Marek Capiński and Ekkehard Kopp

Measure, Integral and Probability

Second Edition

Springer

SPRINGER UNDERGRADUATE MATHEMATICS SERIES

Marek Capin´ski和Ekkehard Kopp

测量、积分和概率

Springer-Verlag

柏林 海德堡 纽约 伦敦 巴黎 东
京
香港 巴塞罗那 布达佩斯

致我们的孩子；孙子们：Piotr, Maciej, Jan, Anna; Lʉkasz Anna, Emily

前言

本书的核心概念是勒贝斯格测量和勒贝斯格积分。它们在英国本科数学课程中的标准角色并不完全稳固；但它们为抽象度量空间的发展提供了主要模型，而这些空间是现代概率论的基础，而勒贝斯格函数空间仍然是检验函数分析方法及其许多应用的主要例子来源，如傅里叶分析和偏微分方程理论。

由此可见，不仅初出茅庐的分析家需要清楚地了解度量和积分的构造和性质，而且那些希望在数学、物理学、电子学、工程学以及最近的金融学等广泛领域认真促进分析方法应用的人，也需要认真研究基础理论。

在目前的文献中，我们发现很少有明确旨在满足这些需求的文本，而且这些文本的水平可供目前的本科生阅读。有许多关于现代概率论的好书，而且它们越来越认识到需要为我们在本书中开发的工具打下坚实的基础，但所有这些处理对于本科生读者来说往往过于高级，或者有些敷衍了事。因此，我们希望目前的文本不会被认为是填补了文献中亟需的空白的文本！"!

在发展对积分的处理时，一个基本的决定是，是从度量开始还是从积分开始，也就是说，是从集合开始还是从函数开始。函数分析家们倾向于后一种方法，而前者显然是概率论发展的必要条件。在这场争论中，我们决定站在概率论者一边，把概率论中基本概念和结果的（合理的）系统发

展作为主要的应用领域--课题的顺序和

我们使用的术语反映了这一选择，每一章最后都有相关概率概念的进一步发展。有时，这种方法似乎不如其他方法 "有效"，但我们选择了直接证明和明确的结构，有时以优雅为代价。我们希望这将增加理解。

在空间和时间的限制下，对度量和积分的处理是自成一体的：有些章节对于最后一年的本科生来说可能显得过于平淡，但在赫尔大学多年来测试大部分材料的经验告诉我们，对于分析中的基本概念的熟悉和信心，在这些听众中经常显得有些不稳定。因此，序言部分包括对黎曼积分的回顾，以及对初级实分析的一些基本概念的提醒。

虽然这里选择了概率论作为度量衡和积分的主要应用领域，但这并不是一篇关于初级概率的文章，在文献中可以找到很多这样的文章。

虽然这不是一本高级读物，但它的目的是为了学习（而不是轻描淡写），它被设计成对有指导的自学以及讲座课程有用。因此，相当一部分标有 "命题 "的结果并没有立即证明，而是留给读者在进一步研究之前尝试（通常有一个如何开始的提示），还有大量的练习题。为了帮助自学，每章末尾都给出了命题的证明，书末尾还给出了习题的概要解答。因此，对于勤奋的人来说，应该没有什么神秘的东西。在介绍性的一章中，为测量和积分的主要定义做了激励和准备，第二章提供了Lebesgue测量的详细构造和它的属性，并着手抽象出适用于概率空间的公理。这为其余各章设定了一个模式，在这些章节中，独立的概念在更广泛的背景下被追求、作为概率论的一个显著特征。

第三章发展了非负可测函数的积分，并介绍了随机变量及其诱导的概率分布，而第四章发展了Lebesgue积分的主要极限定理，并将其与黎曼积分相比较。概率论中的应用导致了对期望的讨论，重点是密度和特征函数的作用。

第五章的动机是更多的函数分析：重点是Lebesgue函数空间，包括讨论空间L的特殊作用[2] ，即平方不可捉摸的函数。第六章是对度量理论的回归，详细阐述了积度量和福比尼定理，现在又引出了概率中联合分布和条件的作用。最后、

在讨论了可积分函数序列的主要收敛模式之后，第7章采用了一种毫无顾忌的概率论倾向，对主要极限定理进行了处理，最终形成了林德伯格-费勒版的中心极限定理。

处理方法绝不是详尽的，因为这是一本教科书，而不是一篇论文。然而，对于三年级的一个学期的课程来说，主题的范围可能略微过大：前五章可以为这些学生提供一个有用的课程，最后两章留作自学，或者作为希望继续学习概率论的学生的阅读课程的一部分。另外，在分析方面有较强准备的学生可以把前两章作为背景材料，在一个学期的课程中完成本书的其余部分。

1998年5月

Marek Capin´ski
Ekkehard Kopp

第二版序言

经过五年的时间和六次印刷，我们似乎应该对读者的意见作出回应，并纠正我们在审查文本时发现的错误和不完善之处。第二版还引入了早期由于时间和空间的限制而无法实现的额外材料，而在我们看来，随着我们潜在读者群的构成越来越清晰，这些材料变得更加重要。我们希望，我们能够本着保留文本基本特征的精神来做这件事，即严格地提供材料，并以适合指导自学的形式提供材料。因此，我们的重点仍然是可及性、明确性和对具体例子的强调，其风格旨在鼓励读者直接参与到材料中去，并挑战他们自己证明许多结果（知道文本中也给出了解决方案！）。

除了进一步的例子和练习之外，这里的新材料有两种截然不同的类型。新的第7章增加了对一般度量的比较的讨论，以拉登-尼科迪姆定理为重点。这里给出的证明虽然不是新的，但在我们看来比通常的证明更具建设性和基本性，我们一致地利用这个结果来阐述线上的Lebesgue-Stieltjes度量的结构，并推导有符号度量的Hahn-Jordan分解。澄清了函数和度量的变异和绝对连续性概念的共同起源。主要的概率论应用是条件期望，第五章还提供了通过正交投影的替代构造。在第七章中，这又被应用于推导离散时间中马汀格的基本属性。

另一个补充出现在每一章的末尾（第1章和第5章除外）。由于很明显，我们目前有很大一部分的

由于本书的读者群是新兴的数学金融领域的学生，因此每一个相关章节的结尾都对该主题的观点进行了简要讨论。在这些章节中，我们偏离了保持全书自成一体的目标，因为我们很难重新发展这整个学科。因此，我们既不定义也不解释我们所讨论的金融概念的起源，而是试图将它们在数学上定位在计量和概率的概念框架内。这使我们得出的结论具有数学上的精确性，而这种精确性有时让从金融角度写作的作者望尘莫及。

为了避免误解，我们重申，本书的目的仍然是发展度量和积分的思想，特别是考虑到它们在概率和（简要）金融中的应用。因此，这既不是一本概率论的教科书，也不是一本数学金融的教科书。这两门学科都有自己的大量专业文献，我们对这些应用领域的评论旨在帮助学生理解支撑它们的数学框架。

我们感谢那些指出第一版中许多错误的读者和同事，包括印刷上的和概念上的。对于不可避免的错误，我们要负全责。为了便于迅速纠正这些错误，我们建立了一个网页，用于通知错误、不准确和查询，网址是http://www.springer.co.uk/MIP，我们鼓励读者毫不留情地使用它。我们还要感谢伦敦施普林格出版社的Stephanie Harding和Karen Borthwick，感谢他们在制作本版时给予的持续关注和帮助。

2003年10月

Marek Capin´ski
Ekkehard Kopp

内容

1

动机和序言

生活是一项不确定的事业。我们很少能确定我们的计划会按照我们的意图进行，因此，我们从小就有条件用某些事件发生的*可能性来*思考，而且哪些事件比其他事件 "更有可能"。将这种模糊的描述变成一个*概率模型*，相当于构建了一个思考不确定性的合理框架。这个框架应该是一个通用的框架，它使我们能够同样地处理那些我们必须筛选大量先验信息的情况，以及那些我们几乎没有什么可依据的情况。在所有情况下都需要一定程度的判断；但我们寻求一个有序的理论框架和方法，使我们能够以定量的方式制定一般规律。

这使我们想到了概率的数学模型，也就是说，经验实践的理想化抽象，但它必须满足广泛适用性、准确性和简单性的标准。在本书中，我们关注的是普遍适用的概率模型的构建和使用，在这些模型中，我们也可以考虑无限的样本空间和无限的试验序列：当我们试图了解这些模型的意义时，就很容易看出这些模型的必要性。

诸如 "从区间[0，1]中随机抽出一个数字 "这样明显简单的概念，以及试图理解一连串相同试验的极限行为。就像初级概率是通过寻找

我们会发现，要解决的基本问题是如何测量一个具有无限多元素的集合的 "大小"。至少对于实线上的集合来说，基本实分析的思想为我们提供了一

个令人信服的答案，这包含了抽象公理框架所需要的所有思想，而概率理论就建立在这个框架之上。对于

因此，*测量*概念的发展，以及*Lebesgue*测量*对*
特别是R，在这本书中占有很重要的位置。

1.1 符号和基本集合 理论

在度量理论中，我们通常处理一些任意给定集合的子集的族，并考虑将实数分配给属于这些族的集合的函数。因此，我们需要回顾一些基本的集合符号和对集合的操作，以及讨论可数和不可数之间的区别。
无限集，特别是实线R的子集。 我们还将
需要来自分析的概念，如序列的极限、系列和开放集。
假设读者对这些材料基本熟悉，因此可以轻松跳过这一节，这一节是为了介绍符号，使文本合理地自成一体，从而有利于自学。讨论仍然是相当非正式的，没有提到基础问题，读者可以参考关于分析的基本文本来获得大多数证明。这里我们只提到两本最近的介绍性教科书：[8]和[11]。

1.1.1 设置和 功能

在我们对集合的操作中，我们将总是处理某个普遍集合Ω的子集的集合；这个集合的性质将从上下文中清楚地看到--通常Ω将是实数集合R或它的一个子集。我们把
的概念是未定义的和给定的，而我们只关注
集体成员和操作。空集用$\varnothing$表示；它没有成员。集合一般用大写字母表示。

集合成员资格用 ∈，所以$x \in A$表示元素x是集合A的成员。集合包容，$A \subset B$，表示A的每个成员都是B的成员。这包括A和B相等的情况；如果包容是严格的，即$A \subset B$和B包含不在A中的元素（写成$x \notin A$），这将单独说明。符号$\{x \in A : P(x)\}$是用来表示

A的所有子集（其*幂集*）用P（A）表示。

我们定义*交集*$A \cap B = \{x : x \in A$和$x \in B\}$和*并集*
$A \cup B = \{x : x \in A$或$x \in B\}$。A的*补体*A^c由不属于A的Ω元素组成；我们也写$A^c = \Omega \setminus A$，更一般地说，我们有$B \setminus A = \{x \in B : x \notin A\} = B \cap A^c$和对称*差*$A \Delta B =$（$A \setminus B$）$\cup$（$B \setminus A$）。请注意，如果$A \Delta B = \varnothing$

且仅当$A=B$时。

交叉（responsive）表达了逻辑连接词 "和"（responsive），并且，通过逻辑符号（存在）和（对于所有），它们有对任意集合的扩展；以某个集合Λ为索引。这些被赋予了

$$\bigcap_{\alpha\in\Lambda} A_\alpha = \{x : x\in A_\alpha \text{ for all } \alpha\in\Lambda\} = \{x : \forall\alpha\in\Lambda, x\in A_\alpha\}$$

$$\bigcup_{\alpha\in\Lambda} A_\alpha = \{x : x\in A_\alpha \text{ for some } \alpha\in\Lambda\} = \{x : \exists\alpha\in\Lambda, x\in A_\alpha\}。$$

这些都是*由德摩根定律*联系起来的：

$$\left(\bigcup_\alpha A_\alpha\right)^c = \bigcap_\alpha A_\alpha^c;\quad \left(\bigcap_\alpha A_\alpha\right)^c = \bigcup_\alpha A_\alpha^c。$$

如果$A\cap B=\varnothing$，那么A和B是*不相交的*。如果$A_\alpha \cap A_\beta = \varnothing$，只要$\alpha \neq \beta$ ($\alpha, \beta \in \Lambda$)，一个集合族$(A_\alpha)_{\alpha\in\Lambda}$是*成对不相交的*。

集合A和B的*笛卡尔积*$A\times B$是有序对$A\times B = \{(a, b) : a\in A, b\in B\}$*的*集合。如前所述，我们用N、Z、Q、R分别表示自然数、整数、有理数和实数的基本数系。R中的区间通过每个端点，用一个方括号表示表示它的包含，一个开放括号的排除，例如，$[a, b) = \{x\in \mathrm{R} : a \leq x < b\}$。我们使用$\infty$和$-\infty$来描述无界区间，例如$(-\infty, b) = \{x\in \mathrm{R} : x < b\}$，$[0, \infty) = \{x\in \mathrm{R} : x\geq 0\} = \mathrm{R}^+$。$\mathrm{R}^2 = \mathrm{R}\times\mathrm{R}$ 表示平面，更广泛地说，R^n 是 R 与自身的 n *倍*笛卡尔积，即由实数组成的所有 n *个元组*（$x_1, \dots, x_n$）的集合。区间的乘积，称为*矩形，*也是这样表示的。

形式上，一个*函数f*：$A\to B$是$A\times B$的一个子集，其中每个第一坐标决定第二坐标：如果（a，b），（a，c）$\in f$，那么$b=c$。*其域*$\mathrm{D}_f = \{a\in A : \exists b\in B, (a, b)\in f\}$，且*范围* $\mathrm{R}_f = \{b\in B : \exists a\in A, (a, b)\in f\}$描述其范围。非正式地，$f$将$B$的元素与$A$的元素联系起来，这样每个$a\in A$最多只有一个$b\in B$*的图像*，我们写成$b=f$（$a$）。集合$X\subset A$的*图像$f$*$(X) = \{b\in B : b = f(a) \text{ for some } a\in X\}$，集合$Y\subset B$的*反图像*是$f^{-1}(Y) = \{a\in A : f(a)\in Y\}$。$f_2\circ f_1$ 的*组成*是 $f_1 : A\to B$ 和 $f_2 : B\to C$ 是函数 $h : A\to C$ 定义为$h(a) = f_2(f_1(a))$。当$A = B = C$时，$x\mapsto (f_1\circ f_2)(x) = f_1(f_2(x))$ 和 $x\mapsto (f_2\circ f_1)(x) = f_2(f_1(x))$ 都定义从A到A*的*函数。一般来说，这些不会是相同的：例如，让$f_1(x) = \sin x, f_2(x) = x^2$，那么$x\mapsto \sin(x^2)$和$x\mapsto (\sin x)^2$就不相等。

如果$\mathrm{D}_f\subset \mathrm{D}_g$，并且$g = f$在$\mathrm{D}_f$，则函数$g$*扩展了$f$*；或者我们说即$f$*将$g$限制在*$\mathrm{D}_f$。这些概念将经常被用于实值的集合函数，其中域是集合的集合，范围是R的一个子集。

实数函数的代数是点状定义的，即$f+g$*之和与乘积*$f\cdot g$由（$f+g$）（x）$=f$（x）$+g$（x），（$f\cdot g$）（x）$=f$（x）$\cdot g$（x）给出。

集合A的*指示*函数1_A，即为函数

$$1_A(x)=\begin{cases}1 & 对于x\in A\\ 0 & 对于x\notin A。\end{cases}$$

请注意，$1_{A\cap B}=1_A\cdot 1_B$，$1_{A\cup B}=1_A+1_B-1_A 1_B$，以及$1_{A^c}=1-1_A$。

我们还需要一个来自基本集合理论的概念，它应该是熟悉的：对于任何集合E，E上的*等价关系*是一种关系（即$E\times E$的一个子集R，我们用$x\sim y$来表示（x，y）$\in R$），具有以下属性：

1. *反射性*：对于所有$x\in E$，$x\sim x$、
2. *对称性*：$X\sim Y$意味着$Y\sim X$、
3. *传递性*：$X\sim Y$和$Y\sim Z$意味着$X\sim Z$。

E上的等价关系将E划分为互不相干的*等价类*：给定$x\in E$，写$[x]=\{z:z\sim x\}$为x的等价类，即E中与x等价的所有元素的集合，因此$x\in[x]$，因此$E=\bigcup_{x\in E}[x]$。这是一个不相交的联合：如果$[x]\cap[y]\neq\varnothing$，那么有$z\in E$，有$x\sim z$和$z\sim y$，因此$x\sim y$，所以$[x]=[y]$。我们将用$E/\sim$来表示这样得到的所有等价类的集合。

1.1.2 R中可数和不可数的集合

我们说，如果一个集合A和N的一个子集之间有一个一一对应的关系，即一个函数$f:A\to$ N，它把不同的点带到不同的点。非正式地讲，如果这种对应关系可以只用N的一个初始段$\{1，2，\ldots，N\}$来建立（对于某些$N\in N$），那么A就是*有限的，*而我们称A
如果使用N的全部，则是*可数的无限*或*可数的。*不难看出
可数集合的可数联盟是可数的；特别是，集合Q
的有理数是可数的。

康托尔表明，集合R*不能与*N的（子集）一一对应，因此它是一个*不可数集的*例子。康托尔的证明
假设我们可以把每个实数唯一地写成一个小数（总是选择非终止的版本）。我们也可以把自己限制在显示区间[0,1]是不可数的范围内(为什么？)。

如果这个集合是可数的，那么我们可以把它的元素写成一个序列(x

$)_{n\geq 1}$，由于每个x_n都有一个唯一的十进制展开形式

$$x_n = 0.a_{n1}a_{n2}a_{n3}\ldots a_{nn}\ldots。$$

对于从0, 1, 2..., 9的集合中选择的数字a_{ij}，因此我们可以写下数组

$$x_1 = 0.a\ a\ a_{111213}\ . . .$$

$$x_2 = 0.a\ a\ a_{212223}\ . .$$

$$x_3 = 0.a\ a\ a_{313233}\ . .$$

.. .

现在写下$y = 0.b\ b\ b_{123} ...$，其中数字b_n被选择为与a_{nn}不同。这样的十进制扩展定义了一个与x_n中的每一个不同的数字$y\in[0，1]$（因为它的扩展与x_n的*第*n位不同。）

因此，我们的序列并没有穷尽[0,1]，而矛盾的是，[0,1]不可能是可数的。

由于两个可数集的结合必须是可数的，而Q是可数的，因此R Q是不可数的，也就是说，无理数比有理数 "多 "得多!让这个问题看起来更容易消化的一个方法是考虑从R的一个区间中随机选择数字的问题。

回顾一下，有理数正是那些实数，它们的解------。

IMAL扩展重复出现（我们在 "重复出现 "下包括 "终止"）。现在想象一下，从[0,1]中随机选择一个实数：把集合R想象成一个包含所有实数的池塘，并想象你在这个池塘中'钓鱼'，拉出

每次都是一个数字。

第一个数字有多大可能是有理的，也就是说，我们有多大可能找到一个其扩张重复出现的数字？这就好比无限次地掷出一个十面体的骰子，并期望在有限次的掷出后，可以肯定地说，*所有*后续的掷出都会得到相同的数字。这确实

因此，当我们以 "无偏 "或统一的方式测量实线上的集合时，我们不应惊讶地发现，可数集合（包括Q）将是我们可以 "忽略 "的集合之一。

到目前为止，我们已经使用了 "随机 "这个术语。然而，可能更令人惊讶的是，发现即使是一些不可数的集合，从这里采用的观点来看也是'可忽略不计'的。

1.1.3 R中集合的拓扑学特性

回顾*开放集O*⊂R的定义：

定义1.1

实线R的一个子集O是*开放的*，如果它是开放区间的联盟，即对于区间$(I)_{\alpha\alpha\in\Lambda}$，其中$\Lambda$是一些索引集（可数或不可数）。

$$O = \bigcup_{\alpha\in\Lambda} I_\alpha \text{。}$$

如果一个集合的补数是开放的，那么它就是*封闭的*。R中的开放集^n (n>1)可以定义为区间的n*次*方乘积的联合体。

这个定义似乎比实际情况更普遍，因为在R上，可数的联合总是足够的--尽管在以后的工作中，使用一般联合的自由将是很方便的。如果Λ是一个索引集，I_α 是一个开放区间，对于

每个$\alpha\in\Lambda$，那么存在一个可数的这些区间的集合$(I_{\alpha k})_{k\geq1}$，其联合体等于$_{\alpha\in\Lambda} I_\alpha$。更重要的是，区间的序列可以被选择为成对的不相交。

很容易看出，开放集的有限相交是开放的；然而，开放集的可数相交不一定是开放的：让O_n = $(-\frac{1}{n}, 1)$ 为

$n\geq1$，那么$E = \bigcap_{n=1}^{\infty} O_n$ = [0, 1]就不是开放的。

请注意，R与R^n 或更一般的空间不同，它有一个*线性顺序*，即给定x，y R，我们可以决定是x y还是y x。因此，如果对所有a A*来说*u是一个集合A R的*上界*，而*下界*的定义与此类似。那么，*上界*（或最小上界）就是所有上界的最小值，而

下限（或最大下限）inf A被定义为所有下限的最大值。R的*完备性属性*可以用这样一句话来表达：每一个有上界的集合都有一个上界。

如果$f^{-1}(O)$对每个开放集O来说都是开放的，则称一个实数函数f是*连续的*。每个定义在封闭有界集上的连续实数函数*都能达到*

*它*在这样一个集合上的*界限*，也就是说，在那里有一个最小值和最大值。例如，如果$f : [a, b] \to$ R是连续的，$M = sup\{f(x) : x\in [a, b]\} = f(x_{\max})$, $m = \inf\{f(x) : x\in [a, b]\} = f(x_{\min})$ 对于某些点$x_{\max}$, $x_{\min}$ $\in[a, b]$。中间值定理说，一个连续的函数把所有的

在极端值之间的中间值，即对于每个$y\in[m, M]$，有一个$\vartheta\in[a, b]$，使$y=f(\vartheta)$。

专门针对实数序列（x_n），我们可以进一步定义*上限*

lim $_{supn}$ x_n 作为

$$\inf\{\sup_{m\geq n} x_m : n \in \mathbb{N}\}$$

和 *下限*lim $\inf_n x_n$ 为

$$\sup\{\inf_{m\geq n} x_m : n\in\mathrm{N}\}。$$

序列x_n收敛，当且仅当这些量重合，并且它们的共同值是它的极限。回顾一下，一个序列（x_n）*收敛*，实数x是它的*极限*，写成$x = \lim_{x\to\infty} x_n$，如果对于每一个$\varepsilon > 0$，存在一个$N$，只要$n \geq N$，$|x_n - x| < \varepsilon$。一个数列 $\sum_{n\geq 1} a_n$ 收敛

如果序列$x_m = \sum_{n=1}^{m} a_n$，其部分和的收敛，其极限是则和$\sum_{n=1}^{\infty} a_n$的系列。

1.2 黎曼积分：范围和 限制

在这一节中，我们对构成入门分析课程主食的黎曼积分做一个简单的回顾，并考虑它不足以满足更高级应用的一些原因。

让$f : [a, b] \to \mathbb{R}$是一个有界实数函数，其中$a, b$，$a < b$，是实数。$[a, b]$的*分区*是一个有限集合$P = \{a_0, a_1, a_2, \ldots, a_n\}$，其中

$$a = a_0 < a_1 < a_2 < \ldots < a_n = b.$$

分区P产生了*上、下黎曼和*

$$U(P, f) = \sum_{i=1}^{n} M_i \Delta a_i、\qquad L(P, f) = \sum_{i=1}^{n} m_i \Delta a_i$$

其中，$\Delta a_i = a_i - a_{i-1}$

$$M_i = \sup_{a_{i-1}\leq x\leq a_i} f(x)$$

，并且

$$m_i = \inf_{a_{i-1}\leq x\leq a_i} f(x)$$

(注意，M_i和m_i是定义良好的实数，因为f在每个区间$[a_{i-1}, a_i]$上是有界的。)

为了定义f的黎曼积分，首先表明对于任何给定的分区P，$L(P, f) \leq U(P, f)$，接下来对于任何*细化*，即一个分区$P' \supseteq P$，我们必须有$L(P, f) \leq L(P', f)$和$U(P', f) \leq U(P, f)$。最后，由于对于任何两个分区P_1和P_2，它们的联合体$P_1 \cup P_2$是两者的细化，我们看到，对于任何一个分区，$L(P, f) \leq U(Q, f)$对于*任何*分区P，Q。$\{L(P, f) : P$是$[a, b]$的一个分区$\}$因此在$\mathbb{R}$中是有界的，我们把它的最高值称为f在$[a, b]$上的*下积分*$\underline{\int_a^b} f$。同样地，在$[a, b]$上的下限的集合是*上层积分*$\overline{\int_a^b} f$。现在说函数f是

在$[a, b]$上，如果这两个数字重合，并且它们的共同值是f的*黎曼积分*，用$\int_a^b f$表示，或者，更常见、

$$\int_a^b f(x)\,dx。$$

这个定义没有为检查特定函数的可整性提供一个方便的标准；然而，下面的表述提供了一个有用的可整性标准--证明见[8]。

定理1.1（黎曼准则）

$f:[a, b] \to \mathbb{R}$是黎曼不稳定的，当且仅当对于每一个$\varepsilon > 0$，存在一个分区P_ε，使得$U(P_\varepsilon, f) - L(P_\varepsilon, f) < \varepsilon$。

例1.1

我们计算$\int_0^1 f(x)\,dx$，当$f(x) = \sqrt{x}$时：我们的直接问题是，除了完全平方外，很难找到平方根。因此我们取

属于完全正方形的分区点，尽管这意味着不同区间的区间长度不会保持不变（没有什么规定说它们应该这样做，即使它常常简化了计算）。事实上，以分区的序列为例

$$P_n = \{0、\left(\frac{1}{n}\right)^2、\left(\frac{2}{n}\right)^2, \ldots, \left(\frac{i}{n}\right)^2, \ldots, 1\}$$

并考虑上下和，利用f是增加的事实：

$$U(P_n, f) = \sum_{i=1}^{n} \left(\frac{i}{n}\right)\left\{\left(\frac{i}{n}\right)^2 - \left(\frac{i-1}{n}\right)^2\right\} = \frac{1}{n^3}\sum_{i=1}^{n}(2i^2 - i)$$

$$L(P_n, f) = \sum_{i=1}^{n} \left(\frac{i-1}{n}\right)\left\{\left(\frac{i}{n}\right)^2 - \left(\frac{i-1}{n}\right)^2\right\} = \frac{1}{n^3}\sum_{i=1}^{n}(2i^2 - 3i + 1).$$

因此

$$U(P_n, f) - L(P_n, f) = \frac{1}{n^3}\sum_{i=1}^{n}(2i - 1) = \frac{1}{n^3}\{n(n+1) - n\} = \frac{1}{n}.$$

通过选择足够大的n，我们可以使这个差异小于任何给定的$\varepsilon>0$，因此f是可整数的。该积分必须是$\frac{2}{3}$，因为$U(P_n, f)$和$L(P_n, f)$收敛到这个值，这很容易看出。

黎曼的标准仍然没有给我们提供黎曼不可捉摸的函数*类的*精确图景。然而，我们很容易证明（见[8]），任何有界*单调*函数都属于这一类，而且只需要多一点的

很难看到，任何连续函数f：[a，b]R（当然，它是 →
自动受限）将是黎曼不稳定的。

这为许多实际的目的提供了相当充分的信息，并且可以通过证明来避免诸如上述计算的繁琐性

定理1.2（微积分的基本定理）

如果$f : [a, b] \to \mathbb{R}$是连续的，并且函数$F : [a, b] \to \mathbb{R}$有导数$f$
(即$F' = f$ on (a, b)) 则

$$F（b）-F（a）= \int_a^b f(x)\,dx。$$

因此，这一结果将黎曼积分与微分联系起来，并将F显示为f的一个*基元*（也称为 "反衍生"）：

$$F(x) = \int_{-a}^{x} f(x)$$

到一个常数，从而证明了构成任何微积分课程一部分的基本积分技术的合理性。

我们可以放松对连续性的要求。一个微不足道的步骤是假设f在[a, b]上有界且连续，除了有限的几个点。那么，f是黎曼可积分的。为了看到这一点，将区间分割成f是连续的部分。那么f在每一块上都是可整除的，因此我们可以推导出f在整个区间上的可整除性。作为一个例子，考虑一个函数f对所有$x \in [0, 1]$都等于零，除了$a_1, \ldots, a_n$，在那里它等于1。它是可积分的，在[0, 1]上的积分等于0。

然而，进一步研究这个问题需要Lebesgue理论的力量：在定理4.23中，我们表明当且仅当f在[a, b]的 "几乎所有 "点上是连续的，它就是黎曼不稳定的。这个结果绝不是微不足道的，如果你试图直接证明以下函数f，

由于*Dirichlet*，在[0，1]上是黎曼不稳定的，你会发现：

$$f(x)=\begin{cases}\frac{1}{n} & \text{如果}\ x=\frac{m}{n}\in \mathrm{Q}\\ 0 & \text{如果}\ x\notin \mathrm{Q}\text{。}\end{cases}$$

事实上，不难看出，见[8]，f在每个无理点是连续的，在每个有理点是不连续的，因此（正如我们将看到的）在[0，1]的 "几乎所有 "点是连续的。

由于本书的目的是介绍*Lebesgue*的积分理论，我们应该讨论*为什么*我们需要一个新的积分理论：如果有的话，上面描述的简单黎曼积分有什么问题?

首先，范围：它没有处理我们希望处理的所有种类的函数。

最容易证明的结果依赖于有界区间上的连续函数；为了处理无界区间上的积分，例如

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx$$

或一个无界函数的积分：

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx、$$

我们不得不求助于 "不适当的 "黎曼积分，由一个极限过程来定义：例如，考虑到积分

$$\int_{-n}^{n} e^{-x^2}\,dx、\qquad \int_{\varepsilon}^{1} \frac{1}{\sqrt{x}}\,dx、$$

并分别让$n\to\infty$或$\varepsilon\to 0$。这并不是什么严重的缺陷。

第二，对区间的依赖：我们没有简单的方法对更多的一般集合进行积分，或对其值分布为 "awk-"的函数进行积分。

在与区间差别很大的集合上'wardly'。例如，考虑Q在[0, 1]上的指标函数1_Q 的上和和；无论我们如何划分[0, 1]，每个子区间必须包含有理和无理的

因此，我们无法计算f在区间[0,1]上的黎曼积分；它只是 "太不连续了"。(你可以很容易地说服自己，f在[0,1]的所有点上是不连续的）。

第三，缺乏完整性：从应用的角度来看，更重要的是，黎曼积分与取函数序列的极限没有很好的互动。我们可以期待以下形式的结果：如果一个黎曼积分函数序列f_n 收敛（在某些适当的

感）到f，那么$\int^b f_n\,dx \to \int^b f\,dx$。

We give two counterexamples showing what difficulties can arise if the func-

的关系(f_n)点状收敛于f，即$f_n(x) \to f(x)$ 对于所有x。

1. 极限不需要是黎曼可积分的，因此收敛性问题甚至没有意义。这里我们可以把$f = 1_{Q}$，$f_n = 1_{A_n}$，其中

$A_n = \{q_1, ..., q_n\}$，而序列(q_n)，$n\geq1$是一个有理数的枚举，所以(f_n)是偶数单调增加的。

2. 极限是黎曼整数，但黎曼整数的收敛性并不成立。设f=0，考虑[a，b]=[0，1]，并把

$$f_n(x) = \begin{cases} 4n x^2 & 如果0\le x<\frac{1}{2n} \\ 4n - 4n x^2 & 如果\frac{1}{2n}\le x<\frac{1}{n} \\ 0 & 如果\frac{1}{n}\le x\le 1。\end{cases}$$

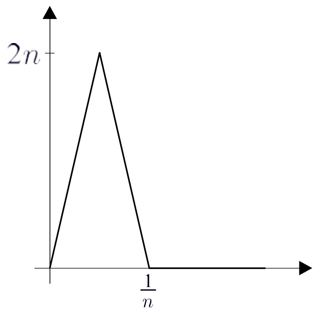

图1.1 f_n的图形。

这是一个积分为1的连续函数。另一方面，序列$f_n(x)$收敛于$f = 0$，因为对于所有的x，$f_n(x)= 0$，对于n足够大（如$\frac{1}{n} < x$）。见图1.1。

为了避免这类问题，我们可以引入*均匀*收敛的概念：如果序列$a_n = \sup\{|f_n(x) - f(x)|: 0\le x\le 1\}$，则$C[0, 1]$中的序列（$f_n$）均匀地收敛于$f$：在这种情况下，我们可以很容易地证明黎曼积分的收敛性：

$$\int_0^1 f_n(x)\,dx \to \int_0^1 f(x)\,dx。$$

然而，"距离 "$\sup\{|f(x) - g(x)|: 0\le x\le 1\}$与积分本身没有关系，而且均匀收敛对许多人来说限制性太大。

应用。一个更自然的"距离 "概念，由$\int_0^1 |f(x) - g(x)|\,dx$给出、导致了另一个问题。定义

$$g_n(x) = \begin{cases} 0 & 如果0\le x\le\frac{1}{2} \\ n(x-\frac{1}{2}) & 如果\frac{1}{2}<x\le\frac{1}{2}+\frac{1}{n} \\ 1 & 否则 \end{cases}$$

$\int$

可以证明：$^{1}|g_n(x) - g_m(x)|\, dx \to 0$，因为$m, n \to \infty$；在图1.2中，阴影部分消失了。(我们说$(f_n)$在这个距离中是一个*考奇序列*)。

然而，没有一个连续的函数f是这个序列收敛的，因为在$x>\frac{1}{2}$， 点状极限是f（x）=1，否则是0，所以$f=1_{(\frac{1}{2},1]}$。
所以所有连续函数f : [0, 1]的空间C([0, 1]) → R太小，从这个观点。

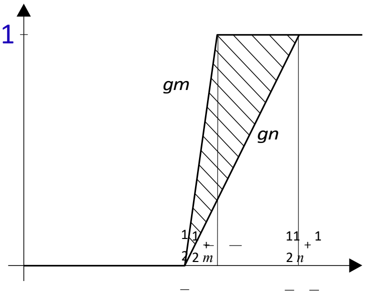

图1.2 g的图形$_n$，g_m

这相当类似于导致人们用R工作而不是用其他方式工作的情况。

而不仅仅是有理数集Q（在Q中存在没有极限的Cauchy序列，例如$\sqrt{2}$的有理数近似序列）。回顾一下
在R的情况下，完整性的关键重要性，我们自然要寻找一个
融合的理论，它没有这个缺点。在这个过程中，我们
我们将发现我们的新理论，包括作为特例的黎曼积分，也解决了所列的其他问题。

1.3 以 随机选择数字

在我们开始发展*Lebesgue度量*的理论来理解R的一般子集的 "长度 "之前，让我们停下来考虑一些实际的动机。有限样本空间的基本概率的简单性
当我们有*无限多*的结果时，例如当我们 "在0和1之间随机挑选一个数字 "时，"概率 "迅速消失。我们面对的是使一个给定的x[0，1]被选中的'概率'的意义。一个类似的、略微宽泛的问题是：我们挑选的数字是有理性的概率是多少？

首先是一个先验问题：我们说随机抽取数字x是什么意思？随机 "似乎意味着在每一次试验中，每个实数被选中的可能性是 "相同的"，因此我们在[0，1]上施加了均匀的概率分布。但是可能的选择的 "数量 "是无限的

。因此，事件A_x，一个固定的x被选中的概率应该是零。从另一个角度看

因此，一个集合$A_x = \varnothing$可以有$P(A_x) = 0$。因此，我们 "测量 "概率的方法不需要能够完全区分集合--如果我们想处理无限的集合，我们不可能真正期望这样。

我们可以再进一步：一个*有限的*集合的任何一个的概率
实数$A = \{x_1, x_2, . . ., x_n\}$被选中也应该是0，因为这似乎很自然
这个概率P（A）应该等于$\sum_{i=1}^{n} P(\{x_i\})$。我们可以将其扩展为
主张概率函数$A' \to P$（A）的有限可加性属性，即
如果$A_1, A_2, ..., A_n$是*不相交的*集合，那么$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$。这
这种说法看起来很有道理，我们将看到它成为任何概率计算的合理基础的一个基本特征。

不太明显的是，在均匀分布下，任何*可数的无限*集，如Q，也必须带有概率0--然而这正是对函数1的 "图下面积 "的分析$_Q$。我们可以将其重新解释为映射$A' \to P$（A）的'连续性属性'的结果，当我们让n 中：如果R的子集的序列（A_i）$\to \infty$是不相交的，那么我们就希望有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i)。$$

我们将在第二章看到，这个条件确实被实线R上的Lebesgue度量所满足，它将被用作任意集合上抽象度量的定义属性。

概率论的内容比本书所阐述的要多得多：例如，我们没有讨论有限样本空间和优雅的组合学思想，这些思想是概率论良好介绍的特点，如[6]和[9]。我们自始至终关注的焦点仍然是勒贝斯格模型在描述基于无限样本空间的概率现象方面所发挥的重要作用。这使我们把这些文章中的许多有趣的例子和应用搁置一边，而是提供一个有密度的随机变量的理论基础的一致发展。

2
衡量标准

2.1 Null 集

正如我们在上一章看到的，"可忽略 "集的概念与黎曼积分的一个限制有关。由于函数$f = 1_{\mathbb{Q}}$只在Q上取非零值，并且在那里等于1，"其图形下的面积"（如果这样做有意义的话）必须与集合Q的 "长度 "密切相关。这就是为什么我们不能在黎曼意义上对f进行积分：集合Q和R Q与区间有很大不同，我们不清楚应该如何测量它们的 "长度"，很明显，f在[0, 1]上的 "积分 "应

等于[0，1]中的有理数集的 "长度"。那么，对于更一般的集合，我们*应该*如何定义这个概念呢?

定义一个集合的 "长度 "的明显方法还是从区间开始。假设I是一个任何种类的有界区间，即I=[a, b]，I=[a, b]，I=（a, b）或I=（a, b）。我们简单地将I的长度定义为$l(I)=b\ a$，在每种情况下都是如此。

作为一个特殊的例子，我们有$l(\{a\}) = l([a, a]) = 0$.然后很自然地说一个单元素集合是'空'的。在我们将这一想法扩展到更一般的集合之前，首先考虑一个有限集合的长度。一个有限集不是一个区间，但是由于一个单点的长度为0，将有限多个这样的长度相加仍应得到0。这里的基本概念是，如果我们将一个集分解成有限个不相交的区间，我们通过将各部分的长度相加来计算这个集的长度。

正如我们所看到的，在一般情况下，它可能并不总是能够实际分解----。

将一个集合摆成区间。因此，我们考虑覆盖一个给定集合的区间系统。我们将通过允许有可数的覆盖区间来概括上述想法。因此，我们得出以下关于 "零长度 "集合的更一般的定义：

定义2.1

空集A R是一个可能被总长度任意小的区间序列覆盖的集合，即给定任何 $\varepsilon>0$，我们可以找到一个序列
$\{I_n : n \geq 1\}$的区间，以便

$$A \subseteq \bigcup_{n=1}^{\infty} I_n$$

和

$$\sum_{n=1}^{\infty} l(I_n) < \varepsilon。$$

(我们也可以简单地说，'*A是空的*）。

练习2.1

证明如果我们在上述定义中用以下任何一个词代替'区间'，我们将得到一个等同的概念：'开放区间', '封闭区间', '形式为(a, b)的区间', '形式为$[a, b]$的区间'。

请注意，这些区间不需要是不相交的。从定义中可以立即看出，空集是空的。

接下来，任何一个单元素集合{x}都是一个空集。因为，让$\varepsilon>0$，取$I_1 = (x - \frac{\varepsilon}{4}, x + \frac{\varepsilon}{4})$, $I_n = [0, 0]$ for $n \geq 2$. （为什么取$I_n = [0, 0]$ for $n \geq 2$？那么，为什么
不是!我们同样可以采取$I_n = (0, 0) = \varnothing$，当然！)现在

$$\sum_{n=1}^{\infty} l(I_n) = l(I_1) = \frac{\varepsilon}{2} < \varepsilon。$$

更一般地说，任何可数集$A = \{x_1, x_2, \ldots\}$是空的。表明这一点的最简单方法是，对所有n而言，$I_n = [x_n, x_n]$。然而，作为对下一个定理的温和介绍，我们将用开放区间覆盖A。这样就更有趣了。

因为，让$\varepsilon>0$，用以下的区间序列覆盖A

$$I_1=\left(x_1-\frac{\varepsilon}{8},\ x_1+\frac{\varepsilon}{8}\right)\qquad l(I_1)=\frac{1}{2}\varepsilon\cdot\frac{1}{2^1}$$
$$I_2=\left(x_2-\frac{\varepsilon}{16},\ x_2+\frac{\varepsilon}{16}\right)\qquad l(I_2)=\frac{1}{2}\varepsilon\cdot\frac{1}{2^2}$$
$$I_3=\left(x_3-\frac{\varepsilon}{32},\ x_3+\frac{\varepsilon}{32}\right)\qquad l(I_3)=\frac{1}{2}\varepsilon\cdot\frac{1}{2^3}$$
$$\cdots\cdots$$
$$I_n=\left(x_n-\frac{\varepsilon}{2^{2+n}},\ x_n+\frac{\varepsilon}{2^{2+n}}\right)\qquad l(I_n)=\frac{1}{2}\varepsilon\cdot\frac{1}{2^n}$$

由于$\sum_{n=1}^{\infty}\frac{1}{2^n}=1$,

$$\sum_{n=1}^{\infty} l(I_n)=\frac{\varepsilon}{2}<\varepsilon$$

根据需要

。

这里我们有以下情况：A是无数个单元素集合的联盟。它们中的每一个都是空的，A*也*变成了空的。

我们可以将这一简单的观察加以概括：

定理2.1

如果$(N_n)_{n\geq 1}$是一连串的空集，那么它们的联合体

$$N=\bigcup_{n=1}^{\infty} N_n$$

也是空的。

证明

我们假设所有的N_n，$n\geq 1$都是空的，为了证明N也是如此，我们采取任何$\varepsilon>0$。我们的目标是用总长度小于ε*的*可数个区间来覆盖N这个集合。

证明分三步进行，每一步都有一点难度。

步骤1。我们仔细地用间隔覆盖每一个N_n。

仔细'意味着长度要小。'小'意味着我们以后要把它们加起来，最后得到一个小数（这里的'小'意味着小于ε）。

由于N_1是空的，所以存在区间I_k^1，$k\geq 1$，使得

$$\sum_{k=1}^{\infty} l(I_k^1) < \frac{\varepsilon}{2}, \quad N_1 \subseteq \bigcup_{k=1}^{\infty} I_k^1 .$$

对于N_2，我们找到一个区间系统I^2_k，$k\ge1$，其中有

$$\sum_{k=1}^{\infty} l(I^2_k) < \frac{\varepsilon}{4}, \quad N_2 \subseteq \bigcup_{k=1}^{\infty} I^2_k .$$

你可以看到一个狡猾的计划，即在每一步以几何速度使总长度变小。
一般来说，我们用区间I^n_k，$k\ge1$来覆盖N_n，其总的长度小于$\frac{\varepsilon}{2^n}$：

$$\sum_{k=1}^{\infty} l(I^n_k) < \frac{\varepsilon}{2^n}, \quad N_n \subseteq \bigcup_{k=1}^{\infty} I^n_k .$$

第2步。间隔I^n_k形成一个序列。

我们将可数的区间家族$\{I^n_k\}_{k\ge1,n\ge1}$排成一个序列J_j，$j\ge1$。例如，我们把$J_1 = I^1_1$，$J_2 = I^1_2$，$J_3 = I^2_1$，$J_4 = I^1_3$，等等。没有一个I^n_k被跳过。新的区间系统的联合与旧的区间系统的联合是一样的，所以

$$N = \bigcup_{n=1}^{\infty} N_n \subseteq \bigcup_{n=1}^{\infty}\bigcup_{k=1}^{\infty} I^n_k = \bigcup_{j=1}^{\infty} J_j 。$$

第3步。计算J_j的总长度。

这很棘手，因为我们有一系列有两个指数的数字：

$$\sum_{j=1}^{\infty} l(J_j) = \sum_{n=1,k=1}^{\infty} l(I^n_k)。$$

现在我们希望把它写成一个数列，每个数列都是一个数列的和。我们可以重新排列双和，因为各部分都是非负的（这是初级微积分的事实）。

$$\sum_{n=1,k=1}^{\infty} l(I^n_k) = \sum_{n=1}^{\infty}\left(\sum_{k=1}^{\infty} l(I^n_k)\right) < \sum_{n=1}^{\infty}\frac{\varepsilon}{2^n} = \varepsilon、$$

这就完成了证明。 □

因此，任何可数集都是空集，而空集似乎与可数集密切相关--这并不奇怪，因为任何适当的区间都是不可数的，所以与区间相比，任何可数子集都是相当 "稀疏 "的，因此对其 "长度 "没有实际贡献。(你可能也注意到了

上述证明中的步骤2与 "对角线论证 "有相似之处，后者通常用于证明Q是一个可数集。）

然而，不可数的集合*可以*是空的，只要它们的点有足够的 "稀疏分布" ，正如下面这个著名的例子，由于康托尔，显示：

1. 从区间[0，1]开始，去除 "中间三分之一"，即区间 ($\frac{1}{3}$,$\frac{2}{3}$)，得到集合C_1，它由两个区间[0,$\frac{1}{3}$]和[$\frac{2}{3}$, 1].
2. 接下来去掉这两个音程的中间三分之一，留下C_2，由四个音程组成，每个音程的长度为$\frac{1}{9}$，等等（见图2.1）。
3. 在第n个阶段，我们有一个集合C_n，由2^n个不相交的封闭区间组成、每个长度为$\frac{1}{3^n}$。因此，C_n的总长度为$\left(\frac{2}{3}\right)^n$。

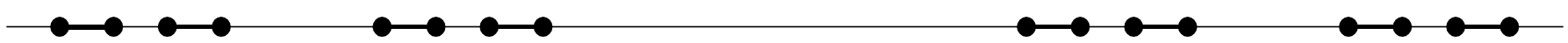

图2.1 康托尔集结构 $(C)_3$

我们称之为

$$C = \bigcap_{n=1}^{\infty} C_n$$

康托尔集。

现在我们表明，C是空的，如约而至。

给定任何$\varepsilon > 0$，选择n，使$\left(\frac{2}{3}\right)^n < \varepsilon$，由于$C \subseteq C_n$，且$C_n$由一个总长度小于$\varepsilon$的（有限的）区间序列组成，我们看到$C$是一个空集。

剩下的就是检查C是一个不可数的集合。这一点留给了你，因为

练习2.2

证明C是不可数的。

提示 适应R的不可数性的证明：首先表达每个x在[0，1]中的三元形式：

∞ x

$3k$ 1

$$= \sum_{k=1} \frac{a_k}{} = 0.a\ a_2 \cdots$$

与$a_k = 0$，1或2。请注意，如果x C的所有a_k都等于0或2。

*为什么*康托尔集是空的，尽管它是不可数的？显然，是它的点的分布，即它"散布"在[0,1]上的事实，造成了麻烦。这使得它成为许多例子的来源，这些例子表明直觉上"显而易见"的事情并不总是真实的!例如，我们可以用康托尔集来定义一个函数，由于勒贝斯格，具有非常奇怪的属性：

如果$x\in[0，1]$有三元扩展(a_n)，即$x = 0.a\,a_{12}\ldots$，$a_n = 0$，1或2，定义N为$a_n = 1$的第一个索引n，如果没有，设$N = \infty$。的a_n是1（即当$x\in C$）。现在设$b_n = \frac{a_n}{2}$为$n < N$，$b_N = 1$、并让$F(x) = \sum_{n=1}^{N} \frac{b_n}{2^n}$，对于每个$x\in[0，1]$。显然，这个函数是单调的增加，并且有$F(0) = 0$，$F(1) = 1$。然而，它在中间的三分之一处是恒定的(即$C$的补数)，所以它的所有增长都发生在康托尔集上。由于我们已经证明了C是一个空集，F从0到1的"增长"完全发生在一个"可忽略不计"的集合上。下面的练习表明，它没有跳跃!

练习2.3

证明Lebesgue函数F是连续的，并画出其图形。

2.2 外部 措施

空集的简单概念为我们的长度概念提供了关键，因为它告诉我们可以"忽略"什么。现在，一个相当普遍的"长度"概念被提供给了：

定义2.2

任何集合$A\subseteq\mathbb{R}$的（Lebesgue）*外部度量*由以下公式给出

$$m^*(A) = \inf Z_A$$

其中

$$Z_A = \left\{\sum_{n=1}^{\infty} l(I_n)：I_n \text{ 是区间，} A\subseteq \bigcup_{n=1}^{\infty} I_n\right\}。$$

我们说$(I_n)_{n\geq 1}$ *覆盖了*集合A。因此，外部测量是A的所有可能覆盖的长

度的下限。（再次注意，一些I_n 可能是

空；这就避免了担心序列（I_n ）是否有有限或无限多的不同成员）。

显然，对于任何$A\subseteq R$，$m^*(A)\geq 0$。对于某些集合A，系列$\sum_{n=1}^{\infty} l(I_n)$对于$A$的任何覆盖都可能发散，所以$m^*(A)$可能等于$\infty$。由于我们希望能够增加各种集合的外部度量，我们必须采用一个

惯例来处理无穷大的问题。一个明显的选择是$a+\infty=\infty$, $\infty+\infty=\infty$，而一个不太明显但相当实用的假设是$0\times\infty=0$，正如我们已经看到的。

集合Z_A自下而上以0为界，所以下限总是存在。如果$r\in Z_A$，那么$[r,+\infty]\subseteq Z_A$（显然，我们可以扩大任何覆盖的第一个区间，以增加总长度的任何数字）。这表明Z_A要么是$\{+\infty\}$，要么是某个实数x的区间（x, $+\infty$）或$[x,+\infty]$。所以Z_A的下限只是x。

首先我们表明，空集的概念与外测的概念是一致的：

定理2.2

$A\subseteq R$是一个空集，当且仅当$m^*(A)=0$。

证明

假设A是一个空集。我们希望证明$\inf Z_A=0$。为此，我们表明，对于任何$\varepsilon>0$，我们可以找到一个元素$z\in Z_A$，使得$z<\varepsilon$。

根据空集的定义，我们可以找到一个涵盖以下内容的区间序列(I_n) A的$\sum_{n=1}^{\infty} l(I_n)<\varepsilon$，所以$\sum_{n=1}^{\infty} l(I_n)$是$Z_A$的必要元素$z$。

反过来说，如果$A\subseteq R$有$m^*(A)=0$，那么根据inf的定义，给定任何$\varepsilon>0$，有$z\in Z_A$，$z<\varepsilon$。但是Z_A的一个成员是A的某个覆盖的总长度。也就是说，有一个A的覆盖（I_n），其总长度小于

比ε，所以A是空的。□

这就把我们的一般外部度量与 "零度量 "的特殊情况结合起来。请注意，对于任何$x\in R$，$m^*(\varnothing)=0$，$m^*(\{x\})=0$，以及$m^*(Q)=0$（事实上，对于任何可数X，$m^*(X)=0$）。

接下来我们观察一下，m^*是单调的：集合越大，它的外量就越大。

命题2.3

如果$A \subset B$，那么m^* （A） $\leq m^*$ （B） 。

提示 证明$Z_B \subseteq Z_A$ 并使用inf的定义。

第二步是将外量子与区间的长度联系起来。这个无辜的结果包含了理论的关键，因为它表明m的正式定义*，它适用于R的*所有*子集，与间的直观想法，我们的思维过程开始于此。我们必须
因此，我们期望证明包含一些隐藏的深度，我们必须分阶段解决这些问题：艰巨的工作在于证明区间的长度不能大于它的外量：为此我们需要求助于著名的海涅-伯勒定理，该定理指出，每个封闭的、有边界的R的子集B是*紧凑的*：给定任何涵盖B的开放集O_α（即$B\subset_\alpha O_\alpha$），有一个有限的子集合$(O_\alpha i)_{i\le n}$，它仍然覆盖B，即$B\subset \bigcup_{i=1}^{n} O_{\alpha i}$（证明见[1]）。

定理2.4

一个区间的外量等于其长度。

证明

如果I是无界的，那么很明显，它不能被一个总长度有限的区间系统所覆盖。这表明，$m^*(I)=\infty$，因此$m^*(I)=l(I)=\infty$。

所以我们把自己限制在有界区间。

步骤1. $m^*(I)\le l(I)$。

我们声称，$l(I)\in Z_I$。取以下的区间序列：$I_1=I$, $I_n=[0,0]$ for $n\ge 2$. 这个序列覆盖了集合I，其总长度等于I的长度，因此$l(I)\in Z_I$。这就足够了，因为Z_I的下限不能超过它的任何元素。

第二步，$l(I)\le m^*(I)$。

(1) $I=[a,b]$。我们将证明，对于任何$\varepsilon>0$

$$l([a,b])\le m^*([a,b])+\varepsilon。\tag{2.1}$$

这就足够了，因为我们可以得到所需的不等式，通过极限，$\varepsilon\to 0$。（注意，如果x，$y\in$R，$y>x$，那么有一个$\varepsilon>0$，有$y>x+\varepsilon$，例如$\varepsilon=\frac{1}{2}(y-x)$）

所以我们取一个任意的$\varepsilon>0$.根据外测量的定义，我们可以找到一个覆盖$[a,b]$的区间序列I_n，以便

$$\sum_{n=1}^{\infty} l(I_n)\le m^*([a,b])+\frac{\varepsilon}{2}.\tag{2.2}$$

$n=1$

我们将把每个区间略微增加到一个开放区间。让I_n的端点为a_n，b_n，我们取

$$J_n = \left(a_n - \frac{\varepsilon}{2^{n+2}}, b_n + \frac{\varepsilon}{2^{n+2}}\right).$$

显而易见的是

$$l(I_n) = l(J_n) - \frac{\varepsilon}{2^{n+1}}$$

以致于

$$\sum_{n=1}^{\infty} l(I_n) = \sum_{n=1}^{\infty} l(J_n) - \frac{\varepsilon}{2}.$$

我们将其插入(2.2)中，我们有

$$\sum_{n=1}^{\infty} l(J_n) \leq m^*([a, b]) + \varepsilon。 \tag{2.3}$$

新的区间序列当然覆盖了$[a, b]$，所以根据海涅-博雷尔定理，我们可以选择有限数量的J_n来覆盖$[a, b]$（集合$[a, b]$在R中是紧凑的）。我们可以在这个有限的家族中加入一些区间，形成一个
序列的初始段(J_n)--只是为了符号的简单性。所以对于
某个有限指数m，我们有

$$[a, b] \subseteq \bigcup_{n=1}^{m} J_n。 \tag{2.4}$$

让$J_n = (c_n, d_n)$。将$c = \min\{c_1, \ldots, c_m\}$, $d = \max\{d_1, \ldots, d_m\}$。覆盖（2.4）意味着$c<a$，$b<d$，因此$l([a, b])<d-c$。

接下来，数字$d - c$肯定小于J的总长度$_n$, $n = 1, \ldots, m$（发生了一些重叠）和

$$l([a, b]) < d - c < \sum_{n=1}^{m} l(J_n)。 \tag{2.5}$$

现在只要把(2.3)和(2.5)放在一起，就可以推导出(2.1)(由于所有项都是非负的，所以有限和小于或等于系列之和)。

(2) $I=(a, b)$。和以前一样，只要证明（2.1）就可以了。让我们固定任何$\varepsilon>0$。

$$
\begin{aligned}
l((a,b)) &= l\left(\left[a+\frac{\varepsilon}{2}, b-\frac{\varepsilon}{2}\right]\right)+\varepsilon \\
&\le m^*\left(\left[a+\frac{\varepsilon}{2}, b-\frac{\varepsilon}{2}\right]\right)+\varepsilon \ \text{(通过(1))} \\
&\le m^*((a,b))+\varepsilon \ \text{(由命题2.3)}。
\end{aligned}
$$

(3) $I = [a, b]$或$I = (a, b)$。

$$
\begin{aligned}
l(I) = l((a, b)) &\le m^*((a, b)) \quad (\text{通过}(2)) \\
&\le m^*(I) \quad (\text{根据命题}2.3)
\end{aligned}
$$

这就完成了证明。 □

在证明了外测量与区间的自然长度概念相吻合之后，我们现在需要研究它的属性。下一个定理给了我们一个重要的技术工具，在许多证明中都会用到。

定理2.5

外部测量是可数次加性的，即对于任何序列的集合$\{E_n\}$来说

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \le \sum_{n=1}^{\infty} m^*(E_n)。$$

(注意，这里的两边都可能是无限的）。

证明(热身)

让我们首先证明一个更简单的声明：

$$m^*(E_1 \cup E_2) \le m^*(E_1) + m^*(E_2)。$$

取一个$\varepsilon>0$，我们显示一个更容易的不等式

$$m^*(E_1 \cup E_2) \le m^*(E_1) + m^*(E_2) + \varepsilon。$$

但这是充分的，因为取$\varepsilon=\frac{1}{n}$，让$n\to\infty$，我们得到的是我们需要。

因此，对于任何$\varepsilon>0$，我们发现覆盖序列$(I_k^1)_{k\ge1}$的E_1和$(I_k^2)_{k\ge1}$的E_2，这样，

$$\sum_{k=1}^{\infty} l(I_k^1) \le m^*(E_1) + \frac{\varepsilon}{2}$$

$$\sum_{k=1}^{\infty} l(I_k^2) \le m^*(E_2) + \frac{\varepsilon}{2}$$

k 2 2

$k=1$

因此，加起来、

$$\sum_{k=1}^{\infty} l(I_k^1) + \sum_{k=1}^{\infty} l(I_k^2) \le m^*(E_1) + m^*(E_2) + \varepsilon。$$

区间序列（$I_1^1, I_1^2, I_2^1, I_2^2, I_3^1, I_3^2, \ldots$）覆盖$E_1 \cup E_2$因此

$$m^*(E_1 \cup E_2) \le \sum_{k=1}^{\infty} l(I_k^1) + \sum_{k=1}^{\infty} l(I_k^2)$$

这与前面的不等式结合起来，就得到了结果。 □

证明（该定理的证明）

如果有手边是无限的，那么这个不等式当然是真的。因此，假设$\sum_{n=1}^{\infty} m^*(E_n) < \infty$。对于每个给定的$\varepsilon>0$和$n\ge1$，找到$E_n$的一个覆盖序列$(I_k^n)_{k\ge1}$，其中有

$$\sum_{k=1}^{\infty} l(I_k^n) \le m^*(E_n) + \frac{\varepsilon}{2^n}.$$

迭代后的序列收敛：

$$\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} l(I_k^n) \le \sum_{n=1}^{\infty} m^*(E_n) + \varepsilon < \infty$$

并且由于它的所有项都是非负的、

$$\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} l(I_k^n) = \sum_{n,k=1}^{\infty} l(I_k^n)。$$

区间系统$(I_k^n)_{k,n\ge1}$覆盖$\bigcup_{n=1}^{\infty} E_n$因此

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \le \sum_{n,k=1}^{\infty} l(I_k^n) \le \sum_{n=1}^{\infty} m^*(E_n) + \varepsilon$$

为了完成证明，我们让$\varepsilon\to0$。 □

类似的结果当然也适用于有限族$(E_n)_{n=1}^m$：

$$m^*\left(\bigcup_{n=1}^{m} E_n\right) \le \sum_{n=1}^{m} m^*(E_n)。$$

这是对定理2.5的一个推论，对于$k>m$，$E_k=\varnothing$。

练习2.4

证明如果$m^*(A)=0$，那么对于每个B，$m^*(A\cup B)=m^*(B)$。

提示 同时使用外部测量的单调性和次可加性。

练习2.5

证明如果$m^*(A\Delta B) = 0$，那么$m^*(A) = m^*(B)$。

提示 注意，$A\subseteq B\cup(A\Delta B)$。

在这一节的最后，我们将以一个简单而直观的外延属性来结束。请注意，如果我们沿实线移动一个区间，其长度不会改变：例如，$l([a, b]) = l([a + t, b + t]) = b\ a$。由于外量子是以区间的长度来定义的，所以很自然地期望它能共享

这个属性。对于$A\subset$R和$t\in$R，我们把$A+t=\{a+t : a\in A\}$。

命题2.6

外围测量是翻译不变的，即

$$m^*(A) = m^*(A + t)$$

对于每个A和t。

提示 结合两个事实：当区间被移位时，区间的长度不会改变，而外部测量是由覆盖物的长度决定的。

2.3 Lebesgue可测量的集合和Lebesgue 措施

有了外测量，次可加性（如定理2.5中）是我们能得到的最大限度。然而，我们希望确保如果集合（E_n）是成对不相交的（即如果$i = j$，$E_i \cap E_j = \varnothing$），那么定理2.5中的不等式就变成了相等。事实证明，这在一般情况下对外部测量来说不会是真的，尽管它失败的例子是相当难以构建的（我们在附录中给出了这样的例子）。但我们的愿望是完全合理的：任何 "长度函数 "至少应该是有限加性的，因为将一个集合分解成有限多的不相交的片

断不应该改变其长度。此外，由于我们通过用 "更简单 "的集合（即区间）来逼近复杂的集合来构造我们的长度函数，要求一个*连续性的属性*似乎是公平的：如果成对的不相交的（E_n ）有并集E、

那么集合B的长度$_n = E\setminus \bigcup_{k=1}^{n} E_k$，可望减少到0，因为$n\to\infty$.将这一点与有限可加性结合起来，就很自然地要求 "长度 "应该是*可数可加的*，也就是

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} m^*(E_n) \quad 时，E_i \cap E_j = \varnothing \text{ for } i \neq j.$$

因此，我们的任务是找到R中具有这种情况的集合的类别。属性。事实证明，这也是抽象概念的关键属性，即度量，而我们将用它来为概率提供数学基础。为了定义具有这种性质的 "好 "集合，似乎也可以说这样的集合应该分摊R中*每一个*集合的外部度量值正如我们在下面的定义2.3中所说的那样，适当的。值得注意的是，这个简单的要求就足以保证我们的 "好 "集合具有我们对它们要求的所有属性！"！

定义2.3

如果对于每一个$A\subseteq$R的集合，我们有：$E\subseteq$R是（Lebesgue）*可测量的*

$$m^*(A) = m^*(A\cap E) + m^*(A\cap E)^c \tag{2.6}$$

其中$E^c = R\setminus E$，我们写$E\in$M。

显然，我们有$A=（A\cap E）\cup（A\cap E^c）$，因此根据定理2.5，我们有

$$m^*(A) \le m^*(A\cap E) + m^*(A\cap E)^c$$

所以我们未来验证（2.6）的任务已经简化：$E \in$ M

当且仅当以下不等式成立

$$m^*(A) \ge m^*(A\cap E) + m^*(A\cap E^c) \text{ 对于所有} A\subseteq \text{R。} \tag{2.7}$$

现在我们给出一些可测集的例子。

定理2.7

(i) 任何空集都是可测量的。

(ii) 任何区间都是可以测量的。

证明

(i) 如果N是一个空集，那么（命题2.2）$m^*(N)=0$。所以对于任何$A \in$ R，我们有

$$m^*(A \cap N) \le m^*(N) = 0，因为A \cap N \subseteq N$$

$$m^*(A\cap N^c)\le m^*(A)\quad 因为 A\cap N^c\subseteq A$$

并加在一起，我们已经证明了（2.7）。

(ii) 让$E=I$是一个区间。例如，假设$I=[a,b]$。取任意一个$A\subseteq\mathbb{R}$和$\varepsilon>0$，找一个A的覆盖，其中

$$m^*(A)\le\sum_{n=1}^{\infty}l(I_n)\le m^*(A)+\varepsilon。$$

显然，区间$I_n'=I_n\cap[a,b]$覆盖$A\cap[a,b]$因此

$$m^*(A\cap[a,b])\le\sum_{n=1}^{\infty}l(I_n')。$$

区间$I_n'=I_n\cap(-\infty, a)$，$I_n''=I_n\cap(b,+\infty)$覆盖$A\cap[a,b]^c$，所以

$$m^*(A\cap[a,b]^c)\le\sum_{n=1}^{\infty}l(I_n')+\sum_{n=1}^{\infty}l(I_n'')。$$

将上述三个不等式放在一起，我们得到（2.7）。

如果I是无界的，比如$I=[a,\infty]$，那么证明就更简单了，因为它是只要考虑$I_n'=I_n\cap[a,\infty]$和$I_n''=I_n\cap(-\infty,a)$即可。 □

现在可以证明R的所有Lebesgue可度量M集类的基本属性。它们分为两类：首先，我们证明M中的某些集合操作会再次产生M中的集合（这些就是我们所说的'闭合属性'）；其次，我们证明对于M中的集合，外部度量m^*具有上面宣布的可数加性的属性。

定理2.8

(i) $\mathbb{R}\in\mathcal{M}$、

(ii) 如果$E\in\mathcal{M}$，那么$E^c\in\mathcal{M}$、

(iii) 如果$E_n\in\mathcal{M}$，对于所有$n=1,2,\ldots$，那么$\bigcup_{n=1}^{\infty}E_n\in\mathcal{M}$。

此外，如果$E_n\in\mathcal{M}$，$n=1,2,\ldots$，$E_j\cap E_k=\varnothing$，为$j\neq k$，则

$$m^*\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}m^*(E_n)。\tag{2.8}$$

备注2.1

这个结果是本章中最重要的定理，为后面的所有内容提供了基础。它还使我们能够给正在讨论的量命名。

条件（i）-（iii）意味着是一个M σ 场。换句话说，如果一个集合族包含基集，并且在补数和可数联合下是封闭的，我们就说它是一个σ-场。一个定义在σ-场上的[0,]值函数，如果对成对不相邻的集合来说满足(2.8)，即它是*可数加性的，则*被称为*度量。*

对*度量空间的*另一种更为抽象和普遍的方法是以上述属性作为*公理*开始，也就是说，如果*Ω是*一个抽象的给定集合，F是Ω的子集的σ*场，*而μ：F [0,] 是一个满足(2.8)的函数(用μ代替m^*)。那么，在R上定义Lebesgue度量的任务就变成了验证，用M 和$m = m^*$ 上如上定义，即三联体（R, M, m）满足这些公理，即成为一个测量空间。

尽管概率论的要求意味着我们必须在适当的时候考虑这样的一般度量空间，但我们还是选择了以勒贝斯格度量为基本例子的更为具体的方法，以证明这一重要的度量空间是如何从以下方面自然产生的
对R中集合的 "长度 "的考虑，并导致了一个整合的理论
这大大扩展了黎曼的观点。它也足以让我们发展
大多数重要的概率分布的例子。

证明（该定理的证明）

(i) 让$A \subseteq \mathbb{R}$。注意$A \cap \mathbb{R} = A$，$\mathbb{R}^c = \varnothing$，所以$A \cap \mathbb{R}^c = \varnothing$。现在（2.6）读作$m^*(A) = m^*(A) + m^*(\varnothing)$，当然是真的，因为$m^*(\varnothing) = 0$。

(ii) 假设$E \in$ M，并取任何$A \subseteq \mathbb{R}$。我们必须证明(2.6)对E^c，即。

$$m^*(A) = m^*(A \cap E^c) + m^*(A \cap (E^c)^c)$$

但由于$(E^c)^c = E$，这就减少了E的条件，而这一条件通过hy-pothesis成立。

我们把③的证明分成几个步骤。但首先：

一个热身。 假设$E_1 \cap E_2 = \varnothing$, E_1 , $E_2 \in$M。我们将证明：

$E_1 \cup E_2 \in$M 和 $m^*(E_1 \cup E_2) = m^*(E_1) + m^*(E_2)$ 。

设$A\subseteq \mathbb{R}$。我们有E的条件$_1$：

$$m^*(A) = m^*(A\cap E_1) + m^*(A\cap E_1^c)。\qquad (2.9)$$

现在，将（2.6）应用于E_2，用$A\cap E_1^c$代替A：

$$m^*(A\cap E_1^c) = m^*((A\cap E_1^c)\cap E_2) + m^*((A\cap E_1^c)\cap E_2^c)。$$
$$= m^*(A\cap(E_1^c\cap E_2)) + m^*(A\cap(E_1^c\cap E_2^c))$$

(情况如图2.2所示）。

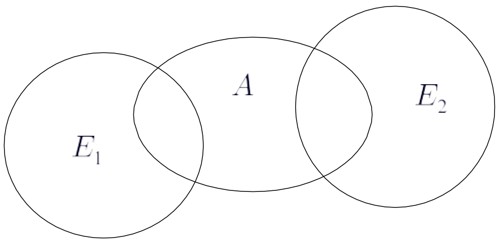

图2.2 集合A, E_1, E_2

由于E_1和E_2是不相交的，所以$E_1^c\cap E_2 = E_2$。根据德摩根定律，$E_1^c\cap E_2^c = (E_1\cup E_2)^c$。我们代入，我们有

$$m^*(A\cap E_1^c) = m^*(A\cap E_2) + m^*(A\cap(E_1\cup E_2)^c)。$$

将其代入（2.9），我们得到

$$m^*(A) = m^*(A\cap E_1) + m^*(A\cap E_2) + m^*(A\cap (E_1\cup E_2)^c)。\qquad (2.10)$$

Now by the subadditivity property of m^* we have

$$m^*(A\cap E_1) + m^*(A\cap E_2) \geq m^*((A\cap E_1)\cup(A\cap E_2))$$
$$= m^*(A\cap(E_1\cup E_2))$$

所以（2.10）得出

$$m^*(A) \geq m^*(A\cap(E_1\cup E_2)) + m^*(A\cap(E_1\cup E_2)^c)$$

这足以使$E_1\cup E_2$属于$\mathcal{M}$（反不等式总是真实的，正如之前观察到的（2.7））。

最后，在（2.10）中把$A=E_1\cup E_2$，得到$m^*(E_1\cup E_2) = m^*(E_1) + m^*(E_2)$、这就完成了论证。 □

我们回到定理的证明上来。

证明

步骤1。如果成对不相交的E_k，$k = 1, 2, \ldots$，都在M中，那么它们的联盟就在
M和（2.8）成立。

我们像暖和的证明中一样开始，我们有

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c)$$

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap (E_1 \cup E_2)^c)$$

（见（2.10）），经过n个步骤，我们预计

$$m^*(A) = \sum_{k=1}^{n} m^*(A \cap E_k) + m^*\Big(A \cap \Big(\bigcup_{k=1}^{n} E_k\Big)^c\Big)。\qquad (2.11)$$

让我们通过归纳法来证明这一点。n=1的情况是上面的第一行。假设

$$m^*(A) = \sum_{k=1}^{n-1} m^*(A \cap E_k) + m^*\Big(A \cap \Big(\bigcup_{k=1}^{n-1} E_k\Big)^c\Big)。\qquad (2.12)$$

由于$E_n \in$M，我们可以应用（2.6），用$A \cap \big(\bigcup_{k=1}^{n-1} E_k\big)^c$来代替$A$：

$$m^*\Big(A \cap \Big(\bigcup_{k=1}^{n-1} E_k\Big)^c\Big) = m^*\Big(A \cap \Big(\bigcup_{k=1}^{n-1} E_k\Big)^c \cap E_n\Big) + m^*\Big(A \cap \Big(\bigcup_{k=1}^{n-1} E_k\Big)^c \cap E_n^c\Big)。\qquad (2.13)$$

现在我们进行与热身赛相同的观察：

$$\Big(\bigcup_{k=1}^{n-1} E_k\Big)^c \cap E_n = E_n \qquad (E_i\text{是成对不相交的})、$$

$$\Big(\bigcup_{k=1}^{n-1} E_k\Big)^c \cap E_n^c = \Big(\bigcup_{k=1}^{n} E_k\Big)^c \qquad (\text{根据德摩根定律})。$$

将这些插入（2.13），我们得到

$$m^*\Big(A \cap \Big(\bigcup_{k=1}^{n-1} E_k\Big)^c\Big) = m^*(A \cap E_n) + m^*\Big(A \cap \Big(\bigcup_{k=1}^{n} E_k\Big)^c\Big)、$$

并将其插入归纳假设（2.12）中，我们得到

$$m^*(A) = \sum_{k=1}^{n-1} m^*(A \cap E_k) + m^*(A \cap E_n) + m^*\Big(A \cap \Big(\bigcup_{k=1}^{n} E_k\Big)^c\Big)$$

作为完成归纳步骤的要求。因此(2.11)通过归纳法对所有n都成立。

正如下一步将看到的那样，E_k 是成对不相交的这一事实对于确保它们的结合属于$\mathcal{M}$并不是必需的。然而，有了这个假设，我们就有了(2.11)中的平等，否则就不成立了。这个平等将使我们能够证明可数可加性（2.8）。

自

$$\bigcap_{k=1}^{n} E_k^c \supseteq \bigcap_{k=1}^{\infty} E_k^c、$$

从（2.11）通过单调性（命题2.3），我们得到

$$m^*(A) \geq \sum_{k=1}^{n} m^*(A \cap E_k) + m^*\Big(A \cap \bigcap_{k=1}^{\infty} E_k^c\Big).$$

在我们传递到极限$n\to\infty$之后，这个不等式仍然成立：

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\Big(A \cap \bigcap_{k=1}^{\infty} E_k^c\Big)。 \tag{2.14}$$

根据可数次加性（定理2.5）

$$\sum_{k=1}^{\infty} m^*(A \cap E_k) \geq m^*\Big(A \cap \bigcup_{k=1}^{\infty} E_k\Big)$$

于是

$$m^*(A) \geq m^*\Big(A \cap \bigcup_{k=1}^{\infty} E_k\Big) + m^*\Big(A \cap \bigcap_{k=1}^{\infty} E_k^c\Big) \tag{2.15}$$

如要求的那样。所以我们已经证明$\bigcup_{k=1}^{\infty} E_k \in \mathcal{M}$，因此（2.15）的两边是相等的。(2.14)的右边被挤压在左边和右边之间。

(2.15)的右边，得到的是

$$m^*(A) = \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\Big(A \cap \bigcap_{k=1}^{\infty} E_k^c\Big)。 \tag{2.16}$$

这里的平等是假设E_k 是成对不相交的结果。它对任何集合A都成立，所以我们可以插入$A = \bigcup_{j=1}^{\infty} E_j$。右边的最后一项是零，因为我们有$m^*$（$\varnothing$）。接下来$\left(\bigcup_{j=1}^{\infty} E_j\right) \cap E_n = E_n$ 和

所以我们有（2.8）。

第二步。如果E_1，$E_2 \in \mathcal{M}$，那么$E_1 \cup E_2 \in \mathcal{M}$（不一定不相交）。

我们再次像热身那样开始：

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c)。 \tag{2.17}$$

接下来，将（2.6）应用于E_2，用$A \cap E_1^c$代替A，我们得到

$$m^*(A \cap E_1^c) = m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c)。$$

我们将其插入（2.17）中，得到

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c)。\qquad (2.18)$$

根据德摩根定律，$E_1^c \cap E_2^c = (E_1 \cup E_2)^c$ 所以（如前所述）。

$$m^*(A \cap E_1^c \cap E_2^c) = m^*(A \cap (E_1 \cup E_2)^c)。\qquad (2.19)$$

根据m的次可加性*，我们有

$$m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) \geq m^*(A \cap E_1 \cup E_2)。\qquad (2.20)$$

将（2.19）和（2.20）插入（2.18）中，我们得到

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c)$$

根据需要。

步骤3。如果$E_k \in$ M，$k = 1, \ldots, n$，那么$E_1 \cup \ldots \cup E_n \in$ M （不一定不相交）。

我们通过归纳法进行论证。对于n=1，没有什么可证明的。假设要求对n-1来说是真的。那么

$$E_1 \cup \ldots \cup e_n = (e_1 \cup \ldots \cup e_{n-1}) \cup e_n$$

因此，结果由步骤2得出。

步骤4。如果E_1，$E_2 \in$M，那么$E_1 \cap E_2 \in$M。

我们有$E_1^c, E_2^c \in$M由(ii)，$E_1^c \cup E_2^c \in$M由步骤2，$(E_1^c \cup E_2^c)^c \in$M由(ii)再次，但根据德摩根定律，最后一组等于$E_1 \cap E_2$。

步骤5。一般情况：如果$E_1, E_2, \ldots$ 都在M中，那么$\bigcup_{k=1}^{\infty} E_k$也是如此。

让$E_k \in$M, $k = 1, 2, \ldots$。我们定义一个辅助序列，由成对不相交的集合F_k，其并集与E_k相同：

$$F_1 = E_1$$
$$f_2 = e_2 \setminus e_1 = e_2 \cap e_1^c$$
$$f_3 = e_3 \setminus (e_1 \cup e_2) = e_3 \cap (e_1 \cup e_2)^c$$
$$\ldots$$

$$f_k = e_k \setminus (e_1 \cup \dots \cup e_{k-1}) = e_k \cap (e_1 \cup \dots \cup e_{k-1})^c$$、

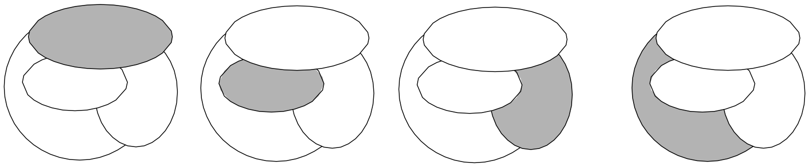

图2.3 集合F_k

见图2.3。

根据步骤3和4，我们知道所有的F_k都在M中。根据结构，它们是成对不相交的，所以根据步骤1，它们的联合体也在M中。我们将证明：

$$\bigcup_{k=1}^{\infty} F_k = \bigcup_{k=1}^{\infty} E_k \text{。}$$

这将完成证明，因为后者现在是在M中。

$$\bigcup_{k=1}^{\infty} F_k \subseteq \bigcup_{k=1}^{\infty} E_k$$

是显而易见的，因为对于每个k，F_k E的定义。对于逆，让$a_k \in \bigcup_{\infty=1} E_k$。把$S = \{n \in \mathbb{N} : a \in E_n\}$，它是非空的，因为$a$属于联盟。让$n_0 = \min S$ $\in S$.如果$n_0 = 1$，那么$a \in E_1 = F_1$。假设$n_0 > 1$。$a \in E_{n_0}$ 而且，根据n的定义，$a \notin E_1$，……。，$a \notin E_{n_0-1}$.根据F_{n_0}的定义，这意味着$a \in F_{n_0}$，所以a是在$\bigcup_{k=1}^{\infty} F_k$。 □

使用德摩根定律，你应该很容易验证一个额外的属性

M.

