PSTAT 160B Lecture Notes

Professor: Dr. Moritz Voss

Spring 2020

Bryan Xu

Introduction

These are the lecture notes for PSTAT 160B - Applied Stochastic Processes II, from the Spring 2020 quarter taught by Moritz Voss. This course covers continuous models. Continuous time stochastic processes: Poisson process, Markov chains, Renewal process, Brownian motion, including simulation of these processes. Applications to Black-Scholes model, insurance and ruin problems and related topics.

Contents

1 Poisson Processes

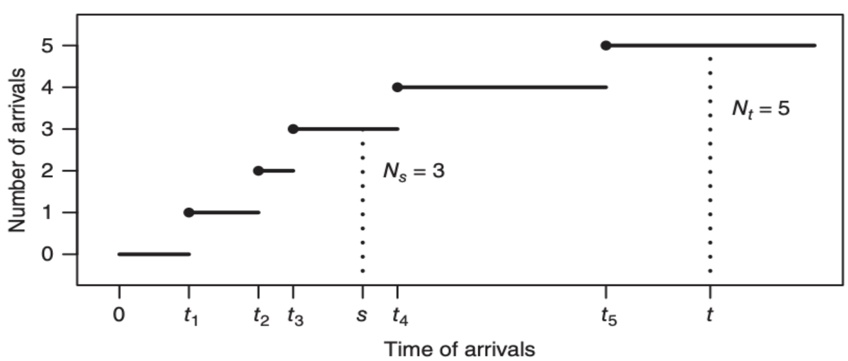
What is our motivation? We want to model the occurrence, or arrival, of events in continuous time and count them. For example, text messages arriving on a phone throughout a day, claims reported to an insurance company over a year, visits of a website over an hour...etc. We can model these with the Poisson process.

1.1 Introduction

Definition 1 (Counting Process)**.** A *counting process* $(N_t)_{t\geq 0}$ is a collection of non-negative integer-valued random variables such that if $0 \leq s \leq t$, then $N_s \leq N_t$.

Here, N_t is the number of arrivals that occur by time t, or the number of events in $[0,t]$. Typically, we have that $N_0 = 0$. Note that for each $t \geq 0$, N_t is a random variable, and thus $(N_t)_{t\geq 0}$ is a continuous-time, integer-valued stochastic process.

Here we show an example of a counting process:



Thus for all $0 \leq s \leq t$, $N_t - N_s$ is equal to the number of events in $(s,t]$.

Definition 2 (Poisson Process)**.** A *Poisson process* with parameter $\lambda > 0$ is a counting process $(N_t)_{t\geq 0}$ with the following properties:

(1) $N_0 = 0$

(2) (*Independent Increments*)
For all $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n$, the random variables $N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.

(3) (*Stationary Increments*)
For all $0 \leq s < t$, the random variable $N_t - N_s$ is Poisson distributed with parameter $\lambda(t-s)$.

In other words, a Poisson process is a counting process for which increments are independent and Poisson distributed random variables. Some important properties Poisson processes arise from the definition:

- $N_t = N_t - 0 = N_t - N_0 \sim \text{Poisson}(\lambda t)$, for all $t > 0$
- $\mathbb{E}[N_t] = \lambda t$, or in other words,

$$\frac{\mathbb{E}[N_t]}{t} = \lambda,$$

  where λ is called the *arrival rate*.

This means that, for example, $N_7 - N_4$, $N_4 - N_2$, $N_2 - N_1$ are independent random variables.

Proposition 1. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda > 0$. For $s > 0$, define

$$\tilde{N}_t = N_{t+s} - N_s \quad t \geq 0$$

Then $(\tilde{N}_t)_{t\geq 0}$ is again a Poisson process with parameter λ.

We call $(N_{t+s} - N_s)_{t\geq 0}$ a *translated process*, and it has the same probabilistic properties as $(N_t)_{t\geq 0}$.

Example 1. Starting at 6 a.m., students arrive at the gym according to a Poisson process at a rate of 30 students per hour. Find the probability that more than 65 students arrive between 9 and 11 a.m.

Solution. We have that $t = 0$ is 6 a.m., so 9 a.m. corresponds to $t = 3$ and 11 a.m. corresponds to $t = 5$. Thus we want to find $\mathbb{P}[N_5 - N_3 > 65]$. We use the properties of Poisson process then to get

$$
\begin{aligned}
\mathbb{P}[N_5 - N_3 > 65] &= \mathbb{P}[N_2 > 65] \\
&= 1 - \mathbb{P}[N_2 \leq 65] \\
&= 1 - \sum_{k=0}^{65} \mathbb{P}[N_2 = k] \quad (\text{As } N_2 \sim \text{Poisson}(2\lambda) = \text{Poisson}(60)) \\
&= e^{-60} \cdot \frac{(60)^k}{k!} \\
&\approx \boxed{0.2355}
\end{aligned}
$$

□

Example 2. You receive text messages starting at 10 a.m. every morning at the rate of 10 messages per hour according to a Poisson process.

(a) Find the probability that you will receive exactly 18 messages by noon and 70 messages by 5 p.m.

(b) Given that you received 18 messages by noon, find the probability that you will receive 70 messages by 5 p.m.

Solution. (a) We are trying to find $\mathbb{P}[N_2 = 18, N_7 = 70]$.

$$
\begin{aligned}
\mathbb{P}[N_2 = 18, N_7 = 70] &= \mathbb{P}[N_2 = 18, N_7 - N_2 = 62] \\
&= \mathbb{P}[N_2 = 18] \cdot \mathbb{P}[N_7 - N_2 = 62] \\
&= e^{-2\lambda} \cdot \frac{(2\lambda)^{18}}{18!} \cdot e^{-5\lambda} \cdot \frac{(5\lambda)^{52}}{52!} \\
&= (\text{Plug in } \lambda = 10) \\
&= \boxed{0.0045}
\end{aligned}
$$

(b) We are trying to find $\mathbb{P}[N_5 = 70 \mid N_2 = 18]$.

$$
\begin{aligned}
\mathbb{P}[N_5 = 70 \mid N_2 = 18] &= \frac{\mathbb{P}[N_2 = 18, N_7 = 70]}{\mathbb{P}[N_2 = 18]} \\
&= \frac{\mathbb{P}[N_2 = 18] \cdot \mathbb{P}[N_7 - N_2 = 52]}{\mathbb{P}[N_2 = 18]} \\
&= \mathbb{P}[N_5 = 52] \\
&= \boxed{0.0531}
\end{aligned}
$$

□

1.2 Arrival Times

Let $(N_t)_{t\geq 0}$ denote a Poisson process with parameter $\lambda > 0$. Let X_1 denote the time of the first arrival, X_2 the waiting time between the first and second arrival, and so on. Can we say something about the distribution of $X_1, X_2, \ldots$? Are they necessarily independent?

Definition 3 (Poisson Process v2)**.** Let $X_1, X_2, \ldots$ be a sequence of iid exponential random variables with parameter $\lambda > 0$. For $t > 0$, let

$$N_t = \max\{n \geq 1 : X_1 + \cdots + X_n \leq t\}$$

with $N_0 = 0$. Then $(N_t)_{t\geq 0}$ defines a *Poisson process* with parameter $\lambda > 0$. Let

$$S_n = X_1 + \cdots + X_n, \quad n = 1, 2, \ldots$$

We call $S_1, S_2, \ldots$ the *arrival times* of the process, where S_k is the time of the k-th arrival. Furthermore,

$$X_k = S_k - S_{k-1} \quad k = 1, 2, \ldots$$

is the *interarrival time* between the $(k-1)$-th and k-th arrival, with $S_0 = 0$.

The two definitions of Poisson process are mathematically equivalent. Why? Because a Poisson process is just a counting process for which interarrival times are independent and identically distributed exponential random variables.

Definition 4 (Memoryless)**.** A random variable X is *memoryless* if, for all $s, t > 0$, we have

$$\mathbb{P}[X > s + t \mid X > s] = \mathbb{P}[X > t]$$

The exponential distribution is the only continuous distribution which is memoryless.

Proposition 2. Let $X_1, \dots, X_n$ be independent exponential random variables with parameters $\lambda_1, \dots, \lambda_n$. Let $M = \min\{X_1, \dots, X_n\}$.

(a) For $t > 0$ we have
$$\mathbb{P}[M > t] = e^{-t(\lambda_1 + \cdots + \lambda_n)}$$
That is, M has exponential distribution with parameter $\lambda_1 + \cdots + \lambda_n$.

(b) For $k = 1, \dots, n$ we have
$$\mathbb{P}[M = X_k] = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}$$

Proof. Let $X_1 \sim \text{Exp}(\lambda_1), \dots, X_n \sim \text{Exp}(\lambda_n)$ be independent, and $M = \min\{X_1, \dots, X_n\}$.

(a) We get that
$$\begin{aligned}
\mathbb{P}[M > t] &= \mathbb{P}[\min\{X_1, \dots, X_n\} > t] \\
&= \mathbb{P}[X_1 > t, \dots, X_n > t] \\
&= \mathbb{P}[X_1 > t] \cdot \cdots \cdot \mathbb{P}[X_n > t] \\
&= e^{-\lambda_1 t} \cdot \cdots \cdot e^{-\lambda_n t} \\
&= e^{-(\lambda_1 + \cdots + \lambda_n)t}
\end{aligned}$$

Then the CDF of M, $F_M(t)$, is given by
$$\begin{aligned}
F_M(t) &= \mathbb{P}[M \le t] \\
&= 1 - \mathbb{P}[M > t] \\
&= 1 - e^{-(\lambda_1 + \cdots + \lambda_n)t}
\end{aligned}$$

Thus, $M \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$.

(b) For $k \in [1, n]$,
$$\begin{aligned}
\mathbb{P}[M = X_K] &= \mathbb{P}[\min\{X_1, \dots, X_n\} = X_k] \\
&= \mathbb{P}[X_1 \ge X_k, \dots, X_n \ge X_k] \\
&= \int_0^\infty \mathbb{P}[X_1 \ge X_k, \dots, X_n \ge X_k \mid X_n = t] \cdot f_{X_k}(t)\, dt
\end{aligned}$$

□

Proposition 3. For $n = 1, 2, \ldots$ let S_n be the time of the n-th arrival in a Poisson process with parameter λ. Then S_n has a gamma distribution with parameters n and λ. The density function of S_n is given by

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

The mean and variance are

$$\mathbb{E}[S_n] = \frac{n}{\lambda}, \quad \mathrm{Var}[S_n] = \frac{n}{\lambda^2}$$

Example 3. A transit center services three lines, $24X$, $12X$, and 20. The buses on each line arrive at the transit center according to three independent Poisson processes. On average, there is the $24X$ every 10 minutes, the $12X$ every 15 minutes, and the line 20 every 20 minutes.

(a) When you arrive at the transit center, what is the probability that the first bus that arrives is the $12X$?

(b) How long will you wait, on average, before some bus arrives?

(c) You have been waiting 20 minutes for the $24X$ and have watched three line 20 buses go by. What is the expected additional time you will wait for your bus $24X$?

Solution. We have three independent Poisson processes:

$$N^{(1)}: \text{ counting } 24X \text{ with } \lambda^{(1)}, \quad X_1^{(1)}: \text{ arrival time of the 1st } 24X$$

$$N^{(2)}: \text{ counting } 12X \text{ with } \lambda^{(2)}, \quad X_1^{(2)}: \text{ arrival time of the 1st } 12X$$

$$N^{(3)}: \text{ counting } 20 \text{ with } \lambda^{(3)}, \quad X_1^{(3)}: \text{ arrival time of the 1st } 20$$

We are also given/know

$$\mathbb{E}[X_1^{(1)}] = 10, \quad \mathbb{E}[X_1^{(2)}] = 15, \quad \mathbb{E}[X_1^{(3)}] = 20$$

$$X_1^{(1)} \sim \mathrm{Exp}(1/10), \quad X_1^{(2)} \sim \mathrm{Exp}(1/15), \quad X_1^{(3)} \sim \mathrm{Exp}(1/20)$$

Thus, in particular,

$$\lambda^{(1)} = \frac{1}{10}, \quad \lambda^{(2)} = \frac{1}{15}, \quad \lambda^{(3)} = \frac{1}{20}$$

(a) The probability that the first bus is the $12X$ is

$$\mathbb{P}\left[\min\{X_1^{(1)}, X_1^{(2)}, X_1^{(3)}\} = X_1^{(2)}\right] = \frac{\lambda^{(2)}}{\lambda^{(1)} + \lambda^{(2)} + \lambda^{(3)}} = \boxed{0.81}$$

(b) By Proposition 1, we know $\min\{X_1^{(1)}, X_1^{(2)}, X_1^{(3)}\} \sim \text{Exp}\left(\lambda^{(1)} + \lambda^{(2)} + \lambda^{(3)}\right)$. This has mean

$$\frac{1}{\lambda^{(1)} + \lambda^{(2)} + \lambda^{(3)}} = \boxed{4.615}$$

Thus we wait, on average, 4.615 minutes.

(c) Our waiting time is independent of line 20 bus arrivals. By memorylessness of waiting times, we will always wait

$$\frac{1}{\lambda^{(1)}} = 10 \text{ minutes,}$$

on average, regardless of elapsed waiting time.

□

Example 4. The times when goals are scored in hockey are modeled as a Poisson process. For such a process, assume that the average time between goals is 15 minutes.

(a) In a 60 minute game, find the probability that a fourth goal occurs in the last 5 minutes of the game

(b) Assume that at least three goals are scored in a game. What is the mean time of the third goal?

Solution. We know that we have $\lambda = \frac{1}{15}$. Let N_t be the number of goals by time t, in minutes.

(a) We compute

$$\begin{aligned}
\mathbb{P}[55 < S_4 \leq 60] &= \int_{55}^{60} f_{S_4}(t)\,dt \quad \text{(By Proposition 2)} \\
&= \frac{1}{6}\int_{55}^{60} \left(\frac{1}{15}\right)^4 t^3 e^{-t/15}\,dt \quad (\text{Since } S_4 \sim \text{Gamma}(n = 4, \lambda = 1/15)) \\
&= \boxed{0.068}
\end{aligned}$$

(b) We get

$$\begin{aligned}
\mathbb{E}[S_3 \mid S_3 < 60] &= \frac{1}{\mathbb{P}[S_3 < 60]} \cdot \int_0^{60} t \cdot f_{S_3}(t)\,dt \\
&= \frac{1}{\mathbb{P}[S_3 < 60]} \cdot \int_0^{60} t \cdot \frac{(\frac{1}{15})^3 t^2 e^{-t/15}}{2}\,dt \\
&= \boxed{33.461 \text{ minutes}}
\end{aligned}$$

□

1.3 Order Statistics and Arrival Times

Suppose that we know the number of jumps of a Poisson process on the time interval $[0,t]$, say $N_t = n \in \mathbb{N}$. How are the arrival times $0 < S_1 < S_2 < \cdots < S_n < t$ distributed in $[0,t]$? To answer this, we need order statistics.

Definition 5 (Order Statistics)**.** Given a set of $n \in \mathbb{N}$ real-valued i.i.d. random variables $X_1, \dots, X_n$, the *order statistics* $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are defined by sorting the values of $X_1, \dots, X_n$ in increasing order.

In other words, $X_{(1)}, \dots X_{(n)}$ are random variables satisfying

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

$X_{(k)}$ is the k-th order statistic, where $k \in [1,n]$. In particular,

$$X_{(1)} = \min\{X_1, \dots, X_n\}, \quad X_{(n)} = \max\{X_1, \dots, X_n\}$$

Note however, that $X_{(k)}$ and X_k don't necessarily have the same distribution.

Proposition 4. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda > 0$. For some $t > 0$ suppose that $N_t = n \in \mathbb{N}$ and denote by $S_1, S_2, \dots, S_n$ the corresponding arrival times. Moreover, let $U_1, \dots, U_n$ be n i.i.d. uniform random variables on $[0,t]$. Then, conditional on the event $\{N_t = n\}$ the joint distribution of $(S_1, \dots, S_n)$ is the joint distribution of the order statistics $(U_{(1)}, \dots, U_{(n)}$ of $U_1, \dots, U_n$. In other words, conditional on the event $\{N_t = n\}$ the joint density function of $(S_1, \dots, S_n)$ is given by

$$f(s_1, \dots, s_n) = \begin{cases} \frac{n!}{t^n}, & 0 \leq s_1 \leq s_2 \leq \cdots \leq s_n < t \\ 0, & \text{otherwise} \end{cases}$$

Corollary 1. Let $U_1, \dots, U_n$ be i.i.d. random variables uniformly distributed on $[0,t]$, and let $U_{(1)} \leq \cdots \leq U_{(n)}$ be the corresponding order statistics. Then

$$(S_1, \dots, S_n \mid N_t = n) \stackrel{d}{=} (U_{(1)}, \dots, U_{(n)})$$

Some important implications of the corollary:

- $\mathbb{P}[(S_1, \dots, S_n) \in A \mid N_t = n] = \mathbb{P}[(U_{(1)}, \dots, U_{(n)} \in A]$ for some set A
- $\mathbb{E}[g(S_1, \dots, S_n) \mid N_t = n] = \mathbb{E}[g(U_{(1)}, \dots, U_{(n)}]$ for some function $g : \mathbb{R}^n \to \mathbb{R}$

Example 5. Starting at time $t = 0$, patrons arrive at a restaurant according to a Poisson process with rate 20 customers per hour. If 60 people arrive by time $t = 3$, find the probability that the 60-th customer arrives in the interval $[2.9, 3]$.

