《概率论》笔记

Ver 1.414

Fiddie

April 9, 2024

Contents

注: 这是 2018-2019 春季学期大二**数学系《概率论基础》课程**笔记, 上课老师是宋玉林, 笔记包括他上课的内容 (并作了重新组织) 以及丘赛试题、MSE 上的题等等.

本人水平有限, 内容难免有错, 如果遇到有疑问的地方可以加微信 Fiddie_Math 进行交流, 如果需要下面的参考书, 也可以加微信交流.

参考书:

- 李贤平《概率论基础》(第三版), 复旦. (我们的教材)
- 陈希孺《概率论与数理统计》. (据说不错)
- 盛骤《概率论与数理统计》(第四版), 浙大. (可以简单看看它的习题集)
- 严加安, 《测度论基础》. (学到后面条件期望可以选看, 但下个学期不讲条件期望, 不作要求)
- Dekking. A modern Introduction to Probability and Statistics.
- **(推荐)**Grimmet, Stirzaker,. One Thousand Exercises in Probability. (里面有很多不错的题, 建议挑一些题来做)
- **(推荐)**Ash. Probability & Measure Theory, Second Edition. (这本书也有不错的例子, 也有习题. 太深的内容可以不看)

学习方法:

- 所有定理的思想都要掌握 (尤其是 Chebyshev 不等式、Nice 引理、Borel-Cantelli 引理等等).
- 重要的定义都要牢记 (尤其是概率的公理化定义、分布函数、数学期望的定义).
- 适当做题来练习, 但不需要做太多.

特别鸣谢: 感谢南京大学 17 级数学系的 myh 同学以及 18 级数学系的 zst 同学指出了许多本笔记出现的笔误!

第 1 章 Kolmogorov 概率公理化定义

§ 1.1 σ-代数的定义

设 Ω 是个抽象集合, $\mathscr{F}$ 是 Ω 上一些子集构成的集类 (即: 由一些集合构成的集合)

定义 1.1.1. σ-代数

如果集类 $\mathscr{F}$ 满足:

(1) $\Omega \in \mathscr{F}$,

(2) (对逆运算封闭) 如果 $A \in \mathscr{F}$, 则 $\overline{A} \in \mathscr{F}$,

(3) (对可列并运算封闭) 如果 $\{A_n\}_{n\geq 1} \subset \mathscr{F}$, 则 $\bigcup\limits_{n=1}^{+\infty} A_n \in \mathscr{F}$,

则把集类 $\mathscr{F}$ 称为 Ω 上的 **σ-代数 (σ-域)**.

注: 如果 $X \in \mathscr{F}$, 则意味着 X 就是个事件. 而 $(\Omega, \mathscr{F})$ 是个样本空间.

定理 1.1.1

$\sigma-$ 代数满足下面的性质.

- 上面条件 (1) 可以由 $\varnothing \in \mathscr{F}$ 代替, 这是因为有 (2).
- 上面条件 (3) 可以由下面的 (3') 代替:

  (对可列交运算封闭) 如果 $\{A_n\}_{n\geq 1} \subset \mathscr{F}$, 则 $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{F}$.

  这是因为 De Morgan 律 $\overline{\bigcap\limits_{n=1}^{+\infty} A_n} = \bigcup\limits_{n=1}^{\infty} \overline{A_n} \in \mathscr{F}$, 再利用 (2) 即可.
- 对于某个 N, 让 (3) 中的 $A_n = \varnothing, (n \geq N)$, 即可推出 σ-代数对有限并运算封闭,
- 对于某个 N, 让 (3') 中的 $A_n = \Omega, (n \geq N)$, 即可推出 σ-代数对有限交运算封闭.
- σ-代数对减法运算封闭.

定理 1.1.2

如果 $\{\mathscr{F}_i\}_{i\in I}$ 是一族 $\sigma-$ 代数, 则 $\bigcap\limits_{i\in I} \mathscr{F}_i$ 还是 $\sigma-$ 代数.

证明: 根据定义来验证. (1)$\Omega \in \bigcap\limits_{i\in I} \mathscr{F}_i$, 是显然的.

(2) 对取逆运算封闭: 若 $A \in \bigcap\limits_{i\in I} \mathscr{F}_i$, 则 $A \in \mathscr{F}_i, \forall i \in I$, 则 $\overline{A} \in \mathscr{F}_i, \forall i \in I$, 所以 $\overline{A} \in \bigcap\limits_{i\in I} \mathscr{F}_i$.

(3) 对可列并运算封闭: 若 $\{A_n\}_{n\geq 1} \subset \bigcap\limits_{i\in I} \mathscr{F}_i$, 则 $\{A_n\} \subset \mathscr{F}_i, \forall i \in I$. 注意到 $\mathscr{F}_i (i \in I)$ 是 $\sigma-$ 代数, 则

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}_i, \forall i \in I,$$

所以 $\bigcup\limits_{n=1}^{+\infty} A_n \in \bigcap\limits_{i\in I} \mathscr{F}_i$. □

定义 1.1.2. 生成的最小 σ-代数

设 $\mathscr{C}$ 是 Ω 的一个集类, 称包含 $\mathscr{C}$ 的所有 $\sigma-$ 代数之交为 $\mathscr{C}$ 生成的最小 $\sigma-$ 代数, 记为 $\sigma(\mathscr{C})$.

为了方便, 对于 $a, b \in \mathbb{R}^n$, 记 $a < b$ 为 $[a_i < b_i, \forall 1 \le i \le n]$. 记 $[a, b)$ 为 $\{(x_1, \cdots, x_n) | a_i \le x_i < b_i, 1 \le i \le n\}$.

定义 1.1.3. Borel σ-代数

记 $\mathscr{C} = \{[a, b) | a, b \in \mathbb{R}^n, a < b\}$. 称 $\sigma(\mathscr{C})$ 是 $\mathbb{R}^n$ 上的 Borel $\sigma-$ 代数. 通常记为 $\mathscr{B}(\mathbb{R}^n)$.

注: 上面的 $\mathscr{C}$ 是个集类, 但不是 σ-代数, 它不含空集和全集.

根据 $\sigma-$ 代数的性质, 可以推出:

定理 1.1.3

$\forall x, y \in \mathbb{R}^n$, 都有 $\{x\} = \bigcap\limits_{n=1}^{\infty} \left[x, x + \dfrac{1}{n}\right) \in \mathscr{B}(\mathbb{R}^n)$. 且 $[x, y], (x, y], [x, y) \in \mathscr{B}(\mathbb{R}^n)$.

注: $\mathscr{B}(\mathbb{R}^n)$ 就是由 $\mathbb{R}^n$ 的所有子集构成的集类.

下面的定理说明对任何区间 $[a, b)$(其中方括号和圆括号可换), 由它生成的最小 $\sigma-$ 代数都是 $\mathscr{B}(\mathbb{R}^n)$.

定理 1.1.4

令 $\mathscr{C}_1 = \{A | A\text{是}\mathbb{R}^n\text{中开集}\}, \mathscr{C}_1 = \{A | A\text{是}\mathbb{R}^n\text{中闭集}\}, \sigma(\mathscr{C}_1) = \sigma(\mathscr{C}_2) = \mathscr{B}(\mathbb{R}^n)$.

§ 1.2 概率的定义

当映射 $f : A \to B$ 中的 A 是个集类 (由若干个集合构成的集合), 则说 f 是个集函数.

1933 年, Kolmogorov 提出了如下的概率论的公理化定义. 教我们概率论基础的宋玉林老师说: 如果这个都不会就很丢脸了.

把 $(\Omega, \mathscr{F})$ 称为**可测空间**, $\mathscr{F}$ 中元素为**可测集**, 也叫 **(随机) 事件**.

定义 1.2.1. Kolmogorov

称 $\mathscr{F}$ 上集函数 P 为概率, 如果 P 满足

(1) 非负性, 即 $\forall A \in \mathscr{F}$, $P(A) \ge 0$.

(2) 规范性, 即 $P(\Omega) = 1$.

(3) 可列可加性 (σ 可加性), 设 $\{A_n\}_{n \ge 1} (\subset \mathscr{F})$ 之间的交集为空集, 则 $P\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} P(A_n)$.

注: P 是定义域为集类的函数, 取值是实数. $\mathscr{F}$ 为定义域. 若定义域为集类的函数可取无穷, 则一般只能从 $+\infty$ 与 $-\infty$ 取一个, 不能都取. 这是因为正无穷大与负无穷大之和没有定义.

注: 概率定义在 $\sigma-$ 代数上, 而不是定义在事件空间上.

命题 1.2.1

$P(\varnothing) = 0$.

证明: 对 (3) 取 $A_1 = \Omega, A_2 = A_3 = \cdots = \varnothing$. 则由 $P(\Omega) = 1$ 得 $0 = \sum\limits_{n=2}^{+\infty} P(A_n)$. 由非负性得 $P(\Omega) = 0$. □

命题 1.2.2. 有限可加性

设 $\{A_i\}_{i=1,\cdots,n}(\subset \mathscr{F})$ 之间的交集为空集, 则 $P\left(\bigcup_{i=1}^{n} A_n\right)=\sum_{i=1}^{n} P(A_n)$.

证明: 只需取 $A_{n+1}=\cdots=\varnothing$ 即可. □

命题 1.2.3. 对减法封闭

如果 $A, B \in \mathscr{F}$ 且 $A \subset B$, 则 $P(B \setminus A)=P(B)-P(A)$.

证明: 作不交并处理, 即 $B=A \cup(B \setminus A)$, 则由有限可加性得 $P(B)=P(A)+P(B \setminus A)$. □

命题 1.2.4. 单调性

如果 $A, B \in \mathscr{F}$ 且 $A \subset B$, 则 $P(A) \leq P(B)$.

证明: 作不交并处理, $B=A \cup(B \setminus A)$, 由有限可加性得 $P(B)=P(A)+P(B \setminus A) \geq P(A)$. □

命题 1.2.5. 从下连续性

对 $\{A_n\}_{n \geq 1} \subset \mathscr{F}$, 若 $\{A_n\}$ 单调递增趋于 A, 即 $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots \subset A$ 且 $\bigcup_{n=1}^{\infty} A_n=A$, 则 $P(A_n)$ 单调递增趋于 $P(A)$.

证明: 作不交并处理, 注意到 $A=\bigcup_{n=1}^{+\infty} A_n=A_1 \cup \bigcup_{n=2}^{+\infty}(A_n \setminus A_{n-1})$, 而诸 $A_n \setminus A_{n-1}$ 之间两两不交, 所以根据可列可加性得

$$
\begin{aligned}
P(A)= & P\left(A_1 \cup \bigcup_{n=2}^{+\infty}(A_n \setminus A_{n-1})\right) \\
= & P(A_1)+\sum_{n=2}^{+\infty} P(A_n \setminus A_{n-1}) \\
= & P(A_1)+\sum_{n=2}^{+\infty}(P(A_n)-P(A_{n-1})) \\
= & \lim_{N \rightarrow \infty}\left[P(A_1)+\sum_{n=2}^{N}(P(A_n)-P(A_{n-1}))\right] \\
= & \lim_{N \rightarrow \infty} P(A_N).
\end{aligned}
$$

上面第 3 个等号 $P(A_n \setminus A_{n-1})=P(A_n)-P(A_{n-1})$ 是因为有限可加性 (A_n-A_{n-1} 与 A_{n-1} 不交). □

命题 1.2.6. 从上连续性

对 $\{A_n\}_{n \geq 1} \subset \mathscr{F}$, 若 $\{A_n\}$ 单调递减趋于 A, 即 $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \supset A$ 且 $\bigcap_{n=1}^{\infty} A_n=A$, 则 $P(A_n)$ 单调递减趋于 $P(A)$.

证明: 注意到 A_n 单调递减趋于 A, 则 $\overline{A_n}$ 单调递增趋于 $\overline{A}$. 根据从下连续性,

$$
\lim_{n \rightarrow \infty}(1-P(A_n))=\lim_{n \rightarrow \infty} P(\overline{A_n})=P(\overline{A})=1-P(A).
$$

所以 $\lim_{n \rightarrow \infty} P(A_n)=P(A)$. □

命题 1.2.7. 次 σ 可加性

对 $\{A_n\}_{n\geq 1}\subset\mathscr{F}$, $P\left(\bigcup\limits_{n=1}^{+\infty}A_n\right)\leq\sum\limits_{n=1}^{+\infty}P(A_n)$.

证明: 作不交并处理, 令

$$\boxed{\begin{aligned}&B_1=A_1,\\&B_n=A_n\setminus(A_1\cup\cdots\cup A_{n-1})\\&\quad(=A_n\cup\overline{A_1}\cup\overline{A_2}\cup\cdots\cup\overline{A_n})\end{aligned}}$$

则 $B_n\subset A_n\in\mathscr{F}$, 且 $\bigcup\limits_{n=1}^{+\infty}A_n=\bigcup\limits_{n=1}^{+\infty}B_n$, 且 $\{B_n\}$ 两两不交. 根据单调性, 有

$$P\left(\bigcup_{n=1}^{+\infty}A_n\right)=P\left(\bigcup_{n=1}^{+\infty}B_n\right)=\sum_{n=1}^{+\infty}P(B_n)\leq\sum_{n=1}^{+\infty}P(A_n).$$

□

定理 1.2.8

设 P 是 $\mathscr{F}$ 上满足 $P(\Omega)=1$ 的非负集函数, 则下面命题等价:

(1) P 具有 $\sigma-$ 可加性.

(2) P 具有有限可加性且 P 从下连续.

证明: (1) $\Rightarrow$ (2) 已证. 下面看 (2) $\Rightarrow$ (1).

$$\begin{aligned}P\left(\bigcup_{n=1}^{\infty}A_n\right)&=P\left(\lim_{m\to\infty}\bigcup_{n=1}^{m}A_n\right)\\&=\lim_{m\to\infty}P\left(\bigcup_{n=1}^{m}A_n\right)&&(\text{从下连续性})\\&=\lim_{m\to\infty}\sum_{n=1}^{m}P(A_n)&&(\text{有限可加性})\\&=\sum_{n=1}^{\infty}P(A_n).\end{aligned}$$

□

定理 1.2.9

设 $A,B\in\mathscr{F}$, 则 $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.

证明: 作不交并处理.

$$A\cup B=A\cup(B\setminus A)=A\cup(B\setminus A)=A\cup(B\cap\overline{A})=A\cup(B\setminus AB).$$

. 根据有限可加性和可减性得

$$P(A\cup B)=P(A)+P(B\setminus AB)=P(A)+P(B)-P(AB).$$

□

推论 1.2.10. Bonferroni 不等式

设 $A, B \in \mathscr{F}$, 则 $P(A \cup B) \geq P(A) + P(B) - 1$.

定理 1.2.11

设 $\{A_i\}_{1\leq i\leq n} \subset \mathscr{F}$, 则

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \cdots + (-1)^{n-1}P(A_1A_2\cdots A_n).$$

§ 1.3 条件概率、独立性

定义 1.3.1. 条件概率

设 $(\Omega, \mathscr{F}, P)$ 是概率空间, $B \in \mathscr{F}$, 满足 $P(B) > 0$. 对 $\forall A \in \mathscr{F}$, 称 $P(A|B) \triangleq \dfrac{P(AB)}{P(B)}$ 为事件 A 在 B 发生条件下发生的概率.

注: $P(A|B)$ 与 $P(A)$ 无大小关系.

容易验证 $P(\bullet|B)$ 是 $\mathscr{F} \to [0,1]$ 上的概率测度 (用定义, 留作习题).

注: 稍微作移项可以得到**乘法公式**: $P(AB) = P(B)P(A|B)$. 若 $P(B) = 0$, 可规定 $P(AB) = 0$. 可以推广到多元情形:

$$\begin{aligned}
P(A_1A_2\cdots A_n) &= P(A_1A_2\cdots A_{n-1})P(A_n|A_1\cdots A_{n-1}) \\
&= P(A_1A_2\cdots A_{n-2})P(A_{n-1}|A_1\cdots A_{n-2})P(A_n|A_1\cdots A_{n-1}) \\
&= \cdots \\
&= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1}).
\end{aligned}$$

为了把全概率公式和 Bayes 公式展示出来, 先引入个定义.

定义 1.3.2. 可数分割

设 $(\Omega, \mathscr{F}, P)$ 是概率空间, $\{A_n\}_{n\geq 1} \subset \mathscr{F}$ 两两不交, 且 $\bigcup\limits_{n=1}^{+\infty} A_n = \Omega$, 则把 $\{A_n\}_{n\geq 1}$ 叫 Ω 的一个可数分割.

定理 1.3.1. 全概率公式

设 $\{A_n\}_{n\geq 1}$ 是 Ω 的一个可数分割, $B \in \mathscr{F}$. 则 $P(B) = \sum\limits_{n=1}^{+\infty} P(B|A_n)P(A_n)$.

证明: 利用概率的可列可加性,

$$P(B) = P\left(\bigcup_{n=1}^{+\infty} A_nB\right) = \sum_{n=1}^{+\infty} P(A_nB) = \sum_{n=1}^{+\infty} P(B|A_n)P(A_n).$$

注: 特别地, $P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A})$. $\square$

定理 1.3.2. Bayes 公式

设 $\{A_n\}_{n\geq 1}$ 是 Ω 的一个可数分割, $B\in\mathscr{F}$ 且 $P(B)>0$. 则

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum\limits_{n=1}^{+\infty}P(B|A_n)P(A_n)}.$$

证明: 用乘法公式 + 全概率公式即可. □

注: 特别地, 利用

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

可以推得

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

Bayes 公式又称 "由结果推原因". 这里把 $P(A_i)$ 叫先验概率 (是已知的), $P(A_i|B)$ 是后验概率.

定义 1.3.3. 独立性

设 $(\Omega,\mathscr{F},P)$ 是概率空间, 若 $A,B\in\mathscr{F}$ 有 $P(AB)=P(A)P(B)$, 则称 A 与 B **独立**.

定义 1.3.4

$(\Omega,\mathscr{F},P)$ 是概率空间, 称集类 $\mathscr{C}_1,\mathscr{C}_2$ 独立, 如果 $\forall A\in\mathscr{C}_1,B\in\mathscr{C}_2$, A 与 B 独立.

对于独立性, 有如下性质.

命题 1.3.3

设 $P(B)>0$, 则 A,B 独立的充分必要条件是 $P(A|B)=P(A)$.

证明: 由条件概率的定义立得. □

命题 1.3.4

若 A,B 独立, 则 A 与 $\overline{B}$, $\overline{A}$ 与 B, $\overline{A}$ 与 $\overline{B}$ 独立.

证明: $P(A\overline{B})=P(A)-P(AB)=P(A)-P(A)P(B)=P(A)P(\overline{B})$. □

命题 1.3.5

零概率事件及其对立事件与任一事件独立.

证明: 设 $N,A\in\mathscr{F}$, $P(N)=0$. 则 $NA\in\mathscr{F}$ 且 $NA\subset N$. 根据概率的单调性, $P(NA)\leq P(N)=0$, 从而由非负性, $P(NA)=0=P(N)P(A)$, 从而零概率事件与任一事件独立. 由第 2 个性质, 对于它的对立事件也正确. □

定义 1.3.5. n 个事件相互独立

设 $(\Omega,\mathscr{F},P)$ 是概率空间, $\{A_i\}_{1\leq i\leq n}\subset\mathscr{F}$. 称 $\{A_i\}_{1\leq i\leq n}$ **相互独立**, 若

$$P\left(\bigcap_{j=1}^{k}A_{i_j}\right)=\prod_{j=1}^{k}P(A_{i_j}),1\leq i\leq n.$$

其中 $\{A_{i_j}\}_{1\leq j\leq k}\subset\{A_i\}_{1\leq i\leq n}$.

注: 上面有 2^n-n-1 条式子.

注: 设 $\{\mathscr{C}_i\}$ 为一族集类, $\{\mathscr{C}_i\}\subset\mathscr{F},\forall i\in I$. 若对任意 $A_i\in\mathscr{C}_i$, $\{A_i\}_{i\in I}$ 相互独立, 则称 $\{\mathscr{C}_i\}$ 之间相互独立.

注: 注意区分相互独立 (所有合在一起是独立的) 与两两独立 (任取两个都独立).

§ 1.4 (*) 独立类扩张定理简介

定义 1.4.1. π 类

若集类 $\mathscr{C}$ 关于有限交运算封闭, 即 $A,B\in\mathscr{C}\Rightarrow A\cap B\in\mathscr{C}$, 则称 $\mathscr{C}$ 是 π 类.

定义 1.4.2. λ 类

若集类 $\mathscr{C}$ 满足下面三个条件:

(1)$\Omega\in\mathscr{C}$;

(2) 对减法封闭, 即 $A,B\in\mathscr{C},B\subset A\Rightarrow A-B\in\mathscr{C}$.

(3) 对单调增运算封闭, 即 $\{A_n\}\subset\mathscr{C},A_n\nearrow A\Rightarrow A\in\mathscr{C}$. 则称 $\mathscr{C}$ 是 λ 类.

定理 1.4.1

$\mathscr{F}$ 是 $\sigma-$ 代数 $\Leftrightarrow$ $\mathscr{F}$ 既为 π 类又为 λ 类.

证明: "$\Rightarrow$", $\mathscr{F}$ 是 π 类显然, 且 $\Omega\in\mathscr{F}$ 显然.

若 $A,B\in\mathscr{F}$, 且 $B\subset A$, 则 $A-B=AB^c\in\mathscr{F}$.(σ-代数对有限交运算封闭)

若 $\{A_n\}\in\mathscr{F}$, $A_n\nearrow A$, 则 $\bigcup\limits_{n=1}^{+\infty}A_n=A\subset\mathscr{F}$. 所以 $\mathscr{F}$ 是 λ 类.

"$\Leftarrow$", 只需证 $\mathscr{F}$ 关于可列并运算封闭. 构造一个具有单调性的 $\{B_n\}$ 即可运用 λ 类定义中的第 3 个条件. 令

$$B_1=A_1,B_n=\bigcup_{i=1}^{n}A_i.$$

则 $B_n\nearrow A$. 因为 $\overline{B_n}=\bigcap\limits_{i=1}^{n}\overline{A_i}$, 根据 π 类有限交运算封闭的条件, $\overline{B_n}\in\mathscr{F}$, 则 $B_n\in\mathscr{F}$. 根据 λ 类的条件 (3), 由 $B_n\nearrow A$, 则 $\bigcup\limits_{n=1}^{+\infty}B_n=A\in\mathscr{F}$ 则 $\bigcup\limits_{n=1}^{+\infty}A_n=\bigcup\limits_{n=1}^{+\infty}B_n=A\in\mathscr{F}$, 证完. □

回顾: 称包含 $\mathscr{C}$ 的所有 $\sigma-$ 代数之交为 $\mathscr{C}$ 生成的最小 $\sigma-$ 代数, 记为 $\sigma(\mathscr{C})$.

定理 1.4.2. 独立类扩张定理

设 $\mathscr{C}_1,\mathscr{C}_2$ 为 π 类, 若 $\mathscr{C}_1,\mathscr{C}_2$ 独立, 则 $\sigma(\mathscr{C}_1),\sigma(\mathscr{C}_2)$ 独立.

如果 $\mathscr{C}_1,\mathscr{C}_2$ 都是 $\sigma-$ 代数, 则不需要证了.

证明: 只需证 $\sigma(\mathscr{C}_1),\mathscr{C}_2$ 独立, 可立即推出 $\sigma(\mathscr{C}_1),\sigma(\mathscr{C}_2)$ 独立.

令 $\mathscr{G}=\{A\in\sigma(\mathscr{C}_1)|P(AB)=P(A)P(B),\forall B\in\mathscr{C}_2\}$. 则 $\mathscr{C}_1\subset\mathscr{G}$. ($\mathscr{G}$ 表示花体的 G). 下证 $\sigma(\mathscr{C}_1)\subset\mathscr{G}$, 即可完成证明.

先证 $\mathscr{G}$ 是 λ 类, 事实上

(1)$\Omega\in\mathscr{G}$(不妨设 $\Omega\in\mathscr{C}_1\cup\mathscr{C}_2$, 全集对独立性没有影响)

(2) 若 $A_1, A_2 \in \mathscr{G}, A_2 \subset A_1$, 下证 $A_1 - A_2 \in \mathscr{G}$ 即 $P((A_1 - A_2)B) = P(A_1 - A_2)P(B)$.

$$\begin{aligned} P((A_1 - A_2)B) =& P(A_1B) - P(A_2B) && \text{(概率的可减性)} \\ =& P(A_1)P(B) - P(A_2)P(B) && (\mathscr{G}\text{的定义}) \\ =& P(A_1 - A_2)P(B) && \text{(概率的可减性)}. \end{aligned}$$

所以 $A_1 - A_2 \in \mathscr{G}$.

(3) 若 $\{A_n\} \in \mathscr{G}$ 且 $A_n \nearrow A$, 下证 $A \in \mathscr{G}$.

$$\begin{aligned} P(AB) =& \lim_{n\to\infty} P(A_nB) && \text{(从下连续性)} \\ =& \lim_{n\to\infty} P(A_n)P(B) && (\mathscr{G}\text{的定义}) \\ =& P(A)P(B) && \text{(从下连续性)} \end{aligned}$$

所以 $\mathscr{G}$ 是 λ 类. 由于 $\mathscr{C}_1 \subset \mathscr{G}$, 则 $\lambda(\mathscr{C}_1) = \sigma(\mathscr{C}_1) \subset \mathscr{G}$. 所以 $\mathscr{G} = \sigma(\mathscr{C}_1)$, 则 $\sigma(\mathscr{C}_1)$ 和 $\mathscr{C}_2$ 独立. □

注: $A \subset B$, 则 B 为 λ 类不可推出 A 为 λ 类. (反例：A 只有 1 个元素)

第 2 章　随机变量与分布

§ 2.1　随机变量、分布、分布函数的定义

定义 2.1.1. 随机变量

设 $(\Omega, \mathscr{F})$ 是个样本空间, 把 $X:\Omega\to\mathbb{R}$ 称为 $(\Omega, \mathscr{F})$ 上的随机变量 (random variable, 简称 r.v.), 如果对于任意集合 $B\in\mathscr{B}(\mathbb{R})$,

$$X^{-1}(B)=\{\omega\in\Omega|X(\omega)\in B\}\subseteq\mathscr{F}.$$

根据定义, r.v. 的定义与概率无关.

注: $X^{-1}(B)$ 表示 B 在 X 上的原像, 而不是取逆 (倒数) 运算. 回忆 Borel 代数定义, 这里的 B 是 $\mathbb{R}$ 中的任意一个子集.

定义 2.1.2. 分布

令 X 是 r.v., 那么

$$P\circ X^{-1}(B)=P(X^{-1}(B))=P(X\in B)=P(\{\omega|X(\omega)\in B\}),$$

称 $P\circ X^{-1}$ 是样本空间 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率, 称为 X **在** P **下的分布**.

注: 对于相同的 r.v., 不同的 P 表示不同的分布.

称 ξ,η **同分布**, 指对任意 $A\in\mathscr{B}(\mathbb{R})$, $P(\xi\in A)=P(\eta\in A)$, 即 $P\circ\xi^{-1}=P\circ\eta^{-1}$, 两个测度一样.

下面定理将会展示: 分布就是概率!

定理 2.1.1

设 X 是 r.v., 则 $P\circ X^{-1}$ 是个概率测度.

证明: 按照概率的公理化定义, 对三个条件一一验证即可.

(1) 非负性: 由概率的非负性可保证, $P(X^{-1}(B))\geq 0,\forall B\in\mathscr{B}(\mathbb{R})$.

(2) 规范性: $\Omega=\mathbb{R}$, 注意到 $P\circ X^{-1}(\Omega)=P(X\in\Omega)=1$. ($X\in\mathbb{R}$ 是必然事件)

(3) 可列可加性: $\forall\{B_n\}_{n\geq1}\subset\mathscr{F}$, 其中 $B_m\cap B_n=\varnothing,\forall m\neq n$. 要证

$$P\circ X^{-1}\left(\bigcup_{n=1}^{+\infty}B_n\right)=\sum_{n=1}^{+\infty}P\circ X^{-1}(B_n)$$

即证

$$P\left(X^{-1}\left(\bigcup_{n=1}^{+\infty}B_n\right)\right)=P\left(\bigcup_{n=1}^{+\infty}X^{-1}(B_n)\right).$$

由于 $\bigcup\limits_{n=1}^{+\infty}X^{-1}(B_n)$ 相当于把所有满足 $\omega\in B_n, n=1,2,\cdots$ 的 ω 都并起来, 由诸 B_n 不交, 则

$$\bigcup_{n=1}^{+\infty}X^{-1}(B_n)=\left\{\omega\in\Omega|X(\omega)\in\sum_{n=1}^{\infty}B_n\right\}=X^{-1}\left(\bigcup_{n=1}^{+\infty}B_n\right)$$

等号两边求概率即可证完. □

例 2.1.2

设 $B \in \mathscr{B}(\mathbb{R})$, $X:\Omega \to \mathbb{R}$ 是 $(\Omega, \mathscr{B}(\mathbb{R}))$ 上的 r.v., $\Omega = \mathbb{R}$. 如果对任意的 $x \in \Omega$ 都有 $X(x) \equiv C$, 则

$$X^{-1}(B) = \begin{cases} \mathbb{R}, & C \in B, \\ \varnothing, & C \notin B. \end{cases}$$

这部分我一开始被老师绕糊涂了, 下面按照我的理解解释一遍:

如果 $C \in B$, 则 B 在 X 中的原像恰好为整个 $\Omega = \mathbb{R}$, 也就是说 $X(\Omega) \equiv \{C\} \subset B$ 恒成立.

如果 $C \notin B$, 则 $\forall x \in \Omega$, $X(x)$ 都不在 B 内, 即 B 在 X 中的原像为空集, 也就是说 $X(\Omega) \equiv \{C\} \cap B = \varnothing$.

定义 2.1.3. 分布函数

设 X 是概率空间 $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$ 中 r.v., 定义 $F(x) = P(X \le x), \forall x \in \mathbb{R}$ 是 X 的分布函数 (distribution function).

注: $P(X \le x)$ 表示事件 $[X \le x]$ 发生的概率.

注: 如果定义成 $P(X < x)$, 对它的性质有影响.

注: 易知 F 是 $F:\mathbb{R} \to [0,1]$ 函数.

注: 分布函数与分布的关系: $F(x) = P(X \le x) = P(X \in (-\infty, x]) = P \circ X^{-1}((-\infty, x])$.【重要】

下面介绍分布函数的性质 (事实上, 把满足下面三个性质的函数都叫**分布函数**). 设 ξ 是 r.v.

命题 2.1.3

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

证明: 用从上、下连续性.

$$\lim_{n \to -\infty} P \circ \xi^{-1}((-\infty, -n]) = P \circ \xi^{-1}\left(\bigcap_{n=1}^{+\infty} (-\infty, -n]\right) = 0.$$

$$\lim_{n \to +\infty} P \circ \xi^{-1}((-\infty, m]) = P \circ \xi^{-1}\left(\bigcup_{n=1}^{+\infty} (-\infty, m]\right) = P \circ \xi^{-1}(\mathbb{R}) = 1. \square$$

命题 2.1.4. 单调不降性

$\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2$, 则 $F(x_1) \le F(x_2)$.

证明: 用概率的单调性并注意到 $(-\infty, x_1] \subset (-\infty, x_2]$ 即可. $\square$

命题 2.1.5. 右连续

即 $F(x+0) = F(x)$.

证明: 只需证 $x_n \searrow x$ 时 $F(x_n) \to F(x)$. 用概率的从上连续性,

$$\begin{aligned} \lim_{n \to \infty} F(x_n) = & \lim_{n \to \infty} P \circ \xi^{-1}((-\infty, x_n]) \\ = & P \circ \xi^{-1}\left(\bigcap_{n=1}^{+\infty} (-\infty, x_n]\right) \\ = & P \circ \xi^{-1}((-\infty, x]) = F(x). \end{aligned}$$

□

注: (1)$F(x)$ 的不连续点个数至多可数.

注: (2)$\bigcap\limits_{n=1}^{+\infty}(-\infty,x_n]=(-\infty,x]$ 用相互包含可证.

注: (3)$\{x|F(x)\neq F(x-0)\}=\{x|F(x)-F(x-0)>0\}=\bigcup\limits_{n=1}^{+\infty}\{x|F(x)-F(x-0)\geq\frac{1}{n}\}$, 其中集合 $\{x|F(x)-F(x-0)\geq\frac{1}{n}\}$ 的元素只有有限个 (不多于 n 个, 考虑到概率的规范性 $P(\Omega)=1$).

§ 2.2 离散型随机变量

2.2.1 离散型随机变量的分布列与分布函数的关系

设 X 为离散型 r.v., 其可能取值为 $\{x_k\}_{k\geq1}$, 且 $P(X=x_k)=p_k, 0\leq p_k\leq1$, 则分布函数是

$$F(x)=P(X\leq x)=\sum_{x_i\leq x}P(X=x_i)=\sum_{x_i\leq x}p_i.$$

另外,

$$p_k=P(X=x_k)=P(X\leq x_k)-P(X<x_k)=F(x_k)-F(x_k-0),$$

可以观察出 $F(x)$ 是个阶梯函数, 在 x_k 处跳的高度是 p_k. 它有左极限且右连续.

2.2.2 几个重要的离散型随机变量分布

下面记 $q=1-p$.

定义 2.2.1

若随机试验只有两种可能结果 A 或 $\overline{A}$, 则称该试验为 **Bernoulli 试验**. 将 Bernoulli 试验**独立**出来进行 n 次, 称为 **n 重 Bernoulli 试验**, 记为 E^n.

n 重 Bernoulli 试验的特点: ①每次试验只有两种可能结果: A 或 $\overline{A}$; ② A 在每次试验中出现的概率 p 不变; ③共进行 n 次相同的试验 (相互独立).

例 2.2.1. Bernoulli 分布, 两点分布, 0-1 分布

若只进行一次 Bernoulli 试验, 如果成功记为 1, 失败记为 0, r.v.X 满足

$$P(X=1)=P(A)=p\in(0,1),$$

则称 X 服从参数 p 的 Bernoulli 分布.

例 2.2.2. 二项分布

$X\sim b(n,p)$. 把 n 重 Bernoulli 试验中事件 A 出现 k 次的概率记为 $b(k;n,p)$, 记X 为 n 次试验中事件 A 发生的次
则

$$P(X=k)=C_n^kp^kq^{n-k}.$$

例 2.2.3. 几何分布

记X 为 Bernoulli 试验中 A 首次发生的时刻, 由 $[X=k]=[$ 前 $k-1$ 次未发生, 第 k 次发生 $]$, 则

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

注: 几何分布具有无记忆性, 已知前 m 次试验中 A 未发生, 记 ξ 为 A 首次发生还需等待的时间. 则

$$P(\xi=k)=P(X=m+k|X>m)=\frac{P(X=m+k)}{P(X>m)}=\cdots=P(X=k).$$

例 2.2.4. Pascal 分布

在 Bernoulli 试验中, 需进行多少次试验, 事件 A 第 r 次出现. 记 X 为事件 A第 r 次发生的时刻. 则前 $k-1$ 次试验中事件 A 发生了 $r-1$ 次, 第 k 次试验中事件 A 发生.

$$P(X=k)=C_{k-1}^{r-1}p^rq^{k-r}.$$

例 2.2.5. 多项分布

做 n 重独立试验, 每次试验有若干结果出现. 设 $P(A_i)=p_i, 1\le i\le r$ 且 $p_1+p_2+\cdots+p_r=1, p_i\in(0,1)$. 记$X_i$ 为 n 重独立试验中 A_i 发生的次数, 则 $(X_1,\cdots,X_r)$ 服从如下多项分布:

$$P(X_1=k_1,X_2=k_2,\cdots,X_r=k_r)=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}, k_i\ge 0, \sum_{k=1}^r k_i=n.$$

例 2.2.6. Poisson 分布

$X\sim P(\lambda)$. 若离散型 r.v. X 满足 $P(X=k)=e^{-\lambda}\dfrac{\lambda^k}{k!}, \lambda>0$, 则称 X 服从参数为 λ 的 Poisson 分布.

定理 2.2.7. 二项分布逼近 Poisson 分布

在独立试验中, 以 p_n 代表事件 A 在试验中发生的概率, 它与试验总数 n 有关. 若 $\lim\limits_{n\to\infty}np_n=\lambda$, 则

$$b(k;n,p)=C_n^kp_n^k(1-p_n)^{n-k}\to\frac{e^{-\lambda}}{k!}\lambda^k.$$

证明: 注意到

$$\begin{aligned}
&C_n^kp_n^k(1-p_n)^{n-k}\\
=&\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{np_n}{n}\right)^k\left(1-\frac{np_n}{n}\right)^{n-k}\\
=&\frac{n(n-1)\cdots(n-k+1)}{n^kk!}(np_n)^k\left(1-\frac{np_n}{n}\right)^{n-k}\\
\to&\frac{1}{k!}\lambda^ke^{-\lambda}(n\to\infty).
\end{aligned}$$

2.2.3 一些例子

例 2.2.8

在 N 件产品中有 M 件次品, 进行 n 次有放回的抽样调查. 问：抽得 k 件次品的概率是多少？

解: 设 X 为 n 次抽检中次品件数, 则 $P(X=k)=C_n^k\left(\dfrac{M}{N}\right)^k\left(1-\dfrac{M}{N}\right)^{n-k}$.

例 2.2.9

一个醉汉开门, 共有 n 把钥匙, 其中仅有一把能将门打开, 他随机选取 1 把钥匙开门. 此人在第 k 次开时首次成功的概率是

$$\left(1-\frac{1}{n}\right)^{k-1}\frac{1}{n}.$$

例 2.2.10. Banach 火柴盒

一个数学家的左、右口袋各放一盒装有 N 根火柴的盒, 每次抽烟时以 $\frac{3}{5}$ 的概率拿左盒并用 1 根. 求发现一盒用完时, 另一盒有 r 根的概率.

解: 先求左边空、右边剩 r 根的概率: 此时左边摸了 $N+1$ 次, 右边摸了 $N-r$ 次, 共 $2N-r+1$ 次. 记 A: 从左袋取一根火柴, 则所求概率 P_1 为事件 A 第 $N+1$ 次发生的时刻为 $2N+1-r$ 的概率. 则

$$P_1 = C_{2N-r}^{N}\left(\frac{3}{5}\right)^{N+1}\left(\frac{2}{5}\right)^{N-r}.$$

同样有

$$P_2 = C_{2N-r}^{N}\left(\frac{2}{5}\right)^{N+1}\left(\frac{3}{5}\right)^{N-r}.$$

所以发现一盒用完时, 另一盒有 r 根的概率为 P_1+P_2. □

例 2.2.11. 直线上的随机游动

设 S_n 为 n 时刻所在位置, $S_0=0$, 每次以相等的概率向直线的左或右边移动, $S_n=k$ 表示在 n 时刻与 0 时刻相比向右走了 k 个单位. 求 $P(S_n=k)$.

解: 容易知道向右比向左多走 k 次. 设 x,y 为向右、向左次数, 则 $x-y=k$ 且 $x+y=n$, 所以 $x=\frac{k+n}{2}, y=\frac{n-k}{2}$. 当 n,k 奇偶性不同时,

$$P(S_n=k)=0;$$

当 n,k 奇偶性相同时,

$$P(S_n=k)=C_n^{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n.$$

例 2.2.12. 平面上的随机游动

质点在平面上等可能向上、下、左、右移动, 每次移动距离为 1, 求经过 $2n$ 次移动回到原点的概率.

解: $p_{上}=p_{下}=p_{左}=p_{右}=\frac{1}{4}$. 记 k 为向上运动次数, 则向下运动了 k 次、向左、右都移动了 $n-k$ 次. 则

$$\begin{aligned} P &= \sum_{k=0}^{n}\frac{(2n)!}{k!k!(n-k)!(n-k)!}\left(\frac{1}{4}\right)^{2n} \\ &= \frac{(2n)!}{4^{2n}}\cdot\frac{1}{(n!)^2}\sum_{k=0}^{n}(C_n^k)^2 = \frac{1}{4^{2n}}(C_{2n}^n)^2. \end{aligned}$$

其中用到了 $\sum_{k=0}^{n}(C_n^k)^2 = C_{2n}^n$, 对 $(1+x)^{2n}=(1+x)^n(1+x)^n$ 两端展开即可.

定理 2.2.13. 二项分布中最可能成功次数

证明当 $k=[(n+1)p]$ 时 $b(k;n,p)$ 取最大值.

证明: 只需考察 $\dfrac{b(k;n,p)}{b(k-1;n,p)}=1+\dfrac{(n+1)p-k}{kq}$ 与 1 的大小关系. 若 $(n+1)p$ 为整数, 则 $b((n+1)p;n,p)=b(np;n,p)$ 为最大值;

否则, 当 $k=[(n+1)p]$ 时, $b(k;n,p)$ 取最大值. 这是因为

$$(n+1)p-[(n+1)p]>0>(n+1)p-[(n+1)p]-1.$$

取 $k\le[(n+1)p]$ 时, $\dfrac{b(k;n,p)}{b(k-1;n,p)}>1$; 取 $k>[(n+1)p]$ 时, $\dfrac{b(k;n,p)}{b(k-1;n,p)}<1$. □

例 2.2.14

设某种疾病的发病率是 0.01, 则在 500 人的社区中进行普查最可能的发病人数是 $[(n+1)p]=[5.01]=5$.

