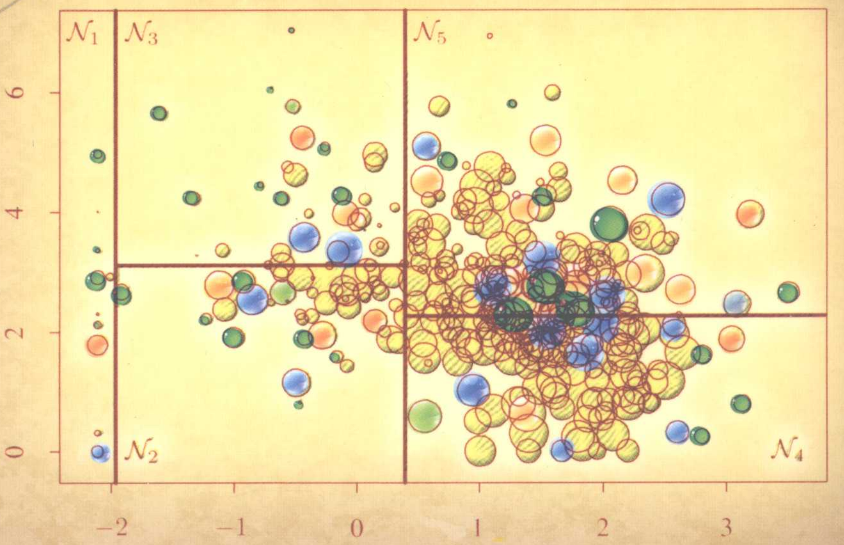
图灵数学 · 统计学丛书 37



Computational Statistics

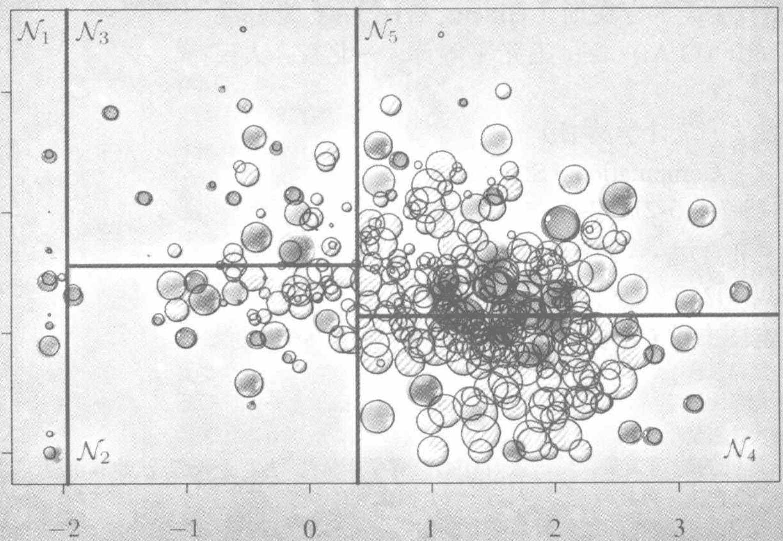
计算统计

[美] Geof H. Givens Jennifer A. Hoeting 著

王兆军 刘民千 邹长亮 杨建峰 译

人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵数学 · 统计学丛书 37



Computational Statistics

计算统计

[美] Geof H. Givens
Jennifer A. Hoeting 著
王兆军 刘民千 邹长亮 杨建峰 译

人民邮电出版社
北京

图书在版编目(CIP)数据

计算统计/（美）吉文斯（Givens, G.H.），（美）霍特伊 (Hoeting, J.A.) 著；王兆军等译. —北京：人民邮电出版社, 2009.9

（图灵数学 · 统计学丛书）

书名原文：Computational Statistics

ISBN 978-7-115-21182-8

Ⅰ.计··· Ⅱ.①吉··· ②霍··· ③王··· Ⅲ.数理统计-计算方法 Ⅳ. O212

中国版本图书馆 CIP 数据核字（2009）第 128672 号

内 容 提 要

随着计算机的快速发展，数理统计中许多涉及大计算量的有效方法也得到了广泛应用与迅猛发展，可以说，计算统计已是统计中一个很重要的研究方向.

本书既包含一些经典的统计计算方法，如求解非线性方程组的牛顿方法、传统的随机模拟方法等，又全面地介绍了近些年来发展起来的某些新方法，如模拟退火算法、基因算法、EM 算法、MCMC 方法、Bootstrap 方法等，并通过某些实例，对这些方法的应用进行了较详细的说明. 本书最后还提供了各种难度的习题.

本书可作为数学、统计学、科学计算等专业的本科生教材，也可供统计学方向的研究生、工程技术人员和应用工作者参考使用.

