Seoul National University

Lecture Note

# Introduction to Stochastic Differential Equations

Lecture by Seo Insuk
Notes taken by Lee Youngjae

October 6, 2018

# Contents

# Chapter 0

# Introduction

E-mail: *insuk.seo@snu.ac.kr, 27-212*
Office Hour: Tuesday 15:00 - 16:00
Grading

- Mid-terms 1 (15%, 10/10 or 17)
- Mid-terms 2 (15%, 11/7)
- Fianl-term (40%)
- Assignment (20%, 8-10 times)
- Attendance (10%, absent: -2%, late: -1%)

Let $X$ be a standard normal random variable in $\mathbb{R}$. i.e., $\mathbb{P}[X \in [a,b]] = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.
(Central Limit Theorem) If $x_1, x_2, \cdots, x_n \in X, E(x_i) = m, Var(x_i) = \sigma^2$, then

$$\frac{\frac{x_1 - m}{\sigma} + \frac{x_2 - m}{\sigma} + \cdots + \frac{x_n - m}{\sigma}}{\sqrt{n}} \to X$$

In this class, we study dynamic version of this theorem. If $(W_t)_{t \geq 0}$ be a fluctuation, then $(W_t)_{t \geq 0}$ be a random variable in $C[0, T]$

*Example.* $\frac{dX_t}{dt} = rX_t; dX_1 = rX_t dt$. Then, $X_t = X_0 e^{rt}$ (unrisky assets, bank)
$dX_t = rX_t dt + \sigma X_t dW_t$, $\sigma$ : volatility (risky assets, stock)

We will study:

1. Probability Space
2. Random Variable
3. Expectation

Textbooks:

1. Stochastic Calculus for Finance II (Shreve), covering chapter 1-3 or 4
2. Introduction to Stochastic Integration (Hui-Hsiung Kuo)

# Chapter 1

# General Probability Theory

## 1.1 Infinite Probability Spaces

There are three elements consisting probability space:

- $S$: Sample space
- $\mathcal{E}$: Family of events $\mathcal{E} \subseteq 2^S$ ($\sigma$-algebra in measure theory)
- $\mathbb{P}$: probability $\Rightarrow \mathbb{P}(E)$ is defined for all $E \in \mathcal{E}$ ($\mu$ with $\mu(S) = 1$)

*Example.*

1. Toss a coin twice (H for Head, T for Tail)
   Then, $S = \{HH, HT, TT, TT\}$
2. Uniform random variable in $[0, 1]^3$
   Then, $S = [0, 1]^3$. If $E = [0, \frac{1}{2}]^3$, then $\mathbb{P}(E) = Vol(E) = \frac{1}{8}$

How to define $\mathcal{E}$?
In example 2, let $\mathcal{E}$ = family of all subsets of $[0, 1]^3$ naively. But Banach-Tarski Paradox says there are disjoint sets $E, F$ with $\mathbb{P}(E \cup F) \neq \mathbb{P}(E) + \mathbb{P}(F)$ in this $\mathcal{E}$. Therefore we cannot naively set $\mathcal{E}$ (Use measure theory)

In example 1, suppose that we cannot see the second flip. If $\{HH\} \notin \mathcal{E}$ and $\{HT, HH\} \in \mathcal{E}$, then $\mathcal{E} = \{\phi, \{HH, HT\}, \{TH, TT\}, \{HH, HT, TH, TT\}\}$

**Definition 1.1.1** (Measure)
Let $\Omega$ be a non-empty set and $\mathcal{F}$ be family of subsets of $\Omega$ with

1. $\phi \in \mathcal{F}$

2. $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$

3. $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

We say $\mathcal{F}$ as **$\sigma$-algebra** or **$\sigma$-field**, $A \in \mathcal{F}$ as **measurable**, and $\Omega$ as **measurable space**.

*Exercises.*

1) $\Omega \in \mathcal{F}$

2) $A_1, A_2, \cdots \in \mathcal{F}$, then $A_1 \cap A_2 \cdots \in \mathcal{F}$

3) $A_1, A_2, \cdots \in A_n \in \mathcal{F}$, then $A_1 \cup \cdots \cup A_n, A_1 \cap \cdots \cap A_n \in \mathcal{F}$.

4) $A, B \in \mathcal{F}$, then $A - B \in \mathcal{F}$

**Definition 1.1.2** (Topological Space)
(See Rudin: *Real and Complex Analysis, Chapter 1.*) Let $\Theta$ be non-empty set and $\tau$ be family of subsets of $\Theta$ with

1. $\phi, \Theta \in \tau$

2. $V_1, \cdots V_n \in \tau \Rightarrow V_1 \cap \cdots \cap V_n \in \tau$

3. $V_\alpha \in \tau \;\; \forall \alpha \in I \Rightarrow \bigcup_{\alpha \in I} V_\alpha \in \tau$.

We say $V \in \tau$ be an **open set**, and $(\Theta, \tau)$ be a **topological space**.

**Definition 1.1.3** (Measurable Function)
$f : (\Omega, \mathcal{F}) \to (\Theta, \tau)$ is **measurable** if $f^{-1}(V) \in \mathcal{F} \;\; \forall V \in \tau$

**Definition 1.1.4** (Positive Measure)
Let $\Omega$ be non-empty set and $\mathcal{F}$ be $\sigma$-algebra. Then $\mu : \mathcal{F} \to [0, \infty]$ is called **measurable** if

1. $A_1, A_2, \cdots$: disjoint members of $\mathcal{F} \Rightarrow \mu(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} \mu(A_i)$

2. $\mu(A) < \infty$ for some $A \in \mathcal{F}$,

and $(\Omega, \mathcal{F}, \mu)$ is called a **measure space**.

**Definition 1.1.5** (probability space, random variable)

1. $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space** if $\mathbb{P}(\Omega) = 1$.

2. $X$ is called a **random variable** if it is a function from $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$

Next Class

- Borel sets on $\mathbb{R}$ or $\mathbb{R}^d$
- Lebesgue Measure
- Lebesgue Integral (Define Expectation of random variable)

Last class, we define a sample space $\Omega$, a $\sigma$-algebra $\mathcal{F}$, and a (positive) measure $\mu : \mathcal{F} \to [0, \infty]$.

