Quals Revision Notes Stochastics Sequence

Tyler Chen

1 Introduction

This document contains personal revision notes for the 2018 Stochastics Qualification Exam in the Applied Mathematics department at the University of Washington. These notes are heavily based on the class notes [?] and texts [?] used in the AMATH 561/562/563 courses.

For notational convenience any result which was determined to be worth memorizing was marked as a "Theorem" even if the result is not usually classified as a Theorem.

Contents

2 Probability Fundamentals

2.1 Events as sets

Definition. (Sample Space)

The set of all possible outcomes of an experiment is called the sample space and is generally denoted by Ω.

Definition. (Event)

An event is a subset of the sample space.

Definition. (σ-algebra)

A collection $\mathcal{F}$ of subsets of Ω is a σ-algebra if it satisfies,

1. contains the empty set: $\emptyset \in \mathcal{F}$
2. closed under unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then $\cup_i A_i \in \mathcal{F}$
3. closed under compliments: if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

Lemma.

If $\mathcal{F}$ and $\mathcal{G}$ are σ-algebras, then their intersection is also a σ-algebra. More generally, if $\{\mathcal{F}_i\}_{i \in I}$ is a family of σ-algebras, then $\mathcal{F} = \cap_{i \in I} \mathcal{F}_i$ is also σ-algebra.

Definition. (σ-algebra generated by a set)

Let $\mathcal{G} \in 2^{\Omega}$. The σ-algebra generated by $\mathcal{G}$, denoted $\sigma(\mathcal{G})$, is the smallest σ-algebra containing $\mathcal{G}$.

Theorem.

Given $\mathcal{G} \in 2^{\Omega}$, $\sigma(\mathcal{G})$ is the intersection of all σ-algebras containing $\mathcal{G}$.

Definition. (Measurable Space)

The pair $(\Omega, \mathcal{F})$ where Ω is a sample space and $\mathcal{F}$ is a σ-algebra of Ω is called a measurable space.

2.2 Probability

Definition. (Probability Measure)

A probability measure defined on a measurable space $(\Sigma, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ satisfying,

1. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$
2. If $A_1, A_2, \ldots$ is a collection of disjoint elemebts of $\mathcal{F}$, in that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Definition. (Probability Space)

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra of Ω, and $\mathbb{P}$ is a probability measure defined on $(\Omega, \mathcal{F})$ is called a measurable space.

Lemma.

Let $A_1, A_2, \ldots$ be a sequence of increasing sets, so that $A_1 \subseteq A_2 \subseteq \cdots$, and write A for their limit,

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i$$

Then $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$.

Proof.

Write $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots$, the union of disjoint sets. Then,

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A_1) + \sum_{i=1}^{\infty} \mathbb{P}(A_{i+1} \setminus A_i) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)] \\
&= \lim_{n\to\infty} \mathbb{P}(A_n)
\end{aligned}$$

□

Definition. (Conditional Probability)

If $\mathbb{P}(B) > 0$then the conditional probability that A occurs given that B occurs is defined to be,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

2.3 Infinite Probability Spaces

2.3.1 Uniform Lebesgue Measure on (0,1)

We construct a mathematical model for choosing a number at random on the open interval $(0, 1)$. Therefore, set $\Omega = (0, 1)$. Define,

$$\mathbb{P}(\{\omega : \omega \in (a, b)\}) = \mu((a, b)) := b - a, \qquad 0 < a \leq b < 1$$

where μ is the Lebesgue measure.

A natural question is what are all subsets of $(0, 1)$ whose probabilities are determined by our probability measure and the properties of probability measures.

This turns out not to be 2^{Ω}, which has some sets whose probabilities are not determined by this measure.

Note: WHY/HOW?

The correct set is the σ-algebra generated by the open intervals. That is,

$$\mathcal{B}((0, 1)) := \sigma(\mathcal{O}), \qquad \mathcal{O} := \{A \subseteq (0, 1) : A = (a, b), 0 \leq a < b \leq 1\}$$

We call $\mathcal{B}((0, 1))$ the Borel σ-algebra on $(0, 1)$.

Thus, the appropriate sample space for our experiment is $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}(0, 1), \mu)$.

Definition. (Borel σ-algebra)

Let Ω be some topological space and let $\mathcal{O}(\Omega)$ be the set of open sets in Ω. The Borel σ-algebra on Ω is defined as $\mathcal{B}(\Omega) := \sigma(\mathcal{O}(\Omega))$.

2.3.2 Infinite sequence of coin tosses

We now consider an infinite sequence of coin tosses. Therefore,

$$\Omega = \text{the set of infinte sequences of Hs and Ts}$$

Note that this set is uncountable (Cantor diagonalization argument).

We would like to construct a σ-algebra for this experiment.

First, consider the trivial σ-algebra,

$$\mathcal{F}_0 := \{\emptyset, \Omega\}$$

Given no information, we can tell if ω is in the sets in $\mathcal{F}_0$ because we know $\omega \in \Omega$ and $\omega \notin \emptyset$.

Now define,

$$A_{\mathrm{H}} = \{\omega \in \Omega : \omega_1 = \mathrm{H}\}, \qquad A_{\mathrm{T}} = \{\omega \in \Omega : \omega_1 = \mathrm{T}\}$$

Noting that $A_{\mathrm{H}} = A_{\mathrm{T}}^c$ we see that,

$$\mathcal{F}_1 := \{\emptyset, \Omega, A_{\mathrm{H}}, A_{\mathrm{T}}\}$$

is a σ-algebra. Given ω_1 it is possible to say whether or not ω is in each of the sets in $\mathcal{F}_1$.

Now define,

$$\begin{aligned} A_{\mathrm{HH}} &= \{\omega \in \Omega : \omega_1 = \mathrm{H}, \omega_2 = \mathrm{H}\}, & A_{\mathrm{TT}} &= \{\omega \in \Omega : \omega_1 = \mathrm{T}, \omega_2 = \mathrm{T}\} \\ A_{\mathrm{TT}} &= \{\omega \in \Omega : \omega_1 = \mathrm{T}, \omega_2 = \mathrm{T}\}, & A_{\mathrm{TH}} &= \{\omega \in \Omega : \omega_1 = \mathrm{T}, \omega_2 = \mathrm{H}\} \end{aligned}$$

The σ-algebra generated by these sets and the sets in $\mathcal{F}_1$ is,

$$\mathcal{F}_2 = \left\{ \begin{array}{l} \emptyset, \Omega, A_{\mathrm{H}}, A_{\mathrm{T}}, A_{\mathrm{HH}}, A_{\mathrm{HT}}, A_{\mathrm{TT}}, A_{\mathrm{TH}}, A_{\mathrm{HH}}^c, A_{\mathrm{HT}}^c, A_{\mathrm{TT}}^c, A_{\mathrm{TH}}^c \\ A_{\mathrm{HH}} \cup A_{\mathrm{TH}}, A_{\mathrm{HH}} \cup A_{\mathrm{TT}}, A_{\mathrm{HT}} \cup A_{\mathrm{TH}}, A_{\mathrm{HT}} \cup A_{\mathrm{TT}} \end{array} \right\}$$

Note that for instance, $A_{\mathrm{HH}} \cup A_{\mathrm{HT}} = A_{\mathrm{H}}$.

Now, given ω_1 and ω_2 we can say if ω belongs to each of the sets in $\mathcal{F}_2$. Continuing this way we can define σ-algebras $\mathcal{F}_n$ for all $n \in \mathbb{N}$. Finally, define,

$$\mathcal{F} := \sigma(\mathcal{F}_\infty), \qquad \mathcal{F}_\infty := \cup_n \mathcal{F}_n$$

Why not take $\mathcal{F} = \mathcal{F}_\infty$? This would tell us about what happens in finitely many coin tosses, but couldn't have information about sets such as "sequences for which 40% of the coin tosses are heads". However, these are in $\mathcal{F}$ as define.

We now construct a probability measure on $\mathcal{F}$. Supose there is a probability p of heads and $q = 1 - p$ of tails. Then,

$$\begin{array}{cccc} \mathbb{P}(\emptyset) = 0, & \mathbb{P}(\Omega) = 1, & \mathbb{P}(A_{\mathrm{H}}) = p, & \mathbb{P}(A_{\mathrm{T}}) = q, \\ \mathbb{P}(A_{\mathrm{HH}}) = p^2, & \mathbb{P}(A_{\mathrm{HT}}) = pq, & \mathbb{P}(A_{\mathrm{TH}}) = pq, & \mathbb{P}(A_{\mathrm{TT}}) = q^2, \ldots \end{array}$$

In this way we can define $\mathbb{P}(A)$ for each $A \in \mathcal{F}_\infty$. It turns out that this uniquely defines $\mathbb{P}$ for sets in $\mathcal{F}$.

Now define,

$$A = \left\{ \omega : \lim_{n\to\infty} \frac{\#\text{ of H in first } n \text{ coin tosses}}{n} = \frac{1}{2} \right\}$$

By the strong law of large numbers we have, $\mathbb{P}(A) = 1$ if $p = 1/2$ and $\mathbb{P}(A) = 0$ otherwise.

Note: Understand this rigorously

> **Definition.** ($\mathbb{P}$ almost surely)
>
> Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$ we say that the event A occurs $\mathbb{P}$ almost surely.

2.4 Random Variables and Distributions

A random variable maps the outcome of an experiment to $\mathbb{R}$.

> **Definition.** (Random Variable)
>
> A random variable X defined on $(\Omega, \mathcal{F})$ is a function $X : \Omega \to \mathbb{R}$ with the property that,
>
> $$\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}, \qquad \forall A \in \mathcal{B}(\mathbb{R})$$
>
> Such a function is said to be $\mathcal{F}$-measurable.

> **Definition.** (Distribution Function)
>
> The distribution function $F_X : \mathbb{R} \to [0, 1]$ of a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given by,
>
> $$F_X(x) = \mathbb{P}(X \leq x)$$

Note that while a random variable does not depend on $\mathbb{P}$, the distribution does.

> **Lemma.**
>
> A distribution function F_X has the following properties:
>
> 1. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$
> 2. if $x < y$ then $F(x) \leq F(y)$ (non-decreasing)
> 3. $F(x + h) \to F(x)$ as $h \to 0^+$ (right continuous)

Lemma.

Let F be the distribution function of X. Then,

1. $\mathbb{P}(X > x) = 1 - F(x)$
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$
3. $\mathbb{P}(X = x) = F(x) - \lim_{y \to x^-} F(y)$

Definition. (Discrete Random Variable)

A random variable X is called discrete if it takes values in some countable set $A = \{x_1, x_2, \ldots \subset \mathbb{R}$. We associate a discrete random variable with a probability mass function $f_X : A \to \mathbb{R}$ defined by $f_X(x_i) = \mathbb{P}(X = x_i)$.

Definition. (Continuous Random Variable)

A random variable X is called continuous if its distribution function F_X can be written as,

$$F_X(x) = \int_{-\infty}^{x} f_X(u) \mathrm{d}u$$

for some $f_X : \mathbb{R} \to [0, \infty)$, called the probability density function.

