

概率论学习笔记

杨勇,2019110294

Yang945841548@bupt.edu.cn

2019 年 10 月 7 日

前言

文学家王蒙曾说过：最高的数学和最高的诗一样，都充满了理想，充满了智慧，充满了创造，充满了章法，充满了和谐也充满了挑战。诗和数学又都充满了灵感，充满了激情，充满了人类的精神力量。那些从诗中体验到数学的诗人是好诗人，那些从数学中体会到诗意的人是好数学家。最高的诗是数学，这是文学家对数学的推崇。亲爱的同学们，诗言志，希望你们能从这里读出诗的理想，接受诗的智慧，学会诗的创造，掌握诗的章法，体验诗的和谐，勇敢地迎接未来的挑战。

基础数学分为代数、几何、分析三大类。而概率论则是一门应用数学，同学们在前面的学习中惟独没有这种应用数学。初次走进测度论，第一个感觉是难，这个感觉恰如第一次接触“诗经”的“风、雅、颂”一样。不要怕难，这是诗意的挑战。在学习这门课的过程中，抽象概念会变得形象直观，给我们打下深刻的烙印。通过严格的逻辑推理和抽象思维的训练，伴随着概率的直觉与对随机性的想象力，概率论中的基础知识将融化在你的血液中。你会感觉概率论和诗一样，充满了灵感，充满了激情，充满了人类的精神力量，你的数学生涯也就化为诗样的人生。

本书是本人在学习测度论的过程中形成的学习笔记。本书在许多证明上的推导为了保持数学的严格性而变得很冗长，这也要请同学们根据自己的情况适当地选择，不必每个证明都看；更不要因为有几个证明想不明白而失去信心。对作者而言，完善一个定理的证明，有时也累日始成，可以说不时就遇到了不小的困难，倾注了不少心力。对有能力的读者而言，应尝试独立去证明每一个结论。如碰到一个证明一时想不明白，宁肯暂时搁一搁，也不要轻易地直接相信书本上的内容。譬如登山，经过艰苦努力上了峰顶，自有其乐趣和成就感。反之，如在未尽全力之前就任人抬上去，则不惟无益，实足以挫折信心。

说句实话，理解这样的课程内容并获得灵活使用它解决具体问题的能力是个复杂的过程。这门课程的难度相比数学分析会呈现出一个飞跃。初学者不能指望这个过程会在完全不遇到困难的情况下轻松地通过。大家会遇到许多与前不同的、大大小小的困境，这是完全可以理解的，在一定意义上来说也是正常的。具有强烈求知欲的学而不厌者方能成功地完成这个困难的过程。如果读者在阅读本书时遇到了困难，可与作者联系，电子邮件：Yang945841548@bupt.edu.cn

我的老师仔细审阅了本书初稿，并对许多部分的推导提供了很好的修改意见，特致谢意。作者在参考了这些意见的基础上，对本书的初稿进行了修改，使得笔记的质量有了很大的提高。作者还要感谢母校的同学们给予的支持和帮助。

囿于作者的水平和经验，书中不妥以至谬误之处，在所难免。尚祈读者不吝指教。最后，赠送一句话与读者共勉：

纸上得来终觉浅，绝知此事要躬行。

杨 勇

2019 年 10 月于北京昌平

目 录

第零章 引言	1
0.1 主要内容 & 主要参考资料	1
0.2 概率论评述	1
 第一部分 测度论	 13
第一章 测度空间与概率空间	15
1.1 可测空间	15
1.1.1 集合及其运算	15
1.1.2 集类	17
1.1.3 集类的生成	23
1.2 测度与测度的构造	28
1.2.1 测度、可测空间与测度空间	28
1.2.2 测度在集代数上的扩张	28
1.2.3 外测度	28
1.2.4 从半集代数到代数上的测度扩张	28
1.3 测度的性质	28
1.3.1 测度的运算性质	28
1.3.2 测度空间的完全化	28
1.3.3 Lebesgue-Stieltjes 测度	28
 第二章 可测函数与随机变量	 29
2.1 可测函数与分布	29
2.2 分布与分布函数	30
2.2.1 可测函数的运算	30
2.3 复合映射的可测性	30
2.4 可测映射列极限的可测性	30
2.5 可测函数的构造	30

第三章 积分与数学期望	31
3.1 积分的定义	31
3.2 积分的性质	31
3.3 空间 $L_p(X, \mathcal{F}, \mu)$	31
3.4 概率空间的积分	31
第四章 不定积分与条件期望	33
4.1 符号测度	33
4.2 符号测度的分解	33
4.2.1 Hahn 分解和 Jordan 分解	33
4.2.2 Radon-Nikodym 定理	33
4.2.3 Lebesgue 分解	33
4.3 条件期望与条件概率	33
第五章 乘积空间	35
5.1 有限维乘积空间	35
5.2 多维 Lebesgue-Stieltjes 测度	35
5.3 可列维乘积空间的概率测度	35
5.4 任意无穷维乘积空间的概率测度	35
第六章 离散鞅论	37
6.1 基本概念	37
6.2 停时定理	37
6.3 收敛定理	37
6.4 鞅的不等式	37
第二部分 概率极限理论	39
第七章 大数定律	41
7.1 零壹律和三级数定理	41
7.2 弱大数律	41
7.3 强大数律	41
7.4 随机序列的收敛性	41
7.5 更新过程	41
7.6 大偏差理论	41

第八章 中心极限定理	43
8.0.1 De Moivre-Laplace 定理	43
8.0.2 特征函数	43
8.0.3 中心极限定理	43
8.0.4 局部极限定理	43
8.0.5 Poisson 收敛和 Poisson 过程	43
第九章 无穷可分分布和普适极限定理	45
9.1 无穷可分分布函数	45
9.2 独立随机变量和的分布	45
9.3 L 族和稳定分布族	45
第十章 弱收敛理论	47
10.1 距离空间概率测度的弱收敛	47
10.2 鞅的极限定理	47
第十一章 离散鞅论	49
第三部分 随机过程论	51
第十二章 Markov 链	53
第十三章 遍历定理	55
第十四章 布朗运动	57
第十五章 Skorokhod 嵌入定理	59
第十六章 多维布朗运动	61
第十七章 概率不等式	63
17.1 随机事件的不等式	63
17.2 分布函数、特征函数的不等式	63
17.3 随机变量的概率不等式	63
17.4 矩不等式	63
17.5 随机变量的概率与矩之间的不等式	63
17.6 部分和及其极值的矩估计	63
17.7 指数不等式	63

参考文献	65
附 录	67
附录	67
附录 1 集合论的温习	67
附录 2 中国的概率论之父——许宝騄先生	73

第零章 引言

0.1 主要内容 & 主要参考资料

- 测度论
- 随机过程论

1. R.Durrett, "Probability: Theory and Examples", Cambridge University Press, 2019.
2. 程士宏, "测度论与概率论基础", 北京: 北京大学出版社, 2004.
3. 严加安, "测度论讲义"(第二版), 北京: 科学出版社, 2005.
4. Terence Tao, "Analysis I", Springer-Verlag, New York, 2016.

0.2 概率论评述

起源

概率论这门学科可以说起源于赌博。尽管早在 15 世纪与 16 世纪意大利的一些数学家(如 Cardano, Pacioli, Tartaglia 等)已经对一些靠运气的游戏中的特定概率进行了计算,但是,概率论作为一门学科起源于 17 世纪。1654 年,一个名叫 A.G.C. de Méré 的法国贵族对赌博以及赌博中的问题很感兴趣,但他对一些问题感到很困惑,为了解决自己的困惑,他向数学家 B.Pascal (1623-1662) 求助。为了解答 de Méré 提出的问题, Pascal 与法国数学家 P.Fermat (1601-1655) 进行了通信讨论。1655 年,荷兰数学家 C.Huygens(1629-1695) 首次访问巴黎,期间他学习了 Pascal 与 Fermat 关于概率论的工作。Huygens 是一个名声和 Newton 相当的大科学家。人们熟知他的贡献之一是物理中的单摆公式。他在概率论的早期发展历史上也占有重要的地位。1657 年,当他回到荷兰后,他写了一本小册子,名叫《De Ratiociniis in Ludo Aleae》(可译为《机遇的规律》),这是关于概率论的第一本书。在这部著作中,他首先引进了“期望”这个术语。在此,“数学期望”这个基本概念以及关于概率的可加性、可乘性已经建立。他基于这些术语解决了一些当时感兴趣的博弈问题。他在这部著作中提出了 14 条命题,第一条命题是:

如果某人在赌博中以一半的可能性赢 a 元,以一半的可能性输 b 元,则他的期望是

$$\frac{a-b}{2}(\text{元}). \quad (0.1)$$

18-19 世纪

概率论在 18 世纪得到了快速发展,这个期间的主要贡献者是 J.Bernoulli(1654-1705) 和 A. de Moivre (1667-1754)。在 19 世纪,概率论的早期理论得到了进一步的发展与推广,这个期间的主要贡献者是 P.S.M. Laplace (1749-1827), S.D. Poisson (1781-1840), C.F. Gauss (1777-1855), P.L. Чебышёв(切比雪夫)(1821-1894), 马尔可夫 (А.А.Марков)(1856-1922) 与李雅普诺夫 (А.М.Ляпунов)(1857-1918)。这个时期的研究主要围绕着概率极限定理展开。

Jacob Bernoulli 是一位瑞士数学家, 是 Bernoulli 家族的第一位数学家. 1705 年, 他在瑞士 Basel 去世. 8 年后, 也就是 1713 年, 他在概率论领域的代表作《Ars Conjectandi》一书正式出版(可译为《猜测的艺术》). 他的这本书正式出版标志着概率论学科的开始. 该书仅包含一个数学定理, 即著名的 Bernoulli 大数定律, 这是概率论的第一个极限定理.

Bernoulli 大数定律: 假设 $\{\xi_n\}$ 为一个独立同分布的随机变量序列, $P(\xi_n = 1) = p, P(\xi_n = 0) = 1 - p$, 其中 $0 < p < 1$. 令 $S_n = \sum_{k=1}^n \xi_k$, 那么

$$P\left(\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty. \quad (0.2)$$

17 世纪下半叶, Newton 刚发明微积分不久, 人们对计算各种数列的极限有着相当大的兴趣, 并发展了不少有效的方法和技巧. 但是, “ $\varepsilon - N$ ” 语言并不能很好地解释“随机试验中频率是否收敛到概率”这样的问题. 正是 Bernoulli 大数律首次给出了“频率收敛到概率”的数学解释和严格证明.

Bernoulli 大数律内涵丰富, 成为了后人发展概率论的源泉. (0.2) 式充分肯定了经验观测可以揭示随机现象规律的基本思想; 提出了随机变量序列依概率收敛到常数(甚至收敛到随机变量)的基本概念. 受此启发, 人们自然会问: 如果考虑一般的随便变量的平均值, 情况会如何? 特别, 假设 $\{\xi_n\}$ 为一个独立同分布的随机变量序列, $E\xi_n = \mu$. 令 $S_n = \sum_{k=1}^n \xi_k$, 那么

$$P\left(\omega : \left|\frac{S_n(\omega)}{n} - \mu\right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty \quad (0.3)$$

成立么?

用观测平均值去计算真值的思想很早以前就已出现, 并一直用于日常生活和社会实践, 关键在于能否给出一个严格的数学证明吗?

回顾 (0.2) 式的证明, Bernoulli 二项分布起着关键作用. 事实上, $S_n \sim \mathcal{B}(n, p)$. 因此

$$P\left(\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right) = \sum_{k: |k-np| > n\varepsilon} \binom{n}{k} p^k (1-p)^{n-k}. \quad (0.4)$$

利用杨辉三角或者 Pascal 二项组合式, 容易得到

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1. \quad (0.5)$$

但是, (0.4) 的困难在于计算部分和, 而不是对所有 $k = 0, 1, \dots, n$ 求和. 为此, Bernoulli 利用了 $n!$ 的渐进计算公式. 显然, 这样一个计算技巧对 (0.4) 不合适, 因为 ξ_n 的分布并不知道. 事实上, 即使 ξ_n 只取三个值, Bernoulli 的计算方法仍然显得笨拙而不可行!

为了证明 (0.4), 需要用到下面的 Чебышёв(切比雪夫) 不等式: 假设 X 是随机变量, $EX = \mu$, $\text{var}(X) = \sigma^2$, 那么对任意 $x > 0$

$$P(|X - EX| \geq x) \leq \frac{\text{var}(X)}{x^2}. \quad (0.6)$$

在当代概率论中, 与 Poisson 相关的有 Poisson 分布、Poisson 过程. Gauss 创立了误差理论, 特别地, 创立了最小二乘的基本方法. 切比雪夫(Чебышёв), 马尔可夫(Марков) 与李雅普诺夫(Ляпунов) 在研究独立但不同分布的随机变量和的极限定理方面发展了有效的方法. Чебышёв(切比雪夫) 被看作

俄国现代数学之父. 上述切比雪夫 (Чебышёв) 不等式在概率论学科发展中起着举足轻重的作用. 事实上, 概率自 16 世纪由赌博游戏引入以来, 到 19 世纪末, 历经 300 年. 尽管由不少的大数学家热衷于概率的研究并积极倡导和应用, 但相对于这一时期的分析、代数、几何等其它数学分支而言, 概率论的发展可以说是十分缓慢, 并且, 在总体上来说, 概率论学科还停留在一些具体事件的概率计算上. 在切比雪夫 (Чебышёв) 之前, 概率论的主要兴趣在于对随机试验的概率进行计算. 而切比雪夫 (Чебышёв) 是第一个清晰认识并充分利用随机变量及其数学期望的人. 切比雪夫 (Чебышёв) 思想的主要倡导者是他忠诚的学生马尔可夫 (Марков), 他将老师的结果完整清晰地展现出来. 马尔可夫 (Марков) 自己对概率论的重大贡献之一是创立了概率论的一个分支: 研究相依随机变量的理论, 成为“马尔可夫 (Марков) 过程”. 可以说, 切比雪夫 (Чебышёв) 不等式可以看作一个时代的转折点.

作为应用, 可以用 (0.6) 证明 (0.3):

$$P\left(\omega : \left| \frac{S_n(\omega)}{n} - p \right| > \varepsilon\right) \leq \frac{\text{var}(S_n/n)}{\varepsilon^2} = \frac{\text{var}(S_n)}{n^2\varepsilon^2}. \quad (0.7)$$

既然 $\varepsilon > 0$ 是任意给定的正数, 那么只要验证 $\text{var}(S_n) = o(n^2)$ 就够了. 在二项分布情形,

$$\text{var}(S_n) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} = np(1-p). \quad (0.8)$$

注意, 正如 (0.7) 一样, (0.8) 的计算比 (0.4) 要容易得多. 更为重要的是, 上述讨论不局限于 Bernoulli 二项分布, 而适用于非常广泛的随机变量序列. 以下是切比雪夫 (Чебышёв) 大数定律: 假设 $\{\xi_n\}$ 为一个随机变量序列, $E\xi_n = \mu_n$, $\text{var}(\xi_n) = \sigma^2$. 如果

$$\text{var}(S_n) = o(n^2), \quad (0.9)$$

那么

$$P\left(\omega : \left| \frac{S_n(\omega)}{n} - \frac{1}{n} \sum_{k=1}^n \mu_k \right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty. \quad (0.10)$$

该定理的条件 (17) 包含的范围非常广泛. 如独立同分布且方差有限; 独立不同分布且方差有界; 不独立但互不相关且方差有界; 其它相依情形. 后几种情形更为常见, 在实际观测时, 并不能苛求试验环境, 数据之间不可避免地存在一定联系, 并且不能保证同分布.

切比雪夫 (Чебышёв) 大数律显然是 Bernoulli 大数律的极大推广. 但是条件 (17) 对于一个大数律来说无疑有点强. 辛钦 (Хинчин) 改进了切比雪夫 (Чебышёв) 大数律, 在独立同分布且数学期望存在有限的情况下, 证明了 (10). 当然, 此时无法直接使用切比雪夫 (Чебышёв) 不等式 (13), 截尾方法应运而生.

Bernoulli 大数律告诉人们: 给定任意精度, 只要试验重复足够多次, 频率就很有可能接近概率真值, 以致误差在给定精度内. 人们自然会问, 试验次数究竟多大合适? 该如何确定呢? 当然, 给定精度 $\varepsilon > 0$, 无论 n 多么大, 都无法保证独立重复试验后 $|S_n/n - p| \leq \varepsilon$. n 的大小取决于事先给定的可靠度 (置信度), 关键在于如何由精度和可靠度来确定 n . 即, 给定 $\varepsilon > 0, \eta < 1$, 如何有效地表达 (近似)

$$P\left(\omega : \left| \frac{S_n(\omega)}{n} - p \right| > \varepsilon\right) = \eta. \quad (0.11)$$

de Moivre 和 Laplace 考虑了上面的问题, 并对独立二点分布的随机变量序列证明了以下中心极限定理: 对任意实数 $a < b$,

$$P\left(\omega : a < \frac{S_n(\omega) - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \Phi(b) - \Phi(a). \quad (0.12)$$

直接应用该结果, (18) 可写成

$$P\left(\omega: \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right) \approx 2\left(1 - \Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)\right) =: \eta. \quad (0.13)$$

由此可计算 n 的大小 (依赖于 ε, n, p).

de Moivre 是一位法国数学家, 但是大部分时间他住在英国. de Moivre 在前人, 特别是 Bernoulli 家族和 Huygens 的基础上, 研究和发展概率论, 可以说, 他开创了概率论的现代方法: 1711 年出版了《The Doctrine of Chances》. 在此书中, 统计独立性的定义首次出现. 该书在 1738 年与 1756 年出了扩版, 生日问题出现在 1738 年的版本中, 赌徒破产模型出现在 1756 年的版本中, 并在赌徒中很有影响和地位. 1730 年, de Moivre 的另一本专著《Miscellanea Analytica Supplementum》(可译为《解析方法》) 正式出版. 其中, 关于对称 Bernoulli 试验的中心极限定理首次提出并得到证明. 他首先考虑了 $p = 1/2$ 情形, 并和他的好朋友 Stirling 同时发现了下列公式

$$n! \sim \sqrt{2\pi n} n^n e^{-n}. \quad (0.14)$$

通常称 (21) 为 Stirling 公式, 实际上应该为 de Moivre-Stirling 公式.

差不多 40 年后, Laplace 考虑了 $p \neq 1/2$ 情形. 应该说, Laplace 对概率统计和天体力学的贡献巨大. 他在 1799-1825 年间出版了五卷本《Celestial Mechanics》. 1812 年, Laplace 的伟大专著《Théorie Analytique des Probabilités》(可译为《概率论的解析理论》) 诞生, 其中, 他阐述了他自己及前辈在概率论方面的结果. 特别地, 他将 de Moivre 的定理推广到 Bernoulli 试验非对称情形. Laplace 最重要的工作是将概率方法应用到观测误差, 在很一般的条件下证明了观测误差的分布一定是渐进正态的. 直到今天, 人们在概率极限理论方面的研究还受到 Laplace 的影响.

正如 Bernoulli 大数定律一样, de Moivre-Laplace 中心极限定理对概率论学科的发展影响深远, 受 (19) 启发, 人们提出了随机变量序列依分布收敛的概念, 并给出了一般形式的中心极限定理: 假设 ξ_1, ξ_2, \dots 是一独立同分布的随机序列, $E\xi_n = \mu, \text{var}(\xi_n) = \sigma^2$, 那么

$$P\left(\omega: \frac{S_n(\omega) - n\mu}{\sqrt{n}\sigma} \leq x\right) \rightarrow \Phi(x), \quad n \rightarrow \infty. \quad (0.15)$$

但是, 如何证明呢? 我们需要找到一个有效的工具和判别准则. 随着调和分析的发展, 人们发现 Fourier 变换是研究分布函数的一个有效工具. 在概率论学科中, 大家称分布函数的 Fourier 变换为特征函数. 为证明概率论的中心极限定理, 切比雪夫 (Чебышёв) 和马尔可夫 (Марков) 利用的是矩方法, 而李雅普诺夫 (Ляпунов) 利用了特征函数的方法. 极限定理的后续发展表明特征函数方法是一种强大的解析工具.

假设 X 是个随机变量, 定义 $\phi(t) = Ee^{itX}$ 为其特征函数 (c.f.). 任意的随机变量都存在特征函数, 并且它具有非常良好的分析性质、运算性质和唯一性. 随着特征函数的引入, 20 世纪 20-30 年代概率论的发展进入了一段黄金时期. 法国数学家 Lévy 建立了连续性定理: $X_n \xrightarrow{d} X$ 当且仅当 $\phi_n(t) \rightarrow \phi(t)$. 其中, ϕ_n 和 ϕ 分别为 X_n 和 X 的 c.f.. 由此, 可以证明 (22). 事实上, 令 $X_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$, 那么

$$Ee^{itX_n} = \prod_{k=1}^n Ee^{it\frac{\xi_k - \mu}{\sqrt{n}\sigma}} = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n \rightarrow e^{-t^2/2}. \quad (0.16)$$

所以, (22) 成立. 这被称为 Lévy-Lindeberg 中心极限定理, 这个证明已经写入许多的本科生概率论教材, 具有微积分基础的同学们都能掌握.

