

开放教育与

Kaifang Jiaoyu Yu

Gaozhi Jiaoyu Yitihua Banxue Linian Yu Shijian

高职教育一体化办学理念与实践

肖 坤◇著

 人 民 出 版 社

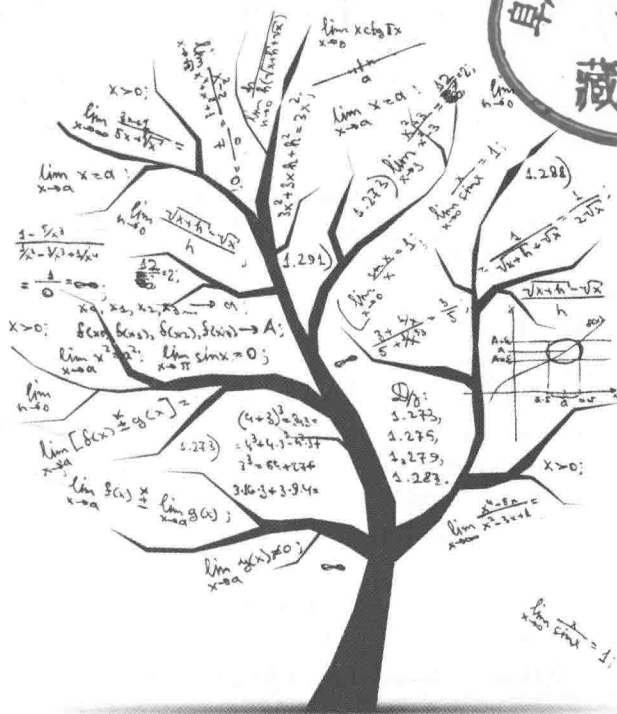
TURING

图灵程序
设计丛书

程序员的数学②

概率统计

[日] 平冈和幸 堀玄 / 著 陈筱烟 / 译



人民邮电出版社
北京

图书在版编目(CIP)数据

程序员的数学. 2, 概率统计 / (日) 平冈和幸,
(日) 堀玄著; 陈筱烟译. — 北京: 人民邮电出版社,
2015.8

(图灵程序设计丛书)
ISBN 978-7-115-40051-2

I. ①程… II. ①平… ②堀… ③陈… III. ①电子计
算机—数学基础②概率统计 IV. ①TP301.6②0211

中国版本图书馆CIP数据核字(2015)第176373号

内 容 提 要

本书沿袭《程序员的数学》平易近人的风格,用通俗的语言和具体的图表深入讲解程序员必须掌握的各类概率统计知识,例证丰富,讲解明晰,且提供了大量扩展内容,引导读者进一步深入学习。

本书涉及随机变量、贝叶斯公式、离散值和连续值的概率分布、协方差矩阵、多元正态分布、估计与检验理论、伪随机数以及概率论的各类应用,适合程序设计人员与数学爱好者阅读,也可作为高中或大学非数学专业学生的概率论入门读物。

-
- ◆ 著 [日] 平冈和幸 堀 玄
译 陈筱烟
责任编辑 徐 骞
执行编辑 高宇涵
责任印制 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京天宇星印刷厂印刷
 - ◆ 开本: 800×1000 1/16
印张: 26.5
字数: 492千字 2015年8月第1版
印数: 1-4000册 2015年8月北京第1次印刷
- 著作权合同登记号 图字: 01-2013-8603号

定价: 79.00元

读者服务热线: (010)51095186转600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京崇工商广字第0021号

说到与程序设计关系紧密的数学学科，肯定有不少人首先就会想到概率论与统计学吧。从信息论到机器学习，从模式识别到数据挖掘，概率与统计的概念和原理活跃于计算机科学的各个领域，发挥着重要的作用。同时，概率与统计也都是应用性极强的学科。它们源于现实需求，在不断发展成熟后又反过来推进了各类问题的解决。纯粹的理论推演无法展现概率论与统计学的全部魅力，计算机科学的出现，为概率统计搭建了绝佳的应用舞台。

话虽如此，充满了大量概念与公式的概率与统计理论并非总是轻松有趣。概率论中既有简单明了的结论，也有与直觉相违的原理。只有准确地理解并掌握所有这些，才能真正在解决实际问题时灵活运用它们。幸运的是，这本《程序员的数学2：概率统计》另辟蹊径，借助不同于传统数学教材的独特编排思路，将概率泛化为面积与体积，强调随机变量的函数属性，深入浅出，巧妙地阐述了概率论与统计学的基本理论。全书穿插了大量形象有趣的实例，并辅以难度恰当的练习题，使没有经过大学数学系课程专业训练的读者也能理解与掌握概率与统计的精髓。