例 2.2.15

某公司制造某种芯片, 次品率为 0.001, 各芯片成为次品相互独立. 求在 1000 次产品抽检中至少有 2 个次品的概率.

解: $n=1000, p=0.001$, 则 $\lambda=np=1$. 则所求概率是

$$P=1-C_{1000}^{1}0.001\times0.999^{999}-C_{1000}^{0}0.999^{1000}\approx1-\frac{1^1}{1!}e^{-1}-\frac{1^0}{0!}e^{-1}=1-\frac{2}{e}.$$

§ 2.3 连续型随机变量

定义 2.3.1. 概率密度函数

对于连续型 r.v. X 的分布函数 $F(x)$, 若存在非负可积函数 $p(x)$ 使得

$$F(x)=\int_{-\infty}^{x}p(y)\mathrm{d}y,$$

则称 $p(x)$ 为 X 的**概率密度函数** (density function).

注: (1) 这里积分是 Lebesgue 积分, dy 表示测度. $p(x)$ 中不大于 0 的点放在一起构成 Lebesgue 测度上的零测度集.

注: (2) 这里 F 连续但不一定可导, 只有当 p 连续时 F 才可导.

注: (3) 一般给的 p 都较好, 可以当作 Riemann 积分来做.

定理 2.3.1. 积分的绝对连续性

设 $f\in L^1(X,\mu)$, 则 $\forall\varepsilon>0,\exists\delta>0$, 若 E 可测且 $\mu(E)<\delta$, 则 $\displaystyle\int_E|f|d\mu<\varepsilon$.

证明: 实变讲义. □

定理 2.3.2

对于连续型 r.v. X, 概率密度函数有如下性质.

(1) $p(x) \geq 0$;

(2) $\int_{-\infty}^{+\infty} p(x)dx = 1$;

(3) $\forall a, b \in \mathbb{R}, a < b, \int_a^b p(x)dx = F(b) - F(a)$.

(4) $\forall a \in \mathbb{R}, P(X = a) = 0$.

证明: (4)$0 \leq P(X = a) \leq P(a - h < x \leq a) = \int_{a-h}^a p(x)dx$, 根据**积分的绝对连续性**, 即当 $h \to 0$ 时 $[a - h, a]$ 的测度趋于 0, 从而 Lebesgue 积分趋于 0, 对最右边式子取极限得

$$0 \leq P(X = a) \leq \lim_{h\to 0} \int_{a-h}^a p(x)dx = 0. \qquad \square$$

例 2.3.3

设随机变量 X 具有概率密度

$$p(x) = \begin{cases} kx, & 0 \leq x < 3, \\ 2 - \frac{x}{2}, & 3 \leq x < 4, \\ 0, & \text{其他}. \end{cases}$$

(1) 确定常数 k; (2) 求 x 的分布函数; (3) 求 $P\left(1 \leq x \leq \frac{7}{2}\right)$.

解: (1) 利用 $\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$ 求得 $k = \frac{1}{6}$. (2)$F(x) = \int_{-\infty}^{x} p(y)\mathrm{d}y = \begin{cases} 0, & x < 0, \\ \frac{1}{12}x^2, & 0 \leq x < 3, \\ 2x - \frac{x^2}{4} - 3, & 3 \leq x \leq 4, \\ 1, & x > 4. \end{cases}$

(3) 计算 $F\left(\frac{7}{2}\right) - F(1)$.

例 2.3.4

设连续性随机变量 X 的密度函数

$$p(x) = \begin{cases} 4x^3, & 0 < x < 1, \\ 0, & \text{其他}. \end{cases},$$

求常数 $a \in \mathbb{R}$ 使得 $P(X > a) = P(X < a)$.

解: 相当于 $P(X > a) = P(X < a) = \frac{1}{2}$. 只需要解方程 $\int_{-\infty}^{a} 4x^3\mathrm{d}x = \frac{1}{2}$.

注: 把满足 $P(X > a) = P(X < a)$ 的 a 叫**中位数**.

2.3.1 几种重要的连续型分布

传统的几个连续型分布为:

名称	记号	密度函数
$[a,b]$ 上的均匀分布	$X\sim U[a,b]$	$p(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b],\\ 0, & \text{其他},\end{cases}$
指数分布 $(\lambda>0)$	$X\sim E(\lambda)$	$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}$
Γ 分布 $(\lambda>0)$	$X\sim\Gamma(r,\lambda)$	$p(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}$
参数为 μ,σ 的正态分布	$X\sim N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},(\sigma>0,\mu\in\mathbb{R})$
Cauchy 分布		$p(x)=\dfrac{\theta}{\pi(x^2+\theta^2)},\theta>0.$

一些补充说明:

(1) **均匀分布 (uniform density)** 中的概率仅与区间测度有关, 与区间位置无关. 均匀分布的分布函数是

$$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leq x\leq b\\ 1, & x>b.\end{cases}$$

(2) **指数分布 (exponential density)** 的分布函数是

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0,\\ 0, & x\leq 0.\end{cases}$$

指数分布的一些性质如下:

定理 2.3.5. 指数分布的无记忆性

若 $X\sim E(\lambda)$, 则 $P(X>t+s|X>s)=P(X>t)$.

证明: $P(X>t+s|X>s)=\dfrac{P(X>t+s)}{P(X>s)}=\dfrac{1-(1-e^{-(t+s)\lambda})}{1-(1-e^{s\lambda})}=P(X>t).$ □

Γ 分布中, $X\sim\Gamma(r,\lambda)$ 中的 r 叫**形状系数**, λ 叫**尺度参数**. 回顾

$$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}\mathrm{d}t.$$

满足 $\Gamma(n)=(n-1)!$, 则当 $r=1$ 时, Γ 分布变成指数分布.

定理 2.3.6

设 $X_1\sim E(\lambda_1)$，$X_2\sim E(\lambda_2)$，且 X_1,X_2 相互独立．设 $X_3=\min\{X_1,X_2\}$，则 $X_3\sim E(\lambda_1+\lambda_2)$.

证明: 设 X_i 的分布函数为 $F_i(x)$，$i=1,2,3$．则对 $i=1,2$，当 $x>0$ 时，有

$$F_i(x)=\int_0^x\lambda_i\mathrm{e}^{-\lambda_i t}\mathrm{e}t=1-\mathrm{e}^{-\lambda_i x}.$$

所以 $P(X_i > x) = 1 - P(X_i \le x) = 1 - F_i(x) = \mathrm{e}^{-\lambda_i x}$. 从而

$$\begin{aligned} F_3(x) = P(X_3 \le x) &= P(\min\{X_1, X_2\} \le x) \\ &= 1 - P(\min\{X_1, X_2\} > x) \\ &= 1 - P(X_1 > x, X_2 > x) \\ &= 1 - P(X_1 > x)P(X_2 > x) \text{（相互独立）} \\ &= 1 - \mathrm{e}^{-\lambda_1 x} \cdot \mathrm{e}^{-\lambda_2 x}. \end{aligned}$$

所以 X_3 的密度函数是 $p_3(x) = (\lambda_1 + \lambda_2)\mathrm{e}^{-(\lambda_1+\lambda_2)x}$，即 $X_3 \sim E(\lambda_1 + \lambda_2)$. □

定理 2.3.7. 指数分布变成 Γ 分布

设 $\xi_1, \cdots, \xi_n$ 是独立同分布的随机变量, 且 $\xi_i \sim E(\lambda)$, 则

$$\xi_1 + \cdots + \xi_n \sim \Gamma(n, \lambda).$$

注: 要注意分布相同不代表随机变量相同, 分布相同指的是 $P \circ \xi_i^{-1}$ 测度相同!

定理 2.3.8. 指数分布与 Poisson 分布的关系

设脑子在任何长为 t 的时间 $[0, t]$ 内短路的次数 $N(t)$ 服从参数为 λt 的 Poisson 分布, 则相继两次短路之间的时间间隔 T 服从参数为 λ 的指数分布.

证明: $P(N(t) = k) = \dfrac{e^{-\lambda t}}{k!}(\lambda t)^k$. 注意到 $\boxed{[T > t] = [N(t) = 0]}$, 因此

$$P(T > t) = P(N(t) = 0) = e^{-\lambda t}, \Rightarrow P(T \le t) = 1 - e^{-\lambda t}, \forall t > 0.$$

这就是指数分布的分布函数. □

(3) **正态分布 (normal density)** 又称 **Gauss 分布**. 当 $\mu = 0, \sigma = 1$ 即 $X \sim N(0, 1)$ 时, 又称**标准正态分布**, 此时

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

(4) **关于正态分布的密度函数，有如下几个性质:**

① $p(x)$ 关于 $x = \mu$ 对称, 即 $P(\mu - h < x \le \mu) = P(\mu \le x < \mu + h)$.

②当 σ 小的时候, 波动小, 图像尖; 当 σ 大的时候, 波动大, 图像平. 波动越大, $\displaystyle\int_t^{t+\Delta t} p(x)\mathrm{d}x$ 越大, 数据越分散.

③ $p(x)$ 在 $x = \mu$ 处取最大值.

④ $p(x)$ 在 $x = \mu \pm \sigma$ 处取拐点, 且以 x 轴为渐近线.

定理 2.3.9. 标准正态分布与一般正态分布的关系

有如下结论:

(1) 设 $\xi \sim N(\mu, \sigma^2)$, 则 $\dfrac{\xi - \mu}{\sigma} \sim N(0, 1)$.

(2) 设 $\eta \sim N(0, 1)$, 则 $\sigma\eta + \mu \sim N(\mu, \sigma^2)$.

(3) 定义**标准正态分布函数**为

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

那么 $\Phi(-x) + \Phi(x) = 1$.

注: 根据这个结论, 若 $X\sim N(\mu,\sigma^2)$, 则

$$P(X\le x)=P\left(\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

可以查正态分布函数表来计算. 此外有

$$P(a\le X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right).$$

§ 2.4 多维随机变量

2.4.1 随机向量的定义

定义 2.4.1. 随机向量

设 $\xi_1,\cdots,\xi_n$ 是定义在 $(\Omega,\mathscr{F})$ 上的随机变量, 则称 $\xi=(\xi_1,\cdots,\xi_n)$ 为 n 元随机变量（n 元随机向量）.

注: ξ_i 是定义在同一个可测空间 $(\Omega,\mathscr{F})$ 上的随机变量. 若不然, ξ_i 是不同的可测空间 $(\Omega_i,\mathscr{F}_i)$ 上的随机变量, 那么就需要在 $\left(\prod\limits_{i=1}^n\Omega_i,\prod\limits_{i=1}^n\mathscr{F}_i\right)$ 上考虑随机向量 $(\xi_1,\cdots,\xi_n)$.

注: 一个等价定义: 对于 $\xi:\Omega\to\mathbb{R}^n$, 若对任意 $B\in\mathscr{B}(\mathbb{R}^n)$, 都有 $[\xi\in B]\in\mathscr{F}$(即任意 Borel 可测集的原像都在 $\mathscr{F}$ 里), 则称 ξ 为**随机向量**.

定义 2.4.2. 联合分布函数

设 $(\xi_1,\cdots,\xi_n)$ 为 $(\Omega,\mathscr{F},P)$ 上的随机变量. 称

$$F(x_1,\cdots,x_n)=P(\xi_1\le x_1,\cdots,\xi_n\le x_n)$$

为 $(\xi_1,\cdots,\xi_n)$ 的**联合分布函数**.

命题 2.4.1

联合分布函数有如下性质:

(1) 单调性: 关于每一个变量单调不减.

(2) 连续性: 关于每一个分量连续.

(3)$0\le F\le 1$ 且让某个分量趋于 $-\infty$ 得到

$$F(x_1,x_2,\cdots,x_{i-1},\infty,x_{i+1},\cdots,x_n)=0,$$

所有分量趋于 $+\infty$ 得到

$$F(+\infty,\cdots,+\infty)=1.$$

(4) 对任意满足 $a_i<b_i(1\le i\le n)$ 的 $a=(a_1,\cdots,a_n),b=(b_1,\cdots,b_n)$,

$$P(\xi\in(a,b])=\Delta^{(1)}_{(b_1,a_1)}\cdots\Delta^{(n)}_{(b_n,a_n)}F.$$

其中, $\xi\in(a,b]$ 指 $\{(x_1,\cdots,x_n)\in\mathbb{R}^n:a_i<x_i\le b_i,1\le i\le n\}$, 而

$$\Delta^{(i)}_{(b_i,a_i)}F=F(x_1,\cdots,x_{i-1},b_i,x_{i+1},\cdots,x_n)-F(x_1,\cdots,x_{i-1},a_i,x_{i+1},\cdots,x_n).$$

注: 一般来说,

$$P(\xi \in (a,b]) \neq \prod_{i=1}^{n} P(\xi_i \in (a_i, b_i]).$$

特别地当 $n=2$ 时,

$$P((x,y) \in (a_1,b_1] \times (a_2,b_2]) = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) \geq 0.$$

$n=2$ 时, 把**联合密度**定义为

$$p(x,y) \triangleq \frac{\partial^2}{\partial x \partial y} F(x,y).$$

定义 2.4.3

称 $(x_1,\cdots,x_n)$ 为 $(\Omega,\mathscr{F},P)$ 上的**连续型随机变量**, 若存在非负可积函数 $p:\mathbb{R}^n \to [0,+\infty)$, 使得对任意 $x=(x_1,\cdots,x_n)\in\mathbb{R}^n$, 有

$$F(x_1,\cdots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n.$$

或记为

$$F(x) = \int_{-\infty}^{x} p(y)\mathrm{d}y.$$

命题 2.4.2

联合密度函数的性质:

(1)$p \geq 0, a.s.$

(2) $\int_{\mathbb{R}^n} p(x)\mathrm{d}x = 1.$

(3) 对任意 $D \in \mathscr{B}(\mathbb{R}^n)$ 有

$$P(x \in D) = \int_D p(x)\mathrm{d}x.$$

注: $\{A_1 \times A_2 \times \cdots \times A_n : \forall i, A_i \in \mathscr{B}(\mathbb{R})\}$ 不构成 σ-代数. 即

$$\mathscr{B}(\mathbb{R}^2) \neq \{A_1 \times A_2 : A_1, A_2 \in \mathscr{B}(\mathbb{R})\}.$$

但是,

$$\sigma(\{A_1 \times \cdots \times A_n : \forall i, A_i \in \mathscr{B}(\mathbb{R})\}) = \sigma(\{(a,b] : a,b \in \mathbb{R}^n, a < b\}).$$

例 2.4.3

设二维随机变量 (X,Y) 有如下的密度:

$$p(x,y) = \begin{cases} 2e^{-(2x+y)}, & x,y>0, \\ 0, & \text{其他}. \end{cases}$$

求分布函数 $F(x,y)$ 并求 $P(X \leq Y)$.

解: 代入公式即可, 得到

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\mathrm{d}u\mathrm{d}v=\begin{cases}(1-e^{-2x})(1-e^{-y}), & x,y>0,\\ 0, & \text{其他}.\end{cases}$$

例 2.4.4. 均匀分布

若 $G\in\mathscr{B}(\mathbb{R}^n)$ 且 $m(G)>0$, 称多元随机变量 $X=(X_1,\cdots,X_n)$ 服从 G 上均匀分布, 若

$$p(x)=\begin{cases}\dfrac{1}{m(G)}, & x\in G,\\ 0, & x\notin G.\end{cases}$$

例 2.4.5. 多元正态分布

若 $X=(X_1,\cdots,X_n)$ 的密度函数是

$$p(x)=\frac{1}{(\sqrt{2\pi})^n(\det\Sigma)^{\frac{1}{2}}}\exp\left\{\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right\}.$$

其中 $\mu\in\mathbb{R}^n,\Sigma$ 是 n 阶正定矩阵. 则称 X 服从参数为 μ,Σ 的正态分布, 记为

$$(x_1,\cdots,x_n)\sim N(\mu,\Sigma).$$

2.4.2 边际分布

定义 2.4.4. 边际分布

设 (X,Y) 是二维随机变量, 称 $P\circ X^{-1}$ 与 $P\circ Y^{-1}$ 为 (x,y) 的边际分布.

注: $P\circ(X,Y)^{-1}$ 是 $\mathscr{B}(\mathbb{R}^2)$ 上的分布.

我们用 $F(x,y)$ 表示 (X,Y) 的联合分布函数, 那么也有**边际分布函数**:

$$F_X(x)\triangleq\lim_{y\to+\infty}F(x,y)=P(X\le x,Y\le+\infty)=P(X\le x).$$
$$F_Y(y)\triangleq\lim_{x\to+\infty}F(x,y)=P(X\le+\infty,Y\le y)=P(Y\le y).$$

若连续性随机变量 (X,Y) 的密度函数为 $p(x,y)$, 则

$$F_X(x)=P(X\le x,Y\le+\infty)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}p(u,v)\mathrm{d}v\mathrm{d}u.$$

这样也可以定义**边际密度函数**:

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y,$$
$$p_Y(Y)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x,$$

若 (X,Y) 是离散型随机变量, 其联合分布列定义为 $p_{ij}=P(X=x_i,Y=y_j)$, 那么 (X,Y) 的**边际**

分布列定义为:

$$p_{i\bullet} \triangleq P(X=x_i)=\sum_j P(X=x_i,Y=y_j)=\sum_j p_{ij}.$$
$$p_{\bullet j} \triangleq P(Y=y_j)=\sum_i P(X=x_i,Y=y_j)=\sum_i p_{ij}.$$

例 2.4.6

设 (X,Y) 的联合密度为 $p(x,y)=\begin{cases}1, & 0<x<1,|y|<x,\\ 0, & \text{其他}.\end{cases}$, 那么

$$p_X(x)=\begin{cases}2x, & 0<x<1,\\ 0, & \text{其他},\end{cases} p_Y(y)=\begin{cases}1-|y|, & -1<y<1,\\ 0, & \text{其他},\end{cases}$$

例 2.4.7

二维正态分布的边际分布是正态分布.

二维正态分布中 ρ 表示 X,Y 之间的相关性, 这是因为边际分布得到的正态分布没有 ρ. 若 $\rho=0$, 则 X,Y 独立.

注: 二维正态分布: 若二维随机变量 $(X,Y)\sim N(\mu,\Sigma)$, 其中

$$\mu=(\mu_1,\mu_2)\in\mathbb{R}^2,\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\sigma_1,\sigma_2>0,\rho\in(-1,1).$$

那么二维正态分布的联合密度函数是

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}.$$

2.4.3 条件分布

首先考虑离散型随机变量的条件分布, 下面设 j 满足 $p_{\bullet j}>0$. 考虑在 $[Y=y_j]$ 的条件下 $[X=x_i]$ 发生的概率, 即求

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\bullet j}}=\frac{p_{ij}}{\sum_k p_{kj}}.$$

易知上述条件概率的分布列满足:

- $P(X=x_i|Y=y_j)\geq 0,\forall i.$
- $\sum\limits_i P(X=x_i|Y=y_j)=1.$

定义 2.4.5. 条件分布列

设 (X,Y) 是离散型随机变量, 对固定的 j, 若 $P(Y=y_j)>0$, 则称 $\left\{\frac{p_{ij}}{p_{\bullet j}}\right\}_{i\geq 1}$ 为 $y=y_j$ 条件下随机变量 X 的条件分布列.

下面看连续型随机向量的条件分布: 设 (X,Y) 是连续型随机向量, 如何定义在 $Y=y$ 条件下 X

的条件分布 $F(x|y)$? 若定义

$$F(x|y) = P(X \leq x|Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}$$

这是不可以的, 因为 $P(Y = y) = 0$. 所以我们换种方式定义如下:

$$\begin{aligned}
F(x|y) &= P(X \leq x|Y = y) \\
&\triangleq \lim_{h\to 0^+} P(X \leq x|y \leq Y \leq y+h) \\
&= \lim_{h\to 0^+} \frac{P(X \leq x, y \leq Y \leq y+h)}{P(y \leq Y \leq y+h)} \\
&= \lim_{h\to 0^+} \frac{\int_{-\infty}^{x} \mathrm{d}u \int_{y}^{y+h} p(u,v)\mathrm{d}v}{\int_{y}^{y+h} p_Y(v)\mathrm{d}v} \\
&= \frac{\int_{-\infty}^{x} p(u,y)\mathrm{d}u}{p_Y(y)} \qquad \text{(积分中值定理)} \\
&= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)}\mathrm{d}u = \frac{\int_{-\infty}^{x} p(u,y)\mathrm{d}u}{\int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x}.
\end{aligned}$$

定义 2.4.6. 条件分布

把条件分布函数与条件密度函数分别定义为:

$$F(x|y) \triangleq \frac{\int_{-\infty}^{x} p(u,y)\mathrm{d}u}{\int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x} = \frac{F(x,y)}{F_Y(y)},$$

$$p(x|y) \triangleq \frac{p(u,y)}{p_Y(y)}.$$

注: 根据此定义,

$$F(x|y) = \int_{-\infty}^{x} p(u|y)\mathrm{d}u.$$

此外, 我们有:

$$\boxed{p(x,y) = p(x|y)p_Y(y)}.$$

即边际密度函数 × 条件密度函数 = 联合密度函数!

例 2.4.8

二维正态分布的条件分布也是正态分布.

证明: 容易计算得到

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2\right\}.$$

那么 $p(y|x)$ 对应的是 $N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$ 的密度函数. □

例 2.4.9

设 (X,Y) 服从 $G = \{(x,y)|x^2+y^2 \leq 1\}$ 上的均匀分布, 求 $p(x|y)$.

解: $p(x,y)=\frac{1}{\pi}I_G$, 那么

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x = \frac{2}{\pi}\sqrt{1-y^2}I_{[-1,1]}, \\ p(x|y) &= \frac{p(x,y)}{p_Y(y)} = \frac{1}{2\sqrt{1-y^2}}I_G. \end{aligned}.$$

例 2.4.10

设随机变量 $X\sim U(0,1)$. 观察到 $X=x(0<x<1)$ 时, 随机变量 $Y\sim U(x,1)$, 求 Y 的概率密度 $P_Y(y)$.

解: 由条件, $P_X(x)=I_{[0,1]}$, 而当 $x\in(0,1)$ 时, $p(y|x)=\frac{1}{1-x}I_{[x,1]}$. 因此

$$p(x,y)=p_X(x)p(y|x)=\frac{1}{1-x}I_{[0<x<y<1]}.$$

所以

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,Y)\mathrm{d}x=-\ln(1-y)I_{[0,1]}.$$

例 2.4.11. 既非离散又非连续的例子

设 $\xi\sim U(0,1)$, $\eta=\xi^2$, 则二维随机变量 (ξ,η) 没有密度函数.

证明: (反证) 若 η 有密度函数, 记为 $p(x,y)$, 则

$$P((\xi,\eta)\in B)=\int_B p(x,y)\mathrm{d}x\mathrm{d}y.$$

选 $B=\{(x,y):y=x^2,\forall x\in\mathbb{R}\}$, 则 $B\in\mathscr{B}(\mathbb{R}^2)$, 于是 $1=P((\xi,\eta)\in B)$, 但是 $m(B)=0$, 故

$$\int_B p(x,y)\mathrm{d}x\mathrm{d}y=0,$$

矛盾. □

注: 最后用了测度的绝对连续性, 参考实变函数书.

连续与离散之间的关系:

(1)$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y=\int_{-\infty}^{+\infty}p_Y(y)p(x|y)\mathrm{d}y$, 对应全概率公式 $P(A)=\sum\limits_{i=1}^{+\infty}P(B_i)P(A|B_i)$.

(2)$p(y|x)=\frac{p(x,y)}{p_X(x)}=\frac{p(x,y)}{\int_{-\infty}^{+\infty}p_Y(y)p(x,y)\mathrm{d}y}$. 对应于 Bayes 公式 $P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_i P(A_i)P(B|A_i)}$.

注: 求和与积分是同一回事.

§ 2.5 随机变量的独立性

定义 2.5.1

设 $X_1,\cdots,X_n$ 是 $(\Omega,\mathscr{F},P)$ 上的随机变量, 若 $\sigma(X_1),\cdots,\sigma(X_n)$ 相互独立, 则称 $X_1,\cdots,X_n$ **相互独立**, 即对任意 $B_1,\cdots,B_n\in\mathscr{B}(\mathbb{R})$, 都有

$$P(X_1\in B_1,\cdots,X_n\in B_n)=\prod_{i=1}^n P(X_i\in B_i).$$

注: $\sigma(X_i)$ 是 $\mathscr{F}$ 的子 $\sigma-$ 代数.

等价定义: **联合分布等于边际分布的乘积**, 即

$$\begin{array}{ccc} F(x_1,\cdots,x_n) & = & F_{X_1}(x_1)_{X_2}(x_2)\cdots F_{X_n}(x_n) \\ \| & & \| \\ P(X_1\in(-\infty,x_1),\cdots,X_n\in(-\infty,x_n)) & = & \prod_{i=1}^n P(X_i^{-1}((-\infty,x_i])) \end{array}$$

其中用到了 $\mathscr{C}_i\triangleq\{X_i^{-1}((-\infty,x]):x\in\mathbb{R}\}$ 是一个 π 类, 且 $\sigma(\mathscr{C}_i)=\sigma(X_i)$, 再用独立类扩张定理.

特别地, 若 $(X_1,\cdots,X_n)$ 有联合密度 $p(x_1,\cdots,x_n)$, 则

$$X_1,\cdots,X_n\text{相互独立}\Leftrightarrow p(x_1,\cdots,x_n)=p_{X_1}(x_1)\cdots p_{X_n}(x_n),\forall x=(x_1,\cdots,x_n)\in\mathbb{R}^n.$$

此时

$$\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n=\prod_{i=1}^n\int_{-\infty}^{x_i}p_{X_i}(y_i)\mathrm{d}y_i.$$

注: 离散随机变量的独立性的定义是类似的: 对于离散随机变量 X,Y, 这两个随机变量是独立的等价于 (X,Y) 的联合分布列为边际分布列的乘积.

关于随机变量的独立性, 有如下几个性质:

命题 2.5.1

若 $\xi,\eta\in\mathbb{R}$ 是独立的, 那么

$$P(\xi\le x|\eta=y)=P(\xi\le x).$$

特别地, 如果 (ξ,η) 为连续型随机变量, 那么

$$p(x|y)=p_X(x).$$

命题 2.5.2

若 ξ,η 为 $\mathbb{R}^n$ 上独立的随机向量, 那么

$$P(\xi\in A,\eta\in B)=P(\xi\in A)P(\eta\in B),A,B\in\mathscr{B}(\mathbb{R}^n).$$

注: 分量之间不一定独立, 仅为向量之间的独立.

命题 2.5.3

若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立, 则其中任意 r 个仍相互独立.

证明: 让其中的 $n-r$ 个 $B_i=\mathbb{R}$ 即可. $\square$

定义 2.5.2

称 $f:\mathbb{R}^n\to\mathbb{R}$ 是 **Borel 可测**的, 若对任意 $B\in\mathscr{B}(\mathbb{R})$, 都有 $f^{-1}(B)\in\mathscr{B}(\mathbb{R}^n)$.

注: 连续函数都是 Borel 可测的.

命题 2.5.4

若 ξ,η 相互独立, g,h 都是从 $\mathbb{R}^n$ 映往 $\mathbb{R}$ 的 Borel 可测函数, 那么 $h(\xi),g(\eta)$ 也相互独立.

证明: 利用 Borel 可测的定义即可:

$$\begin{aligned}P(h(\xi)\in B_1,g(\eta)\in B_2)&=P(\xi\in h^{-1}(B_1),\eta\in g^{-1}(B_2))\\&=P(\xi\in h^{-1}(B_1))P(\eta\in g^{-1}(B_2))\qquad(\text{独立性})\\&=P(h(\xi)\in B_1)P(g(\eta)\in B_2).\end{aligned}$$

例 2.5.5

设 $(\xi,\eta)\sim N(\mu,\Sigma)$. 则 ξ 与 η 相互独立等价于相关系数 $\rho=0$.

注: 正态分布满足 $\rho=0$ 等价于相互独立, 但是其他分布不一定有此性质! 一般来说, 若 (ξ,η) 的相关系数为 0, 那么 ξ,η 未必独立; 但是如果 ξ,η 相互独立, 则相关系数为 0.

例 2.5.6

若 (ξ,η) 服从 $G=\{(x,y):a\le x\le b,c\le y\le d\}$ 上的均匀分布, 则 $\xi\sim U(a,b),\eta\sim U(c,d)$, 且 ξ,η 相互独立.

证明: 易知 $p(x,y)=\dfrac{1}{(b-a)(d-c)}I_G(x,y)$, 则

$$\begin{aligned}p_X(x)&=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y=\frac{1}{b-a}I_{[a,b]}(x),\\p_Y(y)&=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x=\frac{1}{d-c}I_{[c,d]}(y),\end{aligned}$$

故 $p(x,y)=p_X(x)p_Y(y)$, 从而 ξ,η 相互独立. □

例 2.5.7

若 (X,Y) 的联合密度为

$$p(x,y)=\begin{cases}8xy, & 0\le x\le y\le 1,\\ 0, & \text{其他}.\end{cases}$$

问 X,Y 是否独立.

解: 容易求得

$$\begin{aligned}p_X(x)&=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y=4x(1-x^2)I_{[0,1]}(x),\\p_Y(y)&=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x=4y^3I_{[0,1]}(y),\end{aligned}$$

所以 $p_X(x)p_Y(y)\neq p(x,y)$, 故 X,Y 不独立. □

§ 2.6 随机变量的函数及其分布

设 ξ 是随机变量, g 是 $\mathbb{R}$ 上的函数, 问 $g(\xi)$ 何时是个随机变量.

定义 2.6.1

设 $(\Omega,\mathscr{F})$ 和 $(E,\mathscr{E})$ 是两个可测空间, $f:\Omega\to E$ 是映射. 若对任意 $B\in\mathscr{E}$ 都有 $f^{-1}(B)\in\mathscr{F}$, 称 f 为 $\mathscr{F}/\mathscr{E}$ **可测的**.

特别地, 若 $(\Omega,\mathscr{F})=(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$, $(E,\mathscr{E})=(\mathbb{R},\mathscr{B}(\mathbb{R}))$, 那么 f 为 **Borel 可测函数**.

若 ξ 是随机变量, $g:\mathbb{R}\to\mathbb{R}$ 是 Borel 可测函数, 那么 $g(\xi)$ 就是随机变量! 即对任意 $B\in\mathscr{B}(\mathbb{R}),[g(\xi)\in B]=[\xi\in g^{-1}(B)]\in\mathscr{F}$.

例 2.6.1

分段连续函数、分段单调函数（如分布函数）都是 Borel 可测函数.

命题 2.6.2

设 $f:\mathbb{R}^n\to\mathbb{R}$ 是个函数, 下面命题等价:

(1)f 为可测函数.

(2)$\forall a\in\mathbb{R},[f<a]\in\mathscr{B}(\mathbb{R}^n)$. （这里 $[f<a]$ 表示 $(-\infty,a)$ 在 f 上的原像, $f^{-1}(-\infty,a)$. ）

(3)$\forall a\in\mathbb{R},[f\le a]\in\mathscr{B}(\mathbb{R}^n)$.

(4)$\forall a\in\mathbb{R},[f\ge a]\in\mathscr{B}(\mathbb{R}^n)$.

(5)$\forall a\in\mathbb{R},[f>a]\in\mathscr{B}(\mathbb{R}^n)$.

(6)$\forall a,b\in\mathbb{R}$ 且 $a<b$, 都有 $[a<f\le b]\in\mathscr{B}(\mathbb{R}^n)$.

注: 这与 Borel σ-代数的生成很像.

2.6.1 离散型随机变量函数的分布

若离散型随机变量 ξ 的分布列为 $\{x_i;p_i\}$, $g:\mathbb{R}\to\mathbb{R}$ 为 Borel 可测函数, 那么 $g(\xi)$ 是离散型随机变量, 分布列为 $\{g(x_i);p_i\}$. 若 g 不是单射, 可以合并其中的相同项.

定理 2.6.3. 离散卷积公式

若 ξ,η 是相互独立的随机变量, **且取非负整数值**, 分布列分别为 $\{k;a_k\}$ 和 $\{k;b_k\}$. 则随机变量 $\zeta=\xi+\eta$ 的分布列为 $P(\zeta=k)=\sum\limits_{i=0}^{k}a_ib_{k-i}$. 这叫**卷积公式**.

证明: 只需注意到 $[\zeta=k]=[\xi=0,\eta=k]+[\xi=1,\eta=k-1]+\cdots+[\xi=k,\eta=0]$. □

例 2.6.4. Poisson 分布可加性

设 $X\sim P(\lambda_1),Y\sim P(\lambda_2)$, 且 X,Y 相互独立, 证明: $X+Y\sim P(\lambda_1+\lambda_2)$.

证明: $P(X=k)=\dfrac{\lambda_1^k}{k!}e^{-\lambda_1},P(Y=k)=\dfrac{\lambda_2^k}{k!}e^{-\lambda_2}$, 由卷积公式,

$$\begin{aligned}P(X+Y=k)&=\sum_{i=0}^{k}P(X=i)P(X=k-i)=\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}e^{-\lambda_1}\frac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2}\\&=e^{-(\lambda_1+\lambda_2)}k!\sum_{i=0}^{k}\frac{\lambda_1^i\lambda_2^{k-i}k!}{i!(k-i)!}=e^{-(\lambda_1+\lambda_2)}k!(\lambda_1+\lambda_2)^k.(\text{二项式展开})\square\end{aligned}$$

注: 推广：有限个独立 Poisson 分布随机变量之和的分布仍为 Poisson 分布:

$$P(\lambda_1)*\cdots*P(\lambda_n)=P(\lambda_1+\cdots+\lambda_n).$$

2.6.2 连续型随机变量函数的分布

已知 ξ 的分布函数 $F(x)$ 或密度函数 $p(x)$, 求 $\eta=g(\xi)$ 的分布函数 $G(y)$ 或密度函数 $\varphi(y)$. 注意到

$$P(\eta\le y)=P(g(\xi)\le y)=P(\xi\in g^{-1}(-\infty,y])=\int_{g^{-1}(-\infty,y]}p(x)\mathrm{d}x=\int_{\{g(x)\le y\}}p(x)\mathrm{d}x.$$

命题 2.6.5

设 ξ 为连续型随机变量, 密度函数为 $p(x)$. 若 $g:\mathbb{R}\to\mathbb{R}$ 严格单调, 且其反函数 g^{-1} 有连续导数, 则 $\eta=g(\xi)$ 是密度为 $p(g^{-1}(y))|g^{-1}(y)'|$ 连续型 r.v..

证明: 对任意 $y\in\mathbb{R}$, 不妨设 g 严格单调递增, 则

$$P(\eta\le y)=P(g(\xi)\le y)=\int_{\{g(x)\le y\}}p(x)\mathrm{d}x\overset{t=g(x)}{=\!=\!=}\int_{-\infty}^{y}p(g^{-1}(t))|g^{-1}(t)'|\mathrm{d}t.\qquad\square$$

例 2.6.6

若随机变量 ξ 的密度函数为 $p(x)$, 令 $\eta=a\xi+b,a\neq0$, 求 η 的密度函数 $q(y)$.

解: $g^{-1}(y)=\dfrac{y-b}{a}$, 则 $q(y)=p\left(\dfrac{y-b}{a}\right)\dfrac{1}{|a|}$. $\square$

例 2.6.7

设 $\xi\sim N(\mu,\sigma^2)$, 求 $\eta=e^{\xi}$ 的密度函数.

解: $g^{-1}(y)=\ln y(y>0)$, 则 $p_\eta(y)=p(\ln y)\dfrac{1}{y}=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}\dfrac{1}{y},y>0$. 这叫**对数正态分布**. $\square$

例 2.6.8

设 $\theta\sim U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right),\psi=\tan\theta$, 求 ψ 的密度函数.

解: $g^{-1}(x)=\arctan x$, 则 $p_\psi(y)=p(\arctan y)\dfrac{1}{1+y^2}=\dfrac{1}{\pi(1+y^2)},y\in\mathbb{R}$. 称 ψ 符合 **Cauchy 分布**.

命题 2.6.9

设随机变量 ξ 的密度函数为 $p(x)$, g 在互不相交的区间 $I_1,\cdots,I_n$ 上分段严格单调, 且反函数分别为 $h_1,\cdots,h_n$, 满足 $h_1',\cdots,h_n'$ 连续, 则 $\eta=g(\xi)$ 的密度函数为 $\sum\limits_{i=1}^{n}p(h_i(y))|h_i'(y)|$.

证明: 对任意 $y \in \mathbb{R}$, 我们有

$$
\begin{aligned}
P(\eta \leq y) = P(g(\xi) \leq y) &= \int_{\{g(x)\leq y\}} p(x)\mathrm{d}x = \int_{\sum\limits_{i=1}^{n}[g(x)\leq y]\cap I_i} p(x)\mathrm{d}x \\
&= \sum_{i=1}^{n}\int_{[g(x)\leq y]\cap I_i} p(x)\mathrm{d}x = \sum_{i=1}^{n}\int_{-\infty}^{y} p(h_i(y))|h_i'(y)|\mathrm{d}y \\
&= \int_{-\infty}^{y}\sum_{i=1}^{n} p(h_i(y))|h_i'(y)|\mathrm{d}y. \qquad \square
\end{aligned}
$$

例 2.6.10

设 $\xi \sim N(0,1)$, 求 $\eta = \xi^2$ 的密度函数.

解: 【方法一】直接运用前一命题, 可得: 当 $y>0$ 时, $p_\eta(y) = \dfrac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}$; 当 $y \leq 0$ 时, $p_\eta(y) = 0$.

【方法二】直接计算也行: 当 $y \leq 0$ 时, $P(\xi^2 \leq y) = 0$, 则 $p_{\xi^2}(y) = 0$. 当 $y > 0$ 时, $P(\xi^2 \leq y) = P(-\sqrt{y} \leq \xi \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$; 求导得

$$
p_{\xi^2}(y) = p(\sqrt{y})\frac{1}{2\sqrt{y}} + p(-\sqrt{y})\frac{1}{2\sqrt{y}} = \cdots.
$$

注: 对 Γ 分布取 $r = \dfrac{n}{2}, \lambda = \dfrac{1}{2}$ 可以得到 χ^2 **分布**: $\Gamma\left(\dfrac{n}{2}, \dfrac{1}{2}\right) = \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, x > 0$. 而 χ^2 分布取 $n=1$ 可以本例子的分布.

注: n 个 i.i.d. 标准正态分布的平方和为 χ^2 分布.

2.6.3 随机向量的函数的分布

假定 $(\xi_1, \cdots, \xi_n)$ 的联合密度为 $p(x_1, \cdots, x_n)$, $g : \mathbb{R}^n \to \mathbb{R}$ 为 Borel 可测函数, 下面讨论 $\eta = g(\xi_1, \cdots, \xi_n)$ 的分布函数或密度函数.

$$
G(y) = P(\eta \leq y) = P(g(\xi_1, \cdots, \xi_n) \leq y) = \int_{[g(x_1,\cdots,x_n)\leq y]} p(x_1, \cdots, x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n.
$$

1. 和的分布

设 (ξ_1, ξ_2) 的联合密度为 $p(x_1, x_2)$, 求 $\eta = \xi_1 + \xi_2$ 的密度.

$$
\begin{aligned}
G(y) = P(\xi_1 + \xi_2 \leq y) &= \int_{x_1+x_2\leq y} p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \int_{-\infty}^{+\infty}\mathrm{d}x_2\int_{-\infty}^{y-x_2} p(x_1, x_2)\mathrm{d}x_1 = \int_{-\infty}^{+\infty}\mathrm{d}x_2\int_{-\infty}^{y} p(x_1 - x_2, x_2)\mathrm{d}x_1 \\
&= \int_{-\infty}^{y}\mathrm{d}x_1\int_{-\infty}^{+\infty} p(x_1 - x_2, x_2)\mathrm{d}x_2.
\end{aligned}
$$

因此

$$
p_\eta(y) = \int_{-\infty}^{+\infty} p(y-x, x)\mathrm{d}x = \int_{-\infty}^{+\infty} p(x, y-x)\mathrm{d}x.
$$

当 ξ_1, ξ_2 独立时,

$$
p_\eta(y) = \int_{-\infty}^{+\infty} p_1(y-x)p_2(x)\mathrm{d}x = \int_{-\infty}^{+\infty} p_1(x)p_2(y-x)\mathrm{d}x.
$$

这就是连续型随机变量的**卷积公式**.

例 2.6.11. 正态分布可加性

设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 且 X, Y 独立, 则

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

注: 不可以不独立, 例如 $X \sim N(0,1), Y = -X \sim N(0,1)$, 但 $X + Y$ 恒为 0.

证明: 记 $A = \dfrac{1}{\sigma_1^2 + \sigma_2^2}, B = \dfrac{y - \mu_1}{\sigma_1^2} + \dfrac{\mu_2}{\sigma_2^2}$. 用卷积公式,

$$\begin{aligned}
p_Y(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{(y-x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}\right\} \mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[A\left(x - \frac{B}{A}\right)^2 + A(y-\mu_1-\mu_2)^2\right]\right\} \mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{A}{2}(y-\mu_1-\mu_2)^2\right\}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[A\left(x - \frac{B}{A}\right)^2\right]\right\} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left\{-\frac{A}{2}(y-\mu_1-\mu_2)^2\right\}.
\end{aligned}$$

因此根据独立性即可得 $p(x,y) = p_X(x)p_Y(y)$, 从而 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. □

注: 可以推广到 n 个的情形!