Lindeberg 和 Feller 研究了独立不同分布情形, 他们再次运用了特征函数的方法证明了下列定理: 假设 $\{\xi_n; n \geq 1\}$ 为一独立的随机序列, 相应的分布函数分别为 F_n , 并且 $E\xi_n = \mu_n$, $\text{var}(\xi_n) = \sigma_n^2 < \infty$. 若 $B_n^2 = \sum_{k=1}^n \sigma_k^2 \rightarrow \infty$. 那么

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{B_n^2} \rightarrow 0 \quad (0.17)$$

和

$$\frac{1}{B_n} \sum_{k=1}^n (\xi_k - \mu_k) \xrightarrow{d} \mathcal{N}(0, 1) \quad (0.18)$$

成立当且仅当对任意 $\varepsilon > 0$,

$$\frac{1}{B_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \varepsilon B_n} (x - \mu_k)^2 dF_k(x) \rightarrow 0. \quad (0.19)$$

称 (24) 为 Feller 条件, (26) 为 Lindeberg 条件. 当随机变量同分布且方差都存在时, 这些条件都满足. 当随机变量不同分布时, (24) 意味着各个随机变量 ξ_k/B_n “一致地无穷小, 没有一个起显著作用”. 正如切比雪夫 (Чебышев) 大数律一样, Lindeberg-Feller 定理应用非常广泛, 譬如说各类测量误差可以近似地用正态分布描述.

中心极限定理不仅适用于独立随机变量的部分和, 而且可以推广到许多相依随机变量序列情形, 如鞅差序列, 马尔可夫 (Марков) 链, 各类混合序列, 正、负相依 (伴) 序列等, 从而发展了许多新的方法, 如 20 世纪 70 年代提出的 Stein 方法.

除 Bernoulli 大数律和 de Moivre-Laplace 中心极限定理外, 另一个经典极限定理是 Poisson 极限定理. 它讨论的仍然是二项分布. 假设 $\{S_n; n \geq 1\}$ 为一列二项分布随机变量, $S_n \sim \mathcal{B}(n, p_n)$. 如果 $np_n \rightarrow \lambda > 0$, 那么对每个 $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(\omega : S_n(\omega) = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (0.20)$$

注意, $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$. 因此, 可以构造一个随机变量 X , 具有分布

$$P(\omega : X(\omega) = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \quad (0.21)$$

称 X 是服从 Poisson 分布的随机变量, 记作 $X \sim \mathcal{P}(\lambda)$. (27) 可写作

$$S_n \xrightarrow{d} \mathcal{P}(\lambda). \quad (0.22)$$

(29) 的证明并不难, 利用 Stirling 公式 (21) 直接计算或利用 Lévy 连续性定理均可.

以上三大极限定理讨论的都是 Bernoulli 二项分布的随机变量.

20 世纪

20 世纪可称为概率论发展的现代时期, 本时期开始于概率论的公理化. 在这个方向上的早期贡献者有 S.N. Bernstein (1880-1968), R. von Mises (1883-1953) 与 E. Borel (1871-1956). 1933 年, 俄罗斯著名数学家柯尔莫哥洛夫 (А.Н. Колмогоров) 出版了他的伟大专著《Foundations of the Theory of Probability》. 书中, 他为概率论建立了目前广泛采纳的公理化体系. 这一时期, 中国数学家、概率论先驱许先生 (Paolu Hsu, 1910-1970) 在内曼-皮尔逊理论、参数估计理论、多元分析、极限理论等方面取得卓越成就, 许宝騄先生是多元统计分析学科的开拓者之一.

20 世纪 20-30 年代, 人们试图寻求一种普适极限定理, 来描述随机现象的内在规律. 假设 $\{\xi_{n,k}; 1 \leq k \leq k_n, n \geq 1\}$ 为行内独立的三角组列, 令

$$S_n = \sum_{k=1}^{k_n} \xi_{n,k} \quad (0.23)$$

表示组列的第 n 行随机变量的和. 如果对任意 $\varepsilon > 0$,

$$\max_{1 \leq k \leq k_n} P(\omega : |\xi_{n,k}(\omega)| > \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty), \quad (0.24)$$

那么称 $\{\xi_{n,k}\}$ 满足无穷小条件. 该条件意味着每行内的随机变量“一致地小, 没有一个起显著作用”. 如果 S_n 解释为测量误差, 那么 $\{\xi_{n,k}; 1 \leq k \leq k_n\}$ 可看作是造成误差的诸多细微的因素, 每一个因素都会导致测量误差, 但没有系统误差. 人们自然会问: 用什么描述 S_n 的分布比较合适? 事实上, 当误差的因素可以细分, 并且满足无穷小条件 (31) 时, 误差 S_n 的分布是无穷可分分布.

假设 X 是随机变量, $F(x)$ 是它的分布函数, $\phi(t)$ 是它的特征函数. 如果对任意 $n \geq 1$, 存在随机变量 $\xi_{n,1}, \dots, \xi_{n,n}$ 使得

$$X \triangleq \sum_{k=1}^n \xi_{n,k}, \quad (0.25)$$

那么称 X 为无穷可分随机变量. 等价地, 可以用分布函数或特征函数来描述. 无穷可分分布族包括退化单点分布、正态分布、Poisson 分布以及它们的混合分布; 但不包括有界非退化随机变量. 一个分布可以是无穷可分分布当且仅当它的特征函数 $\phi(t)$ 可以写成下列 Lévy-Хи́нчин 表示:

$$\phi(t) = \exp \left(i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x) \right), \quad (0.26)$$

其中, γ 为常数, 函数 $G(x)$ 满足单调不减右连续左极限存在, $G(-\infty) = 0$ 并且 $G(\infty) < \infty$.

普适极限定理断言: 如果无穷小三角组列 $\{\xi_{n,k}\}$ 的行和 S_n 依分布收敛到某个随机变量 X , 那么 X 一定是无穷可分分布的随机变量. 进而, 可以给出收敛到某给定的无穷可分分布的充要条件. 事实上, 三大经典极限定理都是普适极限定理的特殊情况. 可以说, 普适极限定理是 20 世纪 20-30 年代的杰作, 它将特征函数方法运用到极致. 然而, 由于理论过于一般化, 证明相当繁琐, 初学者不易掌握.

几乎处处收敛. 假设 $X, X_n, n \geq 1$ 为一列定义在概率空间 (Ω, \mathcal{A}, P) 上的随机变量, 如果存在一个 Ω_0 , 使得 $P(\Omega_0) = 1$, 并且对每一个 $\omega \in \Omega_0$ 都有

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega), \quad (0.27)$$

则称 X_n 几乎处处收敛到 X , 记作 $X_n \rightarrow X, \text{a.s.}$. 就概念本身而言, 几乎处处收敛是容易理解的. 在随机变量序列的各种各样收敛性中, 几乎必然收敛是最强的收敛性之一. 如何判别某随机变量序列几乎处处收敛呢? 不难看出 $X_n \rightarrow X, \text{a.s.}$ 当且仅当对任意 $\varepsilon > 0$,

$$P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon, \text{i.o.}) = 0. \quad (0.28)$$

这等价于

$$\lim_{n \rightarrow \infty} P(\omega : \sup_{k \geq n} |X_k(\omega) - X(\omega)| > \varepsilon) = 0. \quad (0.29)$$

显然, 一个充分条件为

$$\sum_{n=1}^{+\infty} P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon) < \infty. \quad (0.30)$$

更一般地, 有下列 Borel-Cantelli 引理: 假设 $\{A_n; n \geq 1\}$ 是一列事件, 如果 $\sum_{n=1}^{+\infty} P(A_n) < \infty$, 那么 $P(A_n, \text{i.o.}) = 0$. 当然, 级数收敛这一条件要强得多. 不过, 如果 $\{A_n\}$ 是一列两两独立的事件, 并且 $P(A_n, \text{i.o.}) = 0$, 那么 $\sum_{n=1}^{+\infty} P(A_n) < \infty$.

Borel 强大数律: 假设 $\{\xi_n, n \geq 1\}$ 是一列独立同分布的随机变量, $P(\xi_n = 1) = p, P(\xi_n = 0) = 1 - p$. 令 $S_n = \sum_{k=1}^n \xi_k$, 那么

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = p, \text{ a.s.} \quad (0.31)$$

显然, Borel 强大数律更为深刻地解释了“频率收敛到概率的基本事实”. 值得强调, 从 Bernoulli 大数律到 Borel 大数律历经两百年, 无数人对此进步付出了毕生精力. 证明是 Borel-Cantelli 引理的简单应用. 事实上,

$$\sum_{n=1}^{+\infty} P(\omega : |\frac{S_n(\omega)}{n} - p| > \varepsilon) \leq \sum_{n=1}^{+\infty} \frac{E|S_n - np|^4}{n^4 \varepsilon^4} < \infty. \quad (0.32)$$

如果有什么需要注意的话, 那就是 (38) 中所有的 ξ_n 都定义在同一个概率空间上, 以至于 $S_{n+1}(\omega) = S_n(\omega) + \xi_{n+1}(\omega)$.

正如我们所注意到的那样, Borel 大数律的证明之所以简单, 在于 $S_n \sim \mathcal{B}(n, p)$, 因此, 它的四阶矩存在, 并且 $E|S_n - np|^4 = o(n^2)$. 能否将辛钦 (Хинчин) 大数律加强为 a.s. 收敛么? 柯尔莫哥洛夫 (Колмогоров) 强大数律肯定地回答了这个问题.

柯尔莫哥洛夫 (Колмогоров) 是 20 世纪最伟大的数学家之一, 也是现代概率论的奠基人. 1920 年, 他进入莫斯科国立大学学习, 1929 年获得博士学位. 博士期间他完成了多篇论文, 其中最具代表性的工作有: 强大数律、三级数定理, 以及重对数律.

柯尔莫哥洛夫 (Колмогоров) 强大数律: 假设 $\{\xi_n; n \geq 1\}$ 为独立同分布的随机序列, 令 $S_n = \sum_{k=1}^n \xi_k$, 那么

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu, \text{ a.s.} \quad (0.33)$$

当且仅当 $E|\xi_1| < \infty, E\xi_k = \mu$.

条件的必要性是 Borel-Cantelli 引理的简单推论; 但充分性的证明要复杂得多. 为此, 柯尔莫哥洛夫 (Колмогоров) 创造了许多新的方法, 包括子序列方法和独立随机变量部分和的最大值不等式.

1929 年, 柯尔莫哥洛夫 (Колмогоров) 证明了著名的有界重对数律: 假设 $\{\xi_n; n \geq 1\}$ 为一独立随机变量序列. $E\xi_n = 0, \text{var}(\xi_n) = \sigma_n^2 < \infty$. 令 $S_n = \sum_{k=1}^n \xi_k, B_n^2 = \sum_{k=1}^n \sigma_k^2 \rightarrow \infty$. 如果 $|\xi| \leq M_n = o(\frac{B_n}{\sqrt{\ln \ln B_n^2}})$, a.s., 那么

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{\sqrt{2B_n^2 \ln \ln B_n^2}} = 1, \text{ a.s.} \quad (0.34)$$

这个定理的意义是多方面的. 它是以一种完全不同于大数律和中心极限定理的形式刻画了独立随机变量和的渐进性质. 它将 Borel-Cantelli 引理用到了极致, 还创建了指数型不等式.

对于独立随机变量序列 $\{\xi_n; n \geq 1\}$,

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{\sqrt{2n \ln \ln B_n^2}} = \sigma, \text{ a.s.} \quad (0.35)$$

当且仅当 $E\xi_k = 0, E\xi_k^2 = \sigma^2 < \infty$.

令人惊讶的是, 该定理直到 1941 年才由 Hartman-Wintner 证明.

19 世纪末、20 世纪初, 整个数学正在经历着一场革命. Lebesgue 测度论、Poincaré 的拓扑学、Hilbert 的 23 个问题, 都给 20 世纪的数学家留下了巨大的发展空间. 概率论学科经过了 200 多年的孕育后, 于 20 世纪初迎来了发展的黄金时期. 柯尔莫哥洛夫 (Колмогоров) 于 1933 年出版的著作《Foundations of the Theory of Probability》标志着现代概率论的开始. 此书共分四章, 主要内容包括一、构建概率论的公理化体系; 二、发展条件概率和条件期望; 三、给出无穷维分布的相容性条件; 四、证明独立随机变量和的极限定理. 它们为后来整个概率论学科的发展奠定了基础.

正如大家注意到的那样, 无论大数律、中心极限定理, 还是重对数律, 讨论的都是收敛性的问题. 以 Lévy-Feller 定理为例, 假设 $\{\xi_n; n \geq 1\}$ 独立同分布, $E\xi_n = 0, \text{var}(\xi_n) = 1$, 那么对每个 x ,

$$P\left(\omega : \frac{S_n}{\sqrt{n}} \leq x\right) \rightarrow \Phi(x), (n \rightarrow \infty). \quad (0.36)$$

由于 $\Phi(x)$ 是单调递增的连续函数, 所以上述收敛一致成立, 即

$$\Delta_n \triangleq \sup_{-\infty < x < \infty} |P\left(\omega : \frac{S_n}{\sqrt{n}} \leq x\right) - \Phi(x)| \rightarrow 0, (n \rightarrow \infty). \quad (0.37)$$

一个自然的问题是: Δ_n 趋于 0 的速度如何? 这不仅仅是个理论问题, 而且在数理统计中有着重要应用. Berry-Esseen 给出了下列结果: 假设 $\xi_k, 1 \leq k \leq n$ 独立同分布, $E\xi_k = 0, \text{var}(\xi_k) = 1, E|\xi_k|^3 < \infty$, 那么存在一个数值常数 $A > 0$, 使

$$\Delta_n \leq A \frac{E|\xi_1|^3}{\sqrt{n}} \quad (0.38)$$

对于独立不同分布情形, 类似结果成立:

$$\sup_{-\infty < x < \infty} |P\left(\omega : \frac{S_n - ES_n}{B_n} \leq x\right) - \Phi(x)| \leq A \frac{\sum_{k=1}^n E|\xi_k|^3}{B_n^3}. \quad (0.39)$$

注意, (45) 和 (46) 对任意 $n \geq 1$ 都成立. 有趣的是, 这个定理中, 如果不对 ξ_k 的分布加以假设, (45) 中的上界的阶是不能再改进了的. 例如, 如果 $P(\xi_n = \pm 1) = 1/2$, 那么 $P(S_n = 0) \sim n^{-1/2}$. 但是对于有界对称连续型随机变量, (45) 中上界的阶可以改进到 n^{-1} .

根据 (45),

$$P(\omega : \frac{S_n(\omega)}{\sqrt{n}} \leq x) = \Phi(x) + O(n^{-1/2}). \quad (0.40)$$

类似地

$$P(\omega : \frac{S_n(\omega)}{\sqrt{n}} > x) = 1 - \Phi(x) + O(n^{-1/2}). \quad (0.41)$$

这些结果对统计推断中置信区间估计和假设检验都非常有用. 但是, 当 x 比较大时, $1 - \Phi(x)$ 本身就很小, 有可能远比 $n^{-1/2}$ 小. 例如, 保险精算中破产概率及其风险的估计中就可能出现这种问题. 这样, 用 (48) 来估计 $P(\omega : S_n > \sqrt{n}x)$ 就当然不准确了. Cramér 证明了

$$\frac{P(\omega : S_n > \sqrt{n}x)}{1 - \Phi(x)} = (1 + o(1)) \exp\left(\frac{x^3}{\sqrt{n}} \lambda\left(\frac{x}{\sqrt{n}}\right)\right), \quad (0.42)$$

其中 $x = o(\sqrt{n})$, $\lambda(\cdot)$ 为 Cramér 级数. 称 (49) 为 Cramér 型大偏差, 用于描述小概率事件的渐进性质. 20 世纪 60-70 年代, Donsker, Varadhan 等进一步考虑了多维随机变量和随机过程的大偏差. 大偏差理论现在已经成为了概率极限理论的一个重要分支.

为了精确刻画大数律的收敛性, 许宝騄和 Robbins 在 1947 年证明了下列完全收敛性: 假设 $\{\xi_n; n \geq 1\}$ 是一列独立同分布的随机变量, $E\xi_1 = 0$, $E\xi_1^2 < \infty$, 那么对任意 $\varepsilon > 0$,

$$\sum_{n=1}^{+\infty} P\left(\omega : \left|\frac{S_n(\omega)}{n}\right| > \varepsilon\right) < \infty. \quad (0.43)$$

显然, 从上式可推出强大数律成立, 从而 $E|\xi_1| < \infty$ 并且 $E\xi_1 = 0$. 但是, 方差的存在性并不明显. 事实上, Erdős 在 1950 年证明了方差存在有限是 (50) 成立的必要条件. 而 Baum-Katz 在 1968 年证明了

$$\sum_{n=1}^{+\infty} \frac{1}{n} P\left(\omega : \left|\frac{S_n(\omega)}{n}\right| > \varepsilon\right) < \infty \quad (0.44)$$

当且仅当 $E|\xi_1| < \infty$ 并且 $E\xi_1 = 0$.

随后, 在上世纪 70-80 年代, 有大量文献作着关于各种各样的完全收敛性的讨论. 其中一个非常有意思的问题是: (50) 的左边究竟有多大? 它是如何以来 $\varepsilon > 0$ 的? Hedey 在 1980 年考虑了这个问题并证明了下列结论: 假设 $\{\xi_n; n \geq 1\}$ 是个独立同分布的随机序列, $E\xi_1 = 0$, $E\xi_1^2 = \sigma^2 < \infty$, 那么

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \sum_{n=1}^{+\infty} P\left(\omega : \left|\frac{S_n(\omega)}{n}\right| > \varepsilon\right) = \frac{\sigma^2}{2}. \quad (0.45)$$

其证明利用了中心极限定理以及正态分布的尾概率估计. 当然, 可以进一步讨论其它类似的问题. 现在, 文献中称这类问题为精确渐进性.

在 20 世纪, 随机过程理论 (马尔可夫 (Марков) 过程, 平稳过程, Martingales (鞅论), 随机过程的极限定理等) 得到了快速的发展. 另外, 还有许多分支, 比如 (排名不分先后) 随机微分方程、随机偏微分方程、倒向随机微分方程、随机微分几何、Malliavin 变分、白噪声分析、狄氏型理论、遍历理论、数理金大偏差理论、交互粒子系统、测度值过程、概率不等式、泛函不等式、渗流、最有传输、SLE、随机矩阵、随机优化、随机控制、随机动力系统等众多概率论、随机分析及相关领域中的分支得到了快速的发展.

随机过程的迅猛发展是概率论学科乃至整个数学在 20 世纪取得的最大成就之一. Wiener 过程、Guass 过程、Марков 过程、Lévy 过程、鞅 (martingales) 等都是概率论学家最为熟悉的概念. 在这些过程的研究中, 有关独立随机变量序列的部分和的各种经典概率极限定理得到了广泛的应用和推广. 著名的结果包括 Колмогоров-Smirnov 定理、Donsker 弱不变原理和 Skorohod 强不变原理.

在数理统计中, 总体分布 $F(x)$ 往往并不知道; 人们通常用经验分布 $F_n(x)$ 做统计推断. 对每个固定的 x , 经典的大数律和中心极限定理表明: $F_n(x) \rightarrow F(x)$, a.s. 并且, $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, F(x)(1 - F(x)))$. 但逐点收敛并不能很好地描述概率分布 $F(x)$ 的整体性质, 人们需要一致收敛性. Гливі́нко (Glivenko)-Cantelli 定理证明了 $\sup_x |F_n(x) - F(x)| \rightarrow 0$, a.s.; 进而, Колмогоров 和 Smirnov 给出了其渐进分布. 具体地说, $(F_n(x) - F(x), -\infty < x < \infty)$ 作为随机过程序列依分布收敛到 Brown 桥, 并且

$$\lim_{n \rightarrow \infty} P\left(\sup_{-\infty < x < \infty} \sqrt{n}(F_n(x) - F(x)) > t\right) = e^{-2t^2} \quad (0.46)$$

和

$$\lim_{n \rightarrow \infty} P\left(\sup_{-\infty < x < \infty} \sqrt{n}|F_n(x) - F(x)| > t\right) = 2 \sum_{k=1}^{+\infty} (-1)^{k-1} e^{-2k^2 t^2}. \quad (0.47)$$

这些结果可以用来检验总体分布 $F(x)$. 因此, 通常称 $\sqrt{n}|F_n(x) - F(x)|$ 为 Колмогоров-Smirnov 检验统计量.

