

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

概率理论基础

编者注

前言

初等概率论

公理

与实验数据的关系

术语说明

公理的直接推论; 条件概率; 贝叶斯理论

独立性

作为随机变量的条件概率; 马尔可夫 (Markov) 链

无限概率域

连续性公理

概率的波莱尔 (Borel) 域

无限概率域举例

随机变量

概率函数

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

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

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, \dots 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 I-V 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：

$$\begin{aligned} P(E) &= 1 \\ P(\emptyset) &= 0 \end{aligned}$$

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

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

以上公式中， p_1, p_2, \dots, p_k 称为基本事件 $\xi_1, \xi_2, \dots, \xi_k$ 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） ξ_1, ξ_2, \dots 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, \dots 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. $AB \dots N = 0$ 。	2. 事件 A, B, \dots, N 都互斥。
3. $AB \dots N = X$ 。	3. 当事件 A, B, \dots, N 同时发生时，事件 X 发生。
4. $A + B + \dots + N = X$ 。	4. 当事件 A, B, \dots, 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。

集合论	随机事件
7. $A = E$ 。	7. 事件 A 必然发生。
8. 当 $A_1 + A_2 + \dots + A_n = E$ 时, 称集合 A_1, A_2, \dots, 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, \dots, A_n 中哪些发生, 则称事件 A_1, A_2, \dots, 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:

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

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

因为 $\bar{E} = 0$, 所以有:

$$P(0) = 0 \quad (3)$$

如果事件 A, B, \dots, N 互不相容, 则由公理 V 可得:

$$P(A + B + \dots + N) = P(A) + P(B) + \dots + P(N) \quad (4)$$

如果 $P(A) > 0$, 则公式

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

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:

$$P(AB) = P(A)P_A(B) \quad (6)$$

采用归纳法, 可得上式的更广义的形式 (乘法定理) :

$$P(A_1 A_2 \dots A_n) = P(A_1)P_{A_1}(A_2)P_{A_1 A_2}(A_3) \dots P_{A_1 A_2 \dots A_{n-1}}(A_n) \quad (7)$$

条件概率的广义形式。当有 n 个事件 A_1, A_2, \dots, 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, \dots, 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:

$$P_A(B) \geq 0 \quad (8)$$

$$P_A(E) = 1 \quad (9)$$

$$P_A(B + C) = P_A(B) + P_A(C) \quad (10)$$

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:

$$P_A(A) = 1 \quad (11)$$

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:

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

即，**贝叶斯定理** (the Theorem of Bayes)。

全概率定理 (The Theorem on Total Probability)

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

$$P(X) = P(A_1)P_{A_1}(X) + P(A_2)P_{A_2}(X) + \dots + P(A_n)P_{A_n}(X) \quad (13)$$

证明:

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

由公式(4)可得

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

由公式(6)可得

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

贝叶斯定理

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

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

上式中, A_1, A_2, \dots, 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 Fréchet⁹ 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, $\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \dots, \mathcal{U}^{(n)}$, 即基本集合 E 的 n 个分解:

$$E = A_1^{(i)} + A_2^{(i)} + \dots + A_{\tau_i}^{(i)}, \quad i = 1, 2, \dots, n$$

由此定义 $r = r_1 r_2 \dots r_n$ 种概率:

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

这些概率是任意值, 但需要满足下面的条件¹⁰:

$$\sum_{q_1, q_2, \dots, q_n} p_{q_1 q_2 \dots q_n} = 1 \quad (1)$$

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 $\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \dots, \mathcal{U}^{(n)}$ 相互独立:

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

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

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

(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:

$$\begin{aligned} P(A_{i_1} A_{i_2} \dots A_{i_m}) &= P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_m}) \\ m &= 1, 2, \dots, n, \\ 1 \leq i_1 \leq i_2 < \dots < i_m &\leq n \end{aligned} \quad (4)$$

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 ($2^2 - 2 - 1 = 1$):

$$P(A_1 A_2) = P(A_1) P(A_2) \quad (5)$$

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):

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

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

$$P(A_i A_j) = P(A_i) P(A_j) \quad (i \neq j)$$

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

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

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 $\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \dots, \mathcal{U}^{(n)}$ 独立的必要和充分条件是, 在其他试验 $\mathcal{U}^{(i_1)}, \mathcal{U}^{(i_2)}, \dots, \mathcal{U}^{(i_k)}$ 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 $\mathcal{U}^{(i)}$ 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:

$$P_{A_{i_1} A_{i_2} \dots A_{i_k}}(A_i) = P(A_i) \quad (7)$$

当 $n = 2$ 时, 公式 (7) 退化为两个等式:

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

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 \mathcal{U} 为基本集合 E 的一个分解:

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

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 $P_{A_q}(B)$, 我们称其为当试验 \mathcal{U} 给定时事件 B 的条件概率, 并用 $P_{\mathcal{U}}(B)$ 。当且仅当下面公式成立时, 两个试验 $\mathcal{U}^{(1)}$ 和 $\mathcal{U}^{(2)}$ 独立。

$$P_{\mathcal{U}^{(1)}}(A_q^{(2)}) = P(A_q^{(2)}) \quad q = 1, 2, \dots, r_2$$

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

$$\mathcal{U}^{(1)} \mathcal{U}^{(2)} \dots \mathcal{U}^{(n)}$$

表示基本集合 E 的分解的乘积:

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

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

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

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

$$P_{\mathcal{U}^{(1)} \mathcal{U}^{(2)} \dots \mathcal{U}^{(n-1)}}(A_q^{(n)}) = P_{\mathcal{U}^{(n-1)}}(A_q^{(n)})$$

成立, 则序列 $\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \dots, \mathcal{U}^{(n)}, \dots$ 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

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

则马尔可夫链理论的基本方程为以下形式:

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

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:

$$p(k, n) = p(k, m) p(m, n) \quad k < m < n \quad (2)$$

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

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

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 $1/4$)。

$$\begin{aligned} P(A) &= P(B) = P(C) = 1/2 \\ P(AB) &= P(BC) = P(AC) = 1/4 = (1/2)^2 \\ P(ABC) &= 1/4 \neq (1/2)^3 \end{aligned}$$

无限概率域

连续性公理

按照惯例, 我们用符号 $\bigcap_m A_m$ 表示集合 A_m 的乘积 (无论是有限个还是无限个), 用符号 $\bigcup_m A_m$ 表示集合 A_m 的和。仅当集合 A_m 为不相交集 (disjoint sets, 集互斥集合), 集合的和的形式记为 $\sum_m A_m$ 。由此可得:

$$\begin{aligned} \bigcup_m A_m &= A_1 + A_2 + \dots \\ \sum_m A_m &= A_1 + A_2 + \dots \\ \bigcap_m A_m &= A_1 A_2 \dots \end{aligned}$$

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

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots \quad (1)$$

$$\bigcap_n A_n = 0 \quad (2)$$

当公式 (2) 成立时, 以下等式 (公式 (3)) 成立。

$$\lim_{n \rightarrow \infty} P(A_n) = 0 \quad (3)$$

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:

$$\begin{aligned} A_k &= A_{k+p} = \bigcap_n A_n = 0 \\ \lim P(A_n) &= P(0) = 0 \end{aligned}$$

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 (consistent 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)

$$A = \sum_n A_n \quad (4)$$

可得公式(5)

$$P(A) = \sum_n P(A_n) \quad (5)$$

证明：令

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

则显然

$$\mathcal{D}_n(R_n) = 0$$

因此根据公理 VI 得

$$\lim P(R_n) = 0 \quad n \rightarrow \infty \quad (6)$$

另一方面，根据加法定理得

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_n) + P(R_n) \quad (7)$$

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 $A_1, A_2, \dots, A_n, \dots$ 以及 A 属于 \mathcal{E} ，并且

$$A \subset \bigcup_n A_n \quad (8)$$

则

$$P(A) \leq \sum_n P(A_n) \quad (9)$$

证明：

$$\begin{aligned} A &= A \bigcup_n A_n = A(A_1 + A_2(1 - A_1) + A_3(1 - A_2 - A_1)) + \cdots = AA_1 + A(A_2 - A_2 A_1) + A(A_3 - A_3 A_2 - A_3 A_1) + \dots \\ P(A) &= P(AA_1) + P(A(A_2 - A_2 A_1)) + \cdots \leq P(A_1) + P(A_2) + \dots \end{aligned}$$

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

$$\bigcup_n A_n = A_1 + (A_2 - A_2 A_1) + (A_3 - A_3 A_2 - A_3 A_1) + \dots \quad (1)$$

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

由以下等式

$$\mathcal{D} A_n = E - \bigcup_n A_n \quad (2)$$

可知，对于集合 A_n 的乘积 $\mathcal{D} A_n$ 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)：

$$A \subset \bigcup_n A_n$$

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 $E = \xi_1, \xi_2, \dots, \xi_n, \dots$ 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 ：

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

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 $[a; b) = \{a \leq \xi < b\}$ 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：

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

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) : $\xi = \{x_1, x_2, \dots, x_n\}$ 。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, \dots, a_n} 确定:

$$P(L_{a_1, a_2, \dots, a_n}) = F(a_1, a_2, \dots, a_n) \quad (3)$$

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

对于 $F(a_1, a_2, \dots, a_n)$, 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:

$$\begin{aligned} \lim_{a_i \rightarrow -\infty} F(a_1, a_2, \dots, a_n) &= F(a_1, \dots, a_{i-1}, -\infty, a_{i+1}, \dots, a_n) = 0, \\ &\quad (i = 1, 2, \dots, n) \\ \lim_{a_1 \rightarrow +\infty, a_2 \rightarrow +\infty, \dots, a_n \rightarrow +\infty} F(x) &= F(+\infty, +\infty, \dots, +\infty) = 1 \end{aligned} \quad (4)$$

