Stochastic Modeling Compendium

Torbjørn Baadsvik

November 24, 2020

Contents

1 Conditional probability

if X and Y are random variables on the same probability space, and the variance of Y is finite, then

$$\mathrm{Var}\left[Y\right] = \mathrm{E}\left[\mathrm{Var}\left[Y \mid x\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[Y \mid x\right]\right]$$

2 Discrete-Time Markov Chains (DTMCs)

2.1 Definition of the DTMC

A DTMC $\{X_n\}$ is a stohastic process with the following properties:

- **Markov property**

$$\begin{aligned}&\Pr\left\{X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1} \ldots, X_0 = i_0\right\}\\ =&\Pr\left\{X_{n+1} = j \mid X_n = i\right\}\end{aligned}$$

- **Stationary transition probabilities**

$$\begin{aligned}&\Pr\left\{X_{n+1} = j \mid X_n = i_n, n = m\right\}\\ =&\Pr\left\{X_{n+1} = j \mid X_n = i_n\right\}\end{aligned}$$

- **Countable state space** The state space $\mathcal{S}$ is countable

2.2 Transition Probability Matrices (TPMs)

2.2.1 General TPMs

Let the state space $\mathcal{S}$ of a DTMC be represented by the set

$$\mathcal{S} = \{0, 1, 2, \ldots\}$$

We introduce the notation

$$P_{ij} = \Pr\left\{X_{n+1} = j \mid X_n = i\right\}, \quad i, j \in \mathcal{S}$$

The *transition probability matrix*(TPM) of a DTMC is

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \ldots \\ P_{10} & P_{11} & P_{12} & \ldots \\ P_{20} & P_{21} & P_{22} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

All TPMs satisfy

$$\sum_{j \in \mathcal{S}} \mathbf{P}_{ij} = 1, \quad \forall i \in \mathcal{S}$$

2.2.2 Doubly Stochastic TPMs

A TPM $\mathbf{P}$ is called *doubly stochastic* if

$$\mathbf{P}_{ij} \geq 0, \quad \sum_{k \in \mathcal{S}} \mathbf{P}_{ik} = \sum_{k \in \mathcal{S}} \mathbf{P}_{kj} = 1 \quad \forall i, j \in \mathcal{S}$$

2.2.3 Regular TPMs

A TPM $\mathbf{P}$ is *regular* if

$$(\mathbf{P}^k)_{ij} > 0 \quad \forall i, j \in \mathcal{S}, \quad \forall k \in \mathbb{N}^+$$

2.3 Classification of States

- **Period of a State** The *period* of a state i is the greatest common divisor(gcd) of the set $\{n\}_{n \in \mathbb{N}^+}$ for which $P_{ii}^{(n)} > 0$.
- **Aperiodic** A state with period 1 is called *aperiodic*.
- **Recurrent State** A state i is *recurrent* if the probability of returning to i at some point in time is 1. Let $X_0 = i$. Let N_i be the time until the process first returns to state i. Then the probability distribution of N_i is
$$\Pr\{N_i = n\} = \Pr\{X_n = i \mid X_v \neq i, \quad \forall i \in \{1, 2, \ldots, n-1\}\}$$
If a state i is recurrent, then by the law of total probability we have
$$\sum_{n=1}^{\infty} \Pr\{N_i = n\} \ = \ 1$$
- **Transient State** A state is *transient* if it is not recurrent. Then
$$\sum_{n=1}^{\infty} \Pr\{N_i = n\} \ < \ 1$$
- **Accessible State** A state j is said to be *accessible* from state i if there exists some $n \in \mathbb{N}$ such that
$$P_{ij}^{(n)} > 0$$
- **Communicating States** If two states i and j are both accessible from each other they are said to *communicate* and we write $i \leftrightarrow j$. We have the following properties
  - **Reflexivity**
  $$i \leftrightarrow i$$
  - **Symmetry**
  $$i \leftrightarrow j \quad \Longleftrightarrow \quad j \leftrightarrow i$$
  - **Transitivity**
  $$i \leftrightarrow j \text{ and } j \leftrightarrow k \quad \Longrightarrow \quad i \leftrightarrow k$$
- **Equivalence Class** Let $\mathcal{S}_1 \subseteq \mathcal{S}$ be a subset of $\mathcal{S}$ satisfying the property
$$i \leftrightarrow j, \quad \forall i, j \in \mathcal{S}_1$$
If adding any additional state $j \in \mathcal{S} \backslash \mathcal{S}_1$ to S_1 would break this property, then $\mathcal{S}_1$ is an *equivalence class* of the Markov chain. Just as with states, classes are either recurrent or transient. They can also be assigned a period.
- **Irreducible Markov Chain** A Markov chain is *irreducible* if it has only a single equivalence class $\mathcal{S}_1$ such that $\mathcal{S}_1 = \mathcal{S}$
- **Reducible Markov Chain** A Markov chain is *reducible* if it contains more than one equivalence class.

2.4 First Step Analysis

Assume that a TPM for a DTMC is given.

2.4.1 Expected Time Until Absorption

We say that a state is *absorbing* if, once the process has reached the state, it can never leave. We wish to find the expected time(number of steps) until absorption given $X_0 = j$. Let $\{X_n\}$ be one particular *realization* of the process. We introduce the following notation

- $t_i(n)$: time until absorption from time n given that the process is in state i.
- $T_i(n)$: $\mathrm{E}\left[t_i(n)\right]$

From the Markov property we have

$$T_i(n) = T_i(m), \quad \forall m, n \in \mathbb{N}$$

Hence, we replace $T_i(n)$ by

$$T_i(n) = T_i$$

Let $N + 1$ be the number of states in the state space(assuming it is finite). Assume the process is not already in an absorbing state at time n. Let $\mathcal{A}$ denote the set of absorbing states. Then $T_r = 0 \quad \forall r \in \mathcal{A}$. Now consider the outcome after a single step. We have from the law of total expectation

$$\begin{aligned}
T_i &= 1 + \sum_{k=0}^{N} P_{ik} T_k \\
T_i &= 1 + P_{ii} T_i + \sum_{k \neq i} P_{ik} T_k \\
T_i &= \frac{1 + \sum_{k \neq i} P_{ik} T_k}{1 - P_{ii}}
\end{aligned}$$

On the other hand, we have

$$T_k = \frac{1 + \sum_{j \neq k} P_{kj} T_j}{1 - P_{kk}}, \quad \forall k \in \mathcal{S}$$

Writing out this equation for every $k \in \mathcal{S}$ gives rise to a system of $N + 1$ equations which may be solved to find T_i and $T_k \quad \forall k \in \mathcal{S}$.