We can think of the density function f_X as $f_X(x)\mathrm{d}x = \mathbb{P}(X \in \mathrm{d}x)$

2.5 Stochastic Processes

We can intuitively think of a stochastic process as a process which evolves randomly in time.

Definition. (Stochastic Process)

A stochastic Processes is a collection of random variables $X = (X_t)_{t \in \mathbb{T}}$, where $\mathbb{T}$ is some index set. If the set $\mathbb{T}$ is countable we say X is a discrete time process. If the set $\mathbb{T}$ is uncountable we say X is a continuous time process.

The state space of a stochastic process process X is the union of the state spaces of $(X_t)_{t \in \mathbb{T}}$.

Two common ways to think of a stochastic process are 1. for any time $t \in \mathbb{T}$ we have that $X_t : \Omega \to \mathbb{R}$ is a random variable, and 2. for any $\omega \in \Omega$, $X(\omega) : \mathbb{T} \to \mathbb{R}$ is a function of time.

2.6 Expectation

Definition. (Expectation)

Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X, written $\mathbb{E}[X]$, is defined as,

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega)$$

2.6.1 Lebesgue Integration

Note: Probably will want to expand this to understand this better. Maybe even move to a new chapter

Definition. (Indicator Random Variable)

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $A \in \mathcal{F}$. The Indicator random variable is defined as,

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

For disjoint sets A and B we have

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B$$

and for any two sets A and B we have,

$$\mathbb{1}_{A\cap B} = \mathbb{1}_A \mathbb{1}_B$$

Definition. (Simple Random Variable)

A non-negative random variable X is called simple if it can be written as,

$$X(\omega) = \sum_{i=1}^n x_i \mathbb{1}_{A_i}(\omega), \qquad x_i \geq 0, \qquad A_i \in \mathcal{F}$$

where $(A_i)_{i=1}^n$ is a partition of Ω.

Definition. (Expectation of Simple Random Variable)

Let X be a simple random variable. We define the expectation of X as,

$$\mathbb{E}[X] := \sum_{i=1}^{n} x_i \mathbb{P}(A_i)$$

Note that this means $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

Now consider a random variable X which is not necessarily simple. Let $(X_n)_{n\geq 0}$ be ain increasing sequence of simple random variables that converges almost surely to X.

Note: how do we know such sequences exists....

Note: Is any of this actually relevant??

2.6.2 Computing Expectations

Definition. (Expectation)

The mean value, or expectation, or expected value of the random variable X with mass function f_X is defined to be,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x F_X(\mathrm{d}x) := \lim_{\|\Pi\|\to 0} \sum_i \left(\frac{x_{i+1} - x_i}{2}\right) (F_X(x_{i+1}) - F_X(x_i))$$

where Π is a partition of $\mathbb{R}$, meaning,

$$\Pi = \{x_1, x_2, \ldots, x_n\}, \qquad x_i < x_{i+1}, \qquad \|\Pi\| := \sup_i (x_{i+1} - x_i)$$

In the discrete or continuous this becomes,

$$\mathbb{E}[X] = \sum_i x_i f_X(x_i), \qquad \mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) \mathrm{d}x$$

Definition. (Expectation of Function)

The mean value, or expectation, or expected value of the function of a random variable $g(X)$ with mass function f_X is,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) F_X(\mathrm{d}x)$$

where Π is a partition of $\mathbb{R}$, meaning,

In the discrete or continuous this becomes,

$$\mathbb{E}[X] = \sum_i g(x_i) f_X(x_i), \qquad \mathbb{E}[X] = \int_{\mathbb{R}} g(x) f_X(x) \mathrm{d}x$$

2.7 Change of Measure

Theorem.

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $Z \geq 0$ be a random variable with $\mathbb{E}[Z] = 1$. Define $\tilde{\mathbb{P}} : \mathcal{F} \to [0, 1]$ by,

$$\tilde{\mathbb{P}}(A) := \mathbb{E}[Z \mathbb{1}_A]$$

Then $\tilde{\mathbb{P}}$ is a probability measure on $(\Omega, \mathcal{F})$. Denote the expectation taken with respect to $\tilde{\mathbb{P}}$ by $\tilde{\mathbb{E}}$. Then,

$$\tilde{\mathbb{E}}[X] = \mathbb{E}[ZX], \qquad \text{and if } Z > 0 \text{ then} \qquad \mathbb{E}[X] = \mathbb{E}\left[\frac{1}{Z} X\right]$$

where X is a random variable defined on $(\Omega, \mathcal{F})$.

Definition. (Radon–Nikodyn Derivative)

We call the random variable Z the Radon–Nikodyn Derivative of $\tilde{P}$ with respect to $\mathbb{P}$.

Definition. (Absolutely Continuous)

A probability measure $\mathbb{P}$ defined on $(\Omega, \mathcal{F})$ is absolutely continuous with respect to another probability measure $\tilde{\mathbb{P}}$, written $\mathbb{P} \gg \tilde{\mathbb{P}}$, if,

$$\mathbb{P}(A) = 0 \qquad \Longrightarrow \qquad \tilde{\mathbb{P}}(A) = 0$$

Definition. (Equivalent)

Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ defined on $(\Omega, \mathcal{F})$ are equivalent, written

$\mathbb{P} \sim \tilde{\mathbb{P}}$ if,

$$\mathbb{P}(A) = 0 \qquad \iff \qquad \tilde{\mathbb{P}}(A) = 0$$

Two probability measures are equivalent if and only if the Radon–Nikodyn derivative which relates them is strictly positive.

3 Information and Conditioning

3.1 Information and σ-algebras

Definition. (Filtration)

A filtration is an increasing sequence of σ-algebras.

Definition. (σ-algebra generated by X)

Let X be a random variable defined on a non-empty space Ω. The σ-algebra generated by X, denoted $\sigma(X)$, is the collection of subsets of Ω of the form $\{X \in A\}$ where $A \in \mathcal{B}(\mathbb{R})$.

Definition. ($\mathcal{G}$-measurable)

Let X be a random variable defined on a non-empty space Ω. Let $\mathcal{G}$ be a σ-algbera of subsets of Ω. If $\sigma(X) \subset \mathcal{G}$ then we say that X is $\mathcal{G}$-measurable, and write $X \in \mathcal{G}$.

A random variable is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X.

Definition. ($\mathbb{F}$-adapted)

Let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration and let $X = (X_t)_{t\in[0,T]}$ be a sequence of random variables. We say this collection of random variables is $\mathbb{F}$-adapted if $X_t \in \mathcal{F}_t$ for all $t \in [0, T]$.

3.2 Independence

Definition. (Independent Sets)

Two events A and B are independent if,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

More generally, a family $\{A_i\}_{i \in I}$ is called independent if,

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

for all finite subsets J of I.

Definition. (Independent σ-algebras)

Let $\{\mathcal{G}_i\}_{i=1}^n$ be a family of sub-σ-algebras of $\mathcal{F}$. We say these σ-algebras are independent if,

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i), \qquad \forall A_1 \in \mathcal{G}_1, \dots, \forall A_n \in \mathcal{G}_n$$

Definition. (Independent Random Variables)

Let $\{X_i\}_{i=1}^n$ be a family of random variables. We say $\{X_i\}_{i=1}^n$ are independent if $\{\sigma(X_i)\}_{i=1}^n$ are independent.

We say the random variable X is independent of the σ-algebra $\mathcal{G}$ if $\sigma(X) \perp\!\!\!\perp \mathcal{G}$.

Lemma.

If X and Y are independent random variables then,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Definition. (Joint Distribution)

The Joint Distribution $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ of two random variables X and Y is given by,

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

Variables can be jointly discrete or jointly continuous in natural generalization.

Theorem.

Let X and Y be random variables. The following are equivalent:

1. $X \perp\!\!\!\perp Y$
2. $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
3. $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ discrete: for every $(x,y) \in \mathbb{R}^2$, continuous: for almost every $(x,y) \in \mathbb{R}^2$
4. $\mathbb{E}[\exp(iuX + ivY) = \mathbb{E}[\exp(iuX)]\mathbb{E}[\exp(ivY)]$ for all $(u,v) \in \mathbb{R}^2$

Definition. (Variance)

The variance of a random variable X is defined as,

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X]^2 - (\mathbb{E}[X])^2$$

whenever this expectation exists.

Definition. (Covariance)

The covariance of random variables X and Y is defined as,

$$\mathrm{Co}\mathbb{V}[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

whenever this expectation exists.

Lemma.

For constant a and b,

$$\mathbb{V}[aX + bY] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] + 2ab\mathrm{Co}\mathbb{V}[X,Y]$$

3.3 Conditional Expectation

When (X,Y) are jointly discrete or jointly continuous we have conditional probability mass functions,

$$\text{discrete:} \qquad f_{X|Y}(x_i, y_j) := \mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i \cap Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)}$$

$$\text{continuous:} \qquad f_{X|Y}(x,y)\mathrm{d}x := \mathbb{P}(X \in \mathrm{d}x | Y = y) = \frac{\mathbb{P}(X =\in \mathrm{d}x \cap Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}\mathrm{d}x$$

From this we can defined the conditional expectation $\mathbb{E}[X|Y=y]$ as,

$$\text{discrete:} \qquad \mathbb{E}[X|Y=y_j] := \sum_i x_i f_{X|Y}(x_i, y_j)$$

$$\text{continuous:} \qquad \mathbb{E}[X|Y=y] := \int_{\mathbb{R}} x f_{X|Y}(x, y)\mathrm{d}x$$

Note that $\mathbb{E}[X|Y=y]$ is a function of y; there is nothing random about it.

Unfortunately lots of times (X, Y) will not be jointly discrete or jointly continuous. We need a more general definition of conditional expectation.

Definition. (Conditional Expectation)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algbera of $\mathcal{F}$, and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$ is any random variable satisfying,

1. Measurability: $\mathbb{E}[X|\mathcal{G}] \in \mathcal{G}$
2. Partial averaging: $\mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_A X]$ for all $A \in \mathcal{G}$. Alternatively, $\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[ZX]$ for all $Z \in \mathcal{G}$.

