Stochastic Modelling Notes

Anthony Catterwell

March 11, 2019

Contents

1 Preliminaries **2**
1.1 Conditional Probability 2
1.2 Conditional Expectation 2
1.3 Stochastic Processes 2

2 Discrete Time Markov Chains **3**
2.1 Basic Definitions 3
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

3 Poisson Processes **7**
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
3.3 Super-positioning and Splitting 9
3.4 Campbell's Theorem: Uniform Order Statistics 9
3.5 Non-homogeneous Poisson Process 9
3.5.1 Event Times for Non-homogeneous PP 10
3.6 Compound Poisson Processes 10

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**

$$\begin{aligned}
p(x) &= \sum_y p(x|Y=y)p(y) && X, Y \, \text{discrete} \\
p(x) &= \int p(x|Y=y)f(y)\mathrm{d}y && X \, \text{discrete}, Y \, \text{continuous} \\
f(x) &= \sum_y f(x|Y=y)p(y) && X \, \text{continuous}, Y \, \text{discrete} \\
f(x) &= \int f(x|Y=y)f(y)\mathrm{d}y && X, Y \, \text{continuous}
\end{aligned}$$

1.2 Conditional Expectation

$$\begin{aligned}
E(X|Y=y) &= \sum_{x \in S} xp(x|Y=y) && \text{if } X \text{ is discrete} \\
E(X|Y=y) &= \int_{-\infty}^{\infty} xf(x|Y=y)\mathrm{d}x && \text{if } X \text{ is continuous}
\end{aligned}$$

- **Theorem 1.2.3** *Tower Property*
For X and Y random variables

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

or in detail:

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

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 $(X_n)_{n\in N}$ 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 nth 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 $i \to j$, 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

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

- **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 $i \neq j,\ \varrho_{ij} > 0 \iff i \to j$

- **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 $i \to j$ but $\varrho_{ji} < 1$ then i is transient.

- **Corollary 2.6.4**
  1. If $i \to j$ and i is recurrent then $\varrho_{ji} = 1$
  2. If $i \to j$ and i is recurrent then j is also recurrent
  3. If $i \to j$ 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 $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} < 1$, and it diverges if $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} > 1$

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

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

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

- **Lemma 2.7.2**

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

  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 $\pi = (\pi_1, \pi_2, \ldots) \geq 0$ is *stationary* if it satisfies the *global balance equations*, i.e.
$$\pi = \pi P, \quad \text{and} \sum_{j \in S} \pi_j = 1$$
(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
$$\lim_{n \to \infty} p_{jj}^{(n)} = p_{jj}^* = \frac{1}{m_{jj}} > 0$$
Think about our simple weather example and ask for the probability than 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 $\forall i, j, \in S$,
  $$\lim_{n \to \infty} p_{jj}^{(n)} = \lim_{n \to \infty} p_{ij}^{(n)} \equiv \pi_j$$
  and we call this limit π_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 π_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 π is the unique stationary distribution

- Summary

	Aperiodic	*Periodic*
$\lim_{n\to\infty} p_{ij}^{(n)}$	Exist	Do not exist
p_{ij}^*	Exist	Exist
Interpretation for π_j	Stationary probability, Limiting probability	Stationary probability

$$\pi_j \equiv p_{jj}^* = p_{ij}^* = \lim_{n\to\infty} p_{jj}^{(n)} = \lim_{n\to\infty} p_{ij}^{(n)} \qquad \text{(aperiodic)}$$
$$\pi_j \equiv p_{jj}^* = p_{ij}^* \qquad \text{(periodic)}$$

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

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

Note that $\widehat{T}_A = 0$ 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* $\widehat{T}_A$ 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 $P(\min\{\widehat{T}_A, \widehat{T}_B\} < \infty | X_0 = i) = 1,\ \forall i$.
Then then probability $h_i \equiv P(\widehat{T}_A < \widehat{T}_B | X_0 = i)$ of reaching set A before set B when starting from state i satisfies

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

- **Theorem 2.11.2**
Let $A \subset S$, with $P(\widehat{T}_A < \infty | X_0 = i) = 1,\ \forall i$. Then the mean time $g_i \equiv E(\widehat{T}_A | X_0 = i)$ to reach set A when starting from state i satisfies

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

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

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

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*

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

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 $(X_n)_{n\in N}$ is a stationary DTMC and we fix an m, then the process $(\widetilde{X}_n)_{0\leq n\leq m}$ where $\widetilde{X}_n = X_{m-n}$ is called the reversed process of X.
- **Theorem 2.13.2**
  The reversed process $(\widetilde{X}_n)_{0\leq n\leq m}$ is a DTMC with transition probabilities

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

- **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 $\widetilde{p}_{ij} = p_{ij}$, which implies

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

  This equation together with $\sum_{i\in S}\pi_i = 1$ are called the *detailed balance equations.*
- **Corollary 2.13.4**
  If π 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 λ 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 λ is given by

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

3.1.1 Memoryless property

The *memoryless property* of a stochastic distribution can be stated as:

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

- **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

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$
$$Z = \min(X_1, X_2) \implies P(Z > x) = \mathrm{e}^{-(\lambda_1+\lambda_2)x}$$
$$P(Z > x, X_1 < X_2) = P(Z > x)P(X_1 < X_2) \quad \text{i.e. they are independent}$$