*Exercises.*

- $A_1 \subseteq A_2 \subseteq \cdots \Rightarrow \mu(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(A_n)$
- $A_1 \supseteq A_2 \supseteq \cdots, \mu(A_1) < \infty \Rightarrow \mu(\bigcap_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(A_n)$

**Theorem 1.1.6** (Rudin 1.10)
*Let $\mathcal{F}_0$ be a collection of subset of $\Omega$. Then, $\exists! \mathcal{F}^*$ minimal $\sigma$-algebra containing $\mathcal{F}_0$.*

*Proof.* Let $\{\mathcal{F}_\alpha, \alpha \in I\}$ be a family of $\sigma$-algebra containing $\mathcal{F}_0$. Then, $\mathcal{F}^* = \bigcap_{\alpha\in I} F_\alpha$ satisfies the three condition: 1) contain $\mathcal{F}_0$ 2) $\sigma$-algebra 3) minimal (trivial, $\mathcal{F}^* \subseteq \mathcal{F}_\alpha$) $\square$

**Definition 1.1.7** (Borel measurable)
$\mathcal{B}$ is a **Borel $\sigma$-algebra** on the topological space $(\Theta, \tau)$ if $\mathcal{B}$ is a minimal $\sigma$-algebra containing $\tau$, and $B$ is a **Borel measurable** if $B \in \mathcal{B}$.

*Remark* (Completion of measure space, Rudin 1.15).
Consider an extension $(\Omega, \mathcal{F}, \mu) \to (\Omega, \overline{\mathcal{F}}, \mu)$ where

1. $\overline{\mathcal{F}} = \{A \cup N : A \in \mathcal{F}, N \subseteq A_0 \in \mathcal{F}, \mu(A_0) = 0\}$
2. $\mu(A \cup N) = \mu(A)$

Then, (Check!)

1. (well-definedness) $A_1 \cup N_1 = A_2 \cup N_2 \Rightarrow \mu(A_1) = \mu(A_2)$
2. $\mu : \overline{\mathcal{F}}$ is $\sigma$-algebra.
3. $\mu : \overline{\mathcal{F}} \to [0, \infty]$ is a measure

*Example.*

1) $\mathbb{R}$

$$\mathcal{F}_0 = \tau \xrightarrow{1.10} \mathcal{B} \xrightarrow{\text{completion}} \overline{\mathcal{B}}$$

$$\mathcal{L} \xrightarrow{\text{Rudin CH 2}} \mathcal{L} \xrightarrow{\text{completion}} \mathcal{L}$$

2) $C[0,T] = \Omega = \{f; f : [0,T] \to \mathbb{R}, \text{continuous}\}$.
Define $\mathcal{F}_0 = \{\bigcup_{t_1,t_2,\cdots,t_k}(A_1, A_2, \cdots, A_k) : 0 \le t_1 < t_2 < \cdots < t_k \le T; A_1, \cdots A_k \in \overline{\mathcal{B}}\}$. We call $\{f \in C[0,T] : f(t_1) \in A_1, f(t_2) \in A_2, \cdots, f(t_k) \in A_k\}$ as **cylindrical set**. Consider

$$\mathcal{F}_0 \xrightarrow{1.10} \quad \mathcal{B} \xrightarrow{\text{completion}} \quad \overline{\mathcal{B}}$$

$$\mathbb{P}_{\text{BM}} \xrightarrow{\text{KET}} \mathbb{P}_{\text{BM}} \xrightarrow{\text{completion}} \mathbb{P}^*_{\text{BM}}$$

(KET refers Kolmogorov's Extension Thm)

## 1.2 Random Variables and Distributions

**Definition 1.2.1**
$f : \Omega \to \mathbb{R}$ is measurable if $f^{-1}(V) \in \mathcal{F}$ for any open set $V \subseteq \mathbb{R}$.

*Remark.* $\mathcal{B}(\mathbb{R})$ = Borel $\sigma$-algebra in $\mathbb{R}$.

*Remark.* If $f$: measurable, then $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}(\mathbb{R})$.

*Proof.* Let $G = \{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathcal{F}\}$. Then, $\tau \subseteq G$, $G$ : $\sigma$-algebra (check!), hence $\mathcal{B}(\mathbb{R}) \subseteq G$. □

**Definition 1.2.2**

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a **probability space** if $(\mathbb{P}(\Omega) = 1$.
- $X$ is a **random variable** if $X : \Omega \to \mathbb{R}$ is measurable.

*Example.*

1. Toss a coin Twice.
$\Omega = \{HH, HT, TH, TT\}$, $\mathcal{F} = 2^\Omega = \{\text{all subsets of } \Omega\}$, $\mathbb{P}(A) = \frac{1}{4}|A|, \ \ A \in \mathcal{F}$.
Then, $X$ = the number of $H$'s is a random variable with $X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 1$.

2. Uniform random variable in $[0,1]$
$\Omega = [0,1]$, $\mathcal{F} = \{B \in \mathcal{B}(\mathbb{R}) : B \subseteq [0,1]\}$, $\mathbb{P}(B) = \mathcal{L}(B)$ $(\mathbb{P}([0,1]) = \mathcal{L}([0,1]) = 1)$.
Then, $X : [0,1] \to \mathbb{R}$ with $X(x) = 1$ is a (uniform) random variable in $[0,1]$.

*Remark.* $\mathcal{L}$: Lebesgue measure on $\mathbb{R}$. i.e.,$\mathcal{L}(a,b) = b - a$. Then, $\mathcal{L}(\{a\}) = 0$
$(\because \{a\} = \bigcap_{i=1}^{\infty}(a - \frac{1}{n}, a + \frac{1}{n}) \Rightarrow \mathcal{L}(\{a\}) = \lim_{n\to\infty} \mathcal{L}((a - \frac{1}{n}, a + \frac{1}{n})) = 0)$
Similarly, $\mathcal{L}([a,b]) = \mathcal{L}([a,b)) = \mathcal{L}((a,b]) = b - a$, $\mathcal{L}(\mathbb{Q}) = \sum_{q\in\mathbb{Q}} \mathcal{L}(\{q\}) = 0$.

Return to uniform random variable,

$$\mathbb{P}[X \in (a,b)] = \mathbb{P}[\{x : X(x) \in (a,b)\}] = \mathbb{P}[(a,b)] = b - a.$$

**Definition 1.2.3** (Distribution measure on $X$)
$X$ is a random variable in $(\Omega, \mathcal{F}, \mathbb{P})$. $\mu_X$ is a **distribution measure** on $X$ if $\mu_X$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mu_X(B) = \mathbb{P}[X \in B] = \mathbb{P}[\{\omega : X(\omega) \in B\}] = \mathbb{P}[X^{-1}(B)] \; \forall B \in \mathcal{B}(\mathbb{R})$$

**Definition 1.2.4** (Probability density function)
$f$ is a **probability density function** of $X$ if $\mu_X((a,b)) = \int_a^b f(x)dx$

*Remark.* There is a measure with no pdf: Dirac measure

*Remark.* Lebesgue-Radon-Nikodym decomposition implies that any measure can be decomposed as density part and singular part.