When $\mathcal{G} = \sigma(Y)$ we will often use the short hand notation $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$

Definition. (Conditional Probability)

The conditional probability of A given $\mathcal{G}$ is,

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$$

Theorem.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Conditional expecations satisfy,

1. Linearity: $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$.
2. Taking out what is known: if $X \in \mathcal{G}$ then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$.
3. Iteration Conditioning: if $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$
4. Independence: if $X \perp\!\!\!\perp \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$

Theorem. (Jensen's Inequality)

Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sum-σ-algebra of $\mathcal{F}$. Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function. Then,

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}], \qquad \mathbb{P}\text{-a.s}$$

Definition. (Martingale)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $T > 0$, and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration of sum-σ-algebras of $\mathcal{F}$. Consider an $\mathbb{F}$-adapted stochastic process $M = (M_t)_{t\in[0,T]}$. We say that M is,

$$\begin{aligned} &\text{a martingale if:} && \mathbb{E}[M_t|\mathcal{F}_s] = M_s, && \forall 0 \leq s \leq t \leq T, \\ &\text{a sub-martingale if:} && \mathbb{E}[M_t|\mathcal{F}_s] \geq M_s, && \forall 0 \leq s \leq t \leq T, \\ &\text{a super-martingale if:} && \mathbb{E}[M_t|\mathcal{F}_s] \leq M_s, && \forall 0 \leq s \leq t \leq T, \end{aligned}$$

Definition. (Markov)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $T > 0$, and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration of sum-σ-algebras of $\mathcal{F}$. Consider an $\mathbb{F}$-adapted stochastic process $X = (X_t)_{t\in[0,T]}$.

We say that X is Markov if for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g (which depends on s, t, and f) such that,

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = g(X_s).$$

Identifying $g(X_s) = \mathbb{E}[f(X_t)|X_s]$ we can write the markov property as follows:

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \mathbb{E}[f(X_t)|X_s]$$

Note: I don't understand this last bit

Definition. (Transition Kernel)

If X_t is a discrete or continuous random variabel for every t then we have transition kernel,

$$\begin{aligned} &\text{discrete:} && P(s, x; t, y) := \mathbb{P}(X_t = y|X_s = x) \\ &\text{continuous:} && \Gamma(s, x; t, y)\mathrm{d}y := \mathbb{P}(X_t \in \mathrm{d}y|X_s = x) \end{aligned}$$

If you can write down the transition kernel explicitly then you have essentially proved the process is Markov.

4 Generating and Characteristic Functions

4.1 Generating Functions

Definition. (Probability Generating Function)

Suppose X is a discrete random variable taking values in $\mathbb{Z}_{\geq 0}$. We defined the probability generating function of X by,

$$G_X(s) := \mathbb{E}[s^X] = \sum_k s^k f_X(k)$$

where $f_X(k)$ is the probability mass function of X.

Since $\sum_k f_X(k) = 1$ we know $G_X(s)$ exists when $|s| \leq 1$. Generally we only care about G_X and its derivatives at the point $s = 1$. Power series can be integrated and differentiated term by term within their radius of convergence.

We call G_X the probaility generating function because the coefficient $f_X(k)$ of the $\mathcal{O}(s^k)$ term is precisely $\mathbb{P}(X = k)$. So we can use G_X to obtain f_X.

Note that,

$$G'_X(1) = \mathbb{E}[X], \qquad G''_X(1) = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X]$$

Therefore,

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = G''_X(1) - (G'_X(1))^2 + G'_X(1)$$

Theorem.

Suppose $X \perp\!\!\!\perp Y$. Then $G_{X+Y}(s) = G_X(s)G_Y(s)$.

Theorem.

Suppose $\{X_i\}_{i\in\mathbb{N}}$ are iid with common distribution X. Define $S_n = \sum_{i=1}^n X_i$. Then,

$$G_{S_n}(s) = (G_X(s))^n$$

Proof.

We have,

$$G_{S_n}(s) = \mathbb{E}\left[s^{X_1+X_2+\cdots X_n}\right] = \left(\mathbb{E}[s^X]\right)^n = (G_X(s))^n$$

Theorem.

Suppose $\{X_i\}_{i\in\mathbb{N}}$ are iid with common distribution X. Define $S_n = \sum_{i=1}^n X_i$. Let N be independent of $\{X_i\}_{i\in\mathbb{N}}$. Then,

$$G_{S_n}(s) = G_N(G_X(s))$$

Proof.

We have,

$$\begin{aligned} G_{S_N}(s) &= \mathbb{E}\left[s^{X_1+X_2+\cdots X_N}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{X_1+X_2+\cdots+X_N}|N\right]\right] \\ &= \mathbb{E}\left[\left(\mathbb{E}\left[s^X\right]\right)^N\right] = \mathbb{E}\left[(G_X(s))^N\right] = G_N(G_X(s)) \end{aligned}$$

Definition. (Joint Probability Distribution)

Suppose X and Y are discrete random variables taking values in $\mathbb{Z}_{\geq 0}$. The joint probaiblity generating function of (X, Y) is defined as,

$$G_{X,Y}(s,t) = \mathbb{E}\left[s^X t^Y\right] = \sum_k \sum_m s^k t^m f_{X,Y}(k,m)$$

where $f_{X,Y}$ is the joint probability mass function of (X, Y).

4.2 Branching Processes

Suppose a population evolves in generations. Let Z_n be the size of the n-th generation, and assume $Z_0 = 1$. Let $X_{n,i}$ be the number of children of the i-th member of the n-th generation. Then, clearly, the number of individuals in the $(n+1)$-th generation is given by,

$$Z_{n+1} = X_{n,1} + X_{n,2} + \cdots + X_{n,Z_n}$$

Theorem.

Assume Z_{n+1} is as above and that the $\{X_{n,i}\}$ are iid. Define $G_n(s) := \mathbb{E}\left[s^{Z_n}\right]$ and $G(s) := \mathbb{E}\left[s^X\right]$. Then,

$$G_{n+m}(s) = G_n(G_m(s)) = G_m(G_n(s)), \quad \text{and thus} \quad \underbrace{G_n(s) = G(G(\cdots G(s)\cdots))}_{n\text{-fold iteration}}$$

Theorem.

Suppose $\mathbb{E}[Z_1] = \mu$ and $\mathbb{V}[Z_1] = \sigma^2$. Then,

$$\mathbb{E}[Z_n] = \mu^n, \qquad \mathbb{V}[Z_n] = \begin{cases} n\sigma^2 & \mu = 1 \\ \dfrac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \mu \neq 1 \end{cases}$$

4.3 Characteristic Functions

Definition. (Characteristic Function)

The characteristic function of a random variable X is the function $\phi_X : \mathbb{R} \to \mathbb{C}$ defined by,

$$\phi_X(t) := \mathbb{E}\left[\exp(itX)\right]$$

The characteristic function always exists since $\mathbb{E}\left[|\exp(itX)|\right] = 1$. Clearly, if G_X and ϕ_X both exist then,

$$\varphi_X(t) = G_X(\exp(it))$$

Theorem.

Characteristic functions have the following properties:

1. $\phi_{aX+b}(t) = \exp(ibt)\phi_X(at)$ for all constants a, b
2. if X and Y are independent, $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
3. if $\{X_i\}_{i=1}^n$ are independent and $S_n = \sum_{i=1}^n X_i$, then $\phi_{S_n}(t) = (\phi_X(t))^n$.

Theorem.

Let $\phi_X(t)$ be the characteristic function of a random variable X with $\mathbb{E}[|X|^n] <$

∞. Then,

$$\phi^{(n)}(0) = i^n \mathbb{E}[X^n]$$

Proof.

We have,

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}\phi_X(t) = \mathbb{E}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}\exp(itX)\right] = i^n \mathbb{E}\left[X^n \exp(itX)\right]$$

Now take $t = 0$ to obtain the theorem. $\square$

Theorem. (Inversion)

Suppose X is a continuous random variable with density f_X. Then,

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-itx)\phi_X(t)\mathrm{d}t$$

for all x where f_X is differentiable.

Definition. (Convergence in Distribution)

We say a sequence of distribution functions $\{F_n\}_{n\in\mathbb{N}}$ converges to a distribution function F if $\lim_{n\to\infty} F_n(x) = F(x)$ for all points where F is continuous.

Theorem. (Continuity)

Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions with corresponding characteristic functions $\{\phi_n\}_{n\in\mathbb{N}}$.

1. If $F_n \to F$, where F is a distribution function with corresponding characteristic function ϕ, then $\phi_n \to \phi$ pointwise.
2. Conversely, if $\phi(t) = \lim_{n\to\infty}\phi_n(t)$ exists and is continuous at time $t = 0$, then ϕ is the characteristic function of some distribution F, and $F_n \to F$.

4.4 LLN and CLT

Definition. (Convergence in Distribution of Random Variables)

We say a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges in distribution to a random variable X, written $X_n \xrightarrow{\mathcal{D}} X$, if $F_{X_n} \to F_X$.

Theorem. (Law of Large Numbers)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of iid random variables with $\mathbb{E}[X_n] = \mu$. Define $S_n = \frac{1}{n}\sum_{i=1} nX_i$. Then $S_n \xrightarrow{\mathcal{D}} \mu$

Theorem. (Central Limit Theorem)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of iid random variables with $\mathbb{E}[X_n] = \mu$ and $\mathbb{V}[X_n] = \sigma^2$. Define,

$$S_n = \frac{1}{n}\sum_{i=1} nX_i, \qquad U_n = \frac{S_n - \mu n}{\sqrt{n\sigma}}$$

Then $U_n \xrightarrow{\mathcal{D}} Z$, where $Z \sim \mathcal{N}(0,1)$.

4.5 Large Deviations Principle

Definition. (Cumulant Generating Function)

The cumulant generating function Λ_X of X is defined as,

$$\Lambda_X(t) = \log(M_X(t)), \qquad M_X(t) = \mathbb{E}[\exp(tX)]$$

We note that $\Lambda'_X(0) = mu$ as,

$$\Lambda'_X(0) = \frac{M'_X(0)}{M_X(0)} = \left.\frac{\mathbb{E}[X\exp(tX)]}{\mathbb{E}[\exp(tX)]}\right|_{t=0} = \mathbb{E}[X] = \mu$$

We also note that Λ_X is convex as,

$$\Lambda''_X(t) = \frac{M_X(t)M''_X(t) - (M'_X(t))^2}{M_x^2(t)} = \frac{\mathbb{E}[\exp(tX)]\mathbb{E}[X^2\exp(tX) - \mathbb{E}[X\exp(tX)]^2}{M_X^2(t)} \geq 0$$

where we have used the Cauchy-Schwarz inequality,

$$\mathbb{E}[YZ]^2 \leq \mathbb{E}[Y^2]\mathbb{E}[Z^2], \qquad \text{with} \qquad Y = X\exp(tX/2), \qquad Z = \exp(tX/2)$$

Definition. (Fenchel-Legendre Transform)

We define the Fenchel-Legendre transform Λ_X^* for Λ_X as,

$$\Lambda_X^* = \sup_{t \in \mathbb{R}} \{at - \Lambda_X(t)\}, \qquad a \in \mathbb{R}$$

Theorem. (Large Deviation Principle)

Let $\{X_i\}$ be a sequence of iid random variables with common distribution X. Define $\mu := \mathbb{E}[X]$ and suppose the moment generating function $M_X(t) := \mathbb{E}[\exp(tX)]$ is finite in some neighborhood of $t = 0$. Let Γ_X and Γ_X^* be defined as above. Suppose $a > \mu$ and $\mathbb{P}(X > a) > 0$. Then $\Gamma_X^*(0) > 0$ and,

$$\lim_{n \to \infty} \frac{1}{n} \log(\mathbb{P}(S_n > na)) = -\Lambda_X^*(a), \qquad S_n = \sum_{i=1}^{n} X_i$$

5 Discrete Time Markov Chains

A Markov Chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

5.1 Overview

Definition. (Discrete Time Markov Chain)

A discrete time Markov chain $X = (X_n)_{n \in \mathbb{N}}$ is a discrete time Markov process with countable state space S.

