概率论与数理统计

Теория Вероятностей и Математическая Статистика

作者：Galois 爱求五次根

组织：深北莫数学学社分析小组

时间：2023/7/5

版本：1.0

宗旨：执象而求，咫尺千里

美是首要的试金石：丑陋的数学不可能永存。——G. H. Hardy 哈代

目录

第 1 章 环 (Ring)

1.1 集环与 σ 环

Ring of Sets and σ-Ring

1.1.1 集环与半环

Ring of Sets and Semi-Ring

定义 1.1 (集环)

设 $\mathfrak{R}$ 为非空集族，若

$$(\forall A, B \in \mathfrak{R}) : [(A \triangle B \in \mathfrak{R}) \cap (A \cap B \in \mathfrak{R})]$$

则称 $(\mathfrak{R}, \cap, \Delta)$ 为集合环 (кольцо множества/ring of sets) 或简称集环

♣

命题 1.1 (集环定义合理性)

因为对于任何 A 与 B 均有

$$A \cup B = (A \triangle B) \triangle (A \cap B), \quad A \backslash B = A \triangle (A \cap B)$$

而由 $A \in \mathfrak{R}$ 与 $B \in \mathfrak{R}$ 可推出集 $A \cup B$ 与 $A \backslash B$ 也属于 $\mathfrak{R}$，则 $\mathfrak{R}$ 关于有限交与对称差的运算为封闭集族，即 $(\mathfrak{R}, \cap, \Delta)$ 为一个环

♠

注 另外，命题也指出集环关于任意有限并与有限交也封闭

命题 1.2 (最小集环存在性)

存在最小集环

♠

证明

定义 1.2 (单位与集代数)

设 $\mathfrak{S}$ 为集族，$E \in \mathfrak{S}$，若

$$(\forall A \in \mathfrak{S}) : A \cap E = A$$

则称 E 为 $\mathfrak{S}$ 的单位 (unit)。称具有单位的集环为集代数 (algebra of sets)

♣

定理 1.1 (非空集族生成环存在唯一性)

对于任意非空集族 $\mathfrak{S}$ ，下列命题成立：

1) (非空集族生成环存在性) 存在包含集族 $\mathfrak{S}$ 的集环
2) (非空集合生成环极小性 (唯一性)) 若包含集族 $\mathfrak{S}$ 的集环 $\mathfrak{R}(\mathfrak{S})$ 为任意包含 $\mathfrak{R}$ 的集环的子集环，则集环 $\mathfrak{R}(\mathfrak{S})$ 唯一

♡

证明 考虑 $X = \{\bigcup\limits_{A \in \mathfrak{S}} A | A \in \mathfrak{S}\}$，设 Σ 为属于 $\mathcal{P}(X)$ 且又包含 $\mathfrak{S}$ 的一切集环的总体，则显然所有这些环的交 $\mathfrak{B} = \bigcup\limits_{\mathfrak{R} \in \Sigma} \mathfrak{R}$ 即为所要求的环 $\mathfrak{R}(\mathfrak{S})$，其显然也是唯一的

注 该定理表明，对于任意包含非空集族 $\mathfrak{S}$ 的集环 $\mathfrak{R}^*$，交 $\mathfrak{R} = \mathfrak{R}^* \cap \mathcal{P}(X)$ 为 Σ 中的集环，则 $\mathfrak{S} \subset \mathfrak{R} \subset \mathfrak{R}^*$，即 $\mathfrak{R}$ 满足极小性要求，由此称该环为 $\mathfrak{S}$ 上的**极小环** (minimal ring) 或 $\mathfrak{S}$ 的**生成环** (порождённое кольцо)，记为 $\mathfrak{R}(\mathfrak{S})$

定义 1.3 (集半环)

设集族 $\mathfrak{S}$，若满足下列条件：

1) (含幺性) $\varnothing \in \mathfrak{S}$
2) (有限交封闭性) $(\forall A, B \in \mathfrak{S}): A \cap B \in \mathfrak{S}$
3) (有限覆盖性)

$$(\forall A, A_1 \in \mathfrak{S})\left[(A_1 \subset A) \Rightarrow (\exists A_2, A_3, \cdots, A_n \in \mathfrak{S})\left(A = \bigcup_{k=1}^{n} A_k\right)\right]$$

则称 $\mathfrak{S}$ 为集半环 (semi-ring of sets)，经常简称为半环 (semi-ring/полукольцо)

♣

命题 1.3 (集环半环性)

集环均为半环

♠

证明 若 A 与 $A_1 \subset A$ 属于集环 $\mathfrak{R}$，则成立分解式 $A = A_1 \cup A_2$，其中 $A_2 = A \backslash A_1 \in \mathfrak{R}$

定义 1.4 (有限分解)

若任一组两两不相交的集 $A_1, A_2, \cdots, A_n$ 的并为集 A，称该并为集 A 的有限分解

♣

引理 1.1 (集有限分解存在性)

设集 $A_1, A_2, \cdots, A_n, A$ 属于半环 $\mathfrak{S}$，其中集 A_i 两两不交且皆包含于 A，则存在集 $A_{n+1}, \cdots, A_s \in \mathfrak{S}$ 使得集 A 的有限分解为

$$A = \bigcup_{k=1}^{s} A_k \quad (s \geqslant n)$$

♡

证明 归纳：当 $n = 1$ 时由半环定义引理显然成立。假定当 $n = m$ 时引理成立，考虑满足引理条件的 $m + 1$ 个集 $A_1, \cdots, A_m, A_{m+1}$。由条件有 $A = A_1 \cup A_2 \cup \cdots \cup A_m \cup B_1 \cup B_2 \cup \cdots \cup B_p$，其中 $(\forall q \in \overline{1;p}): B_q \in \mathfrak{S}$，下记 $B_{q1} = A_{m+1} \cap B_q$，则有有限分解 $B_q = B_{q1} \cup B_{q2} \cup \cdots \cup B_{qr_q}$，其中任意 B_{qj} 皆属于 $\mathfrak{S}$。不难看出

$$A = A_1 \cup \cdots \cup A_m \cup A_{m+1} \cup \left(\bigcup_{q=1}^{p}\left(\bigcup_{j=2}^{r_q} B_{qj}\right)\right)$$

则当 $n = m + 1$ 时引理断言得证，由归纳原理，引理即证

引理 1.2 (有限集族有限分解存在性)

任意有限个集 $A_1, \cdots, A_n$ 属于半环 $\mathfrak{S}$，则存在有限个两两不交集 $B_1, \cdots, B_t \in \mathfrak{S}$ 使得任意 $A_k(k = 1, 2, \cdots)$ 均可表示为某些集 B_s 的并：

$$A_k = \bigcup_{s \in M_k} B_s$$

其中 M_k 为指标集

♡

证明 归纳：当 $n = 1$ 时，仅需令 $t = 1, B_1 = A_1$ 即有引理成立。假设当 $n = m$ 时引理成立，考虑 $\mathfrak{S}$ 中一组集 $A_1, \cdots, A_m, A_{m+1}$。设 $B_1, B_2, \cdots, B_t$ 为满足引理关于 $A_1, A_2, \cdots, A_m$ 条件的 $\mathfrak{S}$ 中的集，记 $B_{s1} = A_{m+1} \cap B_s$，则由集有限分解存在性引理 (1.1) 有 $\mathfrak{S}$ 上有限分解

$$A_{m+1} = \left(\bigcup_{s=1}^{t} B_{s1}\right) \cup \left(\bigcup_{p=1}^{q} B'_p\right)$$

而由半环定义有表达式 $B_s = B_{s1} \cup B_{s2} \cup \cdots \cup B_{sf_s}$，其中 $B_{sj} \in \mathfrak{S}$，由归纳假设有

$$(\forall k \in \overline{1;m}) : A_k = \bigcup_{s \in M_k} \bigcup_{j=1}^{f_s} B_{sj}$$

则集 B_{sj}, B'_p 两两不交，则有集 B_{sj}, B'_p 满足引理关于 $A_1, \cdots, A_m, A_{m+1}$ 的条件，由归纳原理，定理即证

定理 1.2 (半环生成环充分条件)

设 $\mathfrak{S}$ 为半环，若

$$(\forall A \in \mathfrak{R}(\mathfrak{S}))(\exists A_1, \cdots, A_n \in \mathfrak{S}) : A = \bigcup_{k=1}^{n} A_k$$

则 $\mathfrak{R}(\mathfrak{S})$ 为半环 $\mathfrak{S}$ 的生成环 (кольцо,порождённое полукольиом)

证明 设集合 $\mathfrak{R} = \left\{ A = \bigcup\limits_{k=1}^{n} A_k, A_k \in \mathfrak{S} \right\}$，下证该集为环。设 $A = \bigcup\limits_{k=1}^{n} A_k, B = \bigcup\limits_{i=1}^{m} B_i$，其中 $A_k, B_i \in \mathfrak{S}$，则有

$$A \backslash B = \left(\bigcup_{k=1}^{n} A_k \right) \backslash \left(\bigcup_{i=1}^{m} B_i \right) = \bigcup_{k=1}^{n} \left(A_k \backslash \left(\bigcup_{i=1}^{m} B_i \right) \right) = \bigcup_{k=1}^{n} \left(A_k \backslash \left(\bigcup_{i=1}^{m} (A_k \cap B_i) \right) \right)$$

即有 $A \backslash B = \bigcup\limits_{k=1} \bigcup\limits_{i=1}^{n} \tilde{A}_{ik}$，其中 $\tilde{A}_{ik} \in \mathfrak{S}$，则有 $A \backslash B \in \mathfrak{R}$

若 $A \cap B = \varnothing$，则 $A \cup B \in \mathfrak{R}$；若 $A \cap B \neq \varnothing$，则 $A \cup B = (A \backslash B) \cup B \in \mathfrak{R}$，于是有

$$A \Delta B = (A \backslash B) \cup (B \backslash A) \in \mathfrak{R}, \quad A \cap B = (A \cup B) \Delta (A \Delta B) \in \mathfrak{R}$$

则 $\mathfrak{R}$ 为集环

1.1.2 σ 环与 σ 代数
σ-Ring and σ-Algebra

定义 1.5 (σ 环与 σ 代数)

设集环 $\mathfrak{R}$，若满足对于可数并运算封闭，即

$$(\forall A_1, A_2, \cdots, A_n, \cdots \in \mathfrak{R}) : \bigcup_{i=1}^{+\infty} A_i \in \mathfrak{R}$$

则称集环为 σ 环 (σ-кольцо)，称具有单位的 σ 环 (σ-кольцо с единицей) 为 σ 代数 (σ-алгебра)

定义 1.6

(β 环与 β 代数) 若满足对于可数交运算封闭，即

$$(\forall A_1, A_2, \cdots, A_n, \cdots \in \mathfrak{R}) : \bigcap_{i=1}^{+\infty} A_i \in \mathfrak{R}$$

则称集环为 β 环 (β-кольцо)，称具有单位的 β 环 (β-кольцо с единицей) 为 β 代数 (β-алгебра)

注 （σ 代数与 β 代数定义等价性）

从对偶关系

$$\bigcup_{i=1}^{+\infty} A_i = E \backslash \left(\bigcap_{i=1}^{+\infty} (E \backslash A_i) \right), \quad \bigcap_{i=1}^{+\infty} A_i = E \backslash \left(\bigcup_{i}^{+\infty} (E \backslash A_i) \right)$$

即得任意 σ 代数同时为 δ 代数，而任意 δ 代数同时也为 σ 代数。由此今后仅限于研究 σ 代数

例题 1.1 （σ 代数）

任意集的幂集为 σ 代数

命题 1.4 (集族上 σ 代数存在性)

设 $\mathfrak{S}$ 为集族，则存在一个包含该集族的 σ 代数

证明 令 $X=\bigcup\limits_{A\in\mathfrak{S}}A$，则有 X 的幂集 $\mathcal{P}$ 为包含 $\mathfrak{S}$ 的 σ 代数

定义 1.7 (不可约 σ 代数)

设集族 $\mathfrak{S}$，令 $X=\bigcup\limits_{A\in\mathfrak{S}}A$，则存在包含集族 $\mathfrak{S}$ 的 σ 代数 $\mathfrak{R}$，其为 X 的幂集。若 $X=\bigcup\limits_{A\in\mathfrak{S}}A$ 为该 σ 代数的单位，则称该 σ 代数对于族 $\mathfrak{S}$ 不可约 (неприводимой по отношению к системе $\mathfrak{S}$)，不引起歧义时，简称为不可约 σ 代数 (неприводимая σ-алгебра)

注 (不可约 σ 代数)

不可约 σ 代数为不包含不属于任意 $A\in S$ 的点的 σ 代数。暂限于研究这种 σ 代数

定理 1.3 (非空集族生成不可约 σ 代数)

对于任意非空集族 $\mathfrak{S}$，都存在（关于该族）不可约的 σ 代数 $\mathfrak{B}(\mathfrak{S})$，其包含 $\mathfrak{S}$ 且属于包含 $\mathfrak{S}$ 的任意 σ 代数

证明 证明方法同非空集族生成环存在唯一性定理 (1.1)

注 称满足定理的 σ 代数 $\mathfrak{B}(\mathfrak{S})$ 为族 $\mathfrak{S}$ 上的**极小 σ 代数**，或由族 $\mathfrak{S}$ 生成的 σ 代数 (σ-алгебра,порожденная системой $\mathfrak{S}$)，简称族 $\mathfrak{S}$ 的**生成 σ 代数** (порожденная σ-алгебра)，记为 $\sigma(\mathfrak{S})$

定义 1.8 (Borel 运算)

称对集族有限或可数次取补集、取交集、取并集以及它们的混合运算为 Borel 运算

设 $\mathcal{A}$ 为空间 Ω 的子集族，则 $\mathcal{A}$ 中集合由所有 Borel 运算得到的 σ 代数 $\mathcal{F}$ 为由 $\mathcal{A}$ 生成的 σ 代数或 $\mathcal{A}$ 的生成 σ 代数，记为 $\mathcal{F}=\sigma(\mathcal{A})$

注 (集函数符号约定)

设 $y=f(x)$ 为在集 M 上的定义的且在集 N 中取值的函数，并设 $\mathfrak{M}$ 为集 M 的子集组成的某一集族，用 $f(\mathfrak{M})$ 表示所有属于 $\mathfrak{M}$ 的集的像 $f(A)$ 组成的族。此外，设 $\mathfrak{N}$ 为包含在 N 中的集的某一集族，$f^{-1}(\mathfrak{N})$ 为属于 $\mathfrak{N}$ 的集的一切原像 $f^{-1}(A)$ 组成的族

性质 (集函数关于集环、集代数、σ 代数性质)

下列命题成立：

1) 若 $\mathfrak{N}$ 为集环，则 $f(\mathfrak{N})$ 也为集环
2) 若 $\mathfrak{N}$ 为集代数，则 $f^{-1}(\mathfrak{N})$ 也为集代数
3) 若 $\mathfrak{N}$ 为 σ 代数，则 $f^{-1}(\mathfrak{N})$ 也为 σ 代数
4) $\mathfrak{N}\left(f^{-1}(\mathfrak{N})\right)=f^{-1}(\mathfrak{N}(\mathfrak{N}))$
5) $\mathfrak{N}\left(f^{-1}(\mathfrak{N})\right)=f^{-1}(\mathfrak{N}(\mathfrak{N}))$

定义 1.9 (Borel 子集与 Borel σ 代数)

设 $\mathfrak{M}$ 为 R^n 上所有开球的集合，则称 $\mathfrak{A}=\sigma(\mathfrak{M})$ 为在 R^n 上子集的（n 维）Borel σ-代数 (борелевская σ-алгебра)，称 $\mathfrak{A}$ 的元素为在 R^n 上的 Borel 子集 (борелевское подмножество)

例题 1.2 (Borel 子集与 Borel σ 代数)

设 $\mathfrak{M}$ 为 R^1 上所有积分的集合，则 $\mathfrak{A}=\sigma(\mathfrak{M})$ 为 Borel σ 代数，而 $A\in\mathfrak{A}$ 为 Borel 子集

命题 1.5 (一维 Borel σ 代数)

设实直线上集合 $\mathfrak{A}_1=\{(-\infty,x)\mid x\in\mathbb{R}\},\mathfrak{A}_2=\{(a,b)\mid -\infty<a<b<\infty\}$，则有下列命题成立：

1)$\sigma(\mathfrak{A}_1)=\sigma(\mathfrak{A}_2)$

2)$\sigma(\mathfrak{A}_1)$ 为 Borel σ 代数

证明 1) 记 $\mathfrak{F}_1=\sigma(\mathfrak{A}_1),\mathfrak{F}_2=\sigma(\mathfrak{A}_2)$，首先有

$$(\forall x\in\mathbb{R}):(-\infty,x)=\bigcup_{n=1}^{\infty}(x-n,x)$$

其是可数个有界开区间的并集，则有 $(\forall x\in\mathbb{N}):(-\infty,x)\in\sigma(\mathfrak{A}_2)=\mathfrak{F}_2$，即有 $\mathfrak{A}_1\subset\mathfrak{F}_2$，则有 $\mathfrak{F}_1=\sigma(\mathfrak{A}_1)\subset\mathfrak{F}_2$。反之，对任意满足 $a<b$ 的实数 a 与 b 都有

$$[a;b)=(-\infty;b)-(-\infty;a)\in\sigma(\mathfrak{A}_1)=\mathfrak{F}_1$$

以及

$$\{a\}=\bigcap_{n=1}^{\infty}\left[a;a+\frac{1}{n}\right)\in\sigma(\mathcal{A}_1)=\mathcal{F}_1$$

则有 $(a;b)=[a;b)-\{a\}=(a;b)\in\sigma(\mathfrak{A}_1)=\mathfrak{F}_1$，即 $\mathfrak{A}_2\subset\mathfrak{F}_1$，则有 $\mathfrak{F}_2=\sigma(\mathcal{A}_2)\subset\mathcal{F}_1$。综上即得 $\sigma(\mathfrak{A}_1)=\sigma(\mathfrak{A}_2)$

2) 由实直线上任意开集均为至多可数个有界开区间的并集，则在 σ 代数 $\sigma(\mathfrak{A}_1)$ 中包含了所有开集，则 $\sigma(\mathcal{A}_1)$ 包含了 Borel σ 代数。反之，由 $\mathfrak{A}_1=\{(-\infty,x)\mid x\in\mathbb{R}\}$ 为实直线上开集族的子族，则 $\sigma(\mathcal{A}_1)$ 必包含在 Borel σ 代数中，综上即知 $\sigma(\mathcal{A}_1)$ 为 Borel σ 代数

注 该命题表明，在由 $\mathfrak{A}_1=\{(-\infty,x)\mid x\in\mathbb{R}\}$ 生成的 σ 代数 $\sigma(\mathfrak{A}_1)$ 中包含了所有的有界开区间和单点集，因此也包含了所有有界区间

定理 1.4 (一维 Borel σ 代数结构)

实直线上下述 σ 代数等价，且均为一维 Borel σ 代数

1) 全体有界开区间生成的 σ 代数
2) 子集族 $\{(-\infty,x)\mid x\in\mathbb{R}\}$ 生成的 σ 代数
3) 全体开集和闭集生成的 σ 代数

注 一般将实直线上的 Borel σ 代数记为 $\mathfrak{B}^1$

第 2 章 测度 (Measure)

2.1 基本概念 Basic Concept

2.1.1 集半环上测度与扩张 Measure on Semi-Ring and Extention

定义 2.1 (集半环测度)

若集函数 $\mu(A): \mathfrak{S} \to \mathbb{R}$ 满足下列条件:

1) (集半环定义域) 集族 $\mathfrak{S}$ 为集半环
2) (值域非负性) 集函数 $\mu(A)$ 的值为非负实数
3) (有限可加性 (свойство аддитивности)) 基函数 $\mu(A)$ 有限可加, 即

$$(\forall A_1, A_2, \cdots A_n \in \mathfrak{S}): \left[\left(A = \bigcup_{i=1}^{n} A_i\right) \Rightarrow \left(\mu(A) = \sum_{k=1}^{n} \mu(A_k)\right)\right]$$

其中 $A_i \cap A_j = \varnothing$

则称该集函数 $\mu(A)$ 为集 A 上测度 (measure/мера) ♣

性质 (空集零测度性)

设集函数 $\mu(A)$ 为测度, 则有 $\mu(\varnothing) = 0$

证明 从 $\varnothing = \varnothing \cup \varnothing$ 有 $\mu(\varnothing) = 2\mu(\varnothing)$, 即 $\mu(\varnothing) = 0$

定义 2.2 (σ 可加性/可数可加性)

设集函数 $\mu(A): \mathfrak{S} \to \mathbb{R}$ 为定义在集半环 $\mathfrak{S}$ 上测度, 若

$$(\forall A_1, A_2, \cdots A_n, \cdot \in \mathfrak{S}): \left[\left(A = \bigcup_{i=1}^{+\infty} A_i\right) \Rightarrow \left(\mu(A) = \sum_{k=1}^{+\infty} \mu(A_k)\right)\right]$$

其中 $A_i \cap A_j = \varnothing$, 则称 $\mu(A)$ 可数可加 (счётно-аддитивная) 或 σ 可加 (σ-аддитивная) ♣

定义 2.3 (外测度)

设 $\mathfrak{S}_m$ 为半环, $B_n \in \mathfrak{S}_m$, 称数

$$\mu^*(A) = \inf \sum_n m(B_n)$$

为集 $A \subset E$ 的外测度, 其中下确界从集 A 的有限覆盖或可数覆盖中取 ♣

定理 2.1 (半环测度保序性)

设 $\mathfrak{S}$ 为半环, 则有

$$(\forall A, B \in \mathfrak{S})[(A \subset B) \Rightarrow (\mu(A) \leqslant \mu(B)]$$

(半环外测度可数保序性): 若集 $A \subset \bigcup_n A_n$, 其中 $\{A_n\}$ 为有限集族或可数集族, 则有

$$\mu^*(A) \leqslant \sum_n \mu^*(A_n)$$

♡

证明 由 $B = A \cup (B \backslash A) = A \cup \left(\bigcup_{i=1}^{n} A_i\right), A_i \in \mathfrak{S}$ 即得 $\mu(B) = \mu(A) + \sum_{i=1}^{n} \mu(A_i) \geqslant \mu(A)$

定理 2.2 (半开半闭区间测度可数可加性)
设半环 $\mathfrak{S}=\{[a;b)|a\leqslant b\}$，定义半开半闭区间的测度为 $m[a;b)=b-a$，则半环 $\mathfrak{S}$ 上测度 m 可数可加 ♡

证明 设 $[a;b)=\bigcup\limits_{i=1}^{\infty}[a_i;b_i)$，则有

$$\bigcup_{i=1}^{\infty}[a_i;b_i)\subset[a;b)\Rightarrow\sum_{i=1}^{\infty}m\left[a_i;b_i\right)\leqslant m[a;b)\Rightarrow\sum_{i=1}^{\infty}(b_i-a_i)\leqslant b-a$$

不妨设 $b>a$，另设 $(\forall i\in\mathbb{N}^*):b_i>a_i$，则对于 $\forall\varepsilon>0$ 有

$$[a;b-\varepsilon]\subset[a;b)\subset\bigcup_{i=1}^{\infty}[a_i;b_i)\subset\bigcup_{i=1}^{\infty}\left(a_i-\frac{\varepsilon}{2^i};b_i\right)$$

由 Heine-Borel 引理有 $(\exists m>0)$ 满足

$$[a;b-\varepsilon)\subset\bigcup_{i=1}^{m}\left(a_i-\frac{\varepsilon}{2^i};b_i\right)\subset\bigcup_{i=1}^{m}\left[a_i-\frac{\varepsilon}{2^i};b_i\right)\Rightarrow$$

$$m[a,b-\varepsilon)\leqslant\sum_{i=1}^{m}m\left[a_i-\frac{\varepsilon}{2^i},b_i\right)\Rightarrow b-a-\varepsilon\leqslant\sum_{i=1}^{m}\left(b_i-a_i+\frac{\varepsilon}{2^i}\right)\leqslant\sum_{i=1}^{m}(b_i-a_i)+\varepsilon$$

则令 $\varepsilon\to0$ 即有 $b-a\leqslant\sum\limits_{i=1}^{+\infty}(b_i-a_i)$

例题 2.1 (Cantor 集/множество Кантора/测度)

考虑数 $x\in I=[0;1)$ 的三进制表示

$$x=\frac{a_1}{3}+\frac{a_2}{3^2}+\frac{a_3}{3^3}+\ldots,\quad a_i\in\{0,1,2\}$$

现设 G_1 为数集，其中 $a_1=1,K_1=I\backslash G_1$；设 G_2 为数集，其中 $a_1\neq1,a_2=1,K_2=K_1\backslash G_2$；设 G_3 为数集，其中 $a_1\neq1,a_2\neq1,a_3=1,K_3=K_2\backslash G_3$，以此类推⋯⋯则得到了 Cantor 集：

$$K=\bigcap_{i=1}^{\infty}K_i\neq\varnothing$$

该集合不包含三进制表示法至少包含一位数字 1 的数

设 $G=\bigcup G_n$，下计算该集合的测度：

$$\mu(G)=\sum_{n=1}^{\infty}\mu\left(G_n\right)=\sum_{n=1}^{\infty}2^{n-1}\cdot\frac{1}{3^n}=1\Rightarrow\mu(K)=0$$

注 另外，Cantor 集 K 有连续统基数：

$$K=\{x=(0,x_1x_2x_3\ldots)_3,x_i\in\{0,2\}\}\sim\{x=(0,x_1x_2x_3\ldots)_2,x_i\in\{0,1\}\}=(0,1)$$

定义 2.4 (测度扩张)
设集族 $\mathfrak{S}_m\subset\mathfrak{S}_\mu$，若

$$(\forall A\in\mathfrak{S}_m):\mu(A)=m(A)$$

则称测度 μ 为测度 m 的扩张 (продолжение) ♣

定理 2.3 (半环测度生成环上扩张)
设半环 $\mathfrak{S}$ 上有测度 $\hat{\mu}$，则下列命题成立：

1) (半环测度在生成环上扩张存在唯一性) 在环 $\mathfrak{R}(\mathfrak{S})$ 上存在且仅存在唯一的测度 μ 为测度 $\hat{\mu}$ 的扩张
2) (半环上可数可加测度生成环上扩张可数可加性) 若测度 m 可数可加，则测度 $\hat{\mu}$ 在环 $\mathfrak{R}(\mathfrak{S})$ 上扩张 μ 也可数可加 ♡

证明

1) 唯一性：由引理 (1.1) 有 $\mathfrak{R}(\mathfrak{S})=\left\{\bigcup\limits_{i=1}^{n}A_i, A_i\in\mathcal{S}\right\}$，设 μ 为 $\hat{\mu}$ 在 $\mathfrak{R}\mathfrak{S})$ 上扩张，则

$$(\forall B\in\mathfrak{R}(\mathfrak{S})):B=\bigcup_{i=1}^{n}A_i, A_i\in\mathcal{S}\Rightarrow\mu(B)=\sum_{i=1}^{n}\mu\left(A_i\right)=\sum_{i=1}^{n}\hat{\mu}\left(A_i\right)$$

则若测度 μ 存在扩张，该扩张唯一

存在性：设 $B\in\mathfrak{R}(\mathfrak{S})$，若 $B=\bigcup\limits_{i=1}^{n}A_i, A_i\in\mathfrak{S}, B=\bigcup\limits_{j=1}^{m}C_j, C_j\in\mathfrak{S}$，则有

$$B=\bigcup_{i=1}^{n}\bigcup_{j=1}^{m}\left(A_i\cap C_j\right), A_i=\bigcup_{j=1}^{m}\left(A_i\cap C_j\right), C_j=\bigcup_{i=1}^{n}\left(A_i\cap C_j\right)$$

$$\mu(B)=\sum_{i=1}^{n}\mu\left(A_i\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}\mu\left(A_i\cap C_j\right)=\sum_{j=1}^{m}\sum_{i=1}^{n}\mu\left(A_i\cap C_j\right)=\sum_{j=1}^{m}\mu\left(C_j\right)$$

则 B 上测度不依赖于半环上有限分解式的选择，则

$$(\forall B\in\mathfrak{R}(\mathfrak{S})):\left[\left(B=\bigcup_{i=1}^{n}A_i\right)\Rightarrow\left(\mu(B)=\sum_{i=1}^{n}\hat{\mu}\left(A_i\right)\right)\right]$$

显然有 $(\forall B\in\mathfrak{R}(\mathfrak{S}):\mu(B)\geqslant 0$，则 μ 为 $\mathfrak{R}(\mathfrak{S})$ 上测度

设 $C\in\mathfrak{R}(\mathfrak{S})$，且 $C=\bigcup\limits_{j=1}^{m}D_j, D_j\in\mathfrak{S}, C\cap B=\varnothing$ ，则有

$$B\cup C=\left(\bigcup_{i=1}^{n}A_i\right)\cup\left(\bigcup_{j=1}^{m}D_j\right)\in\mathfrak{R}(\mathfrak{S})$$

$$\mu(B\cup C)=\sum_{i=1}^{n}\hat{\mu}\left(A_i\right)+\sum_{j=1}^{m}\hat{\mu}\left(D_j\right)=\mu(B)+\mu(C)$$

2) 对于任意集 $A, A_i\in\mathfrak{R}(\mathfrak{S}):A=\bigcup\limits_{i=1}^{\infty}A_i$，设 $A=\bigcup\limits_{s=1}^{m}B_s, B_s\in\mathfrak{S}, A_i=\bigcup\limits_{j=1}^{m_i}B_{ij}, B_{ij}\in\mathfrak{S}$，则

$$B_s=\bigcup_{i=1}^{\infty}\left(A_i\cap B_s\right)=\bigcup_{i=1}^{\infty}\bigcup_{j=1}^{m_i}\left(B_{ij}\cap B_s\right), \hat{\mu}\left(B_s\right)=\sum_{i=1}^{\infty}\sum_{j=1}^{m_i}\hat{\mu}\left(B_{ij}\cap B_s\right), (\forall i,j):B_{ij}\cap A=B_{ij}$$

$$\mu(A)=\sum_{s=1}^{m}\hat{\mu}\left(B_s\right)=\sum_{s=1}^{m}\sum_{i=1}^{\infty}\sum_{j=1}^{m_i}\hat{\mu}\left(B_{ij}\cap B_s\right)=\sum_{i=1}^{\infty}\sum_{j=1}^{m_i}\sum_{s=1}^{m}\hat{\mu}\left(B_{ij}\cap B_s\right)=$$

$$=\sum_{i=1}^{\infty}\sum_{j=1}^{m_i}\hat{\mu}\left(B_{ij}\cap A\right)=\sum_{i=1}^{\infty}\sum_{j=1}^{m_i}\hat{\mu}\left(B_{ij}\right)=\sum_{i=1}^{\infty}\mu\left(A_i\right)$$

则扩张可数可加性得证

2.1.2 环上测度与连续测度
Measure on Semi-Ring and Continuous Measure

定义 2.5 (连续测度)

设在环 $\mathfrak{R}$ 上给定测度 μ，若

$$(\forall A, A_1, A_2,\ldots,A_n\ldots\in\mathcal{K}):\left[(\forall i):\left(A_i\subset A_{i+1}\right)\cap\left(A=\bigcup_{i=1}^{+\infty}A_i\right)\Rightarrow\left(\mu(A)=\lim_{i\to+\infty}\mu\left(A_i\right)\right)\right]$$

则称测度为连续测度 (continuous measure)

注 （符号约定）

今后约定若 $A_i\subset A_{i+1}, A=\bigcup\limits_{i=1}^{+\infty}A_i$，则 $A=\lim A_i$，这时有 $\mu\left(\lim A_i\right)=\lim\mu\left(A_i\right)$

定理 2.4 (环上测度连续性充要条件)

设 μ 为环 $\mathfrak{R}$ 上测度，则下列命题等价：

1) (连续性) μ 连续
2) (可数可加性) μ 可数可加
3) (集列形式 Heine 定理) 设集列 $\{B_i\}$ 满足 $B_i \supset B_{i+1}, B = \bigcap_{i=1}^{+\infty} B_i, \lim_{i\to\infty} B_i = B$ 且 $B, B_i \in \mathfrak{R}$，则有 $\mu(B) = \lim_{i\to\infty} \mu(B_i)$

证明

2)⇒1) 设 μ 为可数可加测度，观察集族 $A, A_i \in \mathfrak{R}$，其满足 $A_i \subset A_{i+1}, A = \bigcup_{i=1}^{+\infty} A_i$，则有 $A = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \cup \ldots$，则

$$\mu(A) = \mu(A_1) + \mu(A_2 \backslash A_1) + \ldots = \lim_{n\to\infty} (\mu(A_1) + \mu(A_2 \backslash A_1) + \ldots + \mu(A_n \backslash A_{n-1}))$$

再由

$$A_i = A_{i-1} \cup (A_i \backslash A_{i-1}) \Rightarrow \mu(A_i) = \mu(A_{i-1}) + \mu(A_i \backslash A_{i-1}) \Rightarrow \mu(A_i \backslash A_{i-1}) = \mu(A_i) - \mu(A_{i-1})$$

即得 $\mu(A) = \lim_{n\to\infty} \mu(A_n)$

1)⇒2) 设测度 μ 连续，观察集族 $A, A_i \in \mathfrak{R}, A = \bigcup_{i=1}^{\infty} A_i$ ，下记 $B_n = \bigcup_{i=1}^{n} A_i$ 且 $(\forall n \in \mathbb{N}^*) : B_n \subset B_{n+1}$，则 $\bigcup_{n=1}^{\infty} B_n = A$，由测度连续性即有

$$\mu(A) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

1)⇒ 3) 显然有 $\bigcup_{i=1}^{+\infty} (B_1 \backslash B_i) = B_1 \backslash \bigcap_{i=1}^{+\infty} B_i$ 且 $(B_1 \backslash B_i) \subset (B_1 \backslash B_{i+1})$，即有

$$\mu\left(B_1 \backslash \bigcap_{i=1}^{+\infty} B_i\right) = \lim_{i\to\infty} \mu(B_1 \backslash B_i)$$

则得

$$\mu(B_1) - \mu\left(\bigcap_{i=1}^{+\infty} B_i\right) = \lim_{i\to\infty} (\mu(B_1) - \mu(B_i)) \Rightarrow \lim_{i\to\infty} \mu(B_i) = \mu\left(\bigcap_{i=1}^{+\infty} B_i\right) = \mu(B)$$

3)⇒ 1) 设 $A = \bigcup_{i=1}^{\infty} A_i, A_i \subset A_{i+1}, A_i, A \in \mathfrak{R}$，且 $B_i = A \backslash A_i, B_i \supset B_{i+1}$

$$\mu\left(\bigcap_{i=1}^{+\infty} B_i\right) = \bigcap_{i=1}^{+\infty} (A \backslash A_i) = A \backslash \bigcup_{i=1}^{\infty} A_i = \varnothing$$

$$\lim_{i\to\infty} \mu(B_i) = \lim_{i\to\infty} (\mu(A) - \mu(A_i)) = 0 \Rightarrow \lim_{i\to\infty} \mu(A_i) = \mu(A)$$

上述命题等价性得证

定理 2.5 (环上测度保序性)

设 $\mathfrak{R}$ 为环，μ 为环 $\mathfrak{R}$ 上测度，则下列命题成立：

1) 设 $A, A_i \in \mathfrak{R}$，若 $\bigcup_{i=1}^{\infty} A_i \subset A$，则有 $\sum_{i=1}^{\infty} \mu(A_i) \leqslant \mu(A)$
2) 设 $A, A_i \in \mathfrak{R}$，若 $A \subset \bigcup_{i=1}^{\infty} A_i$ 且 μ 可数可加，则 $\mu(A) \leqslant \sum_{i=1}^{\infty} \mu(A_i)$

证明

1) 由条件有 $(\forall n \in \mathbb{N}^*) : \bigcup\limits_{i=1}^{n} A_i \subset A$ ，则有

$$\sum_{i=1}^{n} \mu(A_i) = \mu\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \mu(A)$$

令 $n \to \infty$，则有 $\sum\limits_{i=1}^{\infty} \mu(A_i) \leqslant \mu(A)$

2) 由条件有 $A = \bigcup\limits_{i=1}^{\infty} (A_i \cap A) = \bigcup\limits_{i=1}^{\infty} \tilde{A}_i$，其中 $\tilde{A}_i = A_i \cap A \in \Re$，则有表达式

$$A = \tilde{A}_1 \cup \left(\tilde{A}_2 \backslash \tilde{A}_1\right) \cup \left(\tilde{A}_3 \backslash \left(\tilde{A}_1 \cup \tilde{A}_2\right)\right) \cup \ldots$$

进而有

$$\begin{aligned}\mu(A) =& \mu\left(\tilde{A}_1\right) + \mu\left(\tilde{A}_2 \backslash \tilde{A}_1\right) + \mu\left(\tilde{A}_3 \backslash \left(\tilde{A}_1 \cup \tilde{A}_2\right)\right) + \ldots \\ \leqslant& \mu\left(\tilde{A}_1\right) + \mu\left(\tilde{A}_2\right) + \mu\left(\tilde{A}_3\right) + \ldots \leqslant \mu(A_1) + \mu(A_2) + \ldots\end{aligned}$$

则有 $\mu(A) \leqslant \sum\limits_{i=1}^{\infty} \mu(A_i)$

2.2 Lebesgue 测度论
Lebesgue Measure Theory

借助上测度和下测度的概念给出了 Lebesgue 可测与 Lebesgue 测度的定义，另外给出了上测度可数半加性 (2.7) 及其推论，然后给出了 Lebesgue 可测第一充要条件 (2.8) 以及其更好用的推论 Lebesgue 可测第二充要条件 (2.5)。然后给出了作为基本性质的 Lebesgue 可测集基本集运算的算术封闭性 (2.2.1)，Lebesgue 可测集可数并的 Lebesgue 可测性 (2.10) 与 Lebesgue 可测集可数交的 Lebesgue 可测性 (2.7)

借助抽象可测函数引入常用的可测函数的定义，指出结构可测函数体 (2.14)。最后指出相对于 Lebesgue 可测集的等价函数的概念。引入“几乎处处”概念，依次证明了逐点收敛可测函数序列极限函数可测性，并得其推论一致收敛可测函数序列极限函数可测性和几乎处处收敛可测函数序列极限函数可测性。另外给出了揭示可测函数序列一致收敛与几乎处处收敛间联系的 Egorov 定理 (2.17)。

另外，引入“沿测度收敛”的定义，并指出作为可测函数序列沿测度收敛必要条件的从几乎处处收敛到沿测度收敛的 Lebesgue 定理 (2.18) 与作为可测函数序列沿测度收敛（有限制的）充分条件的从沿测度收敛到几乎处处收敛的 Riesz 定理 (2.19)，并举例 (2.5) 验证可测函数序列沿测度收敛为其几乎处处收敛的必要不充分条件

2.2.1 Lebesgue 测度与 Lebesgue 可测
Lebesgue Measure and Lebesgue Measurable

借助上测度和下测度的概念给出了 Lebesgue 可测与 Lebesgue 测度的定义，另外给出了上测度可数半加性 (2.7) 及其推论，然后给出了 Lebesgue 可测第一充要条件 (2.8) 以及其更好用的推论 Lebesgue 可测第二充要条件 (2.5)

最后给出作为基本性质的 Lebesgue 可测集基本集运算的算术封闭性 (2.2.1)，Lebesgue 可测集可数并的 Lebesgue 可测性 (2.10) 与 Lebesgue 可测集可数交的 Lebesgue 可测性 (2.7)

定义 2.6 (Lebesgue 零测度)

设 $\Re$ 为环，该环上测度 μ 可数可加，若

$$(\forall \varepsilon>0)(\exists A_i \in \Re):\left[\left(E \subset \bigcup_{i=1}^{\infty} A_i\right) \Rightarrow\left(\sum_{i=1}^{+\infty} \mu\left(A_i\right)<\varepsilon\right)\right]$$

则称集合 E 有 Lebesgue 零测度 (Lebesgue zero measure/Лебега нуль)

♣

定理 2.6 (Lebesgue 零测度可数可加性)

设 $\Re$ 为集环，集 $A_i \in \Re$ 有 Lebesgue 零测度且 $A=\bigcup\limits_{i=1}^{\infty} A_i$，则集 A 也有 Lebesgue 零测度

♡

证明 由条件有

$$(\forall \varepsilon>0)(\exists B_{ij} \in \Re):\left[\left(A_i \subset \bigcup_{j=1}^{\infty} B_{ij}\right) \Rightarrow\left(\sum_{j=1}^{\infty} \mu\left(B_{ij}\right)<\frac{\varepsilon}{2^i}\right)\right]$$

则由 $A=\bigcup\limits_{i=1}^{\infty} A_i \subset \bigcup\limits_{i=1}^{\infty} \bigcup\limits_{j=1}^{\infty} B_{ij}$ 有 $\sum\limits_{i=1}^{\infty} \sum\limits_{j=1}^{\infty} \mu\left(B_{ij}\right)<\sum\limits_{i=1}^{\infty} \frac{\varepsilon}{2^i}=\varepsilon$，则集 A 有 Lebesgue 零测度

定义 2.7 (上测度与下测度)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，在集代数 $\Re$ 上定义了可数可加的测度 m。对于任意 $A \subset X$ 定义

$$\mu^*(A)=\inf \left\{\sum_{i=1}^{+\infty} m\left(B_i\right) \mid A \subset \bigcup_{i=1}^{\infty} B_i, B_i \in \Re\right\}$$

称 $\mu^*(A)$ 为上测度 (верхняя мера) 或外测度 (outer measure)，定义

$$\mu_*(A)=m(X)-\mu^*(X \backslash A)$$

称 $\mu_*(A)$ 为下测度 (нижняя мера) 或内测度 (interior measure)

另外，若

$$\mu^*(A)+\mu^*(X \backslash A)=m(X) \Leftrightarrow \mu^*(A)=\mu_*(A)$$

则称集 $A \subset X$ 为 Lebsgue 可测 (Lebesgue measurable/измеримое по Лебегу)，这时称 $\mu(A)=\mu^*(A)$ 为集合 A 的 Lebsgue 测度 (Lebesgue measure/мера Лебега)

♣

注 (上测度)

设 $A \in \Re$，则由上测度定义有 $\mu^*(A) \leqslant m(A)$。另一方面，设 $A \subset \bigcup\limits_{i=1}^{\infty} B_i, B_i \in \Re$，这时由测度 m 可数可加性有

$$m(A) \leqslant \sum_{i=1}^{+\infty} m\left(B_i\right) \Rightarrow m(A) \leqslant \mu^*(A)$$

则有 $(\forall A \in \Re): \mu^*(A)=m(A)$

定理 2.7 (上测度可数半加性/полуаддитивность)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，在集代数 $\Re$ 上定义了可数可加的测度 m，$A \subset \bigcup\limits_{i=1}^{+\infty} A_i$，其中 $A, A_i \subset X$，则

$$\mu^*(A) \leqslant \sum_{i=1}^{+\infty} \mu^*\left(A_i\right)$$

♡

证明 若 $A_i \in \Re$，则该断言从上测度的定义中直接得出。下证对于任意 $A_i \subset X$ 满足下确界的定义，这时有

$(\forall\varepsilon>0)(\exists B_{ik}\in\Re)$，当 $A_i\subset\bigcup_k B_{ik}$ 时满足

$$\mu^*(A_i)\leqslant\sum_k m(B_{ik})\leqslant\mu^*(A_i)+\frac{\varepsilon}{2^i}\Rightarrow\sum_{i,k}\mu^*(B_{ik})\leqslant\sum_i\mu^*(A_i)+\varepsilon$$

$$A\subset\bigcup_i A_i\subset\bigcup_{i,k}B_{ik}\Rightarrow\mu^*(A)\leqslant\sum_{i,k}m(B_{ik})=\sum_{i,k}\mu^*(B_{ik})\leqslant\sum_i\mu^*(A_i)+\varepsilon$$

定理即证

推论 2.1 (上下测度基本不等式)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，在集代数 $\Re$ 上定义了可数可加的测度 m，$A\subset X$，则 $0\leqslant\mu_*(A)\leqslant\mu^*(A)$

证明 显然有 $X\subset A\cup(X\backslash A)$，则由上测度可数半加性 (2.7) 即有 $m(X)\leqslant\mu^*(A)+\mu^*(X\backslash A)$，则得 $\mu_*(A)=m(X)-\mu^*(X\backslash A)\leqslant\mu^*(A)$

推论 2.2 (上测度绝对值不等式/上测度差的对称差估计)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，$A,B\subset X$，则有 $|\mu^*(A)-\mu^*(B)|\leqslant\mu^*(A\Delta B)$

证明 显然有 $A\subset B\cup(B\Delta A)$ 且 $B\subset A\cup(B\Delta A)$，则由上测度可数半加性 (2.7) 即有

$$\mu^*(A)\leqslant\mu^*(B)+\mu^*(B\Delta A),\quad\mu^*(B)\leqslant\mu^*(A)+\mu^*(B\Delta A)$$

推论即证

推论 2.3 (上测度三角不等式)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，$A,B\subset X$，则有 $\mu^*(A\Delta B)\leqslant\mu^*(A\Delta C)+\mu^*(C\Delta B)$

证明 显然有 $A\Delta B\subset(A\Delta C)\cup(C\Delta B)$，则由上测度可数半加性 (2.7) 即有 $\mu^*(A\Delta B)\leqslant\mu^*(A\Delta C)+\mu^*(C\Delta B)$

推论 2.4 (Lebesgue 零测度 Lebesgue 可测性)

Lebesgue 零测度集为 Lebesgue 可测集，即若 A 为 Lebesgue 零测度集，则有 $\mu_*(A)=\mu^*(A)=0$

定理 2.8 (Lebesgue 可测第一充要条件)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，在集代数 $\Re$ 上定义了可数可加的测度 m，$A\subset X$。则集 A 为 Lebesgue 可测集充要条件为

$$(\forall\varepsilon>0)(\exists B\in\Re):\mu^*(A\Delta B)<\varepsilon$$

证明 充分性：对于集 B 由上测度绝对值不等式 (2.2) 有 $|\mu^*(A)-\mu^*(B)|\leqslant\mu^*(A\Delta B)<\varepsilon$，进而有 $\mu^*(B)-\varepsilon\leqslant\mu^*(A)\leqslant\mu^*(B)+\varepsilon$，同理由 $(X\backslash A)\Delta(X\backslash B)=A\Delta B$ 即得 $\mu^*(X\backslash B)-\varepsilon\leqslant\mu^*(X\backslash A)\leqslant\mu^*(X\backslash B)+\varepsilon$

由 $B\in\Re$，则有 $\mu^*(B)+\mu^*(X\backslash B)=m(X)$，则

$$(\forall\varepsilon>0):m(X)-2\varepsilon<\mu^*(A)+\mu^*(X\backslash A)\leqslant m(X)+2\varepsilon\Rightarrow\mu^*(A)+\mu^*(X\backslash A)=m(X)$$

则 A 为 Lebesgue 可测集

必要性：设 A 为 Lebesgue 可测集，则 $(\forall\varepsilon>0)(\exists B_k\in\Re):A\subset\bigcup_{k=1}^{\infty}B_k$ 有

$$\mu^*(A)\leqslant\sum_{k=1}^{+\infty}m(B_k)\leqslant\mu^*(A)+\frac{\varepsilon}{3},\quad(\exists N\in\mathbb{N}):\sum_{k=N+1}^{+\infty}m(B_k)<\frac{\varepsilon}{3}$$

设 $B=\bigcup_{k=1}^{N}B_k$，下证 $\mu^*(A\Delta B)<\varepsilon$。显然有 $A\Delta B=(A\backslash B)\cup(B\backslash A)$，则有 $\mu^*(A\Delta B)\leqslant\mu^*(A\backslash B)+\mu^*(B\backslash A)$，

因此
$$A\backslash B\subset\bigcup_{k=N+1}^{\infty}B_k\Rightarrow\mu^*(A\backslash B)\leqslant\sum_{k=N+1}^{+\infty}m\left(B_k\right)<\frac{\varepsilon}{3}$$
由集合 $X\backslash A$ 可测，则有 $(\exists C_k\in\Re)$ 满足
$$\mu^*(X\backslash A)\leqslant\sum_{k=1}^{\infty}m\left(C_k\right)\leqslant\mu^*(X\backslash A)+\frac{\varepsilon}{3},\quad X\backslash A\subset\bigcup_{k=1}^{\infty}C_k$$
进而有
$$X=A\cup(X\backslash A)\subset\bigcup_k B_k\cup\bigcup_k\left(C_k\backslash B_k\right)\Rightarrow m(X)\leqslant\sum_{k=1}^{\infty}m\left(B_k\right)+\sum_{k=1}^{\infty}m\left(C_k\backslash B_k\right)$$
$$B\backslash A=B\cap(X\backslash A)\subset B\cap\bigcup_{k=1}^{\infty}C_k=\bigcup_{k=1}^{\infty}\left(C_k\cap B\right)\Rightarrow\mu^*(B\backslash A)\leqslant\sum_{k=1}^{\infty}m\left(C_k\cap B\right)$$
$$m\left(C_k\cap B\right)=m\left(C_k\right)-m\left(C_k\backslash B\right)\Rightarrow\mu^*(B\backslash A)\leqslant\sum_{k=1}^{\infty}m\left(C_k\right)-\sum_{k=1}^{\infty}m\left(C_k\backslash B\right)$$
则有
$$\begin{aligned}\sum_{k=1}^{\infty}m\left(B_k\right)\leqslant\mu^*(A)+\frac{\varepsilon}{3},\quad\sum_{k=1}^{\infty}m\left(C_k\right)\leqslant\mu^*(X\backslash A)+\frac{\varepsilon}{3}\\\Rightarrow\sum_{k=1}^{\infty}m\left(B_k\right)+\sum_{k=1}^{\infty}m\left(C_k\right)\leqslant m(X)+\frac{2\varepsilon}{3}\\\Rightarrow\sum_{k=1}^{\infty}m\left(B_k\right)+\sum_{k=1}^{\infty}m\left(C_k\right)\leqslant\sum_{k=1}^{\infty}m\left(B_k\right)+\sum_{k=1}^{\infty}m\left(C_k\backslash B_k\right)+\frac{2\varepsilon}{3}\\\Rightarrow\sum_{k=1}^{\infty}m\left(C_k\right)\leqslant\sum_{k=1}^{\infty}m\left(C_k\backslash B_k\right)+\frac{2\varepsilon}{3}\Rightarrow\mu^*(B\backslash A)\leqslant\frac{2\varepsilon}{3}\Rightarrow\mu^*(A\Delta B)<\varepsilon\end{aligned}$$
综上，定理得证

推论 2.5 (Lebesgue 可测第二充要条件)

设 X 为集合，$A\subset X$，则集合 A 为 Lebesgue 可测的充要条件为 $(\forall\varepsilon>0)$ 存在 Lebesgue 可测集 B 满足 $\mu^*(A\Delta B)<\varepsilon$ ♡

证明 设集环 $\Re$ 为 X 的子集的集代数，若 B 存在则有对于 $(\forall B_1\in\Re)$ 由上测度三角不等式 (2.3) 有
$$\mu^*\left(B\Delta B_1\right)<\varepsilon\Rightarrow\mu^*\left(A\Delta B_1\right)\leqslant\mu^*(A\Delta B)+\mu^*\left(B\Delta B_1\right)<2\varepsilon$$
结合定理 (2.8)，该推论即证

性质 (Lebesgue 可测集集运算封闭性)

设 X 为集合，集环 $\Re$ 为 X 的子集的集代数，$A,B\subset X$，集合 A,B 为 Lebesgue 可测集，则下列命题成立：

1)$A\cup B$ 为 Lebesgue 可测集

2)$A\cap B$ 为 Lebesgue 可测集

3)$A\backslash B$ 为 Lebesgue 可测集

4)$A\Delta B$ 为 Lebesgue 可测集

证明 1) 若 A,B 集合 A,B 为 Lebesgue 可测的，则有 $(\exists A_1\in\Re):\mu^*\left(A\Delta A_1\right)<\varepsilon$ 且 $(\exists B_1\in\Re):\mu^*\left(B\Delta B_1\right)<\varepsilon$，则有
$$(A\cup B)\Delta\left(A_1\cup B_1\right)\subset\left(A\Delta A_1\right)\cup\left(B\Delta B_1\right)\Rightarrow\mu^*\left((A\cup B)\Delta\left(A_1\cup B_1\right)\right)<2\varepsilon$$
由此 $A\cup B$ 为 Lebesgue 可测的

2) 若集 A 为 Lebesgue 可测的，则由 Lebesgue 可测定义即得 $X\backslash A$ 也为 Lebesgue 可测集，因此 $A\cap B=X\backslash((X\backslash A)\cup(X\backslash B))$ 表明 $A\cap B$ 为 Lebesgue 可测集

3) 由 $A \backslash B = A \cap (X \backslash B)$ 即得 $A \backslash B$ 为 Lebesgue 可测集

4) 由 $A \Delta B = (A \backslash B) \cup (B \backslash A)$ 即得 $A \Delta B$ 为 Lebesgue 可测集

定理 2.9 (Lebesgue 测度有限可加性)

设 X 为集合，集环 $\mathfrak{R}$ 为 X 的子集的集代数，$A \subset X$，$A=\bigcup\limits_{k=1}^{n} A_k$，其中 A, A_k 为 Lebesgue 可测集，则

$$\mu^*(A)=\sum_{k=1}^{n} \mu^*\left(A_k\right)$$

♡

证明 当 $n=2$ 时，A_1, A_2, A 为 Lebesgue 可测集，则 $A=A_1 \cup A_2$，则

$$(\exists B_i \in \mathfrak{R}): \mu^*\left(A_i \Delta B_i\right)<\varepsilon, i=1,2$$

由定理 (2.7) 有 $\mu^*(A) \leqslant \mu^*\left(A_1\right)+\mu^*\left(A_2\right)$。下证反向不等式，设 $B=B_1 \cup B_2$，则有

$$\begin{aligned}
\left|\mu^*(A)-\mu^*(B)\right| & \leqslant \mu^*(A \Delta B)=\mu^*\left(\left(A_1 \sqcup A_2\right) \Delta\left(B_1 \cup B_2\right)\right) \\
& \leqslant \mu^*\left(A_1 \Delta B_1\right)+\mu^*\left(A_2 \Delta B_2\right)<2 \varepsilon
\end{aligned}$$

$$\Rightarrow \mu^*(A) \geqslant \mu^*(B)-2 \varepsilon=m\left(B_1 \cup B_2\right)-2 \varepsilon=m\left(B_1\right)+m\left(B_2\right)-m\left(B_1 \cap B_2\right)-2 \varepsilon$$

另外有

$$\begin{gathered}
B_1 \cap B_2 \subset\left(A_1 \Delta B_1\right) \cup\left(A_2 \Delta B_2\right) \Rightarrow \mu^*\left(B_1 \cap B_2\right)=m\left(B_1 \cap B_2\right)<2 \varepsilon \\
\Rightarrow \mu^*(A) \geqslant \mu^*\left(B_1\right)+\mu^*\left(B_2\right)-4 \varepsilon
\end{gathered}$$

则有

$$\mu^*\left(B_i\right) \geqslant \mu^*\left(A_i\right)-\varepsilon \Rightarrow \mu^*(A) \geqslant \mu^*\left(A_1\right)+\mu^*\left(A_2\right)-6 \varepsilon$$

由 $\varepsilon>0$ 任意性则有 $\mu^*(A) \geqslant \mu^*\left(A_1\right)+\mu^*\left(A_2\right)$

推论 2.6 (Lebesgue 测度可数可加性)

设 X 为集合，集环 $\mathfrak{R}$ 为 X 的子集的集代数，$A \subset X$，$A=\bigcup\limits_{k=1}^{+\infty} A_k$，其中 A, A_k 为 Lebesgue 可测集，则

$$\mu^*(A)=\sum_{k=1}^{\infty} \mu^*\left(A_k\right)$$

♡

定理 2.10 (Lebesgue 可测集可数并 Lebesgue 可测性)

Lebesgue 可测集的可数并集也为 Lebesgue 可测集

♡

证明 设 A_i 为 Lebesgue 可测集，$A=\bigcup\limits_{i=1}^{\infty} A_i$，则有

$$A=\bigcup_{k=1}^{\infty} A_k^{\prime}, \quad A_1^{\prime}=A_1, A_2^{\prime}=A_2 \backslash A_1, \ldots, A_k^{\prime}=A_k \backslash\left(\bigcup_{i=1}^{k-1} A_i\right)$$

对于任意正整数 n 有 $\bigcup\limits_{i=1}^{n} A_i^{\prime} \subset A$，则有 $\sum\limits_{i=1}^{n} \mu\left(A_i^{\prime}\right) \leqslant \mu^*(A)$，进而 $\sum\limits_{i=1}^{\infty} \mu\left(A_i^{\prime}\right)$ 收敛，则有

$$\begin{gathered}
(\forall \varepsilon>0)(\exists N \in \mathbb{N}): \sum_{i=N+1}^{\infty} \mu\left(A_i^{\prime}\right)<\varepsilon, \quad A \backslash \bigcup_{k=1}^{N} A_k^{\prime} \subset \bigcup_{k=N+1}^{\infty} A_k^{\prime} \\
\Rightarrow \mu^*\left(A \backslash \bigcup_{k=1}^{N} A_k^{\prime}\right)<\varepsilon, \quad \bigcup_{k=1}^{N} A_k^{\prime} \subset A \Rightarrow \bigcup_{k=1}^{N} A_k^{\prime} \backslash A=\varnothing
\end{gathered}$$

即有

$$A\backslash \bigcup_{k=1}^{N} A_k' = A\Delta \bigcup_{k=1}^{N} A_k' \Rightarrow \mu^*\left(A\Delta \bigcup_{k=1}^{N} A_k'\right) < \varepsilon$$

则 A 为 Lebesgue 可测集

推论 2.7 (Lebesgue 可测集可数交 Lebesgue 可测性)

若集 A_i 为 Lebesgue 可测集，则 $A = \bigcap_{i=1}^{\infty} A_i$ 也为 Lebesgue 可测集

♡

证明 应用对偶原理 (принцип двойственности)，由 $A = X\backslash \bigcup_{i=1}^{\infty}(X\backslash A_i)$ 即得

注 现考虑集 X 不存在有限测度的情况。则 $\mathfrak{R}$ 为环，但不是集代数

称测度 μ 为 σ 有限的，若

$$\exists X_k : X = \bigsqcup_{k=1}^{\infty} X_k, \mu\left(X_k\right) < \infty$$

例题 2.2 $X = \mathbb{R} = \bigcup_{k=-\infty}^{+\infty}[k;k+1), \mu([a;b)) = b-a$

定义 2.8 (Lebesgue 不可测集)

称集 $A \subset X$ 为可测的，若 $\forall k$ 集 $A\cap X_k$ 为 Lebesgue 可测的，设 $A_k = A\cap X_k$，这时有

$$A = \bigcup_{k=1}^{+\infty} A_k, \quad \mu(A) = \sum_{k=1}^{\infty}\mu\left(A_k\right)$$

♣

例题 2.3 （Lebegue 不可测集/Пример Витали）

设 $I = [0;1)$，定义等价关系 $x\sim y$，其中 $x-y\in\mathbb{Q}\cap[-1;1)$，另外任意 $x\in I$ 都会属于某个等价类