*Example* (Standard Normal random variable).
Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Define $F : (0,1) \to \mathbb{R}$ by $F(x) = N^{-1}(x)$ for $N(X) = \int_{-\infty}^{x} \phi(y)dy$.
Let $\Omega = (0,1), \mathcal{F} = \{B \in \mathcal{B}(\mathbb{R}) : B \subseteq (0,1)\}, \mathbb{P}(A) = \mathcal{L}(A) : A \in \mathcal{B}(\mathbb{R})$.
Then, $Y : \Omega \ni x \mapsto F(x) \in \mathbb{R}$ is a random variable with

$$\begin{aligned} \mathbb{P}[Y \in (a,b)] &= \mathbb{P}[\{x : Y(x) \in (a,b)\}] \\ &= \mathbb{P}[\{x \in (N(a), N(b))\}] \\ &= N(b) - N(a) = \int_a^b \phi(x)dx, \end{aligned}$$

and a density function is $\phi$.

Previous Question: In the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable $X : \Omega \to \mathbb{R}$, the random element or random realization $\omega \in \Omega$ is a element of events in sample space. For example, $\omega = HHTTH$ is a random element in tossing a coin five times, and $X(\omega) = 3$. ($X(\omega)$ = # of Heads)
In the previous example(Standard Normal random variable), define $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \mathbb{P})$, $\mathbb{P}((a,b)) = b - a$, $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$, $X : (0,1) \ni \omega \mapsto F^{-1}(\omega) \in \mathbb{R}$. Then, $X$ is called a standard normal random variable.

## 1.3 Expectations

In the following, let $\Omega = (\Omega, \mathcal{F}, \mathbb{P})$. Let $X : \Omega \to \mathbb{R}$. Then the expection $\mathbb{E}(X)$ is a mean of $X(\omega)$ with respect to the randomness of $\omega$ (given by $\mathbb{P}$)

**Definition 1.3.1** (Lebesgue Integration)
$(\Omega, \mathcal{F}, \mu)$ is a measure space, and $f : \Omega \to \mathbb{R}$ is a measurable function.

(1) $f : \Omega \to [0, \infty)$
Let $0 = y_0 < y_1 < y_2 < \cdots \to \mathbb{R}$ be a partition of $[0, \infty)$,
$\Pi = \{y_0, y_1, y_2, \cdots\} : \|\Pi\| = \sup_{i \geq 1} |y_i - y_{i-1}|$, and
$\mathrm{LS}_\pi = \sum_{i=0}^{\infty} y_i \mu[f^{-1}([y_i, y_{i+1}))]$.
In Rudin's book, $\lim_{\|\Pi\| \to 0} \mathrm{LS}_\Pi$ converges to an element belonging to $[0, \infty]$.
Now, $\int f d\mu := \lim_{\|\Pi\| \to 0} \mathrm{LS}_\Pi$ is called a **Lebesgue Integral**.

(2) $f : \Omega \to \mathbb{R}$
Let $f^+ = \max\{f, 0\} \geq 0$, and $f^- = -\min\{f, 0\} \geq 0$. Then, $f = f^+ - f^-$, and $|f| = f^+ + f^-$. If $\int f^+ d\mu < \infty$ and $\int f^- d\mu < \infty$, then we say $f$ is Lebesgue integrable and $f \in L^1(\mu)$. The Lebesgue integral of $f = \int f d\mu$ is defined as $\int f^+ d\mu - \int f^- d\mu$

*Remark.*

1. $\int f^+ d\mu < \infty$ and $\int f^- d\mu = \infty$, then $\int f d\mu = -\infty$. The others are defined similarly.
2. $f \in L^1(\mu) \Leftrightarrow \int |f| d\mu < \infty$.

*Example* (Riemann vs Lebesgue Integral (p19-22)).

- $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L})$ Lebesgue measure where $\mathcal{L}((a, b)) = b - a$.
- $f : \mathbb{R} \to \mathbb{R} \in L^1(\mathcal{L})$
- (Def) $A \subseteq \mathbb{R}$, $\int_A f d\mu := \int f \mathbb{1}_A d\mu$, where $\mathbb{1}_A(x) = 1$ if $x \in A$, and 0 otherwise.

If $f$ is Riemann integrable, then $\int_{[a,b]} f d\mathcal{L} = \int_a^b f dx$.
Riemann integral is a limit of approximation by a partition of $x$-axis. On the other hand, Lebesgue integral is a limit of approximation by a partition of $y$-axis with preimage. Partition of $x$-axis is sensitive to fluctuation and restricted to Euclidean space, while partition of $y$-axis is not. For example, $f(x) = \mathbb{1}_{\mathbb{Q}}(x)$ is Lebesgue integrable, but it is not Riemann integrable since it is sensitive to fluctuation.

**Definition 1.3.2** (Almost everywhere, 1.1.5 in Textbook)
$P(x)$ is a property at $x \in \mathbb{R}$. We say $P$ holds **almost everywhere** (or a.e.) in $\mathbb{R}$ if and only if $\mathcal{L}(\{x : P(x) \text{ does not hold }\} = 0$.

*Example.* $f(x) = [x]$ is continuous almost everywhere.

**Theorem 1.3.3**
*$f$ is Riemann integrable if and only if $f$ is continuous a.e.*

*Exercises.* $f = g$ a.e. $\Rightarrow \int f d\mathcal{L} = \int g d\mathcal{L}$.

**Definition 1.3.4** (Almost surely)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The event $A (\in \mathcal{F})$ occurs **almost surely** (a.s.) if $\mathbb{P}(A) = 1$.

*Example.* Let $X$ be a uniform random variable in $(0,1)$. Let $A = \{X(\omega) \neq \frac{1}{2}\}$; $\mathbb{P}(A) = 1$.