命题2.9

如果$E_k \in$M，k=1，2，...，则

$$E = \bigcap_{k=1}^{\infty} E_k \in \text{M。}$$

M 我们可以将

Lebesgue可测量集的 家族的属性总结如下：

1. M *在可数的联合、可数的交叉和补数下是封闭的。它包含区间和所有*

空集。

定义2.4

我们将写$m(E)$而不是$m^*(E)$，对于任何E在 $\mathcal{M}$ 中的任何一个E，并称$m(E)$为集合E的*Lebesgue度量*。

因此，定理2.8和2.4现在读起来如下，并描述了我们辛辛苦苦建立的结构：

1. *Lebesgue度量m* $\mathcal{M} \to [0,\infty]$是一个可数加性的集合函数，定义在可测集合的σ场上$\mathcal{M}$。一个区间的Lebesgue度量等于其长度。空集的Lebesgue度量为零。

2.4 Lebesgue的基本属性 措施

由于Lebesgue度量无非是限制在一类特殊集合上的外度量，所以外度量的一些属性会自动被Lebesgue度量所继承：

命题2.10

假设A、$B\in\mathcal{M}$。

(i) 如果$A\subset B$，那么$m(A)\le m(B)$。

(ii) 如果$A\subset B$且$m(A)$是有限的，那么$m(B\setminus A)=m(B)-m(A)$。

(iii) m是翻译不变的。

由于$\varnothing\in\mathcal{M}$，我们可以在（2.8）中对所有$i>n$取$E_i=\varnothing$，从而得出Lebesgue度量是加法的结论：如果$E_i\in\mathcal{M}$是成对不相交的，那么

$$m\left(\bigcup_{i=1}^{n} E_i\right)=\sum_{i=1}^{n} m(E_i)。$$

练习2.6

找到一个公式，用各个集合和它们的交集的度量来描述$m(A\cup B)$和

$m(A\ B\ C)$（我们不假设这些集合是成对不相交的）。

忆及两个集合的*对称差*$A\Delta B$的定义为
$A\Delta B = (A \setminus B)$? $(B \setminus A)$ 下面的结果也很容易检查：

命题2.11

如果$A\in\mathcal{M}$，m（$A\Delta B$）=0，那么$B\in\mathcal{M}$，m（A）=m（B）。

提示 回顾一下，空集属于$\mathcal{M}$，空集的子集是空的。

正如我们在第一章中指出的，R中的每一个开放集都可以表示为可数的开放区间的联合。这确保了R中的开放集是可测量的，因为它包含区间，并且在可数区间下是封闭的$\mathcal{M}$。

联。我们可以通过包含A的开放集序列的度量来近似任何$A\in\mathcal{M}$的Lebesgue度量。 的 Lebgue

度量，可以通过包含A的一连串开放集的度量从上面近似：

定理2.12

(i) 对于任何$\varepsilon>0$，$A\subset$R，我们可以找到一个开放集O，以便

$$A\subset O、\quad m(O)\le m^*(A)+\varepsilon。$$

因此，对于任何$E\in\mathcal{M}$，我们可以找到一个包含E的开放集O，使得m（$O\backslash E$）$<\varepsilon$。

(ii) 对于任何一个$A\subset$R，我们可以找到一个开放集O的序列$_n$，以便

$$A\subset\bigcap_n O_n、\quad m(\bigcap_n O_n)=m^*(A)。$$

证明

(i) 根据$m^*(A)$的定义，我们可以找到一个序列(I_n)的区间，其中$A\subset\bigcup_n I_n$和$\sum_{n=1}^{\infty} l(I_n)-\frac{\varepsilon}{2}\le m^*(A)$。每个$I_n$都包含在一个开放的区间内，其长度与$I_n$的长度非常接近；如果$I_n$的左右端点是$a_n$和$b_n$分别设$J_n=(a_n-\frac{\varepsilon}{2^{n+2}}, b_n+\frac{\varepsilon}{2^{n+2}})$。设$O=\bigcup_n J_n$，其中是开放的。那么，$A\subset O$

和

$$m(O)\le\sum_{n=1}^{\infty} l(J_n)\le\sum_{n=1}^{\infty} l(I_n)+\frac{\varepsilon}{2}\le m^*(A)+\varepsilon。$$

当$m(E)<\infty$的情况下，最后一句话马上就可以从提议中的(ii)得出

。

第2.10条，因为此时$m(O\setminus E) = m(O) - m(E) \leq \varepsilon$.当$m(E)= \infty$ 我们首先把R写成有限区间的可数联盟：R = $\bigcup_n (-n, n)$。现在

$E_n = E \cap (n, n)$ 具有有限的度量，所以我们可以找到一个开放的$O_n \supset E_n$，其

$m(O_n \setminus E_n) \le \frac{\varepsilon}{2^n}$。集合$O = \bigcup_n O_n$是开放的并且包含$E$。

$$O \setminus E = \left(\bigcup_n O_n\right) \setminus \left(\bigcup_n E_n\right) \subset \bigcup_n (O_n \setminus E_n)$$

因此，$m(O\setminus E) \le \sum_n m(O_n \setminus E_n) \le \varepsilon$，符合要求。

(ii) 在(i)中，使用$\varepsilon = \frac{1}{n}$，让$O_n$是这样得到的开放集。用$E = \bigcap_n O_n$
我们得到一个包含A的可测量集，使得$m(E) < m(O_n) \le m^*(A) + \frac{1}{n}$
对于每一个n，因此，结果如下。 □

备注2.2

定理2.12说明了σ场的封闭特性所允许的运动自由是如何被利用的
$\mathcal{M}$可以通过对任何集合$A \subset \mathbb{R}$产生一个可测量的集合$O \supset A$，它是由具有两个的开放区间得到的
操作（可数的联合，然后是可数的交叉），其
措施等于A的外部措施。

最后我们表明，可测量集的单调序列在*m方面*的表现正如人们所期望的那样
。

定理2.13

假设在所有$n \ge 1$的情况下，$A_n \in \mathcal{M}$。那么我们有
(i) 如果$A_n \subset A_{n+1}$ 对于所有的n，那么

$$m\left(\bigcup_n A_n\right) = \lim_{n\to\infty} m(A_n)、$$

(ii) 如果$A_n \supset A_{n+1}$ 为所有n，且$m(A_1) < \infty$，则

$$m\left(\bigcap_n A_n\right) = \lim_{n\to\infty} m(A_n)。$$

证明

(i) 设$B_1 = A_1$，$B_i = A_i - A_{i-1}$，$i > 1$。然后$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$，则$B_i \in$M是成对不相交的，所以

$$m(\bigcup_i A_i) = m(\bigcup_i B_i)$$

$$= \sum_{i=1}^{\infty} m(B_i) \text{ (通过可数加性)}$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} m(B_i)$$

$$=\text{四分之一}_{n\to\infty}\, m(\bigcup_{n=1}^{n} B_i) \text{ (通过可加性)}$$

$$=\text{四分之一}_{n\to\infty}\, m(A_n)\text{、}$$

因为$A_n = \bigcup_{i=1}^{n} B_i$，通过建设--见图2.4。

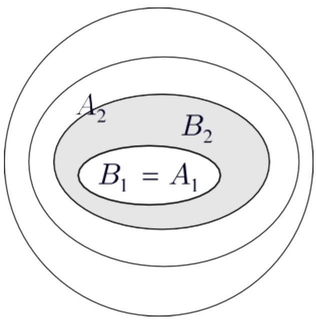

图2.4 集合A_n, B_n

(ii) $A_1 \quad (A_1 = \varnothing \subset A_1 \quad (A_2 \ldots \subset A_1 \quad (A_n \ldots\ldots)$ 对于所有的n，所以由（i）可知

$$m(\bigcup_n (A_1 \setminus A_n)) = \lim_{n\to} m(A_1 \setminus A_n)$$

而由于$m(A_1)$是有限的，所以$m(A_1 \setminus A_n) = m(A_1) - m(A_n)$。另一方面、$\bigcup_n (A_1 \setminus A_n) = A_1 \setminus \bigcap_n A_n$，所以

$$m(\bigcup (A_1 \setminus A_n)) = m(A_1) - m(\bigcap A_n) = m(A_1) - \lim m(A_n)\text{。}$$

n n $n\to\infty$

结果如下。 □

备注2.3

定理2.13的证明仅仅依赖于m的可数加性和$[0,\infty]$中数列之和的定义，即

$$\sum_{i=1}^{\infty} m(A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} m(A_i)。$$

因此，这个结果不仅对我们所构筑的集合函数m是真的，而且对任何定义在σ场上的可数加性集合函数也是真的。这也使我们得出下面的主张，虽然我们在这里只考虑到m，但它实际上是可数加性集合函数的特征。

定理2.14

集合函数m满足：

(i) m是有限加性的，即对于成对不相交的集合(A_i)，我们有

$$m\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} m(A_i)$$

对于每一个n；

(ii) m在∅处是连续的，即如果(B_n)减少到∅，则$m(B_n)$减少到0。

证明

为了证明这一主张，请回顾，m：$\mathcal{M}' \to [0,\infty]$是可数加性的。这意味着(i)，正如我们已经看到的。为了证明(ii)，考虑一个序列(B_n)，其中递减到∅。然后$A_n = B_n \setminus B_{n+1}$在$\mathcal{M}$中定义了一个不相交的序列，并且$\bigcup_n A_n = B_1$。我们可以假设$B_1$是有界的，所以$m(B_n)$对所有$n$都是有限的，因此，根据命题2.10(ii)，$m(A_n) = m(B_n) - m(B_{n+1}) \geq 0$，因此我们有

$$\begin{aligned} m(B_1) &= \sum_{n=1}^{\infty} m(A_n) \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} [m(B_n) - m(B_{n+1})] \\ &= m(B_1) - \lim_{n\to\infty} m(B_n) \end{aligned}$$

这表明$m(B_n) \to 0$，符合要求。

$m(B)_n$

□

2.5 Borel 集

定义 $\mathcal{M}$ 的定义并不容易用来验证某个特定的集合是否属于 $\mathcal{M}$ $\mathcal{M}$；在我们的证明中，我们不得不相当努力地证明在各种操作下是封闭的。因此，在我们的武器库中增加另一个结构是很有用的；这个结构更直接地显示了开放集（实际上是开放区间）和σ场的结构是如何处于我们的核心。

我们已经开发的许多概念。

我们从一个辅助结构开始，使我们能够产生新的σ-场。

定理2.15

σ-场家族的交集是一个σ-场。

证明

让F_α是$\alpha\in\Lambda$的σ-场（索引集Λ可以是任意的）。把

$$F = \bigcap_{\alpha\in\Lambda} F_\alpha。$$

我们验证定义的条件。

1. $R\in F_\alpha$，对于所有$\alpha\in\Lambda$，所以$R\in F$。
2. 如果$E\in F$，那么对于所有$\alpha\in\Lambda$，$E\in F_\alpha$。由于F_α是σ-场，所以$E^c \in F_\alpha$. 所以 $E^c \in F$
3. 如果$E_k \in F$ for $k = 1, 2, \ldots$，则$E_k \in F_\alpha$，所有α，k，因此$\bigcup_{k=1}^{\infty} E_k \in F_\alpha$ 所有α，所 $\bigcup_{k=1}^{\infty} E_k \in F$ □

定义2.5

把

$$\mathcal{B} = \bigcap \{F：F是一个包含所有区间的σ场\}。$$

我们说 $\mathcal{B}$ 是由所有区间生成的σ场，我们称其元素为 *Borel*集（以 $\mathcal{B}$ Emile Borel 1871-1956年命名）。显然，它是最小的σ-

包含所有区间的字段。一般来说，我们说 G 如果G=T{F :F是一个σ场，使得F⊃A}，那么G就是由一个集合族A生成的σ场。

例2.1

(Borel集)下面的例子说明了如何利用σ-场B的封闭特性来验证R中大多数熟悉的集属于B。

(i) 根据构造，所有的区间都属于 ，并且由于是一个σ 场，所有的开放集都必须属于 ，因为任何开放集都是（开放）区间的可数联盟。

(ii) 可数集是伯乐集，因为每一个都是[*a, a*]形式的封闭区间的可数联合；特别是N和Q是伯乐集。因此，作为一个伯乐集的补数，无理数集也是伯乐集。同样地，有限集和共富集也是博勒集。

的定义也非常灵活--只要我们从所有特定类型的区间开始，这些区间集合就会产生相同的Borel σ场。 的定义也非常灵活--只要我们从某一特定类型的*所有*区间开始，这些集合就会产生相同的博雷尔σ-*场*：

定理2.16

如果我们不取所有区间族，而是取所有开放区间、所有封闭区间、所有形式为（*a*，∞）（或形式为[*a*，∞]，（-∞，*b*），或（-∞，*b*））*的*区间、所有开放集或所有封闭集，那么由它们生成的σ*场*与B相同。

证明

例如，考虑由开放区间族*OI*产生的σ场
并用C表示：

$$C = \bigcap \{F \supseteq OI，F是一个σ\text{-}场\}。$$

我们必须证明 B = C 由于开放区间是区间，*OI* ⊂ *I*（所有区间的族），那么

{fnFangSong_GB2312bord1shad1pos(200,288)}{fnFangSong_GB2312bord1shad1pos(200,288)}就可以了。

即包含*I*的所有σ域的集合小于包含较小家族*OI的*所有*σ域*的集合，因为包含较大的家族是一个更苛刻的要求，所以这样的对象较少。在我们取两边的交集后，包容就会被颠覆，因此 C⊂B

(小家庭的交集更大，因为属于其每个成员的要求不那么严格）。

我们将表明，包含C所有的区间。这就足够了，因为它是这些σ*场*的交点，所以它包含在每个σ场中，所以 B B⊂C .

为此，考虑区间[*a, b*], [*a, b*], (*a, b*)（根据定义，形式为(*a, b*)的区间在C中）：

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right)$$

$$[a, b] = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b + \frac{1}{n}),$$

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + \frac{1}{n}).$$

C 作为一个σ*场，*在可数交集方面是封闭的，所以它包含右边的集合。对无界区间的论证也类似。该证明是完整的。 □

练习2.7

证明形式为(*a, b*)的区间族也生成了博勒集的σ场。证明所有区间族[*a, b]*的情况也是如此。

备注2.4

由于是M一个包含所有区间的σ*场*，并且是最小的B这样的σ*场*，我们有这样的包容 B⊂M 也就是说，R中的每一个Borel集都是Lebesgue可测的。因此，问题出现了，这些σ-场是否可能是相同的。事实上

包容是适当的。在M/B中构造一个集合并不完全简单，我们在此不做尝试（但见附录）。然而，根据定理2.12（ii），给定任何$E \in$M，我们可以找到一个Borel集$B \supset E$，其形式为$B = \bigcap_n O_n$，其中（O_n）是开放集，并且使m（E）$= m$（B）。

特别是、

$$m(B \Delta E) = m(B \backslash E) = 0。$$

因此，*M*不能区分可测量集*E*和博勒集

我们已经构建了*B*。

因此，给定一个Lebesgue可测量集*E*，我们可以找到一个Borel集*B*，使得它们的对称差 *EΔB* 是一个空集。现在我们知道，$E \Delta B \in$ M B

，很明显，空集的子集也是空的，因此在 。然而，B M 我们不能断定每一个空集都是博勒集（如果 B 确实包含所有的空集，那么根据定理2.12(ii)，我们将得到=），这就指出了其中的 "不完全性"，这就解释了为什么即使我们从

在区间上定义m开始，然后将定义扩展到博勒集，我们也需要进一步扩展，以便能够准确地确定哪些集对我们来说是 "可忽略的"。另一方面，将度量m扩展到

σ场M就足够了，因为M确实包含了所有的m 个空集，所有空集的子集也都属于M。

我们表明M是R上具有这种性质的最小的σ场，我们说M是B相对于m的*完成度*，（R，M，m）是完全的（而度量空间（R，B，m）不是完全的）。更确切地说，如果对于所有的F∈F，$\mu(F)=0$，对于所有的$N \subset F$，我们有N▯ F(所以$\mu(N)=0$)，那么一个度量空间(X, F, μ)是*完全的。*

相对于一个给定的度量μ，一个σ场G的*完成*被定义为包含G的最小σ场F，如果$N \subset G$▯ G且$\mu(G)=0$，那么N∈F。

命题2.17

G的完成度的形式是{G∪N：G∈F，$N \subset F$∈F与$\mu(F)=0$}。

这使得我们可以通过设置$\bar{\mu}(G \cup N)=\mu(G)$，对$G$▯G，将措施$\mu$唯一地扩展为F上的措施μ。

定理2.18

M是B的完成。

证明

我们首先表明，M包含B中所有的空集子集：所以让$N \subset B$∈B、
为了证明N∈M，我们需要证明

$$m^*(A) \geq m^*(A \cap N) + m^*(A \cap N^c)\text{。}$$

首先注意到，$m^*(A \cap N) \leq m^*(N) \leq m^*(B)=0$。因此，仍然需要证明：$m^*(A) \geq m^*(A \cap N^c)$，但这马上就可以从$m^*$的单调性中得出。

因此，我们已经证明，$N \in$ M.由于 M 是一个完整的σ场，包含B，这意味着 M 也包含了完成 C 的B.

最后，我们表明，是N这样的最小σ*场*，即 T M⊂C：首先考虑E∈M，$m^*(E)<\infty$，并选择$B = {}_n O_n$ ∈B，如上所述，使$B \supset = E$，$m(B) = m^*(E)$。(在整个论证过程中，我们保留对B中的集合使用m的做法)。

考虑$N = B \setminus E$，它在M中，并且有$m^*(N) = 0$，因为m^*在M上是加性的。根据定理2.12(ii)，我们可以找到$L \supset N$，L∈B，$m(L) = 0$。换句话说，N是B中一个空集的子集，因此$E = B \setminus N$属于B的完成度C。对于E∈M，$m^*(E) = \infty$，对$E_n = E \cap [-n, n]$，对于每个n∈N，应用上述方法。每个$m^*(E_n)$是有限的，所以E_n所有

属于 C 因此，它们的可数联盟E也是如此。 M⊂C 所以 □
它们是相等的。

尽管有这些技术上的差异，可测集从未远离 "好 "集，而且，除了像定理2.12中观察到的那样，通过开放集从上到下近似地测量任何$E \in$ M 的度量，我们可以通过封闭子集的度量从下往上近似。

定理2.19

如果$E \in$ M 那么对于给定的$\varepsilon>0$，存在一个封闭集$F \subseteq E$，使$m(E\backslash F)<\varepsilon$。因此存在$B\subset E$ 的形式$B=_n F_n$，其中所有的F_n是封闭集，并且$m(E\backslash B)=0$。

证明

补集E^c 是可测的，根据定理2.12，我们可以找到一个包含E^c 的开集O，使得$m(O \backslash E^c) \leq \varepsilon$。但是$O \backslash E^c = O \cap E = E \backslash O^c$，而$F = O^c$ 是封闭的，包含在E中。因此这个F是我们需要的。最后一部分类似于定理2.12（二），证明留给读者。 □

练习2.8

证明以下两句话中的每一句都等同于说E⊡M：

(i) 给定$\varepsilon>0$，有一个开放的集合$O\supseteq E$，$m^* (O\backslash E) < \varepsilon$、

(ii) 给定$\varepsilon>0$，有一个封闭的集合$F\subset E$，其中$m^* (E\backslash F) <\varepsilon$。

备注2.5

上述练习中的两个陈述是一个相当大的生成的关键，将度量理论的思想与拓扑学的思想联系起来：

一个非负数可加性集合函数μ定义在 B 被称为
如果对每一个博莱尔集B来说，我们有*正规的博莱尔测量*：

$$\mu(B) = inf\{\mu(O)：O是开放的，O \supset B\}、$$
$$\mu(B) = sup\{\mu(F)：F封闭，F \subset B\}。$$

在定理2.12和2.19中，我们已经验证了Lebesgue度量的这些关系。我们将在后面考虑其他有规律的Borel度量的具体例子。

2.6 P robability

导致Lebesgue度量的思想可以被用来构建任意集合上的度量：任何携带外度量的集合Ω（即从P （Ω）$_\infty$到[0，]的单调和可数次加性的映射）可以配备一个定义在其子集的 适当σ域上的 度量μ。如备注2.1所述,F由此产生的三联体（Ω，F，μ）被称为度量空间。请注意，在Lebesgue度量的构造中，我们只使用了属性，而没有使用外度量的特定形式。

然而，就目前而言，我们将满足于简单地指出如何将勒贝斯格测量限制在R的任何勒贝斯格可测量子集B上，且m（B）>0：

给定Lebesgue σ -场 M上的Lebesgue度量m，让

$$\mathrm{m}_B = \{a \cap b : a \in \mathrm{m}\}$$

而对于$A \in \mathrm{M}_B$ 写为

$$m_B(A) = m(A)。$$

命题2.20

(B, M_B, m_B) 是一个完整的度量空间。

提示$\bigcup_{i}(A_i \cap B) = (\bigcup_i A_i) \cap B$ 和 $(A_1 \cap B) \cap B = (A_1 \setminus A_2) \cap B$。

我们最后可以准确地说明我们所说的 "从[0,1]中随机选择一个数字 "的意思：将勒贝斯格度量m限制在区间B=[0,1]，并考虑M的σ 场$_{[0,1]}$ ，即[0,1]的可测量子集。那么$m_{[0,1]}$ 是一个
在$\mathrm{M}_{[0,1]}$ ，"总质量 "为1。由于[0,1]的所有子区间具有
相同长度的数字具有相同的度量，$m_{[0,1]}$ 的 "质量 "均匀地分布在[0,1]上，因此，例如，从[0,$\frac{1}{10}$]中选择一个数字的 "概率
与从[$\frac{6}{10}$,$\frac{7}{10}$)中选择一个数字的情况相同，即$\frac{1}{10}$ 。因此，所有
数字作为所选数字的十进制扩展的第一个数字出现的可能性相同。另一方面，在这种措施下，所选数字是有理数的概率是0，就像抽到康托尔集C的一个元素的概率一样。

我们现在有了一些概率理论的基础，尽管一般的发展仍然需要将度量的概念从R扩展到抽象的集合。然而，这些构件已经在详细的

Lebesgue度量的例子的发展。其主要思想是提供一个

概率论的数学基础是使用度量的概念来提供概率的直观概念的数学模型。概率的显著特征是*独立性*的概念，我们在下面介绍。我们首先定义了一般的框架。

2.6.1 概率 空间

定义2.6

*概率空间*是一个三元组（Ω，F，P），其中Ω*是*一个任意集合，F是一个Ω的子集的σ场，而P是F上的一个度量，使得

$$P(\Omega)=1、$$

称为*概率测量*或简要的*概率*。

备注2.6

Kolmogorov在1932年给出的原始定义是上述定义的一个变体（见定理2.14）：如果（Ω，F，P）如上所述是一个概率空间，并且P*是*一个有限加性集合函数，P（$\varnothing$）=0和P（Ω）=1
这样，只要F中的(B_n)减少到$\varnothing$，$P(B_n)\searrow 0$。

例2.2

我们马上就可以看到，限制在[0，1]的Lebesgue度量是一种概率的确定。更一般地说：假设我们得到一个任意的Lebesgue可测量集$\Omega\subset\mathbb{R}$，$m(\Omega)>0$。那么P=c m_Ω，其中$c=\frac{1}{m(\Omega)}$，$m=m_\Omega$ d-。
注意到Lebesgue测量对Ω的可测量子集的限制，提供了
是Ω上的概率度量，因为P是完整的，P（Ω）=1。
例如，如果$\Omega=[a,b]$，我们得到$c=\frac{1}{b-a}$，而P成为"统一的分布"在[a,b]。然而，我们也可以使用不太熟悉的集合作为我们的基础空间；例如，$\Omega=[a,b]\cap(\mathbb{R}\setminus\mathbb{Q})$，$c=\frac{1}{b-a}$给出相同的分布在[$a,b$]的无理数上。

2.6.2 事件：调节和 独立性

事件 "这个词是用来表示某些事情正在发生。在概率学中，一个典型的事件是从一个集合中抽取元素，然后事件关注的是属于一个特定子集的结果。因此，如上所述，如果

$\Omega=[0,1]$，我们可能对从[0，1]中随机抽取的一个数字属于某个A$\subset$[0，1]的事实感兴趣。我们想估计这种情况发生的概率，在数学设置中，这就是数字P（A），在这里

$m_{[0,1]}(A)$。因此，很自然地要求A应该属于$_{[0,1]}$，因为这些是我们可以测量的集合。通过对语言的轻微滥用，概率论者

倾向于将实际的 "事件 "与事件中的集合A相提并论。接下来的定义只是证实了这种语言的滥用。

定义2.7

给定一个概率空间（Ω，J，P），我们说J的元素是*事件*。

假设从[0，1]中抽出一个数字，但还没有被揭开。我们想赌它在[0,$\frac{1}{4}$]中，我们得到一个提示，它肯定属于[0,$\frac{1}{2}$]。很明显，考虑到这个 "内部信息"，可能的情况是

现在成功的能力是$\frac{1}{2}$，而不是$\frac{1}{4}$。这促使了以下的一般定义。

定义2.8

假设$P(B)>0$，那么数

$$P(A/B)=\frac{p(a\cap b)}{p(b)}$$

被称为*给定B的A的条件概率*。

命题2.21

映射$A'\to P(A/B)$在σ场J上是可数加性的$_B$。

提示 利用$A'\to P$（$A\cap B$）在J上是可数加性的事实。

条件概率的一个经典应用是总概率公式，它可以通过给定一些不相干的假设的条件概率来计算一个事件的概率：

练习2.9

证明如果H_i是成对不相交的事件，使$\bigcup_{i=1}^{\infty} H_i = \Omega$、$P(H_i) \neq 0$，则

$$P(A) = \sum_{i=1}^{\infty} P(A|H_i)\,P(H_i)。$$

很自然地说，如果*B的*事实是*独立于B的，那么*事件*A*就是*独立于B的*。的发生对*A*的机会没有影响，即

$$p（a/b）=p（a）。$$

根据$P(A/B)$的定义，这立即意味着关系

$$p(a\cap b)=p(a)-p(b)$$

这通常被看作是独立的定义。这种做法的好处是，我们可以放弃假设$P（B）>0$。

定义2.9

事件*A*、*B*是*独立的*，如果

$$p（a\cap b）=p（a）-p（b）。$$

练习2.10

假设*A*和*B*是独立事件。证明A^c和*B*也是独立的。

练习表明，如果*A*和*B*是独立的事件，那么它们产生的σ-*场的*所有元素都是相互独立的，因为这些σ-场只是集合$F_A=\{\emptyset, A, A^c, \Omega\}$和$F_B=\{\emptyset, B, B^c, \Omega\}$分别。这使我们得出定义的自然延伸：如果对于任何选择的集合$A_1\in F_1$和$A_2\in F_2$，我们有$P(A_1\cap A_2)=P(A_1)P(A_2)$，则两个σ场$F_1$和$F_2$是*独立的*。

然而，将这些定义扩展到三个或更多的事件（或几个σ*场*）需要注意一下，正如下面的简单例子所示：

例2.3

让$\Omega=[0,1]$，$A=[0,\frac{1}{4}]$如前；那么A独立于$B=[\frac{1}{8},\frac{5}{8}]$和$c=[\frac{1}{8},\frac{3}{8}]\cup[\frac{3}{4},1]$。此外，*B*和*C*是独立的。然而、

$$p（a\cap b\cap c）/= p（a）-p（b）-p（c）。$$

因此，给定三个事件，三个可能对中的每一个的成对独立性并不足以将 "独立性

"扩展到所有三个事件。

另一方面，$A=[0,\frac{1}{16}]$，$B=C=[0,\frac{1}{4}]\cup[\frac{1}{4},\frac{11}{16}]$，（或另一种方式与$C=[0,\frac{1}{16}]\cup[\frac{9}{16},1]$）相对应。

$$p（a\cap b\cap c）=p（a）-p（b）-p（c）。\tag{2.21}$$

但没有一对是独立的事件。

这进一步证实，如果我们想扩展上述定义，我们需要提出更多的要求--成对的独立性是不够的，(2.21)也是不够的；因此我们需要要求这两个条件同时得到满足。将此扩展到n个事件，就可以得出：

定义2.10

事件$A_1, \ldots, A_n$是*独立的*，如果对于所有$k \leq n$，对于每个选择的k事件，它们相交的概率是概率的乘积。

同样，对于σ-场也有一个强大的对应物（可以扩展到序列，甚至是任意的族）：

定义2.11

定义在给定概率空间（Ω，$\mathcal{F}$，P）上的σ场$\mathcal{F}_1$，$\mathcal{F}_2$，…，$\mathcal{F}_n$是*独立的*，如果对于从{1，2，…，n}中选择的所有不同指数i_1，i_2，…，i_k和所有选择的集合$F_{i_n} \in \mathcal{F}_{i_n}$我们有

$$P(F_{i_1} \cap F_{i_2} \cap \ldots \cap F_{i_k}) = P(F_{i_1}) \cdot P(F_{i_2}) P(F_{i_k})。$$

独立性的问题将在随后的章节中重新讨论，在这些章节中我们将开发一些更多的工具来计算概率

2.6.3 在数学上的应用 金融

如序言所述，我们将简要地探讨如何将每一章中提出的观点应用于迅速发展的数学金融领域。这并不是作为这一主题的介绍，但希望它能充分展示一个一致的数学表述如何有助于澄清许多学科的核心思想。不熟悉数理金融的读者应该查阅诸如[4]、[5]、[7]等文本，以了解该主题的定义和主要思想的讨论。

金融学中的概率建模的核心是分析交易资产价值的演变模型，如*股票*

或*债券*，并寻求确定其未来行为的趋势。大部分现代理论涉及评估*衍生证券*，如*期权*，其价值由一些基础证券，如股票的（随机）未来价值决定。

我们在一个经典的股票价格模型，即*二叉树*上说明了上述概率思想。这个模型是基于价格可能发生变化的有限多个时间时刻，而且这些变化的性质非常简单。假设步数为N，用$S(k)$表示第k步的价格，$0 \leq k \leq N$。在每一步，股票价格以如下方式变化：某一步的价格是前一步的价格乘以U，其中有

概率为p或D，概率为$q=1-p$，其中$0<D<U$。因此，最终价格取决于序列

$\omega=(\omega_1, \omega_2, \ldots, \omega_N)$，其中$\omega_i=1$

表示应用因子U或$\omega_i=0$，表示应用因子D。这样的序列被称为*路径*，我们认为ω包括所有可能的路径。换句话说、

$$S(k)=S(0)\times\eta(1)\times\cdots\times\eta(k)、$$

其中

$$\eta(k)=\begin{cases} U的概率为p，D的\\ 概率为q。\end{cases}$$

练习2.11

假设$N=5$，$U=1.2$，$D=0.9$，$S(0)=500$。求所有路径的数量。有多少条路径会导致价格$S(5)=524.88$？如果单步上升的概率是0.5，那么$S(5)>900$的概率是多少？

一般来说，路径的总数显然是2^N，在步骤k有

$k+1$个可能的价格。

我们通过给Ω配备西格玛场2^Ω来构建一个概率空间。

的所有子集，以及定义在单元素集上的概率$P(\{\omega\})=p^k q^{N-k}$，其中$k=\sum_{i=1}^N \omega_i$。

随着时间的推移，我们收集关于股票价格的信息，或者说，相当于收集关于路径的信息。这意味着，在观察了一些价格之后，未来可能的发展范围受到了限制。我们的信息随着时间的推移而增加，这个想法可以通过以下的*σ场*系列来体现。

固定$m<n$，定义一个σ场$F_m=\{A:\omega,\omega'\in A \Rightarrow \omega_1=\omega_1', \omega_2=\omega_2', \ldots, \omega_m=\omega_m'\}$。因此，在这个西格玛场中，所有来自特定集合$A$的路径都有相同的初始段，而其余的坐标是任意的。注意

那

$F_0 = \{\Omega, \varnothing\}$、

$F_1 = \{A_1, A^c_1, \Omega, \varnothing\}$，其中$A_1 = \{\omega : \omega_1 = 1\}$，即$S(1) = S(0)U$，和$A^c_1 = \{\omega : \omega_1 = 0\}$即$S(1) = S(0)D$。

练习2.12

证明$\mathcal{F}_m$有2^{2m}个元素。

练习2.13

证明序列$\mathcal{F}_m$是增加的。

这个序列是一个*过滤的*例子（其识别特征是西格玛字段应该包含在其中并形成一个递增链），这个概念我们将在后面重新审视。

股票价格的连续选择与抛掷硬币密切相关。直觉告诉我们，后者是独立的。这可以通过引入另一个σ*场来*正式看到，描述在某一步我们有一个特定的结果。假设ω是这样的：$\omega_k = 1$。那么我们可以确定具有此属性的所有路径的集合 $A_k = \{\omega : \omega_k = 1\}$，并扩展为一个σ场：$\mathcal{G}_k = \{A_k, A^c_k, \Omega, \varnothing\}$。事实上，$A^c_k = \{\omega : \omega_k = 0\}$。

练习2.14

证明$\mathcal{G}_m$和$\mathcal{G}_k$是独立的，如果$m \neq k$。

2.7 命题的证明

证明（命题2.3的证明）

如果区间I_n覆盖B，那么它们也覆盖A：$A \subset B \subset \bigcup_n I_n$，因此$Z_B \subset Z_A$。较大集合的下限不能大于较小集合的下限（琐碎的说明：$\inf\{0,1,2\} < \inf\{1,2\}$，$\inf\{0,1,2\} = \inf\{0,2\}$）

因此，该结果。 □

证明（命题2.6的证明）

如果一个区间系统I_n覆盖A，那么区间$I_n + t$覆盖$A + t$。反之，如果J_n覆盖$A + t$，那么$J_n - t$覆盖A。此外，当我们把每个区间转移一个数字时，区间家族的总长度不会改变。所以我们在A和$A+t$的区间覆盖之间有一个一

一对应的关系，这个对应关系保留了覆盖的总长度。这意味着 Z_A 和 Z_{A+t} 的集合是相同的，所以它们的infima是相等的。 □

证明（命题2.9的证明）

根据德摩根定律

$$\bigcap_{k=1}^{\infty} E_k = \left(\bigcup_{k=1}^{\infty} E_k^c\right)^c$$

根据定理2.8(ii)，所有E_k^c都在M中，因此根据(iii)，也可以说是在M中。联盟$\bigcup_{k=1}^{\infty} E_k^c$。最后，根据(ii)，这个联盟的补数又是在M，所以交集$\bigcap_{k=1}^{\infty} E_k$是在M中。 □

证明（命题2.10的证明）

(i) 命题2.3告诉我们，外部度量是单调的，但由于m只是m^*对M的限制，那么m也是如此：$A\subset B$意味着$m(A) = m^*(A) \le m^*(B) = m(B)$。

(ii) 我们把B写成一个不相交的联盟$B = A \cup (B\setminus A)$，然后根据以下的可加性

我们有$m(B) = m(A) + m(B\setminus A)$。减去$m(A)$（这里重要的是$m(A)$是有限的），我们得到结果。

(iii) m的平移不变性立即从外部措施的平移不变性中得出，其方式与上述（i）相同。 □

证明（命题2.11的证明）

集合$A\Delta B$是空的，因此它的子集$A\setminus B$和$B\setminus A$也是空的。因此这些集合是可测的，$A\cap B = A\setminus$（$A\setminus B$），因此$B =$（$A\cap B$）∪（$B\setminus A$）∈M。但$m(B\setminus A) = 0 = m(A\setminus B)$，所以$m(B) = m(A\cap B) = m(A\cap B) + m(A\setminus B) = m((A\cap B)\cup(A\setminus B)) = m(A)$。 □

证明（命题2.17的证明）

族G = {$G\cup N$: $G\in$F, $N\subset F\in$F与$\mu(F)=0$}包含X，$X\in$F的集合。如果$G_i\cup N_i \in$G，$N_i\subset F_i$，$\mu(F_i) = 0$，则$\bigcup_i (G_i\cup N_i) = \bigcup_i G_i \cup \bigcup_i N_i$是在G中，因为右边的第一个集合在F中，第二个集合是一个空集的子集$\bigcup_i F_i \in$F。如果$G\cup N\in$G，$N\subset F$、则$(G\cup N)^c = (G\cup F)^c \cup (F\setminus N \cap G^c)$.其中也在G中，因此G是一个σ场。考虑一下包含F和所有空集子集的任何其他σ场H。由于H对于联合体来说是封闭的，它包含G，因此G是具有这一性质的最小的σ场。 □

证明（命题2.20的证明）

从定义和提示中可以立即看出，$\mathcal{M}_B$ 是一个σ场。为了看到m_B 是一个度量，我们检查可数可加性：在$\mathcal{M}_i = A_i \cap B$成对不相交的情况下，我们有

$$m_B\left(\bigcup_i C_i\right) = m\left(\bigcup_i (A_i \cap B)\right) = \sum_i m(A_i \cap B) = \sum_i m(C_i) = \sum_i m_B(C_i)。$$

因此$(B, \mathcal{M}_B, m_B)$是一个度量空间。它是完整的，因为根据定义，B中包含的空集的子集是m_B-可度量的。 □

证明（命题2.21的证明）

假设A_n 是可测量的，并且是成对不相交的。根据条件概率的定义

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} A_n \,|\, b\right) &= \frac{1}{P(B)} p\left(\left(\bigcup_{n=1}^{\infty} a_n\right) \cap B\right) \\
&= \frac{1}{P(B)} P\left(\bigcup_{n=1}^{\infty} (A_n \cap \mathrm{B})\right) \\
&= \frac{1}{P(B)} \sum_{n=1}^{\infty} P(A_n \cap B) \\
&= \sum_{n=1}^{\infty} P(A_n / B)
\end{aligned}$$

因为$A_n \cap B$也是成对不相交的，而且P是可数加的。 □

3

可测量的功能

3.1 扩展的真实 行

R的长度在上面是无界的，也就是 "无限"。为了处理这个问题，我们为无限以及有限度量的集合定义了Lebesgue度量。为了全面地处理这些集合之间的函数，方便地允许函数

它们的值是无限的：我们认为它们的范围是 "扩展实线 "R = [*,] 的* （一部分），通过添加 "无限大的点 "得到的 $-\infty$ 和+到∞
R.正如在第2.2节中已经观察到的那样，在这个集合中的算术需要稍微注意一下：
我们假设$a+\infty = \infty$ 对于所有实数a，$a \times \infty = \infty$ 对于$a > 0$，$a \times \infty = -\infty$ 对于 $a < 0$, $\infty \times \infty = \infty$ 和 $0 \times \infty = 0$，以及类似的定义。这些在直觉上都是' 显而易见'的（可能除了$0 \times \infty$），而且（与度量衡一样）我们避免形成$+\infty + (-\infty)$ 形式的'和'。在这些假设下，"算术工作与以前一样"。

3.2 Lebesgue-可测量的 函数

现在我们可以自由地定义*f*，只限于 "空集"：一旦我们证明两个函数*f*和*g*在R \ *E*上是相等的，其中*E*是某个空集，那么就所有实际目的而言，*f* = *g*。为了正式说明这一点，我们说*f*有一个属性(*P*)

如果*f*在其域的所有点上都具有这一属性，则*几乎无处不在*（a.e），但

可能是在某个空集上。

例如，函数

$$f(x) = \begin{cases} 1 & 对于 x \neq 0 \\ 0 & 对于 x = 0 \end{cases}$$

几乎到处都是连续的，因为它在R 0上是连续的，而例外集0是空的。(注：概率论者倾向于说 "几乎肯定"（a.s.）而不是 "几乎到处"（a.e.），我们将在下文中跟随他们的思路。

专门讨论概率的章节）。

下一个定义将介绍Lebesgue可测量的函数类。对......施加的条件将是必要的（尽管不是充分的），以赋予（Lebesgue）积分fm以意义：R R将是赋予（Lebesgue）积分$f\,dm$意义的必要条件（尽管不是充分条件）。让我们首先给出

一些动机。

积分总是涉及到近似的过程。在黎曼积分中，我们将区间$I=[a, b]$分割成小块I_n - 再次分割成区间。最简单的方法是将区间分成N个相等的部分。然后，我们通过将小区间的长度乘以一定的数字c_n（与此有关）来构建近似和。

有关函数的值；例如，$c_n = \inf_{I_n} f$，$c_n = \sup_{I_n} f$、或$c_n = f(x)$ for some $x \in I_n$）：

$$\sum_{n=1}^{N} c_n\, l(I_n)。$$

对于大的n，这个和接近于黎曼积分$\int_a^b f(x)\,dx$（给定一些f的正则性）。

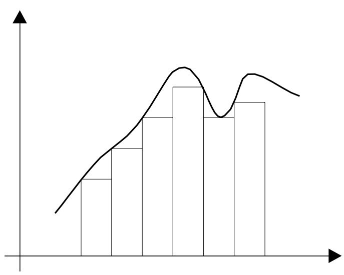
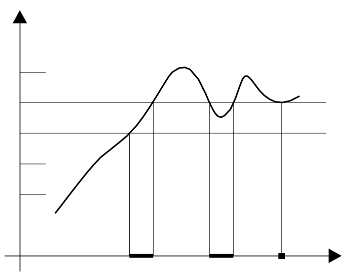

图3.1 黎曼与勒贝斯格的关系

处理Lebesgue积分的方法类似，但有一个关键的区别。我们不是将积分域分割成小部分，而是将其分解为

函数的范围。同样，一个简单的方法是引入等长的短区间J_n。为了建立近似和，我们首先通过f取J_n的反图像，即$f^{-1}(J_n)$。这些可能是复杂的集合，不一定是
间隔。在这里，之前开发的测量理论开始发挥作用。
我们能够测量集合，只要它们是可测量的，即它们是在……。鉴于此，M我们计算出

$$\sum_{n=1}^{N} c_n \, m(f^{-1}(J_n))$$

其中$c\ J_n \in n$或$c_n = \inf J_n$，例如。

下面的定义保证了这个程序是有意义的（尽管在$N\to\infty$时可能需要一些额外的小心来得出一个有限的数字）。

定义3.1

假设E是一个可测量的集合。我们说一个函数f：$E\to\mathbb{R}$是（*Lebesgue-*）*可测的*，如果对于任何区间$I\subseteq\mathbb{R}$

$$f^{-1}(I) = \{x\in\mathbb{R} : f(x) \in I\} \in \mathcal{M}。$$

在下文中，*可测量*（没有限定）一词将指Lebesgue-可测量的函数。

如果所有的集合$f^{-1}(I)\in\mathcal{B}$，即如果它们是博勒集，我们称f为*博勒-可测量的*，或者简单地说是一个*Borel*函数。

其基本理念是各种数学概念所共有的：一个好集的反像就是*好的*。请记住连续函数，例如，任何开放集的反像都被要求是开放的。*nice*这个词的实际含义取决于数学的特定分支。在上面的定义中，请注意，由于$\mathcal{B}\subset\mathcal{M}$，每个Borel函数都是（Lebesgue-）可测量的。

备注3.1

这个术语有些不幸。可测量的 "对象应该被测量（如可测量的集合）。然而，可测量的函数将被整合。这种混淆源于这样一个事实，即可能最适合在这里使用的*可积分*一词带有更多的限制性含义，正如我们将在后面看到

的。这个术语被广泛接受，我们不打算在这里和整个世界对抗。

我们给出了一些等价的公式：

定理3.1

以下条件是等同的

(a) f是可以测量的、

(b) 对于所有的a，f^{-1}（（a，∞））是可测量的、

(c) 对于所有的a，f^{-1}（$[a,\infty]$）是可测量的、

(d) 对于所有的a，f^{-1}（（$-\infty$，a））是可测量的、

(e) 对于所有a，f^{-1} $((-\infty, a])$是可测量的。

证明

当然，（a）意味着其他任何条件。我们表明，(b)意味着

(a).其他含义的证明也类似，留作练习（你应该尝试）。

我们必须证明，对于任何区间I，f^{-1}（I）$\in$M。根据（b）我们有，对于特殊情况$I=$（a，∞）。假设$I=$（$-\infty$，a]。那么

$$f^{-1}((-\infty, a]) = f^{-1}(\mathbb{R} \setminus (a, \infty)) = E \setminus f^{-1}((a, \infty)) \in \mathcal{M} \qquad (3.1)$$

因为E和f^{-1}（（a、∞)都在 M (我们使用之前建立的封闭属性 M)。接下来

$$f^{-1}((-\infty, b)) = f^{-1}\Big(\bigcup_{n=1}^{\infty}(-\infty, b - \frac{1}{n}]\Big)$$

$$= \bigcup_{n=1}^{\infty} f^{-1}\big((-\infty, b - \frac{1}{n}]\big)。$$

根据（3.1），$f^{-1}((-\infty, b-\frac{1}{n}]) \in$ M 而对于可数联盟也是如此。
由此我们可以很容易地推断出

$$f^{-1}([b, \infty)) \in \mathcal{M}。$$

现在让$I=$（a，b），并且

$$f^{-1}((a, b)) = f^{-1}((-\infty, b) \cap (a, \infty))$$

$$= f^{-1}((-\infty, b)) \cap f^{-1}((a, \infty))$$

是在M 的两个元素的交点。 M.通过同样的推理，包 M

含

$$f^{-1}([a, b]) = f^{-1}((-\infty, b] \cap [a, \infty])$$

$$= f^{-1}((-\infty, b]) \cap f^{-1}([a, \infty))$$

和半开区间的处理方式类似。 □

3.3 E xamples

下面的简单结果表明，"实践中 "遇到的大多数函数都是可测量的。

(i) 恒定函数是可测量的。让$f(x) \equiv c$. 那么

$$f^{-1}((a,\infty)) = \begin{cases} \mathbb{R} & \text{如果} a < c \\ \varnothing & \text{否则} \end{cases}$$

而在这两种情况下，我们都有可测量的集合。

(ii) 连续函数是可测的。因为我们注意到（a，∞）是一个开放集，所以 f^{-1}（（a，∞））。正如我们所知，所有开放集都是可测的。

(iii) 定义一个集合A的指标函数为

$$1_A(x) = \begin{cases} 1 & \text{如果} x \in A \\ & \text{否则为0。} \end{cases}$$

那么

$$A \in \mathcal{M} \quad \Leftrightarrow \quad 1_A \text{ 是可计量的}$$

因为

$$1_A^{-1}((a,\infty)) = \begin{cases} \mathbb{R} & \text{如果} a<0 \\ A & \text{如果} 0 \le a < 1 \\ \varnothing & \text{如果} a \ge 1\text{。} \end{cases}$$

练习3.1

证明每个单调函数都是可测量的。

练习3.2

证明如果f是一个可测量的函数，那么水平集$\{x : f(x) = a\}$。对每一个$a \in \bar{\mathbb{R}}$来说是可测量的。

提示 当a是无限的时候，不要忘记案例!

备注3.2

在附录中，假设 "选择公理 "有效，我们表明R的一些子集不能成为Lebesgue

可测量的，并且有

是Lebesgue可测量的集合，它们不是Borel集合。因此，如果(R)表示R的所有子集的σ场，那么以下的内涵是严格的

$$B\subset M\subset P(R)。$$

通过考虑这些集合的指标函数，这些（相当深奥的）事实可以被用来构建非可测函数和非伯乐函数的可测函数的例子。虽然为了理解为什么要引入这些不同的概念，意识到这些区别是很重要的，但这些例子不会在我们所考虑的理论应用中出现。

3.4 P roperties

可测量函数的类别非常丰富，正如以下结果所示。

定理3.2

定义在E上的实值可测函数的集合是一个矢量空间，在乘法下是封闭的。 是一个向量空间，在乘法条件下是封闭的，也就是说，如果f和g是可测量的函数，那么$f+g$和fg也是可测量的（特别是，如果g是一个常数函数$g\equiv c$，cf对所有实数c而言是可测量的）。

证明

首先考虑$f+g$。我们的目标是表明，对于每个$a\in R$、

$$B=(f+g)^{-1}(-\infty,a)=\{t:f(t)+g(t)<a\}\in M.$$

假设所有的有理数都排列成一个序列$\{q_n\}$。现在

$$B=\bigcup_{n=1}^{\infty}\{t:f(t)<q_n,\ g(t)<a-q_n\}$$

- 我们将直线$x+y=a$下方的半平面分解为一个可数的无界 "盒子 "的联盟：$\{(x,y):x<q_n,\ y<a-q_n\}$。很明显

$$\{t:f(t)<q_n,\ g(t)<a-q_n\}=\{t:f(t)<q_n\}\cap\{t:g(t)<a-q_n\}=\{t:f(t)<q_n\}$$

是可测量的，是可测量集的交集。因此，$B\in$M是M的元素的一个可数的联盟。

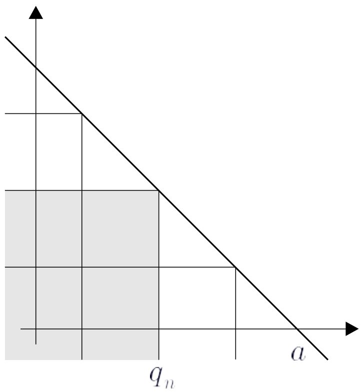

图3.2 箱子

为了处理fg，我们采用了一种略微间接的方法，以便保持 "一维"：首先注意到，如果g是可测的，那么$-g$也是可测的。由于$fg = \frac{1}{4}\{(f+g)^2 - (f-g)^2\}$，因此只需证明一个可测函数的平方是可测的即可。所以取一个可测量的$h : E \to \mathbb{R}$并考虑$\{x \in E : h^2(x) > a\}$。对于$a<0$，这个集合是$E \in \mathcal{M}$，而对于$a \geq 0$

$$\{x : h^2(x) > a\} = \{x : h(x) > \sqrt{a}\} \cup \{x : h(x) < -\sqrt{a}\}.$$

右边的两个集合都是可测的，因此我们已经证明h^2是可测的。分别用$h=f+g$和$h=f-g$来应用，得出fg是可测的结论。由此可见，对于常数c，cf是可测的，因此，实值可测函数类在加法下形成一个向量空间。

□

备注3.3

该定理的一个优雅的证明是基于以下定理，该定理在后面也会有用。它的证明利用了一个简单的拓扑学事实，即R中的每一个开放集[2]，分解成一个可数的矩形联盟，准确地说，是
与R中的开放集和区间进行类比。

定理3.3

假设$F : \mathbb{R}\ \mathbb{R}$是一个连续函数。如果$f$和$g$是可测的，那么$h(x)=F(f(x), g(x))$也是可测的。

现在只要把$F(u, v)=u+v$，$F(u, v)=uv$，就可以得到定理3.2的第二个证明。

证明(该定理的证明)

对于任何实数a

$$\{x : h(x) > a\} = \{x : (f(x), g(x)) \in G\}_a$$

其中$G_a = \{(u, v)：F(u, v) > a\} = F^{-1}((a, \infty))$。暂时假设我们很幸运，$G_a$是一个矩形：$G_a = (a_1, b_1) \times (c_1, d_1)$。

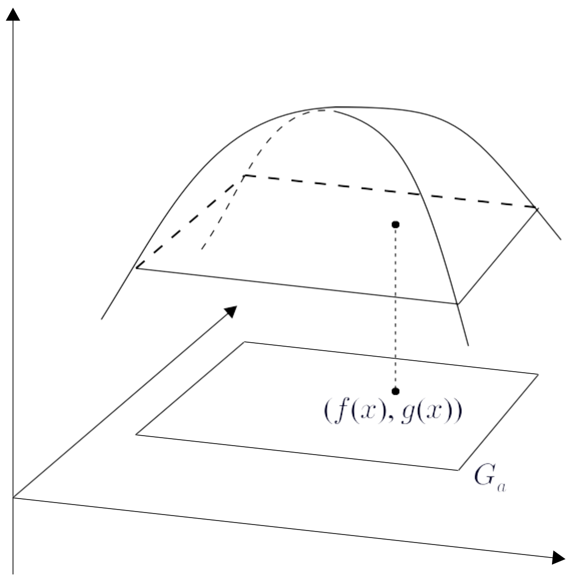

图3.3 集合G_a

从图3.3可以看出，

$$\{x : h(x) > a\} = \{x : f(x) \in (a_1, b_1) \text{ and } g(x) \in (c_1, d_1)\}$$
$$= \{x : f(x) \in (a_1, b_1)\} \cap \{x : g(x) \in (c_1, d_1)\}。$$

一般来说，我们必须将集合G_a分解为矩形的联盟。集合G_a是R R的一个开放子集，因为F是连续的。因此，它可以被写成

$$G_a = \bigcup^{\infty} R_n$$

$n=1$

其中R_n是开放的矩形 $R_n = (a_n, b_n) \times (c_n, d_n)$。所以

$$\{x : h(x) > a\} = \bigcup_{n=1}^{\infty} \{x : f(x) \in (a_n, b_n)\} \cap \{x : g(x) \in (c_n, d_n)\}$$

由于M的稳定特性，它是可测量的。 □

定理3.2的一个简单应用是考虑积$f \cdot 1_A$。如果f是一个可测量的函数，A是一个可测量的集合，那么$f\,1_A$是可测量的。这个函数在A上只是f，在A之外是0。将此应用于集合

$A = \{x \in E : f(x) > 0\}$我们看到，一个可测量的函数的正部分$f^+$是可测量的：我们有

$$f^+(x) = \begin{cases} f(x) & \text{如果} f(x) > 0 \\ 0 & \text{如果} f(x) \le 0。\end{cases}$$

同样，f的负数部分f^-是可测量的，因为

$$f^-(x) = \begin{cases} 0 & \text{如果} f(x) > 0 \\ -f(x), & \text{如果} f(x) \le 0。\end{cases}$$

命题3.4

让E是R的一个可测量的子集。

(i) f:当且仅当f^+和f^-都是可测量的。

(ii) 如果f是可测量的，那么$|f|$也是可测量的；但反之则是错误的。

提示第(ii)部分要求存在不可测量的集合（如附录中证明的），而不是它们的特定形式。

练习3.3

证明如果f是可测量的，那么f的截断：

$$f^a(x) = \begin{cases} f(x) & \text{如果} f(x) > a \\ a & \text{如果} f(x) \le a，\text{则} f(x) \le a \end{cases}$$

也是可以测量的。

练习3.4

找到一个不可测量的f，使f^2是可测量的。

通往极限并不破坏可测性--所有需要的工作在我们建立M的稳定性属性时就已经完成了!

定理3.5

如果$\{f_n\}$是定义在R中的集合E上的可测函数序列，那么以下也是可测函数：

$$\max_{n\le k} f_n、\quad \min_{n\le k} f_n、\quad \sup_{n\in N} f_n、\quad \inf_{n\in N} f_n、\quad \limsup_{n\to\infty} f_n、\quad \liminf_{n\to\infty} f_n。$$

证明

只要注意到以下是可测量的集合，就足够了：

$$\{x : (\max_{n\le k} f_n)(x) > a\} = \bigcup_{n=1}^{k} \{x : f_n(x) > a\}、$$

$$\{x : (\min_{n\le k} f_n)(x) > a\} = \bigcap_{n=1}^{k} \{x : f_n(x) > a\}、$$

$$\{x : (\sup_{n\ge k} f_n)(x) > a\} = \bigcup_{n=k}^{\infty} \{x : f_n(x) > a\}、$$

$$\{x : (\inf_{n\ge k} f_n)(x) \ge a\} = \bigcap_{n=k}^{\infty} \{x : f_n(x) \ge a\}.$$

对于上限，根据定义

$$\limsup_{n\to\infty} f_n = \inf_{n\ge 1}\{\sup_{m\ge n} f_m\}$$

和上述关系表明，$h_n = \sup_{m\ge n} f_m$ 是可测的，因此 $\inf_{n\ge 1} h_n(x)$ 是可测的。下限也是这样做的。 □

推论3.6

如果一个可测量函数的序列f_n （点状）收敛，那么极限就是一个可测量函数。

证明

这是直接的，因为$\lim_{n\to\infty} f_n = \limsup_{n\to\infty} f_n$，这是可以测量的。 □

备注3.4

请注意，定理3.2和3.5对伯乐函数有对应的规定，也就是说，用伯乐函数取代 "可测量 "后，它们仍然有效。

当我们考虑空集的作用时，事情就稍微复杂了一些。一方面，在空集上改变一个函数不能破坏其可测性，也就是说，任何可测函数在空集上被改变后仍然是可测的。然而，由于并非所有的空集都是Borel集，我们不能对Borel集得出类似的结论，因此下面的结果没有自然的 "Borel "对应物。

定理3.7

如果$f : E \to \mathbb{R}$是可测量的，$E \in \mathcal{M}$，$g : E \to \mathbb{R}$是任意的，并且集合$\{x : f(x) = g(x)\}$为空，那么g是可测的。

证明

考虑差值$d(x)=g(x)-f(x)$。除了在一个空集上，它是零，所以

$$\{x : d(x) > a\} = \begin{cases} \text{一个空集，如果 } a \geq 0 \\ \text{一个完整的集合 如果 } a<0 \end{cases}$$

其中满集是空集的补集。空集和全集都是可测的，因此d是一个可测的函数。因此，$g=f+$d是可测的。 □

推论3.8

如果(f_n)是一个可测量的函数序列，并且$f_n(x)\ f(x)$几乎每-在E中的x，那么f是可测量的。

证明

设A为空集，使$f_n(x)$对所有$x \in E \setminus A$都收敛。那么$1_{A^c} f_n$到处都收敛于$g = 1_{A^c} f$，因此是可测的。但$f=g$几乎无处不在，所以f也是可测的。 □

练习3.5

让f_n是一个可测量函数序列。证明集合$E=$

$\{x : f_n(x)$ 收敛$\}$是可衡量的。

由于我们能够在空集上随意调整一个函数*f*而不影响其可测性属性，下面的定义是一种有用的手段，通过确定其 "空集之外 "的界限，将注意力集中在对积分理论 "真正重要 "的*f*的值上：

定义3.2

假设$f : E \to \bar{\mathbb{R}}$是可测的。*基本最高值*ess sup *f*被定义为*inf*{z : $f \le z$ a.e.}，*基本最低值*ess inf *f*是*sup*{z : $f \ge z$ a.e.}。

请注意，ess sup *f*可以是+∞。如果ess sup $f = -\infty$，那么$f = -\infty$ a.e.，因为根据ess sup的定义，$f \le -n$ a.e. for all $n \ge 1$。现在，如果ess sup *f*是有限的，并且$A = \{x : \text{ess sup} f < f(x)\}$，对于$n \ge 1$，定义$A_n$：

$$A_n = \{x : \text{ess sup} f < f(x) - \frac{1}{n}\}。$$

这些都是空集，因此，$A = \bigcup_n A_n$也是空集。，因此我们已经验证了：

$$f \le \text{ess sup} f \text{ a.e.}$$

现在可以直接证明以下内容。

命题3.9

如果*f*，*g*是可测量的函数，那么

$$\text{ess sup}(f + g) \le \text{ess sup} f + \text{ess sup}\, g。$$

练习3.6

证明对于可测量的*f*，ess sup *f* ≤ sup *f*。证明当*f*是连续的时候，这些量值是重合的。

3.5 P robability

3.5.1 随机 变量

在概率空间的特殊情况下，我们使用*随机变量*这一短语来指可测量的函

数。也就是说，如果（Ω，F，P）是一个概率空间，那么X：$\Omega\to R$是一个随机变量，如果对于所有$a\in R$，集合X^{-1}（$[a,\infty]$）是在F中：

$$\{\omega\in\Omega: X(\omega)\geq a\}\in F。$$

在ΩR 是一个可测量的集合，并且$\mathcal{F}$ = 是Ω的Borel子集的σ场的情况下，随机变量只是Borel函数$\mathbb{R} \to \mathbb{R}$。

在应用概率中，集合Ω代表随机实验的结果，可以通过各种测量方法来观察。这些测量方法将数字分配给结果，因此我们以一种自然的方式得出了随机变量的概念。所施加的条件保证了以下问题的意义：随机变量的值在给定范围内的概率是多少？

3.5.2 由随机 变量产生的西格玛场

如前所述，我们遇到的随机变量实际上将是博莱尔可测量的函数。随机变量X的值不会导致我们进入非博雷尔集合；事实上，它们很可能导致我们讨论比博雷尔σ场的复杂性中已有的集合之间更粗略的区分。因此，我们应该准备好考虑包含在$\mathcal{F}$中的不同σ场。准确地说：

集合的家族

$$X^{-1}(\mathcal{B}) = \{S \subset \mathcal{F} : S = X^{-1}(B) \text{ for some } B \in \mathcal{B}\}$$

是一个σ-场。如果X是一个随机变量，$X^{-1}(\mathcal{B}) \subset \mathcal{F}$ 但它可能是一个更小的子集，这取决于X的复杂程度。我们用$\mathcal{F}_X$来表示这个σ场，并称它为*由X生成的σ场*。

最简单的可能情况是X是常数，$X \equiv a$。$X^{-1}(B)$是根据$a \in B$与否，要么是Ω，要么是$\varnothing$，生成的σ场是微不足道的：$\mathcal{F} = \{\varnothing, \Omega\}$。

如果X取两个值$a \neq b$，那么$\mathcal{F}_X$包含四个元素：$\mathcal{F}_X = \{\varnothing, \Omega, X^{-1}(\{a\}), X^{-1}(\{b\})\}$.如果$X$取有限多的值，$\mathcal{F}_X$是有限的。如果$X$取值无数，则$\mathcal{F}_X$是不可数的（它可以与可数集的所有子集的σ场相鉴别）。我们可以看到$\mathcal{F}_X$与X的复杂程度一起增长。

练习3.7

证明$\mathcal{F}_X$是包含反像的最小的σ-场$X^{-1}(B)$的所有Borel集合B。

练习3.8

集合族$\{X(A)$：$A\in F\}$是一个σ-场吗？

$_X$ 的概念有如下解释。测量值X的值是我们所能观察到的。从这些值中，我们推断出关于随机实验的复杂程度的一些信息，也就是Ω和F 的大小$_X$、而我们可以通过统计方法来估计$_X$ 中各组的概率。
产生的σ场代表随机变量产生的信息量。例如，假设抛出一个骰子，根据显示的数字是奇数还是偶数，只报告0和1。我们永远不会把这个实验与掷硬币区分开来。测量所提供的信息不足以探索实验的复杂性（有六个可能的结果，这里归为两组）。

3.5.3 概率 分布

对于任何随机变量X，我们可以通过设置引入一个关于博勒集B的σ场的度量，即

$$p_X(b) = p(x^{-1}(b))。$$

我们称P_X 为随机变量X的*概率分布*。

定理3.10

集合函数P_X 是可数的加性。

证明

给定成对不相交的Borel集B_i，它们的反像$X^{-1}(B_i)$是成对不相交的，并且$X^{-1}(\bigcup_i B_i) = \bigcup_i X^{-1}(B_i)$，所以

$$p_X\left(\bigcup_i b_i\right) = p\left(x^{-1}\left(\bigcup_i b_i\right)\right) = p\left(\bigcup_i x^{-1}(b_i)\right) = \sum_i p(x^{-1}(b_i))$$
$$= \sum_i P_X(B_i)$$

根据需要。 □

因此（R，B，P_X）是一个概率空间。对于这一点，只需注意以下几点
$p_X(\mathrm{r}) = p(\omega) = 1$。

我们考虑一些简单的例子。假设X是常数，即$X\equiv a$。
然后我们称P_X 为集中在a处的*狄拉克度量*，并以δ_a 表示。显然

$$\delta_a(B) = \begin{cases} 1 & \text{如果} a \in B \\ 0 & \text{如果} a \notin B \text{。} \end{cases}$$

特别是，$\delta_a(\{a\}) = 1$。

如果X取2个值：

$$X(\omega) = \begin{cases} a, & \text{概率为}p \\ b & \text{概率为}1-p\text{、} \end{cases}$$

然后

$$P_X(B) = \begin{cases} 1 & \text{如果}a, b \in B \\ p & \text{如果}a \in B, b \notin B \\ 1-p & \text{如果}b \in B, a \notin B \\ 0 & \text{否则、} \end{cases}$$

于是

$$P_X(B) = p\delta_a(B) + (1-p)\delta_b(B)\text{。}$$

一般离散随机变量的分布（即只取有限多的不同值，可能在某些空集上除外）的形式是：如果X的值是以概率$p_i > 0$，$i = 1, 2, \ldots$ 取的a_i。$\sum p_i = 1$，那么

$$P_X(B) = \sum_{i=1}^{\infty} p_i \delta_{a_i}(B).$$

经典的例子是：

(i) 几何分布，其中$p_i = (1-q)q^i$，对于某些$q \in (0, 1)$、

(ii) 泊松分布，其中$p_i = \frac{\lambda^i}{i!}e^{-\lambda}$。

我们将不进一步讨论离散情况，因为这不是我们本文的主要目标，而且在许多关于概率论的初级文本中都有涉及（如[9]）。

现在考虑经典的概率空间，$\Omega=[0，1]$，F=B，P=$m_{[0,1]}$ - Lebesgue度量限制在[0,1]。我们可以举出用明确公式给出的随机变量的例子。

例如，让$X(\omega) = a\omega + b$。那么[0, 1]的图像是区间$[b, a+b]$，$P_X = \frac{1}{a} m|_{[b,a+b]}$，即对于Borel B

$$P_X(B) = \frac{m(B \cap [b, a+b])}{a}\text{。}$$

例3.1

假设一辆汽车在上午12点到下午1点之间随机离开A市。它以50英里/小时的速度向距离A市25英里的B市行驶。下午1时，汽车与B市之间距离的概率分布是什么？

显然，这个距离是0，概率为$\frac{1}{2}$，即如果汽车在离开之前12.30.作为起始时间的函数（表示为$\omega \in [0, 1]$）的函数，距离的形式为

$$X(\omega) = \begin{cases} 0 & \text{如果}\omega \in [0, \frac{1}{2}] \\ 50\omega - 25 & \text{如果}\omega \in (\frac{1}{2}, 1] \end{cases}$$

和$P_X = \frac{1}{2}P_1 + \frac{1}{2}P_2$其中$P_1 = \delta_0$，$P_2 = \frac{1}{25}m_{[0,25]}$。因此，在这个例子中、$P_X$是Dirac和Lebesgue措施的组合。

在以后的章节中，我们将在开发出处理计算所需的进一步机器后，探索更复杂的X和cor-responding分布的形式。

3.5.4 随机 变量的独立性

定义3.3

如果X、Y产生的σ场是独立的，那么它们就是独立的。

换句话说，对于R中的任何Borel集B，C、

$$p(x^{-1}(b) \cap y^{-1}(c)) = p(x^{-1}(b))p(y^{-1}(c))。$$

例3.2

让（Ω=[0，1]，M）配备Lebesgue度量。考虑$X = 1_{[0,\frac{1}{2}]}$、$Y = 1_{[\frac{1}{4},\frac{3}{4}]}$.那么$\mathcal{F}_X = \{\emptyset, [0,1], [0,\frac{1}{2}], (\frac{1}{2},1]\}$, $\mathcal{F}_Y = \{\emptyset, [0,1], [\frac{1}{4},\frac{3}{4}], [0,\frac{1}{4}) \cup (\frac{3}{4},1]\}$显然是独立的。

例3.3

设Ω如上，并设X（Ω）=Ω，Y（Ω）=1Ω。那么$\mathcal{F}_X = \mathcal{F}_Y = M$。一个σ场不能与自己独立（除非它是微不足道的）：以$A \in \mathcal{F}$为例，那么独立性要求P（$A \cap A$）=P（A）P（A）（集合A属于'两个'σ-场），即P（A）=P（A）^{2}，这只有在P（A）=0或P（A）=1时才会发生。所以一个独立于自身的σ场由度量为0或1的集合组成。

3.5.5 在数学上的应用 金融

考虑一个股票价格的模型，在时间上是离散的，即假设股票价格由随机变量序列$S(n)$给出，$n=1, 2, \ldots, N$。如果一个步骤的长度是h，那么我们的时间范围是$T=Nh$，我们经常写$S(T)$而不是$S(N)$。这种模型的一个例子是上一章中考虑的二叉树。回顾一下，欧式看涨期权是$(S(N)K)^+$（N是行权时间，K是行权价格，S是标的资产）形式的随机变量。它的一个自然的概括是一个随机变量，其形式为$f(S(N))$，用于一些可衡量的函数

f :这个随机变量当然是可测量的，相对于
由$S(N)$生成的σ场。这使我们能够制定一个一般的定义：

定义3.4

一个基础资产由序列$S(n)$代表、行使时间为N的欧洲衍生证券（或有债权）是一个随机变量X，*相对于*$S(N)$生成的σ场F是可测量的。

命题3.11

一个欧洲派生安全X必须具有$X=f(S(N))$的形式，适用于一些可衡量的实函数f。

上述定义对应用来说是不够的。例如，它没有涵盖基本衍生工具之一，即期货。回顾一下，*期货*合约的持有者有权根据标的证券的价值收到（或在负值的情况下有义务支付）一定序列（X（1），....$X(N)$）的现金支付，这取决于基础证券的价值。具体来说，如果一个步骤的长度是一年，r是年复利的无风险利率，那么

$$X(n) = S(n)(1 + r)^{N-n} - X(n - 1)(1 + r)^{N-n+1}。$$

为了引入一个涵盖期货的衍生证券的一般概念，我们首先考虑一个自然的概括

$$X(n) = f_n (S(0), S(1), \ldots, S(n))$$

然后我们再进一步推动一般性的水平：

定义3.5

一个衍生证券（或有债权），其标的资产由一个序列（$S(n)$）代表，到期时间为N，是一个序列（$X(1), \ldots X(N)$）的随机变量，这样$X(n)$就σ场而言是可测量的

F_n，由($S(0), S(1), \ldots, S(n)$)，对于每个$n = 1, \ldots, N$.

命题3.12

一个导数安全X必须是$X = f(S(0), S(1), \ldots, S(N))$，对于一些可衡量的$f$: $R^{N+1} \to R$。

我们可以再做一步，处理掉基础的随机变量。基础对象的作用将由一个递增的σ 场序列来扮演 F$_n$而我们会说，一个或有要求（在此避免使用另一个术语）是一个随机变量$X(n)$的序列，使得$X(n)$是F$_n$-可测量的，但在实际应用中几乎不需要这样的一般性。唯一与这种表述有关的情况是，没有数字观测而只有一些信息流的情况以事件和σ 场为模型。

例3.4

异国期权的回报取决于连续股票价格的整个路径。例如，行使时间为N的欧式回溯期权的回报由以下因素决定

$$f(x_0, x_1, \ldots, x_N) = max\{x_0, x_1, \ldots, x_N\} - x_N$$

练习3.9

找出跌停板看涨期权的函数f（这是一个欧式看涨期权，但如果在行使日期前的任何时候股票价格低于障碍物$L<S(0)$，它就不存在了）。

例3.5

考虑一个二项式模型中的美式看跌期权。我们将看到它符合上述的抽象方案。回顾一下，美式期权可以在到期前的任何时间行使，在n时间行使的看跌期权的回报是（$K\ S(n)$）。$^{+}$ -

为了简洁起见，写成$g(S(n))$，$g(x)=(K\ x)^{+}$。这个选项为持有人提供了与股票性质相同的现金流。后者是由股票决定的

价格和股票可以在任何时候出售，当然只有一次。美式期权也只能卖出或行使一次。这个期权的价值将用$P^A(n)$表示，我们将证明它是定义3.5中的衍生证券。

我们将证明，有可能写成

$$P^A(n) = f_n(S(n))$$

对于一些函数f_n。考虑一个在N=2时到期的期权。很明显

$$f_2(x) = g(x)$$

在n=1的时候，期权的持有者可以行使或者等待到n=2。等待的价值与n=1时发行的行使时间为N=2的欧式看跌期权的价值相同（众所周知，在第7.4.3节中将会详细介绍），可以计算出相对于某种概率p的折现回报的期望值）。美式看跌期权的价值是两者中较大的一个，因此

$$f_1(x) = \max n g(x), \frac{1}{1+r}[pf_2(xU) + (1-p)f_2(xD)] 。$$

同样的论证得出

$$f_0(x) = \max n g(x), \frac{1}{1+r}[pf_1(xU) + (1-p)f_1(xD)] .$$

一般来说，对于一个在N时间到期的美式期权，我们有以下一连串的递归公式：

$$f_N(x) = g(x)、$$

$$f_{n-1}(x) = \max n g(x), \frac{1}{1+r}[pf_n(xU)+(1-p)f_n(xD)] 。$$

3.6 命题的证明

证明（命题3.4的证明）

(i) 我们已经证明，如果f是可测的，那么f^+，f^-也是可测的。反过来说，注意$f(x) = f^+(x) - f^-(x)$所以定理3.2给出了结果。

(ii) 函数$u\mapsto|u|$是连续的，所以用F（u，v）=|u|的Lemma 3.3给出了$|f|$的可测性（另一种方法是用$|f|=f^+ + f^-$）。为了说明

反之亦然，取一个不可测量的集合A，让$f = 1_A\ 1_{A^c}$。它是不可测的

{ } | |

□

，因为$x : f(x) > 0 = A$是不可测的。但$f = 1$显然是可测的。

证明（命题3.9的证明）

由于$f \le$ ess sup$f$，$g \le$ ess sup$g$ a.e.，通过加法，我们有$f+g \le$ ess sup$f$+ess sup$g$ a.e.，所以ess sup f+ess sup g这个数字属于$\{z: f+g \le z$ a.e.$\}$这个集合的下限比这个数字小。 □

证明（命题3.11的证明）

首先要注意，由$S(N)$生成的σ场的形式是$=\{S(N)^{-1}(B): B \in$ Borel$\}$，因为这些集合形成了一个σ场，任何其他的σ场，使$S(N)$相对于它是可测的，都必须包含所有Borel集合的反像。接下来我们分三步进行：

1）假设$X = 1_A$ 为$A \in F$。那么$A = S(N)^{-1}(B)$为一个Borel子集的R. 把$f=1_B$，显然$X=f\,S(N)$。

2）如果X是一个阶梯函数，$X = \sum c_i 1_{A_i}$，那么取$f = \sum c_i 1_{B_i}$，其中$a_i = s(n)^{-1}(b_i)$。

3）一般来说，一个可测量的函数X可以通过阶梯函数来近似。观念$X_n = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} - 1_{Y^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))}$（详见命题4.10）
我们取$f = \limsup f_n$，其中f_n 与步骤2中的Y_n 相对应），以及序列显然收敛于$S(N)$的范围。 □

证明（命题3.12的证明）

1) 假设$X = 1_A$ 为$A \in F$。那么$A = (S(1), \ldots, S(N))^{-1}(B)$ 为Borel $B \subset R^N$，而$f = 1_B$，满足要求。

步骤2）和3）与前一个命题的证明相同。 □

4
整体性

为了简单起见，下面所发展的理论涉及到Lebesgue度量。然而，我们所需要的（除了我们讨论里曼积分的部分）是m是一个度量的属性，即一个定义在a的子集的σ场上的可数ad-ditive（扩展-）实值函数μ。F

固定集Ω。因此，为度量空间（R, , m）发展的理论 M
以下各节的内容几乎可以在不改变的情况下扩展到一个AB------。
严格给定的度量空间（Ω, , μF）。

我们鼓励读者牢记这种生成的可能性。我们将在本章末尾的概率部分，以及后面的章节中需要它。

4.1 积分的定义

我们现在能够解决我们之前发现的一个问题：如何整合像1_Q 这样的函数，它只取有限的值，但这些值的集合根本不是 "像区间"。

定义4.1

一个非负函数j : R → R，只取有限的值，即j的范围是一个不同的非负实数的*有限*集合{a_1 , a_2 , . . ., a_n }，是一个

简单函数，如果所有的集合

$$A_i = j^{-1}(\{a_i\}) = \{x : j(x) = a_i\}、\quad i = 1, 2, ..., n,$$

是可测量的集合。请注意，集合$A_i \in$ M 是成对不相交的，它们的联盟是R。

很明显，我们可以写出

$$\phi(x) = \sum_{i=1}^{n} a_i 1_{A_i}(x)$$

所以，（根据定理3.2）每个简单的函数都是可测的。

定义4.2

简单函数j在$E\in$M上的（Lebesgue）*积分*由以下方式给出：

$$\int_E \phi\, dm = \sum_{1}^{n} a_i\, m(A_i \cap E)。$$

(注：由于我们将允许$m(A_i) = +\infty$，我们在此使用惯例$0 \times \infty = 0$）。

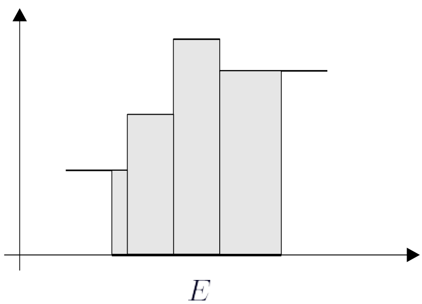

图4.1 一个简单函数的积分

例4.1

考虑简单的函数1_Q，它在Q上取值为1，在R Q上取值为0。根据上述定义，我们有

$$\int_{\mathbb{R}} 1_{\mathbb{Q}}\, dm = 1 \times m(\mathbb{Q}) + 0 \times m(\mathbb{R}\backslash\mathbb{Q}) = 0$$

因为Q是一个空集。回顾一下，这个函数*不是*黎曼不稳定的。同样地，1_C的积分为0，其中*C*是康托集。

练习4.1

求*j*在*E*上的积分，其中

(a) $\phi(x) = \mathrm{Int}(x), E = [0, 10]$

(b) $\phi(x) = \mathrm{Int}(x^2), E = [0, 2]$

(c) $\phi(x) = \mathrm{Int}(\sin x), E = [0, 2\pi]$

而Int表示一个实数的整数部分。(注意：许多文本使用符号[*x*]来表示Int(*x*)。我们倾向于使用Int，以提高清晰度）。

为了将积分扩展到更普遍的函数，亨利-勒贝斯格（1902年）采用了一个表面上很明显，但却很微妙的方法：他没有将一个有界函数*f*的域划分为许多小区间，而是划分为
其范围分为有限数量的小区间，其形式为$A_i = [a_{i-1}, a_i)$、
并通过*上位和*近似计算*f*的图形下的 "面积"。

$$S(n) = \sum_{i=1}^{n} a_i \, m(f^{-1}(A_i))$$

和 *下限之和*

$$s(n) = \sum_{i=1}^{n} a_{i-1} \, m(f^{-1}(A_i))$$

分别是；那么可整定的函数具有这样的特性：所有上层和的下限等于所有下层和的上限--反映了黎曼的构造（也见图3.1）。

一个世纪以来，关于Lebesgue积分的经验导致了许多等价的定义，其中一些在技术上（如果不总是在概念上）更简单。我们将遵循一个版本，它虽然与Lebesgue的原始结构非常相似，但允许我们充分利用已经开发的度量理论。首先我们停留在非负函数上：

定义4.3

$\in \mathcal{M}$ $\int$

对于任何非负的可测量函数f和E 的*积分*$\int_E f\, dm$被定义为

$$\int_E f\, dm = \sup Y(E, f)$$

其中

$$Y(E,f)=\left\{\int_E j\,dm:0\le j\le f，j是简单的\right\}。$$

注意，积分可以是+∞，而且总是非负的。显然，集合$Y(E,f)$的形式总是$[0,x]$或$[0,x)$，其中$x=+\infty$的值是允许的。

如果$E=[a,b]$，我们将积分写为

$$\int_a^b f\,\mathrm{dm}、\quad \int_a^b f(x)\,dm(x)、$$

或甚至作为$\int_a^b f(x)\,dx$，当不可能发生混淆时（我们设定$\int_a^b f\,dm=-\int_b^a f\,dm$如果$a>b$）。符号$\int f\,dm$意味着$\int_{\mathrm{R}} f\,dm$

显然，如果对于某个$A\in\mathcal{M}$和一个非负可测函数g，我们在A^c上有$g=0$，那么任何位于g以下的非负简单函数在A^c上一定是零。将此应用于$g=f.1_A$我们得到重要的特征

$$\int_A f\,dm=\int f\,1_A\,\mathrm{dm}。$$

练习4.2

假设$f:[0,1]\to\mathrm{R}$的定义是：让$f(x)=0$，在康托尔集和$f(x)=k$，对于从$[0,1]$中删除的长度为3的每个区间的所有x^{-k}。计算$\int_0^1 f\,dm$。

提示 回顾一下，$\sum_{k=1}^{\infty} kx^{k-1}=\frac{d}{dx}\left(\sum_{k=0}^{\infty}x^k\right)=\frac{1}{(1-x)^2}$，当$|x|<1$时。

如果f是一个简单的函数，我们现在有两个积分的定义；因此为了保持一致性，你应该仔细检查上述定义是否重合。

命题4.1

对于简单函数，定义4.2和4.3是等同的。

此外，我们可以证明以下简单函数积分的基本属性：

定理4.2

让j，ψ是简单函数。那么：

(i) 如果$\phi \le \psi$，则$\int_E j\,dm \le \int_E \psi\,dm$、

(ii) 如果A、B是M中不相交的集合，那么

$$\int_{A\cup B} \phi\,dm = \int_A \phi\,dm + \int_B \phi\,dm、$$

(iii) 对于所有常数$a>0$

$$\int_E aj\,dm = \int_E aj\,dm。$$

证明

(i) 请注意，$Y(E,\phi) \subseteq Y(E,\psi)$（我们使用定义4.3）。

(ii) 利用m的特性，我们有（$j = \sum c\,1_{iD}$ i）

$$\begin{aligned}\int_{A\cup B} j\,dm &= \sum c_i\,m(D_i \cap (A\cup B))\\ &= \sum c_i\,m(D_i \cap A) + m(D_i \cap B)\\ &= \sum c_i\,m(D_i \cap A) + \sum c_i\,(D_i \cap B)\\ &= \int_A \phi\,dm + \int_B \phi\,dm。\end{aligned}$$

(iii) 如果$j = \sum c\,1_{iA}$ i，那么$aj = \sum ac\,1_{iA}$ i，并且

$$\int_E aj\,dm = \Sigma\, ac_i\,m(E\cap A_i) = a \sum c_i\,m(E\cap A_i) = a\int_E \phi\,dm$$

根据需要。 □

接下来我们表明，简单函数的积分的属性延伸到非负可测量函数的积分：

定理4.3

假设f和g是非负的可测量函数。

(i) 如果$A\in$M，且A上$f\le g$，则

$$\int_A f\,dm \le \int_A g\,dm。$$

(ii) 如果$B\subseteq A$，A，$B\in$M，则

$$\int_B f\,dm \le \int_A f\,dm。$$

(iii) 对于a≥0、
$$\int_A af\,dm = a\int_A f\,\mathrm{dm}。$$

(iv) 如果A是空的，那么
$$\int_A f\,dm = 0。$$

(v) 如果A，B?M，$A∩B$=∅，则
$$\int_{A\cup B} f\,dm = \int_A f\,dm + \int_B f\,\mathrm{dm}。$$

证明

(i) 请注意，$Y(A, f)$ Y ⊆ (A, g) (在g下比在f下有更多的空间来挤压简单的函数)，更大的集合的sup更大。

(ii) 如果j是一个在B上低于f的简单函数，那么在B外将其扩展为零，我们得到一个在A上低于f的简单函数。这些简单函数的积分是相同的，所以Y（B，f）⊆ Y（A，f），我们的结论如（i）。

(iii) 集合$Y(A, af)$的元素的形式是$a\,x$，其中x $Y(A, f)$，所以它们的上位数之间也有同样的关系。

(iv) 对于任何一个简单的函数j，$\int_A j\,dm = 0$。要看到这一点，取$j = \sum c_i 1_{E_i}$，比如，那么$m(A\cap E_i) = 0$，对于每个i，所以$Y(A, f) = \{0\}$。

(v) Y（A?B，f）的元素的形式是$\int_{A?B} j\,dm$，所以根据定理4.2
(ii) 它们的形式是$\int_A \phi\,dm + \int_B \int \phi\,dm$。所以$Y$（$A\cup B$，$f$）=$Y$（$A$，$f$）+ $Y(B, f)$，取其上位数，可得 $\int_{A\cup B} f\,dm \le \int_A f\,dm + \int_B f\,dm$。对于

假设简单函数j和ψ满足以下条件：j≤f在A上，j=0离开A，而ψ≤f在B上，ψ=0离开B：由于$A∩B$=∅，我们可以通过设置γ=j在A上，γ=ψ在B上，以及γ=0在$A\cup B$外，构造一个新的简单函数γ≤f。

$$\begin{aligned}\int_A \phi\,dm + \int_B \psi\,dm &= \int_A \gamma\,dm + \int_B \gamma\,dm \\ &= \int_{A?B} \gamma\,dm \\ &\le \int_{A?B} f\,\mathrm{dm}。\end{aligned}$$

在右边，我们有一个上界，它对所有位于下面的简单函数仍然有效。f 在 $A\cup B$.因此，取上界在 ϕ 和 ψ 分别在左边得到$\int_A f\,dm + \int_B f\,dm \le \int_{A\cup B} f\,dm$。

□

练习4.3

证明以下积分的均值定理：如果 $a \le f(x) \le b$ 为$x\in A$，则$am(A)\le\int_A f\,dm\le bm(A)$。

我们现在确认，空集正是整数理论的 "可忽略的集"：

定理4.4

假设f是一个非负的可测量函数。那么，当且仅当$\int_{\mathbb{R}} f\,dm = 0$时，$f = 0$ a.e.。

证明

首先，请注意，如果 $f = 0$ a.e.和$0 \le \phi \le f$ 是一个简单的函数，那么 $\phi = 0$ a.e. 因为f和ϕ都不取负值。因此，$\int_{\mathbb{R}} \phi\,dm = 0$，对于所有这些$\phi$，所以 $\int_{\mathbb{R}} f\,dm = 0$也是如此。

反过来说，给定 $\int_{\mathbb{R}} f\,dm = 0$，让$E = \{x : f(x) > 0\}$。我们的目标是要证明$m(E)=0$。

$$E_n = f^{-1}\left(\left[\frac{1}{n}, \infty\right)\right) \text{ 为 } n \ge 1.$$

显然，$\{E_n\}$增加到E与

$$E = \bigcup_{n=1}^{\infty} E_n$$

为了证明$m(E)=0$，只需证明$m(E_n)=0$，对所有的n。

定理2.13.)函数$\phi = \frac{1}{n} 1_{E_n}$ 是简单的，并且根据定义$\phi\le f$的E_n。

所以

$$\int_{\mathbb{R}} \phi\,dm = \frac{1}{n} m(E_n) \le \int_{\mathbb{R}} f\,dm = 0$$

因此，对于所有的n，$m(E_n) = 0$。 □

利用到目前为止所证明的结果，下面的 "a.e. "版本的积分的单调性不难证明：

命题4.5

如果f和g是可测量的，那么$f \leq g$ a.e.意味着$\int f \, dm \leq \int g \, dm$。

提示 让$A = \{x : f(x) \le g(x)\}$，那么$B = A^c$是空的，$f 1_A \le g1_A$。现在使用定理4.3和4.4。

利用定理3.2和3.5，你现在应该对我们在命题3.4中已经注意到的一个结果进行第二次证明，但为了强调，在此重复一下：

命题4.6

函数f：如果f^+和f^-都是可测量的，则R → R是可测量的。

4.2 单调收敛 定理

Lebesgue积分的核心是其收敛理论。我们可以通过给出一个著名的结果来开始研究这个问题

定理4.7（法图定理）

如果$\{f_n\}$是一个非负可测量的函数序列，那么

$$\liminf_{n\to} \int_E f_n\,dm \ge \int_E \liminf_{n\to} f_n\,\mathrm{dm}。$$

证明

撰写
$$f = \liminf_{n\to\infty} f_n$$

并回顾说
$$f = \lim_{n\to\infty} g_n$$
其中$g_n = \inf_{k\ge n} f_k$（序列g_n是不递减的）。让j是一个简单的函数，$j \le f$。为了证明
$$\int_E f\,dm \le \liminf_{n\to\infty} \int_E f_n\,dm$$
足以见得

$$\int_E \phi \, dm \leq \liminf_{n\to\infty} \int_E f_n \, dm$$

对于任何这样的j。

f=0的集合是不相关的，因为它对$\int_E f\,dm$没有贡献，所以我们可以在不丧失一般性的情况下，假设E上的f>0。

$$\underline{\phi}(x) = \begin{cases} \phi(x) - \varepsilon > 0 & \text{如果}\,j(x)>0 \\ 0 & \text{如果}\,j(x)=0\text{或}\,x \notin E \end{cases}$$

其中ε足够小，以确保j≥0。

现在$\bar{j} < f$，g_n 3 f，所以'最终'$g_n \geq \bar{\varphi}$。我们做最后的陈述更确切地说，是把

$$A_k = \{x : g_k(x) \geq \bar{\phi}(x)\}$$

而我们有

$$A_k \subseteq A_{k+1}、\quad \bigcup_{k=1} A_k = R_{\text{o}}$$

下一步、

$$\int_{A_n \cap E} \bar{\phi}\,dm \leq \int_{A_n \cap E} g_n\,dm \text{ (因为}g_n\text{ 在}A\text{上支配}j\text{)}_k$$

$$f_k\,dm \quad \text{对于}k \geq n\text{（根据}g\text{的定义}_n\text{）。}$$

$$\leq \int_E f_k\,dm \text{ (因为}E\text{是较大的集合)}$$

对于k≥n.因此

$$\int \phi\,dm \leq \liminf_{k \to \infty} \int_E f_k\,dm_{\text{o}} \tag{4.1}$$

现在我们让$n \to \infty$：写$\bar{j} = \sum_{i=1} c_i 1_{B_i}$ 对于一些$c_i \geq 0, B_i \in$ M, $i \leq l$

$$\int_{A_n \cap E} \bar{\phi}\,dm = \sum_1 c_i\,m(A_n \cap E \cap B_i) \to \sum_1 c_i\,m(E \cap B_i) = \int_E \bar{\phi}\,dm$$

而不等式（4.1）在极限中仍然是真实的：

$$\int_E \bar{\phi}\,dm \leq \liminf_{k \to} \int_E f_k\,dm_{\text{o}}$$

我们已经很接近了--我们所需要的就是在最后一个关系中用j代替$\bar{j}$。这将通过让ε→0来实现，但需要注意一些问题。

假设$m(\{x : j(x) > 0\}) < \infty$。那么

$$\int_E \bar{\phi}\, dm = \int_E j\, dm - \varepsilon m(\{x : j(x) > 0\})$$

而我们通过让$\varepsilon \to 0$得到结果。

该案例 $m(\{x : \phi(x) > 0\}) = \infty$ 必须单独处理。这里 $\int_E \varphi\, dm = \infty$，所以 $\int_E f\, dm = \infty$。我们必须证明

$$\liminf_{k\to} \int_E f_k\, dm = \infty。$$

让c_i 是j的值，让$a = \frac{1}{2}\min\{c_i\}$ （$\{c_i\}$是一个有限集！）。同样地在上面把

$$D_n = \{x : g_n(x) > a\}$$

和

$$\int_{D_n \cap E} g_n\, dm \to \infty$$

因为D_n 3 R。如前所述

$$\int_{D_n \cap E} g_n\, dm \le \int_{D_n \cap E} f_k\, dm \le \int_E f_k\, dm$$

对于$k \ge n$，所以$\liminf \int_E f_k\, dm$必须是无限的。 □

例4.2

让$f_n = 1_{[n,n+1]}$。显然 $\int f_n\, dm = 1$ for all n, $\liminf f_n = 0$ (= $\lim f_n$), 所以上述不等式可能是严格的，我们有

$$\int (\lim f_n)\, dm \neq \lim \int f_n\, dm。$$

练习4.4

构建一个具有上述严格不等式的函数序列的例子，这样所有的f_n，在区间[0，1]外为零。

现在很容易证明两个主要收敛定理中的一个。

定理4.8（单调收敛定理）

如果$\{f_n\}$是一个非负可测量函数序列，并且$\{f_n(x) : n \ge 1\}$对于每个x都单调地增加到$f(x)$，即f_n 3 f点，那么

$$\lim_{n\to\infty} \int_E f_n(x)\, dm = \int_E f\, dm。$$

证明

由于$f_n \le f$，$\int_E f_n\, dm \le \int_E f\, dm$，所以

$$\limsup_{n\to\infty} \int_E f_n\, dm \le \int_E f\, dm。$$

法图的定理给出

$$\int_E f\, dm \le \liminf_{n\to} \int_E f_n\, dm$$

这与基本关系一起

$$\liminf_{n\to\infty} \int_E f_n\, dm \le \limsup_{n\to\infty} \int_E f_n\, dm$$

给予

$$\int_E f\, dm = \liminf_{n\to} \int_E f_n\, dm = \limsup_{n\to} \int_E f_n\, dm$$

因此，序列$\int_E f_n\, dm$收敛于$\int_E f\, dm$。 □

推论4.9

假设 $\{f_n\}$和 f 是非负的和可测量的。如果 $\{f_n\}$ 增加到几乎f无处不在，那么我们仍然有 $\int_E f_n\, dm$ 3 $\int_E f\, dm$ 对于所有可测的 E.