设 M 为恰包含每个等价类中的一个点的集合，考虑

$$A = \bigcup_{k=1}^{\infty}\left\{\left(M+r_k\right)|\left(r_k\in\mathbb{Q}\cap[-1;1)\right)\right\} \Rightarrow [0;1)\subset A\subset[-1;2)$$

由 $(\forall x\in[0;1))(\exists y\in M): x-y=r_k \Rightarrow x=y+r_k\in M+r_k$ ，则集 $M+r_k, k=1,2,\ldots$ 不交（反之设 $x\in\left(M+r_k\right)\cap\left(M+r_m\right)$，则有 $x=y_1+r_k=y_2+r_m, r_m\neq r_k, \Rightarrow y_1-y_2=r_m-r_k\in\mathbb{Q}\Rightarrow y_1\sim y_2$，这与 M 仅仅包含等价类中一个点矛盾）。反证：假设 M 为 Lebesgue 可测集，则 $M+r_k$ 为 Lebesgue 可测集，$\mu(M)=\mu\left(M+r_k\right)$，下分类讨论

1) 设 $\mu(M)=0$，则有 $1=\mu([0,1))\leqslant\sum_{n=1}^{\infty}\mu\left(M+r_n\right)=\sum_{n=1}^{\infty}\mu(M)=0$ ，导出矛盾

2) 设 $\mu(M)>0$，则 $3=\mu([-1,2))\geqslant\sum_{n=1}^{\infty}\mu\left(M+r_n\right)=\sum_{n=1}^{\infty}\mu(M)=+\infty$ ，导出矛盾

则集 M 为 Lebesgue 不可测集

2.2.2 可测函数与可测性
Measurable Function and Measurability

借助抽象可测函数引入常用的可测函数的定义，指出结构可测函数体 (2.14)。最后指出相对于 Lebesgue 可测集的等价函数的概念

定义 2.9 (抽象可测函数)

设 X 和 Y 为二任意集，$\mathfrak{S}_X$ 和 $\mathfrak{S}_Y$ 为对应子集族，另设抽象函数 $f: X\to Y, x\mapsto y=f(x)$，若

$$\left(A\in\mathfrak{S}_Y\right)\Rightarrow f^{-1}(A)\in\mathfrak{S}_X$$

则称 f 为 $(\mathfrak{S}_X,\mathfrak{S}_Y)$ 可测函数 (measurable function)

定理 2.11 (抽象可测函数复合可测性) 设 X,Y,Z 为三个任意集，$\mathfrak{S}_X,\mathfrak{S}_Y,\mathfrak{S}_Z$ 为对应子集族，又设函数 f 为 $(\mathfrak{S}_X,\mathfrak{S}_Y)$ 可测函数，函数 g 为 $(\mathfrak{S}_Y,\mathfrak{S}_Z)$ 可测函数，则函数 $z=\varphi(x)\equiv g(f(x))$ 为 $(\mathfrak{S}_X,\mathfrak{S}_Z)$ 可测函数

证明 若 $A\in\mathfrak{S}_Z$，则由函数 g 的 $(\mathfrak{S}_Y,\mathfrak{S}_Z)$ 可测性有 $g^{-1}(A)=B\in\mathfrak{S}_Y$。又由 f 的 $(\mathfrak{S}_X,\mathfrak{S}_Y)$ 可测性，集 $f^{-1}(B)\in\mathfrak{S}_X$，即 $f^{-1}\left(g^{-1}(A)\right)=\varphi^{-1}(A)\in\mathfrak{S}_X$，即函数 φ 为 $(\mathfrak{S}_X,\mathfrak{S}_Z)$ 可测函数

定义 2.10 (可测函数第一定义) 设 X 为一个在其上给定了一个在 σ 代数 $\mathfrak{S}_\mu$ 上有定义的 σ 可加测度 μ 的集，函数 $f(x)$ 为实函数，若对于实直线上任意 Borel 集 A 都有

$$f^{-1}(A)\in\mathfrak{S}_\mu$$

则称实函数 $f(x)$ 在 X 上为 μ-可测函数 (μ-measurable function)，不引起歧义时称可测函数 (measurable function)

特别地，若实直线上任意 Borel 集 A 的原像都为 Borel 集，则称 $f(x)$ 为 Borel 可测函数 (Borel measurable function)

注 （可测函数复变情形）

类似地，只需将实直线改为复平面，若实部和虚部都可测，则称复函数可测

定理 2.12 (实变函数可测性充要条件) 实变函数 $f(x)$ 为可测函数的充要条件为对于任意实数 c，集 $\{x:f(x)<c\}$ 为 Lebesgue 可测集

证明 必要性显然，因为 $(-\infty;c)$ 为一个 Borel 集。充分性：注意到，由一切射线 $(-\infty;c)$ 的集族 Σ 所生成的 σ 代数与直线上的一切 Borel 集的 σ 代数完全相同。但每一个 Borel 集的原像属于由那些属于 Σ 的半直线的原像所生成的 σ 代数，即实函数 f 为可测函数

注 由该定理导出更为常用的可测函数的第二定义

定义 2.11 (可测函数第二定义) 设 $f:X\to\mathbb{R}$，若 $(\forall c\in\mathbb{R})$ 都满足集 $M_c=\{x\in\mathbb{R}:f(x)<c\}$ 为 Lebesgue 可测集，则称实变函数 $f(x)$ 为可测函数 (measurable function)

定理 2.13 (可测函数必要条件) 设 $f(x)$ 为可测函数，则对于任意 $c\in\mathbb{R}$ 有集合

$$\bar{M}_c=\{x\in\mathbb{R}|f(x)\leqslant c\}$$

为 Lebesgue 可测集

注 结论也可改为

$$\{x\in\mathbb{R}|f(x)\leqslant c\}=\bigcap_{k=1}^{\infty}\left\{x\in\mathbb{R}|f(x)<c+\frac{1}{k}\right\}$$

例题 2.4 （可测函数/Dirichlet 函数）

Dirichlet 函数

$$D(x)=\begin{cases}1, & x\in\mathbb{Q}\\ 0, & x\notin\mathbb{Q}\end{cases}$$

为可测函数

解 当 $c\leqslant 0$ 时有 $M_c=\varnothing$；当 $c\in(0;1]$，则有 $M_c=\mathbb{R}\backslash\mathbb{Q}$；当 $c>1$，则有 $M_c=\mathbb{R}$。上述集合均为 Lebesgue 可测集，则 $D(x)$ 为可测函数

注 类似地，设 A 为 Lebesgue 可测集，则指标函数 χ_A：

$$\chi_A(x)=\begin{cases}1, & x\in A\\ 0, & x\notin A\end{cases}$$

为可测函数

定理 2.14 (可测函数体)

全体可测函数关于函数的和与乘法运算构成体 ♡

证明

1) 若 f 可测，则对于任意常数 k 和 a，函数 kf 和 $a+f$ 显然可测

2) 若 f 和 g 可测，则集 $\{x:f(x)>g(x)\}$ 可测，因为

$$\{x:f(x)>g(x)\}=\bigcup_{k=1}^{\infty}\{x:f(x)>r_k\}\cap\{x:g(x)<r_k\}$$

其中求和为对按任意次序编号的一切有理数 r_k 遍历。由此得

$$\{x:f(x)>a-g(x)\}=\{x:f(x)+g(x)>a\}$$

可测，即可测函数之和可测

3) 由 1) 和 2) 即得差 $f-g$ 的可测性

4) 由 1)2)3) 和抽象可测函数复合可测性 (2.11) 即有恒等式

$$fg=\frac{1}{4}\left[(f+g)^2-(f-g)^2\right]$$

右端为可测函数

5) 设 $g(x)\neq 0$，若 $c>0$，则

$$\left\{x:\frac{1}{f(x)}<c\right\}=\left\{x:f(x)>\frac{1}{c}\right\}\cup\{x:f(x)<0\}$$

而若 $c<0$，则

$$\left\{x:\frac{1}{f(x)}<c\right\}=\left\{x:0>f(x)>\frac{1}{c}\right\}$$

而若 $c=0$，则

$$\left\{x:\frac{1}{f(x)}<c\right\}=\{x:f(x)<c\}$$

右端均为 Lebesgue 可测集，则若 $f(x)\neq 0$ 可测，$\dfrac{1}{f(x)}$ 也可测。由 4) 和 5) 即得商 $\dfrac{f(x)}{g(x)}$ 的可测性（在 $g(x)\neq 0$ 的条件下）

定义 2.12 (完备测度)

设 μ 为测度，若

$$(\forall A,B):\{[(\mu(A)=0)\wedge(B\subset A)]\Rightarrow(\mu(B)=0)\}$$

则称 μ 为完备测度 (complete measure/полная мера) ♣

注 （完备测度）

以下可测函数部分均假设测度为完备测度

定义 2.13 (等价函数)

若两个定义在同一 Lebesgue 可测集 E 上的函数 f 与 g 满足条件

$$\mu\{x : f(x) \neq g(x)\} = 0$$

则称函数 f 和 g 等价 (equivalent)

♣

注 （等价性）

在古典分析里函数的等价性概念不起重要作用，因为在古典分析里主要讨论一元或多元连续函数，而对这些函数等价性就相当于恒等性。确切地说，若两个在某一闭区间 E 上连续的函数 f 和 g 等价（Lebesgue 测度意义下），则它们恒等。事实上，若在任一点 x_0 上 $f(x_0) \neq g(x_0)$，则由 f 和 g 的连续性可找到点 x_0 的一个邻域，在该邻域内的一切点上有 $f(x) \neq g(x)$。这样的邻域的测度为正，因此连续函数不可能等价，除非它们恒等。

而对于任意可测函数，两个函数的等价性绝不表示两个函数恒等。例如，在实直线的有理点等于 1 而在无理点等于零的 Dirichlet 函数就与恒等于零的函数等价

定理 2.15 (等价函数可测性传递性)

若定义在 Lebesgue 可测集 E 上的函数 $f(x)$ 与在 E 上的某一可测函数 $g(x)$ 等价，则函数 $f(x)$ 也为可测函数

♡

证明 由等价性定义推出集 $\{x : f(x) < a\}$ 与 $\{x : g(x) < a\}$ 彼此仅可能相差某一测度为零的集，则若它们中的第二个集可测，则第一个集也为可测的

2.2.3 可测函数序列逐点收敛，一致收敛，几乎处处收敛与沿测度收敛

Sequence of Measurable Functions Point-by-Point Convergence,Uniform Convergence,Convergence Almost Everywhere and Convergence in Measure

本节依次证明了逐点收敛可测函数序列极限函数可测性，并得其推论一致收敛可测函数序列极限函数可测性和几乎处处收敛可测函数序列极限函数可测性。另外给出了揭示可测函数序列一致收敛与几乎处处收敛间联系的 Egorov 定理 (2.17)

最后指出作为可测函数序列沿测度收敛必要条件的从几乎处处收敛到沿测度收敛的 Lebesgue 定理 (2.18) 与作为可测函数序列沿测度收敛（有限制的）充分条件的从沿测度收敛到几乎处处收敛的 Riesz 定理 (2.19)，并举例 (2.5) 验证可测函数序列沿测度收敛为其几乎处处收敛的必要不充分条件

定理 2.16 (逐点收敛可测函数序列极限函数可测性)

设可测函数序列 $f_n(x)$ 逐点 (по-точечно) 收敛到 $f(x)$，则极限函数 $f(x)$ 为可测函数

♡

证明 下考虑

$$M_c = \{x \in \mathbb{R} : f(x) < c\} = \bigcup_{k=1}^{+\infty} \bigcup_{n=1}^{+\infty} \bigcap_{m=n+1}^{+\infty} \left\{x \in \mathbb{R} : f_m(x) < c - \frac{1}{k}\right\}$$

设 $x \in M_c$，则对于充分大的 k 有 $f(x) < c - \dfrac{1}{k}$。因为当 $m \to \infty$ 时 $f_m(x) \to f(x)$，则

$$(\exists k \in \mathbb{N})(\exists n \in \mathbb{N})(\forall m > n) : f_m(x) < c - \frac{1}{k} \Rightarrow x \in \bigcup_{k=1}^{+\infty} \bigcup_{n=1}^{+\infty} \bigcap_{m=n+1}^{+\infty} \left\{x \in \mathbb{R} : f_m(x) < c - \frac{1}{k}\right\}$$

下设

$$x \in \bigcup_{k=1}^{+\infty} \bigcup_{n=1}^{+\infty} \bigcap_{m=n+1}^{+\infty} \left\{x \in \mathbb{R} : f_m(x) < c - \frac{1}{k}\right\}$$

则有

$$(\exists k \in \mathbb{N})(\exists n \in \mathbb{N})(\forall m > n) : f_m(x) < c - \frac{1}{k} \Rightarrow f(x) \leqslant c - \frac{1}{k} \Rightarrow f(x) < c \Rightarrow x \in M_c$$

则集合 $\left\{x \in \mathbb{R} : f_m(x) < c - \frac{1}{k}\right\}$ 中每一个均可测，因此 M_c 也可测，则 $f(x)$ 为可测函数

推论 2.8 (一致收敛可测函数序列极限函数可测性)

设可测函数序列 $f_n(x) \rightrightarrows f(x)$，则极限函数 $f(x)$ 为可测函数

定义 2.14 (函数序列几乎处处收敛)

设函数序列 $f_n(x)$ 定义在某一具有测度 μ 的集 X 上，若

$$\mu\left\{x \in X : \nexists \lim_{n\to+\infty} f_n(x) = f(x)\right\} = 0$$

则称函数序列几乎处处收敛 (convergence almost everywhere/почти всюду сходящаяся) 到函数 $f(x)$，记为 $f_n(x) \xrightarrow{\text{п.в}} f(x)$

推论 2.9 (几乎处处收敛可测函数序列极限函数可测性)

设可测函数序列 $f_n(x) \xrightarrow{\text{п.в.}} f(x)$，则极限函数 $f(x)$ 为可测函数

证明 设 $(\forall x \in X_0) : f_n(x) \to f(x)$ 且 $\mu(X \backslash X_0) = 0$，则令

$$M_c = \{x \in \mathbb{R} : f(x) < c\} = (\{x : f(x) < c\} \cap X_0) \cup (\{x : f(x) < c\} \cap (X \backslash X_0)) = X_1 \cup X_2$$

由逐点收敛可测函数序列极限函数可测性 (2.16) 有 X_1 为 Lebesgue 可测集，而 X_2 为 Lebesgue 零测度集，则 X_2 也可测，因此 M_c 为 Lebesgue 可测集，且极限函数 $f(x)$ 在 X 上可测

定理 2.17 (Egorov 定理)

(Egorov 定理/теорема Егоров[a]) 设可测函数序列 $f_n(x)$ 在一个具有有限测度的集 E 上几乎处处收敛到 $f(x)$，则对于任意 $\delta > 0$ 总存在 Lebesgue 可测集 $E_\delta \subset E$ 满足

1) $\mu(E_\delta) > \mu(E) - \delta$
2) 在集 E_δ 上序列 $f_n(x)$ 一致收敛到 $f(x)$

[a]德米特里 · 叶戈罗夫 (Дми́трий Фёдорович Егóров,1869—1931) 俄罗斯及苏联数学家，主要贡献在微分几何、数学分析等领域。1911 年，叶戈罗夫发表了 Егоров 定理。1921 年，当选莫斯科数学学会会长。1923 年，成为莫斯科大学数学与力学学院院长

证明 由几乎处处收敛可测函数序列极限函数可测性 (2.9) 有，函数 $f(x)$ 为可测函数。下设

$$E_n^m = \bigcap_{i \geqslant n}\left\{x : |f_i(x) - f(x)| < \frac{1}{m}\right\}$$

又设 $E^m = \bigcup\limits_{n=1}^{\infty} E_n^m$，由集 E_n^m 的定义，对于固定的 m 显然有

$$E_1^m \subset E_2^m \subset \cdots \subset E_n^m \subset \cdots$$

则由测度连续性充要条件 (2.4)，根据 σ 可加测度的连续性有

$$(\forall m)(\forall \delta > 0)(\exists n_0(m) : \mu\left(E^m \backslash E_{n_0(m)}^m\right) < \frac{\delta}{2^m}$$

令 $E_\delta = \bigcap\limits_{m=1}^{\infty} E_{n_0(m)}^m$，下证，这样构造的 E_δ 满足定理的要求

首先证明在 E_δ 上序列 $\{f_i(x)\}$ 一致收敛于函数 $f(x)$。若 $x \in E_\delta$，则有

$$(\forall m)(\forall i \geqslant n_0(m)) : |f_i(x) - f(x)| < \frac{1}{m}$$

现估计集 $E\backslash E_\delta$ 的测度. 注意到，$(\forall m):\mu\left(E\backslash E^m\right)=0$。若 $x_0\in E\backslash E^m$，则有充分大的 i 存在，满足

$$\left|f_i\left(x_0\right)-f\left(x_0\right)\right|\geqslant\frac{1}{m}$$

即序列 $\{f_n(x)\}$ 在点 x_0 不收敛于 $f(x)$。由条件，$\{f_n(x)\}$ 几乎处处收敛于 $f(x)$，则 $\mu\left(E\backslash E^m\right)=0$ ，由此推出

$$\mu\left(E\backslash E^m_{n_0(m)}\right)=\mu\left(E^m\backslash E^m_{n_0(m)}\right)<\frac{\delta}{2^m}$$

因此

$$\mu\left(E\backslash E_s\right)=\mu\left(E\backslash\bigcap_{m=1}^{\infty}E^m_{n_0(m)}\right)=\mu\left(\bigcap_{m=1}^{\infty}\left(E\backslash E^m_{n_0(m)}\right)\right)\leqslant\sum_{m=1}^{\infty}\mu\left(E\backslash E^m_{n_0(m)}\right)<\sum_{m=1}^{\infty}\frac{\delta}{2^m}=\delta$$

定理得证

注 该定理揭示了几乎处处收敛与一致收敛间的关系

定理 2.18 (Lebesgue 定理/теорема Лебега)

若可测函数序列 $\{f_n(x)\}_{n=1}^{\infty}$ 在集 X 上几乎处处收敛到 $f(x)$，则有 $f_n(x)$ 沿测度收敛到 $f(x)$

♡

证明 仅需考虑逐点收敛的情形，设 $(\forall x\in X):f_n(x)\to f(x)$，则有对于 $(\forall\varepsilon>0)$

$$E_n=\{x\in X:|f_n(x)-f(x)|\geqslant\varepsilon\},R_n=\bigcup_{k=n}^{\infty}E_k,R_1\supset R_2\supset R_3\supset\ldots$$

则有 $S=\bigcap\limits_{n=1}^{\infty}R_n\Rightarrow\mu(S)=\lim\limits_{n\to\infty}\mu\left(R_n\right)$ ，下证 $S=\varnothing$

反证：设 $(\exists x_0\in S=\cap R_n)$，则有

$$(\forall k_m)(\exists\{E_{k_m}\}):x_0\in E_{k_m}\Rightarrow(\forall k_m):\left|f_{k_m}\left(x_0\right)-f\left(x_0\right)\right|\geqslant\varepsilon$$

与收敛条件矛盾。

若 $S=\varnothing$，则 $\mu(S)=0$ 且 $\lim\limits_{n\to\infty}\mu\left(R_n\right)=0$，则由 $E_n\subset R_n$ 有

$$\mu\left(E_n\right)\leqslant\mu\left(R_n\right)\underset{n\to\infty}{\longrightarrow}0\Rightarrow\mu\left(E_n\right)\underset{n\to\infty}{\longrightarrow}0\Rightarrow f_n(x)\xrightarrow{\mu}f(x)$$

由此定理即证

定理 2.19 (Riesz 定理/теорема Рисса)

(Riesz 定理/теорема Рисса[a]) 若可测函数序列 $\{f_n(x)\}$ 沿测度收敛到函数 $f(x)$，则存在子序列 $\{f_{n_k}(x)\}$ 在集 X 上几乎处处收敛到 $f(x)$

[a]弗里杰什 · 里斯 (匈牙利语：Riesz Frigyes,1880.1.22-1956.2.28) 匈牙利数学家，出生于奥地利（今匈牙利杰尔），卒于匈牙利布达佩斯，是匈牙利与瑞典数学家 Riesz Marcel 的哥哥。Riesz Frigyes 为现代泛函分析的主要创始人之一，其主要工作已与波兰数学家 Banach 合并，最出名的工作有 Riesz 表示定理，Riesz- Fischer 定理和 Riesz 空间

♡

证明 由沿测度收敛性，可以构造集列

$$E_k=\left\{x\in X:\left|f_{n_k}(x)-f(x)\right|\geqslant\frac{1}{k}\right\}$$

并满足 $\mu\left(E_k\right)\leqslant\frac{1}{2^k}$，下证 $f_{n_k}(x)$ 几乎处处收敛到 $f(x)$

设 $R_n=\bigcup\limits_{k=n}^{\infty}E_k,R_1\supset R_2\supset R_3\supset\ldots,\lim\limits_{n\to\infty}R_n=\bigcap\limits_{n=1}^{\infty}R_n=S$，则有

$$\mu\left(R_n\right)\leqslant\sum_{k=n}^{\infty}\mu\left(E_k\right)\leqslant\sum_{k=n}^{\infty}\frac{1}{2^k}=\frac{1}{2^{n-1}}\underset{n\to\infty}{\longrightarrow}0$$

则由测度连续性充要条件 (2.4) 有 $\mu(S)=\lim\limits_{n\to\infty}\mu\left(R_n\right)=0$

注意到 $(\forall x_0\notin S)(\exists m\in\mathbb{N}^*):x_0\notin R_m=\bigcup\limits_{k=m}^{\infty}E_k$，则有 $(\forall k\geqslant m):x_0\notin E_k$。这时由集列 E_k 的定义有

$$(\forall k\geqslant m):\left|f_{n_k}\left(x_0\right)-f\left(x_0\right)\right|<\frac{1}{k}$$

即有 $f_{n_k}(x_0) \underset{n\to\infty}{\longrightarrow} f(x_0)$，由 $x_0 \notin S$ 任意性即证 $\{f_{n_k}(x)\}$ 在 X 上几乎处处收敛到 $f(x)$

例题 2.5 （可测函数序列沿测度收敛但非几乎处处收敛例子）

从可测函数列按测度收敛一般推不出其为几乎处处收敛的，对于任意 $k \in \mathbb{N}$，则在半开区间 $(0;1]$ 上定义 k 个函数 $f_1^{(k)}, f_2^{(k)}, \cdots, f_k^{(k)}$，其中

$$f_i^{(k)}(x) = \begin{cases} 1, & \text{当}\dfrac{i-1}{k} < x \leqslant \dfrac{i}{k} \text{ 时} \\ 0, & \text{当}x \text{ 为其余值时} \end{cases}$$

可以验证其按测度收敛到 0，但同时在每一点都不收敛

注 该例表明可测函数序列沿测度收敛为其几乎处处收敛的必要不充分条件

第 3 章 组合学 (Combinatorics)

组合数学 (combinatorics) 又称组合学，近几十年来发展迅速的一门数学分支。它的发展得益于计算机科学的推动，也得益于与其他数学分支越来越密切的联系。另一方面，组合数学在工程技术及其他自然科学学科和社会科学中也有广泛的应用。

组合数学是研究离散结构性质的一门学科，其研究对象包括排列、整数分拆、集合划分、偏序集、图、拟阵、区组设计、结合方案、编码、凸多胞形等。组合数学虽然有很多分支，但是一些不同分支之间又互相交叉，很难给出一个明确的分类。根据研究的内容，组合数学可分为计数组合学 (enumeratlve combinatorics)、几何组合学、极值组合学、拉姆齐理论、组合□化、计算复杂性、组合数论、组合矩阵论、组合化学等分支。根据研究的方法，组合数学又可分为代数组合学、拓扑组合学、概率组合学、分析组合学、组合恒等式机器证明等分支。

组合数学最初起源于东方，中国和印度在古代组合数学中都占据重要地位。选择与排序的基本思想起源于印度文化，而洛河图、杨辉三角（西方又称为 Pascal 三角）、朱-范德蒙德恒等式等则展示了中国悠久的组合数学历史。一般认为，近现代组合数学的研究始于 17 世纪，以 Pascal 和 G.Leibniz 的工作为代表。Pascal 在他的专著中给出了 Pascal 三角的许多应用。Leibniz 开创了组合数学的整数分拆、位置几何学 (geometry of position) 等多个方向。在组合数学的发展史上，许多著名数学家都做出了里程碑式的工作，如在图论、拉丁方和整数分拆等方面做出开创性贡献的 L.P.Euler，树的计数方面的先驱 A.Cayley，在利用图表研究分拆、不变量理论等多个方面做出重要贡献的 J.J.Sylvester，在排列计数、平面分拆等方面做出重要贡献的 P.A.MacMahon，现代组合设计理论的奠基人 T.P.Kirkman，Boole 代数的奠基人 G.Boole，拟阵理论的代表人物 H.Whitney、W.T.Tutte 和 P.Seymour 等，Pólya 计数定理的发现者 G.Pólya，引领组合集合论、组合数论和组合数学中的概率方法发展的 P.Erdös、L.Lovász、E.Szemerédi 和 N.Alon 等，以及深刻影响着当代计数组合学和代数组合学发展的 G.C.Rota、M.P.Schüzenberger、R.P.Stanley、G.E.Andrews 等。

中国数学家在组合数学方面也做出了突出贡献。例如，清代数学家李善兰在《垛积比类》中给出了中国恒等式 (Chinese identity)，陆家羲解决了施泰纳三元系大集的存在性难题，徐利治提出了关于证明和发现组合恒等式的古尔□-徐反演等。

国际数学大师 I. M. Gelfand 曾语言几何学和组合数学将是 21 世纪数学研究的前沿领域。在美国数学会“21 世纪数学的挑战”大型研讨会上，“组合学中的代数方法”被列为纯数学中最具挑战性的方向之一。不仅如此，近年菲尔兹奖获得者的工作也都与组合数学密切相关。例如，1990 年菲尔兹奖得主 V.Jones 发现的一个重要扭结不变量（现称为 Jones 多项式）与图论中的塔特多项式密切相关。P.G.Tait 在 1898 年猜想链的约化交错投影中的相交数是链自身的拓扑不变量，利用 Jones 多项式与塔特多项式之间的联系，这个长期悬而未决的猜想得以解决。1998 年，T.Gowers 因为把泛函分析理论和组合数学理论联系起来而获得菲尔兹奖，A.Okounkov 在平面分拆方面做出了重要贡献。2006 年菲尔兹奖获得者陶哲轩最著名的工作是组合数论领域的重大讲展，陶哲轩还在几何组合中的挂谷猜想 (Kakeya conjecture) 以及代数组合中的饱和猜想 (saturation conjecture) 方面做出了杰出的贡献。

3.1 排列组合
Permutation and Combination

3.1.1 基本概念
Basic Concept

定理 3.1 (牧羊人原理)

设 f 为从集合 X 到一个集合 Y 上的映射，Y 为 q 元集，且对于所有 $y \in \mathrm{Y}$，集合 $f^{-1}(y) \subset \mathrm{X}$ 为 p 元集，则 X 为 pq 元集

证明 取定一个 p 元集 F，对于每个 $y \in \mathrm{Y}$ 取定一个从 F 到 $f^{-1}(y)$ 上的一个映射 u_y。对于 $y \in \mathrm{Y}$ 和 $z \in \mathrm{F}$ 令 $u(y,z) = u_y(z)$，则定义了映射 $u : \mathrm{Y} \times \mathrm{F} \to \mathrm{X}$ 。由 f 为满射，则可验证 u 为双射，则有

$$\mathrm{Card}(\mathrm{X}) = \mathrm{Card}(\mathrm{Y} \times \mathrm{F}) = \mathrm{Card}(\mathrm{Y}) \times \mathrm{Card}(\mathrm{F}) = qp$$

定理即证

定理 3.2 (排列引入/p 元集到 q 元集映射个数)

设 X 为 p 元集，Y 为 q 元集，则从 Y 到 X 内映射的集合为 p^q 元集

证明 由基数取幂运算即得

注 取 Y 为满足 $1 \leqslant i \leqslant q$ 的自然数 i 的集合，从 Y 到 X 内的一个映射为 X 的元素的一个族 $(x_i)_{i \in \mathrm{Y}}$，记为 $(x_i)_{1 \leqslant i \leqslant q}$，称族为 X 的 q 个元素的一个**排列** (permutation)。该定理表明，若 X 为 p 元集，则这些排列的个数为 p^q

定理 3.3 (p 元集到 q 元集单射个数)

设 X 为 p 元集，Y 为 q 元集，若 $p \leqslant q$，则从 X 到 Y 内的单射的个数为

$$\frac{q!}{(q-p)!}$$

其中对于自然数 n，若 $n \neq 0$，令 $n! = 1 \cdot 2 \cdot 3 \cdots n$，若 $n = 0$，令 $n! = 1$

证明 归纳：若 $p = 0$，集 X 为空集，仅存在一个从 X 到 Y 内单射，定理显然。假设 $\mathrm{Card}(\mathrm{X}) = p + 1$ 和 $\mathrm{Card}(\mathrm{Y}) = q \geqslant p + 1$。选定一个 $a \in \mathrm{X}$ 并令 $\mathrm{X}' = \mathrm{X} - \{a\}$，则 X' 具有 p 个元素。记从 X 到 Y 内单射的集合为 I。若对于所有 $f \in \mathrm{I}$，$u(f) = f(a)$，则定义了映射 $u : \mathrm{I} \to \mathrm{Y}$，该从 I 到 Y 内的映射显然为满射（即对于所有 $b \in \mathrm{Y}$ 存在一个单射使得 $f(a) = b$）。对于一个给定的 $b \in \mathrm{Y}$，考虑使得 $f(a) = b$ 的 $f \in \mathrm{I}$；令 $\mathrm{Y}' = \mathrm{Y} - \{b\}$，一个这样的 f 显然诱导出一个从 X' 到 Y' 内的单射 f'，反之所有从 X' 到 Y' 内的单射 f' 可以延拓为一个从 X 到 Y 内的单射 f，使得 $f(a) = b$。由归纳假设，使得 $u(f)$ 给定的 $f \in \mathrm{I}$ 的个数为

$$\frac{(q-1)!}{[(q-1)-(p-1)]!} = \frac{(q-1)!}{(q-p)!}$$

由于 u 映射 I 到 q 元集 Y 上，由牧羊人原理 (3.1) 有

$$\mathrm{Card}(I) = q \cdot \frac{(q-1)!}{(q-p)!}$$

则有 $q! = q \cdot (q-1)!$

定义 3.1 (置换)

称从有限集 X 到有限集 X 内的双射为置换

推论 3.1 (置换个数)

n 元集 X 的置换的个数为 $n!$

♡

证明 由于 X 为有限集，则由定理 (??)，则 X 的置换也是为从 X 到 X 内的一个单射。又由定理 (3.2)，其中取 $Y = X$ 和 $p = q = n$，并且注意到 $(n-n)! = 0! = 1$

注 称数 $n!$ 为 n 的**阶乘** (factorial/факториал)，显然有关系

$$0! = 1, \quad 1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24, \quad 5! = 120, \quad \cdots$$

定理 3.4 (p 元子集个数)

设 X 为 n 元集，p 为小于或等于 n 的自然数，则包含在 X 内的 p 元集的个数为

$$\frac{n(n-1)\cdots(n-p+1)}{p!} = \frac{n!}{p!(n-p)!}$$

♡

证明 选定一个有 p 个元素的集合 E，用 I 表示从 E 到 X 内的单射的集合，用 P 表示 X 的 p 元子集的集合。对于所有 $f \in \mathrm{I}$，显然 $f(\mathrm{E})$ 为 X 的一个 p 元子集。令 $u(f) = f(\mathrm{E})$ 则定义了映射 $u : \mathrm{I} \to \mathrm{P}$ 且该映射为满射，因为 X 的一个 p 元子集 Y 等势于 E，故为从 E 到 X 内的一个单射 E 的像。

计算使得 $f(\mathrm{E}) = \mathrm{Y}$ 的 $f \in \mathrm{I}$ 的个数，这里 Y 为 X 的一个给定的子集。设 f_0 为一个如下映射，把 f_0 与集合 E 的一个任意的置换复合就可以得到其他的映射：若 $f(\mathrm{E}) = f_0(\mathrm{E})$，则对于所有 $x \in \mathrm{E}$ 存在唯一的 $s(x) \in \mathrm{E}$ 使得 $f(x) = f_0(s(x)), s$ 为 E 的一个置换。则令 E 的每个置换对应从 E 到 X 内的映射 $f = f_0 \circ s$，则有从 E 的置换的集合到使得 $f(\mathrm{E}) = \mathrm{Y}$ 的 $f \in \mathrm{I}$ 的集合上的一个双射。由定理 (3.2)，对于给定的 Y，使得 $f(\mathrm{E}) = \mathrm{Y}$ 的 $f \in \mathrm{I}$ 的个数为 $p!$ 。由牧羊人原理 (3.1) 有 $\mathrm{Card}(\mathrm{I}) = p!\,\mathrm{Card}(\mathrm{P})$，则由定理 (3.2) 有

$$\mathrm{Card}(\mathrm{P}) = \frac{\mathrm{Card}(\mathrm{I})}{p!} = \frac{n!}{p!(n-p)!}$$

定理即证

注 习惯上记

$$\frac{n!}{p!(n-p)!} = \binom{n}{p}$$

称其为**二项式系数** (binomial coefficient)

定理 3.5 (Newton 二项式定理/binomial theorem)

设 n 为正整数，则有

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

♡

证明 由归纳法即证

推论 3.2 (二项式系数基本等式)

设 $n \geqslant k \geqslant 0$，则有

$$\binom{n}{k} = \binom{n}{n-k}$$

♡

推论 3.3

设 $n \geqslant 0$，则有

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

♡

推论 3.4
设 $n \geqslant 1$，则有

$$0=\sum_{k=0}^{n}(-1)^k\binom{n}{k}$$

♡

定理 3.6
设 $n \geqslant 1$，则有

$$\sum_{k\text{ 为奇数}}\binom{n}{k}=\sum_{k\text{ 为偶数}}\binom{n}{k}$$

♡

证明 设 $X=\{1,2,\cdots,n\}, A=\{S\subseteq X||S|\text{ 为偶数且 }1\in S\}, B=\{S\subseteq X||S|\text{ 为奇数且 }1\in S\}, C=\{S\subseteq X||S|\text{ 为偶数且 }1\notin S\}$，$D=\{S\subseteq X||S|\text{ 为奇数且 }1\notin S\}$。构造映射 $f:A\to D$ 为 $f(S)=S\backslash\{1\}$，显然 f 为双射，则由 $|A|=|D|$。类似地，有 $|B|=|C|$，因此有

$$\sum_{k}\text{ 为奇数}\binom{n}{k}=|B|+|D|=|A|+|C|=\sum_{k\text{ 为偶数}}\binom{n}{k}$$

定理即证

注 该定理除了集合论方法也可以使用二项式定理的推论 (3.4) 证明

定理 3.7 (Vandermonde 恒等式)
设 $n,m\geqslant 0$，则有

$$\binom{m+n}{k}=\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}$$

♡

证明

证明一：(比较系数法)

比较等式 $(x+1)^{m+n}=(x+1)^m(x+1)^n$ 两边 x^k 系数即证

证明二：

$\binom{m+n}{k}$ 为 $(m+n)$ 元集 $A\cup B$ 中 k 元子集的个数，这里 $A=\{1,\cdots,m\}, B=\{m+1,\cdots,m+n\}$，而其中包含 A 中 i 个元素的 k 元子集的个数为 $\binom{m}{i}\binom{n}{k-i}$，则有 $\sum\limits_{i=0}^{m}\binom{m}{i}\binom{n}{k-i}$ 即为对所有的 i 来计这些子集的个数

推论 3.5
在 Vandermonde 恒等式 (3.7) 中，令 $m=n=k$，则有

$$\binom{2n}{n}=\sum_{i=0}^{n}\binom{n}{i}^2$$

♡

推论 3.6 (Pascal 恒等式/杨辉恒等式)
在 Vandermonde 恒等式 (3.7) 中，令 $m=1$，则

$$\binom{n+1}{k}=\binom{n}{k}+\binom{n}{k-1}$$

♡

定理 3.8 (朱世杰恒等式)

设 $n,m \geqslant 0$，则有

$$\binom{m+n+1}{n+1} = \sum_{i=0}^{m} \binom{n+i}{n}$$

证明 比较等式

$$(x+1)^{m+n+1} = \underbrace{(x+1)(x+1)\cdots(x+1)}_{m+n+1}$$

两端 x^{n+1} 的系数，左端 x^{n+1} 的系数为 $c(m+n+1, n+1)$，右端 x^{n+1} 的系数可看做从 $m+n+1$ 项中选取 $n+1$ 项利用其中的 x 的方法数，这时按照产生第一个 x 的位置分类：第一个 x 选自第 j 项时，在剩余的 $m+n+1-j$ 项内选取其他 n 个 x，则剩余的项数至少为 n，至多为 $m+n$，且有 $c(m+n+1-j, n)$ 种选取方式得到 x^n，从而右端 x^{n+1} 的系数为

$$\sum_{j=1}^{m+1} \binom{m+n+1-j}{n} = \sum_{i=0}^{m} \binom{n+i}{n}$$

命题即证

推论 3.7

在朱世杰恒等式 (3.8) 中，令 $n=k, m=n-k$，则有

$$\binom{n+1}{k+1} = \sum_{i=0}^{n-k} \binom{k+i}{k} = \sum_{i=k}^{n} \binom{i}{k}$$

定理 3.9 (多项式定理/multinomial theorem)

设 n 为正整数，则有

$$(x_1+x_2+\cdots+x_k)^n = \sum_{n_1+n_2+\cdots+n_k=n} \binom{n}{n_1,n_2,\cdots,n_k} x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$$

其中

$$\binom{n}{n_1,n_2,\cdots,n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$$

证明 仅需考虑 $x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$ 在 $(x_1+x_2+\cdots+x_k)^n$ 展开式中系数即得

3.1.2 递推关系

Recurrence Relation

例题 3.1 (Fibonacci 问题)

意大利比萨的斐波那契 (Leonardo Fibonacci) 在 1202 年出版的书《珠算原理》(Liber Abaci) 里提出一个有趣的问题：假定一对刚出生的小兔一个月后就能长成大兔，再过一个月便能生下一对小兔，并且此后每个月都生一对小兔。若不考虑死亡问题，则一对刚出生的兔子，一年内能繁殖成多少对兔子?

解 记 $f_n(n \geqslant 0)$ 为第 n 个月时大兔的对数，则 $f_0=0, f_1=1$。考察第 $n(n \geqslant 2)$ 个月时的大兔对数由两部分组成：一部分为第 $n-1$ 个月时就已是大兔的兔子，共 f_{n-1} 对；另一部分为第 n 个月时刚长成的大兔，即第 $n-1$ 个月时出生的兔子，它们由第 $n-2$ 个月时的大兔生出，共有 f_{n-2} 对。于是有

$$f_n = f_{n-1} + f_{n-2}, \quad n \geqslant 2$$

故第 n 个月时所有兔子的对数为 $f_n + f_{n-1}$，则所求即为 $f_{12} + f_{11}$ $(= f_{13})$，由初始值与等式计算得到 $f_{13} = 233$，即问题答案为 233 对

定理 3.10 (Fibonacci 序列)
设 $f_n = f_{n-1} + f_{n-2}, n \geqslant 2$ 且 $f_0 = 0, f_1 = 1$，则当 $n \geqslant 0$ 时有

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

证明 记

$$g_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n, \quad n \geqslant 0$$

则有 $g_0 = f_0 = 0, g_1 = f_1 = 1$，且当 $n \geqslant 2$ 时，$g_n = g_{n-1} + g_{n-2}$。注意到 f_n 与 g_n 可由相同的初始值及关系得出，则有

$$f_n = g_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n, \quad n \geqslant 0$$

定理即证

定义 3.2 (常系数线性齐次递推关系)
设数列 $\{h_n\}_{n=0}^{\infty}$，若对所有的 $n \geqslant k$ 有

$$h_n = a_1 h_{n-1} + a_2 h_{n-2} + \cdots + a_k h_{n-k}$$

其中 $a_k \neq 0 (k \geqslant 1)$ 为常数，则称数列满足 k 阶常系数线性齐次递推关系 (linear homogeneous recurrence relation with constant coefficients)

设 q 为非零复数，则 $h_n = q^n$ 为上述 k 阶常系数线性齐次递推关系的解当且仅当 q 为 k 次多项式方程

$$g(x) = x^k - a_1 x^{k-1} - \cdots - a_{k-1} x - a_k = 0$$

的根，则称方程 $g(x) = x^k - a_1 x^{k-1} - \cdots - a_{k-1} x - a_k = 0$ 为上述常系数线性齐次递推关系的特征方程 (characteristic equation)

定理 3.11 (简单 k 阶常系数线性齐次递推关系解结构)
若 k 阶常系数线性齐次递推关系

$$h_n = a_1 h_{n-1} + a_2 h_{n-2} + \cdots + a_k h_{n-k}, \quad a_k \neq 0, k \geqslant 1$$

的特征方程 $g(x) = 0$ 有 k 个互不相同的根 $q_1, q_2, \cdots, q_k$，则

$$h_n = c_1 q_1^n + c_2 q_2^n + \cdots + c_k q_k^n, \quad n \geqslant 0 \tag{3.1}$$

为下述意义下的通解，亦即无论给定怎样的初始值 $h_0, h_1, \cdots, h_{k-1}$，都存在相应的常数 $c_1, c_2, \cdots, c_k$ 使得式 (3.1) 为满足上述递推关系和初始条件的唯一数列

证明 显然 $g(x) = 0$ 无零根（$a_k \neq 0$）。故对任意 $1 \leqslant i \leqslant k$ 数列 $\{q_i^n\}_{n=0}^{\infty}$ 满足上述递推关系，则任意常系数线性组合所得数列 $\{d_1 q_1^n + d_2 q_2^n + \cdots + d_k q_k^n\}_{n=0}^{\infty}$ 也满足上述递推关系。

对任意给定的初始值 $h_0, h_1, \cdots, h_{k-1}$，方程组

$$x_1 q_1^i + x_2 q_2^i + \cdots + x_k q_k^i = h_i, \quad 0 \leqslant i \leqslant k-1$$

的系数矩阵 $\boldsymbol{A}$ 为 Vandermonde 矩阵，则有

$$\det(\boldsymbol{A}) = \prod_{1 \leqslant i < j \leqslant k} (q_j - q_i) \neq 0$$

则上述方程组有唯一解 $x_i = c_i (0 \leqslant i \leqslant k-1)$。令

$$h_n = c_1 q_1^n + c_2 q_2^n + \cdots + c_k q_k^n, \quad n \geqslant 0$$

则数列 $\{h_n\}_{n=0}^{\infty}$ 既满足递推关系又满足初始条件，由上述推导知 $c_1, c_2, \cdots, c_k$ 唯一

定理 3.12 (k **阶常系数线性齐次递推关系解结构**)
设 k 阶常系数线性齐次递推关系

$$h_n = a_1 h_{n-1} + a_2 h_{n-2} + \cdots + a_k h_{n-k}, \quad a_k \neq 0, k \geqslant 1$$

的特征方程 $g(x) = 0$ 有互异根 $q_1, q_2, \cdots, q_t$，其中 q_i 为 s_i 重根 ($1 \leqslant i \leqslant t, s_1 + s_2 + \cdots + s_t = k$)，则有

$$h_n = \sum_{i=1}^{t} P_i(n) q_i^n$$

为该递推关系的一般解，其中 $P_i(n)$ 为关于 n 的次数小于 s_i 的多项式 ♡

注 该定理的证明较为复杂，在微分方程处给出

例题 3.2 （Hanoi 塔问题）

设有三根木柱和套在其中第一根木柱上的 n 个圆盘，圆盘自上而下尺寸递增（即最大的圆盘在底部）。现欲将圆盘全部转移到第二根木柱上，且仍保持圆盘上下顺序不变。规定每次只能移动一片圆盘，移动时不允许将大圆盘放在小圆盘上面，且可利用第三根木柱存放圆盘。问：最少需要多少次移动，才能完成满足要求的转移?

解 用 h_n 表示移动的次数，显然 $h_1 = 1$。对于 n 个圆盘，先把上面的 $n-1$ 个圆盘转移到第三根木柱上，再把最下面的那个圆盘移到第二根木柱上，最后把第三根木柱上的 $n-1$ 个圆盘转移到第二根木柱上，由此可得 $h_n = 2h_{n-1} + 1$ 。显然，当 $n \geqslant 0$ 时，有 $h_n = 2^n - 1$

定义 3.3 (**常系数线性非齐次递推关系**)
设数列 $\{h_n\}_{n=0}^{\infty}$，若对所有的 $n \geqslant k$ 有

$$h_n = a_1 h_{n-1} + a_2 h_{n-2} + \cdots + a_k h_{n-k} + f(n)$$

其中 $a_k \neq 0 (k \geqslant 1)$ 为常数，$f(n) \neq 0$ 为关于 n 的函数，则称数列 $\{h_n\}_{n=0}^{\infty}$ 满足 k 阶常系数线性非齐次递推关系 (linear nonhomogeneous recurrence relation with constant coefficients)
对于线性非齐次递推关系的通解，可通过求解其齐次部分的通解，再找到一个满足原递推关系的特解来得到，即线性非齐次递推关系的通解 = 线性齐次递推关系的通解 + 线性非齐次递推关系的一个特解 ♣

3.1.3 计数原理
Counting Principle

注 将介绍著名的十二重计数方法，如表 (3.1) 所示

其中 $(m)_n$ 表示 $m(m-1)\cdots(m-n+1)$，$S(n,i)$ 和 $p_i(n)$ 分别为第二类 Stirling 数与分拆数，另外

$$\delta(n \leqslant m) = \begin{cases} 1, & n \leqslant m, \\ 0, & \text{其他} \end{cases}$$

定义 3.4 (**排列数**)
称形如

$$P(n,r) = n(n-1)\cdots(n-r-1) = \frac{n!}{(n-r)!}$$

的数为排列数 (number of permutation)。该排列数计数在 n 个元素中取出 r 个序列的方法数 ♣

集合 A 中的元素（基数为 n ）	集合 B 中的元素（基数为 m ）	映射 f	f 的个数
各自不同	各自不同	不加限制	m^n
各自不同	各自不同	单射	$(m)_n$
各自不同	各自不同	满射	$m!S(n,m)$
各自不同	全部相同	不加限制	$\sum\limits_{i=1}^{m} S(n,i)$
各自不同	全部相同	单射	$\delta(n \leqslant m)$
各自不同	全部相同	满射	$S(n,m)$
全部相同	各自不同	不加限制	$\begin{pmatrix} n+m-1 \\ n \end{pmatrix}$
全部相同	各自不同	单射	$\begin{pmatrix} m \\ n \end{pmatrix}$
全部相同	各自不同	满射	$\begin{pmatrix} n-1 \\ m-1 \end{pmatrix}$
全部相同	全部相同	不加限制	$\sum\limits_{i=1}^{m} p_i(n)$
全部相同	全部相同	单射	$\delta(n \leqslant m)$
全部相同	全部相同	满射	$p_m(n)$

表 3.1: 十二重计数方法

定义 3.5 (组合数)

称形如
$$C(n,r)=\frac{P(n,r)}{r!}=\frac{n!}{(n-r)!r!}=\binom{n}{r}$$
的数为组合数 (combinatorial number)，读为"n 选取 r"。该组合数计数从 n 个元素中选取 r 个的方法数。若 $n<r$，则认为 $P(n,r)=C(n,r)=0$。即有
$$\binom{n}{k}=\begin{cases}\dfrac{n!}{k!(n-k)!}, & n\geqslant k,\\ 0, & n<k.\end{cases}$$
称形如
$$\binom{n}{r_1,r_2,\cdots,r_k}=\frac{n!}{r_1!r_2!\cdots r_k!},$$
的数为参数为 $n,r_1,\cdots,r_k$ 的多重组合数，其中 $n=\sum\limits_{i=1}^{k}r_i$
称形如
$$\frac{n!}{(n-r)!}\cdot\frac{1}{r}$$
的数为循环排列数 (number of circular permutation) 或圆排列数或环排列数。该环排列数计数从 n 个不同元素中取出 r 个排成一个圆环的方法数 ♣

例题 3.3 （消序法）

把集合 $\{1,2,\cdots,n\}$ 划分为 b_1 个 1 元子集，b_2 个 2 元子集，$\cdots,b_k$ 个 k 元子集，其中 $\sum\limits_{i=1}^{k}ib_i=n$，这样的分法有多少种?

解 n 个元素的排列有 $n!$ 种，而对于每个划分其中 b_i 个 i 元子集无序，且划分中每个集合的元素也无序，因此每个划分对应 $b_1!b_2!\cdots b_k!(1!)^{b_1}(2!)^{b_2}\cdots(k!)^{b_k}$ 个不同的 n-排列
$$\frac{n!}{b_1!b_2!\cdots b_k!(1!)^{b_1}(2!)^{b_2}\cdots(k!)^{b_k}}$$
即为方法数

注 进一步有，一个 n 元集的全体划分数为
$$\sum_{b_1+2b_2+\cdots+nb_n=n}\frac{n!}{b_1!b_2!\cdots b_n!(1!)^{b_1}(2!)^{b_2}\cdots(n!)^{b_n}}$$

定理 3.13

令 S 为具有 n 类元素的一个多重集，每种元素均可以被重复选取任意多次，则 S 的 r-组合数为 $C(r+n-1,r)$ ♡

证明 (Euler)设这 n 个物体与自然数 $1,2,\cdots,n$ 一一对应，则所考虑组合便可看成含 r 个数的多重集 $\{c_1,c_2,\cdots,c_r\}$。因为与次序无关，不妨设 $c_1\leqslant c_2\leqslant\cdots\leqslant c_r$。构造另一组数 $\{d_1,d_2,\cdots,d_r\}$，其中 $d_i=c_i+i-1,1\leqslant i\leqslant r$，则即使有的 c_i 有重复，但 d_i 无重复，且任一种 $\{c_1,c_2,\cdots,c_r\}$ 的取法对应一种 $\{d_1,d_2,\cdots,d_r\}$ 的取法，反之亦然，从而这两个组合计数问题等价。注意到 c_i 最大可取 n，则 d_i 最大可取 $n+r-1$。因此从 n 个不同物体中可重复地任意多次选取一个基数是 r 的多重集的方法数，等于从 $n+r-1$ 个不同元素中不重复地选取 r 个元素的方法数，即 $C(r+n-1,r)$

定义 3.6 (第一类 Stirling 数)

对于正整数 n,k，定义 $c(n,k)$ 为 n 元对称群 S_n 中恰含 k 个循环（即恰可写成 k 个不交循环的乘积）的置换个数（不动点也看做一个循环），称 $s(n,k)=(-1)^{n-k}c(n,k)$ 为第一类 Stirling 数 (Stirling number of the first kind)，也常称 $c(n,k)$ 为无符号的第一类 Stirling 数
特别地，约定 $c(0,0)=1$，以及当 $n\geqslant 1$ 时，$c(n,0)=c(0,n)=0$ ♣

引理 3.1

对任意 $n\geqslant 1, k\geqslant 1, c(n,k)$ 满足递推关系

$$c(n,k)=(n-1)c(n-1,k)+c(n-1,k-1)$$

♡

证明 设置换 σ 是 S_n 中恰有 k 个轮换的置换。若 $\sigma(n)=n$，则 n 在 σ 中为一个单独的轮换，从而这样的 σ 的个数等于 S_{n-1} 中恰有 $k-1$ 个轮换的置换个数，即 $c(n-1,k-1)$。若 $\sigma(n)\neq n$，则将轮换 σ 中的 n 去掉就得到 S_{n-1} 中含 k 个轮换的置换。又将 n $\boxed{\text{F}}$人 S_{n-1} 中含 k 个轮换的置换中时，可得到 $n-1$ 个 S_n 中含 k 个轮换的置换，从而这样的 σ 的个数等于 S_{n-1} 中恰有 k 个轮换的置换个数的 $n-1$ 倍，即 $(n-1)c(n-1,k)$。因此

$$c(n,k)=(n-1)c(n-1,k)+c(n-1,k-1)$$

定理 3.14

$\{c(n,k)\}_{n=1}^{\infty}$ 满足如下的函数方程：

$$\sum_{k=1}^{n} c(n,k)x^k = x(x+1)\cdots(x+n-1)$$

♡

证明

对 n 用归纳法证明命题。当 $n=1$ 时，命题即 $x=x$，显然成立。

设 $n\geqslant 2$，且命题对 $n-1$ 成立，则对 n，由归纳假设及 $c(n,k)$ 的递推性质知，对任意 $1\leqslant k\leqslant n$，有

$$\begin{aligned}
&\left[x^k\right] x(x+1)\cdots(x+n-1)\\
&=\left[x^k\right] x(x+1)\cdots(x+n-2)x+(n-1)\left(\left[x^k\right] x(x+1)\cdots(x+n-2)\right)\\
&=\left[x^{k-1}\right] x(x+1)\cdots(x+n-2)+(n-1)\left(\left[x^k\right] x(x+1)\cdots(x+n-2)\right)\\
&=c(n-1,k-1)+(n-1)c(n-1,k)\\
&=c(n,k),
\end{aligned}$$

从而

$$\sum_{k=1}^{n} c(n,k)x^k = x(x+1)\cdots(x+n-1)$$

即命题对 n 成立。由归纳原理知，命题对一切正整数 n 成立，即 $\{c(n,k)\}_{n=1}^{\infty}$ 满足定理中所述的函数方程

定理 3.15

$\{s(n,k)\}_{n=1}^{\infty}$ 满足如下的函数方程：

$$\sum_{k=1}^{n} s(n,k)x^k = (x)_n$$

这里 $(x)_n = x(x-1)\cdots(x-n+1)$ ♡

证明 在上述定理中，用 $-x$ 代替 x，得

$$\sum_{k=1}^{n} c(n,k)(-x)^k = (-x)(-x+1)\cdots(-x+n-1)$$

上式两边同时乘以 $(-1)^n$，得

$$\sum_{k=1}^{n}(-1)^n c(n,k)(-x)^k = x(x-1)\cdots(x-n+1)$$

即

$$\sum_{k=1}^{n} s(n,k)x^k = (x)_n$$

第 4 章 概率空间 (Probability Space)

概率是事件发生的可能性大小的数量指标，其值介于 0 与 1 之间。概率的概念形成于 16 世纪初，最早流行于意大利数学家广为讨论的掷骰子相关问题，早期概率定义具体出自何人已难以考证，现今为人所知的是 Cardano。事件可以分为两种，其一是可重复事件，例如“抛掷骰子”，其二是一次性事件，例如“2023 年 1 月 19 日某人与恋人分手”。

早期的概率 (probability) 与机遇 (chance) 两词的用法略有区别：前者用于主观概率而后者用于客观概率，直到 18 世纪初才逐渐统一。另外早期也经常用“胜率 (odds) 一词：若甲与乙赌而甲胜的机遇为 $\frac{1}{3}$，则说他的胜率为 $1:2$ （胜率为双方获胜概率之比，早期只谈论整数比），该词直到现在仍然常用。

现今称可重复事件的概率为客观概率，而称一次性事件的概率为主观概率。依据这两种概率划分了两大数理统计学派，主观概率是 Bayes 学派的基础理论，客观概率的支持者则属于频率学派。在现代，大多数统计学者大多以兼容并蓄的态度对待频率学派与 Bayes 学派各具争议和优点的理论和方法。

主观概率取决于人的主观看法。有哲学家认为这是由于人的知识的不完全性，若关于某事件有关的知识完全掌握，则一次性事件的概率只有 0 异或 1。另外，主观概率可以反映某种信仰，例如“菩萨娘娘保佑”的概率对于佛教徒和基督教徒并不一样，因此逻辑学家和神学家对主观概率也各有其解释。

客观概率有两种形式。第一种形式是，试验的可能结果仅有有限个，且任一种结果都没有比其他结果占优势，则认为各结果有同等出现的机会（经常称第一种方式定义的概率为古典概率）。这种概率适用的经典场合即为以抛掷骰子为代表的博彩游戏。法国数学家 Fermat (1601-1665) 和 Blaise Pascal (1623-1662) 间的七封信被认为是古典概率的起点，而促使 Blaise Pascal 和 Fermat 通信的人则是法国贵族 De Mere，De Mere 曾向 Blaise Pascal 请教了几个有关赌博的问题。1654 年 7 月 29 日，Pascal 给 Fermat 写信转达了著名的 De Mere 问题 (4.9)，该问题是现代概率论起源相关的第一个问题。

据文献记载，在公元 960 年前后，怀特尔德大主教计算出抛掷三个骰子时不计骰子次序的组合数为 56 。而同样作为博彩工具的纸牌，至 1350 年仍无文献纪录。此后由于造纸术传入西方，纸牌在欧洲富裕阶级中逐渐常见，但由于教会反对以及一些国家明令禁止，纸牌在很长时期内远不及骰子流行。直到 18 世纪初纸牌才取代骰子作为主要工具。因此，促进概率论诞生的功劳应该归功于骰子。

客观概率的第二种形式是依据某事件在大量重复试验中出现的频率来定义概率，尽管这仍然只是对该概率的一种估计（经常称第二种方式定义的概率为统计概率）。著名瑞士数学家 Jakob Bernoulli (1654-1705) 在该领域做出了关键贡献，在 Jakob Bernoulli 的家族中，其中至少有 5 人在概率论方面做出过贡献。Jakob Bernoulli 与荷兰数学家 Huygens 长期保持通信，并研读过 Huygens (1629-1695) 在 1657 年出版的著作《机遇的规律》，启发了其对概率论的兴趣。在 1713 年 Jakob Bernoulli 出版了划时代著作《推测术》，该书一共分为四个部分。前三部分是古典概率的系统化和深化，第二部分首次提出了“排列”的概念，同样是组合学史的里程碑，成为了当时排列组合理论的标准教材。而最为经典的第四部分则是概率论在各领域的应用，包含了后世著名的 Bernoulli 大数定律 (5.15)。对于均匀硬币大量重复投掷，正面出现的频率大约是 0.5，Bernoulli 曾说：“哪怕最笨的人，不通过别人的教诲也能理解频率大约是二分之一”。历史上也曾有不少数学家做过试验：

概率史家通常把《推测术》的出版作为现代概率论的起点，根据 Jakob Bernoulli 与 Leibniz 的通信可知该书在 Jakob Bernoulli 生命最后两年完成，后来由 Leibniz 敦促 Jakob Bernoulli 的侄子，也是当时重要的数学家 Nicolaus(II) Bernoulli (1695-1726) 整理出版。

对于统计问题，Jakob Bernoulli 对事物采取了一种机械决定论的观点，亦即世界上的一切事物都受到严格的因果律的支配，后来大数学家 Laplace 也采取了相同的思想立场，因此站在哲学史和统计学史的角度《推测术》也依然是重要的著作，许多统计概念和统计方法至今仍沿用。《推测术》中引进了所谓“道德确定性”(moral certainty) 的概念，即一个事件虽不能确定其必然发生，但它若被认为以极大的可能性以至几乎不会不发生，就称它有道德确定性，这就是今天所说的实际推断原理 (5.11)。在《推测术》中 Jakob Bernoulli 还把古典概率中“等可能性”的思想推广到了主观概率。即若没有任何理由可以认为众多可能性中的某一个或某一些比其他可能性更具优势时，应给予这些可能性以同等的主观概率。例如，当人们对 A, B 两位围棋棋手的水平一无所知时，应给予 A 强于 B 及“B 强于 A”这两种事件相同的主观概率 $\frac{1}{2}$。后世把该原则称为“同等无知原则”。