**Definition 1.3.5** (Expectation, 1.3.3. in Textbook)
**Expectation** of $X : \Omega \to \mathbb{R}$ is defined by

$$\mathbb{E}(X) := \int_\Omega X d\mathbb{P} \quad \text{if} \quad \int_\Omega |X| d\mathbb{P} < \infty$$

**Theorem 1.3.6** (1.3.4 in Textbook)

1. *$X$ takes finite number of values $\{x_1, x_2, \cdots, x_n\} \Rightarrow \mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i)$*
2. *$X, Y$: random variables, $E(|X|), E(|Y|) < \infty$,*
   *(i) $X \leq Y$ a.s. (i.e. $\mathbb{P}[\{X(\omega) \leq Y(\omega)\}] = 1$), then $\mathbb{E}(X) \leq \mathbb{E}(Y)$*
   *(ii) $X = Y$ a.s. $\Rightarrow \mathbb{E}(X) = \mathbb{E}(Y)$*
3. *$X, Y$: random variables, $\mathbb{E}(|X|), \mathbb{E}(|Y|) < \infty \Rightarrow \mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$.*
4. *Jensen's Inequality: $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function $\Rightarrow \phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$*
   *(c.f. $\phi(t) = t^2$)*

*Proof of 4.* Define $S_\phi = \{(a,b) \in \mathbb{R}^2 : a + bt \leq \phi(t) \ \ \forall t\}$. Then $\forall t \in \mathbb{R}, \phi(t) = \sup_{(a,b) \in S_\phi} \{a + bt\}$. In fact, it is a equivalent condition. Now,

$$\begin{aligned} \phi(\mathbb{E}[X]) &= \sup_{a,b \in S_\phi} \{a + b\mathbb{E}[X]\} \\ &= \sup_{a,b \in S_\phi} \mathbb{E}[a + bX] \\ &\leq \mathbb{E}[\sup_{a,b \in S_\phi} (a + bX)] = \mathbb{E}[\phi(X)] \quad \text{(Check!)} \end{aligned}$$

$\square$

*Example* (Dirac Measure in $\mathbb{R}$). $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \delta_y)$ $(y \in \mathbb{R})$ is a probability space with $\delta_y(A) = 1$ if $y \in A$, and 0 otherwise. Then, $\int_{\mathbb{R}} f d\delta_y = f(y)$ (Check!)

Consider modeling: $X$: random variable such that probability of $x_i = p_i$ with $\sum_{i=1}^n p_i = 1$. Then, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ with $\mu = \sum_{i=1}^n p_i \delta_{x_i}$ is a probability space, and $P(X = x_i) = p_i$ for $X : \mathbb{R} \ni \omega \mapsto \omega \in \mathbb{R}$: Example of thm 1.3.4.

*Summary:*

- Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$
- Random variables: $X : \Omega \to \mathbb{R}$
- Expectation: $E(X) = \int X d\mathbb{P}$

## 1.4 Convention of Integrals

We will use this section when we define the Brownian motion.

**Definition 1.4.1**

(1) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $f, f_1, f_2, \cdots$ be measurable ($\Omega \to \mathbb{R}$). Then, $f_n \to f$ **almost everywhere** (a.e.) if

$$\mu[\{\omega : (f_n(\omega))_{n=1}^{\infty} \text{ does not converge to } f(\omega)\}] = 0$$

(2) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X, X_1, X_2, \cdots$ be random variables. Then, $X_n \to X$ **almost surely** (a.s.) if

$$\mathbb{P}[\{\omega : (X_n(\omega))_{n=1}^{\infty} \text{ does not converge to } X(\omega)\}] = 0$$

*Question:* $f_n \to$ a.e. Then, $\int f_n d\mu \to \int f d\mu$? $X_n \to X$ a.s. Then, $\mathbb{E}[X_n] \to \mathbb{E}[X]$?

**Theorem 1.4.2** (Monotone Convergence Theorem. 1.4.5 in Textbook)
*$0 \le f_1 \le f_2 \le \cdots$ (or decreasing), and $f_n \to f$ a.e. Then, $\int f_n d\mu \to \int f d\mu$.*

**Theorem 1.4.3** (Dominated Convergence Theorem. 1.4.9 in Textbook)
*$\exists g \in L^1(\mu)$ such that $|f_n| \le g$ for all $n$, and $f_n \to f$ a.e. Then, $\int f_n d\mu \to \int f d\mu$.*

**Corollary 1.4.4**
$\exists Y \in L^1(\mathbb{P})$ such that $|X_n| \le Y$ for all $n$, and $X_n \to X$ a.s. Then, $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

*Example.* Let $f_n(x) = \begin{cases} n^2 x & \text{if } 0 \le x \le \frac{1}{n}, \\ -n^2 x + n & \text{if } \frac{1}{n} < x \le \frac{2}{n}, \\ 0 & \text{otherwise.} \end{cases}$ Then, $f_n \to 0$ a.e. and $\int f_n dx = 1$.

## 1.5 Computation of Expectations

**Notation:** ($X : \Omega \ni \omega \mapsto X(\omega) \in \mathbb{R}$)

- $\mathbb{E}[X] = \int X d\mathbb{P} = \int_\Omega X(\omega) d\mathbb{P}(\omega)$
- $\int_B X(\omega) d\mathbb{P}(\omega) := \int \mathbb{1}_B(\omega) X(\omega) d\mathbb{P}(\omega)$

**Recall:** $X$: random variable, $\mu_X$: distribution measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mu_X(B) = \mathbb{P}(X \in B)$.

**Theorem 1.5.1**
*$g \in L^1(\mu_X)$. Then, $\mathbb{E}[g(X)] = \int_\mathbb{R} g(x) d\mu_X(x) (:= \int g d\mu_X)$.*

*Example.* $g(x) = x$. $\int |x| d\mu_X(x) < \infty \Rightarrow \mathbb{E}[X] = \int x d\mu_X(x)$.

*Proof.* First, prove the thm holds for $g \geq 0$, then prove for general $g$ by $g = g^+ - g^-$.

(1) $g = \mathbb{1}_B$
By thm 1.3.4. (1), $E[\mathbb{1}_B(X)] = 1 \cdot \mathbb{P}[\mathbb{1}_B(X) = 1] = \mathbb{P}(X \in B) = \mu_X(B) = \int \mathbb{1}_B(x) d\mu_X(x)$.

(2) $g = \sum_{k=1}^n \alpha_k \mathbb{1}_{B_k}$
Trivial by linearity.

(3) $g \geq 0$
By MCT. See *Rudin* chapter 1 for details.

□

**Recall:** $X$ : random variable, $X$ has density function $f_X$ if

$$\mu_X((a,b)) = \int_a^b f_X(x) dx \ \ \forall a, b.$$
$$\mu_X(B) = \int_B f_X d\mathcal{L} = \int_B f_X(x) d\mathcal{L}(x) = \int_B f_X(x) dx.$$

**Theorem 1.5.2**
*$g \in L^1(\mu_X)$. Then, $E(g(X)) = \int_\mathbb{R} g(x) f_X(x) dx$.*

*Example.* Let $X$ be standard normal. i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (regardless what $X$ be). Then, $E(X^4) = \int_\mathbb{R} x^4 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = 3$.

