Applied Probability Revision

03 May, 2023

1 Discrete Time Markov Chains

A **discrete-time stochastic process** is defined as a sequence of discrete random variables $\{X_n\}_{n\in\mathbb{N}_0}$, each taking values in a countable state space E

A discrete time stochastic process satisfying the **Markov condition** is called a **Markov Chain**

$$\Pr(X_n = j|X_{n-1} = i, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0) = \Pr(X_n = j|X_{n-1} = i)$$

for all $n \in \mathbb{N}$ and for all $x_0, \ldots, x_{n-2}, i, j \in E$.

The Markov chain is **time-homogenous** if

$$\Pr(X_{n+1} = j|X_n = i) = \Pr(X_1 = j|X_0 = i)$$

for every $n \in \mathbb{N}_0$ and $i, j \in E$

1.1 Chapman-Kolmogorov (CK) equations

The **n-step** transition probability is

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

For a time homogenous Markov chain we have the **CK-equations**

$$p_{ij}(m+n) = \sum_{l\in E} p_{il}(m)p_{lj}(n)$$

where $m \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$

The formula for **n-step transition matrix** follows:

$$P_{m+n} = P_m P_n$$

and in particular

$$P_n = P^n$$

1.2 First passage and hitting times

The **first passage time** is

$$T_j = \min\{n \in \mathbb{N} : X_n = j\}$$

In other words, $\{T_j = n\} = \{X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j\}$, if $X_n \neq j, \forall n \in \mathbb{N}$, then $T_j = \infty$.

The **first passage probability** is

$$f_{ij}(n) = \Pr(T_j = n | X_0 = i), n \in \mathbb{N}_0$$

from which the hitting probability follows

$$f_{ij} = \Pr(T_j < \infty | X_0 = i) = \sum_{n=0}^{\infty} f_{ij}(n)$$

With the **special case** being $f_{ij}(0) = 0$.

Decomposing the **n-step transition probability**

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l)$$

this is the same as starting from $l = 0$.

1.3 Generating Functions of Markov Chain

Recall the **probability generating function**

$$G_X(s) = \sum_{x=0}^{\infty} s^x \Pr(X = x)$$

where this holds on the support

$$\mathcal{S}_{\mathcal{X}} = \left\{ s \in \mathbb{R} : \sum_{x=0}^{\infty} |s|^x \Pr(X = x) < \infty \right\}$$

The generating functions here are

$$G_{p_{ij}(n)} = \sum_{n=0}^{\infty} p_{ij}(n) s^n$$
$$G_{f_{ij}(n)} = \sum_{n=0}^{\infty} f_{ij}(n) s^n$$

By arguing using equating coefficients and an identity, we have a **theorem**

$$G_{p_{ij}(n)} = \delta_{ij} + G_{f_{ij}(n)}(s) G_{p_{ij}(n)}$$

The identity used is decomposition of n-step transition probability.

2 Recurrence and Transience

A state j is **recurrent** if and only if

$$\sum_{n=1}^{\infty} p_{jj}(n) = \infty$$

A state j is **transient** if and only if

$$\sum_{n=1}^{\infty} p_{jj}(n) < \infty$$

Examples: Examples of transient, irreducible chains

The **number of periods** that the chain is in state j (or **number of visits** to j) is

$$N_j = \sum_{n=0}^{\infty} I_n(j)$$

where $I_n(j)$ is the indicator function taking value 1 if $X_n = j$ and 0 otherwise.

The **expected number of visits** to state j given $X_0 = j$ is

$$\mathbb{E}[N_j | X_0 = j] = \sum_{n=0}^{\infty} p_{jj}(n)$$

proof using generating functions:

Taking $s \to 1$ and using Abel's theorem, we can deduce...

2.1 Properties of recurrent/transient states

Theorem (Number of visits is geometric for transient states)

If j is transient, then

$$\Pr(N_j = n | X_0 = j) = f_{jj}^{n-1}(1 - f_{jj}), n \in \mathbb{N}$$

Let $i \neq j$, then

$$\Pr(N_j = n | X_0 = i) = \begin{cases} 1 - f_{ij} & n = 0 \\ f_{ij} f_{jj}^{n-1}(1 - f_{jj}) & n \geq 1 \end{cases}$$

Intuition is that the chain visits j for the first time and returns to it for $n-1$ times, then leaves it. (note the N_j starts from 0)

Therefore, it follows that for $i \neq j$ by the mean of geometric distribution,

$$\mathbb{E}[N_j | X_0 = i] = \frac{f_{ij}}{1 - f_{jj}}$$

and

$$\mathbb{E}[N_j|X_0 = j] = \frac{1}{1 - f_{jj}}$$

Theorem(Unlikely to visit a transient state)

If j is transient, then

$$\lim_{n\to\infty} p_{ij}(n) = 0, \forall j \in E$$

Similar results hold for null recurrent states.

2.2 Mean recurrence time, null and positive recurrence

The **mean recurrence time** μ_j is

$$\mu_j = \mathbb{E}[T_j|X_0 = j] = \sum_{n=1}^{\infty} n f_{jj}(n)$$

where we recall that $\{T_j = n\} = \{X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j\}$.

Similarly, we can define the **mean first passage time** μ_{ij}:

$$\mu_{ij} = \mathbb{E}[T_j|X_0 = i] = \sum_{n=1}^{\infty} n f_{ij}(n)$$

those expectations can be finite or infinite; for transient states, they must be infinite.