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:

$$\left. \begin{aligned} \lim_{a_i \rightarrow -\infty} F(a_1, a_2, \dots, a_n) &= F(a_1, \dots, a_{i-1}, -\infty, a_{i+1}, \dots, a_n) = 0, \\ &\quad i = 1, 2, \dots, n \\ \lim_{a_1 \rightarrow +\infty, a_2 \rightarrow +\infty, \dots, a_n \rightarrow +\infty} F(a_1, a_2, \dots, a_n) &= F(+\infty, +\infty, \dots, +\infty) = 1. \\ \sum_{\substack{i=1 \\ \epsilon_i=0,1}}^n (-1)^{\epsilon_1 + \epsilon_2 + \dots + \epsilon_n} F(a_1 - \epsilon_1 c_1, a_2 - \epsilon_2 c_2, \dots, a_n - \epsilon_n c_n) &\geq 0, \\ c_i > 0, \quad i &= 1, 2, 3, \dots, n. \end{aligned} \right\} 4)$$

$F(a_1, a_2, \dots, a_n)$ 称为变量 x_1, x_2, \dots, 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:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$$

并且令

$$P(A) = \int \int \dots \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (5)$$

$f(x_1, x_2, \dots, x_n)$ 此时称为在点 (x_1, x_2, \dots, 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：

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

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

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

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 $u_2 u_1(\xi)$ 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 $u(\xi) = u_2 u_1(\xi)$ 。很容易得出，这两个概率函数有以下关联：

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

随机变量和分布函数的定义

定义：定义在基本集合 E 上的实单值函数 $x(\xi)$ ，当对于任意实数 a ，满足 $x < a$ 的所有 ξ 组成的集合属于集合系统 \mathcal{E} ，则这个实单值函数 $x(\xi)$ 称为随机变量（random variable）。

函数 $x(\xi)$ 将基本集合 E 映射到集合 R^1 ，即全体实数。如本章第 1 节所言，这个函数确定了一个实数集 R^1 的子集组成的域 $\mathcal{E}^{(x)}$ 。我们可以用以下方式重新表述随机变量的定义：

对于实函数 $x(\xi)$ ，当且仅当域 $\mathcal{E}^{(x)}$ 包含形如 $(-\infty, a)$ 的所有区间时， $x(\xi)$ 为一个随机变量。

因为 $\mathcal{E}^{(x)}$ 是一个域，所以连同区间 $(-\infty, a)$ 一起，这个域包含了半开区间 $[a; b)$ 的所有可能的有限和。如果我们的概率域是波莱尔域，则域 \mathcal{E} 和 $\mathcal{E}^{(x)}$ 也是波莱尔域。因此域 $\mathcal{E}^{(x)}$ 包含集合 R^1 的所有波莱尔集。

后面我们会用 $P^{(x)}(A')$ 来表示随机变量的概率函数。它是定义在域 $\mathcal{E}^{(x)}$ 的所有集合上的。特别地，对于最重要的情形——概率的波莱尔域， $P^{(x)}$ 定义在实数集 R^1 的所有波莱尔集上。

定义：函数

$$F^{(x)}(a) = P^{(x)}(-\infty, a) = P\{x < a\}$$

其中 $-\infty$ 和 $+\infty$ 是 a 的可取值。此时该函数称为随机变量 x 的分布函数（distribution function of the random variable x ）。

由定义马上可知

$$\begin{aligned} F^{(x)}(-\infty) &= 0 \\ F^{(x)}(+\infty) &= 1 \end{aligned} \tag{1}$$

满足不等式 $a \leq x < b$ 的 x 概率由下式给出

$$P\{x \in [a; b]\} = F^{(x)}(b) - F^{(x)}(a) \tag{2}$$

由此可得, 对于 $a < b$, 有

$$F^{(x)}(a) \leq F^{(x)}(b)$$

上式表明, 函数 $F^{(x)}(a)$ 是个非递减函数。现在假设 $a_1 < a_2 < \dots < a_n < \dots < b$, 则有

$$\mathcal{D}_n\{x \in [a_n; b]\} = 0$$

因此, 根据连续性公理可得, 当 $n \rightarrow +\infty$ 时,

$$F^{(x)}(b) - F^{(x)}(a_n) = P\{x \in [a_n; b]\}$$

上式趋近于 0。由此可得 函数 $F^{(x)}(b)$ 为左连续函数。

采用类比的方法可以证明:

$$\lim_{a \rightarrow -\infty} F^{(x)}(a) = F^{(x)}(-\infty) = 0, \quad a \rightarrow -\infty \tag{3}$$

$$\lim_{a \rightarrow +\infty} F^{(x)}(a) = F^{(x)}(+\infty) = 1, \quad a \rightarrow +\infty \tag{4}$$

如果概率域 (\mathcal{E}, P) 是波莱尔域, 则对于所有实数集 R^1 上的波莱尔集合 A , 概率函数 $P^{(x)}(A)$ 的值由分布函数 $F^{(x)}(a)$ 唯一确定 (参考第二章第 3 节中的第 III 个例子)。本文重点关注 $P^{(x)}(A)$ 的值, 所以分布函数在后续研究中将发挥极其重要的作用。

如果分布函数 $F^{(x)}(a)$ 可微, 则其相对于参数 a 的导数 (如下) 称为 x 在点 a 处的概率密度 (probability density)。

$$f^{(x)}(a) = \frac{d}{da} F^{(x)}(a)$$

如果对于每一个 a , 都有

$$\frac{d}{da} F^{(x)}(a) = \int_{-\infty}^a f^{(x)}(a) da$$

则任意波莱尔集 A 的概率函数 $P^{(x)}(a)$ 可以用以下形式表示

$$P^{(x)}(A) = \int_A f^{(x)}(a) da \tag{5}$$

这种情况下我们称 x 的分布是连续的。上式可以写为另一种更一般的形式

$$P^{(x)}(A) = \int_A dF^{(x)}(a) \tag{6}$$

所有刚刚介绍到的概念都可以推广到条件概率的情形。

集合函数

$$P_B^{(x)}(A) = P_B\{x \in A\}$$

表示在假设 B 下 x 的条件概率。非递减函数

$$F_B^{(x)}(a) = P_B\{x < a\}$$

为对应的分布函数。并且, 当 $F_B^{(x)}(a)$ 可微时,

$$f_B^{(x)}(a) = \frac{d}{da} F_B^{(x)}(a)$$

是在假设 B 的条件下, x 在 a 点的条件概率密度。

多维分布函数

现在给定 n 个随机变量 X_1, X_2, \dots, X_n 。 n 维空间 R^n 中的点 $x = (X_1, X_2, \dots, X_n)$ 是基本事件 ξ 的函数。则根据本章第一节的内容，可得到定义在空间 R^n 上的域 $\mathcal{E}^{(x_1, x_2, \dots, x_n)}$ ，这个域包括了空间 R^n 的子集；也可以得到定义在 \mathcal{E}' 上的概率函数 $P^{(x_1, x_2, \dots, x_n)}(A')$ 。这个概率函数称为随机变量 x_1, x_2, \dots, x_n 的 n 维概率函数。

对于任意选择的 i 和 a_i ($i = 1, 2, \dots, n$)，由随机变量的定义可以直接得出 R^n 中满足 $x_i < a_i$ 的所有的点组成的集合。因此 \mathcal{E}' 也包括了以上集合的交集，例如集合 $L_{a_1 a_2 \dots a_n}$ 表示 R^n 中满足所有不等式 $x_i < a_i$ ($i = 1, 2, \dots, n$) 的点的集合²⁷。

如果把 R^n 空间中满足不等式 $a_i \leq x_i < b_i$ 的所有点组成的集合记为 n 维半开区间 $[a_1, a_2, \dots, a_n; b_1, b_2, \dots, b_n]$ ，则可以发现每一个这样的区间都属于域 \mathcal{E}' ，因为

$$[a_1, a_2, \dots, a_n; b_1, b_2, \dots, b_n) = L_{b_1 b_2 \dots b_n} - L_{a_1 b_2 \dots b_n} - L_{b_1 a_2 \dots b_n} - \dots - L_{b_1 b_2 \dots b_{n-1} a_n}$$

所有的 n 维半开区间系统的波莱尔扩展，包含 R^n 中的所有波莱尔集合。由此可得出，在波莱尔概率域下，域 \mathcal{E} 包含了 R^n 空间的所有的波莱尔集合。

【关于波莱尔扩展 (Borel extension) 和波莱尔集合 (Borel sets) 的区别】

定理：在波莱尔概率域下，每一个定义在有限个随机变量 x_1, x_2, \dots, x_n 上的波莱尔函数 $x = f(x_1, x_2, \dots, x_3)$ 也是个随机变量。

要证明以上定理，只需证明 R^n 上满足 $x = f(x_1, x_2, \dots, x_n) < a$ 的所有的点 (x_1, x_2, \dots, x_n) 组成的集合是波莱尔集。特别地，所有的对随机变量进行有限的求和和求乘积的操作得到的变量，也是随机变量。

定义：函数

$$F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_n) = P^{(x_1, x_2, \dots, x_n)}(L_{a_1 a_2 \dots a_n})$$

称为随机变量 x_1, x_2, \dots, x_n 的 n 维分布函数。

与在一维下的情形类似，我们证明证明 n 维分布函数 $F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_n)$ 是一个非减函数，并且对于任意一个变量都是左连续的。类比第二节中的公式 (3) 和 (4)，可得：

$$\lim_{a_i \rightarrow -\infty} F(a_1, a_2, \dots, a_n) = F(a_1, a_2, \dots, a_{i-1}, -\infty, a_{i+1}, \dots, a_n) = 0 \quad (7)$$

$$\lim_{a_1 \rightarrow +\infty, a_2 \rightarrow +\infty, \dots, a_n \rightarrow +\infty} F(a_1, a_2, \dots, a_n) = F(+\infty, +\infty, \dots, +\infty) = 1 \quad (8)$$

分布函数 $F^{(x_1, x_2, \dots, x_n)}$ 只针对于特定集合 $L_{a_1 a_2 \dots a_n}$ 给出了概率 $P^{(x_1, x_2, \dots, x_n)}$ 的值。但如果概率域是波莱尔域，则²⁸ 对于所有 R^n 上的波莱尔集， $P^{(x_1, x_2, \dots, x_n)}$ 都可以由分布函数 $F^{(x_1, x_2, \dots, x_n)}$ 唯一确定 (uniquely determined)。

如果存在导数

$$f(a_1, a_2, \dots, a_n) = \frac{\partial}{\partial a_1 \partial a_2 \dots \partial a_n} F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_n)$$