2.4.2 Probability of Absorption into Specific State

We shall maintain the notation from section (2.4.1). Let j be an absorbing state and i be a state that is not absorbing. We wish to find

$$a_{ij} = \Pr\left\{X_{n+t_i(n)} = j \mid X_n = i\right\}$$

By the law of total probability we get

$$a_{ij} = \sum_{k=0}^{N} P_{ik} a_{kj}$$

where $a_{jj} = 1$ and $a_{rj} = 0, \quad \forall r \in \mathcal{A} \setminus \{j\}$. Hence, we again arrive at a system of $N + 1$ equations.

2.5 Long Run Behaviour

2.5.1 Limiting Distribution

Any set $\{\pi_j\}_{j \in \mathcal{S}}$ satisfying

$$\pi_j = \sum_{k \in \mathcal{S}} \pi_k P_{kj}, \quad \sum_{k \in \mathcal{S}} \pi_k = 1, \quad \lim_{n \to \infty} P_{ij}^{(n)} = \pi_j > 0, \quad \forall i \in \mathcal{S} \tag{1}$$

is a *limiting distribution* of the Markov chain.

2.5.2 Stationary Distribution

Any set $\{\pi_i\}_{i\in\mathcal{S}}$ satisfying

$$\pi_j = \sum_{i\in\mathcal{S}} \pi_i P_{ij}, \quad \sum_{i\in\mathcal{S}} \pi_i = 1, \quad \pi_j \geq 0, \quad \forall j \in \mathcal{S} \tag{2}$$

is a *stationary distribution* of the Markov chain.
Note: Not every stationary distribution is a limiting distribution. As an example, consider the Markov chain with TPM

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This process has the stationary distribution $\pi = \left[\frac{1}{2}, \frac{1}{2}\right]$. However, the limit in (1) doesn't exist since $P_{01}^{(n)} = P_{10}^{(n)}$ is 1 for odd n and 0 for even n, $\quad \forall n \in \mathbb{N}$. Hence, we can't find any limiting distribution for the process.

2.5.3 The Basic Limit Theorem of Markov Chains

Theorem 1 (The Basic Limit Theorem of Markov Chains)**.** *Let i be a recurrent state of a MC. Then*

$$lim_{n\to\infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=1}^{\infty} nPr\{N_i = n\}} = \frac{1}{\mathrm{E}[N_i]}$$

Proof. Since i is recurrent by definition we have

$$\sum_{n=1}^{\infty} \Pr\{N_i = n\} = 1$$

Then

$$\mathrm{E}[N_i] = \sum_{n=1}^{\infty} n\Pr\{N_i = n\}$$

and $\mathrm{E}[N_i]$ is the expected time between two successive visits to state i. Then

$$\lim_{n\to\infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=1}^{\infty} n\Pr\{N_i = n\}} = \frac{1}{\mathrm{E}[N_i]}$$

□

2.6 Chapman-Kolmogorov Equation

Let $\mathcal{S}$ denote the state space of a DTMC, and let $P_{ij}^{(n)}$ denote the probability of transitioning from state i to state j in n steps. Then

$$P_{ij}^{(n+m)} = \sum_{k\in\mathcal{S}} P_{ik}^{(n)} P_{kj}^{(m)}, \quad \forall m, n \in \mathbb{N} \tag{3}$$

3 Poisson Processes

3.1 The Poisson Distribution

3.1.1 Probability Density Function (PDF)

The *Poisson distribution* is given by

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}, \quad \forall x \in \mathbb{N}$$

With $\mathrm{E}[x] = \mathrm{Var}[x] = \mu$.

3.1.2 Properties

The Poission distribution has the following properties

- **Theorem 2** (PDF of Sum of Independent Poisson Random Variables). *Let $\{X_i\}_{i\in\{1,\ldots,N\}}$ be a set of N independent Poisson random variables with respective rate parameters $\{\mu_i\}_{i\in\{1,\ldots,N\}}$. Then the random variable*

$$Z = \sum_{i=1}^{N} X_i$$

is a Poisson random variable with rate parameter

$$\mu_Z = \sum_{i=1}^{N} \mu_i$$

- **Binomial Variable Conditional on Poisson Variable** Let N be a Poisson variable with rate parameter μ. Let M be a binomial variable with parameters N and p. Then the unconditional distribution of M is Poisson with rate parameter μp.

Proof.

$$\begin{aligned}
\Pr\{M = m\} &= \sum_{n=m}^{\infty} \Pr\{M = m \mid N = n\} \Pr\{N = n\} \\
&= \sum_{n=m}^{\infty} \left\{ \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \right\} \left\{ \frac{\mu^n e^{-\mu}}{n!} \right\} \\
&= \frac{e^{-\mu}(\mu p)^m}{m!} \sum_{n=m}^{\infty} \frac{(\mu(1-p))^{n-m}}{(n-m)!}
\end{aligned}$$

Noting that the sum is the series expansion of $e^{\mu(1-p)}$ we arrive at

$$\Pr\{M = m\} = \frac{e^{-\mu}(\mu p)^m}{m!} e^{\mu(1-p)} = \frac{e^{-\mu p}(\mu p)^m}{m!}, \quad \forall k \in \mathbb{N}$$

□

3.2 The Poisson Process

3.2.1 Properties

A *Poisson process* $\{X(t)\}_{t\geq 0}$ with rate parameter $\lambda > 0$ is a stochastic process with the following properties

- **Independent process increments** Let $h > 0$ be a fixed time increment. For all t_0 and t_1 satisfying $0 < t_0 < t_1 - h$ we have

$$\Pr\{X(t_0 + h) - X(t_0) = k\} = \Pr\{X(t_1 + h) - X(t_1) = k\}$$

- **Process increments are Poisson** For $h > 0$ and $t > 0$ we have

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

3.2.2 Postulates

The Poisson process has the following postulates

1. $X(0) = 0$
2. $\Pr\{X(t+h) - X(t) = 1\} = \lambda h + o(h) \quad \text{as } h \downarrow 0$

   Therefore

$$\lim_{h\to 0^+} \frac{\Pr\{X(t+h) - X(t) = 1 \mid X(t) = x\}}{h} = \lambda$$

3. $\Pr\{X(t+h) - X(t) = 0\} = 1 - \lambda h + o(h) \quad \text{as } h \downarrow 0$

   Therefore
   $$\lim_{h\to 0^+} \frac{\Pr\{X(t+h) - X(t) = 0 \mid X(t) = x\}}{h} = 1 - \lambda$$

Proof. We prove (2).

$$\begin{aligned}\Pr\{X(t+h) - X(t) = 1\} &= \frac{(\lambda h)e^{-\lambda h}}{1!} \\ &= (\lambda h)\left(1 - \lambda h + \frac{1}{2}(\lambda h)^2 - \dots\right) \\ &= \lambda h + o(h) \quad \text{as } h \downarrow 0\end{aligned}$$

where we used the series expansion of $e^{-\lambda h}$. □

The postulate (3) follows readily by employing the same procedure.

