

NCKU 112.1  
Probability

Eric Liu

# CONTENTS

## CHAPTER 0 PROBABILITY SPACE PAGE 2

- 0.1 Basic Definition for Background Space 2
- 0.2 Binomial Asset Pricing Model 8

## CHAPTER 1 DISCRETE RANDOM VARIABLE PAGE 11

- 1.1 Basic Discrete Random Variable and Their Expected Values and Variance 11

## CHAPTER 2 CONTINUOUS RANDOM VARIABLE PAGE 21

- 2.1 Basic Knowledge 21
- 2.2 Law of Unconscious Statistician 26
- 2.3 Uniform, Exponential and Normal 27
- 2.4 Exercises 35
- 2.5 Central Limit Theorem 38

## CHAPTER 3 JOINT DISTRIBUTION PAGE 40

- 3.1 Independence of Event 40
- 3.2 Independence of Random Variable 43
- 3.3 Bayes' Theorem 45
- 3.4 Independent Disjoint Grouping Theorem 47

# Chapter 0

## Probability Space

### 0.1 Basic Definition for Background Space

**Definition 0.1.1. (Definition of  $\sigma$ -Algebra)** We say  $(\Omega, \mathcal{G} \subseteq \mathcal{P}(\Omega))$  is a  $\sigma$ -algebra if

$$\Omega \in \mathcal{G}$$

$$X \in \mathcal{G} \implies \Omega \setminus X \in \mathcal{G} \text{ (Closed under complement)}$$

$$\mathcal{A} \subseteq \mathcal{G} \text{ and } |\mathcal{A}| \leq |\mathbb{N}| \implies \bigcup \mathcal{A} \in \mathcal{G} \text{ (Closed under countable union)}$$

From now, we denote  $\Omega \setminus X$  by  $X^c$ . We say  $\mathcal{G} = \{\emptyset, \Omega\}$  is the trivial  $\sigma$ -algebra on  $\Omega$ .

**Theorem 0.1.2. (Basic Property of  $\sigma$ -Algebra)** Let  $(\Omega, \mathcal{G})$  be a  $\sigma$ -algebra. Then we have

$$\emptyset \in \mathcal{G}$$

$$\mathcal{A} \subseteq \mathcal{G} \implies \bigcap \mathcal{A} \in \mathcal{G}$$

$$A, B \in \mathcal{G} \implies A \setminus B \in \mathcal{G}$$

*Proof.* Observe  $\emptyset = \Omega^c$ , and observe  $\bigcap \mathcal{A} = (\bigcup_{X \in \mathcal{A}} X^c)^c$ , and observe  $A \setminus B = A \cap B^c$  ■

**Lemma 0.1.3. (Intersection of  $\sigma$ -Algebras is a  $\sigma$ -Algebra)** Let  $S$  be a set of  $\sigma$ -algebra over  $\Omega$ , then  $\bigcap S$  is a  $\sigma$ -algebra.

*Proof.*  $\Omega \in \bigcap S$  is trivial. Observe  $A \in \bigcap S \implies \forall \mathcal{G} \in S, A \in \mathcal{G} \implies \forall \mathcal{G} \in S, A^c \in \mathcal{G} \implies A^c \in \bigcap S$ . Observe  $\mathcal{A} \subseteq \bigcap S$  and  $|\mathcal{A}| \leq |\mathbb{N}| \implies \forall \mathcal{G} \in S, \mathcal{A} \subseteq \mathcal{G} \implies \bigcup \mathcal{A} \in \mathcal{G} \implies \bigcup \mathcal{A} \in \bigcap S$ . ■

**Definition 0.1.4. (Definition of Generating  $\sigma$ -Algebra)** Let  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ . The  $\sigma$ -algebra generated by  $\mathcal{F}$  is defined to be the smallest  $\sigma$ -algebra that contain  $\mathcal{F}$

We have now gave enough tools to generate  $\sigma$ -algebra.

The f

**Definition 0.1.5. (Definition of Measure)** Let  $\mathcal{G}$  be a  $\sigma$ -algebra over  $\Omega$ . Function  $\mu : \mathcal{G} \rightarrow \mathbb{R}$  is called a measure if

$$\forall E \in \mathcal{G}, \mu(E) \geq 0 \text{ (Nonnegative)}$$

$$\mu(\emptyset) = 0$$

$$\mathcal{F} \subseteq \mathcal{G} \text{ and } |\mathcal{F}| \leq |\mathbb{N}| \text{ and } \mathcal{F} \text{ is disjoint} \implies \mu\left(\bigcup \mathcal{F}\right) = \sum_{X \in \mathcal{F}} \mu(X) \text{ (}\sigma \text{ additivity)}$$

**Theorem 0.1.6. (Intended Property of Measure)** We have

$$A \subseteq B \in \mathcal{F} \implies \mu(A) \leq \mu(B)$$

*Proof.* Observe  $\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \implies \mu(B) - \mu(A) = \mu(B \setminus A) \geq 0$  ■

**Definition 0.1.7. (Definition of Probability Measure)** A probability measure  $P$  is a measure that satisfy  $P(\Omega) = 1$

**Definition 0.1.8. (Definition of Probability Space)** A probability space is a triple  $(\Omega, \mathcal{G}, P)$  where

$\Omega$  is a set called *sample space*

$\mathcal{G}$  is a  $\sigma$ -algebra over  $\Omega$  called event space

$P : \mathcal{G} \rightarrow [0, 1]$  is a measure called probability measure

Basically, probability measure "measure" the probability an event will happen. Normally we don't measure the probability a sample will happen, since it is undoable in some cases, for instance, shooting a dart to an infinite set. We only measure the probability that the outcome sample is in a set of certain samples, where the assignment of measure is outside of the system.

A simple example of a  $\sigma$ -algebra is

$$\Omega_2 = \{HH, HT, TH, TT\}, \mathcal{G} = \{\emptyset, X, \{HT, HH\}, \{TH, TT\}\}$$

Notice in this example,  $\Omega$  is ought to be interpreted as tossing two coins and  $\mathcal{G}$  is to observe the first coin is head or tail.

To expand the first example, we have another simple example:

$$\Omega_3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Define

$$A_H := \{HHH, HHT, HTH, HTT\} \text{ and } A_T := \{THH, THT, TTH, TTT\}$$

which is the information of tossing head or tail on first try.

Notice  $A_T = A_H^c$ . Define

$$A_{HH} := \{HHH, HHT\} \text{ and } A_{HT} := \{HTH, HTT\}$$

$$A_{TH} := \{THH, THT\} \text{ and } A_{TT} := \{TTH, TTT\}$$

so we have

$$A_H = A_{HH} \cup A_{HT} \text{ and } A_T = A_{TH} \cup A_{TT}$$

Then we can define

$$\mathcal{G} := \left\{ \bigcup N : N \in \mathcal{P}(M) \right\}$$

where  $M = \{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$

Notice we can define four  $\sigma$ -algebras by

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \{\emptyset, \Omega, A_T, A_H\}, \mathcal{F}_2 = \mathcal{G}, \mathcal{F}_3 = \mathcal{P}(\Omega)$$

then we have

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$$

The following concern Borel  $\sigma$ -algebra

**Definition 0.1.9. (Definition of a Random Variable)** We say the function from  $\Omega$  to  $\mathbb{R}$  is a random variable, and often a random variable is defined on a fixed probability space, since the usage of random variable are mostly associated with a probability measure.

We now define 3 random variable for example from the last example of  $\sigma$ -algebra.

Let  $S_0, u, d \in \mathbb{R}^+$  and let  $d < 1 < u$ . We define three random variables  $S_1, S_2, S_3$  on  $\Omega_3$

$$S_1(\omega) = \begin{cases} uS_0 & \text{if } \omega \in A_H \\ dS_0 & \text{if } \omega \in A_T \end{cases} \quad S_2(\omega) = \begin{cases} u^2S_0 & \text{if } \omega \in A_{HH} \\ udS_0 & \text{if } \omega \in A_{HT} \cup A_{TH} \\ d^2S_0 & \text{if } \omega \in A_{TT} \end{cases}$$

$$S_3(\omega) = \begin{cases} u^3S_0 & \text{if } \omega \in \{HHH\} \\ u^2dS_0 & \text{if } \omega \in \{HHT, HTH, THH\} \\ ud^2S_0 & \text{if } \omega \in \{HTT, THT, TTH\} \\ d^3S_0 & \text{if } \omega \in \{TTT\} \end{cases}$$

Often, we just use  $S$  to denote  $S(\omega)$ .

**Definition 0.1.10. (Definition of Borel-Algebra)** The Borel-Algebra on  $\mathbb{R}$ , which we denote  $\mathcal{B}(\mathbb{R})$  is the  $\sigma$ -algebra generated by all open interval of  $\mathbb{R}$ .