总所周知, 经典随机游走 (随机徘徊) 和随机过程有着简单且自然的联系. 假设 $\{\xi_n; n \geq 1\}$ 是一列独立同分布的随机变量, $P(\xi_n = \pm 1) = 1/2$. 令 $S_0 = 0$, $S_n = \sum_{j=1}^n \xi_j$, 那么 (S_0, S_1, S_2, \dots) 构成一个随机过程. 事实上, 它是 Марков 过程、独立增量过程, 还是鞅. 它成为了研究随机过程各种性质的基本例子. 如果在时间-位置坐标系中描点, 便得到随机游走的轨迹. 如果限于前 n 个时刻, 那么获得 2^n 条不同的轨迹, 每条轨迹出现的概率均是 $1/2^n$. 如果把相邻的点连起来, 便得到连续的轨迹. 定义部分和过程

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j + \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}, \quad 0 \leq t \leq 1. \quad (0.48)$$

显然, 根据 Feller-Lévy 中心极限定理知道, 对每个 t 成立 $X_n(t) \rightarrow \mathcal{N}(0, t)$. 进而, 作为随机过程来说, $(X_n(t), 0 \leq t \leq 1)$ 依分布收敛. 特别, Donsker 证明了下列结果

$$X_n \Rightarrow W, \quad n \rightarrow \infty, \quad (0.49)$$

其中, $W = (W(t), 0 \leq t \leq 1)$ 为标准 Wiener 过程. 正如中心极限那样, 这一结果不仅仅对简单随机游动成立, 而且可以推广到一般的随机变量序列, 并成为 Donsker 不变原理. 它的重要性体现在两方面: (1) 证明了 Brown 运动的存在性; (2) 结合依分布收敛的连续性, 得到了随机序列部分和产生的各种统计量的极限定理和渐进分布, 从而在数理统计、计量经济、金融数学等学科中有广泛应用.

上述 Колмогоров-Smirnov 定理和 Donsker 不变原理成为了研究一般拓扑空间和距离空间上概率测度弱收敛的基本例子和主要动机. 根本样本轨迹的正则性和有界性, 可以把随机过程看成某个函数空间上的随机元 (r.e.). 例如, 部分和过程可以看成 $C([0, 1])$ 上的随机元, 经验过程可以看作 $D([0, 1])$ 空间上的随机元. 前者赋有一致拓扑, 后者赋有 Skorohod 拓扑. 距离空间上的概率测度弱收敛定义如下:

假设 (S, ρ) 为距离空间, $P, P_n, n \geq 1$ 为 S 上的一列概率测度, 如果对每一有界连续函数 f ,

$$\int_S f(s) dP_n \rightarrow \int_S f(s) dP, \quad (n \rightarrow \infty), \quad (0.50)$$

那么称 P_n 弱收敛到 P , 记作 $P_n \Rightarrow P$. 这一概念是分布函数弱收敛的推广. 一个自然的问题是, 如何判别概率测度弱收敛呢? 通常分两步: (1) 概率测度序列弱相对紧, (2) 所有子序列的极限测度都相同. Prohorov 定理对于验证概率测度序列弱相对紧起着重要作用. 假设 Π 是 S 上一族概率测度, 如果对任意 $\varepsilon > 0$, 存在一个紧子集 K , 使得

$$\sup_{P \in \Pi} P(K^c) < \varepsilon, \quad (0.51)$$

则称 Π 是一致胎紧的 (uniformly tight). Prohorov 定理给出概率测度族弱相对紧的充分条件: 如果 Π 是一致胎紧的, 那么 Π 是弱相对紧的. 其实, 在可分完备距离空间上, 弱相对紧的概率测度族是一致胎紧的. 该定理的兴趣在于: 它将概率测度族的弱收敛性和距离空间的紧致子集的刻画联系起来. 有了 Prohorov 定理, 不同距离空间上的概率测度弱收敛有着各具特色的判别法则.

自 1960 年以来, Hilbert 空间值和 Banach 空间值随机变量的概率极限理论逐渐发展起来. 许多实数值随机变量的极限定理都被推广到 Banach 空间值随机变量的情形, 并获得新的结果. 特别有趣的成果是概率极限理论可以研究 Banach 空间的局部几何结构, 例如 p -型和 q -余型空间. 为了研究 Banach

空间值随机变量的概率极限定理, 人们建立了许多新型的概率不等式, 如 Hoffman-Jørgensen 不等式, Rosenthal 矩不等式, Talagrand 等周不等式. 这些不等式甚至对实值随机变量的研究都有巨大的帮助. 1991 年, 由 Ledoux 和 Talagrand 编著的《Probability in Banach Spaces》总结了 90 年代以前的主要研究成果, 是 Banach 空间上概率论的经典著作.

1970 年代, 匈牙利 Major 等学者利用 Skorohod 嵌入定理, 建立了与弱不变原理 (56) 相应的强不变原理: 假设 $\{\xi_n; n \geq 1\}$ 是概率空间 (Ω, \mathcal{F}, P) 上的一列独立同分布的随机变量, $E\xi_1 = 0$, $E\xi_1^2 = 1$. 记 S_n 是其部分和, 那么可以构造一个新的概率空间 $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, 在其上存在一个 Wiener 过程 W 和一系列独立同分布的随机变量序列 $\{\tilde{\xi}_n; n \geq 1\}$, 使得 $\{\tilde{S}_n\}$ 和 $\{S_n\}$ 同分布, 且

$$\lim_{n \rightarrow \infty} \frac{|\tilde{S}_n - W(n)|}{\sqrt{n \ln \ln n}} \rightarrow 0, \text{ a.s.} \quad (0.52)$$

随机过程的样本轨道性质的研究自 1930 年开始. Lévy 在 1937 年首先讨论了 Brown 运动样本曲线的连续大小, 并证明

$$\limsup_{\delta \rightarrow 0} \sup_{0 \leq t_2 - t_1 \leq \delta} \frac{W(t_2) - W(t_1)}{\sqrt{-2\delta \ln \delta}} = 1, \text{ a.s.}, \quad (0.53)$$

其中, $W = (W(t), t \geq 0)$ 是标准 Brown 运动. 该结果精确地刻画了 Wiener 过程样本曲线的不正则性. 后来, 有关 Wiener 过程和 Guass 过程样本曲线的连续模大小成为了一个当时的热门课题, 吸引了许多的概率论学者的关注. 特别, 以 Csörgö 和 Révész 为代表的匈牙利学派在该领域做了大量工作, 完整且清晰地刻画了一大类 Guass 过程的样本曲线性质.

第一部分

测度论

第一章 测度空间与概率空间

简言之, 测度论可以理解为再抽象空间上建立类似于实变函数中测度、积分和导数那样的分析系统. 建立近代测度与积分理论的途径大致有两条. 一条是从简单图形的测度 (如矩体的体积) 出发, 逐步把测度的定义域扩张称包含所有”可测集”的一个集合类. 然后在测度理论的基础上建立积分的理论. 这是 Lebesgue 和 Carathéodory 等人的方法. 另一条是把把积分看成一类常见函数构成的线性空间上的具有某种性质的”线性泛函”, 然后逐步把线性泛函的定义域扩张成包含所有”可积函数”组成的线性空间. 这是 F.Riesz, Daniell 和 Kakutani 等人的方法.

概率论的传统是使用第一条途径, 因为先引进概率后引进期望在直观上较易接受. 但是, 值得强调的是, 第二条途径其实更加简便, 并且对泛函分析的学习更方便一些. 对这有兴趣可以去读严加安先生的《测度论讲义》.

1.1 可测空间

1.1.1 集合及其运算

集合是数学中最原始的概念. 若要给它下定义, 不得不引入新的概念来说明它. 若要给这些新的概念下定义, 又不得不引入另外的新的概念. 这样就会导致无休止的讨论. 因此, 在直观的朴素集合论 (naive set theory) 中, 集合被看成是无需下定义的基本概念. 为了方便, 我们愿意给集合概念一个直观的描述, 但这不是给它下定义. 使用朴素集合论有时会导致悖论, 因而后来有了以 ZF (Zermelo-Fraenkel set theory) 为代表的公理化集合论. 如果对集合论有兴趣, 请见本书附录. 在此, 应当相信我们所提到的构造集合的方法都不会导致悖论. 集合论中涉及数学基础的那些深层问题, 也不会自己跳出来颠覆人们所发展的概率论方法.

考虑一个任意非空集合 Ω , 称之为**样本空间**. Ω 的子集以大写英文字母 A, B, C, \dots 等记之, 称之为这个样本空间的**集合**. 元素 x 属于 (belong to) 集合 A , 记作 $x \in A$; 反之, x 不属于集合 A , 则用记号 $x \notin A$ 来表示. Ω 上定义的实函数

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases} \quad (1.1)$$

称为 A 的**示性函数**(indicator function). 集合

$$A^c \triangleq \{x : x \notin A\} \quad (1.2)$$

称为 A 的**余**. 如果

$$x \in A \Rightarrow x \in B \quad (1.3)$$

则说集合 A **包含于**(be included in) B , 或集合 B **包含**(include) 集合 A , 记为 $A \subset B$ 或 $B \supset A$. 如果 $A \subset B$ 且 $B \subset A$, 则称集合 A 等于集合 B , 记为 $A = B$.

给定集合 A 和 B , 集合

$$\begin{aligned} A \cup B &\triangleq \{x : x \in A \text{ 或 } x \in B\}, & AB &\triangleq \{x : x \in A \text{ 且 } x \in B\}, \\ A \setminus B &\triangleq \{x : x \in A \text{ 且 } x \notin B\}, & A \Delta B &\triangleq (A \setminus B) \cup (B \setminus A) \end{aligned} \quad (1.4)$$

分别称为集合 A 与 B 的**并**、**交**、**差**、**对称差**. 如果 $B \subset A$, 则还说 $A \setminus B$ 为 A 与 B 的**真差**.

集合的并与交的运算满足**交换律**和**结合律**, 还满足下面两个分配律

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C); \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C). \quad (1.5)$$

如果两个集合 A 与 B 满足 $AB = \emptyset$, 则称它们为**不交的**.

并和交的概念可以推广到任意多个集合的情形. 对一族集合 $\{A_t, t \in T\}$ (T 表示一个集合, 它的元素用 t 表示. $\{A_t, t \in T\}$ 意味着每一个 T 中的元素 t , 都对应着 Ω 中的一个集合 A_t), 集合

$$\bigcup_{t \in T} A_t \triangleq \{x : \exists t \in T, \text{ 使得 } x \in A_t\} \quad (1.6)$$

称为它们的**并**; 集合

$$\bigcap_{t \in T} A_t \triangleq \{x : \forall t \in T, x \in A_t\} \quad (1.7)$$

称为它们的**交**. 如果对任何 $s, t \in T$, 均有 $A_s A_t = \emptyset$, 那么称这族集合 $\{A_t, t \in T\}$ 是**两两不交的**. 注意, 反映并和交运算之间关系的有下列的 De-Morgan 法则:

$$\left(\bigcup_{t \in T} A_t \right)^c = \bigcap_{t \in T} A_t^c; \quad \left(\bigcap_{t \in T} A_t \right)^c = \bigcup_{t \in T} A_t^c. \quad (1.8)$$

设 $\{A_n, n = 1, 2, \dots\}$ 是一个集合序列. 如果对每个 $n = 1, 2, \dots$, 有

$$A_n \subset A_{n+1}, \quad (1.9)$$

则称 $\{A_n\}$ 是**非降**的, 记为 $A_n \uparrow$, 并把集合 $\lim_{n \rightarrow \infty} A_n \triangleq \bigcup_{n=1}^{\infty} A_n$ 叫做它的**极限**; 如果对每个 $n = 1, 2, \dots$, 有

$$A_n \supset A_{n+1}, \quad (1.10)$$

则称 $\{A_n\}$ 是**非增**的, 记为 $A_n \downarrow$, 并称 $\lim_{n \rightarrow \infty} A_n \triangleq \bigcap_{n=1}^{\infty} A_n$ 为它的**极限**. 非降或非增的集合序列统称为**单调序列**. 因此, **单调序列总有极限**. 对于任意给定的一个集合序列 $\{A_n, n = 1, 2, \dots\}$, 集合序列

$$\left\{ \bigcap_{k=n}^{\infty} A_k, n = 1, 2, \dots \right\} \text{ 和 } \left\{ \bigcup_{k=n}^{\infty} A_k, n = 1, 2, \dots \right\} \quad (1.11)$$

分别是非降的和非增的, 因而分别有极限

$$\liminf_{n \rightarrow \infty} A_n \triangleq \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \text{ 和 } \limsup_{n \rightarrow \infty} A_n \triangleq \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad (1.12)$$

我们把 $\liminf_{n \rightarrow \infty} A_n$ 和 $\limsup_{n \rightarrow \infty} A_n$ 分别叫做 $\{A_n, n = 1, 2, \dots\}$ 的**下极限**和**上极限**. 显然, 记号 $\omega \in \limsup_{n \rightarrow \infty} A_n$ 意味着元素 ω 属于序列 $\{A_n, n = 1, 2, \dots\}$ 的无穷多个集合, 而记号 $\omega \in \liminf_{n \rightarrow \infty} A_n$ 则表明除去 $\{A_n, n = 1, 2, \dots\}$ 中的有限个集合外, 元素 ω 属于该序列的其余集合. 于是我们有

$$\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n. \quad (1.13)$$

如果 $\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$, 我们将认为 $\{A_n, n = 1, 2, \dots\}$ 的**极限存在**, 并把

$$\lim_{n \rightarrow \infty} A_n \triangleq \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n \quad (1.14)$$

称为它的**极限**.

1.1.2 集类

以空间 Ω 中的一些集合为元素组成的集合称为 Ω 上的**集合类**. 换句话说, 集合类就是由集合组成的集合. 集合类一般用花体字母 $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{G}, \mathcal{H}, \dots$ 来表示. 为什么不仅要讨论集合, 还要讨论集合类呢? 道理和实分析中的情形一样: 为了研究一般集合的测度理论, 也就是为了建立测度, 需要先明确研究对象和研究范围. 即必须确定出一些**可测集**, 而这些可测集的全体就组成了一个集合类. 在抽象空间中确定可测集时可能用到的集合类有 π 类、半环、半代数、环、代数、 σ 环、 σ 代数、单调类、 λ 类等. 这些集合类中最重要的是 σ 代数, 下面一一介绍这些集合类.

定义 1.1.1 (π 类) 如果 Ω 上的非空集合类 \mathcal{P} 对交运算封闭, 即

$$A, B \in \mathcal{P} \implies AB \in \mathcal{P}. \quad (1.15)$$

则称 \mathcal{P} 为一个 π 类.

例 1.1.1 $\mathcal{P}_{\mathbb{R}} = \{(-\infty, a] : a \in \mathbb{R}\}$ 对有限交的运算封闭, 因而组成实数空间 \mathbb{R} 上的 π 类.

定义 1.1.2 (半环) 满足下列条件的 π 类 \mathcal{Q} 称为半环: 对任意的 $A, B \in \mathcal{Q}$ 且 $A \supset B$, 存在有限个两两不交的 $\{C_k \in \mathcal{Q}, k = 1, \dots, n\}$, 使得

$$A \setminus B = \bigcup_{k=1}^n C_k. \quad (1.16)$$

例 1.1.2 容易看出, 由实数轴 \mathbb{R} 上的开区间全体组成的集合类、左开右闭区间全体组成的集合类、左闭右开区间全体组成的集合类和闭区间全体组成的集合类都是 π 类. 记由全体左开右闭区间全体组成的集合为

$$\mathcal{Q}_{\mathbb{R}} \triangleq \{(a, b] : a, b \in \mathbb{R}\}, \quad (1.17)$$

则对任何 $(a, b], (c, d] \in \mathcal{Q}_{\mathbb{R}}$, 容易验证 $(a, b] \setminus (c, d]$ 可表成 $\mathcal{Q}_{\mathbb{R}}$ 中至多两个不交集合之并. 因此, 它是 \mathbb{R} 上的半环.

例 1.1.3 如果 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ 是由有限个元素组成的集合, 则有 Ω 上的所有单点集组成的集合类 $\mathcal{P} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \dots, \{\omega_n\}\}$ 是一个 π 类, 也是一个半环.

定义 1.1.3 (半(集)代数) 设 Ω 是给定的一个非空集合, 它的一些子集组成的集类 \mathcal{S} 称为 Ω 的一个半代数, 如果它满足

- (i) $\Omega \in \mathcal{S}$;
- (ii) (是 π 类) 若 $A, B \in \mathcal{S}$, 则 $AB \in \mathcal{S}$;
- (iii) (真差可分割) 若 $A, A_1 \in \mathcal{S}$, 并且 $A_1 \subset A$, 则存在 $A_j \in \mathcal{S}$, 使得 $A_1, \dots, A_n \in \mathcal{S}$ 互不相交, 且 $A = \bigcup_{j=1}^n A_j$.

半集代数由较好的集合运算结构, 它构成了测度理论的基本研究对象和研究范围, 这个结构保证了交运算封闭性, 以及可分割性: 大的研究对象可以分割成小的研究对象之并. 由下面的引理可以得到半集代数的一个等价定义.

引理 1.1.1 Ω 的集类 \mathcal{S} 为半集代数的充分必要条件是满足:

- (i) $\Omega \in \mathcal{S}$;
- (ii) (是 π 类) 若 $A, B \in \mathcal{S}$, 则 $AB \in \mathcal{S}$;
- (iii)' 若 $A \in \mathcal{S}$, 则存在 $A_j \in \mathcal{S}$, 使得 $A_1, \dots, A_n \in \mathcal{S}$ 互不相交, 且 $A^c = \bigcup_{j=1}^n A_j$.

证明. 只需要在 (i) 和 (ii) 成立的前提下证明 (iii) 和 (iii)' 等价.

假设 (iii) 成立, 往证 (iii)' 成立. 事实上, 当 $A \in \mathcal{S}$, 注意到 $A \subset \Omega \in \mathcal{S}$, 由 (iii) 知存在 $A_j \in \mathcal{S}$, 使得

$$A = A_0, A_1, \dots, A_n \in \mathcal{S} \quad (1.18)$$

互不相交, 且 $\Omega = \bigcup_{j=0}^n A_j$, 即 $A^c = \bigcup_{j=1}^n A_j$.

假设 (iii)' 成立, 往证 (iii) 成立. 若 \mathcal{S} 中集 A 和 A_1 满足条件 $A_1 \subset A$, 则由 (iii)' 知存在 $B_j \in \mathcal{S}$, 使得 B_2, B_3, \dots, B_n 互不相交, 且 $A_1^c = \bigcup_{j=2}^n B_j$, 因此

$$\begin{aligned} A &= A\Omega = A(A_1 \cup A_1^c) \\ &= (AA_1) \cup \left(\bigcup_{j=2}^n (AB_j) \right) = \bigcup_{j=1}^n A_j, \end{aligned} \quad (1.19)$$

其中 $A_j = AB_j$, $2 \leq j \leq n$. 显然 $A_1, A_2, \dots, A_n \in \mathcal{S}$ 互不相交, 因此 (iii) 成立. \square

例 1.1.4 $\Omega = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$, 试证明

$$\mathcal{S} = \{\mathbb{Z}_+ \cap [a, b) : a \in \mathbb{R}^1, a \leq b \in \mathbb{R}^1 \cup \{+\infty\}\} \quad (1.20)$$

为半集代数.

证明. 显然, $\Omega, \emptyset \in \mathcal{S}$, 并且 \mathcal{S} 对于交运算封闭. 若

$$A_1 = \mathbb{Z}_+ \cap [a_1, b_1) \subset A = \mathbb{Z}_+ \cap [a, b) \quad (1.21)$$

取

$$A_1 = \mathbb{Z}_+ \cap [a, a_1), \quad A_3 = \mathbb{Z}_+ \cap [b_1, b) \quad (1.22)$$

则 A_1, A_2, A_3 互不相交, 且 $A = A_1 \cup A_2 \cup A_3$, 即 \mathcal{S} 为半集代数. \square

例 1.1.5 $\Omega = \mathbb{R}^1$, 试证明

$$\mathcal{S} = \{(a, b] : a \in \mathbb{R}, a \leq b \in \mathbb{R}^1 \cup \{+\infty\}\} \quad (1.23)$$

为半集代数.

证明. Omitted. \square

定义 1.1.4 (环) 非空集合类 \mathcal{R} 称为一个**环**, 如果它对对并和差的运算是**封闭的**, 即若 $A, B \in \mathcal{R}$, 则 $A \cup B, A \setminus B \in \mathcal{R}$,

例 1.1.6 不难验证

$$\mathcal{R}_{\mathbb{R}} = \left\{ \bigcup_{k=1}^n (a_k, b_k] : a_k, b_k \in \mathbb{R}, n \in \mathbb{N}^* \right\} \quad (1.24)$$

是 \mathbb{R} 上的环.

例 1.1.7 由有限个元素组成的集合 Ω 的一切子集组成的集合类 \mathcal{S} 形成一个环.

引理 1.1.2 非空集合类 \mathcal{R} 是一个环的充分必要条件是: 若 $A, B \in \mathcal{R}$, 则 $AB, A \triangle B \in \mathcal{R}$.

