A GRADUATE COURSE

IN

STOCHASTIC PROCESSES

A GRADUATE COURSE

IN

STOCHASTIC PROCESSES

DDA6001 Notebook

Lecturer

Prof. Masakiyo Miyazawa

Institute for Data and Decision Analytics

Tex Written By

Mr. Jie Wang

The Chinese University of Hongkong, Shenzhen

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments

This book is taken notes from the DDA6001 in spring semester, 2020. These lecture notes were taken and compiled in LaTeX by Jie Wang, an undergraduate student in Spring 2020. Prof. Masakiyo Miyazawa has not edited this document. Students taking this course may use the notes as part of their reading and reference materials. This version of the lecture notes were revised and extended for many times, but may still contain many mistakes and typos, including English grammatical and spelling errors, in the notes. It would be greatly appreciated if those students, who will use the notes as their reading or reference material, tell any mistakes and typos to Jie Wang for improving this notebook.

Notations and Conventions

$\sup$	least upper bound
$\inf$	greatest lower bound
$\overline{E}$	closure of E
$f \circ g$	composition
$\limsup(\liminf)$	upper (lower) limit
$L(\mathcal{P}, f), U(\mathcal{P}, f)$	Riemann sums
$\mathcal{R}[a, b]$	classes of Riemann integrable functions on $[a, b]$
$\underline{\int_a^b} f(x)\,\mathrm{d}x, \overline{\int_a^b} f(x)\,\mathrm{d}x$	Riemann integrals
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$	inner product
$\omega(f; E)$	oscillation of f over set E
$\\| \cdot \\|$	norm
∇f	gradient
$\frac{\partial f}{\partial x_i}, f_{x_i}, f_i, \partial_i f, D_i f$	partial derivatives
$D_{\boldsymbol{v}} f$	directional derivative at direction $\boldsymbol{v}$
$\frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_n)}$	Jacobian
$\mathbb{S}^n$	set of real symmetric $n \times n$ matrices
$\succ$ $(\succeq)$	positive (semi)-definite
$\mathcal{C}^m$	classes of m-th order continuously differentiable functions
$\mathcal{C}(E; \mathbb{R}^m)$	set of $\mathcal{C}^1$ mapping from E to $\mathbb{R}^m$
$(\mathcal{H}, d)$	metric space

Chapter 1

Week1

1.1. Monday

1.1.1. Introduction to Stochastic Process

Course Information.

- Instructor: Prof. Masakiyo Miyazawa
- Course Venue: Chengdao 208
- Office Hour: 3-5pm, Monday, at Daoyuan 514

Motivation of Stochastic Process.

- capture randomness;
- characterize randomness influences in real systems;
- improve the design and operation of such systems.

■ **Example 1.1** [Coin Tossing Example] Consider a scenario where someone bets money by coin tossing. Assume that

- Head and tail occur likely and independently;
- Get or loss money by head or tail, respectively;
- There is no limit for money to bet.

The goal of this people is to double the money he initialy obtains. Therefore, the best policy is that he should stop the betting when the money becomes double of the initial. The question is that is this policy **realizable**?

- By probability theory, we could argue that such event will eventually happen with probability one.
- However, this policy is actually not realizable. This is because the expected time for the realization of this event is infinite! See the Figure. 1.1 for three simulation cases of betting.