Equivalently,

$$\mathbb{P}(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and all $s, x_1, x_2, \ldots, x_{n-1} \in S$.

Definition. (Homogeneous DTMC)

A DTMC X is called homogeneous if,

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

for all n, j, i.

We assume Markov chains are homogeneous unless otherwise stated.

Definition. (One-Step Transition Matrix)

The one-step transitoin matrix of a DTMC X is the $|S| \times |S|$ matrix defined as,

$$p(i,j) = \mathbb{P}(X_1 = j | X_0 = i)$$

Theorem.

The transition matrix P is a stochastic matrix (all entries are non-negative, and rows sum to one).

Definition. (n-Step Transition Matrix)

The n-step transitoin matrix of a DTMC X is the $|S| \times |S|$ matrix defined as,

$$p_n(i,j) = \mathbb{P}(X_1 = j | X_0 = i)$$

Theorem. (Chapman-Kolmogorov Equation)

Let P and P_n be the one-step and n-step tranition matrices of a homogeneous DTMC. Then,

$$P_{m+n} = P_m P_n, \qquad\qquad P_n = P^n$$

Proof.

For any $i, j \in S$ we have,

$$\begin{aligned} p_{m+n}(i,j) &= \mathbb{P}(X_{m+n} = j | X_0 = i) \\ &= \sum_k \mathbb{P}(X_{m+n} = j | X_m = k)\mathbb{P}(X_m = k | X_0 = i) \\ &= \sum_k \mathbb{P}(X_n = j | X_0 = k)\mathbb{P}(X_m = k | X_0 = i) \\ &= \sum_k p_n(k,j) p_m(i,k) \end{aligned}$$

This proves $P_{m+n} = P_m P_n$. Then clealry,

$$P_n = PP_{n-1} = P^2 P_{n-2} = \ldots = P^n \qquad \square$$

Lemma.

Let X be a homogeneous DTMC. Denote the probability mass function of X_n by,

$$\mu_n(i) := \mathbb{P}(X_n = i), \qquad \mu_n = (\mu_n(1), \mu_n(2), \ldots, \mu_n(|S|))$$

Then $\mu_{n+m} = \mu_n P_m$ and thus $\mu_m = \mu_0 p^m$

5.2 Classification of States

Definition. (Persistent/Transient)

Let X be a DTMC. We say the state i is persistent or recurrent if,

$$\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$$

Otherwise we say the state i is transient.

Definition. (First Passage Time)

Let X be a DTMC. We define the first passage to state j by,

$$\tau_j := \inf\{n \geq 1 : X_n = j\}$$

and denote the probability mass function of τ_j given $X_0 = i$ by f_{ij}. That is,

$$f_{ij} = \mathbb{P}(\tau_j = n | X_0 = i)$$

Theorem.

Define the following generating functions

$$P_{ij} := \sum_{=0}^{\infty} s^n p_n(i,j), \quad p_0(i,j) = \delta_{i,j}, \quad F_{ij}(S) := \sum_{n=0}^{\infty} s^n f_{ij}(n), \quad f_{i,j}(0) = 0$$

Then, for any $|s| < 1$ we have,

1. $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$
2. $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ if $i \neq j$

Theorem.

1. State j is persistent if $\sum_n p_n(j,j) = \infty$, and if this holds then $\sum_n p_n(i,j) = \infty$ for all i such that $f_{ij} > 0$.

2. State j is transient if $\sum_n p_n(j,j) < \infty$, and if this holds then $\sum_n p_n(i,j) < \infty$ for all i.

Theorem.

If j is transient then $p_n(i,j) \to 0$ as $n \to \infty$ for all i.

Definition. (Mean Recurrence Time)

The mean recurrence time of a state j is defined as,

$$\hat{\tau}_j = \mathbb{E}[\tau_j | X_0 = j]$$

Clearly $\overline{\tau}_j = \infty$ if state j is transient, however $\overline{\tau}_j$ may be infinite even if j is recurrent.

Definition.

A recurrent state j is said to be null if $\overline{\tau}_j = \infty$, and non-nullor positive if $\overline{\tau}_j < \infty$.

Theorem.

A persistent state j is null if and only if $p - n(i,j) \to 0$ as $n \to \infty$; If this holds then $p_n(i,j) \to 0$ as $n \to \infty$ for all i.

Definition. (Period)

The period of a state i is defined as $d(i) = \gcd\{n : p_n(i,j) > 0\}$. If $d(i) = 1$ we say that the state i is aperiodic.

Definition. (Ergodic)

A state is called ergodic if it is persistent, non-null, and aperiodic.

5.3 Classification of Chains

Definition. (Communicates)

We say that state i communicates with state j, written $i \to j$, if $p_n(i,j) > 0$ for some $n \geq 1$. We say states i and j intercommunicate, written $i \leftrightarrow j$ if $i \to j$ and $j \to i$.

Theorem.

Suppose $i \leftrightarrow j$. Then,

1. i is transient if and only if j is transient.
2. i is null persistent if and only if j is null persistent.
3. i and j have the same period ($d(i) = d(j)$).

Definition. (Closed)

A set C of states is called closed if $p(i,j) = 0$ for all $i \in C$ and $j \notin C$.

Definition. (Irreducible)

A set C of states is called irreducible if $i \leftrightarrow j$ for all $i, j \in C$.

Theorem. (Markov Chain Decomposition)

The state space S of a Markov chain can be uniquely partitioned as,

$$S = T \cup C_1 \cup C_2 \cup \dots$$

where T is the set of transient states, and $C_1, C_2, \dots$ are closed sets of persistent states.

Proof.

Let $C_1, C_2, \dots$ be the persistent equivalence classes of $\leftrightarrow$. We need only show that each C_r us closed. Suppose on the contrary that there exists $i \in C_r$ and $j \notin C_r$ with $p(i,j) > 0$. Now, $j \not\to i$ so,

$$\mathbb{P}(X_n \neq i \text{ for all } n \geq 1 | X_0 = 1) \geq \mathbb{P}(X_1 = j | X_0 = i) = p(i,j) > 0$$

This contradicts the assumption that i is persistent. □

Lemma.

If S is finite, then at least one state is persistent and all persistent states are non-null.

5.4 Stationary Distributions and the Limit Theorem

Definition. (Stationary Distribution)

Let X be a Markov chain with one-step transition matrix P. We say that a row vector $\pi = (\pi(1), \pi(2), \ldots, \pi(|S|))$ is a stationary or invariant distribution if,

$$\pi P = \pi, \qquad \sum_i \pi(i) = 1, \qquad \text{and} \qquad \pi(i) \geq 0$$

Theorem.

An irreducible chain X has a stationary distribution π if and only if all states are non-null persistent. In this case π is unique and given by,

$$\pi(i) = 1/\overline{\tau}_i$$

5.5 Reversibility

Definition. (Time Reversal)

Let $X = (X_n)_{0 \leq n \leq N}$ be a Markov chain. The time reversal of X is the process $Y = (X_{N-n})_{0 \leq n \leq N}$.

Theorem.

Let $X = (X_n)_{0 \leq n \leq N}$ be an irreducible Markov chain with one step transition matrix P and invariant distribution π. Suppose $X_0 = \pi$ (so that $\mu_n = \pi$ for every n). Then Y, the time reversal of X, is a Markov chain and its one-step transition matrix $Q = (q(i,j))_{ij}$ is given by,

$$q(i,j) = \frac{\pi(i)}{\pi(j)} p(j,i)$$

Definition. (Reversible Markov Chain)

Let $X = (X_n)_{n\geq 0}$ be an irreducible Markov chain with one-step transition matrix P and invariant distribution π. Suppose $X_0 = \pi$ (so that $\mu_n = \pi$ for every n). Let Y be the time reversal of X and denote by Q the one-step transition matrix of Y. We say that X is reversible if $P = Q$, or equivalently if,

$$\pi(i)p(i,j) = \pi(j)p(j,i)$$

Definition. (Detailed Balance)

Let P be an $|S| \times |S|$ one-step transition matrix. We say that a distribution $\lambda = (\lambda(1), \lambda(2), \ldots, \lambda(|S|))$ is in detailed balance with P if for all $i, j \in S$,

$$\lambda(i)p(i,j) = \lambda(j)p(j,i)$$

Theorem.

Let P be the one-step transition matrix of an irreducible Markov chain X. Suppose there exists a distribution λ such that $\lambda(i)p(i,j) = \lambda(j)p(j,i)$ for all $i, j \in S$. Then λ is a stationary distribution for X.

5.6 Chains with Finitely Many States

Theorem. (Perron-Frobenius)

If P is a one-step transition matrix of a finite irreducible Markov chain X with period d, then,

1. $\lambda_1 = 1$ is an eigenvalue of P
2. The d complex roots of unity,

$$\lambda_n = \omega^{n-1}, \qquad \omega = e^{2\pi i/d}, \qquad n = 1, 2, \ldots, d$$

   are eigenvalues of P,
3. The remaining eigenvalues $\lambda_{d+1}, \lambda_{d+2}, \ldots, \lambda_{|S|}$ all have modulus less than one.

6 Continuous Time Markov Chains

6.1 The Poisson Process

Definition. (Counting Process)

A counting process is a stochastic process $N = (N_t)_{t\geq 0}$ taking values in $S = \{0, 1, 2, \ldots\}$ such that:

1. $N_0 = 0$
2. if $s < t$ then $N_s \leq N_t$

A counting process counts the number of times an event occurs propr to a given time.

Definition. (Poisson Process)

A Poisson process with intensity λ is a stochastic process $N = (N_t)_{t\geq 0}$ taking values in $S = \{0, 1, 2, \ldots\}$ such that:

1. $N_0 = 0$
2. if $s < t$ then $N_s \leq N_t$
3. if $s < t$ then $(N_t - N_s) \perp\!\!\!\perp N_s$
4. As $s \to 0^+$,

$$\mathbb{P}(N_{t+s} = n + m | N_t = m) = \begin{cases} \lambda s + \mathcal{O}(s^2) & m = 1 \\ \mathcal{O}(s^2) & m \geq 1 \\ 1 - \lambda s + \mathcal{O}(s^2) & m = 0 \end{cases}$$

Lemma.

A Poisson process N is Markov.

Proof.

Let $\mathcal{F}_t := \sigma(N_s : 0 \leq s \leq t)$. Then,

$$\mathbb{E}[g(N_t)|\mathcal{F}_s] = \mathbb{E}[g(N_t - N_s + N_s)|N_s] = \sum_{n=0}^{\infty} g(n + N_s) f_{N_t - N_s}(n)$$

Theorem.