例 2.6.12. Γ 分布的可加性

设 $X \sim \Gamma(r_1, \lambda), Y \sim \Gamma(r_2, \lambda)$, 且 X, Y 独立. 则

$$X + Y \sim \Gamma(r_1 + r_2, \lambda).$$

证明: 设 $Z = X + Y$. 当 $z \le 0$ 时, $p(z) = 0$. 当 $z > 0$ 时, 用卷积公式,

$$\begin{aligned}
p(z) &= \int_{-\infty}^{+\infty} p_X(z-x)p_Y(x)\mathrm{d}x \\
&= \int_0^z p_X(z-x)p_Y(x)\mathrm{d}x \\
&= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)}\int_0^z (z-x)^{r_1-1}x^{r_2-1}e^{-\lambda(z-x)}e^{-\lambda x}\mathrm{d}x \\
&= \frac{\lambda^{r_1+r_2}e^{-\lambda z}z^{r_1+r_2-1}}{\Gamma(r_1)\Gamma(r_2)}\underbrace{\int_0^1 (1-t)^{r_1-1}t^{r_2-1}\mathrm{d}t}_{Beta\text{函数}} \\
&= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)}e^{-\lambda z}z^{r_1+r_2-1}.
\end{aligned}$$

最后用了 Beta 函数与 Γ 函数关系式:

$$B(r_1, r_2) = \frac{\Gamma(r_1)\Gamma(r_2)}{\Gamma(r_1+r_2)}.$$

注: (1)$\underbrace{E(\lambda) * E(\lambda) * \cdots * E(\lambda)}_{m\text{个}} = \Gamma(m, \lambda)$.

(2)m 个独立的 χ^2 分布之和仍为 χ^2 分布: 即

$$\xi_1^2+\cdots+\xi_n^2\sim\Gamma\left(\frac{n}{2},\frac{1}{2}\right).$$

2. 商的分布

若 (ξ_1,ξ_2) 的联合密度为 $p(x_1,x_2)$, 令 $\eta=\dfrac{\xi_1}{\xi_2}$, 求 η 的密度函数. 对任意 $y\in\mathbb{R}$, 有

$$\begin{aligned}
F_\eta(y)=P(\eta\leq y)&=P\left(\frac{\xi_1}{\xi_2}\leq y\right)\\
&=\int_{\left\{\frac{x_1}{x_2}\leq y\right\}}p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\
&=\int_{-\infty}^{0}\mathrm{d}x_2\int_{yx_2}^{+\infty}p(x_1,x_2)\mathrm{d}x_1+\int_{0}^{+\infty}\mathrm{d}x_2\int_{-\infty}^{yx_2}p(x_1,x_2)\mathrm{d}x_1\\
&=\int_{-\infty}^{0}\mathrm{d}x_2\int_{y}^{+\infty}x_2p(x_2t,x_2)\mathrm{d}t+\int_{0}^{+\infty}\mathrm{d}x_2\int_{-\infty}^{y}x_2p(x_2t,x_2)\mathrm{d}t\\
&=\int_{-\infty}^{t}\int_{-\infty}^{+\infty}|x_2|p(x_2t,x_2)\mathrm{d}x_2.
\end{aligned}$$

则

$$p_\eta(y)=F_\eta'(y)=\int_{-\infty}^{+\infty}|x_2|p(x_2y,x_2)\mathrm{d}x_2.$$

3. 顺序统计量的分布

若 $\{\xi_i\}_{i\leq n}$ 是 i.i.d., 分布函数 $F(x)$, 密度函数 $p(x)$. 令

$$\xi_n^*=\max\{\xi_1,\cdots,\xi_n\},\xi_1^*=\min\{\xi_1,\cdots,\xi_n\}.$$

则 $\xi_1\vee\xi_2(\omega)\leq\max\{\xi_1(\omega),\xi_2(\omega)\}$, 它是 Borel 可测的, 因为

$$[\xi_1\vee\xi_2\in B]\in\mathscr{F}\Leftrightarrow[\xi_1\leq x\text{且}\xi_2\leq x]\in\mathscr{F},\forall B\in\mathscr{B}(\mathbb{R}),x\in\mathbb{R}.$$

同理, $\xi_1\wedge\xi_2(\omega)$ 也是 Borel 可测的, 回顾 Borel 可测的等价命题.

下面考虑 ξ_1^* 和 ξ_2^* 的密度函数.

(1) 对任意 $y\in\mathbb{R}$, 由独立性,

$$P(\xi_n^*\leq y)=P(\xi_1\leq y,\cdots,\xi_n\leq y)=P(\xi_1\leq y)P(\xi_2\leq y)\cdots P(\xi_n\leq y)=(F(y))^n.$$

从而

$$p_{\xi_n^*}(y)=n(F(y))^{n-1}p(y).$$

(2) 对任意 $y\in\mathbb{R}$, 由独立性,

$$P(\xi_1^*\leq y)=1-P(\xi_1^*>y)=1-P(\xi_1>y)P(\xi_2>y)\cdots P(\xi_n>y)=1-(1-F(y))^n.$$

所以

$$p_{\xi_1^*}(y)=n(1-F(y))^{n-1}p(y).$$

(3)(ξ_1^*,ξ_n^*) 的联合分布. 令 $G(x,y)=P(\xi_1^*\leq x,\xi_n^*\leq y)$ 为联合分布函数.

若 $x \geq y$, 则

$$G(x,y) = P(\xi_n^* \leq y) = F^n(y).$$

若 $x \leq y$, 则根据事件的关系式 $AB = A - A\overline{B}$ 可得

$$G(x,y) = P(\xi_n^* \leq y) - P(\xi_n^* \leq y, \xi_1^* > x) = F^n(y) - (F(y) - F(x))^n.$$

所以

$$p_{(\xi_1^*,\xi_n^*)}(x,y) = \begin{cases} 0, & x \geq y, \\ n(n-1)[F(y)-F(x)]^{n-2}p(x)p(y), & x < y. \end{cases}$$

(4)$Y = \xi_n^* - \xi_1^*$ 的联合分布:

$$G(y) = P(\xi_n^* - \xi_1^* \leq y) = \int_{\{x_2 - x_1 \leq y\}} p_{(\xi_1^*,\xi_n^*)}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2.$$

再利用 (3) 的结论即可得到.

2.6.4 随机向量的变量替换

命题 2.6.13

假设 (ξ,η) 的联合密度函数为 $p(x,y)$, 则

$$\xi与\eta独立 \Leftrightarrow 存在可测函数 h,g:\mathbb{R}\to[0,+\infty), p(x,y) = h(x)g(y).$$

证明: “$\Rightarrow$” : 记 $h(x) = \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}y, g(y) = \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x$, 由 ξ,η 独立, 故

$$p(x,y) = h(x)g(y).$$

“$\Leftarrow$” : 由于

$$\begin{aligned} p_\xi(x) &= \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}y = h(x)\int_{-\infty}^{+\infty} g(y)\mathrm{d}y \\ p_\eta(y) &= \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x = g(y)\int_{-\infty}^{+\infty} h(x)\mathrm{d}x \end{aligned}$$

且

$$p_\xi(x)p_\eta(y) = h(x)g(y)\int_{-\infty}^{+\infty} h(x)g(y)\mathrm{d}y\mathrm{d}x = h(x)g(y)\cdot 1 = p(x,y).$$

结论证完. □

注: 注意这里 h,g 并不是密度函数, 它们与密度函数相差了个常数.

下面设 $(\xi_1,\cdots,\xi_n)$ 的密度函数为 $p(x) \triangleq p(x_1,\cdots,x_n)$, 其中 $x = (x_1,\cdots,x_n) \in \mathbb{R}^n$. 且 $g_i:\mathbb{R}^n \to \mathbb{R}$ 是 Borel 可测函数, $1 \leq i \leq m$. 令 $\eta_i \triangleq g_i(\xi_1,\cdots,\xi_n), 1 \leq i \leq m$. 则

$$P(\eta_1 \leq y_1,\cdots,\eta_n \leq y_n) = \int_{\{x:g_i(x)\leq y_i, 1\leq i\leq m\}} p(x)\mathrm{d}x.$$

若 g_i^{-1} 存在且有连续偏导数, 且 $\boxed{m = n}$, 此时令

$$u_i = g_i(x) = g_i(x_1,\cdots,x_n), 1 \leq i \leq n,$$

那么 $x_i = g_i^{-1}(u_1, \cdots, u_n), 1 \le i \le n$. 从而

$$P(\eta_1 \le y_1, \cdots, \eta_n \le y_n) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} p\left(g_1^{-1}(u_1, \cdots, u_n), \cdots, g_n^{-1}(u_1, \cdots, u_n)\right) |\det J| du_1 \cdots du_n.$$

其中,

$$J = \begin{pmatrix} \dfrac{\partial g_1^{-1}}{\partial x_1} & \cdots & \dfrac{\partial g_1^{-1}}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n^{-1}}{\partial x_1} & \cdots & \dfrac{\partial g_n^{-1}}{\partial x_n} \end{pmatrix}$$

所以

$$p_{(\eta_1, \cdots, \eta_n)}(y_1, \cdots, y_n) = \begin{cases} p(g_1^{-1}(y_1, \cdots, y_n), \cdots, g_n^{-1}(y_1, \cdots, y_n)) |\det J|, & y_i \in g_i(\mathbb{R}), \\ 0, & \text{其他}. \end{cases}$$

例 2.6.14. 2019 期末

假定随机向量 (ξ_1, ξ_2) 的联合密度函数为 $p(x_1, x_2)$, 令

$$\eta_1 = a\xi_1 + b\xi_2, \eta_2 = c\xi_1 + d\xi_2,$$

其中 $ad - bc \ne 0$, 求 (η_1, η_2) 的密度函数 $q(y_1, y_2)$.

解: 易知 $g_1(x_1, x_2) = ax_1 + bx_2, g_2(x_1, x_2) = cx_1 + dx_2$, 所以

$$g_1^{-1}(y_1, y_2) = \frac{d}{ad - bc} y_1 - \frac{b}{ad - bc} y_2$$
$$g_2^{-1}(y_1, y_2) = -\frac{c}{ad - bc} y_1 + \frac{a}{ad - bc} y_2.$$

从而

$$J = \det \begin{pmatrix} \dfrac{d}{ad - bc} & -\dfrac{b}{ad - bc} \\ -\dfrac{c}{ad - bc} & \dfrac{a}{ad - bc} \end{pmatrix} = \frac{1}{ad - bc}.$$

故

$$q(y_1, y_2) = p\left(\frac{d}{ad - bc} y_1 - \frac{b}{ad - bc} y_2, -\frac{c}{ad - bc} y_1 + \frac{a}{ad - bc} y_2\right) \frac{1}{|ad - bc|}.$$

例 2.6.15

若 $\xi \sim \chi^2(m), \eta \sim \chi^2(n)$, 且 ξ, η 独立. 求 $\alpha = \xi + \eta$ 和 $\beta = \frac{\frac{\xi}{m}}{\frac{\eta}{n}}$ 的密度函数 $q(u, v)$.

解: 由题设可知

$$p_{(\xi, \eta)}(x, y) = p_\xi(x) p_\eta(y) = \left(2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)\right)^{-1} x^{\frac{m}{2} - 1} y^{\frac{n}{2} - 1} e^{-\frac{x+y}{2}} (x, y > 0).$$

作 $u=x+y, v=\frac{\frac{x}{m}}{\frac{y}{n}}$, 则 $x=\frac{muv}{n+mv}, y=\frac{nu}{n+mv}$, 从而

$$J=\det\begin{pmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{pmatrix}=-\frac{m}{n}\frac{u}{1+(\frac{m}{n}v)^2}.$$

故

$$q(u,v)=\underbrace{\left(2^{\frac{m+n}{2}}\Gamma\left(\frac{m+n}{2}\right)\right)^{-1}u^{\frac{m+n}{2}-1}e^{-\frac{u}{2}}}_{u\text{的函数},\sim\chi^2(m+n)}\cdot\underbrace{\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}v^{\frac{m}{2}-1}}{(1+\frac{m}{n}v)^{\frac{m+n}{2}}}}_{v\text{的函数}}.$$

注: 根据 $q(u,v)$ 表达式的构成, α,β 是独立的.

对 $\xi_1,\cdots,\xi_n$, $\eta_i=g_i(\xi_1,\cdots,\xi_n)$, $1\le i\le m$, 当 $m<n$ 时无法定义 g_i^{-1}. 可以重新定义

$$\bar{\eta}_i=\begin{cases}\eta_i, & 1\le i\le m,\\ f_i(\xi_1,\cdots,\xi_n), & m+1\le i\le n,\end{cases}$$

那么 $\{\bar{\eta}_i\}$ 有 n 个: $g_1,\cdots,g_m,f_{m+1},\cdots,f_n$. 若 $\{\bar{\eta}_i\}$ 都有反函数且它们的导数连续, 可考虑算 $(\bar{\eta}_1,\cdots,\bar{\eta}_n)$ 的联合密度, 自然知道 $\bar{\eta}_1,\cdots,\bar{\eta}_n$ 的密度.

例 2.6.16

设 $\xi\sim N(0,1),\eta\sim\chi^2(n)$ 且 ξ 与 η 独立. 令 $T=\frac{\xi}{\sqrt{\frac{\eta}{n}}}$, 求 T 的密度函数.

解: 令 $S=\eta$. 作 $t=\frac{x}{\sqrt{y/n}}, s=y$, 则 $x=t\left(\frac{s}{n}\right)^{1/2}, y=s, J=\det\begin{pmatrix}\frac{\partial x}{\partial s} & \frac{\partial x}{\partial t}\\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t}\end{pmatrix}=\left(\frac{s}{n}\right)^{1/2}$ 从而

$$p_{(S,T)}(s,t)=p_{\xi,\eta}\left(t\left(\frac{s}{n}\right)^{1/2},s\right)|J|=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2s}{2n}}\cdot\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}s^{\frac{n}{2}-1}e^{-\frac{s}{2}}\left(\frac{s}{n}\right)^{1/2}.$$

则 T 的密度函数为

$$p_T(t)=\int_0^\infty p_{(S,T)}(s,t)\mathrm{d}s=\cdots=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

(复习 Gamma 函数的一些性质!) 称 T 符合**自由度是** t **的分布**. □

下面设 F 是分布函数, 令

$$F^{-1}(y)=\sup\{x|F(x)<y\}, y\in(0,1).$$

引理 2.6.17

$F^{-1}(y)\le x\Leftrightarrow y\le F(x)$.

证明: 这个命题不显然!

"$\Leftarrow$": 由 $y>F(x)$ 与 F 的右连续性可知存在 $\delta>0$, 使得 $y>F(x+\delta)$. 由 F^{-1} 定义可知 $x+\delta\le F^{-1}(y)$, 因此 $x\le F^{-1}(y)-\delta\le F^{-1}(y)$.

“$\Rightarrow$”: 由 $x < F^{-1}(y)$ 可知存在 $x^* \in \{x|F(x) < y\}$ 使得 $x < x^*$, (若不然, $x = F^{-1}(y)$, 那么 x 就是 $\{x|F(x) < y\}$ 的上确界). 从而 $F(x) \le F(x^*) < y$. □

定理 2.6.18

设 $F(x)$ 是随机变量 X 的分布函数, $X \sim U(0,1)$, 那么 $F^{-1}(X)$ 的分布函数为 $F(x)$.

证明: $P(F^{-1}(X) \le y) = P(X \le F(y))$,
若 $y < 0$, 则 $F(y) = 0$, 此时 $P(X \le 0) = F(0) = 0$.
若 $y > 1$, 则 $F(y) = 1$, 此时 $P(X \le 1) = F(1) = 1$.
若 $0 \le y \le 1$, 则 $F(y) = y$, $P(X \le y) = F(y)$, 则 $P(X \le F(y)) = F(y)$. □

定理 2.6.19

设随机变量 ξ 具有连续的分布函数 $F(x)$, 则 $\theta = F(\xi)$ 服从均匀分布 $U[0,1]$.

证明: 由于 $F(x)$ 单调不减且连续, 所以

$$P(\theta < y) = P(F(\xi) < y) = P(\xi < F^{-1}(y)) = F(F^{-1}(y)) = y.$$

因此 $\theta \sim U[0,1]$. □

注: F^{-1} 的分析性质:

- $F(F^{-1}(y)) \ge y$;
- $F^{-1}(F(x)) \le x$;
- 若 F 在 $x = F^{-1}(y)$ 处连续, 则 $F(F^{-1}(y)) = y$;
- F^{-1} 是左连续的.

§ 2.7 习题

1. 已知随机向量 (X,Y) 的联合密度

$$p(x,y) = \begin{cases} 6e^{-2x-3y}, & x > 0, y > 0, \\ 0, & \text{其他}. \end{cases},$$

求 $P(X < 1, Y > 1)$ 与 $P(X > Y)$.

2. **(2019 期末)** 已知随机向量 (X,Y) 的联合密度函数为

$$p(x,y) = \begin{cases} 2 - x - y, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{其他}. \end{cases}$$

求 $P(X > 2Y)$.

3. 一射手进行射击, 击中目标的概率为 $p(0 < p < 1)$. 射击直至击中目标两次为止. 设 X 表示首次击中目标所进行的射击次数, 以 Y 表示总的进行射击的次数. 试求 X 与 Y 的联合分布列与条件分布列. (提示: X 服从几何分布)

4. 设随机变量 X, Y 独立, 且 $X \sim U(0,1), Y \sim E(1)$, 求 $P(Y \le X)$ 与 $P(X + Y \le 1)$.

5. (二项分布的可加性) 设 $X \sim b(n,p), Y \sim b(m,p)$, 且 X, Y 相互独立. 证明: $X + Y \sim b(n+m,p)$.

6. 若 ξ, η 是相互独立的随机变量, $\xi \sim N(0,1), \eta \sim N(0,1)$.
   (1) **(2019 期末)** 求 $\rho = \sqrt{\xi^2 + \eta^2}$ 的密度.

(2) **(2012 期中)** 证明: $\rho=\sqrt{\xi^2+\eta^2}$ 与 $\varphi=\arctan\dfrac{\eta}{\xi}$ 是相互独立的.

7. **(2022 南京大学推免)** 设 X,Y 为独立同分布的随机变量, 且 X 服从参数为 1 的指数分布, 求 $\dfrac{X}{X+Y}$ 的密度函数. (这是李贤平《概率论基础》第四章 38 题的简化版.)

8. **(2022 南京大学推免)** 设 $X_1,X_2,\cdots,X_n$ 为独立同分布随机变量, 且 X_1 的密度函数为 $p(x)$. 证明:
   (1) $P(\max\{X_1,\cdots,X_n\}=X_1)=\dfrac{1}{n}$;
   (2) 随机变量 $\max\{X_1,\cdots,X_n\}$ 与 $I_{[X_1=\max\{X_1,\cdots,X_n\}]}$ 相互独立.

9. **(难)** 设 ξ 是随机变量（不知道是离散还是连续）, $\eta\sim N(0,1)$, 且 ξ 与 η 独立, 则 $\xi+\eta$ 为连续型随机变量. (即有密度函数).
   注: 可以推出 $\xi+\dfrac{1}{n}\eta$ 也有密度, 以后学联合分布积分就可以做.

10. **(2024 北京大学期中)** 构造随机变量 $X_1\sim E(\lambda_1)$，$X_2\sim E(\lambda_2)$，$\lambda_1<\lambda_2$，使得 $X_1\geq X_2$ a.s.

11. **(2022 年丘成桐大学生数学竞赛决赛)** 设 $\mathcal{N}=\{1,2,3,4,\cdots\}$ 表示正整数全体, $\mathcal{P}=\{2,3,5,7,\cdots\}$ 表示素数全体. 记 $a\mid b$ 表示 a 整除 b. 固定实数 $s>1$, 令 $\zeta(s)=\sum\limits_{n=1}^{\infty}n^{-s}$, 定义 $\mathcal{N}$ 上的概率测度为 $P_s(\{n\})=\dfrac{1}{\zeta(s)}n^{-s}, n\in\mathcal{N}$. 对任意 $p\in\mathcal{P}$, 定义 $\mathcal{N}$ 上的随机变量 X_p 为 $X_p(n)=\mathbf{1}_{\{p|n\}}(n)$, $n\in\mathcal{N}$, 其中 $\{p|n\}$ 表示事件 $\{n:p|n\}\subset\mathcal{N}$.
    (1) 集合 $\{X_p:p\in\mathcal{P}\}$ 中的随机变量在 P_s 的意义下是否相互独立?
    (2) 用概率方法证明 Euler 恒等式 $\dfrac{1}{\zeta(s)}=\prod\limits_{p\in\mathcal{P}}(1-p^{-s})$.

第 3 章　数字特征与特征函数

§ 3.1　数学期望

定义 3.1.1. 离散型随机变量的数学期望

设 ξ 是离散型随机变量, 分布列为 $\{x_i : p_i\}_{i\geq 1}$. 若级数 $\sum_{i=1}^{+\infty} x_i p_i$ 绝对收敛, 则称该级数为 ξ 的**数学期望 (均值)**, 记为 $\mathbb{E}\xi$.

当 $\sum_{i=1}^{\infty} |x_i| p_i = \infty$ 时, 称 ξ 的数学期望不存在. 期望由分布决定, 如果两个随机变量的分布一样, 那么数学期望也一样.

例 3.1.1

常见离散型随机变量分布的分布列与数学期望, 其中 $p+q=1$. （请自行证明, 复习一下级数计算）

名称	记号	分布列	数学期望	方差
两点分布	$\xi \sim b(1,p)$		p	pq
二项分布	$\xi \sim b(n,p)$	$P(x=k) = C_n^k p^k q^{n-k}$	np	npq
Poisson 分布	$\xi \sim P(\lambda)$	$P(x=k) = e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ
几何分布	$\xi \sim g(k,p)$	$P(x=k) = q^{k-1}p$	$\frac{1}{p}$	

定义 3.1.2. 连续型随机变量的数学期望

设 ξ 是密度函数为 $p(x)$ 的随机变量. 若

$$\int_{-\infty}^{+\infty} |x| p(x) \mathrm{d}x < \infty,$$

称 $\int_{-\infty}^{+\infty} x p(x) \mathrm{d}x$ 为 ξ 的数学期望, 记为 $\mathbb{E}\xi$.

例 3.1.2

常见连续型随机变量的密度函数与数学期望.

名称	记号	密度函数	数学期望	方差
均匀分布	$X \sim U[a,b]$	$p(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b],\\ 0, & 其他,\end{cases}$	$\dfrac{b+a}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布	$X \sim E(\lambda)$	$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}$	$\dfrac{1}{\lambda}$	
Γ 分布	$X \sim \Gamma(r,\lambda)$	$p(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}$		
正态分布	$X \sim N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$	μ	σ^2
Cauchy 分布		$p(x)=\dfrac{\theta}{\pi(x^2+\theta^2)}, \theta>0.$	不存在	不存在

期望的最本质定义是:

$$\boxed{\mathbb{E}\xi \triangleq \int_\Omega \xi(\omega)P(\mathrm{d}\omega)}$$

这是随机变量关于概率测度的积分.

设 ξ 是 $(\Omega,\mathscr{F},P)$ 上的随机变量, $F(x)$ 是 ξ 的分布函数. 下面用 $I_A(x)=\begin{cases}1, & x\in A\\ 0, & x\notin A\end{cases}$ 为示性函数.

定义 3.1.3. 非负简单函数

指形如 $\sum\limits_{i=1}^n a_iI_{A_i}$ 的函数, 其中 A_i 是 Borel 可测集, $a_i\geq 0, 1\leq i\leq n$.

对于 Lebesgue 可测的非负函数 $f(x)$, 可以找一列简单函数 $\{f_n\}\nearrow f$. 把 f_n 的测度定义为

$$\lambda(f_n)\triangleq\sum_{i=1}^n a_i\lambda(A_i), 其中\lambda(A)\triangleq\int_{I_A(x)}\mathrm{d}x.$$

如此定义, 在不同 $\{f_n\}$ 的极限都相同, 把这个极限定义为

$$\lambda(f)\triangleq\lim_{n\to\infty}\lambda(f_n).$$

我们把它写成一个定理如下:

定理 3.1.3. 可测函数构造

设 X 是可测空间, $f\geq 0$ 是 X 上 Lebesgue 可测函数（其中 f 可以取正无穷）, 则存在 X 上的非负简单函数序列 $\{f_n\}$ 满足: f_n 关于 n 单调上升, 且

$$\lim_{n\to\infty}f_n(x)=f(x), \forall x\in X.$$

特别地, 若 f 有界, 则上述收敛是一致的.

证明: 构造性证明. 定义

$$f_n=\sum_{i=0}^{n\cdot 2^{n-1}}\frac{i}{2^n}I_{E_{n_i}}+nI_{F_n}.$$

其中,

$$E_{n_i} = f^{-1}\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right]\right), F_n = f^{-1}[n, +\infty).$$

它们分别用来处理有界部分与无穷部分. 按照此定义, 显然有 f_n 关于 n 单调上升.

(1) 当 $f(x) < \infty$ 时, 对充分大的 n, $F_n = \varnothing$, 所以

$$f_n(x) \leq f(x) \leq f_n(x) + \frac{1}{2^n}, \Leftrightarrow |f(x) - f_n(x)| \leq \frac{1}{2^n}.$$

从而 $f_n(x) \to f(x)$ 且为一致收敛.

(2) 当 $f(x) = \infty$ 时, 对充分大的 n, $S_n(x) = n \to \infty = f(x)(n \to \infty)$.

综上, $\lim\limits_{n\to\infty} f_n(x) = f(x)$. □

注: 一般地, $\lambda(f) = \lambda(f^+) - \lambda(f^-)$. (正负与负部相减). 这里 $\lambda(f)$ 可以为无穷.

下面再回来考虑概率空间与概率测度, 把上面的 f_n 换成 ξ_n, 把 f 换成 ξ, 那么

$$\begin{aligned}
\mathbb{E}\xi &= \sum_{i=0}^{n\cdot 2^{n-1}} \frac{i}{2^n}\mathbb{E}I_{E_{n_i}} + n\mathbb{E}I_{F_n} \\
&= \sum_{i=0}^{n\cdot 2^{n-1}} \frac{i}{2^n}\left[F\left(\frac{i+1}{2^n}\right) - F\left(\frac{i}{2^n}\right)\right] + n(1 - F(n)),
\end{aligned}$$

(注意均匀分布 $\mathbb{E}I_A = P(A)$), 当 $n \to \infty$ 时, 这就是个积分.

注: 回忆

$$F(x) = P(\xi \leq x) = P \circ \xi^{-1}((-\infty, x]),$$

利用 Lebesgue-Stieltjes 积分来重写期望就是:

$$\begin{aligned}
\mathbb{E}\xi &\triangleq \int_\Omega \xi(\omega) P(\mathrm{d}\omega) \\
&= \int_{-\infty}^{+\infty} x\mathrm{d}F(x) \qquad \text{(Lebesgue-Stieltjes 积分)} \\
&= \int_{-\infty}^{+\infty} x \cdot P \circ \xi^{-1}(\mathrm{d}x).
\end{aligned}$$

若 $\xi = I_A \cdot A \in \mathscr{F}$, 则

$$\mathbb{E}\xi \triangleq \int_\Omega \xi(\omega) P(\mathrm{d}\omega) = P(A).$$

(1) 若 $F(x)$ 是阶梯状函数, 则

$$\int_{-\infty}^{+\infty} x\mathrm{d}F(x) = \sum_i x_i(F(x_i) - F(x_{i-1})) = \sum_i x_i p_i.$$

其中 x_i 是 f 的跳跃点.

(2) 若 $F(x)$ 有导数 $p(x)$, 则

$$\int_{-\infty}^{+\infty} x\mathrm{d}F(x) = \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x$$

其中 $p(x)$ 是关于 Lebesgue 测度 F 的导数.

(3) **线性性**: 利用 Lebesgue-Stieltjes 积分的线性性可得

$$\mathbb{E}(a_1f_1(\xi)+a_2f_2(\xi))=a_1\mathbb{E}f_1(\xi)+a_2\mathbb{E}f_2(\xi).$$

(4) **积分区间的可加性**:

$$\int_a^b g(x)\mathrm{d}F(x)=\int_a^c g(x)\mathrm{d}F(x)+\int_c^b g(x)\mathrm{d}F(x).$$

(5) 若 $g(x)\geq 0, F(x)$ 非减, $b\geq a$, 则 $\displaystyle\int_a^b g(x)\mathrm{d}F(x)\geq 0$.

注: $\mathrm{d}F(x)$ 与 $F(\mathrm{d}x)$ 表示的都一样, F 是测度. 例如 $\xi\sim N(0,1)$, 那么方框内的部分看作测度:

$$\mathbb{E}f(\xi)=\int_{\mathbb{R}}f(x)\boxed{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x}$$

定理 3.1.4

设 ξ 是随机变量, g 是 Borel 可测函数, 令 $\eta=g(\xi)$, 则

$$\int_{-\infty}^{+\infty}y\mathrm{d}F_\eta(y)=\int_{-\infty}^{+\infty}g(x)\mathrm{d}F_\xi(x).$$

证明: 要用到实变函数. 标准方法是对示性函数验证, 然后简单函数逼近. □

注: 事实上,

$$\int_{-\infty}^{+\infty}g(x)\mathrm{d}F_\xi(x)=\mathbb{E}g(\xi)=\int_{-\infty}^{+\infty}g(x)\mathrm{d}F_\xi(x).$$

下面把数学期望推广到多维:

定理 3.1.5

设 $(\xi_1,\cdots,\xi_n)$ 的联合分布函数为 $F(x_1,x_2,\cdots,x_n)$, $g:\mathbb{R}^n\to\mathbb{R}^m$ 是 Borel 可测的, 且

$$\int_{\mathbb{R}^n}|g(x_1,\cdots,x_n)|\mathrm{d}F(x_1,\cdots,x_n)<\infty,$$

则

$$\boxed{\mathbb{E}g(\xi_1,\cdots,\xi_n)=\int_{\mathbb{R}^n}g(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)}.$$

注: $\mathbb{E}(\xi_1,\cdots,\xi_n)\triangleq(\mathbb{E}\xi_1,\cdots,\mathbb{E}\xi_n)$. 方框部分的 F 不是函数, 而是由函数定义出来的测度.

基本性质:

(1) 边际分布的期望: 让 $g(x)=x_1$, 其中 $x=(x_1,\cdots,x_n)\in\mathbb{R}^n$, 那么

$$\begin{aligned}\mathbb{E}\xi_1&=\int_{\mathbb{R}^n}x_1\mathrm{d}F(x_1,\cdots,x_n)\\&=\int_{-\infty}^{+\infty}x_1\underbrace{\left(\int_{\mathbb{R}^{n-1}}\mathrm{d}F(x_1,\cdots,x_n)\right)}_{\xi_1\text{边际分布}}\\&=\int_{-\infty}^{+\infty}\mathrm{d}F_{\xi_1}(x).\end{aligned}$$

(2) 若 $n=2, g(x,y)=xy$ 且 ξ_1,ξ_2 独立, 则 $\mathrm{d}F(x,y)=\mathrm{d}F_X(x)\mathrm{d}F_Y(y)$, 从而

$$\mathbb{E}(\xi_1\xi_2)=\mathbb{E}\xi_1\mathbb{E}\xi_2.$$

(乘积的期望等于期望的乘积)

(3) 数学期望有**线性性**: 若 $g(x,y)=ax+by$, 则

$$\begin{aligned}\mathbb{E}(aX+bY)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(ax+by)\mathrm{d}F(x,y)\\&=a\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\mathrm{d}F(x,y)+b\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y\mathrm{d}F(x,y)\\&=a\mathbb{E}X+b\mathbb{E}Y.\end{aligned}$$

(4) 若一维随机变量 X,Y 满足 $X\le Y$, 则 $\mathbb{E}X\le\mathbb{E}Y$.

§ 3.2 方差

定义 3.2.1

设 $(\Omega,\mathscr{F},P)$ 是概率空间, ξ 是随机变量. 称 $\mathbb{E}(\xi-\mathbb{E}\xi)^2$ 是 ξ 的**方差 (variance)**, 记为 $\mathbb{D}\xi$ 或 $\mathrm{Var}(\xi)$. 把 $\sqrt{\mathbb{D}\xi}$ 叫 ξ 的**标准差**.

注: 标准差的定义是为了保证量纲与 ξ 一致. 方差与标准差都是刻画了随机变量与均值之间的偏离程度.

注: 一个重要的公式:

$$\begin{aligned}\mathbb{E}(\xi-\mathbb{E}\xi)^2&=\mathbb{E}(\xi^2-2\xi\mathbb{E}\xi+(\mathbb{E}\xi)^2)\\&=\mathbb{E}(\xi^2)-2\mathbb{E}\xi\mathbb{E}\xi+(\mathbb{E}\xi)^2\\&=\mathbb{E}(\xi^2)-(\mathbb{E}\xi)^2.\end{aligned}$$

写成积分:

$$\mathbb{E}(\xi-\mathbb{E}\xi)^2=\int_{-\infty}^{\infty}\left(x-\int_{-\infty}^{\infty}x\mathrm{d}F(x)\right)^2\mathrm{d}F(x).$$

方差只与 F 有关, 与 ξ 无关. (分布决定方差, 与随机变量无关).

命题 3.2.1

若 $\xi\equiv C$, 则 $\mathbb{D}\xi=0$.

命题 3.2.2

$\mathbb{D}(\xi+c)=\mathbb{D}(\xi)$, 其中 c 是任意常数.

注: 方差与随机变量无关.

命题 3.2.3

$\mathbb{D}(c\xi)=c^2\mathbb{D}(\xi)$, 其中 c 是任意常数.

注: 方差无线性性, 标准差也没有线性性: $\sqrt{\mathbb{D}(c\xi)}=|c|\sqrt{\mathbb{D}\xi}$.

命题 3.2.4

$\mathbb{D}(\xi) = \inf\limits_{c\in\mathbb{R}} \mathbb{E}(\xi - c)^2$.

证明: 【方法一】利用 $\mathbb{E}(\xi + c) = \mathbb{E}\xi + c$ 与 $\mathbb{D}(\xi + c) = \mathbb{D}(\xi)$, 可得

$$\mathbb{D}\xi = \mathbb{D}(\xi - c) = \mathbb{E}(\xi - c)^2 - (\mathbb{E}(\xi - c))^2 \le \mathbb{E}(\xi - c)^2.$$

等号成立条件是 $\mathbb{E}\xi = c$.

【方法二】利用

$$\begin{aligned}\mathbb{E}(\xi - c)^2 &= \mathbb{E}(\xi - \mathbb{E}\xi + \mathbb{E}\xi - c)^2 \\ &= \mathbb{D}\xi + 2\mathbb{E}[(\xi - \mathbb{E}\xi)(\mathbb{E}\xi - c)] + (\mathbb{E}\xi - c)^2 \\ &= \mathbb{D}\xi + (\mathbb{E}\xi - c)^2.\end{aligned}$$

命题 3.2.5

若 ξ, η **独立**, 则 $\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta$.

证明: 用定义验证, 独立性用在了“乘积的期望等于期望的乘积”一处. □

§ 3.3 重要不等式

定理 3.3.1. Chebyshev

设 $\mathbb{E}\xi^2 < +\infty$, 则 $\forall \varepsilon > 0, P(|\xi - \mathbb{E}\xi| \ge \varepsilon) \le \dfrac{\mathbb{D}\xi}{\varepsilon^2}$.

证明: 注意到 $\dfrac{(\xi - \mathbb{E}\xi)^2}{\varepsilon^2} \ge 1$ 以及 $I_A \le 1$, 则

$$\begin{aligned}P(|\xi - \mathbb{E}\xi| \ge \varepsilon) &= \mathbb{E}I_{[|\xi - \mathbb{E}\xi| \ge \varepsilon]} \quad \text{(看作均匀分布)} \\ &\le \mathbb{E}\left(\frac{(\xi - \mathbb{E}\xi)^2}{\varepsilon^2} I_{[|\xi - \mathbb{E}\xi| \ge \varepsilon]}\right) \\ &\le \mathbb{E}\left(\frac{(\xi - \mathbb{E}\xi)^2}{\varepsilon^2}\right) = \frac{\mathbb{D}\xi}{\varepsilon^2}. \quad \square\end{aligned}$$

定理 3.3.2. 推广的 Chebyshev 不等式

$\forall \varepsilon > 0, P(|\xi - \mathbb{E}\xi| \ge \varepsilon) \le \dfrac{\mathbb{E}|\xi - \mathbb{E}\xi|^p}{\varepsilon^p}$.

证明: 思路是同上的, 只需注意到 $\dfrac{(\xi - \mathbb{E}\xi)^p}{\varepsilon^p} \ge 1$ 即可. □

注: 此不等式又叫 Markov 不等式.

例 3.3.3

设 ξ 是 r.v., 若 $\mathbb{D}\xi = 0$, 则 $P(\xi = c) = 1$. （即 ξ 几乎处处是同一常数）

证明: 主要思路是利用 $\{x:x>0\}=\bigcup_{n=1}^{\infty}\{x:x\geq\frac{1}{n}\}$. 记 $c=\mathbb{E}\xi$, 则

$$\begin{aligned}P(\xi\neq c)&=P(|\xi-c|>0)\\&=P\left(\bigcup_{n=1}^{\infty}\left[|\xi-c|\geq\frac{1}{n}\right]\right)\\&\leq\sum_{n=1}^{\infty}P\left(|\xi-c|\geq\frac{1}{n}\right)\quad(\text{次 }\sigma\text{ 可加性})\\&\leq\sum_{n=1}^{\infty}n^2\mathbb{D}\xi=0.\quad(\text{前面定理})\end{aligned}$$

注: 可以推广为 $P(\xi=c)=1\Leftrightarrow\mathbb{E}|\xi-c|^p=0,(p>0)$. □

定理 3.3.4. Cauchy-Schwarz 不等式

设 $r.v.\xi,\eta$ 满足 $\mathbb{E}\xi^2<+\infty,\mathbb{E}\eta^2<+\infty$. 则

$$\mathbb{E}|\xi\eta|\leq(\mathbb{E}\xi^2)^{1/2}(\mathbb{E}\eta^2)^{1/2}.$$

证明: 考虑二次函数 $f(t)=\mathbb{E}(t|\xi|+|\eta|)^2\geq0$, 则它的判别式 $\Delta=(2\mathbb{E}|\xi\eta|)^2-4\mathbb{E}\xi^2\mathbb{E}\eta^2\leq0$. □

定理 3.3.5. 正态分布估计

若 X 是标准正态分布 r.v., 则 $P(|X|\geq t)\leq\sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t},t>0$.

证明: 利用正态分布的对称性, 只需证其中一边. 利用 $X\geq t$, 有

$$P(X\geq t)=\frac{1}{\sqrt{2\pi}}\int_t^{\infty}e^{-x^2/2}dx\leq\frac{1}{\sqrt{2\pi}}\int_t^{\infty}\frac{x}{t}e^{-x^2/2}dx=\frac{1}{\sqrt{2\pi}}\frac{e^{-t^2/2}}{t}.$$

这样, $P(|X|\geq t)=2P(X\geq t)\leq\sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t}$. □

定理 3.3.6. 多项式逼近

设 $f\in C^0[0,1]$, 令

$$f_n(x)=\sum_{m=0}^{n}C_n^m x^m(1-x)^{n-m}f\left(\frac{m}{n}\right),$$

则 $\lim\limits_{n\to\infty}\sup\limits_{x\in[0,1]}|f_n(x)-f(x)|=0$.

证明: $\forall x\in[0,1]$, 设 $\{x_i\}$ 为 i.i.d.r.v., 满足 $P(\xi_1=1)=x,P(\xi_1=0)=1-x$. (两点分布) 则

$$\mathbb{E}f\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\sum_{m=0}^{n}P\left(\sum_{i=1}^{n}\xi_i=m\right)f\left(\frac{m}{n}\right)=f_n(x).$$

由于 f 是一致连续的, 则 $\forall\varepsilon>0,\exists\delta>0,\forall x,y\in[0,1]$, 如果 $|x-y|\leq\delta$, 则有 $|f(x)-f(y)|\leq\varepsilon$. 从而

(记 $\|f\|_\infty \triangleq \sup f$)

$$
\begin{aligned}
|f_n(x)-f(x)| =& \left|\mathbb{E}f\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)-f(x)\right| \\
\leq& \left|\mathbb{E}\left[f\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)-f(x)\right] I_{\left[\left|\frac{1}{n}\sum\limits_{i=1}^n \xi_i - x\right|\geq\delta\right]}\right| \\
&+\left|\mathbb{E}\left[f\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)-f(x)\right] I_{\left[\left|\frac{1}{n}\sum\limits_{i=1}^n \xi_i - x\right|<\delta\right]}\right| \\
\leq& 2\|f\|_\infty \mathbb{E} I_{\left[\left|\frac{1}{n}\sum\limits_{i=1}^n \xi_i - x\right|\geq\delta\right]} + \varepsilon \quad (\text{对前一行分别作放缩}) \\
=& 2\|f\|_\infty P\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i - x\right|\geq\delta\right)+\varepsilon \\
\leq& 2\|f\|_\infty \frac{1}{n\delta^2} n\mathbb{D}\xi_1 + \varepsilon \quad (\text{用 Chebyshev 不等式}) \\
=& 2\|f\|_\infty \frac{1}{n\delta^2} x(1-x)+\varepsilon \\
\leq& 2\|f\|_\infty \frac{1}{n\delta^2}\cdot\frac{1}{4}+\varepsilon.
\end{aligned}
$$

这样 $\sup\limits_{x\in[0,1]}|f_n(x)-f(x)| \leq \dfrac{\|f\|_\infty}{2n\delta^2}+\varepsilon$, 两边取 $n\to\infty$ 以及利用 ε 的任意性即可证完. □

定理 3.3.7. Young 不等式

设 $1<p<\infty, \dfrac{1}{p}+\dfrac{1}{q}=1$. 则 $ab \leq \dfrac{a^p}{p}+\dfrac{b^q}{q}, \forall a,b>0$.