# Chapter 2

# Information and Conditioning

## 2.1 Information and $\sigma$-algebras

*Example.* Toss a coin Three times. $\Omega = \{HHH, HHT, \cdots, TTT\}$.
$A_H = \{HHH, HHT, HTH, HTT\}$, $A_T = \{THT, THT, TTH, TTT\}$.
Let $\mathcal{F}(1) = \{\phi, \Omega, A_H, A_T\}$ so that it is a $\sigma$-algebra containing the randomness up to time 1.
Similarly, define $A_{HH}, A_{HT}, A_{TH}, A_{TT}$.
Let $\mathcal{F}(2) = \{\phi, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A_{HH} \cup A_{HT}, \cdots, A_{TT}^C\}$ so that it is a $\sigma$-algebra containing the randomness up to time 2, and define $\mathcal{F}(0)$ similarly, and let $\mathcal{F}(0) = \{\phi, \Omega\}$.
Then, $\mathcal{F}(0) \subseteq \mathcal{F}(1) \subseteq \mathcal{F}(2) \subseteq \mathcal{F}(3)$. Let $X_t$ = # of heads until time $t$. Then, $X_t$ is $\mathcal{F}(t)$=measurable. for each $t$.
Now, $\{X_1 = 1\} = \{\omega : X_1(\omega) = 1\} = A_H$, and $\{X_1 = 0\} = \{\omega : X_1(\omega) = 0\} = A_T$.

**Definition 2.1.1** ($\sigma$-algebra generated by $X$)
$\Omega$ is a set, $X : \Omega \to \mathbb{R}$. $\sigma(X) = \{A \subseteq \Omega : A = X^{-1}(B)$ for some $B \in \mathcal{B}(\mathbb{R})\}$. Then, $\sigma(X)$ is a $\sigma-$algebra(exercise) and it is called a **$\sigma$-algebra generated by** $X$.

*Remark.* $X$ is a random variable in $(\Omega, \sigma(X))$.
$X$ is a random variable in $(\Omega, \mathcal{F})$, then $\sigma(X) \subseteq \mathcal{F}$ (exercise)

**Definition 2.1.2** ($\mathcal{F}$-measurable)
$(\Omega, \mathcal{F})$: measure space. $X : \Omega \to \mathbb{R}$. $X$ is called **$\mathcal{F}$-measurable** if $\sigma(X) \subseteq \mathcal{F}$. i.e., $X$: measurable with respect to $(\Omega, \mathcal{F})$.

In example, $X(t)$ is $\mathcal{F}(t)$-measurable $\forall t$ (check!)
*cf.* $X(t) : \Omega \to \mathbb{R}$. $(X(t))^{-1}(B) \in \mathcal{F}(t)\ \ \forall B \in \mathcal{B}(\mathbb{R})$.

Enough to check $(X(t))^{-1}(\{0\}), (X(t))^{-1}(\{1\}), \cdots, (X(t))^{-1}(\{t\})$.
$\mathcal{F}(t)$ has enough information to determine $X(t)$ in the sense that $\{\omega : (X(t))(\omega) \in B\} \in \mathcal{F}(t)\ \ \forall B \in \mathcal{B}(\mathbb{R})$.

**Definition 2.1.3** (Filtration, Stochastic Process)
$\Omega$: non-empty set, $T > 0$.

1. If $\mathcal{F}(t)$ is a $\sigma$-algebra $\forall t \in [0, T] \in T$ and $s < t \Rightarrow \mathcal{F}(s) \subseteq \mathcal{F}(t)$, then $(\mathcal{F}(t) : t \in [0, T])$ is called a **filtration**
2. If $X(t) : \Omega \to \mathbb{R}$ is $\mathcal{F}(t)$-measurable $\forall t \in [0, T]$, then $(X(t) : t \in [0, T])$ is called **Stochastic Process adopted to the filtration** $\mathcal{F}(t)$.

## 2.2 Independence

$X : \Omega \to \mathbb{R}$, $\mathcal{F}$: $\sigma$-algebra on $\Omega$.

1. $\mathcal{F}$ has full information to determine $X \Rightarrow X$ is $\mathcal{F}$-measurable. (2.1)
2. $\mathcal{F}$ has no information to determine $X \Rightarrow X$ is independent to $\mathcal{F}$. (2.2)
3. $\mathcal{F}$ has a partition information to determine $X \Rightarrow$ (2.3)

**Definition 2.2.1** (independent)
$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $A, B \in \mathcal{F}$ is **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

*Question:* $X, Y$ are random variables in $(\Omega, \mathcal{F}, \mathbb{P})$. If $X, Y$ are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, but the converse does not hold.

**Definition 2.2.2**
$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are sub $\sigma$-algebras of $\mathcal{F}$. $X, Y : \Omega \to \mathbb{R}$ are random variables in $(\Omega, \mathcal{F}, \mathbb{P})$.

1. $\mathcal{G}, \mathcal{H}$: independent iff $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\ \ \forall A \in \mathcal{G}, B \in \mathcal{H}$.
2. $X, Y$: independent iff $\sigma(X), \sigma(Y)$ are independent.
3. $X, \mathcal{G}$: independent iff $\sigma(X), \mathcal{G}$ are independent.

**Definition 2.2.3**
$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.
$\mathcal{G}_1, \mathcal{G}_2, \cdots \mathcal{G}_n, \cdots$: sub $\sigma$-algebra of $\mathcal{F}$. $X_1, X_2, \cdots, X_n, \cdots$: random variable in $(\Omega, \mathcal{F}, \mathbb{P})$.

1. $\mathcal{G}_1, \cdots, \mathcal{G}_2$ are independent iff $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$ for $A_1 \in \mathcal{G}_1, \cdots, A_n \in \mathcal{G}_n$.

2. $X_1, \cdots, X_n$ are independent iff $\sigma(X_1) \sim \sigma(X_n)$ are independent.

3. $\mathcal{G}_1, \mathcal{G}_2, \cdots$ are independent iff $\mathcal{G}_1 \sim \mathcal{G}_n$ are independent $\forall n$.

4. $X_1, X_2, \cdots$ are independent iff $X_1 \sim X_n$ are independent $\forall n$.

*Example.* Toss a coin three times.

1. $X(2), X(3)$ are not independent.
   $\mathbb{P}(\{X(2) = 2\} \cap \{X(3) = 1\}) \neq \mathbb{P}(X(2) = 2)\mathbb{P}(X(3) = 1)$.

2. $X(2), X(3) - X(2)$ are independent.
   Why: $X(2)$ is an information at tossing first, second times, and $X(3)$ is an information at tossing third time.

