

高维概率及其在 数据科学中的应用

High-Dimensional Probability An Introduction with
Applications in Data Science



[美] 罗曼·韦尔希宁 著 冉启康 译
(Roman Vershynin)



机械工业出版社
China Machine Press

这是一本优秀的、非常及时的教材，书中以一种容易接受和面向实际应用的方式展示了高维几何和概率这一现代工具，且包含大量的练习。这本书融入了作者的洞察力和直觉，并广泛参考了该领域的新发展。对新手和专业研究人员来说，本书都非常有用。

—— 陶哲轩，加州大学洛杉矶分校

只要具备概率论和线性代数第一门课程的知识就可阅读本书。韦尔希宁指导读者贯穿主题并始终如一地通过现代数据科学应用来说明书中理论的实用性。这本书应该是概率论、数据科学及相关领域的学生 and 研究人员必备之书。

—— Ramon van Handel，普林斯顿大学

这本书因对与当代统计科学和机器学习特别相关的几个主题的简洁介绍而广受欢迎。作者在呈现深刻的理论与保持非专业读者的可读性之间做到了很好的平衡。

—— Richard Nickl，剑桥大学

韦尔希宁是高维概率领域的世界领先专家之一，他的这本书为该领域的许多关键工具及其在数据科学领域中的应用进行了优美且全面的介绍。本书涉及的主题对于任何想在这个领域做数学研究，包括从事机器学习、算法和理论计算机科学、信号处理和应用数学工作的人来说都是必需的。

—— Jelani Nelson，哈佛大学

本书无论是对新人还是熟悉该领域的人都提供了优质的资源。我相信，正如作者所希望的，本书所涵盖的内容确实是数据科学发展中必不可少的。

—— Santosh Vempala，佐治亚理工学院

在本书中，罗曼·韦尔希宁对该领域的重要概念、工具和技术进行了循序渐进、清晰的阐述。高年级学生及对数据科学的数学基础感兴趣的实际工作者将在本书中看到许多相关的工作实例和有趣的练习。

—— Rémi Gribonval，法国国家信息与自动化研究所研究室主任/高级研究员

罗曼·韦尔希宁是高维概率研究中的领军人物，且是一位语言表达大师，在本书中，他提供了一些基本的工具和高维概率的主要结果及应用。

—— Elchanan Mosse，麻省理工学院

高维概率的基本知识对应用数学、统计学和计算机科学交叉领域的研究人员来说是必不可少的。广泛的内容和循序渐进的介绍将使得这本书成为经典，基础数据科学领域的每个人都应该阅读它。

—— Alfred Hero，密歇根大学

韦尔希宁的这本书是对现代信号处理和数据科学所需的核心数学知识的精彩介绍。重点是测度集中及其在随机矩阵、随机图、降维和随机过程的界等方面的应用。

—— Andrea Montanari，斯坦福大学

 **CRC Press**
Taylor & Francis Group



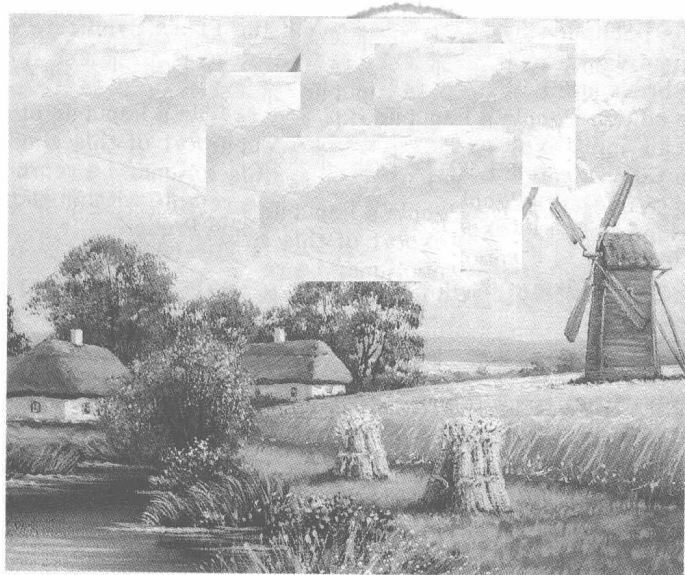
投稿热线：(010) 88379604
读者信箱：hzsj@hzbook.com
客服电话：(010) 88361066 88379833 68326294

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字阅读：www.hzmedia.com.cn

统计学精品译丛

高维概率及其在 数据科学中的应用

High-Dimensional Probability An Introduction with
Applications in Data Science



[美] 罗曼·韦尔希宁 著 冉启康 译
(Roman Vershynin)



机械工业出版社
China Machine Press

此为试读, 需要完整PDF请访问: www.ertongbook.com

图书在版编目 (CIP) 数据

高维概率及其在数据科学中的应用 / (美) 罗曼·韦尔希宁 (Roman Vershynin) 著; 冉启康译. —北京: 机械工业出版社, 2020.4

(统计学精品译丛)

书名原文: High-Dimensional Probability: An Introduction with Applications in Data Science

ISBN 978-7-111-65209-0

I. 高… II. ①罗… ②冉… III. 概率论—应用—数据处理 IV. TP274

中国版本图书馆 CIP 数据核字 (2020) 第 052981 号

本书版权登记号: 图字 01-2019-0936

This is a Simplified-Chinese edition of the following title published by Cambridge University Press:

Roman Vershynin, High-Dimensional Probability: An Introduction with Applications in Data Science, 978-1-108-41519-4.

