



应用随机过程

吴卫星 & 邓军

对外经济贸易大学

金融学院

1/126



GoBack

FullScreen

Close

Quit



第2章： 概率论预备知识

- 概率空间
- 随机变量与分布函数；
- 数字特征、矩母函数与特征函数
- 收敛性
- 独立性与条件期望

2/126



GoBack

FullScreen

Close

Quit

学科历史

古典概率论

- 瑞士数学家 Jacob Bernoulli 受朋友信件的启发，从1685年起发表关于赌博游戏中输赢次数问题的论文，后来写成巨著《猜度术》。



- Bernoulli分布和弱大数定律



3/126



GoBack

FullScreen

Close

Quit

- 古典概率（又称传统概率、古典概型、Laplace概率）假定事件集为有限，每个事件发生的机会相同。
- 一般古典概率教学中会大量使用排列组合知识



4/126



GoBack

FullScreen

Close

Quit



公理化

- 如何定义概率，如何把概率论建立在严格的逻辑基础上，是概率理论发展的困难所在，对这一问题的探索一直持续了3个世纪。
- 20世纪初完成的Lebesgue测度与积分理论及随后发展的抽象测度和积分理论，为概率公理体系的建立奠定了基础。
- Lebesgue 和他的著作Lecons sur l'integration et la recherche des fonctions primitives, 1904

5/126

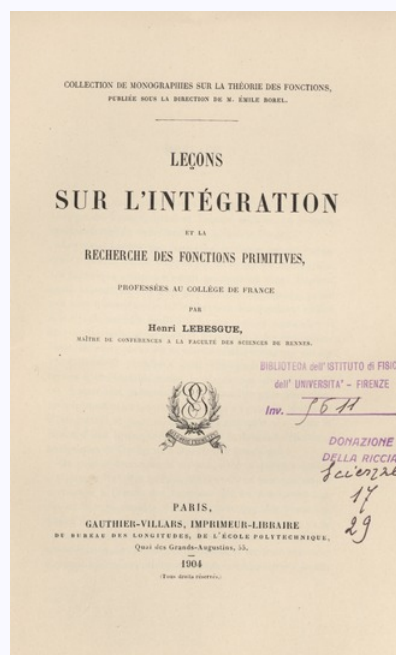


GoBack

FullScreen

Close

Quit



6/126

- 1933年，苏联数学家A. Kolmogorov在他的《概率论基础》一书中第一次给出了概率的测度论的定义和一套严密的公理体系。他的公理化方法成为现代概率论的基础，使概率论成为严谨的数学分支，对概率论的迅速发展起了积极的作用。



GoBack

FullScreen

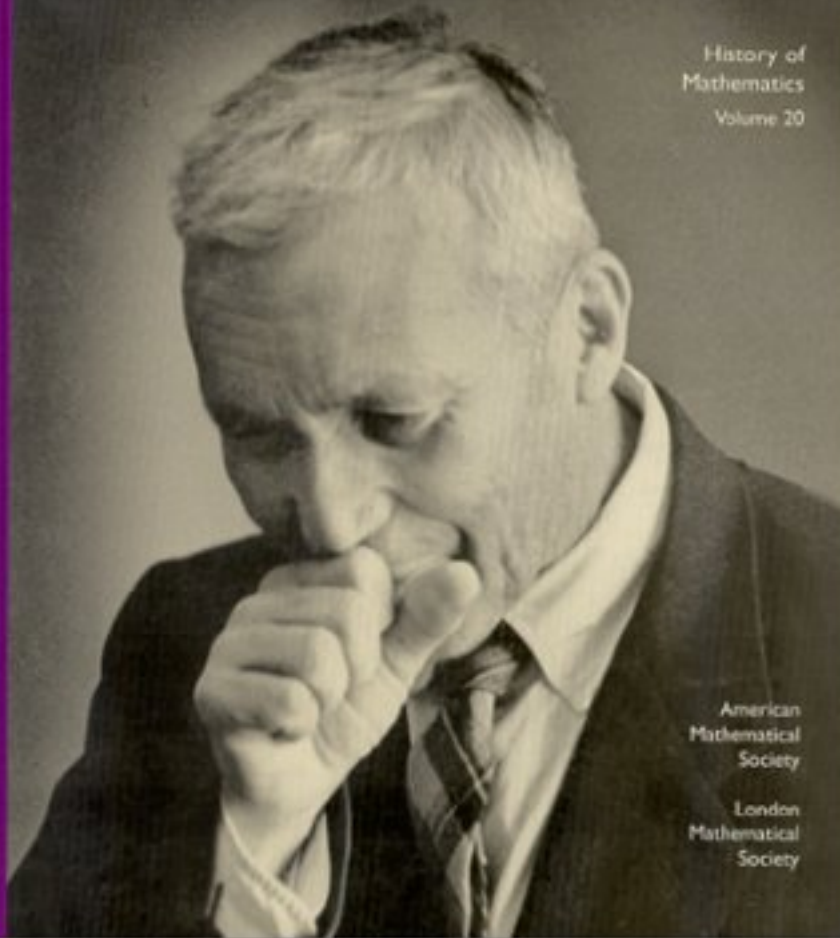
Close

Quit

KOLMOGOROV

IN PERSPECTIVE

History of
Mathematics
Volume 20



American
Mathematical
Society

London
Mathematical
Society



7/126



GoBack

FullScreen

Close

Quit



随机过程

- **随机过程**是指以时间为参数的一族随机变量。由于现代概率论对随机变量赋予了可测结构，所以对随机过程的分析比对一般二元函数的分析要复杂得多。一系列理论由此产生。
- 随机过程领域的创始人包括A. Kolmogorov, A. Markov, P. Lévy 和 N. Wiener等人。
- 1953年, J. Doob 的名著《随机过程论》问世, 它系统且严格地叙述了随机过程的基本理论。

8/126



GoBack

FullScreen

Close

Quit

钟开莱 Kai Lai Chung 1917年9月19日－2009年6月2日



9/126



GoBack

FullScreen

Close

Quit

§2.1 概率空间

随机试验是概率论的基本概念，试验的结果事先不能准确地预言，但具有如下三个特性：

- (1) 可以在相同的条件下重复进行；
- (2) 每次试验的结果不止一个，但预先知道试验的所有可能的结果；
- (3) 每次试验前不能确定哪个结果会出现。

$$(\Omega, \mathcal{F}, P)$$



10/126



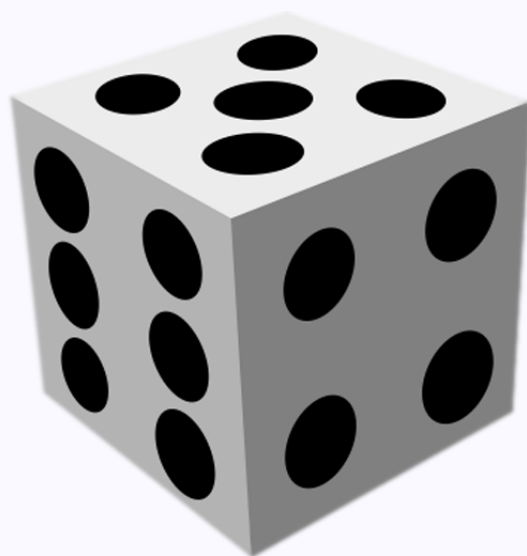
GoBack

FullScreen

Close

Quit

样本点(ω): 随机试验的基本结果
样本空间(Ω): 随机试验所有可能结果组成的集合
基本事件: Ω 中的样本点 ω
必然事件: 样本空间 Ω
不可能事件: 空集 $\emptyset$
事件: 由基本事件组成的 Ω 中的子集 A





样本空间 Ω

- 样本空间 Ω 是所考虑的概率论问题的出发点，涵盖了人们研究的不确定现象的一切可能性。
- 样本空间的元素 $\omega \in \Omega$ ，称为样本点或基本事件，有时也称为状态，从而样本空间有时也可称为状态空间。人们研究的不确定现象的一切可能性。
- 在同一个问题的研究框架下，通常只是针对同一个样本空间

12/126



GoBack

FullScreen

Close

Quit

有限样本空间

- 例1. 掷硬币实验。掷一次硬币，其样本空间为 $\Omega \in \{H(Head), T(Tail)\}$.

试问：独立地掷 n 次硬币，其样本空间是什么？



- 例2. 明天某股票收盘价的样本空间是什么？



无限样本空间

- 例3. 独立地掷无穷多次硬币，其样本空间是什么？

$$\Omega = \{\text{H和T组成的一切无穷序列}\}.$$

- 例4. 在单位区间 $[0, 1]$ 中随机选择一个数。

$$\Omega = [0, 1].$$

- 如果对每个样本 ω 都赋予了概率值，则知道了所有样本事件集发生的概率。古典概型即属于此类。
- 共同点： Ω 中的元素无穷多且不可数。因此，对单个样本赋予概率的刻画方式将不再有效。我们需要新的方式去在 Ω 上描述概率。
- **思考：随机等可能的抽取一个自然数???**

14/126



GoBack

FullScreen

Close

Quit



研究对象客观的抽象： σ -代数

- 子集定义： 样本空间 Ω 中元素构成的集合称为子集。例如： $\Omega = \{\text{随机过程课程上面所有的学生}\}$ ， $A = \{\text{学生1, 学生2, 学生3}\}$ 。则， A 为一个子集。
- 子集族 $\mathcal{F}$ 定义： 样本空间 Ω 的子集构成的集合叫做集族。例如： $A = \{\text{学生1, 学生2, 学生3}\}$ ， $B = \{\text{学生1, 学生4, 学生5}\}$ ， $C = \{\text{学生7, 学生8, 学生10}\}$ 。则： $\mathcal{F} := \{A, B, C\}$ 为一个集族。
- 思考：集族的意义？并不是什么样的集族对我们研究都是有用的！

15/126



GoBack

FullScreen

Close

Quit



定义 2.1.1 设 Ω 是一个样本空间(或任意一个集合),
 $\mathcal{F}$ 是 Ω 的某些子集组成的集合族. 如果满足:

- (1) $\Omega \in \mathcal{F}$;
- (2) 若 $A \in \mathcal{F}$, 则 $A^c = \Omega \setminus A \in \mathcal{F}$ (对取补封闭);
- (3) 若 $A_n \in \mathcal{F}, n = 1, 2, \dots$, 则 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (对可列并封闭);

则称 $\mathcal{F}$ 为 σ 代数. $(\Omega, \mathcal{F})$ 称为**可测空间**, $\mathcal{F}$ 中的元素称为**随机事件**.

注: 如果 $\mathcal{F}$ 是事件的 σ 代数, 则(1) $\emptyset \in \mathcal{F}$; (2) 当 $A_n \in \mathcal{F}, n = 1, 2, \dots$, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

16/126



GoBack

FullScreen

Close

Quit



σ -代数的例子

- 最小的 σ -代数: 平凡 (trivial) σ -代数 $\{\emptyset, \Omega\}$.
- 最大的 σ -代数: 全体子集 $2^\Omega := \{A : A \subset \Omega\}$.
- 分割形成的 σ -代数: 令 $A_1, A_2, \dots, A_n$ 为 Ω 的不相交的子集, 并有 $\bigcup_{i=1}^n A_i = \Omega$. 则由 $A_1, A_2, \dots, A_n$ 进行有限并、补运算产生一个 σ -代数, 其中包含 2^n 个集合。
- $\mathcal{F} = \{A, A^c, \emptyset, \Omega\}$.

17/126



GoBack

FullScreen

Close

Quit



例 2.1.2 在编号为 $1, 2, \dots, n$ 的 n 只股票中取一只.
考虑股票的编号, 则全体基本事件为

$$A_k = \{k\}, \quad k = 1, 2, \dots, n.$$

样本空间为:

$$\Omega = \{1, 2, \dots, n\}.$$

构造如下的事件:

$$A_{k,s} = A_k \cup A_s$$

$$A_{i,k,s} = A_i \cup A_k \cup A_s$$

$$\vdots$$

$$A_{i_1,i_2,\dots,i_j} = A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j}$$

$$\vdots$$

$$A_{i_1,i_2,\dots,i_{n-1}} = A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_{n-1}}$$

可以验证: $\{\emptyset, \Omega, A_k, A_{i,k}, A_{i_1,i_2,\dots,i_j}, \dots, A_{i_1,i_2,\dots,i_{n-1}}\}$ 构成一个 σ 代数。



19/126



GoBack

FullScreen

Close

Quit



例 2.1.3 仅考虑股票是深股或者沪股，则基本事件为

$$A_1 = \{\text{取到深股}\}, A_2 = \{\text{取到沪股}\}, \Omega = \{A_1, A_2\}$$

则，

$$\mathcal{F} = \{\emptyset, \Omega, A_1, A_2\}$$

构成一个 σ 代数。

20/126



GoBack

FullScreen

Close

Quit



σ -代数的生成

- 对于 Ω 的一个子集族 $\mathcal{S}$, 如果存在 Ω 上的 σ -代数 $\mathcal{F}$ 使得
 - 1. $\mathcal{S} \subset \mathcal{F}$;
 - 2. 对于任意包含 $\mathcal{S}$ 的 Ω 上的 σ -代数 $\mathcal{F}'$, 都有 $\mathcal{F} \subset \mathcal{F}'$,则称 $\mathcal{F}$ 是由子集系 $\mathcal{S}$ 生成的 (最小的) σ -代数, 记作

$$\mathcal{F} := \sigma(\mathcal{S}).$$

- 命题: 对 Ω 的任何子集系 $\mathcal{S}$, 都存在 $\mathcal{F} := \sigma(\mathcal{S})$.
- 命题: 如果 $\mathcal{S}$ 本身是 Ω 上的一个 σ -代数, 则 $\sigma(\mathcal{S}) = \mathcal{S}$.