3.2.3 Nonhomogeneous Processes

A Poisson process is *nonhomogeneous* if it's rate parameter varies with time, i.e. $\lambda = \lambda(t)$

3.3 The Law of Rare Events

Consider a binomial distribution

$$\Pr\{X = k; N, p\} = \frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}$$

Now let $N \to \infty$ and $p \to 0$. Then the distribution becomes the Poisson distribution, i.e.

$$\Pr\{X = k; \mu\} = \frac{e^{-\mu}\mu^k}{k!}$$

3.3.1 Approximation using the Poisson Distribution

Let $\{\epsilon_i\}$ be a set of independent Bernoulli random variables, where

$$\Pr\{\mathcal{E}_i = 1\} = p_i, \quad \Pr\{\epsilon_i = 0\} = 1 - p_i$$

Let $S_n = \sum_{i=1}^n \epsilon_i$ and set $\mu = \sum_{i=1}^n p_i$. Then

$$\left|\Pr\{S_n = k\} - \frac{\mu^k e^{-\mu}}{k!}\right| \leq \sum_{i=1}^n p_i^2$$

3.4 Distributions Associated with the Poisson Process

We introduce the following terms:

- **Waiting Time** The *waiting time* W_n is the time of occurrence of the nth event, and is gamma distributed
  $$f_{W_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}e^{-\lambda t}, \quad \forall n \in \mathbb{N}$$
- **Sojourn Time** The differences $S_n = W_{n+1} - W_n$ are called the *sojourn times* of the process and are exponentially distributed
  $$f_{S_k}(s) = \lambda e^{-\lambda s}, \quad s \geq 0$$

Let $\{X(t)\}$ be a Poisson process with rate $\lambda > 0$. Then for $0 < \mu < t$ and $0 \leq k \leq n$

$$\Pr\{X(u) = k \mid X(t) = n\} = \frac{n!}{k!(n-k)!}\left(\frac{u}{t}\right)^k\left(1 - \frac{u}{t}\right)^{n-k}$$

3.5 The Uniform Distribution and Poisson Processes

Theorem 3 (PDF of Waiting Times)**.** *Let $\{W_i\}_{i\in N^+}$ be the occurence times in a Poisson process of rate $\lambda > 0$. Conditioned on $N(t) = n$ the random variables $\{W_i\}_{i\in N^+}$ have the joint PDF*

$$f_{\{W_i\}|X(t)=n}(\{W_i\}) = n!t^{-n}$$

4 Birth and Death Processes (BDPs)

4.1 Postulates

A *birth and death process* is a a generalization of the Poisson process. BDPs constitute a class of *continuous time Markov chains (CTMCs)*. By analogy to the development of a population over time, we let the BDP $X(t)$ denote the population, or state, at time t and define two types of events: *births* and *deaths*. A *birth* corresponds to an *increase* in the population by 1, and a *death* corresponds to a *decrease* in the population by 1. Let i be the current state of the process and define

$$P_{i,k}(h) = \Pr\{X(t+h) - X(t) = k\}$$

which is the probability that the population changes by k during the time interval h. We define λ_i and μ_i to be the *birth rate* and *death rate* respectively of the process in state i. A BDP is defined by the following postulates

$$P_{i,i+1}(h) = \lambda_i h + o(h) \text{ as } h \downarrow 0, \quad i \geq 0 \tag{4a}$$
$$P_{i,i-1}(h) = \mu_i h + o(h) \text{ as } h \downarrow 0, \quad i \geq 1 \tag{4b}$$
$$P_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h) \text{ as } h \downarrow 0, \quad i \geq 1 \tag{4c}$$
$$P_{ij}(0) = \delta_{ij} \tag{4d}$$
$$\mu_0 = 0, \quad \lambda_0 > 0, \quad \mu_i, \lambda_i > 0, \quad \forall i \in N^+ \tag{4e}$$

4.2 Sojourn Times

The *sojourn times* $\{S_i\}$ of a BDP are exponentially distributed with parameter $(\lambda_i + \mu_i)$ and PDF

$$f_{S_i}(t) = (\lambda_i + \mu_i)\exp[-(\lambda_i + \mu_i)] \tag{5}$$

Proof. Define $G_i(t) = \Pr\{S_i > t\}$ and let $h \downarrow 0$. By the Markov property and the postulates (4) we have

$$\begin{aligned} G_i(t+h) &= G_i(t)G_i(h) = G_i(P_{i,i}(h) + o(h)) \\ &= G_i(t)(1 - (\lambda_i + \mu_i)h) + o(h)) \end{aligned}$$

Rearranging yields

$$\begin{aligned} \frac{G_i(t+h) - G_i(t)}{h} &= -(\lambda_i + \mu_i)G_i(t) + o(1) \\ G_i'(t) &= -(\lambda_i + \mu_i)G_i(t) \end{aligned}$$

which is a differential equation with solution

$$G_i(t) = \exp[-(\lambda_i + \mu_i)]$$

Then

$$\Pr\{S_i \leq t\} = 1 - \exp[-(\lambda_i + \mu_i)]$$

Taking the derivative we get the pfd of the sojourn time in state i

$$f_{S_i}(t) = (\lambda_i + \mu_i)\exp[-(\lambda_i + \mu_i)]$$

□

4.3 Transition Probabilities

Let W_n and W_{n+1} be any two subsequent waiting times of the process as defined in (3.4). We state without proof that the *transition probabilities* are given by:

$$\Pr\left\{X(W_{n+1}) - X(W_n) = 1\right\} = \frac{\lambda_i}{\lambda_i + \mu_i}$$
$$\Pr\left\{X(W_{n+1}) - X(W_n) = -1\right\} = \frac{\mu_i}{\lambda_i + \mu_i}$$

4.4 Differential Equations

Any BDP satisfies the *backward Kolmogorov differential equations*

$$P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t) \tag{6a}$$
$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t) \tag{6b}$$

Proof. We begin by proving (6a). Let $h \downarrow 0$. By the *Chapman-Kolmogorov equation* (3) and the postulates (4) we have

$$\begin{aligned} P_{0j}(t+h) &= \sum_{i=0}^{\infty} P_{0k}(h) P_{kj}(t) = \sum_{i=0}^{1} P_{0k}(h) P_{kj}(t) + \sum_{i=2}^{\infty} P_{0k}(h) P_{kj}(t) \\ &= \lambda_0 h P_{1j}(t) + (1 - \lambda_0 h) P_{0j}(t) + o(h) + \sum_{k=2}^{\infty} P_{0k}(h) P_{kj}(t) \end{aligned}$$

where for the last summation we have

$$\begin{aligned} \sum_{k=2}^{\infty} P_{0k}(h) P_{kj}(t) &\leq \sum_{k=2}^{\infty} P_{0k}(h) \\ &= 1 - P_{00}(h) - P_{01}(h) \\ &= 1 - (1 - \lambda_0 h) - \lambda_0 h + o(h) \\ &= o(h) \end{aligned}$$

