Probability Lecture Note

Author: Lv Changle

Institute: University of Science and Technology of China

Date: August 8, 2023

Probability is the measure theory with a soul.

Contents

Chapter 1 Basic Definitions **2**

1.1 Probability Space 2

1.2 Conditional Probability and Independence 4

1.3 Classic Models 7

1.4 Random Variables 9

1.5 Random Vectors 12

Chapter 2 Discrete Random Variables **14**

2.1 Classical Distributions and Independence 14

2.2 Expectation 16

2.3 Probabilistic Method 19

2.4 Conditional Expectation 21

2.5 Random Walk 24

2.6 Generating Function 26

Chapter 3 Continuous Random Variables **29**

3.1 Basic Definitions 29

3.2 (Conditional) Expectation 31

3.3 Multivariate Normal Distribution 35

Chapter 4 Law of Big Numbers **38**

4.1 Expectation Revisited 38

4.2 Modes of Convergence 41

4.3 Almost surely convergence and Borel-Cantelli Lemma 46

4.4 Law of Large Numbers 48

Chapter 5 Central Limit Theorem **52**

5.1 Characteristic Functions 52

5.2 Inversion and Continuity Theorem 55

5.3 Limit Theorems 57

Introduction

Probability Theory(MATH3007) is the first lecture about the probability and statistics major in USTC. The author of the note is learning the course during the autumn semester of 2023 and the lecturer is Prof. Dangzheng Liu. In this note, the main content of the course (c.f. Grimmett, Stirzaker: Probability and Random Process, Chapter 1-5, 7.1-7.6) will be covered. The important definitions, propositions and theorems will be talked about here. There will also be some examples for a easier understanding. In the end of each chapter, the author will give a summary about the most important things in the chapter and some exercises (and answers maybe) will be listed.

This course need some knowledge about real analysis(mainly about the measure and integration theory), which can be found in the textbook *Real Analysis:Modern Techniques and Their Applications* by B. Folland. And the textbook by Grimmett and Stirzaker mentioned above will be the main reference textbook in this course.

Since the expertise of the author and the time of writing the note are limited, it's unavoidable that there are some mistakes in the note. For the sake of future readers, the author hopes that each reader will take the time to keep notes of any mistakes or passages that are awkward or unclear, and let the author know about them as soon as it is convenient for you. Happy reading!

Chapter 1 Basic Definitions

1.1 Probability Space

Definition 1.1

The result of an experiment is called its outcome. *The set of all possible outcomes of an experiment is called the* sample space *and is denoted by* Ω. *Events are the subsets of the sample space* Ω.

♣

Definition 1.2

A subcollection $\mathcal{F}$ *of the set of all subsets of* Ω *is called a* σ-algebra *(or* σ-field*) if* $\mathcal{F}$ *satisfies the following properties:*

(a) $\Omega \in \mathcal{F}$

(b) If $A \in \mathcal{F}$*, then* $A^c \in \mathcal{F}$

(c) If $A_n \in \mathcal{F}$ *for* $n = 1, 2 \ldots$*, then* $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$

♣

It follows from the definition that the σ-fields are closed under the operation of taking countable intersections and unions.

Definition 1.3

A probability measure $\mathbb{P}$ *on* $(\Omega,\mathcal{F})$ *is a function* $\mathbb{P} : \mathcal{F} \to \mathbb{R}$*, so that:*

(a) $\forall A \in \mathcal{F}$, $\mathbb{P}(A) \geqslant 0$

(b) $\mathbb{P}(\Omega) = 1$

(c) If $A_1, A_2 \ldots$ *is a collection of disjoint members of* $\mathcal{F}$*, then*

$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

♣

There are some basic properties of the probability measure.

Lemma 1.1

The probability measure $\mathbb{P}$ satisfies the following:

(a) $\mathbb{P}(A^c) + \mathbb{P}(A) = 1$

(b) If $A \subset B$, then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geqslant \mathbb{P}(A)$

(c) $\mathbb{P}(A \bigcup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \bigcap B)$

(d) $\mathbb{P}(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{k=1}^{n} (-1)^{k+1} \sum\limits_{1 \leq i_1 < i_2 \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \bigcap A_{i_2} \cdots \bigcap A_{i_k})$

(e) For $A_1 \subset A_2 \subset \ldots$, then $\mathbb{P}(\bigcup\limits_{n=1}^{\infty} A_n) = \lim\limits_{n \to \infty} \mathbb{P}(A_n)$

(f) For $B_1 \supset B_2 \supset \ldots$, then $\mathbb{P}(\bigcap\limits_{n=1}^{\infty} B_n) = \lim\limits_{n \to \infty} \mathbb{P}(B_n)$

♡

Proof (a)(b)(c) Trivial.

(d) We use induction. The case where $n = 1$ or 2 is easy. Assume the indentity holds for $n - 1$, then

$$
\begin{aligned}
\mathbb{P}(\bigcup_{i=1}^{n} A_i) &= \mathbb{P}(\bigcup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) - \mathbb{P}((\bigcup_{i=1}^{n-1} A_i) \bigcap A_n) \\
&= \mathbb{P}(\bigcup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) - \mathbb{P}(\bigcup_{i=1}^{n-1} (A_i \bigcap A_n))
\end{aligned}
$$

Then the result follows from expanding the first and thrid terms on the right by using induction hypothesis.

(e) Denote $A = \bigcup\limits_{n=1}^{\infty} A_n = A_1 \bigcup (A_2 \backslash A_1) \bigcup (A_3 \backslash A_2) \bigcup \ldots$, then by the definition of probability measure, we have

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

(f) Take complement and use (e). □

1.2 Conditional Probability and Independence

Definition 1.4

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

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

♣

A family $B_1, B_2, \dots, B_n$ of events is called a *partition* of the set Ω if

$$B_i \bigcap B_j \neq \emptyset \quad \text{when} \quad i \neq j, \quad \text{and} \quad \bigcup_{i=1}^{n} B_i = \Omega$$

Lemma 1.2 (Law of Total Probability)

$B_1, B_2, \dots, B_n$ *is a partition of* Ω *and* $\mathbb{P}(B_i) > 0$ *for all* i*, then*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

In particular, for B *such that* $0 < \mathbb{P}(B) < 1$*, one has*

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

♡

Proof $A = (A \bigcap B) \bigcup (A \bigcap B^c)$ is a disjoint union, so

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \bigcup B) + \mathbb{P}(A \bigcup B^c) \\ &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)\end{aligned}$$

The case when there are a collection of events is similar. □

Lemma 1.3 (Bayes' Formula)

$A_1, A_2, \dots, A_n$ *is a partition of* Ω *and* $\mathbb{P}(A_j) > 0$ *for all* j *and* $\mathbb{P}(B) > 0$*, then*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

♡

Proof Trivial. □

Definition 1.5

Events A and B are called independent *if*

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

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

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

for all finite subsets J of I. ♣

Example 1.1 If the family $\{A_i|i\in I\}$ has the property that

$$\mathbb{P}(A_i\bigcap A_j)=\mathbb{P}(A_i)\mathbb{P}(A_j)\quad \text{for}\quad \text{all}\quad i\neq j$$

then it is called *pairwise independent*. It is clear that independent families are pairwise independent, but NOT vice versa:

Suppose $\Omega=\{abc,acb,cab,cba,bca,bac,aaa,bbb,ccc\}$, and each of the events in Ω occurs with equal probability. Let A_k be the event that the kth letter is a, then one can check that the family $\{A_1,A_2,A_3\}$ is pairwise independent but not independent.

Example 1.2 Let A be an event that occurs with a probability $\epsilon\in(0,1)$, do the experiment for an infinite number of times, then $\mathbb{P}(A_\infty)=\mathbb{P}(A$ occurs for at least one time)=1.

Proof Let A_k denote the event that A occurs in the kth experiment, then

$$\begin{aligned}\mathbb{P}(A_\infty)&=\lim_{k\to\infty}\mathbb{P}(\bigcup_{i=1}^{k}A_i)=\lim_{k\to\infty}\mathbb{P}((\bigcap_{i=1}^{k}A_i^c)^c)\\&=\lim_{k\to\infty}1-\mathbb{P}((\bigcap_{i=1}^{k}A_i^c))=\lim_{k\to\infty}1-\prod_{i=1}^{k}(1-\mathbb{P}(A))\\&=\lim_{k\to\infty}1-(1-\epsilon)^k=1\end{aligned}$$

Note that the independence between distinct experiments is used in the proof. □

Lemma 1.4

If A, B are independent, then the families $\{A, B^c\}$,$\{A^c, B\}$ and $\{A^c, B^c\}$ are also independent. ♡

Proof Just check the definition of independence.

$$
\begin{aligned}
\mathbb{P}(A\bigcap B^c) &= \mathbb{P}(A) - \mathbb{P}(A\bigcap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\mathbb{P}(B^c)
\end{aligned}
$$

So A and B^c are independent. The other two independence can be proved similarly. □

1.3 Classic Models

Example 1.3 Let A={there are at least two students having the same birthday from n students}, then $\mathbb{P}(A) =?$

Proof It's difficult to directly compute the probability of A since A is the union of many events(three students having the same birthday, ...). However, the complement event is easy: the birthdays are all different.

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{A_{365}^n}{365^n}$$

Example 1.4 Put n balls into $N(\geq n)$ boxes. Each way of putting the balls has the same probability. A={there is one ball in each of the first n boxes}. Then $\mathbb{P}(A) =?$

Proof Firstly we will look at the most basic counting problems: there are n elements $a_1, a_2 \ldots, a_n$, and we pick $m(\leq n)$ elements from those. How many ways of picking in total? There are four cases:

(1) it's allowed to pick one element more than once, and there is an order of picking, then the answer is n^m

(2) it's allowed to pick one element more than once, and there is no order of picking, then the answer is C_{n-1+m}^{n-1}

(3) it's not allowed to pick one element more than once, and there is an order of picking, then the answer is A_n^m

(4) it's not allowed to pick one element more than once, and there is no order of picking, then the answer is C_n^m

Back to the original problem, there are also many cases: whether the balls are the same, and whether the boxes have a mximal capacity.

Case 1: the balls are not the same, and there is no maximal capacity:$\mathbb{P}(A) = \frac{n!}{N^n}$

Case 2: the balls are the same, and there is no maximal capacity:$\mathbb{P}(A) = \frac{1}{C_{N+1-n}^n}$

Case 3: the balls are the same, and there is at most one ball in each box:$\mathbb{P}(A) = \frac{1}{C_N^n}$

Example 1.5 n couples are sitting at the two sides of the desk. Males are all sitting at one side. A={there is at least one couple are sitting face to face}. Then $\mathbb{P}(A) =?$

Proof Just give the males a permutation $1, 2, \ldots, n$, the females a permutation $i_1, i_2 \ldots i_n$. Then $\mathbb{P}(A)$=$\mathbb{P}(\{$there is at least one k so that $i_k = k\}$=$\mathbb{P}(\bigcup\limits_{k=1}^{n} A_k)$, where $A_k = \{i_k = k\}$. Then we have

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\bigcup_{k=1}^{n} A_k) \\
&= \sum_{k=1}^{n}(-1)^{k+1} \sum_{1\le i_1 < i_2 \cdots < i_k \le n} \mathbb{P}(A_{i_1} \bigcap A_{i_2} \cdots \bigcap A_{i_k}) \\
&= \sum_{k=1}^{n}(-1)^{k+1} \frac{(n-k)!}{n!} C_n^k \\
&= \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!}
\end{aligned}$$

Example 1.6 In the beginning, the player has k points and the dealer has $n - k$ points. Toss a coin each time, if the result is head, then the player gains a point, otherwise the player lose a point. The gamble doesn't end until the player or the dealer has no points. Then $\mathbb{P}(A) = \mathbb{P}(\{$it's the player who has no point in the end$\}$)=?

Proof Let A_i be the event that it's the player who has no point in the end when the player has i points in the beginning, B be the event that the first toss turns out to be head, then

$$\mathbb{P}(A_i) = \mathbb{P}(A_i|B)\mathbb{P}(B) + \mathbb{P}(A_i|B^c)\mathbb{P}(B^c)$$

That is

$$\mathbb{P}(A_i) = \frac{1}{2}\mathbb{P}(A_{i+1}) + \frac{1}{2}\mathbb{P}(A_{i-1})$$

Combine with the initial values $\mathbb{P}(A_0) = 1, \mathbb{P}(A_n) = 0$, we obtain that $\mathbb{P}(A_k) = 1 - \frac{k}{n}$.

Example 1.7 There are b black balls and r red balls in one box, pick one ball then put it back and put c balls that have the same color with the ball picked. $B_n = \{$ the nth picked ball is black$\}$. Then $\mathbb{P}(B_n)$=?

