概率理论基础

根据英文第二版翻译。

作者：Andrey Nikolaevich Kolmogorov

英文版本译者：Nathan Morrison

作者小传：A.T. Bharucha-Reid

ISBN：9780486821597

编者注

在准备Kolmogorov（柯尔莫戈洛夫）教授这项基础工作的英文翻译稿的过程中，我们参考了1933年发表在《Ergebnisse Der Mathematik》上的德文原版专题论文**Grundbegriffe der Wahrscheinlichkeitrechnung**（概率论的基本概念），以及1936年出版的俄文译本（由G. M. Bavli翻译）。

编者感谢他两位朋友和他侄女的帮助。

编者感谢 Roy Kuebler 提供的自己由德文翻译的英文版本。

前言

本专题论文的目的在于，给概率理论以公理化的基础。作者给自己设定的任务是，将概率论的基本概念（直到最近[1]这些概念还被认为是相当奇怪的概念）置于现代数学的一般概念之中。

如果不先介绍 Lebesgue（勒贝格）关于测度和积分的理论，以上任务几乎是无法完成的。在 Lebesgue 发表他的研究成果之后，集合的度量与事件的概率之间的联系，以及函数的积分与随机变量的数学期望之间的联系，变得显而易见。这些联系为进一步的推广提供了可能，例如，独立随机变量的各种性质被看作与正交函数的相应性质完全类似。但如果要将概率论置于以上的可类比的联系的基础之上，仍然有必要使度量和积分理论独立于勒贝格所突出的几何元素。这项工作由 Fréchet 完成[^2]。

虽然基于上述一般观点的概率论理论在某些数学家中已经存在了一段时间，但缺乏一部完整的、不依赖外部复杂概念的系统性阐述。（参考文献中的 Fréchet 的著作[2]，给出了一个比较完整的论述。）

我想请大家注意那些超出了专家们早已熟知的以上提到的内容，包括：

1. 有限维空间中的概率分布（第三章，第 4 节）；
2. 对某个参数的数学期望的微分和积分（第四章，第 5 节）；
3. 条件概率和条件期望（第五章）；

这些新问题是随着一些完全具体的物理问题产生的[3]。

第六章包含了对 A. Khinchine 和作者关于普通和强大数定律适用性限制的一些结果的概述，但没有提供证明。参考文献中列举了一些近期的作品。

作者对 Khinchine 的致谢，后者通读全文后给予了一些改进建议。

Kljasma near Moscow

1933年复活节

初等概率论

初等概率理论，定义为：

> 研究有限个（随机）事件的概率的理论。

当然，在加入一些必要的新原则之后，这一理论也可用于研究无限个随机变量的问题。第二章将会介绍研究无限个随机变量概率的数学理论时用到的一个公理（公理六）。

作为一门数学分支的概率理论，可以，并且应该像几何和代数一样由一些基本公理构建起来。这意味着，在我们定义好研究的基本要素和它们的基本关系，并且定义好规定这些关系的公理之后，所有进一步的阐述都必须完全基于这些公理，而与这些基本要素及其关系的具体的现实含义无关。

在第一节中，概率域（field of probabilities）定义为满足一些特定条件的集合系统。在纯数学角度发展概率理论的过程中，集合中元素的具体含义是什么，一点都不重要（参考希尔伯特（Hilbert）的 *Foundations of Geometry* 一书中对于基本几何概念的介绍，或者抽象代数中对于群、环和域的定义）。

众所周知，每一个公理化（抽象）理论，除了直接产生它的现实世界的物理域景，都还可以有无数个现实世界中的其他的物理域景（应用）。因此，我们发现了概率论在科学领域中的一些应用，这些应用与**随机事件**和**概率**这两个词的严格含义没有关系。

根据选择的公理不同，以及基本概念和概念之间关联的不同，概率理论的公设基础可以采用不同的方式构建。但如果我们的目标是在公理系统和概率理论未来发展这两方面都达到最大程度的简洁，那么**随机变量**及**概率**这两个基本公设性质的概念似乎是最佳的选择。概率论也有其他的公设系统，尤其在一些公设系统中，**概率**并不被当作基本概念，而是在其他概念的基础上得出的[4]。构建这类公设系统的目的是把数学理论与概率理论的经验发展（empirical development）尽可能紧密地结合起来。

> empirical 用来描述基于观察、实验或经验得出的事实或结论。它强调了通过实证研究或实践经验获得的知识，而不是仅仅依靠理论推导或逻辑推理。在科学、社会科学和哲学等领域中，“empirical”通常指的是基于实际观察和实验的数据或现象，以区别于理论或假设。

公理 [5]

假设 E 是一组元素 ξ, η, ζ, ... 的集合，集合中的元素称为**基本事件**（elementary events）。$\mathcal{E}$ 是集合 E 的子集的集合，那么集合 $\mathcal{E}$ 中的元素称为**随机变量**（random events）。

公理

I. $\mathcal{E}$ 是一个集合的域（a field of sets）；

II. $\mathcal{E}$ 包含集合 E [6]；

III. 对于集合 $\mathcal{E}$ 中的任意一个集合 A，赋予它一个非负实数 $P(A)$。这个实数 $P(A)$ 称为事件 A 的概率。

IV. $P(E)=1$。

V. 如果集合 A 和集合 B 没有公共的元素，则

$$P(A+B)=P(A)+P(B)$$

一个集合系统 $\mathcal{E}$，加上确定的作为概率的实数 $P(A)$，当它们满足公理 1-5 时，就称为一个**概率域**（a field of probability）。

由下面的例子可以证明，公理系统I-V是一致的（consistent）。假设集合E 包含一个元素 ξ，且集合 $\mathcal{E}$ 包括集合 E 和空集 $\emptyset$，则：

$$P(E)=1$$
$$P(\emptyset)=0$$

但是以上公理系统是不完备的（complete），因为在概率理论的各种问题中，需要考虑不同的概率域。

构造概率域

按以下方式构造一个最简单的概率域：

给定任意一个有限集合$E=\{\xi_1,\xi_2,...,\xi_k\}$，以及任意一个非负实数的集合 $\{p_1,p_2,...,p_k\}$，后者满足约束 $p_1+p_2+...+p_k=1$。$\mathcal{E}$ 为所有集合 E 的子集组成的集合，则有：

$$P\{\xi_{i1},\xi_{i2},...,\xi_{i\lambda}\}=p_{i1}+p_{i2}+\cdots+p_{i\lambda}$$

以上公式中，$p_1,p_2,...,p_k$ 称为基本事件 $\xi_1,\xi_2,...,\xi_k$ 的概率，或简称为基本概率（elementary probabilities）。用这种方式，可以推导出所有可能的有限概率域。在这个概率域中， $\mathcal{E}$ 由所有集合 E 的子集组成。当集合 E 为有限集合时，这样产生的概率域称为是有限的。关于概率域更多的例子，参考第二章第三节。

与实验数据的关系 [7]