21/126



GoBack

FullScreen

Close

Quit



σ -代数的生成: 例子

例 2.1.4 • $\Omega = \{\omega_1, \omega_2, \dots, \omega_4\}$, $\mathcal{S} = \{\{\omega_1\}, \{\omega_2, \omega_3\}\}$ 。
则: 由 $\mathcal{S}$ 生成的 σ -代数为:

$$\begin{aligned}\sigma(\mathcal{S}) = \{ & \emptyset, \Omega, \{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}, \\ & \{\omega_1, \omega_2, \omega_3\}, \{\omega_4\}, \\ & \{\omega_2, \omega_3\}, \{\omega_1, \omega_4\} \}\end{aligned}$$

• 如果 $\mathcal{S} = \{\{\omega_1\}, \{\omega_2\}\}$ 。则: 由 $\mathcal{S}$ 生成的 σ -代数为:

$$\begin{aligned}\sigma(\mathcal{S}) = \{ & \emptyset, \Omega, \{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}, \\ & \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \\ & \{\omega_2\}, \{\omega_1, \omega_3, \omega_4\} \}\end{aligned}$$

• 思考: 当 $\mathcal{S}$ 不同时, 产生的 σ -代数的区别的意义是什么??



可测空间

- Ω 与其上的一个 σ -代数 $\mathcal{F}$ 形成的二元组 $(\Omega, \mathcal{F})$ 称为一个**可测空间**。 $\mathcal{F}$ 中的任意集合 A 称为 $(\Omega, \mathcal{F})$ 中的**可测集**，或 $\mathcal{F}$ -可测集。
- 一句有用的废话：对于一个 σ -代数可测的集合 A ，对于更小的 σ -代数可能会不可测。
- 注意：可测空间的定义中，并不需要引入测度。在定义可测空间之后，我们可以在其上开始定义测度（或概率），甚至可以定义多个不同的测度。

23/126



GoBack

FullScreen

Close

Quit

可测空间

- 一个比喻：可测空间好比一堆鹅卵石，测度好比我们怎么去量化这堆石头，我们可以测度她们质量，也可以测度她们的个数，还可以测度她们的体积，密度等等。这些不同的测度方法就类似于我们在可测空间上面，可以定义不同的测度。





引入 σ -代数的意义

- 突破古典概率对样本空间的限制，着眼于无穷空间
- 选择“可列并”作为突破口，引入可测结构
- 与相关领域（尤其是实分析）中的进展相互印证启发
- 作为现代概率论体系的基石，最终集大成而获得了大数定律和中心极限定理
- 为金融信息的“信息流”概念做准备

25/126



GoBack

FullScreen

Close

Quit



可测空间 $(\Omega, \mathcal{F})$ 的性质

1. $\emptyset \in \mathcal{F}$
2. 对可列交运算封闭, 如果 $A_i \in \mathcal{F}$, for all $i = 1, 2, \dots$, 则有

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

3. 对有限交封闭, 如果 $A_i \in \mathcal{F}$, for all $i = 1, 2, \dots$, 则有

$$\bigcup_{i=1}^n A_i \in \mathcal{F}, \quad \bigcap_{i=1}^n A_i \in \mathcal{F}.$$

4. 对差运算封闭, 如果 $A, B \in \mathcal{F}$, for all $i = 1, 2, \dots$, 则有

$$A - B := A \cap B^c \in \mathcal{F}.$$

26/126



GoBack

FullScreen

Close

Quit

不可测集合 — Many

$$\mathcal{F} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}.$$

简单可见: $\{1\}$ 不是 $\mathcal{F}$ 可测集合。



27/126



GoBack

FullScreen

Close

Quit



定义 2.1.5 设 $(\Omega, \mathcal{F})$ 是可测空间, $P(\cdot)$ 是定义在 $\mathcal{F}$ 上的实值函数.如果

- (1) $P(\Omega) = 1$;
- (2) $\forall A \in \mathcal{F}, 0 \leq P(A) \leq 1$;
- (3) 对两两互不相容事件 $A_1, A_2, \dots$, (即当 $i \neq j$ 时, $A_i \cap A_j = \emptyset$)有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

则称 P 是 $(\Omega, \mathcal{F})$ 上的概率, $(\Omega, \mathcal{F}, P)$ 称为概率空间, $\mathcal{F}$ 中的元素称为事件, $P(A)$ 称为事件 A 的概率.



事件的概率有如下性质:

(1) 若 $A, B \in \mathcal{F}$, 则 $P(A \cup B) + P(A \cap B) = P(A) + P(B)$.

(2) 若 $A, B \in \mathcal{F}$, 且 $A \subset B$, 则 $P(B - A) = P(B) - P(A)$ (可减性).

(3) 若 $A, B \in \mathcal{F}$, 且 $A \subset B$, 则 $P(A) \leq P(B)$ (单调性).

(4) 若 $A_n \in \mathcal{F}, n \geq 1$, 则 $P(\bigcup_{n \geq 1} A_n) \leq \sum_{n \geq 1} P(A_n)$.

(5) 从下连续:

若 $A_n \in \mathcal{F}$ 且 $A_n \uparrow A \in \mathcal{F}$, 则 $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.

(6) 从上连续:

若 $A_n \in \mathcal{F}$ 且 $A_n \downarrow A \in \mathcal{F}$, 则 $P(A) = \lim_{n \rightarrow \infty} P(A_n)$



例 2.1.6 设某股票一天的成交笔数为 m , 基本事件为 $\{m\}$, 样本空间 $\mathcal{F}$ 是 Ω 的一切子集组成的集族, 则 $\mathcal{F}$ 是一个 Ω 代数. 定义 $P(\emptyset) = 0$, 并对 $A \in \mathcal{F}$ 令

$$P(A) = \sum_{k \in A} e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda > 0.$$

证明: P 为可测空间 $(\Omega, \mathcal{F}, P)$ 上的概率测度.

30/126



GoBack

FullScreen

Close

Quit



条件概率

定义 2.1.7 在概率空间 $(\Omega, \mathcal{F}, P)$ 上面, 如果 $A, B \in \mathcal{F}$ 且 $P(B) > 0$, 则定义

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

称为已知事件 B 发生的条件下, 事件 A 发生的条件概率.

条件概率告诉了我们已经给定了一个事件下, 另外一个事件发生的概率。例如下面图片所示, 可以设想在一个样本空间中事件发生就是在平板上方掉落的球。如果球落在了横板上面, 说明该事件就发生了。图片中红色挡板就是事件 A , 蓝色的挡板就是的事件 B 。在给定事件 A 条件下, 事件 B 的条件概率就是下落的小球首先已经触碰到红色挡板 A 的条件下, 并且再次触碰到挡板 B 的概率。该图例的动态演示可以

31/126



GoBack

FullScreen

Close

Quit

在 <http://setosa.io/conditional/> 查看.

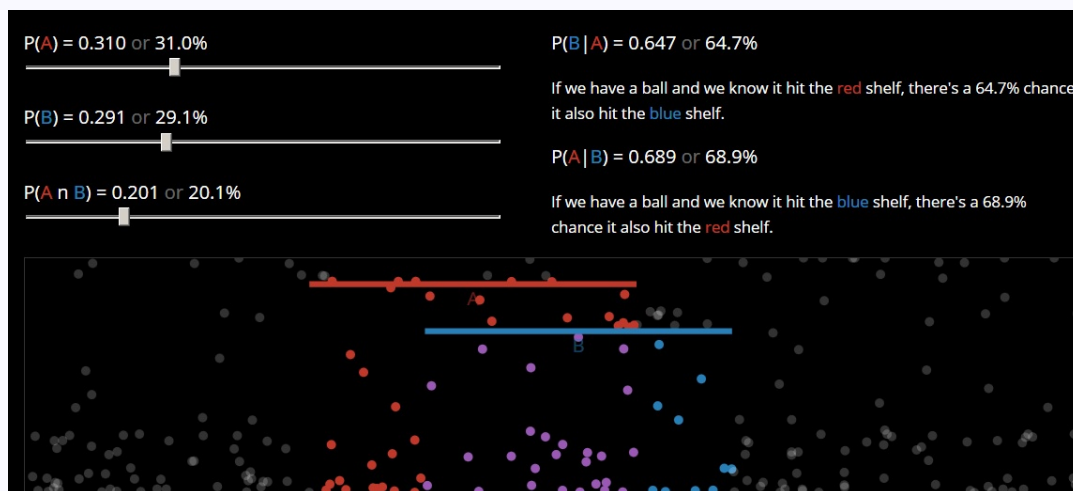


Figure 2.1: 条件概率

定理 2.1.8 在概率空间 $(\Omega, \mathcal{F}, P)$ 上面, $B \in \mathcal{F}$, 且 $P(B) > 0$, 则对所有的 $A \in \mathcal{F}$, 有 $P(A|B)$ 相对应, 且函数 $P(\cdot|B)$ 满足如下的三条公理:

- $\forall A \in \mathcal{F}, 0 \leq P(A|B) \leq 1$





- $P(\Omega|B) = 1$
- $A_i \in \mathcal{F}, i = 1, 2, \dots, n$, and $A_i \cap A_j = \emptyset, i \neq j$, then

$$P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B).$$

定义 2.1.9 记 $P_B := P(\cdot|B)$, 则 P_B 是可测空间 $(\Omega, \mathcal{F})$ 上的概率, 称 $(\Omega, \mathcal{F}, P_B)$ 是条件概率空间.

定理 2.1.10 设 A 是概率空间 $(\Omega, \mathcal{F}, P_B)$ 上的正概率事件, $B \in \mathcal{F}$, 且 $P_A(B) > 0$, 则对任意 $C \in \mathcal{F}$ 有

$$P_A(C|B) = P(C|A \cap B).$$

33/126



GoBack

FullScreen

Close

Quit



Proof We have

$$\begin{aligned} P_A(C|B) &= \frac{P_A(C \cap B)}{P_A(B)} = \frac{P(C \cap B|A)}{P(B|A)} \\ &= \frac{P(ABC)}{P(A)} / \frac{P(AB)}{P(A)} = \frac{P(ABC)}{P(AB)} = P(C|A \cap B). \end{aligned}$$

34/126



GoBack

FullScreen

Close

Quit



全概率公式和Bayes公式

定理 2.1.11 在概率空间 $(\Omega, \mathcal{F}, P)$ 上, 若

- $A_i \in \mathcal{F}$ and $P(A_i) > 0$ for all $i = 1, 2, \dots$
- $\bigcup_{i=1}^{\infty} A_i = \Omega$, $A_i \cap A_j = \emptyset$.

则对任意的 $B \in \mathcal{F}$, 有

$$P(B) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i)$$
$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)}.$$

$B : \{outcome\}, \quad A_j : \{condition\}$

35/126



GoBack

FullScreen

Close

Quit



36/126

|| ◀ ▶ ||

◀ ▶

GoBack

FullScreen

Close

Quit



例 2.1.12 高射炮向敌机发射三发炮弹，每弹击中与否相互独立且每发炮弹击中的概率均为0.3，又知敌机若中一弹，坠毁的概率为0.2，若中两弹，坠毁的概率为0.6，若中三弹，敌机必坠毁。求敌机坠毁的概率。

$B = \{ \text{敌机坠毁} \}, A_i = \{ \text{敌机中了} i \text{ 个子弹} \}$

因此，我们可以计算

$$\begin{aligned} P(B) &= \sum_{i=0}^3 P(A_i)P(B|A_i) \\ &= (1 - 0.3)^3 * 0 + C_3^1 * 0.3 * (1 - 0.3)^2 * 0.2 \\ &\quad + C_3^2 * 0.3^2 * (1 - 0.3) * 0.6 + C_3^3 * 0.3^3 * 1 \\ &= 0.2286. \end{aligned}$$

37/126



GoBack

FullScreen

Close

Quit

计算在敌机坠毁的条件下，该机中了一发炮弹的概率。

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1)P(B|A_1)}{\sum_{i=0}^3 P(A_i)P(B|A_i)} \\ &= \frac{C_3^1 * 0.3 * (1 - 0.3)^2 * 0.2}{0.2286} = 0.385826772. \end{aligned}$$



38/126



GoBack

FullScreen

Close

Quit



§2.2 随机变量和分布函数

定义 2.2.1 设 $\Omega = \mathbb{R}$. 由所有半无限区间 $(-\infty, x)$ 生成的 σ 代数(即包含集族 $\{(-\infty, x), x \in \mathbb{R}\}$ 的最小 σ 代数)称为 $\mathbb{R}$ 上的 Borel σ 代数, 记为 $\mathcal{B}(\mathbb{R})$, 其中的元素称为 Borel 集合. 类似地, 可定义 $\mathbb{R}^n$ 上的 Borel σ 代数 $\mathcal{B}(\mathbb{R}^n)$.