图灵数学 • 统计学丛书

计算统计

◆ 著　　[美] Geof H. Givens　Jennifer A. Hoeting
　译　　王兆军　刘民千　邹长亮　杨建峰
　责任编辑　明永玲
　执行编辑　边晓娜

◆ 人民邮电出版社出版发行　　北京市崇文区夕照寺街 14 号
　邮编 100061　电子函件　315@ptpress.com.cn
　网址：http://www.ptpress.com.cn
　北京铭成印刷有限公司印刷

◆ 开本：700×1000　1/16
　印张：22.5
　字数：454 千字　　2009 年 9 月第 1 版
　印数：1–3 000 册　　2009 年 9 月北京第 1 次印刷

著作权合同登记号　图字：01-2005-5223 号

ISBN 978-7-115-21182-8/O1

定价：59.00 元

读者服务热线：(010)51095186　印装质量热线：(010) 67129223

反盗版热线：(010)67171154

版 权 声 明

译 者 简 介

王兆军 南开大学教授、博士生导师. 现任南开大学数学科学学院副院长、中国概率统计学会常务理事、中国现场统计研究会理事、天津市现场统计研究会副理事长、天津数学会秘书长.

刘民千 南开大学教授、博士生导师, 教育部新世纪优秀人才. 现任中国数学会均匀设计分会副理事长、中国现场统计研究会试验设计分会副理事长、天津市现场统计研究会秘书长.

邹长亮 南开大学数学科学学院在读博士研究生.

杨建峰 南开大学数学科学学院统计系教师.

译　者　序

统计计算不仅是统计学专业本科生的一门重要基础课程, 而且越来越多的理工科、商学、经济学、医学专业本科生及研究生也都开始选修此课程. 虽然国内关于统计计算的教材已有若干本, 但这些教材多是介绍传统的、经典的统计计算方法. 近些年, 随着计算机技术的快速发展和统计方法的不断丰富, 统计计算方法发展很快, 并大受重视, 产生了许多得到广泛应用的统计计算方法, 如 EM 算法、Bootstrap 方法、MCMC 方法、模拟退火方法等. 然而, 到目前为止, 国内还没有一本系统地介绍这些新方法的统计计算教材或专著, 而这本由 Wiley 出版社出版的《计算统计》恰好填补了这一空白.

本书既包含了一些经典的统计计算方法, 如非线性方程组的求解方法、传统的 Monte-Carlo 方法等, 也详细地介绍了近些年发展起来的许多常用统计计算方法, 如模拟退火算法、遗传算法、EM 算法、MCMC 方法、Bootstrap 方法及某些光滑技术等.

本书在讲述方法的同时, 还注重这些方法在金融、优化等方面的应用, 并给出了非常丰富的参考文献. 另外, 虽然全书内容较丰富, 但因其所需的概率统计知识相对较少, 所以很适合低年级本科生自学或课堂学习, 而且其中某些高等内容也可供统计专业的本科生、研究生参考.

我们很高兴能有机会将该书推荐给国内的读者, 也非常感谢人民邮电出版社图灵公司的编辑在此书翻译过程中给予我们的大力支持和帮助.

本书的翻译工作由 4 名老师合作完成, 其中第 1~3 章由王兆军翻译, 第 4~6 章由刘民千翻译, 第 7~9 章由邹长亮翻译, 第 10~12 章由杨建峰翻译, 全书由王兆军、刘民千统校.

由于译者英文、中文水平有限, 专业知识也有待提高, 翻译之中难免会有不妥之处, 欢迎广大读者批评指正.

译　者

2008 年 8 月于南开园

前　言

要广泛深入地学习当今统计计算和计算统计学, 所需了解的大多数内容本书均有涉及. 我们力求让读者理解现有方法的机理, 使读者能够有效地使用这些现代统计方法. 由于许多新方法都是从现有的技术构建出来的, 故我们的最终目的是向科学工作者提供必要的工具, 帮助他们为此领域贡献新的思想.

想要达到这些目的, 就必须精通统计计算、计算统计、计算机科学和数值分析等各方面内容. 我们选取了那些我们认为是本领域中心的内容, 也会是读者感兴趣和认为有用的内容. 另外, 我们从注重实效的角度优先考虑了使学生和研究者受益最多、收效最快的内容.

考虑到出现了一些高质量的软件, 我们省略了本领域过往以来的某些重要的研究内容. 例如, 伪随机数的产生是一个经典的课题, 但我们更倾向于让学生使用可靠的软件来解决问题. 还有一些内容如数值线性代数, 属于讲与不讲两可. 这些内容对于很多应用来说是很关键的, 但是通常都有不错的计算机软件可用. 按我们的判断, 人们不会经常抛开程序而去探究数值线性代数的细节, 因而 (刚好) 不足以让我们把这些内容写到书里. 这些经典内容我们只写了优化和数值积分, 这么做的原因是: (i) 二者是频率学派和 Bayes 推断的基石; (ii) 现有软件程序往往不能应付此方面的难题; (iii) 这些方法本身还是其他统计计算方法的基础.

