC.

**Solution.** We have three equivalence classes  $\{0\}, \{1\}, \{2\}$ . Since  $P_{i,i} > 0 \forall i = 0, 1, 2$  each state is aperiodic.

$$\sum_{n=1}^{\infty} P_{0,0}^{(n)} = \sum_{n=1}^{\infty} P_{2,2}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty$$

which implies state 0 and 2 are recurrent.

We can also show that

$$\sum_{n=1}^{\infty} P_{1,1}^{(n)} = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2}} = 1 < \infty$$

which means state 1 is transient.

In fact, states 0 and 2 are examples of **absorbing states** (i.e.  $P_{0,0} = P_{2,2} = 1$ ).

If one begins in state 1, note that with probability  $\frac{1}{2}$  one can end up in one of the two recurrent classes in the next transition i.e. the DTMC will eventually be absorbed into either state 0 or 2 so state 1 is *transient*.

It can be shown that

$$\lim_{n \rightarrow \infty} P^{(n)} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

Intuitively this makes sense: eventually state 1 will go into 0 or 2. The relative probability of going into state 0 and going into state 2 is equivalent to our limiting probability entries.

**Remark 18.1.** Unlike the previous example, it *does matter* from which state one begins in this DTMC. Furthermore the second column of the above contains all zeros.

### 18.2 Theorem 3.6 solution

**Theorem 18.1.** For any state  $i$  and transient state  $j$  of a DTMC,  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$ .

*Proof.* Recall that we defined the “first visit probability in  $n$  steps” as

$$f_{i,j}^{(n)} = P(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j \mid X_0 = i)$$

Furthermore we defined

$$f_{i,j} = P(\text{DTMC ever makes a future visit to state } j \mid X_0 = i) = \sum_{n=1}^{\infty} f_{i,j}^{(n)}$$

We can then use the fact that

$$P_{i,j}^{(n)} = \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)}$$

So we have

$$\begin{aligned} \sum_{n=1}^{\infty} P_{i,j}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\ &= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{n=k}^{\infty} P_{j,j}^{(n-k)} \\ &= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{l=0}^{\infty} P_{j,j}^{(l)} \quad l = n - k \\ &= \left( \sum_{k=1}^{\infty} f_{i,j}^{(k)} \right) \left( 1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \right) \quad P_{i,i}^{(0)} = 1 \quad \forall i \\ &= f_{i,j} \left( 1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \right) \\ &\leq 1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \quad f_{i,j} \leq 1 \text{ since it's a probability} \\ &< \infty \text{ since } j \text{ is transient so the summation is finite} \end{aligned}$$

Recall the **nth term test**: If  $\lim_{n \rightarrow \infty} a_n \neq 0$  or if the limit does not exist, then  $\sum_{n=1}^{\infty} a_n$  diverges.

We can show by contradiction that if the series converges then the limit is 0.

By the **nth term test** for infinite series

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$$

□

### 18.3 Example 3.10 (continued) solution

Recall we were given the DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Find the limiting probability of this DTMC.

**Solution.** Clearly DTMC is irreducible, aperiodic and positive recurrent. The BLT conditions are met and  $\pi = (\pi_0, \pi_1, \pi_2)$  exists and satisfies

$$\begin{aligned} \pi &= \pi P \\ \pi e' &= 1 \end{aligned} \quad e' = (1, 1, \dots, 1)^T$$

Thus we have

$$(\pi_0, \pi_1, \pi_2) = (\pi_0, \pi_1, \pi_2) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Expanding the matrix multiplication we get

$$\begin{aligned} \pi_0 &= \frac{1}{2}\pi_0 + \frac{1}{2}\pi_1 \\ \pi_1 &= \frac{1}{2}\pi_0 + \frac{1}{4}\pi_1 + \frac{1}{3}\pi_2 \\ \pi_2 &= \frac{1}{4}\pi_1 + \frac{2}{3}\pi_2 \\ 1 &= \pi_0 + \pi_1 + \pi_2 \end{aligned}$$

(we can disregard the  $\pi_1 = \dots$  equation since it requires solving for the most terms and we have 4 equations for 3 unknowns). Solving the system we get

$$\begin{aligned} \pi_0 &= \pi_1 = \frac{4}{11} \\ \pi_2 &= \frac{3}{11} \end{aligned}$$

Recall we had

$$\lim_{n \rightarrow \infty} P^{(n)} = \begin{bmatrix} \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \end{bmatrix} = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \end{bmatrix}$$

We observe that each row of  $P^{(n)}$  converges to  $\pi$  as  $n \rightarrow \infty$ .

## 18.4 Example 3.12 solution (Gambler's Ruin)

Consider a gambler who, at each play of a game, has probability  $p \in (0, 1)$  of winning one unit and probability  $q = 1 - p$  of losing one unit. Assume that successive plays of the game are independent. If the gambler initially begins with  $i$  units, what is the probability that the gambler's fortune will reach  $N$  units ( $N < \infty$ ) before reaching 0 units? This problem is often referred to as the Gambler's Ruin Problem, with state 0 representing bankruptcy and state  $N$  representing the jackpot.