则我们称这个导数为随机变量 x_1, x_2, \dots, x_n 在点 a_1, a_2, \dots, a_n 处的 n 维概率密度。并且如果对于每个点 (a_1, a_2, \dots, a_n) ，都有

$$F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_n) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \dots \int_{-\infty}^{a_n} f(a_1, a_2, \dots, a_n) da_1 da_2 \dots da_n$$

则 x_1, x_2, \dots, x_n 的分布称为连续的。对于每一个波莱尔集合 $A \subset R^n$ ，有以下等式

$$P^{(x_1, x_2, \dots, x_n)}(A) = \int \int \dots \int f(a_1, a_2, \dots, a_n) da_1 da_2 \dots da_n \quad (9)$$

在本章最后我们对各种概率函数和分布函数的关系再做一个说明。

给定如下代换

$$S = \binom{1, 2, \dots, n}{i_1, i_2, \dots, i_n}$$

令 r_s 表示如下的空间 R^n 到其自身的一个变换：

$$x'_k = x_{ik} \quad (k = 1, 2, \dots, n)$$

显然可得

$$\mathrm{P}^{(x_{i_1}, x_{i_2}, \dots, x_{i_n})}(A) = \mathrm{P}^{(x_1, x_2, \dots, x_n)}\{r_s^{-1}(A)\} \quad (10)$$

现在令 $x' = p_k(x)$ 为空间 R^n 在空间 R^k ($k < n$) 的投影，所以空间 R^n 中的点 x_1, x_2, \dots, x_n 映射到空间 R^k 中为 x_1, x_2, \dots, x_k 。所以，与第 1 节中的公式 (2) 类似，有

$$\mathrm{P}^{(x_1, x_2, \dots, x_k)}(A) = \mathrm{P}^{(x_1, x_2, \dots, x_n)}(A)\{p_k^{-1}(A)\} \quad (11)$$

对应的分布函数，由公式 (10) 和公式 (11) 可得以下两个方程：

$$F^{(x_{i_1}, x_{i_2}, \dots, x_{i_n})}(a_{i_1}, a_{i_2}, \dots, a_{i_n}) = F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_n) \quad (12)$$

$$F^{(x_1, x_2, \dots, x_k)}(a_1, a_2, \dots, a_k) = F^{(x_1, x_2, \dots, x_n)}(a_1, a_2, \dots, a_k, +\infty, \dots, +\infty) \quad (13)$$

无限维空间中的概率

在第二章第 3 节我们已经看到如何构造概率论中常见的各种概率域。不过我们仍可以想象某一类问题，在这类问题中基本事件是由无限多个坐标定义的。假定一个由任意基数 m 中的索引 μ 构成的集合 M 。我们把实数 x_μ 构成的系统的总体（其中 μ 可以遍历整个集合 M ）

$$\xi = \{x_\mu\}$$

称为空间 R^M （为了定义空间 R^M 中的一个元素 ξ ，我们必须将集合 M 中的每一个元素 μ 与一个实数 x_μ 对应，或者等效地给每一个元素 μ 赋予一个定义在 M 上的单值实函数 x_μ ）²⁹。如果集合 M 中所包含的是 n 个自然数 $1, 2, \dots, n$ ，则 R^M 就是普通的 n 维空间 R^n 。如果集合 M 中是所有的实数 R^1 ，则对应的空间 $R^M = R^{R^1}$ 包含了实变量 μ 的所有的实函数：

$$\xi(\mu) = x_\mu$$

现在我们将集合 R^M （其中 M 为任意集合）作为基本集合 E 。令 $\xi = \{x_\mu\}$ 为集合 E 中的一个元素。则我们可以用符号 $p_{\mu_1 \mu_2 \dots \mu_n}(\xi)$ 表示 n 维空间 R^n 中的点 $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ 。如果 E 的子集 A 可以表示成如下形式，我们称集合 A 为圆柱集合（cylinder set）：

$$A = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(\xi)(A')$$

其中 A' 是 R^n 的子集。因此，所有圆柱集组成的类（class），与那些可以按如下形式的关系定义的集合组成的类一致：

$$f(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n}) = 0 \quad (1)$$

为按照以上关系定义任意圆柱集 $p_{\mu_1 \mu_2 \dots \mu_n}(A)$ ，我们只需构造一个函数 f ，使它在 A' 处等于 0，并且在除 A' 以外的其他地方等于 1（unity）。

当 A' 是波莱尔集合时，对应的这个圆柱集是波莱尔圆柱集（Borel cylinder set）。空间 R^M 中的所有波莱尔圆柱集构成一个域，记为 \mathcal{E}^M ³⁰。

我们把域 \mathcal{E}^M 的波莱尔扩展记为 $B\mathcal{E}^M$ 。 $B\mathcal{E}^M$ 中的集合称为空间 R^M 的波莱尔集合。

稍后我们会给出一种在 \mathcal{E}^M 上构造和操作概率函数的方法，并且借由扩展定理（Extension Theorem），该方法同样可用于在 $B\mathcal{E}^M$ 上构造和操作概率函数。由此方式得到的概率域，在当集合 M 为可数的情况下，可以满足所有的（研究）目的。由此我们可以处理涉及到可数随机变量序列的所有问题。但当 M 不可数时， R^M 中的很多简单又有趣的子集就超出 $B\mathcal{E}^M$ 的范围。例如，当集合 M 不可数时，对于所有的索引 μ ，那些满足 x_μ 小于某个固定常数的所有元素 ξ 构成的集合，就不属于系统 $B\mathcal{E}^M$ 。

因此，尽可能地把每一个问题转化为一种形式，使得在这种形式下所有基本事件 ξ 的空间只有一个可数的坐标集（a denumerable set of coordinates），这种方法是非常可取的。

假设概率函数 $P(A)$ 定义在 \mathcal{E}^M 上，则我们可以把基本事件 ξ 的每一个坐标 x_μ 都当作一个随机变量。因此这些坐标的每个有限群（group） $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ 都有一个 n 维的概率函数 $P_{\mu_1 \mu_2 \dots \mu_n}(A)$ 和与之对应的分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}(a_1, a_2, \dots, a_n)$ 。显然，对每一个波莱尔圆柱集 A ：

$$A = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(A')$$

有如下等式成立：

$$P(A) = P_{\mu_1 \mu_2 \dots \mu_n}(A')$$

其中 A' 是空间 R^n 的一个波莱尔集。这样，空间 \mathcal{E}^M 上所有圆柱集的概率函数 P 可以由空间 R^n 上的所有波莱尔集的优点概率函数 $P_{\mu_1 \mu_2 \dots \mu_n}$ 确定。然而，对于波莱尔集，概率函数 $P_{\mu_1 \mu_2 \dots \mu_n}$ 又是由相应的分布函数唯一确定的。由此我们证明了以下定理：

所有有限维分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}$ 的集合唯一确定了所有在空间 \mathcal{E}^M 中的集合的概率函数 $P(A)$ 。如果 $P(A)$ 定义在 \mathcal{E}^M 上，则（根据扩展定理）它由分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}$ 的值在 $B\mathcal{E}^M$ 上唯一确定。

接下来读者或许想问：在哪种情况下，一个分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}$ 系统能够先验地定义一个空间 \mathcal{E}^M 上（或者空间 $B\mathcal{E}^M$ ）的概率域呢？

首先需要说明，分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}$ 必须满足第二章第 3 节的第 III 个例子（无限概率域的例子）。事实上这些条件包含在分布函数的概念当中。此外，作为本章第 2 节方程 (13) 和 (14) 的结果（注，推测应为第 3 节的方程 (12) 和 (13)），我们有如下两个等式关系：

$$F_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_n}}(a_{i_1}, a_{i_2}, \dots, a_{i_n}) = F_{\mu_1 \mu_2 \dots \mu_n}(a_1, a_2, \dots, a_n) \quad (2)$$

$$F_{\mu_1 \mu_2 \dots \mu_k}(a_1, a_2, \dots, a_k) = F_{\mu_1 \mu_2 \dots \mu_n}(a_1, a_2, \dots, a_k, +\infty, \dots, +\infty) \quad (3)$$

其中 $k < n$ ，并且 $\binom{1, 2, \dots, n}{i_1, i_2, \dots, i_n}$ 是一个任意的排列（permutation）。这些必要条件也被证明为是充分条件。这一点由下面的定理可以得出。

基本定理：每一个满足条件 (2) 和 (3) 的分布函数系统，都定义了一个 \mathcal{E}^M 上的概率函数 $P(A)$ ，它满足公理 I - VI。并且这个概率函数 $P(A)$ 也可以（通过扩展定理）扩展到空间 $B\mathcal{E}^M$ 上。

证明

给定满足第二章第 3 节例子 III 的一般条件，以及满足条件 (2) 和 (3) 的分布函数 $F_{\mu_1 \mu_2 \dots \mu_n}$ 。分布函数都唯一确定了空间 R^n 上的所有波莱尔集的概率函数 $P_{\mu_1 \mu_2 \dots \mu_n}$ （参考本章第 3 节）。后面的讨论中，我们只关注空间 R^n 中的波莱尔集和空间 E 中的波莱尔圆柱集。

对于每一个圆柱集 $A = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(A')$ ，我们令

$$P(A) = P_{\mu_1 \mu_2 \dots \mu_n}(A'') \quad (4)$$

由于可以由不同的集合 A' 经过构造得到相同的圆柱集 A ，所以我们首先需要确定公式 (4) 总是得到相同的 $P(A)$ 。

假设 $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ 为随机变量 x_μ 的有限系统。根据这些随机变量的概率函数 $P_{\mu_1 \mu_2 \dots \mu_n}$ ，再结合第 3 节提到的规则，我们可以定义每一个子系统 $(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}})$ 的概率函数 $P_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_k}}$ 。由等式 (2) 和 (3) 可得，根据第 3 章内容定义的概率函数与先验给出的函数 $P_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_k}}$ 相同。现在我们假设圆柱集 A 是通过下面等式定义的：

$$A = p_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_k}}^{-1}(A')$$

同时也是通过下面等式定义的：

$$A = p_{\mu_{j_1} \mu_{j_2} \dots \mu_{j_m}}^{-1}(A'')$$

其中所有的随机变量 x_{μ_i} 和 x_{μ_j} 都属于系统 $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ ，这显然不是一个本质上的限制。条件