© Roman Vershynin 2018.

This Simplified-Chinese edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2020.

This Simplified-Chinese edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

本书原版由剑桥大学出版社出版.

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版. 未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分.

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售.

本书全面介绍高维概率的理论、关键工具和现代应用, 涵盖霍夫丁不等式和切尔诺夫不等式等经典结果以及矩阵伯恩斯坦不等式等现代发展, 还介绍了基于随机过程的强大方法, 包括 Slepian 不等式、Sudakov 不等式和 Dudley 不等式等, 以及基于 VC 维数的泛型链接和界限. 全书使用了大量插图, 包括协方差估计、聚类、网络、半正定规划、纠错码、降维、矩阵补全、机器学习、压缩感知以及稀疏回归的经典和现代结果.

本书适合数学、统计学、电气工程、计算生物学及相关领域的博士生和高年级硕士生以及初级研究人员使用.

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 柯敬贤

责任校对: 殷虹

印刷: 大厂回族自治县益利印刷有限公司

版次: 2020 年 5 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 15

书号: ISBN 978-7-111-65209-0

定价: 79.00 元

客服电话: (010) 88361066 88379833 68326294

投稿热线: (010) 88379604

华章网站: www.hzbook.com

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

本书赞誉

“这是一本优秀的、非常及时的教材，书中以一种容易接受和面向实际应用的方式展示了高维几何和概率这一现代工具，且包含大量的练习。这本书融入了作者在该领域的洞察力和直觉，并广泛参考了该领域的新发展。对新手和专业研究人员来说，本书都非常有用。”

——陶哲轩，加州大学洛杉矶分校

“在概率论理论研究及其在数学、统计学、计算机科学和电气工程领域的应用等众多问题中，高维概率的方法已经成为不可缺少的工具。罗曼·韦尔希宁这本教材用一种非常易于接受的方式介绍这个领域，填补了该领域著作上的空白。本书开始只用到了概率论和线性代数第一门课程的知识。作者始终如一地通过现代数据科学应用来说明书中理论的实用性。可以说，本书是概率论、数据科学及相关领域的学生和研究人员的必不可少的读物。”

——拉蒙·范汉德尔，普林斯顿大学

“这本非常受欢迎的著作简洁介绍了与当代统计科学和机器学习相关性很大的几个主题。作者在展示理论的深度和对于非专业读者的可读性方面做了很好的平衡——强烈推荐希望了解更多现代数学中不可缺少的这方面知识的研究生和研究人员阅读本书。”

——理查德·尼克尔，剑桥大学

“韦尔希宁是高维概率领域的世界领先专家之一，本书对该领域的许多关键工具及其在数据科学领域中的应用进行了优美且全面的介绍。本书涉及的主题对于任何想在这个领域做数学研究，包括从事机器学习、算法和理论计算机科学、信号处理和应用数学工作的人来说都是必需的。”

——杰拉尼·尼尔森，哈佛大学

“高维概率是对概率和数据分析中现代方法的一种很好的处理。韦尔希宁既是概率分析专家又是泛函分析专家，他的观点是独特和深刻的。他对内容的讨论是循序渐进的、全面的、引人入胜的。本书无论是对新人还是熟悉该领域的人都提供了优质的资源。我相信，正如作者所希望的，本书所涵盖的内容确实是数据科学发展中必不可少的。”

——桑托什·维帕拉，佐治亚理工学院

“罗曼·韦尔希宁因对高维概率的杰出贡献而闻名。本书中，他循序渐进地阐述了该领域的重要概念、工具和技术。高年级学生及对数据科学的数学基础感兴趣的实际工作者将在本书中看到许多相关的工作实例和有趣的练习。本书是我一直期待的参考书。”

——雷米·格里邦瓦尔，法国国家信息与自动化研究所研究室主任/高级研究员

“高维概率是一个迷人的数学理论，近年来发展迅速。它是高维统计、机器学习和数据科学的基础。罗曼·韦尔希宁是高维概率研究中的领军人物，且是一位语言表达大师，在本书中，他提供了一些基本的工具和高维概率的主要结果及应用。本书可作为数学、统计学、计算机科学和工程学学生的教科书，也可作为从事高维概率和统计工作的研究人员的参考书。”

——埃尔哈南·莫塞尔，麻省理工学院

“这本关于高维概率的理论和应用的书描述得异常清晰，对数据科学基础感兴趣的学生和研究人员来说是非常有价值的。高维概率的基本知识对应用数学、统计学和计算机科学交叉领域的研究人员来说是必不可少的。本书内容广泛，讲述循序渐进，基础数据科学领域的每个人都应该阅读它。”