证明. 设 \mathcal{R} 是一个环, 则

$$\begin{aligned} A \triangle B &= (A \setminus B) \cup (B \setminus A) \in \mathcal{R}, \\ AB &= (A \cup B) \setminus (A \triangle B) \in \mathcal{R}. \end{aligned} \quad (1.25)$$

另一方面, 设 \mathcal{R} 满足引理中的条件, 则

$$\begin{aligned} A \setminus B &= A(A \triangle B) \in \mathcal{R}, \\ A \cup B &= (AB) \triangle (A \triangle B) \in \mathcal{R}. \end{aligned} \quad (1.26)$$

因此, \mathcal{R} 是一个环当且仅当 $A, B \in \mathcal{R} \Rightarrow AB, A \triangle B \in \mathcal{R}$. \square

定理 1.1.1 设 \mathcal{R} 是一个环, 并定义加法 $A + B \triangleq A \triangle B$ 和乘法 $A \cdot B \triangleq AB$, 则 $(\mathcal{R}, +, \cdot)$ 构成一个抽象代数中的“环”.

证明. 须证: $(\mathcal{R}, +)$ 构成一个 Abel 群, \mathcal{R} 对乘法满足结合律, 对加法和乘法满足分配律.

容易按照集合的运算规律逐一验证:

- $(\mathcal{R}, +)$ 满足结合律: $(A \triangle B) \triangle C = A \triangle (B \triangle C)$;
- 存在幺元 \emptyset , 使 $\forall A \in \mathcal{R}, A \triangle \emptyset = \emptyset \triangle A = A$;

- 对每个 $A \in \mathcal{R}$, 存在加法的逆元 $B = A$, 使 $A \triangle B = B \triangle A = \emptyset$;
- \mathcal{R} 对加法有交换律 $A \triangle B = B \triangle A$;
- \mathcal{R} 对乘法满足结合律: $AB = BA$;
- \mathcal{R} 中的加法对乘法有分配律:

$$A(B \triangle C) = (AB) \triangle (AC); (B \triangle C)A = (BA) \triangle (CA). \quad (1.27)$$

□

我们需要更好的集合运算结构, 需要一个对于集合的交、并、差和补这些运算封闭的一个结构, 所以引入下面的集代数.

定义 1.1.5 ((集) 代数) Ω 的非空子集类 \mathcal{A} 称为 Ω 的一个集代数, 如果它满足

- (i) $\Omega \in \mathcal{A}$;
- (ii) (是 π 类) 若 $A, B \in \mathcal{A}$, 则 $AB \in \mathcal{A}$;
- (iii) (对补运算封闭) 若 $A \in \mathcal{A}$, 则 $A^c \in \mathcal{A}$.

下面给出几个集代数的等价定义

引理 1.1.3 对于 Ω 上的集类 \mathcal{A} , 下列条件都是 \mathcal{A} 为集代数的充分必要条件:

- (I) 定义 1.1.5 中的 (i) 和 (iii) 成立, 并且满足 (ii)' (对并运算封闭) 若 $A, B \in \mathcal{A}$, 则 $A \cup B \in \mathcal{A}$;
- (II) 定义 1.1.5 中的 (i) 成立, 并且满足 (iv) (对差运算封闭) 若 $A, B \in \mathcal{A}$, 则 $A \setminus B \in \mathcal{A}$.

证明. 在定义 1.1.5 中的 (i) 和 (iii) 成立的情况下, 由集合运算的对偶公式

$$(AB)^c = A^c \cup B^c, (A \cup B)^c = A^c \cap B^c \quad (1.28)$$

知定义 1.1.5 中的 (ii) 与 (ii)' 等价. 因此 (I) 是集类 \mathcal{A} 为代数的充分必要条件.

当定义 1.1.5 中的 (i) 成立, 并且满足 (iv) 时, 有

$$\begin{aligned} A^c &= \Omega \setminus A \in \mathcal{A}, \forall A \in \mathcal{A}, \\ AB &= A \setminus B^c \in \mathcal{A}, \forall A, B \in \mathcal{A}. \end{aligned} \quad (1.29)$$

即 \mathcal{A} 为集代数; 反之, 当 \mathcal{A} 为集代数时, 定义 1.1.5 中的 (i)、(ii) 和 (iii) 成立. 由 (ii) 和 (iii) 知

$$A \setminus B = AB^c \in \mathcal{A}, \forall A, B \in \mathcal{A}. \quad (1.30)$$

综上所述, (I) 和 (II) 都是集类 \mathcal{A} 为集代数的充分必要条件. □

† 有的文献中, 也把代数叫做域.

引理 1.1.4 半环必是 π 类, 半代数必是半环, 环必是半环, 代数必是环.

证明. 根据定义, 半环必然满足 π 类的条件; 半代数也满足半环的条件.

若 \mathcal{R} 是一个环, 则

$$\begin{aligned} A, B \in \mathcal{R} &\Rightarrow A \cup B, A \setminus B, B \setminus A \in \mathcal{R} \\ &\Rightarrow A \triangle B = [(A \setminus B) \cup (B \setminus A)] \in \mathcal{R} \\ &\Rightarrow AB = (A \cup B) \setminus (A \triangle B) \in \mathcal{R}, \end{aligned} \quad (1.31)$$

由此可见, 环必是半环.

又设 \mathcal{A} 是一个代数, 则

$$\begin{aligned} A, B \in \mathcal{A} &\Rightarrow A \cup B = (A^c \cap B^c)^c \in \mathcal{A} \\ A, B \in \mathcal{A} &\Rightarrow A \setminus B = AB^c \in \mathcal{A} \end{aligned} \quad (1.32)$$

因而代数必是环. □

可以看到, 前面的几个集合运算结构都只是对于集合的有限次运算封闭, 但是不能保证它对于可数次集合运算封闭. 对于建立测度来说, 只有有限运算是不够的. 因此, 还必须引进一些在可列运算下的集合类.

定义 1.1.6 (单调类) 集合类 \mathcal{M} 称为一个**单调类**, 如果对 \mathcal{M} 中的任何单调序列 $\{A_n, n = 1, 2, \dots\}$, 均有 $\lim_{n \rightarrow \infty} A_n \in \mathcal{M}$.

定义 1.1.7 (λ 类) 集合类 \mathcal{L} 称为一个 **λ 类**, 如果它满足下列条件:

- (i) $\Omega \in \mathcal{L}$;
- (ii) (对真差封闭) 当 $A, B \in \mathcal{L}$ 且 $A \subset B$ 时有 $B \setminus A \in \mathcal{L}$;
- (iii) (对非降列极限封闭) 对于 \mathcal{L} 中的非降序列 $\{A_n, n = 1, 2, \dots\}$, 有 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

例 1.1.8 对于 Ω 上的集类 \mathcal{L} , 以下条件为 \mathcal{L} 为 λ 类的充分必要条件: 定义 1.1.7 中的 (iii) 成立, 并且满足:

- (i)' 若 $A \in \mathcal{L}$, 则 $A^c \in \mathcal{L}$;
- (ii)' 若 $A, B \in \mathcal{L}$, 且 $AB = \emptyset$, 则 $A \cup B \in \mathcal{L}$.

定义 1.1.8 (σ 环) 称非空集合类 \mathcal{R} 是一个 σ 环, 如果

$$\begin{aligned} A, B \in \mathcal{R} &\Rightarrow A \setminus B \in \mathcal{R}; \\ A_n \in \mathcal{R}, n \in \mathbb{N}^* &\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{R}. \end{aligned} \quad (1.33)$$

易见: 一个对可列并运算封闭的环是 σ 环.

定义 1.1.9 (σ 代数) 满足下列三个条件的集合类 \mathcal{F} 称为 σ 代数:

- (i) $\Omega \in \mathcal{F}$;
- (ii) (对补运算封闭) 若 $A \in \mathcal{F}$, 则 $A^c \in \mathcal{F}$;
- (iii) (对可列并运算封闭) 若 $A_m \in \mathcal{F}, n \in \mathbb{N}^*$, 则 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

易见: 一个包括 Ω 的 σ 环是 σ 代数.

有的文献中, 也把 σ 代数称作 σ 域. 有两个很特殊的 σ 代数, 它们分别是 Ω 上含集合最少的 σ 代数 $\{\emptyset, \Omega\}$ 和 Ω 上含集合最多的 σ 代数 $\mathcal{F} \triangleq \{A: A \subset \Omega\}$. 因此, 例 1.7 中的环也是 σ 代数. 但是, 沿例 1.1, 例 1.2, 例 1.6 的那条线出来的 σ 代数会比较复杂, 要在下一节才能说清楚.

引理 1.1.5 \mathcal{F} 是一个 σ 代数的充分必要条件是它满足定义中的 (i)、(ii) 和如下条件 (iii)' 若 $A_n \in \mathcal{F}, n \in \mathbb{N}^*$, 则 $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

证明. 可由集合运算的 De-Morgan 法则 (公式 1.8) 得到. □

这一事实又可进一步推出 $A, B \in \mathcal{F} \Rightarrow AB = A \cap B \in \mathcal{F}$. 因此, σ 代数是代数.

σ 代数的集合运算结构合理, 对于集合的可列次运算封闭, 能够满足具有可列可加性的测度的建立. 下面, 我们来讨论单调类、 λ 类和 σ 代数三者之间的关系.

引理 1.1.6 λ 类是单调类; σ 代数是 λ 类.

证明. 设 \mathcal{L} 是 λ 类, 如果对每个 $n \in \mathbb{N}^*$, 有 $A_n \in \mathcal{L}$ 而且 $A_n \downarrow$, 则由 λ 类的定义知

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{L}. \quad (1.34)$$

这一事实加上 λ 类定义的第二条便说明了 λ 类是单调类.

设 \mathcal{F} 是一个 σ 代数, 则 \mathcal{F} 显然满足 λ 类定义的第一条和第三条. 由于 σ 代数是代数, 所以它也就满足 λ 类定义的第二条, 因此, σ 代数是 λ 类. □

至于什么时候其它的集合类能够称为 σ 代数, 有如下两个定理.

定理 1.1.2 ((集合形式的) 单调类定理) 一个既是单调类又是代数的集合类必是 σ 代数.

证明. 把这个集合类记为 \mathcal{F} . 由于 \mathcal{F} 是代数, 故它对有限并是封闭的. 由因为 \mathcal{F} 是单调类, 所以它对非降序列的极限也是封闭的. 因此,

$$\begin{aligned} A_n \in \mathcal{F}, n \in \mathbb{N}^* &\Rightarrow \bigcup_{k=1}^n A_k \in \mathcal{F}, n \in \mathbb{N}^* \\ &\Rightarrow \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{k=1}^n A_k = \lim_{n \rightarrow \infty} \bigcup_{k=1}^n A_k \in \mathcal{F}, \end{aligned} \quad (1.35)$$

可见, \mathcal{F} 确是 σ 代数. □

† 有的文献中, 也把这些符号记成作 $\mathcal{R}(\mathcal{E}), \mathcal{A}(\mathcal{E}), \mathcal{M}(\mathcal{E}), \mathcal{L}(\mathcal{E})$ or $\Lambda(\mathcal{E})$ or $l(\mathcal{E})$ 和 $\sigma(\mathcal{E})$. 这些都只是记号, 各种形式都有人用, 无特殊意义. 阅读时, 应从上下文通晓其意.

定理 1.1.4 设 \mathcal{S} 是一个半环, 则

$$r(\mathcal{S}) = \left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \text{ 为 } \mathcal{S} \text{ 互不相交集, } n \in \mathbb{N}^* \right\}. \quad (1.37)$$

证明. 记式1.37右端的集合类为 \mathcal{R} . 由于环对于有限并的运算是封闭的, 故 $r(\mathcal{S}) \supset \mathcal{R}$. 因此, 要完成定理的证明, 必须且只需证 $r(\mathcal{S}) \subset \mathcal{R}$. 为此, 又只需证明 \mathcal{R} 为一个环. 若 $A, B \in \mathcal{R}$, 则存在互不相交的 $\{A_i \in \mathcal{S}, i = 1, \dots, n\}$ 和互不相交的 $\{B_j \in \mathcal{S}, j = 1, \dots, m\}$ 使得

$$A = \bigcup_{i=1}^n A_i \quad \text{和} \quad \bigcup_{j=1}^m B_j. \quad (1.38)$$

注意到对每对 (i, j) , 存在 $k_{i,j}$ 个互不相交的集合 $\{C_l^{i,j} \in \mathcal{S}, l = 1, \dots, k_{i,j}\}$, 使得

$$A_i \setminus B_j = A_i \setminus (A_i B_j) = \bigcup_{l=1}^{k_{i,j}} C_l^{i,j}, \quad (1.39)$$

由此, 可把 $A \setminus B$ 表示成 \mathcal{S} 中有限个互不相交的集合的并:

$$\begin{aligned} A \setminus B &= \bigcup_{i=1}^n \bigcap_{j=1}^m [A_i \setminus B_j] \\ &= \bigcup_{i=1}^n \bigcap_{j=1}^m \bigcup_{l=1}^{k_{i,j}} C_l^{i,j} \\ &= \bigcup_{i=1}^n \bigcup_{\substack{l_1=1, \dots, k_{i,1} \\ \vdots \\ l_m=1, \dots, k_{i,m}}} (C_{l_1}^{i,1} \cap \dots \cap C_{l_m}^{i,m}) \end{aligned} \quad (1.40)$$

这表明 \mathcal{R} 对于差运是封闭的. 在此基础上, 又可以把 $A \cup B$ 按下列方式表示成 \mathcal{S} 中有限个互不相交的集合的并:

$$\begin{aligned} A \cup B &= B \cup (A \setminus B) \\ &= \left(\bigcup_{j=1}^m B_j \right) \cup (A \setminus B) \\ &= \left(\bigcup_{j=1}^m B_j \right) \cup \left[\bigcup_{i=1}^n \bigcup_{\substack{l_1=1, \dots, k_{i,1} \\ \vdots \\ l_m=1, \dots, k_{i,m}}} (C_{l_1}^{i,1} \cap \dots \cap C_{l_m}^{i,m}) \right] \end{aligned} \quad (1.41)$$

可见, \mathcal{R} 对有限并也是封闭的. 这个法子, 我们就证明了 \mathcal{R} 确实是一个环. □

定理 1.1.5 设 \mathcal{S} 是一个半集代数, 则

$$a(\mathcal{S}) = \left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \text{ 为 } \mathcal{S} \text{ 上的互不相交集, } n \in \mathbb{N}^* \right\}. \quad (1.42)$$

证明. 记式1.42右端为 \mathcal{A} . 由于代数对于有限并的运算是封闭的, 故 $a(\mathcal{S}) \supset \mathcal{A}$. 因此, 要完成定理的证明, 必须且只需证 $a(\mathcal{S}) \subset \mathcal{A}$. 为此, 又只需证明 \mathcal{A} 为集代数. 显然, $\Omega \in \mathcal{S} \subset \mathcal{A}$. 若 $A, B \in \mathcal{A}$, 则存在互不相交的 $\{A_i \in \mathcal{S}, i = 1, \dots, n\}$ 和互不相交的 $\{B_j \in \mathcal{S}, j = 1, \dots, m\}$ 使得

$$A = \bigcup_{i=1}^n A_i \text{ 和 } B = \bigcup_{j=1}^m B_j. \quad (1.43)$$

从而,

$$AB = \bigcup_{i=1}^n \bigcup_{j=1}^m [A_i B_j] \quad (1.44)$$

注意到互不相交集构成的集类

$$\{A_i B_j : 1 \leq i \leq n, 1 \leq j \leq m\} \subset \mathcal{S} \quad (1.45)$$

可得 $AB \in \mathcal{A}$. 进一步, 由 De-Morgan 律 (式1.8) 得

$$A^c = \bigcap_{i=1}^n A_i^c \quad (1.46)$$

而由 \mathcal{S} 为半代数知道 $A_i^c \in \mathcal{S}$. 因此, $A^c \in \mathcal{S} \subset \mathcal{A}$. 综上所述, \mathcal{A} 为集代数. \square

定理 1.1.6 设 \mathcal{S} 是一个半集代数, 则

$$a(\mathcal{S}) = \left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \in \mathcal{S}, n \in \mathbb{N}^* \right\}. \quad (1.47)$$

证明. 根据定理1.1.5知道

$$a(\mathcal{S}) \subset \left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \in \mathcal{S}, n \in \mathbb{N}^* \right\}. \quad (1.48)$$

另一方面, 对于任何

$$A \in \left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \in \mathcal{S}, n \in \mathbb{N}^* \right\}, \quad (1.49)$$

存在 $A_1, \dots, A_n \in \mathcal{S}$, 使得 $A = \bigcup_{i=1}^n A_i$. 注意到 $\mathcal{S} \subset a(\mathcal{S})$, 且 $a(\mathcal{S})$ 为代数, 可得 $A \in a(\mathcal{S})$, 即

$$\left\{ \bigcup_{j=1}^n A_j : A_1, A_2, \dots, A_n \in \mathcal{S}, n \in \mathbb{N}^* \right\} \subset a(\mathcal{S}), \quad (1.50)$$

所以命题中的等式成立. \square

定理 1.1.7 如果 \mathcal{A} 是一个集代数, 则 $\sigma(\mathcal{A}) = m(\mathcal{A})$.

证明. 由于 $\sigma(\mathcal{A})$ 是一个 σ 代数, 从而也就是一个单调类. 由于 $m(\mathcal{A})$ 是包含 \mathcal{A} 的最小单调类. 所以有

$$\sigma(\mathcal{A}) \supset m(\mathcal{A}). \quad (1.51)$$

往证 $\sigma(\mathcal{A}) \subset m(\mathcal{A})$. 则必须且只需证明 $m(\mathcal{A})$ 为一个 σ 代数. 注意到 $m(\mathcal{A})$ 是一个单调系, 所以又只需证明 $m(\mathcal{A})$ 是一个集代数.

(i) 由于 \mathcal{A} 是集代数, 所以 $\Omega \in \mathcal{A} \subset m(\mathcal{A})$;

(ii) 对任意 $A \in \mathcal{A}$, 令 $\mathcal{G}_A = \{B \in m(\mathcal{A}) : A \setminus B \in m(\mathcal{A})\}$. 则 $\mathcal{A} \subset \mathcal{G}_A$.

下证 \mathcal{G}_A 是个单调类. 设单调列 $\{B_n, n \in \mathbb{N}^*\} \subset \mathcal{G}_A$. 则 $B_n, A \setminus B_n \in m(\mathcal{A})$.

由于 B_n 单调, 所以 $A \setminus B_n$ 也单调. 由于 $m(\mathcal{A})$ 是单调类, 所以

$$\lim_{n \rightarrow \infty} B_n, A \setminus \left(\lim_{n \rightarrow \infty} B_n \right) = \lim_{n \rightarrow \infty} (A \setminus B_n) \in m(\mathcal{A}). \quad (1.52)$$

这说明 $\lim_{n \rightarrow \infty} B_n \in \mathcal{G}_A$. 因此, \mathcal{G}_A 是个单调类. 这进一步说明了 $m(\mathcal{A}) \subset \mathcal{G}_A$.

即: 若 $A \in \mathcal{A}, B \in m(\mathcal{A})$, 则 $A \setminus B \in m(\mathcal{A})$.

再对任意 $B \in m(\mathcal{A})$, 令 $\mathcal{H}_B = \{A \in m(\mathcal{A}) : A \setminus B \in m(\mathcal{A})\}$.

由前面的讨论知道 $\mathcal{A} \subset \mathcal{H}_B$. 下面继续证明 \mathcal{H}_B 也是一个单调类.

任选单调列 $\{A_n, n \in \mathbb{N}^*\} \subset \mathcal{H}_B$. 则 $A_n, A_n \setminus B \in m(\mathcal{A})$. 从而根据 $m(\mathcal{A})$ 为单调类知道

$$\lim_{n \rightarrow \infty} A_n, \left(\lim_{n \rightarrow \infty} A_n \right) \setminus B = \lim_{n \rightarrow \infty} (A_n \setminus B) \in m(\mathcal{A}). \quad (1.53)$$

这说明 $\lim_{n \rightarrow \infty} A_n \in \mathcal{H}_B$. 即 \mathcal{H}_B 也是单调类.

所以有, $m(\mathcal{A}) \subset \mathcal{H}_B$.

$$\forall A, B \in m(\mathcal{A}), \text{ 总有 } A \setminus B \in m(\mathcal{A}). \quad (1.54)$$

根据 (i) 和 (ii) 说明, $m(\mathcal{A})$ 也是一个集代数. 证毕. \square

推论 1.1.1 如果 \mathcal{A} 是集代数, \mathcal{M} 是单调类, 则

$$\mathcal{A} \subset \mathcal{M} \Rightarrow \sigma(\mathcal{A}) \subset \mathcal{M}. \quad (1.55)$$

定理 1.1.8 如果 \mathcal{P} 是一个 π 类, 则 $\sigma(\mathcal{P}) = \lambda(\mathcal{P})$.

证明. 由于 σ 代数是 λ 类. 所以总有 $\sigma(\mathcal{P}) \supset \lambda(\mathcal{P})$.

往证 $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P})$, 则只需证 $\lambda(\mathcal{P})$ 是一个 σ 代数. 由于它本身是一个 λ 类, 所以又只需证明它是一个 π 类.

对任意 $A \in \mathcal{P}$, 令 $\mathcal{G}_A = \{B \in \lambda(\mathcal{P}) : AB \in \lambda(\mathcal{P})\}$. 不难验证: $\mathcal{P} \subset \mathcal{G}_A$ 并且 \mathcal{G}_A 是一个 λ 类. 因此, $\lambda(\mathcal{P}) \subset \mathcal{G}_A$. 这说明 $\forall A \in \mathcal{P}, B \in \lambda(\mathcal{P})$, 有 $AB \in \lambda(\mathcal{P})$.