我们用以下的方式把概率理论应用于现实世界：

1. 假设有一个条件复合体（a complex of conditions） $\mathcal{E}$，其中的条件全部确定时，我们称为一次实现；这里允许发生多次实现，即允许这个条件复合体中的条件重复实现多次（可能出现不同结果）。
2. 我们研究一组明确定义的事件，每当条件 $\mathcal{E}$ 中的条件确定时，这一组事件就会发生。对于每一组（或许是不同）的确定条件，不同的事件发生。假设 E 是所有给定的事件变量（给每一个事件赋予一个变量）$\xi_1,\xi_2,...$ 的集合，集合中的一些变量可能永远也不会发生。我们利用先验知识，尽可能多地在集合 E 中包含可能的变量。
3. 如果在条件实现时所发生的事件变量属于集合 A（以任何方式定义），那么我们称事件 A 已经发生。

   例子：假设条件复合体 $\mathcal{E}$ 是扔一枚硬币 2 次，则第 2 段提到的事件集合包括每次抛硬币时可能出现正面或反面的情况。由此得出，只可能有 4 种事件变量（随机事件），分别为：正正、正反、反正、反反。如果事件 A 表示**重复结果发生**，则它包括正正和反反两种基本事件。这样，**每一个事件都可以视为随机事件的集合**。
4. 在某些条件下（这里不会讨论），我们可以假设对于在条件 $\mathcal{E}$ 下可能发生或可能不发生的事件 A，给它分配一个实数 $P(A)$，这个实数具有以下特征：

   a）可以确定，当条件复合体 $\mathcal{E}$ 中的条件重复足够多次数 n 时，如果 m 表示事件A 发生的次数，则 m/n 与 $P(A)$ 的差别会非常小。

   b）如果 $P(A)$ 很小，则几乎可以确定，当条件只重复一次时事件 A 不会发生。

从经验角度推导公理

一般来说，我们会假设观测到的事件 $A, B, C, ...$ 组成的系统 $\mathcal{E}$ 构成一个域，其中每个事件都被赋予了一个确定的概率。这个事件系统 $\mathcal{E}$ 包含了集合 E （公理 I, II，以及公理 III 的第一部分，假定概率的存在）。很显然 $0 \leq m/n \leq 1$，因此自然推导出公理 III 的第二部分。对于事件 E，m 总是等于 n，因此自然推导出 $P(E) = 1$ （公理 IV）。最后，如果集合 A 与集合 B 没有重叠（不兼容），则：

$$m = m_1 + m_2$$

其中 m_1，m_2 分别表示事件 A，B 发生的次数。由此：

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n}$$

因此可以推出$P(A + B) = P(A) + P(B)$ （公理 V）。

备注 1

如果两个独立的陈述各自都是实际可靠的，那么我们可以说它们同时都是可靠的，尽管可靠程度在这个过程中稍微降低了。然而，如果这样的陈述的数量很大，那么从每一个的实际可靠性中就不能推断出它们是否全部同时正确【不知所云】。

【chatgpt的解释】

这段话探讨了关于多个陈述同时可靠性的问题。首先，它指出，如果两个独立的陈述都是可靠的，那么我们可以认为它们同时都是可靠的，尽管它们的可靠程度可能会在这个过程中略微降低。这是因为独立事件的同时发生概率通常低于单个事件的发生概率。但是，当涉及到大量的陈述时，情况就不同了。作者指出，即使每个陈述单独看都是可靠的，我们不能从这些陈述的可靠性推断出它们全部同时正确的可能性。这是因为随着事件数量的增加，出现任何一种错误的可能性也在增加，尤其是在大量事件中。因此，即使每个单独的测试中的结果都接近预期的概率，但当测试次数增多时，某些测试可能会出现与预期结果有所偏差的情况，这就是所谓的 "实际可靠性"。

因此，由原则（a）中陈述的原则，无法说明在大量重复试验中，每一次计算的 m/n 都会与 $P(A)$ 非常接近。【不会因为这次抛硬币的结果是正面，下次就更可能是反面；也不会因为这次刮彩票不中，下次就更可能中】

备注 2

对于不可能事件（空集），根据公理可得 $P(\emptyset) = 0$。但是相反的说法，由 $P(A) = 0$ [8] 无法推出事件 A 不可能发生。事实上，当$P(A) = 0$时，由原则（b）我们可以确定的是，当条件只重复实现一次时，事件 A 几乎是不可能发生的，但并不能说在大量重复试验的情况下 A 不会发生。另一方面，由原则（a），我们只能推导出，当 $P(A) = 0$ 并且 n 很大时，m/n 的值会非常小（例如等于 $1/n$）。

【概率为 0 的含义，只能说明是一个事件几乎不可能发生】。

术语说明

我们已经将未来研究的对象——随机事件——定义为集合。然而，在概率论中，许多集合论的概念被用其他术语来表示。下面我们将简要列出这些概念。

集合论	随机事件
1. 集合 A 与集合 B 没有交集（do not intersect），即 $AB = 0$。	1. 事件 A 与事件 B 是互斥的（incompatible）。
2. $AB \ldots N = 0$。	2. 事件 $A, B, \ldots, N$ 都互斥。
3. $AB \ldots N = X$。	3. 当事件 $A, B, \ldots, N$ 同时发生时，事件 X 发生。
4. $A \dotplus B \dotplus \ldots \dotplus N = X$。	4. 当事件 $A, B, \ldots, N$ 中至少有一个发生时，事件 X 发生。
5. 集合 A 的补集（complementary）$\bar{A}$。	5. 事件 A 的对立事件 $\bar{A}$（opposite event）由所有当事件 A 不发生时（必然）发生的事件组成。
6. $A = 0$。	6. 事件 A 不可能发生。
7. $A = E$。	7. 事件 A 必然发生。

集合论	随机事件
8. 当 $A_1+A_2+\ldots A_n=E$ 时，称集合 $A_1,A_2,\ldots,A_n$ 组成的系统 $\mathcal{U}$ 组成集合 E 的一个分解（decomposition）。	8. 试验 $\mathcal{U}$ 表示事件 $A_1,A_2,\ldots,A_n$ 中哪些发生，则称事件 $A_1,A_2,\ldots,A_n$ 为试验 $\mathcal{U}$ 的可能结果。
9. B 是 A 的子集：$B\subset A$。	9. 由事件 B 发生，推断出事件 A 必然发生。

公理的直接推论；条件概率；贝叶斯理论

由 $A+\bar{A}=E$，结合公理 IV 和 V 可得：

$$P(A)+P(\bar{A})=1 \tag{1}$$

$$P(\bar{A})=1-P(A) \tag{2}$$

因为 $\bar{E}=0$，所以有：

$$P(0)=0 \tag{3}$$

如果事件 $A,B,\ldots,N$ 互不相容，则由公理 V 可得：

$$P(A+B+\cdots+N)=P(A)+P(B)+\cdots+P(N) \tag{4}$$

如果$P(A)>0$，则公式

$$P_A(B)=\frac{P(AB)}{P(A)} \tag{5}$$

定义为在事件 A 发生的条件下，事件 B 发生的条件概率（conditional probability）。由上面的条件概率定义公式可直接推出：

$$P(AB)=P(A)P_A(B) \tag{6}$$

采用归纳法，可得上式的更广义的形式（乘法定理）：

$$P(A_1A_2\ldots A_n)=P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)\ldots P_{A_1A_2\ldots A_{n-1}}(A_n) \tag{7}$$