Solutiom. We are trying to find $\mathbb{P}[2.9 < S_{60} < 3 \mid N_3 = 60]$. By the above proposition,

$$\begin{aligned}
\mathbb{P}[2.9 < S_{60} < 3 \mid N_3 = 60] &= \mathbb{P}[2.9 < U_{(60)} < 3] \\
&= 1 - \mathbb{P}[U_{(60)} \leq 2.9] \\
&= 1 - \mathbb{P}[U_{(1)} \leq 2.9, \ldots, U_{(60)} \leq 2.9] \\
&= 1 - \mathbb{P}[U_1 \leq 2.9, \ldots, U_{60} \leq 2.9] \\
&= 1 - \mathbb{P}[U_1 \leq 2.9] \cdot \cdots \cdot \mathbb{P}[U_{60} \leq 2.9] \\
&= 1 - (\mathbb{P}[U_1 \leq 2.9])^{60} \\
&= 1 - \left(\frac{2.9}{3}\right)^{60} \\
&= \boxed{0.869}
\end{aligned}$$

□

Example 6. Concertgoers arrive at a show according to a Poisson process with parameter $\lambda > 0$. The band starts playing at time $t > 0$. The k-th person to arrive in $[0, t]$ waits $t - S_k$ time units for the start of the concert, where S_k is the k-th arrival time. Find the expected total waiting time of concertgoers who arrive before the band starts.

Solution. By the Law of Total Expectation,

$$\mathbb{E}\left[\sum_{k=1}^{N_t}(t - S_k)\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{N_t}(t - S_k) \mid N_t\right]\right]$$

The inner part of this is a random variable, and the values of it are given by

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{N_t}(t-S_k) \mid N_t = n\right] &= \mathbb{E}\left[\sum_{k=1}^{n}(t-S_k) \mid N_t = n\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{N_t} t \mid N_t = n\right] - \mathbb{E}\left[\sum_{k=1}^{N_t} S_k \mid N_t = n\right] \\
&= nt - \mathbb{E}\left[\sum_{k=1}^{n} U_{(k)}\right] \\
&= nt - \mathbb{E}\left[\sum_{k=1}^{n} U_k\right] \\
&= nt - \sum_{k=1}^{n} \mathbb{E}[U_k] \\
&= nt - n \cdot \mathbb{E}[U_1] \\
&= nt - n \cdot \frac{t}{2} \\
&= n \cdot \frac{t}{2}
\end{aligned}$$

Then $\mathbb{E}\left[\sum_{k=1}^{N_t}(t-S_k) \mid N_t\right] = N_t \cdot \frac{t}{2}$, and we plug back in to get

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{N_t}(t-S_k) \mid N_t\right]\right] &= \mathbb{E}\left[N_t \cdot \frac{t}{2}\right] \\
&= \frac{t}{2} \cdot \mathbb{E}[N_t] \\
&= \frac{t}{2} \cdot \lambda t \\
&= \boxed{\lambda \frac{t^2}{2}}
\end{aligned}$$

□

Definition 6 (Poisson Process v3)**.** A *Poisson process* with parameter λ is a counting process $(N_t)_{t \geq 0}$ with the following properties:

a) $N_0 = 0$

b) The process has stationary and independent increments

c) $\mathbb{P}[N_h = 0] = 1 - \lambda h + o(h)$

d) $\mathbb{P}[N_h = 1] = \lambda h + o(h)$

e) $\mathbb{P}[N_h > 1] = o(h)$

What this essentially mean is that in some time interval $[0,h]$, where h is small, there occurs at most one event.

Simulating a Poisson Process

We want to simulate the trajectory of a Poisson process $(N_t)_{0\leq t\leq T}$ with parameter λ on $[0,T]$.

(a) **Method 1**:

(1) We split the interval $[0,T]$ into n small subintervals of length $\Delta t = T/n$.

(2) Generate n i.i.d. Poisson random variables $\Delta N_1, \Delta N_2, \ldots, \Delta N_n$ with parameter $\lambda \Delta t$.

(3) Let $N_0 = 0$. For each $i \in [0, n-1]$, let $N_{(i+1)\Delta t} = N_{i\Delta t} + \Delta N_{i+1}$.

(4) For each $i \in [0, n-1]$ and each $t \in [i \cdot \Delta t, (i+1) \cdot \Delta t)]$ set $N_t = N_{i\Delta t}$.

(b) **Method 2**:

(1) Let $S_0 = 0$.

(2) Generate i.i.d exponential random variables $X_1, X_2, \ldots$ with parameter λ.

(3) Let $S_n = X_1 + \cdots + X_n$ for $n = 1, 2, \ldots$

(4) For each $k = 0, 1, \ldots$ let $N_t = k$ for $S_k \leq t < \min\{S_{k+1}, T\}$.

(c) **Method 3**:

(1) Simulate the total number of arrivals N_T in $[0,T]$ from a Poisson distribution with parameter λT.

(2) Generate $N_T = n$ i.i.d. random variables $U_1, \ldots, U_n$ uniformly distributed on $(0,T)$.

(3) Sort the variables in increasing order $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ to get the arrival times $S_1 = U_{(1)}, S_2 = U_{(2)}, \ldots, S_n = U_{(n)}$.

(4) For each $k \in [0, n-1]$ set $N_t = k$ for $S_k \leq t < S_{k+1}$, and set $N_t = n$ for $S_n \leq t < T$.

1.4 Thinning and Superposition

Let's assume that babies are born at a hospital according to a Poisson process $(N_t)_{t\geq 0}$ with rate parameter $\lambda > 0$. Suppose that the babies' sex is independent of each other. The worldwide sex ratio at birth is 108 boys to 100 girls. Hence a simple estimate for the probability that any birth is a boy would be

$$p = \frac{108}{108+100} = 0.519$$

How can the number of male and female births happening at the hospital be described?

Definition 7 (Thinned Poisson Process)**.** Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda > 0$. Assume that each arrival, independent of other arrivals, is marked as a type k event with probability $p_k \in (0,1)$, for $k \in [1,n]$, where $p_1 + \cdots + p_n = 1$. Let $N_t^{(k)}$ be the number of type k events in $[0,t]$. Then $(N_t^{(k)})_{t\geq 0}$ is a Poisson process with parameter $\lambda p_k > 0$. Furthermore, the processes

$$\left(N_t^{(1)}\right)_{t\geq 0}, \left(N_t^{(2)}\right)_{t\geq 0}, \ldots, \left(N_t^{(n)}\right)_{t\geq 0}$$

are independent. Each process is called a *thinned Poisson process*.

A quick observation to be made:

$$N_t^{(1)} + N_t^{(2)} + \cdots + N_t^{(n)} = N_t$$

Then recall that if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent, then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Proposition 5 (Superposition Process)**.** Assume that

$$\left(N_t^{(1)}\right)_{t\geq 0}, \left(N_t^{(2)}\right)_{t\geq 0}, \ldots, \left(N_t^{(n)}\right)_{t\geq 0}$$

are n independent Poisson processes with respective parameters $\lambda_1, \ldots, \lambda_n > 0$. Let

$$N_t = N_t^{(1)} + N_t^{(2)} + \cdots + N_t^{(n)}, \quad t \geq 0$$

Then $(N_t)_{t\geq 0}$ is a Poisson process with parameter $\lambda = \lambda_1 + \cdots + \lambda_n$.

Going back to the hospital scenario, we assume that the babies' sex are independent of each other. Then p is the probability for a boy, and $1-p$ is the probaiblity for a girl. Then we introduce

$$M_t = \text{ number of male babies by time } t$$

$$F_t = \text{ number of female babies by time } t$$

Obviously $N_t = M_t + F_t$ for all $t > 0$. Then the joint pmf of (M_t, F_t) for $m, t \in \mathbb{N}_0$ will be

$$\begin{aligned}\mathbb{P}[M_t = m, F_t = f] &= \mathbb{P}[M_t = m, F_t = f, N_t = m + f] \\ &= \mathbb{P}[M_t = m, N_t = m + f] \\ &= \mathbb{P}[M_t = m \mid N_t = m + f] \cdot \mathbb{P}[N_t = m + f]\end{aligned}$$

The first term is equal to

$$\mathbb{P}[M_t = m \mid N_t = m + f] = \binom{m+f}{m} p^m (1-p)^f,$$

because $(M_t = m \mid N_t = m + f) \sim \text{Binom}(m + f, p)$ due to the independence of babies' sex. Then the second term has $N_t \sim \text{Poisson}(\lambda t)$, and so

$$\begin{aligned}\mathbb{P}[M_t = m \mid N_t = m + f] \cdot \mathbb{P}[N_t = m + f] &= \binom{m+f}{m} p^m (1-p)^f \cdot e^{-\lambda t} \frac{(\lambda t)^{m+f}}{(m+f)!} \\ &= e^{-\lambda p t} \frac{(\lambda p t)^m}{m!} \cdot e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^f}{f!}\end{aligned}$$

Thus M_t and F_t are independent and $M_t \sim \text{Poisson}(\lambda p t)$ and $F_t \sim \text{Poisson}(\lambda(1-p)t)$.

Example 7. Assume that births occur at a hospital at the average rate of 2 births per hour.

(a) On an 8-hour shift, what is the expectation and standard deviation of the number of female births?

(b) Find the probability that only girls were born between 2 and 5 p.m.

(c) Assume that five babies were born on the ward yesterday. Find the probability that two are boys.

Solution. We have three Poisson processes:

$(N_t)_{t \geq 0}$, a Poisson process with $\lambda = 2$

$(F_t)_{t \geq 0}$, a Poisson process with $\lambda^{(F)} = \lambda(1-p) = 2(0.481) = 0.962$

$(M_t)_{t \geq 0}$, a Poisson process with $\lambda^{(M)} = \lambda p = 2(0.519) = 1.038$

We also know that $(F_t)_{t \geq 0}$ and $(M_t)_{t \geq 0}$ are independent.

(a) We want to compute $\mathbb{E}[F_8]$ and $\sqrt{\text{Var}[F_8]}$. Since $F_8 \sim \text{Poisson}(\lambda^{(F)} \cdot 8)$,

$$\mathbb{E}[F_8] = \lambda^{(F)} \cdot 8 = 8(0.962) = \boxed{7.696}$$

$$\sqrt{\text{Var}[F_8]} = \sqrt{\lambda^{(F)} 8} = \sqrt{7.696} = \boxed{2.774}$$

(b) We want to find $\mathbb{P}[F_3 > 0, M_3 = 0]$. We can compute this:

$$\begin{aligned}\mathbb{P}[F_3 > 0, M_3 = 0] &= \mathbb{P}[F_3 > 0] \cdot \mathbb{P}[M_3 = 0] \\ &= (1 - \mathbb{P}[F_3 = 0])\mathbb{P}[M_3 = 0] \\ &= \left(1 - e^{-\lambda^{(F)}3}\right)e^{-\lambda^{(M)}3} \\ &= (0.944) \cdot (0.044) \\ &= \boxed{0.042}\end{aligned}$$

(c) We want to find $\mathbb{P}[M_{24} = 2 \mid N_{24} = 5]$. Using the fact that $(M_{24} \mid N_{24} = n) \sim \text{Binom}(n, p)$,

$$\begin{aligned}\mathbb{P}[M_{24} = 2 \mid N_{24} = 2] &= \binom{5}{2} p^2 (1-p)^3 \\ &= \binom{5}{2} (0.519)^2 (0.481)^3 \\ &= \boxed{0.30}\end{aligned}$$

□

1.5 Compound Poisson Process

Let's say that the arrival of claims which are reported to an insurance agency during a specific time interval $[0, t]$ can be modeled by a Poisson process. However, each of these claims typically come with a different and in advance not known damage sum. How can we model the total amount of money needed by an insurance company to cover all claims which are reported by time t?

Definition 8 (Compound Poisson Process)**.** Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda > 0$. Furthermore, let $Y_1, Y_2, Y_3, \ldots$ be a sequence of i.i.d. random variables with density function f_Y, also independent of $(N_t)_{t\geq 0}$. Then

$$C_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0$$

is called a *compound Poisson process* with jump-distribution f_Y and jump-intensity λ.

What this means is that C_t is a random sum, with $(C_t)_{t\geq 0}$ having jumps of random height given by $Y_1, Y_2, \ldots$.

Example 8. Going back to the insurance company problem mentioned above, an interpretation in terms of a compound Poisson process would be:

- N_t: the number of claims reported by time t
- Y_i: the dollar amount of the i-th claim (or the height of the i-th jump) following a given distribution with density function f_Y, or

$$\mathbb{P}[Y_i \in [a,b]] = \int_a^b f_Y(x)\,dx, \quad i = 1,2,\ldots$$

- C_t: total sum of claims by time t

It is important to see that a compound Poisson process $(C_t)_{t\geq 0}$ is again a stochastic process with stationary and independent increments:

(1) $C_0 = 0$

(2) For all $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n$, the random variables $C_{t_2} - C_{t_1}, C_{t_3} - C_{t_2}, \ldots, C_{t_n} - C_{t_{n-1}}$ are independent

(3) For all $0 \leq s < t$ the distribution of the random variable $C_t - C_s$ only depends on $t - s$

However, in contrast with a Poisson process, $C_t - C_s$ is not Poisson distributed anymore. Here the distribution depends on the jump distribution f_Y.

Example 9. Let $(C_t)_{t\geq 0}$ be a compound Poisson process. Show that

(a) $\mathbb{E}[C_t] = \mathbb{E}[N_t] \cdot \mathbb{E}[Y_1] = \lambda t \cdot \mathbb{E}[Y_1]$.

(b) $\mathrm{Var}[C_t] = \mathbb{E}[N_t] \cdot \mathbb{E}[Y_1^2] = \lambda t \cdot \mathbb{E}[Y_1^2]$.

Solution. □

Now, given a compound Poisson process $(C_t)_{t\geq 0}$ with claims $Y_1, Y_2, \ldots$, what if we want to count the number of claims with losses below and above a certain threshold? This will result again in two independent Poisson processes, much like a thinned Poisson process.

Proposition 6. Let $(C_t)_{t\geq 0}$ be a compound Poisson process with jump-intensity λ and jump-distribution f_Y. Furthermore, let $A_1, A_2, \dots, A_n$ be pair-wise disjoint subsets of $\mathbb{R}$, and let

$$N_t^{(k)} = \sum_{i=1}^{N_t} \mathbb{1}_{\{Y_i \in A_k\}}, \quad t \geq 0$$

for all $k \in [0, n]$. Then

$$\left(N_t^{(1)}\right)_{t\geq 0}, \left(N_t^{(2)}\right)_{t\geq 0}, \dots, \left(N_t^{(n)}\right)_{t\geq 0}$$

are independent Poisson processes with rate $\lambda \cdot \mathbb{P}[Y_1 \in A_k]$, respectively.

Example 10. If we again consider the insurance company example, and our threshold is \$1000, then we can interpret the problem as follows:

- $A_1 = [0, 1000)$, $A_2 = [1000, \infty)$
- $N_t^{(1)}$: the number of claims reported by time t with damage sum below 1000. The arrival rate is

$$\lambda \cdot \mathbb{P}[Y_1 < 1000] = \lambda \cdot \int_0^{1000} f_Y(x)\,dx$$

- $N_t^{(2)}$: the number of claims reported by time t with damage sum above 1000. The arrival rate is

$$\lambda \cdot \mathbb{P}[Y_1 \geq 1000] = \lambda \cdot \int_{1000}^{\infty} f_Y(x)\,dx$$

1.6 Nonhomogeneous Poisson Process

The assumption that arrivals occur at a constant arrival rate λ, which is independent of time, is a very unrealistic assumption. To remedy this, we can make λ depend on t, to get a *intensity function*, $\lambda(t)$, for $t \geq 0$.

Definition 9 (Nonhomogeneous Poisson Process)**.** A counting process $(N_t)_{t\geq 0}$ is a *nonhomogeneous Poisson process* with intensity function $\lambda(t)$, if

(1) $N_0 = 0$

(2) For all $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n$, the random variables $N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \dots, N_{t_n} - N_{t_{n-1}}$ are independent

(3) For all $0 \leq s < t$ the distribution of the random variable $N_t - N_s$ is Poisson distributed with parameter $\int_s^t \lambda(u)\,du$

Now the distribution of increments is no longer stationary. However, if $\lambda(t) = \lambda$ is constant, then we just have a normal Poisson process again.