我们这里使用 "现代" 这个字眼, 可能面临如下矛盾, 其实这本书不可能囊括所有的最新、最好的技术. 事实上, 我们也从未打算这么做. 有些领域实在变化得太快, 比如启发式搜索和 MCMC. 我们只是努力提供这些领域主要内容的近期概况, 而把其多样性和专业性让读者自己去探索回味. 还有的内容 (如主曲线和 tabu 搜索) 我们写在书中, 仅仅是因为这些内容很有意思, 可以从全新的角度去看熟悉的问题. 也许研究者将来能从这些内容出发设计出有创意且有效的新算法来.

本书的目标读者为统计和相关专业的研究生、应用统计工作者和其他领域做定量分析的科学工作者. 我们希望这些读者在应用标准方法和研发新方法的时候, 能够用到本书.

本书不要求读者具有高深的数学水平, 但要了解 Taylor 级数和线性代数方面的知识. 读者数学训练的广度比深度更有用. 第 1 章回顾了基础知识, 较高级的读者可以在与具体内容相关的很多其他书中找到更多的数学细节, 我们在书中列举出了这些参考文献. 其他读者如果对分析的细节不太关注, 则看懂本书的算法和例子讲解就够了.

本书要求的统计知识仅限于一年级研究生所学的统计和概率论内容, 其中最重要的基础知识是极大似然方法、Bayes 方法、基本渐近理论、马氏链和线性模型. 大部分这些内容都会在第 1 章提到.

至于计算机编程, 我们发现好学生可以按需自学. 当然, 了解一门合适的语言有助于快速地把本书中的概念加以实现. 我们在书中摈弃了那些针对具体语言的例子、算法和编码. 对于那些在学习本书的同时还想学习语言的人, 建议他们选择一个高水平的交互式软件包, 即可以灵活设计图形化显示并包含基本的统计和概率函数的软件包. 目前在本书写作阶段, 我们推荐使用 S-Plus,R 和 MATLAB①, 这些都是研究人员在开发新的统计计算技术时经常用到的软件, 也适用于实现我们书中描述的绝大部分方法, 除了个别特大型复杂问题以外. 当然, 你也可以用一些低级语言如 C++, 在研究人员把方法琢磨好后, 通常可以用低级语言把它们作成一个专业版的软件.

即使是编程的老手, 对于数学运算是如何在计算机的二进制世界里实现的细节, 也可能不甚了解. 各种稀奇古怪的问题其实并不少见, 比如满秩矩阵似乎不可逆, 积分和似然是退化的, 数值近似比实际情况还精确等等. 我们一方面不能忽视计算机运算和稳定的数值计算的重要性, 另一方面更要重视算法原理的大局观, 而不去拘泥某些数值计算的细枝末节.

本书共分为 3 个主要部分: 优化 (第 2 章到第 4 章), 积分 (第 5 章到第 8 章), 光滑技术 (第 10 章到第 12 章). 第 9 章穿插介绍了另一个重要内容 Bootstrap 方法. 每章的内容都是独立的, 老师可以根据课程需要自由选取章节. 如果是一学期的课程, 通常我们选取第 2 章, 第 5 章到第 7 章, 第 9 章到第 11 章. 如果想讲得更为从容或深入, 还可以进一步缩小范围. 对于一学年的课程来说, 本书的内容也足够丰富, 何况老师可能还想讲些补充内容.

每章后面都有大量的课后作业. 有些题目直截了当, 但有些题目则需要学生对学过的模型或方法有深入的了解, 仔细 (甚至机灵) 地编写出适当的程序, 并且充分注重对于结果的分析.

正文和习题中涉及的数据集可以从本书网站获得: `www.stat.colostate.edu/computationalstatistics`. 网站上还有本书的勘误表. 作者对于书中的错误负全部责任.

Geof H.Givens, Jennifer A.Hoeting
于科罗拉多州福特科林斯

① 这些软件包的主页分别为: `www.insightful.com`,`www.r-project.org`和`www.mathworks.com`, 其中 R 是一种免费的能运行 S-Plus 部分功能的软件, 而其他的均为商业软件.