对任意 $B \in \lambda(\mathcal{P})$, 令 $\mathcal{H}_B = \{A \in \lambda(\mathcal{P}), AB \in \lambda(\mathcal{P})\}$. 根据前面的论证知道 $\mathcal{P} \subset \mathcal{H}_B$. 又不难验证 \mathcal{H}_B 是一个 λ 类, 所以有 $\lambda(\mathcal{P}) \subset \mathcal{H}_B$.

这说明 $\forall A, B \in \lambda(\mathcal{P}), AB \in \lambda(\mathcal{P})$. 即 $\lambda(\mathcal{P})$ 是一个 π 类. 证毕. \square

推论 1.1.2 如果 \mathcal{P} 是 π 类, \mathcal{L} 是 λ 类, 则

$$\mathcal{P} \subset \mathcal{L} \Rightarrow \sigma(\mathcal{P}) \subset \mathcal{L}. \quad (1.56)$$

例 1.1.9 对例 1.4 中的 \mathcal{S} , $\sigma(\mathcal{S})$ 是 \mathbb{Z}_+ 的所有子集全体. 对于例子 1.1 中的 $\mathcal{P}_{\mathbb{R}}$, 称 $\sigma(\mathcal{P}_{\mathbb{R}})$ 是一维 Borel σ 代数记 $\mathcal{O}_{\mathbb{R}}$ 是由 \mathbb{R} 中的开集组成的集合类, 则容易证明 $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{O}_{\mathbb{R}})$. 由此出发, 可将 Borel σ 代数的概念一般化: 对于拓扑空间 (topological space) (X, τ) . 其中 τ 是其所有开集构成的集合类. 我们将把

$$\mathcal{B} \triangleq \sigma(\tau) \quad (1.57)$$

称为拓扑空间 X 上的 Borel σ 代数, 其中的集合称为 X 中的 Borel 集, 而 (X, \mathcal{B}) 就是所谓的拓扑可测空间.

在测度论中, 为了讨论可测函数. 一般还会引入**广义实数集** $\overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{+\infty, -\infty\}$. 今后, $\overline{\mathbb{R}}$ 中的元素将称为**广义实数**. 关于 $\overline{\mathbb{R}}$ 中元素的顺序, 除了实数按原有顺序外, 还规定

$$-\infty < a < +\infty, \quad \forall a \in \mathbb{R}. \quad (1.58)$$

根据这种顺序, 又可定出 $\overline{\mathbb{R}}$ 中的区间: 对任何 $a, b \in \overline{\mathbb{R}}$, 令

$$\begin{aligned} (a, b) &= \{x \in \overline{\mathbb{R}} : a < x < b\}; \\ [a, b) &= \{x \in \overline{\mathbb{R}} : a \leq x < b\}; \\ (a, b] &= \{x \in \overline{\mathbb{R}} : a < x \leq b\}; \\ [a, b] &= \{x \in \overline{\mathbb{R}} : a \leq x \leq b\}; \end{aligned} \quad (1.59)$$

关于 $\overline{\mathbb{R}}$ 中的运算, 规定

$$\begin{aligned} (\pm\infty) + a &= a + (\pm\infty) = a - (\mp\infty) \\ &= \pm\infty, \quad \forall a \in \mathbb{R}; \end{aligned} \quad (1.60)$$

$$(\pm\infty) + (\pm\infty) = (\pm\infty) - (\mp\infty) = \pm\infty; \quad (1.61)$$

$$\begin{aligned} (\pm\infty) \cdot a &= a \cdot (\pm\infty) \\ &= \begin{cases} \pm\infty, & 0 < a \leq +\infty, \\ 0, & a = 0, \\ \mp\infty, & -\infty \leq a < 0. \end{cases} \end{aligned} \quad (1.62)$$

注意: 诸如 $(\pm\infty) - (\pm\infty)$, $(\pm\infty)/(\pm\infty)$, \dots 等是没有定义的. 对 $a \in \overline{\mathbb{R}}$, 记

$$a^+ = \max(a, 0) \quad \text{和} \quad a^- = \max(-a, 0), \quad (1.63)$$

分别把它们叫做 a 的**正部**和**负部**. 易见, $a = a^+ - a^-$. 另外, 还记

$$\mathcal{B}_{\overline{\mathbb{R}}} = \sigma(\mathcal{B}_{\mathbb{R}}, \{-\infty, +\infty\}). \quad (1.64)$$

定理 1.1.9 下列等式成立:

$$\begin{aligned}
 \mathcal{B}_{\mathbb{R}} &= \sigma(\{[-\infty, a] : a \in \mathbb{R}\}) \\
 &= \sigma(\{[-\infty, a) : a \in \mathbb{R}\}) \\
 &= \sigma(\{[a, \infty] : a \in \mathbb{R}\}) \\
 &= \sigma(\{(a, \infty] : a \in \mathbb{R}\})
 \end{aligned} \tag{1.65}$$

1.2 测度与测度的构造

1.2.1 测度、可测空间与测度空间

1.2.2 测度在集代数上的扩张

1.2.3 外测度

1.2.4 从半集代数到代数上的测度扩张

1.3 测度的性质

1.3.1 测度的运算性质

1.3.2 测度空间的完全化

1.3.3 Lebesgue-Stieltjes 测度

第二章 可测函数与随机变量

2.1 可测函数与分布

定义 2.1.1 设 (Ω, \mathcal{F}) 为可测空间, $A \in \mathcal{F}$, 记

$$A \cap \mathcal{F} \triangleq \{AB : B \in \mathcal{F}\} \quad (2.1)$$

如果函数 $f : A \rightarrow \bar{\mathbb{R}}$ 满足条件

$$f^{-1}(B) \triangleq \{\omega \in A : f(\omega) \in B\} \in A \cap \mathcal{F}, \quad \forall B \in \mathcal{B}_{\bar{\mathbb{R}}}, \quad (2.2)$$

其中 $\mathcal{B}_{\bar{\mathbb{R}}} = \sigma(\{\mathcal{B}_{\mathbb{R}}, \bar{\mathbb{R}}\})$, 则称 f 为定义在 A 上的 \mathcal{F} **可测函数**, 或者 **广义随机变量**, 简称为 **可测函数**; 如果 A 上的可测函数 f 满足条件 $f(A) \subset \mathbb{R}$, 则称 f 为 A 上的 **有限实值随机变量**; 将 A 上的有限实值随机变量简称为 **随机变量**, 记为 r.v.

上述概念可以进一步推广到一般映射的情况.

定义 2.1.2 对于映射 $f : \Omega \rightarrow E$, 称

$$f^{-1}(A) \triangleq \{f \in A\} \triangleq \{\omega \in \Omega : f(\omega) \in A\}, \quad \forall A \subset E \quad (2.3)$$

为 A 对 f 的 **原像**, 简记为 $\{f \in A\}$; 称

$$f^{-1}(\mathcal{E}) \triangleq \{f^{-1}(A) : A \in \mathcal{E}\} \quad (2.4)$$

为集类 \mathcal{E} 对 f 的原像, 其中 \mathcal{E} 是 E 的子集类; 如果 (Ω, \mathcal{F}) 和 (E, \mathcal{E}) 为可测空间, f 满足

$$f^{-1}(\mathcal{E}) \subset \mathcal{F} \quad (2.5)$$

则称 f 为 (Ω, \mathcal{F}) 到 (E, \mathcal{E}) 的 \mathcal{F} **可测映射**, 简称为 **可测映射**; 而 $\sigma(f) \triangleq f^{-1}(\mathcal{E})$ 叫做 **使映射 f 可测的最小 σ 代数**. 如果 (Ω, \mathcal{F}) 和 (E, \mathcal{E}) 为可测空间, \mathcal{F} 上有概率测度 \mathbf{P} , 则称映射 f 为 $(\Omega, \mathcal{F}, \mathbf{P})$ 到 (E, \mathcal{E}) 的 **随机元**. 如果 f 是概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 到 $(\mathbb{R}^n, \mathcal{B}^n)$ 的随机元, 则称 f 为 **n 维随机变量**.

定理 2.1.1 集合的原像有以下性质: 设 f 是 Ω 到 E 上的映射, 则有

$$\begin{aligned} f^{-1}(E) &= \Omega, \quad f^{-1}(\emptyset) = \emptyset \\ B_1 \subset B_2 &\Rightarrow f^{-1}(B_1) \subset f^{-1}(B_2) \\ f^{-1}(B^c) &= (f^{-1}(B))^c, \quad \forall B \subset E \end{aligned}$$

$$\begin{aligned}
f^{-1}(A \setminus B) &= (f^{-1}A) \setminus (f^{-1}B), \forall A, B \subset E \\
f^{-1} \bigcup_{t \in T} A_t &= \bigcup_{t \in T} f^{-1}A_t \\
f^{-1} \bigcap_{t \in T} A_t &= \bigcap_{t \in T} f^{-1}A_t,
\end{aligned} \tag{2.6}$$

其中, $\{A_t : t \in T\}$ 为 Y 的子集类.

证明. Omitted. □

引理 2.1.1 设 f 是 $\Omega \rightarrow E$ 的映射, \mathcal{E} 是 E 上的 σ 代数, 则 $f^{-1}(\mathcal{E})$ 是 Ω 的 σ 代数.

证明. 可由定理2.1.1得结论. □

定理 2.1.2 设 f 是 Ω 到 E 上的映射, 对 E 上的非空子集类 \mathcal{E} , 有

$$\sigma(f^{-1}\mathcal{E}) = f^{-1}\sigma(\mathcal{E}). \tag{2.7}$$

证明. 由引理2.1.1易见 $f^{-1}\sigma(\mathcal{E})$ 是一个 σ 代数. 因此 $\sigma(f^{-1}\mathcal{E}) \subset f^{-1}\sigma(\mathcal{E})$. 另一方面, 令

$$\mathcal{G} = \{B \subset E : f^{-1}B \in \sigma(f^{-1}\mathcal{E})\} \tag{2.8}$$

则 \mathcal{G} 是一个 σ 代数, 且 $\mathcal{E} \subset \mathcal{G}$, 因此 $\sigma(f^{-1}\mathcal{E}) \supset f^{-1}\sigma(\mathcal{E})$, 即结论成立. □

2.2 分布与分布函数

2.2.1 可测函数的运算

测度论中的典型方法 在测度论和概率论中, 为了证明一个关于可测函数的命题, 常常分解为如下几个比较容易的步骤进行:

- (1) 证明该命题对最简单的函数——示性函数成立.
- (2) 证明该命题对非负简单函数——示性函数的线性组合成立.
- (3) 证明该命题对非负可测函数——非降的非负简单函数列的极限成立.
- (4) 证明该命题对一般的可测函数——两个非负可测函数, 即它的正部和负部之差成立.

按上述步骤证明命题的方法叫做测度论中的典型方法. 典型方法符合人们的认识过程, 是一种具有普遍意义、行之有效的方法, 必须熟练掌握.

2.3 复合映射的可测性

2.4 可测映射列极限的可测性

2.5 可测函数的构造

第三章 积分与数学期望

3.1 积分的定义

3.2 积分的性质

3.3 空间 $L_p(X, \mathcal{F}, \mu)$

3.4 概率空间的积分

第四章 不定积分与条件期望

4.1 符号测度

4.2 符号测度的分解

4.2.1 Hahn 分解和 Jordan 分解

4.2.2 Radon-Nikodym 定理

4.2.3 Lebesgue 分解

4.3 条件期望与条件概率

第五章 乘积空间

5.1 有限维乘积空间

5.2 多维 Lebesgue-Stieltjes 测度

5.3 可列维乘积空间的概率测度

5.4 任意无穷维乘积空间的概率测度

第六章 离散鞅论

6.1 基本概念

6.2 停时定理

6.3 收敛定理

6.4 鞅的不等式

第二部分

概率极限理论

第七章 大数定律

7.1 零壹律和三级数定理

7.2 弱大数律

7.3 强大数律

7.4 随机序列的收敛性

7.5 更新过程

7.6 大偏差理论

第八章 中心极限定理

8.0.1 De Moivre-Laplace 定理

8.0.2 特征函数

8.0.3 中心极限定理

8.0.4 局部极限定理

8.0.5 Poisson 收敛和 Poisson 过程

第九章 无穷可分分布和普适极限定理

9.1 无穷可分分布函数

9.2 独立随机变量和的分布

9.3 L 族和稳定分布族

第十章 弱收敛理论

10.1 距离空间概率测度的弱收敛

10.2 鞅的极限定理

第十一章 离散鞅论

第三部分

随机过程论

第十二章 Markov 链

第十三章 遍历定理

第十四章 布朗运动

第十五章 Skorokhod 嵌入定理

第十六章 多维布朗运动

第十七章 概率不等式

概率不等式是概率论的理论研究中的重要工具. 有力的概率不等式, 常常是创建一个定理的重要依据. 对于概率的极限定理等, 几乎所有的重要结果的论证或是借助于建立一些概率不等式, 或是通过巧妙应用已有的概率不等式. 在此, 我们将概率论中常见的一些基本不等式分类罗列如下.

17.1 随机事件的不等式

17.2 分布函数、特征函数的不等式

17.3 随机变量的概率不等式

17.4 矩不等式

17.5 随机变量的概率与矩之间的不等式

17.6 部分和及其极值的矩估计

17.7 指数不等式

参考文献

- [1] Gregory Lawler, Introduction to Stochastic Processes , CRC Press. 中译本《随机过程导论》张景肖译, 机械工业出版社, 2010 年
- [2] Geoffrey Grimmett and David Stirzaker, One Thousand Exercises in Probability, Oxford University Press, 2001.
- [3] Athreya, K.B. and Ney, P.E. , Branching Processes., Springer-Verlag, New York, 1972.
- [4] Allen Linda J.S., An Introduction to Stochastic Processes with Applications to Biology. , Pearson Prentice Hall, Upper Saddle River, NJ, 2003.
- [5] Bailey, N.T.J., The Elements of Stochastic Processes: With Applications to the Natural Sciences., John Wiley and Sons, 1964.
- [6] R. Durrett, Probability: Theory and Examples, 5th Edition, Cambridge University Press, 2019.
- [7] R. Durrett, Essentials of Stochastic Processes, Third Edition, Springer-Verlag, New York, 2016.
- [8] 胡迪鹤 高等概率论及其应用, 北京: 高等教育出版社, 2008.
- [9] 严加安, 测度论讲义 (第二版), 北京: 科学出版社, 2005.
- [10] 汪嘉冈 现代概率论基础, 复旦大学出版社, 1988.
- [11] 程士宏, 高等概率论, 北京: 北京大学出版社, 1996.
- [12] 严士健, 王隽骧, 刘秀芳, 概率论基础, 北京: 科学出版社, 1982.
- [13] 严士健, 刘秀芳, 测度与概率, 北京: 北京师范大学出版社, 1994.
- [14] 林正炎, 陆传荣, 苏中根, 概率极限理论基础, 北京: 高等教育出版社, 1999.

附录

附录 1 集合论的温习

在此复习一下集合论的一些简单知识. 这些知识构成了概率论语言的基础. 但是在其它课程对于朴素的集合论已经具有了一定的了解, 因此只将这些介绍蜻蜓点水式地简单回顾. 进一步的资料可以去查找陶哲轩写的 [《Analysis I》](#)

- 集合论自 19 世纪 80 年代由 G.Cantor 创立以来, 现在已经发展成独立的数学分支.
- Russell 悖论 (Russell's paradox 或 Russell's antinomy) 的出现促使数学家产生了许多改进方案, 就是把集合论公理化.
- 最著名是 ZF(Zermelo-Fraenkel) 公理集合论体系, 及其保守扩张 GB(Gödel 以及 Bernays).

定义 17.7.1 (集合) • 集合就是一些东西的总体.

- 总体中的东西称为这个集合的元素
- 元素 ω 是集合 A 的一个元素, 称作元素 ω 属于集合 A , 记作 $\omega \in A$ 或者 $A \ni \omega$.
- 在一个数学问题中, 常有这样一个集合 Ω , 使得我们打算研究的所有对象都是这个特定集合的元素. 这个集合 Ω 常被称为 (这个问题的) 空间/全集 (universal set).

众所周知, 一个集合 (set) 就是把一堆元素 (element) 放在一起当作一个整体来看待. 当 x 是 A 的元素的时候, 我们也称 x 属于 (belong to) A , 记为 $x \in A$, 否则称 x 不属于 A , 记为 $x \notin A$. 当两个集合相等当且仅当所含元素完全相同. 于是证明两个集合 $A = B$ 的标准办法就是去证明

$$x \in A \iff x \in B. \quad (17.1)$$

不含任何元素的集合称为空集 (empty set), 记做 \emptyset .

但并不是说随便把一堆对象拿出来放在一起就可以构成一个集合, 这会导致悖论. 比如著名的 Russell 悖论 (Russell's paradox 或 Russell's antinomy) 是说, 如果定义 X 为“又所有不属于自己的集合构成的集合”, 则从 $X \in X$ 就能推出 $X \notin X$, 反过来从 $X \notin X$ 也能推出 $X \in X$, 这是自相矛盾的. 换言之有些构造集合的方法是安全的, 有些则是不安全的, 为了避免悖论的产生, 无法用安全的方法实现的构造我们都不应当承认它是集合. 下面介绍几种常规 (即安全的) 构造方法.

一种最常用的构造方法是从一个已知的集合中筛选出满足特定条件的元素, 构成新的集合. 已知 X 是一个集合, 设 $P(x)$ 是一个关于 x 的命题, 则 X 中的元素使得命题 $P(x)$ 成立的那些元素 x 就构成一个集合, 记为

$$\{x \in X | P(x)\}. \quad (17.2)$$

有些时候我们打算研究的对象都是某个特定集合 X 的元素, 这时就把 X 称为全集 (universal set). 所以 Russell 悖论和上述合法构造的差别就在于: 悖论里构造的那个东西没有一个全集限定元素的选择

范围.

设 A, B 是两个集合, 如果 $x \in A$ 蕴含 $x \in B$, 则称 A 为 B 的子集 (subset) 或者 A 包含于 (be included in) B , 并称 B 包含 A , 记作 $A \subset B$ 或 $B \supset A$. 如果 $A \subset B$, 并且 $A \neq B$, 则称 A 为 B 的真子集 (proper subset).

X 的全体子集也构成一个集合, 称其为幂集 (power set), 记为 2^X . 用这个记号是因为当 X 是恰好有 n 个元素的有限集时, 2^X 是恰好有 2^n 个元素的有限集.

一个所有元素都是集合的集合称为集合族 (collection of sets). 设 \mathcal{A} 是集合族, 则可以定义 \mathcal{A} 的所有元素 (都是集合) 的交集 (intersection) $\cap \mathcal{A}$ 以及并集 (union) $\cup \mathcal{A}$ 如下:

- $x \in \cap \mathcal{A}$ 当且仅当 x 是每一个集合 $A \in \mathcal{A}$ 的元素;
- $x \in \cup \mathcal{A}$ 当且仅当存在集合 $A \in \mathcal{A}$, 使得 x 是 A 的元素.

这里有个小小的问题: 如果 \mathcal{A} 不含任何元素, 那么如何判断一个元素 x 是不是每个集合 $A \in \mathcal{A}$ 的元素呢? 为此我们规定 $\cap \emptyset = \emptyset$.

如果集合族 \mathcal{A} 只有两个元素 A 和 B , 则交集和并集就记为 $A \cap B$ 和 $A \cup B$. 对于一般的集合族, 人们也往往喜欢弄出一个指标集 Λ , 然后把集合族写成 $\{A_\lambda\}_{\lambda \in \Lambda}$, 并把交集和并集写成 $\bigcap_{\lambda \in \Lambda} A_\lambda$ 和 $\bigcup_{\lambda \in \Lambda} A_\lambda$ 的样子. 请注意, 指标集可以是任何集合, 而不一定是自然数集或其子集, 所以一般来说是不能对指标集应用数学归纳法的.

交和并的运算满足各自的交换律和结合律, 也满足分配律, 即

$$A \cup \left(\bigcap_{\lambda \in \Lambda} B_\lambda \right) = \bigcap_{\lambda \in \Lambda} (A \cup B_\lambda) \quad (17.3)$$

$$A \cap \left(\bigcup_{\lambda \in \Lambda} B_\lambda \right) = \bigcup_{\lambda \in \Lambda} (A \cap B_\lambda) \quad (17.4)$$

两个集合 A, B 的差集 (difference) 定义为

$$A \setminus B = \{x \in A \mid x \notin B\}, \quad (17.5)$$

有时候, 差集也记做 $A - B$.

差集满足 De Morgan 定律 (De Morgan's law)

$$B \setminus \left(\bigcap_{\lambda \in \Lambda} A_\lambda \right) = \bigcup_{\lambda \in \Lambda} (B \setminus A_\lambda) \quad (17.6)$$

$$B \setminus \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right) = \bigcap_{\lambda \in \Lambda} (B \setminus A_\lambda) \quad (17.7)$$

特别地, 如果对于我们研究的问题有一个预先取定的全集 X , 则差集 $X \setminus A$ 称为 A 对全集 X 的余集或补集 (complement), 记为 A^c . 注意, 有些文献会习惯使用 \overline{A} 表示 A 的余集, 但我们如果也使用这个符号, 就会在后面讨论 A 的“闭包”时, 会产生歧义, 因而此处只使用 A^c 来表示 A 的余集.