定义 2.2.2 设 $(\Omega, \mathcal{F}, P)$ 是(完备的)概率空间, X 是定义在 Ω 上取值于实数集 $\mathbb{R}$ 的函数, 如果对任意实数 $x \in \mathbb{R}$, $\{\omega : X(\omega) < x\} \in \mathcal{F}$, 则称 $X(\omega)$ 是 $\mathcal{F}$ 上的随机变量, 简称为随机变量.

$$F(x) = P(\omega : X(\omega) < x), \quad -\infty < x < \infty$$

称为随机变量 X 的分布函数.



分布函数 $F(x)$ 函数的性质。

- $F(x)$ 是单调不减函数
- $0 \leq F(x) \leq 1, \lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1.$
- $F(x)$ 是左连续函数, 即

$$\forall x \in R, \quad F(x-) = F(x).$$

40/126



GoBack

FullScreen

Close

Quit



Proof 注意到:

$$\cup_{n=1}^{\infty} \{X < x - \frac{1}{n}\} = \{X < x\}$$

由概率测度的从上连续可得:

$$\begin{aligned} F(x) &= P(\{X < x\}) = P(\cup_{n=1}^{\infty} \{X < x - \frac{1}{n}\}) \\ &= \lim_{n \rightarrow +\infty} P(\{X < x - \frac{1}{n}\}) = F(x-) \end{aligned}$$

41/126



GoBack

FullScreen

Close

Quit



如果存在函数 $f(x)$, 满足

$$F(x) = \int_{-\infty}^x f(t)dt$$

则称 $f(x)$ 为随机变量 X 或其分布函数 $F(x)$ 的分布密度.

如果 X 具有分布密度, 则称 X 为连续型随机变量; 如果 X 最多以正概率取可数多个值, 则称 X 为离散型随机变量.

42/126



GoBack

FullScreen

Close

Quit



常用的两种类型随机变量：

(1) 离散型随机变量 X 的概率分布用分布列描述：

$$p_k = P(X = x_k), \quad k = 1, 2, \dots$$

其分布函数 $F(x) = \sum_{x_k \leq x} p_k$.

(2) 连续型随机变量 X 的概率分布用概率密度 $f(x)$ 描述，其分布函数

$$F(x) = \int_{-\infty}^x f(t) dt.$$

43/126



GoBack

FullScreen

Close

Quit



定义 2.2.3 两个随机变量 X 与 Y , 如果满足 $P(\omega \in \Omega : X(\omega) \neq Y(\omega)) = 0$, 则称它们是等价的.

注: 对于两个等价的随机变量, 我们视为同一.

例 2.2.4 $(\Omega, \mathcal{F}, P)$, where

$$\mathcal{F} = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4\}, \dots\}, \quad P(\omega_4) = 0.$$
$$X = \begin{cases} 1, & \{\omega_1, \omega_2\} \\ 2, & \{\omega_3\} \\ 3, & \{\omega_4\} \end{cases} \quad Y = \begin{cases} 1, & \{\omega_1, \omega_2\} \\ 2, & \{\omega_3\} \\ -1, & \{\omega_4\} \end{cases}$$

定理 2.2.5 下列命题等价:

- (1) X 是随机变量;
- (2) $\{\omega : X(\omega) \geq a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R};$
- (3) $\{\omega : X(\omega) > a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R};$



(4) $\{\omega : X(\omega) < a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R}.$

注：习惯上将 $\{\omega : X(\omega) \geq a\}$ 记为 $\{X \geq a\}$.

$$\{X = a\} = \bigcap_{n=1}^{\infty} \{X < a + \frac{1}{n}\} - \{X < a\} \in \mathcal{F}.$$

45/126



GoBack

FullScreen

Close

Quit



定理 2.2.6 (1) 若 X, Y 是随机变量, 则 $\{X < Y\}, \{X \leq Y\}, \{X = Y\}$ 及 $\{X \neq Y\}$ 都属于 $\mathcal{F}$;

(2) 若 X, Y 是随机变量, 则 $X \pm Y$ 与 XY 亦然;

(3) 若 $\{X_n\}$ 是随机变量序列, 则 $\sup_n X_n, \inf_n X_n, \limsup_{n \rightarrow \infty} X_n$ 和 $\liminf_{n \rightarrow \infty} X_n$ 都是随机变量.

Proof

$$\begin{aligned}\{X + Y < a\} &= \{X < a - Y\} \\ &= \bigcup_{n=1}^{\infty} \{X < r_n\} \cap \{a - Y \geq r_n\} \in \mathcal{F}.\end{aligned}$$

$$\{X^2 < a\} = \{X < \sqrt{a}\} \cap \{X > -\sqrt{a}\} \in \mathcal{F}.$$

$$XY = \frac{(X + Y)^2 - (X - Y)^2}{4}.$$

$$\limsup_{n \rightarrow \infty} X_n := \lim_{n \rightarrow \infty} \left(\sup_{m \geq n} X_m \right)$$

$$\liminf_{n \rightarrow \infty} X_n := \lim_{n \rightarrow \infty} \left(\inf_{m \geq n} X_m \right)$$



二维随机变量

定义 2.2.7 如果 X 和 Y 是定义在同一概率空间 $(\Omega, \mathcal{F}, P)$ 上的两个随机变量, 称 (X, Y) 为二维随机变量(向量).

思考:

- 如何理解二维的含义
- 如何理解定义在同一概率空间

对随机变量的理解:

- 随机变量是从样本空间 Ω 到实数集合 R 的映射。
- 对于所有的 $x \in R$, 有:

$$\{\omega : X(\omega) < a\} = \{\omega : \omega \in X^{-1}(-\infty, a)\} \in \mathcal{F}.$$

- 未涉及到概率 P .

二维随机变量的联合分布:



定义 2.2.8 设 (X, Y) 是定义在 $(\Omega, \mathcal{F}, P)$ 的随机向量,对任意的 $(x, y) \in R^2$,

$$F(x, y) := P(\{\omega : X(\omega) < x, Y(\omega) < y\})$$

称为 (X, Y) 的联合分布函数.

定理 2.2.9 若 $F(x, y)$ 是联合分布函数, 则

- (1) $F(x, y)$ 对每个变量都是单调递增的;
- (2) $F(x, y)$ 对每个变量都是左连续的;
- (3) 对 $i = 1, 2, \dots, d$

$$\lim_{x \rightarrow -\infty} F(x, y) = 0, \quad \lim_{y \rightarrow -\infty} F(x, y) = 0,$$

$$\lim_{x, y \rightarrow \infty} F(x, y) = 1.$$



高维随机变量

映射 $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$, 表示为 $\mathbf{X} = (X_1, \dots, X_d)$, 若对所有的 $k, 1 \leq k \leq d$, X_k 都是随机变量, 则称 $\mathbf{X}$ 为随机向量. 对于随机向量 $\mathbf{X} = (X_1, \dots, X_d)$, 它的 (d 维) 分布函数 (或联合分布函数) 定义为

$$F(x_1, \dots, x_d) = P(X_1 < x_1, \dots, X_d < x_d)$$

这里 d 为正整数, $x_k \in \mathbb{R}, k = 1, 2, \dots, d$.

49/126



GoBack

FullScreen

Close

Quit



定理 2.2.10 若 $F(x_1, \dots, x_d)$ 是联合分布函数, 则

- (1) $F(x_1, \dots, x_d)$ 对每个变量都是单调的;
- (2) $F(x_1, \dots, x_d)$ 对每个变量都是左连续的;
- (3) 对 $i = 1, 2, \dots, d$

$$\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_i, \dots, x_d) = 0,$$

$$\lim_{x_1, x_2, \dots, x_d \rightarrow \infty} F(x_1, x_2, \dots, x_d) = 1.$$

50/126



GoBack

FullScreen

Close

Quit



如果 $f(x_1, \dots, x_d) = \frac{\partial^d F}{\partial x_1 \dots \partial x_d}$ 对所有的 $(x_1, \dots, x_d) \in R^d$ 存在, 则称函数 $f(x_1, \dots, x_d)$ 为 $F(x_1, \dots, x_d)$ 或 $X = (X_1, \dots, X_d)$ 的联合密度函数, 并且

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(t_1, \dots, t_d) dt_d \dots dt_1$$

引理 2.2.11 设 $F(x_1, \dots, x_d)$ 为 $X_1, \dots, X_d$ 的联合分布函数 $1 \leq k_1 < \dots < k_n \leq d$, 则 $X_1, \dots, X_d$ 的边际分布 $F_{k_1, \dots, k_n}(x_{k_1}, \dots, x_{k_n})$ 为

$$F_{k_1, \dots, k_n}(x_{k_1}, \dots, x_{k_n}) \\ = F(\infty, \dots, \infty, x_{k_1}, \infty, \dots, \infty, x_{k_2}, \infty, \dots, \infty, x_{k_n}, \infty, \dots, \infty)$$



一些常见分布:

1. 退化分布:若随机变量 X 只取常数 c , 即

$$P\{X = c\} = 1$$

则 X 并不随机, 但我们把它看作随机变量的退化情况更为方便, 因此称之为退化分布, 又称单点分布.

52/126



GoBack

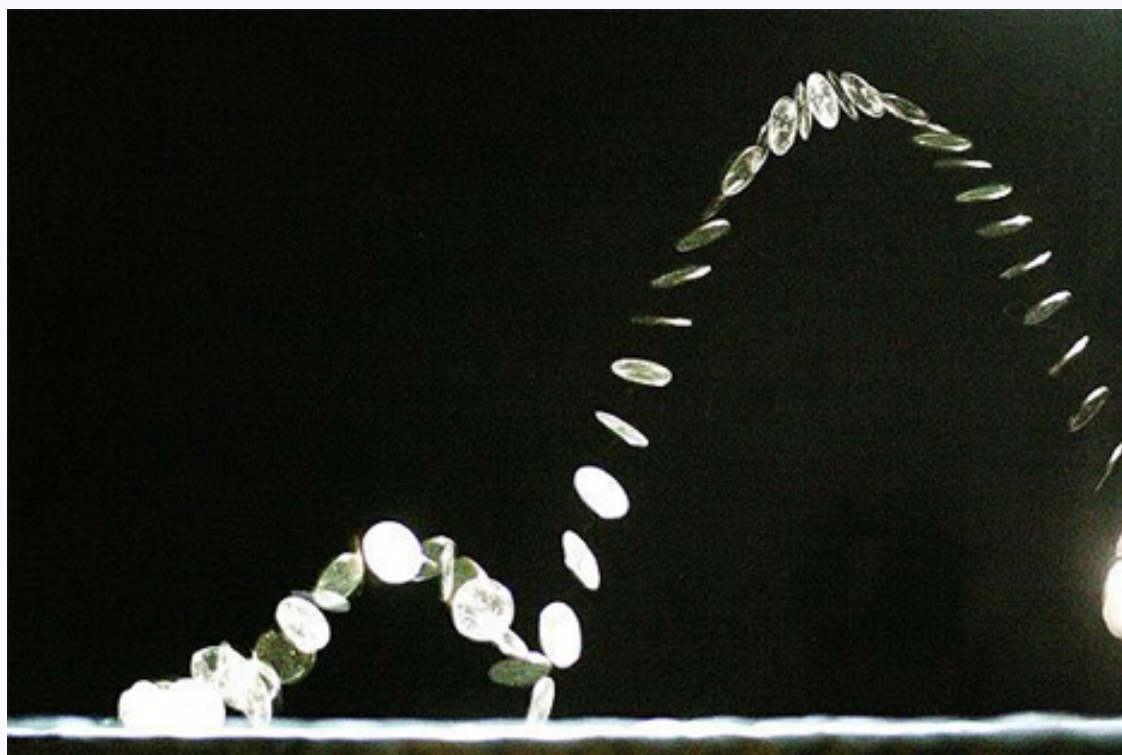
FullScreen

Close

Quit

2. Bernoulli分布:在一次试验中, 设事件 A 出现的概率为 p , $0 < p < 1$, 不出现的概率为 $1 - p$, 若以 X 记事件 A 出现的次数, 则 X 的可能取值仅为 $0, 1$, 其对应的概率为

$$P\{X = k\} = p^k(1 - p)^{1-k}, \quad k = 0, 1$$





3. 二项分布:在 n 重Bernoulli试验中, 设事件 A 在每次试验中出现的概率均为 $p, 0 < p < 1$, 以 X 记事件 A 出现的次数, X 的可能取值为 $0, 1, 2, \dots, n$, 其对应的概率为

$$P\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

则称之为以 n 和 p 为参数的二项分布, 简记为 $X \sim B(n, p)$.

54/126



GoBack

FullScreen

Close

Quit



55/126



GoBack

FullScreen

Close

Quit



4. **Poisson 分布**: 若随机变量 X 可取一切非负整数值, 且

$$P\{X = k\} = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots$$

其中 $\lambda > 0$, 则称 X 服从Poisson分布, 记为 $X \sim P(\lambda)$.