> 条件概率的广义形式。当有 n 个事件 $A_1,A_2,\ldots,A_n$ 时。它们同时发生的概率等于事件 A_1 发生的概率，乘事件 A_1 发生的条件下，事件 A_2 发生的概率；乘事件 A_1,A_2 同时发生的条件下，事件 A_3 发生的概率；……；乘 事件 $A_1,A_2,\ldots A_{n-1}$ 同时发生的条件下，事件 A_n 发生的概率。

显然，以下定理可直接推出：

$$P_A(B)\geq 0 \tag{8}$$

$$P_A(E)=1 \tag{9}$$

$$P_A(B+C)=P_A(B)+P_A(C) \tag{10}$$

将上面公式 $(8-10)$ 与公理 III 到 公理 V 对比，可知当集合 A 是一个固定的集合时，给定函数 $P_A(B)$后，集合系统 $\mathcal{E}$ 构成一个概率域，因此当事件 A 给定时，所有以上关于事件 B 的概率 $P(B)$ 的定理/推论同样适用于条件概率 $P_A(B)$。

由条件概率的定义可得：

$$P_A(A)=1 \tag{11}$$

> 当已知事件 A 发生时，事件 A 必然发生。

由公式(6)可类比得到：

$$P(AB)=P(B)P_B(A)$$

进一步得到以下这个非常重要的公式：

$$P_B(A)=\frac{P(A)P_A(B)}{P(B)} \tag{12}$$

即，**贝叶斯定理**（the Theorem of Bayes）。

全概率定理（The Theorem on Total Probability）

假设 $A_1+A_2+\ldots+A_n=E$，并且事件 $A_1,A_2,\ldots,A_n$ 相互独立，假设 X 为任意事件，则有：

$$P(X)=P(A_1)P_{A_1}(X)+P(A_2)P_{A_2}(X)+\cdots+P(A_n)P_{A_n}(X) \tag{13}$$

证明：

$$X=A_1X+A_2X+\cdots+A_nX$$

由公式(4)可得

$$P(X)=P(A_1X)+P(A_2X)+\cdots+P(A_nX)$$

由公式(6)可得

$$P(A_iX)=P(A_i)P_{A_i}(X)$$

贝叶斯定理

假设 $A_1+A_2+\ldots+A_n=E$，并假设 X 为任意事件，则有：

$$P_X(A_i)=\frac{P(A_i)P_{A_i}(X)}{P(A_1)P_{A_1}(X)+P(A_2)P_{A_2}(X)+\cdots+P(A_n)P_{A_n}(X)} \tag{14}$$

上式中，$A_1,A_2,\ldots,A_n$ 称为假说（hypotheses），公式(14)表示当事件 X 发生时，假说 A_i 成立的概率 $P_X(A_i)$。$P(A_i)$ 表示事件 A_i 的先验概率（a priori probability）。

证明：

由公式(12)可得：

$$P_X(A_i)=\frac{P(A_i)P_{A_i}(X)}{P(X)}$$

把上式中的$P(X)$用公式(13)替换，即可证明。

独立性

在某种意义上，两个或多个试验的相互独立性的概念，占据着概率论的核心地位。事实上，正如我们已经看到的，从数学的角度来看，概率论可以被视为一般的可加集函数理论（general theory of additive set functions）的一种特殊应用。既然如此，概率论是如何发展成为一门拥有特定研究方法的科学的呢？

想要回答这个问题，首先必须指出当可加集函数理论中的一般问题在概率理论领域提出时，它所经历的特殊化过程。

可加集函数 $P(A)$ 为非负值且满足 $P(E)=1$，本身并不会导致新的困难。从数学角度看，随机变量（random variables，见第 III 章）仅仅表示相对于 $P(A)$ 可测的（measurable）函数，并且它的数学期望是抽象勒贝格积分（Lebesgue integrals）（这个类比在 Fréchet [9] 的工作中第一次得到了详尽的阐述）。仅仅以上这些概念的介绍，根本不足以产生一门新科学发展的基础。

历史上，试验和随机变量的独立性代表着一个给概率论打上了特别的印记概念。LaPlace、Poisson、Tchebychev、Markov、Liapounov、 Mises、以及 Bernstein 的经典工作致力于对独立随机变量序列的基本研究。尽管最新的论文（Markov、 Bernstein 等人）常常不假设完全独立性，但它们仍然揭示了引入类似的、弱条件的独立性的必要性，以获得足够显著的结果（见第 6 节，马尔可夫链）。

由此可见，独立性的概念是概率论中这一类独特问题的起源。然而本书中并不强调这一点，我们主要关注的是概率理论专业研究的逻辑基础。因此，在自然科学的基本原理（philosophy）中，最重要的问题之一是明确那些使得任意给定的真实事件可以被视为独立的前提——除了众所周知的关于概率这一概念的本质的问题之外。然而，这个问题超出了本书的讨论范围。

独立性的定义：给定 n 个试验，$\mathscr{U}^{(1)},\mathscr{U}^{(2)},\ldots,\mathscr{U}^{(n)}$，即基本集合 E 的 n 个分解：

$$E=A_1^{(i)}+A_2^{(i)}+\cdots+A_{r_i}^{(i)},\quad i=1,2,\ldots,n$$

由此定义 $r=r_1r_2\ldots r_n$ 种概率：

$$p_{q_1q_2\ldots q_n}=P(A_{q_1}^{(1)}A_{q_2}^{(2)}\ldots A_{q_n}^{(n)})\geq 0$$

这些概率是任意值，但需要满足下面的条件[10]：

$$\sum_{q_1,q_2,\dots,q_n} p_{q_1q_2\dots q_n} = 1 \tag{1}$$

定义 I　当对于任意的 $q_1,q_2,\dots,q_n$，以下等式总成立时，称 n 个试验 $\mathscr{U}^{(1)},\mathscr{U}^{(2)},\dots,\mathscr{U}^{(n)}$ 相互独立：

$$P(A_{q_1}^{(1)}A_{q_2}^{(2)}\dots A_{q_n}^{(n)}) = P(A_{q_1}^{(1)})P(A_{q_2}^{(2)})\dots P(A_{q_n}^{(n)}) \tag{2}$$