Some members of  $\mathcal{B}(\mathbb{R})$  :

$$\begin{aligned} (b, a), (a, \infty), \mathbb{R} \\ (b, a] &= (b, \infty) \setminus (a, \infty) \\ [a, \infty) &= \mathbb{R} \setminus (-\infty, a) \\ [a, b] &= [a, \infty) \setminus (b, \infty) \\ \{a\} &= \mathbb{R} \setminus (-\infty, a) \cup (a, \infty) \end{aligned}$$

**Theorem 0.1.11. (Construct  $\sigma$ -Algebra with Random Variable)** Let  $X$  be a random variable on  $\Omega$ . We define

$$X^{-1}[B] = \{\omega \in \Omega : X(\omega) \in B\}$$

and define the  $\sigma$ -algebra  $\sigma(X)$  by

$$\sigma(X) = \{X^{-1}[B] : B \in \mathcal{B}(\mathbb{R})\}$$

We can verify  $\sigma(X)$  is a  $\sigma$ -algebra.

*Proof.* Notice  $\Omega = X^{-1}[\mathbb{R}] \in \sigma(X)$ . Observe  $(X^{-1}[B])^c = X^{-1}[B^c] \in \sigma(X)$ . Let  $C = \{X^{-1}[B] : B \in \mathcal{C}\} \subseteq \sigma$  be countable. Observe  $\bigcup_{B \in C} X^{-1}[B] = X^{-1}[\bigcup_{B \in C} B] \in \sigma(X)$  ■

Notice  $\sigma(S_1) = \mathcal{F}_1, \sigma(S_2) \neq \mathcal{F}_2, \sigma(S_3) \neq \mathcal{F}_3$ .

Normally, we write  $P(X = x_i)$  to denote  $P(X^{-1}[x_i])$

**Theorem 0.1.12. (Induce Measure on Borel Algebra by a Random Variable and a Probability Space)** Let  $(\Omega, \mathcal{F}, P : \mathcal{F} \rightarrow \mathbb{R})$  be a probability space and  $X : \Omega \rightarrow \mathbb{R}$  be a random variable. We can induce a measure on  $\mathcal{B}(\mathbb{R})$  (we say this measure is induced by  $X$  and  $P$ )

$$\mathcal{L}_X : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}, B \mapsto P(X^{-1}[B])$$

*Proof.* Because  $P$  is nonnegative, we know  $\mathcal{L}_X$  is nonnegative. Notice that  $X^{-1}[\emptyset] = \emptyset$ , so we know  $\mathcal{L}_X(\emptyset) = P(\emptyset) = 0$ . Let  $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R})$  and  $\mathcal{A}$  be countable and disjoint. We have

$$\begin{aligned} \mathcal{L}_X\left(\bigcup \mathcal{A}\right) &= P(X^{-1}[\bigcup \mathcal{A}]) \\ &= P\left(\bigcup \mathcal{C}\right) \text{ where } \mathcal{C} = \{X^{-1}[a] : a \in \mathcal{A}\} \\ &= \sum_{c \in \mathcal{C}} P(c) \\ &= \sum_{a \in \mathcal{A}} P(X^{-1}[a]) = \sum_{a \in \mathcal{A}} \mathcal{L}_X(a) \end{aligned}$$

Notice that  $\mathcal{C}$  is countable because  $\mathcal{A}$  is countable, and  $\mathcal{C}$  is disjoint because  $x \in X^{-1}[a] \cap X^{-1}[b] \implies X(x) \in a \cap b$ . ■

**Definition 0.1.13. (Definition of Cumulative Distribution Function)** We say  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  is the cumulative distribution function of  $X$  if

$$F_X(x) = \mathcal{L}_X((-\infty, x]) = P(X^{-1}[(-\infty, x]])$$

Distribution function is a function on the image of a random variable  $X$ , that map a real number to the probability the random variable  $X$  is the real number  $x_1$ . How? Probability is the measure of the set that can be mapped to  $x_1$ .

In a more formal language, the random variable doesn't specify the probability of each event, but the one does, probability measure is often implicitly given by the random variable.

We now use the random variable  $S_3 : \Omega \rightarrow \mathbb{R}$  on  $(\Omega, \mathcal{F}, P : \mathcal{P}(\Omega) \rightarrow \mathbb{R})$  to serve as an example of induced measure and cumulative distribution function.

Notice that the range of  $S_3$  is  $\{d^3S_0 < d^2uS_0 < du^2S_0 < u^3S_0\}$ . We have

$$F_{S_3}(d^3S_0) = \mathcal{L}_{S_3}((-\infty, d^3S_0]) = P(S_3^{-1}[(-\infty, d^3S_0)]) = P(\{TTT\})$$

$$F_{S_3}(d^2uS_0) = \mathcal{L}_{S_3}((-\infty, d^2uS_0]) = P(S_3^{-1}[(-\infty, d^2uS_0)]) = P(\{TTT, TTH, THT, HTT\})$$

$$F_{S_3}(du^2S_0) = \mathcal{L}_{S_3}((-\infty, du^2S_0]) = P(S_3^{-1}[(-\infty, du^2S_0)]) = P(\Omega \setminus \{HHH\})$$

$$F_{S_3}(u^3S_0) = \mathcal{L}_{S_3}((-\infty, u^3S_0]) = P(S_3^{-1}[(-\infty, u^3S_0)]) = P(\Omega)$$

With above deduction, we can precisely deduce

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3S_0 \\ P(\{TTT\}) & \text{if } d^3S_0 \leq x < d^2uS_0 \\ P(\{TTT, TTH, THT, HTT\}) & \text{if } d^2uS_0 \leq x < du^2S_0 \\ P(\Omega \setminus \{HHH\}) & \text{if } du^2S_0 \leq x < u^3S_0 \\ P(\Omega) & \text{if } u^3S_0 \leq x \end{cases}$$

Notice that we can define measure  $P$  in multiple ways, and then we will have different induced measures  $\mathcal{L}_{S_3}$  and different cumulative distribution function  $F_{S_3}(x)$ . For instance, one can check that  $P : \mathcal{P}(\Omega) = \mathcal{F}_3 \rightarrow \mathbb{R}, X \mapsto \frac{|X|}{8}$  is a measure, and  $P : \mathcal{F}_3 \rightarrow \mathbb{R}, X \mapsto \sum_{x \in X} f(x)$  where  $f$  is defined by  $\omega_1\omega_2\omega_3 \mapsto \frac{2^i}{3^3} : i = |\{\omega_j : \omega_j = T, j = 1, 2, 3\}|$  are all measures, but the former is

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3S_0 \\ \frac{1}{8} & \text{if } d^3S_0 \leq x < d^2uS_0 \\ \frac{1}{2} & \text{if } d^2uS_0 \leq x < du^2S_0 \\ \frac{7}{8} & \text{if } du^2S_0 \leq x < u^3S_0 \\ 1 & \text{if } u^3S_0 \leq x \end{cases}$$

and the latter is

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3S_0 \\ \frac{8}{27} & \text{if } d^3S_0 \leq x < d^2uS_0 \\ \frac{8}{27} + \binom{3}{1}\frac{4}{27} & \text{if } d^2uS_0 \leq x < du^2S_0 \\ \frac{8}{27} + \binom{3}{1}\frac{4}{27} + \binom{3}{2}\frac{2}{27} & \text{if } du^2S_0 \leq x < u^3S_0 \\ 1 & \text{if } u^3S_0 \leq x \end{cases}$$



## 0.2 Binomial Asset Pricing Model

**Definition 0.2.1. (Definition of Stochastic Process)** Given a common probability space  $(\Omega, \mathcal{F}, P : \mathcal{F} \rightarrow \mathbb{R})$ , we say an indexed family of random variables defined on our given probability space is a stochastic process, denoted

$$\{X_t : t \in \Lambda, \omega \in \Omega\}$$

where the index set  $\Lambda$  is called time horizon. So, if someone say  $X$  is a stochastic process, we can say  $X_t$  is a random variable, and we can say  $X(t, \omega) = X_t(\omega) \in \mathbb{R}$ .

Additionally, we say  $\Omega$  is "background space" and  $\Lambda \times \mathbb{R}$ , where we can draw a trajectory to record the sequence of outcome, is "foreground space" or "configuration space", and we say  $\Theta = \{X(t, \omega) : t \in \Lambda, \omega \in \Omega\}$  is the "state space" that contains "state of process"

The "frequency" of a state of the process  $X(t, \omega)$  is ?

We now give an example of stochastic process. Let  $\Omega = \{H, T\}$  and  $\Lambda = [0, 2]$ . The stochastic process if the indexed set  $\{X_t : t \in \Lambda\}$  of random variables specified by

$$X_t(H) = \sin(2\pi t) \text{ and } X_t(T) = \sin(4\pi t)$$

Then

$$\Theta = [-1, 1]$$

Another example is: Let  $\Lambda = \{1, 2, 3, 4, 5\}$  and let  $\Omega = \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) : \forall i \in \Lambda, \omega_i \in \{H, T\}\}$ . The configuration space is then  $\{1, 2, 3, 4, 5\} \times \mathbb{R}$ , and we can specify the stochastic process  $\{S_t : t \in \Lambda\}$  by

$$S_t((\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)) = u^i d^{t-i} S_0 : i = |\{\omega_j : 1 \leq j \leq t, \omega_j = H\}|$$

Then

$$\Theta = \{u^i d^j S_0 : i, j \in \mathbb{N} \cup \{0\}, 1 \leq i + j \leq 5\}$$

To buy an option from someone means acquiring the right, but not the obligation, to purchase a stock at a predetermined price,  $K$ . This comes with an associated cost called the “risk premium” or  $V_0$ , which is paid at the time of the transaction.