3.1.3 Strong memoryless property

- **Theorem 3.14**
  If X_2 is an exponential random variable with rate λ and X_1 is an independent non-negative continuous random variable, then $\forall x \geq 0$

$$P(X_2 > X_{1x} | X_2 > X_1) = P(X_2 > x) = e^{-\lambda x}$$

3.1.4 Sums of I.I.D exponentials

- **Theorem 3.1.5**
  If $Z = X_1 + X_2 + \cdots + X_n$, where $X_i \approx \exp(\lambda)$ for all i and independent, then Z is called the gamma (n, λ) random variable and its density function is given by

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

3.2 Poisson Processes

- **Definition 3.2.1**
  Let τ_i be independent exponential (λ) random variables, $S_0 = 0, s_n = \tau_1 + \tau_2 + \cdots + \tau_n$ 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 λ, or briefly PP(λ).
- **Theorem 3.2.2**
  If $(N-t)_{t \geq 0}$ is a PP(λ), then N_t follows a Poisson distribution with rate λt for any t, that is

$$P(N_t = k) = \frac{\mathrm{e}^{-\lambda t}(\lambda t)^k}{k!}$$

- **Definition**

$$\left(N_t^{(s)}\right)_{t \geq 0} = N_{s+t} - N_s \quad ; \quad \forall t \geq 0$$

  I.e. *resetting* the counter at s.

- **Theorem 3.2.4** The process $(N_t^{(s)})_{t\geq 0}$ is a PP(λ), and it is independent of $(N_u)_{0\leq u\leq s}$.
- **Definition 3.2.6**
  1. A process $(N_t)_{t\geq 0}$ is said to have *stationary increments* if $N_{s+t}-N_s$ is identically distributed for all s, that is the distribution does not depend on s.
  2. A process $(N_t)_{t\geq 0}$ is said to have independent increments if the increments of the distribution is independent for non-overlapping intervals, that is $N_{s_1+t_1}-N_{s_1}$ and $N_{s_2+t_2}-N_{s_2}$ are independent if $[s_1, s_2+t_1]\cap[s_2, s_2+t_2]=\emptyset$.
- **Theorem 3.2.7** $(N_t)_{t\geq 0}$ is a PP(λ) if and only if
  1. it has stationary and independent increments
  2. N_t is a Poisson (t) random variable for all t.
- **Definition 3.2.8** A function $f(x)$ is said to be a $o(x)$ function, if $\lim_{x\to 0}\frac{f(x)}{x}=0$.
- **Theorem 3.2.10** $(N_t)_{t\geq 0}$ is a PP(λ) if and only if
  1. it has stationary and independent increments
  2. $$\begin{aligned}P(N_h=0)&=1-\lambda h+o(h)\\ P(N_h=1)&=\lambda h+o(h)\\ P(N_h\geq 2)&=o(h)\end{aligned}$$

3.3 Super-positioning and Splitting

- **Theorem 3.3.1** Let $(N_t^{(i)})_{t\geq 0}$ be PP(λ_I) for each $i=1,\dots,k$. Then, with $N_t=N_t^{(1)}+\dots+N_t^{(k)}$, $(N_t)_{t\geq 0}$ is a PP($\lambda_1+\dots+\lambda_k$).
- **Theorem 3.3.2** Suppose $(N_t)_mt\geq 0$ is a PP(λ). If an event in this process is of type i with probability p_i independent of other events, where $\sum_{i=1}^k p_i=1$, then the processes

  $(N_t^{(1)})_{t\geq 0}, (N_t^{(2)})_{t\geq 0}, \dots, (N_t^{(k)})_{t\geq 0}$ are independent Poisson processes with rates $\lambda p_1, \lambda p_2, \dots, \lambda p_k$ respectively.

3.4 Campbell's Theorem: Uniform Order Statistics

- **Theorem 3.4.1 Campbell's Theorem**
  Let S_n be the event times for a Poisson process.

  If $N_t=n$ is given, then the vector $(S_1, S_2, \dots, \S_n)$ follows the distribution of ordered independent uniform variables $(U_{(1)}, U_{(2)}, \dots, U_{(n)})$.

  Consequently, the unordered set of arrival times $\{S_1,\dots,S_n\}$ has the same distribution as $\{U_1,\dots,U_n\}$.

  In other words, $\{S_1,\dots,S_n\}$ is uniformly distributed.

3.5 Non-homogeneous Poisson Process

Having now relaxed the stationary assumption, $\lambda(t)$ depends on time.

- **Definition 3.5.1** A counting process $(N_t)_{t\geq 0}$ is a non-homogeneous Poisson process if
  1. it has independent increments

2.

$$
\begin{aligned}
P(N_{t+h} - N_t = 0) &= 1 - \lambda(t)h + o(h) \\
P(N_{t+h} - N_t = 1) &= \lambda(t)h + o(h) \\
P(N_{t+h} - N_t \geq 2) &= o(h)
\end{aligned}
$$

Interestingly, the number of arrivals at any fixed time t is still Poisson.

- **Theorem 3.5.2** Define
$$\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$$
Then N_t is a Poisson $\Lambda(t)$ random variable for any t.
- **Corollary 3.5.3** $N_t - N_s$ is a Poisson random variable with parameter $\Lambda(t) - \Lambda(s) = \int_s^t \lambda(u)\mathrm{d}u$.
- **Corollary 3.5.5** The super-positioning and splitting properties are true for non-homogeneous Poisson processes as well.

3.5.1 Event Times for Non-homogeneous PP

- **Theorem 3.5.6** Let $(N_t)_{t\geq 0}$ be a non-homogeneous PP, then if $N_t = k$ is given, then
$$(S_1, S_2, \ldots, S_k) \sim (U_{(1)}, U_{(2)}, \ldots, U_{(k)})$$

3.6 Compound Poisson Processes