图灵数学·统计学丛书 18

WILEY

An Introduction to Probability Theory and Its Applications

概率论及其应用

（第 2 卷 · 第 2 版）

[美] 威廉·费勒 著

郑元禄 译

人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵数学·统计学丛书 18

An Introduction to Probability Theory and Its Applications

概率论及其应用

（第2卷·第2版）

[美] 威廉·费勒 著
郑元禄 译

人民邮电出版社
北 京

图书在版编目（CIP）数据

概率论及其应用：第 2 版，第 2 卷 /（美）费勒著；郑元禄译. —北京：人民邮电出版社，2008.1
（图灵数学·统计学丛书）
ISBN 978-7-115-16735-4

Ⅰ. 概… Ⅱ. ①费…②郑… Ⅲ. 概率论 Ⅳ.O211

中国版本图书馆 CIP 数据核字（2007）第 133313 号

内 容 提 要

本书是威廉·费勒的著作《概率论及其应用》第 1 卷的续篇. 曾经影响了包括中国在内的世界各国几代概率论及其相关领域的学生和研究者. 即使用今天的标准来衡量，该书仍是一本经典佳作. 本书包括各种重要的分布和随机过程、大数定律、中心极限定理、无穷可分分布、半群方法与无穷可分分布和马尔可夫过程的关系、更新理论、随机游动及傅里叶方法的应用、拉普拉斯变换及其应用、特征函数以及调和分析等 19 章内容.

本书既可作为概率论及相关学科的教学参考书，亦可作为相关科学研究的引导书.

图灵数学·统计学丛书

概率论及其应用（第 2 卷·第 2 版）

◆ 著 [美] 威廉·费勒
译 郑元禄
责任编辑 明永玲

◆ 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
北京隆昌伟业印刷有限公司印刷
新华书店总店北京发行所经销

◆ 开本：700×1000 1/16
印张：38.25
字数：771 千字 2008 年 1 月第 1 版
印数：1 – 4 000 册 2008 年 1 月北京第 1 次印刷

著作权合同登记号 图字：01-2007-2131 号

ISBN 978-7-115-16735-4/O1

定价：79.00 元

读者服务热线：(010)88593802 印装质量热线：(010)67129223

反盗版热线：(010) 67171154

版权声明

译 者 序

威廉 · 费勒 (William Feller) 著的《概率论及其应用》第 2 卷第 1 版出版于 1966 年, 第 2 版出版于 1971 年. 此译本是根据原著第 2 卷第 2 版翻译的.

该书论述了连续空间上的概率论及其应用. 理论内容包括: 测度论基础、概率分布、基本极限定理、特征函数、大数定律、中心极限定理、无穷可分分布、随机过程、更新理论、半群方法、傅里叶方法、拉普拉斯变换和调和分析等. 应用内容包括: 概率论在物理学、化学、生物学、医学、遗传学、天文学、对策论、排队论、数理统计、交通运输、电信工程、经济学、人口学和统计学等领域中的应用.

该书的内容十分丰富, 论述极其精辟, 行文优美生动. 原著 (共 2 卷) 已经问世 50 多年, 风行全世界, 培养和教育了许多国家不计其数的概率论和有关领域的专家学者, 对概率论的教学、科研、普及和应用做出了卓越的贡献.

该书多年前曾在科学出版社出版过, 王梓坤院士在百忙之中审校了此书, 并评价: "这本书是世界一流水平的概率论经典巨著, 作者不愧是世界概率论大师." 原译稿曾经李志阐教授审校, 并提出了许多宝贵意见. 我再次把该书译成中文出版, 希望能为我国高等学校和科研机构的教师、大学生、研究生和研究人员提供一部优秀的概率论教学参考书. 限于译者水平, 译本难免有缺点和错误, 敬请读者批评指正.