另外，国际上的 Bernoulli 统计期刊和 Bernoulli 统计学会就以 Bernoulli 名字纪念 Bernoulli 在统计学的贡献。

英国哲学家 Bayes (1702-1761) 由朋友在 1763 年发表的遗作 An essay towards solving a problem in the doctrine of chances (机遇理论中一个问题的解) 开创了统计学中的 Bayes 学派。该论文的思想也同样基于 Bernoulli 的“同等无知原则”，而大数学家 Laplace (1749-1827) 后来提出的“不充分理由原则”，其思想并无本质不同。不同于频率学派，Bayes 学派对概率统计问题有自己的独特理解，但也有一些观念上的难以自圆其说之处。Bayes 学派对先验概率的解释和处理经常受频率学派的批评和指责。尤其是他们在一无所知的情况下，主张按照“同等无知原则”赋予各个可能的成因以相同的概率。20 世纪频率学派与 Bayes 学派进行了不少的辩论和驳斥，但至今仍然无法达成一致，这也是 20 世纪数学史上引人注目的亮点。如今所说的 Bayes 学派，大都指坚持主观概率的学者。

随着概率论的发展，一些对于概率的理解出现了歧义，例如法国数学家 Bertrand 于 1889 年提出的 Bertrand 奇论 (4.16)。Bertrand 奇论对于相同问题甚至给出了不同的概率，这使得概率论理论的统一和公理化迫在眉睫。1933 年，前苏联大数学家 Kolmogorov[a]在选用数量很少且极为简单的公理的情况下实现了概率论的公理化。现今不论是谈论客观概率还是主观概率，基本都遵守 Kolmogorov 公理体系。

在中国，许宝□ (1910-1970) 教授是概率论和数理统计研究的先驱，作出了杰出的贡献，在国际上享有盛誉。1979 年，《数理统计年鉴》(The Annnals of Statistics) 邀请了一些著名学者撰文介绍了许宝□的生平，高度评价了许宝□在概率论和数理统计两方面的研究工作。

[a]安德雷 · 柯尔莫哥洛夫 (俄语：Андре́й Никола́евич Колмого́ров，1903.4.25-1987.10.20) 俄罗斯数学家，主要在概率论、算法信息论和拓扑学方面做出了重大贡献，最为人所称道的是对概率论公理化所作出的贡献

4.1 基本概念
Basic Concept

4.1.1 试验与事件
Experiment and Event

> Whatever begins to exist, must have a cause of existence.
> （凡开始存在的东西皆有其存在的原因）
>
> David Hume　大卫 · 休谟

注（试验）

若在一个偶发性介入的动作结束后可能有多种不同的现象出现，并且事前无法知道究竟会出现哪一种现象，则称这些现象为**随机现象** (random phenomenon)，称为了解随机现象所可能导致的所有不同影响及其发生规律而进行的实验或观察为**试验** (experiment/эксперимент)

例题 4.1 （试验）

以下动作为一个试验：掷一枚硬币，观察是否正面朝上

注 （试验）

在概率论中，一般不称“试验在一组条件下进行”，而默认这样的“条件”存在

定义 4.1 (结果)

称试验后可能发生的现象为结果 (исход) ♣

例题 4.2 （结果）

掷一枚硬币，若用 H(*head*) 表示硬币正面朝上，用 T(*tail*) 表示硬币反面朝上，则试验有两个可能的结果：H 和 T

定义 4.2

（随机试验）若试验满足下列条件：

1) （可重复性）可以在相同的条件下重复地进行
2) （空间确定性）每次试验存在可能的结果，且能事先明确试验的所有可能结果
3) （元素不确定性）进行一次试验之前不能确定哪一个结果会出现

则称该试验为随机试验 (random experiment/случайный эксперимент) ♣

注 （随机试验）

在概率论中，今后除特殊说明，均把“试验”理解为“随机试验”

定义 4.3 (样本空间)

若某些结果的集合满足下列条件：

1) 每次试验至少出现一个集合中的结果
2) 每次试验至多出现一个集合中的结果

换言之，每次试验出现且唯一出现集合中的一个结果，则称该集合为基本结果空间 (пространство элементарных исходов) 或样本空间 (sample space)，称其中的结果为基本结果 (элементарный исход) 或样本 (sample/выборка) 或样本点 (sample point)，基本结果空间记为 Ω，其元素即基本结果记为 ω ♣

例题 4.3 （样本空间）

将一枚硬币重复抛掷 n 次，则样本空间为

$$\Omega = \{\omega | \omega = (a_1, \cdots, a_n)\}$$

其中 a_i 表示正面或反面，则事件总数为 $\mathrm{Card}(\Omega) = 2^n$

注 对于结果的顺序不影响事件发生的情况，经常把样本空间记为

$$\Omega = \{\omega | \omega = [a_1, \cdots, a_n]\}$$

定义 4.4 (事件)

设样本空间 Ω，若 A 为 Ω 的子集，则称 A 为样本空间 Ω 的（基本）事件 (event/событие)。特别地，空集 $\varnothing$ 为 Ω 的子集，称 $\varnothing$ 为不可能事件 (impossible event/невозможное событие)；而 Ω 也为样本空间 Ω 的子集，则称 Ω 为确定事件 (certain event/достоверное событие) ♣

例题 4.4 （事件）

将一枚硬币抛掷三次，观察正面 H 与反面 T 出现的情况，事件“三次出现同一面”即为 $\{HHH, TTT\}$

定义 4.5 (并事件与交事件/和事件与积事件)

设事件 A,B，称 $A\cup B$ 为事件 A,B 的并事件 (union of event)；称 $A\cap B$ 为事件 A,B 的交事件 (intersection of event) 或积事件 (product event)，经常记交事件 $A\cap B$ 为 AB ♣

定义 4.6 (补事件)

设样本空间 Ω 上有事件 A,B, 称 A 在 Ω 的补集 A^c 为事件 A 的补事件 (complementary event) 或余事件；若事件 A,B 满足 $A\cup B=\Omega$，则称事件 A,B 互为对立事件 (complementary event/противоположное событие) ♣

定义 4.7 (互斥事件)

若事件 A,B 满足 $A\cap B=\varnothing$, 则称事件 A,B 不相容 (несовместны) 或互为互斥事件 (mutually exclusive event)，一般称互斥事件 A,B 的并事件为和事件 (sum event/сумма)，记为 $A+B$ ♣

定义 4.8 (子事件)

设事件 A,B，若 $A\subset B$，则称事件 A 蕴涵事件 B，也称事件 A 为事件 B 的子事件 ♣

4.1.2 概率测度与概率空间
Probability Measure and Probability Space

> 概率论作为数学学科，可以而且应该从公理开始建设，和几何、代数的路一样
> Андре́й Никола́евич Колмого́ров A. H. 柯尔莫戈洛夫

古典定义中经常使用事件代数的概念来定义概率空间，但德国数学家 Caratheodory 证明的概率扩张定理 (4.1) 表明，其可以任意地延拓到 σ 代数的情况

定义 4.9 (事件代数)

若事件族 $\mathcal{A}$ 满足：
1)（规范性）$\Omega\in\mathcal{A}$
2)（可逆性）$A\in\mathcal{A}\Rightarrow A^c\in\mathcal{A}$
3)（有限可并性）$A,B\in\mathcal{A}\Rightarrow A\cup B\in\mathcal{A}$
则称 $\mathcal{A}$ 为事件代数 (алгебра событий) ♣

定义 4.10 (弱概率空间)

设样本空间 Ω，$\mathcal{A}$ 为 Ω 上事件代数，定义函数 $P:\mathcal{A}\to\mathbb{R}$，若满足下列条件
1*)（非负性）$(\forall A\in\mathcal{A}):P(A)\geqslant 0$
2*)（规范性）$P(\Omega)=1$
3*)（有限可加性）若 $\{A_n|n\in\mathbb{N}\}$ 为一列两两不交的事件，则有

$$P\left(\bigcup_{k=1}^{n}A_k\right)=\sum_{k=1}^{n}P\left(A_k\right)$$

称空间 $(\Omega,\mathcal{A},P)$ 为（弱）概率空间 (probability space/вероятностное пространство) ♣

定理 4.1 (概率扩张定理/теорема о продолжении вероятности)

设 $(\Omega,\mathcal{A}_0,P)$ 为弱概率空间，且函数 P 满足可数可加性，则在 $\mathcal{A}=\sigma(\mathcal{A}_0)$ 上存在唯一的可数可加的函数 Q 满足

$$(\forall A\in\mathcal{A}_0):P(A)=Q(A)$$

证明 见附录 (7.1)

注 该定理首先由 Caratheodory[1]证明

定义 4.11 (概率空间)

设样本空间 Ω，$\mathcal{A}$ 为 Ω 子集的 σ 代数，定义测度 $P:\Omega\to\mathbb{R}$，若满足

1) （非负性）对于任意事件 E 都有 $P(E)\geqslant 0$
2) （规范性）$P(\Omega)=1$
3) （可数可加性）若 $\{E_n|n\in\mathbb{N}\}$ 为一列两两不交的事件，则有

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}P(E_n)$$

称定义了测度 P 的可测空间 $(\Omega,\mathcal{A})$ 为概率空间[a](probability space/вероятностное пространство)，记为 $(\Omega,\mathcal{A},P)$，称测度 P 为概率测度 (probability measure)

[a]该定义最早由苏联数学家 А.Н.Колмогоров 于 1933 年在德语撰写的《概率论基本概念》中提出，于 1936 年出版俄语版 “Основные понятия теории вероятностей”

注 （并事件与交事件）根据新的概率空间，对于可数个事件 $A_1,A_2,\cdots$，称

$$A_1\cup A_2\cup\cdots=\bigcup_{k=1}^{\infty}A_k=\lim_{n\to\infty}\bigcup_{k=1}^{n}A_k$$

为事件 $A_1,A_2,\cdots$ 的**并事件** (union of event)；称

$$A_1\cap A_2\cap\cdots=\bigcap_{k=1}^{\infty}A_k=\lim_{n\to\infty}\bigcap_{k=1}^{n}A_k$$

为事件 $A_1,A_2,\cdots$ 的**交事件** (intersection of event)

性质 （概率空间性质）若概率空间满足条件 1)-3)，则满足下列性质

4) （规范性）$P(\varnothing)=0$
5) （有限可加性）若 $A_k(k=1,2,\cdots,n)$ 为 n 个两两不交的事件，则有

$$P\left(\bigcup_{k=1}^{n}A_k\right)=\sum_{k=1}^{n}P(A_k)$$

6) （可减性）若 $A_1\in\mathcal{A},A_2\in\mathcal{A}$，并且 $A_1\supset A_2$，则有

$$P(A_1-A_2)=P(A_1)-P(A_2)$$

7) （单调性）若 $A_1\in\mathcal{A},A_2\in\mathcal{A}$ 并且 $A_1\supset A_2$，则有

$$P(A_1)\geqslant P(A_2)$$

8) （基本等式）$P(A^c)=1-P(A)$
9) (Kolmogorov 加法公理/容斥原理/principle of inclusion-exclusion/формула включения и исключения)

[1]卡拉西奥多里 (Caratheodory.Constantin,1873.9.13-1950.2.2) 希腊裔德国数学家。1902 年在 Minkowski 指导下取得博士学位，曾是《数学年刊》的编辑

若 $E_k(k=1,2,\cdots,n)$ 为任意 n 个事件，则

$$P\left(\bigcup_{k=1}^{n} A_k\right)=\sum_{k=1}^{n} P\left(A_k\right)-\sum_{1\leqslant i<j\leqslant n} P\left(A_i A_j\right)+\sum_{1\leqslant i<j<k\leqslant n} P\left(A_i A_j A_k\right)$$
$$+\cdots+(-1)^{n-1} P\left(A_1 A_2 \cdots A_n\right)$$

特别地，若 $n=2$，则有

$$P\left(E_1 \cup E_2\right)=P\left(E_1\right)+P\left(E_2\right)-P\left(E_1 E_2\right)$$

10) （下连续性）若 $E_n(n\in\mathbb{N})$ 为上升的事件序列，即 $E_n\subset E_{n+1}, n=1,2,\cdots$，则有

$$P\left(\bigcup_{n=1}^{\infty} E_n\right)=\lim_{n\to\infty} P\left(E_n\right)$$

11) （上连续性）若 $E_n(n\in\mathbb{N})$ 为下降的事件序列，即 $E_n\supset E_{n+1}, n=1,2,\cdots$，则有

$$P\left(\bigcap_{n=1}^{\infty} E_n\right)=\lim_{n\to\infty} P\left(E_n\right)$$

12) （半可加性）若 $E_n(n\in\mathbb{N})$ 为一列事件，则有

$$P\left(\bigcup_{n=1}^{\infty} E_n\right)\leqslant\sum_{n=1}^{\infty} P\left(E_n\right)$$

证明 4) 若取 $A_1=A_2=\cdots=\varnothing$，则有 $(\forall i\neq j): A_i\cap A_j=\varnothing$ ，因此 $A_n(n\in\mathbb{N})$ 为一列两两不交的事件，则由 3) 得

$$P(\varnothing)=P\left(\bigcup_{n=1}^{\infty} A_n\right)=\sum_{n=1}^{\infty} P\left(A_n\right)=\sum_{n=1}^{\infty} P(\varnothing)$$

该等式当且仅当 $P(\varnothing)=0$ 时才能成立

5) 事实上，只需在 3) 中令 $A_{n+1}=A_{n+2}=\cdots=\varnothing$ 即得

6) 因为此时 $A_1A_2=A_2$，而 $A_1A_2\cup A_1A_2^{c}=A_1$，则由有限可加性

$$P\left(A_1\right)=P\left(A_1A_2\right)+P\left(A_1A_2^{c}\right)=P\left(A_2\right)+P\left(A_1-A_2\right)$$

7) 此由可减性 6) 立得

8) 在 6) 中令 $A_1=\Omega, A_2=A$ 即得

9) 归纳：当 $n=2$ 时易知事件 $E_1E_2, E_1E_2^c, E_1^cE_2$ 两两不交且

$$E_1\cup E_2=E_1E_2\cup E_1E_2^c\cup E_1^cE_2,\quad E_1E_2\cup E_1E_2^c=E_1,\quad E_1E_2\cup E_1^cE_2=E_2$$

则有

$$\begin{aligned}
P\left(A_1\cup A_2\right)&=P\left(A_1A_2\right)+P\left(A_1A_2^c\right)+P\left(A_1^cA_2\right)\\
&=\left\{P\left(A_1A_2\right)+P\left(A_1A_2^c\right)\right\}+\left\{P\left(A_1A_2\right)+P\left(A_1^cE_2\right)\right\}-P\left(A_1A_2\right)\\
&=P\left(A_1\right)+P\left(A_2\right)-P\left(A_1A_2\right)
\end{aligned}$$

假设当 $n=k$ 时结论成立，当 $n=k+1$ 时令 $A=\bigcup\limits_{j=1}^{k} A_j, B=A_{k+1}$，则由 $n=2$ 时结论得

$$P\left(\bigcup_{j=1}^{k+1} A_j\right)=P\left(\bigcup_{j=1}^{k} A_j\right)+P\left(A_{k+1}\right)-P\left(\bigcup_{j=1}^{k}\left(A_j\cap A_{k+1}\right)\right)$$

再代入 $n=k$ 时结论代入其中，整理后即知 $n=k+1$ 时结论也成立

10) 记 $A_1=E_1, A_n=E_nE_{n-1}^{c}, n=2,3,\cdots$，则事件 $A_1,A_2,\cdots$ 两两不交且有 $\bigcup\limits_{n=1}^{\infty} A_n=\bigcup\limits_{n=1}^{\infty} E_n$ 及

$E_m=\bigcup\limits_{n=1}^{m}A_n,m=1,2,\cdots$ 则由 3) 和 5) 知

$$P\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P\left(A_n\right)=\lim_{m\to\infty}\sum_{n=1}^{m}P\left(A_n\right)=\lim_{m\to\infty}P\left(\bigcup_{n=1}^{m}A_n\right)=\lim_{m\to\infty}P\left(A_m\right)$$

11) 由 $E_n^c(n\in\mathbb{N})$ 为上升事件序列，则由下连续性 10) 和 De Morgan 定律即得

12) 先由加法公理，证得关于有限个事件的半可加性：

$$P\left(\bigcup_{n=1}^{m}E_n\right)\leqslant\sum_{n=1}^{m}P\left(E_n\right)$$

再令 $m\to\infty$ 并根据下连续性 10) 即得

性质 （概率空间等价公理化条件）

若仅满足对于概率空间定义的条件 1) 和 2)，考虑重新证明下列命题等价

3) （可数可加性/счетная аддитивность） 若 $\{E_n|n\in\mathbb{N}\}$ 为一列两两不交的事件，则有

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}P\left(E_n\right)$$

10) （下连续性/непрерывность снизу/continuity from below） 若 $E_n(n\in\mathbb{N})$ 为上升的事件序列，即 $E_n\subset E_{n+1},n=1,2,\cdots$ 且 $E=\bigcup\limits_{n=1}^{\infty}E_n\in\mathcal{A}$，则有

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=\lim_{n\to\infty}P\left(E_n\right)$$

11) （上连续性/непрерывность сверху/continuity from above） 若 $E_n(n\in\mathbb{N})$ 为下降的事件序列，即 $E_n\supset E_{n+1},n=1,2,\cdots$ 且 $E=\bigcup\limits_{n=1}^{\infty}E_n\in\mathcal{A}$，则有

$$P\left(\bigcap_{n=1}^{\infty}E_n\right)=\lim_{n\to\infty}P\left(E_n\right)$$

13)（空集上的上连续性/непрерывность сверху на $\varnothing$）若 $E_n(n\in\mathbb{N})$ 为上升的事件序列，即 $E_n\subset E_{n+1},n=1,2,\cdots$，且 $E=\bigcup\limits_{n=1}^{\infty}E_n=\varnothing$ 则有

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=\lim_{n\to\infty}P\left(E_n\right)$$

证明 3)⇒10) 定义 $B_1=E_1,B_2=E_2\backslash E_1,\ldots,B_k=E_k\backslash E_{k-1},\ldots$，这时事件 $B_1,B_2,\ldots$ 两两不交，由 3) 即得

$$\lim_{n\to\infty}P\left(\sum_{k=1}^{n}B_k\right)=\lim_{n\to\infty}\sum_{k=1}^{n}P\left(B_k\right)=\sum_{k=1}^{\infty}P\left(B_k\right)=P\left(\sum_{k=1}^{\infty}B_k\right)$$

由于事件 $B_1,B_2,\ldots$ 两两不交，则可以改写为

$$P\left(\sum_{k=1}^{\infty}B_k\right)=P\left(\bigcup_{k=1}^{\infty}B_k\right)=P\left(\bigcup_{k=1}^{\infty}E_k\right)=\lim_{n\to\infty}P\left(\sum_{k=1}^{n}B_k\right)=\lim_{n\to\infty}P(E_n)$$

10)⇒11) 在 10) 的条件下定义 $B_n=E_n^c\in\mathcal{A}$，这时由 10) 即有

$$P\left(\bigcap_{n=1}^{\infty}E_n\right)=1-P\left(\left(\bigcap_{n=1}^{\infty}E_n\right)^c\right)=1-P\left(\bigcup_{n=1}^{\infty}B_n\right)=1-\lim_{n\to\infty}P\left(B_n\right)=\lim_{n\to\infty}P\left(E_n\right)$$

11)⇒13) 显然

13)⇒3) 在 13) 的条件下记 $B_n=E_1+\ldots E_n,C_n=\sum\limits_{k=n+1}^{\infty}E_k$，则序列 C_n 满足 $C_1\supset C_2\supset\ldots$ 且 $\bigcap\limits_{n=1}^{\infty}C_n=\emptyset$，此外 $E=B_n+C_n$，因此由 13) 即得 $\lim\limits_{n\to\infty}P\left(C_n\right)=0$，又由 $B_n\cap C_n=\varnothing$ 即有

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=P\left(B_n\right)+P\left(C_n\right)=\sum_{k=1}^{n}P\left(E_k\right)+P\left(C_n\right)=\sum_{k=1}^{\infty}P\left(E_k\right)+0$$

命题即证

注 证明过程中实际使用了概率的有限可加性 5)，由此概率空间的条件 3) 实际上应改为有限可加性。只需修改条件 3) 为条件 5) 后如 3)⇒10) 中那样做的即可从有限可加性 5) 得到可数可加性 3)，因此考虑到更广泛的应用场景选用可数可加性作为基础公理来定义概率空间

例题 4.5 （Bernoulli 空间/概率空间）

设 Ω 为样本空间，A 为 Ω 非空真子集，$\mathcal{A}=\{\Omega,\varnothing,A,A^c\}$，$p$ 为小于 1 的正数，记 $q=1-p$，令 $P(A)=p,P\left(A^c\right)=q,P(\Omega)=1,P(\varnothing)=0$。如此定义的集函数 P 满足性质 1)-3)，则 $(\Omega,\mathcal{A},P)$ 为一个概率空间

注 称该概率空间为 Bernoulli **空间**，Bernoulli 在其《推测术》中首先确立了这种概率模型

例题 4.6 （概率空间）

设样本空间 Ω 为全体正整数的集合，用 A_m 表示能被 m 整除的正整数，形式上定义比例函数为 P

1) 求 Ω 中被 3 整除，不被 5 整除且又能被 4 或 6 整除的数所占的比例 P

2) 设 $\mathcal{A}$ 为 A_3,A_4, A_5 和 A_6 的生成 σ 代数，试证明 $(\Omega,\mathcal{A},P)$ 为概率空间

3) 设 $\mathcal{A}$ 为 Ω 子集的 σ 代数，试证明 $(\Omega,\mathcal{A},P)$ 不为概率空间

解 1) 用 A_m 表示能被 m 整除的正整数，则 $A_m=\{jm\mid j=1,2,\cdots\}$ ，再设 $C_n=\{1,2,\cdots,n\}$，则可形式上定义比例

$$P(A_m)=\lim_{n\to\infty}\frac{\mathrm{Card}(A_m\cap C_n)}{\mathrm{Card}(C_n)}=\frac{1}{m}\tag{4.1}$$

为 A_m 在 Ω 中所占的比例

对于 $A,B\subset\Omega,P(\cdot)$ 满足加法公式 $P(A\cup B)=P(A)+P(B)-P(AB)$ 。下计算 $P\left(A_3A_5^c\left(A_4\cup A_6\right)\right)$。利用式 (4.1) 即得

$$\begin{aligned}
P\left(A_3A_5^c\left(A_4\cup A_6\right)\right)=&P\left(\left(A_3A_5^cA_4\right)\cup\left(A_3A_5^cA_6\right)\right)\\
=&P\left(A_3A_4A_5^c\right)+P\left(A_3A_6A_5^c\right)-P\left(A_3A_4A_6A_5^c\right)\\
=&P\left(A_3A_4\right)-P\left(A_3A_4A_5\right)+P\left(A_3A_6\right)-P\left(A_3A_6A_5\right)\\
&-P\left(A_3A_4A_6\right)+P\left(A_3A_4A_5A_6\right)\\
=&P\left(A_{12}\right)-P\left(A_{60}\right)+P\left(A_6\right)-P\left(A_{30}\right)-P\left(A_{12}\right)+P\left(A_{60}\right)\\
=&\frac{1}{6}-\frac{1}{30}=\frac{2}{15}
\end{aligned}$$

2) 由式 (4.1) 即得，对 $A,B\in\mathcal{A}$，有非负性 $P(A)\geqslant 0$ 和规范性 $P(\Omega)=1$。当 $AB=\varnothing,P(A+B)=P(A)+P(B)$ 时，由于 $\mathcal{A}$ 中仅有有限个元素，则 P 有可数可加性，则 $(\Omega,\mathcal{A},P)$ 为概率空间

3) 取 $\mathcal{A}$ 为 Ω 子集的全体，则由式 (4.1) 定义的 P 不为 $(\Omega,\mathcal{A})$ 上的概率测度。因为若 P 为概率测度，则对任意的 $\{j\}$ 有 $P(\{j\})=0$，再由概率测度的可数可加性即得

$$P(\Omega)=P\left(\bigcup_{j=1}^{\infty}\{j\}\right)=\sum_{j=1}^{\infty}P(\{j\})=0$$

这与 P 为概率测度矛盾

4.2 基本概型

Elementary Models of Probability

4.2.1 古典概型

Classical Model of Probability

我们把只涉及有限个事件概率的那一部分概率论称作初等概率论
Андре́й Никола́евич Колмого́ров　А. Н. 柯尔莫戈洛夫《概率论的基本概念》

关于古典概率的现存最早著作是 Cardano 的《机遇博弈》(The Book of Games of Chance)。该书约成于 1564 年，但在 Cardano 去世很久后的 1663 年得以发表，这时关于概率论的若干重要著作已经出版，削弱了该著作及其作者在概率史上的地位和影响。不过从书中可以看出，对于 Jakob Bernoulli 所谓 “哪怕最笨的人，不通过别人的教诲也能理解频率大约是二分之一”，即频率与概率的关系，Cardano 尚且没有认识，而且组合数学的知识也较为贫乏。

此外，1539 年，Cardano 还在另一本著作中提出了后世著名的分赌注问题 (5.15) 的一种解法（尽管并不完全正确）

例题 4.7 （有限概率空间）

设样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$ 为有限集，取 $\mathcal{A}$ 为 Ω 的 σ 代数。在每一个基本结果 $w_i(i=1,\cdots,N)$ 上定义函数 $p(\omega_i)$（或记为 p_i），若满足下列条件

1)（非负性）$0\leqslant p(\omega_i)\leqslant 1$

2)（规范性）$p(\omega_1)+\cdots+p(\omega_N)=1$

则定义事件 E 的**概率** (probability/вероятность)：

$$P(E)=\sum_{j:\omega_j\in E}p_j$$

不难验证 P 满足概率测度的性质 1)-性质 3)，则 P 为一个概率测度，则 $(\Omega,\mathcal{A},P)$ 为一个概率空间，称 $(\Omega,\mathcal{A},P)$ 为**有限概率空间** (finite probability space)

注 特别地，若对任意 $\{\omega_j\}$ 都有 $P(\{\omega_j\})=\dfrac{1}{n}$，则称 $(\Omega,\mathcal{A},P)$ 为**古典概率空间** (classical probability space) 或**古典概率模型** (classical probabilistic model)，也简称**古典概型**

注 若改设样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$ 为可数集，则改称该概率空间为**离散概率空间** (discrete probability space/дискретное вероятностное пространство)

例题 4.8 （古典概型）

考虑如下规则的抽彩票。假设总共有 M 张编号为 $1,2,\cdots,M$ 的彩票，其中编号为 $1,2,\cdots,n(M\geqslant 2n)$ 的中奖。一人买了 n 张彩票，求至少有一张中奖的概率（记为 P）

解 由于抽取彩票的顺序对于所购买的彩票是否中奖无影响，则认为样本空间为

$$\Omega=\{\omega=[a_1,\cdots,a_n]\,|(\forall k\neq l):a_k\neq a_l,a_i=1,\cdots,M\}$$

由计数原理知 $N(\Omega)=C_M^n$，下假设

$$A_0=\{\omega=[a_1,\cdots,a_n]\,|(\forall k\neq l):a_k\neq a_l,a_i=n+1,\cdots,M\}$$

为事件“所买彩票中没有中奖彩票”，仍由计数原理知 $N(A_0)=\mathrm{C}_{M-n}^n$，则

$$P(A_0)=\frac{C_{M-n}^n}{C_M^n}=\frac{(M-n)_n}{(M)_n}=\left(1-\frac{n}{M}\right)\left(1-\frac{n}{M-1}\right)\cdots\left(1-\frac{n}{M-n+1}\right)$$

即

$$P=1-P(A_0)=1-\left(1-\frac{n}{M}\right)\left(1-\frac{n}{M-1}\right)\cdots\left(1-\frac{n}{M-n+1}\right)$$

即为所需概率

例题 4.9 （De Mere **问题**）

1654 年 7 月 29 日，热爱赌博的法国贵族 De Mere 向当时杰出的科学家之一 Pascal 提出了这项问题。“两个骰子掷了 24 次。您可以至少赌一次两个 6 的出现，也可以赌投掷 4 次骰子时至少出现一个 6。”

De Mere 在投掷 4 次骰子时押注至少出现一个 6，并且通常赢的次数比输的次数多。他的推理如下：两个 6

出现在一对骰子上的概率为 $\frac{1}{36}$，比单次骰子出现一个 6 的概率（即 $\frac{1}{6}$）小 6 倍，但是现在有 6 倍的尝试，24 对 4。因此，结果应该相同。但游戏经验表明，事实并非如此。在 24 次掷两个骰子时，他输的比赢的多。

Blaise Pascal 和 Pierre Fermat 相互独立地找到了正确的解答。令 A_1 为在 4 次掷骰中至少得到一个 6 的概率，而 A_2 为在 24 次掷两个骰子中至少得到两个 6 的概率。实际上

$$P(A_1) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.518 > \frac{1}{2}$$

$$P(A_2) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.491 < \frac{1}{2}$$

这两个数字非常接近，因此，很难立即注意到得到了不同的结果。

注 在信件中，Pascal 当时给出的解法使用了所谓全概率公式 (4.3)，而 Fermat 的解法则使用了所谓概率乘法定理 (4.1)，从一个侧面也揭示了两人通信的历史意义

例题 4.10 （古典概型）一个笼子里关着 10 只猫，其中有 7 只白猫、3 只黑猫. 每次只能钻出 1 只猫，10 只猫都钻出了笼子，以 A_k 表示第 k 只出笼的猫是黑猫的事件，试求 $P(A_k), k = 1, 2, \cdots, 10$

解

* 方法一：排列数法 10 只猫出笼的先后顺序共有 10! 种不同可能，则 $|\Omega| = 10!$。在事件 A_k 中第 k 只出笼的猫是黑猫，有 $A_3^1 = 3$ 种可能，在其余 9 个位置上剩下的猫可任意排列，有 9! 种可能的顺序，则 $|A_k| = 3 \cdot 9!$，则有

$$P(A_k) = \frac{3 \cdot 9!}{10!} = \frac{3}{10}, \quad k = 1, 2, \cdots, 10$$

即为所求概率

* 方法二：组合数法

视为仅有两种不同的元素，一种有 7 个（7 只白猫），另一种有 3 个（3 只黑猫），由计数原理知共有 C_{10}^3 种可能的排列顺序，则 $|\Omega| = C_{10}^3 = 120$。在事件 A_k 中第 k 只出笼的猫是黑猫，在其余 9 个位置上有两个位置为黑猫，则 $C_9^2 = 36$，则有

$$P(A_k) = \frac{36}{120} = \frac{3}{10}, \quad k = 1, 2, \cdots, 10$$

即为所求概率

* 方法三：经典方法

由于仅需求出第 k 只出笼的猫是黑猫的概率，则仅需考虑前 k 只出笼的猫的排列情况。在 Ω 中包括了由 10 只不同的猫中选出 k 只猫的所有方式，则 $|\Omega| = \frac{10!}{(10-k)!}$。在事件 A_k 中第 k 只出笼的猫是黑猫，有 $C_3^1 = 3$ 种可能，在其前面的 $k-1$ 个位置上则是由其余 9 只不同的猫中选出 $k-1$ 只猫的所有可能方式，则 $|A_k| = \frac{3 \cdot 9!}{(10-k)!}$，则有

$$P(A_k) = \frac{3 \cdot 9!}{10!} = \frac{3}{10}, \quad k = 1, 2, \cdots, 10$$

即为所求概率

注 该例方法一和方法二表明，尽管采用了两种不同的计数模式计算 $|\Omega|$ 和 $|A_k|$，但计算出的概率值却是相等的。因此，概率计算问题不同于普通的计数问题。在概率计算问题中，重要的不是采用何种计数模式来计算样本空间和随机事件中的样本点的数目，而是要保证对两者采用同一种计数模式。

例题 4.11 （绳结问题/古典概型）有 n 根短绳，现把它们的 $2n$ 个端头两两任意连接。试求事件：n 根短绳恰结成 n 个圈的概率

解

方法一：排列数法以 A 表示上述事件。设把 $2n$ 个端头排成一行，然后规定将第 $2k-1$ 个端头与第 $2k$ 个端头相连，$k = 1, 2, \cdots, n$，则每一种排法对应一种连法，即得 $\Omega = (2n)!$。在事件 A 中每根短绳的两端自行连接，这等价于在 $2n$ 个端头的排列中，每根短绳的两端都相邻放置。因此可先对 n 根短绳进行排列，以确定各

根短绳的先后位置，再考虑每根短绳的两端的前后位置，即知 $|A| = n!(2!)^n$，则有

$$P(A) = \frac{n!2^n}{(2n)!} = \frac{1}{(2n-1)!!}$$

即为所求概率

方法二：组合数法

将 $2n$ 个端头分为 n 个无编号的组，每组两个端头，同一组内的两个端头连接，不同分组方式数量为

$$|\Omega| = \frac{(2n)!}{n!(2!)^n} = \frac{(2n)!}{n!(2)^n} = (2n-1)!!$$

在事件 A 发生时，每根短绳的两个端头都分在同一组，只有一种分法，即 $|A| = 1$，则有

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{(2n-1)!!}$$

即为所求概率

4.2.2 几何概型
Geometric Model of Probability

例题 4.12 （几何概型）

向长度等于 1 的线段上随机抛掷两个点，把线段分成长度分别为 x, y 和 z 的 3 段，试求可以用这 3 条线段为边构成三角形的概率

解 用向量 (x, y, z) 来表示结果，则所有可能结果的集合为

$$\Omega = \{(x, y, z) \mid x + y + z = 1, x \geqslant 0, y \geqslant 0, z \geqslant 0\}$$

其为平面 $x + y + z = 1$ 中的一个三角形，若用 A 表示 3 条线段可以形成三角形的事件，则当事件 A 发生时必有 $x + y > z, y + z > x, z + x > y$ ，则有

$$A = \{(x, y, z) \mid (x, y, z) \in \Omega, x + y > z, y + z > x, z + x > y\}$$

其也为平面 $x + y + z = 1$ 中的一个三角形，如图 (4.1)

试验者	抛硬币次数	出现正面次数	出现正面频率
Buffon	4040	2048	0.5069
De Morgan	4092	2048	0.5005
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

由于两个点随机抛掷，则样本点 (x, y, z) 在 Ω 中均匀分布，认为 $P(A)$ 应当等于三角形 A 的面积与三角形 Ω 的面积之比。分别用 $L(\Omega)$ 和 $L(A)$ 表示 Ω 与 A 的面积的缩写，则有

$$P(A) = \frac{L(A)}{L(\Omega)} = \frac{1}{4}$$

即为所求概率

注 最后的公式即为几何概型概率计算公式。当 Ω 与 A 为 n 维空间 $\mathbb{R}^n$ 中 Lebesgue 可测集时，$L(\Omega)$ 和 $L(A)$ 相应地表示它们的 n 维 Lebesgue 测度

例题 4.13 （几何概型）

甲、乙二人于某日下午 6 时至 7 时之间到达某处，每人都只在该处停留 10 分钟。试求他们可在该处相遇的概率

解 以 (x, y) 表示两人到达该处的时间，则易知 $\Omega = \{(x, y) \mid 6 \leqslant x \leqslant 7, 6 \leqslant y \leqslant 7\}$。若以 A 表示两人在该处相遇的事件，则有

$$A = \left\{(x, y) | (x, y) \in \Omega, |x - y| \leqslant \frac{1}{6}\right\}$$

可认为两人均随机地在 6 时与 7 时之间到达该处，则样本点 (x,y) 在 Ω 中均匀分布。显然，Ω 为一个边长为 1 的正方形，而 A^c 为直角边长为 $\frac{5}{6}$ 的两个直角等腰三角形之并如图 (4.2)

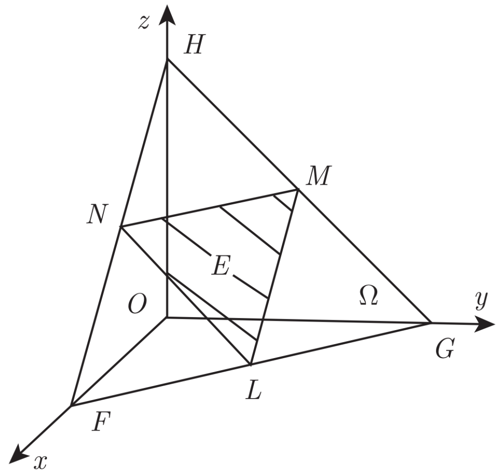

图 4.1: 例 (4.12) 示意图

则有 $P(A)=\dfrac{L(A)}{L(\Omega)}=\dfrac{11}{36}$

注 $L(A)=L(\Omega)-L\left(A^c\right)$ 从几何直观看显然，从概率论角度即为

$$P(A)=\frac{L(\Omega)-L\left(A^c\right)}{L(\Omega)}=1-P\left(A^c\right)$$

例题 4.14 （Buffon 投针问题/Buffon's needle problem）

平面上画满间距为 a 的平行直线，向该平面随机投掷一枚长度为 l 的针 $(l<a)$，试求针与直线相交的概率

解 以 E 表示针与直线相交的事件，如图 (4.3) 所示

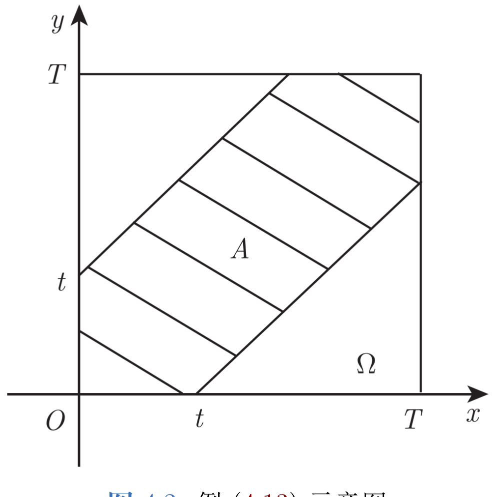

图 4.2: 例 (4.13) 示意图

则有针的位置可由它的中点到最近的直线的距离 ρ 以及它与直线的夹角 θ 决定，则

$$\Omega=\left\{(\rho,\theta)\mid 0\leqslant\rho\leqslant\frac{a}{2},0\leqslant\theta\leqslant\frac{\pi}{2}\right\}$$

而针与直线相交当且仅当 $\rho\leqslant\dfrac{l}{2}\sin\theta$，则有

$$E=\left\{(\rho,\theta)\mid(\rho,\theta)\in\Omega,\rho\leqslant\frac{l}{2}\sin\theta\right\}$$

由随机投掷则有样本点 (ρ, θ) 在 Ω 中均匀分布，则有

$$L(\Omega)=\frac{\pi a}{4}, \quad L(E)=\int_0^{\frac{\pi}{2}} \frac{l}{2}\sin\theta d\theta=\frac{l}{2}$$

则有 $P(E)=\dfrac{L(E)}{L(\Omega)}=\dfrac{2l}{\pi a}$

注 利用这例结果和“频率接近于概率”的规律可知，只要抛掷的次数 n 充分多，频率 $f_n(E)$ 会充分地接近于概率 $P(E)$。换言之，若在 n 次抛掷中针与直线相交了 m_n 次，则当 n 充分大时就有

$$\frac{m_n}{n} \approx \frac{2l}{\pi a}$$

即有 $\pi \approx \dfrac{2nl}{m_n a}$ ，则通过大量重复抛针就可估计 π 的近似值。称这样的方法为 Monte Carlo **方法** (Monte Carlo method)

例题 4.15 （几何概型）

试求下列三个问题

1)（零概率事件对不可能事件不充分性）往区间 $[0;1]$ 中随机抛掷一个质点，求该质点落在区间中点的概率

2)（零概率事件对不可能事件不充分性）往区间 $[0;1]$ 中随机抛掷一个质点，求质点落在区间中有理点上的概率

3)（一概率事件对确定事件不充分性）往区间 $[0;1]$ 中随机抛掷一个质点，求质点落在区间 $(0;1)$ 中的概率

解 1) 以 E 表示质点落在区间中点的事件，其中 $\Omega=[0;1], E=\left\{\frac{1}{2}\right\}$，则 $L(\Omega)=1, L(E)=0$，则有

$$P(E)=\frac{L(E)}{L(\Omega)}=0$$

2) 以 E 表示质点落在区间中有理点上的事件，其中 $\Omega=[0;1]$，E 为 $[0;1]$ 中的有理点集，则 $L(\Omega)=1, L(E)=0$，则有

$$P(E)=\frac{L(E)}{L(\Omega)}=0$$

3) 类似 1) 即得概率为 1

注 该例 1)2) 表明“往区间 $[0;1]$ 中随机抛掷一个质点，质点落在区间中点”和“往区间 $[0;1]$ 中随机抛掷一个质点，质点落在区间中有理点上”均为零概率事件，但这两个事件显然并非不可能事件

注 该例 3) 表明“往区间 $[0;1]$ 中随机抛掷一个质点，质点落在区间 $(0;1)$ 中”的概率为 1，但该事件显然并非确定事件

例题 4.16 （Bertrand 奇论）

在单位圆内任作一弦，试求弦长大于 $\sqrt{3}$ 的概率

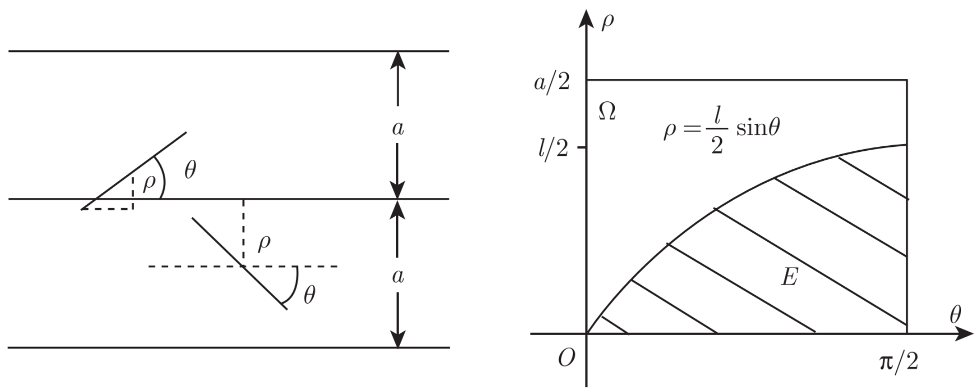

图 4.3: 例 (4.14) 示意图

解

* 解法一

以 E 表示弦长大于 $\sqrt{3}$ 的事件。不妨认为弦的一个端点 A 已经取定，问题变为在圆周上取另一端点 B，则 Ω 为整个圆周。由于单位圆的内接正三角形的边长等于 $\sqrt{3}$，则若以 A 为一个顶点作单位圆的内接正三角形 $\triangle AMN$，则当且仅当弦 AB 与边 MN 相交时弦 AB 的长度大于 $\sqrt{3}$，而此时端点 B 位于弧 $\overset{\frown}{MN}$ 上，则 E 即为弧 $\overset{\frown}{MN}$，则有

$$P(E)=\frac{L(\overset{\frown}{MN})}{L(\Omega)}=\frac{1}{3}$$

即为所求概率

* 解法二

在单位圆内取定一条直径 MN，仅考虑与 MN 垂直的弦 AB，此时弦 AB 的中点 K 必在该直径上，所以可把直径 MN 作为 Ω。易知当且仅当弦 AB 的中点 K 与圆心 O 的距离小于 $\frac{1}{2}$ 时弦 AB 的长度大于 $\sqrt{3}$，则有

$$E=\left\{K|K\in MN,|KO|<\frac{1}{2}\right\}$$

则 $L(\Omega)=|MN|=2,L(E)=1$，则有

$$P(E)=\frac{L(E)}{L(\Omega)}=\frac{1}{2}$$

即为所求概率

* 解法三

弦 AB 的长度仅由它的中点 K 的位置确定。易知当且仅当中点 K 位于以 $\frac{1}{2}$ 为半径的同心圆之内时弦 AB 的长度大于 $\sqrt{3}$，则 Ω 为整个单位圆，E 为以 $\frac{1}{2}$ 为半径的同心圆，L 为它们的面积，则有

$$P(E)=\frac{L(E)}{L(\Omega)}=\frac{1}{4}$$

即为所求概率

注 由于题目中“任作一弦”的定义不清晰，因此不同的理解产生了不同的问题，所以得到了不同的概率。该奇论由法国数学家 Bertrand 于 1889 年提出，该奇论的出现成为了促使概率论公理化体系诞生的原因之一

4.3 基本公式
Basic Formulas

4.3.1 条件概率与概率乘法公式
Conditional Probability and Product Theorem

关于条件概率，在 Bayes 的遗作 An essay towards solving a problem in the doctrine of chances 中 Bayes 从概率概念开始，把概率定义为对某种末知情况出现可能性大小的一个主观指标，这与在 Bayes 以前就流行的主观概率定义并无差别。接着 Bayes 对这种概率证明了几个命题，其中有关于条件概率的命题 3 和命题 5：

命题 3：设 E_1 和 E_2 是按时间先后的两个事件，则有

$$P\left(E_2\mid E_1\right)=\frac{P\left(E_1\cap E_2\right)}{P\left(E_1\right)}$$

命题 5：设 E_1 和 E_2 是按时间先后的两个事件，则有

$$P\left(E_1\mid E_2\right)=\frac{P\left(E_1\cap E_2\right)}{P\left(E_2\right)}$$

之所以分为两个命题，除了公理化的证明思路，由于事件 E_1 和 E_z 有先后时间顺序，两个命题一个是由过去 (E_1) 测未来，而另一个则是由未来反测过去，这也反映了 Bayes 的哲学理念

定义 4.12 (条件概率)

设 $(\Omega, \mathcal{A}, P)$ 为概率空间，$A, B \in \mathcal{A}$，其中 $P(A) > 0$，则称

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

为在已知事件 A 发生的情况下事件 B 发生的条件概率 (условная вероятность)

命题 4.1 (条件概率定义正确性)

设概率空间 $(\Omega, \mathcal{A}, P)$，$A, B \in \mathcal{A}$，其中 $P(A) > 0$，则

$$P(B \mid A) = \frac{P(AB)}{P(A)} \tag{4.2}$$

为概率测度

证明 非负性和规范性显然，仅需证明可数可加性。设 $\{B_n | n \in \mathbb{N}\}$ 为一列两两不交的事件，$A \in \mathcal{A}, P(A) > 0$，则 $\{AB_n | n \in \mathbb{N}\}$ 也为一列两两不交的事件，则有

$$P\left(A \bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} AB_n\right) = \sum_{n=1}^{\infty} P(AB_n)$$

上式两端同时除以 $P(A)$ 由式 (4.2) 即得

$$P\left(\bigcup_{n=1}^{\infty} B_n \mid A\right) = \sum_{n=1}^{\infty} P(B_n \mid A)$$

定理即证

注 值得一提的是，在式 (4.2) 中要求 $P(A) > 0$，是为了保证分母不为 0。在后文进一步拓展条件概率的概念，将式 (4.2) 变形为

$$P(AB) = P(A)P(B \mid A)$$

由对称性也有 $P(AB) = P(B)P(A \mid B)$

推论 4.1 (概率乘法定理/product theorem/теорема умножения)

设概率空间 $(\Omega, \mathcal{F}, P)$，$\{A_k, k = 1, 2, \cdots, n\} \subset \mathcal{F}$。若

$$P\left(\bigcap_{k=1}^{n} A_k\right) > 0$$

则有概率乘法公式：

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1A_2) \cdots P(A_n \mid A_1A_2 \cdots A_{n-1})$$

证明 记 $B = \bigcap_{k=1}^{n-1} A_k, A = A_n$，由条件概率定义即得

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P\left(\bigcap_{k=1}^{n-1} A_k\right) P\left(A_n \bigcap_{k=1}^{n-1} A_k\right)$$

归纳即证

例题 4.17 (Polyá[2] 罐子模型/Polya's urn scheme/概率乘法定理)

罐中放有 a 个白球和 b 个黑球，每次从罐中随机抽取一个球并连同 c 个同色球一起放回罐中，反复进行。

[2]波利亚 (George Polya，1887.12.13-1985.9.7) 美籍匈牙利数学家，法国科学院、美国国家科学院和匈牙利科学院院士。1963 年 Polya 获美国数学会杰出贡献奖。为了纪念 Polya，美国工业与应用数学学会设立了组合理论及其应用的 Polya 奖

证明在前 $n=n_1+n_2$ 次取球中取出 n_1 个白球和 n_2 个黑球的概率为

$$\binom{n}{n_1}\frac{a(a+c)(a+2c)\cdots(a+n_1c-c)\,b(b+c)(b+2c)\cdots(b+n_2c-c)}{(a+b)(a+b+c)(a+b+2c)\cdots(a+b+nc-c)}$$

解 (概率乘法定理) 以 A 表示 "在前 $n=n_1+n_2$ 次取球中, 取出了 n_1 个白球和 n_2 个黑球" 的事件。再以 A_k 表示在第 k 次取球时取出白球的事件, 则 A_k^c 即为在第 k 次取球时取出黑球的事件。由概率乘法定理 (4.1) 得

$$P\left(A_1\cdots A_{n_1}A_{n_1+1}^c\cdots A_n^c\right)=\frac{a(a+c)(a+2c)\cdots(a+n_1c-c)\,b(b+c)(b+2c)\cdots(b+n_2c-c)}{(a+b)(a+b+c)(a+b+2c)\cdots(a+b+nc-c)}$$

该概率为 "在前 n_1 次取球中取出了白球, 在接下来的 n_2 次取球中取出了黑球" 的事件的概率。而事件 A 为在 "前 n 次取球中有 n_1 次取出了白球, 其余 n_2 次取出了黑球" 的事件。因此需要考虑 "哪 n_1 次取出了白球, 哪 n_2 次取出了黑球", 共有 $C_n^{n_1}$ 种不同的可能情况, 各种情况所代表的事件互不相交, 则概率 $P(A)$ 等于这 $C_n^{n_1}$ 个事件的概率之和。对于每一种指定的情况, 由概率乘法定理 (4.1) 得

$$\frac{a(a+c)(a+2c)\cdots(a+n_1c-c)\,b(b+c)(b+2c)\cdots(b+n_2c-c)}{(a+b)(a+b+c)(a+b+2c)\cdots(a+b+nc-c)}$$

即都与 $P\left(A_1\cdots A_{n_1}A_{n_1+1}^c\cdots A_n^c\right)$ 相等, 则 $P(A)$ 等于 $P\left(A_1\cdots A_{n_1}A_{n_1+1}^c\cdots A_n^c\right)$ 的 $C_n^{n_1}$ 倍

注 该模型可以用来粗略描述流行病的传播规律

例题 4.18 (绳结问题/概率乘法定理)

有 n 根短绳, 现把它们的 $2n$ 个端头两两任意连接。试求事件: n 根短绳恰结成 n 个圈的概率

证明

方法三: 概率乘法定理

以 Ω 表示所有不同连接结果的集合, 设想把 $2n$ 个端头排成一行, 然后规定将第 $2k-1$ 个端头与第 $2k$ 个端头相连接, $k=1,2,\cdots,n$, 于是每一种排法对应一种连接结果, 则 $|\Omega|=(2n)!$。以 A 表示恰好连成 n 个圈的事件。设已将 n 根短绳作了编号, 以 A_k 表示第 k 号短绳被连成 1 个圈的事件, 则有 $A=\bigcap\limits_{k=1}^{n}A_k$

当 A_1 发生时, 1 号短绳被连成 1 个圈, 这等价于存在 $k\in\{1,2,\cdots,n\}$ 使得在 $2n$ 个端头的排列中, 1 号短绳的两个端头排在第 $2k-1$ 和第 $2k$ 个位置上, 则 $|A_1|=2n(2n-2)!$, 则

$$P\left(A_1\right)=\frac{|A_1|}{|\Omega|}=\frac{1}{2n-1}$$

下求 $P\left(A_2\mid A_1\right)$, 即要在已知 1 号短绳被连成 1 个圈的情况下, 求 2 号短绳也被连成 1 个圈的概率。由于 1 号短绳已经自成 1 个圈, 则只需对剩下的 $n-1$ 根短绳讨论其中的头一号短绳被连成 1 个圈的问题。由于现在的情况与原来的情况完全类似, 总绳数变为 $n-1$ 根, 则类比即知

$$P\left(A_2\mid A_1\right)=\frac{1}{2(n-1)-1}=\frac{1}{2n-3}$$

同理可得

$$P\left(A_k\mid A_1A_2\cdots A_{k-1}\right)=\frac{1}{2[n-(k-1)]-1}=\frac{1}{2n-2k+1},\quad k=3,4,\cdots,n$$

则由概率乘法定理 (4.1) 得

$$P(A)=P\left(\bigcap_{k=1}^{n}A_k\right)=\prod_{k=1}^{n}\frac{1}{2n-2k+1}=\frac{1}{(2n-1)!!}$$

即为所求概率

4.3.2 独立性 Independence

注 (独立性引入)

设两个事件 A 和 B, 若事件 A 的出现对事件 B 出现的概率不产生影响, 自然应该认为事件 B 不依赖于事件 A。换言之, 假设 $P(A)>0$, 若 $P(B\mid A)=P(B)$, 则称事件 B 对事件 A 独立。

由条件概率定义有 $P(B \mid A)=\dfrac{P(AB)}{P(A)}$，则有 $P(AB)=P(A)P(B)$。对称地，设 $P(B)>0$，若 $P(A \mid B)=P(A)$，自然称事件 A 对事件 B 独立。由此仍得到关系式 $P(AB)=P(A)P(B)$，则该式关于事件 A 和 B 对称且当事件 A 和 B 的概率（一个或两个）可能为 0 时仍然成立

定义 4.13 (独立)

设两个事件 A 和 B，若

$$P(AB)=P(A)P(B)$$

则称事件 A 和 B（关于概率 P）独立 (independent/независимы)（或统计独立）

设 Ω 上 σ 代数 $\mathcal{A}_1$ 和 $\mathcal{A}_2$，若对应属于 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 的两个任意子集 A_1 和 A_2 独立，则称 Ω 上 σ 代数 $\mathcal{A}_1$ 和 $\mathcal{A}_2$（关于概率 P）独立 (independent/независимы)（或统计独立）

♣

例题 4.19 （独立）

证明 σ 代数 $\mathcal{A}_1=\{A_1,\bar{A}_1,\varnothing,\Omega\}$ 和 $\mathcal{A}_2=\{A_2,\bar{A}_2,\varnothing,\Omega\}$ 独立当且仅当对任意事件 A_1 和 A_2 独立

解 必要性：$\mathcal{A}_1$ 和 $\mathcal{A}_2$ 独立说明，16 个事件序偶 A_1 和 A_2，A_1 和 $\bar{A}_2,\cdots,\Omega$ 和 Ω 独立。因此 A_1 和 A_2 独立。充分性：若 A_1 和 A_2 独立，则需证明其余 15 个事件序偶也独立。现验证 A_1 和 $\bar{A}_2$ 独立，即

$$\begin{aligned} P\left(A_1\bar{A}_2\right) &= P\left(A_1\right)-P\left(A_1A_2\right)=P\left(A_1\right)-P\left(A_1\right)P\left(A_2\right) \\ &= P\left(A_1\right)\left[1-P\left(A_2\right)\right]=P\left(A_1\right)P\left(\bar{A}_2\right) \end{aligned}$$

同理可证其余事件序偶独立

例题 4.20

（独立）1) 向区间 $[0;1)$ 中随机抛掷一个点，分别以 A_1 和 A_2 表示该点落在区间 $\left[0;\dfrac{1}{2}\right)$ 中和 $\left[\dfrac{1}{4};\dfrac{3}{4}\right)$ 中的事件。试问事件 A_1 与 A_2 是否相互独立?

2) 设样本空间 $\Omega=[0;1)$ 及其 Borel 子集上 σ 代数 $\mathcal{F}$，$E\in\mathcal{F}$，令

$$P(E)=\sum_{k:\frac{1}{2^k}\in E}\frac{1}{2^k}$$

亦即事件 E 的概率等于 E 中的所有形如 $\dfrac{1}{2^k}$ 的数的和。(不难验证, $(\Omega,\mathcal{F},P)$ 为一个概率空间) 以 A_1 表示事件 $\left[0;\dfrac{1}{2}\right)$，以 A_2 表示事件 $\left[\dfrac{1}{4};\dfrac{3}{4}\right)$。试问事件 A_1 与 A_2 是否相互独立?