致　谢

我们借用了 Adrian Raftery 大量的知识, 在此特别致谢他并不仅仅是由于他的教学与指导, 而且还由于他坚定的支持和取之不尽的好思想. 另外, 我们要感谢华盛顿大学统计系那些极富影响力的导师们, 包括 David Madigan, Werner Stuetzle 和 Judy Zeh. 当然, 本书的每一章内容都能拓展成一本独立的书, 并且一些著名学者已这样做了. 我们这门课的讲授及本讲义的编写都依赖于这些学者的努力, 我们深深地感谢他们.

从 1994 年起, 我们就已在科罗拉多州立大学讲授基于本书内容的课程. 因此, 我们感谢统计系同事们的不断支持, 还要特别感谢我们职业生涯最初几年给予我们指导的已故 Richard Tweedie 先生. 我们也要感谢那些多年来寒窗听课的学生. 本书的部分内容是在新西兰的奥特加大学数学与统计系完成的, 期间我们受到了全体教员的热情款待.

John Bickham, Kate Cowles, Jan Hanning, Alan Herlihy, David Hunter, Devin Johnson, Michael Newton, Doug Nychka, Steve Sain, David W. Scott, N. Scott Urquhart, Haonan Wang 和 Darrell Whitley 及八位匿名审稿者的建设性意见极大地改进了原稿. 本书的出版还得到了本书编辑 Steve Quigley 及 Wiley 出版社的编辑们的支持与帮助. 我们要感谢 Nélida Pohl 允许我们采用她设计的封面. 我们还要感谢 Zube(又名 John Dzubera) 使我们的计算机始终能正常运转.

本书第一作者要感谢国家自然科学基金 (NSF)CAREER(资助号为 #SBR-9875508) 在本书写作过程中给予的大力支持, 也要感谢他在阿拉斯加北斯路普自治市 (North Slope Borough, Alaska) 野生动植物管理部门的同事与朋友们的长期研究的支持. 第二作者还十分感谢由美国环保局 (EPA) 授予科罗拉多州立大学的 STAR 研究助理协议的支持 (协议号为 CR-829095). 书中表述的仅是作者自己的观点, 所提到的产品或商业服务并没有得到 NSF 和 EPA 的核准.

最后, 我们要题献此书给我们的父母, 感谢他们能够让我们学习且支持我们学习, 感谢他们带给我们的韧力, 这些韧力是研究生阶段、获取终身教授职位及出版本书所必需的.

目　录

第 1 章　回　　顾

本章将回顾一些有关数学、概率和统计中的记号和背景资料. 读者可以跳过本章直接阅读第 2 章.

1.1　某些数学记号

为与一个常变量 x 或常数 M 相区别, 我们用黑体表示向量 $\boldsymbol{x} = (x_1, \cdots, x_p)$ 或矩阵 $\boldsymbol{M}$. 在点 $\boldsymbol{x}$ 取值的向量函数也是黑体, 即 $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_p(\boldsymbol{x}))$. 以 $\boldsymbol{M}^{\mathrm{T}}$ 表示矩阵 $\boldsymbol{M}$ 的转置.

除非特别指出, 所有向量均为列向量. 因此, 一个 $n \times p$ 阶矩阵可以写成 $\boldsymbol{M} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^{\mathrm{T}}$. 以 $\boldsymbol{I}$ 表示单位矩阵, $\mathbf{1}$ 和 $\mathbf{0}$ 分别表示 1 和 0 的向量.

如果对所有非零向量 $\boldsymbol{x}$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} > 0$, 则称对称方阵 $\boldsymbol{M}$ *正定*. 正定的等价条件是其所有的特征根为正. 如果对所有非零向量 $\boldsymbol{x}$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} \geqslant 0$, 则称 $\boldsymbol{M}$ *非负定或半正定*.

记函数 f 在点 x 的导数为 $f'(x)$. 当 $\boldsymbol{x} = (x_1, \cdots, x_p)$ 时, 函数 f 在 $\boldsymbol{x}$ 点的*梯度*为 $\boldsymbol{f}'(\boldsymbol{x}) = \left(\frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}x_1}, \cdots, \frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}x_p}\right)$. 函数 f 在 $\boldsymbol{x}$ 点的*Hessian 矩阵*记为 $\boldsymbol{f}''(\boldsymbol{x})$, 其第 (i,j) 元素为 $\frac{\mathrm{d}^2 f(\boldsymbol{x})}{\mathrm{d}x_i \mathrm{d}x_j}$. 负的 Hessian 阵在统计推断中具有重要的应用.

以 $\boldsymbol{J}(\boldsymbol{x})$ 表示一对一映射 $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ 在点 $\boldsymbol{x}$ 处的*Jacobian 矩阵*, 其第 (i,j) 元素为 $\frac{\mathrm{d}f_i(\boldsymbol{x})}{\mathrm{d}x_j}$.