——阿尔弗雷德·海罗，密歇根大学

“韦尔希宁的这本书对现代信号处理和数据科学所需的核心数学知识进行了精彩介绍，重点是测度集中及其在随机矩阵、随机图、降维和随机过程的界等方面的应用。本书的处理非常干脆利落，读者将学到不含冗余公式的深奥数学并获得美的享受。”

——安德烈亚·蒙塔纳里，斯坦福大学

序 言

本书从一个有趣的实例——B. Maurey 提出的求集合凸包的平均近似点的经验方法开始，使读者了解概率论可以优美地解决那些初看上去和概率无关的问题。从本书中学习概率论如何打开其他（那些之前被发现是很难研究的）数学领域的大门是一段令人非常满意的经历。

在展示了必要的背景材料后，第 3 章直接进入了本书的核心部分：以一种具有启发性的方法处理高维集中问题。例如，注 3.1.2 中公式 $\sqrt{n \pm O(\sqrt{n})} = \sqrt{n} \pm O(1)$ 的表述很简洁。同样在图 3.6 中，一个高斯点云在高维中表出：它集中在一个半径为 $\sqrt{n}$ 的球面上。这种形状与二维或三维中的钟形形状几乎没有任何共同之处——我们的低维直觉是无用的！作为概率论可以让生活更容易的另一个例子，本书给出了 Grothendieck 不等式的富有洞察力的证明。要理解该不等式的任何其他形式（带有“好”的常数）的证明，我们可能需要几年时间。

再提及一个本书处理得很好的主题：等周不等式以及它如何导致放大。如果球面的一个子集覆盖了至少 50% 的球面，那么它的覆盖范围指数地接近于 100%。本书还介绍了一些扩展到其他度量空间的内容，例如，格拉斯曼流形上的集中。通过这种方式，为读者提供了这个领域的一个进入点，如果读者对这个领域感兴趣，可参考各章后注中提供的丰富材料。

本书读起来很有趣。作者把材料当作激动人心的故事来讲述，使人欲罢不能。许多分散在书中的练习鼓励读者参与故事情节的发展。

本书中讨论的其他主题包括随机矩阵、经验过程理论和稀疏恢复等，这些结果对数据科学的研究很重要，其本身也很漂亮。许多学生和研究人员可能已经听出了话外音，这是一本他们苦苦寻求的书。

Sara van de Geer, ETH Zürich

前言

读者对象

这是一本着眼于数据科学应用的高维概率论教材，它面向数学、统计学、电气工程、计算生物学及相关领域的博士生和高年级硕士生以及初级研究人员，为扩展他们在现代数据科学研究中使用的理论方法而写。

关于本书

数据科学正在快速发展，概率方法经常为其提供基础和灵感。如今，一门经典的研究生概率论课程已经不足以达到数据科学研究人员所期望的数学复杂程度。本书旨在部分地填补这一空白。它提出了一些关键的概率方法和结果，这些方法和结果为数学数据科学家提供了必要的理论工具。它可以作为概率论第二门课程的教材，以使学生对数据科学的应用有所了解。本书也适合自学。

本书内容

高维概率是概率论中一个研究 $\mathbb{R}^n$ 中的随机对象的分支，其中维数 n 可能非常大。本书重点介绍随机向量、随机矩阵和随机投影。它讲授分析这些对象的基本理论技能，包括集中不等式、覆盖与填充理论、解耦和对称化技巧、随机过程的链和比较技术、基于 VC 维数的组合推理等。

高维概率的研究为数据科学应用提供了重要的理论工具。本书将理论与协方差估计、半正定规划、网络、统计学习要素、纠错码、聚类、矩阵补全、降维、稀疏信号恢复和稀疏回归等应用结合起来。

预备知识

阅读本书的基本前提是具备扎实的概率论基础(硕士或博士水平)，对本科阶段的线性代数有很好的掌握，对度量空间、赋范空间和希尔伯特空间以及线性算子的基本概念有全面的了解。对测度论是否了解并不重要，但会有所帮助。

关于练习

练习穿插在正文中。对文中所提的问题，读者可以立即进行验证，以检验对该问题的

理解，并为接下来的应用做更好的准备。练习的难度用咖啡杯的数量来表示，排列顺序由易(☺)到难(☹☹☹☹)。带指向的手(☞)意味着该练习在本书末尾有提示。

相关阅读

本书只涵盖了高维概率理论内容的一小部分，并且其应用仅限于数据科学中的一些例子。本书的每章结尾都有一个后注，给出与本章内容相关的其他文献，也给出了一些特别有用的信息。现代经典的文献[8]全面介绍了概率方法在离散数学和计算机科学中的应用。文献[19]呈现了数学数据科学的全景图，其重点在计算机科学中的应用上。研究生和高年级本科生都可以阅读这两本书。文献[206]是面向研究生的，更多地介绍了高维概率的理论。