解

1) 情形为几何概型，易知 $A_1A_2=\left[\dfrac{1}{4};\dfrac{1}{2}\right)$，则 $P\left(A_1A_2\right)=\dfrac{1}{4}$。而 $P\left(A_1\right)=P\left(A_2\right)=\dfrac{1}{2}$，则 $P(A_1A_2)=P(A_1)P(A_2)$ 成立，则 A_1 与 A_2 相互独立

2) 由定义的概率有

$$P\left(A_1\right)=P\left(\left[0;\frac{1}{2}\right)\right)=\sum_{k=2}^{\infty}\frac{1}{2^k}=\frac{1}{2};\quad P\left(A_2\right)=P\left(\left[\frac{1}{4},\frac{3}{4}\right)\right)=\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$$

但这时

$$P\left(A_1A_2\right)=P\left(\left[\frac{1}{4},\frac{1}{2}\right)\right)=\frac{1}{4}$$

则有 $P\left(A_1A_2\right)\neq P\left(A_1\right)P\left(A_2\right)$，则 A_1 与 A_2 不相互独立

例题 4.21 （独立）

设 $0\leqslant\alpha\leqslant\dfrac{\pi}{2}$，证明：$0\leqslant\sin\alpha+\cos\alpha-\sin\alpha\cos\alpha\leqslant 1$

解 由于 $0\leqslant\alpha\leqslant\dfrac{\pi}{2}$，则 $0\leqslant\sin\alpha,\cos\alpha\leqslant 1$，故可把它们视为概率。设事件 A 与事件 B 相互独立，且

$P(A)=\sin\alpha, P(B)=\cos\alpha$，则有

$$P(A\cup B)=P(A)+P(B)-P(A)P(B)=\sin\alpha+\cos\alpha-\sin\alpha\cos\alpha$$

由于 $0\leqslant P(A\cup B)\leqslant 1$，则 $0\leqslant \sin\alpha+\cos\alpha-\sin\alpha\cos\alpha\leqslant 1$

例题 4.22 （独立）

设 a,b,c 为三角形三边之长，有 $a+b+c=1$，证明

$$a^2+b^2+c^2+4abc\leqslant\frac{1}{2}$$

解 由三角形中任意两边的和都大于第三边，则有 $1=a+b+c>2a$，同理 $2b<1,2c<1$，则可将 $2a,2b,2c$ 视为概率。设 A,B,C 为三个相互独立的事件且 $P(A)=2a,P=2b,P(C)=2c$，则有

$$1\geqslant P(A\cup B\cup C)=2(a+b+c)-4(ab+bc+ca)+8abc=2-2\left(1-a^2-b^2-c^2\right)+8abc$$

整理后即得 $a^2+b^2+c^2+4abc\leqslant\dfrac{1}{2}$

性质 （相互独立性质）

若事件 A 与 B 相互独立，则有下列性质成立

1) A 与 B^c 相互独立
2) A^c 与 B 相互独立
3) A^c 与 B^c 相互独立

证明 1) 下证 A 与 B^c 相互独立，由全概率公式与事件 A 与 B 相互独立即得

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

2)3) 同理即得

定义 4.14 (全体独立)

设集合 $A_1,\cdots,A_n$，若对于任意 $k=1,2,\cdots,n$ 和 $1\leqslant i_1<i_2<\cdots<i_k\leqslant n$ 满足

$$P\left(A_{i_1},\cdots,A_{i_k}\right)=P\left(A_{i_1}\right)\cdots P\left(A_{i_k}\right)$$

则称 $A_1,\cdots,A_n$（关于概率 P）全体独立 (независимы в совокупности)（或互相统计独立）

若属于 $\mathcal{A}_1,\cdots,\mathcal{A}_n$ 的对应的集合 $A_1,\cdots,A_n$ 全体独立，则称 σ 代数 $\mathcal{A}_1,\cdots,\mathcal{A}_n$（关于概率 P）全体独立 (независимы в совокупности)（或全体统计独立）

♣

例题 4.23 （两两独立对全体独立的不充分性/С.Н.Бернштейна[3]例题）

正四面体涂有三种颜色：一种是蓝色的，第二种是红色的，第三种是绿色的，第四种是三种颜色的，抛起来并标出它从哪一侧掉下来。设 A_1 为蓝色，A_2 为红色，A_3 为绿色，则

$$P\left(A_1\right)=\frac{1}{2},P\left(A_2\right)=\frac{1}{2},P\left(A_3\right)=\frac{1}{2},P\left(A_1A_2\right)=P\left(A_2A_3\right)=P\left(A_1A_3\right)=P\left(A_1A_2A_3\right)=\frac{1}{4}$$

由此即有

$$P\left(A_1A_2\right)=\frac{1}{4}=\frac{1}{2}\cdot\frac{1}{2}=P\left(A_1\right)P\left(A_2\right)$$

同理即得两两独立，但是

$$P\left(A_1A_2A_3\right)=\frac{1}{4}\neq\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=P\left(A_1\right)P\left(A_2\right)P\left(A_3\right)$$

例题 4.24 （全体独立对两两独立的不充分性）

设样本空间 Ω 由有序数偶 (i,j) 组成，其中 $i,j=1,2,\cdots,6$，且这些数偶都等可能，则对于事件 $A=\{(i,j):$

[3]谢尔盖 · 纳塔诺维奇 · 伯恩施坦（Sergei Natanovich Bernstein，俄语：Серге́й Ната́нович Бернште́йн，1880.3.5 - 1968.10.26）苏联数学家。Bernstein 在 1904 年在巴黎大学上交的博士论文解决了椭圆微分方程的希尔伯特第十九问题。之后，他发表了许多涉及概率论、构造性功能理论以及基因学的数学基础的著作

$j=1,2$ 或 $5\}, B=\{(i,j): j=4,5$ 或 $6\}, C=\{(i,j): i+j=9\}$ 即有

$$P(AB)=\frac{1}{6} \neq \frac{1}{4}=P(A)P(B)$$
$$P(AC)=\frac{1}{36} \neq \frac{1}{18}=P(A)P(C)$$
$$P(BC)=\frac{1}{12} \neq \frac{1}{18}=P(B)P(C)$$

但这时有

$$P(ABC)=\frac{1}{26}=P(A)P(B)P(C)$$

即事件 A,B,C 全体独立，但非两两独立

性质 （全体独立性质）

设事件序列 $A_1, A_2, \cdots, A_n$ 全体独立，则下列命题成立：

1) 对于任意 $1 \leqslant j_1<j_2<\cdots<j_k \leqslant n$，都有 $A_{j_1}, A_{j_2}, \cdots, A_{j_k}$ 全体独立
2) 用 B_i 表示 A_i 或 $\bar{A}_i$，则 $B_1, B_2, \cdots, B_n$ 全体独立
3) $\left(A_1 \cap A_2\right), A_3, \cdots, A_n$ 全体独立
4) $\left(A_1 \cup A_2\right), A_3, \cdots, A_n$ 全体独立

定义 4.15 (有限概率空间直积)

设 n 个有限概率空间 $\left(\Omega_1, \mathcal{B}_1, P_1\right), \cdots,\left(\Omega_n, \mathcal{B}_n, P_n\right)$，组成点为 $\omega=\left(a_1, \cdots, a_n\right)$ 的空间 $\Omega=\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$，其中 $a_i \in \Omega_i$，记为

$$\mathcal{A}=\mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \cdots \otimes \mathcal{B}_n$$

$\mathcal{A}$ 为 Ω 上 σ 代数，由形如 $A=B_1 \times B_2 \times \cdots \times B_n\left(B_i \in \mathcal{B}_i\right)$ 的集合构成。对于 $\omega=\left(a_1, \cdots, a_n\right)$，令 $p(\omega)=p_1\left(a_1\right) \cdots p_n\left(a_n\right)$ 且把集合 $A=B_1 \times B_2 \times \cdots \times B_n$ 的概率定义为

$$P(A)=\sum_{\left\{a_1 \in B_1, \cdots, a_n \in B_n\right\}} p_1\left(a_1\right) \cdots p_n\left(a_n\right)$$

不难验证 $P(\Omega)=1$，因此 $(\Omega, \mathcal{A}, P)$ 决定某个概率空间，称该概率空间为有限概率空间

$$\left(\Omega_1, \mathcal{B}_1, P_1\right), \cdots,\left(\Omega_n, \mathcal{B}_n, P_n\right)$$

的直积 (direct product)

♣

性质 （有限概率空间直积独立性）

设 n 个有限概率空间 $\left(\Omega_1, \mathcal{B}_1, P_1\right), \cdots,\left(\Omega_n, \mathcal{B}_n, P_n\right)$，组成点为 $\omega=\left(a_1, \cdots, a_n\right)$ 的空间 $\Omega=\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$，其中 $a_i \in \Omega_i$，则有

1) 事件 $A_1=\left\{\omega: a_1 \in B_1\right\}, \cdots, A_n=\left\{\omega: a_n \in B_n\right\}$ （关于概率 P）全体独立，其中 $B_i \in \mathcal{B}_i$
2) 空间 Ω 上 σ 代数

$$\mathcal{B}_1=\left\{A_1: A_1=\left\{\omega: a_1 \in B_1\right\}, B_1 \in \mathcal{B}_1\right\}$$
$$\cdots\cdots\cdots\cdots$$
$$\mathcal{B}_n=\left\{A_n: A_n=\left\{\omega: a_n \in B_n\right\}, B_n \in \mathcal{B}_n\right\}$$

全体独立

4.3.3 单调性与 Borel-Cantelli 引理 Monotonicity and Borel-Cantelli Lemma

定义 4.16 (集列单调性)

若集族 $\{A_n|n\in\mathbb{N}^*\}$ 满足 $A_1\supset A_2\supset\cdots$，则称集族 A_n 单调下降，记为 $A_n\downarrow$；若集族 $\{A_n|n\in\mathbb{N}^*\}$ 满足 $A_1\subset A_2\subset\cdots$，则称集族 A_n 单调上升，记为 $A_n\uparrow$

♣

定义 4.17 (集列上极限和下极限)

对于集列 $\{A_j\}$，定义 $C_n=\bigcup\limits_{j=n}^{\infty}A_j, D_n=\bigcap\limits_{j=n}^{\infty}A_j$ ，称单调递减序列 $\{C_n\}$ 的极限

$$\bigcap_{n=1}^{\infty}C_n=\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty}A_j$$

为 $\{A_j\}$ 的上极限，可记为 $\overline{\lim\limits_{n\to\infty}}A_n$。

对称地，称单调递增序列 $\{D_n\}$ 的极限

$$\bigcup_{n=1}^{\infty}D_n=\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}A_j$$

为 $\{A_j\}$ 的下极限，可记为 $\underline{\lim\limits_{n\to\infty}}A_n$。

♣

注 (集列上极限和下极限)

上极限 $\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{j=n}^{\infty}A_j$ 发生当且仅当有无穷个 A_j 发生。下极限 $\bigcup\limits_{n=1}^{\infty}\bigcap\limits_{j=n}^{\infty}A_j$ 发生当且仅当至多有限个 A_j 不发生

性质

1) (集列上极限和下极限基本不等式)

对于集列 $\{A_j\}$ 有

$$\underline{\lim_{n\to\infty}}A_n=\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}A_j\subset\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty}A_j=\overline{\lim_{n\to\infty}}A_n$$

2) (集列上极限和下极限基本等式)

$$P\left(\overline{\lim_{n\to\infty}}A_n\right)=\lim_{n\to\infty}P\left(\bigcup_{j=n}^{\infty}A_j\right)$$

$$P\left(\underline{\lim_{n\to\infty}}A_n\right)=\lim_{n\to\infty}P\left(\bigcap_{j=n}^{\infty}A_j\right)$$

性质 (事件序列上极限和下极限基本等式)

对于事件序列 $\{A_j\}$ 有

$$P\left(\overline{\lim_{n\to\infty}}A_n\right)=\lim_{n\to\infty}P\left(\bigcup_{j=n}^{\infty}A_j\right)$$

$$P\left(\underline{\lim_{n\to\infty}}A_n\right)=\lim_{n\to\infty}P\left(\bigcap_{j=n}^{\infty}A_j\right)$$

证明 由概率的上连续性和下连续性即证

性质 (事件序列上极限性质)

设事件序列 $\{A_n\}$ 全体独立，则有下列命题成立且仅成立其中一个：

1) $P\left(\overline{\lim_{n\to\infty}} A_n\right)=1$
2) $P\left(\overline{\lim_{n\to\infty}} A_n\right)=0$

证明 当 $\sum\limits_{j=1}^{\infty} P\left(A_j\right)=\infty$，则 $P\left(\overline{\lim_{n\to\infty}} A_n\right)=1$；当 $\sum\limits_{j=1}^{\infty} P\left(A_j\right)$ 有限，则 $P\left(\overline{\lim_{n\to\infty}} A_n\right)=0$

定理 4.2 (Borel-Cantelli 引理)

设事件序列 $\{A_j\}$，则下列命题成立：

1) (Borel-Cantelli 第一引理) 若 $\sum\limits_{j=1}^{\infty} P\left(A_j\right)$ 有限，则有 $P\left(\overline{\lim_{n\to\infty}} A_n\right)=0$
2) (Borel-Cantelli 第二引理) 若 $\{A_j\}$ 两两独立，$\sum\limits_{j=1}^{\infty} P\left(A_j\right)=\infty$，则有 $P\left(\overline{\lim_{n\to\infty}} A_n\right)=1$

♡

证明

1) 由事件序列性质即得

$$P\left(\overline{\lim_{n\to\infty}} A_n\right) \leqslant \lim_{n\to\infty} \sum_{j=n}^{\infty} P\left(A_j\right)=0$$

由概率非负性即证

2) 用概率的上连续性和不等式 $1-|x| \leqslant \mathrm{e}^{-|x|}$ 即得

$$\begin{aligned} P\left(\bigcap_{j=n}^{m} A_j^c\right) &= \prod_{j=n}^{m} P\left(A_j^c\right)=\prod_{j=n}^{m}\left(1-P\left(A_j\right)\right) \leqslant \prod_{j=n}^{m} \exp\left(-P\left(A_j\right)\right) \\ &= \exp\left(-\sum_{j=n}^{m} P\left(A_j\right)\right) \xrightarrow[m\to\infty]{} 0 \end{aligned}$$

则由事件序列性质即知

$$P\left(\bigcup_{j=n}^{\infty} A_j\right)=\lim_{m\to\infty} P\left(\bigcup_{j=n}^{m} A_j\right)=\lim_{m\to\infty}\left(1-P\left(\bigcap_{j=n}^{m} A_j^c\right)\right)=1$$

命题即证

4.3.4 全概率公式
Formula of Total Probability

在 Pascal 与 Fermat 的通信中，已经不加证明地使用了全概率公式，另外 Huygens 在 1657 年出版的著作《机遇的规律》的命题中也使用了全概率公式

定义 4.18 (完备事件组)

设样本空间 Ω，事件 $A_1,\cdots,A_n\in\Omega$，若 $A_1,\cdots,A_n\in\Omega$ 为样本空间 Ω 的划分，亦即 $(\forall i,j\in\overline{1;n})$：$A_iA_j=\varnothing$ 且 $A_1+\cdots+A_n=\Omega$，则称 $A_1,\cdots,A_n$ 为完备事件组 (полная группа событий)

♣

定理 4.3 (全概率公式)

考虑样本空间 Ω 的某个完备事件组 $A_1,\cdots,A_n$ 且 $P\left(A_i\right)>0(i=1,2,\cdots,n)$，则有全概率公式 (Formula of Total Probability/формула полной вероятности)：

$$P(B)=\sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right)$$

♡

证明 由 $A_1, \cdots, A_n$ 为完备事件组，则有 $(AA_i) \cap (BA_j) = BA_iA_j = \varnothing$ 且

$$B = B \cap \Omega = B\left(A_1 + \ldots + A_n\right) = BA_1 + \ldots + BA_n$$

由此即得

$$P(B) = \sum_{i=1}^{n} P\left(BA_i\right) = \sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right)$$

定理即证

注 特别地，若 $0 < P(A) < 1$，完备事件组退化为仅有两个事件 A, A^c，则有

$$P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$$

例题 4.25 （全概率公式）

甲、乙二人抛掷一枚均匀的硬币，甲抛了 n 次，乙抛了 $n+1$ 次。求乙抛出的正面次数比甲多的概率

解 若甲和乙都抛掷这枚均匀的硬币 n 次，以 A_0 表示事件“甲乙抛出的正面次数一样多”，A_1 表示事件“甲抛出的正面次数比乙多”，A_2 表示事件“乙抛出的正面次数比甲多”。这时有 $P\left(A_0\right) + P\left(A_1\right) + P\left(A_2\right) = 1$，而由对称性有 $P\left(A_1\right) = P\left(A_2\right)$。

以 E 表示问题中，乙抛出的正面次数比甲多的事件，再以 B 表示乙第一次抛掷时抛出的是正面的事件，则 B^c 为乙第一次抛掷时抛出的是反面的事件。显然，若 B 发生，则乙只要在接下来的 n 次抛掷中，抛出的正面次数不比甲少，则他所抛出的正面总次数就比甲多；而若 B^c 发生，则乙需要在接下来的 n 次抛掷中，抛出的正面次数比甲多，亦即

$$P(E \mid B) = P\left(A_0\right) + P\left(A_2\right), \quad P\left(E \mid B^c\right) = P\left(A_2\right) = P\left(A_1\right)$$

则由全概率公式知

$$P(E) = P(B)P(E \mid B) + P\left(B^c\right) P\left(E \mid B^c\right) = \frac{1}{2}\left(P\left(A_0\right) + P\left(A_1\right) + P\left(A_2\right)\right) = \frac{1}{2}$$

即为乙抛出的正面次数比甲多的概率

例题 4.26 （敏感问题调查/全概率公式）

在调查家庭暴力（或婚外恋、服用兴奋剂、吸毒等敏感问题）所占家庭的比例 p 时，被调查者往往不愿回答真相，这使得调查数据失真。为得到实际的 p 同时又不侵犯个人隐私，调查人员将袋中放入比例为 p_0 的红球和比例为 $q_0 = 1 - p_0$ 的白球。被调查者在袋中任取一球窥视后放回，并承诺取得红球就讲真话，取到白球就讲假话。被调查者只需在匿名调查表中选“是”（有家庭暴力）或“否”，然后将表放入投票箱。若声称有家庭暴力的家庭比例为 p_1，试求 p

解 对任选的一个家庭，用 B 表示回答“是”，用 A 表示实际“是”，由全概率公式即得

$$\begin{aligned} p_1 = P(B) &= P(B \mid A)P(A) + P(B \mid A^c)P(A^c) \\ &= p_0 P(A) + q_0(1 - P(A)) = q_0 + \left(p_0 - q_0\right) P(A) \end{aligned}$$

则只要 $p_0 \neq q_0$ 就有

$$p = P(A) = \frac{p_1 - q_0}{p_0 - q_0}$$

假定调查了 n 个家庭，其中有 k 个家庭回答“是”，则可以用 $\hat{p}_1 = \dfrac{k}{n}$ 估计 p_1，则可用

$$\hat{p} = \frac{\hat{p}_1 - q_0}{p_0 - q_0}$$

估计 p

注 可以证明 $|p_0 - q_0|$ 越大，结论越可靠。但 $|p_0 - q_0|$ 越大，调查方案也越不易被被调查者接受

例题 4.27 （秘书问题/全概率公式）

某公司需招收秘书一名，共有 n 个人报名应聘，公司希望从中挑选出最佳人选，要对他们逐个面试。面试的当时就需对应聘者表态是否录用，一旦对某人表态不录用便不可反悔。公司打算按照如下的策略行事：不录用前 $k(1 \leqslant k < n)$ 个面试者，自第 $k+1$ 个开始，只要发现某人比他前面的所有面试者都好就录用他，否则就

录用最后一个。试对该公司的策略作概率分析。

解 以 A 表示最佳人选被录用的事件，以 B_j 表示最佳人选在面试顺序中排在第 j 个，则

$$P\left(B_j\right)=\frac{(n-1)!}{n!}=\frac{1}{n}, \quad j=1, \cdots, n$$

即最佳人选出现在各个位置上的概率相等。以给定所在位置的序号作为条件有

$$P(A)=\sum_{j=1}^n P\left(A \mid B_j\right) P\left(B_j\right)=\frac{1}{n} \sum_{j=1}^n P\left(A \mid B_j\right)$$

当 $1 \leqslant j \leqslant k$ 时，最佳人选位于前 k 个面试者之中，不会被录用，此时 $P\left(A \mid B_j\right)=0$。当 $k+1 \leqslant j \leqslant n$ 时，最佳人选被录用，当且仅当前 $j-1$ 个面试者中的最佳者在前 k 个人中，故则 $P\left(A \mid B_j\right)=\dfrac{k}{j-1}$。由全概率公式知

$$P(A)=\frac{1}{n} \sum_{j=k+1}^n \frac{k}{j-1}=\frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j} \sim \frac{k}{n} \ln \frac{n}{k}$$

现在考虑可导函数 $g(x)=\dfrac{x}{n} \ln \dfrac{n}{x}, x>0$，则有 $g^{\prime}(x)=\dfrac{1}{n} \ln \dfrac{n}{x}-\dfrac{1}{n}$，这时有

$$g^{\prime}(x)=0 \leftrightarrow x=\frac{n}{\mathrm{e}}$$

既然 $P(A) \sim g(k)$，故知若将 k 取为最靠近 $\dfrac{n}{\mathrm{e}}$ 的正整数时，可使 $P(A)$ 达到最大，由于 $g\left(\dfrac{n}{\mathrm{e}}\right)=\dfrac{1}{\mathrm{e}} \approx 0.36788$，该概率最大值与 n 无关。这表明即使对很大的 n 采用所说的策略，也可以有 0.36788 的概率录用到最佳人选

例题 4.28 （基因遗传问题/全概率公式）

设群体中每个个体带且仅带有以下三种基因之一：AA, aa, Aa。任意两个个体随机相遇后，各自以 $\dfrac{1}{2}$ 的概率将自己基因的一半遗传给后代（例如若基因是 BD，则等可能提供基因 B 异或 D）。群体中这三种基因比例为

$$u: w: 2 v, \quad u v w>0, \quad u+w+2 v=1 .$$

设群体数目充分大，求“子”一代群体中这三种基因的比例。

解 由题意，“父”“母”按下列方式将基因遗传给后代：

$$\left\{\begin{array}{l}
P(S A A \mid F A A, M A A)=1, \\
P(S A A \mid F A A, M A a)=1 / 2, \\
P(S A A \mid F A a, M A A)=1 / 2, \\
P(S A A \mid F A a, M A a)=1 / 4 .
\end{array}\right.$$

其中 S 表示“子”，F 表示“父”，M 表示“母”。由于群体充分大，两个个体随机相遇等同于有放回地任选两个个体，则有

$$\left\{\begin{array}{l}
P(F A A, M A A)=P(F A A) P(M A A)=u^2, \\
P(F A A, M A a)=P(F A A) P(M A a)=2 u v, \\
P(F A a, M A A)=P(F A a) P(M A A)=2 u v \\
P(F A a, M A a)=P(F A a) P(M A a)=4 v^2 .
\end{array}\right.$$

用 $u_1: w_1: 2 v_1$ 表示“子”一代群体中基因的比例，再利用全概率公式即得

$$\begin{aligned}
u_1=P(S A A)= & P(S A A \mid F A A, M A A) P(F A A, M A A) \\
& +P(S A A \mid F A A, M A a) P(F A A, M A a) \\
& +P(S A A \mid F A a, M A A) P(F A a, M A A) \\
& +P(S A A \mid F A a, M A a) P(F A a, M A a) \\
= & u^2+u v+u v+v^2 \\
= & (u+v)^2 .
\end{aligned}$$

对称即得 $w_1 = P(Saa) = (w+v)^2$ 。再由 $\sqrt{u_1}+\sqrt{w_1} = u+2v+w = 1$ 即得

$$2v_1 = P(SAa) = 1-u_1-w_1 = (\sqrt{u_1}+\sqrt{w_1})^2 - u_1 - w_1 = 2\sqrt{u_1}\sqrt{w_1}$$

则“子”一代中这三种基因的比例是 $u_1 : w_1 : 2\sqrt{u_1 w_1}$。综上即得，“子”二代中三种基因的比例为

$$\begin{cases} u_2 = \left(u_1+\sqrt{u_1}\sqrt{w_1}\right)^2 = u_1, \\ w_2 = \left(w_1+\sqrt{u_1}\sqrt{w_1}\right)^2 = w_1, \\ 2v_2 = 2\sqrt{u_2 w_2} = 2v_1. \end{cases}$$

亦即只要群体充分大，从第二代开始，每个基因都将按固定比例永远遗传

例题 4.29 （Polyá 罐子模型/全概率公式）

罐中放有 a 个白球和 b 个黑球，每次从罐中随机抽取一个球并连同 c 个同色球一起放回罐中，反复进行。试证明：在第 n 次取球时取出白球的概率为 $\dfrac{a}{a+b}$

解 （全概率公式）以 A_k 表示在第 k 次取球时取出白球的事件，则 A_k^c 为在第 k 次取球时取出黑球的事件。对 n 归纳：显然有 $P(A_1) = \dfrac{a}{a+b}$。假设 $n=k-1, k\geqslant 2$ 时结论成立，证结论对 $n=k$ 也成立。

以 A_1 和 A_1^c 作为对 Ω 的一个划分，此时可将 $P(A_k \mid A_1)$ 看成从原来放有 $a+c$ 个白球和 b 个黑球的罐中按规则取球并且在第 $k-1$ 次取球时取出白球的概率，则由归纳假设有 $P(A_k \mid A_1) = \dfrac{a+c}{a+b+c}$，同理亦有 $P(A_k \mid A_1^c) = \dfrac{a}{a+b+c}$，则由全概率公式得

$$\begin{aligned} P(A_k) &= P(A_1)P(A_k \mid A_1) + P(A_1^c)P(A_k \mid A_1^c) \\ &= \frac{a}{a+b}\cdot\frac{a+c}{a+b+c} + \frac{b}{a+b}\cdot\frac{a}{a+b+c} = \frac{a}{a+b} \end{aligned}$$

则结论对一切 n 成立

例题 4.30 （全概率公式）

甲、乙二人进行某项体育比赛，每回合胜者得 1 分，败者不得分。比赛进行到有一人比另外一个人多 2 分就终止，多 2 分者获胜。现知每回合甲胜概率为 p，乙胜概率为 $1-p$，其中 $0<p<1$。试求甲最终获胜的概率

解 方法一：经典方法

以 A 表示甲最终获胜的事件，甲只能在偶数个回合比赛后获胜，以 A_{2n} 表示甲在 $2n$ 个回合比赛后获胜的事件，则 $P(A_{2n}) = (2p(1-p))^{n-1}p^2$，则

$$P(A) = \sum_{n=1}^{\infty} P(A_{2n}) = p^2\sum_{n=1}^{\infty}(2p(1-p))^{n-1} = p^2\sum_{n=0}^{\infty}(2p(1-p))^n = \frac{p^2}{1-2p(1-p)}$$

即为甲最终获胜概率

解 方法二：全概率公式

以 A 表示甲最终获胜的事件，并记 $P_1 = P(A)$。考虑前两个回合，分别以 B_1, B_2, B_3 表示甲二胜、一胜一败、二败的事件，则有

$$P(B_1) = p^2, \quad P(B_2) = 2p(1-p), \quad P(B_3) = (1-p)^2$$

$$P(A \mid B_1) = 1, \quad P(A \mid B_2) = P_1, \quad P(A \mid B_3) = 0$$

则由全概率公式得

$$P_1 = P(A) = \sum_{i=1}^{3} P(B_i)P(A \mid B_i) = p^2 + 2p(1-p)P_1$$

解得 $P(A) = P_1 = \dfrac{p^2}{1-2p(1-p)}$

解 方法三：随机游动

考察质点在数轴整点上的随机游动。每一步由所在整点 $x=n$ 以概率 p 向右移动到整点 $x=n+1$，以概率 $1-p$ 向左移动到整点 $x=n-1$。质点由原点 $x=0$ 出发，一旦到达 $x=2$ 或 $x=-2$ 便停止不动。若以 p_n 表示质点由整点 $x=n$ 出发能够到达 $x=2$ 的概率，则 p_0 即为甲取胜的概率。显然，$p_{-2}=0, p_2=1$，而由全

概率公式即得

$$\begin{cases} p_0 = pp_1 + (1-p)p_{-1} \\ p_1 = pp_2 + (1-p)p_0 = p + (1-p)p_0 \\ p_{-1} = pp_0 + (1-p)p_{-2} = pp_0 \end{cases}$$

把后面两式代入第一式即得

$$p_0 = p\left(p + (1-p)p_0\right) + (1-p)\left(pp_0\right), \quad p_0 = p^2 + 2p(1-p)p_0$$

解得 $p_0 = \dfrac{p^2}{1-2p(1-p)}$

例题 4.31 （Simpson 悖论/全概率公式）

有两种治疗肾结石的方案。在接受方案一治疗的患者中，小结石患者占 25%，大结石患者占 75%，小结石患者的治愈率为 93%，大结石患者的治愈率为 73%。在接受方案二治疗的患者中，小结石患者占 77%，大结石患者占 23%，小结石患者的治愈率为 87%，大结石患者的治愈率为 69%。试问：方案一优于方案二吗?

解 以 A 表示患者为小结石患者的事件，以 B 表示被治愈的事件。对于方案一，由全概率公式有

$$P(B) = P(A)P(B \mid A) + P\left(A^c\right)P\left(B \mid A^c\right) = 0.25 \cdot 0.93 + 0.75 \cdot 0.73 = 0.78$$

对于方案二，由全概率公式有

$$P(B) = P(A)P(B \mid A) + P\left(A^c\right)P\left(B \mid A^c\right) = 0.77 \cdot 0.87 + 0.23 \cdot 0.69 \approx 0.83$$

则方案二的治愈率高于方案一，可见方案一并不优于方案二

注 这类问题在统计界被称为 Simpson **悖论** (Simpson's paradox)，其表明光凭直觉难以直接作出判断。尽管所看到的统计数据真实，但它们只是对各个方案分类统计出来的数据，仅为条件概率。

例题 4.32 （绳结问题/全概率公式）

有 n 根短绳，现把它们的 $2n$ 个端头两两任意连接。试求事件：n 根短绳恰结成 n 个圈的概率

解

方法四：全概率公式

以 A_n 表示 n 根短绳恰好连成 n 个圈的事件，记 $p_n = P\left(A_n\right)$，再以 B 表示第 1 根短绳连成 1 个圈的事件，用 B 和 B^c 作为对 Ω 的一个划分，则由全概率公式得

$$p_n = P\left(A_n\right) = P(B)P\left(A_n \mid B\right) + P\left(B^c\right)P\left(A_n \mid B^c\right)$$

在方法三中已求得 $P(B) = \dfrac{1}{2n-1}$，易知 $P\left(A_n \mid B^c\right) = 0$，而 $P\left(A_n \mid B\right)$ 则是在已知第 1 根短绳连成 1 个圈的条件下，其余 $n-1$ 根短绳连成 $n-1$ 个圈的概率，此时第 1 根短绳已经与其余 $n-1$ 根短绳无关，则 $P\left(A_n \mid B\right) = P\left(A_{n-1}\right) = p_{n-1}$，代入全概率公式即得

$$p_n = P\left(A_n\right) = \frac{1}{2n-1}p_{n-1}, \quad n = 2, 3, \cdots$$

反复利用该式，并注意 $p_1 = 1$ 即得 $p_n = \dfrac{1}{(2n-1)!!}$，$\quad n = 1, 2, \cdots$

注 全概率公式是概率论前期发展中的一个重要里程碑，其意义和价值远远超出了时间的局限。其要点是在 Ω 中引入一个适当的划分，把概率条件化以达到化难为易的目的。这就为利用递推方法解答概率问题提供了途径

4.3.5 Bayes 公式与 Bayes 定理
Bayes Formula and Bayes Theorem

> 推广到极处可以说：
> “正概率”是由原因推结果，是概率论；
> “逆概率”是由结果推原因，是数理统计。
>
> 陈希孺《数理统计学简史》

Bayes 的遗作 An essay towards solving a problem in the doctrine of chances 的前言中明确提出了该文要讨论的核心问题：给定了一个事件在一系列观察中出现的次数和不出现的次数，并给定两个数，求该事件在一次观察中出现的概率 θ 落在此两数之间的机遇。

假设某个过程具有 $A_1, A_2, \cdots, A_n$ 这样 n 个可能的前提，而 $\{P(A_k), k=1,2,\cdots,n\}$ 为人们对这 n 个可能前提的可能性大小的一种事前估计，称其为先验概率。当该过程有了一个结果 B 之后，便会通过条件概率 $\{P(A_k \mid B), k=1,2,\cdots,n\}$ 来对这 n 个可能前提的可能性大小作出一种新的认识，因此称这些条件概率为后验概率。而后世所谓 Bayes 公式恰恰提供了一种计算后验概率的工具。后世从这种先验概率和后验概率的理念中发展出了一整套统计理论和方法，并形成了概率统计中的 Bayes 学派。

英国数学家 Harold Jeffreys[a]在 1939 年出版的《概率论》是 Bayes 学派的经典著作，在 Bayes 理论的复兴中起到了重要作用。，近代对于 Bayes 理论去除糟粕后发展出经验 Bayes 方法，由于计算机强大的运算功能和快速的运算能力，Bayes 方法和经验 Bayes 方法的实用价值大大提高。在人工智能的研究和开发中，由于 Bayes 理论对于智能机器人的深度学习具有重要的指导意义，Bayes 统计学迎来了前所未有的机会。

[a]哈罗德 · 杰弗里斯 (Harold Jeffreys, 1891.4.22-1989.3.18) 英国数学家，统计学家，地球物理学家与天文学家，1937 年或皇家天文学会金奖，1948 年获 Lagrange 奖，1960 年获皇家学会 Copley 奖，1962 年获皇家统计学会金奖。主要数学贡献是概率的 Bayes 方法，以及经常被忽略的关于近似解线性二阶微分方程一种解法（常称为 WKB 方法）

定理 4.4 (Bayes 公式)

设事件 A 和 B 的概率大于 0，即 $P(A)>0, P(B)>0$，则有

$$P(A \mid B)=\frac{P(A) P(B \mid A)}{P(B)}$$

♡

证明 由条件概率有 $P(AB)=P(A \mid B)P(B), P(AB)=P(B)P(A \mid B)$ ，联立即证

注 称该公式为 Bayes **公式** (формула Байеса)，其为 Beyes 定理 (4.5) 的二维形式

定理 4.5 (Bayes 定理)

设事件组 $A_1, \cdots, A_n$ 为 Ω 的一个划分，B 的概率大于 0，则有

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

♡

证明 由全概率公式与 Bayes 公式即证

注 统计学中，称事件组 $A_1, \cdots, A_n$ 为**假设**，称 $P(A_i)$ 为假设 A_i 的**先验概率** (prior probability/априорная вероятность)，称 $P(A_i \mid B)$ 为假设 A_i 在事件 B 出现后的**后验概率** (posterior probability/апостериорная вероятность)

例题 4.33 （Bayes 定理）

甲、乙二人之间经常用 E-mail 相互联系，他们约定在收到对方信件的当天即给回信。由于线路问题，每 n 份 E-mail 中会有 1 份不能在当天送达收件人。甲在某日发了 1 份 E-mail 给乙，但未在当天收到乙的回音，试求乙在当天收到了甲发给他的 E-mail 的概率

解 该问题中，甲发给乙的 E-mail 不一定能在当天到达乙处，乙所回的 E-mail 不一定能在当天到达甲处。而乙是否回 E-mail，则完全取决于他是否收到了甲发来的 E-mail。

以 A 表示乙在当天收到了甲发给他的 E-mail 的事件，以 B 表示甲在当天收到了乙回给他的 E-mail 的事件，下求条件概率 $P(A \mid B^c)$。显然 A 和 A^c 为一个 Ω 的划分，由条件知

$$P(A)=\frac{n-1}{n}, \quad P(B^c \mid A)=\frac{1}{n}, \quad P(B^c \mid A^c)=1$$

则由 Bayes 定理 (4.5) 得

$$P(A \mid B^c) = \frac{P(A)P(B^c \mid A)}{P(A)P(B^c \mid A) + P(A^c)P(B^c \mid A^c)} = \frac{\frac{n-1}{n} \cdot \frac{1}{n}}{\frac{n-1}{n} \cdot \frac{1}{n} + \frac{1}{n} \cdot 1} = \frac{n-1}{2n-1}$$

即为乙在当天收到了甲发给他的 E-mail 的概率

注 该例表明，Bayes 定理可以用来解决"已知结果，分析原因"的问题。甲在当天没有收到乙回给他的 E-mail，这是已经发生的结果，则通过计算条件概率 $P(A \mid B^c)$ 估计乙收到了他的 E-mail 的可能性为 $P(A \mid B^c) = \frac{n-1}{2n-1} < \frac{1}{2}$

例题 4.34 (Bayes 定理/先验概率/后验概率)

对以往数据分析结果表明，当机器调整得良好时，产品的合格率为 98%，而当机器发生某种故障时其合格率为 55%。每天早上机器开动时，机器调整良好的概率为 95%。试求已知某日早上第一件产品是合格品时，机器调整良好的概率为多少?

解 设 A 为事件"产品合格"，B 为事件"机器调整良好"。已知

$$P(A \mid B) = 0.98, P(A \mid B^c) = 0.55, P(B) = 0.95, P(B^c) = 0.05$$

需求概率为 $P(B \mid A)$，则由 Bayes 定理 (4.5) 即有

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)} = \frac{0.98 \times 0.95}{0.98 \times 0.95 + 0.55 \times 0.05} = 0.97$$

则当生产出第一件产品是合格品时，此时机器调整良好的概率为 0.97

注 该例中概率 0.95 为由以往的数据分析得到的，即为先验概率。而在得到信息，即生产出的第一件产品是合格品，之后再重新加以修正的概率，即 0.97 为后验概率

例题 4.35 (Bayes 定理)

某种疾病的患病率为 0.5%，通过验血诊断该病的误诊率为 5% (即非患者中有 5% 的人验血结果为阳性，患者中有 5% 的人验血结果为阴性)。现知某人验血结果为阳性，试求他确患有此病的概率

解 以 A 表示患有此病的事件，以 B 表示验血结果为阳性的事件。求条件概率 $P(A \mid B)$。由已知条件知

$$P(A) = 0.5\%, \quad P(B \mid A) = 95\%, \quad P(B \mid A^c) = 5\%$$

注意到 A 和 A^c 为事件空间 Ω 的一个划分，则由 Bayes 定理 (4.5) 得

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^c)P(B \mid A^c)} = \frac{0.5\% \times 95\%}{0.5\% \times 95\% + 99.5\% \times 5\%} \approx 0.087$$

即为他确患有此病的概率

注 该结果较小的原因在于人群中该病的患病率很低，仅为 0.5%，所以尽管通过验血诊断该病的误诊率仅为 5%，但与患病率相比为 10 倍，从而 Bayes 公式中分母几乎为分子的 11 倍。该例表明，当验血结果为阳性时，确患有此病的概率并不一定就很大，因为患病的概率除了依赖于验血时的准确率之外，还与人群中该病的患病率有关——而这一点对于罕见病的诊断尤为重要。

第 5 章 随机变量 (Random Variable)

6

随机变量的概念最先在 19 世纪下半叶由俄罗斯圣彼得堡学派引入——这一概念的引入标志着概率论已经推进到了现代化的门槛。对随机变量概念的严格表述，则是 20 世纪 30 年代，在概率论的公理化体系建立以后才得以完成。

5.1 基本概念

Basic concept

5.1.1 随机变量

Random Variable

定义 5.1 (随机变量第一定义)

设概率空间 $(\Omega, \mathcal{A}, P)$，函数 $\xi : \Omega \to \mathbb{R}^1$，且对于任意 $x \in R$ 都有 $(\xi \leqslant x)$ 为随机事件 (random event)，亦即

$$(\xi \leqslant x) = \{\omega | \xi(\omega) \leqslant x\} \in \mathcal{A} \tag{5.1}$$

则称函数 ξ 为随机变量 (random variable/случайная величина)

♣

注 （随机变量第二定义引入）

当 ξ 为离散型随机变量时，其一共只可能取至多可数个值。而若其一共只能取有限个可能值 $\{a_1, a_2, \cdots, a_n\}$，则对于其中每一个 a_j 都有 $(\xi = a_j) = \{\omega \mid X(\omega) = a_j\}$ 为事件。这时显然 $(\xi = a_1), (\xi = a_2), \cdots, (\xi = a_n)$ 为 Ω 的划分，并且有

$$(\forall x \in \mathbb{R}) : (\xi \leqslant x) = \bigcup_{j \leqslant a_j \leqslant x} (\xi = a_j) \tag{5.2}$$

而当 X 可取可数个值 $\{a_n | n \in \mathbb{N}\}$ 时 $\{(\xi = a_n) | n \in \mathbb{N}\}$ 为 Ω 的划分，且式 (5.2) 成立。则对于离散型随机变量 ξ，式 (5.1) 等价于对 ξ 的任何一个可能值 a 都有

$$(\xi = a) = \{\omega \mid \xi(\omega) = a\} \in \mathcal{A}$$

对于非离散型随机变量，上述定义说明了随机变量即为定义在概率空间上的 Borel 可测函数

命题 5.1 (随机变量对 Borel 可测函数保构性)

若 $\xi : \Omega \to R^1$ 为一个随机变量并且 $g : R^1 \to R^1$ 为一个 Borel 可测函数，则 $\eta = g(\xi)$ 也为一个随机变量

♠

证明 由 $\Omega \xrightarrow{\xi} R^1 \xrightarrow{g} R^1$ 即有 $\eta^{-1}(B) = \xi^{-1}\left(g^{-1}(B)\right)$ 为任意 $B \in \mathcal{B}$ 的随机事件

注 实际上有

$$P_\eta(B) = P(\eta \in B) = P(g(\xi) \in B) = P\left(\xi \in g^{-1}(B)\right) = P_\xi\left(g^{-1}(B)\right)$$

即 $P_\eta(B) = P_\xi\left(g^{-1}(B)\right)$

定义 5.2 (随机变量第二定义)

设概率空间 $(\Omega, \mathcal{A}, P)$，而图 $(\mathbb{R}^1, \mathcal{B})$ 为样本空间 Ω 子集 σ 代数上定义的实值函数，亦即函数 $\xi : \Omega \to \mathbb{R}^1$，其满足

$$(\forall B \in \mathcal{B}) : \left[\left(\xi^{-1}(B) := \{\omega \in \Omega | \xi(\omega) \in B\}\right) \in \mathcal{A}\right]$$

则称函数 ξ 为随机变量 (random variable/случайная величина)。为简单起见，经常也直接称定义在（有限）样本空间 Ω 上的实变函数 $\xi=\xi(\omega)$ 为随机变量

定义 5.3

（概率分布）设随机变量 ξ，函数 P_ξ 定义在 Borel 子集 σ 代数 $\mathcal{B}$ 上，其满足

$$(\forall B\in\mathcal{B}):(P_\xi(B):=P\{\omega|\xi(\omega)\in B\})$$

则称函数 P_ξ 为随机变量 ξ 的概率分布律 (law of probability distribution) 或简称概率分布 (probability distribution) 或简称分布 (distribution/распределение)

5.1.2 分布函数

Distribution Function

定义 5.4 (分布函数)

设随机变量 ξ 与概率空间 $(\Omega,\mathcal{A},P_\xi)$，$x\in\mathbb{R}^1$，则称函数

$$F_\xi(x)=P\{\omega:\xi(\omega)\leqslant x\}$$

为随机变量 ξ 的概率分布函数 (probability distribution function) 或累积分布函数 (cumulative distribution function) 或简称分布函数 (distribution function/функция распределения)

注 （分布函数）

经常也定义

$$F_\xi(x)=P\{\omega:\xi(\omega)<x\}$$

为随机变量 ξ 的分布函数

命题 5.2 (分布函数必要条件)

分布函数具有下列性质：

1)（保序性）$x_1\leqslant x_2\Rightarrow F_\xi(x_1)\leqslant F_\xi(x_2)$

2)（右连续性）$\lim\limits_{\delta\to0^+}F_\xi(x+\delta)=F_\xi(x)$

3)（规范性）$\lim\limits_{x\to-\infty}F_\xi(x)=0,\ \lim\limits_{x\to\infty}F_\xi(x)=1$

证明

1) 定义事件 $A(x):=(\xi<x)$，若 $x_1\leqslant x_2$，则有 $A(x_1)\subset A(x_2)$ 且

$$F_\xi(x_1)=P(A(x_1))\leqslant P(A(x_2))=F_\xi(x_2)$$

2) 对任意 $x\in\mathbb{R}$，事件序列 $\left\{X\leqslant x+\dfrac{1}{n},n\in\mathbb{N}\right\}$ 为非升集序列，其序列极限为 $(X\leqslant x)$，则由概率上连续性得 $F(x)=\lim\limits_{n\to\infty}F\left(x+\dfrac{1}{n}\right)$。注意到

$$(\exists V\in\mathcal{V}(x))(\forall\delta>0):\left[(x+\delta\in V)\Rightarrow(\exists n\in\mathbb{N})\left(x+\frac{1}{n+1}<x+\delta\leqslant x+\frac{1}{n}\right)\right]$$

则有

$$F(x)=\lim_{n\to\infty}F\left(x+\frac{1}{n+1}\right)\leqslant\liminf_{\delta\to0^+}F(x+\delta)\leqslant\limsup_{\delta\to0^+}F(x+\delta)\leqslant\lim_{n\to\infty}F\left(x+\frac{1}{n}\right)=F(x)$$

即有 $\lim\limits_{\delta\to0^+}F_\xi(x+\delta)=F_\xi(x)$，右连续性得证

3) 注意到

$$\lim_{x\to-\infty}(\xi<x)=\varnothing,\quad \lim_{x\to\infty}(\xi<x)=\Omega$$

则由概率的上下连续性得证

注 如果把分布函数定义改为 $F_\xi(x)=P\{\omega:\xi(\omega)<x\}$，同理只需将命题改为

2^*)（左连续性）$\lim\limits_{\delta\to0^-}F_\xi(x+\delta)=F_\xi(x)$

设序列 $\{x_n\}$ 单调递增且 $\lim\limits_n x_n=x$，则事件序列 $\{A(x_n)\}$ 也为单调递增且有 $\bigcup\limits_{n=1}^{\infty}A(x_n)=A(x)$。利用概率的连续性即有 $F_\xi(x_n)=P(A(x_n))\to P(A(x))=F_\xi(x)$

定义 5.5 (分布函数形式反函数)

设 $F(x)$ 为定义在实数 $\mathbb{R}$ 上的满足 1) 保序性、2) 右连续性和 3) 规范性的函数，则称

$$F^{-1}(u)=\inf\{x\mid F(x)\geqslant u\},\quad u\in(0,1) \tag{5.3}$$

为 $F(x)$ 的（形式）反函数

♣

注（分布函数形式反函数）

如此定义的反函数为通常的反函数概念的推广。通常定义反函数时要求原函数严格单调，但分布函数未必总是严格单调递增。不难证明，所定义的 $F^{-1}(u)$ 为定义在区间 $(0,1)$ 上的保序的右连续函数。当 $F(x)$ 为严格单调递增的连续函数时，$F^{-1}(u)$ 为 $F(x)$ 的通常意义下的反函数

注（分布函数形式反函数）

同样地，也可以如下定义分布函数形式反函数：

定义二：将 $F(x)$ 的反函数 F^{-1} 定义为

$$F^{-1}(u)=\sup\{x\mid F(x)\leqslant u\},\quad u\in[0;1]$$

由于 $F(x)$ 为右连续的，则该定义下 F^{-1} 左连续

定义三：将 $F(x)$ 的反函数 F^{-1} 定义为

$$F^{-1}(u)=\inf\{x\mid F(x)>u\},\quad u\in(0;1)$$

由于 $F(x)$ 为右连续的，则该定义下 F^{-1} 右连续

定义四：将 $F(x)$ 的反函数 F^{-1} 定义为

$$F^{-1}(u)=\sup\{x\mid F(x)<u\},\quad u\in(0;1)$$

由于 $F(x)$ 为右连续的，则该定义下的 F^{-1} 左连续

命题 5.3 (分布函数充分条件)

定义在实数 $\mathbb{R}$ 上的函数 $F(x)$，若满足

1)（保序性）$x_1\leqslant x_2\Rightarrow F_\xi(x_1)\leqslant F_\xi(x_2)$

2)（右连续性）$\lim\limits_{\delta\to0^+}F_\xi(x+\delta)=F_\xi(x)$

3)（规范性）$\lim\limits_{x\to-\infty}F_\xi(x)=0,\ \lim\limits_{x\to\infty}F_\xi(x)=1$

则为某个随机变量的分布函数

♠

证明 将概率空间 $(\Omega,\mathcal{A},P)$ 选为区间 $(0;1)$ 上的几何概率空间，在其上定义一个随机变量 Y，再令 $X=F^{-1}(Y)$，则可证 X 为以 $F(x)$ 作为自己的分布函数的随机变量

由式 (5.3) 易知 $F(x)\geqslant u\leftrightarrow x\geqslant F^{-1}(u)$，而 $(\forall x\in\mathbb{R}):F(x)\in\mathbb{R}$，故由 Y 为随机变量知

$$(X\leqslant x)=\left(F^{-1}(Y)\leqslant x\right)=(Y\leqslant F(x))\in\mathcal{A}$$

则 $X=F^{-1}(Y)$ 为一个随机变量。注意到 Y 为 $(0;1)$ 上的几何概型随机变量，则有

$$P(X\leqslant x)=P\left(F^{-1}(Y)\leqslant x\right)=P(Y\leqslant F(x))=F(x),\quad x\in\mathbb{R}$$

则 $F(x)$ 为随机变量 X 的分布函数

定理 5.1 (分布函数充要条件)

定义在实数 $\mathbb{R}$ 上的函数 $F(x)$ 为一个分布函数的充要条件为它满足

1)（保序性）$x_1 \leqslant x_2 \Rightarrow F_\xi(x_1) \leqslant F_\xi(x_2)$

2)（右连续性）$\lim\limits_{\delta \to 0^+} F_\xi(x+\delta) = F_\xi(x)$

3)（规范性）$\lim\limits_{x \to -\infty} F_\xi(x) = 0, \lim\limits_{x \to \infty} F_\xi(x) = 1$

♡

例题 5.1 （分布函数充要条件）

试判断下面的函数是否为随机变量 X 的分布函数:

1)

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ \dfrac{x}{1+x}, & x > 0 \end{cases}$$

2) $F(x) = \dfrac{\pi}{2} + \operatorname{arctg} x$

解 1) 显然，函数 $\dfrac{x}{1+x} = 1 - \dfrac{1}{1+x}$ 严格递增且处处连续。注意到

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

则有函数为随机变量 X 的分布函数

2) 显然，函数处处不减且处处连续，但这时有

$$\lim_{x \to -\infty} F(x) = \frac{\pi}{2} - \frac{\pi}{2} = 0, \lim_{x \to \infty} F(x) = \frac{\pi}{2} + \frac{\pi}{2} = \pi \neq 1$$

不满足规范性，则该函数不为随机变量 X 的分布函数

性质 （分布函数基本性质）

分布函数具有下列性质:

4)（分布函数基本值域）$(\forall x \in \mathbb{R}^1): 0 \leqslant F_\xi(x) \leqslant 1$

5)（分布函数基本等式）$P(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a)$

证明 4) 利用概率的规范性即证

5) 注意到 $(a \leqslant \xi < b) = (\xi < b) - (\xi < a)$ ，则有

$$P(a \leqslant \xi < b) = P(\xi < b) - P(\xi < a) = F_\xi(b) - F_\xi(a)$$

命题即证

例题 5.2 （分布函数基本等式）

对于任意 $a \in \mathbb{R}$ 都有

$$F(\xi = a) = \lim_{n \to \infty} P\left(a - \frac{1}{n} < \xi \leqslant a\right) = F(a) - \lim_{n \to \infty} F\left(a - \frac{1}{n}\right) = F(a) - F(a-0)$$

注 当 $x = a$ 为分布函数 $F(x)$ 的间断点时，随机变量 ξ 取 a 为值的概率即为 $F(x)$ 在该点处的跃度；而当 $x = a$ 为分布函数 $F(x)$ 的连续点时，随机变量 ξ 取 a 值的概率为 0；若分布函数 $F(x)$ 处处连续，则随机变量 X 取任意单点值的概率均为 0

定义 5.6 (分布函数凸组合)

设 $\{F_n(x), n \in \mathbb{N}\}$ 为一列分布函数，$\{a_n, n \in \mathbb{N}\}$ 为一列非负实数，若 $\sum\limits_{n=1}^{\infty} a_n = 1$，则称

$$F(x) = \sum_{n=1}^{\infty} a_n F_n(x)$$

为分布函数的凸组合 (convex combination)

♣

定理 5.2 (分布函数凸组合保构)

分布函数的凸组合为分布函数 ♡

证明 1) 保序性显然。有限和情况定理显然，下仅考虑可数和情形。设 $F(x)=\sum\limits_{n=1}^{\infty} a_n F_n(x)$，任意给定 $0<\varepsilon<1$。由 $\sum\limits_{n=1}^{\infty} a_n=1$，则可取 m 充分大使得 $\sum\limits_{n=1}^{m} a_n>1-\frac{1}{2}\varepsilon$，亦即 $\sum\limits_{n=m+1}^{\infty} a_n<\frac{1}{2}\varepsilon$

任意给定 $x_0 \in \mathbb{R}$，由于 $F_1(x), F_2(x), \cdots, F_m(x)$ 都在 $x=x_0$ 处右连续，则只要 $x>x_0$ 充分接近 x_0 便有

$$F_n\left(x_0\right) \leqslant F_n(x)<F_n\left(x_0\right)+\frac{1}{2}\varepsilon, n=1,2,\cdots,m$$

则有

$$\begin{aligned}
F(x) & \leqslant \sum_{n=1}^{m} a_n F_n(x)+\frac{1}{2}\varepsilon \leqslant \sum_{n=1}^{m} a_n\left(F_n\left(x_0\right)+\frac{1}{2}\varepsilon\right)+\frac{1}{2}\varepsilon \\
& \leqslant \sum_{n=1}^{m} a_n F_n\left(x_0\right)+\varepsilon \leqslant \sum_{n=1}^{\infty} a_n F_n\left(x_0\right)+\varepsilon=F\left(x_0\right)+\varepsilon
\end{aligned}$$

则 $F(x)$ 具有 2) 右连续性，类似可证 $F(x)$ 的 3) 规范性

5.1.3 Bernoulli 随机变量与二项随机变量 Bernoulli Random Varible and Binomial Random Variable

实际上，在 Jakob Bernoulli 的《推测术》中就已经证明了二项分布的概率公式，并且 Jakob Bernoulli 开创了通过无穷级数计算概率的方法，而值得一提的是，在 Jakob Bernoulli 的时代无穷级数是崭新的数学领域，他本人做出了重要贡献。另外，Jakob Bernoulli 在《推测术》中也首先明确指出了具有确定概率的事件重复独立试验的模型，因此也把这种基础模型称为 Bernoulli 概型，称对应试验为 Bernoulli 试验，称对应随机变量为 Bernoulli 随机变量。

在很长一个时期中，统计方法在社会问题中的应用主要限于人口统计，特别是男、女出生的比例问题，而这是一个典型的二项分布相关的统计问题。不妨说，在 19 世纪以前，数理统计是二项分布的天下，而进入 19 世纪后，随着 Gauss 误差分布理论的建立，正态分布才逐渐在数理统计学中取得中心地位。

二项分布在数理统计历史上另一个重大作用是，正是由于对二项分布中未知概率的推断的探讨，在 18 世纪中叶导致了 Bayes 推断思想的建立，催生了 Bayes 学派。

定义 5.7 (示性函数)

设 X 为某个集合，$A \subset X$，则称函数

$$I_A(x)=\begin{cases}1, & x \in A, \\ 0, & x \in A^c.\end{cases}$$

为集合 A 的示性函数或指标 (index/индикатор)，可简记为 I_A ♣

定义 5.8 (Bernoulli 随机变量)

设 $(\Omega, \mathcal{A}, P)$ 为概率空间，$A \in \mathcal{A}$，并设

$$X(\omega)=I_A(\omega)=\begin{cases}1, & \omega \in A, \\ 0, & \omega \in A^c.\end{cases}$$

即有 $(X=1)=\{\omega \mid X(\omega)=1\}=A,(X=0)=\{\omega \mid X(\omega)=0\}=A^c$

若 $P(A)=p, 0<p<1$，则有

$$P(X=1)=P(A)=p, \quad P(X=0)=P\left(A^c\right)=1-p:=q$$

即 $X=I_A$ 具有概率分布

X	0	1
P	q	p

或记为 $\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$

称具有上述概率分布的随机变量为参数为 p 的 Bernoulli 随机变量，称上述概率分布为参数为 p 的两点分布 ((two-point distribution) 或参数为 p 的 Bernoulli 分布 (Bernoulli distribution/распределение Бернулли) ♣

定义 5.9 (二项分布)

设 $(\Omega,\mathcal{A},P)$ 为概率空间，$A\in\mathcal{A}$，并设 $X=aI_A+bI_{A^c}$，则有

$$X(\omega)=I_A(\omega)=\begin{cases} a, & \omega\in A, \\ b, & \omega\in A^c \end{cases}$$

若 $P(A)=p, 0<p<1$，则有

$$P(X=1)=P(A)=p,\quad P(X=0)=P\left(A^c\right)=1-p:=q$$

即 $X=I_A$ 具有概率分布

X	a	b
P	q	p

或记为 $\begin{pmatrix} a & b \\ q & p \end{pmatrix}$

称具有上述概率分布的随机变量为参数为 p 的二项式随机变量，简称参数为 p 的 二项随机变量 (binomial random variable)，称上述概率分布为参数为 p 的二项式分布，简称参数为 p 的二项分布 (binomial distribution/биномиальное распределение) ♣

注 (二项分布一般化)

设 $(\Omega,\mathcal{A},P)$ 为概率空间，可数完备事件组 $A_1,A_2\cdots,\in\mathcal{A}$ 分别具有概率 $P\left(A_k\right)=p_k, k=1,2,\cdots$，又若 $a_1,a_2,\cdots$ 为可数个实数，则 $X=\sum\limits_{k=1}^{\infty}a_kI_{A_k}$ 为一个定义在该概率空间上的可以取 $a_1,a_2,\cdots$ 这可数个不同值的随机变量。当 $\omega\in A_k$ 时有 $X(\omega)=a_k, k=1,2,\cdots$，则 X 的概率分布为

$$P\left(X=a_k\right)=p_k,\quad k=1,2,\cdots$$

5.2 离散型随机变量 Discrete Random Variable

5.2.1 离散分布 Discrete Distribution

定义 5.10 (离散分布)

设随机变量 ξ 的值仅能取至多可数集 $X=\{x_1,x_2,\ldots\}$ 且满足 $P(\xi\in X)=1$，则称

$$P\left(\xi=x_k\right)=p_k$$

为这些值对应的概率 (probability)，称随机变量 ξ 为离散型随机变量简称离散随机变量 (discrete random variable)，称概率组

$$(P(\xi=x_1),P(\xi=x_2),\cdots,P(\xi=x_k),\cdots)$$

为离散（概率）分布 (discrete distribution/дискретное распределение) ♣

注 （离散分布）

对于离散型随机变量 ξ 有

$$F_{\xi}(x)=P(\xi \leqslant x)=\sum_{j: a_{j} \leqslant x} P\left(\xi=a_{j}\right)$$

则 $F_{\xi}(x)$ 为一个右连续的不减阶梯函数

注 （离散分布）

在实际应用中通常不把离散分布写成分布函数的形式，而是以概率分布的形式表示，例如记

$$p_{j}=F\left(a_{j}\right)-F\left(a_{j}-0\right), \quad j=1,2, \cdots$$

其中 $p_{j} \geqslant 0, \sum_{j} p_{j}=1$ ，以 $P\left(\xi=a_{j}\right)=p_{j}(j=1,2, \cdots)$ 表示其概率分布，$\left\{a_{j}, j=1,2, \cdots\right\}$ 为 ξ 的取值集合。也可以表示为矩阵的形式，例如

$$\left(\begin{array}{ccc}a_{1} & a_{2} & \cdots \\ p_{1} & p_{2} & \cdots\end{array}\right)$$

也称离散型随机变量的概率分布为其**概率质量函数** (probability mass function)

例题 5.3 （离散分布）

盒中有 n 个白球和 1 个黑球。每次从盒中取出一个球，并放入一个白球，直到取出黑球为止，试求所需取球次数的概率分布

解 用 X 表示所需取球次数，其取值集合为可数集 $\{1,2, \cdots\}$。若在所考虑的随机试验中以 A_{k} 表示直到第 k 次取球时才取到黑球的事件，则事件 $(X=k)$ 即为 A_{k}。以 Ω 表示 k 次取球的所有可能结果，则有 $|\Omega|=(n+1)^{k},\left|A_{k}\right|=n^{k-1}$，则

$$P\left(A_{k}\right)=\frac{\left|A_{k}\right|}{|\Omega|}=\frac{n^{k-1}}{(n+1)^{k}}=\frac{1}{n+1}\left(1-\frac{1}{n+1}\right)^{k-1}$$

又由该式对任意 k 都满足，则有

$$P(X=k)=\frac{1}{n+1}\left(1-\frac{1}{n+1}\right)^{k-1}, \quad k=1,2, \cdots,$$

即为所需取球次数 X 的概率分布

注 显然也有

$$\sum_{k=1}^{\infty} P(X=k)=\frac{1}{n+1} \sum_{k=0}^{\infty}\left(1-\frac{1}{n+1}\right)^{k}=1$$

