

ACTL2102 Course Notes

Lecturer Email: j.k.woo@unsw.edu.au

Introduction to Stochastic Modelling

Stochastic vs Deterministic Models

- **Stochastic Model:** recognises random nature of inputs and outputs
- **Deterministic Model:** no random component. Inputs and outputs are fixed, and relationships are defined
 - For example, the output of a stochastic model could be a distribution rather than a fixed value

Stochastic Process

- Definition: A stochastic process is any collection of random variables $X(t)$, $t \in T$
 - ▶ T is called the *index set* of the process and is usually a subset of $[0, \infty)$ and is the time parameter.
 - ▶ $X(t)$ is the *state* of the process at time t .
 - ▶ The set of values that $X(t)$ can take is called the *state space* of the process and is often denoted by S .
 - ▶ The stochastic process is denoted as

$$\{X(t), t \in T\}.$$

Stochastic Process Classification

- Classification according to index set:
 - Discrete Time Process: the index set is finite or countably infinite
 - Continuous Time Process: the index set is continuous
- Classification according to state space:
 - Discrete State Space: $X(t)$ is a discrete random variable
 - Continuous State Space: $X(t)$ is a continuous random variable

More Definitions

- **Sample Path/Realisation:** a particular assignment of possible values for a stochastic process
- **Independent Increments:** a stochastic process has independent increments if the random variables $X(t_0)$, $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, $X(t_n) - X(t_{n-1})$ are independent.
 - OR, if future increases are independent of the past and the present
- **Stationary Increments:** a stochastic process has stationary increments if all increments over different time periods of the same length have the same distribution.

Discrete Time Markov Chains

Markov Chains

- Definition: A Markov process has the Markov property which states:
$$\Pr(X(t_n) \leq x_n | X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_{n-1}) = x_{n-1}) \\ = \Pr(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1})$$
 - i.e. Given the present state, the past states do not have an influence on the future.

- Markov Chain: A Markov process on a discrete index set $T = \{0, 1, 2, \dots\}$ which is denoted by X_n .
 - The state space can either be finite or countably infinite.
 - $X_n = k$ means the process is in state k at time n .
- The conditional probabilities are called **one-step transition probabilities** of the Markov Chain.

$$P_{ij}^{(n,n+1)} = \Pr(X_{n+1} = j | X_n = i)$$

Example:

- The wealth of a gambler at day n is denoted by X_n . The gambler's initial wealth is $X_0 = 100$. The gambler will win Z_n on day n ($n = 1, 2, \dots$) and the amount that he wins each day is 1 with probability p or -1 with probability $1 - p$. Assume that Z_1, \dots, Z_n, \dots are independent.
- We can see that $X_n = X_0 + \sum_{k=1}^n Z_k$ and $X_n = X_k + Z_{k+1} + \dots + Z_n$. Then for $n > k$

$$\begin{aligned} \Pr(X_n = x_n | X_0 = x_0, \dots, X_k = x_k) \\ = \Pr(X_k + Z_{k+1} + \dots + Z_n = x_n | X_k = x_k) \\ = \Pr(X_n = x_n | X_k = x_k) \end{aligned}$$

Then the stochastic process $\{X_t : t = 0, 1, \dots\}$ is a Markov chain.

Transition Probability

- If the one-step transition probabilities do not depend on time, we drop the n notation:
- $P_{ij} = P_{ij}^{(n,n+1)} = \Pr(X_{n+1} = j | X_n = i)$
 - This is called a **homogeneous Markov chain** and the transition probabilities are **stationary**.
- The matrix consisting of these transition probabilities is called the transition matrix:

$$P = (P_{ij}) = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ P_{20} & P_{21} & P_{22} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

- It satisfies:

$$P_{ij} \geq 0 \text{ for } i, j = 0, 1, 2, \dots \text{ and } \sum_{j=0}^{\infty} P_{ij} = 1 \text{ for } i = 0, 1, 2, \dots$$

Chapman-Kolmogorov Equations

- The n -step transition probability is the probability that a process in state i will be in state j in n steps, that is:

$$P_{ij}^n = \Pr(X_{n+m} = j | X_m = i)$$

- Chapman-Kolmogorov Equations provide a procedure to compute n-step transition probabilities:
- As well as:

▶ **Kolmogorov Forward equations:**

$$P_{ij}^{n+1} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj} \text{ for all } n \geq 0, \text{ all } i, j.$$

▶ **Kolmogorov Backward equations:**

$$P_{ij}^{n+1} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj} \text{ for all } n \geq 0, \text{ all } i, j.$$

▶ Note that beginning with

$$P^{(2)} = P^{(1+1)} = P \bullet P = P^2$$

and continuing by induction, we can obtain the n -step ($n > 0$) transition matrix

$$P^{(n)} = P^{(n-1+1)} = P^{n-1} \bullet P = P^n.$$

Example:

a. Suppose that:

- ▶ coin 1 has probability 0.7 of coming up heads;
- ▶ coin 2 has probability 0.6 of coming up heads.

If the coin flipped today comes up:

- ▶ heads, then we select coin 1 to flip tomorrow;
- ▶ tails, then we select coin 2 to flip tomorrow.

If the coin initially flipped is equally likely to be coin 1 or coin 2, then what is the probability that the coin flipped on the third day after the initial flip is coin 1?

b. Suppose in a. that the coin flipped on Monday comes up heads. What is the probability that the coin flipped on Friday of the same week comes up heads?



Solution:

a. Let X_n denote the label of the coin that is flipped on the n th day after the initial flip. X_n is a Markov chain.

The transition probability matrix is given by

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{pmatrix}$$

So

$$P^{(2)} = P^2 = \begin{pmatrix} 0.67 & 0.33 \\ 0.66 & 0.34 \end{pmatrix}$$

$$P^{(3)} = P^3 = \begin{pmatrix} 0.667 & 0.333 \\ 0.666 & 0.334 \end{pmatrix}.$$

Hence the required probability is

$$\begin{aligned} & \Pr(X_3 = 1) \\ &= \Pr(X_3 = 1|X_0 = 1)\Pr(X_0 = 1) + \Pr(X_3 = 1|X_0 = 2)\Pr(X_0 = 2) \\ &= \frac{1}{2}(P_{11}^3 + P_{21}^3) = 0.6665 \end{aligned}$$



b. Let X_n represents the result of the flip on day n , that is,

$$X_n = \begin{cases} 1 & \text{if heads} \\ 2 & \text{otherwise} \end{cases}$$

$\{X_n, n = 0, 1, \dots\}$ is a Markov chain.

We have

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{pmatrix}.$$

$$P^{(4)} = P^4 = \begin{pmatrix} 0.6667 & 0.3333 \\ 0.6666 & 0.3334 \end{pmatrix}.$$

Then the desired probability is $P_{11}^4 = 0.6667$.

Example:

Suppose that whether or not it rains today depends on previous weather conditions over the last two days. Specifically, suppose that if:

- ▶ it has rained for the past two days, then it will rain tomorrow with probability 0.7;
- ▶ if it rained today but not yesterday, then it will rain tomorrow with probability 0.5;
- ▶ if it rained yesterday but not today, then it will rain tomorrow with probability 0.4;
- ▶ if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

If we know that it rained on both Monday and Tuesday, what is the probability that it will rain on Thursday.

Solution:

Let X_n be in:

- ▶ state 0: if it rained both today and yesterday,
- ▶ state 1: if it rained today but not yesterday,
- ▶ state 2: if it rained yesterday but not today,
- ▶ state 3: if it did not rain either yesterday or today .

The transition probability matrix

$$P = \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix}$$

Then the two-step transition matrix is given by

$$P^{(2)} = P^2 = \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix} \cdot \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix}$$

$$= \begin{pmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{pmatrix}$$

Since raining on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the desired probability is given by

$$P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61.$$

Classification of States

- ▶ **Definition:** A state i is said to be an *absorbing state* if $P_{ii} = 1$ or, equivalently $P_{ij} = 0$ for all $j \neq i$.
- ▶ **Definition:** State j is *accessible* from state i if $P_{ij}^n > 0$ for some $n \geq 0$. This is written as $i \rightarrow j$, i leads to j or j is accessible from i .
 - ▶ Note that if $i \rightarrow j$ and $j \rightarrow k$ then $i \rightarrow k$.
- ▶ **Definition:** States i and j *communicate* if $i \rightarrow j$ and $j \rightarrow i$. This is written as $i \leftrightarrow j$.
 - ▶ Note that $i \leftrightarrow i$ for all i , if $i \leftrightarrow j$ then $j \leftrightarrow i$, and if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.
- ▶ **Definition:** The *class* of states that communicate with state i is $C(i) = \{j \in S : i \leftrightarrow j\}$.
 - ▶ Two states that communicate are in the same class.
 - ▶ Note that for any two states $i, j \in S$ either $C(i) = C(j)$ or $C(i) \cap C(j) = \emptyset$. Also if $j \in C(i)$ then $C(i) = C(j)$.

Irreducible Markov Chains

- Definition: A Markov is irreducible if there is only one class.
 - i.e. If all states communicate with each other.

Recurrent vs Transient States

- It is important to know how many times a process visits a given state. We need to distinguish between two types of states:
 - Denote by f_i the probability that the process starting in state i will at some future time return to state i .
 - Definition:** State i is said to be *recurrent* if $f_i = 1$.
 - Definition:** State i is said to be *transient* if $f_i < 1$.
 - If state i is recurrent then, starting in state i , the process will enter state i infinitely often.
 - Starting in state i , if state i is transient, then the probability that the process will be in state i for exactly $n \geq 1$ time periods is given by a geometric distribution $f_i^{n-1} (1 - f_i)$. The expected number of time periods that the process is in state i is $\frac{1}{(1 - f_i)}$.

Example:

Consider a Markov chain with state space $S = \{0, 1, 2, \dots, 5\}$ and transition matrix

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

Classify the states and identify the recurrent and transient states, if any.

Solution:

Accessibility table:

0	\rightarrow	0, 2, 5
1	\rightarrow	0, 1, 2, 3, 4, 5
2	\rightarrow	0, 2, 5
3	\rightarrow	3
4	\rightarrow	0, 1, 2, 3, 4, 5
5	\rightarrow	0, 2, 5

- State 3 is clearly absorbing, therefore recurrent.
- The class $\{0, 2, 5\}$ is finite and closed, therefore consists of recurrent states.
- The class $\{1, 4\}$ is open (because they can lead to other states outside this set), therefore must consist of transient states.

In summary, 0, 2, 3, and 5 are recurrent states while 1 and 4 are transient states.

- The number of periods that the process is in state i is given by:

$$\sum_{n=0}^{\infty} I_n \text{ where } I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

so that

$$\begin{aligned} E \left[\sum_{n=0}^{\infty} I_n | X_0 = i \right] &= \sum_{n=0}^{\infty} E [I_n | X_0 = i] \\ &= \sum_{n=0}^{\infty} \Pr [X_n = i | X_0 = i] = \sum_{n=0}^{\infty} P_{ii}^n \end{aligned}$$

- State i is:

recurrent if $\sum_{n=0}^{\infty} P_{ii}^n = \infty$

transient if $\sum_{n=0}^{\infty} P_{ii}^n < \infty$.

- A transient state will only be visited a finite number of times
- In a finite state Markov Chain, at least **one** state must be recurrent, and not all states can be transient.
 - Otherwise, after a finite number of times, no states will be visited which is impossible.
- If state i is recurrent and state i communicates with state j , then state j is recurrent.
- If state i is transient and it communicates with state j , then state j is transient.
- All states in a finite irreducible Markov chain are recurrent.
- Definition: A class of states is **recurrent** if all states in the class are recurrent
- Definition: A class of states is **transient** if all states in the class are transient.

Period

- State i has period $d(i)$ if:
 - $P_{ii}^n = 0$ whenever n is not divisible by $d(i)$,
 - and $d(i)$ is the largest integer with this property;
 - or equivalently, if $d(i)$ is the greatest common divisor of all $n \geq 1$ for which $P_{ii}^n > 0$.
- If $P_{ii}^n = 0$ for all $n \geq 1$ then define $d(i) = 0$.