致谢

许多同事的反馈对准备本书很有帮助。特别感谢 Florent Benaych-Georges、Jennifer Bryson、Lukas Grätz、Rémi Gribonval、Ping Hsu、Mike Izbicki、George Linderman Cong Ma、Galyna Livshyts、Jelani Nelson、Ekkehard Schnoor、Martin Spingler、Dominik Stöger、Tim Sullivan、Terence Tao、Joel Tropp、Katarzyna Wyczesany、Yifei Shen 和 Haoshu Xu 提出的许多有价值的建议和更正，特别是 Sjoerd Dirksen、Larry Goldstein、Wu Han、Han Wu 和 Mahdi Soltanolkotabi 对本书的详细校对。很感谢 Can Le、Jennifer Bryson 和我的儿子 Ivan Vershynin 在许多图片上的帮助。

目 录

本书赞誉
序言
前言

第 0 章 预备知识：用概率覆盖一个几何集 1

0.1 后注 3

第 1 章 随机变量的预备知识 4

1.1 随机变量的数字特征 4

1.2 一些经典不等式 5

1.3 极限理论 7

1.4 后注 8

第 2 章 独立随机变量和的集中 9

2.1 集中不等式的由来 9

2.2 霍夫丁不等式 11

2.3 切尔诺夫不等式 14

2.4 应用：随机图的度数 16

2.5 次高斯分布 17

2.6 广义霍夫丁不等式和辛钦不等式 22

2.7 次指数分布 24

2.8 伯恩斯坦不等式 28

2.9 后注 30

第 3 章 高维空间的随机向量 32

3.1 范数的集中 32

3.2 协方差矩阵与主成分分析法 34

3.3 高维分布举例 38

3.4 高维次高斯分布 42

3.5 应用：Grothendieck 不等式与半正定规划 46

3.6 应用：图的最大分割 50

3.7 核技巧与 Grothendieck 不等式的改良 52

3.8 后注 55

第 4 章 随机矩阵 57

4.1 矩阵基础知识 57

4.2 网、覆盖数和填充数 61

4.3 应用：纠错码 64

4.4 随机次高斯矩阵的上界 67

4.5 应用：网络中的社区发现 70

4.6 次高斯矩阵的双侧界 74

4.7 应用：协方差估计与聚类算法 75

4.8 后注 78

第 5 章 没有独立性的集中 80

5.1 球面上利普希茨函数的集中 80

5.2 其他度量空间的集中 85

5.3 应用：Johnson-Lindenstrauss 引理 89

5.4 矩阵伯恩斯坦不等式 92

5.5 应用：用稀疏网络进行社区发现 98

5.6 应用：一般分布的协方差估计 99

5.7 后注	101	8.8 后注	172
第 6 章 二次型、对称化和压缩 ...	103	第 9 章 随机矩阵的偏差与几何结论	174
6.1 解耦	103	9.1 矩阵偏差不等式	174
6.2 Hanson-Wright 不等式	106	9.2 随机矩阵、随机投影及协方差估计	179
6.3 各向异性随机向量的集中	109	9.3 无限集上的 Johnson-Lindenstrauss 引理	181
6.4 对称化	110	9.4 随机截面: M^* 界和逃逸定理 ...	183
6.5 元素不是独立同分布的随机矩阵	112	9.5 后注	186
6.6 应用: 矩阵补全	114	第 10 章 稀疏恢复	187
6.7 压缩原理	116	10.1 高维信号恢复问题	187
6.8 后注	118	10.2 基于 M^* 界的信号恢复	188
第 7 章 随机过程	119	10.3 稀疏信号的恢复	189
7.1 基本概念与例子	119	10.4 低秩矩阵的恢复	192
7.2 Slepian 不等式	122	10.5 精确恢复和 RIP	194
7.3 高斯矩阵的精确界	127	10.6 稀疏回归的 Lasso 算法	199
7.4 Sudakov 最小值不等式	129	10.7 后注	203
7.5 高斯宽度	131	第 11 章 Dvoretzky-Milman 定理 ...	204
7.6 稳定维数、稳定秩和高斯复杂度	135	11.1 随机矩阵关于一般范数的偏差	204
7.7 集合的随机投影	137	11.2 Johnson-Lindenstrauss 嵌入和更精确的 Chevet 不等式	206
7.8 后注	140	11.3 Dvoretzky-Milman 定理	208
第 8 章 链	142	11.4 后注	211
8.1 Dudley 不等式	142	练习提示	212
8.2 应用: 经验过程	148	参考文献	217
8.3 VC 维数	152	索引	226
8.4 应用: 统计学习理论	161		
8.5 通用链	166		
8.6 Talagrand 优化测度和比较定理	169		
8.7 Chevet 不等式	170		

第 0 章 预备知识：用概率覆盖一个几何集

我们从概率方法在几何学中的应用的一个漂亮结果来开始高维概率学习.