Notice when we say net earn or profit, we are comparing his money with and without making the deal.

If the stock ends up at price  $S_1(\omega)$  lower than  $K$ , the buyer won’t buy the stock at the price  $K$ , so for seller, he profit  $V_0$ .

If the stock ends up at price  $S_1(\omega)$  higher than  $K$ , then the seller net earn:  
 $-(S_1(\omega) - K) + V_0$

Now, we consider the profit we can make for depositing money into bank and get interest. The interest rate is  $r$ .

If the stock ends up at price  $S_1(\omega)$  lower than  $K$ , the buyer won’t buy the stock at the price  $K$ , so for seller, he profit  $(1 + r)V_0$ .

If the stock ends up at price  $S_1(\omega)$  higher than  $K$ , then the seller net earn:  
 $-(S_1(\omega) - K) + (1 + r)V_0$

Now, we wish to lower the risk so we buy  $\Delta_0$  share of the stock we short at price  $S_0$  to form a portfolio.

If the stock ends up at price  $S_1(\omega)$  lower than  $K$ , the buyer won’t buy the stock at the price  $K$ , so for seller, he profit  $(1 + r)V_0 + \Delta_0(S_1(\omega) - (1 + r)S_0)$ .

If the stock ends up at price  $S_1(\omega)$  higher than  $K$ , then the seller net earn:  
 $-(S_1(\omega) - K) + (1 + r)V_0 + \Delta_0(S_1(\omega) - (1 + r)S_0)$

From now, we denote the value of the option at time  $t$

$$V_t(\omega) := \max\{S_t(\omega) - K, 0\}$$

So we can say the value of our portfolio  $X_t(\omega)$ , our action of selling the option and buy the stock, at time 1 is

$$X_1(\omega) = -V_1(\omega) + (1+r)V_0 + \Delta_0(S_1(\omega) - (1+r)S_0)$$

Now we wish, for all  $\omega$ , we have  $X_1(\omega) \geq 0$ , by negotiating with the buyer on  $V_0$  and deciding how many share  $\Delta_0$  we want to buy.

Notice that if we negotiate  $V_0$  at

$$V_0 := \frac{1}{1+r} \left[ \frac{1+r-d}{u-d} V_1(H) + \frac{u-(1+r)}{u-d} V_1(T) \right]$$

and buy share  $\Delta_0$

$$\Delta_0 := \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}$$

We see

$$\begin{aligned} X_1(H) = -V_1(H) + (1+r) \frac{1}{1+r} \left[ \frac{1+r-d}{u-d} V_1(H) + \frac{u-(1+r)}{u-d} V_1(T) \right] \\ + \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} (S_1(H) - (1+r)S_0) \end{aligned}$$

equals 0 by substituting  $S_1(H)$  and  $S_1(T)$  with  $S_0u$  and  $S_0d$

# Chapter 1

## Discrete Random Variable

### 1.1 Basic Discrete Random Variable and Their Expected Values and Variance

**Definition 1.1.1. (Definition of Discrete Random Variable and Probability Mass Function)** Let  $(\Omega, \mathcal{F}, P : \mathcal{F} \rightarrow \mathbb{R})$  be a probability space. We say a random variable  $X : \Omega \rightarrow \mathbb{R}$  is a discrete random variable, if

$$X[\Omega] \text{ if countable}$$

and we define the probability mass function  $p_X$  as

$$p_X(r) = P(X = r)$$

In this section,  $X$  is always a discrete random variable.

**Theorem 1.1.2. (Property of Probability Mass Function)** Let  $X[\Omega] = \{x_1, \dots\}$ . We have

$$P(a \leq X \leq b) = \sum_{x_i \in [a, b]} p_X(x_i)$$

This reads: The probability  $X$  is in  $[a, b]$  equals the sum of the probability  $X$  is a number in  $[a, b]$

**Theorem 1.1.3. (Property of Probability Mass Function)** We have

$$\sum_{x_i \in \mathbb{R}} p_X(x_i) = 1$$

Now we introduce expected value.

**Definition 1.1.4. (Definition of Expected Value)** We say the expected value of  $X$  is

$$E(X) = \sum x_i p(x_i)$$

If possible, normally we write

$$\mu = E(X)$$

**Theorem 1.1.5. (Expected Value)** For function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have

$$E(g \circ X) = \sum g(x_i) p(x_i)$$

**Theorem 1.1.6. (Linearity of Expected Value of Discrete Random Variable)** We have

$$E(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 E(X_1) + \alpha_2 E(X_2)$$

We now introduce variance.

**Definition 1.1.7. (Definition of Variance)** We define variance of  $X$  to be

$$\text{Var}(X) := E((X - \mu)^2)$$

If possible, normally we write

$$\sigma(X) := \sqrt{\text{Var}(X)}$$

**Theorem 1.1.8. (Calculation of Variance)** We have

$$\text{Var}(X) = E(X^2) - \mu^2$$

*Proof.*

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

■

**Theorem 1.1.9. (Variance)** We have

$$\text{Var}[\alpha_1 X_1] = \alpha_1^2 \text{Var}[X_1]$$

*Proof.*

$$\begin{aligned} \text{Var}[\alpha_1 X_1] &= E[(\alpha_1 X_1)^2] - E[\alpha_1 X_1]^2 \\ &= \alpha_1^2 E[X_1^2] - \alpha_1^2 E[X_1]^2 \\ &= \alpha_1^2 (E[X_1^2] - E[X_1]^2) = \alpha_1^2 \text{Var}[X_1] \end{aligned}$$

■

We now introduce Bernoulli random variable.

**Definition 1.1.10. (Definition of Bernoulli Random Variable)** Let  $A$  be an event and  $P(A) = p$ . We define

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

**Theorem 1.1.11. (Properties of Bernoulli Random Variable)** We have

$$P(I_A = 1) = p \text{ and } P(I_A = 0) = 1 - p$$

$$E[I_A] = p$$

$$\text{Var}[I_A] = p(1 - p)$$

*Proof.*

$$\text{Var}[I_A] = E[I_A^2] - E[I_A]^2 = 1^2p - p^2 = p - p^2 = p(1 - p)$$

■

**Definition 1.1.12. (Definition of Binomial Random Variable)** We say the binomial random variable  $B_{n,p}$  is the number of success in  $n$  independent trials, each of which is a success with probability  $p$ . We have

$$P(B_{n,p} = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

**Theorem 1.1.13. (Properties of Binomial Random Variable)** We have

$$E[B_{n,p}] = np \text{ and } \text{Var}[B_{n,p}] = np(1 - p)$$

*Proof.*

$$E[B_{n,p}] = E\left[\sum I_A\right] = \sum E[I_A] = np$$

That of variance is left unproved.

■

Now we introduce the notion of Poisson random variable.

**Definition 1.1.14. (Definition of Poisson Random Variable)** For  $\lambda > 0$ , we define Poisson random variable  $X_\lambda$  by

$$P(X_\lambda = i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

**Theorem 1.1.15. (Verification of Poisson Random Variable)** Let  $X$  be a Poisson random variable with parameter  $\lambda$ . We verify

$$\sum_{k=0}^{\infty} P(X = k) = 1$$

*Proof.*

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

■

**Theorem 1.1.16. (Expectation Value of Poisson Random Variable)** Let  $X$  be a Poisson random variable with parameter  $\lambda$ . We have

$$E[X] = \lambda$$

*Proof.* Compute

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k P(x = k) \\ &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\ &= e^{\lambda} \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} \\ &= e^{\lambda} \lambda \sum_{u=0}^{\infty} \frac{\lambda^u}{u!} \\ &= \lambda e^{\lambda} e^{-\lambda} = \lambda \end{aligned}$$

■

**Theorem 1.1.17. (Variance of Poisson Random Variable)** Let  $X$  be a Poisson random variable with parameter  $\lambda$ . We have

$$\text{Var}[X] = \lambda$$

*Proof.* Compute

$$\begin{aligned}
E[X^2] &= \sum_{k=0}^{\infty} k^2 P(x = k) \\
&= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \left[ \sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] \\
&= \lambda e^{-\lambda} \left[ \sum_{k=2}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} \right] \\
&= \lambda e^{-\lambda} \left[ \sum_{k=2}^{\infty} \lambda \frac{\lambda^{k-2}}{(k-2)!} + e^{\lambda} \right] \\
&= \lambda e^{-\lambda} \left[ \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + e^{\lambda} \right] \\
&= \lambda e^{-\lambda} [\lambda e^{\lambda} + e^{\lambda}] \\
&= \lambda(\lambda + 1)
\end{aligned}$$

So we have

$$\text{Var}[X] = E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

■

Now we introduce Poisson Approximation of Binomial Random Variable

**Theorem 1.1.18. (Poisson Approximation of Binomial Random Variable)** Let  $X$  be a binomial  $B_{n,p}$  and let  $\lambda = np$ . We have

$$\lim_{n \rightarrow \infty} P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$



*Proof.* Compute

$$\begin{aligned}
P(X = k) &= \binom{n}{k} p^k (1 - p)^{n-k} \\
&= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(\frac{n(n-1) \cdots (n-k+1)}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^{-k}
\end{aligned}$$

So we have

$$\begin{aligned}
\lim_{n \rightarrow \infty} P(X = k) &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} e^{-\lambda} (1)(1) = e^{-\lambda} \frac{\lambda^k}{k!}
\end{aligned}$$

■

Now we introduce the geometric random variable.