定理 3.3.8. Hölder 不等式

设 $1<p<+\infty$. 且 $\dfrac{1}{p}+\dfrac{1}{q}=1$. 若 $\mathbb{E}|\xi|^p<+\infty$ 且 $\mathbb{E}|\eta|^q<+\infty$, 则

$$
\mathbb{E}|\xi\eta| \leq (\mathbb{E}|\xi|^p)^{1/p}(\mathbb{E}|\eta|^q)^{1/q}.
$$

证明: 不妨设 $\mathbb{E}|\xi|^p=\mathbb{E}|\eta|^q=1$(否则可以乘一个倍数). 由 Young 不等式,

$$
\mathbb{E}|\xi\eta| \leq \mathbb{E}\left(\frac{1}{p}|\xi|^p+\frac{1}{q}|\eta|^q\right)=\frac{1}{p}\mathbb{E}|\xi|^p+\frac{1}{q}\mathbb{E}|\eta|^q=1=(\mathbb{E}|\xi|^p)^{1/p}(\mathbb{E}|\eta|^q)^{1/q}.
$$

□

§ 3.4 协方差、相关系数、矩

定义 3.4.1

设 (X,Y) 是二维随机变量, 若 $\mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y))$ 存在, 则称它为 (X,Y) 的**协方差 (covariance)**, 记为 $\mathrm{cov}(X,Y)$.

若 $\mathrm{cov}(X,Y)>0$, 称 X,Y **正相关**.

若 $\mathrm{cov}(X,Y)<0$, 称 X,Y **负相关**.

若 $\mathrm{cov}(X,Y)=0$, 称 X,Y **不相关**.

命题 3.4.1

$\mathrm{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}X\mathbb{E}Y$.

证明: $\mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y))=\mathbb{E}(XY-X\mathbb{E}Y-Y\mathbb{E}X+\mathbb{E}X\mathbb{E}Y)=\mathbb{E}(XY)-\mathbb{E}X\mathbb{E}Y$. □

注: 若 X 与 Y 不相关, 则 $\mathbb{E}(XY)=\mathbb{E}X\mathbb{E}Y$, 不代表 X,Y 独立! 独立 $\Rightarrow$ 不相关, 但不相关不能推出独立.

例 3.4.2

设 $X\sim N(0,1), Y=X^2$, 则

$$P(X\in(0,1),Y\in(2,3))=0\neq P(X\in(0,1))P(Y\in(2,3)).$$

故 X,Y 不独立. 但

$$\mathrm{cov}(X,Y)=\mathbb{E}X^3-\mathbb{E}X\mathbb{E}(X^2)=0.$$

这是因为 X^3 与 X 都是奇函数. 故 X,Y 不相关.

命题 3.4.3

$\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$.

命题 3.4.4

$\mathbb{D}(X\pm Y)=\mathbb{D}X\pm 2\mathrm{cov}(X,Y)+\mathbb{D}(Y)$.

命题 3.4.5

$\mathrm{cov}(aX+b,Y)=a\mathrm{cov}(X,Y)$. $\mathrm{cov}(X_1+X_2,Y)=\mathrm{cov}(X_1,Y)+\mathrm{cov}(X_2,Y)$.

注: 协方差有双线性函数的味道.

例 3.4.6

设 (ξ,η) 的联合密度为 $p(x,y)=\begin{cases}3x, & 0\le y<x<1,\\ 0, & \text{其他}.\end{cases}$, 求 $\mathrm{cov}(\xi,\eta)$.

解: 由于

$$\begin{aligned}\mathbb{E}\xi&=\int_0^1\int_0^x x\cdot(3x)\mathrm{d}y\mathrm{d}x=\frac{3}{4},\\ \mathbb{E}\eta&=\int_0^1\int_0^x y\cdot(3x)\mathrm{d}x\mathrm{d}y=\frac{3}{8},\\ \mathbb{E}(\xi\eta)&=\int_0^1\int_0^x xy\cdot(3x)\mathrm{d}y\mathrm{d}x=\frac{3}{10},\end{aligned}$$

所以 $\mathrm{cov}(\xi,\eta)=\dfrac{3}{160}$. 根据计算结果, ξ,η 正相关, 当然不独立. □

注: 牢记

$$\mathbb{E}f(x_1,x_2)=\int_{\mathbb{R}^2}f(x_1,x_2)p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2.$$

定义 3.4.2. 相关系数

设 (ξ,η) 是二维随机变量, 满足 $\mathbb{D}\xi>0,\mathbb{D}\eta>0$, 则称

$$\rho(\xi,\eta)=\frac{\mathrm{cov}(\xi,\eta)}{\sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta}}$$

为 (ξ,η) 的**相关系数**.

注: 相关系数的引入是为了让协方差与 ξ,η 量纲保持一致, 便于比较!

标准化的协方差:

$$\rho(\xi,\eta)=\mathrm{cov}\left(\frac{\xi-\mathbb{E}\xi}{\sqrt{\mathbb{D}\xi}},\frac{\eta-\mathbb{E}\eta}{\sqrt{\mathbb{D}\eta}}\right).$$

其中, $\dfrac{\xi-\mathbb{E}\xi}{\sqrt{\mathbb{D}\xi}}$ 是期望为 0, 方差为 1 的随机变量.

例 3.4.7

二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的相关系数是 ρ.

注: 二维正态分布中, 独立 $\Leftrightarrow$ 不相关 $\Leftrightarrow \rho=0$.

命题 3.4.8

相关系数不超过 1.

证明: 注意

$$|\mathrm{cov}(\xi,\eta)|=|\mathbb{E}(\xi-\mathbb{E}\xi)(\eta-\mathbb{E}\eta)|\le\mathbb{E}|(\xi-\mathbb{E}\xi)(\eta-\mathbb{E}\eta)|\le\sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta},$$

最后一个不等号用了 Cauchy-Schwarz 不等式. □

命题 3.4.9

$\rho(\xi,\eta)=\pm1\Leftrightarrow\exists a\ne0,b\in\mathbb{R},P(a\xi+b=\eta)=1$.（即几乎必然有 $a\xi+b=\eta$.）

特别地, 若 $\rho=1$ 则 $a>0$; 若 $\rho=-1$ 则 $a<0$.

注: 如果 $\mathbb{E}|\xi|^p=0$, 则 $P(\xi=0)=1$, 用 Chebyshev 不等式可以验证如下:

$$\begin{aligned}P(|\xi|>0)&=P\left(\bigcup_{n=1}^{\infty}\left[|\xi|\ge\frac{1}{n}\right]\right)\\&\le\sum_{n=1}^{\infty}P\left(|\xi|\ge\frac{1}{n}\right)\le\sum_{n=1}^{\infty}n^p\mathbb{E}|\xi|^p=0.\end{aligned}$$

例 3.4.10

随机变量不相关但也可能会有非线性的关系.

答: 考虑 $\theta \sim (0,2\pi), \xi = \cos\theta, \eta = \cos(\theta + a), a$ 是常数. 则

$$\mathbb{E}\xi = 0, \mathbb{E}\eta = 0, \mathbb{E}\xi^2 = \frac{1}{2}, \mathbb{E}\eta^2 = \frac{1}{2}, \mathbb{E}\xi\eta = \frac{1}{2}\cos a.$$

所以相关系数为

$$\rho = \frac{\mathbb{E}\xi\eta}{\sqrt{\mathbb{E}\xi^2 - (\mathbb{E}\xi)^2}\sqrt{\mathbb{E}\eta^2 - (\mathbb{E}\eta)^2}} = \cos a.$$

(1) 当 $a = \frac{\pi}{2}$ 或 $\frac{3\pi}{2}$ 时, $\rho = 0$, 此时 ξ 与 η 不相关, 但此时 $\xi^2 + \eta^2 = 1$, ξ, η 不独立, 且有非线性的关系.

(2) 当 $a = 0$ 时, $\rho = 1, \xi = \eta$; 当 $a = \pi$ 时,$\rho = -1, \xi = -\eta$, 此时有线性关系. □

例 3.4.11

袋中有 N 张卡片, 各记以标号 $x_1, x_2, \cdots, x_N$, 不放回从中抽 n 张, 求其和的期望与方差.

解: 当 $n = 1$ 时, ξ 表示抽中卡片的数字, 则 ξ 的分布列为

$$\left\{x_i : \frac{1}{N}\right\}_{1 \le i \le N}.$$

所以

$$\mathbb{E}\xi = \frac{x_1 + \cdots + x_N}{N} = \bar{x}, \mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \frac{1}{N}\sum_{i=1}^{N}(\xi - \bar{x})^2 = \sigma^2.$$

当 $n > 1$ 时, 用 ξ_i 表示第 i 次抽中卡片上的数字, 则

$$\begin{aligned}
\mathbb{E}\sum_{i=1}^{n}\xi_i &= \sum_{i=1}^{n}\mathbb{E}\xi_i = \sum_{i=1}^{n}\bar{x} = n\bar{x}, \\
\mathbb{D}\sum_{i=1}^{n}\xi_i &= \mathbb{E}\left(\sum_{i=1}^{n}\xi_i\right)^2 - \left(\mathbb{E}\sum_{i=1}^{n}\xi_i\right)^2 \\
&= \sum_{i=1}^{n}\mathbb{D}(\xi_i) + \sum_{i \ne j}\mathrm{cov}(\xi_i, \xi_j) \\
&= n\sigma^2 - \frac{n(n-1)}{N-1}\sigma^2 = n\frac{(N-n)}{N-1}\sigma^2.
\end{aligned}$$

注意

$$\mathrm{cov}(\xi_1, \xi_2) = -\frac{\sigma^2}{N-1}.$$

当 $n = N$ 时,

$$\mathbb{D}\sum_{i=1}^{n}\xi_i = 0 = N\sigma^2 + N(N-1)\mathrm{cov}(\xi_1, \xi_2),$$

定义 3.4.3. 矩

对任意非零正整数 k, 把 $\mathbb{E}\xi^k$ 称为 ξ 的 k 阶**原点矩**, 把 $\mathbb{E}(\xi - \mathbb{E}\xi)^k$ 称为 ξ 的 k 阶**中心矩**.

注: 用二项式展开可以把 k 阶原点矩与中心矩进行相互转换.

如果高阶矩存在, 那么低阶矩也存在, 即 $p > q > 0$ 时,

$$\mathbb{E}|\xi|^p < +\infty \Rightarrow \mathbb{E}|\xi|^q < +\infty.$$

例 3.4.12

若 $\xi \sim N(0,\sigma^2)$, 则

$$\mathbb{E}\xi^k = \begin{cases} 0, & k\text{为奇数}, \\ \sigma^k(k-1)!!, & k\text{为偶数}. \end{cases}$$

证明: k 为奇数时显然, 只看 k 为偶数: 事实上

$$\mathbb{E}\xi^k = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} x^k e^{-\frac{x^2}{2\sigma^2}}\mathrm{d}x \overset{\text{换元}}{=\!=} \cdots = \sqrt{\frac{2}{\pi}}\sigma^k 2^{\frac{k-1}{2}}\Gamma\left(\frac{k+1}{2}\right).$$

注: 回顾 $\Gamma(x+1)=x\Gamma(x)$, 而且 $\Gamma(1/2)=\sqrt{\pi}$ 用余元公式 $\Gamma(x)\Gamma(1-x)=\dfrac{\pi}{\sin\pi x}$ 可证明.

§ 3.5 特征函数

定义 3.5.1. 特征函数

设 r.v.ξ 的分布是 $F(x)$, 称

$$f(t) \triangleq \mathbb{E}e^{it\xi} = \int_{-\infty}^{\infty} e^{itx}\mathrm{d}F(x)$$

为 ξ 的**特征函数 (characteristic function)**.

注: 特征函数只与分布相关, 因此也称为某个分布函数的特征函数.

注: 容易观察有如下性质: $f(0)=1, |f(t)|\le 1, f(-t)=\overline{f(t)}$.

引理 3.5.1. 控制收敛定理

设 $(\Omega,\mathscr{F})$ 是可测空间, μ 是它上的测度, $f_n:\Omega\to\mathbb{R}$. 若存在可测函数 $f\ge 0$ 使得 $|f_n|\le f, \forall n\ge 1$ (即 f 控制 f_n), 同时

$$\int_\Omega |f(x)|\mu(\mathrm{d}x) < +\infty, f_n \xrightarrow{a.s.} f,$$

则

$$\lim_{n\to\infty}\int_\Omega f_n(x)\mu(\mathrm{d}x) = \int_\Omega \lim_{n\to\infty} f_n(x)\mu(\mathrm{d}x).$$

定理 3.5.2. 特征函数的一致连续性

特征函数 f 有如下的增量不等式:

$$|f(t)-f(t+u)|^2 \le 2(1-Re[f(u)]), \forall t,u\in\mathbb{R}$$

由于 $Re[f(u)]$ 是连续的, 从而 $f(t)$ 一致连续.

证明: 作如下放缩:

$$
\begin{aligned}
|f(t)-f(t+u)|^2 &= \left|\int_{-\infty}^{\infty}(e^{itx}-e^{i(t+u)x})\mathrm{d}F(x)\right|^2 \\
&= \left|\int_{-\infty}^{\infty}e^{itx}(1-e^{iux})\mathrm{d}F(x)\right|^2 \\
&\le \int_{-\infty}^{\infty}e^{itx}|^2 dF(x)\int_{-\infty}^{\infty}|1-e^{iux}|^2\mathrm{d}F(x)\ \text{(Cauchy-Schwarz)} \\
&= \int_{-\infty}^{\infty}[(1-\cos ux)^2+\sin^2 ux]\mathrm{d}F(x) \\
&= 2-2\int_{-\infty}^{\infty}\cos ux\mathrm{d}F(x) = 2(1-Re[f(u)]).
\end{aligned}
$$

只需要看 $Re[f(u)]$ 有没有被控制住. 由于

$$
\lim_{\delta u\to 0} Re[f(u+\delta u)] - Re[f(u)] = \lim_{\delta u\to 0}\int_{-\infty}^{+\infty}\cos(u+\delta u)x-\cos ux\mathrm{d}F(x),
$$

由 $|\cos(u+\Delta u)x-\cos ux|\le 2$ 以及控制收敛定理可知上述积分与极限可交换顺序, 再由 $\cos x$ 的连续性可知 $Re[f(u)]$ 是连续的, 从而 $f(t)$ 一致连续. □

定理 3.5.3. 非负定性

$\forall n\ge 2, t_1\cdots,t_n\in\mathbb{R},\alpha_1,\cdots,\alpha_n\in\mathbb{C}$, 有

$$
\sum_{k=1}^{n}\sum_{j=1}^{n} f(t_k-t_j)\alpha_k\overline{\alpha_j}\ge 0.
$$

证明: 利用 $\alpha\overline{\alpha}=|\alpha|^2$ 即可. □

定理 3.5.4

若 ξ,η 独立, 则 $\mathbb{E}e^{i(\xi+\eta)t}=\mathbb{E}e^{it\xi}\mathbb{E}e^{it\eta}$.

证明: 注意到对于所有的 $\mathbb{R}^n\to\mathbb{R}$ 的 Borel 可测函数 g,h, 都有 $g(\xi),h(\eta)$ 独立. □

注: 利用该性质可以推出二项分布、Poisson 分布、正态分布、Gamma 分布的再生性. 这里从略.

定理 3.5.5

若 $\exists n\in\mathbb{N},\mathbb{E}|\xi|^n<+\infty$, 则 $\mathbb{E}e^{it\xi}$ 关于自变量 t 是 n 阶可导的, 且 $\forall k\le n, f^{(k)}(0)=i^k\mathbb{E}\xi^k$.

证明: 利用导数的定义以及控制收敛定理即可. □

注: 若 $\mathbb{E}X^2<+\infty, f(x)$ 是特征函数, 则 $\mathbb{E}X=-if'(0),\mathrm{D}X=\mathbb{E}X^2-(\mathbb{E}X)^2=-f''(0)+(f'(0))^2$.

定理 3.5.6

设 $\eta=a\xi+b,a,b\in\mathbb{R}$, 则 $f_\eta(t)=e^{ibt}f_\xi(at)$.

定理 3.5.7. 逆转公式

设分布函数 $F(x)$ 的特征函数是 $f(t), x_1, x_2$ 是 F 的连续点, 则

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t)\mathrm{d}t.$$

进一步, 若 f 在 $(-\infty, +\infty)$ 上是 Lebesgue 可积的, 即 $\displaystyle\int_{-\infty}^{\infty} |f(t)|\mathrm{d}t < +\infty$, 则函数

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t)\mathrm{d}t$$

是 F 的密度函数.

证明: 套定义验证即可, 其中要用到控制收敛定理.

注: 定理表明分布函数与特征函数可以相互转化.

§ 3.6 习题

1. **(Mills's ratio)** 设 $\phi(x)$ 表示标准正态分布的密度函数, 证明 $\phi'(x)+x\phi(x)=0$, 并证明

$$\frac{1}{x}-\frac{1}{x^3}<\frac{1-\Phi(x)}{\phi(x)}<\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}, x>0.$$

这个估计式我们非常感兴趣, 因为 $\Phi(x)$ 没有封闭形式的表达式.

2. 在长为 a 的线段上任取两点 X 和 Y, 求此两点之间的平均长度. (即求距离的期望值)
3. 设 X_1, X_2 是独立同分布的随机变量, $X_1 \sim E(\lambda_1)$. 求 $Y=\max\{X_1, X_2\}$ 的数学期望.
4. **(2019 期末)**X, Y 是独立的随机变量, $\mathbb{E}X=0, \mathbb{E}|Y|<+\infty, \mathbb{E}(|X+Y|)<+\infty$. 证明:

$$\mathbb{E}(|Y|) \leq \mathbb{E}(|X+Y|).$$

5. **(2022 某校推免)** 假定随机变量 X 服从参数为 1 的泊松分布, 现在对其独立观测 n 次, 设 Y_n 为 X 大于 1 的次数, 求 Y_n^2 的期望.
6. 设随机变量 X, Y 的期望分别为 $-2, 2$, 方差分别为 $1, 4$, 且 $\mathbb{E}(X+2)(Y-2)=-1$. 请用所学知识给出 $P(|X+Y| \geq 6)$ 的一个非平凡上界.
7. 整理并列出常见离散型随机变量与连续型随机变量分布的特征函数.
8. **(2022 某校推免)** 设 ξ 为取自然数值的随机变量, φ 为其特征函数, 证明:

$$P(\xi=k)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\varphi(t)\mathrm{d}t, \qquad k=0,1,2,\cdots.$$

9. 如果 X 是取非负整数值的 r.v., 则

$$\mathbb{E}X=\sum_{n=0}^{\infty} P(X \geq n).$$

10. **(2019 期末)** 假定 X 是非负 r.v., $p \geq 1$ 为常数, 则

$$\mathbb{E}X^p = p\int_0^{\infty} y^{p-1}P(X>y)dy.$$

11. 设 $X \geq 0, Y \geq 0$ 是随机变量, $p>1$. 证明: $\mathbb{E}((X+Y)^p) \leq 2^{p-1}(\mathbb{E}(X^p)+\mathbb{E}(Y^p))$.
参考: `https://math.stackexchange.com/questions/1532907/`
12. **(123 Theorem)** 设 X, Y 是 i.i.d.r.v., 证明 $P[|X-Y| \leq 2] < 3P[|X-Y| \leq 1]$.
13. **(2016 丘赛 Team)** 在单位圆 $x^2+y^2=1$ 均匀随机取 n 个点 $(n \geq 2)$, 这 n 个点可以把单位圆分成 n 段圆弧. 求包含点 $(1,0)$ 的圆弧的长度的数学期望.

第 4 章　随机变量的收敛性

§ 4.1　几种收敛性的定义

记 $(\Omega, \mathscr{F}, P)$ 是概率空间, $\{X_n\}_{n\geq 1}$ 是 r.v. 列.

定义 4.1.1. 几乎必然收敛

如果 $\exists A \in \mathscr{F}$ 满足 $P(A) = 0$（即 A 是个零测集）, 使得 $\forall \omega \in A^c, \xi_n(\omega) \to \xi(\omega)$, 则称 ξ_n 几乎必然收敛到 ξ, 或者记为 $\xi_n \xrightarrow{a.s.} \xi(n \to \infty)$.

定义 4.1.2. 依概率收敛

如果 $\forall \varepsilon > 0$ 都有 $\lim\limits_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0$, 则称 ξ_n 依概率收敛到 ξ, 或者记为 $\xi_n \xrightarrow{P} \xi(n \to \infty)$.

定义 4.1.3. 依分布收敛

$\forall f \in C^0(\mathbb{R}), |f| < +\infty$, 如果有 $Ef(\xi_n) \longrightarrow Ef(\xi)$, 则称 $\{\xi_n\}$ 依分布收敛到 ξ, 记为 $\xi_n \xrightarrow{d} \xi$.

注: 这里的 $\{\xi_n\}$ 可能不是定义在同一个概率空间上, 但是依概率收敛中的 r.v. 必定定义在同一个概率空间上.

$L^p(\Omega, \mathscr{F}, P) \triangleq \{\xi : \mathscr{F}\text{可测且}E|\xi|^p < +\infty\}$.

定义 4.1.4. 概率空间上的范数

设 $p \in (0, +\infty)$, 令 $L^p(\Omega, \mathscr{F}, P) \triangleq \{\xi : E|\xi|^p < +\infty\}$, 这里 ξ 是 r.v., 定义 $\|\xi\|_p \triangleq (E|\xi|^p)^{1/p}$, 则 $\|\cdot\|_p$ 是 $L^p(\Omega, \mathscr{F}, P)$ 上的范数.

这里的范数满足一般范数的三条性质 (非负性、齐次性、三角不等式), 且 $(L^p(\Omega, \mathscr{F}, P), \|\cdot\|_p)$ 是 Banach space.

定义 4.1.5. p 阶收敛 (L^p 收敛)

设 $\{\xi_n; \xi\} \subset L^p(\Omega, \mathscr{F}, P)$, 若 $\lim\limits_{n\to\infty} \|\xi_n - \xi\|_p = 0$, 则称 $\{\xi_n\}$ 为 p 阶收敛（或称 L^p 收敛）于 ξ, 记为 $\xi_n \xrightarrow{L^p} \xi$.

定义 4.1.6. 上极限集与下极限集

(1) 把

$$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{\omega : \omega\text{在无穷多个}A_n\text{中}\} = \{\omega : \forall j \in \mathbb{N}, \exists k \geq j(x \in A_k)\}$$

称为 A_n 的上极限集, 记为 $\limsup\limits_{n\to\infty} A_n$.

(2) 把

$$\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = \{\omega : \omega\text{不在至多有限个}A_n\text{中}\} = \{\omega : \exists j_0 \in \mathbb{N}, \forall k \geq j_0(x \in A_k)\}$$

称为 A_n 的下极限集, 记为 $\liminf\limits_{n\to\infty} A_n$.

注: 显然 $\liminf\limits_{n\to\infty} A_n \subset \limsup\limits_{n\to\infty} A_n$.

§ 4.2 几个重要的定理

下面这个定理很 Nice, 所以在这里暂时叫做 Nice 引理.

引理 4.2.1

[Nice] 假定 Y 是非负 r.v., 则

$$\sum_{n=1}^{\infty} P(Y \geq n) \leq EY \leq \sum_{n=1}^{\infty} P(Y \geq n) + 1.$$

证明: 主要思想是根据概率的非负性, 从而对级数进行顺序交换, 以及利用 $P(A) = I_A$ 的思想 (均匀分布).

$$\begin{aligned}
\sum_{n=1}^{\infty} P(Y \geq n) &= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} P(m \leq Y < m+1) \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{m} P(m \leq Y < m+1) \\
&= \sum_{m=1}^{\infty} mP(m \leq Y < m+1) \\
&= \sum_{m=1}^{\infty} mEI_{[m \leq Y < m+1]} \\
&\leq \sum_{m=1}^{\infty} EYI_{[m \leq Y < m+1]} \ (Y \geq m) \\
&\leq \boxed{\text{EY}} \\
&= \sum_{m=0}^{\infty} EYI_{[m \leq Y < m+1]} \\
&\leq \sum_{m=0}^{\infty} (m+1)P(m \leq Y < m+1) \ (Y < m+1) \\
&= \sum_{m=1}^{\infty} mP(m \leq Y < m+1) + \sum_{m=0}^{\infty} P(m \leq Y < m+1) \\
&= \sum_{n=1}^{\infty} P(Y \geq n) + P(Y \geq 0) \ (\text{凑出欲证命题的式子}) \\
&= \sum_{n=1}^{\infty} P(Y \geq n) + 1. \square
\end{aligned}$$

定理 4.2.2. Borel-Cantelli

(1) 我们有

$$\sum_{n=1}^{\infty} P(A_n) < +\infty \Longrightarrow P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = 0.$$

即 A_n 不可能发生无穷多次.

(2) 若 $\{A_n\}$ 相互独立, 则有

$$\sum_{n=1}^{\infty} P(A_n) = +\infty \Longrightarrow P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = 1.$$

即 A_n 必然发生无穷多次.

证明: (1) 只需注意到

$$\begin{aligned} P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) &= \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty} A_n\right) \quad (\text{序列 } \{\bigcup_{n=k}^{\infty} A_n\}_{k\geq 1} \text{ 单调递减}) \\ &\leq \lim_{k\to\infty}\sum_{n=k}^{\infty} P(A_n) \quad (\text{次 } \sigma\text{-可加性}) \\ &\leq 0. \quad (\text{利用条件以及 Cauchy 准则}) \end{aligned}$$

(2) 欲证命题可以作如下转化:

$$\begin{aligned} & P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = 1 \\ \Longleftrightarrow & \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty} A_n\right) = 1 \\ \Longleftrightarrow & P\left(\bigcup_{n=k}^{\infty} A_n\right) = 1, \forall k \geq 1 \quad (\text{单调递减趋于 } 1\text{, 只能为 } 1) \\ \Longleftrightarrow & P\left(\bigcap_{n=k}^{\infty} A_n^c\right) = 0, \forall k \geq 1. \quad (\text{de Morgan}) \end{aligned}$$

事实上,

$$\begin{aligned} 0 \leq P\left(\bigcap_{n=k}^{\infty} A_n^c\right) &= \lim_{m\to\infty} P\left(\bigcap_{n=k}^{m} A_n^c\right) \quad (\text{从下连续性}) \\ &= \lim_{m\to\infty}\prod_{n=k}^{m}(1-P(A_n)) \quad (\text{独立性}) \\ &\leq \lim_{m\to\infty}\prod_{n=k}^{m}(e^{-P(A_n)}) \quad (e^{-x} \geq -x+1) \\ &= \exp\left(-\lim_{m\to\infty}\sum_{n=k}^{m} P(A_n)\right) = 0. \end{aligned}$$

根据夹逼定理可以证明完毕. □

注: 如果把相互独立减弱为两两不相关或者两两负相关, 结论仍成立.

注: Borel-Cantelli 给出了几乎处处收敛的充分条件 (把上述 $A_n \triangleq [|\xi_n - \xi| \geq \varepsilon]$). 另外如果加上相互独立又给出了个几乎处处收敛的必要条件. 常用这个定理来证几乎处处收敛.

下面给出一个换了一种高大上表述的“推论”——Borel 0-1 律.

推论 4.2.3. Borel 0-1 律

若 $\{A_n\}$ 相互独立, 则

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = 0.$$
$$\sum_{n=1}^{\infty} P(A_n) = +\infty \Longrightarrow P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = 1.$$

例 4.2.4

设 $\{X_n\}$ 相互独立, 则对任意 $\varepsilon > 0$, $X_n \xrightarrow{a.s.} 0 \Leftrightarrow \sum\limits_{n=1}^{+\infty} P(|X_n| \geq \varepsilon) < +\infty$.

证明: 取 $A_n = [|X_n| \geq \varepsilon]$, 用 Borel-Cantelli 引理和 a.s. 收敛的等价刻画即可. $\square$

引理 4.2.5. Chung-Erdös 不等式

设 $\{A_k\}_{k\leq n} \subset \mathscr{F}$, 且 $P\Big(\bigcup\limits_{k=1}^{n} A_k\Big) > 0$, 则

$$P\Big(\bigcup_{k=1}^{n} A_k\Big) \geq \frac{\Big(\sum\limits_{k=1}^{n} P(A_k)\Big)^2}{\sum\limits_{i,k=1}^{n} P(A_iA_k)}.$$

证明: 只需注意到

$$\begin{aligned}
\Big(\sum_{k=1}^{n} P(A_k)\Big)^2 &= \Big(\mathbb{E}\sum_{k=1}^{n} I_{A_k}\Big)^2 = \Big(\mathbb{E}\sum_{k=1}^{n} I_{A_k} I_{\bigcup_{k=1}^{n} A_k}\Big)^2 \\
&\leq \mathbb{E}\Big(\sum_{k=1}^{n} I_{A_k}\Big)^2 \mathbb{E}\Big(I_{\bigcup_{k=1}^{n} A_k}\Big)^2 = \mathbb{E}\sum_{i,k=1}^{n} I_{A_i} I_{A_k} P\Big(\bigcup_{k=1}^{n} A_k\Big) \\
&= \left(\sum_{i,k=1}^{n} P(A_iA_k)\right) P\Big(\bigcup_{k=1}^{n} A_k\Big). \qquad \square
\end{aligned}$$

定理 4.2.6. Kochen-Stone

设 $\{A_n\} \subset \mathscr{F}$, $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$, 则

$$P(A_n, i.o.) \geq \limsup_{n\to\infty} \frac{\Big(\sum\limits_{k=1}^{n} P(A_k)\Big)^2}{\sum\limits_{i,k=1}^{n} P(A_iA_k)} = \limsup_{n\to\infty} \frac{\sum\limits_{1\leq i<k\leq n} P(A_i)P(A_k)}{\sum\limits_{1\leq i<k\leq n} P(A_iA_k)}.$$

证明: (1) 设 $a_n = \Big(\sum\limits_{k=1}^{n} P(A_k)\Big)^2, b_n = \sum\limits_{i,k=1}^{n} P(A_iA_k)$, 则 $\lim\limits_{n\to\infty} a_n = \infty$. 由引理 4.2.5, $\lim\limits_{n\to\infty} b_n = \infty$.

设 $m \geq 1$. 由引理 4.2.5, 以及 $\sum\limits_{i,k=m+1}^{n} P(A_iA_k) \leq b_n - b_m$, 可得

$$P\Big(\bigcup_{k=m+1}^{\infty} A_k\Big) = \lim_{n\to\infty} P\Big(\bigcup_{k=m+1}^{\infty} A_k\Big) \geq \limsup_{n\to\infty} \frac{(\sqrt{a_n}-\sqrt{a_m})^2}{b_n - b_m} = \limsup_{n\to\infty} \frac{a_n}{b_n}.$$

由 m 的任意性与测度的从上连续性, 可得

$$P\Big(\bigcap_{m=1}^{\infty}\bigcup_{k=m+1}^{\infty} A_k\Big) = \lim_{m\to\infty} P\Big(\bigcup_{k=m+1}^{\infty} A_k\Big) = \limsup_{n\to\infty} \frac{a_n}{b_n}.$$

不等式证明完毕.

(2) 由 $a_n \to \infty$ 且

$$a_n = 2\sum_{1\leq i<k\leq n} P(A_i)P(A_k) + \sum_{i=1}^{n} [P(A_i)]^2,$$

$$b_n = 2\sum_{1\leq i<k\leq n} P(A_iA_k) + \sum_{i=1}^{n} P(A_i),$$

所以经过一系列放缩可得 $\limsup\limits_{n\to\infty} \frac{a_n}{b_n} = \text{RHS}$. □

定理 4.2.7. Lévy

设分布函数列 $\{F_n\}$ 弱收敛到分布函数 F, 则相应的特征函数 $\{f_n\}$ 点点收敛到 f, 且在 t 的任一有限区间上一致收敛. 反之, 若 $\{f_n\}$ 逐点收敛到一个复值函数 f, 且 f 在 $t = 0$ 处连续, 则 f 为其分布函数 F 的特征函数, 且 $F_n \xrightarrow{W} F$.

定理的证明要用到 Helly 定理, 如果展开来写的话篇幅过长, 在这儿从略.

注: 定理揭示了分布函数列弱收敛与对应特征函数逐点收敛的关系.

§ 4.3 几种收敛性的关系

注: 利用极限的唯一性可以证明依概率收敛与 a.s. 收敛的极限几乎必然唯一.

下面两个定理很重要, 它刻画了几乎必然收敛与依概率收敛.

定理 4.3.1. 几乎必然收敛的刻画

$$\xi_n \xrightarrow{a.s.} \xi \Longleftrightarrow \forall \varepsilon > 0, P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} [|\xi_n - \xi| \geq \varepsilon]\right) = 0.$$

证明: 只需利用 $\{x : x < \varepsilon\} = \bigcap_{n=1}^{\infty} \{x : x \geq \frac{1}{n}\}$, 并注意到

$$
\begin{aligned}
\xi_n \xrightarrow{a.s.} \xi \iff & P\left(\lim_{n\to\infty} \xi_n = \xi\right) = 1 \\
\iff & P(\{\omega \in \Omega : \xi_n(\omega) \to \xi(\omega)\}) = 1 \\
\iff & P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \left[|\xi_i - \xi| < \frac{1}{k}\right]\right) = 1 \\
\iff & P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \left[|\xi_i - \xi| \geq \frac{1}{k}\right]\right) = 0 \\
\iff & P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \left[|\xi_i - \xi| \geq \frac{1}{k}\right]\right) = 0, \forall k \geq 1 (\text{次 } \sigma \text{ 可加性与单调性}) \\
\iff & P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} [|\xi_i - \xi| \geq \varepsilon]\right) = 0, \forall \varepsilon > 0. \square
\end{aligned}
$$

定理 4.3.2. 依概率收敛的刻画

$$\xi_n \xrightarrow{P} \xi \iff \text{对 } \{\xi_n\} \text{ 的任一子列 } \{\xi_n'\}, \text{ 都存在它的子列 } \{\xi_{n_k}'\}_{k\geq 1}, \text{ 使得} \xi_{n_k}' \xrightarrow{a.s.} \xi.$$

证明: "$\Rightarrow$": 假定 $\xi_n \xrightarrow{P} \xi$, 根据定义, 对它的任一子列 $\{\xi_n'\}$ 都有 $\xi_n' \xrightarrow{P} \xi$, 从而 $\forall k \geq 1, \exists \{\xi_{n_k'}\} \subset \{\xi_n'\}$, 使得

$$
\begin{aligned}
& P\left(|\xi_{n_k}' - \xi| \geq \frac{1}{k}\right) \leq \frac{1}{2^k} \\
\Rightarrow & P\left(\bigcup_{k=m}^{\infty} \left[|\xi_{n_k}' - \xi| \geq \frac{1}{k}\right]\right) \leq \sum_{k=m}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-1}}.
\end{aligned}
$$

$\forall \varepsilon > 0$, 取 m 充分大, 使得 $\varepsilon > \frac{1}{k}, k = m, m+1, \cdots$, 则对于 $j \geq \max\left\{m, \left[\frac{1}{\varepsilon}\right] + 1\right\}$, 有

$$
\begin{aligned}
& P\left(\bigcup_{k=m}^{\infty} [|\xi_{n_k}' - \xi| > \varepsilon]\right) \leq \frac{1}{2^{j-1}} \\
\Longrightarrow & \lim_{j\to\infty} P\left(\bigcup_{k=j}^{\infty} [|\xi_{n_k}' - \xi| \geq \varepsilon]\right) = 0 \\
\iff & P\left(\bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} [|\xi_{n_k}' - \xi| \geq \varepsilon]\right) = 0.
\end{aligned}
$$

"$\Leftarrow$" (反证) 假设 $\xi_n \xrightarrow{P} \xi$ 不成立, 即 $\exists \varepsilon_0 > 0, \delta > 0$ 以及 $\{\xi_{n_m}\} \subset \{\xi_n\}$, 使得

$$P(|\xi_{n_m} - \xi| \geq \varepsilon_0) > \delta > 0, \forall n \geq 1.$$

(即这个子列不是依概率收敛). 对此子列而言,

$$\begin{aligned}&P\left(\bigcap_{k=1}^{\infty}\bigcup_{m=k}^{\infty}[|\xi_{n_m}-\xi|\geq\varepsilon_0]\right)\\&=\lim_{k\to\infty}P\left(\bigcup_{m=k}^{\infty}[|\xi_{n_m}-\xi|\geq\varepsilon_0]\right)\\&\geq\limsup_{m\to\infty}P(|\xi_{n_m}-\xi|\geq\varepsilon_0)>\delta>0,\end{aligned}$$

与 $\xi'_{n_k}\xrightarrow{a.s.}\xi$ 矛盾. □

注: 以后将会频繁利用上面两个定理, 尤其是几乎必然收敛的刻画. 利用这个刻画再结合概率测度的从上 (下?) 连续性可以进行放缩.

例 4.3.3

设 $\xi_n\xrightarrow{P}\xi$, $f:\mathbb{R}\to\mathbb{R}$ 连续, 则 $f(\xi_n)\xrightarrow{P}f(\xi)$.

证明: $\xi_n\xrightarrow{P}\xi$, 则 $\forall\{\xi'_n\}\subset\{\xi_n\},\exists\{\xi'_{n_k}\}\subset\{\xi'_n\}$, 使得 $\xi'_{n_k}\xrightarrow{a.s.}\xi$. 由于 $f\in C(\mathbb{R})$, 故 $f(\xi'_{n_k})\xrightarrow{a.s.}f(\xi)$ （复合函数连续性, 可参考数学分析书）, 即得 $f(\xi_n)\xrightarrow{P}f(\xi)$. □

注: a.s. 收敛可推出依概率收敛, 反之不成立. 见下面定理与例子:

定理 4.3.4

如果 $\xi_n\xrightarrow{a.s.}\xi(n\to\infty)$, 则 $\xi_n\xrightarrow{P}\xi(n\to\infty)$.

证明: 利用几乎必然收敛的刻画, 即

$$\xi_n\xrightarrow{a.s.}\xi\Longleftrightarrow\forall\varepsilon>0,P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right)=0.$$

由于 $\forall\varepsilon>0$,

$$\lim_{n\to\infty}P(|\xi_n-\xi|\geq\varepsilon)\leq\lim_{k\to\infty}P\left(\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right)=P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right)=0,$$

所以 $\xi_n\xrightarrow{P}\xi(n\to\infty)$. □

例 4.3.5

令 $\Omega=(0,1],\mathscr{F}=(0,1]\cap\mathscr{B}(\mathbb{R})$ 是个 Borel-σ 代数. 把 P 取为 $\mathscr{B}(\mathbb{R})$ 上的 Lebesgue 测度. 令

$$\eta_{ki}(\omega)=\begin{cases}1,&\omega\in\left(\dfrac{i-1}{k},\dfrac{i}{k}\right]\\0,&\omega\notin\left(\dfrac{i-1}{k},\dfrac{i}{k}\right]\end{cases}$$

这里 $k=1,2,\cdots;i=1,2,\cdots,k$, 先取定 k 再取定 i. 定义 $\xi_n\triangleq\eta_{ki},n=i+\dfrac{k(k-1)}{2}$, 该定义合理（没重复的 n 对应不同的 k,i）.

注意到 $\forall\omega\in\Omega$, 必有无穷个 n 使得 $\xi_n(\omega)=0$, 也有无穷多个 m 使得 $\xi_m(\omega)=1$, 所以不可能几乎必然收敛为 0, 即

$$\xi_n\overset{a.s.}{\nrightarrow}0(n\to\infty).$$

另一方面,

$$\forall\varepsilon\in(0,1),P(\xi_n(\omega)>\varepsilon)=P(\eta_{ki}(\omega)>\varepsilon)=P\left(\frac{i-1}{k}<\omega\leq\frac{i}{k}\right)=\frac{1}{k},$$

让 $n\to\infty$, 则根据 $n=i+\dfrac{k(k-1)}{2}\leq\dfrac{k(k+1)}{2}$ 可知 $k\to\infty$, 则 $\lim\limits_{n\to\infty}P(\xi_n>\varepsilon)=0$. 从而

$$\xi_n\xrightarrow{P}0(n\to\infty).$$

定理 4.3.6

r.v. 列 $\xi_n \xrightarrow{L^p} \xi$, 等价于"$\xi_n$ 依概率收敛到 ξ, 且 $E|\xi_n|^p \to E|\xi|^p$."

证明: "$\Rightarrow$": 利用 Chebyshev 不等式, $\forall \varepsilon > 0$,

$$P(|\xi_n - \xi| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} E|\xi_n - \xi|^p = \frac{1}{\varepsilon^p} \|\xi_n - \xi\|_p^p \to 0(n \to \infty).$$

所以 ξ_n 依概率收敛到 ξ. 而根据范数的三角不等式有

$$|\|\xi_n\|_p - \|\xi\|_p| \leq \|\xi_n - \xi\|_p \to 0(n \to \infty),$$

则 $E|\xi_n|^p \to E|\xi|^p$.