郑元禄

2007 年 6 月

郑元禄　福建省泉州五中高级教师, 视力残疾人. 因家庭问题未能进入高等学府深造, 自学成才. 1986 年获福建省自学成才奖, 1991 年获全国自学成才荣誉证, 1993 年获中国科学出版社基金. 多年来在泉州各高校和中学授课之余, 坚持翻译不辍, 从事英、俄文翻译 50 年, 发表数理译著 500 万字, 其中包括《高等数学例题与习题集 (四) 常微分方程》、《断裂力学》和《新英汉数学词汇》等, 并在《应用数学与计算数学》《数学通报》和《数理天地》等多家杂志上发表译文 200 篇左右. 1963~1983 年间担任中国科学技术情报研究所《数学文摘 (概率论与数理统计)》和《力学译丛》等杂志的翻译工作. 生平事迹被收入《中国科技翻译家辞典》、《中国教育专家名典》和美国《世界名人录》等.

第 1 版 序

在本书第 1 卷写作期间 (1941 年和 1948 年之间)，概率尚未得到人们的普遍关注. 概率的教学工作还只在非常有限的规模上开展，如马尔可夫链这样的专题，现在已经广泛应用于一些学科中，而那时却是纯粹数学中很专的篇章. 因此，第 1 卷可以比作是一本去一个陌生国家旅行的通用指南. 为了说清概率的本质，它着重于理论的数学内容以及各种各样潜在的应用. 曾经有人预言，本书难度上的起伏会局限它的使用. 实际上，它甚至在今天还被广泛使用，虽然它的新颖已经渐渐消逝，其形式和内容也为更多新书所借鉴. 本书似乎还得到了一些新读者. 学数学的学生存在困难的一些章节，却难不住外行人，这个事实表明，难度不能客观地予以判断，它依赖于我们要找的资料类型和我们准备忽略的细节. 为了到达山顶，旅行者可以选择自己攀登，也可以选择乘坐缆车.

鉴于第 1 卷的成功，第 2 卷也以同样的风格撰写. 它包括了较难的数学内容，但是多数读者可以看懂其中大部分内容. 测度论的处理可以说明这一点. 第 4 章包括了对测度论基本概念的简略介绍和概率概念的基础. 这一章列举了以后各章所用的少量测度论事实，以便用最简单的形式表达分析定理，并避免对正则条件作没有价值的讨论. 测度论在这方面的主要作用是说明形式运算和到极限的过渡，而非数学专业的读者们从未对这些内容提出过问题. 因此，主要对实际结果感兴趣的读者并未感到对测度论有任何需求.

为了更好地理解各个专题，我们使各章内容尽可能是独立的，一些特殊情形往往在一般理论前分别论述. 各种专题 (例如稳定分布和更新理论) 从不同角度在几个地方讨论. 为了避免重复，定义和说明性例子都集中在第 6 章，这一章可以被称为以下各章的共同导言. 本书的关键是第 5、8 和 15 章. 读者可自行决定阅读哪些准备性章节，并确定采用哪一条阅读路径.

专家们将发现新的结果和证明，但更重要的是努力把一般方法论巩固并统一起来. 实际上，概率的某些作用受损于缺乏条理性，因为问题的一般分类和论述基本上是依赖于历史发展的偶然性. 在由此引起的混乱中，密切相关的问题未被我们看出，简单的东西被复杂的方法弄得难以理解. 最佳技巧的系统应用和改进可带来大量的简化. 这特别对于众所周知的凌乱的极限定理是如此 (第 16 章和第 17 章). 此外，简化可由顺应自然的前后关系去处理问题而实现. 例如，对特殊随机游动的初等研究就导致了渐近估计的推广，而渐近估计在风险理论中原来是用困难又麻烦的方法推导出来的 (在排队论中，则在更严格的条件下独立地推导出来).