Hence

$$P_{0j}(t+h) = \lambda_0 h P_{1j}(t) + (1 - \lambda_0 h) P_{0j}(t) + o(h)$$

Moving P_{0j} to the left hand side and dividing by h as $h \downarrow 0$ we obtain

$$P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t)$$

We now derive the second equation (6b) in a similar manner

$$\begin{aligned} P_{ij}(t+h) &= \sum_{k=0}^{\infty} P_{ik}(h) P_{kj}(t) \\ &= P_{i,i-1}(h) P_{i-1,j}(t) + P_{i,i}(h) P_{ij}(t) + P_{i,i+1}(h) P_{i+1,j}(t) + \sum_{k}' P_{ik}(h) P_{kj}(t) \end{aligned}$$

where the last summation is over all $k \in \mathbb{N} \backslash \{i-1, i+1\}$. From (4) we get

$$\begin{aligned} \sum_{k}' P_{ik}(h) P_{kj}(t) &\leq \sum_{k}' P_{ik}(h) \\ &= 1 - [P_{i,i}(h) + P_{i,i-1}(h) + P_{i,i+1}(h)] \\ &= 1 - [1 - (\lambda_i + \mu_i)h + o(h) + \mu_i h + o(h) + \lambda_i h + o(h)] \\ &= o(h) \end{aligned}$$

Hence

$$P_{ij}(t+h) = \mu_i h P_{i-1,j}(t) + [1 - (\lambda_i + \mu_i)h] \, P_{ij}(t) + \lambda_i h P_{i+1,j}(t) + o(h)$$

Moving $P_{ij}(t)$ to the left-hand side and dividing by h as $h \downarrow 0$ we obtain

$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t)$$

□

4.5 Limiting Behaviour

A BDP with no absorbing states has a *stationary distribution* that satisfies

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}(t)$$

In this case, the limiting distribution is also a *limiting dstribution* since it satisfies

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j \geq 0 \tag{7}$$

which we will not prove.
Let $t \to \infty$. Inserting (7) into the backward Kolmogorov differential equations (6a) and (6b) we get

$$0 = -\lambda_0 \pi_0 + \mu_1 \pi_1 \tag{8a}$$
$$0 = \lambda_{j-1}\pi_{j-1} - (\lambda_j + \mu_j)\pi_j + \mu_{j+1}\pi_{j+1} \tag{8b}$$

where the left hand sides are zero by (7). The solution of this system is

$$\pi_k = \theta_k \pi_0, \quad \theta_k = \prod_{i=0}^{k-1} \frac{\lambda_k}{\mu_{k+1}}, \quad \theta_0 = 1 \tag{9}$$

Proof. We prove (9) by induction. Assume (9) holds. Then by (8a) we have

$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1} = \theta_1 \pi_0$$

Inserting into (8b) yields

$$\begin{aligned}\mu_{j+1}\pi_{j+1} &= (\lambda_j + \mu_j)\theta_j \pi_0 - \lambda_{j-1}\theta_{j-1}\pi_0 \\ &= \lambda_j \theta_j \pi_0 + (\mu_j \theta_j - \lambda_{j-1}\theta_{j-1})\pi_0 \\ &= \lambda_j \theta_j \pi_0\end{aligned}$$

which, when rearranged, leads to

$$\pi_{j+1} = \theta_{j+1}\pi_0$$

which completes the induction part of the the proof. □

What remains is to find an expression for π_0. Assuming the limiting distribution sums to one, we have

$$\sum_{k=0}^{\infty} \pi_k = \pi_0 \sum_{k=0}^{\infty} \theta_k = 1$$

Hence

$$\pi_0 = \frac{1}{\sum_{k=0}^{\infty} \theta_k}$$

4.6 Probability of Absorption into State 0

Just as for DTMCs, one may analyse problems that arrive from processes with absorbing states. Consider a process with $\lambda_0 = \mu_0 = 0$. One example of such a process is a linear-growth BDP without immigration for which $\lambda_n = n\lambda$ and $\mu_n = n\mu$. We introduce the following notation:

- a_i: probability of absorption into state 0 from the initial state i
- b_i: probability that the next jump of the process is to state $i+1$ (birth) given that the current state is i.
- d_i: probability that the next jump of the process is to state $i-1$ (death) given that the current state is i.

Absorption into state 0 is not a certain event, since the process may wander forever among the states $\mathbb{N}^+$, or it may drift toward infinity. Hence, it isn't generally true that $a_i = 1, \forall i \in \mathbb{N}$. The probability of absorption into state 0 given initial state i is

$$a_i = \frac{\sum_{i=m}^{\infty} \rho_i}{1 + \sum_{i=1}^{\infty} \rho_i}, \quad \rho_i = \prod_{j=0}^{i} \frac{\mu_j}{\lambda_j} \tag{10}$$

Proof. From (4.3) we have

$$b_i = \frac{\lambda_i}{\mu_i + \lambda_i}, \quad d_i = \frac{\mu_i}{\mu_i + \lambda_i}$$

Hence

$$a_i = \frac{\lambda_i}{\mu_i + \lambda_i} a_{i+1} + \frac{\mu_i}{\mu_i + \lambda_i} a_{i-1} \tag{11}$$

Rewriting (11) we get

$$(a_{i+1} - a_i) = \frac{\mu_i}{\lambda_i}(a_i - a_{i-1})$$

Let $v_i = a_{i+1} - a_i$. Then we obtain the recursive formula

$$v_i = \frac{\mu_i}{\lambda_i} v_{i-1}$$

Writing out this recursion we get

$$v_i = \rho_i v_0, \quad \rho_i = \prod_{j=0}^{i} \frac{\mu_j}{\lambda_j}$$

Then

$$v_i = \rho_i v_0 = \rho_i(a_1 - a_0) = \rho_i(a_i - 1), \quad \forall i \geq 1$$

Summing over v_i from $i = 1$ to $i = m - 1$ yields

$$\sum_{i=1}^{m-1} v_i = a_m - a_1 = (a_1 - 1) \sum_{i=1}^{m-1} \rho_i \tag{12}$$

Since $a_m < 1, \forall m \in \mathbb{N}^+$ we have

$$\sum_{i=1}^{\infty} \rho_i < \infty$$

Now let $m \to \infty$. Then, it can be shown that $a_m \to 0$. Therefore (12) gives

$$a_1 = \frac{\sum_{i=1}^{\infty} \rho_i}{1 + \sum_{i=1}^{\infty} \rho_i} \tag{13}$$