**Definition 1.1.19. (Definition of Geometric Random Variable)** A geometric random variable counts the number of trials required for the first success in independent trials with success probability  $p$ . Let  $X$  be a geometric random variable. We have

$$P(X = k) = (1 - p)^{k-1} p$$

**Theorem 1.1.20. (Verification of Geometric Random Variable)** Let  $X$  be a geometric random variable with parameter  $p$ . We have

$$\sum_{k=1}^{\infty} P(X = k) = 1$$

*Proof.*

$$\begin{aligned}
 \sum_{k=1}^{\infty} P(X = k) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\
 &= p \sum_{k=1}^{\infty} (1-p)^{k-1} \\
 &= p \sum_{k=0}^{\infty} (1-p)^k \\
 &= p \left( \frac{1}{1-(1-p)} \right) = 1
 \end{aligned}$$

■

**Theorem 1.1.21. (Expected Value of Geometric Random Variable)** Let  $X$  be a geometric random variable with parameter  $p$ . We have

$$E[X] = \frac{1}{p}$$

*Proof.* Notice that the trials are independent. We have

$$E[X] = (p)(1) + (1-p)(E[X] + 1)$$

So we have

$$\begin{aligned}
 p(E[X]) &= p + 1 - p \\
 \implies E[X] &= \frac{1}{p}
 \end{aligned}$$

■

**Theorem 1.1.22. (Variance of Geometric Random Variable)** Let  $X$  be a geometric random variable with parameter  $p$ . We have

$$\text{Var}[X] = \frac{1-p}{p^2}$$

*Proof.* Denote  $1-p$  by  $q$ . First notice

$$\sum_{k=1}^{\infty} k q^{k-1} p = E[X] = \frac{1}{p}$$

Then deduce

$$\begin{aligned}
E[X^2] &= \sum_{k=1}^{\infty} k^2 q^{k-1} p \\
&= \sum_{k=1}^{\infty} (k-1+1)^2 q^{k-1} p \\
&= \sum_{k=1}^{\infty} (k-1)^2 q^{k-1} p + \sum_{k=1}^{\infty} 2(k-1) q^{k-1} p + \sum_{k=1}^{\infty} q^{k-1} p \\
&= q \sum_{k=1}^{\infty} (k-1)^2 q^{k-2} p + 2 \sum_{k=1}^{\infty} k q^{k-1} p - \sum_{k=1}^{\infty} q^{k-1} p \\
&= q \sum_{u=1}^{\infty} u^2 q^{u-1} p + 2E[X] - p \sum_{k=1}^{\infty} q^{k-1} \\
&= qE[X^2] + 2E[X] - \frac{p}{1-q}
\end{aligned}$$

Then we can deduce

$$p(E[X^2]) = (1-q)E[X^2] = 2E[X] - \frac{p}{1-q} = \frac{2}{p} - \frac{p}{p} = \frac{2-p}{p}$$

So in summary we have

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

■

**Theorem 1.1.23. (Geometric Random Variable)** Let  $X$  be a geometric random variable with parameter  $p$ . We have

$$P(X > n) = (1-p)^n$$

*Proof.* Let  $q = 1 - p$

$$\begin{aligned}
P(X > n) &= 1 - \sum_{k=1}^n P(X = k) \\
&= 1 - \sum_{k=1}^n q^{k-1} p \\
&= 1 - (1 - q) \sum_{k=1}^n q^{k-1} \\
&= 1 - \sum_{k=1}^n q^{k-1} + \sum_{k=1}^n q^k \\
&= 1 - \sum_{k=0}^{n-1} q^k + \sum_{k=1}^n q^k \\
&= 1 + (q^n - 1) = q^n
\end{aligned}$$

■

**Corollary 1.1.24.** We have

$$P(X > n + k \text{ while } X > n) = \frac{q^{n+k}}{q^n} = q^k$$

**Theorem 1.1.25. (Summary)** We have

A Bernoulli random variable  $I$  is a building block

Parameter  $p$  is the probability it happens

$$E[I] = p \text{ and } \text{Var}[X] = p - p^2$$

A Binomial random variable  $B_{n,p}$  counts the success of multiple independent Bernoulli random variable (trials) of parameter  $p$ .

Parameter  $n$  is the number of trials and parameter  $p$  is the success probability.

$$E[B_{n,p}] = np \text{ and } \text{Var}[B_{n,p}] = np(1 - p)$$

A Poisson random variable  $X_\lambda$  is a random variable that counts the number of success in a given time unit when it is expected  $\lambda$  amount of success would occur.

Parameter  $\lambda$  is the expected amount of success in the given time unit.

$$E[X_\lambda] = \lambda \text{ and } \text{Var}[X_\lambda] = \lambda$$

A geometric random variable  $X_p$  is a random variable that counts the number of trials required for the first success.

Parameter  $p$  is the success rate of each independent trials.

$$E[X_p] = \frac{1}{p} \text{ and } \text{Var}[X_p] = \frac{1-p}{p^2}$$

$P$  measure is specified by the natural condition of experiment (if the coin is fair).  $X$  is specified by what do you want to observe in the experiment.

# Chapter 2

## Continuous Random Variable

### 2.1 Basic Knowledge

**Definition 2.1.1. (Definition of Continuous Random Variable)** A continuous random variable  $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is a random variable such that

$$X[\Omega] \text{ is uncountable}$$

**Definition 2.1.2. (Definition of Cumulative Distribution Function)** The cumulative distribution function  $F_X$  is

$$F_X(x) = P(X \leq x)$$

**Definition 2.1.3. (Definition of Absolutely Continuous Random Variable)** By an absolutely continuous random variable, we mean a continuous random variable  $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that

$$F_X \text{ is differentiable everywhere}$$

**Definition 2.1.4. (Definition of Probability Density Function)** Let  $X$  be an absolutely continuous random variable. The probability density function  $f_X$  for  $X$  is defined to be

$$f_X(x) = F'_X(x)$$

Above is how we begin in our argument, we now develop some simple yet crucial tool. These tool enable us too calculate the probability of  $X$  taking value in some Borel set  $B$  by taking

$$P(X \in B) = \int_B f_X(s) ds$$

**Theorem 2.1.5. (Fundamental Theorem of Calculus)** We have

$$P(a \leq X \leq b) = \int_a^b f_X(s)ds$$

*Proof.*

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X < a) \\ &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b F'_X(s)ds \\ &= \int_a^b f_X(x)ds \end{aligned}$$

■

**Corollary 2.1.6. (Fundamental Theorem of Calculus)** We have

$$P(a \leq X) = \int_a^\infty f_X(s)ds$$

*Proof.* Because measure is countable additive, we have

$$\begin{aligned} P(a \leq X) &= \lim_{n \rightarrow \infty} \sum_{k=0}^n P(a+k \leq X < a+k+1)? \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n \int_{a+k}^{a+k+1} f_X(s)ds \\ &= \lim_{n \rightarrow \infty} \int_a^{a+n+1} f_X(s)ds \\ &= \int_a^\infty f_X(s)ds \end{aligned}$$

■

**Corollary 2.1.7. (Fundamental Theorem of Calculus)** We have

$$P(X \leq a) = \int_{-\infty}^a f_X(s)ds$$

*Proof.* Because measure is countable additive, we have

$$\begin{aligned}
P(X \leq a) &= \lim_{n \rightarrow \infty} \sum_{k=0}^n P(a - k - 1 < X \leq a - k) \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^n \int_{a-k-1}^{a-k} f_X(s) ds \\
&= \lim_{n \rightarrow \infty} \int_{a-n-1}^a f_X(s) ds \\
&= \int_{-\infty}^a f_X(s) ds
\end{aligned}$$

■

The definition of expected value and variance of continuous random variable present.