**Definition 2.2.4** (Joint distribution)
$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $X, Y$ are random variables in $\Omega$. $(X, Y) : \Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$

1. Joint Distribution Measure in $\mathbb{R}^2$
   $\mu_{X,Y}(C) = \mathbb{P}((X, Y) \in C)$ for $C \in \mathcal{B}(\mathbb{R}^2)$.
   (Note: We checked that $\{\omega : (X(\omega), Y(\omega)) \in C\} \in \mathcal{F}$ in real analysis.)

2. Joint Cumulative Distribution Function
   $F_{X,Y}(a, b) = \mathbb{P}(X \leq a, Y \leq b) = \mu_{X,Y}((-\infty, a] \times (-\infty, b])$ (check!)

3. Joint Probability Distribution Function
   If $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ is Borel-measurable and satisfies $\mu_{X,Y}(A \times B) = \int_B \int_A f_{X,Y}(x, y)dxdy$ for all $A, B \in \mathcal{B}(\mathbb{R})$, then $f_{X,Y}$ is called a joint probability density function (jpdf)

**Theorem 2.2.5**
*$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $X, Y$ are random variables in $\Omega$. Then, the followings are equivalent.*

*(i)* *$X, Y$ are independent*

*(ii)* $\mu_{X,Y}(A \times B) = \mu_X(A)\mu_Y(B) \ \ \forall A, B \in \mathcal{B}(\mathbb{R})$

*(iii)* $F_{X,Y}(a, b) = F_X(a)F_Y(b) \ \ \forall a, b \in \mathbb{R}$

*(iv)* $\mathbb{E}[e^{uX+vY}] = \mathbb{E}[e^{uX}]\mathbb{E}[e^{vY}]$

*Remark.* If JPDF $f_{X,Y}$ exists, then (i) to (iv) $\Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$ a.e.

**Theorem 2.2.6**
*$X, Y$ are independent if and only if $f, g : \mathbb{R} \to \mathbb{R}$ Borel-measurable, $\mathbb{E}[|f(X)g(Y)|] < \infty$ implies that $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$.*

*Remark.* $f(x) = g(x) = x : \mathbb{E}[|XY|] < \infty \Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

*Proof.* Details are exercises.

(1) $f = \mathbb{1}_A, g = \mathbb{1}_B$

(2) $f, g$ are simple functions

(3) $f, g \geq 0$

(4) $f, g$ are general.

□

Review
$\mathcal{G}, \mathcal{H}$ are independent if $\forall A \in \mathcal{G}, B \in \mathcal{H} \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
$X, Y$ are independent if $\sigma(X), \sigma(Y)$ are independent.
* $\sigma(X) = \{A \in \Omega : A = X^{-1}(B)$ for some $B \in \mathcal{B}(\mathbb{R})\}$.
* $\mu_{X,Y}(C) = \mathbb{P}((X,Y) \in C) \ \ \forall C \in \mathcal{B}(\mathbb{R}^2)$.

Thm. T.F.A.E.C:

1. $X, Y$ are independent
2. $\mu_{X,Y}(A \times B) = \mu_X(A)\mu_Y(B)$
3. $\mathcal{F}_{X,Y}(x,y) = \mathcal{F}_X(x)\mathcal{F}_Y(y)$
4. (If JPDF $f_{X,Y}$ exists) $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

**Theorem 2.2.7**
*$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $X, Y$ are independent random variables, $f, g : \mathbb{R} \to \mathbb{R}$ are Borel measurable. Then, $f(X), g(Y)$ are independent.*

*Proof.* $A \in \sigma(f(X))$; $A = (f \circ X)^{-1}(B)$ for some $B \in \mathbb{R} = X^{-1}(f^{-1}(B)) \in \sigma(X)$.
$\therefore \sigma(f(X)) \subseteq \sigma(X)$, $\sigma(g(Y)) \subseteq \sigma(Y) \Rightarrow \sigma(f(X)), \sigma(g(Y))$ are independent. □

**Corollary 2.2.8**
$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X))\mathbb{E}(g(Y)]$.

**Definition 2.2.9**

$X, Y$ are random variables in $(\Omega, \mathcal{F}, \mathbb{P})$.

1. $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
2. $std(X) = \sqrt{Var(X)}$
3. $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
4. $corr(X, Y) = cov(X, Y)/(std(X)std(Y))$

*Example.*

- $X$: standard normal random variable ($N(0, 1^2)$)
- $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$ ($X, Z$ are independent)
- $Y = XZ$. Then
  1) Y is standard normal,
  2) $corr(X, Y) = 0$.
  3) $X, Y$ are not independent.

**Definition 2.2.10** (Jointly normal)

$X, Y$ are **jointly normal** with mean $m = (m_X, m_Y)$, $Var(C) = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ if

$$f_{X,Y}(z) = \frac{1}{\sqrt{(2\pi)^2 \det C}} e^{-\frac{1}{2}(z-m)C^{-1}(z-m)^\intercal}$$

**Theorem 2.2.11**

*$X, Y$ are jointly normal and uncorrelated ($C_{12} = C_{21} = 0$). Then, they are independent.*

## 2.3 Conditional Expectation

$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $\int_A X d\mathbb{P} := \int \mathbb{1}_A X d\mathbb{P} = \int \mathbb{1}_A(\omega) X(\omega) d\mathbb{P}(\omega)$.

*Lemma.* $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$ for all $A \in \mathcal{F}$ if and only if $X = Y$ a.s.

*Proof.* $A_n = \{\omega : X(\omega) - Y(\omega) > \frac{1}{n}\}, B_n = \{\omega : X(\omega) - Y(\omega) < -\frac{1}{n}\}$. Then,

$$0 = \int_{A_n} (X - Y) d\mathbb{P} \geq \int_{A_n} \frac{1}{n} d\mathbb{P} = \frac{1}{n} \int \mathbb{1}_{A_n} d\mathbb{P} = \frac{1}{n} \mathbb{P}(A_n)$$

Thus, $\mathbb{P}(A_n) = 0$ $\forall n$. Similarly, $\mathbb{P}(B_n) = 0$. Now, $\{\omega : X(\omega) \neq Y(\omega)\} = (\bigcup_{n=1}^\infty A_n) \cup (\bigcup_{n=1}^\infty B_n) \Rightarrow$ measure 0. $\square$

*Intuition.* $(\Omega, \mathcal{F}, \mathbb{P})$ is given, $X : \mathcal{F}$-measurable random variable, $\mathcal{G} \subseteq \mathcal{F}$ is a sub $\sigma$-algebra. If we know nothing, then we expect $X$ as $\mathbb{E}[X]$. If we know $\mathcal{F}$, then we expect $X$ as $X$. Now, if we know $\mathcal{G}$, then we expect $X$ as $\mathbb{E}[X|\mathcal{G}]$ (what is it?)

