

# Stochastic Modelling Notes

Anthony Catterwell

February 26, 2019

## Contents

<b>1</b>	<b>Preliminaries</b>	<b>2</b>
1.1	Conditional Probability . . . . .	2
1.2	Conditional Expectation . . . . .	2
1.3	Stochastic Processes . . . . .	2
<b>2</b>	<b>Discrete Time Markov Chains</b>	<b>2</b>
2.1	Basic Definitions . . . . .	2
2.2	Modelling examples . . . . .	3
2.3	More complicated examples . . . . .	3
2.4	Chapman-Kolmogorov equations . . . . .	3
2.5	Classification of states . . . . .	3
2.6	Transience and recurrence . . . . .	4
2.7	Positive and null recurrence . . . . .	4
2.8	Periodicity of chains . . . . .	5
2.9	Stationary probabilities: Aperiodic case . . . . .	5
2.10	Stationary probabilities: Periodic case . . . . .	5
2.11	First passage probabilities and times . . . . .	6
2.12	Costs and rewards . . . . .	6
	2.12.1 Long-run average cost . . . . .	6
	2.12.2 Cost in transient states . . . . .	7
2.13	Reversibility . . . . .	7
<b>3</b>	<b>Poisson Processes</b>	<b>7</b>
3.1	Exponential Random Variable . . . . .	7
	3.1.1 Memoryless property . . . . .	8
	3.1.2 Properties of minimum of two exponentials . . . . .	8
	3.1.3 Strong memoryless property . . . . .	8
	3.1.4 Sums of I.I.D exponentials . . . . .	8
3.2	Poisson Processes . . . . .	8

# 1 Preliminaries

## 1.1 Conditional Probability

- **Definition 1.1.2** *Conditional probability*

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}$$

- **Theorem 1.1.4** *Law of Total Probability*

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

- **Theorem 1.1.8**

$$p(x) = \sum_y p(x|Y=y)p(y) \quad X, Y \text{ discrete}$$

$$p(x) = \int p(x|Y=y)f(y)dy \quad X \text{ discrete, } Y \text{ continuous}$$

$$f(x) = \sum_y f(x|Y=y)p(y) \quad X \text{ continuous, } Y \text{ discrete}$$

$$f(x) = \int f(x|Y=y)f(y)dy \quad X, Y \text{ continuous}$$

## 1.2 Conditional Expectation

$$E(X|Y=y) = \sum_{x \in S} xp(x|Y=y) \quad \text{if } X \text{ is discrete}$$

$$E(X|Y=y) = \int_{-\infty}^{\infty} xf(x|Y=y)dx \quad \text{if } X \text{ is continuous}$$

- **Theorem 1.2.3** *Tower Property*

For  $X$  and  $Y$  random variables

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

or in detail:

$$E[E(X|Y)] = \begin{cases} \sum_{y \in S} E(X|Y=y)p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy & \text{if } Y \text{ is continuous} \end{cases}$$

## 1.3 Stochastic Processes

- **Definition 1.3.1** *Stochastic process*

A *stochastic process*  $(X_t)_{t \in T}$  is an indexed collection of random variables. Set  $T$  is called the *index set*. The set  $S$  of all possible states is referred to as the *state space* of the process.

# 2 Discrete Time Markov Chains

## 2.1 Basic Definitions

- **Definition 2.1.1** *Markov property*

A stochastic process is said to have the *Markov property* if, given the present state, the future

events are independent of the past. For discrete-time discrete-space processes  $(X_n)_{n \in \mathbb{N}}$  this property can be stated as

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

for all  $j, i, i_{n-1}, \dots, i_0 \in S$  and  $n \in \mathbb{N}$ , and we also define  $p_{ij}(n) \equiv P(X_{n+1} = j | X_n = i)$  and refer to it as the (one step) transition probability from  $i$  to  $j$  at time  $n$ .

## 2.2 Modelling examples

## 2.3 More complicated examples

## 2.4 Chapman-Kolmogorov equations

- **Notation**

$p_{ij}^{(n)}$  denotes the probability of reaching state  $j$  from state  $i$  in  $n$  periods

- **Theorem 2.4.1** *Chapman-Kolmogorov equations*

$$p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^n P_{kj}^m$$

also

$$p^{n+m} = p^n p^m$$

- **Corollary 2.4.2**

If  $P^n$  is defined to be the  $n$ th power of a matrix  $P$ , then

$$P^{(n)} = P^n$$

- **Theorem 2.4.7**

A one-step transition matrix  $P$  and the initial distribution  $a^{(0)}$  completely characterises the DTMC, that is, all finite-dimensional probabilities can be calculated.