For a recurrent state j, it is called **null recurrent** if $\mu_j = \infty$ and **positive recurrent** if $\mu_j < \infty$.

Theorem (unlikely to visit null recurrent state)

A state j is null recurrent, if and only if

$$\lim_{n\to\infty} p_{jj}(n) = 0$$

In addition if the above equation holds, then

$$\lim_{n\to\infty} p_{ij}(n) = 0, \forall i \neq j \in E$$

3 Aperiodicity and Ergodicity

The **period** of a state j is

$$d(j) = gcd\{n \in \mathbb{N} : p_{jj}(n) > 0\}$$

It is not necessarily true that $p_{jj}(d(j)) > 0$ (cf. Notes Pg. 36).

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

A state is **ergodic** if it is positive recurrent and aperiodic.

4 Communicating classes

- We say that a state j is **accessible** from state i if the chain can reach j at some time, written as $i \to j$.
- Two states i and j are **communicating** if there exists a state k such that $i \to k$ and $k \to j$, we write $i \leftrightarrow j$; this is an **equivalence relation**.
- If $i \neq j$, then $i \to j$ if and only if $f_{ij} > 0$.

4.1 Properties preserved by Communicating Classes

- Same period
- Same transience/recurrence
- Null recurrence

For a **set of states** C:

- C is **closed** if $\forall i \in C$, $j \notin C$, $p_{ij} = 0$
- C is **irreducible** if all states in the set communicate with each other

Therefore, an irreducible set of states share the same properties described above.

Theorem (Recurrence and closed)
If C is a communicating class of recurrent states, then C is closed.

Theorem (Stochastic matrix on closed states)
The stochastic matrix P restricted to a closed set of closed states C is still a stochastic matrix.

4.2 Decomposition of Chains

The state space can be partitioned into communicating classes.

$$E = T \cup \left(\bigcup_i C_i \right)$$

where T is the set of transient states and C_i's are irreducible closed sets of recurrent states.

4.3 Class Properties

The **classes** refer to communicating classes.

Theorem (Finite Chains have recurrent)
When state space is **finite**, at least one state is *recurrent* and all *recurrent* states are **positive**

Remark This combined with later results on stationarity makes a chain with finite state space particularly nice.

Remark It follows that there are no null recurrent states in a finite state space.

Theorem (Finite and closed)
If C is a finite, closed communicating class, then all states are positive recurrent.

Communicating class properties

Type of Class	Finite	Infinite
Closed	positive recurrent	positive/null recurrent, transient
Not closed	transient	transient

5 Gambler's Ruin

- Starting at state $i \in \{0, 1, \ldots, N\}$
- p of winning one unit and $1-p$ of losing one unit

- Assume successive games are independent

Question: What is the probability of reaching N before reaching 0?

Strategy: Use **first step analysis**:

- Define $V_i = \min\{n \in \mathbb{N}_0 : X_n = i\}$, the first time being present at state i, desired event is $V_N < V_0$
- Consider the conditional probability of starting at i and ending up at N

$$h_i = \Pr(V_N < V_0 \mid X_0 = i)$$

- Consider the first step

$$\begin{aligned} h_i &= \Pr(V_N < V_0 \mid X_1 = i+1, X_0 = i) \Pr(X_1 = i+1 \mid X_0 = i) \\ &\quad + \Pr(V_N < V_0 \mid X_1 = i-1, X_0 = i) \Pr(X_1 = i-1 \mid X_0 = i) \\ &\quad = h_{i+1} p + h_{i-1}(1-p) \end{aligned}$$

Finish by solving this recurrence relation

$$h_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N} & p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & p = \dfrac{1}{2} \end{cases}$$

Stationarity

We are interested in the equilibrium states of a chain

- A **distribution** is a row vector λ with $\Sigma_j \lambda_j = 1$

- If $\lambda P = \lambda$ then it is called ***invariant***

5.1 Stationary distributions of irreducible chains

Theorem Every irreducible chain has a **stationary distribution** π if and only if all states are **positive recurrent**

- π is unique
- $\pi = \mu_i^{-1}$ the inverse of mean recurrence time

We first have some lemmas:

$$l_{ji}(n) = \Pr(X_n = i, T_j \geq n | X_0 = j)$$

being the probability that the chain reaches i in n steps without returning to j
Lemma (Decomposing the first hitting)

$$f_{jj}(m+n) = \sum_{i \in E, i \neq j} l_{ji}(m) f_{ij}(n)$$

from which $f_{jj}(m+n) \geq l_{ji}(m) f_{ij}(n)$ follows

Lemma (Formula for hitting) We also have the following recurrence relation for $l_{ji}(n+1)$

$$l_{ji}(n+1) = \sum_{r \in E, r \neq j} p_{ri} l_{jr}(n)$$

with $l_{ji}(1) = p_{ji}$.

Lemma (Existence): A positive recurrent chain has a stationary distribution.