回忆一下, 点 $z_1, \dots, z_m \in \mathbb{R}^n$ 的凸组合是系数为非负且系数的和为 1 的线性组合, 即它的形式为

$$\sum_{i=1}^m \lambda_i z_i \quad \text{其中 } \lambda_i \geq 0, \quad \sum_{i=1}^m \lambda_i = 1 \quad (0.1)$$

集合 $T \subset \mathbb{R}^n$ 的凸包是指 T 中所有有限点集的所有凸组合的集合:

$$\text{conv}(T) := \{z_1, \dots, z_m \in T \text{ 的凸组合, 对所有 } m \in \mathbb{N}\}$$

见图 0.1 所示.

$\mathbb{R}^n$ 中定义凸组合的元素数目 m 是没有先验限制的. 但是, 经典的卡拉特奥多里 (Caratheodory) 定理指出: 我们总是可以取 $m \leq n+1$.

0.0.1 定理 (卡拉特奥多里定理) 集合 $T \subset \mathbb{R}^n$ 的凸包中的每一个点都可以表示为 T 中最多 $n+1$ 个点的凸组合.

界 $n+1$ 是不能被改进的, 因为显然能找到一个单纯形 T (一般位置上的 $n+1$ 个点的集合). 可是, 如果我们仅仅是想将一个点 x 近似表示为 $x \in \text{conv}(T)$, 而不是将它精确地表示为凸组合, 能用少于 $n+1$ 个点来实现吗? 我们下面证明这是可能的, 实际上所需点的数量根本不需要依赖于 n 的维数!

0.0.2 定理 (卡拉特奥多里定理的近似形式) 考虑一个集合 $T \subset \mathbb{R}^n$, 其直径[⊖]以 1 为界. 那么, 对每一个点 $x \in \text{conv}(T)$ 和每个整数 k , 可以找到点 $x_1, \dots, x_k \in T$, 使得

$$\left\| x - \frac{1}{k} \sum_{j=1}^k x_j \right\|_2 \leq \frac{1}{\sqrt{k}}$$

有两个理由使得这一结果是令人惊讶的. 首先, 凸组合中点的数量 k 不依赖于维数 n . 其次, 凸组合的系数可以全部相等 (可是, 需要注意的是: 点 x_i 之间是可以重复的).

证明 我们的证明方法被称为 B. Maurey 的经验方法.

如有必要, 变换 T , 不用直径, 而是假定 T 的半径以 1 为界, 即

$$\|t\|_2 \leq 1, \text{ 对所有 } t \in T \quad (0.2)$$

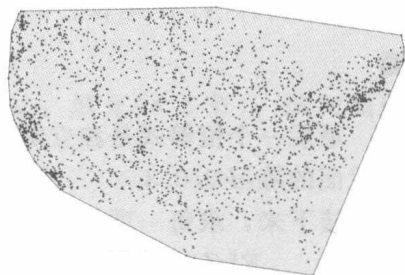


图 0.1 表示美国主要城市的点集的凸包

⊖ T 的直径定义为 $\text{diam}(T) = \sup\{\|s-t\|_2 : s, t \in T\}$. 为了简单起见, 假设 $\text{diam}(T) = 1$. 对于一般集 T , 定理中的界变为 $\frac{\text{diam}(T)}{\sqrt{k}}$ (自己验证!).

固定一个点 $x \in \text{conv}(T)$, 并将其表示为一些向量 $z_1, \dots, z_m \in T$ 的形如 (0.1) 的凸组合. 现在, 将 λ_i 视为概率, 我们从概率的角度来解释凸组合 (0.1) 的定义. 具体来说, 我们可以定义一个随机向量 Z , 它取值 z_i 的概率为 λ_i , 即

$$\mathbb{P}\{Z = z_i\} = \lambda_i, \quad i = 1, \dots, m$$

(这是可能的, 因为权数 λ_i 是非负的, 且和为 1). 则有

$$\mathbb{E}Z = \sum_{i=1}^m \lambda_i z_i = x$$

考虑 Z 的独立副本 $Z_1, Z_2, \dots$, 由强大数定律, 有

$$\frac{1}{k} \sum_{j=1}^k Z_j \xrightarrow{\text{几乎处处}} x, \quad k \rightarrow \infty$$

为了得到结果, 我们计算 $\frac{1}{k} \sum_{j=1}^k Z_j$ 的方差 (顺便说一下, 这个计算是证明弱大数定律的核心), 得到

$$\begin{aligned} \mathbb{E} \left\| x - \frac{1}{k} \sum_{j=1}^k Z_j \right\|_2^2 &= \frac{1}{k^2} \mathbb{E} \left\| \sum_{j=1}^k (Z_j - x) \right\|_2^2 \quad (\text{由于 } \mathbb{E}(Z_i - x) = 0) \\ &= \frac{1}{k^2} \sum_{j=1}^k \mathbb{E} \| Z_j - x \|_2^2 \end{aligned}$$

最后一个恒等式是独立随机变量之和的方差等于方差之和这一基本事实的高维版本, 见后面的练习 0.0.3.