Let $N = (N_t)_{t\geq 0}$ be a Poisson process with parameter λ. Then, for all $t \geq 0$ we have $N_t \sim \text{Poi}(\lambda t)$. That is,

$$p_t(j) := \mathbb{P}(N_t = j) = \frac{(\lambda t)^j}{j!} \exp(-\lambda t)$$

Proof.

p54 of pdf understand this.

Definition. (Arrival Time)

Let $N = (N_t)_{t\geq 0}$ be a counting process. We define S_n, the n-th arrival time by,

$$S_0 = 0, \qquad S_n = \inf\{t \geq 0 : N_t = n\}, \qquad n \geq 1$$

with $\inf \emptyset = \infty$.

Definition. (Inter-Arrival Time)

Let $N = (N_t)_{t\geq 0}$ be a counting process. We define τ_n, the n-th inter-arrival time by,

$$\tau_n = S_n - S_{n-1}, \qquad n \geq 1$$

Theorem.

Suppose the inter-arrival times τ_i of a counting process $N = (N_t)_{t\geq 0}$ are iid and exponentially distributed with parameter λ. Then N is a Poisson process with parameter λ.

Proof.

also may be worth understanding

6.2 Overview of Continuous Time Markov Chains

Definition. (Continuous Time Markov Chain (CTMC))

A continuous time Markov chain $X = (X_t)_{t\geq 0}$ is a Markov process witha countable state space S.

Definition. (Homogeneous)

A CTMC X is called homogeneous if,

$$\mathbb{P}(X_{t+s} = j | X_s = i) = \mathbb{P}(X_t = j | X_0 = i) =: p_t(i, j), \qquad \forall t, s > 0, \forall i, j \in S$$

Definition. (Transition Semigroup)

Let P_t be the $|S| \times |S|$ matrix $P_t = (p_t(i,j))$. We call the collection of matrices $(P_t)_{t\geq 0}$ the transition semigroup.

Theorem.

The transition semigroup satisfies the follow:

1. $P_0 = I$
2. $\sum_j p_t(i,j) = 1$
3. $P_t P_s = P_{t+s}$

Definition. (Generator of CTMC)

A CTMC with generator $G = (g(i,j))$ is a stochastic process satisfying,

1. if $s < t$ then $(X_t - X_s) \perp\!\!\!\perp X_s$
2. As $s \to 0$ we have $\mathbb{P}(X_{t+s} = j | X_t = i) = \delta_{i,j} + g(i,j)s + \mathcal{O}(s^2)$

More compactly,

$$G = \lim_{s \to 0^+} \frac{1}{s}(P_s - I)$$

Theorem. (Kolmogorov Forward and Backward Equations)

Let $X = (X_t)_{t\geq 0}$ be a CTMC with generator G. The transition semigroup of X satisfies the following ODEs:

$$\text{Kolmogorov Forward Equation:} \qquad \frac{\mathrm{d}}{\mathrm{d}t} P_t = P_t G,$$
$$\text{Kolmogorov Backward Equation:} \qquad \frac{\mathrm{d}}{\mathrm{d}t} P_t = G P_t$$

Theorem.

Let $X = (X_t)_{t\geq 0}$ be a CTMC with generator G and let P_t be the semigroup generated by G. Then the invariant distribution π satisfies,

$$\pi = \pi P_t \qquad \iff \qquad \pi G = 0$$

7 Brownian Motion

7.1 Scaled Random Walk

Define,

$$M_0 = 0, \qquad M_k = \sum_{i=0}^{k} X_i, \qquad X_i = \begin{cases} +1 & \omega_i = \mathrm{H} \\ -1 & \omega_i = \mathrm{T} \end{cases}$$

Definition. (Brownian Motion)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A Brownian motion is a stochastic process $W = (W_t)_{t \geq 0}$ satisfying,

1. $W_0 = 0$.
2. if $0 \leq r < s < t < u < \infty$ then $(W_u - W_t) \perp\!\!\!\perp (W_s - W_r)$.
3. if $0 \leq r < s < \infty$ then $(W_s - W_r) \sim \mathcal{N}(0, r - s)$.
4. the map $t \to W_t(\omega)$ is continuous for every ω.

Definition. (Filtration for a Brownian Motion)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a Brownian motion $W = (W_t)_{t \geq 0}$ is defined. A filtration for the Brownian motion W is a collection of σ-algebras satisfying,

1. information accumulates: if $0 \leq s < t < \infty$ then $\mathcal{F}_s \subseteq \mathcal{F}_t$.
2. adaptivity: for all $t \geq 0$, $W_t \in \mathcal{F}_t$.
3. independence of future increments: if $0 \leq t < u < \infty$, $W_u - W_t) \perp\!\!\!\perp \mathcal{F}_t$

Definition. (Natural Filtration for Brownian Motion)

The natural filtration for a Brownian motion $W = (W_t)_{t \geq 0}$ is defined as, $\mathcal{F}_t = \sigma(W_u, 0 \leq u \leq t)$.

In theory a filtration could contain more information that by observing the Brownian motion, however information in the filtration is not allowed to destroy the independence of future increments.

If $\mathcal{F}_t$ is a filtration for W, then W is a martingale with respect to this filtration as,

$$\mathbb{E}[W_T | \mathcal{F}_t] = \mathbb{E}[W_T - W_t + W_t) | \mathcal{F}_t] = \mathbb{E}[W_T - W_t | \mathcal{F}_t] + \mathbb{E}[W_t | \mathcal{F}_t] = 0 + W_t = W_t$$

Definition. (Quadratic Variation)

Let $f : [0, T] \to \mathbb{R}$. The quadratic variation of f up to time T is defined as,

$$[f, f]_T := \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2$$

where $\Pi = \{t_i : 0 = t_0 < t_1 < \cdots < t_n = T\}$ and $\|\Pi\| = \max_i(t_{i+1} - t_i)$.

Lemma.

Suppose $f : [0, T] \to \mathbb{R}$ has a continuous first derivative. Then $[f, f]_T = 0$.

Theorem.

Let $W = (W_t)_{t \geq 0}$ be a Brownian motion. Then for all $T \geq 0$, $[W, W]_T = T$ almost surely.

Definition. (Covariation)

Let $f, g : [0, T] \to \mathbb{R}$. The covariation of f and g up to time T is defined as,

$$[f, g]_T := \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)] [g(t_{j+1}) - g(t_j)]$$

where $\Pi = \{t_i : 0 = t_0 < t_1 < \cdots < t_n = T\}$ and $\|\Pi\| = \max_i(t_{i+1} - t_i)$.

Theorem.

Let $W = (W_t)_{t \geq 0}$ be a Brownian motion. Then for all $T \geq 0$, $[W, \mathrm{Id}]_T = 0$ almost surely.

This gives us the Heuristics,

$$\mathrm{d}W_t \mathrm{d}W_t = \mathrm{d}t, \qquad \mathrm{d}W_t \mathrm{d}t = 0, \qquad \mathrm{d}t \mathrm{d}t = 0$$

7.2 Markov Property of Brownian Motion

Theorem.

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration for W. Then W is a Markov process.

Proof.

Let $0 \leq t \leq T$ and let f be a non-negative Borel measurable function. Then, since $W_t \in \mathcal{F}_t$ and $(W_T - W_t) \perp\!\!\!\perp \mathcal{F}_t$

$$\mathbb{E}[f(W_T)|\mathcal{F}_t] = \mathbb{E}[f(W_T - W_t + W_t)|\mathcal{F}_t] = \int_{\mathbb{R}} f(y + W_t)\Gamma(t, W_t; T, y)\mathrm{d}y$$

where $\Gamma(t, W_t; T, \cdot)$ is the density of a normal random variable with mean W_t and variance $T - t$. Thus, setting $g(w) = \int_{\mathbb{R}} f(y + w)\Gamma(t, w; T, y)\mathrm{d}y$ we have,

$$\mathbb{E}[f(W_t)|\mathcal{F}_t] = g(W_t)$$

□

7.3 First Hitting Time of Brownian Motion

Theorem.

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration for W. Define a process $Z = (Z_t)_{t\geq 0}$ by,

$$Z_t = \exp\left(-\frac{1}{2}\sigma^2 t + \sigma W_t\right)$$

Then Z is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$.

Proof.

Let $T \geq t$. Then,

$$\begin{aligned}
\mathbb{E}[Z_T|\mathcal{F}_t] &= \mathbb{E}\left[\exp\left(-\frac{1}{2}\sigma^2 T + \sigma W_T\right)\middle|\mathcal{F}_t\right] \\
&= \exp\left(-\frac{1}{2}\sigma^2 T + \sigma W_t\right)\mathbb{E}\left[\exp\left(\sigma(W_T - W_t)\right)\middle|\mathcal{F}_t\right] \\
&= \exp\left(-\frac{1}{2}\sigma^2 T + \sigma W_t\right)\exp\left(\frac{1}{2}\sigma^2(T - t)\right) \\
&= Z_t
\end{aligned}$$

Note: good to know $\mathbb{E}[\exp(X)] = \exp(\mu + \sigma^2/2)$ when $X \sim \mathcal{N}(0, \sigma^2)$

Definition. (First Hitting Time)

The first hitting time of a Brownian motion W to level m is defined as,

$$\tau_m := \inf\{t \geq 0 : W_t = m\}$$

Theorem.

For any $m \in \mathbb{R}$ and $\alpha > 0$ we have,

$$\mathbb{E}[\exp(-\alpha\tau_m)] = \exp(-|m|\sqrt{2\alpha})$$

7.4 Reflection Principle

For every trajectory of a Brownian motion that hits level m prior to time t and finishes at level $W_t = w \leq m$, there is an equally likely path that finishes at a level $W_t = 2m - w$. Then,

$$\mathbb{P}(\tau_m \leq t, W_t \leq w) = \mathbb{P}(W_t \geq 2m - w), \qquad m > 0, \qquad w \leq m$$

Theorem.

For all $m \neq 0$, the first hitting time of a brownian motion has density f_{τ_m} give by,

$$f_{\tau_m}(t) = \mathbb{1}_{t\geq 0}\frac{|m|}{t\sqrt{2\pi t}}\exp(-m^2/2t)$$

Theorem.

For any $t > 0$ the joint density of a Brownian motion W_t and its running maximum $\overline{W}_t$ is,

$$f_{W_t,\overline{W}_t}(w, m) = \frac{2(2m-w)}{t\sqrt{2\pi t}}\exp(-(2m-w)^2/2t), \qquad m > 0, \qquad w \leq m$$

8 Stochastic Calculus

8.1 Itô Integrals

Definition. (Simple Process)

The process $\Delta = (\Delta_t)_{t\geq 0}$ si called simple if it is of the form,

$$\Delta_t = \sum_{j=0}^{n-1} \Delta_{t_j} \mathbb{1}_{t_j \leq t \leq t_{j+1}}, \qquad 0 = t_0 < t_1 < \cdots < t_n = T, \qquad \Delta_{t_j} \in \mathcal{F}_{t_j}$$

Definition. (Integral of Simple Process)

Given a simple process Δ define,

$$I_T = \int_0^T \Delta_t \mathrm{d}W_t := \sum_{j=0}^{n-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j})$$

Theorem.