既然题为“程序员的数学”，本书自然也专门介绍了一些与计算机科学相关的主题。本书涵盖了随机数理论、贝叶斯公式、马尔可夫链、信息熵等计算机专业课程中涉及的知识点，不但能够作为相关课程教材的补充，还能启发读者理解甚至设计相关的计算机算法，解决与概率统计相关的计算问题。本书的习题量不大，但都很典型，是检测相关知识点理解程度的理想材料。为了深入掌握本书内容，建议读者准备纸笔，随时通过笔头推算与演练，必能获得更好的学习效果。

数学是一门严谨的学科。翻译以数学为主题的书籍，不仅需要准确地理解原文，将其转换为恰达的中文，更重要的是严格精确地表述各条数学表达式的含义。中日数学表述体系间存在的细微差异无形中为此增加了难度，遑论国内数学教材对部分名词术语的翻译尚未达成统一，一字一句，都必须推敲琢磨。原则上，本书选择了当前学界较为流行的中文术语与公式表述方式，关键的公式与理论也经过了审校与确认。然限于水平，谬误之处在所难免，敬请广大读者给予批评与斧正。

本书能顺利完稿，离不开许多人的帮助与支持。家人的理解与关心，使自己能够始终安心地翻译。当我在生活与工作中遇到困难时，好友的支持帮助我攻克了一个个难关。最

后还要感谢图灵公司的编辑提出了大量有价值的建议与意见，帮助本书顺利完成并最终问世。最后，希望对概率统计有兴趣的读者们都能从本书中获益。

陈筱烟

2015年6月于上海

本书的编写目标

本书的目标读者是那些希望在自己感兴趣的领域中运用概率统计知识，但并不打算成为数学专业人士的人。本书名称沿用笔者的另一本书《程序员的数学3：线性代数》，但内容更加普适而通用。

说起概率论与统计理论，很多人都会对此抱有一些负面印象：计算公式复杂，检验类型繁多，辛苦学习很久后却发现真正对工作有帮助的其实是那些电子表格软件。不少人在运用统计理论时态度也很消极，他们只是为了让别人相信自己的结论客观合理而被迫使用那些繁琐的处理。

然而概率统计理论并非没有价值。事实上，近年来为了更好地处理数据，人们开始在各个学科中积极应用概率统计理论。数据挖掘、垃圾邮件自动筛选、文档的自动分类与类似文件搜索、非法使用的鉴别（例如根据信用卡消费历史得出持卡人的消费模式，以检测非正常消费行为，或监视网络流量识别可疑的访问模式等）、语音识别与图像识别、通信工程（例如信号清晰的高音质通话技术）、基因分析、金融工程中的资产组合分析、受生物学启发得到的高效信息处理方式（神经网络、遗传算法等）以及基于蒙特卡罗模拟的循证式日程规划等各领域中都活跃着概率统计理论的身影。

为了了解这些技术，我们必须掌握一些概率统计基础知识。然而，目前面向非理科背景读者的教材少之又少，他们往往不知道应该怎样学习概率统计知识。

概率论的专业教材需要一定的数学功底才能阅读，普通读者可能难以理解；而轻松易懂的入门书又常常不成体系，或是无法提供足够的信息。如果要应用概率统计理论，我们需要掌握更多的基础知识。例如，在实际应用中我们常会遇到多个事件同时发生的情况，如何才能胸有成竹地处理这类问题，而不是仅凭直觉或模糊的概念妄加猜测呢？为此，除了了解基本的计算步骤，我们还需要深刻理解以下两条概念。

- 概率是面积与体积的泛化
- 随机变量是一种以变量为名的函数

不了解这方面的读者可能难以理解这两句话，不过它们已经成为现代概率论的基础。

另一方面，与统计相关的参考书通常会罗列各类分布与检验方法。可惜这种编排方式难以帮助我们将概率统计理论应用至上述当前热门的领域。然而，直接阅读专门介绍垃圾邮

件分类的技术书也并非良策。换个角度讲，我们不可能仅靠从应用中学到的知识来改进已有的应用方式。如果基础知识不足，在实际应用中就不能深入理解问题，从而无法触及问题的本质。何况这些技术至今仍在不断发展，只学习一些阶段性的未成熟的技术并没有太大的价值。相比研究这些具体的检验方法与应用技术，我们更应该首先强化相关的基础知识，以应对(包括当前方式在内的)更广泛的应用情景。