Proof: (constructive)

- **(Step1 Construction)** Let $N_i(j)$ be the number of visits to state i before visiting state j (again); the sum of such numbers over i is equal to the hitting time T_j

- Define $\rho_i(j)$ to be the expected number of visits to the state i between two successive visits to state j (in this step the **recurrence** of the chain is used, as the T_j is finite with probability 1)

$$\begin{aligned}
\rho_i(j) &= \mathbb{E}[N_i(j)|X_0 = j] \\
&= \sum_n \Pr(X_n = i, T_j \geq n | X_0 = j) \\
&= \sum_n l_{ji}(n)
\end{aligned}$$

- Now the mean hitting time can be computed as

$$\begin{aligned}
\mu_j &= \mathbb{E}\big[\sum_i N_i(j)|X_0 = j\big] \\
&= \sum_i \rho_i(j)
\end{aligned}$$

- which can be written as sum of $\rho_i(j)$ by Tonelli and linearity of conditional expectation
- **(Step2 Finiteness)** Use a lemma to bound $\rho_i(j)$ so it's finite
- Namely write $\rho_i(j) = \sum_n l_{ji}(n)$ and bound using the fact that the chain is irreducible, so there exists $f_{ij}(n^*) > 0$, so $f_{jj}(m + n^*) \geq l_{ji}(m) f_{ij}(n^*)$
- **(Step3 Stationarity)** Use recurrence to show

$$\begin{aligned}
\rho_i(j) &= \sum_n l_{ji}(n) \\
&= p_{ji} + \sum_{n=2} \sum_{r, r \neq j} p_{ri} l_{jr}(n-1) \\
&= p_{ji}\rho_j(j) + \sum_{n=1} \sum_{r \neq j} p_{ri} l_{jr}(n) \\
&= p_{ji}\rho_j(j) + \sum_{r, r \neq j} p_{ri} \sum_{n=1} l_{jr}(n) \\
&= \sum_r \rho_r(j) p_{ri}
\end{aligned}$$

- which uses the fact that $\rho_j(j) = 1$
- This $\rho_i(j)$ does not necessarily give a probability vector when the chain is not positive recurrent.

- Now if the chain is positive recurrent, we have μ_j finite for every j, we have

$$\pi_i = \frac{\rho_i(j)}{\mu_j}$$

Therefore, we conclude that

Every **irreducible, recurrent** chain has a positive solution to the equation $\mathbf{x} = \mathbf{x}P$, which is unique up to a multiplicative constant (see PS2 for proof).
Moreover, the chain is

- positive recurrent if $\sum_i x_i < \infty$
- null recurrent if $\sum_i x_i = \infty$

Also, from the proof, we conclude 3 **identities**:

(Sum of expected visits)

$$\mu_j = \sum_i \rho_i(j)$$

(Sum of hitting prob)

$$\rho_i(j) = \sum_n l_{ji}(n)$$

(Sum of number of visits)

$$T_j = \sum_i N_i(j)$$

Lemma (Tail probability is expectation)
T is a nonnegative integer-valued random variable and A is an event with $\Pr(A) > 0$. Then

$$\mathbb{E}[T|A] = \sum_{n=0}^{\infty} \Pr(T \geq n|A)$$

Lemma If a stationary distribution exists, then the chain is positive recurrent and the distribution must be given by $\pi_i = \mu_i^{-1}$

proof: ...

In a **reducible chain**, the following results are useful:

- $\pi_i = 0$ for stationary distribution π if i is transient or null recurrent, so we can only compute the positive recurrent states and set the rest of π_i to 0
- A discrete time Markov chain with **finite** state space always has at least one stationary distribution.
- This distribution is **unique** unless it has two or more closed communicating classes.
- Every stationary distribution is a **linear combination** of the stationary distributions of the closed communicating classes, with coefficients adding up to 1.

5.2 Limiting Distribution

A distribution π is a **limiting distribution** of a chain if π satisfies

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j$$

for any $i \in E$.

Theorem For an irreducible, aperiodic chain

$$\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\mu_j}$$

It follows that for an irreducible, aperiodic, and positive recurrent state, the limiting distribution is its unique stationary distribution

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j = \frac{1}{\mu_j}$$

Example of Chain with no limiting distribution

Consider the transistion matrix of two alternating states

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

the even and odd powers differ, but it has stationary distribution $\pi = (1/2, 1/2)$.

5.3 Ergodic Theorem

The **number of visits to i before time n** is defined as

$$V_i(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k = i\}}$$

Theorem If a chain is irreducible, $V_i(n)/n$ denotes the proportion of time the chain spent in state i before time n

$$\Pr\left(\frac{V_i(n)}{n} \to \frac{1}{\mu_j} \text{ as } n \to \infty\right) = 1$$

5.4 Summary of properties of irreducible chains

1 Positive Recurrent

- Stationary distribution **exists** and is **unique**
- All mean recurrence times $\mu_j = \mathbb{E}[T_j | x_0 = j]$ are finite and $\pi_j = \dfrac{1}{\mu_j}$
- $V_i(n)/n \to \pi_i$
- If the chain is aperiodic, the limiting distribution is the stationary distribution

2 Null Recurrent

- All mean recurrence times are infinite

- No stationary distribution
- $V_i(n)/n \to 0$
- The limiting distribution is 0

3 Transient

- All mean recurrence times are infinite (any state is eventually never visited)
- No stationary distribution
- $V_i(n)/n \to 0$
- The limiting distribution is 0

6 Time reversibility

In this section, we assume the Markov chains are **irreducible and positive recurrent**, therefore there is a unique stationary distribution π.

The **reversed chain** for some $N \in \mathbb{N}$ is defined as

$$Y_n = X_{N-n}$$

Theorem (Reversed still Markov) The reversed chain is a Markov chain

$$\Pr(Y_{n+1} = j | Y_n = i) = \Pr(X_{N-n-1} = j | X_{N-n} = i) = \frac{\pi_j}{\pi_i} p_{ji}$$

A Markov chain (X_n) is called **time-reversible** if its transition matrix is the same as the transition matrix of its reversed chain.

