Chapter 1 Distribution Function

Probability Theory and Stochastic Processes

WANG Zhuoran

16320216@stu.njau.edu.cn

香港科技大學 (廣州)

FTEC5031

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Course Introduction

This is a one-semester graduate-level course in probability theory that discusses a wide range of topics in probability theory based on measure theory.

Prerequisite. Real Analysis (MATH4123), Functional Analysis (MATH4122).

Textbook.

1. Chung, Kai Lai. *A Course in Probability Theory*. Academic press, 2001. Ch 1. – 7.
2. Le Gall, Jean-François. *Measure Theory, Probability, and Stochastic Processes*. Vol. 295. Springer Nature, 2022. Ch 8. – 10.

Reference textbook.

1. Durrett, Rick. *Probability: Theory and Examples*, 5th Edition. CUP, 2019.
2. Billingsley, Patrick. *Probability and Measure*, Anniversary edition. Wiley. 2012.

Course Introduction

Syllabus.

- Chapter 1. Distribution Function. 回顾经典分析中的诸多内容, 包括单调函数、分布函数、绝对连续函数, 并介绍奇异分布函数.
- Chapter 2. Measure Theory. 回顾实分析 (抽象测度) 中的基本内容, 包括集类与单调类定理、概率测度及其分布、Lebesgue-Stieltjes 测度等内容.
- Chapter 3. Random Variable. Expectation. Independence. 介绍随机变量 (可测函数) 的定义与性质、数学期望 (积分) 的定义与性质、独立性 (概率测度中崭新的概念).
- Chapter 4. Convergence Concepts. 介绍各种收敛性, 几乎处处收敛与 Borel-Cantelli 引理, 淡收敛及其性质, 连续性, 一直可积性与矩收敛.
- Chapter 5. Law of Large Numbers. Random Series. 介绍简单的极限定理, 讨论弱大数律与强大数律, 随机序列的收敛性, 最后讨论它们的应用.
- Chapter 6. Characteristic Function. 介绍特征函数与卷积的定义与性质, 特征函数的唯一性定理与逆转定理, 收敛定理与表示定理, 最后介绍高维情形与 Laplace 变换.
- Chapter 7. Central Limit Theorem and Its Ramifications. 介绍 Liapounov 定理, Lindeberg-Feller 定理, 讨论中心极限定理的一些分支结论, 最后, 讨论误差估计、重对数律与无限可分分布.

Content

Monotone Functions

Distribution Functions

Absolutely Continuous and Singular Distribution

Section 1

Monotone Functions

- We begin with a discussion of distribution functions as a traditional way of introducing probability measures.
- It serves as a convenient bridge from elementary analysis to probability theory, upon which the beginner may pause to review his mathematical background and test his mental agility (机敏).
- Some of the methods as well as results in this chapter are also useful in the theory of stochastic processes.

Properties of Increasing Functions

- Let f be an **increasing function** defined on the real line $(-\infty, \infty)$. Thus for any two real numbers x_1 and x_2,
$$x_1 < x_2 \quad \Longrightarrow \quad f(x_1) \leqslant f(x_2). \tag{1}$$
We begin by reviewing some properties of such a function.
- **(i)**. For each x, both *unilateral limits* (单侧极限)
$$\lim_{t \uparrow x} f(t) = f(x-) \quad \text{and} \quad \lim_{t \downarrow x} f(t) = f(x+) \tag{2}$$
exist and are finite. Furthermore the *limits at infinity* (无穷远点处的极限)
$$\lim_{t \downarrow -\infty} f(t) = f(-\infty) \quad \text{and} \quad \lim_{t \uparrow +\infty} f(t) = f(+\infty)$$
exists; the former may be $-\infty$, the latter may be $+\infty$[1].
- PROOF. 只需要使用单调性, i.e.
$$f(x-) = \sup_{-\infty < t < x} f(t), \quad f(x+) = \inf_{x < t < \infty} f(t). \quad \blacksquare$$

[1] 证明可参考张筑生数学分析新讲 (重排本 第一册) 定理 2.6.2.

Properties of Increasing Functions

- **(ii).** For each x, f is continuous at x if and only if

$$f(x-) = f(x) = f(x+).$$

- PROOF. 注意到单调函数 f 在 x 处连续等价于

$$\lim_{t \uparrow x} f(t) = f(x) = \lim_{t \downarrow x} f(t).$$

由性质 (i), 上述两个极限存在, 分别为 $f(x-)$ 与 $f(x+)$, 并且有

$$f(x-) \leqslant f(x) \leqslant f(x+), \tag{3}$$

从而 (ii) 成立. ■

- In general, we say that the function f has a **jump (跳跃)** at x if and only if the two limits in (2) both exists but are unequal. The value of f at x itself may be arbitrary, but for an increasing f the relation (3) must hold.
- **(iii).** The only possible kind of discontinuity of an increasing function is a jump.
- If there is a jump at x, we call x a **point of jump (跳跃点)** of f and the number $f(x+) - f(x-)$ the **size of the jump (跃度)** or simply "the jump" of f at x.

Example. 跳跃点的聚点不一定是间断点

例 1 (跳跃点的聚点不一定是间断点)

Let x_0 be an arbitrary real number, and define a function f as follows:

$$f(x)=\begin{cases} 0, & \text{for } x\leqslant x_0-1, \\ 1-\dfrac{1}{n}, & \text{for } x_0-\dfrac{1}{n}\leqslant x<x_0-\dfrac{1}{n+1},\ n\in\mathbb{N} \\ 1, & \text{for } x\geqslant x_0. \end{cases}$$

The point x_0 is a point of accumulation (聚点) of the points of jump $\{x_0-1/n : n\geqslant 1\}$, but f is continuous at x_0.

- 在介绍下一个例子之前, 我们首先定义在本讲义中常用的一个记号.
- For any real number t, we set

$$\delta_t(x)=\begin{cases} 0, & \text{for } x<t, \\ 1, & \text{for } x\geqslant t. \end{cases} \tag{4}$$

We shall call the function δ_t the **point mass at** t.

Example. 单调函数的跳跃点至多可数

- 如下例子告诉我们, $\mathbb{R}$ 上的单调函数的跳跃点可以是可数集.

例 2 (单调函数的跳跃点至多可数)

设 $\mathbb{Q}=\{a_n : n\geqslant 1\}$ 为有理数域, $\{b_n : b_n\geqslant 0, n\geqslant 1\}$ 满足 $\sum_{n=1}^{\infty} b_n<\infty$. 令

$$f(x)=\sum_{n=1}^{\infty} b_n\delta_{a_n}(x), \tag{5}$$

则函数项级数绝对收敛:

$$\sum_{n=1}^{\infty}\left|b_n\delta_{a_n}(x)\right|\leqslant\sum_{n=1}^{\infty} b_n<\infty, \quad \text{其中} \quad \left|\delta_{a_n}(x)\right|\leqslant 1.$$

由函数项级数的 Weierstrass 判别法, 可知 $f(x)$ 一致收敛. 另外, 因 $\delta_{a_n}(x)$ 是增函数, 任取 $x_1<x_2$, 有

$$f(x_2)-f(x_1)=\sum_{n=1}^{\infty} b_n\left[\delta_{a_n}(x_2)-\delta_{a_n}(x_1)\right]\geqslant 0.$$

于是 $f(x)$ 是增函数.

Example. 单调函数的跳跃点至多可数

例 2 续 (单调函数的跳跃点至多可数)

由一致收敛函数项级数的分析性质, 对任一 $x \in \mathbb{R}$, 便有

$$f(x+) - f(x-) = \sum_{n=1}^{\infty} b_n \big[\delta_{a_n}(x+) - \delta_{a_n}(x-)\big]. \tag{6}$$

Case I. 如果 $x \notin (a_n)_{n\geqslant 1}$, 则 $\delta_{a_n}(x+) - \delta_{a_n}(x-) = 0$, 此时 $f(x+) - f(x-) = 0$. 再由 (3) 式可知此时有 $f(x-) = f(x) = f(x+)$, 即 f 在点 x 处连续.

Case II. 如果 $x = a_k$ (for some $k \in \mathbb{N}$), 则 $\delta_{a_n}(x+) - \delta_{a_n}(x-) = 0$ ($\forall\ n \neq k$), 同时 $\delta_{a_k}(x+) = 1$, $\delta_{a_k}(x-) = 0$, 此时

$$f(x+) - f(x-) = b_k \geqslant 0.$$

即每一个 a_n 都是函数 f 的跳跃点.

结合上述两种情况, 函数 f 仅在每一个有理点处跳跃.

- REMARK. This example shows that the set of points of jump of an increasing function may be everywhere dense.

- **(iv)**. The set of discontinuous of an increasing function is countable[2] (至多可数).
- Proof. 设 f 是单调增函数, x 是 f 的任一跳跃点. 定义开区间 $I_x = \big(f(x-), f(x+)\big)$. 如果 x' 是异于 x 的另一跳跃点且 $x < x'$, 则存在 $\tilde{x}$, s.t. $x < \tilde{x} < x'$. 由 f 单调可知

$$f(x+) \leqslant f(\tilde{x}) \leqslant f(x'-).$$

  于是两个开区间 I_x 与 $I_{x'}$ 不交, i.e. $I_x \cap I_{x'} = \varnothing$. 现在我们将 f 定义域中的跳跃点与 f 值域中的互不相交的开区间构成了一一对应. 再由有理数集在实数集中的稠密性, 对每个开区间 I_x, 一定存在 $r_x \in \mathbb{Q}$ 且 $r_x \in I_x$. 又因为 $(r_x) \subseteq \mathbb{Q}$ 至多可数, 于是 f 的跳跃点至多可数. ■
- **(v)**. Let f_1 and f_2 be two increasing functions and D a set that is dense in $\mathbb{R}$. Suppose that

$$\forall\ x \in D: \ f_1(x) = f_2(x).$$

  Then f_1 and f_2 have the same points of jump of the same size, and they coincide (重合) except possibly at some of these points of jump.

[2] By "countable" we mean always "finite (possibly empty) or countably infinite".

Properties of Increasing Functions

- Proof. $\forall\ x\in\mathbb{R}$, 因为 D 在 $\mathbb{R}$ 中稠密, 于是可以找到 $(t_n)_{n\geqslant 1}\subseteq D$ 和 $(t_n')_{n\geqslant 1}\subseteq D$, s.t. $t_n\uparrow x$, $t_n'\downarrow x$. 由单调函数的性质 (i) 和 Heine 原理可知

$$f_1(x-)=\lim_{n\to\infty}f_1(t_n)=\lim_{n\to\infty}f_2(t_n)=f_2(t-),$$
$$f_1(x+)=\lim_{n\to\infty}f_1(t_n')=\lim_{n\to\infty}f_2(t_n')=f_2(t+).$$

  于是对任一 $x\in\mathbb{R}$, 均有 $f_1(x+)-f_1(x-)=f_2(x+)-f_2(x-)$, 由单调函数的性质 (ii) 我们可知函数 f_1 与 f_2 具有相同的跳跃点, 并且跃度相同. 再若 f_1 在点 x 处连续, 则 f_2 也在点 x 处连续, 并且我们有 $f_1(x)=f_1(x-)=f_2(x-)=f_2(x)$, 于是 f_1 与 f_2 只可能在某些跳跃点处不重合 (不相等). ■

- Remark. f_1 与 f_2 有可能会在 $\big(f_1(x-),f_1(x+)\big)=\big(f_2(x-),f_2(x+)\big)$ 中取不同的值, 即在跳跃点处选取不同的值. 在第二章中我们将会看到, 跳跃点处的函数值并不重要, 所以我们可以根据我们的需求更改这些函数值, 例如:

$$\widetilde{f}(x)=f(x-),\quad \widetilde{f}(x)=f(x+),\quad \widetilde{f}(x)=\frac{f(x-)+f(x+)}{2},$$

  在 Fourier 分析中, 第三种更改是更方便的, 但是前两种更改在概率论中是更方便的. 我们选取第二种更改, 即让函数 $\widetilde{f}(x)$ 右连续.

Properties of Increasing Functions

- **(vi)**. If we put $\forall\ x: \widetilde{f}(x) = f(x+)$, where f is increasing, then $\widetilde{f}$ is increasing and right continuous everywhere.
- PROOF. 首先函数 g 在点 x 处右连续的定义是 $\lim_{t\downarrow x} g(t) = g(x)$, i.e. $g(x+) = g(x)$. 于是要证 $\widetilde{f}$ 右连续, 即证
$$\forall\ x:\ \lim_{t\downarrow x} f(t+) = f(x+).$$
首先由 f 在点 x 处右极限存在的定义, $\forall\ \varepsilon > 0$, $\exists\ \delta(\varepsilon) > 0$, s.t. $\forall\ s \in (x, x+\delta)$, 有 $\left|f(s) - f(x+)\right| < \varepsilon$. 再令 $s \downarrow t$, 我们便有
$$\left|f(t+) - f(x+)\right| \leqslant \left|f(t+) - f(s)\right| + \left|f(s) - f(x+)\right| \leqslant 2\varepsilon.$$
这就说明 $\widetilde{f}$ 是右连续的. $\widetilde{f}$ 的单调性是显然的. ∎
- Let D be dense in $\mathbb{R}$, and suppose that f is a function with the domain D. We may speak of the monotonicity, continuity, uniform continuity, and so on of f on its domain of definition if in usual definitions we restrict ourselves to the points of D. Even if f is defined in a larger domain, we may still speak of these properties "on D" by considering the "restriction of f to D".

Properties of Increasing Functions

- Property (vii) is a generalization of (vi).
- **(vii)**. Let f be increasing on D, and define $\widetilde{f}$ on $\mathbb{R}$ as follows:

$$\forall\ x:\ \widetilde{f}(x) = \inf_{x<t\in D} f(t).$$

  Then $\widetilde{f}$ is increasing and right continuous everywhere.
- PROOF. $\forall\ x_0 \in \mathbb{R}$, $\forall\ \varepsilon > 0$, 由下确界的定义, $\exists\ t_0 \in D$ 且 $t_0 > x_0$, s.t.

$$f(t_0) \leqslant \inf_{x_0<t\in D} f(t) + \varepsilon,$$

  i.e. $f(t_0) - \varepsilon \leqslant \widetilde{f}(x_0) \leqslant f(t_0)$. 所以如果 $t \in D$ 满足 $x_0 < t < t_0$, 则有

$$0 \leqslant f(t) - \widetilde{f}(x_0) \leqslant f(t_0) - \widetilde{f}(x_0) \leqslant \varepsilon.$$

  再由 $\widetilde{f}$ 的定义便可知, 若 $x_0 < x < t_0$, 有 $0 \leqslant \widetilde{f}(x) - \widetilde{f}(x_0) \leqslant \varepsilon$. 由 ε 的任意性可知 $\widetilde{f}$ 在 x_0 处右连续. ■

Back to Content.

Section 2

Distribution Functions

- Suppose now that f is bounded as well as *increasing* and *not constant*.
- Consider the "normalized" function:

$$\widetilde{f}(x) = \frac{f(x) - f(-\infty)}{f(+\infty) - f(-\infty)}$$

  which is bounded and increasing with

$$\widetilde{f}(-\infty) = 0, \quad \widetilde{f}(+\infty) = 1. \quad (\clubsuit)$$

- Without loss of generality, assume the conditions ($\clubsuit$) in dealing with a bounded increasing function.
- We shall also assume the *right continuity* of f.

Definition of A Distribution Function

定义 3 (Distribution Functions)

A *real-valued* function F with domain $(-\infty, \infty)$ that is *increasing* and *right continuous* with $F(-\infty) = 0$, $F(+\infty) = 1$ is called a **distribution function (分布函数)**, to be abbreviated hereafter as "d.f.".

- A d.f. that is a point mass as defined in (4) is said to be "degenerate" (**退化分布**).
- All properties given in Section 1.1 hold for a d.f.
- In particular, let (a_j) be the countable set of points of jump of F and b_j the size of jump at a_j, then $F(a_j) - F(a_j-) = b_j$ since $F(a_j+) = F(a_j)$. Consider the function

$$F_d(x) = \sum_{j=1}^{\infty} b_j \delta_{a_j}(x)$$

which represents the sum of all the jumps of F in the half-line $(-\infty, x]$. It is clearly increasing, right continuous (since δ_{a_j} is right continuous), with

$$F_d(-\infty) = 0, \quad F_d(+\infty) = \sum_{j=1}^{\infty} b_j \leqslant 1. \tag{7}$$

Hence F_d is a bounded increasing function. It should constitute the "jumping part" of F.

分布函数可以分解为跳跃函数与连续增函数的和

定理 4 (分布函数可以分解为跳跃函数与连续增函数的和)

Let

$$F_c(x) = F(x) - F_d(x),$$

where F_d is the jumping part of F. Then F_c is positive, increasing, and continuous.

PROOF. 任取 $x < x'$, 则

$$0 \leqslant F_d(x') - F_d(x) = \sum_{x<a_j\leqslant x'} b_j = \sum_{x<a_j\leqslant x'} \left[F(a_j) - F(a_j-)\right] \leqslant F(x') - F(x).$$

由上式左边第一个不等号可知 F_d 是增函数. 又因为 $F(x) - F_d(x) \leqslant F(x') - F_d(x')$, 于是 F_c 也是增函数. 若令 $x = -\infty$, 由 $F_d(-\infty) = F(-\infty) = 0$ 可知 $F_d \leqslant F$, 从而 $F_c \geqslant 0$. 最后, 因为 F_d 与 F 均右连续, 从而 F_c 也右连续; 另外, 定义 F_d 的函数项级数在点 x 处一致收敛, 由例 2 可知

$$F_d(x) - F_d(x-) = \begin{cases} b_j, & \text{if } x = a_j, \\ 0, & \text{otherwise.} \end{cases}$$

事实上, 由 a_j 与 b_j 的定义, 将上式中的 F_d 替换成 F 也成立. 从而对任意的 x, 有

$$F_c(x) - F_c(x-) = F(x) - F(x-) - \left[F_d(x) - F_d(x-)\right] = 0.$$

于是 F_c 是左连续的. 从而 F_c 是连续函数. ∎

分布函数的前述分解是唯一的

定理 5 (分布函数的前述分解是唯一的)

Let F be a d.f. Suppose that there exist a continuous function G_c and a function G_d of the form

$$G_d(x) = \sum_{j=1}^{\infty} b'_j \delta_{a'_j}(x),$$

where $(a'_j)_{j\geqslant 1}$ is a countable set of real numbers and $\sum_j |b'_j| < \infty$, such that $F = G_c + G_d$, then $G_c = F_c$, $G_d = F_d$, where F_c and F_d are defined as before.

Proof. **反证法**. 若 $F_d \neq G_d$, 那么要么 (a_j) 与 (a'_j) 不相等; 要么对 (a'_j) 重排后, (a_j) 与 (a'_j) 相等, 但存在某个 $j \geqslant 1$, s.t. $b_j \neq b'_j$. 无论是哪种情形, 都存在至少一个 $j \geqslant 1$, 对于 $\tilde{a} = a_j$ 或 a'_j, 有

$$\left[F_d(\tilde{a}) - F_d(\tilde{a}-)\right] - \left[G_d(\tilde{a}) - G_d(\tilde{a}-)\right] \neq 0.$$

又因为 $F_c - G_c = G_d - F_d$, 于是

$$F_c(\tilde{a}) - G_c(\tilde{a}) - \left[F_c(\tilde{a}-) - G_c(\tilde{a}-)\right] \neq 0,$$

这与 $F_c - G_c$ 的连续性矛盾! 于是 $F_d = G_d$, 进而可知 $F_c = G_c$. ∎

分布函数可以写为离散型分布和连续型分布的凸组合

定义 6 (Discrete d.f. and Continuous d.f.)

A d.f. F that can be represented in the form $F = \sum_{j=1}^{\infty} b_j \delta_{a_j}$, where $(a_j)_{j \geqslant 1}$ is a countable set of real numbers, $b_j > 0$ for every j and $\sum_j b_j = 1$, is called a **discrete** d.f. A d.f. that is continuous everywhere is called a **continuous** d.f.

定理 7 (分布函数可以写为离散型分布和连续型分布的凸组合)

Every d.f. can be written as the convex combination of a discrete d.f. and a continuous d.f. Such a decomposition is unique.

PROOF. 设 $F_c \neq 0$, $F_d \neq 0$. 令 $\alpha = F_d(\infty)$, 便有 $0 < \alpha < 1$. 若不然, 若 $\alpha = F_d(\infty) = 1$, 则 $F_c(\infty) = F(\infty) - F_d(\infty) = 0$. 但由定理 4 可知 $F_c \geqslant 0$ 且 F_c 单增, 于是 $F_c \equiv 0$, 矛盾! 现设

$$F_1 = \frac{1}{\alpha} F_d, \quad F_2 = \frac{1}{1-\alpha} F_c,$$

于是 $F = \alpha F_1 + (1-\alpha) F_2$, 其中 F_1 是离散型分布, F_2 是连续型分布, 并且 F 是 F_1 与 F_2 的凸组合. Back to Content. ∎

Section 3

Absolutely Continuous and Singular Distribution

- Throughout the note the Lebesgue measure will be denoted by m.
- "almost everywhere" on the real line will be abbreviated to "a.e."
- An integral written in the form $\int \cdots dt$ is a Lebesgue integral, and a measurable function f is integrable in (a,b) iff

$$\int_a^b f(t)\,dt$$

  is defined and finite. The class of such functions will be denoted by $L^1(a,b)$.
- The main conclusions of this section have all been proved in MATH4123.

Definition of Absolutely Continuous

定义 8 (Absolutely Continuous 绝对连续)

A function F is called **absolutely continuous** in $(-\infty, \infty)$ and with respect to the Lebesgue measure if and only if there exists a function f in L^1 such that we have for every $x < x'$:

$$F(x') - F(x) = \int_x^{x'} f(t)\,dt. \tag{8}$$

- THEOREM. Such an absolutely continuous function F has a derivative equal to f a.e.
- In particular, if F is a d.f., then

$$f \geqslant 0 \text{ a.e.} \quad \text{and} \quad \int_{-\infty}^{\infty} f(t)\,dt = 1. \tag{9}$$

Conversely, given any f in $L^1(\mathbb{R})$ satisfying the conditions (9), the function F defined by

$$\forall\ x:\ F(x) = \int_{-\infty}^{x} f(t)\,dt \tag{10}$$

is easily seen to be a d.f. that is absolutely continuous.

Definition of Singular

定义 9 (Singular 奇异)

A function F is called **singular** if and only if it is not identically zero and F' exists and equals zero a.e.

定理 10 (积分与微分定理)

Let F be bounded increasing with $F(-\infty)=0$, and let F' denote its derivative wherever existing. Then the following assertions are true.

1. *If S denotes the set of all x for which $F'(x)$ exists with $0 \leqslant F'(x) < \infty$, then $m(S^c)=0$.*
2. *This F' belongs to $L^1(\mathbb{R})$, and we have for every $x < x'$:*

$$\int_x^{x'} F'(t)\,dt \leqslant F(x') - F(x). \tag{11}$$

3. *If we put*

$$\forall\ x:\ F_{ac}(x) = \int_{-\infty}^{x} F'(t)\,dt, \quad F_s(x) = F(x) - F_{ac}(x), \tag{12}$$

then $F'_{ac} = F'$ a.e. so that $F'_s = F' - F'_{ac} = 0$ a.e. and consequently F_s is singular if it is not identically zero.

分布函数可分解为离散、绝对连续和奇异连续分布之和

定义 11 (Density; Absolutely Continuous and Singular Part)

Any positive function f that is equal to F' a.e. is called a **density** of F. F_{ac} is called **the absolutely continuous part**, F_s **the singular part** of F. Note that the previous F_d is part of F_s as defined here.

- It is clear that F_{ac} is increasing and $F_{ac} \leqslant F$. From (11) it follows that if $x < x'$,

$$F_s(x') - F_s(x) = F(x') - F(x) - \int_x^{x'} f(t)\,dt \geqslant 0.$$

  Hence F_s is also increasing and $F_s \leqslant F$.
- We are now in a position to announce the following result, which is a refinement (改进) of Theorem 7.

定理 12 (分布函数可分解为离散、奇异连续和绝对连续分布之和)

Every d.f. F can be written as the convex combination of a discrete, a singular continuous, and an absolutely continuous d.f. Such a decomposition is unique.

Back to Content.

参考文献

- Chung, Kai Lai. *A Course in Probability Theory*, Third Edition. Academic Press, 2001.
- Durrett, Rick. *Probability: Theory and Examples*, 5th edition. Cambridge university press, 2019.
- Billingsley, Patrick. *Probability and Measure*, Anniversary edition. John Wiley & Sons, Inc. 2012.
- Varadhan, SR Srinivasa. *Probability theory*. No. 7. American Mathematical Soc., 2001.
- Williams, David. *Probability with martingales*. Cambridge university press, 1991.
- 任佳刚 巫静. 测度与概率教程. 科学出版社, 2018.
- 严士健 刘秀芳. 测度与概率, 第二版. 北京师范大学出版社, 2023.

FINANCIAL TECHNOLOGY

The Thrust nurtures creative minds to explore frontiers in Financial Technology, to shape the future financial industry that promotes social-financial health.

香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

社会枢纽
SOCIETY HUB

Society Hub • Financial Technology Thrust

Chapter 2 Measure Theory

Probability Theory and Stochastic Processes

WANG Zhuoran

🖂 16320216@stu.njau.edu.cn

香港科技大學 (廣州)

FTEC5031

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Content

Classes of Sets

Probability Measure and Their Distribution Functions

Probability Measure and Its Properties

Relations between D.F. and P.M.

Section 1

Classes of Sets

- Probability measures are defined on measurable spaces, so first we need to define the set class.
- We will briefly review the definitions of the various set classes in this section. And introduce the very important monotone class theorem.

- Let Ω be an "abstract space", namely a nonempty set of elements to be called "points" and denoted generically by ω.
- A nonempty collection $\mathscr{A}$ of subsets of Ω may have certain "closure properties".

(i) $E \in \mathscr{A} \implies E^c \in \mathscr{A}$.

(ii) $E_1 \in \mathscr{A},\ E_2 \in \mathscr{A} \implies E_1 \cup E_2 \in \mathscr{A}$.

(iii) $E_1 \in \mathscr{A},\ E_2 \in \mathscr{A} \implies E_1 \cap E_2 \in \mathscr{A}$.

(iv) $\forall\ n \geqslant 2\colon\ E_j \in \mathscr{A},\ 1 \leqslant j \leqslant n \implies \bigcup_{j=1}^{n} E_j \in \mathscr{A}$.

(v) $\forall\ n \geqslant 2\colon\ E_j \in \mathscr{A},\ 1 \leqslant j \leqslant n \implies \bigcap_{j=1}^{n} E_j \in \mathscr{A}$.

(vi) $E_j \in \mathscr{A};\ E_j \subseteq E_{j+1},\ 1 \leqslant j < \infty \implies \bigcup_{j=1}^{\infty} E_n \in \mathscr{A}$.

(vii) $E_j \in \mathscr{A};\ E_j \supseteq E_{j+1},\ 1 \leqslant j < \infty \implies \bigcap_{j=1}^{\infty} E_n \in \mathscr{A}$.

(viii) $E_j \in \mathscr{A},\ 1 \leqslant j < \infty \implies \bigcup_{j=1}^{\infty} E_n \in \mathscr{A}$.

(ix) $E_j \in \mathscr{A},\ 1 \leqslant j < \infty \implies \bigcap_{j=1}^{\infty} E_n \in \mathscr{A}$.

(x) $E_1 \in \mathscr{A},\ E_2 \in \mathscr{A},\ E_1 \subseteq E_2 \implies E_2 \backslash E_1 \in \mathscr{A}$.

- 当 (i) 成立时, 由 De Morgan 律, (ii) 与 (iii) 等价; (vi) 与 (vii) 等价; (viii) 与 (ix) 等价.
- 由数学归纳法可知, (ii) 可推出 (iv), (iii) 可推出 (v).

Classes of Sets

定义 1 (Several Classes of Sets)

Following the numbering of the set operations on the previous page, we have the following definitions: a nonempty collection $\mathscr{F}$ of subsets of Ω is called a

- **ring** iff (x) and (ii) holds. 环
- **algebra** (or **field**) iff (i) and (ii) holds. 代数 (域)
- σ-**ring** iff (x) and (viii) holds. σ-环
- σ-**algebra** (or **Borel algebra**) iff (i) and (viii) holds. σ-代数 (博雷尔域)
- **monotone class** iff (vi) and (vii) holds. 单调类

定理 2 (一个代数是 σ-代数当且仅当它是单调类)

An algebra is a σ-algebra if and only if it is also an monotone class.

Proof. ($\Rightarrow$). 显然, 因为 σ-代数对任意并封闭, 自然对单调列的并或交封闭.

($\Leftarrow$). 我们用 (iv) 与 (vi) 推出 (viii) 即可. 任取 $(E_j)_{j\geqslant 1} \subseteq \mathscr{A}$, 令 $F_n = \bigcup_{j=1}^{n} E_j \in \mathscr{A}$ (iv), 又因为 $F_n \subseteq F_{n+1}$ 且 $\bigcup_{j=1}^{\infty} E_j = \bigcup_{j=1}^{\infty} F_j$, 于是 $\bigcup_{j=1}^{\infty} E_j \subseteq \mathscr{A}$ (vi). ■

- The collection of all subsets of Ω is a σ-algebra called the *total* σ-algebra; the collection of the two sets $\{\varnothing, \Omega\}$ is a σ-algebra called the *trivial* σ-algebra.
- **Theorem**. If A is any index set and if for every $\alpha \in A$, $\mathscr{F}_\alpha$ is a σ-algebra (or monotone class) then the intersection $\bigcap_{\alpha \in A} \mathscr{F}_\alpha$ of all theses σ-algebra's (or monotone class's) is also a σ-algebra (or monotone class).
- **Theorem**. Given any nonempty collection $\mathscr{C}$ of sets, there is a *minimal* σ-algebra (or algebra, or monotone class) containing it; this is just the intersection of all σ-algebra's (or algebra's, or monotone class's) containing $\mathscr{C}$, of which there is at least one, namely the total σ-algebra mentioned above.
- This minimal σ-algebra (or algebra, or monotone class) is also said to be **generated by** $\mathscr{C}$. Back to Content.

定理 3 (Monotone Class Theorem 单调类定理)

Let $\mathscr{F}_0$ be an algebra, $\mathscr{G}$ the minimal monotone class containing $\mathscr{F}_0$, $\mathscr{F}$ the minimal σ-algebra containing $\mathscr{F}_0$, then $\mathscr{F} = \mathscr{G}$.

- We just skip the proof because it can be found in MATH4123 Chapter 3, Theorem 14.

推论 4 (代数上生成 σ-代数是包含该代数的最小的满足可列交 (并) 封闭的集类)

Let $\mathscr{F}_0$ be an algebra, $\mathscr{F}$ the minimal σ-algebra containing $\mathscr{F}_0$; $\mathscr{C}$ a class of sets containing $\mathscr{F}_0$ and having the closure properties (vi) and (vii), then $\mathscr{C}$ contains $\mathscr{F}$.

Section 2

Probability Measure and Their Distribution Functions

- First, we introduce the definition of probability measure and its properties, and discuss the equivalent conditions of countable additivity.
- After defining the probability measure, we discuss the relationship between probability measure and distribution function, that is, Lebesgue-Stieltjes measure is introduced.

Definition of Probability Measures

定义 5 (Probability Measures 概率测度)

Let Ω be a space, $\mathscr{F}$ a σ-algebra of subsets of Ω. A **probability measure** $\mathbb{P}$ on $\mathscr{F}$ is a numerically valued set function with domain $\mathscr{F}$, satisfying the following axioms:

1. $\forall\ E \in \mathscr{F}:\ \mathbb{P}(E) \geqslant 0$.
2. *Countable additivity*. If $(E_j)_{j\geqslant 1}$ is a countable collection of (pairwise) disjoint sets in $\mathscr{F}$, then
$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(E_j).$$
3. $\mathbb{P}(\Omega) = 1$.

The abbreviation “p.m.” will be used for “probability measure”.

The triple $(\Omega, \mathscr{F}, \mathbb{P})$ is called a **probability space**; Ω alone is called the **sample space**, and ω is then a **sample point**.

- The axiom corresponding to axiom (ii) restricted to a finite collection (E_j) is called **finite additivity**.

定理 6 (Properties of Probability Measures)

These axioms imply the following consequences, where all sets are members of $\mathscr{F}$.

1. $\mathbb{P}(E) \leqslant 1$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
4. $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.
5. $E \subseteq F \;\Rightarrow\; \mathbb{P}(E) = \mathbb{P}(F) - \mathbb{P}(F \backslash E) \leqslant \mathbb{P}(F)$.
6. *Monotone property.* $E_n \uparrow E$ *or* $E_n \downarrow E \;\Rightarrow\; \mathbb{P}(E_n) \to \mathbb{P}(E)$.
7. *Sub-additivity.* $\mathbb{P}\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(E_j)$, *here we do not make any requirements on the set sequence* $(E_j)_{j \geqslant 1}$.

- The following proposition

$$E_n \downarrow \emptyset \quad \Longrightarrow \quad \mathbb{P}(E_n) \to 0 \tag{1}$$

is called the "**axiom of continuity**". It is a particular case of the monotone property (6) above.

可列可加性的等价条件

定理 7 (可列可加性的等价条件)

The axiom of finite additivity and of continuity together are equivalent to the axiom of countable additivity.

PROOF. $(\Leftarrow)$. 假设 $\mathbb{P}$ 满足可列可加性, 则 $\mathbb{P}$ 一定满足有限可加性. 为了证明 $\mathbb{P}$ 满足 (1) 式的连续公理, 不妨设 $(E_n)_{n\geqslant 1}$ 是单调递减集合列, 于是有

$$E_n = \bigcup_{k=n}^{\infty}(E_k \backslash E_{k+1}) \cup \bigcap_{k=1}^{\infty} E_k.$$

如果 $E_n \downarrow \varnothing$, 则 $\bigcap_{k=1}^{\infty} E_k = \varnothing$. 因此, 由可列可加性便有

$$\forall\ n \geqslant 1:\ \mathbb{P}(E_n) = \sum_{k=n}^{\infty} \mathbb{P}(E_k \backslash E_{k+1}).$$

上述级数是收敛的, 于是 $\lim\limits_{n\to\infty} \mathbb{P}(E_n) = 0$, 从而连续性公理得证.

可列可加性的等价条件

(⇒). 现在假设 $\mathbb{P}$ 满足有限可加性与连续性公理. 为了证明 $\mathbb{P}$ 满足可列可加性, 不妨设 $(E_n)_{n\geqslant 1}$ 是互不相交集合列, 于是

$$\bigcup_{k=n+1}^{\infty} E_k \downarrow \emptyset.$$

上式成立是因为: 首先, $(\bigcup_{k=n+1}^{\infty} E_k)_n$ 单调递减是显然的, 若 $x \in \bigcap_{n=1}^{\infty} \bigcup_{k=n+1}^{\infty} E_k \neq \emptyset$, 则 $\forall\ n \in \mathbb{N}$, $\exists\ k_0 \geqslant n$, s.t. $x \in E_{k_0}$, 于是 (E_n) 中有无限项包含 x, 这与 (E_n) 互不相交矛盾! 现在, 有连续性公理, 便有

$$\lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k=n+1}^{\infty} E_k\Big) = 0. \quad (\clubsuit)$$

现在, 利用有限可加性, 便有

$$\forall\ n \geqslant 1:\ \mathbb{P}\Big(\bigcup_{k=1}^{\infty} E_k\Big) = \mathbb{P}\Big(\bigcup_{k=1}^{n} E_k\Big) + \mathbb{P}\Big(\bigcup_{k=n+1}^{\infty} E_k\Big) = \sum_{k=1}^{n} \mathbb{P}(E_k) + \mathbb{P}\Big(\bigcup_{k=n+1}^{\infty} E_k\Big).$$

上式说明级数 $\sum_{k=1}^{\infty} \mathbb{P}(E_k)$ 收敛 (因为其被上式左端控制). 在上式两端同时令 $n \to \infty$, 由 $(\clubsuit)$ 式便有可列可加性成立. ■

The Trace of Probability Space

- Let $\Delta \subseteq \Omega$, then the **trace** of the σ-algebra $\mathscr{F}$ on Δ is the collection of all sets of the form $\Delta \cap F$, where $F \in \mathscr{F}$.
- It is easy to see that this is a σ-algebra of subsets of Δ, and we shall denote it by $\Delta \cap \mathscr{F}$.
- Suppose $\Delta \in \mathscr{F}$ and $\mathbb{P}(\Delta) > 0$; then we may define the set function $\mathbb{P}_\Delta$ on $\Delta \cap \mathscr{F}$ as follows:

$$\forall\ E \in \Delta \cap \mathscr{F}: \quad \mathbb{P}_\Delta(E) = \frac{\mathbb{P}(E)}{\mathbb{P}(\Delta)}.$$

- It is easy to see that $\mathbb{P}_\Delta$ is a p.m. on $\Delta \cap \mathscr{F}$. The triple $(\Delta, \Delta \cap \mathscr{F}, \mathbb{P}_\Delta)$ will be called the **trace** of $(\Omega, \mathscr{F}, \mathbb{P})$ on Δ.

Discrete Probability Space

例 8 (Discrete Probability Space)

Let Ω be a countable set: $\Omega = \{\omega_j : j \in J\}$, where J is a countable index set, and let $\mathscr{C}$ be the total σ-algebra of Ω. Choose any sequence of numbers $\{p_j : j \in J\}$ satisfying

$$\forall\ j \in J:\ p_j \geqslant 0; \quad \sum_{j \in J} p_j = 1; \tag{2}$$

and define a set function $\mathbb{P}$ on $\mathscr{C}$ as follows:

$$\forall\ E \in \mathscr{C}: \quad \mathbb{P}(E) = \sum_{\omega_j \in E} p_j. \tag{3}$$

In words, we assign p_j as the value of the "probability" of the singleton $\{\omega_j\}$, and for an arbitrary set of ω_j's we assign as its probability the sum of all the probabilities assigned to its elements. Clearly axioms (1 - 3) are satisfied. Hence $\mathbb{P}$ so defined is a p.m.

Conversely, let any such $\mathbb{P}$ be given on $\mathscr{C}$. Since $\{\omega_j\} \in \mathscr{C}$ for every j, $\mathbb{P}(\{\omega_j\})$ is defined, let its value be p_j. Then (2) is satisfied. We have thus exhibited all the possible p.m.'s on Ω, this will be called a *discrete probability space*.

We've covered a lot of discrete probability spaces in elementary probability theory.

P.M. Defined on $(0, 1]$

例 9

Let $\mathscr{U} = (0, 1]$, $\mathscr{S}$ the collection of intervals:

$$\mathscr{S} = \{(a, b] : 0 < a < b \leqslant 1\}; \quad \text{Semi-ring (we have defined it in MATH4123)}$$

$\mathscr{B}$ the minimal σ-algebra containing $\mathscr{S}$, m the Borel-Lebesgue measure on $\mathscr{B}$. Then $(\mathscr{U}, \mathscr{B}, m)$ is a probability space.

Let $\mathscr{B}_0$ be the collection of subsets of $\mathscr{U}$ each of which is the union of a finite number of members of $\mathscr{S}$. Thus a typical set B in $\mathscr{B}_0$ is of the form

$$B = \bigcup_{j=1}^{n} (a_j, b_j] \quad \text{where } a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n.$$

It is easily seen that $\mathscr{B}_0$ is an algebra that is generated by $\mathscr{S}$ and in turn generates $\mathscr{B}$.

If we take $\mathscr{U} = [0, 1]$ instead, then $\mathscr{B}_0$ is no longer an algebra since $\mathscr{U} \notin \mathscr{B}_0$, but $\mathscr{B}$ and m may be defined as before. The new $\mathscr{B}$ is generated by the old $\mathscr{B}$ and the singleton $\{0\}$.

Lebesgue Measure Defined on Borel σ-Algebra

例 10 (Lebesgue Measure Defined on Borel σ-Algebra)

Let $\mathbb{R} = (-\infty, +\infty)$, $\mathscr{S}$ the collection of intervals of the form $(a, b]$, $-\infty < a < b < +\infty$. The algebra $\mathscr{B}_0$ generated by $\mathscr{S}$ consists of finite unions of disjoint sets of the form $(a, b]$, $(-\infty, a]$ or $(b, +\infty)$. The **Borel σ-algebra** $\mathscr{B}^1(\mathbb{R})$ on $\mathscr{R}$ is the σ-algebra generated by $\mathscr{S}$.

A set in $\mathscr{B}^1(\mathbb{R})$ will be called a (**linear**) **Borel set** when there is no danger of ambiguity.

However, the Borel-Lebesgue measure m on $\mathbb{R}$ is not a p.m.; indeed $m(\mathbb{R}) = +\infty$ so that m is not a finite measure but it is **σ-finite** on $\mathscr{B}_0$, namely: there exists a sequence of sets $E_n \in \mathscr{B}_0$, $E_n \uparrow \mathbb{R}$ with $m(E_n) < \infty$ for each n.

定理 11 (Borel σ-代数的等价定义)

The σ-algebra $\mathscr{B}^1(\mathbb{R})$ of Borel subsets of $\mathbb{R}$ is generated by each of the following collections of sets:

1. *the collection of all closed subsets of $\mathbb{R}$;*
2. *the collection of all subintervals of $\mathbb{R}$ of the form $(-\infty, b]$;*
3. *the collection of all subintervals of $\mathbb{R}$ of the form $(a, b]$.*

概率测度与分布函数之间的关系

- The question of probability measures on $\mathscr{B}^1(\mathbb{R})$ is closely related to the theory of distribution functions studied in Chapter 1.
- There is in fact a one-to-one correspondence between the set functions (p.m.) on the one hand, and the point functions (d.f.) on the other.
- Both points of view are useful in probability theory.
- Let's start with the simpler first relation: the p.m. determines the d.f.

定理 12 (P.M. Determines D.F.)

Each p.m. μ on $\mathscr{B}^1(\mathbb{R})$ determines a unique d.f. F through the correspondence

$$\forall\ x \in \mathbb{R}: \quad \mu\big((-\infty, x]\big) = F(x). \tag{4}$$

As a consequence, we have for $-\infty < a < b < +\infty$:

$$\begin{aligned} \mu\big((a,b]\big) &= F(b) - F(a), \quad & \mu\big((a,b)\big) &= F(b-) - F(a), \\ \mu\big([a,b)\big) &= F(b-) - F(a-), \quad & \mu\big([a,b]\big) &= F(b) - F(a-). \end{aligned} \tag{5}$$

Furthermore, let D be any dense subset of $\mathbb{R}$, then the correspondence is already determined by that in (4) restricted to $x \in D$, or by any of the four relations in (5) when a and b are both restricted to D.

Relation 1: P.M. Determines D.F.

PROOF. **Step 1. 去证明 $\mu\big((-\infty,x]\big)$ 是分布函数**. 首先令 $\forall\ x\in\mathbb{R}$: $I_x=(-\infty,x]$. 于是 $I_x\in\mathscr{B}(\mathbb{R})$, 从而我们可以定义 $\mu(I_x)$. 令 $F(x)=\mu(I_x)$, 于是我们定义了 $\mathbb{R}$ 上的函数, 我们去证明这个函数是分布函数.

(i). 首先, 由概率测度 μ 的单调性可知 F 是单调递增函数.

(ii). 再取 $x_n\downarrow x$, 于是 $I_{x_n}\downarrow I_x$. 利用概率测度的连续性, 就有

$$F(x_n)=\mu(I_{x_n})\downarrow\mu(I_x)=F(x). \tag{6}$$

于是函数 F 是右连续的.

(iii). 类似地, $I_x\downarrow\varnothing$ $(x\downarrow-\infty)$, 且 $I_x\uparrow\mathbb{R}$ $(x\uparrow+\infty)$. 再次利用概率测度的连续性, 有

$$\lim_{x\downarrow-\infty}F(x)=\lim_{x\downarrow-\infty}\mu(I_x)=\mu(\varnothing)=0;$$
$$\lim_{x\uparrow-\infty}F(x)=\lim_{x\uparrow-\infty}\mu(I_x)=\mu(\mathbb{R})=1.$$

Step 2. 验证 (5) 式成立. 事实上, 我们只需要证明 $\mu\big((-\infty,x)\big)=F(x)$ 成立即可. 为此, 取 $x_n<x$ 且 $x_n\uparrow x$. 因为 $I_{x_n}\uparrow(-\infty,x)$, 由概率测度的单调性, 便有

$$F(x-)=\lim_{n\to\infty}F(x_n)=\lim_{n\to\infty}\mu(I_{x_n})=\mu\big((-\infty,x)\big).$$

Step 3. 去证明限制在 D 上的 (4) 式可以推出无限制的 (4) 式. 为此, 注意到 $F|_D(x)$ 作为 $x\in D$ 的函数是右连续的. 由第一章单调函数的性质 (v), 因为 $F|_D$ 与 F 在每个 $x\in D$ 重合, 于是这两个函数只可能在跳跃点处不重合. 但是由于 $F|_D$ 与 F 的有连续性, 这两个函数在跳跃点处的函数值也相等. 于是由 $F|_D$ 可以直接决定 F. ■

Relation 2: D.F. Determines P.M.