56/126



GoBack

FullScreen

Close

Quit



57/126

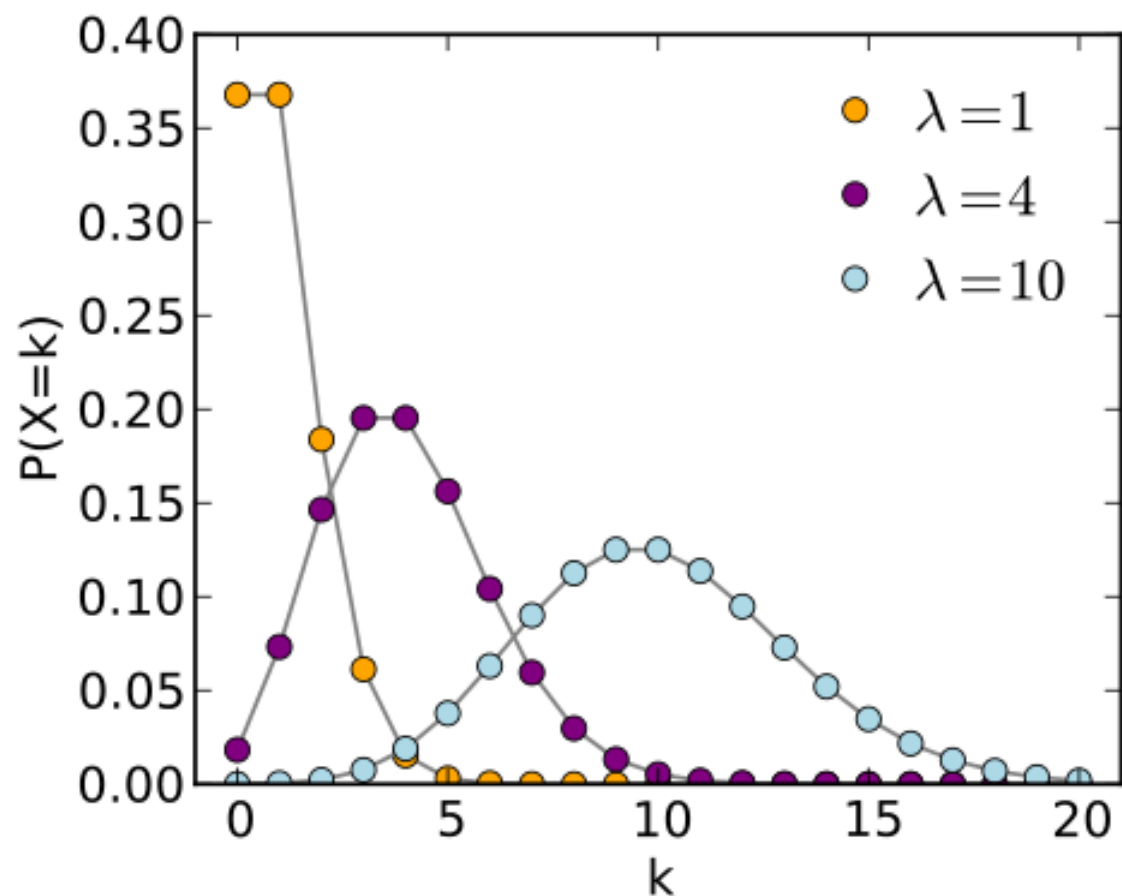


GoBack

FullScreen

Close

Quit



5. 几何分布 (Geometric distribution):在Bernoulli试验中, 设事件 A 在每次试验中出现的概率均为 $p, 0 < p < 1$, 以 X 记事件 A 首次出现的试验次数, X 的可能取值为 $0, 1, 2, \dots$, 其对应的概率为

$$P\{X = k\} = p(1 - p)^{k-1}, \quad k \geq 1$$

则称之为几何分布.





6. Pascal分布:在Bernoulli试验中, 设事件 A 在每次试验中出现的概率均为 $p, 0 < p < 1$, 以 X 记事件 A 第 r 次出现的试验次数, X 的可能取值为 $r, r + 1, \dots$, 其对应的概率为

$$P\{X = k\} = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r + 1, \dots$$

则称之为Pascal分布.

60/126



GoBack

FullScreen

Close

Quit



7. 负二项分布 (negative binomial distribution): 对于任意实数 $r > 0$, 称

$$P\{X = k\} = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \dots, \quad r > 0$$

为负二项分布.

负二项分布通常用于替换Poisson分布. 同Poisson分布一样, 它也在非负整数上取值, 但因为它包含两个参数, 相比Poisson分布其变化更灵活. Poisson分布的方差和均值相等, 但负二项分布的方差大于均值, 这说明当某类数据集观测到的方差大于均值时, 负二项分布要比Poisson分布更合适.

8. 离散均匀分布: 如果分布列为

$$p_k = \frac{1}{n}, \quad k = 1, 2, \dots, n$$

61/126



GoBack

FullScreen

Close

Quit



9. 连续均匀分布(简称均匀分布):如果密度函数为

$$f(x) = \begin{cases} [b - a]^{-1}, & \text{若 } a \leq x \leq b \\ 0, & \text{其他} \end{cases}$$

其中 $a < b$, 则称之为区间 $[a, b]$ 上的均匀分布.

10. 正态分布: 如果密度函数为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\{-(x - \mu)^2/2\sigma^2\}, x \in \mathbb{R}$$

则称之为参数为 μ 和 σ^2 的正态分布, 也称为 Gauss 分布, 记为 $X \sim N(\mu, \sigma^2)$.

Simulation:

<http://www.mathsisfun.com/data/quincunx.html>

62/126



GoBack

FullScreen

Close

Quit



11. d 维正态分布: 设 $\mu = (\mu_1, \dots, \mu_d)$, Σ 是 d 阶正定对称矩阵, 并且其行列式为 $|\Sigma|$. 如果其联合密度函数为

$$f(x_1, \dots, x_d) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

则称之为 d 维正态分布, 记为 $X \sim N(\mu, \Sigma)$.

12. Γ 分布: 如果密度函数为

$$f(x) = \begin{cases} \frac{\lambda^s}{\Gamma(s)} x^{s-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称之为以 $s > 0$, $\lambda > 0$ 为参数的 Γ 分布, 其中 Γ 函数定义为

$$\Gamma(s) = \int_0^{\infty} e^{-x} x^{s-1} dx, \quad s > 0$$



13. 指数分布: 如果在 Γ 分布中令 $s = 1$, 即密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称之为指数分布.

14. χ^2 分布: 如果在 Γ 分布中取 $s = \frac{n}{2}$, n 是正整数, $\lambda = \frac{1}{2}$, 即

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x > 0$$

则称之为自由度是 n 的 χ^2 分布.

64/126



GoBack

FullScreen

Close

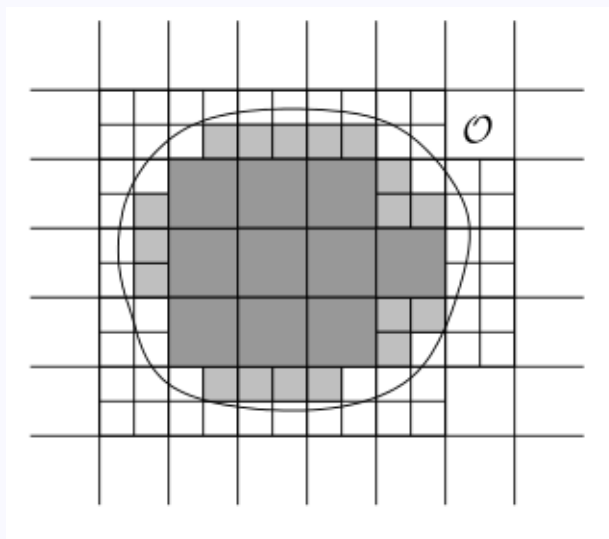
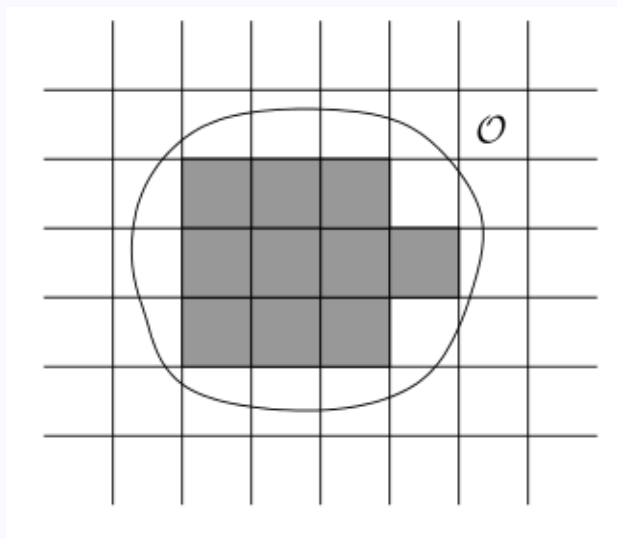
Quit

§2.3 数字特征、矩母函数与特征函数

Lebesgue积分

一些测度的思想：

计算二维平面的任意图形的面积：



65/126



GoBack

FullScreen

Close

Quit

外测度 (Exterior Measure)

$$m^*(E) = \inf \sum_{j=1}^{\infty} |Q_j|$$

其中, E 是 R^d 中的子集, (Q_j) 是 E 的一个覆盖序列。

Lebesgue测度的思想

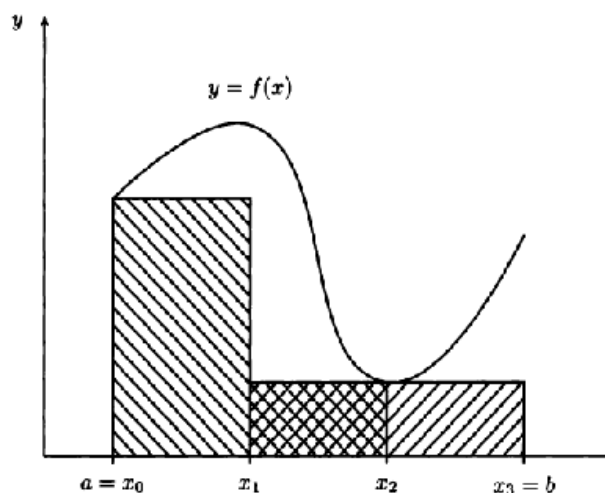


Fig. 1.3.1. Lower Riemann sum.

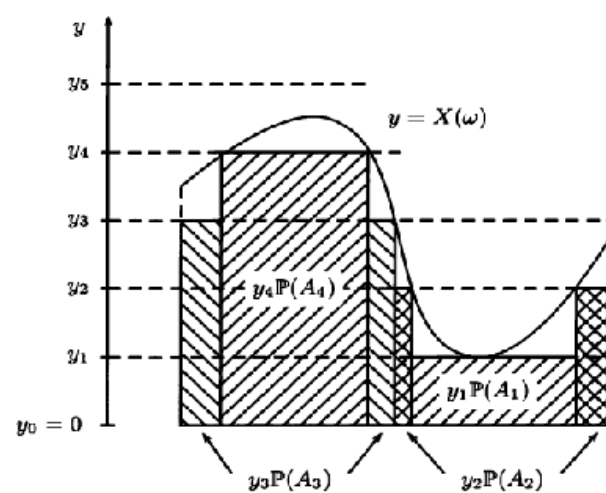


Fig. 1.3.2. Lower Lebesgue sum.



Lebesgue测度

- $\mathcal{B}$ 是 $\mathbb{R}$ 上的 σ -代数。但在理论上 $\mathcal{B}$ 这个集合系还不够大，有一些应该是零测集的集合并不属于 $\mathcal{B}$ 。换句话说， $\mathcal{B}$ 并不是 $\mathbb{R}$ 的一个“完备的”集合系。实分析理论希望考查一个 $\mathbb{R}$ 上的完备 σ -代数。
- Caratheodory扩张定理指出，可以把 $\mathcal{B}$ 进行完备化扩张，得到的更大的 σ -代数称为Lebesgue可测集合系。
- Lebesgue测度是Borel测度的一种推广，所有Borel集的Lebesgue测度都等于先前自然定义的测度。此外还有许多 $\mathcal{B}$ 以外的集合也有Lebesgue测度。

67/126



GoBack

FullScreen

Close

Quit



Lebesgue下和

- 对非负随机变量 X 的取值进行划分 $\Pi := \{y_0, y_1, \dots\}$, 其中 $0 = y_0 < y_1 < \dots$, 并令

$$A_k := \{\omega \in \Omega : y_k \leq X(\omega) < y_{k+1}\} \in \mathcal{F},$$

并定义如下的Lebesgue下和:

$$LS_{\Pi}^{-}(X) := \sum_{k=0}^{+\infty} y_k P(A_k).$$

- 能这样定义是基于随机变量可测的性质。这里, P 可以指定义域 Ω 上的概率测度。

68/126



GoBack

FullScreen

Close

Quit

Lebesgue积分

- 令分割 Π 变得无限密集:

$$||\Pi|| := \sup_{k \geq 0} (y_{k+1} - y_k) \rightarrow 0,$$

如果Lebesgue下和收敛, 则定义其极限为 X 在 Ω 上关于概率 P 的 Lebesgue积分, 记为

$$\begin{aligned} \lim_{||\Pi|| \rightarrow 0} LS_{\Pi}^{-}(X) &:= \sum_{k=0}^{+\infty} y_k P(A_k) \\ &=: \int_{\Omega} X(\omega) dP = E[X]. \end{aligned}$$



69/126



GoBack

FullScreen

Close

Quit

分割的是值域 与黎曼积分分割定义域不同, Lebesgue积分分割的是被积函数的值域。

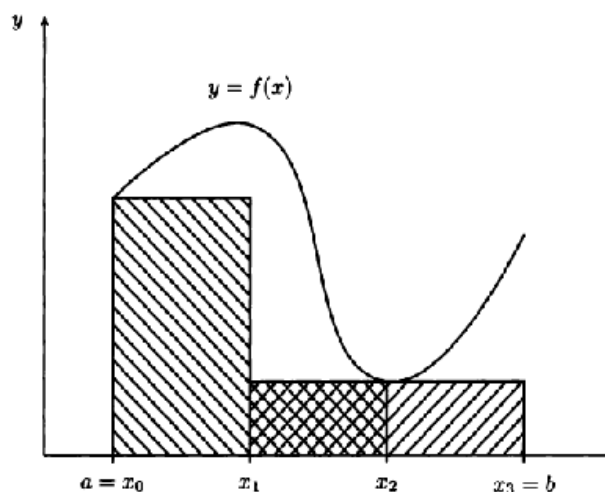


Fig. 1.3.1. Lower Riemann sum.