Let $S_L^U = \sum_{i=L}^{U} \rho_i$. For general m we get from (12) and (13)

$$\begin{aligned}
a_m &= (a_1 - 1)S_1^{m-1} + a_1 \\
&= \frac{S_1^\infty S_1^{m-1}}{1 + S_1^\infty} - S_1^{m-1} + \frac{S_1^\infty}{1 + S_1^\infty} \\
&= \frac{S_1^\infty S_1^{m-1} - (1 + S_1^\infty)S_1^{m-1} + S_1^\infty}{1 + S_1^\infty} \\
&= \frac{S_m^\infty}{1 + S_1^\infty} \\
&= \frac{\sum_{i=m}^{\infty} \rho_i}{1 + \sum_{i=1}^{\infty} \rho_i}
\end{aligned}$$

□

4.7 Mean Time Until Absorption

Let w_i be the mean time until absorption starting from state i. Then

$$w_m = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} + \sum_{k=1}^{m-1} \rho_k \sum_{i=k+1}^{\infty} \frac{1}{\lambda_i \rho_i}, \quad \rho_k = \prod_{j=1} \frac{\mu_j}{\lambda_j} \tag{14}$$

Proof. Since the sojourn times of the process are exponentially distributed with parameter $\lambda_i + \mu_i$ we have

$$w_i = \frac{1}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i} w_{i+1} + \frac{\mu_i}{\lambda_i + \mu_i} w_{i-1} \tag{15}$$

Let $z_i = w_i - w_{i+1}$ then rearranging (15) leads to the recursive formula

$$z_i = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} z_{i-1}, \quad i \geq 1$$

Writing out the recursion leads to

$$z_m = \sum_{i=1}^{m} \frac{1}{\lambda_i} \frac{\rho_m}{\rho_i} + \rho_m z_0, \quad \rho_k = \prod_{j=1}^{k} \frac{\mu_j}{\lambda_j} \tag{16}$$

Since $z_m = w_m - w_{m+1}$ and $z_0 = w_0 - w_1 = -w_1$ we get

$$\frac{1}{\rho_m}(w_m - w_{m+1}) = \sum_{i=1}^{m} \frac{1}{\lambda_i \rho_i} - w_1 \tag{17}$$

If $\sum_{i=1}^{m} \frac{1}{\lambda_i \rho_i} = \infty$ then $w_1 = \infty$. Suppose $\sum_{i=1}^{m} \frac{1}{\lambda_i \rho_i} < \infty$ and let $m \to \infty$. Then

$$w_1 = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} - \lim_{m\to\infty} \frac{1}{\rho_m}(w_m - w_{m+1})$$

We state without proof that

$$\lim_{m\to\infty} \frac{1}{\rho_m}(w_m - w_{m+1}) = 0$$

Then

$$w_1 = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} \tag{18}$$

Hence, from (16) we have

$$w_m = \sum_{i=1}^{m} \frac{1}{\lambda_i} \frac{\rho_m}{\rho_i} - \rho_m w_1$$

Summing $(w_k - w_{k+1})$, from $k = 1$ to $k = m - 1$ yields a telescoping sum

$$\sum_{k=1}^{m-1} (w_k - w_{k+1}) = w_1 - w_m \tag{19}$$

This sum can be evaluated by (17)

$$\begin{aligned} w_m &= w_1 - \sum_{k=1}^{m-1} \rho_k \left(\sum_{i=1}^{k} \frac{1}{\lambda_i \rho_i} - \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} \right) \\ &= \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} + \sum_{k=1}^{m-1} \rho_k \sum_{i=k+1}^{\infty} \frac{1}{\lambda_i \rho_i} \end{aligned}$$

□

4.8 Finite-State Continuous Time Markov Chains

Assume we are given a CTMC (not necessarily BDP) with finite state space. Then we can represent the transition probabilities $P_{ij}(t)$ by the matrix $\mathbf{P}(t)$ such that

$$\mathbf{P}(t)_{ij} = P_{ij}(t)$$

From the Chapman-Kolmogorov equation we have

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

We employ the postulate

$$\lim_{t\to 0^+} P_{ij}(t) = \begin{cases} 1, & i = 1 \\ 0, & i \neq 1 \end{cases}$$

Hence, if we approach t from the *right* we obtain

$$\lim_{h\to 0^+} \mathbf{P}(t+h) = \mathbf{P}(t) \lim_{h\to 0^+} \mathbf{P}(h) = \mathbf{P}(t)\mathbf{I} = \mathbf{P}(t)$$

If instead we approach t from the *left* we get

$$\lim_{h\to 0^+} \mathbf{P}(t-h) = \mathbf{P}(t) \lim_{h\to 0^+} \mathbf{P}(h) = \mathbf{P}(t)\mathbf{I} = \mathbf{P}(t)$$

By definition, this shows that $P(t)$ is continuous. We will show that the matrix $P(t)$ is also differentiable.

$$\lim_{h\to 0^+} \frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \lim_{h\to 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h}\mathbf{P}(t) = \mathbf{A}\mathbf{P}(t) \tag{20}$$

It remains to show that $\mathbf{A}$ exists. Let

$$\begin{aligned} q_i &= \lim_{h\to 0^+} \frac{1 - P_{ii}(h)}{h} \\ q_{ij} &= \lim_{h\to 0^+} \frac{P_{ij}(h)}{h} \end{aligned}$$

Note the relation

$$1 - P_{ii}(h) = \sum_{j=0, j\neq i}^{N} P_{ij}(h)$$

Dividing by h and letting $h \to 0$ we obtain

$$q_i = \sum_{j=0, j\neq i}^{N} q_{ij}$$

Taylor expanding $P_{ij}(0)$ and letting $h \downarrow 0$ in the Taylor expansion we obtain

$$\begin{aligned} \Pr\{X(h) = j \mid X(0) = i\} &= q_{ij}h + o(h) \\ \Pr\{X(h) = i \mid X(0) = i\} &= 1 - q_i h + o(h) \end{aligned}$$

These probabilites constitute the elements of the matrix $\lim_{h\to 0} \mathbf{P}(h)$. Subtracting $\mathbf{I}$ and dividing by h we obtain

$$\mathbf{A} = \begin{bmatrix} -q_0 & q_{01} & \cdots & q_{0N} \\ q_{10} & -q_1 & \cdots & q_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N0} & q_{N1} & \cdots & -q_N \end{bmatrix}$$

By (20) we have the differential equation

$$\mathbf{P}'(t) = \mathbf{A}\mathbf{P}(t)$$

with solution

$$\mathbf{P}(t) = e^{\mathbf{A}t} = \mathbf{I} + \sum_{n=1}^{\infty} \frac{\mathbf{A}^n t^n}{n!}$$