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}}) \subset A'$$

和条件

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \dots, x_{\mu_{j_m}}) \subset A''$$

是等价的。因此

$$P_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_k}}(A') = P(A) = P_{\mu_1 \mu_2 \dots \mu_n}\{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}}) \subset A'\} = P_{\mu_1 \mu_2 \dots \mu_m}\{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \dots, x_{\mu_{j_m}}) \subset A''\} = P_{\mu_{j_1} \mu_{j_2} \dots \mu_{j_m}}(A'')$$

由此证明了 $P(A)$ 定义的唯一性。

接下来证明概率域 (\mathcal{E}^M, P) 满足所有的公理 I - VI。公理 I 只要求 \mathcal{E}^M 为一个域——这一事实已经在上面的内容中证明。并且，对于任意索引 μ ，有如下关系：

$$E = p_\mu^{-1}(R^1)$$

$$\mathbf{P}(E) = \mathbf{P}_\mu(R^1) = 1$$

由此证明了公理 II 和公理 IV 也满足。最后，由 $\mathbf{P}(A)$ 的定义可得 $\mathbf{P}(A)$ 是非负的（满足公理 III）。

证明公理 V 略微复杂一些。为此我们需要考虑两个圆柱集

$$A = p_{\mu_{i_1} \mu_{i_2} \dots \mu_{i_k}}^{-1}(A')$$

和

$$B = p_{\mu_{j_1} \mu_{j_2} \dots \mu_{j_m}}^{-1}(B')$$

我们假设随机变量 x_{μ_i} 和 x_{μ_j} 属于有限系统（原文为inclusive finite system, inclusive 不知道该怎么翻译） $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ 。如果集合 A 和集合 B 没有交集，则以下两个关系

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}}) \subset A'$$

和（下面的等式，最后的下标 k ，或许应该为 m ，待确定）

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \dots, x_{\mu_{j_k}}) \subset B'$$

是互斥的。因此

$$\mathbf{P}(A + B) = \mathbf{P}_{\mu_1 \mu_2 \dots \mu_n} \{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}}) \subset A'\} \text{ OR } \{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \dots, x_{\mu_{j_m}}) \subset B'\} = \mathbf{P}_{\mu_1 \mu_2 \dots \mu_n} \{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \dots, x_{\mu_{i_k}}) \subset A' \text{ AND } (x_{\mu_{j_1}}, x_{\mu_{j_2}}, \dots, x_{\mu_{j_m}}) \subset B'\} = \mathbf{P}(A) + \mathbf{P}(B)$$

由此得证公理 V。

此时只剩下公理 VI。令

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$$

为一个圆柱集的递减序列，并且满足以下条件

$$\lim \mathbf{P}(A_n) = L > 0$$

我们需要证明，所有集合 A_n 的乘积为非空。在不会实际限制原有问题的情况下，我们可以假设在前 n 个圆柱集 A_k 的定义中，序列 $x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_{i_n}}, \dots$ 中只有前 n 个坐标 x_{μ_k} 存在，即

$$A_n = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(B_n)$$

为简写记，记为

$$\mathbf{P}_{\mu_1 \mu_2 \dots \mu_n}(B) = \mathbf{P}_n(B)$$

显然有

$$\mathbf{P}_n(B_n) = \mathbf{P}(A_n) \geq L > 0$$

在每一个集合 B_n 中都有可能找到一个具有闭区间的有界集合 U_n ，使得

$$\mathbf{P}_n(B_n - U_n) \leq \frac{\epsilon}{2^n}$$

根据以上不等式，对于集合 V_n

$$V_n = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(U_n)$$

可得不等式

$$\mathbf{P}(A_n - V_n) \leq \frac{\epsilon}{2^n} \tag{5}$$

进一步，令

$$W_n = V_1 V_2 \dots V_n$$

由公式 (5) 可得

$$\mathbf{P}(A_n - W_n) \leq \epsilon$$

由于 $W_n \subset V_n \subset A_n$ ，所以有如下关系

$$P(W_n) \geq P(A_n) - \epsilon \geq L - \epsilon$$

如果 ϵ 足够小, $P(W_n) > 0$ 并且 W_n 不为空。我们可以在每个集合 W_n 中选择一个坐标为 $x_\mu^{(n)}$ 的点 $\xi^{(n)}$ 。每个点 $\xi^{(n+p)}$, 其中 $p = 0, 1, 2, \dots$, 都属于集合 V_n 。因此有

$$(x_{\mu_1}^{(n+p)}, x_{\mu_2}^{(n+p)}, \dots, x_{\mu_n}^{(n+p)}) = p_{\mu_1 \mu_2 \dots \mu_n}^{-1}(\xi^{(n+p)}) \subset U_n$$

因为集合 U_n 是有界的, 所以我们可以从序列 $\{\xi^{(n)}\}$ 中挑选出一个子序列 (采用对角线法, by the diagonal method) :

$$\xi^{(n_1)}, \xi^{(n_2)}, \dots, \xi^{(n_i)}, \dots$$

在这个子序列中, 对于任意的 k , 每一个点的坐标 $x_{\mu_k}^{(n_i)}$ 都趋向于一个确定的极限 x_k 。最后我们令 ξ 为集合 E 中的一点, 它的坐标为

$$\begin{aligned} x_{\mu_k} &= x_k \\ x_m &= 0, \quad \mu \neq \mu_k \quad k = 1, 2, 3, \dots \end{aligned}$$

作为序列 $x_1^{(n_i)}, x_2^{(n_i)}, \dots, x_k^{(n_i)}$, $i = 1, 2, 3, \dots$ 的极限, 点 (x_1, x_2, \dots, x_k) 属于集合 U_k 。因此对于任意 k , ξ 属于

$$A_k \subset V_k = p_{\mu_1 \mu_2 \dots \mu_k}^{-1}(U_k)$$

因此证明

$$A = \bigcap_k A_k$$

等价随机变量及多种收敛

从现在开始, 我们将专门讨论概率的波莱尔域 (Borel fields of probability)。正如在第 2 节中已经解释过的, 这一讨论范围的限定实际上并不会构成对我们研究问题的限制。【有点奇怪】

如果两个随机变量 x 和 y , $x \neq y$ 的概率等于 0, 则称这两个随机变量等价 (equivalent)。显然, 两个等价的随机变量具有相同的概率函数:

$$P^{(x)}(A) = P^{(y)}(A)$$

因此, 它们对应的分布函数 $F^{(x)}$ 和 $F^{(y)}$ 也相同。在概率理论的许多问题中, 我们都可以把某一个随机变量替换为任何其他等价随机变量。

现在, 令

$$x_1, x_2, \dots, x_n, \dots \tag{1}$$

为一个随机变量序列。现在来研究满足序列 (1) 为收敛序列的所有基本事件 ξ 构成的集合 A 。将满足以下不等式的基本事件 ξ 的集合记为 $A_{n,p}^{(m)}$:

$$|x_{n+k} - x_n| < \frac{1}{m} \quad k = 1, 2, \dots, p$$

由此可得

$$A = \bigcap_{m,n,p} \mathcal{DGA}_{n,p}^{(m)} \tag{2}$$

由第 3 节可知, 集合 $A_{n,p}^{(m)}$ 总是属于域 \mathcal{E} 。等式 (2) 表明, 集合 A 也属于 \mathcal{E} 。因此, 我们可以讨论随机变量序列收敛的概率, 因为它总是具有明确的意义。

现在令收敛集合 A 的概率 $P(A)$ 等于 1, 则序列 (1) 依概率 1 收敛于一个随机变量 x 。其中 x 除了等价关系外是唯一确定的。为确定随机变量 x , 令集合 A 中的

$$x = \lim x_n \quad n \rightarrow \infty$$

且集合 A 外有 $x = 0$ 。我们需要证明, x 是一个随机变量, 即满足 $x < a$ 的所有元素 ξ 构成的集合 $A(a)$ 属于 \mathcal{E} 。但当 $a \leq 0$ 时有

$$A(a) = \bigcap_{n,p} \mathcal{D}\{x_{n+p} < a\}$$

相反，当 $a > 0$ 时有

$$A(a) = \underset{n, p}{\mathcal{AGD}} \{x_{n+p} < a\} + \bar{A}$$

由此，可立即得出序列 (1) 依概率 1 收敛于一个随机变量 x 的结论。

如果收敛序列 (1) 收敛于 x 的概率为 1，则称序列 (1) 几乎必然收敛到 x (converges almost surely to x)。但对概率理论来说，另一种收敛的概念可能更为重要。

定义：如果对于任意 $\epsilon > 0$ ，随机变量序列 $x_1, x_2, \dots, x_n, \dots$ 依概率收敛于随机变量 x ，则当 $n \rightarrow \infty$ 时，概率

$$\text{P}\{|x_n - x| > \epsilon\}$$

趋近于 0³¹。

I. 如果序列 (1) 依概率收敛于 x 和 x' ，则 x 与 x' 等价。事实上

$$\text{P}\{|x - x'| > \frac{1}{m}\} \leq \text{P}\{|x_n - x| > \frac{1}{2m}\} + \text{P}\{|x_n - x'| > \frac{1}{2m}\}$$

由于对于足够大的 n ，最后这些概率可以任意小，因此可以得出

$$\text{P}\{|x - x'| > \frac{1}{m}\} = 0$$

并且可立即得出

$$\text{P}\{x \neq x'\} \leq \sum_m \text{P}\{|x - x'| > \frac{1}{m}\} = 0$$

II. 如果序列 (1) 几乎必然收敛于 x ，则它也依概率收敛于 x 。令 A 为序列 (1) 的收敛集合，则

$$1 = \text{P}(A) \leq \lim_{n \rightarrow \infty} \{\text{P}\{|x_{n+p} - x| < \epsilon, p = 0, 1, 2, \dots\} \leq \lim_{n \rightarrow \infty} \text{P}\{|x_n - x| < \epsilon\}$$