5.2.2 Bernoulli 试验与二项概型
Bernoulli Experiment and Binomial Model of Probability

例题 5.4 （Bernoulli 试验/二项概型第一模型）

假设将一枚硬币接连掷 n 次，观测结果用有序数组 $\left(a_{1}, \cdots, a_{n}\right)$ 表示，其中当第 i 次抛掷出现正面时 $a_{i}=1$，当第 i 次出现反面时 $a_{i}=0$，样本空间可记为 $\Omega=\left\{\omega: \omega=\left(a_{1}, \cdots, a_{n}\right), a_{i}=0,1\right\}$ ，在每一个基本结果 $\omega=\left(a_{1}, \cdots, a_{n}\right)$ 上定义函数

$$p(\omega)=p^{\sum a_{i}} q^{n-\sum a_{i}}$$

其中 $p, q \geqslant 0, p+q=1$，则该函数上定义了一个有限概率空间

解 只需验证 $\sum_{\omega \in \Omega} p(\omega)=1$。考虑所有满足 $\sum_{i} a_{i}=k,(k=0,1, \cdots, n)$ 的基本结果 $\omega=\left(a_{1}, \cdots, a_{n}\right)$，由计数原理，这样的基本结果个数等于 C_{n}^{k}，则有

$$\sum_{\omega \in \Omega} p(\omega)=\sum_{k=0}^{n} C_{n}^{k} p^{k} q^{n-k}=(p+q)^{n}=1$$

而同样可设 $\mathcal{A}$ 为空间 Ω 上 σ 代数，即可在 A 上定义概率 $P(A)=\sum_{\omega \in A} p(\omega), A \in \mathcal{A}$ ，其中 $P(\{\omega\})=p(\omega), \omega \in \Omega$,

从而定义了一个概率模型

定义 5.11 (Bernoulli 试验)

设离散概率空间 $(\Omega, \mathcal{A}, P)$，其中样本空间 Ω 的样本点形如 $\omega = (\omega_1, \dots, \omega_n), \omega_k = 0, 1, k = \overline{1, n}$，而样本点 ω 的概率定义为

$$P(\omega) = p^m (1-p)^{n-m}$$

其中 m 为结果 ω 中取 1 的次数。则称该离散概率空间为参数为 n 和 p 的 Bernoulli 试验 (схема Бернулли/Bernoulli experiment) ♣

定义 5.12 (二项概型)

设离散概率空间 $(\Omega, \mathcal{A}, P)$，其中样本空间 Ω 为 $\{\omega_0, \dots, \omega_n\}$，而样本点 ω_m 的概率定义为

$$P(\omega_m) = b(n, p, m) = C_n^m p^m (1-p)^{n-m}$$

其中 $m \in \overline{0:n}, 0 < p < 1$，则称该离散概率空间为参数为 n 和 p 的二项概型 (биномиальная модель)，其概率分布为二项分布 ♣

性质 （二项概型性质）

二项概型满足下列性质：

1) （非负性）$b(n, p, m) \geqslant 0$
2) （规范性）$\sum\limits_{m=0}^{n} b(n, p, m) = 1$
3) （对称性）$b(n, p, m) = b(n, 1-p, n-m)$
4) （单调性）若 $m < (n+1)p - 1$，则当 m 趋近于 $m+1$ 时概率增加；若 $m > (n+1)p - 1$，则当 m 趋近于 $m+1$ 时概率减小
5) （最值）若 $(n+1)p$ 为整数，则有两个可能的最大值 $(n+1)p - 1$ 与 $(n+1)p$；若 $(n+1)p$ 为小数，则有一个可能的最大值 $m_0 = [(n+1)p]$

证明 1)2)3) 显然，仅证 4)5)。由比值

$$\frac{b(n, p, m+1)}{b(n, p, m)} = \frac{C_n^{m+1} p^{m+1} (1-p)^{n-m-1}}{C_n^m p^m (1-p)^{n-m}} = \frac{n-m}{m+1} \frac{p}{1-p}$$

当 $m < (n+1)p + 1$ 时大于 1 与当 $m > (n+1)p - 1$ 时小于 1 即证

例题 5.5 （二项概型第二模型/随机游动）

引进一个与 Bernoulli 试验等价的模型，其描绘某一质点的随机游动。设质点从 0 出发，经过单位时间向上或向下移动一步，如图 (5.1) 所示。

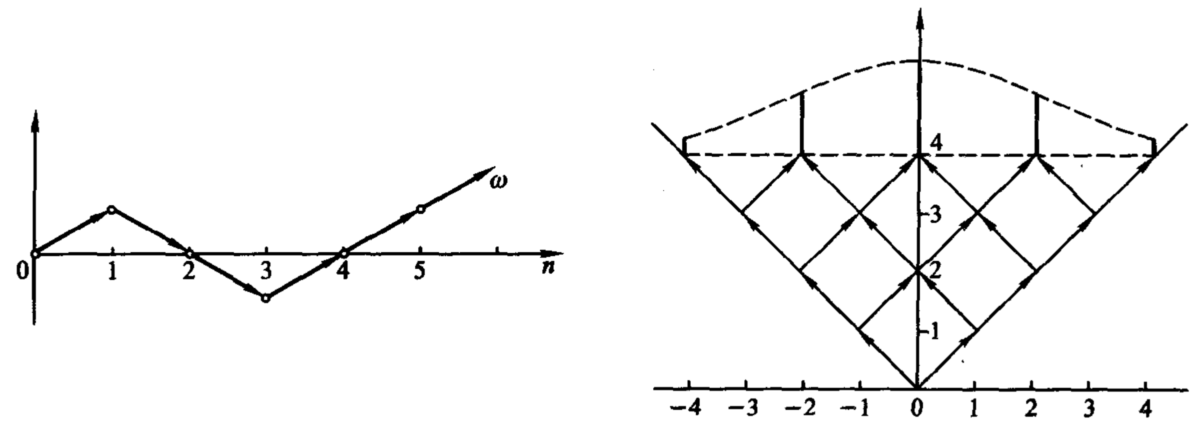

图 5.1: 随机游动示意图

1)（概率空间）经过 n 步后质点最多向上移动 n 步或向下移动 n 步。显然，质点运动的每一条“轨道” ω 可由数组 $(a_1,\cdots,a_n)$ 表示，其中当质点在第 i 步向上移动时 $a_i=+1$，而当质点在第 i 步向下移动时 $a_i=-1$。在每条“轨道” ω 上定义概率 $p(\omega)=p^{\nu(\omega)}q^{n-\nu(\omega)}$ ，其中 $\nu(\omega)$ 为数组 $\omega=(a_1,\cdots,a_n)$ 中 “+1” 的个数，即

$$\nu(\omega)=\frac{(a_1+\cdots+a_n)+n}{2}$$

注意到 $p\geqslant 0,q\geqslant 0,p+q=1$，则由 $\sum\limits_{\omega\in\Omega}p(\omega)=1$ 即知概率组 $\{p(\omega)\}$ 与事件 $\omega=(a_1,\cdots,a_n)$ 的空间 Ω 及其子集决定质点 n 步运动的概率模型。

2)（A_k 概率）对于事件 A_k “质点经 n 步到达纵坐标为 k 的点”，一切满足

$$\nu(\omega)-[n-\nu(\omega)]=k$$

亦即满足 $\nu(\omega)=\dfrac{n+k}{2}$ 的轨道都满足事件条件，其数量由计数原理即知为 $C_n^{\frac{n+k}{2}}$ ，则有

$$P\left(A_k\right)=\mathrm{C}_n^{\frac{n+k}{2}}p^{\frac{n+k}{2}}q^{\frac{n-k}{2}}$$

则可认为二项分布 $\{P\left(A_{-n}\right),\cdots,P\left(A_0\right),\cdots,P\left(A_n\right)\}$ 描绘质点移动 n 步后位置的概率分布。特别地，对于 $p=q=\dfrac{1}{2}$ 的情形，每条轨道的概率等于 2^{-n}，则有

$$P\left(A_k\right)=\mathrm{C}_n^{\frac{n+k}{2}}2^{-n}$$

3)（渐进性质）若移动步数等于 $2n$，则由二项式系数性质即知，在概率 $P\left(A_k\right)(|k|\leqslant 2n)$ 中最大的概率为 $P\left(A_0\right)=\mathrm{C}_{2n}^n 2^{-2n}$ 。注意到 Stirling 公式则有 $n!\sim\sqrt{2\pi n}\mathrm{e}^{-n}n^n$ ，则

$$\mathrm{C}_{2n}^n=\frac{(2n)!}{(n!)^2}\sim\frac{2^{2n}}{\sqrt{\pi n}}$$

即对充分大的 n 有 $P\left(A_0\right)\sim\dfrac{1}{\sqrt{\pi n}}$

定义 5.13 (Bernoulli 随机变量独立)

设 A_1 和 A_2 为概率空间 $(\Omega,\mathcal{A},P)$ 中的两个相互独立的随机事件，则 $X_1=I_{A_1}$ 与 $X_2=I_{A_2}$ 为定义在同一个概率空间上的两个不同的 Bernoulli 随机变量，称 Bernoulli 随机变量 X_1 与 X_2 独立 (independent) ♣

定义 5.14 (Bernoulli 随机变量独立和)

设 $X_1=I_{A_1}$ 与 $X_2=I_{A_2}$ 为两个相互独立的 Bernoulli 随机变量，定义 $X=X_1+X_2$，满足

$$X(\omega)=X_1(\omega)+X_2(\omega)=\begin{cases}0, & \omega\in A_1^{\mathrm{c}}A_2^{\mathrm{c}}\\ 1, & \omega\in A_1A_2^{\mathrm{c}}\cup A_1^{\mathrm{c}}A_2\\ 2, & \omega\in A_1A_2\end{cases}$$

称 $X=X_1+X_2$ 为相互独立的 Bernoulli 随机变量 X_1 与 X_2 的和，简称为 Bernoulli 变量 X_1 与 X_2 的独立和

♣

例题 5.6 （二项概型第三模型/Bernoulli 随机变量独立和）

设 $X_1=I_{A_1}$ 与 $X_2=I_{A_2}$ 为两个相互独立的 Bernoulli 随机变量，若 $P\left(A_j\right)=p_j,j=1,2$ 且记 $q_j=1-p_j$，则 $X=X_1+X_2$ 的概率分布为

$$\begin{pmatrix}0 & 1 & 2\\ q_1q_2 & p_1q_2+q_1p_2 & p_1p_2\end{pmatrix}$$

若 $P\left(A_j\right)=p,j=1,2$ 且记 $q=1-p$，则 $X=X_1+X_2$ 的概率分布为

$$\begin{pmatrix}0 & 1 & 2\\ q^2 & 2pq & p^2\end{pmatrix}$$

归纳即得，若 $\xi_1,\cdots,\xi_r$ 为两两独立的 Bernoulli 随机变量，且 $P\left\{\xi_i=1\right\}=p,P\left\{\xi_i=0\right\}=q$，则随机变量

$\zeta = \xi_1 + \cdots + \xi_r$ 服从二项分布

$$P_\zeta(k) = \mathrm{C}_n^k p^k q^{n-k} \quad (k = 0, 1, \cdots, n)$$

5.2.3 Poisson 分布 Poisson Distribution

定理 5.3 (Poisson 定理)

设近似二项概型参数为 n, p_n，若 $\lim\limits_{n\to\infty} np_n = \lambda \in \mathbb{R}^+$，则有

$$(\forall m \in \mathbb{N}) : b(n, p_n, m) \underset{n\to\infty}{\longrightarrow} \pi_m(\lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$$

证明 固定某个非负整数 m，则有

$$\begin{aligned} b(n, p_n, m) &= C_n^m p_n^m (1-p_n)^{n-m} = \frac{n!}{m!(n-m)!} \cdot p^m (1-p_n)^{n-m} \\ &= \frac{1}{m!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)(1-p_n)^{-m}(np_n)^m(1-p_n)^n \end{aligned}$$

由 $\lim\limits_{n\to\infty} np_n = \lambda \in \mathbb{R}^+$ 则有当 m 为非负整数时

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)(1-p_n)^{-m} = 1$$

且 $\lim\limits_{n\to\infty}(np_n)^k = \lambda^k$，由此仅需证 $\lim\limits_{n\to\infty}(1-p_n)^n = \mathrm{e}^{-\lambda}$ ，即仅需证

$$\lim_{n\to\infty}(1-p_n)^n = \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n$$

注意到 $|1-p_n| < 1$，且当 n 充分大时也有 $\left|1-\frac{\lambda}{n}\right| < 1$，考虑当 $|a| \leqslant 1, |b| \leqslant 1$ 时有不等式 $|a^n - b^n| \leqslant n|a-b|$，则有当 $n \to \infty$ 时有

$$\left|(1-p_n)^n - \left(1-\frac{\lambda}{n}\right)^n\right| \leqslant n\left|p_n - \frac{\lambda}{n}\right| = |np_n - \lambda| \underset{n\to\infty}{\longrightarrow} 0$$

定理即证

注 更精确的分析指出，若 $np = \lambda$ 且 $(\forall m \geqslant n) : b(n, p, m) = 0$，则有

$$\sum_{m=0}^{\infty} |b(n, p, m) - \pi_m(\lambda)| \leqslant np^2$$

若 np^2 明显小于 $\pi_m(\lambda)$，则该近似值适用

例题 5.7 （Poisson 定理）

一个电话交换机有 10000 个号码，该站平均每天接到 30000 个电话。求一个特定号码恰好接到两个电话的概率

解 假设呼叫任意号码的可能性相同，并且在每次呼叫中选择一个号码，独立于其他呼叫。则有一个参数为 $n = 3 \cdot 10^4, p = 10^{-4}$ 的 Bernoulli 试验。则由 Poisson 定理 (5.3)，$\lambda = 3$，即有

$$b(n, p, 2) = C_{30000}^2 p^2 (1-p)^{29998} \underset{n\to\infty}{\longrightarrow} \frac{3^2}{2!} e^{-3} \approx 0.1804$$

另外，$np^2 = 0.0003$，则该近似精确到小数点后前三位

命题 5.4 (Poisson 分布定义正确性)

设样本空间为 $\Omega = \{0, 1, 2\cdots\}$，$\mathcal{A}$ 为 Ω 上 σ 代数另设 $\lambda > 0$，并在样本点上定义函数

$$p_m = \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}, \quad m = 0, 1, 2, \cdots$$

则有 $(\Omega, \mathcal{A}, p_m)$ 为离散概率空间

证明 非负性显然。由

$$\sum_{m=0}^{\infty} p_n = \mathrm{e}^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \mathrm{e}^{-\lambda} \cdot \mathrm{e}^{\lambda} = 1$$

即得规范性

定义 5.15 (Poisson 分布)

设样本空间为 $\Omega = \{0, 1, 2 \cdots\}$，另设 $\lambda > 0$，并在样本点上定义概率

$$p_m = \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}, \quad m = 0, 1, 2, \cdots$$

则称概率组

$$\left\{ p_m = \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda} \right\}_{m=0}^{\infty}$$

为参数为 λ 的 Poisson 分布 (Poisson distribution/распределение Пуассона)，称概率分布为 Poisson 分布的随机变量为 Poisson 随机变量 ♣

性质 （Poisson 分布性质）

参数为 λ 的 Poisson 分布满足下列性质：

1) （单调性）当 $m < \lambda$ 时，p_m 随 m 增大而增大；当 $m > \lambda$ 时，p_m 随 m 增大而减小
2) （最值）p_m 在 $k = [\lambda]$ 处达最大值。当 λ 为整数时，在 $k = \lambda$ 和 $\lambda - 1$ 处达最大值

证明 由比值

$$\frac{p_k}{p_{k-1}} = \frac{\lambda^k (k-1)! \mathrm{e}^{-\lambda}}{\lambda^{k-1} k! \mathrm{e}^{-\lambda}} = \frac{\lambda}{k}, \quad k \geqslant 1$$

性质即证

例题 5.8 （Poisson 分布随机选择不变性）

假设要记录一块放射性物质一段时间内所放射出的粒子数目。已知它在单位时间内放射出的粒子数目 ν 服从参数为 $\lambda > 0$ 的 Poisson 分布。但是由于仪器原因，并非每一粒放射出的粒子都可被记录下来，现知每粒粒子可被记录下来的概率均为 $p, 0 < p < 1$，并且各粒子能否被记录下来的事件相互独立。试求单位时间内被记录下来的粒子数目 X 的概率分布

解 对每一粒放射出的粒子都可以用一个参数为 p 的 Bernoulli 随机变量来表示它是否被记录下来：若被记录下来，令该 Bernoulli 变量的值为 1；而若未被记录下来，就令其为 0。由于各粒子能否被记录下来的事件相互独立，从而当单位时间内放射出的粒子数目已知时，例如为 n 时，被记录下来的粒子数目就服从二项分布 $b(n, p, m)$。但单位时间内放射出的粒子数目也为一个随机变量 ν，则只能在 ν 值给定的条件下考虑被记录下来的粒子数目 X 的分布。则当 $\nu = n$ 时 X 服从二项分布 $b(n, p, m)$

$$P(X = m \mid \nu = n) = b(n, p, m) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \cdots, n$$

其中 $(\nu = n)$ 为一个随机事件。由于 ν 服从 Poisson 分布，则 ν 可能为任意非负整数。若记 $q = 1 - p$，则由全概率公式即得

$$\begin{aligned} P(X = m) &= \sum_{n=0}^{\infty} P(\nu = n) P(X = m \mid \nu = n) = \sum_{n=k}^{\infty} \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda} b(n, p, m) \\ &= \mathrm{e}^{-\lambda} \sum_{n=m}^{\infty} \frac{(\lambda q)^{n-m}}{(n-m)!} \cdot \frac{1}{m!} \cdot (\lambda p)^m = \frac{1}{m!}(\lambda p)^m \mathrm{e}^{-\lambda p}, \quad k = 0, 1, 2, \cdots \end{aligned}$$

即为所求概率分布

注 该例表明在对 Poisson 变量 ν 作了随机选择之后仍然服从 Poisson 分布，称该性质为 Poisson 分布在随机选择之下的不变性

例题 5.9 （Poisson 分布）

设做有三个结果的 n 次试验，结果对应概率分别为 p_1, p_2, p_3，假设 $n \to \infty, p_1 \to 0, p_2 \to 0$ 使得 $np_1 \to$

$\lambda_1, np_2 \to \lambda_2$，则固定 m_1 与 m_2 即有

$$P\left(m_1, m_2, n-m_1-m_2\right) \to \frac{\lambda^{m_1}}{m_1!} e^{-\lambda_1} \frac{\lambda^{m_2}}{m_2!} e^{-\lambda_2}$$

5.2.4 几何分布 Geometric Distribution

定义 5.16 (几何分布)

设离散随机变量 ξ，若其取值为至多可数集，$0<p<1$ 且有

$$P(\xi=m)=p(1-p)^{m-1}, \quad m=1,2,\ldots$$

则称该随机变量的概率分布为参数为 p 的几何分布 (geometric distribution/геометрическое распределение)，经常记为 $\operatorname{Geo}(p)$ ♣

定理 5.4 (正整数集取值几何分布无记忆性)

设随机变量 ξ 的取值为正整数集，则下列命题等价：

1) (几何分布) 随机变量 ξ 的概率分布为参数为 p 的几何分布
2) (无记忆性/memoryless property/свойство отсутствия) 设 $j=1,2,\cdots, \quad m=1,2,\cdots$，则有 $P(\xi=m+j \mid \xi>m)=P(\xi=j)$ ♡

证明 1)⇒2)：由条件概率定义即得

$$\begin{aligned} P(\xi=m+j \mid \xi>k) &= \frac{P(\xi=m+j \mid \xi>k)}{P(\xi>k)} \\ &= \frac{P(\xi=k+j)}{P\left(\bigcup_{j=m+1}^{\infty}\{\xi=j\}\right)} = \frac{(1-p)^{m+j-1} p}{\sum_{j=m+1}^{\infty}(1-p)^{j-1} p} = (1-p)^{j-1} p = P(\xi=j) \end{aligned}$$

2)⇒1)：设 $r_m=P(\xi>m)$，由 $\{\xi=m+1\}=\{\xi>m\}-\{\xi>m+1\}$ 即得

$$p=P(\xi=m+1 \mid \xi>m)=\frac{P(\xi>m)-P(\xi>m+1)}{P(\xi>m)}=1-\frac{r_{m+1}}{r_m}$$

而由 $r_0=P(\xi>0)=1$ 即得

$$r_{m+1}=(1-p) r_m=(1-p)^2 r_{m-1}=\cdots=(1-p)^{m+1}$$

即得 $P(\xi=m)=P(\xi>m-1)-P(\xi>m)=r_{m-1}-r_m=(1-p)^{m-1} p$

注 另一个几何分布的充要条件的陈述方式为

$$P(\xi>n)=(1-p)^n, n \geqslant 0, n \in \mathbb{N}$$

注 实际上可以证明几何分布为唯一具有无记忆性的取值集合为正整数集的离散分布

例题 5.10 (几何分布)

甲向一个目标射击，直到击中为止，用 X 表示首次击中目标时的射击次数。若甲每次击中目标的概率为 p，则 X 的概率分布为几何分布

解 显然 X 的取值集合为 $\mathbb{N}=\{1,2,\cdots\}$。设一个概率空间 $\{\Omega, \mathcal{F}, P\}$，其中存在全体独立的事件列 $\{A_k \mid k \in \mathbb{N}\}$，并且有 $P\left(A_k\right)=p, k=1,2,\cdots$。若以 A_k 表示第 k 次射击命中目标的事件，且令 $X_k=I_{A_k}(k=1,2,\cdots)$，则 $\{X_k \mid k \in \mathbb{N}\}$ 为一列全体独立的以 p 为参数的 Bernoulli 随机变量，且

$$(X=n)=\left(X_1=0, \cdots, X_{n-1}=0, X_n=1\right), \quad n=1,2,\cdots.$$

下记 $q=1-p$，由

$$\left(X_k=1\right)=\left\{\omega \mid X_k(\omega)=1\right\}=A_k, \quad\left(X_k=0\right)=\left\{\omega \mid X_k(\omega)=0\right\}=A_k^c$$

则由事件序列的 $\{A_k, k\in\mathbb{N}\}$ 的全体独立性立即得对于 $n=1,2,\cdots$ 有

$$P(X=n)=P\left(X_1=0,\cdots,X_{n-1}=0,X_n=1\right)=P\left(A_1^c\cdots A_{n-1}^cA_n\right)=(1-p)^{n-1}p$$

则该概率分布为几何分布

例题 5.11 （几何分布/实际推断原理）

说明"在重复独立试验中，小概率事件必会发生"

解 将某小概率事件在试验中发生概率记为 p。对每次试验结果都用一个 Bernoulli 随机变量表示：若该事件发生，则该 Bernoulli 随机变量的值为 1；若该事件不发生，则令其为 0。由于试验独立重复则历次试验结果形成一个定义在同一概率空间上的全体独立的参数为 p 的 Bernoulli 随机变量序列。将该事件首次发生时的试验次数记为 Y，则由 (5.10)Y 的概率分布为几何分布 $\mathrm{Geo}(p)$。而事件会发生，即指 Y 等于某个正整数 n，亦即 $Y<\infty$。记 $q=1-p$，则显然有

$$P(Y<\infty)=\sum_{n=1}^{\infty}P(Y=n)=\sum_{n=1}^{\infty}p(1-p)^{n-1}=1$$

则不论 p 为多么小的正数，事件 $(Y<\infty)$ 发生的概率均等于 1

例题 5.12 （几何分布/二项分布）

设 100 张签中有 5 张可以中奖，3 人参加抽签，每人有放回地抽取一张，试计算有人中奖的概率

解

方法一：几何分布

有放回抽取等价于独立重复 Bernoulli 试验。若以 X 表示第一张有奖签出现时的抽取次数，则 X 的概率分布为几何分布 $\mathrm{Geo}(0.05)$。当 $X\leqslant 3$ 时有人中奖，则有人中奖的概率为 $P(X\leqslant 3)=1-P(X>3)=1-0.95^3$

方法二：二项分布

若以 Y 表示所取出的 3 张签中"有奖签"的张数，则 Y 的概率分布为二项分布 $B(3,0.05)$。当 $Y\neq 0$ 时有人中奖，则有人中奖的概率为 $P(Y\neq 0)=1-P(Y=0)=1-0.95^3$

5.2.5 Pascal 分布（负二项分布） Pascal Distribution(Negative Binomial Distribution)

定义 5.17 (Pascal 分布)

设离散随机变量 ξ，其取值为至多可数集，$0<p<1$ 且有

$$P(X=n)=\begin{pmatrix} n-1 \\ r-1 \end{pmatrix}p^rq^{n-r}$$

则称该随机变量概率分布为参数为 p 和 r 的 Pascal 分布 (Pascal distribution) ♣

注 （Pascal 分布）

参数为 p 和 1 的 Pascal 分布即为几何分布。不仅可以通过几何分布为第一次取得成功时的试验次数的分布，还可通过在 Pascal 中令 $r=1$ 得到几何分布 $\mathrm{Geo}(p)$ 的分布律

命题 5.5 (Pascal 分布定义正确性)

设概率空间 $\{\Omega,\mathcal{A},P\}$ 上定义一列全体独立的以 p 为参数的 Bernoulli 随机变量 $\{X_k|k\in\mathbb{N}\}$，记 $q=1-p$，则 $P(X=n)=\mathrm{C}_{n-1}^{r-1}p^rq^{n-r}$ 为离散型随机变量的概率分布 ♠

证明 对每个正整数 $n\geqslant r$ 有，当且仅当"在随机变量 $X_1,\cdots,X_{n-1}$ 中有 $r-1$ 个等于 1，其余为 0，且 X_n 等于 1"时有 $X=n$，则事件 $(X=n)$ 等价于"在事件 $A_1,\cdots,A_{n-1}$ 中恰有 $r-1$ 个发生并且事件 A_n 发生"，由于事件 $A_1,\cdots,A_n$ 全体独立，则有 $P(X=n)=\mathrm{C}_{n-1}^{r-1}p^rq^{n-r}$，若记

$$p_n=\mathrm{C}_{n-1}^{r-1}p^rq^{n-r},\quad n=r,r+1,\cdots$$

则显然有

$$\sum_{n=r}^{\infty} p_n = \sum_{n=r}^{\infty} \mathrm{C}_{n-1}^{r-1} p^r q^{n-r} = p^r \sum_{k=0}^{\infty} \mathrm{C}_{r+k-1}^{r-1} q^k = p^r (1-q)^{-r} = 1$$

则上述概率分布的确为一个离散型随机变量的分布律

注 该命题证明过程表明，Pascal 分布可以用负二项展开式中的各项表示，所以又称 Pascal 分布为**负二项分布** (negative binomial distribution)

例题 5.13 （Pascal 分布第一模型）

若以 Z_1 表示出现第 1 次成功时的试验次数，以 Z_2 表示自第 1 次成功之后到出现第 2 次成功之间的试验次数，$\cdots$，以 Z_r 表示自第 $r-1$ 次成功之后到出现第 r 次成功之间的试验次数，则它们均为表示取得一次成功所需的试验次数，即概率分布均为几何分布 $\mathrm{Geo}(p)$。

又由于各次试验相互独立，则 $Z_j, j = 1, 2, \cdots, r$ 为定义在同一概率空间上的既相互独立又服从相同分布的(简记为 i.i.d.)随机变量。因此对于表示取得第 r 次成功时的总试验次数的随机变量 X 就有 $X = Z_1+Z_2+\cdots+Z_r$ ，则对每个正整数 $n \geqslant r$ 有

$$(X = n) = (Z_1 + Z_2 + \cdots + Z_r = n)$$

注意到，Z_j 均为只取正整数值的随机变量，因此

$$(X = n) = \bigcup_{(n_1, n_2, \cdots, n_r)} (Z_1 = n_1, Z_2 = n_2, \cdots, Z_r = n_r)$$

其中右端为对满足关系式 $n_1 + n_2 + \cdots + n_r = n$ 的所有有序正整数组 $(n_1, n_2, \cdots, n_r)$ 求并，且右端事件两两不交

由“把 n 个相同的小球分入 r 个不同的盒子，且无空盒出现”的计数原理即得共有 C_{n-1}^{r-1} 组。而对于每一个这样的有序正整数组 $(n_1, n_2, \cdots, n_r)$ 都有

$$P(Z_1 = n_1, Z_2 = n_2, \cdots, Z_r = n_r) = \prod_{j=1}^{r} P(Z_j = n_j) = p^r (1-p)^{n-r} = p^r q^{n-r}$$

综上即得 $P(X = n) = C_{n-1}^{r-1} p^r q^{n-r}$ ，即为 Pascal 分布

例题 5.14 （Pascal 分布第二模型）

试求在独立重复 Bernoulli 试验中，第 n 次成功发生在第 m 次失败之前的概率

解 以 X 表示取得第 n 次成功时的试验次数，则 X 的概率分布为参数为 p 和 n 的 Pascal 分布，其中 p 为每次 Bernoulli 试验中取得成功的概率。以 A 表示“第 n 次成功在第 m 次失败之前发生”的随机事件。显然 A 等价于“在少于 $n+m$ 次试验中取得了 n 次成功”，则

$$P(A) = P(X < n+m) = \sum_{k=n}^{n+m-1} P(X = k) = \sum_{k=n}^{n+m-1} \mathrm{C}_{k-1}^{n-1} p^n q^{k-n}$$

即为所求概率

例题 5.15 （Pascal 分布/分赌注问题）

甲、乙二人乒乓球水平相同，约定 5 局 3 胜，胜者得奖金 800 元。现因故在甲胜了第一局后终止比赛，试问应当如何分配奖金？

解 合理的方案为按照“若把球打完，甲、乙二人各自取胜的概率”的比例来分配奖金。由于甲已先胜一局，则甲取胜的事件为“在接下来的比赛中第 3 次失败之前获得两次成功”。以 X 表示甲取得两次成功所需的局数，则 X 的概率分布为参数为 $p = 0.5$ 和 $r = 2$ 的 Pascal 分布，则甲赢乙的概率为

$$P(X < 5) = p^2 + 2p^2 q + 3p^2 q^2 = \frac{11}{16}$$

而由甲赢乙的对立事件为乙赢甲，则乙赢甲的概率为 $\dfrac{5}{16}$，则应按 $11:5$ 的比例分配奖金，故甲得 550 元，乙得 250 元。

注 该问题起源于概率论发展史上的一个著名的所谓“分赌注问题”。1654 年，一个职业赌徒向法国大数学家 Pascal 提出问题：甲、乙二人各下赌注 d 元，约定先胜 3 局者赢得全部赌金。假设在每一局中两人获胜机会相

等且各局输赢相互独立。若甲获胜一局而乙尚未获胜时赌博被迫终止，应当如何分配赌金？而 Pascal 给出的就是这里的分配方案

例题 5.16 （Pascal 分布/Banach 火柴问题）

某人口袋里放有两盒火柴，每盒装有火柴 n 根。每次随机取出一盒并从中拿出一根火柴使用。试求取出一盒发现已空，而此时另一盒中剩余 r 根火柴的概率

解 以 A 表示甲盒已空、乙盒中尚余 r 根火柴的事件，由对称性即得所求概率等于 $2P(A)$。将每取出甲盒一次视为取得一次成功，以 X 表示取得第 $n+1$ 次成功时的取盒次数，则 X 的概率分布为参数为 $p=0.5$ 和 $n+1$ 的 Pascal 分布。易知事件 A 发生，当且仅当 X 等于 $2n-r+1$，则

$$2P(A)=2P(X=2n-r+1)=\begin{pmatrix}2n-r\\n\end{pmatrix}2^{r-2n}$$

即为所求概率

5.3 连续型随机变量 Continous Random Variable

5.3.1 绝对连续分布与奇异分布 Absolutely Continuous Distribution and Singular Distribution

定义 5.18 (绝对连续分布)

设随机变量 ξ 定义在概率空间 $(\Omega,\mathcal{B},P)$ 上，若存在实值函数 $\rho_\xi(x)$ 满足

$$(\forall B\in\mathcal{B}):P_\xi(B)=P(\xi\in B)=\int_B\rho_\xi(x)dx$$

则称随机变量 ξ 的概率分布为绝对连续分布 (absolutely continuous distribution/абсолютно непрерывное распределение)，或连续概率分布 (continuous probability distribution) 简称连续型分布，连续分布，称随机变量 ξ 为连续型随机变量 (continuous random variable) 简称连续随机变量，称函数 $\rho_\xi(x)$ 为随机变量 ξ 的概率分布密度简称分布密度 (distribution density /плотность распределения) 或概率密度函数 (probability density function) 简称密度函数，经常利用英语缩写记为 p.d.f. ♣

注 （绝对连续分布）

俄罗斯教材中，有时也称绝对连续分布为**简单连续分布** (просто непрерывное распределение)

注 （概率密度函数）

随机变量 X 的概率密度函数 $p(x)$ 在 $x=x_0$ 处的值 $p\left(x_0\right)$ 反映了该随机变量在 x_0 取值的概率大小。由定义即知

$$P\left(x_0<X\leqslant x_0+\Delta x\right)=F\left(x_0+\Delta x\right)-F\left(x_0\right)=\int_{x_0}^{x_0+\Delta x}p(u)\mathrm{d}u\approx p\left(x_0\right)\Delta x$$

由此概率密度函数类似于离散随机变量的概率质量函数，因此也有教材把概率密度函数和概率质量函数统称为**概率（分布）函数** (probability(distribution)function)

定理 5.5 (随机变量分布函数与密度函数关系)

设随机变量 X 的分布函数 F 连续，数集 A 中任意两点之间的距离大于 $\delta>0$。若在 A 外导数 $F'(x)$ 存在且连续，则有

$$f(x)=\begin{cases}F'(x), & x\notin A,\\0, & x\in A\end{cases}$$

为 X 的密度函数

证明 对任意 $a<b$，不妨设 $(a;b)$ 中只有集合 A 中的 $a_1,a_2,\cdots,a_k$ 并且

$$a_0=a<a_1<a_2<\cdots<a_k<b=a_{k+1}$$

这时有

$$\begin{aligned}P(a<X\leqslant b)&=\sum_{j=0}^{k}P\left(a_j<X\leqslant a_{j+1}\right)=\sum_{j=0}^{k}\left[F\left(a_{j+1}\right)-F\left(a_j\right)\right]\\&=\sum_{j=0}^{k}\int_{a_j}^{a_{j+1}}f(t)\mathrm{d}t=\int_a^b f(t)\mathrm{d}t\end{aligned}$$

则 $f(x)$ 为 X 的概率密度

注 当 A 仅有有限个极限点时也可证明该结论。该定理中 F 连续的条件至关重要，尽管 F 连续不能保证 X 为连续随机变量，但 F 不连续能保证 X 不是连续随机变量。当 F 为二项随机变量的分布函数时，除去有限个点外 $F'(x)=0$，则 X 的概率密度不存在

性质 （绝对连续分布性质）

设随机变量 ξ 具有绝对连续分布，分布密度为 $\rho_\xi(x)$，则有下列性质成立：

1) （非负性）$(\forall x\in\mathbb{R}^1):\rho_\xi(x)\geqslant 0$

2) （规范性）

$$\int_{-\infty}^{\infty}\rho_\xi(x)dx=1$$

3) （分布函数表示）

$$(\forall x\in\mathbb{R}^1):F_\xi(x)=\int_{-\infty}^{x}\rho_\xi(y)dy$$

相反地有

$$\rho_\xi(x)=\frac{d}{dx}F_\xi(x)$$

连续

4) （分布密度基本等式）

$$P(a\leqslant\xi<b)=F_\xi(b)-F_\xi(a)=\int_a^b\rho_\xi(x)dx$$

5) （零测度性）$(\forall x\in\mathbb{R}^1):P(\xi=x)=0$

注 实际上性质 1)2) 为分布密度的特征性质，也可以作为分布密度的定义

定义 5.19 (奇异分布)

设随机变量 ξ，若存在一个不可数 Borel 子集 $B\subset\mathbb{R}^1$，其 Lebesgue 测度为零，使得 $P(\xi\in B)$，则称随机变量 ξ 的分布为奇异分布 (singular distribution/сингулярное распределение)

注 （奇异分布）

本文不考虑奇异分布。此外，奇异分布在实用中并不常见

例题 5.17 （奇异分布/Cantor 集）

定义函数 $F(x)$，令 $F(x)=0,x<0;F(x)=1,x\geqslant 1$，再令

$$F(x)=\frac{1}{2},\quad\frac{1}{3}<x<\frac{2}{3}$$

然后令

$$F(x)=\begin{cases}\frac{1}{4}, & \frac{1}{9}<x<\frac{2}{9},\\ \frac{3}{4}, & \frac{7}{9}<x<\frac{8}{9}.\end{cases}$$

接着再令

$$F(x)=\begin{cases}\frac{1}{8}, & \frac{1}{27}<x<\frac{2}{27},\\ \frac{3}{8}, & \frac{7}{27}<x<\frac{8}{27},\\ \frac{5}{8}, & \frac{19}{27}<x<\frac{20}{27},\\ \frac{7}{8}, & \frac{25}{27}<x<\frac{26}{27}.\end{cases}$$

以此类推每次都把 $(0;1)$ 区间中未定义的每个小区间 3 等分，在中间一个开区间上按递增原则将 $F(x)$ 定义为常数，则在区间 $(0;1)$ 中的 Cantor 集中的每一点上 $F(x)$ 都有了定义。最后，再按右连续原则定义出其余点上的 $F(x)$ 的值

上述定义的 $F(x)$ 显然为一个分布函数。但当 $x<0,x>1$，以及当 x 属于 $(0;1)$ 中的 Cantor 集时，$F(x)$ 都可导且导数都为 0，由于 $(0;1)$ 中的 Cantor 集的 Lebesgue 测度等于 0，则 $F(x)$ 的导数几乎处处为 0，从而 $F(x)$ 不能由它的导函数积分得到，即为一个奇异分布

注 此外，可以证明 $F(x)$ 为一个连续函数

定义 5.20 (混合分布)

设有限或可数集 $X=\{x_1,x_2,\cdots\}$，$\{p_k\}$ 为某个正数集，而 $\rho_\xi(x)$ 为非负函数，若

$$P_\xi(B)=P(\xi\in B):=\sum_{x_k\in B}p_k+\int_B\rho_\xi(x)dx$$

则称随机变量 ξ 的分布为混合分布 (mixture distribution/смешанное распределение)，称 $(X,\{p_k\})$ 为混合分布的离散分量 (дискретный компонент)，而称分布密度 $\rho_\xi(x)$ 为混合分布的连续分量 (непрерывный компонент)。若混合分布仅包含一个分量，则称其为纯净分布 (чистое распределение) 分别称数

$$\alpha_d:=\sum_k p_k,\quad \alpha_c:=\int_{-\infty}^{\infty}\rho_\xi(x)dx$$

为离散分量及连续分量的权重 (weight/вес)，显然 $\alpha_d,\alpha_c\geqslant 0,\alpha_d+\alpha_c=1$ ♣

例题 5.18 （混合分布）

从区间 $[0;1]$ 中随机选择点 ω，得到随机变量

$$\xi=\begin{cases}0, & \omega<\frac{1}{4}\\ \omega-\frac{1}{4}, & \frac{1}{4}\leqslant\omega<\frac{3}{4}\\ \frac{1}{2}, & \omega\geqslant\frac{3}{4}\end{cases}$$

则随机变量 ξ 有两个离散点 $x_1=0$ 与 $x_2=\frac{1}{2}$，其对应概率为 $p_1=P(\xi=0)=\frac{1}{4},p_2=P(\xi=\frac{1}{2})=\frac{1}{4}$ 且在 $\left[\frac{1}{4};\frac{3}{4}\right]$ 上均匀分布 (равномерное распределение)，则有当 $x\in\left[\frac{1}{4};\frac{3}{4}\right]$ 时 $\rho_\xi(x)=1$ 。分量权重为 $\alpha_d=\frac{1}{2}$ 与 $\alpha_c=\frac{1}{2}$

注 （Lebesgue 分解理论）

根据实变函数论中的 Lebesgue 分解理论，任意一个（一元）分布函数 $F(x)$ 都具有如下形式的分解式：

$$F(x)=a_1F_1(x)+a_2F_2(x)+a_3F_3(x),\quad x\in\mathbb{R}$$

其中 $F_1(x),F_2(x),F_3(x)$ 分别为离散型、绝对连续型和奇异型分布函数，$a_j\geqslant 0$ 且 $a_1+a_2+a_3=1$

因此也有，当 a_1,a_2,a_3 中至少有两者不为 0 时，称 $F(x)$ 为混合型的分布函数；当其中仅有一者不为 0（一定为 1）时，称 $F(x)$ 为纯净型的分布函数

5.3.2 随机变量的截尾
Truncation of Random Variable

定义 5.21 (无界随机变量)

设 X 为随机变量，若 $\forall M>0:P(X>M)>0$，则称 X 上无界；若 $\forall M>0:P(X<-M)>0$，则称 X 下无界 ♣

定义 5.22 (限尾随机变量)

设 X 为随机变量（一般为上下无界的随机变量），$a<b$，若

$$Y=\begin{cases}a, & X<a,\\ X, & a\leqslant X\leqslant b,\\ b, & X>b.\end{cases}$$

则称 Y 为 X 的限尾随机变量，称这样的变换为随机变量 X 的限尾 ♣

注（限尾随机变量）

可用示性函数 I 将 X 的限尾随机变量 Y 表示为

$$Y=aI(X<a)+XI(a\leqslant X\leqslant b)+bI(X>b)$$

不过通常采用对称的限尾方式，即对某个 $c>0$，取 $a=-c,b=c$，此时有

$$\begin{aligned}Y&=-cI(X<-c)+XI(-c\leqslant X\leqslant c)+cI(X>c)\\&=-cI(X<-c)+XI(|X|\leqslant c)+cI(X>c)\end{aligned}$$

另外，若 $F_X(x)$ 原来为连续型分布，则 $F_Y(x)$ 变为混合型分布

定义 5.23 (切尾随机变量)

设 X 为无界随机变量，$a<b$，若

$$Y=XI(a\leqslant X\leqslant b)$$

则称 Y 为 X 的切尾随机变量，称这样的变换为随机变量 X 的切尾 ♣

注（切尾随机变量）

通常对某个 $c>0$，取 $a=-c,b=c$。对 $c>0$，将随机变量 X 在 $\pm c$ 处切尾后所得的随机变量 Y 的分布函数 $F_Y(x)$ 与 X 的分布函数 $F_X(x)$ 满足

$$F_Y(x)=\begin{cases}0, & x<-c,\\ F_X(x)-F_X(-c-0), & -c\leqslant x<0,\\ F_X(x)+1-F_X(c), & 0\leqslant x<c,\\ 1, & x\geqslant c.\end{cases}$$

因此通常 $F_Y(x)$ 在 $x=0$ 处有一个跃度:

$$F_Y(0)-F_Y(0-0)=F_X(0)-F_X(0-0)+F_X(-c-0)+1-F_X(c)$$

注 (截尾随机变量)

通常把限尾随机变量和切尾随机变量都称为**截尾随机变量** (truncated random variable), 称这样的变换为随机变量 X 的**截尾** (truncation)

5.3.3 均匀分布与 Cauchy 分布
Uniform Distribution and Cauchy Distribution

定义 5.24 (均匀分布)

设随机变量 X, 若随机变量 X 的分布函数为

$$F(X)=P(X\leqslant x)=\begin{cases}0, & x<0\\ x, & 0\leqslant x<1\\ 1, & x\geqslant 1\end{cases}$$

则称随机变量 X 的概率分布为区间 $[0;1]$ 上的均匀分布 (uniform distribution/равномерное распределение), 记为 $U[0;1]$。若一个随机变量 X 具有该形式的分布函数 $F(x)$, 则称 X 服从区间 $[0;1]$ 上的均匀分布, 简记为 $X\sim U[0;1]$ ♣

例题 5.19 (均匀分布第一模型/Bernoulli 随机变量)

设 $\{X_n|n\in\mathbb{N}\}$ 为某个概率空间 $(\Omega,\mathcal{A},P)$ 中的一列相互独立的同以 $p=\dfrac{1}{2}$ 为参数的 Bernoulli 随机变量, 令

$$Z=\sum_{n=1}^{\infty}\frac{X_n}{2^n}$$

试证 $Z\sim U[0;1]$

解 首先, 对于每个 $\omega\in\Omega$, 级数 $\sum\limits_{n=1}^{\infty}\dfrac{X_n(\omega)}{2^n}$ 均为非负级数且有

$$\sum_{n=1}^{\infty}\frac{X_n(\omega)}{2^n}\leqslant\sum_{n=1}^{\infty}\frac{1}{2^n}=1$$

则级数在 Ω 中处处收敛, 即对每个 $\omega\in\Omega, Z(\omega)=\sum_{n=1}^{\infty}\dfrac{X_n(\omega)}{2^n}$ 都有定义且 $0\leqslant Z(\omega)\leqslant 1$

记部分和序列 $Z_m=\sum\limits_{n=1}^{m}\dfrac{X_n}{2^n}$, 则有

$$Z_m(\omega)=\sum_{n=1}^{m}\frac{X_n(\omega)}{2^n}\uparrow Z(\omega),\quad m\to\infty,\quad \forall\omega\in\Omega$$

简记为

$$Z_m=\sum_{n=1}^{m}\frac{X_n}{2^n}\uparrow Z,\quad m\to\infty$$

则对任意 $x\in\mathbb{R}$ 都有 $(Z_m\leqslant x)\supset(Z_{m+1}\leqslant x), m=1,2,\cdots$, 则有

$$(Z\leqslant x)=\lim_{m\to\infty}(Z_m\leqslant x)=\bigcap_{m=1}^{\infty}(Z_m\leqslant x)$$

再由概率上连续性即知

$$P(Z\leqslant x)=P\left\{\lim_{m\to\infty}(Z_m\leqslant x)\right\}=\lim_{m\to\infty}P(Z_m\leqslant x)$$

注意到, Z_m 为一个取 2^m 个不同值 $\left\{0,\dfrac{1}{2^m},\dfrac{2}{2^m},\cdots,\dfrac{2^m-1}{2^m}\right\}$ 的离散型随机变量, 且取每个值的概率均为

2^{-m}，则

$$(0 \leqslant x \leqslant 1): P\left(Z_m \leqslant x\right)=\sum_{k \leqslant 2^m x} P\left(2^m Z_m=k\right)=\frac{\left[2^m x\right]}{2^m}$$

其中 $[2^m x]$ 为不大于 $2^m x$ 的最大整数，则 $\lim\limits_{m \to \infty} \dfrac{[2^m x]}{2^m}=x$，则

$$(0 \leqslant x \leqslant 1): P(Z \leqslant x)=\lim _{m \rightarrow \infty} P\left(Z_m \leqslant x\right)=\lim _{m \rightarrow \infty} \frac{\left[2^m x\right]}{2^m}=x$$

则 $Z=\sum\limits_{n=1}^{\infty} \dfrac{X_n}{2^n}$ 为一个 $U[0;1]$ 随机变量，且 $(Z \leqslant x)$ 为可数个事件 $\{(Z_m \leqslant x)\,|m \in \mathbb{N}\}$ 的交集，则也为一个随机事件

定义 5.25 (均匀分布)

设 $a<b$，若分布函数 $F(x)$ 具有密度函数

$$p(x)=\frac{1}{b-a}, \quad a \leqslant x \leqslant b$$

则称该分布为区间 $[a;b]$ 上的均匀分布 (uniform distribution)，记为 $U[a;b]$

$U[a;b]$ 的分布函数为

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

♣

注 （均匀分布）

由于 $U[a;b]$ 的分布函数 $F(x)$ 连续，因此无论怎样改变自变量 x 在区间 $[a;b]$ 的端点 a 和 b 处的开闭情况，所得到的 $F(x)$ 都一样，因此无须考虑在区间的开闭问题

例题 5.20 （Cauchy 分布第一模型）

设随机变量 X 概率分布为均匀分布 $U\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, $Y=\tan X$，试求随机变量 Y 的分布函数和密度函数

解 易知，对任意实数 x 有

$$F_Y(x)=P(Y \leqslant x)=P(\tan X \leqslant x)=P(X \leqslant \arctan x)=\int_{-\frac{\pi}{2}}^{\arctan x} \frac{1}{\pi} \mathrm{d} t=\frac{1}{\pi} \arctan x+\frac{1}{2}$$

即为 Y 的分布函数。由此即得

$$p_Y(x)=\frac{\mathrm{d}}{\mathrm{d} x} F_Y(x)=\frac{1}{\pi \cdot\left(1+x^2\right)}, \quad x \in \mathbb{R}$$

即为 Y 的密度函数

定义 5.26 (Cauchy 分布)

设连续随机变量 X，若其密度函数为

$$p(x)=\frac{1}{\pi} \frac{1}{1+x^2}$$

则称对应分布为 Cauchy 分布 (Cauchy distribution/распределение Коши)

更一般地，若其密度函数为

$$p(x)=\frac{1}{\pi} \frac{\lambda^2}{\lambda^2+(x-u)^2}$$

则称对应分布为参数为 $\lambda>0, \mu \in \mathbb{R}$ 的 Cauchy-Lorenz 分布 (Cauchy-Lorenz distribution) 或 Breit-Wigner 分布 (Breit–Wigner distribution)

♣

5.3.4 随机变量的变换 Transformation of Random Variable

推论 5.1 (分布函数充分条件推论)
若随机变量 Y 服从均匀分布 $U(0;1)$，则对任意一元分布函数 $F(x)$，随机变量 $X=F^{-1}(Y)$ 服从分布 $F(x)$，其中 $F^{-1}(u)(u\in(0,1))$ 为 $F(x)$ 的形式反函数

证明 由命题 (5.3) 即证

定理 5.6 (随机变量连续分布函数必要条件)
若随机变量 X 的分布函数 $F_X(x)$ 连续，则随机变量 $Y=F_X(X)$ 服从均匀分布 $U(0;1)$

证明 显然有 $0\leqslant Y=F_X(X)\leqslant 1$ 。由 $F_X(x)$ 连续，则 $Y=F_X(X)$ 为连续随机变量，且对 $t\in(0;1)$ 有

$$F_X^{-1}(t)=\inf\{x\mid F_X(x)\geqslant t\}\geqslant y\Longleftrightarrow F_X(y)\leqslant t$$
$$F_X\left(F_X^{-1}(t)\right)=\inf\{F_X(x)\mid F_X(x)\geqslant t\}=t$$

则有

$$P(Y\leqslant t)=P\left(F_X(X)\leqslant t\right)=P\left(F_X^{-1}(t)\geqslant X\right)=P\left(X\leqslant F_X^{-1}(t)\right)=F_X\left(F_X^{-1}(t)\right)=t$$

因此 $Y=F_X(X)$ 服从均匀分布 $U(0;1)$

定理 5.7 (连续随机变量反函数密度函数)
设随机变量 ξ 服从连续分布，且其对应密度函数为 $\rho_\xi(x)$，另设 $y=g(x)$ 为严格单调可微函数，则随机变量 $\eta=g(\xi)$ 服从连续分布，且其对应密度函数为

$$\rho_\eta(y)=\rho_\xi\left(g^{-1}(y)\right)\left|\frac{d}{dx}g\left(g^{-1}(y)\right)\right|^{-1}$$

证明 仅对 $y=g(x)$ 严格递增的情形证明。设 $y\in R^1$ 为满足 $0<P(\eta<y)<1$ 的点。这时 $F_\eta(y)=P(\eta<y)=P(g(\xi)<y)=P\left(\xi<g^{-1}(y)\right)=F_\xi\left(g^{-1}(y)\right)$

利用复变函数和反函数的微分法则即得

$$\rho_\eta(y)=\frac{d}{dy}F_\eta(y)=\rho_\xi\left(g^{-1}(y)\right)\left(\frac{d}{dx}g\left(g^{-1}(y)\right)\right)^{-1}$$

而严格递减函数情形同理，定理即证

注 对于非单调函数的情况，也可以给出类似的定理。但这种结果在实际问题中很少使用。在每种特定情况下重新进行所有计算通常更容易。若 ξ 服从连续分布，其对应密度函数为 $\rho_\xi(x)$，则随机变量 $\eta=g(\xi)$ 的分布可能连续也可能离散

例题 5.21 （连续随机变量反函数密度函数）

设函数 $g(x)=a+bx,x\in\mathbb{R},b\neq 0$，$X$ 为随机变量，随机变量 $Y=g(X)$。试分别对 (1) 随机变量 X 的概率分布为均匀分布 $U(0;1)$；(2) 随机变量 X 的概率分布为标准正态分布 $\mathcal{N}(0,1)$，求 Y 的密度函数

解 1) 当 X 的概率分布为均匀分布 $U(0;1)$ 时，由定理 (5.7) 得，Y 为连续随机变量。若 $b>0$，则密度函数为

$$p_Y(x)=p_X\left(\frac{x-a}{b}\right)\frac{1}{b}=\frac{1}{b},\quad x\in(a;a+b)$$

即 Y 服从均匀分布 $U(a;a+b)$，若 $b<0$，则密度函数为

$$p_Y(x)=p_X\left(\frac{x-a}{b}\right)\frac{1}{|b|}=\frac{1}{|b|},\quad x\in(a+b;a)$$

即 Y 服从均匀分布 $U(a+b;a)$

2) 当 X 的概率分布为标准正态分布 $\mathcal{N}(0,1)$ 时，由定理 (5.7) 得，Y 为连续随机变量且密度函数为

$$p_Y(x)=p_X\left(\frac{x-a}{b}\right)\frac{1}{|b|}=\frac{1}{\sqrt{2\pi}|b|}\exp\left\{-\frac{(x-a)^2}{2b^2}\right\},\quad x\in\mathbb{R}$$

则不论 $b>0$ 或 $b<0$，都有 Y 服从正态分布 $\mathcal{N}\left(a,b^2\right)$

命题 5.6 (随机变量反函数分布函数)

设随机变量 ξ，而 $y=g(x)$ 为严格单调函数，则对于 $\eta=g(\xi)$，若 $g(x)$ 严格递增则有

$$F_\eta(y)=F_\xi\left(g^{-1}(y)\right)$$

而若 $g(x)$ 严格递减则有

$$F_\eta(y)=1-F_\xi\left(g^{-1}(y)+0\right)$$

证明 仅证明严格递增的情况。由分布函数定义即得

$$F_\eta(y)=P(\eta<y)=P(g(\xi)<y)=P\left(\xi<g^{-1}(y)\right)=F_\xi\left(g^{-1}(y)\right)$$

严格递减情况类似

例题 5.22 (随机变量反函数分布函数)

设随机变量 X 的概率分布为标准正态分布 $\mathcal{N}(0,1), Y=X^2$，试求 Y 的分布函数和密度函数

解 显然 Y 为非负随机变量，则有 $F_Y(x)=0, x\leqslant 0$。当 $x>0$ 时有

$$F_Y(x)=P(Y\leqslant x)=P\left(X^2\leqslant x\right)=P(-\sqrt{x}\leqslant X\leqslant\sqrt{x})=\Phi(\sqrt{x})-\Phi(-\sqrt{x})$$

则有

即为 Y 的密度函数

例题 5.23 (随机变量变换)

设随机变量 X 的概率分布为连续分布，其分布函数为 $F_X(x)$，对应的密度函数为 $p_X(x)$，$Y=\sin X$，求 Y 的分布函数和密度函数

解 显然，$-1\leqslant Y\leqslant 1$，则仅需对 $x\in[-1;1]$ 考虑 $F_Y(x)$ 和 $p_Y(x)$ 。设 $-1<x<1$，则有

$$\begin{aligned}F_Y(x)=P(Y\leqslant x)&=P(-1\leqslant\sin X\leqslant x)\\&=\sum_{k=-\infty}^{\infty}P((2k-1)\pi-\arcsin x\leqslant X\leqslant 2k\pi+\arcsin x)\\&=\sum_{k=-\infty}^{\infty}\{F_X(2k\pi+\arcsin x)-F_X((2k-1)\pi-\arcsin x)\}\end{aligned}$$

则

$$\begin{aligned}p_Y(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_Y(x)&=\sum_{k=-\infty}^{\infty}\left(\frac{\mathrm{d}}{\mathrm{d}x}F_X(2k\pi+\arcsin x)-\frac{\mathrm{d}}{\mathrm{d}x}F_X((2k-1)\pi-\arcsin x)\right)\\&=\sum_{k=-\infty}^{\infty}\frac{1}{\sqrt{1-x^2}}\{p_X(2k\pi+\arcsin x)+p_X((2k-1)\pi-\arcsin x)\}\end{aligned}$$

即为密度函数

例题 5.24 (Weibull[1]分布/随机变量变换)

设随机变量 ξ 为参数为 $\lambda>0$ 的指数分布，则其为非负随机变量。当 $x>0$ 时对应密度函数为 $\rho_\xi(x)=\lambda e^{-\lambda x}$，则 $x>0$ 时的分布函数为

$$F(x)=P(\xi<x)=\int_0^x\lambda e^{-\lambda y}dy=1-e^{-\lambda x}$$

[1]瓦洛迪 · 威布尔 (Waloddi Weibull,1887.6.18-1979.10.12) 瑞典工程师，数学家。美国机械工程师协会于 1972 年授予 Weibull 金质奖章。瑞典国王 Carl XVI Gustaf 于 1978 年亲自颁发给 Weibull 瑞典皇家工程科学院的伟大金质奖章

设 $\alpha > 0$，考虑随机变量 $\eta = \xi^{\frac{1}{\alpha}}$，其为非负随机变量，则仅需考虑 $y > 0$ 的分布，这时

$$F_\eta(y) = P(\eta < y) = P\left(\xi^{\frac{1}{\alpha}} < y\right) = P\left(\xi < y^\alpha\right) = 1 - \exp\left(-\lambda y^\alpha\right)$$

再微分即得新的随机变量 η 的密度函数 $\rho_\eta(y) = \lambda\alpha y^{\alpha-1}e^{-\lambda y^\alpha}$

定义 5.27 (Weibull 分布)

设随机变量 X，若 X 具有密度函数

$$\rho_\eta(x) = \lambda\alpha x^{\alpha-1}e^{-\lambda x^\alpha}$$

则称随机变量 X 的概率分布为参数为 (λ, α) 的 Weibull 分布 (Weibull distribution/распределение Вейбулла)

♣

注 （Weibull 分布）

Weibull 分布常用于排队论 (теория массового обслуживания)、可靠性理论 (теория надежности) 等应用领域中。1927 年，Fréchet 首先给出该分布的定义。1933 年，Rosin 和 Rammler 在研究碎末的分布时，第一次应用了 Weibull 分布[2]。1951 年，瑞典数学家 Weibull 详细解释了该分布，因此该分布便以他的名字命名

例题 5.25 （Lévy 分布/随机变量变换）

设随机变量 ξ 的概率分布为标准正态分布 $(a = 0, \sigma^2 = 1)$ 并定义新的随机变量 $\eta = |\xi|^{-2}$，由定义 η 为非负随机变量，这时对于 $x > 0$ 有

$$F_\eta(x) = P(\eta < x) = P\left(\frac{1}{\xi^2} < x\right) = P\left(|\xi| > \frac{1}{\sqrt{x}}\right) = 2 \cdot \left(1 - \Phi\left(\frac{1}{\sqrt{x}}\right)\right)$$

这时

$$\rho_\eta = \frac{d}{dx}F_\eta(x) = \frac{1}{\sqrt{2\pi}}x^{-\frac{3}{2}}e^{-\frac{1}{2x}}, x > 0$$

定义 5.28 (Lévy 分布)

设随机变量 X，若 X 具有密度函数

$$\rho_\eta = \frac{d}{dx}F_\eta(x) = \frac{1}{\sqrt{2\pi}}x^{-\frac{3}{2}}e^{-\frac{1}{2x}}, x > 0$$

则称随机变量 X 的概率分布为 Lévy 分布 (Lévy distribution/распределение Леви)

♣

5.3.5 指数分布与 Γ 分布
Exponential Distribution and Γ Distribution

英国著名统计学家 Pearson 在研究物理、生物及经济中的随机变量时发现很多连续型随机变量的分布不是正态分布．这些随机变量的特点是只取非负值，于是 Pearson 致力于这类随机变量的研究。从 1895 年至 1916 年间，Pearson 连续发表了一系列的连续分布密度曲线，认为这些曲线可以包括常见的单峰分布，其中就有 Γ 分布。在气象学中，干旱地区的年、季或月降水量被认为服从 Γ 分布，指定时间段内的最大风速等也被认为服从 Γ 分布

定义 5.29 (指数分布)

设 $\lambda > 0$，若随机变量 X 的概率密度为

$$p(x) = \lambda \mathrm{e}^{-\lambda x}, \quad x > 0$$

则称随机变量 X 的概率分布为参数为 λ 的指数分布 (exponential distribution/показательное

[2] Rosin, P.; Rammler, E. (1933), "The Laws Governing the Fineness of Powdered Coal", Journal of the Institute of Fuel 7: 29 - 36.