在(2)中的 r 个方程中，仅有 $r-r_1-r_2-\dots-r_n+n-1$ 个独立方程[11]。

定理 I　如果 n 个试验 $\mathscr{U}^{(1)},\mathscr{U}^{(2)},\dots,\mathscr{U}^{(n)}$ 相互独立，则其中的任意 m 个（$m<n$）试验 $\mathscr{U}^{(i_1)},\mathscr{U}^{(i_2)},\dots,\mathscr{U}^{(m)}$ 也相互独立[12]。在独立性的情况下，有以下方程：

$$P(A_{q_1}^{(i_1)}A_{q_2}^{(i_2)}\dots A_{q_m}^{(i_m)}) = P(A_{q_1}^{(i_1)})P(A_{q_2}^{(i_2)})\dots P(A_{q_m}^{(i_m)}) \tag{3}$$

（所有的 i_k 都互不相同）。

定义 II　对于 $k=1,2,\dots,n$，如果以下分解/试验成立：

$$E = A_k + \bar{A}_k$$

则称 n 个事件 $A_1,A_2,\dots,A_n$ 相互独立。

在这种情况下，$r_1=r_2=\dots=r_n$，$r=2^n$；因此 公式(2)中的 2^n 个方程中，只有 2^n-n-1 个是独立的。事件 $A_1,A_2,\dots,A_n$ 独立的充分必要条件是满足以下的 2^n-n-1[13]：

$$\begin{gathered} P(A_{i_1}A_{i_2}\dots A_{i_m}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_m}) \\ m=1,2,\dots,n, \\ 1\le i_1\le i_2<\dots<i_m\le n \end{gathered} \tag{4}$$

以上所有方程都相互独立。

当 $n=2$ 时，由公式 (4) 可得两个事件 A_1 和 A_2 独立的一个条件（2^2 - 2 - 1 = 1）：

$$P(A_1A_2) = P(A_1)P(A_2) \tag{5}$$

在这种情况下（$n=2$），方程系统 (2) 退化为三个方程（除方程(5)外）：

$$\begin{gathered} P(A_1\bar{A}_2) = P(A_1)P(\bar{A}_2) \\ P(\bar{A}_1A_2) = P(\bar{A}_1)P(A_2) \\ P(\bar{A}_1\bar{A}_2) = P(\bar{A}_1)P(\bar{A}_2) \end{gathered}$$

以上公式显然可以由公式 (5) 导出[14]。

需要强调的是，事件 $A_1,A_2,\dots,A_n$ 两两相互独立，例如以下关系

$$P(A_iA_j) = P(A_i)P(A_j) \qquad (i\ne j)$$

并不能推导出当 $n>2$ 时这些事件相互独立[15]。（当公式 (4) 中的方程全部成立时，事件$A_1,A_2,\dots,A_n$ 才相互独立）。

在引入独立性的概念时，并没有使用条件概率。我们的目标是用纯数学的方式，尽可能清楚地解释这个概念的含义。然而，它的应用通常依赖于某些条件概率的性质。

如果我们假设所有的概率 $P(A_q^{(i)})$ 都为正，则从公式 (3) 可以推导出[16]

$$P_{A_{q_1}^{(i_1)}A_{q_2}^{(i_2)}\dots A_{q_{m-1}}^{(i_{m-1})}}(A_{q_m}^{(i_m)}) = P(A_{q_m}^{(i_m)}) \tag{6}$$

由公式 (6) 成立，以及第 1.4 节（公理的直接推论；条件概率；贝叶斯理论）公式 (7) 可得公式 (2)。因此，我们得到定理 II。

定理 II：当概率 $P(A_q^{(i)})$ 为正时，试验 $\mathscr{U}^{(1)},\mathscr{U}^{(2)},\dots,\mathscr{U}^{(n)}$ 独立的必要和充分条件是，在其他试验 $\mathscr{U}^{(i_1)},\mathscr{U}^{(i_2)},\dots,\mathscr{U}^{(i_k)}$ 有确定的结果 $A_{q_1}^{(i_1)},A_{q_2}^{(i_2)},A_{q_3}^{(i_3)},\dots,A_{q_k}^{(i_k)}$ 的假设下，试验 $\mathscr{U}^{(i)}$ 的结果 $A_q^{(i)}$ 的条件概率等于该结果的绝对概率 $P(A_q^{(i)})$。

在公式 (4) 的基础上，我们可以以类比的方式证明如下定理：

定理 III：如果所有的概率 $P(A_K)$ 都为正值，则事件 $A_1,A_2,\dots,A_n$ 相互独立的必要和充分条件是，对于任意不同的 $i_1,i_2,\dots,i_k,i$，如下等式成立：

$$P_{A_{i_1}A_{i_2}\ldots A_{i_k}}(A_i) = P(A_i) \tag{7}$$

当 $n=2$ 时，公式 (7) 退化为两个等式：

$$\begin{aligned} P_{A_1}(A_2) = P(A_2) \\ P_{A_2}(A_1) = P(A_1) \end{aligned} \tag{8}$$

很容易看出，公式 (8) 中的第一个等式是当 $P(A_i) > 0$ 时，事件 A_1 和 A_2 独立的必要和充分条件。

作为随机变量的条件概率；马尔可夫（Markov）链

假设 $\mathscr{U}$ 为基本集合 E 的一个分解：

$$E = A_1 + A_2 + \ldots + A_n$$

并假设 x 为基本事件 ξ 的一个实函数（real function），即对于每一个集合 A_q，都有一个常数 a_q 与它对应。x 称为随机变量（random variable），以下等式

$$E(x) = \sum_q a_q P(A_q)$$

称为变量 x 的数学期望（mathematical expectation）。随机变量的理论将在第三章和第四章介绍。我们不应该仅局限于只能取有限个值的随机变量。

对应集合 A_q 的随机变量记为 $P_{A_{q_i}}(B)$，我们称其为当试验 $\mathscr{U}$ 给定时事件 B 的条件概率，并用 $P_{\mathscr{U}}(B)$。当且仅当下面公式成立时，两个试验 $\mathscr{U}^{(1)}$ 和 $\mathscr{U}^{(2)}$ 独立。

$$P_{\mathscr{U}^{(1)}}(A_q^{(2)}) = P(A_q^{(2)}) \qquad q = 1, 2, \ldots, r_2$$

给定任意试验的分解 $\mathscr{U}^{(1)}, \mathscr{U}^{(2)}, \ldots, \mathscr{U}^{(n)}$，我们用以下公式

$$\mathscr{U}^{(1)}\mathscr{U}^{(2)}\ldots\mathscr{U}^{(n)}$$

表示基本集合 E 的分解的乘积：

$$A_{q_1}^{(1)}A_{q_2}^{(2)}\ldots A_{q_n}^{(n)}$$

试验 $\mathscr{U}^{(1)}, \mathscr{U}^{(2)}, \ldots, \mathscr{U}^{(n)}$ 相互独立，当且仅当以下公式成立：

$$P_{\mathscr{U}^{(1)}\mathscr{U}^{(2)}\ldots\mathscr{U}^{(k-1)}}(A_q^{(k)}) = P(A_q^{(k)})$$