由此可以推得依概率收敛。

III. 对于序列 (1) 依概率收敛的情况，以下条件既是充分的，也是必要的：对任意 $\xi > 0$ ，总存在一个 n ，使得对于任意 $p > 0$ ，以下等式成立【这里有问题，因为都没有 p 】：

$$\text{P}\{|x_{n+p} - x_n| > \epsilon\} < \epsilon$$

令 $F_1(a), F_2(a), \dots, F_n(a), \dots, F(a)$ 为随机变量 $x_1, x_2, \dots, x_n, \dots, x$ 的分布函数。如果序列 x_n 依概率收敛于 x ，分布函数 $F(a)$ 由 $F_n(a)$ 唯一确定。事实上有以下定理

定理：如果序列 $x_1, x_2, \dots, x_n, \dots$ 依概率收敛于 x ，则对应的分布函数 $F_n(a)$ 的序列在 $F(a)$ 的每一个连续点处收敛于 x 的分布函数 $F(a)$ 。

$F(a)$ 是左连续的单调函数，由连续点处的值唯一确定³²。由此可得， $F(a)$ 由 $F_n(a)$ 确定。为证明这一定理，假设 F 在点 a 连续。令 $a' < a$ ，则当 $x < a'$ 、 $x_n \geq a$ 时，必须有 $|x_n - x| > a - a'$ 。因此有

$$\begin{aligned} \lim P(x < a', x_n \geq a) &= 0, \\ F(a') = P(x < a') &\leq P(x_n < a) + P(x < a', x_n \geq a) = F_n(a) + P(x < a', x_n \geq a), \\ F(a') &\leq \liminf F_n(a) + \lim P(x < a', x_n \geq a), \\ F(a') &\leq \liminf F_n(a) \end{aligned} \quad (3)$$

注： \liminf 表示下极限 (limit inferior)，即序列 $F_n(a)$ 在所有子序列极限中的最小极限值。下文中的 \limsup 表示上极限。【感慨一下，翻译到这里时，deepseek 已经声名鹊起】

由以上结果类比，当 $a'' > a$ 时有

$$F(a'') \geq \limsup F_n(a) \quad (4)$$

因为当 $a' \rightarrow a$ ，且 $a'' \rightarrow a$ 时， $F(a')$ 与 $F(a'')$ 都收敛于 $F(a)$ ，所以由公式 (3) 和 (4) 可得

$$\lim F_n(a) = F(a)$$

由此证明以上定理。

$$\begin{aligned} f(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n}) &= 0 \\ g(x_{\lambda_1}, x_{\lambda_2}, \dots, x_{\lambda_m}) &= 0 \end{aligned}$$

由此可根据以下关系定义集合 $A + B$, AB , 以及 $A - B$:

$$\begin{aligned} f \cdot g &= 0 \\ f^2 + g^2 &= 0 \\ f^2 + \omega(g) &= 0 \end{aligned}$$

其中当 $x \neq 0$ 时 $\omega(x) = 0$, 并且 $\omega(0) = 1$ 。如果 f 和 g 都是波莱尔函数, 则 $f \cdot g$ 、 $f^2 + g^2$ 和 $f^2 + \omega(g)$ 也是波莱尔函数。因此 $A + B$, AB , 以及 $A - B$ 都是波莱尔圆柱集。由此我们证明集合 \mathcal{E}^M 的系统为一个域。

数学期望 33

抽象勒贝格 (Lebesgue) 积分

令 x 为随机变量, A 为 \mathcal{E} 构成的集合。给定一个正的 λ , 构造以下和式:

$$S_\lambda = \sum_{k=-\infty}^{k=+\infty} k\lambda P\{k\lambda \leq x < (k+1)\lambda, \xi \in A\} \quad (1)$$

如果对每一个 λ , 这个级数都绝对收敛, 则当 $\lambda \rightarrow 0$ 时, S_λ 趋近于一个确定值, 且这个确定值为一个积分:

$$\int_A x P(dE) \quad (2)$$

在这一抽象形式下, 积分的概念由 Fréchet³⁴ 提出。它对概率理论是必不可少的 (dispensable)。(后文中读者会发现, 除去一个常数因子外, 变量 x 在假设 A 下的条件数学期望的常规定义与积分 (2) 中的定义恰巧相同)。

此处简要介绍下形式 (2) 中的积分的最重要的一些性质。读者会在任何一本关于实变量的习题册中找到它们的证明, 尽管这些证明通常都是基于 $P(A)$ 是空间 R^n 上集合的勒贝格测度 (Lebesgue measure) 这一假设做出的。将这些证明推广到更广义情形的过程并不包含任何新的数学问题, 大多数证明 (与勒贝格测度前提相比) 都是完全一样的。

- I. 如果随机变量 x 在 A 上可积, 则它在 A 的每一个子集 A' 上可积 ($A' \subset \mathcal{E}$)。
- II. 如果 x 在 A 上可积, 且 A 可以分解为 \mathcal{E} 上的不超过可数个的不相交集合 A_n , 则

$$\int_A x P(dE) = \sum_n \int_{A_n} x P(dE)$$

- III. 如果 x 可积, 则 $|x|$ 也可积, 并且在这种情况下

$$|\int_A x P(dE)| \leq \int_A |x| P(dE)$$

- IV. 如果对于每一个事件 ξ , 都有 $0 \leq y \leq x$ 成立, 则 x 、 y 也可积³⁵, 并且在这种情况下

$$\int_A y P(dE) \leq \int_A x P(dE)$$

- V. 如果 $m \leq x \leq M$, 其中 m 和 M 为两个常数, 则

$$mP(A) \leq \int_A x P(dE) \leq MP(A)$$

- VI. 如果 x 和 y 可积, 且 K 和 L 为两个实常数, 则 $Kx + Ly$ 也可积, 并且在这种情况下

$$\int_A (Kx + Ly) P(dE) = K \int_A x P(dE) + L \int_A y P(dE)$$

- VII. 如果级数

$$\sum_n \int_A |x_n| P(dE)$$

收敛，则级数

$$\sum_n x_n = x$$

在除使得 $P(B) = 0$ 成立的集合 B 之外，在集合 A 的任意一点处都收敛。如果在 $A - B$ 外的任意位置都令 $x = 0$ ，则有

$$\int_A x P(dE) = \sum_n \int_A x_n P(dE)$$

VIII. 如果 x 和 y 等价（即 $P\{x \neq y\} = 0$ ），则对于任意的 $A \subset \mathcal{E}$ ，都有

$$\int_A x P(dE) = \int_A y P(dE) \quad (3)$$

IX. 如果对于任意的 $A \subset \mathcal{E}$ 等式 (3) 都成立，则 x 和 y 等价。

根据前述的积分定义还可以得出以下通常勒贝格理论中没有的性质：

X. 令 $P_1(A)$ 和 $P_2(A)$ 为两个定义在相同 \mathcal{E} 上的概率函数， $P(A) = P_1(A) + P_2(A)$ ，并且 x 对于 $P_1(A)$ 和 $P_2(A)$ 在 A 上是可积的，则有

$$\int_A x P(dE) = \int_A x P_1(dE) + \int_A x P_2(dE)$$

XI. 每一个有界随机变量都是可积的。

绝对数学期望/条件数学期望

令 x 为一随机变量，则积分

$$E(x) = \int_E x P(dE)$$

称为变量 x 的数学期望 (mathematical expectation)。由性质 III、IV、V、VI、VII、VIII、XI，可得

I. $|E(x)| \leq E(|x|)$;

II. 如果 $0 \leq y \leq x$ ，则 $E(y) \leq E(x)$;

III. $\inf(x) \leq E(x) \leq \sup(x)$;

IV. $E(Kx + Ly) = KE(x) + LE(y)$;

V. 如果级数 $\sum_n E(|x_n|)$ 收敛，则 $E(\sum_n x_n) = \sum_n E(x_n)$;

VI. 如果 x 和 y 等价，则 $E(x) = E(y)$;

VII. 每个有界的随机变量都有数学期望。

由积分的定义可得

$$\begin{aligned} E(x) &= \lim_{k=-\infty}^{k=+\infty} km P\{km \leq x < (k+1)m\} \\ &= \lim_{k=-\infty}^{k=+\infty} \{F((k+1)m) - F(km)\} \end{aligned}$$

第二行是 Stieltjes 积分¹⁹ 的定义

$$\int_{-\infty}^{+\infty} adF^{(x)}(a) = E(x) \quad (1)$$

公式 (1) 可以作为数学期望 $E(x)$ 的定义。

现在令 u 为基本事件 ξ 的函数，且 $x = x(u)$ 为 u 的单值函数，则

$$P\{km \leq x < (k+1)m\} = P^{(u)}\{km \leq x(u) < (k+1)m\}$$

其中 $P^{(u)}(A)$ 是变量 u 的概率函数。由积分的定义可得

$$\int_E x P(dE) = \int_{E^{(u)}} x P^{(u)}(dE^{(u)})$$

因此

$$E(x) = \int_{E^{(u)}} x(u) P^{(u)}(dE^{(u)}) \quad (2)$$

其中 $E^{(u)}$ 表示所有 u 的可能取值构成的集合。

特别地, 当 u 本身是个随机变量时, 可得

$$E(x) = \int_E x P(dE) = \int_{R^1} x(u) P^{(u)}(dR^1) = \int_{-\infty}^{+\infty} x(a) dF^{(u)}(a) \quad (3)$$

当 $x(u)$ 连续时, 公式 (3) 的最后一项积分为常规的 Stieltjes 积分。此处有必要指出, 即使数学期望 $E(x)$ 不存在, 但积分

$$\int_{-\infty}^{+\infty} x(a) dF^{(u)}(a)$$

是存在的。 $E(x)$ 存在的充要条件是, 积分

$$\int_{-\infty}^{+\infty} |x(a)| dF^{(u)}(a)$$

是有限的³⁶。

如果 u 是空间 R^n 中的点 (u_1, u_2, \dots, u_n) , 则由公式 (2) 可得

$$E(x) = \int \int \cdots \int_{R^n} x(u_1, u_2, \dots, u_n) P^{(u_1, u_2, \dots, u_n)}(dR^n) \quad (4)$$