Example 11. Students arrive at the UCen for lunch according to a nonhomogeneous Poisson process. The doors open at 11 a.m. The arrival rate increases linearly from 100 to 200 students per hour between 11 a.m. and noon. The rate stays constant for the next 2 hours, and then decreases linearly down to 100 from 2 to 3 p.m. Find the probability that there are at least 400 people in the cafeteria between 11:30 a.m. and 1:30 p.m.

Solution. We have that our intensity function is given by

$$\lambda(t) = \begin{cases} 100 + 100t, & 0 \leq t \leq 1 \\ 200, & 1 \leq t \leq 3 \\ 500 - 100t, & 3 \leq t \leq 4 \end{cases}$$

We want to compute $\mathbb{P}[N_{2.5} - N_{0.5} \geq 400]$. Here $N_{2.5} - N_{0.5}$ is Poisson distributed with parameter

$$\int_{0.5}^{2.5} \lambda(t)\,dt = \int_{0.5}^{1} (100 + 100t)\,dt + \int_{1}^{2.5} 200\,dt = 387.5$$

Then

$$\begin{aligned} \mathbb{P}[N_{2.5} - N_{0.5} \geq 400] &= 1 - \mathbb{P}[N_{2.5} - N_{0.5} \leq 399] \\ &= 1 - \sum_{k=0}^{399} \mathbb{P}[N_{2.5} - N_{0.5} = k] \\ &= 1 - \sum_{k=0}^{399} e^{-387.5} \cdot \frac{(387.5)^k}{k!} \\ &= \boxed{0.269} \end{aligned}$$

□

1.7 Spatial Poisson Process

What if we want to model the random distribution of points in two or higher dimensional spaces? Some examples of this are the location of trees in a forest, galaxies in the night sky, or cancer clusters across the US, etc. We first define our notation:

- $d \geq 1$ and $A \subset \mathbb{R}^d$
- The random variable N_A is the number of points in the set A
- $|A|$ = size of A (length in $\mathbb{R}^1$, area in $\mathbb{R}^2$, volume in $\mathbb{R}^3$)

Definition 10 (Spatial Poisson Process)**.** A collection of random variables $(N_A)_{A\subset\mathbb{R}^d}$ is a *spatial Poisson process* in dimension d with parameter $\lambda > 0$ if

(1) Whenever A and B are disjoint sets, N_A and N_B are independent random variables

(2) For each bounded set $A \subset \mathbb{R}^d$, N_A has a Poisson distribution with parameter $\lambda|A|$.

This generalizes the regular one-dimensional Poisson process where $N_t = N_{[0,t]} \sim \text{Poisson}(\lambda \cdot |[0,t]|)$. It is a fact that given a bounded set $A \subset \mathbb{R}^d$, conditional on $N_A = n$, meaning there are n points in N_A, the locations of the points are uniformly distributed in A. Hence a spatial Poisson process is a model of complete spatial randomness.

1.8 Renewal Process

Recall that a Poisson process has independent and identically distributed interarrival times following an exponential distribution with parameter $\lambda > 0$. A renewal process is a counting process where interarrival times follow a more general distribution. Hence a renewal process is a generalization of the Poisson process.

Definition 11 (Renewal Process)**.** Let $L_1, L_2, \ldots$ be a sequence of positive, independent and identically distributed random variables. We refer to L_i as the i-th holding time. Define for all $n \geq 1$,

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

Each J_n is referred to as the n-th jump time and $[J_n, J_{n+1}]$ is called renewal interval. Define

$$X_t = \sum_{n=1}^{\infty} \mathbb{1}_{\{J_n \leq t\}}, \quad t \geq 0$$

Then X_t represents the number of jumps that have occurred by time $t > 0$, and the continuous-time stochastic process $(X_t)_{t\geq 0}$ is called a *renewal process*.

We can see that a Poisson process is a renewal process where $L_1, L_2, \ldots$ are i.i.d. Exp(λ)-distributed. A renewal process, for example, can also be used to model the number of breakdowns of worn-out machinery:

- $L_1, L_2 \ldots$: lifetime of machinery/device
- J_n: breakdown time of n-th device
- X_t: number of replaced devices by time $t > 0$

Further Generalizations of Poisson Process

Poisson processes and compound Poisson processes are building blocks for more general classes of processes:

- *Lévy processes*: Continuous-time stochastic processes with stationary and independent increments.
- *Mixed Poisson processes*: The rate parameter Λ is itself a random variable (Cox process, doubly stochastic process).

$$(N_t \mid \Lambda = \lambda) \sim \text{Poisson}(\lambda t)$$

- Even more general has the model intensity function $\lambda(t)$ itself as a continuous time stochastic process.
- *Hawkes process*: The intensity function $\lambda(t)$ depends on $(N_s)_{0\leq s\leq t}$ itself (self-exciting process).

2 Continuous-Time Markov Chains

We want to extend the discrete-time Markov chain model that we developed in 160A. A continuous-time process allows us to model not only the transitions between states, but also the duration of time in each state. The state space $\mathcal{S}$ remains discrete, and the Markov property still continues to hold, but time evolves continuously.

2.1 Introduction

Let's say that we have a three-state weather chain Markov model with state space $\mathcal{S} = \{\text{rain, snow, clear}\}$. Changes in weather states are described by the transition matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \\ 1/4 & 3/4 & 0 \end{pmatrix}$$

Then the continuous-time extension would be: Duration of time for each weather state is exponentially distributed with parameters λ_r, λ_s, and λ_c:

- Rainfall lasts, on average, 3 hours at a time ($\lambda_r = 1/3$)
- Snow lasts, on average, 6 hours at a time ($\lambda_s = 1/6$)
- Weather stays clear, on average, 12 hours at a time ($\lambda_c = 1/12$)

We denote by X_t the weather at time $t \geq 0$. Then $(X_t)_{t\geq 0}$ is a continuous-time Markov chain. The matrix P, the exponential time parameters $(\lambda_r, \lambda_s, \lambda_c)$, and initial distribution completely specify the process.

Definition 12 (Markov Property)**.** A continuous-time stochastic process $(X_t)_{t\geq 0}$ with discrete state space $\mathcal{S}$ is a *continuous-time Markov chain* if

$$\mathbb{P}[X_{t+s} = j \mid X_s = i, X_u = x_u \text{ for some } 0 \leq u \leq s] = \mathbb{P}[X_{t+s} = j \mid X_s = i]$$

for all $s, t \geq 0$, $i, j, x_u \in \mathcal{S}$, and $0 \leq u \leq s$. The process is said to be *time-*

homogeneous if this probability does not depend on s. That is,

$$\mathbb{P}[X_{t+s} = j \mid X_s = i] = \mathbb{P}[X_t = j \mid X_0 = i]$$

Note: In this class, we only consider the time-homogeneous case.

Definition 13 (Transition Function)**.** For each $t \geq 0$ the transition probabilities of a (homogeneous) continuous-time Markov chain can be arranged in a matrix function

$$P_{ij}(t) = \mathbb{P}[X_t = j \mid X_0 = i], \quad t \geq 0$$

We call the matrix $P(t) = (P_{ij}(t))_{i,j\in\mathcal{S}}$ the *transition function*.

Keep in mind that for each $t \geq 0$, we have a matrix $P(t)$. Hence the given P in the weather example is *not* the transition function. We also have that $P(0) = I$, the identity matrix.

Example 12. A Poisson process $(N_t)_{t\geq 0}$ with parameter λ is a continuous-time Markov chain with countably infinite state space $\mathcal{S} = \mathbb{N}_0$. The Markov property holds as a consequence of stationary and independent increments. Compute the transition function $P(t)$ for all $t \geq 0$.

Solution. The transition probability function can be calculated to be, for $0 \leq i \leq j$,

$$\begin{aligned}
P_{ij}(t) &= \mathbb{P}[N_{t+j} = j \mid N_j = i] \\
&= \frac{\mathbb{P}[N_s = i, N_{t+j} = j]}{\mathbb{P}[N_s = i]} \\
&= \frac{\mathbb{P}[N_s = i, N_{t+j} - N_s = j - i]}{\mathbb{P}[N_s = i]} \\
&= \frac{\mathbb{P}[N_s = i]\mathbb{P}[N_{t+s} - N_s = j - i]}{\mathbb{P}[N_s = i]} \\
&= \mathbb{P}[N_t - N_0 = j - 1] \\
&= e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}
\end{aligned}$$

Hence for all $t \geq 0$,

$$\boxed{P(t) = \begin{pmatrix} e^{-\lambda t} & e^{-\lambda t}\frac{(\lambda t)}{1!} & e^{-\lambda t}\frac{(\lambda t)^2}{2!} & \cdots \\ 0 & e^{-\lambda t} & e^{-\lambda t}\frac{(\lambda t)}{1!} & \cdots \\ 0 & 0 & e^{-\lambda t} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}}$$

□

Proposition 7 (Chapman-Kolmogorov Equations)**.** For a continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $P(t)$, it holds that

$$P(s+t) = P(s) \cdot P(t)$$

for all $s, t \geq 0$. That is,

$$P_{ij}(s+t) = [P(s)P(t)]_{ij} = \sum_{k \in \mathcal{S}} P_{ik}(s) \cdot P_{kj}(t)$$

for all states $i, j \in \mathcal{S}$ and $s, t \geq 0$.

Proof. We have that

$$\begin{aligned}
P_{ij}(s+t) &= \mathbb{P}[X_{s+t} = j \mid X_0 = i] \\
&= \sum_k \mathbb{P}[X_{s+t} = j \mid X_0 = i, X_j = k] \cdot \mathbb{P}[X_s = k \mid X_0 = i] \\
&= \mathbb{P}[X_{s+t} = j \mid X_j = k] \cdot P_{ik}(s) \\
&= \mathbb{P}[X_t = j \mid X_0 = k] \cdot P_{ik}(s) \\
&= \sum_k P_{ik}(s) P_{kj}(t) \\
&= (P(s)P(t))_{ij}
\end{aligned}$$

Then the special case gives us that $P(s+t) = P(s)P(t)$, as desired. □

2.2 Holding Times

Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain with finite or countably infinite state space $\mathcal{S}$.

Definition 14 (Holding Time)**.** We denote by T_i the *holding time* at state i, that is, the length of time a continuous-time Markov chain stays in state i before transitioning to a new state.

As a consequence of time-homogeneity and the Markov property,

Proposition 8. The holding times T_i for any state i are independent and exponentially distributed with holding time parameter $q_i \in [0, \infty)$.

Consequently, expected duration of time the continuous-time Markov chain spends in state i is given by

$$\mathbb{E}[T_i] = \frac{1}{q_i}$$

Definition 15 (Absorbing State)**.** A state $i \in \mathcal{S}$ is called an *absorbing state* if $q_i = 0$. A Markov chain with at least one absorbing state is called an *absorbing Markov chain*.

When an absorbing state is visited the process never leaves that state.

Definition 16 (Embedded Chain)**.** Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain and let $\tilde{P} = (p_{ij})_{i,j\in\mathcal{S}}$ describe the discrete transition probabilities from state i to state j. If we ignore time, and just watch state to state transitions, we see a sequence $Y_0, Y_1, \ldots$ where Y_n is the n-th state visited by the continuous process $(X_t)_{t\geq 0}$. The sequence $Y_0, Y_1, \ldots$ is a discrete-time Markov chain called the *embedded chain* with transition probability matrix $\tilde{P}$.

$\tilde{P}$ is a stochastic matrix whose diagonal entries are 0. Hence we can piece together the evolution of a continuous-time Markov chain:

(1) Starting from i, the process stays in i for an exponentially distributed length of time (on average $1/q_i$ time units)

(2) Then it jumps to a new state $j \neq i$ with probability p_{ij}

(3) Then process stays in j for an exponentially distributed length of time (on average $1/q_j$ time units)

(4) Then it jumps to a new state $k \neq j$ with probability p_{jk}

(5) And so on...

Example 13. For a Poisson process with parameter λ specify the holding time parameters and the transition matrix of the embedded chain.

Solution. The holding time paraemters are

$$q_i = \lambda, \quad \text{for all states } i \in \mathbb{N}_0$$

Then the transition matrix of the embedded chain will be

$$\tilde{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

□

2.3 Transition Rates

We can also describe a continuous-time Markov chain by specifying transition rates between pairs of states. Consider for each state i the pairs (i, j) for all states $j \neq i$ which can be visited after i. Each pair (i, j) is associated with an exponential alarm clock T_{ij} with parameter q_{ij}, so $T_{ij} \sim \text{Exp}(q_{ij})$. All alarm clocks are independent and ring after an exponentially distributed length of time.

When our Markov chain hits state i, all clocks T_{ij} are started simultaneously and the first alarm that rings determines the next state to visit. If T_{ij} clock rings first, then $T_i = T_{ij}$ and the process moves from state i to state j. Hitting state j will trigger a new set of exponential alarm clocks with rates $q_{j1}, q_{j2}, \ldots$. The first alarm that rings determines the time spent in j and the next state to hit after state j.

Definition 17 (Transition Rates)**.** The q_{ij} are called the *transition rates* or *instantaneous rates* of the continuous-time Markov chain.

As a convention, if $q_{ij} = 0$, then state j cannot be hit from state i. By Proposition 2, we get that

- Holding time parameters:

$$q_i = \sum_{k=1}^{n_i} q_{ik}$$

- Transition probabilities of the embedded chain $\tilde{P} = (p_{ij})_{i,j \in \mathcal{S}}$:

$$p_{ij} = \frac{q_{ij}}{\sum_{k=1}^{n_i} q_{ik}} = \frac{q_{ij}}{q_i}$$

- In particular, $q_{ij} = q_i \cdot p_{ij}$.

Definition 18 (Generator Matrix)**.** The matrix Q with entries

$$Q_{ij} = \begin{cases} q_{ij}, & i \neq j \\ -q_i, & i = j \end{cases} \quad (i, j \in \mathcal{S})$$

with transition rates q_{ij} and holding rates $q_i = \sum_{j \neq i} q_{ij}$ is called the *generator matrix*.

This is the most important matrix for continuous-time Markov chains. However, Q is not a stochastic matrix; the q_{ij} are not probabilities.

Example 14. It is time for students to register for classes, and a line is forming at the registrar's office for those who need assistance. It takes the registrar an exponentially distributed amount of time to service each student, at the rate of one student every 5 minutes. Students arrive at the office and get in line according to a Poisson process at the rate of one student every 4 minutes. Line size is capped at 4 people. If an arriving student finds that there are already 4 people in line, then they try again later. As soon as there is at least one person in line, the registrar starts assisting the first available student. The arrival times of the students are independent of the registrar's service time.

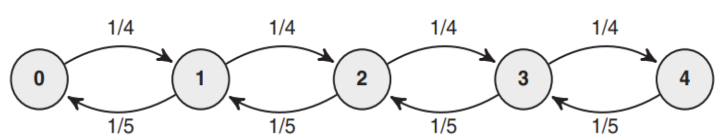
(a) Provide the transition rate graph with the transition rates.

(b) Compute the holding time parameters.

(c) Provide the generator matrix.

(d) Provide the transition probability matrix $\tilde{P}$ for the embedded chain.

Solution. (a) Let $(X_t)_{t \geq 0}$ be our continuous time Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4\}$, and X_t be the number of students in line at time t. The students' arrival is modeled with a Poisson process with $\lambda = \frac{1}{4}$, and the registrar has service rate $\mu = \frac{1}{5}$. Our transition rate graph will be



(b) The holding time parameters:

$$T_i = \text{duration of time spent in state } i \in [0, 4], \quad T_i \sim \text{Exp}(q_i)$$

Since

$$q_i = \sum_k q_{ik},$$

this gives us

$$q_0 = q_{01} = \frac{1}{4}, \quad q_1 = q_{12} + q_{10} = \frac{1}{4} + \frac{1}{5} = \frac{9}{20}, \quad q_2 = q_3 = \frac{9}{20}, \quad q_4 = q_{43} = \frac{1}{5}$$

(c) The generator matrix is

$$Q = \begin{pmatrix} -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{5} & -\frac{9}{20} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & -\frac{9}{20} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{5} & -\frac{9}{20} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{5} & -\frac{1}{5} \end{pmatrix}$$

(d) The transition probability matrix $\tilde{P}$ for the embedded chain will be given by

$$p_{ij} = \frac{q_{ij}}{q_i}, \quad i, j \in \mathcal{S}$$

Hence

$$\tilde{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \\ \frac{4}{9} & 0 & \frac{5}{9} & 0 & 0 & \\ 0 & \frac{4}{9} & 0 & \frac{5}{9} & 0 & \\ 0 & 0 & \frac{4}{9} & 0 & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 1 & 0 & \end{pmatrix}$$

□

2.4 Infinitesimal Generator

Given a generator matrix Q of a continuous-time Markov chain, how do we get the transition function $P(t)$? We examine the infinitesimal behavior of $P_{ij}(t)$ as $t \downarrow 0$.