证明

假设f_n f a.e.，A是收敛性成立的集合，因此

A^c is null. We can define

$$g_n = \begin{cases} f_n & \text{on } A \\ 0 & 在A上^c \end{cases}、$$

$$g = \begin{cases} f & 在A上 \\ 0 & 在A上^c \end{cases}。$$

然后用$E=[E\cap A^c]\cup[E\cap A]$我们得到

$$\int_E g_n\, dm = \int_{E\cap A} f_n\, dm + \int_{E\cap A^c} 0\, dm$$

$$= \int_{E\cap A} f_n\, dm + \int_{E\cap A^c} f_n\, dm$$

$$= \int_E f_n\, dm$$ (因为$E\cap A^c$是空的)，同样

地，$\int g\,dm = \int$

$f\,dm$。收敛性$g_n \to g$

处处成立，所以根据定理4.8、 $\int_E g_n\,dm \to \int_E g\,dm$ □

为了应用单调收敛定理，通过简单函数的增加序列来近似非负可测量函数是很方便的。

命题4.10

对于任何非负可测量的f，都有一个非负简单函数的序列s_n，以便s_n ３ f。

提示把

$$s_n = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} - 1_{f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))}.$$

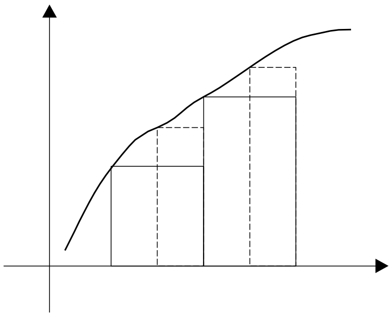

图4.2 通过简单的函数进行逼近

4.3 可积分的 函数

所有的艰苦工作都完成了：我们可以非常容易地将积分扩展到一般的实数函数，使用任何可测函数f的正数部分$f^+ = \max(f, 0)$，以及负数部分$f^- = \max(-f, 0)$：我们将不单独使用非负的可测函数f：正如我们在命题3.4中看到的，$|f|$

可以是可测量的，而不需要f是可测量的!

定义4.4

如果$E\in$▮和可测量的函数f同时具有$\int_E f^+ \, dm$和$\int_E f^- \, dm$有限的，那么我们说f是*可整定的*，并定义为

$$\int_E f \, dm = \int_E f^+ \, dm - \int_E f^- \, dm。$$

所有在E上可整定的函数的集合用$L^1(E)$表示。在下文中，E将是固定的，我们常常简单地把L^1写成$L^1(E)$。

练习4.5

在$L^1(E)$中，(a)$E=(0，1)$；(b)$E=(1，\infty)$，对于哪个α，是$f(x)=x^\alpha$?

请注意，f是可整定的，如果$|f|$是可整定的，并且

$$\int_E |f|\,dm = \int_E f^+\,dm + \int_E f^-\,dm。$$

因此，Lebesgue积分是一个 "绝对 "积分：我们不能通过取消大的正负部分来使一个函数可积分。这就造成了一些具有不适当的黎曼积分的函数不能成为Lebesgue积分的结果（见第4.5节）。

非负函数的积分的属性延伸到任何不一定是非负的可积分函数。

命题4.11

如果f和g是可整定的，$f\leq g$，那么$\int f\,dm\leq \int g\,dm$。

提示 如果$f\leq g$，则$f^+ \leq g^+$，但$f^- \geq g^-$。

备注4.1

我们观察到（根据[12]，5.12），许多关于可控函数的结果的证明都遵循一个标准模式，利用线性和单调收敛特性。为了证明一个 "线性 "的结果对所有函数在

在$L^1(E)$这样的空间中，我们分四个步骤进行：

(i) 验证所要求的属性对指标函数是成立的--根据定义通常是这样、

(ii) 用线性来扩展该属性到非负的简单函数、

(iii) 然后用单调收敛性来证明该属性是所有非负可测量函数所共有的、

(iv) 最后，通过写$f = f^+ - f^-$并再

次使用线性，扩展到整个函数类。

下一个结果很好地说明了这一技术。

我们希望证明，映射$f\mapsto\int_A f\,dm$是线性的。这一事实本身就很有趣，但也将使我们能够证明L^1是一个矢量空间。

定理4.12

对于任何可整定的函数f，g，它们的总和$f+g$也是可整定的，并且

$$\int_E (f+g)\,dm=\int_E f\,dm+\int_E g\,dm。$$

证明

我们采用备注4.1中描述的技术。

步骤1.首先假设f和g是非负的简单函数。其结果是一个常规计算的问题：让$f=\sum a_i 1_{A_i}$，$g=\sum b_j 1_{B_j}$。总和$f+g$也是一个简单函数，可以写成以下形式

$$f+g=\sum_{i,j}(a_i+b_j)1_{A_i\cap B_j}。$$

因此

$$
\begin{aligned}
\int_E (f+g)\,dm&=\sum_{i,j}(a_i+b_j)m(A_i\cap B_j\cap E)\\
&=\sum_i\sum_j a_i\,m(A_i\cap B_j\cap E)+\sum_j\sum_i b_j\,m(A_i\cap B_j\cap E)\\
&=\sum_i a_i\sum_j m(A_i\cap B_j\cap E)+\sum_j b_j\sum_i m(A_i\cap B_j\cap E)\\
&=\sum_i a_i\,m\Big(\bigcup_j(A_i\cap B_j\cap E)\Big)+\sum_j b_j\,m\Big(\bigcup_i(A_i\cap B_j\cap E)\Big)\\
&=\sum_i a_i\,m\Big(A_i\cap\bigcup_j B_j\cap E\Big)+\sum_j b_j\,m\Big(B_j\cap\bigcup_i A_i\cap E\Big)\\
&=\sum_i a_i\,m(A_i\cap E)+\sum_j b_j\,m(B_j\cap E)\\
&=\int_E f\,dm+\int_E g\,dm
\end{aligned}
$$

其中我们使用了m的可加性以及A_i覆盖R的事实，对于B_j也是如此。

第二步。现在假设f, g是非负的可测量（不一定是简单）的函数。根据命题4.10，我们可以找到简单函数的序列s_n, t_n，使得s_n f和t_n g。显然，$s_n + t_n$ $f + g$，因此利用单调收敛定理和简单函数的可加性属性，我们得到

$$\int_E (f+g)\, dm = \lim_{n\to\infty} \int_E (s_n + t_n)\, dm$$

$$＝四分之一 \lim_{n\to\infty} \int_E s_n\, dm + \lim_{n\to\infty} \int_E t_n\, dm$$

$$= \int_E f\, dm + \int_E g\, dm。$$

这尤其意味着，如果f和g的积分是有限的，那么$f+g$的积分就是有限的。

第3步。最后，让f，g是任意可整定的函数。由于

$$\int_E |f + g|\, dm \le \int_E (|f| + |g|)\, dm、$$

我们可以用步骤2来推导出左手边是有限的。

我们有

$$f + g = (f + g)^+ - (f + g)^-$$

$$f + g = (f^+ - f^-) + (g^+ - g)^-$$

那么

$$(f + g)^+ - (f + g)^- = f^+ - f^- + g^+ - g^- 。$$

我们重新排列等式，使其两边只有加数

$$(f + g)^+ + f^- + g^- = f^+ + g^+ + (f + g)^- 。$$

我们在两边都有非负的函数，所以根据我们到目前为止所证明的

$$\int_E (f+g)^+\, dm + \int_E f^-\, dm + \int_E g^-\, dm = \int_E f^+\, dm + \int_E g^+\, dm + \int_E (f+g)^-\, dm$$

因此

$$\int_E (f+g)^+\, dm - \int_E (f+g)^-\, dm = \int_E f^+\, dm - \int_E f^-\, dm + \int_E g^+\, dm - \int_E g^-\, dm。$$

根据积分的定义，最后一个关系意味着定理的要求。

□

下面的结果是单调收敛的一个常规应用：

命题4.13

如果f是可整定的，$c\in\mathbb{R}$，那么

$$\int_E (cf)\,dm = c\int_E f\,dm。$$

提示 用一连串的简单函数来近似f。我们完成了$\mathcal{L}^1$是

一个矢量空间的证明：

定理4.14

对于任何可测量的E，$\mathcal{L}^1(E)$是一个矢量空间。

证明

让f，$g\in\mathcal{L}^1$。为了证明$f+g\in\mathcal{L}^1$，我们必须证明$|f+g|$是可整数的：

$$\int_E |f+g|\,dm \le \int_E (|f|+|g|)\,dm = \int_E |f|\,dm + \int_E |g|\,dm < \infty。$$

现在让c是一个常数：

$$\int_E |cf|\,dm = \int_E |c|\,|f|\,dm = |c|\int_E |f|\,dm < \infty$$

因此，$cf\in\mathcal{L}^1(E)$。 □

现在我们可以回答一个重要的问题，即积分在多大程度上决定了积分体
。

定理4.15

如果$\int_A f\,dm \le \int_A g\,dm$ 为所有$A\in\mathcal{M}$，则 $f\le g$ 几乎无处不在。特别是，如果$\int_A f\,dm = \int_A g\,dm$ 对于所有$A\in\mathcal{M}$，那么$f=g$几乎无处不在。

证明

根据积分的可加性（以及下面的命题4.12），只需证明对于所有$A\in\mathcal{M}$的$\int_A h\,dm\geq 0$意味着$h\geq 0$（然后取$h=g-f$）。

写$A = \{x : h(x) < 0\}$; 然后$A = \bigcup A_n$ 其中$A_n = \{x : h(x) \le -\frac{1}{n}\}$。通过积分的单调性

$$\int_{A_n} h\, dm \le \int_{A_n} \frac{-1}{n} dm = -\frac{1}{n} m(A_n)、$$

是非负的，但这只能在$m(A_n)=0$时发生。

集 A_n随着 n 增加，因此 $m(A) = 0$，因此$h(x) \ge 0$几乎无处不在。类似的论证表明，如果$\int_A h\, dm \le 0$，对于所有A，那么$h \le 0$，即。这就意味着定理的第二个主张：把$h=g-f$和 $\int_A h\, dm$ 既非负数又非正数，因此$h \ge 0$，而$h \le 0$ a.e. 因此$h = 0$ a.e. □

下一个命题列出了可整数函数的其他重要属性，其直接证明是迄今为止所证明的结果的典型应用。

命题4.16

(i) 一个可整定的函数是a.e.有限的。

(ii) 对于可测量的f和A

$$m(A) \inf_A f \le \int_A f\, dm \le m(A) \sup_A f。$$

(iii) $\left|\int f\, dm\right| \le \int |f|\, dm$。

(iv) 假设$f \ge 0$和 $\int f\, dm = 0$. 那么$f = 0$ a.e.

下面的定理使我们有可能构建许多有趣的度量，并且对概率分布的发展至关重要。

定理4.17

设$f \ge 0$，则$A' \to \int_A f\, dm$是一个措施。

证明

记为$\mu(A) = \int_A f\, dm$。我们的目标是要证明

$$\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i)$$

对于成对不相交的E_i。为此，考虑序列$g_n = f\mathbf{1}_{\bigcup_{i=1}^{n} E_i}$和注意，$g_n \uparrow g$，其中$g = f\mathbf{1}_{\bigcup_{i=1}^{\infty} E_i}$。现在

$$\int g\,dm = \mu\Big(\bigcup_{i=1}^{\infty} E_i\Big)、$$

$$\int g_n\,dm = \int_{\bigcup_{i=1}^{n} E_i} f\,dm = \sum_{i=1}^{n} \int_{E_i} f\,dm = \sum_{i=1}^{n} \mu(E_i)$$

和单调收敛定理完成了证明。 □

4.4 主导收敛性 定理

分析学中的许多问题都集中在应用于某些函数的两个极限过程的顺序可以互换的条件上。由于积分是应用于可测量函数的极限过程，因此很自然地要问，在什么条件下，在一个点式（或点式a.e.）收敛序列（f_n）上，积分的极限是点式极限函数f的积分、

也就是说，我们什么时候可以说$\lim \int f_n\,dm = \int (\lim f_n)\,dm$？单调递增定理（定理4.8）提供了答案，即这个结论对非负可测函数的单调递增序列有效，当然，在这种情况下，极限可能等于+。下面的例子表明，对于一般的可整数函数序列，如果没有一些进一步的条件，该结论将不成立：

例4.3

让$f_n(x) = n\mathbf{1}_{[0,\frac{1}{n}]}(x)$。显然，对于所有的$x$，$f_n(x) \to 0$，但$\int f_n(x)\,dx = 1$。

在实践中最有用的极限定理指出，对于一个被可整数函数*支配的等价*收敛序列，收敛性是成立的。法图尔定理再次成为证明的关键。

定理4.18（支配性收敛定理）

假设$E \in \mathcal{M}$。设(f_n)是一连串可测量的函数，使得
$|f_n| \le g$ a.e. on E for all $n \ge 1$, where g is integrable over E. If $f = \lim_{n\to\infty} f_n$ a.e. 那么f在E上是可整定的，并且

$$\lim_{n\to\infty}\int_E f_n(x)\, dm = \int_E f \,dm。$$

证明

假设$f_n \geq 0$。Fatou's lemma给出了

$$\int_E f\,dm \leq \liminf_{n\to\infty} \int_E f_n\,dm。$$

因此，只需证明

$$\limsup_{n\to} \int_E f_n\,dm \leq \int_E f\,dm。 \tag{4.2}$$

法图定理适用于$g - f_n$，得到的是

$$\int_E \lim_{n\to\infty} (g - f_n) \leq \liminf_{n\to\infty} \int_E (g - f_n)\,dm。$$

在左边，我们有

$$\int_E (g - f)\,dm = \int_E g\,dm - \int_E f\,dm。$$

在右边

$$\begin{aligned}
&\liminf_{n\to} \int_E (g - f_n)\,dm \\
&= \liminf_{n\to} \int_E g\,dm - \int_E f_n\,dm \\
&= \int g\,dm - \limsup_{n\to} \int_E f_n\,dm、
\end{aligned}$$

其中我们使用了一个基本事实，即

$$\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n。$$

把这些放在一起，我们得到

$$\int_E g\,dm - \int_E f\,dm \leq \int_E g\,dm - \limsup_{n\to} \int_E f_n\,dm。$$

最后，减去$\int_E g\,dm$（它是有限的）并乘以-1，得出（4.2）。

现在考虑一个一般的、不一定是非负数的序列（f_n）。因为根据假设

$$-g(x) \leq f_n(x) \leq g(x)$$

我们有

$$0 \leq f_n(x) + g(x) \leq 2g(x)$$

我们可以把对非负函数证明的结果应用到序列上

$f_n(x) + g(x)$ （函数$2g$当然是可整数的）。 □

例4.4

回到定理前面的例子，$f_n = n1_{[0,\frac{1}{n}]}$，我们可以看到，无法找到一个可整定的$g$来支配$f_n$。最小的上界是$g(x) = \sup_n f_n(x)$, $g(x) = k$ on $(\frac{1}{k+1}, \frac{1}{k}]$ 所以

$$\int g(x)\,dx = \sum_{k=1}^{\infty} k\left(\frac{1}{k} - \frac{1}{k+1}\right) = \sum_{k=1}^{\infty} \frac{1}{k+1} = +\infty.$$

对于一个典型的正面例子，请考虑

$$f_n(x) = \frac{n \sin x}{1 + n^2 x^{1/2}}$$

对于$x \in$（0，1）。显然，$f_n(x) \to 0$。要得出$\lim_n \int f_n\,dm = 0$的结论，我们需要一个可整定的支配函数。这通常是需要一些巧思的地方，然而在本例中，最直接的估计就足够了：

$$\frac{n \sin x}{1 + n^2 x^{1/2}} \le \frac{n}{1 + n^2 x^{1/2}} \le \frac{n}{n^2 x^{1/2}} = \frac{1}{n x^{1/2}} \le \frac{1}{x^{1/2}}.$$

(从第一原理看，支配函数$g : x \mapsto \frac{1}{\sqrt{x}}$是整在[0，1]上的抓取可能是相当乏味的--参见第一章中关于$x \mapsto \sqrt{x}$的黎曼积分的工作实例。然而，我们将很快表明，如果一个有界函数的Lebesgue积分和Riemann积分存在，那么后者就会重合，因此我们可以应用微积分基本定理来确认g的可整定性)。

以下事实以后会很有用。

命题4.19

假设f是可整定的，并定义$g_n = f\,1_{[-n,n]}$，$h_n = \min(f, n)$ (两者都截断) f以某种方式：g_n 在有界区间外消失，h_n 是有界的）。那么$\int |f - g_n|\,dm \to 0$，$\int |f - h_n|\,dm \to 0$。

提示 使用支配性收敛定理。

练习4.6

使用支配收敛定理，求出

$$\underset{n\to\infty}{\text{肢体}} \int_{1}^{\infty} f_n(x)\, dx$$

其中

$$f_n(x) = \frac{\sqrt{x}}{1 + nx^3}。$$

练习4.7

调研收敛的

$$\int_a^\infty \frac{n^2 x e^{-n^2 x^2}}{1 + x^2}\, dx$$

对于a>0，以及对于a=0。

练习4.8

调研收敛的

$$\int_0^\infty \frac{1}{(1+\frac{x}{n})^n \sqrt{x}}\, dx。$$

我们将需要对定理4.12进行如下扩展：

命题4.20

对于一连串的非负可测量的函数f_n，我们有

$$\int \sum_{n=1}^\infty f_n\, dm = \sum_{n=1}^\infty \int f_n\, dm。$$

提示 序列$g_k = \sum_{n=1}^k f_n$是增加的，并且收敛于$\sum_{n=1}^\infty f_n$。

我们还不能得出结论说右边的系列之和是 a.e. 有限，所以$\sum_{n=1}^\infty f_n$不需要是可整数的。然而：

定理4.21（Beppo-Levi）

假设

$$\sum_{k=1}^\infty \int |f_k|\, dm 是有限的。$$

那么数列$\sum_{k=1}^\infty f_k(x)$几乎对所有的x都收敛，其和是可整数的、和

$$\int \sum_{k=1}^\infty f_k\, dm = \sum_{k=1}^\infty \int f_k\, dm。$$

证明

函数$\phi(x) = \sum_{k=1}^{\infty} |f_k(x)|$是非负的、可测量的，并且通过Propo-第4.20段

$$\int \phi \, dm = \sum_{k=1}^{\infty} \int |f_k| \, dm。$$

这是有限的，所以j是可整数的。因此，j是有限的等价物。$\sum_{k=1}^{\infty} |f_k(x)|$ a.e.收敛，因此系列$\sum_{k=1}^{\infty} f_k(x)$几乎对所有的$x$都收敛（因为它绝对收敛）。让$f(x) = \sum_{k=1}^{\infty} f_k(x)$（对于系列发散的$x$，把$f(x) = 0$ - 我们选择的值不重要，因为
这种x的集合是空的）。对于所有的部分和，我们有

$$\left|\sum_{k=1}^{n} f_k(x)\right| \leq j(x)$$

因此，我们可以应用主导收敛定理来找到

$$\begin{aligned} \int f \, dm &= \int 限于_{n\to\infty} \sum_{k=1}^{n} f_k \, dm \\ &= 四分之一_{n\to\infty} \int \sum_{k=1}^{n} f_k \, dm \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int f_k \, dm \\ &= \sum_{k=1}^{\infty} \int f_k \, dm \end{aligned}$$

根据需要。 □

例4.5

忆及$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$，我们可以使用贝波-列维定理来评估积分$\int_0^1 \left(\frac{\log x}{1-x}\right)^2 dx$；首先让$f_n(x) = nx^{n-1}(\log x)^2$为$n \geq 1$，$x \in (0, 1)$，所以$f_n \geq 0$，$f_n$是连续的，因此是可测量的，并且$\sum_{n=1}^{\infty} f_n(x) = \left(\frac{\log x}{1-x}\right)^2 = f(x)$对$x$来说是有限的$\in (0, 1)$.根据Beppo-Levi的和是可整定的，且$\int_0^1 f(x)\,dx = \sum_{n=1}^{\infty} \int_0^1 f_n(x)\,dx$。为了计算$\int_0^1 f_n(x)\,dx$我们首先使用分项积分，得到$\int_0^1 x^{n-1}(\log x)^2\,dx = \frac{2}{n^3}$。因此$\int_0^1 f(x)\,dx = 2\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{3}$.

练习4.9

以下是上述主题的变化：

(a) 对于$\alpha\in\mathbb{R}$的哪些值，幂级数 $\sum_{n\geq 0} n\, x^{\alpha n}$ 在[-1, 1]上定义一个可整定的函数？

(b) 证明 $\int_0^{\infty} \frac{x}{e^x-1}dx = \frac{\pi^2}{6}$。

4.5 与Riemann 积分的关系

我们引入Lebesgue积分的主要动机是为度量和积分这两个概念提供一个健全的理论基础，并作为建立度量空间的抽象理论的模型。这样的一般理论有许多应用，其中一个主要的应用是概率论的数学基础。同时，Lebesgue积分在处理极限运算方面比它的Riemann对应理论有更大的范围和更大的灵活性。

然而，就像黎曼积分一样，从第一原理计算具体的积分是很费力的，而且我们还没有处理特定函数的简单 "配方"。因此，为了将该理论与初级微积分的便捷技术联系起来，我们需要进一步采取两个步骤：证明第一章所述的微积分基本定理，并证明只要后者存在，Lebesgue和黎曼积分就会重合。在此过程中，我们将找到黎曼积分存在的必要和充分条件。

事实上，鉴于命题4.16，基本定理的证明是对连续函数中间值定理的简单应用，因此留给读者：

命题4.22

如果 $f:[a, b]\to\mathbb{R}$ 是连续的，那么 f 就是可整定的，函数 F 由 $F(x) = \int_a^x f\, dm$ 对于$x\in(a, b)$是可微的，导数$F' = f$。

提示：注意，如果$f\in L^1$和A、$B\in\mathcal{M}$是不相交的，那么 $\int_{A\cup B} f\, dm = \int_A f\, dm + \int_B f\, dm$。因此表明，我们可以写出 $F(x+h) - F(x) = \int_x^{x+h} f\, dm$ 对于固定的$[x, x+h]\subset(a, b)$。

我们转而表明，Lebesgue的理论扩展了黎曼的理论：

定理4.23

让f : [a, b] '→R是有界的。

(i) 当且仅当f相对于[a, b]上的Lebesgue度量是a.e.连续的时候，f是黎曼不可逆的。

(ii) 在[a, b]上的黎曼可积分函数对于[a, b]上的Lebes- gue度量来说是可积分的，并且积分是相同的。

证明

我们需要为证明做一点准备，回顾一下符号和一些基本事实。回顾第一章的内容，任何分区

$$P = \{a_i : a = a_0 < a_1 < ... < a_n = b\}$$

的区间[a, b]， $\Delta_i = a_i - a_{i-1}$ (i = 1, 2, ..., n) 和 M_i (response)。m_i)，f在$I_i = [a_{i-1}, a_i]$上的超（resp.inf），诱导出上、下Riemann和 $U_P = \Sigma^n_{i=1} M\,\Delta_{ii}$ 和 $L_P = \Sigma n_{i=1}\ m\,\Delta_{ii}$。但这些只是Lebesgue简单函数的积分 $u_P = \Sigma_n{}_{i=1}$ 密$_{1Ii}$ 和$l_P = \Sigma^n{}_{i=1}\ m\,1_{I_i}$, i，由这类函数的积分定义。

选择一个分区序列(P_n)，使每个P_{n+1} 细化P_n，并且P_n 中最大子区间的长度为0；用u_n 表示u_{P} n，用l_n 表示l_P n，我们有$l_n \le f \le u_n$ for all n。在度量空间（[a, b], $M_{[a,b]}$, m）上应用这个方法，其中$m = m_{[a,b]}$ 表示限制在[a, b]的李比西格量。那么$u = \inf_n u_n$ 和$l = \sup_n l_n$ 是可测量的函数，并且都是

序列是单调的，因为

$$l_1 \le l_2 \le ... \le f \le ... \le u_2 \le u_1 。 \tag{4.3}$$

因此，$u = \lim_n u_n$ ，$l = \lim_n l_n$ (pointwise)，(4.3)中的所有函数在[a, b]上都被$M = \sup \{f(x) : x \in [a, b]\}$ 所约束，它在[a, b]上是可整定的。通过支配性收敛，我们得出结论：

$$\lim_n U_n = \lim_n \int_a^b u_n\, dm = \int_a u\, dm、\qquad \lim_n L_n = \lim_n \int_a^b l_n\, dm = \int_a l\, dm$$

并且极限函数u和l是（Lebesgue-）可整定的。

现在假设x不是任何一个区间的端点(P_n)--它只排除了[a, b]的可数个点

。那么我们有：

如果$u(x)=f(x)=l(x)$，*则f*在x处是连续的。

这从连续性的定义中一下子就可以看出来，因为每个子区间的长度都接近于0，所以f在包含x*的*区间上的变化接近于0，如果f在x处是连续的。

黎曼积分 $\int_a^b f(x)\,dx$被定义为$\lim_n U_n = \int_a^b u\,dm$和$\lim_n L_n = \int_a^b l\,dm$的共同值，只要这些极限是相等的。

为了证明(i)，首先假设f是黎曼不稳定的，因此，上和下限积分重合：$\int_a^b u\,dm = \int_a^b l\,dm$。但$l \le f \le u$，因此$\int_a^b (u - l)\,dm = 0$意味着根据定理4.15，$u = l = f$ a.e. 。因此，f是连续的即通过上述f在x处的连续性的特征，它只排除了另一个无效的分割点集。

反过来说，如果f是a.e.连续的，那么$u=f=l$ a.e.，u和l是可测量的Lebesgue，因此f也是可测量的（注意，这使用了Lebesgue测量的完备性！）。但根据假设，f也是有界的，所以它在$[a, b]$上是可整定的，而且由于积分是a.e.相等的，所以积分是重合的

(但请注意，$\int_a^b f\,dm$表示f的*Lebesgue*积分！）：

$$\int_a^b l\,dm = \int_a^b f\,dm = \int_a^b u\,\mathrm{dm}。 \tag{4.4}$$

由于外积分是相同的，根据定义，f也是黎曼可积分的，这就证明了(i)。

为了证明(ii)，请注意，如果f是黎曼不可控的，(i)表明f是a.e.连续的，因此是可测的，然后(4.4)表明其Lebesgue积分与两个外部积分重合，因此与黎曼积分重合。

例4.6

回顾第1.2节中的以下例子：迪里切特函数在[0，1]上的定义为

$$f(x) = \begin{cases} \frac{1}{n} & 如果x = \frac{m}{n} \in \mathbb{Q} \\ 0 & 如果x \notin \mathbb{Q} \end{cases}$$

是a.e.连续的，因此是黎曼不可控的，它的黎曼积分等于它的勒贝斯格积分，后者是0，因为f在空集Q之外是零。

现在我们已经证明了在评估积分时在前面的例子中提出的未经证实的主张，因为至少对于有界区间上的任何连续函数，初级微积分的技术也给出了有关函数的Lebesgue积分。由于积分在不相连的域上是加法的，所以

这些技术的使用也延伸到了片状连续函数。

例4.7（不恰当的黎曼积分）

处理不恰当的黎曼积分涉及到一个额外的极限运算

我们通过以下方式定义这样一个积分：

$$\int_{-\infty}^{\infty} f(x)\,dx := \lim_{a\to-\infty,b\to\infty} \int_a^b f(x)\,dx$$

只要存在双重极限。(其他 "不当积分 "的情况将在备注4.2中讨论）。

现在假设对于函数f： R R这个不恰当的黎曼积分存在。那么黎曼积分$\int_a^b f(x)\,dx$对于每个有界区间[a, b]都存在，所以f在每个[a, b]上都是等价连续的，因此在R上也是如此。然而，反之是错误的：当n为1时，函数f在[$n, n+1$]上取值为1。

偶数，当n为奇数时为1，是a.e.连续的（因而在R上是可测量的），但显然上述极限未能存在。

更一般地说，不难证明，如果$f \in \mathcal{L}^1$ (R)，那么上述双重极限将始终存在。另一方面，双重极限的存在本身并不能保证$f \in \mathcal{L}^1$而没有进一步的条件：考虑

$$f(x) = \begin{cases} \frac{(-1)^n}{n+1} & \text{如果}x\in[n, n+1]，n\geq 0 \\ 0 & \text{如果}x<0。\end{cases}$$

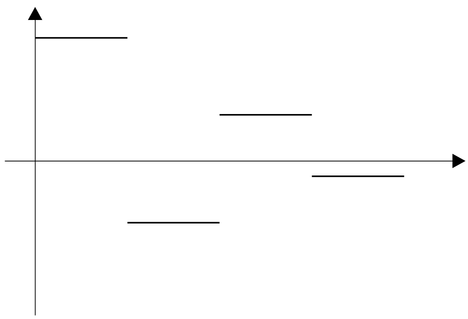

图4.3 f的图表

显然，不恰当的黎曼积分存在、

$$\int_{-\infty}^{\infty} f(x)\,dx = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1}$$

且该系列收敛。然而，$f \notin \mathcal{L}^1$，因为$\int_{\mathbb{R}} |f|\,dm = \sum_{n=0}^{\infty} \frac{1}{n+1} = \infty$。，这分歧。

这就产生了Lebesgue整数的 "绝对 "性质的另一个说明：$f \in \mathcal{L}^1$ iff $|f| \in \mathcal{L}^1$，所以我们不能指望一个积分的 "碎片 "构成一个有条件收敛的数列的有限和。对于

非负的函数

这些问题不会出现；我们有：

定理4.24

如果$f \geq 0$，并且上述不恰当黎曼积分的 f 存在，那么Lebesgue积分$\int_{\mathbb{R}} f\, dm$总是存在的，并且等于不恰当的积分。

证明

要看到这一点，只需注意，序列(f_n)中$f_n = f\, 1_{[-n,n]}$ 单调地增加到f，因此f是Lebesgue可测的。由于f_n 在[-n, n]上是黎曼可积分的，所以积分在那里是重合的，也就是说。

$$\int_{\mathbb{R}} f_n\, dm = \int_{-n}^{n} f(x)\, dx$$

对于每一个n，所以$f_n \in L^1(\mathbb{R})$ 对于所有的n。根据假设，双极限

$$\lim_n \int_{-n}^{n} f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx$$

存在。另一方面

$$\lim_n \int_{\mathbb{R}} f_n\, dm = \int_{\mathbb{R}} f\, dm$$

通过单调收敛，所以$f \in L^1(\mathbb{R})$和

$$\int_{\mathbb{R}} f\, dm = \int_{-\infty}^{\infty} f(x)\, dx$$

根据需要。 □

练习4.10

证明由$f(x) = \frac{\sin x}{x}$ $(x \neq 0)$给出的函数f有一个不正确的R上的黎曼积分，但不在L^1中。

备注4.2

第二种不恰当的黎曼积分是为了处理在有界区间上有渐近线的函数，比如$f(x) = \frac{1}{x}$ on (0, 1).对于在这种情况下，我们可以定义

$$\int_a^b f(x)\, dx = \lim_{\varepsilon \searrow 0} \int_{a+\varepsilon}^{b} f(x)\, dx$$

当极限存在时。(类似的说法也适用于积分的上限）。

4.6 可衡量的 函数的逼近

上一节说明了通过发展勒贝斯格积分所获得的额外 "自由 "的程度：黎曼积分将我们约束在那些不连续的函数上，而我们仍然可以找到那些*不*连续的函数的黎贝斯格积分，例如

作为1_Q。然而，我们可能会问，这种额外的通用性有多真实：例如，我们可以通过连续函数*来近似*任意的f L？ 事实上、

由于连续性是一个局部属性，我们是否可以对任意可测函数进行这样的处理？而这又提供了一个与简单函数的联系，因为每个可测函数都是简单函数的极限。我们可以进一步问，对于一个简单的函数*g*，是否接近于一个给定的可测函数*f*

我们可以选择*g*的范围内的每个元素的反像$g^{-1}(\{a\})$，以是一个区间（这样的*g*通常被称为*阶梯函数*；$g=\sum_n c_n 1_{I_n}$，其中I_n是区间）。我们将首先解决这个问题：

定理4.25

如果*f*是$[a, b]$上的一个有界可测函数，并且$\varepsilon>0$是给定的，那么存在一个阶梯函数*h*，使得$\int_a^b |f-h|\,dm<\varepsilon$。

证明

首先额外假设$f\geq 0$，那么$\int_a^b f\,dm$的定义是很好的，即

$$\sup\{\int_a^b \phi\,dm : 0\leq\phi\leq f \quad \text{简单}\}。$$

由于$f\geq\phi$，我们有$|f-\phi|=f-\phi$，所以我们可以找到一个简单的函数ϕ

令人满意的

$$\int_a^b |f-\phi|\,dm=\int_a^b f\,dm-\int_a^b \phi\,dm<\frac{\varepsilon}{2}.$$

然后，剩下的就是用一个阶梯函数*h*来近似一个在$[a, b]$外消失的任意简单函数*j*。函数*j*的有限范围$\{a_1, a_2, \ldots, a_n\}$划分了$[a, b]$，产生了不相交的可测量集合$E_i=j^{-1}(\{a_i\})$，从而$\bigcup_{i=1}^n E_i=[a, b]$.现在我们用区间来近似每个$E_i$：注意，由于*j*

是简单的，$M = sup\{j(x) : x \in [a, b]\} < \infty$。根据定理2.12，我们可以找到开放集$O_i$，使得$E_i \subset O_i$，并且$m(O_i \setminus E_i) < \frac{\varepsilon}{2nM}$，对于$i \leq n$。由于每个$E_i$具有有限度量，$O_i$也是如此，因此每个$O_i$可以依次被一个不相交的开放区间的*有限*联合：我们知道$O_i = \bigcup_{j=1}^{\infty} I_{ij}$，其中开放区间可以选择不相交的，所以$m(O_i) = \sum_{j=1}^{\infty} m(I_{ij}) < \infty$。由于

系列收敛，我们可以找到k_i，使$m(O_i) - m(\bigcup_{j=1}^{k_i} I_{ij}) < \frac{\varepsilon}{2nM}$。因此$G_i = \bigcup_{j=1}^{k_i} I_{ij}$ 我们有 $\int_a^b |1_{E_i} - 1_{G_i}|\, dm = m(E_i \Delta G_i) < \frac{\varepsilon}{nM}$ 对于每个 $i \le n$. 所以设$h = \sum_{i=1}^{n} a_i 1_{G_i}$，这个步骤函数满足$\int_a^b |j - h|\, dm < \frac{\varepsilon}{2}$。因此，$\int_a^b |f - h|\, dm < \varepsilon$。

对一般f的扩展很清楚：f^+和f^-可以通过阶梯函数h_1和h_2来近似到$\frac{\varepsilon}{2}$内，所以用$h = h_1 - h_2$，我们得到

$$\int_a^b |f - h|\, dm \le \int_a^b |f^+ - h_1|\, dm + \int_a^b |f^- - h_2|\, dm < \varepsilon$$

这就完成了证明。 □

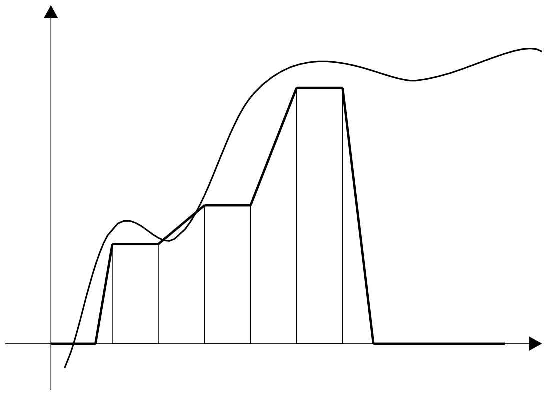

图4.4 用连续函数进行逼近

现在的 "回报 "是直接的：有了上面的f和h，我们可以把区间I_{ij}重新排序为一个单一的有限序列$(J_m)_{m \le n}$，其中$J_m = (c_m, d_m)$，$h = \sum_{m=1}^{n} a_m 1_{J_m}$.我们可以假设，$l(J_m) = (d_m - c_m) > \frac{\varepsilon}{2}$，并近似$1_{J_m}$由一个连续的函数$g_m$，通过设置$g_m = 1$在稍小的区间$(c_m + \varepsilon, d_m - \varepsilon)$和$J_m$外的0，同时在两者之间线性延伸。(见图4.4)。很明显，g_m是连续的，并且$\int_a^b |1_{J_m} - g_m|\, dm < \frac{\varepsilon'}{2}$. 对每个$J_m$重复，并取$\varepsilon' < \frac{\varepsilon}{nK}$，其中$K = \max_{m \le n} |a_m|$，表明连续函数$g = \sum_{m=1}^{n} a_m g_m$，满足$\int_a^b |h - g|\, dm < \frac{\varepsilon}{2}$。

将这个不等式与定理4.25结合起来，可以得到：

定理4.26

鉴于 $f\in L^1$ 和 $\varepsilon>0$，我们可以找到一个连续函数 g，在某个有限区间外消失，从而使 $\int |f-g|\, dm<\varepsilon$。

证明

当f是一个在某个区间$[a, b]$上消失的有界可测函数时，前面的论证已经验证了这一点。对于一个给定的$f \in L^1[a, b]$，我们可以再次假设$f \geq 0$。让$f_n = \min(f, n)$；那么f_n是有界的。由f支配的可测函数，$f_n \to f$，因此，$\int_a^b |f - f_N| \, dm < \frac{\varepsilon}{2}$为某个$N$。现在我们可以找到一个连续的$g$，在一个有限区间外消失、这样，$\int_a^b |f_N - g| \, dm < \frac{\varepsilon}{2}$。因此，$\int_a^b |f - g| \, dm < \varepsilon$。

最后，让$f \in L^1(\mathbb{R})$和$f \geq 0$是给定的。选择足够大的n以保证即$\int_{\{|x| \geq n\}} f \, dm < \frac{\varepsilon}{3}$（我们可以做到这一点，因为$\int_{\mathbb{R}} |f| \, dm$是有限的；命题4.19）、并同时选择一个连续的$g$，$\int_{\{|x| \geq n\}} g \, dm < \frac{\varepsilon}{3}$满足以下条件$\int_{-n}^{n} |f - g| \, dm < \frac{\varepsilon}{3}$。因此，$\int_{\mathbb{R}} |f - g| \, dm < \varepsilon$。 □

著名的*黎曼-勒贝斯格定理*在讨论傅里叶级数时非常有用，它很容易从上述近似定理中推导出来：

Lemma 4.27 (Riemann-Lebesgue)
假设$f \in L^1(\mathbb{R})$。那么序列$s_k = \int_{-\infty}^{\infty} f(x) \sin kx \, dx$ 和 $c_k = \int_{-\infty}^{\infty} f(x) \cos kx \, dx$都收敛到0，因为$k \to \infty$.

证明

我们对(s_k)进行证明，把其他类似的情况留给读者。为了符号的简单性，把$\int$写成$\int_{-\infty}^{\infty}$。变换$x = y + \frac{\pi}{k}$表明：

$$s_k = \int f\left(y + \frac{\pi}{k}\right) \sin(ky + \pi) \, dy = -\int f\left(y + \frac{\pi}{k}\right) \sin(ky) \, dy。$$

因为$|\sin x| \leq 1$、

$$\int \left|f(x) - f\left(x + \frac{\pi}{k}\right)\right| dx \geq \left|\int \left(f(x) - f\left(x + \frac{\pi}{k}\right)\right) \sin kx \, dx\right| = 2|s_k|。$$

因此，只要证明以下几点就足够了$\int |f(x) - f(x + h)| \, dx \to 0$，当$h \to 0$.

这一点最容易做到，即用一个连续的g来逼近f，而这个g是消失的在某个有限区间$[a, b]$之外，并且在给定的情况下，$\int |f - g| \, dm < \frac{\varepsilon}{3}$ $\varepsilon > 0$。对于$|h| < 1$，连续函数$g_h(x) = g(x + h)$然后消失。
$[a - 1, b + 1]$ 和

$$\int |f(x+h) - f(x)|\, dm \leq \int |f(x+h) - g(x+h)|\, dm$$

$$+ \int |g(x+h) - g(x)|\, dm + \int |g(x) - f(x)|\, dm。$$

右边的第一个和最后一个积分都小于$\frac{\varepsilon}{3}$，而的情况下，可以使第二种情况小于 $\frac{\varepsilon}{3(b-a+2)}$ 只要$|h| < \delta$，通过适当的选择 $\delta>0$，因为g是连续的。由于g在$[a-1, b+1]$之外消失，所以第二个积分也小于$\frac{\varepsilon}{3}$。因此，如果$|h|<\delta$，$\int |f(x+h) - f(x)|\, dm<\varepsilon$。这证明了$\lim_{k\to\infty} \int f(x) \sin kx\, dx = 0$. □

4.7 P robability

4.7.1 与概率 分布有关的整合

设X是一个随机变量，其概率分布 为P_X。下面的定理显示了在对X的一个函数进行积分时如何进行变量的改变。换句话说，它显示了如何改变积分中的度量。这是将积分理论应用于概率的基础。我们再次强调，对于我们将在这里应用的定理，只需要σ场的封闭性和度量的可数可加性，因此我们在讨论它们的应用时可以使用概率空间$(\Omega, \mathcal{F}, P)$的抽象表述。

定理4.28

给出一个随机变量X：$\Omega\to\mathbb{R}$、

$$\int_\Omega g(X(\omega))\, dP(\omega) = \int_{\mathbb{R}} g(x)\, dP_X(x). \tag{4.5}$$

证明

我们采用备注4.1中描述的技术。对于指标函数$g = 1_A$，我们在两边都有$P(X\in A)$。然后通过线性关系，我们就有了简单函数的结果。通过简单函数的单调序列与单调收敛相结合对非负可测g进行逼近

一般$g\in L^1$的情况与前面一样，由积分的线性关系得出，使用$g = g^+ - g^-$。 □

该公式在P_X的形式已知的情况下很有用，并允许人们进行明确的计算。

在我们继续讨论这些情况之前，考虑一个非常简单的情况作为公式的说明。假设X是常数，即$X(\omega)\equiv a$ 。那么在(4.5)的左边我们有一个常数函数的积分，根据指标函数积分的一般方案，它等于$g(a)P(\omega)=g(a)$。在右边$P_X=\delta_a$，因此我们有一种计算关于狄拉克计量的积分的方法：$\int g(x)\,d\delta_a = g(a)$。

对于离散的X取值a_i，概率p_i，我们有

$$\int g(X)\,dP = \sum_i g(a_i)p_i$$

这是初级概率论中的一个著名公式（也见第3.5.3节）。在这种情况下，我们有$P_X=\sum_i p_i\delta_{a_i}$，在右边，关于措施的组合的积分是积分的组合：

$$\int g(x)\,dP_X = \sum_i p_i \int g(x)\,d\delta_{a_i}(x)。$$

事实上，这是一个一般的属性。

定理4.29

如果$P_X=\sum_i p_i P_i$，其中P_i是概率度量，$\sum p_i=1$，$p_i\geq 0$，则

$$\int g(x)\,dP_X(x) = \sum_i p_i \int g(x)\,dP_i。$$

证明

方法同上：首先考虑指标函数1_A，要求只是P_X的定义：在左边我们有$P_X(A)$，在右边$\sum_i p_i P_i(A)$。然后通过可加性，我们得到简单函数的公式，最后，近似和使用单调收敛定理就完成了的证明与以前一样。 □

4.7.2 绝对连续度量：例子 密度的例子

形式的措施P

$$A\mapsto P(A)=\int_A f\,dm$$

的非负整数f将被称为*绝对连续*，而函数f将被称为*P相对于Lebesgue度量的密度*，或者简称为*密度*。显然，为了使P成为一个概率，我们必须规定条件

$$\int f\,dm = 1。$$

概率论的学生往往对随机变量的世界有一个过于简化的心理想象，认为随机变量要么是离散的，要么是绝对连续的。这种印象源于许多初级教科书的实际计算方法，这些教科书在介绍概率的时候没有考虑到必要的度量理论背景。我们已经提供了一个简单的例子，表明这是一个错误的二分法（例3.1）。

最简单的密度例子是这样的：让ΩR 是一个 具有有限Lebesgue度的Borel集，并把

$$f(x) = \begin{cases} \frac{1}{m(\Omega)} & 如果x\in\Omega \\ 0 & 否则。\end{cases}$$

我们在上一章已经接触过这种度量，也就是一个特定随机变量的概率分布。我们说，在这种情况下，测量（分布）是*均匀的*。它对应的情况是，随机变量的值均匀地分布在某个集合上，通常是一个区间，例如在随机选择一个数字时（例2.2）。

稍微复杂一点的是所谓的*三角形*分布，其密度形式如图4.5所示。

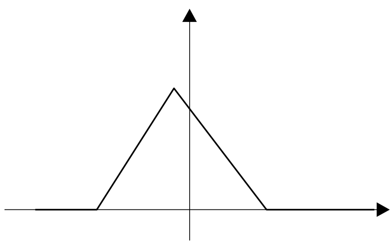

图4.5 三角形分布

最著名的是*高斯*或*正态*密度

$$n(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4.6}$$

这个函数对于$x=\mu$*来说*是对称的，并且在无穷大时消失，即$\lim_{x\to-\infty} n(x)$

$= 0 = \lim_{x\to\infty} n(x)$。

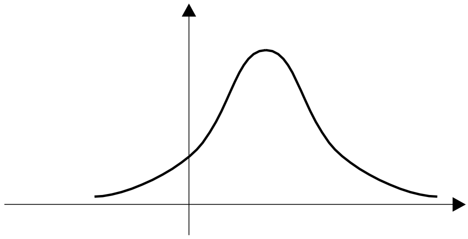

图4.6 高斯分布

练习4.11

表明 $\int_{-\infty}^{\infty} n(x)\, dx = 1$。

提示 首先考虑μ=0，σ=1的情况，然后将一般情况转化为这种情况
。

数字μ的含义将在下文中变得清晰，而σ将在下一章中得到阐述。

另一个广泛使用的例子是*Cauchy*密度：

$$c(x)=\frac{1}{\pi}\frac{1}{1+x^2}.$$

这种密度产生了许多对 "定理 "的反例，这些 "定理 "好得不象是真的。

练习4.12

表明 $\int_{-\infty}^{\infty} c(x)\, dx = 1$。

*指数*密度由以下公式给出

$$f(x) = \begin{cases} c^{-\lambda x} & \text{如果}x\geq 0 \\ 0 & \text{否则。} \end{cases}$$

练习4.13

寻找常数c，使f成为概率分布的密度。

*伽马*分布实际上是一个大的分布系列，由参数t>0来标示。它包含指数分布，作为t=1的特殊情况。它的密度被定义为

$$f(x)=\begin{cases}\frac{1}{\Gamma(t)}\lambda^t x^{t-1}e^{-\lambda x} & \text{如果}x\geq 0\\ 0 & \text{否则}\end{cases}$$

其中伽马函数Γ（t）$=\int_0^\infty x^{t-1}e^{-x}\,dx$。

伽马分布包含另一个广泛使用的分布，作为特例：当$\lambda=\frac{1}{2}$，$t=\frac{d}{2}$时，从密度f得到的分布。对于某些$d\in\mathbb{N}$表示为$\chi^2(d)$，称为*卡方分布，其特征为d的自由度*。

与密度相对应的（累积）*分布函数*是这样的由

$$F(y)=\int_{-\infty}^{y} f(x)\,dx$$

如果f是连续的，那么F是可微的，根据微积分基础定理，$F'(x)=f(x)$（见命题4.22）。如果这个关系在可整定的f中成立，我们就说f是绝对连续的，那么f就是密度的

是由F诱导的概率测量。下面这个由Lebesgue提出的例子表明，F的连续性对于密度的存在是不够的。

例4.8

回顾第20页上定义的Lebesgue函数F。我们有F（y）=0，对于$y\leq 0$、$F(y)=1$ 对于$y\geq 1$, $F(y)=\frac{1}{2}$ 对于$y\in[\frac{1}{3},\frac{2}{3})$, $F(y)=\frac{1}{4}$ 对于$y\in[\frac{1}{9},\frac{2}{9})$, $F(y)=\frac{3}{4}$ 对于$y\in[\frac{7}{9},\frac{8}{9})$，以此类推。函数$F$在删除的区间上是常数在构建康托尔集的过程中。

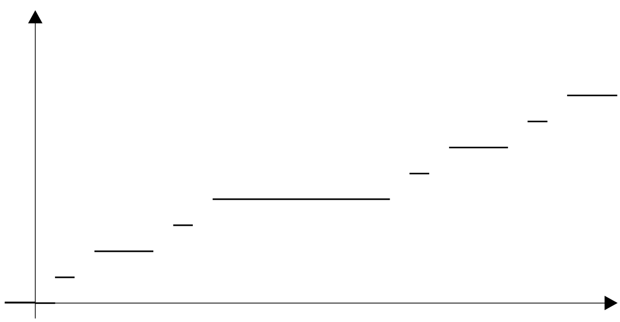

图4.7 勒贝斯格的函数

它几乎处处可微，导数为零。所以F不可能是绝对连续的，因为那样的话F几乎到处都是零，但另一方面，它的积分是1。

我们现在定义一个随机变量的（累积）分布函数
X：$\Omega\to$R，如上所述，（Ω，F，P）是一个给定的概率空间：

$$F_X(y) = P(\{\omega : X(\omega) \le y\}) = P_X((-\infty, y]) 。$$

命题4.30

(i) F_X是非递减的（$y_1 \le y_2$意味着$F_X(y_1) \le F_X(y_2)$）、
(ii) $\lim_{y\to\infty} F_X(y) = 1$, $\lim_{y\to-\infty} F_X(y) = 0$、
(iii) F_X是右连续的（如果$y\to y_0$，$y\ge y_0$，那么$F_X(y)\to F(y_0)$）。

练习4.14

证明F_X是连续的，当且仅当$P_X(\{y\}) = 0$对所有y。

练习4.15

找到F_X，用于

(a) 一个常数随机变量X，$X(\omega)=a$，对于所有的ω

(b) $X : [0, 1] \to$ R，由$X(\omega)=\min\{\omega, 1-\omega\}$给出（到区间[0, 1]的最近端点的距离）

(c) $X : [0, 1]^2 \to$ R，到正方形$[0, 1]^2$最近的边的距离。

我们在R的子集上做概率[n]，作为样本空间，这一点原来是没有限制性的。事实上，区间[0，1]是足够的，正如下面的斯科罗霍德表示定理所显示的那样。

定理4.31

如果一个函数F : R → [0, 1]满足命题4.30的条件(i)-(iii)，那么有一个定义在概率空间([0, 1], B, $m_{[0,1]}$)上的随机变量$X : [0, 1] \to$ R，这样$F = F_X$。

证明

我们写道，对于$\omega \in [0，1]$、

$$X^{+}(\omega) = inf\{x : F(x) > \omega\}、\quad X^{-}(\omega) = sup\{x : F(x) < \omega\}。$$

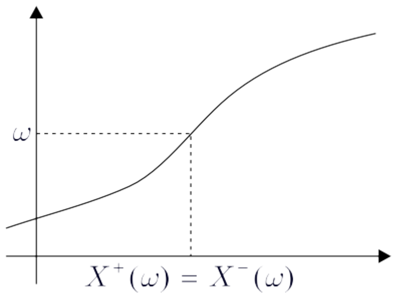

图4.8 构建X^{-}；连续性点

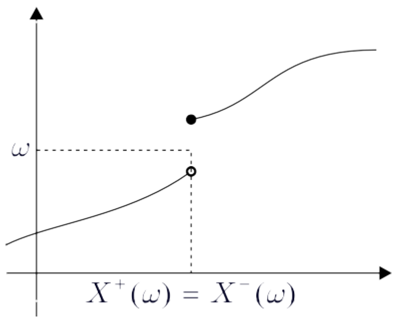

图4.9 构建X^{-}；不连续点

图4.8、4.9和4.10中说明了三种可能的情况。我们证明$F_{X^{-}} = F$，为此我们必须证明F（y）$= m$（$\{\omega : X^{-}(\omega) \leq y\}$）。集合$\{\omega : X^{-}(\omega) \leq y\}$是一个左端点为0的区间。如果我们证明其右端点为$F(y)$，即如果$X^{-}(\omega) \leq y$等同于$\omega \leq F(y)$，我们就完成了。

假设$\omega \leq F$（y）。那么

$$\{x : F(x) < \omega\} \subset \{x : F(x) < F(y)\} \subset \{x : x \leq y\}。$$

(最后一个包含在F的单调性中），因此$X^{-}(\omega) = sup\{x : F(x) < \omega\} \leq y$。

假设$X^{-}(\omega) \leq y$。根据单调性，$F(X^{-}(\omega)) \leq F(y)$。根据$F$的右连续性，$\omega \leq F$（$X^{-}(\omega)$）（如果$\omega > F$（$X^{-}(\omega)$），则有$x_0 > X^{-}(\omega)$）。

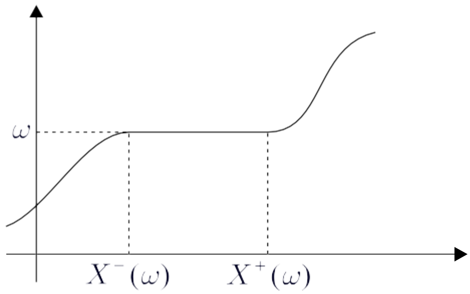

图4.10 X的构造⁻；"平 "片

这样，$F(X^-(\omega)) < F(x_0) < \omega$，这是不可能的，因为$x_0$是在集合中，其最高值被取来得到$X^-(\omega)$） 所以$\omega \le F(y)$。

为了今后的使用，我们还表明，$F_X + = F$。只需看到$m(\{\omega$：
$X^-(\omega) < X^+(\omega)\}) = 0$（这在直觉上很清楚，因为只有在以下情况下才可能发生这种情况
F的图形是 "平的"，而且有无数的值对应于
的 "平 "件，它们的李贝斯格度量为零）。更严格地说、

$$\{\omega : X^-(\omega) < X^+(\omega)\} = \bigcup_{q\in Q} \{\omega : X^-(\omega) \le q < X^+(\omega)\}$$

和$m(\{\omega : X^-(\omega) \le q < X^+(\omega)\}) = m(\{\omega : X^-(\omega) \le q\}.\backslash \{\omega : X^+(\omega) \le q$
$q\})$ $=F$（q） $-F$（q） $=0$。 □

下面的定理为计算相对于绝对连续分布的积分提供了一种强有力的方法。这个结果对一般的度量来说是成立的，但我们对随机变量的概率分布进行表述，以便不至于使符号过重或混乱。

定理4.32

^{n}如果定义在R上的P_X是绝对连续的，密度为f_X，g： $\mathbb{R}^n \to \mathbb{R}$对于$P_X$是可整定的，那么

$$\int_{R^n} g(x)\, dP_X(x) = \int_{R^n} f_X(x) g(x)\, dx。$$

证明

对于一个指标函数$g(x)=1_A(x)$，我们在左边有$P_X(A)$，等于

$\int_A f_X(x)\,dx$ 由P的形式，因此等于$\int_{R^n} 1_A(x) f_X(x)\,dx$、即右手边。通过线性扩展到简单函数，并扩展到一般的通过极限通道对可整定的g*进行处理*是常规的。 □

推论4.33

在前一个定理的情况下，我们有

$$\int_\Omega g(X)\,dP = \int_{R^n} f_X(x)\,g(x)\,dx。$$

证明

这是上述定理和定理4.28的一个直接结果。 □

在这一节的最后，我们用一个给定密度的随机变量的函数的密度公式来总结。假设f_X是已知的，我们想找到$Y = g(X)$的密度。

定理4.34

如果g： R → R是增加的和可微的（因此是可逆的），那么

$$f_{g(X)}(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$$

染料

证明

考虑一下分布函数：

$$F_{g(X)}(y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))。$$

对y进行微分，即可得到结果。 □

备注4.3

如果g是递减的，类似的结果也成立。与上面的论证相同，可以得到

$$f_{g(X)}(y) = \text{-}f_X \quad (g^{-1}(y)) \frac{d}{} g^{-1}(y)$$

染料

例4.9

如果X具有*标准*正态分布

$$n(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

(即(4.6)中的μ=0和σ=1)，那么$Y=\mu+\sigma X$的密度由(4.6)给出。这从定理4.34中立即得出：$g^{-1}(y)=\frac{y-\mu}{\sigma}$；其导数为等于$\frac{1}{\sigma}$。

练习4.16

找到$Y=X$的密度^3，其中$f_X=1_{[0,1]}$。

4.7.3 随机 变量的期望值

如果X是一个定义在概率空间（Ω， ， P）上的随机变量，F那么我们引入以下符号：

$$E(X)=\int_\Omega X\,dP$$

我们称这个抽象的积分为X的*数学期望值*。

利用上一节的结果，我们立即得到以下公式：可以用概率分布来计算期望值：

$$E(X)=\int_{-\infty}^{\infty}x\,dP_X(x)、$$

而对于绝对连续的X，我们有

$$E(X)=\int_{-\infty}^{\infty}xf_X(x)\,dx。$$

例4.10

假设$P_X=\frac{1}{2}P_1+\frac{1}{2}P_2$，其中$P_1=\delta_a$，$P_2$有一个密度$f_2$。那么

$$E(X)=\frac{1}{2}a+\frac{1}{2}\int xf(x)\,dx。$$

因此，回到例3.1，我们可以计算那里考虑的随机变量的期望值：

$$e(x)=\frac{1}{2}0+\frac{1}{2}\frac{1}{25}\int_0^{25}x\,dx=6.25$$

练习4.17

找出以下的期望值

(a) 一个常数随机变量X，$X(\omega)=a$，对于所有的ω

(b) $X:[0,1]\to\mathbb{R}$，由$X(\omega)=\min\{\omega,1-\omega\}$给出（到区间[0, 1]的最近端点的距离）

(c) $X:[0,1]^2\to\mathbb{R}$，到正方形$[0,1]^2$最近的边的距离。

练习4.18

求一个随机变量的数学期望值，该变量有

(a) 在区间$[a,b]$上均匀分布、

(b) 三角形分布、

(c) 指数分布。

4.7.4 特征 功能

在下面的内容中，我们将需要一些复数函数的积分。该理论是对实数情况的直接扩展。

设$Z=X+iY$，其中X，Y是实值随机变量，并定义

$$\int Z\,dP=\int X\,dP+i\int Y\,dP。$$

显然，积分的线性和支配收敛定理在复杂情况下是成立的。另一个保持不变的重要关系是：

$$\left|\int Z\,dP\right|\le\int|Z|\,dP。$$

要看到这一点，请考虑以下的极坐标分解 $\int Z\,dP=\left|\int Z\,dP\right|e^{-i\theta}$。然后，以$\Re(z)$为复数$z$的实部、$\left|\int Z\,dP\right|=e^{i\theta}\int Z\,dP$

$\int e^{i\vartheta}Z\,dP$是实数，因此等于$\int\Re(e^{i\theta}Z)\,dP$，但$\Re(e^{i\vartheta}Z)\le|e^{i\vartheta}Z|=|Z|$而我们已经完

我们希望整合的函数是exp{itX}，其中X是一个实数随机变量，$t\in\mathbb{R}$

。

$$\int \exp\{itX\}\, dP = \int \cos(tX)\, dP + i \int \sin(tX)\, dP,$$

根据$x' \to \exp\{itx\}$的有界性，它始终存在。

定义4.5

对于一个随机变量X，我们写成

$$\phi_X(t) = E(e^{itX})$$

对于$t\in R$。我们称j_X为X的*特征函数*。

要计算ϕ_X，只需知道X的分布即可：

$$\phi_X(t) = \int e^{itx}\, dP_X(x)$$

而在绝对连续的情况下

$$\phi_X(t) = \int e^{itx} f_X(x)\, dx。$$

下面给出了特征函数的一些基本属性。其他属性将在第6章和第8章中进行探讨。

定理4.35

函数ϕ_X，满足以下条件

(i) $\phi_X(0) = 1$, $|\phi_X(t)| \le 1$、

(ii) $\phi_{aX+b}(t) = e^{itb} j_X(at)$.

证明

(i) 0处的值是1，因为常数函数的期望值是它的值。该估计由命题4.16（iii）得出：$\left|\int e^{itx}\, dP_X(x)\right| \le \int |e^{itx}|\, dP_X(x) = 1$。

(ii) 这里我们使用期望值的线性：

$$j_{aX+b}(t) = E(e^{it(aX+b)}) = E(e^{itaX} e^{itb}) = e^{itb}E(e^{i(ta)X}) = e^{itb} j_X(ta)，$$

如要求。 □

练习4.19

求一个随机变量的特征函数，该变量有

(a) 在区间$[a, b]$上均匀分布、

(b) 指数分布、

(c) 高斯分布。

4.7.5 在数学上的应用 金融

考虑欧洲类型的衍生证券，即形式为$f(S(N))$的随机变量，其中$S(n)$, $n = 1, . . .$，N，是基础证券的价格，为了简单起见，我们称之为股票。(或者我们写$f(S(T))$，其中标的证券在连续时间t [0, T]中描述，价格为$S(t)$。)金融学中的一个关键问题是找到这种证券的价格Y（0）。在此，我们假设读者熟悉以下事实，该事实对于由概率空间和代表股票价格的随机变量组成的某些特定模型是真实的：

$$Y(0)=\exp\{-rT\}\mathrm{E}(f(S(T)))。\tag{4.7}$$

其中r是连续复利的无风险利率。这将在第7.4.3节中详细解释，但在这里我们只想利用本章中收集的经验从这个公式中得出一些结论。

特别是，考虑到欧洲看涨期权（$f(x) = (x K)^+$）和看跌期权（$f(x) = (K x)^+$）的报酬函数的形式，我们分别有以下看涨和看跌期权价值的一般公式：

$$C = \exp\{-rT\}\mathrm{E}(S(T) - K)^+、$$
$$P = \exp\{-rT\}\mathrm{E}(K - S(T))^+。$$

在不依赖任何特定模型的情况下，我们可以证明以下关系，即所谓的调用-投入奇偶性（例如，见[4]）：

$$S(0) = C - P + K\exp\{-rT\}。\tag{4.8}$$

命题4.36

调用-投入奇偶性的右手边与K无关。

备注4.4

这个命题使我们能够提出一个有趣的看法，它是金融学中一个著名结果的版本，即米勒-莫迪利安尼定理，该定理说，一个公司的价值并不取决于它的融资方式。让我们简单回顾一下，一个公司的价值是股权（以股票为代

表）和债务的总和，所以该定理说，债务水平对公司价值没有影响。

公司的价值。假设该公司借入$K\exp\{-rT\}$的利率为

等于r，并且它必须在T时间偿还K的金额。如果它失败了，公司就会破产。所以控制公司的股东可以通过支付K来"买回 "公司。只有当公司的价值$S(T)$超过K时，这才有意义。

C.因此，债务的价值是$K \exp rT P$,{¬小于}K的现值，它反映了总和K可能无法全部收回的风险。

我们现在评估期望值，在两个最广泛使用的模型中建立一般公式（4.7）的明确形式。

首先考虑第2.6.3节中介绍的二项式模型。假设概率空间有一个由概率决定的度量，即

$p = \frac{R-D}{U-D}$ 为单步上升运动，其中R=exp{rh}，h为一个步骤的长度。(这个概率被称为风险中性；观察一下

$E(\eta) = R$.)为了确保$0 \leq p \leq 1$，我们自始至终假定$D \leq R \leq U$。

命题4.37

在二项式模型中，行使时间为$T=hN$的看涨期权的价格C

由Cox-Ross-Rubinstein公式给出

$$C = S(0)\Psi(A, N, pU e^{-rT}) - Ke^{-rT}\Psi(A, N, p)$$

其中$\Psi(A, N, p) = \sum_{k=A}^{N} \binom{N}{k} p^k (1-p)^{N-k}$，而$A$是第一个整数$k$，所以即$S(0)U^k D^{N-k} > K$。

在著名的连续时间Black-Scholes模型中，股票价格在时间T的形式是

$$S(T) = S(0)\exp\{(r - \frac{\sigma^2}{2})T + \sigma w(T)\}、$$

其中r是无风险利率，$\sigma>0$，$w(T)$是一个高斯分布的随机变量，均值为0，方差为T。

命题4.38

我们有以下关于C的Black-Scholes公式：

$$C = S(0)N(d_1) - Ke^{-rT}N(d_2)。$$

其中

$$d_1 = \frac{\ln \frac{S(0)}{K^{-rT}} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \quad d_2 = \frac{\ln \frac{S(0)}{K^{-rT}} - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

练习4.20

找到看跌期权的公式。

4.8 命题的证明

证明（命题4.1的证明）

让$f = \sum c_i 1_{A_i}$。我们必须证明

$$\sum c_i\, m(A_i \cap E) = \sup Y(E, f).$$

首先，我们可以在Y（E，f）的定义中取$j=f$，所以左边的数字（$\sum c_i\, m(A_i \cap E)$）属于Y（E，f），所以

$$\sum c_i\, m(A_i \cap E) \leq \sup Y(E, f).$$

反之，取任何$a \in Y$（E，f）。所以

$$a = \int_E \psi\, dm = \sum d_j\, m(E \cap B_j)$$

对于一些简单的$\psi \leq f$。现在

$$a = \sum_j \sum_i d_j\, m(E \cap B_j \cap A_i)$$

根据度量的特性（A_i构成R的一个分区）。对于$x \in B_j \cap A_i$，f（x）$=c_i$，ψ（x）$=d_j$，所以$d_j \leq c_i$（如果只有$B_j \cap A_i \neq \varnothing$）。因此

$$a \leq \sum_i \sum_j c_i\, m(E \cap B_j \cap A_i) = \sum_i c_i\, m(E \cap A_i)$$

因为B_j分区R。 □

证明（命题4.5的证明）

让$A = \{x : f(x) \leq g(x)\}$，则$A^c$为空，且$1_A f_A \leq g_A 1_A$.所以$\int f_A\, dm \leq \int g 1_A\, dm$，根据定理4.3。但由于$A^c$是空的、$\int f 1_{A^c}\, dm = 0 = \int g 1_{A^c}\, dm$。所以根据同一定理的(v)项

$$\int_{\mathbb{R}} f\, dm = \int_A f\, dm + \int_{A^c} f\, dm = \int_A f\, dm$$

$$\leq \int_A g\, dm = \int_A g\, dm + \int_{A^c} g\, dm = \int_{\mathbb{R}} g\, dm$$ □

$_{c} g\ dm = \int$

$g\,\mathrm{dm}$。

证明（命题4.6的证明）

如果f^+和f^-都是可测量的，那么对f来说也是如此，因为$f = f^+ - f^-$。反过来说，如果a为$0, \infty$ $(f)^{+-1}([a,\infty)) = \mathbb{R}$，$(f)^{+-1}([a,\infty)) = f^{-1}([a,\infty))$；在每种情况下都是可测集。类似地，对于$f^-$。 □

证明（命题4.10的证明）

把

$$s_n = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} 1_{f^{-1}([\frac{k}{2^n},\frac{k+1}{2^n}))}$$

因为集合$A_k = f^{-1}([\frac{k}{2^n},\frac{k+1}{2^n}))$是可测量的。的

序列增加，因为如果我们取$n+1$，那么每个A_k，被分成两半，而对于和的每个组成部分，有两个新的组成部分。这两个分数的值分别等于或大于旧的分数。收敛性成立，因为对于每一个x，值$s_n(x)$将是一个形式为$\frac{k}{2^n}$，近似于$f(x)$的分数。图4.2说明了上述论点。 □

证明（命题4.11的证明）

如果$f \le g$，则 $f^+ \le g^+$不过$f^- \ge g^-$.这些不等式意味着 $\int f^+ \mathrm{d}m \le \int g^+ dm$ 和 $\int g^- dm \le \int f^- dm$。加法和重排得到的结果。 □

证明（命题4.13的证明）

对于简单的函数$f = \sum a_i 1_{A_i}$来说，这个要求是明显的，它只是初级代数。对于非负的可测量的f，和正的c，取$s_n \nearrow f$，并注意$cs_n \nearrow cf$，所以

$$\int cf\, dm = \lim \int cs_n\, dm = \lim c \int s_n\, dm = c \lim \int s_n\, dm = c \int f\, \mathrm{d}m。$$

最后，对于任何f和c，我们采用通常的技巧，引入正负两部分。 □

证明（命题4.16的证明）

(i) 假设 $f(x) = \infty$ 为 $x \in A$ 与 $m(A) > 0$.那么简单的职能 $s_n = n 1_A$ 满足 $s_n \le f$，但 $\int s_n\, dm = nm(A)$，这里的最高值是

∞。因此，$\int f\,dm = \infty$ - 矛盾。

(ii) 简单函数$s(x)$=c1_A，$c=\inf_A f$有积分$\inf_A f m(A)$，并且满足 $s \le f$，这证明了第一个不等式。把$t(x) = d1_A$与$d = \sup_A f$，$f \le t$，所以$\int f\,dm \le \int t\,dm$，这是第二个不等式。

(iii) 请注意，$-|f| \le f \le |f|$ 因此$-\int |f|\,dm \le \int f\,dm \le \int |f|\,dm$，我们都完成了

(iv) 设$E_n = f^{-1}([\frac{1}{n}, \infty))$，和$E = \bigcup_{n=1} E_n$。$E_i$ 的集合是可测量的E也是如此。函数$s_n = \frac{1}{n} 1_{E_n}$是一个简单函数，$s \le f$。因此$\int s_n\,dm \le \int f\,dm = 0$，所以$\int s_n\,dm = 0$，因此$\frac{1}{n} m(E_n) = 0$。最后，$m(E_n) = 0$ 因为$E_n \subset E_{n+1}$，所以$m(E) = \lim m(E_n) = 0$。但是$E = \{x : f(x) > 0\}$，所以f在空集E之外是零。 □

证明（命题4.19的证明）

如果$n \to \infty$，那么$1_{[-n,n]} \to 1$，因此$g_n = f\,1_{[-n,n]} \to f$。收敛性是受控的：$g_n \le |f|$，根据受控收敛定理，我们有$\int |f - g_n|\,dm \to 0$。同样地，$h_n = \min(f, n) \to f$，因为$n \to \infty$，$h_n \le |f|$ 所以$\int |f - h_n|\,dm \to 0$. □

证明（命题4.20的证明）

利用$\int (f+g)\,dm = \int f\,dm + \int g\,dm$，我们可以很容易地得到（通过归纳）。

$$\int \sum_{k=1}^{n} f_k\,dm = \sum_{k=1}^{n} \int f_k\,dm$$

对于任何n。序列$\sum_{k=1}^{n} f_k$是递增的（$f_k \ge 0$），并收敛于$\sum_{k=1}^{\infty} f_k$。所以单调收敛定理得出

$$\int \sum_{k=1}^{\infty} f_k\,dm = \lim_{n\to\infty} \int \sum_{k=1}^{n} f_k\,dm = \lim_{n\to\infty} \sum_{k=1}^{n} \int f_k\,dm = \sum_{k=1}^{\infty} \int f_k\,dm$$

根据需要。 □

证明（命题4.22的证明）

连续函数是可测量的，f在$[a, b]$上是有界的，因此f $\in$
$L^1[a, b]$。固定$a < x < x + h < b$，那么$F(x \pm h)$ $F(x)$ $\int_x^{x+h} f\,dm$，因为区间$[a, x]$和$(x, x + h]$是不相连的，所以积分对上端点是加法的。根据均值特性，右图的值为
手部积分包含在区间$[Ah, Bh]$中，其中$A = \inf\{f(t) : t \in$

$[x, x+h]\}$和$B = sup\{f(t) : t \in [x, x+h]\}$。两个极值都达到了，因为$f$是连续的，所以我们可以在$[x, x+h]$中找到$t_1, t_2$，$A = f(t_1), B = f(t_2)$。因此

$$f(t_1) \leq \frac{1}{h}\int_x^{x+h} f\,dm \leq f(t_2)。$$

中间值定理规定 $\vartheta \in [0,1]$，这样，$f(x+\vartheta h) = \frac{1}{h}\int_x^{x+h} f\,dm = \frac{F(x+h)-F(x)}{h}$。假设$h\to 0$，$f$的连续性保证了 $F'(x) = f$ □

证明（命题4.30的证明）

(i) 如果 $y_1 \leq y_2$，那么$\{\omega : X(\omega) \leq y_1\} \subset \{\omega : X(\omega) \leq y_2\}$，并且根据措施的单调性

$$F_X(y_1) = P(\{\omega : X(\omega) \leq y_1\}) \leq P(\{\omega : X(\omega) \leq y_2\}) = F_X(y_2)。$$

(ii) 让$n \to \infty$；那么$\bigcup_n\{\omega : X(\omega) \leq n\} = \Omega$（集合增加）。因此根据定理2.13（i），$P$（$\{\omega$：$X$（$\omega$）$\leq n\}$）$\to P$（$\omega$）$=1$，所以$\lim_{y\to\infty} F_X(y) = 1$。对于第二个要求，考虑$F_X(-n) = P(\{\omega : X(\omega) \leq -n\})$ 并注意到即$\lim_{y\to-\infty} F_X(y) = P(\bigcap_n\{\omega : X(\omega) \leq -n\}) = P(\emptyset) = 0$.