一个*泛函*就是一个函数空间中的实值函数. 例如, 如果 $T(f) = \displaystyle\int_0^1 f(x)\mathrm{d}x$, 则泛函 T 为可积函数到一维实数的映射.

示性函数 $1_{\{A\}}$ 等于 1 , 如果 A 成立, 否则就等于 0. 一维实空间记为 $\Re$, p 维实空间记为 $\Re^p$.

1.2　Taylor 定理和数学极限理论

为了描述函数收敛的相对阶数, 我们首先定义记号 O 与 o. 设 f, g 为两个定义在同一区间 (区间可能无限) 上的函数, z_0 为此区间内或边界上一点 (即 $-\infty$ 或 ∞). 我们要求函数 $g(z) \neq 0$, 其中 在 z_0 的一个邻域内 $z \neq z_0$. 如果存在一个常数

M 满足：当 $z \to z_0$ 时，$|f(z)| \leqslant M|g(z)|$，则称

$$f(z) = O(g(z)). \tag{1.1}$$

例如，当 $n \to \infty$ 时，$\frac{n+1}{3n^2} = O(n^{-1})$. 如果 $\lim\limits_{z \to z_0} f(z)/g(z) = 0$，则称

$$f(z) = o(g(z)). \tag{1.2}$$

例如，如果 f 在 x_0 点可微，则当 $h \to 0$ 时，$f(x_0+h) - f(x_0) = hf'(x_0) + o(h)$. 如取 $f(n) = x_n$，则关于序列 $\{x_n\}$ 的收敛性，同样有上述记号.

Taylor 定理给出了一个函数 f 的多项式近似. 设 f 在区间 (a,b) 上具有有限的 $(n+1)$ 阶导数，在区间 $[a,b]$ 上有连续的 n 阶导数. 则对于任意一个不同于 x 的 $x_0 \in [a,b]$，函数 f 在点 x_0 的 Taylor 级数展开为

$$f(x) = \sum_{i=1}^{n} \frac{1}{i!} f^{(i)}(x_0)(x-x_0)^i + R_n, \tag{1.3}$$

其中 $f^{(i)}(x_0)$ 为函数 f 在点 x_0 处的 i 阶导数，且

$$R_n = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-x_0)^{n+1}, \tag{1.4}$$

其中 ξ 在由 x 与 x_0 构成的区间内. 注意到当 $|x-x_0| \to 0$ 时，$R_n = O(|x-x_0|^{n+1})$.

多元的 Taylor 定理与之类似. 设 f 为一关于 $\boldsymbol{x}$ 的 p 元实值函数，它在包含 $\boldsymbol{x}$ 和 $\boldsymbol{x}_0 \neq \boldsymbol{x}$ 的一个开的凸集中具有 $n+1$ 阶连续偏导数. 则

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{i=1}^{n} \frac{1}{i!} D^{(i)}(f; \boldsymbol{x}_0, \boldsymbol{x}-\boldsymbol{x}_0) + R_n, \tag{1.5}$$

其中

$$D^{(i)}(f; \boldsymbol{x}, \boldsymbol{y}) = \sum_{j_1=1}^{p} \cdots \sum_{j_i=1}^{p} \left\{ \left(\frac{\mathrm{d}^i}{\mathrm{d}t_{j_1} \cdots \mathrm{d}t_{j_i}} f(\boldsymbol{t})|\boldsymbol{t} = \boldsymbol{x} \right) \prod_{k=1}^{i} y_{j_k} \right\}, \tag{1.6}$$

$$R_n = \frac{1}{(n+1)!} D^{(n+1)}(f; \boldsymbol{\xi}, \boldsymbol{x}-\boldsymbol{x}_0), \tag{1.7}$$

其中 $\boldsymbol{\xi}$ 在由点 $\boldsymbol{x}$ 和 $\boldsymbol{x}_0$ 连成的直线段上. 当 $|\boldsymbol{x}-\boldsymbol{x}_0| \to 0$ 时，$R_n = O(|\boldsymbol{x}-\boldsymbol{x}_0|^{n+1})$.

Euler-Maclaurin 公式在渐近分析中很有用. 如果 f 在 [0,1] 上具有 $2n$ 阶连续导数，则

$$\int_0^1 f(x)\mathrm{d}x = \frac{f(0)+f(1)}{2} - \sum_{i=0}^{n-1} \frac{b_{2i}(f^{(2i-1)}(1) - f^{(2i-1)}(0))}{(2i)!} - \frac{b_{2n} f^{(2n)}(\xi)}{(2n)!}, \tag{1.8}$$

其中 $0 \leqslant \xi \leqslant 1$，$f^{(j)}$ 是 f 的 j 阶导数，$b_j = B_j(0)$ 由下列迭代关系确定

$$\sum_{j=0}^{m} \binom{m+1}{j} B_j(z) = (m+1)z^m, \tag{1.9}$$

其初值 $B_0(z) = 1$. 此结论可由分部积分证得 ([328]).