Theorem A Markov chain is time-reversible if and only if

$$\pi_i p_{ij} = \pi_j p_{ji}$$

this condition is called **detailed balance**.

Theorem (Detailed balance implies positive recurrence)
For an irreducible chain, if **there is a vector** π such that the detailed balance equation holds for all i, j, then the chain is **time-reversible and positive recurrent** with stationary distribution π.

Proof: Note that the detailed balance conditions imply the chain has a stationary distribution (summing w.r.t. i), hence positive recurrent by previous theorems.

7 Continuous Time Markov Chains

7.1 Types of Processes

A **right continuous** stochastic process $\{X_t\}_{t\geq 0}$ is such that for any $\omega \in \Omega$ and $t \geq 0$, there is $\varepsilon > 0$, such that

$$X_t(\omega) = X_s(\omega) \qquad \forall s \in [t, t+\varepsilon]$$

Can be thought as closed point on the left and open point on the right.

There are three types of right continuous processes

- **Normal**: infinitely many jumps but only finitely many in a finite time interval
- **Absorption**: Only has finitely many jumps, gets absorbed at some point (stay at one state)
- **Explosion**: Infinitely many jumps in a finite time interval.

The **jump times** are random variables $J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}$.

The **holding times** are random variables defined as:

$$H_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise} \end{cases}$$

from which it follows that $J_n = \sum_{i=1}^{n} H_i$.

The **explosion time** is

$$J_\infty = \sup_{n \in \mathbb{N}_0} J_n = \sum_{n=1}^{\infty} H_n$$

A **jump process** or jump chain is a discrete time stochastic process $Z_n = X_{J_n}$, where J_n is the nth jump time.

7.1.1 Relating continuous process to its jump process

A **counting process** is a stochastic process $\{N_t\}_{t \geq 0}$ satisfying

- $N_0 = 0$
- $\forall t \geq 0, N_t \in \mathbb{N}_0$
- (Non-decreasing) If $0 \leq s \leq t$, $N_s \leq N_t$
- (Counting) When $s < t$, $N_t - N_s$ equals the no. of events in $(s, t]$
- (Right continuous) The process is piecewise constant and has upward jumps (single step) of size 1, therefore

$$N_{t^-} = \lim_{s \uparrow t} N_s$$

A **counting process associated the sequence** $(J_n)_{n \in \mathbb{N}_0}$

7.2 Properties of random variables

The **Poisson random variable** has (pmf)

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x \in \mathbb{N}_0$$

It has expectation

$$\mathbb{E}[X] = \lambda$$

and variance

$$\mathrm{Var}[X] = \lambda$$

The **exponential random variable** has

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

and c.d.f.

$$F_X(x) = 1 - e^{-\lambda x}$$

with a nonnegative support.

It has expectation

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

and variance

$$\mathrm{Var}[X] = \frac{1}{\lambda^2}$$

The **memoryless property** of a random variable refers to the fact:

$$\Pr(X > x + y \mid X > x) = \Pr(X > y)$$

- A continuous random variable is memoryless iff it is $\mathrm{Exp}(\lambda)$
- A discrete random variable is memoryless iff it is $\mathrm{Geom}(p)$

The **sum of exponential** $\mathrm{Exp}(\lambda)$ is a $\mathrm{Gamma}(n, \lambda)$ distribution

$$f_{J_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \qquad t > 0$$

The **convergence for infinite sum of exponential** has the following criteria

- If $\sum \frac{1}{\lambda_i} < \infty$, then $\Pr(J_\infty < \infty) < 1$
- If $\sum \frac{1}{\lambda_i} = \infty$, then $\Pr(J_\infty = \infty) = 1$

The **minimum of exponential** is

$$H \sim \mathrm{Exp}(\sum_{i=1}^{n} \lambda_i)$$

and the probability of any of the k variables being the minimum is

$$\Pr(H = H_k) = \frac{\lambda_k}{\sum_{i=1}^{n} \lambda_i}$$

The **Laplace Transform** of a random variable X is given by

$$\mathcal{L}_X(u) = \mathbb{E}[e^{-uX}]$$

A list of transformations for common random variables:

- (Poisson) $\exp(\lambda t[e^{-u} - 1])$
- (Exponential) $\dfrac{\lambda}{\lambda + u}$

The **characteristic function** of a random variable X is given by

$$\phi_X(t) = \mathbb{E}[e^{itX}]$$

8 Poisson Processes

8.1 Definitions

A **Poisson process**, denoted $\{N_t\}_{t\geq 0}$, is a non-decreasing stochastic process with nonnegative values satisfying

- $N_0 = 0$
- The increments are independent, $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$, the random variables $N_{t_0}, N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent
- The increments are stationary

$$\Pr(N_t - N_s = k) = \Pr(N_{t-s} = k)$$

- There is a single arrival (only one arrives in a small interval), for all $t \geq 0$ and $\delta > 0$, $\delta \to 0$

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

This also ensures that a Poisson process is continuous in probability.