Proof Define R_n in the same way as B_n. Straightforward computation shows that $\mathbb{P}(B_1B_2R_3) = \mathbb{P}(R_1B_2B_3) = \mathbb{P}(B_1R_2B_3)$. Let $A_k(b) = \{k$ black balls are picked in the first n picked$\}$, then it's clear that each case has the same probability, which gives

$$A_k(b) = \frac{b(b+c)\dots(b+(k-1)c)r(r+c)\dots(r+(n-k-1)c)}{(b+r)(b+r+1)\dots(b+r+(n-1)c)}C_n^k$$

Then we have

$$\begin{aligned}
\mathbb{P}(B_{n+1}) &= \sum_{k=1}^{n} \mathbb{P}(A_k)\mathbb{P}(B_{n+1}|A_k) \\
&= \sum_{k=1}^{n} \frac{b(b+c)\dots(b+(k-1)c)r(r+c)\dots(r+(n-k-1)c)}{(b+r)(b+r+1)\dots(b+r+(n-1)c)}C_n^k \frac{b+kc}{b+r+nc} \\
&= \frac{b}{b+r}\sum_{k} \frac{(b+c)(b+c+c)\dots(b+c+(k-1)c)r(r+c)\dots(r+(n-k-1)c)}{(b+c+r)(b+c+r+1)\dots(b+c+r+(n-1)c)}C_n^k \\
&= \frac{b}{b+r}
\end{aligned}$$

Some identities above are just some calculations, the details are omitted here.

1.4 Random Variables

Definition 1.6

A random variable *is a function* $X : \Omega \to \mathbb{R}$ *with the property that* $\{\omega \in \Omega | X(\omega) \le x\} \in \mathcal{F}$ *for each* $x \in \mathbb{R}$. ♣

Definition 1.7

A probability distribution function *of a random variable* X *is a function* $F : \mathbb{R} \to [0, 1]$ *given by* $F(x) = \mathbb{P}(X \le x)$. ♣

Theorem 1.1

If F is a probability distribution function, then:

(i) If $x < y$, then $F(x) \le F(y)$

(ii) $\lim_{x\to\infty} F(x) = 1, \ \lim_{x\to-\infty} F(x) = 0$

(iii) F is right continuous, i.e. $\lim_{h \to 0^+} F(x+h) = F(x)$ ♡

Proof

(i) This is straightly from the fact that $\{X \le x\} \subset \{X \le y\}$

(ii) Let $A_n = \{X \le n\}$, then $\lim_{n\to\infty} F(n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n\to\infty} A_n) = 1$. The other identity is similar.

(iii) Let $B_n = \{X \le x + \frac{1}{n}\}$, then $\{B_n\}$ descends and $\bigcap_{n=1}^{\infty} B_n = \{X \le x\}$, then

$$\lim_{h\to 0^+} F(x+h) = \lim_{n\to\infty} F(x+\frac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} B_n) = F(x)$$

□

The function satisfying the (i)(ii)(iii) above is also called *distribution function*. In fact, every distribution function is the probability function of some random variable in some probability space.

Theorem 1.2

If F is a probability distribution function, then:

(i) $\mathbb{P}(X > x) = 1 - F(x)$

(ii) $\mathbb{P}(x < X \le y) = F(y) - F(x)$

(iii) $\mathbb{P}(x = y) = F(y) - F(y - 0)$ ♡

Definition 1.8

The minimal σ-field in $\mathbb{R}$ generated by the intervals in the form of $(a, b]$ is called the one dimensional Borel field, *denoted by $\mathcal{B}(\mathbb{R})$.* ♣

Then one can check that the one point set, the open intervals and the closed intervals are also in $\mathcal{B}(\mathbb{R})$.

Theorem 1.3

X is a random variable in the probability space $(X, \mathcal{F}, \mathbb{P})$, then $\forall B \in \mathcal{B}(\mathbb{R})$. One has

$$X^{-1}(B) := \{\omega \in \Omega | X(\omega) \in B\} \in \mathcal{F}$$

♡

Proof

Let $\mathcal{A} = \{A \subset \mathbb{R} | X^{-1}(A) \in \mathcal{F}\}$. Then $\mathcal{A}$ is a σ-field:

(i)$X^{-1}(\mathbb{R}) = \Omega \in \mathcal{F}$, so $\mathbb{R} \in \mathcal{A}$

(ii)If $A \in \mathcal{A}$, namely $X^{-1}(A) \in \mathcal{F}$, then $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$, so $A^c \in \mathcal{A}$

(iii)If $A_n(n = 1, 2 \cdots) \in \mathcal{A}$, namely $X^{-1}(A_n) \in \mathcal{F}$, then $X^{-1}(\bigcup\limits_n A_n) = \bigcup\limits_n X^{-1}(A_n) \in \mathcal{F}$, so $\bigcup\limits_n A_n \in \mathcal{F}$.

By the definition of X, $(-\infty, x] \in \mathcal{A}$, then $(a, b] \in \mathcal{A}$. By the minimal property of $\mathcal{B}(\mathbb{R})$, $\mathcal{B}(\mathbb{R}) \subset \mathcal{A}$. □

Theorem 1.4

If X, Y are random variables, then so is $X + Y$. ♡

Proof Just check that $\forall x \in \mathbb{R}$, $\{X + Y \leq x\} = \bigcap\limits_{r \in \mathbb{Q}} (\{X \leq r\} \bigcup \{Y \leq x - r\}) \in \mathcal{F}$.

1.5 Random Vectors

Definition 1.9

$X_1, X_2 \cdots, X_n$ are the random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then the vector $\overrightarrow{X} = (X_1, X_2, \cdots, X_n)$ is called the n dimensional random vector. *And the function $F(x_1, x_2, \cdots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$ is the* joint distribution function *of $\overrightarrow{X}$.* ♣

Focusing on the case where $n = 2$, we have the following properties of the joint distribution function $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$.

Theorem 1.5

(i) If $(x_1, y_1) \leq (x_2, y_2)$, then $F(x_1, y_1) \leq F(x_2, y_2)$

(ii) F is right continuous, i.e. $\lim\limits_{u \to 0^+, v \to 0^+} F(x+u, y+v) = F(x, y)$

(iii) $\lim\limits_{x \to -\infty} F(x, y) = \lim\limits_{y \to -\infty} F(x, y) = 0, \lim\limits_{x, y \to \infty} F(x, y) = 1$

(iv) $\mathbb{P}(x \in (x_1, x_2], y \in (y_1, y_2]) = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1)$ ♡

The properties (ii)(iii) and (iv) can determine a joint distribution, for (ii)(iii) and (iv) together imply (i). However, the properties (i)(ii) and (iii) together can't determine a joint distribution function. For example, the function

$$F(x, y) = \begin{cases} 1 & x + y \geq 0 \\ 0 & x + y < 0 \end{cases}$$

satisfies (i)(ii) and (iii), but doesn't satisfy (iv).

Definition 1.10

If $\overrightarrow{X} = (X_1, X_2, \cdots, X_n)$ can only take values in some countably subset of $\mathbb{R}^n$, then $\overrightarrow{X}$ is said to be a discrete random vector. *And $f(x_1, x_2, \cdots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$ is called the* joint mass function *of $\overrightarrow{X}$.* ♣

Definition 1.11

If there exists an integrable function $f : \mathbb{R}^n \to [0, \infty)$, *so that*

$$F(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, u_2, \cdots, u_n) du_1 \cdots du_n$$

Then $\overrightarrow{X}$ *is called to be a* continuous random vector, *and* f *is called to be the* joint density function. ♣

Definition 1.12

$\overrightarrow{X} = (X_1, X_2, \cdots, X_n)$ *is a random vector, for* $1 \le k \le n$, $(X_1, X_2, \cdots, X_k)$ *is the* marginal distribution *of* $\overrightarrow{X}$. *And* $\mathbb{P}(X_1 \le x_1, X_2 \le x_2, \cdots, X_k \le x_k)$ *is the* marginal distribution function *of* $F(x_1, x_2, \cdots, x_n)$. ♣

It's clear that $\mathbb{P}(X_1 \le x_1, X_2 \le x_2, \cdot, X_k \le x_k) = \lim\limits_{x_{k+1}, \cdots, x_n \to -\infty} F(x_1, x_2, \cdots, x_n)$.

Now we focus on the case where $n = 1$, assume X is a continuous random variable, and f is the density function of X. Then

(i) $\frac{F(x_0+\Delta x)-F(x_0)}{\Delta x} = \frac{1}{\Delta x}\int_{x_0}^{x_0+\Delta x} f(x)dx$. If f is continuous at x, then $\mathbb{P}(x_0 < x \le x_0 + \Delta x) \sim \Delta x f(x_0)$.

(ii) $F(x) = \int_{-\infty}^{x} f(u)du$. The density function is not unique, since changing finitely many values of f doesn't change the distribution of X.

(iii) $\mathbb{P}(x = a) \le \int_{a-\frac{1}{n}}^{a} f(u)du \to 0$, so $\mathbb{P}(x = a) = 0$.

(iv) If F is continuous, and $F'(x)$ exists and is continuous except for finitely many x. Then F is a continuous distribution function, and F' is the density function.

Chapter 2 Discrete Random Variables

2.1 Classical Distributions and Independence

Example 2.1 The discrete random variable X is said to have the *binomial distribution* with parameters n and p, written $B(n, p)$, if the mass function of X is

$$f(k) = \binom{n}{k} p^k q^{n-k} \quad 0 \leq k \leq n, p + q = 1$$

Example 2.2 The discrete random variable X is said to have the *geometric distribution* with parameter $p(0 < p < 1)$,if the mass function of X is

$$f(k) = pq^{k-1} \quad k \in \mathbb{N}^*$$

Theorem 2.1

X is a random variable that takes values in $\mathbb{N}^*$, and $\mathbb{P}(X = m + 1|X \geq m)$ is independent of m, then X has the geometric distribution. ♡

Proof

By the condition, we assume

$$\mathbb{P}(X = m + 1) = (1 - q)q\mathbb{P}(X \geq m)$$

where q is a constant. Also note that

$$\mathbb{P}(X = m + 1) = \mathbb{P}(X \geq m + 2) - \mathbb{P}(X \geq m + 1)$$

let $a_m = \mathbb{P}(X \geq m)$, then we have

$$(1 - q)qa_m = a_{m+1} - a_{m+2}$$

Then by the extra condition that $a_\infty = 1$, one can have that $a_m = q^{m-1}$, then $\mathbb{P}(X = m) =$

$a_m - a_{m+1} = pq^{m-1}$, where $p + q = 1$, which means that X has the geometric distribution.

Example 2.3 The discrete random variable X is said to have the *Poisson distribution* with parameter $\lambda(\lambda > 0)$,if the mass function of X is

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k \in \mathbb{N}$$

Definition 2.1

$X_1, X_2, \cdots, X_n$ *are random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$. *If*

$$\forall x_1, \cdots, x_n \in \mathbb{R}, \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_n = x_n)$$

, then we say $X_1, X_2, \cdots, X_n$ *are* independent. ♣

Lemma 2.1

$X_1, X_2, \cdots, X_n$ *are independent if and only if the distribution functions satisfy*

$$F(x_1, x_2, \cdots x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

♡

2.2 Expectation

Definition 2.2

The expectation *of the random variable X with mass function f is defined to be*

$$\mathbb{E}[X] = \sum_x x f(x)$$

when the sum is absolutely convergent. ♣

Theorem 2.2

$\mathbb{E}[g(X)] = \sum_x g(x)f(x)$ *when the sum on the right is absolutely convergent.* ♡

Proof Let $Y = g(X)$, then $f_Y(y) = \mathbb{P}(\bigcup_{x:y=g(x)} \{X = x\}) = \sum_{x:y=g(x)} f(x)$.

Thus $\mathbb{E}[Y] = \sum_y y \sum_{x:y=g(x)} f(x) = \sum_x g(x)f(x)$. □

Definition 2.3

The kth moment *m_k of X is defined to be $m_k = \mathbb{E}[X^k]$. The kth* central moment *of X is defined to be $\sigma_k = \mathbb{E}[(X-\mu)^k]$.* ♣

We also denote the expectation $\mathbb{E}[X] = m_1 = \mu$. We call the *variance* of X is $\mathrm{Var}(X) = \sigma_2 = \mathbb{E}[(X-\mu)^2]$. The *standard deviation* of X is $\sigma = \sqrt{\mathrm{Var}(X)}$. Then a simple computation gives that $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}[X])^2 = m_2 - \mu^2 \leq m_2$.

Example 2.4 X is of $B(n,p)$. Compute $\mathbb{E}[X], \mathbb{E}[X^2]$.

Proof

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k} \\ &= \sum_{k=0}^{n-1} n\binom{n-1}{k}p^{k+1}q^{n-k-1} = np\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}\binom{n}{k}p^k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2}\binom{n-2}{k}p^k q^{n-2-k} \\
&= n(n-1)p^2
\end{aligned}
$$

Then $\mathbb{E}[X^2] = np[1+(n-1)p], \mathrm{Var}(X) = np(1-p) = npq.$ □

Theorem 2.3

The expectation $\mathbb{E}$ can be regarded as a linear operator, that is:

(i)$X \geq 0$, *then* $\mathbb{E}[X] \geq 0$ *(ii)*$\mathbb{E}[1] = 1$

(iii)$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ *for* $a, b \in \mathbb{R}$.

Proof We only check (iii) here.

Let $A_x = \{X = x\}, B_y = \{Y = y\}$, then $X = \sum_x xI_{A_x}, Y = \sum_y yI_{B_y}$.

$aX + bY = a\sum_{x,y} IA_xB_y + b\sum_{x,y} yIA_xB_y = \sum_{x,y}(ax+by)I_{A_x+B_y}$.