我力求不拘泥于文字而达到数学的严密性. 例如，在我看来，说 “$1/(1+\xi^2)$ 是 $\frac{1}{2}e^{-|x|}$ 的特征函数” 似乎是以下语句合乎要求并合乎逻辑的缩写：“在点 ξ 上取值

$1/(1+\xi^2)$ 的函数是在点 x 上取值 $\frac{1}{2}e^{-|x|}$ 的函数的特征函数. ”

我担心简单的历史评论和引述不能给许多对概率论作出贡献的作者以公正的评价，而我只要有可能就尽力给予赞扬. 最初的著作的很多内容已经为更新的研究成果所取代，作为一项规则，参考文献通常只给出读者为了查阅补充资料可能需要的论文. 例如，参考文献没有列出我自己关于极限定理的著作，而叙述某实例的观察结果及其理论的一篇论文却被引用了，即使它不包含数学.[①] 在这些情况下，作者索引并不表示他们在概率论中的地位. 另一个困难是公正评价先驱者的工作，我们要把新的研究方向、新的途径和新的方法归功于这些先驱者的工作. 一些曾经被认为是独创的、深奥的定理，现在以简单的证明在更精确的结果中出现. 我们今天很难在历史的背景下去认识这些定理，它们曾是举足轻重的起步.

感谢美国陆军研究室支持我们在普林斯顿大学进行的概率论研究. 我得到了戈德曼 (J. Goldman)，皮特 (L. Pitt)，西尔维斯坦 (M. Silverstein)，特别是拉奥 (M. M. Rao) 的帮助. 他们删去许多不精确的内容和含糊之处. 所有各章都反复重写过几次，各章的初稿曾经在朋友们中间流传. 这样，我从埃里奥特 (J. Elliott)，平克哈姆 (R. S. Pinkham) 和沙韦奇 (L. J. Savage) 所提出的意见中得到教益. 我特别感谢杜布 (J. L. Doob) 和沃尔福威兹 (J. Wolfowitz) 的建议和批评. 柯西随机游动图是特罗特尔 (H. Trotter) 供给的. 印制是由麦科杜加尔 (H. McDougal) 夫人负责监督的，本书的出版在很大程度上应归功于她.

威廉 · 费勒

1965 年 10 月

作者辞世时手稿已经完成，但未收到校样. 感谢出版社指定专人负责校对工作，并编制了索引. 戈德曼、格伦堡、麦基恩、皮特和皮腾革分工合作，检验书中数学内容. 每个数学工作者都明白，这项任务工作量有多大. 我深深地感激他们，谢谢他们满怀爱心的劳动.

克拉拉 · 费勒

1970 年 5 月

① 这种方式也曾经应用于第 1 卷中，但是被后来的一些作者所误会；他们现在把该书中所用的方法归功于他们并不认识的早期科学家们.

引　言

本书的特点和编排方式仍然保持不变，但是整部书经过了彻底的修订. 完全改写了许多章 (特别是第 17 章)，并增加了部分章节. 在许多地方，叙述被简化为更高效的论证 (有时是新的论证). 一些新的材料被编入正文.

在写第 1 版时，我因为担心书的篇幅太长而顾虑重重. 很遗憾，这使我为缩短原文和节省版面而白白花费了好几个月的时间. 这个损失现在已经弥补. 为使阅读更加容易，我尽了很大的努力. 偶然的重复还使得直接阅读一些特殊章节变得容易，并且能在本书中阅读到某些与第 1 卷有关联的章节.

关于材料的编排方式，请见第 1 版序的第 2 段 (这里再次重提一下).

我感谢许多读者指出了错误和遗漏之处. 特别感谢芝加哥的海哈尔 (D. A. Hejhal)，他提出了详尽无遗的勘误表和涉及全书的修改建议.

威廉 · 费勒

1970 年 1 月

于新泽西州普林斯顿

符号与约定

时　刻　这一术语用来表示时间轴上的点，而时间则用来表示间隔和持续时间.(在随机过程中,“时间”一词负担太重.系统地使用 Riordan 首创的 “epoch” 可避免反复交替使用 “moment ”，“instant”，“point”.)

区　间　以横线表示：$\overline{a,b}$ 是开区间，$\overline{a,b}$ 是闭区间，半开区间以 $\overline{a,b}$ 和 $\overline{a,b}$ 表示. 这种记号也用于高维空间中. 关于向量记号和次序关系的约定放在 5.1 节 (亦在 4.2 节) 中. 符号 (a,b) 用来表示对偶和点.