最后, 我们注意到有时会利用有限差分来数值近似一个函数的导数. 例如, 函数 f 在点 $\boldsymbol{x}$ 处梯度的第 i 个分量为

$$\frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}x_i} \approx \frac{f(\boldsymbol{x}+\epsilon_i\boldsymbol{e}_i) - f(\boldsymbol{x}-\epsilon_i\boldsymbol{e}_i)}{2\epsilon_i}, \tag{1.10}$$

其中 ϵ_i 是一任意小数, $\boldsymbol{e}_i$ 是第 i 个梯度方向的单位向量. 一般地, 人们可从 $\epsilon_i = 0.01$ 或 0.001 开始, 采用逐步减少的 ϵ_i 序列来近似所求导数. 且这种近似方法一般均可逐步得到改进, 直到当 ϵ_i 非常小时导致计算退化且计算完全由计算机的四舍五入所控制. 关于此方法的介绍和可以得到较高精度的 Richadson 外推法请见 [328]. 有限差分法仍可用来近似函数 f 在 $\boldsymbol{x}$ 处的二阶导数, 即

$$\begin{aligned}\frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}x_i\mathrm{d}x_j} \approx \frac{1}{4\epsilon_i\epsilon_j}&\left(f(\boldsymbol{x}+\epsilon_i\boldsymbol{e}_i+\epsilon_j\boldsymbol{e}_j) - f(\boldsymbol{x}+\epsilon_i\boldsymbol{e}_i-\epsilon_j\boldsymbol{e}_j)\right.\\ &\left.-f(\boldsymbol{x}-\epsilon_i\boldsymbol{e}_i+\epsilon_j\boldsymbol{e}_j) + f(\boldsymbol{x}-\epsilon_i\boldsymbol{e}_i-\epsilon_j\boldsymbol{e}_j)\right),\end{aligned} \tag{1.11}$$

它仍可用类似的 ϵ_i 序列来改进近似精度.

1.3 某些统计记号和概率分布

我们用大写字母表示随机变量, 如 Y 或 X; 用小写字母表示随机变量的取值, 如 y 或 x. 记 f 和 F 分别为 X 的概率密度函数和累积分布函数. 我们以记号 $X \sim f(x)$ 表示 X 服从密度为 $f(x)$ 的分布. 一般地, 以一条竖线, 如 $f(x|\alpha,\beta)$ 表示密度函数 $f(x)$ 依赖于一个或多个参数. 由于本书内容较多, 故应注意到 $f(x|\alpha)$ 也表示此密度函数在 x 处的取值. 当所用记号的含义清楚时, 我们则不加以区别, 如 $f(\cdot|\alpha)$ 就表示此函数. 当有多个随机变量的密度需要加以区别时, 可加下标以示区别, 即分别用 f_X 和 f_Y 表示 X 和 Y 的密度函数. 对于离散随机变量和有关 Bayes 的内容, 我们使用同样的记号.

给定 $Y = y$ 时 X 的条件密度记为 $f(x|y)$ 或 $f_{X|Y}(x|y)$, 此时也称 $X|Y$ 具有密度 $f(x|Y)$. 为了记号的简单, 我们允许密度函数由其变量所决定, 于是, 我们可以用同一个记号, 如 f 表示不同的函数, 如下面的方程: $f(x,y|\mu) = f(x|y,\mu)f(y|\mu)$. 最后, $f(X)$ 和 $F(X)$ 均是随机变量, 它表示密度函数和累积分布函数在随机自变量 X 处的取值.

以 $\mathrm{E}\{X\}$ 表示随机变量的期望. 除非特别指出, 求期望所用的分布均指 X 的分布. 我们以 $P[A]$ 表示事件 A 的概率, 且 $P[A] = \mathrm{E}\{1_{\{A\}}\}$. 用 $\mathrm{E}\{X|y\}$ 表示 $X|Y = y$

的期望. 当 Y 未知时, $\mathrm{E}\{X|Y\}$ 是依赖于 Y 的随机变量. 关于 X 和 Y 的其他分布特征有 $\mathrm{var}\{X\}, \mathrm{cov}\{X,Y\}, \mathrm{cor}\{X,Y\}$ 和 $\mathrm{cv}\{X\} = \mathrm{var}\{X\}^{1/2}/\mathrm{E}\{X\}$, 它们分别表示 X 的方差、X 和 Y 的协方差和 X 的变异系数.