## 2.5 Classification of states

- **Definition 2.5.1** *Accessibility*

A state  $j$  is said to be accessible from state  $i$ , denoted  $i \rightarrow j$ , if  $\exists n \geq 0 \ni p_{ij}^{(n)} > 0$

- **Theorem 2.5.3** *Communication is an equivalence relation* that is

1.  $i \leftrightarrow i \quad \forall i \in S$  (reflexive)
2.  $i \leftrightarrow j \implies j \leftrightarrow i$  (symmetric)
3.  $i \leftrightarrow j, j \leftrightarrow i \implies i \leftrightarrow k$  (transitive)

- **Definition 2.5.4** *Communicating class*

Let  $C \subseteq S$ .  $C$  is a *communicating class* if

1.  $i \in C, j \in C \implies i \leftrightarrow j$
2.  $i \in C, i \leftrightarrow j \implies j \in C$

If, in addition to these properties, we cannot leave  $C$ , that is

$$\forall i \in C, \forall k \notin C \ni i \not\rightarrow k \implies C \text{ is a closed, communicating class}$$

- **Definition 2.5.5** *Irreducibility*

A DTMC is *irreducible* if the state space  $S$  is a single (closed) communicating class, and it is called *reducible* if it is composed of several communicating classes

## 2.6 Transience and recurrence

- **Notation**

$$T_j = \min\{n \geq 1 : X_n = j\}$$

$$\varrho_{ij} = P(T_j < \infty | X_0 = i)$$

Note that for  $i \neq j$ ,  $\varrho_{ij} > 0 \iff i \rightarrow j$

- **Definition 2.6.1** *Recurrence and transience*

State  $i$  is *recurrent* if  $\varrho_{ii} = 1$ , and *transient* if  $\varrho < 1$

- **Lemma 2.6.2**

$$P(N_i = \infty | X_0 = i) = \begin{cases} 1 & \text{if } i \text{ is recurrent} \\ 0 & \text{if } i \text{ is transient} \end{cases}$$

$$E(N_i | X_0 = i) = \begin{cases} \infty & \text{if } i \text{ is recurrent} \\ \frac{1}{1-\varrho_{ii}} & \text{if } i \text{ is transient} \end{cases}$$

- **Theorem 2.6.3**

If  $i \rightarrow j$  but  $\varrho_{ji} < 1$  then  $i$  is transient.

- **Corollary 2.6.4**

1. If  $i \rightarrow j$  and  $i$  is recurrent then  $\varrho_{ji} = 1$
2. If  $i \rightarrow j$  and  $i$  is recurrent then  $j$  is also recurrent
3. If  $i \rightarrow j$  and  $j$  is transient then  $j$  is also transient
4. Recurrence and transience are class properties

- **Theorem 2.6.7**

State  $i$  is recurrent if and only if  $\sum_{n=0}^{(n)} = \infty$

- **Lemma 2.6.9**

$\sum_{k=0}^{\infty} a_k$  converges if  $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} < 1$ , and it diverges if  $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} > 1$

## 2.7 Positive and null recurrence

- **Theorem 2.7.1** A recurrent state  $i$  is *positive recurrent* if and only if

$$p_{ii}^* = \frac{1}{m_{ii}} > 0$$

where

$$m_{ij} = E(T_j | X_0 = i) \quad \text{and} \quad p_{ij}^* = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N p_{ij}^{(n)}$$

$p_{ij}^*$  is the mean proportion of time spent at  $j$  when starting from  $i$ .

- **Lemma 2.7.2**

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N f_n = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N f_{n+m}$$

for any sequence  $(f_n)_{n \in \mathbb{N}}$  which is bounded  $|f_n| \leq b$  for all  $n$ , and for any  $m \in \mathbb{N}$

- **Theorem 2.7.3**

Positive recurrence and null recurrence are class properties, that is if recurrent states  $i \leftrightarrow j$  then  $i$  and  $j$  are both positive or null recurrent.

- **Theorem 2.7.4** All states in a finite closed communicating class are positive recurrent
- **Theorem 2.7.5** All states in an open communicating class are transient.

## 2.8 Periodicity of chains

- **Definition 2.8.1** *Period*

The *period*  $d$  of a state  $i$  is the greatest common factor of  $\{n \geq 0 : p_{ii}^{(n)} > 0\}$ . If  $d = 1$ , the state is called aperiodic, and for  $d \geq 2$ , it is called periodic.