其中 k 和 q 为任意值[17]。

定义：当对于任意的 n 和 q，如果以下等式

$$P_{\mathscr{U}^{(1)}\mathscr{U}^{(2)}\ldots\mathscr{U}^{(n-1)}}(A_q^{(n)}) = P_{\mathscr{U}^{(n-1)}}(A_q^{(n)})$$

成立，则序列 $\mathscr{U}^{(1)}, \mathscr{U}^{(2)}, \ldots, \mathscr{U}^{(n)}, \ldots$ 构成一个马尔可夫链（Markov chain）。

因此，马尔可夫链是相互独立的试验的序列的推广。如果假设

$$p_{q_m,q_n}(m,n) = P_{A_m^{(m)}}(A_{q_n}^{(n)}) \qquad m < n$$

则马尔可夫链理论的基本方程为以下形式：

$$p_{q_k,q_n}(k,n) = \sum_{q_m} p_{q_k,q_m}(k,m)\, p_{q_m,q_n}(m,n) \qquad k < m < n \tag{1}$$

如果把矩阵 $||p_{q_m,q_n}(m,n)||$ 记为 $p(m,n)$，则公式 (1) 可以记为[18]：

$$p(k,n) = p(k,m)\, p(m,n) \qquad k < m < n \tag{2}$$

$$P(A_{q_1}^{(1)}A_{q_2}^{(2)}\ldots A_{q_{n-1}}^{(n-1)}) = \sum_{q_n} P(A_{q_1}^{(1)}A_{q_2}^{(2)}\ldots A_{q_n}^{(n)}) = P(A_{q_1}^{(1)})P(A_{q_2}^{(2)})\ldots P(A_{q_{n-1}}^{(n-1)}) \sum_{q_n} P(A_{q_n}^{(n)}) = P(A_{q_1}^{(1)})P(A_{q_2}^{(2)})\ldots I$$

$$A = \{\xi_1, \xi_2\}, \quad B = \{\xi_1, \xi_3\}, \quad C = \{\xi_1, \xi_4\}$$

由此很容易可以算出（注意 $P(AB)$ 表示事件 A 和 B 同时发生，即基本事件 ξ_1 发生，所以为 $1/4$）。

$$P(A)=P(B)=P(C)=1/2$$
$$P(AB)=P(BC)=P(AC)=1/4=(1/2)^2$$
$$P(ABC)=1/4\neq(1/2)^3$$

无限概率域

连续性公理

按照惯例，我们用符号 $\underset{m}{\mathscr{D}}A_m$ 表示集合 A_m 的乘积（无论是有限个还是无限个），用符号 $\underset{m}{\mathscr{G}}A_m$ 表示集合 A_m 的和。仅当集合 A_m 为不相交集（disjont sets，集互斥集合），集合的和的形式记为 $\sum\limits_m A_m$。由此可得：

$$\underset{m}{\mathscr{G}}A_m=A_1\dot{+}+A_2\dot{+}\dots$$
$$\sum_m A_m=A_1+A_2+\dots$$
$$\underset{m}{\mathscr{D}}A_m=A_1A_2\dots$$

在后续的研究中，除了公理 I-V 外，我们认为以下公理也成立：

公理 VI：对于集合 $\mathcal{E}$ 中的一个递减的事件序列

$$A_1\supset A_2\supset\cdots\supset A_n\supset\dots\tag{1}$$
$$\underset{n}{\mathscr{D}}A_n=0\tag{2}$$

当公式 (2) 成立时，以下等式（公式 (3)）成立。

$$\lim_{n\to\infty}P(A_n)=0\tag{3}$$

在后续的研究中，我们将满足第一章第1节中五个公理，以及本章开头提到的公理 VI 的**集合系统**及其对应的$P(A)$，称为概率域。而第一章中原本定义的概率域称为广义概率域（generalized fields of probability）。

如果集合系统 $\mathcal{E}$ 是有限的，公理 VI 可以由公理 I-V 推导得出。事实上，这种情况下在序列 (1) 中仅存在有限个不同的集合。假设 A_k 是该序列中最小的（集合的大小指集合的**基数**），则所有与 A_k 恰巧（基数）相等的集合 A_{p+k}，有以下关系：

$$A_k=A_{k+p}=\underset{n}{\mathscr{D}}A_n=0$$
$$\lim P(A_n)=P(0)=0$$

因此，第一章中提到的所有有限概率域的例子都满足公理 VI。已证明，公理系统 I-VI 是一致且不完备的（consitent and incomplete）。

然而，对于无限概率域，连续性公理（公理VI）已被证明是独立于公理 I-V 的。因为这一新的公理仅对于无限概率域是必要的，所以几乎无法像第一章第 2 节那样中，对公理 I-V 那样，解释清楚公理 VI 的经验意义（empirical meaning）。因为在描述任何可观测的随机过程时，我们只能得到有限概率域。无限概率域只出现在现实中的随机过程的理想化模型中。后续的讨论，我们仅限于满足公理 VI 的模型。这一限制（尽管有些随意）在研究绝大多数的模型时有效且可行的。

广义加法定理：如果 $A_1,A_2,\dots,A_n,\dots$ 以及 A 属于 $\mathcal{E}$，则由公式(4)

$$A=\sum_n A_n\tag{4}$$

可得公式(5)

$$P(A)=\sum_n P(A_n)\tag{5}$$

证明：令

$$R_n=\sum_{m>n}A_m$$

则显然

$$\underset{n}{\mathscr{D}}(R_n)=0$$

因此根据公理 VI 得

$$\lim P(R_n) = 0 \qquad n \to \infty \tag{6}$$

另一方面，根据加法定理得

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_n) + P(R_n) \tag{7}$$

由公式 (6) 和公式 (7) 可得公式 (5)。

由此已表明，概率 $P(A)$ 是一个集合系统 $\mathcal{E}$ 上的完全可加集函数。公理 V 和 VI 对于定义在任意 $\mathcal{E}$ 上的完全可加集函数都成立[19]。因此可以通过以下方式定义概率域的概念：

> 令 E 为任意集合，$\mathcal{E}$ 为 E 的包括其本身在内的子集组成的域，以及 $P(A)$ 为一个定义在 $\mathcal{E}$ 上的非负的完全可加集函数。则 $\mathcal{E}$ 与集合函数 $P(A)$ 共同构成一个概率域。

收敛性定理：如果 $A_1, A_2, \ldots, A_n, \ldots$ 以及 A 属于 $\mathcal{E}$，并且

$$A \subset \underset{n}{\mathscr{S}} A_n \tag{8}$$

则

$$P(A) \leq \sum_n P(A_n) \tag{9}$$

证明：

$$A = A \underset{n}{\mathscr{S}} A_n = A(A_1 + A_2(1 - A_1) + A_3(1 - A_2 - A_1)) + \cdots = AA_1 + A(A_2 - A_2A_1) + A(A_3 - A_3A_2 - A_3A_1) -$$

$$P(A) = P(AA_1) + P(A(A_2 - A_2A_1)) + \cdots \leq P(A_1) + P(A_2) + \ldots$$

概率的波莱尔（Borel）域

对于 $\mathcal{E}$ 中的集合 A_n，如果集合 A_n 的可数和 $\sum_n A_n$ 也属于 $\mathcal{E}$，则称 $\mathcal{E}$ 为一个波莱尔域。波莱尔域也称为**完全可加的集合系统**（completely additive systems of sets）。由以下等式

$$\underset{n}{\mathscr{S}} A_n = A_1 + (A_2 - A_2A_1) + (A_3 - A_3A_2 - A_3A_1) + \ldots \tag{1}$$