Jensen 不等式是关于期望的一个有用结果. 设 g 在某可能无限的开区间 I 内是凸函数, 则对于所有的 $x, y \in I$ 和 $0 < \lambda < 1$, 有

$$g(\lambda x + (1-\lambda)y) \leqslant \lambda g(x) + (1-\lambda)g(y). \tag{1.12}$$

Jensen 不等式指出, 如果随机变量 X 满足 $P[X \in I] = 1$, 则 $\mathrm{E}\{g(X)\} \geqslant g(\mathrm{E}\{X\})$.

表 1.1, 表 1.2 和表 1.3 给出了本书中用到的多个离散和连续随机变量的相关信息. 我们有如下常用的组合常数:

表 1.1 某些常用离散随机变量概率分布的记号和描述

名 称	记号和参数空间	密度和样本空间	均值与方差
Bernoulli	$X \sim \mathrm{B}(p)$	$f(x) = p^x(1-p)^{1-x}$	$\mathrm{E}\{X\} = p$
	$0 \leqslant p \leqslant 1$	$x = 0$ 或 1	$\mathrm{var}\{X\} = p(1-p)$
二项	$X \sim \mathrm{B}(n,p)$	$f(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$\mathrm{E}\{X\} = np$
	$0 \leqslant p \leqslant 1, n = 1, 2, \cdots$	$x = 0, 1, \cdots, n$	$\mathrm{var}\{X\} = np(1-p)$
多项	$\boldsymbol{X} \sim \mathrm{MB}(n, \boldsymbol{p})$	$f(\boldsymbol{x}) = \binom{n}{x_1, \cdots, x_k} \prod_{i=1}^k p_i^{x_i}$	$\mathrm{E}\{\boldsymbol{X}\} = n\boldsymbol{p}$
	$\boldsymbol{p} = (p_1, \cdots, p_k)$, $0 \leqslant p_i \leqslant 1$	$\boldsymbol{x} = (x_1, \cdots, x_k), x_i = 0, 1, \cdots, n$	$\mathrm{var} X_i = np_i(1-p_i)$
	$\sum_{i=1}^k p_i = 1, n = 1, 2, \cdots$	$\sum_{i=1}^k x_i = n$	$\mathrm{cov}\{X_i, X_j\} = -np_ip_j$
负二项	$X \sim \mathrm{NB}(r,p)$	$f(x) = \binom{x+r-1}{r-1} p^r(1-p)^x$	$\mathrm{E}\{X\} = r(1-p)/p$
	$0 \leqslant p \leqslant 1, r = 1, 2, \cdots$	$x = 0, 1, \cdots$	$\mathrm{var}\{X\} = r(1-p)/p^2$
Poisson	$X \sim \mathrm{P}(\lambda)$	$f(x) = \frac{\lambda^x}{x!}\exp\{-\lambda\}$	$\mathrm{E}\{X\} = \lambda$
	$\lambda > 0$	$x = 0, 1, 2, \cdots$	$\mathrm{var} X = \lambda$

$$n! = n(n-1)(n-2)\cdots(3)(2)(1), \ (\text{注意 } 0! = 1), \tag{1.13}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \tag{1.14}$$

$$\binom{n}{k_1 \cdots k_m} = \frac{n!}{\prod_{i=1}^m k_i!}, \ \text{其中 } n = \sum_{i=1}^m k_i, \tag{1.15}$$

$$\Gamma(r) = \begin{cases} (r-1)!, & \text{如果} r = 1, 2, \cdots, \\ \int_0^\infty t^{r-1}\exp\{-t\}\mathrm{d}t, & \text{如果 } r > 0. \end{cases} \tag{1.16}$$

注意到 $\Gamma(1/2) = \sqrt{\pi}$, 且对于任意的正整数 n, $\Gamma(n+\frac{1}{2}) = \frac{1\times3\times5\times\cdots\times(2n-1)\sqrt{\pi}}{2^n}$.