(iii) 这直接源于定理2.13(ii)，$A_n = \{\omega : X(\omega) \leq y_n\}$，$y_n \searrow y$，因为$F_X(y) = P(\bigcap_n\{\omega : X(\omega) \leq y_n\})$。 □

证明（命题4.36的证明）

插入看涨-看跌平价期权价格的公式，我们有

$$\begin{aligned} S(0) &= \exp\{-rT\}\int_\Omega (S(T)-K)^+\,dP - \int_\Omega (K-S(T))^+\,dP + K \\ &= \exp\{-rT\}\int_{\{S(T)\geq K\}} (S(T)-K)dP - \int_{\{S(T)<K\}} (K-S(T))dP + K \\ &= \exp\{-rT\}\int_\Omega S(T)dP、 \end{aligned}$$

如所称，它与K无关。 □

证明（命题4.37的证明）

一般公式$C = e^{-rT} E(S(N)-K)^+$，其中$S(N)$有二项分布，可得出

$$C = e^{-rT}\sum_{k=0}^{N}\binom{N}{k} p^k (1-q)^{N-k} (S(0)U^k D^{N-k} - K)^+$$

$$= e^{-rT}\sum_{k=A}^{N}\binom{N}{k} p^k (1-q)^{N-k} (S(0)U^k D^{N-k} - K)$$

我们可以重写如下：注意，如果我们设定$q=PU\ e^{-rT}$，那么$1-q=(1-p)U\ e^{-rT}$，因此，$0\leq q\leq 1$和

$$C = S(0)\sum_{k=A}^{N}\binom{N}{k} q^k (1-q)^{N-k} - Ke^{-rT}\sum_{k=A}^{N}\binom{N}{k} p^k (1-p)^{N-k}$$

我们将其简洁地写为

$$C = S(0)\Psi(A, N, pU e^{-rT}) - Ke^{-rT}\Psi(A, N, p) \qquad (4.9)$$

其中$\Psi(A, N, p) = \sum_{k=A}^{N}\binom{N}{k} p^k (1-p)^{N-k}$是互补二项式的分布函数。 □

证明（命题4.38的证明）

为了计算期望值$E(S(T)-K)^+$，我们采用高斯分布的密度，因此

$$C = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}y^2}\left(S(0)e^{-\frac{1}{2}\sigma^2 T} e^{y\sigma\sqrt{T}} - Ke^{-rT}\right)^+ dy$$

整合后的结果就是满足以下条件的y值

$$S(0)e^{-\frac{1}{2}\sigma^2 T} e^{y\sigma\sqrt{T}} - Ke^{-rT} \geq 0$$，否

则积分为零。求解y可以得到

$$y \geq d = \frac{\ln\frac{Ke^{-rT}}{S(0)} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

对于这些y，我们可以放弃正数部分，采用线性。右边的第一项的形式是

$$S(0)\int_d^{+\infty} e^{y\sigma\sqrt{T}-\frac{1}{2}\sigma^2 T} g(y)dy$$

其中g表示高斯密度，均值为零，标准变量为单位。将$z = y - \sigma\sqrt{T}$代入，得到

$$S(0)\int_{d-\sigma\sqrt{T}}^{+\infty} n(z)dz = S(0)N(-d + \sigma\sqrt{T})$$

其中N是高斯（正态）分布的累积分布函数。第二个项的形式是

$$-Ke^{-rT}\int_{d}^{+\infty} n(y)dy = -Ke^{-rT}N(-d)$$

所以写出$d_1 = -d + \sigma\sqrt{T}$，$d_2 = -d$，我们就完成了。 □

5

可整定函数的空间

到目前为止，我们把度量空间(R, $\mathcal{M}$, m)的点，以及更一般的，任何抽象概率空间(Ω, $\mathcal{F}$, P)的点，当作基本对象、

并将可测或可整函数视为与实数有关的映射。我们现在稍微改变一下我们的观点，把可积分函数当作函数空间中的一个 "点"，或者更准确地说，当作*规范化矢量空间*的一个元素。为此，我们需要在我们处理的函数空间上有一些额外的结构，而且我们需要接受这样一个事实，即度量和积分不能区分几乎到处相等的函数。

我们需要的额外结构是在给定的可积分函数之间定义一个*距离*的概念（即一个*度量*）--通过与熟悉的R中矢量的欧几里得距离的类比^n，我们将得到两个函数之间的距离。
职能的长度，或*规范*，他们的差异 - 因此，利用向量
职能空间的空间结构。我们将能够以各种方式做到这一点，每一种方式都有其自身的优势--与R^n 中的情况不同，所有的规范都变成了等价的，我们现在得到的是真正不同的距离
职能。

值得注意的是，我们将要讨论的函数空间都是*无限维的*向量空间：通过考虑C定义在[a, b]上的实值连续函数的向量空间（[a, b], R）和非-的向量空间，已经可以看出这一点。
在此基础上，一个n度的多项式函数不能被表示为一个线性的

低度多项式的组合。

最后，回顾一下，在介绍特征函数时，在最后的

在上一章中，我们需要将可整性的概念扩展到复值函数。我们注意到，对于$f = u + iv$，由$\int_E f\, dm = \int_E u\, dm + i\int_E v\, dm$定义的积分是线性的，并且不等式$|\int_E f\, dm| \leq \int_E |f|\, dm$仍然有效。当考虑到可衡量的函数$f : E \to C$ 在本章定义可整定函数的适当空间时，这个不等式将表明$\int_E f\, dm \in C$是定义明确的。

下面证明的结果延伸到复值函数的情况、
除非另有说明。当希望强调我们在特定的应用或例子中处理的是复值函数时，我们将使用$f \in L^1(E, C)$这样的符号来表示。复值函数
当我们考虑重要的空间$L^2(E)$的时候，会有特别的兴趣。
'平方可积分'函数。

5.1 空间L^1

首先，我们回顾一下一个集合的各点之间 "距离 "的一般概念的定义：

定义5.1

设X是任何集合。如果函数$d : X \times X \to R$是X上的一个*度量*（并且（X，d）被称为一个*度量空间*），它满足：：

(i) 对于所有的x，$y \in X$，$d(x, y) \geq 0$、

(ii) $d(x, y) = 0$，当且仅当$x = y$、

(iii) $d(y, x) = d(x, y)$ 对于所有$x, y \in X$、

(iv) $d(x, z) \leq d(x, y) + d(y, z)$ 对于所有x、y、$z \in X$。

最后一个属性被称为*三角形不等式*，它概括了众所周知的R^n 中的向量的该名称的不等式。当X是一个向量空间时（在我们几乎所有的例子中都是如此），有一个非常简单的方法
通过将两个向量之间的距离定义为它们之差的 "长度 "来生成一个度量。为此我们需要一个进一步的定义：

定义5.2

设X是R（或C）上的一个向量空间。函数$x \mapsto \|x\|$从X进入R是X上的*准则*，如果它满足：

(i) 对于所有的$x \in X$，$\|x\| \geq 0$、

(ii) $\|x\| = 0$，当且仅当 $x = 0$、

(iii) $\|\alpha x\| = |\alpha|\,\|x\|$ 对于所有 $\alpha\in\mathbb{R}$（或C），$x\in X$、

(iv) $\|x+y\| \le \|x\| + \|y\|$，对于所有 x，$y\in X$。

显然，一个规范 $x\mapsto\|x\|$ 在 X 上通过设置 $d(x, y) = \|x-y\|$ 诱导出一个度量。一旦我们观察到 $\|x-z\| = \|(x-y)+(y-z)\|$ 并应用(iv)，三角不等式就会出现。

我们自然希望用积分来定义 $\mathcal{L}^1(E)$ 中函数之间的距离概念，对于可测量的 $E\subset\mathbb{R}$。

矢量 f 的'长度'为 $\int_E |f|\, dm$，因此遇到了麻烦，因为 $\mathcal{L}^1(E)$ 的非零元素有可能有'零长度'。

所采用的解决方案是通过在 $\mathcal{L}^1(E)$ 上定义一个等价关系，并为所产生的函数等价类定义长度函数，而不是为函数本身定义长度函数，以此来识别等价的函数。

因此，我们定义

$$L^1(E) = \mathcal{L}^1(E)/\equiv$$

其中，等价关系由以下方式给出：

当且仅当 $f(x)=g(x)$ 时，几乎所有的 $x\in E$，$f\equiv g$。

(也就是说，$\{x\in E : f(x) \neq g(x)\}$ 为空)。

将 $[f]$ 写为包含函数 $f\in\mathcal{L}^1(E)$ 的等价类。因此

$h\in[f]$ iff $h(x) = f(x)$ a.e.

练习5.1

检查 $\equiv$ 是 $\mathcal{L}^1(E)$ 上的一个等价关系。

我们现在证明 $L^1(E)$ 是一个矢量空间，因为根据定理4.14，$\mathcal{L}^1$ 是一个矢量空间。然而，这要求我们解释我们所说的等价类的线性组合是什么意思。这对于任何等价关系都可以很普遍地做到；然而，我们将专注于在我们的特定情况下所需要的：将 $[f]+[g]$ 定义为 $f+g$ 的 $[f+g]$ 类，即 $h\in[f]+[g]$ iff $h(x)=f(x)+g(x)$，可能在某些空集上除外。这是一致的，因为两个空集的联合是空的。这个定义显然不取决于从每个等价类

中选择的代表。同样，对于

与常数相乘：$a[f]=[af]$为a R。因此，$L^1(E)$是一个向量
这些操作的空间。

严格地说，我们应该继续区分等价类[f]L_1 (E)和函数f_1 (E)，后者是这个类的代表。然而，在下面的所有内容中一致地这样做会掩盖基本的想法，而且根据上下文，将f互换为L^1 和^{1}的成员，并不会严重损失清晰度。换句话说，通过把等价类[f]当作函数f来处理，我们隐含地识别了两个函数，只要它们是当量相等。有了这个约定，我们就会明白，下面定义的 "长度函数 "是L^1 (E)上的一个真正的规范。

我们给L^1 (E)配备了准则

$$\|f\|_1 = \int_E |f| \, dm。$$

这是L^1 (E)上的一个规范：

1. $\|f\|_1 = 0$，当且仅当$f = 0$ a.e.，所以$f \equiv 0$作为L^1 (E) 的一个元素、
2. $\|cf\|_1 = \int_E |cf| \, dm = |c| \int_E |f| \, dm = |c| \, \|f\|_1$， ($c \in$ R),
3. $\|f + g\|_1 = \int_E |f + g| \, dm \le \int_E |f| \, dm + \int_E |g| \, dm = \|f\|_1 + \|g\|_1$。

从我们现在的角度来看，L^1 (E)最重要的特点是它是一个*完整的*规范化向量空间。下面给出了精确的定义。实线R和欧几里得空间R的完备性^n，是指导

在这里，我们寻求一种类似的方法，它具有类似的功能。

在函数空间所提供的无限维背景下的影响。该定义将针对一般规范化向量空间进行阐述：

定义5.3

设X是一个具有规范$\| \; \|_X$的向量空间。我们说，一个序列$f_n \in$ X是*Cauchy*如果

$$\forall \varepsilon > 0 \; \exists N : \forall n, m \ge N \quad \|f_n - f_m\|_X < \varepsilon。$$

如果每个Cauchy序列都能收敛到X的某个元素，那么我们说

X是*完整的*。

例5.1

设$f_n(x) = \frac{1}{x} 1_{[n,n+1]}(x)$，并假设$n \neq m$。

$$\|F_n - F_m\|_1 = \int_0^\infty \frac{1}{x}\left|1_{[n,n+1]} - 1_{[m,m+1]}\right| dx$$
$$= \int_n^{n+1} \frac{1}{x} dx + \int_m^{m+1} \frac{1}{x} dx$$
$$= \log \frac{n+1}{n} + \log \frac{m+1}{m}。$$

如果$a_n \to 1$，那么对数$a_n \to 0$，右手边可以小到我们希望：对于$\varepsilon > 0$，取N，使$\log\frac{N+1}{N} < \frac{\varepsilon}{2}$。所以$f_n$是一个考奇序列，在$L^1$（0，∞）。(当$E = (a, b)$时，我们把$L^1(a, b)$写成$L^1(E)$，等等。)

练习5.2

决定下列各项是否是考奇的序列，在$L^1(0, \infty)$

(a) $f_n = 1_{[n,n+1]}$

(b) $f_n = \frac{1}{x} 1_{(0,n)}$

(c) $f_n = \frac{1}{x^2} 1_{(0,n)}$

下面主要结果的证明主要利用了贝波-列维定理，以便将主要收敛问题转移到数列的问题上。

它的作用主要是提供一个类似的事实，即在R中（因此在C中）绝对收敛的数列总是收敛的。(Beppo-Levi定理显然延伸到了复值函数，就像我们的显示为主导收敛定理，但在下面的证明中我们将集中于实际情况，因为对C的扩展是直接的）。

我们简单地回顾一下序列的这一属性是如何确保R中的完整性的：让(x_n)是R中的一个Cauchy序列，并提取一个子序列(x_{n_k})，使得$|x_n - x_{n_k}| < 2^{-k}$对于所有$n \geq n_k$，如下所示：

1. 找到n_1，使$|x_n - x_{n_1}| < 2^{-1}$，所有$n \geq n_1$、
2. 找到$n_2 > n_1$，使$|x_n - x_{n_2}| < 2^{-2}$，所有$n \geq n_2$、
3. ...
4. 找到$n_k > n_{k-1}$，并且$|x_n - x_{n_k}| < 2^{-k}$，所有$n \geq n_k$。

Cauchy属性确保每次都能找到这样的n_k。现在考虑具有部分和的伸缩数列

$$y_k = x_{n_1} + (x_{n_2} - x_{n_1}) + \cdots + (x_{n_k} - x_{n_{k-1}}) = x_{n_k}$$

其中有

$$|y_k| \le |x_{n1}| + \sum_{i=1}^{k} |x_{n_i} - x_{n_{i-1}}| < |x_{n1}| + \sum_{i=1}^{k} \frac{1}{2^i}.$$

因此，这个序列收敛，换句话说，(x_{n_k})在R中收敛，其极限也是整个考奇序列(x_n)的极限。

因此，为了应用下面的Beppo-Levi定理，我们需要从$L^1(E)$中的给定Cauchy序列中提取一个 "快速收敛序列"。这就为原始序列提供了一个极限，而法图定理则是
休息。

定理5.1

空间L^1 （E） 是完整的。

证明

假设f_n是一个Cauchy序列。让$\varepsilon=\frac{1}{2}$。有N个$_1$，这样，对于$n \ge N_1$

$$\|f_n - f_{N_1}\| \le \frac{1}{2}。$$

接下来，让$\varepsilon=\frac{1}{2^2}$，对于一些$N_2 > N_1$，我们有

$$\|f_n - f_{N_2}\|_1 \le \frac{1}{2^2}$$

对于$n \ge N_2$。这样，我们构建一个子序列f_{N_k}，满足

$$\|f_{N_{n+1}} - f_{N_n}\|_1 \le \frac{1}{2^n}$$

对于所有的n，因此，该系列$\sum_{n \ge 1} \|f_{N_{n+1}} - f_{N_n}\|_1$收敛，根据贝波-列维定理，该系列

$$f_{N_1}(x) + \sum_{n=1}^{\infty} [f_{N_{n+1}}(x) - f_{N_n}(x)]$$

收敛，用$f(x)$表示其总和。由于

$$f_{N_1}(x) + \sum_{n=1}^{k} [f_{N_{n+1}}(x) - f_{N_n}(x)] = f_{N_{k+1}}$$

左手边收敛于f（x），所以$f_{N_{k+1}}$（x）收敛于f（x）。由于实数序列$f_n(x)$是Cauchy，且上述子序列收敛，所以整个序列收敛于同一极限$f(x)$。

我们必须证明，$f \in L^1$ 和$\|f - f_k\|_1 \to 0$。

让$\varepsilon>0$，考奇条件给出一个N，使得

$$\forall n, m \geq N, \|f_n - f_m\|_1 < \varepsilon。$$

根据Fatou's lemma

$$\|f - f_m\|_1 = \int |f - f_m| \, dm \leq \liminf_{k\to\infty} \int |f_{N_k} - f_m| \, dm = \lim_{k\to\infty} \|f_{N_k} - f_m\|_1 < \varepsilon$$

(5.1)

所以$f - f_m \in L^1$，这意味着$f = (f - f_m) + f_m \in L^1$，但(5.1)还可以得出$\|f - f_m\|_1 \to 0$。 □

5.2 希尔伯特空间L²

我们现在介绍的空间在该理论中起着特殊的作用。在函数空间中，它提供了最接近欧几里得空间$\mathbb{R}^n$ 的类似物，其几何形状紧密地模仿了$\mathbb{R}^n$ 。通过积分，可以通过内积诱导规范，这又提供了一个

函数之间的正交性（以及因此而产生的 "角度"）的概念。这使L^2 具有许多令人愉快的特性，如 "毕达哥拉斯定理 "和正交投影的概念，这在许多应用中发挥着重要作用。

为了定义规范，因此对于一个给定的可测量集，空间$L^2(E)$ $E\subset\mathbb{R}$，让

$$\|f\|_2 = \left(\int_E |f|^2 \, dm\right)^{\frac{1}{2}}$$

并将$\mathcal{L}^2(E)$定义为这个量是有限的可测量函数的集合。(注意，与L^1 一样，我们需要非负的积分；为了使平方根有意义，积分必须是非负的。尽管当$f(x)$为实数时，我们总是有$f^2(x) = (f(x))^2 \geq 0$，但为了包括复值函数$f: E \to C$ 的情况，需要使用模子。

这也使得我们的符号与其他L^p -空间的符号一致。

考虑到以下情况，其中$|f|^2$ 被任意$p \geq 1$的$|f|^p$ 所取代)。

我们引入$L^2(E)$作为$\mathcal{L}^2(E)$元素的等价类的集合，在等价关系$f \sim g$ iff $f = g$ a.e.下，与$L^1(E)$完全一样，并继续采用将等价类视为函数的惯例。如果$f: E\to\mathbb{C}$满足$\int_E |f|^2 \, dm < \infty$，我们写$f \in L^2(E, \mathbb{C})$ --再次使用f

可互换地作为其等价类的代表，并表示该类本身。

直接证明L^2 （E）是一个矢量空间：显然，对于$a\in$R、
$|af|^2$是可整定的，如果$|f|^2$是，而

$$|f+g|^2 \le 2^2 \max\{|f|^2, |g|^2\} \le 4(|f|^2 +$$

$|g|^2$)表明L^2 (E)在加法下是封闭的。

5.2.1 L的属性 -2 norm

我们提供了一个简单的证明，即映射$f \mapsto \|f\|_2$ 是一个规范：要看到它满足三角形不等式，需要做一点工作，但其思路将是
在R^n 的初级分析中非常熟悉，术语也是如此，尽管背景相当不同。我们陈述并证明一般情况下的结果
的L^2 （E，C）。

定理5.2（施瓦兹不等式）

如果f，$g\in L^2$ （E，C），那么$fg\in L^1$ （E，C）和

$$\left|\int_E f\bar{g}\,dm\right| \le \|fg\|_1 \le \|f\|_2\|g\|_2 \tag{5.2}$$

其中$\bar{g}$表示g的复共轭。

证明

用，代替$|f|$和$|g|$，我们可以假设 和 都是非负的f（第一个不等式已经被验证了，因为$\|fg\|_1 = \int_E |fg|\,dm$，第二个不等式只涉及每种情况下的模数）。由于我们事先不知道
$\int_E fg\,dm$是有限的，我们将首先通过设置$f_n = f \wedge n$和$g_n = g \wedge n$来限制对有界可测函数的关注，并将我们的积分域限制在有界集合$E\cap[-k, k] = E_k$。

对于任何$t\in$R，我们有

$$0 \le \int_{E_k} (f_n + tg_n)^2\,dm = \int_{E_k} f_n^2\,dm + 2t\int_{E_k} f_n g_n\,dm + t^2\int_{E_k} g_n^2\,dm。$$

作为t的二次方，它没有两个不同的解，所以判别式是非正的。因此，对于所有$n\ge1$

$$\left(2\int_{E_k} f_n g_n\,dm\right)^2 \le 4\left(\int_{E_k} f_n^2\,dm\right)\left(\int_{E_k} g_n^2\,dm\right)^2$$

$$\le 4\left(\int_E |f|^2\, dm\right)\left(\int_E |g|^2\, dm\right)$$

$$= \|f\|_2^2 \|g\|_2^2 \text{。}$$

单调收敛现在可以得到

$$\left(\int_{E_k} fg\ dm\right)^2 \leq \|f\|_2^2 \|g\|_2^2$$

对于每个k，由于$E = \bigcup_k E_k$，我们最后得到：

$$\left(\int_E fg\ dm\right)^2 \leq \|f\|_2^2 \|g\|_2^2,$$

这意味着施瓦兹不等式。 □

$L^2(E, \mathbb{C})$上的规范的三角形不等式现在一下子就出来了--我们需要证明 $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ 对于$f, g \in L^2(E, \mathbb{C})$：

$$\|f + g\|_2^2 = \int_E |f + g|^2\ dm = \int_E (f + g)\overline{(f + g)}dm = \int_E (f + g)(\bar{f} + \bar{g})\ dm。$$

后者的积分为

$$\int_E |f|^2\ dm + \int_E (f\bar{g} + \bar{f}g)\ dm + \int_E |g|^2\ dm,$$

它被$(\|f\|_2 + \|g\|_2)^2$所支配，因为施瓦兹不等式给出了

$$\int_E (f\bar{g} + g\bar{f})\ dm \leq 2\int_E |fg|\ dm \leq 2\|f\|_2\|g\|_2。$$

结果如下。

其他属性是直接的：

(i) 显然，$\|f\|_2 = 0$意味着$|f|^2 = 0$ a.e.，因此$f = 0$ a.e.、

(ii) 对于$a \in \mathbb{C}$，$\|af\|_2 = \left(\int_E |af|^2\ dm\right)^{\frac{1}{2}} = |a|\|f\|_2$。

因此，映射$f \mapsto \|f\|_2$是$L^2(E，\mathbb{C})$上的一个规范。

$L^2(E)$在此规范下是完整的，其证明类似于对$L^1(E)$，并将在下面的定理5.11中给出任意的L^p-空间（$1 < p < \infty$）。

一般来说，在不限制域集E的情况下，$L^1 \subseteq L^2$或$L^2 \subseteq L^1$。为了看到这一点，考虑$E = [1, \infty]$，$f(x) = \frac{1}{x}$。那么$f \in L^2(E)$，但$f \notin L^1(E)$。接下来把$F=（0，1）$，$g（x）= \frac{1}{\sqrt{x}}$。现在$g \in L^1(F)$，但$g \notin L^2(F)$。

对于有限度量空间--因此也是概率空间！我们确实有一个有用的包容物！-我们确实有一个有用的包容：

命题5.3

如果集合D具有有限度量（即$m(D) < \infty$），那么$L^2(D) \subset L^1(D)$。

提示 通过$|f|^2$来估计$|f|$，然后利用$|f|^2$的积分是有限的这一事实。

在探索L^2 的规范所引起的几何学之前，我们考虑L^2 中的序列的例子，以提供一点实践，确定哪些是L^2 -规范的考奇序列，并与它们作为L^1 的元素的行为进行比较。

例5.2

我们表明，序列$f_n = \frac{1}{x} 1_{[n,n+1]}$ 在$L^2(0,\infty)$中是Cauchy。

$$
\begin{aligned}
\|F_n - F_m\|_2^2 &= \int_0^\infty \left| \frac{1}{x^2} 1_{[n,n+1]} - 1_{[m,m+1]} \right|^2 dx \\
&= \int_n^{n+1} \frac{1}{x^2} dx + \int_m^{m+1} \frac{1}{x^2} dx \\
&= \frac{1}{n} - \frac{1}{n+1} + \frac{1}{m} - \frac{1}{m+1} \\
&\leq \frac{2}{n} + \frac{2}{m}
\end{aligned}
$$

而对于$\varepsilon>0$，让N是这样的， 即 $\frac{2}{N} < \frac{\varepsilon}{2}$ 那么$\|f_m - f_n\| < \varepsilon$，只要$m, n \geq N$.

练习5.3

顺序是

$$g_n(x) = 1_{(n,\infty)}(x) \frac{1}{x^2}$$

$L^2(R)$中的考奇序列?

练习5.4

决定下列各项是否是考奇的序列，在$L^2(0,\infty)$.

(a) $f_n = 1_{(0,n)}$

(b) $f_n = \frac{1}{x} 1_{(0,n)}$

(c) $f_n = \frac{1}{x^2} 1_{(0,n)}$

5.2.2 内积 空间

我们已经准备好了（在Lebesgue函数空间中）针对L^2的额外结构：

$$\forall f, g \in L^2(E, \mathbb{C}) \qquad (f, g) = \int fg\, dm \tag{5.3}$$

定义了一个*内积*，它诱导了L^2-norm：

$$\sqrt{(f, f)} = \left(\int_E f\bar{f}\, dm\right)^{\frac{1}{2}} = \left(\int_E |f|^2\, dm\right)^{\frac{1}{2}} = \|f\|_2 。$$

为了解释这意味着什么，我们验证了以下属性，所有这些都很容易从我们所开发的积分理论中得到：

命题5.4

直线性（在第一个参数中）

$$(f + g, h) = (f, h) + (g, h)、$$

$$(cf, h) = c(f, h)。$$

共轭对称的正定性

$$(f, g) = \overline{(g, f)}。$$

$$(f, f) \geq 0、\quad (f, f) = 0 \Leftrightarrow f = 0。$$

提示 在第一部分中使用积分的可加性，在最后一部分中回忆一下，如果f=0 a.e.，那么f是L的零元素^2 （E， C）。

作为一个直接的结果，我们得到关于第二个参数的共轭线性

$$(f, cg + h) = \bar{c}(f, g) + (f, h)。$$

当然，如果f，$g \in L^2$是实值的，内积是实值的，在第二个参数中也是线性的。

对施瓦兹不等式证明的研究表明，这里定义在$L^2(E, \mathbb{C})$上的内积的特殊形式与这个结果完全无关：我们需要证明的是，(5.3)中定义的映射有在最后一个命题中为其证明的属性。

因此，我们将作出以下重要的定义，从R^n的有限维背景中我们将熟悉这个定义，现在我们希望更普遍地应用这个定义。

定义5.4

C上的向量空间H的*内积*是一个映射（，）：H H C，它满足命题5.4中所列的三个条件。这一对$(H, (,))$被称为*内积空间*。

例5.3

R中通常的标量积^n，使这个空间成为一个实数内积空间，而C^n，配备（z，w）$=\sum_{i=1}^{n} z_i \overline{w_i}$ 是一个复杂的。

命题5.4表明，L^2 （E，C）是一个（复数）内积空间。随着定义的明显简化，向量空间$L^2 (E, \mathrm{R}) = L^2 (E)$ 是一个*实数*内积空间，即以R为标量集合。

下面的特性是上述定义的直接后果。

命题5.5

让$(H, (\cdot, \cdot))$是一个复杂的内积空间，有诱导规范$\|\cdot\|$。对于所有的$h_1, h_2 \in H$，下面的相同点成立：

(i) 平行四边形法：

$$\|h_1 + h_2\|^2 + \|h_1 - h_2\|^2 = 2(\|h_1\|^2 + \|h_2\|^2)。$$

(ii) 极化身份：

$$4(h_1, h_2) = \|h_1 + h_2\|^2 - \|h_1 - h_2\|^2 + i\{\|h_1 + ih_2\|^2 - \|h_1 - ih_2\|^2\}。$$

备注5.1

这些特性虽然是定义的微不足道的结果，但在检查某些规范*不能*由内积引起时是很有用的。下面的练习中给出了一个例子。随着完整性的增加，这些特性可以用来描述内积规范：可以证明，在完全规范化空间（称为*Banach*空间）中，那些规范由内积诱导的空间（即*希尔伯特*空间）正是那些平行四边形定律成立的空间，然后通过极化特性从规范中恢复内积。我们不在此证明这一点。

练习5.5

证明不可能在空间$C[0, 1]$上定义一个内积

的连续函数$f : [0, 1] \to \mathbb{R}$，它将诱导出超规范
$\|f\|_\infty = sup\{|f(x)| : x\in [0, 1]\}$。

提示 试着用函数$f, g\ C[0, 1]$验证平行四边形定律，函数$f(x)=1$,
$g(x)=x$为所有x。

练习5.6

证明不可能在空间上定义一个内积
$L^1([0, 1])$，规范为 $\|\ \|_1$。

提示 试着用下列函数来验证平行四边形定律
$f(x) = \frac{1}{2} - x,\ g(x) = x - \frac{1}{2}$

5.2.3 正交性和 投影

我们以一种有点迂回的方式介绍了内积空间的概念，以便强调这种结构是空间L^2 中可用的自然的附加工具，它仍然是我们的主要兴趣来源。然而，这种额外的结构确实使我们能够简化许多论证，并证明其他函数空间所没有的结果。从某种意义上说，L^2 中的数学生活在一个无限维的矢量中是 "好得不能再好 "的。
空间，因为它的结构与更熟悉的空间$\mathbb{R}^n$
和$\mathbb{C}^n$ 。

作为这个新工具的力量的一个例子，回顾一下，对于$\mathbb{R}^n$ 中的向量，我们有一个重要的正交概念，这意味着两个向量的标量积为零。这可以延伸到任何内积空间，尽管
我们将首先陈述它，并对L产生明确的例子²：函数f，g
是*正交*的，如果

$$(f, g) = 0。$$

例5.4

如果 $_{[0,1]}$，那么（f，g）=0，当且仅当$\int_1 g(x)\,dx=0$，例如，如果
$f=1$

0

$g(x) = x - \frac{1}{2}$。

练习5.7

证明$f(x)$=sin nx，$g(x)$=cos mx，对于$x\in[-\pi，\pi]$，外面是0、

是正交的。

证明$f(x)=\sin nx$，$g(x)=\sin mx$，对于$x\in[-\pi，\pi]$，外侧为0，对于n /=m来说是正交的。

事实上，在任何内积空间中，我们可以通过设置来定义元素g，h之间的*角度*

$$\cos\vartheta=\frac{(g,h)}{\|g\|\|h\|}.$$

请注意，根据施瓦兹不等式，这个量--正如我们将在下面看到的，它也有一个自然的解释，即两个（中心）随机变量之间的*相关性*--位于[1, 1]中，并且（g, h）=0意味着$\cos\vartheta=0$，即ϑ是$\frac{\pi}{2}$的奇数倍。因此，自然可以说，g是*正交于*

h，如果（g，h）=0。

复数内积空间（H，（，））中向量的正交性提供了一种在H中表述毕达哥拉斯定理的方法：因为$\|g+h\|^2=(g+h,g+h)=(g,g)+(h,h)+(g,h)+(h,g)=\|g\|^2+\|h\|^2+(g,h)+(h,g)$我们马上看到，如果$g$和$h$在$H$中是正交的，那么$\|g+h\|^2=\|g\|^2+\|h\|^2$。

现在把注意力限制在$(H,(\cdot,\cdot))$在内积规范中是完全的情况。记得这意味着（见定义5.3），如果（h_n）是H中的一个Cauchy序列，那么存在$h\in H$，使得$\lim_{n\to\infty}\|h_n-h\|=0$。如备注5.1所述，如果这一点成立，我们称$H$为*希尔伯特空间*。我们的内容

我们有一个关于这种空间的基本事实：

让K是H的一个完整的子空间，因此上述条件对于K中的(h_n)也是成立的，然后得到$h\in K$。正如一个直角三角形在(x, y)平面的标准位置上的水平边是斜边对水平轴的投影，而垂直边与该轴正交，我们现在证明H中的矢量对子空间K的正交投影的存在。

定理5.6

让K是希尔伯特空间H的一个完整子空间。对于每一个$h\in H$，我们可以找到一个唯一的$h'\in K$，使得$h''=h-h'$是正交于H的每一个元素。

K.等价地，$\|h-h'\|=\inf\{\|h-k\|:k\in K\}$。

证明

定义h'的两个条件是等价的：假设已经找到$h' \in K$，从而使$h'' = h - h'$与每一个$k \in K$正交。给定$k \in K$，注意

由于$(h' - k)\in K$，$(h'', h' - k) - 0$，所以毕达哥拉斯定理的含义是

$$\|h - k\|^2 = \|(h - h') + (h' - k)\|^2 = \|h''\|^2 + \|h' - k\|^2 > \|h''\|^2$$

除非$k = h'$。因此，$\|h''\| = \|h - h'\| = \inf\{\|h - k\| : k\in K\} = \delta_K$，说。

反过来说，在找到$h' \in K$，使$\|h - h'\| = \delta_K$，那么对于任何实的t和$k\in K$，$h' + tk\in K$，所以

$$\|h - (h' + tk)\|^2 \geq \|h - h'\|^2。$$

将内积相乘，写成$h'' = h - h'$，这意味着：
$-t[(h'', k)+(k, h'')]+t^2\|k\|^2 \geq 0$。只有在$(h'', k)=0$的情况下，这才能对所有接近0的$t$成立，因此，对于每一个$k\in K$，$h''\perp k$。

要找到$h' \in K$，且$\|h - h'\| = \delta_K$，首先在K中选择一个序列（k_n），使$\|h - k_n\| \to \delta_K$，因为$n \to \infty$。然后将平行四边形定律（命题5.5(i)）应用于向量$h_1 = h - \frac{1}{2}(k_m + k_n)$和$h_2 = \frac{1}{2}(k_m - k_n)$。

请注意，$h_1 + h_2 = h - k_n$和$h_1 - h_2 = h - k_m$。因此，平行四边形法律内容如下

$$\|h - k_n\|^2 + \|h - k_m\|^2 = 2\left(\left\|h - \frac{1}{2}(k_m + k_n)\right\|^2 + \left\|\frac{1}{2}(k_m - k_n)\right\|^2\right)$$

由于$\frac{1}{2}(k_m + k_n) \in K$，*所以*$\|h - \frac{1}{2}(k_m + k_n)\|^2 \geq \delta_K^2$。当$m, n \to \infty$时，左边的

边收敛到$2\delta_K^2$，因此右边的最后一项必须收敛到0。因此序列(k_n)在K中是Cauchy，所以收敛到一个元素h'的

K.但由于$\|h - k_n\| \to \delta_K$，而$\|k_n - h'\| \to 0$，因为$n \to \infty$，$\|h - h'\| \leq \|h - k_n\| + \|k_n - h'\|$表明，$\|h - h'\| = \delta_K$。这就完成了证明。□

在写$h = h' + h''$时，我们将向量$h \in H$分解为两个向量之和，第一个是它在K上的正交投影，而第二个是与K中所有向量的正交。我们说h''是*正交于K的*，用$K^\perp$表示正交于K*的*所有向量的集合。这就把H表现为一个*直接的和*$H = K \oplus K^\perp$，第一个因子的每个向量都正交于第二个因子的每个向量。

我们将在第5.4.3节中使用对$L^2(\Omega, \mathcal{F}, P)$子空间的正交投影的存在来构造随机变量相对于σ*场的*条件期望。

备注5.2

前面的讨论几乎没有触及内积空间结构的表面，例如$L^2(E)$，它在例如[10]中被优雅地解释。一方面，内积空间中的正交概念导致了对*正交集*的考虑，即H

中的族（e_α）是相互的

的正交性，并且具有规范1。一个自然的问题是，H的每一个元素是否可以由（e_α）的线性组合来表示（或至少是近似的）。在L^2 ([π, π])中，这导致，例如，*傅里叶级数代表*

$f(x)=\sum_{n=0}^{\infty}(f, \psi_n)\psi_n$，其中正态函数

函数是$\psi_0 = \frac{1}{\sqrt{2\pi}}$，$\psi_{2n}(x) = \frac{1}{\sqrt{\pi}}\cos nx$，$\psi_{2n-1}(x) = \frac{1}{\sqrt{\pi}}\sin nx$，而

系列在L^2-norm中收敛。L^2的完整性对于确保这样一个*正态基础的*存在至关重要。

5.3 L^p 空间： 完全性

更一般地说，当我们对f的p^{th}次方进行积分时，就会得到空间$L^p(E)$。对于$p \geq 1$，我们说$f \in L^p$（类似地，对于L^p（E）和L^p（E，C）），如果$|f|^p$是可积分的（当f和g*是可积分的*时候，用同样的约定来识别它们是a.e.相等的）。需要做一些工作来检查L^p是一个向量空间，并且为L^2介绍的规范的 "自然 "概括实际上是一个规范。我们将需要$p \geq 1$来实现这一点。

定义5.5

对于每个$p \geq 1$，$p < \infty$，我们定义（识别类和函数）

$$L^p(E) = \{f : \int_E |f|^p \, dm \text{是有限的}\}。$$

而L^p上的规范，定义为

$$\|f\|_p = \left(\int_E |f|^p \, dm\right)^{\frac{1}{p}}.$$

(考虑到这一点，我们用$\|f\|_1$表示$L^1(E)$中的规范，用$\|f\|_2$(E)表示L_2。)

回顾定义3.2：对于任何可衡量的函数f：$E \to [0, \infty]$

$$\text{ess } sup f := inf\{c : |f| \leq c \text{ a.e.}\}。$$

更确切地说，如果$F = \{c \geq 0 : m(\{|f| > c\}) = 0\}$，我们设置ess $sup f = \inf F$ (用惯例 $\inf \varnothing = +\infty$)。很容易看出，下限属于$F$。

定义5.6

一个满足ess sup f < 的可测量函数f被称为*基本有界*，E上所有基本有界函数的集合用$L^\infty(E)$表示（同样用通常的函数等价识别法）。
类），其规范$\|f\|_\infty$ = ess sup$f$。

从命题3.9和明显的同一性可以看出，sup(cf) = c(ess sup$f$)，即$L^\infty(E)$是一个矢量空间。

我们需要证明，对于每个p（1 ≤ p ≤ ∞），($L^p(E)$, ‖‖)是一个规范的向量空间。

首先我们观察到，L^p （E） 是一个矢量空间，对于1 ≤ p < ∞.如果f和g
属于L^p，那么它们是可测量的，因此cf和$f+g$也是可测量的。我们有
$|cf(x)|^p = |c|^p|f(x)|^p$ 因此

$$\|cf\|_p = \left(\int |cf(x)|^p\,dx\right)^{\frac{1}{p}} = |c|\left(\int |f(x)|^p\,dx\right)^{\frac{1}{p}} = |c|\|f\|_p。$$

接下来$|f(x)+g(x)|^p \le 2^p \max\{|f(x)|^p, |g(x)|^p\}$，因此$\|f+g\|_p$是有限的，只要
$\|f\|_p$和$\|g\|_p$是。此外，如果$\|f\|_p = 0$，那么$|f(x)|^p = 0$几乎无处不在
所以$f(x)$=0几乎无处不在。反之亦然。

三角形不等式

$$\|f+g\|_p \le \|f\|_p + \|g\|_p$$

对于一般的P 1来说，这绝不是显而易见的：我们需要推导出一个著名的Hölder的不等式，它在许多情况下也是非常有用的，它概括了Schwarz不等式。

备注5.3

在解决这个问题之前，我们注意到$L^\infty(E)$的情况相当容易：$|f+g| \le |f|+|g|$ 一次性意味着$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$，同样，$|af| = |a||f|$给出$\|af\|_\infty = |a|\|f\|_\infty$。因此，$L^\infty(E)$是一个向量空间，由于$\|f\|_\infty = 0$显然成立，当且仅当$f = 0$ a.e.，因此，$\|\cdot\|_\infty$是$L^\infty(E)$上的一个规范。练习3.6和5.6表明，这个规范不能由L^∞上的任何内积引起。

定理5.7

对于任何非负实数x，y和所有α，β（0，1），$\alpha+\beta=1$，我们有

$$x\,y^{\alpha\beta} \leq \alpha x + \beta y。$$

证明

如果$x=0$，要求是明显的。所以取$x>0$。考虑$f(t)=(1-\beta)+\beta t-t^{\beta}$，$t\ge 0$，$\beta$为给定值。我们有$f'(t)=\beta-\beta t^{\beta-1}=\beta(1-t^{\beta-1})$，由于$0<\beta<1$，$f'(t)<0$，1)。所以$f$在$[0，1]$上减少，而$f'(t)$在$(1，\infty)$上$>0$，因此$f$在$[1，\infty)$上增加。所以$f(1)=0$是$f$在$[1,\infty]$上的唯一最小点。$[0,\infty]$，即$t\ge 0$时$f(t)\ge 0$。现在设$t=\frac{y}{x}$，则$(1-\beta)+\beta\frac{y}{x}-\left(\frac{y}{x}\right)^{\beta}\ge 0$、

即$\left(\frac{y}{x}\right)^{\beta}\le\alpha+\beta\frac{y}{x}$。写作$x=x^{\alpha+\beta}$，我们有$x^{\alpha+\beta}\left(\frac{y}{x}\right)^{\beta}\le\alpha x+\beta x\frac{y}{x}$，所以
即$x^{\alpha}y^{\beta}\le\alpha x+\beta y$，符合要求。□

定理5.8（Hölder不等式）

如果$\frac{1}{p}+\frac{1}{q}=1, p>1$，那么对于$f\in L^p(E), g\in L^q(E)$，我们有$fg\in L^1$和

$$\|fg\|_1\le\|f\|_p\|g\|_q$$

就是说

$$\int|fg|\,dm\le\left(\int|f|^p\,dm\right)^{\frac{1}{p}}\left(\int|g|^q\,dm\right)^{\frac{1}{q}}.$$

证明

步骤1。假设$\|f\|_p=\|g\|_q=1$，所以我们只需要证明$\|fg\|_1\le 1$。
我们应用定理5.7，$\alpha=\frac{1}{p}$，$\beta=\frac{1}{q}$，$x=|f|^p$，$y=|g|^q$，那么我们有

$$|fg|=x^{\frac{1}{p}}y^{\frac{1}{q}}\le\frac{1}{p}|f|^p+\frac{1}{q}|g|^q。$$

通过积分，我们得到

$$\int|fg|\,dm\le\frac{1}{p}\int|f|^p\,dm+\frac{1}{q}\int|g|^q\,dm=\frac{1}{p}+\frac{1}{q}=1$$

因为$\int|f|^p\,dm=1$，$\int|g|^q\,dm=1$。所以我们有$\|fg\|_1\le 1$，符合要求。

第2步。对于一般的$f\in L^p$和$g\in L^q$，我们写$\|f\|_p=a, \|g\|_q=b$，对于某些$a，b>0$。（如果a或b为零，那么其中一个函数几乎为零。到处都是量，而且不等式是微不足道的）。因此，函数$\tilde{f}=\frac{1}{a}f$，$\tilde{g}=\frac{1}{b}g$满足步骤1的假设，所以$\|\tilde{f}\tilde{g}\|_1\le\|\tilde{f}\|_p\|\tilde{g}\|_q$。由此可得

$$\frac{1}{\ }\|\qquad\frac{1}{\ }\|\quad\frac{1}{\ }\|$$

$$\frac{\|fg\|_1}{ab} \leq \frac{\|f\|_p}{a}\frac{\|g\|_q}{b}$$

并乘以ab后，结果得到证明。 □

假设$p = q = 2$，并回顾L中标量积的定义²
我们得到以下现在熟悉的Hölder不等式的特例。

推论5.9（施瓦兹不等式）

如果f，$g \in L^2$，则

$$|(f, g)| \le \|f\|_2 \|g\|_2 。$$

我们现在可以完成验证，即 $\|\cdot\|_p$ 是$L^p(E)$上的一个规范。

定理5.10（Minkowski不等式）

对于每个$p \ge 1$，f，$g \in L^p(E)$

$$\|f + g\|_p \le \|f\|_p + \|g\|_p 。$$

证明

假设$1 < p < \infty$（前面做了$p = 1$的情况）。我们有

$$|f + g|^p = |(f + g)(f + g)^{p-1}| \le |f||f + g|^{p-1} + |g||f + g|^{p-1}、$$

同时取q，使$\frac{1}{p} + \frac{1}{q} = 1$，换言之，$p+q=pq$，我们得到

$$|f + g|^{(p-1)q} = |f + g|^p < \infty 。$$

因此，$(f+g)^{p-1} \in L^q$，并且

$$\|(f + g)^{p-1}\|_q = \left(\int |f + g|^p \, dm\right)^{\frac{1}{q}} .$$

我们可以应用Hölder不等式：

$$\begin{aligned}
\int |f + g|^p \, dm &\le \int |f||f + g|^{p-1} \, dm + \int |g||f + g|^{p-1} \, dm \\
&\le \left(\int |f|^p \, dm\right)^{\frac{1}{p}} \left(\int |f + g|^p \, dm\right)^{\frac{1}{q}} \\
&\quad + \left(\int |g|^p \, dm\right)^{\frac{1}{p}} \left(\int |f + g|^p \, dm\right)^{\frac{1}{q}} \\
&= \left(\left(\int |f|^p \, dm\right)^{\frac{1}{p}} + \left(\int |g|^p \, dm\right)^{\frac{1}{p}}\right)
\end{aligned}$$

$\frac{1}{}$

$-A$ p

与$A = \left(\int |f+g|^p dm\right)^{\frac{1}{q}}$ 如果A=0，那么$\|f+g\|_p = 0$，而且没有什么可以证明。所以假设A>0，除以A

$$\|f+g\|_p = \left(\int |f+g|^p dm\right)^{1-\frac{1}{q}}$$
$$= \frac{1}{A}\int |f+g|^p dm$$
$$\leq \left(\int |f|^p dm\right)^{\frac{1}{p}} + \left(\int |g|^p dm\right)^{\frac{1}{p}}$$
$$= \|f\|_p + \|g\|_p$$

这是要证明的。 □

接下来我们证明$L^p(E)$是一个完全规范化空间（即巴拿赫空间）的例子，对于$1 < p < \infty$，即$L^p(E)$中的每个考奇序列在规范上都收敛于$L^p(E)$的一个元素。我们有时将序列在L^p规范中的收敛称为p^{th} *平均值的收敛*。

该证明与p=1的情况非常相似。

定理5.11

空间$L^p(E)$在$1 < p < \infty$时是完整的。

证明

给定一个考奇序列f_n（即$\|f_n - f_m\|_p \to 0$，因为n，$m \to \infty$），我们找到一个子序列f_{n_k}，其中

$$\|f_n - f_{n_k}\|_p < \frac{1}{2^k}$$

对于所有k≥1，我们设定

$$g_k = \sum_{i=1}^{k} |f_{n_{i+1}} - f_{n_i}|、\qquad g = \lim_{k\to\infty} g_k = \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}|。$$

三角形不等式得到$\|g_k\|_p \leq \sum_{i=1}^{k} \frac{1}{2^i} < 1$，我们可以应用法图的对非负可测量函数$g^p_k$，$k$≥1，所以

$$\|g\|_p^p=\int\lim_{n\to}g^p\,dm\le\liminf_{k\to}\int g^p\,dm\le 1。$$

因此，g几乎到处都是有限的，$f_{n_1}+\sum_{i=1}(f_{n_{i+1}}-f_{n_i})$几乎到处绝对收敛，将一个可测量的函数$f$定义为它的总和。

我们需要证明$f\in L^p$。首先注意，$f=\lim_{k\to\infty} f_{n_k}$ a.e.，并且给定$\varepsilon>0$，我们可以找到N，使$\|f_n - f_m\|_p<\varepsilon$，对于$m$，$n\geq N$。应用对序列$(|f_{n_i}-f_m|)_{i\geq 1}$，让$i\to\infty$，我们有

$$\int |f-f_m|^p\,dm \leq \liminf_{i\to} \int |f_{n_i}-f_m|^p\,dm \leq \varepsilon^p。$$

因此，$f-f_m \in L^p$，所以$f=f_m+(f-f_m)\in L^p$，我们有$\|f-f_m\|_p<\varepsilon$对于所有$m\geq N$。因此，$f_m\to f$在L^p-norm中符合要求。 □

空间L^∞（E）也是完整的，因为对于L^∞（E）中的任何考奇序列（f_n），其中$|f_k(x)|>\|f\|_\infty$或$|f_n(x)-f_m(x)|>\|f_n-f_m\|_\infty$，对于$k$、$m$、$n\in\mathbb{N}$，仍然是一个空集，$F$说。在$F$之外，序列（$f_n$）均匀地收敛于一个有界函数，$f$说。很明显，$\|f_n-f\|_\infty\to 0$，$f\in L^\infty(E)$，所以我们就完成了。

练习5.8

顺序是

$$g_n(x)=1_{(0,\frac{1}{n}]}(x)\frac{1}{\sqrt{x}}$$

Cauchy in L^4?

我们在不同p的L^p空间之间有以下关系，这是对命题5.3的概括。

定理5.12

如果E具有有限的李布斯格度量，那么当$1\leq p\leq q\leq\infty$时，L^q（E）$\subseteq L^p$（E）。

证明

请注意，如果$|f(x)|^p\leq 1$，那么$|f(x)|\leq 1$。如果$|f(x)|\geq 1$，那么$|f(x)|^p\leq|f(x)|^q$。因此

$$|f(x)|^p\leq 1+|f(x)|^q、$$

$$\int_E |f|^p\,dm \leq \int_E 1\,dm + \int_E |f|^q\,dm = m(E)+\int_E |f|^q\,dm<\infty、$$

所以如果$m(E)$和$\int_E |f|^q\,dm$是有限的， □

对于$\int |f|^p \, dm$也是如此。

5.4 P robability

5.4.1 瞬间

属于空间*L的*随机变量^p （Ω），其中指数p $\in$ N，在概率中起着重要作用。

定义5.7

随机变量$X \in L^n$ (Ω)的*n阶矩*是指

$$E(X^n)、\quad n = 1, 2, ...$$

写出E(X)=μ；那么*中心矩*由以下公式给出

$$E(X - \mu)^n、\quad n = 1, 2, ...$$

根据定理4.28，矩是由概率分布决定的：

$$E(X^n) = \int x^n \, dP_X(x)、$$

$$E((X - \mu)^n) = \int (x - \mu)^n \, dP_X(x),$$

如果X有一个密度f_X，那么根据定理4.32，我们有

$$E(X^n) = \int x \, f^n{}_X(x) \, dx、$$

$$E((X - \mu)^n) = \int (x - \mu) \, f^n{}_X(x) \, dx。$$

命题5.13

如果$E(X^n)$对于某些n是有限的，那么对于$k \le n$，$E(X^k)$是有限的。如果$E(X^n)$是无限的，那么对于$k \ge n$的$E(X^k)$也是如此。

提示 使用定理5.12。

练习5.9

找到X，使$E(X^2) = \infty$， $E(X) < \infty$。这样的X能否有E(X)=0?

提示 你可以使用本章中以前的一些例子。

定义5.8

一个随机变量的*方差*是二阶的中心矩：

$$\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 \text{。}$$

显然，写出μ=E(X)、

$$\mathrm{Var}(X) = \mathrm{E}(X^2 - 2\mu X + \mu^2) = \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - \mu^2 \text{。}$$

这表明，前两个时刻决定了第二个中心摩尔。这可以被推广到任意顺序，更重要的是，这种关系也是反过来的。

命题5.14

n阶的中心矩是由$k \leq n$的k阶矩决定的。

提示 使用二项式定理和积分的线性。

命题5.15

n阶矩由$k \leq n$的k阶中心矩决定。

提示 将$\mathrm{E}(X^n)$写成$\mathrm{E}((X - \mu + \mu)^n)$，然后使用二项式定理。

练习5.10

求Var(aX)与Var(X)的关系。

例5.5

如果X在[a, b]上具有均匀分布，即$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x)$ 那么

$$\int x f_X(x)\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\frac{1}{2}x^2\Big|_a^b = \frac{1}{2}(a+b) \text{。}$$

练习5.11

证明对于均匀分布的X，$\mathrm{Var}X = \frac{1}{12}(b-a)^2$。

练习5.12

找出以下的方差

(a) 一个常数随机变量X，$X(\omega)=a$，对于所有的ω

(b) $X:[0,1]\to\mathbb{R}$，由$X(\omega)=\min\{\omega, 1-\omega\}$给出（到区间[0, 1]的最近端点的距离）

(c) $X:[0,1]^2\to\mathbb{R}$，到正方形$[0,1]^2$最近的边的距离。

我们将看到，对于高斯分布，前两个时刻决定了其余的时刻。首先，我们计算一下期望值：

定理5.16

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\mu.$$

证明

进行替换 $z=\frac{x-\mu}{\sigma}$，然后，用$\int$来表示$\int_{\mathbb{R}}$、

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int xe^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{\sigma}{\sqrt{2\pi}}\int ze^{-\frac{z^2}{2}}dz+\frac{\mu}{\sqrt{2\pi}}\int e^{-\frac{z^2}{2}}dz.$$

请注意，第一个积分是零，因为积分项是一个奇数函数。第二个积分是$\sqrt{2\pi}$，因此有了结果。

所以密度中的参数μ是数学期望值。我们现在表明，σ^2是方差。

定理5.17

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{\mathbb{R}}(x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\sigma^2$$

证明

进行与之前相同的替换：$z=\frac{x-\mu}{\sigma}$; 那么

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int(x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{\sigma^2}{\sqrt{2\pi}}\int z^2 e^{-\frac{z^2}{2}}dz.$$

$2 - z^2$

通过部分积分$u=z$，$v=ze^{-z^2/2}$，得到

$$\frac{\sigma^2}{\sqrt{2\pi}}\int z^2 e^{-\frac{z^2}{2}}dz = -\frac{\sigma^2}{\sqrt{2\pi}}ze^{-\frac{z^2}{2}}\Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}}\int e^{-\frac{z^2}{2}}dz = \sigma^2$$

因为第一个项消失了。□

请注意，高斯随机变量的奇数中心矩为零：积分

$$\frac{1}{\sqrt{2\pi}\sigma}\int (x-\mu)^{2k+1} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

消失，因为经过上面的替换，我们对一个奇数函数进行了积分。通过重复部分积分的论证，可以证明

$$\mathrm{E}(X-\mu)^{2k} = 1 \cdot 3 \cdot 5 \cdots (2k-1)\sigma^k。$$

例5.6

让我们考虑考奇密度 $\frac{1}{\pi}\frac{1}{1+x^2}$ 并尝试计算期望值(我们将看到这是不可能的)：

$$\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x}{1+x^2}dx = \frac{1}{2\pi}\text{极限}_{x_n\to+\infty}\ln(1+x_n^2) - \text{肢体}_{y_n\to-\infty}\ln(1+y_n^2)$$

对于一些序列x_n，y_n。结果，如果是有限的，应该不取决于他们的选择，然而，如果我们设置例如$x_n = ay_n$，那么我们有

$$\text{肢体}_{x_n\to+\infty}\ln(1+x_n^2) - \text{肢体}_{y_n\to-\infty}\ln(1+y_n^2) = \text{极限}_{y_n\to\infty}\ln\frac{1+a y_n^2}{1+y_n^2} = \ln a$$

这是个矛盾。因此，我们看到，对于考奇密度来说，矩不存在。

备注5.4

我们不需要证明就给出了特征函数和时刻之间的简单关系：(回顾一下，j_X $(t) = \mathrm{E}(e^{itX})$ --见定义4.5)。

和 $_X$如果j_00是k次连续可微的，那么X有有限的k次矩

$$\mathrm{E}(X^k) = \frac{1}{i^k}\frac{d^k}{dt^k}\phi_X(0).$$

反过来说，如果X有第k个有限时刻，那么$j_X(t)$是k次可微，上式成立。

5.4.2 独立性

期望值为两个随机变量的独立性提供了一个有用的标准。

定理5.18

随机变量X、Y是独立的，当且仅当

$$\mathrm{E}(f(X)g(Y)) = \mathrm{E}(f(X))\mathrm{E}(g(Y)) \tag{5.4}$$

对所有Borel可测量的有界函数f, g都成立。

证明

假设(5.4)成立，并取任何博勒集B_1, B_2。让$f = 1_{B_1}, g = 1_{B_2}$并应用（5.4）得到

$$\int_\Omega 1_{B_1}(X(\omega))1_{B_2}(Y(\omega))\, dP(\omega) = \int_\Omega 1_{B_1}(X(\omega))\, dP(\omega) \int_\Omega 1_{B_2}(Y(\omega))\, dP(\omega)$$

左手边等于

$$\int_\Omega 1_{B_1 \times B_2}(X(\omega), Y(\omega))\, dP(\omega) = P((X \in B_1) \cap (Y \in B_2))、$$

而右手边是$P(X \in B_1)P(Y \in B_2)$，从而证明了X和Y的不相干性。

现在假设X，Y是独立的。那么(5.4)对于$f = 1_{B_1}, g = 1_{B_2}, B_1, B_2$ Borel集，通过上面的论证而成立。通过线性，我们将该公式扩展到简单函数：$\varphi = \sum b_i 1_{B_i}, \psi = \sum_j c_j 1_{C_j}$、

$$\begin{aligned}
\mathrm{E}(\varphi(X)\psi(Y)) &= \mathrm{E}\left(\sum b_i 1_{B_i}(X) \sum c_j 1_{C_j}(Y)\right) \\
&= \sum_{i,j} b_i c_j\, \mathrm{E}(1_{B_i}(X)1_{C_j}(Y)) \\
&= \sum_{i,j} b_i c_j\, \mathrm{E}(1_{B_i}(X))\mathrm{E}(1_{C_j}(Y)) \\
&= \sum_i b_i\, \mathrm{E}(1_{B_i}(X)) \sum_j c_j\, \mathrm{E}(1_{C_j}(Y)) \\
&= \mathrm{E}(\varphi(X))\mathrm{E}(\psi(Y))。
\end{aligned}$$

我们用简单的函数来近似一般的f, g，而支配的收敛定理（f, g是有界的）将该公式扩展到f, g。□

命题5.19

假设X、Y是独立的随机变量。证明如果E(X)=0、
E(Y)=0，则E(XY)=0。

提示 上述定理不能适用于$f(x)=x$, $g(x)=x$(这些函数是没有边界的)。所以需要一些近似的方法。

期望只不过是一个积分，所以数字(X, Y)=E(XY)是随机变量空间$L^2(\Omega)$中的内积，相对于P而言是平方整数。因此独立性意味着这个空间中的正交性。

如果一个随机变量的期望值是非零的，我们就会修改正交性的概念。我们的想法是，增加（或减少）一个数字并不破坏或改善独立性。

定义5.9

对于一个具有有限μ=E(X)的随机变量，我们写成$X_c = X$ - E(X)，我们称X_c为*中心*随机变量（显然E(X_c)=0）。X和Y的*协方差*被定义为

$$\mathrm{Cov}(X, Y) = (X_c, Y_c) = \mathrm{E}\ (X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\ .$$

*相关性*是X_c和Y_c之间角度的余弦：

$$\rho X_{,Y} = \frac{(X_c\ ,\ Y_c)}{\|X\|_2\|Y\|_2} = \frac{Cov(X, Y)}{\|X_2\|\ \|Y\ \|_2}.$$

如果ρ=0，我们说X，Y是*不相关的*。

请注意，一些初级代数给出了一个更方便的协方差表达：

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)。$$

因此，不相关的X，Y满足E(XY)=E(X)E(Y)。显然，独立的随机变量是不相关的；在定理5.18中，只要把f（x）=x E（X），g（x）- =x E（Y）。一般来说，反之亦然，尽管--正如我们将在第六章中看到的--它对高斯随机变量是成立的。

例5.7

让Ω=[-1，1]，具有Lebesgue度量：$P = \frac{1}{2} m|_{[-1,1]}$ ，$X = x$，$Y = x^2$ 。

那么E（X）=0，E（XY）=$\int_{-1}^{1} x^3\,dx = 0$，因此$Cov(X, Y) = 0$，因此$\rho_{X,Y} = 0$。然而，$X$，$Y$并不独立。直观上这是很清楚的，因为

$Y = X^2$，因此，X、Y中的每一个都是另一个的函数。具体来说，取 $A = B = [-\frac{1}{2}, \frac{1}{2}]$，并（按照定义3.3的要求）比较概率 $P(X^{-1}(A) \cap Y^{-1}(A))$ 和$P(X^{-1}(A))P(Y^{-1}(A))$。我们得到$X^{-1}(A) = A$、$Y^{-1}(A) = [-\frac{\sqrt{1}}{\sqrt{2}}, \frac{\sqrt{1}}{\sqrt{2}}]$，因此 $P(X^{-1}(A) \cap Y^{-1}(A)) = \frac{\sqrt{1}}{\sqrt{2}}$，$P(X^{-1}(A)) = \frac{1}{2}$、$p(y^{-1}(a)) = \frac{\sqrt{1}}{\sqrt{2}}$ 因此，X和Y并不独立。

练习5.13

如果$X = 2Y + 1$，求相关的$\rho_{X,Y}$。

练习5.14

取$\Omega=[0，1]$，具有Lebesgue度量，让$X(\omega)=\sin 2\pi\omega$，$Y(\omega)=\cos 2\pi\omega$。证明$X$、$Y$是不相关的，但不是独立的。

我们以两个进一步的应用来结束本节。

命题5.20

不相关的随机变量之和的方差是它们的方差之和：

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)。$$

提示 为了避免繁琐的符号，先证明两个随机变量的公式

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

使用公式 $\mathrm{Var}(X) = \mathrm{E}(X^2) - (EX)^2$。

命题5.21

假设X、Y是独立随机变量。那么我们有如下的特征函数公式：

$$\phi_{X+Y}(t) = \varphi_X(t)，\phi_Y(t)。$$

更一般地说，如果$X_1, \ldots, X_n$是独立的，那么

$$\phi_{X_1+\cdots+X_n}(t)=\phi_{X_1}(t)\cdots\phi_{X_n}(t)。$$

提示 使用特征函数的定义和定理5.18。

5.4.3 条件期望值（第一 建设）

在第5.2.3节中进行的完全内积空间的正交投影的构造，使我们能够预览现代概率论中也许最重要的概念：一个可测量的整数随机变量X的条件期望，给定一个包含在其中的σ场（即，使其中的每个集合也属于）。我们将在第七章中详细研究这个想法，在那里我们还将通过参考更熟悉的概念来证明下面的定义，但对于任何可整数的X，用我们手头的工具可以实现条件期望值的构造，作为一个可测量的随机变量。我们的论证在很大程度上归功于[12]中给出的优雅构造。

F G F G

定义5.10

假设(Ω, F, P)是一个概率空间，并假设G是F的一个子σ场，给定$X \in L^1(\Omega, F, P) = L^1(F)$，存在$Y \in L^1(\Omega, G, P) = L^1(G)$，这样即$\int_G Y\, dP = \int_G X\, dP$ 对于每一个$G \in G$.我们写成$Y = E(X|G)$，并称Y是X的条件期望，给定.这些条件唯一地定义了Y，直到P-null集。

定理4.15，适用于P而不是m，G而不是M，意味着Y是P-a.s.唯一的：如果$Z \in L^1(G)$也满足$\int_G Z\, dP = \int_G X\, dP$，因为每一个$G \in G$，那么$Z = Y$ P-a.s.这通常通过说Y是$E(X|G)$ 的一个*版本来*表达：根据$L^1(\Omega, G, P)$的定义，唯一性要求是所有版本都属于$L^1(\Omega, G, P)$中的同一个等价类。

当且仅当$P(\{\omega \in \Omega : f(\omega) = g(\omega)\}) = 1$的等价关系$f \sim g$。按照我们的惯例（第5.1节），我们将继续用函数而不是用等价类来工作。

为了构建Y，我们首先将注意力限制在$X \in L^2(F) = L^2(\Omega, F, P)$。根据定理5.11，内积空间$H = L^2(F)$是完整的，而矢量子空间$K = L^2(G)$是一个完整的子空间。因此5.2.3节中正交投影的构造适用，并提供了L的一个元素$^2(G)$，根据我们的惯例，我们将其表示为一个函数$Y \in L^2(G)$，这样$(X - Y)$是正交于K的。产品在L^2这意味着

$$(X - Y, Z) = \int_\Omega (X - Y) Z\, dP = 0$$

对于每一个$Z \in L^2(G)$。特别是，由于$1_G \in L^2(G)$对于每一个$G \in G$，我们有$\int_G Y\, dP = \int_G X\, dP$。

为了构建一个任意的X的$Y \in L^1(\mathcal{G})$，我们分四个阶段进行：

(i) 首先，请注意，如果结果已被证明为非负函数在$L^1(\mathcal{F})$中，我们可以考虑$X = X^+ - X^-$。根据假设，有Y^+，$Y^- \in L^1(\mathcal{G})$，这样对于$G \in \mathcal{G}$来说，$\int_G Y^+ \, dP = \int_G X^+ \, dP$和$\int_G Y^- \, dP = \int_G X^- \, dP$。两边相减，我们得到$\int_G Y \, dP = \int_G X \, dP$，其中$Y = Y^+ - Y^- \in L^1(\mathcal{G})$。所以我们只需要验证非负整数$X$的结果。

(ii) 其次，如果一个随机变量Z是有界的，那么它在$L^2(\mathcal{F})$，因为$P(\Omega)$是有限的。因此，Z有一个条件期望，也就是说，有存在$W \in L^2(\mathcal{G})$，使$\int_G W \, dP = \int_G Z \, dP$，对$G \in \mathcal{G}$而言。另外：如果$Z \geq 0$，则$W \geq 0$ P-a.s.为了看到这一点，假设W取负值，正值概率。那么，存在$n \geq 1$，使得集合$G = \{W < -\frac{1}{n}\} \in \mathcal{G}$有$P(G) > 0$.因此$\int_G W \, dP < -\frac{1}{n} P(G) < 0$. 但$\int_G W \, dP = \int_G Z \, dP \geq 0$，因为$Z \geq 0$（命题4.11）。矛盾之处表明，$W \geq 0$，$P$-a.s.

(iii) 现在在$L^1(\mathcal{F})$中取一个任意的$X \geq 0$，对于$n \geq 1$，设$X_n = \min(X, n)$。那么X_n是有界的和非负的，所以第（二）部分适用于X_n、产生一个非负的$Y_n \in L^2(\mathcal{G})$，$\int_G Y_n \, dP = \int_G X_n \, dP$。由于$(X_n)$是随着n的增加，对于任何固定的$n$，有界随机变量$Z = X_{n+1} - X_n$是非负的，并且有一个条件期望值$W \geq 0$，如(ii)。但$X_{n+1}$和$X_n$也是如此，因此a.s.唯一性属性意味着$W = Y_{n+1} - Y_n$ P-a.s.。因此（Y_n）也随n增加（a.s）。

(iv) 最后，设$Y(\omega) = \limsup_{n\to\infty} Y_n(\omega)$，对于每个$\omega \in \Omega$。根据定理3.5 Y是可测量的，而且序列(Y_n)对Y来说是绝对增加的。此外，$0 \leq X \in L^1(\mathcal{F})$，对于$G \in \mathcal{G}$，$\int_G Y_n \, dP = \int_G X_n \, dP \leq \int_G X \, dP < \infty$，对于所有$n$。单调收敛定理表明，积分$(\int_G Y_n \, dP)_{n \geq 1}$增加到$\int_G Y \, dP$，而且最后的积分是有限的，所以$Y \in L^1(\mathcal{G})$。关于另一方面、$\int_G Y_n \, dP = \int_G X_n \, dP$，后者的积分也增加到$\int_G X \, dP$，所以我们已经证明：$\int_G Y \, dP = \int_G X \, dP$，对所有$G \in \mathcal{G}$。

备注5.5

请注意，由于正交投影Y可以最小化各点之间的距离。向量$X \in L^2(\mathcal{F})$和子空间$L^2(\mathcal{G})$，L^2 - norm，因此其平方为$\int_\Omega (X - Z)^2 \, dP$，对于$L^2(\mathcal{G})$中的元素来说是最小的，如果我们取$Z = Y$，写成这是用期望值来表示的：$E([X - E(X|\mathcal{G})]^2) = \inf\{E([X - Z]^2) : Z \in L^2(\mathcal{G})\}$。将σ场视为包含关于随机事件的 "信息"，我们可以将$E(X|\mathcal{G})$解释为在可测量的平方不可变函数类别中对X的 "最佳预测"，因为它最小化了最小均方差。

这个想法导致了在许多科学领域的许多有用的应用。

5.5 命题的证明

证明（命题5.3的证明）

假设$\int_D f(x)\,dx$是有限的。那么用$a \le \frac{1+a^2}{2}$（这由$(a-1)^2 \ge 0$)，我们有

$$\int_D f(x)\,dx \le \int_D 1\,dx + \int_D f^2(x)\,dx = m(D) + \int_D f^2(x)\,d(x) < \infty。$$

□

证明（命题5.4的证明）

我们用积分$(f+g, h) = \int (f(x)+g(x))h(x)\,dx$的线性来验证前两个属性

$$= \int f(x)h(x)\,dx + \int g(x)h(x)\,dx$$

$$= (f, h) + (g, h)、$$

$$(cf, h) = \int cf(x)h(x)\,dx = c\int f(x)h(x)\,dx = c(f, h)。$$

对称性是显而易见的，因为在积分下$f(x)g(x) = g(x)f(x)$。 □

证明（命题5.5的证明）

平行四边形定律：$\|h_1 + h_2\|^2 = (h_1 + h_2, h_1 + h_2) = (h_1, h_1) + (h_1, h_2) + (h_2, h_1) + (h_2, h_2)$，$\|h_1 - h_2\|^2 = (h_1 - h_2, h_1 - h_2) = (h_1, h_1) - (h_1, h_2) - (h_2, h_1) + (h_2, h_2)$, 加起来就可以得到结果。

偏振特性：减去上面的$\|h_1 + h_2\|^2 - \|h_1 - h_2\|^2 = 2(h_1, h_2) + 2(h_2, h_1)$，用$ih_2$替换$h_2$，得到$\|h_1 + ih_2\|^2 - \|h_1 - ih_2\|^2 = 2(h_1, ih_2) + 2(ih_2, h_1)$。将得到的表达式插入同位素的右侧。

的问题。在左边我们有$2[(h_1, h_2)+(h_2, h_1)+i(h_1, ih_2)+i(ih_2, h_1)] = 2[(h_1, h_2) + (h_2, h_1) + i(-i)(h_1, h_2) + i^2(h_2, h_1)] = 4(h_1, h_2)$。 □

证明（命题5.13的证明）

假设$E(X^n)<\infty$，这意味着$X\in L^n(\Omega)$，那么由于Ω的度量*是*有限的，我们可以应用定理5.12，所以X⍰$L^k(\Omega)$ 对于所有$k\leq n$。如果$E(X^n)=\infty$，那么对于$k\geq n$的$E(X^k)$也必须如此，因为否则$E(X^k)<\infty$将意味着$E(X^n)<\infty$，这是一个矛盾。 □

证明（命题5.14的证明）

使用二项式扩展，我们有

$$(X - \mu)^n = \sum_{i=0}^{n} \binom{n}{i} X^i (-\mu)^{n-i}、$$

因此，根据期望值的线性关系

$$E(X - \mu)^n = \sum_{0}^{n} \binom{n}{i} (-\mu)^{n-i} E(X^i)。$$

□

证明（命题5.15的证明）

我们有

$$E(X^n) = E((X - \mu + \mu)^n) \sum_{i=1}^{n} \binom{n}{i} E(X - \mu)^i \mu^{n-i}$$

□

证明（命题5.19的证明）

让$f_n(x) = \max\{-n, \min\{x, n\}\}$。根据定理5.18，由于$X$，$Y$是独立的，我们有$E(f_n(X)f_n(Y)) = E(f_n(X))E(f_n(Y))$。$X$和$Y$的整数性使我们能够进入极限，即右边的0。 □

证明（命题5.20的证明）

Let X, Y be uncorrelated random variables. Then

$$\begin{aligned}
\mathrm{Var}(X + Y) &= E((X + Y) - E(X + Y))^2 \\
&= e(x + y)^2 - (e(x) + e(y))^2 \\
&= ex^2 + 2e(xy) + ey^2 - (ex)^2 - 2e(x)e(y) - (ey)^2 \\
&= ex^2 - (ex)^2 + ey^2 - (ey)^2 + 2[e(xy) - e(x)e(y)] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

因为$E(XY)=E(X)E(Y)$。对于n个随机变量的一般情况，可通过以下方式得出

诱导或通过重复使用两个人的公式：

$$\begin{aligned}\operatorname{Var}(X_1 + \cdots + X_n) &= \operatorname{Var}(X_1 + [X_2 + \cdots + X_n])\\ &= \operatorname{Var}(X_1) + \operatorname{Var}(X_2 + [X_3 + \cdots + X_n])\\ &\cdots\\ &= \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n)。\end{aligned}$$

□

证明（命题5.21的证明）

根据定义

$$\begin{aligned}\phi_{X+Y}(t) &= E(e^{it(X+Y)})\\ &= E(e^{itX}e^{itY})\\ &= E(e^{itX})E(e^{itY}) \quad \text{(根据定理5.18)}\\ &= \phi_X(t)\phi_Y(t)。\end{aligned}$$

对n个组件的概括是直接的--归纳或一步步应用两个组件的结果。

□

6

产品措施

6.1 多维Lebesgue 措施

在第二章中，我们构建了实线上的Lebesgue度量。这方面的基础是区间的长度概念。现在考虑用平面R^2 来代替R。这里的区间我们理解为任何类型的矩形：

$$R = I_1 \times I_2$$

其中I_1 , I_2 是任何区间。一个长方形的 "长度 "就是它的面积

$$a(R) = l(I_1) \times l(I_2) 。$$

空集的概念与一维的情况一样被引入。和以前一样，可数集是空集。值得注意的是，在平面上，我们有更复杂的空集，例如，一条线段或一个函数的图形。

整个构造没有变化，结果是在矩形产生的σ*场上*定义的平面上的勒贝斯格测量m_2 。

有一点很微妙，它清楚地说明了与线性测量的区别：任何形式的$A \times \{a\}$, $a \in R$ 的集合都是空的，因此在平面上是可测量的Lebesgue。一个有趣的例子是，当A是一个
实线上的不可测量的集合!

接下来，我们考虑R^3 。这里的 "区间 "指的是一个立方体，"长度 "是其体积：

$c = i_1 \times i_2 \times i_3$ 、

159

$$v(C) = l(I_1) \times l(I_2) \times l(I_3)。$$

现在曲面是空集的例子。按照同样的构造，我们得到Lebesgue度量m_3 在R^3 。

最后，我们考虑R^n （这包括n=1，2，3的特殊情况，所以对于一个真正的数学家，这是唯一值得关注的案例，因为它涵盖了所有的其他）。'间隔'现在是n*维的*立方体：

$$i = i_1 \times - - - \times i_n$$

和广义的 "长度 "是由

$$l(I) = l(I_1) \times . . . \times l(I_n)。$$

空集的一个有趣的例子是超平面。

上述内容为下文提供了动力。多维Lebesgue度量将在下面的考虑中再次出现，在那里我们将与一般的度量空间一起工作。考虑到我们主要对概率的应用感兴趣，我们坚持使用概率论的符号。

6.2 产品 σ - 领域

让(Ω_1, F_1, P_1), (Ω_2, F_2, P_2) 是两个度量空间。把

$$\Omega = \Omega_1 \times \Omega_2。$$

我们希望在Ω上定义一个措施P，与Ω上的措施 "一致 "$_1$, Ω_2。

在构建P之前，我们需要指定它的领域，也就是在P上的一个σ场。Ω.