**Definition 2.1.8. (Definition of Expected Value)** Let  $X$  be an absolutely continuous random variable. We define

$$E[X] := \int_{-\infty}^{\infty} s \times f_X(s) ds$$

**Definition 2.1.9. (Definition of Variance)** Let  $X$  be an absolutely continuous random variable. We define

$$\text{Var}[X] = \int_{-\infty}^{\infty} (s - E[X])^2 \times f_X(s) ds$$

**Theorem 2.1.10. (Formula for Variance)**

$$\text{Var}[X] = E[X^2] - E^2[X]$$

*Proof.*

$$\begin{aligned}
\text{Var}[X] &= \int_{-\infty}^{\infty} (s - E[X])^2 \times f_X(s) ds \\
&= \int_{-\infty}^{\infty} s^2 \times f_X(s) ds - 2E[X] \int_{-\infty}^{\infty} s \times f_X(s) ds + \int_{-\infty}^{\infty} E^2[X] \times f_X(s) ds \\
&= E[X^2] - 2E^2[X] + E^2[X] \\
&= E[X^2] - E^2[X]
\end{aligned}$$

■

The below show case when a variable  $Y$  depend on a random variable  $X$  (thus  $Y$  is a random variable): what is the correct way to compute  $f_Y$  from  $f_X$ . Notice that Theorem 2.1.11 is inspired by Theorem 2.1.12



**Theorem 2.1.11. (Example)** Define

$$f_X(x) := \begin{cases} 3x^2 & \text{if } x \in [0, 1] \\ 0 & \text{if otherwise} \end{cases} \quad \text{and } g(x) := 1 - x^4 \text{ and } Y = g \circ X$$

We don't have

$$f_Y(y) = g^{-1}(f_X(g^{-1}(y)))$$

*Proof.* We have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(1 - X^4 \leq y) \\ &= P(1 - y \leq X^4) \\ &= P(X \geq (1 - y)^{\frac{1}{4}}) \\ &= \int_{(1-y)^{\frac{1}{4}}}^{\infty} f_X(s) ds \\ &= \int_{(1-y)^{\frac{1}{4}}}^1 3s^2 ds \\ &= \left[ s^3 \right]_{s=(1-y)^{\frac{1}{4}}}^1 \\ &= 1 - (1 - y)^{\frac{3}{4}} \end{aligned}$$

This give us

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} 1 - (1 - y)^{\frac{3}{4}} \\ &= \frac{3}{4} (1 - y)^{-\frac{1}{4}} \end{aligned}$$

While

$$\begin{aligned} g^{-1}(f_X(g^{-1}(y))) &= g^{-1}(f_X((1 - y)^{\frac{1}{4}})) \\ &= g^{-1}(3(1 - y)^{\frac{1}{2}}) \\ &= (1 - 3(1 - y)^{\frac{1}{2}})^{\frac{1}{4}} \end{aligned}$$

■

**Theorem 2.1.12. (Example)** Define

$$f_X(x) := \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{if elsewhere} \end{cases} \quad \text{and } g(x) := 3x \text{ and } Y = g \circ X$$

*Proof.* We see

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(3X \leq y) \\&= P(X \leq \frac{y}{3}) \\&= F_X(\frac{y}{3}) \\&= \frac{y}{3} \text{ for } y \in [0, 3]\end{aligned}$$

Then

$$f_Y(y) = \begin{cases} \frac{1}{3} & \text{if } y \in [0, 3] \\ 0 & \text{if otherwise} \end{cases}$$

And

$$g^{-1}(f_X(g^{-1}(y))) = g^{-1}f_X(\frac{y}{3}) = g^{-1}(1) = \frac{1}{3}$$

■

## 2.2 Law of Unconscious Statistician

**Theorem 2.2.1. (Law of Unconscious Statistician)** Let  $Y = g(X)$ . If  $X$  is discrete, we have

$$E[Y] = \sum_{x \in X[\Omega]} g(x)p_X(x)$$

If  $X$  is continuous, we have

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

*Proof.* We only prove when  $X$  is continuous and  $g$  is differentiable and monotone and  $g^{-1}$  exists and is differentiable and monotone.

First notice

$$x = g^{-1}(y) \implies \frac{dx}{dy} = \frac{d}{dy}g^{-1}(y) \implies dx = \frac{d}{dy}g^{-1}(y)dy$$

Notice

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \end{aligned}$$

Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) \\ &= \frac{d}{dy}F_X(g^{-1}(y)) \\ &= \frac{d}{dg^{-1}(y)}F_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} \\ &= f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} \end{aligned}$$

We now have

$$\begin{aligned} \int_{-\infty}^{\infty} yf_Y(y)dy &= \int_{-\infty}^{\infty} yf_X(g^{-1}(y))dg^{-1}(y) \\ &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \end{aligned}$$

■

## 2.3 Uniform, Exponential and Normal

We now go into each continuous random variable. Start with the easiest, uniform random variable.

**Definition 2.3.1. (Definition of Uniform Random Variable)** We say an absolutely continuous random variable  $X \sim \text{Uni}(\alpha, \beta)$  is uniform if there exists  $\alpha < \beta \in \mathbb{R}$  such that

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{if elsewhere} \end{cases}$$

**Theorem 2.3.2. (Expected Value and Variance of Uniform Random Variable)**  
Let  $X \sim \text{Uni}(\alpha, \beta)$ . We have

$$E[X] = \frac{\alpha + \beta}{2} \text{ and } \text{Var}[X] = \frac{(\beta - \alpha)^2}{12}$$

*Proof.* First verify for

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dx = \frac{\beta - \alpha}{\beta - \alpha} = 1$$

Then

$$\begin{aligned} E[X] &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{1}{\beta - \alpha} \times \frac{x^2}{2} \Big|_{x=\alpha}^{\beta} \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

Then

$$\begin{aligned}
 E[X^2] &= \int_{\alpha}^{\beta} \frac{x^2}{\beta - \alpha} dx \\
 &= \frac{1}{\beta - \alpha} \times \frac{x^3}{3} \Big|_{x=\alpha}^{\beta} \\
 &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\
 &= \frac{(\beta - \alpha)(\beta^2 + \beta\alpha + \alpha^2)}{3(\beta - \alpha)} \\
 &= \frac{\beta^2 + \beta\alpha + \alpha^2}{3}
 \end{aligned}$$

Then

$$\begin{aligned}
 \text{Var}[X] &= E[X^2] - E^2[X] \\
 &= \frac{\beta^2 + \beta\alpha + \alpha^2}{3} - \frac{\beta^2 + 2\beta\alpha + \alpha^2}{4} \\
 &= \frac{\beta^2 - 2\beta\alpha + \alpha^2}{12} \\
 &= \frac{(\beta - \alpha)^2}{12}
 \end{aligned}$$

■

Second, exponential.

**Definition 2.3.3. (Definition of Exponential Random Variable)** We say an absolute continuous random variable  $X \sim \text{Exp}(\lambda)$  is exponential if there exists  $\lambda \in \mathbb{R}^+$  such that

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if elsewhere} \end{cases}$$

**Theorem 2.3.4. (Expected Value and Variance of Exponential Random Variable)** Let  $X \sim \text{Exp}(\lambda)$ . We have

*Proof.* First, verify for

$$\begin{aligned}
 \int_{-\infty}^{\infty} f_X(x) dx &= \int_0^{\infty} \lambda e^{-\lambda x} dx \\
 &= -e^{-\lambda x} \Big|_{x=0}^{\infty} \\
 &= -0 - (-1) \\
 &= 1
 \end{aligned}$$

We have

$$\begin{aligned}
E[X] &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\
&= -x e^{-\lambda x} \Big|_{x=0}^{\infty} - \int_0^{\infty} -e^{-\lambda x} dx \\
&= \int_0^{\infty} e^{-\lambda x} dx \\
&= \frac{e^{-\lambda x}}{-\lambda} \Big|_{x=0}^{\infty} \\
&= 0 - \left( \frac{1}{-\lambda} \right) \\
&= \frac{1}{\lambda}
\end{aligned}$$

And have

$$\begin{aligned}
E[X^2] &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\
&= -x^2 e^{-\lambda x} \Big|_{x=0}^{\infty} - \int_0^{\infty} -2x e^{-\lambda x} dx \\
&= 2 \int_0^{\infty} x e^{-\lambda x} dx \\
&= 2 \left( \frac{-1}{\lambda} \right) x e^{-\lambda x} \Big|_{x=0}^{\infty} - 2 \int_0^{\infty} \left( \frac{-1}{\lambda} \right) e^{-\lambda x} dx \\
&= \frac{2}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \\
&= \frac{2}{\lambda} \left( \frac{-1}{\lambda} \right) e^{-\lambda x} \Big|_{x=0}^{\infty} \\
&= \left( \frac{-2}{\lambda^2} \right) (0 - 1) = \frac{2}{\lambda^2}
\end{aligned}$$

Then we have

$$\text{Var}[X] = E[X^2] - E^2[X] = \frac{2}{\lambda^2} - \left( \frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

■

**Theorem 2.3.5. (Property of Exponential Random Variable)** Let  $X \sim \text{Exp}(\lambda)$ . We see

$$P(X \geq a) = e^{-a\lambda}$$

*Proof.*

$$\begin{aligned}
 P(X \geq a) &= \int_a^{\infty} \lambda e^{-\lambda x} dx \\
 &= -e^{-\lambda x} \Big|_{x=a}^{\infty} \\
 &= e^{-a\lambda}
 \end{aligned}$$

■

**Theorem 2.3.6. (Memory-less Property of Exponential Random Variable)** Let  $X \sim \text{Exp}(\lambda)$ . We see

$$P(X \geq a + b \mid X \geq b) = \frac{e^{-(a+b)\lambda}}{e^{-b\lambda}} = e^{-a\lambda}$$

If  $X \sim \text{Exp}(\lambda)$ , it means the measurement of waiting time of an occurrence of event that is expected to happen after  $\lambda$  time period.