定理 13 (D.F. Determines P.M.)

Each d.f. F determines a unique p.m. μ on $\mathscr{B}(\mathbb{R})$ through any one of the relations given in (5), or alternatively through (4). Uniqueness is trivial

Remark (D.F. Determines P.M.)

- This is the classical theory of *Stieltjes measure*; see, e.g., [YSJ23], Proposition 3.2.4.
- The F in the above theorem can be weakened to a nondecreasing right-continuous real valued function, at which μ is no longer a probability measure, but a σ-finite measure.
- The above theorem can be seen as a direct corollary of the measure extension theorem, and we give an alternative proof here.

* PROOF. 假设 F 单调不减右连续，令 $F(+\infty) = \lim_{x\to\infty} F(x)$, $F(-\infty) = \lim_{x\to-\infty} F(x)$,

$$\mu\big((a,b]\big) = F(b) - F(a), \quad a, b \in \mathbb{R} \cup \{\pm\infty\},\ a \leqslant b.$$

事实上，上式右端是有意义的，因为 $F(\infty) > -\infty$, $F(-\infty) < \infty$ (此时杜绝了 $\infty - \infty$ 不定式出现的可能). 显然，μ 是定义在 $\mathscr{S} = \{(a,b] : -\infty \leqslant a \leqslant b \leqslant \infty\}$ 上的非负集函数[1].

[1]其中 $b = \infty$ 理解为 b 可以取任意大的实数，但不能取作 ∞, $a = -\infty$ 也可以类似理解.

Relation 2: D.F. Determines P.M.

Step 1. 证明 μ 是有限可加的. 事实上, 若 $(a,b]=\bigcup_{k=1}^{n}(a_k,b_k]$, 其中 $(a_k,b_k]$ $(1\leqslant k\leqslant n)$ 互不相交, 重排后不妨设 $a=a_1<b_1=a_2<b_2=\cdots<b_{n-1}=a_n<b_n=b$. 于是

$$\mu\big((a,b]\big)=F(b)-F(a)=F(b_n)-F(a_1)=\sum_{k=1}^{n}\big(F(b_k)-F(a_k)\big)=\sum_{k=1}^{n}\mu\big((a_k,b_k]\big).$$

Step 2. 利用 $\mathbb{R}$ 中有限闭区间的紧性证明 μ 是可列可加的. 不妨设

$$(a,b]=\bigcup_{k=1}^{n}(a_k,b_k], \tag{7}$$

其中 $-\infty\leqslant a<b\leqslant\infty$, $(a_k,b_k]$ 互不相交, 我们去证明

$$\mu\big((a,b]\big)=\sum_{k=1}^{\infty}\mu\big((a_k,b_k]\big). \tag{8}$$

(1). 首先证明 $\mu\big((a,b]\big)\geqslant\sum_{k=1}^{\infty}\mu\big((a_k,b_k]\big)$. $\forall\ n\in\mathbb{N}$, 由 (7) 式可知 $(a,b]\supseteq\bigcup_{k=1}^{n}(a_k,b_k]$, 由于 $(a_k,b_k]$ $(1\leqslant k\leqslant n)$ 两两不交, 不妨设 $a\leqslant a_1<b_1\leqslant a_2<b_2\leqslant\cdots<b_{n-1}\leqslant a_n<b_n\leqslant b$, 于是

$$\mu\big((a,b]\big)=F(b)-F(a)\geqslant F(b_n)-F(a_1)=\sum_{k=1}^{n}\big(F(b_k)-F(a_k)\big)+\sum_{k=2}^{n}\big(F(a_k)-F(b_{k-1})\big)$$
$$\geqslant\sum_{k=1}^{n}\big(F(b_k)-F(a_k)\big)=\sum_{k=1}^{n}\mu\big((a_k,b_k]\big).$$

在上式两端同时令 $n\to\infty$, 便有

$$\mu\big((a,b]\big)\geqslant\sum_{k=1}^{\infty}\mu\big((a_k,b_k]\big).$$

(2.1). 在 $-\infty<a<b<\infty$ 的情况下证明反向不等式. 首先任取 $\varepsilon>0$, s.t. $a+\varepsilon<b$. 由 F 的右连续性, $\forall\ k\in\mathbb{N}$, $\exists\ \delta_k>0$, s.t.

$$F(b_k+\delta_k)-F(b_k)<\frac{\varepsilon}{2^k}. \tag{9}$$

于是 $\{(a_k,b_k+\delta_k):k\in\mathbb{N}\}$ 是 $[a+\varepsilon,b]$ 的一个开覆盖. 由于 $[a+\varepsilon,b]$ 是 $\mathbb{R}$ 中闭 (紧) 区间, 由有限覆盖引理, $\exists\ N\in\mathbb{N}$, s.t.

$$[a+\varepsilon,b]\subseteq\bigcup_{k=1}^{N}(a_k,b_k+\delta_k). \tag{10}$$

由 (10) 式可知, $\exists\ k_1\leqslant N$, s.t. $a+\varepsilon\in(a_{k_1},b_{k_1}+\delta_{k_1})$. 若 $b_{k_1}+\delta_{k_1}\in[a+\varepsilon,b]$, 则由 (10) 式可知, $\exists\ k_2\leqslant N$, s.t. $b_{k_1}+\delta_{k_1}\in(a_{k_2},b_{k_2}+\delta_{k_2})$. 一般地, 若 $k_1,\cdots,k_l$ 已选定, s.t.

$$b_{k_i}+\delta_{k_i}\in(a_{k_{i+1}},b_{k_{i+1}}+\delta_{k_{i+1}}),\quad 其中\quad 1\leqslant i\leqslant l-1, b_{k_l}+\delta_{k_l}\leqslant b, \tag{11}$$

再由 (10) 式可知 $\exists\, k_{l+1} \leqslant N$, s.t. $b_{k_l}+\delta_{k_l} \in (a_{k_{l+1}}, b_{k_{l+1}}+\delta_{k_{l+1}})$. 由 (10) 式及归纳法可知, $\exists\, k_m \leqslant N$, s.t. $b \in (a_{k_m}, b_{k_m}+\delta_{k_m})$. 现在由 (9) 式便可知

$$
\begin{aligned}
& F(b)-F(a+\varepsilon) \leqslant F(b_{k_m}+\delta_{k_m})-F(a_{k_1}) \\
= & F(b_{k_m}+\delta_{k_m})-F(b_{k_{m-1}}+\delta_{k_{m-1}})+F(b_{k_{m-1}}+\delta_{k_{m-1}})-\cdots-F(b_{k_1}+\delta_{k_1})+F(b_{k_1}+\delta_{k_1})-F(a_{k_1}) \\
= & \sum_{l=1}^{m-1}\left[F(b_{k_{l+1}}+\delta_{k_{l+1}})-F(b_{k_l}+\delta_{k_l})\right]+\left[F(b_{k_1}+\delta_{k_1})-F(a_{k_1})\right] \\
\leqslant & \sum_{l=1}^{m}\left[F(b_{k_l}+\delta_{k_l})-F(a_{k_l})\right] \leqslant \sum_{k=1}^{\infty}\left[F(b_k+\delta_k)-F(b_k)\right]+\sum_{k=1}^{\infty}\left[F(b_k)-F(a_k)\right] \quad \text{[由 (9) 式]} \\
\leqslant & \sum_{k=1}^{\infty}\mu((a_k,b_k])+\sum_{k=1}^{\infty}\frac{\varepsilon}{2^k}=\sum_{k=1}^{\infty}\mu((a_k,b_k])+\varepsilon,
\end{aligned}
$$

$$
\text{即} \quad F(b)-F(a+\varepsilon) \leqslant \sum_{k=1}^{\infty}\mu\big((a_k,b_k]\big)+\varepsilon. \tag{12}
$$

在上式中令 $\varepsilon \to 0$, 由 F 的右连续性便可知反向不等式成立.

(2.2). 在 $a=-\infty$, $b<\infty$ 的情况下证明反向不等式. $\forall\, N \in \mathbb{N}$, 有 $[-N,b] \subseteq \bigcup_{k=1}^{\infty}(a_k,b_k]$. 于是 (2.1) 中直到 (12) 式的证明当 $a+\varepsilon$ 换成 $-N$ 后完全适用, 所以有 $F(b)-F(-N) \leqslant \sum_{k=1}^{\infty}\mu\big((a_k,b_k]\big)+\varepsilon$. 先令 $\varepsilon \to 0$, 再令 $N \to \infty$ 便可知此情形下反向不等式成立. $a>-\infty$, $b=\infty$ 以及 $a=-\infty$, $b=\infty$ 的情形式类似的. 现在, 可列可加性得证. ▸ Back to Content. ■

Bibliography

- Chung, Kai Lai. *A Course in Probability Theory*, Third Edition. Academic Press, 2001.
- Durrett, Rick. *Probability: Theory and Examples*, 5th edition. Cambridge university press, 2019.
- Billingsley, Patrick. *Probability and Measure*, Anniversary edition. John Wiley & Sons, Inc. 2012.
- Varadhan, SR Srinivasa. *Probability theory*. No. 7. American Mathematical Soc., 2001.
- Williams, David. *Probability with martingales*. Cambridge university press, 1991.
- 任佳刚 巫静. 测度与概率教程. 科学出版社, 2018.
- 严士健 刘秀芳. 测度与概率, 第二版. 北京师范大学出版社, 2023.

FINANCIAL TECHNOLOGY

The Thrust nurtures creative minds to explore frontiers in Financial Technology, to shape the future financial industry that promotes social-financial health.

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

社会枢纽
SOCIETY HUB

Society Hub • Financial Technology Thrust

Chapter 3 Random Variable. Expectation. Independence

Probability Theory and Stochastic Processes

WANG Zhuoran

16320216@stu.njau.edu.cn

香港科技大學 (廣州)

FTEC5031

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Content

Random Variables

- Relations Between R.V. and P.M. of R.V.
- Functions of Random Variables
- Some Notations

Expectation

- Definition and Properties
- Change of Variables Formula
- Moments

Independence

- Definition and Properties
- Product Measures and Kolmogorov's Extension Theorem

Section 1

Random Variables

- First, we recall the definition of a random variable and show the relationship between r.v.'s on Ω and r.v.'s on $\mathbb{R}$.
- Next, we discuss functions of r.v.'s. Define random vectors and their functions later.
- Finally, we present some of the notations that are commonly used in this handout.

Definition of a Random Variable

- Let the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ be given. $\mathbb{R} = (-\infty, +\infty)$ the (finite) real line, $\mathbb{R}^* = [-\infty, +\infty]$ the (extended) real line, $\mathscr{B}^1 = \mathscr{B}^1(\mathbb{R})$ the Borel σ-algebra on $\mathbb{R}$, $\mathscr{B}^*$ the extended Borel σ-algebra. A set in $\mathscr{B}^*$ is just a set in $\mathscr{B}^1$ possibly enlarged by one or both points $\pm\infty$.

定义 1 (Random Variable)

A real, extended-valued **random variable** is a function X whose domain is a set Δ in $\mathscr{F}$ and whose range is contained in $\mathbb{R}^*$ such that for each extended Borel set $B \in \mathscr{B}^*$, we have

$$\{\omega : X(\omega) \in B\} \in \Delta \cap \mathscr{F}, \tag{1}$$

where $\Delta \cap \mathscr{F}$ is the trace of $\mathscr{F}$ on Δ. A complex-valued random variable is a function on a set Δ in $\mathscr{F}$ to the complex plane whose real and imaginary parts are both real, finite-valued random variables.

- For a discussion of basic properties *we may suppose $\Delta = \Omega$ and that X is real and finite-valued with probability one*, i.e.

$$\mathbb{P}\{\omega : |X(\omega)| < \infty\} = 1 \quad \Longleftrightarrow \quad \mathbb{P}\{\omega : |X(\omega)| = \infty\} = 0,$$

that means X is finite "almost surely", abbreviated as "a.s."

Definition of a Random Variable

- The restricted meaning of a "random variable", abbreviated as "r.v.", will be understood in this notes unless otherwise specified.
- The general case may be reduced to this one by considering the trace of $(\Omega, \mathscr{F}, \mathbb{P})$ on Δ, or on the "domain of finiteness" $\Delta_0 = \{\omega : |X(\omega)| < \infty\}$, and taking real and imaginary parts.

Remark (Inverse Mapping of a R.V.)

The **inverse mapping** X^{-1} from $\mathbb{R}$ to Ω, defined as follows:

$$\forall \ A \subseteq \mathbb{R}: \ X^{-1}(A) = \{\omega : X(\omega) \in A\}.$$

Condition (1) then states that X^{-1} carries members of $\mathscr{B}^1$ onto members of $\mathscr{F}$:

$$\boxed{\forall \ B \in \mathscr{B}^1 : \ X^{-1}(B) \in \mathscr{F};} \tag{2}$$

or in the briefest notation:

$$\boxed{X^{-1}(\mathscr{B}^1) \in \mathscr{F}.}$$

Such a function is said to be **measurable** (**with respect to** $\mathscr{F}$). Thus, and r.v. is just a measurable function from Ω to $\mathbb{R}$ (or $\mathbb{R}^*$).

General Properties of Inverse Images

Reference. Sheldon Axler. *Measure, Integration & Real Analysis*. Theorem 2.33 and 2.34.

- Inverse Images have good algebraic properties, as is shown in the next results.

引理 2 (Algebra of Inverse Images)

Suppose $f: X \to Y$ and $g: Y \to W$ are functions. Then

1. $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$ *for every* $A \subseteq Y$.
2. $f^{-1}\left(\bigcup_{A \in \mathscr{A}} A\right) = \bigcup_{A \in \mathscr{A}} f^{-1}(A)$ *for every set* $\mathscr{A}$ *of subsets of* Y.
3. $f^{-1}\left(\bigcap_{A \in \mathscr{A}} A\right) = \bigcap_{A \in \mathscr{A}} f^{-1}(A)$ *for every set* $\mathscr{A}$ *of subsets of* Y.
4. $(g \circ f)^{-1}(A) = f^{-1}\left(g^{-1}(A)\right)$ *for every* $A \subseteq W$.

PROOF. (1). $\forall\ x \in X$, 我们有

$$\begin{aligned} x \in f^{-1}(Y \setminus A) &\iff f(x) \in Y \setminus A \\ &\iff f(x) \notin A \\ &\iff x \notin f^{-1}(A) \\ &\iff x \in X \setminus f^{-1}(A). \end{aligned}$$

这便说明 $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$. 其余三条留作练习. ■

随机变量的等价命题

定理 3 (随机变量的等价命题)

X is a r.v. if and only if for each real number x, or each real number x in a dense subset of $\mathbb{R}$, we have

$$\{\omega : X(\omega) \leqslant x\} \in \mathscr{F}.$$

Proof. $(\Rightarrow)$. 显然, 因 $(-\infty, x] \in \mathscr{B}^1$, 故 $\{\omega : X(\omega) \in (-\infty, x]\} = \{\omega : X(\omega) \leqslant x\} \in \mathscr{F}$.

$(\Leftarrow)$. 设 $\forall\ x : X^{-1}\big((-\infty, x]\big) \in \mathscr{F}$. 令 $\mathscr{A} = \{S \subseteq \mathbb{R} : X^{-1}(S) \in \mathscr{F}\}$, 则由引理 2,

(i). $S \in \mathscr{A} \ \Rightarrow\ X^{-1}(S^c) = \big(X^{-1}(S)\big)^c \in \mathscr{F} \ \Rightarrow\ S^c \in \mathscr{A}$.

(ii). $(S_k)_{k\geqslant 1} \subseteq \mathscr{A} \ \Rightarrow\ X^{-1}\big(\bigcup_{k=1}^{\infty} S_k\big) = \bigcup_{k=1}^{\infty} X^{-1}(S_k) \in \mathscr{F} \ \Rightarrow\ \bigcup_{k=1}^{\infty} S_k \in \mathscr{A}$.

故 $\mathscr{A}$ 是 $\mathbb{R}$ 上 σ-代数. 又因为 $\{\omega : X(\omega) \leqslant x\} = X^{-1}\big((-\infty, x]\big) \in \mathscr{F}$, 于是 $(-\infty, x] \in \mathscr{A}$. 而 $\mathscr{B}^1$ 可由 $\{(-\infty, x] : x \in \mathbb{R}\}$ 或者 $\{(-\infty, x] : x \in D$ 其中 $D \subseteq \mathbb{R}$ 稠密$\}$ 生成, 由生成集类的性质可知 $\mathscr{B}^1 \subseteq \mathscr{A}$. 于是 $\forall\ B \in \mathscr{B}^1$, 均有 $X^{-1}(B) \in \mathscr{F}$. 这就说明 X 是随机变量. ■

$(\Omega, \mathscr{F}, \mathbb{P})$ 上的随机变量可导出 $(\mathbb{R}, \mathscr{B}^1)$ 上的概率测度

- Since $\mathbb{P}$ is defined on $\mathscr{F}$, the probability of the set in (1) is defined and will be written as

$$\mathbb{P}\{\omega : X(\omega) \in B\} \quad \text{or} \quad \mathbb{P}(X \in B).$$

- The next theorem relates the p.m. $\mathbb{P}$ to a p.m. on $(\mathbb{R}, \mathscr{B}^1)$ as discussed in Section 2.2.

定理 4 ($(\Omega, \mathscr{F}, \mathbb{P})$ 上的随机变量可导出 $(\mathbb{R}, \mathscr{B}^1)$ 上的概率测度)

Each r.v. X on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ induces a probability space $(\mathbb{R}, \mathscr{B}^1, \mu)$ by means of the following correspondence:

$$\forall\ B \in \mathscr{B}^1: \quad \mu(B) = \mathbb{P}\{X^{-1}(B)\} = \mathbb{P}\{X \in B\}. \tag{3}$$

Proof. **只需验证 μ 是概率测度**. 首先, $\forall\ B \in \mathscr{B}^1$, 有 $\mu(B) \geqslant 0$. 再设 $(B_k)_{k\geqslant 1} \subseteq \mathscr{B}^1$ 是互不相交集合列, 由引理 2 可知 $\left(X^{-1}(B_k)\right)$ 互不相交. 则

$$\mu\Big(\bigcup_{k=1}^{\infty} B_k\Big) = \mathbb{P}\Big(X^{-1}\big(\bigcup_{k=1}^{\infty} B_k\big)\Big) = \mathbb{P}\Big(\bigcup_{k=1}^{\infty} X^{-1}(B_n)\Big) = \sum_{k=1}^{\infty} \mathbb{P}(X^{-1}(B_n)) = \sum_{k=1}^{\infty} \mu(B_n).$$

最后, 又因为 $X^{-1}(\mathbb{R}) = \Omega$, 于是 $\mu(\mathbb{R}) = 1$. 这就说明 μ 是概率测度. ∎

随机变量与分布函数之间的关系

Remark (随机变量与分布函数之间的关系)

- The collection of sets $\{X^{-1}(S), S \subseteq \mathbb{R}\}$ is a σ-algebra for any function X. trivial
- If X is a r.v, then the collection $\{X^{-1}(B), B \in \mathscr{B}^1\}$ is called the σ-**algebra generated by** X. It is the smallest σ-algebra of $\mathscr{F}$ which contains all sets of the form $\{\omega : X(\omega) \leqslant x\}$, where $x \in \mathbb{R}$, i.e. the smallest σ-algebra on Ω that makes X a random variable.
- Thus (3) is a convenient way of representing the measure $\mathbb{P}$ when it is restricted to this σ-algebra; symbolically we may write it as follows:

$$\mu = \mathbb{P} \circ X^{-1}.$$

- This μ is called the **probability distribution measure** or p.m. of X, and its associated d.f. F according to Theorem 2.12[1] will be called the d.f. of X.
- Specifically, F is given by

$$F(x) = \mu\big((-\infty, x]\big) = \mathbb{P}(X \leqslant x).$$

While *the r.v. X determines μ and therefore F, the converse is obviously false*.

[1]Theorem 12 in Chapter 2 of this note.

分布函数并不能够唯一决定随机变量

Remark (随机变量与分布函数之间的关系)

- Why a d.f. F cannot determines a r.v. X?
- Example. Suppose $X \sim N(0,1)$, then

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

But $F_X(x)$ is not just the d.f. for the r.v. X. In fact, $F_X(x)$ is also the d.f. of $Y = -X$.

定义 5 (Identically Distributed)

A family of r.v.'s having the same distribution is said to be **identically distributed**.

- The condition that X is identically distributed with Y is weaker than $X = Y$ a.s. The former may not have much bearing (与 $\cdots$ 有关) on the latter.
- One could have $X =_d Y$[2] even if $\mathbb{P}(X \neq Y) = 1$. The examples on this page just satisfy this assertion.

[2]This notation indicates that the r.v. X is identically distributed with Y.

随机变量的例子

例 6 (离散概率空间中任一数值型函数均是随机变量)

Let $(\Omega, \mathscr{S})$ be a discrete sample space (see Example 8 of Chapter 2). Every numerically valued function is an r.v.

例 7 (第二章例 9 续)

Consider the probability space $((0,1], \mathscr{B}, m)$. In this case an r.v. is by definition just a **Borel measurable function**. According to the usual definition, f on $(0,1]$ is a Borel measurable function iff $f^{-1}(\mathscr{B}^1) \subseteq \mathscr{B}$. In particular, the function f given by $f(\omega) \equiv \omega$ is an r.v. The two r.v.'s ω and $1-\omega$ are not identical but are identically distributed; in fact their common distribution is the underlying measure m.

例 8 $((\mathbb{R}, \mathscr{B}^1, \mu))$

The definition of a Borel measurable function is not affected, since no measure is involved; so any such function is an r.v., whatever the given p.m. μ may be. As in Example 7, there exists an r.v. with the underlying μ as its p.m.

- We proceed to produce new r.v.'s from given ones.

定理 9 (Function of a Random Variable is Still a Random Variable)

If X is an r.v., f is a Borel measurable function on $(\mathbb{R}, \mathscr{B}^1)$, then $f(X)$ is an r.v.

Proof. 随机变量 X 是 ω 的函数, 于是 $f(X)$ 可以看作 ω 的复合函数:

$$f \circ X : \Omega \to \mathbb{R}, \quad \omega \mapsto f(X(\omega)),$$

另外, 对于复合函数, 我们有 $(f \circ X)^{-1} = X^{-1} \circ f^{-1}$, 于是

$$(f \circ X)^{-1}(\mathscr{B}^1) = X^{-1}\big(f^{-1}(\mathscr{B}^1)\big) \subseteq X^{-1}(\mathscr{B}^1) \subseteq \mathscr{F}.$$

这就说明 $f(X)$ 是随机变量. ■

Remark. The reader who is not familiar with operations of this kind is advised to spell out the proof above in the old-fashioned manner, which takes only a little longer.

Borel σ-algebra of 2-dimension

- **Random vector** is just a vector each of whose components is an r.v.
- It is sufficient to consider the case of two dimensions, since there is no essential difference in higher dimensions apart from complication in notation.
- the Borel σ-algebra $\mathscr{B}^2$ is generated by rectangles of the form

$$\{(x,y) : a < x \leqslant b, c < y \leqslant d\}.$$

  It is also generated by *product sets* of the form

$$B_1 \times B_2 = \{(x,y) : x \in B_1, y \in B_2\},$$

  where B_1 and B_2 belongs to $\mathscr{B}^1$.
- The collection of sets, each of which is a finite union of disjoint product sets, forms an algebra $\mathscr{B}_0^2$.
- A function from $\mathbb{R}^2$ into $\mathbb{R}$ is called a **Borel measurable function (of two variables)** iff $f^{-1}(\mathscr{B}^1) \subseteq \mathscr{B}^2$. Written out, this says that for each 1-dimensional Borel set B, the set

$$\{(x,y) : f(x,y) \in B\}$$

  is a 2-dimensional Borel set, viz. a member of $\mathscr{B}^2$.

Random Vector

- Now let X and Y be two r.v.'s on $(\Omega, \mathscr{F}, \mathbb{P})$. The random vector (X, Y) induces a probability ν on $\mathscr{B}^2$ as follows:

$$\forall\ A \in \mathscr{B}^2:\ \nu(A) = \mathbb{P}\{(X, Y) \in A\} = \mathbb{P}\{\omega : (X(\omega), Y(\omega)) \in A\}. \tag{4}$$

  This ν is called the (2-dimensional, probability) **distribution** or simply the p.m. of (X, Y).

- Let us define the **inverse mapping** $(X, Y)^{-1}$ by the following formula:

$$\forall\ A \in \mathscr{B}^2:\ (X, Y)^{-1}(A) = \{\omega : (X(\omega), Y(\omega)) \in A\}.$$

  The mapping has properties analogous to those of X^{-1} given in lemma 2.

- We can now easily generalize Theorem 9.

定理 10 (Function of a Random Vector is a Random Variable)

If X and Y are r.v.'s and f is a Borel measurable function of two variables, then $f(X, Y)$ is an r.v.

Proof. 类似于定理 9 的证明, 我们有

$$[f\circ(X,Y)]^{-1}(\mathscr{B}^1)=(X,Y)^{-1}\circ f^{-1}(\mathscr{B}^1)\subseteq(X,Y)^{-1}(\mathscr{B}^2)\subseteq\mathscr{F}.$$

其中上式最后一个包含关系说明 (X,Y) 将每个二维 Borel 集映到 $\mathscr{F}$ 中去. 我们来证明这个事实. 不妨设 $B=B_1\times B_2$, 其中 $B_1\in\mathscr{B}^1$, $B_2\in\mathscr{B}^1$, 从而

$$\begin{aligned}(X,Y)^{-1}(B)&=\{\omega:(X(\omega),Y(\omega))\in B_1\times B_2\}\\&=\{\omega:X(\omega)\in B_1,Y(\omega)\in B_2\}\\&=X^{-1}(B_1)\cap Y^{-1}(B_2)\in\mathscr{F}.\end{aligned}$$

现在, 设 $\mathscr{A}=\{A\subseteq\mathbb{R}^2:(X,Y)^{-1}(A)\in\mathscr{F}\}$, 由引理 2, 类似于定理 3 的证明, $\mathscr{A}$ 是一个 σ-代数. 又因为所有形如 $B=B_1\times B_2$ 的集合均属于 $\mathscr{A}$, 于是 $\mathscr{B}_0^2\subseteq\mathscr{A}$. 最后再利用生成集类的性质, 有 $\mathscr{B}^2\subseteq\mathscr{A}$. 于是包含关系得证. ■

Special Cases of Theorem 9 and 10

- Throughout the notes we shall use the notation for numbers as well as functions:

$$x \vee y = \max(x, y), \quad x \wedge y = \min(x, y). \tag{5}$$

推论 11 (Special Cases of Theorem 9 and 10)

If X is an r.v. and f is a continuous function on $\mathbb{R}$, then $f(X)$ is an r.v.; in particular X^r for positive integer r, $|X|^r$ for positive real r, $e^{-\lambda X}$, $e^{\mathrm{i}tX}$ for real λ and t, are all r.v.'s (the last being complex-valued). If X and Y are r.v.'s, then

$$X \vee Y, \quad X \wedge Y, \quad X + Y, \quad X - Y, \quad XY, \quad X/Y$$

are r.v.'s, the last provided Y does not vanish.

- Generalization to a finite number of r.v.'s is immediate.

Generalization to an Infinite Sequence

- Passing to an infinite sequence, let us state the following theorem, although its analogue in real analysis should be well known to the reader.

定理 12 (Generalization to an Infinite Sequence)

If $(X_k)_{k\geqslant 1}$ is a sequence of r.v.'s, then

$$\inf_{k\geqslant 1} X_k, \quad \sup_{k\geqslant 1} X_k, \quad \liminf_{k\to\infty} X_k, \quad \limsup_{k\to\infty} X_k$$

are r.v.'s, not necessarily finite-valued with probability one though every defined, and

$$\lim_{k\to\infty} X_k$$

is an r.v. on the set Δ on which there is either convergence or divergence to $\pm\infty$.

- The proof is omitted. The reader is referred to the proof of Theorem 13 in Chapter 4 of the MATH4123 lecture notes.

Discrete Random Variables

- Here already we see the necessity of the general definition of an r.v. given at the beginning of this section.

定义 13 (Discrete Random Variables)

An r.v. X is called **discrete** (or **countably valued**) iff there is a countable set $B \subseteq \mathbb{R}$ such that $\mathbb{P}(X \in B) = 1$.

定理 14 (X is discrete iff its d.f. is)

An r.v. X is discrete iff its d.f. F is discrete.

Proof. ($\Leftarrow$). 若 F 离散, 则 $F(x) = \sum_{k=1}^{\infty} p_k \delta_{a_k}(x)$, 其中 $\sum_{k=1}^{\infty} p_k = 1$. 令 $C = (a_k)_{k \geqslant 1}$, 则有

$$\mathbb{P}(C) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} \{a_k\}\right) = \sum_{k=1}^{\infty} \mathbb{P}(\{a_k\}) = \sum_{k=1}^{\infty} (F(a_k) - F(a_k-)) = \sum_{k=1}^{\infty} p_k = 1.$$

($\Rightarrow$). 若 X 离散, 则存在 $C = (a_k)_{k \geqslant 1}$, s.t. $\mathbb{P}(X \in C) = 1$. 于是

$$F(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X \in (-\infty, x] \cap C) = \sum_{a_k \leqslant x} \mathbb{P}(\{a_k\}) = \sum_{k=1}^{\infty} \mathbb{P}(\{a_k\}) \delta_{a_k}(x).$$

令 $p_k = \mathbb{P}(\{a_k\})$, 于是便可知 F 是离散的. ■

- The following terminology and notation will be used throughout the book for an arbitrary Ω, not necessarily the sample space.

定义 15 (Indicator Function)

For each $\Delta \subseteq \Omega$, the function $\mathbf{1}_\Delta(\cdot)$ defined as follows:

$$\forall\ \omega \in \Omega: \quad \mathbf{1}_\Delta(\omega) = \begin{cases} 0, & \text{if } \omega \in \Omega \backslash \Delta, \\ 1, & \text{if } \omega \in \Delta, \end{cases}$$

is called the **indicator** (**function**) of Δ.

- Clearly $\mathbf{1}_\Delta$ is an r.v. if and only if $\Delta \in \mathscr{F}$.
- A **countable partition** of Ω is a countable family of disjoint sets $(\Lambda_k)_{k\geqslant 1} \subseteq \mathscr{F}$ such that $\Omega = \bigcup\limits_{k\geqslant 1} \Lambda_k$. We have then

$$\mathbf{1} = \mathbf{1}_\Omega = \sum_{k\geqslant 1} \mathbf{1}_{\Lambda_k}.$$

R.V. of A Weighted Partition

- More generally, let b_k be arbitrary real numbers, then the function φ defined below:

$$\forall\ \omega \in \Omega:\ \varphi(\omega) = \sum_{k=1}^{\infty} b_k \mathbf{1}_{\Lambda_k}(\omega),$$

  is a discrete r.v. We shall call φ the r.v. **belonging to the weighted partition** $(\Lambda_k; b_k)$.
- Each discrete r.v. X belongs to a certain weighted partition (**任一离散随机变量 X 对应于一个加权分割**).
- For let (b_k) be the countable set in the definition of X and let $\Lambda_k = \{\omega : X(\omega) = b_k\}$, then X belongs to the weighted partition $(\Lambda_k; b_k)$.
- If k ranges over a finite index set, the partition is called **finite** and the r.v. belonging to it **simple**. At this point X is the *simple function* defined in MATH4123.

Back to Content.

Section 2

Expectation

- Mathematical expectations correspond to integrals in real analysis, so we already have a basic understanding of their definitions and properties.
- Next, we discuss the relationship between the integrals with respect to $\mathbb{P}$ and the integrals with respect to the Stieltjes measure.
- Finally, the definition and properties of moments are discussed.

Step1. Expectation of Positive Discrete R.V.

- The concept of "(mathematical) expectation" is the same as that of integration in the probability space with respect to the measure $\mathbb{P}$.
- The r.v.'s below will be tacitly assumed to be *finite everywhere* to avoid trivial complications.

- **For each positive discrete r.v. X belonging to the weighted partition** (Λ_k, b_k), we define its *expectation* to be

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} b_k \mathbb{P}(\Lambda_k). \tag{6}$$

  This is either a positive finite number or $+\infty$.
- *Well-defined*. It is trivial that if X belongs to different partitions, the corresponding values given by (6) agree. *The proof can be found in MATH4123.*

Step 2.1. Expectation of Arbitrary Positive R.V.

- **Let X be an arbitrary positive r.v.** For any two positive integers m and n, the set

$$\Lambda_{mn} = \left\{\omega : \frac{n}{2^m} \leqslant X(\omega) < \frac{n+1}{2^m}\right\} \in \mathscr{F}.$$

For each m, let X_m denote the r.v. belonging to the weighted partition $\{\Lambda_{mn}; n/2^m\}$; thus $X_m = n/2^m$ if and only if $n/2^m \leqslant X < (n+1)/2^m$. It is easy to see that we have for each m:

$$\forall\ \omega:\ X_m(\omega) \leqslant X_{m+1}(\omega); \quad 0 \leqslant X(\omega) - X_m(\omega) < \frac{1}{2^m}.$$

Consequently there is *monotone convergence*:

$$\forall\ \omega:\ \lim_{m\to\infty} X_m(\omega) = X(\omega).$$

- The above operation shows that: *any positive r.v. can be approximated by an increasing non-negative discrete r.v. sequence*.

Step 2.2. - 3. Expectation of Arbitrary (Positive) R.V.

- The expectation X_m has just been defined: it is

$$\mathbb{E}(X_m) = \sum_{n=0}^{\infty} \frac{n}{2^m} \mathbb{P}\left\{\frac{n}{2^m} \leqslant X < \frac{n+1}{2^m}\right\}.$$

If for one value of m we have $\mathbb{E}(X_m) = +\infty$, then we define $\mathbb{E}(X) = +\infty$; otherwise we define

$$\mathbb{E}(X) = \lim_{m\to\infty} \mathbb{E}(X_m),$$

the limit existing, finite or infinite, since $\mathbb{E}(X_m)$ is an increasing sequence of real numbers.

- **For an arbitrary r.v.** X, put as usual

$$X = X^+ - X^- \quad \text{where} \quad X^+ = X \vee 0, \quad X^- = (-X) \vee 0. \tag{7}$$

Both X^+ and X^- are positive r.v.'s, and so their expectations are defined. Unless both $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are $+\infty$, we define

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \tag{8}$$

with the usual convention regarding ∞.

General Definition of Expectation; Lebesgue Integration

- We say X has a **finite** or **infinite expectation** (or **expected value**) according as $\mathbb{E}(X)$ is a finite number or $\pm\infty$.
- In the expected case we say that the expectation of X **does not exist**, i.e. $\mathbb{E}(X) = \pm\infty$.
- The expectation, when it exists, is also denoted by

$$\int_\Omega X(\omega)\mathbb{P}(d\omega).$$

More generally, for each Λ in $\mathscr{F}$, we define

$$\boxed{\int_\Lambda X(\omega)\mathbb{P}(d\omega) = \mathbb{E}(X \cdot \mathbf{1}_\Lambda)} \tag{9}$$

and call it **the integral of** X **(with respect to** $\mathbb{P}$**) over the set** Λ. We shall say that X is **integrable with respect to** $\mathbb{P}$ **over** Λ iff the integral above exists and is finite.
- **In the case of** $(\mathscr{U}, \mathscr{B}, m)$, the integral reduces to the ordinary Lebesgue integral

$$\int_a^b f(x)m(dx) = \int_a^b f(x)\,dx.$$

Here m is atomless, so the notation is adequate and there is no need to distinguish between the different kinds of intervals.

Integral with Respect to Stieltjes Measure

- **In the case of** $(\mathbb{R}, \mathscr{B}^1, \mu)$, if we write $X = f$, $\omega = x$, the integral

$$\int_\Lambda X(\omega)\mathbb{P}(d\omega) = \int_\Lambda f(x)\mu(dx)$$

is just the ordinary **Lebesgue-Stieltjes integral of f with respect to μ**.

- If F is the d.f. of μ and $\Lambda = (a, b]$, this is also written as

$$\int_{(a,b]} f(x)\, dF(x).$$

This classical notation is really an anachronism (不合时宜), originated in the days when point function was more popular than a Set function.

- The notation above must then amend to

$$\int_{a+0}^{b+0}, \int_{a-0}^{b+0}, \int_{a+0}^{b-0}, \int_{a-0}^{b-0}$$

to distinguish clearly between the four kinds of intervals $(a, b]$, $[a, b]$, (a, b), $[a, b)$.

Properties of Integral with Respect to $\mathbb{P}$

- The general integral has the familiar properties of the Lebesgue integral on $[0,1]$.
- As a general notation, the left member of (9) will be abbreviated to $\int_\Lambda X\, d\mathbb{P}$.
- In the following, X, Y are r.v.'s; a, b are constants, Λ is a set in $\mathscr{F}$.
- **Absolute Integrability**. $\int_\Lambda X\, d\mathbb{P}$ is finite if and only if

$$\int_\Lambda |X|\, d\mathbb{P} < \infty.$$

- **Linearity**.

$$\int_\Lambda (aX + bY)\, d\mathbb{P} = a\int_\Lambda X\, d\mathbb{P} + b\int_\Lambda Y\, d\mathbb{P}$$

  provided that the right side is meaningful, namely not $+\infty - \infty$ or $-\infty + \infty$.
- **Additivity over Sets**. If the Λ_n's are disjoint, then

$$\int_{\cup_n \Lambda_n} X\, d\mathbb{P} = \sum_{n=1}^{\infty} \int_{\Lambda_n} X\, d\mathbb{P}.$$

- **Positivity**. If $X \geqslant 0$ a.s. on Λ, then

$$\int_\Lambda X\, d\mathbb{P} \geqslant 0.$$

Properties of Integral with Respect to $\mathbb{P}$

- **Monotonicity**. If $X_1 \leqslant X \leqslant X_2$ a.s. on Λ, then

$$\int_\Lambda X_1 \, d\mathbb{P} \leqslant \int_\Lambda X \, d\mathbb{P} \leqslant \int_\Lambda X_2 \, d\mathbb{P}.$$

- **Mean Value Theorem**. If $a \leqslant X \leqslant b$ a.s. on Λ, then

$$a\mathbb{P}(\Lambda) \leqslant \int_\Lambda X \, d\mathbb{P} \leqslant b\mathbb{P}(\Lambda).$$

- **Modulus Inequality**.

$$\left| \int_\Lambda X \, d\mathbb{P} \right| \leqslant \int_\Lambda |X| \, d\mathbb{P}.$$

- **Dominated Convergence Theorem**. If $\lim_{n\to\infty} X_n = X$ a.s. or merely in measure on Λ and $\forall\ n: |X_n| \leqslant Y$ a.s. on Λ, with $\int_\Lambda Y \, d\mathbb{P} < \infty$, then

$$\lim_{n\to\infty} \int_\Lambda X_n \, d\mathbb{P} = \int_\Lambda X \, d\mathbb{P} = \int_\Lambda \lim_{n\to\infty} X_n \, d\mathbb{P}. \tag{10}$$

- **Bounded Convergence Theorem**. If $\lim_{n\to\infty} X_n = X$ a.s. or merely in measure on Λ and there exists a constant M such that $\forall\ n$: $|X_n| \leqslant M$ a.s. on Λ, then (10) is true.
- **Monotone Convergence Theorem**. If $X_n \geqslant 0$ and $X_n \uparrow X$ a.s. on Λ, then (10) is again true provided that $+\infty$ is allowed as a value for either member. The condition "$X_n \geqslant 0$" may be weakened to : "$\mathbb{E}(X_n) > -\infty$ for some n".
- **Integration Term by Term**. If

$$\sum_{n=1}^{\infty} \int_{\Lambda} |X_n| \, d\mathbb{P} < \infty,$$

then $\sum_{n=1}^{\infty} |X_n| < \infty$ a.s. on Λ so that $\sum_{n=1}^{\infty} X_n$ converges a.s. on Λ and

$$\int_{\Lambda} \sum_{n=1}^{\infty} X_n \, d\mathbb{P} = \sum_{n=1}^{\infty} \int_{\Lambda} X_n \, d\mathbb{P}.$$

- **Fatou's Lemma**. If $x_n \geqslant 0$ a.s. on Λ, then

$$\int_{\Lambda} \underline{\lim}_{n\to\infty} X_n \, d\mathbb{P} \leqslant \underline{\lim}_{n\to\infty} \int_{\Lambda} X_n \, d\mathbb{P}.$$

A Useful Estimate of Mathematical Expectation

- Let us prove the following useful theorem as an instructive example.

定理 16 (A Useful Estimate of Mathematical Expectation)

We have

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geqslant n) \leqslant \mathbb{E}(|X|) \leqslant 1 + \sum_{n=1}^{\infty} \mathbb{P}(|X| \geqslant n) \tag{11}$$

so that $\mathbb{E}(|X|) < \infty$ *if and only if the series above converges.*

Proof. 利用积分的可列可加性, 若 $\Lambda_n = \{n \leqslant |X| < n+1\}$, 则

$$\mathbb{E}(|X|) = \sum_{n=0}^{\infty} \int_{\Lambda_n} |X| \, d\mathbb{P}.$$

再将积分中值定理应用于每个 Λ_n, 就有

$$\sum_{n=0}^{\infty} n\mathbb{P}(\Lambda_n) \leqslant \mathbb{E}(|X|) \leqslant \sum_{n=0}^{\infty} (n+1)\mathbb{P}(\Lambda_n) = 1 + \sum_{n=0}^{\infty} n\mathbb{P}(\Lambda_n). \tag{12}$$

(12) 式中等号成立是因为 $\sum_{n=0}^{\infty} \mathbb{P}(\Lambda_n) = \mathbb{P}(\Omega) = 1$.