前文已经证明过, 条件概率 $P_B(A)$ 具有所有概率函数的性质。对应的积分

$$E_B(x) = \int_E x P_B(dE) \quad (5)$$

称为随机变量 x 在事件 B 下的条件数学期望。由于

$$\begin{aligned} P(\bar{B}) &= 0 \\ \int_{\bar{B}} x P_B(dE) &= 0 \end{aligned}$$

所以由公式 (5) 可得

$$E_B(x) = \int_E x P_B(dE) = \int_B x P_B(dE) + \int_{\bar{B}} x P_B(dE) = \int_B x P_B(dE)$$

已知当 $A \subset B$ 时有

$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)}$$

由此得

$$E_B(x) = \frac{1}{P(B)} \int_B x P(dE) \quad (6)$$

$$\int_B P(dE) = P(B) E_B(x) \quad (7)$$

由公式 (6) 和以下等式

$$\int_{A+B} x P(dE) = \int_A x P(dE) + \int_B x P(dE)$$

最终可得

$$E_{A+B}(x) = \frac{P(A)E_A(x) + P(B)E_B(x)}{P(A+B)} \quad (9)$$

特别地, 有如下等式

$$E(x) = P(A)E_A(x) + P(\bar{A})E_{\bar{A}}(x) \quad (9)$$

切比雪夫 (Tchebycheff) 不等式

$f(x)$ 为实数 x 的非负函数。如果对于 $x \geq a$, 恒有 $f(x) \geq b > 0$, 则如果数学期望 $E\{f(x)\}$ 存在, 则对于任意随机变量 x , 有

$$P(x \geq a) \leq \frac{E\{f(x)\}}{b} \quad (1)$$

证明:

$$E\{f(x)\} = \int_E f(x)P(dE) \geq \int_{\{x \geq a\}} f(x)P(dE) \geq bP(x \geq a)$$

例如, 对任意正数 c ,

$$P(x \geq a) \leq \frac{E(e^{cx})}{e^{ca}} \quad (2)$$

现在令 $f(x)$ 为非负的偶函数, 并且对于 $x > 0$, $f(x)$ 为非减函数。则对于任意随机变量 x 和任意选择的 $a > 0$, 以下不等式成立:

$$P(|x| \geq a) \leq \frac{E\{f(x)\}}{f(a)} \quad (3)$$

特别地,

$$P(|x - E(x)| \geq a) \leq \frac{E\{x - E(x)\}}{f(a)} \quad (4)$$

当 $f(x) = x^2$ 时此关系尤为重要, 由公式 (3) 和 (4) 可得

$$P(|x| \geq a) \leq \frac{E(x^2)}{a^2} \quad (5)$$

$$P(|x - E(x)| \geq a) \leq \frac{E\{x - E(x)\}^2}{a^2} = \frac{\sigma^2(x)}{a^2} \quad (6)$$

其中

$$\sigma^2(x) = E\{x - E(x)\}^2$$

称为变量 x 的方差 (variance)。很容易计算得出

$$\sigma^2(x) = E(x^2) - \{E(x)\}^2$$

如果 $f(x)$ 有界, 即

$$|f(x)| \leq K$$

则可以找到 $P(|x| \geq a)$ 的下界。

$$E(f(x)) = \int_E f(x)P(dE) = \int_{\{|x| < a\}} f(x)P(dE) + \int_{\{|x| \geq a\}} f(x)P(dE) \leq f(a)P(|x| < a) + KP(|x| \geq a) \leq f(a) + KP(|x| \geq a)$$

因此

$$P(|x| \geq a) \geq \frac{E\{f(x)\} - f(a)}{K} \quad (7)$$

如果不是 $f(x)$ 有界, 而是 x 本身有界, 即

$$|x| \leq M$$

则 $f(x) \leq f(M)$, 并且可得如下不等式

$$P(|x| \geq a) \geq \frac{E\{f(x)\} - f(a)}{f(M)} \quad (8)$$

当 $f(x) = x^2$ 时, 由

公式 (8) 可得

$$P(|x| \geq a) \geq \frac{E(x^2) - a^2}{M^2} \quad (9)$$

收敛判据

假定

$$x_1, x_2, \dots, x_n, \dots \quad (1)$$

为随机变量序列, $f(x)$ 为非负偶函数, 且当 $x > 0$ 时 $f(x)$ 单调递增³⁷, 则以下定理成立:

定理 I: 序列 (1) 依概率收敛的充分条件是: 对任意 $\epsilon > 0$, 总存在 n , 使得对每一个 $p > 0$, 以下不等式成立:

$$E\{f(x_{n+p} - x_n)\} < \epsilon \quad (2)$$

定理 II: 序列 (1) 依概率收敛到随机变量 x 的充分条件是:

$$\lim_{n \rightarrow +\infty} \{f(x_n - x)\} = 0 \quad (3)$$

定理 III: 如果 $f(x)$ 有界且连续, 并且 $f(0) = 0$, 则定理 I 和 II 中的充分条件也为必要条件。

定理 IV: 如果 $f(x)$ 连续, $f(0) = 0$, 且序列 $x_1, x_2, \dots, x_n, \dots, x$ 是有界的, 则定理 I 和 II 中的充分条件也为必要条件。

由定理 II 和定理 IV 可得定理 V:

定理 V: 序列 (1) 依概率收敛的充分条件是:

$$\lim E(x_n - x)^2 = 0 \quad (4)$$

如果序列 $x_1, x_2, \dots, x_n, \dots, x$ 是有界的, 则此充分条件也为必要条件。

定理 I - IV 的证明见 Slutsky [1]、Fréchet [1]。不过, 由上一节的公式 (3) 和 (8) 也可立即推得这些定理。

数学期望对参数的微分和积分

现在将每个基本事件 ξ 对应到一个定义在实变量 t 上的确定的实函数 $x(t)$ 。如果对于每一个固定的 t , 变量 $x(t)$ 是随机变量, 则称 $x(t)$ 为随机函数 (random variable)。现在的问题是, 在什么条件下数学期望符号 ($E(x)$) 可以与积分符号 (\int) 及微分符号 (dx) 互换。接下来的两个定理, 尽管无法穷尽这个问题, 但它们可以在许多简单的情况下给出足够令人满意的答案。

定理 I: 如果对于任意的 t , 数学期望 $E[x(t)]$ 是有限的, 且 $x(t)$ 总是可微的, 同时 $x(t)$ 对 t 的导数 $x'(t)$ 的绝对值总是小于某个常数 M , 则有

$$\frac{d}{dt} E[x(t)] = E[x'(t)]$$

定理 II: 如果 $x(t)$ 的绝对值总是小于某个常数 K , 且黎曼可积 (integrable in the Riemann sense¹⁹), 并且 $E[x(t)]$ 也是黎曼可积的, 则

$$\int_a^b E(x(t)) dt = E[\int_a^b x(t) dt]$$

定理 I 的证明: 假设 $x'(t)$ 为以下随机变量的极限:

$$\frac{x(t+h) - x(t)}{h} \quad h = 1, \frac{1}{2}, \dots, \frac{1}{n}, \dots$$

$x'(t)$ 也是随机变量。因为 $x'(t)$ 是有界的，所以数学期望 $E[x'(t)]$ 存在（第 2 节，数学期望的性质 VII）。现在固定 t ，记以下事件为 A ：

$$|\frac{x(t+h) - x(t)}{h} - x'(t)| > \epsilon$$

对于每一个 $\epsilon > 0$ ，当 $h \rightarrow 0$ 时，概率 $P(A)$ 趋近于 0。因为

$$|\frac{x(t+h) - x(t)}{h}| \leq M$$

和 $|x(t)| \leq M$ 总是成立的【前一个有定理 I 中导数有界可知，第二个为何总是成立？】，并且事件 A 的对立事件 \bar{A} 为

$$|\frac{x(t+h) - x(t)}{h} - x'(t)| \leq \epsilon$$

因此

$$|\frac{E[x(t+h)] - E[x(t)]}{h} - E[x'(t)]| \leq E|\frac{x(t+h) - x(t)}{h} - x'(t)| = P(A)E_A|\frac{x(t+h) - x(t)}{h} - x'(t)| + P(\bar{A})E_{\bar{A}}|\frac{x(t+h) - x(t)}{h} - x'(t)|$$

选择任意的 $\epsilon > 0$ ，对于任何足够小的 h ， $P(A)$ 可以任意小。因此有

$$\frac{d}{dt}E[x(t)] = \lim_{h \rightarrow 0} \frac{E[x(t+h)] - E[x(t)]}{h} = E[x'(t)]$$

定理 I 得证。

定理 II 的证明：令

$$S_n = \frac{1}{h} \sum_{k=1}^{k=n} x(t + kh), \quad h = \frac{b-a}{n}$$

因为 S_n 收敛于 $J = \int_a^b x(t)dt$ ，我们可以选择任意的 $\epsilon > 0$ 和 N ，使得当 $n \geq N$ 时，以下不等式成立

$$P(A) = P\{|S_n - J| > \epsilon\} < \epsilon$$

如果令

$$S_n^* = \frac{1}{h} \sum_{k=1}^{k=n} E[x(t + kh)] = E(S_n)$$

则

$$|S_n^* - E(J)| = |E(S_n - J)| \leq E|S_n - J| = P(A)E_A|S_n - J| + P(\bar{A})E_{\bar{A}}|S_n - J| \leq 2KP(A) + \epsilon \leq (2K+1)\epsilon$$

因此 S_n^* 收敛于 $E(J)$ ，由此得

$$\int_a^b E[x(t)]dt = \lim S_n^* = E(J)$$

定理 II 可以很容易推广到二重、三重以及更高阶的多重积分。此处给出这个定理在几何概率中的一个应用。令 G 为某一平面上的可测度区域，该平面的形状是随机的。换种方式表述即为，对概率域中的每个基本事件 ξ ，为其指定一个可测平面 G 。将区域 G 的面积记为 J ，点 (x, y) 属于区域 G 的概率记为 $P(x, y)$ 。则

$$E(J) = \int \int P(x, y)dxdy$$

要证明这一关系，只需记

$$J = \int \int f(x, y)dxdy$$

$$P(x, y) = Ef(x, y)$$

其中 $f(x, y)$ 是区域 G 的示性函数 (characteristic function) (在 G 上 $f(x, y) = 1$ ，在 G 外 $f(x, y) = 0$)

$$\int_a^b f(x) dx = \lim_{\|P\| \rightarrow 0} \sum_{i=1}^n f(\xi_i) \Delta x_i$$