如果读者已经具备了一定的数学基础(不但能够理解数学表达式，还能领会那些抽象而巧妙的描述方式)，教材的选择就会轻松很多。注重学习效率的教材、基础知识详细的教材、讲解生动有趣的教材，不一而足，大家可随意挑选。然而，由于存在这一门槛，普通读者缺乏合适的教材，于是能够掌握稍微系统一点的概率统计知识的人就一下子少了很多。

为此，本书在编写时考虑了以下三个原则。

- 精选了大量非数学专业读者也应当掌握的知识点
- 内容的深度与大学非数学系学生的学习能力相符
- 知识的讲解力求详细具体

因此，与传统教材相比，本书的着力点与讲解思路都较为独特。尽管本书面向初学者，却保持了一定的深度。虽然前半部分花了较长篇幅详解条件概率的概念，但省略了各种类型的分布，也没有具体介绍各类计算技巧及特征函数等内容。这是权衡了各个知识点的重要程度之后做出的决定。此外，有些部分的讲解看似冗长，实则语言明晰，非常易于理解。

在讲解各类估计与检验方式等统计学核心知识点时，本书也下了一番功夫。本书希望读者不仅能了解它们的用法，还能理解其背后的原理，因此详细介绍了相关的基础概念与思维方式。上文提到过，本书并不打算罗列所有的统计方法(尤其是数据的采集方式与区间估计等内容)。如果读者想要学习各种具体的估计与检验方法，请参考专门的统计学教材。本书将着重说明这些方法背后的理论依据，解释这些方法背后共通的原理。

0.1 在阅读本书前，需要预先掌握哪些数学知识^①

- 需要具备高中理科级别的数学知识(包括向量、微积分等的概念与基本计算能力)
- 偶尔会用到大学级别的数学知识(主要是多元微积分)，不过本书会为不具备这方面知识的读者提供必要的解释
- 不会使用大学数学系级别的知识(主要是测度论)。如果讲解时无法避免，本书将明确提示，并给出易于理解的说明，帮助读者建立直观印象

此外，第5章与8.1节需要用到一些大学程度的线性代数知识(例如，需要通过正交矩阵将对称矩阵对角化)。如果读者之前不了解线性代数理论的应用价值，也许会惊叹它的巨大威力。

最后，附录A简单总结了求和符号、指数与对数的一些基础知识，读者如有需要可适当参考。

^① 该部分在正文中称为“回答专栏”。——译者注

0.2 为什么要在写线性代数教材的同时再写一本概率统计教材

撰写这两本教材都是因为该学科的初学者与专家之间存在人才断层现象。在这两个领域中，一些有经验的人能够轻松解决的问题，初学者却常常感到过于复杂。产生这一问题的根本原因在于，无论是矩阵还是概率，不同水平的人对它们的理解程度往往大相径庭。

例如，只要具备一定的线性代数基础，就一定会知道矩阵的本质是一种映射。然而，考虑到初学者不一定能够理解这种抽象的解释，入门教材一般会使用更浅显的说明方式。因此，笔者希望打破这一惯例，教授初学者矩阵更加本质的含义，即，

矩阵是一种映射！

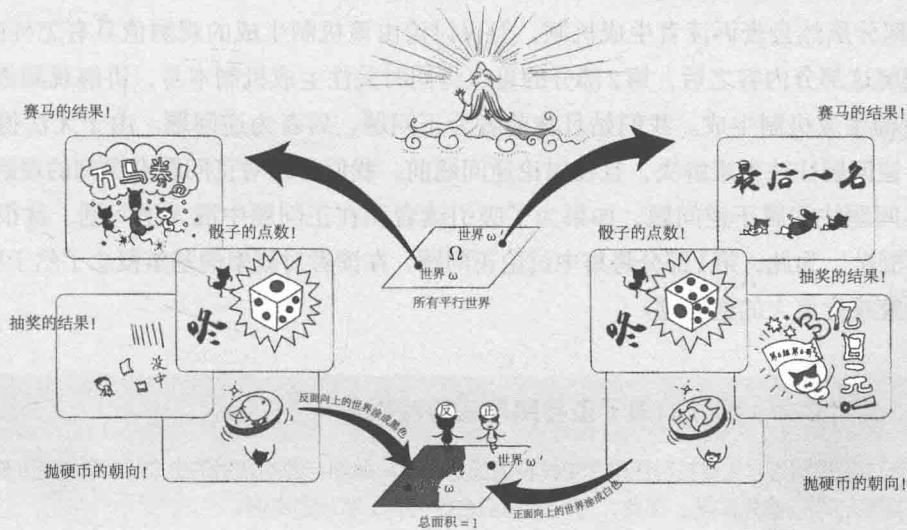
笔者相信，这本书证明了如下几点。

- 只要讲解详细，说明充分，初学者也能理解抽象概念的本质
- 即使不打算成为数学方面的专家，掌握这些本质也大有裨益
- 事实上，首先阐明概念的本质，往往更利于读者理解