A Useful Estimate of Mathematical Expectation

现在我们只需证明

$$\sum_{n=0}^{\infty} n\mathbb{P}(\Lambda_n) = \sum_{n=0}^{\infty} \mathbb{P}(|X| \geqslant n) \quad \text{有限或无限.} \tag{13}$$

注意到

$$\begin{aligned}
\mathbb{P}(\Lambda_n) &= \mathbb{P}\{n \leqslant |X| < n+1\} = \mathbb{P}(|X| < n+1) - \mathbb{P}(|X| < n) \\
&= \big[1 - \mathbb{P}(|X| \geqslant n+1)\big] - \big[1 - \mathbb{P}(|X| \geqslant n)\big] \\
&= \mathbb{P}(|X| \geqslant n) - \mathbb{P}(|X| \geqslant n+1),
\end{aligned}$$

于是

$$\begin{aligned}
\sum_{n=0}^{N} n\mathbb{P}(\Lambda_n) &= \sum_{n=0}^{N} n\big[\mathbb{P}(|X| \geqslant n) - \mathbb{P}(|X| \geqslant n+1)\big] \\
&= \sum_{n=0}^{N} n\mathbb{P}(|X| \geqslant n) - \sum_{n=1}^{N+1} (n-1)\mathbb{P}(|X| \geqslant n) \\
&= \sum_{n=1}^{N} \big(n - (n-1)\big)\mathbb{P}(|X| \geqslant n) - N\mathbb{P}(|X| \geqslant N+1) \\
&= \sum_{n=1}^{N} \mathbb{P}(|X| \geqslant n) - N\mathbb{P}(|X| \geqslant N+1).
\end{aligned}$$

A Useful Estimate of Mathematical Expectation

现在我们就有

$$\sum_{n=1}^{N} n\mathbb{P}(\Lambda_n) \leqslant \sum_{n=1}^{N} \mathbb{P}(|X| \geqslant n) \leqslant \sum_{n=1}^{N} n\mathbb{P}(\Lambda_n) + N\mathbb{P}(|X| \geqslant N+1). \tag{14}$$

对 $N\mathbb{P}(|X| \geqslant N+1)$ 继续使用中值定理, 我们有

$$N\mathbb{P}(|X| \geqslant N+1) \leqslant \int_{\{|X| \geqslant N+1\}} |X|\, d\mathbb{P} \leqslant \mathbb{E}(|X|).$$

于是, 若 $\mathbb{E}(|X|) < \infty$, 于是 (14) 式中最后一项 $N\mathbb{P}(|X| \geqslant N+1) \to 0$ $(N \to \infty)$. 此时

$$\sum_{n=1}^{\infty} n\mathbb{P}(\Lambda_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(|X| \geqslant n) \leqslant \sum_{n=1}^{\infty} n\mathbb{P}(\Lambda_n).$$

于是 (13) 式成立. 若 $\mathbb{E}(|X|) = \infty$, 则由 (12) 式中第二个不等式可知

$$\infty = \mathbb{E}(|X|) \leqslant 1 + \sum_{n=0}^{\infty} n\mathbb{P}(\Lambda_n),$$

从而 $\sum_{n=1}^{\infty} n\mathbb{P}(\Lambda_n) \to \infty$ $(N \to \infty)$. 再由 (14) 式可知 (13) 式成立. 定理得证. ■

A Useful Estimate of Mathematical Expectation

- If the r.v. X takes the value of a positive integer, its expectation has the following representation:

推论 17

If X takes only positive integer values, then

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n).$$

PROOF. 如果随机变数 X 取值为正整数, 则 X 是离散型随机变量, 由 (6) 式, 便有

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=1}^{\infty} n\mathbb{P}(X = n) \\
&= \sum_{n=1}^{\infty} n\mathbb{P}(n \leqslant X < n+1) \quad \text{[利用 (13) 式]} \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n). \quad \blacksquare
\end{aligned}$$

- There is a basic relation between the abstract integral with respect to $\mathbb{P}$ over sets in $\mathscr{F}$ on the one hand, and the Lebesgue-Stieltjes integral with respect to μ over sets in $\mathscr{B}^1$ on the other, induced by each r.v.

定理 18 (Change of Variables Formula)

Let X on $(\Omega, \mathscr{F}, \mathbb{P})$ induce the probability space $(\mathbb{R}, \mathscr{B}^1, \mu)$ according to Theorem 4 and let f be Borel measurable. Then we have

$$\mathbb{E} f(X) = \int_{\Omega} f(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}} f(x) \mu(dx) = \int_{\mathbb{R}} f(x)\, dF(x) \tag{15}$$

provided that either side exists. 前提是任何一边都存在.

HINT. 本定理的证明技巧是标准流程.

1. 首先证明示性函数 (indicator function) 满足定理结论.
2. 利用积分的线性证明离散型随机变量满足定理结论.
3. 利用积分的单调收敛定理证明非负可测函数满足定理结论.
4. 最后由 $f = f^+ - f^-$ 证明一般可测函数满足定理结论.

Proof. Step 1. **示性函数**. 令 $B \in \mathscr{B}^1$ 且 $f = \mathbf{1}_B$. 则

$$\int_\Omega f(X(\omega))\mathbb{P}(d\omega) = \int_{\{X\in B\}} 1\,\mathbb{P}(d\omega) = \mathbb{P}(X \in B).$$

$$\int_{\mathbb{R}} f(x)\mu(dx) = \int_B 1\,\mu(dx) = \mu(B).$$

由定理 4 中 (3) 式, 可知 $\mathbb{P}(X \in X) = \mu(B)$, 进而示性函数满足 (15) 式.

Step 2. **离散型随机变量**. 由积分的线性, 不难证明 (15) 式对示性函数的线性组合成立:

$$f = \sum_{k=1}^{\infty} b_k \mathbf{1}_{B_k}, \quad \text{其中 } (B_k; b_k) \text{ 是 weighted partition.}$$

Step 3. **非负可测函数**. 任取非负 Borel 可测函数 f, 由可测函数的构造, 可以找到单调递增广义简单函数列 $(f_m)_{m\geq 1}$ 使得 $f_m \uparrow f$, 其中每个 f_m 均满足

$$\int_\Omega f_m(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}} f_m(x)\mu(dx).$$

令 $m \to \infty$, 由单调收敛定理, 便可知 (15) 式对一般非负可测函数成立. 一般情形参考上页提示. 定理得证. ■

General Case of Change of Variables Formula

- Although we have not yet defined the integral of a multivariate function, we place the relevant conclusions here first.
- We shall need the generalization of the preceding theorem in several dimensions. No change is necessary except for notation, which we will give in two dimensions.
- Instead of the ν in (4), let us write the "mass element" as $\mu^2(dx, dy)$ so that

$$\nu(A) = \iint_A \mu^2(dx, dy).$$

定理 19 (General Case of Change of Variables Formula)

Let (X, Y) on $(\Omega, \mathscr{F}, \mathbb{P})$ induce the probability space $(\mathbb{R}^2, \mathscr{B}^2, \mu^2)$ and let f be a Borel measurable function of two variables. Then we have

$$\mathbb{E} f(X, Y) = \int_\Omega f(X(\omega), Y(\omega)) \mathbb{P}(d\omega) = \iint_{\mathbb{R}^2} f(x, y) \mu^2(dx, dy). \tag{16}$$

- Note that $f(X, Y)$ is a r.v. by Theorem 10.

Remark of Change of Variables Formula

- As a consequence of Theorem 18, we have: if μ_X and F_X denote, respectively, the p.m. and d.f. induced by X, then we have

$$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x)\mu_X(dx) = \int_{-\infty}^{\infty} f(x)\, dF(x) \tag{17}$$

  with the usual proviso regarding existence and finiteness.
- Let μ^2 be as in Theorem 19 and take $f(x,y) = x + y$ here. We obtain

$$\mathbb{E}(X+Y) = \iint_{\mathbb{R}^2} (x+y)\mu^2(dx,dy) = \iint_{\mathbb{R}^2} x\mu^2(dx,dy) + \iint_{\mathbb{R}^2} y\mu^2(dx,dy).$$

  On the other hand, if we take $f(x,y)$ to be x or y, respectively, then we have

$$\mathbb{E}(X) = \iint_{\mathbb{R}^2} x\mu^2(dx,dy), \quad \mathbb{E}(Y) = \iint_{\mathbb{R}^2} y\mu^2(dx,dy),$$

  and consequently $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. We reduce the additivity of abstract integrals to the additivity of Stieltjes integrals, such **reduction** is *helpful* when there are technical difficulties in the abstract treatment.[3]

[3] This is not quite rigorous, since we have not yet defined the integral of a multivariate function.

Definition of Moments

- Let a be real, r positive, then $\mathbb{E}(|X-a|^r)$ is called the **absolute moment of X of order r, about a**. It may be $+\infty$;
- otherwise, if r is an integer, $\mathbb{E}(X-a)^r = \mathbb{E}\big((X-a)^r\big)$ is the correspond **moment** (矩).
- If μ and F are, respectively, the p.m. and d.f. of X, then we have by Theorem 18:

$$\mathbb{E}(|X-a|^r) = \int_{\mathbb{R}} |x-a|^r \mu(dx) = \int_{-\infty}^{\infty} |x-a|^r \, dF(x),$$
$$\mathbb{E}(X-a)^r = \int_{\mathbb{R}} (x-a)^r \mu(dx) = \int_{-\infty}^{\infty} (x-a)^r \, dF(x).$$

- For $r=1$, $a=0$, this reduces to $\mathbb{E}(X)$, which is also called the **mean** of X.
- The moments about the mean are called **central moments** (中心矩).
- Central moment of order 2 is called the **variance** (方差), $\mathrm{var}(X)$; its positive square root the **standard deviation** (标准差). $\sigma(X)$:

$$\mathrm{var}(X) = \sigma^2(X) = \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

We note the inequality $\sigma^2(X) \leqslant \mathbb{E}(X^2)$, which will be used a good deal in Law of Large Numbers.

L^p Space

- For any positive number p, X is said to belong to $L^p = L^p(\Omega, \mathscr{F}, \mathbb{P})$ iff $\mathbb{E}(|X|^p) < \infty$.
- The well known inequalities of Hölder and Minkowski may be written as follows.
- Let X and Y be r.v.'s, $1 < p < \infty$ and $1/p + 1/q = 1$, then

$$\text{Hölder's Inequality}: \quad |\mathbb{E}(XY)| \leqslant \mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^p)^{\frac{1}{p}} \mathbb{E}(|X|^q)^{\frac{1}{q}},$$

$$\text{Minkowski's Inequality}: \quad \mathbb{E}(|X+Y|^p)^{\frac{1}{p}} \leqslant \mathbb{E}(|X|^p)^{\frac{1}{p}} + \mathbb{E}(|X|^p)^{\frac{1}{p}}.$$

- If $Y \equiv 1$ in Hölder's inequality, we obtain

$$\mathbb{E}(|X|) \leqslant \mathbb{E}(|X|^p)^{\frac{1}{p}}; \tag{18}$$

  for $p = 2$, Hölder's inequality is called the **Cauchy-Schwarz** inequality.
- Replacing $|X|$ by $|X|^r$, where $0 < r < p$, and writing $r' = pr$ in (18) we obtain

$$\mathbb{E}(|X|^r)^{\frac{1}{r}} \leqslant \mathbb{E}(|X|^{r'})^{\frac{1}{r'}}, \quad 0 < r < r' < \infty. \tag{19}$$

  The last will be referred to as the **Liapounov inequality**.

定理 20 (Jensen's Inequality)

If φ *is a convex function on* $\mathbb{R}$, *that is,*

$$\varphi(\lambda x+(1-\lambda)y)\leqslant\lambda\varphi(x)+(1-\lambda)\varphi(y)$$

for all $\lambda\in(0,1)$, $x,y\in\mathbb{R}$; X *and* $\varphi(X)$ *are integrable r.v.'s, then*

$$\varphi(\mathbb{E}X)\leqslant\mathbb{E}\varphi(X). \tag{20}$$

- We first need the following lemma: *Reference: [YSJ23] Lemma 8.3.11.*

引理 21

If φ *is a convex function on* $\mathbb{R}$, *then for all* $x\in\mathbb{R}$, $\varphi'_-(x)$ *and* $\varphi'_+(x)$ *both exist and finite, and for all* $x<y$, *we have*

$$\varphi'_-(x)\leqslant\varphi'_+(x)\leqslant\frac{\varphi(y)-\varphi(x)}{y-x}\leqslant\varphi'_-(y)\leqslant\varphi'_+(y). \tag{21}$$

Proof of Lemma 21. 由凸函数的定义知 $\forall\ x \leqslant y$, φ 在 (a,b) 区间上的图像都在连接 $(x,\varphi(x))$ 与 $(y,\varphi(y))$ 的直线下方. 任取 $x_1 < x < x_2$, 有

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leqslant \frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1} \leqslant \frac{\varphi(x_2)-\varphi(x)}{x_2-x}. \tag{22}$$

事实上, 因 $x_1 < x < x_2$, 所以存在 $\lambda \in (0,1)$, s.t. $x = \lambda x_1 + (1-\lambda)x_2$. 由凸函数的定义,

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leqslant \frac{\lambda\varphi(x_1)+(1-\lambda)\varphi(x_2)-\varphi(x_1)}{\lambda x_1+(1-\lambda)x_2-x_1} = \frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1}.$$

(22) 中另一不等式类似可得. 其次, 由 (22) 可知, $\forall\ x_1' \in (x_1,x)$, $\forall\ x_2' \in (x,x_2)$, 有

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leqslant \frac{\varphi(x_1')-\varphi(x)}{x_1'-x} \leqslant \frac{\varphi(x_2')-\varphi(x)}{x_2'-x} \leqslant \frac{\varphi(x_2)-\varphi(x)}{x_2-x},$$

在上式中令 $x_1 \uparrow x$ 即可知 $\varphi_-'(x)$ 存在且 $\varphi_-'(x) \leqslant \dfrac{\varphi(x_2')-\varphi(x)}{x_2'-x}$ $(\forall\ x_2' > x)$. 再令 $x_2 \downarrow x$ 即可知 $\varphi_+'(x)$ 存在且 $\varphi_+'(x) \geqslant \dfrac{\varphi(x_1')-\varphi(x)}{x_1'-x}$ $(\forall\ x_1' < x)$, 于是就有 $\varphi_-'(x) \leqslant \varphi_+'(x)$. 至于 $\varphi_+'(x) \leqslant \varphi_-'(y)$ 可由 (22) 式与左右导数的存在性得到. 引理得证. ■

PROOF of Theorem 20. 由引理 21, 可知

$$(y\downarrow x)\quad \frac{\varphi(y)-\varphi(x)}{y-x}\downarrow \varphi'_+(x)\geqslant \varphi'_-(x)\uparrow \frac{\varphi(y)-\varphi(x)}{y-x}\quad (y\uparrow x).$$

并且 φ 在任一点 $x\in\mathbb{R}$ 处存在有限的左导数和右导数, 于是 φ 在任一点 x 处左连续和右连续, 于是 φ 在 $\mathbb{R}$ 上连续. 连续函数均 Borel 可测 (因为开集的原像是开集, 从而 Borel 可测), 所以 $\varphi(X)$ 是随机变量, 且

$$\varphi(y)-\varphi(x)\geqslant \varphi'_+(x)(y-x),\quad \forall\ y\in\mathbb{R}.$$

令 $y=X$, $x=\mathbb{E}X$, 则有

$$\varphi(X)-\varphi(\mathbb{E}X)\geqslant \varphi'_+(\mathbb{E}X)(X-\mathbb{E}X).$$

在上式不等式两端取期望, 由期望 (积分) 的单调性, 注意到 $\mathbb{E}(X-\mathbb{E}X)=\mathbb{E}X-\mathbb{E}X=0$, 于是 $\mathbb{E}\varphi(X)\geqslant \mathbb{E}(\varphi(\mathbb{E}X))=\varphi(\mathbb{E}X)$. 于是 Jensen 不等式得证. ■

Chebyshev's Inequality

- Finally, we prove a famous inequality which is almost trivial but very useful.

定理 22 (Chebyshev's Inequality)

If φ is a strictly positive and increasing function on $(0,\infty)$, $\varphi(u)=\varphi(-u)$, and X is a r.v. such that $\mathbb{E}\varphi(X)<\infty$, then for each $u>0$:

$$\mathbb{P}(|X|\geqslant u)\leqslant\frac{\mathbb{E}\varphi(X)}{\varphi(u)}.$$

PROOF. 由积分中值定理, 我们有

$$\mathbb{E}\varphi(X)=\int_{\Omega}\varphi(X)\,d\mathbb{P}\geqslant\int_{\{|X|\geqslant u\}}\varphi(X)\,d\mathbb{P}\geqslant\varphi(u)\mathbb{P}(|X|\geqslant u).$$

移项便可得 Chebyshev 不等式. ■

- *The most familiar application* is when $\varphi(u)=|u|^p$ for $0<p<\infty$, so that the inequality yields an upper bound for the "tail" probability in terms of an absolute moment.

Back to Content.

Section 3

Independence

- Mathematical expectations correspond to integrals in real analysis, so we already have a basic understanding of their definitions and properties.
- Next, we discuss the relationship between the integrals with respect to $\mathbb{P}$ and the integrals with respect to the Stieltjes measure.
- Finally, the definition and properties of moments are discussed.

Independence of Random Variables

定义 23 (Independence; Pairwise Independence)

The r.v.'s $(X_k)_{k=1}^n$ are said to be (**totally**) **independent** iff for any Borel sets $(B_k)_{k=1}^n$ we have

$$\mathbb{P}\left\{\bigcap_{k=1}^n (X_k \in B_k)\right\} = \prod_{k=1}^n \mathbb{P}(X_k \in B_k). \tag{23}$$

The r.v.'s of an infinite family are said to be independent iff those in every finite subfamily are. They are said to be **pairwise independent** iff every two of them are independent.

The r.v.'s of an *infinite* (not necessarily countable) family are said to be independent iff those in every finite subfamily are.

- Note that (23) implies that the r.v.'s in every subset of $(X_k)_{k=1}^n$ are also independent, since we may take some of the B_k's as $\mathbb{R}$.
- On the other hand, (23) is implied by the apparently weaker hypothesis: for every set of real numbers $(x_k)_{k=1}^n$:

$$\mathbb{P}\left\{\bigcap_{k=1}^n (X_k \leqslant x_k)\right\} = \prod_{k=1}^n \mathbb{P}(X_k \leqslant x_k). \tag{24}$$

This can be derived from Theorem 26 directly. We first recall the $\pi - \lambda$ theorem.

Appendix. $\pi-\lambda$ Theorem

定义 24 (π-System; λ-System)

A class $\mathscr{C}$ of subsets is called a π-**system** if it satisfies $A, B \in \mathscr{C} \;\Rightarrow\; A \cap B \in \mathscr{C}$.

A class Λ of subsets is called a λ-**system** (or **Dynkin system**) if it satisfies

1. $\Omega \in \Lambda$;
2. $A, B \in \Lambda,\ A \supseteq B \;\Rightarrow\; A \backslash B \in \Lambda$;
3. for any monotone increasing sequence $(E_n) \subseteq \Lambda$, we have $\lim E_n \in \Lambda$.

定理 25 ($\pi-\lambda$ 定理: 集合形式的单调类定理)

Let a class $\mathscr{C}$ of subsets of Ω be a $\underline{\pi\text{-system}}$, $\Lambda(\mathscr{C})$ be the smallest λ-system containing $\mathscr{C}$. Then $\Lambda(\mathscr{C}) = \sigma(\mathscr{C})$, where $\sigma(\mathscr{C})$ denotes the σ-algebra generated by $\mathscr{C}$.

- $\pi-\lambda$ 定理相较之前的单调类定理更好用. 因为我们弱化了 $\mathscr{C}$ 需要满足的条件.
- 由上述定理可知: **任何包含 $\mathscr{C}$ 的 λ-系** $\Lambda \supseteq \sigma(\mathscr{C})$.
- 它的用法如下: 通常是已知 $\mathscr{C}$ 中元具有性质 S, 我们需要证明 $\sigma(\mathscr{C})$ 中的元也具有性质 S, 为此令 $\Lambda := \{B \subseteq X : B$ 具有性质 $S\}$, 于是 $\Lambda \supseteq \mathscr{C}$. 再去证明 $\mathscr{C}$ 是 π-系, 再证明 Λ 是 λ-系, 由 $\pi-\lambda$ 定理便可知 $\Lambda \supseteq \sigma(\mathscr{C})$, 即 $\sigma(\mathscr{C})$ 中元素都具有性质 S. 这种方法称为 λ-系方法.

独立事件类的扩张定理

- Our main task now is to show that (24) can imply (23). In fact, we can prove a more general conclusion. *Reference. [YSJ23] Theorem 5.3.3.*

定理 26 (独立事件类的扩张定理)

Let $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space;* $\mathscr{C}_k \subseteq \mathscr{F}$ $(1 \leqslant k \leqslant n)$ *be* π*-systems which contain* Ω*. For all* $A_k \in \mathscr{C}_k$ $(1 \leqslant k \leqslant n)$ *we have*

$$\mathbb{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} \mathbb{P}(A_k), \tag{25}$$

then (25) is true for all $A_k \in \sigma(\mathscr{C}_k)$ $(1 \leqslant k \leqslant n)$*.*

PROOF. Step 1. 对任一给定的 $1 \leqslant l \leqslant n$, 令

$$\Lambda_l = \{A_l \in \sigma(\mathscr{C}_l) : \forall\ A_k \in \mathscr{C}_k, k \neq l, (25)\ 成立\}.$$

则显然有 $\mathscr{C}_l \subseteq \Lambda_l \subseteq \sigma(\mathscr{C}_l)$. 现在去证明 Λ_l 为 $\mathscr{F}$ 中的 λ-系.

(1). 首先, 由假设可知 $\Omega \in \mathscr{C}_l \subseteq \Lambda_l$, 于是 $\Omega \in \Lambda_l$.

独立事件类的扩张定理

(2). 任取 $A, B \in \Lambda_l$ 且 $A \supseteq B$, 则 $\forall\ A_k \in \mathscr{C}_k\ (k \neq l)$, 由 Λ_l 的定义可知 Λ_l **对真差封闭**:

$$\mathbb{P}\Big(\Big(\bigcap_{l\neq k=1}^{n} A_k\Big) \cap (A \backslash B)\Big) = \mathbb{P}\Big(\Big(\bigcap_{l\neq k=1}^{n} A_k\Big) \cap A\Big) - \mathbb{P}\Big(\Big(\bigcap_{l\neq k=1}^{n} A_k\Big) \cap B\Big)$$
$$= \prod_{l\neq k=1}^{n} \mathbb{P}(A_k)\mathbb{P}(A) - \prod_{l\neq k=1}^{n} \mathbb{P}(A_k)\mathbb{P}(B) = \prod_{l\neq k=1}^{n} \mathbb{P}(A_k)\mathbb{P}(A\backslash B).$$

(3). 再设 $(A_{l,m})_{m\geq 1} \subseteq \Lambda_l$ 且 $A_{l,m} \uparrow A_l\ (m \to \infty)$. 则由 $\mathbb{P}$ 的连续性及 Λ_l 的定义可知

$$\mathbb{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \lim_{m\to\infty} \mathbb{P}\Big(\Big(\bigcap_{l\neq k=1}^{n} A_k\Big) \cap A_{l,m}\Big) = \lim_{m\to\infty} \prod_{l\neq k=1}^{n} \mathbb{P}(A_k)\mathbb{P}(A_{l,m}) = \prod_{k=1}^{n} \mathbb{P}(A_k).$$

于是 Λ_l **对不降集列的极限封闭**. 于是 Λ_l 是 λ-系, 由 $\pi-\lambda$ 定理可知 $\Lambda_l \supseteq \sigma(\mathscr{C}_l)$. 于是 $\Lambda_l = \sigma(\mathscr{C}_l)$, 即将假设中的任一 $\mathscr{C}_l$ 替换成 $\sigma(\mathscr{C}_l)$, (25) 式仍然成立.

Step 2. 由第一步可知, 先将假设中的 $\mathscr{C}_1$ 换为 $\sigma(\mathscr{C}_1)$ 仍然可以保证 (25) 式成立. 由于 $\Omega \in \sigma(\mathscr{C}_1)$ 且 $\sigma(\mathscr{C}_1)$ 仍为 π-系, 于是 $\sigma(\mathscr{C}_1), \mathscr{C}_2, \cdots, \mathscr{C}_n$ 仍然满足定理假设, 于是再利用第一步, 将 $\mathscr{C}_2$ 替换为 $\sigma(\mathscr{C}_2)$ 后 (25) 式仍成立. 多次应用第一步, 经过 n 次即可得证: (25) 式对一切 $A_k \in \sigma(\mathscr{C}_k)\ (1 \leq k \leq n)$ 成立. ■

n-Dimensional Distribution Function

- In terms of the p.m. μ^n induced by the random vector $(X_1, \cdots, X_n)$ on $(\mathbb{R}^n, \mathscr{B}^n)$, and the p.m.'s $(\mu_k)_{k=1}^n$ induced by each X_k on $(\mathbb{R}, \mathscr{B}^1)$, the relation (25) may be written as

$$\mu^n\Big(\prod_{k=1}^n B_k\Big) = \prod_{k=1}^n \mu_k(B_k), \tag{26}$$

where $\prod_{k=1}^n B_k = B_1 \times \cdots \times B_n$.

- Finally, we may introduce the n-**dimensional distribution function** corresponding to μ^n, which is defined by the left side of (24) or in alternative notation:

$$F(x_1, \cdots, x_n) = \mathbb{P}\{X_1 \leqslant x_1, \cdots, X_n \leqslant x_n\} = \mu^n\Big(\prod_{k=1}^n (-\infty, x_k]\Big);$$

then (24) may be written as

$$F(x_1, \cdots, x_n) = \prod_{k=1}^n F_k(x_k).$$

Independence of Events

- From now on when the probability space is fixed, a set in $\mathscr{F}$ will also be called an **event**.
- The events $(E_k)_{k=1}^n$ are said to be **mutually independent (相互独立)** iff their indicators are independent; this is equivalent to: for any subset $\{k_1, \cdots, k_\ell\}$ of $\{1, \cdots, n\}$, we have

$$\mathbb{P}\Big(\bigcap_{m=1}^{\ell} E_{k_m}\Big) = \prod_{m=1}^{\ell} \mathbb{P}(E_{k_m}). \tag{27}$$

- It is clear that mutually independent events are pairwise independent. The following example shows that the converse is not true.

例 27 (两两独立无法推出相互独立)

Let X_1, X_2, X_3 be independent random variables with $\mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 1) = 1/2$. Let $A_1 = \{X_2 = X_3\}$, $A_2 = \{X_3 = X_1\}$, and $A_3 = \{X_1 = X_2\}$. These events are pairwise independent since if $i \neq j$, then

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(X_1 = X_2 = X_3) = 1/4 = \mathbb{P}(A_i)\mathbb{P}(A_j),$$

but they are not independent since

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/4 \neq 1/8 = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Functions of Independent R.V.'s

定理 28 (Functions of Independent R.V.'s Are Independent R.V.'s)

If $(X_k)_{k=1}^n$ are independent r.v.'s and $(f_k)_{k=1}^n$ are Borel measurable functions, then $\big(f_k(X_k)\big)_{k=1}^n$ are independent r.v.'s.

PROOF. 任取 $A_k \in \mathscr{B}^1$, 因为 f_k 是 Borel 可测函数, 于是 $f_k^{-1}(A_k) \in \mathscr{B}^1$. 又

$$\bigcap_{k=1}^n \{f_k(X_k) \in A_k\} = \bigcap_{k=1}^n \{X_k \in f_k^{-1}(A_k)\},$$

于是我们有

$$\begin{aligned}\mathbb{P}\Big\{\bigcap_{k=1}^n \{f_k(X_k) \in A_k\}\Big\} &= \mathbb{P}\Big\{\bigcap_{k=1}^n \{X_k \in f_k^{-1}(A_k)\}\Big\} = \prod_{k=1}^n \mathbb{P}\{X_k \in f_k^{-1}(A_k)\} \\ &= \prod_{k=1}^n \mathbb{P}\{f_k(X_k) \in A_k\}.\end{aligned}$$

由 A_k $(1 \leqslant k \leqslant n)$ 的任意性便可知 $f_k(X_k)$ $(1 \leqslant k \leqslant n)$ 是独立随机变量. ■

Functions of Independent R.V.'s

- The proof of the next theorem is similar and is left as an exercise.

定理 29 (Functions of Independent R.V.'s Are Independent R.V.'s)

Let $1 \leqslant n_1 < n_2 < \cdots < n_k = n$; f_1 a Borel measurable function of n_1 variables, f_2 one of $n_2 - n_1$ variables, $\cdots$, f_k one of $n_k - n_{k-1}$ variables. If $(X_j)_{j=1}^n$ are independent r.v.'s then the k r.v.'s

$$f_1(X_1, \cdots, X_{n_1}), f_2(X_{n_1+1}, \cdots, X_{n_2}), \cdots, f_k(X_{n_{k-1}+1}, \cdots, X_{n_k})$$

are independent.

Expectation of the Product of Independent R.V.'s

定理 30 (Expectation of the Product of Independent R.V.'s)

If X and Y are independent and both have finite expectations, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{28}$$

Proof. 按照标准流程证明. **Step 1**. 设 X 与 Y 均为非负离散随机变量, 不妨设 X 对应的加权分割为 $(\Lambda_j; c_j)$; Y 对应的加权分割为 $(M_k; d_k)$, 于是 $\Lambda_j = \{X = c_j\}$, $M_k = \{Y = d_k\}$. 由离散型随机变量的数学期望可知

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} c_j \mathbb{P}(\Lambda_j), \quad \mathbb{E}(Y) = \sum_{k=1}^{\infty} d_k \mathbb{P}(M_k).$$

又因为

$$\Omega = \left(\bigcup_{j=1}^{\infty} \Lambda_j\right) \cap \left(\bigcup_{k=1}^{\infty} M_k\right) = \bigcup_{j=1}^{\infty}\bigcup_{k=1}^{\infty} (\Lambda_j \cap M_k) = \bigcup_{j=1}^{\infty}\bigcup_{k=1}^{\infty} (\Lambda_j M_k),$$

且 $X(\omega)Y(\omega) = c_j d_k$ (若 $\omega \in \Lambda_j M_k$). 于是随机变量 XY 是离散型随机变量, 其对应的加权分割为 $(\Lambda_j M_k; c_j d_k)_{j,k\geqslant 1}$.

Expectation of the Product of Independent R.V.'s

因为 X 与 Y 独立, 于是, 对任意的 $j,k\geqslant 1$, 有

$$\mathbb{P}(\Lambda_j M_k)=\mathbb{P}(X=c_j,Y=d_k)=\mathbb{P}(X=c_j)\mathbb{P}(Y=d_k)=\mathbb{P}(\Lambda_j)\mathbb{P}(Y=d_k).$$

于是, 再次利用离散型随机变量的数学期望可知

$$\mathbb{E}(XY)=\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}c_jd_k\mathbb{P}(\Lambda_jM_k)=\left(\sum_{j=1}^{\infty}c_j\mathbb{P}(\Lambda_j)\right)\left(\sum_{k=1}^{\infty}d_k\mathbb{P}(M_k)\right)=\mathbb{E}(X)\mathbb{E}(Y).$$

Step 2. 设 X 与 Y 均为非负随机变量, 则存在单调不减非负离散型随机变量序列 $(X_m)_{m\geqslant 1}$ 与 $(Y_m)_{m\geqslant 1}$, s.t. $\mathbb{E}(X_m)\uparrow\mathbb{E}(X)$, $\mathbb{E}(Y_m)\uparrow\mathbb{E}(Y)$. **事实上, 对任一 $m\in\mathbb{N}$, X_m 与 Y_m 独立**. 理由如下: 由本章第二节的构造,

$$X_m=\frac{n}{2^m}\iff\frac{n}{2^m}\leqslant X<\frac{n+1}{2^m}\iff n\leqslant 2^mX<n+1\iff[2^mX]=n.$$ [4]

于是 $X_m=\dfrac{[2^mX]}{2^m}$ 是随机变量 X 的函数 (Borel 可测), 同理, $Y_m=\dfrac{[2^mY]}{2^m}$ 也是随机变量 Y 的函数. 由定理 28 便可知 X_m 与 Y_m 独立, 于是 $\mathbb{E}(X_mY_m)=\mathbb{E}(X_m)\mathbb{E}(Y_m)$.

[4] $[X]$ denotes the greatest integer in X. 取整函数

另外, $(X_mY_m)_{m\geqslant 1}$ 是关于 m 的单调不减随机变量序列:

$$\begin{aligned}X_{m+1}Y_{m+1}-X_mY_m &= (X_{m+1}Y_{m+1}-X_{m+1}Y_m)+(X_{m+1}Y_m-X_mY_m)\\ &= X_{m+1}(Y_{m+1}-Y_m)+(X_{m+1}-X_m)Y_m \geqslant 0.\end{aligned}$$

并且 $0\leqslant XY-X_mY_m = X(Y-Y_m)+Y_m(X-X_m)\to 0$, 于是由单调收敛定理可知

$$\begin{aligned}\mathbb{E}(XY) &= \lim_{m\to\infty}\mathbb{E}(X_mY_m) = \lim_{m\to\infty}\mathbb{E}(X_m)\mathbb{E}(Y_m)\\ &= \lim_{m\to\infty}\mathbb{E}(X_m)\lim_{m\to\infty}\mathbb{E}(Y_m) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

其中上式第三个等号成立是因为 X 与 Y 的数学期望存在.

Step 3. 设 X 与 Y 为任一随机变量. 由 X 与 Y 的独立性可知 X^+, X^-; Y^+, Y^- 之间独立[5]. 则

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}\{(X^+-X^-)(Y^+-Y^-)\} = \mathbb{E}(X^+Y^+-X^+Y^--X^-Y^++X^-Y^-)\\ &= \mathbb{E}(X^+)\mathbb{E}(Y^+)-\mathbb{E}(X^+)\mathbb{E}(Y^-)-\mathbb{E}(X^-)\mathbb{E}(Y^+)+\mathbb{E}(X^-)\mathbb{E}(Y^-)\\ &= \left[\mathbb{E}(X^-)-\mathbb{E}(X^-)\right]\left[\mathbb{E}(Y^+)-\mathbb{E}(Y^-)\right] = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

其中上式用到了期望的线性与这些随机变量之间的独立性. 定理得证. ■

[5] X 的任一 Borel 可测函数与 Y 的任一 Borel 可测函数之间是独立的, 这个读者应该能够意识到吧.

Expectation of the Product of Independent R.V.'s

Second Proof. 考虑随机向量 (X,Y) 和由其导出的概率测度 $\mu^2(dx,dy)$. 由定理 19, 有

$$\mathbb{E}(XY) = \int_\Omega XY\,d\mathbb{P} = \iint_{\mathbb{R}^2} xy\,\mu^2(dx,dy).$$

又因为 X 与 Y 独立, 所以由 (26) 式便得

$$\int_{\mathbb{R}}\int_{\mathbb{R}} xy\,\mu_1(dx)\mu_2(dy) = \int_{\mathbb{R}} x\,\mu_1(dx)\int_{\mathbb{R}} y\,\mu_2(dy) = \mathbb{E}(X)\mathbb{E}(Y).$$

这就证明了 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, 定理得证[6]. ■

[6] 证法 2 中使用了乘积测度与 Fubini 定理的结论. 可参考 MATH4123.

Existence of Independent Random Variables

- Do independent random variables exist?
- We discussed an example in the classical probability model where a die is thrown consistently, so that the number of points on the die after each throw constitutes a random vector.
- It is easy to conceive of the idea that the number of points appearing on the previous die does *not affect* the number of points appearing on the next; indeed, we need imaginative rules if we want the result of the second throw to be affected by the result of the first.
- In this circumstance, idealized, the trials are carried out "independently of one another" and the corresponding r.v.'s are "independent" according to definition.
- *Our aim in this section is to construct independent random variables rigorously.*
- As for the construction of the general product measure, we have already discussed it in Chapter 6 of MATH4123, and the reader is encouraged to review it.

Independent R.V.'s on a Discrete Product Probability Space

例 31 (Independent R.V.'s on a Discrete Product Probability Space)

Let $n \geqslant 2$ and $(\Omega_j, \mathscr{C}_j, \mathbb{P}_j)$ be n discrete probability spaces, where $\mathscr{C}_j$ is the *total* σ-algebra of Ω_j. We define the *product space*

$$\Omega^n = \Omega_1 \times \cdots \times \Omega_n \quad (n \text{ factors}) \tag{29}$$

to be the space of all *ordered* n-tuples $\omega^n = (\omega_1, \cdots, \omega_n)$, where each $\omega_j \in \Omega_j$. The *product σ-algebra* $\mathscr{C}^n$ is simply the collection of all subsets of Ω^n (the total σ-algebra of Ω^n). Since Ω^n is a countable set, we may define a p.m. $\mathbb{P}^n$ on $\mathscr{C}^n$ by the following assignment:

$$\mathbb{P}^n(\{\omega^n\}) = \prod_{j=1}^{n} \mathbb{P}_j(\{\omega_j\}), \tag{30}$$

this p.m. will be called the *product measure* derived from the p.m.'s $(P_j)_{j=1}^n$ and denoted by $\prod_{j=1}^n \mathbb{P}_j$. It is trivial to verify that this is a p.m. Furthermore, it has the following product property, extending its definition (30): if $S_j \in \mathscr{C}_j$, $1 \leqslant j \leqslant n$, then

$$\mathbb{P}^n\Big(\prod_{j=1}^{n} S_j\Big) = \prod_{j=1}^{n} \mathbb{P}_j(S_j). \tag{31}$$

Independent R.V.'s on a Discrete Product Probability Space

例 31 续 (Independent R.V.'s on a Discrete Product Probability Space)

To see (31), we observe that the left side is, by definition, equal to

$$\begin{aligned}\sum_{\omega_1 \in S_1} \cdots \sum_{\omega_n \in S_n} \mathbb{P}^n(\{\omega_1, \cdots, \omega_n\}) &= \sum_{\omega_1 \in S_1} \cdots \sum_{\omega_n \in S_n} \prod_{j=1}^n \mathbb{P}_j(S_j) \\ &= \prod_{j=1}^n \left\{ \sum_{\omega_j \in S_j} \mathbb{P}(\{\omega_j\}) \right\} = \prod_{j=1}^n \mathbb{P}_j(S_j).\end{aligned}$$

Now let X_j be an r.v. on Ω_j; B_j be an arbitrary Borel set; and $S_j = X_j^{-1}(B_j)$ so that $S_j \in \mathscr{C}_j$. We have then by (30):

$$\mathbb{P}^n\left(\prod_{j=1}^n \{X_j \in B_j\}\right) = \mathbb{P}^n\left(\prod_{j=1}^n S_j\right) = \prod_{j=1}^n \mathbb{P}_j(S_j) = \prod_{j=1}^n \mathbb{P}_j\{X_j \in B_j\}. \tag{32}$$

To each function X_j on Ω_j let the correspond function $\widetilde{X}_j$ on Ω^n defined below, in which $\omega = (\omega_1, \cdots, \omega_n)$ and each "coordinate" ω_j is regarded as a *function* of the point of ω:

$$\forall\ \omega \in \Omega^n: \quad \widetilde{X}_j(\omega) = X_j(\omega_j).$$

Independent R.V.'s on a Discrete Product Probability Space

例 31 续 (Independent R.V.'s on a Discrete Product Probability Space)

Then we have

$$\bigcap_{j=1}^{n}\{\omega : \widetilde{X}_j(\omega) \in B_j\} = \prod_{j=1}^{n}\{\omega_j : X_j(\omega_j) \in B_j\}$$

since

$$\{\omega : \widetilde{X}_j(\omega) \in B_j\} = \Omega_1 \times \cdots \times \Omega_{j-1} \times \{\omega_j : X_j(\omega_j) \in B_j\} \times \Omega_{j+1} \times \cdots \times \Omega_n.$$

It follows from (32) that

$$\mathbb{P}^n\Big(\bigcap_{j=1}^{n}\{\widetilde{X}_j \in B_j\}\Big) = \prod_{j=1}^{n}\mathbb{P}^n\{\widetilde{X}_j \in B_j\}.$$

Therefore the r.v.'s $(\widetilde{X}_j)_{j=1}^{n}$ are independent.

Independent R.V.'s on a n-dimensional Cube

例 32 (Independent R.V.'s on a n-dimensional Cube)

Let $\mathscr{U}^n$ be the n-dimensional cube:

$$\mathscr{U}^n = \{(x_1, \cdots, x_n) : 0 \leqslant x_j \leqslant 1; 1 \leqslant j \leqslant n\}.$$

The trace on $\mathscr{U}^n$ of $(\mathbb{R}^n, \mathscr{B}^n, m^n)$, where $\mathbb{R}^n$ is the n-dimensional Euclidean space, $\mathscr{B}^n$ and m^n the usual Borel σ-algebra and Lebesgue measure, is a probability space. The p.m. m^n on $\mathscr{U}^n$ is a product measure and having the property analogous to (31). Let $(f_j)_{j=1}^n$ be n Borel measurable functions of one variable, and

$$X_j((x_1, \cdots, x_n)) = f_j(x_j).$$

Then $(X_j)_{j=1}^n$ are independent r.v.'s.[7] In particular if $f_j(x_j) \equiv x_j$, we obtain the n *coordinate variables* in the cube. The reader may recall the term "independent variables" used in calculus, particular for integration in several variables.

[7] 独立性的证明可参考例 31 中 $\tilde{X}_j$ 独立性的证明.

Independent R.V.'s on $(\mathscr{U}, \mathscr{B}, m)$

- Can we construct r.v.'s on the probability space $(\mathscr{U}, \mathscr{B}, m)$ itself, without going to a product space?
- Indeed we can, but only by *imbedding* (嵌入) a product structure in $\mathscr{U}$.

例 33 (Independent R.V.'s on $(\mathscr{U}, \mathscr{B}, m)$)

For each real number in $(0,1]$, consider its binary digital expansion (二进制展开)

$$x = .\varepsilon_1\varepsilon_2\cdots\varepsilon_n\cdots_2 = \sum_{n=1}^{\infty}\frac{\varepsilon_n}{2^n}, \quad \text{each } \varepsilon_n = 0 \text{ or } 1. \tag{33}$$

This expansion is sometimes not unique. For the sake of definiteness, let us agree that only expansions with infinitely many digits "1" are used. Now each digit ε_j of x is a function of x taking the values 0 and 1 on two Borel sets. Hence they are r.v.'s.

Let $(c_j)_{j\geqslant 1}$ be a given sequence of 0's and 1's. Then the set

$$\{x : \varepsilon_j(x) = c_j, 1 \leqslant j \leqslant n\} = \bigcap_{j=1}^{n}\{x : \varepsilon_j(x) = c_j\}$$

is the set of numbers x whose first n digits are the given c_j's.

例 33 续 (Independent R.V.'s on $(\mathscr{U}, \mathscr{B}, m)$)

Thus

$$x = .c_1 c_2 \cdots c_n \varepsilon_{n+1} \varepsilon_{n+2} \cdots$$

with the digits from the $(n+1)$st on completely arbitrary. It is clear that this set is just an interval of length $1/2^n$, hence of probability $1/2^n$. On the other hand for each j, the set $\{x : \varepsilon_j(x) = c_j\}$ has probability $1/2$ for a similar reason. We have therefore

$$\mathbb{P}\{\varepsilon_j = c_j, 1 \leqslant j \leqslant n\} = \frac{1}{2^n} = \prod_{j=1}^{n} \left(\frac{1}{2}\right) = \prod_{j=1}^{n} \mathbb{P}\{\varepsilon_j = c_j\}.$$

This being true for every choice of the c_j's, the r.v.'s $(\varepsilon_j)_{j \geqslant 1}$ are independent.

Let $(f_j)_{j \geqslant 1}$ be arbitrary functions with domain the two points $\{0, 1\}$, then $(f_j(\varepsilon_j))_{j \geqslant 1}$ are also independent r.v.'s.

The Fundamental Existence Theorem of Product Measures

- We are ready to prove the fundamental existence theorem of product measures.
- In fact, this is the one-dimensional case of the *Kolmogorov's extension theorem*.

定理 34 (The Fundamental Existence Theorem of Product Measures)

Let a finite or infinite sequence of p.m.'s (μ_j) on $(\mathbb{R},\mathscr{B}^1)$, or equivalently their d.f.'s, be given. There exists a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and a sequence of independent r.v.'s (X_j) defined on it such that for each j, μ_j is the p.m. of X_j.

PROOF. 不失一般性, 我们只需讨论 (μ_j) 是无限序列的情形. 现在, 对任意的 $n \in \mathbb{N}$, 存在概率空间 $(\Omega_n,\mathscr{F}_n,\mathbb{P}_n)$ 上的随机变量 X_n, s.t. μ_n 是 X_n 对应的概率测度 (即满足关系式 (3)). 上述命题成立, 如果我们令 $(\Omega_n,\mathscr{F}_n,\mathbb{P}_n) = (\mathbb{R},\mathscr{B}^1,\mu_n)$, 并且令随机变量 X_n 是样本点 $x \in \mathbb{R}$ 的恒同映射. 现在我们将样本点 $x \in \mathbb{R}$ 写作 $\omega_n \in \Omega_n$.

Step 1. 首先定义无穷乘积空间及其上代数

$$\Omega = \prod_{n=1}^{\infty} \Omega_n = \{(\omega_1,\omega_2,\cdots,\omega_n,\cdots) : \omega_n \in \Omega_n, \forall\ n \in \mathbb{N}\}.$$

其中 Ω 的子集 E 称作**有限乘积集** (finite-product set) 当且仅当 E 有如下表示:

$$E = \prod_{n=1}^{\infty} F_n, \tag{34}$$

其中 $F_n \in \mathscr{F}_n$, 并且只有有限个 $F_n \neq \Omega_n$. 因此 $\omega \in E$ 当且仅当 $\omega_n \in F_n$ $(\forall\ n \in \mathbb{N})$, 但事实上, 这一限制的个数是有限的. 再定义

$$\mathscr{F}_0 = \left\{ \bigcup_{j=1}^{k} E_j : E_j \text{ 是互不相交的有限乘积集}, \forall\ k \in \mathbb{N} \right\}.$$

容易证明 $\mathscr{F}_0$ 中的元素关于补集和交集运算封闭, 于是 $\mathscr{F}_0$ 是一个代数. 现在, 我们可取定理中的 $\mathscr{F}$ 为由代数 $\mathscr{F}_0$ 生成的 σ-代数, 将其称作 $(\mathscr{F}_n)$ **的乘积 σ-代数** (the product σ-algebra of the sequence $(\mathscr{F}_n)$).