The process $I = (I_t)_{t\geq 0}$ is a martingale with respect to $\mathcal{F}$.

Theorem.

The process $I = (I_t)_{t\geq 0}$ satisfies,

$$\mathbb{V}[I_T] = \mathbb{E}[I_T^2] = \mathbb{E}\left[\int_0^T \Delta_t^2 \mathrm{d}t\right]$$

Theorem.

The process $I = (I_t)_{t\geq 0}$ satisfies,

$$[I, I]_T^2 = \mathbb{E}\left[\int_0^T \Delta_t^2 \mathrm{d}t\right]$$

8.1.1 Itô Integrals for General Integrands

We generalize our integral by defining the integral of a $\mathbb{F}$-adapted process Δ which is square integrable as the limit of the integrals of a sequence converging to Δ in the sense that,

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left(\Delta_t - \Delta_t^{(n)}\right)^2 \mathrm{d}t\right] = 0$$

Note: How have we defined the integral of the square???

Theorem.

Let W be a Brownian motion and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration for W. Let $\Delta = (\Delta_t)_{0\leq t\leq T}$ be adapted to the filtration $\mathbb{F}$. Let $I = (I_t)_{0\leq t\leq T}$ be give by $I_t = \int_0^t \Delta_s \mathrm{d}s$. Then the process I has the following properties,

1. The sample paths of I are continuous.
2. The process I is $\mathbb{F}$-adapted. That is, $I_t \in \mathcal{F}_t$ for all t.
3. If $\Gamma = (\Gamma_t)_{0\leq t\leq T}$ satisfies the same conditions as Δ, then,

$$\int_0^T (a\Delta_t + b\Gamma_t)\mathrm{d}W_t = a\int_0^T \Delta_t \mathrm{d}W_t + b\int_0^T \Gamma_t \mathrm{d}W_t$$

4. The process I is a martingale with respect to $\mathbb{F}$.
5. The Itô isometry $\mathbb{E}[I_t^2]] = \mathbb{E}\left[\int_0^T \Delta_t^2 \mathrm{d}t\right]$ holds.
6. The quadratic variation of I is give by $[I, I]_T = \int_0^T \Delta_t^2 \mathrm{d}t$.

8.2 Itô Formula

Since W_t is not differentiable we no longer have the chain rule,

$$\mathrm{d}f(g(t)) = f'(g(t))g'(t)\mathrm{d}t$$

Theorem.

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and suppose $f : \mathbb{R} \to \mathbb{R}$ satisfies $f \in C^2(\mathbb{R})$. Then, for any $T \geq 0$,

$$f(W_T) - f(W_0) = \int_0^T f'(W_t)\mathrm{d}W_t + \frac{1}{2}\int_0^T f''(W_t)\mathrm{d}t$$

We can use this to compute $\int_0^T W_t \mathrm{d}t$. By the above we have,

$$W_T^2 - W_0^2 = \int_0^T 2W_t \mathrm{d}t + \int_0^T \mathrm{d}t$$

where we have chosen $f(x) = x^2$. Thus,

$$\int_0^T W_t \mathrm{d}t = \frac{1}{2}W_T^2 - \frac{1}{2}T$$

Definition. (Itô Process)

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration for W. An Itô process is any process $X = (X_t)_{t\geq 0}$ for the form,

$$X_t = X_0 + \int_0^t \Theta_s \mathrm{d}s + \int_0^t \Delta_s \mathrm{d}W_s$$

where $\Theta = (\Theta_t)_{t\geq 0}$ and $\Delta = (\Delta_t)_{t\geq 0}$ are adapted tot he filtration $\mathbb{F}$ and satisfy,

$$\int_0^T |\Theta_t| \mathrm{d}t < \infty, \qquad \int_0^T \Delta_t^2 \mathrm{d}t < \infty, \qquad \forall T \geq 0$$

and X_0 is not random.

Definition. (Differential Form of Itô Process)

We can write a process in its differential form,

$$\mathrm{d}X_t \Theta_t \mathrm{d}t + \Delta_t \mathrm{d}W_t$$

This expression literally means that X satisifes the itnegral form.

Lemma.

The quadratic variation $[X, X]_T$ of an Itô process is give by,

$$[X, X]_T = \int_0^T \Delta_t^2 \mathrm{d}t$$

Definition. (Integral with respect to Itô Process)

Let $X = (X_t)_{t\geq 0}$ be an Itô process and let $\Gamma = (\Gamma_t)_{t\geq 0}$ be adapted tot he filtration of the Brownian motion $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. We define,

$$\int_0^T \Gamma_t \mathrm{d}X_t := \int_0^T \Gamma_t \Theta_t \mathrm{d}t + \int_0^T \Gamma_t \Delta_t \mathrm{d}W_t$$

where we assume,

$$\int_0^T |\Gamma_t \Theta_t| \mathrm{d}t < \infty, \qquad \int_0^T (\Gamma_t \Delta_t)^2 < \infty, \qquad \forall T \geq 0$$

Theorem. (Itô Formula)

Let $X = (X_t)_{t\geq 0}$ be an Itô process and suppose $f : \mathbb{R} \to \mathbb{R}$ satisfies $f \in C^2(\mathbb{R})$. Then, for any $T \geq 0$,

$$f(X_T) - f(X_0) = \int_0^T f'(X_t)\mathrm{d}X_t + \frac{1}{2}\int_0^T f''(X_t)\mathrm{d}[X,X]_t$$

Lemma.

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion. Suppose $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a deterministic function. Then,

$$I_T := \int_0^T g(t)\mathrm{d}W_t \sim \mathcal{N}(0, v(T)), \qquad v(T) = \int_0^T g^2(t)\mathrm{d}t$$

8.3 Multivaraite Stochastic Calculus

Definition. (Multidimensional Brownian Motion)

A d-dimensional Brownian motion is a process, $W = (W_t^1, W_t^2, \ldots, W_t^d)_{t\geq 0}$ satisfying,

1. Each W_t^i is a one dimensional Brownian motion
2. The processes $\{W_t^i\}_{i=1}^d$ are independent

Definition. (Filtration for Multidimensional Brownian Motion)

A filtration for W is a collection of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ such that,

1. information accumulates: $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $0 \leq s < t$
2. adaptivity: $W \in \mathcal{F}_t$ for all $t \geq 0$
3. independent increments: $(W_t - W_s) \perp\!\!\!\perp \mathcal{F}_t$ for all $0 \leq s < t$

Theorem.

Let $W = (W_t^1, W_t^2, \ldots, W_t^d)_{t\geq 0}$ be a d-dimensional Brownian motion. The covariation of independent components of W is zero. That is, $[W^i, W^j]_T = 0$ for all $T \geq 0$.

Theorem.

Let $W = (W_t^1, W_t^2, \dots, W_t^d)_{t\geq 0}$ be a d-dimensional Brownian motion and let $X^i = (X_t^i)_{t\geq 0}$, $i = 1, 2, \dots, n$ be the Itô procsses given by,

$$\mathrm{d}X_t^i = \Theta_t^i \mathrm{d}t + \sum_{j=1}^{d} \sigma_t^{ij} \mathrm{d}W_t^j, \qquad i = 1, 2, \dots, n$$

Then,

$$\mathrm{d}[X^i, X^j] = \sum_{k=1}^{d} \sigma_t^{ik} \sigma_t^{jk} \mathrm{d}t$$

Theorem. (Multidimensional Itô formula)

Let $X = (X_t^1, X_t^2, \dots, X_t^n)_{t\geq 0}$ be an n-dimensional Itô process and suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $f \in C^2(\mathbb{R}^n)$. Then, for any $T \geq 0$,

$$\mathrm{d}f(X_t) = \sum_{i=1}^{n} \frac{\partial f(X_t)}{\partial x_i} \mathrm{d}X_t^i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j-1}^{n} \frac{\partial^2 f(X_t)}{\partial x_i \partial x_j} \mathrm{d}[X^i, X^j]_t$$

Lemma. (Product Rule)

Let X and Y be two one-dimensional Itô processes. Then,

$$\mathrm{d}(X_t Y_t) = Y_t \mathrm{d}X_t + X_t \mathrm{d}Y_t + \mathrm{d}[X, Y]_t$$

Theorem. (Lévy Characterization of Brownian Motion)

Let $M = (M_t)_{t\geq 0}$ be a martingale with respect to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. Suppose M has continuous sample paths and satisfies $M_0 = 0$ and $[M, M]_t = t$ for all $t \geq 0$. Then M is a Brownian motion.

8.4 Brownian Bridge

Definition. (Gaussian Process)

A Gaussian process $X = (X_t)_{t\geq 0}$ is any process that, for arbitarty times $t_1 < t_2 < \cdots < t_n$, the joint distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ is jointly normal.

Definition. (Brownian Bridge (Version 1))

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and fix $T > 0$. We define a Brownian bridge from a to b on $[0,T]$, $X^{a\to b} = (X_t^{a\to b})_{t\in[0,T]}$, by,

$$X_t^{a\to b} = a + \frac{t}{T}(b-a) + W_t - \frac{t}{T}W_T, \qquad t \in [0,T]$$

Definition. (Brownian Bridge (Version 2))

Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and fix $T > 0$. We defined a Brownian bridge from a to b on $[0,T]$, $Y^{a\to b} = (Y_t^{a\to b})_{t\in[0,T]}$, by,

$$Y_t^{a\to b} = a + \frac{t}{T}(b-a) + (T-t)\int_0^t \frac{1}{T-s}\mathrm{d}W_s, \qquad t \in [0,T]$$

Note: OTHER DEFINITIONS AND THEOREMS

8.5 Girsanov's Theorem

Definition. (Radon–Nikodym Derivative Process)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathbb{F} = (cF_t)_{t\in[0,T]}$ be a filtration on this space. A Radon–Nikodym derivative process$(Z_t)_{t\in[0,T]}$ is any process of the form,

$$Z_t := \mathbb{E}[Z|\mathcal{F}_t]$$

where Z is a random variable satisfying $Z > 0$ and $\mathbb{E}[Z] = 1$.

Lemma.

A Radon–Nikodym derivative process is a martingale.

Proof.

For $0 < s < t < T$ we have,

$$\mathbb{E}[Z_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[Z|\mathcal{F}_s] = Z_s \qquad \square$$

Lemma.

Let $(Z_t)_{t\in[0,T]}$ be a Radon-Nikodym derivative process and define $\mathrm{d}\tilde{\mathbb{P}}/\mathrm{d}\mathbb{P} = Z$. Suppose $Y \in \mathcal{F}_s$ where $s \in [0,T]$. Then,

$$\tilde{\mathbb{E}}[Y] = \mathbb{E}[Z_s Y]$$

Proof.