распределение)，简记为 Exp(λ)，其分布函数为

$$F(x)=\begin{cases}0, & -\infty<x\leqslant 0,\\ 1-\mathrm{e}^{-\lambda x}, & 0<x<\infty.\end{cases}$$

♣

定理 5.8 (非负指数分布无后效性)

设 X 为连续型非负随机变量，则系列命题等价:

1) (指数分布) X 的概率分布为指数分布

2) (无后效性) $(\forall s,t\geqslant 0):P(X>s+t\mid X>s)=P(X>t)$

♡

证明 1)⇒2) 设 $X\sim \mathrm{Exp}(\lambda)$，则有

$$P(X>x)=\int_x^{\infty}\lambda\mathrm{e}^{-\lambda s}\,\mathrm{d}s=\mathrm{e}^{-\lambda x}$$

又由条件概率即得

$$P(X>s+t\mid X>s)=\frac{P(X>s+t)}{P(X>s)}=\frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}}=\mathrm{e}^{-\lambda t}=P(X>t)$$

2)⇒1) 设 X 有无后效性，用 $f(x)$ 表示 X 的概率密度，则有

$$G(x)\overset{\text{def}}{=}P(X>x)=\int_x^{\infty}f(t)\mathrm{d}t$$

为连续函数。由无后效性得到 $G(s+t)=G(s)G(t)>0$，则 $G(x)$ 为指数函数 $G(x)=\mathrm{e}^{-\lambda x}$。注意到对于任意 $0\leqslant a<b$, 从

$$P(a<X\leqslant b)=P(X>a)-P(X>b)=\mathrm{e}^{-\lambda a}-\mathrm{e}^{-\lambda b}=\int_a^b\lambda\mathrm{e}^{-\lambda s}\,\mathrm{d}s$$

即得 X 的概率分布为指数分布

注 若 X 表示某仪器的工作寿命，无后效性可以解释为：当仪器工作了 s 小时后再继续工作 t 小时的概率等于该仪器刚开始就能工作 t 小时的概率，说明该仪器的使用寿命不随使用时间的增加发生变化。一般来说，电子元件和计算机软件等具备这种性质，它们本身的老化可以忽略不计

注 在应用概率中，习惯采用记号

$$\bar{F}(x)=1-F(x)=P(X>x),\quad x>0$$

并称 $\bar{F}(x)$ 为**生存函数** (survival function)，意为 X 在不小于 x 之后继续生存的概率

由上述计算可知，对于指数分布 Exp(λ) 及其服从该分布的随机变量 X，其生存函数为

$$\bar{F}(x)=P(X>x)=\mathrm{e}^{-\lambda x},\quad x>0$$

注 指数分布也是唯一的具有无后效性的非负连续分布。另外也指出，服从指数分布的随机变量只能取非负实数值

定义 5.30 (随机过程)

若当 $t\geqslant 0$ 时，以 X_t 表示在时刻 t 以前到来的质点数目，亦即表示在时间区间 $[0;t)$ 中到来的质点数目。可证明在一定的条件下，对任意 $t\geqslant 0$，随机变量 X_t 都服从 Poisson 分布，并且其参数与 t 有关。显然，对不同的 t，X_t 为不同的随机变量，则 $\{X_t|t\geqslant 0\}$ 为一族以 t 为参数的随机变量，称该随机变量族为随机过程

♣

注 (Poisson 过程)

可证明在一定的条件下随机质点流的数目形成 Poisson **过程**

例题 5.26 (Poisson 过程/指数分布/Γ 分布)

设随机质点流的计数过程 $\{X_t|t\geqslant 0\}$ 为强度为 λ 的 Poisson 过程，以 Z_r 表示第一个质点的到来时刻。易

知，对于 $t>0$，事件 $(Z_r>t)$ 表示第一个质点在时刻 t 以后到来，而事件 $(X_t=0)$ 表示到时刻 t 为止，到来的质点数目为 0，则上述两个事件相等，则有

$$P\left(Z_1>t\right)=P\left(X_t=0\right)=\mathrm{e}^{-\lambda t}, \quad t>0$$

其形式与参数为 λ 的指数分布的“生存函数”全等，则 Z_r 的分布即为参数为 λ 的指数分布

考虑第 r 个质点到来时刻的分布。若在强度为 λ 的 Poisson 过程 $\{X_t|t\geqslant 0\}$ 中，以 Z_r 表示第 r 个质点的到来时刻，则对于 $t>0$，事件 $(Z_r\leqslant t)$ 表示第 r 个质点的到来不迟于时刻 t，而事件 $(X_t\geqslant r)$ 表示到时刻 t 为止到来的质点数目不少于 r 个，则这两个事件相等，则有

$$F_r(t)=P\left(Z_r\leqslant t\right)=P\left(X_t\geqslant r\right)=\mathrm{e}^{-\lambda t}\sum_{k=r}^{\infty}\frac{(\lambda t)^k}{k!}, \quad t>0$$

上式右端为 t 的可导函数并可逐项求导。对其求导即得

$$p_r(t)=\frac{\mathrm{d}}{\mathrm{d}t}F_r(t)=\mathrm{e}^{-\lambda t}\sum_{k=r}^{\infty}\frac{\lambda^k t^k-1}{(k-1)!}-\mathrm{e}^{-\lambda t}\sum_{k=r}^{\infty}\frac{\lambda^{k+1}t^{lk}}{k!}=\mathrm{e}^{-\lambda t}\frac{\lambda^r}{(r-1)!}t^{r-1}, \quad t>0$$

利用 Γ 函数的定义，即

$$\Gamma(r)=\int_0^{\infty}x^{r-1}\mathrm{e}^{-x}\,\mathrm{d}x$$

该积分对一切 $r>0$ 收敛。特别地，对正整数 n 有 $\Gamma(n+1)=n!$，则 $p_r(t)$ 为一个概率密度函数，且可以将其记为

$$p(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, \quad x>0 \tag{5.4}$$

注 值得一提的是，若把参数 r 由正整数推广到任意正数，式 (5.4) 中的 $p(x)$ 仍然为一个概率密度函数。由此给出 Γ 分布的定义

定义 5.31 (Γ 分布)

称以

$$p(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, \quad x>0$$

作为密度函数的连续分布为以 $\lambda>0$ 和 $r>0$ 为参数的 Γ 分布 (Gamma distribution)，记为 $\Gamma(\lambda,r)$，其中

$$\Gamma(r)=\int_0^{\infty}x^{r-1}\mathrm{e}^{-x}\,\mathrm{d}x$$

为 Euler Γ 函数

♣

例题 5.27 （Γ 分布/指数分布）

对于参数 r 为正整数的 $\Gamma(\lambda,r)$ 分布，其产生背景为：在计数过程 $\{X_t|t\geqslant 0\}$ 为强度为 λ 的 Poisson 过程时，第 r 个质点到来的时间 Z_r 服从分布 $\Gamma(\lambda,r)$

若以 Y_1 表示上述质点流中第一个质点的到来时刻，以 Y_k 表示第 k 个质点与第 $k-1$ 个质点到来的时间间隔，则显然

$$Z_r=Y_1+\cdots+Y_r, \quad r\in\mathbb{N}$$

而利用计数过程 $\{X_t|t\geqslant 0\}$ 的平稳增量性和独立增量性可证明 $\{Y_k|k\in\mathbb{N}\}$ 为独立同分布的随机变量序列，则它们都与 Y_1 同分布，即服从指数分布 $\mathrm{Exp}(\lambda)$

则参数 r 为正整数时的 Γ 分布 $\Gamma(\lambda,r)$ 即为 r 个独立同分布的 $\mathrm{Exp}(\lambda)$ 随机变量之和的分布。

注 注意到，参数为 p 和 r 的 Pascal 分布为 r 个独立同分布的同时服从几何分布 $\mathrm{Geo}(p)$ 的随机变量之和的分布。从该角度看，两者的关系完全相似。而事实上，它们都是等待成功的时间的分布

5.3.6 正态分布
Normal Distribution

De Moivre 在赌博问题中近似二项概率时，利用同时代的 Stirling 公式近似计算阶乘后最早得到了今天所说的正态分布的密度函数表达式。关于二项概率的近似，Johann Bernoulli 曾利用分析方法证明今天所说的 Bernoulli 大数定律 (5.15)，但方法繁杂，且较难应用。而 Laplace 使用概率方法不仅证明了 Bernoulli 大数定律 (5.15)，且进一步得到了二项分布收敛于正态分布的结论，这一结论在 De Moivre 最初成果公布之后四十余年得到，由此称为 De Moivre-Laplace 局部极限定理 (5.16)

Gauss 首先把正态分布应用于随机误差的刻画。历史上，为了处理测量数据中的误差，许多天文学家和数学家从十八世纪开始寻找误差分布的曲线，其中 Laplace 做了许多有意义的工作，Gauss 在十九世纪初介入了这一工作。

1801 年 1 月，一位天文学家发现了一颗从未见过的光度 8 等的星在移动，这颗小行星今天被称为谷神星，它当时在夜空中出现了 6 个星期，旋转过 8 度角以后就消失在了太阳的光芒之下无处寻找。由于留下的观察数据有限，不足以计算出它的运行轨道，甚至都无法确定它究竟是行星还是彗星。正当天文学家们束手无策之际，Gauss 闻知了此事并产生了浓厚兴趣，他创造了一种崭新的计算行星轨道的方法，利用这种方法代入观察数据后不过一个小时就算出了这颗新星的运行轨道，预言了它在天空再次出现的时间与位置。1801 年 12 月 31 日，一位德国天文爱好者在 Gauss 预言的时间，用天文望远镜对准了那片天空，那颗新星出现在了他的镜头中!

当时 Gauss 拒绝透露计算轨道的方法，原因可能是 Gauss 认为自己方法的理论基础还不够成熟。直到 1809 年，Gauss 公布了自己的方法——用今天的话即为以正态分布为基础的最小二乘法。Gauss 的这一成就是十九世纪统计学中的最重要的成就。正态分布也随着 Gauss 的这一成就而被人们称为了 Gauss 分布

定义 5.32 (正态分布)

设 $a \in \mathbb{R}, \sigma > 0$，若连续分布的密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$$

则称该分布为以 a 和 σ^2 为参数的正态分布 (normal distribution/нормальное распределение) 或 Gauss 分布/Gaussian distribution/Гауссовское распределение，记为 $\mathcal{N}\left(a, \sigma^2\right)$

正态分布 $\mathcal{N}\left(a, \sigma^2\right)$ 的分布函数为

$$\Phi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left\{-\frac{(u-a)^2}{2\sigma^2}\right\} \mathrm{d}u, \quad x \in \mathbb{R}$$

特别地，将正态 $\mathcal{N}\left(a, \sigma^2\right)$ 的分布函数和密度函数分别记为 $\Phi_{a,\sigma}(x)$ 和 $\varphi_{a,\sigma}(x)$。称 $a = 0, \sigma = 1$ 的正态分布为标准正态分布 (standard normal distribution/стандартное нормальное распределение)。标准正态分布 $\mathcal{N}(0,1)$ 的分布函数和密度函数分别用 $\Phi(x)$ 和 $\varphi(x)$ 表示，即

$$\rho_\xi(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in R^1$$

$$F_\xi(x) = \Phi(x) = \int_{-\infty}^{x} \varphi(y) dy$$

♣

注 (正态分布分布函数)

正态分布分布函数较难积分，对于标准正态分布分布函数，有可以查阅的表，利用变换

$$\Phi_{a,\sigma}(x) = \Phi\left(\frac{x-a}{\sigma}\right), \quad x \in \mathbb{R}$$

即可计算一般正态分布的分布函数

命题 5.7 (正态分布定义正确性)

函数

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\},\quad x\in\mathbb{R}$$

为一个概率密度函数

证明 显然 $p(x)$ 非负，下证 $\int\limits_{-\infty}^{\infty}p(x)\mathrm{d}x=1$，记 $I=\int\limits_{-\infty}^{\infty}p(x)\mathrm{d}x$ ，由于 $p(x)$ 的原函数不为初等函数，转而考虑 I^2。易知

$$\begin{aligned}I^2&=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\mathrm{d}x\cdot\int_{-\infty}^{\infty}\exp\left\{-\frac{(y-a)^2}{2\sigma^2}\right\}\mathrm{d}y\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-\frac{u^2+v^2}{2}\right\}\mathrm{d}u\ \mathrm{d}v=\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\infty}\exp\left\{-\frac{r^2}{2}\right\}r\ \mathrm{d}r=1\end{aligned}$$

其中先做了坐标变换 $u=\dfrac{x-a}{\sigma},v=\dfrac{y-a}{\sigma}$，再做极坐标变换。由 $I>0$ 即得 $I=1$。则 $p(x)$ 为一个概率密度函数

定义 5.33 (对数正态分布)

设一个随机变量 ξ ，若可表示为 $\xi=e^{\eta}$，其中随机变量 η 的概率分布为参数为 $a\in R^1$ 和 $\sigma^2>0$ 的正态分布，则称随机变量 ξ 的概率分布为参数为 $a\in R^1$ 和 $\sigma^2>0$ 的对数正态分布 (logarithmic normal distribution/логнормальное распределение)

性质 （标准正态分布分布函数性质）

标准正态分布分布函数满足如下性质：

1) （规范性） $\Phi(-\infty)=0,\Phi(\infty)=1,\Phi(0)=0.5$
2) （对称性） $\Phi(-x)=1-\Phi(x)$

注 由于 2)，编制函数值表时通常只考虑定义域只为正或只为负。由于 $\Phi(3.9)=0.999$，则对于 $x\geqslant 4$ 可认为 $\Phi(x)\approx 1$

注 （概率积分）

经常也使用

$$\Phi_0(x)=\int_0^x\varphi(y)dy$$

来代替 $\Phi(x)$，称其为 Laplace **函数** (функция Лапласа) 或**概率积分** (интеграл вероятностей)

性质 （概率积分性质）

概率积分满足下列性质：

1) $\Phi_0(0)=0,\Phi_0(\infty)=0.5$
2) （奇性） $\Phi_0(-x)=-\Phi_0(x)$
3) $\Phi(x)=\Phi_0(x)+0.5$

性质 （正态分布密度函数）

正态分布密度函数满足下列命题：

1) （对称性） $\varphi_{a,\sigma}(x)$ 关于 $x=a$ 对称，即有

$$(\forall x\in\mathbb{R}):\varphi_{a,\sigma}(a-x)=\varphi_{a,\sigma}(a+x)$$

特别地，标准正态分布 $\mathcal{N}(0,1)$ 的密度函数

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\quad x\in\mathbb{R}$$

关于 $x=0$ 对称，即为偶函数

2) （最值与离散程度） $\varphi_{a,\sigma}(x)$ 在 $x=a$ 处达到最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$。$\sigma$ 的值越小，$\varphi_{a,\sigma}(x)$ 的峰越陡峭；反之，σ 的

值越大, $\varphi_{a,\sigma}(x)$ 的峰越平缓

证明 1) 显然

2) 若随机变量 X 的概率分布为正态分布 $\mathcal{N}\left(a,\sigma^2\right)$, 则对任何 $x_1<a<x_2$, 就有

$$P\left(x_1<X<x_2\right)=\int_{x_1}^{x_2}\varphi_{a,\sigma}(u)\mathrm{d}u$$

则 σ 的值越小, 概率 $P\left(x_1<X<x_2\right)$ 的值就越大; 反之, σ 的值越大, 概率 $P\left(x_1<X<x_2\right)$ 的值就越小

注 称 a 为正态分布 $\mathcal{N}\left(a,\sigma^2\right)$ 的**位置参数**。称 σ 为正态分布 $\mathcal{N}\left(a,\sigma^2\right)$ 的**形状参数**

例题 5.28 (正态分布)

由学校到机场有两条路线, 所需时间随交通堵塞状况有所变化, 若以分钟计算, 第一条路线所需时间 X_1 服从正态分布 $\mathcal{N}(50,100)$, 第二条路线所需时间 X_2 服从正态分布 $\mathcal{N}(60,16)$, 下问 (1) 在 70 分钟内赶到机场应选择哪条路线 (2) 在 65 分钟内赶到机场应选择哪条路线?

解 利用变换, 分别出计算如下概率

$$\begin{aligned}
P\left(X_1\leqslant 70\right)&=\Phi\left(\frac{70-50}{10}\right)=\Phi(2)=0.9772\\
P\left(X_2\leqslant 70\right)&=\Phi\left(\frac{70-60}{4}\right)=\Phi(2.5)=0.9938\\
P\left(X_1\leqslant 65\right)&=\Phi\left(\frac{65-50}{10}\right)=\Phi(1.5)=0.9332\\
P\left(X_2\leqslant 65\right)&=\Phi\left(\frac{65-60}{4}\right)=\Phi(1.35)=0.8944
\end{aligned}$$

比较即知: 若要求在 70 分钟内赶到机场, 则应选择第二条路线; 而若要求在 65 分钟内赶到机场, 则应选择第一条路线

命题 5.8 (标准正态分布尾分布)

对于标准正态随机变量 X 有

$$P(|X|>x)\sim\sqrt{\frac{2}{\pi}}\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\},\quad x\to\infty$$

其中 $a(x)\sim b(x),x\to\infty$ 表示 $\lim\limits_{x\to\infty}\dfrac{a(x)}{b(x)}=1$

证明 由于对任意 $x>0$ 都有

$$P(|X|>x)=\sqrt{\frac{2}{\pi}}\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u$$

则仅需证

$$\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u\sim\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\},\quad x\to\infty$$

注意到

$$\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u=\int_x^{\infty}\frac{1}{u}\cdot u\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u=-\int_x^{\infty}\frac{1}{u}\,\mathrm{d}\exp\left\{-\frac{1}{u^2}\right\}$$

由分部积分即得

$$\begin{aligned}
\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u&=-\int_x^{\infty}\frac{1}{u}\,\mathrm{d}\exp\left\{-\frac{1}{u^2}\right\}=\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}+\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}\frac{u^2}{2}\\
&=\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}-\int_x^{\infty}\frac{1}{u^2}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u\leqslant\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}
\end{aligned}$$

其中最后一步等式还可得

$$\begin{aligned}\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u&=\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}-\int_x^{\infty}\frac{1}{u^2}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u\\&=\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}-\int_x^{\infty}\frac{1}{u^3}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}\frac{u^2}{2}\\&=\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}+\int_x^{\infty}\frac{1}{u^3}\,\mathrm{d}\exp\left\{-\frac{1}{u^2}\right\}\\&=\left(\frac{1}{x}-\frac{1}{x^3}\right)\exp\left\{-\frac{1}{x^2}\right\}+3\int_x^{\infty}\frac{1}{u^4}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u\\&\geqslant\left(\frac{1}{x}-\frac{1}{x^3}\right)\exp\left\{-\frac{1}{x^2}\right\}\end{aligned}$$

综上可得

$$\left(\frac{1}{x}-\frac{1}{x^3}\right)\exp\left\{-\frac{1}{x^2}\right\}\leqslant\int_x^{\infty}\exp\left\{-\frac{1}{u^2}\right\}\mathrm{d}u\leqslant\frac{1}{x}\exp\left\{-\frac{1}{x^2}\right\}$$

命题即证

例题 5.29 （Pauta 准则/Pauta criterion/ 3σ 准则/Three sigma guidelines/правило трех σ）

注意到

$$P(|Y|\leqslant 3)=\Phi(3)-\Phi(-3)=2\Phi(3)-1\approx 0.9973$$

利用变换即得

$$\begin{aligned}P(|X-a|\leqslant 3\sigma)&=P(a-3\sigma\leqslant X\leqslant a+3\sigma)=\Phi_{a,\sigma}(a+3\sigma)-\Phi_{a,\sigma}(a-3\sigma)\\&=\Phi(3)-\Phi(-3)\approx 0.9973002039367398\end{aligned}$$

类似地，若随机变量 X 概率分布为正态分布 $\mathcal{N}\left(a,\sigma^2\right)$，则有

$$P(|X-a|\leqslant\sigma)=\Phi(1)-\Phi(-1)\approx 0.6826894921370859$$
$$P(|X-a|\leqslant 2\sigma)=\Phi(2)-\Phi(-2)\approx 0.9544997361036416$$

例题 5.30 （6σ 管理原则[3]）

对某件产品上的某个数据的设计要求为 M，但由于各种随机因素的影响，根据实际情况规定了该数据的允许上限 T_U 和允许下限 T_L，凡是该数据落在范围 $(T_L;T_U)$ 中的都认为合格。通常，M 为区间 $(T_L;T_U)$ 的中点，而称 $T=:T_U-T_L$ 为**容差** (tolerance)。提高产品合格率的关键为控制产品加工中的误差方差 σ^2，而 6σ 原则即为利用 T 来确定 σ 的值。该原则引入了称为**能力指数** (capability index/CP) 的量 C_p

$$C_p=\frac{T}{6\sigma}\approx\frac{T}{6s}$$

其中 σ 为该数据分布作为正态分布的均方差，而 s 为通过实际数据求出的样本均方差

若 $C_p=1$，则意味着 $6\sigma=T$。从理论上，只要能保证加工出来的工件在该数据上的均方差为 $\sigma=\dfrac{T}{6}$，则该项产品在该数据上的合格率就可达到 99.73%。但由于实际加工出来的产品的该项数据的均值 μ 未必为 M，必须考虑偏移值 $|M-\mu|$ 所带来的影响。因为落在区间 $(\mu-3s;\mu+3s)$ 中与落在 $(M-3\sigma;M+3\sigma)=(T_L,T_U)$ 内的误差可能会相差很远。为了避免这点，有的企业就在进一步缩小加工产品的均方差 σ 努力。例如，规定 $C_p=1.5$，即 $\sigma=\dfrac{T}{9}$，则变为 $9\sigma=T,(T_L,T_U)=(M-4.5\sigma,M+4.5\sigma)$ 。这时即使偏移达到 $\mu-M=1.5\sigma$ 或 $\mu-M=-1.5\sigma$，则亦有可能为

$$(\mu-3\sigma,\mu+3\sigma)\subset(M-4.5\sigma,M+4.5\sigma)=(T_L,T_U)$$

事实上，C_p 越大，工件加工时所允许的均方差越小，从而产品尺寸落在可允许区间中的可能性就越大，以至于保证 99.73% 的合格率

[3]在企业质量管理上广泛采用一种 6σ 管理原则。二十世纪七八十年代，美国摩托罗拉公司在与日本公司的竞争中遭受一系列挫折，丢掉了许多市场。他们痛定思痛，在企业内部开展以“零缺陷”为目标的质量改进运动。他们提出：减小误差方差 σ^2 并把误差控制在 $\pm3\sigma$ 的范围内。正是这一措施使得摩托罗拉不仅摆脱了濒临破产的境地，而且一跃成为质量与利润都领先的世界著名公司。他们在此过程中形成了一整套基于概率统计的系统化管理方法，被人们称为 6σ 管理原则，后来被许多企业采用并逐步演变为一个强有力的管理系统 [ref8][ref9]

5.4 数值特征与极限定理 Numerical Characteristics and Limit Theorem

5.4.1 数学期望与 Riemann-Stieltjes 积分 Mathematic Expectation and Riemann-Stieltjes Integral

Huygens[a]在概率论的早期发展历史上也占有重要的地位，其主要著作《机遇的规律》在 1657 年出版。在这部著作中，Huygens 首先引进了“期望”这个术语，基于该术语解决了一些当时感兴趣的博弈问题。Huygens 在这部著作中提出了 3 条定理和 11 个命题，其中前三条命题为：

命题 1：若某人在赌博中以等概率 $\frac{1}{2}$ 得 a 元，则其期望为 $\frac{a-b}{2}$ 元

命题 2：若某人在赌博中以等概率 $\frac{1}{3}$ 得 a,b 和 c 元，则其期望为 $\frac{(a+b+c)}{3}$ 元

命题 3：若某人在赌博中以等概率 $p,q(p+q=1)$ 得 a,b 元，则其期望为 $pa+qb$ 元

[a]克里斯蒂安 · 惠更斯 (Christiaan Huygens, 1629.4.14—1695.7.8) 荷兰物理学家、天文学家、数学家。Huygens 是介于 Galileo 与 Newton 之间的重要的物理学先驱，对力学的发展和光学的研究有杰出的贡献。Huygens 在数学和天文学方面也有卓越的成就。他建立向心力定律，提出动量守恒原理，并改进了计时器。

定义 5.34 (离散随机变量数学期望)

设离散随机变量 $\xi=\sum\limits_{i=1}^{k} x_i I_{A_i}$ 与概率空间 $(\Omega,\mathcal{A},P)$，则称实数

$$E(\xi)=\sum_{i=1}^{k} x_i P\left(A_i\right)$$

为离散随机变量 ξ 的数学期望 (mathematic expectation/математичесеиое ожидание) ♣

注 （离散随机变量数学期望）

由 $A_i=\{\omega:\xi(\omega)=x_i\}$，而 $P_\xi\left(x_i\right)=P\left(A_i\right)$，则有

$$E(\xi)=\sum_{i=1}^{k} x_i P_\xi\left(x_i\right)$$

由离散随机变量的定义记 $\Delta F_\xi(x)=F_\xi(x)-F_\xi(x-)$ ，即得 $P_\xi\left(x_i\right)=\Delta F_\xi\left(x_i\right)$，进而亦有

$$E(\xi)=\sum_{i=1}^{k} x_i \Delta F_\xi\left(x_i\right)$$

性质 （离散随机变量数学期望性质）

设 ξ,η 为离散随机变量，则下列命题成立：

1) （保号性）若 $\xi\geqslant 0$，则有 $E(\xi)\geqslant 0$
2) （线性性）$(\forall a,b\in\mathbb{R}):E(a\xi+b\eta)=aE(\xi)+bE(\eta)$
3) （保序性）若 $\xi\geqslant\eta$，则有 $E(\xi)\geqslant E(\eta)$
4) （绝对值不等式）$|E(\xi)|\leqslant E(|\xi|)$
5) （独立同态性）若 ξ 和 η 独立，则有 $E(\xi\eta)=E(\xi)E(\eta)$
6) （Cauchy-Schwarz-Буняковский 不等式）$(E(|\xi\eta|))^2\leqslant E(\xi^2)E(\eta^2)$
7) （示性函数基本等式）若 $\xi=I(A)$，则 $E(\xi)=P(A)$

证明 1) 显然

2) 设 $\xi=\sum x_iI(A_i),\eta=\sum y_jI(B_j)$，则有

$$a\xi+b\eta=a\sum_{i,j}x_iI(A_i\cap B_j)+b\sum_{i,j}y_jI(A_i\cap B_j)=\sum_{i,j}(ax_i+by_j)I(A_i\cap B_j)$$

$$\begin{aligned}E(a\xi+b\eta)&=\sum_{i,j}(ax_i+by_j)P(A_i\cap B_j)=\sum_i ax_iP(A_i)+\sum_j by_jP(B_j)\\&=a\sum_i x_iP(A_i)+b\sum_j y_jP(B_j)=aE(\xi)+bE(\eta)\end{aligned}$$

3) 由 1) 和 2) 立得

4) 显然有 $-|\xi|\leqslant\xi\leqslant|\xi|$ ，则由 3) 即得

$$E(-|\xi|)\leqslant E(\xi)\leqslant E(|\xi|)$$

再由 2) 即得

$$-E(|\xi|)\leqslant E(\xi)\leqslant E(|\xi|)$$

5) 对于相互独立的随机变量 ξ 和 η，事件 $A_i=\{\omega:\xi(\omega)=x_i\}$ 和 $B_j=\{\omega:\eta(\omega)=y_j\}$ 相互独立，则有 $P(A_i\cap B_j)=P(A_i)P(B_j)$，则有

$$\begin{aligned}E(\xi\eta)&=E\left[\left(\sum_i x_iI(A_i)\right)\left(\sum_j y_jI(B_j)\right)\right]=E\left(\sum_{i,j}x_iy_jI(A_i\cap B_j)\right)\\&=\sum_{i,j}x_iy_jP(A_i\cap B_j)=\sum_{i,j}x_iy_jP(A_i)P(B_j)\\&=\left(\sum_i x_iP(A_i)\right)\left(\sum_j y_jP(B_j)\right)=E(\xi)E(\eta)\end{aligned}$$

6) 设 $E(\xi^2)>0,E(\eta^2)>0$，记

$$\tilde{\xi}=\frac{\xi}{\sqrt{E(\xi^2)}},\quad\tilde{\eta}=\frac{\eta}{\sqrt{E(\eta^2)}}$$

由 $2|\tilde{\xi}\tilde{\eta}|\leqslant\tilde{\xi}^2+\tilde{\eta}^2$，即由 2)3) 得 $2E(|\tilde{\xi}\tilde{\eta}|)\leqslant E(\tilde{\xi}^2)+E(\tilde{\eta}^2)=2$，则有

$$E(|\tilde{\xi}\tilde{\eta}|)\leqslant 1,\quad(E(|\xi\eta|))^2\leqslant E(\xi^2)\times E(\eta^2)$$

若 $E(\xi^2)=0,$，则有 0 为 ξ 的可能值并且 $P\{\omega:\xi(\omega)=0\}=1$ 。则若 $E(\xi^2)$ 或 $E(\eta^2)$ 之一等于 0，则显然 $E(|\xi\eta|)=0$，不等式仍然成立

7) 由示性函数定义，随机变量 ξ 取两个值：以概率 $P(A)$ 取 1，以 $1-P(A)$ 概率取 0，概则根据离散随机变量的数学期望的定义即有

$$E(\xi)=1\cdot P(A)+0\cdot(1-P(A))=P(A)$$

即得所求等式

例题 5.31 （Bernoulli 随机变量数学期望）

设 ξ 为参数为 p 的 Bernoulli 随机变量，另设概率空间 $(\Omega,\mathcal{A},P)$ 和随机事件 A，满足 $P(A)=p$，并令 $\xi=I_A$，则 Bernoulli 随机变量 ξ 的数学期望为

$$E(\xi)=1\times P\{\xi=1\}+0\times P\{\xi=0\}=p=P(A)$$

例题 5.32 （Bernoulli 随机变量和数学期望）

设 $\xi_1,\cdots,\xi_n$ 为 n 个参数为 p 的 Bernoulli 随机变量，试求随机变量 $S_n=\xi_1+\cdots+\xi_n$ 的数学期望

解

方法一：Bernoulli 随机变量

另设概率空间 $(\Omega,\mathcal{A},P)$ 和随机事件 A_i，满足 $P(A_i)=p$，并令 $\xi_i=I_{A_i}$，则对于随机变量 $S_n=\xi_1+\cdots+\xi_n$，其数学期望为 $E(S_n)=np$

方法二：经典方法

若假设 $\xi_1,\cdots,\xi_n$ 为全体独立的 Bernoulli 随机变量,数学期望 $E(S_n)$ 不变。这时有 $P\{S_n=k\}=\mathrm{C}_n^k p^k q^{n-k}(k=0,1,\cdots,n)$ ，则有

$$\begin{aligned}E(S_n)&=\sum_{k=0}^{n}kP\{S_n=k\}=\sum_{k=0}^{n}k\mathrm{C}_n^k p^k q^{n-k}=\sum_{k=0}^{n}\frac{kn!}{k!(n-k)!}p^k q^{n-k}\\&=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!}p^{k-1}q^{(n-1)-(k-1)}\\&=np\sum_{l=0}^{n-1}\frac{(n-1)!}{l!((n-1)-l]!}p^l q^{(n-1)-l}=np\end{aligned}$$

即随机变量 $S_n=\xi_1+\cdots+\xi_n$ 的数学期望为 $E(S_n)=np$

例题 5.33 （离散随机变量数学期望）

m 个 0 和 n 个 1 随机地排成一行，求其中首次出现 1 的序号的概率分布及其数学期望

解 1) 将 m 个 0 和 n 个 1 随机排成的序列记作 $a_1,a_2,\cdots,a_{m+n}$, 以 W 表示首次出现 1 的序号。则事件 $(W=1)$ 表示 $a_1=1$，而对 $k\geqslant 2$，事件 $(W=k)$ 表示 $a_1=\cdots=a_{k-1}=0,a_k=1$。由计数原理即知

$$\begin{aligned}&P(W=1)=\frac{C_{m+n-1}^m}{C_{m+n}^m}=\frac{n}{m+n}\\&P(W=k)=\frac{C_{m+n-k}^{m-k+1}}{C_{m+n}^m}=\frac{(m+n-k)!m!n}{(m+n)!(m+1-k)!},\quad k=2,\cdots,m+1\\&P(W=k)=0,\quad k=m+2,\cdots,m+n\end{aligned}$$

即为所求概率分布

2) (Bernoulli 随机变量) 将 m 个 0 分别记作 $0_1,\cdots,0_m$，以 B_k 表示 0_k 出现在所有的 1 之前的事件。记 $W_k=I_{B_k}$，则 $\sum\limits_{k=1}^{m}W_k$ 即为第 1 个出现的 1 前面的 0 的个数，则首次出现 1 的序号为

$$W=1+\sum_{k=1}^{m}W_k$$

下求 $E(W_k)$，即 $P(B_k)$。由于 0_k 等可能地出现在 n 个 1 所形成的 $n-1$ 个间隔和头尾两个位置上，则它出现在任何一个位置上的概率均为 $\dfrac{1}{n+1}$，则有

$$E(W_k)=P(B_k)=\frac{1}{n+1},\quad k=1,\cdots,m$$

再由数学期望线性性和 Bernoulli 随机变量数学期望即得

$$E(W)=1+\sum_{k=1}^{m}E(W_k)=1+\frac{m}{n+1}=\frac{m+n+1}{n+1}$$

即为所求数学期望

引理 5.1 (非负随机变量离散随机变量序列逼近第一引理)

设 ξ 为任意非负随机变量，则存在离散随机变量序列 $\{\xi_n\}$ 满足

1) (非负性) $(\forall n):\xi_n(\omega)\geqslant 0$
2) (非降性) $\xi_n(\omega)\leqslant\xi_{n+1}(\omega)$
3) (渐近性质) $(\forall\omega\in\Omega):|\xi_n(\omega)-\xi(\omega)|\underset{n\to\infty}{\longrightarrow}0$

♡

证明 将 $[0;\infty)$ 分为长度为 2^{-n} 的相等的线段。定义若 $k2^{-n}\leqslant\xi(\omega)<(k+1)2^{-n}$，则有 $\xi_n(\omega)=\dfrac{k}{2^n}$ 。则由该定义即有 1)2) 且注意到

$$(\forall\omega\in\Omega):|\xi_n(\omega)-\xi(\omega)|\leqslant 2^{-n}\underset{n\to\infty}{\longrightarrow}0$$

即有引理成立

引理 5.2 (非负随机变量离散随机变量序列逼近第二引理)

设 ξ 为非负随机变量，$\{\xi_n\}$ 与 $\{\eta_n\}$ 为两个满足引理 (5.1) 条件

1)（非负性）$(\forall n):\xi_n(\omega)\geqslant 0,$

2)（非降性）$\xi_n(\omega)\leqslant\xi_{n+1}(\omega)$

3)（渐近性质）$(\forall\omega\in\Omega):|\xi_n(\omega)-\xi(\omega)|\underset{n\to\infty}{\longrightarrow}0$ 的离散随机变量序列，则有

$$\lim_n E(\xi_n)=\lim_n E(\eta_n)$$

证明 由 3) 知存在一个正数序列 $\{\varepsilon_n\}$ 满足 $|\xi_n(\omega)-\eta_n(\omega)|\leqslant\varepsilon_n\underset{n\to\infty}{\longrightarrow}0$ ，由此即得

$$(\forall\omega\in\Omega):\xi_n(\omega)-\varepsilon_n\leqslant\eta_n(\omega)\leqslant\xi_n(\omega)+\varepsilon_n$$

再由离散随机变量数学期望性质得 $E(\xi_n)-\varepsilon_n\leqslant E(\eta_n)\leqslant E(\xi_n)+\varepsilon_n$ ，令 $n\to\infty$ 即证

定义 5.35 (非负随机变量数学期望)

设 ξ 为任意非负随机变量，若离散随机变量序列 $\{\xi_n\}$ 满足

1)（非负性）$(\forall n):\xi_n(\omega)\geqslant 0$

2)（非降性）$\xi_n(\omega)\leqslant\xi_{n+1}(\omega)$

3)（渐近性质）$(\forall\omega\in\Omega):|\xi_n(\omega)-\xi(\omega)|\underset{n\to\infty}{\longrightarrow}0$

则称

$$E(\xi)=\lim_n E(\eta_n)$$

为非负随机变量 ξ 的数学期望 (mathematic expectation/математичесеиое ожидание)

性质 （随机变量正负部分分解）

设 ξ 为任意随机变量，定义

$$\xi^+(\omega)=\begin{cases}\xi(\omega) & ,\xi(\omega)\geqslant 0\\ 0 & ,\xi(\omega)<0\end{cases}\quad,\quad\xi^-(\omega)=\begin{cases}0 & ,\xi(\omega)\geqslant 0\\ -\xi(\omega) & ,\xi(\omega)<0\end{cases}$$

则有

1)（非负性）$\xi^+(\omega)\geqslant 0,\xi^-(\omega)\geqslant 0$

2)（随机变量表示）$\xi(\omega)=\xi^+(\omega)-\xi^-(\omega)$

定义 5.36 (随机变量数学期望)

对于任意随机变量 ξ 定义

$$\xi^+(\omega)=\begin{cases}\xi(\omega) & ,\xi(\omega)\geqslant 0\\ 0 & ,\xi(\omega)<0\end{cases}\quad,\quad\xi^-(\omega)=\begin{cases}0 & ,\xi(\omega)\geqslant 0\\ -\xi(\omega) & ,\xi(\omega)<0\end{cases}$$

若 $M(\xi^+)$ 和 $M(\xi^-)$ 至少一个有限，则称

$$M(\xi)=M(\xi^+)-M(\xi^-)$$

为随机变量 ξ 的数学期望 (mathematic expectation/математичесеиое ожидание)

若 $M(\xi^+)$ 和 $M(\xi^-)$ 均有限，则称随机向量 ξ 可积

注 （随机变量数学期望）

下文除特殊说明，总是假设计算随机变量 ξ 数学期望时，$M(\xi^+)$ 和 $M(\xi^-)$ 均有限

定义 5.37 (Riemann-Stieltjes 可积)

设 $F(x)$ 为不减有界函数，$f(x)$ 为定义在 $[a;b)$ 上的 Borel 可测函数。设划分 $a=t_0<t_1<\ldots<t_k<$

$t_{k+1} \ldots < t_n = b$ 并选择点 $s_k \in [t_k; t_{k+1})$，则有积分和

$$S_n = \sum_{k=0}^{n-1} f(s_k)\left[F(t_{k+1}) - F(t_k)\right]$$

若积分和 S_n 当 $\max_k |t_{k+1} - t_k| \to 0$ 时存在极限，且该极限与 $[a; b)$ 的划分方法和点 s_k 的选择无关，则称该极限为函数 $f(x)$ 相对于函数 $F(x)$ 的 Riemann-Stieltjes 积分 (Riemann-Stieltjes integral)，称函数 $f(x)$ 在 $[a; b)$ 上相对于函数 $F(x)$ 按 Riemann-Stieltjes 意义可积 (интегрируема в смысле Римана - Стилтьеса относительно функции $F(x)$ на $[a; b)$) 或在 $[a; b)$ 上相对于函数 $F(x)$ 为 Riemann-Stieltjes 可积 (Riemann-Stieltjes integrable) 的，记为

$$\int_a^b f(x) dF(x)$$

♣

注 (Riemann-Stieltjes 积分)

在 Riemann-Stieltjes 积分中，若 $F(x) = x$，则得到了 Riemann 积分的定义。另外，若存在 $\rho_\xi(x) = F'(x)$，则

$$\int_a^b f(x) dF(x) = \int_a^b f(x) \rho_\xi(x) dx$$

其中最后一个积分为 Riemann 积分

注 (Riemann-Stieltjes 积分)

若积分区间为无穷区间，则首先在有限区间上确定积分，然后当区间的一端或两端趋于无穷时，取极限求得。在有限区间上 Riemann-Stieltjes 可积的充分条件为函数 $f(x)$ 在闭区间 $[a; b]$ 上的连续性

定理 5.9 (随机变量数学期望 Riemann-Stieltjes 积分表示)

设 ξ 为随机变量，则有

$$E(\xi) = \int_{-\infty}^{\infty} x dF_\xi(x)$$

其中右端为 Riemann-Stieltjes 积分

若 ξ 为离散随机变量，则可以写为

$$E(\xi) = \sum_n x_n p_n$$

若 ξ 为连续随机变量，则可以写为

$$E(\xi) = \int_{-\infty}^{\infty} x \rho_\xi(x) dx$$

♡

证明 假设 ξ 为非负随机向量，构造离散随机变量序列 ξ_n 满足若 $\dfrac{k}{2^n} \leqslant \xi(\omega) < \dfrac{k+1}{2^n}$，则有 $\xi_n(\omega) = \dfrac{k}{2^n}$，这时由随机变量数学期望定义即得

$$E(\xi) = \lim_n E(\xi_n) = \lim_n \sum_{k=0}^{\infty} \frac{k}{2^n}\left[F_\xi\left(\frac{k+1}{2^n}\right) - F_\xi\left(\frac{k}{2^n}\right)\right]$$

将最后一个等式与 $f(x) = x$ 和 $F(x) = F_\xi(x)$ 时的 Riemann-Stieltjes 积分的定义比较即得

$$E(\xi) = \int_0^{\infty} x dF_\xi(x)$$

利用对称性展开随机变量 $\xi(x) = \xi^+(x) - \xi^-(x)$ 即得所求结论

推论 5.2 (积分变换定理)

若随机变量 ξ 有分布函数 $F_\xi(x)$，而 $y = g(x)$ 为 Borel 可测函数，则随机变量 $\eta = g(\xi)$ 的正负部分期望

均有限，且有

$$E(\eta)=\int_{-\infty}^{\infty} g(x) d F_{\xi}(x)$$

其中右端为 Riemann-Stieltjes 积分
若 ξ 为离散随机变量，则可以写为

$$E(\eta)=\sum_{n} g\left(x_{n}\right) p_{n}$$

若 ξ 为连续随机变量，则可以写为

$$E(\eta)=\int_{-\infty}^{\infty} g(x) \rho_{\xi}(x) d x$$

♡

例题 5.34 (Cauchy-Lorenz 分布/随机变量数学期望)

设随机变量 X 概率分布为参数为 $\lambda>0, \mu \in \mathbb{R}$ 的 Cauchy-Lorenz 分布，即

$$\rho(x)=\frac{1}{\pi} \frac{\lambda^{2}}{\lambda^{2}+(x-\mu)^{2}}, \quad x \in \mathbb{R}$$

试证明：Cauchy-Lorenz 分布有无穷数学期望

解 注意到

$$\int_{-\infty}^{\infty} x \rho(x) \mathrm{d} x=\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+\left(\frac{x}{\lambda}-\frac{\mu}{\lambda}\right)^{2}} \mathrm{~d} x=\left\{\begin{array}{l}\lambda y=x-\mu \\ d x=\lambda d y\end{array}\right\}=\frac{\lambda^{2}}{\pi} \int_{-\infty}^{\infty} \frac{y}{1+y^{2}} \mathrm{~d} y+\lambda \mu$$

而由

$$\int_{0}^{\infty} \frac{y}{1+y^{2}} \mathrm{~d} y=\lim _{A \rightarrow \infty} \int_{0}^{A} \frac{y}{1+y^{2}} \mathrm{~d} y=\frac{1}{2} \lim _{A \rightarrow \infty} \ln \left(1+A^{2}\right)=\infty$$

由定理 (5.9) 即得

$$E(X)=\int_{-\infty}^{\infty} x p(x) \mathrm{d} x=\infty$$

则 Cauchy-Lorenz 分布有无穷数学期望

例题 5.35 (Peter-Paul 分布/圣彼得堡游戏/St Petersburg game)

设一个人连续抛掷一枚均匀的硬币，直到抛出反面为止。若此前他抛出了 n 次正面，则发给他 2^{n} 元奖金。以 X 表示他所得到的奖金，试证明 X 有无穷数学期望

解 显然 X 的概率分布为

$$P\left(X=2^{n-1}\right)=\frac{1}{2^{n}}, \quad n \in \mathbb{N}$$

则有

$$\sum_{n=1}^{\infty} a_{n} p_{n}=\sum_{n=1}^{\infty} \frac{2^{n-1}}{2^{n}}=\sum_{n=1}^{\infty} \frac{1}{2}=\infty$$

则 X 有无穷数学期望

注 该例中所涉及的 Peter-Paul 分布为概率论中一个著名的分布，也称为“圣彼得堡游戏”。若赌场开设该游戏，虽然从数学期望看，赌徒的平均收益无穷，但很少有人愿意花 32 元以上去玩一次这样的游戏，因为能够赢回这 32 元的概率很小：

$$P(X \geqslant 32)=P\left(X \geqslant 2^{5}\right)=\sum_{n=6}^{\infty} \frac{1}{2^{n}}=\frac{1}{2^{5}}=\frac{1}{32}=0.03125$$

性质 (随机变量数学期望性质)

设 ξ, η 为随机变量，则下列命题成立：

1) (保号性) 若 $\xi \geqslant 0$，则有 $E(\xi) \geqslant 0$
2) (线性性) $(\forall a, b \in \mathbb{R}): E(a \xi+b \eta)=a E(\xi)+b E(\eta)$
3) (保序性) 若 $\xi \geqslant \eta$，则有 $E(\xi) \geqslant E(\eta)$

4)（绝对值不等式）$|E(\xi)| \leqslant E(|\xi|)$

5)（独立同态性）若 ξ 和 η 独立，则有 $E(\xi\eta)=E(\xi)E(\eta)$

6)（Cauchy-Schwarz-Буняковский 不等式）$(E(|\xi\eta|))^2 \leqslant E(\xi^2)E(\eta^2)$

7)（示性函数基本等式）若 $\xi=I(A)$，则 $E(\xi)=P(A)$

证明 （仅证明连续分布情况）1) 显然

2) 由定理 (5.9) 即得

$$E(a\xi+b\eta)=\int_{-\infty}^{\infty} x(a\rho_\xi(x)+b\rho_\eta(x))dx=a\int_{-\infty}^{\infty} x\rho_\xi(x)dx+b\int_{-\infty}^{\infty} x\rho_\eta(x)dx=aE(\xi)+bE(\eta)$$

3) 由 1)2) 立得

4) 同离散随机变量数学期望性质证明

5) a) 设 ξ_1 为 ξ_2 为非负随机变量，且正负部分数学期望有限。构造两个满足引理 (5.1) 条件的离散随机变量序列 $\left\{\xi_1^{(n)}\right\}$ 和 $\left\{\xi_2^{(n)}\right\}$ 由引理 (5.1) 即得 $\xi_1^{(n)}$ 和 $\xi_2^{(n)}$ 可以逼近随机变量 ξ_1 和 ξ_2。由此即得二者独立

假设

$$\left|\xi_1^{(n)}-\xi_1\right| \leqslant \frac{1}{n}, \quad \left|\xi_2^{(n)}-\xi_2\right| \leqslant \frac{1}{n}$$

则由非负随机变量数学期望定义有 $E(\xi_1)=\lim_n E(\xi_1^{(n)}), E(\xi_2)=\lim_n E(\xi_2^{(n)})$ ，进而则有

$$\begin{aligned}
&|E(\xi_1\xi_2)-E(\xi_1)\cdot E(\xi_2)| \\
=&\,|\,E(\xi_1\xi_2)-E\left(\xi_1^{(n)}\xi_2\right)+E\left(\xi_1^{(n)}\xi_2\right)-E\left(\xi_1^{(n)}\xi_2^{(n)}\right)+E\left(\xi_1^{(n)}\xi_2^{(n)}\right)-E(\xi_1)\cdot E(\xi_2)\,| \\
\leqslant& E\left(\left|\xi_1-\xi_1^{(n)}\right||\xi_2|\right)+E\left(\left|\xi_1^{(n)}\right|\left|\xi_2-\xi_2^{(n)}\right|\right)+\left|E(\xi_1^{(n)})\cdot E(\xi_2^{(n)})-E(\xi_1)\cdot E(\xi_2)\right| \\
\leqslant&\frac{1}{n}E(|\xi_2|)+\frac{1}{n}E(|\xi_1|)+\left|E(\xi_1^{(n)})\cdot E(\xi_2^{(n)})-E(\xi_1)\cdot E(\xi_2)\right| \underset{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

b) 设任意随机变量 ξ_1 和 ξ_2 表示为 $\xi_1=\xi_1^+-\xi_1^-$ 和 $\xi_2=\xi_2^+-\xi_2^-$ 。由 ξ_1 与 ξ_2 的独立性，则 (ξ_1^+,ξ_1^-) 与 (ξ_2^+,ξ_2^-) 同样独立。由 a) 即得

$$\begin{aligned}
E(\xi_1\xi_2)=&E\left[(\xi_1^+-\xi_1^-)(\xi_2^+-\xi_2^-)\right]=E\left(\xi_1^+\xi_2^+-\xi_1^+\xi_2^--\xi_1^-\xi_2^++\xi_1^-\xi_2^-\right) \\
=&E(\xi_1^+)E(\xi_2^+)-E(\xi_1^+)E(\xi_2^-)-E(\xi_1^-)E(\xi_2^+)+E(\xi_1^-)E(\xi_2^-) \\
=&E(\xi_1^+-\xi_1^-)E(\xi_2^+-\xi_2^-)=E(\xi_1)\cdot E(\xi_2)
\end{aligned}$$

6)7) 同离散随机变量数学期望性质证明

定理 5.10 (数学期望形式 Jensen 不等式/Jensen Inequality/неравенство Иенсена)

设 $y=g(x)$ 为凸 Borel 可测函数，ξ 为随机变量，则有

$$g(E(\xi)) \leqslant E(g(\xi))$$

♡

证明 由凸函数性质即有

$$(\forall x_0\in R^1)(\exists\lambda(x_0)\in R^1)(\forall x\in R^1): g(x)-g(x_0)\geqslant\lambda(x_0)(x-x_0)$$

用 ξ 和 $E(\xi)$ 代替 x 和 x_0 即得 $g(\xi)-g(E(\xi))\geqslant\lambda(E(\xi))(\xi-E(\xi))$ ，再由随机变量数学期望性质即得

$$E(g(\xi))-g(E(\xi))\geqslant\lambda(E(\xi))(E(\xi)-E(\xi))=0$$

定理即证

推论 5.3 (加权算术-加权幂平均均值不等式)

若 $a_1,a_2,\cdots,a_n$ 与 $p_1,p_2,\cdots,p_n$ 均为正实数，$p_1+p_2+\cdots+p_n=1$，则有

$$\sum_{i=1}^{n} p_i a_i \geqslant \prod_{i=1}^{n} a_i^{p_i}$$

♡

证明 设随机变量 X 的概率分布为 $P\left(X=\ln a_i\right)=p_i, i=1,2,\cdots,n$。取 $g(x)=\mathrm{e}^x$ 则有

$$E(g(X))=\sum_{i=1}^{n}\mathrm{e}^{\ln a_i}p_i=\sum_{i=1}^{n}p_i a_i$$

$$g(E(X))=\exp\left\{\sum_{i=1}^{n}p_i\ln a_i\right\}=\exp\left\{\sum_{i=1}^{n}\ln a_i^{p_i}\right\}=\exp\left\{\ln\prod_{i=1}^{n}a_i^{p_i}\right\}=\prod_{i=1}^{n}a_i^{p_i}$$

而 $g(x)$ 为凸 Borel 可测函数，则由 Jensen 不等式 (5.10) 得 $E(g(X))\geqslant g(E(X))$

例题 5.36 （信息熵/Jensen 不等式）

若离散随机变量 X 的概率分布为 $P(X=k)=p_k, k=1,2,\cdots,n$，$X$ 的信息熵为 $H(X)=-\sum\limits_{k=1}^{n}p_k\ln p_k$，则有

$$H(X)\leqslant\ln n$$

解 令 $f(x)=x\ln x, x>0$，其为 Borel 可测函数。由 $f'(x)=\ln x+1, f''(x)=\dfrac{1}{x}>0$，则 $f(x)$ 为定义在 $(0;\infty)$ 上的凸函数，则由 Jensen 不等式 (5.10) 即知

$$\frac{1}{n}\sum_{k=1}^{n}f\left(p_k\right)\geqslant f\left(\frac{1}{n}\sum_{k=1}^{n}p_k\right)=f\left(\frac{1}{n}\right)$$

亦即

$$\frac{1}{n}\sum_{k=1}^{n}p_k\ln p_k\geqslant\frac{1}{n}\ln\frac{1}{n}$$

则有

$$H(X)=-\sum_{k=1}^{n}p_k\ln p_k\leqslant-\ln\frac{1}{n}=\ln n$$

不等式即证

5.4.2 矩

Moment

定义 5.38 (随机变量矩)

设随机变量 ξ，则称

$$E((\xi-a)^k)$$

为随机变量 ξ 相对于点 a 的 k 阶矩 (moment/момент)。特别地，若 $a=0$，则称 $\alpha_k=E\left(\xi^k\right)$ 为 k 阶原点矩 (начальный момент)；若 $a=E(\xi)$，则称 $\mu_k=E((\xi-E(\xi))^k)$ 为 k 阶中心矩 (центральный момент)

另外，称 $E(|\xi-a|^k)$ 为相对于点 a 的 k 阶绝对矩 (абсолютный момент)，称 $\beta_k=E(|\xi|^k)$ 为 k 阶原点绝对矩 (абсолютный момент)，称 $E(|\xi-E(\xi)|^k)$ 为 k 阶中心绝对矩 (центральный абсолютный момент)

♣

性质 （随机变量原点矩与中心矩关系）

随机变量 ξ 的 k 阶原点矩和 k 阶中心矩满足关系：

$$\mu_k=\sum_{m=0}^{k}(-1)^m\mathrm{C}_k^m\alpha_m\alpha_{k-m}$$

定理 5.11 (随机变量矩序列充要条件)

数值序列 $\{\alpha_k\}$ 为某个随机变量的矩的充要条件为 $\alpha_0=1$ 且对于任意正整数 N 与复数 $c_0,\dots,c_N$ 满足

$$\sum_{m,n=0}^{N}\alpha_{m+n}c_m\overline{c_n}\geqslant 0$$

♡

例题 5.37 （随机变量原点矩）

设随机变量 X 概率分布为 $\mathcal{N}(0,1)$，试求 X 的正整数阶原点矩

解 设 n 为正整数。当 $r=2n-1$ 时有

$$\begin{aligned}E(X^{2n-1})&=0\\ E(|X|^{2n-1})&=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^{2n-1}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=\sqrt{\frac{2}{\pi}}\int_0^{\infty}2^{n-1}t^{n-1}\mathrm{e}^{-t}\,\mathrm{d}t\\ &=\sqrt{\frac{2}{\pi}}2^{n-1}\Gamma(n)=\sqrt{\frac{2}{\pi}}(2n-2)!!\end{aligned}$$

当 $r=2n$ 时有

$$\begin{aligned}E(X^{2n})=E(|X|^{2n})&=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^{2n}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=\frac{2^n}{\sqrt{\pi}}\int_0^{\infty}t^{n-\frac{1}{2}}\mathrm{e}^{-t}\,\mathrm{d}t\\ &=\frac{2^n}{\sqrt{\pi}}\Gamma\left(n+\frac{1}{2}\right)=(2n-1)!!\end{aligned}$$

综上即得 X 的所有正整数阶原点矩

注 值得一提的是，由于对任意 $r>0$ 都有

$$E|X|^r=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|^r\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^r\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x<\infty$$

则有 X 的任意阶矩都存在且有限

定理 5.12 (Lyapunov 不等式/неравенство Ляпунова)

若 ξ 为随机变量，$0<s<t<\infty$，则有

$$\left(E(|\xi|^s)\right)^{\frac{1}{s}}\leqslant\left(E(|\xi|^t)\right)^{\frac{1}{t}}$$

♡

证明 设 $r=\dfrac{t}{s}>1, g(x)=|x|^r, \eta=|\xi|^s$，则 $g(x)$ 为凸 Borel 可测函数，由数学期望形式 Jensen 不等式 (5.10) 即得

$$g(E(\eta))=\left|E(|\xi|^s)\right|^{\frac{t}{s}}\leqslant E(g(\eta))=E\left(|\xi|^{s\cdot\frac{t}{s}}\right)=E\left(|\xi|^t\right)$$

由此即得 $\left(E(|\xi|^s)\right)^{\frac{1}{s}}\leqslant\left(E(|\xi|^t)\right)^{\frac{1}{t}}$

注 该定理给出了随机变量的矩的一种常用的估计

定义 5.39 (随机变量矩母函数)

设 ξ 为随机变量，则称函数

$$\psi_\xi(t)=E\left(e^{t\xi}\right)$$

为随机变量 ξ 的矩母函数 (производящая функция моментов)

♣

性质 （随机变量矩母函数性质）

设 $\psi_\xi(t)$ 为随机变量 ξ 的矩母函数，则有下列命题成立：

1) （规范性）$\psi_\xi(0)=1$

2) （不相容随机变量和与矩母函数关系）若随机变量 ξ_1 和 ξ_2 不相容，则有

$$\psi_{\xi_1+\xi_2}(t)=\psi_{\xi_1}(t)\cdot\psi_{\xi_2}(t)$$

3) (矩母函数基本等式) 设 $t_0>0$，则若矩母函数 $\psi(t)$ 对任意 $t:|t|<t_0$ 都存在，即有

$$(\forall k \geqslant 1): \left.\frac{d^k}{dt^k}\psi_{\xi}(t)\right|_{t=0}=E\left(\xi^k\right)$$

例题 5.38 (随机变量矩母函数)

设 $S_n=\varepsilon_1+\ldots+\varepsilon_n$，其中 $\varepsilon_1,\ldots,\varepsilon_n$ 的概率分布为参数为 p 的 Bernoulli 分布，试求 $S_n=\varepsilon_1+\ldots+\varepsilon_n$ 的一阶原点矩和二阶原点矩

解 由矩母函数定义即有

$$\psi_{\varepsilon_k}(t)=e^{t\cdot 1}p+e^{t\cdot 0}(1-p)=1+p\left(e^t-1\right)$$

$$\psi_{S_n}(t)=\prod_{k=1}^{n}\psi_{\varepsilon_k}(t)=\left[1+p\left(e^t-1\right)\right]^n$$

微分即得

$$\frac{d}{dt}\psi_{S_n}(t)=n\left[1+p\left(e^t-1\right)\right]^{n-1}pe^t$$

$$\frac{d^2}{dt^2}\psi_{S_n}(t)=n(n-1)\left[1+p\left(e^t-1\right)\right]^{n-2}p^2e^{2t}+n\left[1+p\left(e^t-1\right)\right]^{n-1}pe^t$$

由矩母函数性质即得 $E\left(S_n\right)=np, E\left(S_n^2\right)=n(n-1)p^2+np$

定理 5.13 (矩母函数应用)

设 $\{\alpha_n\}$ 为随机变量 ξ 的矩序列，若存在 $r>0$ 满足下列等价条件其一：

1) 级数 $\sum\limits_{n=0}^{\infty}\frac{\alpha_n}{n!}r^n$ 绝对收敛

2) $\psi_{\xi}(r)=E\left(e^{r\xi}\right)$ 有限

则随机变量 ξ 的概率分布由自身的矩单值定义

♡

注 结论意味着若以某种方式计算出分布的矩，则可称其为这是已知的一种分布

例题 5.39 (矩母函数应用)

设随机变量 ξ 的概率分布为标准正态分布，则有

$$\psi_{\xi}(r)=E\left(e^{r\xi}\right)=\int_{-\infty}^{\infty}e^{rx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-r)^2}{2}}dx\cdot e^{\frac{r^2}{2}}=e^{\frac{r^2}{2}}<\infty$$

则正态分布由自身的矩单值定义

5.4.3 方差
Variance

早期文献经常选用 dispertion(离散程度) 的首字母 D 来表示随机变量的方差，而"方差 (variance)"一词的解释和严格定义最早由英国统计学家 Fisher[a]于 1918 年在论文《The correlation between relatives on the supposition of Mendelian inheritance》中提出[b]，因此现代概率文献经常也使用 Var 来表示方差

[a]罗纳德 · 费希尔 (Ronald Aylmer Fisher,1890-1962) 英国统计学家，遗传学家。Fisher 对达尔文进化论作了基础澄清的工作，为现代统计科学的奠基人之一

[b]Fisher R A. XV.—The correlation between relatives on the supposition of Mendelian inheritance[J]. Earth and Environmental Science Transactions of the Royal Society of Edinburgh, 1918, 52(2): 399-433.