本书也将以此为目标。对于具备一定的概率统计基础的人来说，以下这句话是众所周知的常识。

概率是一种面积！

本书希望让更多的人理解这一概念^①。1.3节将借助“上帝视角”具体讲解这句话的含义（它也许与读者的想象有些差别）。在理解了这句话之后，概率论中的很多问题都会迎刃而解，敬请期待。



① 准确地讲，概率是一种测度。测度可以简单理解为面积与体积等可测量的量的泛化概念。

本书的构成

本书包含以下两大部分。

- 第1部分：聊聊概率这件事
- 第2部分：探讨概率的应用

其中，第2部分中的各个主题相对独立，读者可依个人兴趣与喜好有选择地阅读。

不过，这并不是这样分类的真正原因，这背后还有一个更为深层的理由，即“不希望读者混淆观测值与该观测值背后的生成机制”。

我们来看一个极为简单的例子。

- 背后的生成机制：算命箱中有七成是吉，三成是凶
- 观测值：5人抽签后，4人抽到了吉，1人抽到了凶

我们可以从 $7/10$ 与 $4/5$ 这两种角度来讨论吉凶的比例，但两者的含义与意义完全不同。我们可以直接观测到观测值，但无法观测该值生成的机制。要真正理解并应用统计方法，必须明确区分两者。

第1部分虽然会告诉读者生成机制，但仅讨论由该机制生成的观测值具有怎样的性质。在充分理解这部分内容之后，第2部分的重点将同时关注生成机制本身，讲解观测值怎样通过其背后的生成机制生成。我们姑且称前者为正问题，后者为逆问题。由于无法得到完整的信息，逆问题往往更难解决，且在讨论逆问题前，我们必须对正问题有深刻的理解。

现实问题往往属于逆问题，如果为了吸引读者而在正问题中混入逆问题，就很可能使读者感到混乱。为此，第1部分将集中讨论正问题，在读者对概率的基本概念了然于胸之后，再开始研究难度更大的逆问题。

0.3 为什么第1部分的例子多与赌局或游戏相关

上文所说的观测值与生成机制正是产生这种情况的原因。赌局与游戏背后的生成机制即是它们的规则，而这些规则我们可以事先知道。因此，它们非常适合作为第1部分的实例。

在线资源

读者可以从图灵社区的网站上下载在各章末专栏的实验中执行的 Ruby 脚本^①。

谢辞

原启介、高桥信、小山达树与堀英彰审读了本书的原稿，并对内容、编排与表述方式提供了很多有价值的建议。^②小柴健史给伪随机数部分提出了宝贵意见。使用过本书 Beta 版的学习小组也提供了很多有益的反馈。kogai、studio-rain、mia、pige 以及其他一些匿名网友为本书在互联网上公开的草稿提出了许多勘误与问题。欧姆社开发部的诸位从各方面为本书的写作提供了支持^③，使本书达到了仅凭笔者一人绝对无法达到的水准。mUDA 为本书的关键概念绘制了形象有趣又令人印象深刻的插图。笔者在此向诸位表示衷心的感谢。本书得以面世，离不开大家的努力（因笔者水平所限，本书难免还有不足之处，望读者见谅并批评）。