- Case 1: $i \neq j$

$$\lim_{t \downarrow 0} \frac{\mathbb{P}[X_t = j \mid X_0 = i]}{t} = \lim_{t \downarrow 0} \frac{P_{ij}(t) - P_{ij}(0)}{t} = P'_{ij}(0) = q_{ij},$$

where $P_{ij}(0) = 0$.

- Case 2: $i = j$

$$P'_{ii}(0) = \lim_{t \downarrow 0} \frac{P_{ii}(t) - P_{ii}(0)}{t} = \lim_{t \downarrow 0} \frac{-\sum_{j \neq i} P_{ij}(t)}{t} = -\sum_{j \neq i} q_{ij} = -q_i,$$

where $P_{ii}(0) = 1$.

Proposition 9. The generator matrix Q satisfies the property

$$Q = P'(0)$$

where $P(t)$ is the transition function.

As said before, Q is not a stochastic matrix; q_{ij} are not probabilities. However, in terms of infinitesimals, for $i \neq j$,

$$\mathbb{P}[X_{t+\Delta t} = j \mid X_t = i] \approx q_{ij} \cdot \Delta t, \quad \Delta t \text{ very small}$$

Hence q_{ij} is an instantaneous rate from i to j.

For discrete-time Markov chains, there is no generator matrix, because all probabilistic properties are captured by the transition probability matrix. For continuous-time Markov chains, Q gives a complete description of the dynamics of the process. As shown before, we can derive Q from $P(t)$ via the proposition, but in a modeling context we are usually given Q and we want to find $P(t)$. We can do this by solving a coupled system of linear ordinary differential equations.

Proposition 10 (Chapman-Kolgomorov Forward/Backward Equations)**.** A continuous-time Markov chain with transition function $P(t)$ and infinitesimal generator Q satisfies the *forward equation*

$$P'(t) = P(t) \cdot Q$$

and the *backward equation*

$$P'(t) = Q \cdot P(t)$$

with $P(0) = I$, the identity matrix. Equivalently, for all states i and j,

$$P'_{ij}(t) = \sum_k P_{ik}(t) q_{kj} = -P_{ij}(t) q_j + \sum_{k \neq j} P_{ik}(t) q_{kj}$$

and

$$P'_{ij}(t) = \sum_k q_{ik} P_{kj}(t) = -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik} P_{kj}(t).$$

Proof. □

Example 15. Consider a general two-state continuous-time Markov chain with state space $\mathcal{S} = \{1, 2\}$ and transition rates $q_{12} = \lambda$ and $q_{21} = \mu$. Show that the transition function $P(t)$ is given by

$$P(t) = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix} \quad (t \geq 0)$$

Solution. The generator matrix will be

$$Q = \begin{pmatrix} -q_1 & q_{12} \\ q_{21} & -q_2 \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Then

$$P(t) = \begin{pmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{pmatrix} = \begin{pmatrix} P_{11}(t) & 1 - P_{11}(t) \\ 1 - P_{22}(t) & P_{22}(t) \end{pmatrix}$$

By the forward equation, $P'(t) = P(t)Q$, so

$$\begin{pmatrix} P'_{11}(t) & P'_{12}(t) \\ P'_{21}(t) & P'_{22}(t) \end{pmatrix} = \begin{pmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{pmatrix} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Then we get a system of equations:

$$\begin{cases} P'_{11}(t) = -\lambda P_{11}(t) + \mu(1 - P_{11}(t)) \\ P'_{22}(t) = -\mu P_{22}(t) + \lambda(1 - P_{22}(t)) \end{cases}$$

However, since we know that the solution to the linear ODE $f'(x) = af(x) + b$, $f(0) = c$, with constants $a, b, c \in \mathbb{R}$ will be given by

$$f(x) = \left(c + \frac{b}{a}\right) \cdot e^{ax} - \frac{b}{a}$$

Thus

$$P_{11}(t) = \frac{\mu}{\lambda + \mu} + \left(1 - \frac{\mu}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}$$

$$P_{22}(t) = \frac{\lambda}{\lambda + \mu} + \left(\frac{\mu}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}$$

One can then easily solve for P_{12} and P_{21}. □

The backward equation $P'(t) = Q \cdot P(t)$ is a matrix-valued linear ODE. For scalars, the solution to

$$f'(t) = q \cdot f(t), \quad f(0) = 1$$

is just given by $f(t) = e^{qt}$. For matrices, it is a different case.

Definition 19 (Matrix Exponential)**.** Let A be a $k \times k$ quadratic matrix. The *matrix exponential* e^A is a $k \times k$ matrix defined as

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2} A^2 + \frac{1}{6} A^3 + \cdots$$

Some properties for square matrices $A, B \in \mathbb{R}^{n\times n}$:

(1) The matrix e^A is well defined because the series converges.

(2) $e^0 = I$ with zero matrix $0 \in \mathbb{R}^{k\times k}$.

(3) $e^A e^{-A} = I$.

(4) $e^{(s+t)A} = e^{sA}e^{tA}$ for $s, t \in \mathbb{R}$.

(5) If $AB = BA$, then $e^{A+B} = e^A \cdot e^B = e^B \cdot e^A$.

(6) $\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}\cdot A$.

(7) If $D \in \mathbb{R}^{n\times n}$ is diagonal with diagonal entries $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then

$$e^D = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix}$$

Proposition 11. For a continuous-time Markov chain with transition function $P(t)$ and infinitesimal generator Q, it holds that

$$P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{1}{n!}(tQ)^n = I + tQ + \frac{t^2}{2}Q^2 + \frac{t^3}{6}Q^3 + \cdots$$

In particular, e^{tQ} is the unique solution to the forward/backward equations.

Computing the matrix exponential is often challenging; we have to use numerical approximation methods. In general, there does not exist a closed-form for the transition function $P(t)$, unless Q is diagonalizable. Recall from linear algebra:

- A square matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* if there exists an invertible matrix $S \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that

$$A = SDS^{-1}$$

- If $A \in \mathbb{R}^{n\times n}$ is diagonalizable, then the entries of D are the eigenvalues of A and the columns of S are corresponding eigenvectors.
- If $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, then A is diagonalizable.

Using this fact,

Proposition 12. If the generator matrix Q is diagonalizable with $Q = SDS^{-1}$, then the transition function is given by

$$P(t) = e^{tQ} = Se^{tD}S^{-1}.$$

Proof. Assume $Q = SDS^{-1}$. Then

$$\begin{aligned} P(t) &= e^{tQ} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(tQ)^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(tSDS^{-1})^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}t^n(SD^nS^{-1}) \\ &= S\left(\sum_{n=0}^{\infty} \frac{1}{n!}(tD)^n\right)S^{-1} \\ &= Se^{tD}S^{-1} \end{aligned}$$

□

Example 16. For the two-state Markov chain with state space $\mathcal{S} = \{1, 2\}$ and generator matrix

$$Q = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$$

compute the transition function $P(t)$ via diagonalization of Q.

Solution. First we compute the eigenvalues:

$$\det(Q - \lambda I) = \det\begin{pmatrix} -1-\lambda & 1 \\ 2 & -2-\lambda \end{pmatrix} = \lambda(\lambda + 3)$$

$$\implies \lambda_1 = 0, \quad \lambda_2 = -3$$

Then we find the eigenvectors:

- $\lambda_1 = 0$:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

  We will always have $\lambda_1 = 0$ with corresponding eigenvector $\binom{1}{1}$ because the rows sum to zero.

- $\lambda_2 = -3$: We solve

$$Qv_2 = -3v_2 \iff (Q+3I)v_2 = 0 \iff \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} v_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\implies v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

Now we calculate the required matrices for the matrix exponential. We know that

$$D = \begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$$

Using Gaussian elimination, we calculate S^{-1} to be

$$S^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$$

Finally,

$$\begin{aligned} P(t) &= \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} e^{0t} & 0 \\ 0 & e^{-3t} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 1 & e^{-3t} \\ 1 & -2e^{-3t} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \boxed{\frac{1}{3} \begin{pmatrix} 2+e^{-3t} & 1-e^{-3t} \\ 2-2e^{-3t} & 1+2e^{-3t} \end{pmatrix}} \end{aligned}$$

□

2.5 Long-Term Behavior

Definition 20 (Limiting Distribution)**.** A probability distribution π is the *limiting distribution* of a continuous-time Markov chain if for all states i and j,

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

π_j is the long-term proportion of time the chain spends in state j. In others words,

$$\mathbb{P}[X_t = j] \approx \pi_j, \quad t \text{ large}$$

Definition 21 (Stationary Distribution)**.** A probability distribution π is a *stationary distribution* if

$$\pi^T \cdot P(t) = \pi^T, \quad \forall t \geq 0$$

That is, for all states j,

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i \cdot P_{ij}(t), \quad \forall t \geq 0$$

The limiting distribution, if it exists, is a stationary distribution, but not vice versa.

Definition 22 (Irreducibility)**.** A continuous-time Markov chain is *irreducible* if for all states i and j it holds that $P_{ij}(t) > 0$ for some $t > 0$.

Proposition 13. A finite-state continuous-time Markov chain is irreducible if all the holding time parameters are strictly positive, i.e., $q_i > 0$ for all $i \in \mathcal{S}$.

Unlike the discrete case, periodicity is not an issue in continuous-time.

Theorem 1 (Fundamental Limit Theorem)**.** Let $(X_t)_{t\geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition function $P(t)$. Then there exists a unique stationary distribution π, which is the limiting distribution. That is, for all j,

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j, \quad \text{for all initial } i$$

Equivalently,

$$\lim_{t\to\infty} P(t) = \Pi,$$

where Π is a matrix all of whose rows are equal to π.

Example 17. Consider the general two-state continuous-time Markov chain with state space $\mathcal{S} = \{1, 2\}$, and transition rates $q_{12} = \lambda$ and $q_{21} = \mu$. We know that the transition function is

$$P(t) = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix} \quad (t \geq 0)$$

Solution. We have

$$\begin{aligned} \lim_{t\to\infty} P(t) &= \lim_{t\to\infty} \frac{1}{\lambda+\mu} \begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix} \\ &= \frac{1}{\mu+\lambda} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix} \end{aligned}$$

Hence the limiting distribution is

$$\boxed{\pi^T = \begin{pmatrix} \frac{\mu}{\mu+\lambda} & \frac{\lambda}{\mu+\lambda} \end{pmatrix}}$$

Note that π is also the unique stationary distribution of this irreducible Markov chain with finite state space. □

Proposition 14. A probability distribution π is a stationary distribution of a continuous-time Markov chain with generator Q if and only if

$$\pi^T \cdot Q = 0.$$

That is,

$$\sum_{i \in \mathcal{S}} \pi_i \cdot Q_{ij} = 0, \quad \forall j \in \mathcal{S}$$

Proof. $(\Longrightarrow)$: Let π be a stationary distribution. Then

$$\pi^T P(t) = \pi^T, \quad \forall t \geq 0$$

We differentiate with respect to t to get

$$\pi^T P'(t) = 0, \quad \forall t \geq 0$$

Then we plug in $t = 0$ to get

$$0 = \pi^T P'(0) = \pi^T Q$$

$(\Longleftarrow)$: Let π satisfy $\pi^T Q = 0$. Then we multiply $P(t)$ on the right for both sides, and then by Kolgomorov's Backward Equation,

$$0 = \pi^T Q P(t) = \pi^T P'(t) \implies \pi^T P(t) = \text{constant}, \quad \forall t \geq 0$$

In particular, since $\pi^T P(0) = \pi^T$, it must hold that $\pi^T P(t) = \pi^T$. Hence π is a stationary distribution. □

If we rearrange the latter equation, we get what is called the *Global Balance Equation*:

$$\sum_{i \neq j} \pi_i q_{ij} = \pi_j q_j \quad \forall j.$$

Example 18. During this remote quarter Tom's life as a college student can be described by a continuous-time Markov chain. He is always in one of three states: eat, study, and sleep. He eats on average for 1 hour at a time; studies on average for 5 hours; and sleeps on average for 8 hours. After eating, there is a 50-50 chance he will fall asleep or study. After studying, there is a 50-50 chance he will eat or sleep. And after sleeping, Tom always eats. What proportion of the day does Tom sleep in the long-run?

Solution. Let $(X_t)_{t\geq 0}$ be our continuous-time Markov chain with state space $\mathcal{S} = \{\text{eat (1), study (2), sleep (3)}\}$. Let X_t be Tom's state at time t. We have

$$\tilde{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

The holding time parameters are, in hours,

$$q_1 = 1, \quad q_2 = \frac{1}{5}, \quad q_3 = \frac{1}{8}$$

Then we can solve for q_{ij} by taking

$$q_{ij} = p_{ij} q_i.$$

Hence

$$Q = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{10} & -\frac{1}{5} & \frac{1}{10} \\ \frac{1}{8} & 0 & -\frac{1}{8} \end{pmatrix}$$

Let $\pi^T = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix}$. We solve $\pi^T Q = 0$ and $\pi_1 + \pi_2 + \pi_3 = 1$ with a system of equations:

$$\begin{cases} -\pi_1 + \frac{1}{10}\pi_2 + \frac{1}{8}\pi_3 = 0 \\ \frac{1}{2}\pi_1 - \frac{1}{5}\pi_2 = 0 \\ \frac{1}{2}\pi_1 + \frac{1}{10}\pi_2 - \frac{1}{8}\pi_3 = 0 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

Solving this, we get

$$\pi^T = \begin{pmatrix} \frac{2}{19} & \frac{5}{19} & \frac{12}{19} \end{pmatrix}$$

Since the chain is irreducible and the state space is finite, the stationary distribution is also the unique limiting distribution, and we can see that Tom spends around $\pi_3 = \frac{12}{19} \approx \boxed{0.63}$ of his day sleeping. □

Consider the embedded discrete-time Markov chain for $\tilde{P}$ in our continuous-time Markov chain, denoted $Y_0, Y_1, \ldots$. Recall that a probability distribution ϕ is a stationary distribution of the embedded chain if and only if

$$\phi^T \cdot \tilde{P} = \phi^T.$$

In other words,

$$\phi_j = \sum_{i \in \mathcal{S}} \phi_i \cdot \tilde{P_{ij}}, \quad \forall j \in \mathcal{S}$$

The stationary distribution π of the continuous-time Markov chain is not equal to the stationary distribution of ϕ of the embedded chain!

Proposition 15. (1) Let π be given. Then

$$\phi_j = \frac{\pi_j q_j}{\sum_k \pi_k q_k}, \quad \forall j$$

is the stationary distribution of the embedded chain.

(2) Let ϕ be given. Then

$$\pi_j = \frac{\phi_j / q_j}{\sum_k \phi_k / q_k}, \quad \forall j$$

is the stationary distribution of the continuous-time Markov chain.

In other words, ϕ_j is the long-term proportion of transitions the chain makes into state j.

Example 19. Consider the Markov chain of Tom's life from before. Compute the stationary distribution ϕ of the embedded chain from the transition matrix $\tilde{P}$.

Solution. To compute the stationary distribution, we solve $\phi^T \tilde{P} = \phi^T$ and $\phi_1 + \phi_2 + \phi_3 = 1$, where

$$\tilde{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

Then we get the linear system

$$\begin{cases} -\phi_1 + \frac{1}{2}\phi_2 + \phi_3 = 0 \\ \frac{1}{2}\phi_1 - \phi_2 = 0 \\ \frac{1}{2}\phi_1 + \frac{1}{2}\phi_2 - \phi_3 = 0 \\ \phi_1 + \phi_2 + \phi_3 = 1 \end{cases}$$

Solving this, we get

$$\boxed{\phi^T = \begin{pmatrix} \frac{4}{9} & \frac{2}{9} & \frac{3}{9} \end{pmatrix}}$$

We can show that Proposition 15 is satisfied:

$$\pi_1 = \frac{4/9 \cdot 1}{\frac{4}{9} \cdot 1 + \frac{2}{9} \cdot 5 + \frac{3}{9} \cdot 8} = \frac{2}{19}$$

$$\pi_2 = \frac{2/9 \cdot 5}{\frac{4}{9} \cdot 1 + \frac{2}{9} \cdot 5 + \frac{3}{9} \cdot 8} = \frac{5}{19}$$

$$\pi_3 = \frac{3/9 \cdot 8}{\frac{4}{9} \cdot 1 + \frac{2}{9} \cdot 5 + \frac{3}{9} \cdot 8} = \frac{12}{19}$$

□

2.6 Absorbing States

Let $(X_t)_{t\geq 0}$ be an absorbing Markov chain on $\{1,\ldots,k\}$ with one absorbing state a. Similar to the discrete case, all non-absorbing states are *transient*. Let T denote the set of transient states. We write the generator matrix in canonical block matrix form:

$$Q = \begin{pmatrix} 0 & 0 \\ \star & V \end{pmatrix},$$

where V is a $(k-1)\times(k-1)$ matrix. For transient state i, what can we say about the expected time until absorption?