Step 2. 再定义 $(\Omega, \mathscr{F}_0)$ 上的概率测度 $\mathbb{P}$. 首先, 对于任一形如 (34) 式的有限乘积集 E, 定义

$$\mathbb{P}(E) = \prod_{n=1}^{\infty} \mathbb{P}_n(F_n), \tag{35}$$

当然上式右端只有有限项不为 1. 接下来, 如果 $E \in \mathscr{F}_0$ 并且 $E = \bigcup_{k=1}^{n} E^{(k)}$, 其中 $E^{(k)}$ 为互不相交的有限乘积集, 定义

$$\mathbb{P}(E) = \sum_{k=1}^{n} \mathbb{P}(E^{(k)}). \tag{36}$$

事实上, 若 $E \in \mathscr{F}_0$ 有两种表示, 不难验证 $\mathbb{P}$ 是良定义的 (验证挺冗长的).

现在, 集合函数 $\mathbb{P}$ 被唯一定义在 $\mathscr{F}_0$ 上, 满足

$$\mathbb{P}(\Omega)=\prod_{n=1}^{\infty}\mathbb{P}_n(\Omega_n)=1,$$

和有限可加性: 若 $E_j\in\mathscr{F}_0$ $(1\leqslant j\leqslant m)$ 互不相交, 于是

$$E_i\cap E_j=\left(\bigcup_{k=1}^{m_i}E_i^{(k)}\right)\cap\left(\bigcup_{l=1}^{m_j}E_j^{(l)}\right)=\bigcup_{k=1}^{m_i}\bigcup_{l=1}^{m_j}(E_i^{(k)}\cap E_j^{(l)})=\varnothing.$$

这说明每个小的有限乘积集均互不相交[8], 于是可以利用定义 (36) 式:

$$\mathbb{P}\left(\bigcup_{j=1}^{m}E_j\right)=\mathbb{P}\left(\bigcup_{j=1}^{m}\bigcup_{k=1}^{n_j}E_j^{(k)}\right)=\sum_{j=1}^{m}\sum_{k=1}^{n_j}\mathbb{P}(E_j^{(k)})=\sum_{j=1}^{m}\mathbb{P}(E_j).$$

上式中第一个等式中将每个 $E_j\in\mathscr{F}_0$ 展开为有限个互不相交有限乘积集的并集, 注意到上式第二部分中的并集实际上还是有限个有限乘积集的无交并, 于是可以利用 (36) 式进行展开. 最后, 第三个等式中绿色式再次利用了 (36) 式, 从而证明了 $\mathbb{P}$ 是有限可加的.

[8]这里读者应该不难看出每个 E_i 的表示, 就不再正文中再次赘述了.

现在已知集函数 $\mathbb{P}$ 满足 $\mathbb{P}(\Omega)=1$ 并且是有限可加的. 为了证明 $\mathbb{P}$ 是可列可加的, 只需证明 $\mathbb{P}$ 满足 "连续性公理[9]". 为此, 我们去证明 $\mathbb{P}$ 满足连续性公理的逆否命题, 即: 若存在 $\delta>0$ 与一个单调递减集合列 $(C_n)_{n\geqslant 1}\subseteq\mathscr{F}_0$ 满足 $\mathbb{P}(C_n)\geqslant\delta>0$, 则有 $\bigcap\limits_{n=1}^{\infty}C_n\neq\varnothing$.

首先注意到 $\mathscr{F}_0$ 中任一集合 E, 或任一有限乘积集均被某一个有限坐标唯一决定: $\forall\ E\in\mathscr{F}_0$, $\exists\ k\in\mathbb{N}$ 仅依赖于 E, s.t. 若 $\omega=(\omega_1,\omega_2,\cdots),\omega'=(\omega_1',\omega_2',\cdots)\in\Omega$ 且满足 $\omega_j=\omega_j'$ $(\forall\ 1\leqslant j\leqslant k)$, 则或者 $\omega,\omega'\in E$, 或者 $\omega,\omega'\notin E$.

不失一般性, 因 (C_n) 是 $\mathscr{F}_0$ 中的单调递减集合列, 不妨设 $\forall\ n\in\mathbb{N}$, C_n 被其前 n 个坐标唯一决定. 现在给定 ω_1^0, $\forall\ E\subseteq\Omega$, 定义 $(E|\omega_1^0)=\Omega_1\times E_2$, 其中

$$E_2=\Big\{(\omega_2,\omega_3,\cdots)\in\prod_{n=2}^{\infty}\Omega_n:(\omega_1^0,\omega_2,\omega_3,\cdots)\in E\Big\}.$$

如果 $E_2=\varnothing$, i.e. ω_1^0 不等于 E 中任一元素的第一个坐标, 定义 $(E|\omega_1^0)=\varnothing$.

容易看出, 若 $E\in\mathscr{F}_0$, 则 $(E|\omega_1^0)\in\mathscr{F}_0$ $(\forall\ \omega_1^0\in\Omega_1)$.

[9]可见 Chapter 2 定理 7.

断言: $\exists\ \omega_1^0 \in \Omega_1$, s.t. $\forall\ n \in \mathbb{N}$, 有 $\mathbb{P}((C_n|\omega_1^0)) \geqslant \delta/2$.

◀ 首先, 若 C_n 是有限乘积集 i.e. $C_n = \prod_{k=1}^{\infty} C_n^{(k)}$, 则

$$(C_n|\omega_1) = \begin{cases} \Omega_1 \times \prod_{k=2}^{\infty} C_n^{(k)}, & \text{if } \omega_1 \in C_n^{(1)}, \\ \emptyset, & \text{if } \omega_1 \notin C_n^{(1)}. \end{cases} \qquad (\clubsuit)$$

于是

$$\begin{aligned}\mathbb{P}(C_n) &= \prod_{k=1}^{\infty} \mathbb{P}_k(C_n^{(k)}) = \mathbb{P}_1(C_n^{(1)}) \prod_{k=2}^{\infty} \mathbb{P}_k(C_n^{(k)}) = \mathbb{P}_1(C_n^{(1)}) \mathbb{P}_1(\Omega_1) \prod_{k=2}^{\infty} \mathbb{P}_k(C_n^{(k)}) \\ &= \mathbb{P}\Big(\Omega_1 \times \prod_{k=2}^{\infty} C_n^{(k)}\Big) \int_{C_n^{(1)}} 1\, \mathbb{P}_1(d\omega_1) = \int_{C_n^{(1)}} \mathbb{P}\Big(\Omega_1 \times \prod_{k=2}^{\infty} C_n^{(k)}\Big) \mathbb{P}_1(d\omega_1) \\ &= \int_{\Omega_1} \mathbb{P}((C_n|\omega_1)) \mathbb{P}_1(d\omega_1). \quad [\text{利用 } (\clubsuit) \text{ 式}]\end{aligned}$$

一般地, 若 $C_n \in \mathscr{F}_0$ 是有限个互不相交有限乘积集的并, 由 $\mathbb{P}$ 与积分的有限可加性, 也有

$$\mathbb{P}(C_n) = \int_{\Omega_1} \mathbb{P}((C_n|\omega_1)) \mathbb{P}_1(d\omega_1). \tag{37}$$

现在令 $B_n=\{\omega_1:\mathbb{P}((C_n|\omega_1))\geqslant\delta/2\}\subseteq\Omega_1$，则

$$\begin{aligned}\delta\leqslant\mathbb{P}(C_n)&=\int_{\Omega_1}\mathbb{P}((C_n|\omega_1))\mathbb{P}_1(d\omega_1)\\&=\int_{B_n}\mathbb{P}((C_n|\omega_1))\mathbb{P}_1(d\omega_1)+\int_{B_n^c}\mathbb{P}((C_n|\omega_1))\mathbb{P}_1(d\omega_1)\\&\leqslant\int_{B_n}1\,\mathbb{P}_1(d\omega_1)+\int_{B_n^c}\frac{\delta}{2}\mathbb{P}_1(d\omega_1)=\mathbb{P}_1(B_n)+\frac{\delta}{2}(1-\mathbb{P}_1(B_n)).\end{aligned}$$

于是就有 $\mathbb{P}_1(B_n)\geqslant\delta-\frac{\delta}{2}+\frac{\delta}{2}\mathbb{P}_1(B_n)\geqslant\frac{\delta}{2}$ $(\forall\ n\in\mathbb{N})$. 因为 C_n 是单调递减集合列，于是 B_n 也是单调递减集合列，由概率测度的连续性，便有 $\mathbb{P}\left(\bigcap_{n=1}^{\infty}B_n\right)\geqslant\frac{\delta}{2}$，从而 $\bigcap_{n=1}^{\infty}B_n\neq\emptyset$. 现在，任取 $\omega_1^0\in\bigcap_{n=1}^{\infty}B_n$，由 B_n 的定义可知 $\mathbb{P}((C_n|\omega_1^0))\geqslant\frac{\delta}{2}$ $(\forall\ n\in\mathbb{N})$. ▶

对 $(C_n|\omega_1^0)\in\mathscr{F}_0$ 重复上述操作，则 $\exists\ \omega_2^0\in\Omega_2$, s.t. $\forall\ n\in\mathbb{N}$，有 $\mathbb{P}((C_n|\omega_1^0,\omega_2^0))\geqslant\delta/4$. 其中 $(C_n|\omega_1^0,\omega_2^0)=((C_n|\omega_1^0)|\omega_2^0)=\Omega_1\times\Omega_2\times E_3$，其中

$$E_3=\left\{(\omega_3,\omega_4,\cdots)\in\prod_{n=3}^{\infty}\Omega_n:(\omega_1^0,\omega_2^0,\omega_3,\omega_4,\cdots)\in C_n\right\}.$$

一般地, 由数学归纳法, $\forall\ k \geqslant 1$, $\exists\ \omega_k^0 \in \Omega_k$, s.t.

$$\forall\ n \in \mathbb{N}:\ \mathbb{P}((C_n|\omega_1^0,\cdots,\omega_k^0)) \geqslant \frac{\delta}{2^k}.$$

考虑 $\omega^0 = (\omega_1^0, \omega_2^0, \cdots, \omega_k^0, \cdots)$, 因 $(C_k|\omega_1^0,\cdots,\omega_k^0) \neq \varnothing$, 故 C_k 中有一点满足该点的前 k 个坐标恰为 ω^0 的前 k 个坐标. 而 C_k 由其前 k 个坐标决定, 于是 $\omega^0 \in C_k$ $(\forall\ k \in \mathbb{N})$. 这就说明 $\omega^0 \in \bigcap\limits_{n=1}^{\infty} C_k$. 于是我们最终证明了 $\mathbb{P}$ 满足连续性公理的逆否命题, 从而 $\mathbb{P}$ 是 $(\Omega, \mathscr{F}_0)$ 上的概率测度.

Step 3. 将 $(\Omega, \mathscr{F}_0)$ 上的概率测度 $\mathbb{P}$ 唯一的扩张到 $(\Omega, \sigma(\mathscr{F}_0))$ 上去, 其中 $\sigma(\mathscr{F}_0)$ 是由代数 $\mathscr{F}_0$ 生成的 σ-代数. 而这可直接由测度论中的 Carathéodory 扩张定理得到, 扩张的唯一性成立, 因为 $\mathbb{P}$ 是有限测度[10].

Step 4. 最后证明由 (μ_j) 导出的随机变量 X_j 之间的独立性. 由本定理最一开始的叙述, 这些随机变量可取作样本点 ω_n 的恒同映射, 于是随机变量之间的完全独立性可由 (35) 式直接导出. ■

REMARK. 乘积测度并不一定是完备测度, 但是我们可以对乘积测度进行完备化. 详细内容可参考 MATH4123 第三章最后一节.

[10] Carathéodory 扩张定理可参考 MATH4123 第三章定理 40, 唯一性可参考定理 42.

Kolmogorov's Extension Theorem: General Case

- We close this section by stating a generalization of Theorem 34 to the case where the finite-dimensional joint distribution of the sequence (X_j) are arbitrary given, subject only to *mutual consistency* (相互一致性).
- This result is a particular case of *Kolmogorov's extension theorem*, which is valid for an arbitrary family of r.v.'s. ▸ Back to Content.
- Let m and n be integers: $1 \leqslant m \leqslant n$, and define π_{mn} to be the "projection map" of $\mathscr{B}^m$ onto $\mathscr{B}^n$ given by

$$\forall \ B \in \mathscr{B}^m: \ \pi_{mn}(B) = \{(x_1, \cdots, x_n) : (x_1, \cdots, x_m) \in B\}.$$

定理 35 (Kolmogorov's Extension Theorem: General Case)

For each $n \geqslant 1$, let μ^n be a p.m. on $(\mathbb{R}^n, \mathscr{B}^n)$ such that

$$\forall \ m < n: \quad \mu^n \circ \pi_{mn} = \mu^m. \quad \textit{consistency condition} \tag{38}$$

Then there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sequence of r.v.'s (X_j) on it such that for each n, μ^n is the n-dimensional p.m. of the vector $(X_1, \cdots, X_n)$.

- We'll come back to this theorem at the beginning of the stochastic process.

参考文献

- Chung, Kai Lai. *A Course in Probability Theory*, Third Edition. Academic Press, 2001.
- Durrett, Rick. *Probability: Theory and Examples*, 5th edition. Cambridge university press, 2019.
- Billingsley, Patrick. *Probability and Measure*, Anniversary edition. John Wiley & Sons, Inc. 2012.
- Varadhan, SR Srinivasa. *Probability theory*. No. 7. American Mathematical Soc., 2001.
- Williams, David. *Probability with martingales*. Cambridge university press, 1991.
- 任佳刚 巫静. 测度与概率教程. 科学出版社, 2018.
- 严士健 刘秀芳. 测度与概率, 第二版. 北京师范大学出版社, 2023.

FINANCIAL TECHNOLOGY

The Thrust nurtures creative minds to explore frontiers in Financial Technology, to shape the future financial industry that promotes social-financial health.

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

社会枢纽
SOCIETY HUB

Society Hub • Financial Technology Thrust

Chapter 4 Convergence Concepts

Probability Theory and Stochastic Processes

WANG Zhuoran

16320216@stu.njau.edu.cn

香港科技大學 (廣州)

FTEC5031

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Content

Various Modes of Convergence

- Convergence Almost Everywhere
- Convergence in Probability and in L^p

Borel-Cantelli Lemma

Vague Convergence

Continuation; Convergence in Distribution

- Continuation
- Convergence in Distribution

Uniform Integrability; Convergence of Moments

Section 1

Various Modes of Convergence

- This section focuses on some convergence modes and their properties of r.v.'s.
- Including: convergence "almost everywhere"; convergence "in probability" and convergence in L^p.
- Finally, we will recall the definition of weak convergence which we have known in functional analysis.

Definition of Convergence "Almost Everywhere"

- As numerical-valued functions, the convergence of a sequence of r.v.'s $\{X_n, n \geqslant 1\}$, to be denoted simply by (X_n) below, is a well-defined concept.
- The term "convergence" will be used to mean convergence to a *finite* limit.
- Thus it make sense to say: for every $\omega \in \Delta$, where $\Delta \in \mathscr{F}$, the sequence $(X_n(\omega))$ converges. The limit is then a finite-valued r.v. (see Theorem 3.12), say $X(\omega)$, defined on Δ. If $\Omega = \Delta$, then we have **convergence everywhere**, but a more useful concept is the following one.

定义 1 (Definition of Convergence "Almost Everywhere" (a.e.))

The sequence of r.v. (X_n) is said to converge **almost everywhere** (also **with probability** 1 (w.p.1.)) [to the r.v. X] iff there exists a null set N such that

$$\forall\ \omega \in \Omega \backslash N: \quad \lim_{n \to \infty} X_n(\omega) = X(\omega) \text{ finite.} \tag{1}$$

This will be denoted by $X_n \to X$ a.e.

- Recall that our convention stated at the beginning of Chapter 3.1. allows each r.v. a null set on which it may be $\pm\infty$. The union of all these sets being still a null set, it may be included in the set N in (1) without modifying the conclusion. $\Rightarrow$ *When dealing with a countable set of r.v.'s that are finite a.e., to regard them as finite everywhere*.

Equivalent Definition of Convergence "Almost Everywhere"

定理 2 (Equivalent Definition of Convergence "Almost Everywhere")

The sequence (X_n) *converges a.e. to* X *iff for every* $\varepsilon > 0$ *we have*

$$\lim_{m\to\infty} \mathbb{P}\{|X_n - X| \leqslant \varepsilon \text{ for } \underline{\text{all}}\ n \geqslant m\} = 1; \tag{2}$$

or equivalently

$$\lim_{m\to\infty} \mathbb{P}\{|X_n - X| > \varepsilon \text{ for } \underline{\text{some}}\ n \geqslant m\} = 0. \tag{3}$$

Proof. $(\Rightarrow)$. 设 (X_n) 在 Ω 上几乎处处收敛于 X, 更具体地, 存在零测度集 N, 而 X_n 在 $\Omega\backslash N$ 上点点收敛于 X. 现任取 $m \geqslant 1$, 令

$$A_m(\varepsilon) = \{|X_n - X| \leqslant \varepsilon \text{ for } \underline{\text{all}}\ n \geqslant m\} = \bigcap_{n=m}^{\infty} \{|X_n - X| \leqslant \varepsilon\},$$

则 $(A_m(\varepsilon))_{m\geqslant 1}$ 是单调递增集合列. 任取 $\omega_0 \in \Omega_0$, 由数列 $X_n(\omega_0) \to X(\omega_0)$ 可知对于给定的 $\varepsilon > 0$, $\exists\ m(\omega_0, \varepsilon)$, $\forall\ n \geqslant m(\omega_0, \varepsilon)$, 有 $|X_n(\omega_0) - X(\omega_0)| \leqslant \varepsilon$. 再由 $A_m(\varepsilon)$ 的定义可知 $\omega_0 \in A_{m(\omega_0,\varepsilon)}(\varepsilon)$, 由 $\omega_0 \in \Omega_0$ 的任意性可知 $\Omega_0 \subseteq \bigcup_{m=1}^{\infty} A_m(\varepsilon)$.

Equivalent Definition of Convergence "Almost Everywhere"

于是由测度的单调性可知

$$1=\mathbb{P}(\Omega_0)\leqslant \mathbb{P}\Big(\bigcup_{m=1}^{\infty}A_m(\varepsilon)\Big)=\lim_{m\to\infty}\mathbb{P}(A_m(\varepsilon))\leqslant 1,$$

于是 $\lim\limits_{m\to\infty}\mathbb{P}(A_m(\varepsilon))=1$, 这就证明了 (2) 式.

($\Leftarrow$). 若 (2) 式成立, 则 $A(\varepsilon)=\bigcup\limits_{m=1}^{\infty}A_m(\varepsilon)$ 满足 $\mathbb{P}(A(\varepsilon))=\lim\limits_{m\to\infty}\mathbb{P}(A_m(\varepsilon))=1$. 现在任取 $\omega_0\in A(\varepsilon)$, 则 $\exists\ m_0\geqslant 1$, s.t. $\omega_0\in A_{m_0}(\varepsilon)$. 于是 $\forall\ n\geqslant m_0$ 有 $|X_n(\omega_0)-X(\omega_0)|\leqslant\varepsilon$. 现在特取 $A=\bigcap\limits_{k=1}^{\infty}A(1/k)$, 注意到 $\big(A(1/k)\big)_{k\geqslant 1}$ 是单调递减集合列, 于是

$$\mathbb{P}(A)=\lim_{k\to\infty}\mathbb{P}(A(1/k))=1.$$

现在, $\forall\ \omega_0\in A$, 则 $\forall\ \varepsilon>0$, $\exists\ k_0\in\mathbb{N}$, s.t. $1/k_0<\varepsilon$. 对于这一 k_0, 因 $\omega_0\in A(1/k_0)$, 则 $\exists\ m(k_0)\geqslant 1$, s.t. $\forall\ n\geqslant m(k_0)$, 有 $|X_n(\omega_0)-X(\omega_0)|\leqslant 1/k_0<\varepsilon$. 这就说明 X_n 在 A 上点点收敛于 X. 又因为 $\mathbb{P}(A)=1$, 于是 $X_n\to X$ a.e. ∎

Definition of Convergence "in Probability"

- A *weaker* concept of convergence is of basic importance in probability theory.

定义 3 (Definition of Convergence "in Probability" (in pr.))

The sequence (X_n) is said to converge **in probability** to X iff for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \varepsilon\} = 0. \tag{4}$$

This will be denoted by $X_n \xrightarrow{\mathbb{P}} X$.

- 需要注意的是, (4) 式只在 X_n 与 X 几乎处处有限时才有意义, 此时我们可以忽略掉 $\infty - \infty$ 这些无法定义的点 (零测度集).

定理 4 (Convergence a.e. $\Rightarrow$ Convergence in pr.)

Convergence a.e. [to X] implies convergence in pr. [to X].

Proof. 若 $X_n \to X$ a.e., 则 (3) 式成立, i.e. $\lim\limits_{m\to\infty} \mathbb{P}\left(\bigcup\limits_{n=m}^{\infty} \{|X_n - X| > \varepsilon\}\right) = 0$, 于是

$$0 \leqslant \mathbb{P}\{|X_n - X| > \varepsilon\} \leqslant \mathbb{P}\left(\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) \to 0 \ (n \to \infty).$$

这就说明 $X_n \xrightarrow{\mathbb{P}} X$. ■

Definition of Convergence "in L^p" $(0 < p < \infty)$

定义 5 (Definition of Convergence "in L^p" $(0 < p < \infty)$)

The sequence (X_n) is said to converge **in** L^p to X iff $X_n \in L^p$, $X \in L^p$ and

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = \lim_{n\to\infty} \int_\Omega |X_n - X|^p \, d\mathbb{P} = 0. \tag{5}$$

This will be denoted by $X_n \xrightarrow{p} X$.

- In all these definitions above, *X_n converges to X iff $X_n - X$ converges to 0*. Hence there is no loss of generality if we put $X \equiv 0$ in the discussion, provided that any hypothesis involved can be similarly reduced to this case.
- We say that X is **dominated** by Y if $|X| \leqslant Y$ a.e., and that the sequence (X_n) is dominated by Y iff this is true for each X_n with the same Y.
- We say that X or (X_n) is **uniformly dominated** iff the Y above may be taken to be a constant.

定理 6 (Relations Between Convergence to 0 in L^p and Convergence to 0 in pr.)

If X_n converges to 0 in L^p, then it converges to 0 in pr. The converse is true provided that (X_n) is dominated by some Y that belongs to L^p.

Proof of Theorem 6

Remark

If $X_n \xrightarrow{p} X$, and (X_n) is dominated by Y, then $(X_n - X)$ is dominated by $Y + |X|$, which is in L^p. Hence there is no loss of generality to assume $X \equiv 0$.

Proof. (⇒). 由 Chebyshev 不等式, 令 $\varphi(x) \equiv |x|^p$, 则

$$\mathbb{P}(|X_n| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(|X_n|^p)}{\varepsilon^p}.$$

因 $X_n \xrightarrow{p} 0$, 于是 $\mathbb{E}(|X_n - 0|^p) \to 0$, 于是 $\mathbb{P}(|X_n - 0| \geqslant \varepsilon) \to 0$ $(n \to \infty)$, i.e. $X_n \xrightarrow{\mathbb{P}} 0$.

(⇐). 若 $|X_n| \leqslant Y$ a.e., 其中 $Y \in L^p$, 于是 $\mathbb{E}(Y^p) < \infty$, 则

$$\mathbb{E}(|X_n|^p) = \int_{\{|X_n| < \varepsilon\}} |X_n|^p \, d\mathbb{P} + \int_{\{|X_n| \geqslant \varepsilon\}} |X_n|^p \, d\mathbb{P} \leqslant \varepsilon^p + \int_{\{|X_n| \geqslant \varepsilon\}} Y^p \, d\mathbb{P}.$$

又因为 $\mathbb{P}(|X_n| \geqslant \varepsilon) \to 0$, 则 $\int_{\{|X_n| \geqslant \varepsilon\}} Y^p \, d\mathbb{P} \leqslant \sup_{\omega \in \Omega}\{Y^p(\omega)\} \mathbb{P}(|X_n| \geqslant \varepsilon) \to 0$. 代入上式, 并令 $\varepsilon \to 0$ 便有 $\mathbb{E}(|X_n|^p) \to 0$ $(n \to \infty)$, 于是 $X_n \xrightarrow{p} 0$. ■

- As a corollary, *for a uniformly bounded sequence (X_n) convergence in pr. and in L^p are equivalent*.
- The general result is as follows.

定理 7 (Equivalent Condition of Convergence in pr.)

$X_n \xrightarrow{\mathbb{P}} 0$ if and only if

$$\mathbb{E}\Big(\frac{|X_n|}{1+|X_n|}\Big) \to 0. \tag{6}$$

Furthermore, the functional $\rho(\cdot,\cdot)$ given by

$$\rho(X,Y) = \mathbb{E}\Big(\frac{|X-Y|}{1+|X-Y|}\Big)$$

is a metric in the space of r.v.'s, provided that we identify that are equal a.e.

PROOF. 首先注意到 $\rho(X,Y)=0 \;\Leftrightarrow\; \mathbb{E}(|X-Y|)=0 \;\Leftrightarrow\; X=Y$ a.e. (积分的性质), 于是, 我们将两个几乎处处相等的随机变量看为同一[1].

[1]类似的处理我们在实分析中已经讨论过了.

Equivalent Condition of Convergence in pr.

现在为了证明 $\rho(\cdot,\cdot)$ 是距离，只需证明其满足三角不等式：$\rho(X,Y) \leqslant \rho(X,Z)+\rho(Y,Z)$. 因

$$\frac{|X-Y|}{1+|X-Y|} \leqslant \frac{|X-Z|+|Y-Z|}{1+|X-Z|+|Y-Z|} \leqslant \frac{|X-Z|}{1+|X-Z|}+\frac{|Y-Z|}{1+|Y-Z|},$$

其中第一个不等式成立是因为 $f(t)=\frac{t}{1+t}$ 是单调递增函数. 再利用积分的性质便有

$$\mathbb{E}\left(\frac{|X-Y|}{1+|X-Y|}\right) \leqslant \mathbb{E}\left(\frac{|X-Z|}{1+|X-Z|}\right)+\mathbb{E}\left(\frac{|Y-Z|}{1+|Y-Z|}\right).$$

这就说明 $\rho(\cdot,\cdot)$ 是距离. 又因为对任一随机变量 X, $\frac{|X|}{1+|X|} \leqslant 1$, 由定理 6 可知

$$\frac{|X_n|}{1+|X_n|} \to 0 \text{ in pr.} \quad \Longleftrightarrow \quad \mathbb{E}\left(\frac{|X_n|}{1+|X_n|}\right) \to 0.$$

所以我们只需证 $|X_n| \xrightarrow{\mathbb{P}} 0 \;\Leftrightarrow\; \frac{|X_n|}{1+|X_n|} \xrightarrow{\mathbb{P}} 0$. 而这是成立的因为 $|x| \leqslant \varepsilon$ 等价于 $\frac{|x|}{1+|x|} \leqslant \frac{\varepsilon}{1+\varepsilon}$. 于是 $\mathbb{P}(|X_n| \leqslant \varepsilon) \to 1 \;\Leftrightarrow\; \mathbb{P}\left(\frac{|X_n|}{1+|X_n|} \leqslant \frac{\varepsilon}{1+\varepsilon}\right) \to 1$. ■

Convergence in pr. $\not\Rightarrow$ Convergence in L^p $\not\Rightarrow$ Convergence a.e.

例 8 (Convergence in pr. $\not\Rightarrow$ Convergence in L^p $\not\Rightarrow$ Convergence a.e.)

(1). Convergence in L^p cannot imply convergence a.e. Take the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to be $(\mathscr{U}, \mathscr{B}, m)$. Let φ_{kj} be the indicator of the interval $\left(\frac{j-1}{k}, \frac{k}{k}\right)$, $k \geqslant 1$, $1 \leqslant j \leqslant k$. Order these functions lexicographically (字典序) first according to k increasing, and then for each k according to j increasing, into one sequence (X_n) so that $X_n = \varphi_{k_n j_n}$, then $k_n \to \infty$ as $n \to \infty$. Thus for each $p > 0$:

$$\mathbb{E}(X_n^p) = \frac{1}{k_n} \to 0,$$

and so $X_n \xrightarrow{p} 0$. But for each ω and every k, there exists a j such that $\varphi_{kj}(\omega) = 1$; hence there exists infinitely many values of n such that $X_n(\omega) = 1$. Similarly there exist infinitely many values of n such that $X_n(\omega) = 0$. It follows that the sequence $(X_n(\omega))$ of 0's and 1's cannot converge for any ω, i.e. $X_n \nrightarrow 0$ a.e.

(2). Convergence in pr. cannot imply convergence L^p. Now if we replace φ_{kj} by $k^{\frac{1}{p}}\varphi_{kj}$, where $p > 0$, then $\mathbb{P}\{X_n > 0\} = 1/k_n \to 0$ so that $X_n \xrightarrow{\mathbb{P}} 0$, but for each n, we have $\mathbb{E}(X_n^p) = 1$. Consequently $\lim_{n\to\infty} \mathbb{E}(|X_n - 0|^p) = 1$ and $X_n \nrightarrow 0$ in L^p.

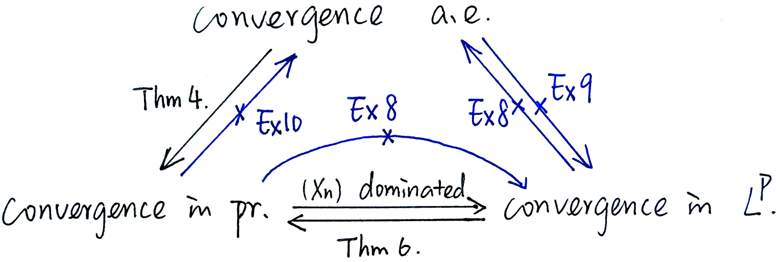
Convergence a.e. $\not\Rightarrow$ Convergence in L^p

例 9 (Convergence a.e. $\not\Rightarrow$ Convergence in L^p)

In $(\mathscr{U}, \mathscr{B}, m)$ define

$$X_n = 2^n \mathbf{1}_{(0,1/n)};$$

then $\mathbb{E}(|X_n|^p) = 2^{np}/n \to \infty$ for each $p > 0$, but $X_n \to 0$ everywhere.



例 10 (Convergence in pr. $\not\Rightarrow$ Convergence a.e.)

Let
$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}, \quad \mathbb{P}(X_n = 1) = \frac{1}{n},$$
and X_n's are independent. Then $X_n \xrightarrow{\mathbb{P}} 0$ since $\mathbb{P}(|X_n - 0| > \varepsilon) \leqslant 1/n \to 0$. However, $X_n \nrightarrow 0$ a.e. since for any $0 < \varepsilon < 1$ we have

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{m \geqslant n} \{|X_m - 0| \leqslant \varepsilon\}\Big) &= \mathbb{P}\Big(\lim_{r\to\infty} \bigcap_{m=n}^{r} \{|X_m| \leqslant \varepsilon\}\Big) = \lim_{r\to\infty} \mathbb{P}\Big(\bigcap_{m=n}^{r} \{|X_m| \leqslant \varepsilon\}\Big) \\
&= \lim_{r\to\infty} \prod_{m=n}^{r} \mathbb{P}\{|X_m| \leqslant \varepsilon\} = \lim_{r\to\infty} \prod_{m=n}^{r} \Big(1 - \frac{1}{m}\Big) \\
&= \lim_{r\to\infty} \frac{n-1}{n} \frac{n}{n+1} \cdots \frac{r-1}{r} = \lim_{r\to\infty} \frac{n-1}{r} = 0.
\end{aligned}$$

By the equivalent definition of convergence a.e., we see that $X_n \nrightarrow 0$ a.e.

Back to Content.

Section 2

Borel-Cantelli Lemma

- This section has two main tasks.
- First we recall the notion of "$\overline{\lim}$ and $\underline{\lim}$" of a sequence of sets introduced in real analysis, which leads to the notation "i.o." commonly used in probability theory;
- Next, we introduce the Borel-Cantelli lemma and its generalizations.

Limits of a Sequence of Sets

- An important concept in set theory is that of the "$\overline{\lim}$" and "$\underline{\lim}$" of a sequence of sets.

定义 11 (Limits of a Sequence of Sets)

Let (E_n) be any sequence of subsets of Ω; we define

$$\overline{\lim_{n\to\infty}} E_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n, \quad \underline{\lim}_{n\to\infty} E_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n.$$

- Let us observe at one that

$$\underline{\lim}_{n\to\infty} E_n = \Big(\overline{\lim_{n\to\infty}} E_n^c \Big)^c, \quad \textit{Use De Morgan's Law} \tag{7}$$

so that in a sense one of the two notions suffices, but it is convenient to employ both.

- Refer to Chapter 1 of MATH4123 for a detailed discussion of this concept.

Infinitely Often, i.e. "**i.o.**"

引理 12 (Equivalent Conditions of Limits of a Sequence of Sets)

- $\omega \in \overline{\lim_{n\to\infty}} E_n$ *iff for all* $n \in \mathbb{N}$, *there exists a* $k_0 \geqslant n$ *such that* $\omega \in E_{k_0}$, *i.e., there are infinite terms in* (E_n) *that contain* ω.
- $\omega \in \underline{\lim_{n\to\infty}} E_n$ *iff there exists a* $n_0 \in \mathbb{N}$ *such that for all* $k \geqslant n_0$, *we have* $\omega \in E_k$, *i.e., there are only finite terms in* (E_n) *that do not contain* ω.

- In more intuitive language: *the event* $\overline{\lim_{n\to\infty}} E_n$ *occurs if and only if the events* E_n *occur infinitely often*. Thus we may write

$$\boxed{\mathbb{P}\left(\overline{\lim_{n}} E_n\right) = \mathbb{P}(E_n \text{ i.o.})}$$

where the abbreviation "i.o." stands for "infinitely often".

Borel-Cantelli Lemma (Convergence Part)

定理 13 (Borel-Cantelli Lemma (Convergence Part))

We have for arbitrary events (E_n)*:*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \quad \Rightarrow \quad \mathbb{P}(E_n \text{ i.o.}) = 0. \tag{8}$$

PROOF. 利用概率测度的次可加性, 任取事件列 (E_n), 有

$$\mathbb{P}\Big(\bigcup_{n=m}^{\infty} E_n\Big) \leqslant \sum_{n=m}^{\infty} \mathbb{P}(E_n) \to 0, \quad (m \to \infty).$$

上式右端趋于零是因为 $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ (收敛级数的性质). 又因为 $\bigcup_{n=m}^{\infty} E_n$ 关于 m 是单调递减集合列, 于是

$$\mathbb{P}(E_n \text{ i.o.}) = \mathbb{P}\Big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n\Big) = \lim_{m\to\infty} \mathbb{P}\Big(\bigcup_{n=m}^{\infty} E_n\Big) = 0. \quad \blacksquare$$

- As an illustration of the convenience of the new notions, we may restate Theorem 2 as follows. The intuitive content of condition (3) below is the point being stressed here.

定理 14 (Restatement of Theorem 2)

$X_n \to 0$ *a.e. if and only if*

$$\forall\ \varepsilon > 0: \quad \mathbb{P}\{|X_n| > \varepsilon \text{ i.o.}\} = 0. \tag{9}$$

Proof. 类似地, 令 $A_m = \bigcap_{n=m}^{\infty}\{|X_n| \leqslant \varepsilon\}$ $(X=0)$, 于是我们有

$$\{|X_n| > \varepsilon \text{ i.o.}\} = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{|X_n| > \varepsilon\} = \bigcap_{m=1}^{\infty} A_m^c.$$

由定理 2, $X_n \to 0$ a.e. 当且仅当 $\forall\ \varepsilon > 0$, $\mathbb{P}(A_m^c) \to 0$ $(m \to \infty)$. 又因为 $(A_m^c)_{m\geqslant 1}$ 关于 m 是单调递减集列, 于是 $\mathbb{P}(A_m^c) \to 0$ 等价于 $\mathbb{P}\{|X_n| > \varepsilon \text{ i.o.}\} = 0$. ■

定理 15 (Convergence in pr. ⇒ Convergence a.e. Along a Subsequence)

If $X_n \to X$ in pr., then there exists a sequence (n_k) of integers increasing to infinity such that $X_{n_k} \to X$ a.e. Briefly, convergence in pr. implies convergence a.e. along a subsequence.

PROOF. 不失一般性, 设 $X \equiv 0$. 于是 $X_n \to 0$ in pr. 可写作

$$\forall\ k \in \mathbb{N}: \qquad \lim_{n\to\infty} \mathbb{P}\left(|X_n| > \frac{1}{2^k}\right) = 0.$$

于是, $\forall\ k \in \mathbb{N}$, 可取 n_k 使得

$$\mathbb{P}\left(|X_n| > \frac{1}{2^k}\right) \leqslant \frac{1}{2^k}.$$

于是

$$\sum_{k=1}^{\infty} \mathbb{P}\left(|X_{n_k}| > \frac{1}{2^k}\right) \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty.$$

对于这一子列 $(n_k)_{k\geqslant 1}$, 由 Borel-Cantelli 引理, 便有 $\mathbb{P}(|X_{n_k}| > 2^{-k} \text{ i.o.}) = 0$, 再由定理 14 就可知 $X_{n_k} \to 0$ a.e. (因为对任一 $\varepsilon > 0$ 总能够找到 $k \in \mathbb{N}$, s.t. $2^{-k} > \varepsilon$). ∎

Borel-Cantelli Lemma (Divergence Part)

- Under the assumption of independence, Theorem 13 has a striking complement.

定理 16 (Borel-Cantelli Lemma (Divergence Part))

If the events (E_n) *are* *independent*, *then*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty \quad \Rightarrow \quad \mathbb{P}(E_n \text{ i.o.}) = 1. \tag{10}$$

PROOF. 首先 $\mathbb{P}(\varliminf_n E_n^c) = \lim_{m\to\infty} \mathbb{P}(\bigcap_{n=m}^{\infty} E_n^c)$. 由 (E_n) 的完全独立性可知 (E_n^c) 的完全独立性，于是

$$\mathbb{P}\Big(\bigcap_{n=m}^{k} E_n^c\Big) = \prod_{n=m}^{k} \mathbb{P}(E_n^c) = \prod_{n=m}^{k} (1 - \mathbb{P}(E_n)).$$

又因为 $1 - x \leqslant e^{-x}$, 于是上式右端各项满足 $1 - \mathbb{P}(E_n) \leqslant e^{-\mathbb{P}(E_n)}$, 于是上式右端不会超过

$$\prod_{n=m}^{k} e^{-\mathbb{P}(E_n)} = \exp\Big\{-\sum_{n=m}^{k} \mathbb{P}(E_n)\Big\} \to 0 \quad (k \to \infty).$$

Borel-Cantelli Lemma (Divergence Part)

上式右端趋于零是因为 $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$ 这一假设. 又因为 $\bigcap_{n=m}^{k} E_n^c$ 关于 k 是单调递减集合列, 于是

$$\mathbb{P}\Big(\bigcap_{n=m}^{\infty} E_n^c\Big) = \lim_{k\to\infty} \mathbb{P}\Big(\bigcap_{n=m}^{k} E_n^c\Big) = 0.$$

这就说明

$$\mathbb{P}(\underline{\lim}_{n} E_n^c) = \lim_{m\to\infty} \mathbb{P}\Big(\bigcap_{n=m}^{\infty} E_n^c\Big) = 0.$$

由 (7) 式, 这等价于 $\mathbb{P}(E_n \text{ i.o.}) = 1$. 定理得证. ∎

Remark (Borel-Cantelli Lemma)

- Theorem 13 and 16 together will be referred to as the **Borel-Cantelli lemma**, the former the "convergence part" and the latter the "divergence part".
- The first is more useful since the events there may be completely arbitrary.
- The second has an extension to *pairwise* independent r.v.'s; although the result is of some interest, it is the method of proof to be given below that is more important.

Extension of “Divergence Part” of Borel-Cantelli Lemma

Reference. Gut, Allan. *Probability: a graduate course*. New York: Springer, 2006. Theorem 2.18.5.

定理 17 (Extension of “Divergence Part” of Borel-Cantelli Lemma)

The implication (10) remains true if the events (E_n) are pairwise independent.

Proof. 令 $I_n = \mathbf{1}_{E_n}$, 由 (E_n) 两两独立便可知

$$\forall\ m \neq n: \quad \mathbb{E}(I_m I_n) = \mathbb{E}(I_m)\mathbb{E}(I_n). \tag{11}$$

另外, 考虑 $\sum_{n=1}^{\infty} I_n$, 我们有

$$\begin{aligned} \sum_{n=1}^{\infty} I_n(\omega) = \infty &\iff I_n(\omega) \text{ 中有无限项等于 } 1 \\ &\iff \omega \text{ 属于无限项 } E_n \iff \omega \in \overline{\lim_{n\to\infty}} E_n. \end{aligned}$$

于是 $\mathbb{P}(E_n \text{ i.o.}) = \mathbb{P}\{\sum_{n=1}^{\infty} I_n = +\infty\}$. 另外, 注意到 $\mathbb{P}(E_n) = \mathbb{E}(I_n)$, 于是 (10) 式等价于:

$$\sum_{n=1}^{\infty} \mathbb{E}(I_n) = \infty \quad \Rightarrow \quad \mathbb{P}\left\{\sum_{n=1}^{\infty} I_n = +\infty\right\} = 1. \tag{12}$$

由 Chebyshev 不等式，令 $X=\sum_{n=1}^k I_n-\mathbb{E}(\sum_{n=1}^k I_n)$，$\varphi(u)=u^2$，我们有

$$\mathbb{P}\left\{\left|\sum_{n=1}^k (I_n-\mathbb{E}(I_n))\right|>\frac{1}{2}\sum_{n=1}^k \mathbb{P}(E_n)\right\}\leqslant\frac{\operatorname{Var}(\sum_{n=1}^k I_n)}{(\frac{1}{2}\sum_{n=1}^k \mathbb{P}(E_n))^2}$$

$$=\frac{4\sum_{n=1}^k \mathbb{P}(E_n)(1-\mathbb{P}(E_n))}{(\sum_{n=1}^k \mathbb{P}(E_n))^2}\leqslant\frac{4}{\sum_{n=1}^k \mathbb{P}(E_n)}\to 0\quad(k\to\infty).$$

上式成立只需要注意到 $\mathbb{E}(X-\mathbb{E}X)^2=\operatorname{Var}X$，并且 $\operatorname{Var}I_n=\mathbb{P}(E_n)(1-\mathbb{P}(E_n))$. 由上式，

$$\mathbb{P}\left\{\left|\sum_{n=1}^k (I_n-\mathbb{E}(I_n))\right|\leqslant\frac{1}{2}\sum_{n=1}^k \mathbb{P}(E_n)\right\}\to 1\quad(k\to\infty).$$

将上式中的绝对值符号打开并注意到 $\mathbb{E}(I_n)=\mathbb{P}(E_n)$，于是有

$$\mathbb{P}\left\{\sum_{n=1}^k I_n\geqslant\frac{1}{2}\sum_{n=1}^k \mathbb{E}(I_n)\right\}\leqslant\mathbb{P}\left\{\frac{1}{2}\sum_{n=1}^k \mathbb{E}(I_n)\leqslant\sum_{n=1}^k I_n\leqslant\frac{3}{2}\sum_{n=1}^k \mathbb{E}(I_n)\right\}\to 1\quad(k\to\infty).$$

紫色式中的两个求和号均随 k 的增加而增加，于是首先令 $\sum_{n=1}^k I_n$ 中的 $k\to\infty$，再令 $\sum_{n=1}^k \mathbb{E}(I_n)$ 中的 $k\to\infty$ 就有 $\mathbb{P}\{\sum_{n=1}^\infty I_n=+\infty\}=1$. 定理得证. ∎

Typical Example of "Zero-or-One" Law

- This is an example of a "zero-or-one" law to be discussed in Chapter 5, though it is not included in any of the general results there.

推论 18 (Typical Example of "Zero-or-One" Law)

If the events (E_n) *are pairwise independent, then*

$$\mathbb{P}(E_n \text{ i.o.}) = 0 \text{ or } 1$$

according as $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ *or* $= \infty$.