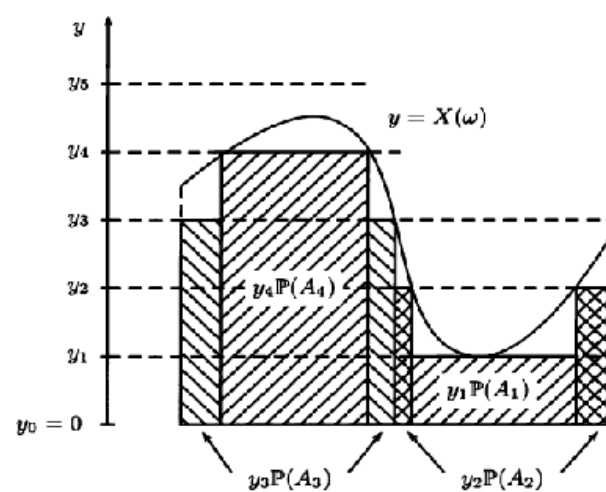


Fig. 1.3.2. Lower Lebesgue sum.



§2.3.1 Riemann-Stieltjes积分

设 $g(x), F(x)$ 为有限区间 $(a, b]$ 上的实值函数, $a = x_0 < x_1 < \cdots < x_n = b$ 为 $(a, b]$ 的一个分割, 令

$$\Delta F(x_i) = F(x_i) - F(x_{i-1}), \xi_i \in [x_{i-1}, x_i], 1 \leq i \leq n,$$

$\lambda = \max_{1 \leq i \leq n} (x_i - x_{i-1})$, 如果当 $\lambda \rightarrow 0$ 时, 极限

$$\lim_{\lambda \rightarrow 0} \sum_{i=1}^n g(\xi_i) \Delta F(x_i)$$

存在, 且与分割的选择以及 $\xi_i \in [x_{i-1}, x_i]$ 的取法无关, 则称该极限值为函数 $g(x)$ 关于 $F(x)$ 在 $(a, b]$ 上的Riemann-Stieltjes积分, 记为

$$\int_a^b g(x) dF(x) = \lim_{\lambda \rightarrow 0} \sum_{i=1}^n g(\xi_i) \Delta F(x_i) \quad (2.3.1)$$

71/126



GoBack

FullScreen

Close

Quit



当 $F(x) = x$ 时, Riemann-Stieltjes积分即为Riemann积分. 关于Riemann-Stieltjes积分存在的条件, 这里不做更进一步的讨论, 只给出一个简单的充分条件: 若函数 $g(x)$ 连续, $F(x)$ 单调, 则Riemann-Stieltjes积分存在. 本书中用到的 $g(x)$ 为连续函数, $F(x)$ 为分布函数, 因此积分的存在性不成问题. 为了后面的需要, 将积分推广到无限区间上:

$$\begin{aligned}\int_a^{+\infty} g(x) dF(x) &\doteq \lim_{b \rightarrow +\infty} \int_a^b g(x) dF(x) \\ \int_{-\infty}^b g(x) dF(x) &\doteq \lim_{a \rightarrow -\infty} \int_a^b g(x) dF(x) \\ \int_{-\infty}^{+\infty} g(x) dF(x) &\doteq \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b g(x) dF(x)\end{aligned}$$

72/126



GoBack

FullScreen

Close

Quit



与Riemann积分不同的是

$$\int_a^{a+} dF(x) = \lim_{\delta \rightarrow 0+} \int_a^{a+\delta} dF(x) = F(a+) - F(a)$$

当 $F(x)$ 在 $x = a$ 处有跳跃时, 上式的值等于 $F(x)$ 在 a 点的跃度. 当 $F(x)$ 是一个阶梯函数时, Riemann-Stieltjes积分成为一个级数, 即设 $F(x)$ 在 $x = x_i$ 处有跃度 $p_i, i = 1, 2, \dots$, 则

$$\int_{-\infty}^{+\infty} g(x) dF(x) = \sum_{i=1}^{+\infty} g(x_i) p_i$$

Riemann-Stieltjes积分的一些基本性质:

(1) 线性性质:

$$\int_a^b [\alpha g_1(x) \pm \beta g_2(x)] dF(x) = \alpha \int_a^b g_1(x) dF(x) \pm \beta \int_a^b g_2(x) dF(x)$$

73/126



GoBack

FullScreen

Close

Quit



(2) 区间可加性:

$$\int_a^b g(x) dF(x) = \int_a^c g(x) dF(x) + \int_c^b g(x) dF(x)$$

(3) $\int_a^b dF(x) = F(b) - F(a)$, 其中 a, b 均可有限或无穷大.

$$(4) \int_a^b g(x) d[\alpha F_1(x) + \beta F_2(x)] = \alpha \int_a^b g(x) dF_1(x) + \beta \int_a^b g(x) dF_2(x)$$

(5) 若 $g(x) \geq 0$, $F(x)$ 单调不减, $b > a$, 则 $\int_a^b g(x) dF(x) \geq 0$.

74/126



GoBack

FullScreen

Close

Quit



§2.3.2 数字特征

定义 2.3.1 (1) 取值为 $\{s_k\}$ 的离散型随机变量 X 的数学期望 (简称为期望) $E[X]$ 定义为

$$E[X] = \sum_k s_k p_k = \sum_k s_k P(X = s_k)$$

如果 $\sum |s_k| p_k < \infty$.

(2) 连续型随机变量 X 的数学期望 (**Expectation**) $E[X]$ 定义为

$$E[X] = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx$$

如果 $\int_{-\infty}^{\infty} |x| dF(x) < \infty$, 这里 $F(x)$ 是 X 的分布函数, $f(x)$ 是其密度函数.

利用 Riemann-Stieltjes 积分, 我们可以对离散型随机



变量和连续型随机变量的期望给出一个统一的表达式:

$$\int_{\Omega} X(\omega) dP(\omega) := E[X] = \int_{-\infty}^{+\infty} x dF(x)$$

(3) 设 X 为任一随机变量, 对正整数 k , 称 $m_k = E[X^k]$ 为 X 的 k 阶原点矩. 数学期望是一阶原点矩.

(4) 设 X 为任一随机变量, 对正整数 k , 称 $c_k = E[X - E[X]]^k$ 为 X 的 k 阶中心矩 (**k -th moment function**). 方差是二阶原点矩.

(5) 设 X, Y 为两个随机变量, 对正整数 k, l , 称 $E[X - E[X]]^k [Y - E[Y]]^l$ 为 X 的 $k + l$ 阶混合中心矩. 协方差是二阶混合中心矩.

76/126



GoBack

FullScreen

Close

Quit



§2.3.3 选读 关于概率测度的积分 选读

定义 2.3.2 设 $(\Omega, \mathcal{F})$ 为一可测空间, $\mathbb{R}$ 为实数域, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, 分别用 $\mathcal{B}(\mathbb{R})$ 及 $\mathcal{B}(\bar{\mathbb{R}})$ 表示 $\mathbb{R}$ 及 $\bar{\mathbb{R}}$ 上的 Borel σ 代数. 令 f 为 Ω 到 $\mathbb{R}$ 中的一个映射, 如果 $f^{-1}(\mathcal{B}(\bar{\mathbb{R}})) \subset \mathcal{F}$, 则称 f 为 **Borel 可测函数**, 简称可测函数. 若进一步 f 只取实值, 则称 f 为实值可测函数.

定义 2.3.3 如果存在实数 $a_k, 1 \leq k \leq n$ 和 Ω 的分割 $A_k \in \mathcal{F}, 1 \leq k \leq n$ (即 $\bigcup_{k=1}^n A_k = \Omega$, 且 $A_i \cap A_j = \emptyset, i \neq j, 1 \leq i, j \leq n$), 使得 $f(\omega) = \sum_{k=1}^n a_k I_{A_k}(\omega)$, 则称 f 为简单可测函数. 这里 $I_A(\cdot)$ 表示集合 A 的示性函数.

若 f 还可以表示为 $f = \sum_{j=1}^m b_j I_{B_j}$, 则当 $B_j \cap A_k \neq \emptyset$ 时, $b_j = a_k$. 于是, 通过将分割中的集合合并可以得到 f 的最简单

77/126



GoBack

FullScreen

Close

Quit



表达式, 即表达式中的系数 a_k 互不相同.

以下给定一个完备的概率空间 $(\Omega, \mathcal{F}, P)$. 用 $\mathcal{S}^+$ 表示 Ω 上非负简单可测函数全体, $\mathcal{L}^+$ 表示 Ω 上非负可测函数全体, $\mathcal{L}$ 表示 Ω 上可测函数全体.

首先我们定义非负简单可测函数关于概率测度的积分.

定义 2.3.4 令 $\mathbf{S}$ 表示所有非负简单可测函数 $h : \Omega \rightarrow R_+$. 如果 $h \in \mathbf{S}$, 则定义其关于 P 的积分为

$$\int h dP = \sum_{k=1}^n a_k P(A_k) = E[h]$$

易知 $\int_{\Omega} f dP$ 不依赖于 f 的具体表达. 称 $\int_{\Omega} f dP$ 为 f 关于概率测度 P 的积分.

定理 2.3.5 (简单函数逼近定理) 设 $(\Omega, \mathcal{F})$ 为一可测空间, $f \in \mathcal{L}$.

(1) 存在一简单可测函数序列 $(f_n, n \geq 1)$, 使得对一切 $n \geq 1$, 有 $|f_n| \leq |f|$, 且 $\lim_{n \rightarrow \infty} f_n = f$.

(2) 若 f 非负, 则存在非负简单可测函数的增序列 (f_n) , 使得 $\lim_{n \rightarrow \infty} f_n = f$.

接下来定义非负可测函数的积分.

定义 2.3.6 设 $f \in \mathcal{L}^+$. 任取 $f_n \in \mathcal{S}^+$ 使得 $f_n \uparrow f$, 令

$$\int_{\Omega} f dP = \lim_{n \rightarrow \infty} \int_{\Omega} f_n dP$$

上述右端极限存在, 且不依赖于序列 (f_n) 的选取, 称 $\int_{\Omega} f dP$ 为 f 关于概率测度 P 的积分.

最后定义可测函数的积分.

定义 2.3.7 设 $f \in \mathcal{L}$. 令

$$f^+ = \max\{f, 0\} \quad \text{及} \quad f^- = -\min\{f, 0\}$$



注意到 $f^+ \geq 0, f^- \geq 0, f = f^+ - f^-, |f| = f^+ + f^-$. 如果 $\int_{\Omega} f^+ dP < \infty$ 或 $\int_{\Omega} f^- dP < \infty$, 则称 f 关于 P 的积分存在. 令

$$\int_{\Omega} f dP = \int_{\Omega} f^+ dP - \int_{\Omega} f^- dP$$

如果 $\int_{\Omega} f^+ dP < \infty$ 且 $\int_{\Omega} f^- dP < \infty$, 则称 f 关于概率测度 P 可积.

于是, 当随机变量 X 可积时, 它的期望就可以定义为

$$E[X] = \int_{\Omega} X dP$$

令 $\mathcal{L}^p(\Omega), p \geq 1$ 表示所有使得 $E[|X|^p] < \infty$ 的随机变量(等价类)全体, 简记为 $\mathcal{L}^p$.

关于 Ω 中样本点的某种性质 Π , 如果使得 Π 不成立的点的集合的概率是零, 则称 Π 几乎必然(almost surely) 或



以概率1 (with probability one) 成立, 记为 $a.s.$ 或 $w.p.1.$

定理 2.3.8 设 f, g 积分存在.