Proposition 16 (Mean Time Until Absorption)**.** For an absorbing continuous-time Markov chain, define a square matrix F on the set T of transient states, where F_{ij} is the expected time, for the chain started in $i \in T$, that the process spends in j until absorption. Then,

$$F = -V^{-1}.$$

For the chain started in $i \in T$, the mean time until absorption is,

$$a_i = \sum_{j\in T} F_{ij}$$

The matrix F is called the *fundamental matrix*.

As an example, multi-state Markov models are used in medicine to model the course of diseases. A patient may advance into, or recover from, successively more severe stages of a disease until some terminal state. Each stage represents a state of an absorbing continuous-time Markov chain.

Example 20. Bartolomelo et al. (2011) developed such a model to study the progression of liver disease among patients diagnosed with cirrhosis of the liver. The general form of the infinitesimal generator matrix for their three-parameter model is

$$Q = \begin{pmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ 0 & -q_{23} & q_{23} \\ 0 & 0 & 0 \end{pmatrix},$$

where state 1 represents cirrhosis, state 2 denotes liver cancer, and state 3 is death. Compute the mean absorption times for states 1 and 2.

Solution. We can rewrite our generator matrix to be

$$Q = \begin{pmatrix} 0 & 0 & 0 \\ q_{13} & -(q_{13}+q_{12}) & q_{12} \\ q_{23} & 0 & -q_{23} \end{pmatrix}$$

Then V will be the bottom left quadrant:

$$V = \begin{pmatrix} -(q_{13}+q_{12}) & q_{12} \\ 0 & -q_{23} \end{pmatrix}$$

Then the fundamental matrix is

$$F = -V^{-1} = \frac{1}{q_{23}(q_{13}+q_{12})} \begin{pmatrix} q_{23} & q_{12} \\ 0 & q_{13}+q_{12} \end{pmatrix}$$

Then the mean absorption rates will be

$$a_1 = \frac{1}{q_{13}+q_{12}} + \frac{q_{12}}{q_{23}(q_{13}+q_{12})}, \quad a_2 = \frac{1}{q_{23}}$$

Then if $q_{12} = 0.0151$, $q_{13} = 0.0071$, $q_{23} = 0.0284$, where t is in months, then

$$\boxed{a_1 = 69, \quad a_2 = 35.21}$$

□

Definition 23 (Detailed Balance Condition)**.** Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain with generator Q. A probability distribution λ satisfies the *detailed balance condition* (or local balance condition) if

$$\lambda_i q_{ij} = \lambda_j q_{ij}, \quad \forall i, j$$

As a consequence,

Proposition 17. If a probability distribution λ satisfies the detailed balance condition, then λ is a stationary distribution.

Proof. Suppose λ satisfies the detailed balance condition. We want to show $\lambda^T Q = 0$, or

$$\sum_i \lambda_i q_{ij} = 0$$

Indeed,

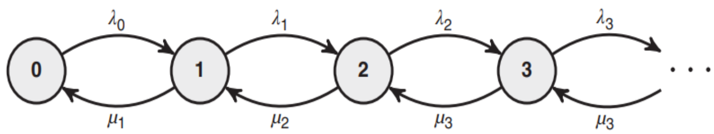
$$\sum_i \lambda_i q_{ij} = \sum_i \lambda_j q_{ji} = \lambda_j \sum_i q_{ji} = 0$$

□

Definition 24 (Birth-and-Death Process)**.** A *birth-and-death process* is a continuous-time Markov chain with state space $\mathcal{S} = \{0, 1, 2, \dots\}$ where transitions only occur to neighboring states, i.e. "births" occur from i

to $i+1$ at the rate λ_i and "deaths" occur from i to $i-1$ at the rate μ_i.

This transition graph illustrates a birth-and-death process:



Notice that the state space $\mathcal{S}$ does not need to be finite.

Proposition 18. For a birth-and-death process with birth rates λ_i and death rates μ_i for $i = 0, 1, 2, \ldots,$ assume that

$$\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} < \infty.$$

Then the unique stationary distribution π is given by

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \quad \forall k = 0, 1, 2, \ldots,$$

where

$$\pi_0 = \left(\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} \right)^{-1}$$

Proof. By the detailed balance condition,

$$\pi_i q_{ij} = \pi_j q_{ji}, \quad \forall i, j$$

Here,

$$\pi_i q_{i,i+1} = \pi_{i+1} q_{i+1,i} \quad i = 0, 1, \ldots$$

and $q_{i,i+1} = \lambda_i$, and $q_{i+1,i} = \mu_{i+1}$. Then

$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1}, \quad \pi_2 = \pi_1 \frac{\lambda_1}{\mu_2} = \pi_0 \frac{\lambda_0}{\mu_1} \frac{\lambda_1}{\mu_2}$$

$$\implies \pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i=1}}{\mu_i}, \quad k = 0, 1, \ldots$$

Moreover,

$$1 = \sum_{k=0}^{\infty} \pi_k = \sum_{k=0}^{\infty} \left(\pi_0 \sum_{i=1}^{k} \frac{\lambda_{k-1}}{\mu_k} \right) = \pi_0 \sum_{k=0}^{\infty} \left(\prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} \right)$$

$$\Longrightarrow \sum_{k=0}^{\infty}\left(\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i}\right) < \infty, \quad \pi_0 = \left(\sum_{k=0}^{\infty}\left(\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i}\right)\right)^{-1}$$

□

It is a fact that if the stationary distribution in the above proposition exists, then it is also the limiting distribution of the birth-and-death process.

Example 21. Consider a continuous-time version of the simple random walk on $\{0,1,2,\dots\}$ with reflecting boundary at 0. From 0, the walk moves to 1 after an exponentially distributed length of time with rate λ. From $i > 0$, transitions to the left occur at rate μ, and transitions to the right occur at rate λ. Find the stationary distribution.

Solution. The simple random walk is a birth-and-death process with $\lambda_i = \lambda$ and $\mu_i = \mu$. By Proposition 18,

$$\pi_k = \pi_0 \prod_{i=1}^{k}\frac{\lambda}{\mu} = \pi_0\left(\frac{\lambda}{\mu}\right)^k \quad k = 0, 1, 2$$

and

$$\pi_0 = \left(\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k\right)^{-1} = \left(\frac{1}{1-\lambda/\mu}\right)^{-1} = 1 - \frac{\lambda}{\mu}, \quad \text{if } \frac{\lambda}{\mu} < 1$$

This comes from the geometric series expansion:

$$\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k = \begin{cases}\frac{1}{1-\lambda/\mu}, & \frac{\lambda}{\mu} < 1\\ \infty, & \text{otherwise}\end{cases}$$

In conclusion,

- $\lambda < \mu$:

  $$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k, \quad k = 0, 1, \dots$$

  will be the stationary distribution (and also the limiting distribution).

- $\lambda \geq \mu$: The stationary distribution does not exist.

□

2.7 Queueing Theory

Queueing theory is the study of waiting lines or queues. For example, customers arrive at a facility for a service. If the service is not immediately available, they wait for service, and leave the system when the service is complete. The general queueing model is quite broad, with notation:

$$A/B/n = \text{arrival time distribution/service time distribution/\# of servers}$$

In the context of Markov chains, we focus on birth-and-death processes, where $(X_t)_{t\geq 0}$ is a Markov chain, and X_t is the number of customers in the system at time t. This means we deal with an *M/M/c* queue, where M stands for either Markov or memoryless, which means the arrival and service times are both exponentially distributed.

Proposition 19 (Little's Formula)**.** In a queueing system, let L denote the long-term average number of customers in the system, λ the rate of arrivals, and W the long-term average time that a customer is in the system. Then,

$$L = \lambda W.$$

Example 22. Cars arrive at a drive-through carwash (with one spot) according to a Poisson process at the rate of nine customers per hour. The time to wash a car has exponential distribution with mean 5 minutes.

(a) How many cars, on average, are at the carwash in the long-run?

(b) How long, on average, is a customer at the carwash in the long-run?

(c) How long, on average, does a customer wait to be served?

(d) What is the expected number of cars waiting to be served?

Solution. Let X_t be the number of cars in the carwash at time t. This is a M/M/1 queueing system, a birth-and-death process. The arrival rate is $\lambda = 9$ and the service rate is $\mu = 12$. The limiting distribution is

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k = \left(1 - \frac{9}{12}\right)\left(\frac{9}{12}\right)^k = \frac{1}{4} \cdot \left(\frac{3}{4}\right)^k$$

(a) Since $\pi_k \sim \text{Geo}(p = 1/4)$, the expected value is

$$\frac{1-p}{p} = \frac{3/4}{1/4} = 3$$

Hence, $L = 3$.

(b) By Little's Formula,

$$W = \frac{L}{\lambda} = \frac{3}{9} = 20 \text{ minutes}$$

(c) We have $W = W_q + W_s$, where W_q is the long-term average waiting time and W_s is the long-term average service time. Here,

$$W_s = \frac{1}{\mu} = \frac{1}{12}$$

Hence

$$W_q = W - W_s = \frac{1}{3} - \frac{1}{12} = \frac{1}{4}$$

So a customer will wait on average 15 minutes before being served.

(d) Consider a process restricted to the queue as its own queueing system. Then $L_q = \lambda W_q$, and so

$$L_q = 9 \cdot \frac{1}{4} = \frac{9}{4}$$

Hence on average, there are 2.25 cars in the queue waiting to be served.

□

Definition 25 (M/M/c Queue)**.** A *M/M/c* queue has exponentially distributed arrival and service times, with c independent servers. Death-rate depends on number of customers i in the system:

- $0 < i \leq c$: all customers are being served with rate $\mu_i = i \cdot \mu$
- $i > c$ only c customers are being served with rate $\mu_i = c \cdot \mu$

$$\mu_i = \begin{cases} i \cdot \mu, & i = 1, \ldots, c-1 \\ c \cdot \mu & i = c, c+1, \ldots \end{cases}$$

for $i = 1, 2, \ldots$

Proposition 20. For $0 < \lambda < c\mu$ the stationary distribution π of the M/M/c queue exists and is given by

$$\pi_k = \begin{cases} \frac{\pi_0}{k!}(\frac{\lambda}{\mu})^k, & 0 \leq k < c \\ \frac{\pi_0}{c^{k-c}c!}(\frac{\lambda}{\mu})^k, & k \geq c \end{cases}$$

where

$$\pi_0 = \left(\sum_{k=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!} \left(\frac{1}{1 - \lambda/(c\mu)} \right) \right)^{-1}$$

Proof. We check the assumption:

$$\begin{aligned}
\sum_{k=0}^{\infty}\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i} &= \sum_{k=0}^{c-1}\prod_{i=1}^{k}\frac{\lambda}{i\mu} + \sum_{k=c}^{\infty}\left(\prod_{i=1}^{c}\frac{\lambda}{i\mu}\right)\left(\prod_{i=c+1}^{k}\frac{\lambda}{c\mu}\right) \\
&= \sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{1}{c!}\sum_{k=c}^{\infty}\left(\frac{\lambda}{\mu}\right)^k\left(\frac{1}{c}\right)^{k-c} \\
&= \sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!}\sum_{k=c}^{\infty}\left(\frac{\lambda}{c\mu}\right)^{k-c} \\
&= \frac{1}{1-\frac{\lambda}{c\mu}}, \quad \lambda < c\mu
\end{aligned}$$

Hence if $\lambda < c\mu$ the stationary distribution exists and is given by

$$\pi_0 = \left(\sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!}\frac{1}{1-\frac{\lambda}{c\mu}}\right)^{-1}$$

and

$$\pi_k = \pi_0\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i} = \begin{cases}\frac{\pi_0}{k!}(\frac{\lambda}{\mu})^k, & 0 \le k < c \\ \frac{\pi_0}{c^{k-c}c!}(\frac{\lambda}{\mu})^k, & k \ge c\end{cases}$$

□

Example 23. A hair salon has five chairs. Customers arrive at the salon at the rate of 6 per hour. The hair stylists each take, on average, half an hour to service a customer, independent of arrival times.

(a) Bill, the owner, wants to know the long-term probability that no customers are in the salon.

(b) Leslie, a potential customer, wants to know the average waiting time for a haircut.

(c) Dennis, a hair stylist, wants to know the long-term expected number of customers in the salon.

Solution. This is a M/M/5 queue with $\lambda = 6$ and generator matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(a) This will just be π_0. Using the formula in the Proposition, with $\lambda = 6$, $\mu = 2$, and $c = 5$, we find

$$\pi_0 = \frac{16}{343} \approx \boxed{0.0466}$$

(b) Since $L_q = W_q\lambda$, and we know that

$$L_q = \sum_{k=c}^{\infty}(k-c)\pi_k,$$

at $k = c$ all 5 seats are taken and there are 0 spots in the queue. Then

$$\begin{aligned}
\sum_{k=c}^{\infty}(k-c)\pi_k &= \sum_{k=0}^{\infty}(k-c)\frac{\pi_0}{c^{k-c}c!}\left(\frac{\lambda}{\mu}\right)^k \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{k=c}^{\infty}(k-c)\frac{1}{c^{k-c}}\left(\frac{\lambda}{\mu}\right)^{k-c} \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{l=0}^{\infty}\frac{l}{c^l}\left(\frac{\lambda}{\mu}\right)^{l} \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{l=0}^{\infty} l\left(\frac{\lambda}{c\mu}\right)^{l} \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \left(\frac{\lambda}{c\mu}\right)\left(\frac{1}{1-\lambda/c\mu}\right)^2 \\
&= 0.35423
\end{aligned}$$

Thus we can see that

$$W_q = \frac{L_q}{\lambda} = \frac{0.35423}{6} = \boxed{0.059 \text{ hours}} \approx 3.54 \text{ minutes}$$

(c) This represents L in Little's Formula for the whole chain. Since

$$W = W_q + W_s,$$

where we know W_q from above and

$$W_s = \frac{1}{\mu} = \frac{1}{2},$$

we deduce that $W = 0.059 + 0.5 = 0.559$ hours. Then

$$L = \lambda W = 6(0.559) = \boxed{3.354 \text{ customers}}$$

□

3 Brownian Motion

Brownian motion is apart of a bigger class of continuous-time, continuous-state stochastic processes, called *Wiener processes*.

3.1 Introduction

Definition 26 (Brownian Motion)**.** A continuous-time stochastic process $(B_t)_{t\geq 0}$ is called a *standard Brownian motion* if it satisfies the following properties:

(1) $B_0 = 0$

(2) (*Independent Increments*)
For all $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n$, the random variables $B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.

(3) (*Stationary Increments*)
For all $0 \leq s < t$ the random variable $B_t - B_s$ is normally distributed with mean $\mu = 0$ and variance $\sigma^2 = t - s$.

(4) (*Continuous paths)*
The function $t \mapsto B_t$ is continuous.

Some consequences:

- $B_t = B_t - 0 = B_t - B_0 \sim N(0,t)$ for all $t > 0$.
- $\mathbb{E}[B_t] = 0$, $\mathrm{Var}[B_t] = t$.
- $B_5 - B_3 \overset{d}{=} B_4 - B_2 \overset{d}{=} B_2 - B_0 \sim N(0,2)$.
- B_t is not independent of B_s for any $(0 \leq s < t)$, but $B_t - B_s$ is independent of $B_s = B_s - B_0$.
- We find that
$$\mathbb{P}[B_t \leq c] = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}\, dx$$

Brownian motion can be thought of as the motion of a particle that diffuses randomly in time along the real line $\mathbb{R}$. As t increases, the particle's position is more diffuse. Many applications in practice, namely modeling the evolution of stock prices in mathematical finance.

Example 24. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion.

(a) For $0 < s < t$, find the distribution of $B_s + B_t$.

(b) For $s, t > 0$, find the covariance of B_s and B_t.

Solution. (a) Let $0 < s < t$. Observe that

$$B_s + B_t = 2B_s + B_t - B_s$$

Due to independent increments, B_s and $B_t - B_s$ are independent. Hence, $2B_s$ and $B_t - B_s$ are independent. Moreover, the sum of independent normal random variables is again normal. Hence $2B_s + (B_t - B_s) = B_s + B_t$ is normally distributed, with

$$\mathbb{E}[B_s + B_t] = 0$$

Then

$$\begin{aligned}
\mathrm{Var}[B_s + B_t] &= \mathrm{Var}[2B_s + (B_t - B_s)] \\
&= \mathrm{Var}[2B_s] + \mathrm{Var}[B_t - B_s] \\
&= 4\,\mathrm{Var}[B_s] + (t - s) \\
&= 4s + (t - s) \\
&= 3s + t
\end{aligned}$$

So $\boxed{B_s + B_t \sim N(0, 3s + t)}$.