Back to Content.

Section 3

Vague Convergence

- In the next two sections we will study another important type of convergence of r.v.'s — convergence in distribution (or weak convergence).
- In this section we begin with a discussion of *vague convergence*, which is a type of convergence of p.m.'s of r.v.'s.

- If a sequence of r.v.'s (X_n) tends to a limit, the corresponding sequence of p.m.'s (μ_n) ought to tend to a limit in some sense.
- QUESTION. Is it true that $\lim_{n\to\infty} \mu_n(A)$ exists for all $A \in \mathscr{B}^1$ or at least for all intervals A? The answer is no from trivial examples.

例 19

Let $X_n = c_n$ where the c_n's are constants tending to zero, then $X_n \to 0$. For any interval I such that $0 \notin \overline{I}$, where $\overline{I}$ is the closure of I, we have $\lim_{n\to\infty} \mu_n(I) = 0 = \mu(I)$; for any interval such that $0 \in \mathring{I}$, where $\mathring{I}$ is the interior of I, we have $\lim_{n\to\infty} \mu_n(I) = 1 = \mu(I)$.

But if (c_n) oscillates (摆动) between strictly positive and strictly negative values, and $I = (a,0)$ or $(0,b)$, where $a < 0 < b$, then $\mu_n(I)$ oscillates between 0 and 1, while $\mu(I) = 0$. At this time $\lim_{n\to\infty} \mu_n(I)$ not exists! On the other hand, if $I = (a,0]$ or $[0,b)$, then $\mu_n(I)$ oscillates as before but $\mu(I) = 1$.

- HINT. $\mu_n(I) = \mathbb{P}\{\omega : X_n(\omega) \in I\}$.

Subprobability Measure and Vague Convergence

定义 20 (Subprobability Measure)

A measure μ on $(\mathbb{R}, \mathscr{B}^1)$ with $\mu(\mathbb{R}) \leqslant 1$ will be called a **subprobability measure** (s.p.m.).

定义 21 (Vague Convergence)

A sequence (μ_n) of s.p.m.'s is said to **converge vaguely (淡收敛)** to an s.p.m. μ iff there exists a dense subset D of $\mathbb{R}$ such that

$$\forall\ a \in D, b \in D, a < b: \quad \mu_n((a,b]) \to \mu((a,b]). \tag{13}$$

This will be denoted by $\mu_n \xrightarrow{v} \mu$ and μ is called the **vague limit** of (μ_n).

- We will write $\mu((a,b])$ as $\mu(a,b]$ below, and similarly for other kinds of intervals.

定义 22 (Atom)

An **atom (原子)** of any measure μ on $\mathscr{B}^1$ is a singleton $\{x\}$ such that $\mu(\{x\}) > 0$.

- An interval (a,b) is called a **continuity interval (连续区间)** of μ iff neither a nor b is an atom of μ; in other words iff $\mu(a,b) = \mu[a,b]$.
- As a notational convention, $\mu(a,b) = 0$ when $a > b$.

Equivalent Propositions of Vague Convergence

定理 23 (Equivalent Propositions of Vague Convergence)

Let (μ_n) and μ be s.p.m.'s. The following propositions are equivalent.

i. *For every finite interval (a,b) and $\varepsilon>0$, there exists an $n_0(a,b,\varepsilon)$ such that if $n\geqslant n_0$, then*

$$\mu(a+\varepsilon,b-\varepsilon)-\varepsilon\leqslant\mu_n(a,b)\leqslant\mu(a-\varepsilon,b+\varepsilon)+\varepsilon. \tag{14}$$

Here and hereafter the first term is interpreted as 0 if $a+\varepsilon>b-\varepsilon$.

ii. *For every continuity interval (a,b) of μ, we have $\mu_n(a,b)\to\mu(a,b)$.*

iii. $\mu_n\xrightarrow{v}\mu$.

PROOF. (i) $\Rightarrow$ (ii). 不妨设 (a,b) 是 μ 的连续区间. 由测度的单调性, 有

$$\lim_{\varepsilon\downarrow 0}\mu(a+\varepsilon,b-\varepsilon)=\mu(a,b)=\mu[a,b]=\lim_{\varepsilon\downarrow 0}\mu(a-\varepsilon,b+\varepsilon).$$

在 (14) 式中先令 $n\to\infty$, 再令 $\varepsilon\downarrow 0$, 并注意到 $\mu_n(a,b)\leqslant\mu_n[a,b]$ (测度的单调性), 便有

$$\mu(a,b)\leqslant\varliminf_{n\to\infty}\mu_n(a,b)\leqslant\varlimsup_{n\to\infty}\mu_n[a,b]\leqslant\mu[a,b]=\mu(a,b).$$

这就说明 $\mu(a,b)=\lim\limits_{n\to\infty}\mu_n(a,b)$. 事实上此处 $(a,b]$ 可以替换为 $(a,b]$ 或 $[a,b)$ 或 $[a,b]$.

Equivalent Propositions of Vague Convergence

(ii) $\Rightarrow$ (iii). 设 $C\subseteq\mathbb{R}$ 是 μ 的原子构成的集合, i.e. $\forall\ c\in C$, $\mu(\{c\})>0$. 注意到 $\mu(-\infty,x]$ 是关于 x 的单调不减右连续函数, 原子 c 就是函数 $\mu(-\infty,x]$ 的间断点, 从而至多可列. 取 $D=\mathbb{R}\backslash C$, 则 D 是 $\mathbb{R}$ 上的稠密集, 且 $\forall\ a,b\in D$, $a<b$, 有 $\mu_n(a,b)\to\mu(a,b)$. 又因为 $\mu(\{b\})=0$, 于是 $\mu_n(a,b]\to\mu(a,b]$, i.e. $\mu_n\xrightarrow{v}\mu$.

(iii) $\Rightarrow$ (i). 现设 $\mu_n\xrightarrow{v}\mu$, 任取 $(a,b)\subseteq\mathbb{R}$, 任取 $\varepsilon>0$, 则存在 $a_1,a_2,b_1,b_2\in D$, s.t.

$$a-\varepsilon<a_1<a<a_2<a+\varepsilon,\quad b-\varepsilon<b_1<b<b_2<b+\varepsilon.$$

由 (13) 式, 可知存在 $n_0\in\mathbb{N}$, s.t. $\forall\ n\geqslant n_0$, 有

$$\left|\mu_n(a_i,b_j]-\mu(a_i,b_j]\right|\leqslant\varepsilon,\quad i=1,2;\ j=1,2.$$

于是

$$\begin{aligned}\mu(a+\varepsilon,b-\varepsilon)-\varepsilon&\leqslant\mu(a_2,b_1]-\varepsilon\leqslant\mu_n(a_2,b_1]\leqslant\mu_n(a,b)\leqslant\mu_n(a_1,b_2]\\&\leqslant\mu(a_1,b_2]+\varepsilon\leqslant\mu(a-\varepsilon,b+\varepsilon)+\varepsilon.\end{aligned}$$

自此定理得证. $\blacksquare$

The Vague Limit is Unique

- As an immediate consequence, the *vague limit is unique*.

推论 24 (The Vague Limit is Unique)

If besides (13) we have also a dense subset D' of $\mathbb{R}$ such that

$$\forall\ a \in D', b \in D', a < b: \quad \mu_n(a,b] \to \mu'(a,b],$$

then $\mu \equiv \mu'$.

Proof. 现设 A 是测度 μ 与 μ' 的原子组成的集合，则 A 至多可列。若 $a, b \in A^c$，则由定理 23 (ii) 可知

$$\mu'(a,b) \leftarrow \mu_n(a,b) \to \mu(a,b) \quad \Rightarrow \quad \mu(a,b) = \mu'(a,b).$$

从而就有 $\mu(a,b] = \mu'(a,b]$。注意到 A^c 在 $\mathbb{R}$ 中稠密，于是设 $\mathscr{S} = \{(a,b] : a, b \in A^c\}$，则

$$\forall\ I \in \mathscr{S}: \quad \mu(I) = \mu'(I).$$

由 μ 与 μ' 的可列可加性可知二者在由 $\mathscr{S}$ 生成的代数 $\mathscr{A}(\mathscr{S})$ 上一致。又因为 μ 与 μ' 是有限测度，由测度扩张的唯一性可知 μ 与 μ' 在 $\mathscr{B}^1$ 上一致。从而 $\mu \equiv \mu'$. ∎

- Recall the definition of **sequentially compact**: Given any sequence of numbers in the set, there is a subsequence which converges, and the limit is also a number in the set.
- We have the following analogue which states: *The set of all s.p.m.'s is sequentially compact with respect to vague convergence.*

定理 25 (Helly's Selection Principle)

Given any sequence of s.p.m.'s, there is a subsequence that converges vaguely to an s.p.m.

Proof. 类似于概率测度, 我们可以定义次分布函数 (subdistribution function) (s.d.f.) F_n:

$$\forall\ x: \quad F_n(x) = \mu_n(-\infty, x].$$

易知 F_n 为单调不减右连续函数, 并且 $F_n(-\infty) = 0$, $F_n(+\infty) = \mu_n(\mathbb{R}) \leqslant 1$.

Step 1. 证明在 $\mathbb{R}$ 上某一可列稠密集 D 上存在子列 $(F_{n_k})_{k\geqslant 1}$ 与不减函数 G, s.t. $F_{n_k} \to G$.
设 $D = (r_k)_{k\geqslant 1}$ 是 $\mathbb{R}$ 上的可列稠密子集. 现在, 数列 $\{F_n(r_1), n \geqslant 1\}$ 有界, 于是由实数的 Bolzano-Weierstrass 定理, 存在 $\{F_n, n \geqslant 1\}$ 的子列 $\{F_{1k}, k \geqslant 1\}$ s.t.

$$\lim_{k\to\infty} F_{1k}(r_1) = \ell_1 \text{ 存在}.$$

接下来, 数列 $\{F_{1k}(r_2), k \geqslant 1\}$ 有界, 于是存在子列 $\{F_{2k}, k \geqslant 1\}$ s.t. $\lim_{k\to\infty} F_{2k}(r_2) = \ell_2$ 存在. 而 F_{2k} 是 F_{1k} 的子列, 于是 $\lim_{k\to\infty} F_{2k}(r_1) = \ell_1$. 一般地, 我们有

$$
\begin{array}{rl}
F_{11}, F_{12}, \cdots, F_{1k}, \cdots & \text{收敛于 } r_1; \\
F_{21}, F_{22}, \cdots, F_{2k}, \cdots & \text{收敛于 } r_1, r_2; \\
\cdots\cdots\cdots\cdots\cdots & \\
F_{j1}, F_{j2}, \cdots, F_{jk}, \cdots & \text{收敛于 } r_1, \cdots, r_j; \\
\cdots\cdots\cdots\cdots\cdots &
\end{array}
$$

现在考虑对角线序列 $\{F_{kk}, k \geqslant 1\}$, 事实上, $\forall\ j \geqslant 1$, 数列 $\{F_{kk}(r_j), k \geqslant 1\}$ 收敛于 ℓ_j. 这是因为对于固定的 $j \in \mathbb{N}$, $\{F_{kk}, k \geqslant j\}$ 是 $\{F_{jk}, k \geqslant 1\}$ 的子列, 于是二者收敛于同一极限.

现在, 在可列稠子集 D 上, 我们已经找到了 (F_n) 的子列 $(F_{n_k})_{k\geqslant 1}$ 与函数 G, s.t.

$$\forall\ r \in D: \quad \lim_{k\to\infty} F_{n_k}(r) = G(r).$$

因为 $\forall\ k \geqslant 1$, F_{n_k} 单调不减, 于是 G 也是不减函数. 因为 $\forall\ s < t \in D$,

$$0 \leqslant F_{n_k}(t) - F_{n_k}(s) \to G(t) - G(s) \quad (k \to \infty).$$

Step 2. 证明子列 $(F_{n_k})_{k\geqslant 1}$ 在 $\mathbb{R}$ 上收敛于单调不减右连续函数 F. 定义 $\mathbb{R}$ 上的函数

$$\forall\ x \in \mathbb{R}: \quad F(x) = \inf_{x<r\in D} G(r).$$

由第一章单调函数的性质 (vii) 可知 F 单调不减右连续. 令 C 是 F 的连续点构成的集合, 则 C^c 至多可列, 于是 $C \subseteq \mathbb{R}$ 稠密并且有

$$\forall\ x \in C: \quad \lim_{k\to\infty} F_{n_k}(x) = F(x). \tag{15}$$

事实上, $\forall\ x \in C$, $\forall\ \varepsilon > 0$, 则 $\exists\ r, r', r'' \in D$, s.t. $r < r' < x < r''$ 且 $F(r'') - F(r) \leqslant \varepsilon$. 于是我们有

$$F(r) \leqslant G(r') \leqslant F(x) \leqslant G(r'') \leqslant F(r'') \leqslant F(r) + \varepsilon;$$

还有

$$G(r') \leftarrow F_{n_k}(r') \leqslant F_{n_k}(x) \leqslant F_{n_k}(r'') \to G(r''),$$

于是由 ε 的任意性可知 (15) 式成立.

单调不减右连续函数 F 唯一决定了一个次概率测度 μ s.t. $F(x) - F(-\infty) = \mu(-\infty, x]$. 再由 (15) 式便有

$$\forall\ a, b \in C, a < b: \quad \lim_{k\to\infty} \mu_{n_k}(a, b] = \mu(a, b],$$

因此 $\mu_{n_k} \xrightarrow{v} \mu$, 定理得证. ■

- We say that F_n **converges vaguely** to F and write $F_n \xrightarrow{v} F$ for $\mu_n \xrightarrow{v} \mu$ where μ_n and μ are the s.p.m.'s corresponding to the s.d.f's F_n and F.

定理 26 (Properties of convergence of any subsequence)

If every vaguely convergent subsequence of the sequence of s.p.m.'s (μ_n) converges to the same μ, then $\mu_n \xrightarrow{v} \mu$.

PROOF. 反证法. 设 μ_n 不淡收敛于 μ, 由定理 23 (ii), 存在 μ 的连续区间 (a,b), s.t. $\mu_n(a,b) \nrightarrow \mu(a,b)$. 再由 Bolzano-Weierstrass 定理, 存在收敛子列 $\mu_{n_k}(a,b)$, 不妨设 $\mu_{n_k}(a,b) \to L \neq \mu(a,b)$, 于是由定理 25, 存在 $(\mu_{n_k})_{k\geqslant 1}$ 的子列 $\mu_{n'_k}$, 由定理条件, $\mu_{n'_k}$ 作为 μ_n 的子列淡收敛于 μ, 于是再由定理 23 (ii) 可知

$$\mu_{n'_k}(a,b) \to \mu(a,b) \quad (k \to \infty).$$

但是数列 $\mu_{n'_k}(a,b)$ 作为 $\mu_{n_k}(a,b)$ 的子列, 二者极限一致, 于是

$$\mu_{n'_k}(a,b) \to L \neq \mu(a,b) \quad (k \to \infty).$$

于是构成矛盾! 定理得证. Back to Content. ■

Section 4

Continuation; Convergence in Distribution

- This section continues with a discussion of *equivalent propositions* for vague convergence.
- Next, the definition and basic properties of convergence in distribution are presented.

Classes of Continuous Functions on $\mathbb{R}$

- We proceed to discuss another kind of criterion, which is becoming ever more popular in measure theory as well as functional analysis.
- This has to do with classes of continuous functions on $\mathbb{R}$.

$$\begin{aligned}
C_K &= \text{ the class of continuous functions } f \text{ each vanishing outside a } \textit{compact set } K(f); \\
C_0 &= \text{ the class of continuous functions } f \text{ such that } \lim_{|x|\to\infty} f(x) = 0; \\
C_B &= \text{ the class of bounded continuous functions;} \\
C &= \text{ the class of continuous functions.}
\end{aligned}$$

  We have $C_K \subseteq C_0 \subseteq C_B \subseteq C$. It is well known that C_0 is the closure of C_K with respect to uniform convergence.[2]
- An arbitrary function f on an arbitrary space is said to **have support in** a subset S of the space iff it vanishes outside S. Thus if $f \in C_K$, then it has support in a certain compact set, hence also in a certain compact interval.
- A **step function on a finite or infinite interval** (a, b) is one with support in it such that $f(x) = c_j$ for $x \in (a_j, a_{j+1})$ for $1 \leqslant j \leqslant \ell$, where ℓ is finite, $a = a_1 < \cdots < a_\ell = b$, and the c_j's are arbitrary numbers.

[2] A subset K of a metric space X is said to be **compact** if any sequence of K has a subsequence converges to an element of K.

- A step function will be called D-**valued** (D-值) iff all the a_j's and c_j's are belong to a given set D.
- Note that the values of step function f at the points a_j are left unspecified to allow for flexibility; frequently they are defined by right or left continuity.
- The following lemma is basic.

定理 27 (Approximation Lemma)

Suppose that $f \in C_K$ has support in the compact interval $[a,b]$. Given any dense subset A of $\mathbb{R}$ and $\varepsilon > 0$, there exists an A-valued step function f_ε on (a,b) such that

$$\sup_{x\in\mathbb{R}} \left| f(x) - f_\varepsilon(x) \right| \leqslant \varepsilon. \tag{16}$$

If $f \in C_0$, the same is true if (a,b) is replaced by $\mathbb{R}$.

- In fact, for any $f \in C_K$, one may even require that either $f_\varepsilon \leqslant f$ or $f_\varepsilon \geqslant f$.
- This theorem is also a particular case of the Stone-Weierstrass theorem.

Equivalent Proposition 1 of Vague Convergence

- The discussion in this section is meant in part to introduce some modern terminology to the relevant applications in probability theory.

定理 28 (Alternative Criterion for Vague Convergence)

Let (μ_n) and μ be s.p.m.'s. Then $\mu_n \xrightarrow{v} \mu$ if and only if

$$\forall\ f \in C_K\ [\text{or } C_0]:\quad \int_{\mathbb{R}} f(x)\,\mu_n(dx) \to \int_{\mathbb{R}} f(x)\,\mu(dx). \tag{17}$$

PROOF. (⇒). 设 $\mu_n \xrightarrow{v} \mu$, 不妨设 $D \subseteq \mathbb{R}$ 稠密. 首先, 对任意的 $a, b \in D$, 当 f 是区间 $(a,b]$ 的示性函数时, (17) 式成立:

$$\int_{(a,b]} \mu_n(dx) = \mu_n(a,b] \to \mu(a,b] = \int_{(a,b]} \mu(dx).$$

于是由积分的线性, 当 f 为 D-值阶梯函数时, (17) 式仍然成立. 现在 $\forall\ f \in C_0$, $\forall\ \varepsilon > 0$, 由逼近定理 27, $\exists\ D$-值阶梯函数 f_ε s.t. (16) 式成立.

现在我们有

$$\left|\int f\,d\mu_n-\int f\,d\mu\right|\leqslant\left|\int(f-f_\varepsilon)\,d\mu_n\right|+\left|\int f_\varepsilon\,d\mu_n-\int f_\varepsilon\,d\mu\right|+\left|\int(f_\varepsilon-f)\,d\mu\right|.$$

其中上式右端第一项, 由积分的性质, 有如下估计:

$$\left|\int(f-f_\varepsilon)\,d\mu_n\right|\leqslant\int|f-f_\varepsilon|\,d\mu_n\leqslant\varepsilon\int 1\,d\mu_n=\varepsilon\,\mu_n(\mathbb{R})\leqslant\varepsilon.$$

上式右端第三项也类似. 另外, 因为 f_ε 为阶梯函数, 于是 (17) 式成立, 从而上式右端第二项趋于零. 现在, 当 $n\to\infty$ 时, 由 ε 的任意性便有

$$\left|\int f\,d\mu_n-\int f\,d\mu\right|\leqslant 2\varepsilon\to 0.$$

($\Leftarrow$). 设 (17) 式对 $\forall\ f\in C_K$ 成立, 并令 A 为 μ 的原子构成的集合, 去证明 $\mu_n\xrightarrow{v}\mu$ (稠密集取 $D=A^c$). 首先, $\forall\ a,b\in D$, 令 $g=\mathbf{1}_{(a,b]}$. 现在, $\forall\ \varepsilon>0$, $\exists\ \delta>0$, s.t. $a+\delta<b-\delta$ 并且 $\mu(U)<\varepsilon$, 其中 $U=(a-\delta,a+\delta)\cup(b-\delta,b+\delta)$.

Equivalent Proposition 1 of Vague Convergence

定义 $g_1, g_2 \in C_K$ 如下:

$$g_1 = \begin{cases} g, & x \in (-\infty, a] \cup [a+\delta, b-\delta] \cup [b, \infty), \\ \text{linear function}, & \text{otherwise;} \end{cases}$$

$$g_2 = \begin{cases} g, & x \in (-\infty, a-\delta] \cup [a, b] \cup [b+\delta, \infty), \\ \text{linear function}, & \text{otherwise.} \end{cases}$$

从而 $g_1 \leqslant g \leqslant g_2 \leqslant g_1 + 1$, 于是 $\int g_1 \, d\mu_n \leqslant \int g \, d\mu_n \leqslant \int g_2 \, d\mu_n$ (μ_n 替换为 μ 也成立). 又因为 $g_1, g_2 \in C_K$, 于是由 (17) 式便有

$$\int g_1 \, d\mu_n \to \int g_1 \, d\mu, \quad \int g_2 \, d\mu_n \to \int g_2 \, d\mu.$$

并且

$$\int g_2 \, d\mu - \int g_1 \, d\mu \leqslant \int_U 1 \, d\mu = \mu(U) < \varepsilon.$$

于是 $\int g \, d\mu_n \to \int g \, d\mu$, i.e. $\mu_n(a, b] = \int_{(a,b]} 1 \, d\mu_n \to \int_{(a,b]} 1 \, d\mu = \mu(a, b]$ ($\forall \ a, b \in D$), 这就说明 $\mu_n \xrightarrow{v} \mu$. 定理得证. ∎

Equivalent Proposition 1 of Vague Convergence

推论 29 (Corollary: Equivalent Proposition 1 of Vague Convergence)

If (μ_n) *is a sequence of s.p.m.'s such that for every* $f \in C_K$*,*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\,\mu_n(dx)$$

exists, then (μ_n) *converges vaguely.*

PROOF. 由定理 25, 存在 (μ_n) 的子列 (μ_{n_k}) 淡收敛于 μ. 再由定理 28, 有

$$\int_{\mathbb{R}} f(x)\,\mu_{n_k}(dx) \to \int_{\mathbb{R}} f(x)\,\mu(dx). \quad \text{数列收敛}$$

又数列极限 $\lim\limits_{n\to\infty} \int_{\mathbb{R}} f(x)\,\mu_n(dx)$ 存在, 于是 $\lim\limits_{k\to\infty} \int_{\mathbb{R}} f(x)\,\mu_{n_k}(dx) = \lim\limits_{n\to\infty} \int_{\mathbb{R}} f(x)\,\mu_n(dx)$. 所以对 (μ_n) 的任一淡收敛的子列 $(\mu_{n'_k})$ 一定有

$$\lim_{k\to\infty} \int_{\mathbb{R}} f(x)\,\mu_{n'_k}(dx) = \int_{\mathbb{R}} f(x)\,\mu(dx).$$

于是对 (μ_n) 的任一淡收敛的子列 $(\mu_{n'_k})$ 一定有 $\mu_{n'_k} \xrightarrow{v} \mu$. 再由定理 26 便可知 $\mu_n \xrightarrow{v} \mu$. ■

Equivalent Proposition 1: Case of P.M.'s

定理 30 (Equivalent Proposition 1: Case of Probability Measures)

Let (μ_n) and μ be p.m.'s. Then $\mu_n \xrightarrow{v} \mu$ if and only if

$$\forall\ f \in C_B: \quad \int_{\mathbb{R}} f(x)\,\mu_n(dx) \to \int_{\mathbb{R}} f(x)\,\mu(dx). \tag{18}$$

Proof. ($\Leftarrow$) 在定理 28 中已经证明完毕, 因为 $C_K \subseteq C_0 \subseteq C_B$. ($\Rightarrow$). 现设 $\mu_n \xrightarrow{v} \mu$ (稠密集取 D). $\forall\ \varepsilon > 0$, $\exists\ a, b \in D$, s.t. $\mu((a,b]^c) = 1 - \mu(a,b] < \varepsilon/2$. 又因为 $\mu_n \xrightarrow{v} \mu$, 于是对于上述区间 $(a,b]$ 与 $\varepsilon > 0$, $\exists\ n_0(\varepsilon) > 0$, s.t. $\forall\ n \geqslant n_0(\varepsilon)$, 有 $|\mu_n(a,b] - \mu(a,b]| < \varepsilon/2$. 放缩之后便有

$$\mu_n((a,b]^c) = 1 - \mu_n(a,b] < \varepsilon.$$

现在任取 $f \in C_B$, 不妨设 $|f| \leqslant M < \infty$, 定义函数 $f_\varepsilon \in C_K$ 如下:

$$f_\varepsilon(x) = \begin{cases} f(x), & x \in [a,b], \\ 0, & x \in (-\infty, a-1) \cup (b+1, \infty), \\ \text{linear function}, & \text{otherwise}. \end{cases}$$

则 $|f - f_\varepsilon| \leqslant 2M$.

又因为 $f_\varepsilon \in C_K$, 于是由定理 28 可知

$$\int_{\mathbb{R}} f_\varepsilon \, d\mu_n \to \int_{\mathbb{R}} f_\varepsilon \, d\mu.$$

另外,

$$\int_{\mathbb{R}} |f - f_\varepsilon| \, d\mu_n \leqslant \int_{(a,b]^c} 2M \, d\mu_n \leqslant 2M\varepsilon,$$

上式中将测度 μ_n 替换为 μ 也成立. 现在, 类似于定理 28 必要性证明中的估计 (P40),

$$\begin{aligned} \left| \int f \, d\mu_n - \int f \, d\mu \right| &\leqslant \left| \int (f - f_\varepsilon) \, d\mu_n \right| + \left| \int f_\varepsilon \, d\mu_n - \int f_\varepsilon \, d\mu \right| + \left| \int (f_\varepsilon - f) \, d\mu \right| \\ &\leqslant 4M\varepsilon \quad (n \to \infty). \end{aligned}$$

再由 ε 的任意性便可知 $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$. 必要性得证. ∎

When the Vague Limit is a P.M.

Remark

Theorem 25 and 26 deal with s.p.m.'s. Even if the given sequence (μ_n) consists only of strict p.m.'s, the sequential vague limit may not be so. 概率测度列的淡极限不一定仍是概率测度

例 31 (Counterexample of the Remark)

Let $X_n = c_n$ where $c_n \to \infty$. Then $X_n \to \infty$ deterministically. According to the definition of r.v., the constant $+\infty$ is a r.v. But for any finite interval (a,b) to have $\lim_{n\to\infty} \mu_n(a,b) = 0$, so any limiting measure must also be identically zero (not a p.m.).

- It is sometimes demanded that such a limit be a p.m. The following criterion is not deep, but applicable.

定理 32 (When the Vague Limit is a P.M.)

Let a family of p.m.'s $(\mu_\alpha)_{\alpha\in A}$ be given on an arbitrary index set A. In order that every sequence of them contains a subsequence which converges vaguely to a p.m., it is necessary and sufficient that the following condition be satisfied: for any $\varepsilon > 0$, there exists a finite interval I such that

$$\inf_{\alpha\in A} \mu_\alpha(I) > 1 - \varepsilon. \tag{19}$$

PROOF. $(\Leftarrow)$ 充分性. 若 $\forall\ \varepsilon>0$, 存在有限区间 I 满足 (19) 式, 任取序列 $(\mu_n)_{n\geqslant 1}\subseteq(\mu_\alpha)$, 由定理 25, 存在子列 $(\mu'_n)\subseteq(\mu_n)$ s.t. $\mu'_n \xrightarrow{v} \mu$. 现在去证 μ 为概率测度. 现取 J 为 μ 的某一连续区间且 $J\supseteq I$, 则

$$
\begin{aligned}
\mu(\mathbb{R}) \geqslant \mu(J) &= \lim_{n\to\infty}\mu'_n(J) = \varlimsup_{n\to\infty}\mu'_n(J) \geqslant \varlimsup_{n\to\infty}\mu'_n(I)\\
&= \lim_{k\to\infty}\sup_{n\geqslant k}\mu'_n(I) \geqslant \lim_{k\to\infty}\inf_{n\geqslant k}\mu'_n(I) \geqslant 1-\varepsilon.
\end{aligned}
$$

上式最后一个不等式成立因为 (19) 式. 再由 ε 的任意性可知 $\mu(\mathbb{R})=1$, 故 μ 为概率测度.

$(\Rightarrow)$ 必要性. 去证逆否命题: $\exists\ \varepsilon>0$, 对任一有限区间 I 均有 $\inf_{\alpha\in A}\mu_\alpha(I)\leqslant 1-\varepsilon\ \Rightarrow$ 存在序列 (μ_n), 其某一子列的淡极限是次概率测度. 现在, 存在 $\varepsilon>0$, 任取单调上升有限区间序列 $I_n\uparrow\mathbb{R}$, 则 $\forall\ n\in\mathbb{N}$, $\inf_{\alpha\in A}\mu_\alpha(I_n)\leqslant 1-\varepsilon$. 由下确界的定义, 对于上述 $\varepsilon>0$, $\exists\ \alpha_0\in A$, s.t.

$$
\mu_{\alpha_0}(I_n)\leqslant \inf_{\alpha\in A}\mu_\alpha(I_n)+\varepsilon/2\leqslant 1-\varepsilon/2.
$$

将每个 μ_{α_0} 记作 μ_n, 于是我们得到序列 (μ_n) 满足 $\mu_n(I_n)\leqslant 1-\varepsilon/2$ $(\forall\ n\in\mathbb{N})$. 对于序列 (μ_n), 存在子列 (μ'_n), s.t. $\mu'_n \xrightarrow{v} \mu$. 现设 J 为 μ 的任一连续区间, 则对于充分大的 $n\in\mathbb{N}$, 有 $J\subseteq I$, 于是

$$
\mu(J)=\lim_{n\to\infty}\mu'_n(J)\leqslant \varliminf_{n\to\infty}\mu'_n(I_n)\leqslant 1-\varepsilon/2.
$$

由 J 的任意性, 便有 $\mu(\mathbb{R})\leqslant 1-\varepsilon/2$, 于是 μ 是次概率测度. 定理得证. ∎

Vague Convergence in Topological Space

Remark

- A family of p.m.'s satisfying the condition above involving (19) is said to be **tight**.
- Theorem 32 can be stated as follows: *a family of p.m.'s is relatively compact iff it is tight*. The word "relatively" purports that the limit need not belong to the family; the word "compact" is an abbreviation of "sequentially vaguely convergent to a strict p.m."

- Extension of the result to p.m.'s in more general topological spaces is straight-forward but plays an important role in the convergence of stochastic processes.
- The new definition of vague convergence in Theorem 28 has the advantage over the older ones in that *it can be at once carried over to measures in more general topological spaces*.
- There is no substitute for "intervals" in such a space but the classes C_K, C_0 and C_B are readily available.
- We will illustrate the general approach by indicating one more result in this direction.

Vague Convergence in Topological Space

- A **lower semicontinuous** function on $\mathbb{R}$ is defined by

$$\forall\ x \in \mathbb{R}: \quad f(x) \leqslant \varliminf_{x \neq y \to x} f(y). \tag{20}$$

- f is bounded and lower semicontinuous if and only if there exists a sequence of functions $f_k \in C_B$ which increases to f everywhere, and we call f **upper semicontinuous** if and only if $-f$ is lower semicontinuous.
- Usually f is allowed to be extended-valued; but to avoid complications we will deal with *bounded functions* only and denote by L and U respectively the classes of *bounded lower semicontinuous* and *bounded upper semicontinuous* functions.

定理 33 (Vague Convergence in Topological Space)

If (μ_n) and μ are p.m.'s, then $\mu_n \xrightarrow{v} \mu$ if and only if one of the two conditions below is satisfied:

$$\begin{aligned} \forall\ f \in L: \quad & \varliminf_{n \to \infty} \int_{\mathbb{R}} f(x)\, \mu_n(dx) \geqslant \int_{\mathbb{R}} f(x)\, \mu(dx), \\ \forall\ g \in U: \quad & \varlimsup_{n \to \infty} \int_{\mathbb{R}} g(x)\, \mu_n(dx) \leqslant \int_{\mathbb{R}} g(x)\, \mu(dx). \end{aligned} \tag{21}$$

Vague Convergence in Topological Space

Proof. 首先我们注意到 (21) 式中的两个条件是相互等价的, 只需令 $f=-g$ 即可.

($\Rightarrow$). 设 $\mu_n \xrightarrow{v} \mu$, 再取 $f_k \in C_B$ 且 $f_k \uparrow f$, 则由定理 30

$$\varliminf_{n\to\infty} \int f\,d\mu_n \geqslant \lim_{n\to\infty} \int f_k\,d\mu_n = \int f_k\,d\mu.$$

令 $k\to\infty$, 由积分的单调收敛定理, 我们有 $\int f_k\,d\mu \to \int f\,d\mu$, 于是

$$\varliminf_{n\to\infty} \int f\,d\mu_n \geqslant \int f\,d\mu.$$

($\Leftarrow$). 设 (21) 式成立, 任取 $\varphi \in C_B$, 则 $\varphi \in L$ 且 $\varphi \in U$. 于是由 (21) 式:

$$\int \varphi\,d\mu \leqslant \varliminf_{n\to\infty} \int \varphi\,d\mu_n \leqslant \varlimsup_{n\to\infty} \varphi\,d\mu_n \leqslant \int \varepsilon\,d\mu.$$

这就说明

$$\forall\ \varphi \in C_B: \quad \lim_{n\to\infty} \int \varphi\,d\mu_n = \int \varphi\,d\mu.$$

再由定理 30 便可知 $\mu_n \xrightarrow{v} \mu$. 定理得证. ■

REMARK. (21) remains true if, e.g., f is lower semicontinuous, with $+\infty$ as a possible value but bounded below.

推论 34 (Equivalent Conditions of Vague Convergence in Topological Space)

The conditions in (21) may be replaced by the following:

$$\text{for every open } O: \quad \varliminf_{n\to\infty} \mu_n(O) \geqslant \mu(O);$$
$$\text{for every closed } C: \quad \varlimsup_{n\to\infty} \mu_n(C) \leqslant \mu(C).$$

PROOF. 分别令 $f = \mathbf{1}_O$, $g = \mathbf{1}_C$, 不难验证 $f \in L$, $g \in U$. 再利用定理 33 便可以证明上述推论成立:

$$\int_{\mathbb{R}} \mathbf{1}_O(x)\,\mu_n(dx) = \int_O 1\,\mu_n(dx) = \mu_n(O). \quad \blacksquare$$

Convergence in Distribution

- Finally, we return to the connection between the convergence of r.v.'s and that of their distributions.

定义 35 (Convergence in Distribution (in dist.))

A sequence of r.v.'s (X_n) is said to **converge in distribution** (or **weakly**) to X, denoted by $X_n \xrightarrow{d} X$, iff the sequence (F_n) of corresponding d.f.'s *converges vaguely* to the d.f. F.

- In the above definition, F_n's is the d.f.'s of X_n's and F is the d.f. of X.

定理 36 (Convergence in pr. $\Rightarrow$ Convergence in dist.)

Let (F_n), F be the d.f.'s of the r.v.'s (X_n), X. If $X_n \xrightarrow{\mathbb{P}} X$, then $F_n \xrightarrow{v} F$. More briefly stated, convergence in pr. implies convergence in dist.

- In general, we have

$$\begin{array}{l} X_n \to X \text{ a.e.} \\ X_n \xrightarrow{p} X \end{array} \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

No other implications hold in general.

Convergence in pr. $\Rightarrow$ Convergence in dist.

Proof. 若 $X_n \xrightarrow{\mathbb{P}} X$, 则对任一 $f \in C_K$, 有 $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. 首先证明这一断言.

◀ $\forall\ f \in C \supseteq C_K$, $\forall\ \varepsilon > 0$, $\forall\ k > 0$, 我们有

$$\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leqslant \mathbb{P}(|f(X_n) - f(X)| > \varepsilon, |X| \leqslant k) + \mathbb{P}(|X| > k).$$

又因为 f 在 $[-k, k]$ 上一致连续, 则对上述 $\varepsilon > 0$, 存在 $\delta > 0$, s.t. $\forall\ x, y \in [-k, k]$, 只要 $|x - y| < \delta$, 就有 $|f(x) - f(y)| \leqslant \varepsilon$. 从而

$$\{|f(X_n) - f(X)| > \varepsilon, |X| \leqslant k\} \subseteq \{|X_n - X| > \delta, |X| \leqslant k\} \subseteq \{|X_n - X| > \delta\}.$$

故

$$\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leqslant \mathbb{P}(|X_n - X| > \delta) + \mathbb{P}(|X| > k). \tag{22}$$

现在, $\forall\ \varepsilon' > 0$, 可取充分大的 k, s.t. $\mathbb{P}(|X| > k) < \varepsilon'$. 现在, 在 (22) 式中令 $n \to \infty$, 由 $\varepsilon' > 0$ 的任意性便有 $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. ▶

又因为 $f \in C_K$, 于是 f 有界. 由定理 6, $f(X_n) \xrightarrow{p} f(X)$. 特取 $p = 1$, 便有

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X).$$

再由淡收敛的等价定义, 便知上式是 (17) 式的等价定义, 故 $\mu_n \xrightarrow{v} \mu$. 定理得证. ■

The Sum of Sequences Converging in dist.

- Convergence of r.v.'s in dist. does not have the usual properties associated with convergence.
- For instance, if $X_n \to X$ in dist. and $Y_n \to Y$ in dist., it does not follow by any means that $X_n + Y_n$ will converge in dist. to $X + Y$.
- *But if X_n and Y_n are independent, then the preceding assertion is indeed true as a property of the convergence of convolutions of distributions*. (see Chapter 6)
- However, in the simple situation of the next theorem no independence assumption is needed.

定理 37 (The Sum of Sequences Converging in dist.)

If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} 0$, then

i. $X_n + Y_n \xrightarrow{d} X$;

ii. $X_n Y_n \xrightarrow{d} 0$.

推论 38 (Corollary: The Sum of Sequences Converging in dist.)

If $X_n \xrightarrow{d} X$, $\alpha_n \xrightarrow{d} a$, $\beta_n \xrightarrow{d} b$, where a and b are constants, then $\alpha_n X_n + \beta_n \xrightarrow{d} aX + b$.

PROOF. (i). $\forall f \in C_K$, 不妨设 $|f| \leqslant M$, 因为 f 一致连续, 则 $\forall \varepsilon > 0$, $\exists \delta > 0$, s.t. $\forall |x-y| \leqslant \delta$, 有 $|f(x)-f(y)| \leqslant \varepsilon$. 则

$$\begin{aligned}
&\mathbb{E}\{|f(X_n+Y_n)-f(X_n)|\} \\
&\quad \leqslant \varepsilon \mathbb{P}(|f(X_n+Y_n)-f(X_n)| \leqslant \varepsilon) + 2M\mathbb{P}(|f(X_n+Y_n)-f(X_n)| > \varepsilon) \\
&\quad \leqslant \varepsilon + 2M\mathbb{P}(|Y_n| > \delta) \to 0, \quad (n \to \infty).
\end{aligned} \tag{23}$$

于是 $\lim_{n\to\infty} \mathbb{E}f(X_n+Y_n) = \lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$. 上式第一个等式成立是因为 (23) 式, 第二个不等式成立是因为 $X_n \xrightarrow{d} X$ (定理 28). 再由定理 28, 由上式可知 $X_n + Y_n \xrightarrow{d} X$.

(ii). $\forall \varepsilon > 0$, 取充分大的 A_0 使得 $\pm A_0$ 均为 X 的分布函数的连续点, 且

$$\lim_{n\to\infty} \mathbb{P}(|X_n| > A_0) = \mathbb{P}(|X| > A_0) < \varepsilon.$$

自行验证紫色式. 由上式可知 $\mathbb{P}\{|X_n| > A_0\} < \varepsilon$ 对 $n \geqslant n_0(\varepsilon)$ 成立. 故当 $n \geqslant n_0(\varepsilon)$ 时,

$$\mathbb{P}\{|X_nY_n| > \varepsilon\} \leqslant \mathbb{P}\{|X_n| > A_0\} + \mathbb{P}\left\{|Y_n| > \frac{\varepsilon}{A_0}\right\} \leqslant \varepsilon + \mathbb{P}\left\{|Y_n| > \frac{\varepsilon}{A_0}\right\} \to 0 \ (n \to \infty).$$

于是 $X_nY_n \xrightarrow{\mathbb{P}} 0$, 由定理 36 可知 $X_nY_n \xrightarrow{d} 0$. ▸ Back to Content. ■

Section 5

Uniform Integrability; Convergence of Moments

- Sometimes we need to establish that moments of some sequence (X_n), or at least some lower-order moments, converge to moments of X.
- The functions $|x|^r$, $r > 0$, is in C but not in C_B, hence Theorem 30 does not apply to it. Indeed, we have seen in Example 9 that even convergence a.e. does not imply convergence of any moment of order $r > 0$.
- An extra condition that ensures convergence of appropriate moments is *uniform integrability*. On the other hand, convergence of moments is useful for proving convergence in distribution itself.
- This is related to a classic and extremely deep problem in analysis and probability called the *moment problem*.

Convergence in L^r $\Rightarrow$ Convergence of Moments

It is useful to have conditions to ensure the convergence of moments when X_n converges a.e.

定理 39 (Convergence in L^r $\Rightarrow$ Convergence of Moments)

If $X_n \to X$ a.e., then <u>for every $r > 0$</u>:

$$\mathbb{E}(|X|^r) \leqslant \varliminf_{n\to\infty} \mathbb{E}(|X_n|^r). \quad \textit{Fatou's Lemma} \tag{24}$$

If $X_n \xrightarrow{r} X$, and $X \in L^r$, then $\mathbb{E}(|X_n|^r) \to \mathbb{E}(|X|^r)$.

PROOF. 若 $X_n \xrightarrow{r} X \in L^r$ ($r > 1$), 由 Minkowski 不等式, 有

$$\mathbb{E}(|X_n|^r)^{\frac{1}{r}} - \mathbb{E}(|X_n - X|^r)^{\frac{1}{r}} \leqslant \mathbb{E}(|X|^r)^{\frac{1}{r}} \leqslant \mathbb{E}(|X_n|^r)^{\frac{1}{r}} + \mathbb{E}(|X_n - X|^r)^{\frac{1}{r}},$$

其中 $X = X_n - (X_n - X)$. 由 $X_n \xrightarrow{r} X$, 令 $n \to \infty$ 便有 $\mathbb{E}(|X_n|^r) \to \mathbb{E}(|X|^r)$.

若 $X_n \xrightarrow{r} X \in L^r$ ($r \leqslant 1$), 由 $|x+y|^r \leqslant |x|^r + |y|^r$ 可知

$$\mathbb{E}(|X_n|^r) - \mathbb{E}(|X_n - X|^r) \leqslant \mathbb{E}(|X|^r) \leqslant \mathbb{E}(|X_n|^r) + \mathbb{E}(|X_n - X|^r),$$

类似地便可完成定理证明. ∎

- The next theorem should be compared with Theorem 6.