接下来, 求每一项的方差的界. 我们有

$$\begin{aligned} \mathbb{E} \| Z_j - x \|_2^2 &= \mathbb{E} \| Z - \mathbb{E}Z \|_2^2 \\ &= \mathbb{E} \| Z \|_2^2 - \| \mathbb{E}Z \|_2^2 \quad (\text{另一个方差恒等式, 见练习 0.0.3}) \\ &\leq \mathbb{E} \| Z \|_2^2 \leq 1 \quad (\text{由于 } Z \in T \text{ 并使用 (0.2)}) \end{aligned}$$

我们已经证明了

$$\mathbb{E} \left\| x - \frac{1}{k} \sum_{j=1}^k Z_j \right\|_2^2 \leq \frac{1}{k}$$

因此, 存在随机变量 $Z_1, \dots, Z_k$ 的一条路径, 使得

$$\left\| x - \frac{1}{k} \sum_{j=1}^k Z_j \right\|_2^2 \leq \frac{1}{k}$$

由构造可知每个 Z_j 在 T 中取值, 证毕. ■

0.0.3 练习 验证下面的方差恒等式, 这是我们在定理 0.0.2 的证明中使用过的.

(a) 设 $Z_1, \dots, Z_k$ 为 $\mathbb{R}^n$ 中的独立、零均值随机向量, 求证

$$\mathbb{E} \left\| \sum_{j=1}^k Z_j \right\|_2^2 = \sum_{j=1}^k \mathbb{E} \| Z_j \|_2^2$$

(b) 设 Z 为 $\mathbb{R}^n$ 中的随机向量, 求证

$$\mathbb{E} \| Z - \mathbb{E}Z \|_2^2 = \mathbb{E} \| Z \|_2^2 - \| \mathbb{E}Z \|_2^2$$

下面给出定理 0.0.2 在计算几何中的一个应用. 假设我们有一个子集 $P \subset \mathbb{R}^n$, 并要求用给

定半径 ϵ 的球覆盖它，如图 0.2 所示. 最少需要多少球，应该如何放置它们？

0.0.4 推论(用球覆盖多面体) 设 P 为 $\mathbb{R}^n$ 中有 N 个顶点的多面体，其直径以 1 为界. 那么， P 最多可被半径为 $\epsilon > 0$ 的 $N^{\lceil 1/\epsilon^2 \rceil}$ 个欧几里得球所覆盖.

证明 定义球的中心如下：设 $k := \lceil 1/\epsilon^2 \rceil$ 并考虑集合

$$\mathcal{N} := \left\{ \frac{1}{k} \sum_{j=1}^k x_j : x_j \text{ 为 } P \text{ 的顶点} \right\}$$

我们宣称中心在 $\mathcal{N}$ 上的 ϵ 球族满足推论的结论. 为了验证这一点，注意到多面体 P 是其顶点集的凸包，我们用 T 表示这个凸包. 因此，可将定理 0.0.2 应用于任意点

$x \in P = \text{conv}(T)$ ，并推导出 x 与 $\mathcal{N}$ 中某一点的距离为 $\frac{1}{\sqrt{k}} \leq \epsilon$ ，这表明中心在 $\mathcal{N}$ 上的 ϵ 球确实覆盖了 P .

为了求 $\mathcal{N}$ 的势，注意到，从 N 个顶点中允许重复地选择 k 个，共有 N^k 种方法. 因此 $|\mathcal{N}| \leq N^k = N^{\lceil 1/\epsilon^2 \rceil}$ ，证毕. ■

在本书中，我们将学习与覆盖问题有关的其他几种方法：填充(4.2 节)、熵和编码(4.3 节)以及随机过程(第 7 章和第 8 章).

在结束这一节之前，我们稍微改进一下推论 0.0.4.

0.0.5 练习(二项式系数之和) 证明不等式

$$\left(\frac{n}{m}\right)^m \leq \binom{n}{m} \leq \sum_{k=0}^m \binom{n}{k} \leq \left(\frac{en}{m}\right)^m$$

对于所有的整数 $m \in [1, n]$ 成立.

0.0.6 练习(改进覆盖) 验证：在推论 0.0.4 中

$$(C + C\epsilon^2 N)^{\lceil 1/\epsilon^2 \rceil}$$

个欧几里得球就足够了. 这里 C 是一个合适的绝对常数(注意，对于小的 ϵ ，这个界略强于 $N^{\lceil 1/\epsilon^2 \rceil}$). ■

0.1 后注

在本章中，我们说明了概率方法，利用随机性构造了一个有用的对象. 文献[8]提供了许多概率方法的例子，主要集中在组合数学方面.

本章介绍的 B. Maurey 的经验方法最初在[162]中提出. B. Carl[48]用它得到了覆盖数的界，包括推论 0.0.4 和练习 0.0.6 中所述的那些界. 练习 0.0.6 中的界是精确的，见[48-49].