定义6.1

让F成为包含 "矩形 "$A_1 \times A$的Ω子集的最小的σ场。$_2$
对于所有的$A_1 \in F_1$, $A_2 \in F_2$。我们称 F的*乘积σ-场* F_1和 F_2.在其他

话说，乘积σ场是由集合族（'矩形'）产生的

$$r = \{a_1 \times a_2 : a_1 \in f_1, a_2 \in f_2\}。$$

用于乘积σ场的符号是简单的：$F = F_1 \times F_2$。

有许多方法可以产生相同的乘积σ-*场*，每一种方法都会在后面证明是有用的。

定理6.1

(i) 乘积 σ-场 $F_1\times F_2$是由一系列的集合（"圆柱体 "或 "条带"）产生的

$$c = \{a_1 \times \omega_2 : a_1 \in f_1\} \cup \{\omega_1 \times a_2 : a_2 \in f_2\}。$$

(ii) 乘积σ场F是最小的σ场，使得投影

$$Pr_1 : \omega \to \omega_1、\quad Pr_1(\omega_1, \omega_2) = \omega_1$$

$$Pr_2 : \omega \to \omega_2、\quad Pr_2(\omega_1, \omega_2) = \omega_2$$

是可以衡量的。

证明

回顾一下，我们用σ(E)表示由族E产生的σ 场。

显然，C包含在R中，因此C产生的σ场比R产生的σ 场要小、

$$a_1 \times a_2 = (a_1 \times \omega_2) \cap (\omega_1 \times a_2)$$

因此，矩形属于由圆柱体产生的σ -场： R⊂σ(C)。这意味着σ(R)⊂σ(σ(C))=σ(C)，从而完成了(i)的证明。

对于(ii)，请注意：

$$Pr1^{-1}(A_1) = A_1 \times \Omega_2、\qquad Pr2^{-1}(A_2) = \Omega_1 \times A_2$$

因此，根据(i)，这些投影是可测的。但是，使它们都可测量的最小σ*场*是包含圆柱体的最小σ*场*，这又是F的（i）。 □

考虑Ω_1，Ω_2 =R， F_1 =F_2 =B Borel集的一个特殊情况。那么我们有两种方法可以产生 $F =_2 B$平面上的博莱尔集：我们可以使用Borel集的乘积族或区间的乘积族。下面的结果表明，它们给出了相同的R的子集集合[2]。

命题6.2

所产生的σ-场

$$r = \{b_1 \times b_2 : b_1, b_2 \in b\}、$$

$$I = \{I_1 \times I_2 : I_1, I_2 \text{ 是区间}\}。$$

是一样的。

提示 使用前述定理的证明思路。

我们可以很容易地推广到n个因子：假设我们得到n个度量空间（Ω_i $,_i, P_i$），$\Gamma = 1, \ldots, n$，那么在$\Omega = \Omega_1$ ×- - - Ω_n 中的乘积σ*场就是*由集合生成的σ*场*

$$\{a_1 \times - - - \times a_n : a_i \in f_i\}。$$

6.3 产品的构建 措施

回顾一下，(Ω_1, F_1, P_1), (Ω_2, F_2, P_2) 是任意的度量空间。我们将在$\Omega = \Omega_1 \times \Omega_2$ 上构造一个度量P，它由P_1, P_2 以自然方式决定。这里需要一个技术假设：P_1 和P_2 被认为是σ-*无限的*，也就是说，有一个可测量集A_n 的序列，其$_{n \prec}$ $A_n = \Omega_1$, $P_1(A_n)$是有限的（对P_2, Ω_2 也是如此）。当然，对于概率度量和Lebesgue度量也是如此（在后一种情况下，例如$A_n = [-n, n]$ 就可以了）。为简单起见，我们假设P_1, P_2 是有限的。(扩展到σ-finite的情况是通过一个常规的极限通道$n \to \infty$的结果得到的。
得到的是对A的限制措施$_n$)。

多维Lebesgue度量的构造所提供的动机给了P以下的自然条件：

$$p(a_1 \times a_2) = p_1(a_1)\, p_2(a)_2 \tag{6.1}$$

对于$A_1 \in F_1, A_2 \in F_2$。

我们想把（6.1）推广到来自乘积σ场的所有集合。为此，我们引入了$\Omega_1 \times \Omega_2$ 的子集A的*截面*概念：对于$\Omega_2 \in \Omega_2$、

$$A_\omega 2 = \{\omega_1 \,?\, \omega_1 : (\omega_1, \omega_2) \in A\} \subset \omega_1。$$

对于$\omega_1 \in \omega_1$，也进行了类似的构造：

$$A_{\omega 1} = \{\omega_2 \in \omega_2 : (\omega_1, \omega_2) \in A\} \subset \omega_2。$$

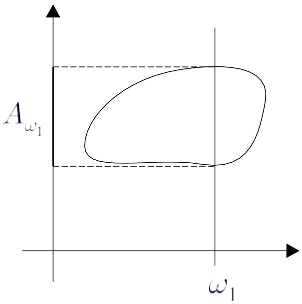

图6.1 ω_2 集的部分

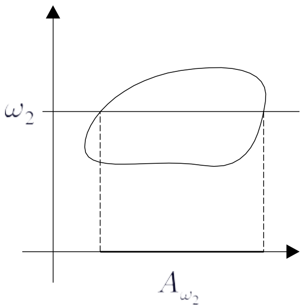

图6.2 ω_1 集的部分

定理6.3

如果A在乘积σ场F中，那么对于每个ω_2，A_{ω} 2∈F$_1$，对于每个ω_1，A_{ω}
$_1$ ⍰ F$_2$。

证明

让

$$G = \{A \in F : 对于所有\omega_2，A_{\omega}\ 2 \in F_1\}。$$

如果我们证明 G是一个包含矩形的σ*场*，那么 G = F 因为 F 是具有这一性质的最小的σ*场*。

If $A = A_1 \times A_2$, $A_1 \in F_1$, then

$$A\omega_2 = \begin{cases} A_1 \text{ 如果 } \omega_2 ⍰ A_2 \\ \varnothing \text{ 如果 } \omega_2 ⍰/ A_2 \end{cases}$$

是在F$_1$，所以矩形是在G中。

如果$A\in G$，那么$\Omega A\in G$，因为

$$(\omega \backslash A)_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in (\omega \backslash A)\}$$
$$= \Omega_1 \{\omega_1 : (\omega_1, \omega_2) \in A\}$$
$$= \Omega_1 / A_{\omega_2}。$$

最后，让$A_n \in G$. 因为

$$\left(\bigcup_{n=1}^{\infty} A_n\right)_{\omega_2} = \bigcup_{n=1}^{\infty} (A_n)_{\omega_2},$$

联盟$_{n=1}^{\infty}$ A_n 也在G中。

A_{ω} 1的证明是完全相同的。 □

If $A = A_1 \times A_2$, then the function $\omega_2 \mapsto P(A_{\omega_2})$ is a step function:

$$P(A_{2\omega}) = \begin{cases} P(A_1) & 如果\omega_2 \in A_2 \\ 0 & 如果\omega_2 \notin A_2 \end{cases}$$

因此，我们有

$$p(a) = p_1(a_1)p_2(a_2) = \int_{\Omega 2} P(A_{\omega} 2)\, dP_2(\omega_2)。$$

这促使了一般的公式；对于任何A，我们写成

$$P(A) = \int_{\Omega 2} P_1(A_{\omega} 2)\, dP_2(\omega_2)。 \tag{6.2}$$

我们称P为*乘积度量*，有时我们会用$P=P_1 \times P_2$来表示它。我们已经知道，$P_1(A_{\omega} 2)$是有意义的，因为$A_{2\omega} \in F_1$，如前所示。为了使积分有意义，我们需要更多：

定理6.4

假设P_1，P_2是有限的。如果$A\in F$，那么函数

$$\omega_2 \mapsto P_1(A_{\omega} 2)、\qquad \omega_1 \mapsto P_2(A_{\omega} 1)$$

对F_2，F_1，分别是可测量的，并且

$$\int_{\Omega 1} P_1(A_{\omega} 2)\, dP_2(\omega_2) = \int_{\Omega 2} P_2(A_{\omega} 1)\, dP_1(\omega_1)。 \tag{6.3}$$

在我们证明这个定理之前，我们允许自己对产生σ场的方式进行一次离题。这将大大简化几个证明。给定一个集合族A，让它成为包含F的最小的σ场。假设A它是一个场，即它对于有限联合和差分来说是封闭的：$A, B \in \mathcal{A}$ 意味着$A \cup B$，$A \setminus B \in \mathcal{A}$。那么我们有另一种方法，通过所谓的单调类来描述σ-场的特征。

定义6.2

*单调类*是指在可数的递增联合体下封闭的集合家族。

集和递减集的可数交集。就是说 G 是一个单调的类，如果

$$a_1 \subset a_2 \subset \ldots, \quad A_i \in G \quad \Rightarrow \quad \bigcup_{i=1}^{\infty} A_i \in G、$$

$$a_1 \supset a_2 \supset \ldots, \quad A_i \in G \quad \Rightarrow \quad \bigcap_{i=1}^{\infty} A_i \in G。$$

Lemma 6.5（单调类定理）

包含场A的最小单调类G_A，与σ场重合。
F_A 由A产生。

证明

σ-场是一个单调类，所以$G_A \subset F_A$（因为G_A 是包含A的最小单调类）。

为了证明反过来，我们首先证明G_A 是一个场。集合的家族

$$\{A : A^c \in G_A\}$$

包含 A 因为$A \in \mathcal{A}$ 意味着$A^c \in \mathcal{A} \subset G_A$。我们观察到它是一个单调的类。假设$A_1 \subset A_2 \subset \ldots$是这样的：$A_i^c \in G_A$。我们必须证明$(\bigcup A_i)^c \in G_A$。我们有 $A_1^c \supset A_2^c \supset \ldots$，因此$\bigcap A_i^c \in G_A$ 因为G_A 是一个单调的类。根据德摩根定律$\bigcap A_i^c = (\bigcup A_i)^c$，所以联盟满足于所需的条件。交叉点的证明也类似：$A_1 \supset A_2 \supset \ldots$ 意味着$A_1^c \subset A_2^c \subset \ldots$，因此$\bigcup A_i^c \in G_A$，所以$(\bigcap A_i)^c = \bigcup A_i^c$ 也属于到G_A。

我们的结论是

$$G_A \subset \{A : A^c \in G_A\}$$

因此，G_A 在取补数方面是封闭的。

现在考虑联合体。首先固定$A\in$A并考虑

$$\{b : a \cup b \in g_A\}。$$

这个家族包含（如果$A\ B$，则$A\ B\ \cup\ \in A \subset G_A$)，并且是一个单调的类。因为⊊让$B\ldots$是这样的：$A\cup B$ G $\cup_A$.那么$A\ B\ A\ B\ \cup\subset_2$ ..因此$(A\ B\ )_{iA}$，因此$A\ B\ _{iA}$。类似的论证也适用于递减的集合链的交集，所以对于这个固定的A

$$g_A \,\{b : a \cup b \in g_A\}。$$

这意味着对于AA和$B$$G_A$，我们有$A\cup B$$G_A$。现在取任意的$A\in G_A$。根据我们刚才的观察、

$$a\subset\{b : a\cup b\in g\}_A$$

并通过与之前相同的论证，后者的家族是一个单调的类。所以

$$g_A \,\{b : a \cup b \in g\}_A$$

这次是针对一般的A，这就完成了G_A是一个场的证明。

现在，在表明G_A是一个场之后，我们观察到它是一个σ*场*。这是显而易见的，因为对于一个序列$A_i\in G_A$，我们有$A_1\subset A_1\cup A_2\ldots$，它们都在$G_A$（根据场属性），它们的联合也是如此（因为$G_A$是一个单调的类）。

因此，G_A是一个包含A的σ场，所以它比A生成的σ场要大：

$$F_A \subset G_A$$

这就完成了证明。 □

上面介绍的R矩形族不是一个场，所以它不能用于上述结果。因此，我们认为是所有不相交的矩形的联合体的家族A。

证明（定理6.4的证明）

撰写

$$G = nA : \omega_2 \mapsto P_1(A_{\omega_2})、\omega_1 \mapsto P_2(A_{\omega_1})\text{是可测量的，并且}$$

$$\int_{\Omega_2} P_1(A_{\omega_2})\,dP_2(\omega_2) = \int_{\Omega_1} P_2(A_{\omega_1})\,dP_1(\omega_1)\text{）。}$$

证明的思路是这样的。首先我们表明R⊂G，然后A⊂G，最后我们表明G是一个单调的类。根据定理6.5，G=F，这意味着定理的要求对F的所有集合

都成立。

如果A是一个矩形，$A = A_1 \times A_2$，那么正如我们之前注意到的，$\omega_2 \mapsto P_1(A_{\omega 2}), \omega_1 \mapsto P_2(A_{\omega 1})$是乘以一些常数的指示函数，（6.3）成立，每边等于$P_1(A_1)P_2(A_2)$。

接下来让$A = (A_1 \times A_2) \cup (B_1 \times B_2)$是不相交的矩形的联合。不相交意味着要么$A_1 B_1 = \varnothing$，要么$A_2 B_2 = \varnothing$。假设是前者、

比如说。然后

$$A_{\omega 2} = \begin{cases} A_1 \cup B_1 & \text{if } \omega_2 \in A_2 \cap B_2 \\ A_1 & \text{if } \omega_2 \in A_2 \setminus B_2 \\ B_1 & \text{if } \omega_2 \in B_2 \setminus A_2 \\ \varnothing & \text{否则} \end{cases}$$

和

$$\begin{aligned} \int_{\Omega_2} P_1(A_{\omega 2})\, dP_2(\omega_2) &= [P_1(A_1) + P_1(B_1)]P_2(A_2 \cap B_2) \\ &\quad + p_1(a_1)p_2(a_2 / b_2) + p_1(b_1)p_2(b_2 / a_2) \\ &= p_1(a_1)[p_2(a_2 \cap b_2) + p_2(a_2 \cap b_2)] \\ &\quad + p_1(b_1)[p_2(a_2 \cap b_2) + p_2(b_2 \cap a_2)] \\ &= p_1(a_1)p_2(a_2) + p_1(b_1)p_2(b_2)。 \end{aligned}$$

另一方面

$$A_{\omega 1} = \begin{cases} A_2 & \text{if } \omega_1 \in A_1 \\ B_2 & \text{如果 } \omega_1 \in B_1 \\ \varnothing & \text{否则} \end{cases}$$

和

$$\int_{\Omega_1} P_2(A_{\omega 1})\, dP_1(\omega_1) = P_1(A_1)P_2(A_2) + P_1(B_1)P_2(B_2)$$

和以前一样。

有限多个矩形的一般情况也可以用同样的方法证明。这很容易，但很繁琐，我们跳过这个论证，希望对两个矩形的详细介绍足以在一般情况下指导读者。仍然是这样，函数$\omega_2 \mapsto P_1(A_{\omega 2}), \omega_1 \mapsto P_2(A_{\omega 1})$都是简单函数，所以(6.3)的验证只是简单代数。

剩下的就是验证G是一个单调的类。让$A_1 \subset A_2 \subset \dots$是来自G的集合；因此函数$\omega_2 \mapsto P_1((A)_{i\omega 2}), \omega_1 \mapsto P_2((A)_{i\omega 1})$是可测的。由于$(A)_{i\omega 2}, (A)_{i\omega 1}$的部分是增加的，所以它们随$i$增加。如果$i \to \infty$，然后

$$p_1((a)_{i\omega 2}) \to p_1\Big(\bigcup_i (a)_{i\omega 2}\Big) = p_1\Big(\big(\bigcup_i a\big)_{i\omega 2}\Big)$$

$2 \mapsto 1$ ...

因此，函数ω P ((A)是可测的。同样的论证表明，函数ω_1 '$\to P_2$ $((S_i A)_{i\omega}$ 1) 是可测的。等式(6.3)

对每个i都成立，根据单调收敛定理，S它在极限中被保留下来。因此，(6.3)对联合体A成立$_i$。

对于相交点来说，论证是类似的。有关序列是递增的；函数$\omega_2 P_1((_iA)_{i\omega 2}), \omega_1 P_2((_iA)_{i\omega 1})$的极限是可以测量的，根据单调收敛定理，（6.3）成立。

定理6.6

假设P_1, P_2是有限度量。由（6.2）给出的集合函数P是可数加性的。任何与P在矩形上重合的其他度量都等于P在积σ场上的度量。

证明

让$A_i \in$F是成对不相交的。那么（A）$_{i\omega 2}$也是成对不相交的，并且

$$\begin{aligned}
P(\bigcup_i A_i) &= \int_{\Omega_2} P_1((\bigcup_i A)_{i\omega 2})\,dP_2(\omega)_2 \\
&= \int_{\Omega_2} P_1(\bigcup_i (A)_{i\omega 2})\,dP_2(\omega)_2 \\
&= \int_{\Omega_2} \sum_i P_1((A)_{i\omega 2})\,dP_2(\omega)_2 \\
&= \sum_i \int_{\Omega_2} P_1((A)_{i\omega 2})\,dP_2(\omega)_2 \\
&= \sum_i P(A)_i
\end{aligned}$$

其中我们采用了联合的部分是部分的联合的事实（见图6.3）和贝波-列维定理。

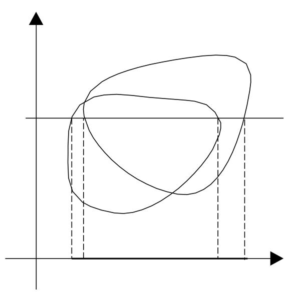

图6.3 工会的截面

为了唯一性，让Q是定义在乘积σ场F上的度量，这样P（$A_1 \times A_2$） = Q（$A_1 \times A_2$），$A_1 \in F_1$，$A_2 \in F_2$。让

$$h = \{a \subset \omega_1 \times \omega_2 : a \in f, p(a) = q(a)\}.$$

根据假设，这个家族包含所有的矩形。根据P和Q的可加性，它包含不相交的矩形的联合体；换句话说，它包含领域.

H

A

剩下的就是证明它是一个单调类，因为这样它就和Lemma 6.5相吻合

F

。这是很简单的，再次利用P和

Q是措施。如果$A_1 \subset A_2 \subset .. ., A_i \in H$，那么$P(\bigcup A_i) = Q(\bigcup A_i)$，$P(\bigcup A_i) = \lim P(A_i)$，$Q(\bigcup A_i) = \lim Q(A_i)$，因此$P(A_i) = Q(A_i)$，这意味着H相对于单调的联合而言是封闭的。关于单调相交的论证也完全相同：如果$A_1 \supset A_2 \supset ..$，那么$P(\bigcap A_i) = \lim P(A_i)$、

$Q(\bigcap A_i)=\lim Q(A_i)$，因此$P(A_i)=Q(A_i)$意味着$P(\bigcap A_i)=Q(\bigcap A_i)$。 □

备注6.1

定理6.6证明的唯一性部分说明了一个重要的技术诀窍：为了证明σ场上的两个度量是重合的，只要证明它们在该σ *场*的生成集上是重合的就可以了，这是单调类定理的一个应用。

作为定理6.4的一个直接结果，我们有

$$P(A) = \int_{\Omega 1} P_2(A_{\omega} 1)\, dP_1(\omega_1)。$$

由Lebesgue可测量集的σ场建立的乘积σ场M×M的完成度是σ场$_2$，在此基础上定义了m_2。 M

很容易看出，二维的Lebesgue度量与一维的乘积的完成度相吻合：$m_2 = c(m \times m)$。首先，它们在由区间构建的矩形上一致。因此，他们同意由这些矩形产生的σ 场，也就是平面上的博勒尔σ场。Borel集的完成给出了Lebesgue可测量集的σ场，其方式与一维的情况相同。

6.4 富比尼的 定理

我们希望整合定义在空间的乘积上的函数（Ω_1, F_1, P_1），（$\Omega_2, {}_2, P_2$），利用关Fl于度量$\mathrm{P}_1, \mathrm{P}_2$

单独。

我们首先解决可测量性的问题。

定理6.7

如果一个非负的函数f：$\Omega_1 \times \Omega_2 \rightarrow$ R对于$J_1 \times J_2$来说是可测的，那么对于每一个$\omega_1 \in \Omega_1$的函数（我们将称之为f*的*一个*部分*）$\omega_2 \mapsto f(\omega_1$、$\omega_2$）是$J_2$-可测量的，而对于每个$\omega_2$ ⍰ ω_2，$\omega_1 \mapsto f$（ω_1，ω_2）是J_1-可测量的。

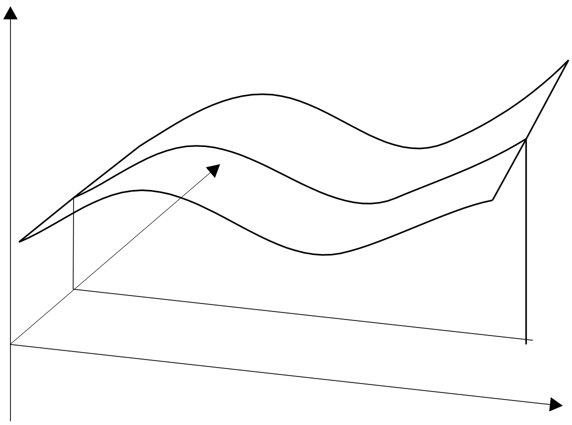

图6.4 f的截面

证明

首先，我们以类似于命题4.10的方式用简单的函数对f进行近似；我们写道

$$f_n(\omega_1,\omega_2)=\begin{cases}\frac{k}{n} & \text{如果} f(\omega_1,\omega_2)\in[\frac{k}{n},\frac{k+1}{n}),\ k<n^2\\ n & \text{如果} f(\omega_1,\omega_2)>n\end{cases}$$

而当$n\rightarrow\infty$时，f_n 3 f。

简单可测函数的截面是简单和可测的。这对上面观察到的指标函数来说是很清楚的，接下来我们利用和的部分是部分之和的事实。

最后，很明显，f的截面$_n$，收敛于f的截面，由于可测性在极限中被保留，该定理被证明。□

推论6.8

职能

$$\omega_1 \mapsto \int_\Omega f(\omega_1, \omega_2)\, dP_2(\omega_2)、\quad \omega_2 \mapsto \int_\Omega f(\omega_1, \omega_2)\, dP_1(\omega_1)$$

分别是F_1, F_2-可测量的。

证明

由于相关函数的可测量性，积分可以被抽取（可能是无限的）。根据单调收敛定理，它们是.的部分的积分的极限。这些积分 $\int_{\Omega_1} f_n(\omega_1, \omega_2)\, dP_1(\omega_1)$，$\int_{\Omega_2} f_n(\omega_1, \omega_2)\, dP_2(\omega_2)$是简单的函数，因此极限是可测量的。
能够 □

定理6.9

设f是一个定义在$\Omega_1 \times \Omega_2$ 的可测量非负函数。那么

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2)\, d(P_1 \times P_2)(\omega_1, \omega_2) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, dP_2(\omega_2)\, dP_1(\omega_1)$$

$$= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\, dP_1(\omega_1)\, dP_2(\omega_2). \tag{6.4}$$

证明

对于矩形A的指标函数$_1 \times A_2$，(6.4)的每条边正好成为
$p_1(a_1)p_2(a_2)$。那么根据积分的可加性，该公式对简单的
函数。通过简单函数对任何可测量的*f进行*单调的逼近，使我们能够将这个公式扩展到一般情况。 □

定理6.10（福比尼定理）

如果$f \in L^1(\Omega_1 \times \Omega_2)$，那么在适当的空间中，各部分是可整定的，函数

$$\omega_1 \mapsto \int_\Omega f(\omega_1, \omega_2)\, dP_2(\omega_2)、\quad \omega_2 \mapsto \int_\Omega f(\omega_1, \omega_2)\, dP_1(\omega_1)$$

分别在$L^1(\Omega_1)$, $L^1(\Omega_2)$中，且(6.4)成立：简而言之，即为

$$\int_{\Omega_1\times\Omega_2} f\,d(P_1\times P_2) = \int_{\Omega_1}\int_{\Omega_2} f\,dP_2\,dP_1 = \int_{\Omega_2}\int_{\Omega_1} f\,dP_1\,dP_2$$
。

证明

通过分解$f = f^+ - f^-$ 和对非负函数证明的结果，这一关系是直接的。这些积分是有限的，因为如果$f\in L^1$，那么f^+，$f^- \in L^1$，右边的所有积分都是有限的。 □

备注6.2

整个过程可以扩展到任意数量的空间的乘积。特别是，我们有一种方法来构建n维的Lebesgue测量，作为一维Lebesgue测量的n *个副本的*乘积的完成。

例6.1

Let $\Omega_1 = \Omega_2 = [0, 1]$, $P_1 = P_2 = m_{[0,1]}$,

$$f(x, y) = \begin{cases} \frac{1}{x} & \text{如果}0 < y < x < 1 \\ 0 & \text{否则。} \end{cases}$$

我们将看到，f在广场上的积分是无限的。为此，我们取一个由f支配的非负的简单函数并计算其积分。设

如果$f(x, y) \in [n, n+1]$，则$j(x, y) = n$，如果$x>y$，$x\in \left(\frac{1}{n+1}, \frac{1}{n}\right)$。

这个集合的面积是$\frac{1}{2}\left(\frac{1}{n} - \frac{1}{n+1}\right)$，而

$$\int_{[0,1]^2} \phi \, dm_2 = \sum_{n=1}^{\infty} n \frac{1}{2}\left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^{\infty} \frac{1}{2} \frac{1}{n+1} = \infty.$$

因此，该函数

$$g(x, y) = \begin{cases} \frac{1}{x^2} & \text{如果}0 < y < x < 1 \\ -\frac{1}{y^2} & \text{如果}0 < x < y < 1 \\ 0 & \text{否则} \end{cases}$$

是不可积分的，因为g的积分$^+$，是无限的（g的积分$^-$，也是如此）。

练习6.1

对于上述例子中的g，表明

$$\int_0^1 \int_0^1 g(x, y)\, dx\, dy = -1$$

、

$$\int_0^1 \int_0^1 g(x, y)\, dy\, dx = 1$$

这表明，如果违反了富比尼定理的条件，迭代积分可能不相等。

下面的命题为积计量和福比尼定理的许多应用开辟了道路。

命题6.11

让f： R $\to$ R是可测量的和正的。考虑上半平面中位于f的图形之下的所有点的集合：

$$A_f = \{(x, y)：0 \leq y < f(x)\}。$$

证明A_f是m_2-可测量的，并且$m_2(A_f) = \int f(x)\,dx$。

提示 对于可测量性，"用矩形填充 "A_f，使用近似值为
f的简单函数。然后应用乘积测量的定义。

练习6.2

计算 $\int_{[0,3]\times[-1,2]} x^2 y\,dm_2$。

练习6.3

计算椭圆内区域的面积$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$。

6.5 P robability

6.5.1 联合 分布

让X，Y是两个定义在同一概率空间（Ω，F，P）上的随机变量。考虑随机向量

$$(x, y): \omega \to r^2。$$

它的分布是为平面上的博莱尔集定义的度量，其定义为

$$p_{(x,y)}(b) = p((x, y) \in b)、\ B \subset R^2。$$

如果这个措施可以写成

$$P_{(X,Y)}(B) = \int_B$$

$f_{(X,Y)}(x, y)\, dm_2(x, y)$

对于一些可整定的$f_{(X,Y)}$，那么我们说X，Y有一个*联合密度*。

联合分布决定了一维的Ran-dom变量X、Y的分布：

$$p_X(a) = p_{(X,Y)}(a \times \mathrm{r})、$$
$$p_Y(a) = p_{(X,Y)}(\mathrm{r} \times a)、$$

对于Borel $A \subset \mathbb{R}$，这些被称为*边际分布*。如果X、Y有一个联合密度，那么X和Y都是绝对连续的，其密度由以下公式给出

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy、$$
$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dx。$$

下面的例子表明，在一般情况下，反之亦然并不正确。

例6.2

让$\Omega=[0，1]$，P=$m_{[0,1]}$，让$(X，Y)(\omega)=(\omega，\omega)$。这个向量并不有密度，因为$P_{(X,Y)}(\{(x, y) : x = y\}) = 1$，对于任何可整定的函数$f : \mathbb{R}^2 \to \mathbb{R}$，$\int_{\{(x,y):x=y\}} f(x, y)\, dm_2(x, y) = 0$；一个矛盾。然而，该边际分布$P_X, P_Y$是绝对连续的，密度为

$f_X = f_Y = 1_{[0,1]}$。

例6.3

联合密度的一个简单例子是均匀密度：$f = \frac{1}{m(A)} 1_A$，博莱尔$A \subset \mathbb{R}^2$。一个特殊情况是$A = [0, 1] \times [0, 1]$，那么显然边缘密度是$1_{[0,1]}$。

练习6.4

取A为角在（0，1），（1，0），（2，1），（1，2）的正方形。求$f=1_A$的边际密度。

练习6.5

设$f_{X,Y}(x, y) = \frac{1}{50}(x^2 + y^2)$ 如果$0 < x < 2, 1 < y < 4$，否则为零。求P（$X+Y>4$），P（$Y>X$）。

二维高斯（正态）密度由以下公式给出

$$n(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\}。 \tag{6.5}$$

可以证明，ρ是X，Y，随机变量的相关性，其密度是n（x，y）的边际密度，（见[9]）。

联合密度使我们能够计算随机变量的各种函数的分布。这里有一个重要的例子。

定理6.12

如果X，Y有联合密度$f_{X,Y}$，那么它们之和的密度由以下公式给出

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_{X,Y}(x, z-x)\, dx。 \tag{6.6}$$

证明

我们采用分布函数：

$$\begin{aligned}
F_{X+Y}(z) &= P(X+Y \leq z) \\
&= P_{X,Y}(\{(x,y) : x+y \leq z\}) \\
&= \int\int_{\{(x,y):x+y\leq z\}} f_{X,Y}(x,y)\, dxdy \\
&= \int_{\mathbb{R}} \int_{-\infty}^{z-x} f_{X,Y}(x,y)\, dydx \\
&= \int_{-\infty}^{z} \int_{\mathbb{R}} f_{X,Y}(x, y'-x)\, dxdy'
\end{aligned}$$

(我们使用了$y' = y + x$和福比尼定理的替换)，通过微分得出结果。

□

练习6.6

如果$f_{X,Y}=1_{[0,1]\times[0,1]}$，求$f_{X+Y}$。

6.5.2 独立 再次

假设随机变量X，Y是独立的。那么对于一个博雷尔矩形：$B = B_1 \times B_2$ 我们有

$$\begin{aligned}
p_{(X,Y)}(b_1 \times b_2) &= p((x,y) \in b_1 \times b_2) \\
&= p((x \in b_1) \cap (y \in b_2))
\end{aligned}$$

$$= p\ (x \in b_1\ )p\ (y \in b\ )_2$$

$$= p_X\ (b_1\ )p_Y\ (b\ )_2$$

所以分布$P_{(X,Y)}$与矩形上的乘积度量$P\ P_X \times_Y$相吻合，因此它们是相同的。反之亦然：

定理6.13

随机变量X、Y是独立的，当且仅当

$$p_{(x,y)} = p_X \times p_Y。$$

证明

上面显示了 "只有当 "的部分。假设$P_{(X,Y)} = P_X \times P_Y$，并取任何博勒集$B_1, B_2$。同样的计算表明，$P((X\in B_1) \cap (Y\in B_2)) = P(X\in B_1)P(Y\in B_2)$，即$X$和$Y$是独立的。□

在绝对连续随机变量的情况下，我们有这个定理的一个有用版本。

定理6.14

如果X、Y有一个联合密度，那么它们是独立的，当且仅当

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y)。 \tag{6.7}$$

如果X和Y是绝对连续和独立的，那么它们有一个联合密度，它由（6.7）给出。

证明

假设$f_{(X,Y)}$是X, Y的联合密度。如果它们是独立的，那么

$$\begin{aligned}\int_{B_1\times B_2} f_{(X,Y)}(x, y)\, dm_2(x, y) &= P((X, Y) \in B_1 \times B_2) \\ &= P(X\in B_1)P(Y\in B_2) \\ &= \int_{B_1} f_X(x)\, dm(x) \int_{B_2} f_Y(y)\, dm(y) \\ &= \int_{B_1\times B_2} f_X(x) f_Y(y)\, dm_2(x, y)\end{aligned}$$

这意味着（6.7）。同样的计算显示了反面的情况：

$$p((x,y)\in b_1\times b_2)=\int_{B1\times B2} f_{(X,Y)}(x,y)\,dm_2(x,y)$$
$$=\int_{B1\times B2} f_X(x)f_Y(y)\,dm_2(x,y)$$
$$=p(x\in b_1)p(y\in b_2)。$$

对于最后的主张，请注意，如果X和Y是独立的，函数$f_X(x)f_Y(y)$起到联合密度的作用。 □

推论6.15

如果高斯随机变量是正交的，那么它们就是独立的。

证明

将ρ=0插入(6.5)，我们立即看到二维高斯密度是一维密度的乘积。 □

命题6.16

密度为f的独立随机变量之和的密度$_X$，f_Y

是由

$$f_{X+Y}(z)=\int_R f_X(x)f_Y(z-x)\,dx。$$

练习6.7

假设X、Y的联合密度为1_A，其中A是角在（0，1）、（1，0）、（2，1）、（1，2）的正方形。X, Y是独立的吗？

练习6.8

求P（$Y>X$）和P（$X+Y>1$），如果X，Y是独立的，有
$f_X=1_{[0,1]}$, $f_Y=\frac{1}{2}1_{[0,2]}$。

6.5.3 条件 概率

我们考虑两个随机变量X、Y的情况，其联合密度$f_{X,Y}(x, y)$。给定博莱尔集A、B，我们计算一下

$$
\begin{aligned}
p（y\in b \mid x\in a）&= \frac{p（x\in a，y\in b）。}{P(X\in A)}\\
&= \frac{\int_{A\times B} f_{(X,Y)}(x, y)\, dm_2(x, y)}{\int_A f_X(x)\, dm(x)}\\
&= \int_B \frac{\int_A f_{(X,Y)}(x, y)\, dx}{\int_A f_X(x)\, dx}\, dy
\end{aligned}
$$

使用富比尼定理。因此，给定X的Y的条件分布 $\in A$有一个密度

$$h(y \mid X\in A) = \frac{\int_A f_{(X,Y)}(x, y)\, dx}{\int_A f_X(x)\, dx}。$$

$A=a$的情况在这里没有意义，因为那样的话，我们的分母中就会有零。然而，从形式上看，我们可以把

$$h(y \mid X = a) = \frac{f_{(X,Y)}(a, y)}{f_X(a)}$$

如果只有$f_X(a)=0$的话，这是有薏义的。这个限制被证明是不相关的，因为

$$
\begin{aligned}
P((X, Y) \in \{(x, y) : f_X(x) = 0\}) &= \int_{\{(x,y): f_X(x)=0\}} f_{(X,Y)}(x, y)\, dx\, dy\\
&= \int_{\{x: f_X(x)=0\}} \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy\, dx\\
&= \int_{\{x: f_X(x)=0\}} f_X(x)\, dx\\
&= 0.
\end{aligned}
$$

因此，我们可以通过$h(y \mid X=a)$来定义$Y\in B$的条件概率，我们把它简写为$h(y \mid a)$：

$$P(Y\in B \mid X = a) = \int_B h(y \mid a)\, dy$$

和条件期望值

$$\mathrm{E}(Y \mid X = a) = \int_{\mathbb{R}} y h(y \mid a)\, dy。$$

这可以被看作是一个以X为随机性来源的随机变量。也就是说，对于$\omega\in\omega$，我们写成

$$E(Y/X)(\omega) = \int_{R} yh(y/X(\omega))\, dy。$$

这个函数当然是可测量的，相对于由下列因素产生的σ场而言
X.

这个随机变量的期望值可以用富比尼定理来计算：

$$
\begin{aligned}
E(E(Y/X)) &= E\left(\int_{R} yh(y/X(\omega))\, dy\right)\\
&= \int_{R}\int_{R} yh(y/x)\, dy f_X(x)\, dx\\
&= \int_{R}\int_{R} yf_{(X,Y)}(x, y)\, dx\, dy\\
&= \int_{R} yf_Y(y)\, dy\\
&= E(Y)。
\end{aligned}
$$

更一般地说，对于$A\subset\Omega$，$A=X^{-1}(B)$，B *Borel*、

$$
\begin{aligned}
\int_{A} E(Y/X)\, dP &= \int_{\Omega} 1_B(X)E(Y/X)\, dP\\
&= \int_{\Omega} 1_B(X(\omega))\left(\int_{R} yh(y/X(\omega))\, dy\right) dP(\omega)\\
&= \int_{R}\int_{R} 1_B(x)yh(y/x)\, dy f_X(x)\, dx\\
&= \int_{R}\int_{R} 1_B(x)yf_{(X,Y)}(x, y)\, dx\, dy\\
&= \int_{\Omega} 1_A(X)Y\, dP\\
&= \int_{A} Y\, dP。
\end{aligned}
$$

这为给定随机变量X的随机变量Y的条件期望的一般概念提供了动机：$E(Y/X)$是一个相对于X生成的σ场F_X可测量的随机变量，并且对于所有$A\in F_X$

$$\int_{A} E(Y/X)\, dP = \int_{A} Y\, dP。$$

我们将在下一章进一步探讨这些想法。

练习6.9

设$f_{X,Y} = 1_A$，其中A是角在（0，0）、（2，0）、（0，1）的三角形。求条件密度$h(y, x)$和条件期望值$E(Y \mid X = 1)$。

练习6.10

设$f_{X,Y}(x, y) = (x + y)1_A$，其中$A = [0, 1] \times [0, 1]$。求每个$y$∈R的$E(X \mid Y = y)$。

6.5.4 特征函数决定了 分布

现在我们有足够的工具来证明特征函数的一个基本属性。

定理6.17（反演公式）

如果一个随机变量X的累积分布函数在a，b∈R处是连续的，那么

$$F_X(b) - F_X(a) = \lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iua} - e^{-iub}}{iu} \phi_X(u)\, du.$$

证明

首先，根据ϕ_X的定义、

$$\frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iua} - e^{-iub}}{iu} \phi_X(u)\, du = \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iua} - e^{-iub}}{iu} \int_{R} e^{iux}\, dP_X(x)\, du.$$

我们可以应用富比尼定理，因为

$$\left|\frac{e^{-iua} - e^{-iub}}{iu} e^{iux}\right| = \left|\int_a^b e^{iux}\, d(x)\right| \le b - a$$

对于$P_X \times m_{[-c,c]}$，它是可整定的。我们计算u中的积分

$$\frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iua} - e^{-iub}}{iu} e^{iux}\, du = \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{iu(x-a)} - e^{iu(x-b)}}{iu}\, du = \frac{1}{2\pi}\int_{-c}^{c} \frac{\sin u(x-a) - \sin u(x-b)}{u}\, du.$$

$$+\frac{1}{}\int_{c}\frac{\cos u(x-a)-\cos u(x-b)}{d}$$

第二个积分消失了，因为积分是一个奇数函数。我们改变第一个变量：$y=u(x-a)$，$z=u(x-b)$，然后它的形式是

$$I(x, c) = \frac{1}{2\pi}\int_{-c(x-a)}^{c(x-a)} \frac{孽y}{y}\, dy - \frac{1}{2\pi}\int_{-c(x-b)}^{c(x-b)} \frac{\sin z}{z}\, dz = I_1(x, c) - I_2(x,$$

说。我们采用以下基本事实，无需证明：

$$\int_s^t \frac{\sin y}{y} \mathrm{d} y \to \pi \quad 如\ t \to \infty, \quad s \to -\cdot$$

请考虑以下情况：

1. $x < a$，则x也$< b$，且$c\ (x-a) \to -\infty$，$c\ (x-b) \to -\infty$，$-c\ (x-a) \to \infty$、$-c(x-b) \to \infty$，因为$c \to \infty$。因此，$I_1(x, c) \to -\frac{1}{2}$, $I_2(x, c) \to -\frac{1}{2}$，所以$I(x, c) \to 0$。
2. $x>b$，则x也$>a$，且$c\ (x-a) \to \infty$，$c\ (x-b) \to \infty$，$-c\ (x-a) \to -\infty$、$-c(x-b) \to -\infty$，因为$c \to \infty$所以$I_1(x, c) \to \frac{1}{2}$, $I_2(x, c) \to \frac{1}{2}$，结果是与1中相同。
3. $a < x < b$ 因此$I_1(x, c) \to \frac{1}{2}$, $I_2(x, c) \to -\frac{1}{2}$ 和整体的极限表达式为1。

写出$f(x) = \lim_{c\to\infty} I(x, c)$（我们还没有讨论$x = a, x = b$的值但正如我们将看到的那样，它们是不相关的）。

$$肢体_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iua} - e^{-iub}}{iu} \phi_X(u)\, \mathrm{d}\, u = \lim_{c\to} \int_{\mathbb{R}} I(x, c)\, \mathrm{d}P_X(x)$$
$$= \int_{\mathbb{R}} f(x)\, dP_X(x)$$

通过Lebesgue的支配收敛定理。由于f是一个简单的函数，f的积分可以很容易地计算出来：

$$\int_{\mathbb{R}} f(x)\, dP_X(x) = P_X((a, b]) = F_X(b) - F_X(a)$$

($P_X(\{a\}) = P_X(\{b\}) = 0$，因为$F_X$在$a$和$b$处是连续的)。 □

推论6.18

特征函数决定了概率分布。

证明

由于F_X是单调的，它是连续的，除了（可能）在可数个点上是右连续的。它在不连续点的值可以是

从上面的连续性点的值来近似。后者是由特征函数通过反转公式确定的。

最后，我们看到，F_X 决定了度量P_X。对于$B = (a, b]$，这当然是如此：$P_X((a, b]) = F_X(b) - F_X(a)$。接下来，我们对任何区间，然后对区间的有限联合，以及最后对任何博莱尔集的扩展都是通过单调类定理来显示的。

定理6.19

如果φ_X 是可整定的，那么X有一个密度，该密度由以下公式给出

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\varphi_X(u)\, du.$$

证明

该函数f是定义明确的。为了证明它是X的密度，我们首先证明它给出了区间（a，b）的概率分布的正确值，其中F_X是连续的：

$$\begin{aligned}
\int_a^b f_X(x)\, dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(u) \left(\int_a^b e^{-iux}\, dx\right) du \\
&= \lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \phi_X(u) \left(\int_a^b e^{-iux}\, dx\right) du \\
&= \lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \phi_X(u) \frac{e^{-iua} - e^{-iub}}{iu}\, du \\
&= F_X(b) - F_X(a)
\end{aligned}$$

通过反转公式。这延伸到所有的a，b，因为F_X是右边连续的，左边的积分对于a和b是连续的。此外、F_X 是非递减的，所以$\int_a^b f_X(x)\, dx \geq 0$，对于所有$a \leq b$，因此$f_X \geq 0$. 最后

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \lim_{b\to\infty} F_X(b) - \lim_{a\to-\infty} F_X(a) = 1$$

所以f_X 是一个密度。 □

6.5.5 应用于数学 金融

经典的投资组合理论关注的是对风险和收益之间的平衡的分析。这种平衡具有根本的重要性，特别是在

公司金融，其中的关键概念是资本成本，这是一个基于投资风险水平的回报率。在概率论方面，收益由期望值表示，风险由方差表示。如果我们假设有关随机变量的正态（高斯）分布，那么只处理随机变量的两个时刻的理论是相关的，因为在这种情况下，这两个时刻决定了分布的唯一性。我们简要介绍一下这种假设下的投资组合理论的基本事实。

设k是某项投资在单期的回报，即$k(\omega)=\frac{V(1,\omega)-V(0)}{V(0)}$ 其中V（0）是开始时的已知投资额，而$V(1)$是随机终端值。一个典型的例子，应该保持在心灵是买入和卖出某只股票的一股。在有许多股票可用的情况下，我们面临着股票S_i、$k_i=\frac{S_i(1,\omega)-S_i(0)}{S_i(0)}$的回报序列$k_1$、$k_2$ ��但为了��� *单起见，我们的注意*��只限于两个，k、k。通过选择权重$w=(w_1,w_2)$, $w_1+w_2=1$来决定初始财富$V(0)$在S_1和S_2中的持股比例，从而形成一个投资组合。然后、正如众所周知的和基本的验证，n个$_1=\frac{w_1V(0)}{S_1(0)}$股的组合属于第一种股票的$n_2=\frac{w_2V(0)}{S_2(0)}$股第二种股票，其收益率为

$$k_w=w_1k_1+w_2k_2。$$

我们假设向量（k_1,k_2）是具有相关系数ρ的联合正态。我们用$\mu_i=\mathrm{E}(k_i)$, $\sigma_i^2=\mathrm{Var}(k_i)$来表示成分的期望和变异。引入以下矩阵是很方便的

$$C=\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22}\end{bmatrix}=\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$$

其中$c_{12}=c_{21}$是k_1和k_2之间的协方差。假设（这样做并不优雅，但省去了我们的代数弯路）C是可逆的，$C^{-1}=[d_{ij}]$。根据定义，联合密度的形式为

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{\det C}}\exp\{-\frac{1}{2}\sum_{i,j=1}^{2}d_{ij}(x_i-\mu_i)(x_j-\mu_j)\}$$

很容易看出，(6.5)是这个公式的一个特例，$\mu_i=0$, $\sigma_i=1$, $-1<\rho<1$。众所周知，特征函数$\phi(t_1,t_2)=\mathrm{E}(\exp\{i(t_1k_1+t_2k_2)\})$的矢量（$k_1,k_2$）的形式是

$$\phi(t_1,t_2)=\exp\{i\sum_{i=1}^{2}\mu_it_i-\frac{1}{2}\sum_{i,j=1}^{2}c_{ij}t_it_j\}。\tag{6.8}$$

我们将证明投资组合的回报率也是正态分布的，我们将找到期望值和

标准差。这一切都可以在一个步骤中完成

定理6.20

k_w 的特征ϕ函数$_w$，其形式为

$$\phi_w(t) = \exp\{it(w_1\mu_1 + w_2\mu_2) - \frac{1}{2}t^2(w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2)\}$$

证明

根据定义，$j_w(t) = E(\exp\{itk_w\})$，并使用$k_w$ 的形式，我们有

$$\begin{aligned}\phi_w(t) &= E(exp\{it(w_1k_1 + w_2k_2)\}\\ &= E(exp\{itw_1k_1 + itw_2k_2\})\\ &= \phi(tw_1, tw_2)\end{aligned}$$

根据向量的特征函数的定义。由于该向量是正常的，(6.8)立即得到了结果。□

推论6.18的多维版本（在掌握了一维情况后很容易相信，但证明起来略显繁琐，所以我们认为这是理所当然的，请读者参阅任何概率教科书）表明，k_w 具有正态分布，具有

$$\mu_w = w_1\mu_1 + w_2\mu_2$$
$$\sigma_w^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2$$

投资组合的方差可以低于各组成部分的方差，这一点至关重要。这些公式在一般情况下是有效的（即没有正态分布的假设），并且可以很容易地用k的公式证明$_w$。 本节的主要目标是看到投资组合收益率是正态分布。

例6.4

假设第二部分不是随机的，即S_2（1）是一个独立于ω的常数。那么回报率k_2是无风险的，它被表示为r（的符号通常保留给周期长度为一年的情况）。它可以被认为是一个银行账户，假设S_2（0）=1是很方便的。那么以S_1(0)的价格购买的n股组合和m单位的银行账户在期末的价值为V（1）$=nS_1$（1）$+m$（1+r），预期收益为$k_w = w_1\mu_1 + w_2 r$，$w_1 = \frac{nS_1(0)}{V(0)}$、$w_2 = 1 - w_1$。 违反了正常联合收益的假设，但这个投资组合的标准差可以直接很容易地计算出来。

定义给出$\sigma_{\mathrm{W}} = w\,\sigma_{11}$ （当然，$\sigma_2 = 0$，该公式与上述一致）。

备注6.3

上述考虑可以立即推广到由任何有限数量的成分组成的投资组合，关键公式如下

$$k_{\mathrm{W}} = \sum w\, k_{ii}\,,$$
$$\mu_{\mathrm{W}} = \sum w_i\,\mu_i$$
$$\sigma_{\mathrm{W}}^{2} = \sum_{i,j} w_i w_j c_{ij}\,.$$

这只是诺贝尔奖得主哈里-马科维茨在20世纪50年代开始的故事的开头。此后，关于这个主题的大量论文和书籍被写出来，证明了 "简单就是美 " 的普遍看法。

6.6 命题的证明

证明（命题6.2的证明）

用F_R 表示由Borel "矩形 "R = {$B_1 \times B_2 : B_1, B_2 \in$B}产生的σ场，用$F_I$ 表示由真正的矩形I = {$I_1 \times I_2 : I_1, I_2$ 是区间}产生的σ场。

由于I⊂R，显然$F_I \subset F_R$。

为了显示逆向包容，我们表明博勒圆柱$B_1 \times \Omega_2$ 和 $\Omega_1 \times B_2$ 都在F_I。为此，写D = {A : A $\times\,\Omega_2 \in F_I$ }，注意这是一个包含所有区间的σ场，因此B⊂D符合要求。 □

证明（命题6.11的证明）

让$s = \sum^{n} c_k 1_{A_k}$ 是一个递增的简单函数序列，趋近于 $\int$ ß dm

f .让$R_k = A_k \times [0, c_k]$，这种矩形的联合实际上是

。那么$\bigcup_{n\to\infty} \bigcup_k R_k = A_f$ 所以A_f 是可测的。

对于第二个要求，取A的一个y段$_f$ ，该段为区间[0，f（x）]。它的度量是$f(x)$，根据乘积度量的定义$m_2(A_f) =$

$\int f(x)\,dx$。 □

证明（命题6.16的证明）

联合密度是密度的乘积：$f_{X,Y}(x, y) = f_X(x)f_Y(y)$，将其代入（6.6）立即得到结果。 $\square$

7 拉登-尼科迪姆定理

在本章中，我们将考虑实伯勒测量v和勒贝斯格测量m之间的关系。这种关系的关键是定理4.17，它表明对于每个非负整数实函数f，集合函数

$$A \mapsto v(A) = \int f \, dm \tag{7.1}$$

定义了(R,) 上的(Borel)度量v。我们要问的自然是反过来的问题：究竟哪些实的Borel度量可以通过这种方式找到？我们将在本章中找到这个问题的完整答案，并且为了与我们的观点保持一致
在第5章和第6章中，我们将用抽象集Ω上的一般度量来表述我们的结果。

7.1 密集度和 调节

我们将在本章发展的结果也使我们能够更详细地研究概率密度（在第4.7.2节中介绍）、条件概率和条件期望（见第5.4.3和6.5.3节）。对于上面定义的v是一个概率度量，我们显然要求$\int f \, dm = 1$。特别是，如果$v=P_X$是随机变量X的分布，那么（7.1）中与v对应的函数$f=f_X$被称为X的*密度*。

以类似的方式，我们在第6.5.1节中定义了两个随机变量的联合密度$f_{(X,Y)}$，通过将它们的联合分布引用到二维的

Lebesgue度量m_2：如果X和Y是定义在某个概率空间（Ω，F，P）上的实数随机变量，它们的联合分布是定义在R 的Borel子集B 上的度量[2]：$P_{(X,Y)}$（B）= P（（X，Y）∈B）。在特殊情况下
其中，相对于m_2，这个度量是由上述的一个可整定的函数给出的
$f_{(X,Y)}$，我们说X和Y有这个函数作为其联合密度。

这又自然而然地导致了（见第6.5.3节）条件密度的概念。

$$h(y|a) = h(y|X = a) = \frac{f_{(X,Y)}(a, y)}{f_X(a)}$$

和条件期望值

$$E(Y|X = a) = \int_R y h(y|a)\, dy。$$

回顾一下：$X\ \Omega \to R$，最后一个方程可以写成$E(\ )(\ ) = \int_R yh(y|X(\omega))\, dy$，将条件期望显示为随机变量$E(Y|X) : \Omega \to R$，相对于X产生的σ场$F_X$，是可以测量的。应用福比尼定理可以得到一个基本特性，对于所有$A \in F_X$

$$\int_A E(Y|X)\, dP = \int_A Y\, dP。 \tag{7.2}$$

在一般情况下，无论联合密度是否存在，这个随机变量的存在在理论和应用中都是非常重要的--威廉姆斯[12]称之为 "现代概率的核心定义"。它对于在许多应用中起着关键作用的马丁格尔概念至关重要，我们在本章末尾会介绍这个概念。

正如我们在第5.4.3节中所描述的那样，L^2 中存在的正交投影允许我们进一步扩展定义的范围：而不是将我们自己限制在关于σ-场的可测量的随机变量上。
形式的F_X，我们指定F的任何子σ场G，并要求一个G可测量的随机变量E(Y|G)来扮演(7.2)中的$E(Y|X) = E(Y|F_X)$。如同乘积度量的情况一样，建立G测度的最自然的背景是

条件期望的属性是一般度量的属性；注意，定理4.17的证明只需要单调收敛来建立P的可数可加性。因此，我们进一步发展抽象度量的比较，一如既往地以随机变量和分布的具体例子为指导。

7.2 Radon-Nikodym 定理

$\int$

在特殊情况下，即度量v具有$v(A)={}_A f\ dm$的形式，对于某个非负的整数函数f，我们说（第4.7.2节），v是绝对的。

只要$m(A)=0$，$\int_A f\,dm$就立即$=0$（见定理4.3（iv））。因此，当度量ν由密度给出时，$m(A)=0$意味着$\nu(A)=0$。我们将此作为两个给定度量的一般情况下的定义。

定义7.1

设Ω是一个集合，让$\mathcal{F}$是其子集的σ场。(对（Ω，$\mathcal{F}$）是一个*可测空间*）假设ν和μ是（Ω，$\mathcal{F}$）上的度量。如果$\mu(A)=0$意味着$\nu(A)=0$，我们说ν*相对于*μ*是绝对连续的*，因为

$A\in\mathcal{F}$。我们把它写成$\nu \ll \mu$。

练习7.1

让λ_1, λ_2 和 μ 是$(\Omega, \mathcal{F})$上的度量。证明如果$\lambda_1 \ll \mu$ 和 $\lambda_2 \ll \mu$ 那么$(\lambda_1+\lambda_2) \ll \mu$。

这个定义与通常的函数连续性概念有什么关系并不明显。我们将在本章后面看到它如何与实数函数的绝对连续性概念相适应。就目前而言，我们注意到以下定义的重新表述，这对于我们将要证明的主要结果来说是不需要的，但有助于使ν和μ之间的关系更 "贴近"，并且在许多应用中是有用的：

命题7.1

让ν和μ是可测空间（Ω，$\mathcal{F}$）上的有限度量。那么，当且仅当对于每一个$\varepsilon>0$，存在一个$\delta>0$，使得对于$F\in\mathcal{F}$，$\mu(F)<\delta$意味着$\nu(F)<\varepsilon$时，$\nu \ll \mu$。

提示 假设(ε, δ)的条件失败。那么我们可以找到$\varepsilon>0$，并设定(F_n)，这样对于所有$n\geq 1$，$\mu(F_n)<\frac{1}{2^n}$，但$\nu(F_n)>\varepsilon$。考虑$\mu(A)$ 和$\nu(A)$为$A=\bigcap_{n\geq 1}(\bigcup_{i\geq n} F_i)$。

我们从Lebesgue度量的特殊情况进行概括：如果μ是$(\Omega, \mathcal{F})$上的任何度量，并且$f: \Omega \to \mathbb{R}$是一个存在$\int f\,d\mu$的可测函数，那么$\nu(F)=\int_F f\,d\mu$定义了一个度量ν，$\nu \ll \mu$.(这与对$\mu=m$的情况完全相同，因为

$\mu(F)$=0意味着$_F f\, d\mu$=0。注意，我们采用的是0=0的惯例。）对于σ-finite度量，下面的关键结果断言了反面的情况：

定理7.2（Radon-Nikodym）

给定可测空间（Ω，）上的两个σ-finite度量v，μ，其中v μ，那么就有一个非负的可测函数 : $h\Omega$ R，使得（）=dv F，对于 $\int_F h\,\mu$ 每一个 $F\in$ F .该函数 h 是唯一的，直到μ-空集：如果g 也对于所有的$F\in$F，满足$v(F)=\int_F g\,\mathrm{d}\mu$，那么$g=h$ a.e. (μ)。

由于应用中最有趣的情况出现在概率空间，然后$h^1(\mu)$，我们最初将把注意力限制在μ和v是有限度量的情况下。事实上，最初把μ看作是一个概率度量，即$\mu(\Omega)$=1是有帮助的。在这个非常重要的定理的几种不同方法中，我们以R.C.布拉德利在《美国数学月刊》（第96卷第5期，1989年5月，第437-440页）中提出的方法为基础进行论证，因为它提供了我们所知道的最 "建设性 "和最基本的处理。

从一个特殊情况开始是有启发的：假设（在进一步通知之前），$\mu(\Omega)$=1。我们说，当$0\leq v(F)\leq\mu(F)$对每一个$F\in$F来说都*支配v时，*措施μ*就支配了*。这显然意味着v μ 。在这种简化的情况下，我们将明确地构建所需的函数h。首先，我们概括一下以下的想法

我们在构建黎曼积分时，对(Ω, F)中的可测子集进行了很好的划分和细化。

定义7.2

让（Ω，F）是一个可测量的空间。Ω的有限(可测)*分区*是F中不相交的子集P = $(A_i)_{i\leq n}$ 的有限集合，其联盟是Ω。如果P中的每个集合都是P′中的集合的不相交的联合，那么有限分区P′就是P的细化。

练习7.2

让P_1和P_2是Ω的有限分区。证明完善这两个分区的最粗分区（即集合数最少）由所有的交点$A\cap B$组成，其中$A\in P_1$，$B\in P_2$。

下面是一个简化的 "拉登-尼科迪姆定理"，用于支配性测量：

定理7.3

假设$\mu(\Omega)=1$，且$0\leq \nu(F)\leq \mu(F)$，对于每一个$F\in \mathcal{F}$，则有

在Ω上存在一个非负的F可测量函数h，使得$\nu(F)=\int_F h\,d\mu$，对于所有F∈F。

我们将分三步来证明这一点：在步骤1中，我们为有限分区P中的集合定义所需的函数h_P，并比较函数h_{P_1}和h_{P_2}，当分区P_2完善了P_1。这使我们能够表明，积分$\int_\Omega h_P^2\,d\mu$是不递减的，如果我们采取连续的细化方法。由于它们是上面也是有界的（通过1），$c=\int_\Omega h_P^2\,d\mu$存在于R中。在步骤2中，我们通过仔细的极限论证，利用第四章的收敛定理，构建所需的函数h。在步骤3中，我们表明h具有所需的特性。

第1步：函数h_P为有限分区

假设对于每一个$F\in$F，$0\le\nu(F)\le\mu(F)$。让$P=\{A_1,A_2,...,A_k\}$是Ω的一个有限分区，使得每个$A_i\in$F。通过设定简单函数$h_P:\Omega\to\mathbb{R}$来定义

$$h_P(\omega)=c_i=\frac{\nu(A_i)}{\mu(A_i)}\text{对于}\omega\in A_i\text{，当}\mu(A_i)>0\text{时，}h_P(\omega)=0\text{。}$$

由于h_P在每个"原子"A_i上是常数，$\nu(A_i)=\int_{A_i}h_P\,d\mu$。那么$h_P$具有以下性质：

(i) 对于Ω的每个有限分区P，$0\le h_P(\Omega)\le 1$，对于所有$\Omega\in\Omega$。

(ii) 如果$A=\bigcup_{j\in J}A_j$，对于一个索引集$J\subset\{1,2,...k\}$，那么$\nu(A)=\int_F h_P\,d\mu$。因此$\nu(\Omega)=\int_\Omega h_P\,d\mu$。

(iii) 如果P_1和P_2是Ω的有限分区，并且P_2完善了P_1，那么，随着$h_n=h_{P_n}$，（$n=1,2$）我们有

(a) 对于所有$A\in P_1$，$\int_A h_1\,d\mu=\nu(A)=\int_A h_2\,d\mu$、

(b) 对于所有$\int_A h_1h_2\,d\mu=\int_A h_1^2\,d\mu$

(iv) $\int_\Omega(h_2^2-h_1^2)\,d\mu=\int_\Omega(h_2-h_1)^2\,d\mu$，因此

$$\int_\Omega h_2^2\,d\mu=\int_\Omega h_1^2\,d\mu+\int_\Omega(h_2-h_1)^2\,d\mu\ge\int_\Omega h_1^2\,d\mu\text{。}$$

现在我们依次证明这些论断。

(i) 通过构建h_P，这一点是*微不足道*的，因为μ支配ν。

(ii) 让$A = \bigcup_{j\in J} A_j$，对于某个索引集$J\subset\{1，2，...，k\}$.由于$\{A_j\}$的是不相交的，并且只要$\mu(A_j)=0$，$\nu(A_j)$就=0，我们有

$$\nu(A) = \sum_{j\in J}\nu(A_j) = \sum_{j\in J,\mu(A_j)>0}\frac{\nu(A_j)}{\mu(A_j)}\mu(A_j) = \sum_{j\in J,\mu(A_j)>0} c_j\,\mu(A_j) = \sum_{j\in J}\int_{A_j} h_P\,d\mu = \int_A h_P\,d\mu。$$

特别是，由于 P 分割Ω，这在$A=\Omega$时成立。

(iii) (a) 以上述的P_n, h_n ($n = 1, 2$)，我们可以写出$A = \bigcup_{j\in J} B_j$，对于每个$A\in P_1$，其中J是一个有限索引集，$B_j\in P_2$。B_j的集合是成对不相交的，同样$\nu(B_j)=0$，当$\mu(B_j)=0$，所以

$$\int_A h_1\,d\mu = \nu(A) = \sum_{j\in J}\nu(B_j) = \sum_{j\in J,\mu(B_j)>0}\frac{\nu(B_j)}{\mu(B_j)}\mu(B_j) = \sum_{j\in J}\int_{B_j} h_2\,d\mu = \int_A h_2\,d\mu。$$

(b) A与(a)部分相同，$\mu(A)>0$，注意$h_1 = \frac{\nu(A)}{\mu(A)}$在$A$上是常数。因此，

$$\int_A h_1 h_2\,d\mu = \frac{\nu(A)}{\mu(A)}\int_A h_2\,d\mu = \frac{(\nu(A))^2}{\mu(A)} = \int_A\left(\frac{\nu(A)}{\mu(A)}\right)^2 d\mu = \int_A h_1^2\,d\mu$$

(iv) 根据(iii)(b)，$\int_A h_1(h_2-h_1)\,d\mu = 0$，对于每个$A\in P_1$.由于$A_i\in P_1$分割$\Omega$，我们也有

$$\int_\Omega h_1(h_2-h_1)\,d\mu = \sum\int_{A_i} h_1(h_2-h_1)\,d\mu = 0。$$

因此

$$\begin{aligned}\int_\Omega (h_2-h_1)^2\,d\mu &= \int_\Omega (h_2^2 - 2h_1h_2 + h_1^2)\,d\mu\\ &= \int_\Omega [h_2^2 - 2h_1(h_2-h_1) - h_1^2]\,d\mu\\ &= \int_\Omega (h_2^2 - h_1^2)\,d\mu、\end{aligned}$$

因此

$$\int h_2^2\,d\mu = \int h_1^2\,d\mu + \int$$

$$\int_{\Omega} (h_2 - h_1)^2 \, d\mu \geq \int_{\Omega} h^2 \, d\mu。$$

第2步：通向极限--构建h。

在步骤1中，我们表明，积分$\int_\Omega h^2_{\mathrm{P}}\, d\mu$在苏----的过程中是不递减的。对一个有限分区的逐次细化 Ω.此外，根据上面的(i)，每个function h_{P} 满足 $0 \le h_{\mathrm{P}}(\omega) \le 1$ 对于所有$\omega\in\omega$。因此，设定$c = \sup \int_\Omega h^2_{\mathrm{P}} d\mu$，其中最高值取自所有的有限分区的 Ω，我们有$0 \le c \le 1$ (这里我们使用假设，即$\mu(\Omega)=1$）。

对于每一个$n\ge1$，让P_n 是Ω的一个有限可衡量的分区，以便 $\int_\Omega h^2_{\mathrm{P}n}\, d\mu > c-\frac{1}{4n}$。让$\mathrm{Q}_n$ 是分区的最小的共同细化 P_1，P_2，$\ldots$，P_n。对于每一个n，Q_n 在结构上细化了P_n，Q_{n+1} 细化了Q_n，因为每个Q_k 由所有的交集$A_1\cap A_2\cap\ldots\cap A_k$，其中$A_i\in\mathrm{P}_i$，$i\ k$。因此，$\mathrm{Q}_n$ 中的每个集合都是Q_{n+1} 中的集合的不相交联合。因此，我们有以下不等式：

$$c-\frac{1}{4n}<\int_\Omega h^2_{\mathrm{P}n}\, d\mu \le \int_\Omega h^2_{\text{秦}}\, d\mu \le \int_\Omega h^2_{\mathrm{Q}n+1}\, d\mu\le c。$$

利用步骤1(iv)中证明的特性，我们现在有

$$\int_\Omega (h_{\mathrm{Q}\,n+1}-h_{n\mathrm{Q}})^2\, d\mu=\int_\Omega (h^2_{\mathrm{Q}n+1}-h^2_{\mathrm{Q}n})\, d\mu<\frac{1}{4n}。$$

施瓦兹不等式适用于$f=|h_{\mathrm{Q}\,n+1}-h_{\mathrm{Q}\,n}|$和$g\equiv1$，那么就可以得到对于每一个$n\ge1$、

$$\int_\Omega |h_{\mathrm{Q}\,n+1}-h_{\mathrm{Q}\,n}|\, d\mu<\frac{1}{2n}。$$

根据Beppo-Levi定理，由于 $\sum_{n\ge1}\int_\Omega |h_{\mathrm{Q}\,n+1}-h_{\mathrm{Q}\,n}|\, d\mu$是有限的，我们得出结论：数列$\sum_{n\ge1}(h_{\mathrm{Q}\,n+1}-h_{\mathrm{Q}\,n})$几乎到处收敛($\mu$)，因此，极限函数

$$h=h_{\mathrm{P}\,1}+\sum_{n\ge1}(h_{\mathrm{Q}\,n+1}-h_{\mathrm{Q}\,n})=\lim_n h_{n\mathrm{Q}}$$

(注意到 $\mathrm{Q}_1=\mathrm{P}_1$）几乎在任何地方都是很好定义的（$\mu$）。我们通过在例外的$\mu$空集上设置$h$=0来完成构造。

第3步：验证h的特性。

由步骤1（i）可知，$0 \le h(\omega) \le 1$，从它的构造中可以看出 h 是F-可衡量的。

我们需要证明，对于每一个$F\in\mathrm{F}$，$v(F)=\int_F h\, d\mu$。可衡量的集合F，并让$n\ge1$.定义为R_n作为最小的共同细化的两个分区 Q_n (定义同步骤2)和 $\{F, F^c\}$由于 F 是R_n中集合的有限不相交的联合，我们有$v(F)=\int_F h_{\mathrm{R}\,n}\, d\mu$从步骤1（ii）。

根据步骤2，$c-\frac{1}{4n}<\int_\Omega h^2_{Qn}\,d\mu\le\int_\Omega h^2_{Rn}\,d\mu\le c$，所以，如前所述，我们可以

断定 $\int_\Omega(h_{4n}-h_{nQ})^2\,d\mu<\frac{1}{4n}$，并使用一次施瓦兹不等式

此外，这一次$a=1_F$，我们有

$$\left|\int_F(h_{n_R}-h_{n_Q})\,d\mu\right|\le\int_F|h_{n_R}-h_{n_Q}|\,d\mu<\frac{1}{2n}。$$

对于所有n，$\nu(F)=\int_F h_{R_n}\,d\mu=\int_F(h_{R_n}-h_{n_Q})\,d\mu+\int_F h_{Q_n}\,d\mu$。右边的第一个积分在$n\to\infty$时收敛于0，而第二个积分根据支配收敛定理收敛于$\int_F h\,d\mu$（因为对于所有$n\ge1$，$0\le h_{Q_n}\le1$和$\mu(\Omega)$是有限的）。因此，我们已经验证了$\nu(F)=\int_F h\,d\mu$，符合要求。

直接检查一下，假设$\mu(\Omega)=1$并不重要，因为对于任何有限的正度量μ，我们可以重复上述论证 $\frac{\mu}{\mu(\Omega)}$

我们把上面定义的函数h写成$\frac{d\nu}{d\mu}$，并调用它是ν相对于μ的*Radon-Nikodym导数*。当我们考虑有界变化的实数函数时，它与函数导数的关系将变得清晰。

练习7.3

设$\Omega=[0,1]$，具有Lebesgue度量，考虑分别由密度1_A，1_B给出的度量μ,ν。在集合A、B上找到一个条件，使μ支配ν，并找到Radon-Nikodym导数$\frac{d\nu}{d\mu}$。

以上是对函数h的定义。

练习7.4

假设Ω是一个有限集，配有所有子集的代数。让μ和ν是Ω上的两个度量，使得$\mu(\{\omega\})/=0$，$\nu(\{\omega\})/=0$，对于所有$\omega\in\Omega$。决定在哪些条件下μ支配ν，并找到$\frac{d\nu}{d\mu}$。

接下来的观察是对备注4.1中高亮的一般程序的一个简单应用：

命题7.4

如果μ和j是有限度量，$0\le\mu\le j$，如果$h_\mu=\frac{d\mu}{dj}$，如上构造，那么对于Ω上的

任何非负的F度量函数g，我们有

$$\int_{\Omega} g \, \mathrm{d}\mu = \int_{\Omega} g h_{\mu} \, \mathrm{d}j。$$

对于任何$g \in L^{1}(\mu)$，同样的特性也成立。

提示 从指标函数开始，利用积分的线性扩展到简单函数，以及一般非负g的单调收敛，其余的从定义中可以看出。

对于有限措施，我们现在可以证明前面宣布的一般结果：

定理7.5（Radon-Nikodym）

设v和μ是可测量空间（Ω，）上的有限度量，并假设 F

那 v μ .那么有一个非负的 F-可测量的函数 h 在 Ω

$v(A)=\int_A h\, d\mu$ 对于所有$A\in$F。

证明

让$j = v + \mu$. 那么j是一个正的有限度量，它同时支配了

v和μ。因此，Radon-Nikodym导数$h_v = \frac{dv}{dj}$ 和$h_\mu = \frac{d\mu}{dj}$ 是由前面的构造很好地定义了。考虑F $= h > 0$和$G = \{h_\mu = 0\}$中的集合。显然，$\mu(G) = \int_G h_\mu\, dj = 0$，因此$v(G) = 0$，因为$v$ μ。定义$h = \frac{h_v}{h_\mu} 1_F$，并让$A\in$F，$A\subset F$。根据前面的命题，用$h1_A$而不是$g$，我们有

$$v(A) = \int_A h_v\, dj = \int_A hh_\mu\, dj = \int_A h\, d\mu$$

的要求。由于μ和v在G上都是空的，这就证明了该定理。 □

练习7.5

Let $\Omega = [0, 1]$ with Lebesgue measure and consider probability measures μ, v given by densities f, g respectively. Find a condition characterising the absolute continuity v μ and find the Radon-Nikodym derivative $\frac{dv}{d\mu}$.

练习7.6

假设Ω是一个装有所有子集的代数的有限集，让μ和v是Ω上的两个度量。确定绝对连续性v μ的特征，并找到$\frac{dv}{d\mu}$。

现在你可以很容易地完成对σ-无限量的描述，并验证函数h是"本质上唯一的"：

命题7.6

如果度量v和μ是σ-有限的，则拉登-尼科迪姆定理仍然有效：对于任何两个具有 $v \ll \mu$ 我们可以在Ω上找到一个有限值的非负可测函数f，这样$v(F)=\int_F h\, d\mu$，对于所有$F \in \mathcal{F}$。这样定义的函数在μ-空集之前是唯一的，也就是说，如果$g: \Omega \to \mathbb{R}^+$ 也满足$v(F) = \int_F g\, d\mu$ 对于所有的$F \in \mathcal{F}$，那么$g = h$ $a.e.$ （关于μ）。

提示 有一些序列(A_n)、(B_m)的集合$\mathcal{F}$，其中$\mu(A_n)$，$v(B_m)$是有限的。对于所有的m，$n \geq 1$和$\bigcup_{n\geq 1} A_n = \Omega = \bigcup_{m \geq 1} B_m$。我们可以选择这些是不相交集合的序列（为什么？）因此，将Ω显示为以下的互不相干的联盟集$A_n \cap B_m$ $(m, n \geq 1)$，从而找到一个具有联盟Ω的不相交集的序列(C_n)，其所有成员在μ和v下都有有限度量。固定n并将上述结果应用于可测空间（Ω，$\mathcal{F}_n$），其中

$\mathcal{F}_n = \{F \cap C_n : F \in \mathcal{F}\}$，然后对所有$n$的结果函数进行 "粘贴"。

度量的拉登-尼科迪姆导数服从简单的组合规则，这些规则来自于唯一性属性。我们用两个拉登-尼科迪姆导数的和与组合来说明这一点，并将 " 逆规则 "作为一个练习。

命题7.7

Assume we are given σ-finite measures λ, v, μ satisfying $\lambda \ll \mu$ and $v \ll \mu$ with Radon-Nikodym derivatives $\frac{d\lambda}{d\mu}$ and $\frac{dv}{d\mu}$, respectively.

(i) 随着$\varphi=\lambda+v$，我们有$\frac{d\varphi}{d\mu} = \frac{d\lambda}{d\mu} + \frac{dv}{d\mu}$ a.s. (μ)、

(ii) 如果$\lambda \ll v$，那么$\frac{d\lambda}{d\mu} = \frac{d\lambda}{dv}\frac{dv}{d\mu}$ a.s. (μ).

练习7.7

Show that if μ, v are equivalent measures, i.e. both $v \ll \mu$ and $\mu \ll v$ 为真，则

$$\frac{d\mu}{dv} = \left(\frac{dv}{d\mu}\right)^{-1} \text{ a.s. } (\mu)。$$

给定一对在$(\Omega, \mathcal{F})$ 上的σ-finite度量λ, μ，我们自然会问，我们是否能确

定$\mu(E)=0$意味着$\lambda(E)=0$的集合。这意味着我们可以将λ的质量分成两块，一块由μ *积分*表示，另一块 "集中 "在μ空集上，即远离μ的质量。我们将这种 "分离 "两个度量的质量的想法转化为

定义7.3

如果有一个集合$E\in\mathcal{F}$ $\lambda(F)=\lambda(E\cap F)$，对每一个$F$来说都是如此$\in\mathcal{F}$ 如 果 两个度量μ、ν都集中在Ω的不相交的子集上，我们就说它们是*互为奇异的*，写成$\mu\perp\nu$。

显然，如果λ集中在E上，并且$E\cap F=\varnothing$，那么$\lambda(F)=\lambda(E\cap F)=0$。反之，如果对于所有的$F\in\mathcal{F}$，$F\cap E=\varnothing$意味着$\lambda(F)=0$，考虑$\lambda(F)=\lambda(F\cap E)+\lambda(F\setminus E)$。因为$(F\setminus E)\cap E=\varnothing$，我们必须有$\lambda(F\setminus E)=0$，所以$\lambda(F)=\lambda(F\cap E)$。我们已经证明，$\lambda$在$E$上是集中的，当且仅当对于所有的$F\in\mathcal{F}$，$F\cap E=\varnothing$意味着$\lambda(F)=0$。我们收集一些关于互为奇异度量的简单事实：

命题7.8

如果$\mu,\nu,\lambda\ \lambda_{1,2}$是σ-场F上的度量，则以下情况为真：

(i) 如果$\lambda_1\perp\mu$，$\lambda_2\perp\mu$，那么也可以（$\lambda_1+\lambda_2$）$\perp\mu$。

(ii) 如果$\lambda_1\ \ll\mu$和$\lambda_2\perp\mu$，那么$\lambda_2\perp\lambda_1$。

(iii) 如果$\nu\ll\mu$和$\nu\perp\mu$，那么$\nu=0$。

提示 对于(i)，$i=1,2$，让A_i，B_i是不相交的集合，λ_i集中在A_i，μ*集中*在B_i。考虑$A_1\cup A_2$和$B_1\cap B_2$。对于(ii)使用命题前的注释。

下一个结果表明，一个σ-finite度量相对于另一个σ-finite度量的唯一 "质量分割 "总是可能的：

定理7.9（Lebesgue 分解法）

让λ，μ是（Ω，$\mathcal{F}$）上的σ-finite度量。那么λ可以唯一地表示为两个度量之和，$\lambda=\lambda_a+\lambda_s$，其中$\lambda_a\ll\mu$和$\lambda_s\perp\mu$。

证明

存在性：我们考虑有限度量；扩展到σ-无限的情况是常规的。由于$0\leq\lambda\leq\lambda+\mu=\varphi$，即$\varphi$支配λ，*所以*有$0\leq h\leq 1$，使得$\lambda(E)=\int_E h\,d\varphi$，适用于所有

可测量的E. 让$A = \{\omega : h(\omega < 1\}$和$B =$
$\{\omega : h(\omega) = 1\}$。设$\lambda_a(E) = \lambda(A \cap E)$，$\lambda_s(E) = \lambda(B \cap E)$，对于每个$E \in F$。现在，如果 $E \subset A$ 和 $\mu(E) = 0$，那么 $\lambda(E) = \int_E h\, d\phi = \int_E h\, d\lambda$ 这样，$\int_E (1-h)\, d\lambda = 0$。但在$A$上$h<1$，因此在$E$上也是如此。
$\lambda(E)=0$。因此，如果$E \in F$且$\mu(E)=0$，$\lambda_a(E)=\lambda(A \cap E)=0$，因为$A \cap E \subset A$。

所以λ_a λ 。另一方面，如果$E\subset B$，我们得到λ（E）$=_{\int}$E $h\,dj$ =
$\int_E$ 1 d($\lambda+\mu$) = λ(E) + μ(E)，因此，μ(E) = 0。由于$A = B^c$，我们已表
只要$E\cap A=\varnothing$，μ(E)=0，所以μ在A上是集中的。$_s$
是集中在B上的，这表明λ_s和μ是互为奇异的。

唯一性留给读者。(提示：使用命题7.8。）theo- rem被证明。

□

结合这一点和Radon-Nikodym定理，我们可以把λ*相对于*μ*的*结构描述为 "基础度量"：

推论7.10

与定理中的$\mu, \lambda, \lambda_a, \lambda_s$一样，有一个$\mu$-$a$.s. 唯一的非负可测函数$h$，使得
$\lambda(E) = {}_{\int E}\, h\, \mathrm{d}\mu + \lambda_s\,(E)$ 对于每个$E\in$F。

备注7.1

这个结果让人想起了有限维矢量的结构理论
空间：如果$x\in$R^n，且$m < n$，我们可以写成$x = y + z$，其中$y = {}^{\Sigma}$m $y\, e_{ii}$
$_{i=1}$
是对R的正交投影^m，z是对这个子空间的正交。
我们也利用了希尔伯特空间的类似想法。在这个意义上，措施μ
具有提供 "线性组合 "的 "基 "的作用，它描述了度量λ向Ω上的度量空间的一个子空间的投影。

练习7.8

考虑以下实线上的措施：$P_1 = \delta_0$, P_2 =
$\frac{1}{25}$m$/_{[0,25]}$，$P_3 =\frac{1}{2} P_1 + \frac{1}{2}P_2$ （见例3.1）。对于哪个i /= j，我们
有$P_i\, P_j$?在每个这样的情况下，找出拉登-尼科迪姆导数。

练习7.9

设$\lambda = \delta_0 + m_{[1,3]}$，$\mu = \delta_1 + m_{[2,4]}$，并按推论7.10中的方法找到
λ_a，λ_s，和h。

7.3 Lebesgue-Stieltjes 措施

回顾一下（见第3.5.3节），给定任何随机变量X：Ω-→R，我们将其概

率分布定义为Borel上的措施$P_X = P \circ X^{-1}$。

R上的集合（即我们设定P（$X \le x$）$= P \circ X^{-1}$（（$-\infty$，x]）$= P_X$（（$-\infty$，x]），并将其扩展到B。）设定$F_X(x) = P_X((-\infty, x])$ 我们在命题4.30中验证了这样定义的分布函数F_X是单调递增的、右连续的，其极限为无穷大 $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$ 和 $F_X(+\infty) = \lim_{x\to\infty} F_X(x) = 1$。

在第四章中，我们研究了$F_X(x) = P_X((-\infty, x]) =$ 的特殊情况。$\int^x f_X\, dm$ 对于某个实数函数f_X，P_X的密度相对于L-eb∞esgue度量m，命题4.22表明，如果f_X是连续的，那么F_X是可微的，并且在每个x R处都有密度f_X作为其导数。另一方面，例4.8中的Lebesgue函数说明了F_X的连续性并不足以保证密度的存在。

此外，当F_X有一个密度f_X，测量P_X被说成是关于m的 "绝对连续"。在Radon-Nikodym定理的背景下，我们应该把这个特例的术语与基因的术语相协调。

在本章中考虑的第二个问题。显而易见，当$P_X(B) = \int_B f_X\, dm$

dm

关于m的唯一性确保$\frac{dP_X}{} = f_X$ a.s. 与

我们有P_X m，因此，P_X有一个拉登-尼科迪姆导数 X^{dP}

在本章后面我们将建立对累积分布函数F的精确分析要求$_X$，这将保证密度的存在。

7.3.1 Lebesgue-Stieltjes的构建 措施

为了做到这一点，我们首先研究定义在(R, B)上的度量，这些度量以类似的方式对应于R上的递增、右连续函数。它们的构造与Lebesgue度量的构造一样，只做了一些改变，概括了 "区间长度 "的概念。这些措施

我们得到的分布函数被称为*Lebesgue-Stieltjes度量*。在这种情况下，如果一个函数F：R-→R是单调递增和右连续的，我们称之为*分布函数*。很明显，每一个定义在(R, B)上的有限度量μ都通过$F(x)=\mu((-\infty, x])$来定义这样一个函数，其中$F(-\infty)=\lim_{x\to-\infty} F(x)=0$，$F(+\infty)=\lim_{x\to\infty} F(x)=\mu(\Omega)$。

然而，我们主要关心的是相反的问题：给定一个单调的右连续的F：R R，我们能否总是与F相关联一个关于（Ω，B）的 度量，如果可以，它与Lebesgue度量有什么关系？

第一个问题可以通过仔细回顾第二章中关于R上Lebesgue度量m的构造来回答：首先我们定义了区间长度的自然概念，$l(I)=b-a$，对于任何有端点a，b（$a<b$）的区间I、

并与我们对空集的讨论相类比，我们将R的任意子集的Lebesgue外量m^*定

义为总长度的下限值

$\sum_{n=1}^{\infty} l(I_n)$的所有序列$(I_n)_{n\geq 1}$覆盖$A$的区间。为了概括这个想法，我们应该明确地用$F(b)F(a)$代替$b$ a，-得到一个'广-义的

相对于F的"区间长度"，但由于F只是右连续的，我们将需要照顾到可能的不连续。因此，我们需要确定单调递增函数的可能的不连续性--幸运的是，这些函数是相当乖巧的，正如你可以在下面轻松验证：

命题7.11

如果$F : \mathbb{R} \to \mathbb{R}$是单调递增的（即$x_1 \leq x_2$意味着$F(x_1) \leq F(x_2)$），那么左极限$F(x-)$和右极限$F(x+)$在每个$x \in \mathbb{R}$处都存在，$F(x-) \leq F(x) \leq F(x+)$。因此，$F$最多有无数个不连续点，而且这些不连续点是跳跃不连续点，即$F(x-) < F(x+)$。

提示 对于任何x，考虑$\sup\{F(y) : y < x\}$和$\inf\{F(y) : x < y\}$来验证第一个说法。对于第二个说法，请注意，如果F在x处有不连续，则$F(x-) < F(x+)$. 利用Q在R中是密集的这一事实来表明，只能有无数个许多这样的点。

由于F是单调的，它在有界集上保持有界。为简单起见，我们假设$\lim_{x\to-\infty} F(x) = 0$。我们定义有界区间（$a, b$）的"相对于$F$的长度"为

$$l_F(a, b] = F(b) - F(a)。$$

注意，我们把自己限制在左开右闭的区间。由于F是右连续的，对于所有的x，包括a，b，$F(x+) = F(x)$。因此$l_F(a, b] = F(b+) - F(a+)$，$F$的所有跳跃都有$F(x) - F(x-)$的形式。
通过限制在这种类型的区间内，我们也可以确保l_F是加在
毗连区间：如果$a < c < b$，那么$l_F(a, b] = l_F(a, c] + l_F(c, b]$。

我们对定义2.2做如下概括：

定义7.4

任何集合$A \subseteq \mathbb{R}$的F-*外测量*是$[0, \infty]$的元素

$$m^*_F(A) = \inf Z_F(A)$$

其中

∞ ∞

$$Z_F(A) = \{\sum I_F(I_n) : I_n = (a_n, b_n], a_n \le b_n, A \subseteq \bigcup I_n\}.$$

我们的 "覆盖区间 "现在也被限制为左开和右闭。这对于 "使事情合为一体 "至关重要，但并不影响可测量性：回顾一下（定理2.16），无论我们从所有区间的家族还是从各种子家族开始，都会产生博勒尔σ*场*。

现在更详细地考虑定理2.4的证明：我们在那里的目的是证明区间的外量等于其长度。我们展示了如何调整证明，使这一主张对m^*_F和l_F适用于形式的区间（*a*，*b*）。因此，复习一下以下的证明将是有帮助的

在继续阅读之前，请先阅读定理2.4!