$\mathbf{R}^1,\mathbf{R}^2,\mathbf{R}^r$　表示直线、平面和 r 维笛卡儿 (Cartesian) 空间.

▶　表示证明完毕或一组例子的结束.

$\mathfrak{n}$ 和 $\mathfrak{N}$　分别表示期望为 0、方差为 1 的正态密度和正态分布函数.

O,o 和 $\sim$　令 u 和 v 依赖于参数 x, x 趋于 a. 假定 v 是正的，我们记

$$\left.\begin{array}{l} u=O(v)\\ u=o(v)\\ u\sim v \end{array}\right\}\text{如果}\frac{u}{v}\left\{\begin{array}{l}\text{保持有界}\\ \longrightarrow 0\\ \longrightarrow 1\end{array}\right.$$

$f(x)U\{\mathrm{d}x\}$　对于这一符号参看 5.3 节.

关于博雷尔 (Borel) 集与贝尔 (Baire) 函数，参看第 5 章的有关内容.

目　录

注：左上角有星号的诸节对理解下文是不需要的，初读时可略去.

第 1 章　指数密度与均匀密度

1.1　引　　言

在本书第 1 卷中，我们反复讨论了由很多小项之和所定义的概率，并且利用下列形式的逼近：

$$\mathrm{P}\{a < X < b\} \approx \int_a^b f(x)\mathrm{d}x. \tag{1.1}$$

一个最初的例子是二项分布的正态逼近.① 这类逼近通常以极限定理的形式来表述，该定理包含一系列越来越精细的离散概率模型. 在许多情形下，通过这种极限在概念上引入一个新的样本空间，而后者在直观上可以比原来的离散模型更简单.

例 (a)　指数型等待时间. 为了用离散模型来表述等待时间，我们把时间离散化，并假定变化仅发生在时刻② $\delta, 2\delta, \cdots$. 最简单的等待时间 T，是在具有成功概率 p_δ 的伯努利 (Bernoulli) 试验序列中第一次成功的等待时间. 于是 $\mathrm{P}\{T > n\delta\} = (1-p_\delta)^n$，等待时间的期望是 $\mathrm{E}(T) = \delta/p_\delta$. 这个模型的精确化，是在保持期望值 $\delta/p_\delta = \alpha$ 不变的条件下让 δ 减小而得到. 在长度为 t 的时间内，相应地有 $n \approx t/\delta$ 次试验，取对数后可以看出，当 δ 很小时近似地有

$$\mathrm{P}\{T > t\} \approx (1-\delta/\alpha)^{t/\delta} \approx \mathrm{e}^{-t/\alpha}. \tag{1.2}$$

该模型把等待时间作为几何分布的离散随机变量来研究，而 (1.2) 说明我们“依极限”得出一个指数分布. 直观上看，似乎从以实数为元素的样本空间出发直接引入指数分布比较自然.

例 (b)　随机选择. 在区间③ $\overline{0,1}$ 中“随机地选择一点”是一个有明显直观意义的理想实验. 它可用离散逼近来表述，但比较容易的是用整个区间作为样本空间，并指定每个子区间的长度为其概率. 作两次独立的随机选择 $\overline{0,1}$ 中点的理想实验，得到一对实数. 因此，自然的样本空间是一个单位正方形. 在这种样本空间中，自然把“概率”和“面积”同等看待. 这样做对于某些基本目的是十分令人满意的，但是往后进一步就会产生“‘面积’究竟意味着什么”这种疑问.

---

① 更多的例子取自：第 1 卷 3.4 节中的反正弦分布; 3.7 节中的返回原点次数和初过时间的分布; 第 14 章中的随机游动的极限定理; 11.7 节习题 20 中的均匀分布.

② 关于术语“时刻”的使用，参看本书的“符号与约定”表.

③ 区间是带有横线的，保留符号 (a, b) 表示平面上点的坐标. 参考本书的“符号与约定”表.