5 Queueing Systems

5.1 Queueing Processes

Queueing systems are classified according to the following terms

- **Input Process** The PDF of the pattern of customer arrivals in time
- **Service Distribution** The PDF of the random time to serve a customer
- **Queue Discipline** The number of servers and the order of customer service

5.1.1 The Queueing Formula (Little's Law)

Consider a queueing system that has been operating sufficiently long to have reached an appropriate steady state, or a position of statistical equilibrium. Let

- $L =$ the average number of customers in the system (average load)
- $\lambda =$ the rate of arrival of customers to the system
- $W =$ the average time spent by a customer in the system

Then *Little's law* states that

$$L = \lambda W \tag{21}$$

5.2 Poisson Arrivals, Exponential Service Times

Let $X(t)$ denote the number of customers in the system at time t. Let the *input process* be Poisson with rate parameter λ, and the *service distribution* be exponential with parameter μ. Series expanding both distributions as in (3.2.2) we obtain

$$\begin{aligned}\Pr\{X(t+h) - X(t) = 1\} &= [\lambda h + o(h)]\,[1 - \mu h + o(h)] \\ &= \lambda h + o(h)\end{aligned}$$

and similarly

$$\Pr\{X(t+h) - X(t) = -1\} = \mu h + o(h) \tag{22}$$

which by the postulates (4) shows that $X(t)$ is a BDP with birth rate $\lambda_k = \lambda$ and death rate $\mu_k = \mu$ for all $k \in \mathbb{N}^+$. From the definition of the process, we must have $\mu_0 = 0$.

5.2.1 Limiting Distribution

Since $X(t)$ is a BDP we find it's limiting distribution $\{\pi_k\}$ using the same method and notation as in section (4.5). Assuming that the geometric series

$$\theta_k = \left(\frac{\lambda}{\mu}\right)^k, \quad k \geq 1$$

converges as $k \to \infty$, we have

$$\lim_{k\to\infty} \theta_k = \frac{1}{1 - \lambda/\mu}$$

Hence

$$\pi_0 = \frac{1}{\sum_{i=0}^{\infty} \theta_k} = 1 - \lambda/\mu$$

and finally

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k, \quad k \in \mathbb{N}$$

5.2.2 Applying Little's law

Note that π_k has the same form as the PDF of the geometric distribution. If we let $p = 1 - \lambda/\mu$ then by the geometric distribution we have

$$L = \mathrm{E}[X] = \frac{1 - (1 - \lambda/\mu)}{1 - \lambda/\mu} = \frac{\lambda}{\mu - \lambda}$$

The ratio $\rho = \lambda/\mu$ is called the *traffic intensity*. Note that $L \to \infty$ as $\rho \to 1$.

5.2.3 Waiting Times

The *waiting time* of a customer is defined as the time between a customer arrives to the server and the time it becomes served. Let T be the waiting time of a given customer. Then T is exponentially distributed with parameter $(\mu - \lambda)$

$$\Pr\{T = t\} = (\mu - \lambda)\exp[-t(\mu - \lambda)] \tag{23}$$

Proof. It must be the case that T is equal to the sum of the exponentially distributed serving times for all the N people in front of him in the queue. Therefore T has a gamma distribution of order $N+1$ with scale parameter μ

$$\Pr\{T \leq t \mid N = n\} = \int_0^t \frac{\mu^{n+1}\tau^n e^{-\mu\tau}}{\Gamma(n+1)} d\tau \tag{24}$$

By the law of total probability we obtain

$$\Pr\{T \leq t\} = \sum_{n=0}^{\infty} \left[\Pr\{T \leq t \mid N = n\} \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right)\right] \tag{25}$$

Substituting (24) into (25) yields

$$\begin{aligned}
\Pr\{T \leq t\} &= \sum_{n=0}^{\infty} \int_0^t \frac{\mu^{n+1}\tau^n e^{-\mu\tau}}{\Gamma(n+1)} \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) d\tau \\
&= \int_0^t \mu e^{-\mu\tau} \left(1 - \frac{\lambda}{\mu}\right) \sum_{n=0}^{\infty} \frac{\tau^n \lambda^n}{\Gamma(n+1)} d\tau \\
&= \int_0^t \left(1 - \frac{\lambda}{\mu}\right) \mu \exp\left\{-\tau\mu\left(1 - \frac{\lambda}{\mu}\right)\right\} d\tau = 1 - \exp\left[-t(\mu - \lambda)\right]
\end{aligned}$$

which shows that T is exponentially distributed with parameter $(\mu - \lambda)$. □

6 Brownian Motion

6.1 Introduction

The *Brownian motion* stochastic process is a continuous-time and continuous-state process. In this introduction we shall let $X(t)$ denote the position of a particle in 1 dimension at time t. We let the initial position of the particle at time $t = 0$ be x_0. During a time period t, the particle wanders randomly to a position x. Intuitively, we would expect the particle to wander further from x_0 if we increase t. Let σ^2 be the variance of x given $t = 1$, and let $p(x, t \mid x_0, \sigma^2)$ be the PDF of the position x given an initial position x_0, time period t and variance σ^2. The PDF has the form

$$p(x, t \mid x_0, \sigma^2) = \frac{1}{\sqrt{2\pi t\sigma^2}} \exp\left(-\frac{1}{2t\sigma^2}(x - x_0)^2\right) \tag{26}$$

The PDF satisfies

$$p(x, t \mid x_0, \sigma^2) \geq 0, \qquad \int_{-\infty}^{\infty} p(x, t \mid x_0, \sigma^2) dx = 1, \quad \forall x \in \mathbb{R}$$

Additionally, we employ the postulate

$$\lim_{t \to 0} (x, t \mid x_0, \sigma^2) = 0, \quad x \neq x_0$$

Throughout most of this section we shall assume that $\sigma^2 = 1$. We introduce the notation

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

for the standard normal PDF ($\sigma^2 = t = 1$), and

$$\Phi(z) = \int_{-\infty}^{z} \phi(x) dx$$

for the corresponding CDF. If we allow t to vary, the PDF becomes

$$\phi_t(z) = \frac{1}{\sqrt{t}} \phi(z/\sqrt{t})$$

and the CDF becomes

$$\Phi_t(z) = \Phi(z/\sqrt{t})$$

Hence

$$\Pr\{X(t) \leq x \mid X(0) = x_0\} = \Phi\left(\frac{x - x_0}{\sqrt{t}}\right)$$

6.2 Properties

Brownian motion satisfies the following properties

1. Every increment $X(s+t) - X(s)$ is normally distributed with mean zero and variance $\sigma^2 t$
2. For every pair of disjoint time intervals $(t_1, t_2], (t_3, t_4]$, with $0 \leq t_1 < t_2 \leq t_3 < t_4$ the increments $X(t_4) - X(t_3)$ and $X(t_2) - X(t_1)$ are independent random variables. This applies for any number of disjoint time intervals.
3. $X(0) = 0$, and $X(t)$ is continuous as a function of t