定义 5.40 (方差)

称随机变量 ξ 的二阶中心矩

$$\mu_2=E((\xi-E(\xi))^2)$$

为随机变量 ξ 的方差 (variance/дисперсия)，称 $\sigma(\xi)=\sqrt{Var(\xi)}$ 为随机变量 ξ 的标准离差 (standard deviation)，简称为标准差 ♣

注 （方差）

若随机变量 ξ 的分布函数为 $F(x)$，则有

$$Var(\xi)=\int_{-\infty}^{\infty}(x-E(\xi))^2dF(x)$$

离散情况即为

$$Var(\xi)=\sum_{n}(x_n-E(\xi))^2\cdot p_n$$

若绝对连续分布密度函数为 $\rho(x)$ ，则有

$$Var(\xi)=\int_{-\infty}^{\infty}(x-E(\xi))^2\rho(x)dx$$

性质 （方差性质）设随机变量 ξ,η，方差满足下列性质：

1) （方差第二定义）$Var(\xi)=E(\xi^2)-(E(\xi))^2$
2) （方差基本性质）$Var(\xi)\geqslant 0$，当且仅当 ξ 退化为常数时取等
3) （随机变量平移和伸缩）$(\forall a,b\in\mathbb{R}):Var(a+b\xi)=b^2Var(\xi)$
4) （随机变量和）$Var(\xi+\eta)=Var(\xi)+Var(\eta)+2E(\xi-E(\xi))(\eta-E(\eta))$

证明 1) 直接利用定义即得

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

2) 由 $(\xi-E(\xi))^2\geqslant 0$ 即得

3) 由 1) 即有

$$\begin{aligned}Var(a+b\xi)&=E((a+b\xi)^2)-(E(a+b\xi))^2\\&=E(a^2+2ab\xi+b^2\xi^2)-a^2-2abE(\xi)-b^2(E(\xi))^2\\&=b^2E(\xi^2)-b^2(E(\xi))^2=b^2Var(\xi)\end{aligned}$$

4) 直接利用定义即得

$$Var(\xi+\eta)=E[(\xi-E(\xi))+(\eta-E(\eta))]^2=Var(\xi)+Var(\eta)+2E(\xi-E(\xi))(\eta-E(\eta))$$

命题即证

例题 5.40 （方差）

设 $0<\alpha,\beta,\gamma<\dfrac{\pi}{2}$ 且 $\sin^2\alpha+\sin^2\beta+\sin^2\gamma=1$，试证明

$$\frac{\sin^2\alpha}{\sin\beta}+\frac{\sin^2\beta}{\sin\gamma}+\frac{\sin^2\gamma}{\sin\alpha}>1$$

证明 由 $0<\alpha,\beta,\gamma<\dfrac{\pi}{2}$，则有 $0<\sin\alpha,\sin\beta,\sin\gamma<1$。由 $\sin^2\alpha+\sin^2\beta+\sin^2\gamma=1$ 和不等式 $2ab\leqslant a^2+b^2$，轮换即得 $0<\sin\alpha\sin\beta+\sin\beta\sin\gamma+\sin\gamma\sin\alpha\leqslant 1$。由此可以构造随机变量 X，其概率分布为

$$P\left(X=\frac{\sin\alpha}{\sin\beta}\right)=\sin\alpha\sin\beta$$
$$P\left(X=\frac{\sin\beta}{\sin\gamma}\right)=\sin\beta\sin\gamma$$
$$P\left(X=\frac{\sin\gamma}{\sin\alpha}\right)=\sin\gamma\sin\alpha$$
$$P(X=0)=1-\sin\alpha\sin\beta-\sin\beta\sin\gamma-\sin\gamma\sin\alpha$$

则有

$$E(X) = \sin^2\alpha + \sin^2\beta + \sin^2\gamma = 1$$

$$E(X^2) = \frac{\sin^3\alpha}{\sin\beta} + \frac{\sin^3\beta}{\sin\gamma} + \frac{\sin^3\gamma}{\sin\alpha}$$

则由随机变量 X 的方差 $\operatorname{Var} X = E(X^2) - (E(X))^2 \geqslant 0$，则有

$$\frac{\sin^3\alpha}{\sin\beta} + \frac{\sin^3\beta}{\sin\gamma} + \frac{\sin^3\gamma}{\sin\alpha} \geqslant 1$$

再由 $0 < \sin\alpha, \sin\beta, \sin\gamma \geqslant 1$ 即推出

$$\frac{\sin^2\alpha}{\sin\beta} + \frac{\sin^2\beta}{\sin\gamma} + \frac{\sin^2\gamma}{\sin\alpha} > 1$$

命题即证

例题 5.41 （Pascal 分布数学期望和方差）

若随机变量 X 的概率分布为参数为 p 和 r 的 Pascal 分布，则有

$$E(X) = \frac{r}{p}, \quad Var(X) = \frac{rq}{p^2}$$

解 若 X 的概率分布为参数为 p 和 r 的 Pascal 分布，则存在 r 个相互独立的概率分布为几何分布 $\mathrm{Geo}(p)$ 的随机变量 $Z_1, \cdots, Z_r$，使得 $X = Z_1 + \cdots + Z_r$。由于对于服从几何分布 $\mathrm{Geo}(p)$ 的随机变量 Z 有

$$E(Z) = \frac{1}{p} \quad Var(Z) = \frac{q}{p^2}$$

则有

$$E(X) = E(Z_1) + \cdots + E(Z_r) = \frac{r}{p}$$

$$Var(X) = Var(Z_1) + \cdots + Var(Z_r) = \frac{rq}{p^2}$$

即为所求数学期望与方差

例题 5.42 （均匀分布数学期望与方差）

对于 $U[a;b]$ 随机变量 X 有

$$E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}$$

解 由定义直接计算即得

$$E(X) = \frac{1}{b-a}\int_a^b x \ \mathrm{d}x = \frac{a+b}{2}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{b-a}\int_a^b x^2 \ \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

即为所求数学期望与方差

例题 5.43 （指数分布数学期望与方差）

指数分布 $\mathrm{Exp}(\lambda)$ 的随机变量 X 的数学期望与方差为

$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}$$

解 由定义直接计算

$$E(X) = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x = \lambda\int_0^{\infty} x\mathrm{e}^{-\lambda x} \ \mathrm{d}x = \frac{1}{\lambda}\int_0^{\infty} t\mathrm{e}^{-t} \ \mathrm{d}t$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2p(x)\mathrm{d}x = \lambda\int_0^{\infty} x^2\mathrm{e}^{-\lambda x} \ \mathrm{d}x = \frac{1}{\lambda^2}\int_0^{\infty} t^2\mathrm{e}^{-t} \ \mathrm{d}t$$

其中 $t = \lambda x$。另外，对正整数 n 有

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

则 $E(X) = \frac{1}{\lambda}$，而 $D(X) = E(X^2) - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

例题 5.44 （正态分布数学期望与方差）

设随机变量 $X \sim \mathcal{N}\left(a, \sigma^2\right)$，则有

$$E(X) = a, \quad D(X) = \sigma^2$$

解 根据定义即得

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x\varphi_{a,\sigma}(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-a)\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\mathrm{d}x + a\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\mathrm{d}x \\ &= I_1 + aI_2 \end{aligned}$$

其中令 $t = x - a$，则由奇函数性质有

$$I_1 = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} t\exp\left\{-\frac{t^2}{2\sigma^2}\right\}\mathrm{d}t = 0$$

另外有

$$I_2 = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\mathrm{d}x = \int_{-\infty}^{\infty}\varphi_{a,\sigma}(x)\mathrm{d}x = 1$$

即得 $E(X) = a$

利用定义即有

$$\begin{aligned} D(X) &= (E(X-E(X)))^2 = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-a)^2\exp\left\{-\frac{(x-a)^2}{\sigma^2}\right\}\mathrm{d}x \\ &= \sigma^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u = \sigma^2\frac{2}{\sqrt{2\pi}}\int_0^{\infty}2t\mathrm{e}^{-t}\,\mathrm{d}\sqrt{2t} = \frac{2\sigma^2}{\sqrt{\pi}}\int_0^{\infty}t^{\frac{1}{2}}\mathrm{e}^{-t}\,\mathrm{d}t \\ &= \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right) = \sigma^2 \end{aligned}$$

即得 $D(X) = \sigma^2$

定义 5.41 (标准化随机变量)

设非退化随机变量 X，若其数学期望与方差存在，则称

$$X^* = \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - E(X)}{\sigma}$$

为 X 的标准化随机变量 (standardized random variable)

♣

注 （标准化随机变量）

标准化有两层含义：其一为用随机变量减去它的期望，通常称这一步为中心化；其二为除以标准差，通常把这一步叫做正则化。称中心化与正则化全体为标准化。标准化后的随机变量期望为 0，方差为 1

例题 5.45 （标准化随机变量）

正态分布 $\mathcal{N}\left(a, \sigma^2\right)$ 随机变量 X 的标准化随机变量为

$$X^* = \frac{X - a}{\sigma}$$

即 X^* 的概率分布为标准正态分布 $\mathcal{N}(0, 1)$

5.4.4 Bernoulli 大数定律与 De Moivre-Laplace 局部极限定理

Bernoulli Law of Large Numbers and De Moivre-Laplace Local Limit Theorem

> 如果我们能把一切事件永恒地观察下去，则我们终将发现：世间的一切事物都受到因果律的支配，而我们也注定会在种种极其纷纭杂乱的事象中认识到某种必然。

Jakob Bernoulli 雅各布 · 伯努利《推测术》

定理 5.14 (Markov **不等式**/Chebyshev **定理**)

(Markov[a]不等式/Chebyshev 定理) 设概率空间 $(\Omega, \mathcal{A}, P)$, $\xi = \xi(\omega)$ 为非负随机变量, 则

$$(\forall \varepsilon > 0) : P\{\xi \geqslant \varepsilon\} \leqslant \frac{E(\xi)}{\varepsilon}$$

[a]安德雷 · 安德耶维齐 · 马尔可夫 (俄语: Андрей Андреевич Марков,1856.6.14-1922.7.20) 俄罗斯数学家。1874 年入读圣彼得堡大学, 后师从 Chebyshev, 圣彼得堡数学学派代表人物, 逝世于圣彼得堡。Markov 以数论和概率论领域的工作闻名, 开创了随机过程领域, 以其名字命名的 Markov 链被广泛应用在工程、自然科学和社会科学领域

♡

证明 注意到

$$\xi = \xi I(\xi \geqslant \varepsilon) + \xi I(\xi < \varepsilon) \geqslant \xi I(\xi \geqslant \varepsilon) \geqslant \varepsilon I(\xi \geqslant \varepsilon)$$

其中 $I(A)$ 为集合 A 的示性函数。则由根据随机变量数学期望性质有

$$E(\xi) \geqslant \varepsilon E(I(\xi \geqslant \varepsilon)) = \varepsilon P\{\xi \geqslant \varepsilon\}$$

所需不等式即证

注 尽管该不等式以俄罗斯数学家 Markov 的名字命名,但该不等式曾出现在一些更早的文献中,其中包括 Markov 的老师 Chebyshev, 因此有时也称为 Chebyshev 定理

定理 5.15 (Bernoulli **大数定律**/Bernoulli law of large numbers)

设随机变量 $\xi_1, \cdots, \xi_n$, 其中 $\xi_i = \xi_i(\omega) = a_i, i = 1, \cdots, n$, 而 $\omega = (a_1, \cdots, a_n)$。已知 Bernoulli 随机变量 $\xi_i = \xi_i(\omega) = a_i, i = 1, \cdots, n$ 两两独立且概率分布均为

$$P\left\{\xi_i = 1\right\} = p, P\left\{\xi_i = 0\right\} = 1 - p = q, \quad i = 1, \cdots, n$$

对于一切 n 和 $1 \leqslant k \leqslant n$, 记 $P_n(k) = \mathrm{C}_n^k p^k q^{n-k}$, 则有

$$\sum_{\left\{k: \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k) \underset{n \to \infty}{\longrightarrow} 0$$

♡

证明 (概率方法) 设 $S_0(\omega) \equiv 0, S_k = \xi_1 + \cdots + \xi_k (k = 1, \cdots, n)$, 由例 (5.32) 已证明 $E(S_n) = np$, 则有 $E\left(\frac{S_n}{n}\right) = p$ 。而若 ξ 为任意随机变量, 则由 Markov 不等式 (5.14) 有对任意 $\varepsilon > 0$ 都满足

$$P\{|\xi| \geqslant \varepsilon\} \leqslant \frac{E(|\xi|)}{\varepsilon}$$

$$P\{|\xi| \geqslant \varepsilon\} = P\left\{\xi^2 \geqslant \varepsilon^2\right\} \leqslant \frac{E(\xi^2)}{\varepsilon^2}$$

$$P\{|\xi - E(\xi)| \geqslant \varepsilon\} \leqslant \frac{D(\xi)}{\varepsilon^2}$$

利用最后一个不等式, 令 $\xi = \dfrac{S_n}{n}$, 则有

$$P\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{D\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{D(S_n)}{n^2\varepsilon^2} = \frac{npq}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2}$$

进而

$$P\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}$$

对于一切 n 和 $1 \leqslant k \leqslant n$, 记 $P_n(k) = \mathrm{C}_n^k p^k q^{n-k}$, 则有

$$P\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\left\{k: \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k) \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \underset{n \to \infty}{\longrightarrow} 0$$

命题即证

注 Johann Bernoulli 曾利用分析方法来证明该定理，但方法繁杂，且较难应用。而 Laplace 使用概率方法得到的概率 $P_n(k)$ 的渐进公式特别重要，不仅证明了 Bernoulli 大数定律，且得到了更精确的 De Moivre-Laplace 局部极限定理 (5.16) 以及 De Moivre-Laplace 积分定理 (5.18)

注 Bernoulli 大数定律的估计比较粗糙，其不够精确是由于 Markov 不等式 (5.14) 较弱

该定理的实质在于，对于充分大 n 的和至少满足 $k \sim np$ 的 k，有

$$P_n(k) \sim \frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}$$

而

$$\sum_{\left\{k:\left|\frac{k}{n}-p\right|\leqslant\varepsilon\right\}} P_n(k) \sim \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$$

定理 5.16 (De Moivre-Laplace **局部极限定理**/De Moivre-Laplace local limit theorem/локальная теорема Муавра-Лап

对于一切 n 和 $1 \leqslant k \leqslant n$，记 $P_n(k) = \mathrm{C}_n^k p^k q^{n-k}, q = 1-p$，则对于满足 $|k-np| = o(npq)^{\frac{2}{3}}$ 的所有 k，一致地有

$$P_n(k) \sim \frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}$$

即当 $n \to \infty$ 时

$$\sup_{\{k:|k-np|\leqslant\varphi(n)\}}\left|\frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}}-1\right| \to 0$$

♡

证明 注意到 Stirling 公式，即 $n! = \sqrt{2\pi n}\mathrm{e}^{-n}n^n(1+R(n))$，其中 $R(n) \underset{n\to\infty}{\longrightarrow} 0$。由 Stering 公式即得若 $n\to\infty, k\to\infty, n-k\to\infty$，则有

$$\begin{aligned}
\mathrm{C}_n^k &= \frac{n!}{k!(n-k)!} = \frac{\sqrt{2\pi n}\mathrm{e}^{-n}n^n(1+R(n))}{\sqrt{2\pi k\times 2\pi(n-k)}\mathrm{e}^{-k}k^k(n-k)^{n-k}\mathrm{e}^{-(n-k)}(1+R(k))(1+R(n-k))}\\
&= \frac{1}{\sqrt{2\pi n\frac{k}{n}\left(1-\frac{k}{n}\right)}}\times\frac{1+\varepsilon(n,k,n-k)}{\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k}}
\end{aligned}$$

其中当 $n\to\infty, k\to\infty, n-k\to\infty$ 时，$\varepsilon(n,k,n-k)\to 0$。这时有

$$P_n(k) = \mathrm{C}_n^k p^k q^{n-k} = \frac{1}{\sqrt{2\pi n\frac{k}{n}\left(1-\frac{k}{n}\right)}}\times\frac{p^k(1-p)^{n-k}(1+\varepsilon)}{\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k}}$$

记 $\widehat{p} = \dfrac{k}{n}$，则有

$$\begin{aligned}
P_n(k) &= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\left(\frac{p}{\widehat{p}}\right)^k\left(\frac{1-p}{1-\widehat{p}}\right)^{n-k}(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{k\ln\frac{p}{\widehat{p}}+(n-k)\ln\frac{1-p}{1-\widehat{p}}\right\}\times(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{n\left[\frac{k}{n}\ln\frac{p}{\widehat{p}}+\left(1-\frac{k}{n}\right)\ln\frac{1-p}{1-\widehat{p}}\right]\right\}\times(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\{-nH(\widehat{p})\}\times(1+\varepsilon)
\end{aligned}$$

其中

$$H(x) = x\ln\frac{x}{p}+(1-x)\ln\frac{1-x}{1-p}$$

这时所考虑的 k 满足 $|k-np| = o(npq)^{\frac{2}{3}}$，故 $p-\widehat{p}\underset{n\to\infty}{\longrightarrow}0$。又由当 $0<x<1$ 时

$$H'(x) = \ln\frac{x}{p}-\ln\frac{1-x}{1-p},\quad H''(x) = \frac{1}{x}+\frac{1}{1-x},\quad H'''(x) = -\frac{1}{x^2}+\frac{1}{(1-x)^2}$$

因此，若将 $H(\widehat{p})$ 表示为 $H(p+(\widehat{p}-p))$，并利用 Taylor 公式有当 n 充分大时有

$$\begin{aligned} H(\widehat{p}) &= H(p)+H'(p)(\widehat{p}-p)+\frac{1}{2}H''(p)(\widehat{p}-p)^2+O\left(|\widehat{p}-p|^3\right) \\ &= \frac{1}{2}\left(\frac{1}{p}+\frac{1}{q}\right)(\widehat{p}-p)^2+O\left(|\widehat{p}-p|^3\right) \end{aligned}$$

则有

$$P_n(k)=\frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{-\frac{n}{2pq}(\widehat{p}-p)^2+nO\left(|\widehat{p}-p|^3\right)\right\}\times(1+\varepsilon)$$

注意到

$$\frac{n}{2pq}(\widehat{p}-p)^2=\frac{n}{2pq}\left(\frac{k}{n}-p\right)^2=\frac{(k-np)^2}{2npq}$$

则有

$$P_n(k)=\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}\left(1+\varepsilon'(n,k,n-k)\right)$$

其中

$$1+\varepsilon'(n,k,n-k)=[1+\varepsilon(n,k,n-k)]\mathrm{e}^{nO\left(|\hat{p}-p|^3\right)}\sqrt{\frac{p(1-p)}{\widehat{p}(1-\widehat{p})}}$$

由上述讨论即得 $\sup|\varepsilon'(n,k,n-k)|\underset{n\to\infty}{\longrightarrow}0$，其中 sup 为对满足 $|k-np|\leqslant\varphi(n)=o(npq)^{\frac{2}{3}}$ 的 k 来求上确界

注 该定理于 1730 年得到

注 该定理的结论，可表述为如下等价形式：

对于一切 $x\in\mathbb{R}^1$，若 $x=o(npq)^{\frac{1}{6}}$，而 $np+x\sqrt{npq}$ 为集合 $\{0,1,\cdots,n\}$ 中的整数，则

$$P_n(np+x\sqrt{npq})\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}}$$

即当 $n\to\infty$ 时，有

$$\sup_{\{x:|x|\leqslant\psi(n)\}}\left|\frac{P_n(np+x\sqrt{npq})}{\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}}}-1\right|\to 0$$

其中 $\psi(n)=o(npq)^{\frac{1}{6}}$

再利用概率的语言将上面得到的结果表述为：

$$\begin{aligned} P\left\{S_n=k\right\} &\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}},\quad |k-np|=o(npq)^{\frac{2}{3}} \\ P\left\{\frac{S_n-np}{\sqrt{npq}}=x\right\} &\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}},\quad x=o(npq)^{\frac{1}{6}} \end{aligned}$$

假如设

$$t_k=\frac{k-np}{\sqrt{npq}},\quad \Delta t_k=t_{k+1}-t_k=\frac{1}{\sqrt{npq}}$$

则形式变为

$$P\left\{\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}=t_k\right\}\sim\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}},\quad t_k=o(npq)^{\frac{1}{6}}$$

这时有

$$\Delta t_k=\frac{1}{\sqrt{npq}}\underset{n\to\infty}{\longrightarrow}0$$

由此即得

$$P\left\{a<\frac{S_n-np}{\sqrt{npq}}\leqslant b\right\}\sim\frac{1}{\sqrt{2\pi}\cdot\sqrt{npq}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x,\quad -\infty<a\leqslant b<\infty$$

综上所述定理 (5.16) 可以改述为定理 (5.17) 的形式

定理 5.17 (De Moivre-Laplace 局部极限定理等价形式)

设参数为 n,p 的 Bernoulli 试验，则对于任意有限区间 $(a;b)$，若

$$a < x_{n,m} = \frac{m-np}{\sqrt{np(1-p)}} < b$$

固定 p，当 $n\to\infty$ 时，则对于所有 m 一致有

$$b(n,p,m) \sim \frac{1}{\sqrt{np(1-p)}}\varphi\left(x_{n,m}\right) = \frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x, \quad -\infty < a \leqslant b < \infty$$

其中

$$\varphi(x_{n,m}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

例题 5.46 (De Moivre-Laplace 局部极限定理)

一个对称的硬币被抛掷 100 次，求恰好出现 50 次的概率

解 问题为满足 $n=100, p=0.5, m=50$ 的二项概型，则计算

$$x_{n,m} = \frac{m-np}{\sqrt{np(1-p)}} = \frac{50-100\cdot 0.5}{\sqrt{100\cdot 0.5\cdot 0.5}} = 0$$

由 De Moivre-Laplace 局部极限定理 (5.17) 即得

$$b(100,0.5,50) \approx \varphi(0)\frac{1}{\sqrt{100\cdot 0.5\cdot 0.5}} = \frac{0.3989}{5} \approx 0.08$$

即为所求概率

定理 5.18 (De Moivre-Laplace 积分定理/интегральная теорема Муавра-Лапласа)

设 $0<p<1$，若设

$$P_n(k) = \mathrm{C}_n^k p^k q^{n-k}, \quad P_n(a,b] = \sum_{a<x\leqslant b} P_n(np + x\sqrt{npq})$$

则有

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|P_n(a;b] - \frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right| \underset{n\to\infty}{\longrightarrow} 0$$

证明 对于 $-\infty < a \leqslant b < \infty$，设

$$P_n(a;b] = \sum_{a<x\leqslant b} P_n(np + x\sqrt{npq})$$

其中求和对象为一切使 $np + x\sqrt{npq}$ 为整数的 x。由 De Moivre-Laplace 局部极限定理 (5.16) 有，对于由 $k = np + t_k\sqrt{npq}$ 确定且满足条件 $|t_k| \leqslant T < \infty$ 的 t_k，有

$$P_n\left(np + t_k\sqrt{npq}\right) = \frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}\left[1 + \varepsilon\left(t_k, n\right)\right]$$

其中 $\sup\limits_{|t_k|\leqslant T}\left|\varepsilon\left(t_k,n\right)\right| \underset{n\to 0}{\longrightarrow} \infty$，则对于固定的 $a,b(-T\leqslant a\leqslant b\leqslant T, T<\infty)$ 满足

$$\begin{aligned}\sum_{a<t_k\leqslant b} P_n\left(np + t_k\sqrt{npq}\right) &= \sum_{a<t_k\leqslant b}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}} + \sum_{a<t_k\leqslant b}\varepsilon\left(t_k,n\right)\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x + R_n^{(1)}(a,b) + R_n^{(2)}(a,b)\end{aligned} \tag{5.5}$$

其中

$$R_n^{(1)}(a,b) = \sum_{a<t_k\leqslant b}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}} - \frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$$

$$R_n^{(2)}(a,b) = \sum_{a<t_k\leqslant b}\varepsilon\left(t_k,n\right)\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}$$

注意到

$$\sup_{-T\leqslant a\leqslant b\leqslant T}\left|R_n^{(1)}(a,b)\right| \underset{n\to 0}{\longrightarrow} 0 \tag{5.6}$$

且有

$$\begin{aligned}\sup_{-T\leqslant a\leqslant b\leqslant T}\left|R_n^{(2)}(a,b)\right| &\leqslant \sup_{|t_k|\leqslant T}\left|\varepsilon\left(t_k,n\right)\right|\sum_{|t_k|\leqslant T}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}\\ &\leqslant \sup_{|t_k|\leqslant T}\left|\varepsilon\left(t_k,n\right)\right|\times\left[\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x+\sup_{-T\leqslant a\leqslant b\leqslant T}\left|R_n^{(1)}(a,b)\right|\right]\underset{n\to 0}{\longrightarrow} 0\end{aligned} \tag{5.7}$$

其右侧收敛于 0 是因为式 (6.6) 与式 (5.8)

$$\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\leqslant\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=1 \tag{5.8}$$

下记

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

则由式 (6.5)(6.6)(5.7) 有

$$\sup_{-T\leqslant a\leqslant b\leqslant T}\left|P_n(a,b]-[\Phi(b)-\Phi(a)]\right|\underset{n\to\infty}{\longrightarrow} 0 \tag{5.9}$$

下证式 (5.9) 不仅对于有限 T 成立，且对于 $T=\infty$ 也成立

由式 (5.8) 有

$$(\forall\varepsilon>0)(\exists T=T(\varepsilon)\in\mathbb{R}):\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x>1-\frac{\varepsilon}{4} \tag{5.10}$$

再由式 (5.9) 有

$$(\forall\varepsilon>0)(\exists N\in\mathbb{N})(\forall n>N)(\forall T=T(\varepsilon)):\sup_{-T\leqslant a\leqslant b\leqslant T}\left|P_n(a,b]-[\Phi(b)-\Phi(a)]\right|<\frac{\varepsilon}{4} \tag{5.11}$$

因此由式 (5.10) 和式 (5.11) 即得 $P_n(-T;T]>1-\dfrac{\varepsilon}{2}$，因此

$$P_n(-\infty;-T]+P_n(T;\infty)\leqslant\frac{\varepsilon}{2}$$

其中

$$P_n(-\infty;T]=\lim_{S\downarrow-\infty}P_n(S;T],P_n(T;\infty)=\lim_{S\uparrow\infty}P_n(T;S]$$

综上对于任意 $-\infty\leqslant -T\leqslant a\leqslant b\leqslant T\leqslant\infty$ 有

$$\begin{aligned}&\left|P_n(a,b]-\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right|\leqslant\left|P_n(-T,T]-\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right|\\ &+\left|P_n(a,-T]-\frac{1}{\sqrt{2\pi}}\int_a^{-T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right|+\left|P_n(T,b]-\frac{1}{\sqrt{2\pi}}\int_T^b\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right|\\ \leqslant&\frac{\varepsilon}{4}+P_n(-\infty,-T]+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-T}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x+P_n(T,\infty)+\frac{1}{\sqrt{2\pi}}\int_T^{\infty}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\\ \leqslant&\frac{\varepsilon}{4}+\frac{\varepsilon}{2}+\frac{\varepsilon}{8}+\frac{\varepsilon}{8}=\varepsilon\end{aligned}$$

再注意到式 (5.9)，由此易证 $P_n(a,b]$ 关于 $-\infty\leqslant a\leqslant b\leqslant\infty$ 一致趋向 $\Phi(b)-\Phi(a)$

注 该定理可以改述为定理 (5.19) 的形式

定理 5.19 (De Moivre-Laplace 积分定理等价形式)

设 $0<p<1$，若设

$$P_n(k)=\mathrm{C}_n^k p^k q^{n-k},\quad P_n(a,b]=\sum_{a<x\leqslant b}P_n(np+x\sqrt{npq})$$

则有

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|P\left\{a<\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}\leqslant b\right\}-\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right|\underset{n\to\infty}{\longrightarrow}0 \tag{5.12}$$

并且对于任意 $-\infty\leqslant A<B\leqslant\infty$ 有

$$P\{A<S_n\leqslant B\}-\left[\Phi\left(\frac{B-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{A-np}{\sqrt{np(1-p)}}\right)\right]\underset{n\to\infty}{\longrightarrow}0$$

♡

注 也可以改述为设参数为 n,p 的 Bernoulli 试验，$n\to\infty$，且 p 固定，则对于所有 $m_1<m_2$ 一致有

$$P\left(m_1\leqslant S_n<m_2\right)\sim\Phi\left(x_{n,m_2}\right)-\Phi\left(x_{n,m_1}\right)$$

其中

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

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|P\left\{a<\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}\leqslant b\right\}-\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right|\underset{n\to\infty}{\longrightarrow}0 \tag{5.13}$$

并且对于任意 $-\infty\leqslant A<B\leqslant\infty$ 有

$$P\{A<S_n\leqslant B\}-\left[\Phi\left(\frac{B-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{A-np}{\sqrt{np(1-p)}}\right)\right]\underset{n\to\infty}{\longrightarrow}0$$

例题 5.47 （De Moivre-Laplace 积分定理）

一个对称的硬币被抛掷 100 次，求正面出现 40 到 60 次范围内的概率

解 取 $n=100,p=0.5,m_1=40,m_2=60$，由 De Moivre-Laplace 积分定理 (5.18) 可知

$$P\left(m_1\leqslant S_n<m_2\right)\approx\Phi_0\left(x_{n,m_2}\right)-\Phi_0\left(x_{n,m_1}\right)$$

$$x_{n,m_2}=\frac{60-100\cdot 0.5}{\sqrt{100\cdot 0.5\cdot 0.5}}=\frac{10}{5}=2,\quad x_{n,m_1}=\frac{40-100\cdot 0.5}{\sqrt{100\cdot 0.5\cdot 0.5}}=-\frac{10}{5}=-2$$

$$P\left(40\leqslant S_n<60\right)\approx\Phi_0(2)-\Phi_0(-2)=2\Phi_0(2)=2\cdot 0.4772=0.9544$$

即为所求概率

定理 5.20 (二项概率正态逼近)

设 $0<p<1$，若设 $P_n(k)=\mathrm{C}_n^k p^k q^{n-k}$ 及

$$P_n(a,b]=\sum_{a<x\leqslant b}P_n(np+x\sqrt{npq}),\quad P_n(-\infty;x]=P\left\{\frac{S_n-np}{\sqrt{np(1-p)}}\leqslant x\right\}$$

则有

$$\sup_{-\infty\leqslant x\leqslant\infty}\left|P_n(-\infty;x]-\Phi(x)\right|\leqslant\frac{p^2+(1-p)^2}{\sqrt{np(1-p)}}$$

♡

证明 由 De Moivre-Laplace 局部极限定理局部极限定理 (5.16) 有对于 $x=o(np(1-p))^{\frac{1}{6}}$，概率 $P_n(np+x\sqrt{np(1-p)})$ 的值近似地位于曲线

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

上，而由 De Moivre-Laplace 积分定理 (5.18) 知，概率

$$\begin{aligned}P_n(a;b]&=P\left\{a\sqrt{np(1-p)}<S_n-np\leqslant b\sqrt{np(1-p)}\right\}\\&=P\left\{np+a\sqrt{np(1-p)}<S_n\leqslant np+b\sqrt{np(1-p)}\right\}\end{aligned}$$

的值可以较好地由积分

$$\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$$

逼近，则由式 (5.13) 可见

$$\sup_{-\infty\leqslant x\leqslant\infty}|F_n(x)-\Phi(x)| \underset{n\to\infty}{\longrightarrow} 0$$

利用 A. C. Berry-C. G. Esseen 定理得

$$\sup_{-\infty\leqslant x\leqslant\infty}|F_n(x)-\Phi(x)| \leqslant \frac{p^2+(1-p)^2}{\sqrt{np(1-p)}}$$

定理即证

注 后文给出 A. C. Berry-C. G. Esseen 定理的证明

注 定理也可以改述为对于任意 $m_1 < m_2$ 有估计

$$\left|P\left(m_1 \leqslant S_n < m_2\right)-\left(\Phi\left(x_{n,m_2}\right)-\Phi\left(x_{n,m_1}\right)\right)\right| \leqslant \frac{p^2+(1-p)^2}{\sqrt{np(1-p)}}$$

第 6 章 随机向量 (Random Vector)

6.1 基本概念

6.1.1 随机向量及其分布 Random Vector and Its Distribution

定义 6.1 (随机向量)

若向量 $\xi=(\xi_1,\cdots,\xi_n)$ 其各分量均为定义在同一概率空间 $(\Omega,\mathcal{A},P)$ 上的随机变量，则称该向量为定义在概率空间 $(\Omega,\mathcal{A},P)$ 上的 n 维随机向量 (random vector) ♣

定理 6.1 (随机向量充要条件)

向量 $(X_1,\cdots,X_n)$ 为定义在概率空间 $(\Omega,\mathcal{A},P)$ 上的 n 维随机向量的充要条件为 $(\forall\,(x_1,\cdots,x_n)\in\mathbb{R}^n)$:

$$\{\omega\mid X_1(\omega)\leqslant x_1,\cdots,X_n(\omega)\leqslant x_n\}=(X_1\leqslant x_1,\cdots,X_n\leqslant x_n)\in\mathcal{A} \tag{6.1}$$

♡

证明 必要性：若 $(X_1,\cdots,X_n)$ 为一个概率空间 $(\Omega,\mathcal{A},P)$ 上的 n 维随机向量，则 $X_1,\cdots,X_n$ 即为定义在同一个概率空间上的 n 个随机变量，则

$$(X_1\leqslant x_1,\cdots,X_n\leqslant x_n)=\bigcap_{j=1}^{n}(X_j\leqslant x_j)\in\mathcal{A}$$

充分性：设对任意 n 维实值向量 $(x_1,\cdots,x_n)$ 都有式 (6.1) 成立，则 $(\forall k)(\forall x\in\mathbb{R})(\forall M>0)$:

$$(X_1\leqslant M,\cdots,X_{k-1}\leqslant M,X_k\leqslant x,X_{k+1}\leqslant M,\cdots,X_n\leqslant M)\quad(X_k\leqslant x)\bigcap_{j\neq k}(X_j\leqslant M)\in\mathcal{A}$$

若令 $M\uparrow\infty$，则有

$$\bigcap_{j\neq k}(X_j\leqslant M)\uparrow\Omega,$$

则上式中的事件趋于 $(X_k\leqslant x)$，则有 $(X_k\leqslant x)\in\mathcal{A}$，即 X_k 为随机变量

注 实际上式 (6.1) 等价于对任意 n 维 Borel 集合 B, 都有

$$((X_1,\cdots,X_n)\in B)=\{\omega\mid(X_1(\omega),\cdots,X_n(\omega))\in B\}\in\mathcal{A}$$

定义 6.2 (随机向量概率分布)

设 n 维随机向量 ξ，若在 $\mathcal{B}_n$ 上 σ 代数上定义函数 P_ξ 满足

$$P_\xi(B)=P(\xi\in B)$$

则称 P_ξ 为随机向量 ξ 的联合概率分布律 (joint law of probability distribution)，不引起歧义时可简称概率分布 (probability distribution) ♣

定义 6.3 (随机向量联合分布函数)

设 n 维随机向量 $\xi=(\xi_1,\ldots,\xi_n)$，若函数 $F_\xi(x):\mathbb{R}^n\to\mathbb{R}$ 满足

$$(\forall x=(x_1,\ldots,x_n)\in\mathbb{R}^n):F_\xi\left(x_1,\ldots,x_n\right)=P\left(\xi_1\leqslant x_1,\ldots,\xi_n\leqslant x_n\right)$$

则称函数 $F_\xi(x)$ 为 n 维随机向量 ξ 的联合分布函数 (joint distribution function) ♣

注 （随机向量联合分布函数）

也可以把随机向量联合分布函数的定义改为

$$(\forall x=(x_1,\ldots,x_n)\in\mathbb{R}^n): F_\xi(x_1,\ldots,x_n)=P(\xi_1<x_1,\ldots,\xi_n<x_n)$$

性质 （随机向量联合分布函数特征性质）

随机向量 $\xi=(\xi_1,\ldots,\xi_n)$ 的联合分布函数 $F_\xi(x)=F_\xi(x_1,\ldots,x_n)$ 有下列性质：

1) （非降性）$F_\xi(x)$ 在每个分量 $x_i, i=\overline{1;n}$ 上不减
2) （左连续性）$F_\xi(x)$ 在每个分量 $x_i, i=\overline{1;n}$ 上左连续
3) （规范性）对每个分量 $x_i, i=\overline{1;n}$ 都有 $F_\xi(x)\underset{x_i\to-\infty}{\longrightarrow}0$；
   当所有分量 $x_i, i=\overline{1;n}$ 同时有 $x_i\to\infty$ 时有 $F_\xi(x)\to 1$
4) （增量非负性）对于 $\forall x=(x_1,\ldots,x_n)\in\mathbb{R}^n$ 有

$$\Delta h_1\ldots\Delta h_n F_\xi(x_1,\ldots,x_n)=\sum\operatorname{sgn}(x)F(x)\geqslant 0$$

其中

$$\Delta h_i F_\xi(x_1,\ldots,x_n)=F_\xi(x_1,\ldots x_i+h_i,\ldots,x_n)-F_\xi(x_1,\ldots x_i,\ldots,x_n)$$

证明 1)2) 显然

3) 对任意 $j, 1\leqslant j\leqslant n$ 都有

$$(X_1\leqslant x_1,\cdots,X_n\leqslant x_n)=(X_j\leqslant x_j)\bigcap_{k\neq j}(X_k\leqslant x_k)\underset{x_j\to-\infty}{\longrightarrow}\varnothing$$

当且仅当对所有的 $j, 1\leqslant j\leqslant n$ 都同时有 $x_j\to\infty$ 时有

$$(X_1\leqslant x_1,\cdots,X_n\leqslant x_n)=\bigcap_{j=1}^{n}(X_j\leqslant x_j)\to\Omega$$

4) 由 $P(x_1\leqslant\xi_1<x_1+h_1,\ldots,x_n\leqslant\xi_n<x_n+h_n)=\Delta h_1\ldots\Delta h_n F_\xi(x_1,\ldots,x_n)$ 即得

注 同一维情况一样，对于任意一个具有性质 1)2)3)4) 的 n 元函数 $F(x_1,\cdots,x_n)$ 都可以找到一个恰当的概率空间 $(\Omega,\mathcal{A},P)$ 在其上定义一个 n 维随机向量，使得该随机向量的分布函数即为 $F(x_1,\cdots,x_n)$

例题 6.1 （随机向量联合分布函数特征性质）

试求如下函数是否为某个二维随机向量的联合分布函数：

$$F(x,y)=I_{(x+y\geqslant 1)},\quad (x,y)\in\mathbb{R}^2$$

解 二维随机向量的联合分布函数必须满足性质 1)2)3)4)，而对该函数却有

$$\Delta_{(0,0)}^{(1,1)}F=F(1,1)-F(1,0)-F(0,1)+F(0,0)=-1$$

即不满足增量非负性，则其不是二维分布函数

性质 （随机向量联合分布函数性质）

随机向量 $\xi=(\xi_1,\ldots,\xi_n)$ 的分布函数 $F_\xi(x)=F_\xi(x_1,\ldots,x_n)$ 有下列性质

5) （联合分布函数基本值域）$(\forall x\in\mathbb{R}^n): 0\leqslant F_\xi(x)\leqslant 1$
6) （降维公式）$F_\xi(x_1,\ldots x_{i-1},\infty,x_{i+1},\ldots,x_n)$ 为随机向量 $(\xi_1,\ldots\xi_{i-1},\xi_{i+1},\ldots,\xi_n)$ 的联合分布函数

6.1.2 离散型随机向量

Discrete Random Vector

定义 6.4 (随机向量离散分布)

设随机向量 $\xi=(\xi_1,\ldots,\xi_n)$，存在至多可数集 $X\subset R^n$ 满足 $P(\xi\in X)=1$，若 $x=(x_1,\ldots,x_n)\in X$ 为随机向量 ξ 可能的值，则称 $p(x)=P(\xi=x)$ 为出现值 x 的 概率 ((probability)，称该随机向量的概率分布为离散分布 ((discrete distribution/дискретное распределение)，称概率分布为离散分布的随机向量为离散型随机向量简称离散随机向量 (discrete random variable)

♣

例题 6.2 （二维离散随机向量联合分布律）

将两个不同的小球随机地放入 3 个带有编号 1,2,3 的盒子。以 X 表示空盒的个数，以 Y 表示放有小球的盒子的最小编号，求 (X,Y) 的联合概率分布律

解 易知 X 的取值集合为 $\{1,2\}$，而 Y 的取值集合为 $\{1,2,3\}$。当 $X=2$ 时，两个小球放在同一个盒中，并且该盒子的编号 j 即为放有小球的盒子的最小编号，即 Y 的值。由于每个小球等可能地放入每个盒子，则有

$$p_{2j}=P(X=2,Y=j)=\frac{1}{9},\quad j=1,2,3$$

当 $X=1$ 时，两个小球放在两个不同盒中，此时放有小球的盒子的最小编号只能为 1 或 2，则有

$$p_{13}=P(X=1,Y=3)=0$$

而当放有小球的盒子的最小编号为 1 时，有一个球放在 1 号盒中，另一个球可以放在 2 号或 3 号盒中，则有

$$p_{11}=P(X=1,Y=1)=\frac{4}{9}$$

而当放有小球的盒子的最小编号为 2 时，两个球分别放在 2 号和 3 号盒中，则有

$$p_{12}=P(X=1,Y=2)=\frac{2}{9}$$

列一个矩阵表示上述结果：

$$\begin{pmatrix} (1,1) & (1,2) & (2,1) & (2,2) & (2,3) \\ \frac{4}{9} & \frac{2}{9} & \frac{1}{9} & \frac{1}{9} & \frac{4}{9} \end{pmatrix}$$

也可以表示为

Y/X	1	2	3
1	$\frac{4}{9}$	$\frac{2}{9}$	0
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

其为 (X,Y) 的联合概率分布律的分布表

注 （离散随机向量标准形）

随机向量 ξ 的每个坐标 ξ_k 的概率分布均为离散分布。设 $X^{(k)}=\left\{x_1^{(k)},x_2^{(k)},\ldots,x_{i_k}^{(k)},\ldots\right\}$ 为随机变量 ξ_k 的值的集合，则构成 R^n 上集合

$$X=X^{(1)}\times\ldots X^{(n)}=\left\{x=(x_1,\ldots,x_n):x_k\in X^{(k)},k=\overline{1,n}\right\}$$

可证明 $P(\xi\in X)=1$，则 X 可作为随机向量 ξ 的值的集合。而对于任意向量 $x=\left(x_{i_1}^{(1)},\ldots,x_{i_n}^{(n)}\right)$，其中 $x_{i_k}^{(k)}\in X^{(k)}$，可表示为

$$P_{i_1\ldots i_n}=P(\xi=x)=P\left(\xi_1=x_{i_1}^{(1)},\ldots,\xi_n=x_{i_n}^{(n)}\right)$$

即为随机向量 ξ 的值 x 出现的概率。通过这样选择集合 X，它的一些元素将以 0 的概率出现

性质 （离散随机向量性质）

离散随机向量 $\xi=(\xi_1,\ldots,\xi_n)$ 的分布 $(X,\{P_{i_1,\ldots,i_n}\})$ 具有下列性质：

1) （非负性） $(\forall x=\left(x_{i_1}^{(1)},\ldots,x_{i_n}^{(n)}\right)\in X):P_{i_1,\ldots,i_n}\geqslant 0$
2) （规范性） $\sum\limits_{i_1,\ldots,i_n}P_{i_1,\ldots,i_n}=1$
3) （降维公式）

$$P\left(\xi_1=x_{i_1}^{(1)},\ldots\xi_{k-1}=x_{i_{k-1}}^{(k-1)},\xi_{k+1}=x_{i_{k+1}}^{(k+1)},\ldots,\xi_n=x_{i_n}^{(n)}\right)=\sum_{i_k}P_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_n}$$

例题 6.3 （多项分布/随机向量离散分布）

设离散概率空间 $(\Omega,\mathcal{A},P)$，样本空间上 r 个不同基本结果出现的概率分别为 $p_1,\cdots,p_r$。设 ξ_k 为 n 次独立重复试验中第 k 个结果出现的次数，则 $\xi=(\xi_1,\cdots,\xi_r)$ 为一个离散随机向量，其值为 $m=(m_1,\ldots,m_r)$，其中 m_i 为非负整数并且 $m_1+\ldots+m_r=n$，

注意到

$$\sum_{m_1+\cdots+m_r=n} C_n\left(n_1,\cdots,n_r\right)p_1^{n_1}\cdots p_r^{n_r}=\left(p_1+\cdots+p_r\right)^n=1$$

即的确定义了一个概率模型

定义 6.5 (多项分布/随机向量离散分布)

设离散概率空间 $(\Omega,\mathcal{A},P)$，在其上定义了离散随机向量 $\xi=(\xi_1,\cdots,\xi_r)$。离散随机向量 ξ 的值定义为 $m=(m_1,\ldots,m_r)$，其中 m_i 为非负整数并且 $m_1+\ldots+m_r=n$，定义概率

$$P(\xi=m)=P\left(m_1,\ldots,m_r\right)=\frac{n!}{m_1!\ldots m_r!}p_1^{m_1}\ldots p_r^{m_r}$$

称其为参数为 $(n;p_1,\ldots p_r)$ 的多项式分布简称参数为 $(n;p_1,\ldots p_r)$ 多项分布 (multinomial distribution/полиномиальное распределение)

♣

例题 6.4 （多元超几何分布）

假设一箱子中有编号为 $1,2,\cdots,M$ 的 M 个不同的球，其中 M_1 个球具有颜色 $b_1,\cdots,M_r$ 个球具有颜色 b_r，且 $M_1+\cdots+M_r=M$。现从箱中进行 $n(n<M)$ 次不放回抽样。假设基本事件等可能，而 $B_{n_1,\cdots,n_r}$ 表示事件“n_1 个球具有颜色 $b_1,\cdots,n_r$ 个球具有颜色 b_r”，且 $n_1+\cdots+n_r=n$，求事件 $B_{n_1,\cdots,n_r}$ 的概率。

其基本事件空间为

$$\Omega=\{\omega:\omega=\left(a_1,\cdots,a_n\right),a_k\neq a_l,k\neq l;a_i=1,\cdots,M\}$$

解 显然 $\operatorname{Card}(\Omega)=(M)_n$。易证 $\operatorname{Card}\left(B_{n_1,\cdots,n_r}\right)=\mathrm{C}_n\left(n_1,\cdots,n_r\right)\left(M_1\right)_{n_1}\cdots\left(M_r\right)_{n_r}$，则有

$$P\left(B_{n_1,\cdots,n_r}\right)=\frac{\operatorname{Card}\left(B_{n_1,\cdots,n_r}\right)}{\operatorname{Card}(\Omega)}=\frac{\mathrm{C}_{M_1}^{n_1}\cdots\mathrm{C}_{M_r}^{n_r}}{\mathrm{C}_M^n}$$

即为事件 $B_{n_1,\cdots,n_r}$ 的概率。

注 称概率组 $\{P\left(B_{n_1,\cdots,n_r}\right)\}$ 为**多元超几何分布** (multivariate hypergeometric distribution)。特别地，当 $r=2$ 时称此分布为**超几何分布** (hypergeometric distribution)（因为其母函数为超几何函数）。注意到

$$P\left(B_{n_1,n_2}\right)=\frac{\mathrm{C}_{M_1}^{n_1}\mathrm{C}_{M_2}^{n_2}}{\mathrm{C}_M^n},n_1+n_2=n,M_1+M_2=M$$

若当 $M\to\infty,M_1\to\infty$ 时有 $\dfrac{M_1}{M}\to p$，则有 $\dfrac{M_2}{M}\to 1-p$，则 $P\left(B_{n_1,n_2}\right)\to \mathrm{C}_{n_1+n_2}^{n_2}p^{n_1}(1-p)^{n_2}$。换言之，在上述条件下超几何分布逼近二项分布。这在直观上显然，因为当 M 和 M_1 较大时，由不放回抽样得到几乎与放回抽样一样的结果

6.1.3 连续型随机向量

Continuous Random Vector

定义 6.6 (随机向量绝对连续分布)

设随机向量 $\xi=(\xi_1,\ldots,\xi_n)$ 定义在概率空间 $(\Omega,\mathcal{B},P)$ 上，若存在实值函数 $\rho_\xi(x),x\in R^n$ 满足

$$(\forall B\in\mathcal{B}):P_\xi(B)=P(\xi\in B)=\int_B\rho_\xi(x)dx$$

则称随机向量 ξ 的概率分布为绝对连续分布 (absolutely continuous distribution/абсолютно непрерывное распределение)，或连续概率分布 (continuous probability distribution) 简称连续型分布，连续分布，称随机变量 ξ 为连续型随机变量 (continuous random variable) 简称连续随机变量，称函数 $\rho_\xi(x)$ 为随机变量 ξ 的联合概率分布密度简称联合分布密度，不引起歧义时可简称分布密度 (distribution density /плотность распределения)，或称为联合概率密度函数 (joint probability density function) 简称联合密度函数，不引起歧义时可简称密度函数

性质 （连续随机向量性质）

设随机向量 $\xi=(\xi_1,\ldots,\xi_n)$ 的概率分布为绝对连续分布，其密度函数为 $\rho_\xi(x), x\in R^n$，则有下列性质成立：

1)（非负性）$(\forall x\in R^n):\rho_\xi(x)\geqslant 0$

2)（规范性）

$$\int_{R^n}\rho_\xi(x)dx=1$$

3)（降维公式）连续随机向量 $\tilde{\xi}=(\xi_1,\ldots,\xi_{k-1},\xi_{k+1},\ldots,\xi_n)$ 的密度函数为

$$\rho_{\tilde{\xi}}\left(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n\right)=\int_{-\infty}^{\infty}\rho_\xi\left(x_1,\ldots,x_{k-1},x_k,x_{k+1},\ldots,x_n\right)dx_k$$

4)（联合分布函数密度函数积分表示）$(\forall x=(x_1,\ldots,x_n)\in R^n)$：

$$F_\xi\left(x_1,\ldots,x_n\right)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}\rho_\xi\left(y_1,\ldots,y_n\right)dy_n\ldots dy_1$$

5)（联合分布函数与密度函数关系）若 $(x_1,\ldots,x_n)$ 为密度函数 $\rho_\xi(x)$ 的连续点，则有

$$\rho_\xi\left(x_1,\ldots,x_n\right)=\frac{\partial^n F_\xi\left(x_1,\ldots,x_n\right)}{\partial x_1\ldots\partial x_n}$$

定理 6.2 (Fubini 定理)

设 D 为 $\mathbb{R}^n$ 的子区域，$\varphi\left(x_1,x_2,\cdots,x_n\right)$ 为 D 上的非负函数或绝对可积函数，则对区域 D 上的 n 重积分

$$\int_D\varphi\left(x_1,x_2,\cdots,x_n\right)\mathrm{d}x_1\;\mathrm{d}x_2\cdots\mathrm{d}x_n$$

可计算累次积分且积分次序可以交换

定理 6.3

设 (X,Y) 有连续的联合分布函数 $F(x,y)$，定义

$$f(x,y)=\begin{cases}\dfrac{\partial^2F(x,y)}{\partial x\partial y}, & \text{当该混合偏导数存在,}\\ 0, & \text{其他}\end{cases}$$

若

$$\int_{\mathbb{R}^2}f(x,y)\mathrm{d}x\;\mathrm{d}y=1$$

则 $f(x,y)$ 为 (X,Y) 的联合密度函数

命题 6.1 (联合分布函数连续充分条件)

若随机变量 X,Y 的分布函数连续，则它们的联合分布函数 $F(x,y)$ 连续

证明 对于任意一点 (x_0,y_0) 和 $\varepsilon>0$，存在 $\delta>0$ 使得当 $|x-x_0|<\delta,|y-y_0|<\delta$ 时有

$$|F_X(x)-F_X\left(x_0\right)|<\frac{\varepsilon}{2},\quad |F_Y(y)-F_Y\left(y_0\right)|<\frac{\varepsilon}{2}.$$

用 I_1 表示 x,x_0 构成的左开右闭区间，用 I_2 表示 y,y_0 构成的左开右闭区间。仅需 $|x-x_0|<\delta,|y-y_0|<\delta$

就有

$$
\begin{aligned}
|F(x,y)-F(x_0,y_0)| \leqslant & |F(x,y)-F(x_0,y)|+|F(x_0,y)-F(x_0,y_0)| \\
= & |P(X \leqslant x, Y \leqslant y)-P(X \leqslant x_0, Y \leqslant y)| \\
& +|P(X \leqslant x_0, Y \leqslant y)-P(X \leqslant x_0, Y \leqslant y_0)| \\
= & P(X \in I_1, Y \leqslant y)+P(X \leqslant x_0, Y \in I_2) \\
\leqslant & P(X \in I_1)+P(Y \in I_2) \\
= & |F_X(x)-F_X(x_0)|+|F_Y(y)-F_Y(y_0)| \\
< & \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon
\end{aligned}
$$

命题即证

例题 6.5 （连续联合分布函数对连续随机变量不充分性）

设随机向量 (X,Y)，X 在 $[0;1]$ 上均匀分布，$Y=X$，则该随机向量有连续的联合分布函数，但其不是连续型随机变量

解 随机向量 (X,Y) 有连续的联合分布函数

$$
F(x,y)=P(X \leqslant \min\{x,y\})=\begin{cases} 0, & \min\{x,y\} \leqslant 0, \\ \min\{x,y\}, & \min\{x,y\} \in (0;1], \\ 1, & \min\{x,y\}>1. \end{cases}
$$

若 (X,Y) 有联合密度 $f(x,y)$，用 D 表示直线 $y=x$ 上点的全体，则有

$$
1=P(X=Y)=P((X,Y) \in D)=\int_D f(x,y)\mathrm{d}x\ \mathrm{d}y=0
$$

导出矛盾。则有 (X,Y) 没有联合密度，不为连续型随机变量

注 该例表明，即使 $F(x,y)$ 连续，除去有限条直线外 $\dfrac{\partial^2 F(x,y)}{\partial x \partial y}$ 存在且连续仍不能保证 (X,Y) 有联合密度函数

6.1.4 边缘分布与条件分布

Marginal Distribution and Conditional Distribution

定义 6.7 (边缘分布)

设随机向量 $\xi=(\xi_1,\ldots,\xi_n)$，则通过选择它的 k 个坐标，即得新的随机向量 $\tilde{\xi}=(\xi_{i_1},\ldots,\xi_{i_k})$，称 $\tilde{\xi}=(\xi_{i_1},\ldots,\xi_{i_k})$ 为随机向量 ξ 的 k 维边缘随机向量，其中 $k=1,\cdots,n-1$。在不引起歧义时，经常也把一维边缘分布 (одномерное маргинальное распределение) 简称边缘分布 (marginal distribution)

k 维边缘随机向量的分布函数可表示为

$$
\begin{aligned}
F_{i_1,\cdots,i_k}(x_{i_1},\cdots,x_{i_k}) &= P(\xi_{i_1} \leqslant x_{i_1},\cdots,\xi_{i_k} \leqslant x_k)=P\left(\bigcap_{j=1}^{k}(\xi_{i_j} \leqslant x_{i_j})\right) \\
&= P\left(\bigcap_{j=1}^{k}(\xi_{i_j} \leqslant x_{i_j}) \bigcap_{j=k+1}^{n}(\xi_j<\infty)\right) \\
&= F(x_1,\cdots,x_k,\infty,\cdots,\infty) \\
&= \lim_{x_{k+1}\to\infty,\cdots,x_n\to\infty} F(x_1,\cdots,x_k,x_{k+1},\cdots,x_n)
\end{aligned}
$$