(b) Since $\mathbb{E}[B_s]$ and $\mathbb{E}[B_t]$ are zero,

$$\mathrm{Cov}(B_s, B_t) = \mathbb{E}[B_s B_t] - \mathbb{E}[B_s]\mathbb{E}[B_t] = \mathbb{E}[B_s B_t]$$

For $s < t$, $B_t = B_s + (B_t - B_s)$. Then

$$\begin{aligned}
\mathbb{E}[B_s B_t] &= \mathbb{E}[B_s(B_s + (B_t - B_s))] \\
&= \mathbb{E}[B_s^2 + B_s(B_t - B_s)] \\
&= \mathbb{E}[B_s^2] + \mathbb{E}[B_s(B_t - B_s)] \\
&= \mathrm{Var}[B_s] + \mathbb{E}[B_s]\mathbb{E}[B_t - B - s] \\
&= s
\end{aligned}$$

For $t < s$, $\mathbb{E}[B_t B_s] = t$ through similar computations. Hence

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

□

3.2 Simulating Brownian Motion

Consider simulating Brownian motion on $[0,t]$:

- We have a grid of discrete time points $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$.
- By stationary/independent increments, with $B_{t_0} = B_0 = 0$,

$$B_{t_i} = B_{t_{i-1}} + (B_{t_i} - B_{t_{i-1}}) \stackrel{d}{=} B_{t_{i-1}} + X_i \quad i = 1,2,\ldots,n$$

  where $X_i \sim N(0, t_i - t_{i-1})$ independent of $B_{t_{i-1}}$.

- The recursive representation would be, where $Z_1, \ldots Z_n \stackrel{i.i.d.}{\sim} N(0,1)$,

$$B_{t_i} = B_{t_{i-1}} + \sqrt{t_i - t_{i-1}} \cdot Z_i \quad i = 1,2,\ldots,n$$

- This generates the Brownian motion random variables $B_{t_0}, B_{t_1}, \ldots, B_{t_n}$ on the discrete grid.
- Typically equally spaced time points: $t_i = i \cdot \frac{t}{n}$ and hence $t_i - t_{i-1} = t/n$.

Recall that a *simple symmetric random walk* is a sequence of random variables $(S_n)_{n\geq 0}$ with $S_0 = 0$ such that

$$S_n = \sum_{i=1}^{n} X_i,$$

where X_i are a sequence of i.i.d. random variables satisfying

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$$

Then observe that

- $(S_n)_{n\geq 0}$ is a discrete-time, discrete-state stochastic process with stationary and independent increments.
- By the Central Limit Theorem,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} = \frac{S_n}{\sqrt{n}} \longrightarrow Z \sim N(0,1) \quad \text{as } n \to \infty$$

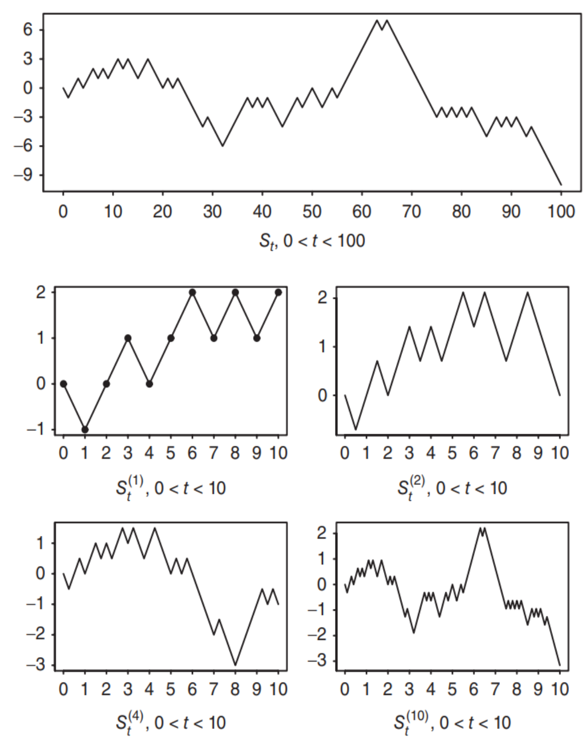
This motivates us to define for each $n \geq 1$ the continuous-time stochastic process $(B_t^{(n)})_{t\geq 0}$ by

$$(B_t^{(n)})_{t\geq 0} = \frac{1}{\sqrt{n}} \cdot \left(S_{\lfloor nt \rfloor} + X_{\lfloor nt \rfloor + 1} \cdot (nt - \lfloor nt \rfloor) \right), \quad t \geq 0$$

via linearly interpolating the discrete values $S_0, S_1, S_2, \ldots$. That is,

- A simple symmetric random walk is scaled both horizontally (speeding up steps by factor n) and vertically (shrinking values by factor $1/\sqrt{n}$)
- We have $\mathbb{E}[B_t^{(n)}] = 0$ and $\mathrm{Var}[B_t^{(n)}] \approx t$ for n large
- via CLT, $B_t^{(n)} \longrightarrow Z \sim N(0,1)$ as $n \to \infty$

Graphically, we can see various instances of the linear interpolation:



Proposition 21 (Donsker's Invariance Principle)**.** We have

$$(B_t^{(n)})_{t\geq 0} \longrightarrow (B_t)_{t\geq 0} \quad \text{for } n \to \infty$$

where $(B_t)_{t\geq 0}$ denotes a standard Brownian motion.

Here we have a convergence of stochastic processes; a "functional CLT" for the entire path.

3.3 Gaussian Process

Definition 27 (Multivariate Normal Distribution)**.** The random vector $(X_1, \dots, X_n) \in \mathbb{R}^n$ follows a *multivariate normal distibution* if for all numbers $a_1, \dots, a_n \in \mathbb{R}$, the linear combination

$$a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_n \cdot X_n \in \mathbb{R}$$

is normally distributed on $\mathbb{R}$ (univariate normal distribution). A multivariate normal distribution is completely determined by its *mean vector*

$$\mu = (\mu_1, \dots, \mu_n)^T = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])^T \in \mathbb{R}^n$$

and *covariance matrix* $\Sigma \in \mathbb{R}^{n \times n}$, where

$$\Sigma_{ij} = \operatorname{Cov}(X_i, X_j) \quad \text{for all } 1 \le i,\ j \le n.$$

In the case that Σ is invertible, the joint density function of the multivariate normal distribution is given by

$$f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \cdot \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\Big)$$

where $x = (x_1, \dots, x_n)^T$ and $|\Sigma|$ is the determinant of Σ.

Our notation for this will be $X = (X_1, \dots, X_n)^T \sim N(\mu, \Sigma)$. Some properties:

- If $X_1, \dots, X_n$ are independent with $X_i \sim N(\mu_i, \sigma_i^2)$ for all $i \in [1, n]$, then $X = (X_1, \dots, X_n)^T$ is multivariate normally distributed with

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

- If $X = (X_1, \dots, X_n)^T \sim N(\mu, \Sigma)$, $A \in \mathbb{R}^{k \times n}$, $a \in \mathbb{R}^k$, then $Y = A \cdot X + a$ is multivariate normal distributed with mean $A \cdot \mu + a$ and covariance matrix $A \cdot \Sigma \cdot A^T$.

- If $X_1, \dots, X_n$ are i.i.d. with $X_i \sim N(0, 1)$ for all $i \in [1, n]$, $A \in \mathbb{R}^{n \times n}$, $\mu \in \mathbb{R}^n$, then $Y = A \cdot X + \mu \sim N(\mu, \Sigma)$, where $\Sigma = A \cdot A^T$.

This leads us to introduce the following concept:

Definition 28 (Gaussian Process)**.** A *Gaussian process* $(X_t)_{t\geq 0}$ is a continuous-time stochastic process with the property that for all $n = 1, 2, \dots$ and all $0 \leq t_1 < \cdots < t_n$, the random vector

$$(X_{t_1}, \dots, X_{t_n}) \in \mathbb{R}^n$$

follows a multivariate normal distribution. A Gaussian process is completely determined by its *mean function*

$$m(t) = \mathbb{E}[X_t] \quad t \geq 0$$

and *covariance function*

$$c(s,t) = \mathrm{Cov}(X_s, X_t) \quad s,t \geq 0.$$

The Gaussian process extends the multivariate normal distribution to stochastic processes. The standard Brownian motion is a specific Gaussian process.

Proposition 22. A stochastic process $(B_t)_{t\geq 0}$ is a standard Brownian motion if and only if it is a Gaussian process with the following properties:

(1) $B_0 = 0$

(2) (*Mean function*):
$$\mathbb{E}[B_t] = 0 \quad \forall t \geq 0$$

(3) (*Covariance function*):
$$\mathrm{Cov}(B_s, B_t) = \min\{s,t\} \quad \forall s,t \geq 0$$

(4) (*Continuous paths*):

The function $t \mapsto B_t$ is continuous

Proof. $(\Longrightarrow)$: Let $(B_t)_{t\geq 0}$ be a standard Brownian motion.

(i) Gaussian process: Let $0 < t_1 < \cdots < t_n$ and $a_1, \dots, a_n \in \mathbb{R}$ be arbitrary. By independent increments,

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots B_{t_n} - B_{t_{n-1}}$$

are independent and normally distributed. Then

$$a_1 B_{t_1} + \cdots + a_n B_{t_n} = (a_1 + a_2 + \cdots + a_n) B_{t_1} + (a_2 + \cdots + a_n) B_{t_2} + \cdots + a_n (B_{t_n} - B_{t_{n-1}}),$$

which is univariate normally distributed as a linear combination of independent normally distributed random variables.

(ii) Desired properties:

(1) $B_0 = 0$

(2) $\mathbb{E}[B_t] = 0$, for all $t \geq 0$

(3) $\mathrm{Cov}(B_t, B_s) = \min\{s, t\}$, for all $s, t > 0$

(4) $(B_t)_{t\geq 0}$ has continuous paths

($\Longleftarrow$) : Let $(B_t)_{t\geq 0}$ be a Gaussian process with the four properties. We want to show that $(B_t)_{t\geq 0}$ has stationary and independent increments.

(i) Since $(B_t)_{t\geq 0}$ is Gaussian, $B_t - B_0 = B_t \sim N(0, t)$ for all $t > 0$. Similarly, $B_{t+s} - B_s$ is normally distributed with $\mathbb{E}[B_{t+s} - B_s] = 0$, and

$$\begin{aligned}
\mathrm{Var}[B_{t+s} - B_s] &= \mathrm{Var}[B_{t+s}] + \mathrm{Var}[-B_s] + 2\,\mathrm{Cov}(B_{t+s}, -B_s) \\
&= \mathrm{Var}[B_{t+s}] + \mathrm{Var}[B_s] - 2\,\mathrm{Cov}(B_{t+s}, B_s) \\
&= t + s + s - 2s \\
&= t
\end{aligned}$$

Hence $B_{t+s} - B_s \overset{d}{=} B_t - B_0$, so we have stationary increments.

(ii) Let $0 \leq q, r \leq s < t$. Then

$$\begin{aligned}
\mathrm{Cov}(B_r - B_q, B_t - B_s) &= \mathbb{E}[(B_r - B_q)(B_t - B_s)] - \mathbb{E}[B_r - B_q]\mathbb{E}[B_t - B_s] \\
&= \mathbb{E}[B_r B_t] - \mathbb{E}[B_r B_s] - \mathbb{E}[B_q B_t] + \mathbb{E}[B_q B_s] \\
&= r - r - q + q \\
&= 0
\end{aligned}$$

Hence $(B_r - B_q)$ and $(B_t - B_s)$ are uncorrelated, so they are independent and normally distributed. Thus we have independent increments.

□

3.4 Transformations of Brownian Motion

Proposition 23. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then, each of the following transformations is again a standard Brownian motion:

(1) (*Reflection*): $(-B_t)_{t\geq 0}$.

(2) (*Translation*): $(B_{t+s} - B_s)_{t\geq 0}$ for all $s \geq 0$.

(3) (*Rescaling*): $(a^{-1/2} B_{at})_{t\geq 0}$ for all $a > 0$.

(4) (*Time Inversion*):

$$W_t = \begin{cases} W_t = 0, & t = 0 \\ t \cdot B_{1/t}, & t > 0 \end{cases}$$

Proof. We only prove the rescaling transformation. Let $X_t = B_{at}/\sqrt{a}$. We want to show $(X_t)_{t\geq 0}$ is a Brownian motion. To do this, we show $(X_t)_{t\geq 0}$ is a Gaussian process with

(i) $X_0 = 0$

(ii) $\mathbb{E}[X_t] = 0$

(iii) $\text{Cov}(X_s, X_t) = \min\{s,t\}$

(iv) Continuous sample paths

First, for $0 < t_1 < \cdots < t_n$ and $a_1, \ldots, a_n \in \mathbb{R}$ arbitrary,

$$\sum_{i=1}^{n} a_i X_i = \sum_{i=1}^{n} \frac{a_i}{\sqrt{a}} B_{at_i}$$

is univariate normally distributed because $(B_t)_{t\geq 0}$ is a Gaussian process. Hence, $(X_t)_{t\geq 0}$ is a Gaussian process. Moreover,

(i) $X_0 = \frac{1}{\sqrt{a}} B_0 = 0$

(ii) $\mathbb{E}[X_t] = \frac{1}{\sqrt{a}} \mathbb{E}[B_{at}] = 0$

(iii) $\text{Cov}(X_s, X_t) = \text{Cov}(\frac{1}{\sqrt{a}} B_{as}, \frac{1}{\sqrt{a}} B_{at}) = \frac{1}{a} \text{Cov}(B_{as}, B_{at}) = \frac{1}{a} \min\{as, at\} = \min\{s,t\}$

(iv) Continuity of $t \mapsto X_t$ follows from continuity of $(B_t)_{t\geq 0}$.

□

A remarkable fact about Brownian motion is that the sample paths are everywhere-continuous, yet nowhere-differentiable. This is due to the fractal structure of Brownian motion sample paths: jagged character of the paths remain jagged at all time scales.

If we want to change the starting value,

Proposition 24. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. For $x \in \mathbb{R}$ the process $(X_t)_{t\geq 0}$ defined by

$$X_t = x + B_t \quad t \geq 0$$

is called *Brownian motion started at x*.

Some properties:

- $(X_t)_{t\geq 0}$ retains all properties of standard Brownian motion, like independent/stationary increments, and continous sample paths.
- $X_t \sim N(x,t)$, for all $t > 0$.
- $X_t - X_s \sim N(0, t-s)$, for all $t \geq s > 0$.

3.5 Stopping Times and Distributions

Definition 29 (Markov Process)**.** A continuous-state, continuous-time stochastic process $(X_t)_{t\geq 0}$ is a *Markov process* if

$$\mathbb{P}[X_{s+t} \leq y \mid X_u,\ 0 \leq u \leq s] = \mathbb{P}[X_{s+t} \leq y \mid X_s]$$

for all $s, t \geq 0$ and all $y \in \mathbb{R}$. In addition, the process is time homogeneous if

$$\mathbb{P}[X_{s+t} \leq y \mid X_s] = \mathbb{P}[X_t \leq y \mid X_0]$$

In words, it is conditional on the present, past and future independent. As a consequence of stationary and independent increments, Brownian motion is an example of a Markov process.

Definition 30 (Stopping Time)**.** A positive random variable S is called a *stopping time* for a stochastic process $(X_t)_{t\geq 0}$ if, for each $s \in \mathbb{R}_+$, the occurrence of the event $\{S \leq s\}$ can be determined from the stochastic process $(X_t)_{0\leq t\leq s}$ up to time s.

An important example of a stopping time is as follows:

Definition 31 (First Hitting Time)**.** For $a \in \mathbb{R}$ let

$$T_a = \min\{t \geq 0 : B_t = a\}$$

be the *first hitting time* that Brownian motion hits level a.

The random variable T_a is a stopping time for a Brownian motion $(B_t)_{t\geq 0}$. Indeed, for any $s > 0$, the occurrence of the event $\{T_a \leq s\}$ can be determined from the Brownian motion $(B_t)_{0\leq t\leq s}$ up to time s.

A counterexample would be the random time

$$L = \max\{0 \leq t \leq 1 : B_t = 0\},$$

or the last visit to 0 in $[0,1]$. This is not a stopping time for a Brownian motion $(B_t)_{t\geq 0}$. Indeed, for any $0 < s < 1$, the occurrence of the event $\{L \leq s\}$ can only be determined from the Brownian motion $(B_t)_{0\leq t\leq 1}$ but not just from the Brownian motion $(B_t)_{0\leq t\leq s}$ up to time s.