6.3 Covariance Function

Using the independent increments property, we will determine the covariance of the process. Let $0 \leq s < t$. Since $\mathrm{E}[X(t)] = \mu(t) = 0, \forall t \geq 0$ we have

$$\begin{aligned}
\mathrm{Cov}[X(s), X(t)] &= \mathrm{E}[(X(s) - \mu(s))(X(t) - \mu(t))] = \mathrm{E}[X(s)X(t)] \\
&= \mathrm{E}[X(s)X(t) - X(s) + X(s)] \\
&= \mathrm{E}\left[X(s)^2\right] + \mathrm{E}[X(s)\{X(t) - X(s)\}] \\
&= \sigma^2 s + \mathrm{E}[X(s)]\,\mathrm{E}[X(t) - X(s)] \\
&= \sigma^2 s
\end{aligned}$$

For $0 \leq t < s$ we obtain $\mathrm{Cov}[X(s), X(t)] = \sigma^2 t$. The general formula is

$$\mathrm{Cov}[X(s), X(t)] = \sigma^2 \min\{s, t\}, \quad s, t \geq 0$$

The correlation is

$$\mathrm{Corr}[X(s), X(t)] = \frac{\mathrm{Cov}[X(s), X(t)]}{\sqrt{s}\sigma \cdot \sqrt{t}\sigma} = \frac{\min\{s, t\}}{\sqrt{st}} = \sqrt{\frac{\min\{s, t\}}{\max\{s, t\}}} \tag{27}$$

6.4 The Central Limit Theorem and The Invariance Principle

Let $S_n = \sum_{i=1}^{n} \xi_i$ be the sum of n independent and identically distributed random variables having zero means and unit variances. The *central theorem* asserts that

$$\lim_{n \to \infty} \Pr\left\{\frac{S_n}{\sqrt{n}} \leq x\right\} = \Phi(x), \quad \forall x \tag{28}$$

Consider a function $X_n(t)$ of the continuous variable t

$$X_n(t) = \frac{S_{[nt]}}{\sqrt{n}}, \quad t \geq 0$$

where $[x]$ is the greatest integer less than or equal to x. Observe that

$$X_n(t) = \frac{S_k}{\sqrt{n}} = \frac{\sqrt{[nt]}}{\sqrt{n}} \frac{S_k}{\sqrt{k}}, \quad \frac{k}{n} \leq t < \frac{k}{n} + \frac{1}{n}$$

Because $S_k/\sqrt{k}$ has unit variance, the variance of $X_n(t)$ is $\{nt\}/n$, which converges to t as $n \to \infty$. When n is large, then $k = [nt]$ is large, and $S_k/\sqrt{k}$ is approximately normally distributed according to (28). It's therefore reasonable to believe that $X_n(t)$ should behave much like a standard Brownian motion when n is large. This is indeed true, although we shall not prove it here. The convergence of a sum of a sequence of stochastic processes (6.4) to a standard Brownian motion is termed the *invariance principle*.

6.5 The Reflection Principle

We refer to the book for a deeper and more illustrative look into the *reflection principle*. Here, we shall only present some important results.
Let $X(t)$ be a Brownian motion process and define

$$M(t) = \max_{0 \leq u \leq t} X(u)$$

Then

$$\begin{aligned}
\Pr\{M(t) > y\} &= 2\Pr\{X(t) > y\} \\
&= 2[1 - \Phi_t(y)]
\end{aligned}$$

6.6 Hitting Times

The *hitting time* of a process is defined as

$$\tau_y = \min\{u \geq 0 : X(u) = y\}$$

Logically, we must have

$$\begin{aligned}\Pr\{\tau_y \leq t\} &= \Pr\{M(t) \geq y\} = 2\left[1 - \Phi_t(y)\right] \\ &= \frac{2}{\sqrt{2\pi t}} \int_y^{\infty} e^{-\xi^2/(2t)} d\xi\end{aligned}$$

The change of variable $\xi = \eta\sqrt{t}$ leads to

$$\Pr\{\tau_y \leq t\} = \sqrt{\frac{2}{\pi}} \int_{y/\sqrt{t}}^{\infty} e^{-\eta^2/2} d\eta$$

Differentiating with respect to t we obtain the PDF

$$f_{\tau_y}(t) = \frac{y t^{-3/2}}{\sqrt{2\pi}} e^{-y^2/(2t)}$$

7 Introduction to Gaussian Processes (GPs)

7.1 Definition

Let $X(t)$ be a continuous-time, continuous-state stochastic process with $X(0) = 0$ and $\mathrm{E}\left[X(t)\right] = \mu(t)$. We sample the process at some finite set of n time-points $\{t_i\}$ to get the corresponding states $\{X(t_i)\}$. Then we may construct a linear combination of the n states $\{X(t_i)\}$ by setting

$$S = \sum_{i=1}^{n} \alpha_i X(t_i)$$

If S is normally distributed for *any* set of scalar coefficients $\{\alpha_i\}$, then $X(t)$ is a *Gaussian process* (GP).

7.2 Univariate GP

Let $X(t)$ be a univariate process. By (7.1) S is normally distributed for *any* pair of equally sized sets $\{\alpha_i\}$ and $\{t_i\}$. Therefore, it must also hold true when the sets contain only a single element, that is, when we sample only a single time-point. It inevitably follows that $X(t)$ must be normally distributed for all $t > 0$ and we have

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{1}{2\sigma^2(t)}(x - \mu(t))^2\right), \quad x \in \mathbb{R}, t \geq 0$$

Note: We need not necessarily have $\mu(t) = 0$. Brownian motion is just a special type of GP where $\mu(t) = 0$ and $\sigma^2(t) = t\sigma^2$. Then, disjoint process increments are *independent* and we say that the process is *stationary.*

Throughout most of this section we let $\mu(t)$ and $\sigma^2(t)$ be constants, i.e. $\mu(t) = \mu$ (first order stationary) and $\sigma^2(t) = \sigma^2$ (second order stationary).

7.3 Multivariate GP

The multivariate Gaussian PDF for a random variable $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ viewed as an $n \times 1$ vector, is

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^n$$

The size $n \times 1$ *mean vector* is $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n), \mathrm{E}\left[x_i\right] = \mu_i$, and the *covariance matrix* is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{1,1} & \ldots & \Sigma_{1,n} \\ \ldots & \ldots & \ldots \\ \Sigma_{n,1} & \ldots & \Sigma_{n,n} \end{bmatrix}$$

For short, the PDF of $\boldsymbol{x}$ is denoted by

$$p(\boldsymbol{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The *correlation* between x_i and x_j is then

$$\text{Corr}\left[x_i, x_j\right] = \frac{\boldsymbol{\Sigma}_{i,j}}{\sigma_i \sigma_j}$$

If all $x_i, x_j \in \boldsymbol{x}$ are *independent*, then $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ are diagonal.