**Definition 2.3.1** (Conditional Expectation)
$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $X \in L^1(\mathbb{P})$ is a random variable. $\mathcal{G}$ is a sub $\sigma$-algebra of $\mathcal{F}$. We define $\mathbb{E}[X|\mathcal{G}]$ as

1. $\mathcal{G}$-measurable random variable
2. $\int_A \mathbb{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega)$.

*Question.* $\mathbb{E}[X|\mathcal{G}]$ exists? (Yes! proof skip). unique? (Yes! up to a.s.)

*Remark.* Lemma implies determine $X$ (a.s.) is equivalent to know $\int_A X d\mathbb{P} \ \ \forall A \in \mathcal{F}$.
In this sense, conditional expectation $Y = \mathbb{E}[X|\mathcal{G}]$ is knowing $\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \ \ \forall A \in \mathcal{G}$.

*Example.* Toss a coin three times.
$\mathcal{F}(0) \subseteq \mathcal{F}(1) \subseteq \mathcal{F}(2) \subseteq \mathcal{F}(3)$. $X(t)$ is a number of heads until $t$ times; $X(t)$ is $\mathcal{F}(t)$-measurable. If $\mathcal{F}(1) = \{\phi, \Omega, A_H, A_T\}$, then $\mathbb{E}[X(2)|\mathcal{F}(1)] = X(1) + \frac{1}{2}$, since we know the information of 1st flip.

*Proof.* Want: $\int_A (X(1) + \frac{1}{2}) d\mathbb{P} = \int_A X(2) d\mathbb{P}$ for all $A \in \mathcal{F}(1)$ (*c.f.* $\mathbb{P}(\omega) = \frac{1}{8} \ \forall \omega \in \Omega$).
For $A = A_H$, $\int \mathbb{1}_{A_H}(\omega)(X(1)(\omega) + \frac{1}{2}) d\mathbb{P}(\omega) = \frac{3}{2}\mathbb{P}(A_H) = \frac{3}{4}$.
$\int \mathbb{1}_{A_H}(\omega)(X(2))(\omega) d\mathbb{P}(\omega) = \sum_{\omega \in A_H} (X(2))(\omega)\mathbb{P}(\omega) = \frac{1}{8}(2 + 2 + 1 + 1) = \frac{3}{4}$.
□

*Remark.* $\mathcal{G} = \sigma(Y)$; $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|\sigma(Y)] := \mathbb{E}[X|Y]$

**Theorem 2.3.2**
*$X, Y$ are independent random variable in $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}$ is a sub $\sigma$-algebra of $\mathcal{F}$.*

1. $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$
2. *$X$ is $\mathcal{G}$-measurable. Then,* $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$.
3. *$\mathcal{H}$ is a sub $\sigma$-algebra of $\mathcal{G}$. Then,* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.
4. *$X, \mathcal{G}$ are independent, then* $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

*Proof.*
1. Exercise
2. We only need to show that $X \geq 0, Y \geq 0$ implies 2.

(a) $X = \mathbb{1}_B$
Want: $\mathbb{E}[\mathbb{1}_B Y|\mathcal{G}] = \mathbb{1}_B \mathbb{E}[Y|\mathcal{G}]$ for $B \in \mathcal{G}$.

(b) $X = \sum_{i=1}^n \alpha + i\mathbb{1}_{B_i}$
Use linearity.

(c) $X \geq 0$
Use MCT

3. Want: $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$
Let $A \in \mathcal{H}$. Then, $\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}](\omega)d\mathbb{P}(\omega) = \int_A \mathbb{E}[X|\mathcal{G}]d\mathbb{P} = \int_A Xd\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{H}]d\mathbb{P}$

4. Can be shown similarly as in 2. Check $X = \mathbb{1}_B$ case. (Hint: $A \in \mathcal{G} \Rightarrow A, B$ are independent.)

□

*Example* (Revisit).

$$\begin{aligned}
\mathbb{E}[X(2)|\mathcal{F}(1)] &= \mathbb{E}[X(2) - X(1) + X(1)|\mathcal{F}(1)] \\
&= \mathbb{E}[X(2) - X(1)|\mathcal{F}(1)] + X(1) \\
&= \mathbb{E}[X(2) - X(1)] + X(1) \\
&= \frac{1}{2} + X(1)
\end{aligned}$$

# Appendices

# Appendix A

# TA Session

*Example* (1.2.2). Let $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P})$ be the independent, infinite coin-toss space. Define stock price by

$$S_0(\omega) = 4 \quad \text{for all } \omega \in \Omega_\infty$$

$$S_1(\omega) = \begin{cases} 8 & \text{if } \omega_1 = H \\ 2 & \text{if } \omega_1 = T \end{cases}$$

$$S_2(\omega) = \begin{cases} 16 & \text{if } \omega_1 = \omega_2 = H \\ 4 & \text{if } \omega_1 \neq \omega_2 \\ 1 & \text{if } \omega_1 = \omega_2 = T \end{cases}$$

and in general

$$S_{n+1}(\omega) = \begin{cases} 2S_n(\omega) & \text{if } \omega_{n+1} = H \\ \frac{1}{2}S_n(\omega) & \text{if } \omega_{n+1} = T \end{cases}$$

Then, $S_0, S_1, \cdots$, are random variable.
For example, $\mathbb{P}(S_2 = 4) = \mathbb{P}(A_{HT} \cup A_{TH}) = 2pq$

*Example* (2.2.2). Let $\Omega$ be a three independent coin-toss space. Stock price random variables $S_0, S_1, \cdots$, are the same as the previous example. Let the probability measure $\mathbb{P}$ be given by

$$\mathbb{P}(HHH) = p^3, \mathbb{P}(HHT) = p^2q, \cdots, \mathbb{P}(TTT) = q^3.$$

Assume $0 < p < 1$. Then, the random variables $S_2$ and $S_3$ are not independent.
$\because$ Consider the sets $\{S_3 = 32\} = \{HHH\}$ and $\{S_2 = 16\} = \{HHH, HHT\}$ whose probabilities are $\mathbb{P}(S_3 = 32) = p^3$ and $\mathbb{P}(S_2 = 16) = p^2$. In order to have Independence, $p^3 = \mathbb{P}(S_3 = 32) = \mathbb{P}(S_2 = 16 \text{ and } S_3 = 32) = \mathbb{P}(S_2 = 16)\mathbb{P}(S_3 = 32) = p^5 \Rightarrow\Leftarrow$.