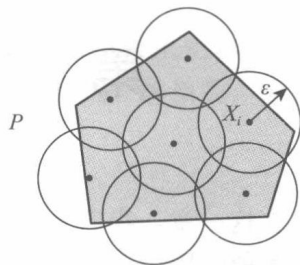


图 0.2 覆盖问题：需要多少个半径为 ϵ 的球去覆盖 $\mathbb{R}^n$ 中给定的集合 P ，以及如何放置这些球

第 1 章 随机变量的预备知识

在这一章, 我们复习概率论中的一些基本概念和结果. 读者对其中的大部分知识应该是熟悉的, 因为这些知识通常会在概率论的入门课程中讲授.

1.1 节介绍随机变量的期望、方差和矩. 一些经典等式可以在 1.2 节中找到. 1.3 节回顾两个基本的概率极限理论: 大数定律与中心极限定理.

1.1 随机变量的数字特征

在概率论的基础课程中, 我们学习了随机变量 X 的两个最重要的数字特征: 期望[⊖](也称为均值)和方差. 在本书中将它们表示为[⊕]

$$\mathbb{E}X \text{ 和 } \text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$$

我们回顾一下描述概率分布的其他经典数字特征和函数. X 的矩母函数被定义为

$$M_X(t) = \mathbb{E}e^{tX}, \quad t \in \mathbb{R}$$

对于 $p > 0$, X 的 p 阶矩记为 $\mathbb{E}X^p$, 且 p 阶绝对矩记为 $\mathbb{E}|X|^p$.

p 阶绝对矩的 p 次根是有用的, 这引出了随机变量的 L^p 范数的概念:

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{\frac{1}{p}}, \quad p \in (0, \infty)$$

通过 $|X|$ 的本性上确界, 这个定义可以扩展到 $p = \infty$ 上去, 即

$$\|X\|_{L^\infty} = \text{ess sup } |X|$$

对于给定的 p 和概率空间 $(\Omega, \Sigma, \mathbb{P})$, 经典向量空间 $L^p = L^p(\Omega, \Sigma, \mathbb{P})$ 由定义在 Ω 上的所有具有有限 L^p 范数的随机变量 X 构成, 即

$$L^p = \{X : \|X\|_{L^p} < \infty\}$$

如果 $p \in [1, \infty)$, 则 $\|X\|_{L^p}$ 是一个范数, L^p 是一个巴拿赫(Banach)空间. 这一事实源于闵可夫斯基(Minkowski)不等式, 我们将在下面的(1.4)中复习它. 对于 $p < 1$, 由于三角不等式不成立, 因此 $\|X\|_{L^p}$ 不是范数.

L^p 空间中, 当指数 $p = 2$ 时情况很特殊, 因为 L^2 不仅是巴拿赫空间还是希尔伯特(Hilbert)空间. L^2 上的内积和相应的范数由下式给出:

$$\langle X, Y \rangle_{L^2} = \mathbb{E}XY, \quad \|X\|_{L^2} = (\mathbb{E}|X|^2)^{\frac{1}{2}} \quad (1.1)$$

因而 X 的标准差可表示为

$$\|X - \mathbb{E}X\|_{L^2} = \sqrt{\text{Var}(X)} = \sigma(X)$$

⊖ 如果你已经学习了测度论, 你会记得概率空间 $(\Omega, \Sigma, \mathbb{P})$ 上的随机变量 X 的期望 $\mathbb{E}X$, 是由函数 $X: \Omega \rightarrow \mathbb{R}$ 的勒贝格(Lebesgue)积分定义的, 这使得勒贝格积分的所有定理都适用于概率论中随机变量的期望.

⊕ 在本书中, 我们省略了括号而简单地写为 $\mathbb{E}f(X)$. 因此, 非线性函数直接写在期望后面.

同样, 可以将随机变量 X 和 Y 的协方差表示为

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \langle X - \mathbb{E}X, Y - \mathbb{E}Y \rangle_{L^2} \quad (1.2)$$

1.1.1 注(随机变量中的几何学) 当我们将随机变量视为 L^2 空间(希尔伯特空间)中的向量时, 等式(1.2)给出了协方差的几何解释: 向量 $X - \mathbb{E}X$ 和 $Y - \mathbb{E}Y$ 越相关, 它们的内积和协方差越大.

1.2 一些经典不等式

詹森(Jensen)不等式是指: 对于任意随机变量 X 和凸函数[⊖] $\phi: \mathbb{R} \rightarrow \mathbb{R}$, 我们有

$$\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X)$$

由詹森不等式直接得到: $\|X\|_{L^p}$ 是关于 p 的增函数, 即

$$\|X\|_{L^p} \leq \|X\|_{L^q}, \quad \text{对于任何 } 0 \leq p \leq q < \infty \quad (1.3)$$

这个不等式成立是因为当 $\frac{q}{p} \geq 1$ 时, $\varphi(x) = x^{\frac{q}{p}}$ 是凸函数.