♣

例题 6.6 （条件分布引入）

假设二维随机向量 (X,Y) 的联合分布函数为 $F(x,y)$，则随机变量 X 和 Y 的边缘分布分别为 $F_1(x)=$

$F(x,\infty), F_2(y)=F(\infty,y)$ 。若 $a<b$，则有

$$P(a<X\leqslant b)=F_1(b)-F_1(a)=F(b,\infty)-F(a,\infty)$$

$$P(a<Y\leqslant b)=F_2(b)-F_2(a)=F(\infty,b)-F(\infty,a)$$

若 $P(a<Y\leqslant b)>0$，则可对一切 $x\in\mathbb{R}$ 计算条件概率

$$P(X\leqslant x\mid a<Y\leqslant b)=\frac{P(X\leqslant x,a<Y\leqslant b)}{P(a<Y\leqslant b)}$$

不难证明，由此得到的 x 的函数 $P(X\leqslant x\mid a<Y\leqslant b)$ 具有非降性、规范性和右连续性，因此为一个一元分布函数，称其为随机变量 X 在条件 $a<Y\leqslant b$ 下的条件分布函数，记为 $F_1(x\mid a<Y\leqslant b)$

注 特别地，还可以定义诸如 $F_1(x\mid Y\leqslant b), F_1(x\mid Y>a)$ 形式的条件分布函数，仅需作为条件的事件的概率大于 0。对于 Y 的条件分布函数，也可以类似地给出定义

定义 6.8 (条件分布)

设随机向量 $\xi=(\xi_1,\ldots,\xi_m,\xi_{m+1},\ldots,\xi_{m+n})$ ，若固定前 m 个坐标，则称

$$P\left(x_{m+1},\ldots,x_{m+n}\mid x_1,\ldots,x_m\right)=\frac{P\left(\xi_1=x_1,\ldots,\xi_m=x_m,\xi_{m+1}=x_{m+1},\ldots,\xi_{m+n}=x_{m+n}\right)}{P\left(\xi_1=x_1,\ldots,\xi_m=x_m\right)}$$

为 $\xi_{m+1},\ldots,\xi_{m+n}$ 在条件 $\xi_1=x_1,\ldots,\xi_m=x_m$ 下的条件分布 (условное распределение/conditional distribution) ♣

注 （条件分布）更换固定的 m 个坐标的序号以及条件的形式，可得类似的定义

例题 6.7 （边缘分布与条件分布）

设有 n 个编有号码 $1,2,\cdots,n$ 的盒子 $(n\geqslant 3)$ 和 m 个不同的小球，每个小球落入第 k 号盒子的概率为 $p_k(k=1,2,\cdots,n)$，其中 $p_k\geqslant 0, \sum\limits_{k=1}^{n}p_k=1$，各个小球落入各个盒子的事件两两独立，分别以 $X_1,X_2,\cdots,X_n$ 表示落入各个盒子的球数，试求：

1) 随机向量 $(X_1,X_2,\cdots,X_n)$ 的联合分布
2) X_k 的边缘分布, 其中 $k=1,2,\cdots,n$
3) (X_1,X_2) 的边缘分布
4) 在条件 $X_1=m_1$ 下随机向量 $(X_2,\cdots,X_n)$ 的条件分布

解 显然 $(X_1,X_2,\cdots,X_n)$ 的取值集合为

$$\left\{(m_1,m_2,\cdots,m_n)\mid m_1+m_2+\cdots+m_n=m, m_k\in\overline{\mathbb{N}}, k=1,2,\cdots,n\right\}$$

其中 $\overline{\mathbb{N}}$ 表示非负整数集

1）由独立性假设即得：对于任意属于 $(X_1,X_2,\cdots,X_n)$ 取值集合的有序非负整数组 $(m_1,m_2,\cdots,m_n)$ 都有

$$P\left(X_1=m_1,X_2=m_2,\cdots,X_n=m_n\right)=\frac{m!}{m_1!m_2!\cdots m_n!}p_1^{m_1}p_2^{m_2}\cdots p_n^{m_n}$$

2) 对任意 $m_1\in\{0,1,\cdots,n\}, P\left(X_1=m_1\right)$ 等于所有满足条件 $m_2+\cdots+m_n=m-m_1$ 的概率值

$P\left(X_1=m_1, X_2=m_2, \cdots, X_n=m_n\right)$ 的和，则有

$$
\begin{aligned}
P\left(X_1=m_1\right)= & \sum_{\left(m_2, \cdots, m_n\right): m_2+\cdots+m_n=m-m_1} P\left(X_1=m_1, X_2=m_2, \cdots, X_n=m_n\right) \\
= & \sum_{\left(m_2, \cdots, m_n\right): m_2+\cdots+m_n=m-m_1} \frac{m!}{m_1!m_2!\cdots m_n!} p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n} \\
= & \frac{m!}{m_1!\left(m-m_1\right)!} p_1^{m_1}\left(1-p_1\right)^{m-m_1} \\
& \times \sum_{m_2+\cdots+m_n=m-m_1} \frac{\left(m-m_1\right)!}{m_2!\cdots m_n!}\left(\frac{p_2}{1-p_1}\right)^{m_2} \cdots\left(\frac{p_n}{1-p_1}\right)^{m_n} \\
= & \mathrm{C}_m^{m_1} p_1^{m_1}\left(1-p_1\right)^{m-m_1}
\end{aligned}
$$

最后一步是因为倒数第二步中乘号之后的和式即为等式

$$
\left(\frac{p_2}{1-p_1}+\cdots+\frac{p_n}{1-p_1}\right)^{m-m_1}=\left(\frac{1-p_1}{1-p_1}\right)^{m-m_1}=1
$$

左端的展开式。类似可求出其他的 $P\left(X_k=m_k\right)$，亦即

$$
P\left(X_k=j\right)=\mathrm{C}_m^j p_k^j\left(1-p_k\right)^{m-j}, \quad j=0,1, \cdots, n, \quad k=1, \cdots, n
$$

则多项分布 $M_n\left(m ; p_1, p_2, \cdots, p_n\right)$ 的一维边缘分布为二项分布，其中第 k 个一维边缘分布为二项分布 $B\left(m ; p_k\right)$

3) 以此类推，对任意满足 $m_1+m_2 \leqslant m$ 的非负整数 m_1, m_2，概率 $P\left(X_1=m_1, X_2=m_2\right)$ 的值为

$$
\begin{aligned}
P\left(X_1=m_1, X_2=m_2\right) & =\sum_{\left(m_3, \cdots, m_n\right): m_3+\cdots+m_n=m-m_1-m_2} P\left(X_1=m_1, X_2=m_2, \cdots, X_n=m_n\right) \\
& =\sum_{\left(m_3, \cdots, m_n\right): m_3+\cdots+m_n=m-m_1-m_2} \frac{m!}{m_1!m_2!\cdots m_n!} p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n} \\
& =\frac{m!}{m_1!m_2!\left(m-m_1-m_2\right)!} p_1^{m_1} p_2^{m_2}\left(1-p_1-p_2\right)^{m-m_1-m_2}
\end{aligned}
$$

其余二维边缘分布 $P\left(X_i=m_i, X_j=m_j\right)$ 类似即得。

综上，多项分布 $M_n\left(m ; p_1, p_2, \cdots, p_n\right)$ 的二维边缘分布为 $n=3$ 的多项分布，其中 $\left(X_1, X_2\right)$ 的边缘分布为 $M_3\left(m ; p_1, p_2, 1-p_1-p_2\right)$

显然，多项分布 $M_n\left(m ; p_1, p_2, \cdots, p_n\right)$ 的 k 维边缘分布为 $n=k+1$ 的多项分布，而二项分布 $B(m ; p)$ 即为 1 $M_2(m ; p, 1-p)$

4) 对任意满足 $m_2+\cdots+m_n=m-m_1$ 的非负有序整数组 $\left(m_2, \cdots, m_n\right)$ 有

$$
\begin{aligned}
P\left(X_2=m_2, \cdots, X_n=m_n \mid X_1=m_1\right) & =\frac{P\left(X_1=m_1, X_2=m_2, \cdots, X_n=m_n\right)}{P\left(X_1=m_1\right)} \\
& =\frac{\dfrac{m!}{\dfrac{m!}{m_1!m_2!\cdots m_n!} p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n}}}{\dfrac{m!}{m_1!\left(m-m_1\right)!} p_1^{m_1}\left(1-p_1\right)^{m-m_1}} \\
& =\frac{\left(m-m_1\right)!}{m_2!\cdots m_n!}\left(\frac{p_2}{1-p_1}\right)^{m_2} \cdots\left(\frac{p_n}{1-p_1}\right)^{m_n}
\end{aligned}
$$

这恰好为 2) 解答中倒数第二步乘号之后和式中的加项

注 (边缘密度引入)

设随机向量 (X, Y) 的联合密度函数为 $p(x, y)$，于是则有

$$
F(x, y)=P(X \leqslant x, Y \leqslant y)=\int_{-\infty}^x \mathrm{d} u \int_{-\infty}^y p(u, v) \mathrm{d} v, \quad(x, y) \in \mathbb{R}^2
$$

则 X_1 和 X_2 的边缘分布分别为

$$F_1(x)=\int_{-\infty}^{x}\mathrm{d}u\int_{-\infty}^{\infty}p(u,v)\mathrm{d}v,\quad x\in\mathbb{R}$$
$$F_2(y)=\int_{-\infty}^{\infty}\mathrm{d}u\int_{-\infty}^{y}p(u,v)\mathrm{d}v,\quad y\in\mathbb{R}$$

若记

$$p_1(x)=\int_{-\infty}^{\infty}p(x,v)\mathrm{d}v,\quad x\in\mathbb{R};\quad p_2(y)=\int_{-\infty}^{\infty}p(u,y)\mathrm{d}u,\quad y\in\mathbb{R},$$

则 $p_1(x)\geqslant 0,p_2(y)\geqslant 0$ 且有

$$F_1(x)=\int_{-\infty}^{x}p_1(u)\mathrm{d}u,\quad x\in\mathbb{R}$$

则 $p_1(x)$ 为 X 的密度函数，同理有 $p_2(y)$ 为 Y 的密度函数。称函数 $p_1(x)$ 为 X 的边缘密度函数，称 $p_2(x)$ 为 Y 的边缘密度函数

定义 6.9 (边缘密度函数)

设随机向量 $(X_1,\cdots,X_n)$ 的联合密度函数为 $p(x_1,\cdots,x_n)$，则对任意 $1\leqslant k<n$，称函数

$$p_{1,\cdots,k}(x_1,\cdots,x_k)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_k,u_{k+1},\cdots,u_n)\,\mathrm{d}u_{k+1}\cdots\mathrm{d}u_n,\quad (x_1,\cdots,x_k)\in\mathbb{R}^k$$

为 k 维随机向量 $(X_1,\cdots,X_k)$ 的边缘密度函数 (marginal density function) ♣

例题 6.8 （边缘密度函数对联合密度函数不充分性）

设二维随机向量 (X,Y) 的联合密度函数 $p(x,y)$ 满足

$$p(x,y)=x+y,\quad 0<x,y<1$$

1) 试分别求出 X 和 Y 的边缘密度函数和边缘分布函数

2) 如下的函数 $q(x,y)$ 是否为某个二维随机向量 (X,Y) 的联合密度：

$$q(x,y)=\left(x+\frac{1}{2}\right)\left(y+\frac{1}{2}\right),\quad 0<x,y<1$$

解 1) 由边缘密度函数定义即得

$$\begin{aligned}p_1(x)&=\int_{-\infty}^{\infty}p(x,v)\mathrm{d}v=\int_0^1(x+v)\mathrm{d}v=x+\frac{1}{2},\quad x\in(0,1)\\ p_2(y)&=\int_{-\infty}^{\infty}p(u,y)\mathrm{d}u=\int_0^1(u+y)\mathrm{d}u=y+\frac{1}{2},\quad y\in(0,1)\end{aligned}\tag{6.2}$$

即得边缘分布函数

$$F_1(x)=\int_{-\infty}^{x}p_1(u)\mathrm{d}u=\int_0^x\left(u+\frac{1}{2}\right)\mathrm{d}u=\frac{1}{2}\left(x^2+x\right),\quad 0<x\leqslant 1$$
$$F_2(y)=\int_{-\infty}^{y}p_2(v)\mathrm{d}v=\int_0^y\left(v+\frac{1}{2}\right)\mathrm{d}v=\frac{1}{2}\left(y^2+y\right),\quad 0<y\leqslant 1$$

2) 容易验证 $q(x,y)$ 为某个二维随机向量 (X,Y) 的联合密度函数，且不难求出 X 和 Y 的边缘密度函数如式 (6.2) 所示

注 1) 表明边缘密度函数可由联合密度函数唯一确定，但是其逆命题不真。例如 1)2) 随机向量 (X,Y) 的联合密度函数并不相同，但 X 和 Y 的边缘密度函数相同，这就表明：联合密度函数不能由边缘密度函数所确定。

例题 6.9 （边缘密度函数对联合密度函数不充分性）

设 $\{\Omega,\mathcal{A},P\}$ 为区间 $(0;1)$ 上的几何概率空间，令

$$(\forall\omega\in(0;1)):X(\omega)=Y(\omega)=\omega$$

则 X 与 Y 都服从均匀分布 $U(0;1)$，即为连续随机变量，但二维随机向量 (X,Y) 没有联合密度函数，取值集合仅为平面上的一条线段

注 若 $(X_1,\cdots,X_n)$ 为 n 维连续随机向量，则其中任意 k 个随机变量都形成 k 维连续随机向量，且其边缘密度可以由 $(X_1,\cdots,X_n)$ 的联合密度唯一确定。但反过来，即使随机变量 $X_1,\cdots,X_n$ 为定义在同一个概率空间上的 n 个连续随机变量，随机向量 $(X_1,\cdots,X_n)$ 也未必为连续随机向量

6.1.5 条件分布与条件密度 Conditional Distribution and Conditional Density

定理 6.4 (条件密度函数存在充分条件)

若二维连续随机向量 (X,Y) 的联合密度函数为 $p(x,y)$，则当 $p_2(y)>0$ 时，函数

$$p_1(x\mid y)=\frac{p(x,y)}{p_2(y)},\quad x\in\mathbb{R}$$

即为随机变量 X 在条件 $Y=y$ 下的条件密度函数；而若 $p_1(x)>0$，则函数

$$p_2(y\mid x)=\frac{p(x,y)}{p_1(x)},\quad x\in\mathbb{R}$$

即为 Y 在条件 $X=x$ 下的条件密度函数

证明 设二维连续随机向量 (X,Y) 的联合密度函数为 $p(x,y)$，若对 $a<b$，有 $P(a\leqslant Y\leqslant b)>0$，则可计算出 X 在条件 $a\leqslant Y\leqslant b$ 下的条件分布：

$$P(X\leqslant x\mid a\leqslant Y\leqslant b)=\frac{P(X\leqslant x,a\leqslant Y\leqslant b)}{P(a\leqslant Y\leqslant b)}=\frac{\int\limits_{-\infty}^{x}\mathrm{d}u\int\limits_{a}^{b}p(u,v)\mathrm{d}v}{\int\limits_{a}^{b}p_2(v)\mathrm{d}v}=\int\limits_{-\infty}^{x}\frac{\int\limits_{a}^{b}p(u,v)\mathrm{d}v}{\int\limits_{a}^{b}p_2(v)\mathrm{d}v}\mathrm{d}u$$

若记

$$p_1(x\mid[a;b])=\frac{\int\limits_{a}^{b}p(x,v)\mathrm{d}v}{\int\limits_{a}^{b}p_2(v)\mathrm{d}v},\quad x\in\mathbb{R}$$

则式 $P(X\leqslant x\mid a\leqslant Y\leqslant b)$ 与表明 $p_1(x\mid[a;b])$ 即为 X 在条件 $a\leqslant Y\leqslant b$ 下的条件密度函数

下考虑区间 $[a;b]$ 的长度趋于 0 的情况。设 $a=y,b=y+\Delta y$，其中 $\Delta y>0$，并设 $p_2(y)>0$，考虑 $\Delta y\to 0$，类似即得

$$P(X<x\mid y\leqslant Y\leqslant y+\Delta y)=\frac{\int\limits_{-\infty}^{x}\left\{\int\limits_{y}^{y+\Delta y}p(u,v)\mathrm{d}v\right\}\mathrm{d}u}{\int\limits_{y}^{y+\Delta y}p_2(v)\mathrm{d}v}=\frac{\int\limits_{-\infty}^{x}\left\{\frac{1}{\Delta y}\int\limits_{y}^{y+\Delta y}p(u,v)\mathrm{d}v\right\}\mathrm{d}u}{\frac{1}{\Delta y}\int\limits_{y}^{y+\Delta y}p_2(v)\mathrm{d}v}$$

则有

$$\lim_{\Delta y\to 0}\frac{1}{\Delta y}\int_{y}^{y+\Delta y}p_2(v)\mathrm{d}v=p_2(y)$$

而若对任意 $u\in\mathbb{R}$ 都有极限 $\lim\limits_{\Delta y\to 0}\frac{1}{\Delta y}\int\limits_{y}^{y+\Delta y}p(u,v)\mathrm{d}v$ 存在，则有

$$\lim_{\Delta y\to 0}\frac{1}{\Delta y}\int_{y}^{y+\Delta y}p(u,v)\mathrm{d}v=p(u,y)$$

令 $\Delta y\to 0$，则有

$$P(X\leqslant x\mid Y=y)=\int_{-\infty}^{x}\frac{p(u,y)}{p_2(y)}\mathrm{d}u,\quad x\in\mathbb{R}$$

这时有 $p_1(x\mid y)$ 为一个密度函数，并且对任意 $x\in\mathbb{R},p_1(u\mid y)$ 在区间 $(-\infty;x]$ 上对 u 的积分即为条件分布

$P(X \leqslant x \mid Y = y)$

例题 6.10 （条件密度函数）

设 $\lambda > 0$，随机变量 X 的密度函数为

$$p_1(x) = \lambda^2 x \mathrm{e}^{-\lambda x}, \quad x > 0$$

而随机变量 Y 服从区间 $(0; X)$ 上的均匀分布，试求：

1) 随机向量 (X, Y) 的联合密度；

2) 随机变量 Y 的密度函数

解 在给定 $X = x$ 的条件下，Y 具有条件密度 $p_2(y \mid x) = \dfrac{1}{x}, 0 < y < x$，则 (X, Y) 的联合密度函数为

$$p(x, y) = p_1(x) p_2(y \mid x) = \lambda^2 \mathrm{e}^{-\lambda x}, \quad , 0 < y < x$$

而 Y 的密度函数为

$$p_2(y) = \int_{-\infty}^{\infty} p(u, y) \mathrm{d}u = \int_{y}^{\infty} \lambda^2 \mathrm{e}^{-\lambda u} \mathrm{d}u = \lambda \mathrm{e}^{-\lambda y}, \quad y > 0$$

即 Y 服从参数为 λ 的指数分布

6.1.6 独立性

Independence

定义 6.10 (全体独立)

设 n 维随机向量 $(\xi_1, \ldots, \xi_n)$ 定义在概率空间 $(\Omega, \mathcal{B}, P)$ 上，若对于任意 Borel 子集 $B_1, \ldots, B_n \in \mathcal{B}$ 都有

$$P(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = P(\xi_1 \in B_1) \cdot \ldots \cdot P(\xi_n \in B_n) \tag{6.3}$$

则称随机变量 $\xi_1, \ldots, \xi_n$（关于概率 P）（全体）独立 (independent/независимы) ♣

定理 6.5 (独立性判别法/критерий независимости)

设随机向量 $\xi = (\xi_1, \ldots, \xi_n)$，其分量 $\xi_1, \ldots, \xi_n$ 独立的充要条件为

$$F_\xi(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n) \tag{6.4}$$

对于离散随机向量，该条件等价于

$$P\left(\xi_1 = x_{i_1}^{(1)}, \ldots, \xi_n = x_{i_n}^{(n)}\right) = P\left(\xi_1 = x_{i_1}^{(1)}\right) \cdot \ldots \cdot P\left(\xi_n = x_{i_n}^{(n)}\right) \tag{6.5}$$

对于连续随机向量，该条件等价于

$$\rho_\xi(x_1, \ldots, x_n) = \rho_{\xi_1}(x_1) \cdot \ldots \cdot \rho_{\xi_n}(x_n) \tag{6.6}$$

♡

证明 离散随机向量情况显然，下设 $(X_1, \cdots, X_n)$ 为 n 维连续型随机向量。仅以证明 $n = 2$ 的情形，证明不难推广到 n 维场合

充分性：若随机变量 X 和 Y 的联合密度函数等于其两个边缘密度函数的乘积，即

$$\rho(x, y) = \rho_1(x) \rho_2(y)$$

则显然对一切 $(x, y) \in \mathbb{R}^2$ 都有

$$\begin{aligned} F(x, y) &= \int_{-\infty}^{x} \mathrm{d}u \int_{-\infty}^{y} p(u, v) \mathrm{d}v = \int_{-\infty}^{x} \mathrm{d}u \int_{-\infty}^{y} \rho_1(u) \rho_2(v) \mathrm{d}v \\ &= \int_{-\infty}^{x} \rho_1(u) \mathrm{d}u \int_{-\infty}^{y} \rho_2(v) \mathrm{d}v = F_1(x) F_2(y) \end{aligned}$$

则由联合分布函数与分布函数定义推出式 (6.3)，即 X 与 Y 相互独立

必要性：若 (X,Y) 为连续随机向量且有 $F(x,y)=F_1(x)F_2(y)$，则对任意 $(x,y)\in\mathbb{R}^2$ 有

$$\begin{aligned}\int_{-\infty}^{x}\mathrm{d}u\int_{-\infty}^{y}\rho(u,v)\mathrm{d}v=F(x,y)=F_1(x)F_2(y)=\int_{-\infty}^{x}\rho_1(u)\mathrm{d}u\int_{-\infty}^{y}\rho_2(v)\mathrm{d}v\\=\int_{-\infty}^{x}\mathrm{d}u\int_{-\infty}^{y}\rho_1(u)\rho_2(v)\mathrm{d}v\end{aligned}$$

比较两端，定理即证

例题 6.11 （几何分布/独立）

设相互独立的随机变量 ξ_1 与 ξ_2 的概率分布分别为参数为 p_1 和 p_2 的几何分布，即

$$P\left(\xi_1=m\right)=p_1\left(1-p_1\right)^{m-1}\quad 1,2,\ldots,m,\ldots$$
$$P\left(\xi_2=m\right)=p_2\left(1-p_1\right)^{m-1}\quad 1,2,\ldots,m,\ldots$$

试证明随机变量 $\eta=\min\left(\xi_1,\xi_2\right)$ 的概率分布为参数为 p_1p_2 的几何分布

解 考虑事件 $(\xi_1>m)$ 的概率，事件当且仅当前 m 次试验失败时发生。由试验独立性，概率即为 $(1-p_1)^m$，由此即得

$$\begin{aligned}P(\eta=m)&=P\left(\xi_1=m,\xi_2=m\right)+P\left(\xi_1=m,\xi_2>m\right)+P\left(\xi_1>m,\xi_2=m\right)\\&=P\left(\xi_1=m\right)P\left(\xi_2=m\right)+P\left(\xi_1=m\right)P\left(\xi_2>m\right)+P\left(\xi_1>m\right)P\left(\xi_2=m\right)\\&=p_1\left(1-p_1\right)^{m-1}p_2\left(1-p_2\right)^{m-1}+p_1\left(1-p_1\right)^{m-1}p_2^m+p_1^mp_2\left(1-p_2\right)^{m-1}\\&=p_1p_2\left(1-p_1p_2\right)^{m-1}\end{aligned}$$

这即为参数为 p_1p_2 的几何分布

例题 6.12 （联合密度函数/独立）

设随机向量 $\xi=(\xi_1,\xi_2)$ 的联合密度函数为

$$\rho_\xi\left(x_1,x_2\right)=\begin{cases}c\cdot\left|x_1\right|\left|x_2\right|, & \left|x_1\right|+\left|x_2\right|\leqslant 1\\0, & \left|x_1\right|+\left|x_2\right|>1\end{cases}$$

1) 求常数 c
2) 求每个坐标的密度函数
3) 求 $P\left(\xi_1+\xi_2>0.5\right)$
4) 检验坐标的独立性

解 1) 由密度函数定义即得随机向量值取在顶点为 $(0,1),(1,0),(0,-1),(-1,0)$ 的正方形 K 上，即有 $\int\limits_K\rho_\xi dx_1dx_2=1$ 则由规范性即得

$$\begin{aligned}\int_K\rho_\xi dx_1dx_2=4\cdot\int_0^1\left(\int_0^{1-x_2}c\cdot x_1\cdot x_2dx_1\right)dx_2=4\cdot c\cdot\int_0^1x_2\cdot\frac{1}{2}\cdot\left(1-x_2\right)^2dx_2\\=2c\int_0^1\left(x_2-2x_2^2+x_2^3\right)dx_2=2c\left(\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right)=\frac{c}{6}=1.\end{aligned}$$

则有 $c=6$

2) 设 $x_2>0$，对坐标积分即得

$$\rho_\xi\left(x_2\right)=\int_{-\infty}^{\infty}\rho_\xi\left(x_1,x_2\right)dx_1$$

当 $x_2>1$ 时即得 0。若 $0<x_2<1$ 则有

$$\rho_{\xi_2}\left(x_2\right)=\int_{-(1-x_2)}^{(1-x_2)}6\cdot x_2\cdot\left|x_1\right|dx_1=2\cdot\int_0^{(1-x_2)}x_2\cdot x_1dx_1=x_2\left(1-x_2\right)^2$$

而由对称性对于 $-1<x_2<0$ 即得 $\rho_{\xi_2}\left(x_2\right)=\left|x_2\right|\left(1-\left|x_2\right|\right)^2$ 。由对称性对 x_1 得到同样的结果

3) 由定义即得

$$P\left(\xi_1+\xi_2>0.5\right)=P(\xi\in B)$$

将区域 B 划分为 B_1, B_2, B_3, B_4 并计算出现在每个区域的概率

$$P(\xi \in B_1) = P(\xi \in B_4) = \int_{1/2}^{1} \left(\int_{1/2-x_2}^{x_2-1} 6x_2 |x_1| \, dx_1 \right) dx_2 = \frac{1}{32}$$

$$P(\xi \in B_2) = \int_{1/2}^{1} \left(\int_{1/2-x_1}^{1-x_1} 6x_1 x_2 dx_2 \right) dx_1 = \frac{5}{32}$$

$$P(\xi \in B_3) = \int_{1/2}^{1} \left(\int_{0}^{1-x_1} 6x_1 x_2 dx_2 \right) dx_1 = \frac{11}{64}$$

由此则有 $P(\xi \in B) = \dfrac{23}{64}$

4) 选择点 $x = (\frac{3}{4}, \frac{3}{4})$，在该点有

$$0 = \rho_\xi \left(\frac{3}{4}, \frac{3}{4} \right) \neq \rho_{\xi_1} \left(\frac{3}{4} \right) \cdot \rho_{\xi_2} \left(\frac{3}{4} \right) > 0$$

则两个坐标不互相独立

命题 6.2

若 $X_1, \cdots, X_n$ 为 n 个全体独立的随机变量，A_1 和 A_2 为集合 $\{1, \cdots, n\}$ 的两个互不相交的非空子集，其中分别含有 k_1 和 k_2 个元素，则对任意两个分别为 k_1 元和 k_2 元的 Borel 可测函数 f 和 g，$Y_1 = f(X_i, i \in A_1)$ 和 $Y_2 = g(X_j, j \in A_2)$ 为两个独立的随机变量

♠

定义 6.11 (随机变量全体独立)

设组 $\xi_1, \cdots, \xi_r$ 为在 $\mathbb{R}^1$ 中（有限）集合 X 上取值的随机变量。记 $\mathcal{B}$ 为 X 中所有 σ 代数。若对于任意 $x_1, \cdots, x_r \in X$,

$$P\{\xi_1 = x_1, \cdots, \xi_r = x_r\} = P\{\xi_1 = x_1\} \cdots P\{\xi_r = x_r\}$$

或等价地：对于任意 $B_1, \cdots, B_r \in \mathcal{B}$,

$$P\{\xi_1 \in B_1, \cdots, \xi_r \in B_r\} = P\{\xi_1 \in B_1\} \cdots P\{\xi_r \in B_r\}$$

则称随机变量 $\xi_1, \cdots, \xi_r$ 全体独立

♣

随机变量 $\xi_1, \cdots, \xi_r$ 的函数的随机变量 $f(\xi_1, \cdots, \xi_r)$ 。考虑求随机变量之和 $\xi + \eta$ 的分布的情形。

6.1.7 卷积公式

Convolution formula

定义 6.12 (离散随机变量和)

若离散随机变量 ξ 的值域为 $X = \{x_1, \cdots, x_k\}$，随机变量 η 的值域为 $Y = \{y_1, \cdots, y_l\}$，则 $\zeta = \xi + \eta$ 的值域为 $Z = \{z : z = x_i + y_j, i = 1, \cdots, k, j = 1, \cdots, l\}$，且显然

$$P_\zeta(z) = P\{\zeta = z\} = P\{\xi + \eta = z\} = \sum_{\{(i,j): x_i + y_j = z\}} P\{\xi = x_i, \eta = y_j\}$$

称 $\zeta = \xi + \eta$ 为离散随机变量 ξ 与 η 的和 (sum)

♣

定理 6.6 (离散分布卷积公式)

设离散随机变量 ξ_1 与 ξ_2 相互独立，记 $p(x, y) = P(\xi_1 = x, \xi_2 = y)$ 。设 $z = g(x, y) = x + y$，这时

$\eta=\xi_1+\xi_2$ 为离散随机变量且值为 $\eta=z$，其中 z 为 $\xi_1+\xi_2$ 可能的值，则有公式

$$P(\eta=z)=\sum_{(x,y):x+y=z} p(x,y)=\sum_x p(x,z-x)$$
$$P(\eta=z)=\sum_{(x,y):x+y=z} p(x,y)=\sum_x p(z-y,y)$$

证明 若 ξ_1 与 ξ_2 相互独立，则有

$$p(x,y)=P(\xi_1=x,\xi_2=y)=P(\xi_1=x)\cdot P(\xi_2=y)=P_1(x)P_2(y)$$

则有

$$P(\xi_1+\xi_2=z)=\sum_x P_1(x)P_2(z-x)$$

交换 ξ_1 和 ξ_2 的位置即证

注 经常称

$$P(\eta=z)=\sum_{(x,y):x+y=z} p(x,y)=\sum_x p(x,z-x)$$
$$P(\eta=z)=\sum_{(x,y):x+y=z} p(x,y)=\sum_x p(z-y,y)$$

为**离散分布卷积公式** (формула свертки для дискретных распредений)

例题 6.13 (二项分布/离散分布卷积公式)

设随机变量 ξ_1 的概率分布为参数为 (n_1,p) 的二项分布，随机变量 ξ_2 的概率分布为参数为 (n_2,p) 的二项分布，且二随机变量独立。试证明随机变量 $\eta=\xi_1+\xi_2$ 的概率分布为参数为 (n_1+n_2,p) 的二项分布

解 对于随机变量 ξ_1 有

$$P_1(m)=C_{n_1}^m p^m(1-p)^{n_1-m}, m=0,1,\ldots n_1$$

对于随机变量 ξ_2 有

$$P_2(m)=C_{n_2}^m p^m(1-p)^{n_2-m}, m=0,1,\ldots n_2$$

利用卷积公式即得 $k=0,1,\ldots n_1+n_2$

$$\begin{aligned}P(\eta=k)&=\sum_{m=0}^{n_1} P_1(m)P_2(k-m)=\sum_{m=0}^{k} P_1(m)P_2(k-m)\\&=\sum_{m=0}^{k} C_{n_1}^m p^m(1-p)^{n_1-m}\cdot C_{n_2}^{k-m}p^{k-m}(1-p)^{n_2-k+m}\\&=p^k(1-p)^{n_1+n_2-k}\sum_{m=0}^{k} C_{n_1}^m C_{n_2}^{k-m}=C_{n_1+n_2}^k p^k(1-p)^{n_1+n_2-k}\end{aligned}$$

即有随机变量 $\eta=\xi_1+\xi_2$ 为参数为 (n_1+n_2,p) 的二项分布

例题 6.14 (Poisson 分布/二维离散分布卷积公式)

设两个独立的随机变量 ξ_1 与 ξ_2 的概率分布为参数为 λ_1,λ_2 的 Poisson 分布。试证明随机变量 $\eta=\xi_1+\xi_2$ 的概率分布为参数为 $\lambda_1+\lambda_2$ 的 Poisson 分布

解 利用二维离散分布卷积公式即得

$$\begin{aligned}P(\eta=n)&=\sum_{k=0}^{n} P(\xi_1=k)\cdot P(\xi_2=n-k)=\sum_{k=0}^{n}\frac{\lambda_1^k}{k!}e^{-\lambda_1}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}\\&=e^{-(\lambda_1+\lambda_2)}\frac{1}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}=\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}\end{aligned}$$

即为参数为 $\lambda_1+\lambda_2$ 的 Poisson 分布

定理 6.7 (连续随机向量变换密度函数)
设 $(X_1,\cdots,X_n)$ 为 n 维连续随机向量，其联合密度函数为 $p(x_1,\cdots,x_n)$，若存在 n 个 Borel 可测函数

$$y_j = f_j(x_1,\cdots,x_n),\quad j=1,\cdots,n,$$

满足

$$Y_j = f_j(X_1,\cdots,X_n),\quad j=1,\cdots,n,$$

并且对于 $(Y_1,\cdots,Y_n)$ 的每一组可能值 $(y_1,\cdots,y_n)$，方程组

$$\begin{cases} y_1 = f_1(x_1,\cdots,x_n) \\ \quad\cdots\cdots \\ y_n = f_n(x_1,\cdots,x_n) \end{cases}$$

都有唯一解

$$\begin{cases} x_1 = h_1(y_1,\cdots,y_n) \\ \quad\cdots\cdots \\ x_n = h_n(y_1,\cdots,y_n) \end{cases} \tag{6.7}$$

其中每个 $h_j(y_1,\cdots,y_n)$ 都有一阶连续偏导数，则随机向量 $(Y_1,\cdots,Y_n)$ 为连续随机向量，具有联合密度函数

$$\rho(y_1,\cdots,y_n) = p(h_1(y_1,\cdots,y_n),\cdots,h_n(y_1,\cdots,y_n))|J|,\quad (y_1,\cdots,y_n)\in D \tag{6.8}$$

其中 D 为随机向量 $(Y_1,\cdots,Y_n)$ 的所有可能值的集合，J 为变换的 Jacobi 行列式，即

$$J = \frac{\partial(x_1,\cdots,x_n)}{\partial(y_1,\cdots,y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

♡

证明 由于随机向量 $(Y_1,\cdots,Y_n)$ 的联合分布为

$$\begin{aligned} F(y_1,\cdots,y_n) &= P(Y_1\leqslant y_1,\cdots,Y_n\leqslant y_n) \\ &= \int_{f_n(x_1,\cdots,x_n)\leqslant y_n}\cdots\int_{f_1(x_1,\cdots,x_n)\leqslant y_1} p(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &= \int_{-\infty}^{y_n}\cdots\int_{-\infty}^{y_1} p(h_1(u_1,\cdots,u_n),\cdots,h_n(u_1,\cdots,u_n))|J|\mathrm{d}u_1\cdots\mathrm{d}u_n \end{aligned}$$

最后一步得积分变量替换 (6.7)，则随机向量 $(Y_1,\cdots,Y_n)$ 的联合密度函数即为式 (6.8)

注 式 (6.8) 在变换 $\eta=g(\xi)$ 下经常也简记为

$$\rho_\eta(y) = \rho_\xi\left(g^{-1}(y)\right)\left|J\left(g^{-1}(y)\right)\right|^{-1}$$

其中 J 为 $y=g(x)$ 的 Jacobi 行列式

定理 6.8 (连续分布卷积公式)
设随机向量 $\xi=(\xi_1,\xi_2)$ 的密度函数为 $\rho_\xi(x_1,x_2)$，则随机变量 $\eta_1=\xi_1+\xi_2$ 的密度函数可表示为

$$\rho_{\xi_1+\xi_2}(y) = \int_{-\infty}^{\infty}\rho_{\xi_1}(y-x)\rho_{\xi_2}(x)dx$$

♡

证明

情况一：随机变量相互独立

设 $y=Ax$，其中 A 为 n 维非退化矩阵，则由定理 (6.7) 有

$$\rho_\eta(y) = \rho_\xi\left(A^{-1}y\right)\cdot|\det A|^{-1}$$

考虑另一个随机变量 $\eta_2=\xi_2$。这时有线性映射 $y_1=x_1+x_2, y_2=x_2$，该映射的矩阵 A 为

$$A=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$$

其 $\det A=1$，而逆矩阵 A^{-1} 等于

$$A^{-1}=\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix}$$

则由 $\rho_\eta\left(y_1,y_2\right)=\rho_\xi\left(A^{-1}y\right)\cdot|\det A|^{-1}$ ，积分即得

$$\rho_{\eta_1}\left(y_1\right)=\int_{-\infty}^{\infty}\rho_\xi\left(y_1-y_2,y_2\right)dy_2$$

由 ξ_1 与 ξ_2 相互独立即得 $\rho_\xi\left(x_1,x_2\right)=\rho_{\xi_1}\left(x_1\right)\rho_{\xi_2}\left(x_2\right)$。将 y_1 替换为 y 即得

$$\rho_{\xi_1+\xi_2}(y)=\int_{-\infty}^{\infty}\rho_{\xi_1}(y-x)\rho_{\xi_2}(x)dx$$

即为随机变量 $\eta_1=\xi_1+\xi_2$ 的密度函数

情况二：一般情形

计算随机变量 $\eta=\xi_1+\xi_2$ 的分布函数

$$F_\eta(y)=P(\eta<y)=P\left(\xi_1+\xi_2<y\right)$$

即需找到随机向量 $\xi=(\xi_1,\xi_2)$ 在集合 $B(y)=\{(x_1,x_2):x_1+x_2<y\}\subset R^2$ 上概率。这时有

$$F_\eta(y)=P\left(\left(\xi_1,\xi_2\right)\in B(y)\right)=\iint_{B(y)}\rho_\xi\left(x_1,x_2\right)dx_1dx_2$$

改写为二重积分即得

$$\begin{aligned}F_\eta(y)&=\int_{-\infty}^{\infty}\int_{-\infty}^{y-x_2}\rho_\xi\left(x_1,x_2\right)dx_1dx_2=\left|\begin{array}{c}x=x_1+x_2\\ z=x_2\end{array}\right|\\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{y}\rho_\xi(x-z,z)dxdz=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty}\rho_\xi(x-z,z)dz\right]dx\end{aligned}$$

对 y 微分即得

$$\rho_\eta(y)=\frac{d}{dy}F_\eta(y)=\int_{-\infty}^{\infty}\rho_\xi(y-z,z)dz$$

由此得到相同的结果

注 经常称

$$\rho_{\xi_1+\xi_2}(y)=\int_{-\infty}^{+\infty}\rho_{\xi_1}\left(x_1\right)\rho_{\xi_2}\left(y-x_1\right)\mathrm{d}x_1$$

$$\rho_{\xi_1+\xi_2}(y)=\int_{-\infty}^{+\infty}\rho_{\xi_1}(y-x_2)\rho_{\xi_2}(x_2)\mathrm{d}x_2$$

为**连续分布卷积公式** (формула свертки для непрерывных распределений) 或 **密度卷积公式** (формула свертки для плотностей)

例题 6.15 （均匀分布独立和/三角分布/二维连续分布卷积公式）

设随机变量 ξ_1 与 ξ_2 相互独立，对应密度函数为

$$\rho_1(x)=\rho_2(x)=1,x\in[0;1]$$

试求随机变量 $\eta=\xi_1+\xi_2$ 的密度函数

解 显然其密度函数在 [0;2] 上取值，利用连续分布卷积公式即得

$$\rho_\eta(y)=\int_{-\infty}^{\infty}\rho_1(x)\rho_2(y-x)dx$$

当 $0<y\leqslant 1$ 时有

$$\rho_\eta(y)=\int_0^y 1\cdot 1dx=y$$

当 $1<y\leqslant 2$ 时有

$$\rho_\eta(y)=\int_{y-1}^{1}1\cdot 1dx=2-y$$

综上即得随机变量 $\eta=\xi_1+\xi_2$ 的密度函数

注 经常也称这种分布为**三角分布** (треугольное распределение)

例题 6.16 （指数分布独立和/Laplace 分布/二维连续分布卷积公式）

设随机变量 ξ_1 与 ξ_2 相互独立且概率分布为指数分布，即对应密度函数为

$$\rho_{\xi_1}(x)=\rho_{\xi_2}(x)=\lambda e^{-\lambda x},x>0$$

试求随机变量 $\eta=\xi_1-\xi_2$ 的密度函数

解 若记 $\eta=\xi_1-\xi_2=\xi_1+(-\xi_2)=\xi_1+(\xi_2')$ 则有

$$\rho_{\xi_2'}(x)=\lambda e^{\lambda x},x<0$$

显然分布关于零对称，其可能的值的集合为 $(-\infty,+\infty)$。取 $y>0$ 并由连续分布卷积公式即得

$$\rho_\eta(y)=\int_{-\infty}^{\infty}\rho_{\xi_1}(x)\rho_{\xi_2'}(y-x)dx=\int_{y}^{\infty}\rho_{\xi_1}(x)\rho_{\xi_2'}(y-x)dx=\frac{\lambda}{2}e^{-\lambda y}$$

即为随机变量 $\eta=\xi_1-\xi_2$ 的密度函数

注 经常称该分布为 Laplace **分布** (распределение Лапласа)

例题 6.17 （Γ 分布/密度卷积公式）

设连续随机变量 ξ_1 与 ξ_2 的概率分布分别为参数为 (α_1,β) 与 (α_2,β) 的 Γ 分布。则两个随机变量的和的概率分布为参数为 $(\alpha_1+\alpha_2,\beta)$ 的 Γ 分布

解 由定义即得

$$\rho_{\xi_1}(x_1)=\frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)}x_1^{\alpha_1-1}e^{-\beta x_1},\quad \rho_{\xi_2}(x_2)=\frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)}x_2^{\alpha_2-1}e^{-\beta x_2},\quad x_1,x_2>0$$

由密度卷积公式有

$$\begin{aligned}\rho_\eta(y)&=\int_{-\infty}^{+\infty}\rho_{\xi_1}(x_1)\cdot\rho_{\xi_2}(y-x_1)\,dx_1\\&=\int_0^y\frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)}x_1^{\alpha_1-1}e^{-\beta x_1}\cdot\frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)}(y-x_1)^{\alpha_2-1}e^{-\beta(y-x_1}dx_1\\&=\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-\beta y}\int_0^y x_1^{\alpha_1-1}(y-x_1)^{\alpha_2-1}dx_1\end{aligned}$$

再换元 $u=\dfrac{x_1}{y}$ 即得

$$\begin{aligned}\rho_\eta(y)&=\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-\beta y}y^{\alpha_1+\alpha_2-1}\int_0^1u^{\alpha_1-1}(1-u)^{\alpha_2-1}dx_1\\&=\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-\beta y}y^{\alpha_1+\alpha_2-1}\cdot B(\alpha_1,\alpha_2)=\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-\beta y}y^{\alpha_1+\alpha_2-1}\cdot\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\\&=\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}y^{\alpha_1+\alpha_2-1}e^{-\beta y},\quad y>0\end{aligned}$$

即为参数为 $(\alpha_1+\alpha_2,\beta)$ 的 Γ 分布

定义 6.13 (二维连续分布)

若 $F(x,y)$ 为一个二维分布函数，并且存在一个 Lebesgue 可积非负函数 $p(x,y)$ 满足

$$(\forall(x,y)\in\mathbb{R}^2):F(x,y)=\int_{-\infty}^{x}\mathrm{d}u\int_{-\infty}^{y}p(u,v)\mathrm{d}v$$

则称 $F(x,y)$ 为一个二维连续型分布简称二维连续分布，称 $p(x,y)$ 为 $F(x,y)$ 的密度函数。

若一个二维随机向量 (X_1,X_2) 的概率分布为连续分布 $F(x,y)$，则称 (X_1,X_2) 为二维连续型随机向量简称二维连续向量，并称 $F(x,y)$ 的密度函数 $p(x,y)$ 为 (X_1,X_2) 的密度函数

♣

由式 (4.1.5) 可知：一个二元函数 $p(x,y)$ 为密度函数，当且仅当

$$p(x,y) \geqslant 0, \quad \forall (x,y) \in \mathbb{R}^2, \quad \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} p(u,v)\mathrm{d}v = 1$$

6.2 数值特征 Numerical Characteristics

6.2.1 协方差与相关系数 Covariance and Correlation Coefficient

定义 6.14 (随机向量矩)

设随机向量 $\xi = (\xi_1, \cdots, \xi_d)$，$k_1 + \ldots + k_d = k$，则称实数

$$E\left[(\xi_1 - a_1)^{k_1} \cdot \ldots \cdot (\xi_d - a_d)^{k_d}\right]$$

为随机向量 ξ 相对于点 $a \in R^d$ 的 k 阶矩 (moment/момент)。特别地，若 $a_1 = \cdots = a_d = 0$，则称其为 k 阶原点矩 (начальный момент)；若 $a_1 = E(\xi_1), \cdots, a_d = E(\xi_d)$，则称其为 k 阶中心矩 (центральный момент) ♣

定义 6.15 (协方差)

设随机向量 $\xi = (\xi_1, \cdots, \xi_d)$ 的全部二阶原点矩表示为

$$\sigma_{11} = Var(\xi_1) = E\left[(\xi_1 - E(\xi_1))^2\right]$$

$$\cdots$$

$$\sigma_{dd} = Var(\xi_d) = E\left[(\xi_d - E(\xi_d))^2\right]$$

$$\sigma_{ij}(\xi_i, \xi_j) = E\left[(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))\right]$$

也称 $E(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))$ 为随机变量 ξ_i 和 ξ_j 的协方差 (covariance/ковариация)，记为 $\mathrm{cov}(\xi_1, \xi_2)$，并称矩阵 $(\sigma_{ij}) \in R^{d\times d}$ 为随机向量 ξ 的协方差矩阵 (covariance matrix/матрица ковариаций) ♣

定义 6.16 (Pearson 相关系数)

（Pearson[a]相关系数）若 $Var(\xi) \geqslant 0, Var(\eta) \geqslant 0$，则称

$$\rho(\xi,\eta) = \frac{\mathrm{cov}(\xi,\eta)}{\sqrt{Var(\xi)\times Var(\eta)}} = \frac{\mathrm{cov}(\xi,\eta)}{\sigma(\xi)\sigma(\eta)}$$

为随机变量 ξ 和 η 的 Pearson 相关系数 (Pearson correlation coefficient)。由于 Pearson 相关系数为最常见的相关系数，因此不引起歧义时，有时也直接简称为相关系数 (correlation coefficient/коэффициент корреляции)

[a]卡尔 · 皮尔逊 (Karl Pearson,1857.3.27-1936.4.27) 英国数学家，生物统计学家。Pearson 为 20 世纪科学革命和哲学革命的先驱，对生物统计学、气象学、社会达尔文主义理论和优生学做出了重大贡献。Pearson 被公认为旧理学派和描述统计学派的代表人物，现代数理统计学的创立者，“批判学派”的代表人物之一 ♣

定理 6.9 (Pearson 相关系数性质)

设 ξ_1 与 ξ_2 为非退化随机变量，则有下列命题成立

1)（Pearson 相关系数基本值域）$|\rho(\xi_1,\xi_2)| \leqslant 1$

2)（线性无关充分条件）若 ξ_1 与 ξ_2 不相关，则有 $\rho(\xi_1,\xi_2) = 0$

3) (线性相关必要条件) 设 $\eta_1 = a_1\xi_1 + b_1, \eta_2 = a_2\xi_2 + b_2, a_1 \cdot a_2 \neq 0$, 则有

$$\rho\left(\eta_1, \eta_2\right) = \rho\left(\xi_1, \xi_2\right) \cdot \operatorname{sgn}\left(a_1 a_2\right)$$

4) (线性相关充分条件) 若 $\rho\left(\xi_1, \xi_2\right) = 1$, 则存在 $a > 0, b \in R^1$ 满足 $\xi_1 = a\xi_2 + b$; 若 $\rho\left(\xi_1, \xi_2\right) = -1$, 则存在 $a < 0, b \in R^1$ 满足 $\xi_1 = a\xi_2 + b$

6.2.2 条件期望

例题 6.18 (条件期望)

设随机变量 X 的概率分布为参数为 $\lambda > 0$ 的指数分布，试对任意 $x > 0$ 求条件期望 $E(X - x \mid X > x)$

解 由条件期望公式有

$$E(X - x \mid X > x) = \frac{\lambda \int_x^{\infty} (u - x) \mathrm{e}^{-\lambda u} \mathrm{~d}u}{\lambda \int_x^{\infty} \mathrm{e}^{-\lambda u} \mathrm{~d}u} = \frac{\lambda \mathrm{e}^{-\lambda x} \int_0^{\infty} y \mathrm{e}^{-\lambda y} \mathrm{~d}y}{\lambda \int_x^{\infty} \mathrm{e}^{-\lambda u} \mathrm{~d}u} = \frac{1}{\lambda}$$

即为所求条件期望

注 前文提到过，指数分布具有无记忆性，并且介绍过，当参数为 $\lambda > 0$ 时，其期望等于 $\frac{1}{\lambda}$。该例的结果再一次解释了指数分布的无记忆性。若把 X 视为寿命，该例表明，无论已经活了多久，剩余寿命的期望都跟初出生时的平均寿命相等

第 7 章 证明附录

7.1 单调类定理与概率扩张定理

注 （单调集列序列极限符号约定）

若集族 $A_n \uparrow$，则记 $\lim \uparrow A_n = \bigcup\limits_{n=1}^{\infty} A_n$；若集族 $A_n \downarrow$，则记 $\lim \downarrow A_n = \bigcap\limits_{n=1}^{\infty} A_n$

定义 7.1 (单调类)

$\mathcal{C}$ 为 Ω 的一个子集族，若 $\mathcal{C}$ 在 $\lim \uparrow$ 和 $\lim \downarrow$ 运算下封闭，则称 $\mathcal{C}$ 为 Ω 的一个单调类 ♣

定理 7.1 (事件代数为 σ 代数充要条件)

Ω 上事件代数 $\mathcal{A}$ 为 σ 代数的充要条件为 $\mathcal{A}$ 为单调类 ♡

证明 必要性显然。充分性：设事件代数 $\mathcal{A}$ 为单调类，事件族 $\{A_n | n \in \mathbb{N}^*\}$，记 $B_n = \bigcup\limits_{i=1}^{n} A_i \uparrow$，则有

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$$

则 $\mathcal{A}$ 在可数并运算下封闭，其他条件由 $\mathcal{A}$ 为事件代数满足，则 $\mathcal{A}$ 为 σ 代数

定义 7.2 (生成单调类)

设 $\mathcal{C}$ 为 Ω 子集族，称包含 $\mathcal{C}$ 的最小单调类为由 $\mathcal{C}$ 生成的单调类，记为 $\mathcal{M}(\mathcal{C})$. ♣

定理 7.2 (事件代数上生成单调类与生成 σ 代数等价性)

若 $\mathcal{A}$ 为 Ω 上事件代数，则有 $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$ ♡

证明 记 $\mathcal{B} = \sigma(\mathcal{A})$，由定理 (7.1) 有 $\mathcal{B}$ 为单调类，则有 $\mathcal{M}(\mathcal{A}) \subset \mathcal{B}$

1) $\mathcal{M}(\mathcal{A})$ 在补运算下封闭。令

$$\mathcal{M}'(\mathcal{A}) = \{B | (B \in \mathcal{M}(\mathcal{A})) \wedge (B^c \in \mathcal{M}(\mathcal{A}))\}$$

由 $\mathcal{A}$ 为事件代数知 $\mathcal{A} \subset \mathcal{M}'(\mathcal{A}) \subset \mathcal{M}(\mathcal{A})$。设 $B_n \uparrow B, B_n \in \mathcal{M}'(\mathcal{A})$，则有 $B \in \mathcal{M}(\mathcal{A})$ 且

$$B^c = (\lim \uparrow B_n)^c = \lim \downarrow B_n^c \in \mathcal{M}(\mathcal{A})$$

类似地，若 $B_n \downarrow B, B_n \in \mathcal{M}'(\mathcal{A})$，则有 $B \in \mathcal{M}(\mathcal{A})$ 且

$$B^c = (\lim \downarrow B_n)^c = \lim \uparrow B_n^c \in \mathcal{M}(\mathcal{A})$$

则 $\mathcal{M}'(\mathcal{A})$ 为单调类，由 $\mathcal{M}(\mathcal{A})$ 最小性即得 $\mathcal{M}'(\mathcal{A}) = \mathcal{M}(\mathcal{A})$.

2) $\mathcal{M}(\mathcal{A})$ 在交运算下封闭。设 $A \in \mathcal{M}(\mathcal{A})$，令

$$\mathcal{M}_A = \{C | (C \in \mathcal{M}(\mathcal{A})) \wedge (A \cap C \in \mathcal{M}(\mathcal{A}))\}$$

再设 $\{C_n | (n \in \mathbb{N}^*\}$ 为 $\mathcal{M}_A$ 中单调序列，由 $\lim\limits_{n\to\infty} AC_n = A \cap \lim\limits_{n\to\infty} C_n$，则 $\mathcal{M}_A$ 为单调类。又若 $A \in \mathcal{A}$，则由 $\mathcal{M}_A$ 定义得 $\mathcal{A} \subset \mathcal{M}_A$，则 $\mathcal{M}_A$ 为包含 $\mathcal{A}$ 的单调类，由 $\mathcal{M}(\mathcal{A})$ 的最小性即得 $\mathcal{M}_A = \mathcal{M}$ 。注意到 $(\forall A, B \in \mathcal{M}) : [(A \in \mathcal{M}_B) \Leftrightarrow (B \in \mathcal{M}_A)]$，则当 $A \in \mathcal{A}, B \in \mathcal{M}(\mathcal{A})$ 时，由 $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$ 得 $B \in \mathcal{M}_A$，则 $A \in \mathcal{M}_B$。由此 $\mathcal{A} \subset \mathcal{M}_B$，又 $\mathcal{M}_B$ 为单调类，由 $\mathcal{M}$ 的最小性即得 $\mathcal{M}_B = \mathcal{M}$

由 1) 和 2) 由定理 (7.1) 即有 $\mathcal{B} \subset \mathcal{M}(\mathcal{A})$，则 $\mathcal{B} = \mathcal{M}(\mathcal{A})$

注 下利用 Zorn 引理指出更一般的定理

定理 7.3 (单调类定理/monotone class theorem)

设 $\mathcal{C}$ 为 Ω 上包含 $\varnothing$ 且对于运算 $(\cup f, \cap f)$ 封闭的集类。设 $\mathcal{M}$ 为 Ω 上包含 $\mathcal{C}$ 的单调类的类，$\mathcal{S}$ 为包含 $\mathcal{C}$ 且对运算 $(\cup d, \cap d)$ 封闭的最小集类，则 $\mathcal{M}$ 包含 $\mathcal{S}$。进一步，若 $\mathcal{C}$ 对于补运算封闭，则 $\mathcal{M}$ 包含由 $\mathcal{C}$ 生成的 σ 代数

证明 为简单起见，该证明称对于运算 $(\cup f, \cap f)$ 封闭的集类为族。由 Zorn 引理[1]，设 $\mathcal{H}$ 为 $\mathcal{M}$ 中包含 $\mathcal{C}$ 的一个极大族，下证 $\mathcal{H}$ 对于运算 ($\cup d, \cap d)$ 封闭。

设 (A_n) 为 $\mathcal{H}$ 中的一个单调下降序列且令 $A = \cap A_n$。所有形如 $(H \cap A) \cup H'$ （其中 $H \in \mathcal{H} \cup \{\Omega\}$, $H' \in \mathcal{H}$) 的子集组成的集类包含 $\mathcal{H}$ (令 $H = \varnothing$) 和 A （令 $H = \Omega, H' = \varnothing$)，并且它为 $\mathcal{M}$ 中的族。由 $\mathcal{H}$ 为极大族，则上述构造的族等于 $\mathcal{H}$，则有 $A \in \mathcal{H}$，亦即 $\mathcal{H}$ 对于运算 $(\cap d)$ 封闭，则包含 $\mathcal{C}$ 且对于运算 $(\cap d)$ 封闭的最小集类包含在 $\mathcal{M}$ 中。同理可证 $(\cup d)$ 的情形。

设 $\mathcal{T}$ 为满足 $A \in \mathcal{S}$ 和 $A^c \in \mathcal{S}$ 的集合 A 的全体。若 $\mathcal{C}$ 中每一元素的补集仍属于 $\mathcal{C}$ （或者更一般地考虑 $\mathcal{S}$)，则 $\mathcal{T}$ 包含 $\mathcal{C}$。特别地，$\varnothing \in \mathcal{T}$。显然 $\mathcal{T}$ 为 $\mathcal{M}$ 中的 σ 代数，定理得证

推论 7.1 (概率扩张定理)

设 $\mathcal{F}_0$ 为 Ω 上对于运算 $(\cup f, {}^c)$ 封闭的一个集类。记 $\mathcal{T}(\mathcal{F}_0) = \mathcal{F}$. 设 $\mathbb{P}$ 和 $\mathbb{P}'$ 为 $\mathcal{F}$ 上的两个概率测度满足 $(\forall A \in \mathcal{F}_0) : \mathbb{P}(A) = \mathbb{P}'(A)$，则有 $\mathbb{P}$ 和 $\mathbb{P}'$ 在 $\mathcal{F}$ 上相等.

证明 在单调类定理 (7.3) 中令 $\mathcal{C} = \mathcal{F}_0$，并令 $\mathcal{M}$ 为 $\mathcal{F}$ 中使得 $\mathbb{P}(A) = \mathbb{P}'(A)$ 的元素 A 组成的集合，定理即证

[1]该定理可不使用 Zorn 引理证明，见 Chung K L. A Course in Probability Theory Harcourt[J]. Brace and World, New York, 1968. 的第 17 页

Bibliography

[1] 冯荣权, 宋春伟. 组合数学. 北京: 北京大学出版社,2015.8.

[2] Stanley R P. Enumerative Combinatorics(Vol 1). Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press,1997,4.

[3] 苏淳, 冯群强. 概率论（第三版）. 北京: 科学出版社,2020.4

[4] 施利亚耶夫. 概率（第 1 卷）. 北京: 高等教育出版社,2007.7.

[5] 盛骤, 谢式千, 潘承毅. 概率论与数理统计（第 4 版）. 北京: 高等教育出版社,2008.6.

[6] 何书元. 概率论. 北京: 北京大学出版社,2006.1.

[7] 缪柏其, 胡太忠. 概率论教程. 合肥: 中国科学技术大学出版社,2009.5.

[8] 陈希孺. 概率论与数理统计. 合肥: 中国科学技术大学出版社,1992.5.

[9] 陈希孺. 数理统计学简史. 长沙: 湖南教育出版社,2002.

[10] Chung K L. A Course in Probability Theory Harcourt[J]. Brace and World, New York, 1968.

[11] 马林. 六西格玛管理. 北京: 中国人民大学出版社,2004.

[12] 韦博成. 漫话信息时代的统计学. 北京: 中国统计出版社,2011.

[13] Sidney I. Resnick. USA:A Probability Path. School of Operations Research and Information Engineering in Cornell University,1998