7.4 Linear Transformations

Let $\boldsymbol{x}$ be a $n \times 1$ random vector, $\boldsymbol{F}$ be a fixed $m \times n$ matrix, and $\boldsymbol{b}$ be a fixed $m \times 1$ vector. Then

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{b}$$

defines a linear transformation of $\boldsymbol{x}$, and has Gaussian PDF

$$p(\boldsymbol{y}) = N\left(\boldsymbol{F}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^T\right)$$

Hence, a linear transformation of a Gaussian variable is Gaussian itself. We will show that a Gaussian variable can be transformed to independent zero-mean and unit-variance variables $\boldsymbol{z} = (z_1, \ldots, z_n)$. Define $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$ where $\boldsymbol{L}$ is the lower triangular Cholesky factor of $\boldsymbol{\Sigma}$. Set $\boldsymbol{F} = \boldsymbol{L}^{-1}$ and $\boldsymbol{b} = -\boldsymbol{L}^{-1}\boldsymbol{\mu}$. Let

$$\boldsymbol{z} = \boldsymbol{y} = \boldsymbol{L}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \quad \boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{z} \tag{29}$$

7.5 Conditioning

We may split the random vector $\boldsymbol{x}$ in two blocks of variables $\boldsymbol{x}_A = (x_{A,1}, \ldots, x_{A,n_A})$ and $\boldsymbol{x}_B = (x_{B,1}, \ldots, x_{B,n_B})$ where $n_A + n_B = n$. Then

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_A, \boldsymbol{\mu}_B), \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{A,B} \\ \boldsymbol{\Sigma}_{B,A} & \boldsymbol{\Sigma}_B \end{bmatrix}$$

where $\boldsymbol{\Sigma}_{A,B} = \boldsymbol{\Sigma}_{B,A}^T$ is the size $n_A \times n_B$ cross-covariance matrix between $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. Assume the variables in $\boldsymbol{x}_B$ are known. Let $\boldsymbol{\mu}_{A|B} = \mathrm{E}\left[\boldsymbol{x}_A \mid \boldsymbol{x}_B\right]$, and let $\boldsymbol{\Sigma}_{A|B}$ be the covariance matrix of $\boldsymbol{x}_A$ conditional on $\boldsymbol{x}_B$. Then the conditional PDF of $\boldsymbol{x}_A$ is also Gaussian with

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) \tag{30a}$$

$$\boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{B,A} \tag{30b}$$

7.6 Sampling

In this section we provide two methods for generating random realizations from the multivariate Gaussian PDF. The first method uses sequential sampling

$$p(\boldsymbol{x}) = p(x_1)p(x_2 \mid x_1)\ldots p(x_n \mid x_{n-1}, \ldots x_1)$$

The second method uses (29)

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{z}, \quad p(\boldsymbol{z}) = N(\boldsymbol{0}_n, \boldsymbol{I}_n)$$

where $\boldsymbol{0}_n$ is a size $n \times 1$ vector of 0 entries and $\boldsymbol{I}_n$ is the size n identity matrix.

8 Stationary Gaussian Processes

8.1 Correlation Function

Let a stationary GP $X(t)$ be given. A *correlation function* describes the correlation between $X(t)$ and $X(s)$ as a function of the time-distance $d = |t - s|$. What follows is a list of some correlation functions

- **Exponential**

$$\rho(d) = \exp\left(-\phi_E d\right)$$

- **Matern Type**

$$\rho(d) = (1 + \phi_M d)\exp\left(-\phi_M d\right)$$

- **Squared Exponential**

$$\rho(d) = \exp\left(-\phi_G d^2\right)$$

In the special case of the exponential correlation function, the GP satisfies the Markov property.

8.2 Sampling

Consider a GP $X(t)$ with $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$. Let $\boldsymbol{t} = (t_1, t_2, \dots, t_n)^T$ be a $n \times 1$ vector of n time-points. Then $\boldsymbol{t}$ forms a *discretized grid* of the time-domain. One can simulate the corresponding $n \times 1$ vector $\boldsymbol{x} = (X(t_0), X(t_1), \dots, X(t_n))^T$ using the $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ which depends on the correlation function $\rho(d)$. We begin by computing $\boldsymbol{\Sigma}$. Assume a correlation function $\rho(d)$ is given. The covariance between $X(t_i)$ and $X(t_j)$ depends on the time-distance $d_{ij} = |t_i - t_j|$ for all $i, j \in \{0, 1, \dots, n\}$. Therefore, the covariance matrix $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{\Sigma}_{ij} = \sigma^2 \rho(d_{ij})$$

which may be computed by

$$\boldsymbol{\Sigma} = \sigma^2 \rho(\boldsymbol{H}), \quad \boldsymbol{H} = |\boldsymbol{t} \otimes \mathbf{1}_n - \mathbf{1}_n \otimes \boldsymbol{t}|, \quad \rho(\boldsymbol{H})_{ij} = \rho(\boldsymbol{H}_{ij}) \tag{31}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones and $\otimes$ denotes the *outer product*.

8.3 Conditional Process

Assume the process is known for $\boldsymbol{t}_B = (t_{B,1}, \dots, t_{B,n_B})^T$ with corresponding variables $\boldsymbol{x}_B = (x_{B,1}, \dots, x_{B,n_B})^T$, and that the process is unknown for the time-domain of interest $\boldsymbol{t}_A = (t_{A,1}, \dots, t_{A,n_A})^T$ with corresponding variables $\boldsymbol{x}_A = (x_{A,1}, \dots, x_{A,n_A})^T$. Let $\boldsymbol{\mu}_{A|B} = \mathrm{E}[\boldsymbol{x}_A | \boldsymbol{x}_B]$ and $\boldsymbol{\Sigma}_{A|B}$ be the conditional covariance matrix of $\boldsymbol{x}_A$ given $\boldsymbol{x}_B$. We want to compute $\boldsymbol{\mu}_{A|B}$ and $\boldsymbol{\Sigma}_{A|B}$.
The $n_A \times n_B$ covariance matrix $\boldsymbol{\Sigma}_{A,B}$ is

$$\boldsymbol{\Sigma}_{A,B} = \sigma^2 \rho(\boldsymbol{H}), \quad \boldsymbol{H} = |\boldsymbol{t}_A \otimes \mathbf{1}_{n_B} - \mathbf{1}_{n_A} \otimes \boldsymbol{t}_B| \tag{32}$$

We then compute $\boldsymbol{\mu}_{A|B}$ and $\boldsymbol{\Sigma}_{A|B}$ according to (30).