定理 40 (Relations Between Convergence in dist. and Convergence of Moments)

If (X_n) *converges in dist. to* X*, and for some* $p>0$*,* $\sup_n \mathbb{E}(|X_n|^p)=M<\infty$*, then for each* $r<p$*:*

$$\lim_{n\to\infty}\mathbb{E}(|X_n|^r)=\mathbb{E}(|X|^r)<\infty. \tag{25}$$

If r *is a* <u>*positive integer*</u>*, then we may replace* $|X_n|^r$ *and* $|X|^r$ *above by* X_n^r *and* X^r*.*

PROOF. **我们只讨论 $r\in\mathbb{N}$ 的情形**. 设 F_n 与 F 是 X_n 与 X 的分布函数, 由 $X_n\xrightarrow{d}X$ 可知 $F_n\xrightarrow{v}F$. 任取 $A>0$, 定义 $f_A\in C_B$ 如下:

$$f_A(x)=\begin{cases} x^r, & \text{if } |x|\leqslant A;\\ A^r, & \text{if } x>A;\\ (-A)^r, & \text{if } x<-A.\end{cases}$$

于是由定理 33 可知 $f_A\in L$ 且 $f_A\in U$, 于是

$$\int_{-\infty}^{\infty}f_A(x)\,dF(x)\leqslant\varliminf_{n\to\infty}\int_{-\infty}^{\infty}f_A(x)\,dF_n(x)\leqslant\varlimsup_{n\to\infty}\int_{-\infty}^{\infty}f_A(x)\,dF_n(x)\leqslant\int_{-\infty}^{\infty}f_A(x)\,dF(x).$$

这说明 $\int_{-\infty}^{\infty} f_A(x)\,dF_n(x) \to \int_{-\infty}^{\infty} f_A(x)\,dF(x)$. 接下来, 我们还有

$$\begin{aligned}\int_{-\infty}^{\infty} |f_A(x) - x^r|\,dF_n(x) &\leqslant \int_{|x|>A} |x|^r\,dF_n(x) = \int_{|X_n|>A} |X_n|^r\,d\mathbb{P} \\ &\leqslant \frac{1}{A^{p-r}} \int_{\Omega} |X_n|^p\,d\mathbb{P} \leqslant \frac{M}{A^{p-r}}. \quad (\clubsuit)\end{aligned}$$

其中上式第一个不等号成立只需注意到 $r \in \mathbb{N}$, 再分类讨论即可. 第二个不等号成立:

$$\int_{\Omega} |X_n|^p\,d\mathbb{P} \geqslant \int_{|X_n|>A} |X_n|^{p-r}|X_n|^r\,d\mathbb{P} \geqslant A^{p-r} \int_{|X_n|>A} |X_n|^r\,d\mathbb{P}.$$

另外, $(\clubsuit)$ 式最后一项不依赖于 $n \in \mathbb{N}$, 于是当 $A \to \infty$ 时, $\int_{-\infty}^{\infty} f_A\,dF_n$ 关于 $n \in \mathbb{N}$ 一致收敛于 $\int_{-\infty}^{\infty} x^r\,dF$. 现在由重极限与累次极限的关系, 有

$$\begin{aligned}\int_{-\infty}^{\infty} x^r\,dF &= \lim_{A\to\infty} \int_{-\infty}^{\infty} f_A\,dF = \lim_{A\to\infty} \lim_{n\to\infty} \int_{-\infty}^{\infty} f_A\,dF_n \\ &= \lim_{n\to\infty} \lim_{A\to\infty} \int_{-\infty}^{\infty} f_A\,dF_n = \lim_{n\to\infty} \int_{-\infty}^{\infty} x^r\,dF_n.\end{aligned}$$

上式便说明 $\mathbb{E}(|X_n|^r) \to \mathbb{E}(|X|^r)$. 命题得证. ■

Uniform Integrability

- We now introduce the concept of *uniform integrability*, which is of basic importance in probability theory. It is also an essential hypothesis in certain convergence questions arising in the theory of martingales (Chapter 9).

定义 41 (Uniform Integrability)

A family of r.v.'s $\{X_t, t \in T\}$, where T is an arbitrary index set, is said to be **uniformly integrable** iff

$$\lim_{A\to\infty} \sup_{t\in T} \left\{ \int_{|X_t|>A} |X_t| \, d\mathbb{P} \right\} = 0. \tag{26}$$

定理 42 (An equivalent Definition of Uniform Integrability)

The family (X_t) is uniformly integrable iff the following two conditions are satisfied:

i. L^1 BOUNDED. $\mathbb{E}(|X_t|)$ *is bounded in* $t \in T$*;*

ii. ABSOLUTELY CONTINUOUS. *For every* $\varepsilon > 0$*, there exists* $\delta > 0$ *such that for any* $E \in \mathscr{F}$*:*

$$\mathbb{P}(E) < \delta \quad \Longrightarrow \quad \int_E |X_t| \, d\mathbb{P} < \varepsilon \text{ for every } t \in T.$$

PROOF. (⇒). 若 (26) 式成立, 于是可取充分大的 A, s.t. $\sup_{t\in T}\int_{|X_t|>A}|X_t|\,d\mathbb{P}<1$. 于是

$$\mathbb{E}(|X_t|)=\left(\int_{|X_t|\leqslant A}+\int_{|X_t|>A}\right)|X_t|\,d\mathbb{P}\leqslant A\mathbb{P}(\Omega)+1=A+1.$$

于是 $\sup_{t\in T}\mathbb{E}(|X_t|)\leqslant A+1$. 另外, 任取 $E\in\mathscr{F}$, 令 $E_t=\{\omega:|X_t(\omega)|>A\}$. 由中值定理,

$$\int_E|X_t|\,d\mathbb{P}=\left(\int_{E\cap E_t}+\int_{E\setminus E_t}\right)|X_t|\,d\mathbb{P}\leqslant\int_{E_t}|X_t|\,d\mathbb{P}+A\mathbb{P}(E).$$

现任取 $\varepsilon>0$, 由 (26) 式, 存在 $A=A(\varepsilon)$ 使得上式积分小于 $\varepsilon/2$ ($\forall\ t\in T$). 现在只需令 $\delta=\varepsilon/2A$ 便有 (ii) 成立.

(⇐). 若条件 (i, ii) 成立, 由 Chebyshev 不等式, $\forall\ t\in T$, 有

$$\mathbb{P}(|X_t|>A)\leqslant\frac{\mathbb{E}(|X_t|)}{A}\leqslant\frac{M}{A}.$$

其中 $\sup_{t\in T}\mathbb{E}(|X_t|)\leqslant M<\infty$. 现若 $A>M/\delta$, 则 $\mathbb{P}(E_t)<\delta$, 由 (ii), 便有 $\int_{E_t}|X_t|\,d\mathbb{P}<\varepsilon$. 由 ε 的任意性便可知 (26) 式成立. 定理成立. ∎

定理 43 (Convergence in pr. + u.i. ⇒ Convergence in L^r)

Let $0<r<\infty$, $X_n\in L^r$, and $X_n\xrightarrow{\mathbb{P}}X$. Then the following three propositions are equivalent:

i. $\{|X_n|^r\}$ *is uniformly integrable;*
ii. $X_n\xrightarrow{r}X$;
iii. $\mathbb{E}(|X_n|^r)\to\mathbb{E}(|X|^r)<\infty$.

Proof. (i) ⇒ (ii). 设 (i) 成立, 我们分三步证明命题 (ii) 成立.

Step 1. **证明** $\mathbb{E}(|X|^r)<\infty$. 因为 $X_n\xrightarrow{\mathbb{P}}X$, 由定理 15, 存在子列 X_{n_k}, s.t. $X_{n_k}\to X$ a.e. 再由定理 39 可知

$$\mathbb{E}(|X|^r)=\mathbb{E}\Big[\lim_{k\to\infty}|X_{n_k}|^r\Big]=\mathbb{E}\Big[\varliminf_{k\to\infty}|X_{n_k}|^r\Big]\leqslant\varliminf_{k\to\infty}\mathbb{E}(|X_{n_k}|^r)\leqslant\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|^r)<\infty,$$

其中上式最后一个不等式利用了定理 42 中的条件 (i). 这就说明 $X\in L^r$.

Step 2. **证明 $\{|X_n - X|^r\}$ 一致可积**. $\forall\ r > 0$, 我们总有如下不等式:

$$
\begin{aligned}
|X_n \pm X|^r &\leqslant 2^{r-1}(|X_n|^r + |X|^r), \qquad \text{if } r \geqslant 1; \\
&\leqslant |X_n|^r + |X|^r, \qquad\qquad \text{if } 0 < r < 1.
\end{aligned} \tag{27}
$$

又因为 $\{|X_n|^r\}$ 一致可积, 且 $\{|X_n - X|^r\}$ 被 $\{|X_n|^r\}$ 与 $|X|^r$ 控制, 于是 $\{|X_n - X|^r\}$ 一致可积.

Step 3. **证明 $X_n \xrightarrow{r} X$**. 现在任取 $\varepsilon > 0$, 我们有

$$
\begin{aligned}
\int_\Omega |X_n - X|^r \, d\mathbb{P} &= \int_{|X_n - X| > \varepsilon} |X_n - X|^r \, d\mathbb{P} + \int_{|X_n - X| \leqslant \varepsilon} |X_n - X|^r \, d\mathbb{P} \\
&\leqslant \int_{|X_n - X| > \varepsilon} |X_n - X|^r \, d\mathbb{P} + \varepsilon^r \, \mathbb{P}(|X_n - X| \leqslant \varepsilon) \\
&\leqslant \int_{|X_n - X| > \varepsilon} |X_n - X|^r \, d\mathbb{P} + \varepsilon^r.
\end{aligned}
$$

因为 $\{|X_n - X|^r\}$ 一致可积, 于是由定理 42 (ii) 中的绝对连续性可知:

$$
X_n \xrightarrow{\mathbb{P}} X \ \Rightarrow\ \mathbb{P}(|X_n - X| > \varepsilon) \to 0 \quad \Longrightarrow \quad \int_{|X_n - X| > \varepsilon} |X_n - X|^r \, d\mathbb{P} \to 0.
$$

最后再由 ε 的任意性可知 $X_n \xrightarrow{r} X$. (ii) $\Rightarrow$ (iii). 定理 39 已证.

(iii) $\Rightarrow$ (i). 任取 $A>0$, 定义非负连续函数 $f_A \in C_K$ 如下:

$$\begin{aligned} f_A(x) &= |x|^r, \quad \text{if } |x|^r \leqslant A; \\ &\leqslant |x|^r, \quad \text{if } A<|x|^r \leqslant A+1; \\ &= \ 0 \ \ , \quad \text{if } |x|^r > A+1. \end{aligned}$$

注意到 $0 \leqslant |x|^r \mathbf{1}_{|x|^r \leqslant A} \leqslant f_A(x) \leqslant |x|^r \mathbf{1}_{|x|^r \leqslant A+1}$, 于是

$$\varliminf_{n\to\infty} \int_{|X_n|^r \leqslant A+1} |X_n|^r \, d\mathbb{P} \geqslant \lim_{n\to\infty} \mathbb{E} f_A(X_n) = \mathbb{E} f_A(X) \geqslant \int_{|X|^r \leqslant A} |X|^r \, d\mathbb{P}.$$

其中 $\lim\limits_{n\to\infty} \mathbb{E} f_A(X_n) = \mathbb{E} f_A(X)$ 成立利用了定理 36 与 28. 再由命题 (iii) 的假设:

$$\begin{aligned} \varlimsup_{n\to\infty} \int_{|X_n|^r > A+1} |X_n|^r \, d\mathbb{P} &= \varlimsup_{n\to\infty} \left[\mathbb{E}(|X_n|^r) - \int_{|X_n|^r \leqslant A+1} |X_n|^r \, d\mathbb{P} \right] \\ &\leqslant \mathbb{E}(|X|^r) - \varliminf_{n\to\infty} \int_{|X_n|^r \leqslant A+1} |X_n|^r \, d\mathbb{P} \\ &\leqslant \mathbb{E}(|X|^r) - \int_{|X|^r \leqslant A} |X|^r \, d\mathbb{P} \\ &= \int_{|X|^r > A} |X|^r \, d\mathbb{P}. \end{aligned}$$

注意到 $\int_{|X|^r>A} |X|^r \, d\mathbb{P}$ 不依赖于 $n \in \mathbb{N}$ 且

$$\int_{|X|^r>A} |X|^r \, d\mathbb{P} \to 0 \quad (A \to \infty).$$

这意味着: $\forall\ \varepsilon > 0$, $\exists\ A_0 = A_0(\varepsilon)$ 与 $n_0 = n_0(A_0)$ 使得当 $A > A_0$ 时有

$$\sup_{n>n_0} \int_{|X_n|^r>A+1} |X_n|^r \, d\mathbb{P} < \varepsilon.$$

又因为 $|X_n|^r$ 可积, 于是存在 $A_1 = A_1(\varepsilon)$ 使得

$$\sup_{n\geqslant 1} \int_{|X_n|^r>A+1} |X_n|^r \, d\mathbb{P} < \varepsilon, \qquad A > \max\{A_0, A_1\}.$$

由一致可积性的定义可知命题 (i) 成立. 定理得证. ■

The Moment Problem

- In the remainder of this section *the term "moment" will be restricted to a moment of positive integer order*.
- In fact, on $(\mathscr{U}, \mathscr{B})$ any p.m. or equivalent its d.f. is uniquely determined by its moments of all orders.
- The corresponding result is **false** in $(\mathbb{R}, \mathscr{B}^1)$ and a further condition on the moments is required to ensure uniqueness.
- When a given sequence of numbers $\{m_r, r \geqslant 1\}$ uniquely determines a d.f. F such that

$$m_r = \int_{-\infty}^{\infty} x^r \, dF(x), \tag{28}$$

  we say that "the moment problem is determinate" for the sequence.
- In this section we are not concerned with the sufficient condition of the moment problem, but instead study some useful conclusions when the moment problem is determinate.

定理 44 (Method of Moments)

Suppose there is a unique d.f. F with the moments $\{m^{(r)}, r \geqslant 1\}$, all finite. Suppose that (F_n) is a sequence of d.f.'s, each of which has all its moments finite:

$$m_n^{(r)} = \int_{-\infty}^{\infty} x^r \, dF_n(x).$$

Finally, suppose that for every $r \geqslant 1$:

$$\lim_{n\to\infty} m_n^{(r)} = m^{(r)}. \tag{29}$$

Then $F_n \xrightarrow{v} F$.

Proof. 设 μ_n 是 F_n 所对应的概率测度. 由 Helly 选择定理 25, 存在子列 $\{m_{n_k}\}$ 弱收敛. 不妨设 $\mu_{n_k} \xrightarrow{v} \mu$. 我们来证明 μ 是概率测度且其分布函数为 F. 由 Chebyshev 不等式, $\forall\ A > 0$:

$$\mu(-A, A) \geqslant 1 - \frac{m_{n_k}^{(2)}}{A^2}, \quad 其中\ \varphi(u) = u^2. \tag{30}$$

又因为 $m_{n_k}^{(2)} \to m^{(2)} < \infty$, 令 $A \to \infty$, 可知 (30) 式右端关于 k 一致收敛于 1. 再设 μ_{n_k} 于稠密集 D 上淡收敛于 μ, 若 $\pm A \in D$, 再令 $A \to \infty$, 有

$$\begin{aligned}\mu(\mathbb{R}) &= \lim_{A\to\infty} \mu(-A, A) = \lim_{A\to\infty} \lim_{k\to\infty} \mu_{n_k}(-A, A) \\ &= \lim_{k\to\infty} \lim_{A\to\infty} \mu_{n_k}(-A, A) = \lim_{k\to\infty} \mu_{n_k}(\mathbb{R}) = 1.\end{aligned}$$

现在, $\forall\ r \in \mathbb{N}$, 取 p 为大于 r 的第一个偶数, 于是有

$$\int_{-\infty}^{\infty} x^p \, d\mu_{n_k} = m_{n_k}^{(p)} \to m^{(p)},$$

由定理假设, $\sup\limits_k \{m_{n_k}^{(p)}\} < \infty$, 于是由定理 40 与 $\mu_{n_k} \to \mu$ 可知:

$$\int_{-\infty}^{\infty} x^r \, d\mu_{n_k} \to \int_{-\infty}^{\infty} x^r \, d\mu.$$

同时上式左端还满足 $m_{n_k}^{(r)} \to m^{(r)}$, 于是由定理唯一性的假设可知 μ 就是分布函数 F 所对应的概率测度. 现在我们就证明了 (μ_n) (或 (F_n)) 的任一淡收敛子列有相同的淡极限 μ (或 F), 再次利用定理 26, 便可知 $F_n \xrightarrow{v} F$. 定理得证. Back to Content. ■

Bibliography

- Chung, Kai Lai. *A Course in Probability Theory*, Third Edition. Academic Press, 2001.
- Durrett, Rick. *Probability: Theory and Examples*, 5th edition. Cambridge university press, 2019.
- Billingsley, Patrick. *Probability and Measure*, Anniversary edition. John Wiley & Sons, Inc. 2012.
- Varadhan, SR Srinivasa. *Probability theory*. No. 7. American Mathematical Soc., 2001.
- Williams, David. *Probability with martingales*. Cambridge university press, 1991.
- 任佳刚 巫静. 测度与概率教程. 科学出版社, 2018.
- 严士健 刘秀芳. 测度与概率, 第二版. 北京师范大学出版社, 2023.

FINANCIAL TECHNOLOGY

The Thrust nurtures creative minds to explore frontiers in Financial Technology, to shape the future financial industry that promotes social-financial health.

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

社会枢纽
SOCIETY HUB

Society Hub • Financial Technology Thrust

Chapter 5 Law of Large Numbers. Random Series

Probability Theory and Stochastic Processes

WANG Zhuoran

16320216@stu.njau.edu.cn

香港科技大學 (廣州)

FTEC5031

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Content

Simple Limit Theorems (Chebyshev)

Weak Law of Large Numbers (Khintchine)

Convergence of Series
- Kolmogorov's Zero-One Law
- Kolmogorov's Maximal Inequalities
- Kolmogorov's Three Series Theorem

Strong Law of Large Numbers
- Kronecker's Lemma
- Independent Cases
- I.I.D. Cases (Kolmogorov)

* Applications
- Empiric Distributions. Glivenko-Cantelli Theorem
- Renewal Theory

Section 1

Simple Limit Theorems (Chebyshev)

- This section begins with an introduction to the object of study of the law of large numbers: convergence of partial sums of r.v's.
- The concepts of uncorrelated r.v.'s and orthogonal r.v.'s will be introduced in this section.
- Finally, the simplest law of large numbers (a sequence of uncorrelated r.v.'s whose second moments are uniformly bounded) due to Chebyshev and Rajchman is introduced, and a common technique, the subsequence method, is introduced.

Introduction

- The various concepts of Chapter 4 will be applied to the so-called "law of large numbers" — a famous name in the theory of probability.
- This has to do with partial sums

$$\boxed{S_n = \sum_{i=1}^{n} X_i}$$

  of a sequence of r.v.'s. In the most classical formulation, the "weak" or the "strong" law of large numbers is said to hold for the sequence according as

$$\frac{S_n - \mathbb{E}(S_n)}{n} \to 0 \tag{1}$$

  in pr. or a.e. This, of course, presuppose the finiteness of $\mathbb{E}(S_n)$.
- A natural generalization is as follows:

$$\frac{S_n - a_n}{b_n} \to 0,$$

  where (a_n) is a sequence of real numbers and (b_n) a sequence of positive numbers tending to infinity.

Simplest Case of Law of Large Numbers

- Recall that

$$\begin{array}{ccccc} X_n \xrightarrow{r} X & \Rightarrow & X_n \xrightarrow{\mathbb{P}} X & \Rightarrow & \exists\ X_{n_k} \to X \text{ a.e.} \\ & & \Uparrow & & \\ & & X_n \to X \text{ a.e.} & & \end{array}$$

- If Z_n is any sequence of r.v.'s, then $\mathbb{E}(Z_n^2) = \mathbb{E}(|Z_n - 0|^2) \to 0$ implies that $Z_n \xrightarrow{\mathbb{P}} 0$ and $Z_{n_k} \to 0$ a.e. for a subsequence (n_k).
- Applied to $Z_n = S_n/n$, the first assertion becomes

$$\mathbb{E}(S_n^2) = o(n^2) \text{ i.e. } \lim_{n\to\infty} \frac{\mathbb{E}(S_n^2)}{n^2} = 0 \quad \Longrightarrow \quad \frac{S_n}{n} \xrightarrow{\mathbb{P}} 0. \tag{2}$$

- We can calculate $\mathbb{E}(S_n^2)$ more explicitly as follows:

$$\mathbb{E}(S_n^2) = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}(X_i^2) + 2 \sum_{1\leqslant i<j\leqslant n} \mathbb{E}(X_i X_j). \tag{3}$$

  There are n^2 terms above, so that even if all of them are bounded by a fixed constant, only $\mathbb{E}(S_n^2) = O(n^2)$ will result, which falls critically short of the hypothesis in (2).
- *The idea then is to introduce certain assumptions to cause enough cancellation among the "mixed terms" in (3).*

Definitions of Uncorrelation and Orthogonality

- A salient feature of probability theory and its applications is that such assumptions are not only permissible but realistic. We begin with the simplest of its kind.

定义 1 (Definitions of Uncorrelation and Orthogonality)

Two r.v.'s X and Y are said to be **uncorrelated** iff both have finite second moments and

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{4}$$

They are said to be **orthogonal** iff (4) is replaced by

$$\mathbb{E}(XY) = 0. \tag{5}$$

The r.v.'s of any family are said to be uncorrelated (or orthogonal) iff every two of them are.

- Eq. (4) is equivalent to $\mathrm{Cov}(X,Y) := \mathbb{E}\{(X-\mathbb{E}X)(Y-\mathbb{E}Y)\} = 0$ (covariance), which reduces to (5) when $\mathbb{E}(X) = \mathbb{E}(Y) = 0$.
- The requirement of finite second moments seems unnecessary, but it does ensure the finiteness of $\mathbb{E}(XY)$ (Cauchy-Schwarz's inequality) as well as that of $\mathbb{E}(X)$ and $\mathbb{E}(Y)$, and without it the definitions are hardly useful.
- It is obvious that *pairwise independence implies uncorrelatedness*, provided second moments are finite.

Uncorrelated Case of Law of Large Numbers

定理 2 (Chebyshev: Uncorrelated Case of Law of Large Numbers)

If the X_i's are uncorrelated and their second moments have a common bound, then (1) is true in L^2 and hence also in pr.

PROOF. 设 (X_n) 是一列不相关随机事件列, 则 $\{X_n - \mathbb{E}(X_n)\}$ 是正交的. 于是我们有

$$\sigma^2(S_n) = \mathbb{E}[(S_n - \mathbb{E}S_n)^2] = \mathbb{E}(S_n^2) - \mathbb{E}^2(S_n) = \sum_{i=1}^{n} \mathbb{E}(X_i^2) - \mathbb{E}^2(X_i) = \sum_{i=1}^{n} \sigma^2(X_i). \tag{6}$$

反之, 若上式成立, 则可知随机事件列 (X_n) 中元素两两不相关.

因随机事件列 (X_n) 的二阶矩上有界, 并且 (6) 式右端只有 n 项, 于是我们有

$$\sigma^2(S_n) = O(n) = o(n^2).$$

于是

$$\mathbb{E}\left[\frac{S_n - \mathbb{E}S_n}{n}\right]^2 = \frac{\mathbb{E}(S_n - \mathbb{E}S_n)^2}{n^2} \to 0 \quad (n \to \infty).$$

从而 $\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{L^2} 0$, 进而 $\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0$. 定理得证. ■

Uncorrelated Case of Law of Large Numbers

- Theorem 2 is due to Chebyshev, who invented his famous inequalities for its proof.
- The next result, due to Rajchman (1932), strengths the conclusion by proving convergence a.e. This result is interesting by virtue of its simplicity, and serves well to introduce an important method, that of taking subsequence.
- The method used in the proof is called "subsequence method", very useful in other contexts as well. *It first proves the result for a subsequence and then fill in the gap*.

定理 3 (Rajchman: Uncorrelated Case of Law of Large Numbers)

Under the same hypothesis as in Theorem 2, (1) holds also a.e.

PROOF. 不失一般性, 设 $\mathbb{E}(X_i)=0$, 否则只需考虑 $X_i-\mathbb{E}(X_i)$. 此时随机变量序列 (X_n) 是正交的. 由 (6) 式不妨设

$$\mathbb{E}(S_n^2)=\sigma^2(S_n)\leqslant Mn, \quad \text{其中} \quad \sup_n \sigma^2(S_n)=\sup_n \mathbb{E}(X_n^2)\leqslant M<\infty.$$

则由 Chebyshev 不等式, $\forall\ \varepsilon>0$, 有

$$\mathbb{P}(|S_n|>n\varepsilon)\leqslant \frac{\mathbb{E}(S_n^2)}{(n\varepsilon)^2}\leqslant \frac{Mn}{n^2\varepsilon^2}=\frac{M}{n\varepsilon^2}.$$

对上式左右两边求和:

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n| > n\varepsilon) \leqslant \sum_{n=1}^{\infty} \frac{M}{n\varepsilon^2},$$

但是上式右端级数发散. 于是我们对子列 (n^2) 求和, 于是有

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_{n^2}| > n^2\varepsilon) \leqslant \sum_{n=1}^{\infty} \frac{M}{n^2\varepsilon^2}.$$

上式右端级数收敛, 于是有 Borel-Cantelli 引理 (Chapter 4 Theorem 13 and 14):

$$\mathbb{P}(|S_{n^2}| > n^2\varepsilon \text{ i.o.}) = 0 \iff \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon \text{ i.o.}\right) = 0 \iff \frac{S_{n^2}}{n^2} \to 0 \text{ a.e.}$$

于是我们对子列 (n^2) 证明了定理结论, 为了最终证明定理, 我们只需证明 S_k 与其最近的 S_{n^2} 之间没有太大的差异即可. $\forall\ n \geqslant 1$, 令

$$D_n = \max_{n^2 \leqslant k < (n+1)^2} |S_k - S_{n^2}|.$$

则

$$\mathbb{E}(D_n^2) = \mathbb{E}\Big[\max_{n^2 \leqslant k < (n+1)^2} |S_k - S_{n^2}|\Big]^2 \leqslant \sum_{k=n^2}^{(n+1)^2-1} \mathbb{E}\big(|S_k - S_{n^2}|^2\big)$$

$$= \sum_{k=n^2}^{(n+1)^2-1} \sum_{j=n^2+1}^{k} \sigma^2(X_j) \leqslant \sum_{k=n^2}^{(n+1)^2-1} \sum_{j=n^2+1}^{(n+1)^2-1} \sigma^2 M \leqslant 4n^2 M.$$

于是再由 Chebyshev 不等式可知

$$\mathbb{P}(D_n > n^2\varepsilon) \leqslant \frac{4M}{\varepsilon^2 n^2}.$$

于是类似于刚才的讨论便有 $D_n/n^2 \to 0$ a.e. 现在, $\forall\ n^2 \leqslant k < (n+1)^2$, 有

$$0 \leqslant \frac{|S_k|}{k} = \frac{|S_{n^2} + (S_k - S_{n^2})|}{k} \leqslant \frac{|S_{n^2}| + |S_k - S_{n^2}|}{k} \leqslant \frac{|S_{n^2}| + |D_n|}{n^2} \to 0 \text{ a.e.}$$

于是 (1) 式成立, 定理得证. Back to Content. ■

Remark

The hypotheses of Theorem 2 and 3 are certainly satisfied for a sequence of *independent r.v.'s that are uniformly bounded* or that are *identically distributed with a finite second moment*.

Section 2

Weak Law of Large Numbers (Khintchine)

- This section discusses the weak law of large numbers, i.e., Khintchine's law of large numbers.
- In order to prove Khintchine's law of large numbers, we first need to discuss the definition of the equivalent sequence and its properties.
- Finally, we generalize the condition of *pairwise* independence to the case of *total* independence, i.e., Kolmogorov-Feller's law of large numbers.

Equivalent Sequences

- The law of large numbers in the form (1) involves only the first moment, but so far we have operated with the second.
- In order to drop any assumption on the second moment, we need a new device, that of "equivalent sequences", due to Khintchine (辛钦, 1894 – 1959).

定义 4 (Equivalent Sequences)

Two sequences of r.v.'s (X_n) and (Y_n) are said to be **equivalent** iff

$$\sum_{n=1}^{\infty} \mathbb{P}\{X_n \neq Y_n\} < \infty. \tag{7}$$

- In practice, an equivalent sequence is obtained by "truncating (截断)" in various ways, as we shall see presently.

定理 5 (Properties of Equivalent Sequences)

If (X_n) *and* (Y_n) *are equivalent, then*

i). $\sum_{n=1}^{\infty}(X_n - Y_n)$ *converges a.e.;*

ii). *If* $a_n \uparrow \infty$, *then* $\frac{1}{a_n}\sum_{i=1}^{n}(X_i - Y_i) \to 0$ *a.e.*

PROOF. 若序列 (X_n) 与 (Y_n) 等价, 由定义, (7) 式成立. 再由 Borel-Cantelli 引理, 我们有 $\mathbb{P}\{X_n \neq Y_n \text{ i.o.}\} = 0$. 这就说明存在零测度集 N 满足: $\mathbb{P}(N) = 0$; 若 $\omega \in \Omega \setminus N$, 则存在 $n_0(\omega)$ 使得

$$\forall\ n \geqslant n_0(\omega) \quad \Rightarrow \quad X_n(\omega) = Y_n(\omega).$$

于是对于上述 $\omega \in \Omega \setminus N$, 数列 $\{X_n(\omega)\}$ 与 $\{Y_n(\omega)\}$ 仅在有限项处不同, 于是级数

$$\sum_{n=1}^{\infty}[X_n(\omega) - Y_n(\omega)]$$

在某一项之后由零组成. 这便可推的定理结论成立. ∎

Properties of Equivalent Sequences

推论 6 (Properties of Equivalent Sequences)

Suppose that (X_n) *and* (Y_n) *are equivalent, and* $a_n \uparrow \infty$*. Then with probability one (a.e.):*

i). $\sum_{n=1}^{\infty} X_n$ *or* $\frac{1}{a_n}\sum_{i=1}^{n} X_i$ *converges, diverges to* $+\infty$*,* $-\infty$*, or fluctuates in the same way as* $\sum_{n=1}^{\infty} Y_n$ *or* $\frac{1}{a_n}\sum_{i=1}^{n} Y_i$ *respectively;*

ii). *In particular, if* $\frac{1}{a_n}\sum_{i=1}^{n} X_i$ *converges to* X *in pr., then so does* $\frac{1}{a_n}\sum_{i=1}^{n} Y_i$*.*

Proof. (i). 由定理 5 的证明过程可以直接看出.

(ii). 由第四章定理 4, $a_n^{-1}\sum_{i=1}^{n}(X_i - Y_i) \to 0$ a.e. 可推得 $a_n^{-1}\sum_{i=1}^{n}(X_i - Y_i) \xrightarrow{\mathbb{P}} 0$. 从而若 $a_n^{-1}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} X$, 则

$$\frac{1}{a_n}\sum_{i=1}^{n} Y_i = \frac{1}{a_n}\sum_{i=1}^{n} X_i + \frac{1}{a_n}\sum_{i=1}^{n}(Y_i - X_i) \xrightarrow{\mathbb{P}} X + 0 = X. \quad \blacksquare$$

Khintchine's Law of Large Numbers

- The next law of large numbers is due to Khintchine.
- Under the stronger hypothesis of total independence, it will be proved again by an entirely different method in Chapter 6 (Use the tool of characteristic functions).

定理 7 (Khintchine's Law of Large Numbers 辛钦大数定律)

Let (X_n) be ***pairwise*** *independent and identically distributed (i.i.d.) r.v.'s with finite mean m. Then we have*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} m. \tag{8}$$

Proof. 因序列 (X_n) 独立同分布, 不妨设 X_n 的分布函数为 F. 而 $\mathbb{E}(X_n) = m < \infty$, 于是由第三章定理 16 可知:

$$\left|\mathbb{E}(X_1)\right| < \infty \iff \mathbb{E}(|X_1|) < \infty \iff \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) < \infty.$$

现在定义 “截断” 随机变量 Y_n:

$$Y_n := X_n \mathbf{1}_{\{|X_n| \leqslant n\}}.$$

由 Y_n 的定义可知 (X_n) 与 (Y_n) 等价: 因为 $\mathbb{P}(X_n \neq Y_n) = \mathbb{P}(|X_n| > n)$ 于是

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) < \infty.$$

现在令 $T_n = \sum_{i=1}^{n} Y_i$, 于是由推论 6, S_n/n 与 T_n/n 的敛散性相同, 且

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} m \quad \Longleftrightarrow \quad \frac{T_n}{n} \xrightarrow{\mathbb{P}} m.$$

于是我们只需证 $\frac{T_n}{n} \xrightarrow{\mathbb{P}} m$, 进而只需证:

$$\text{(i). } \mathbb{E}\left(\frac{T_n}{n}\right) \to m, \quad \text{(ii). } \sigma^2\left(\frac{T_n}{n}\right) \to 0.$$

这是因为由 Chebyshev 不等式:

$$0 \leqslant \mathbb{P}\left(\left|\frac{T_n}{n} - m\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\frac{T_n}{n} - m\right)^2\right] = \frac{1}{\varepsilon^2}\left\{\mathbb{E}\left(\frac{T_n}{n}\right)^2 - 2m\mathbb{E}\left(\frac{T_n}{n}\right) + m^2\right\}$$

$$= \frac{1}{\varepsilon^2}\left\{\sigma^2\left(\frac{T_n}{n}\right) + \mathbb{E}^2\left(\frac{T_n}{n}\right) - 2m\mathbb{E}\left(\frac{T_n}{n}\right) + m^2\right\} \xrightarrow{n\to\infty} \frac{0 + m^2 - 2m\cdot m + m^2}{\varepsilon^2} = 0.$$

Proof of (i). 注意到 $\forall\ n \geqslant 1$: $\mathbb{E}(X_n) = m$, 于是

$$0 \leqslant \left|\mathbb{E}\left(\frac{T_n}{n}\right) - m\right| = \frac{1}{n}\left|\sum_{i=1}^{n}\left(\mathbb{E}(Y_i) - \mathbb{E}(X_i)\right)\right| = \frac{1}{n}\left|\sum_{i=1}^{n}\left\{\mathbb{E}(X_i \mathbf{1}_{\{|X_i| \leqslant i\}}) - \mathbb{E}(X_i)\right\}\right|$$
$$= \frac{1}{n}\left|\sum_{i=1}^{n}\left\{-\mathbb{E}(X_i \mathbf{1}_{\{|X_i| > i\}})\right\}\right| \leqslant \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(|X_i| \mathbf{1}_{\{|X_i| > i\}}\right) \to 0.$$

首先注意到 $\mathbb{E}\left(|X_n| \mathbf{1}_{\{|X_n| > n\}}\right) \to 0$, 因为 $\{\omega : |X_n(\omega)| > n\} \downarrow \varnothing\ (n \to \infty)$, 再利用积分的绝对连续性[1]即可. 另外, 我们还利用了数学分析中的习题: $a_n \to 0 \ \Rightarrow\ n^{-1}\sum_n a_n \to 0$, 这里我们只需要取 $a_n = \mathbb{E}\left(|X_n| \mathbf{1}_{\{|X_n| > n\}}\right)$.

Proof of (ii). 由第三章定理 28 可知 (Y_n) 相互独立, 又因为 $|Y_n| \leqslant n$, 于是 Y_n 有有限二阶矩. 现在来计算 $\sigma^2(T_n)$:

$$\sigma^2(T_n) = \sum_{i=1}^{n}\sigma^2(Y_i) \leqslant \sum_{i=1}^{n}\mathbb{E}(Y_i^2) = \sum_{i=1}^{n}\int_{|x| \leqslant i} x^2\, dF(x).$$

且

$$\sum_{i=1}^{n}\int_{|x| \leqslant i} x^2\, dF(x) \leqslant \sum_{i=1}^{n} i \int_{|x| \leqslant i} |x|\, dF(x) \leqslant \frac{n(n+1)}{2}\int_{-\infty}^{\infty} |x|\, dF(x) < \infty.$$

[1] 参考 MATH4123 第五章推论 23.

于是

$$\sum_{i=1}^{n}\int_{|x|\leqslant i} x^2\,dF(x)=O(n^2),\quad \text{但}\quad \sum_{i=1}^{n}\int_{|x|\leqslant i} x^2\,dF(x)\neq o(n^2).$$

上述结论并不足够好，因为只有 $o(n^2)$ 才能利用第一节的结论. 所以我们需要使用其它形式的“截断”. 现设 $(a_n)\subseteq\mathbb{N}$ 满足 $0<a_n<n$ 且 $a_n\uparrow\infty$, $a_n=o(n)$, 于是我们有

$$\begin{aligned}
\sum_{i=1}^{n}\int_{|x|\leqslant i} x^2\,dF(x) &= \sum_{i=1}^{a_n}+\sum_{i=a_n+1}^{n}\\
&\leqslant \sum_{i=1}^{a_n} a_n\int_{|x|\leqslant a_n}|x|\,dF(x)+\sum_{i=a_n+1}^{n} a_n\int_{|x|\leqslant a_n}|x|\,dF(x)+\sum_{i=a_n+1}^{n} n\int_{a_n<|x|\leqslant n}|x|\,dF(x)\\
&\leqslant a_n\sum_{i=1}^{a_n}\int_{-\infty}^{\infty}|x|\,dF(x)+a_n\sum_{i=a_n+1}^{n}\int_{-\infty}^{\infty}|x|\,dF(x)+n\sum_{i=a_n+1}^{n}\int_{|x|>a_n}|x|\,dF(x)\\
&\leqslant na_n\int_{-\infty}^{\infty}|x|\,dF(x)+n^2\int_{|x|>a_n}|x|\,dF(x).
\end{aligned}$$

其中 $na_n\int_{-\infty}^{\infty}|x|\,dF(x)=O(na_n)=o(n^2)$, $n^2\int_{|x|>a_n}|x|\,dF(x)=n^2\,o(1)=o(n^2)$. 其中 $\int_{|x|>a_n}|x|\,dF(x)=o(1)$ 只需再次利用积分的绝对连续性. 于是综合上述所有结论便有 $\sigma^2(T_n)=o(n^2)$, (ii) 得证. ■

Classical Forms of The Weak Law of Large Numbers

定理 8 (Kolmogorov-Feller: Classical Forms of The Weak Law of Large Numbers)

Let (X_n) *be a sequence of (totally) independent r.v.'s with d.f.'s* (F_n)*; and* $S_n = \sum_{i=1}^{n} X_i$*. Let* (b_n) *be a given sequence of real numbers increasing to* $+\infty$*.*

Suppose that we have

i). $\displaystyle\sum_{i=1}^{n} \int_{|x|>b_n} dF_i(x) = o(1)$,

ii). $\displaystyle\frac{1}{b_n^2} \sum_{i=1}^{n} \int_{|x|\leqslant b_n} x^2 \, dF_i(x) = o(1)$;

then if we put

$$a_n = \sum_{i=1}^{n} \int_{|x|\leqslant b_n} x \, dF_i(x), \tag{9}$$

we have

$$\frac{S_n - a_n}{b_n} \xrightarrow{\mathbb{P}} 0. \tag{10}$$

Next suppose that the F_n*'s have the property that there exists a* $\lambda > 0$ *such that*

$$\forall\ n: \ F_n(0) \geqslant \lambda, \quad 1 - F_n(0-) \geqslant \lambda. \tag{11}$$

Then if (10) holds for the given (b_n) *and any sequence of real numbers* (a_n)*, the conditions (1) and (2) must hold.*

Classical Forms of The Weak Law of Large Numbers

- For totally independent r.v.'s, necessary and sufficient conditions for the weak law of large numbers in the most general formulation, due to Kolmogorov and Feller, are known.
- The sufficiency of Theorem 8 is easily proved, but we omit the proof of its necessity.

Remark

- Condition (11) may be written as

$$\mathbb{P}\{X_n \leqslant 0\} \geqslant \lambda, \quad \mathbb{P}\{X_n \geqslant 0\} \geqslant \lambda;$$

When $\lambda = 1/2$ this means that 0 is a *median* for each X_n. In general it ensures that none of the distribution is too far *off center*, and it is certainly satisfied if all F_n are the same.

- It is possible to replace the a_n in (9) by

$$\sum_{i=1}^{n} \int_{|x| \leqslant b_i} x \, dF_i(x)$$

and maintain (10).

Proof of Sufficiency. $\forall\ n \geqslant 1$, $\forall\ 1 \leqslant i \leqslant n$, 定义 $Y_{n,i} := X_i \mathbf{1}_{|X_i| \leqslant b_n}$, 令 $T_n = \sum_{i=1}^n Y_{n,i}$. 于是条件 (i) 可写为

$$\sum_{i=1}^{n} \mathbb{P}(Y_{n,j} \neq X_i) = o(1).$$

再利用概率测度的次可加性, 便有

$$\mathbb{P}(T_n \neq S_n) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{n} \{Y_{n,i} \neq X_i\}\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(Y_{n,i} \neq X_i) = o(1). \tag{12}$$

接下来, 条件 (ii) 也可以写为

$$\sum_{i=1}^{n} \mathbb{E}\left[\left(\frac{Y_{n,i}}{b_n}\right)^2\right] = o(1),$$

又因为 $\{Y_{n,i} : 1 \leqslant i \leqslant n\}$ 完全独立 (利用 $\{X_n\}$ 的完全独立性), 于是有

$$\sigma^2\left(\frac{T_n}{b_n}\right) = \sum_{i=1}^{n} \sigma^2\left(\frac{Y_{n,j}}{b_n}\right) \leqslant \sum_{i=1}^{n} \mathbb{E}\left[\left(\frac{Y_{n,i}}{b_n}\right)^2\right] = o(1).$$

于是继续利用 (2) 式, 便有

$$\frac{T_n - \mathbb{E}(T_n)}{b_n} \xrightarrow{\mathbb{P}} 0. \tag{13}$$

又因为

$$\left|\frac{S_n - \mathbb{E}(T_n)}{b_n}\right| \leqslant \left|\frac{S_n - T_n}{b_n}\right| + \left|\frac{T_n - \mathbb{E}(T_n)}{b_n}\right|,$$

所以

$$\begin{aligned}
\left\{\left|\frac{S_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\} &\subseteq \left\{\left|\frac{S_n - T_n}{b_n}\right| + \left|\frac{T_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\} \\
&\subseteq \left\{\left|\frac{S_n - T_n}{b_n}\right| > 0\right\} \cup \left\{\left|\frac{T_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\} \\
&= \{S_n \neq T_n\} \cup \left\{\left|\frac{T_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\}.
\end{aligned}$$

再由 (12) 与 (13) 式便可知

$$\mathbb{P}\left\{\left|\frac{S_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\} \leqslant \mathbb{P}\{S_n \neq T_n\} + \mathbb{P}\left\{\left|\frac{T_n - \mathbb{E}(T_n)}{b_n}\right| > \varepsilon\right\} \to 0 \quad (n \to \infty).$$

又因为

$$\mathbb{E}(T_n) = \sum_{i=1}^{n} \mathbb{E}(Y_{n,i}) = \sum_{i=1}^{n} \int_{|x| \leqslant b_n} x \, dF_i(x) = a_n,$$

于是 (10) 式得证. ▸ Back to Content. ■

Section 3

Convergence of Series

- One magic result in probability theory concerns situations in which the probability of an event can only be 0 or 1, i.e. Kolmogorov's zero-one law.
- Essential to the study of random series are maximal inequalities — inequalities concerning the maxima of partial sums.
- Finally, let's discuss Kolmogorov's three series theorem on how to determine whether a general independent r.v.'s series converges or not.

The σ-Algebra Generated by R.V.'S

- Let $(X_\lambda)_{\lambda\in\Lambda}$ be a nonempty family of r.v.'s on $(\Omega, \mathscr{F})$ (Λ may not be countable). Define

$$\begin{aligned}\sigma(X_\lambda, \lambda\in\Lambda) &:= \sigma(X_\lambda \in B, B\in\mathscr{B}, \lambda\in\Lambda) = \sigma(X_\lambda^{-1}(\mathscr{B}^1), \lambda\in\Lambda)\\ &= \sigma\Big(\bigcup_{\lambda\in\Lambda} X_\lambda^{-1}(\mathscr{B}^1)\Big),\end{aligned}$$

  which is called **the σ-algebra generated by** $(X_\lambda)_{\lambda\in\Lambda}$.
- $\sigma(X_\lambda, \lambda\in\Lambda)$ is the smallest σ-algebra with respect to which *each* X_λ is measurable.
- For $\Lambda = \{1,2,\cdots,n\}$ (n may be ∞), we have

$$\sigma(X_i) = \sigma(X^{-1}(\mathscr{B}^1))$$
$$\sigma(X_1,\cdots,X_n) = \sigma\Big(\bigcup_{i=1}^n X_i^{-1}(\mathscr{B}^1)\Big) = \sigma\Big(\bigcup_{i=1}^n \sigma(X_i)\Big).$$

- For $\Lambda = \{1,2,\cdots\}$, it is easy to check that

$$\sigma(X_1) \subseteq \sigma(X_1,X_2) \subseteq \cdots \subseteq \sigma(X_1,\cdots,X_n),$$
$$\sigma(X_1,X_2,\cdots) \supseteq \sigma(X_2,X_3,\cdots) \supseteq \cdots \supseteq \sigma(X_n,X_{n+1},\cdots).$$

- Consider the set A:

$$A = \left\{\omega : \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{n\to\infty} 0\right\}.$$

- For each $m \in \mathbb{N}$, the values of $X_1(\omega), \cdots, X_{m-1}(\omega)$ are irrelevant to the question of whether or not ω lies in A, and so A ought to lie in the σ-algebra $\sigma(X_m, X_{m+1}, \cdots)$.
- In fact, $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{m-1} X_i(\omega) = 0$ for a fixed m, and hence $\omega \in A$ iff $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=m}^{n} X_i(\omega) = 0$. Therefore,

$$A = \bigcap_{\varepsilon} \bigcup_{N \geqslant m} \bigcap_{n \geqslant N} \left\{\omega : \left|\frac{1}{n}\sum_{i=m}^{n} X_i(\omega)\right| < \varepsilon\right\},^{2} \tag{14}$$

  the first intersection extending over positive rational ε.