Example:

Consider the simple symmetric random walk defined by

$$X_n = Y_1 + Y_2 + \cdots + Y_n = \sum_{j=1}^n Y_j,$$

where the random variables Y_j , for $j = 1, 2, \dots$ are *i.i.d* and are defined by

$$Y_j = \begin{cases} 1, & \text{w.p. } 1/2 \\ -1, & \text{w.p. } 1/2 \end{cases} .$$

Starting in state 0, it is possible for the process to enter state 0, only at times $n = 0, 2, 4, \dots$, that is, $P_{00}^n > 0$ only if $n = 0, 2, \dots$

So $d(0) = 2$.

Important Results

- A state with period 1 is called **aperiodic**
- Periodicity is a class property. That is, if $i \leftrightarrow j$, then $d(i) = d(j)$.
- If state i is recurrent and, starting in state i , the expected time until the process returns to state i is finite, then state i is **positive recurrent**.
 - Positive recurrence is a class property
- In a finite Markov chain, all recurrent states are positive recurrent.
- If a state is positive recurrent and aperiodic, then it is called **ergodic**.

Limiting Probabilities

- For an irreducible ergodic Markov chain: $\lim_{n \rightarrow \infty} P_{ij}^n$ exists and is independent of i .
- ▶ Letting $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n$, $j \geq 0$, then
 - ▶ π_j is the unique nonnegative solution to the set of equations:

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \text{ for } j \geq 0; \quad \sum_{j=0}^{\infty} \pi_j = 1,$$

- ▶ π_j equals the long run proportion of time the Markov chain is in state j .
- ▶ π_j equals the limiting probability that the process will be in state j .

Stationary Probabilities

- The long run proportions $\pi_j, j \geq 0$, are called **stationary probabilities**.
If the initial state is chosen so that $\Pr(X_0 = j) = \pi_j$, then

$$\Pr(X_n = j) = \pi_j \text{ for all } n, j \geq 0,$$

or in matrix form $\pi = \pi P$, where $\pi = (\pi_0, \pi_1, \dots)$.

- If m_{jj} is the expected number of transitions until a Markov Chain which starts in state j returns to state j (the mean time between visits to state j), then:

$$\pi_j = \frac{1}{m_{jj}}.$$

- i.e. The proportion of time spent in state j equals the inverse of the mean time between visits to state j
- If the Markov chain is transient then:
$$\lim_{n \rightarrow \infty} P_{ij}^n = 0.$$

No Claim Discount Insurance

- The NCD system in car insurance is a system where the premium charged depends on the policyholder's claim record (usually modelled with a Markov chain)
- Initially, they will be charged the full premium for their risk class based on certain factors (age, engine, etc)
- If a policyholder has a claim free year, then they will move up to a higher discount level. If they have one or more claims, they will move to a lower level.

Example:

An insurance company allows the following NCD levels:

NCD Level	Discount %
0	0
1	25
2	50

If a policyholder has no claims in a year then they move to the next higher level of discount (unless they are already on the highest level, in which case they stay there). If a policyholder has one or more claims in a year then they move to the next lower level of discount (unless they are already on the lowest level, in which case they stay there). The probability of having no claims in a year is assumed to be 0.9.

- a. Give the states and probability transition matrix for the NCD scheme.
- b. Determine the stationary probabilities for the states.
- c. Determine the mean time between the different NCD levels.

Solution:

- a. The state space of the Markov chain is $\{0, 1, 2\}$ where state i corresponds to NCD level i . The transition matrix is

$$P = \begin{pmatrix} 0.1 & 0.9 & 0 \\ 0.1 & 0 & 0.9 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

b. Stationary probabilities are obtained by solving

$$\pi_j = \sum_i \pi_i P_{ij}, \quad 1 = \pi_0 + \pi_1 + \pi_2$$

That is,

$$\begin{aligned}\pi_0 &= P_{00}\pi_0 + P_{10}\pi_1 + P_{20}\pi_2 \\ \pi_1 &= P_{01}\pi_0 + P_{11}\pi_1 + P_{21}\pi_2 \\ \pi_2 &= P_{02}\pi_0 + P_{12}\pi_1 + P_{22}\pi_2\end{aligned}$$

Solving

$$\begin{aligned}\pi_0 &= 0.1\pi_0 + 0.1\pi_1 \\ \pi_1 &= 0.9\pi_0 + 0.1\pi_2 \\ \pi_2 &= 0.9\pi_1 + 0.9\pi_2\end{aligned}$$

subject to

$$1 = \pi_0 + \pi_1 + \pi_2$$

results in

$$\begin{aligned}\pi_0 &= \frac{1}{91} = 0.010989 \\ \pi_1 &= \frac{9}{91} = 0.098901 \\ \pi_2 &= \frac{81}{91} = 0.890110\end{aligned}$$

c. Mean Time Between Visits to State j , $m_{jj} = \frac{1}{\pi_j}$

$$\begin{aligned}m_{00} &= 91 \\ m_{11} &= \frac{91}{9} = 10.11 \\ m_{22} &= \frac{91}{81} = 1.12.\end{aligned}$$

Mean time Spent in Transient States

- For a finite state Markov Chain, denote the transient states by $T = \{1, 2, \dots, t\}$. Let:

$$P_T = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1t} \\ P_{21} & P_{22} & \dots & P_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ P_{t1} & P_{t2} & \dots & P_{tt} \end{pmatrix}$$

- Where the rows sum to less than one
- For transient states i and j , the s_{ij} be the expected number of time periods that the Markov chain is in state j , given that it starts in state i .
- Conditioning on the initial transition, we have for $i, j \in T$,

$$s_{ij} = \sum_{k=1}^t P_{ik} s_{kj}, \text{ if } i \neq j$$

$$s_{ii} = 1 + \sum_{k=1}^t P_{ik} s_{ki},$$

- $s_{kj} = 0$ for recurrent states (because it is impossible to go from a recurrent state to a transient state)

- ▶ Define

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1t} \\ s_{21} & s_{22} & \dots & s_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ s_{t1} & s_{t2} & \dots & s_{tt} \end{pmatrix}$$

- ▶ We have

$$S = I + P_T S,$$

where I is the $t \times t$ identity matrix.

- ▶ Therefore

$$S = (I - P_T)^{-1}.$$

That is, the mean time spent in transient states can be obtained by inverting the matrix $I - P_T$.

Gamblers Ruin

- Consider a Gambler with probability p of winning \$1 and probability $q = 1 - p$ is losing \$1.
 - The game stops when the gambler has N or 0 wealth.
- ▶ Denote by X_n the gambler's wealth at time n . Then $\{X_n, n = 0, 1, 2, \dots\}$ is a Markov chain.
- ▶ The transition probabilities are

$$P_{00} = P_{NN} = 1,$$

$$P_{i,i+1} = p, \quad P_{i,i-1} = q = 1 - p \text{ for } i = 1, 2, \dots, N-1.$$

- ▶ 3 classes of states: $\{0\}, \{N\}, \{1, 2, \dots, N-1\}$
The first two are recurrent and the last is transient.

The probability that a gambler starting with wealth i will eventually reach N is

$$P_i = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, & \text{if } p \neq \frac{1}{2} \\ \frac{i}{N}, & \text{if } p = \frac{1}{2} \end{cases}.$$

Example:

Assume that a player starts with \$3, $N = \$7$ and the probability of winning \$1 is 0.4 and losing \$1 is 0.6.

- What is the probability that they will reach \$7 before going broke?
- Classify the states and give the probability transition matrix for the Markov Chain.
- What is the expected amount of time that starting with i dollars the gambler will have j dollars?

Solution:

a. The required probability is

$$\begin{aligned} P_3 &= \frac{1 - \left(\frac{0.6}{0.4}\right)^3}{1 - \left(\frac{0.6}{0.4}\right)^7} \\ &= \frac{-2.375}{-16.08594} \\ &= 0.14764 \end{aligned}$$

b. The Markov Chain has 3 classes $\{0\}, \{7\}, \{1, 2, 3, 4, 5, 6\}$.

$\{0\}, \{7\}$ are recurrent while $\{1, 2, 3, 4, 5, 6\}$ are transient.

The transition probability matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

c. For the transient states we have

$$P_T = \begin{pmatrix} 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0.6 & 0 \end{pmatrix}$$

We require $S = (I - P_T)^{-1}$.

In general we need the (i, j) element in the S matrix.

Introduction to Simulation

- Many Problems cannot be solved analytically
 - Simulation is used to numerically analyse these problems

Monte Carlo Simulation

- **Problem:**

Suppose $\mathbf{X} = (X_1, \dots, X_n)$ is a random vector with given density function $f(x_1, \dots, x_n)$ and we are interested in computing

$$E[g(\mathbf{X})] = \int \int \cdots \int g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

for some function g which may be impossible to compute.

One way to approximate it is to use Monte Carlo simulation.

- **Method:**

- ▶ Generate a random vector $\mathbf{X}^{(1)} = (X_1^{(1)}, \dots, X_n^{(1)})$ with joint density f and compute $\mathbf{Y}^{(1)} = g(\mathbf{X}^{(1)})$.
- ▶ Generate a second, independent random vector $\mathbf{X}^{(2)} = (X_1^{(2)}, \dots, X_n^{(2)})$ with joint density f and compute $\mathbf{Y}^{(2)} = g(\mathbf{X}^{(2)})$.
- ▶ Repeat this process until you have generated r (a fixed number) i.i.d. random vectors

$$\mathbf{Y}^{(i)} = g(\mathbf{X}^{(i)}) \text{ for } i = 1, 2, \dots, r.$$

- ▶ Use $\frac{\mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(r)}}{r}$ (the average of the generated Y 's) to estimate $E[g(\mathbf{X})]$. By the strong law of large numbers,

$$\lim_{r \rightarrow \infty} \frac{\mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(r)}}{r} = E(\mathbf{Y}^{(i)}) = E[g(\mathbf{X})]$$

Pseudo-Random Numbers

- To generate random vectors from a specified distribution, the first step is to generate random variables from a uniform $(0, 1)$ distribution.
- Generate Pseudo-random numbers:
 - ▶ Start with a seed X_0 and specify positive integers a , c and m .
 - ▶ Use the linear congruential formula to calculate recursively

$$X_{n+1} = (aX_n + c) \pmod{m}$$

(The above modulo function means that the value for X_{n+1} is the remainder after dividing $aX_n + c$ by m .)

- ▶ $\frac{X_n}{m}$ will be an approximation to a uniform $(0, 1)$ random variable (since X_m is one of the numbers $0, 1, \dots, m - 1$).

Example:

Use the algorithm with parameter values $a = 11$, $c = 37$, $m = 100$ and seed $x_0 = 85$ to generate 3 random numbers from the $U(0, 1)$ distribution.

Solution:

$$x_1 = (ax_0 + c) \bmod m = (11 \times 85 + 37) \bmod 100 = 972 \bmod 100 = 72$$

$$x_2 = (ax_1 + c) \bmod m = (11 \times 72 + 37) \bmod 100 = 829 \bmod 100 = 29$$

$$x_3 = (ax_2 + c) \bmod m = (11 \times 29 + 37) \bmod 100 = 356 \bmod 100 = 56$$

These are then divided by m to generate the random numbers, which are: 0.72, 0.29 and 0.56.

Simulating Continuous RVs

- Two methods can be used to simulate continuous random variables:
 - Inverse transform method
 - Acceptance-Rejection Method

Inverse Transform Method

► **Rationale:** Consider a continuous random variable with cumulative distribution function F and a uniform $(0, 1)$ random variable U .

Define $X = F^{-1}(U)$. Then, X will have distribution function

$$\begin{aligned}F_X(x) &= \Pr(X \leq x) = \Pr[F^{-1}(U) \leq x] \\&= \Pr[U \leq F(x)] = F(x).\end{aligned}$$

► **Method:**

- Compute F^{-1} , if this is computable
- Generate a uniform $(0, 1)$ random variable U
- Set $X = F^{-1}(U)$.

► To apply the inverse transform method, the inverse of the distribution function F must have an explicit expression.

Acceptance Rejection Method

- Used when Inverse Transform is not applicable
- In this method, suppose we have a method for simulating a RV with density $g(x)$. Then we can use this as a basis for simulating from the continuous distribution $f(x)$. We simulate a random variable Y with density function $g(y)$ and accept this as the random number with a probability proportional to $\frac{f(Y)}{g(Y)}$

To generate a random variable X that has density function f

Procedure: Let c be some constant such that $\frac{f(y)}{g(y)} \leq c$ for all y .

- Simulate Y with density g .
- Simulate a uniform $(0, 1)$ random number U
- If $U \leq \frac{f(Y)}{cg(Y)}$ then set $X = Y$, otherwise reject and go to 1.