An **equivalent definition** replaces the last condition with the variable being Poisson with rate N_t

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

Another **equivalent definition** characterizes Poisson process $\{N_t\}_{t\geq 0}$ explicitly

- Let $H_1, H_2, \ldots$ denote i.i.d. $\text{Exp}(\lambda)$ random variables
- Let $J_0 = 0$ and $J_n = \sum_{i=1}^n H_i$
- We define

$$N_t = \sup\{n \in \mathbb{N}_0 : J_n \leq t\}$$

8.2 Properties of Poisson Process

8.2.1 Inter-arrival times

The inter-arrival times are **i.i.d.** $\text{Exp}(\lambda)$ random variables

8.2.2 Time to n^{th} event

The time to n^{th} event is defined as

$$J_n = \sum_{i=1}^{n} H_i$$

which follows a Gamma(n, λ) distribution

$$f_{J_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \qquad t > 0$$

8.2.3 Conditional distribution of arrival times

The conditional joint density of $(J_1, \ldots, J_n)$ is given by the order statistic

$$f_{(J_1,\ldots,J_n)}(t_1, \ldots, t_n \mid N_t = n) = \begin{cases} \dfrac{n!}{t^n} & 0 < t_1 < \ldots < t_n \\ 0 & \text{otherwise} \end{cases}$$

The expectation of the k^{th} value of n uniformly distributed order statistics on $[0, t]$ is

$$\mathbb{E}[X_{(k)}] = \frac{tk}{n+1} = \mathbb{E}[J_k \mid N_t = n]$$

8.3 Extensions to Poisson Processes

8.3.1 Superposition

Given n *independent* Poisson processes $\{N_t^{(1)}\}_{t\geq 0}, \ldots, \{N_t^{(n)}\}_{t\geq 0}$, with respective rates $\lambda_1, \ldots, \lambda_n > 0$,

$$N_t = \sum_{i=1}^{n} N_t^{(i)}$$

is also a Poisson process with rate $\lambda = \sum_{i=1}^{n} \lambda_i$.

This is called a **superposition of Poisson processes**.

8.3.2 Thinning

- Each arrival of a Poisson Process $\{N_t\}_{t\geq 0}$ is marked as a type k event with probability p_k, for $k = 1, \ldots, n$, where $\sum_{k=1}^{n} p_k = 1$.
- Then let $N_t^{(k)}$ denote the number of type k events up to time t (in $[0, t]$).
- Every $N_t^{(k)}$ is a Poisson process with rate λp_k.

Each process is called a **thinned Poisson Process**.

8.4 Non-homogenous Poisson processes

Let $\lambda : [0, \infty) \to (0, \infty)$ denote a non-negative and locally integrable function. Then the process $N = \{N_t\}_{t\geq 0}$ is a **non-homogenous Poisson process** with intensity function $\lambda(t)$ if

- $N_0 = 0$

- N has independent increments
- Single arrival; for all $t \geq 0$ and $\delta > 0$,

$$\Pr(N_{t+\delta} - N_t = 1) = \lambda(t)\delta + o(\delta)$$
$$\Pr(N_{t+\delta} - N_t \geq 2) = o(\delta)$$

Each N_t follows a **Poisson distribution with rate** $m(t)$, where

$$m(t) = \int_0^t \lambda(s)ds$$

The stationarity also changes. We have

$$N_t - N_s \sim \text{Poisson}(\int_s^t \lambda(u)du) = \text{Poisson}(m(t) - m(s))$$

8.4.1 Deriving the forward equations

An important technique for deriving concrete probability mass functions using the single arrival property.

$$\begin{aligned}
p_n(t+\delta) = \Pr(N_{t+\delta} = n) &= \sum_{k=0}^{n} \Pr(N_{t+\delta} = n \mid N_t = k)\Pr(N_t = k) \\
&= \sum_{k=0}^{n} \Pr(N_{t+\delta} - N_t = n - k \mid N_t = k)\Pr(N_t = k) \\
&= \sum_{k=0}^{n} \Pr(N_{t+\delta} - N_t = n - k)\Pr(N_t = k) \\
&= (1 - \lambda(t)\delta)p_n(t) + \lambda(t)\delta p_{n-1}(t) + o(\delta)
\end{aligned}$$

Note the use of independence of increments and the single arrival property.

This gives the differential equation

$$\frac{dp_n(t)}{dt} = \lambda(t)p_{n-1}(t) - \lambda(t)p_n(t)$$

When $n = 0$,

$$\frac{dp_0(t)}{dt} = -\lambda(t)p_0(t)$$

8.5 Compound Poisson processes

Let $\{N_t\}_{t\geq 0}$ be a Poisson process with rate $\lambda > 0$ and $\{Y_n\}_n$ be a sequence of identically, indepdently distributed random variables that are also *independent* of $\{N_t\}_{t\geq 0}$.

$$S_t = \sum_{n=1}^{N_t} Y_n$$

$\{S_t\}_{t\geq 0}$ is a **compound Poisson process**.

The mean and variance of S_t are

$$
\begin{aligned}
\mathbb{E}[S_t] &= \lambda t \ \mathbb{E}[Y_1] \\
\mathrm{Var}[S_t] &= \lambda t \ \mathbb{E}[Y_1^2]
\end{aligned}
$$

This is proven by conditioning on N_t and using the fact that Y_n are independent.

We also recall the laws of total expectation and total variance.

$$
\begin{aligned}
\mathbb{E}[S_t] &= \mathbb{E}[\mathbb{E}[S_t \mid N_t]] \\
\mathrm{Var}[S_t] &= \mathbb{E}[\mathrm{Var}[S_t \mid N_t]] + \mathrm{Var}[\mathbb{E}[S_t \mid N_t]]
\end{aligned}
$$

8.6 Cramer-Lundberg

An application of the compound Poisson process is the Cramer-Lundberg model.