笔者

2009年10月

^① <http://www.ituring.com.cn/book/1254>，点击“随书下载”。——编者注

^② 阅读尚不成熟的原稿十分辛苦。多亏了几位的早期审读，笔者才能对包括本书整体框架在内的方方面面做出巨大改进。

^③ 大家除了准备邮件列表、版本管理系统、错误追踪系统等基本的软件环境，还提供了可以直接通过 make 指令确认运行结果的实验模拟程序，技术水平非常高。

第1部分 聊聊概率这件事

第1章 概率的定义	3
1.1 概率的数学定义	3
1.2 三扇门(蒙提霍尔问题)——飞艇视角	4
1.2.1 蒙提霍尔问题	5
1.2.2 正确答案与常见错误	6
1.2.3 以飞艇视角表述	6
1.3 三元组(Ω, \mathcal{F}, P)——上帝视角	9
1.4 随机变量	13
1.5 概率分布	17
1.6 适于实际使用的简记方式	19
1.6.1 随机变量的表示方法	19
1.6.2 概率的表示方法	20
1.7 Ω 是幕后角色	21
1.7.1 不必在意 Ω 究竟是什么	21
1.7.2 Ω 的习惯处理方式	22
1.7.3 不含 Ω (不含上帝视角)的概率论	23
1.8 一些注意事项	23
1.8.1 想做什么	23
1.8.2 因为是面积……	24
1.8.3 解释	26
第2章 多个随机变量之间的关系	29
2.1 各县的土地使用情况(面积计算的预热)	29
2.1.1 不同县、不同用途的统计(联合概率与边缘概率的预热)	30
2.1.2 特定县、特定用途的比例(条件概率的预热)	31
2.1.3 倒推比例(贝叶斯公式的预热)	32

2.1.4 比例相同的情况(独立性的预热)	34
2.1.5 预热结束	38
2.2 联合概率与边缘概率	38
2.2.1 两个随机变量	38
2.2.2 三个随机变量	41
2.3 条件概率	42
2.3.1 条件概率的定义	42
2.3.2 联合分布、边缘分布与条件分布的关系	45
2.3.3 即使条件中使用的不是等号也一样适用	50
2.3.4 三个或更多的随机变量	51
2.4 贝叶斯公式	55
2.4.1 问题设置	56
2.4.2 贝叶斯的作图曲	57
2.4.3 贝叶斯公式	61
2.5 独立性	63
2.5.1 事件的独立性(定义)	64
2.5.2 事件的独立性(等价表述)	67
2.5.3 随机变量的独立性	70
2.5.4 三个或更多随机变量的独立性(需多加注意)	73
第3章 离散值的概率分布	79
3.1 一些简单的例子	79
3.2 二项分布	82
3.2.1 二项分布的推导	82
3.2.2 补充: 排列 ${}_nP_k$ 、组合 ${}_nC_k$	83
3.3 期望值	85
3.3.1 期望值的定义	85
3.3.2 期望值的基本性质	87
3.3.3 期望值乘法运算的注意事项	91
3.3.4 期望值不存在的情况	93
3.4 方差与标准差	99
3.4.1 即使期望值相同	99

3.4.2	方差即“期望值离散程度”的期望值	100
3.4.3	标准差	102
3.4.4	常量的加法、乘法及标准化	104
3.4.5	各项独立时, 和的方差等于方差的和	108
3.4.6	平方的期望值与方差	110
3.5	大数定律	112
3.5.1	独立同分布	114
3.5.2	平均值的期望值与平均值的方差	116
3.5.3	大数定律	117
3.5.4	大数定律的相关注意事项	118
3.6	补充内容: 条件期望与最小二乘法	120
3.6.1	条件期望的定义	120
3.6.2	最小二乘法	121
3.6.3	上帝视角	122
3.6.4	条件方差	123
第4章	连续值的概率分布	127
4.1	渐变色打印问题(密度计算的预热)	128
4.1.1	用图表描述油墨的消耗量(累积分布函数的预热)	128
4.1.2	用图表描述油墨的打印浓度(概率密度函数预热)	129
4.1.3	拉伸打印成品对油墨浓度的影响(变量变换的预热)	133
4.2	概率为零的情况	136
4.2.1	出现概率恰好为零的情况	137
4.2.2	概率为零将带来什么问题	139
4.3	概率密度函数	140
4.3.1	概率密度函数	140
4.3.2	均匀分布	146
4.3.3	概率密度函数的变量变换	147
4.4	联合分布·边缘分布·条件分布	152
4.4.1	联合分布	152
4.4.2	本小节之后的阅读方式	155
4.4.3	边缘分布	155