"$\Leftarrow$":(本科范围内不作要求.)需要先证明下面的定理:

定理 4.3.7. 推广的控制收敛定理

设 $\{f_n\}$ 是 $(\Omega, \mathcal{F}, \mu)$ 上的可测实值函数, 且 $f_n \xrightarrow{a.s} f$ 或 $f_n \xrightarrow{\mu} f$. 又设 g_n 是非负实值函数, 满足 $g_n \xrightarrow{a.s} g$ 或 $g_n \xrightarrow{\mu} g$. 若 g 以及每个 g_n 都可积, $\mu(g_n) \to \mu(g)$, 且 $|f_n| \leq g_n, a.e., \forall n \geq 1$, 则 $\lim\limits_{n\to\infty} \mu(|f_n - f|) = 0$. 特别地, $\lim\limits_{n\to\infty} \mu(f_n) = \mu(f)$.

其中 $\mu(f) = \int_\Omega f \mathrm{d}\mu$. 用概率测度的语言来写就是: (注意 a.s. 收敛可以推出依概率收敛)

设 $\{X_n\}$ 是 $(\Omega, \mathcal{F}, P)$ 上的 r.v. 列, 且 $X_n \xrightarrow{P} X$. 又设 $\{Y_n\}$ 是非负 r.v. 列, 满足 $Y_n \xrightarrow{P} Y$. 若 $\lim\limits_{n\to\infty} EY_n = EY$, 且 $|X_n| \leq Y_n, a.e., \forall n \geq 1$, 则 $\lim\limits_{n\to\infty} E|X_n - X| = 0$.

由于

$$|\xi_n - \xi|^p \leq 2^{p-1}(|\xi_n|^p + |\xi|^p),$$

对 $X_n = (\xi_n - \xi)^p, Y_n = 2^{p-1}(|\xi_n|^p + |\xi|^p)$ 使用前面的定理即可:

由条件, $X_n \xrightarrow{P} 0$, $\{Y_n\}$ 是非负 r.v. 列, $Y_n \xrightarrow{P} 2^p|\xi|^p \triangleq Y$, 且

$$EY_n = 2^{p-1}(E|\xi_n|^p + E|\xi|^p) \to 2^p E|\xi|^p = EY,$$

由推广的控制收敛定理, 则 $E|X_n| \to 0$, 即 $\|\xi_n - \xi\|_p \to 0$. □

注: "推广的控制收敛定理"证明过程参考严加安的《测度论讲义》.

定理 4.3.8. 控制收敛定理

$\{\xi_n, \xi\}$ 是 r.v., η 为 r.v., 满足:

① $\xi_n \xrightarrow{P} \xi$ 或 $\xi_n \xrightarrow{a.s} \xi$,

② $|\xi_n| \leq \eta, \forall n \geq 1$, a.s.,

③ η 可积 $(E|\eta| < +\infty)$.

则 $\lim\limits_{n\to\infty} E\xi_n = E\xi$.

定理 4.3.9

$X_n \xrightarrow{P} X \Longrightarrow X_n \xrightarrow{d} X$.

利用控制收敛定理立即得证. 反之不成立, 见下面例子.

例 4.3.10. 反例

设 $\Omega=\{\omega_1,\omega_2\}, P(\omega_1)=P(\omega_2)=\dfrac{1}{2}$. 令

$$\xi(\omega)=\begin{cases}1, & \omega=\omega_1,\\ -1, & \omega=\omega_2.\end{cases},$$

此时 $\{\xi_n\}_{n\geq 1}$ 同分布, 但是 $P(|\xi_n-\xi|=2)=1$, 不依概率收敛.

定理 4.3.11

如果 $\{\xi_n\}$ 是单调下降 r.v. 正序列, 且 $\xi_n\xrightarrow{P}\xi(n\to\infty)$, 则 $\xi_n\xrightarrow{a.s.}\xi(n\to\infty)$.

证明: 此时在前面定理中,

$$\lim_{n\to\infty}P(|\xi_n-\xi|\geq\varepsilon)=\lim_{k\to\infty}P\left(\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right),$$

(即第 2 条式子的第一个不等号可以变为等号). □

定理 4.3.12

如果 $\xi_n\xrightarrow{P}\xi(n\to\infty)$, 则 $\xi_n\xrightarrow{L}\xi(n\to\infty)$.

证明: $\forall f\in C(\mathbb{R}),|f|<\infty$, 有

$$\xi_n\xrightarrow{P}\xi\Rightarrow f(\xi_n)\xrightarrow{P}f(\xi).$$

根据控制收敛定理, 可得

$$Ef(\xi_n)\xrightarrow{P}Ef(\xi).$$

根据 Levy 定理, 可得 $\xi_n\xrightarrow{L}\xi(n\to\infty)$. □

注: Levy 定理可以用来处理依分布收敛.

定理 4.3.13

$\xi_n\xrightarrow{P}C(n\to\infty)$ 的充分必要条件是 $\xi_n\xrightarrow{L}C(n\to\infty)$.

证明: "$\Rightarrow$": 前面定理已证.

"$\Leftarrow$": $\forall\varepsilon>0$,

$$\begin{aligned}&P(|\xi_n-c|\geq\varepsilon)\leq P(\xi_n\geq C+\varepsilon)+P(\xi_n\leq C-\varepsilon)\\ =&1-P(\xi_n<C+\varepsilon)+P(\xi_n\leq C-\varepsilon)\\ \leq&1-P(\xi_n\leq C+\varepsilon/2)+P(\xi_n\leq C-\varepsilon)\\ =&1-F_n(C+\varepsilon/2)+F_n(C-\varepsilon)\\ \to&1-F(C+\varepsilon/2)+F(C-\varepsilon)=1-1+0=0(n\to\infty).\end{aligned}$$

故 $\lim\limits_{n\to\infty}P(|\xi_n-c|\geq\varepsilon)=0$. □

例 4.3.14

设 $\Omega=[0,1],(\Omega,\mathscr{F},P)$ 是概率空间, $\Omega(\omega)\equiv 0$, 且

$$\xi_n(\omega)=\begin{cases} n^{1/r}, & 0<\omega\le\dfrac{1}{n} \\ 0, & \dfrac{1}{n}<\omega<1 \end{cases},$$

则 $\forall\omega\in\Omega,\xi_n(\omega)\to\xi(\omega)$, 即 $\xi_n\xrightarrow{a.s}\xi$. 又 $\forall\varepsilon>0$,

$$P(|\xi_n(\omega)-\xi(\omega)|\ge\varepsilon)\le\frac{1}{n},$$

因此 $\xi_n\xrightarrow{P}\xi$, 然而 $E|\xi_n-\xi|^p=(n^{1/r})^r\cdot\dfrac{1}{n}=1$, 故不是 p 阶收敛.

定理 4.3.15

$\xi_n\xrightarrow{L^{p+1}}\xi\Rightarrow\xi_n\xrightarrow{L^p}\xi$.

证明: 记 $a_k=E|\xi|^k$, 下面我们证 $\sqrt[k]{a_k}\le\sqrt[k+1]{a_{k+1}}$. 利用 Cauchy-Schwarz 不等式有

$$\left|E|\xi|^{\frac{k-1}{2}}|\xi|^{\frac{k+1}{2}}\right|^2\le E|\xi|^{k-1}E|\xi|^{k+1},$$

即 $a_k^2\le a_{k-1}a_{k+1}$. 于是 $a_k^{2k}\le a_{k-1}^k a_{k+1}^k$. 由于 $a_0=1$, 取 $k=1,2,\cdots,n$ 有

$$a_1^2\le a_2^1\qquad a_2^4\le a_1^2a_3^2\qquad a_3^6\le a_2^3a_4^3\qquad\cdots\cdots\qquad a_n^{2n}\le a_{n-1}^n,a_{n+1}^n$$

把上面 n 个不等式相乘, 可得

$$a_1^2a_2^4a_3^6\cdots a_{n-1}^{2n-2}a_n^{2n}\le a_1^2a_2^4a_3^6\cdots a_{n-1}^{2n-2}a_n^{n-1}a_{n+1}^n,$$

即 $a_n^{2n}\le a_n^{n-1}a_{n+1}^n\Leftrightarrow a_n^{n+1}\le a_{n+1}^n\Leftrightarrow\sqrt[k]{a_k}\le\sqrt[k+1]{a_{k+1}}$. □

§ 4.4 习题

1. 设 $\{X_n\}$ 是 r.v. 列, 满足两点分布, 且 $P(X_n=1)=\dfrac{1}{n^2}$. 证明: $\{X_n\}$ 几乎必然收敛到 0.
2. **(2019 期末)** 设 $\{X_n\}$ 相互独立, 则对任意 $\varepsilon>0$,

$$P(\lim_{n\to\infty}X_n=0)=1\Leftrightarrow\sum_{n=1}^{+\infty}P(|X_n|\ge\varepsilon)<+\infty.$$

3. 设 $\{X_n\}$ 是 i.i.d.r.v, $X_1\sim U(0,1)$. 令 $Z_n=\left(\prod_{i=1}^n X_i\right)^{1/n}$, 证明; 存在 C 使得 $Z_n\xrightarrow{P}C$.
4. 如果 $\sum_{n=1}^{\infty}\|\xi_n-\xi\|_p^p<+\infty$, 则 $\xi_n\xrightarrow{a.s.}\xi$.
5. **(2019 期末)** 设 ξ 是随机变量. 证明: $E\xi^2<\infty$ 当且仅当 $\sum_{n=1}^{\infty}nP(|\xi|>n)<\infty$.

6. 让 $\{X_n\}$ 为正 r.v. 列, 并假定 $X_n \xrightarrow{P} 0$ 且 $\lim_{n\to\infty} EX_n = 2$. 证明 $\lim_{n\to\infty} E|X_n - 1|$ 存在, 并求这个值.

7. **(2019 期末)** 设 $\{X_n\}$ 是两两不相关, $\sum_{n=1}^{+\infty} nD(X_n) < +\infty$. 证明: $S_n = \sum_{i=1}^{n}(X_i - EX_i)$ 几乎必然收敛.

8. **(2014 丘赛 Individual)** 设 $\{X_n\}$ 是一列不相关的 r.v. 列且均值为 0, 满足 $\sum_{n=1}^{\infty} nE|X_n|^2 < \infty$. 则 $S_n = \sum_{i=1}^{n} X_i$ 几乎必然收敛.

9. **(2016 丘赛 Team)** 让 $\{X_n\}_{n\geq 1}$ 为 i.i.d. 实值 r.v. 列, 证明或否定: 若 $\overline{\lim_{n\to\infty}} \frac{|X_n|}{n} \leq 1$,a.s., 则

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) < \infty.$$

10. **(2019 丘赛 Team)** 假定 $\{X_n\}_{n\geq 1}$ 是 i.i.d.r.v. 列且共同分布是参数为 1 的指数分布, 则

$$P\left(\limsup_{n\to\infty} \frac{X_n}{\log n} = 1\right) = 1.$$

11. 设 $\{X_n : n \geq 1\}$ 是独立的 $N(0,1)$ 随机变量, 证明:
(1) $P\left(\limsup_{n\to\infty} \frac{|X_n|}{\sqrt{\log n}} = \sqrt{2}\right) = 1;$
(2) $P\left(X_n > a_n \quad i.o.\right) = \begin{cases} 0, & \text{若} \sum_n P(X_1 > a_n) < \infty, \\ 1, & \text{若} \sum_n P(X_1 > a_n) = \infty, \end{cases}$
(提示: 可以用 Mills's Ratio 来估计 $\Phi(x)$.)

12. **(2019 丘赛 Individual)** 设 $\{X_n\}_{n\geq 1}$ 是一列正 r.v., 存在常数 $C > 0$ 使得

$$EX_n \leq C, E\max\{0, -\log X_n\} \leq C, \forall n \geq 1.$$

则 $\limsup_{n\to\infty} X_n^{\frac{1}{n}} = 1$.

13. **(2019 期末)** 设 $f : (0,\infty) \to (0,\infty)$ 单调不降, 若随机变量序列 $\{X_n, X\}_{n\geq 1}$ 满足

$$\sum_{n=1}^{\infty} Ef(|X_n - X|) < \infty,$$

证明: $X_n \xrightarrow{a.s.} X$.

14. **(2010 期末)** 设 $\{\xi_n\}$ 为独立随机变量序列, $P(\xi_n = 0) = \frac{1}{2} = P(\xi_n = 1)$. 令 $\eta_n = \sum_{i=1}^{n} \frac{\xi_i}{2^i}$, 证明 η_n 依分布收敛于 $(0,1)$ 的均匀分布. (提示: 可以考虑特征函数.)

15. **(2013 年丘成桐大学生数学竞赛团体赛)** 设实数 $\varepsilon > 0$, 证明: 对几乎所有的 $x \in [0,1]$, 只有有限个有理数 $\frac{p}{q} \in (0,1)$(其中 $p, q \in \mathbb{N}^+$) 满足

$$\left|x - \frac{p}{q}\right| < \frac{1}{q^2(\log q)^{1+\varepsilon}}.$$

16. 设 $(X_n, n \geq 1)$ 是独立 r.v. 序列, 则 $\limsup_{n\to\infty} X_n$ 与 $\liminf_{n\to\infty} X_n$ 是退化随机变量 (即 a.s. 等于某个常数).

17. 设 $\{X_n\}$ 是 i.i.d., 则 $\dfrac{X_n}{\log n} \xrightarrow{a.s.} 0 \Leftrightarrow \forall c > 0, \mathbb{E}e^{c|X_1|} < \infty$.

18. 设 X, Y 相互独立, X 有密度函数, 则 $X + Y$ 也有密度函数.

19. 设 $(X_n, n \geq 1)$ 是 i.i.d.r.v. 列, X_n 都服从指数为 1 的指数分布, 即 $P(X_n > x) = e^{-x}, x \geq 0$.

    (1) 证明:$P(X_n > \alpha \log n, i.o.) = \begin{cases} 0, & 若 \alpha > 1, \\ 1, & 若 \alpha \leq 1. \end{cases}$

    (2) 令 $L = \limsup_{n\to\infty} \dfrac{X_n}{\log n}$, 证明: $P(L = 1) = 1$. (提示: 证明 $P(L \geq 1) = 1, P(L > 1) = 0$.)

20. 设 (ξ_n) 是一列非负实值随机变量, 则存在一正实数序列 (c_n), 使得 $\sum_{n=1}^{\infty} c_n \xi_n < \infty$, a.s..

    提示: 取正实数序列 (a_n), 使得 $\sum_{n=1}^{\infty} P(\xi_n > a_n) < \infty$, 用 Borel-Cantelli 引理并令 $c_n = (2^n a_n)^{-1}$.

21. 设 $(\Omega, \mathscr{F}, P)$ 是概率空间.

    (1) 设 $\xi_1, \cdots, \xi_n$ 是非负随机变量, $\mathbb{E}\xi_i = 1, 1 \leq i \leq n$. 则 $\prod_{i=1}^{n} \dfrac{1}{n} \sum_{j=1}^{n} \mathbb{E}(\xi_i \xi_j) \geq 1$.

    (2) 设 $A_1, \cdots, A_n \in \mathscr{F}$, 且每个 $P(A_i) > 0$. 令 $A = \bigcup_{i=1}^{n} A_i$, 则

$$\prod_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} \frac{P(A_i A_j)}{P(A_i) P(A_j)} \geq \left(\frac{1}{P(A)} \right)^n .$$

第 5 章　大数定律与极限定理

§ 5.1　大数定律

定理 5.1.1. Chebyshev 弱大数定律

设 $\{x_n\}_{n\geq 1}$ 是两两不相关的 r.v. 序列, 且 $\sup\limits_n D\xi_n$ 有界, 则 $\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\frac{1}{n}\left|\sum_{i=1}^{n}(\xi_i - E\xi_i)\right| \geq \varepsilon\right) = 0.$$

证明: 利用 Chebyshev 不等式, 得

$$\begin{aligned}\lim_{n\to\infty} P\left(\frac{1}{n}\left|\sum_{i=1}^{n}(\xi_i - E\xi_i)\right| \geq \varepsilon\right) \leq& \frac{1}{\varepsilon^2}D\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i - E\xi_i)\right)\\ =&\frac{1}{n^2\varepsilon^2}D\left(\sum_{i=1}^{n}(\xi_i - E\xi_i)\right) \leq \frac{1}{n\varepsilon}\sup_i D\xi_i \to 0(n\to\infty).\end{aligned}$$

定理 5.1.2. Borel 强大数定律

设 μ_n 是事件 A 在 n 次独立试验中出现的次数, p 是 A 在每次试验中发生的概率, 则 $\dfrac{\mu_n}{n} \xrightarrow{a.s.} p$.

证明: $\forall \varepsilon > 0$, 根据几乎必然收敛的刻画, 我们只需要证

$$P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left[\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right]\right) = 0.$$

再根据 Borel-Cantelli 引理, 我们只需证

$$\sum_{n=1}^{\infty} P\left(\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right) < \infty.$$

记 $\xi_i = \begin{cases}1, & \text{第 i 次 A 发生}\\ 0, & \text{第 i 次 A 不发生}\end{cases}$, 则 $E\xi_i = p, \dfrac{\mu_n}{n} - p = \dfrac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - p)$, 相当于作了中心化处理. 由于 $\{\xi_n\}$ 之间相互独立, 则有

$$E(\xi_i - p)(\xi_j - p) = E(\xi_i - p)E(\xi_j - p) = 0, \forall i \neq j.$$

根据推广 Chebyshev 不等式以及题目所给的 i.i.d 条件,

$$\begin{aligned}P\left(\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right) \leq& \frac{1}{\varepsilon^4}E\left|\frac{\mu_n}{n} - p\right|^4\\ =&\frac{1}{\varepsilon^4 n^4}E\left(\sum_{i=1}^{n}(\xi_i - p)\right)^4\\ =&\frac{1}{\varepsilon^4 n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E(\xi_i - p)(\xi_j - p)(\xi_k - p)(\xi_l - p)\\ =&\frac{1}{\varepsilon^4 n^4}\left[nE|\xi_i - p|^4 + \mathrm{C}_n^2\mathrm{C}_4^2 E(\xi_1 - p)^2(\xi_2 - p)^2\right]\\ =&\frac{1}{\varepsilon^4 n^4}\left[npq(p^3 + q^3) + 3n(n-1)p^2q^2\right] \leq \frac{Cn^2}{\varepsilon^4 n^4} = \frac{C}{n^2\varepsilon^4}, \text{for some C.}\end{aligned}$$

从而

$$\sum_{n=1}^{\infty} P\left(\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{C}{\varepsilon^4}\cdot\frac{\pi^2}{6} < +\infty.$$

证明完毕. □

事实上, Borel-Cantelli 引理还应用于 Komogorov 强大数定律,

定理 5.1.3. Kolmogorov 不等式

设 $X_1,\cdots,X_n$ 独立且 $E|X_i|<+\infty$, 令

$$S_n=\sum_{i=1}^{n}X_i, i=1,2,\cdots,n,$$

则 $\forall\varepsilon>0$, 有

$$P\left(\max_{1\le k\le n}|S_k-ES_k|\ge\varepsilon\right)\le\frac{1}{\varepsilon^2}DS_n.$$

证明: 令

$$\lambda_j=\left[\max_{1\le k\le j-1}|S_k-ES_k|<\varepsilon,|S_j-ES_j|\ge\varepsilon\right],\forall j=1,2,\cdots,n,$$

并约定 $\lambda_1=0$, 则 $\lambda_j\in\sigma(x_1,\cdots,x_j)$（由 $x_1,\cdots,x_j$ 生成的 sigma-代数）易知 λ_j 与 $\sigma(x_{j+1},\cdots,x_n)$ 相互独立, 且 λ_j 之间两两不交 (作了不交并处理), 且满足下面的等式:

$$\sum_{j=1}^{n}\lambda_j=\bigcup_{j=1}^{n}[|S_j-ES_j|\ge\varepsilon]=\left[\max_{1\le j\le n}|S_j-ES_j|\ge\varepsilon\right].$$

记 $\lambda=\sum_{j=1}^{n}\lambda_j\subset\Omega$, 于是

$$\begin{aligned}
DS_n&=E(S_n-ES_n)^2\\
&\ge\int_{\lambda}(S_n-ES_n)^2dP\\
&=\sum_{j=1}^{n}\int_{\lambda_j}(S_n-ES_n)^2dP\ (\text{之前作了不交并处理})\\
&=\sum_{j=1}^{n}\int_{\lambda_j}[(S_n-ES_n)-(S_j-ES_j)+(S_j-ES_j)]^2dP\\
&=\sum_{j=1}^{n}\int_{\Omega}\left[\left(\sum_{i=j+1}^{n}(X_i-EX_i)\right)+(S_j-ES_j)\right]^2I_{\lambda_j}dP\\
&\ge\sum_{j=1}^{n}\int_{\Omega}2\sum_{i=j+1}^{n}(X_i-EX_i)(S_j-ES_j)I_{\lambda_j}+(S_j-ES_j)^2I_{\lambda_j}dP\\
&=\sum_{j=1}^{n}\int_{\Omega}(S_j-ES_j)^2I_{\lambda_j}dP\ (\text{前一行中第一项为 0, 由独立性})\\
&\ge\sum_{j=1}^{n}\int_{\lambda_j}\varepsilon^2dP\ (\lambda_j\ \text{定义})\\
&=\varepsilon^2\sum_{j=1}^{n}P(\lambda_j)\ (\text{概率定义})\\
&=\varepsilon^2P\left(\max_{1\le k\le n}|S_k-ES_k|\ge\varepsilon\right).\square
\end{aligned}$$

注: 令 $n=1$ 可得 Chebyshev 不等式.

定理 5.1.4

设 $\{X_n\}_{n\geq 1}$ 是独立 r.v. 序列, 且期望有限, 若 $\sum\limits_{n=1}^{\infty} DX_n < +\infty$, 则 $\sum(X_n - EX_n)$ 几乎必然收敛.

证明: 不妨设 $EX_n = 0, \forall n \geq 1$. 令 $S_n = \sum\limits_{i=1}^{n} X_i, \forall n \geq 1$. 先对欲证命题进行转化:

$$
\begin{aligned}
& S_n \text{a.s. 收敛} \\
\Leftrightarrow & S_n - S_m \xrightarrow{a.s.} 0(n, m \to \infty) \text{ (Cauchy 准则)} \\
\Leftrightarrow & \forall \varepsilon > 0, P\left(\bigcap_{k=1}^{\infty} \bigcup_{m,n=k}^{\infty} [|S_n - S_m| \geq \varepsilon]\right) = 0 \text{ (a.s. 收敛刻画)} \\
\Leftrightarrow & \forall \varepsilon > 0, \lim_{n\to\infty} P\left(\bigcup_{m,n=k}^{\infty} [|S_n - S_m| \geq \varepsilon]\right) = 0 \qquad \text{(单调性)} \\
\Leftrightarrow & \forall \varepsilon > 0, P\left(\bigcup_{l=1}^{\infty} [|S_{l+k} - S_k| \geq \varepsilon]\right) \to 0(k \to \infty).
\end{aligned}
$$

固定上面的 k, 注意到

$$
\begin{aligned}
& P\left(\bigcup_{l=1}^{\infty} [|S_{l+k} - S_k| \geq \varepsilon]\right) \\
= & \lim_{m\to\infty} P\left(\bigcup_{l=1}^{m} [|S_{l+k} - S_k| \geq \varepsilon]\right) \text{ (从下连续性)} \\
= & \lim_{m\to\infty} P\left(\max_{1\leq l\leq m} |S_{l+k} - S_k| \geq \varepsilon\right) \\
\leq & \lim_{m\to\infty} \frac{1}{\varepsilon^2} D(S_{m+k} - S_k) \text{ (前一定理)} \\
= & \frac{1}{\varepsilon^2} \lim_{m\to\infty} \sum_{i=k+1}^{k+m} DX_i \\
= & \frac{1}{\varepsilon^2} \sum_{i=k+1}^{\infty} DX_i \to 0(k \to \infty). \text{ (Cauchy 准则)}
\end{aligned}
$$

这样证明已完成. □

引理 5.1.5. Kronecker

设 $\sum\limits_{k=1}^{\infty} a_k$ 存在, 且 $\{p_n\}$ 单调递增趋于正无穷, 则

$$\lim_{n\to\infty} \frac{p_1 a_1 + \cdots + p_n a_n}{p_n} = 0.$$

证明: 略.

定理 5.1.6. Kolmogorov 强大数定律

设 $\{X_n\}_{n\geq 1}$ 为独立 r.v. 列, $\{b_n\}$ 单调递增趋于正无穷, 若

$$\sum_{n=1}^{\infty}\frac{D(X_n)}{b_n^2}<+\infty,$$

则 $\displaystyle\frac{1}{b_n}\sum_{i=1}^{n}(X_i-EX_i)\xrightarrow{a.s.}0$.

证明: 根据定理 5.1.4, $\displaystyle\sum_{n=1}^{\infty}\frac{X_n}{b_n}$ 几乎必然收敛. 根据 Kronecker 引理,

$$\frac{S_n-ES_n}{b_n}=\frac{1}{b_n}\sum_{k=1}^{n}\left(\frac{X_k-EX_k}{b_k}\right)b_k\xrightarrow{a.s.}0.$$

□

推论 5.1.7

设 $\{X_n\}_{n\geq 1}$ 相互独立 r.v. 列, 且有相同的均值与方差, 则 $\displaystyle\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{a.s.}EX_1$.

证明: 在 Kolmogorov 强大数定律中令 $b_n=n, D(X_n)=C$ 即可. □

例 5.1.8

设 $\{\xi_n\}_{n\geq 1}$ 是 i.i.d.r.v. 列, 则

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{a.s.}a(<+\infty)\Leftrightarrow E|\xi_1|<+\infty\text{且}a=E\xi_1.$$

证明: “$\Rightarrow$”: 如下.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{a.s.}a\Rightarrow\frac{\xi_n}{n}&=\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{n-1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\xi_i\right)\xrightarrow{a.s.}0\\
&\Rightarrow P\left(\left[\left|\frac{\xi_n}{n}\right|\geq 1\right]\text{发生无穷次}\right)=0\qquad(\text{a.s. 收敛的刻画})\\
&\Rightarrow\sum_{n=1}^{\infty}P\left(\left|\frac{\xi_n}{n}\right|\geq 1\right)<+\infty\qquad(\text{独立性, Borel-Cantelli})\\
&\Rightarrow\sum_{n=1}^{\infty}P(|\xi_n|\geq n)<+\infty\qquad(\text{同分布})\\
&\Rightarrow E|\xi_1|\leq\sum_{n=1}^{\infty}P(|\xi_n|\geq n)+1<+\infty.\qquad(\text{Nice 引理})
\end{aligned}$$

另外,

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{a.s.}a\Rightarrow E\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{a.s.}Ea=a,$$

根据同分布可得 $E\xi_i\xrightarrow{a.s.}a$, 而 $E\xi_i$ 是常数, 所以 $E\xi_i=a$. □

"$\Leftarrow$": 不妨设 $E\xi_1=0$(否则可以作中心化处理). 令 $\eta_i=\xi_i I_{[|\xi_i|<i]}, \forall i\geq 1$(作了一个截断, 留下 $-i<\xi_i<i$ 部分), 则

$$\begin{aligned}\sum_{n=1}^{\infty}P(\xi_n\neq\eta_n)&\leq\sum_{n=1}^{\infty}P(|\xi_n|\geq n)\ \text{(i.i.d)}\\&\leq E|\xi_1|<\infty\ \text{(Nice 引理)}.\end{aligned}$$

根据 Borel-Cantelli 引理, 可得 $P(\xi\neq\eta_i\text{发生无穷多次})=0$. 也就是说只有有限个 $\xi_i\neq\eta_i$, 于是

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{a.s.}0\\\Leftrightarrow&\frac{1}{n}\sum_{i=1}^{n}\eta_i\xrightarrow{a.s.}0\\\Leftrightarrow&\frac{1}{n}\sum_{i=1}^{n}(\eta_i-E\eta_i)\xrightarrow{a.s.}0.\end{aligned}$$

考虑到

$$\begin{aligned}&\sum_{n=1}^{\infty}\frac{1}{n^2}D\eta_n<\infty\\\Rightarrow&\sum_{n=1}^{\infty}\frac{\eta_n-E\eta_n}{n}\text{a.s.}\qquad\text{(定理 5.1.4)}\\\Rightarrow&\frac{1}{n}\sum_{i=1}^{n}(\eta_i-E\eta_i)\xrightarrow{a.s.}0.\quad\text{(Kronecker)}\end{aligned}$$

只需证 $\sum_{n=1}^{\infty}\frac{1}{n^2}D\eta_n<\infty$. 事实上,

$$
\begin{aligned}
&\sum_{n=1}^{\infty}\frac{1}{n^2}D\eta_n\\
\leq&\sum_{n=1}^{\infty}\frac{1}{n^2}E\xi_n^2 I_{[|\xi_n|<n]} && \text{(方差定义)}\\
=&\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=0}^{n-1}E\xi_n^2 I_{[k\leq|\xi_n|<k+1]}\\
=&\sum_{k=0}^{\infty}\sum_{n=k+1}^{\infty}\frac{1}{n^2}E\xi_n^2 I_{[k\leq|\xi_n|<k+1]} && \text{(正项级数换序)}\\
<&\sum_{k=1}^{\infty}\sum_{n=k+1}^{\infty}\left(\frac{1}{n(n-1)}E\xi_n^2 I_{[k\leq|\xi_n|<k+1]}\right)+\sum_{n=1}^{\infty}\frac{1}{n^2}E\xi_n^2 I_{[0\leq|\xi_n|<1]} && \text{(提取 } k=0)\\
\leq&\sum_{k=1}^{\infty}\left(\frac{1}{k}E\xi_n^2 I_{[k\leq|\xi_n|<k+1]}\right)+\sum_{n=1}^{\infty}\frac{1}{n^2} && \text{(因为 } |\xi_n|<1)\\
\leq&\sum_{k=1}^{\infty}\left(\frac{(k+1)^2}{k}P(k\leq|\xi_n|<k+1)\right)+\frac{\pi^2}{6} && \text{(因为 } |\xi_n|<k+1)\\
\leq&3\sum_{k=1}^{\infty}kP(k\leq|\xi_n|<k+1)+\frac{\pi^2}{6}\\
\leq&3\sum_{k=1}^{\infty}E|\xi_n|I_{[k\leq|\xi_n|<k+1]}+\frac{\pi^2}{6} && \text{(因为 } k\leq|\xi_n|)\\
\leq&3E|\xi_n|+\frac{\pi^2}{6}<+\infty.\square
\end{aligned}
$$

§ 5.2 中心极限定理

定理 5.2.1. Lindberg-Lévy

设 $\{\xi_n\}_{n\geq 1}$ 是 i.i.d.r.v. 列, 且 $E\xi_1=\mu,\sigma^2=D\xi_1<+\infty$, 令

$$\eta_n\triangleq\frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(\xi_i-\mu)$$

（相当于标准化）, 则

$$\lim_{n\to\infty}P(\eta_n\leq y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-x^2/2}dx(=\Phi(y)).$$

注: 结论等价于 $\eta_n\xrightarrow{L}\mu$, 或者用 Lévy 定理,

$$Ef(\eta_n)\to\int_{-\infty}^{+\infty}f(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

证明: 设 $g(t)$ 是 $\xi_1-\mu$ 的特征函数, 记

$$\varphi_n(t)\triangleq Ee^{i\eta_n t}=\cdots=\left[g\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right]^n.$$

利用特征函数的性质, $g(0)=1, g'(0)=iE(\xi_1-\mu)=0, g''(0)=i^2E(\xi_1-\mu)^2=-D\xi_1$. 当 t 充分小时, 我们有

$$g(t)=g(0)+g'(0)t+\frac{1}{2}g''(0)t^2+o(t^2)=1-\frac{\sigma^2}{2}t^2+o(t^2).(t\to 0)$$

特别地, 对任一固定 $t\in\mathbb{R}$ 以及充分大的 n, 都有

$$\begin{aligned}&\lim_{n\to\infty}\left[g\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right]\\ =&\lim_{n\to\infty}\left[1-\frac{\sigma^2}{2}\frac{t^2}{n\sigma^2}+o\left(\frac{t^2}{n\sigma^2}\right)\right]^n\\ =&\exp\left(\lim_{n\to\infty}n\left[-\frac{t^2}{2n}+o\left(\frac{t^2}{n\sigma^2}\right)\right]\right)=e^{-t^2/2}.\end{aligned}$$

由于 $t\mapsto e^{-t^2/2}$ 是连续的, 利用 Lévy 定理可得 $P(\eta_n\le y)\xrightarrow{W}\Phi(y)$. □

注: (1) 如果一个量由大量相互独立的随机因素所构成, 而每一个个别因素所起到的作用不是很大, 则这个量近似正态分布.

(2) 中心极限定理比大数定律要精细, 事实上,

$$\begin{aligned}P\left(\left|\frac{1}{n}\sum_{i=1}^n\xi_i-\mu\right|<\varepsilon\right)&=P\left(\frac{1}{\sqrt{n\sigma^2}}\left|\sum_{i=1}^n(\xi_i-\mu)\right|\le\frac{\sqrt{n}}{\sigma}\varepsilon\right)\\ &=P\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\le\eta_n\le\frac{\sqrt{n}\varepsilon}{\sigma}\right)\\ &\sim\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)-\Phi\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right)\\ &=2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)-1.\end{aligned}$$

这给了个定量估计, 但是大数定律只给了定性估计.

下面的例子很有意思，把区间 [0,1] 中的数的二进制表示与概率论结合起来.

例 5.2.2. Borel

若 $\{x_n\}_{n\ge1}$ 为 i.i.d.r.v. 列, 且有相同的分布: $P(\xi_n=1)=P(\xi_n=0)=\frac{1}{2}$. 令 $\eta_n=\sum_{k=1}^n\frac{\xi_k}{2^k}$, 则 η_n 的分布收敛于 $[0,1]$ 上的均匀分布.

证明: 用 Lévy 定理, 只需转化为特征函数. $[0,1]$ 上的均匀分布的分布函数是 $F(x)=1, x\in[0,1]$, 则它的特征函数为

$$f(t)=Ee^{i\xi t}=\int_{-\infty}^{\infty}e^{ixt}dF(x)=\int_0^1 e^{ixt}dx=\frac{1-e^{it}}{it}.$$

而 η_n 的特征函数为

$$\begin{aligned}
f_n(t) &= Ee^{i\sum\limits_{k=1}^{n}\frac{\xi_k}{2^k}t} \\
&= \prod_{k=1}^{n} Ee^{i\frac{\xi_k}{2^k}t} \\
&= \prod_{k=1}^{n} \frac{1}{2}\left(e^{i\frac{0}{2^k}t} + e^{i\frac{1}{2^k}t}\right) \\
&= \prod_{k=1}^{n} \frac{1}{2}\left(1 + \cos\frac{t}{2^k} + i\sin\frac{t}{2^k}\right) \\
&= \prod_{k=1}^{n} \left(\cos^2\frac{t}{2^{k+1}} + i\sin\frac{t}{2^{k+1}}\cos\frac{t}{2^{k+1}}\right) \\
&= \prod_{k=1}^{n} \cos\frac{t}{2^{k+1}} e^{\frac{it}{2^{k+1}}} \\
&= \prod_{k=1}^{n} \cos\frac{t}{2^{k+1}} \prod_{k=1}^{n} e^{\frac{it}{2^{k+1}}} \\
&\to \frac{2}{t}\sin\frac{t}{2}e^{\frac{it}{2}} = \frac{1-e^{it}}{it}.(n\to\infty)\square
\end{aligned}$$

例 5.2.3

用特征函数法证明二项分布的 Poisson 逼近定理.

证明: 同样用 Lévy 定理来证明. 设 $\eta_n \sim B(n,p)$. 又记 $\eta \sim P(\lambda)$, 则 $P(\eta = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, 则 η 的特征函数是

$$f(t) = Ee^{i\eta t} = \sum_{k=1}^{\infty} e^{ikt}\cdot\frac{e^{-\lambda}\lambda^k}{k!} = \cdots = e^{\lambda(e^{it}-1)}.$$

另外, η_n 的特征函数是

$$f_n(t) = Ee^{i\eta t} = \sum_{k=0}^{n} e^{itk}\mathrm{C}_n^k p^k(1-p)^{n-k} = (pe^{it}+1-p)^n = \left[1+\frac{np(e^{it}-1)}{n}\right]^n$$

这样, 如果 $np \to \lambda(n\to\infty)$, 则 $f_n(t)\to f(t)$. $\square$

例 5.2.4

用特征函数法证明泊松分布当 $\lambda\to\infty$ 时, 渐近正态分布.

证明: 同样用 Lévy 定理来证明. 设 $\eta_\lambda \sim P(\lambda)$, 则 $P(\eta_\lambda = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, 则 η_λ 的特征函数是

$$f_\lambda(t) = Ee^{i\eta t} = \sum_{k=1}^{\infty} e^{ikt}\cdot\frac{e^{-\lambda}\lambda^k}{k!} = \cdots = e^{\lambda(e^{it}-1)}.$$

又设 $\xi \sim N(0,1)$, 则 $\eta = \sigma\xi + \mu \sim N(\mu,\sigma^2)$, ξ 的特征函数是

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(tx)e^{-x^2/2}dx.$$

所以

$$f'(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x(-\sin tx)e^{-x^2/2}dx=\cdots=-tf(t).$$

解微分方程并且由 $f(0)=1\Rightarrow f(t)=e^{-t^2/2}$. η 的特征函数是 $g(t)=e^{i\mu t}f(t)=e^{i\mu t}e^{-\sigma^2t^2/2}$. 取 $\mu=\lambda,\sigma=\sqrt{\lambda}$ 即可. □

下面记

$$\Phi(y)=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$$

是标准正态分布函数, $\{\xi_n\}_{n\geq 1}$ 是一列独立 r.v. 列, 均值与方差有限, 记

$$a_k=E\xi_k, b_k^2=D\xi_k, B_n^2=\sum_{k=1}^{n}b_k^2.$$

记 $\eta_k=\dfrac{1}{B_k}\sum\limits_{i=1}^{k}(\xi_i-a_i)$ 为标准化的 r.v.

Lindberg 条件指的是对于 $\forall\varepsilon>0$, 都有 $\lim\limits_{n\to\infty}\dfrac{1}{B_n^2}\sum\limits_{k=1}^{n}\int_{[x-a_k|\geq\varepsilon B_k]}(x-a_k)^2dF_k(x)=0$,

Feller 条件指的是 $\lim\limits_{n\to\infty}\dfrac{1}{B_n}\max\limits_{1\leq k\leq n}b_k=0$.

定理 5.2.5. Lindberg-Feller 中心极限定理

$$\begin{cases}\lim\limits_{n\to\infty}P(\eta_n\leq y)=\psi(y)\\ \{\xi_n\}\text{满足 Feller 条件}\end{cases}\Leftrightarrow\{\xi_n\}\text{满足 Lindberg 条件}.$$

定理 5.2.6. Lyapunov

记 $\{\xi_n\}$ 是独立 r.v. 列. 若 $\exists\delta>0$, 使得

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E(\xi_k-a_k)^{2+\delta}=0,$$

则 $P(\eta_n\leq x)\to\Phi(x)$.

证明: 利用 Lindberg-Feller 中心极限定理以及 Chebyshev 不等式即可.

$$\begin{aligned}&\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{[|x-a_k|\geq\varepsilon B_n]}(x-a_k)^2dF_k(x)\\ \leq&\frac{1}{\varepsilon^\delta B_n^{\delta+2}}\sum_{k=1}^{n}\int_{\mathbb{R}}(x-a_k)^{2+\delta}dF_k(x)\\ =&\frac{1}{\varepsilon^\delta B_n^{\delta+2}}\sum_{k=1}^{n}E(\xi_k-a_k)^{2+\delta}\to 0(n\to\infty).\end{aligned}$$

§ 5.3 习题

1. **(2019 期末)** 某车间有同型号的机床 200 台, 在 1 小时内每台机床约有 70% 时间是工作的. 假定每个机床工作是相互独立的, 工作时每台机床要消耗电能 15 千瓦, 问至少要多少电能才可以有 95% 可能性保证此车间正常生产?
   $(\Phi(1.645) = 0.95)$.

2. **(2019 期末)** 一个复杂系统由 100 个相互独立工作的部件组成, 每个部件正常工作的概率为 0.9. 已知整个系统中至少有 85 个部件正常工作时系统才正常工作, 求系统正常工作的概率.
   $(\Phi(1.83) = 0.9656, \Phi(1.67) = 0.9525)$.

3. 设 $X_2, X_3, \cdots$ 是 i.i.d.r.v., 满足

$$P(X_n = n) = P(X_n = -n) = \frac{1}{2n\log n}, P(X_n = 0) = 1 - \frac{1}{n\log n}.$$

   通过验证 $\dfrac{1}{n}\sum\limits_{i=2}^{n} X_i$ 依概率收敛到 0 而不几乎处处收敛到 0, 来证明这个序列满足弱大数定律, 但不满足强大数定律.