For an insurance company, there are - **Claims** S_t (expense to pay when there are accidents) modelled by a **compound Poisson process**

- **Initial capital** u
- **Premiums** ct (money collected from customers with rate c)

We define the **risk process** to be

$$
U_t = u + ct - S_t, \qquad t \geq 0
$$

The company goes bankrupt if $U_t < 0$.

Thus, the **ruin probability** is defined as

$$
\psi(u, T) = \Pr(U_t < 0 \text{ for } \text{ some } \text{ t} \leq T), \qquad T > 0, u \geq 0
$$

The **total claim amount** $\{S_t\}_{t \geq 0}$ is

$$
S_t = \begin{cases} \sum_{n=1}^{N_t} Y_n & \text{if } N_t \geq 1 \\ 0 & \text{otherwise} \end{cases}
$$

where N_t is a Poisson process with rate λ and Y_n are independent and identically distributed random variables with finite mean μ and variance σ^2.

We can compute the expected value of the risk process.

$$
\mathbb{E}[U_t] = u + ct - \lambda t \mu
$$

Therefore, a minimal requirement for this company to choose premium rate could be

$$
c > \lambda \mu
$$

this is called the **net profit condition**.

8.7 Coalescent Process

The coalescent process describes the merging of n offspring into a single ancestor occuring at random times.

- We have n individuals at time $t = 0$
- Each pair of individuals merge according to a Poisson process with rate $\lambda = 1$ and there are $\binom{n}{2}$ pairs
- The time of first coalescence follows $\text{Exp}(\binom{n}{2})$ distribution
- There are $n - 1$ coalescences
- The process is in fact a death process

We can compute the time to the most recent common ancestor (i.e. the time of the last coalescence).

$$\mathbb{E}\left(\sum_{k=1}^{n-1} H_k\right) \qquad n \in \mathbb{N}, n \geq 2$$

with

$$H_k \sim \text{Exp}(\binom{n-(k-1)}{2})$$

So it follows that

$$\mathbb{E}\left(\sum_{k=1}^{n-1} H_k\right) = \sum_{k=1}^{n-1} \frac{2}{k(k+1)} = 2(1 - \frac{1}{n})$$

Comparing with the last coalescence time, we have

$$\mathbb{E}(H_{n-1}) = 1 > 2(1 - \frac{1}{n})$$

showing that the last coalescence time is larger than half of the expected total coalescence time.

9 Continuous-time Markov chains

A continuous-time stochastic process $\{X_t\}_{t \in [0,\infty)}$ satisfies the **Markov property** if

$$\Pr(X_{t_n} = j \mid X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}) = \Pr(X_{t_n} = j \mid X_{t_{n-1}} = i_{n-1})$$

for all $j, i_1, \ldots, i_{n-1} \in E$ and for **any** sequence $0 \leq t_1 < \cdots < t_n < \infty$.

The **transition probability** is $p_{ij}(s,t)$, for $s \leq t$, $i, j \in E$

$$p_{ij}(s,t) = \Pr(X_t = j \mid X_s = i)$$

The chain is **homogeneous** if

$$p_{ij}(s,t) = p_{ij}(0, t-s)$$

In this course, it is always assumed that the chain is homogeneous, thus we always denote $p_{ij}(t) = p_{ij}(0,t)$.

Theorem The family is a **stochastic semigroup** if:

- $\mathbf{P}_0 = I_{K\times K}$
- $\mathbf{P}_t$ is stochastic
- Chapman-Kolmogorov equations are satisfied

$$p_{ij}(s+t) = \sum_{k\in E} p_{ik}(s)p_{kj}(t)$$

The semigroup $\{P_t\}$ is called **standard** if

$$\lim_{t\downarrow 0} \mathbf{P}_t = \mathbf{I}$$

The Poisson process is a continuous time Markov chain.

9.1 Holding times

We define the **holding time at state i** as

$$H_{|i} = \inf\{s \geq 0 : X_{t+s} \neq i\}$$

Theorem The holding time follows an **exponential distribution** (due to its memoryless property)

9.1.1 Exponential Alarm Clocks

- For each state $i \in E$, it can reach n_i states
- Set n_i independent exponential alarm clocks with rates q_{ij}
- The state transfers to the index of the first alarm clock that rings
- Transfer to state j with probability $\frac{q_{ij}}{\sum_k q_{ik}}$ (ordering of exponential random variables)

9.2 The generator

The **generator** $\mathbf{G} = (g_{ij})_{i,j\in E}$ of the Markov chain with stochastic semigroup $\mathbf{P}_t$ is defined as the card(E) $\times$ card(E) matrix

$$\mathbf{G} = \lim_{\delta\downarrow 0} \frac{1}{\delta}[\mathbf{P}_\delta - \mathbf{I}]$$

where $\mathbf{P}_t$ is differentiable at $t = 0$.