Before we start to talk about normal distribution. We first have to talk about Gauss Integral.

**Theorem 2.3.7. (Gauss Integral)**

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

*Proof.* Let  $I = \int_{-\infty}^{\infty} e^{-x^2} dx$ . We have

$$\begin{aligned}
 I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy \\
 &= \int_0^{\infty} \int_0^{2\pi} r e^{-r^2} d\theta dr \\
 &= 2\pi \int_0^{\infty} r e^{-r^2} dr \\
 &= 2\pi \frac{1}{-2} e^{-r^2} \Big|_{r=0}^{\infty} \\
 &= \pi
 \end{aligned}$$

Then

$$I^2 = \pi \implies I = \sqrt{\pi}$$

■

**Theorem 2.3.8. (Gauss Integral)**

$$\int_{-\infty}^{\infty} e^{\frac{-(x-a)^2}{b}} dx = \sqrt{b\pi}$$

*Proof.* Let  $x = a + \sqrt{b}y$ . We have

$$\frac{dx}{dy} = \sqrt{b}$$

Then we have

$$\begin{aligned} \int_{-\infty}^{\infty} e^{\frac{-(x-a)^2}{b}} dx &= \int_{-\infty}^{\infty} e^{-y^2} \sqrt{b} dy \\ &= \sqrt{b\pi} \end{aligned}$$

■

**Theorem 2.3.9. (Gamma Function)** For  $z \in \mathbb{R}$ , let

$$\Gamma(z) = \int_0^{\infty} t^z e^{-t} dt$$

Now we start to discuss normal distribution.

**Definition 2.3.10. (Definition of Normal Distribution)** A normal random variable  $X \sim N(\mu, \sigma^2)$  is defined by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

**Theorem 2.3.11. (Expected Value and Variance of Normal Random Variable)** Let  $X \sim N(\mu, \sigma^2)$ . We have

$$E[X] = \mu \text{ and } \text{Var}[X] = \sigma^2$$

*Proof.* We first verify

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sqrt{2\sigma^2\pi} = 1 \end{aligned}$$



Observe

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu + \mu) e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu) e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2\sigma^2}} dz + \mu \\
&= \mu
\end{aligned}$$

Where

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2\sigma^2}} dz = 0 \text{ because } z e^{\frac{-z^2}{2\sigma^2}} \text{ is odd function}$$

Observe

$$\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx + 2\mu \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx - \mu^2 \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} z^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-z^2}{2\sigma^2}} dz + \mu^2 \\
&= 2 \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} s^2 e^{-s^2} ds + \mu^2 \text{ where } s = \frac{z}{\sqrt{2}\sigma} \\
&= 4 \frac{\sigma^2}{\sqrt{\pi}} \int_0^{\infty} s^2 e^{-s^2} ds + \mu^2 \\
&= 2 \frac{\sigma^2}{\sqrt{\pi}} \int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt + \mu^2 \text{ where } t = s^2 \\
&= 2 \frac{\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) + \mu^2 \\
&= 2 \frac{\sigma^2}{\sqrt{\pi}} \left(\frac{\sqrt{\pi}}{2}\right) + \mu^2 \\
&= \sigma^2 + \mu^2
\end{aligned}$$

Then we have

$$\text{Var}[X] = E[X^2] - E[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

■

Now, we introduce how to do linear change of variable for normal distribution.

**Theorem 2.3.12. (Linear Change of Normal Random Variable)** Let

$$X \sim N(\mu, \sigma^2) \text{ and } Y = \alpha X + \beta \text{ where } \alpha > 0$$

We have

$$Y \sim N(\alpha\mu + \beta, (\alpha\sigma)^2)$$

*Proof.* Observe

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\alpha X + \beta \leq y) \\ &= P(X \leq \frac{y - \beta}{\alpha}) \\ &= \int_{-\infty}^{\frac{y - \beta}{\alpha}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \end{aligned}$$

This give us

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d\frac{y - \beta}{\alpha}}{dy} \frac{d}{d\frac{y - \beta}{\alpha}} \int_{-\infty}^{\frac{y - \beta}{\alpha}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \\ &= \frac{1}{\alpha} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{y - \beta}{\alpha} - \mu)^2}{2\sigma^2}} \\ &= \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{-\frac{(y - \beta - \alpha\mu)^2}{2\alpha^2\sigma^2}} \end{aligned}$$

This give us

$$Y \sim N(\alpha\mu + \beta, (\alpha\sigma)^2)$$

■

**Corollary 2.3.13.**

$$X \sim N(\mu, \sigma^2) \text{ and } Z = \frac{X - \mu}{\sigma} \implies Z \sim N(0, 1)$$

**Definition 2.3.14. (Definition of Standard Normal Random Variable)** Let  $Z \sim N(0, 1)$ . We denote

$$\Phi(z) := F_Z(z)$$

Thus

$$\Phi(z) = F_Z(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{(s-0)^2}{2 \cdot 1^2}} ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{s^2}{2}} ds$$

**Theorem 2.3.15. (Probability of Normal Variable within Confidence Level)** Let  $X \sim N(\mu, \sigma^2)$ . We have

$$\begin{aligned}P(X \in [\mu - \sigma, \mu + \sigma]) &= 0.3173 \\P(X \in [\mu - 2\sigma, \mu + 2\sigma]) &= 0.0455 \\P(X \in [\mu - 3\sigma, \mu + 3\sigma]) &= 0.0027\end{aligned}$$

*Proof.* Let  $Z = \frac{X-\mu}{\sigma}$ . We have  $Z \sim N(0, 1)$ . This give us

$$\begin{aligned}P(X \in [\mu - \sigma, \mu + \sigma]) &= P(Z \in [-1, 1]) = 0.3173 \\P(X \in [\mu - 2\sigma, \mu + 2\sigma]) &= P(Z \in [-2, 2]) = 0.0455 \\P(X \in [\mu - \sigma, \mu + \sigma]) &= P(Z \in [-3, 3]) = 0.0027\end{aligned}$$

■

**Theorem 2.3.16. (Example)** Let

$$X \sim N(2, 25)$$

Compute

$$P(1 \leq X \leq 4)$$

*Proof.*

$$\mu = 2 \text{ and } \sigma = 5$$

■

## 2.4 Exercises

### Question 1

A “dyadic rational number” is a real number of the form  $\frac{m}{2^k}$  where  $k$  and  $m$  are integers. Suppose we set  $p = q = \frac{1}{2}$  in the construction for a probability measure on  $\Omega_\infty$  and  $X(\omega) = \sum_{i=1}^{\infty} \frac{Y_i(\omega)}{2^i}$  is a random variable on  $\Omega_\infty$ .

- Show that, the induced measure  $\mathcal{L}_X$  by the random variable  $X$  on  $\Omega$  satisfies that, for any positive integers  $k$  and  $m$  such that  $0 \leq \frac{m-1}{2^k} < \frac{m}{2^k} \leq 1$ , we have

$$\mathcal{L}_X \left[ \frac{m-1}{2^k}, \frac{m}{2^k} \right] = \frac{1}{2^k}.$$

In other words, the induced measure  $\mathcal{L}_X$  on all intervals in  $[0, 1]$  whose endpoints are dyadic rational numbers is the same as the Lebesgue measure of these intervals. The only possible way is that  $\mathcal{L}_X$  is indeed the Lebesgue measure.

- Show that, in this case ( $p = \frac{1}{2}$ ), the random variable  $X(\omega) = \sum_{k=1}^{\infty} \frac{Y_k(\omega)}{2^k}$  is uniformly distributed on  $[0, 1]$ .

*Proof. (Part 1)* First notice

$$\begin{aligned} \mathcal{L}_X \left[ \frac{m-1}{2^k}, \frac{m}{2^k} \right] &= P \left( X \in \left[ \frac{m-1}{2^k}, \frac{m}{2^k} \right] \right) \\ &= P \left( (2^k) X \in [m-1, m] \right) \end{aligned}$$

Then we express  $X$  in the form

$$X(\omega) = \sum_{i=1}^{\infty} \frac{Y_i(\omega)}{2^i} = \sum_{i=1}^k \frac{Y_i(\omega)}{2^i} + \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i}$$

Notice that because  $Y_i(\omega) \in \{0, 1\}$ , we can compute

$$\sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in \left[ 0, \frac{1}{2^k} \right]$$

Then

$$2^k X(\omega) = \sum_{i=1}^k (2^{k-i}) Y_i(\omega) + 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \text{ where } 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in [0, 1]$$

$$\text{and } \sum_{i=1}^k (2^{k-i} Y_i(\omega)) \in \mathbb{Z}$$

Notice that we have

$$P\left(\sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} = 0\right) = 0$$

Because of such, we can deduce

$$\begin{aligned} P((2^k) X \in [m-1, m]) &= P\left(\sum_{i=1}^k (2^{k-i}) Y_i(\omega) + 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in [m-1, m]\right) \\ &= P\left(\sum_{i=1}^k (2^{k-i}) Y_i(\omega) = m-1\right) \end{aligned}$$