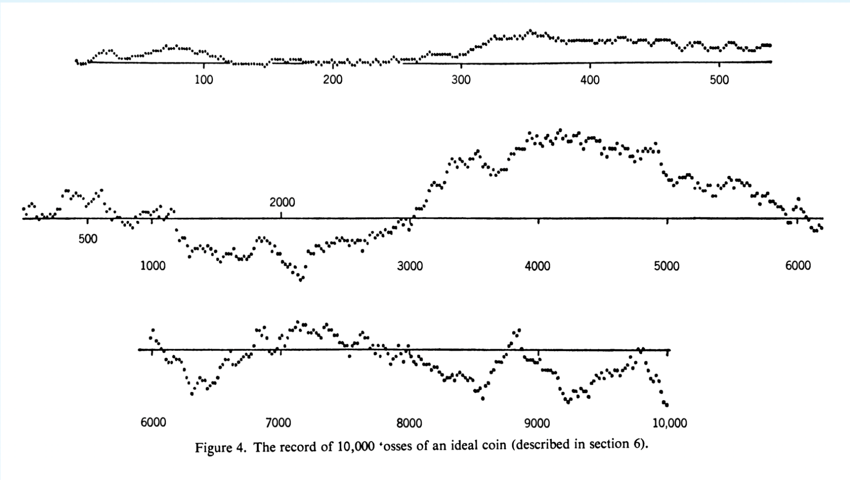

Figure 1.1: Simulation of coin tossing for 10000 times

Course Outline. Mathematical Model for random phenomena: probability space.

-

Two ways to constructor a probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

- Measure theory based: $\mathcal{F} = \mathcal{B}(\Omega)$.

  Standard for stochastic processes. (indexed by time)

- Functional Analysis based: Define the probability $\mathbb{P}$ by functional mapping:

$$\begin{aligned} \phi : \quad & \mathcal{C}_b(\Omega) \to \mathbb{R}_+ \\ \text{with} \quad & f \in \mathcal{C}_b(\Omega) : \Omega \to S \text{ (state space)} \end{aligned}$$

  Also we could define the norm for f and ϕ:

$$\|f\| = \sup_{x \in \Omega} f(x), \quad \|\phi\| = \sup_{\|f\| \leq 1} \|f\|$$

  Suitable for Gaussian Processes (indexed by vector).

Discrete time stochastic processes (DTSP):

- random walk and limit theorems
- Conditional expectation, martingale, stopping time
- DTMC and applications;

Continuous time stochastic processes (CTSP):

- Possion process, Brownian motion, martingale
- Stochastic analysis; predictability, semi-martingale, stochastic integral, Ito's formula
- CTMC and applications.

W. Rudin Real and Complex Analysis.

1.1.2. Mathematical Background

Logics. Get familar with the operations $\wedge, \vee, \neg, \equiv, \Rightarrow$:

- $P \Rightarrow Q$ is T iff $\neg P \vee Q$ is T;
- $\neg(P \Rightarrow Q) \equiv P \wedge (\neg Q)$;
- $\neg(\forall x \in A, P(x)) \equiv \exists x \in A, \neg P(x)$

Limits. Get familar with $\sup, \inf, \limsup, \liminf, \lim$:

- $\sup\limits_{n\geq 1} a_n = \bar{a}$ iff
$$\begin{cases} \forall \varepsilon > 0, \exists n_0, \forall n \geq n_0, a_n < \bar{a} + \varepsilon \\ \forall \varepsilon > 0, \forall n \geq 1, \exists n_1 \geq n, \bar{a} - \varepsilon < a_{n_1} \end{cases}$$
and $\inf\limits_{n\geq 1} a_n = -\sup\limits_{n\geq 1}(-a_n)$.

- $\limsup\limits_{n\to\infty} a_n$ and $\liminf\limits_{n\to\infty} a_n$.

- $\lim\limits_{n\to\infty} a_n = a$ iff $\limsup\limits_{n\to\infty} a_n = a = \liminf\limits_{n\to\infty} a_n$. Or equivalently,
$$\forall \varepsilon > 0, \exists n_0 \geq 1, \forall n \geq n_0, \qquad |a_n - a| < \varepsilon.$$

Set Theory. Given mapping $f : A \to B$

- Onto mapping: $f(A) = B$;

- One-to-one mapping: $f(x) = f(y)$ implies $x = y$

- Berstein's theorem: if there are both one-to-one mappings from A to B and from B to A, then $\|A\| = \|B\|$.

1.1.3. Random Phenomena

The key feature for random phenomena is **multiple outcomes**. Define possible outcomes as follows.