Informally, we have $g_{ij} = q_{ij} = p'_{ij}(0)$, so when the time interval δ is small enough, we have the estimates for transition probabilities:

$$p_{ij}(\delta) \approx g_{ij}\delta = q_{ij}\delta p_{ii}(\delta) \approx 1 + g_{ii}\delta = 1 - \sum_{j\in E} q_{ij}\delta$$

9.3 Forward and backward equations

Theorem
Subject to regularity conditions, a continuous-time Markov chain with stochastic semigroup $\{P_t\}$ and generator $\mathbf{G}$ satisfies the **Kolmogorov forward equation** and the **Kolmogorov backward equation**

$$\mathbf{P}_t' = \mathbf{P}_t\mathbf{G}$$
$$\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t$$

This allows us to write

$$\mathbf{P}_t = \exp(t\mathbf{G})$$

using matrix exponential.

9.4 Irreducibility and stationarity

The chain is **irreducible** if for all $i, j \in E$, there exists $t > 0$ such that $p_{ij}(t) > 0$.

Theorem (No periodicity in continuous)
If $p_{ij}(t) > 0$ for some $t > 0$, then $p_{ij}(t) > 0$ for all $t > 0$.

A distribution is the **stationary distribution** if it satisfies

$$\pi\mathbf{P}_t = \pi$$

for all $t \geq 0$.

A distribution π is the **limiting distribution** if for all $i, j \in E$

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j$$

Theorem (find stationary distr)
Subject to regularity conditions, $\pi = \pi\mathbf{P}_t$ for all $t \geq 0$ if and only if $\pi\mathbf{G} = 0$,.

Theorem (Ergodicity in continuous time)

1. If there exists a stationary distribution, then it is *unique* and $\forall i, j \in E$
$$\lim_{t\to+\infty} p_{ij}(t) = \pi_j$$
2. If there is no stationary distribution then
$$\lim_{t\to+\infty} p_{ij}(t) = 0$$

9.5 Jump chain and explosion

9.5.1 From continuous to discrete

Assume the generator is known.

- J_n being the n^{th} change in value of the chain X, $J_0 = 0$
- Values right after the jump $Z_n = X_{J_n+}$ form a discrete time Markov chain
- Construct transition matrix $p_{ij}^Z = \dfrac{g_{ij}}{-g_{ii}}$ and 0 if absorption (all the diagonal entries are 0)
- $\{Z_n\}_{n\geq 0}$ is the **jump chain**

9.5.2 From discrete to continuous

Assume the transition matrix is known.

- Let $p^Z_{ii} = 0$ to avoid jumps to itself in the discrete chain
- Construct generator matrix with arbitrary nonnegative g_i for each i

$$g_{ij} = \begin{cases} g_i p^Z_{ij} & i \neq j \\ -g_i & i = j \end{cases}$$

- Condition on Z_i, let $H_i \sim \mathrm{Exp}(\mathrm{g_{Z_{i-1}}})$ be the 'holding times'
- Then at time t, check if between two jump times

$$X_t = \begin{cases} Z_n & J_n \leq t < J_{n+1} \\ \infty & \text{otherwise} \end{cases}$$

The chain explodes if $\Pr(J_\infty < \infty) > 0$.

9.6 Relation between common quantities

Notation	Element	Meaning and Conditions
q_{ij}	$q_i := q_{ii} = \sum_{j \in E} q_{ij}$	The **exponential rates** $q_{ij} > 0$ when $i \neq j$ and $i \leftrightarrow j$, zero otherwise
$\mathbf{G}$	$g_{ij} = q_{ij}$ and $g_{ii} = -q_{ii}$	**generator**, $\mathbf{P}_t = \exp(t\mathbf{G})$, not stochastic, row sum is 0
$\mathbf{P}_t$	$p_{ij}(t) = \exp(tG)_{ij}$	the **stochastic semigroup**, transition matrix at time t, a stochastic matrix
$\mathbf{P}^Z$	$p^Z_{ij} = -g_{ij}/g_{ii} = q_{ij}/q_{ii}$	transition matrix of **jump chain**, a stochastic matrix

9.7 Birth Processes

A **birth process** with intensities $\lambda_1, \lambda_2, \ldots$ is a continuous time Markov chain $\{N_t\}_{t \geq 0}$ with nonnegative values such that

- It is non-decreasing
- There is 'single arrival'

$$\Pr(N_{t+\delta} = n + m \mid N_t = n) = \begin{cases} 1 - \lambda_n \delta + o(\delta) & m = 0 \\ \lambda_n \delta + o(\delta) & m = 1 \\ o(\delta) & m > 1 \end{cases}$$

- Conditional on N_s, the increment $N_t - N_s$ is independent of all arrivals prior to time s, where $t > s$.

A birth process with constant intensity is a Poisson process. (Poisson process is a special case of birth process.)

It has generator $\mathbf{G}$

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

9.7.1 Simple Birth Process

- We take intensities $\lambda_n = n\lambda$

$$\Pr(N_{t+\delta} = n+m \mid N_t = n) = \binom{n}{m}(\lambda\delta)^m(1-\lambda\delta)^{n-m} + o(\delta)$$

which gives

$$\Pr(N_{t+\delta} = n+m \mid N_t = n) = \begin{cases} (1-\lambda\delta)^n + o(\delta) & m = 0 \\ n\lambda\delta(1-\lambda\delta)^{n-1} + o(\delta) & m = 1 \\ o(\delta) & m > 1 \end{cases}$$

Note that the higher order terms are $o(\delta)$, so we have $1 - n\lambda\delta + o(\delta)$ and $n\lambda\delta + o(\delta)$.