**Solution.** For  $n \in \mathbb{N}$ , define  $X_n$  as the gambler's fortune after the  $n$ th play of the game. Therefore  $\{X_n, n \in \mathbb{N}\}$  is a DTMC with TPM

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \vdots & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

Where the TPM has entries  $P_{ij}$  where  $i, j \in [0, N]$  (for each  $i \in [1, N - 1]$ , we have  $P_{i(i-1)} = q$  and  $P_{i(i+1)} = p$ .  $P_{00} = P_{NN} = 1$  and all other  $P_{mn} = 0$ ).

**Note:** States 0 and  $N$  are treated as **absorbing states**, therefore they are both recurrent.

States  $\{1, 2, \dots, N - 1\}$  belong to the same equivalence class and it is straightforward to see it is a transient class.

**Goal:** Determine  $P(i)$  for  $i = 0, 1, \dots, N$  which represents the probability that starting with  $i$  units, the gambler's fortune will eventually reach  $N$  units.

It follows that (for the limiting TPM)

$$\lim_{n \rightarrow \infty} P^{(n)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 - P(1) & 0 & 0 & 0 & \dots & 0 & 0 & P(1) \\ 1 - P(2) & 0 & 0 & 0 & \dots & 0 & 0 & P(2) \\ \vdots & & & & & & & \\ 1 - P(N-1) & 0 & 0 & 0 & \dots & 0 & 0 & P(N-1) \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

where the  $(0,0)$  entry is 1 (0 state is recurrent),  $(0,N)$  is 0 (can never reach state  $N$  from 0) and similarly for  $(N,N)$  and  $(N,0)$ .

Entries  $(i,N) = P(i)$  by definition where  $P(0) = 0$  and  $P(N) = 1$ . Since the sum of the entries for a given row must be 0, we get for entry  $(i,0)$  as  $1 - P(i)$ .

Conditioning on the outcome of the very first game  $i = 1, 2, \dots, N-1$ , we have

$$P(i) = p \cdot P(i+1) + q \cdot P(i-1)$$

(if we go up by 1 with probability  $p$ , then the probability of winning afterwards is  $P(i+1)$  i.e.  $P(i+1) = P(i \mid \text{win})$  and  $p = P(\text{win})$ ). Similarly for  $q$  and  $P(i-1)$ ).

Note that  $p + q = 1$ , so we have

$$\begin{aligned} P(i) &= pP(i) + qP(i) \\ \Rightarrow p(P(i+1) - P(i)) &= q(P(i) - P(i-1)) \\ P(i+1) - P(i) &= \frac{q}{p}(P(i) - P(i-1)) \end{aligned}$$

To determine if an explicit solution exists, consider several values of  $i$

$$\begin{aligned} i = 1 &\Rightarrow P(2) - P(1) = \frac{q}{p}(P(1) - P(0)) = \frac{q}{p}P(1) \\ i = 2 &\Rightarrow P(3) - P(2) = \frac{q}{p}(P(2) - P(1)) = \left(\frac{q}{p}\right)^2 P(1) \\ i = 3 &\Rightarrow P(4) - P(3) = \frac{q}{p}(P(3) - P(2)) = \left(\frac{q}{p}\right)^3 P(1) \\ &\vdots \\ i = k &\Rightarrow P(k+1) - P(k) = \frac{q}{p}(P(k) - P(k-1)) = \left(\frac{q}{p}\right)^k P(1) \end{aligned}$$

Note: the above  $k$  equations are linear in terms of the unknowns  $P(1), P(2), \dots, P(k+1)$ . Summing these  $k$

equations yields

$$\begin{aligned}
 P(k+1) - P(1) &= \sum_{i=1}^k \left(\frac{q}{p}\right)^i P(1) \\
 \Rightarrow P(k+1) &= \sum_{i=0}^k \left(\frac{q}{p}\right)^i P(1) \quad k = 0, 1, \dots, N-1 \\
 \Rightarrow P(k) &= \sum_{i=0}^{k-1} \left(\frac{q}{p}\right)^i P(1) \quad k = 1, 2, \dots, N
 \end{aligned}$$

Applying the formula for a finite geometric series

$$P(k) = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \frac{q}{p}} P(1) & p \neq \frac{1}{2} \\ kP(1) & p = \frac{1}{2} \end{cases}$$

Letting  $k = N$ , we obtain for  $p \neq \frac{1}{2}$

$$1 = P(N) = \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} P(1) \Rightarrow P(1) = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}$$

Similarly for  $p = \frac{1}{2}$ , we have  $P(1) = \frac{1}{N}$ .

So for  $k = 0, 1, \dots, N$  we have

$$P(k) = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2} \\ \frac{k}{N} & p = \frac{1}{2} \end{cases}$$

Think about what happens when  $N \rightarrow \infty$  for each case.