(1) $\forall a \in \mathbb{R}$, af 的积分存在, 且 $\int_{\Omega} af dP = a \int_{\Omega} f dP$.

(2) 若 $f + g$ 处处有定义, 且 $\int_{\Omega} f dP + \int_{\Omega} g dP$ 处处有意义 (即不出现 $\infty - \infty$), 则 $f + g$ 的积分存在, 且有 $\int_{\Omega} (f + g) dP = \int_{\Omega} f dP + \int_{\Omega} g dP$.

(3) 若取非负整数值的随机变量 X 的期望存在, 则 $E[X] = \sum_{k=1}^{\infty} P\{X \geq k\}$.

(4) 若 N 为一零测集, 则 $\int_{\Omega} f I_N dP = 0$.

(5) 若 $f \leq g, a.s.$, 则 $\int_{\Omega} f dP \leq \int_{\Omega} g dP$.

(6) $|\int_{\Omega} f dP| \leq \int_{\Omega} |f| dP$.

(7) 设 $X \in \mathcal{L}^+$, 则 $E[X] = 0$ 当且仅当 $X = 0$ a.s.

81/126



GoBack

FullScreen

Close

Quit



§2.3.4 矩母函数(Moment Generating Function)

定义 2.3.9 若随机变量 X 的分布函数为 $F_X(x)$, 则称

$$\phi_X(t) = E[e^{tX}] = \int_{\Omega} e^{tX(\omega)} P(d\omega) = \int_{-\infty}^{\infty} e^{tx} dF_X(x) \quad (2.3.2)$$

为 X 的矩母函数.

注: X 的各阶矩与矩母函数的关系 (假设对 $\phi(t)$ 求导时, 求导运算与求期望运算可以交换次序)

$$\phi'(t) = E(Xe^{tX})$$

$$\phi''(t) = E(X^2e^{tX})$$

$$\vdots$$

$$\phi^{(n)}(t) = E(X^n e^{tX})$$

82/126



GoBack

FullScreen

Close

Quit



令 $t = 0$, 得到 $\phi^{(n)}(0) = E[X^n]$, $n \geq 1$. 当矩母函数存在时, 它唯一地决定分布, 因此我们能够用矩母函数刻画随机变量的概率分布. 但有时随机变量的矩母函数不一定存在, 在这种情况下, 更方便的是特征函数.

§2.3.5 特征函数(Characteristic function)

定义 2.3.10 若随机变量 X 的分布函数为 $F_X(x)$, 则称

$$\psi_X(t) = E[e^{itX}] = \int_{\Omega} e^{itX(\omega)} P(d\omega) = \int_{-\infty}^{\infty} e^{itx} dF_X(x)$$

为 X 的特征函数. 如果 F_X 有密度 $f(x)$, 则 $\psi_X(t)$ 就是 $f(x)$ 的 Fourier 变换

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$



特征函数是一个实变量的复值函数, 因为 $|e^{itx}| = 1$, 所以它对一切实数 t 都有定义. 特征函数有如下常用性质:

(1) 有界性: $|\psi(t)| \leq 1 = \psi(0)$;

(2) 共轭对称性: $\psi(-t) = \overline{\psi(t)}$;

(3) 一致连续性:

$$|\psi(t+h) - \psi(t)| \leq \int_{-\infty}^{\infty} |e^{ihx} - 1| dF(x);$$

(4) 线性变换的作用: 设 $Y = aX + b$, 则 Y 的特征函数是 $\psi_Y(t) = e^{ibt}\psi_X(at)$;

(5) 两个相互独立的随机变量之和的特征函数等于它们的特征函数之积.

(6) 非负定性: 对于任意的正整数 n , 任意实数 $t_1, \dots, t_n$ 及复数 $\lambda_1, \dots, \lambda_n$, 有 $\sum_{k=1}^n \sum_{j=1}^n \psi(t_k - t_j) \lambda_k \overline{\lambda_j} \geq 0$.

(7) 设随机变量 X 有 n 阶矩存在, 则它的特征函数可微



分 n 次, 且当 $k \leq n$ 时, 有

$$\psi^{(k)}(0) = i^k E[X^k]$$

特别地, 特征函数可作如下带皮阿诺型余项的Taylor展开:

$$\psi(t) = 1 + itE[X] + \frac{(it)^2}{2!}E[X^2] + \cdots + \frac{(it)^n}{n!}E[X^n] + o(t)$$

例 2.3.11 求单点分布 $P(X = c) = 1$ 的特征函数。

$$\psi(t) = E(e^{itc}) = e^{itc}, \quad t \in R.$$

例 2.3.12 求两点分布 $P(X = 0) = p, \quad P(X = 1) = 1 - p$ 的特征函数。

$$\psi(t) = e^{it*0} * p + e^{it*1} * (1 - p) = p + e^{it} * (1 - p), \quad t \in R.$$

例 2.3.13 求指数分布的特征函数。

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$



$$\begin{aligned}\psi(t) &= \int_0^{\infty} \lambda e^{itx} e^{-\lambda x} dx \\ &= \int_0^{\infty} \lambda e^{-\lambda x} \cos(tx) dx + i\lambda \int_0^{\infty} e^{-\lambda x} \sin(tx) dx \\ &= \lambda \frac{\lambda}{\lambda^2 + t^2} + i\lambda \frac{t}{\lambda^2 + t^2} = \left(1 - \frac{it}{\lambda}\right)^{-1}.\end{aligned}$$

例 2.3.14 求标准正态分布 $N(0, 1)$ 的特征函数.

解: 由定义 $\psi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx$

86/126



GoBack

FullScreen

Close

Quit



从而

$$\begin{aligned}\psi'(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ix e^{itx} e^{-x^2/2} dx \\ &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} d(-e^{-x^2/2}) dx \\ &= \frac{i}{\sqrt{2\pi}} \left(-e^{itx-x^2/2} \Big|_{-\infty}^{\infty} + it \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx \right) \\ &= -t\psi(t)\end{aligned}$$

87/126

解初值问题 $\psi'(t) + t\psi(t) = 0$, $\psi(0) = 1$ 得 $\psi(t) = e^{-\frac{1}{2}t^2}$.

例 2.3.15 求正态分布 $N(\mu, \sigma^2)$ 的特征函数.

解: 设 $Y \sim N(\mu, \sigma^2)$, $X \sim N(0, 1)$, 则有 $Y = \sigma X + \mu$.

利用性质(4)有 $\psi_Y(t) = e^{it\mu} \psi_X(\sigma t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$.

由于特征函数只与分布函数有关, 所以称为分布的特征



GoBack

FullScreen

Close

Quit



函数. 另一方面, 有下述定理.

定理 2.3.16 (唯一性定理) 分布函数由其特征函数唯一决定.

从而说明特征函数与分布函数是相互唯一确定的.

定理 2.3.17 特征函数与矩的关系, 若随机变量 X 的 n 阶矩存在, 则 X 的特征函数 $\psi(t)$ 的 k 阶导数 $\psi^{(k)}(t)$ 存在, 且

$$E(X^k) = i^{(-k)} \psi^{(k)}(0).$$

例 2.3.18 随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} \frac{1}{2} \cos(x), & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0, & \text{others.} \end{cases}$$

求 $E(X)$, $D(X)$.



Proof Recall

$$\cos(x) + \cos(y) = 2 \cos\left(\frac{x+y}{2}\right) \cos\left(\frac{x-y}{2}\right).$$

省略了虚部

$$\begin{aligned}\psi(t) &= 2 \int_0^{\frac{\pi}{2}} \frac{1}{2} \cos(x) \cos(tx) dx \\ &= \frac{1}{2} \int_0^{\frac{\pi}{2}} (\cos(t+1)x + \cos(t-1)x) dx \\ &= \frac{1}{2} \left(\frac{1}{t+1} \sin(t+1) \frac{\pi}{2} + \frac{1}{t-1} \sin(t-1) \frac{\pi}{2} \right).\end{aligned}$$

Since $\psi^{(1)}(0) = 0$ and $\psi^{(2)}(0) = 2 - \pi^2/4$, we get

$$E(X) = i^{-1} \psi^{(1)}(0) = 0, \quad D(X) = i^{-2} \psi^{(2)}(0) = -2 + \pi^2/4.$$

89/126



GoBack

FullScreen

Close

Quit

反演公式及唯一性定理

由随机变量 X 的分布函数可惟一确定其特征函数, 能否由 X 的特征函数惟一确定其分布函数??

定理 2.3.19 (反演公式) 设随机变量 X 的分布函数和特征函数分别为 $F(x)$ 和 $\psi(t)$. 则对 $F(x)$ 的任意连续点 x_1, x_2 , 有:

$$F(x_2) - F(x_1) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \psi(t) dt.$$

Recall:

$$\int_0^\infty \frac{\sin(t)}{t} dt = \frac{\pi}{2}, \text{ by } k(\lambda) := \int_0^\infty e^{-\lambda t} \frac{\sin(t)}{t} dt$$

推论 2.3.20 (唯一性定理) 分布函数由其特征函数惟一确定.



Proof:

$$\begin{aligned}
& \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \psi(t) dt \\
&= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \int_{-\infty}^{\infty} f(y) e^{ity} dy dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-it(x_1-y)} - e^{-it(x_2-y)}}{it} dy dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) \lim_{T \rightarrow \infty} \int_{-T}^T \frac{\cos(t(x_1-y)) - i \sin(t(x_1-y)) - \cos(t(x_2-y)) + i \sin(t(x_2-y))}{it} dy dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) 2 \lim_{T \rightarrow \infty} \int_0^T \frac{i \sin(t(x_2-y)) - i \sin(t(x_1-y))}{it} dy dt \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} f(y) \int_0^{\infty} \frac{\sin(t(x_2-y)) - \sin(t(x_1-y))}{t} dy dt \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} f(y) \frac{\pi}{2} (\operatorname{sgn}(x_2-y) - \operatorname{sgn}(x_1-y)) dy \\
&= F(x_2) - F(x_1).
\end{aligned}$$

91/126



GoBack

FullScreen

Close

Quit



推论 2.3.21 若随机变量 X 的特征函数 $\psi(t)$ 在 R 上绝对可积, 则 X 为连续型随机变量, 其概率密度为

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi(t) dt.$$

例 2.3.22 若随机变量 X 是离散型的, 其分布律为

$$p_k = P(X = k), \quad k \in \mathbb{Z}.$$

其特征函数为

$$\psi(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikt}, \quad t \in R.$$

且:

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \psi(t) dt.$$

Proof

$$\begin{aligned}
 \int_{-\pi}^{\pi} e^{-imt} \psi(t) dt &= \int_{-\pi}^{\pi} e^{-imt} \sum_{k=-\infty}^{\infty} p_k e^{ikt} dt \\
 &= \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} p_k e^{-imt} e^{ikt} dt \\
 &= p_m \int_{-\pi}^{\pi} dt + \sum_{k=-\infty, k \neq m}^{\infty} \int_{-\pi}^{\pi} p_k e^{-imt} e^{ikt} dt \\
 &= 2\pi p_m.
 \end{aligned}$$

93/126

例 2.3.23 随机变量 X 在 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 上服从均匀分布, $Y = \cos(X)$, 利用特征函数求 Y 的概率密度.

Proof X 的密度函数为:

$$f(x) = \begin{cases} \frac{1}{\pi}, & x \in [-\pi/2, \pi/2] \\ 0, & \text{others.} \end{cases}$$



GoBack

FullScreen

Close

Quit



由此的的 Y 的特征函数为：

$$\begin{aligned}\psi_Y(t) &= E(e^{itY}) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{it \cos(x)} \frac{1}{\pi} dx \\ &= \int_0^1 e^{itu} \frac{2}{\pi} \frac{1}{\sqrt{1-u^2}} du.\end{aligned}$$

利用特征函数与分布函数一一对应的惟一性定理，知随机变量 Y 的概率密度函数为：

$$f_Y(u) = \frac{2}{\pi} \frac{1}{\sqrt{1-u^2}}, \quad 0 < u < 1.$$

94/126



GoBack

FullScreen

Close

Quit



独立随机变量和的特征函数

定理 2.3.24 若随机变量 $X_1, X_2, \dots, X_n$ 相互独立, 令 $Y = \sum_{k=1}^n X_k$, 则

$$\psi_Y(t) = \prod_{k=1}^n \psi_{X_k}(t).$$

例 2.3.25 随机变量 $Y \sim B(n, p)$, 写出其特征函数.

Proof 由于

$$Y = \sum_{k=1}^n X_k,$$

where, $X_k \sim B(1, p)$ 为二项分布. 由此可得:

$$\psi_Y(t) = \prod_{k=1}^n \psi_{X_k}(t) = (q + pe^{it})^n.$$



例 2.3.26 若 $X_1, X_2, \dots, X_n$ 相互独立, 且 $X_k \sim N(0, 1)$.

证明

$$Y = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k.$$

也服从 $N(0, 1)$ 分布.

Proof 标准正态分布 X_k 的特征函数为 $\psi_{X_k}(t) = e^{-\frac{t^2}{2}}$.
则,

$$\psi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \psi_{X_k}(t) = e^{-n\frac{t^2}{2}}.$$

由此可以得到:

$$\psi_Y(t) = \psi_{\sum_{k=1}^n X_k}(t/\sqrt{n}) = e^{-\frac{t^2}{2}}.$$

由唯一性定理知, $Y \sim N(0, 1)$.