The **Forward & Backward** equations are given by

$$\begin{aligned} p'_{ij}(t) &= -\lambda_j p_{ij}(t) + \lambda_{j-1} p_{i,j-1}(t) \\ p'_{ij}(t) &= -\lambda_i p_{ij}(t) + \lambda_i p_{i+1,j}(t) \end{aligned}$$

9.7.2 Deriving the Forward & Backward Equations

Note here we are looking at the transition probabilities $p_{ij}(t)$, not the value of the process N_t.

We need to use the Chapman-Kolmogorov equations

$$p_{ij}(t+\delta) = \sum_{l \in E} p_{il}(t) p_{lj}(\delta)$$

which gives the forward direction with $p_{i,j-1}(t)\lambda_{j-1}\delta + p_{ij}(t)(1-\lambda_j\delta) + o(\delta)$.

The backward direction is similar but 'splitting' in a different way.

$$p_{ij}(t+\delta) = \sum_{l \in E} p_{il}(\delta) p_{lj}(t)$$

with $p_{i+1,j}(t)\lambda_i\delta + p_{ij}(t)(1-\lambda_i\delta) + o(\delta)$.

Theorem The forward equation has a unique solution, which is also satisfied by the backward equation.

9.8 Birth-Death Processes

The **birth-death process** $\{X_t\}_{t\geq 0}$ is a continuous-time Markov chain taking values in $\mathbb{N}_0$ such that

- The birth rates λ_n and death rates μ_n are nonnegative with $\mu_0 = 0$
- The infinitesimal transition probabilities are

$$\Pr(X_{t+\delta} = n + m \mid X_t = n) = \begin{cases} 1 - (\lambda_n + \mu_n)\delta + o(\delta) & m = 0 \\ \lambda_n \delta + o(\delta) & m = 1 \\ \mu_n \delta + o(\delta) & m = -1 \\ o(\delta) & |m| > 1 \end{cases}$$

The single arrival property rids us of the cancellation of birth and death.

The **stationary distribution** of a birth-death process is

$$\pi_n = \frac{\lambda_0 \times \cdots \lambda_{n-1}}{\mu_1 \times \cdots \times \mu_n} \pi_0$$

with normalizing constant when the sum $\sum_{n=0}^{\infty} \pi_n < \infty$

$$\pi_0 = \frac{1}{\sum_{n=0}^{\infty} \pi_n}$$

9.8.1 Immigration

- Constant immigration(birth) rate λ
- Varying death rate $\mu_n = \mu n$

equivalent to a birth-death process with $\lambda_n = \lambda$ and $\mu_n = n\mu$. Same formulas above.

10 Brownian Motions

A real-valued stochastic process $B = \{B_t\}_{t\geq 0}$ is a **Brownian Motion** if

- $B_0 = 0$ almost surely
- B has independent increments
- B has stationary increments
- The increments are Gaussian, for $0 \leq s < t$

$$B_t - B_s \sim N(0, t - s)$$

- The samples paths are a.s. continuous. ($t \mapsto B_t$ is a.s. continuous)

A Brownian motion with **drift** μ and **variance** σ^2 is given by

$$Y_t = \mu t + \sigma B_t$$

then we have

$$Y_t - Y_s \sim N(\mu(t - s), \sigma^2(t - s))$$

10.1 Construction of Brownian Motion

Consider the random walk $X_n = \sum_{i=1}^{n} Y_n$ with $Y_i \in \{-1, 1\}$, from the central limit theorem, we have

$$\frac{X_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

We define the Brownian motion as a limit when $n \to \infty$

$$B_t^{(n)} = \frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} = \sqrt{t}\frac{X_k}{\sqrt{k}} \xrightarrow{d} N(0,t)$$

where k is such that $k \leq nt < k+1$ and this follows by Slutsky's Theorem. So $\dfrac{X_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} B_t$

10.2 Properties

The covariance of B_t and B_s is

$$\mathrm{Cov}(B_t, B_s) = \min(t, s)$$

10.2.1 The symmetries of Brownian motion

Let B_t be a standard Brownian motion, then each of the following is also a Brownian motion:

- (Reflection) $\{-B_t\}$
- (Translation) $\{B_{t+s} - B_s\}$
- (Rescaling) For $a > 0$, $\{aB_{t/a^2}\}$
- (Inversion) $\{tB_{1/t}\}$

10.2.2 Reflection

The **stopping-time** τ is the first time B_t hits x for some $x > 0$.

$$\tau = \inf\{t \geq 0 \mid B_t \geq x\}$$

The **reflected Brownian motion** B_t'' is given by

$$B_t'' = \begin{cases} B_t & t \leq \tau \\ x - (B_t - x) & t > \tau \end{cases}$$

This is also a Brownian motion.

The **maximum and minimum processes** of a Brownian motion are given by

$$M_t^+ = \max_{0 \leq s \leq t} B_s$$
$$M_t^- = \min_{0 \leq s \leq t} B_s$$

The distribution of M_t^+ is given by

$$\Pr(M_t^+ \geq x) = \Pr(\tau \leq t) = 2 - 2\Phi(x/\sqrt{t})$$

whence the density of τ is given by

$$p_\tau(t) = \frac{x}{\sqrt{2\pi t^3}} \exp(-\frac{x^2}{2t})$$

10.3 A model for assest prices

Let S_t be the price of an asset at time t. We can model the price as:

$$S_t = S_0 \exp\big((\mu - \sigma^2/2)t + \sigma B_t\big)$$

where S_0 is the initial price, μ is the risk-free interest rate and σ is the volatility (the instantaneous standard deviation of the stock).