We have,

$$\tilde{\mathbb{E}}[Y] = \mathbb{E}[ZY] = \mathbb{E}[Y\mathbb{E}[Z|\mathcal{F}_s]] = \mathbb{E}[Z_s Y] \qquad \square$$

Note: I guess this makes sense that it shouldn't matter what s is since Z_t is a margingale

Lemma.

Let $(Z_t)_{t\in[0,T]}$ be a Radon–Nikodym derivative process and define $\mathrm{d}\tilde{\mathbb{P}}/\mathrm{d}\mathbb{P} = Z$. Suppose $Y \in \mathcal{F}_t$ where $t \in [0,T]$ and let $s \in [0,t]$. Then,

$$\tilde{\mathbb{E}}[Y|\mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}[Z_t Y|\mathcal{F}_s]$$

Theorem. (Girsanov)

Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration for W. Suppose $\Theta = (\Theta_t)_{t\in[0,T]}$ is adapted to the filtration $\mathbb{F}$. Define $(Z_t)_{t\in[0,T]}$ and $\tilde{W} = (\tilde{W}_t)_{t\in[0,T]}$ by,

$$Z_t = \exp\left(-\int_0^t \frac{1}{2}\Theta_s^2 \mathrm{d}s - \int_0^t \Theta_s \mathrm{d}W_s\right), \quad \mathrm{d}\tilde{W}_t = \Theta_t \mathrm{d}t + \mathrm{d}W_t, \quad \tilde{W}_0 = 0$$

Assume that,

$$\mathbb{E}\left[\int_0^T \Theta_t^2 Z_t^2 \mathrm{d}t\right] < \infty$$

Define a Radon-Nikodym derivative $Z = \mathrm{d}\tilde{\mathbb{P}}/\mathrm{d}\mathbb{P} = Z_T$. Then the process $\tilde{W}$ is a Brownian motion under $\tilde{P}$.

Theorem. (Martingale Representation)

Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration generated by W. Let M be a martingale with respect to $\mathbb{F}$. Then there exists an $\mathbb{F}$-adapted process $\Gamma = (\Gamma_t)_{t\in[0,T]}$ such that,

$$M_t = M_0 + \int_0^t \Gamma_s \mathrm{d}W_s, \qquad t \in [0, T]$$

9 SDEs and PDEs

9.1 Stochastic Differential Equations

Definition. (Stochastic Differential Equation)

A stochastic differential equation (SDE) is an equation of the form,

$$\mathrm{d}X_s = \mu(s, X_s)\mathrm{d}s + \sigma(s, X_s)\mathrm{d}W_s, \qquad X_t = x$$

where we refer to the functions μ and σ as the drift and diffusion respectively, and call $X_t = x$ the initial condition.

Definition. (Strong Solution)

A strong solution of an SDE is a stochastic process $X = (X_s)_{s\geq t}$ such that,

$$X_T = x + \int_t^T \mu(s, X_s)\mathrm{d}s + \int_t^T \sigma(s, X_s)\mathrm{d}W_s, \qquad \forall T \geq t$$

A strong solution means that given a sample path of the Brownian motion, we can construct a unique sample path of the solution.

Theorem.

Consider the SDE,

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

Suppose μ and σ satisfy,

$$|\mu(t, x)| + |\sigma(t, x)| < C_1(1 + |x|)$$

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) + \sigma(t,y)| < C_2|x-y|$$

for all t, x, y and some constant C_1 and C_2. Then the SDE has a unique solution which is adapted to the filtration generated by W and satisfies $\mathbb{E}\int_0^T X_t^2 \mathrm{d}t < \infty$ for all $T < \infty$.

Theorem. (Markov Property of Solutions of SDEs)

Let $X = (X_t)_{t\geq 0}$ be the solution of an SDE of the form,

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

Then X is a Markov process. That is, for $t \leq T$ and for some suitable function φ, there exists a function g (dependent on t, T, and φ) for which,

$$\mathbb{E}[\varphi(X_T)|\mathcal{F}_t] = g(X_t)$$

where $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is any filtration to which X is adapted.

9.2 Connection to Partial Differential Equations

Theorem.

Let $X = (X_t)_{t\geq 0}$ be the solution of an SDE of the form,

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

For some suitible function φ, define,

$$u(t, X_t) := \mathbb{E}[\varphi(X_T)|\mathcal{F}_t]$$

If $u \in C^{1,2}$ then it satisfies the Kolmogorov Backward Equation,

$$(\partial_t \mathcal{A}(t))u(t,\cdot) = 0, \qquad u(T,\cdot) = \varphi, \qquad \mathcal{A}(t) = \mu(t,x)\partial_x + \frac{1}{2}\sigma^2(t,x)\partial_x^2$$

9.2.1 Killing a Diffusion

Note: IDK why this example was relevant

9.3 Kolmogorov Forward and Backward Equations

Definition. (Transition Density)

The transition density of a process X is,

$$\Gamma(t,x;T,y)\mathrm{d}y = \mathbb{P}(X_T \in \mathrm{d}y | X_t = x)$$

Definition. (Two Parameter Semigroup)

The two parameter semigroup $(\mathcal{P}_t)_{t\in[0,T]}$ of a Markov diffusion X is defined as,

$$\mathcal{P}(t,T)\varphi(x) = \mathbb{E}[\varphi(X_T)|X_t = x] = \int \Gamma(t,x;T,y)\varphi(y)\mathrm{d}y$$

where φ is integrable with respect to the transition density Γ.

Theorem.

The semigroup satisifes,

$$\mathcal{P}(t,t) = I, \qquad \mathcal{P}(t,s)\mathcal{P}(s,T) = \mathcal{P}(t,T), \qquad t \in [0,T], s \in [t,T]$$

Definition. (Infinitesimal Generator)

The infinitesimal generator of a semigroup of operators $(\mathcal{P}(t,s))_{t\in[0,T]}$ is defined as,

$$\begin{aligned}
\mathcal{A}(t)\varphi(x) &= \lim_{s\to t^+} \frac{1}{s-t}\left(\mathcal{P}(t,s)\varphi(x) - \varphi(x)\right) \\
&= \lim_{s\to t^+} \frac{1}{s-t}\left(\mathbb{E}[\varphi(X_s)|X_t = x] - \varphi(x)\right)
\end{aligned}$$

where φ is any function for which this limit exists.

Theorem.

If $\varphi \in C_0^2$ (bounded and twice differentiable) and X is the solution of,

$$\mathrm{d}X_t = \mu(t,X_t)\mathrm{d}t + \sigma(t,X_t)\mathrm{d}W_t$$

then the generator $\mathcal{A}(t)$ of the semigroup $(P(s,t))_{t\in[0,T]}$ of X is given by,

$$\mathcal{A}(t) = \mu(t,X_t)\partial_x + \frac{1}{2}\sigma^2(t,X_t)\partial_x^2$$

Theorem. (KBE and KFE)

Let $X = (X_t)_{t\geq 0}$ be the unique solution (assumed to exist) of,

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

Let Γ denote the transition density of X,

$$\Gamma(t, x; T, y)\mathrm{d}y = \mathbb{P}(X_T \in \mathrm{d}y | X_t = x)$$

Then, seen as a function of the backwards variables (t, x), the transition density $\Gamma(\cdot, \cdot, ; T, y)$ satisfies the KBE,

$$(\partial_t + \mathcal{A}(t))\Gamma(\cdot, t; T, y) = 0, \qquad \Gamma(T, \cdots; T, y) = \delta_y$$

where $\mathcal{A}(t)$ is the infinitesimal generator of X,

$$\mathcal{A}(t) = \mu(t, x)\partial_x + \frac{1}{2}\sigma^2(t, x)\partial_x^2$$

Seen as a function of the forward variables (T, y), the transition density $\Gamma(t, x; \cdot, \cdot)$ satisfies the KFE,

$$(-\partial_T + \mathcal{A}^*(T))\Gamma(t, x; T, \cdot) = 0, \qquad \Gamma(t, x; t, \cdot) = \delta_x$$

where $\mathcal{A}^*(T)$ is the $L^2(\mathrm{d}x)$ adjoint of $\mathcal{A}(T)$,

$$\mathcal{A}^*(T) = -\partial_y \mu(T, y) + \frac{1}{2}\partial_y^2 \sigma(T, y)$$

Theorem.

Let $X = (X_t)_{t\geq 0}$ be defined on some interval $I = (l, r)$ and $\tau = \inf\{t \geq 0 : X_t \notin I\}$. Suppose X satisfies,

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$$

Define,

$$u(x) := \mathbb{E}[\exp(-\lambda(\tau - t))\varphi(X_\tau) + \int_t^\tau \exp(-\lambda(s - t))g(X_s)\mathrm{d}s | X_T = x], \quad t \leq \tau$$

Then the function u satisfies,

$$\begin{aligned}(\mathcal{A}-\lambda)u+g&=0 && \text{in } I\\ u&=\varphi && \text{on } \partial I\end{aligned}$$

where $\mathcal{A}=\mu(X)\partial_x+\frac{1}{2}\sigma^2(x)\partial_x^2$ is the infinitesimal generator of X.

Lemma.

The Laplace transform of τ, given by,

$$u(x):=\mathbb{E}[\exp(-\lambda t)|X_0=x]$$

satisfies,

$$\begin{aligned}(\mathcal{A}-\lambda)u&=0 && \text{in } I\\ u&=1 && \text{on } \partial I\end{aligned}$$

9.4 Scalar Time-Homogeneous Diffusions

Consider a diffusion $X=(X_t)_{t\geq 0}$ that lives on some interval I with endpoints l and j, where $-\infty\leq l<r\leq\infty$. Suppose X satisfies,

$$\mathrm{d}X_t=\mu(X_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}W_t$$

The generator $\mathcal{A}$ of X is given by,

$$\mathcal{A}=\mu(x)\partial_x+\frac{1}{2}\sigma^2(x)\partial_x^2$$

We can write this in the form,

$$\mathcal{A}=\frac{1}{m(x)}\partial_x\left(\frac{1}{s(x)}\partial_x\right)$$

where s and m are the speed and scale densities, and are given by,

$$s(x)=\exp\left(-\inf\frac{2\mu(x)}{\sigma^2(x)}\mathrm{d}x\right),\qquad m(x)=\frac{2}{\sigma^2(x)}\exp\left(\int\frac{2\mu(x)}{\sigma^2(x)}\mathrm{d}x\right)$$

If $m(x)$ is normalizable then m is a time-homogeneous solution of the KBE. Thus, m is a stationary density for X.

If we define a scale function $S(x)=\int s(x)\mathrm{d}x$, then the process $S(X)$ is a martingale.