多维随机变量的特征函数

定义 2.3.27 二维随机变量 (X, Y) 的特征函数定义为

$$\psi(t_1, t_2) = E \left(e^{i(t_1 X + t_2 Y)} \right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(t_1 x + t_2 y)} dF(x, y).$$

若 (X, Y) 为连续型的随机变量, 则:

$$\psi(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(t_1 x + t_2 y)} f(x, y) dx dy.$$

若 (X, Y) 为离散型的随机变量, 则:

$$\psi(t_1, t_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} e^{i(t_1 m + t_2 n)} P(X = m, Y = n).$$

定义 2.3.28 n 维随机向量 $(X_1, X_2, \dots, X_n)$ 的分布函数为 $F(X_1, X_2, \dots, X_n)$, 则它的特征函数为

$$\psi(t_1, t_2, \dots, t_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i(t_1 x_1 + t_2 x_2 + \cdots + t_n x_n)} dF(x_1, x_2, \dots, x_n).$$



引理 2.3.29 随机变量 $X_1, X_2, \dots, X_n$ 相互独立的充要条件是

$$\psi(t_1, t_2, \dots, t_n) = \prod_{k=1}^n \psi_{X_k}(t_k).$$

引理 2.3.30 二维随机变量 (X, Y) 的特征函数为 $\psi(t_1, t_2)$. 则, $Z = aX + bY + c$ 的特征函数为

$$\psi_Z(t) = e^{itc} \psi(at, bt).$$

特别的有:

$$\psi_{X+Y}(t) = \psi(t, t).$$

98/126



GoBack

FullScreen

Close

Quit



若随机向量 $(X_1, \cdots, X_n)$ 的分布函数为 $F(x_1, \cdots, x_n)$, 与随机变量相仿, 类似地定义它的特征函数

$$\psi(t_1, \cdots, t_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i(t_1 x_1 + \cdots + t_n x_n)} dF(x_1, \cdots, x_n)$$

可以类似于一元的场合, 建立起 n 元特征函数的理论.

例如, 我们考虑在上1.2节中定义的 d 维正态分布, 简称为多元正态分布, 这个定义在推导很多性质时并不方便, 同时还不能考虑 Σ 不是正定的情形, 为此, 我们采用下面的定义. 如果随机变量 X 的特征函数为

$$\psi_X(t) = \exp\{it'\mu - \frac{1}{2}t'\Sigma t\}$$

则称 $\mathbf{X}$ 服从多元正态分布 $N(\mu, \Sigma)$. 其中 $\mathbf{X} = (X_1, \cdots, X_d)'$, $(\cdot)'$ 表示 $(\cdot)$ 的转置, $\mathbf{t} = (t_1, \cdots, t_d)'$, $\mu = (\mu_1, \cdots, \mu_d)'$, Σ 是非负定 $d \times d$ 矩阵. 当 $\mu = \mathbf{0}$, $\Sigma = \mathbf{I}_d$ 时, 称为标准多



元正态分布, 记为 $X \sim N(\mathbf{0}, \mathbf{I}_d)$.

利用特征函数方法不难证明下面几个命题:

命题 2.3.31 若 $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_d)$, 则 $\mathbf{X}$ 的任一线性函数 $\mathbf{Y} = \mathbf{A}_{n \times d} \mathbf{X} + \mu$ 服从 n 维正态分布 $N(\mu, \mathbf{A} \mathbf{A}')$.

命题 2.3.32 若 $\mathbf{Y} \sim N(\mu, \Sigma)$, 则 $\mathbf{A} \mathbf{Y} + \mathbf{b} \sim N(\mathbf{A} \mu + \mathbf{b}, \mathbf{A} \Sigma \mathbf{A}')$.

§2.4 收敛性 (该内容为选读)

定义 2.4.1 (1) 设 $\{X_n, n \geq 1\}$ 是随机变量序列, 若存在随机变量 X 使得

$$P\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\} = 1$$

100/126



GoBack

FullScreen

Close

Quit



则称随机变量序列 $\{X_n, n \geq 1\}$ 几乎必然收敛(或以概率1收敛)于 X , 记为 $X_n \rightarrow X, a.s.$ 或 $X_n \xrightarrow{a.s.} X$.

(2) 设 $\{X_n, n \geq 1\}$ 是随机变量序列, 若存在随机变量 X 使得 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

则称随机变量序列 $\{X_n, n \geq 1\}$ 依概率收敛于 X , 记为 $X_n \xrightarrow{P} X$.

(3) 设随机变量序列 $\{X_n\} \subset \mathcal{L}^p, p \geq 1, X \in \mathcal{L}^p$, 若有

$$\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0$$

则称随机变量序列 $\{X_n, n \geq 1\}$ p 次平均收敛于 X , 或称 $\{X_n\}$ 在 $\mathcal{L}^p$ 中强收敛于 X . 当 $p = 2$ 时, 称为均方收敛.

(4) 设 $\{F_n(x)\}$ 是分布函数列, 如果存在一个单调不减函



数 $F(x)$,使得在 $F(x)$ 的所有连续点 x 上均有

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记为 $F_n(x) \xrightarrow{W} F(x)$.

设随机变量 X_n, X 的分布函数分别为 $F_n(x)$ 及 $F(x)$, 若 $F_n(x) \xrightarrow{W} F(x)$, 则称 $\{X_n\}$ 依分布收敛于 X , 记为 $X_n \xrightarrow{L} X$.

定理 2.4.2 (1) 随机变量序列 $X_n \xrightarrow{a.s.} X$ 的充分必要条件是 $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{\sup_{m \geq n} |X_m - X| \geq \varepsilon\} = 0$$

(2) 随机变量序列 $X_n \xrightarrow{P} X$ 的充分必要条件是 $\{X_n\}$ 的任意子序列都包含几乎必然收敛于 X 的子序列.



随机变量序列的这4种收敛性之间的关系可以总结为下面的关系图:

几乎必然收敛 $\implies$ 依概率收敛 $\implies$ 依分布收敛;

p 次平均收敛 $\implies$ 依概率收敛 $\implies$ 依分布收敛.

注: 几乎必然收敛与 p 阶矩收敛之间没有蕴含关系.

例 2.4.3 取 $\Omega = (0, 1]$, $\mathcal{F}$ 为 $(0, 1]$ 中全体Borel子集所构成的 σ -代数, P 为Lebesgue测度, 我们可以构造出2个随机变量序列, 其中之一是 r 阶矩收敛的, 但是不几乎必然收敛; 另外一个则几乎必然收敛, 但不是 r 阶矩收敛的.

令

$$Y_{11} = 1; \quad Y_{21} = \begin{cases} 1, & \omega \in (0, \frac{1}{2}] \\ 0, & \omega \in (\frac{1}{2}, 1] \end{cases}$$

103/126

||◀ ▶||

◀ ▶

GoBack

FullScreen

Close

Quit



$$Y_{22} = \begin{cases} 0, & \omega \in (0, \frac{1}{2}] \\ 1, & \omega \in (\frac{1}{2}, 1] \end{cases}$$

一般地, 将 $(0, 1]$ 分成 k 个等长区间, 并且令

$$Y_{ki} = \begin{cases} 1, & \omega \in (\frac{i-1}{k}, \frac{i}{k}] \\ 0, & \omega \notin (\frac{i-1}{k}, \frac{i}{k}] \end{cases} \quad i = 1, \dots, k, \quad k = 1, 2, \dots$$

定义随机变量序列

$$X_1 = Y_{11}, \quad X_2 = Y_{21}, \quad X_3 = Y_{22}, \quad X_4 = Y_{31}, \quad X_5 = Y_{32}, \dots$$

对任意 $\varepsilon > 0$, 由于

$$E|Y_{ki} - 0|^r = \frac{1}{k} \rightarrow 0, \quad (k \rightarrow \infty)$$

可见 $\{X_n\}$ 为 r 阶矩收敛, 但是对任意固定的 $\omega \in \Omega$, 任一自然数 k , 恰有一个 i , 使得 $Y_{ki}(\omega) = 1$, 而对其余的 j 有 $Y_{kj}(\omega) =$



0. 由此知 $\{X_n(\omega)\}$ 中有无穷多个1及无穷多个0, 于是 $\{X_n(\omega)\}$ 对每个 $\omega \in \Omega$ 都不收敛.

如果取

$$Z_n = \begin{cases} n^{\frac{1}{r}}, & \text{若 } \omega \in (0, \frac{1}{n}] \\ 0, & \text{若 } \omega \notin (0, \frac{1}{n}] \end{cases}$$
$$Z = 0, \quad \forall \omega \in \Omega$$

易见, $Z_n(\omega) \rightarrow Z(\omega) \quad \forall \omega \in \Omega$, 所以

$$Z_n \rightarrow Z, \quad a.s.$$

但是 $E(|Z_n - Z|^r) = n \cdot \frac{1}{n} = 1 \not\rightarrow 0$.

下面我们给出积分号下取极限的三大基本定理.

定理 2.4.4 (单调收敛定理) 设 $f_n \in \mathcal{L}$ 的积分存在, $n \geq 1$, 则:

105/126

||◀ ||▶||

◀ ▶

GoBack

FullScreen

Close

Quit



(1) 设 $f_n \uparrow f, a.s.$, 若 $\int_{\Omega} f_1 dP > -\infty$, 则 f 的积分存在, 且 $\int_{\Omega} f_n dP \uparrow \int_{\Omega} f dP$;

(2) 设 $f_n \downarrow f, a.s.$, 若 $\int_{\Omega} f_1 dP < \infty$, 则 f 的积分存在, 且 $\int_{\Omega} f_n dP \downarrow \int_{\Omega} f dP$.

定理 2.4.5 (Fatou引理) 设随机变量序列 $\{X_n\}$ 的期望存在, $n \geq 1$, 则

$$E[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} E[X_n] \leq \limsup_{n \rightarrow \infty} E[X_n] \leq E[\limsup_{n \rightarrow \infty} X_n]$$

定理 2.4.6 (Lebesgue控制收敛定理) 设 $f_n, f \in \mathcal{L}$, $f_n \xrightarrow{a.s.} f$ 或 $f_n \xrightarrow{P} f$. 若存在一非负可积函数 g , 使得 $\forall n \geq 1$ 有 $|f_n| \leq g, a.s.$, 则 f 可积, 且有 $\lim_{n \rightarrow \infty} \int_{\Omega} f_n dP = \int_{\Omega} f dP$.

106/126



GoBack

FullScreen

Close

Quit



§2.5 独立性与条件期望

§2.5.1 独立性

定义 2.5.1 (1) 设 A, B 为两个事件, 若 $P(A \cap B) = P(A)P(B)$, 则称 A 与 B 独立. 更一般地, 设 $A_1, A_2, \dots, A_n$ 为 n 个事件, 如果对任何 $m \leq n$ 及 $1 \leq k_1 < k_2 < \dots < k_m \leq n$, 有

$$P\left(\bigcap_{j=1}^m A_{k_j}\right) = \prod_{j=1}^m P(A_{k_j})$$

则称 $A_1, A_2, \dots, A_n$ 相互独立. $A_1, A_2, \dots, A_n$ 两两独立不一定相互独立.

(2) 设 $\{A_i, i \in I\}$ 是一族事件, 若 对 I 的任意有限子

107/126



GoBack

FullScreen

Close

Quit

集 $\{i_1, \dots, i_k\} \neq \emptyset$ 有

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}) \quad (2.5.1)$$

则称 $\{A_i, i \in I\}$ 是相互独立的.

(3) 设 $\{\mathcal{A}_i, i \in I\}$ 是一族事件类, 如果对 I 的任意有限子集 $\{i_1, \dots, i_k\} \neq \emptyset$, 任意 $A_{i_j} \in \mathcal{A}_{i_j}$ 有(2.5.1)式成立, 则称 $\{\mathcal{A}_i, i \in I\}$ 是独立事件类.

(4) 设 $\{X_i, i \in I\}$ 是 Ω 上一族随机变量, 如果 σ 代数族 $\{\sigma(X_i), i \in I\}$ 是独立事件类, 则称 $\{X_i, i \in I\}$ 相互独立.

容易证明随机变量 $X_1, \dots, X_n$ 独立的充分必要条件是它们的联合分布函数可以分解为

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



定理 2.5.2 (1) 设随机变量 $X_1, \dots, X_n \in \mathcal{L}^1$ 是独立的, 则 $E[\prod_{k=1}^n X_k] = \prod_{k=1}^n E[X_k]$.

(2) 设随机变量 $X_1, \dots, X_n \in \mathcal{L}^2$ 是独立的, 则 $Var[\sum_{k=1}^n X_k] = \sum_{k=1}^n Var[X_k]$.