Definition 1.1 [Sample Space] An outcome is called a sample. The set of all possible samples is called a **sample space**, denoted by Ω. ■

■ **Example 1.2** Toss a coin for n times. Define H for head and T for tail. The possible outcome can be denoted as $\omega = (\omega_i)_{i=1}^n$, with $\omega_i \in \{H, T\}$. The sample space Ω is the set of all these ω, with 2^n elements. ■

In order to characterize behaviours such as coin tossing, we need to define how to choose a sample from Ω. The statistical way is to assign a number to each sample which represents the corrresponding likelihood to choose it, called a **probability**. The formal definition of probability relies on the distribution function:

Definition 1.2 [Distribution] Assume Ω to be finite or countably infinite set. Define a function $p:\Omega \to [0,1]$ as follows:

- $p(\omega) \geq 0, \forall \omega \in \Omega$
- $\sum_{\omega\in\Omega} p(\omega) = 1$

then $p(\cdot)$ is called a **distribution** over Ω. Moreover, $(\Omega, p(\cdot))$ is called a **discrete probability space**, and $p(\omega)$ denotes the probability that the event ω occurs.

R This definition cannot be directly applied into the uncountable sample space Ω. For instance, consider the unlimited coin tossing behaviour. The uncountable Ω defined as follows:

$$\Omega = \{0,1\}^{\infty} \supseteq [0,1].$$

Since any event ω happens with the same probability as the coin is fair, one can argue that $p(\omega) = 0, \forall \omega \in \Omega$, which contradicts to the second condition in Definition 1.2.

Ideas to define a probability for uncountable sample space. Use limiting operation when faced uncountable.

Definition 1.3 [σ-field] A set $\mathcal{F}$ containing subsets of Ω is called a σ-field if:

1. $\Omega \in \mathcal{F}$;
2. $\mathcal{F}$ is closed under complement, i.e., $A \in \mathcal{F}$ implies $\Omega \setminus A \in \mathcal{F}$;
3. $\mathcal{F}$ is closed under infinite union operation, i.e., $A_i \in \mathcal{F}, i \geq 1$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition 1.4 [Probability Measure] A function $\mathbb{P}:\mathcal{F}\to\mathbb{R}$ is called a **probability measure** on $(\Omega,\mathcal{F})$ if

- $\mathbb{P}(\Omega)=1$;
- $\mathbb{P}(A)\ge 0, \forall A\in\mathcal{F}$;
- $\mathbb{P}$ is σ-additive, i.e., when $A_i\in\mathcal{F}, i\ge 1$ and $A_i\cap A_j=\emptyset, \forall i\neq j$,

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

where $\mathbb{P}(A)$ is called the **probability of the event** A.

Definition 1.5 [Probability Space] A **probability space** is a triple $(\Omega,\mathcal{F},\mathbb{P})$, where

- Ω is a set of samples, called the sample space;
- $\mathcal{F}$ is a σ-field, a collection of events;
- $\mathbb{P}$ is a probability measure, which assigns probability to events.

R Here we give an explicit construction of the **canonical** probability space, as the σ-field induced by Ω may not be unique.

- When Ω is discrete, we can construct **the** probability space $(\Omega,\mathcal{F},\mathbb{P})$ such that

$$\mathcal{F}=2^{\Omega},\quad \mathbb{P}(A)=\sum_{\omega\in A}p(\omega),\quad A\in\mathcal{F}.$$

- Otherwise, we define $\mathcal{F}$ with no "redundancy". When Ω is a **topological space**, with $\mathcal{O}$ being the **topology** of Ω, define

$$\mathcal{B}(\Omega)=\sigma(\mathcal{O})\equiv\text{the minimal }\sigma\text{-field on }\Omega\text{ containing }\mathcal{O}.$$

Let $\mathcal{F}=\mathcal{B}(\Omega)$ and construct the associated probability measure $\mathbb{P}$. Then

$(\Omega, \mathcal{F}, \mathbb{P})$ is the **canonical** probability space.

Next we give some reviewings on the topological space:

Definition 1.6 A **topological space** $(X, \mathcal{T})$ consists of a (non-empty) set X, and a family of subsets of X ("open sets" $\mathcal{T}$) such that

1. $\emptyset, X \in \mathcal{T}$
2. $U, V \in \mathcal{T}$ implies $U \bigcap V \in \mathcal{T}$
3. If $U_\alpha \in \mathcal{T}$ for all $\alpha \in \mathcal{A}$, then $\bigcup_{\alpha \in \mathcal{A}} U_\alpha \in \mathcal{T}$.