Remarks: The value for X is Y_N where N is the number of iterations until a random number is accepted.

- **Proof:**

$$\begin{aligned}
 \Pr(X \leq x) &= \Pr(Y_N \leq x) = \Pr\left[Y \leq x \mid U \leq \frac{f(Y)}{cg(Y)}\right] \\
 &= \frac{\Pr\left[Y \leq x, U \leq \frac{f(Y)}{cg(Y)}\right]}{\Pr\left[U \leq \frac{f(Y)}{cg(Y)}\right]} = \frac{\int_{-\infty}^x \left[\frac{f(y)}{cg(y)}\right] g(y) dy}{\Pr\left[U \leq \frac{f(Y)}{cg(Y)}\right]} \\
 &= \frac{\int_{-\infty}^x f(y) dy}{c \times \Pr\left[U \leq \frac{f(Y)}{cg(Y)}\right]}
 \end{aligned}$$

For $x \rightarrow \infty$, the LHS is 1 and so is the numerator on the RHS.

Thus, we have

$$\Pr\left[U \leq \frac{f(Y)}{cg(Y)}\right] = \frac{1}{c}$$

so that

$$\Pr[X \leq x] = \int_{-\infty}^x f(y) dy$$

as required.

Example:

Simulating a standard normal random variable Z .

Solution:

Note that the absolute value of Z has density function

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad 0 < x < \infty.$$

Let $g(x) = e^{-x}$, $0 < x < \infty$.

Notice that $\frac{f(x)}{g(x)} = \sqrt{\frac{2e}{\pi}} \exp\{-\frac{(x-1)^2}{2}\} \leq \sqrt{\frac{2e}{\pi}}$. So we can let $c = \sqrt{\frac{2e}{\pi}}$.

1. Generate uniform random numbers U_1 and U_2 .
2. Compute $Y = -\ln U_1$, which is a RV with density function g .
3. If $U_2 \leq e^{-\frac{1}{2}(Y-1)^2}$ (that is, $(Y-1)^2 \leq -2 \ln U_2$) then set $X = Y$ otherwise go to Step 1.

The Box-Muller Approach (Normal Dist. Simulation)

If U_1 and U_2 are independent uniform $(0, 1)$ random numbers then

$$X = (-2 \ln U_1)^{\frac{1}{2}} \cos(2\pi U_2)$$

and

$$Y = (-2 \ln U_1)^{\frac{1}{2}} \sin(2\pi U_2)$$

are a pair of independent standard normal random variables. A pair of uncorrelated normal (μ, σ) random numbers can then be derived using $\mu + \sigma X$ and $\mu + \sigma Y$.

Example:

Simulating a Gamma distribution.

Solution:

Recall that the waiting time to the n th event in the continuous time Markov chains has a Gamma distribution.

In order to simulate from a Gamma distribution with parameters (n, λ) , we use the result that this is the sum of n independent exponential random variables each with rate λ .

The procedure is: Generate n independent uniform $(0, 1)$ random variables U_1, U_2, \dots, U_n , and set $X = -\frac{1}{\lambda} \sum_{i=1}^n \ln(U_i)$.

There are other methods and tricks to speed up the simulation of these random variables.

Simulating Discrete RVs

- ▶ Simulate the discrete random variable with probability mass function

$$\begin{aligned}\Pr(X = x_j) &= p_j, \quad j = 1, 2, \dots \\ \sum_{j=1}^{\infty} p_j &= 1\end{aligned}$$

- ▶ Method:

Let U be uniform $(0, 1)$.

Set

$$X = \begin{cases} x_1 & \text{if } U \leq p_1 \\ x_2 & \text{if } p_1 < U \leq p_1 + p_2 \\ \vdots & \\ x_j & \text{if } \sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^j p_i \end{cases}$$

Example:

Simulating a Geometric distribution.

Solution:

Consider the Geometric random variable X with

$$p_j = \Pr(X = j) = p(1-p)^{j-1} \quad j \geq 1.$$

Notice

$$\begin{aligned}\sum_{i=1}^{j-1} p_i &= 1 - \Pr(X > j-1) = 1 - \sum_{k=j}^{\infty} p(1-p)^{k-1} \\ &= 1 - \frac{p(1-p)^{j-1}}{1 - (1-p)} = 1 - (1-p)^{j-1}.\end{aligned}$$

$1 - (1-p)^{j-1} < U \leq 1 - (1-p)^j$, that is, $(1-p)^j \leq 1 - U < (1-p)^{j-1}$

or equivalently,

$$(1-p)^j \leq U < (1-p)^{j-1}$$

since $1 - U$ is also uniform.

Since

$$\sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^j p_i$$

is equivalent to

$$(1-p)^j \leq U < (1-p)^{j-1},$$

the following technique can be used:

1. Generate a uniform random number U
2. The random number required will be the first integer j for which

$$j \geq \frac{\ln U}{\ln(1-p)}.$$

Example:

Simulating a Poisson distribution.

Solution:

Note that if U is uniform $(0, 1)$ then $-\ln U$ is exponential with rate 1.

This is the interarrival time for a Poisson process with rate 1.

If we select

$$N = \max \left\{ n : \sum_{i=1}^n -\ln U_i < \lambda \right\}$$

then we will have simulated the number of events by time λ for a Poisson process with rate 1. This is the same distribution as a Poisson with mean λ .

Variance Reduction Techniques

Suppose we are interested in computing

$$P = E[g(X_1, X_2, \dots, X_n)].$$

We can use the Monte Carlo simulation to estimate these expected values which we recall as follows:

1. Generate $X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)}$ from the joint distribution of X_1, X_2, \dots, X_n and determine $Y_1 = g(X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)})$
2. Repeat this k times and form $Y_k = g(X_1^{(k)}, X_2^{(k)}, \dots, X_n^{(k)})$
3. Calculate the arithmetic average of the Y 's

$$\bar{Y} = \frac{1}{k} \sum_{i=1}^k Y_i$$

Then:

- \bar{Y} is an estimate of $E[g(X_1, X_2, \dots, X_n)]$ with expectation

$$E(\bar{Y}) = E\left(\frac{1}{k} \sum_{i=1}^k Y_i\right) = E[g(X_1, X_2, \dots, X_n)]$$

since each Y_i is independent and identically distributed with the distribution $g(X_1, X_2, \dots, X_n)$.

- The estimator is an unbiased estimator. Its variance can be computed as

$$\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{k} \sum_{i=1}^k Y_i\right) = \frac{\text{Var}(Y_i)}{k}$$

- Since we do not usually know $\text{Var}(Y_i)$, we estimate it using the sample variance

$$\frac{1}{k-1} \sum_{i=1}^k (Y_i - \bar{Y})^2$$

- There are two methods to reduce variance of a random variable:
 - Antithetic variates
 - Control variates

Antithetic Variates

- Rationale: This method involves generating estimates with negative correlation and then adding these estimates to obtain the final estimate.

Assume that k is even and that $n = \frac{k}{2}$

$$\bar{Y}_1 = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{and} \quad \bar{Y}_2 = \frac{1}{n} \sum_{i=n+1}^k Y_i$$

where the simulated estimates of \bar{Y}_1 and \bar{Y}_2 have correlation ρ .

We then use

$$\bar{Y} = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

as our final estimate.

- Note that:
- Procedure Example:
 1. Generate a set of $n = \frac{k}{2}$ uniform $(0, 1)$ variates U_1, \dots, U_n
 2. Use $X_i = F^{-1}(U_i)$ to determine the estimate $\bar{Y}_1 = \frac{\sum_{i=1}^n g(X_i)}{n}$
 3. Take $1 - U_1, \dots, 1 - U_n$ with $X_i = F^{-1}(1 - U_i)$ to determine the estimate $\bar{Y}_2 = \frac{\sum_{i=1}^n g(X_i)}{n}$.
 4. Use
$$\bar{Y} = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

as our final estimate for $E[Y] = E[g(X)]$.

We will obtain a variance reduction as long as $Y = g(X)$ is a monotone increasing or decreasing function. $)$

$$\begin{aligned} &= \frac{1}{2} \text{Var}(\bar{Y}_1)(1 + \rho) \\ &= \frac{\text{Var}(Y_i)}{k}(1 + \rho) \end{aligned}$$

So if ρ is negative we can gain a variance reduction.

Control Variates

- If we wish to evaluate the expected value, $E[g(X)]$, and there is a function f such that the expected value of $f(X)$ can be evaluated analytically with:

$$E[f(X)] = \mu,$$

then we have:

$$\begin{aligned} E[g(X)] &= E[g(X) + a(\mu - f(X))] \\ &= a\mu + E[g(X) - af(X)] \end{aligned}$$

If we use simulation to estimate

$$E[g(X) - af(X)]$$

This variance is minimized for

$$a = \frac{\text{Cov}[g(X_i), f(X_i)]}{\text{Var}[f(X_i)]}$$

and for this value, we have

$$\text{Var}(\bar{Y}_{cv}) = \frac{\text{Var}[g(X_i)] - \frac{[\text{Cov}[g(X_i), f(X_i)]]^2}{\text{Var}[f(X_i)]}}{k}.$$

* Note: The estimator for $E[g(X)]$

$$\bar{Y}_g = \frac{1}{k} \sum_{i=1}^k g(X_i)$$

has variance

$$\text{Var}(\bar{Y}_g) = \frac{\text{Var}[g(X_i)]}{k}$$

Number of simulations

- We need to carry out simulations as efficiently as possible since the large number of simulations required for accuracy can be quite large and the computing time required can be excessive as a result.
- Assume that we will generate n simulations and use:

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

as an estimate for an expected value of interest. The accuracy of the estimate is given by:

$$Var[\bar{Y}_n] = \frac{Var[Y_i]}{n}$$

- Since we do not know $Var[Y_i]$, we need to use a sample variance from the first k runs where $k < n$ and usually at least 30.
- Our estimate of the variance becomes:

$$Var[\bar{Y}_n] = \frac{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$$

- We can then select n so that our estimate is within a desired accuracy of the true mean (e.g. within 5% of true mean with 95% probability)
- For large values of n we use the CLT to assume the average is approximately normally distributed and solve for n , e.g.

$$Pr[|\mu - \bar{Y}_n| < 0.05\mu] = 0.95$$

Example:

Random variates are generated from a gamma distribution

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

Twenty values are generated for each sample and the mean and standard deviation of each of the samples, to 2 decimal places, are as follows

Sample	1	2	3	4	5	6	7
Mean	12.01	11.79	13.43	14.01	11.44	11.19	11.24
Standard Deviation	6.15	4.73	6.42	7.02	4.30	3.90	3.84

Sample	8	9	10
Mean	12.42	12.91	12.29
Standard deviation	4.35	3.59	4.60

Determining the number of simulated values required in order for the
Solution: $\frac{1}{2}$ of the true

Require n such that



$$Pr(|\bar{X} - \mu| < 0.05\mu) = 0.95$$

This is

$$\begin{aligned} Pr\left(\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} < \frac{0.05\mu}{\frac{\sigma}{\sqrt{n}}}\right) &= 0.95 \\ Pr\left(|Z| < \frac{0.05\mu}{\frac{\sigma}{\sqrt{n}}}\right) &= 0.95 \end{aligned}$$

where Z is standard normal. Thus require n such that

$$\frac{0.05\mu}{\frac{\sigma}{\sqrt{n}}} = 1.96$$

Let i be the sample number. Since all samples have 20 sample values, the estimate of the mean is

$$\bar{X} = \frac{1}{200} \sum_{i=1}^{10} \sum_{j=1}^{20} X_{ij} = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i = 12.273$$

The estimates for the variance is the sample standard deviation

$$\begin{aligned} S_i &= \sqrt{\frac{1}{(20 \times 10) - 1} \left[\sum_{i=1}^{10} \sum_{j=1}^{20} x_{ij}^2 - (20 \times 10) \bar{X}^2 \right]} \\ &= 4.9877 \end{aligned}$$

Now substitute the sample estimates for the mean and standard deviation to get

$$\frac{0.05 \times 12.273}{\frac{4.9877}{\sqrt{n}}} = 1.96 \quad n = 254$$

Exponential Distribution and the Poisson Process

- **Poisson process:** Counting process for the number of events that have occurred up to a particular time
 - It is a “jump” process as it jumps to a higher state every time an event occurs
 - Continuous time Markov Chain
- The time between events/jumps follows an exponential distribution.
- Applications:
 - Individual insurance claims
 - Defaults on loans

The Exponential Distribution

PDF	$f(x) = \lambda e^{-\lambda x}, x \geq 0, \lambda > 0.$
Mean	$E[X] = \frac{1}{\lambda}$
Variance	$Var[X] = \frac{1}{\lambda^2}$
CDF	$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$
MGF	$\phi(t) = E[e^{tx}]$ $= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda$
Survival Function	$S(x) = 1 - F(x)$ $= \begin{cases} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
Hazard/Failure Rate	$\mu(x) = \frac{-\frac{d}{dx}S(x)}{S(x)}$ $= \frac{f(x)}{1 - F(x)}$ $= \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}}$ $= \lambda, \quad x \geq 0$

- No memory Property: X is said to be memoryless if for all $s, t \geq 0$,

$$\Pr(X > s + t | X > t) = \Pr(X > s)$$
- If we think of X as the lifetime of an instrument, then the memoryless property states that if the instrument is alive at time t , then the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution.
- The exponential RV is the only continuous RV that has this property.