这些例子表明，连续样本空间在概念上可以比离散模型简单，但是其上概率的定义却依赖于积分和测度论这类的工具. 在可数样本空间中，可以对一切可想象的事件赋以概率，但在一般空间中，这一直觉的方式会导致逻辑上的矛盾，而我们的直观必须适应形式逻辑的要求. 我们将很快看到，直觉的方法即使在相当简单的问题中也可能产生困难. 恰当地说，许多有概率意义的问题不需要明确的概率定义. 有时它们是具有分析特征的问题，而概率背景主要是作为对我们的直观的一种支撑. 更为中肯的是，某些具有复杂样本空间的复杂随机过程可导致很有意思而且容易理解的问题. 这些问题不依赖于在整个过程的分析中所用的巧妙工具. 典型的论述如下：如果过程可以被完全地描述出来，则随机变量 Z 应当具有如此这般的性质，因此其分布应当满足如此这般的积分方程. 虽然概率的论证能极大影响所述问题中方程的分析处理，但是在原则上后者与概率公理无关. 各领域的专家们常常熟悉这类问题，对此他们不用测度论，因为他们不了解其他类型的问题和不严格论证所带来错误结果的情况. ①

这种情况在本章 (全书的非正式引论) 中变得更明显了. 它叙述了本书通篇将用到的两种重要分布的某些分析性质. 本章也包括一些特殊论题，部分是由于它们的重要应用，部分是为了说明我们将要遇到的新问题和对适当工具的需求. 但不必要系统地或按其出现顺序去研究它们.

在本章中，概率由初等积分定义，同时承认这个定义的局限性. 概率论术语如随机变量或期望的使用，在两种意义下可认为是正当的. 其一，与第 1 卷同等情况相似，它们可以解释成为直观的技术助手；其二，本章中的内容可以用逻辑上无缺陷的方式由 1.2 节例 (a) 所述的离散模型通过取极限来说明. 虽然原则上既不必要又不值得追求，但后一方式为初学者提供了一个很好的练习机会.

1.2 密度和卷积

直线(或 $\mathbf{R}^1$)*上的概率密度*是满足下列条件的函数 f, 使得

$$f(x) \geqslant 0, \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1. \tag{2.1}$$

目前我们只考虑逐段连续密度 (一般情况可参看 5.3 节). 对于每一密度 f，它所对

① 严格和直观两者的作用容易被误解. 如第 1 卷指出，朴素的直观和朴素的思维是无力的，但是它们随着数学理论的发展而增加了力量. 今天的直观和应用依赖于昨天的最难于理解的理论. 而严格的理论加强了思维的效率. 的确，经验表明，在应用中大多数人依靠冗长的计算而不是简单的论证，因为这些论证看来是危险的. (最恰当的说明在 1.5 节例 (a) 中.)

应的分布函数[1] F 为

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y. \tag{2.2}$$

它是从 0 增加到 1 的单调连续函数. 我们说 f 和 F 集中于区间 $a \leqslant x \leqslant b$ 上, 如果 f 在这区间外为 0. 密度 f 是作为直线上区间的概率的赋值而被研究. 区间 $\overline{a,b} = \{a < x < b\}$ 有概率

$$F(b) - F(a) = \int_{a}^{b} f(x)\mathrm{d}x. \tag{2.3}$$

这个概率有时用 $\mathrm{P}\{\overline{a,b}\}$ 来表示. 在这种赋值下, 单点的概率是 0, 闭区间 $a \leqslant x \leqslant b$ 和 $\overline{a,b}$ 有相同的概率.

在最简单的情形中, 以实直线作为"样本空间", 也就是说, 以数来表示理想实验的结果. (和第 1 卷一样, 在表示实验序列的样本空间的建立中, 这只是第一步.) 随机变量是定义在样本空间上的函数. 为了简单起见, 目前我们仅把这样的函数 U 作为随机变量: 对于每一个 t, 它使事件 $\{U \leqslant t\}$ 由有限多个区间组成. 于是

$$G(t) = \mathrm{P}\{U \leqslant t\}, \tag{2.4}$$

正好被定义为 f 在这些区间上的积分. (2.4) 定义的函数 G 称为 U 的分布函数. 如果 G 是函数 g 的积分, 则 g 称为分布 G 的密度或变量 U 的密度.