表 1.2 某些常用连续随机变量概率分布的记号和描述

名 称	记号和参数空间	密度和样本空间	均值与方差
Beta	$X \sim \mathrm{Beta}(\alpha,\beta)$	$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\mathrm{E}\{X\}=\frac{\alpha}{\alpha+\beta}$
	$\alpha>0,\beta>0$	$0\leqslant x\leqslant 1$	$\mathrm{var}\{X\}=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Cauchy	$X\sim \mathrm{Cauchy}(\alpha,\beta)$	$f(x)=\frac{1}{\pi\beta\left[1+\left(\frac{x-\alpha}{\beta}\right)^2\right]}$	$\mathrm{E}\{X\}$ 不存在
	$\alpha\in\Re,\beta>0$	$x\in\Re$	$\mathrm{var}\{X\}$ 不存在
χ^2	$X\sim\chi^2_\nu$	$f(x)=\mathrm{Gamma}(\nu/2,1/2)$	$\mathrm{E}\{X\}=\nu$
	$\nu>0$	$x>0$	$\mathrm{var}\{X\}=2\nu$
Dirichlet	$\boldsymbol{X}\sim\mathrm{Dirichlet}(\boldsymbol{\alpha})$	$f(\boldsymbol{x})=\frac{\Gamma(\alpha_0)\prod_{i=1}^k x_i^{\alpha_i-1}}{\prod_{i=1}^k\Gamma(\alpha_i)}$	$\mathrm{E}\{\boldsymbol{X}\}=\boldsymbol{\alpha}/\alpha_0$
	$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_k)$	$\boldsymbol{x}=(x_1,\cdots,x_k),0\leqslant x_i\leqslant 1$	$\mathrm{var}\{X_i\}=\frac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(\alpha_0+1)}$
	$\alpha_i>0,\alpha_0=\sum_{i=1}^k\alpha_i$	$\sum_{i=1}^k x_i=1$	$\mathrm{cov}\{X_1,X_j\}=\frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$
指数	$X\sim\mathrm{Exp}(\lambda)$	$f(x)=\lambda\exp\{-\lambda x\}$	$\mathrm{E}\{X\}=1/\lambda$
	$\lambda>0$	$x>0$	$\mathrm{var}\{X\}=1/\lambda^2$
Gamma	$X\sim\mathrm{Gamma}(r,\lambda)$	$f(x)=\frac{\lambda^r x^{r-1}}{\Gamma(r)}\exp\{-\lambda x\}$	$\mathrm{E}\{X\}=r/\lambda$
	$\lambda>0,r>0$	$x>0$	$\mathrm{var}\{X\}=r/\lambda^2$

表 1.3 其他一些常用连续随机变量概率分布的记号和描述

名 称	记号和参数空间	密度和样本空间	均值与方差
对数正态	$X\sim\mathrm{LN}(\mu,\sigma^2)$	$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{\log\{x\}-\mu}{\sigma}\right)^2\right\}$	$\mathrm{E}\{X\}=\exp\{\mu+\sigma^2/2\}$
	$\mu\in\Re,\sigma>0$	$x\in\Re$	$\mathrm{var}\{X\}=\exp\{2\mu+2\sigma^2\}$ $-\exp\{2\mu+\sigma^2\}$
多元正态	$\boldsymbol{X}\sim\mathrm{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$	$f(\boldsymbol{x})=\frac{\exp\{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\}}{(2\pi)^{k/2}\lvert\boldsymbol{\Sigma}\rvert^{1/2}}$	$\mathrm{E}\{\boldsymbol{X}\}=\boldsymbol{\mu}$
	$\boldsymbol{\mu}=(\mu_1,\cdots,\mu_k)\in\Re^k$ $\boldsymbol{\Sigma}$ 正定	$\boldsymbol{x}=(x_1,\cdots,x_k)\in\Re^k$	$\mathrm{var}\{\boldsymbol{X}\}=\boldsymbol{\Sigma}$
正态	$X\sim\mathrm{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	$\mathrm{E}\{X\}=\mu$
	$\mu\in\Re,\sigma>0$	$x\in\Re$	$\mathrm{var}\{X\}=\sigma^2$
学生 -t	$X\sim t_\nu$	$f(x)=\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1+x^2/\nu)^{-(\nu+1)/2}$	$\mathrm{E}\{X\}=0(\nu>1)$
	$\nu>0$	$x\in\Re$	$\mathrm{var}\{X\}=\frac{\nu}{\nu+2}(\nu>2)$
均匀	$X\sim\mathrm{U}(a,b)$	$f(x)=\frac{1}{b-a}$	$\mathrm{E}\{X\}=(a+b)/2$
	$a,b\in\Re,a<b$	$x\in[a,b]$	$\mathrm{var}\{X\}=(b-a)^2/12$
Weibull	$X\sim\mathrm{Weibull}(a,b)$	$f(x)=abx^{b-1}\exp\{-ax^b\}$	$\mathrm{E}\{X\}=\frac{\Gamma(1+1/b)}{a^{1/b}}$
	$a>0,b>0$	$x>0$	$\mathrm{var}\{X\}=\frac{\Gamma(1+2/b)-\Gamma(1+1/b)^2}{a^{2/b}}$

统计中常用的多数分布都属于指数分布族. 一个具有 k 个参数的指数分布族函数可表示成

$$f(x|\gamma)=c_1(x)c_2(\gamma)\exp\left\{\sum_{i=1}^k y_i(x)\theta_i(\gamma)\right\},\tag{1.17}$$

其中 c_1,c_2 为非负函数; 向量 γ 为常用参数, 如 Poisson 分布中的 λ 及二项分布中