Example:

- Suppose that the amount of time a customer will spend in a bank queue is exponentially distributed with mean 10 minutes. What is the probability that a customer will spend more than 15 minutes in the bank? What is the probability that the customer will spend 15 minutes given that he has already spent 10 minutes?

Solution:

$$\Pr(X > 15) = \Pr(X > 10 + 15 | X > 10)$$

- Important Results:

- ▶ Let X_1, X_2, \dots, X_n be n **independent** exponential random variables with rate λ_i , or mean $1/\lambda_i$, for $i = 1, 2, \dots, n$.

- i) If all the rates are equal, i.e. $\lambda_i = \lambda$ for all $i = 1, 2, \dots, n$, then $Y_n = X_1 + X_2 + \dots + X_n$ has a $\text{gamma}(n, \lambda)$ distribution with pdf

$$f_{Y_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

- ii) $X_{(1)} = \min(X_1, X_2, \dots, X_n)$ also has an exponential distribution with rate $\lambda_1 + \lambda_2 + \dots + \lambda_n = \sum_{i=1}^n \lambda_i$, or mean $\frac{1}{\sum_{i=1}^n \lambda_i}$.
- iii) The probability that X_i is the smallest is given by

$$\frac{\lambda_i}{\sum_{j=1}^n \lambda_j}.$$

Convolutions of Exponential Variables

- ▶ Let X_1, X_2, \dots, X_n be n **independent** exponential random variables with rate λ_i or mean $1/\lambda_i$ for $i = 1, 2, \dots, n$. Suppose $\lambda_i \neq \lambda_j$ for $i \neq j$. Then

$$X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

is said to have a **convolution of exponentials** distribution.

To find the form of its density, consider the case where $n = 2$. Note

$$\begin{aligned} f_{X_1+X_2}(y) &= \int_0^y f_{X_1}(s) f_{X_2}(y-s) ds \\ &= \int_0^y \lambda_1 e^{-\lambda_1 s} \lambda_2 e^{-\lambda_2(y-s)} ds \\ &= \lambda_1 \lambda_2 e^{-\lambda_2 y} \int_0^y e^{-(\lambda_1 - \lambda_2)s} ds \\ &= \frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_2 e^{-\lambda_2 y} + \frac{\lambda_2}{\lambda_2 - \lambda_1} \lambda_1 e^{-\lambda_1 y}. \end{aligned}$$

In general, we have

$$f_{X_1+X_2+\dots+X_n}(y) = \sum_{i=1}^n C_{i,n} \lambda_i e^{-\lambda_i y}$$

where

$$C_{i,n} = \left(\prod_{j \neq i}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \right).$$

Counting Processes

- Counting process: A stochastic process $\{N(t), t \geq 0\}$ that represents the number of events that occur up to time t .
- A counting process must satisfy:
 1. $N(t) \geq 0$.
 2. $N(t)$ is integer-valued.
 3. $N(s) \leq N(t)$ for any $s < t$, i.e. it must be non-decreasing.
 4. For $s < t$, $N(t) - N(s)$ is the number of events that have occurred in the interval $(s, t]$.
- A counting process has **independent** increments if the number of events that occur between time s and t is independent of the number of events that occur up to time s .
- A counting process has **stationary** increments if the number of events that occur in any interval of time depends only on the length of the time interval.

The Poisson Process

- **Definition 1:** A counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate λ if:
 1. $N(0) = 0$;
 2. it has independent increments; and
 3. the number of events in any interval of length t has a Poisson distribution with mean λt . That is, for all $s, t \geq 0$,

$$\Pr [N(t+s) - N(s) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \dots$$

- **Definition 2:** A counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate λ , for $\lambda > 0$, if:
 1. $N(0) = 0$;
 2. it has independent increments;
 3. it has stationary increments; and
 4. it satisfies

$$\Pr [N(t+h) - N(t) = 1] = \lambda h + o(h)$$

and

$$\Pr [N(t+h) - N(t) \geq 2] = o(h).$$

A function is said to be $o(h)$ (order h or “little o ” h) if

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$$

- **Note:** The two definitions are equivalent

Interarrival Time

- Let T_n be the time between the $(n - 1)^{th}$ and n^{th} events.
 $\{T_n, n = 1, 2, \dots\}$ are the **inter-arrival times or holding times**.
- The inter-arrival times $T_n, n = 1, 2, \dots$ are independent and identically distributed exponential random variables, each with expected value $\frac{1}{\lambda}$, i.e. $\Pr(T_n > t) = e^{-\lambda t}$.

Proof:

The event $\{T_1 > t\}$ is equivalent to the event that no events occur in the Poisson process during $[0, t]$. Therefore,

$$\Pr(T_1 > t) = \Pr[N(t) = 0] = e^{-\lambda t}.$$

Thus, T_1 has an exponential distribution with expected value $\frac{1}{\lambda}$.

$$\begin{aligned} & \Pr(T_2 > t | T_1 = s) \\ &= \Pr(0 \text{ events in } (s, s+t] | T_1 = s) \\ &= \Pr(0 \text{ events in } (s, s+t]) \text{ (Independence)} \\ &= \Pr(0 \text{ events in } (0, t]) \text{ (Stationary)} \\ &= e^{-\lambda t}. \end{aligned}$$

Thus, T_2 also has an exponential distribution with expected value λ^{-1} . Note also that T_1 and T_2 are independent. This argument may be generalized.

Arrival/Waiting Time

- The arrival/waiting time of the n th event, denoted by S_n is:

$$S_n = T_1 + T_2 + \dots + T_n = \sum_{i=1}^n T_i.$$

- S_n is the sum of n iid exponential RVs, each with the mean $\frac{1}{\lambda}$.
- Hence, S_n has a $\text{gamma}(n, \lambda)$ distribution.
- Therefore:

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geq 0.$$

Example:

A certain scientific theory proposes that mistakes in cell division occur according to a Poisson process with rate 2.5 per year, and that an individual dies when 196 such mistakes have occurred. Assuming this theory, find:

- The mean lifetime of an individual
- The variance of the lifetime of an individual
- The probability that an individual dies before age 67.2
- The probability that an individual reaches age 90

Solution:

Life time of the individual is S_{196} .

We are interested in the distribution of S_{196} defined as the waiting time until the 196th event. As cell division mistakes occur according to a Poisson Process with rate 2.5 per year, we know that

$$S_{196} \sim \text{Gamma}(196, 2.5).$$

Hence, we have

1. Mean:

$$E(S_{196}) = \frac{196}{2.5} = 78.4.$$

2. Variance:

$$\text{Var}(S_{196}) = \frac{196}{2.5^2} = 31.36$$

For 3 and 4, we can use the CLT to approximate these probabilities. Denoting by $Z \sim N(0, 1)$ a standard normal variable.

3.

$$\begin{aligned}\Pr(S_{196} < 67.2) &\approx \Pr\left\{\frac{S_{196} - 78.4}{\sqrt{31.36}} < \frac{67.2 - 78.4}{\sqrt{31.36}}\right\} \\ &= \Pr\{Z < -2\} = 0.0227.\end{aligned}$$

4.

$$\Pr(S_{196} > 90) \approx 0.0192.$$

Conditional Distribution of Arrival Times

- ▶ Let $Y_{(i)}$ be the i th smallest value among Y_1, \dots, Y_n . Then $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ is called the order statistics.
- ▶ If Y_i , for $i = 1, \dots, n$, are independent, identically distributed continuous random variables with probability density $f(y_i)$, then the joint density of the order statistics $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ is

$$f(y_1, y_2, \dots, y_n) = n! \prod_{i=1}^n f(y_i),$$

where $y_1 < y_2 < \dots < y_n$.

- Proof:

(i) $(Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}) = (y_1, y_2, \dots, y_n)$ if (Y_1, \dots, Y_n) is equal to any of the $n!$ permutation of (y_1, y_2, \dots, y_n) ,

ii) the probability density that (Y_1, \dots, Y_n) is equal to $(y_{i_1}, y_{i_2}, \dots, y_{i_n})$ is $\prod_{j=1}^n f(y_{i_j}) = \prod_{j=1}^n f(y_j)$, where i_1, \dots, i_n is a permutation of $1, 2, \dots, n$.

If the Y_i 's, $i = 1, \dots, n$, are uniformly distributed over $(0, t)$ then the joint density of the order statistics is

$$f(y_1, y_2, \dots, y_n) = \frac{n!}{t^n},$$

where $0 < y_1 < y_2 < \dots < y_n < t$.

- Theorem:
 - ▶ Given $N(t) = n$, the n arrival times S_1, \dots, S_n have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on $(0, t)$.
 - ▶ For $0 < s_1 < s_2 < \dots < s_n < t$, noticing that $\{S_1 = s_1, S_2 = s_2, \dots, S_n = s_n, N(t) = n\} = \{T_1 = s_1, T_2 = s_2 - s_1, \dots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n\}$, we get

$$\begin{aligned}
 & f(s_1, s_2, \dots, s_n | N(t) = n) \\
 &= \frac{f(s_1, s_2, \dots, s_n, n)}{\Pr[N(t) = n]} \\
 &= \frac{(\lambda e^{-\lambda s_1}) \dots (\lambda e^{-\lambda(s_n - s_{n-1})}) (e^{-\lambda(t - s_n)})}{e^{-\lambda t} \frac{(\lambda t)^n}{(n)!}} = \frac{n!}{t^n}.
 \end{aligned}$$

Sum of Independent Poisson Processes

Theorem: If $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ are two independent Poisson processes with rates λ_1 and λ_2 respectively, then $\{N(t), t \geq 0\}$ where

$$N(t) = N_1(t) + N_2(t)$$

is also a Poisson process with rate $\lambda_1 + \lambda_2$.

Thinning of Poisson Processes

- Consider a Poisson process with rate λ and suppose that each time an event occurs, it is classified as either a
 - Type I event, with probability p
 - Type II event, with probability $1 - p$
 independently of all other events
- Let $N_1(t)$ and $N_2(t)$ denote respectively the type I and type II events occurring in $[0, t]$. Note that:
 $N(t) = N_1(t) + N_2(t)$
- An important result:
 $\{N_1(t), t \geq 0\}$ is a Poisson process with rate λp and
 $\{N_2(t), t \geq 0\}$ is also a Poisson process with rate $\lambda(1 - p)$.
 The two Poisson processes are also independent.

Example:

Consider an insurance company that has two types of policy: Policy A and Policy B. Total claims from the company arrive according to a Poisson process at the rate of 9 per day. A randomly selected claim has a 1/3 chance that it is of policy A.

1. Calculate the probability that claims from policy A will be fewer than 2 on a given day.
2. Calculate the probability that claims policy B will be fewer than 2 on a given day.
3. Calculate the probability that total claims from the company will be fewer than 2 on a given day.
4. A randomly selected claim from policy A has a 2/3 probability of being over \$10,000 while a randomly selected claim from policy B has probability 2/9 of being over \$10,000.