步骤1。$m^*_F((a, b]) \le l_F(a, b]$的证明仍然大致相同：

为了看到$l_F(a, b] \in Z_F((a, b])$，我们用$(I_n)$覆盖(a, *b*]，$I_1 = (a, b]$, $I_n = (a, a] = \varnothing$, $n > 1$。这个序列的总长度是$F(b) - F(a) = l_F(a, b]$，因此结果由inf的定义得出。

第二步。剩下的就是证明$l_F(a, b]$ $m^*_F((a, b])$。在这里，我们需要注意总是 "从右边接近点"，以便利用*F*的右边连续性，从而避免其跳跃。

根据inf的定义，我们可以通过区间$I_n = (a_n, b_n]$ 找到$I = (a, b]$的一个覆盖，这样$\sum_{n\ge 1} l_F(I_n) < m^*_F(I) + \frac{\varepsilon}{2}$。接下来，让$J_n = (a_n, b'_n)$，根据*F*的右连续性，对于每个$n \ge 1$，我们可以选择$b'_n > b_n$，并且$F(b'_n) - F(b_n) < \frac{\varepsilon}{2^{n+1}}$。那么$F(b'_n) - F(a_n) < \{F(b_n) - F(a_n)\} + \frac{\varepsilon}{2^{n+1}}$。

$(J_n)_{n\ge 1}$然后形成一个紧凑区间$[a + \delta, b]$的开放覆盖，因此根据海涅-伯勒定理，有一个有限子家庭$(J_n)_{n\le N}$，它

也覆盖$[a+\delta, b]$。将这N个区间J_n重新排序，我们可以假设它们的右侧端点形成一个递增序列，然后

$$
\begin{aligned}
F(b) - F(a+\delta) = l_F(a+\delta, b] &\le \sum_{n=1}^{N} \{F(b'_n) - F(a_n)\} \\
&< \sum_{n=1}^{N} \{F(b_n) - F(a_n) + \frac{\varepsilon}{2^{n+1}}\} < \sum_{n=1}^{\infty} l_F(I_n) + \frac{\varepsilon}{2} \\
&< m^*_F(I) + \varepsilon。
\end{aligned}
$$

这对所有的$\varepsilon>0$都成立，因此$F(b) - F(a+\delta) \le m^*_F(I)$，对每一个$\delta>0$。根据*F*的右连续性，让$\delta\downarrow 0$，我们得到$l_F(a, b] = \lim_{\delta\downarrow 0} l_F(a+\delta, b] \le m^*_F(a, b)$。这就完成了$m^*_F((a, b]) = l_F(a, b]$的证明。

这是对导致Lebesgue度量的构造唯一需要的实质性改变。证明m^*_F是一个外测量，即

$m^* F(A) \geq 0$、 $m^* F(\varnothing) = 0$、 $m^* F(A) \leq m^* F(B)$ 如果$A \subseteq B$

、

$$m^*_F\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} m^*_F(A_i)、$$

与Lebesgue外在度量的字句相同（Propo- sition 2.3, Theorem 2.5）。因此，如定义2.3，我们说一个集合E是*可测的*（对于外测量m^*_F），如果对于每个$A \subseteq \mathbb{R}$

$$m^*_F(A) = m^*_F(A \cap E) + m^*_F(A \cap E^c)。$$

同样，定理2.8的证明也是逐字进行的，我们用m_F来表示所得到的Lebesgue-Stieltjes度量，即m^*_F限制在*Lebesgue-Stieltjes可测量集*的σ场$\mathcal{M}_F$。根据结构，就像Lebesgue度量一样，m_F是一个完全度量：m_F的子集--空集都在$\mathcal{M}_F$。然而，正如我们稍后将看到的，$\mathcal{M}_F$并不总是与$\mathcal{M}$的σ场相吻合。

Lebesgue可测集，尽管两者都包含所有的Borel集。也可以直接验证第2.4节中所证明的Lebesgue度量的属性对一般的Lebesgue-Stieltjes度量是成立的，只有一个例外：

的外部度量m^*_F，一般来说，不会是平移不变的。我们可以看到对于区间来说，这一点是一次性的，因为$l_F((a+t, b+t]) = F(b+t)\ F(a+t)$通常不会等于$F(b)\ F(a)$；简单来说，以$F(x) = x^3$，为例。事实上，可以证明Lebesgue度量是R上唯一的平移不变的度量。

此外，请注意，一个单子{a}现在不一定是一个空集，因为m_F：根据定理2.13的类比，我们可以得出：

$$m_F(\{a\}) = \lim_{n\to\infty} m_F\left(\left(a - \frac{1}{n}, a\right]\right) = F(a) - \lim_{n\to\infty} F\left(a - \frac{1}{n}\right) = F(a) - F(a-)。$$

因此，集合a的度量恰恰是a处跳跃的大小（如果有的话）。由此，通过类似的论证，我们很容易看到一个区间的 "长度 "是如何取决于其端点的存在与否的：鉴于$m_F((a, b]) = F(b) - F(a)$，我们看到：$m_F((a, b)) = F(b-) - F(a)$，$m_F([a, b]) = F(b) - F(a-)$、$m_F([a, b]) = F(b-) - F(a)$。

例7.1

当$F = 1_{[a,\infty)}$，我们得到$m_F = \delta_a$，狄拉克测量集中在a.同样，我们可以描述一个一般的离散概率分布，其中随机变量X取值$\{a_i : i =$

1, 2, . . ., n}，其概率为
$\{p_i = 1, 2, . . ., n\}$作为Lebesgue-Stieltjes度量，由函数产生的
$F = \sum_{i=1}^{n} p_i 1_{[a_i,\infty]}$。

离散分布和连续分布的混合物，例如在Ex- ample 3.1中描述的，显然也适合于这种情况。当然，Lebesgue度量m是

分布均匀的特殊情况，即如果F（x）=x对所有$x\in$R
那么$m_F = m$。

例7.2

仅仅更一般一点，R上的每一个有限博勒尔度量μ*都*对应于一个Lebesgue-Stieltjes度量，因为分布函数$F(x)=\mu((, x])$显然是增加的，并且通过适用于μ*的*定理2.13，是右连续的。
和区间$I_n = (-\infty, x + \frac{1}{n}]$）。相应的Lebesgue-Stieltjes度量$m_F = \mu$，因为它们在上述形式的区间的生成族上是重合的。因此它们在Borel集的σ场B上是重合的。根据我们对作为完全度量的m_F的构造，可以看出m_F是完成度
的μ。

例7.3

回到例4.8中讨论的Lebesgue函数F。由于F是连续和单调增加的，它在区间[0, 1]上诱导了一个Lebesgue-Stieltjes度量m_F，我们现在来研究它的性质。在每个 "中间三分之二 "的集合上，F都是常数，因此这些区间对于m_F来说是空集，而且由于它们的数量可数，它们的联合也是空集，即 "中间三分之二 "集合，D。因此，康托尔集$C = D^c$满足

$$1 = F(1) - F(0) = m_F([0, 1]) = m_F(C)$$

(注意，由于F是连续的，$m_F(\{0\}) = F(0) - F(0-) = 0$；事实上，每个单子都是$m_F$-空的。)因此，我们得出结论，$m_F$集中在Lebesgue度量$m$的空集上，即$m_F$ ⊥ m，而且在m_F相对于m的Lebesgue分解中，没有绝对连续的成分（通过唯一性
的分解）。

练习7.10

假设单调递增函数F在最多无数个点上是非常数的（如离散分布的情况）。证明R的每个子集都是m_F-可测量的。

提示 先考虑有界区间[-M, M]上的m_F。

练习7.11

找出Lebesgue-Stieltjes度量m_F产生的

$$F(x)=\begin{cases}0, & \text{如果 } x<0、\\ 2x, & \text{如果 } x\in[0,1],\\ 2 & \text{如果 } x\geq 1。\end{cases}$$

7.3.2 函数的绝对连续性

现在我们来讨论一下对分布F的要求，以确保它有一个密度。正如我们在例4.8中所看到的，概率分布函数的连续性并不能保证密度的存在。然而，下面这个更强的限制条件可以解决这个问题：

定义7.5

一个实数函数F在区间$[a,b]$上是*绝对连续的*，如果，给定$\varepsilon>0$、有$\delta>0$，这样对于每一个有限的不相交的区间集合$J_k=(x_k,y_k)$, $k\leq n$，包含在$[a,b]$中，并且$\sum_{k=1}^{n}(y_k-x_k)<\delta$，我们有$\sum_{k=1}^{n}|F(x_k)-F(y_k)|<\varepsilon$。

这个条件将使我们能够确定那些产生相对于Lebesgue度量而言绝对连续（在度量的意义上）的Lebesgue-Stieltjes度量的分布函数。我们很快就会看到，绝对连续的函数也是 "有界变化的"：这描述了在小区间内不会 "变化太大 "的函数。首先我们验证相对于密度的不定积分（见命题4.22）是绝对连续的。

命题7.12

如果$f\in L^1([a,b])$，其中区间$[a,b]$是有限的，那么函数$F(x)=\int_a^x f\,dm$是绝对连续的。

提示 使用$\mu(G)=\int_G f\,dm$相对于Lebesgue度量m的绝对连续性。

练习7.12

决定以下哪些函数是绝对连续的：(a)

$f(x) = x|, x \in [-1, 1]$, (b) $g(x) = \sqrt{x}, x \in [0, 1]$, (c) Lebesgue函数。

下一个结果是上述例子的重要反面，它表明所有由绝对连续函数产生的斯蒂尔杰斯积分都会导致相对于勒贝斯格度量而言是绝对连续的度量，因此有一个密度。与该例一起，这也是密度所产生的分布的特征（在我们对分布函数施加的条件下）。

定理7.13

如果F在R上是单调增加和绝对连续的，让m_F是它产生的Lebesgue-Stieltjes度量。那么每个Lebesgue可测集都是m_F-可测的，在这些集上m_F m。

证明

我们首先证明，如果Borel集B有$m(B)=0$，那么也有$m_F(B)=0$。回顾一下，给定$\delta>0$，我们可以找到一个包含B的开放集O，且$m(O)<\delta$(定理2.12)，有一个不相交的开放区间序列$(I_k)_{k\geq 1}$、
$I_k = (a_k, b_k)$，其联合体为O。由于区间不相交，其总长度为
根据F的绝对连续性，给定任何$\varepsilon>0$，我们可以找到$\delta>0$。
这样，对于每一个有限的区间序列$J_k = (x_k, y_k)$, $k \leq n$, 总数为
长度$\sum^n (y_k - x_k) < \delta$，我们有$\sum^n \{F(y_k) - F(x_k)\} < \frac{\varepsilon}{2}$。应用这个
对固定n的序列$(I_k)_{k\leq n}$，我们得到$\sum_{k=1}^{n} \{F(b_k) - F(a_k)\} < \frac{\varepsilon}{2}$。
由于这对每个n都成立，我们也有$\sum_{k=1}^{\infty} \{F(b_k) - F(a_k)\} \leq \frac{\varepsilon}{2} < \varepsilon$。这是覆盖
$O \supset B$的不相交区间序列的总长度，因此
$m_F(B) < \varepsilon$ 对于每一个$\varepsilon > 0$，所以$m_F(B) = 0$。

现在对于每一个Lebesgue可测量集E，$m(E)=0$，我们可以找到一个Borel集$B \supseteq E$，$m(B)=0$。因此，$m_F(B)=0$。现在E是一个m_F-null集的子集，因此它也是m_F-null。因此，所有 *可*测量 *的*集合都是m_F-的。
可测量的和*m-null的*集合是m_F-null，即当两者都被视为M上的测量时，m_F m。□

与拉登-尼科迪姆定理一起，上述结果有助于澄清由单调增加的右连续函数产生的勒贝斯格度量和勒贝斯格-斯蒂尔杰斯度量之间的结构关系，因此，特别是对概率分布而言：当函数F是绝对连续的，它有一个

密度f，因此可以写成

作为其 "不定积分"。由于Lebesgue-Stieltjes度量m_F m，所以

Radon-Nikodym导数$\frac{dm_F}{dm}$是定义良好的。反之，对于密度f

F的存在，函数F必须是绝对连续的。现在剩下的就是

澄清Radon-Nikodym导数$\frac{dm_F}{dm}$和

密度f。从连续的f的例子（命题4.22）可以很自然地想到，f应该是F的导数（至少是m-a.e）。因此我们需要了解F的哪些条件将确保$F'(x)$存在于m-almost

所有的x R。

我们将在更广泛的背景下讨论这一问题，即

积分器 "函数F不再一定是单调增加的，而是具有有界的变化，正如下一节所介绍的。

7.3.3 有界的函数 变化的函数

因为在一般情况下，我们需要处理可以取负值的集合函数，例如，映射

$$E \dashrightarrow \int_E g\, dm，其中 g \in L^1(m)、$$

因此，我们需要一个 "广义长度函数 "的概念，它被认为是两个单调递增函数的差。我们首先需要确定这种函数的特征。这可以通过引入以下内容来实现

定义7.6

一个实数函数F在$[a, b]$上是*有界变化的*（简言之$F \in BV[a, b]$），如果$T_F[a, b] < \infty$， 其中对于任何$x \in [a, b]$

$$T_F[a, x] = \sup\{\sum_{k=1}^{n} |F(x_k) - F(x_{k-1})|\}。$$

的所有有限分区中取最高值，其中$a = x_0 < x_1 < \ldots < x_n = x_o$

我们再引入两个非负函数，设定为

和

$$P_F[a, x] = \sup\{\sum_{k=1}^{n}[F(x_k) - F(x_{k-1})]\}^+$$

$$N_F[a, x] = \sup\{\sum_{k=1}^{n}[F(x_k) - F(x_{k-1})]\}^-$$

其中最高值又是取自$[a, x]$的所有分区。函数T_F (P_F, N_F)分别被称为F的*总（正，负）变化函数*。在下文中，我们将保持a的固定，并将其视为x的函数，用于x a。≥

我们可以很容易地验证这些定义之间的以下基本关系：

命题7.14

如果F在$[a，b]$上是有界变化的，我们有$F（x）-F（a）=P_F（x）-N_F（x）$，而$T_F（x）=P_F（x）+N_F（x）$，对于$x\in[a，b]$。

提示 考虑$p(x) = \Sigma^n_{k=1} [F(x_k) - F(x_{k-1})]^+$和$n(x) = \Sigma^n_{k=1} [F(x_k) - F(x_{k-1})]^-$对于$[a, x]$的一个固定分区，注意$F(x) - F(a) = p(x) - n(x)$。现在使用上位数的定义。对于第二个特征，考虑$T_F(x) \geq p(x) + n(x) = 2p(x) - F(x) + F(a)$ 并使用第一个特征。

命题7.15

如果F是有界变化的，并且$a\leq x\leq b$，那么$T_F[a, b] = T_F[a, x] + T_F[x, b]$。类似的结果也适用于$P_F$和$N_F$。因此，对于固定的$a\in$R，所有三个变化函数在$x$中都是单调递增的。此外，如果$F$在$[a, b]$上具有有界变化，那么它在任何$[c, d]\subset[a, b]$上都具有有界变化。

提示 在一个分区中增加一个点会增加所有三个分区的总和。另一方面，将$[a, c]$和$[c, b]$的分区放在一起，我们得到$[a, b]$的分区。

我们表明，有限区间上的有界变化函数正是我们要找的：

定理7.16

设$[a，b]$是一个有限区间。当且仅当一个实数函数是$[a, b]$上两个单调递增的实数函数之差时，它在$[a, b]$上是有界变化的。

证明

如果F是有界变化的，用$F（x）=[F（a）+P_F（x）]- N_F（x）$从命题7.14来表示$F$是两个单调递增函数之差。

反过来说，如果$F=g-h$是两个单调递增的函数之差，那么对于[a，b]的任何分区$a=x_0 < x_1 < . . . < x_n = b$，我们得到，因为$g$，$h$是递增的、

$$\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| = \sum_{i=1}^{n} |g(x_i) - h(x_i) - g(x_{i-1}) + h(x_{i-1})|$$
$$\leq \sum_{i=1}^{n} [g(x_i) - g(x_{i-1})] + \sum_{i=1}^{n} [h(x_i) - h(x_{i-1})]$$
$$\leq g(b) - g(a) + h(b) - h(a)。$$

因此，$M = g(b) - g(a) + h(b) - h(a)$ 是一个独立于分区选择的上界，所以$T_F [a, b] \leq M < \infty$， 符合要求。 □

这种分解是最小的：如果$F = F_1 - F_2$ ，并且F_1 , F_2 是增加的，那么对于任何分区$a = x_0 < x_1 . . . < x_n = b$，我们可以写，对于固定的$i \leq n$

$$\{F(x_i) - F(x_{i-1})\}^+ - \{F(x_i) - F(x_{i-1})\}^- = F(x_i) - F(x_{i-1})$$
$$= \{F_1(x_i) - F_1(x_{i-1})\} - \{F_2(x_i) - F_2(x_{i-1})\}$$

从$x = x^+ x^-$ 的最小属性可以看出，右边的差额中的每项都支配着左边的对应项。加起来并取上位数，我们得出结论：P_F 被F_1 的总变化所支配，N_F 被F_2 的总变化所支配。换言之，在增函数的集合中
其差值为F，函数($F(a)+P_F$)和N_F 的和最小。
在[a, b]的每一个点上。

练习7.13

(a) 设F在[a, b]上是单调递增的。求$T_F [a, b]$。

(b) 证明：如果$F \in BV [a, b]$，则F是连续的，即(m)和Lebesgue可测的。

(c) 找到一个不在BV[0, 1]中的可微函数。

(d) 证明如果有一个（Lipschitz）常数M>0，使得$|F(x) - F(y)| \leq M |x - y|$ 对于所有的x，$y \in [a, b]$，那么$F \in BV [a, b]$。

以下简单事实将有界变化和有界区间[a, b]上的函数的绝对连续性联系起来：

命题7.17

假设实数函数F在$[a, b]$上绝对连续；那么我们有：

(i) $F \in BV[a, b]$、

(ii) 如果$F = F_1 - F_2$是F作为定理7.16中描述的两个单调递增函数之差的最小分解，那么F_1和F_2在$[a, b]$上都是绝对连续。

提示 给定$\varepsilon>0$，选择$\delta>0$，如定义7.5中。在(i)中，从$[a, b]$的任意分区(x_i)开始，我们不能使用F的绝对连续性，除非我们知道子区间的长度为δ。因此增加足够的新分区点以保证这一点，并考虑它们产生的和。对于(ii)，比较在间隔长度之和受δ约束的分区上求和时的各种变化函数。

定义7.7

如果$F \in BV[a, b]$，其中$a, b \in$ R，让$F = F_1 - F_2$是其最小分解为单调递增函数。定义F的Lebesgue-Stieltjes*符号度量*为Borel集的σ场B上给出的可加集函数m_F，即$m_F = m_{F_1} - m_{F_2}$，其中m_{F_i}为Lebesgue-Stieltjes度量的F_i，（$i = 1$，2）。

我们将在下一节更广泛地研究签名措施。就目前而言，我们注意到以下几点

例7.4

当考虑由密度f_X在R上诱导的度量$P_X(E) = \int_E f_X\,dm$时，我们将注意力限制在$f_X \geq 0$，以确保P_X是非负的。但是对于一个可测量的函数f：R→R，我们设定（定义4.4）

$$\int_E f\,dm = \int_E f^+\,dm - \int_E f^-\,dm \text{ 每当 } \int_E |f|\,dm < \infty。$$

由$v(E)=\int_E f\,dm$定义的集合函数v然后自然分裂为两种措施之差，即$v = v^+ - v^-$，其中$v^+(E) = \int_E f^+\,dm$和$v^-(E) = \int_E f^-\,dm$。限制在$[a, b]$上支持的函数f，并设定$F(x) = \int_a^x f\,dm$我们得到$m_F = v$，如果$F = F_1 - F_2$如上所述定义，那么$m_{F_1} = v^+$，$m_{F_2} = v^-$，根据最小化的特性。F的拆分。

7.3.4 已签署的 措施

上面的例子和由BV函数生成的Lebesgue-Stieltjes度量的定义促使了下面的抽象定义和随后对两个度量之差的类似分解的探索。我们需要在抽象环境中简要概述有符号度量的结构，这为上述Stieltjes整数和分布函数的发展提供了一般背景。我们的结果将使我们能够通过参考有符号度量的 "正负部分 "的分解来定义相对于有符号度量的函数积分，就像上面所说的。我们还得到了更普遍的Lebesgue分解和Radon-Nikodym定理，从而完成了对有界有符号度量相对于给定σ-无限度量的结构的描述。这就导致了前面提到的微积分基本定理的一般版本。

定义7.8

可测量空间（Ω，$\mathcal{F}$）上的*有符号度量*是一个集合函数ν : $\mathcal{F} \to (-\infty, +\infty]$，满足于

(i) $\nu(\varnothing) = 0$

(ii) $\nu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \nu(E_i)$ 如果$E_i \in \mathcal{F}$ 和$E_i \cap E_j = \varnothing$ for $i \neq j$。

我们需要通过要求v最多取一个值来避免类似的模糊性 要求ν最多只能取其中一个值$+\infty$ $-\infty$；因此我们一致要求$\nu(E) > -\infty$

对于其域中的所有集合E。还请注意，在(ii)中，要么两边都是$+\infty$，要么都是有限的，因此数列在R中收敛。由于左边不受数列项的任何重新排列的影响，因此，只要数列收敛，就会绝对收敛，即$\sum_i |\nu(E_i)| < \infty$ 如果并且只有当$|\nu(\bigcup_{i=1}^{\infty} E_i)| < \infty$。在激励性的例子中，收敛性很明显，因为对于任何$E \subseteq \mathbb{R}$，我们有

$$|\nu(E)| = |\int_E f\,dm| \leq \int_E |f|\,dm < \infty \text{ 当} f \in L^1(\mathbb{R}) \text{时。}$$

请注意，ν是有限加性的（让$E_i = \varnothing$，对于(ii)中的所有$i > n$，那么(i)意味着

$\nu(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \nu(E_i)$，如果$E_i \in \mathcal{F}$和$E_i \cap E_j = \varnothing$ for $i \neq j, i, j \leq n$)。

因此，如果$F \subseteq E$，$F \in \mathcal{F}$，且$|\nu(E)| < \infty$，则$|\nu(F)| < \infty$，因为 的两边

$\nu(E) = \nu(F) + \nu(E \setminus F)$ 是有限的，并且$\nu(E \setminus F) > -\infty$的假设。

有符号的措施不继承没有变化的措施的属性：

作为否定的结果，我们有

命题7.18

定义在σ场F上的有符号度量v是单调增加的（$F\subset E$
意味着$v(F)\leq v(E)$)，如果并且只有v是F上的一个度量。

提示∅是每个$E\in$F的子集！

另一方面，一个有符号的度量在……中的某些集合达到了它的界限。F
更确切地说：给定$(\Omega,$ F$)$ 上的有符号度量v，可以在其中找到集合A和B，
使$v(A)=\inf\{v(F): F\in F\}$ 而$v(B)=\sup\{v(F): F\in F\}$.

我们将不直接证明这个结果，而是从哈恩-乔丹分解定理中推导出它。这个基本结果表明，集合A和它的补集可以用来定义两个（正）度量
v^+,v^-，以便
$v=v^+-v^-$，其中$v^+(F)=v(F\cap A^c)$，$v^-(F)=-v(F\cap A)$ 对于所有$F\in$F。
分解是最小的：如果$v=\lambda_1-\lambda_2$，其中λ_i是措施，那么
$v^+\leq\lambda_1$ 和 $v^-\leq\lambda_2$。

将注意力限制在有界的有符号措施上（这对于应用来说已经足够了）。我们可以通过应用拉登-尼科迪姆定理推导出这种分解。(我们的说法是处理的一个特例

在[10]第6章中给出了复值集合函数的方法）。首先，给定一个有界的有符号度量v：F-→R，我们寻求支配v的最小（正）度量μ，即满足$\mu(E)\geq|v(E)|$对于所有$E\in$F。

$$|v|(E)=\sup\left\{\sum_{i=1}|v(E_i)|:\{E_i\}\subset F,\ E=\bigcup_{i\geq1}E_i,\ E_i\cap E_j=\varnothing \text{ if } i\neq j\right\}.$$

产生一个集合函数，对每一个E都满足$|v|(E)\geq|v(E)|$。然后，对所有i的再要求$\mu(E_i)\geq|v(E_i)|$，就可以得到

$$\mu(E)=\sum_{i=1}^{\infty}\mu(E_i)\geq\sum_{i=1}^{\infty}|v(E_i)||$$

对于任何支配v的度量μ。因此，为了证明$|v|$具有所需的特性，我们只需要证明它是可数加性的。我们称$|v|$为v的*总变化*。请注意，我们在这里使用Ω的可数分区，就像我们使用
在定义R中的Lebesgue度量时，区间的序列。

定理7.19

有界有符号度量的总变异|v|是一个（正）度量，在
F.

证明

将E划分为集合E_i，在R^+中选择(a_i)，使$a_i < v(E_i)$为所有i。
分为E_i，在$\mathbb{R}^+$中选择$\{a_i\}$，使$a_i < v(E_i)$ for all i。
将每个E依次划分为A集，并通过sup的定义
我们可以选择这些以确保$a_i < \sum_j |v(A_{ij})|$的每一个$i \geq 1$。但$\{A_{ij}\}$也会分割E，因此$\sum_i a_i < \sum_{i,j} |v(A_{i,j})| < |v|(E)$。对所有满足这些要求的序列$(a_i)$取最高值可以确保

$$\sum_i |v|(E_i) = \sup \sum_i a_i \leq |v|(E)。$$

对于相反的不等式，考虑E的任何分区$\{B_k\}$，并注意到对于固定的k，$\{B_k \cap E_i\}_{i\geq 1}$分区$B_k$，而对于固定的$i$，$\{B_k \cap E_i\}_{k\geq 1}$分区$E_i$。这意味着

$$\sum_{k\geq 1} |v(B_k)| = \sum_{k\geq 1} \Big| \sum_{i\geq 1} v(B_k \cap E_i)\Big| \leq \sum_{k\geq 1}\sum_{i\geq 1} |v(B_k \cap E_i)|。$$

由于双系列的项都是非负的，我们可以交换求和的顺序，所以最后

$$\sum_{k\geq 1} |v(B_k)| \leq \sum_{i\geq 1}\sum_{k\geq 1} |v(B_k \cap E_i)| \leq \sum_{i\geq 1} |v|(E_i)。$$

但E的分区$\{B_k\}$是任意的，所以右边的估计也支配了$|v|(E)$。这就完成了$|v|$是一个度量的证明。 □

我们现在定义有符号措施的*正（或负）变化*
v通过设置：

$$v^+ = \frac{1}{2}(|v| + v)、\quad v^- = \frac{1}{2}(|v| - v)。$$

显然，两者都是F上的正度量，而且我们有

$$v = v^+ - v^-，\quad |v| = v^+ + v^-。$$

有了这些定义，我们可以立即将Radon-Nikodym和Lebesgue分解定理扩展到v是有界有符号的情况（我们保留7.3.2节中使用的符号，所以这里μ仍然是正的！）：

定理7.20

设μ是σ-无限（正）度量，并假设v是一个有界的有符号度量。那么就有

唯一的分解 $\nu=\nu_a+\nu_s$，分解为两个有符号的
措施，其中$\nu_a\ \mu$和ν_s ▯ μ。此外，有一个唯一的（至于集合的
0）$h\in L^1(\mu)$，使得ν_a （F）$=\int_F h\,\mathrm{d}\mu$，对于所有$F\in\mathcal{F}$。

证明

给定$v=v^+ - v^-$，我们希望将Lebesgue分解和Radon Nikodym定理应用于有限度量对（v^+, μ）和（v^-, μ）。首先我们需要检查，对于一个有符号的度量λ μ，我们也有$|\lambda|$ μ（因为那时显然λ^+ μ和λ^- μ）。但如果$\mu(E)=0$，F对E进行分区，那么每个$\mu(F_i)=0$，因此$\lambda(F_i)=0$，所以$\sum_{i\,1}|\lambda(F_i)| = 0$。由于这对每个分区都成立，$|\lambda|(E)| = 0$。

同样地，如果λ集中在一个集合A上，并且$A\cap E=\varnothing$，那么对于任何E的分区$\{F_i\}$，我们将有$\lambda(F_i)=0$，每i 1。因此，$|\lambda|(E)=0$、所以$|\lambda|$也集中在A上。因此，如果两个有符号的度量是互为奇异的，那么它们的总变化度量也是互为奇异的，从而也是它们的正

和负变化。将Lebesgue分解和Radon Nikodym定理应用于度量v^+和v^-，可以得到（正）度量$(v)^+_a, (v)^+_s, (v)^-_a, (v)^-_s$，这样$v^+ = (v)^+_a + (v)^+_s$和$(v)^+_a(F) = \int_F h' \,d\mu$，而$v^- = (v)^-_a + (v)^-_s$和$(v)^-_a(F) = \int_F h'' \,d\mu$，对于非负函数$h', h'' \in L^1(\mu)$，并且措施$(v)^+_s, (v)^-_s$各自相互之间以$\mu$为奇数。让$v_a = (v)^+_a - (v)^-_a$我们得到一个有符号的度量$v_a$ μ、和一个函数$h = h' - h'' \in L^1(\mu)$，对于所有$F\in$F，$v_a(F) = \int_F h \,d\mu$。有符号的度量$v_s = (v)^+_s - (v)^-_s$对$\mu$来说显然是奇异的，而且$h$在$\mu$空集之前是唯一的，因为这对$h', h''$来说是成立的，而且分解$v = v^+ - v^-$是最小的。 □

例7.5

如果$g\in L^1(\mu)$，那么$v(E) = \int_E g \,d\mu$是一个有符号的度量，并且v μ。Nikodym theorem shows that (with our conventions) all signed measures v μ 有这种形式。

我们几乎准备好了微积分基本定理的一般形式。首先，我们确认，正如从拉登-尼科迪姆定理的证明中可以预料的那样，有界的导数之间的密切关系

由R上有界的有符号（Borel）度量v和导数$f = F'$引起的变化函数F：

定理7.21

如果v是R上有界的有符号度量，$F(x)=v((-\infty, x])$，那么对于任何$a\in$R，以下是等价的：

(i) F在a处是可微的，并且$F'(a) = L$。

(ii) 给定$\varepsilon>0$，存在$\delta>0$，使得$\left|\frac{v(J)}{m(J)}-L\right|<\varepsilon$，如果开放区间

J包含a且$l(J) < \varepsilon$。

证明

我们可以假设L=0；否则就考虑$\rho = \nu\ Lm$，而限制在包含a的有界区间。

如果(i)在L=0时成立，并且给定了$\varepsilon>0$，我们可以找到$\delta>0$，从而

$$|F(y) - F(x)| < \varepsilon|y - x| \text{ 只要} |y - x| < \delta。$$

设J=（x，y）是一个包含a的开放区间，且（y-x）$<\delta$。对于足够的大的N，我们可以确保$a>x+\frac{1}{N}>x$，因此对于$k\geq1$，$y_k = x+\frac{1}{N}$，是上面以a为界，当$k\to\infty$时，减小到x。因此

$$\begin{aligned}|\nu(y_k, y])| &= |F(y) - F(y_k)| \leq |F(y) - F(a)| + |F(a) - F(y_k)| \\ &\leq \varepsilon\{(y - a) + (a - y_k)\} < \varepsilon m(J)。\end{aligned}$$

但是由于$y_k \to x$，$\nu(y_k, y] \to \nu(x, y]$，我们已经证明$\left|\frac{\nu(J)}{m(J)}\right| < \varepsilon$，因此(ii) 持。对于反面，让$\varepsilon$，$\delta$与(ii)中一样，以便在$x < a < y$和$y - x < \delta$ (ii)意味着$|\nu(x, y+\frac{1}{n})| < \varepsilon(y+\frac{1}{n} - x)$ 对于所有足够大的n。

如$(x,] = \bigcap_n (x, y+\frac{1}{n})$，我们也有

$$|\nu(x, y]| < |F(y) - F(x)| < \varepsilon(y - x)。 \tag{7.3}$$

最后，由于(ii)成立，对于任何包含a的足够小的开放区间I，$\nu(a)$ $\nu(I) < \varepsilon l(I)$。因此$F(a) = F(a^-)$，所以$F$在$a$处是连续的。

这表明左手边的导数也是如此。因此，$F'(a) = 0$，所以(i)持有。 □

定理7.22（微积分基本定理）

让F在$[a, b]$上绝对连续。那么F是可微的m-a.e.，其

Lebesgue-Stieltjes有符号度量m_F有Radon-Nikodym导数$\frac{dm_F}{dm}=$ F' m-a.e. 此外，对于每个$x\in[a, b]$、

$$F(x) - F(a) = m_F[a, x] = \int_a^x F'(t)\,dt。$$

证明

Radon-Nikodym定理提供了$\frac{dm_F}{dm} = h \in L^1(m)$，这样$m_F(E) = \int_E h\, dm$，对所有$E \in$ B来说，选择分区

$$P_n = \{(t_i, t_{i+1}] : t_i = a + \frac{i}{2^n}(b - a), i \le 2^n\}$$

我们依次得到每个P_n作为分区$P_1, P_2, \ldots\ldots$的最小共同细化。, P_{n-1}。因此，设定$h_n(a) = 0$和

$$h_n(x) = \sum_{i=1}^{2^n} \frac{m_F[(t_i, t_{i+1}]]}{m(t_i, t_{i+1}]} 1_{(t_i, t_{i+1}]} = \sum_{i=1}^{2^n} \frac{F(t_{i+1}) - F(t_i)}{t_{i+1} - t_i} 1_{(t_i, t_{i+1}]} \text{ 对于} a < x \le b、$$

我们得到一个序列(h_n)，对应于在证明Radon-Nikodym定理的步骤2中构建的序列$(h_Q\ n)$。由此可见，$h_n(x) \to h(x)$ m-a.e.。但对于任何固定的$x \in (a, b)$，定理4.2中的条件(ii)适用于长度小于δ的每个区间(t_i, t_{i+1})上的函数F，并且$L = h(x)$，表明$h = F'$ m-a.e.。现在从定义上看，最后的要求是明显的。

□

因此，下面的结果是直接的，它证明了在这种一般情况下 "不确定积分 "这一术语的合理性。

推论7.23

如果F在$[a, b]$上是绝对连续的，并且$F' = 0$ m-a.e.，那么F是常数。

最后一个推论现在完成了分布函数及其密度的思路圈：

推论7.24

如果$f \in L^1([a, b])$，并且$F(x) = \int_a^x f\, dm$，对于每个$x \in [a, b]$，那么F是可微的。

m-a.e.和$F'(x) = f(x)$ 对于几乎每一个$x \in [a, b]$。

作为Radon-Nikodym定理的进一步应用，我们推导出前面概述过的ν的Hahn-Jordan分解。首先我们需要以下条件

定理7.25

让ν是一个有界的有符号度量，让$|\nu|$是它的总变化。那么

我们可以找到一个可衡量的函数 h 这样 $|h(\omega)| = 1$，对于所有 $\omega \in \omega$ 和 $\nu(E) = \int_E h \, d|\nu|$ 对于所有$E\in$F。

证明

Radon-Nikodym定理提供了一个可测量的函数h，对于所有$E\in$F，$\nu(E)=\int_E h \, d|\nu|$，因为每个|ν|空集都是ν空 的 （{E, ∅, ∅, . .}是E的一个分区）。让$C_\alpha = \{\omega : |h(\omega)| < \alpha\}$，$\alpha > 0$。然后，对于任何分区 $\{E_i\}$的C_α、

$$\sum_{i\geq 1} |\nu(E_i)| = \sum_{i\geq 1} \int_{E_i} h \, d|\nu| \leq \sum_{i\geq 1} \alpha|\nu|(E_i) = \alpha|\nu|(C_\alpha)。$$

由于这对任何分区都是成立的，所以对它们的最高值也是成立的，即$|\nu|(C_\alpha) \leq \alpha |\nu|(C_\alpha)$。对于$\alpha<1$，我们必须得出结论，$C_\alpha$ 是ν-null，因此也是ν-*null*。因此，$|h| \geq 1$ ν-*a*.e.

为了证明$|h| \leq 1$ ν-*a*.e.我们注意到，如果E具有正的|ν|度量，那么，根据h的定义、

$$\frac{\int_E h \, d|\nu|}{|\nu|(E)} = \frac{|\nu(E)|}{|\nu|(E)} \leq 1。$$

这意味着|h| ≤ 1 ν-*a*.e.从下面的命题中可以看出，应用$\rho = |\nu|$。因此，|h| /= 1的集合是|ν|空的，因此也是ν空的，我们可以在那里重新定义h，这样$|h(\omega)| = 1$对于所有$\omega\in\omega$。 □

命题7.26

给定一个有限度量ρ和一个函数$f\in L^1(\rho)$，假设对每个 $E\in$F，$\rho(E)>0$，我们有$\left|\frac{1}{\rho(E)}\int_E f \, d\rho\right|\leq 1$。那么$|f(\omega)| \leq 1$, ρ-a.e.

提示 让$E=\{f>1\}$。如果$\rho(E)>0$，考虑$\int_E \frac{f}{\rho(E)} d\rho$。

我们准备非常简单地推导出哈恩-乔丹的分解：

命题7.27

让ν是一个有界的有符号度量。有一些不相交的可测量集A，B，使得$A\cup B=\Omega$，并且$\nu^+(F)=\nu(B\cap F)$，$\nu^-(F)=\nu(A\cap F)$，对于所有F F。因此，如果$\nu=\lambda_1 - \lambda_2$，对于措施$\lambda_1$，$\lambda_2$，那么$\lambda_1 \geq \nu^+$ 和 $\lambda_2 \geq \nu^-$。

提示 由于$d\nu=h\,d|\nu|$和$|h|=1$，所以让$A=\{h=-1\}$，$B=\{h=1\}$。使用ν的定义$^+$，以证明$\nu^+(F)=\frac{1}{2}\int_F(1+h)\,d|\nu|=\nu(F\cap B)$对于每一个$F$。

练习7.14

让ν是一个有界的有符号度量。证明对于所有的F，$\nu^+(F)=\sup_{G\subset F}\nu(G)$, $\nu^-(F)=-\inf_{G\subset F}\nu(G)$, 所有的相关集合都是F的成员。

提示 $\nu(G)\le \nu^+(G)\le \nu(B\cap G)+\nu((B\cap F)\setminus(B\cap G))=\nu(B\cap F)$。

练习7.15

证明当$\nu(F)=\int_F f\,d\mu$，其中$f\in L^1(\mu)$，其中μ是一个（正）的衡量，哈恩分解集是$A=\{f<0\}$和$B=\{f\ge 0\}$，并且$\nu^+(F)=\int_F f^+\,d\nu$，而$\nu^-(F)=\int_F f^-\,d\nu$。

我们最终得出了一个相对于有符号测量的积分的一般定义：

定义7.9

设 μ 是有符号的度量，并且f上的一个可测函数。 $F\in\mathcal{F}$。 定义积分$\int_F f\,d\mu$如下

$$\int_F f\,d\mu=\int_F f\,d\mu^+-\int_F f\,d\mu^-$$

只要右边的两个项都是有限的或者不是$\pm(\infty-\infty)$的形式。

如果这样定义的积分是有限的，该函数有时被称为*可求和的*。请注意，先前对Lebesgue-Stieltjes有符号度量的定义适合于这个一般框架。我们通常把注意力限制在两个项都是有限的情况下，当μ是有界的时候，这显然成立。

练习7.16

验证如下：设μ是一个有限度量，定义有符号的$\nu(F)=\int_F g\,d\mu$来衡量ν。证明$f\in L^1(\nu)$，当且仅当$fg\in L^1(\mu)$ 和 $\int_E f\,d\nu=\int_E fg\,d\mu$ 对于所有μ*可*测量*的*集合E

7.4 P robability

7.4.1 相对于σ- 场的条件期望值

假设我们得到了一个随机变量$X \in L^1(P)$，其中(Ω, F, P)是一个概率空间。在第五章中，我们将$X \in L^2(P)$相对于F的子σ场G的条件期望值$E(X|G)$定义为满足以下条件的a.s.唯一随机变量$Y \in L_2(G)$。

$$\int_G Y\,dP = \int_G X\,dP\text{，对于所有}G \in \mathcal{G}。\tag{7.4}$$

该结构是希尔伯特空间L^2中正交投影的结果，其扩展到所有可整数的随机变量是 "手工 "进行的，这需要一点小心。有了Radon-Nikodym定理在手，我们可以非常简单地验证可整数随机变量的条件期望的存在：

有界度量$v(F)=\int_F X\,dP$对于P*来说*是绝对连续的。将两种度量都限制在(Ω, G)上，就能保持这一点。

关系，因此有一个可测量的G P-但根据定义，$v(G)=\int_G X\,dP$，Y 所以$Y=E(X|G)$的定义方程(7.4)已经得到了验证。

备注7.2

特别是，这表明对于$X \in L^2(F)$，其对$L^2(G)$的正交投影是措施v：$F \to \int_F X\,dP$的Radon-Nikodym导数的一个版本。

从现在开始，我们将用$E(X|G)$代替Y，并始终牢记我们有选择特定 "版本 "的自由，也就是说，只要我们寻求的结果只要求有关$E(X|G)$的关系成立P-a.s.，我们就可以在空集上改变这个随机变量而不影响定义的真实性。

方程式：

定义7.10

一个随机变量$E(X|G)$被称为X相对于σ-场G的*条件期望*，如果

(1) $E(X|G)$是可测量的G、

(2) $\int_G E(X|G)\,dP = \int_G X\,dP$，对于所有$G \in \mathcal{G}$。

我们研究条件期望值的属性。首先，最简单的作为一个命题留给读者。在这个和随后的定理中，我们做了以下假设：

(i) 所有相关的随机变量都定义在一个概率空间（Ω，F，P）上；

(ii) 下面使用的X、Y和所有(X_n)都假定在$L^1(\Omega, F, P)$中；

(iii) G和H是F的子σ-场。

下一个命题中列出的属性是基本的，而且被反复使用。在适当的地方，我们用关于X的信息对其 "意义 "进行口头描述。

命题7.28

条件期望值$E(X|G)$具有以下特性：

(i) $E(E(X|G))=E(X)$
(更确切地说：X的任何版本的条件期望值具有与X的期望值相同）。

(ii) 如果X是G可测量的，那么$E(X|G) = X$
(如果给定G，我们已经 "知道 "X，我们对它的 "最佳估计 "是完美的）。

(iii) 如果X是独立于G的，那么$E(X|G)=E(X)$
(如果G对X "一无所知"， 我们对X的最佳猜测就是其平均值）。

(iv) (线性) $E((aX + bY)|G)= aE(X|G)+ bE(Y|G)$ 对于任何实数a，b
(注意，这实际上是说，每一个线性组合的Ver-解释为右边的版本是左边的版本）。

定理7.29

以下属性对上述定义的$E(X|G)$是成立的：

(i) 如果$X \geq 0$，则$E(X|G) \geq 0$ a.s.
(positivity)。

(ii) 如果$\{X_n\}_{n\geq 1}$ 是非负的，并且对X来说是增加的，那么$\{E(X_n|G)\}$逐渐增加到$E(X|G)$
(条件期望的'单调收敛')。

(iii) 如果Y是G可测量的，XY是可整定的，那么$E(XY|G) = YE(X|G)$("去掉一个已知因素")。

(iv) 如果$H \subset G$，则$E([E(X|G)]|H) = E(X|H)$
(塔楼的财产)。

(v) 如果$j:\mathbb{R}\to\mathbb{R}$是一个凸函数，并且$j(X)\in L^1(P)$，则

$$E(j(X)|\mathcal{G})\ge j(E(X|\mathcal{G}))。$$

(这被称为条件詹森不等式--一个类似的结果也适用于期望。回顾一下，如果对于所有的$x，y\in(a，b)$，$j(px+(1-p)y)\le pj(x)+(1-p)j(y)$，一个实数函数$j$在（$a$，$b$）上是*凸的；j*的图形保持在连接（$x$，$j(x)$），（$y$，$j(y)$）的直线上或下面）。

证明

(i) 对于每一个$k\ge1$的集合$E_k=\{E(X/\mathcal{G})<-\frac{1}{k}\}\in\mathcal{G}$，所以

$$\int_{E_k} X\,dP=\int_{E_k} E(X|\mathcal{G})\,dP。$$

由于$X\ge0$，左手边是非负的，而右手边由$-\frac{1}{k}P(E_k)$来约束。这就迫使$P(E_k)$对每一个k都$=0$，因此也是
$P(E(X/\mathcal{G})<0\}=P(\bigcup_k E_k)=0$。因此，$E(X/\mathcal{G})$

(ii) 对于每个n，让Y_n是$E(X_n|\mathcal{G})$的一个版本。根据(i)和第5.4.3节，(Y_n)是非负的，并且是绝对增加的。让$Y=\limsup_n Y_n$提供

a $\mathcal{G}$-可测量的随机变量，使得实数序列$(Y_n(\omega))_n$趋向于$Y(\omega)$，几乎对所有ω.然后，推论4.9表明，$(\int_G Y_n\,dP)$增加到$\int_G Y\,dP$。但是我们有$\int_G Y_n\,dP=\int_G X_n\,dP$，对于每一个$n$，并且$(X_n)$以点方式增加到$X$。根据单调收敛定理，可以得出$(\int_G X_n\,dP)_{n\ge1}$增加到$\int_G X\,dP$，所以$\int_G X\,dP=\int_G Y\,dP$。这表明
Y是$E(X|\mathcal{G})$的一个版本，因此证明了我们的主张。

(iii) 我们可以把注意力限制在$X\ge0$上，因为一般情况下，通过线性关系可以得出结论。现在首先考虑指标的情况：如果$Y=1_E$，对于一些$E\in\mathcal{G}$，我们有，对于所有$G\in\mathcal{G}$、

$$\int_G 1_E\,E(X|\mathcal{G})\,dP=\int_{E\cap G}E(X|\mathcal{G})\,dP=\int_{E\cap G}X\,dP=\int_G 1_E X\,dP$$

因此，$1_E\,E(X/\mathcal{G})$满足定义方程，因此是乘积XY的条件期望的一个版本。因此，当$Y=1_E$，$E\in\mathcal{G}$时，$E(XY|\mathcal{G})=Y\,E(X/\mathcal{G})$已经得到验证。根据线性性质，这延伸到了简单函数，对于任意的$Y\ge0$，我们现在使用(ii)和简单函数的序列(Y_n)增加到Y，从而推断，对于非负的X，$E(XY_n|\mathcal{G})=Y_n\,E(X|\mathcal{G})$一方面增加到$E(XY|\mathcal{G})$，另一方面增加到$Y\,E(X|\mathcal{G})$。因此，如果$X$和$Y$都是非负的，我们就验证了（iii）。

线性允许我们将其扩展到一般的$Y = Y^+ - Y^-$。

$\int$ (iv) 对于$G \in \mathcal{G}$，我们有$\int_G E(X/\mathcal{G})\, dP = \int_G X\, dP$，$\int_H E(X/\mathcal{H})\, dP =$

$_H$对于$H \in \mathcal{H} \subset \mathcal{G}$，$X\, dP$，因此对于$H \in \mathcal{H}$，我们得到 $_H E(X|\mathcal{G})\, dP =$

$\int_H E(X|G)dP$。因此$E(X|G)$满足定义$E(X|H)$相对于H的条件期望，所以$E[E(X|G)|H]=E(X|H)$。

(v) 凸函数可以写成一连串仿生函数的最高值，即有一连串(a_n), (b_n)的实数，使$j(x)=\sup_n(a_n x+b_n)$，对于每一个$x\in R$。固定n，那么由于$j(X(\omega))\geq a_n X(\omega)+b_n$ for
所有的ω，正性和线性特性保证了

$$E(j(X)|G)(\omega)\geq E([a_n X+b_n]|G)(\omega)=a_n E(X|G)(\omega)+b_n$$

对于所有$\omega\in\omega/A_n$，其中$P(A_n)=0$。由于$A=\bigcup_n A_n$也是空的，因此，对于所有n≥1，$E(\varphi(X)|G)(\omega)\geq a_n E(X|G)(\omega)+b_n$ a.s。因此，当我们取右边的最高值时，不等式也成立，所以($E(j(X)|G)(\omega)\geq\phi[(E(X/G)(\omega)]$a.s. 这证明了（v）。 □

(v)的一个直接后果是，$E(X/G)$的L^p-norm被$p\geq1$的X的norm所约束，因为函数$\phi(x)=|x|^p$是凸的：我们得到

$$|E(X/G)|^p=\phi(E(X/G))\leq E(\phi(X)|G)=E(|X|^p|G)\ \text{a.s.}$$

以致

于 $\|E(X/G)\|_p^p=E(|E(X/G)|^p)\leq E(E(|X|^p|G))=E(|X|^p)=\|X\|_p^p$、

其中倒数第二步对$|X|^p$适用属性(1)。以p^{th}根为例，有
$\|E(X|G)\|_p\leq\|X\|_p$。

练习7.17

设$\Omega=[0,1]$，具有Lebesgue度量，设$X(\Omega)=\Omega$。求$E(X|G)$，如果
(a) $G=\{[0,\frac{1}{2}],(\frac{1}{2},1],[0,1],\varnothing\}$, (b) G是由集合族产生的
$\{B\subset[0,\frac{1}{2}]$：Borel$\}$.

7.4.2 马廷格

假设我们希望通过随机变量的序列（X_n）来模拟某些物理现象的行为。$X^{th}_n(\omega)$的值可能是一枚"公平"的硬币抛掷1000次的结果，"正面"被记录下来。

那么，$Y(\omega)=\sum_1^{1000}X_n(\omega)$将记录的次数是
硬币落地为"头"。通常情况下，我们会执行这个随机
在冒险对这枚硬币的"正面"概率做出陈述之前，我们要做大量的实验。我们可以平均我们的结果，即寻求计算E（Y）。但我们也可能对猜测"Y"的

价值感兴趣。

$X_n(\omega)$可能是在进行了$k < n$次抛掷之后，即对于一个固定的$\omega \in \omega$、

知道$(X_i(\omega))_{i\le k}$的值是否有助于预测$n>k$的$X_n(\omega)$的值？在一个"理想化"的抛硬币实验中，我们假定它不会，也就是说，假定连续的抛硬币是独立的--a

这个事实常常使概率论的初学者感到困惑。

有许多情况下，(X_n)将代表被建模过程的过去行为可以合理地被认为是影响其未来行为的结果，例如，如果X_n记录*第n*天是否下雨。我们寻求一种数学描述，说明我们对(X_n)过去行为的知识可以被编码的方式。一个自然的想法是使用σ-场

$F_k=\sigma\{X_i:0\le i\le k\}$由序列$(X_n)_{n\ge0}$产生，代表

从知道我们实验的前*k个*结果中获得的知识。我们把$(X_n)_{n\ge0}$称为（离散的）*随机过程*，以强调我们现在的重点是结果序列的"动态"，因为它正在展开。我们包括一个第0

阶段，以方便记述，所以在实验开始之前有一个"起点"，然后F_0，代表我们在观察到任何结果之前的知识。

因此，我们在"时间"*k*（即在记录了*k个*结果之后）可获得的关于"世界状态"ω的信息是由值$(X_i(\omega))_{0\le i\le k}$，这被封装在知道$F_k$的哪些集合中。点ω。但我们可以假设一连串的σ场$(F_n)_{n\ge0}$，非常普遍、

而不参考任何随机变量的序列。同样，我们对任何特定的ω*的*知识在$k\ge1$阶段通过知道F_k中的哪些集合包含ω来表示。二项式股票价格模型提供了一个简单的例子

第2.6.3节（见练习2.13）。在这个例子的指导下，我们把它变成一个一般的

定义7.11

给定一个概率空间(Ω, F, P)，一个（离散的）*过滤*是F的子σ场$(F_n)_{n\ge0}$的递增序列；也就是说。

$$F_0\subset F_1\subset F_2\subset\ldots\subset F_n\subset\ldots\subset F.$$

我们写$F=(F_n)_{n\ge0}$。我们说随机变量的序列$(X_n)_{n\ge0}$适应滤波F，如果X_n对每一个$n\ge0$都是F_n-可测量的。元组$(\Omega, F, (F_n)_{n\ge0}, P)$被称为*滤波概率空间*。

在我们的应用中，我们通常会假设$F_0=\{\emptyset,\Omega\}$，所以我们从"无信息"

开始，而且很多时候我们会假设整个序列产生的 "最终 "σ场，即$_\infty$ ＝σ(

$_{n\geq 0n}$)，是所有(所以
在实验结束时，"我们知道所有要知道的事情"）。很明显
(X_n)适应其*自然*滤波(F)$_{nn}$ ， 其中$F_n = \sigma(X_i : 0 \leq i \leq n\}$

对于每一个n，它适应于每一个包含这个的过滤。但同样地，如果$F_n = \sigma(X_i : 0 \le i \le n\}$ 对于某些过程$(X)_{nn}$，可能对于其他过程$(Y)_{nn}$，每个Y_n是F_n-可测量的，即$(Y)_{nn}$是适应$(F)_{nn}$。回顾一下，根据命题3.12，这意味着对于每一个$n \ge 1$，都有一个Borel-可测量的函数f_n： $R^{n+1} \to R$，这样$Y_n = f(X_0, X_1, X_2, \ldots, X_n)$。

我们来看看本节介绍的主要概念：

定义7.12

让$(\Omega, F, (F_n)_{n\ge 0}, P)$是一个过滤的概率空间。在$(\Omega, F, P)$上的随机变量序列$(X)_{nn\ge 0}$是相对于所提供的滤波$F = (F_n)_{n\ge 0}$的一个*马太效应*:

(i) $(X)_{nn}$是适应F的；

(ii) 每个X_n都在$L^1(P)$中、

(iii) 对于每个$n \ge 0$，$E(X \,|F_{n+1n}) = X_n$。

我们注意到这个定义的两个直接后果，它们被反复使用：

1) 如果$m > n \ge 0$，那么$E(X_m|F_n) = X_n$。这由塔的属性得出的条件期望值，因为（A.S.）

$$e(x \,|f_{mn}) = e(e(x \,|f_{mm-1})|f_n) = e(x \,|f_{m-1n}) = \ldots = e(x \,|f_{n+1n}) = x_n .$$

2) 任何马丁格尔（X_n）都有恒定的期望值：

$$e(x_n) = e(e(x \,|f_{n0})) = e(x)_0$$

对每一个$n \ge 0$都成立，由1）和（i）的命题7.28决定。

马丁格尔法代表了赌博中的 "公平游戏"：例如，以抛硬币的结果为赌注，我们在 "游戏n "中的赢利（n^{th}抛硬币的结果）将是 $\Delta X_n = X_n - X_{n-1}$，这是与
我们在那场比赛之前和之后都有。(我们假设$X_0 = 0$。) 如果游戏
是公平的，我们会在时间（n - 1）上预测，在知道n^{th}的结果之前，E（$\Delta X \,|F_{nn-1}$） =0，其中$F_k = \sigma\{X_i : i \le k\}$是过程（$X_n$） 的自然过滤的σ域。这是因为我们的知识在
时间（n 1）被封装在$_{n-1}$，在F一个公平的游戏中，我们期望我们在任何阶段的增量赢利平均为0。因此，在这种情况下，(X_n)形成了一个马丁格尔。

同样，在一个对赌徒有利的游戏中，我们应该期望$E(\Delta X_{nn-1}) 0$，即
|F ≥ |F ≥

$E(X_{n}|\mathcal{F}_{n-1}) \geq X_{n-1}$ a.s.。我们把满足这个不等式的序列（以及定义7.12的(i)、(ii)）称为*亚鞅*，而a

对赌徒不利的游戏（因此对赌场有利！）同样被*超鞅所*代表，对于每一个n，它有$E(X_n|F_{n-1}) \le X_{n-1}$ a.s.。注意，对于一个亚鞅，(X_n)的期望值增加了

而对超鞅来说，它们会减少。最后，请注意，如果我们用$X_n - X_0$（只要$X_0 \in L^1(F_0)$，以保留整数性和适应性）取代X_n，这些过程的特性不会改变，因此我们可以在不丧失一般性的情况下，以$X_0 = 0$开始工作。

例7.6

最明显的，但在某些方面相当普遍的马丁格尔的例子包括一连串的条件期望：给定一个随机变量$X \in L^1(F)$和F的子σ场的过滤$(F_n)_{n\ge0}$，让$X_n = E(X|F_n)$

那么$E(X_{n+1}|F_n)=E(E(X|F_{n+1})|F_n)=E(X|F_n)=X_n$，利用
塔的特性。我们可以通过将每个X_n解释为给予

我们在时间n可用的信息，即包含在σ场F_n，关于随机变量X的信息。（记住，条件期望值是X的 "最佳猜测"，关于均方误差，当我们在L^2中工作时）。
对于一个有限滤波$\{F_n : 0 \le n \le N\}$，$F_N = F$，很明显的是
$E(X|F_N)=X$。对于一个无限的序列，我们可能同样希望 "在
限"，我们将拥有关于X的 "全部 "信息，这表明我们应该

能够在某种意义上将X检索为(X_n)的极限。极限存在的条件需要仔细研究--

详见例如[12]，[8]。

马丁格尔的第二个标准例子是：

例7.7

假设$(Z_n)_{n\ge1}$是一串独立的随机变量，其零
意思是说。设$X_0 = 0$，$F_0 = \{\varnothing, \Omega\}$，设$X_n = \sum_{k=1}^{n} Z_k$，并定义$F_n = \sigma\{Z_k : k\le n\}$，对于每个$n\ge1$。那么$(X_n)_{n\ge0}$相对于过滤$(F_n)$来说是一个马特格尔。要看到这一点，请回忆一下，对于每一个n，Z_n是独立于F_{n-1}的，所以$E(Z_n|F_{n-1}) = E(Z_n) = 0$。因此$E(X_n|F_{n-1}) = E(X_{n-1}|F_{n-1}) + E(Z_n) =$
X_{n-1}，因为X_{n-1}是F_{n-1}-可测量的。(你应该仔细检查哪些道具
我们在这里使用的条件期望值的特性！)

这个例子的 "乘法 "版本如下：

练习7.18

让$Z_n \geq 0$是一串独立的随机变量，$\mathrm{E}(Z_n) =$

$\mu = 1$ 让$F_n = \sigma\{Z_k : k \le n\}$并表明，$X_0 = 1, X_n = Z Z_{12} \ldots Z_n$ $(n \ge 1)$为(F_n)定义了一个鞅，条件是所有的乘积都是可整定的随机变量，例如，如果所有的$Z_n \in L^\infty(\Omega, F, P)$，这一点就成立。

练习7.19

让$(Z)_{nn\ge1}$ 是一串独立的随机变量，其平均值$\mu = E(Z_n) \ne 0$，适用于所有n。表明它们的部分和的序列$X_n = Z_1 + Z_2 + \cdots + Z_n$ 不是过滤$(F)_{nn}$ 的马蒂格，其中$F_n = \sigma\{Z_k : k \le n\}$。我们如何通过改变$X_n$ 来 "补偿 "这个问题？

让$X = (X)_{nn\ge0}$ 是过滤$F = ( )_{nn\ge0}$ （用我们上面的惯例）的马太效应；简而言之，我们只是指马太效应（X，F）。函数$\varphi(x)=x^2$ 是凸的，因此根据詹森不等式（定理7.29）：
我们有$E(X^2_{n+1} | F_n) \ge (E(X | F_{n+1n}))^2 = X^2$ ，所以X^2 是一个亚马特。我们研究是否有可能像练习7.19中那样进行 "补偿"，以使所产生的过程再次成为一个马太效应。请注意，X^2 的期望值是增加的，所以我们需要从X^2 中减去一个增加的过程来实现这一点。

事实上，这种 "补偿器 "过程的构造是相当普遍的。让$Y = (Y_n)$是任何适应的过程，每个$Y_n \in L^1$。对于任何过程Z，将其增量写成$\Delta Z_n = Z_n - Z_{n-1}$ ，对于所有n。回顾一下，在这个符号中，马太效应特性可以简明地表示为$E(\Delta Z_n | F_{n-1})=0$--我们将在下文中反复使用。

我们定义两个新的过程$A = (A_n)$和$M = (M_n)$，$A_0 = 0$，$M_0 = 0$，通过他们的连续增量

$$\Delta A_n = E(\Delta Y_n | F_{n-1})，\ \Delta M_n = \Delta Y_n - \Delta A_n ，\ n\ge1。$$

我们得到$E(\Delta M_n | F_{n-1}) = E([\Delta Y_n - E(\Delta Y_n | F_{n-1})] | F_{n-1}) = E(\Delta Y_n | F_{n-1}) - E(\Delta Y_n | F_{n-1}) = 0$，因为$E(\Delta Y_n | F_{n-1})$ 是可以测量的。因此，M是一个马拉松。此外，当且仅当$0 \le \Delta A_n =$时，过程A是增加的。
$E(\Delta Y_n | F_{n-1}) = E(Y_n | F_{n-1}) - Y_{n-1}$，当且仅当$Y$是一个次martingale。请注意，$A_n = \sum_{k=1}^n \Delta A_k = \sum_{k=1}^n [E(Y_k | F_{k-1}) - Y_{k-1}]$ 是F_{n-1}可测量的。因此，A的值$_n$，在时间$n-1$之前是 "已知 "的。一个具有这种性质的过程被称为*可预测的*，因为我们可以提前一步 "预测 "它的未来值。马太效应的一个基本属性是它们*不是*预先可支配的：事实上，如果X是一个可预测的马太效应，那么我们有

$$X_{n-1} = E(X_n | F_{n-1}) = X_n \text{ a.s. for every } n$$

其中第一个等式是马太效应的定义，而第二个等式是由于X_n 是$_{n-1}$ -可测量的。因此，一^{F}个可预测的马太效应是一个常数，如果它开始于0，就会停留在那里。这一事实给出了一个可预测马太效应的分解。

适应的过程Y变成了一个鞅和一个可预测过程的总和，这是一个有用的唯一性属性：首先，由于$M_0 = 0 = A_0$，我们有$Y_n = Y_0 + M_n + A_n$，对于上面定义的过程M, A。如果还有$Y_n = Y_0 + M_n' + A'_n$，其中
M_n' 是一个马太效应，并且A'_n是可预测的，那么

$$M_n - M_n' = A'_n - A_n \text{ a.s.}$$

是一个可预测的鞅，在时间0时为0。因此两边对每一个n都是0，所以分解是唯一的。

我们把这称为适应过程的*Doob分解*。它在应用于从马太效应X产生的子马太效应$Y = X^2$ 时具有特殊的重要性。在这种情况下，正如我们在上面看到的，可预测过程A是增加的，因此，对于每一个n，$A_n \leq A_{n+1}$ a.s.，而Doob分解为：

$$x^2 = x_0^2 + m + a$$

特别是，如果$X_0 = 0$（我们可以不失时机地假设），我们把$X^2 = M + A$写成一个鞅的M和一个可预测的增加过程A的总和。这一点的意义在一个非常有用的属性中显示出来。

马太效应，这是证明拉登-尼科迪姆定理的一个关键部分（见定理7.3的步骤1（iv），其中马太效应的联系被很好地隐藏了！）：对于任何马太效应X，我们可以写出，$(\Delta X_n)^2 = (X_n - X_{n-1})^2$：

$$\begin{aligned} e((\Delta x_n)^2 | f_{n-1}) &= e([x_n^2 - 2x_n x_{n-1} + x_{n-1}^2] | f_{n-1}) \\ &= e(x_n^2 | f_{n-1}) - 2x_{n-1} e(\Delta x_n | f_{n-1}) - x_{n-1}^2 \\ &= E([X_n^2 - X_{n-1}^2] | F_{n-1}) \end{aligned}$$

。

因此，给定马汀格尔X，$X_0 = 0$，分解$X^2 = M + A$
产生，因为M也是一个马太效应：

$$\begin{aligned} 0 = e(\Delta m_n | f_{n-1}) &= e((\Delta x_n)^2 - \Delta a_n) | f_{n-1}) \\ &= e([x_n^2 - x_{n-1}^2] | f_{n-1}) - e(\Delta a_n | f_{n-1})。\end{aligned}$$

换句话说，由于A是可预测的、

$$e((\Delta x_n)^2 | f_{n-1}) = e(\Delta a_n | f_{n-1}) = \Delta a_n \tag{7.5}$$

它将过程A表现为原始马太效应X的条件 "二次变化 "过程：$e((\Delta x_n)^2) = e(\Delta a_n)$。

例7.8

还请注意，$E(X^2)_{\overline{n}} E(M_n) + E(A_n) = E(A_n)$ (为什么？)，因此，当且仅当鞅的X作为$L^2(\omega, F, P)$中的一个序列被约束时，两边对所有n都被约束。由于(A_n)是增加的，所以存在a.s.极限$A_\infty(\omega) = \lim_{n\to\infty} A_n(\omega)$，并且积分的有界性确保在这种情况下$E(A_\infty) <$
∞.

练习7.20

假设$(Z)_{nn\geq 1}$是一个伯努利随机变量序列，每个Z_n取值为1和-1，每个概率为$\frac{1}{}$。设$X_0 = 0$，$X_n = Z_1 + Z_2 +\cdots + Z_n$，$_2$设$(F)_{nn}$是由$(Z_n)$产生的自然滤波。验证$(X^2)$是一个亚鞅，并在其Doob分解中找到增加过程$(A_n)$。什么是 "意外 "的属性？

(A_n)你能在这个例子中发现吗？

在离散设置中，我们现在有了构建 "随机积分 "的工具，并表明它们保留了马丁格尔属性。事实上，正如我们对Lebesgue-Stieltjes度量所看到的那样，对于离散分布，"积分 "只是分布函数增量的适当线性组合。如果我们想用一个马太效应的X作为积分器，那么我们需要

来处理增量的线性组合 $\Delta X_n = X_n X_{n-1}$。因为 -
我们现在要处理的是随机过程（即n与n之间的函数）。
ω）而不是实数函数，可测性条件将有助于确定什么构成 "适当的 "线性组合。因此，如果对于$\omega\in\omega$，我们设定$I_0(\omega) = 0$，并形成和

$$I_n(\omega) = \sum_{k=1}^{n} c_k(\omega)(\Delta X_k)(\omega) = \sum_{k=1}^{n} c_k(\omega)(X_k(\omega) - X_{k-1}(\omega)) \text{ for } n \geq 1、$$

我们寻找过程$(c)_{nn}$的可测性属性，以确保新过程$(I)_{nn}$具有有用的属性。当$(c)_{nn}$是一个有界的可预测过程，并且X是一个给定过滤$(F)_{nn}$的马汀格尔时，我们对此进行研究。
有些文献称这个过程$(I)_{nn}$为*马丁格尔变换*--我们更喜欢这个术语
离散的随机积分。我们计算I的条件期望值$_n$：

$$E(I\,|F_{nn-1}) = E([I_{n-1} + c\,\Delta X_{nn}]|F_{n-1}) = I_{n-1} + c_n E(\Delta X\,|F_{nn-1}) = I_{n-1}$$
、

因为c_n是F_{n-1}可测量的，$E(\Delta X\,|F_{nn-1}) = E(X\,|F_{nn-1}) - X_{n-1} = 0$。因

此，当过程$c = (c\ )_{nn}$ 对马廷风$X = (X_n\ )$进行整合时，是可预测的，马廷风属性在以下情况下是保留的。

离散的随机积分：$I = (I\ )_{nn}$ 也是一个相对于以下因素的马太效应：

过滤（F）$_{nn}$。我们将把这个随机积分写成c-X，也就是说，对于所有的n 0⩾，$I_n = (c\ X)_n$。这个结果的重要性足以让我们把它作为一个定理来记录：

定理7.30

让$(\Omega, , F(_nF)_{n\geq 0}, P)$是一个过滤的概率空间。如果$X$是一个鞅，$c$是一个有界的可预测过程，那么离散的随机积分$c\ X$又是一个鞅。 -

请注意，我们使用有界性假设是为了确保$c\ \Delta X_{kk}$是可整数的，这样它的条件期望值才有意义。对于2-马太效应（也就是我们在大多数应用中得到的），我们可以放宽这个条件

条件，只要求对每个n的$c\ _n^2$ ( )。$_{n-}$∈L F

虽然保留马丁格尔属性可能会让数学界感到高兴，但这并不意味着它就会消失。

对赌徒来说，这是个令人沮丧的消息！我们可以把过程c解释为代表赌徒在每场比赛中的赌注大小，因此c_n是他在第n场比赛中的赌注金额。请注意，c_n可能是0，这意味着赌徒 "坐等 "第n场比赛，不下注。同样合理的是，赌注的大小取决于之前游戏的结果，因此

c_n是$_F{}_1$-可测量的，因此c是可预测的。

那么$c\ X$是马丁格尔的结论意味着 "聪明 "的赌博

如果游戏是公平的，那么策略就没有用了。无论赌徒采用什么策略，它都是公平的！当然，如果一开始就对赌徒不利，那也是不公平的！当然，如果一开始就不利于

赌徒，所以X是一个超鞅$(X_{n-1} \geq E(X\ |F_{nn-1}))$，上述计算表明，只要每个$n$的$c_n$ 0，那么$E(I_{\ nn-1})\ I_{n-1}$ ⩾ 因此，无论赌徒用什么样非负的赌注，游戏仍然是不利的。

地方（毕竟负的赌注似乎不太可能被接受...）。当然，你会立即验证，当c是一个非负的过程时，一个亚马逊X产生一个亚马逊$c\ X$。遗憾的是，这种有利的游戏在实践中很难找到。

将$(I\ )_{nn}$的定义与子马太尔X^2的Doob分解结合起来，我们得到了一个特征，说明了为什么马太尔会成为有用的 "积分器"。当$c = (c_n)$是可预测的，$X = (X_n)$是马太效应时，我们计算$(c - X)_n$的平方的期望值：

n n

$$E((c - X)^2) = E([\sum_{k=1} c\, \Delta X\,]_{kk}{}^2) = E(\sum_{j,k=1} c c_{jk}\, \Delta X\, \Delta X_{jk})。$$

分别考虑双重和中的条款：当$j<k$时，我们有

$$E(c c_{jk}\, \Delta X\, \Delta X_{jk}) = E(c c\, \Delta X_{jkjk}\ \ \Delta X|F_{k-1}) = E(c c_{jk}\, \Delta X_j\, E(\Delta X_k\, |F_{k-1}) =$$
0

因为前三个因素都是$_{k-1}$ - 可测量的，而$E(\Delta X_{k}|F_{k-1})=0$，因为$X$是一个马太效应。在$j$、$k$互换的情况下，这也表明，当$k<j$时，这些项都是0。

其余条款的形式为

$$E(c_k^2 (\Delta X)_k^2) = E(c_k^2 E((\Delta X)_k^2 | F_{k-1})) = E(c_k^2 \Delta A_k)。$$

因此，根据线性关系，我们有相对于马廷格的随机积分的基本特性（也称为*伊藤等值*）：

$$E([\sum_{k=1}^{n} c_k \Delta X_k]^2) = E(\sum_{k=1}^{n} c_k^2 \Delta A_k)。$$

备注7.3

右边期望符号内的和是递增过程的 "Stieltjes和"，因此，现在至少有理由相信，这一特性允许我们在连续时间环境下定义马廷格积分，使用简单过程的近似，就像本书中对实数函数所做的那样。在相对于布朗运动等过程的随机积分的定义中，Ito等值是至关重要的：在定义Lebesgue-Stieltjes积分时，我们的积分器是有界变化的。通常情况下，布朗运动的路径（我们在本书中不会定义这一过程
- 见（例如）[3]的基本属性）不是有界变化的，但是伊藤等值线表明，它们的二次变化可以在上述框架的（更微妙的）连续时间版本中得到处理，这使得人们可以用布朗运动（和更普遍的马汀格）作为 "积分器"，来定义一类广泛的函数的积分。

我们最后谈谈在一个随机时间*停止*马丁格尔的想法。

定义7.13

一个随机变量$\tau : \Omega \to \{0, 1, 2, ..., n, ..\} \cup \{\infty\}$是相对于过滤（$F_n$）的*停止时间*，如果对于每一个$n \geq 1$，事件$\{\tau = n\}$属于$F_n$。

请注意，我们包括价值$\tau(\omega) = \infty$，所以我们需要$\{\tau = \infty\} \in F_\infty = \sigma(\cup_n F_n)$。即 "极限*σ场*"。停顿时间也被称为*随机次*，以强调 "时间 "τ是一个随机变量。

对于一个停止时间τ，事件$\{\tau \le n\} = \bigcup_{k=0}^{n} \{\tau = k\}$是在$F_n$，因为对于

每个$k \le n$ $\{\tau = k\} \in F_k$，并且σ场随n增加。另一方面、

鉴于对每个n来说，事件$\{\tau\le n\}\in F_n$，那么

$$\{\tau = n\} = \{\tau \le n\}\{\tau \le n - 1\} \in F_n。$$

因此，我们同样可以把条件τ n $\{\le\}\in F$ $_n$为所有n*的条件*作为停止时间的定义。

例7.9

赌徒可以在随机的游戏次数后决定停止游戏，d取决于他的赢利X是否达到预先确定的水平L（或者他的资金已经用完！）。时间$\tau = \min\{n : X_n \ge L\}$是第一个时间
在这个过程中，X击中了区间$[L,\infty]$；更确切地说，对于$\omega\in\omega$、
$\tau(\omega) = n$，如果$X_n(\omega) \ge L$，而$X_k(\omega) < L$，对于所有的$k < n$。由于$\{\tau = n\}$因此是
由X和X的值决定$_k$，对于$k<n$，现在很清楚了
即τ是一个停顿时间。

例7.10

同样地，如果一只股票S的价值低于其当前（时间0）价格的75%，我们可以决定卖出我们的股票。因此，我们在随机时间$\tau=\min\{n : S_n <\frac{3}{4} S_0\}$，这也是一个止损时间。这是一个 "止损策略 "的例子，在熊市中很常见。

相当普遍地，一个适应过程X对一个Borel集$A\subset R$的*首次命中时间*τ_A是通过设置$\tau_A = min\{n \ge 0： X_n \in A\}$。对于任何$n\ge0$，我们有$\{\tau_A \le n\} = \cup_{k\le n} \{X_k \in A\} \in F_n$。为了照顾$X$从未击中$A$*的*可能性，我们使用惯例$\min \varnothing = \infty$，所以$\{\tau_A = \infty\} = \Omega\setminus(\cup_{n\ge0} \{\tau_A \le n\})$代表这个事件。但是它的补数在$F_\infty = \sigma(\cup F_{n\ge0n})$，因此$\tau_A$ = 也是。我们已经证明，τ_A是一个停止时间。

回到赌博的主题，我们看到停止只是赌博策略的一种特殊形式，因此毫不奇怪，在停止的情况下，马太效应的特性被保留下来（对于超马太效应和亚马太效应也有类似的结论）。对于任何适应的过程X和停止时间τ，我们通过设置$X^\tau(\omega) = X_{n\wedge\tau(\omega)}(\omega)$，在每一个*停止的过程*中定义$X^\tau_n$。
$\omega\in\omega$。(回顾一下，对于实数x，y，我们写$x \wedge y = \min\{x, y\}$。)

停止的过程X^τ 再次适应于过滤（F ）$_{nn}$，因为 $\{X_{\tau\wedge n} \in A\}$意味着，要么$\tau>n$且$X_n$ �A，要么$\tau=k$，对于某些$k\leq n$且X_k �A。现在，事件$\{X_n \in A\} \cap \{\tau>n\}$�$F_n$ �������� *于每个k，事件*$\{\tau = k\}$�F 。$\cap \{X_k \in A\} \in F_k$。因此，对于所有$k\leq n$*的*事件都属于$F_n$。因此，$\{X_{\tau\wedge n} \in A\}$也是如此，这证明$X^\tau$ 是适应的。

定理7.31

让$(\Omega, F, (F_n), P)$是一个过滤的概率空间，让X是一个马太效应，$X_0 = 0$。如果τ是一个停止时间，停止过程X^τ又是一个马太效应。

证明

我们使用离散随机积分（"赌博策略"）下的马丁格尔属性的保存。让$c_n = 1_{\{\tau \geq n\}}$，每个$n \geq 1$。这就定义了一个有界的可预测过程$c = (c_n)$，因为它只取值0、1和

$\{c_n = 0\} = \{\tau \leq n - 1\} \in F_{n-1}$，因此，$\{c_n = 1\} = \Omega \setminus \{c_n = 0\} \in F_{n-1}$。因此，根据定理7.30，过程$c \cdot X$又是一个马太效应。但根据结构，$(c \cdot X)_0 = 0 = X_0 = X^\tau$，而对于任何$n \geq 1$

$$(c \cdot X)_n = c_1 (X_1 - X_0) + c_2 (X_2 - X_1) + \ldots + c_n (X_n - X_{n-1}) = X_{\tau \wedge n}。$$

□

由于在证明中定义的c_n 0，因此，超鞅和亚鞅的特性在停顿下也被保留下来。此外，对于一个鞅，我们还有期望值被保留，即（在一般情况下）$e(x_{\tau\wedge n}) = e(x_0)$。同样地，对于停止的亚马逊的期望值也会增加、而对于停止的超音速来说，则会减少。

然而，这些都不能保证由$X_\tau(\omega) = X_{\tau(\omega)}(\omega)$定义的随机变量$X_\tau$具有有限的期望值--为了得到一个将其期望值与$X_0$的期望值联系起来的结果，我们通常需要满足更严格的条件。对于有界的停止时间（其中有一个统一的上界

N，$\tau(\omega) \leq N$，对于所有$\omega \in \omega$)，事情很简单：如果X是一个鞅，$X_{\tau\wedge n}$对于所有n是可整定的，并且根据上述定理$E(X_{\tau\wedge n}) = E(X_0)$。现在应用这个方法，$n = N$，所以$X_{\tau\wedge n} = X_{\tau\wedge N} = X_\tau$。因此我们有$E(X_\tau) = E(X_0)$，只要$\tau$是一个有界的停止时间。我们将不再深入研究，但

请读者参考主要致力于马丁格尔理论的文本，例如[12]、[8]、[3]。在许多实际应用中，例如在金融中的离散美式期权的分析中，有约束的停止时间就足够了。

7.4.3 在数学上的应用 金融

金融理论的一个主要任务和挑战是为资产和证券定价，建立与市场实践一

致的模型。这种一致性意味着，任何偏离理论价格的行为都应受到

市场。具体来说，如果一个市场参与者报出的价格与模型提供的价格不同，她应该必然会亏损。

问题是未知的未来，它必须以某种方式反映在价格中，因为市场参与者通过商定价格来表达他们对未来的信念。在数学上，一个理想的情况是，价格过程$X(t)$是一个
martingale。那么我们就会有一个明显的定价公式$X(0)=\mathbb{E}(X(T))$
此外，我们还将有一系列的中级计算公式
通过基于收集到的信息的条件性预期的方式来确定价格。

然而，价格遵循马太效应的情况与市场事实不符，即资金可以无风险投资，这为投资于风险证券的投资者创造了一个预期基准。因此，我们修改了我们的任务，坚持认为贴现值$Y(t)=X(t)\exp\{-rt\}$形成一个马太效应。这个修改是通过一个确定的常数来实现的，所以它不会给资产估值带来问题。

一个特殊的目标是对衍生证券进行定价，我们给定的是形式为$f(S(T))$的终端价值(债权)，其中f是已知的，并且假定相关资产$S(t)$的价值的概率能力分布是通过采取一些数学模型而知道的。上述马丁格尔思想将通过构建一个过程$X(t)$来解决定价问题，即$X(T)=f(S(T))$(我们称X为*复制*的主张)。

我们可以总结一下这些任务：建立一个基础证券$S(t)$的价格模型，以便

1. 有一个复制$X(t)$，使$X(T)=f(S(T))$、
2. 过程$Y(t)=X(t)\exp\{-rt\}$是一个马太效应，所以$Y(0)$是f所描述的证券的价格、
3. 任何偏离所得价格的行为都会导致损失。

步骤1和2是数学性质的，但步骤3与实际市场活动有关。

我们为单步二项式模型执行任务。

第一步。回顾一下，这个模型中的价格是$S(0)$，$S(1)=S(0)\eta$，其中$\eta=U$或$\eta=D$有一定的概率。让$f(x)=(x-K)^+$。我们可以很容易地找到$X(0)$，从而使$X(1)=(S(1)-K)^+$，通过建立一个由n个股票组成的投资组合S

和m个单位的银行账户（见第6.5.5节），解出系统后

$$nS(0)U + mR = (S(0)U - K)^{+}、$$
$$nS(0)D + mR = (S(0)D - K)^{+}。$$

注意，$X(1)$是通过时间t=0的数据和模型参数得到的。

第2步。写下$R = \exp\{r\}$。我们需要的马氏条件是微不足道的：$x(0)r$=e(x(1))。这里的任务是找到概率测量（X是

定义，我们对R没有影响，所以这是我们唯一可以调整的参数）。回顾一下，我们假设$D \le R \le U$。我们解决

$$X(0)R = pX(0)U + (1-p)X(0)D$$

从而得出

$$p = \frac{R - D}{U - D}$$

然后我们就完成了。因此，看涨期权的理论价格是

$$C = R^{-1} p(S(0)U - K) = \exp\{-r\}E(S(1) - K)^{+}。$$

第三步。为了说明在我们的模型中，价格是正确的，假设有人愿意以$C' > C$的价格买入一个看涨期权。我们立即卖出它，将收益投资于步骤1中的投资组合。在行使时，我们的投资组合与看涨期权的回报相匹配，并赚取了差额$C' C$。因此，我们继续购买看涨期权，直到卖方意识到错误并提高价格。同样，如果有人以$C' < C$的价格卖出看涨期权，我们通过形成(n, m)投资组合获得现金，买入看涨期权（由于复制的结果，在到期日结算我们的投资组合负债），并

我们有利润，直到报价的看涨价格达到理论价格C。

上述分析总结了一般理论的关键特征。将这个微不足道的模型直接扩展到n个步骤，就可以得到所谓的CRR（Cox-Ross-Rubinstein）模型，对于大的n来说，这个模型对于重新定价是相当充分的。我们评估了二项式模型中的期望值，在命题4.37中确定了欧洲看涨期权的CRR价格。

对于连续时间来说，这个任务变得相当困难。在Black-Scholes模型中，$S(t) = S(0) \exp\{(r - \frac{1}{2}\sigma^2)t + \sigma w(t)\}$，其中$w(t)$是一个随机过程（称为*维纳过程*或*布朗运动*），$w(0) = 0$和独立增量，这样$w(t)-w(s)$是高斯的，均值为零，方差为$t-s$，$s < t$

练习7.21

证明$\exp\{-rt\}S(t)$是一个马太效应。

复制过程$X(t)$的存在并不容易，但可以证明，还有一个事实是，过程$Y(t)$是一个马太效应。这导致了相同的一般定价公式：$\exp\{-rT\}E(f(S(T)))$。

利用$w(T)$的密度，这
对于特定的衍生证券，可以用明确的形式写出数字
(见第 4.7.5 节，在那里得出了看涨和看跌期权的价格公式）。

备注7.4

熟悉金融理论的读者会注意到，我们关注的是衍生证券的定价，这就导致我们要考虑贴现价格形成马丁格尔的模型。这个模型是一个数学上的创造，不一定与现实生活一致，它需要一个不同的概率空间和同一框架内的不同参数。现实世界和马廷格之间的联系是由拉登-尼科迪姆定理的一个特殊应用提供的，该定理将概率度量联系起来，但这是一个故事，我们不会在这里继续讨论，请读者参阅许多专门讨论这个问题的书籍（例如[5]）。

7.5 命题的证明

证明（命题7.1的证明）

假设（ε，δ）条件失败。 与提示中的A一样，我们有$\mu(A) \leq \mu(\bigcup_{i \geq n} F_i) \leq \sum_{i \geq n} \frac{1}{2^i} = \frac{1}{2^{n-1}}$ 对于每一个$n \geq 1$。因此，$\mu(A)=0$。 但是$\nu(F_n)$对每一个n都$\geq \varepsilon$，因此$\nu(E_n) \geq \varepsilon$，其中$E_n = \bigcup_{i \geq n} F_i$。 Se-quence $(E_n)_n$递减，因此，由于$\nu(F_1)$是有限的，定理2.13(i)给出：$\nu(A)=\nu(\lim_n E_n) = \lim_n \nu(E_n) \geq \varepsilon$。因此$\nu$不是绝对连续的。