4. 构造一列 i.i.d.r.v$\{X_r : r \geq 1\}$, 满足下列条件:
   (1) $EX_r = 0, \forall r \geq 1$;
   (2) $\dfrac{1}{n}\sum\limits_{r=1}^{n} X_r \overset{a.s.}{\Longrightarrow} -\infty (n \to \infty)$.

5. 设 $\xi_1, \xi_2, \cdots$ 是 i.i.d, 证明: 如果对某个 $\alpha \in (0, 1)$, 有 $E|\xi_1|^\alpha < \infty (\forall n)$, 则 $\dfrac{S_n}{n^{1/\alpha}}$ 几乎必然收敛到 0; 如果对某个 $\beta \in [1, 2)$ 有 $E|\xi_1|^\beta < \infty$, 则 $\dfrac{S_n - nE\xi_1}{n^{1/\beta}}$ 几乎必然收敛到 0.

6. 设 $\xi_1, \xi_2, \cdots$ 是 i.i.d, 且 $E|\xi_n| = \infty (\forall n)$, 证明：

$$\limsup_{n\to\infty} \left| \frac{S_n}{n} - a_n \right| = \infty (a.s.)$$

   对任意序列 $\{a_n\}$ 都成立.

7. **(2019 期末)** 设 $\{X_n\}$ 是一列独立同分布的随机变量且 $X_1 \sim U(0, 1)$. 令

$$Z_n = \left(\prod_{i=1}^{n} X_i \right)^{\frac{1}{n}},$$

   证明存在常数 C, 使得 $Z_n \overset{P}{\Longrightarrow} C$.

第 6 章 (*) 条件期望

§ 6.1 条件期望的定义

条件期望是现代概率论的基础.

回顾: 设 $(\Omega, \mathscr{F}, P)$ 是概率空间, 若 $A \in \mathscr{F}, P(A) > 0$, 定义 $P(B|A) = \dfrac{P(AB)}{P(A)}, \forall B \in \mathscr{F}$.

则 $P(\cdot|A) : \mathscr{F} \to [0,1]$ 是个概率测度. $\mathbb{E}(\xi|A) = \displaystyle\int_\Omega \xi dP(\cdot|A) = \dfrac{\mathbb{E}(\xi I_A)}{P(A)}$.

最简单的情况: 设 $\mathscr{C} = \{\Omega, \varnothing, A, A^c\}$, 满足 $P(A) > 0, P(A^c) > 0$, 则定义条件期望 $\mathbb{E}(\xi|\mathscr{C}) \triangleq \mathbb{E}(\xi|A)I_A + \mathbb{E}(\xi|A^c)I_{A^c}$, 注意 I_A, I_{A^c} 都是 r.v., 因此 $\mathbb{E}(\xi|\mathscr{C})$ 也是 r.v.

推广最简单的情况: 设 $\{A_n\}_{n\geq 1} \subset \mathscr{F}$ 是 Ω 的划分, 即满足 $\displaystyle\bigcup_n A_n = \Omega$ 且诸 A_i 两两不交, $P(A_n) > 0$, 记 $\mathscr{C} = \sigma(\{A_n\}_{n\geq 1})$ 为包含 $\{A_n\}$ 的最小 $\sigma-$ 代数, 定义 $\mathbb{E}(\xi|\mathscr{C}) \triangleq \displaystyle\sum_{n=1}^{\infty} \mathbb{E}(\xi|A_n)I_{A_n}$,

满足: ① $\mathbb{E}(\xi|\mathscr{C})$ 关于 $\mathscr{C}$ 可测, ② $\forall B \in \mathscr{C}$, $\displaystyle\int_B \mathbb{E}(\xi|\mathscr{C})dP = \int_B \xi dP$.(重要性质)

推广到更一般的 sigma 代数:

定义 6.1.1. 条件期望

设 $\mathscr{C} \subset \mathscr{F}$ 为 $\sigma-$ 代数, ξ 是可积 r.v., 称 (关于)$\mathscr{C}$ 可测的 r.v.η 为 ξ 关于 $\mathscr{C}$ 的条件期望, 若

$$\int_B \xi dP = \int_B \eta dP, \forall B \in \mathscr{C}.$$

此时记 $\eta = \mathbb{E}(\xi|\mathscr{C})$.

注: ①存在性: 令 $\nu(B) = \displaystyle\int_B \xi \mathrm{d}P, \forall B \in \mathscr{C}$, 则 $\nu \ll P|_{\mathscr{C}}$, 故存在唯一的关于 $\mathscr{C}$ 可测函数 g 使得 $\nu = g.P|_{\mathscr{C}}$, 即 $\nu(B) = \displaystyle\int_B g\mathrm{d}P|_{\mathscr{C}}$(Radon-Nikodym 定理). 所以 $\displaystyle\int_B \xi \mathrm{d}P = \int_B g\mathrm{d}P|_{\mathscr{C}} = \int_B g\mathrm{d}P$, 这里的 g 满足条件期望定义, g 是关于 P 的 R-N 导数.

②唯一性成立 (在 $P|_{\mathscr{C}}$ 几乎处处定义, 即 η, ξ 差一个关于 $\mathscr{C}$ 可测的零测集都成立)

③如果 $E\xi$ 存在, 则 ξ 关于 $\mathscr{C}$ 的 $\sigma-$ 代数的条件期望存在 (可测).

性质 6.1.1. 若 $\mathscr{C} = \{\varnothing, \Omega\}$, 则 $\mathbb{E}(\xi|\mathscr{C}) = \mathbb{E}\xi$, a.s..

性质 6.1.2. $\mathbb{E}(\mathbb{E}(\xi|\mathscr{C})) = E\xi$.

证明: 取 $B = \Omega$ 即可. □

性质 6.1.3. 若 ξ 关于 $\mathscr{C}$ 可测, 则 $\mathbb{E}(\xi|\mathscr{C}) = \xi$ a.s., 即上面的唯一性.

性质 6.1.4. 若 $c_1, c_2 \in \mathbb{R}$, 且 $c_1\xi_1 + c_2\xi_2, \xi_1, \xi_2$ 期望都存在, 则 (期望存在意味着期望的正部或者负部都存在)

$$\mathbb{E}(c_1\xi_1 + c_2\xi_2|\mathscr{C}) = c_1\mathbb{E}(\xi_1|\mathscr{C}) + c_2\mathbb{E}(\xi_2|\mathscr{C}), \mathrm{a.s.}$$

性质 6.1.5 (单调性). 若 $X \geq Y$ a.s.(关于 $\mathscr{F}$), 则 $\mathbb{E}(X|\mathscr{C}) \geq \mathbb{E}(Y|\mathscr{C})$ a.s.(关于 $P|_{\mathscr{C}}$).

证明: $\forall B \in \mathscr{C}, X \geq Y \Rightarrow \cdots \Rightarrow \displaystyle\int_B \mathbb{E}(X|\mathscr{C})dP \geq \int_B \mathbb{E}(Y|\mathscr{C})dP$. □

性质 6.1.6. $|\mathbb{E}(\xi|\mathscr{C})| \le \mathbb{E}(|\xi||\mathscr{C})$ a.s..

证明: 利用单调性.

性质 6.1.7. 设 $\{\xi_n\}$ 为非负 r.v. 列, 且 $\xi_n \le \xi_{n-1}$ a.s., 则 $\lim\limits_{n\to\infty} \mathbb{E}(\xi_n|\mathscr{C}) = E\left(\lim\limits_{n\to\infty}\xi_n \middle| \mathscr{C}\right)$, a.s..

证明: 利用单调收敛定理. □

性质 6.1.8 (硬性质, 不能忘)**.** 设 $\xi, \xi\eta$ 的期望均存在且 η 关于 $\mathscr{C}$ 可测, 则

$$\mathbb{E}(\xi\eta|\mathscr{C}) = \eta\mathbb{E}(\xi|\mathscr{C})\text{a.s.}$$

(让 $\xi = 1$ 可推出性质②)

证明: 证明思路: 证 r.v. 成立, 先证对示性函数成立, 再证对非负简单函数成立. 回顾 $\mathbb{E}f(\xi) = \int_\Omega f(x)dF(x)$ 的证明过程.

先设 $\eta = I_A, A \in \mathscr{C}$, 即证 $\mathbb{E}(\xi I_A|\mathscr{C}) = I_A\mathbb{E}(\xi|\mathscr{C})$ a.s. 成立. 由于 $I_A\mathbb{E}(\xi|\mathscr{C})$ 是可测的, (这是因为 I_A 与 $\mathbb{E}(\xi|\mathscr{C})$ 都关于 $\mathscr{C}$ 可测, 它们的乘积也可测),

$\forall B \in \mathscr{C}$, 由于 $\int_B \xi I_A dP = \int_{A\cap B}\xi dP = \int_{A\cap B}\mathbb{E}(\xi|\mathscr{C})dP = \int_B I_A\mathbb{E}(\xi|\mathscr{C})dP$,

根据性质②, 则 $\mathbb{E}(\xi I_A|\mathscr{C}) = I_A\mathbb{E}(\xi|\mathscr{C})$ a.s. 成立.

下面再证非负简单的情形 $\eta = \sum\limits_{i=1}^n a_i I_{A_i}, a_i \ge 0$, 且 $A_i \in \mathscr{C}$ 两两不交, 用线性性 (性质③) 可知这是成立的. 然后证 η 是非负且关于 $\mathscr{C}$ 可测成立, 用非负简单可测 r.v. 列 $\{\eta_n\}$ 逼近即可 (性质⑥). 对于一般的 η, 记为 $\eta = \eta^+ - \eta^-$(正部与负部) 就 OK 了. □

性质 6.1.9. 若 $\xi, \mathscr{C}$ 独立, 则 $\mathbb{E}(\xi|\mathscr{C}) = E\xi$, a.s..

证明: $\forall B \in \mathscr{C}$,

$$\begin{aligned}\int_B \mathbb{E}(\xi|\mathscr{C})dP = \mathbb{E}(\mathbb{E}(\xi|\mathscr{C})I_B) &= \mathbb{E}(\mathbb{E}(\xi I_B|\mathscr{C})) \qquad (\text{B 是可测的})\\ &= \mathbb{E}(\xi I_B) \qquad (\xi, I_B \text{ 独立})\\ &= E\xi EI_B = E\xi\int_B 1dP = \int_B E\xi dP.\end{aligned}$$

这里 $E\xi$ 是常数, 当然是可测的, 根据性质②, $\mathbb{E}(\xi|\mathscr{C}) = E\xi$. □

性质 6.1.10 (平滑性)**.** 设 $\mathscr{C}_1, \mathscr{C}_2 \subset \mathscr{F}$ 是 σ-代数, 且 $\mathscr{C}_1 \subset \mathscr{C}_2$, 则

$$\mathbb{E}(\xi|\mathscr{C}_1) = E[\mathbb{E}(\xi|\mathscr{C}_2)|\mathscr{C}_1], \text{a.s..}$$
$$\mathbb{E}(\xi|\mathscr{C}_1) = E[\mathbb{E}(\xi|\mathscr{C}_1)|\mathscr{C}_2], \text{a.s..}$$

证明: 只证第一条. $\forall A \in \mathscr{C}_1$, 根据定义有

$$\int_A E[\mathbb{E}(\xi|\mathscr{C}_2)|\mathscr{C}_1]\mathrm{d}P = \int_A \mathbb{E}(\xi|\mathscr{C}_2)\mathrm{d}P = \int_A \xi\mathrm{d}P = \int_A \mathbb{E}(\xi|\mathscr{C}_1)\mathrm{d}P.$$

注: 这个性质是很好的. 根据证明过程不难知道, 类似可以证明如果 $\mathscr{C}_2 \subset \mathscr{C}_1$, 则 $\mathbb{E}(\xi|\mathscr{C}_2) = E[\mathbb{E}(\xi|\mathscr{C}_2)|\mathscr{C}_1]$, a.s.. 因此让 ξ 对 $\mathscr{C}_1, \mathscr{C}_2$ 分别取条件期望 (无论顺序), 得到的一定是较小 sigma-代数的条件期望.

例 6.1.11

对 $\Omega=\{a,b,c\}$, 给一个例子使得 $\mathbb{E}(\mathbb{E}(X|\mathscr{C}_1)|\mathscr{C}_2)\neq\mathbb{E}(\mathbb{E}(X|\mathscr{C}_2)|\mathscr{C}_1)$.

下面把实变函数的部分定理推广到条件期望上来.

定理 6.1.12. 控制收敛定理

设 $\{X_n\}_{n\geq 1}$ 是 r.v. 序列, ξ 是可积 r.v.. 若 $X_n\xrightarrow{a.s.}X$ 或 $X_n\xrightarrow{P}X$, 且 $|X_n|\leq\xi,\forall n\geq 1$, a.s., 则 $\lim\limits_{n\to\infty}\mathbb{E}(X_n|\mathscr{C})=\mathbb{E}(\lim\limits_{n\to\infty}X_n|\mathscr{C})=\mathbb{E}(X|\mathscr{C})$.

定理 6.1.13. Fatou

设 $\{X_n\}$ 是 r.v. 序列, 且 $EX_n(n=1,2,\cdots)$ 存在.

(1) 若存在 r.v. Y, 使得 $EY>-\infty$, 且对每个 $n\geq 1$ 有 $X_n\geq Y$, a.s., 则 $\liminf\limits_{n\to\infty}X_n$ 的期望存在, 且满足

$$\mathbb{E}(\liminf_{n\to\infty}X_n|\mathscr{C})\leq\liminf_{n\to\infty}\mathbb{E}(X_n|\mathscr{C}).$$

(2) 若存在 r.v. Y, 使得 $EY<+\infty$, 且对每个 $n\geq 1$ 有 $X_n\leq Y$, a.s., 则 $\limsup\limits_{n\to\infty}X_n$ 的期望存在, 且满足

$$\mathbb{E}(\limsup_{n\to\infty}X_n|\mathscr{C})\geq\limsup_{n\to\infty}\mathbb{E}(X_n|\mathscr{C}).$$

定理 6.1.14. Hölder 不等式

$\forall p,q>1$ 满足 $\dfrac{1}{p}+\dfrac{1}{q}=1$, 则

$$\mathbb{E}(|\xi\eta||\mathscr{C})\leq\mathbb{E}\left(|\xi|^p|\mathscr{C}\right)^{\frac{1}{p}}\mathbb{E}\left(|\xi|^q|\mathscr{C}\right)^{\frac{1}{q}}.$$

注: 当 $p=q=2$ 时为 Cauchy-Schwarz 不等式.

定理 6.1.15. Jensen 不等式

设 $f:\mathbb{R}\to\mathbb{R}$ 是连续凸函数,r.v.ξ 满足 $f(\xi)$ 积分存在, 则

$$f(E\xi)\leq Ef(\xi),f(\mathbb{E}(\xi|\mathscr{C}))\leq\mathbb{E}(f(\xi)|\mathscr{C}),a.s..$$

注: 特别地取 $f(x)=|x|$ 显然成立 (便于记忆). 取 $f(x)=x^2$ 恰好是 Cauchy-Schwarz 不等式.

§ 6.2 条件期望的几何意义

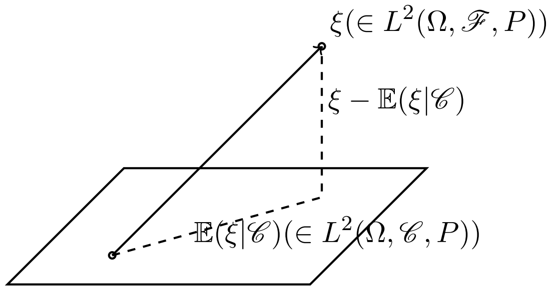
记

$$L^2(\Omega,\mathscr{F},P)=\{\xi:\xi\text{为}\mathscr{F}\text{中的可测 r.v., 且}\mathbb{E}\xi^2<+\infty\},$$
$$L^2(\Omega,\mathscr{C},P)=\{\xi:\xi\text{为}\mathscr{C}\text{中的可测 r.v., 且}\mathbb{E}\xi^2<+\infty\}.$$
$$\langle\xi,\eta\rangle_{L^2}=\mathbb{E}(\xi\eta).$$
$$\|\xi\|_{L^2}=\langle\xi,\xi\rangle_{L^2}^{\frac{1}{2}}=(E\xi^2)^{\frac{1}{2}}.$$

则 $L^2(\Omega,\mathscr{F},P)$ 和内积 $\langle\cdot,\cdot\rangle_{L^2}$ 一起构成 Hilbert 空间, $L^2(\Omega,\mathscr{C},P)$ 是 $L^2(\Omega,\mathscr{F},P)$ 的闭子空间 (要证一下), $\mathbb{E}(\cdot|\mathscr{C})$ 是 $L^2(\Omega,\mathscr{F},P)$ 到 $L^2(\Omega,\mathscr{C},P)$ 的正交投影算子.

定理 6.2.1

$\forall \xi \in L^2(\Omega, \mathscr{F}, P)$, 可对 ξ 做如下的正交分解: $\xi = \mathbb{E}(\xi|\mathscr{C}) + (\xi - \mathbb{E}(\xi|\mathscr{C}))$.



注: $\mathbb{E}(\xi|\mathscr{C}) \in L^2(\Omega, \mathscr{C}, P)$ 是因为 $\mathbb{E}(\xi|\mathscr{C})$ 可测, 且 $\mathbb{E}(\mathbb{E}(\xi|\mathscr{C})^2) \le \mathbb{E}(\mathbb{E}(\xi^2|\mathscr{C})) = E\xi^2 < +\infty$(Jensen 不等式).

证明: 先验证垂直, 再证距离最短.

$$\begin{aligned}
\langle \mathbb{E}(\xi|\mathscr{C}), \xi - \mathbb{E}(\xi|\mathscr{C}) \rangle_{L^2} =& \mathbb{E}\left[\mathbb{E}(\xi|\mathscr{C}) \cdot (\xi - \mathbb{E}(\xi|\mathscr{C}))\right] \\
=& \mathbb{E}\left\{\mathbb{E}\left[\mathbb{E}(\xi|\mathscr{C}) \cdot (\xi - \mathbb{E}(\xi|\mathscr{C}))|\mathscr{C}\right]\right\} && \text{【性质①】} \\
=& \mathbb{E}\left\{\mathbb{E}(\xi|\mathscr{C})\mathbb{E}\left[\xi - \mathbb{E}(\xi|\mathscr{C})|\mathscr{C}\right]\right\} && \text{【}\mathbb{E}(\xi|\mathscr{C})\text{ 可测, 性质⑦】} \\
=& \mathbb{E}\left\{\mathbb{E}(\xi|\mathscr{C})[\mathbb{E}(\xi|\mathscr{C}) - \mathbb{E}(\xi|\mathscr{C})]\right\} && \text{【性质③ (线性性), 性质②】} \\
=& 0.
\end{aligned}$$

下面验证距离最短: 即验证

$$\mathbb{E}(\xi - \mathbb{E}(\xi|\mathscr{C}))^2 = \inf_{y \in L^2(\Omega, \mathscr{C}, P)} \{\mathbb{E}(\xi - y)^2 | y \in L^2(\Omega, \mathscr{C}, P)\}.$$

事实上, $\forall y \in L^2(\Omega, \mathscr{C}, P)$,

$$\begin{aligned}
\mathbb{E}(\xi - \mathbb{E}(\xi|\mathscr{C}))^2 &= \mathbb{E}(\xi - \mathbb{E}(\xi|\mathscr{C}) - y + y)^2 \\
&= \mathbb{E}(\xi - y)^2 - 2\mathbb{E}[(\xi - y)(\mathbb{E}(\xi|\mathscr{C}) - y)] + E[\mathbb{E}(\xi|\mathscr{C}) - y]^2 \\
&= \mathbb{E}(\xi - y)^2 - 2\mathbb{E}\{\mathbb{E}[(\xi - y)(\mathbb{E}(\xi|\mathscr{C}) - y)|\mathscr{C}]\} + E[\mathbb{E}(\xi|\mathscr{C}) - y]^2 \quad \text{【性质①】} \\
&= \mathbb{E}(\xi - y)^2 - E[\mathbb{E}(\xi|\mathscr{C}) - y]^2 \quad \text{【}\mathbb{E}(\xi|\mathscr{C}) - y\text{ 可测, 性质⑦】} \\
&\le \mathbb{E}(\xi - y)^2.
\end{aligned}$$

根据 y 的任意性, 对上式取下确界即可. □

定理 6.2.2

设 $(\Omega, \mathscr{F}, P)$ 是概率空间, $\mathscr{C} \subset \mathscr{F}$ 是 $\sigma-$ 代数. $(G, \mathscr{G}), (E, \mathscr{E})$ 是可测空间. 令 $X : \Omega \to G$ 是 $\mathscr{C}/\mathscr{G}$ 可测, $Y : \Omega \to E$ 是 $\mathscr{F}/\mathscr{E}$ 可测. 若 Y 与 $\mathscr{C}$ 独立, $g : G \times E \to \mathbb{R}$ 可测, 且 $\mathbb{E}|g(X, Y)| < +\infty$, 则

$$\mathbb{E}[g(X, Y)|\mathscr{C}] = \mathbb{E}(g(x, Y))|_{x=X}$$

证明: 设 $f(x) = \mathbb{E}g(x, Y), x \in G$, 下证 $\mathbb{E}(g(X, Y)|\mathscr{G}) = f(X)$, a.s.. 只需证 $f(X)$ 关于 $\mathscr{C}$ 可测 (即 f 为 $\mathscr{G}$-可测), 且 $\int_A f(X)\mathrm{d}P = \int_A g(X, Y)\mathrm{d}P, \forall A \in \mathscr{C}$.

(1) 下证 f 为 $\mathscr{G}$-可测: 等价于 $\forall B \in \mathscr{B}(\mathbb{R}), f^{-1}(B) \in \mathscr{G}$. 注意 $\mathscr{G} \times \mathscr{E} = \sigma(A \times B | A \in \mathscr{G}, B \in \mathscr{C})$, 所以只需证矩形集的情形, 再用单调类定理即可. (步骤略). 所以 $f(X)$ 是 $\mathscr{C}$-可测.

(2) 下证对任一 $\mathscr{C}$-可测 r.v.Z, 有 $\mathbb{E}(g(X,Y)Z)=\mathbb{E}(f(X)Z)$. (等价于 $\int_\Omega f(X)\mathrm{d}P=\int_\Omega g(X,Y)\mathrm{d}P$.) 注意 Z 有界且 $g(X,Y)$ 可积, 所以 $\mathbb{E}(g(X,Y)Z)$ 积分存在, 且

$$\begin{aligned}\mathbb{E}(g(X,Y)Z)&=\int_{G\times E\times\mathbb{R}}g(x,y)z\mathrm{d}F_{(X,Y,Z)}(x,y,z)\\&=\int_{G\times\mathbb{R}}\int_E g(x,y)z\mathrm{d}F_Y(y)\mathrm{d}F_{(X,Z)}(x,z)\qquad\text{(独立性)}\\&=\int_{G\times\mathbb{R}}\int_E g(x,y)\mathrm{d}F_Y(y)z\mathrm{d}F_{(X,Z)}(x,z)\qquad\text{(Fubini)}\\&=\int_{G\times\mathbb{R}}f(x)z\mathrm{d}F_{(X,Z)}(x,z)=\mathbb{E}(f(X)Z).\qquad\square\end{aligned}$$

注: 取 $g(x,y)=g_1(x)g_2(y)$, 则

$$\mathbb{E}[g_1(X)g_2(Y)|\mathscr{C}]=g_1(X)\mathbb{E}[g_2(Y)|\mathscr{C}]=g_1(X)\mathbb{E}g_2(Y)=g_1(x)\mathbb{E}g_2(Y)|_{x=X},$$

所以这个定理推广了条件期望的两条性质.

定理 6.2.3

设 $(\Omega,\mathscr{F},P)$ 是概率空间, Q 是概率测度, 且 $Q\ll P$. 设 $\mathbb{C}$ 是 $\mathscr{F}$ 的子 σ-代数, 则:

(1)$Q\left(\mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)>0\right)=1$.

(2) 若 X 关于 Q 期望存在, 则 $\mathbb{E}_Q(X|\mathscr{C})=\dfrac{\mathbb{E}\left(X\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)}{\mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)}$. (默认 $\mathbb{E}$ 是关于 P 积分的期望.)

证明: (1) 注意到[1]

$$\begin{aligned}Q\left(\mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)>0\right)&=\int_\Omega I_{[\mathbb{E}(\frac{\mathrm{d}Q}{\mathrm{d}P}|\mathscr{C})>0]}\underbrace{\frac{\mathrm{d}Q}{\mathrm{d}P}}_{\text{是}r.v.}\mathrm{d}P=\int_\Omega I_{[\mathbb{E}(\frac{\mathrm{d}Q}{\mathrm{d}P}|\mathscr{C})>0]}\mathbb{E}(\frac{\mathrm{d}Q}{\mathrm{d}P}|\mathscr{C})\mathrm{d}P\\&=\int_\Omega\mathbb{E}(\frac{\mathrm{d}Q}{\mathrm{d}P}|\mathscr{C})\mathrm{d}P=\int_\Omega\frac{\mathrm{d}Q}{\mathrm{d}P}\mathrm{d}P=Q(\Omega)=1.\end{aligned}$$

(2) 下证 $\mathbb{E}_Q(X|\mathscr{C})\mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)=\mathbb{E}\left(X\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)$. 由条件期望的定义, 即证

$$\int_B\mathbb{E}_Q(X|\mathscr{C})\mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)\mathrm{d}P=\int_B\mathbb{E}\left(X\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)\mathrm{d}P,\forall B\in\mathscr{C}.$$

---

[1]若 μ,ν 是测度, 则 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\geq0$, μ-a.s.. 要不然, 记 $A=\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}<0\right]$, 若 $\mu(A)>0$, 则 $\nu(A)=\int_A\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu<0$, 这与 ν 非负矛盾.

事实上,

$$
\begin{aligned}
\text{LHS} &= \int_\Omega \underbrace{I_B \mathbb{E}_Q(X|\mathscr{C})}_{\mathscr{C}\text{ 可测}} \mathbb{E}\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)\mathrm{d}P = \int_\Omega I_B \mathbb{E}_Q(X|\mathscr{C})\frac{\mathrm{d}Q}{\mathrm{d}P}\mathrm{d}P \quad (\text{性质 } 6.1.8)\\
&= \int_\Omega \mathbb{E}_Q(X|\mathscr{C}) I_B \mathrm{d}Q = \int_\Omega X I_B \mathrm{d}Q \quad (\text{性质 } 6.1.8)\\
&= \int_B X\frac{\mathrm{d}Q}{\mathrm{d}P}\mathrm{d}P = \int_B \mathbb{E}\left(X\frac{\mathrm{d}Q}{\mathrm{d}P}\Big|\mathscr{C}\right)\mathrm{d}P \quad (\text{条件期望的定义})
\end{aligned}
$$

□

下设 X 积分存在, Y 是 r.v., 则 $\mathbb{E}(X|Y)$ 关于 $\sigma(Y)$ 可测. 由定理??, 存在可测函数 $h:\mathbb{R}\to\mathbb{R}$ 使得 $\mathbb{E}(X|Y)=h(Y)$. 若 h_1,h_2 满足此 h, 则 $P\circ Y^{-1}(h_1\neq h_2)=0$(条件期望的唯一性), 所以 h 仅在 Y 处 a.s. 唯一.

设 $\mu(A)=P\circ Y^{-1}(A), \nu(A)=\mathbb{E}[\xi I_{Y^{-1}(A)}]$, 其中 $A\in\mathscr{B}(\mathbb{R})$, 则 ν 是符号测度且 $\nu\ll\mu$, 从而 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ 存在且 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}=h(\mu\text{-a.e.}) \Leftrightarrow \forall A\in\mathscr{B}(\mathbb{R})$, 有

$$
\begin{aligned}
\nu(A) &= \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(y)\mu(\mathrm{d}y) = \int_A h(y)\mu(\mathrm{d}y) = \int_{\mathbb{R}} h(y) I_A(y)\mu(\mathrm{d}y)\\
&= \mathbb{E}(h(Y)I_A(Y)) = \mathbb{E}(\mathbb{E}(X|Y)I_A(Y)) = \mathbb{E}XI_A(Y).
\end{aligned}
$$

注意 $\sigma(Y)=\{Y^{-1}(A)|A\in\mathscr{B}(\mathbb{R})\}$, 则 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}=\mathbb{E}(X|Y)=h$, a.e.. (差 $\sigma(Y)$ 中的零测集)

§ 6.3 条件独立

定义 6.3.1

设 $(\Omega,\mathscr{F},P)$ 是概率空间, $\mathscr{C}$ 是子 σ-代数.

(1) 称 $P(B|\mathscr{C})\triangleq\mathbb{E}(I_B|\mathscr{C})$ 是 B 关于 $\mathscr{C}$ 的**条件概率**.

(2) 设 $A,B\in\mathscr{F}$. 称 A 与 B 关于 $\mathscr{C}$ **条件独立**, 若 $P(AB|\mathscr{C})=P(A|\mathscr{C})P(B|\mathscr{C})$.

(3) 设 $\mathscr{C}_1,\mathscr{C}_2\subset\mathscr{F}$. 称 $\mathscr{C}_1,\mathscr{C}_2$ 关于 $\mathscr{C}$ **条件独立**, 若 $P(AB|\mathscr{C})=P(A|\mathscr{C})P(B|\mathscr{C}), \forall A\in\mathscr{C}_1, B\in\mathscr{C}_2$.

命题 6.3.1

设 $\mathscr{C}_1,\mathscr{C}_2$ 是 $\mathscr{F}$ 的子 σ-代数, 则 $\mathscr{C}_1,\mathscr{C}_2$ 关于 $\mathscr{C}$ 条件独立等价于对任意 $B_2\in\mathscr{C}_2$, 有

$$
P(B_2|\mathscr{C})=P(B_2|\mathscr{C}_1\vee\mathscr{C}), a.s.. \tag{6.1}
$$

注: (1) 类比下述事实：若 A,B 独立, 则 $P(A|B)=P(A)$. (2) 可以用单调类定理证明 $\mathscr{C}_1\vee\mathscr{C}\triangleq\sigma(\mathscr{C}_1\cup\mathscr{C})=\sigma(\{A\cap B|A\in\mathscr{C}_1,B\in\mathscr{C}\})$.

证明: (6.1) 式等价于对任意 $B\in\mathscr{C},B_1\in\mathscr{C}_1$, 有 $\int_{B\cap B_1}P(B_2|\mathscr{C})\mathrm{d}P=\int_B I_{B_1}I_{B_2}\mathrm{d}P$.

而 $\mathscr{C}_1,\mathscr{C}_2$ 关于 $\mathscr{C}$ 条件独立 $\Leftrightarrow \forall B_1\in\mathscr{C}_1,B_2\in\mathscr{C}_2$, $\int_B P(B_1|\mathscr{C})P(B_2|\mathscr{C})\mathrm{d}P=\int_B I_{B_1\cap B_2}\mathrm{d}P, B\in\mathscr{C}$.

但是对 $B \in \mathscr{C}, B_1 \in \mathscr{C}_1, B_2 \in \mathscr{C}_2$, 有

$$\begin{aligned}\int_{B\cap B_1} P(B_2|\mathscr{C})\mathrm{d}P &= \int_B I_{B_1}P(B_1|\mathscr{C})\mathrm{d}P\\ &= \int_B \mathbb{E}(I_{B_1}P(B_1|\mathscr{C})|\mathscr{C})\mathrm{d}P = \int_B P(B_1|\mathscr{C})P(B_2|\mathscr{C})\mathrm{d}P\end{aligned}$$

所以欲证命题成立. □

例 6.3.2

设 $(\Omega, \mathscr{F}, P)$ 为概率空间, $\mathscr{C} \subset \mathscr{F}$ 为子 σ-代数, $\Gamma \in \mathscr{F}$ 是事件, 证明以下等价:

(1)$\Gamma, \mathscr{C}$ 独立;

(2) 任一概率测度 Q on $(\Omega, \mathscr{F})$, Q 与 P 等价, 且 $\dfrac{dQ}{dP}$ 为 $\mathscr{C}$ 可测, 则 $Q(\Gamma) = P(\Gamma)$.

证明: “(1)⇒(2)”: $Q(A) = \displaystyle\int_A \frac{dQ}{dP}(w)P(dw)$, 这里 $\dfrac{dQ}{dP}(w)$ 是个 r.v.. 记 E_Q 是关于 Q 积分的期望, E_P 是关于 P 积分的期望.

$\Gamma, \mathscr{C}$ 独立 $\Leftrightarrow P(\Gamma \cap B) = P(\Gamma)P(B), \forall B \in \mathscr{C}$, 而

$$\begin{aligned}Q(\Gamma) &= \int_\Gamma dQ = \int_\Omega I_\Gamma dQ = \int_\Omega I_\Gamma \frac{dQ}{dP}dP && \text{【形式上的转化】}\\ &= E_P\left(I_\Gamma \frac{dQ}{dP}\right)\\ &= E_P(I_\Gamma)E_P\left(\frac{dQ}{dP}\right) && \text{【独立性】}\\ &= P(\Gamma)\int_\Omega \frac{dQ}{dP}dP = P(\Gamma)\int_\Omega dQ = P(\Gamma).\end{aligned}$$

“(1)⇐(2)”: 要证 $P(\Gamma \cap B) = P(\Gamma)P(B), \forall B \in \mathscr{C}$. 事实上, 令 $\boxed{dQ = \dfrac{I_B + 1}{P(B)+1}dP}$(+1 为了保证分子分母不为 0, 除以 $(P(B)+1)$ 这一常数是为了归一化). 下面验证 Q 是概率测度: 根据定义验证.

1° 非负性: $Q(A) = \displaystyle\int_A \frac{I_B+1}{P(B)+1} = \int_\Omega I_A \frac{I_B+1}{P(B)+1}dP \geq 0$, 这里 $I_A \geq 0, \dfrac{I_B+1}{P(B)+1} > 0$.

2° 可列可加性: $\forall \{A_n\}_{n\geq 1} \subset \mathscr{F}, A_n \cap A_m = \varnothing, \forall n \neq m$, 我们有:

$$\begin{aligned}Q\left(\sum_n A_n\right) &= \int_{\sum_n A_n} \frac{I_B+1}{P(B)+1}dP\\ &= \int_\Omega I_{\sum_n A_n}\frac{I_B+1}{P(B)+1}dP\\ &= \int_\Omega \sum_n \left(I_{A_n}\frac{I_B+1}{P(B)+1}\right)dP && \text{【两两不交】}\\ &= \sum_n \int_{A_n}\frac{I_B+1}{P(B)+1}dP && \text{【非负, 积分与求和可调换次序】}\\ &= \sum_n Q(A_n).\end{aligned}$$

3° 规范性: $Q(\Omega) = \displaystyle\int_\Omega \frac{I_B+1}{P(B)+1}dP = \frac{1}{P(B)+1}(P(B)+1) = 1.$

根据 Radon-Nikodym 定理, 因此 $\dfrac{dQ}{dP}$ 是 $\mathscr{C}$ 可测的, 根据条件,

$$\begin{aligned}Q(\Gamma)=\int_\Gamma\frac{I_B+1}{P(B)+1}dP=P(\Gamma)\Longleftrightarrow&\frac{\int_\Omega I_\Gamma(I_B+1)dP}{P(B)+1}=P(\Gamma)\\ \Longleftrightarrow&\int_\Omega I_{\Gamma\cap B}dP+P(\Gamma)=P(\Gamma)(P(B)+1)\\ \Longleftrightarrow&P(\Gamma\cap B)=P(\Gamma)P(B),\forall B\in\mathscr{C}.\end{aligned}$$

证完. □

例 6.3.3

设 X,Y,Z 是 r.v. 且 Y 可积, 证明若 (X,Y) 与 Z 独立, 则 $\mathbb{E}(Y|X,Z)=\mathbb{E}(Y|X)$.

注: $\mathbb{E}(Y|X_1,X_2)$ 表示关于由 X_1,X_2 生成的 sigma-代数 $\sigma(X_1,X_2)=\sigma(\sigma(X_1)\cup\sigma(X_2))$ 的条件期望. (X,Y) 与 Z 独立指 $\sigma(Z),\sigma(X,Y)$ 独立.

证明: 只需证

$$\begin{aligned}&\int_A YdP=\int_A\mathbb{E}(Y|X)dP,\forall A\in\sigma(X,Z)\\ \Longleftrightarrow&\int_A YdP=\int_A\mathbb{E}(Y|X)dP,\forall A\in\sigma(X)\cup\sigma(Z)\ \text{【单调类定理】}\\ &\text{【不需对所有都进行验证, 只需要看子类,}\\ &\qquad\sigma(X)\cup\sigma(Z)=\{X^{-1}(B),Z^{-1}(C):B,C\in\mathscr{B}(\mathbb{R})\}\text{】}\\ \Longleftrightarrow&\int_\Omega I_{X^{-1}(B)}I_{Z^{-1}(C)}YdP=\int_\Omega\mathbb{E}(Y|X)I_{X^{-1}(B)}I_{Z^{-1}(C)}dP,\forall B,C\in\mathscr{B}(\mathbb{R})\\ &\text{【}I_{X^{-1}(B)}\text{ 即 }I_B(X)\text{】}\end{aligned}$$

事实上,

$$\begin{aligned}\int_\Omega I_B(X)I_C(Z)YdP=&P(Z\in C)\int_\Omega I_B(X)YdP &&\text{【}X,Z\text{ 独立】}\\ =&P(Z\in C)\mathbb{E}(I_B(X)Y)\\ =&P(Z\in C)\mathbb{E}(\mathbb{E}(I_B(X)Y|X)) &&\text{【条件期望的期望 = 无条件期望】}\\ =&EI_C(Z)\mathbb{E}(I_B(X)\mathbb{E}(Y|X)) &&\text{【}I_B(X)\text{ 关于 }X\text{ 可测】}\\ =&\mathbb{E}(I_B(X)\mathbb{E}(Y|X)I_C(Z)) &&\text{【独立性】}\\ =&\int_\Omega\mathbb{E}(Y|X)I_B(X)I_C(Z)dP.\quad\square\end{aligned}$$

例 6.3.4

设一列 r.v.$\{X_n\}$ 依分布收敛于一个 r.v. X, 记 $\{N_t\}_{t\geq0}$ 是一列正整数 r.v. 集合, 与 $\{X_n\}$ 独立且依概率收敛为 $\infty(t\to\infty)$. 证明: $X_{N_t}\xrightarrow{d}X,(t\to\infty)$.

证明: 固定 $c\in\mathbb{R}$, 记 $a_n=Ee^{icX_n},a=Ee^{icx}$, 由于 $X_n\xrightarrow{d}X,(t\to\infty)$, 则 $a_n\to a(n\to\infty)$. (依概率收敛与特征函数收敛是一一对应的!)

即 $\forall\varepsilon>0,\exists M\in\mathbb{N}$, 使得当 $n\geq M$ 时, 有 $|a_n-a|\leq\varepsilon$. 因此

$$[|a_{N_t}-a|\leq\varepsilon]\supset[N_t\geq M],\Longleftrightarrow[|a_{N_t}-a|>\varepsilon]\subset[N_t<M].$$

则 $P(|a_{N_t}-a|>\varepsilon)\leq P(N_t<M)\to 0(t\to\infty)$, 则 $a_{N_t}\xrightarrow{d}a,(t\to\infty)$.

下面用条件期望: 注意到 (把 a_{N_t} 看作关于 r.v.N_t 的随机函数. 根据后面"注"的定理,

$$\begin{aligned}Ea_{N_t}&=E[\mathbb{E}(a_{N_t}|N_t)] &&【取条件期望】\\&=E[(Ea_n)|_{n=N_t}]=Ea_{N_t} &&【“注”的定理】\end{aligned}\tag{6.2}$$

则 $Ea_{N_t}\to a=Ea$. 又由于 $|a_{N_t}|\leq 1$, 根据控制收敛定理, $a_{N_t}\xrightarrow{L^1}a$, 特别地 $Ea_{N_t}\to a(t\to\infty)$, 则 $X_{N_t}\xrightarrow{d}X$.

§ 6.4 随机变量族的一致可积性

定义 6.4.1

设 $(\Omega,\mathscr{F},P)$ 是概率空间, $\mathcal{H}$ 是一族随机变量. 若

$$\lim_{c\to\infty}\sup_{\xi\in\mathcal{H}}\mathbb{E}(|\xi|I_{|\xi|\geq c})=0,$$

则称 $\mathcal{H}$ 是一致可积的.

例 6.4.1

(1) 若存在 $\eta\in L^1(\Omega,\mathscr{F},P)$, 使得 $|\xi|\leq\eta,\forall\xi\in\mathcal{H}$, 则 $\mathcal{H}$ 是一致可积的.

(2) 若 $\{\xi_1,\cdots,\xi_n\}\subset L^1(\Omega,\mathscr{F},P)$, 则 $\{\xi_1,\cdots,\xi_n\}$ 是一致可积的.

(3) 若 $\mathcal{H}_1$ 是一致可积族, $\mathcal{H}_2$ 满足对任意 $\xi\in\mathcal{H}_2$, 存在 $\eta\in\mathcal{H}_1$, 使得 $|\xi|\leq|\eta|$, 则 $\mathcal{H}_2$ 是一致可积族.

下面的定理像 Ascoli-Arzela 定理.