Thus

$$
\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_{x,y}(ax+by)\mathbb{P}(A_xB_y) \\
&= a\sum_{x,y} x\mathbb{P}(A_xB_y) + b\sum_{x,y} y\mathbb{P}(A_xB_y) \\
&= a\sum_{x} x\mathbb{P}(A_x) + b\sum_{y} y\mathbb{P}(B_y) \\
&= a\mathbb{E}[X] + b\mathbb{E}[Y]
\end{aligned}
$$

Theorem 2.4

If X, Y are independent with $\mathbb{E}[|X|] < \infty, \mathbb{E}[|Y|] < \infty$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof $XY = \sum_{x,y} xyI_{A_x B_y}$, then

$$\mathbb{E}[XY] = \sum_{x,y} xy\mathbb{P}(A_x B_y)$$
$$= \sum_{x,y} xy\mathbb{P}(A_x)\mathbb{P}(B_y) = \mathbb{E}[X]\mathbb{E}[Y]$$

Theorem 2.5

(i)$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

(ii)$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$

In particular, when X, Y are independent, $Var(X) + Var(Y) = Var(X + Y)$.

Proof By the definition of variance, we have

(i)

$$\mathrm{Var}(aX + b) = \mathbb{E}[(aX + b)^2] - (\mathbb{E}[aX + b])^2$$
$$= \mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[X] + b)^2$$
$$= a^2\mathbb{E} + 2ab\mathbb{E}[X] + b^2 - a^2(\mathbb{E}(X))^2 - b^2 - 2ab\mathbb{E}[X] - b^2 = a^2\mathrm{Var}(X)$$

(ii)

$$\mathrm{Var}(X + Y) - \mathrm{Var}(X) - \mathrm{Var}(Y) = \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 - \mathbb{E}[X^2] + (\mathbb{E}[X])^2 - \mathbb{E}[Y^2] + (\mathbb{E}[Y])^2$$
$$= 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y]$$

Example 2.5 Let $x_k = \frac{1}{k}(-2)^k$, and $\mathbb{P}(X = x_k) = \frac{1}{2^k}(k = 1, 2, \cdots)$, then

$$\sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -\log 2$$

But $\mathbb{E}[X]$ does NOT exist since $\sum_{k=1}^{\infty} |x_k| p_k$ does not converge.

Definition 2.4

The covariance *of X, Y is defined to be*

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

The correlation of X, Y is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X.Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

♣

More generally, $\overrightarrow{X} = (X_1, X_2, \cdots, X_n)$, define its *covariance matrix* to be the $n \times n$ matrix $A = (\sigma_{ij})$. Each entry is $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$, then we have $A \geq 0$:

For any $t_i \in \mathbb{R}(1 \leq i \leq n)$

$$\begin{aligned}\sum_{i,j=1}^{n} t_i t_j \sigma_{ij} &= \sum_{i,j=1}^{n} t_i t_j \mathbb{E}[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] \\ &= \mathbb{E}[\sum_{i=1}^{n} t_i (X_i - \mu_{X_i}) \sum_{j=1}^{n} t_j (X_j - \mu_{X_j})] \\ &= \mathbb{E}[(\sum_{i=1}^{n} t_i (X_i - \mu_{X_i})^2] \geq 0\end{aligned}$$

2.3 Probabilistic Method

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then for any $A \in \mathcal{F}$, the indicator function $I_A : \Omega \to \mathbb{R}$, maps ω to 1 if $\omega \in A$, to 0 otherwise. Then a trivial observation gives that $\mathbb{E}[I_A] = \mathbb{P}(A)$.

Example 2.6 Let S_n denote the set of all the n-permutations, $|S_n| = n!$. $\forall \sigma \in S_n$, define $N(\sigma)$ =the number of fixed points under σ.

Let A_i denote the event that i is a fixed point, and I_i its indicator function. Set

$$X = \sum_{i_1 < \cdots < i_r} I_{i_1} \cdots I_{i_r} (1 - I_{i_{r+1}}) \cdots (1 - I_{i_n})$$

Then it's clear that

$$
\begin{aligned}
\mathbb{P}(N = r) &= \mathbb{E}[X] \\
&= \frac{\binom{n}{r}}{n!}(n-r)!\sum_{k=0}^{n-r}\frac{(-1)^k}{k!} \\
&= \frac{1}{r!}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}
\end{aligned}
$$

$$
\mathbb{E}[N] = \mathbb{E}[\sum_{k=1}^{n} I_k] = n\mathbb{E}[I_1] = 1
$$

$$
Var(N) = \sum_{k,l=1}^{n}\mathbb{E}[I_kI_l] - 1 = n\mathbb{E}[I_1^2] + n(n-1)\mathbb{E}[I_1I_2] - 1 = 1
$$

Example 2.7 A 17-gon has exactly five vertices painted red. Prove that there must exist seven adjacent vertices, so that there are three red vertices among them.

Proof Let $\Omega = \{1, 2, \cdots, 17\}, a_i = 1$ if the i-th vertex is red, and $a_i = 0$ otherwise. And set $X(k) = \sum_{i=1}^{7} a_{k+i}$(we identify 18 with 1, 19 with 2 and so on).

Then $\mathbb{E}[X] = \sum_{k=1}^{17}\frac{1}{17}(a_{k+1} + a_{k+2} + \cdots + a_{k+7}) = \frac{35}{17} > 2$. Therefore there must exist a k_0, so that $X(k_0) > 2$. Since X take value in integers, so $X(k_0) \geq 3$.

2.4 Conditional Expectation

Definition 2.5

If random variables X, Y satisfy that $\mathrm{Cov}(X, Y) = 0$*, then we say that X, Y are* uncorrelated. ♣

Theorem 2.6

(i) $|\rho(X, Y)| \leq 1$

(ii) If X, Y are independent or uncorrelated, then $\rho(X, Y) = 0$

(iii) $\rho(X, Y) = \pm 1$ *if and only if* $\exists a, b \in \mathbb{R}$*, so that* $\mathbb{P}(aX + b = Y) = 1$ ♡

Lemma 2.2

$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$ ♡

Proof (i) If $\mathbb{E}[X^2] = 0$, then $\sum\limits_x x^2 f_X(x) = 0$, so $f_X(x) = 0$ whenever $x \neq 0$, which means $\mathbb{P}(X = 0) = 1$.

Since $f_X(x) = \sum\limits_y f(x, y)$, we have $f(x, y) = 0$ when $x \neq 0$, then $\mathbb{E}[XY] = \sum\limits_{x,y} xyf(x, y) = 0$.

(ii) If $\mathbb{E}[X^2] \neq 0$, then $\mathbb{E}[(Y - tX)^2] = t^2\mathbb{E}[X^2] - 2t\mathbb{E}[XY] = t^2\mathbb{E}[Y^2] \geq 0$. Thus

$$\Delta = 4((\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2]) \leq 0$$

The equality con be obtained if and only if $\exists t \in \mathbb{R}$, so that $\mathbb{E}[Y - tX] = 0$, which means $\mathbb{P}(Y = tX) = 1$ by (i).

Note that we used the two-variable version of Theorem 2.2.2 in the proof above.

Example 2.8 $\overrightarrow{X} = (X_1, X_2, \cdots, X_r), \mathbb{P}(X_1 = k_1, \cdots, X_r = k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}$, where $\sum\limits_i p_i = 1, \sum\limits_i k_i = n$. Compute $\mathrm{Cov}(X_i, X_j), \rho(X_i, X_j)$ for $i \neq j$.

(Identity: $\sum\limits_{k_1 + \cdots + k_r = n} x_1^{k_1} \cdots x_r^{k_r} \frac{n!}{k_1! \cdots k_r!} = (x_1 + \cdots + x_r)^n$)

Proof

$$\begin{aligned}\mathbb{P}(X_i = k_i) &= \sum_{k_1+\cdots+\hat{k_i}+\cdots+k_r=n-k_i} \frac{n!}{k_1!\cdots\cdots k_r!} p_1^{k_1}\cdots p_r^{k_r} \\ &= \frac{n!}{k_i!(n-k_i)!} p_i^{k_i}(1-p_i)^{n-k_i}\end{aligned}$$

So X_i is of $B(n, p_i)$, similarly one can check that $X_i + X_j$ is of $B(n, p_i + p_j)$ for $i \neq j$. Then

$$\mathrm{Cov}(X_i, X_j) = \frac{1}{2}(\mathrm{Var}(X_i + X_j) - \mathrm{Var}(X_i) - \mathrm{Var}(X_j)) = -np_ip_j$$

$$\rho(X_i, X_j) = \frac{-p_ip_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}$$

Definition 2.6

(X, Y) are discrete random variables, when $f_X(x) > 0$, give the distribution of Y under $X = x$:

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}, F_{Y|X}(y) = \mathbb{P}(Y \leq y|X = x)$$

Then the conditional expectation *of Y under $X = x$ is*

$$\psi(x) = \mathbb{E}[Y|X = x] = \sum_y \mathbb{P}(Y \leq y|X = x)$$

And we call $\psi(X)$, which is also a random variable, the conditional expectation of Y under X, which is denoted by $\mathbb{E}[Y|X]$. ♣

Theorem 2.7

$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$ ♡

Proof

$$\begin{aligned}LHS = \mathbb{E}[\psi(X)] &= \sum_x \psi(x) f_X(x) \\ &= \sum_x f_X(x) \sum_y y f_{Y|X}(y|x) \\ &= \sum_x \sum_y y f(x, y) \\ &= \sum_y y f_Y(y) = \mathbb{E}[Y] = RHS\end{aligned}$$

Theorem 2.8
For the "good" measurable function $g : \mathbb{R} \to \mathbb{R}$(the word good here means that g makes the two sides of the below identity have meaning, which requires the two sides to be absolutely convergent), we have

$$\mathbb{E}[g(X)\psi(X)] = \mathbb{E}[Yg(X)]$$

♡

Example 2.9 A bird can lay N eggs, where N is of the Poisson distribution with parameter λ. Each egg has independently the probability p to become a bird. Let K to be the number of birds that come out in the end. Calculate $\mathbb{E}[K|N], \mathbb{E}[K]$ and $\mathbb{E}[N|K]$.

Proof $\mathbb{E}[K|N = n] = np$, thus $\mathbb{E}[K|N] = pN$, $\mathbb{E}[K] = \mathbb{E}[\mathbb{E}[K|N]] = \mathbb{E}[pN] = p\lambda$.

$$\begin{aligned}
f_{N|K}(n|k) &= \frac{\mathbb{P}(K = k|N = n)\mathbb{P}(N = n)}{\sum\limits_{m=k}^{\infty} \mathbb{P}(K = k|N = m)\mathbb{P}(N = m)} \\
&= \frac{\frac{n!}{(n-k)!k!} p^k q^{n-k} \frac{\lambda^n}{n!} e^{-\lambda}}{\sum\limits_{m=k}^{\infty} \frac{m!}{(m-k)!k!} p^k q^{m-k} \frac{\lambda^m}{m!} e^{-\lambda}} \\
&= \frac{\frac{n!}{(n-k)!k!} p^k q^{n-k} \frac{\lambda^n}{n!} e^{-\lambda}}{\sum\limits_{m=0}^{\infty} \frac{1}{m!k!} p^k q^m \lambda^{m+k} e^{-\lambda}} \\
&= \frac{(q\lambda)^{n-k} e^{-\lambda q}}{(n-k)!}
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}[N|K = k] &= \sum_{n=k}^{\infty} \frac{n(q\lambda)^{n-k} e^{-\lambda q}}{(n-k)!} \\
&= \sum_{n=0}^{\infty} \frac{(n+k)(\lambda q)^n e^{-\lambda q}}{n!} \\
&= k + \lambda q
\end{aligned}$$

Thus $\mathbb{E}[N|K] = K + \lambda q$.

2.5 Random Walk

Let $S_0 = a \in \mathbb{Z}^d$, and $S_n = a + \sum_{k=1}^{n} X_k$ with X_k's are independently and identically distributed, then the scene is called the *random walk*. When $d = 1, \mathbb{P}(X_k = 1) = p, \mathbb{P}(X_k = -1) = q, p + q = 1$, then it's called *simple random walk*. Moreover, if $p = \frac{1}{2}$, then it's called the *symmetric simple random walk*.

Theorem 2.9

(i) (Spatially Homogeneous) $\mathbb{P}(S_n = j + b | S_0 = a + b) = \mathbb{P}(S_n = j | S_0 = a)$

(ii) (Temporally Homogeneous) $\mathbb{P}(S_{n+m} = j | S_m = a) = \mathbb{P}(S_{n+m} = j | S_m = j_m)$

(iii) (Markov Property) $\mathbb{P}(S_{n+m} = j | S_0 = j_0, \cdots, S_m = j_m) = \mathbb{P}(S_{n+m} = j | S_m = j_m)$ ♡

Proof (i)

$$
\begin{aligned}
LHS &= \frac{\mathbb{P}(S_n = j + b, S_0 = a + b)}{\mathbb{P}(S_0 = a + b)} = \frac{\mathbb{P}(\sum_{k=1}^{n} X_k = j - a, S_0 = a + b)}{\mathbb{P}(S_0 = a + b)} \\
&= \mathbb{P}(\sum_{k=1}^{n} X_k = j - a) = RHS
\end{aligned}
$$

(ii)

$$
LHS = \mathbb{P}(\sum_{k=m+1}^{m+n} X_k = j - a) = \mathbb{P}(\sum_{k=1}^{n} X_k = j - a) = RHS
$$

(iii)

$$
\begin{aligned}
LHS &= \frac{\mathbb{P}(S_{n+m} = j, S_0 = j_0, \cdots, S_m = j_m)}{\mathbb{P}(S_0 = j_0, \cdots, S_m = j_m)} \\
&= \frac{\mathbb{P}(\sum_{k=m+1}^{m+n} X_k = j - j_m, S_0 = j_0, \cdots, S_m = j_m)}{\mathbb{P}(S_0 = j_0, \cdots, S_m = j_m)} \\
&= \mathbb{P}(\sum_{k=m+1}^{m+n} X_k = j - j_m) = RHS
\end{aligned}
$$

Now introduce the symbols that we will use from now on. We see the simple random walk on a plane, with the two axes representing the time and the position. Let $N_n(a, b)$ be the number of paths from $(0,a)$ to (n,b). Let $N_n^0(a,b)$ ne the number of paths from $(0,a)$ to (n,b) that passes the time axis.