反过来说，如果(ε, δ)条件成立，并且$\mu(F)=0$，那么对于任何$\delta>0$，$\mu(F)<\delta$，因此，对于每一个给定的$\varepsilon>0$，$\nu(F)<\varepsilon$，因此$\nu(F)=0$。 □

证明（命题7.4的证明）

我们按备注4.1的方法进行。以g为指标函数1_G，为$G \in \mathcal{G}$。然后我们有：$\int_\Omega g \, d\mu = \mu(G) = \int_G h_\mu \, dj$ 根据h_μ的构造。 接下来、让$g = \sum_{i=1}^n a_i 1_{G_i}$，对于集合$G_1, G_2, \ldots, G_n$，以及实数$a_1, a_2, \ldots, a_n$；积分的线性关系，得到

$$\int_\Omega g \, d\mu = \sum_{i=1}^n a_i \mu(G_i) = \sum_{i=1}^n a_i \left(\int_{G_i} h_\mu \, dj\right) = \sum_{i=1}^n a_i \left(\int_\Omega 1_{G_i} h_\mu \, dj\right) = \int_\Omega g h_\mu \, dj。$$

纪晓岚

最后，任何G-可测量的非负函数g都可以从下往上近似。由G-simple函数的增加序列$g_n = \sum_{i=1}^n a_i 1_{G_i}$，其积分$\int_\Omega g \, d\mu$是递增序列的极限$(\int_\Omega g_n h_\mu \, dj)_n$。 但由于$0 \leq h_\mu \leq 1$的结构，$(g_n h_\mu)$增加到$g h_\mu$点，因此该序列$(\int_\Omega g_n h_\mu \, dj)_n$也增加到$\int_\Omega g h_\mu \, dj$，所以极限是相等的。对于可整定的

$g = g^+ - g^-$ 将上述内容分别应用于g^+ , g^- 中，并使用线性。 □

证明（命题7.6的证明）

如果 $\bigcup_{n\geq 1} A_n = \Omega$，且$A_n$不是不相交的，则用$E_n$替换，其中 $E_1 = A_1$, $E_n = A_n \setminus (\bigcup_{i=1}^{n-1} E_i)$, $n > 1$。对B也可以这样做，因此我们可以把这两个序列作为不相交的。现在$\Omega = \bigcup_{m,n\geq 1} (A_n \cap B_m)$也是一个不相交的联合，并且$\nu, \mu$在每个$A_n \cap B_m$上都是有限的。将这些集合重新排序为一个单一的序列$(C_n)_{n\geq 1}$，并固定$n\geq 1$。将两个度量都限制在σ场$\mathcal{F}_n = \{F \cap C_n : F \in \mathcal{F}\}$得到它们是$(\Omega, \mathcal{F}_n)$上的有限度量、以便拉登-尼科迪姆定理适用，并提供一个非负的$\mathcal{F}_n$可测量函数h_n，使$\nu(E) = \int_E h_n \, d\mu$，对于每个$E \in \mathcal{F}_n$。但任何集合$F \in \mathcal{F}$的形式是$\bigcup_n F \cap C_n$，所以我们可以通过定义$h = h_n$，每$n\geq 1$。现在$\nu(F) = \sum_{n=1}^{\infty} \int_{F_n} h_n \, d\mu = \int_F h \, d\mu$。该

唯一性很清楚：如果g具有与h相同的属性，那么每个$F\in\mathcal{F}$，$\int_F (h-g)\, d\mu = 0$，因为所以根据定理4.15，$h-g = 0$ a.e.。

□

证明（命题7.7的证明）

(i) 这是微不足道的，因为$\varphi = \lambda + \nu$是σ-无限的，并且相对于μ是绝对连续的，而且我们有，对于$F\in\mathcal{F}$：

$$\int_F \frac{d\varphi}{d\mu} d\mu = \varphi(F) = (\lambda + \nu)(F) = \lambda(F) + \nu(F) = \int_F \left[\frac{d\lambda}{d\mu} + \frac{d\nu}{d\mu}\right] d\mu。$$

因此，左右两端的积分是a.s.(μ)相等的，所以结果就出来了。

(ii) 写下$\frac{d\lambda}{d\nu} = g$ 和$\frac{d\nu}{d\mu} = h$。这些都是非负的可测量的函数而我们需要证明，对于$F\in\mathcal{F}$

$$\lambda(F) = \int_F gh \, d\mu。$$

首先考虑当g被一个形式为$\varphi = \sum_{i=1}^{n} c_i 1_{E_i}$的简单函数取代时。然后我们得到：

$$\int_F \varphi \, d\nu = \sum_{i=1}^{n} c_i \nu(F \cap E_i) = \sum_{i=1}^{n} c_i \int_{F\cap E_i} h \, d\mu = \int_F \varphi h \, d\mu。$$

现在，让(φ_n)是一连串的简单函数，以点的方式增加到g。

通过单调收敛定理：

$$\lambda(F)=\int_F g\,\mathrm{d}\nu=\lim_n\int_F\varphi_n\,\mathrm{d}\nu=\lim_n\int_F\varphi_n h\,d\mu=\int_F gh\,\mathrm{d}\mu、$$

因为（φ_n h）增加到gh。这就证明了我们的主张。 □

证明（命题7.8的证明）

使用提示：$\lambda_1 + \lambda_2$ 集中于A_1 A_2，而μ集中于B_1 B_2。但A_1 A_2 与B_1 B_2 不相干，因此，U措施$\lambda_1 + \lambda_2$ 和μ是相互奇异的。这证明了(i)。对于(ii)，选择一个集合E，对于该集合，$\mu(E)$=0，而λ_2 集中在E上。让$F \subset E$，这样$\mu(F)$=0，因此$\lambda_1(F)$=0（因为λ_1 μ）。这表明λ_1 集中于E^c，因此λ_1 和 λ_2 是互为奇异的。最后，在$\lambda_1 = \lambda_2 = v$*的情况下*应用(ii)表明，$v \perp v$，这只能在$v = 0$时发生。 □

证明（命题7.11的证明）

固定$x \in$R。集合$A = \{F(y) : y < x\}$在上面被$F(x)$所约束，而$B = \{F(y) : x < y\}$在下面被$F(x)$所约束。因此，$\sup A = K_1 \le F(x)$, and $\inf B = K_2 \ge F(x)$ 都存在于R中，对于任何$\varepsilon > 0$，我们可以找到$y_1 < x$，使得$K_1 - \varepsilon < F(y_1)$和$y_2 > x$，使得$K_2 + \varepsilon > F(y_2)$。但由于$F$是递增的，这意味着在整个区间（$y_1, x$）内$K_1 - \varepsilon < F(y) < K_1$，在整个（$y, y_2$）内$K_2 < F(y) < K_2 + \varepsilon$。因此，两个单边极限$F(x-) = \lim_{y\uparrow x} F(y)$和$F(x+) = \lim_{y\downarrow x} F(y)$都是定义明确的，根据其定义，$F(x-) \le F(x) \le F(x+)$。

现在让$C = \{x \in$R : F在x处不连续$\}$。对于任何$x \in C$，我们有$F(x-) < F(x+)$。因此我们可以在开放区间（$F(x-), F(x+)$）内找到一个有理数$r = r(x)$。没有两个不同的x能有相同的$r(x)$，因为如果$x_1 < x_2$ 我们从这些极限的定义中得到$F(x_1+) \le F(x_2-)$。因此，对应关系x $r(x)$定义了C和Q的一个子集之间的一一对应关系，所以C最多是可数的。在每个不连续处，我们有$F(x-) < F(x+)$，所以所有的不连续都是由F的跳跃造成的。

证明（命题7.12的证明）

固定ε>0，让一个有限的不相交的区间集合$J_k = (x_k, y_k)$。让$E = \bigcup_k J_k$。那么

$$\sum_{k}^{n} |F(y_k) - F(x_k)| = \sum_{k}^{n} \left| \int_{x_k}^{y_k} f\, dm \right| \le \sum_{k}^{n} \int_{x_k}^{y_k} |f|\, dm = \int_E |f|\, dm。$$

但是，由于$f \in L^1$，度量$\mu(G) = \int_G |f|\, dm$是绝对连续的，有关于Lebesgue度量m，因此，根据命题7.1，存在δ>0，使得$m(F)<\delta$意味着$\mu(F)<\varepsilon$。但是如果区间J_k 的总长度小于δ，那么$m(F)<\delta$，因此$\mu(F)<\varepsilon$。这证明函数F是绝对连续的。 □

证明（命题7.14的证明）

使用提示中定义的函数：对于$[a, x]$的任何分区（x）$_{kk\leq n}$，我们有

$$F(x) - F(a) = \sum_{k=1}^{n} [F(x_k) - F(x_{k-1})]$$

$$= \sum_{k=1}^{n} [F(x_k) - F(x_{k-1})]^+ - \sum_{k=1}^{n} [F(x_k) - F(x_{k-1})]^-$$

$$= p(x) - n(x)$$

因此，根据sup的定义，p（x）$=n$（x）$+[F$（b）-F（a）$]\leq N_F$（x）$+[F$（（b）$-F$（a））]。这对所有分区都成立，因此$P_F(x) = \sup p(x) \leq N_F(x) + [F((b) - F(a)]$。另一方面，写$n(x)=p(x)+[F(a)-F(b)]$得到$N_F(x)\leq P_F(x)+[F(a)\ F(b)]$。因此，$P_F(x)\ N_F(x) = F(b)\ F(a)$。现在对于任何固定的分区

我们有

$$T_F(x) \geq \sum_{k=1}^{n} |F(x_k) - F(x_{k-1})| = p(x) + n(x) = p(x) + \{p(x) - [F(b) - F(a)]\}$$

$$= 2p(x) - [P_F(x) - N_F(x)] = 2p(x) + N_F(x) - P_F(x)。$$

取右边的最高值：$T_F(x) \geq 2P_F(x) + N_F(x) - P_F(x) = P_F(x) + N_F(x)$。但我们也可以写成$\sum^n |F(x_k) - F(x_{k-1})| = P(x) + n(x) \leq P_F(x) + N_F(x)$ 对于任何分区，所以，取左边的sup提供了$T_F(x) \leq P_F(x) + N_F(x)$。所以两边是相等的。 □

证明（命题7.15的证明）

对于T_F，证明这一点就足够了，因为其他情况也类似。如果$[a, b]$的分区P产生绝对差值的总和t(P)，如果P′=P∪{c}，对于一些$c\in(a, b)$，那么$t(P)[a, b]\leq t(P')[a, c]+t(P')[c, b]$，这在上面由$T_F[a, c] + T_F[c, b]$ 对于所有分区的约束。因此它也约束了$T_F[a, b]$。关于

另一方面，$[a, c]$和$[c, b]$的任何分区共同构成了$[a, b]$的分区，因此，$T_F[a, b]$界于它们的共同和。所以两边一定是相等的。特别是，固定a，T_F（$[a, c] \leq T_F[a, b]$，当$c\leq b$时，因此T_F（x）$= T_F[a, x]$是随x增加的。对于P_F和N_F也是如此。最后的声明是很明显的。 □

证明（命题7.17的证明）

(i) 给定$\varepsilon>0$，找到$\delta>0$，使$\sum_{i=1}^{n}(y_i-x_i)<\delta$，意味着$\sum_{i=1}^{n}|F(y_i)-F(x_i)|<\frac{\varepsilon\delta}{b-a}$。给定$[a, b]$的一个分区$(t_i)_{i\leq K}$，我们进一步添加分区点，间隔均匀，距离为 $\frac{b-a}{M}$彼此之间的关系，以确保

组合分区$(z_i)_{i\leq N}$，其所有的点相距都小于δ。要做到这一点，我们只需要选择M作为$T=\frac{b-a}{\delta}+1$的整数部分。由于(t_j)构成了分区点的一个子集$(z_i)_{i=0,1,\ldots,N}$，因此可以得出

$$\sum_{i=1}^{K}|F(t_i)-F(t_{i-1})|\leq\sum_{i=1}^{N}|F(z_i)-F(z_{i-1})|。$$

后者的总和可以重新排序为M组条款，其中每组以两个连续的新分区点开始和结束：然后k^{th}组总共包含（例如）m_k点，根据它们的结构，它们的连续距离之和（即两个新端点之间的距离！）。

小于δ，所以对于每个$k\leq M$，$\sum^{m_k}|F(w_{i,k})-F(w_{i-1,k})|<\frac{\varepsilon\delta}{b-a}$，其中$w_{i,k}$）是重新排序的点（$z_i$）。因此，整个总和的界限是$M(\frac{\varepsilon\delta}{b-a})\leq T(\frac{\varepsilon\delta}{b-a})<\varepsilon$。这表明$F\in BV[a,b]$，因为约束是独立于原始分区$(t_i)_{i\leq K}$。

对于(ii)，首先注意到，根据(i)，函数F在以下方面具有有界的变化$[a,b]$，因此在任何子区间$[x_i,y_i]$上，总变化函数$T_F[x_i,y_i]$是有限的。再如绝对连续的定义中给出的ε,δ。

众多函数。如果$(x_i,y_i)_{i\leq n}$是子区间，$\sum^{n}|y_i-x_i|<\delta$，则$\sum_{i=1}^{n}|F(y_i)-F(x_i)|<\varepsilon$。如上一个命题，这意味着$T_F[x_i,y_i]\leq\varepsilon$。因此$P_F[x_i,y_i]$和$N_F[x_i,y_i]$都小于$\varepsilon$，所以函数$F_1$和$F_2$是绝对连续的。 □

证明（命题7.18的证明）

显然$v(\varnothing)=0$，对于任何E，$\varnothing\subset E$。所以如果v是单调递增的，$v(E)\geq v(\varnothing)\geq 0$。反过来说，如果$v$是一个度量，$F\subset E$，$v(E)=v(F)+v(E\backslash F)\geq v(E)$。 □

证明（命题7.26的证明）

如果$E=\{f>1\}$有$\rho(E)>0$，$\frac{f}{\rho(E)}$是定义良好的，且$\int_E\frac{f}{\rho(E)}d\rho=\frac{1}{\rho(E)}\int_E f\,d\rho>1$.这与假设相矛盾，所以$\rho(E)=0$.类似地、$F=\{f<-1\}$有$\rho(F)=0$.因此$|f|\leq 1$ ρ-a.e. □

证明（命题7.27的证明）

按照提示选择h、A、B。回顾一下，$v^+=\frac{1}{2}(|v|+v)$，并注意：$\frac{1}{2}(1+h)=h1_B$，因此，对于$F\in$F、

$$v^+(F)=\frac{1}{2}\int_F(1+h)\,d|v|)=\int_F$$

$h\,\mathrm{d}|\nu| = \nu(F \cap B)$。

但这样一来，由于$B=A^c$、

ν^- （F） =-[ν （F） -ν^+ （F）]=-[ν （F） -ν （$F\cap B$）]=-ν （$F\cap A$） 。

最后，如果$\nu = \lambda_1 - \lambda_2$ ，其中λ_i 是度量，那么$\nu \le \lambda_1$ ，所以$\nu^+(F) = \nu(F\cap B) \le \lambda_1(F\cap B) \le \lambda_1(F)$，通过单调性。这证明了命题的最后陈述。

□

证明（命题7.28的证明）

(i) 是直接的，因为$\int_\Omega E(X|G)\, dP = \int_\Omega X\, dP$ 根据定义。

(ii) 如果两个积分都是G可测量的，并且$\int_G E(X|G)\, dP = \int_G X\, dP$为所有的$G\in$G，那么根据定理4.15，积分点是a.s.相等的，因此，X是E($X|G$)的一个版本。

(iii) 对于任何$G\in$G，1_G和X是独立的随机变量，所以

$$\int_G X\, dP = E(X1_G) = E(X)E(1_G) = \int_G E(X)\, dP$$

因此，根据定义，E(X)是E(X|G)的一个版本.但E(X)是常数，所以这个特性在任何地方都成立。

(iv) 使用积分的线性特性：

$$\int_G (aX + bY) = a\int_G X\, dP + b\int_G Y\, dP = a\int_G E(X|G)\, dP + b\int_G E(Y|G)\, dP$$

$$= \int_G [aE(X|G)\, dP + bE(Y|G)]\, dP、$$

所以，结果是这样的。 □

8 极限定理

在这一章中，我们引入了一些节奏上的变化，读者在初读时可以省略那些技术要求较高的证明，以获得对随机变量序列的主要极限定理的概述。我们把重点放在概率上，以推导前述理论的实质性应用。

然而，首先我们讨论实变函数序列的一些基本收敛模式。然后，我们转到本章主要涉及的概率论环境。

8.1 收敛的模式

假设E是R的一个Borel子集^n。对于L^p （E） 中的一个给定序列 （f_n ），$p\geq 1$，我们可以用一些不同的方式来表达'$f_n \to f$ as $n \to \infty$'这句话：

定义8.1

(1) $f_n \to f$*均匀地在E上*：给定$\varepsilon > 0$，存在$N = N(\varepsilon)$，这样，对于所有$n \geq N$、

$$\|f_n - f\|_\infty = \sup_{x \in E}(|f_n(x) - f(x)|) < \varepsilon。$$

(注意，我们需要$f_n \in L^\infty(E)$，才能使sup在一般情况下是有限的)。

(2) $f_n \to f$*在E上的指向性*：对于每个$x \in E$，给定$\varepsilon > 0$，存在$N = N(\varepsilon, x)$，使得$|f_n(x) - f(x)| < \varepsilon$，对于所有$n \geq N$。
(3) $f_n \to f$在E上*几乎无处不在*（a.e）：有一个空集$F \subset E$，使得$f_n \to f$在$E \backslash F$上的指向性。
(4) $f_n \to f$ *in* $L^p - norm$ (in the pth mean): $\|f_n - f\|_p \to 0$ as $n \to \infty$, i.e. for given $\varepsilon > 0$, $\exists N = N(\varepsilon)$ such that

$$\|F_n - F\|_p = \left(\int_E |f_n - f|^p \, dm\right)^{\frac{1}{p}} < \varepsilon$$

对于所有的

$n \geq N$。

处理这些不同的收敛模式需要一些小心；在这一点上，我们甚至还没有证明它们都是真正不同的。显然，点式（和a.e.）极限往往更容易 "猜测"，然而，我们不能总是确定极限函数，如果它存在的话，又是L^p (E)的一员。

然而，对于平均收敛来说，这是由$L^p(E)$的完备性保证的，同样，对于均匀收敛来说，这只是在L^∞-norm中的收敛。请注意，支配性收敛和单调性收敛的结论是

定理产生了a.e.收敛序列(f_n)的平均收敛性，但只有通过对(f_n)施加额外的条件。

定理8.1

在(f_n)如上的情况下，唯一有效的含义是以下几点：(1) ⇒ (2) ⇒ (3).对于有限措施，（1）⇒（4）。

证明

上述含义是显而易见的。同样明显的是，（3）/⇒（2）。为了看到(2)/⇒(1)，取$E = [0, 1]$, $f_n = 1_{(0, \frac{1}{n})}$，它在所有的时候都收敛于$f = 0$点，但对于所有的$n$，$\sup f_n = 1$。

对于(3)/⇒(4)，取$f_n = n1_{(0, \frac{1}{n}]}$; f_n点式收敛到0，但$\int_0^1 f_n^p \, dm = n^p \frac{1}{n} = n^{p-1} \geq 1.$

为了看到（4）/⇒（3），让E=[0，1]，并把

$$g1 = 1_{[0, \frac{1}{2}]} \quad G2 = 1_{[\frac{1}{2}, 1]}$$

$$g_3 = \mathbf{1}_{[0,\frac{1}{4}]} \quad g_4 = \mathbf{1}_{[\frac{1}{4},\frac{2}{4}]} \quad g_5 = \mathbf{1}_{[\frac{2}{4},\frac{3}{4}]} \quad g_6 = \mathbf{1}_{[\frac{3}{4},1]}$$

$\ldots$

那么
$$\int_0^1 g_n^p dm = \int_0^1 g_n\, dm \to 0$$

但对于每个$x \in [0,1]$，$g_n(x)=1$的无限多的n，所以$g_n(x)$在E中的任何x处都不收敛。□

例8.1

我们调查
$$h_n(x) = x^n$$

为在E=[0，1]上的每个模式中的收敛性。

该序列在任何地方都收敛于函数$h(x)$=0，对于$x \in [0,1)$, $h(1)$=1，因此它也几乎到处收敛。

它不会均匀地收敛，因为$\sup_{[0,1]} |h_n(x) - h(x)| = 1$，对于所有的$n$。对于$p$>0，它在$L^p$收敛：
$$\int_0^1 |h_n(x) - 0|^p dx = \int_0^1 x^{pn}\, dx = \frac{1}{pn+1} x^{pn+1}\Big|_0^1 \to 0.$$

备注8.1

对于可测函数序列，还有其他的收敛模式可以考虑，而这些模式与上述模式之间的关系一般来说是相当复杂的。在这里，我们将不对这一主题进行一般性的探讨，而是专门讨论概率空间，在这里我们将推导出不同极限过程之间的额外关系。

练习8.1

对于以下每一项决定$f_n \to 0$ (i)在L^p, (ii)均匀地、(iii)点式，(iv)a.e.。

(a) $f_n = 1_{[n,n+\frac{1}{n}]}$、

(b) $f_n = n1_{[0,\frac{1}{n}]} - n1_{[-\frac{1}{n},0]}$。

8.2 P robability

本章的其余部分专门讨论了概率论中随机变量的基本极限定理。术语的定义是

概率 "依赖于对这种结果的信念，也就是说，我们可以对一连串独立的相同分布的试验中的成功的 "极限平均 "赋予一种意义。那么抛硬币的'无休止的重复'的目的就是用正面的'极限频率'作为正面概率的定义。

同样，高斯密度在统计学中被赋予的关键作用也起源于著名的中心极限定理（实际上有很多不同的定理），该定理表明该密度描述了适当条件下的分布序列的极限分布。因此，分布的收敛为随机变量序列提供了另一个重要的极限概念。

在这两种情况下，独立性的概念起着至关重要的作用，首先我们需要把这个概念扩展到无限的随机变量序列。在下面的内容中、

$$X_1, X_2, \ldots, X_n, \ldots$$

将表示一个定义在概率空间上的随机变量序列。

定义8.2

我们说随机变量$X_1, X_2, \ldots$ 是独立的，如果对于任何$n\in\mathbb{N}$的变量$X_1, \ldots \ldots, X$ *是独立的（见定义3.3）。*, X_n 是独立的（见定义3.3）。

另一种方法是要求X的任何有限集合$_i$， 是独立的。当然，这个条件意味着独立，因为有限集合涵盖了n个变量的初始片段。

反过来说，取任何一个X的有限集合$_i$，让n是这个有限集合的最大索引。现在，$X_1, \ldots, X_n$ 是独立的，那么对于每个子集来说，其元素的集合都是独立的；这尤其包括所选择的那个。

我们研究以下序列

$$s_n = x_1 + \cdots + x_n。$$

如果所有的X_i 具有相同的分布（我们说它们是*相同分布的*），那么$\frac{s_n}{n}$是同一实验重复n次后X_n （或X_1，这并不重要）的平均值。

我们研究S的行为$_n$，因为n到了无穷大？我们要解决的两个主要问题是：

1. $\frac{s_n}{n}$的随机变量何时收敛到某一数字？这里有一个关于适当的收敛方式的

直接问题。这类问题的肯定答案被称为大数法则。

2. 随机变量$\frac{S_n}{}$的分布何时会收敛到一个数学模型。 n
   确定吗？在什么条件下这个极限测量是高斯的？我们得到的回答结果被称为中心极限定理。

8.2.1 概率中的收敛性

我们的第一个额外的收敛模式，即概率上的收敛，有时被称为 "措施上的收敛" 。

定义8.3

一个序列$X_1, X_2, ...$在概率上收敛于X，如果对于每个$\varepsilon > 0$

$$P\left(|X_n - X| > \varepsilon\right) \to 0$$

当$n \to \infty$时。

练习8.2

回到定理8.1的证明（E=[0，1]），看看在那里构建的随机变量序列中，有哪些在概率上是收敛的。

练习8.3

找出一个[0，1]上的随机变量序列在概率上不收敛于0的例子。

我们首先证明，几乎肯定的收敛（即几乎每一个地方）要强于概率上的收敛。但首先我们要证明一个辅助性的结果。

定理8.2

以下条件是等同的

(a) $Y_n \to 0$ 几乎肯定

(b) 对于每个ε>0、

∞

$$\lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty}\{\omega : |Y_n(\omega)| \geq \varepsilon\}\right) = 0。$$

证明

简明扼要地表达，趋同几乎肯定意味着

$$P(\{\omega : \forall \varepsilon > 0\ \exists N \in \mathbb{N} : \forall n \geq N, |Y_n(\omega)| < \varepsilon\}) = 1.$$

用另一种方式书写这组全量，我们有

$$P\Big(\bigcap_{\varepsilon>0}\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big) = 1。$$

外交的概率（在所有$\varepsilon>0$的情况下）小于其任何项的概率，但由于已经是1，它不能增加，因此对于所有$\varepsilon>0$的情况下

$$P\Big(\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big) = 1。$$

我们有一个增加集的联盟，所以

$$\lim_{N\to\infty} P\Big(\bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big) = 1$$

因此

$$\lim_{N\to\infty}\Big(1 - P\Big(\bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big)\Big) = 0$$

但我们可以写成1=$P(\Omega)$，所以

$$P(\Omega) - P\Big(\bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big) = P\Big(\Omega \setminus \bigcap_{n\geq N}\{\omega : |Y_n(\omega)| < \varepsilon\}\Big)$$
$$= P\Big(\bigcup_{n\geq N}\{\omega : |Y_n(\omega)| \geq \varepsilon\}\Big)$$

根据德摩根定律。因此，（a）意味着（b）。向后努力，这些步骤也证明了反面。 □

定理8.3

如果$X_n \to X$几乎肯定，那么$X_n \to X$的概率。

证明

为了简化符号，考虑差值$Y_n = X_n - X$，问题简化为讨论Y_n的收敛性为零。我们有

$$\lim_{k\to\infty} P\Big(\bigcup_{n=k}^{\infty}\{\omega : |Y_n(\omega)| \geq \varepsilon\}\Big) \geq \lim_{k\to\infty} P(\{\omega : |Y_k(\omega)| \geq \varepsilon\})$$

根据Lemma 8.2，左边的极限是零，因此右边的极限也是零。 □

请注意，不等式的两边整齐地概括了收敛性（a.s.）和概率上的区别。对于概率上的收敛，我们考虑单个Y_n与极限至少相差ε的概率、而为了几乎确定的收敛性，我们需要同时考虑整个尾部序列（Y）$_{n n\geq k}$。

下面的例子表明，上述定理中的暗示不能逆转，也表明L^p中的收敛并不意味着几乎肯定收敛。

例8.2

考虑以下定义在Ω=[0，1]上的具有Lebesgue度量的随机变量序列：$Y_1 = 1_{[0,1]}$，$Y_2 = 1_{[0,1/2]}$，$Y_3 = 1_{[1/2,1]}$，$Y_4 = 1_{[0,1/4]}$，$Y_5 = 1_{[1/4,1/2]}$，等等（就像定理8.1的证明中一样）。该序列显然在概率上和L^p上收敛为零，但对于每个$\omega \in [0, 1]$，$Y_n(\omega) = 1$

对于无限多的n，所以它不能顺时收敛。

概率上的收敛还有一个有用的特点：

命题8.4

由$d(X, Y)=\mathrm{E}\left(\frac{|X-Y|}{1+|X-Y|}\right)$定义的函数是一个度量，并且收敛于$d$相当于概率上的收敛。

温馨提示 如果$X_n \to X$的概率，那么将期望值分解为$\int_A + \int_{\Omega \setminus A}$其中$A = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}$。

我们现在通过非负随机变量的矩，对该随机变量在某一特定集合中取值的概率给出一个基本估计。

定理8.5（切比雪夫不等式）

如果Y是一个非负的随机变量，$\varepsilon>0$，$0<p<\infty$，那么

$$P(Y \geq \varepsilon) \leq \frac{\mathrm{E}(Y)^p}{\varepsilon^p}. \tag{8.1}$$

证明

从积分的基本属性来看，这是直接的：让$A = \{\omega : Y(\omega) \geq \varepsilon\}$。

然后

$$E(Y^p) \geq \int_A Y^p\,dP \quad \text{(在一个较小的集合上进行整合)}$$

$$\geq P(A) \cdot \varepsilon^p$$

因为在A上$Y^p(\omega) > \varepsilon^p$，这就得到了除以$\varepsilon$后的结果^p。 □

切比雪夫不等式将主要用于小的ε，但让我们看看如果ε很大会发生什么

。

命题8.6

假设$E(Y^p) < \infty$。那么

$$\varepsilon^p\,P(Y \geq \varepsilon) \to 0 \quad 当\varepsilon \to \infty时。$$

提示写作

$$E(Y^p) = \int_{\{\omega: Y(\omega) \geq \varepsilon\}} Y^p\,dP + \int_{\{\omega: Y(\omega) < \varepsilon\}} Y^p\,dP$$

并按切比雪夫不等式证明中的方法估计第一项。

推论8.7

设X是一个随机变量，具有有限期望值$E(X)=m$和方差σ^2。让$0 < a < \infty$；那么

$$P(|X - m| \geq a\sigma) \leq \frac{1}{a^2}。$$

证明

使用切比雪夫不等式，$Y = |X - m|$和$p=2$，$\varepsilon=a\sigma$，我们发现

$$P(|X - m| \geq a\sigma) \leq \frac{E(|X-m|^2)}{a^2\sigma^2} = \frac{1}{a^2}$$

根据需要。 □

备注8.2

切比雪夫不等式还表明，L^p 中的收敛意味着概率的收敛。因为，让 $Y_n = |X_n - X|$，并假设 $X_n - X_p \to 0$。这意味着 $P(Y_n\ \varepsilon)\ 0$。反之，一般情况下是错误的，因为接下来的

例子显示。

例8.3

让$\Omega=[0, 1]$，具有Lebesgue度量，让$X_n = n1_{[0, \frac{1}{n}]}$。该序列$X_n$点式收敛到0，所以我们取$X = 0$，我们看到$\|X_n - X\|_p^p = \int_0^{\frac{1}{n}} n^p\, dm = n^{p-1}$。如果$p\geq1$，则$\|X_n - X\|_p \not\to 0$，然而正如我们已经见（练习8.2）、

$$P(|X_n - X| \geq \varepsilon) = P(X_n = n) = \frac{1}{n} \to 0$$

显示$X_n \to X$的概率。

8.2.2 大 数字的弱规律

最简单的 "大数法则 "提供了一个L^2 -convergence的结果：

定理8.8

如果$X_1, X_2, \ldots$是独立的，$E(X_i) = m, \mathrm{Var}(X_i) \leq K < \infty$， 那么$\frac{S_n}{n} \to m$在$L^2$，因此在概率上。

证明

首先注意到，$E(S_n) = E(X_1) + \cdots + E(X_n) = nm$，根据表达式的线性的计算。因此，$E(\frac{S_n}{n}) = m$，所以$E(\frac{S_n}{n} - m)^2 = \mathrm{Var}(\frac{S_n}{n})$。根据的方差（命题5.20）。

$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{1}{n^2}(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)) \leq \frac{1}{n^2}nK = \frac{K}{n} \to 0$$

作为$n \to \infty$.这又意味着概率的收敛，正如我们在Re-mark 8.2中看到的那样。□

练习8.4

利用切比雪夫不等式，找出100次抛掷硬币的平均头数与以下 情况不同的概率下限 $\frac{1}{2}$ 0.1。

练习8.5

找出1000次掷骰子所显示的平均数字与3.5相差0.01的概率的下限。

我们给出了弱大数法则的一些经典应用。魏尔斯特拉斯定理说，每个连续函数都可以被多项式均匀地应用。切比雪夫不等式提供了一个简单的证明。

定理8.9（Bernstein-Weierstrass近似定理）

如果$f : [0, 1] \to \mathbb{R}$是连续的，那么伯恩斯坦多项式的序列

$$f_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

均匀地收敛于f。

证明

数字$f_n(x)$有一个概率意义。也就是说，$f_n(x) = \mathrm{E}(f(\frac{S_n}{n}))$ where $S_n = X_1 + \cdots + X_n$,

$$X_i = \begin{cases} 1，概率为x \\ 0，概率为1-x。\end{cases}$$

然后把E写成E_x，而不是E，以强调基本概率能力取决于x的事实，我们有

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| \le \sup_{x\in[0,1]} \left|\mathrm{E}_x\left(f\left(\frac{S_n}{n}\right)\right) - f(x)\right|$$
$$\le \sup_{x\in[0,1]} \mathrm{E}_x \left|f\left(\frac{S_n}{n}\right) - f(x)\right|。$$

取任何$\varepsilon>0$，并找到$\delta>0$，使得如果$|x-y|<\delta$，那么$|f(x)-f(y)|<\frac{\varepsilon}{2}$(这是有可能的，因为$f$是均匀连续的）。

$$\mathrm{E}_x \left|f\left(\frac{S_n}{n}\right) - f(x)\right| = \int_{\{\omega:|\frac{S_n}{n}-x|<\delta\}} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP$$
$$+ \int_{\{\omega:|\frac{S_n}{n}-x|\ge\delta\}} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP$$
$$\le \frac{\varepsilon}{2} + 2\sup_{x\in[0,1]} |f(x)| \cdot P\left(\left|\frac{S_n}{n} - x\right| \ge \delta\right)。$$

最后一项根据大数法则收敛为零，因为$x = \mathrm{E}(\frac{S_n}{n})$、而$\mathrm{Var}(\frac{S_n}{n})$是有限的。这种收敛性在$x$中是均匀的：

$$P\left(\left|\frac{S_n}{n}\right.\right. \qquad \frac{\mathrm{Var}(\frac{S_n}{n})}{}$$

$- x| \geq \delta) \leq \quad \delta^2 \quad = \quad x(1-x)\ n^{\delta}2 \quad \frac{n}{} \quad \frac{}{1} \quad \leq \ ^{4n}\delta 2$

(由于$4x(1 x)$ ≤ 1和$\mathrm{Var}(S_n) = nVar(X_i) = nx(1 x)$)。因此，对于足够大的$n$，右手边小于$\varepsilon$，这就完成了证明。 □

在许多实际情况下，不可能直接计算积分（特别是面积或体积）。大数法是所谓的蒙特卡洛方法的基础，该方法通过随机选点给出一个近似的解决方案。下面的两个例子说明了这种方法。

例8.4

为了简单起见，我们把自己限制在$F\subset[0，1]\times[0，1]$。假设F是Lebesgue可测量，我们的目标是找到它的度量。让X_n, Y_n是inde-垂线，均匀分布在[0，1]。让$M_n = \frac{1}{n}\sum_{k=1}^{n} 1_F(X_k, Y_k)$。如果我们抽出$n$次数字对，那么这个总和就给出了集合$F$的命中率，而$M_n \to m(F)$的概率。要看到这一点，首先观察$\mathrm{E}(1_F(X_k, Y_k)) = P((X_k, Y_k) \in F) = m(F)$由$X_k, Y_k$的分布假设：独立性保证了这对数字的分布是二维Lebesgue度量限制为平方。那么

$$P(|M_n - m(F)| \geq \varepsilon) \leq \frac{m(F)}{n\varepsilon} \to 0.$$

一个类似的例子说明了使用蒙特卡洛方法来计算积分。

例8.5

设f是一个定义在[0, 1]上的可积分函数。在X_n，独立的单一形式分布在[0，1]上，我们取

$$I_n = \frac{1}{n}\sum_{k=1}^{n} f(X_k)$$

并且我们表明，

$$I_n \to \int_0^1 f(x)\,dx$$

在概率上。首先注意到，X的分布$_k$，是Lebesgue度量在[0,1] 因此

$$\mathrm{E}(f(X_k)) = \int f(x)\,dP_{X_k}(x) = \int_0^1 f(x)\,dx、$$

所以$\mathrm{E}(I_n$

$) = \int^{1} f(x)\,dx$ 。弱法提供了所需的收敛性。

$$P\left(\left|I_n - \int_0^1 f(x)\,dx\right| \geq \varepsilon\right) \to 0。$$

我们再来进一步考虑随机变量序列的弱大数法。E(X_k)和Var(X_k)是有限的这一假设可以放宽。然而，施加额外的条件是要付出代价的。

首先我们介绍一些方便的符号。对于一个给定的Ran-dom变量序列(X)$_{kk\geq 1}$ 介绍截断的随机变量

$$x_k(n) = x_k \cdot 1\{\omega : |X_k(\omega)| \leq n\}。$$

并设m_n = E(X_1 (n))。注意，m_n = E(X_k (n)，对于所有$k\geq 1$，因为所有X_k的分布是相同的。还可以写成

$$\hat{S}_m = X_1(n) + \cdots + X_m(n)$$

对于所有的$m\geq 1$。

定理8.10

如果X_k 是独立的同分布的随机变量，以便

$$aP(|X_1| > a) \to 0 \quad 当a \to \infty时、 \tag{8.2}$$

然后

$$\frac{S_n}{n} - m_n \to 0 \quad 在概率上。 \tag{8.3}$$

我们将需要以下定理，它本身就很有意义，而且在下面的内容中也很有用。

定理8.11

如果$Y\geq 0$，$Y\in L^p$ ，$0 < p < \infty$，则

$$E(Y^p) = \int_0^\infty p y^{p-1} P(Y > y)\, dy。$$

特别是(p=1)

$$E(Y) = \int_0^\infty P(Y > y)\, dy。$$

证明(该定理的证明)

这是富比尼定理的一个简单应用：

$$\int_0^\infty py^{p-1}\,\mathrm{P}\,(Y>y)\,dy$$

$$= \int\int^\infty pyp\text{-}11_{\{\omega:Y\,(\omega)>y\}}(\omega)\,dP\,(\omega)\,dy$$

$$= \int^0 \int_\Omega^\infty py\ 1^{p-1}_{\{\omega:Y\,(\omega)>y\}}(\omega)\,dy\,dP\,(\omega)\ \text{(通过Fubini)}$$

$$= \int^\Omega \int_0^{Y\,(\omega)} py^{p-1}\,dy\,dP\,(\omega)$$

$$= \int_\Omega Y^p\,(\omega)\,dP\,(\omega)\ \text{（计算内积分，}\omega\text{固定）。}$$

$$= \mathrm{E}(Y)^p$$

根据需要。 □

证明（该定理的证明）

取$\varepsilon>0$，显然

$$P\left(\left|\frac{S_n}{n}-m_n\right|\geq\varepsilon\right)\leq P\left(\left|\frac{\hat{S}_n}{n}-m_n\right|\geq\varepsilon\right)+P\left(\hat{S}_n\neq S_n\right)。$$

陕西省

我们估计第一个项

$$P\left(\left|\frac{\hat{S}_n}{n}-m_n\right|\geq\varepsilon\right)\leq\frac{E\left(\left|\frac{\hat{S}_n}{n}-m_n\right|^2\right)}{\varepsilon^2}\ \text{(切比雪夫著)}$$

$$=\frac{E\left(\left|\sum X_k\,(n)-nm_n\right|^2\right)}{n2\varepsilon^2}$$

$$=\frac{\mathrm{Var}\left(\sum X_k\,(n)\right)}{n^2\varepsilon 2}.$$

请注意，被截断的随机变量是独立的，是以下的函数

因此，我们可以继续进行估计：

$$
\begin{aligned}
&\le \frac{\sum_k \operatorname{Var}(X_k(n))}{n^2\varepsilon^2}\\
&= \frac{\operatorname{Var}(X_1(n))}{n\varepsilon^2} \qquad (\text{因为}\operatorname{Var}(X_k(n))\text{是相同的})\\
&\le \frac{\mathrm{E}(X_1^2(n))}{n\varepsilon^2} \qquad (\text{通过 }\operatorname{Var}(Z)=\mathrm{E}(Z^2)-(EZ)^2\le \mathrm{E}(Z^2))\\
&= \frac{1}{n\varepsilon^2}\int_0^\infty 2yP(|X_1(n)|>y)\,dy \qquad (\text{根据}p=2\text{的定理})\\
&= \frac{1}{n\varepsilon^2}\int_0^n 2yP(|X_1|>y)\,dy。
\end{aligned}
$$

假设函数$y' \to 2yP(|X_1|>y)$随着$y\to\infty$而收敛到0，因此对于给定的$\delta>0$，有y_0，使得对于$y\ge y_0$，这个量小于$\frac{1}{2}\delta\varepsilon^2$，我们有

$$
\begin{aligned}
&= \frac{1}{n\varepsilon^2}\int_0^{y_0} 2yP(|X_1|>y)\,dy + \frac{1}{n\varepsilon^2}\int_{y_0}^{n} 2yP(|X_1|>y)\,dy\\
&\le \frac{1}{n\varepsilon^2}\, y_0 \max_{y\in[0,\infty]}\{yP(|X_1|>y)\} + \frac{1}{n\varepsilon^2}\frac{\delta\varepsilon^2 n}{2}\\
&\le \delta
\end{aligned}
$$

只要n足够大。所以第一项收敛为0。现在我们处理第二项：

$$
\begin{aligned}
P(\hat{S}_n \neq S_n) &\le P(X_k(n)\neq X_k \text{ for some } k\le n)\\
&\le \sum_{k=1}^{n} P(X_k(n)\neq X_k) \qquad (\text{通过}P\text{的次可加性})\\
&= nP(X_1(n)\neq X_1) \qquad (\text{相同的分布})\\
&= nP(|X_1|>n)\\
&\to 0 \qquad \text{通过假设。}
\end{aligned}
$$

这样就完成了证明。 □

备注8.3

请注意，我们不能将最后一个定理推广到不相关的随机变量的情况，因为我们基本使用了独立性。尽管身份

$$
\operatorname{Var}\sum X_k(n) = \sum \operatorname{Var}(X_k(n))
$$

对于不相关的随机变量来说，我们需要(X_k)的独立性--这意味着($X_k(n)$)的独立性--来得出截断的随机变量是不相关的结论。

定理8.12

如果X_n是独立和相同的分布，$E(|X_1|) < \infty$，那么(8.2)得到满足，$m_n \to m = E(X_1)$，$\frac{S_n}{n} \to m$ 的概率。(注意，我们在此不假设X_1具有有限方差)。

证明

$|X_1|$的有限期望给出条件（8.2）：

$$aP(|X_1| > a) = a\int_\Omega 1_{\{\omega:|X_1(\omega)|>a\}}\,dP$$

$$\leq \int_\Omega \ldots$$

$$= \int_{\{\omega:|X_1(\omega)|>a\}} |X_1|\,dP$$

$$\to 0$$

根据支配收敛定理，$a \to \infty$。因此$\frac{S_n}{n} - m_n \to 0$，但 $m_n = E(X_1 1_{\{\omega:|X_1(\omega)|\leq n\}}) \to E(X_1)$，因为$n \to \infty$，所以结果如下。 □

8.2.3 Borel-Cantelli Lemmas

一个点$\omega\in\Omega$属于给定序列(A_n)的'无限多'事件，这个想法可以很容易地精确化：对于每一个$n \geq 1$，我们需要找到一个$m \geq n$，使$\omega \in A_m$。可以很容易地精确化：对于每一个$n \geq 1$，我们需要能够找到一个$m \geq n$，使$\omega \in A_m$。这就确定了一个子序列

(m_n)的指数，使得对于每一个$n \geq 1$，$\omega \in A_{m_n}$，即$\omega \in \bigcup_{m \geq n} A_m$ 对于每一个$n \geq 1$。因此，我们说$\omega \in A$ *无限频繁*，并写成$\omega \in A$ i.o.，如果$\omega \in \bigcap_{n\geq 1}\bigcup_{m \geq n} A_m$。我们称这个集合为序列的上限（$A_n$）并将其写为

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m.$$

练习8.6

对于一个序列$A_1 = [0, 1]$, $A_2 = [0, \frac{1}{2}]$, $A_3 = [\frac{1}{2}, 1]$, 求$\limsup_{n\to\infty} A_n$、
$A_4 = [0, \frac{1}{4}]$, $A_5 = [\frac{1}{4}, \frac{1}{2}]$ 等等。

给定$\varepsilon>0$，一个随机变量序列(X_n)和一个随机变量X，对于每个n集合$A_n = \{|X_n - X| > \varepsilon\}$。那么$\omega \in A_n$ i.o.正是当对于每一个$\varepsilon > 0$，$|X_n(\omega) - X(\omega)| > \varepsilon$发生在无限子序列$(m_n)$的所有元素的索引上，这意味着$(X_n)$未能收敛到$X$。

因此，可以看出，

$X_n \to X$ a.s.(P) $\iff$ $\forall\varepsilon > 0$ $P(\limsup_{n\to\infty} A_n) = P(|X_n - X| > \varepsilon \text{ i.o.}) = 0$。

同样地，定义

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m。$$

(我们说这个集合是序列(A_n)的*下限*)。

命题8.13

(i) 我们有$\omega \in \liminf_{n\to\infty} A_n$，当且仅当$\omega \in A_n$，除了有限多的$n$.(我们说$\omega \in A_n$ *最终*)。

(ii) $P(X_n \to X) = \lim_{\varepsilon\to 0} P(|X_n - X| < \varepsilon$ 最终)。

(iii) 如果$A = \limsup_{n\to\infty} A_n$，那么$A^c = \liminf_{n\to\infty} A^c$。

(iv) 对于任何事件序列（A_n），P（$\{\omega \in A_n$ ev.$\}$）≤P（$\{\omega \in A_n$ i.o.$\}$）。

提示：使用Fatou关于(iv)中集合的指标函数的定理。

$\liminf_{n\to\infty} A_n$和$\limsup_{n\to\infty} A_n$是序列(A_n)的 "尾部事件"：我们只有通过了解整个序列才能确定一个点是否属于它们。通常情况下，一个尾部事件的概率是这样的

0或1--这样的结果被称为0 1定律。其中最简单的是由两个 "Borel-Cantelli定理 "结合提供的，我们现在就来看看这两个定理：它们共同表明，对于一个独立的事件序列（A_n），$\limsup_{n\to\infty} A_n$的概率是0或1，取决于它们各自的概率系列是收敛的还是发散的。在第一种情况下，我们甚至不需要独立，但可以在一般情况下证明这个结果。

练习8.7

Let $S_n = X_1 + X_2 + ... + X_n$ describe the position after n steps of a symmetric random walk on $\mathbb{Z}^d$. Using the asymptotic formula: $n! \sim \left(\frac{n}{e}\right)^n\sqrt{2\pi n}$和Borel-Cantelli法则表明，概率为

$\{S_n = 0 \text{ i.o.}\}$当d=1，2时为1，d>2时为0。

我们有以下简单但基本的事实。

定理8.14（Borel-Cantelli推论）

如果

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

然后

$$P(\limsup_{n\to\infty} A_n) = 0$$

即'$\omega \in A_n$ 为无限多的n'只以零的概率出现。

证明

首先注意，$\limsup_{n\to\infty} A_n$ ⊂ $\bigcup_{n=k}^{\infty} A_n$ 因此

$$P(\limsup_{n\to\infty} A_n) \le P(\bigcup_{n=k}^{\infty} A_n) \text{ (对于所有}k\text{)}$$

$$\le \sum_{n=k}^{\infty} P(A_n) \text{ (通过次可加性)}$$

$$\to 0$$

因为收敛数列的尾部会收敛到0。 □

该定理的基本应用提供了几乎肯定收敛和概率收敛之间的联系。

定理8.15

如果$X_n \to X$的概率，那么有一个子序列X_{k_n}收敛到X
几乎是肯定的。

证明

我们必须找到一个子序列会收敛的全度量集。因此，整个序列的行为是 "坏"的那个集合应该是度量为零的。为此我们采用Borel-Cantelli法则，其结论正是如此。因此，我们引入序列A_n，囊括了X_n 的'坏'行为，从收敛的角度看，它是由不等式表示的。

它的类型$|X_n(\omega) - X(\omega)| > a$。具体来说，我们取$a = 1$，因为$P(|X_n - X| > 1) \to 0$，我们找到$k_1$，以便

$$p(|x_n - x| > 1) < 1$$

对于 $n \geq k_1$ 。接下来对于 $a = \frac{1}{2}$ ，我们找到 $k_2 > k_1$ ，这样对于所有 $n \geq k_2$

$$p(|X_n - X| > \frac{1}{2}) < \frac{1}{4}.$$

我们继续这个过程，得到一个递增的整数序列k_n，其中有

$$P(|X_{k_n} - X| > \frac{1}{n}) < \frac{1}{n^2}。$$

系列$\Sigma_{n=1}^{\infty} P(A_n)$收敛，其中$A_n = \{\omega: |X_{k_n}(\omega) - X(\omega)| > \frac{1}{n}\}$，因此，$A = \limsup A_n$的概率为零。

我们观察到，对于$\omega \in \Omega \setminus A$，$\limsup X_{k_n}(\omega) = X(\omega)$。因为，如果$\omega \in \omega \setminus A$，那么对于某个$k$，$\omega \in \bigcap_{n=k}^{\infty}(\omega \setminus A_n)$，所以对于所有$n \geq k$，$|X_{k_n}(\omega) - X(\omega)| \leq \frac{1}{n}$。

因此，我们得到了理想的收敛性。 □

第二个Borel-Cantelli法则部分地完善了这一情况。在独立的附加条件下，它表明当无限多事件发生的概率为1时。

定理8.16

假设事件A_n是独立的。我们有

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n\to} A_n) = 1。$$

证明

只需证明对于所有k

$$P(\bigcup_{n=k}^{\infty} A_n) = 1$$

因为在k上的交集也会有1的概率。固定k，并考虑到$m>k$的部分联合。A的补充$_n$，因此也是独立的。

$$p(\bigcap_{n=k}^{m} a^c) = \prod_{n=k}^{m} p(a^c) = \prod_{n=k}^{m}(1 - p(a_n$$

因为$1-x \leq e^{-x}$、

m m m

$n=k$ $n=k$ $n=k$

$$\prod (1 - P(A_n)) \leq \prod e^{-P(A_n)} = \exp(-\sum P(A_n))。$$

根据假设，最后一个表达式在$m\to\infty$时收敛为0，因此

$$P\left(\bigcap_{n=k}^{m} A_n^c\right) \to 0$$

但

$$p\left(\bigcap_{n=k}^{m} A_n^c\right) = p\left(\omega \setminus \bigcup_{n=k}^{m} A_n\right) = 1 - p\left(\bigcup_{n=k}^{m} A_n\right)。$$

集合$B_m = \bigcup_{n=k}^{m} A_n$形成一个递增链，$\bigcup_{m=k}^{\infty} B_m = \bigcup_{n=k}^{\infty} A_n$因此，$P(B_m)$，正如我们知道的那样，收敛于1，收敛于$P\left(\bigcup_{n=k}^{\infty} A_n\right)$。

因此这个量也等于1。 □

下面我们讨论强大数法则，其中概率收敛被加强到几乎肯定收敛。但我们已经可以观察到这些改进的一些局限性。借鉴第二个Borel-Cantelli定理，我们给出了一个否定的结果。

定理8.17

假设$X_1, X_2, \ldots$是独立的同分布随机变量，并假设$E(|X_1|) = \infty$(因此，对于所有n，$E(|X_n|) = \infty$)。那么

(i) $P(\{\omega: |X_n(\omega)| \geq n \text{ for infinitely many } n\}) = 1$、

(ii) $P(\lim_{n\to\infty} \frac{S_n}{n}$存在并且是有限的$) = 0$。

证明

(i) 首先

$$E(|X|) = \int_0^{\infty} P(|X| > x)\,dx \quad \text{(根据定理8.11)}$$

$$= \sum_{k=0}^{\infty} \int_k^{k+1} P(|X_1| > x)\,dx \quad \text{(可数可加性)}$$

$$\leq \sum_{k=0}^{\infty} P(|X_1| > k)$$

因为函数$x \mapsto P(|X_1| > x)$ 在$x = k$时，在$[k, k+1]$上达到最大值，因为$\{\omega: |X_1(\omega)| > k\} \supseteq \{\omega: |X_1(\omega)| > x\}$，如果$x \geq k$。根据假设，这个系列是发散的，但是$P(|X_1| > k) = P(|X_k| > k)$ 因为分布是相同的，所以

$$\sum_{k=0}^{\infty} P(|X_k| > k) = \infty。$$

第二个Borel-Cantelli定理是适用的，从而产生要求。

(ii) 用A表示$\frac{S_n}{n}$的极限存在（并且是有限的）n的集合。一些分数的基本代数给出了

$$\frac{ĀĀĀ}{n} - \frac{蛇+1}{n+1} = \frac{(n+1)S_n - nS_{n+1}}{n(n+1)} = \frac{S_n - nX_{n+1}}{n(n+1)} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}。$$

对于任何$\omega_0 \in A$，左手边收敛为零，同时

$$\frac{S_n(\omega_0)}{n(n+1)} \to 0。$$

因此，也$\frac{X_{n+1}(\omega_0)}{n+1} \to 0$。这意味着

$$\omega_0 \notin \{\omega: |X_k(\omega)| > k \text{ for infinitely many k}\} = B、$$

但根据(i)，P（B）=1，因此P（A）=0。 □

8.2.4 强大的大数法则

我们将考虑几个版本的强大数法则，首先对序列(X_n)的矩施加额外的条件，然后逐渐放宽这些条件，我们得出定理8.21，它提供了最一般的积极结果。

第一个结果是由冯-诺伊曼提出的。请注意，我们没有规定X_n具有相同的分布条件。我们付出的代价是不得不假设高阶矩是有限的。然而，对于许多熟悉的随机变量，例如高斯，这并不是一个严重的限制。

定理8.18

假设随机变量X_n是独立的，$E(X_n) = m$，并且$E(X_n^4) \leq K$.则

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k \to m \text{ a.s.}$$

证明

通过考虑$X_n - m$，我们可以假设$E(X_n)=0$，对于所有的n。这简化了

以下为计算结果

$$E(S_n^4)=E\left(\sum_{k=1}^{n}X_k\right)^4$$

$$=E\left(\sum_{k=1}^{n}X_k^4+\sum_{i\neq j}X_i^2X_j^2+\sum_{i\neq j}X_iX_jX_k^2+\sum_{i,j,k,l\text{不同}}X_iX_jX_kX_l\right)。$$

最后两个条款因独立而消失：

$$e\left(\sum_{i\neq j}x_ix_jx_k^2\right)=\sum_{i\neq j}e(x_ix_jx_k^2)=\sum_{i\neq j}e(x_i)e(x_j)e(x_k^2)=0$$

同样，对于所有指数都不同的术语，也是如此。

$$E\left(\sum X_iX_jX_kX_l\right)=\sum E(X_iX_jX_kX_l)$$

$$=\sum e(x_i)e(x_j)e(x_k)e(x_l)=0。$$

第一个项很容易通过假设来估计

$$E\left(\sum X_k^4\right)=\sum E(X_k^4)\leq nK。$$

对于剩下的项，我们首先应用施瓦兹不等式

$$E\left(\sum_{i\neq j}X_i^2X_j^2\right)=\sum_{i\neq j}E(X_i^2X_j^2)\leq\sum_{i\neq j}\sqrt{E(X_i^4)}\sqrt{E(X_j^4)}\leq NK$$

其中N是这种成分的数量。(我们可以通过利用独立性来做得更好，但这样我们就必须用第四个时刻来估计第二时刻，而这将归结为相同的结果。)

要找到N，首先要注意，两个不同的指数对可以选择在$\binom{n}{2}=\frac{n(n-1)}{2}$方式。在固定的$i$，$j$条件下，$X_i^2X_j^2$出现了6种相关方式。对应于2对2个指数的可能排列：（i，i，j，j），（i，j，i，j），（i，j，j，i），（j，j，i，i），（j，i，j，i），（j，i，j）。所以$N=3n(n-1)$，我们有

$$E(S_n^4)\leq K(n+3n(n-1))=K(n+3n^2-3n)\leq 3Kn^2。$$

根据切比雪夫不等式

$$P\left(\left|\frac{S_n}{n}\right|\geq\varepsilon\right)=P(|S_n|\geq n\varepsilon)\leq\frac{E(S_n)^4}{(n\varepsilon)^4}\leq\frac{3K}{\varepsilon^4}\cdot\frac{1}{n^2}.$$

数列$\sum\frac{1}{n^2}$收敛，根据Borel-Cantelli，集合lim sup A_n，$A_n=$

$\{\omega : |\frac{S_n}{n}| \geq \varepsilon\}$的度量为零。它的补充是全度量的集合，我们

需要在其上，序列$\frac{S_n}{n}$收敛到0。为了看到这一点，让$\omega \notin \limsup A_n$ 这意味着ω在有限多的A_n。因此对于某个n_0，所有$n \geq n_0$、$\omega \notin A_n$，即$\frac{S_n}{n} < \varepsilon$（如之前所观察到的）。而这正是我们需要的。为有关的收敛性。 □

下一个定律将只需要2阶的有限时刻，甚至不一定是均匀有界的。

我们在它之前有一个由Kolmogorov提出的辅助但关键的不等式。它比切比雪夫不等式提供了一个更好的估计。后者说那

$$P(|S_n| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}.$$

在下面的定理中，左手边比较大，因此结果比较强。

定理8.19

如果$X_1, \ldots, X_n$ 是独立的，期望值为0，变差有限，那么对于任何$\varepsilon > 0$

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}$$

其中$S_n = X_1 + \cdots + X_n$。

证明

We fix an $\varepsilon > 0$ and describe the first instance that $|S_k|$ exceeds ε. Namely, we write

$$\phi_k = \begin{cases} 1 & \text{如果} |S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon \\ 0 & \text{如果所有} |S_i| < \varepsilon。\end{cases}$$

对于任何一个ω，$j_k(\omega)$中最多有一个数字可能是1，其余的是0，因此它们的总和不是0就是1。显然

$$\sum_{k=1}^{n} \phi_k = 0 \quad \Leftrightarrow \quad \max_{1\leq k\leq n} |S_k| < \varepsilon、$$

$$\sum_{k=1}^{n} \phi_k = 1 \quad \Leftrightarrow \quad \max_{1\leq k\leq n} |S_k| \geq \varepsilon。$$

因此

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) = P(\sum_{k=1}^{n} \phi_k = 1) = E(\sum_{k=1}^{n} \phi_k)$$

因为期望值是一个指标函数的积分：

$$E\Big(\sum_{k=1}^{n}\phi_k\Big)=\int_{\{\omega:\sum_{k=1}^{n}\phi_k(\omega)=0\}}0\,dP+\int_{\{\omega:\sum_{k=1}^{n}\phi_k(\omega)=1\}}1\,dP。$$

因此，剩下的就是证明

$$E\Big(\sum_{k=1}^{n}\phi_k\Big)\le\frac{1}{\varepsilon^2}\mathrm{Var}(S_n)=\frac{1}{\varepsilon^2}E(S_n^2)、$$

因为$E(S_n)=0$。我们从下面估计$E(S_n^2)$。

$$\begin{aligned}
E(S_n^2) &\ge E\Big(\sum_{k=1}^{n}\phi_k S_n^2\Big) \quad (\text{因为}\sum_{k=1}^{n}\phi_k\le 1)\\
&= E\Big(\sum_{k=1}^{n}[S_k^2+2S_k(S_n-S_k)+(S_n-S_k)^2]j_k\Big) \quad （\text{简单代数}）\\
&\ge E\Big(\sum_{k=1}^{n}[S_k^2+2S_k(S_n-S_k)]j_k\Big) \quad (\text{非负数项删除})\\
&= E\Big(\sum_{k=1}^{n}S_k^2\phi_k\Big)+2E\Big(\sum_{k=1}^{n}S_k(S_n-S_k)\phi_k\Big)。
\end{aligned}$$

我们表明，最后一个期望值等于0。请注意，j_k是随机变量$X_1,\ldots,X_k$，所以它独立于$X_{k+1},\ldots,X_n$，而且S_k也独立于$X_{k+1},\ldots$……由于同样的原因，S_n 也独立于 X，...，X。我们计算一个成分
这最后一笔钱的：

$$\begin{aligned}
E(S_k(S_n-S_k)j_k) &= E\Big(S_k\phi_k\Big(\sum_{i=k+1}^{n}X_i\Big)\Big) \quad (\text{根据}S_n\text{的定义})\\
&= E(S_k\phi_k)E\Big(\sum_{i=k+1}^{n}X_i\Big) \quad （\text{通过独立性}）\\
&= 0 \quad (\text{因为}E(X_i)=0\text{，对所有}i\text{而言})。
\end{aligned}$$

在剩下的和中注意到，对于每个$k\le n$，$jS_k^2\ge j\varepsilon_k^2$（这是$0\ge0$，如果$\phi_k=0$，如果$\phi_k=1$，则$S_k^2\ge\varepsilon^2$，均为真），因此

$$E\Big(\sum_{k=1}^{n}S_k^2\phi_k\Big)\ge E\Big(\varepsilon^2\sum_{k=1}^{n}\phi_k\Big)=\varepsilon^2E\Big(\sum_{k=1}^{n}\phi_k\Big)$$

这就得到了所需的不等式。 □

定理8.20

假设$X_1, X_2, \ldots$是独立的，$E(X_n) = 0$，并且

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathrm{Var}(X_n) < \infty.$$

那么

$$\frac{S_n}{n} \to 0 \quad 几乎可以肯定。$$

证明

我们引入辅助性的随机变量

$$Y_m = 最大_{k \le 2^m} |S_k|$$

而对于$2^{m-1} \le n \le 2^m$

$$\left|\frac{S_n}{n}\right| \le \frac{1}{n} 最大_{k \le 2^m} |S_k| \le \frac{1}{2^{m-1}} Y_m$$

只需证明：$\frac{Y_m}{2^m} \to 0$几乎是肯定的，由Lemma 8.2可知足以证明对于每个$\varepsilon>0$

$$\sum_{m=1}^{\infty} P\left(\left|\frac{Y_m}{2^m}\right| \ge \varepsilon\right) < \infty.$$

首先取一个单项$P(|Y_{2^m}| \ge 2^m \varepsilon)$，并通过Kolmogorov的in- equal（定理8.19）来估计它

$$P(|Y_m| \ge 2^m \varepsilon) \le \frac{曲线(S_{2^m})}{\varepsilon^2 2^{2m}}.$$

这个问题可以简化为证明

$$\sum_{m=1}^{\infty} \frac{1}{4^m} \mathrm{Var}(S_{2^m}) <$$

我们重新排列组件

$$\sum_{m=1}^{\infty}\frac{1}{4^m}\mathrm{Var}(S_{2^m})=\sum_{m=1}^{\infty}\frac{1}{4^m}\sum_{k=1}^{2^m}\mathrm{Var}(X_k)$$

$$=\mathrm{Var}(X_1)\sum_{m=1}^{\infty}\frac{1}{4^m}+\mathrm{Var}(X_2)\sum_{m=1}^{\infty}\frac{1}{4^m}$$

$$+\mathrm{Var}(X_3)\sum_{m=2}^{\infty}\frac{1}{4^m}+\mathrm{Var}(X_4)\sum_{m=2}^{\infty}\frac{1}{4^m}$$

$$+\mathrm{Var}(X_5)\sum_{m=3}^{\infty}\frac{1}{4^m}+\cdots+\mathrm{Var}(X_8)\sum_{m=3}^{\infty}\frac{1}{4^m}+\cdots$$

因为$\mathrm{Var}(X_1)$, $\mathrm{Var}(X_2)$出现在m系列的所有分量中($1, 2\leq 2^1$), $\mathrm{Var}(X_3)$, $\mathrm{Var}(X_4)$出现在除第一个之外的所有分量中($2^1<3, 4\leq 2^2$), $\mathrm{Var}(X_5), \ldots$, $\mathrm{Var}(X_8)$出现在除前两个以外的所有地方（$2^2<5, 6, 7, 8$ 2^3），以此类推。我们得出的系列是

$$\sum_{k=1}^{\infty}\mathrm{Var}(X_k)a_k$$

其中

$$a_k=\sum_{\{m:2^m>k\}}\frac{1}{4^m}.$$

这是一个几何数列，比率为$\frac{1}{4}$，第一项$\frac{1}{4^j}$，其中j是最小的
如果我们用k代替2^j，我们通过增加更多的项来增加总和，那么第一个项就是$\frac{1}{2k}$：

$$a_k\leq\frac{\frac{1}{2k}}{1-\frac{1}{4}}$$

并通过假设

$$\sum_{k=1}^{\infty}\mathrm{Var}(X_k)a_k\leq\frac{4}{3}\sum_{k=1}^{\infty}\mathrm{Var}(X_k)\frac{1}{2k}<\infty$$

这就完成了证明。 □

最后，我们进一步放宽对时刻的条件；同时我们需要施加随机变量是同分布的假设。

定理8.21

假设$X_1, X_2, \ldots$ 是独立的同分布，$E(X_1) = m < \infty$.那么

$$\frac{S_n}{n} \to m \text{几乎是肯定的。}$$

证明

我们的想法是使用以前的定理，在那里我们需要有限的变异。由于我们在这里没有这个条件，所以我们截断X_n

$$Y_n = X_n(n) = X_n 1\{\omega : |X_n(\omega)| \leq n\}。$$

截断的随机变量具有有限的变异性，因为每个变量都是有界的：$|Y_n| \leq n$（如果X_n 敢于上跨n级，则右手边被迫为零）。如果$X_n > n$，则新的变量与原来的变量不同。然而，正如下面的论证所示，这种情况不可能经常发生。首先

$$\begin{aligned}
\sum_{n=1}^{\infty} P(Y_n \neq X_n) &= \sum_{n=1}^{\infty} P(|X_n| > n) \\
&= \sum_{n=1}^{\infty} P(|X_1| > n) \text{ (分布是相同的)} \\
&\leq \sum_{n=1}^{\infty} \int_{n-1}^{n} P(|X_1| > x)\, dx \text{ (因为} P(|X_1| > x) \geq P(|X_1| > n)) \\
&= \int_0^{\infty} P(|X_1| > x)\, dx \\
&= E(|X_1|) \text{ (由Lemma 8.11决定)} \\
&< \infty.
\end{aligned}$$

所以根据Borel-Cantelli，在概率为1的情况下，只有有限的事件$X_n \neq Y_n$ 发生，所以换句话说，有一个集合Ω'，$P(\Omega') = 1$，这样对于$\omega \in \Omega'$，$X_n(\omega) = Y_n(\omega)$，除了有限的n个以外。所以在$\Omega'$，如果$\frac{Y_1 + \cdots + Y_n}{n}$收敛到某个极限，那么对于$\frac{S_n}{n}$也是如此。

为了使用前面的定理，我们必须证明系列的收敛性

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2}$$

但由于$\mathrm{Var}(Y_n) = E(Y_n^2) - (EY_n)^2 \le EY_n^2$，所以只需显示收敛性。的重要性。

$$\sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2}。$$

为此，首先要注意

$$E(Y_n^2) = \int_0^{\infty} 2xP(|Y_n|>x)\,dx \quad (根据定理8.11)$$

$$= \int_0^n 2xP(|Y_n|>x)\,dx \quad (因为 P(|Y_n|>n)=0)$$

$$= \int_0^n 2xP(|X_n|>x)\,dx \quad (如果 |Y_n| \le n 然后 Y_n = X_n)$$

$$= \int_0^n 2xP(|X_1|>x)\,dx \quad (相同的分布)。$$

下一页

$$\sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2}\int_0^n 2xP(|X_1|>x)\,dx$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^2}\int_0^{\infty} 2x1_{[0,n)}(x)P(|X_1|>x)\,dx$$

$$= 2\int_0^{\infty} \sum_{n=1}^{\infty} \frac{1}{n^2} x1_{[0,n)}(x)P(|X_1|>x)\,dx。$$

我们研究一下函数$x \mapsto \sum_{n=1}^{\infty} \frac{1}{n^2} x1_{[0,n)}(x)$.

如果$0 \le x \le 1$，那么数列中的任何一个项都不会被杀死，并且

$$\sum_{n=1}^{\infty} \frac{1}{n^2} x1_{[0,n)}(x) \le \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} < 2$$

正如众所周知的那样。

如果$x>1$，那么我们只有n上的和$\ge x$.让$m=\mathrm{Int}(x)$（x的整数部分）。我们有

$$\sum_{n=1}^{\infty} \frac{1}{n^2} x1_{[0,n)}(x) = x\sum_{n=m+1}^{\infty} \frac{1}{n^2} \le x\int_m^{\infty} \frac{1}{x^2}\,dx = x\frac{1}{m} \le 2.$$

$$_{n=m+1} \quad ^{n2} \qquad _m \quad ^{x2} \qquad m$$

在每一种情况下，有关的函数都是以2为界限的，所以

$$\sum^{\infty} \frac{E(Y_n)^2}{n2} \leq 4 \int_0^{\infty} P(|X_1| > x)\, dx = 4E(|X_1|) < \infty。$$

考虑$Y_n - EY_n$，应用前面的定理，得到

$$\frac{1}{n}\sum_{k=1}^{n}(y_k - ey_k) \to 0 \qquad \text{几乎肯定。}$$

我们有

$$E(Y_k) = E(X_k 1_{\{\omega:|X_k|\le k\}}) = E(X_1 1_{\{\omega:|X_1(\omega)|\le k\}}) \to m$$

因为$X_1 1_{\{\omega:|X_1(\omega)|\le k\}} \to X_1$，该序列被$X_1$所支配，该序列是可整数的。因此，根据三角形不等式，我们有

$$\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - m\right| \le \left|\frac{1}{n}\sum_{k=1}^{n}(Y_k - ey_k)\right| + \left|\frac{1}{n}\sum_{k=1}^{n} ey_k - m\right| \to 0.$$

如前所述，这意味着$\frac{S_n}{n}$几乎肯定会收敛。 □

8.2.5 弱的 收敛性

为了推导出中心极限定理，我们首先需要研究随机变量序列(X_n)分布的一致性。

定义8.4

^{n}R上的Borel概率度量的序列P_n *弱地*收敛于P，当且仅当它们的累积分布函数F_n在F是连续的所有点上收敛于P的分布函数F。如果$P_n = P_{X_n}$, $P = P_X$是一些随机变量的分布，那么我们说，该序列
(X_n) *弱地*收敛于X。

命名为 "弱 "是有道理的，因为这种收敛是由我们迄今为止遇到的最弱的收敛，即概率上的收敛所暗示的。

定理8.22

如果X_n在概率上收敛于X，那么X_n的分布就会弱地收敛。

证明

让$F(y)=P(X\le y)$。固定y为F的一个连续点，$\varepsilon>0$，目标是得到

$$F(y)-\varepsilon<F_n(y)<F(y)+\varepsilon$$

根据F的连续性，我们可以找到$\delta>0$，从而使

$$P\ (X\leq y) - \frac{\varepsilon}{2} < P\ (X\leq y-\delta)\ 、\quad P\,(X\leq y+\delta) < F\,(y) + \frac{\varepsilon}{2}。$$

通过概率的收敛、

$$P\,(|X_n - X| > \delta) < \frac{\varepsilon}{2}.$$

显然，如果$X_n \leq y$并且$|X_n - X| < \delta$，那么$X < y+\delta$

，所以

$$P\,((X_n \leq y) \cap (|X_n - X| < \delta)) \leq P\,(X < y+\delta).$$

我们可以从下面估计出左手边的情况：

$$P\,(X_n \leq y) - \frac{\varepsilon}{2} < P\,((X_n \leq y) \cap (|X_n - X| < \delta))。$$

将所有这些放在一起，我们得到

$$P\ (X_n \leq y)\ < P\ (X \leq y)\ + \varepsilon$$

并让$\varepsilon \to 0$，我们就实现了一半的目标。另一半则以类似方式获得。□

然而事实证明，在某种意义上的弱收敛性几乎肯定意味着收敛性。我们对"在某种意义上 "的意思在下一个定理中作了解释：它也给[0, 1]中的Borel集和Lebesgue度量一个核心作用。

定理8.23（斯科罗霍德表征定理）

如果P_n弱地收敛于P，那么存在X_n，X，定义在概率（[0，1]，B，$m_{[0,1]}$）上的随机变量，这样$P_{X_n} = P_n$，$P_X = P$，$X_n \to X$ a.s.

证明

取X^+，X^-，X_n^+，X_n^- 对应于F_n，F，的分布函数。

P_n，P，如定理4.31中。我们在那里表明，$F_{X^+} = F_{X^-} = F$，这意味着$P\ (X^+ = X^-)\ = 1$。固定一个ω，使得$X^+\ (\omega) = X^-\ (\omega)$。让$y$是$F$的一个连续性点，使$y>X^+\ (\omega)$。那么$F\,(y) > \omega$，并且，根据弱收敛性，对于足够大的$n$，我们有$F_n\,(y) > \omega$。然后，根据结构，$X^+\,(\omega) \leq y$。这个不等式除了有限的$n$之外，对所有的$n$都是成立的，所以如果我

们在左边取上限的话，它就被保留了：

$$\limsup X^+_n(\omega) \le y_o$$

取一个F的连续性点的序列y_k，从上面开始收敛到$X^+(\omega)$（一个单调函数的不连续性点的集合最多可以计数）。

对于$y = y_k$考虑上述不等式，并通过k的极限得到

$$\limsup X^+_n(\omega) \le X^+(\omega)。$$

同样如

$$\liminf X_n^-(\omega) \ge X^-(\omega)$$

此

$$X^-(\omega) \le \liminf X_n^-(\omega) \le \limsup_n X^+(\omega) \le X^+(\omega)。$$

极端值是相等的，所以收敛性也是相等的。 □

斯科罗霍德定理是概率学的一个重要工具。我们只需要在下面的结果中用到它，它将分布的收敛性与相关特征函数的收敛性联系起来。

定理8.24

如果P_{X_n}弱地收敛于P_X，那么$j_{X_n} \to j_X$。

证明

取斯科罗克霍德表示Y_n, Y的度量P_{X_n}, P_X。Y_n几乎肯定地收敛于Y，这意味着根据支配收敛定理，$E(e^{itY_n}) \to E(e^{itY})$。但是$X_n$, X的分布与Y的分布$_n$，Y，所以特征函数是相同的。 □

定理8.25（赫尔利定理）

让F_n是一些概率度量的分布函数的序列。存在F，一个度量的分布函数（不一定是proba-bility），以及一个序列k_n，使得$F_{k_n}(x) \to F(x)$在F的连续性点上。

证明

将有理数排列成一个序列：$Q = \{q_1, q_2, ..\}$。序列$F_n(q_1)$是有界的（分布

函数的值位于[0,1]），因此它有一个收敛的子序列、

$$F_{k_n^1}(q_1) \to y_1 \text{。}$$

接下来考虑序列$F_{k_n^1}(q_2)$，它也是有界的，所以对于k_n^2的子序列k_n^1，我们有收敛性

$$F_{k_n^2}(q_2) \to y_2。$$

当然还有

$$F_{k_n^2}(q_1) \to y_1。$$

以这种方式进行，我们发现$k_n^3, k_n^4, \ldots$，每项都是一个子序列的前一个，有

$$F_{k_n^3}(q_m) \to y_m \text{，} m \le 3$$

$$F_{k_n^4}(q_m) \to y_m \text{，} m \le 4、$$

以此类推。对角线序列$F_{k_n} = F_{k_n^n}$在所有的理性点上都收敛。我们在Q上定义F_Q：

$$F_Q(q) = \lim F_{k_n}(q)$$

接下来我们写出

$$F(x) = inf\{F_Q(q) : q \in Q, q > x\}.$$

我们表明，F是非递减的。由于F_n是非递减的，F_Q也是如此（$q_1 < q_2$意味着$F_{k_n}(q_1) \le F_{k_n}(q_2)$，这在极限中仍然是真的）。现在让$x_1 < x_2$。$F(x_1) \le F_Q(q)$ for all $q > x_1$ hence in particular for all $q > x_2$, so $F(x_1) \le \inf_{q > x_2} F_Q(q) = F(x_2)$.

我们表明，F是右连续的。设$x_n \searrow x$。根据F的单调性，$F(x) \le F(x_n)$因此$F(x) \le \lim F(x_n)$。假设$F(x) < \lim F(x_n)$。根据F的定义，有$q \in Q$，$x < q$，使得$F_Q(q) < \lim F(x_n)$。对于某些n_0，$x \le x_{n_0} < q$，因此$F(x_{n_0}) \le F_Q(q)$同样根据F的定义，因此$F(x_{n_0}) < \lim F(x_n)$，这是一个矛盾。

最后，我们表明，如果F在x处是连续的，那么$F_{k_n}(x) \to F(x)$。让$\varepsilon > 0$是任意的，并找到有理数$q_1 < q_2 < x < q_3$，使得

$$F(x) - \varepsilon < F(q_1) \le F(x) \le F(q_3) < F(x) + \varepsilon。$$

由于$F_{k_n}(q_2) \to F_Q(q_2) \ge F(q_1)$，对于足够大的$n$

$$F(x) - \varepsilon < F_{k_n}(q_2)。$$

但F_{k_n}是不递减的，所以

$$F_{k_n}(q_2) \le F_{k_n}(x) \le F_{k_n}(q_3)。$$

最后，$F_{k_n}(q_3) \to F_Q(q_3) \ge F(q_3)$，所以对于足够大的$n$

$$F_{k_n}(q_3) < F(x) + \varepsilon。$$

将上述三个不等式放在一起，我们可以得到

$$F(x) - \varepsilon < F_k \text{n}(x) < F(x) + \varepsilon$$

这证明了收敛性。 □

备注8.4

极限分布函数不需要对应于概率测量。例子：$F_n = 1_{[n,\infty)}$ ， F_n 0，所以 $F = 0$。这是一个分布函数（非递减，右连续），相应的度量满足P（A）= 0

然后我们非正式地说，质量逃到了无穷大。为了防止这种情况发生，我们引入了以下概念。

定义8.5

$_n$我们说，如果对于每一个ε>0，有M，使得P_n ($\mathbb{R}^d$ \ [-M, M]) < ε，对于所有n，$\mathbb{R}^d$ 上的概率*序列*是*紧的*。

通过R中的区间^n ，我们理解区间的乘积：[-M, M] = $\{x = (x_1 , . . . , x)_n \in \mathbb{R}^n : x_i \in -[M, M]$ 所有$\}$ i。重要的是，M ε的选择对所有n*都*是好的--不等式是均匀的。很容易找到这样一个$M = M_n$ ，对每一个P_n 分别。这源于这样一个事实，即 P_n ([-M, M]) → 1，因为$M \to \infty$。

定理8.26（普罗霍罗夫定理）

如果一个序列P_n 是紧的，那么它就有一个子序列弱地收敛于某个概率测量P。

证明

根据赫尔利定理，一个子序列F_k n收敛到某个分布函数F。我们要做的就是证明F对应于某个概率度量P，这意味着我们必须证明F （∞） =1（即 $\lim_{y\to\infty} F$（y） =1）。固定ε>0，并找到一个连续性点，使得$F_n (y) = P_n$ ((-∞, y]) > 1 - ε，对于

所有的n（从紧密度的定义中找到M，并取F的一个大于M的连续点）。因此$\lim_{n\to\infty} F_k$ n (y) ≥ 1 - ε，但这个极限是F (y)。这证明了$\lim_{y\to\infty} F$ (y) = 1。□

我们需要扩展特征函数的概念。

定义8.6

我们说j是R上博雷尔测量P的特征函数，如果

$j(t) = \int e^{itx} \, dP(x)$.