显然, 基本随机变量是坐标变量[2] X, 而所有其他随机变量都是 X 的函数. X 的分布函数重合于定义概率的分布 F. 不用说, 任一随机变量 $Y = g(X)$ 可作为一新直线上的坐标变量.

上述这些术语, 可用与第 1 卷情形的简单类比来说明其正确, 但是下例表明, 我们的模型可通过离散模型取极限而得到.

例 (a) 数据分组. 令 F 为一给定的分布函数. 选一固定的 $\delta > 0$, 并考虑离散随机变量 X_δ; 对于 $(n-1)\delta < x \leqslant n\delta, X_\delta$ 取常数值 $n\delta$. 其中 $n = 0, \pm1, \pm2, \cdots$. 在第 1 卷中, 我们用 δ 的倍数作为样本空间, 并以下式来描述 X_δ 的概率分布:

$$\mathrm{P}\{X_\delta = n\delta\} = F(n\delta) - F((n-1)\delta). \tag{2.5}$$

现在 X_δ 变成扩大的样本空间中的随机变量, 它的分布函数当 $n\delta \leqslant x < (n+1)\delta$ 时等于 $F(n\delta)$. 在连续模型中, X_δ 作为 X 的近似, X 是把区间与其右端点等同起来

[1] 所谓"分布函数"是指在 $\pm\infty$ 处具有极限 0 和 1 的右连续非减函数. 第 1 卷主要讨论了其增长是纯跳跃式的分布. 现在我们把注意力集中在分布函数上, 它由积分来定义. 一般的分布函数将在第 5 章研究.

[2] 我们尽可能用大写字母来表示随机变量 (即样本空间上的函数), 用小写字母表示数或位置参数. 特别地, 这对于坐标变量 X, 即函数 $X(x) = x$ 也是适合的.

而得到的 (统计学家知道，该方法是作数据分组用的). 根据第 1 卷的精神，我们把 X_δ 作为基本随机变量，把 δ 作为自由参数来处理. 令 $\delta \to 0$，由极限定理得知，F 是 X_δ 的极限分布.

例 (b)　对于 $x>0$，事件 $\{X^2 \leqslant x\}$ 和事件

$$\{-\sqrt{x} \leqslant X \leqslant \sqrt{x}\}$$

相同；随机变量 X^2 有一个集中在 $\overline{0,\infty}$ 上的分布，它由

$$F(\sqrt{x}) - F(-\sqrt{x})$$

给出. 微分后可见，X^2 的密度g 为

$$g(x) = \begin{cases} \dfrac{1}{2}\left[f(\sqrt{x}) + f(-\sqrt{x})\right]/\sqrt{x}, & \text{当 } x>0 \text{ 时}, \\ 0, & \text{当 } x<0 \text{ 时}. \end{cases}$$

X^3 的分布函数对所有 x 由 $F(\sqrt[3]{x})$ 给出，它的密度为

$$\frac{1}{3} f(\sqrt[3]{x})/\sqrt[3]{x^2}.$$

X 的期望(expection) 定义为

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x, \tag{2.6}$$

但是要求积分绝对收敛. 例 (a) 的近似离散变量 X_δ 的期望与它的积分的黎曼 (Riemann) 和相同，因此 $\mathrm{E}(X_\delta) \to \mathrm{E}(X)$. 如果 u 是一有界连续函数，则同样的论证适用于随机变量 $u(X)$，关系式 $\mathrm{E}(u(X_\delta)) \to \mathrm{E}(u(X))$ 蕴含

$$\mathrm{E}(u(X)) = \int_{-\infty}^{+\infty} u(x) f(x)\mathrm{d}x; \tag{2.7}$$

微妙之处在于这个公式不明显地利用 $u(X)$ 的分布. 于是，只要知道了随机变量 X 的分布就完全可以算出它的函数的期望.