闵可夫斯基不等式是指: 对于任意 $p \in [1, \infty)$ 和任意随机变量 $X, Y \in L^p$, 我们有

$$\|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p} \quad (1.4)$$

可以将这个不等式看作三角不等式, 其中当 $p \in [1, \infty)$ 时, $\|\cdot\|_{L^p}$ 是范数.

柯西-施瓦茨(Cauchy-Schwarz)不等式是指: 对任意随机变量 $X, Y \in L^2$, 我们有

$$|\mathbb{E}XY| \leq \|X\|_{L^2} \|Y\|_{L^2}$$

更一般的赫尔德(Hölder)不等式是指: 如果 $p, q \in (1, \infty)$ 是共轭指数, 即 $\frac{1}{p} + \frac{1}{q} = 1$, 则随机变量 $X \in L^p$ 和 $Y \in L^q$ 满足

$$|\mathbb{E}XY| \leq \|X\|_{L^p} \|Y\|_{L^q}$$

该不等式对 $p=1, q=\infty$ 同样适用.

回忆一下, 由概率论的基本概念可知, 随机变量 X 的分布直观地反映了 X 所取值的概率信息. 更确切地说: X 的分布由 X 的累积分布函数(Cumulative Distribution Function, CDF)所确定, X 的累积分布函数定义为

$$F_X(t) = \mathbb{P}\{X \leq t\}, \quad t \in \mathbb{R}$$

使用随机变量的尾分布通常更方便, 即

$$\mathbb{P}\{X > t\} = 1 - F_X(t)$$

尾分布与随机变量的期望(更一般地说, 矩)之间存在着重要的关系. 下列等式通常用于界定尾分布的期望.

1.2.1 引理(积分等式) 设 X 为一个非负随机变量, 则有

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\} dt$$

该等式的两端是同时有限或无限的.

⊖ 凸函数的定义: 如果对于任意 $\lambda \in [0, 1]$ 及任意在 ϕ 的定义域中的向量 x, y , $\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y)$ 恒成立, 则称函数 ϕ 是凸的.

证明 可以通过如下等式[⊖]表示任何非负实数 x ,

$$x = \int_0^x 1 dt = \int_0^\infty \mathbf{1}_{\{t < x\}} dt$$

用随机变量 X 替换 x , 并对等式两边取期望, 得

$$\mathbb{E}X = \mathbb{E} \int_0^\infty \mathbf{1}_{\{t < X\}} dt = \int_0^\infty \mathbb{E} \mathbf{1}_{\{t < X\}} dt = \int_0^\infty \mathbb{P}\{t < X\} dt$$

为了交换第二个等式中期望和积分的顺序, 我们使用了 Fubini-Tonelli 定理, 证毕. ■

1.2.2 练习 (积分等式的推广)  证明引理 1.2.1 的推广, 即求证引理 1.2.1 对任何随机变量 X (不一定非负) 都成立, 即

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\} dt - \int_{-\infty}^0 \mathbb{P}\{X < t\} dt$$

1.2.3 练习 (通过尾分布计算 p 阶矩)  设 X 是随机变量, $p \in (0, \infty)$. 求证:

$$\mathbb{E}|X|^p = \int_0^\infty p t^{p-1} \mathbb{P}\{|X| > t\} dt$$

7 在右端是有限的时恒成立.

另一个经典工具——马尔可夫(Markov)不等式——也可以通过期望来界定尾分布.

1.2.4 命题 (马尔可夫不等式) 对于任何非负随机变量 X 和 $t > 0$, 有

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}X}{t}$$

证明 对任意 $t > 0$, 我们可以通过示性函数

$$x = x \mathbf{1}_{\{x \geq t\}} + x \mathbf{1}_{\{x < t\}}$$

来表示任意实数 x . 用随机变量 X 替换 x , 并对等式两边取期望, 则有

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}X \mathbf{1}_{\{X \geq t\}} + \mathbb{E}X \mathbf{1}_{\{X < t\}} \\ &\geq \mathbb{E}t \mathbf{1}_{\{X \geq t\}} + 0 = t \mathbb{P}\{X \geq t\} \end{aligned}$$

将不等式两边同时除以 t , 结论成立, 证毕. ■

众所周知, 切比雪夫(Chebyshev)不等式是马尔可夫不等式的特殊情况. 它给出了比对 t 更好的依赖性, 即关于 t 是二次依赖的. 且它不是控制单侧尾部, 而是量化了 X 关于其均值的集中度.

1.2.5 推论 (切比雪夫不等式) 设 X 是一个随机变量, 其均值为 μ , 方差为 σ^2 , 则对任意 $t > 0$, 有

$$\mathbb{P}\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}$$

1.2.6 练习  通过将 $|X - \mu| \geq t$ 两端平方, 并应用马尔可夫不等式推导出切比雪夫不等式.

1.2.7 注 第2章的命题 2.5.2 中, 我们将建立三个与随机变量相关的基本量(矩母函数、 L^p 范数、尾分布)之间的联系.

⊖ 这里及本书后面, $\mathbf{1}_E$ 表示事件 E 的示性函数, 如果 E 发生则它取值为 1, 否则为 0.