  Translation. $\forall\ \varepsilon > 0$, $\exists\ N \geqslant m$, s.t. $\forall\ n \geqslant N$, 有 $|\frac{1}{n}\sum_{i=m}^{n} X_i(\omega)| < \varepsilon$.
- The set on the inside lies in $\sigma(X_m, X_{m+1}, \cdots)$, and hence so does A. Similarly, the ω-set where the series $\sum_{n=1}^{\infty} X_n(\omega)$ converges lies in each $\sigma(X_m, X_{m+1}, \cdots)$.

[2] The reader has to be able to translate this equation into language on his own.

Tail σ-Algebra and Kolmogorov's Zero-One Law

- The intersection

$$\mathscr{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \cdots)$$

  is the **tail σ-algebra** associated with the sequence (X_n); its elements are **tail events**.
- Through the analysis on the previous page, we are surprised to find that the *sets* of sample points that make

$$\text{the partial sum } \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \text{the series } \sum_{n=1}^{\infty} X_n$$

  converge is tail events. That's our motivation for introducing tail events.
- The following theorem discusses the properties of tail events in general.

定理 9 (Kolmogorov's Zero-One Law)

Suppose that (X_n) is independent and that $A \in \mathscr{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \cdots)$. Then either $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$.

Kolmogorov's Zero-One Law

PROOF. **去证明 $A \in \mathscr{T}$ 与其自身独立**, i.e. $\mathbb{P}(A\cap A)=\mathbb{P}(A)\mathbb{P}(A)$, 于是 $\mathbb{P}(A)=\mathbb{P}(A)^2$, 这就说明 $\mathbb{P}(A)=0$ 或 1. 现在分两步证明上述断言.

Step 1. 事件 $\Lambda\in\sigma(X_1,\cdots,X_n)$ 与 $M\in\sigma(X_{n+1},X_{n+2},\cdots)$ 独立.

由第三章定理 26, 因 $\mathscr{F}_n=\sigma(X_1,\cdots,X_n)$ 与 $\mathscr{F}'_n=\sigma(X_{n+1},X_{n+2},\cdots)$ 均是 π-系且包含 Ω, 于是可知事件 Λ 与 M 独立.

Step 2. 事件 $\Lambda\in\sigma(X_1,X_2,\cdots)$ 与 $M\in\mathscr{T}$ 独立.

因为 $\mathscr{T}\subseteq\sigma(X_{n+1},X_{n+2},\cdots)$, 若存在 $n\in\mathbb{N}$ 使得 $\Lambda\in\mathscr{F}_n$, 由第一步结论可知第二步成立. 又因为 $\bigcup_n\mathscr{F}_n$ 与 $\mathscr{T}$ 均为包含 Ω 的 π-系, 再由第三章定理 26 可知 Λ 与 M 独立.

又因为 $\mathscr{T}\subseteq\sigma(X_1,X_2,\cdots)$, 由第二步的结论, 取 $\Lambda=A$, $M=A$, 便可知 $A\in\mathscr{T}$ 与其自身独立. 定理得证. ∎

Remark

As noted above, the set where $\sum_n X_n(\omega)$ converges satisfies the hypothesis of Theorem 9, and so does the set where $n^{-1}\sum_{i=1}^n X_i(\omega)\to 0$. In many similar cases it is very easy to prove by this theorem that a set at hand must have probability either 0 or 1. But to determine which of 0 and 1 is, in fact, the probability of the set may be extremely difficult.

Kolmogorov's Maximal Inequalities (Upper Bound)

- If the terms of an infinite series are independent r.v.'s, we have shown that the probability of its convergence is either zero or one.
- Here we shall establish a concrete criterion for the latter alternative.
- Not only is the result a complete answer to the question of convergence of independent r.v.'s, but it yields also a satisfactory form of the strong law of large numbers.

定理 10 (Kolmogorov's Maximal Inequalities (Upper Bound))

Let (X_n) *be independent r.v.'s such that*

$$\forall\ n \in \mathbb{N}: \quad \mathbb{E}(X_n) = 0, \quad \mathbb{E}(X_n^2) = \sigma^2(X_n) < \infty.$$

Then we have for every $\varepsilon > 0$*:*

$$\mathbb{P}\left\{\max_{1 \leqslant i \leqslant n} |S_i| > \varepsilon\right\} \leqslant \frac{\sigma^2(S_n)}{\varepsilon^2}. \tag{15}$$

- Chebyshev's inequality is a special case of Kolmogorov's maximal inequality by taking $n = 1$.

Kolmogorov's Maximal Inequalities (Upper Bound)

PROOF. $\forall\ \varepsilon > 0$, 对于 $\Lambda = \left\{\omega : \max\limits_{1\leqslant i\leqslant n} |S_i(\omega)| > \varepsilon\right\}$, 定义

$$\nu(\omega) = \min\{i : 1 \leqslant i \leqslant n, |S_i(\omega)| > \varepsilon\},$$

则 ν 是定义在 Λ 上的随机变量. 令

$$\Lambda_k = \{\omega : \nu(\omega) = k\} = \left\{\omega : \max_{1\leqslant i\leqslant k-1} |S_i(\omega)| \leqslant \varepsilon, |S_k(\omega)| > \varepsilon\right\},$$

其中对于 $k=1$, 默认 $\max\limits_{1\leqslant i\leqslant 0} |S_i(\omega)| = 0$. 由定义, $(\Lambda_k)_{k\geqslant 1}$ 互不相容, 且 $\Lambda = \bigcup_{k=1}^n \Lambda_k$. 故

$$\begin{aligned}
\int_\Lambda S_n^2\, d\mathbb{P} = \sum_{k=1}^n \int_{\Lambda_k} S_n^2\, d\mathbb{P} = \sum_{k=1}^n \int_{\Lambda_k} \left[S_k + (S_n - S_k)\right]^2 d\mathbb{P} \\
= \sum_{k=1}^n \int_{\Lambda_k} \left[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] d\mathbb{P}.
\end{aligned} \tag{16}$$

设 $\varphi_k = \mathbf{1}_{\Lambda_k}$, 于是由第三章定理 29 可知 $\varphi_k S_k$ 与 $S_n - S_k$ 独立.

Kolmogorov's Maximal Inequalities (Upper Bound)

于是

$$\int_{\Lambda_k} S_k(S_n - S_k)\,d\mathbb{P} = \int_{\Omega} (\varphi_k S_k)(S_n - S_k)\,d\mathbb{P} = \int_{\Omega} \varphi_k S_k\,d\mathbb{P} \int_{\Omega} (S_n - S_k)\,d\mathbb{P}.$$

其中上式第二个等号利用了 $\varphi_k S_k$ 与 $S_n - S_k$ 的独立性. 由定理条件: $\mathbb{E}(X_n) = 0$, 可知

$$\int_{\Omega} (S_n - S_k)\,d\mathbb{P} = \mathbb{E}(S_n - S_k) = \sum_{i=k+1}^{n} \mathbb{E}(X_i) = 0.$$

于是由 (16) 式:

$$\begin{aligned}
\sigma^2(S_n) &= \int_{\Omega} S_n^2\,d\mathbb{P} \geqslant \int_{\Lambda} S_n^2\,d\mathbb{P} = \sum_{k=1}^{n} \int_{\Lambda_k} S_n^2\,d\mathbb{P} \\
&\geqslant \sum_{k=1}^{n} \int_{\Lambda_k} S_k^2\,d\mathbb{P} \quad [\text{在 } \Lambda_k \text{ 上 } |S_k(\omega)| \geqslant \varepsilon \text{ 成立}] \\
&\geqslant \varepsilon^2 \sum_{k=1}^{n} \mathbb{P}(\Lambda_k) = \varepsilon^2 \mathbb{P}(\Lambda).
\end{aligned}$$

将 ε^2 除到上述不等式左边便有 (15) 式成立. ■

Kolmogorov's Maximal Inequalities (Lower Bound)

定理 11 (Kolmogorov's Maximal Inequalities (Lower Bound))

Let (X_n) *be independent r.v.'s with finite mean and suppose that there exsits an* A *such that*

$$\forall\ n \in \mathbb{N}: \quad |X_n - \mathbb{E}(X_n)| \leqslant A < \infty. \tag{17}$$

Then for every $\varepsilon > 0$ *we have*

$$\mathbb{P}\left\{\max_{1 \leqslant i \leqslant n} |S_i| \leqslant \varepsilon\right\} \leqslant \frac{(2A + 4\varepsilon)^2}{\sigma^2(S_n)}. \tag{18}$$

Proof. 不妨设 $M_0 = \Omega$, $\forall\ 1 \leqslant k \leqslant n$, 定义

$$M_k = \{\omega : \max_{1 \leqslant i \leqslant k} |S_i| \leqslant \varepsilon\}, \quad \Delta_k = M_{k-1} \backslash M_k.$$

不妨设 $\mathbb{P}(M_n) > 0$, 否则不等式 (18) 显然满足. 现在, 令 $S_0' = 0$, $\forall\ k \geqslant 1$, 定义

$$X_k' = X_k - \mathbb{E}(X_k), \quad S_k' = \sum_{i=1}^{k} X_i'.$$

再定义 a_k $(0 \leqslant k \leqslant n)$:

$$a_k = \frac{1}{\mathbb{P}(M_k)} \int_{M_k} S'_k \, d\mathbb{P},^{3}$$

于是

$$\int_{M_k} (S'_k - a_k) \, d\mathbb{P} = 0. \tag{19}$$

现在我们将给出如下积分的估计:

$$\begin{aligned} \int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2 \, d\mathbb{P} &= \int_{M_k} (\underline{S'_k - a_k} + \underline{a_k - a_{k-1}} + X'_{k+1})^2 \, d\mathbb{P} \quad (I_1) \\ &\quad - \int_{\Delta_{k+1}} (\underline{S'_k - a_k} + \underline{a_k - a_{k-1}} + X'_{k+1})^2 \, d\mathbb{P} \quad (I_2). \end{aligned} \tag{20}$$

首先利用 $M_k = \{\omega : \max\limits_{1 \leqslant i \leqslant k} |S_i| \leqslant \varepsilon\}$ 的定义, 我们有

$$\begin{aligned} |S'_k - a_k| &= \left| S_k - \mathbb{E}(S_k) - \frac{1}{\mathbb{P}(M_k)} \int_{M_k} [S_k - \mathbb{E}(S_k)] \, d\mathbb{P} \right| = \left| S_k - \frac{1}{\mathbb{P}(M_k)} \int_{M_k} S_k \, d\mathbb{P} \right| \\ &\leqslant |S_k| + \frac{1}{\mathbb{P}(M_k)} \int_{M_k} |S_k| \, d\mathbb{P} \leqslant |S_k| + \varepsilon. \quad [\max_{1 \leqslant i \leqslant k} |S_i| \leqslant \varepsilon \ \Rightarrow \ |S_k| \leqslant \varepsilon] \end{aligned}$$

[3] 容易知道, $\forall\ 0 \leqslant k \leqslant n$, 有 $\max\limits_{1 \leqslant i \leqslant n} |S_i| \geqslant \max\limits_{1 \leqslant i \leqslant k} |S_i|$, 于是 $M_n \subseteq M_k$. 从而 $0 < \mathbb{P}(M_n) \leqslant \mathbb{P}(M_k)$, 这说明 a_k 定义中的分母不为零, 良定义.

类似地, 再次利用三角不等式与 M_k 的定义与 (17) 式, 还有

$$\begin{aligned} |a_k - a_{k-1}| &= \left| \frac{1}{\mathbb{P}(M_k)} \int_{M_k} S_k' \, d\mathbb{P} - \frac{1}{\mathbb{P}(M_{k+1})} \int_{M_{k+1}} S_k' \, d\mathbb{P} - \frac{1}{\mathbb{P}(M_{k+1})} \int_{M_{k+1}} X_{k+1}' \, d\mathbb{P} \right| \\ &\leqslant 2\varepsilon + A. \qquad |X_{k+1}'| = |X_{k+1} - \mathbb{E}(X_{k+1})| \leqslant A \end{aligned} \tag{21}$$

又因为在 Δ_{k+1} 上有 $|S_k| \leqslant \varepsilon$, 于是

$$I_2 \leqslant \int_{\Delta_{k+1}} \big(|S_k| + \varepsilon + 2\varepsilon + A + A\big)^2 \, d\mathbb{P} \leqslant (4\varepsilon + 2A)^2 \mathbb{P}(\Delta_{k+1}).$$

另一方面, 我们还有

$$\begin{aligned} I_1 = \int_{M_k} \big[&(S_k' - a_k)^2 + (a_k - a_{k-1})^2 + (X_{k+1}')^2 + 2(S_k' - a_k)(a_k - a_{k-1}) \\ &+ 2(S_k' - a_k) X_{k+1}' + 2X_{k+1}'(a_k - a_{k-1}) \big] \, d\mathbb{P} \end{aligned}$$

由 (X_n) 的独立性, 利用第三章定理 29 和 (19) 式, 上式蓝色项所对应的积分均为零.

另外,

$$\int_{M_k} 2X'_{k+1}(a_k - a_{k-1})\, d\mathbb{P} = (a_k - a_{k-1}) \int_{M_k} \left[X_{k+1} - \mathbb{E}(X_{k+1})\right] d\mathbb{P}$$

$$= \int_\Omega \mathbf{1}_{M_k} X_{k+1}\, d\mathbb{P} - \mathbb{E}(X_{k+1})\mathbb{P}(M_k)$$

$$(\clubsuit) = \int_\Omega \mathbf{1}_{M_k}\, d\mathbb{P} \cdot \int_\Omega X_{k+1}\, d\mathbb{P} - \mathbb{E}(X_{k+1})\mathbb{P}(M_k) = 0.$$

其中第三个等式成立, 只需要注意到 $\mathbf{1}_{M_k}$ 作为 $X_1, \cdots, X_k$ 的函数与 X_{k+1} 独立即可. 于是

$$I_1 \geqslant \int_{M_k} (S'_k - a_k)^2\, d\mathbb{P} + \int_{M_k} (X'_{k+1})^2\, d\mathbb{P} = \int_{M_k} (S'_k - a_k)^2\, d\mathbb{P} + \mathbb{P}(M_k)\sigma^2(X_{k+1}).$$

其中上式红色项成立类似于 $(\clubsuit)$ 的处理. 现在, 在 (20) 式右端使用 I_1 与 I_2 的估计, 有

$$\int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2\, d\mathbb{P} \geqslant \int_{M_k} (S'_k - a_k)^2\, d\mathbb{P} + \mathbb{P}(M_k)\sigma^2(X_{k+1}) - (4\varepsilon + 2A)^2 \mathbb{P}(\Delta_{k+1})$$

$$[M_n \subseteq M_k] \qquad \geqslant \int_{M_k} (S'_k - a_k)^2\, d\mathbb{P} + \mathbb{P}(M_n)\sigma^2(X_{k+1}) - (4\varepsilon + 2A)^2 \mathbb{P}(\Delta_{k+1}),$$

即 $\displaystyle\int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2\, d\mathbb{P} - \int_{M_k} (S'_k - a_k)^2\, d\mathbb{P} \geqslant \mathbb{P}(M_n)\sigma^2(X_{k+1}) - (4\varepsilon + 2A)^2 \mathbb{P}(\Delta_{k+1}).$

在上式两端同时对 k 从 0 至 $n-1$ 求和，我们有

$$\int_{M_n}(S_n'-a_n)^2\,d\mathbb{P}-\int_{M_0}(S_0'-a_0)^2\,d\mathbb{P}\geqslant\mathbb{P}(M_n)\sum_{i=1}^n\sigma^2(X_i)-(4\varepsilon+2A)^2\sum_{i=1}^n\mathbb{P}(\Delta_i).$$

注意到 $S_0'=a_0=0$，再由 Δ_n 的定义与 (X_n) 的独立性，上式化简为

$$\int_{M_n}(S_n'-a_n)^2\,d\mathbb{P}\geqslant\mathbb{P}(M_n)\sigma^2(S_n)-(4\varepsilon+2A)^2\mathbb{P}(\Omega\setminus M_n)\quad(\spadesuit).$$

另一方面，再次利用 $|S_k'-a_k|$ 的估计，我们有

$$\begin{aligned}\int_{M_n}(S_n'-a_n)^2\,d\mathbb{P}&\leqslant\int_{M_n}(|S_n|+\varepsilon)^2\,d\mathbb{P}=\int_{M_n}(|S_n|^2+2\varepsilon|S_n|+\varepsilon^2)\,d\mathbb{P}\\&\leqslant\int_{M_n}(\varepsilon^2+2\varepsilon\cdot\varepsilon+\varepsilon^2)\,d\mathbb{P}=4\varepsilon^2\mathbb{P}(M_n).\end{aligned}$$

代回 $(\spadesuit)$ 式中我们有

$$4\varepsilon^2\mathbb{P}(M_n)\geqslant\mathbb{P}(M_n)\sigma^2(S_n)-(4\varepsilon+2A)^2\mathbb{P}(\Omega\setminus M_n).$$

又因为 $(4\varepsilon+2A)^2\geqslant 4\varepsilon^2$，移项整理上式便完成了 (18) 式的证明. ■

Lévy's Inequality

Reference. Varadhan, SR Srinivasa. *Probability theory*. American Mathematical Soc., 2001. Lemma 3.8.

定理 12 (Lévy's Inequality)

Let (X_n) *be independent r.v.'s, assume that*

$$\mathbb{P}\left\{\max_{1\leqslant k\leqslant n}|X_k+\cdots+X_n|\geqslant\frac{\ell}{2}\right\}\leqslant\delta, \tag{22}$$

then

$$\mathbb{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\}\leqslant\frac{\delta}{1-\delta}. \tag{23}$$

Proof. 不妨设 $E_k=\left\{\max_{1\leqslant j\leqslant k-1}|S_j|\leqslant\ell,|S_k|\geqslant\ell\right\}$, 则 $\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\}=\bigcup_{k=1}^{n}E_k$ 为无交并. 现在由 (22) 式, 我们有

$$\mathbb{P}\left(\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\}\cap\left\{|S_n|>\frac{\ell}{2}\right\}\right)\leqslant\mathbb{P}\left\{|S_n|>\frac{\ell}{2}\right\}\leqslant\mathbb{P}\left\{|S_n|\geqslant\frac{\ell}{2}\right\}\leqslant\delta.$$

另外，

$$\mathbb{P}\left(\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant \ell\right\}\cap\left\{|S_n|\leqslant\frac{\ell}{2}\right\}\right)=\sum_{k=1}^{n}\mathbb{P}\left(E_k\cap\left\{|S_n|\leqslant\frac{\ell}{2}\right\}\right)$$

$$\leqslant\sum_{k=1}^{n}\mathbb{P}\left(E_k\cap\left\{|S_n-S_k|\geqslant\frac{\ell}{2}\right\}\right)\quad\text{[利用第三章定理 29: 独立性]}$$

$$=\sum_{k=1}^{n}\mathbb{P}(E_k)\mathbb{P}\left\{|S_n-S_k|\geqslant\frac{\ell}{2}\right\}\leqslant\delta\sum_{k=1}^{n}\mathbb{P}(E_k)=\delta\,\mathbb{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\}.$$

现在，将上述两式求和，我们有

$$\mathbb{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\}\leqslant\delta+\delta\,\mathbb{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\ell\right\},$$

化简记得 (23) 式. ■

我们来证明 $E_k\cap\{|S_n|\leqslant\ell/2\}\subseteq E_k\cap\{|S_n-S_k|\geqslant\ell/2\}$. 若 $\omega\in E_k$，则 $|S_k(\omega)|\geqslant\ell$. 另外，若 $\omega\in\{|S_n|\leqslant\ell/2\}$，则

$$\ell\leqslant|S_k(\omega)|\leqslant|S_n(\omega)-S_k(\omega)|+|S_n(\omega)|\leqslant|S_n(\omega)-S_k(\omega)|+\ell/2,$$

从而 $|S_n(\omega)-S_k(\omega)|\geqslant\ell/2$，于是 $\omega\in\{|S_n-S_k|\geqslant\ell/2\}$.

Kolmogorov's One Series Theorem

- For independent X_n, the probability that $\sum X_n$ converges is either 0 or 1. It is natural to try and characterize the two cases in terms of the distributions of the individual X_n.

定理 13 (Kolmogorov's One Series Theorem)

Let (X_n) be independent r.v.'s, each of which has finite mean and variance, satisfy $\mathbb{E}(X_n) = 0$ and $\sum_{n=1}^{\infty} \sigma^2(X_n) < \infty$, then $\sum_{n=1}^{\infty} X_n$ converges a.e. ($\Leftrightarrow$ with probability 1).

Proof. 由 Kolmogorov 不等式 (定理 10),

$$\mathbb{P}\left(\max_{1 \leqslant k \leqslant r} |S_{n+k} - S_n| > \varepsilon \right) \leqslant \frac{1}{\varepsilon^2} \sum_{k=1}^{r} \sigma^2(X_{n+k}).$$

上式右端是关于 r 得不减函数, 令 $r \to \infty$, 有

$$\mathbb{P}\left(\max_{k \geqslant 1} |S_{n+k} - S_n| > \varepsilon \right) \leqslant \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma^2(X_{n+k}).$$

由 $\sum_{n=1}^{\infty}\sigma^2(X_n)<\infty$, 令 $n\to\infty$, 便有

$$\lim_{n\to\infty}\mathbb{P}\left(\max_{k\geqslant 1}|S_{n+k}-S_n|>\varepsilon\right)=0. \tag{24}$$

现在设 $E(n,\varepsilon)=\{\omega:\max\limits_{j,k\geqslant n}|S_j-S_k|>2\varepsilon\}$, 定义 $E(\varepsilon)=\bigcap\limits_{n=1}^{\infty}E(n,\varepsilon)$, 则 $E(n,\varepsilon)\downarrow E(\varepsilon)$. 由概率测度的连续性, 利用 (24) 式, 有

$$\mathbb{P}(E(\varepsilon))=\lim_{n\to\infty}\mathbb{P}(E(n,\varepsilon))=0.$$

现在, 集合 $\bigcup\limits_{\varepsilon\in\mathbb{Q}}E(\varepsilon)$ 包含了使得 $(S_n)_n$ 不收敛的样本点构成的集合 (其概率为 0). 证毕. ■

推论 14 (Kolmogorov's Two Series Theorem)

Let $\mu_n=\mathbb{E}(X_n)$ be the means and $\sigma_n^2=\sigma^2(X_n)$ the variance of a sequence of independent r.v.'s (X_n). Assume that $\sum\limits_{n=1}^{\infty}\mu_n$ and $\sum\limits_{n=1}^{\infty}\sigma_n^2$ converge, then the series $\sum\limits_{n=1}^{\infty}X_n$ converges a.e. Just define $Y_n=X_n-\mu_n$ and apply the previous theorem to Y_n.

Kolmogorov's Three Series Theorem

Remark (Finite Moments & Infinite Moments?)

- In general r.v.'s need not have finite means or variances. If (X_n) is any sequence of r.v.'s we can take a "truncation (截断)" value A and define

$$Y_n = \begin{cases} X_n, & \text{if } |X_n| \leqslant A; \\ 0, & \text{if } |X_n| > A. \end{cases}$$

- The (Y_n) are independent and bounded in absolute value by A. The theorem can be applied to (Y_n) and if we impose the additional condition that

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A) < \infty,$$

  by an application of Borel-Cantelli lemma, with probability 1 ($\Leftrightarrow$ a.e.), $X_n = Y_n$ for all sufficiently large n. Discussed precisely later.

- The convergence of $\sum_{n=1}^{\infty} X_n$ and $\sum_{n=1}^{\infty} Y_n$ are therefore equivalent. Then we have the *Kolmogorov's Three Series Theorem*.

Kolmogorov's Three Series Theorem

- The final result in this section, the *three-series theorem*, provides necessary and sufficient conditions for the convergence of $\sum X_n$ in terms of the individual distribution of the X_n.

定理 15 (Kolmogorov's Three Series Theorem)

Let (X_n) be independent r.v.'s and define for a fixed constant $A > 0$:

$$Y_n(\omega) = \begin{cases} X_n(\omega), & \textit{if } |X_n(\omega)| \leqslant A; \\ 0, & \textit{if } |X_n(\omega)| > A. \end{cases}$$

Then the series $\sum_{n=1}^{\infty} X_n$ converges a.e. if and only if the following three series all converge:

i). $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n)$;

ii). $\sum_{n=1}^{\infty} \mathbb{E}(Y_n)$;

iii). $\sum_{n=1}^{\infty} \sigma^2(Y_n)$.

PROOF. (⇐). 假设定理中的三个级数均收敛. 由 (i) 收敛, 利用 Borel-Cantelli 引理, 我们有

$$\begin{aligned}\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0 \quad &\iff \quad \mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{X_m \neq Y_m\}\Big) = 0 \\ &\iff \quad \mathbb{P}\Big(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{X_m = Y_m\}\Big) = 1. \qquad (\star)\end{aligned}$$

$(\star)$ 式表明: $\exists\ n > 0$, s.t. $\forall\ m \geqslant n$, 有 $X_m = Y_m$ a.e., 即 X_m 与 Y_m 以概率 1 最终相等. 这就说明级数 $\sum_{n=1}^{\infty} X_n$ 与 $\sum_{n=1}^{\infty} Y_n$ 的收敛性一致. 再由级数 (ii) 与 (iii) 的收敛性, 利用推论 14, 可知 $\sum_{n=1}^{\infty} Y_n$ 以概率 1 收敛, 于是 $\sum_{n=1}^{\infty} X_n$ 也以概率 1 收敛.

(⇒). 假设 $\sum_{n=1}^{\infty} X_n$ 以概率 1 收敛, 于是 $X_n \to 0$ a.e. 再由 Borel-Cantelli 引理, $\forall\ A > 0$, 我们有

$$\mathbb{P}(|X_n| > A \text{ i.o.}) = \mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0.$$

再次利用 Borel-Cantelli 引理, 此时有

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) < \infty.$$

即级数 (i) 收敛. 由充分性证明中的同样操作可知 $\sum_{n=1}^{\infty} Y_n$ 以概率 1 收敛.

但因为 $|Y_n-\mathbb{E}(Y_n)|<2A$，由定理 11 可知

$$\mathbb{P}\left\{\max_{n\leqslant k\leqslant r}\left|\sum_{j=n}^{k}Y_j\right|\leqslant\varepsilon\right\}\leqslant\frac{(4A+4\varepsilon)^2}{\sum_{j=n}^{r}\sigma^2(Y_j)}.$$

若 $\sum_{n=1}^{\infty}\sigma^2(Y_n)$ 发散，于是 $\sum_{j=n}^{r}\sigma^2(Y_j)\to\infty$ $(r\to\infty)$. 于是在上式中令 $r\to\infty$，有

$$\forall\ n\in\mathbb{N}:\ 0\leqslant\mathbb{P}\left\{\sup_{k\geqslant n}\left|\sum_{j=n}^{k}Y_j\right|\leqslant\varepsilon\right\}\leqslant\frac{(4A+4\varepsilon)^2}{\sum_{j=n}^{\infty}\sigma^2(Y_j)}=0\ \Rightarrow\ \mathbb{P}\left\{\sup_{n}\left|\sum_{j=n}^{\infty}Y_j\right|\leqslant\varepsilon\right\}=0.$$

等价的：$\mathbb{P}\left\{\sup_{n}\left|\sum_{j=n}^{\infty}Y_j\right|>\varepsilon\right\}=1$. 这说明级数 $\sum_{k=1}^{\infty}Y_k$ 的余项 $\sum_{j=n}^{\infty}Y_j$ 以概率 1 无法被任何常数 ε 控制，于是级数 $\sum_{k=1}^{\infty}Y_k$ 无法几乎处处收敛，矛盾！从而 $\sum_{n=1}^{\infty}\sigma^2(Y_n)$ 收敛.

最后考虑级数 $\sum_{n=1}^{\infty}(Y_n-\mathbb{E}Y_n)$. 首先，$\mathbb{P}(|Y_n-\mathbb{E}Y_n|>2A)=0$ 且 $\mathbb{E}(Y_n-\mathbb{E}Y_n)=0$，于是对应的级数 (i) 与 (ii) 收敛. 另外，级数 (iii) 也收敛 (刚刚证毕)，于是由该定理的充分性可知级数 $\sum_{n=1}^{\infty}(Y_n-\mathbb{E}Y_n)$ 几乎处处收敛. 又因为 $\sum_{n=1}^{\infty}Y_n$ 几乎处处收敛，于是级数

$$\sum_{n=1}^{\infty}\mathbb{E}(Y_n)=\sum_{n=1}^{\infty}Y_n-\sum_{n=1}^{\infty}(Y_n-\mathbb{E}Y_n)$$

几乎处处收敛，于是级数 (ii) 收敛. ■

- The convergence of a series of r.v.'s is, of course, by definition the same as its partial sums, in any sense of convergence discussed in Chapter 4.
- For independent terms we have, however, the following theorem due to Paul Lévy.

引理 16 (Lévy's Theorem)

If (X_n) is a sequence of independent r.v.'s, then the convergence of the series $\sum_{n=1}^{\infty} X_n$ in pr. is equivalent to its convergence in dist. and convergence a.e.

PROOF. Convergence a.e. $\Rightarrow$ convergence in pr. $\Rightarrow$ convergence in dist. are trivial.

Convergence in dist. $\Rightarrow$ convergence in pr. will be proved after we introduced characteristic functions (Chapter 6).

我们现在来证明: convergence in pr. $\Rightarrow$ convergence a.e. 若级数 $\sum_n X_n$ 依概率收敛, 于是 $(S_n)_n$ 是 Cauchy 列: $\forall\ \varepsilon > 0$, $\delta > 0$, $\exists\ p > 0$, s.t. $\forall\ n, m \geqslant p$, 有 $\mathbb{P}(|S_n - S_m| \geqslant \varepsilon) \leqslant \delta$.

现在不妨设 $Y_i = X_{p+i}$, $N = q - p$, $M = k - p$ $(p+1 \leqslant k \leqslant q)$, 于是 $1 \leqslant M \leqslant N$. 由 (S_n) 是 Cauchy 列, 我们有:

$$\mathbb{P}(|Y_M + \cdots + Y_N| \geqslant \varepsilon/2) \leqslant \delta.$$

于是由 Lévy 不等式,

$$\mathbb{P}\left(\max_{1\leqslant M\leqslant N}|Y_1+\cdots+Y_M|\geqslant\varepsilon\right)\leqslant\frac{\delta}{1-\delta},$$

即

$$\mathbb{P}\left(\max_{1\leqslant M\leqslant N}|X_{p+1}+\cdots+X_{p+M}|\geqslant\varepsilon\right)\leqslant\frac{\delta}{1-\delta}\iff\mathbb{P}\left(\max_{p<k\leqslant q}|S_k-S_p|\geqslant\varepsilon\right)\leqslant\frac{\delta}{1-\delta}.$$

由 ε 与 δ 的任意性, 我们有 Back to Content.

$$\lim_{p,q\to\infty}\mathbb{P}\left(\max_{p<k\leqslant q}|S_k-S_p|\geqslant\varepsilon\right)=0,$$

这说明 (S_n) 在某一零测集外是 Cauchy 列, 于是 $\sum_n X_n$ 几乎处处收敛. ■

Remark

Regarding the proof of Lévy's theorem, we also need the following note:

- A sequence of r.v.'s converging in probability (measure) can be inferred that the sequence is a Cauchy sequence in probability (measure). The proof can be found in MATH4123 Chapter 4 Lemma 33.

Section 4

Strong Law of Large Numbers

- This section begins with an introduction to Kronecker lemma which convert convergence results for random series into convergence of averages.
- Next, we give a sufficient condition for a sequence of independent r.v.'s to satisfy strong law of large numbers.
- Finally, we give a sufficient condition for a sequence of independent identically distributed r.v.'s to satisfy strong law of large numbers and study its expanded version.

- The important Kronecker lemma enables us to convert convergence results for random series into convergence of averages, i.e., into laws of large numbers.

引理 17 (Kronecker's Lemma)

Let (x_k) be a sequence of real numbers, (a_k) a sequence of numbers > 0 and $\uparrow \infty$. Then

$$\sum_{n=1}^{\infty} \frac{x_n}{a_n} \text{ converges } \Rightarrow \frac{1}{a_n} \sum_{i=1}^{n} x_i \to 0.$$

Proof. 令 $a_0 = 0$, $b_0 = 0$, 并且 $\forall\ n \geqslant 1$, 定义 $b_n = \sum_{i=1}^{n} x_i / a_i$, 于是 $x_n = a_n(b_n - b_{n-1})$, 且

$$\begin{aligned}\frac{1}{a_n}\sum_{i=1}^{n} x_i &= \frac{1}{a_n}\left(\sum_{i=1}^{n} a_i b_i - \sum_{i=1}^{n} a_i b_{i-1}\right) = \frac{1}{a_n}\left(a_n b_n + \sum_{i=2}^{n} a_{i-1} b_{i-1} - \sum_{i=1}^{n} a_i b_{i-1}\right)\\ &= \frac{1}{a_n}\left(a_n b_n + \sum_{i=2}^{n} (a_{i-1} - a_i) b_{i-1}\right) = b_n - \sum_{i=1}^{n} \frac{a_i - a_{i-1}}{a_n} b_{i-1}. \qquad (\clubsuit)\end{aligned}$$

由 $a_0 = b_0 = 0$ 不难看出上式角标是正确的.

Kronecker's Lemma

现在, 取 $B = \sup_n |b_n| < \infty$, 再取 $M > 0$, s.t. $|b_n - b_\infty| < \varepsilon/2$ ($\forall\ n \geqslant M$). 再取 $N > 0$ 满足

$$\forall\ n \geqslant N: \quad \frac{a_M}{a_n} < \frac{\varepsilon}{4B}.$$

于是, $\forall\ n \geqslant \max\{M, N\}$, 我们有

$$\begin{aligned}
\left|\sum_{i=1}^{n} \frac{a_i - a_{i-1}}{a_n} b_{i-1} - b_\infty\right| &\leqslant \sum_{i=1}^{n} \frac{a_i - a_{i-1}}{a_n} |b_{i-1} - b_\infty| \\
&\leqslant \sum_{i=1}^{M} \frac{a_i - a_{i-1}}{a_n} (|b_{i-1}| + |b_\infty|) + \sum_{i=M+1}^{n} \frac{a_i - a_{i-1}}{a_n} |b_{i-1} - b_\infty| \\
(\star) \quad &\leqslant 2B \frac{a_M}{a_n} + \frac{\varepsilon}{2} \frac{a_n - a_M}{a_n} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

现在, 在 (♣) 式两端同时令 $n \to \infty$, 我们有 $a_n^{-1} \sum_{i=1}^{n} x_i \to b_\infty - b_\infty = 0$. 引理得证. ■

$(\star)$ 式成立, 首先注意到 B 的定义, 将蓝色式放大提出求和号, 求和号各项可以相消, 只剩下 a_M/a_n. 又因为 $n \geqslant N$, 于是 $a_M/a_n < \varepsilon/(4B)$. 第二项类似, 不再赘述.

SLLN. A Sufficient Condition of Independent Case

- Now let φ be a positive, even, and continuous function on $\mathbb{R}$ such that as $|x|$ increase,

$$\frac{\varphi(x)}{|x|}\uparrow, \quad \frac{\varphi(x)}{x^2}\downarrow. \tag{25}$$

定理 18 (Convergence for Random Series. A Sufficient Condition of Independent Case)

Let (X_n) be a sequence of independent r.v.'s with $\mathbb{E}(X_n)=0$ for every n; and $0<a_n\uparrow\infty$. If φ satisfies the condition (25) above and

$$\sum_{n=1}^{\infty}\frac{\mathbb{E}\varphi(X_n)}{\varphi(a_n)}<\infty, \quad \textit{then} \quad \sum_{n=1}^{\infty}\frac{X_n}{a_n} \textit{ converges a.e.} \tag{26}$$

- Applying Kronecker's lemma to (26) for each ω in a set of probability one, we obtain:

推论 19 (SLLN. A Sufficient Condition of Independent Case)

Under the hypothesis of the theorem 18, we have

$$\frac{1}{a_n}\sum_{i=1}^{n}X_i=\frac{S_n}{a_n}\to 0 \text{ a.e.} \tag{27}$$

Particular Cases of SLLN (Independent Case)

i). Let $\varphi(x) = |x|^p$, $1 \leqslant p \leqslant 2$; $a_n = n$. Then we have

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \mathbb{E}(|X_n|^p) < \infty \ \Rightarrow \ \frac{1}{n} \sum_{i=1}^{n} X_i \to 0 \text{ a.e.} \tag{28}$$

For $p = 2$, this is due to Kolmogorov; for $1 \leqslant p < 2$, it is due to Marcinkiewicz and Zygmund.

ii). Suppose for some δ, $0 < \delta \leqslant 1$ and $M < \infty$ we have

$$\forall \ n \in \mathbb{N}: \quad \mathbb{E}(|X_n|^{1+\delta}) \leqslant M.$$

Then the hypothesis in (28) is clearly satisfied with $p = 1 + \delta$. This case is due to Markov.

Particular Cases of SLLN (Independent Case)

By proper choice of (a_n), we can considerably sharpen the conclusion (27). 强化结论

iii). Suppose

$$\forall\ n \in \mathbb{N}: \quad \sigma^2(X_n) = \sigma_n^2 < \infty, \quad \sigma^2(S_n) = s_n^2 = \sum_{i=1}^{n} \sigma_n^2 \to \infty.$$

Choose $\varphi(x) = x^2$ and $a_n = s_n(\log s_n)^{(1/2)+\varepsilon}$, $\varepsilon > 0$, in the corollary to Theorem 18. Then

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}(X_n^2)}{a_n^2} = \sum_{n=1}^{\infty} \frac{\sigma_n^2}{s_n^2(\log s_n)^{1+2\varepsilon}} < \infty$$

by Dini's theorem, and consequently

$$\frac{S_n}{s_n(\log s_n)^{(1/2)+\varepsilon}} \to 0 \text{ a.e.}$$

In case all $\sigma_n^2 = 1$ so that $s_n^2 = n$, the above ratio has a denominator that is close to $n^{1/2}$. Later we shall see that $n^{1/2}$ is a critical order of magnitude for S_n.

Convergence for Random Series: Proof

Proof of Theorem 18. 我们利用 Kolmogorov 三级数定理证明该定理. 设 F_n 是 X_n 的分布函数, 令

$$Y_n(\omega) = \begin{cases} X_n(\omega), & \text{if } |X_n(\omega)| \leqslant a_n; \\ 0, & \text{if } |X_n(\omega)| > a_n. \end{cases}$$

于是

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\frac{Y_n^2}{a_n^2}\right) = \sum_{n=1}^{\infty} \int_{|x| \leqslant a_n} \frac{x^2}{a_n^2}\, dF_n(x).$$

由假设 (25), $\varphi(x)/x^2$ 是单调递减函数, 于是

$$|x| \leqslant a_n \;\Rightarrow\; \frac{\varphi(x)}{x^2} \geqslant \frac{\varphi(a_n)}{a_n^2} \;\Rightarrow\; \frac{x^2}{a_n^2} \leqslant \frac{\varphi(x)}{\varphi(a_n)},$$

从而

$$\sum_{n=1}^{\infty} \sigma^2\left(\frac{Y_n}{a_n}\right) \leqslant \sum_{n=1}^{\infty} \mathbb{E}\left(\frac{Y_n^2}{a_n^2}\right) \leqslant \sum_{n=1}^{\infty} \int_{|x| \leqslant a_n} \frac{\varphi(x)}{\varphi(a_n)}\, dF_n(x) \leqslant \sum_{n=1}^{\infty} \frac{\mathbb{E}\varphi(X_n)}{a_n} < \infty.$$

于是级数 $\sum_{n=1}^{\infty} \sigma^2\left(\frac{Y_n - \mathbb{E}Y_n}{a_n}\right) < \infty$ (注意到 $\mathbb{E}Y_n$ 是常数).

另外, 因 $|Y_n - \mathbb{E}Y_n| \leqslant 2a_n$, 于是

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{Y_n - \mathbb{E}Y_n}{a_n}\right| > 2\right) = 0, \quad \text{另外} \quad \sum_{n=1}^{\infty} \mathbb{E}\left(\frac{Y_n - \mathbb{E}Y_n}{a_n}\right) = 0.$$

于是由 Kolmogorov 三级数定理,

$$\sum_{n=1}^{\infty} \frac{Y_n - \mathbb{E}Y_n}{a_n} \text{ 几乎处处收敛.} \quad (\spadesuit)$$

接下来, 我们有

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{|\mathbb{E}Y_n|}{a_n} &= \sum_{n=1}^{\infty} \frac{1}{a_n}\left|\int_{|x| \leqslant a_n} x\, dF_n(x)\right| \quad [\mathbb{E}(X_n) = 0] \\ &= \sum_{n=1}^{\infty} \frac{1}{a_n}\left|\int_{|x| > a_n} x\, dF_n(x)\right| \leqslant \sum_{n=1}^{\infty} \int_{|x| > a_n} \frac{|x|}{a_n}\, dF_n(x). \end{aligned}$$

在 $|x| > a_n$ 时, 利用假设 (25), $\varphi(x)/|x|$ 时单调递增函数, 于是

$$|x| > a_n \;\Rightarrow\; \frac{\varphi(x)}{|x|} \geqslant \frac{\varphi(a_n)}{a_n} \;\Rightarrow\; \frac{|x|}{a_n} \leqslant \frac{\varphi(x)}{\varphi(a_n)}.$$

于是

$$\sum_{n=1}^{\infty}\frac{|\mathbb{E}Y_n|}{a_n}\leqslant\sum_{n=1}^{\infty}\int_{|x|>a_n}\frac{\varphi(x)}{\varphi(a_n)}\,dF_n(x)\leqslant\sum_{n=1}^{\infty}\frac{\mathbb{E}\varphi(x)}{\varphi(a_n)}<\infty.$$

由此，再利用 (♠) 式，我们便知 $\sum\limits_{n=1}^{\infty}\frac{Y_n}{a_n}$ 几乎处处收敛. 要证明 (26) 式中的结论，此时只需证明 (X_n) 与 (Y_n) 等价. 因为 φ 是单调递增函数，于是

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}(X_n\neq Y_n)&=\sum_{n=1}^{\infty}\int_{|x|>a_n}dF_n(x)\leqslant\sum_{n=1}^{\infty}\int_{|x|>a_n}\frac{\varphi(x)}{\varphi(a_n)}\,dF_n(x)\\&\leqslant\sum_{n=1}^{\infty}\frac{\mathbb{E}\varphi(X_n)}{\varphi(a_n)}<\infty.\qquad |x|>a_n\Rightarrow\varphi(x)\geqslant\varphi(a_n)\end{aligned}$$

这就说明随机变量序列 (X_n) 与 (Y_n) 等价，进而 (X_n/a_n) 与 (Y_n/a_n) 等价. 利用推论 6 与引理 16，可知 $\sum\limits_{n=1}^{\infty}\frac{X_n}{a_n}$ 几乎处处收敛. 定理得证. ■

SLLN. A Sufficient Condition of IID Case

- We now come to the strong version of theorem 18 in the totally independent case. This result is also due to Kolmogorov.