可推断出，波莱尔域包含了所有的 $\underset{n}{\mathscr{S}} A_n$，后者由波莱尔域中的可数个集合 A_n 组成。

由以下等式

$$\underset{n}{\mathscr{D}} A_n = E - \underset{n}{\mathscr{S}} A_n \tag{2}$$

可知，对于集合 A_n 的乘积 $\underset{n}{\mathscr{D}} A_n$ 也有相同的结论。

当域 $\mathcal{E}$ 为波莱尔域时，对应的概率域称为波莱尔概率域。只有在波莱尔概率域下，我们才可以自由地研究概率，而不用担心**没有概率的事件**。接下来我们将从**扩展定理**开始证明，后续讨论都将在波莱尔概率域中进行。

给定一个概率域 $(\mathcal{E}, P)$，可知[20]存在一个包含域 $\mathcal{E}$ 的最小波莱尔域 $B\mathcal{E}$，我们有以下扩展定理：

扩展定理：总是可以将定义在域 $\mathcal{E}$ 上的非负完全可加集函数 $P(A)$ 扩展到所有的最小波莱尔域 $B\mathcal{E}$，且不丢失它的任何性质（非负性、完全可加性）。且只有一种方式可以做到。

扩展域 $B\mathcal{E}$ 与扩展后的集合函数 $P(A)$ 一起构成了概率域 $(B\mathcal{E}, P)$。这个概率域称为域 $(\mathcal{E}, P)$ 的波莱尔扩展。

这个定理的证明属于可加集函数理论的范畴，并且有时以其他的形式出现。此处给出一种形式的证明：

令 A 为集合 E 的任一子集，用 $P^*(A)$表示以下和式的下限

$$\sum_n P(A_n)$$

对于集合 A 的所有由有限个或可数多个 $\mathcal{E}$ 中的集合 A_n 构成的覆盖的覆盖（coverings）：

$$A \subset \underset{n}{\mathscr{S}} A_n$$

很容易证明，$P^*(A)$ 是 Carathéodory 外测度[21]。根据收敛定理（本章第一节），对于所有 $\mathcal{E}$ 中的集合，$P^*(A)$ 恰巧等于 $P(A)$。进一步可以证明，所有 $\mathcal{E}$ 中的集合在 Carathéodory 测度下都是可测的。因为所有的可测集构成了波莱尔域，所以域 $B\mathcal{E}$ 中的集合也是可测的。因此集合函数 $P^*(A)$ 在 $B\mathcal{E}$ 上是完全可加的。在域 $B\mathcal{E}$ 上，我们令

$$P(A) = P^*(A)$$

由此我们证明了扩展域的存在。扩展域的唯一性来源于域 $B\mathcal{E}$ 是最小波莱尔域。

备注：即使域 $\mathcal{E}$ 中的集合（事件）A 是真是存在的、可观测的事件，但这并不意味着扩展域中的集合也是真实存在的、可观测的。

因此存在这样的可能性：当概率域 $(\mathcal{E}, P)$ 可以视为现实中的随机事件的镜像时，扩展概率域 $(B\mathcal{E}, P)$ 仍然只具有数学上的结构，而没有现实的概率意义。

因此域 $(B\mathcal{E}, P)$ 中的集合通常只是理想事件，它们并没有在现实世界中对应的对象。然而，如果利用这些理想事件的概率进行的推演，能够让我们得出现实世界中的随机事件的概率，那么从经验的角度来看，这并不矛盾【利用只在数学概念中存在的概率，计算出现实世界中事件的概率】。

无限概率域举例

I. 在本章第 1 节中，我们构造了很多有限概率域。现在假定 $E = \xi_1, \xi_2, \ldots, \xi_n, \ldots$ 为一个可数集合，并令 $\mathcal{E}$ 为 E 的所有子集的集合。基于 $\mathcal{E}$，所有可能的概率域可以按如下方式得出：

给定一个非负实数序列 p_n：

$$p_1 + p_2 + \cdots + p_n + \cdots = 1$$

且对于每个集合 A，定义

$$P(A) = {\sum_n}' p_n$$

其中求和符号 $\sum'$ 作用于所有的属于 A 的 ξ_n 的下标。这些概率域显然是波莱尔域。

II. 在这个例子中，我们应当假设 E 代表实数轴。首先假设 $\mathcal{E}$ 由所有可能的半开区间 $[a_i; b) = \{a \leq \xi < b\}$ [22] 的有限和组成（不仅仅只考虑由有限实数 a 和 b 构成的常规区间，也考虑如 $[-\infty; a), [a; +\infty), [-\infty; +\infty)$ 这类的非常规区间）。这样，$\mathcal{E}$ 即为一个域。然而，根据扩展定理，任何定义在 $\mathcal{E}$ 上的概率域都可扩展为一个定义在 $B\mathcal{E}$ 上的类似的域。因此在这种情况下，集合系统 $B\mathcal{E}$ 实际上无非是实数轴上所有波莱尔点集的系统。接下来考虑下面的情况。

III. 同样假设 E 为实数轴，$\mathcal{E}$ 由这条轴上的所有波莱尔点集构成。想要构造给定域 $\mathcal{E}$ 上的概率域，只需在 $\mathcal{E}$ 上定义一个任意的非负的完全可加集函数 $P(A)$，且满足 $P(E) = 1$。已知这样的函数由其在区间 $[-\infty; x)$ 上的取值唯一确定 [23]：

$$P[-\infty; x) = F(x) \tag{1}$$

函数 $F(x)$ 称为 ξ 的分布函数。之后在第三章第 2 节我们可以证明 $F(x)$ 是左连续的非减函数，且具有以下的极限值：

$$\begin{gathered}\lim_{x \to -\infty} F(x) = F(-\infty) = 0, \\ \lim_{x \to +\infty} F(x) = F(+\infty) = 1\end{gathered} \tag{2}$$