§2.5.2 条件期望

设 B 是一个事件, 且 $P(B) > 0$. 则事件 B 发生的条件下事件 A 发生的条件概率为

$$P(A|B) = P(A \cap B) / P(B)$$

定理 2.5.3 (全概率公式) 设 $\{B_n\}$ 是 Ω 的一个分割, 且使得 $P(B_n) > 0, \forall n$. 如果 $A \in \mathcal{F}$, 则

$$P(A) = \sum_n P(B_n) P(A|B_n)$$



定理 2.5.4 (Bayes公式) 设 $\{B_n\}$ 是 Ω 的一个分割, 且使得 $P(B_n) > 0, \forall n$, 如果 $P(A) > 0$, 则

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_n P(B_n)P(A|B_n)}, \quad n \geq 1$$

§2.5.3 条件分布

[1]如果 X 与 Y 是离散型随机变量, 对一切使得 $P\{Y = y\} > 0$ 的 y , 给定 $Y = y$ 时, X 的条件概率定义为:

$$P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}}$$

X 的条件分布定义为:

$$F(x|y) = P\{X < x|Y = y\}$$



X 的条件期望定义为:

$$E[X|Y = y] = \int x dF(x|y) = \sum_x x P\{X = x|Y = y\}$$

[2] 如果 X 与 Y 有联合概率密度函数 $f(x, y)$, 则对一切使得 $f_Y(y) \geq 0$ 的 y , 给定 $Y = y$ 时, X 的条件概率密度函数定义为:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

X 的条件分布定义为:

$$F(x|y) = P\{X < x|Y = y\} = \int_{-\infty}^x f(x|y) dx$$

X 的条件期望定义为:

$$E[X|Y = y] = \int x dF(x|y) = \int x f(x|y) dx$$



[3] 我们以 $E[X|Y]$ 表示随机变量 Y 的函数, 它在 $Y = y$ 时, 取值为 $E[X|Y = y]$!!!!

例 2.5.5 设随机变量 (X, Y) 的联合概率密度为

$$f(x, y) = \begin{cases} \frac{1}{2}e^{-y}, & y > |x| \\ 0, & \text{others} \end{cases}$$

求 $E(Y|X = x)$.

Proof We have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \int_{|x|}^{\infty} \frac{1}{2}e^{-y} dy = \frac{1}{2}e^{-|x|}. \end{aligned}$$

Therefore,

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = e^{-y+|x|}, \quad y > |x|.$$

112/126



GoBack

FullScreen

Close

Quit



Then, we get

$$E(Y|X = x) = \int_{|x|}^{\infty} ye^{-y+|x|} dy = 1 + |x|.$$

例 2.5.6 已知 $X \sim N(\mu, \tau^2)$, 给定 $X = x$ 的条件下, Y 的条件分布为 $N(x, \sigma^2)$, 求 Y 的分布及给定 $Y = y$ 的条件下 X 的条件分布.

Proof We know that

$$f_X(x) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{(x-\mu)^2}{2\tau^2}}$$
$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}.$$

Then, we have

$$f(x, y) = f_X(x)f(y|x) = \frac{1}{2\pi\tau\sigma} e^{-\frac{(x-\mu)^2}{2\tau^2} - \frac{(y-x)^2}{2\sigma^2}}.$$

113/126



GoBack

FullScreen

Close

Quit

Furthermore,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \frac{1}{2\pi\tau\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\tau^2} - \frac{(y-x)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2)}} e^{-\frac{(y-\mu)^2}{2(\tau^2 + \sigma^2)}}. \end{aligned}$$

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \dots$$



114/126



GoBack

FullScreen

Close

Quit

条件期望的数学性质

定理 2.5.7 设函数 $g(x)$ 在 R 上连续, 若

$$\int_{-\infty}^{\infty} |g(x)| dF(x|y) < +\infty.$$

则随机变量 $g(X)$ 在 $Y = y$ 条件下的条件数期望为

$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) dF(x|y).$$

定义 2.5.8 定义

$$D(X|Y = y) := E(X - E[X|Y = y])^2$$

为 $Y = y$ 的条件下, 随机变量 X 的条件方差.

定理 2.5.9 设 X, Y, Z 是 $(\Omega, \mathcal{F}, P)$ 上的随机变量, g, h 为 R 上连续函数, 且各数学期望存在. 则有

- $E(c|Y) = c, c$ 为常数。



- $E(aX + bY|Z) = aE(X|Z) + bE(Y|Z)$, a, b 为常数。
- 如果 X 与 Y 独立, 则 $E(X|Y) = E(X)$.
- $E(g(X)h(Y)|X) = g(X)E(h(Y)|X)$ and $E(g(X)h(Y)|Y) = h(Y)E(g(X)|Y)$.
- $E(E(g(X, Y)|Y)) = E(g(X, Y))$.

116/126



GoBack

FullScreen

Close

Quit

全期望公式

定理 2.5.10 对一切随机变量 X 和 Y ，当期望存在时

$$E[X] = E[E[X|Y]] = \int E[X|Y = y]dF_Y(y) \quad (2.5.2)$$

特别的：

- 当 Y 为一个离散随机变量时，(2.5.2)式为

$$E[X] = \sum_y E[X|Y = y]P\{Y = y\}$$

- 当 Y 为一个连续随机变量时，(2.5.2)式为

$$E[X] = \int_{-\infty}^{+\infty} E[X|Y = y]f(y)dy$$

例 2.5.11 (随机个随机变量之和) 设 $X_1, X_2, \dots$ 是一列与 X 独立同分布的随机变量；设 N 为一非负整值随机变



量, 且与序列 $X_1, X_2, \dots$ 独立.

求 $Y = \sum_{i=1}^N X_i$ 的均值和方差.

解 首先在对 N 取条件的情况下来计算 $Y = \sum_{i=1}^N X_i$ 的矩母函数, 即

$$E[\exp\{t \sum_{i=1}^N X_i\} | N = n] = E[\exp\{t \sum_{i=1}^n X_i\}] = (\phi_X(t))^n$$

其中 $\phi_X(t)$ 是随机变量 X 的矩母函数, 因此

$$E[\exp\{t \sum_{i=1}^N X_i\} | N] = (\phi_X(t))^N$$

从而

$$\phi_Y(t) = E[\exp\{t \sum_{i=1}^N X_i\}] = E[(\phi_X(t))^N]$$



现在对 $\phi_Y(t)$ 求导得

$$\phi_Y'(t) = E[N(\phi_X(t))^{N-1}\phi_X'(t)]$$

再求一次导数得

$$\phi_Y''(t) = E[N(N-1)(\phi_X(t))^{N-2}(\phi_X'(t))^2 + N(\phi_X(t))^{N-1}\phi_X''(t)]$$

计算在 $t = 0$ 点的值, 得

$$E[Y] = E[NE[X]] = E[N]E[X] \quad (2.5.3)$$

及

$$\begin{aligned} E[Y^2] &= E[N(N-1)(E[X])^2 + NE[X^2]] \\ &= E[N]Var[X] + E[N^2](E[X])^2. \end{aligned}$$

因此有

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



120/126

例 2.5.12 一个矿工陷进一个有三个门的矿井。第一个门通向一个隧道，沿此隧道走2小时的旅程他可到达安全地；第二个门通向另一个隧道，沿此隧道走3小时的旅程会使他回到矿井；第三个门通向一个隧道，沿此隧道走3小时的旅程会使他回到矿井。假定矿工总是等可能地选取任意一个门，用 X 表示矿工到达安全第所需的时间，求 X 的矩母函数。

Proof 写出 X 的分布列很困难”故无法直接求其均值！令 Y 表示矿工第一次选择的门”则 $\{Y = i\}$ 表示第一次选



GoBack

FullScreen

Close

Quit



择第 i 个门”由题意知

$$P(Y = 1) = P(Y = 2) = P(Y = 3) = \frac{1}{3}.$$

$$\begin{aligned} E[X] &= E[E[X|Y]|Y] \\ &= \frac{1}{3} (E[X|Y = 1] + E[X|Y = 2] + E[X|Y = 3]) \\ &= \frac{1}{3} (2 + 3 + E[X] + 5 + E[X]). \end{aligned}$$

由此可得：

$$E[X] = 10.$$

有条件概率公式得到：

$$\begin{aligned} E[e^{tX}] &= E[E[e^{tX}]|Y] \\ &= \frac{1}{3} (E[e^{tX}|Y = 1] + E[e^{tX}|Y = 2] + E[e^{tX}|Y = 3]). \end{aligned}$$

121/126



GoBack

FullScreen

Close

Quit



注意到:

$$E[e^{tX}|Y=1]=e^{2t}$$

$$E[e^{tX}|Y=2]=E[e^{t(\tilde{X}+3)}]=e^{3t}E[e^{tX}]$$

$$E[e^{tX}|Y=3]=E[e^{t(\tilde{X}+5)}]=e^{5t}E[e^{tX}].$$

其中, $\tilde{X}$ 与 X 有相同的分布。由此可得:

$$E[e^{tX}] = \frac{e^{2t}}{3 - e^{3t} - e^{5t}}.$$

例 2.5.13 设某段时间内到达商场的顾客人数 N 服从参数为 λ 的泊松分布. 每位顾客在该商场的消费额 X 服从 $[a, b]$ 上的均匀分布. 各位顾客之间消费是相互独立的且与 N 独立. 求顾客在该商场总的消费额.

例 2.5.14 设随机变量 Y 的 $E(Y)$ 存在, $Y = \sum_{k=1}^N X_k$ 及 N 都是随机变量, 并且相互独立. N 仅取自然数, $E(N)$ 存



在. 证明:

$$E(Y) = \sum_{k=1}^{\infty} P(N \geq k) E(X_k).$$

例 2.5.15 已知随机变量 X 服从 $[0, a]$ 上的均匀分布, 随机变量 Y 服从 $[X, a]$ 上的均匀分布, 试求

$$E(Y|X = x), \quad 0 < x < a, \quad \text{and} \quad E(Y).$$

Proof 由条件知对 $x > 0$, 有

$$f(y|x) = \begin{cases} \frac{1}{a-x}, & x < y < a \\ 0 & \text{others} \end{cases}$$

对任意的 $0 < x < a$,

$$E(Y|X = x) = \int_x^a \frac{y}{a-x} dy = \frac{a+x}{2}.$$

$$E(Y) = E(E(Y|X)) = E\left(\frac{a+X}{2}\right) = \frac{3a}{4}.$$

123/126



GoBack

FullScreen

Close

Quit



选读

最后我们将条件期望推广到一般随机变量及 σ 代数情形.

定义 2.5.16 设 X 是随机变量且 $E[|X|] < \infty$. 若对每个子 σ 代数 $\mathcal{G} \subset \mathcal{F}$, 存在唯一的(几乎必然相等的意义下)随机变量 X^* , 有 $E[|X^*|] < \infty$, 使得 X^* 是 $\mathcal{G}$ 可测随机变量(即对任何的 $a \in \mathbb{R}$, 有 $\{X^* \leq a\} \in \mathcal{G}$), 且

$$E[X^* I_B] = E[X I_B], \quad \forall B \in \mathcal{G}$$

则称随机变量 X^* 为 X 在给定 $\mathcal{G}$ 下的条件期望, 记为 $X^* = E[X|\mathcal{G}]$. 即

$$\int_B E[X|\mathcal{G}] dP = \int_B X dP, \quad \forall B \in \mathcal{G}$$

124/126



GoBack

FullScreen

Close

Quit



定理 2.5.17 条件期望有如下基本性质:

- (1) $E[E[X|\mathcal{G}]] = E[X]$.
- (2) 若 X 是 $\mathcal{G}$ 可测, 则 $E[X|\mathcal{G}] = X, a.s..$
- (3) 设 $\mathcal{G} = \{\emptyset, \Omega\}$, 则 $E[X|\mathcal{G}] = E[X], a.s..$
- (4) $E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}], a.s..$
- (5) 若 $X \leq Y, a.s..$, 则 $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}], a.s..$
- (6) 若 a, b 为实数, $X, Y, aX + bY$ 的期望存在, 则

$$E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}], a.s..$$

如果右边和式有意义.

- (7) $|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}], a.s..$
- (8) 设 $0 \leq X_n \uparrow X, a.s..$, 则 $E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}], a.s..$
- (9) 设 X 及 XY 的期望存在, 且 Y 为 $\mathcal{G}$ 可测, 则

$$E[XY|\mathcal{G}] = YE[X|\mathcal{G}], a.s..$$

125/126



GoBack

FullScreen

Close

Quit



(10) 若 X 与 $\mathcal{G}$ 相互独立(即 $\sigma(X)$ 与 $\mathcal{G}$ 相互独立), 则有

$$E[X|\mathcal{G}] = E[X], a.s.$$

(11) 若 $\mathcal{G}_1, \mathcal{G}_2$ 是两个子 σ 代数, 使得 $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, 则

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1], a.s.$$

(12) 若 X, Y 是两个独立的随机变量, 函数 $g(x, y)$ 使得 $E[|g(X, Y)|] < +\infty$, 则有

$$E[g(X, Y)|Y] = E[g(X, y)]|_{y=Y}, \quad a.s.$$

这里 $E[g(X, y)]|_{y=Y}$ 的意义是, 先将 y 视为常数, 求得数学期望 $E[g(X, y)]$ 后再将随机变量 Y 代入到 y 的位置.

126/126



GoBack

FullScreen

Close

Quit