Lemma 2.3

$N_n(a,b) = \binom{n}{\frac{n+b-a}{2}}$

Lemma 2.4

If $a, b > 0$, then $N_n^0(a,b) = N_n(-a,b)$.

This lemma is also referred to as *reflection principle*. This can be understood intuitively by reflecting the part above the time axis before the first time the path passes the axis.

Proposition 2.1

If $b > 0$, then the number of paths from $(0,0)$ to (n,b) that doesn't pass the time axis except the starting point is $\frac{b}{n}N_n(0,b)$.

Proof Clearly the first step must be towards right. Thus the answer is equal to the number of paths from $(1,1)$ to (n,b) that doesn't pass the time axis, which is also equal to the number of the kind of paths from $(0,1)$ to $(n-1,b)$, which is

$$N_{n-1}(1,b) - N_{n-1}^0(1,b) = N_{n-1}(1,b) - N_{n-1}(-1,b) = \frac{b}{n}N_n(0,b)$$

The last step is by some simple calculation with Lemma 2.5.2.

Theorem 2.10

$S_0 = 0, n \geq 1$, then

$$\mathbb{P}(S_1 \cdots S_n \neq 0, S_n = b) = \frac{|b|}{n}\mathbb{P}(S_n = b).$$

As a consequence,

$$\mathbb{P}(S_1 \cdots S_n \neq 0) = \frac{1}{n}\mathbb{E}(|S_n|)$$

Proof Just assume $b > 0$, then

$$\mathbb{P}(S_1 \cdots S_n \neq 0, S_n = b) = \frac{b}{n}N_n(0,b)p^{\frac{n+b}{2}}q^{\frac{n-b}{2}} = \frac{b}{n}\mathbb{P}(S_n = b)$$

2.6 Generating Function

For a sequence $\{a_n\}_{n=0}^{\infty}$, define the *generating function* of $\{a_n\}$ to be $G_a(s) = \sum_{n=0}^{\infty} a_n s^n$. For two sequences $\{a_n\}$ and $\{b_n\}$, the *convolution* of them(denoted as $a+b$) is to be the sequence $\{c_n\}$, where $c_n = \sum_{i+j=n} a_i b_j$. A nice property of the convolution is that $G_c(s) = G_a(s)G_b(s)$.

Example 2.10 Consider the symmetric simple random walk, compute the probability $\mathbb{P}(S_0 = S_{2n} = 0, S_i \geq 0, i = 1, \cdots, 2n-1)$.

Proof Denote the probability by c_n.

Then by considering the first hitting time of the x-axis $2k(k = 1, \cdots, n)$, we can get

$$c_n = \sum_{k=1}^{n} c_{k-1}c_{n-k} = \sum_{k=0}^{n-1} c_k c_{n-1-k}, c_0 = 1$$

Now define the corresponding generating function $G(s) = \sum_{n=0}^{\infty} c_n s^n$, then $\frac{G(s)-1}{s} = \sum_{n=1}^{\infty} c_n s^{n-1} = G(s)G(s)$. Then last equality comes from the property of convolution. Since the $G(s)$ should be defined for $s \to 0$, we can have

$$G(s) = \frac{1 - \sqrt{1-4s}}{2s}$$

Then by the Taylor expansion, we have

$$\begin{aligned}
G(s) &= \frac{-1}{2s}\left(\sum_{n=1}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-n+1)}{n!}(-4s)^n\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{2s}\frac{(2n-3)\cdots 1}{n!}(2s)^n \\
&= \frac{1}{2s}\sum_{n=0}^{\infty} \frac{(2n-1)!!}{(n+1)!}(2s)^{n+1} \\
&= \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(n+1)!}2^n s^n \\
&= \sum_{n=0}^{\infty} \frac{(2n-1)!}{n!(n+1)!}s^n
\end{aligned}$$

Thus we have $c_n = \frac{1}{n+1}\binom{2n}{n}$.

In the following we will always assume that X is a non-negative integer-valued random variable.

Definition 2.7

The function $G_X(s) = \mathbb{E}[s^X]$ *is called the* (probability) generating function *of* X.

Note the $G_X(s) = \sum_{k=0}^{\infty} s^k P(X - k)$. It's clear that the convergence radius $R \geq 1$, and the coefficients of $G_X(s)$ are all non-negative.

Example 2.11 (i) $X \sim B(n,p)$, then $G_X(s) = \sum_k \binom{n}{k} p^k q^{n-k} s^k = (sp+q)^n$.

(ii) $P(X = k) = pq^{k-1}(k = 1, 2, \cdots)$, then $G_X(s) = \sum_{k=1}^{\infty} pq^{k-1}s^k = \frac{ps}{1-qs}$.

(iii) $X \sim P(\lambda)$, then $G_X(s) = \sum_k \frac{s^k \lambda^k}{k!} e^{-\lambda} = e^{\lambda s - \lambda}$.

Theorem 2.11

(i) $\mathbb{E}[X] = G'(1)$

(ii) $\mathbb{E}[X(X-1)\cdots(X-k+1)] = G^{(k)}(1)$

(iii) $\mathrm{Var}(X) = G''(1) + G'(1) - G'(1)^2$

The derivative $G^{(k)}(1)$ *is actually* $\lim_{s \to 1^-} G^{(k)}(s)$, *which is well defined by the Abel's Theorem.*

Theorem 2.12

If $X_1, \cdots, X_n$ *are independent, then* $G_{S_n}(s) = \prod_{k=1}^{n} G_{X_k}(s)$, *where* $S_n = \sum_{k=1}^{n} X_k$.

Proof One can check the fact that if X, Y are independent, then $g(X), h(Y)$ are independent(c.f. 3.11.1(1)), thus for any s, $s^{X_1}, \cdots, s^{X_n}$ are independent.

Then

$$G_{S_n}(s) = \mathbb{E}[s^{X_1} \cdots s^{X_n}] = \mathbb{E}[s^{X_1}] \cdots \mathbb{E}[s^{X_n}] = G_{X_1}(s) \cdots G_{X_n}(s)$$

Theorem 2.13

If $\{X_k\}$ *are independent, and* N *is independent with* $\{X_k\}$, *then for* $S_N = \sum_{k=1}^{N} X_k$, *we have*

$$G_{S_N}(s) = G_N(G_{X_1}(s))$$

Proof By using conditional expectation, we have

$$\begin{aligned} G_{S_N}(s) &= \mathbb{E}[s^{S_N}] = \mathbb{E}[\mathbb{E}[s^{S_N}|N]] \\ &= \sum_n f_N(n)\mathbb{E}[s^{S_N}|N = n] \\ &= \sum_n f_N(n)(G_{X_1}(s))^n = G_N(G_{X_1}(s)) \end{aligned}$$

Definition 2.8

The joint (probability) generating function *of* X, Y *is* $G_{X,Y}(s,t) = \mathbb{E}[s^X t^Y]$. ♣

Theorem 2.14

X, Y are independent if and only if $G_X(s)G_Y(t) = G_{X,Y}(s,t)$. ♡

Proof $G_X(s)G_Y(t) = G_{X,Y}(s,t)$ is equivalent to

$$\begin{aligned} \sum_{i,j}^{\infty} \mathbb{P}(X = i, Y = j)s^i t^j &= \sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i \sum_{j=0}^{\infty} \mathbb{P}(Y = j)t^j \\ &= \sum_{i,j}^{\infty} \mathbb{P}(X = i)\mathbb{P}(Y = j)s^i t^j \end{aligned}$$

which is equivalent to the independence of X, Y.

Example 2.12 Throw a fair dice k times, what is the probability that the sum of the numbers is 9?

Proof X_i is the number of the i-th dice, and $S_k = \sum_k X_k$, then

$$G_{S_k}(s) = (G_{X_1}(s))^k = (\frac{1}{6}s\frac{1 - s^6}{1 - s})^k$$

Consider the case where $k = 3$, then

$$G_{S_3}(s) = \frac{s^3}{6^3}\frac{(1 - s^6)^3}{(1 - s)^3} = \frac{s^3}{6^3}(1 - 3s^6 + 3s^{12} - s^{18})\sum_{n=0}^{\infty}\binom{n}{-3}(-s)^n$$

The coefficient of s^9 is $\frac{1}{6^3}(\binom{6}{-3} - 3) = \frac{25}{216}$, which is the probability we want.

For more general random variables(the values are no longer non-negative integers), we can also define the generating function.

Definition 2.9

The moment generating function *of* X *is* $M_X(t) = \mathbb{E}[e^{tX}]$. ♣

Chapter 3 Continuous Random Variables

3.1 Basic Definitions

Recall that a random variable X is *continuous* if for its distribution function F, there exists a non-negative and integrable function $f : \mathbb{R} \to \mathbb{R}$, so that $F(x) = \int_{-\infty}^{x} f(u)du$. Such a function f is called the *density function* of X. Then it's clear that a density function should satisfy that $\int_{-\infty}^{\infty} f(x)dx = 1$. If X has density function f, then for any Borel measurable set B, then

$$\mathbb{P}(X \in B) = \int_B f(x)dx$$

In particular, $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$, and $\mathbb{P}(X = x) = 0$ for all x.

Example 3.1 X is *uniform* on $[a, b]$, denoted by $X \sim U[a, b]$, if $f(x) = \frac{1}{b-a}$ for any $x \in [a, b]$.

Example 3.2 X is *exponential* with parameter λ, denoted by $X \sim \mathrm{Exp}(\lambda)$, if $f(x) = \lambda e^{-\lambda x}$ for any $x \geq 0$. Then the distribution function of X is $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.

Example 3.3 X is of *normal* distribution with parameters μ, σ, denoted by $X \sim N(\mu, \sigma^2)$, if $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. When $\mu = 0, \sigma = 1$, then the distribution is called the *standard* normal distribution. When $x = \mu$, $f(x)$ attains its maximum.

Example 3.4 $f(x) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}(-2\sigma \leq x \leq 2\sigma)$ is the density function in the *Wigner Semicircular Law*.

Definition 3.1
$X_1, X_2, \cdots, X_n$ are random variables in $(\Omega, \mathcal{F}, \mathbb{P})$. They are called independent *if*

$$\forall x_1, \cdots, x_n \in \mathbb{R}, F(x_1, \cdots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

♣

Theorem 3.1

$X_1, \cdots, X_n$ are independent if and only if

$$\forall B_1, \cdots, B_n \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}(X_1 \in B_1, \cdots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i)$$

♡

Theorem 3.2

If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, $X_1, \cdots, X_n$ are independent, then $g(X_1), \cdots, g(X_n)$ are independent.

♡

Proof Let $Y_i = g(X_i)$, then we have

$$\begin{aligned}
\mathbb{P}(Y_i \in (-\infty, y_i], \forall i = 1, \cdots, n) &= \mathbb{P}(X_i \in g_i^{-1}((-\infty, y_i]), \forall i = 1, \cdots, n) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in g_i^{-1}((-\infty, y_i]) \\
&= \prod_{i=1}^{n} \mathbb{P}(y_i \in (-\infty, y_i])
\end{aligned}$$

Thus $Y_1, \cdots, Y_n$ are independent.

Theorem 3.3

$X_1, \cdots, X_n$ have density functions $f_1, \cdots, f_n$ respectively, then $X_1, \cdots, X_n$ are independent if and only if the joint density function $f(\overrightarrow{x}) = \prod_{i=1}^{n} f_i(x_i)$.

♡

If X_1, X_2 has a joint density function f, D is an area in $\mathbb{R}^2$, and T is an one-to-one mapping that takes (x_1, x_2) to (y_1, y_2). Then invert the transformation into $x_1 = g_1(y_1, y_2), x_2 = g_2(y_1, y_2)$, and define

$$J = \det \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} \end{pmatrix}$$

Proposition 3.1

The joint density function of Y_1, Y_2 at (y_1, y_2) is

$$|J| f(g_1(y_1, y_2), g_2(y_1, y_2)) I_{T(D)}$$

♠

Proof For any $B = T(A)$, we can have

$$\begin{aligned}
\mathbb{P}((Y_1, Y_2) \in B) &= \mathbb{P}((X_1, X_2) \in A) \\
&= \int\int_A f(x_1, x_2) dx_1 dx_2 \\
&= \int\int_B f(g_1(y_1, y_2), g_2(y_1, y_2))|J| dy_1 dy_2
\end{aligned}$$

Take $B = ((-\infty, y_1] \times (-\infty, y_2]) \bigcap T(D)$, then the result follows.