The random variables $S_2$ and $S_3/S_2$ are independent. The $\sigma$-algebra generated by $S_2$ comprises $\phi, \Omega$, the atoms
$\{S_2 = 16\} = \{HHH, HHT\}, \{S_2 = 4\} = \{HTH, HTT, THH, THT\}, \{S_2 = 1\} = \{TTH, TTH\}$, and their unions.

The $\sigma$-algebra generated by $S_3/S_2$ comprises $\phi, \Omega$ and
$\{S_3/S_2 = 2\} = \{HHH, HTH, THH, TTH\}, \{S_3/S_2 = \frac{1}{2}\} = \{HHT, HTT, THT, TTT\}$

For $A \in \sigma(S_2), B \in \sigma(S_3/S_2)$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
ex) $p^3 = \mathbb{P}(S_2 = 16 \text{ and } S_3/S_2 = 2\} = \mathbb{P}(S_2 = 16)\mathbb{P}(S_3/S_2 = 2)) = p^2 p = p^3$

*Example* (2.2.10 Uncorrelated, dependent normal random variables).
Let $X, Z$ be random variable satisfying

$$X: \text{ standard normal random variable}$$
$$Z: \text{ independent of } X,\ \mathbb{P}(Z = 1) = \frac{1}{2},\ \mathbb{P}(Z = -1) = \frac{1}{2}$$

Define $Y = ZX$. Show

1. $Y$ is standard normal random variable
2. $X$ and $Y$ are uncorrelated but they are dependent.

*Proof.*

1. $$\begin{aligned} F_Y(b) &= \mathbb{P}(Y \le b) \\ &= \mathbb{P}(Y \le b \text{ and } Z = 1) + \mathbb{P}(Y \le b \text{ and } Z = -1) \\ &= \mathbb{P}(X \le b \text{ and } Z = 1) + \mathbb{P}(X \le b \text{ and } Z = -1) \\ &= \mathbb{P}(X \le b)\mathbb{P}(Z = 1) + \mathbb{P}(X \le b)\mathbb{P}(Z = -1) \\ &= \frac{1}{2}N(b) + \frac{1}{2}N(b) \\ &= N(b) \end{aligned}$$
2. Since $\mathbb{E}X = \mathbb{E}Y = 0$,
$$Cov(X, Y) = \mathbb{E}[XY] = \mathbb{E}[ZX^2] = \mathbb{E}[Z]\mathbb{E}[X^2] = 0$$
$\therefore X$ and $Y$ are uncorrelated.
If $X$ and $Y$ are independent, $|X|$ and $|Y|$ are independent. But $\mathbb{P}(|X| \le 1, |Y| \le 1) = \mathbb{P}(|X| \le 1) = N(1) - N(-1)$, and $\mathbb{P}(|X| \le 1, |Y| \le 1) = \mathbb{P}(|X| \le 1)\mathbb{P}(|Y| \le 1) = (N(1) - N(-1))^2 \Rightarrow\Leftarrow$

$\square$

Let $\mu_{X,Y}$ be a joint distribution measure of $(X,Y)$. Since $|X| = |Y|$, $(X,Y)$ takes values only in the set $C = \{(x,y) : x = \pm y\}$.
It follows that for any measurable function $f$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{1}_C(x,y) f_{X,Y}(x,y)dydx = 0$$

$\therefore$ There is no joint density $f_{X,Y}$ for $(X,Y)$.

$$\begin{aligned}
F_{X,Y}(a,b) &= \mathbb{P}(X \le a, Y \le b) \\
&= \mathbb{P}(X \le a, X \le b, Z = 1) + \mathbb{P}(X \le a, -X \le b, Z = -1) \\
&= \frac{1}{2}\mathbb{P}(X \le a \wedge b) + \frac{1}{2}\mathbb{P}(-b \le X \le a) \\
&= \frac{1}{2}N(A \wedge b) + \frac{1}{2}((N(a) - N(-b)) \vee 0)
\end{aligned}$$

*Example* (2.2.12). Let $(X,Y)$ be jointly normal with the density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right)$$

Define $W = Y - \frac{\rho\sigma_2}{\sigma_1}X$. Then, $X$ and $W$ are independent.
Note that linear combination of jointly normal random variables are jointly normal (i.e., $(X,W)$ is jointly normal).
Thus it suffices to show that $Cov(X,W) = 0$ (by Thm 2.2.9)

$$\begin{aligned}
Cov(X,W) &= \mathbb{E}[(X - \mathbb{E}X)(W - \mathbb{E}W)] \\
&= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] - \mathbb{E}[\frac{\rho\sigma_2}{\sigma_1}(X - \mathbb{E}X)^2] \\
&= Cov(X,Y) - \frac{\rho\sigma_2}{\sigma_1}\sigma_1^2 \\
&= 0
\end{aligned}$$

Let $f_{X,W}$ be joint density of $X$ and $W$.

$$\begin{aligned}
\mathbb{E}[W] &= \mu_2 - \frac{\rho\sigma_2\mu_1}{\sigma_1} =: \mu_3 \\
\mathbb{E}[(W - \mathbb{E}W)^2] &= \mathbb{E}[(Y - \mathbb{E}Y)^2] - \frac{2\rho\sigma_2}{\sigma_1}\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] + \frac{\rho^2\sigma_2^2}{\sigma_1^2}\mathbb{E}[(X - \mathbb{E}X)^2] \\
&= \sigma^2 - \frac{2\rho\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2 + \frac{\rho^2\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= (1-\rho^2)\sigma_2^2 =: \sigma_3^2
\end{aligned}$$

$\therefore f_{X,W}(x,w) = \frac{1}{2\pi\sigma_1\sigma_3} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(w-\mu_3)^2}{2\sigma_3^2}\right)$.
Note that we have decomposed $Y$ into the linear combination $Y = \frac{\rho\sigma_2}{\sigma_1}X + W$ of a pair of independent normal random variables $X$ and $W$.

*Example* (2.3.3)*.* Let $\mathcal{G} = \sigma(X)$. Observe estimate $Y$ based on $X$ and error.

$$\mathbb{E}[Y|X] = \frac{\rho\sigma_2}{\sigma_1} + \mathbb{E}[W] = \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1) + \mu_2.$$
$$Y - \mathbb{E}[Y|X] = W - \mathbb{E}[W]$$

Note that the error is random variable with expected value zero and independent of the estimation $\mathbb{E}[Y|X]$.