反过来，如果一个函数 $F(x)$ 满足这些条件，则它总是可以确定一个非负的完全可加集函数 $P(A)$，使得 $P(E) = 1$ [24]。

IV. 现在假设基本集合 E 为一个 n 维欧几里得空间 R^n，即集合中的每个元素 ξ 是一个由实数构成的含有 n 个元素的有序元组（tuples）：$\xi = \{x_1, x_2, \ldots, x_n\}$。假设 $\mathcal{E}$ 由欧几里得空间 R^n 中所有的波莱尔点集 [25] 组成。根据与示例 II 中使用的类似推理，我们不需要研究更狭义的集合系统，例如 n 维区间的系统。概率函数 $P(A)$ 同样应该是一个定义在 $\mathcal{E}$ 上的非负的完全可加集函数，且满足 $P(E) = 1$。这样的集合函数的取值由特定集合 $L_{a_1, a_2, \ldots, a_n}$ 确定：

$$P(L_{a_1, a_2, \ldots, a_n}) = F(a_1, a_2, \ldots, a_n) \tag{3}$$

其中 $L_{a_1, a_2, \ldots, a_n}$ 表示所有满足 $x_i < a_i (i = 1, 2, \ldots, n)$ 的 ξ 的集合。

对于 $F(a_1, a_2, \ldots, a_n)$，我们选择的函数需要满足对于每一个变量都是左连续的非减函数，且满足以下条件：

$$\begin{gathered}\lim_{a_i \to -\infty} F(a_1, a_2, \ldots, a_n) = F(a_1, \ldots, a_{i-1}, -\infty, a_{i+1}, \ldots, a_n) = 0, \\ (i = 1, 2, \ldots, n) \\ \lim_{a_1 \to +\infty, a_2 \to +\infty, \ldots, a_n \to +\infty} F(x) = F(+\infty, +\infty, \ldots, +\infty) = 1\end{gathered} \tag{4}$$

> 注意以上公式中。$F(x) = 0$ 的极限，只需有其中一个 a_i 趋近于负无穷即可；$F(x) = 1$ 的极限,需要所有的 a_i 都趋近于正无穷。

【注】参考的两个版本在公式 (4) 上有出入，这里贴了一个我能看得懂的，下面图中是另一个版本的我看不懂的公式：

$$\left.\begin{array}{r}\lim\limits_{a_i\to-\infty} F(a_1,a_2,\ldots,a_n)=F(a_1,\ldots,a_{i-1},-\infty,a_{i+1},\ldots,a_n)=0,\\ i=1,2,\ldots,n\\ \lim\ F(a_1,a_2,\ldots,a_n)=F(+\infty,+\infty,\ldots,+\infty)=1.\end{array}\right\}\ 4)$$

$$\sum_{\substack{i=1\\ \varepsilon_i=0,1}}^{n}(-1)^{\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n}F(a_1-\varepsilon_1 c_1,a_2-\varepsilon_2 c_2,\ldots,a_n-\varepsilon_n c_n)\geqq 0,$$
$$c_i>0,\quad i=1,2,3,\ldots,n.$$

$F(a_1,a_2,\ldots,a_n)$ 称为变量 $x_1,x_2,\ldots,x_n$ 的分布函数。

对上述类型的概率域的研究足以解决概率理论中的所有经典问题[26]。特别地，R^n 上的概率函数可以定义为：

定义在 R^n 上的任意非负点函数：

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n=1$$

并且令

$$P(A)=\int\int\cdots\int_A f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n \tag{5}$$

$f(x_1,x_2,\ldots,x_n)$ 此时称为在点 $(x_1,x_2,\ldots,x_n)$ 处的概率密度（参考第三章第 2 节）。

另一类 R^n 上的概率函数可以由以下方式得出：假设 ξ_i 为 R^n 上的一个点的序列，令 p_i 为一个非负实数的序列，使得 $\sum P_i=1$。此时如例子 I 中那样，令

$$P(A)=\sum{}' p_i$$

其中求和符号 $\sum{}'$ 作用于所有属于 A 的 ξ 的下标。

以上提到的两种定义在 R^n 上的概率函数并没有涵盖所有可能的情况，但是通常认为这两种定义对于概率理论的应用已经足够。然而，我们可以设想一些在经典范围之外的应用问题，其中基本事件通过无限多个坐标来定义。我们将在引入为此目的所需的若干概念后，进一步详细研究相应的概率场（参考第三章第 3 节）。

随机变量

概率函数

给定一个由任意类型元素构成的集合 E 到 E' 的映射，例如定义在 E 上的单值函数 $u(\xi)$，其值属于 E'。对于 E' 的每一个子集 A'，我们将与之对应的 E 中的原像（pre-image）记为 $u^{-1}(A')$，它包含所有映射到 A' 中元素的 E 中的元素。令 $\mathcal{E}^{(u)}$ 为E' 的所有子集 A' 组成的系统，其中 A' 的原像属于域 $\mathcal{E}$，则 $\mathcal{E}^{(u)}$ 也是一个域。如果 $\mathcal{E}$ 恰巧是一个波莱尔域，则 $\mathcal{E}^{(u)}$ 也是一个波莱尔域。由此可令：

$$P^{(u)}(A')=P\{u^{-1}(A')\} \tag{1}$$

因为定义在 $\mathcal{E}^{(u)}$ 上的集合函数 $P^{(u)}$ 满足公理 I-VI，所以它可以作为 $\mathcal{E}^{(u)}$ 上的概率函数。在证明以上内容之前，我们需要先给出下面的定义：

定义：给定一个随机事件 ξ 的单值函数 $u(\xi)$，公式 (1) 中定义的函数 $P^{(u)}(A')$ 称为 u 的概率函数。

备注 1：在研究概率域 $(\mathcal{E},P)$ 时，我们只简单地把 $P(A)$ 称为概率函数。但是我们称 $P^{(u)}(A')$ 为u 的概率函数。当 $u(\xi)=\xi$ 时，$P^{(u)}(A')$ 恰巧与 $P(A)$ 相等。

备注 2：事件 $u^{-1}(A')$ 表明 $u(\xi)$ 属于 A'，因此 $P^{(u)}(A')$ 表示 $u(\xi)$ 属于 A' 的概率。

我们仍需证明以上提到的 $\mathcal{E}^{(u)}$ 和 $P^{(u)}$ 的性质，不过它们实际上来源于一个简单的事实：

引理：原像集合 $u^{-1}(A')$ 的和、乘积、差，等于对应原始集合 A' 的和、乘积和差的原像。

以上引理的证明过程留给读者。

【补充证明过程】

令 A' 和 B' 为域 $\mathcal{E}^{(u)}$中 的两个集合，它们的原像 A 和 B 则属于 $\mathcal{E}$。由于 $\mathcal{E}$ 构成一个域，则集合 AB、$A+B$ 、$A-B$ 也属于 $\mathcal{E}$；同时这三个集合也分别是集合 $A'B'$、$A'+B'$、$A'-B'$ 的原像（后三者属于域 $\mathcal{E}^{(u)}$）。由此证明 $\mathcal{E}^{(u)}$ 是一个域。同样地，可以证明如果 $\mathcal{E}$ 是一个波莱尔域，则 $\mathcal{E}^{(u)}$ 也是。进一步可得

$$P^{(u)}(E') = P\{u^{-1}(E')\} = P(E) = 1$$

显然 $P^{(u)}$ 总是非负的。因此接下来只需证明，$P^{(u)}$ 是完全可加的（参考第二章第 1 节结尾）。

假设所有的集合 A'_n，以及它们各自的原像 $u^{-1}(A'_n)$ 是互斥的（对于所有的 n， A'_n 两两互斥，原像同理），则有

$$P^{(u)}(\sum_n A'_n) = P\{u^{-1}(\sum_n A'_n)\} = P\{\sum_n u^{-1}(A'_n)\} = \sum_n P\{u^{-1}(A'_n)\} = \sum_n P^{(u)}(A'_n)$$