3.2 (Conditional) Expectation

Definition 3.2

X has a density function f. When $\int_{\mathbb{R}} |x| f(x)dx < \infty$, we call $\int_{\mathbb{R}} x f(x)dx$ the expectation *of X. Define the moment, variance, covariance, correlation similarly.* ♣

Lemma 3.1

X has a distribution function F, then $\mathbb{E}[X] = \int_0^\infty 1 - F(x)dx - \int_{-\infty}^0 F(x)dx$ ♡

Proof

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \int_0^x 1 dt f(x)dx - \int_{-\infty}^0 \int_x^0 1 dt f(x)dx \\
&= \int_0^\infty \int_t^\infty f(x)dxdt - \int_{\infty}^0 \int_{\infty}^t f(x)dxdt \\
&= \int_0^\infty 1 - F(x)dx - \int_{-\infty}^0 F(x)dx
\end{aligned}$$

Theorem 3.4

If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, $X, g(X)$ are continuous random variables and their expectations exist, then

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

♡

Proof

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_0^\infty \mathbb{P}(g(x) > t)dt - \int_\infty^0 \mathbb{P}(g(x) \le t)dt \\
&= \int_0^\infty \int_{g(x)>t} f(x)dxdt - \int_{-\infty}^0 \int_{g(x)\le t} f(x)dxdt \\
&= \int_{g(x)>0} \int_0^{g(x)} 1dtf(x)dx - \int_{g(x)\le 0} \int_{g(x)}^0 1dtf(x)dx \\
&= \int_{-\infty}^\infty g(x)f(x)dx
\end{aligned}$$

Theorem 3.5

(X, Y) is a continuous random vector, $g : \mathbb{R}^2 \to \mathbb{R}^2$ is Borel measurable, and $g(X, Y)$ is continuous and the expectation exists. Then

$$\mathbb{E}[g(X,Y)] = \int\int g(x,y)f(x,y)dxdy$$

In particular, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

♡

Theorem 3.6

$|E[XY]| \le \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$. *Consequently,* $|\rho(X,Y)| \le 1$.

♡

Example 3.5 $X \sim N(\mu, \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. Then

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

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

Example 3.6 Cauchy distribution is with the density function $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Its expectation doesn't exist.

Definition 3.3

(X, Y) has the joint density $f(x, y)$, the conditional density *is* $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$. *The* conditional expectation *is*

$$\psi(x) = \mathbb{E}[Y|X = x] = \int_{\infty}^{\infty} y f_{Y|X}(y|x) dx$$

♣

Theorem 3.7

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

♡

Proof

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_X(x)} dy dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \\
&= \int_{-\infty}^{\infty} y f_Y(y) dy = \mathbb{E}[Y]
\end{aligned}$$

More generally, for any good function g, we have $\mathbb{E}[g(X)\psi(X)] = \mathbb{E}[Y g(X)]$.

Example 3.7

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

The distribution is called *bivariate standard normal distribution*, where $\rho \in (-1, 1)$.

$$\begin{aligned}
f_X(x) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int e^{-\frac{1}{2}x^2 - \frac{(y-\rho x)^2}{2(1-\rho^2)}} d(y - \rho x) \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}
\end{aligned}$$

So $X \sim N(0, 1)$, similarly $Y \sim N(0, 1)$.

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \int\int xy f(x, y) dx dy \\
&= \int\int (x(y - \rho x) + \rho x^2) f(x, y) dx dy = 0 + \rho = \rho
\end{aligned}$$

Thus ρ is actually the correlation of X and Y.

$$f_{Y|X}(y|x) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$

$$\mathbb{E}[Y|X=x] = \rho x, \mathbb{E}[Y|X] = \rho X$$

3.3 Multivariate Normal Distribution

Definition 3.4

The random vector $\overrightarrow{X} = (X_1, \cdots, X_n)$ *has the* multivariate normal distribution *if, written* $N(\overrightarrow{\mu}, \sum)$*, if its density is*

$$f(\overrightarrow{x}) = \frac{1}{\sqrt{(2\pi)^n |\sum|}} e^{-\frac{1}{2}(\overrightarrow{x}-\overrightarrow{\mu})\sum^{-1}(\overrightarrow{x}-\overrightarrow{\mu})^T}$$

where $\sum$ *is a positive definite symmetric matrix.* ♣

Theorem 3.8

$$\mathbb{E}[\overrightarrow{X}] = \overrightarrow{\mu}, \mathbb{E}[(\overrightarrow{X} - \overrightarrow{\mu})(\overrightarrow{X} - \overrightarrow{\mu})^T] = \sum$$

In other words, $\mathbb{E}[X_i] = \mu_i, \mathrm{Cov}(X_i, X_j) = \sigma_{i,j}$. ♡

Proof There exists an orthogonal matrix B, so that $\sum = B^T \Lambda B$, where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. Let $\overrightarrow{x} = \overrightarrow{\mu} + \overrightarrow{y} B$, then

$$\begin{aligned} \int f(\overrightarrow{x}) dx &= \frac{|B|}{\sqrt{(2\pi)^n |\sum|}} \int e^{-\frac{1}{2}\overrightarrow{y}\Lambda\overrightarrow{y}^T} d\overrightarrow{y} \\ &= \frac{|B|}{\sqrt{(2\pi)^n |\sum|}} \prod \sqrt{2\pi\lambda_n} = 1 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X_i] &= \int x_i f(\overrightarrow{x}) d\overrightarrow{x} \\ &= \int (\mu_i + \sum_j y_j b_{ij}) \frac{e^{-\frac{1}{2}\overrightarrow{y}\Lambda\overrightarrow{y}}}{\sqrt{(2\pi)^n |\sum|}} d\overrightarrow{y} = \mu_i \end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \int (x_i - \mu_i)(x_j - \mu_j) f(\overrightarrow{x}) d\overrightarrow{x} \\
&= \int \sum_{k,l} y_k b_{ki} y_l b_{lj} \frac{e^{-\frac{1}{2}\overrightarrow{y}\Lambda\overrightarrow{y}}}{\sqrt{(2\pi)^n |\sum|}} d\overrightarrow{y} \\
&= \sum_k \int y_k^2 b_{ki} b_{kj} \frac{e^{-\frac{y_k^2}{2\lambda_k}}}{\sqrt{2\pi\lambda_k}} dy_k \\
&= \sum_k b_{ki}\lambda_k b_{kj} = \sigma_{ij}
\end{aligned}$$

Theorem 3.9

$\overrightarrow{X} \sim N(\overrightarrow{\mu}, \sum)$, D is a nonsingular $n \times n$ matrix, then $\overrightarrow{Y} = \overrightarrow{X} D \sim N(\overrightarrow{\mu} D, D^T \sum D)$. ♡

Proof Let $B = \{\overrightarrow{x} | x_i \in (a_i, b_i], \forall i\}$, $A = \{\overrightarrow{x} | \overrightarrow{x} D \in B\}$, then

$$\begin{aligned}
\mathbb{P}(\overrightarrow{Y} \in B) = \mathbb{P}(\overrightarrow{x} \in A) \\
&= \int_A f(\overrightarrow{x}) d\overrightarrow{x} = \int_B f(\overrightarrow{y} D^{-1}) |D^{-1}| d\overrightarrow{y} \\
&= \int_B \frac{1}{\sqrt{(2\pi)^n |D^T \sum D|}} e^{-\frac{1}{2}\overrightarrow{y}(D^T \sum D)^{-1}\overrightarrow{y}^T} d\overrightarrow{y}
\end{aligned}$$

Lemma 3.2

If

$$\sum = \begin{pmatrix} \sum_{11} & 0 \\ 0 & \sum_{22} \end{pmatrix}$$

, and $\overrightarrow{X} = (\overrightarrow{X}^{(1)}, \overrightarrow{X}^{(2)})$, then $\overrightarrow{X}^{(1)} \sim N(\overrightarrow{\mu}^{(1)}, \sum_{11})$, similarly for $\overrightarrow{X}^{(2)}$. ♡

Lemma 3.3

If

$$\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$$

, and $\overrightarrow{X} = (\overrightarrow{X}^{(1)}, \overrightarrow{X}^{(2)})$, then $\overrightarrow{X}^{(1)} \sim N(\overrightarrow{\mu}^{(1)}, \sum_{11})$, similarly for $\overrightarrow{X}^{(2)}$. ♡

Theorem 3.10

$\overrightarrow{X} \sim N(\overrightarrow{\mu}, \sum)$, *A is a* $n \times m$ *matrix with rank* m, *then* $\overrightarrow{Y} = \overrightarrow{X}A \sim N(\overrightarrow{\mu}A, A^T \sum A)$. ♡

Chapter 4 Law of Big Numbers

4.1 Expectation Revisited

Recall that we defined the expectation as

$$\mathbb{E}[X] = \begin{cases} \sum xf(x) & \text{discrete} \\ \int xf(x)dx & \text{continuous} \end{cases}$$

So we define

$$dF(x) = \begin{cases} F(x) - F(x-0) & \text{discrete} \\ f(x)dx & \text{continuous} \end{cases}$$

Then we can uniformly define $\mathbb{E}[X] = \int xdF(x)$ for both discrete and continuous random variables.

How to define expectation for more general probability space?

Step1:

Firstly consider *simple variables*, which take their values in a finite set, then we can write $X = \sum\limits_{i=1}^{n} x_i I_{A_i}$, where $\{A_i\}$ is a partition. Then we can define the expectation as

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

Step2:

Then for non-negative random variables $X \geq 0$, then there exists a sequence of simple variables $\{X_n\}$ with $X_n \geq 0$, so that $X_n \uparrow X$. For example, we can take

$$X_n = nI_{A_n} + \sum_{j=1}^{n2^n} \frac{j-1}{2^n} I_{A_{n,j}}$$

where $A_n = \{X \geq n\}, A_{n,j} = \{\frac{j-1}{2^n} \leq X < \frac{j}{2^n}\}$. Then we define $\mathbb{E}[X] = \lim\limits_{n\to\infty} \mathbb{E}[X_n]$. The expectation is well defined because of the following theorem.

Theorem 4.1 (Levy's Theorem)

If $X_n \uparrow X, Y_n \uparrow X$, *then* $\lim\limits_{n\to\infty} \mathbb{E}[X_n] = \lim\limits_{n\to\infty} \mathbb{E}[Y_n]$ ♡

Step3:

For the most general variable X, decompose it into $X = X^+ - X^-$, where $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$. When $\mathbb{E}[X^+] < \infty$ or $\mathbb{E}[X^-] < \infty$, define $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

In general, we write $\mathbb{E}[X] = \int_\Omega X(\omega)d\mathbb{P}$ or $\int_\Omega X(\omega)\mathbb{P}(d\omega)$. We say the expectation exists if $\mathbb{E}[|X|] < \infty$.

Now we focus on some properties of the expectation.

Proposition 4.1

(i) $X \geq 0$, *then* $\mathbb{E}[X] \geq 0$ *(ii)* $\mathbb{E}[1] = 1$

(iii) $a, b \in \mathbb{R}$, *then* $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. ♠

Then we introduce the continuity of expectation. Assume $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$.

(i)(Monotone Convergence)

If $X_{n+1}(\omega) \geq X_n(\omega) \quad \forall n, \omega$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(ii)(Dominated Convergence)

If $|X_n| \leq Y(\forall n)$, and $\mathbb{E}[Y] < \infty$ then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(iii)(Bounded Convergence)

If $\exists C > 0$, so that $|X_n| \leq C(\forall n)$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(iv)(Fatou's Lemma)

If $X_n \geq 0$ a.s. $\forall n$, then $\mathbb{E}[\underline{\lim}_{n\to\infty} X_n] \leq \underline{\lim}_{n\to\infty} \mathbb{E}[X_n]$.

Lebesgue-Stieltjes Integral:

Define a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as below: $\mu_F(\bigcup\limits_i (a_i, b_i]) = \sum\limits_i F(b_i) - F(a_i)$, and we can extend the definition to all the Borel sets. Then we obtain a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$.

The random variables are the measurable functions $g : \mathbb{R} \to \mathbb{R}$, and the abstract integral $\int g d\mu_F$(also can be written as $\int gdF$) is called the Lebesgue-Stieltjes integral.

Proposition 4.2

$\mathbb{E}[g(X)] = \int_{\mathbb{R}} gdF$

And define $\int_B gdF = \int_{\mathbb{R}} gI_B dF$ when the integration area is not the whole real line.

Theorem 4.2

X, Y are independent and their expectations exist, then $\mathbb{E}[|XY|] < \infty$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof Step1:

The case where X, Y are simple variables is proved before.

Step2:

When X, Y are nonnegative, define X_n, Y_n as before. One can check that X_n, Y_n are independent, and $X_n \uparrow X, Y_n \uparrow Y$, then

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y]$$

Step3:

For general X, Y, write $X = X^+ - X^-, Y = Y^+ - Y^-$, then $XY = (X^+Y^+ - X^-Y^-) - (X^+Y^- + X^-Y^+)$, since $\{X^+, X^-\}$ and $\{Y^+, Y^-\}$ are independent, thus

$$\mathbb{E}[XY] = (\mathbb{E}[X^+] - \mathbb{E}[X^-])(\mathbb{E}[Y^+] - \mathbb{E}[Y^-]) = \mathbb{E}[X]\mathbb{E}[Y]$$

4.2 Modes of Convergence

Firstly introduce the following modes of convergence.