定理 20 (SLLN. A Sufficient Condition of IID Case)

Let (X_n) be a sequence of independent and identically distributed r.v.'s. Then we have

$$\begin{aligned} &i). \quad \mathbb{E}(|X_1|)<\infty \;\Rightarrow\; \frac{S_n}{n}\to \mathbb{E}(X_1) \text{ a.e.} \\ &ii). \quad \mathbb{E}(|X_1|)=\infty \;\Rightarrow\; \varlimsup_{n\to\infty}\frac{|S_n|}{n}=+\infty \text{ a.e.} \end{aligned} \tag{29}$$

Proof. (i). 取 $a_n=n$, 并定义

$$Y_n(\omega)=\begin{cases} X_n(\omega), & \text{if } |X_n(\omega)|\leqslant n; \\ 0, & \text{if } |X_n(\omega)|>n. \end{cases}$$

由第三章定理 16, 我们有

$$\sum_{n=1}^{\infty}\mathbb{P}(X_n\neq Y_n)=\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>n)=\sum_{n=1}^{\infty}\mathbb{P}(|X_1|>n)\leqslant \mathbb{E}(|X_1|)<\infty.$$

这说明 (X_n) 与 (Y_n) 是等价的.

现在取 $\varphi(x) = x^2$, $X_n = Y_n - \mathbb{E}(Y_n)$, 由 (28) 式, 我们想要证明

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\left[(Y_n - \mathbb{E}Y_n)^2\right]}{n^2} < \infty, \qquad (\clubsuit)$$

由此便可推出

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \mathbb{E}Y_i\right] \to 0 \text{ a.e.}$$

又因为 $\mathbb{E}(Y_n) \xrightarrow{n\to\infty} \mathbb{E}(X_1)$ (单调收敛定理), 于是

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) \to \mathbb{E}(X_1),$$

最终, 便有

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}(X_1) \text{ a.e.}$$

由 (X_n) 与 (Y_n) 的等价性, 利用推论 6, 便有 (i) 式成立.

现在来证明 (♣) 式. 首先:

$$\sum_{n=1}^{\infty} \frac{\sigma^2(Y_n)}{n^2} \leqslant \sum_{n=1}^{\infty} \frac{\mathbb{E}(Y_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{|x| \leqslant n} x^2 \, dF(x).$$

接下来的估计难点在于: 我们不得不用一阶矩来估计二阶矩, 因为定理条件中只有一阶矩的信息. 标准技巧是对积分区域进行分划, 然后分别进行估计:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^{n} \int_{j-1<|x| \leqslant j} x^2 \, dF(x) &= \sum_{j=1}^{\infty} \int_{j-1<|x| \leqslant j} x^2 \, dF(x) \sum_{n=j}^{\infty} \frac{1}{n^2} \\
&\leqslant \sum_{j=1}^{\infty} j \int_{j-1<|x| \leqslant j} |x| \, dF(x) \cdot \frac{C}{j} \leqslant C \sum_{j=1}^{\infty} \int_{j-1<|x| \leqslant j} |x| \, dF(x) \\
&= C\mathbb{E}(|X_1|) < \infty.
\end{aligned}$$

其中我们使用了如下简单估计:

$$\sum_{n=j}^{\infty} \frac{1}{n^2} \leqslant \frac{C}{j}.$$

由此便完成了 (♣) 式的证明. (i) 得证!

(ii). 因 $\mathbb{E}(|X_1|)=+\infty$, 于是 $\forall\ A>0$, 我们有 $\mathbb{E}(|X_1|/A)=+\infty$. 于是再利用第三章定理 16 与 (X_n) 独立同分布的假设, 我们有

$$\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>An)=\sum_{n=1}^{\infty}\mathbb{P}(|X_1|>An)=+\infty.$$

现在由 Borel-Cantelli 引理 (第四章定理 16), 我们有

$$\mathbb{P}(|X_n|>An\ \text{i.o.})=1,\ \text{i.e.},\ \mathbb{P}(|S_n-S_{n-1}|/n>A\ \text{i.o.})=1.$$

若 $\omega\in\{|S_n-S_{n-1}|/n>A\ \text{i.o.}\}$, 即: $\forall\ n\in\mathbb{N}$, $\exists\ m\geqslant n$, s.t. $\omega\in\{|S_m-S_{m-1}|/m>A\}$. 于是

$$\frac{|S_m(\omega)|}{m}\geqslant\frac{A}{2}\ \text{或}\ \frac{|S_{m-1}(\omega)|}{m-1}\geqslant\frac{|S_{m-1}(\omega)|}{m}\geqslant\frac{A}{2}\ \text{至少居其一}.$$

这说明 ω 必属于无限项 $\{|S_n|/n\geqslant A/2\}$. 于是 $\omega\in\{|S_n|/n\geqslant A/2\ \text{i.o.}\}$. 这说明

$$\mathbb{P}(|S_n|/n\geqslant A/2\ \text{i.o.})=1.$$

于是, $\forall\ A>0$, 存在零测度集 $Z(A)$, s.t. $\forall\ \omega\in\Omega\backslash Z(A)$, $\varlimsup_{n\to\infty}|S_n(\omega)|/n\geqslant A/2$. 现在令 $Z=\bigcup_{m=1}^{\infty}Z(m)$, 易知 $\mathbb{P}(Z)=0$, 且 $\forall\ \omega\in\Omega\backslash Z$, $\varlimsup_{n\to\infty}|S_n(\omega)|/n\geqslant m/2$ 对一切 $m\in\mathbb{N}$ 成立, 于是 $\varlimsup_{n\to\infty}|S_n|/n=+\infty$ 几乎处处成立. (ii) 得证. ■

Extension of SLLN. (IID Case)

- Here is an interesting extension of the law of large numbers when the mean is infinite, due to Feller (1946).

定理 21 (Extension of SLLN. (IID Case))

Let (X_n) be as in Theorem 20 with $\mathbb{E}(|X_1|) = \infty$. Let (a_n) be a sequence of positive numbers satisfying the condition $a_n/n \uparrow$. Then we have

$$\overline{\lim_{n\to\infty}} \frac{|S_n|}{a_n} = 0 \text{ a.e.,} \quad \textit{or } = \infty \text{ a.e.} \tag{30}$$

according as [the proof of the case "$=\infty$ a.e." is similar to Theorem 20]

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant a_n) = \sum_{n=1}^{\infty} \int_{|x|\geqslant a_n} dF(x) < \infty, \quad \textit{or } = \infty. \tag{31}$$

Proof. 首先注意到 a_n/n 单调增加可推得 (a_n) 单调增加. 于是:

$$\int_{|x|\geqslant a_n} dF(x) = \sum_{k=n}^{\infty} \int_{a_k \leqslant |x| < a_{k+1}} dF(x).$$

于是 (31) 右侧级数可写为:

$$\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}\int_{a_k\leqslant|x|<a_{k+1}} dF(x)=\sum_{k=1}^{\infty}\sum_{n=1}^{k}\int_{a_k\leqslant|x|<a_{k+1}} dF(x)=\sum_{k=1}^{\infty}k\int_{a_k\leqslant|x|<a_{k+1}} dF(x).$$

于是 (31) 式中级数收敛当且仅当

$$\sum_{k=1}^{\infty}k\int_{a_k\leqslant|x|<a_{k+1}} dF(x)<\infty. \tag{32}$$

现在考虑 (31) 式中级数收敛的情形. 令

$$\mu_n=\int_{|x|<a_n} x\,dF(x),\qquad Y_n=\begin{cases} X_n-\mu_n, & \text{if } |X_n|<a_n;\\ -\mu_n, & \text{if } |X_n|\geqslant a_n.\end{cases}$$

于是由下式可知 $\mathbb{E}(Y_n)=0$:

$$\begin{aligned}\mathbb{E}(Y_n)&=\int_{\Omega}Y_n\,d\mathbb{P}=\int_{|X_n|<a_n}(X_n-\mu_n)\,d\mathbb{P}+\int_{|X_n|\geqslant a_n}(-\mu_n)\,d\mathbb{P}\\&=\int_{|X_n|<a_n}X_n\,d\mathbb{P}-\mu_n\mathbb{P}(|X_n|<a_n)-\mu_n\mathbb{P}(|X_n|\geqslant a_n)\\&=\int_{|x|<a_n}x\,dF(x)-\mu_n\big[\mathbb{P}(|X_n|<a_n)+\mathbb{P}(|X_n|\geqslant a_n)\big]=0.\end{aligned}$$

另外, 由 Y_n 的定义与 (31) 式, 知

$$\sum_{n=1}^{\infty} \mathbb{P}(Y_n \neq X_n - \mu_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant a_n) < \infty. \tag{33}$$

再设 $a_0 = 0$, 于是

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{E}\left(\frac{Y_n^2}{a_n^2}\right) &\leqslant \sum_{n=1}^{\infty} \frac{1}{a_n^2} \int_{|x|<a_n} x^2\, dF(x) \quad \text{Verify it by yourself} \\
&= \sum_{n=1}^{\infty} \frac{1}{a_n^2} \sum_{k=1}^{n} \int_{a_{k-1} \leqslant |x| < a_k} x^2\, dF(x) \\
&\leqslant \sum_{k=1}^{\infty} \int_{a_{k-1} \leqslant |x| < a_k} dF(x)\, a_k^2 \sum_{n=k}^{\infty} \frac{1}{a_n^2}. \quad (\star)
\end{aligned}$$

因为 $a_n/n \uparrow$, 于是对 $k \leqslant n$, 有 $a_n/n \geqslant a_k/k$. 于是 $n^2/a_n^2 \leqslant k^2/a_k^2 \;\Rightarrow\; 1/a_n^2 \leqslant k^2/(a_k^2 n^2)$. 于是有

$$\sum_{n=k}^{\infty} \frac{1}{a_n^2} \leqslant \frac{k^2}{a_k^2} \sum_{n=k}^{\infty} \frac{1}{n^2} \leqslant \frac{2k}{a_k^2}.$$

将上式代入 ($\star$) 式, 利用 (32) 式, 我们有

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\frac{Y_n^2}{a_n^2}\right) \leqslant \sum_{k=1}^{\infty} 2k \int_{a_{k-1} \leqslant |x| < a_k} dF(x) < \infty.$$

于是由定理 18 与 Kronecker 引理 17, 我们有

$$\frac{1}{a_n} \sum_{k=1}^{n} Y_k \to 0 \text{ a.e.} \tag{34}$$

由 (33) 式我们可知 (Y_n) 与 $(X_n - \mu_n)$ 等价, 于是由推论 6 可知

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \mu_k) \to 0 \text{ a.e.}$$

于是要证 (30) 式左端就只需证 $a_n^{-1} \sum_{k=1}^{n} \mu_k \to 0$. 首先, 我们有

$$\frac{1}{a_n} \sum_{k=1}^{n} \mu_k = \frac{1}{a_n} \sum_{k=1}^{n} \int_{|x| < a_k} x \, dF(x). \tag{35}$$

现在任取 $N<n$, 我们有

$$(35)=\frac{1}{a_n}\sum_{k=1}^{n}\left[\int_{|x|\leqslant a_N}+\int_{a_N<|x|<a_k}\right]=\frac{n}{a_n}\int_{|x|\leqslant a_N}+\frac{1}{a_n}\sum_{k=1}^{n}\int_{a_N<|x|<a_n}.$$

注意到上式右端最后一项与 k 无关, 于是

$$\begin{aligned}\left|\frac{1}{a_n}\sum_{k=1}^{n}\mu_k\right| &\leqslant \frac{n}{a_n}\left(\int_{|x|\leqslant a_N}|x|\,dF(x)+\int_{a_N<|x|<a_n}|x|\,dF(x)\right)\\ &\leqslant \frac{n}{a_n}\left(a_N\,\mathbb{P}(|X_1|\leqslant a_N)+\int_{a_N<|x|<a_n}|x|\,dF(x)\right)\\ &\leqslant \frac{n}{a_n}\left(a_N+\int_{a_N<|x|<a_n}|x|\,dF(x)\right).\end{aligned}\tag{36}$$

又因 $\mathbb{E}(|X_1|)=\infty$, 若 $a_n/n\uparrow M<\infty$, 则 $a_n/n\leqslant M$. 故 $\mathbb{P}(|X_n|\geqslant a_n)\geqslant\mathbb{P}(|X_n|\geqslant Mn)$. 由第三章定理 16 可知

$$M^{-1}\mathbb{E}(|X_1|)=\mathbb{E}\left(\frac{|X_1|}{M}\right)\leqslant 1+\sum_{n=1}^{\infty}\mathbb{P}(|X_1|\geqslant Mn)\leqslant 1+\sum_{n=1}^{\infty}\mathbb{P}(|X_1|\geqslant a_n),$$

再由 (X_n) 独立同分布可看出 $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|\geqslant a_n)=\infty$. 这与 (33) 式矛盾, 于是 $a_n/n\uparrow\infty$.

因 $a_n/n\uparrow\infty$, 于是 $n/a_n\downarrow 0$. 现在, 对于固定的 N, 我们总有 $(na_N)/a_n\downarrow 0$ $(n\to\infty)$. 另外,

$$
\begin{aligned}
\frac{n}{a_n}\int_{a_N<|x|<a_n}|x|\,dF(x) &= \frac{n}{a_n}\sum_{j=N+1}^{n}\int_{a_{j-1}\leqslant|x|<a_j}|x|\,dF(x)\\
&\leqslant \frac{n}{a_n}\sum_{j=N+1}^{n}a_j\int_{a_{j-1}\leqslant|x|<a_j}dF(x)\leqslant \sum_{j=N+1}^{n}j\int_{a_{j-1}\leqslant|x|<a_j}dF(x)
\end{aligned}
$$

其中上式第二行不等式成立是因为 $a_j/j\leqslant a_n/n$ $(\forall\ j\leqslant n)$. 现在 (36) 式变为

$$
\left|\frac{1}{a_n}\sum_{k=1}^{n}\mu_k\right|\leqslant\frac{na_N}{a_n}+\sum_{j=N+1}^{n}j\int_{a_{j-1}\leqslant|x|<a_j}dF(x).
$$

在上式中令 $n\to\infty$, 有 Back to Content.

$$
\forall\ N\in\mathbb{N}:\quad \lim_{n\to\infty}\left|\frac{1}{a_n}\sum_{k=1}^{n}\mu_k\right|\leqslant\sum_{j=N+1}^{\infty}j\int_{a_{j-1}\leqslant|x|<a_j}dF(x).
$$

由 (32) 式, 在上式右端令 $N\to\infty$, 便可知 $\lim\limits_{n\to\infty}a_n^{-1}\sum\limits_{k=1}^{n}\mu_k=0$. 定理得证. ■

Section 5

* Applications

- This section focuses on two applications: empirical distributions and renewal processes.
- The first purpose in sampling theory is to approximate the true (and often *unknown*) distribution function with the observed empirical distribution function.
- The renewal process is the first stochastic process we encounter, and we will systematically discuss the definition and related properties of the renewal process.

Empiric Distributions

- Let (X_n) be a sequence of independent, identically distributed r.v.'s with common d.f. F.
- F is sometimes referred to as the "underlying" or "theoretical distribution" and is regarded as "unknown" in statistical lingo.
- For each ω, the values $X_n(\omega)$ are called "**samples**" or "observed values".
- The idea is to get some information on F by looking at the samples.
- For each n, and each $\omega \in \Omega$, let n real numbers $(X_i(\omega))_{1 \leqslant i \leqslant n}$ be arranged in increasing order as

$$Y_{n1}(\omega) \leqslant Y_{n2}(\omega) \leqslant \cdots \leqslant Y_{nn}(\omega). \tag{37}$$

Now define a discrete d.f. $F_n(\cdot, \omega)$ as follows:

$$\begin{aligned} F_n(x,\omega) &= 0, \quad \text{if } x < Y_{n1}(\omega); \\ &= \frac{k}{n}, \quad \text{if } Y_{nk}(\omega) \leqslant x < Y_{n,k+1}(\omega),\ 1 \leqslant k \leqslant n-1; \\ &= 1, \quad \text{if } x \geqslant Y_{nn}(\omega). \end{aligned}$$

In other words, for each x, $nF_n(x,\omega)$ is the number of values of i $(1 \leqslant i \leqslant n)$, for which $X_i(\omega) \leqslant x$; or again $F_n(x,\omega)$ is the observed frequency of sample values not exceeding x. The function $F_n(\cdot,\omega)$ is called the **empiric distribution function** (经验分布函数) **based on n samples from F.**

- For each x, $F_n(x,\cdot)$ is an r.v. Let us introduce r.v.'s $\{\xi_i(x), i \geqslant 1\}$ as follows:

$$\xi_i(x,\omega) = \begin{cases} 1, & \text{if } X_i(\omega) \leqslant x; \\ 0, & \text{if } X_i(\omega) > x. \end{cases}$$

  We have then

$$F_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(x,\omega).$$

- For each x, the sequence $\{\xi_i(x)\}$ is totally independent since (X_i) is. Furthermore they have common "Bernoulli distribution":

$$\xi_i(x) \sim \begin{bmatrix} 1 & 0 \\ p = F(x) & q = 1 - F(x) \end{bmatrix}.$$

  Thus $\mathbb{E}\xi_i(x) = F(x)$. The strong law of large numbers in the form Theorem 20 applies, and we concluded that

$$F_n(x,\omega) \to F(x) \text{ a.e. } (\mathbb{P}) \tag{38}$$

- Matters end here if we are interested only in a particular value of x, or a finite number of values, but since both members in the whole line.

Glivenko-Cantelli Theorem

- How much better it would be to make a global statement about the *functions* $F_n(\cdot,\omega)$ and $F(\cdot)$?
- First, let us observe the precise meaning of (38): for each x, there exists a null set $N(x)$ such that (38) holds for $\omega \in \Omega \backslash N(x)$. It follows that (38) also holds simultaneously for all x in any given dense countable set Q, such as the set of rational numbers, for $\omega \in \Omega \backslash N$, where
$$N = \bigcup_{x \in Q} N(x)$$
is again a null set. Hence by the definition of vague convergence in Chapter 4, we can already assert that
$$F_n(\cdot,\omega) \xrightarrow{v} F(\cdot) \text{ for a.e. } \omega.$$
- This will be further strengthened in two ways: convergence for all x and uniformity. The result is due to Glivenko and Cantelli.

定理 22 (Glivenko-Cantelli Theorem)

We have as $n \to \infty$
$$\sup_{-\infty < x < \infty} |F_n(x,\omega) - F(x)| \to 0 \text{ a.e. } (\mathbb{P})$$

PROOF. 设 J 是分布函数 F 的跳跃点构成的集合, 则 J 至多可数. $\forall\ x \in J$, 定义

$$\eta_i(x,\omega) = \begin{cases} 1, & \text{if } x = X_i(\omega); \quad x \text{ 恰为某一样本} \\ 0, & \text{if } x \neq X_i(\omega). \end{cases}$$

于是 $\forall\ x \in J$:

$$F_n(x,\omega) - F_n(x-,\omega) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(x,\omega).$$

事实上, $\sum_{i=1}^{n} \eta_i(x,\omega)$ 统计了恰好等于 x 的样本 (抽样) 的个数. 由 (38) 式, $\forall\ x \in J$, 存在零测度集 $N(x)$, s.t. $\forall\ \omega \in \Omega \backslash N(x)$, 有

$$F_n(x,\omega) - F_n(x-,\omega) \to F(x) - F(x-). \tag{39}$$

现在令 $N_1 = \bigcup_{x \in Q \cup J} N(x)$, 于是 N_1 也是零测度集. 若 $\omega \in \Omega \backslash N_1$, 则 (39) 式对一切 $x \in J$ 成立, 并且

$$F_n(x,\omega) \to F(x) \tag{40}$$

对一切 $x \in Q$ 成立. 由如下引理, 我们便完成了 Glivenko-Cantelli 定理的证明. ■

引理 23

Let F_n and F be (right continuous) d.f.'s, Q and J as before. Suppose that we have

$$
\begin{aligned}
\forall\ x \in Q: &\quad F_n(x) \to F(x);\\
\forall\ x \in J: &\quad F_n(x) - F_n(x-) \to F(x) - F(x-).
\end{aligned}
$$

Then $F_n \to F$ uniformly in $\mathbb{R}$. $\forall\ \varepsilon > 0$, $\exists\ N > 0$, *s.t.* $\forall\ n \geqslant N$, $\forall\ x \in \mathbb{R}$: $|F_n(x) - F(x)| < \varepsilon$.

PROOF. **反证法**. 若 F_n 不一致收敛于 F: $\exists\ \varepsilon > 0$, $\forall\ n \in \mathbb{N}$, $\exists\ n_k \geqslant n$, $\exists\ x_k \in \mathbb{R}$, s.t.

$$|F_{n_k}(x_k) - F(x_k)| \geqslant \varepsilon > 0.$$

事实上, 序列 (x_k) 不可能收敛于 $\pm\infty$, 因为

$$F_n(\infty) = F(\infty) = 1, \quad F_n(-\infty) = F(-\infty) = 0.$$

这就说明 (x_k) 是有界数列. 于是存在 (x_k) 的收敛子列, 不妨仍记作 (x_k), 满足

$$x_k \to \xi \in \mathbb{R}.$$

现在对于充分大的 k, 对于上述 $\xi \in \mathbb{R}$, $\exists\ r_1, r_2 \in Q$ 且 $r_1 < \xi < r_2$, 有如下四种情形:

Case I. $x_k \uparrow \xi$ 且 $x_k < \xi$:

$$\begin{aligned}\varepsilon &\leqslant F_{n_k}(x_k) - F(x_k) \leqslant F_{n_k}(\xi-) - F(r_1)\\ &\leqslant F_{n_k}(\xi-) - F_{n_k}(\xi) + F_{n_k}(r_2) - F(r_2) + F(r_2) - F(r_1).\end{aligned}$$

Case II. $x_k \uparrow \xi$ 且 $x_k < \xi$:

$$\begin{aligned}\varepsilon &\leqslant F(x_k) - F_{n_k}(x_k) \leqslant F(\xi-) - F_{n_k}(r_1)\\ &= F(\xi-) - F(r_1) + F(r_1) - F_{n_k}(r_1).\end{aligned}$$

Case III. $x_k \downarrow \xi$ 且 $x_k \geqslant \xi$:

$$\begin{aligned}\varepsilon &\leqslant F(x_k) - F_{n_k}(x_k) \leqslant F(r_2) - F_{n_k}(\xi)\\ &\leqslant F(r_2) - F(r_1) + F(r_1) - F_{n_k}(r_1) + F_{n_k}(\xi-) - F_{n_k}(\xi).\end{aligned}$$

Case IV. $x_k \downarrow \xi$ 且 $x_k \geqslant \xi$:

$$\begin{aligned}\varepsilon &\leqslant F_{n_k}(x_k) - F(x_k) \leqslant F_{n_k}(r_2) - F(\xi)\\ &\leqslant F_{n_k}(r_2) - F_{n_k}(r_1) + F_{n_k}(r_1) - F(r_1) + F(r_1) - F(\xi).\end{aligned}$$

在上述每种情况中令 $k \to \infty$, 再令 $r_1 \uparrow \xi$, $r_2 \downarrow \xi$. 于是各个不等式的最后一项均趋于零, 与反证假设矛盾! 命题得证. ■

Definition of Renewal Processes

定义 24 (Renewal Processes)

Let (X_n) be a sequence of independent and identically distributed r.v.'s, with $0 < X_n < \infty$. Then the process

$$N(t) := \sup\{n : T_n \leqslant t\}, \quad \text{where } T_n = \sum_{i=1}^{n} X_i, \tag{41}$$

is said to be a **renewal process (更新过程)**.

- For a concrete situation, consider a diligent janitor who replaces a light bulb the instant it burns out. Suppose the first bulb is put in at time 0 and let X_n be the lifetime of the nth light bulb. In this interpretation, T_n is the time the nth light bulb burns out and $N(t)$ is the number of light bulbs that have burnt out by time t.
- The **distribution** of $N(t)$ can be obtained, at least in theory, by first noting the important relationship that the number of renewals by time t is greater than or equal to n iff, the nth renewal occurs before or at time t:

$$N(t) \geqslant n \;\Leftrightarrow\; T_n \leqslant t. \tag{42}$$

From equation (42) we obtain

$$\mathbb{P}\{N(t) = n\} = \mathbb{P}\{N(t) \geqslant n\} - \mathbb{P}\{N(t) \geqslant n+1\} = \mathbb{P}(T_n \leqslant t) - \mathbb{P}(T_{n+1} \leqslant t).$$

Renewal and Interarrival Times

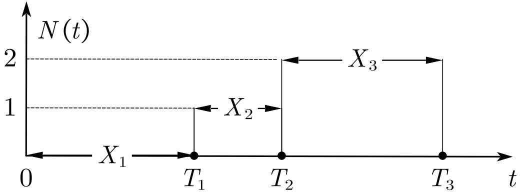


图 1: Renewal and interarrival times

In this section we will examine the following questions:

1. the asymptotic behavior of the number of renewals;
2. the asymptotic behavior of the renewal function (Wald's equation);
3. to further investigate the moment of renewal arrival, we introduce renewal measures and discuss Blackwell's renewal theorem; (Omitted)
4. finally, we discuss the renewal equation. (Omitted)

For more information, the reader is referred to [Dur19] Sec. 2.6; [GGDS20] Chap. 10, P447; [Fel91] Chap. XI.

1. Asymptotic Behavior of the Number of Renewals

定理 25 (Asymptotic Behavior of the Number of Renewals)

Suppose $N(t)$ *is a renewal process,* $0 < \mu = \mathbb{E}(X_1) \leqslant \infty$, *then*

i). $\lim\limits_{t\to\infty} N(t) = +\infty$ *a.e.;*

ii). $\lim\limits_{n\to\infty} T_n/n = \mu$ *a.e.;* *Using the Strong Law of Numbers (Theorem 20)*

iii). $\lim\limits_{t\to\infty} T_{N(t)}/N(t) = \mu$ *a.e..*

Remark (Asymptotic Behavior of the Number of Renewals)

Using the strong law of numbers, we can obtain theorem 25 (ii), from which we can derive the following conclusion:

- $\mu > 0 \Rightarrow T_n \to \infty$ as $n \to \infty$;
- since $\forall\ n:\ 0 < X_n < \infty$, we have $T_n \to \infty \Rightarrow n \to \infty$.

Together, these two points indicate that *only a finite number of renewals can occur in a finite period of time*. But we also indicate that *the total number of renewals becomes infinite with time* (according to (i)).

PROOF of Theorem 25. (i). 不妨设 $N(\infty)=\lim_{t\to\infty}N(t)$, 于是我们有

$$\begin{aligned}\mathbb{P}\{N(\infty)<\infty\}&=\mathbb{P}\{X_n=\infty \text{ for some } n\}=\mathbb{P}\left\{\bigcup_{n=1}^{\infty}\{X_n=\infty\}\right\}\\&\leqslant\sum_{n=1}^{\infty}\mathbb{P}\{X_n=\infty\}=0. \quad [\text{sub-additivity}]\end{aligned}$$

其中最后一个等号成立是因为 $0<X_n<\infty$ 恒成立.

(iii). 由 (i), 存在零测集 Z_1, s.t. $\forall\ \omega\in\Omega\backslash Z_1$, $\lim_{t\to\infty}N(t,\omega)=+\infty$. 由 (ii), 存在零测集 Z_2, s.t. $\forall\ \omega\in\Omega\backslash Z_2$, $\lim_{n\to\infty}T_n(\omega)/n=\mu$. 现在对任一 $\omega_0\in\Omega$, 若数列 $\{a_n(\omega_0),n\geqslant 1\}$ 收敛到极限 (有限或无限) μ, 同时 $\lim_{t\to\infty}N(t,\omega_0)=+\infty$, 于是由数列极限的 Heine 原理,

$$a_{N(t,\omega_0)}(\omega_0)\to\mu \quad (t\to\infty).$$

现在不妨设 $a_n=n^{-1}\sum_{i=1}^{n}X_i$, 对任一 $\omega\in\Omega\backslash(Z_1\cup Z_2)$ 应用上述结论, 我们有

$$\lim_{t\to\infty}\frac{T_{N(t,\omega)}(\omega)}{N(t,\omega)}=\mu,$$

注意到 $\mathbb{P}(Z_1\cup Z_2)=0$, 于是 (iii) 成立. ■

Rates of Divergence of the Number of Renewals

- Based on Theorem 25, we want to know *at what rate does $N(t)$ tend to infinity*?

定理 26 (Rates of Divergence of the Total Number of Renewals)

Suppose $N(t)$ is a renewal process. If $\mathbb{E}(X_1) = \mu \leqslant \infty$, then as $t \to \infty$,

$$\frac{N(t)}{t} \to \frac{1}{\mu} \text{ a.e.} \quad (1/\infty = 0)$$

The number $1/\mu$ is called the ***rate*** *of the renewal process.*

Proof. 由更新过程 $N(t)$ 的定义, 有 $T_{N(t)} \leqslant t < T_{N(t)+1}$. 在不等式中同时除以 $N(t)$, 于是有

$$\frac{T_{N(t)}}{N(t)} \leqslant \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}.$$

由定理 25 (iii) 可知: $T_{N(T)}/N(t) \to \mu$ a.e., 同理, $T_{N(t)+1}/(N(t)+1) \to \mu$ a.e.. 在上式两端同时令 $t \to \infty$, 便有

$$\frac{t}{N(t)} \to \frac{1}{\mu} \text{ a.e.}$$

于是结论得证. ■

2. Wald's Equation and Renewal Function

- Departing slightly from the notation in Definition 24, we let

$$N(t) = \inf\{n : T_n > t\}. \tag{43}$$

  $N(t)$ is the number of renewals in $[0, t]$, counting the renewal at time 0.
- The advantage of this definition is that $N(t)$ is a stopping time[4], i.e., $\{N(t) \leqslant k\}$ is measurable with respect to $\mathscr{F}_k$, where $\mathscr{F}_k$ is the σ-algebra generated by $\{X_i, 1 \leqslant i \leqslant k\}$.
- Our next result concerns the asymptotic behavior of $U(t) = \mathbb{E}[N(t)]$ (**Renewal Function**). To derive the result we need:

定理 27 (Wald's Equation)

Let (X_n) be i.i.d. with $\mathbb{E}(|X_1|) < \infty$. If N is a stopping time with $\mathbb{E}(N) < \infty$, then

$$\mathbb{E}(X_1 + \cdots + X_N) =: \mathbb{E}(T_N) = \mathbb{E}(X_1)\mathbb{E}(N).$$

[4] The concept of stopping time will be encountered frequently in stochastic analysis (Martingale).

PROOF. **Case 1.** $X_n \geqslant 0$.

$$\infty > \mathbb{E}(T_N) = \int_\Omega T_N \, d\mathbb{P} = \sum_{n=1}^{\infty} \int_\Omega T_n \mathbf{1}_{\{N=n\}} \, d\mathbb{P} = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \int_\Omega X_m \mathbf{1}_{\{N=n\}} \, d\mathbb{P} \text{ [正项级数收敛可换序]}$$

$$= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int_\Omega X_m \mathbf{1}_{\{N=n\}} \, d\mathbb{P} = \sum_{m=1}^{\infty} \int_\Omega X_m \mathbf{1}_{\{N \geqslant m\}} \, d\mathbb{P}.$$

现在, $\{N \geqslant m\} = \{N \leqslant m-1\}^c \in \mathscr{F}_{m-1} = \sigma(X_1, \cdots, X_{m-1})$, 于是 $\mathbf{1}_{\{N \geqslant m\}}$ 独立于 X_m. 现在, 利用第三章推论 17, 我们有

$$\mathbb{E}(T_N) = \sum_{m=1}^{\infty} \mathbb{E}(X_m)\mathbb{P}(N \geqslant m) = \mathbb{E}(X_1)\mathbb{E}(N).$$

Case 2. 一般情形. 我们反过来推导. 若 $\mathbb{E}(N) < \infty$, 于是

$$\infty > \sum_{m=1}^{\infty} \mathbb{E}(|X_m|)\mathbb{P}(N \geqslant m) = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int_\Omega |X_m| \mathbf{1}_{\{N=n\}} \, d\mathbb{P}.$$

这说明上述级数绝对收敛, 绝对收敛级数可重排, 于是有

$$\mathbb{E}(T_N) = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \int_\Omega X_m \mathbf{1}_{\{N=n\}} \, d\mathbb{P} = \sum_{n=1}^{\infty} \sum_{n=m}^{\infty} \int_\Omega X_m \mathbf{1}_{\{N=n\}} \, d\mathbb{P}.$$

再由类似地步骤便可证明 $\mathbb{E}(T_N) = \mathbb{E}(X_1)\mathbb{E}(N)$. 定理得证. ∎

Properties of the Renewal Function

- Using Wald's equation for the renewal process requires that $U(t)=\mathbb{E}[N(t)]<\infty$. So we also need to stop and prove the following lemmas. We divide this into two steps.

引理 28 (Representation of the Renewal Function)

Let T_n and $N(t)$ be as in Definition 24, then

$$U(t)=\mathbb{E}[N(t)]=\sum_{n=1}^{\infty}\mathbb{P}(T_n\leqslant t). \tag{44}$$

PROOF. Technique 1.

$$U(t)=\mathbb{E}[N(t)]=\sum_{n=1}^{\infty}n\,\mathbb{P}\{N(t)=n\}=\sum_{n=1}^{\infty}\sum_{m=1}^{n}\mathbb{P}\{N(t)=n\}$$
$$=\sum_{m=1}^{\infty}\sum_{n=m}^{\infty}\mathbb{P}\{N(t)=n\}=\sum_{m=1}^{\infty}\mathbb{P}\{N(t)\geqslant m\}=\sum_{m=1}^{\infty}\mathbb{P}(T_m\leqslant t).$$

Technique 2.

$$
\begin{aligned}
U(t) = \mathbb{E}[N(t)] &= \sum_{n=0}^{\infty} n\mathbb{P}\{N(t) = n\} \\
&= \sum_{n=0}^{\infty} n\left[\mathbb{P}\{N(t) \geqslant n\} - \mathbb{P}\{N(t) \geqslant n+1\}\right] \\
&= \sum_{n=0}^{\infty} n\left[\mathbb{P}(T_n \leqslant t) - \mathbb{P}(T_{n+1} \leqslant t)\right] \\
&= \sum_{n=0}^{\infty} n\mathbb{P}(T_n \leqslant t) - \sum_{n=0}^{\infty} n\mathbb{P}(T_{n+1} \leqslant t) \\
&= \sum_{n=1}^{\infty} n\mathbb{P}(T_n \leqslant t) - \sum_{m=1}^{\infty} (m-1)\mathbb{P}(T_m \leqslant t) \\
&= \sum_{m=1}^{\infty} [n-(n-1)]\mathbb{P}(T_n \leqslant t) = \sum_{n=1}^{\infty} \mathbb{P}(T_n \leqslant t). \qquad \blacksquare
\end{aligned}
$$

引理 29 (The Expectation of Renewal Function)

Let T_n and $N(t)$ be as in Definition 24, then $U(t) = \mathbb{E}[N(t)] < \infty$.

* Proof. 由定义, 有 $\mathbb{P}(X_n > 0) > 0$, 则存在 $a > 0$ s.t. $\mathbb{P}(X_n \geqslant a) > 0$, 从而 $\mathbb{P}(X_n < a) < 1$. 而

$$\mathbb{P}(X_n \leqslant a) = \mathbb{P}(X_n < a) + \mathbb{P}(X_n = a).$$

不妨设 $0 < b < a$, 显然有

$$\mathbb{P}(X_n \leqslant b) \leqslant \mathbb{P}(X_n < a) < 1.$$

对任意固定的 t, 一定存在正整数 k, 使得 $kb \geqslant t$, 所以

$$\{T_k \leqslant t\} \subseteq \{T_k \leqslant kb\}.$$

另外, 因为 $\{X_1 + \cdots + X_k \leqslant kb\}^c = \{X_1 + \cdots + X_k > kb\}$, 而

$$\{X_1 > b, \cdots, X_k > b\} \subseteq \{X_1 + \cdots + X_k > kb\},$$

故 $\{X_1 + \cdots + X_k > kb\}^c \subseteq \{X_1 > b, \cdots, X_k > b\}^c$. 又因为

$$\{X_1 + \cdots + X_k > kb\}^c = \{X_1 + \cdots + X_k \leqslant kb\} = \{T_k \leqslant kb\},$$

所以

$$\{T_k \leqslant t\} \subseteq \{T_k \leqslant kb\} \subseteq \{X_1 > b, \cdots, X_k > b\}^c.$$

于是

$$\mathbb{P}(T_k \leqslant t) \leqslant 1 - \mathbb{P}(X_1 > b, \cdots, X_k > b) = 1 - [\mathbb{P}(X_1 > b)]^k =: 1 - \beta,$$

其中 $\beta = [\mathbb{P}(X_1 > b)]^k > 0$. 又因为 (X_n) 是独立同分布且取值为非负实数的随机变量序列，所以对任意 $1 \leqslant \ell \leqslant m$，必有

$$T_{\ell k} - T_{(\ell-1)k} = \sum_{i=(\ell-1)k+1}^{\ell k} X_i \leqslant T_{mk} = \sum_{i=1}^{mk} X_i.$$

故当 $T_{mk} < t$ 时，必有 $T_{\ell k} - T_{(\ell-1)k} \leqslant t$，$\ell = 1, 2, \cdots, m$. 所以

$$\{T_{mk} \leqslant t\} \subseteq \{T_k - T_0 \leqslant t, T_{2k} - T_k \leqslant t, \cdots, T_{mk} - T_{(m-1)k} \leqslant t\}.$$

再由 (X_n) 的独立性，由第三章定理 29，我们有

$$\mathbb{P}(T_k - T_0 \leqslant t, T_{2k} - T_k \leqslant t, \cdots, T_{mk} - T_{(m-1)k} \leqslant t) = [\mathbb{P}(T_k \leqslant t)]^m.$$

从而有

$$\mathbb{P}(T_{mk} \leqslant t) \leqslant [\mathbb{P}(T_k \leqslant t)]^m \leqslant (1-\beta)^m.$$

对任意的 $j \in \mathbb{N}$，有

$$\{T_{mk+j} \leqslant t\} \subseteq \{T_{mk} \leqslant t\},$$

所以

$$\sum_{n=mk}^{(m+1)k-1} \mathbb{P}(T_n \leqslant t) \leqslant k\mathbb{P}(T_{mk} \leqslant t).$$

综上可得

$$
\begin{aligned}
U(t) &= \sum_{n=1}^{\infty} \mathbb{P}(T_n \leqslant t) \\
&= \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + \sum_{n=k}^{\infty} \mathbb{P}(T_n \leqslant t) \\
&= \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + \sum_{m=1}^{\infty} \sum_{j=0}^{k-1} \mathbb{P}(T_{mk+j} \leqslant t) \\
&= \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + \sum_{m=1}^{\infty} k\mathbb{P}(T_{mk} \leqslant t) \\
&= \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + \sum_{m=1}^{\infty} k(1-\beta)^m \\
&= \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + k\frac{1-\beta}{1-(1-\beta)} \\
&< \sum_{n=1}^{k-1} \mathbb{P}(T_n \leqslant t) + \frac{k}{\beta} < \infty.
\end{aligned}
$$

命题得证. ■

Elementary Renewal Theorem

- The following theorem discusses the asymptotic behavior of the renewal function.

定理 30 (Elementary Renewal Theorem)

As $t \to \infty$, $U(t)/t \to 1/\mu$.

Remark

At first glance it might seem that the elementary renewal theorem should be a simple consequence of Theorem 26. That is, since the average renewal rate will, with probability 1, converge to $1/\mu$, should this not imply that the expected average renewal rate also converges to $1/\mu$? We must, however, be careful; consider the next example.

例 31

Let U be a r.v. which is uniformly distributed on $(0,1)$; and define the random variables Y_n, $n \geqslant 1$, by $Y_n = n\mathbf{1}_{\{U \leqslant 1/n\}}$. It is obvious that $Y_n \to 0$ a.e. as $n \to \infty$. However,

$$\mathbb{E}(Y_n) = n\mathbb{P}(U \leqslant 1/n) = 1.$$

Therefore, even though the sequence of r.v.'s Y_n converges to 0, the expected values of the Y_n are all indentically 1.

Proof of Theorem 30. 证明分三步走. Step 1. 证明 $\varliminf_{t\to\infty} U(t)/t \geqslant 1/\mu$. 注意到 $T_{N(t)+1}$ 是时刻 t 之后第一次更新的时刻, 于是 $T_{N(t)+1} = t + Y(t)$, 其中 $Y(t)$ 称为超额寿命, 其由 t 至下次更新的时间间隔定义. 由 Wald 等式,

$$\mathbb{E}[T_{N(t)+1}] = \mathbb{E}(X_1)\mathbb{E}[N(t)+1] = \mathbb{E}(X_1)\mathbb{E}[N(t)] + \mathbb{E}(X_1) = \mu[U(t)+1].$$

再注意到 $\mathbb{E}[T_{N(t)+1}] = t + \mathbb{E}[Y(t)]$, 于是 $t + \mathbb{E}[Y(t)] = \mu[U(t)+1]$. 化简可得

$$\frac{U(t)}{t} = \frac{1}{\mu} + \frac{\mathbb{E}[Y(t)]}{t\mu} - \frac{1}{t}.$$

由于 $Y(t) \geqslant 0$, 于是 $U(t)/t \geqslant 1/\mu - 1/t$. 于是可得 $\varliminf_{t\to\infty} U(t)/t \geqslant 1/\mu$.

Step 2. 证明 $\varlimsup_{t\to\infty} U(t)/t \leqslant 1/\mu$. 首先设更新间隔 X_n 是几乎处处一致有界的, 即存在 $M > 0$, s.t. $\mathbb{P}(X_n \leqslant M) = 1$. 由此可知 $\mathbb{P}\{Y(t) < M\} = 1$, 于是 $\mathbb{E}[Y(t)] < M$: (积分的绝对连续性)

$$\mathbb{E}[Y(t)] = \int_{Y(t)<M} Y(t)\, d\mathbb{P} + \int_{Y(t)\geqslant M} Y(t)\, d\mathbb{P} \leqslant M\,\mathbb{P}\{Y(t) < M\} + 0 \leqslant M.$$

从而

$$\frac{U(t)}{t} \leqslant \frac{1}{\mu} + \frac{M}{t\mu} - \frac{1}{t} \quad \Rightarrow \quad \varlimsup_{t\to\infty} \frac{U(t)}{t} \leqslant \frac{1}{\mu}.$$

再设更新间隔 X_n 不是几乎处处一致有界的, 此时对于固定的 $M>0$, 定义

$$\overline{X}_n = X_n \wedge M = \min\{X_n, M\} = \begin{cases} X_n, & \text{if } X_n \leqslant M; \\ M, & \text{if } X_n > M. \end{cases}$$

再定义 $\overline{N}(t)$ 为 $(\overline{X}_n)$ 所对应的更新过程. 现在有 $\overline{N}(t) \geqslant N(t)$, 这是因为 $\overline{N}(t)$ 的每个更新间隔的时间不大于 $N(t)$ 对应的更新间隔时间, 直至时刻 t 它必须至少有一样多的更新. 因此 $\mathbb{E}[N(t)] \leqslant \mathbb{E}[\overline{N}(t)]$. 从而

$$\varlimsup_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} \leqslant \varlimsup_{t\to\infty} \frac{\mathbb{E}[\overline{N}(t)]}{t} = \frac{1}{\mathbb{E}[X_n \wedge M]}.$$

其中上式中的等号利用了 $\overline{N}(t)$ 的更新间隔有界的事实. 现在

$$\lim_{M\to\infty} \mathbb{E}[X_n \wedge M] = \mathbb{E}(X_n) = \mu,$$

于是在上式中令 $M\to\infty$, 我们就有 $\varlimsup\limits_{t\to\infty} \mathbb{E}[N(t)]/t \leqslant 1/\mu$.

现在, 结合第一步与第二步, 便可知 $U(t)/t \to 1/\mu$. 定理得证. ■

‣ Back to Content.

Reference

- Chung, Kai Lai. *A Course in Probability Theory*, Third Edition. Academic Press, 2001.
- Durrett, Rick. *Probability: Theory and Examples*, 5th edition. Cambridge university press, 2019.
- Grimmett, Geoffrey, and David Stirzaker. *Probability and Random Processes*, 4th edition. Oxford university press, 2020.
- Feller, William. *An Introduction to Probability Theory and Its Applications*, Volume 2. Vol. 81. John Wiley & Sons, 1991.
- Varadhan, SR Srinivasa. *Probability theory*. No. 7. American Mathematical Soc., 2001.

FINANCIAL TECHNOLOGY

The Thrust nurtures creative minds to explore frontiers in Financial Technology, to shape the future financial industry that promotes social-financial health.

香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

社会枢纽
SOCIETY HUB

Society Hub • Financial Technology Thrust