由此证明了$P^{(u)}$ 的完全可加性。

最后，还有一点需要说明。令 $u_1(\xi)$ 是由 E 映射到 E' 的函数，$u_2(\xi')$ 是由 E' 映射到 E'' 的函数。则乘积函数 $u_2u_1(\xi)$ 从 E 映射到 E''。接下来研究概率函数 $P^{(u_1)}(A')$ 和 $P^{(u)}(A'')$，其中 $u(\xi) = u_2u_1(\xi)$。很容易得出，这两个概率函数有以下关联：

$$P^{(u)}(A'') = P^{(u_1)}\{u_2^{-1}(A'')\} \tag{2}$$

随机变量和分布函数的定义

多维分布函数

无限维空间中的概率

等价随机变量及多种收敛

数学期望

抽象勒贝格（Lebesgue）积分

绝对数学期望/条件数学期望

切比雪夫（Tchebycheff）不等式

收敛判据

数学期望对参数的微分和积分

条件概率和数学期望

条件概率

波莱尔悖论的解释

随机变量的条件概率

条件数学期望

独立性；大数定律

独立性

随机变量的独立性

大数定律

关于数学期望概念的说明

强大数定律；级数收敛

附录：概率理论中的零一定律

参考文献

关于补充参考文献的说明

补充参考文献

注

---

1. 原始德文版本发表于1933年。 ↩
2. 暂不确定此人是谁。 ↩
3. 原文脚注1，Zur Statistik der kontinuierlichen Systeme und des zeitlichen Verlaufes der physikalischen Vorgange. Phys. Jour, of the USSR, Vol. 3, 1933, pp. 35-63. 标题英文为 *On the statistics of continuous systems and the temporal evolution of physical processes*. ↩
4. 见参考文献 R. von Mises [1] and [2] and S. Bernstein [1]. ↩
5. 希望从一开始就为以下公理赋予具体含义的读者，可以参考第二章（与实验数据的关系）。 ↩
6. 参考 HAUSDORFF, Mengenlehre, 1927, p. 78. 如果一个集合系统中的任意两个集合的和、差、积也属于这个集合系统，则这个集合系统称为一个域。任意非空集合包含空集。使用豪斯多夫（Hausdorff）的标记，把 A 和 B 的乘积记为 AB；当 $AB=0$ 时和记为 $A+B$；广义的和记为 $A+B$；A 与 B 的差记为 $A-B$。集合 $E-A$ 表示集合 A 的补集，记为 $\bar{A}$。我们假定读者熟悉集合以及集合的和、差、乘积的基本运算规则。所有 $\mathcal{E}$ 的子集都用大写拉丁字母表示。 ↩
7. 只对概率论的纯数学发展感兴趣的读者可以略过此节。接下来的内容都只基于第一节的公理，并不利用当前的讨论。第二章，我们仅限于简要解释概率论的公理如何产生，并忽略了关于经验世界中概率概念的深刻哲学论述。在建立概率论应用于现实世界的前提条件时，作者在很大程度上使用了冯·米泽斯的工作，[1] 第21-27页。 ↩
8. 第四节公式（3） ↩
9. 见 Fréchet [1] 和 [2]. ↩
10. 在满足上面提到的条件下，可以用任意概率值构建一个概率域：E 由 r 个元素组成，$\xi_{q_1 q_2 \ldots q_n}$。令每一个元素对应的基本概率为 $p_{q_1 q_2 \ldots q_n}$，则 $A_{q_n}^{(i)}$ 表示当 $q_i = q$ 时的所有 $\xi_{q_1 q_2 \ldots q_n}$ 组成的集合。 ↩
11. 事实上在独立的情况下，可以只选取 $r_1 + r_2 + \ldots r_n$ 种概率，$p_q^{(i)} = P(A_q^i)$，以遵循 n 个条件 $\sum_q p_q^{(i)} = 1$，因此在一般情形下有 $r-1$ 个自由度，但在独立情形下只有 $r_1 + r_2 \ldots + r_n - n$ 个自由度。 ↩
12. 要证明这一点，只需表明，从（概率空间的） n 个分解的相互独立性可以推导出前 $n-1$ 个分解的相互独立性（关于 n 个分解的解释：In the context of probability and statistics, **n decompositions** typically refers to the partitioning or breaking down of a probability space into nnn mutually exclusive and exhaustive events or sets.）假如公式 (2) 正确，则有： ↩
13. 参考 S. N. Bernstein [1] pp. 47-57。不过读者应该可以很容易采用数学归纳法自己证明这一推论。 ↩
14. $P(A_1\bar{A}_2) = P(A_1) - P(A_1A_2) = P(A_1) - P(A_1)P(A_2) = P(A_1)\{1 - P(A_2)\} = P(A_1)P(\bar{A}_2)$，其他公式也是类似的推导方法。 ↩
15. 以下为一个简单的证明（来自 S. N. Bernstein）：假设集合 E 包含 4 个元素 $\xi_1, \xi_2, \xi_3, \xi_4$；并假设对应的基本概率 P_1, P_2, P_3, P_4 均为 $1/4$，并且有 ↩
16. 想要证明该公式，读者需要记住条件概率的定义（1.4节公式 (5)），并将乘积的概率 $P(AB)$ 替换为公式 (3)。 ↩
17. 这些条件的必要性可由第 5 节的定理 II 导出；其充分性可以从乘法定理（第 4 节公式 (7)）得出。 ↩
18. 关于马尔可夫链理论更深入的阐述，参考 v. Mises [1], § 16，以及 B. HOSTINSKY, Méthodes générales du calcul des probabilités, “Mém. Sci. Math.” V. 52, Paris 1931。 ↩
19. See, for example, O. NIKODYM, Sur une généralisation des intégrales de M. J. Radon, Fund. Math. v. 15, 1930, p. 136. ↩
20. 参考 HAUSDORFF, Mengenlehre, 1927, p. 85. ↩
21. CARATHÉODORY, Vorlesungen über reelle Funktionen, pp.237-258. (New York, Chelsea Publishing Company). ↩
22. 此处原文，区间表示方式，中间为`;`。 ↩
23. Cf., for example, LEBESGUE, Leçons sur l’intégration, 1928, p. 152-156. ↩

24. 参考前一个注释中的内容。 ↩

25. For a definition of Borel sets in R see HAUSDORFF, Mengenlehre, 1927, pp. 177-181. ↩

26. Cf., for example, R. v. MISES [1], pp. 13-19. Here the existence of probabilities for “all practically possible:” sets of an n-dimensional space is required. ↩