Definition 4.1

$X, X_1, \cdots, X_n. \cdots$ *are random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) Almost surely convergence*:* $\mathbb{P}(\{\omega \in \mathcal{F} : X_n(\omega) \to X(\omega)\}) = 1$*, also denoted as* $X_n \xrightarrow{a.s.} X$

(ii) r order convergence*:* $\mathbb{E}[|X_n|] < \infty(\forall n)$*, and* $\mathbb{E}[|X_n - X|^r] \to 0$*, also denoted as* $X_n \xrightarrow{r} X$

(iii) Convergence in probability*:* $\forall \epsilon, \mathbb{P}(|X_n - X| > \epsilon) \to 0$*, also denoted as* $X_n \xrightarrow{P} X$

(iv) Convergence in distribution*:* C_{F_X} *is the set of the continuous point of the distribution function* F_X*,* $\forall x \in C_{F_X}$*,* $F_{X_n}(x) \to F_X(x)$*, also denoted as* $X_n \xrightarrow{D} X$ ♣

The convergence in distribution is also known as *weak convergence*. Consider the following example:

$X_n = \frac{1}{n}$, then

$$F_{X_n}(x) = \begin{cases} 1 & x \geq \dfrac{1}{n} \\ 0 & x < \dfrac{1}{n} \end{cases}$$

$X = 0$, then

$$F_X(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$\lim\limits_{n \to \infty} F_{X_n}(0) = 0 \neq 1 = F_X(0)$, but 0 is a point where the distribution function F_X is not continuous, so it's still true that $X_n \xrightarrow{D} X$. In this case, we say the distribution functions F_{X_n} converges to F_X *weakly*.

Define $||X||_r = (\mathbb{E}[|X|^r])^{\frac{1}{r}}$. The following are some inequalities, which are useful later.

Lemma 4.1

(i)(Hölder's Inequality) $||XY|| \leq ||X||_p||Y||_q$, *where* $\frac{1}{p}+\frac{1}{q}=1$

(ii)(Minkowski's Inequality) $||X+Y||_p \leq ||X||_p + ||Y||_p$

(iii)(Lyapunov's Inequality) If $r > s \geq 1$, *then* $||X||_r > ||X||_s$

(iv)(Markov's Inequality) For $a > 0$, $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}$

(v)(Chebyshev's Inequality) If X *has a finite expectation* μ, *then* $\mathbb{P}(|X-\mu| < a) \leq \frac{\mathrm{Var}(X)}{a^2}$. ♡

The checking of these inequalities are easy, and are omitted here.

Lemma 4.2

$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$. ♡

Proof

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(X_n \leq x, X \leq x+\epsilon) + \mathbb{P}(X_n \leq x, X > x+\epsilon)$$

$$\leq \mathbb{P}(X \leq x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Swap X_n with X, and $x-\epsilon$ and x, then

$$\mathbb{P}(X \leq x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Take the upper limit and lower limit of all sides, then

$$\mathbb{P}(X \leq x-\epsilon) \leq \varliminf_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \varlimsup_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x+\epsilon)$$

Then let $\epsilon \to 0^+$, the result follows from that $x \in C_{F_X}$.

Example 4.1 $\mathbb{P}(X=0) = \mathbb{P}(X=1) = \frac{1}{2}, X_n = X, Y = 1-X$, then $X_n \xrightarrow{D} Y$, but X_n does not converge to Y in the first three modes.

Lemma 4.3

(i) If $r > s \geq 1$, $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X$.

(ii) If $r \geq 1$, $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{P} X$. ♡

Proof (i) The result follows straightly from Lyapunov's inequality.

(ii) $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|X_n - X|^r > \epsilon^r) \leq \frac{\mathbb{E}[|X_n - X|^r]}{\epsilon^r} \to 0$

Example 4.2 $\Omega = [0,1]$, $\mathbb{P}$ is the Lebesgue measure, set

$$X_n(\omega) = \begin{cases} n^{\frac{1}{r}} & 0 \leq \omega \leq \dfrac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

Then $X_n \xrightarrow{P} X$, but $X_n \xrightarrow{r} X$ doesn't hold.

Theorem 4.3

(i) $X_n \xrightarrow{D} c \in \mathbb{R} \Rightarrow X_n \xrightarrow{P} c$

(ii) If $\exists K$, *so that* $|X_n| \leq K$ *a.s., then* $X_n \xrightarrow{P} c \Rightarrow X_n \xrightarrow{r} c$

Proof (i) Since c is the only point that is not continuous of F_X

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n > c + \epsilon) + \mathbb{P}(X_n < c - \epsilon) \\ &\leq 1 - F_{X_n}(c+\epsilon) + F_{X_n}(c - \frac{\epsilon}{2}) \\ &\to 1 - F_X(c+\epsilon) + F_X(c - \frac{\epsilon}{2}) \to 0 \end{aligned}$$

(ii) Firstly, observe that $\{|X| \leq K + \epsilon\} \supset \{|X_n| \leq K\} \bigcap \{|X_n - X| \leq \epsilon\}$, so $\mathbb{P}(|X| \leq K + \epsilon) = 1$.

Letting $\epsilon \to 0^+$ gives that $|X| \leq K$ a.s..

$$\begin{aligned} \mathbb{E}[|X_n - X|^r] &= \mathbb{E}[|X_n - X|^r I_{|X_n - X| > \epsilon}] + \mathbb{E}[|X_n - X|^r I_{|X_n - X| \leq \epsilon}] \\ &\leq (2K)^r \mathbb{P}(|X_n - X| > \epsilon) + \epsilon^r \end{aligned}$$

Thus $\overline{\lim}_{n\to\infty} \mathbb{E}[|X_n - X|^r] \leq \epsilon^r$, take $\epsilon \to 0^+$, we have $\lim\limits_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0$.

Theorem 4.4

Let $\triangle$ *denote* $a.s.$ *or* r, P, *then*

(i) $X_n \xrightarrow{\triangle} X, X_n \xrightarrow{\triangle} Y \Rightarrow \mathbb{P}(X = Y) = 1$

(ii) $X_n \xrightarrow{\triangle} X, Y_n \xrightarrow{\triangle} Y \Rightarrow X_n + Y_n \xrightarrow{\triangle} X + Y$

(iii) The results don't hold for convergence in distribution in general.

Proof (i)We only prove the case where $\triangle = r$ here. Since

$$||X - Y||_r \leq ||X - X_n||_r + ||Y - X_n||_r \leq 0$$

So $\mathbb{E}[|X - Y|^r] = 0$. By Markov's Inequality, for any $\epsilon > 0$,

$$\mathbb{P}(|X - Y| > \epsilon) \leq \frac{\mathbb{E}[|X - Y|^r]}{\epsilon^r} = 0$$

Then let $\epsilon \to 0^+$, the result comes from the fact that the distribution function is right continuous.

(ii)We only prove the case where $\triangle = P$ here.

$$\begin{aligned}\mathbb{P}(|X_n + Y_n - X - Y| > \epsilon) &\leq \mathbb{P}((|X_n - X| < \frac{\epsilon}{2})\bigcup(|Y_n - Y| < \frac{\epsilon}{2})) \\ &\leq \mathbb{P}(|X_n - X| < \frac{\epsilon}{2}) + \mathbb{P}(|Y_n - Y| < \frac{\epsilon}{2})\end{aligned}$$

Then the result follows.

(iii) Let $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$, and X_n has the same distribution as X. Then $X_n \xrightarrow{D} X, X_n \xrightarrow{D} -X$. But $\mathbb{P}(X = -X) \neq 1$, and $X_n + X_n$ doesn't converges to 0 in distribution.

Theorem 4.5 (Skorokhod's Representation Theorem)

$X_n \xrightarrow{D} X$, then there exists random variables Y_n, Y in $(\Omega, \mathcal{F}, \mathbb{P})$, so that:

(i)Y_n and X_n are identically distributed, Y and X are identically distributed.

(ii)$Y_n \xrightarrow{a.s.} Y$

The proof is omitted here.

Theorem 4.6

$X_n \xrightarrow{D} X$ if and only if $\forall g \in BC(\mathbb{R})$, $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$

Proof If $X_n \xrightarrow{D} X$, take Y_n, Y as in the representation theorem. Then for $\forall g \in BC(\mathbb{R}), g(Y_n) \xrightarrow{a.s.} g(Y)$. Then by the bounded convergence theorem, $\mathbb{E}[g(X_n)] = \mathbb{E}[g(Y_n)] \to \mathbb{E}[g(Y)] = \mathbb{E}[g(X)]$

From the other side, $\forall x \in C_{F_X}$, take $g(y) = I_{(-\infty,x]}(y)$, and take the mollifier $g_{x,\epsilon}$ of g, so that

$g_{x,\epsilon}(x) = 1, g_{x,\epsilon}(x+\epsilon) = 0$.

$$\mathbb{P}(X_n \leq x) = \mathbb{E}[I_{(-\infty,x]}(X_n)] \leq \mathbb{E}[g_{x,\epsilon}(X_n)]$$

Then take the upper limit, we have

$$\overline{\lim_{n\to\infty}} F_n(x) \leq \mathbb{E}[g_{x,\epsilon}(X)] \leq F(x+\epsilon)$$

Similarly we have

$$\underline{\lim_{n\to\infty}} F_n(x) \geq \mathbb{E}[g_{x-\epsilon,\epsilon}(X)] \geq F(x-\epsilon)$$

Then let $\epsilon \to 0$, we obtain the result.

4.3 Almost surely convergence and Borel-Cantelli Lemma

We can rewrite $\{\lim\limits_{n\to\infty} X_n = X\} = \bigcap\limits_{k=1}^{\infty}\bigcup\limits_{m=1}^{\infty}\bigcap\limits_{n=m}^{\infty}\{|X_n - X| \le \frac{1}{k}\}$, so the almost surely convergence is equivalent to the following:

$$\mathbb{P}(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{|X_n - X| \le \frac{1}{k}\}) = 1$$

or

$$\mathbb{P}(\bigcup_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{|X_n - X| > \frac{1}{k}\}) = 0$$

Lemma 4.4

(i)

$$X_n \xrightarrow{a.s.} X \Leftrightarrow \forall \epsilon > 0, \mathbb{P}(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{|X_n - X| > \epsilon\}) = 0$$
$$\Leftrightarrow \forall \epsilon > 0, \lim_{m\to\infty} \mathbb{P}(\bigcup_{n=m}^{\infty}\{|X_n - X| > \epsilon\}) = 0$$

(ii) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$

(iii) If $\forall \epsilon > 0, \sum\limits_{n} \mathbb{P}(|X_n - X| < \epsilon) < \infty$, *then* $X_n \xrightarrow{a.s} X$

♡

Let $\{A_n\}$ be a sequence of events, define the upper limit event to be

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m$$

Define the lower limit event to be

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m$$

The upper limit means that $\{A_n\}$ happens for an infinite number of times, which is also denoted as $\{A_n\ i.o.\}$. And the lower limit means that $\{A_n\}$ doesn't happen for a finite number of times.

Lemma 4.5 (Borel-Cantelli)

(i) If $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, *then* $\mathbb{P}(A_n\ i.o.) = 0$

(ii) If $\{A_n\}$ *are independent, and* $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, *then* $\mathbb{P}(A_n\ i.o.) = 1$

♡

Proof (i)$\mathbb{P}(A_n\ i.o.) \leq \mathbb{P}(\bigcup\limits_{m=n}^{\infty} A_m) \leq \sum\limits_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$

(ii) $\mathbb{P}(\bigcup\limits_{n=1}^{\infty} \bigcap\limits_{m=n}^{\infty} A_m^c) \leq \mathbb{P}(\bigcap\limits_{m=n}^{\infty} A_m^c) = \prod\limits_{m=n}^{\infty} (1 - \mathbb{P}(A_m)) \leq e^{-\sum\limits_{m=n}^{\infty} \mathbb{P}(A_m)} \to 0$

Lemma 4.6

$\{X_n\}$ *are identically distributed and their expectation exists, then*

(i) $\sum\limits_{n=1}^{\infty} \mathbb{P}(|X_1| > n) \leq \mathbb{E}[|X_1|] \leq \sum\limits_{n=0}^{\infty} \mathbb{P}(|X_1| \geq n)$

(ii) Let $Y_n = X_n I_{|X_n| \leq n}$, $a_n \to \infty$, *then* $\sum\limits_{k=1}^{n} \frac{1}{a_n}(Y_k - X_k) \xrightarrow{a.s.} 0.$

♡

Proof (i) $\sum\limits_{n=0}^{\infty} n\mathbb{P}(n \leq |X_1| < n+1) \leq \mathbb{E}[\sum\limits_{n=0}^{\infty} |X_1| I_{[n,n+1)}] \leq \sum\limits_{n=0}^{\infty} (n+1)\mathbb{P}(n \leq |X_1| < n+1)$

The right side is

$$\sum_{n=0}^{\infty} (n+1)(\mathbb{P}(|X_1| \geq n) - \mathbb{P}(|X_1| \geq n+1)) = \sum_{n=0}^{\infty} \mathbb{P}(|X_1| \geq n)$$

and the left side is similar.

(ii)

$$\sum_{k=1}^{\infty} \mathbb{P}(X_k \neq Y_k) = \sum_{k} \mathbb{P}(|X_k| \geq k) = \sum_{k} \mathbb{P}(|X_1| \geq k) < \infty$$

so by the Borel-Cantelli Lemma $\mathbb{P}(X_k \neq Y_k\ i.o) = 0$, the result follows.