定理 6.4.2

设 $\mathcal{H}\subset L^1(\Omega,\mathscr{F},P)$, 则

$$\mathcal{H}\text{是一致可积族}\Leftrightarrow\begin{cases}L^1\text{有界: }\sup\limits_{\xi\in\mathcal{H}}\mathbb{E}|\xi|<\infty,\\ \text{一致绝对连续: }\forall\varepsilon>0,\exists\delta>0,\text{若}P(A)\leq\delta,\text{则}\sup\limits_{\xi\in\mathcal{H}}\displaystyle\int_A|\xi|\mathrm{d}P\leq\varepsilon.\end{cases}$$

证明: "$\Rightarrow$": (1)$\forall\varepsilon>0$, 存在 C_0, 当 $C>C_0$ 时, $\sup\limits_{\xi\in\mathcal{H}}\int_{[|\xi|\geq C]}|\xi|\mathrm{d}P\leq\frac{\varepsilon}{2}$. 所以

$$\sup_{\xi\in\mathcal{H}}\mathbb{E}|\xi|\leq\sup_{\xi\in\mathcal{H}}\int_{[|\xi|\geq C]}|\xi|\mathrm{d}P+\sup_{\xi\in\mathcal{H}}\int_{[|\xi|<C]}|\xi|\mathrm{d}P\leq\frac{\varepsilon}{2}+C.$$

(2) 取 $\delta=\frac{\varepsilon}{2C}$, 则当 $P(A)<\delta$ 时,

$$\int_A|\xi|\mathrm{d}P\leq\int_{[|\xi|<C]\cap A}|\xi|\mathrm{d}P+\int_{[|\xi|\geq C]\cap A}|\xi|\mathrm{d}P\leq CP(A)+\frac{\varepsilon}{2}<\varepsilon.$$

“$\Leftarrow$”: $\forall \varepsilon>0, \exists \delta>0$, 当 $C \geq \frac{1}{\delta} \sup\limits_{\xi \in \mathcal{H}} \mathbb{E}|\xi|$ 时,

$$P(|\xi| \geq C) \leq \frac{\mathbb{E}|\xi|}{C} \leq \frac{a}{C} \leq \delta.$$

所以 $\sup\limits_{\xi \in \mathcal{H}} \int_{[|\xi| \geq C]}|\xi| \mathrm{d} P \leq \varepsilon$, 则 $\mathcal{H}$ 一致可积. □

推论 6.4.3

设 $\eta \in L^1(\Omega, \mathscr{F})$ 是可积随机变量, $\mathcal{H}$ 是一致可积族, 则 $\eta+\mathscr{H}=\{\xi+\eta | \xi \in \mathcal{H}\}$ 是一致可积的.

推论 6.4.4

若 $\mathcal{H} \subset L^1(\Omega, \mathscr{F}, P)$ 是一致可积族, 则 $\mathcal{H}$ 在 $L^1(\Omega, \mathscr{F}, P)$ 中的闭凸包也一致可积.

定理 6.4.5. L^1 收敛准则

设 $\{\xi_n\} \subset L^1(\Omega, \mathscr{F}, P)$, TFAE:

(1)$\xi_n \xrightarrow{L^1} \xi$;

(2)$\xi_n \xrightarrow{P} \xi$, 且 $\{\xi_n\}$ 一致可积;

(3)$\xi_n \xrightarrow{P} \xi$, 且 $\mathbb{E}|\xi_n| \rightarrow \mathbb{E}|\xi|<\infty$.

证明: “$(1) \Leftrightarrow (3)$” 以前证过了. 下面证明 “$(1) \Leftrightarrow (2)$”.

“$\Rightarrow$”: 设 $\xi_n \xrightarrow{L^1} \xi$, 即 $\mathbb{E}|\xi_n-\xi| \rightarrow 0$. 令 $A \in \mathscr{F}$, 则

$$\int_A|\xi_n| \mathrm{d} P \leq \int_A|\xi| \mathrm{d} P+\mathbb{E}|\xi_n-\xi|,$$

给定 $\varepsilon>0$, 取正数 N, 使得当 $n>N$ 时, $\mathbb{E}|\xi_n-\xi| \leq \frac{\varepsilon}{2}$ (处理无限个),

再选取 $\delta>0$, 使得对任何满足 $P(A) \leq \delta$ 的 $A \in \mathscr{F}$, 有

$$\int_A|\xi| \mathrm{d} P \leq \frac{\varepsilon}{2}, \int_A|\xi_n| \mathrm{d} P \leq \frac{\varepsilon}{2}, n=1,2, \cdots, N.$$

(处理有限个). 所以对任何满足 $P(A) \leq \delta$ 的 $A \in \mathscr{F}$, 有 $\sup\limits_n \int_A|\xi_n| \mathrm{d} P \leq \varepsilon$. 此外有 $\sup\limits_n \mathbb{E}|\xi_n|<\infty$. 所以由定理 6.4.2, $\{\xi_n\}$ 是一致可积族. 最后显然有 $\xi_n \xrightarrow{P} \xi$.

“$\Leftarrow$”: 设 $\{\xi_n\}$ 一致可积, 且 $\xi_n \xrightarrow{P} \xi$. 由 Fatou 引理与定理 6.4.2,

$$\mathbb{E}|\xi| \leq \sup_n \mathbb{E}|\xi_n|<+\infty,$$

所以 ξ 可积, 所以 $\xi_n-\xi$ 一致可积.

下证 $\mathbb{E}|\xi_n-\xi| \rightarrow 0$. 对任何 $\varepsilon>0$, 由定理 6.4.2, 存在 $\delta>0$, 使得对任意满足 $P(A)<\delta$ 的 $A \in \mathscr{F}$, 有

$$\sup_n \int_A|\xi_n-\xi| \mathrm{d} P \leq \varepsilon.$$

由于 $\xi_n \xrightarrow{P} \xi$, 可以取 N 充分大使得当 $n \geq N$ 时 $P([|\xi_n-\xi| \geq \varepsilon])<\delta$, 故当 $n \geq N$ 时, 有

$$\mathbb{E}|\xi_n-\xi|=\int_{[|\xi_n-\xi|<\varepsilon]}|\xi_n-\xi| \mathrm{d} P+\int_{[|\xi_n-\xi| \geq \varepsilon]}|\xi_n-\xi| \mathrm{d} P \leq 2 \varepsilon.$$

所以 $\xi_n \xrightarrow{L^1} \xi$. □

把前一定理的 L^1 改为 L^p, 会有如下结果:

定理 6.4.6

设 $(\Omega, \mathscr{F}, P)$ 是概率空间, $0 < p < \infty$. 若 $\xi_n \xrightarrow{P} \xi$ 且 $\{|\xi_n|^p\}$ 一致可积, 则 $\xi_n \xrightarrow{L^p} \xi$.

证明: 只需注意不等式

$$|a-b|^p \leq \gamma_p(|a|^p + |b|^p), \text{其中}\gamma_p = \max\{1, 2^{p-1}\}.$$

来推 $|\xi_n - \xi|^p$ 一致可积. 其他步骤与前一定理类似. 【待补充】

定理 6.4.7

设 $\mathcal{H} \subset L^1(\Omega, \mathscr{F}, P)$, TFAE:

(1)$\mathcal{H}$ 是一致可积的;

(2) 存在函数 $\varphi : [0, +\infty) \to [0, +\infty)$, 满足 $\lim\limits_{t\to\infty} \dfrac{\varphi(t)}{t} = \infty$, 且 $\sup\limits_{\xi\in\mathcal{H}} \mathbb{E}(\varphi \circ |\xi|) < \infty$.

证明: “(1) ⇒ (2)” : 设 $\mathcal{H}$ 是一致可积族, 由于对任何 $a > 0$, 有

$$\int_\Omega (|\xi| - a)^+ \mathrm{d}P \leq \int_{[|\xi|>a]} (|\xi| - a)^+ \mathrm{d}P \leq \int_{[|\xi|>a]} |\xi| \mathrm{d}P \to 0 (a \to \infty),$$

故存在自然数 $n_k \nearrow \infty$, 使得

$$\sup_{\xi\in\mathcal{H}} \int_\Omega (|\xi| - n_k)^+ \mathrm{d}P < 2^{-k}, k \geq 1.$$

令

$$\varphi(t) = \sum_{k\geq1} ([t] - n_k)^+, t \in [0, +\infty)$$

则 φ 非负、单调非降且右连续, 而且由 Fatou 引理,

$$\lim_{n\to\infty} \frac{\varphi(n)}{n} = \lim_{n\to\infty} \sum_{k\geq1} \left(1 - \frac{n_k}{n}\right)^+ \geq \sum_{k\geq1} \inf_{n\to\infty} \left(1 - \frac{n_k}{n}\right)^+ = \sum_{k\geq1} 1 = \infty.$$

所以 $\lim\limits_{t\to\infty} \dfrac{\varphi(t)}{t} = \infty$. 最后有

$$\begin{aligned}
\mathbb{E}(\varphi \circ |\xi|) &= \int_\Omega \varphi(|\xi|) \mathrm{d}P = \sum_{n=0}^{\infty} \int_{n\leq|\xi|<n+1} \varphi(|\xi|) \mathrm{d}P \\
&= \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} (n - n_k)^+ P(n \leq |\xi| < n+1) \\
&= \sum_{k=1}^{\infty} \sum_{n=0}^{\infty} (n - n_k)^+ P(n \leq |\xi| < n+1) \qquad \text{(非负可换序)} \\
&\leq \sum_{k=1}^{\infty} \int_\Omega (|\xi| - n_k)^+ \mathrm{d}P < 1.
\end{aligned}$$

“(2) ⇒ (1)”：设 (2) 成立, 对 $\varepsilon > 0$, 令 $a = M/\varepsilon$, 其中 $M = \sup\limits_{\xi\in\mathcal{H}} \mathbb{E}(\varphi(|\xi|))$. 取充分大的 C 使得当 $t \geq C$ 时, $\dfrac{\varphi(t)}{t} \geq a$. 则在 $[|\xi| \geq C]$ 上, 有 $|\xi| \leq \dfrac{\varphi\circ|\xi|}{a}$, 故

$$\int_{[|\xi|\geq C]} |\xi|\mathrm{d}P \leq \frac{1}{a}\int_{[|\xi|\geq C]} \varphi\circ|\xi|\mathrm{d}P \leq \frac{M}{a} = \varepsilon, \xi\in\mathcal{H}.$$

所以 $\mathcal{H}$ 是一致可积族.

推论 6.4.8

设 $\mathcal{H}\subset L^p(\Omega,\mathscr{F},P), p>1$. 如果 $\sup\limits_{\xi\in\mathcal{H}}\mathbb{E}|\xi|^p<\infty$, 则 $\mathcal{H}$ 是一致可积族.

注: 注意对比定理 6.4.2的条件, 这里少了积分一致绝对连续的条件.

证明: 【方法一】定理 6.4.7中让 $\varphi(t) = t^p$ 立得.

【方法二】设 $a = \sup\limits_{\xi\in\mathcal{H}}\mathbb{E}|\xi|^p$, 则对任意 $C>0$, 有

$$\int_{[|\xi|>C]} |\xi|\mathrm{d}P \leq \int_{[|\xi|>C]} \frac{|\xi|^p}{C^{p-1}}\mathrm{d}P \leq \frac{1}{C^{p-1}}\mathbb{E}|\xi|^p \leq \frac{a}{C^{p-1}} \to 0 (C\to\infty).$$

所以由定义可知 $\mathcal{H}$ 是一致可积族. □

定理 6.4.9

设 $(\Omega,\mathscr{F},P)$ 是概率空间, ξ 是可积随机变量, $\{\mathscr{G}_i\}_{i\in I}$ 是 $\mathscr{F}$ 的一族子 σ-代数. 令 $\eta_i = \mathbb{E}(\xi|\mathscr{G}_i)$, 则 $\{\eta_i\}_{i\in I}$ 是一致可积族.

证明: 注意用条件期望的定义以及“条件期望的期望等于无条件期望”即可.

对任何 $C>0$, 有

$$P(|\eta_i|\geq C) \leq \frac{1}{C}\mathbb{E}|\eta_i| = \frac{1}{C}\mathbb{E}|\mathbb{E}(\xi|\mathscr{G}_i)| \leq \frac{1}{C}\mathbb{E}(\mathbb{E}(|\xi||\mathscr{G}_i)) = \frac{1}{C}\mathbb{E}|\xi|.$$

注意 $[|\eta_i|\geq C]\in\mathscr{G}_i$, 则

$$\begin{aligned}\int_{[|\eta_i|\geq C]} |\eta_i|\mathrm{d}P &\leq \int_{[|\eta_i|\geq C]} |\xi|\mathrm{d}P = \int_{[|\eta_i|\geq C]\cap[|\xi|<\delta]} |\xi|\mathrm{d}P + \int_{[|\eta_i|\geq C]\cap[|\xi|\geq\delta]} |\xi|\mathrm{d}P \\ &\leq \delta P([|\eta_i|\geq C]) + \int_{[|\xi|\geq\delta]} |\xi|\mathrm{d}P \\ &\leq \frac{\delta}{C}\mathbb{E}|\xi| + \int_{[|\xi|\geq\delta]} |\xi|\mathrm{d}P.\end{aligned}$$

由积分的绝对连续性, 对 $\varepsilon>0$, 取 $\delta>0$, 使得 $\int_{[|\xi|\geq\delta]} |\xi|\mathrm{d}P \leq \dfrac{\varepsilon}{2}$, 则当 $C \geq \dfrac{2\delta}{\varepsilon}\mathbb{E}|\xi|$ 时, 有 $\int_{[|\eta_i|\geq C]} |\eta_i|\mathrm{d}P \leq \varepsilon, i\in I$. 所以 $\{\eta_i\}$ 是一致可积族. □

§ 6.5 第六章习题

1. 设 $X,Y\in L^2(\Omega,\mathscr{F},P)$ 满足 $\mathbb{E}(Y|X) = X, \mathbb{E}(X|Y) = Y$, 则 $X=Y$, a.s..
2. 设 $X,Y\in L^1(\Omega,\mathscr{F},P)$ 满足 $\mathbb{E}(Y|X) = X, \mathbb{E}(X|Y) = Y$, 则 $X=Y$, a.s..

3. 已知 X 是一可积 r.v., $\mathscr{C}$ 是 $\mathscr{F}$ 的子 sigma 代数. 令 $Y = \mathbb{E}(X|\mathscr{C})$, 假定 X 与 Y 同分布, 证明:
   (1) 若 X 平方可积, 则 $X = Y$, a.s..
   (2) 若 $\forall a, b \in \mathbb{R}, a < b$, 有 $(X \vee a) \wedge b = (Y \vee a) \wedge b$, a.s., 则 $X = Y$, a.s..
4. 设 $X \in L^2(\Omega, \mathscr{F}, P)$, $\mathscr{C}_1 \subset \mathscr{C}_2$ 是 $\mathscr{F}$ 的子 σ-代数, 则
$$\mathbb{E}(\mathbb{E}(X|\mathscr{C}_1) - X)^2 \geq \mathbb{E}(\mathbb{E}(X|\mathscr{C}_2) - X)^2.$$
5. 设 ξ, η 是 $(\Omega, \mathscr{F}, P)$ 中的随机变量, $\mathbb{E}(\xi|\mathscr{C}) = \eta$, 且 $\mathbb{E}\xi^2 = \mathbb{E}\eta^2 < +\infty$. 证明: $\xi = \eta$, a.s..
6. 设 $\{\xi_n\}$ 是一致可积随机变量序列, 则有 $\lim\limits_{n\to\infty} \mathbb{E}\left(\frac{1}{n}\sup\limits_{1\leq k\leq n}|\xi_k|\right) = 0$.
7. 设 $\mathcal{H} \subset L^1(\Omega, \mathscr{F}, P)$, 若 $\mathcal{H}$ 满足:
$$A_n \in \mathcal{F}, A_n \searrow \varnothing \Rightarrow \lim_{n\to\infty} \sup_{\xi\in\mathcal{H}} \int_{A_n} |\xi| \mathrm{d}P = 0,$$
   则对任给 $\varepsilon > 0$, 存在 $\delta > 0$, 使得
$$A \in \mathscr{F}, P(A) < \delta \Rightarrow \sup_{\xi\in\mathcal{H}} \int_A |\xi| \mathrm{d}P \leq \varepsilon.$$
8. 记 $(\Omega, \mathscr{F})$ 是可测空间, $\mathscr{C}$ 是 $\mathscr{F}$ 的子 $\sigma-$ 代数, 记 P, Q 是两个在 $\mathscr{F}$ 中相互绝对连续（即等价）的概率测度, X_0 是 Q 关于 P 在 $\mathscr{F}$ 上的 Radon-Nikodym density, 证明下面性质成立:
   (1) $0 < E_P(X_0|\mathscr{C}) < +\infty$, a.s..
   (2) 对每个 $\mathscr{F}$ 可测的非负 r.v.f, $E_P(fX_0|\mathscr{C}) = E_Q(f|\mathscr{C})E_P(X_0|\mathscr{C})$.

第 7 章　部分习题的参考答案

§ 7.1　第二章习题

例 7.1.1

已知随机向量 (X,Y) 的密度函数的分布为

$$p(x,y)=\begin{cases}2-x-y, & 0<x<1,0<y<1,\\ 0, & \text{其他}.\end{cases}$$

求 $P(X>2Y)$.

提示: $P(X>2Y)=\mathbb{E}I_A(x,y)=\int_A p(x,y)\mathrm{d}x\mathrm{d}y$. 其中 $A=\{(x,y):x>2y\}\subset\mathbb{R}^2$.

复习: $\mathbb{E}f(\xi)=\int_{\mathbb{R}^n}f(x)\mathrm{d}F_\xi(x)$. 其中 ξ 是 n 维向量.

例 7.1.2. 2012 期中

若 ξ,η 是相互独立的随机变量, $\xi\sim N(0,1),\eta\sim N(0,1)$, 则 $\rho=\sqrt{\xi^2+\eta^2}$ 与 $\varphi=\arctan\frac{\eta}{\xi}$ 是相互独立的.

证明: 令 $x=r\cos\theta,y=r\sin\theta$, 则 $r=\sqrt{x^2+y^2},J=r$, 且

$$\theta=\begin{cases}\arctan\frac{y}{x}, & x>0,\\ \arctan\frac{y}{x}+\pi, & x<0,\end{cases},\theta\in[-\pi/2,3\pi/2].$$

从而 (ρ,φ) 的密度函数为

$$q(r,\theta)=\frac{1}{2\pi}re^{-\frac{r^2}{2}},r>0.$$

而 ρ 的密度为

$$R(r)=\int_{-\pi/2}^{3\pi/2}\frac{1}{2\pi}re^{-\frac{r^2}{2}}\mathrm{d}\theta=re^{-\frac{r^2}{2}},$$

θ 的密度为

$$p(\theta)=\int_0^{\infty}\frac{1}{2\pi}re^{-\frac{r^2}{2}}\mathrm{d}r=\frac{1}{2\pi}.$$

所以 $q(r,\theta)=R(r)p(\theta)$, 从而 θ,ρ 独立. □

例 7.1.3. 2022 南京大学推免

设 X,Y 为独立同分布的随机变量, 且 X 服从参数为 1 的指数分布, 求 $\frac{X}{X+Y}$ 的密度函数.

解:【正解】$X\sim E(1)$, 则 X 的密度函数是 $p(x)=\begin{cases}e^{-x}, & x\geq 0,\\ 0, & x<0.\end{cases}$

$\alpha=X+Y$ 的密度函数是

$$p_{X+Y}(y)=\int_{-\infty}^{+\infty}p(y-x)p(x)\mathrm{d}x.$$

当 $y<0$ 时, $p_\alpha(y)=0$. 当 $y\ge 0$ 时, $p_\alpha(y)=\int_0^y e^{-(y-x)}e^{-x}\mathrm{d}x=ye^{-y}$. 所以

$$p_\alpha(y)=\begin{cases} ye^{-y}, & y\ge 0,\\ 0, & y<0.\end{cases}$$

把 $\beta=\dfrac{X}{X+Y}$ 的密度函数记为 $p_\beta(y)$. 令 $u=x+y$, $v=\dfrac{x}{x+y}$, 则 $x=uv$, $y=u(1-v)$, 并且 Jacobi 矩阵为

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{bmatrix} v & u\\ 1-v & -u\end{bmatrix},\qquad |\det J|=u.$$

根据联合密度的替换公式, u,v 的联合密度是

$$\begin{aligned} q(u,v)&=p(x)p(y)|\det J|\\ &=e^{-x}I_{[x\ge 0]}\cdot e^{-y}I_{[y\ge 0]}\cdot u\\ &=ue^{-u}I_{[x\ge 0,y\ge 0]}.\end{aligned}$$

注意当 $x,y\ge 0$ 时, $u=x+y\ge 0$, 从而由 $x=uv$, $y=u(1-v)$ 可得 $v\in[0,1]$. 反之, 当 $u\ge 0$, $v\in[0,1]$ 时, 也可以推出 $x,y\ge 0$. 所以

$$I_{[x\ge 0,y\ge 0]}=I_{[u\ge 0]}I_{[0\le v\le 1]}.$$

故

$$q(u,v)=p_\alpha(u)I_{[0\le v\le 1]}.$$

从而根据上式可得 α,β 独立, 并且 β 的密度函数是 $p_\beta(v)=I_{[0\le v\le 1]}$, 即 β 服从 $[0,1]$ 上的均匀分布. □

【错解. 错因是什么?】$X\sim E(1)$, 则 X 的密度函数是 $p(x)=\begin{cases} e^{-x}, & x\ge 0,\\ 0, & x<0.\end{cases}$

$X+Y$ 的密度函数是

$$p_{X+Y}(y)=\int_{-\infty}^{+\infty}p(y-x)p(x)\mathrm{d}x.$$

当 $y<0$ 时, $p_{X+Y}(y)=0$. 当 $y\ge 0$ 时, $p_{X+Y}(y)=\int_0^y e^{-(y-x)}e^{-x}\mathrm{d}x=ye^{-y}$. 所以

$$p_{X+Y}(y)=\begin{cases} ye^{-y}, & y\ge 0,\\ 0, & y<0.\end{cases}$$

随机变量 $Z=\dfrac{X}{X+Y}$ 的密度函数是

$$p_Z(y)=\int_{-\infty}^{+\infty}|x|p(xy)p_{X+Y}(x)\mathrm{d}x.$$

当 $y<0$ 时, $p_Z(y)=0$. 当 $y\ge 0$ 时,

$$p_Z(y)=\int_0^{+\infty}|x|e^{-xy}\cdot xe^{-x}\mathrm{d}x=\int_0^{+\infty}x^2e^{-(y+1)x}\mathrm{d}x=\frac{2}{(y+1)^3}.$$

所以 $\dfrac{X}{X+Y}$ 的密度函数是

$$p_Z(y)=\begin{cases}\dfrac{2}{(y+1)^3}, & y\geq 0,\\ 0, & y<0.\end{cases}$$

例 7.1.4. 2022 南京大学推免

设 $X_1, X_2, \cdots, X_n$ 为独立同分布随机变量, 且 X_1 的密度函数为 $p(x)$. 证明:

(1) $P(\max\{X_1,\cdots,X_n\}=X_1)=\dfrac{1}{n}$;

(2) 随机变量 $\max\{X_1,\cdots,X_n\}$ 与 $I_{[X_1=\max\{X_1,\cdots,X_n\}]}$ 相互独立.

证明: (1) 由于 $X_1,\cdots,X_n$ 是独立同分布的, 所以

$$P(\max\{X_1,\cdots,X_n\}=X_1)=P(\max\{X_1,\cdots,X_n\}=X_2)=\cdots=P(\max\{X_1,\cdots,X_n\}=X_n).$$

注意到

$$\sum_{k=1}^{n}P(\max\{X_1,\cdots,X_n\}=X_k)=1,$$

所以 $P(\max\{X_1,\cdots,X_n\}=X_1)=\dfrac{1}{n}$.

(2) 记 $F(x)=\displaystyle\int_{-\infty}^{x}p(t)\mathrm{d}t$. 注意到由于 $X_1,\cdots,X_n$ 是独立同分布的, 故对任意实数 $x\in\mathbb{R}$,

$$P(\max\{X_1,\cdots,X_n\}\leq x)=P(X_1\leq x,\cdots,X_n\leq x)=[P(X_1\leq x)]^n=[F(x)]^n$$

并且由 (1),

$$P(I_{[X_1=\max\{X_1,\cdots,X_n\}]}=1)=P(X_1=\max\{X_1,\cdots,X_n\})=\frac{1}{n}$$

并且

$$\begin{aligned}
&P(\max\{X_1,\cdots,X_n\}\leq x, I_{[X_1=\max\{X_1,\cdots,X_n\}]}=1)\\
=&P(X_1\leq x, X_1=\max\{X_1,\cdots,X_n\})\\
=&P(X_1\leq x, X_2\leq X_1,\cdots,X_n\leq X_1)\\
=&\int_{-\infty}^{x}p(x_1)\left(\int_{-\infty}^{x_1}p(x_2)\mathrm{d}x_2\cdots\int_{-\infty}^{x_1}p(x_n)\mathrm{d}x_n\right)\mathrm{d}x_1\\
=&\int_{-\infty}^{x}p(x_1)F(x_1)^{n-1}\mathrm{d}x_1=\int_{-\infty}^{x}F(x_1)^{n-1}\mathrm{d}F(x_1)\\
\xlongequal{y=F(x_1)}&\int_{-\infty}^{F(x)}y^{n-1}\mathrm{d}y=\frac{1}{n}[F(x)]^n.
\end{aligned}$$

于是我们证明了

$$\begin{aligned}
&P(\max\{X_1,\cdots,X_n\}\leq x, I_{[X_1=\max\{X_1,\cdots,X_n\}]}=1)\\
=&P(\max\{X_1,\cdots,X_n\}\leq x)P(I_{[X_1=\max\{X_1,\cdots,X_n\}]}=1).
\end{aligned}$$

从而随机变量 $\max\{X_1,\cdots,X_n\}$ 与 $I_{[X_1=\max\{X_1,\cdots,X_n\}]}$ 相互独立.

例 7.1.5. 2024 北京大学期中

构造随机变量 $X_1\sim E(\lambda_1)$，$X_2\sim E(\lambda_2)$，$\lambda_1<\lambda_2$，使得 $X_1\geq X_2$ a.s.

设 $X_1\sim E(\lambda_1)$，$X_3\sim E(\lambda_2-\lambda_1)$ 是相互独立的随机变量，并设 $X_2=\min\{X_1,X_3\}$. 则

$X_2 \sim E(\lambda_2)$，且 $X_2 = \min\{X_1, X_3\}$. $\square$

例 7.1.6. 2022 年丘成桐大学生数学竞赛 (决赛) 概率与统计部分

设 $\mathcal{N} = \{1,2,3,4,\cdots\}$ 表示正整数全体, $\mathcal{P} = \{2,3,5,7,\cdots\}$ 表示素数全体. 记 $a \mid b$ 表示 a 整除 b. 固定实数 $s > 1$, 令 $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, 定义 $\mathcal{N}$ 上的概率测度为 $P_s(\{n\}) = \frac{1}{\zeta(s)} n^{-s}, n \in \mathcal{N}$. 对任意 $p \in \mathcal{P}$, 定义 $\mathcal{N}$ 上的随机变量 X_p 为 $X_p(n) = \mathbf{1}_{\{p|n\}}(n)$, $n \in \mathcal{N}$, 其中 $\{p|n\}$ 表示事件 $\{n : p|n\} \subset \mathcal{N}$.

(1) 集合 $\{X_p : p \in \mathcal{P}\}$ 中的随机变量在 P_s 的意义下是否相互独立?

(2) 用概率方法证明 Euler 恒等式 $\frac{1}{\zeta(s)} = \prod_{p \in \mathcal{P}} (1 - p^{-s})$.

证明: 我们省略下标 s.

(1) 对 $p, q \in \mathcal{P} (p \neq q)$, 有

$$P(X_p = 1) = P(\{n : p|n\}) = P(\{pk : k = 1,2,\cdots\}) = \sum_{k=1}^{\infty} P(\{pk\}) = \sum_{k=1}^{\infty} \frac{1}{\zeta(s)} (kp)^{-s} = p^{-s}.$$

并且

$$\begin{aligned} P(X_p = 1 \text{且} X_q = 1) &= P(\{n : p|n \text{且} q|n\}) = P(\{pqk : k = 1,2,\cdots\}) \\ &= \sum_{k=1}^{\infty} P(\{pqk\}) = \sum_{k=1}^{\infty} \frac{1}{\zeta(s)} (kpq)^{-s} = (pq)^{-s}. \end{aligned}$$

因此 $P(X_p = 1 \text{且} X_q = 1) = P(X_p = 1) P(X_q = 1)$. 利用 $P(AB) = P(B) - P(\overline{A}B)$ 可证明对 $x, y \in \{0,1\}$ 都有

$$\begin{aligned} P(X_p = 0 \text{且} X_q = 1) &= P(X_q = 1) - P(X_p = 1 \text{且} X_q = 1) \\ &= q^{-s} - (pq)^{-s} = (1 - p^{-s}) q^{-s} \\ &= P(X_p = 0) P(X_q = 1). \end{aligned}$$

其他两个是类似的. 故 $\{X_p : p \in \mathcal{P}\}$ 中的随机变量都是相互独立的.

(2) 为方便起见, 把素数从小到大排列为 $\mathcal{P} = \{p_i : i = 1,2,\cdots\}$. 于是由独立性可知, 对任意正整数 N, 有

$$P\left(\bigcap_{k=1}^{N} [X_{p_k} = 0] \right) = \prod_{k=1}^{N} P(X_{p_k} = 0) = \prod_{k=1}^{N} (1 - p_k^{-s}).$$

下面记 $A_N = \bigcap_{k=1}^{N} [X_{p_k} = 0]$, $A = \bigcap_{k=1}^{\infty} [X_{p_k} = 0]$, 则 A_N 单调递减趋于 A. (即 $A_1 \supset A_2 \supset \cdots \supset A_N \supset \cdots \supset A$ 且 $\bigcap_{N=1}^{\infty} A_N = A$)

注意到, $n \in A_N = \bigcap_{k=1}^{N} [X_{p_k} = 0] = \bigcap_{k=1}^{N} \{n : p_k \nmid n\}$ 等价于 $p_k \nmid n$ 对任意正整数 $k = 1,2,\cdots,N$ 成立.

根据大于 1 的正整数必定可以分解成一些素数的乘积, 即对任意 $n > 1$, 存在 $p_m \in \mathcal{P}$ 使得 $p_m | n$,

于是 $n \notin \bigcap_{k=1}^{m}[X_{p_k}=0]=A_m$, 所以 $n \notin A(A \subset A_m$, A 是个更小的集合). 于是必有

$$A=\bigcap_{k=1}^{\infty}[X_{p_k}=0]=\{1\}.$$

从而根据从上连续性,

$$\frac{1}{\zeta(s)}=P(\{1\})=P(A)=\lim_{N\to\infty}P(A_N)=\lim_{N\to\infty}\prod_{k=1}^{N}(1-p_k^{-s})=\prod_{k=1}^{\infty}(1-p_k^{-s}).$$

这就完成了证明. □

§ 7.2 第三章习题

引理 7.2.1. Mills's Ratio

设 $\phi(x)=e^{-\frac{x^2}{2}}, \Phi(x)=\int_{-\infty}^{x}\phi(t)\mathrm{d}t$, 则

$$\frac{1}{x}-\frac{1}{x^3}<\frac{1-\Phi(x)}{\phi(x)}<\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}, x>0.$$

这个估计式我们非常感兴趣, 因为 $\Phi(x)$ 没有封闭形式的表达式.

证明: $\phi(x)$ 满足 $\phi'=-x\phi$. 所以不断用分部积分可得

$$\begin{aligned}1-\Phi(x)&=\int_x^{\infty}\phi(t)\mathrm{d}t=-\int_x^{\infty}\frac{\phi'(t)}{t}\mathrm{d}t\\&=\frac{\phi(x)}{x}+\int_x^{\infty}\frac{\phi'(t)}{t^3}\mathrm{d}t\\&=\frac{\phi(x)}{x}-\frac{\phi(x)}{x^3}-\int_x^{\infty}\frac{\phi'(t)}{t^5}\mathrm{d}t\\&=\frac{\phi(x)}{x}-\frac{\phi(x)}{x^3}+\frac{3\phi(x)}{x^5}-\int_x^{\infty}\frac{15\phi(t)}{t^6}\mathrm{d}t.\end{aligned}$$

证明完毕. □

例 7.2.2

在长为 a 的线段上任取两点 X 和 Y, 求此两点之间的平均长度.

解: $X\sim U(0,a), Y\sim U(0,a)$, 且 X, Y 独立, 则

$$\mathbb{E}|X-Y|=\int_{\mathbb{R}^2}|x-y|\mathrm{d}F(x,y)=\int_0^a\int_0^a|x-y|\frac{1}{a}\cdot\frac{1}{a}\mathrm{d}x\mathrm{d}y=\cdots=\frac{a}{3}.$$

注: $\mathrm{d}F(x,y)=p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y$.

例 7.2.3. 2022 某校推免

假定随机变量 X 服从参数为 1 的泊松分布, 现在对其独立观测 n 次, 设 Y_n 为 X 大于 1 的次数, 求 Y_n^2 的期望.

解: $X \sim P(1)$, 则 $P(X=k) = e^{-1}\frac{1}{k!}$, $k = 0, 1, \cdots$. 所以

$$P(X > 1) = 1 - 2e^{-1}.$$

由于每次观测都是独立的, 所以 $Y_n \sim b(n, 1-2e^{-1})$, 即 Y_n 服从二项分布. 故 $\mathbb{E}Y_n = (1-2e^{-1})n$, $\mathbb{D}Y_n = 2e^{-1}(1-2e^{-1})n$. 注意到恒等式

$$\mathbb{D}Y_n = \mathbb{E}(Y_n)^2 - (\mathbb{E}Y_n)^2,$$

所以 Y_n^2 的期望为

$$\mathbb{E}(Y_n^2) = \mathbb{D}Y_n + (\mathbb{E}Y_n)^2 = 2e^{-1}(1-2e^{-1})n + (1-2e^{-1})^2 n^2.$$

例 7.2.4. 2019 期末

X, Y 是独立的随机变量, $EX = 0$, $E|Y| < +\infty$, $E(|X+Y|) < +\infty$. 证明:

$$E(|Y|) \le E(|X+Y|).$$

证明: $EX = 0 \Rightarrow |y| = |E(X+y)|$. 所以

$$\begin{aligned}
E|Y| = \int_{\mathbb{R}} |y| \mathrm{d}F_Y(y) &= \int_{\mathbb{R}} |E(y+X)| \mathrm{d}F_Y(y) \\
&\le \int_{\mathbb{R}} E|y+X| \mathrm{d}F_Y(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |y+x| \mathrm{d}F_X(x) \mathrm{d}F_Y(y) \\
&= \int_{\mathbb{R}^2} |x+y| \mathrm{d}F_{(X,Y)}(x,y) \qquad (\text{独立性}) \\
&= E|X+Y|.
\end{aligned}$$

例 7.2.5

设随机变量 X, Y 的期望分别为 $-2, 2$, 方差分别为 $1, 4$, 且 $\mathbb{E}(X+2)(Y-2) = -1$. 请用所学知识给出 $P(|X+Y| \ge 6)$ 的一个非平凡上界.

证明: 【方法一】注意到

$$\begin{aligned}
\mathbb{P}(|X+Y| \ge 6) &\le \frac{\mathbb{D}(X+Y)}{36} \\
&= \frac{1}{36}\left(\mathbb{E}(X-\mathbb{E}X)^2 + 2\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] + \mathbb{E}(Y-\mathbb{E}Y)^2\right) \\
&= \frac{1}{36}(\mathbb{D}X + \mathbb{D}Y - 2) = \frac{1}{12}.
\end{aligned}$$

【方法二】对任意 $p \in (0, 2)$,

$$P(|X+Y| \ge 6) \le \frac{1}{6^p}\mathbb{E}|X+Y|^p \le \frac{1}{6^p}(\mathbb{E}|X+Y|^2)^{p/2} = \frac{1}{6}^p \cdot 3^{p/2} = \frac{1}{12^{p/2}} < 1.$$

注: 方法二用了 Holder 不等式.

例 7.2.6. 2022 某校推免

设 ξ 为取自然数值的随机变量, φ 为其特征函数, 证明:

$$P(\xi=k)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ikt}\varphi(t)\mathrm{d}t,\qquad k=0,1,2,\cdots.$$

证明: 根据特征函数的定义,

$$\varphi(t)=\mathbb{E}e^{it\xi}=\sum_{j=0}^{\infty}e^{ijt}P(\xi=j).$$

注意到 Fourier 基函数具有正交性, 即

$$\int_{-\pi}^{\pi}e^{-i(k-j)t}\mathrm{d}t=\begin{cases}2\pi, & k=j,\\ 0, & k\neq j,\end{cases}$$

所以

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ikt}\varphi(t)\mathrm{d}t&=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ikt}\sum_{j=0}^{\infty}e^{ijt}P(\xi=j)\mathrm{d}t\\&=\frac{1}{2\pi}\sum_{j=0}^{\infty}P(\xi=j)\int_{-\pi}^{\pi}e^{-ikt}e^{ijt}\mathrm{d}t\\&=\frac{1}{2\pi}\cdot 2\pi P(\xi=k)=P(\xi=k).\end{aligned}$$

其中级数和积分可交换是因为有控制收敛定理, 并且

$$\left|\sum_{j=0}^{\infty}e^{ijt}P(\xi=j)\right|\leq\sum_{j=0}^{\infty}P(\xi=j)=1.$$

而 1 在 $[-\pi,\pi]$ 上的积分有界, 即常函数 1 是可积的. □

例 7.2.7

如果 $\{X_n\}$ 是一列非负整数值 r.v., 则

$$EX=\sum_{n=0}^{\infty}P(X\geq n).$$

证明: 注意到

$$\begin{aligned}EX&=\sum_{n=0}^{\infty}nP(X=n)\\&=\sum_{n=0}^{\infty}\sum_{i=1}^{n}P(X=n)\\&=\sum_{i=1}^{\infty}\sum_{n=i}^{\infty}P(X=n)=\sum_{i=1}^{\infty}P(X\geq i).\end{aligned}$$ □

例 7.2.8

假定 X 是非负 r.v., $p\geq 1$ 为常数, 则 $\mathbb{E}X^p=p\int_0^{\infty}y^{p-1}P(X>y)dy$.

证明: 注意到 Fubini 定理,

$$
\begin{aligned}
\mathbb{E}X^p &= \int_0^\infty x^p dF(x) \\
&= \int_0^\infty \int_0^x py^{p-1} dy dF(x) && \text{(不可以用 Newton-Leibniz 公式)} \\
&= p\int_0^\infty \int_y^\infty df(x) dy && \text{(被积函数非负可换序)} \\
&= p\int_0^\infty y^{p-1} P(X > y) dy.
\end{aligned}
$$

注: 特别地, 当 $p=1$ 时, 可以变成如下等式: $EX = \displaystyle\int_0^\infty P(X > y)dy$.

例 7.2.9. 2016 丘赛 Team

在单位圆 $x^2+y^2=1$ 上按均匀分布随机取 n 个点 $(n \geq 2)$, 这 n 个点可以把单位圆分成 n 段圆弧. 求包含点 $(1,0)$ 的圆弧的长度的数学期望和方差.