Let  $N = m-1$ . Express  $N$  by binary

$$N = \sum_{j=0}^n a_j 2^j \text{ where } a_j \in \{0, 1\} \text{ and } a_n = 1$$

That is

$$N = a_n 2^n + a_{n-1} 2^{n-1} + \dots + a_1 2^1 + a_0 2^0$$

The premise of the problem include  $\frac{m}{2^k} \leq 1$ , so we have  $n < k$ . Then we can express  $N$  by

$$N = a_{k-1} 2^{k-1} + a_{k-2} 2^{k-2} + \dots + a_0 2^0$$

Then we see

$$\begin{aligned} P\left(\sum_{i=1}^k (2^{k-i}) Y_i(\omega) = m-1\right) &= P\left(\sum_{i=1}^k (2^{k-i}) Y_i(\omega) = N\right) \\ &= P(\forall i : 1 \leq i \leq k, Y_i(\omega) = a_{k-i}) \\ &= \prod_{i=1}^k P(Y_i(\omega) = a_{k-i}) \\ &= \prod_{i=1}^k \frac{1}{2} = \frac{1}{2^k} \end{aligned}$$

**(Part 2)** Arbitrarily select an closed interval  $[a, b]$  contained by  $[0, 1]$ , we only wish to prove

$$\mathcal{L}_X [a, b] = b - a$$

Express  $a, b$  by

$$a = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \text{ and } b = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$$

Define

$$A_k = \sum_{n=1}^k \frac{a_n}{2^n} \text{ and } B_k = \sum_{n=1}^k \frac{b_n}{2^n} \text{ where } \{a_n : n \in \mathbb{N}\} \cup \{b_n : n \in \mathbb{N}\} \subseteq \{0, 1\}$$

We have

$$\lim_{k \rightarrow \infty} A_k = a \text{ and } \lim_{k \rightarrow \infty} B_k = b$$

Express  $A_k, B_k$  by

$$A_k = \frac{\sum_{n=1}^k (2^{k-n}) a_n}{2^k} \text{ and } B_k = \frac{\sum_{n=1}^k (2^{k-n}) b_n}{2^k}$$

Define  $A'_k := \sum_{n=1}^k (2^{k-n}) a_n$  and  $B'_k = \sum_{n=1}^k (2^{k-n}) b_n$ . It is easy to check  $A'_k, B'_k \in \mathbb{Z} \cap [0, 2^k]$ . Then by the result in part 1, we have

$$\begin{aligned} \forall k, P(X \in [A_k, B_k]) &= P\left(X \in \left[\frac{A'_k}{2^k}, \frac{B'_k}{2^k}\right]\right) \\ &= B_k - A_k \end{aligned}$$

Then we see

$$\begin{aligned} \mathcal{L}_X [a, b] &= P(X \in [a, b]) \\ &= \lim_{k \rightarrow \infty} P(X \in [A_k, B_k]) \\ &= \lim_{k \rightarrow \infty} B_k - A_k \\ &= b - a \end{aligned}$$

■

## 2.5 Central Limit Theorem

**Theorem 2.5.1. (Central Limit Theorem)** Let  $X$  be an absolutely continuous random variable with expectation value  $\mu$  and variance  $\sigma^2$ . Define

$$Z_n := \frac{X_1 + \cdots + X_n - \mu n}{\sigma \sqrt{n}}$$

We have

$$\forall x, \lim_{n \rightarrow \infty} f_{Z_n}(x) = f_X(x)$$

**Corollary 2.5.2. (De Moivre-Laplace Central Limit Theorem)** Fix  $p$ . Suppose  $S_n \sim \text{Bin}(n, p)$ . Define  $\mu_n := E[S_n] = np$  and  $\sigma_n := \sqrt{\text{Var}[S_n]} = \sqrt{np(1-p)}$ . Define  $Z_n := \frac{S_n - \mu_n}{\sigma_n}$ . We have

$$\forall k \in \mathbb{N}, \lim_{n \rightarrow \infty} \frac{f_{S_n}(k)}{f_{Z_n}(k)} = 1$$

In other words

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

**Corollary 2.5.3. (De Moivre-Laplace Central Limit Theorem)** Recall

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Let

$$S_n \sim \text{Bin}(n, p)$$

Fix  $p$ , we have

$$\forall x \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left(\frac{S_n - \mu_n}{\sigma_n} \leq x\right) = \Phi(x)$$

where

$$\mu_n = E[S_n] = np \text{ and } \sigma_n = \sqrt{\text{Var}[S_n]} = \sqrt{np(1-p)}$$

Then very weakly, we have

$$\lim_{n \rightarrow \infty} \frac{S_n - \mu_n}{\sigma_n} \sim N(0, 1)$$

**Corollary 2.5.4. (Approximation of Large Binomial R.V.)** Let  $S_n \sim \text{Bin}(n, p)$ . If  $n$  is large enough, we can approximate

$$\begin{aligned} P(m_1 \leq S_n \leq m_2) &= P\left(\frac{m_1 - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}}\right) \\ &\approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right) \\ &= \int_{\frac{m_1 - np}{\sqrt{npq}}}^{\frac{m_2 - np}{\sqrt{npq}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

And

$$P(S_n = m) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(m - np)^2}{2npq}}$$

### Question 2

A die is threw 12000 times. What is the probability that there will be exactly 1800 rolls of 6? What is the probability that the number of 6's lies in the interval  $[1950, 2100]$

*Proof.* We have  $n = 12000$  and  $p = \frac{1}{6}$ . Let  $S_n \sim \text{Bin}(n, p)$ . We have

$$P(S_n = 1800) \approx \frac{1}{\sqrt{2\pi \times 12000 \times \frac{5}{36}}} e^{-\frac{(1800 - 2000)^2}{2 \times 12000 \times \frac{5}{36}}} \sim 6.004 \times 10^{-8}$$

Notice that the direct computation of  $P(S_n = 1800)$  is  $\binom{12000}{1800} \left(\frac{1}{6}\right)^{1800} \left(\frac{5}{6}\right)^{10200}$ , in which the computation of  $\binom{12000}{1800}$  is difficult even for computer.

Again, notice that the Poisson approximation  $V \sim \text{Poisson}(2000)$  of  $P(V = 1800)$  is  $\frac{e^{-2000} 2000^{1800}}{1800!}$ , in which the computation of  $1800!$  is difficult even for computer.

Let  $\sigma = \sqrt{npq}$  and  $\mu = np$ . We have

$$\begin{aligned} P(1950 \leq S_n \leq 2100) &\approx \int_{\frac{1950 - \mu}{\sigma}}^{\frac{2100 - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \Phi\left(\frac{2100 - \mu}{\sigma}\right) - \Phi\left(\frac{1950 - \mu}{\sigma}\right) \\ &\approx 0.8825 \end{aligned}$$

■



# Chapter 3

## Joint Distribution

### 3.1 Independence of Event

**Definition 3.1.1. (Definition of Independence of Events)** Let  $A, B$  be two events in  $\mathcal{F}$ . We define

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

We say  $A$  is independent of  $B$  if

$$P(A|B) = P(A)$$

**Theorem 3.1.2. (Independence)**

$$\begin{aligned} A \text{ is independent of } B &\iff P(A \cap B) = P(A)P(B) \\ &\iff B \text{ is independent of } A \end{aligned}$$

*Proof.*

$$\begin{aligned} P(A|B) = P(A) &\iff \frac{P(A \cap B)}{P(B)} = P(A) \\ &\iff P(A \cap B) = P(A)P(B) \\ &\iff \frac{P(A \cap B)}{P(A)} = P(B) \\ &\iff P(B|A) = P(B) \end{aligned}$$

■

**Corollary 3.1.3.**

$$A, A^c \text{ are independent} \iff P(A) = 0 \text{ or } 1$$

Every event is independent of  $\Omega$  and  $\emptyset$

**Corollary 3.1.4. (Independence)**

$$\begin{aligned}
A \text{ is independent of } B &\iff A^c \text{ is independent of } B \\
&\iff A \text{ is independent of } B^c \\
&\iff A^c \text{ is independent of } B^c
\end{aligned}$$

*Proof.* We have

$$\begin{aligned}
P(A^c \cap B) &= P(B \setminus A) \\
&= P(B) - P(A \cap B)
\end{aligned}$$

and

$$P(A^c) = 1 - P(A)$$

Then we have

$$P(A^c)P(B) = P(B) - P(A)P(B)$$

This tell us

$$P(A^c \cap B) = P(A^c)P(B) \iff A, B \text{ are independent}$$

■

**Corollary 3.1.5.**

$$A, B \text{ are independent} \iff \forall A' \in \sigma(A), \forall B' \in \sigma(B), A', B' \text{ are independent}$$

**Definition 3.1.6. (Independence of Sub- $\sigma$ -Algebra)** We say two sub- $\sigma$ -algebras  $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$  is independent if

$$\forall A \in \mathcal{G}, \forall B \in \mathcal{H}, A, B \text{ are independent}$$

**Corollary 3.1.7.** Given two event  $A, B$

$$A, B \text{ are independent} \iff \sigma(A) \text{ and } \sigma(B) \text{ are independent}$$

**Question 3**

Toss a coin twice with  $\Omega_2 = \{HH, HT, TH, TT\}$  and  $\mathcal{F}_2$  (with  $|\mathcal{F}_2| = 16$ ) is the  $\sigma$ -algebra consisting of all information up to time 2.