4.4 Law of Large Numbers

Theorem 4.7

$\{X_n\}$ are i.i.d. r.v.s. $\mu = \mathbb{E}[X_k]$, $\mathrm{Var}[X_k] = \sigma^2 < \infty$, then $\frac{S_n}{n} \xrightarrow{2} \mu$, $\frac{S_n}{n} \xrightarrow{P} \mu$. ♡

Proof

$$\begin{aligned}
\mathbb{E}[(\frac{S_n}{n} - \mu)^2] &= \mathbb{E}[\frac{1}{n^2}(S_n - \mathbb{E}[S_n])^2] \\
&= \frac{1}{n^2}\mathrm{Var}(\sum_{k=1}^{n} X_k) \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_i) \\
&= \frac{1}{n}\mathrm{Var}(X_1) \to 0
\end{aligned}$$

Thus $\frac{S_n}{n} \xrightarrow{2} \mu$.

$$\mathbb{P}(|\frac{S_n}{n} - \mu| > \epsilon) \leq \frac{\mathrm{Var}(\frac{S_n}{n})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

Thus $\frac{S_n}{n} \xrightarrow{P} \mu$.

To deal with some situations where X_k have infinite values, we introduce a method of *truncation*.

We define

$$Y_n(\omega) = \begin{cases} X_n(\omega) & |X_n(\omega)| \leq k \\ 0 & |X_n(\omega)| > k \end{cases}$$

and $T_n = \sum_{k=1}^{n} Y_k$.

Theorem 4.8 (Weak Law of Large Numbers)

$\{X_n\}$ are i.i.d. r.v.s. $\mu = \mathbb{E}[X_k]$, then $\frac{S_n}{n} \xrightarrow{P} \mu$.(The independence condition can be weaken to pairwise independence) ♡

Proof Note that $\{Y_k\}$ are independent but not identically distributed. Take $a_n = n^\delta, \delta \in (0,1)$, then

$$
\begin{aligned}
\mathbb{P}(|T_n - \frac{\mathbb{E}[T_n]}{n}| > \epsilon) &\le \frac{\mathrm{Var}(T_n)}{\epsilon^2 n^2} \\
&= \frac{\sum_{k=1}^{n} \mathrm{Var}(Y_k)}{\epsilon^2 n^2} \\
&\le \frac{\sum_{k=1}^{n} \mathbb{E}[X_k^2 I_{|X_k|\le k}]}{\epsilon^2 n^2} \\
&= \frac{\sum_{k=1}^{n} \mathbb{E}[X^2 I_{|X_k|\le k}]}{\epsilon^2 n^2} \\
&= \frac{\sum_{k=1}^{a_n} \mathbb{E}[X^2 I_{|X_k|\le k}]}{\epsilon^2 n^2} + \frac{\sum_{k=a_n+1}^{n} \mathbb{E}[X^2 I_{|X_k|\le k}]}{\epsilon^2 n^2} \\
&\le a_n \frac{\sum_{k=1}^{a_n} \mathbb{E}[|X| I_{|X_k|\le k}]}{\epsilon^2 n^2} + \frac{\sum_{k=a_n+1}^{n} \mathbb{E}[X^2 (I_{|X_k|\le a_n} + I_{a_n < |X_k| \le k})]}{\epsilon^2 n^2} \\
&\le a_n \frac{\sum_{k=1}^{a_n} \mathbb{E}[|X| I_{|X_k|\le k}]}{\epsilon^2 n^2} + a_n \frac{\sum_{k=a_n+1}^{n} \mathbb{E}[|X| (I_{|X_k|\le a_n})]}{\epsilon^2 n^2} + n \frac{\sum_{k=a_n+1}^{n} \mathbb{E}[|X| (I_{|X_k| > a_n})]}{\epsilon^2 n^2} \\
&\le \frac{n a_n + n^2 \mathbb{E}[|X| I_{|X| > a_n}]}{\epsilon^2 n^2} \to 0
\end{aligned}
$$

The last inequality is because of the fact that $|X| I_{|X| > a_n} \le |X|$ and the dominated convergence theorem.

Theorem 4.9

$\{X_n\}$ are i.i.d. r.v.s. $\mu = \mathbb{E}[X_k], \mathrm{Var}(X_k) = C < \infty$, then $\frac{S_n}{n} \xrightarrow{a.s.} \mu$.

Proof Assume that $\mu = 0$. Since

$$\sum_n \mathbb{P}(\frac{|S_{n^2}|}{n^2} > \epsilon) \le \sum_n \frac{n^2 C}{\epsilon^2 n^4} < \infty$$

$\frac{S_{n^2}}{n^2} \xrightarrow{a.s.} 0$ by Borel-Cantelli Lemma. Introduce $M_n = \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$, then for $n^2 \le k < (n+1)^2$, we have $\frac{S_k}{k} = \frac{S_k - S_{n^2} + S_{n^2}}{k}$. Since

$$\mathbb{E}[M_n^2] \le 2n \mathbb{E}[(S_{(n+1)^2} - S_{n^2})^2] \le 4Cn^2$$

then

$$\sum_n \mathbb{P}(\frac{M_n}{n^2} > \epsilon) \leq \sum_n \frac{4C}{n^2\epsilon^2} < \infty$$

by Borel-Cantelli Lemma $\frac{M_n}{n^2} \xrightarrow{a.s.} 0$, thus

$$\frac{|S_k|}{k} \leq \frac{|S_{n^2}|}{n^2} + \frac{M_n}{n^2} \xrightarrow{a.s.} 0$$

Theorem 4.10 (Strong Law of Large Numbers)

$\{X_n\}$ are i.i.d. r.v.s, then $\mathbb{E}[|X_1|] < \infty$ is equivalent to $\frac{S_n}{n} \xrightarrow{a.s.} \mu = \mathbb{E}[X_k]$. ♡

Proof We can assume that $\{X_k\}$ are non negative. And we do the truncation to obtain Y_k.

Firstly suppose that $\mu = \mathbb{E}[X_k] < \infty$.

For $\alpha > 1$, let $\beta_k = [\alpha^k]$, then $\alpha^k - 1 < \beta_k \leq \alpha^k$, and $\forall m \geq 1$, $\sum_{k=m}^{\infty} \frac{1}{\beta_k^2} \leq \frac{C_\alpha}{\beta_m^2}$. Then

$$\begin{aligned}
\sum_n \mathbb{P}(\frac{1}{\beta_n}|T_{\beta_n} - \mathbb{E}[T_{\beta_n}]| > \epsilon) &\leq \sum_n \frac{1}{\epsilon^2\beta_n^2}\sum_{k=1}^{\beta_n} \mathrm{Var}(Y_k) \\
&\leq \sum_n \frac{1}{\epsilon^2\beta_n^2}\sum_{k=1}^{\beta_n} \mathbb{E}[Y_k^2] \\
&\leq \sum_{k=1}^{\infty} \frac{1}{\epsilon^2}\mathbb{E}[Y_k^2] \sum_{n:\beta_n \geq k} \frac{1}{\beta_n^2} \\
&\leq \frac{C_\alpha}{\epsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\mathbb{E}[Y_k^2] \\
&= \frac{C_\alpha}{\epsilon^2}\sum_k \frac{1}{k^2}\sum_{j=1}^{k}\mathbb{E}[Y_k^2 I_{j-1<X_k\leq j}] \\
&\leq \frac{C_\alpha}{\epsilon^2}\sum_k \frac{1}{k^2}\sum_{j=1}^{k} j^2\mathbb{P}(j-1 < X_k \leq j) \\
&\leq \frac{C_\alpha}{\epsilon^2}\sum_{j=1}^{\infty} j^2 \sum_{k=j}^{\infty}\frac{1}{k^2}\mathbb{P}(j-1 < X_k \leq j) \\
&\leq 2\frac{C_\alpha}{\epsilon^2}\sum_j \mathbb{P}(j-1 < X_1 \leq j) \\
&\leq \frac{C_\alpha}{\epsilon^2}2(\mathbb{E}[X_1]+1) < \infty
\end{aligned}$$

So by the Borel-Cantelli lemma, $\frac{1}{\beta_n}(T_{\beta_n} - \mathbb{E}[T_{\beta_n}]) \xrightarrow{a.s.} 0$. Also by the dominated convergence,

$$\frac{1}{n}\mathbb{E}[T_n] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_1 I_{X_1 \leq k}] \xrightarrow{a.s.} \mathbb{E}[X_1] = \mu$$

Thus $\frac{T_{\beta_n}}{\beta_n} \xrightarrow{a.s.} \mu$. Now for $\beta_n \leq k < \beta_{n+1}$, since

$$\frac{\beta_n}{\beta_{n+1}}\frac{T_{\beta_n}}{\beta_n} \leq \frac{T_k}{k} \leq \frac{\beta_{n+1}}{\beta_n}\frac{T_{\beta_{n+1}}}{\beta_{n+1}}$$

Take upper limit and lower limit, and let $\alpha \to 1^+$, then we have $\frac{T_k}{k} \xrightarrow{a.s.} \mu$.

Now suppose $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} \mu$, then $\frac{X_n}{n} \xrightarrow{a.s.} 0$, then $\sum_{n} \mathbb{P}(|X_n| \geq n) < \infty$. Otherwise, by the Borel-Cantelli lemma, $\mathbb{P}(|X_n| \geq n\ i.o.) = 1$, which contradicts that $\frac{X_n}{n} \to 0$. Thus $\mathbb{E}[|X_1|] < \infty$.

Here is a beautiful result which we will not prove here.

Theorem 4.11 (Khintchine's Law of Iterated Logarithm)

$\{X_k\}$ *are i.i.d r.v.s,* $\mathbb{E}[X_k] = 0, \mathrm{Var}(X_k) = 1$, *then*

$$\overline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{2n\mathrm{loglog}n}} = 1$$

♡

Chapter 5 Central Limit Theorem

5.1 Characteristic Functions

Definition 5.1

X, Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then we call $Z = X + iY$ a complex random variable*,*

and define $\mathbb{E}[Z] = \mathbb{E}[X] + i\mathbb{E}[Y]$. ♣

It's clear that a complex random variable can be regarded as a two-dimension random vector.

Definition 5.2

We say $Z_1 = X_1 + iY_1, Z_2 = X_2 + iY_2$ are independent*, if $(X_1, Y_1), (X_2, Y_2)$ are independent,*

in other words, if

$$\mathbb{P}(X_1 \le x_1, Y_1 \le y_1, X_2 \le x_2, Y_2 \le y_2) = \mathbb{P}(X_1 \le x_1, Y_1 \le y_1)\mathbb{P}(X_2 \le x_2, Y_2 \le y_2)$$ ♣

One can check that, if Z_1, Z_2 are independent, then $\mathbb{E}[Z_1 Z_2] = \mathbb{E}[Z_1]\mathbb{E}[Z_2]$.

Definition 5.3

For a random variable X, we define its characteristic function *to be*

$$\phi_X(t) = \mathbb{E}[e^{itX}](t \in \mathbb{R})$$ ♣

Since $|\phi_X(t)| \le 1$, $\phi_X(t)$ exists for all t and all kinds of random variables X, and $\phi_X(t) = \int_{\mathbb{R}} e^{itX} dF$. For continuous variable X, $\phi_X(t) = \int_{\mathbb{R}} e^{itX} f(x) dx$.

Theorem 5.1

(i) $\phi(0) = 1, \overline{\phi(t)} = \phi(-t), |\phi(t)| \le 1$

(ii) $\phi(t)$ is uniformly continuous on $(-\infty, \infty)$.

(iii) $\phi(t)$ is negative definite, in other words, for any $z_1, \cdots, z_n \in \mathbb{C}, t_1, \cdots, t_n \in \mathbb{R}$,

$$\sum_{j,k} z_j \overline{z_k} \phi(t_j - t_k) \geq 0$$

♡

Proof (i) Trivial.

(ii)$\forall t_0 \in \mathbb{R}$

$$|\phi(t_0 + h) - \phi(t_0)| = |\mathbb{E}[e^{it_0X}(e^{ihX} - 1)]| \leq \mathbb{E}[|e^{ihX} - 1|] \to 0(h \to 0)$$

The last step uses the bounded convergence theorem.

(iii)$\forall z_1, \cdots, z_n \in \mathbb{C}, t_1, \cdots, t_n \in \mathbb{R}$

$$\sum_{j,k} z_j \overline{z_k} \phi(t_j - t_k) = \sum_{j,k} \mathbb{E}[z_j \overline{z_k} e^{i(t_j - t_k)X}] = \mathbb{E}[(\sum_j z_j e^{it_jX} \overline{\sum_k z_k e^{it_kX}}] \geq 0$$

Actually, if a function satisfies (i)(ii)(iii), then it's a characteristic function of some random variable.

Theorem 5.2

If $\mathbb{E}[|X|^k] < \infty$, then $\forall j \leq k$, $\phi^{(j)}(0) = i^j \mathbb{E}[X^j]$, then

$$\phi(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!} \mathbb{E}[X^j] + o(t^k)$$

♡

Proof By exercise 5.6.4, $\forall j \leq k, \mathbb{E}[|X|^j] < \infty$, since $|\frac{d^j}{dt^j} e^{itX}| = |X|^j$, we can change the order of derivation and integration. Then the theorem is the result of Taylor's expansion theorem.

Theorem 5.3

(i)$\phi_{aX+b}(t) = e^{itb}\phi_X(at)$

(ii) If X, Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

♡

This can be checked easily.