Generalisations of the Poisson Process

• Non-homogenous Poisson Process

- The counting process $\{N(t), t \geq 0\}$ is a *non-homogeneous Poisson process* with intensity function $\lambda(t)$, $t \geq 0$, if
 1. $N(0) = 0$;
 2. it has independent increments; and
 3. it has unit jumps i.e.

$$\Pr[N(t+h) - N(t) = 1] = \lambda(t)h + o(h)$$

and

$$\Pr[N(t+h) - N(t) \geq 2] = o(h)$$

- The *mean value function* of a non-homogeneous Poisson process is

$$m(t) = \int_0^t \lambda(y) dy$$

- $N(s+t) - N(s)$ is a Poisson random variable with mean $m(s+t) - m(s)$.
- $N(t)$ is a Poisson random variable with mean $m(t)$.
- If $\{N(t), t \geq 0\}$ is a non-homogeneous Poisson process with mean value function $m(t)$ then $\{N(m^{-1}(t)), t \geq 0\}$ is a homogeneous Poisson process with intensity $\lambda = 1$.

• Compound Poisson Process

- A stochastic process $\{X(t), t \geq 0\}$ is a **compound Poisson**
- Mean and Variance:

$$\begin{aligned} E[X(t)] &= E\{E[X(t)|N(t)]\} \\ &= E\left\{E\left[\sum_{i=1}^{N(t)} Y_i | N(t)\right]\right\} \\ &= E[N(t)E[Y_1]] = \lambda t E[Y_1] \end{aligned}$$

$$\begin{aligned} \text{Var}[X(t)] &= E\{\text{Var}[X(t)|N(t)]\} + \text{Var}\{E[X(t)|N(t)]\} \\ &= E[N(t)\text{Var}(Y_1)] + \text{Var}(N(t)E[Y_1]) \\ &= \lambda t \text{Var}(Y_1) + (E[Y_1])^2 \lambda t \\ &= \lambda t (\text{Var}(Y_1) + (E[Y_1])^2) \\ &= \lambda t E[Y_1^2]. \end{aligned}$$

Example:

Lucky Jack finds coins on the way to work according to a Poisson process with a mean of 22 coins per month.

5% of the time, Jack finds coins worth a total of 10.

15% of the time, Jack finds coins worth a total of 5.

80% of the time, Jack finds coins worth a total of 1.

The amounts of coins he finds are independent, and are independent of the number of coins. Calculate the variance of the total amount he finds in a month.

Solution:

Let $N(t)$ denote the number of coins he finds up to time t . Then $N(t)$ is a Poisson process with rate 22 per month.

Let Y_i denote the amount of the i th coin he finds. Then total amount equals $\sum_{i=1}^{N(1)} Y_i$.

So the required variance equals

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^{N(1)} Y_i\right) &= E\left[\text{Var}\left(\sum_{i=1}^{N(1)} Y_i \mid N(1)\right)\right] + \text{Var}\left(E\left[\sum_{i=1}^{N(1)} Y_i \mid N(1)\right]\right) \\ &= E[N(1)\text{Var}(Y)] + \text{Var}(N(1)E[Y]) \\ &= 22\text{Var}(Y) + 22(E[Y])^2 = 22E[Y^2] \\ &= 22(10^2 \times 0.05 + 5^2 \times 0.15 + 1^2 \times 0.8) = 210.1. \end{aligned}$$

Continuous-Time Markov Chain

- A continuous time Markov chain is a Markov process in continuous-time with discrete state space.
 - AKA Markov Jump Process
- The process $\{X(t), t \geq 0\}$ is a continuous-time Markov chain if for all $s, t \geq 0$ and non-negative integers i, j, k for $0 \leq u \leq s$,

$$\begin{aligned} &\Pr[X(t+s) = j | X(s) = i, X(u) = k, 0 \leq u < s] \\ &= \Pr[X(t+s) = j | X(s) = i]. \end{aligned}$$
- The process has stationary or homogeneous transition probabilities if:

$$\Pr[X(t+s) = j | X(s) = i] = P_{ij}(t)$$
 is independent of s

- **Example:** The Poisson process is a continuous-time Markov chain with

$$\begin{aligned}
& \Pr [X(t+s) = j | X(s) = i] \\
&= \Pr [j - i \text{ jumps in } (s, s+t) | i \text{ jumps in } [0, s]] \\
&= \Pr [j - i \text{ jumps in } (s, s+t)] \\
&= \Pr [j - i \text{ jumps in } (0, t)] \\
&= e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \quad j = i, i+1, \dots
\end{aligned}$$

Time Spent in a State

- For a stationary, continuous-time Markov chain, T_i , the time spent on state i is a memoryless random variable:

$$\begin{aligned}
& \Pr(T_i - s \geq t | T_i \geq s) \\
&= \Pr(T_i \geq t + s | T_i \geq s) \\
&= \Pr[X(v) = i, s < v \leq t + s | X(u) = i, 0 \leq u \leq s] \\
&= \Pr[X(v) = i, s < v \leq t + s | X(s) = i] \\
&= \Pr[X(v) = i, 0 < v \leq t | X(0) = i] \\
&= \Pr(T_i \geq t | T_i \geq 0) \\
&= \Pr(T_i \geq t)
\end{aligned}$$

- The time spent in each state is exponentially distributed
 - In other words,

$$T_i \sim \text{Exponential}(\nu_i),$$

where ν_i is the rate at which the process makes a transition when in state i (i.e. the transition rate out of state i).

So,

$$E(T_i) = \frac{1}{\nu_i}.$$

- The amount of time spent in state i is independent of the next state visited (Markov property)

Transition Probability

- The **transition probability** of the continuous-time Markov chain, $P_{ij}(s, s+t)$, is defined by

$$P_{ij}(s, s+t) = \Pr[X(s+t) = j | X(s) = i].$$

- For a homogeneous or stationary process,

$$\begin{aligned}
P_{ij}(s, s+t) &= \Pr[X(s+t) = j | X(s) = i] \\
&= \Pr[X(t) = j | X(0) = i] = P_{ij}(t).
\end{aligned}$$

- **Notes:**

$$P_{ij}(s, s) = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$$

For a homogeneous continuous-time Markov chain,

$$P_{ij}(0) = P_{ij}(s, s) = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}.$$

Transition Rate

- Let q_{ij} denote the (**instantaneous**) **transition rate** from state i to state j of a continuous-time Markov process, where

$$q_{ij} = \lim_{h \rightarrow 0} \frac{P_{ij}(h) - P_{ij}(0)}{h} = \begin{cases} \lim_{h \rightarrow 0} \frac{P_{ij}(h)}{h}, & \text{if } i \neq j \\ \lim_{h \rightarrow 0} \frac{P_{ij}(h) - 1}{h}, & \text{if } i = j \end{cases}.$$

- For the non-homogeneous process, the (**instantaneous**) **transition rate** from state i to state j at time s is

$$\begin{aligned} q_{ij}(s) &= \lim_{h \rightarrow 0} \frac{P_{ij}(s, s+h) - P_{ij}(s, s)}{h} \\ &= \begin{cases} \lim_{h \rightarrow 0} \frac{P_{ij}(s, s+h)}{h}, & \text{if } i \neq j \\ \lim_{h \rightarrow 0} \frac{P_{ij}(s, s+h) - 1}{h}, & \text{if } i = j \end{cases}. \end{aligned}$$

- For a **homogeneous** continuous-time Markov chain:

►

$$\begin{aligned} \left[\frac{\partial}{\partial t} P_{ij}(t) \right]_{t=0} &= \lim_{h \rightarrow 0} \frac{P_{ij}(h) - P_{ij}(0)}{h} \\ &= \begin{cases} q_{ij}, & \text{if } i \neq j \\ q_{ii} = -\nu_i, & \text{if } i = j \end{cases} \end{aligned}$$

► Therefore,

$$P_{ij}(h) = \begin{cases} q_{ij}h + o(h), & \text{if } i \neq j \\ 1 + q_{ii}h + o(h), & \text{if } i = j \end{cases}$$

for small $h \geq 0$.

- For a **non-homogeneous** CTMC:

►

$$\begin{aligned} \left[\frac{\partial}{\partial t} P_{ij}(s, t) \right]_{t=s} &= \lim_{h \rightarrow 0} \frac{P_{ij}(s, s+h) - P_{ij}(s, s)}{h} \\ &= \begin{cases} q_{ij}(s), & \text{if } i \neq j \\ q_{ii}(s) = -\nu_i(s), & \text{if } i = j \end{cases}. \end{aligned}$$

► Therefore,

$$P_{ij}(s, s+h) = \begin{cases} q_{ij}(s)h + o(h), & \text{if } i \neq j \\ 1 + q_{ii}(s)h + o(h), & \text{if } i = j \end{cases},$$

for small $h \geq 0$.

Transition Rate and Probability

- Let ν_i denote the transition rate out of state i when the process is in state i .
- Let P_{ij} denote the probability that given the continuous-time Markov chain is in state i and a transition occurs, the transition is into state j .

- The rate of transition of the process into state j , when the process is in state i is

$$q_{ij} = \nu_i P_{ij} \quad \text{for } j \neq i.$$

- We have

$$\begin{aligned}\nu_i &= \sum_{j \neq i} q_{ij} = \sum_{j \neq i} \nu_i P_{ij}, \\ q_{ii} &= -\nu_i = -\sum_{j \neq i} q_{ij}.\end{aligned}$$

Therefore,

$$P_{ij} = \frac{q_{ij}}{\nu_i} = \frac{q_{ij}}{\sum_{j \neq i} q_{ij}}.$$

Chapman-Kolmogorov Equations

- For a homogeneous CTMC:
 - Chapman-Kolmogorov equations

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s)$$

- Kolmogorov's backward equations

$$\frac{\partial}{\partial t} P_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t)$$

- Kolmogorov's forward equations

$$\frac{\partial}{\partial t} P_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t)$$

with initial conditions $P_{ij}(0) = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$

- CK Equations (matrix version):

- Define the matrices

$$\begin{aligned}\mathbf{P}(t) &= [P_{ij}(t)], \\ \mathbf{Q} &= [q_{ij}], \text{ and} \\ \frac{\partial}{\partial t} \mathbf{P}(t) &= \left[\frac{\partial}{\partial t} P_{ij}(t) \right].\end{aligned}$$

- Chapman-Kolmogorov equations

$$\mathbf{P}(t+s) = \mathbf{P}(t) \mathbf{P}(s), \quad t \geq 0$$

- Kolmogorov's backward equations

$$\frac{\partial}{\partial t} \mathbf{P}(t) = \mathbf{Q} \mathbf{P}(t), \quad t \geq 0$$

with initial condition $\mathbf{P}(0) = I$.

- Kolmogorov's forward equations

$$\frac{\partial}{\partial t} \mathbf{P}(t) = \mathbf{P}(t) \mathbf{Q}, \quad t \geq 0$$

with initial condition $\mathbf{P}(0) = I$.

- Remark:

- ▶ Since we have

$$q_{ii} = -\nu_i,$$

we can write

$$\mathbf{Q} = \begin{pmatrix} -\nu_0 & q_{01} & q_{02} & \dots \\ q_{10} & -\nu_1 & q_{12} & \dots \\ q_{20} & q_{21} & -\nu_2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

- ▶ The Kolmogorov's backward and forward equations are essentially two forms of differential equations involving the transition probabilities. So, given the transition rates q_{ij} , which can be estimated from data, we can determine the transition probabilities $P_{ij}(t)$ by solving these differential equations.

Limiting Probabilities

- Sufficient conditions must be met for limiting probabilities to exist:
 - All states must communicate
 - The Markov chain is positive recurrent so that starting in any state the mean time to return to that state is finite.
- Assuming the limiting probabilities exist, these limits are independent of initial state.
 - So we can define:

$$P_j \equiv \lim_{t \rightarrow \infty} P_{ij}(t)$$

- The limiting probabilities satisfy:

$$\nu_j P_j = \sum_{k \neq j} q_{kj} P_k \text{ for all states } j, \text{ and } \sum_k P_k = 1$$

- When the limiting probabilities exist, the Markov chain is said to be **ergodic**. If the initial state is chosen according to distribution P_i , (i.e. $\Pr(X(0) = i) = P_i$), then $\Pr(X(t) = i) = P_i$.
- P_j is also interpreted as the long run proportion of time that the process is in state j .
- ▶ **Remarks:** The equation $\nu_j P_j = \sum_{k \neq j} q_{kj} P_k$ implies that the rate at which the process leaves state j = the rate at which the process enters state j , because

- ▶ $\nu_j P_j$ = the rate at which the process leaves state j
since P_j is the proportion of time the process is in state j and
 ν_j is the rate of transition out of state j when the process is in state j .
- ▶ $\sum_{k \neq j} q_{kj} P_k$ = the rate at which the process enters state j
since P_k is the proportion of the time the process is in state k and q_{kj} is the rate of transition from state k to state j .