定义 17.7.2 (映射) • 一个映射 (map 或 mapping) f 是两个集合 X 和 Y 的元素之间的一个对应关系, 使得每个 X 的元素 x 都对应唯一一个 Y 的元素 $f(x)$.

- 我们有时也把映射的定义简写为

$$f : X \rightarrow Y, x \mapsto f(x). \quad (17.8)$$

- 称 X 为 f 的定义域 (domain), Y 为 f 的陪域 (codomain).
- 全体 X 到 Y 的映射也构成一个集合, 记作 Y^X . 用这个记号是因为当 X 有 m 个元素, Y 含 n 个元素时, Y^X 恰好含 n^m 个元素.

† 作为特殊情形, 集合到自身的映射称为**变换**.

注意: 幂集 2^Ω 的元素可以和 $\{0, 1\}^\Omega$ 的元素建立起一一对应关系: 每一子集 $A \subset \Omega$ 对应的映射就定义为

$$\mathbf{1}_A : \Omega \rightarrow \{0, 1\}, \omega \mapsto \begin{cases} 1, & \omega \in A; \\ 0, & \omega \in \Omega \setminus A. \end{cases} \quad (17.9)$$

如果 $y = f(x)$, 则称 y 为 x 的像 (image), 称 x 为 y 的一个原像 (preimage 或 inverse image). 从定义可以看出, 每个 $x \in X$ 只能有一个像, 但是每个 $y \in Y$ 可以有很多原像, 也可以没有原像. y 的所有原像构成的集合记为 $f^{-1}(y)$. 如果 A 是 X 的子集, 则称

$$f(A) = \{y \in Y \mid \text{存在 } x \in A, \text{ 使得 } f(x) = y\} \quad (17.10)$$

为 A 的像集 (image).

如果 B 是 Y 的子集, 则称

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\} \quad (17.11)$$

为 B 的原像集 (preimage 或 inverse image).

不难验证, 原像集满足下述简单的规律:

•

$$f^{-1}\left(\bigcap_{\lambda \in \Lambda} A_\lambda\right) = \bigcap_{\lambda \in \Lambda} f^{-1}(A_\lambda); \quad (17.12)$$

•

$$f^{-1}\left(\bigcup_{\lambda \in \Lambda} A_\lambda\right) = \bigcup_{\lambda \in \Lambda} f^{-1}(A_\lambda); \quad (17.13)$$

•

$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B). \quad (17.14)$$

但是, 像集却完全可能不满足这些规律.

如果任取两个点 $x_1, x_2 \in X, x_1 \neq x_2$ 蕴含 $f(x_1) \neq f(x_2)$, 则称 f 为单射 (injection). 如果任取一个 Y 中的点 $y, f^{-1}(y) \neq \emptyset$, 则称 f 是满射 (surjection). 如果 f 既是单射又是满射, 则称 f 为双射 (bijection) 或一一对应 (one-to-one correspondence). 有时, 也称双射为可逆映射 (invertible map), 因为此时每一个 $y \in Y$ 存在唯一原像, 从而可以定义逆映射 (inverse)

$$f^{-1} : Y \rightarrow X, \quad (17.15)$$

把每个 $y \in Y$ 对应到它关于 f 的那个唯一原像.

任何集合上都可以定义恒同映射 (identity)

$$\text{id}_X : X \rightarrow X, x \mapsto x. \quad (17.16)$$

如果 $A \subset X$, 则可定义含入映射 (inclusion), 也叫包含映射

$$i_A : A \rightarrow X, x \mapsto x. \quad (17.17)$$

若有两个映射 $f : X \rightarrow Y$ 和 $g : Y \rightarrow Z$, 则可以定义复合映射 (composition)

$$g \circ f : X \rightarrow Z, x \mapsto g(f(x)) \quad (17.18)$$

不难验证, 复合应当满足下述两个基本性质:

- 如果 $g \circ f$ 是单射, 则 f 是单射;
- 如果 $g \circ f$ 是满射, 则 g 是满射.

若 X, Y 是两个集合, 则全体有序对 (x, y) (其中 $x \in X, y \in Y$) 也构成一个集合, 称为 X 和 Y 的笛卡儿积 (Cartesian product) 或直积 (direct product), 记作 $X \times Y$ (英文 Cartesian 其实是将 Descartes (笛卡儿) 的名字形容词化之后的结果). 也可以类似地定义有限多个集合的直积 $X_1 \times \cdots \times X_n$, 即所有序列 (x_1, \cdots, x_n) (其中每个 $x_j \in X_j$) 构成的集合, 特别地, n 个 X 的直积记为 X^n , 例如, 实 n 维空间 $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$.

集合上的一个二元运算即是由 $A \times A$ 到 A 的一个映射. A 上的一个二元关系 R 定义为 $A \times A$ 的一个子集. 如果 $(a_1, a_2) \in R$, 就称 a_1 与 a_2 有关系 R , 记为 $a_1 R a_2$.

Cantor 最早发明集合论的时候说得很含糊, 他说集合是“一些确定的、不同的东西的总体”. 他开始想的其实只是实数集 \mathbb{R} 的子集. 当然了, 他更关心的是无穷子集. 他首先发现, 一个集合含有无穷多个元素的充分必要条件是它能和自己的一部分一一对应. 然后他又提出了比较两个无穷集合所含元素的“个数”谁多谁少的方法. 一般集合上这种相当于“个数”的概念称为基数 (cardinality) 或势 (potency). 如果集合 A 和 B 的元素能建立起一一对应, 就称它们的基数相同, 记为

$$\|A\| = \|B\|. \quad (17.19)$$

如果 $\|A\| \neq \|B\|$, 但是 A 能和 B 的一部分一一对应, 则称 A 的基数小于 B 的基数, 记为

$$\|A\| < \|B\|. \quad (17.20)$$

Cantor 很快便证明了 $\|\mathbb{Q}\| = \|\mathbb{N}\|$, 而 $\|\mathbb{R}\| > \|\mathbb{N}\|$. 然后他被一个问题难倒了: 是否存在一个集合 A , 使得 $\|\mathbb{N}\| < \|A\| < \|\mathbb{R}\|$? Cantor 猜想这样的集合不存在, 这个猜想被称为连续统假设 (因为 \mathbb{R} 的

基数被称为连续统),但是他无法证明.

给朴素的集合论想法带来更大打击的是悖论: 不加限制地把“满足某种条件的对象”放在一起, 有的时候构造出来的东西不能当作集合对待, 也就是说, 那些对于集合应该自然成立的结论到这里不一定成立, 否则会导出自相矛盾的结果.

比如 Cantor 就发现, 把所有的对象放在一起后成一个大全集 Ω , Ω 就不能当作集合对待. 因为否则的话, 一方面任何集合的基数都不应该比 Ω 的基数更大; 而另一方面, 如果考虑 Ω 的幂集合 2^Ω , 又可以像证明 $\|\mathbb{R}\| > \|\mathbb{N}\|$ 那样证明 $\|2^\Omega\| > \|\Omega\|$. 这被称为 Cantor 悖论 (Cantor's paradox).

随后 Russell(罗素) 又发现了更简单直接的 Russell 悖论 (Russell's paradox 或 Russell's antinomy): 把所有“不是自己的元素的集合”放在一起构成的那个东西也不能当作集合对待, 因为否则的话记这个集合为 Ω , 则 $(\Omega \in \Omega) \iff (\Omega \notin \Omega)$.

想要避免悖论的发生, 就必须限制构造集合的方法, 需要明确知道那些构造方法是安全的, 只有那些符合安全规范的方法构造出来的对象才全是集合, 才能放心大胆地应用关于集合的那些性质和命题. 像前面零个悖论中的 Ω 就都不算集合了. 那然, 那些最最基础的构造方法是不是安全是没法证明的, 只能把每一条当做一个公理, 这样得到的就是集合论的公理系统. 历史的发展证明, 只要对集合的构造方法加以限制, 集合论就可能成为数学的可靠基础.

最常用的集合论公理系统是 ZF 及其保守扩张 GB(这里 Z,F,G,B 分别是其发明人 Zermelo, Fraenkel, Gödel 以及 Bernays 名字的第一个字母)

ZF 是 Zermelo-Fraenkel 集合论 (Zermelo-Fraenkel set theory) 的简称, 包括九条公理, 每条公理是一个逻辑命题. 注意 ZF 系统的所有命题中提到的对象都可以是集合, 包括其中提到集合的元素时, 这些元素课都可以是集合. 仔细观察以下不难发现, 这些公理中“某某集合存在”表达的其实都是“某某方式构造是对象算是集合”的意思.

外延公理 (axiom of extensionality) 如果 $x \in X$ 当且仅当 $x \in Y$, 则 $X = Y$, 即集合由其元素完全决定

无序对公理 (axiom of pairing) $\forall x, y$, 存在集合 $\{x, y\}$ (无序对), 使得 $z \in \{x, y\}$ 当且仅当 $z = x$ 或 $z = y$.

Rmk: 这里允许 $x = y$, 此时 $\{x, y\} = \{x\}$. 能区分 x, y 地位的有序对则定义为集合

$$(x, y) = \{\{x\}, \{x, y\}\}. \quad (17.21)$$

并集公理 (axiom of union) $\forall \mathcal{A}$, 存在集合 $\cup \mathcal{A}$, 使得 $x \in \cup \mathcal{A}$ 当且仅当存在 $A \in \mathcal{A}$ 满足 $x \in A$.

结合无序对公理和并集公理可以归纳定义无序多元组如下:

$$\{x_1, \dots, x_{n+1}\} = \{x_1, \dots, x_n\} \cup \{x_{n+1}\}. \quad (17.22)$$

幂集公理 (axiom of power set) $\forall X$, 存在集合 2^X (幂集), 使得 $A \in 2^X$ 当且仅当 $A \subset X$. 这里 $A \subset X$ 是逻辑命题

$$\forall x \in A \Rightarrow x \in X \quad (17.23)$$

(即 A 是 X 的子集) 的简写.

分离公理 (axiom schema of separation 或 axiom schema of specification) 任取集合 X 以及关于 x 的命题 $P(x)$, 存在集合 A 使得 $x \in A$ 当且仅当 $x \in X$ 并且 $P(x)$ 成立. 集合 A 通常记为 $\{x \in X | P(x)\}$.

结合分离公理和并集公理, 就可以把交集定义为

$$\cap \mathcal{A} = \{x \in \cup \mathcal{A} | \forall A \in \mathcal{A}, x \in A\}. \quad (17.24)$$

注意到当 $x \in X, y \in Y$ 时, 前面定义的有序对 (x, y) 其实是一个由 $X \cup Y$ 的子集构成的集合, 这样的集合是 $2^{X \cup Y}$ 的子集, 也就是 $2^{(2^{X \cup Y})}$ 的元素, 因此利用分离公理, 可以把直积 (direct product) 定义为

$$X \times Y = \{z \in 2^{(2^{X \cup Y})} | \exists x \in X, y \in Y, \text{使得 } z = (x, y)\}. \quad (17.25)$$

直积也称作笛卡儿积 (Cartesian product), 其中英文 Cartesian 其实是将 Descartes(笛卡儿) 的名字形容词化之后的结果.

每个从 X 射到 Y 的映射 $f: X \rightarrow Y$ 可以用它在 $X \times Y$ 中的函数图像来刻画, 因此也不需要添加新的公理, 就可以把从 X 打到 Y 的映射定义为集合

$$Y^X = \{f \in 2^{X \times Y} | \forall x \in X, \text{存在唯一 } y \in Y \text{ 使得 } (x, y) \in f\} \quad (17.26)$$

的元素. 并且此时如果 $(x, y) \in f$, 则把 y 记为 $f(x)$.

空集公理 (axiom of empty set) 存在集合 \emptyset (空集), 使得 $\forall x, x \notin \emptyset$.

由外延公理可以知道空集是唯一的. 你也许会认为空集公理是多余的, 因为有了分离公理可以随便取个集合, 再取一个该集合中元素永远不满足的性质, 然后就可以构造出空集来了, 比如说 $\emptyset = \{x \in A | x \notin A\}$. 很多的书籍上也确实是这么写的, 但是这实际上依赖于另外的一条其它公理中没有提到的假设: 在这个世界上确实至少存在着那么一个集合 A . 缺少这个假设, 形式逻辑的推理过程就没有了起点. 当然, 也有群体认为下面的一条公理“无穷公理”本身就包含了一定存在一个空集的意思.

无穷公理 (axiom of infinity) 存在一个集合 ω (含有无穷多个元素的集合), 使得 $\emptyset \in \omega$, 并且 $x \in \omega$ 蕴含 $x \cup \{x\} \in \omega$.

在集合论中, 每个自然数 n 被归纳地定义成一个恰好有 n 个元素的特殊集合:

$$0 = \emptyset, 1 = \{0\}, \dots, n+1 = \{0, \dots, n\} = n \cup \{n\}. \quad (17.27)$$

比较一下这个定义和上述无穷公理不难看出, 自然数集 \mathbb{N} 就是满足无穷公理的最小集合. 有趣的是关于自然数的加减乘除运算的各种基本性质都可以从这个定义以及 ZF 系统的其它公理推导出来. 有了自然数之后当然也可以定义整数和有理数, 并把数学分析中定义实数的方法 (比如 Cauchy 序列或者 Dedekind 分割等等) 也改写成符合 ZF 系统的形式.

替换公理 (axiom schema of replacement) 任取集合 X 以及关于 x, y 的逻辑命题 $R(x, y)$, 如果 R 满足 $\forall x \in X$, 存在唯一 y 使得该命题成立, 则存在集合 Y , 使得 $y \in Y$ 当且仅当存在一个 $x \in X$ 使得 $R(x, y)$ 成立.

如果我们把 $R(x, y)$ 理解成一个映射 $f: x \mapsto y$, 那么替换公理构造的 Y 其实就是 X 的象集 $f(X)$.

正则公理 (axiom of regularity) 任取非空集合 X , 其中元素关于 \in 关系存在一个极小元素, 即存在 $x \in X$, 使得 $\forall y \in X, y \notin x$.

注意: 极小元素的选取并不一定唯一, 也不一定是最小元素. 有趣的是, 任取两个自然数 m 和 $n, m < n$ 当且仅当按照前述公理化定义把它们当作集合看待时 $m \in n$. 因此, 这最后一条公理给出了数学归纳法 (mathematical induction) 的一种新的理解: 要想证明一系列命题 P_n 对任意 $n \in \mathbb{N}$ 都成立, 可以取 $X = \{n \in \mathbb{N} | \text{命题 } P_n \text{ 不成立}\}$, 则 X 一定含有一个极小元素 m , 从而任取自然数 $i < m$, 命题 P_i 成立. 于是如果我们能从 P_0, \dots, P_{m-1} 成立推导出命题 P_m 成立, 就完成了证明.

ZF 是集合论的公理化体系中最简单可靠的一个. 当然, 简单可靠的代价就是应用上的局限性. 有许多数学中的著名论断都需要在 ZF 之外再添加一条选择公理才能推导出出来.

选择公理 (axiom of choice) 任取一个由两两不相交的集合构成的集合族 \mathcal{A} , 存在一个集合 C , 它与 \mathcal{A} 的每个元素 (元素是集合) 都恰好交于一点.

选择公理的一个简单推论是: 任取一个集合族 \mathcal{A} (不一定两两不交), 存在一个映射 $f: \mathcal{A} \rightarrow \cup \mathcal{A}$, 使得 $\forall A \in \mathcal{A}, f(A) \in A$. 换言之, 允许同时在集合族 \mathcal{A} 的每个元素 (集合) 里选择一个元素.

选择公理是一个饱受争议的公理, 数学家们一方面对于是否应该允许“同时”进行无穷多项的构造提出了强烈的质疑, 另一方面又利用它证明了很多虚无缥缈、无法构造的东西存在. 比如线性代数中任意线性空间中基的存在性, 或者抽象代数中任意环的极大理想的存在性, 或者实变函数 (测度论) 中不可测集的存在性, 或者泛函分析中的 Hahn-Banach 定理, 它们的证明过程中都直接或者间接地用到了选择公理, 或者用到了在 ZF 中与选择公理等价的良序定理或 Zorn 引理.

值得一提的是与不可测集的存在性相关的一个著名的结论, 人们称之为 Banach-Tarski 悖论 (Banach-Tarski paradox). 这个悖论大意是说, 如果承认选择公理正确, 则可以证明存在三维欧氏空间中的一族有限多个互不相交的子集, 它们并起来是一个半径为 1 的实心球, 但是把每个子集只进行一些旋转和平移, 还可以在新的位置上保持互不相交地重新拼出两个半径为 1 的实心球. 也就是将神话里的“分身术”找到了数学依据! 这看上去显然是非常不符合常识的, 这个结论也曾经一度被认为是选择公理不成立的直接证明, 因此它才被称为“悖论”而不是“定理”.

当然, 这个结论在逻辑上并没有任何矛盾, 它只是和我们从多面体体积那里得来的常识相矛盾, 有什么理由认为这些常识对于那些复杂得不可想象的子集也应该正确呢? 今天的大部分数学家都倾向于站在选择公理一边, 也就是说, 当我们把一大堆没有体积的点胡乱堆在一起的时候, 不应当假定这堆点的“体积”就一定符合积木方砖那种东西带给我们的几何常识. 只不过在每一个需要用到选择公理的结论上, 数学家们都会特意标注一下, 省得将来反悔的时候不知道该丢弃些什么. 我们也会这样处理. 当然, 对于本课的主要内容来说, ZF 已经足够应用了.

有趣的是, 即使是 ZF 加上选择公理构成所谓的 ZFC, 依然解决不了最初难道 Cantor 的连续统假设, 或者说已经解决了, 却不是 Cantor 想要的答案. Gödel 在 1940 年证明了从 ZFC 出发不可能证明连续统假设是错误的, 而 Cohen 则在 1963 年证明了从 ZFC 出发不可能证明连续统假设是正确的.

附录 2

我们在这里介绍一下我国非常著名的概率统计学家——许宝騄先生. 应当说, 他是我们国家概率统计事业的奠基人.