- **Theorem 2.8.2** Period is a class property, that is  $i \leftrightarrow j \implies d_i = d_j$ .

## 2.9 Stationary probabilities: Aperiodic case

- **Definition 2.9.1** *Stationary distribution*

A distribution  $\pi = (\pi_1, \pi_2, \dots) \geq 0$  is *stationary* if it satisfies the *global balance equations*, i.e.

$$\pi = \pi P, \quad \text{and} \quad \sum_{j \in S} \pi_j = 1$$

(i.e. the next state distribution is the same as the current distribution.)

A chain with a stationary distribution is said to be in a stationary state or steady state.

- **Proposition 2.9.2**

If a chain is initially in a stationary distribution,  $a^{(0)} = \pi$ , then  $a^{(n)} = \pi \forall n \geq 0$ .

- Intuition

$$\lim_{n \rightarrow \infty} p_{jj}^{(n)} = p_{jj}^* = \frac{1}{m_{jj}} > 0$$

Think about our simple weather example and ask for the probability that in a million years it will rain if it rains today. Now what is this probability a million years and one day later? Intuitively, we can argue that these probabilities should be equal after a really long time, and it should be equal to the long run average proportions of the days that are rainy.

- **Theorem 2.9.3**

1. For aperiodic, irreducible chains, the *limiting probabilities* are independent of the initial state, that is  $\forall i, j \in S$ ,

$$\lim_{n \rightarrow \infty} p_{jj}^{(n)} = \lim_{n \rightarrow \infty} p_{ij}^{(n)} \equiv \pi_j$$

and we call this limit  $\pi_j$ .

2. If the chain is also positive recurrent then the limiting probability distribution is the unique stationary distribution.

## 2.10 Stationary probabilities: Periodic case

- **Theorem 2.10.1** (analogous to Theorem 2.9.3)

1. For irreducible chains, the steady-state probabilities  $\pi_j$  are independent of the initial state, that is,

$$\pi_j = p_{jj}^* = p_{ij}^*$$

for every  $i, j \in S$ .

2. If the chain is also positive recurrent then the limiting probability distribution  $\pi$  is the unique stationary distribution

- Summary

	<i>Aperiodic</i>	<i>Periodic</i>
$\lim_{n \rightarrow \infty} p_{ij}^{(n)}$	Exist	Do not exist
$p_{ij}^*$	Exist	Exist
Interpretation for $\pi_j$	Stationary probability, Limiting probability	Stationary probability

$$\pi_j \equiv p_{jj}^* = p_{ij}^* = \lim_{n \rightarrow \infty} p_{jj}^{(n)} = \lim_{n \rightarrow \infty} p_{ij}^{(n)} \quad (\text{aperiodic})$$

$$\pi_j \equiv p_{jj}^* = p_{ij}^* \quad (\text{periodic})$$

## 2.11 First passage probabilities and times

- Preamble

$$T_j = \min\{n \geq 1 : X_n = j\}$$

Define the *first passage time* of a chain to a set  $A \subset S$  as

$$\hat{T}_A = \min\{n \geq 0 : X_n \in A\}$$

Note that  $\hat{T}_A = 0$  for  $X_0 \in A$ , that is if we are already in  $A$ , it takes no time to get there.

Otherwise, the *first passage time*  $\hat{T}_A$  is identical to the *first arrival time*

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

for  $X_0 \in S - A$ . We defined this variation for convenience.

- **Theorem 2.11.1**

Let  $A, B \subset S$ , with  $P(\min\{\hat{T}_A, \hat{T}_B\} < \infty | X_0 = i) = 1, \forall i$ .

Then the probability  $h_i \equiv P(\hat{T}_A < \hat{T}_B | X_0 = i)$  of reaching set  $A$  before set  $B$  when starting from state  $i$  satisfies

$$h_i \equiv P(\hat{T}_A < \hat{T}_B | X_0 = i) = \begin{cases} 0 & \text{if } i \in B \\ 1 & \text{if } i \in A \\ \sum_{j \in S} P_{ij} h_j & \text{if } i \in S - (A \cup B) \end{cases}$$

- **Theorem 2.11.2**

Let  $A \subset S$ , with  $P(\hat{T}_A < \infty | X_0 = i) = 1, \forall i$ . Then the mean time  $g_i \equiv E(\hat{T}_A | X_0 = i)$  to reach set  $A$  when starting from state  $i$  satisfies

$$g_i \equiv E(\hat{T}_A | X_0 = i) = \begin{cases} 0 & \text{if } i \in A \\ 1 + \sum_{j \in S} P_{ij} g_j & \text{if } i \in S - A \end{cases}$$