Embedded Markov Chain

- For a CTMC that is ergodic with limiting probabilities, we can consider the sequence of states visited as a discrete-time Markov chain.
 - This DTMC is called the **embedded Markov chain**.
- The embedded Markov chain has transition probabilities P_{ij} (from the CTMC).

- Assume this embedded MC is ergodic and the limiting probabilities are solutions of:

$$\begin{aligned}\pi_i &= \sum_j \pi_j P_{ji} \\ \sum_i \pi_i &= 1.\end{aligned}$$

- The proportion of time the CTMC spends in state i is given by:

$$P_i = \frac{(\pi_i / \nu_i)}{\sum_j (\pi_j / \nu_j)},$$

since π_i is the proportion of the number of transitions into state i and $\frac{1}{\nu_i}$ is the mean time spent in state i during a visit.

Birth and Death Process

- A birth and death process is a CTMC with states $\{0, 1, 2, \dots\}$ for which transitions from state n may go only to either state $n + 1$ (a birth) or $n - 1$ (a death).
- Suppose that if the number of people in a population (or any system) is n
 - ▶ new arrivals enter the population/system at rate λ_n (the arrival or birth rate) with the time until the next arrival exponentially distributed with mean $\frac{1}{\lambda_n}$, and
 - ▶ people leave the population/system at rate μ_n (the departure or death rate) with the time until the next departure exponentially distributed with mean $\frac{1}{\mu_n}$ and independent of the next arrival.
- For the above Birth and Death process we have:

$$\nu_0 = \lambda_0$$

$$\nu_i = \lambda_i + \mu_i, \quad i > 0.$$

- For the corresponding embedded DTMC, the transition probabilities for state 0 will be:

$$P_{01} = 1,$$

since the number in the system cannot be negative and so a birth must be the first jump out of state 0.

- Given that the population is i :

- ▶ the next jump will be from i to $i + 1$ if a birth occurs before a death and from i to $i - 1$ if a death occurs before a birth;
- ▶ the time to a birth, T_b , is exponential with rate λ_i , the time to a death, T_d , is exponential with rate μ_i , and they are independent;

$$\Pr [T_b < T_d] = \frac{\lambda_i}{\lambda_i + \mu_i}.$$

- Hence,
- Recall from the results for the exponential distribution that:
 - ▶ if (X_1, X_2, \dots, X_n) are n independent exponential random variables, X_i , each with different rate λ_i
 - ▶ then the probability that X_i is the smallest is $\frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$.
$$P_{i,i-1} = 1 - P_{i,i+1} = \frac{\mu_i}{\lambda_i + \mu_i}.$$

$$\nu_i = \lambda_i + \mu_i.$$
- Recall also that the time to the next jump, regardless of whether it is a birth or a death, is exponential with rate:

Example:

- Birth and death rates independent of n and an example of a **queue**:
 - ▶ A bank (supermarket) has one server and customers join a queue when they arrive.
 - ▶ Customers arrive at rate λ with inter-arrival times exponentially distributed (a Poisson process). The server serves customers at the rate μ with service times exponentially distributed.
 - ▶ If $X(t)$ is the number of customers in the queue at time t , then $\{X(t); t > 0\}$ is a birth and death process with

$$\begin{aligned}\lambda_n &= \lambda \text{ for } n \geq 0 \text{ and} \\ \mu_n &= \mu \text{ for } n \geq 1.\end{aligned}$$

- In this example:
 - ▶ if there are now s tellers (checkout operators) serving the queue in the bank (supermarket) with customers joining a single queue and going to the first available server,
 - ▶ then we have a birth and death process with

$$\begin{aligned}\lambda_n &= \lambda \text{ for } n \geq 0 \text{ and} \\ \mu_n &= \begin{cases} n\mu, & \text{for } 1 \leq n < s \\ s\mu, & \text{for } n > s. \end{cases}\end{aligned}$$

Example:

- **Population Growth**
Each individual in a population gives birth at an exponential rate λ plus there is immigration at an exponential rate θ . Each individual has an exponential death rate μ .
The population size at time t is a birth and death process with

$$\begin{aligned}\lambda_n &= n\lambda + \theta \text{ for } n \geq 0 \text{ and} \\ \mu_n &= n\mu \text{ for } n \geq 1.\end{aligned}$$

Example:

- **Poisson Process**

- The Poisson process is a Birth and Death process with no deaths.
- $$\lambda_n = \lambda \text{ for } n \geq 0 \text{ and}$$
- $$\mu_n = 0 \text{ for } n \geq 1.$$

Expected Time in States (Birth and Death Process)

- Recursive relationships are used to compute the expected times for going from state to another in the birth and death process.
 - ▶ Let V_i be the time for the process starting from state i to enter state $i + 1$ (a birth).
 - ▶ For $i > 0$ let

$$l_i = \begin{cases} 1, & \text{if first transition is } i \rightarrow i + 1 \\ 0, & \text{if first transition is } i \rightarrow i - 1 \end{cases}.$$

- ▶ Conditioning on the first transition,

$$\begin{aligned} E(V_i) &= E(V_i | l_i = 1) \Pr[l_i = 1] + E(V_i | l_i = 0) \Pr[l_i = 0] \\ &= \frac{1}{\lambda_i + \mu_i} \frac{\lambda_i}{\lambda_i + \mu_i} + \left[\frac{1}{\lambda_i + \mu_i} + E(V_{i-1}) + E(V_i) \right] \frac{\mu_i}{\lambda_i + \mu_i} \end{aligned}$$

since the time to the first transition, independent of whether it is a birth or death, is an exponential variable with rate $\lambda_i + \mu_i$, and the time taken to get to $i + 1$ if there is a death is the time taken to get from $i - 1$ to i and then from i to $i + 1$.

- ▶ Simplifying the above relationship, we get

$$E(V_i) = \frac{1}{\lambda_i + \mu_i} + [E(V_{i-1}) + E(V_i)] \frac{\mu_i}{\lambda_i + \mu_i}$$

- ▶ and solving for $E(V_i)$ to get

$$\begin{aligned} E(V_i) &= \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E(V_{i-1}) \text{ for } i \geq 1. \\ &= \frac{1}{\lambda_i} + \sum_{j=1}^n \frac{1}{\lambda_{j-1}} \cdot \prod_{k=j}^n \frac{\mu_k}{\lambda_k} \end{aligned}$$

Transition Probabilities (Birth and Death Process)

- The transition rates:
 - $q_{01} = \lambda_0$, $q_{00} = -v_0 = -\lambda_0$, and $q_{0i} = 0$ for all $i > 1$
 - and for states $i > 0$, we have:
 - $q_{i,i+1} = \lambda_i$, $q_{i,i-1} = \mu_i$, and $q_{i,i} = -v_i = -(\lambda_i + \mu_i)$
 - and zero otherwise

- The backwards equations are:

$$\frac{\partial}{\partial t} P_{0,j}(t) = \lambda_0 P_{1,j}(t) - \lambda_0 P_{0,j}(t) \text{ and}$$

$$\begin{aligned} \frac{\partial}{\partial t} P_{i,j}(t) &= \lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t) \\ &\quad - (\lambda_i + \mu_i) P_{i,j}(t) \end{aligned}$$

- The forward equations are:

$$\frac{\partial}{\partial t} P_{i,0}(t) = \mu_1 P_{i,1}(t) - \lambda_0 P_{i,0}(t) \text{ and}$$

$$\begin{aligned} \frac{\partial}{\partial t} P_{i,j}(t) &= \lambda_{j-1} P_{i,j-1}(t) + \mu_{j+1} P_{i,j+1}(t) \\ &\quad - (\lambda_j + \mu_j) P_{i,j}(t) \end{aligned}$$

- For the homogeneous case, the backward and forward equations are effectively equivalent.

Limiting Probabilities (Birth and Death Process)

- The limiting probabilities are determined by the *balance equations*. That is:

$$\nu_j P_j = \sum_{k \neq j} q_{kj} P_k \quad \text{for all states } j$$

$$\sum_k P_k = 1$$

- This simply says that the long-run probabilities, P_k , are determined such that the rate at which the process leaves state j is equated to the rate at which the process enters state j .

- For a birth and death process, the balance equations are:

State Rate leaves = rate enters

$$0 \quad \lambda_0 P_0 = \mu_1 P_1$$

$$1 \quad (\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \lambda_0 P_0$$

$$n, n \geq 1 \quad (\lambda_n + \mu_n) P_n = \mu_{n+1} P_{n+1} + \lambda_{n-1} P_{n-1}.$$

- Working forward from state 0 and substituting gives:

$$(\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \lambda_0 P_0$$

$$(\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \mu_1 P_1$$

$$\lambda_1 P_1 = \mu_2 P_2$$

and in general:

$$\lambda_n P_n = \mu_{n+1} P_{n+1} \text{ for } n \geq 0.$$

- Expressing the long-run probabilities in terms of P_0 to get:

$$\lambda_0 P_0 = \mu_1 P_1 \quad \text{or}$$

$$P_1 = \frac{\lambda_0}{\mu_1} P_0$$

$$P_2 = \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1}{\mu_2} \left(\frac{\lambda_0}{\mu_1} P_0 \right).$$

in general:

$$P_n = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} P_0, \quad n \geq 1.$$

- Substituting the above into:

$$\sum_{k=0}^{\infty} P_k = 1$$

we get

$$P_0 + P_0 \sum_{n=1}^{\infty} \left(\frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} \right) = 1$$

or

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \left(\frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} \right)}.$$

- Hence we have:

$$P_n = \frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} P_0 = \frac{\frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1}}{1 + \sum_{n=1}^{\infty} \left(\frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} \right)},$$

for $n \geq 1$.

- This also gives the condition for the long-run probabilities to exist, namely:

$$\sum_{n=1}^{\infty} \left(\frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2} \frac{\lambda_0}{\mu_1} \right) < \infty.$$

Example: (A pure birth process)

- The Poisson process is an example of a pure birth process with:

$$\lambda_n = \lambda \text{ for } n \geq 0$$

$$\mu_n = 0 \text{ for } n \geq 0$$

- The Kolmogorov backward equations are:

By noting that $P_{i,j}(0) = 0$ for $i \neq j$, we obtain

$$\begin{aligned} P_{i,j}(s) &= e^{-\lambda s} \int_0^s e^{\lambda t} \lambda P_{i+1,j}(t) dt \\ &= \int_0^s e^{-\lambda(s-t)} \lambda P_{i+1,j}(t) dt. \end{aligned}$$

$$\frac{\partial}{\partial s} P_{i,j}(s) = \lambda P_{i+1,j}(s) - \lambda P_{i,j}(s), \quad i > 0.$$

$$P_{i,j}(s) = e^{-\lambda s} \frac{(\lambda s)^{j-i}}{(j-i)!}, \quad j > i.$$

$$\frac{\partial}{\partial t} P_{i,j}(t) + \lambda P_{i,j}(t) = \lambda P_{i+1,j}(t)$$

Now multiply both sides by $e^{\lambda t}$ to get

$$e^{\lambda t} \left[\frac{\partial}{\partial t} P_{i,j}(t) + \lambda P_{i,j}(t) \right] = e^{\lambda t} \lambda P_{i+1,j}(t)$$

which is equivalent to

$$\frac{\partial}{\partial t} [e^{\lambda t} P_{i,j}(t)] = e^{\lambda t} \lambda P_{i+1,j}(t).$$

Now integrate both sides with respect to t from 0 to s to get

$$\begin{aligned} \int_0^s \frac{\partial}{\partial t} [e^{\lambda t} P_{i,j}(t)] dt &= \int_0^s e^{\lambda t} \lambda P_{i+1,j}(t) dt \\ e^{\lambda s} P_{i,j}(s) - e^{\lambda 0} P_{i,j}(0) &= \int_0^s e^{\lambda t} \lambda P_{i+1,j}(t) dt. \end{aligned}$$

- The solution of the integral equation is:

Example: (A Simple Sickness Model)