X 的二阶矩(second moment) 定义为

$$\mathrm{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x, \tag{2.8}$$

只要积分收敛. 令 $\mu = \mathrm{E}(X), X$ 的方差又定义为

$$\mathrm{Var}(X) = \mathrm{E}((X-\mu)^2) = \mathrm{E}(X^2) - \mu^2. \tag{2.9}$$

注 如果变量 X 是正的(即密度 f 集中在 $\overline{0,\infty}$), 且 (2.6) 式中积分发散, 为了方便则说 X有无穷期望, 记作 $\mathrm{E}(X)=\infty$. 同样, 当 (2.8) 式中积分发散时, 我们说 X 有无穷方差. 对于取正值与负值的变量, 在积分 (2.6) 发散时, 期望没有定义. 一个典型的例子由密度 $\pi^{-1}(1+x^2)^{-1}$ 给出.

密度的概念可以转用于高维空间中, 但是一般的讨论放在第 3 章. 在那之前, 我们将仅考虑第 1 卷 5.4 节所引入的乘积概率的类似概念, 以便描述独立实验的组合. 换句话说, 在本章中我们仅讨论形如 $f(x)g(y), f(x)g(y)h(z)$ 等的乘积密度, 此处 $f,g,\cdots$ 是直线上的密度. 给定 $\mathbf{R}^2$ 平面上形如 $f(x)g(y)$ 的密度就意味着"概率"与下列积分等同:

$$\mathrm{P}\{A\}=\iint_A f(x)g(y)\mathrm{d}x\mathrm{d}y. \tag{2.10}$$

"两个具有密度 f 和 g 的独立随机变量 X 和 Y"的说法, 意味着在 (X,Y) 平面上由 (2.10) 给出概率. 它蕴含区间的乘法法则, 例如 $\mathrm{P}\{X>a,Y>b\}=\mathrm{P}\{X>a\}\mathrm{P}\{Y>b\}$. 它与离散的情形是如此相似, 以致不需作进一步说明.

许多新的随机变量可定义为 X 和 Y 的函数, 但最起作用的还是和 $S=X+Y$. 事件 $A=\{S\leqslant s\}$ 可由使 $x+y\leqslant s$ 成立的点 (x,y) 组成的半平面来表示. 以 G 表示 Y 的分布函数, 因此有 $g(y)=G'(y)$. 为了得出 $X+Y$ 的分布函数, 在式 (2.10) 中, 在 $y\leqslant s-x$ 上积分得到

$$\mathrm{P}\{X+Y\leqslant s\}=\int_{-\infty}^{+\infty}G(s-x)f(x)\mathrm{d}x. \tag{2.11}$$

由于对称性, F 与 G 所处地位可以交换, 而不影响其结果. 微分可见, $X+Y$ 的密度由下列两积分中任一个给出:

$$\int_{-\infty}^{+\infty}f(s-y)g(y)\mathrm{d}y=\int_{-\infty}^{+\infty}f(y)g(s-y)\mathrm{d}y. \tag{2.12}$$

上式定义的运算是 5.4 节所导出的卷积的特殊情形. 目前我们仅对密度利用卷积这一术语: 两个密度 f 和 g 的卷积是一个由 (2.12) 所定义的函数, 以 $f*g$ 表示.

整个第 1 卷处理的是离散分布的卷积, 且法则是相同的. 依 (2.12) 有 $f*g=g*f$. 给定第 3 个密度 h, 我们可作出 $(f*g)*h$, 这是 3 个分别具有密度 f,g,h 的独立变量之和 $X+Y+Z$ 的密度. 随机变量求和的可交换性与可结合性, 蕴含卷积的同样性质, 因此 $f*g*h$ 与运算顺序无关.

正值随机变量起着重要作用, 因此指出下面一点是有益的, 即如果 f 和 g 集中在 $\overline{0,\infty}$ 上, 则式 (2.12) 的卷积 $f*g$ 就化为

$$f*g(s)=\int_0^s f(s-y)g(y)\mathrm{d}y=\int_0^s f(x)g(s-x)\mathrm{d}x. \tag{2.13}$$