其中：

- 分割 (Partition)：将区间 $[a, b]$ 分为子区间 $[x_{i-1}, x_i]$ ，记分割为 $P = \{x_0, x_1, \dots, x_n\}$ 。
- 样本点 (Tag)：在每个子区间中任取一点 $\xi_i \in [x_{i-1}, x_i]$ 。
- 黎曼和 (Riemann Sum)： $\sum_{i=1}^n f(\xi_i) \Delta x_i$ ，其中 $\Delta x_i = x_i - x_{i-1}$ 。
- 分割的模 (Norm)： $\|P\| = \max \Delta x_i$ ，表示最大子区间长度。

当所有可能的黎曼和在 $\|P\| \rightarrow 0$ 时趋于同一极限，则此极限为黎曼积分。

条件概率和数学期望

条件概率

在第一章第 6 节，我们定义了事件 B 相对于试验 \mathcal{U} 的条件概率 $P_{\mathcal{U}}(B)$ 。当时假设 \mathcal{U} 仅包含有限个不同的可能结果。实际上，也可在 \mathcal{U} 包括无限个可能结果的情况下定义 $P_{\mathcal{U}}(B)$ ，例如集合 E 被划分为无限多个不相交的子集的情况。特别地，如果考虑基本事件 ξ 的任意函数 u ，定义 u 等于常数的集合为划分 \mathcal{U}_u 的元素，则可以得到这样一种划分。条件概率 $P_{\mathcal{U}_u}(B)$ 也记为 $P_u(B)$ 。如果对于每一个基本事件 ξ 都定义一个 $u(\xi)$ ，使得 $u(\xi)$ 表示基本集合 E 的那些划分 \mathcal{U} 中，包括基本事件 ξ 的所有划分的集合，则对基本集合 E 的任意划分 \mathcal{U} 都可以定义为由基本事件 ξ 的函数 u 诱导出的划分 \mathcal{U}_u 。

当且仅当定义在域 $\mathcal{E}^{(u)}$ 上的 u 和定义在 $\mathcal{E}^{(u')}$ 上的 u' 之间存在一对一的对应关系 $u' = f(u)$ ，使得 $u'(\xi)$ 与 $f(u(\xi))$ 相同时，基本事件 ξ 的两个函数 u 和 u' 确定基本集合 E 的两个相同划分 $\mathcal{U}_u = \mathcal{U}_{u'}$ 。读者可以很容易证明，以下定义的随机变量 $P_u(B)$ 与 $P_{u'}(B)$ 在这种情况下是相同的。这实际上是由这两个划分相等决定的，即 $\mathcal{U}_u = \mathcal{U}_{u'}$ 。

为定义 $P_u(B)$ ，需要如下等式：

$$P_{\{u \subset A\}}(B) = E_{\{u \subset A\}}[P_u(B)] \quad (1)$$

如果对所有可能的 u ，集合 $E^{(u)}$ 都是有限的，则可以很容易证明，对于任意的 A （当 $P_u(B)$ 的定义如第一章第 6 节所述时），公式 (1) 总是成立的。在一般情况下（当 $P_u(B)$ 尚未定义时）则需要证明，总存在一个且仅有一个随机变量 $P_u(B)$ （排除等价随机变量的情况），它是 u 的函数，且对 $\mathcal{E}^{(u)}$ 中任意的 A ，当满足 $P^{(u)} > 0$ 时，公式 (1) 也总是成立的。由此在等价意义下确定的 u 的函数 $P_u(B)$ （determined to within equivalence），称为 B 对于给定 u 的条件概率。当 $u = a$ 时 $P_u(B)$ 的值记为 $P_u(a; B)$ 。

关于 $P_u(B)$ 的存在性和唯一性的证明：

在公式 (1) 的两边乘上 $P\{u \subset A\} = P^{(u)}(A)$ 可得：

等式左边：

$$P\{u \subset A\}P_{\{u \subset A\}}(B) = P(B\{u \subset A\}) = P(Bu^{-1}(A))$$

等式右边：

$$P\{u \subset A\}E_{\{u \subset A\}}[P_u(B)] = \int_{\{u \subset A\}} P_u(B)P(dE) = \int_A P_u(B)P^{(u)}(dE^{(u)})$$

由此得出以下等式

$$P(Bu^{-1}(A)) = \int_A P_u(B)P^{(u)}(dE^{(u)}) \quad (2)$$

反之，由 (2) 也可以推得 (1)。当 $P^{(u)}(A) = 0$ 时，公式 (1) 无意义，公式 (2) 显然成立且无需证明。公式 (2) 因此等价于公式 (1)。根据第四章第 1 节的性质 IX 可得，随机变量 x 由所有 \mathcal{E} 上的集合的积分的值唯一确定（排除等价随机变量的情况）

$$\int_A x P d(E)$$

因为 $P_u(B)$ 是定义在概率域 $(\mathcal{E}^{(u)}, P^{(u)})$ 上的随机变量，可知在不考虑等价随机变量的情况下， $P_u(B)$ 由公式 (2) 唯一确定。

现在还需要证明 $P_u(B)$ 是存在的。此处需要用到 Nikodym 定理³⁹：

令 \mathcal{E} 为一个波莱尔域， $P(A)$ 为定义在 \mathcal{E} 上的非负完全可加集函数（概率理论上的术语为，在 (\mathcal{E}, P) 上的随机变量）；令 $Q(A)$ 为定义在 \mathcal{E} 上的另一个完全可加集函数。由 $Q(A) \neq 0$ 可得 $P(A) > 0$ 。于是可知，存在一个在 \mathcal{E} 上可测的函数 $f(\xi)$ （概率论术语，随机变量），且对于每个 \mathcal{E} 中的集合 A ，满足以下公式

$$Q(A) \int_A f(\xi) P(dE)$$

为将这一定理用于本文的情形，需要证明

1. $Q(A) = P(Bu^{-1}(A))$ 是 $\mathcal{E}^{(u)}$ 上的完全可加集。
2. 由 $Q(A) \neq 0$ ，可得不等式 $P^{(u)}(A) > 0$ 。

首先，由

$$0 \leq P(Bu^{-1}(A)) \leq P(u^{-1}(A)) = P^{(u)}(A)$$

可证明第 2 点【大于等于的情况，排除了等于，就只剩下了大于】。

为证明第 1 点，令

$$A = \sum_n A_n$$

则

$$u^{-1}(A) = \sum_n u^{-1}(A_n)$$

并且

$$Bu^{-1}(A) = \sum_n Bu^{-1}(A_n)$$

因为 P 完全可加，所以有⁴⁰

$$P(Bu^{-1}(A)) = \sum_n P(Bu^{-1}(A_n))$$

证毕。

由公式 (1) 可得一个重要关系（如果令 $A = E^{(u)}$ ）

$$P(B) = E[P_u(B)] \quad (3)$$

接下来证明条件概率的两个基本性质。

定理 I: 以下关系几乎必然成立 (It is almost sure)

$$0 \leq P_u(B) \leq 1 \quad (4)$$

定理 II: 如果将集合 B 分解为至多可数个集合 B_n ，则以下等式几乎必然成立

$$P_u(B) = \sum_n P_u(B_n) \quad (5)$$

条件概率 $P_u(B)$ 的这两个人性对应概率函数 $P(B)$ 的两个人性： $0 \leq P(B) \leq 1$ 总是成立，以及 $P(B)$ 完全可加。这使得我们得以将绝对概率 $P(B)$ 的许多其他性质推广到条件概率 $P_u(B)$ 上。但必须时刻谨记，对于确定的集合 B ， $P_u(B)$ 是一个仅在等价意义下唯一确定的随机变量（a random variable determined uniquely only to within equivalence）。

定理 I 的证明：与待证命题相反，如果假设在集合 $M \subset E^{(u)}$ 上，且满足 $P^{(u)}(M) > 0$ 时，不等式 $P_u(B) \geq 1 + \epsilon$ ， $\epsilon > 0$ 总成立，则对应的公式 (1) 变为

$$P_{\{u \subset A\}}(B) = E_{\{u \subset A\}}[P_u(B)] \geq 1 + \epsilon$$

显然这是不可能的。同理可证明 $P_u(B) \geq 0$ 几乎必然成立。

定理 II 的证明：由以下序列的收敛

$$\sum_n E|P_u(B_n)| = \sum_n E(P_u(B_n)) = \sum_n P(B_n) = P(B)$$

并结合数学期望的性质 V (第四章第 2 节) 可得，序列

$$\sum_n P_u(B_n)$$

几乎必然收敛。因为序列

$$\sum_n E_{\{u \subset A\}}|P_u(B_n)| = \sum_n E_{\{u \subset A\}}(P_u(B_n)) = \sum_n P_{\{u \subset A\}}(B_n) = P_{\{u \subset A\}}(B)$$

对于每一个满足 $P^{(u)}(A) > 0$ 的集合 A 都是收敛的，所以由刚刚提到的数学期望的性质可得，对于每一个以上提到的集合 A ，有以下关系：

$$E_{\{u \subset A\}}(\sum_n P_u(B_n)) = \sum_n E_{\{u \subset A\}}(P_u(B_n)) = P_{\{u \subset A\}}(B) = E_{\{u \subset A\}}(P_u(B_n))$$

由此，等式 (5) 得证。

在结束本节前，还需要特别指出两种具体情形。首先，如果 $u(\xi) = c$ ， c 为常数，则 $P_c(A) = P(A)$ 。其次，如果令 $u(\xi) = \xi$ ，则可立即推得 $P_\xi(A)$ 在集合 A 上几乎必然为 1，在集合 \bar{A} 上几乎必然为 0。由此得 $P_\xi(A)$ 为集合 A 的示性函数。