When $A \in \mathcal{T}$, A is called the open subset of X. The $\mathcal{T}$ is called a **topology** on X.

Example 1.3 Let (X, d) be any metric space, and define

$$\mathcal{T} = \{O \subseteq X : \ \forall x \in O, \exists \varepsilon > 0, \mathbb{B}_\varepsilon(x) \subseteq O\}$$

It's clear that $(X, \mathcal{T})$ is a topological space.

Definition 1.7 [Topological Space] A is open when

$$\forall x \in A, \exists B \in \mathcal{O}, x \in B \subseteq A$$

Define the distribution function for $\Omega = \mathbb{R}$:

Define the probability of an event $A \in \mathcal{B}(\mathbb{R})$.

1.1.4. Random variable and test functions

$$\mathbb{E} f(X), \qquad f \in \mathcal{C},$$

where f is called a **test function** of X.

1.2. Thursday

Reviewing. Probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω: collection of data;
- $\mathcal{F}$: collection of events;
- $\mathbb{P}$: probability measure $\mathcal{F} \to [0,1]$ satisfying:
  - $\mathbb{P}(\Omega) = 1$;
  - $A_i \in \mathcal{F}, A_i \cap A_j = \emptyset$, then

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

From observation to number: $\Omega \ni \omega \to X(\omega) \in \mathbb{R}$. Therefore, X should be measurable, i.e.,

$$\{\omega \in \Omega; X(\omega) \in B\} := \{X \in B\} := X^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

If X is measurable, we call X as a random variable. We can also show that

$$\{\omega \in \Omega; X(\omega) \in B\} \in \mathcal{F} \Longleftrightarrow \{X \leq a\} \in \mathcal{F}$$

1.2.1. Representative value of random variable X

We define the random variable by the following three steps:

- Consider first that X is simple, i.e., taking finite many values. Assume that X only takes $x_1, x_2, \dots, x_n \in \mathbb{R}$.

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

  with $\mathbb{P}(X = x_i) = \mathbb{P}(\{X = x_i\}) = \mathbb{P}\{\omega \in \Omega; X(\omega) = x_i\}$.

- Then consider the case where X is non-negative, which is approximated by simple random variables. For each $n \geq 1$,

$$X_n(\omega) = \begin{cases} (i-1)2^{-n}, & \text{if } (i-1)2^{-n} \leq X(\omega) < i2^{-n}, i = 1,2,\dots,n2^n+1 \\ n, & \text{if } X(\omega) \geq n \end{cases}$$

Taking $n \to \infty$, $X_n(\omega) \uparrow X(\omega)$. Define $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

- Consider finally X is in the whole real line.

$$X_+(\omega) = \max\{0, X(\omega)\} \geq 0$$

■ **Example 1.4** Consider the news vender problem: ■

The random variable X may take other types of values than $(-\infty,\infty)$, such as vector values.

$$X: \quad \omega \in \Omega \to X(\omega) \in S,$$

with S being a topological space. A speical case is the metric space. The advantage of topological space is that we can use it to define borel set conveniently:

$$\mathcal{B}(S) := \sigma(\{\text{topology of } S\})$$

X is S-valued random variable if $X: \Omega \to S$ and $\{X \in B\} \in \mathcal{F}, \forall B \in \mathcal{B}(S)$.

It may be difficult to evaluate the expectation of a random variable, and therefore we pick the test function

$$f: S \to \mathbb{R}$$

satisfying measurability:

$$f^{-1}(B') \in \mathcal{B}(S), \forall B' \in \mathcal{B}(\mathbb{R})$$

Define $f(X)(\omega) = f(X(\omega)), \forall \omega \in \Omega$:

$$\Omega \xrightarrow{X} S \xrightarrow{f} \mathbb{R}.$$

We use a family of test functions, denoted by $\mathcal{C}$. If we take a sufficiently large $\mathcal{C}$, we can determine the distribution of X.

- When $S = \mathbb{R}$,

  $$\mathcal{C} = \{f_\theta : \ f_\theta(x) = e^{\theta x}, \theta \in \mathbb{R}\}$$

  and $\mathbb{E}[f_\theta(x)]$ is the moment generating function.