出身钱塘许氏

许宝騄 (1910—1970 年), 字闲若, 祖籍浙江杭州, 1910 年 9 月 1 日生于北京, 因此小名京生。

许宝騄在中国开创了概率论、数理统计的教学与研究。在奈曼—皮尔逊理论、参数估计理论、多元分析、极限理论等方面取得卓越成就, 是多元统计分析学科的开拓者之一。其研究成果推动了概率论与数理统计的发展。至今“许方法”(多元分析统计学家谢菲 H.Scheffe 称之为“数学严密性的范本”)仍被认为是解决检验问题的最实用方法。

许宝騄是 20 世纪最富创造性的统计学家之一, 也正是许宝騄拉开了中国概率论与数理统计学科研究的帷幕, 他被公认为在数理统计和概率论方面第一个具有国际声望的中国数学家。《中国大百科全书数学》称赞说: “许宝騄是中国早期从事数理统计学和概率论研究并达到世界先进水平的一位杰出学者。”1983 年, 德国施普林格出版社刊印了《许宝騄全集》(Pao-Lu Hsu Collected Papers) 由斯普林格 (Springer-Verlag) 出版社刊印, 全集是由钟开莱主编的, 共收集了已发表的、未被发表的论文 40 篇。书评中有这样一句话:

“许宝騄被公认为在数理统计和概率论方面第一个具有国际声望的中国数学家”。

许宝騄的像片悬挂在斯坦福大学统计系的走廊上, 与世界著名的统计学家并列。

许宝騄是二十世纪和华罗庚、陈省身齐名的中国数学家, 中央研究院第一届当选的 5 名数学所院士之一(注: 另外四位首届数学院士是姜立夫、陈省身、华罗庚、苏步青)。当然由于各种原因的结果, 他并不为国人所知。许先生出身于杭州的名门望族。据许氏家谱记载, 杭州许氏本姓沈, 始祖是富阳的沈显荣, 1536 年他在北京经商病故, 遗有两子由杭州的表姑夫许魁带回杭州抚养, 改姓许, 入籍杭州。其中尤以横河桥支系的“积厚轩”堂兴旺。

乾隆三年许家出了第一位举人许钺, 之后钱塘许姓一族显宦至多, 嘉庆、道光、同治间最盛。许氏十世祖许学范为乾隆三十七年(1772 年)进士。许学范生得八个儿子, 按照“学乃身之宝、儒以道得民”的行辈排下来, 这八个儿子便被时人称人钱塘“八乃”, 其中乃济、乃普、乃钊三子为进士, 其余四子也皆中举人, 皇帝给许家赐匾“七子登科”, 意为许家府上出现过七位有才学的人。当时的学者、书法家梁同书曾写对联“世间数百年旧家, 无非积德; 天下第一件好事, 还是读书”赠与许学范。

到光绪二十九年, 共出了 34 位贡生、举人, 其中 9 人中进士。许氏共有 140 多人出任清代的大小官员, 除江西省没有外, 分布全国各省, 名震朝野。老杭州都知道以前文武官员路过横河桥时, “文官下轿, 武官下马”。就是因为许家有许多大官。许乃济(1777 年—1839 年), 字叔舟, 号青士, 为嘉庆十四年(1814 年)进士。选庶吉士, 授编修。嘉庆二十五年(1825 年)任山东道监察御史。道光三年(1823 年)任兵科给事中。道光五年(1825 年)任广东肇罗道, 七年(1827 年)改广东督粮道, 道光九年(1829 年)改高廉道。道光十三年(1833 年)任光禄寺少卿, 官至太常寺少卿。

许乃普(1787 年—1866 年), 字贞锡, 号滇生嘉庆二十五年(1820 年)殿试高居一甲第二名(榜眼), 赐进士及第, 授翰林院编修。官至吏部尚书。同治五年(1866 年)卒, 谥文恪。工书, 法二王, 与祁寯藻、陈孚恩、赵光并称四书家。许乃普子彭寿为道光二十七年(1847 年)丁未科会元, 殿试位列二甲第一名(传胪)。选翰林院庶吉士。散馆授编修。官至内阁学士。

许乃赓为嘉庆二十二年(1817 年)丁丑科二甲进士。改庶吉士, 散馆授编修。历官右庶子。许宝騄的曾祖许乃钊(1799 年—1878 年)行七, 字贞恒, 号信臣, 又号恂普, 为道光十五年(1835

年）殿试二甲、朝考一等第十名，授编修。之后历任国史馆总纂官，河南、广东学政，内阁学士兼礼部侍郎衔，江南大营帮办。咸丰三年（1853年），出任江苏巡抚，镇压上海小刀会起义，次年因无功被革职。咸丰七年（1857年），以三品顶戴帮办江南军务，咸丰八年（1858年），迁光禄寺卿。咸丰十年（1860年），太平军破江南大营，攻克苏州、常州，再次被革职。不久引疾回乡。光绪四年（1878年）卒。

许乃钊的子辈许祐身是许宝騄的祖父，累官至苏州知府。许宝騄父亲许引之从清朝末年至北洋政府时期一直担任中层官员，曾任清朝两浙盐运使。母亲程时嘉，江西新建人。许家为官当中有两人 与慈禧 60 大寿和死亡有关。

许庚身（1825年—1893年），字星叔，号吉珊，崇尚天文、算术、舆地诸学。咸丰二年举人，考取内阁中书。同治元年，以在任官员的身份中进士，任内阁侍读。许庚身在军机处任职近三十年，曾参与策划对太平军和捻军的军事行动。同治十二年，任光禄寺卿。光绪五年，升任礼部侍郎，后分别任户部侍郎和刑部侍郎。中法战争爆发，许以刑部侍郎身份任军机大臣，兼总理各国事务衙门大臣，坚持以“驭夷无上策，可羁縻，不可迁就”的态度对待法国。十四年，升兵部尚书。

许庚身曾奉命总办慈禧 60 大寿，但是在寿前去世，清廷晋赠太子太保，谥恭慎。许宝衡，光绪二十七年举人。他任军机章京时，记了许多日记，成为研究清末衰亡的重要史料。在光绪、慈禧死亡时，他记载了他忙碌的一天。早上五时，他至西苑门吉祥桥时，知道光绪于昨天酉时死了。进内宫后知道，昨天已颁发遗诏，传位于醇亲王子之，由摄政王监国。等到其他大官到场后，由他拟旨尊慈禧皇太后为太皇太后，皇后为皇太后。然后是拟各种皇太后的谕旨。不料，11点慈禧病危。又拟旨由摄政王面请皇太后掌管大事。下午2点，慈禧换衣，随即死亡。他在日记中说“十一时中两遭大丧，亘古所未有，所谓奇变。”就是说，在11个钟头里，皇室死了两人，真是前所未有的大变故，他说在写各种旨章时，“心震手颤，莫知所主”。可见当时朝廷的慌乱。

根据许宝衡日记，在选择新皇帝的纪元年号时，最初拟出四个：宪昌、宪治、宣统、圣宪。一般按清惯例，应首选排在前面的“宪昌”，但不知何故，皇太后圈了“宣统”，在临抄写时，重新把“宣统”列在第一位。

许祐身的孙子一辈也多为知名人士，宝字辈共有兄弟姐妹七人，其中许宝驹、许宝騄、许宝騄被称为“杭州许氏三杰”，许宝驯工于书画、昆曲，是著名学者俞平伯的夫人。一子许宝朴过世较早，但其子许儒鸿却承其世家之风，成就了一代文史兼攻的作家，他便是《慈禧全传》、《胡雪岩》等历史小说的作者、台湾著名作家——高阳。

许宝驹（1898—1960），字昂若，北大国文系毕业，曾积极投身五四运动，历任中国国民党一大代表、国民革命军第十八军党代表、浙江省政府秘书长、铸币局局长、国民党中央执委等职。民盟创始人和领导人之一。1949年后，历任民革中央常委、全国人大代表和政协委员。

许宝騄，1909年4月1日生于浙江杭州，1932年毕业于燕京大学哲学系，后在广州、北京多所大学任教。抗日战争爆发后，他积极追随中国共产党，投身于抗日救亡运动，是“中国民主革命同盟”的重要发起人之一。1945年，许宝騄任《正报》主笔，参与“三民主义同志联合会”地下组织，并在争取北平和平解放的过程中开展了积极有效的工作。新中国成立后，许宝騄同志继续在北京任教，同时担任民革中央宣传部副部长、民革北京市委员会代理秘书长。译有穆勒《论自由》、培根《新工具》等。

七子登科的同时，还有五凤齐飞入翰林，许佑申的姐姐嫁给了礼部尚书廖寿恒，妹妹嫁给了晚清直隶总督陈夔龙。

通过联姻，钱塘许家有许多社会名流姻亲，其中包括俞樾。俞平伯的高祖俞樾（1821-1907）是清朝大学者，浙江德清人，道光进士，官翰林院编修，著有《春在堂全书》二百五十卷。俞樾的二女儿俞绣孙嫁到许家，成了许宝騄的祖母（祖父许佑申），许宝騄的父亲是许引之。许宝騄的一位姑姑是俞樾的孙媳妇，也是俞平伯的母亲。许宝騄的长姊许宝驯工书善画，并善于昆曲，1917年嫁给了著名学者俞平伯。许、俞两家乃故家故亲。许宝騄在家排行第七。许俞两家是亲上加亲，按现代优生学的观点来看，是不符合优生学原理的，值得庆幸的是在两家后代中并没有发生不幸，其后代皆学有所成。

天赋异禀

许宝騄的父亲许引之一生仕宦于晚清、民国，因此幼年的许宝騄随父母出生、成长于宦所。2岁时随家移居天津，八岁时去杭州。1924年，许宝騄十四岁时，父亲病逝于杭州，母亲携全家迁回天津，第二年，移居北京。

许宝騄从小体质就比较孱弱，而其慧敏之秉赋已隐然可见。从5岁开始，许引之就请来家庭教师教授许宝騄在家读书，第一位启蒙师为张树之先生；后改从吴县高德馨（字远香）先生、蕲春陈先生，一直到他十四岁止。期间，许宝騄通读四书五经，涉猎四史及古文辞，大都能琅琅背诵。10岁后许宝騄就学作文言文，因此他的文学修养很深，用语、写作都很精练、准确，塾师评为“简老”。11岁时许宝騄写过以《花生姻缘》和《神花》为题的文言小说。

许宝騄学写小楷后，临摹《玉板十三行》，古朴神似。后来为姐夫俞平伯手写的《古槐书屋词》，有刻本行世。又善就古书诗曲小说制作灯谜，颇具巧思。他善于巧妙地利用古文诗词制作灯谜。十一岁时，许宝騄开始学英文、算术，塾师为杭县邵家驹（字昂士）先生，两年后便能阅读英文的古典文学名著。

许宝騄在幼年之时就已经在数理方面显露出异于常人的天赋。

许宝騄在孩提时就能拼摆“益智图”。到北京后，为了能够考取中学，需要从头开始系统补学新式数学知识，为此专门聘请北京大学数学系吴缉熙老师讲授代数及三角，从头起步，仅仅花了连个月的时间就已经成绩斐然，由此也让许宝騄对数学引起极大的兴趣，其数学天才开始崭露头角。

1925年暑期后，许宝騄顺利考入北京汇文中学高中一年级。高中学习期间，许宝騄利用暑假时间跟着表姐夫徐传元先生（字孟轮，毕业于美国麻省理工学院）钻研数学，得其指点，进步神速。还坚持学习法文，两年后他的法文达到能会话和作短文的程度。清康熙二十九年，褚稼轩《坚瓠集》中载有“移棋相间法”，最初始于清顺治六七年。清末，经学大师俞樾和夫人在闲暇十之时，也做移棋之试，曾推至二十棋子，并作诗记载此事，诗云：

闲将棋子试推移，黑白分明亦一奇。此后空留遗法在，更谁灯下运灵棋。

（注）：褚稼轩《坚瓠集》有移棋相间法，以黑白三子，三移而黑白相间，自三子至十子皆然内人复推广之，自十一子至二十子，今存其法于《春在堂随笔》。

许宝騄在中学读书的时候和俞樾的曾孙俞平伯根据《春在堂随笔》记载，又设法查找《坚瓠集》，知道了最初之法，他们就研究推增至五十子。以后并非不能再推加，因过于繁琐而止。一年后，许宝騄以研究所得移棋相间“合四为一”的新律相告，俞平伯先生有一段文字记此事，文曰：

己巳新正二日之夜，忽以电话觅谈，适余外出，越日访之，乃以合四为一之新律相告，其法简而整，其言明且清，虽其根柢不出四律，而去其繁冗，正其谬误，使人一览豁然贯通，于应用上方便至多，……依新律，则口耳授受一分钟可毕，真庶乎儿童不乱矣。清初以来的“移棋相间法”，本是一种有数学性质的游戏，许宝騄与俞平伯先是推至五十棋子，以后又总结出四条规律，本来还想成一公式，后来许宝騄因为赴英国留学，科学研究任务繁重，对少年时代的游戏也就无暇再问了。

1928年，汇文中学毕业后的许宝騄考入燕京大学理学院学习化学。当时考大学是各校单独命题考试招生，许宝騄当时报考的是燕京大学和清华大学。按许宝騄当时的水平和成绩考入清华大学是没有问题的。不过因为许宝騄文笔快捷，思想敏锐，当时，有不少认识他的人问他题目怎样做，他帮助别人做题而耽误了自己，因而清华大学没有录取，而被燕京大学录取了。

许宝騄在中学期间因为受表姐夫徐传元的影响，对数学颇有兴趣，入大学后了解到清华大学数学系最好，决心转学念数学。由于出众的才能，1929年就转入清华大学数学系继续深造，仍从一年级读起，一起学习的有华罗庚、柯召等人。著名数学家、国际数学界最高荣誉“菲尔兹奖”获得者丘成桐坦言，从1927年到1949年清华数学系是全国最好的数学系。当时清华可算中国数学的一个学术中心，学校当时请了两个著名的外国教授，其中之一便是维纳，清华面向全国办数学培训班，培养了一批教员。

当时清华的教授有杨武之、熊庆来等。杨武之就是杨振宁的父亲，他是一位家教很严的父亲，有时甚至杖责儿子，听说一度因此而使父子关系紧张。

杨武之教授自认学术声望不算高，当时，杨武之公开承认自己的数学成就、解题能力不及他的学生，但他认为解不出难题的教授也可以培养出杰出的学生，因为老师知道哪些题难，哪些题重要，可以布置给学生去想。许宝騄等数学家也许正得益于在清华的这些训练，据华罗庚后来回忆，当时他们一个班有约40人，这门课老师出了几道难题，他（华罗庚）就上图书馆查题鉴、查参考书，向助教请教，每天晚上开夜车，匆匆忙忙完成才发现当天就要交作业了，每次都以为只有他一人按期交作业并得全分，但发下来的，才发现每次都是四个人，另外三人就是陈省身、许宝騄、柯召。许宝騄在清华读书，成绩十分优秀，熊庆来，杨武之等对于他的才华十分看重，杨武之曾作诗书赠先生，首句即为：“许公宝騄，颇有天才”。

许宝騄除了数学特别出色之外，兴趣爱好十分广泛。许宝騄受家庭熏陶，爱好昆曲，工昆旦兼习小生，会拉二胡，精善音律。每听一曲，不出几遍就能写出谱子，表现出良好的艺术才能。30年代，许宝騄与姐夫俞平伯共组清华谷音社。解放后，在教学科研之余，经常参加老君堂俞宅昆曲清唱和北京昆曲研习社活动。

许宝騄还是桥牌高手，当时西方的桥牌刚刚传入清华，所以很多人对打桥牌这种娱乐形式很感兴趣。许宝騄更是利用业余时间钻研桥牌方面的书籍，以至于精通桥牌，牌艺出众，在西南联大任教时曾参加清华的校队，经常参加比赛，还常和朱自清、浦江清、俞平伯在俞家打桥牌。

双博士

许宝騄幼年体质就非常虚弱，以至于对他后来的工作生活造成了很大的影响。在清华读书时时，体重还不到40公斤，并且患旅游肺病和胃病，肺病需要加强营养，但因为同时有胃病，营养得不到吸收。许先生体育成绩大概不及格。当时可用庚子赔款去英国留学，许宝騄所在的清华大学每年有一名名额，这当然是非许宝騄、华罗庚莫属。

1933年，许宝騄于清华大学毕业获理学士学位，经考试录取赴英留学，体检时发现体重太轻不合格，未能成行。在这种情况下，许宝騄到香山疗养过一段时间，当时住在香山脚下的一所房子里，期间注意营养，加强锻炼，使身体状况大为好转。

1934年许宝騄进入北京大学数学系担任正在访问北京大学的美国哈佛大学教授奥斯古德的助教，前后共两年。奥斯古德是分析方面的专家，在他后来出版的书中，提到了许宝騄的帮助。在这两年内许宝騄做了大量的分析方面的习题，并进行了一些研究，于1935年发表了两篇分析方面的论文，其中一篇是与江泽涵合作的。那时芬布尔和阿蒂肯合写的《标准矩阵论》已出版，许宝騄熟练地掌握

了矩阵的工具，尤其精通分块演算的技巧。所以这两年内他在分析和代数两方面都打下了扎实的基础。1936年许宝騄再次考取了赴英留学，派往伦敦大学学院，在统计系学习数理统计，攻读博士学位。

上世纪三十年代，统计学，特别是各种专业统计，例如生物统计、医药统计、农业统计、工业统计首先在英国迅速发展，得到广泛的应用。许宝騄在英国伦敦大学学院攻读博士学位期间，师从著名统计学家 Neyman。此后的几年间，他在统计推断和多元分析等方面做了一系列理论性开创性工作，把许多数学中的分支，如矩阵论、函数论、测度论等引进统计学，使统计学中的许多问题的理论基础更加深厚，逐渐形成了统计学中的一个主流方面——数理统计。在新中国成立后的相当长的一段时间里，在研究机构和一些大学里，统计学往往被称之为数理统计。其实把数理统计视为统计学的一个主要分支似乎更妥当些，可以说，许先生是数理统计这一方向的奠基人之一。

1936年到1940年，伦敦大学学院统计系正处于鼎盛时期，是公认的数理统计研究中心。该系名家汇聚，皮尔逊退休后，由费歇任高尔顿实验室主任，皮尔逊当系主任。一些学者陆续前来访问，包括美国的多元分析专家霍太林，威尔克斯，频率曲线专家克莱格，概率专家费勒。教师中有内曼这样的教授，所以许宝騄很快就接触到数理统计方面科学前沿的情况。自30年代到40年代，正是N. P. 理论（内曼—皮尔逊理论）的形成时期。对于点估计和假设检验，首次提出优良性的概念，明确地提出了应该寻求优良的方法，而优良性有客观的标准。

1938年许宝騄导出了霍太林提出的 T_2 检验在一定意义下是局部最优的，主要的困难是在零假设不成立时，如何导出 T_2 的分布，通常称为非零分布，有了非零分布才能讨论功效函数的大小。他的这一工作在N.P.理论和多元统计分析中都是占有重要地位的先驱性工作。

内曼(Neyman)是著名统计学大师，在其一生中，与许宝騄相遇两次。1936-1938年，在英国大学学院时，内曼是许的老师。他后来称许是他最好的学生。1945-1947年，许先生去美国时，他们又相遇了。内曼在纪念许宝騄的文章中写了如下的这一段话来论述这篇论文的意义：

“这篇文章开创了两个发展方向。一方面，他的学生席玛卡将许的方法用于多元问题（霍太林的 T_2 及多元相关系数）……。另一方面，在这篇文章中，许提供了获得全部相似检验的新方法。在许的建议下，席玛卡和莱曼将这个方法用于其他问题，后来莱曼和谢飞形成了完备性的概念。”

这足以说明许宝騄在这一方面的工作对后来的研究有多大的影响。

1938年许宝騄共发表了3篇论文。当时伦敦大学规定数理统计方向要取得哲学博士学位，必需寻找一个新的统计量，编制一张统计量的临界值表，而许宝騄因成绩优异，研究工作突出，第一个被破格用统计实习的口试来代替。1938年他获得了哲学博士学位。同年，系主任内曼受聘去美国加州大学伯克利分校，他推荐将许宝騄提升为讲师，接替他在伦敦大学讲课。1939年，许宝騄又发表了两篇论文，1940年又发表了3篇。其中两篇文章是数理统计学科的重要文献，在多元统计分析和内曼—皮尔逊理论中是奠基性的工作，因此他获得了科学博士学位。

许宝騄在30年代做的论文，被一些人士认为他可能是继费歇(Fisher)之后执掌统计大旗的人，但是以为身体很差的缘故，学术工作不能不受到一些影响。所以，后来成就就稍逊于皮尔逊(Pearson)和内曼(Neyman)，但不可否认许宝騄是世界第一流的统计学家。

执教西南联合大学

1940年，抗日战争处于最艰难的时候，在英国伦敦大学学院获得双博士学位后的许宝騄，放弃优越的学术环境和生活条件，毅然回国效劳，受聘为北京大学教授，于1940年到昆明，在西南联合大学任教。

当时昆明的工作条件和生活待遇极端艰苦，单身耽在昆明的许宝騄在那段时期的生活是十分艰苦俭朴的。他住一间宿舍，既是卧室又是书房，除了书桌、床铺和一块大黑板外，别无其它大型家具。在这种艰难的环境下，许宝騄喜欢穿一件风雨衣出外散步，在校园附近漫步时学生们常能看到他的儒雅风度。

战争年代生活是十分困难的，1943 年-1944 年许宝騄给 Neymann 的信中曾提到过挨饿之事，但他仍然坚持研究。当时的西南联大图书资讯极端贫乏，连教材都少有，学生听课主要靠记笔记。许宝騄学习非常勤奋、刻苦，因为资料贫乏，以至于想找一本书都困难，他曾手抄过梯其马舍的整本《函数论》，他念过的书，往往都写了不少批注，有的书都被他翻得成零页了。

在 1941 年到 1945 年抗日战争胜利，许宝騄在“Biometrika”、“J. London Math.”、“Ann. Math. Statist.”……等许多国际权威性刊物上发表了十多篇有关数理统计等方面的开创性文章，成为国际上数理统计这一方向的奠基人之一。他曾经说过，我们在某杂志上发文章，不是借该杂志来标榜我们的学术水平，应该让我们在该杂志发表了文章来抬升此杂志的地位。先生这样说了，也确实这样做了。

在西南联大期间，许宝騄对 Neyman-Pearson 理论作出了重要的贡献，他得到了一些重要的非中心分布，论证了 F 检验在上述理论中的优良性，这些都是奠基性的工作；同时他对多元统计分析中的精确分布和极限分布得到了重要的结果，导出正态分布样本协方差矩阵特征根的联合分布和极限分布，这些结果是多元分析中的基石。以上这两方面的工作确立了他在数理统计中的国际上的地位。此外许宝騄在寻求统计量的极限分布，在次序统计量的极限律型方面，都有重要的贡献。西南联大时代，许宝騄开设微分几何课程。

当年在西南联大简陋的教室讲课，往往无法听到上下课的摇铃声，而他又没有手表，于是每次上课时他就将装有一只闹钟的布袋放在讲台上。当上课的学生们第一次忽然听到讲台上布袋里铃声大作，而许先生又笑着说了一声“下课了”时，大家真是惊喜交加。

许宝騄备课极为认真，讲授条理清晰，擅长于在课堂上强调一些细微之处。更因其出生于书香门第，自幼受过很好的文学熏陶，所以他善于运用形象思维的语言来解释现代数学基本概念产生的必然性。