Definition 5.4

If $\overrightarrow{X}$ *is a random vector, define its* characteristic function *to be* $\phi_{\overrightarrow{X}}(\overrightarrow{t}) = \mathbb{E}[e^{i\overrightarrow{t}\overrightarrow{X}}]$ ♣

Example 5.1 (i) X is of Bernoulli distribution with parameter p, then $\phi_X(t) = p + qe^{it}$

(ii) X is of exponential distribution with parameter λ, that is $f(x) = \lambda e^{-\lambda x}(x > 0)$, then $\phi_X(t) = \frac{\lambda}{\lambda - it}$

(iii) $X \sim N(0, 1)$, then $\phi_X(t) = e^{-\frac{1}{2}t^2}$. More generally, if $Y \sim N(\mu, \sigma)$, then $\phi_Y(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$

(iv) $\overrightarrow{X} \sim N(\overrightarrow{\mu}, \sum)$, then $\phi_{\overrightarrow{X}}(\overrightarrow{t}) = e^{i\overrightarrow{\mu}\,\overrightarrow{t}^T - \frac{1}{2}\overrightarrow{t}\sum\overrightarrow{t}^T}$

5.2 Inversion and Continuity Theorem

Theorem 5.4 (Inversion Formula)

For $-\infty < a < b < \infty$, then

$$\frac{1}{2}(F(b)+F(b-0))-\frac{1}{2}(F(a)+F(a-0))=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi(t)dt$$

Proof

$$\begin{aligned}
I_T &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi(t)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\int_{\mathbb{R}}e^{itx}dFdt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\int_{\mathbb{R}}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dFdt \\
&= \int_{\mathbb{R}}g_T(x)dF
\end{aligned}$$

where

$$\begin{aligned}
g_T(x) &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt \\
&= \frac{1}{\pi}\int_{0}^{T}\frac{\sin(x-a)t}{t}-\frac{\sin(x-b)t}{t}dt
\end{aligned}$$

Since

$$\frac{1}{\pi}\int_0^\infty\frac{\sin xt}{t}dt=\begin{cases}\frac{1}{2} & x>0\\ 0 & x=0\\ -\frac{1}{2} & x<0\end{cases}$$

We have $|g_T(x)|$ is bounded, and

$$\lim_{T\to\infty}g_T(x)=\begin{cases}1 & x\in(a,b)\\ \frac{1}{2} & x\in\{a,b\}\\ 0 & \text{otherwise}\end{cases}{}'$$

Then by the dominated convergence theorem

$$
\begin{aligned}
\lim_{T\to\infty} I_T &= \frac{1}{2}\mu_F(a,b) + \mu_F((a,b)) \\
&= \frac{F(b)+F(b-0)}{2} - \frac{F(a)+F(a-0)}{2}
\end{aligned}
$$

Corollary 5.1

$\phi_X = \phi_Y$ if and only if $F_X = F_Y$.

Theorem 5.5 (Multivariable Inversion Theorem)

$R = (a_1, b_1] \times (a_2, b_2] \cdots (a_n, b_n]$ and $\mu_F(\partial R) = 0$. And $\phi(t_1, \cdots, t_n) = \int_{\mathbb{R}^n} e^{i(t_1 x_1 \cdots + t_n x_n)} dF$.

Then

$$
\mu_F(R) = \lim_{T_1,\cdots,T_n\to\infty} \frac{1}{(2\pi)^n} \int_{-T_1}^{T_1} \cdots \int_{-T_n}^{T_n} \prod_{k=1}^{n} \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} \phi(t) dt_1 \cdots dt_n
$$

Specially for $n = 2$, it's clear that if $\phi_X(t_1, t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2)$, then X_1, X_2 are independent.

Theorem 5.6 (Levy-Cramer Continuity Theorem)

If $F_n(x) = \mathbb{P}(X_n \leq x), \phi_n(t) = \int_{\mathbb{R}} e^{itx} dF_n$. Then

(i) If $F_n \xrightarrow{W} F$, F is a distribution function, then $\phi_n(t) \to \phi(t) = \int_{\mathbb{R}} e^{itx} dF$, and the convergence is inner closed uniform convergence.

(ii) If $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ and $\phi(t)$ is continuous at $t = 0$, then $\phi(t)$ is a characteristic function of a distribution function F, and $F_n \xrightarrow{W} F$.

5.3 Limit Theorems

Firstly we use the characteristic function to prove the weak LLN. Since

$$\phi_{\frac{S_n}{n}}(t) = (\phi_{X_1}(\frac{t}{n}))^n = (1 + \frac{it\mu}{n} + o(\frac{t}{n}))^n \to e^{i\mu t}$$

Then by the Levy-Cramer theorem, $\frac{S_n}{n} \xrightarrow{D} \mu$, which is a constant, then by Theorem 4.3, $\frac{S_n}{n} \xrightarrow{P} \mu$.

Similarly we can prove the following central limit theorem.

Theorem 5.7 (CLT)

$\{X_k\}$ *are i.i.d r.v.s,* $\mu = \mathbb{E}[X_k], \sigma^2 = \mathrm{Var}(X_k), \sigma \in (0, \infty)$, *then*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1)$$

Proof Assume $\mu = 0, \sigma = 1$, then

$$\phi_{\frac{S_n}{\sqrt{n}}}(t) = (\phi_{X_1}(\frac{t}{\sqrt{n}}))^n = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-\frac{1}{2}t^2}$$

By Levy-Cramer theorem, $\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1)$.

Now we focus on the Lindeberg condition for CLT:

$\{X_k\}$ are random variables(might not identically distributed) with common expectation $\mu = 0$. $b_k^2 = \mathrm{Var}(X_k), B_n^2 = \sum\limits_{k=1}^{n} b_k^2$, if

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}[X_k^2 I_{|X_k| > \epsilon B_n}] = 0 (\forall \epsilon)$$

Then we say $\{X_k\}$ satisfy the Lindeberg condition(L).

Theorem 5.8

$\{X_k\}$ *are independent and satisfy the Lindeberg condition. Then* $\{X_k\}$ *satisfy the central limit law and the Feller condition(F) as follow:*

$$\lim_{n\to\infty} \frac{1}{B_n^2} \max_{1\le k\le n} b_k^2 = 0$$

The proof is omitted here. Also it's true that if the central limit law and the Feller condition hold, then the Lindeberg condition holds.

Corollary 5.2

If $\{X_k\}$ are independent and satisfy the following Lyapunov condition, then the central limit law holds.

$$\exists \delta > 0, \text{s.t.} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}[|X_k|^{2+\delta}] \to 0$$

Proof It suffices to check the Lindeberg condition, this comes from a straightforward calculation:

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}[X_k^2 I_{|X_k|>\epsilon B_n}] = \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}[\frac{X_k^{2+\delta}}{X_k^{\delta}} I_{|X_k|>\epsilon B_n}] \leq \lim_{n\to\infty} \frac{1}{\epsilon^{\delta} B_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}[|X_k|^{2+\delta}] = 0$$

Since

$$\begin{aligned}
\frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}[X_k^2 I_{|X_k|>\epsilon B_n}] &\geq \epsilon^2 \sum_{k=1}^{n} \mathbb{E}[I_{|X_k|>\epsilon B_n}] \\
&= \epsilon^2 \sum_{k=1}^{n} \mathbb{P}(|X_k| > \epsilon B_n) \\
&\geq \epsilon^2 \mathbb{P}(\bigcup_{k=1}^{n} \{|X_k| > \epsilon B_n\}) \\
&= \epsilon^2 \mathbb{P}(\max_{1\leq k\leq n} \frac{|X_k|}{B_n} < \epsilon)
\end{aligned}$$

So

$$\mathbb{P}(\max_{1\leq k\leq n} \frac{|X_k|}{B_n} < \epsilon) \to 0 (\forall \epsilon)$$

Thus the Lindeberg condition means that the probability of the relative deviations are uniformly small is 1.

The Lindeberg condition can deduce the Feller condition:

$$\frac{b_k^2}{B_n^2} = \frac{1}{B_n^2} \mathbb{E}[I_{|X_k|<\epsilon B_n} + I_{|X_k|\geq \epsilon B_n}] \leq \epsilon^2 + \sum_{k=1}^{n} \frac{1}{B_n^2} \mathbb{E}[X_k^2 I_{|X_k|\geq \epsilon B_n}]$$

Take the upper limit, we have

$$\overline{\lim_{n\to\infty}} \frac{b_k^2}{B_n^2} \le \epsilon^2$$

Then let $\epsilon \to 0$, we obtain the Feller condition.

Theorem 5.9 (Multivariable CLT)

$\{\overrightarrow{X_k}\}$ are i.i.d random vectors, with $\mathbb{E}[\overrightarrow{X_k}] = 0$ and $\mathbb{E}[\overrightarrow{X_k}^T \overrightarrow{X_k}] = \sum > 0$, then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \overrightarrow{X_k} \xrightarrow{D} N(0, \sum)$$

Now we try to obtain the an approximation of CLT from the Bernoulli distribution. Consider X_k are i.i.d Bernoulli random variables with parameter p. Then $\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}$. Also $\mathbb{E}[S_k] = np$, $\mathrm{Var}(S_k) = npq$, thus we introduce $x_k = \frac{k-np}{\sqrt{npq}}$.

Theorem 5.10 (Local CLT)

For $p \in (0,1)$, and for all k so that $|x_k| \le A$, we have uniformly

$$\mathbb{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}x_k^2} (n \to \infty)$$

Proof Since $k = np + \sqrt{npq}x_k$ and $n - k = nq - \sqrt{npq}x_k$, we have $k \sim np, n-k \sim nq$ uniformly. Then by Stirling's Formula

$$\begin{aligned}
\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k} &\sim \frac{(\frac{n}{e})^n \sqrt{2\pi n} p^k q^{n-k}}{(\frac{k}{e})^k \sqrt{2\pi k} (\frac{n-k}{e})^{n-k} \sqrt{2\pi(n-k)}} \\
&= \sqrt{\frac{n}{2\pi k(n-k)}} (\frac{np}{k})^k (\frac{nq}{n-k})^{n-k} \\
&\sim \frac{1}{\sqrt{2\pi npq}} (1 - \frac{\sqrt{npq}}{k} x_k)^k (1 + \frac{\sqrt{npq}}{n-k} x_k)^{n-k} \\
&= \frac{1}{\sqrt{2\pi npq}} e^{k\log(1 - \frac{\sqrt{npq}}{k} x_k) + (n-k)\log(1 + \frac{\sqrt{npq}}{n-k} x_k)} \\
&= \frac{1}{\sqrt{2\pi npq}} e^{-\sqrt{npq}x_k - \frac{npq}{k} x_k^2 + O(\frac{1}{\sqrt{n}}) + \sqrt{npq}x_k - \frac{npq}{n-k} x_k^2 + O(\frac{1}{\sqrt{n}})} \\
&\sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}x_k^2}
\end{aligned}$$

Theorem 5.11 (CLT:Integration Form)

$$\mathbb{P}(a < \frac{S_n - np}{\sqrt{npq}} \leq b) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

♡

Proof

$$\begin{aligned}
\mathbb{P}(a < \frac{S_n - np}{\sqrt{npq}} \leq b) &= \sum_{k: x_k \in (a,b]} \mathbb{P}(S_n = k) \\
&\sim \sum_{k: x_k \in (a,b])} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}x_k^2} \\
&= \sum_{k: x_k \in (a,b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2} (x_{k+1} - x_k) \\
&\to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx
\end{aligned}$$

Lemma 5.1

If $\forall k, n$, $\mathbb{E}[X_n^k]$ exists, and $\mathbb{E}[X_n^k] \to \gamma_k \forall k$, and $\{\gamma_k\}$ satisfy the Riesz condition:

$$\varlimsup_{k \to \infty} \frac{1}{k} (\gamma_{2k})^{\frac{1}{2k}} < \infty$$

Then $X_n \xrightarrow{D} X$, where X is the only random variable that has γ_k as its kth moment.

♡

The proof is omitted here.

One can calculate the moments of $X \sim N(0,1)$, which is

$$\mathbb{E}[X^k] = \gamma_k = \begin{cases} (2m-1)!! & k = 2m-1' \\ 0 & k = 2m \end{cases}$$

As a consequence of the above lemma, we have

Theorem 5.12

$\{X_k\}$ are independent with $\mathbb{E}[X_k] = 0, \mathrm{Var}(X_k) = 1 (\forall k)$, and $\forall m \geq 3$, $\sup_{k \to \infty} \mathrm{E}[|X_k|^m] < \infty$,

then

$$\mathbb{E}[(\frac{S_n}{\sqrt{n}})^k] \to \gamma_k$$

As a result, $\frac{S_n}{\sqrt{n}} \xrightarrow{D} N(0,1)$.

Proof Since $\mathbb{E}[(\frac{S_n}{\sqrt{n}})^k] = n^{-\frac{k}{2}} \sum\limits_{i_1,\cdots,i_k} \mathbb{E}[X_{i_1} \cdots X_{i_k}]$, the non-zero terms must be in the form of $\mathbb{E}[X_{i_1}^{a_1} \cdots X_{i_m}^{a_m}](i_1 \neq \cdots \neq i_m)$, with $a_i \geq 2, \sum\limits_i a_i = k$.

Consider the coefficients of those terms:

If k is odd, then $m \leq \frac{k-1}{2}$, the coefficient is of $\frac{n^m}{n^{\frac{k}{2}}} \to 0$;

Otherwise, m has to be $\frac{k}{2}$, under this case the coefficient is the number of divisions of $2m$ numbers into m pairs, which is γ_k.