## 2.12 Costs and rewards

- Preamble

At step  $n$  we incur a cost of  $c(X_n)$

### 2.12.1 Long-run average cost

We can write the *long-run average cost* starting from state  $i$  to be

$$\psi_i = \lim_{N \rightarrow \infty} \frac{1}{N+1} E \left( \sum_{n=0}^N c(X_n) | X_0 = i \right)$$

The long-run average cost is independent of the initial state and is calculated using steady-state probabilities. Note that you'll get the same expression for the *long-run mean cost*

$$\lim_{n \rightarrow \infty} E(c(X_n) | X_0 = i) = \sum_{j \in S} c(j) \pi_j$$

for chains which are also aperiodic.

### 2.12.2 Cost in transient states

Another situation is when only transient states have non-zero costs. In this case the total final cost can be defined  $\sum_{n=0}^{\infty} c(X_n)$ .

## 2.13 Reversibility

- **Definition 2.13.1** *Reversed process*

If  $(X_n)_{n \in \mathbb{N}}$  is a stationary DTMC and we fix an  $m$ , then the process  $(\tilde{X}_n)_{0 \leq n \leq m}$  where  $\tilde{X}_n = X_{m-n}$  is called the reversed process of  $X$ .

- **Theorem 2.13.2**

The reversed process  $(\tilde{X}_n)_{0 \leq n \leq m}$  is a DTMC with transition probabilities

$$\tilde{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i}$$

- **Definition 2.13.3** *Reversibility*

A stationary DTMC is said to be reversible if the reversed process is stochastically the same as the original process, that is  $\tilde{p}_{ij} = p_{ij}$ , which implies

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i, j \in S$$

This equation together with  $\sum_{i \in S} \pi_i = 1$  are called the *detailed balance equations*.

- **Corollary 2.13.4**

If  $\pi$  satisfies the *detailed balance equations*, then it also satisfies the *global balance equations*.

- A *tree DTMC* has the following properties:

- $p_{ij} > 0 \implies p_{ji} > 0$
- No cycles in its state diagram

- **Theorem 2.13.7** A stationary *tree DTMC* is reversible.

- **Corollary 2.13.8**

A *stationary random walk* that is a positive recurrent random walk in stationarity, is reversible.

## 3 Poisson Processes

### 3.1 Exponential Random Variable

- **Definition 3.1.1:** *Exponential random variable*

A continuous non-negative random variable  $X$  is called exponential with rate  $\lambda$  if its cumulative distribution function is

$$P(X \leq x) = F(x) = 1 - e^{-\lambda x}$$

Consequently, its density is

$$f(x) = \lambda e^{-\lambda x}$$

both for  $x \geq 0$ , and zero otherwise.

- **Theorem 3.1.2**

The  $r$ -th moment of the exponential random variable with rate  $\lambda$  is given by

$$E(X^r) = \frac{r!}{\lambda^r}$$

### 3.1.1 Memoryless property

$$P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

- **Theorem 3.1.3**

The only continuous distribution which has a support  $[0, \infty)$  with memoryless property is the exponential.

### 3.1.2 Properties of minimum of two exponentials

### 3.1.3 Strong memoryless property

- **Theorem 3.14**

If  $X_2$  is an exponential random variable with rate  $\lambda$  and  $X_1$  is an independent non-negative continuous random variable, then  $\forall x \geq 0$

$$P(X_2 > X_1 + x | X_2 > X_1) = P(X_2 > x) = e^{-\lambda x}$$

### 3.1.4 Sums of I.I.D exponentials

- **Theorem 3.1.5**

If  $Z = X_1 + X_2 + \dots + X_n$ , where  $X_i \approx \exp(\lambda)$  for all  $i$  and independent, then  $Z$  is called the gamma  $(n, \lambda)$  random variable and its density function is given by

$$f_n(z) = \lambda e^{-\lambda z} \frac{(\lambda z)^{n-1}}{(n-1)!}$$

## 3.2 Poisson Processes

- **Definition 3.2.1**

Let  $\tau_i$  be independent exponential  $(\lambda)$  random variables,  $S_0 = 0, S_n = \tau_1 + \tau_2 + \dots + \tau_n$  and  $N_t = \max\{n \geq 0 : S_n \leq t\}$ . Then  $(N_t)_{t \in \mathbb{R}_{\geq 0}}$  is a *Poisson process* with rate parameter  $\lambda$ , or briefly  $PP(\lambda)$ .

++i