他对学生说微分几何不过是微积分的一种应用而已，而主要工具是泰勒级数展开，再用一点儿初级代数计算。这些话对学生产生了很大影响。许宝騄在教学中特别强调“直观地理解数学”的重要性，他主张要把数学定理及其证明的“原始思想”告知学生，总是殷切地期望学生能直观地领悟数学命题的来龙去脉。在分析问题，他强调要有一种“内视”能力，认为“数学中的抽象能力很重要，一些问题经过抽象后，不仅简明而且其实质也清楚了”。他循循善诱，对学生的指导具体而细致，对学生的读书笔记逐字逐句地修改，甚至错别字和标点也要改正。他批改作业，不但指出正误，而且给出更好的解法。

宝騄对做一名好教师有独到见解：“要做一个好老师，必须自己有相当的功底，才能讲好。应该做到以十当一，自己会十，但讲出来的是一。”他说道：“教师在台上讲课，就像举重运动员，应该举重若轻，很重的东西，一下就举起来了，让人看了感到舒服；而不应该是举轻若重，一份很轻的分量，举也举不起，两腿发抖，让人感到难受。”

正是许宝騄这种精辟的教学思想，深深地影响了学生。他们中一些人已经成为今日的数学栋梁，如钟开莱、冷生明、王寿仁、徐利治和张尧庭等，以及美国的比乌曼 (I.Biumen)、布拉迪 (R.Bradiey)、罗宾斯 (H.Robbins)、鲍克 (A.Bowker)、莱曼 (E.Lehmann) 和奥利金 (I.Olikin) 等。徐利治曾说，

他从许宝騄身上学到了“不怕计算”和“乐于计算”的习惯,十分乐于从计算中发现规律和提炼一般性公式。许宝騄的教学风格也给教堂山的美国学生留下了深刻印象。比乌曼写道:“许坚持深入浅出,毫不回避困难,特别是深沉、明确而默默地投身于学术的最高目标和水准,这些精神吸引了我们。”布拉迪说:“许堪称后辈的典范。”罗宾斯回忆道:“许是令人难忘的,无以伦比的。”

在这一时期,许宝騄受群众的影响,加入了一个政治团体——中国民主革命同盟,它是受中国共产党影响的。

赴美讲学

许宝騄的老师内曼和小皮尔逊在1933年发表了关于假设检验的论文,把检验问题作为一个数学最优化问题来处理,发展了费希尔的研究工作。由于费希尔对皮尔逊有成见,因而对内曼和小皮尔逊的研究也不以为然,甚至称其编辑的《统计学研究通报》是“一堆破烂货”。这也使得内曼感到在英国难以发展,于1938年4月应聘为美国加州伯克利大学数学系教授,并筹建了统计实验室。

1944年的英国正在遭受着德国的轰炸。希特勒亮出了讹诈已久的“秘密武器”——无人火箭。正是在这种情况下,内曼决定回到英国为国效力。但在离开伯克利大学数学系之前找一个适当的接班人是内曼所面临的问题。耐曼始终认为,在战争时期坚持教育质量是非常重要的,美国显然已没有多余的老资格统计学家。二战时期,美国人才紧缺,内曼培养的学生,一个个地被军方调走。在考虑接班人时,想到了他的学生许宝騄。内曼非常器重许宝騄,认为许宝騄是新一代数理统计学家中的佼佼者,一度选定其为接班人。1943年,内曼收到过许宝騄一篇论文的抽印本。中国的形势一片混乱,天下三分,其一在日本人的统治下,其二由共产主义者领导,第三部分属于民族主义者。许宝騄在某处一个地窖里从事科学研究。他的工作居然能够印出来,简直是个奇迹。他渴望到美国来。

内曼第一脚踏上赴英的旅程,就给许宝騄发了一份电报,邀请他到伯克利来讲六个月课,并在来年担任一个收入有保证的职务。

由于内曼的邀请,许宝騄于1945年八月在北京大学留职停薪,在当年欧战结束之后,赴美参加了伯克利(Berkeley)第一届概率统计的学术交流会。当时,赴美路费是内曼从44年开始设法筹措的。

会后,许宝騄在加州大学伯克利分校教了一个学期。在伯克利任教期间,加州大学主要教员有两位,一位是内曼,另一位就是许宝騄。内曼因有许宝騄这样才华出众的数理统计学家在伯克利而感到高兴。因此,当他看到许宝騄在一张表中仅仅被称为统计学讲师后,便对这个“令人不快的错误”向校方提出了抗议。

他说:“有人也许会说在一张表格中怎么称呼关系不大,但是既然写上了头衔就有其目的,否则写它做什么。对许宝騄来说,在这样的失误中所受到的损害比美国人更多。作为一个中国学者,两所美国的重要大学邀请他来当访问教授,这是一种很高的荣誉。事实上,要是那张表能真实的反映他的工作,那就好了,那样一份东西留在他家中,子孙后代看了也会感到自豪的。”

一学期之后,许宝騄又到哥伦比亚大学教了一个学期。当时美国各大学争相邀请许先生去任教,内曼希望许宝騄继续回到伯克利。但最后,许宝騄决定在哥伦比亚大学的教学工作结束后跟着郝泰林(Hotelling)到北卡罗来纳大学去筹建统计系。许宝騄在郝泰林创办的北卡罗来纳大学的新统计系教了一年。因为统计这个领域正在发展,很缺乏像许这样有成就的学者,他在这三个学校都很有建树。他被选为IMS(Institute of Mathematical Statistics)的委员,这是1946年的事。

许宝騄在伦敦大学学院攻读学位时,熟读了克拉美的《随机变量与概率分布》(1937年出版)掌握了特征函数的工具,所以他对极限理论很有兴趣。1947年他与H.罗宾斯合写的论文《全收敛和大

数定律》，第一次引入全收敛的概念。当时国际上在概率方面主要的兴趣是独立随机变量之和的极限分布，正在从古典的向近代结果转化。一些著名的概率论专家如科尔莫哥罗夫，辛钦，格涅坚科，莱维和费勒等人都在攻这难题。1947年，许宝騄已获得了主要的结果：每行独立的无限小随机变量三角阵列的行和，依分布收敛到一给定的无穷可分律的充分必要条件。由于当时信息不通，他不知道别人的工作情况，当时他写信给钟开莱时说：“……我担心正在进行的工作会和别人相重……”后来，他知道了格涅坚科和科尔莫哥罗夫的工作，就没有再发表自己的研究。实际上许的方法和俄国人还是不同的，许的方法更为直接。1968年，当格涅坚科和科尔莫哥罗夫合写的《独立随机变量之和的极限分布》英译本再版时，钟开莱用附录的方式第一次刊印了许宝騄的工作。然而许在生前并未看到这本书，他始终没有看到自己的这一部分工作的公开发表。

1947年，许宝騄谢绝了一些大学的聘任，回到北京大学任教授。当许宝騄回国时，内曼一再挽留，想把他争回自己的麾下。回国后，许宝騄也与奈曼保持了多年的联系。许宝騄对科学所做的贡献以及孜孜以求的好学精神，是与内曼的教诲和影响分不开的。

1948年，许宝騄当选为中央研究院院士。回国后不久，许宝騄就发现已患肺结核。此后他长期带病工作，教学科研一直未断，在矩阵论，概率论和数理统计方面颇有建树。

执教北京大学数学系

许宝騄在北卡罗莱纳大学开创了统计系之后，于47年底回北大数学系任教。

许宝騄回国可能也为了结婚，但由于种种原因，他终身未婚。关于他的个人婚姻事宜，有各种说法，俞平伯的儿子俞润民的说法可以认为是可靠的。俞润民先生有一姑父郭则澐，是清朝翰林，曾在北洋政府徐世昌处任秘书长。郭有一女儿，与许先生年龄相仿。按中国旧的传统来说，许郭辈份不合，许宝騄比这位郭女士长一辈。郭则澐认为虽然辈份不合，但郭先生与许宝騄先生的父亲在北洋政府时期是同事，作为儿女之事也是合适的。所以郭先生还是赞同这桩婚事。据推算，1936年许宝騄留英以前，两人已有恋爱关系而往来，后来之所以没有成，是郭先生的儿子（郭女士的弟弟）反对这桩婚姻，欲将他的姐姐嫁给国民党的一个官吏。后来，郭女士嫁给了一个铁路局长。这样，许先生的婚事就告吹了。这件事情发生在抗战时期（40-45），许郭之间相处的时间不长。

外面传说，许先生因肺结核未能成婚，这不符合事实，只是对外的一种托辞。另据俞润民回忆，许宝騄在美国时，曾有一女士有意于他，这一次是自由恋爱。据润民回忆，许宝騄给俞平伯先生的信中提到他采取的策略是“以退为进”，现在已搞不清楚以退为进的具体含义。从现在恋爱方式来看，应是双方主动。不久发现身染肺疾，乃作废约。此后终身未娶。

许宝騄因为学与统计学的诸多领域成果丰硕，造诣精深。1948年，当我国中央研究院首批评选院士时，三十八岁的许宝騄就高票当选。

在1948年辽沈战役后，许宝騄已确信“国民党败局已定”，他很欢欣1949年我国的解放，还拍电报给美国的同事表达他对中国新生的喜悦之情。

从1947年至1956年，许宝騄一直在北京大学数学系执教。开设一些国内一般大学不开的课程，像“实变函数论”等，并积极开展多元统计和矩阵论方面的研究，特别是关于矩阵方面所获得的结果，就是基础数学家，也为之赞叹，此乃上上之作。

中共建政初期，不认为概率统计是重要的，他在矩阵论方面的工作，刊登了几篇论文在中国的数学杂志上，发表的论文还有与特征函数有关的内容。1954年，中国科学家代表团访问苏联，柯尔莫果洛夫(Kolmogorov)教授问到了许。人们开始认识到许的国际地位。1955年，他成为中国政治协商会的一名委员。同年他与其他著名的数学家，如华罗庚、苏步青等一起被选为中国科学院学部委员。

1956年,周恩来总理主持制定了“全国科学发展12年远景规划”。规划中把概率论与数理统计作为数学的三大重点发展方向之一。为了落实这一规划,大力发展我国概率论与数理统计,许宝騄殚思极虑,费尽心力,采取了当时条件所能做到的一切措施。

1956年秋,中科院数学研究所的王寿仁先生,张里千先生,中山大学的郑曾国先生、梁之舜先生被借调到北京大学任教。与此同时,从北京大学数学力学系抽出34名四年级学生,从中山大学和南开大学各抽调10名四年级学生来北京大学培养,此外北京大学还接收全国各主要综合大学的概率统计方面的教师来进修。许宝騄亲自主持“独立随机变量族的极限理论”的讨论班,系统地学习了“测度论”、“概率极限理论”、“马尔可夫链”、“数理统计”等课程。这是我国第一批培养的数量可观的概率论与数理统计人才。自此以后,全国各综合性大学绝大多数都设有概率论与数理统计教研室。

1958年以后,先生主持着三个讨论班:数理统计、马尔可夫过程、平稳过程。参加讨论班的人员,不仅有北京大学概率论与数理统计教研室的师生,还有校外的一些人士。

在1956年第一批大规模培养概率论与数理统计人才的实践基础上,许宝騄逐步确定了本专业的必学的基本课程:测度论;概率极限理论(后改称分析概率论);随机过程论;数理统计。根据本专业的不同研究方面,可在下列诸课程中选择一两门:马尔可夫过程,平稳过程,博奕论,排队论,统计试验设计,抽样论……等等。

1956年以前,北京大学数学力学系仅开设一门概率论课程,教材是苏联 Gnedenko 所著“概率论教程”(丁寿田译)。此书作教学参考书可以,完全作教材并不太合适。许先生建议搞教材建设,一是引进,翻译国外一批优秀著作当教学参考书,二是组织力量自己写书。

遗憾的是,由于“文革”前以阶级斗争为纲,这些著译计划,在许宝騄有生之年以前,绝大部分没有完成,有的仅出了一半就夭折了,有的还在酝酿中,等到“文革”后,才出版了一部分。

“文革”前,我国与西方并无学术交流,许宝騄计划在1957和1958两年聘请苏联和东欧的一些著名学者来华讲学。计有:苏联的 Denkin 教授, Prokhorov 教授,波兰的 Fisz 教授和 Urbanik 教授。结果,1957年 Fisz 来北京大学讲了多元分析和抽样论等专题;Urbanik 来北京大学讲了广义随机过程;Denkin 于1958年来北京大学讲了马尔可夫过程的几个前沿课题。由于当时反右和大跃进运动,基本上没有取得什么效果。

49年以后,国内教学方面的杂志,除了少数几所大学的学报以外,就只有“中国科学”、“科学纪录”、“数学学报”和“数学进展”这几种主要杂志。年青人的文章很难有发表的机会。有鉴于此,许宝騄极力主张创办概率统计杂志。还说,如经费不够,可从我的积蓄中资助。无奈那时对出版物的控制很严,就连与政治相隔甚远的杂志也不易获得创刊。

无言的结局

从1956到1959年,在许宝騄的领导下,不变原理(Prokhorov, Donsker 的论文)、多元分析(Anderson 的书)和随机过程(Doob 的书)是他所主持的讨论班讨论的主题。他要求他的学生在讨论班上主讲,在他们讲完后,他再用他特有的简洁的方法给以总结。从1959到1962年,试验设计、抽样调查(cochran 书)都是研究的主题。在他领导下,这一阶段做过林业的调查。从1963到1966年,主题是次序统计量、平稳时间序列(Grenander 和 Rosenblatt 的书),马尔可夫过程以及组合数学。

从1955年起,许宝騄身体非常虚弱。他于1933年开始患肺结核,后来终身未娶,过单身生活。他约有174公分身高,但体重只有43.2公斤。到了六十年代,那时他已因身体条件不能在课堂上讲授概率论。而后的岁月,他只能在家中坐在沙发上对讨论班的研究集体讲课。

当 1966 年 5 月文化大革命冲击来临后，所有的教学活动都停止了，他和学校中其他的同事都遭受到折磨。

从北京大学办公楼往东南行约五六百米，柳林深处坐落着几栋小平房，有的形似老北京的四合院，有的由门字形的三排平房组成，这就是鲜为人知的佟府。

许宝騄从上世纪五十年代初到 1970 年他去世，就一直住在佟府丙八号。这是一所两廊四间的小平房。一进门是一个临时封闭起来的托檐，不到四平方米，用作厨房。由此前进，是一条由北向南的走廊，尽头是一间贮藏室，东西两侧各有两居室。西侧较大，住着张景昭老师一家，东侧两间较小，进门一间大的，也只有十三四平方米，算是许宝騄的客厅，里面的套间就是卧室和卫生间了。客厅其实是一个多功能厅。厅内东面墙上，挂着一块黑板，北面放着两个齐屋顶高的书架和一把双人沙发，西南各放置一把单人沙发，中间是一张一米见方的矮桌，厅内还放着几张小凳和几个竹壳热水瓶。许宝騄主持教研室的讨论班时，这个客厅就是教室；教研室要政治学习或讨论问题时，它变成了小会议室；查阅资料时，它变成了图书馆；用餐时，它又变成了餐厅；只有外客来访或学生向先生问问题，这间小屋才恢复原来的角色——客厅。

许宝騄饮食非常简单，一天三小瓶牛奶，早中晚各一瓶，中餐和晚餐也只有两三碟小菜，由于先生不仅患过肺病，还有胃病，食量很小。张景昭老师曾在西南联大数学系就读，许先生当时是西南联大教授，所以张也可算许先生的学生，为了照顾许先生，她与许宝騄合请一个保姆料理家务，并经常陪先生一起进餐。由于房子小，保姆只能早来晚归。

许宝騄绝大部分时间在卧室工作。靠坐在床上，在一块一尺见方的薄板上把稿纸展开，撰写论文和讲稿。由于睡眠情况不好，黎明前就开始工作，晚上无人照顾，饿了就用一块巧克力和一杯热开水充饥，“三年困难期间”，巧克力也随之困难掉了。累了就听听收音机，然后再睡一会。卧室中总是放着一台“熊猫牌”收音机，这是了解时事和休闲的主要工具。

上世纪五十年代，许宝騄寒暑假还到北京城内的宾馆疗养一两个星期，六十年代以后，几乎是足不出户。自患肺病后，身体一直瘦弱，无论春、秋、冬，在室内总是穿着长衫和毛裤，只有外客来访，才着正装。

没有多少人知道，住在这样简朴的居室，过着如此清贫的生活，忍着如此的孤寂，竟然是为我国科学事业辛勤耕耘一生的一代学术宗师许宝騄先生。

文化大革命中许宝騄并未中断研究，当时看不到任何杂志，直到 1970 年，才允许他看杂志，那时他已瘫痪。在两个月内，他翻阅了 1966 年“文化大革命”以后的全部《数理统计纪事》，了解国际上的学术动态，写下了最后一篇关于 BIB 与编码的论文，并将这篇文章的手稿托付给段学复教授。

1970 年 12 月 18 日许宝騄去世。临终时，在他枕边搁着一枝他用了数十年的派克笔，从床头到地板散落着一大堆演算的草稿。这些遗物，见证了先生临终前一刻还在继续着他终生的事业，还在思考和探索广阔无垠的科学世界。

去世时，身边无一亲朋好友。一代宗师，竟这样走完他的人生之旅。虽然他患过肺结核和胃病，但这些均非不治之症，更何况 50 年代末已治愈。若不是那场“史无前例”的运动，先生何至英年早逝，六十而亡。一位终身独居，为国家的科教事业而呕心沥血的爱国者，一位享誉国内外的学术泰斗，仍难逃此劫难，实为我们国家的不幸，民族的不幸。

如今的佟府丙八号，早已是人去房毁，若仅是人去房空，还可以去那里凭吊先生，现在，只能把崇敬和谢恩之情永铭心间。

许的确是中国统计学家的先驱,他开创了这条道路。他被公认为在概率统计领域具有国际地位的首位数学家(见 Springer-Verlag 出版社 1983 年刊印的许宝騄全集)。许宝騄的像片悬挂在斯坦福大学统计系的走廊上,与世界著名的统计学家并列。

奈曼认为,许绝对地是与阿伯拉罕·瓦尔特 (Abraham Wald) 具有同一水平的人,他们是他下一代中两个杰出的数理统计学家(见 Reid 书奈曼传)。瓦尔特是在 1950 年到印度旅游演讲后死于飞机失事,才 48 岁。许死时 60 岁。不幸的是许在 1950 年后未能发表更多的卓越的论文,但他确实给中国统计的发展打下了坚实的基础。

80 年代中期,在旅美华裔统计学家郑清水的协助下,曾由德国 Springer 出版社委托钟主编出版了英文版的《许宝騄选集》。无疑,这是对世界统计数学文库的重要贡献。后来又在钟开莱、郑清水、徐利治联名倡议下,创立了“许宝騄概率统计数学奖”,并且发布了一次奖。但终于由于基金匮乏和人力不济等原因,未能使这项具有深远意义的数学奖顺利地继续运作下去。

他不仅自己在多元分析方面有很多开创性的工作,他还培养了像安德森、奥肯等国际多元分析学术带头人。许宝騄先生把数学家分成三流,他说:

“第一流的数学家,是有天才的,他们能开闢新的领域,如柯尔莫哥洛夫,冯·诺依曼,维纳这一类人,这些人是可望而不可及的。

第二流数学家是靠刻苦学习而成的,认真消化整理前人的东西,在这个基础上有所创造发现,象欣钦这样的数学家就是这一类的,他写的《公用事业理论的数学方法》、《信息论基础》等就是消化整理的结果。这种工作对后人影响较大,年青人可以在这个基础上较快地进入科学的前沿,中国缺少一批做这一类工作的人。

第三流的数学家只在某一、二个问题上有一点贡献,不能象第二流的那样有系统的工作。剩下的就是不入流的数学家了。”

他认为自己没有才能,是刻苦学习得到的,他也没有经验去培养有天才的人,他只能传授如何认真学习,努力钻研,埋头苦干的经验。他衷心希望他的学生超过他,一次他在讨论班上说:“自古以来,只有能教出状元来的老师是光荣的,至于做状元的学生那就没有什么了。”

许宝騄的大弟子钟开莱和许一样也是一位才智出众的人物。他俩出生于历史名城杭州市,都对古典文学有深深的爱好和素养,他俩都能写出典雅的中文文章和英文文章。可以这么说,是许宝騄造就了钟开莱,若没有许宝騄的提拔,钟开莱恐怕就难以成名。钟开莱在中学时成绩很好,以双百、第一的成绩考进了西南联大,在大学期间成绩也很好,师从华罗庚研究数论。但钟与华都较自负,两人关系不很融洽,华罗庚出的论文题钟觉得不满意,钟就自己找题做。据说他们在讨论数学问题时,都曾拍案而起,互相反问“你有什么了不起?”学位答辩时,结果是全票通过。之后,询问导师的意见,决定他是否留校,华罗庚立即回答“不留”,但许宝騄马上说:“你不留我留”。一般来说,如果一个教授决定不留某一学生,别的教授不便再留,但许先生实在爱惜钟的才华,又知华、钟二人关系不好,结果钟开莱改从许先生。

Years	where	Topics
1935	Beijing	topology
1938-1940	London	statistics
1941-1945	Kunming	Random matrices & Limiting distr.
1946-1947	USA	Matrices & complete convergence
1949-1964	Beijing	Ch.f., Matrices trasf., Markov processes,
		Experimental design & Limiting distr.
1968-	-	Markov chain, Random Mat.,
		Limiting distr., Coding.