Proposition 25. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and S a stopping time for $(B_t)_{t\geq 0}$. Then

$$\tilde{B}_t = B_{s+t} - B_S \quad t \geq 0$$

is a standard Brownian motion independent of $(B_t)_{0\leq t\leq S}$.

This is also known as the *strong Markov property*. This generalizes Proposition 24 and the Translation property. In particular, any constant $S = s \geq 0$ is a stopping time.

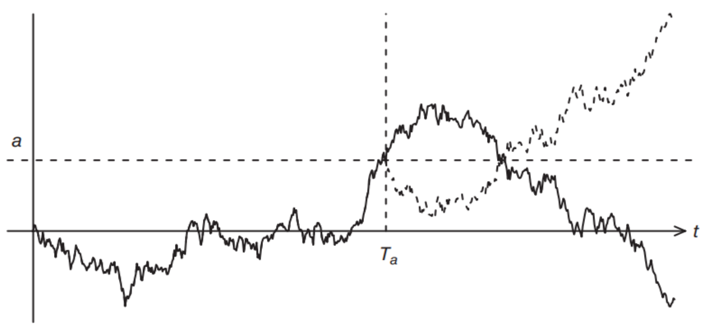
Example 25. For the first hitting time T_a, show that $(B_{T_a} + t)_{t\geq 0}$ is again a Brownian motion starting at $B_{T_a} = a$ and independent of $(B_t)_{0\leq t\leq T_a}$.

Proposition 26 (Reflection Principle)**.** For a standard Brownian motion $(B_t)_{t\geq 0}$, and first hitting time T_a with $a > 0$, the process

$$B_t^* = \begin{cases} B_t, & 0 \leq t \leq T_a \\ a - (B_t - a) = 2a - B_t, & t \geq T_a \end{cases}$$

is again a standard Brownian motion. The process is also called *Brownian motion reflected* at T_a.

This can be illustrated as such:



For a standard Brownian motion $(B_t)_{t\geq 0}$, let

$$M_t = \max_{0\leq s\leq t} B_s, \quad t \geq 0$$

be the maximum of standard Brownian motion on $[0, t]$.

Proposition 27. (1) For $a > 0$, $y \geq 0$, $t > 0$, it holds that

$$\mathbb{P}[M_t \geq a, B_t \leq a - y] = \mathbb{P}[B_t \geq a + y].$$

(2) For $a > 0$ it holds that

$$\mathbb{P}[M_t \geq a] = 2 \cdot \mathbb{P}[B_t \geq a] = \mathbb{P}[|B_t| \geq a] = \mathbb{P}[T_a \leq t].$$

In particular, $M_t \overset{d}{=} |B_t|$.

Proof. 1) We have

$$\begin{aligned}\mathbb{P}[B_t \le a-y, M_t \ge a] &= \mathbb{P}[B_t \le a-y, T_a \le t] \\ &= \mathbb{P}[B_t^* \le a-y, T_a^* \le t] \\ &= \mathbb{P}[2a - B_t \le a-y, T_a \le t] \\ &= \mathbb{P}[B_t \ge a+y, T_a \le t] \\ &= \mathbb{P}[B_t \ge a+y]\end{aligned}$$

2) We have

$$\begin{aligned}\mathbb{P}[M_t \ge a] &= \mathbb{P}[M_t \ge a, B_t \le a] + \mathbb{P}[M_t \ge a, B_t > a] \\ &= \mathbb{P}[B_t \ge a] + \mathbb{P}[B_t > a] \\ &= 2\mathbb{P}[B_t \ge a]\end{aligned}$$

Moreover,

$$\begin{aligned}2\mathbb{P}[B_t \ge a] &= \mathbb{P}[B_t \ge a] + \mathbb{P}[B_t \ge a] \\ &= \mathbb{P}[B_t \ge a] + \mathbb{P}[-B_t \ge a] \\ &= \mathbb{P}[|B_t| \ge a]\end{aligned}$$

Finally, since $\{M_t \ge a\} = \{T_a \le t\}$, we get that $\mathbb{P}[M_t \ge a] = \mathbb{P}[T_a \le t]$.

□

Proposition 28 (Distributions of Maximum and Hitting Times)**.** (1) (M_t, B_t) has density

$$f_{(M_t,B_t)}(m,x) = \sqrt{\frac{2}{\pi t^3}} \cdot (2m - x) \cdot \exp\left(-\frac{(2m-x)^2}{2t}\right)$$

for all $m > 0$, $x \le m$.

(2) M_t has density

$$f_{M_t}(x) = \sqrt{\frac{2}{\pi t}} \cdot \exp\left(-\frac{x^2}{2t}\right), \quad x > 0$$

(3) T_a for $a \in \mathbb{R}$ has density

$$f_{T_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} \cdot \exp\left(-\frac{a^2}{2t}\right)$$

Proof. We only prove (2). The CDF of M_t will be

$$\begin{aligned}
F_{M_t}(x) &= \mathbb{P}[M_t \leq x] \\
&= 1 - \mathbb{P}[M_t > x] \\
&= 1 - \mathbb{P}[M_t \geq x] \\
&= 1 - 2\mathbb{P}[B_t \geq x] \\
&= 1 - \mathbb{P}[B_t \leq x] \\
&= 1 - \mathbb{P}[\sqrt{t}Z \leq x] \\
&= 1 - \Phi\left(\frac{x}{\sqrt{t}}\right)
\end{aligned}$$

This is equivalent to $2\Phi(\frac{x}{\sqrt{t}}) - 1$. Then the density of M_t will be

$$\begin{aligned}
f_{M_t}(x) &= \frac{d}{dx} F_{M_t}(x) \\
&= 2\Phi'\left(\frac{x}{\sqrt{t}}\right)\left(\frac{1}{\sqrt{t}}\right) \\
&= \frac{2}{\sqrt{t}} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\frac{x^2}{t}\right)
\end{aligned}$$

□

It follows that there are some surprising properties about the first time hitting distribution:

- For all $a \in \mathbb{R}$, we have
$$\mathbb{P}[T_a < \infty] = 1,$$
no matter how large a is.
- For all $a \in \mathbb{R}$, we have
$$\mathbb{E}[T_a] = \infty,$$
no matter how small a is.

Definition 32 (Arcsine Distribution)**.** A random variable X on $[0,1]$ is called *arcsine distributed* if the cumulative distribution function is given by

$$\mathbb{P}[X \leq x] = \frac{2}{\pi}\arcsin(\sqrt{x}), \quad 0 \leq x \leq 1$$

and the probability density function is given by

$$f_X(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \quad x \in (0,1)$$

For a standard Brownian motion $(B_t)_{t\geq 0}$, define the following random variables:

- Proportion of time that $(B_t)_{t\geq 0}$ is positive on $[0,1]$:

$$C = |\{0 \leq t \leq 1 : B_t \geq 0\}|.$$

- Time of last visit to 0 in $[0,1]$:

$$L = \max\{0 \leq t \leq 1 : B_t = 0\}.$$

- Time at which $(B_t)_{t\geq 0}$ obtains its maximum on $[0,1]$:

$$M = \operatorname{argmax}\{B_t : 0 \leq t \leq 1\}.$$

It turns out that one can show that

$$\mathbb{P}[C \leq x] \stackrel{d}{=} \mathbb{P}[L \leq x] \stackrel{d}{=} \mathbb{P}[M \leq x] = \frac{2}{\pi}\arcsin(\sqrt{x}),$$

for $x \in [0,1]$. In other words, all these random variables follow the arcsine distribution!

3.6 Variations and Applications

Definition 33 (Brownian Motion with Drift)**.** Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. For $\mu \in \mathbb{R}$ and $\sigma > 0$, the process defined by

$$X_t = \mu t + \sigma B_t \quad t \geq 0$$

is called *Brownian motion with drift parameter μ and variance σ^2*.

Some properties:

- $\mathbb{E}[X_t] = \mu t$, $\mathrm{Var}[X_t] = \sigma^2 t$, for all $t \geq 0$
- Stationary and independent increments with

$$X_t - X_s \stackrel{d}{=} X_{t-s} - X_0 \sim N(\mu(t-s), \sigma^2(t-s))$$

Definition 34 (Geometric Brownian Motion)**.** Let $(X_t)_{t\geq 0}$ be a Brownian motion with drift parameter $\mu \in \mathbb{R}$ and variance parameter $\sigma^2 > 0$. The process $(G_t)_{t\geq 0}$ defined by

$$G_t = G_0 e^{X_t} \quad t \geq 0$$

where $G_0 > 0$, is called *geometric Brownian motion*.

Definition 35 (Lognormal Distribution)**.** A positive random variable X is called *log-normally* distributed with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if

$$\log(X) \sim N(\mu, \sigma^2)$$

We denote this as $X \sim LN(\mu, \sigma^2)$.

Consequently, if $X \sim LN(\mu, \sigma^2)$, then $X = e^Z$, where $Z = N(\mu, \sigma^2)$.

Proposition 29. Let $(G_t)_{t\geq 0}$ be a geometric Brownian motion. Then,

(1) $G_t \sim LN(\log(G_0) + \mu t, \sigma^2 t)$

(2) It holds that

$$\mathbb{E}[G_t] = G_0 e^{t(\mu+\sigma^2/2)}, \quad \mathrm{Var}[G_t] = G_0^2 e^{2t(\mu+\sigma^2/2}(e^{t\sigma^2} - 1)$$

(3) For all $n \in \mathbb{N}$, and $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n$, the random variables

$$\frac{G_{t_2}}{G_{t_1}}, \frac{G_{t_3}}{G_{t_2}}, \ldots, \frac{G_{t_n}}{G_{t_{n-1}}}$$

are independent.

(4) For all $0 \leq s < t$, we have

$$\frac{G_t}{G_s} \stackrel{d}{=} \frac{G_{t-s}}{G_0} \sim LN(\mu(t-s), \sigma^2(t-s))$$

(5) Let $Y_k = G_k / G_{k-1}$ for $k = 1, \ldots, n$, and let Y_k be i.i.d. Then

$$G_n = G_0 \cdot Y_1 \cdot Y_2 \cdots Y_n$$

Proof. We know $G_t = G_0 e^{X_t}$ with $X_t = \mu t + \sigma^2 B_t$, where $(B_t)_{t\geq 0}$ is a standard Brownian motion.

(1) If $G_t \sim LN(\log(G_0)+\mu t, \sigma^2 t)$, then

$$\log(G_t) = \log(G_0 e^{\mu t+\sigma B_t}) = \log(G_0) + \mu t + \sigma B_t \sim N(\mu t + \log(G_0), \sigma^2 t)$$

(2) We have

$$\mathbb{E}[G_t] = \mathbb{E}[G_0 e^{\mu t+\sigma B_t}] = G_0 e^{\mu t}\mathbb{E}[e^{\sigma B_t}] = G_0 e^{\mu t} e^{\frac{1}{2}\sigma^2 t} = G_0 e^{(\mu+\frac{1}{2}\sigma^2)t}$$

Then

$$\begin{aligned}
\mathrm{Var}[G_t] &= \mathbb{E}[G_t^2] - (\mathbb{E}[G_t])^2 \\
&= \mathbb{E}[G_0^2 e^{2\mu t+2\sigma B_t} - G_0^2 e^{2(\mu+\frac{1}{2}\sigma^2)t} \\
&= G_0^2 e^{2\mu t} e^{2\sigma^2 t} - G_0^2 e^{2(\mu+\frac{1}{2}\sigma^2)t} \\
&= G_0^2 e^{2t(\mu+\frac{1}{2}\sigma^2)}(e^{t\sigma^2} - 1)
\end{aligned}$$

(3) Since

$$\begin{aligned}
\frac{G_{t_2}}{G_{t_1}} &= \frac{G_0 e^{\mu t_2+\sigma B_{t_2}}}{G_0 e^{\mu t_1+\sigma B_{t_1}}} = e^{\mu(t_2-t_1)+\sigma(B_{t_2}-B_{t_1})} \\
\frac{G_{t_3}}{G_{t_2}} &= \frac{G_0 e^{\mu t_3+\sigma B_{t_3}}}{G_0 e^{\mu t_2+\sigma B_{t_2}}} = e^{\mu(t_3-t_2)+\sigma(B_{t_3}-B_{t_2})}
\end{aligned}$$

Then $(B_{t_2} - B_{t_1})$ is independent of $(B_{t_3} - B_{t_2})$, so $\frac{G_{t_2}}{G_{t_1}}$ and $\frac{G_{t_3}}{G_{t_2}}$ are independent. Moreover,

$$\frac{G_t}{G_s} = e^{\mu(t-s)+\sigma(B_t-B_s)} \stackrel{d}{=} e^{\mu(t-s)+\sigma B_{t-s}} = \frac{G_{t-s}}{G_0}$$

□

As an application of geometric Brownian motion, stock prices can be modeled as a geometric Brownian motion:

$$S_t = S_0 \cdot e^{R_t}, \quad R_t = \mu t + \sigma B_t, \quad t \geq 0$$

We interpret S_t as the price at time $t > 0$, R_t the "log-return" on $[0,t]$, μ the expected annual return, and σ the annual volatility. The model assumes that the daily log-returns $\log(S_{i\Delta t}/S_{(i-1)\Delta t}$ for $i \in [1,250]$, and $\Delta t = 1/250$ are i.i.d. $N(\mu\Delta t, \sigma^2\Delta t)$-normally distributed. This model is called the Osborne-Samuelson market model, or the Black-Scholes-Merton model. The pros for this model are

- Prices are positive
- It models exponential growth

However, there exist a few cons as well:

- Log-returns from historical prices are typically not independent and normally distributed
- Normal distribution underestimates occurrences of extreme price moves

Example 26 (Black-Scholes Option Price Formula)**.** A K-strike *call option* has payoff at maturity T for the buyer

$$(S_T - K)^+ = \begin{cases} S_T - K, & S_T > K \\ 0, & S_T \leq K \end{cases}$$

where S_T denotes the stock price at time T. Assume that $(S_t)_{t\geq 0}$ follows a geometric Brownian motion with drift

$$\mu = r - \sigma^2/2$$

where $r > 0$ denotes the annual risk-free interest rate. Show that the present value of the expected payoff is given by

$$\mathbb{E}[e^{-rT}(S_T - K)^+] = S_0\Phi(d_+) - Ke^{-rT}\Phi(d_-),$$

where

$$d_\pm = \frac{\log(S_0/K) + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

Solution. We first split the present value of the expected payoff into

$$\mathbb{E}[e^{-rT}(S_t-K)^+] = \mathbb{E}[e^{-rT}(S_T-K)\cdot\mathbb{1}_{\{S_T-K\geq 0\}}] = e^{-rT}\mathbb{E}[S_T\cdot\mathbb{1}_{\{S_T\geq K\}}]-Ke^{-rT}\mathbb{E}[\mathbb{1}_{\{S_T\geq K\}}]$$

The second term can be calculated to be:

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{\{S_T\geq K\}}] &= \mathbb{P}[S_T \geq K] \\
&= \mathbb{P}[S_0 e^{\mu T+\sigma B_T} \geq K] \\
&= \mathbb{P}\left[\mu T + \sigma B_T \geq \log\frac{K}{S_0}\right] \\
&= \mathbb{P}\left[B_T \geq \frac{\log(\frac{K}{S_0}) - \mu T}{\sigma}\right] \\
&= \mathbb{P}\left[Z \geq \frac{\log(\frac{K}{S_0}) - \mu T}{\sigma\sqrt{T}}\right] \\
&= \mathbb{P}\left[-Z \geq \frac{\log(\frac{K}{S_0}) - \mu T}{\sigma\sqrt{T}}\right] \\
&= \mathbb{P}\left[Z \leq \frac{\log(\frac{S_0}{K}) + \mu T}{\sigma\sqrt{T}}\right] \\
&= \Phi(d_-)
\end{aligned}$$

Then the first term can be calculated to be:

$$\mathbb{E}\left[S_T\cdot\mathbb{1}_{\{S_T\geq K\}}\right] = \mathbb{E}\left[S_0 e^{\mu T+\sigma\sqrt{T}Z}\cdot\mathbb{1}_{\left\{Z\geq\frac{\log(K/S_0)-\mu T}{\sigma\sqrt{T}}\right\}}\right]$$

Let $\beta = \frac{\log(K/S_0)-\mu T}{\sigma\sqrt{T}}$. Then this expectation is equal to

$$\begin{aligned}
\mathbb{E}[S_0 e^{\mu T+\sigma\sqrt{T}z}\cdot\mathbb{1}_{\{z\geq\beta\}}] &= S_0 e^{\mu T}\int_\beta^\infty e^{\sigma\sqrt{T}x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx \\
&= S_0 e^{\mu T+\frac{1}{2}\sigma^2 T}\int_\beta^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2+\sigma\sqrt{T}x-\frac{1}{2}\sigma^2 T}\,dx \\
&= S_0 e^{\mu T+\frac{1}{2}\sigma^2 T}\int_{\beta-\sigma\sqrt{T}}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\,dy \\
&= S_0 e^{\mu T+\frac{1}{2}\sigma^2 T}\mathbb{P}\left[Z \geq \frac{\log(\frac{K}{S_0}) - \mu T}{\sigma\sqrt{T}} - \sigma\sqrt{T}\right] \\
&= S_0 e^{\mu T+\frac{1}{2}\sigma^2 T}\mathbb{P}\left[Z \leq \frac{\log(\frac{S_0}{K}) + \mu T + \sigma^2 T}{\sigma\sqrt{T}}\right] \\
&= \Phi(d_+)
\end{aligned}$$

We can finally put it all together and see

$$\begin{aligned}
e^{-rT}\mathbb{E}[S_T \cdot \mathbb{1}_{\{S_T \geq K\}}] - Ke^{-rT}\mathbb{E}[\mathbb{1}_{\{S_T \geq K\}}] &= e^{-rT}S_0 e^{\mu T + \frac{1}{2}\sigma^2 T}\Phi(d_+) - Ke^{-rT}\Phi(d_-) \\
&= S_0\Phi(d_+) - Ke^{-rT}\Phi(d_-)
\end{aligned}$$

□

4 Stochastic Calculus

Suppose we have a standard Brownian motion $(B_t)_{t\geq 0}$. What do the following expressions mean?

$$\int_0^t B_s\,ds, \quad \text{and} \quad \int_0^t B_s\,dB_s.$$

More generally, for any stochastic process $(X_t)_{t\geq 0}$, we want to make sense of

$$\int_0^t X_s\,ds \quad \text{and} \quad \int_0^t X_s\,dB_s.$$

For the latter, we say that $(X_t)_{t\geq 0}$ is *integrated with respect to the Brownian motion* $(B_t)_{t\geq 0}$. This more generally is called a *stochastic integral*.