解: 圆上的点可由其极坐标的角度唯一决定, 故可设 $\xi_1, \cdots, \xi_n \sim U[0,2\pi]$ 是 n 个独立同分布的随机变量. 记 $\xi_1^* = \min\{\xi_1, \cdots, \xi_n\}$, $\xi_n^* = \max\{\xi_1, \cdots, \xi_n\}$. 于是包含点 $(1,0)$ 的圆弧长度是一个随机变量, 如下定义:

$$X = 2\pi + \xi_1^* - \xi_n^*.$$

为了求 X 的数学期望, 只需求 ξ_1^* 和 ξ_n^* 的数学期望, 这可以让我们联想到顺序统计量的分布. 当 $0 \leq x \leq 2\pi$ 时, 我们有

$$
\begin{aligned}
& P(\xi_1^* > x) = \prod_{i=1}^n P(\xi_i \geq x) = \left(1 - \frac{x}{2\pi}\right)^n, \\
& P(\xi_n^* \leq x) = \prod_{i=1}^n P(\xi_i \leq x) = \left(\frac{x}{2\pi}\right)^n, \\
\Rightarrow\quad & P(\xi_n^* > x) = 1 - \left(\frac{x}{2\pi}\right)^n.
\end{aligned}
$$

因此

$$
\begin{aligned}
\mathbb{E}\xi_1^* &= \int_0^{2\pi} P(\xi_1^* > x)\mathrm{d}x = \int_0^{2\pi} \left(1-\frac{x}{2\pi}\right)^n \mathrm{d}x = -\frac{2\pi}{n+1}\left(1-\frac{x}{2\pi}\right)^{n+1}\Big|_0^{2\pi} = \frac{2\pi}{n+1}, \\
\mathbb{E}\xi_n^* &= \int_0^{2\pi} P(\xi_n^* > x)\mathrm{d}x = \int_0^{2\pi} 1-\left(\frac{x}{2\pi}\right)^n \mathrm{d}x = 2\pi - \frac{2\pi}{n+1}\left(\frac{x}{2\pi}\right)^{n+1}\Big|_0^{2\pi} = 2\pi - \frac{2\pi}{n+1},
\end{aligned}
$$

所以

$$
\begin{aligned}
\mathbb{E}X &= \mathbb{E}(2\pi + \xi_1^* - \xi_n^*) \\
&= 2\pi + \mathbb{E}\xi_1^* - \mathbb{E}\xi_n^* = 2\pi + \frac{2\pi}{n+1} - \left(2\pi - \frac{2\pi}{n+1}\right) = \frac{4\pi}{n+1}.
\end{aligned}
$$

计算方差稍微麻烦一点, 因为 $\mathbb{D}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$, 而 X^2 展开后会得到 $\xi_1^*\xi_n^*$, 我们需要计算 $\mathbb{E}(\xi_1^*\xi_n^*)$, 就要知道联合密度是什么. 首先, 当 $0 < x < y < 2\pi$ 时,

$$P(x < \xi_1^*, \xi_n^* \leq y) = \prod_{i=1}^n P(x < \xi_i \leq y) = \left(\frac{y-x}{2\pi}\right)^n.$$

因此联合分布函数是

$$F(x,y)=P(\xi_1^*\le x,\xi_n^*\le y)=P(\xi_n^*\le y)-P(x<\xi_1^*,\xi_n^*\le y)=\left(\frac{y}{2\pi}\right)^n-\left(\frac{y-x}{2\pi}\right)^n.$$

所以密度为

$$p(x,y)=\frac{\partial^2}{\partial x\partial y}=\frac{n(n-1)}{(2\pi)^n}(y-x)^{n-2}I_{[0\le x<y\le 2\pi]}.$$

所以

$$\mathbb{E}(\xi_1^*\xi_n^*)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyp(x,y)\mathrm{d}x\mathrm{d}y=\int_0^{2\pi}\int_0^y xy\cdot\frac{n(n-1)}{(2\pi)^n}(y-x)^{n-2}\mathrm{d}x\mathrm{d}y=\frac{(2\pi)^2}{n+2}.$$

(中间用一下换元 $x=yt$ 会方便一点) 另外,

$$\begin{aligned}\mathbb{E}[(\xi_1^*)^2]&=\int_0^{2\pi}2xP(\xi_1^*>x)\mathrm{d}x=\int_0^{2\pi}2x\left(1-\frac{x}{2\pi}\right)^n\mathrm{d}x=\frac{2(2\pi)^2}{(n+1)(n+2)}\\ \mathbb{E}[(\xi_n^*)^2]&=\int_0^{2\pi}2xP(\xi_n^*>x)\mathrm{d}x=\int_0^{2\pi}2x-2x\left(\frac{x}{2\pi}\right)^n\mathrm{d}x=(2\pi)^2-\frac{2(2\pi)^2}{n+2}.\end{aligned}$$

所以

$$\begin{aligned}\mathbb{D}X&=\mathbb{E}(X^2)-(\mathbb{E}(X))^2\\&=\mathbb{E}[(2\pi)^2+4\pi\xi_1^*-4\pi\xi_n^*+(\xi_1^*)^2+(\xi_n^*)^2-2\xi_1^*\xi_n^*]-\left(\frac{4\pi}{n+1}\right)^2\\&=(2\pi)^2+4\pi\cdot\frac{2\pi}{n+1}-4\pi\cdot\left(2\pi-\frac{2\pi}{n+1}\right)\\&\quad+\frac{2(2\pi)^2}{(n+1)(n+2)}+\left[(2\pi)^2-\frac{2(2\pi)^2}{n+2}\right]-\frac{2(2\pi)^2}{n+2}-\frac{4(2\pi)^2}{(n+1)^2}\\&=\frac{8\pi^2(n-1)}{(n+1)^2(n+2)}.\end{aligned}$$

§ 7.3 第四章习题

例 7.3.1

设 $\{X_n\}$ 是 i.i.d.r.v, $X_1\sim U(0,1)$. 令 $Z_n=\left(\prod_{i=1}^n X_i\right)^{1/n}$, 证明; 存在 C 使得 $Z_n\xrightarrow{P}C$.

证明: 注意到

$$Z_n=\exp\left(\frac{1}{n}\ln\prod_{i=1}^n x_i\right)=\exp\left(\frac{1}{n}\sum_{i=1}^n\ln x_i\right)\xrightarrow{P}e^{E\ln x_1}.$$

(最后一步用了弱大数定律). □

注: X_n 依概率收敛, f 连续, 则 $f(X_n)$ 也依概率收敛.

例 7.3.2

如果 $\sum_{n=1}^{\infty}\|\xi_n-\xi\|_p^p<+\infty$, 则 $\xi_n\xrightarrow{a.s.}\xi$.

证明: 根据几乎必然收敛的刻画, 以及概率测度的单调性、从上连续性, 可以把欲证命题进行等价

转化:

$$\begin{aligned}\xi_n \xrightarrow{a.s.} \xi \Longleftrightarrow & \forall \varepsilon>0, P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right)=0\\ \Longleftrightarrow & \forall\varepsilon>0, \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right)=0.\end{aligned}$$

利用 Cauchy 准则, 可得

$$\sum_{n=1}^{\infty}\|\xi_n-\xi\|_p^p<+\infty \Longleftrightarrow \lim_{k\to\infty}\sum_{n=k}^{\infty}\|\xi_n-\xi\|_p^p=0.$$

则

$$\begin{aligned}P\left(\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\right) &\leq \sum_{n=k}^{\infty}P[|\xi_n-\xi|\geq\varepsilon] \quad (\text{次 } \sigma \text{ 可加性})\\ &\leq \sum_{n=k}^{\infty}\frac{1}{\varepsilon^p}E|\xi_n-\xi|^p \quad (\text{Chebyshev 不等式})\\ &\to 0(k\to\infty)\end{aligned}$$

例 7.3.3

让 $\{X_n\}$ 为正 r.v. 列, 并假定 $X_n \xrightarrow{P} 0$ 且 $\lim\limits_{n\to\infty}\mathbb{E}X_n=2$. 证明 $\lim\limits_{n\to\infty}\mathbb{E}|X_n-1|$ 存在, 并求这个值.

证明: $E|X_n-1|=E(1-X_n)I_{[X_n\leq1]}+E(X_n-1)I_{[X_n>1]}$. 而

$$\begin{aligned}\mathbb{E}(1-X_n)I_{[X_n\leq1]} &= P(X_n\leq1)-\mathbb{E}X_nI_{[X_n\leq1]}\\ &=1-P(X_n>1)-\mathbb{E}X_nI_{[X_n\leq1]}\\ &\to 1(n\to\infty).\end{aligned}$$

上面 $P(X_n>1)\to0$ 是因为题目条件的依概率收敛, 而 $\mathbb{E}X_nI_{[X_n\leq1]}\to0$ 的原因请自己思考.

另一方面,

$$\begin{aligned}E(X_n-1)I_{[X_n>1]} &= EX_nI_{[X_n>1]}-P(X_n>1)\\ &=EX_n-EX_nI_{[X_n\leq1]}-P(X_n>1)\to2(n\to\infty).\end{aligned}$$

则 $\lim\limits_{n\to\infty}E|X_n-1|=3$. □

例 7.3.4. 2014 个人

设 $\{X_n\}$ 是一列不相关的 r.v. 列且均值为 0, 满足

$$\sum_{n=1}^{\infty}nE|X_n|^2<\infty.$$

则 $S_n=\sum\limits_{i=1}^{n}X_i$ 几乎必然收敛.

证明: 易知 $S_n \xrightarrow{L^2} S \triangleq \sum_{i=1}^{\infty} X_i$, 这是因为

$$E\left(\sum_{i=1}^{\infty} X_i - \sum_{i=1}^{n} X_i\right)^2 = E\left(\sum_{i=n+1}^{\infty} X_i\right)^2 = \sum_{i=n+1}^{\infty} EX_i^2 \to 0(n \to \infty).$$

这里最后一个等号用到了不相关性 $(EXY = EXEY)$, 以及 $EX_k = 0, k \in \mathbb{N}$.

从而 $\forall \varepsilon > 0$,

$$P(|S_n - S| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} E|S_n - S|^2 \leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{\infty} E|X_i|^2.$$

故

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|S_n - S| \geq \varepsilon) &\leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{i=n+1}^{\infty} E|X_i|^2 \\
&= \frac{1}{\varepsilon^2} \sum_{i=2}^{\infty} \sum_{n=1}^{i-1} E|X_i|^2 \text{ 【非负, 可换求和次序】} \\
&= \frac{1}{\varepsilon^2} \sum_{i=2}^{\infty} (i-1) E|X_i|^2 \\
&\leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} i E|X_i|^2 < \infty \text{ 【条件】}.
\end{aligned}$$

根据 Borel-Cantelli 引理, $P([|S_n - S| \geq \varepsilon] i.o.) = 0$, 则 $S_n \xrightarrow{a.s.} S$. □

例 7.3.5

设 $\{X_n\}$ 两两不相关, 且 $\sum_{n=1}^{+\infty} nD(X_n) < +\infty$. 证明: $S_n = \sum_{i=1}^{n} (X_i - EX_i)$ 几乎必然收敛.

证明: 由两两不相关性,

$$E|S_m - S_n|^2 = E\left(\sum_{i=n+1}^{m} (X_i - EX_i)\right)^2 = \sum_{i=n+1}^{m} DX_i \to 0(n, m \to +\infty).$$

则 $S_n \xrightarrow{L^2} S$. 由 Chebyshev 不等式,

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|S_n - S| \geq \varepsilon) &= \sum_{n=1}^{+\infty} P\left(\left|\sum_{i=n+1}^{+\infty} (X_i - EX_i)\right| \geq \varepsilon\right) \\
&\leq \sum_{n=1}^{+\infty} \frac{1}{\varepsilon^2} \sum_{i=n+1}^{+\infty} DX_i \\
&= \frac{1}{\varepsilon^2} \sum_{i=2}^{+\infty} \sum_{n=1}^{i-1} DX_n \leq \frac{1}{\varepsilon^2} \sum_{i=2}^{+\infty} i DX_i < +\infty.
\end{aligned}$$

由 Borel-Cantelli 引理, $P(|S_n - S| \geq \varepsilon, i.o.) = 0$. 则 $S_n \xrightarrow{a.s.} S$. □

例 7.3.6. 2016Team,4

让 $\{X_n\}_{n\geq 1}$ 为 i.i.d. 实值 r.v. 列, 证明或否定: 若 $\overline{\lim_{n\to\infty}}\dfrac{|X_n|}{n}\leq 1$,a.s., 则

$$\sum_{n=1}^{\infty}P(|X_n|\geq n)<\infty.$$

证明: 主要用两次 Nice 引理. 根据 i.i.d. 条件,

$$\sum_{n=1}^{\infty}P(|X_1|\geq n)\leq E|X_1|\leq\sum_{n=1}^{\infty}P(|X_1|\geq n)+1.$$

用 $\frac{1}{2}|X_1|$ 代替 $|X_1|$ 有 (事实上, 换成 $k|X_1|,0<k<1$ 都行)

$$\sum_{n=1}^{\infty}P(|X_1|\geq 2n)\leq\frac{1}{2}E|X_1|\leq\sum_{n=1}^{\infty}P(|X_1|\geq 2n)+1.$$

(反证) 若

$$\sum_{n=1}^{\infty}P(|X_n|\geq n)=\infty,$$

则根据前面的两个 Nice 不等式得

$$\sum_{n=1}^{\infty}P(|X_n|\geq 2n)=\infty,$$

根据 Borel-Cantelli 引理, $P\left(\dfrac{|X_n|}{n}\geq 2, i.o.\right)=1$. 则 $\overline{\lim_{n\to\infty}}\dfrac{|X_n|}{n}\geq 2$, a.s., 与条件矛盾. □

例 7.3.7. 2019Team,1

假定 $\{X_n\}_{n\geq 1}$ 是 i.i.d.r.v. 列且共同分布是参数为 1 的指数分布, 则

$$P\left(\limsup_{n\to\infty}\frac{X_n}{\log n}=1\right)=1.$$

注: X_n 的分布函数是 $p(x)=\begin{cases}e^{-x}, & x\geq 0,\\ 0, & x<0.\end{cases}$. 我们的想法是用 Borel-Cantelli 引理. 注意到

$$\begin{aligned}\sum_{n=1}^{\infty}P\left(\frac{X_n}{\log n}\geq a\right)&=\sum_{n=1}^{\infty}P(X_1\geq a\log n)\\&=\sum_{n=1}^{\infty}\int_{a\log n}^{\infty}e^{-x}dx=\sum_{n=1}^{\infty}n^{-a}=\begin{cases}<+\infty, & a>1,\\ =+\infty, & a\leq 1.\end{cases}\end{aligned}$$

则当 $a>1$ 时,

$$P\left(\frac{X_n}{\log n}\geq a, i.o.\right)=0\Longleftrightarrow P\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left[\frac{X_n}{\log n}<a\right]\right)=1.$$

即从某个 n 以后所有事件 $\left[\dfrac{X_n}{\log n}<a\right]$ 都发生, 则根据 $a>1$ 是任意的, 必有 $\overline{\lim_{n\to\infty}}\dfrac{X_n}{\log n}\leq 1$, a.s..

当 $a\leq 1$ 时, 由 Borel-Cantelli 引理以及题目中 i.i.d 条件, $P\left(\dfrac{X_n}{\log n}\geq a, i.o.\right)=1$. 因此

$$\overline{\lim_{n\to\infty}}\frac{X_n}{\log n}\geq 1,\quad \text{a.s.}.$$

综上, $P\left(\limsup\limits_{n\to\infty}\frac{X_n}{\log n}=1\right)=1..$ □

例 7.3.8

设 f 单调不降, $\sum\limits_{n=1}^{\infty}Ef(|X_n-X|)<\infty$. 证明: $X_n\xrightarrow{a.s.}X$.

注: 考虑 Chebyshev 不等式与 Borel-Cantelli 引理.

证明: $\forall\varepsilon>0$, 由几乎必然收敛刻画, 只需证

$$P\left(\bigcap_{n=1}^{+\infty}\bigcup_{k=n}^{+\infty}[|X_k-X|\geq\varepsilon]\right)=0.$$

由 Borel-Cantelli 引理, 只需证

$$\sum_{n=1}^{+\infty}P(|X_n-X|\geq\varepsilon)<+\infty.$$

事实上, 由 Chebyshev 不等式的思想,

$$\sum_{n=1}^{+\infty}P(|X_n-X|\geq\varepsilon)\leq\sum_{n=1}^{+\infty}\frac{1}{f(\varepsilon)}Ef(|X_n-X|)<+\infty.$$

例 7.3.9. 2013 年丘成桐大学生数学竞赛团体赛

设实数 $\varepsilon>0$, 证明: 对几乎所有的 $x\in[0,1]$, 只有有限个有理数 $\frac{p}{q}\in(0,1)$(其中 $p,q\in\mathbb{N}^+$) 满足

$$\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}.$$

注: Dirichlet 定理: 对任意实数 θ, 存在无穷多个既约有理数 $\frac{p}{q}$ 满足

$$\left|\theta-\frac{p}{q}\right|<\frac{1}{q^2}.$$

证明: 记 P 表示 $[0,1]$ 上的概率测度, 那么本题意思是要证明

$$P\left(\left\{x\in[0,1]\middle|\text{只有有限个有理数 }\frac{p}{q}\in(0,1)(\text{其中 }p,q\in\mathbb{N}^+)\text{ 满足 }\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}.\right\}\right)=1.$$

首先注意到

$$\begin{aligned}&\text{对几乎所有的 }x\in[0,1]\text{, 只有有限个有理数 }\frac{p}{q}\in(0,1)\text{, 使得 }\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}.\\ \Leftrightarrow&\text{对几乎所有的 }x\in[0,1]\text{, 只对有限个正整数 }q\geq 2\text{, 存在 }p\in\{1,2,\cdots,q-1\}\text{, 使得 }\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}.\\ \Leftrightarrow&\text{事件 }A_q=\left\{x\in[0,1]\middle|\text{存在}p\in\{1,2,\cdots,q-1\}\text{, 使得 }\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}.\right\}\text{ 发生无限多次的概率为 0.}\end{aligned}$$

所以我们只需要用 Borel-Cantelli 引理来说明最后的事实即可.

事实上,

$$\begin{aligned}P(A_q) &\le P\left(\bigcup_{p=1}^{q-1}\left[\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}\right]\right)\\ &\le \sum_{p=1}^{q-1}P\left(\left|x-\frac{p}{q}\right|<\frac{1}{q^2(\log q)^{1+\varepsilon}}\right)\\ &=(q-1)\cdot\frac{2}{q^2(\log q)^{1+\varepsilon}}<\frac{2}{q(\log q)^{1+\varepsilon}},\end{aligned}$$

所以

$$\sum_{q=2}^{\infty}P(A_q)\le\sum_{q=2}^{\infty}\frac{2}{q(\log q)^{1+\varepsilon}}<\infty.$$

这个级数收敛可以利用积分判别法来说明:

$$\sum_{q=3}^{\infty}\frac{2}{q(\log q)^{1+\varepsilon}}\le\sum_{q=3}^{\infty}\int_{q-1}^{q}\frac{2}{t(\log t)^{1+\varepsilon}}\mathrm{d}t=\int_{2}^{\infty}\frac{2}{t(\log t)^{1+\varepsilon}}\mathrm{d}t=\frac{2}{\varepsilon\cdot(\log 2)^{\varepsilon}}.$$

这就完成了证明. □

注: 辛钦 (Khintchine) 在 1924 年[1]给出如下结论: 若 $\psi:\mathbb{N}\to(0,+\infty)$ 是单调函数, 且 P 是 $[0,1]$ 上的 Lebesgue 测度. 记集合

$$L_\psi=\left\{x\in[0,1]\,\middle|\,\text{有无限多个有理数 }\frac{p}{q}\in(0,1)\text{ 满足 }\left|x-\frac{p}{q}\right|<\frac{\psi(q)}{q}\right\}$$

则

$$P\left(L_\psi\right)=\begin{cases}0, & \text{若}\displaystyle\sum_{q=1}^{\infty}\psi(q)<\infty,\\ 1, & \text{若}\displaystyle\sum_{q=1}^{\infty}\psi(q)=\infty.\end{cases}$$

关于它的介绍可参见 [V. Bernik, V. Beresnevich, F. Götze, O. Kukso, 2013, Distribution of Algebraic Numbers and Metric Theory of Diophantine Approximation.]

7.3.1 随机变量的独立性

例 7.3.10

设 $(X_n,n\ge1)$ 是独立 r.v. 序列, 则 $\limsup\limits_{n\to\infty}X_n$ 与 $\liminf\limits_{n\to\infty}X_n$ 是退化随机变量 (即 a.s. 等于某个常数).

证明: 对任意 $c>0$, $\displaystyle\sum_{n=1}^{\infty}P(X_n\ge c)=\sum_{n=1}^{\infty}P(X_1\ge c)=\begin{cases}+\infty, & P(X_1\ge c)\ne0,\\ 0, & P(X_1\ge c)=0.\end{cases}$ 由 Borel-Cantelli 引理及 (X_n) 是独立 r.v. 列, 则

$$P\Big(X_n\ge c,i.o.\Big)=\begin{cases}0, & P(X_1\ge c)=0,\\ 1, & P(X_1\ge c)\ne0.\end{cases}$$

[1] A.Ya. Khintchine, Einige Satzeüber Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen. Math. Ann. 92, 115-125 (1924)

所以 $P(\limsup\limits_{n\to\infty} X_n = c) = 0$ 或 1. □

例 7.3.11

设 $\{X_n\}$ 是 i.i.d., 则 $\dfrac{X_n}{\log n} \xrightarrow{a.s.} 0 \Leftrightarrow \forall c > 0, \mathbb{E}e^{c|X_1|} < \infty$.

证明: 由 a.s. 收敛的刻画, $\dfrac{X_n}{\log n} \xrightarrow{a.s.} 0$ 等价于对任意 $c > 0$, 有 $\lim\limits_{k\to\infty} P\left(\bigcup\limits_{n\geq k}\left[\dfrac{|X_n|}{\log n} > \dfrac{1}{c}\right]\right) = 0$. 利用独立性 (或者 Borel-Cantelli 引理),

$$P\left(\bigcup_{n\geq k}\left[\frac{|X_n|}{\log n} > \frac{1}{c}\right]\right) = \sum_{n=k}^{\infty} P\left(\frac{|X_n|}{\log n} > \frac{1}{c}\right) = \sum_{n=k}^{\infty} P\left(\frac{|X_1|}{\log n} > \frac{1}{c}\right) = \sum_{n=k}^{\infty} P\left(e^{c|X_1|} > n\right).$$

运用 Nice 引理 (引理 4.2.1) 即可得欲证结论. □

例 7.3.12

设 X, Y 相互独立, X 有密度函数, 则 $X+Y$ 也有密度函数.

证明: 设 $Z = X + Y$, 则由独立性,

$$F_Z(z) = \int_{[x+y\leq z]} \mathrm{d}F_{(X,Y)}(x,y) = \int_{\mathbb{R}}\int_{[x\leq z-y]} \mathrm{d}F_X(x)\mathrm{d}F_Y(y) = \int_{\mathbb{R}} F_X(z-y)\mathrm{d}F_Y(y).$$

若 X 有密度函数 f_X, 则 $F_X(z-y) = \displaystyle\int_0^{z-y} F_X(x)\mathrm{d}x$, 所以

$$\begin{aligned} F_Z(z) &= \int_{\mathbb{R}}\int_{-\infty}^{z-y} f_X(x)\mathrm{d}x\mathrm{d}F_Y(y) \\ &\xlongequal{x=u-y} \int_{\mathbb{R}}\int_{-\infty}^{z} f_X(u-y)\mathrm{d}u\mathrm{d}F_Y(y) \\ &= \int_{-\infty}^{z}\int_{\mathbb{R}} f_X(u-y)\mathrm{d}F_Y(y)\mathrm{d}u \qquad \text{(Fubini)} \end{aligned}$$

所以 $f_Z(z) = \displaystyle\int_{\mathbb{R}} f_X(z-y)\mathrm{d}F_Y(y)$. □

例 7.3.13

设 $(X_n, n \geq 1)$ 是 i.i.d.r.v. 列, X_n 都服从指数为 1 的指数分布, 即 $P(X_n > x) = e^{-x}, x \geq 0$.

(1) 证明:$P(X_n > \alpha\log n, i.o.) = \begin{cases} 0, & 若\alpha > 1, \\ 1, & 若\alpha \leq 1. \end{cases}$

(2) 令 $L = \limsup\limits_{n\to\infty}\dfrac{X_n}{\log n}$, 证明: $P(L = 1) = 1$. (提示: 证明 $P(L \geq 1) = 1, P(L > 1) = 0$.)

证明: (1) 由 Borel-Cantelli 引理, 欲证命题等价于 $\displaystyle\sum_{n=1}^{\infty} P(X_n > \alpha\log n)\begin{cases} < \infty, & 若\alpha > 1, \\ +\infty, & 若\alpha \leq 1. \end{cases}$ 事实上, 有

$$\sum_{n=1}^{\infty} P(X_n > \alpha\log n) = \sum_{n=1}^{\infty} e^{-\alpha\log n} = \sum_{n=1}^{\infty} n^{-\alpha}\begin{cases} < \infty, & 若\alpha > 1, \\ +\infty, & 若\alpha \leq 1. \end{cases}$$

证明完毕.

(2) 回顾: $[\inf_n f_n < x] = \bigcup_{n=1}^{\infty}[f_n < x]$, 而 $[\sup_n f_n > x] = \bigcup_{n=1}^{\infty}[f_n > x]$.

注意到

$$
\begin{aligned}
P(L > 1) = P(\inf_k \sup_{n\ge k} \frac{X_n}{\log n} > 1) &= P\left(\bigcup_{m=1}^{\infty}\left[\inf_k \sup_{n\ge k}\frac{X_n}{\log n} \ge 1+\frac{1}{m}\right]\right) \\
&= P\left(\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\left[\sup_{n\ge k}\frac{X_n}{\log n} \ge 1+\frac{1}{m}\right]\right) \\
&= P\left(\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left[\frac{X_n}{\log n} \ge 1+\frac{1}{m}\right]\right) \\
&= \lim_{m\to\infty} P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left[\frac{X_n}{\log n} \ge 1+\frac{1}{m}\right]\right) = 0, \qquad \text{(由第 (1) 小问)}
\end{aligned}
$$

$$
\begin{aligned}
P(L \ge 1) = P(\inf_k \sup_{n\ge k} \frac{X_n}{\log n} \ge 1) &= P\left(\bigcap_{k=1}^{\infty}\left[\sup_{n\ge k}\frac{X_n}{\log n} \ge 1\right]\right) \\
&= P\left(\bigcap_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\left[\sup_{n\ge k}\frac{X_n}{\log n} > 1-\frac{1}{m}\right]\right) \\
&= P\left(\bigcap_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=k}^{\infty}\left[\frac{X_n}{\log n} > 1-\frac{1}{m}\right]\right) \\
&= \lim_{m\to\infty} P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left[\frac{X_n}{\log n} > 1-\frac{1}{m}\right]\right) = 1.
\end{aligned}
$$

例 7.3.14

设 $\{X_n : n \ge 1\}$ 是独立的 $N(0,1)$ 随机变量, 证明:

$$P\left(\limsup_{n\to\infty}\frac{|X_n|}{\sqrt{\log n}} = \sqrt{2}\right) = 1.$$

证明: 类似前一题 (2), 只需要证明 $P(|X_n| > \alpha\log n, i.o.) = \begin{cases} 0, & \alpha > \sqrt{2}, \\ 1, & \alpha \le \sqrt{2}. \end{cases}$.

根据 Mills' Ratio（第三章习题），当 $\alpha > \sqrt{2}$ 时,

$$\sum_{n=3}^{\infty} P(|X_n| > \alpha\sqrt{\log n}) < \sum_{n=3}^{\infty} n^{-\frac{\alpha^2}{2}} \cdot \frac{1}{\alpha\sqrt{\log n}} \le \sum_{n=3}^{\infty} n^{-\frac{\alpha^2}{2}} < \infty.$$

当 $0 < \alpha \le \sqrt{2}$ 时,

$$\sum_{n=3}^{\infty} P(|X_n| > \alpha\sqrt{\log n}) > \sum_{n=3}^{\infty} n^{-\frac{\alpha^2}{2}} \cdot \frac{1}{\alpha\sqrt{\log n}}\left(1-\frac{1}{\alpha^2\log n}\right) = +\infty.$$

由 Borel-Cantelli 引理, $P(|X_n| > \alpha\sqrt{\log n}, i.o.) = \begin{cases} 0, & 若\alpha > \sqrt{2}, \\ 1, & 若\alpha \le \sqrt{2}. \end{cases}$. 用类似前一题 (2) 的说明可知 $P\left(\limsup_{n\to\infty}\frac{|X_n|}{\sqrt{\log n}} = \sqrt{2}\right) = 1.$ □

§ 7.4 第五章习题

回顾前面笔记没提到的 de Movir-Laplace 定理 (可以用 Lindberg-Lévy 推导):

定理 7.4.1. de Movir-Laplace

在 n 重 Bernoulli 试验中, 事件 A 在每次试验中出现的概率为 $p(0<p<1)$, 记 y_n 为 n 次试验中事件 A 出现的次数, $y_n^*=\dfrac{y_n-np}{\sqrt{npq}}$ (**标准化**), 则

$$\lim_{n\to\infty}P(y_n^*\le y)=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x=\Phi(y),\forall y\in\mathbb{R}.$$

注: 我们有

$$\begin{aligned}P(k_1\le y_n\le k_2)&=P\left(\frac{k_1-np}{\sqrt{npq}}\le\frac{y_n-np}{\sqrt{npq}}\le\frac{k_2-np}{\sqrt{npq}}\right)\\&\approx\Phi\left(\frac{k_2-np}{\sqrt{npq}}\right)-\Phi\left(\frac{k_1-np}{\sqrt{npq}}\right).\end{aligned}$$

有时候修正 0.5 更精确: 采用 $\Phi\left(\dfrac{k_2-np+0.5}{\sqrt{npq}}\right)-\Phi\left(\dfrac{k_1-np-0.5}{\sqrt{npq}}\right)$.

当 $np>5$ 且 $n(1-p)>5$ 时, 用正态分布近似二项分布.

例 7.4.2

某车间有同型号的机床 200 台, 在 1 小时内每台机床约有 70% 时间是工作的. 假定每个机床工作是相互独立的, 工作时每台机床要消耗电能 15 千瓦, 问至少要多少电能才可以有 95% 可能性保证此车间正常生产? $(\Phi(1.645)=0.95)$.

解: $n=200,p=0.7,\beta=95\%$, 记 y 为台数. 则

$$P\left(\frac{y_n-np}{\sqrt{npq}}\le\frac{y}{\sqrt{npq}}-\frac{np}{\sqrt{npq}}\right)\approx\Phi\left(\frac{y-np-0.5}{\sqrt{npq}}\right)\ge 95\%,$$

等价于

$$\Phi\left(\frac{y-140-0.5}{\sqrt{42}}\right)\ge 95\%,\Rightarrow\frac{y-140-0.5}{\sqrt{42}}\ge 1.645,\Rightarrow y\ge 150.16.$$

于是 $15y\ge 2252.4$(千瓦), 至少要 2252.4 千瓦的电能才可以有 95% 可能性保证此车间正常生产.

例 7.4.3

设 $X_2,X_3,\cdots$ 是 i.i.d.r.v., 满足

$$P(X_n=n)=P(X_n=-n)=\frac{1}{2n\log n},P(X_n=0)=1-\frac{1}{n\log n}.$$

通过验证 $\dfrac{1}{n}\displaystyle\sum_{i=2}^{n}X_i$ 依概率收敛到 0 而不几乎处处收敛到 0, 来证明这个序列满足弱大数定律, 但不满足强大数定律.

提示: 设 $S_n=X_1+\cdots+X_n$, 则

$$E(S_n^2)=\sum_{i=2}^{n}\frac{i}{\log i}\le\frac{n^2}{\log n}.$$

因此

$$P\left(\frac{S_n}{n} \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{D}\left(\frac{S_n}{n}\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left(\frac{S_n}{n}\right)^2 = \frac{1}{\varepsilon^2 \log n} \to 0(n \to \infty).$$

另一方面, $\sum\limits_{i=2}^{n} P(|X_i| \geq i) = \sum\limits_{i=2}^{\infty} \frac{1}{i \log i} = +\infty$, 由 Borel-Cantelli 引理可知 $P\left(\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}[|X_n| \geq n]\right) = 1$. 对某个 i, 我们有 $|X_i| = |S_i - S_{i-1}| \geq i$, 推出 $\frac{S_n}{n}$ 几乎必然不收敛. □

例 7.4.4

构造一列 i.i.d.r.v$\{X_r : r \geq 1\}$, 满足下列条件:

(1)$EX_r = 0, \forall r \geq 1$;

(2)$\frac{1}{n}\sum\limits_{r=1}^{n} X_r \xRightarrow{a.s.} -\infty(n \to \infty)$.

提示: 构造

$$P(X_n = -n) = 1 - \frac{1}{n^2}, P(X_n = n^3 - n) = \frac{1}{n^2}.$$

那么 X_n 的期望为 0. 但是

$$\sum_{n=1}^{\infty} P\left(\frac{X_n}{n} \neq -1\right) = \sum_{n=1}^{\infty}\frac{1}{n^2} < \infty.$$

由 Borel-Cantelli 引理, $P(X_n/n \to -1) = 1$.

根据数学分析, 若 $x_n \to -1$, 则 $\frac{1}{n}\sum\limits_{i=1}^{n} x_i \to -1$, 从而推出 (2). □

§ 7.5 第六章习题

7.5.1 条件期望

例 7.5.1

设 $X, Y \in L^2(\Omega, \mathscr{F}, P)$ 满足 $\mathbb{E}(Y|X) = X, \mathbb{E}(X|Y) = Y$, 则 $X = Y$, a.s..

证明: 注意到

$$\begin{aligned}
0 \leq \mathbb{E}(X-Y)^2 &= \mathbb{E}X^2 - 2\mathbb{E}XY + \mathbb{E}Y^2 \\
&= \mathbb{E}X^2 - 2\mathbb{E}[\mathbb{E}(XY|X)] + \mathbb{E}Y^2 \quad \text{【性质①】} \\
&= \mathbb{E}X^2 - 2\mathbb{E}(X\mathbb{E}(Y|X)) + \mathbb{E}Y^2 \quad \text{【性质⑦】} \\
&= \mathbb{E}Y^2 - \mathbb{E}X^2.
\end{aligned}$$

同理, 如果对上面第二行的式子改为作用 Y 的条件期望, 可得 $\mathbb{E}(X-Y)^2 = EX^2 - EY^2$. 一个数同时等于另一个数与它的相反数, 则这个数只能为 0, 即 $\mathbb{E}(X-Y)^2 = 0$, 则 $X = Y$, a.s..

例 7.5.2

设 $X, Y \in L^1(\Omega, \mathscr{F}, P)$ 满足 $\mathbb{E}(Y|X) = X, \mathbb{E}(X|Y) = Y$, 则 $X = Y$, a.s..

提示: 只需考虑 $\mathbb{E}(X-Y)(\arctan X - \arctan Y)$, 这个依然是非负的, 而且 $\arctan x$ 有界, 则 $(X-Y)(\arctan X - \arctan Y)$ 必定可积.

例 7.5.3

已知 X 是一可积 r.v., $\mathscr{C}$ 是 $\mathscr{F}$ 的子 sigma 代数. 令 $Y=\mathbb{E}(X|\mathscr{C})$, 假定 X 与 Y 同分布, 证明:

(1) 若 X 平方可积, 则 $X=Y$, a.s..

(2) 若 $\forall a,b\in\mathbb{R}, a<b$, 有 $(X\vee a)\wedge b=(Y\vee a)\wedge b$, a.s., 则 $X=Y$, a.s..

证明: (1) 注意到

$$\begin{aligned}\mathbb{E}(X-Y)^2&=\mathbb{E}X^2-2\mathbb{E}XY+\mathbb{E}Y^2=2\mathbb{E}Y^2-2\mathbb{E}XY &&\text{【同分布】}\\&=2\mathbb{E}Y^2-2\mathbb{E}[\mathbb{E}(XY|\mathscr{C})] &&\text{【性质①】}\\&=2\mathbb{E}Y^2-2\mathbb{E}(Y\mathbb{E}(X|\mathscr{C})) &&\text{【性质⑦】}\\&=2\mathbb{E}Y^2-2\mathbb{E}Y^2 &&\text{【题目条件】}\\&=0,\Rightarrow X=Y,a.s..\end{aligned}$$

(2) 下证 $\mathbb{E}(X\vee a|\mathscr{C})=Y\vee a(=\mathbb{E}(X|\mathscr{C})\vee a),\forall a\in\mathbb{R}$, 则

$$\mathbb{E}[(X\vee a)\wedge b|\mathscr{C}]=(Y\vee a)\wedge b,a.s.,\forall a,b\in\mathbb{R},a<b.$$

(把并的证明推广到交的证明, 结果是一样的, 所以下面只证明并的情况)

首先考虑到函数 $f(x)=x\vee a$ 是凸的, 根据 Jensen 不等式有

$$\mathbb{E}(X\vee a|\mathscr{C})\geq\mathbb{E}(X|\mathscr{C})\vee a,a.s..$$

只需证 $P[\mathbb{E}((X\vee a)|\mathscr{C})>(\mathbb{E}(X|\mathscr{C})\vee a)]=0$. 事实上,

$$\begin{aligned}\mathbb{E}[(X\vee a)|\mathscr{C}-(\mathbb{E}(X|\mathscr{C})\vee a)]&=\mathbb{E}[\mathbb{E}(X\vee a-\mathbb{E}(X|\mathscr{C})\vee a|\mathscr{C})] &&\text{【性质①】}\\&=\mathbb{E}(X\vee a-\mathbb{E}(X|\mathscr{C})\vee a) &&\text{【线性性】}\\&=\mathbb{E}X\vee a-\mathbb{E}(\mathbb{E}(X|\mathscr{C})\vee a)\\&=\mathbb{E}X\vee a-\mathbb{E}Y\vee a=0. &&\text{【同分布】}\end{aligned}$$

证明完毕. □

例 7.5.4

设 ξ,η 是 $(\Omega,\mathscr{F},P)$ 中的随机变量, $\mathbb{E}(\xi|\mathscr{C})=\eta$, 且 $\mathbb{E}\xi^2=\mathbb{E}\eta^2<+\infty$. 证明: $\xi=\eta$, a.s..

证明: 注意

$$\begin{aligned}\mathbb{E}(\xi-\eta)^2&=\mathbb{E}\xi^2+\mathbb{E}\eta^2-2\mathbb{E}(\xi\eta)\\&=2\mathbb{E}\eta^2-2\mathbb{E}(\mathbb{E}(\xi\eta|\mathscr{C}))\\&=2\mathbb{E}\eta^2-2\mathbb{E}(\eta\mathbb{E}(\xi|\mathscr{C}))\\&=2\mathbb{E}\eta^2-2\mathbb{E}\eta^2=0.\end{aligned}$$

7.5.2 一致可积

例 7.5.5

设 $\{\xi_n\}$ 是一致可积随机变量序列, 则有 $\lim\limits_{n\to\infty}\mathbb{E}\left(\frac{1}{n}\sup\limits_{1\leq k\leq n}|\xi_k|\right)=0.$

证明: (1) 我们有

$$\begin{aligned}
&\mathbb{E}(|X|\vee|Y|I_{[|X|\vee|Y|\geq C]})\\
&=\int_{[|X|\geq|Y|]}|X|\vee|Y|I_{[|X|\vee|Y|\geq C]}\mathrm{d}P+\int_{[|X|\leq|Y|]}|X|\vee|Y|I_{[|X|\vee|Y|\geq C]}\mathrm{d}P\\
&=\int_{[|X|\geq|Y|]}|X|I_{[|X|\geq C]}\mathrm{d}P+\int_{[|X|\leq|Y|]}|Y|I_{[|Y|\geq C]}\mathrm{d}P\\
&\leq\mathbb{E}(|X|I_{[|X|\geq C]})+\mathbb{E}(|Y|I_{[|Y|\geq C]}).
\end{aligned}$$

(2) 注意到

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n}\sup_{1\leq k\leq n}|\xi_k|\right)&=\mathbb{E}\left(\frac{1}{n}\sup_{1\leq k\leq n}|\xi_k|I_{[\sup\limits_{1\leq k\leq n}|\xi_k|\geq C]}\right)+\mathbb{E}\left(\frac{1}{n}\sup_{1\leq k\leq n}|\xi_k|I_{[\sup\limits_{1\leq k\leq n}|\xi_k|<C]}\right)\\
&\leq\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(|\xi_i|I_{[|\xi_i|\geq C]})+\frac{C}{n}\leq\sup_{\xi\in\mathcal{H}}\mathbb{E}(|\xi|I_{[|\xi|\geq C]})+\frac{C}{n}.
\end{aligned}$$

由一致可积性, 对 $\varepsilon>0$, 存在 C_0, 当 $C>C_0$ 时, $\left|\sup\limits_{\xi\in\mathcal{H}}\mathbb{E}(|\xi|I_{[|\xi|\geq C]})\right|<\varepsilon/2$.

对上述 ε, 取 $N=\dfrac{2C}{\varepsilon}$, 则当 $n>N$ 时, $\dfrac{C}{n}<\dfrac{\varepsilon}{2}$. 所以 $\mathbb{E}\Big(\dfrac{1}{n}\sup\limits_{1\leq k\leq n}|\xi_k|\Big)<\varepsilon$. □

例 7.5.6

设 $\mathcal{H}\subset L^1(\Omega,\mathscr{F},P)$, 若 $\mathcal{H}$ 满足:

$$A_n\in\mathcal{F},A_n\searrow\varnothing\Rightarrow\lim_{n\to\infty}\sup_{\xi\in\mathcal{H}}\int_{A_n}|\xi|\mathrm{d}P=0,$$

则对任给 $\varepsilon>0$, 存在 $\delta>0$, 使得

$$A\in\mathscr{F},P(A)<\delta\Rightarrow\sup_{\xi\in\mathcal{H}}\int_A|\xi|\mathrm{d}P\leq\varepsilon.$$

证明: (反证) 若不然, 存在 $\varepsilon>0$, 对 $\delta_n=\dfrac{1}{2^n}$, 都存在 $B_n\in\mathscr{F}$, 使得

$$P(B_n)<\delta_n=\frac{1}{2^n}\text{且}\sup_{\xi\in\mathcal{H}}\int_{B_n}|\xi|\mathrm{d}P\geq\varepsilon.$$

由 Borel-Cantelli 引理, 根据 $\sum\limits_{n=1}^{\infty}P(B_n)<\infty$, 可得 $P\Big(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}B_k\Big)=0$.

取 $C_n=\bigcup\limits_{k=n}^{\infty}B_k$, 则 $\sup\limits_{\xi\in\mathcal{H}}\int_{C_n}|\xi|\mathrm{d}P\geq\varepsilon$, 且 $P\Big(\bigcap\limits_{m=1}^{\infty}C_m\Big)=0$.

取 $A_n=C_n\setminus\bigcap\limits_{m=1}^{\infty}C_m$, 则 A_n 单调下降趋于 $\varnothing$, 但对任意 $n,\sup\limits_{\xi\in\mathcal{H}}\int_{A_n}|\xi|\mathrm{d}P\geq\varepsilon$, 这与条件矛盾. □