在$P = P_X$的情况下，我们显然有$\phi_P = \phi_X$，所以这两个定义是一致的。

定理8.27

假设(P_n)是紧的，让P是普罗霍罗夫定理提供的(P_n)的一个子序列的极限。如果$j_n(u) \to j(u)$ 其中j_n 是P_n 的特征函数，j是P的特征函数，那么$P_n \to P$弱。

证明

固定F的一个连续性点。对于每一个子序列F_k n，都有一个子序列(subsubsequence)l_k n，为简洁起见，l_n，使得F_l n收敛于某个函数H(海利定理)。用P' 表示相应的度量。因此j_l n $\to j_{P'}$、
但另一方面，$j\ n_l$ $\to \phi$.因此，$j_{P'} = j$，因此$P' = P$
(推论6.18)。以上是序列收敛的充分条件
$F_n(y)$。 □

8.2.6 中心极限 定理

下面这条定理在下文中很有用。它显示了如何估计概率测量的 "尾部 " 的 "重量"，并将成为证明紧密度的有用工具。

定理8.28

如果j是P的特征函数，那么

$$p\,(\mathbb{R} \setminus [-m, m]) \leq 7m \int_0^{\frac{1}{m}} [1-\Box j(u)]\, du.$$

证明

$$M\int_0^{\frac{1}{M}}[1-\square j(u)]\,du = M\int_0^{\frac{1}{M}}[1-\square\int_{\mathbb{R}} e^{ixu}\,dP(x)]\,du$$
$$= M\int_0^{\frac{1}{M}}[1-\int_{\mathbb{R}}\cos(xu)\,dP(x)]\,du$$
$$= \int_{\mathbb{R}} M\int_0^{\frac{1}{M}}[1-\cos(xu)]\,du\,dP(x) \quad \text{(福比尼著)}$$
$$= \int_{\mathbb{R}}(1-\frac{\sin(\frac{x}{M})}{\frac{x}{M}})\,d(P\,x)$$
$$\geq \int_{|t|\geq 1}(1-\frac{\sin(\frac{x}{M})}{\frac{x}{M}})\,dP(x)$$
$$\geq \inf_{|t|\geq 1}(1-\frac{\sin t}{t})\int_{|\frac{x}{M}|>1} dP(x)$$
$$\geq \frac{1}{7}\,p(r\backslash[-m, m])$$

因为 $1-\frac{\sin t}{t}\geq 1-\sin 1\geq \frac{1}{7}$。 □

定理8.29（利维定理）

让ϕ_n 是P_n 的特征函数。假设ϕ_n ϕ是在0处连续的函数。那么j是一个度量P的特征函数，$P_n \to P$弱。

证明

只需证明P_n 是紧的（那么P_k n → P弱，对于某些P和k_n ，ϕ n_k ϕ_P ，$\phi = \phi_P$ ，根据前面的定理我们就完成了）。

应用定理8.28，我们有

$$p_n(r\backslash[-m, m]) \leq 7m\int_0^{\frac{1}{M}}[1-\square j_n(u)]\,du.$$

由于$j_n \to \varphi$,$|j_n|\leq 1$，对于固定的M，这个上界是可整数的，所以根据支配收敛定理

$$7M\int_0^{\frac{1}{M}}[1-\square j_n(u)]\,du \to 7M\int_0^{\frac{1}{M}}[1-\square j(u)]\,du.$$

如果$M \to \infty$，那么

$$\frac{1}{7M}\int_0^{\frac{1}{M}}[1-\varphi(u)]\,du \le 7M\cdot\frac{1}{M}\sup_{[0,\frac{1}{M}]}|1-\varphi(u)| \to 0$$

通过φ在0处的连续性（记得$\varphi(0)=1$）。现在让$\varepsilon>0$，并找到M_0，以便即 $7\sup_{[0,\frac{1}{M_0}]}|1-\varphi(u)| < \frac{\varepsilon}{2}$，且 n_0，以便 $n \ge n_0$

$$\left|\int_0^{\frac{1}{M_0}}[1-\varphi_n(u)]\,du - \int_0^{\frac{1}{M_0}}[1-\varphi(u)]\,du\right| < \frac{\varepsilon}{2}$$

因此

$$P_n(\mathbb{R}\setminus[-M_0, M_0]) < \varepsilon$$

对于$n \ge n_0$。现在对于每个$n = 1, 2, \ldots, n_0$找到M_n，使$P_n([-M_n, M_n]) > 1-\varepsilon$，让$M = \max\{M_0, M_1, \ldots, M_{n_0}\}$。当然，由于$M \ge M_k$，$P_n([-M, M]) \ge P_n([-M_k, M_k]) > 1-\varepsilon$ 对于每个n，这证明了P_n 的紧密度。 □

对于一个序列X_k，$m_k = E(X_k)$, $\sigma_k^2 = \mathrm{Var}(X_k)$ 有限，让$S_n = X_1 + \cdots + X_n$ 像往常一样，考虑归一化的随机变量

$$T_n = \frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}}。$$

显然，$E(T_n)=0$，$\mathrm{Var}(T_n)=1$（由$\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$）。写下$c_n^2 = \mathrm{Var}(S_n)$（如果$X_n$ 是独立的，那么正如我们已经知道的，$c_n^2 = \sum_{k=1}^n \sigma_k^2$）。我们

我们证明一个条件，在这个条件下，T的分布序列，收敛到

标准的高斯度量G（其密度为 $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$）：

$$\frac{1}{c_n^2}\sum_{k=1}^n\int_{\{x:|x-m_k|\ge\varepsilon c_n\}}(x-m_k)^2\,dP_{X_k}(x) \to 0，因为n \to \infty。\quad (8.4)$$

特别是，如果X_n 的分布是相同的，$m_k = m$, $\sigma_k = \sigma$，那么这个条件就得到满足。要看到这一点，请注意，假设独立，我们有$c_n^2 = n\sigma^2$，然后

$$\int_{\{x:|x-m|\ge\varepsilon\sigma\sqrt{n}\}}(x-m)^2\,dP_{X_k}(x) = \int_{\{x:|x-m|\ge\varepsilon\sigma\sqrt{n}\}}(x-m)^2\,dP_{X_1}(x)$$

因此

$$\frac{1}{n\sigma^2}\sum_{k=1}^n\int_{\{x:|x-m|\ge\varepsilon\sigma\sqrt{n}\}}(x-m)^2\,dP_{X_k}(x) = \frac{1}{}$$

$$\frac{1}{\sigma^2}\int_{\{x:|x-m|\geq\varepsilon\sigma\sqrt{n}\}}(x-m)^2\,dP_X(x)\to 0$$

为$n \to \infty$因为集合$\{x : |x - m| \geq \varepsilon\sigma\sqrt{n}\}$递减到∅。

我们已经为概率论的主要定理做好了准备。证明是相当技术性和高级的，初读时可以省略。

定理8.30（林德伯格-费勒定理）

让X_n是独立的，具有有限的期望和变异。如果条件（8.4）成立，那么$P_{T_n} \to G$弱。

证明

首先假设$m_k = 0$。只需证明特征函数ϕ_{T_n}收敛于G的特征函数，即证明

$$\phi_{T_n}(u) \to e^{-\frac{1}{2}u^2}.$$

我们计算出

$$\begin{aligned}
\phi_{T_n}(u) &= E(e^{iuT_n}) \text{ (根据}\phi_{T_n}\text{的定义)} \\
&= E(e^{i\frac{u}{c_n}\sum_{k=1}^{n} X_k}) \\
&= E(\prod_{k=1}^{n} e^{i\frac{u}{c_n}X_k}) \\
&= \prod_{k=1}^{n} E(e^{i\frac{u}{c_n}X_k}) \text{ (通过独立性)} \\
&= \prod_{k=1}^{n} \phi_{X_k}\left(\frac{u}{c_n}\right) \text{（根据}\phi_{X_k}\text{的定义）}.
\end{aligned}$$

我们需要证明的是

$$\log \prod_{k=1}^{n} \phi_{X_k}\left(\frac{u}{c_n}\right) \to -\frac{1}{2}u^2。$$

我们将使用以下公式（复变的泰勒公式的特殊情况）。

$$\log(1+z) = z + \vartheta_1 |z|^2 \quad \text{对于一些}\vartheta_1，|\vartheta_1| \leq 1、$$

$$e^{iy} = 1 + iy + \frac{1}{2}\vartheta_2 y^2 \quad \text{对于某些}\vartheta_2，|\vartheta_2| \leq 1、$$

$$e^{iy} = 1 + iy - \frac{1}{2}y^2 + \frac{1}{6}\vartheta_3 |y|^3 \quad \text{对于某些}\vartheta_3，|\vartheta_3| \leq 1$$

所以对于固定的$\varepsilon>0$

$$\phi_{X_k}(u)=\int_{|x|\geq\varepsilon c_n} e^{iux}\,dP_{X_k}(x)+\int_{|x|<\varepsilon c_n} e^{iux}\,dP_{X_k}(x)$$

$$=\int_{|x|\geq\varepsilon c_n}\left(1+iux+\frac{1}{2}\vartheta_2 u^2x^2\right)dP_{X_k}(x)$$

$$+\int_{|x|<\varepsilon c_n}\left(1+iux-\frac{1}{2}u^2x^2+\frac{1}{6}\vartheta_3|u|^3|x|^3\right)dP_{X_k}(x)$$

$$=1+\frac{1}{2}u^2\int_{|x|\geq\varepsilon c_n}\vartheta_2 x^2\,dP_{X_k}(x)-\frac{1}{2}u^2\int_{|x|<\varepsilon c_n}x^2\,dP_{X_k}(x)$$

$$+\frac{1}{6}|u|^3\int_{|x|<\varepsilon c_n}\vartheta_3|x|^3\,dP_{X_k}(x)$$

因为$\int x\,dP_{X_k}(x)=0$（这就是$E(X_k)$）。为清楚起见，我们引入以下符号

$$\alpha_{nk}=\int_{|x|\geq\varepsilon c_n}x^2\,dP_{X_k}(x)、$$

$$\beta_{nk}=\int_{|x|<\varepsilon c_n}x^2\,dP_{X_k}(x).$$

请注意，$\beta_{nk}\leq\varepsilon^2 c_n^2$，因为在我们取得积分的集合上我们有$x^2<\varepsilon^2 c_n^2$，我们对一个概率度量进行积分。

条件（8.4）现在的形式是

$$\sum_{k=1}^{n}\frac{1}{c_n^2}\alpha_{nk}\to 0，因为n\to\infty。\tag{8.5}$$

自

$$\sum_{k=1}^{n}(\alpha_{nk}+\beta_{nk})=\sum_{k=1}^{n}\mathrm{Var}(X_k)=c_n^2$$

我们有

$$\sum_{k=1}^{n}\frac{1}{c_n^2}\beta_{nk}\to 1，因为n\to\infty。\tag{8.6}$$

数字α_{nk}，β_{nk}是正数，所以最后的收敛是单调的（序列是增加的）。上述关系对每个$\varepsilon>0$都成立。

现在我们分析一下ϕ_{X_k}的表达式中的一些条款。由于

$$\left|\int_{|x|\geq\varepsilon c_n}\vartheta_2 x^2\,dP_{X_k}(x)\right|\leq\int_{|x|\geq\varepsilon c_n}x^2\,dP_{X_k}(x)、$$

我们有一个ϑ_2'，$|\vartheta_2'|\leq 1$，从而

$$\int_{|x|\geq\varepsilon c_n}\vartheta_2 x^2\,dP_{X_k}(x)=\vartheta_2'\int_{|x|\geq\varepsilon c_n}x^2\,dP_{X_k}(x)=\vartheta_2'\alpha_{nk}。$$

下一页

$$\left|\int_{|x|<\varepsilon c_n} \vartheta_3 |x|^3 \, dP_{X_k}(x)\right| \le \int_{|x|<\varepsilon c_n} |x|^3 \, dP_{X_k}(x) \le \int_{|x|<\varepsilon c_n} \varepsilon c_n x^2 \, dP_{X_k}(x)$$

(我们用εc代替一个x_n，剩下两个)，因此对于某些θ_3'与$|\theta_3'| \le 1$

$$\left|\int_{|x|<\varepsilon c_n} \vartheta_3 |x|^3 \, dP_{X_k}(x)\right| \le \vartheta_3' \int_{|x|<\varepsilon c_n} \varepsilon c_n x^2 \, dP_{X_k}(x) = \vartheta_3' \varepsilon c_n \beta_{nk}'\text{。}$$

我们将其代入j_X k的表达式，得到

$$\phi_{X_k}(u) = 1 + \frac{1}{2} u \vartheta_2' \alpha_{nk} - \frac{1}{2} u^2 \beta_{nk} + \frac{1}{6} |u|^3 \theta_3' \varepsilon c_n \beta_{nk}\text{。}$$

将u替换为$\frac{u}{c_n}$以获得

$$\phi_{X_k}\left(\frac{u}{c_n}\right) = 1 + \frac{1}{2} u \vartheta_2' \frac{1}{c_n^2} \alpha_{nk} - \frac{1}{2} u^2 \frac{1}{c_n^2} \beta_{nk} + \frac{1}{6} |u|^3 \theta_3' \frac{\varepsilon}{c_n^2} \beta_{nk} = 1 + \gamma_{nk}$$

与

$$\gamma_{nk} = \frac{1}{2} u^2 \vartheta_2' \frac{1}{c_n^2} \alpha_{nk} - \frac{1}{2} u^2 \frac{1}{c_n^2} \beta_{nk} + \frac{1}{6} |u|^3 \theta_3' \frac{\varepsilon}{c_n^2} \beta_{nk}\text{。}$$

关系（8.5），（8.6）得出

$$\sum_{k=1} \gamma_{nk} \to -\frac{1}{2} u^2 + \frac{1}{6} |u|^3 \vartheta_3' \varepsilon\text{。} \tag{8.7}$$

回顾一下，我们真正追求的是

$$\text{对数} \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{u}{c_n}\right) = \sum_{k=1}^{n} \log \phi_{X_k}\left(\frac{u}{c_n}\right) = \sum_{k=1}^{n} (\gamma_{nk} + \vartheta |\gamma_{nk}|^2)$$

其中我们介绍了泰勒的对数公式，所以我们离目标并不遥远。我们所要做的就是证明

$$\left| \text{对数} \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{u}{c_n}\right) + \frac{1}{2} u^2 \right| \to 0$$

当$n \to \infty$时。 因此，让$\delta>0$是任意的，u是固定的。

$$\left| \text{对数} \prod_{k=1}^{n} \phi_{X_k}\left(\frac{u}{c_n}\right) + \frac{1}{2} u^2 \right|$$

$$
\left|\sum_{k=1} \gamma_{nk} + \frac{1}{2}u^2\right| + |\vartheta_1| \left|\sum_{k=1} \gamma_{nk}^2\right|
$$

$$
\le \sum_{k=1}^{n} \gamma_{nk} + \frac{1}{2}u^2 - \frac{1}{6}|u|^3\theta^{3\prime}\varepsilon + \sum_{k=1}^{n} |\gamma_{nk}|^2 + |u|^3\varepsilon|\theta^{3\prime}
$$

根据（8.7），右边的第一项收敛为零，所以它小于$\frac{\delta}{2}$，因为足够大的n，剩下的就是证明

$$\sum_{k=1}^{n} |\gamma_{nk}|^2 + |u|^3 \varepsilon < \frac{\delta}{2}$$

我们选择ε，使其小到$|u|3\varepsilon < \frac{\delta}{4}$。剩下的就是证明

$$\sum_{k=1}^{n} |\gamma_{nk}|^2 < \frac{\delta}{4}$$

对于大的n，我们有一个γ_{nk}的公式，我们知道一些关于$\sum_{k=1}^{n} \gamma_{nk}$的信息因此，我们使用以下技巧

$$\sum_{k=1}^{n} |\gamma_{nk}|^2 \leq \max_{k=1,\dots,n} |\gamma_{nk}| \cdot \sum_{k=1}^{n} |\gamma_{nk}|$$

第一个因素有：

$$\max_{k=1,\dots,n} |\gamma_{nk}| \leq \frac{1}{2} u^2 \text{可能的}\max_{k=1,\dots,n} \left|\frac{\alpha_{nk}}{c_n^2}\right| + \frac{1}{2} u^2 \varepsilon^2 + \frac{1}{6} |u|^3 \varepsilon^3$$

(使用$\beta_{nk} \leq \varepsilon c_n^2$)但

$$\text{最大}_{k=1,\dots,n} \left|\frac{\alpha_{nk}}{c_n^2}\right| \leq \sum_{k=1}^{n} \frac{\alpha_{nk}}{c_n^2} \to 0.$$

第二个因素满足

$$\sum_{k=1}^{n} |\gamma_{nk}| \leq \frac{1}{2} u^2 \sum_{k=1}^{n} \frac{\alpha_{nk}}{c_n^2} + \frac{1}{2} u^2 + \frac{1}{6} |u|^3 \varepsilon$$

其中我们使用的事实是$\sum \frac{\beta_{nk}}{c_n^2} \leq 1$.写为$\sum_{k=1}^{n} \frac{1}{c_n^2} \alpha_{nk} = a_n$，$a_n \to 0$，对于明确我们有，取$\varepsilon \leq 1$、

$$\sum_{k=1}^{n} |\gamma_{nk}|^2 \leq \left(\frac{u^2}{2} a_n + \frac{u^2 \varepsilon^2}{2} + \frac{|u|^3 \varepsilon^3}{6}\right)\left(\frac{u^2}{2} a_n + \frac{u^2}{2} + \frac{|u|^3 \varepsilon}{6}\right) \leq C a_n + D\varepsilon$$

对于一些数字（仅取决于u）C，D，所以最后选择ε，以便$D\varepsilon < \frac{\delta}{8}$，然后取$n_0$，大到$Ca_n < \frac{\delta}{8}$，$n \geq n_0$。 □

作为一个特例，我们立即推导出一个著名的结果：

推论8.31（de Moivre-Laplace定理）

让X_n 是相同分布的独立随机变量，P（X_n = 1） = P （X_n = -1） =$\frac{1}{2}$ 。那么

$$P\left(a < \frac{S_n}{\sqrt{n}} < b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

练习8.8

使用中心极限定理来估计1000次独立抛掷中头像数与500次相差小于2%的概率。

练习8.9

一枚硬币需要抛多少次才能使概率至少达到0.99，平均头数与0.5相差不到1%?

8.2.7 在数学上的应用 金融

我们将证明Black-Scholes模型是一连串适当定义的CRR模型的极限，期权价格也在不断收敛。对于固定的T，回顾一下，在Black-Scholes模型中，股票价格的形式是

$$S(T) = S(0) \exp\{\xi(T)\}$$

其中$\xi(T) = (r - \frac{\sigma^2}{2})T + \sigma w(T)$，$r$是连续com-pounding的无风险利率。随机性只存在于$w(T)$中，它是高斯的，平均值为零，方差为T。

为了构建近似序列，固定n来分解时间段分为n个长度为$\frac{T}{n}$，并写出

$$R_n = \exp\{r\frac{T}{n}\}、$$

这是一个步骤的无风险增长因素。我们构建一个序列$\eta_n(i)$, $i = 1, \ldots, n$ 个独立的、相同分布的随机变量，具有二项分布，因此，二项模型中的价格在n 个步骤之后

$$S_n(T) = S(0) \times \eta_n(1) \times \cdots \times \eta_n(n)$$

趋近于$S(T)$。写下

$$\eta_n(i) = \frac{U_n}{D_n}$$

其中每个值都是以概率$\frac{1}{2}$。我们的任务是找到U_n, D_n、
假设$U_n > D_n$。以下条件

$$r_n = \frac{1}{2}(u_n + d_n) \tag{8.8}$$

保证$S_n(T)$是一个马太效应（见7.4.3节）。我们看一下对数收益：

$$\ln \frac{S_n(T)}{S(0)} = \ln(\eta_n(1) \times \cdots \times \eta_n(n)) = \sum_{i=1}^{n} \ln \eta_n(i).$$

我们希望将中心极限定理应用于序列$\ln \eta_n(i)$，所以我们调整方差。我们希望有

$$\mathrm{Var}(\ln(\eta_n(1) \times \cdots \times \eta_n(n)) = T\sigma^2。$$

在左边，使用独立、

$$\mathrm{Var}(\sum_{i=1}^{n} \ln \eta_n(i)) = \sum_{i=1}^{n} \mathrm{Var}(\ln \eta_n(i)) = nVar(\ln \eta_n(1))。$$

对于二项分布中的$p = \frac{1}{2}$

$$\mathrm{Var}(\ln \eta_n(1)) = \frac{1}{4}(\ln(U_n) - \ln(D_n))^2、$$

所以需要的条件是

$$n\frac{1}{4}(\ln(U_n) - \ln(D_n))^2 = T\sigma^2。$$

由于$U_n > D_n$、

$$\ln(\frac{U_n}{D_n}) = 2\sigma\frac{\sqrt{T}}{\sqrt{n}}、$$

所以最后

$$U_n = D_n \exp\{2\sigma\sqrt{\frac{T}{n}}\}。 \tag{8.9}$$

我们解决系统（8.8），（8.9），得到

$$D_n = e^{r\frac{T}{n}} \frac{2}{1 + e^{2\sigma\sqrt{T/n}}},$$

$$U_n = D_n^{e2}\sigma\sqrt{T}/n。$$

考虑一下预期值

$$E(\ln(\eta_n(1)\times\cdots\times\eta_n(n)) = E(\sum_{i=1}^{n}\ln\eta_n(i)) = nE(\ln\eta_n(1))$$
$$= n\frac{1}{2}(\ln(U_n)+\ln(D_n)) = a_n、$$

说

。

练习8.10

表明

$$a_n \to (r - \frac{1}{2}\sigma^2)T。$$

对于每一个n，我们有一串n个独立同分布的随机变量$\xi_n(i) = \ln\eta_n(i)$形成所谓的*三角阵列*。我们有以下版本的中心极限定理。它可以用与定理8.30完全相同的方式来证明（也见[2]）。

定理8.32

如果$\xi_n(i)$是一个三角形阵列，$\lambda_n = \sum \xi_n(i)$， $E(\lambda_n) \to \mu = (r - \frac{1}{2}\sigma^2)T$，$Var(\lambda_n) \to \sigma^2 T$，那么序列$(\lambda_n)$弱地收敛于一个高斯的Ran-dom变量，平均值为$\mu$，方差*为*$\sigma^2 T$。

正如我们所看到的，定理中所述的条件在上文对U_n和D_n的赋值下是成立的。因此，$\ln S_n(T)$弱地收敛于$\ln S(T)$。二项式模型中的put值由以下公式给出

$$P_n(0) = \exp\{-rT\}E(K - S_n(T))^+ = exp\{-rT\}E(g(\ln S_n(T))$$

其中$g(x)=(K-e^x)^+$是一个有界的连续函数。因此

$$P_n(0) \to \exp\{-rT\}E(g(\ln S(T))$$

正如我们所知（见练习4.20），它给出了Black-Scholes公式。赎回价格的收敛立即形成了在每个模型中普遍存在的认购-认沽奇偶性。

8.3 命题的证明

证明（命题8.4的证明）

显然，d（X，Y）$=d$（Y，X）。如果$d(X,Y)=0$，那么$E(|X-Y|)=0$，因此$X=Y$几乎到处都是，所以在L^1，$X=Y$。三角不等式是由公制$\rho(x,y)=\frac{|x-y|}{1+|x-y|}$的三角不等式。

接下来，假设 $X_n \to X$的概率。让$\varepsilon>0$

$$d(X_n,X)=\int_{|X_n-X|<\frac{\varepsilon}{2}}\frac{|X_n-X|}{1+|X_n-X|}dP+\int_{|X_n-X|\geq\frac{\varepsilon}{2}}\frac{|X_n-X|}{1+|X_n-X|}dP$$
$$\leq\frac{\varepsilon}{2}+P\left(|X_n-X|\geq\frac{\varepsilon}{2}\right)$$

因为第一项的积分小于ε（用1来估计分母），我们用1来估计第二项的积分。对于足够大的n，第二项小于$\frac{\varepsilon}{2}$。

反过来说，让$E_{\varepsilon,n}=\{\omega：|X_n(\omega)-X(\omega)|>\varepsilon\}$并假设$0<\varepsilon<1$。此外，让$A_n=\{\omega：|X_n(\omega)-X(\omega)|<1\}$并写出

$$d(X_n,X)=\int_{A_n}\frac{|X_n-X|}{1+|X_n-X|}dP+\int_{A_n^c}\frac{|X_n-X|}{1+|X_n-X|}dP。$$

我们从下面估计这两个条款的每一个。首先、

$$\int_{A_n}\frac{|X_n-X|}{1+|X_n-X|}dP\geq\int_{A_n\cap E_{\varepsilon,n}}\frac{|X_n-X|}{1+|X_n-X|}dP$$
$$\geq\frac{1}{2}\int_{A_n\cap E_{\varepsilon,n}}\varepsilon\,dP$$
$$=\frac{\varepsilon}{2}P(A_n\cap E_{\varepsilon,n})$$

因为$\frac{a}{1+a}>\frac{a}{2}$，如果$a<1$。第二、

$$\int_{A_n^c}\frac{|X_n-X|}{1+|X_n-X|}dP\geq\int_{A_n^c}\frac{1}{2}dP\geq\int_{A_n^c}\frac{1}{2}dP\geq\frac{\varepsilon}{2}P(A_n^c\cap E_{\varepsilon,n})$$

因此，$d(X_n,X)\geq\frac{\varepsilon}{2}P(E_{\varepsilon,n})\to 0$，所以$(X_n)$收敛于$X$概率。 □

证明（命题8.6的证明）

首先

$$E(Y^p) = \int_{\{\omega:Y(\omega)\geq\varepsilon\}} Y^p\,dP + \int_{\{\omega:Y(\omega)<\varepsilon\}} Y^p\,dP$$

$$\geq \varepsilon^p P(Y \geq \varepsilon) + \int_{\{\omega:Y(\omega)<\varepsilon\}} Y^p\,dP。$$

现在，如果我们让$\varepsilon \to \infty$，那么第二项就会收敛到$\int_\Omega Y^p\,dP = E(Y)^p$

而第一项没有选择，只能收敛到0。 □

证明（命题8.13的证明）

(i) 写$A=\liminf_{n\to\infty} A_n$。如果$\omega\in A$，那么存在一个$N$，使得$\omega\in A_n$，对于所有的$n\geq N$，这意味着$\omega\in A_n$，除了有限多的$n$之外。反之，如果$n$最终$\in A_n$，那么存在$N$，使得$\omega\in A_n$，对于所有$n\geq N$和$\omega\in$A。

(ii) 如果$A^\varepsilon = \{|X_n - X| < \varepsilon\}$，那么$\{|X_n - X| < \varepsilon\ e.v.\} = \{|X - X| < \varepsilon\ e.v.\}$。= $\liminf_{n\to\infty} A_n^\varepsilon = A^\varepsilon$，说。但是

$$\{X_n \to X\} = \bigcap_{\varepsilon>0}\{|X_n - X| < \varepsilon\ \text{ev.}\} = \bigcap_{\varepsilon>0} A^\varepsilon$$

而集合A^ε，随着ε&0而减少。依次取$\varepsilon=\frac{1}{n}$，表明

$$P(X_n \to X) = P\left(\bigcap_{n=1}^{\infty}\liminf A_n^{1/n}\right) = \lim_{\varepsilon\to 0} A^\varepsilon = \lim_{\varepsilon\to 0} P(|X_n - X| < \varepsilon\ \text{ev.})。$$

(iii) 作者：德-摩根

$$\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} a_m\right)^c = \bigcap_{n=1}^{\infty}\left(\bigcap_{m=n}^{\infty} a_m\right)^c = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} a_m^c。$$

(iv) 如果$A = \liminf_{n\to\infty} A_n$，那么$1_A = \liminf_{n\to\infty} 1_{A_n}$（因为$1_A(\omega) = 1$ 当且仅当$1_{A_n} = 1$ ev.），如果$B = \limsup_{n\to\infty} A_n$，那么$1_B = \limsup_{n\to\infty} 1_{A_n}$（因为$1_B(\omega) = 1$ 当且仅当$1_{A_n} = 1$ i.o）。法图定理的含义是

$$P(A_n\ \text{ev.}) = \int_\Omega 边缘\inf_{n\to\infty} 1_{A_n}\,dP \leq \liminf_{n\to\infty}\int_\Omega 1_{A_n}\,dP$$

就是

说

但是

$$P(B) \leq \inf_{n\to\infty} \text{边缘} \; P(A_n) \leq \limsup_{n\to\infty} P(A_n)。$$

$$\limsup_{n\to} P(A_n) = \limsup_{n\to} \int_{\Omega} 1_{A_n}\, dP \leq \int_{\Omega} \limsup_{n\to} 1_{A_n}\, dP$$

由法图反推（见定理4.18的证明），因此

$$\limsup_{n\to\infty} P(A_n) = \int_\Omega 1_A \, dP = P(A) = P(A_n \text{ i.o.})。$$

□

9 练习的解决方案

第二章

2.1 如果我们可以用规定总长度的开放区间覆盖A，那么我们也可以用具有相同端点的封闭区间覆盖A（封闭区间更大），而且总长度是相同的。对于任何其他类型的区间来说也是如此。反过来说，假设在用封闭区间覆盖的意义上，A是空的。设$\varepsilon>0$，取一个覆盖$C_n = [a_n, b_n]$，其中$\sum_n (b_n - a_n) < \frac{\varepsilon}{2}$，让$I_n = (a_n - \varepsilon\frac{1}{2^{n+2}}, b_n + \varepsilon\frac{1}{2^{n+2}})$。它们都比较大，所以它们覆盖了$A$；总长度小于$\varepsilon$。以同样的方式，我们细化了由任何其他种类的间隔来覆盖。

2.2 以三元形式写出C的每个元素。假设C是可数的，因此它们可以被排列成一个序列。定义一个不在这个序列中但有三元扩展的数字，因此在C中，在第n个位置交换0和2。

2.3 为了在x [0, 1]处的连续性，取任何$\varepsilon > 0$，并找到$F(x) - \varepsilon < a < F(x) < b < F(x) + \varepsilon$的形式$a = \frac{k}{2^n}$, $b = \frac{m}{2^n}$。根据F的构造、这些数字是F在某些区间(a_1, a_2), (b_1, b_2)上取的值，有三元的端点，且$a_1 < x < b_2$。取一个δ，使得$a_1 < x - \delta < x + \delta < b_2$以得到连续性条件。图表见图4.7。

2.4 $m^*(B) \le m^*(A \cup B) \le m^*(A) + m^*(B) = m^*(B)$ 根据单调性（Proposition 2.3）和次加性（Theorem 2.5）。因此，$m^*(A \cup B)$ 被挤压在$m^*(B)$ 和$m^*(B)$ 之间，所以没有什么选择。

2.5 由于$A \subset B \cup (A \Delta B)$，$m^*(A) \le m^*(B) + m^*(A \Delta B) = m^*(B)$ (单调-)

ity和subadditivity），反之则以同样方式显示。

2.6 由于$A\cup B=A\cup(B\setminus A)=A\cup(B\setminus(A\cap B))$，利用可加性和命题2.10，我们有$P(A\cup B)=P(A)+P(B)-P(A\cap B)$。同样，$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$。

2.7 只要注意到以下几点就足够了

$$(a,b)=\bigcup_{n=1}^{\infty}(a,b-\tfrac{1}{n}]\text{、}\qquad (a,b)=\bigcup_{n=1}^{\infty}[a+\tfrac{1}{n},b).$$

2.8 如果E是Lebesgue可测量的，那么O和F的存在是由The-分别为公式2.12和2.19。反过来说，让$\varepsilon=\frac{1}{n}$，求O_n，F_n，和然后 $m^*(\bigcap O_n \setminus E)=0$，$m^*(E\setminus\bigcup F_n)=0$，所以 E是Borel至空集。

因此，E是在M中。

2.9 我们可以将A分解为$\bigcup_{i=1}^{\infty}(A\cap H_i)$；这些成分是成对的，所以

$$p(a)=\sum_{i=1}^{\infty}p(a\cap h_i)=\sum_{i=1}^{\infty}p(a/h_i)\cdot p(h_i)$$

使用条件概率的定义。

2.10 由于$A^c\cap B=(\Omega\setminus A)\cap B=B\setminus(A\cap B)$, $P(A^c\cap B)=P(B)-P(A\cap B)=p(b)-p(a)\cdot p(b)=p(b)(1-p(a))=p(b)\cdot p(a^c)$。

2.11 There are 32 paths altogether. $S(5)=524.88=500U^2D^3$ so there are $\binom{5}{2}=10$条路径。$S(5)>900$有两种情况：$S(5)=1244.16=500U^5$或$S(5)=933.12=500U^4D$所以我们有6条路径，每条路径的概率为$0.5^5=0.03125$所以问题的概率是0.1875。

2.12 有2^m条长度为m的路径、$\mathcal{F}_m$可以确定为所有此类路径集合的幂集，因此它有2^{2^m}个元素。

2.13 $\mathcal{F}_m\subset\mathcal{F}_{m+1}$因为如果一个序列的前$m+1$个元素是相同的，那么前$m$个元素也是相同的。

2.14 只要注意到$P(A_m\cap A_k)=\frac{1}{4}$，这与$P(A_m)P(A_k)$是一样的就可以了。

第三章

3.1 如果f是单调的（在这个意义上，$x_1\leq x_2$意味着$f(x_1)\leq f(x_2)$，

区间（a，∞）的反像是[b，∞]或（b，∞），$b=sup${x：f（x）≤a}，所以它显然是可测的。

3.2 水平集{x : f (x) = a}是f^{-1} ([a, ∞])的交点和 f^{-1} (($-\infty$, a])，根据定理3.1，每个都是可测量的。

3.3 如果$b \geq a$，则$(f)^{a-1}((-\infty, b]) = \mathbb{R}$，如果$b<a$，则$(f)^{a-1}((-\infty, b]) = f^{-1}((-\infty, b])$；在每种情况下都是一个可测量的集合。

3.4 设A是一个不可测量的集合，如果$x \in A$，让$f(x)=1$，否则$f(x)=-1$。集合f^{-1}（[1，∞）是不可测量的（它是A），所以f是不可测量的，但$f^2 \equiv 1$显然是可测量的。

3.5 设$g(x)=\limsup f_n(x)$，$h(x)=\liminf f_n(x)$；根据The- orem 3.5，它们是可测的。它们的差值也是可测的，f_n收敛的集合是这个差值的水平集：$\{x : f_n \text{ 收敛}\} = \{x : (h - g)(x)=0\}$，因此是可测的。

3.6 如果sup $f = \infty$，那么就没有什么可证明的。如果sup $f = M$，那么$f(x) \leq M$对所有的x都是如此，所以M是ess sup定义中的z之一。让f是连续的，并假设ess sup$f < M$是有限的。那么我们在这些数字之间取z因此，$A = \{x : f(x) > z\}$是空的。然而，通过连续性，A包含集合$f^{-1}((z, M))$，它是开放的--一个矛盾。如果sup f是无限的，那么对于每个z，条件$f(x) \leq z$ a.e.不成立。空集的下限是$+\infty$，所以我们就完成了。

3.7 只需注意到，如果G是任何包含Borel集反像的σ场，那么$\mathcal{F}_X \subseteq \mathcal{G}$。

3.8 让f：$\mathbb{R} \to \mathbb{R}$由$f(x) = x^2$。集合$f(\mathbb{R})$等于$(-\infty, 0)$的补集不能是$f(A)$的形式，因为每一个都包含在$[0, \infty]$中。

3.9 跌停板欧式看涨的回报是$f((S(0), S(1), \ldots, S(n))$，其中$f(x_0, x_1, \ldots, x_N) = (x_N - K)^+ - 1_A$，其中$A = \{(x_0, x_1, \ldots, x_N) : \min\{x_0, x_1, \ldots, x_N\} \geq L\}$。

第四章

4.1 (a) $\int_0^{10} \mathrm{Int}(x)\,dx = 0m([0, 1)) + 1m([1, 2)) + 2m([2, 3)) + \cdots + 9m([9, 10)) + 10m([10, 10]) = 45$.

(b) $\int_0^{4} \mathrm{Int}(\sqrt{x})\,dx = 0m([0, 1]) + 1m([1, \sqrt{2})) + 2m([\sqrt{2}, \sqrt{3})) + 3m([\sqrt{3}, 4)) = 5 - \sqrt{2} - \sqrt{3}$.

(c) $\int_0^{2\pi} \mathrm{Int}(\sin x)\,dx = 0m([0, \frac{\pi}{2})) + 1m([\frac{\pi}{2}, \frac{\pi}{2}]) + 0m((\frac{\pi}{2}, \pi]) - 1m((\pi, 2\pi]) = -\pi$.

4.2 我们有$f(x) = \sum_{k=1}^{\infty} k 2^{k-1} 1_{A_k}(x)$，其中$A_k$是$2^{k-1}$间的联合。每个长度为$\frac{1}{3^k}$在第k阶段从[0，1]中删除。在第k个阶段，每一个vals的长度被从[0，1]中移除。

收敛性是单调的，所以

$$\int_{[0,1]} f\,dm = \lim_{n\to\infty}\sum_{k=1}^{n} k\frac{2^{k-1}}{3^k} = \frac{1}{3}\sum_{k=1}^{\infty} k\left(\frac{2}{3}\right)^{k-1}.$$

由于$\sum_{k=1}^{\infty}\alpha^k = \frac{1}{1-\alpha}$，对$\alpha$逐项进行微分，所以我们有$\sum_{k=1}^{\infty} k\alpha^{k-1} = \frac{1}{(1-\alpha)^2}$。与$\alpha$到$=\frac{2}{3}$我们得$\int_{[0,1]} f\,dm = 3$.

4.3 简单的函数$a1_A$小于f，所以它的积分，等于$am(A)$，比f的积分小。接下来，$f\le b1_A$，因此有第二个不等式。

4.4 让$f_n = n1_{(0,\frac{1}{n}]}$；$\lim f_n(x) = 0$ for all x but$\int f_n\,dm = 1$。

4.5 让$\alpha \neq -1$。我们有$\int x^\alpha\,dx = \frac{1}{\alpha+1}x^{\alpha+1}$（不确定的积分）。首先CON-考虑到$E=(0，1)$：

$$\int_0^1 x^\alpha\,dx = \frac{1}{\alpha+1}\left[x^{\alpha+1}\right]_0^1$$

如果$\alpha>-1$，它就是有限的。接下来$E=(1，\infty)$：

$$\int_1^\infty x^\alpha\,dx = \frac{1}{\alpha+1}\lim_{n\to\infty}\left[x^{\alpha+1}\right]_1^n$$

而要使之成为有限的，我们需要$\alpha<-1$。

4.6 序列$f_n(x) = \frac{\sqrt{x}}{1+nx^3}$点状收敛于0、$\frac{\sqrt{x}}{1+nx^3}\le\frac{\sqrt{x}}{nx^3}\le\frac{1}{n}x^{-2.5}\le x^{-2.5}$，在$[1,\infty)$上是可整定的、所以积分的序列是收敛到0。

4.7 首先$a=0$，代入$u=nx$：

$$\int_0^\infty \frac{n^2 x e^{-n^2x^2}}{1+x^2}\,dx = \int_0^\infty \frac{u e^{-u^2}}{1+(\frac{u}{n})^2}\,du。$$

积分序列收敛到ue^{-u^2}对于所有的$u\ge0$，它被支配着通过$g(u)=ue^{-u^2}$，所以

$$\int_0^\infty f_n\,dm = \int_0^\infty \lim f_n\,dm = \int_0^\infty ue^{-u^2}\,du = \frac{1}{2}.$$

现在$a>0$.经过同样的替换，我们有

$$\int_a^\infty \frac{n^2 x e^{-n^2x^2}}{1+x^2}\,dx = \int_{\mathbb{R}} \frac{ue^{-u^2}}{1+(\frac{u}{n})^2}1_{[na,\infty]}(u)\,du = \int_{\mathbb{R}} f_n(u)\,du、$$

说，并且$f_n\to0$，$f_n(u)\le ue^{-u^2}$，所以$\lim\int f_n\,dm = 0$。

4.8 序列f_n (x) 在$x\geq 0$时收敛于e^{-x}。我们发现支配性的函数。让$n>1$。对于$x\in$（0，1），$\frac{x^n}{n}\geq \bar{x}^2$，$(1+\frac{x}{n})^n\geq 1$，所以$f_n$ $(x)\leq \frac{1}{\sqrt{x}}$

它在（0，1）上是可整定的。对于$x\in[1,\infty]$，$x^{-\frac{1}{n}}\leq 1$，所以f_n $(x)\leq (1+\frac{x}{n})^{-n}$。下一页

$$1+\frac{x}{n}^n = 1 + x + \frac{n(n-1)}{2!}\frac{x^2}{n} + \ldots > x^2\frac{n-1}{2n}\geq \frac{1}{4}x^2$$

所以$f_n(x)\leq \frac{4}{x^2}$，在$[1,\infty]$上是可整数的。

因此，根据支配性收敛定理、

$$\infty f\int_0^{\infty} f_n\, dm = \int_0^{\infty} e^{-x}\, dx = 1。$$

4.9 (a) $\int_{-1}^{1} |n\, x^{an}|\, dx = n^a \int_{-1}^{1} |x|^n\, dx = n^a \int_0^1 x^n\, dx$ ($|x|^n$ 是偶数函数)

$=\frac{2n^a}{n+1}$。如果$a<0$，那么序列$\sum_{n\geq 1}\frac{2n^a}{n+1}$收敛，与$\frac{1}{n^{1-a}}$，我们可以应用Beppo-Levi定理，问题中的幂级数定义了一个可整定的函数。如果$a=0$，该数列为$\sum_{n\geq 1} x^n = \frac{x}{1-x}$

由于$\int_{-1}^{1}(\sum_{n\geq 1} x^n)\, dx = \sum_{n=1}^{\infty}\int_{-1}^{1} x^n\, dx = \infty$.通过比较一下，如果$a>0$，该系列不能给出一个可整定的函数。

(b) 写作$\frac{x}{e^x-1} = xe^{-x}\frac{1}{1-e^{-x}} = \sum_{n\geq 1} xe^{-nx}$，$\int_0^{\infty} xe^{-nx}\, dx = x(-\frac{1}{n})e^{-nx}\big|_0^{\infty} - \int_0^{\infty}(-\frac{1}{n})e^{-nx}\, dx = \frac{1}{n^2}$ (通过部分整合)，而且，众所周知、$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$。

4.10 我们通过将$f(0)=1$来扩展f，因此f是连续的，因此黎曼在任何有限区间上都是可积分的。让一个$_n = \int_{n\pi}^{(n+1)\pi} f(x)\, dx$。由于$f$是偶数、

$a_{-n} = a_n$，因此，$\int_{-\infty}^{\infty} f(x)\, dx = 2\sum_{n=0}^{\infty} a_n$。该系列收敛，因为$a_n = (-1)^n|a_n|$, $|a_n|\leq \frac{2}{n\pi}$ ($x\geq n\pi$, $|\int_{n\pi}^{(n+1)\pi}\sin x\, dx| = 2$)。然而对于$f$是在$L^1$我们将需要$\int_{\mathbb{R}}|f|\, dm = 2\sum_{n=0}^{\infty} b_n$有限，其中$b_n = \int_{n\pi}^{(n+1)\pi}|f(x)|\, dx$。这是不可能的，因为$b_n\geq \frac{2}{(n+1)\pi}$.

4.11 表示$\int_{-\infty}^{\infty} e^{-x^2}\, dx = I$；那么

$$I^2 = \int\int_{R^2} e^{-(x^2+y^2)}\, dx\, dy = \int_0^{2\pi}\int_0^{\infty} re^{-r^2}\, dr\, d\alpha = \pi$$

使用极坐标和富比尼定理（第6章）。

代入$x = \frac{z-\mu}{\sqrt{2}\sigma}$ 在I $\sqrt{\pi} = \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{1}{\sqrt{2}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$

这就完成了计算。

4.12 $\int_{R} \frac{1}{1+x^2} dx = \arctan x\Big|_{-\infty}^{+\infty} = \pi$ 因此，$\int_{-\infty}^{\infty} c(x)\, dx = 1$。

4.13 $\int_{0}^{\infty} e^{-\lambda x}\, dx = -\frac{1}{\lambda} e^{-\lambda x}\Big|_{0}^{\infty} = \frac{1}{\lambda}$，因此$c = \lambda$。

4.14 让$a_n \to 0$，$a_n \geq 0$。那么$P_X(\{y\}) = \lim_{n\to\infty} P_X((y - a_n, y]) = F_X(y) - \lim_{n\to\infty} F_X(y - a_n)$ 这证明了所需的等价关系。(回顾一下，P_X总是右连续的。)

4.15 (a) $F_X(y) = 1$，对于$y \geq a$，否则为零。

(b) 对于$y < 0$，$F_X(y) = 0$；对于$y \geq \frac{1}{2}$，$F_X(y) = 1$，以及$F_X(y) = 2y$ 否则的话。

(c) 对于$y < 0$，$F_X(y) = 0$，对于$y \geq \frac{1}{2}$，$F_X(y) = 1$，以及$F_X(y) = 1 - (1 - 2y)^2$ 否则的话。

4.16 在这种情况下，$\varphi(x) = x^3$，$\varphi^{-1}(y) = \sqrt[3]{y}$，$\frac{d}{dy}\varphi^{-1}(y) = \frac{1}{3} y^{-\frac{2}{3}}$ 因此$f_{X^3}(y) = 1_{[0,1]}(\sqrt[3]{y}) \frac{1}{3} y^{-\frac{2}{3}} = \frac{1}{3} y^{-\frac{2}{3}} 1_{[0,1]}(y)$.

4.17 (a) $\int_\Omega a\,dP = aP(\Omega) = a$ (常数函数是一个简单的函数)。

(b) 利用练习4.15，$f_X(x) = 2 1_{[0,\frac{1}{2}]}(x)$，$E(X) = \int_0^{\frac{1}{2}} 2x\,dx = \frac{1}{4}$。

(c) 再根据练习4.15，$f_X(x) = 4x 1_{[0,\frac{1}{2}]}(x)$ 所以$E(X) = \int_0^{\frac{1}{2}} 4x^2\,dx = \frac{1}{6}$。

4.18 (a) 随着$f_X = \frac{1}{b-a} 1_{[a,b]}$，$E(X) = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\frac{1}{2}(b^2 - a^2) = \frac{1}{2}(a + b)$.

(b) 考虑简单的三角形分布，$f_X(x) = x + 1$为$x \in [-1, 0]$，$f_X(x) = -x + 1$ for $x \in (0, 1]$and zero elsewhere.那么我最近 $\int_{-1}^{1} x f_X(x)\,dx = 0$。对密度$f_Y$的类似计算，其三角形的底是$[a, b]$，得到$E(X) = \frac{a+b}{2}$。

(c) $\lambda \int_0^\infty x e^{-\lambda x}\,dx = \frac{1}{\lambda}$（通过部分整合）。

4.19 (a) $\phi_X(t) = \frac{1}{(b-a)}(e^{ibt} - e^{iat})$、

(b) $\phi_X(t) = \lambda \int_0^\infty e^{(it-\lambda)x}dx = \frac{\lambda}{\lambda - it}$、

(c) $\phi_X(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$。

4.20 使用调用-投入奇偶性，高斯分布的调用和对称性公式，我们有$P = S(0)(N(d_1) - 1) - Ke^{-rT}(N(d_2) - 1) = -S(0)N(-d_1) + Ke^{-rT}N(-d_2)$

第五章

5.1 首先$f \equiv f$，因为$f = f$无处不在。其次，如果$f = g$ a.e.，那么当然$g = f$ a.e. 第三，如果$f = g$在一个完整的集合$F_1 \subset E$上（$m(E \setminus F_1) = 0$），$g = h$在一个完整的集合$F_2 \subset E$上，那么$f = h$在$F_1 \cap F_2$上也是完整的。

5.2 (a) $\|f_n - f_m\|_1 = m - n$ 如果$m > n$，那么这个序列就不是考奇的。

(b) $\|f_n - f_m\|_1 = \int_n^m \frac{1}{x}dx = \log m - \log n$（为简单起见，假设

$n < m$），设$\varepsilon = 1$，取任意N，设$n = N$，取m，使log m-log $N > 1$（log $m \to \infty$，因为$m \to \infty$） --该序列不是Cauchy。

(c) $\|f_n - f_m\|_1 = \int_n^m \frac{1}{x^2}dx = -\frac{1}{x}\Big|_n^m = \frac{1}{n} - \frac{1}{m}$ ($n < m$如前)，以及对于任何$\varepsilon>0$，取N，使$\frac{1}{N} < \frac{\varepsilon}{2}$，对于$n$，$m \ge N$，显然$\frac{1}{n} - \frac{1}{m} < \varepsilon$ - 序列是Cauchy的。

5.3 $\|g_n - g_m\|_2^2 = \int_n^m \frac{1}{x^4}dx = -\frac{1}{3x^3}\Big|_n^m = \frac{1}{3}\left(\frac{1}{n^3} - \frac{1}{m^3}\right)$ - 该序列是考奇的。

5.4 (a) $\|f_n - f_m\|_2^2 = m - n$ 如果$n \ge m$，那么该序列不是考奇的。

(b) $\|f_n - f_m\|_2^2 = \int_n^m \frac{1}{x^2}dx = \frac{1}{n} - \frac{1}{m} \to 0$ - 该序列是Cauchy。

(c) $\|f_n - f_m\|_2^2 = \int_n^m \frac{1}{x^4}dx = \left(\frac{1}{3n^3} - \frac{1}{3m^3}\right)$ - 该序列是Cauchy。

5.5 $\|f + g\|^2 = 4, \|f - g\|^2 = 1, \|f\|^2 = 1, \|g\|^2 = 1$, 以及平行四边形定律被违反。

5.6 $\|f + g\|_1^2 = 0$，$\|f - g\|_1^2 = 1$，$\|f\|_1^2 = \frac{1}{4}$，$\|g\|_1^2 = \frac{1}{4}$，这与之矛盾。平行四边形法。

5.7 由于$\sin nx \cos mx = \frac{1}{2}[\sin(n + m)x + \sin(n - m)x]$和$\sin nx \sin mx = \frac{1}{2}[\cos(n - m)x + \cos(n + m)x]$，很容易计算出不确定的积分。它们是周期性函数，所以在$[-\pi, \pi]$上的积分为零（对于后者我们需要$n \neq m$）。

5.8 不：取任何n，m（假设$n<m$）并计算出

$$\|g_n - g_m\|_4^4 = \int_{\frac{1}{m}}^{\frac{1}{n}} \left(\frac{1}{\sqrt{x}}\right)^4 dx = -x^{-1}\Big|_{1/m}^{1/n} = (m - n) \ge 1$$

所以这个序列不是考奇的。

5.9 假设$\Omega = [0, 1]$，具有Lebesgue度量，$X(\omega) = \frac{1}{\sqrt{\omega}}$，$E(X) = \int_0^1 X\,dm = \int_0^1 \frac{1}{\sqrt{x}}dx = 2$，$E(X^2) = \int_0^1 \frac{1}{x}dx = \infty$。如果我们把$X(\omega) = \frac{1}{\sqrt{\omega}} - 2$然后$E(X) = 0$，$E(X^2) = \infty$。

5.10 $\mathrm{Var}(aX) = E((aX)^2) - (E(aX))^2 = a^2(E(X^2) - (E(X))^2) = a^2\,\mathrm{Var}(X)$。

5.11 设$f(x) = \frac{1}{b-a}1_{[a,b]}$，$E(X) = \frac{a+b}{2}$，$\mathrm{Var}X = E(X^2) - \frac{(a+b)^2}{4}$，$E(X^2) = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\frac{1}{3}(b^3 - a^3)$，简单的代数可以得到结果。

5.12 (a) $E(X) = a$，$E((X-a)^2) = 0$，因为$X = a$ a.s。

(b) 根据练习4.15 $f_X(x) = 2 1_{[0,\frac{1}{2}]}(x)$，根据练习4.17，$E(X) = \frac{1}{4}$；所以$\mathrm{Var}(X) = \int_0^{\frac{1}{2}} 2\left(x - \frac{1}{4}\right)^2 dx = 2\int_{-\frac{1}{4}}^{\frac{1}{4}} x^2\,dx = \frac{1}{48}$。

(c) 根据练习4.15，$f_X = 4x 1_{[0,\frac{1}{2}]}(x)$，根据练习4.17，$E(X) = \frac{1}{6}$、因此$\mathrm{Var}(X) = 4\int_0^{\frac{1}{2}} x\left(x - \frac{1}{6}\right)^2 dx = \frac{1}{48}$。

5.13 $\mathrm{Cov}(Y, 2Y+1) = E((Y)(2Y+1)) - E(Y)E(2Y+1) = 2E(Y^2) - 2(E(Y))^2 = 2\mathrm{Var}(Y)$, $\mathrm{Var}(2Y+1) = \mathrm{Var}(2Y) = 4\mathrm{Var}(Y)$, 因此$\rho = 1$.

5.14 根据练习5.7，X，Y是不相关的。取$a > 0$，小到集合$A = \{\omega : \sin 2\pi\omega > 1 - a\}$，$B = \{\omega : \cos 2\pi\omega > 1 - a\}$不相交。那么$P$（$A \cap B$）$= 0$，

但P（A）P（B）/=0。

第六章

6.1 职能

$$g(x, y) = \begin{cases} \frac{1}{x^2} & 如果 0 < y < x < 1 \\ -\frac{1}{y^2} & 如果 0 < x < y < 1 \\ 0 & 否则 \end{cases}$$

是不可积分的，因为g的积分$^+$，是无限的（g的积分$^-$，也是如此）。然而、

$$\int_0^1\int_0^1 g(x, y)\, dx\, dy = -1 \text{、} \quad \int_0^1\int_0^1 g(x, y)\, dy\, dx = 1$$

这表明，如果违反了富比尼定理的条件，迭代积分可能不相等。

6.2 $\int\int_{[0,3]\times[-1,2]} x^2 y\, dm_2 = \int_{-1}^{2}\int_0^3 x^2 y\, dx\, dy = \int_{-1}^{2} 9y\, dy = \frac{27}{2}$。

6.3 根据对称性，只需考虑 $x \geq 0, y \geq 0$，因此，面积为 $4\frac{b}{a}\int_0^a \sqrt{a^2 - x^2}\, dx = ab\pi$。

6.4 固定$x\in[0, 2]$，$\int_0^2 1_A(x, y)\, dy = m(A_x)$，因此，对于$x\in[0, 1]$，$f_X(x) = x$，对于$x\in(1, 2)$，$f_X(x) = 2 - x$，否则为零（三角形分布）。根据对称性，对于$f_Y$也是如此。

6.5 $P(X + Y > 4) = P(Y > -X + 4) = \int_A f_{X,Y}(x, y)\, dx\, dy$ 其中$A = \{(x, y) : y > 4 - x\} \cap [0, 2] \times [1, 4]$, 所以$P(X + Y > 4) = \int_0^2\int_{4-x}^4 \frac{1}{50}(x^2 + y^2)\, dy\, dx = \frac{1}{50}\int_0^2(-4x^2 + \frac{4}{3}x^3 + 16x)\, dx = \frac{8}{15}$。

$$P（Y>X） = \int_1^2\int_0^{y}\frac{1}{50}(x^2 + y^2)\, dx\, dy + \int_2^4\int_0^{2}\frac{1}{50}(x^2 + y^2)\, dx\, dy$$

$$= \frac{1}{50}\int_1^2 \frac{4}{3}y^3\, dy + \frac{1}{50}\int_2^4 (\frac{8}{3} + 2y^2)\, dy$$

$$= \frac{143}{150}$$

类似地，$P（Y > X） = \int_A f_{X,Y}(x, y)\, dx\, dy$ 其中$A = \{(x, y) : y > x\} \cap$ $[0，2] \times [1，4]$，所以 $\int_1^2\int_0^y \frac{1}{50}(x^2 + y^2)\, dx\, dy + \int_2^4\int_0^2 \frac{1}{50}(x^2 + y^2)\, dy\, dx = \frac{1}{50}\int_1^2 \frac{4}{3}y^3\, dy + \frac{1}{50}\int_2^4(\frac{8}{3} + 2y^2)\, dy = \frac{143}{150}$

6.6

$$f_{X+Y}(z) = \int_R f_{X,Y}(x, z - x)\, dx = \begin{cases} 0 & z < 0 \\ z & 0 \leq z \leq 1 \\ 2 - z & 1 \leq z \leq 2 \\ 0 & 2 < z \end{cases}$$

6.7 根据练习6.4，$f_{X,Y}(x, y) \neq f_X(x) f_Y(y)$ 所以X, Y不是独立的。

6.8 $f_{Y+(-X)}(z) = \int_{-\infty}^{+\infty} f_Y(y) f_{-X}(z-y)\,dy = \int_{-\infty}^{+\infty} \frac{1}{2} 1_{[0,2]}(y) 1_{[-1,0]}(z-y)\,dy$，所以

$$f_{Y-X}(z) = \begin{cases} 0 & z < -1 \text{ 或 } 2 < z \\ \frac{1}{2}(z+1) & -1 \le z \le 0 \\ \frac{1}{2} & 0 \le z \le 1 \\ \frac{1}{2}(2-z) & 1 \le z \le 2 \end{cases}$$

$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx = \int_{-\infty}^{+\infty} 1_{[0,1]}(x) \frac{1}{2} 1_{[0,2]}(z-x)\,dx$ 因此

$$f_{X+Y}(z) = \begin{cases} 0 & z < 0 \text{ 或 } 3 < z \\ \frac{1}{2}z & 0 \le z \le 1 \\ \frac{1}{2} & 1 \le z \le 2 \\ \frac{1}{2}(3-z) & 2 \le z \le 3 \end{cases}$$

$P(Y > X) = P(Y - X > 0) = \int_0^{\infty} f_{Y-X}(z)\,dz = \frac{1}{2} + \int_1^2 \frac{1}{2}(2-z)\,dz = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$;
$P(X + Y > 1) = \int_1^{\infty} f_{X+Y}(z)\,dz = \frac{1}{2} + \int_2^3 \frac{1}{2}(3-z)\,dz = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$。

6.9 $f_X(x) = \int_0^{1-\frac{1}{2}x} 1_A\,dy = 1 - \frac{1}{2}x$, $h(y, x) = \frac{1_A(x,y)}{1-\frac{1}{2}x}$ 和 $E(Y|X = 1) = 2\int_0^{\frac{1}{2}} x\,dx = \frac{1}{4}$。

6.10 $f_Y(y) = \int_0^1 (x + y)\,dx = \frac{1}{2} + y$, $h(x|y) = \frac{x+y}{\frac{1}{2}+y} 1_A(x, y)$, $E(X|Y = y) = \int_0^1 x \frac{x+y}{\frac{1}{2}+y}\,dx = \frac{\frac{1}{3}+\frac{1}{2}y}{\frac{1}{2}+y}$。

第七章

7.1 如果$\mu(A)=0$，那么$\lambda_1(A)=0$，$\lambda_2(A)=0$，因此$(\lambda_1+\lambda_2)(A)=0$。

7.2 假设Q是Ω的一个有限分区，它同时细化了P_1和P_2。因此，每个集$A\in Q$可以写成一个不相交的联盟$A=\bigcup_{i=1}^{n} E_i = \bigcup_{j=1}^{m} F_j$ 其中$E_i \in P_1$, $F_j \in P_2$. A的每个元素正好属于一个E_i和一个F_j，所以$A = \bigcup_{i,j}(E_i \cap F_j)$也是一个不相交的联盟。因此，Q细化了$R = \{E \cap F : E \in P_1, F \in P_2\}$（正如上述应用于$A = \Omega$的论证所显示的，这是一个分区）。我们只需看到R Refines P_i、$i = 1, 2$. 但$E \in P_1$可以写成$E = E \cap \bigcup_{F\in P_2} F = \bigcup_{F \in P_2}(E \cap F)$，所以$E$是R的元素的不相交的联合。类似的论证表明，R提炼了P_2。

7.3 我们首先要假设$m(A)=0$，那么$B \setminus A$显然意味着μ支配ν。(事实上$m(B \setminus A)=0$的情况略微更普遍）。然后考虑分区$\{B, A\setminus B, \Omega \setminus A\}$。来看，$h = 1_B$. 为了重复检查、

$$\nu(F) = m(F \cap B) = \int_{F \cap B} 1_B \, dm = \int_{F \cap B} 1_B \, d\mu。$$

7.4 显然，$\mu(\{\omega\}) \geq v(\{\omega\})$等同于$\mu$支配v。对于每个$\omega$，我们有$\frac{dv}{d\mu}(\omega) = \frac{v(\{\omega\})}{\mu(\{\omega\})}$。

7.5 由于v（E）$= \int_E g\, dm$，而我们希望v（E）$= \int_E \frac{dv}{d\mu} d\mu = \int_E \frac{dv}{d\mu} f\, dm$

很自然，我们的目标是取$\frac{dv}{d\mu}(x) = \frac{g(x)}{f}$。那么一个充分条件是如果$A = \{x : f(x) = 0\}$，那么$v(A) = 0$，即$g(x) = 0$。a.e. 在$A$上。然后我们把$\frac{dv}{d\mu}(x) = \frac{g(x)}{f}$在$A$上和0在$A^c$上，我们有

$v(E) = \int_{E\cap A^c} g\, dm = \int_{E\cap A^c} \frac{dv}{d\mu}(x) f\, dm = \int_E \frac{dv}{d\mu} d\mu$，符合要求。

7.6 显然，$v \ll \mu$等同于$A = \{\omega : \mu(\{\omega\}) = 0\} \subset \{\omega : v(\{\omega\}) = 0\}$ 然后$\frac{dv}{d\mu}(\omega) = \frac{v(\{\omega\})}{\mu(\{\omega\})}$，在$A$上为零。

7.7 由于$v \ll \mu$，我们可以写成$h = \frac{dv}{d\mu}$，因此$v(F) = \int_F h\, d\mu$。由于$\mu(F)=0$当且仅当$v(F)=0$时，集合$\{h = 0\}$既是μ空的，也是v空的。因此，$h^{-1} = (\frac{dv}{d\mu})^{-1}$是定义良好的，我们可以使用命题7.7中的(ii)和$\lambda = \mu$来得出结论：$1 = h^{-1}$ h意味着$\frac{d\mu}{dv} = h^{-1}$，如要求的那样。

7.8 $\delta_0((0, 25]) = 0$, 但$\frac{1}{25}m|_{[0,25]}((0, 25]) = 1$; $\frac{1}{25}m|_{[0,25]}(\{0\}) = 0$ 但 $\delta_0(\{0\}) = 1$ 因此，$P_1 \ll P_2$ 或 $P_2 \ll P_1$ 都不是。显然，$P_1 \ll P_3$ 与 $\frac{dP_1}{dP_3}(x) = 2 \times 1_{\{0\}}(x)$和$P_2 \ll P_3$ 与$\frac{dP_2}{dP_3}(x) = 2 \times 1_{(0,25]}(x)$。

7.9 $\lambda_a = m|_{[2,3]}$, $\lambda_s = \delta_0 + m|_{(1,2)}$, and $h = 1_{[2,3]}$。

7.10 假设F在一个a_i，有正的跳动c_i，$i = 1, 2, \ldots$的非常数。以$M \neq a_i$，其中-M $\neq a_i$，并让$I = \{i : a_i \in [-M, M]\}$。那么

$$m_F([-M, M]) = F(M) - F(-M) = \sum_{i\in I} c_i = \sum_{i\in I} m_F(\{a_i\})、$$

这是有限的，因为F在一个有界区间上是有界的。所以任何$A \subseteq [-M, M]/\{a_i\}$都是$m_F$-null，因此是可测的。$\bigcup_{i\in I}\{a_i\}$是$m_F$-null，因此是可测的。但是$\{a_i\}$是$m_F$-可测量的，因此$[-M, M]$的每个子集都是$m_F$-可测量的。最后，R的任何子集$E$是$E\cap[-M, M]$形式的集合的联合，所以$E$也是$m_F$-可测的。

7.11 m_F对于$x\in[0，1]$，密度f（x）=2，否则为零。

7.12 (a) $|x| = 1 + \int_{-1}^{x} f(y)\, dy$，其中$f(y) = -1$，用于$y\in[-1, 0]$，$f(y) = 1$，用于$y \in (0, 1]$.

(b) 设$1>\varepsilon>0$，取$\delta=\varepsilon^2$，$\sum_{k=1}^{n}(y_k - x_k) < \delta$，其中$y_k \leq x_{k+1}$；

然后

$$\left(\sum_{k=1}^{n} |\sqrt{x_k} - \sqrt{y_k}|\right)^2 \leq (\sqrt{y_n} - \sqrt{x_1})^2 = y_n - 2\sqrt{y_n x_1} + x_1 < y_n - x_1$$

$\sum^{n}$

$$\leq \sum_{k=1} (y_k - x_k) < \varepsilon^2 。$$

(c) Lebesgue函数f是可微的，$f' = 0$ a.e. 如果它是绝对连续，可以写成$f(x) = \int_0^x f'(y)\,dy = 0$，a 矛盾。

7.13 (a) 如果F在$[a, b]$上是单调增加的，$\sum_{i=1}^k |F(x_i) - F(x_{i-1})| = F(b) - F(a)$ 对于任何分区$a = x_0 < x_1 < \cdots < x_k = b$。因此$T_F[a, b] = F(b) - F(a)$。

(b) 如果$F \in BV[a, b]$，我们可以写$F = F_1 - F_2$，其中F_1，F_2都是单调递增，因此只有数不清的不一致点。

的不确定性。所以F是连续的即，因此是Lebesgue可测量的。

(c) $f(x) = x^2 \cos\frac{\pi}{x}$ 对于$x \neq 0$和$f(0) = 0$是可微的，但不属于$BV[0, 1]$。

(d) 如果F是Lipschitz，$\sum_{i=1}^k |F(x_i) - F(x_{i-1})| \leq M \sum_{i=1}^k |x_i - x_{i-1}| = M(b-a)$ 的任何分区，所以$T_F[a, b] \leq M(b-a)$ 是有限的。

7.14 回顾一下，$v^+(E) = v(E \cap B)$，其中B是哈恩分解中的正集。与提示中一样，如果$G \subseteq F$，$v(G) \leq v^+(G \cap B) \leq v(G \cap B) + v((F \cap B) \setminus (G \cap B)) = v(F \cap B)$。由于$(F \cap B) \setminus (G \cap B) \subseteq B$，它的v-度量是非负的。但是$F \cap B \subseteq F$，所以$\sup\{v(G): G \subseteq F\}$达到了，等于$v(F \cap B) = v^+(F)$。类似的论证表明$v^-(F) = \sup\{-v(G)\} = -\inf_{G \subset F}\{v(G)\}$。

7.15 对于所有$F \in \mathcal{F}$，$v^+(F) = \int_{B \cap F} f\,d\mu = \sup_{G \subset F} \int_G f\,d\mu$。如果在一个集合上$f > 0$

$C \subset A \cap F$，$\mu(C) > 0$，则$\int_C f\,d\mu > 0$，因此，$\int_C f\,d\mu + \int_{B \cap F} f\,d\mu >$ 超$_{G \subset F} \int_G f\,d\mu$。这是个矛盾，因为$C \cup (B \cap F) \subset F$。a.s. (μ) 在$A \cap F$上。我们可以把$\{f = 0\}$这个集合带入B，因为它不影响积分。因此$\{f < 0\} ⊂ A$，$\{f \geq 0\} \subset B$。但这两个较小的集合分割了Ω，所以我们在两种情况下都是相等的。因此，$f^+ = f\,1_B$，$f^- = -f\,1_A$，因此对于所有$F \in \mathcal{F}$

$$v^+(F) = v(B \cap F) = \int_{B \cap F} f\,d\mu = \int_F f^+\,d\mu、$$

$$v^-(F) = -v(A \cap F) = -\int_{A \cap F} f\,d\mu = \int_F f^-\,d\mu。$$

7.16 $f \in L^1(v)$ iff $\int f^+\,dv$ 和 $\int f^-\,dv$ 都是有限的。那么$\int_E f^+ g\,d\mu$和 $\int_E f^- g\,d\mu$是明确定义的、有限的，它们的差值是 $\int_E fg\,d\mu$。$fg \in L^1(\mu)$，因为$\int (f^+ - f^-)|g|\,d\mu < \infty$。反过来说，如果$fg \in L^1(\mu)$，则两者都 $\int_E f^+|g|\,d\mu$和 $\int_E f^-|g|\,d\mu$是有限的，因此它们的差异也是有限的。$\int_E f\,dv$。

7.17 (a) 如果$\omega \in [0, \frac{1}{4}]$，$E(X \mid G)(\omega) = \frac{1}{2}$，否则$E(X \mid G)(\omega) = \frac{3}{4}$。

(b) 如果$\omega \in [0, \frac{1}{2}]$，$E(X \mid G)(\omega) = \omega$，否则$E(X \mid G)(\omega) = \frac{3}{4}$。

7.18 $E(x_n \mid \mathcal{F}_{n-1}) = E(z_1 z_2 \ldots z_n \mid \mathcal{F}_{n-1}) = z_1 z_2 \ldots Z_{n-1} E(Z_n \mid \mathcal{F}_{n-1})$，和由于$Z_n$独立于 $\mathcal{F}_{n-1}$因此，$E(Z_n \mid \mathcal{F}_{n-1}) = E(Z_n) = 1$，因此。

结果。

7.19 $E(X_n) = n\mu \neq \mu = E(X_1)$，所以$X_n$ 不是一个马太效应。显然，$Y_n = X$ $n\mu_n$
是一个马丁格尔。

7.20 $e((z_1 + z_2 + \cdots + z_n)^2 | f_{n-1}) \leq e(z_1 + z_2 + \cdots + z / f_{nn-1})^2 = (z_1 + Z_2 + \cdots + Z_{n-1})^2$使用詹森不等式。补偿器是确定性的：$A_n = n$。

7.21 对于$s<t$，因为增量是独立的，$w(t)-w(s)$与$w(t-s)$的分布相同、

$$E(\exp(-\sigma w(t) - \frac{1}{2}\alpha^2 t)) = e^{-\sigma w(s) - \frac{1}{2}\sigma^2 t} E(\exp(-[\sigma(w(t) - w(s)])|F_s)$$

$$= e^{-\sigma w(s) - \frac{1}{2}\sigma^2 t} E(\exp(-[\sigma(w(t) - w(s)])$$

$$= e^{-\sigma w(s) - \frac{1}{2}\sigma^2 t} E(\exp(-\sigma w(t - s)))。$$

现在$\sigma w(t - s) \sim N(0, \sigma^2 (t - s))$ 所以期望值等于$E(e^{-\sigma \sqrt{t-s} Z}) = e^{-\frac{1}{2}\sigma^2(t-s)}$ (其中$Z \sim N(0, 1)$)，因此结果如下。

第八章

8.1 (a) $f_n = 1_{[n,n+\frac{1}{n}]}$在$L^p$，点式收敛到0，但不是均匀的。
(b) $f_n = n1_{[0,\frac{1}{n}]} - n1_{[-\frac{1}{n},0]}$点状收敛于0，且为即期收敛。
既不在L^p，也不均匀。

8.2 我们有$\Omega=[0，1]$，有Lebesgue度量。序列$X_n = 1_{(0,\frac{1}{n})}$、$X_n = n1_{(0,\frac{1}{n})}$在概率上收敛到0，因为$P(|X_n| > \varepsilon) \leq \frac{1}{n}$和对序列$g$来说也是如此$_n$。

8.3 有无尽的可能性，最简单的是$X_n(\omega)$ 1（但这个序列会收敛到1），或者，为了确保它不收敛到任何东西，$X_n(\omega) \equiv n$。

8.4 让$X_n = 1$表示头，$X_n = 0$表示尾，那么$\frac{S_{100}}{100}$是100次抛掷中的平均头数。显然，$E(X_n) = \frac{1}{2}$，$E(\frac{S_{100}}{100}) = \frac{1}{2}$、$\mathrm{Var}(X_n) = \frac{1}{4}$，$\mathrm{Var}(\frac{S_{100}}{100}) = \frac{1}{100^2} 100 \cdot \frac{1}{4} = \frac{1}{400}$ 所以

$$P(|\frac{S_{100}}{100} - \frac{1}{2}| \geq 0.1) \leq \frac{1}{0.1^2 400}$$

和

$$P(|\frac{S_{100}}{100} - \frac{1}{2}| < 0.1) \geq 1 - \frac{1}{0.1^2 400} = \frac{3}{4}。$$

8.5 让X_n 是骰子上显示的数字，$E(X_n) = 3.5$, $\mathrm{Var}(X_n) \approx 2.9$。

$$p(|\frac{s_{1000}}{1000} - 3.5| < 0.01) \geq 0.29。$$

8.6 联盟 $\bigcup_{m=n}^{\infty} A_m$ 对于所有的m都等于[0, 1]，所以$\limsup_{n\to\infty} A_n$。

8.7 让d=1。有$\binom{2n}{n}$条路径返回到0，所以P（$S_{2n}=0$）$=\binom{2n}{n}\frac{1}{2^{2n}}$。现在

$$\frac{(2n)!}{(n!)^2} \sim \frac{(\frac{2n}{e})^{2n}\sqrt{2\pi 2n}}{(\frac{n}{e})^{2n}2\pi n} = \frac{2^{2n}\sqrt{2}}{\sqrt{\pi n}}$$

所以P（$S_{2n}=0$）$\sim\frac{c}{\sqrt{n}}$，$c=\sqrt{\frac{2}{\pi}}$。因此$\sum_{n=1}^{\infty} P(A_n)$发散，Borel-坎特利适用（因为$(A_n)$是独立的），所以$P(S_{2n}=0 \text{ i.o.})=1$。对于$d=2$也是一样，因为$P(A_n)\sim\frac{1}{n}$。但对于$d>2$，$P(A_n)\sim\frac{1}{n^{d/2}}$，该系列收敛，根据第一个Borel-Cantelli定理，P（$S_{2n}=0$ i.o.）=0。

8.8 写$S=S_{1000}$；$P(|S-500|<10)=P(\frac{|S-500|}{\sqrt{250}}<0.63)\approx 0.47$。

8.9 对n的条件是$P(|\frac{S_n}{n}-0.5|<0.005)=P(\frac{|S_n-0.5n|}{\sqrt{n/4}}<0.01\sqrt{n})\geq 0.99$，因此$n\geq 66615$。

8.10 写下$x_n=e^{\sigma\sqrt{T/n}}$。那么

$$\frac{1}{2}(\ln U_n+\ln D_n)=\frac{1}{2}\ln(U_nD_n)=\frac{1}{2}\ln\frac{(2R_nx_n)^2}{(1+x_n^2)^2}=\ln e^{r\frac{T}{n}}-\ln(\frac{1+x_n^2}{2x_n}).$$

因此，只需证明右边的最后一项是$\frac{\sigma^2T}{2n}+o(\frac{1}{n})$。但是

$$\frac{1+x_n^2}{2x_n}=\frac{x_n^{-1}+x_n}{2}=\frac{e^{-\sigma\sqrt{T/n}}+e^{\sigma\sqrt{T/n}}}{2}=\cosh(\sigma\sqrt{T/n})$$

$$=1+\frac{\sigma^2T}{2n}+o(\frac{1}{n})$$

以致于

$$\ln(\frac{1+x_n^2}{2x_n})=\ln(1+\frac{\sigma T^2}{2n}+o(\frac{1}{n}))=\frac{\sigma T^2}{2n}+o(\frac{1}{n})。$$

10

附录

非可测量和非伯乐集的存在性

在第二章中，我们定义了Borel集的σ场 $\mathcal{B}$ 和较大的σ场 $\mathcal{M}$ 的Lebesgue可测集，我们随后对Lebesgue积分及其性质的所有分析都涉及R的这两个子集家族。集合内涵

$$\mathcal{B} \subset \mathcal{M} \subset \mathcal{P}(\mathbb{R})$$

是微不足道的；然而，乍一看，它们是严格的，这一点并不明显、即在R中存在不属于Lebesgue可测量的集合，以及存在不属于Borel集合的Lebesgue可测量的集合。在本附录中，我们构建了这类集合的例子。利用 A⊂R是可测的(resp.Borel-measurable)iff其指标函数$1_A \in \mathcal{M}$ (resp.) $\mathcal{B}$ 的事实，可以看出

我们将自动拥有不可测（可测但非博雷尔）函数的例子。

构建一个不可测量的集合需要一些集合理论的准备工作。这需要一个公理的形式，虽然对于集合理论的一致发展来说不需要，但却大大地丰富了该理论。它的真假无法从现代集合理论所基于的标准公理中得到证明，但我们将接受它作为一个公理的有效性，而不进一步深入探讨基础性问题。

选择的公理

假设 $= A_\alpha : \alpha \Lambda$ 是{一个非空的}集合，以某个集合Λ为索引，是一个固定集合Ω的非空的不相交的子集，那么存在一个集合$E\subset\Omega$，它恰好包含来自每个集合A_α 的一个元素，即有
是一个*选择函数f*：$\Lambda\to$A。

备注

这个公理看起来很无害，但它可以被证明是独立于集合论的（Zermelo-Fraenkel）公理。如果集合只有有限的几个成员，那么找到一个选择函数当然是没有问题的。为了说明这样一个函数的存在对于无限集合来说是有问题的，请考虑伯特兰-罗素给出的以下说明：想象一下，我们面临着一个无限的鞋子集合和另一个袜子集合。构建由所有左脚鞋组成的集合很简单；而定义所有左脚袜子的集合就不简单了!

为了构建我们的非可测量集的例子，首先在[0, 1]上定义以下等价关系：如果y-x是一个有理数（将在[-1, 1]），则$x\sim y$。这个关系很容易被看作是反身的、对称的和传递的。因此，它将[0, 1]划分为互不相干的等价类(A_α)，其中对于每个

α，A_α 的任何两个元素x，y都相差一个有理数，而不同类的元素总是相差一个无理数。因此，每个A_α 是可数的，因为Q是可数的，但有不可数的不同类，因为[0, 1]是不可数的。

$_\alpha$现在使用选择公理来构造一个新的集合$E\subset[0, 1]$，其中恰好包含 一个来自A_α 的成员。现在列举[-1, 1]中的理数：只有可数的数目，所以我们可以把它们排序为一个序列（q_n ）。用$E_n = E + q_n$ 来定义E的翻译序列。如果E是

Lebesgue可测，那么每个E_n 也是如此，根据命题2.10，它们的度量是相同的。

但(E_n)是不相干的：为了看到这一点，假设$z\in E_m \cap E_n$ ，对于某些 $m \neq m$。那么我们可以写出$a_\alpha + q_m = z = a_\beta + q_n$ ，对于某些a_α，a_β $\in E$，和

它们的差值$a\ a - q\ q$是有理的。由于E只包含每个类的一个元素，$\alpha=\beta$，因此$m=n$。因此$\bigcup$ E_n 是一个包含[0, 1]的不相交的联盟。

因此，我们有 [0, 1] $\subseteq \bigcup$ E_n [1, 2] 和 $m(E_n) = m(E)$ for all n。
这意味着m的可数可加性和单调性：

$\sum$

$$1 = m([0,1]) \le \sum_{n=1} m(E_n) = m(E) + m(E) + \cdots \le 3.$$

这显然是不可能的，因为和必须是0或.因此我们必须得出结论，E是不可测的。

对于一个非Borel的可测量集的例子，让C表示Cantor集，并定义*康托尔函数*f：[0，1]→C如下：对于x∈[0，1]写道$x = 0.a\ a_{12}$ ......以二进制形式，即$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$，其中每个$a_n$ = 0或1（取非终结性扩展，其中存在选择）。函数$x' \to a_n$是由一个有限多的二进制区间系统决定（即a的值$_n$被满足有限多个线性不等式的x所固定），所以是可测量的 - 因此，由f（x）=Σn∞=$_1$给出的函数f也是如此。$\frac{2安}{3n}$.由于f的所有项

$y = \Sigma_n \infty = 1$ $\frac{2an}{3n}$ 的分子为0或2，由此可见，R_f f的是一个

此外，y的值决定了序列（a_n），因此x，唯一的，所以f是可逆的。

现在考虑上面构造的非可测集E在C中的图像，即让B=f（E）。那么B是空集C的一个子集，因此根据m的完备性，它也是可测的和空的。另一方面，E=$f^{-1}(B)$是不可测量的。我们表明，这种情况与B是一个Borel集。

给定一个集合B B 和一个可测量的函数g，那么$g^{-1}(B)$必须是可测量的。因为，根据可测函数的定义，$g^{-1}(I)$对于每个区间I*都*是可测的，我们有

$$g^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} g^{-1}(A_i)\text{、}\quad g^{-1}(A^c) = (g^{-1}(A))^c$$

对于任何集合和函数都是如此。因此，在可测量函数g下的反像又是可测量的集合形成了一个包含区间的σ场，因此也包含了所有的伯乐集合。

但我们已经找到了一个可测函数f和一个Lebesgue可测集B，对其而言，$f^{-1}(B) = E$ *不是*可测的。因此，可测集B不可能是Borel集，也就是说，B⊂M的包含是严格的。

书目

[1] T.M. Apostol, *Mathematical Analysis*, Addison-Wesley, Reading, 1974.

[2]P.Billingsley, *Probability and Measure*, John Wiley and Sons, New York 1995

[3] Z.Brzezniak, T.Zastawniak, *Basic Stochastic Processes*, Springer-Berlag、伦敦1999年

[4]M.Capinski, T.Zastawniak, *Mathematics for Finance, An Introduction to Financial Engineering*, Springer-Verlag, London 2003

[5] R.J.Elliott, P.E.Kopp, *Mathematics of Financial markets*, Springer-Verlag、纽约1999年

[6]G.R. Grimmett和D.R. Stirzaker, *Probability and Random Processes*, Clarendon Press, Oxford, 1982.

[7]J.Hull, *Options, Futures, and Other Derivatives*, Prentice Hall, Upper Saddle River NJ 2000

[8] P.E. Kopp, *Analysis*, Modular Mathematics, Edward Arnold, London, 1996.

[9] J.Pitman, *Probability*, Springer-Verlag, New York, 1995.

[10] W.Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1966.

[11] G. Smith, *Introductory Mathematics：代数与分析》*，Springer-Verlag，SUMS，1998。

[12] D.Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

索引

SPRINGER

UNDERGRADUATE

MATHEMATICS

SERIES

Measure, Integral and Probability is a gentle introduction that makes measure and integration theory accessible to the average third-year undergraduate student. The ideas are developed at an easy pace in a form that is suitable for self-study, with an emphasis on clear explanations and concrete examples rather than abstract theory.

For this second edition, the text has been thoroughly revised and expanded. New features include:

- a substantial new chapter, featuring a constructive proof of the Radon–Nikodym theorem, an analysis of the structure of Lebesgue–Stieltjes measures, the Hahn–Jordan decomposition, and a brief introduction to martingales
- key aspects of financial modelling, including the Black–Scholes formula, discussed briefly from a measure-theoretical perspective to help the reader understand the underlying mathematical framework.

In addition, further exercises and examples are provided to encourage the reader to become directly involved with the material.

From the reviews of the first edition:

"The level of explanation is excellent and great care has gone into providing motivation for the study of all aspects of the material... Overall, this is an excellent and interesting text."

—*Times Higher Education Supplement*

"A clear, understandable treatment of a very problematic area... The authors are to be commended for their lucid writing style."

—*Journal of the American Statistical Association*

The Springer Undergraduate Mathematics Series (SUMS) is designed for undergraduates in the mathematical sciences. From core foundational material to final year topics, SUMS books take a fresh and modern approach and are ideal for self-study or for a one- or two-semester course. Each book includes numerous examples, problems and fully-worked solutions.

ISBN 1-85233-781-8

EAN

9 781852 337810

springeronline.com