4.1 Introduction

We want to define the integral

$$\int_0^t B_s\,ds$$

pathwise. For each scenario/realization ω, the function $s \mapsto B_s(\omega)$ is continuous and the integral $\int_0^t B_s(\omega)\,ds$ is well defined as the limit of a Riemann sum

$$\int_0^t B_s(\omega)\,ds \triangleq \lim_{n\to\infty} \sum_{i=1}^n B_{t_{i-1}}(\omega)(t_i - t_{i-1}), \quad \forall\omega,$$

for any partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ of $[0,t]$ such that the length of the sub-intervals converge to 0 as $n \to \infty$.

Now we would like to define the integral

$$\int_0^t B_s\,dB_s$$

pathwise. Naively, we could, for all scenarios ω, set

$$\int_0^t B_s(\omega)\,dB_s(\omega) \triangleq \lim_{n\to\infty} \sum_{i=1}^n B_{t_{i-1}}(\omega)(B_{t_i}(\omega) - B_{t_{i-1}}(\omega)).$$

The problem with this is that the limit does not exist for all ω. To remedy this, we can consider the limit in the *mean-square* sense.

Definition 36 (Mean-Square Convergence). A sequence of random variables $X_0, X_1, \dots$ is said to converge to X in *mean-square* if

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0.$$

This gives rise to the so-called *Itô integral*. We define the stochastic integral of a process $(X_t)_{t\geq 0}$ with respect to Brownian motion $(B_t)_{t\geq 0}$ as the mean-square limit of

$$\lim_{n\to\infty} \sum_{i=1}^{n} X_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) \stackrel{\Delta}{=} \int_0^t X_s \, dB_s$$

Some things to keep in mind:

- This requires technical assumptions on the process $(X_t)_{t\geq 0}$. (Learn in PSTAT 213!)
- $\int_0^t X_s \, dB_s$ is a random variable for all $t > 0$.
- $\left(\int_0^t X_s \, dB_s\right)_{t\geq 0}$ is a continuous-time stochastic process.
- If $(X_t)_{t\geq 0}$ satisfies $\mathbb{E}\left[\int_0^t X_s^2 \, ds\right] < \infty$, the process $\left(\int_0^t X_s \, dB_s\right)_{t\geq 0}$ is a martingale with respect to Brownian motion.

Stochastic integration gives rise to a larger class of stochastic processes:

Definition 37 (Itô Process). Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Let $(H_t)_{t\geq 0}$ and $(K_t)_{t\geq 0}$ be a stochastic process satisfying $\int_0^t |H_s| \, ds < \infty$ and $\int_0^t K_s^2 \, ds < \infty$. A continuous-time stochastic process of the form

$$X_t = X_0 + \int_0^t K_s \, ds + \int_0^t H_s \, dB_s, \quad t \geq 0$$

is called an *Itô process*.

The shorthand form of this, also called the *differential form* or the *Itô dynamics*, is

$$dX_t = K_t \, dt + H_t \, dB_t,$$

where K_t is the *drift coefficient* and H_t is the *diffusion coefficient*.

Example 27. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. The Brownian motion with drift parameter $\mu \in \mathbb{R}$ and variance parameter $\sigma^2 > 0$, starting at $x \in \mathbb{R}$ given by

$$X_t = x + \mu t + \sigma B_t = x + \int_0^t \mu \, ds + \int_0^t \sigma \, dB_s$$

is an Itô process with $K_t = \mu$ and $H_t = \sigma$. The Itô dynamics of this will be

$$dX_t = \mu\, dt + \sigma\, dB_t.$$

In particular, this shows us that $(B_t)_{t\geq 0}$ is an Itô process with $\mu = 0$ and $\sigma = 1$. Now we would like to explicitly evaluate some stochastic integrals.

Example 28. Evaluate the stochastic integral

$$\int_0^t dB_s.$$

Solution. We have that our function is $g(x) = 1$. Thus our integral is normally distributed with mean 0 and variance $\int_0^t ds = t$, which is identical to B_t. Furthermore, our integral defines a continuous Gaussian process with mean 0 and covariance function

$$\int_0^{\min\{s,t\}} dx = \min\{s,t\}.$$

Thus $\left(\int_0^t dB_s\right)_{t\geq 0}$ is a standard Brownian motion, and we conclude that

$$\int_0^t dB_s = \boxed{B_t}.$$

□

Example 29. Evaluate the stochastic integral

$$\int_0^t e^s\, dB_s.$$

Solution. The integral will be normally distributed with mean 0 and variance

$$\int_0^t (e^s)^2\, ds = \int_0^t e^{2s}\, ds = \boxed{\frac{1}{2}e^{2t}}.$$

□

Stochastic integrals follow many of the same properties as normal integrals do. For example, for functions g and h and constants α and β, we find

$$\int_a^b [\alpha g(s) + \beta h(s)]\, dB_s = \alpha \int_a^b g(s)\, dB_s + \beta \int_a^b h(s)\, dB_s. \quad \textit{(Linearity)}$$

Then for $a < c < b$,

$$\int_a^b g(s)\, dB_s = \int_a^c g(s)\, dB_s + \int_c^b g(s)\, dB_s.$$

Proposition 30 (Integration by Parts)**.** For g, a differentiable function, we have that

$$\int_0^t g(s)\, dB_s = g(t)B_t - \int_0^t B_s g'(s)\, ds.$$

4.2 Itô Calculus

We can define stochastic integration for Itô processes.

Definition 38 (Itô Integral)**.** Let $X_t = X_0 + \int_0^t K_s\,ds + \int_0^t H_s\,dB_s$ be an Itô process and let $(L_t)_{t\geq 0}$ be a stochastic process such that $\int_0^t |L_s K_s|\,ds < \infty$ and $\int_0^t L_s^2 H_s^2\,ds < \infty$. Then we can define the *Itô integral* as

$$Y_t = \int_0^t L_s\,dX_s = \int_0^t L_s K_s\,ds + \int_0^t L_s H_s\,dB_s.$$

Note that the stochastic integral

$$(Y_t)_{t\geq 0} = \left(\int_0^t L_s\,dX_s\right)_{t\geq 0}$$

is again an Itô process with differential form/Itô dynamics of

$$dY_t = L_t dX_t = L_t K_t\,dt + L_t H_t\,dB_t,$$

because $dX_t = K_t\,dt + H_t\,dB_t$.

Proposition 31 (Itô's Lemma)**.** Let $X_t = X_0 + \int_0^t K_s\,ds + \int_0^t H_s\,dB_s$ be an Itô process and let $f : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function. Then, the stochastic process $(f(X_t))_{t\geq 0}$ is again an Itô process of the form

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s)\,dX_s + \frac{1}{2}\int_0^t f''(X_s)H_s^2\,ds \\ &= f(X_0) + \int_0^t \left(f'(X_s)K_s + \frac{1}{2}f''(X_s)H_s^2\right)ds + \int_0^t f'(X_s)H_s\,dB_s. \end{aligned}$$

The Itô dynamics of the process $(f(X_t))_{t\geq 0}$ will be:

$$\begin{aligned} df(X_t) &= f'(X_t)\,dX_t + \frac{1}{2}f''(X_t)H_t^2\,ds \\ &= \left(f'(X_t)K_t + \frac{1}{2}f''(X_t)H_t^2\right)dt + f'(X_t)H_t\,dB_t \end{aligned}$$

Proposition 32 (Extension of Itô's Lemma)**.** Let $X_t = X_0 + \int_0^t K_s\,ds + \int_0^t H_s\,dB_s$ be an Itô process and let $g : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function. Then the stochastic process $(g(t,X_t))_{t\geq 0}$ is again an Itô process of the form

$$\begin{aligned} g(t,X_t) &= g(0,X_0) + \int_0^t \frac{\partial g}{\partial t}(s,X_s)\,ds + \int_0^t \frac{\partial g}{\partial x}(s,X_s)\,dX_s + \frac{1}{2}\int_0^t \frac{\partial^2 g}{\partial x^2}(s,X_s)H_s^2\,ds \\ &= g(0,X_0) + \int_0^t \left(\frac{\partial g}{\partial t}(s,X_s) + \frac{\partial g}{\partial x}(s,X_s)K_s + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(s,X_s)H_s^2\right) ds \\ &\quad + \int_0^t \frac{\partial g}{\partial s}(s,X_s)H_s\,dB_s \end{aligned}$$

Example 30. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Compute the Itô dynamics of $(B_t)_{t\geq 0}$.

Solution. We apply Itô's Lemma to $(B_t)_{t\geq 0}$, which is an Itô process by

$$B_t = B_0 + \int_0^t 0\,ds = \int_0^t 1\,dB_s,$$

and the function $f(x) = x^2$. We have that $f'(x) = 2x$ and $f''(x) = 2$. This yields

$$\begin{aligned} B_t^2 = f(B_t) &= f(B_0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\cdot 1^2\,ds \\ &= 0 + \int_0^t 2\,dB_s + \frac{1}{2}\int_0^t 2\,ds \\ &= 2\int_0^t B_s\,dB_s + t \end{aligned}$$

Hence our Itô dynamics will be

$$\int_0^t dB_s^2 = \int_0^t ds + \int_0^t 2B_s\,dB_s \Longleftrightarrow \boxed{dB_s^2 = ds + 2B_s\,dB_s}$$

In particular, observe that we obtain an explicit expression for the stochastic integral of Brownian motion with respect to Brownian motion, namely

$$\int_0^t B_s\,dB_s = \frac{1}{2}(B_t^2 - t),$$

which is a martingale with respect to $(B_t)_{t\geq 0}$. □

Example 31. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Show that

$$\int_0^t sB_s\,dB_s = \frac{1}{2}tB_t^2 - \frac{1}{2}\int_0^t (B_s^2 + s)\,ds.$$

Solution. We apply the Extension of Itô's Lemma to $(B_t)_{t\geq 0}$ and the function $g(t,x) = tx^2$. We have that

$$\frac{\partial g}{\partial t}(t,x) = x^2, \quad \frac{\partial g}{\partial x}(t,x) = 2+x, \quad \frac{\partial^2 g}{\partial x^2}(t,x) = 2t.$$

This yields:

$$\begin{aligned} g(t,B_t) = tB_t^2 = g(0,B_t) + \int_0^t \frac{\partial g}{\partial t}(s,B_s)\,ds + \int_0^t \frac{\partial g}{\partial x}(s,B_s)\,dB_s + \frac{1}{2}\int_0^t \frac{\partial^2 g}{\partial x^2}(s,B_s)\cdot 1^2\,ds \\ = \int_0^t B_s^2\,ds + \int_0^t 2sB_s\,dB_s + \frac{1}{2}\int_0^t 2s\,ds \\ = 2\int_0^t sB_s\,dB_s + \int_0^t (B_s^2 + s)\,ds \end{aligned}$$

This means

$$\int_0^t sB_s\,dB_s = \frac{1}{2}tB_t^2 - \int_0^t (B_s^2+s)\,ds$$

The Itô dynamics of this will be

$$dg(t,B_t) = d(t,B_t^2) = (B_t^2 + t)\,dt + 2tB_t d\,B_t.$$

□

4.3 Stochastic Differential Equations

Definition 39 (Stochastic Differential Equation)**.** A *stochastic differential equation*, or SDE, is an equation of the form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \quad X_0 = x,$$

where $x \in \mathbb{R}$, $b : \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \to \mathbb{R}$ are given.

We want to find a stochastic process $(X_t)_{t\geq 0}$ which satisfies the integral equation

$$X_t = x + \int_0^t b(X_s)\,ds + \int_0^t \sigma(X_s)\,dB_s, \quad t \geq 0.$$

Example 32. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion, $x_0, \mu \in \mathbb{R}$ and $\sigma > 0$. Find a solution $(X_t)_{t\geq 0}$ to the stochastic differential equation

$$dX_t = \mu X_t\,dt + \sigma X_t\,dB_t, \quad X_0 = x_0.$$

Solution. We claim that geometric Brownian motion,

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}, \quad t \geq 0$$

solves the SDE. Observe that $X_t = f(Z_t)$, where $f(x) = x_0 e^x$ and $(Z_t)_{t\geq 0}$ is an Itô process given by

$$Z_t = \left(\mu - \frac{1}{2}\right)t + \sigma B_t = \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) ds + \int_0^t \sigma \, dB_s$$

$$\iff \quad dZ_t = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma \, dB_t.$$

Then we apply Itô's Lemma to $(Z_t)_{t\geq 0}$ and $f(x) = x_0 e^x$. We have

$$f'(x) = x_0 e^x = f(x), \quad f''(x) = x_0 e^x = f(x).$$

This yields, in Itô dynamics,

$$\begin{aligned} dX_t = df(Z_t) &= f'(Z_t)\, dZ_t + \frac{1}{2} f''(Z_t)\sigma^2 \, dt \\ &= f(Z_t)\left(\mu - \frac{1}{2}\sigma^2\right) dt + f(Z_t)\sigma \, dB_t + \frac{1}{2} f(Z_t)\sigma^2 \, dt \\ &= f(Z_t)\mu \, dt + f(Z_t)\sigma \, dB_t \\ &= X_t \mu \, dt + X_t \sigma \, dB_t \end{aligned}$$

Hence X_t solves the SDE. □

Example 33 (Ornstein-Uhlenbeck Process)**.** Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and $x_0, \mu, \sigma > 0$ are constants. Find a solution $(X_t)_{t\geq 0}$ to the SDE

$$dX_t = -r(X_t - \mu)\, dt + \sigma \, dB_t, \quad X_0 = x_0.$$

Solution. We apply the Extension of Itô's Lemma to the Itô process $(X_t)_{t\geq 0}$ satisfying the SDE above and the function $g(t,x) = e^{rt}x$. We find that

$$\frac{\partial g}{\partial t}(t,x) = re^{rt}x, \quad \frac{\partial g}{\partial x}(t,x) = e^{rt}, \quad \frac{\partial^2 g}{\partial x^2}(t,x) = 0.$$

This yields in differential form

$$\begin{aligned} dg(t,X_t) &= \frac{\partial g}{\partial t}(t,X_t)\, dt + \frac{\partial g}{\partial x}(t,X_t)\, dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)\sigma^2 \, dt \\ &= re^{rt}X_t \, dt - re^{rt}(X_t - \mu)\, dt + e^{rt}\sigma \, dB_t \\ &= \mu r e^{rt} \, dt + \sigma e^{rt} \, dB_t \end{aligned}$$

Hence

$$\begin{aligned} g(t,X_t) = e^{rt}X_t &= g(0,X_0) + \int_0^t \mu r e^{rs} \, ds + \int_0^t \sigma e^{rs} \, dB_s \\ &= x_0 + \mu(e^{rt} - 1) + \int_0^t \sigma e^{rs} \, dB_s \end{aligned}$$

This implies that the solution $(X_t)_{t\geq 0}$ to the SDE is given by

$$X_t = e^{-rt}x_0 + \mu(1 - e^{-rt}) + e^{-rt}\sigma \int_0^t e^{rs}\, dB_s.$$

□