4.4.4	条件分布	159
4.4.5	贝叶斯公式	162
4.4.6	独立性	163
4.4.7	任意区域的概率·均匀分布·变量变换	166
4.4.8	实数值与离散值混合存在的情况	174
4.5	期望值、方差与标准差	174
4.5.1	期望值	175
4.5.2	方差·标准差	179
4.6	正态分布与中心极限定理	180
4.6.1	标准正态分布	181
4.6.2	一般正态分布	184
4.6.3	中心极限定理	187
第5章	协方差矩阵、多元正态分布与椭圆	195
5.1	协方差与相关系数	196
5.1.1	协方差	196
5.1.2	协方差的性质	199
5.1.3	分布倾向的明显程度与相关系数	200
5.1.4	协方差与相关系数的局限性	206
5.2	协方差矩阵	208
5.2.1	协方差矩阵=方差与协方差的一览表	208
5.2.2	协方差矩阵的向量形式表述	209
5.2.3	向量与矩阵的运算及期望值	212
5.2.4	向量值随机变量的补充说明	215
5.2.5	协方差矩阵的变量变换	217
5.2.6	任意方向的发散程度	218
5.3	多元正态分布	220
5.3.1	多元标准正态分布	220
5.3.2	多元一般正态分布	223
5.3.3	多元正态分布的概率密度函数	228
5.3.4	多元正态分布的性质	230
5.3.5	截面与投影	232

5.3.6 补充知识: 卡方分布	239
5.4 协方差矩阵与椭圆的关系	242
5.4.1 (实例一)单位矩阵与圆	242
5.4.2 (实例二)对角矩阵与椭圆	244
5.4.3 (实例三)一般矩阵与倾斜的椭圆	247
5.4.4 协方差矩阵的局限性	251

第2部分 探讨概率的应用

第6章 估计与检验	257
6.1 估计理论	257
6.1.1 描述统计与推断统计	257
6.1.2 描述统计	258
6.1.3 如何理解推断统计中的一些概念	260
6.1.4 问题设定	264
6.1.5 期望罚款金额	265
6.1.6 多目标优化	266
6.1.7 (策略一)减少候选项——最小方差无偏估计	267
6.1.8 (策略二)弱化最优定义——最大似然估计	269
6.1.9 (策略三)以单一数值作为评价基准——贝叶斯估计	272
6.1.10 策略选择的相关注意事项	275
6.2 检验理论	276
6.2.1 检验理论中的逻辑	276
6.2.2 检验理论概述	278
6.2.3 简单假设	279
6.2.4 复合假设	282
第7章 伪随机数	285
7.1 伪随机数的基础知识	285
7.1.1 随机数序列	285
7.1.2 伪随机数序列	286

7.1.3 典型应用：蒙特卡罗方法	287
7.1.4 相关主题：密码理论中的伪随机数序列·低差异序列	289
7.2 遵从特定分布的随机数的生成	291
7.2.1 遵从离散值分布的随机数的生成	292
7.2.2 遵从连续值分布的随机数的生成	293
7.2.3 遵从正态分布的随机数的生成	296
7.2.4 补充知识：三角形内及球面上的均匀分布	298
第8章 概率论的各类应用	305
8.1 回归分析与多变量分析	305
8.1.1 通过最小二乘法拟合直线	305
8.1.2 主成分分析	312
8.2 随机过程	319
8.2.1 随机游走	321
8.2.2 卡尔曼滤波器	326
8.2.3 马尔可夫链	331
8.2.4 关于随机过程的一些补充说明	342
8.3 信息论	343
8.3.1 熵	343
8.3.2 二元熵	347
8.3.3 信源编码	349
8.3.4 信道编码	352
附录A 本书涉及的数学基础知识	359
A.1 希腊字母	359
A.2 数	359
A.2.1 自然数·整数	359
A.2.2 有理数·实数	359
A.2.3 复数	360
A.3 集合	360
A.3.1 集合的表述方式	360

A.3.2	无限集的大小	361
A.3.3	强化练习	361
A.4	求和符号 Σ	362
A.4.1	定义与基本性质	362
A.4.2	双重求和	364
A.4.3	范围指定	366
A.4.4	等比数列	366
A.5	指数与对数	368
A.5.1	指数函数	368
A.5.2	高斯积分	371
A.5.3	对数函数	374
A.6	内积与长度	377
附录 B	近似公式与不等式	381
B.1	斯特林公式	381
B.2	琴生不等式	381
B.3	吉布斯不等式	384
B.4	马尔可夫不等式与切比雪夫不等式	385
B.5	切尔诺夫界	386
B.6	闵可夫斯基不等式与赫尔德不等式	387
B.7	算术平均值 \geq 几何平均值 \geq 调和平均值	390
附录 C	概率论的补充知识	393
C.1	随机变量的收敛	393
C.1.1	依概率 1 收敛	393
C.1.2	依概率收敛	395
C.1.3	均方收敛	396
C.1.4	依分布收敛	396
C.2	特征函数	397
C.3	KL 散度与大偏差原理	399
参考文献		404