- In a simple sickness model, an individual can be in state 0 (healthy) or state 1 (sick).
- Assume that individuals remain healthy for an exponential time with mean $\frac{1}{\lambda}$ before becoming sick, and that it takes an exponential time to recover from sick to healthy again with mean time $\frac{1}{\mu}$.
- It is a birth death process with $\lambda_0 = \lambda$ and $\mu_1 = \mu$
and all other λ_i, μ_i are zero.
- Using the Kolmogorov equations, it can be shown that the transition probabilities are:

$$P_{0,0}(s) = \frac{\mu}{(\mu + \lambda)} + \frac{\lambda}{(\mu + \lambda)} e^{-(\mu + \lambda)s},$$

$$P_{1,0}(s) = \frac{\mu}{(\mu + \lambda)} \left[1 - e^{-(\mu + \lambda)s} \right],$$

$$P_{0,1}(s) = \frac{\lambda}{(\mu + \lambda)} \left[1 - e^{-(\mu + \lambda)s} \right], \text{ and}$$

$$P_{1,1}(s) = \frac{\lambda}{(\mu + \lambda)} + \frac{\mu}{(\mu + \lambda)} e^{-(\mu + \lambda)s}.$$

- We are mainly interested in the probability that a healthy individual will stay healthy for a period of s , or a sick individual will remain sick throughout the same period. These notations are:

$$P_{\overline{0,0}}(s) = e^{-\lambda s} \text{ and } P_{\overline{1,1}}(s) = e^{-\mu s}.$$

- These are called “*Occupancy probabilities*”
- Define the occupation time, $O_i(t)$, as the total time that the process spends in state i during the interval $(0, t)$.
- The total time spent sick during the time interval $(0, t)$ is

$$O_1(t) = \int_0^t I_1(s) ds$$

where

$$I_1(s) = \begin{cases} 1, & \text{if } X(s) = 1 \\ 0, & \text{if } X(s) = 0 \end{cases}$$

- The expected total time spent sick during the time interval $(0, t)$ given that the initial state is healthy:

$$\begin{aligned} E[O_1(t)|X(0)=0] &= E\left[\int_0^t I_1(s) ds | X(0)=0\right] \\ &= \frac{\lambda}{(\mu+\lambda)}t + \frac{\lambda}{(\mu+\lambda)^2} [e^{-(\mu+\lambda)t} - 1]. \end{aligned}$$

Non-homogeneous Markov Jump Process

- A Markov Jump Process is Non-homogeneous if: the transition rates and probabilities are dependent on the **current time (s)**
- The transition rates are denoted by $\sigma_{ij}(s)$

$$\sigma_{ij}(s) = \left[\frac{\partial}{\partial t} P_{ij}(s, t) \right]_{t=s}$$

►

$$P_{ij}(s, s+h) = \begin{cases} h\sigma_{ij}(s) + o(h), & \text{if } i \neq j \\ 1 + h\sigma_{ii}(s) + o(h), & \text{if } i = j \end{cases}$$

and

$$\sigma_{ii}(s) = - \sum_{i \neq j} \sigma_{ij}(s).$$

- Note

$$P_{ij}(s-h, s) = \begin{cases} h\sigma_{ij}(s-h) + o(h), & \text{if } i \neq j \\ 1 + h\sigma_{ii}(s-h) + o(h), & \text{if } i = j \end{cases}$$

So

$$\sigma_{ij}(s-h) = \frac{P_{ij}(s-h, s) - P_{ij}(s, s) - o(h)}{h}$$

Letting $h \rightarrow 0$ we get

$$\begin{aligned} \sigma_{ij}(s) &= - \lim_{h \rightarrow 0} \frac{P_{ij}(s-h, s) - P_{ij}(s, s) - o(h)}{-h} \\ &= - \left[\frac{\partial}{\partial s} P_{ij}(s, t) \right]_{t=s}. \end{aligned}$$

- The Chapman-Kolmogorov equations:

- Kolmogorov Backward Equations:

$$\begin{aligned}\frac{\partial}{\partial s} P_{ij}(s, t) &= \sum_k \left[\frac{\partial}{\partial s} P_{ik}(s, u) \right]_{u=s} P_{kj}(s, t) \\ &= \sum_k -\sigma_{ik}(s) P_{kj}(s, t).\end{aligned}$$

- In matrix notation,

$$\frac{\partial}{\partial s} \mathbf{P}(s, t) = -\mathbf{A}(s) \mathbf{P}(s, t).$$

- The forward and backward equations cannot be solved in the same way as for the homogeneous case. These are linear differential equations with time-dependent coefficients.
 - Need to use numerical methods to solve these equations

First Holding Time (Time Homogeneous Case)

- Define the **first holding time** $T_0 = \inf \{t : X_t \neq X_0\}$, that is, the first time the process jumps out of the initial state.
- Given that the initial state is i , T_0 is exponentially distributed with parameter $\lambda_i := -\sigma_{ii} = \sum_{j \neq i} \sigma_{ij}$. Thus,

$$\Pr(T_0 > t | X_0 = i) = e^{-\lambda_i t}$$

- The probability density of the **first holding time**, given that the initial state is i , is

$$\lambda_i e^{-\lambda_i t}.$$

- The probability of the state to which the process jumps from $X_0 = i$ to $X_{T_0} = j$ is

$$\Pr(X_{T_0} = j | X_0 = i) = P_{ij} = \frac{\sigma_{ij}}{\lambda_i}, \quad j \neq i$$

and X_{T_0} is independent of T_0 .

- In matrix notation: $\frac{\partial}{\partial t} \mathbf{P}(s, t) = \mathbf{P}(s, t) \mathbf{A}(t)$, where

$$\mathbf{A}(t) = \begin{pmatrix} \sigma_{00}(t) & \sigma_{01}(t) & \dots & \sigma_{0n}(t) \\ \sigma_{10}(t) & \sigma_{11}(t) & \dots & \sigma_{1n}(t) \\ \vdots & & \ddots & \\ \sigma_{n0}(t) & \dots & \dots & \sigma_{nn}(t) \end{pmatrix}.$$

Residual Holding Time (Time Inhomogeneous Case)

- ▶ Define the **residual holding time** R_s to be the time between s and the NEXT jump:

$$\{R_s > w, X_s = i\} = \{X_u = i, s \leq u \leq s + w\}$$

i.e. the state remains the same as the state at time s (state i) between s and $s + w$.

- ▶ $\Pr(R_s > w | X_s = i) = e^{- \int_s^{s+w} \lambda_i(u) du},$

where $\lambda_i(u) = -\sigma_{ii}(u)$.

- ▶ The density of $R_s | X_s = i$ at w is given by

$$\lambda_i(s + w) e^{- \int_s^{s+w} \lambda_i(u) du}.$$

- ▶ Define $X_s^{(+)} = X_{s+R_s}$, the state the process jumps to at the next jump.

$$\Pr(X_s^{(+)} = j | X_s = i, R_s = w) = \frac{\sigma_{ij}(s + w)}{\lambda_i(s + w)}.$$

- ▶ An integrated form of the backward equations

$$P_{ij}(s, t) = \Pr(X_t = j | X_s = i) \quad j \neq i \\ = \sum_{l \neq i} \int_0^{t-s} \lambda_l(s + w) e^{- \int_s^{s+w} \lambda_l(u) du} \frac{\sigma_{il}(s + w)}{\lambda_l(s + w)} P_{lj}(s + w, t) dw$$

The process first jumps out of current state i at time $s + w$ (probability $\lambda_i(s + w) e^{- \int_s^{s+w} \lambda_i(u) du}$), the process jumps to state l (probability $\frac{\sigma_{il}(s + w)}{\lambda_i(s + w)}$), then it jumps from state l at time $s + w$ to state j at time t with probability $P_{lj}(s + w, t)$. Sum and integrate over all possible times when the first jump can occur (any time between s and t) and all possible states the jump can be to (any state except i).

- ▶ Define C_t to be the **current holding time**, which is the time between the last jump and t .

$$\{C_t \geq w, X_t = j\} = \{X_u = j, t - w \leq u \leq t\}.$$

$$\Pr(C_t \geq w | X_t = j) = e^{- \int_{t-w}^t \lambda_j(u) du}$$

- ▶ An integrated form of the forward equations

$$P_{ij}(s, t) = \sum_{k \neq j} \int_0^{t-s} P_{ik}(s, t - w) \sigma_{kj}(t - w) e^{- \int_{t-w}^t \lambda_j(u) du} dw,$$

for $i \neq j$.

The Survival Model

- Consider the basic survival model with states A (alive) and D (dead), with the transition intensity matrix:
- $$\mathbf{A}(t) = \begin{pmatrix} \sigma_{AA}(t) = -\mu(t) & \sigma_{AD}(t) = \mu(t) \\ \sigma_{DA}(t) = 0 & \sigma_{DD}(t) = 0 \end{pmatrix}$$
- where $\mu(t)$ is called the force of mortality

- The Kolmogorov Equations:

$$\begin{aligned} & \left(\begin{array}{cc} \frac{\partial}{\partial t} P_{AA}(s, t) & \frac{\partial}{\partial t} P_{AD}(s, t) \\ \frac{\partial}{\partial t} P_{DA}(s, t) & \frac{\partial}{\partial t} P_{DD}(s, t) \end{array} \right) \\ &= \left(\begin{array}{cc} P_{AA}(s, t) & P_{AD}(s, t) \\ P_{DA}(s, t) & P_{DD}(s, t) \end{array} \right) \left(\begin{array}{cc} -\mu(t) & \mu(t) \\ 0 & 0 \end{array} \right). \end{aligned}$$

- The initial conditions at time $t = s$:

$$\left(\begin{array}{cc} P_{AA}(s, s) & P_{AD}(s, s) \\ P_{DA}(s, s) & P_{DD}(s, s) \end{array} \right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right)$$

Example:

Find the probability that an individual is alive at time t given they are alive at time s .

Solution:

The desired probability is $P_{AA}(s, t)$.

Solving

$$\frac{\partial}{\partial t} P_{AA}(s, t) = -\mu(t)P_{AA}(s, t),$$

with initial condition $P_{AA}(s, s) = 1$, we can obtain

$$P_{AA}(s, t) = e^{-\int_s^t \mu(x)dx}.$$

The Death Sickness Model

- Consider a model with states H (healthy), S (sick) or D (dead) and the transition matrix:

$$\mathbf{A}(t) = \begin{bmatrix} -\sigma(t) - \mu(t) & \sigma(t) & \mu(t) \\ \rho(t) & -\rho(t) - v(t) & v(t) \\ 0 & 0 & 0 \end{bmatrix}$$

where

- ▶ $\sigma(t)$ is the transition rate from healthy to sick (i.e. the rate of healthy people getting sick per year),
- ▶ $\mu(t)$ is the force of mortality,
- ▶ $\rho(t)$ is the recovery rate for the sick, $v(t)$ is the mortality rate for the sick.

- The time to jump out of states H and S , are exponentially distributed with:

$$\lambda_H(t) = -\sigma_{HH}(t) = \sigma(t) + \mu(t)$$

$$\lambda_S(t) = -\sigma_{SS}(t) = \rho(t) + v(t)$$

- Some probabilities of interest:

- ▶ the probability of staying healthy between s and t (probability process jumps out of healthy state after time $t - s$)

$$\Pr(R_s > t - s | X_s = H) = e^{- \int_s^t [\sigma(u) + \mu(u)] du}.$$

- ▶ the probability of staying sick between s and t (probability process jumps out of sick state after time $t - s$)

$$\Pr(R_s > t - s | X_s = S) = e^{- \int_s^t [\rho(u) + v(u)] du}.$$

- ▶ the probability that a healthy individual at time s is sick at time t having been sick for at least w

$$\begin{aligned} & \Pr(X_t = S, C_t > w | X_s = H) \\ &= \int_w^{t-s} P_{\overline{H}\overline{H}}(s, t-v) \sigma(t-v) e^{- \int_{t-v}^t [\rho(u) + v(u)] du} dv, \end{aligned}$$

i.e. stays healthy from s to $(t - v)$ (probability $P_{\overline{H}\overline{H}}(s, t-v)$), jumps from healthy to sick at time $(t - v)$ (transition intensity $\sigma(t-v)$), stays sick to time t (probability $e^{- \int_{t-v}^t [\rho(u) + v(u)] du}$) then integrate for all possible times v

- ▶ the probability that a healthy individual at time s will get sick and stay sick until time t

$$= \int_0^{t-s} e^{- \int_s^{s+w} [\sigma(u) + \mu(u)] du} \sigma(s+w) P_{\overline{S}\overline{S}}(s+w, t) dw,$$

i.e. stays healthy from s to $s+w$, gets sick at time $s+w$, and stays sick from $s+w$ to t

Time Series

- A time series is a sequence of observations that are recorded over time
- Usually recorded at discrete time intervals, spaced evenly
- Purpose: Determine a model that describes the pattern of the time series
 - Can be used to describe important features, forecast and monitor future values

Time Series Model

- Let $\{X_t\}$ denote the time series

$$\dots, X_{t-1}, X_t, X_{t+1}, \dots$$

Each X_t is a random variable. Generally we do not know with certainty which value the object will take until it is observed.