Let the probability of tossing a  $H$  in the first trial be  $p_1 \in (0, 1)$  and that for the second trial be  $p_2 \in (0, 1)$ .

Define the probability  $P$  as

$$P(HH) = p_1 p_2, P(HT) = p_1 (1 - p_2) \text{ vice versa}$$

Then,

$(\Omega_2, \mathcal{F}_2, P)$  is a probability space

Let

$$\mathcal{G} = \{\emptyset, \Omega_2, \{HH, HT\}, \{TH, TT\}\} \subset \mathcal{F}_2$$

be the sub- $\sigma$ -algebra consisting of information of the first toss; while

$$\mathcal{H} = \{\emptyset, \Omega_2, \{HH, TH\}, \{HT, TT\}\} \subset \mathcal{F}_2$$

be the sub- $\sigma$ -algebra consisting of information of the second toss.

Show that

$$p_1 = \frac{1}{2} = p_2 \implies \mathcal{G}, \mathcal{H} \text{ are independent}$$

and that

$$P(HH) = \frac{1}{9}, P(HT) = \frac{2}{9}, P(TH) = \frac{1}{3}, P(TT) = \frac{1}{3} \implies \mathcal{G}, \mathcal{H} \text{ are dependent}$$

*Proof.* Let  $A = \{HH, HT\}$  and  $B = \{HH, TH\}$ . We see

$$\mathcal{G} = \{\emptyset, \Omega_2, A, A^c\} \text{ and } \mathcal{H} = \{\emptyset, \Omega_2, B, B^c\}$$

And

$$A \cap B = \{HH\}$$

Then if  $p_1 = \frac{1}{2} = p_2$ , we see

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A) P(B)$$

If  $P(HH) = \frac{1}{9}, P(HT) = \frac{2}{9}, P(TH) = \frac{1}{3}, P(TT) = \frac{1}{3}$ , we see

$$P(A \cap B) = \frac{1}{9} \neq \frac{1}{3} \times \frac{4}{9} = P(A) P(B)$$

■

## 3.2 Independence of Random Variable

**Definition 3.2.1. ( $\sigma$ -Algebra Generated by Random Variable)** The  $\sigma$ -algebra  $\sigma(X)$  generated by random variable  $X$  is defined by

$$\sigma(X) := \{X^{-1}[B] : B \in \mathcal{B}(\mathbb{R})\}$$

**Lemma 3.2.2.** Suppose

$A$  is independent of  $C$  and  $B$  is independent of  $C$  and  $A \cap B = \emptyset$

Then

$A \cup B$  is independent of  $C$

*Proof.*

$$\begin{aligned} P((A \cup B) \cap C) &= P((A \cap C) \cup (B \cap C)) \\ &= P(A \cap C) + P(B \cap C) \\ &= P(A)P(C) + P(B)P(C) \\ &= (P(A) + P(B))P(C) \\ &= P(A \cup B)P(C) \end{aligned}$$

■

### Question 4

Show that in the probability space of the binomial asset pricing model with three independent tosses of the same coin, the two random variable  $S_2$  and  $\frac{S_3}{S_2}$  are independent.

*Proof.* The  $\sigma$ -algebra  $\sigma(S_2)$  generated by  $S_2$  is  $\mathcal{F}_2$

$$\sigma(S_2) = \left\{ \bigcup B : B \subseteq \{A_{HH}, A_{TT}, A_{HT} \cup A_{TH}\} \right\}$$

The  $\sigma$ -algebra  $\sigma\left(\frac{S_3}{S_2}\right)$  generated by  $\frac{S_3}{S_2}$  is

$$\sigma\left(\frac{S_3}{S_2}\right) = \{\emptyset, \Omega, A_{..H}, A_{..T}\}$$

Observe

$$P(A_{HH} \cap A_{..H}) = P(\{HHH\}) = \frac{1}{p^3} = \frac{1}{p^2} \times \frac{1}{p} = P(A_{HH})P(A_{..H})$$

and

$$P(A_{TT} \cap A_{..H}) = P(\{TTH\}) = \frac{1}{q^2p} = \frac{1}{q^2} \times \frac{1}{p} = P(A_{TT}) P(A_{..H})$$

Then by corollary 3.1.3 and lemma 3.2.2, we see

$$\sigma(S_2), \sigma\left(\frac{S_3}{S_2}\right) \text{ are independent}$$

■

**Theorem 3.2.3. (Change of Independent Random Variable)** If  $X, Y$  are independent random variables, then

$$g(X) \text{ and } h(Y) \text{ are independent}$$

*Proof.* This is an immediate consequence of

$$\sigma(g(X)) \subseteq \sigma(X) \text{ and } \sigma(h(Y)) \subseteq \sigma(Y)$$

■

An example for Theorem 3.2.3 is  $X(H) = 1$  and  $X(T) = -1$ . We see

$$\sigma(X) = \{\emptyset, \{H\}, \{T\}, \Omega\}$$

and

$$\sigma(X^2) = \{\emptyset, \Omega\} \subset \sigma(X)$$

### 3.3 Bayes' Theorem

#### Question 5

We have the following knowledge

$$P(\text{tested } H | \text{being } H) = 0.9$$

$$P(\text{tested } T | \text{being } T) = 0.8$$

$$P(\text{being } H) = 0.05$$

Compute

$$P(\text{being } H | \text{tested } H)$$

*Proof.* We have

$$\begin{cases} \frac{P(\text{tested } H \cap \text{being } H)}{P(\text{being } H)} = 0.9 \\ P(\text{being } H) = 0.05 \end{cases}$$

So we have

$$P(\text{tested } H \cap \text{being } H) = 0.045$$

Also, the fact  $P(\text{being } H) = 0.05$  tell us  $P(\text{being } T) = 0.95$ . Then from

$$\begin{cases} \frac{P(\text{tested } T \cap \text{being } T)}{P(\text{being } T)} = 0.8 \\ P(\text{being } T) = 0.95 \end{cases}$$

We have

$$P(\text{tested } T \cap \text{being } T) = 0.76$$

This give us

$$P(\text{tested } H \cap \text{being } T) = P(\text{being } T) - P(\text{tested } T \cap \text{being } T) = 0.19$$

Then we have

$$P(\text{tested } H) = P(\text{tested } H \cap \text{being } H) + P(\text{tested } H \cap \text{being } T) = 0.235$$

Then

$$\frac{P(\text{being } H \cap \text{tested } H)}{P(\text{tested } H)} = \frac{0.045}{0.235}$$

**Theorem 3.3.1. (Bayes's Theorem)**

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

### Question 6

We have the following knowledge

$$P(\text{tested } H | \text{being } H) = 0.9$$

$$P(\text{tested } T | \text{being } T) = 0.8$$

$$P(\text{being } H) = 0.05$$

Compute

$$P(\text{being } H | \text{tested } H)$$

*Proof.* Let  $B = (\text{being } H)$  and  $A = (\text{tested } H)$ . By Bayes' Theorem, we have

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) P(B)}{P(A)} = \frac{0.9 \times 0.05}{0.235}$$

where  $P(A)$  is computed as above. ■

**Theorem 3.3.2. (General Form of Bayes' Theorem)** Let  $\{F_1, F_2, \dots, F_n\} \subseteq \mathcal{F}$  be a partition of  $\Omega$ . We have

$$P(F_j|A) = \frac{P(A|F_j) P(F_j)}{\sum_{i=1}^n P(A|F_i) P(F_i)}$$

*Proof.*

$$\begin{aligned} P(F_j|A) &= \frac{P(F_j \cap A)}{P(A)} \\ &= \frac{P(A|F_j) P(F_j)}{P(A)} \\ &= \frac{P(A|F_j) P(F_j)}{P(A \cap F_1) + P(A \cap F_2) + \dots + P(A \cap F_n)} \\ &= \frac{P(A|F_j) P(F_j)}{\sum_{i=1}^n P(A|F_i) P(F_i)} \end{aligned}$$
■

## 3.4 Independent Disjoint Grouping Theorem

**Definition 3.4.1. (Definition of Sub- $\sigma$ -Algebra Being Independent)** In  $(\Omega, \mathcal{F}, P)$ , we say a finite amount of sub- $\sigma$ -algebra  $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n\}$  is independent if we arbitrarily choose one event  $G_i$  from each sub- $\sigma$ -algebra  $\mathcal{G}_i$ , we have

$$P\left(\bigcap_{i=1}^n G_i\right) = \prod_{i=1}^n P(G_i)$$

**Definition 3.4.2. (Definition of Set of Random Variable Being Independent)** We say a set of random variable  $\{x_1, \dots, x_n\}$  is independent if

$$\{\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)\} \text{ are independent}$$

**Theorem 3.4.3. (Example)**

$$\{X_1, X_2, X_3, X_4\} \text{ are independent} \implies X_2 X_3 \text{ is independent of } X_1 + \cos(2X_4)$$