9.4.1 Boundary Classification and Boundary Conditions

Definition. (Scale Measure)

Let X be a time-homogeneous scalar diffusion on the interval $I = (l, r)$. Let $s(x)$ be the scale density of X. For $x, y \in (l, r)$ define,

$$S[x, y] = \int_x^y s(u)\mathrm{d}u, \qquad S(l, y] = \lim_{x \to l^+} S[x, y], \qquad S[x, r) = \lim_{y \to r^-} S[x, y]$$

Define further,

$$\begin{aligned} I_l &= \int_l^r S(l, u]\mathrm{d}u, & I_r &= \int_x^r S[u, r)\mathrm{d}u \\ J_l &= \int_l^x S[u, x]\mathrm{d}u, & J_r &= \int_x^r S[x, u]\mathrm{d}u \end{aligned}$$

Where x is any point in (l, r). Note that for a given endpoint, whether I and J are finite does not depend on x.

Definition. (Classification of boundaries)

Let X be a time-homogeneous scalar diffusion on the interval (l, r). An endpoint l or r is said to be,

	$I < \infty$ (attainable)	$I = \infty$ (unattainable)
$J < \infty$	regular	entrance
$J = \infty$	exit	natural

The process X can be started from a regular boundary and can reach a regular boundary in a finite time

The process X cannot be started from an exit boundary but can reach an exit boundary in finite time. If X reaches an exit boundary, it does not return.

The process X can be started from an entrance boundary but cannot reach an entrance boundary in finite time.

The process X cannot be started from a natural boundary nor can it reach a natural boundary in finite time.

10 Jump Diffusions

10.1 Basic Definitions and Results on Lévy Processes

Definition. (Lévy Process)

A d-dimensional stochastic process $\eta = (\eta_t)_{t \geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a Lévy process if it satisfies,

1. $\eta_0 = 0$.
2. if $0 \leq r < s < t < u < \infty$ then $(\eta_u - \eta_t) \perp\!\!\!\perp (\eta_s - \eta_r)$.
3. if $0 \leq r < s < \infty$ then $\eta_s - \eta_r \sim \eta_{s-r}$.
4. for any $\epsilon > 0$ and $t \geq 0$, $\lim_{s \to 0^+} \mathbb{P}(|\eta_{t+s} - \eta_t| > \epsilon) = 0$.

Note that the fourth condition only means that the probability of a jump occurring at a fixed time is zero.

We assume Lévy processes are right continuous with left limits.

Definition. (Filtration for Lévy Process)

A filtration for the Lévy process η is a collection of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying,

1. information accumulates: if $0 \leq s < t$ then $\mathcal{F}_s \subseteq \mathcal{F}_t$.

2. adaptivity: for all $t \geq 0$, $\eta_t \in \mathcal{F}_t$.

3. independence of future increments: if $u > t \geq 0$ then $(\eta_u - \eta_t) \perp\!\!\!\perp \mathcal{F}_t$.

Definition. (Jump of Lévy Process)

The jump of a Lévy process η at time t is defined as,

$$\Delta \eta_t = \eta_t - \eta_{t^-}, \qquad \eta_{t^-} = \lim_{s \to t^-} \eta_s$$

Definition. (Poisson Random Measure)

We define the Poisson random measure of η, $N : \mathbb{R}_+ \times \mathcal{B}_0^d \times \Omega \to \mathbb{Z}_{\geq 0}$ by,

$$N(t, U, \omega) = \sum_{s \in [0,t]} \mathbb{1}_{\Delta \eta_s(\omega) \in U}, \qquad U \in \mathcal{B}_0^d = \{B \in \mathcal{B}(\mathbb{R}^d) : \{0\} \notin B\}$$

The Poisson random measure counts the number of jumps of sizes in U which occur prior to time t.

The differential form, $N(\mathrm{d}t, \mathrm{d}z)$ counts the number of jumps of size in $[z, z + \mathrm{d}z]$ over the time interval $\mathrm{d}t$.

Note: It is z+dz right?

Definition. (Lévy Measure)

Let N be the Poisson random measure of a Lévy Process η. Define the Lévy measure of η, $\nu : \mathcal{B}_0^d \to \mathbb{R}_+$ by,

$$\nu(U) = \mathbb{E}[N(1, U)], \qquad\qquad U \in \mathcal{B}_0^d$$

Theorem.

Let $U \in \mathcal{B}_0^d$. Then the process $(N(t, U))_{t\geq 0}$ is a Poisson process with intensity $\nu(U)$.

We see that $\nu(U)$ is the expected rate at which η has a jump of size $z \in U$.

A pure jump L evy process has a finite Levy measure $\nu(\mathbb{R} \setminus \{0\}) < \infty$ if and only if it can be represented by a compound Poisson process. In this case we can express η in as either,

$$\eta_t = \int_{\mathbb{R}^d \setminus \{0\}} zN(t, \mathrm{d}z) \qquad \text{or} \qquad \eta_t = \sum_{t=1}^{P_t} X_n$$

However, there are some Lévy processes for which $\nu(\mathbb{R} \setminus \{0\}) = \infty$. In this case neither of the prior representations makes sense.

Note: WHY THO??

Definition. (Compensated Poisson Random Measure)

Let N be a Poisson random measure with associated Lévy measure ν. The compensated Poisson random measure, denote $\tilde{N}$ is defined as,

$$\tilde{N}(t, U) = N(t, U) - \nu(U)t$$

11 Dynamic Mode Decomposition

We would like to study some dynamical system,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t), \qquad\qquad x(t) \in \mathbb{R}^n$$

In discrete time we have a flow map,

$$x_{k+1} = F(x_k)$$

We would like to approximate $x_t = f(x,t)$ by $u_t = \mathcal{A}x$. Given such an approximation we have,

$$x_{k+1} = Ax_k, \qquad\qquad A = \exp(\mathcal{A}\Delta t)$$

DMD finds A such that,

$$\|x_{k+1} - Ax_k\|$$

is minimized over $k = 1, 2, \ldots, m-1$.

We can write this as,

$$\min_A \|X' - AX\|_F$$

where,

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_{m-1} \\ | & | & & | \end{bmatrix}, \qquad X' = \begin{bmatrix} | & | & & | \\ x_2 & x_3 & \cdots & x_m \\ | & | & & | \end{bmatrix}$$

The solution to this minimization problem is,

$$A = X'X^\dagger$$

We are interested in the eigenvalues and eigenvectors of A. Suppose $A\Phi = \Phi\Lambda$.

Then,

$$x(t) \approx \Phi \exp(\Omega t)\Phi^\dagger x(0), \qquad\qquad \omega_j = \ln(\lambda_j)/\Delta t$$

More usefully,

$$x_{k+1} = Ax_k$$

Since A may be very large, we would like to take advantage of low rank structure of X in order to find eigenmodes for A.

11.1 DMD via SVD

Let the (reduced) SVD of X be given by,

$$X = \hat{U}\hat{\Sigma}\hat{V}^*, \qquad \hat{U}, \hat{V} \in \mathbb{R}^{n\times r}, \hat{\Sigma} \in \mathbb{R}^{r\times r}$$

Then,

$$A = X'X^\dagger = X'\hat{V}\hat{\Sigma}^{-1}\hat{U}^*$$

Define,

$$\hat{S} := \hat{U}^* A \hat{U} = \hat{U}^* X' \hat{V}\hat{\Sigma}^{-1}$$

Now note that $\hat{S} \in \mathbb{R}^{r\times r}$ so that it may be much easier to compute the eigenvalues of $\hat{S}$ than of A.

Moreover, A has rank r so all eigenvalues of A which are not eigenvalues of $\hat{S}$ are zero.

Suppose w is an eigenvalues of $\hat{S}$. Then,

$$\lambda w = \hat{S} w = \hat{U}^* A \hat{U} w \qquad \Longrightarrow \qquad \hat{U}\hat{U}^* A \hat{U} w = \lambda \hat{U} w$$

The $\hat{U}w$ is an eigenvector of A if A lies in the range of X, since $\hat{U}\hat{U}^*$ is an orthogonal projector to the range of X.

In this case,

$$\hat{\varphi} = \hat{U}w = \frac{1}{\lambda} A\hat{U}w = \frac{1}{\lambda} X'\hat{V}\hat{\Sigma}^{-1}\hat{U}^*\hat{U}w = \frac{1}{\lambda} X'\hat{V}\hat{\Sigma}^{-1}w$$

Let $\varphi = \frac{1}{\lambda} X'\hat{V}\hat{\Sigma}^{-1}w$. Then,

$$A\varphi = A(X'\hat{V}\hat{\Sigma}^{-1})w = (X'\hat{V}\hat{\Sigma}^{-1}\hat{U}^*)X'\hat{V}\hat{\Sigma}^{-1}w = \varphi\hat{S}w = \lambda\varphi w$$

Note that this is true regardless of whether A lies in the range of X.

11.2 Projective vs. Exact DMD

We now have two different ways of defining DMD modes. We call $\hat{\varphi} = \hat{U}w$ a projective DMD modes, and $\varphi = \frac{1}{\lambda} X'\hat{V}\hat{\Sigma}^{-1}$ an exact DMD mode.

Projective DMD produces modes contained in the range of X, while exact DMD modes lie in the range of Y.

As noted above, the exact DMD modes are always eigenvectors of A. On the other hand, the projective DMD modes are eigenvectors of A when $\hat{U}\hat{U}^* A = A$.

11.3 DMD with Rank Truncation

We have seen that we can use the reduced SVD to compute the DMD modes. We can instead use a truncated SVD to approximate the DMD modes. This will produce low rank approximations to the DMD modes. How good of approximations these are depends on the singular values of X.

Note: DO we have any way of knowing how good though?

Note: Is projective ever better in the approximate case?

11.4 Implementations

Method. (Dynamic Mode Decomposition)

1. Arrange data $(x_1, x_1'), (x_2, x_2'), \ldots, (x_m, x_m')$ into X and X'
2. Compute (reduced/rank-truncated) SVD $X = U\Sigma V^*$
3. Define, $\tilde{S} = U^* X' V \Sigma^\dagger$
4. Compute eigen-decompoition, $\tilde{S}W = W\Lambda$
5. Define DMD modes $\Phi = X'V\Sigma^\dagger W$ **Note:** scale by Λ?
6. Approximate solution is $x(t) = \Phi e^{\Omega t}\Phi^\dagger x(0)$, where $\omega_i = \ln(\lambda_i)/\Delta t$.

12 Koopman Embedding

12.1 Example

Start with nonlinear system $\dot{x} = f(x)$.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu x_1 \\ \lambda(x_2 - x_1^2) \end{bmatrix}$$

Define $y_1 = x_1$, $y_2 = x_2$, $y_3 = x_1^2$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \mu x_1 \\ \lambda(x_2 - x_2^2) \\ 2x_1\dot{x_1} \end{bmatrix} = \begin{bmatrix} \mu y_1 \\ \lambda(y_2 - y_3) \\ 2\mu y_3 \end{bmatrix} = \begin{bmatrix} \mu & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 2\mu \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

In our new coordinates we now have a linear system $\dot{y} = Ky$.

Note: something about fixed points

13 Time Delay Embedding

14 SINDy