- Let $\{x_t\}$ denote the observed data

$$\{x_1, x_2, \dots, x_n\}.$$

Each x_t is a fixed number.

- A time series model for $\{x_t\}$ is a family of distributions to which the joint distribution of $\{X_t\}$ is assumed to belong.
- A distinctive property of time series data is the serial correlation between $\{X_t\}$ and an analysis of this is essential in time series modelling.

Autocovariance and Autocorrelation functions

- Let $\{X_t\}$ be a stochastic process such that $\text{Var}(X_t) < \infty$ for all t .
 - The autocovariance function (ACVF) is defined by

$$\gamma_X(\tau) = \text{Cov}(X_{t+\tau}, X_t).$$

- The autocorrelation function (ACF) is defined by

$$\rho_X(\tau) = \frac{\gamma_X(\tau)}{\gamma_X(0)} = \text{corr}(X_{t+\tau}, X_t).$$

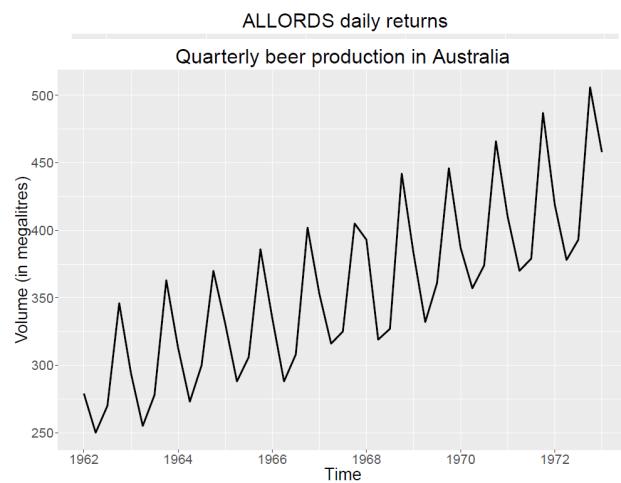
Weakly Stationary

- $\{X_t\}$ is said to be **weakly stationary** (second order stationary, wide sense stationary) if:

$$\begin{aligned} E[X_t] &= \mu \text{ for all } t \\ \text{Cov}(X_{t+\tau}, X_t) &= \gamma(\tau) \text{ for all } t \text{ and } \tau. \end{aligned}$$

That is, mean $E(X_t)$ is a constant and the covariance of the process $\text{Cov}(X_s, X_t)$ depends only on the time difference $t - s$.

- Example:** (Stationary Time Series)



- Example:** (Non-stationary time series)

Stationary Time Series Modelling

- The estimation of the parameters in time series models using historical data can only be performed efficiently in the case of stationary random processes
- A non-stationary process has to be transformed into a stationary one before analysis and modelling.
- Keys steps in stationary time series modelling:
 - Choose a probability model to fit the series (use sample statistics including autocorrelation)
 - Estimate unknown parameters in the model
 - Check model for goodness of fit to the data
 - Use fitted model to enhance understanding of the mechanism generating the series.

Some Operators

- ▶ The backshift operator or lag operator B is defined by

$$BX_t = X_{t-1}.$$

- ▶ Powers of B are defined by

$$B^j X_t = X_{t-j}, \quad t = \dots, -2, -1, 0, 1, 2, \dots$$

- ▶ The difference operator ∇ is defined by

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}.$$

- ▶ Powers of ∇ are defined by

$$\nabla^j X_t = (1 - B)^j X_t,$$

$$\nabla^0 X_t = X_t.$$

- ▶ Define the lag- d difference operator:

$$\nabla_d X_t = (1 - B^d)X_t = X_t - X_{t-d}.$$

IID Noise

- $\{X_t\}$ is IID Noise if X_t and X_{t+h} are independently and identically distributed with mean zero.
- Notation:

$$X_t \sim \text{IID}(0, \sigma^2).$$

Assuming $E(X_t^2) < \infty$, then

$$\begin{aligned} \mu_X &= 0, \\ \gamma_X(0) &= \sigma^2, \\ \gamma_X(h) &= \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0. \end{cases} \end{aligned}$$

White Noise

- $\{X_t\}$ is white noise with zero mean if:

$$\mu_X = 0,$$

$$\gamma_X(0) = \sigma^2,$$

$$\gamma_X(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0. \end{cases}$$

- Notation:

$$X_t \sim \text{WN}(0, \sigma^2).$$

- IID Noise is white noise, but the converse is not true

- White noise is weakly stationary

Linear Processes

- $\{X_t\}$ is a linear process if it can be represented as

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where ψ_j is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$,
 $\{Z_t\} \sim \text{WN}(0, \sigma^2)$.

- The regularity condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ (we say $\sum_{j=-\infty}^{\infty} \psi_j$ is absolutely convergent) ensures that the infinite sum can be manipulated the same way as a finite sum.

For example, two absolutely convergent series can be added or multiplied together in the usual way.

- A linear process can be written compactly as

$$X_t = \psi(B)Z_t$$

where $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$.

Linear processes are stationary.

Proof.

$$\begin{aligned} \text{Cov}(X_t, X_{t+k}) &= \text{Cov}\left(\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \sum_{l=-\infty}^{\infty} \psi_l Z_{t+k-l}\right) \\ &= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \psi_l \text{Cov}(Z_{t-j}, Z_{t+k-l}) \\ &= \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \text{Cov}(Z_{t-j}, Z_{t-j}) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \sigma^2 \end{aligned}$$

- Remarks: Essentially all stationary processes can be represented by a linear process.
- If $\{Y_t\}$ be stationary, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, and $X_t = \psi(B)Y_t$, then $\{X_t\}$ is also stationary.

Moving Average (MA) Model

- $\{X_t\}$ is a first-order moving average process, MA(1), if there is a constant θ and a process $\{Z_t\} \sim WN(0, \sigma^2)$ such that

$$X_t = Z_t + \theta Z_{t-1}.$$

Notation:

$$\{X_t\} \sim MA(1).$$

- The moments and ACVF are

$$\begin{aligned}\mu_X &= 0, \\ \gamma_X(0) &= \sigma^2(1 + \theta^2), \\ \gamma_X(h) &= \begin{cases} \sigma^2(1 + \theta^2) & \text{if } h = 0 \\ \sigma^2\theta & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1. \end{cases}\end{aligned}$$

- $\{X_t\}$ is a moving average process of order q , MA(q), if

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

where $\theta_1, \dots, \theta_q$ are constants, $\{Z_t\} \sim WN(0, \sigma^2)$.

- Remark: (Weakly) stationarity of $\{X_t\}$ follows immediately since X_t is a finite linear combination of stationary variables (white noise).

Autoregressive (AR) Model

- The class of AR models is commonly used
- An AR model has the intuitive appeal that is closely resembles the traditional regression model.

- $\{X_t\}$ is first-order autoregressive process, AR(1), if
 - $\{X_t\}$ is stationary,
 - $X_t = \phi X_{t-1} + Z_t$. where ϕ is a constant and $\{Z_t\} \sim WN(0, \sigma^2)$
 - and Z_t is uncorrelated with X_s for $s < t$.

Notation:

$$\{X_t\} \sim AR(1).$$

- For AR(1) model, we have

$$\begin{aligned}\mu_X &= 0, \\ \gamma_X(0) &= \frac{\sigma^2}{1 - \phi^2}, \\ \gamma_X(h) &= \phi^{|h|} \frac{\sigma^2}{1 - \phi^2}.\end{aligned}$$

AR Model – Condition for Stationary

(ii) If $|\phi| > 1$,

$$\begin{aligned} X_t &= \frac{Z_t}{(1 - \phi B)} = \frac{Z_t}{-\phi B(1 - \frac{1}{\phi B})} = -\frac{1}{\phi B} \sum_{j=0}^{\infty} \left(\frac{1}{\phi B}\right)^j Z_t \\ &= -\sum_{j=0}^{\infty} \left(\frac{1}{\phi}\right)^{j+1} B^{-(j+1)} Z_t = -\sum_{j=0}^{\infty} \left(\frac{1}{\phi}\right)^{j+1} Z_{t+j+1} \end{aligned}$$

Note here $\sum_{j=0}^{\infty} \left|\left(\frac{1}{\phi}\right)^j\right| < \infty$ since $|\phi| > 1$. So stationary.

How about when $\phi = 1$?

When $\phi = 1$, the ACVF $\gamma_X(h)$, $h = 0, 1, \dots$ are not defined. More importantly, when $\phi = 1$ we have a non-stationary process called a **random walk**. We do not consider this an AR(1) process in this course (however, some references do).

Consider the random walk $X_t = X_{t-1} + Z_t = Z_t + Z_{t-1} + \dots + Z_1 + X_0$.

If $X_0 = 0$, then

$$\begin{aligned} E(X_t) &= 0, \\ E(X_t^2) &= t\sigma^2, \\ E(X_{t+h}X_t) &= \begin{cases} t\sigma^2 & \text{if } h = 0 \\ (t+h)\sigma^2 & \text{if } -t < h < 0 \\ t\sigma^2 & \text{if } h > 0. \end{cases} \end{aligned}$$

A random walk is not stationary!

But note that $\nabla X_t = X_t - X_{t-1} = Z_t$ is stationary!

AR Model

- $\{X_t\}$ is an autoregressive process of order p , AR(p), if

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t,$$

where $\{X_t\}$ is stationary, $\{Z_t\} \sim WN(0, \sigma^2)$, Z_t is uncorrelated with X_s for $s < t$, and ϕ_1, \dots, ϕ_p are constants.

- The requirement that $\{X_t\}$ is stationary restricts the values of ϕ_1, \dots, ϕ_p .

Autoregressive Moving Average (ARMA) Model

- ▶ $\{X_t\}$ is an ARMA(1, 1) process if
 - ▶ X_t is stationary
 - ▶

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1},$$
 where ϕ and θ are constants, $\{Z_t\} \sim WN(0, \sigma^2)$.

- ▶ The ARMA(1, 1) process can be written as

$$\phi(B)X_t = \theta(B)Z_t,$$

where

$$\phi(B) = 1 - \phi B \text{ and } \theta(B) = 1 + \theta B.$$

- ▶ Which values of ϕ and θ ensure the existence of a stationary $\{X_t\}$ that satisfies the ARMA(p, q) equation

$$\phi(B)X_t = \theta(B)Z_t?$$

- ▶ $\{X_t\}$ is an ARMA process of order(p, q): ARMA(p, q), if

- ▶ $\{X_t\}$ is stationary,
- ▶

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad (i.e. \phi(B)X_t = \theta(B)Z_t)$$

where

- ▶ $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ are constants
- ▶ $\{Z_t\} \sim WN(0, \sigma^2)$
- ▶ the polynomials $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common factors

- ▶ $\{X_t\}$ is an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

Example:

For an ARMA(1, 1) model, we have $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$.

For an ARMA(1, 1) process $\{X_t\}$, stationarity is guaranteed if $|\phi| \neq 1$.

This is equivalent to saying that for any z with $|z| = 1$,

$$\phi(z) = 1 - \phi z \neq 0.$$

In other words, the root of the polynomial $\phi(z)$ is not ± 1 .

Theorem:

Consider an ARMA(p,q) model

$$\phi(B)X_t = \theta(B)Z_t$$

If $\phi(z)$ has no roots on the unit circle, there exists a filter

$$\frac{1}{\phi(B)} = \sum_{j=-\infty}^{\infty} \chi_j B^j \quad \text{with} \quad \sum_{j=-\infty}^{\infty} |\chi_j| < \infty$$

such that

$$\begin{aligned} \frac{\phi(B)}{\phi(B)} X_t &= \frac{\theta(B)}{\phi(B)} Z_t \\ X_t &= \frac{\theta(B)}{\phi(B)} Z_t. \end{aligned}$$

Since $\theta(B)/\phi(B)$ is absolutely convergent, it implies that $\{X_t\}$ is **stationary**.