波莱尔悖论的解释

取球面上的所有点的集合作为基本集合 E ，则 \mathcal{E} 为球面上所有波莱尔集合的构成的集合。最终， $P(A)$ 与集合 A 的测度成正比。现选择两个径向对称的点 (two diametrically opposite points) 作为球面上的两极，则每一个子午圈 (meridian circle) 都可以由经度 Ψ ， $0 \leq \Psi < \pi$ 唯一定义。因为 Ψ 的取值范围是 0 到 π (即此处考虑的是完整的子午圈而不仅是半子午圈)，纬度 Θ 必须从 $-\pi$ 到 $+\pi$ (不是 $-\pi/2$ 到 $+\pi/2$)。波莱尔提出了这样一个问题：给定经度 Ψ ，给出纬度 Θ ， $-\pi \leq \Theta < \pi$ 的条件概率分布。

很容易计算得【如何得到的】

$$P_\Psi\{\Theta_1 \leq \Theta < \Theta_2\} = \int_{\Theta_1}^{\Theta_2} |\cos \Theta| d\Theta$$

给定 Ψ 时， Θ 的概率分布不是均匀的。

如果假设“假定 ξ 位于给定子午圈上”时 Θ 的条件概率必须是均匀的，则此处产生了矛盾。

这表明，针对一个概率为零的孤立假设 (例如单一点或零测集) 来定义条件概率的概念是不可行的。因为只有当我们把子午圈视为整个球面按给定极点分解为若干子午圈这一分解结构中的一个元素时，才能获得该子午圈上 Ψ 的概率分布。

随机变量的条件概率

如果 X 是随机变量, $P_x(B)$ 为 X 的函数, 且 $P_x(B)$ 是波莱尔可测的, 则 $P_x(B)$ 可以以一种基本的方式定义。我们可以以另一种形式重写第 1 节的公式 (2):

$$P(B)P_B^{(x)}(A) = \int_A P_x(B)P^{(x)}(dE) \quad (1)$$

由此可由公式 (1) 立刻得出

$$P(B)F_B^{(x)}(a) = \int_{-\infty}^a P_x(a; B)dF^{(x)}(a) \quad (2)$$

根据勒贝格 (Lebesgue) 的一个定理⁴¹, 由公式 (2) 可得

$$P_x(a; B) = P(B) \lim \frac{F_B^{(x)}(a+h) - F_B^{(x)}(a)}{F^{(x)}(a+h) - F^{(x)}(a)}, \quad h \rightarrow 0 \quad (3)$$

除在满足 $P^{(x)}(H) = 0$ 的点集 H 外, 公式 (3) 总是成立。

$P_x(a; B)$ 在本章第 1 节定义 (除在满足 $P^{(x)}(G)$ 的集合 G 之外)。如果现在把公式 (3) 作为 $P_x(a; B)$ 的定义 (当公式 (3) 右边的极限不存在时, 令 $P_x(a; B) = 0$) , 则这个新的变量满足第 1 节中所有的要求。

此外, 如果概率密度 $f^{(x)}$ 和 $f_B^{(x)}(a)$ 存在, 且 $f^{(x)}(a) > 0$, 则公式 (3) 又可以写成

$$P_x(a; B) = P(B) \frac{f_B^{(x)}(a)}{f^{(x)}(a)} \quad (4)$$

此外, 如果公式 (3) 中的极限存在, 并且概率密度 $f^{(x)}(a)$ 存在, 则 $f_B^{(x)}(a)$ 存在。因此

$$P(B)f_B^{(x)}(a) \leq f^{(x)}(a) \quad (5)$$

如果 $P(B) > 0$, 则由公式 (4) 可得

$$f_B^{(x)}(a) = \frac{P_x(a; B)f^{(x)}(a)}{P(B)} \quad (6)$$

当 $f^{(x)}(a) = 0$ 时, 由公式 (5) 可得 $f_B^{(x)}(a) = 0$ 并且此时公式 (6) 依然成立。此外, 如果 X 的分布是连续的, 可得

$$P(B) = E(P_x(B)) = \int_{-\infty}^{+\infty} P_x(a; B)dF^{(x)}(a) = \int_{-\infty}^{+\infty} P_x(a; B)f^{(x)}(a)da \quad (7)$$

由公式 (6) 和 (7) 可得

$$f_B^{(x)}(a) = \frac{P_x(a; B)f^{(x)}(a)}{\int_{-\infty}^{+\infty} P_x(a; B)f^{(x)}(a)da} \quad (8)$$

这个公式给出了连续分布的贝叶斯定理 (Bayes Theorem for continuous distributions)。这个定理得以证明的前提假设如下:

- $P_x(B)$ 是波莱尔可测的, 且在点 a 它由公式 (3) 定义;
- X 为连续分布;
- 在点 a 处存在概率密度 $f_B^{(x)}(a)$ 。

条件数学期望

令 u 为变量 ξ 的任意函数, y 为一个随机变量。随机变量 $E_u(y)$ 可以表示为 u 的函数, 且对任意的 $\mathcal{E}^{(u)}$ 中的集合 A , 满足 $P^{(u)}(A) > 0$, 则以下定义 (如果存在)

$$E_{\{u \subset A\}}(y) = E_{\{u \subset A\}}E_u(y) \quad (1)$$

称为给定 u 时, 变量 y 的条件数学期望 (conditional mathematical expectation)。

如果给公式 (1) 两边乘 $P^{(u)}(A)$, 则有

$$\int_{\{u \subset A\}} y P(dE) = \int_A E_u(y) P^{(u)}(dE^{(u)}) \quad (2)$$

相反地，由公式 (2) 也可得公式 (1)。当 $P^{(u)}(A) = 0$ 时，公式 (1) 无意义，公式 (2) 是显然成立的。以与第 1 节中条件概率相同的方式，可以证明 $E_u(y)$ 由公式 (2) 唯一确定（等价情况除外）。

当 $u = a$ 时 $E_u(y)$ 的值记为 $E_a(y)$ 。同样需要说明的是， $E_u(y)$ 和 $P_u(y)$ 仅由分割 \mathcal{U}_u 决定，所以 $E_u(a; y)$ 也可以记为 $E_{\mathcal{U}_u}(y)$ 。

$E(y)$ 的存在性隐含在 $E_u(y)$ 的定义中（令 $A = E^{(u)}$ ，则 $E_{\{u \subset A\}}(y) = E(y)$ ）。

接下来证明 $E(y)$ 的存在足以保证 $E_u(y)$ 的存在。为此只需证明，由 Nikodym 定理（见本章第 1 节），可得 1) 集合函数在 $\mathcal{E}^{(u)}$ 上完全可加，并且 2) 关于 $P^{(u)}(A)$ 绝对连续。条件 1 可以按与第 1 节条件概率的情况几乎完全相同的方式证明。条件 2（绝对连续性）包含在这一事实中：由 $Q(A) \neq 0$ 可得， $P^{(u)}(A) > 0$ 。如果假定 $P^{(u)}(A) = P\{u \subset A\} = 0$ ，则显然有

$$Q(A) = \int_{\{u \subset A\}} y P(dE) = 0$$

由此第二个条件得到满足。

如果在公式 (1) 中令 $A = E^{(u)}$ ，则有

$$E(y) = E E_u(y) \quad (3)$$

进一步可以证明

$$E_u(ay + bz) = aE_u(y) + bE_u(z) \quad (4)$$

其中 a 和 b 为任意常数（证明留给读者）。

如果 u 和 v 是基本事件 ξ 的两个函数，则二元组 (u, v) 可以总是被当作 ξ 的函数。于是有以下重要等式成立

$$E_u E_{(u,v)}(y) = E_u(y) \quad (5)$$

证明如下： $E_u(y)$ 的定义如下：

$$E_{\{u \subset A\}}(y) = E_{\{u \subset A\}} E_u(y)$$

因此必须证明 $E_u E_{(u,v)}(y)$ 满足如下关系：

$$E_{\{u \subset A\}}(y) = E_{\{u \subset A\}} E_u E_{(u,v)}(y) \quad (6)$$

由 $E_{(u,v)}(y)$ 的定义可知

$$E_{\{u \subset A\}}(y) = E_{\{u \subset A\}} E_{(u,v)}(y) \quad (7)$$

由 $E_u E_{(u,v)}(y)$ 的定义可知

$$E_{\{u \subset A\}} E_{(u,v)}(y) = E_{\{u \subset A\}} E_u E_{(u,v)}(y) \quad (8)$$

由公式 (7) 和公式 (8) 可证明公式 (6)，由此证明公式 (5)。

如果令 $y = P_u(B)$ ，在集合 B 上 $y = 1$ ，在集合 B 外 $y = 0$ ，则

$$\begin{aligned} E_u(y) &= P_u(B) \\ E_{(u,v)}(y) &= P_{(u,v)}(B) \end{aligned}$$

此时由公式 (5) 可得

$$E_u P_{(u,v)}(B) = P_u(B) \quad (9)$$

条件数学期望 $E_u(y)$ 也可以直接通过对应的条件概率的方式定义。为此考虑如下和式

$$S_\lambda(u) = \sum_{k=-\infty}^{k=+\infty} k \lambda P_u\{k \lambda \leq y < (k+1)\lambda\} = \sum_k P_k \quad (10)$$

如果 $E(y)$ 存在，则级数 (10) 几乎必定（almost certainly）¹⁹ 是收敛的。由第 1 节的公式 (3) 可知

$$E|R_k| = |k\lambda| P\{k\lambda \leq y < (k+1)\lambda\}$$

以及下面级数的收敛性

$$\sum_{k=-\infty}^{k=+\infty} |k\lambda| P\{k\lambda \leq y < (k+1)\lambda\} = \sum_k E|R_k|$$

是 $E(y)$ 存在的必要条件 (见第四章, 第 1 节)。由这一收敛性可得级数 (10) 几乎必定收敛 (见第四章, 第 2 节, 性质 V)。可以进一步证明, 正如勒贝格积分理论说明的那样, 由级数 (10) 在当 λ 取一些值时收敛可推得, 它对所有的 λ 都收敛; 当级数 (10) 收敛时, 在 $\lambda \rightarrow 0$ 时⁴² $S_\lambda(u)$ 趋向于一个确定的极限, 由此可以定义

$$E_u(y) = \lim_{\lambda \rightarrow 0} S_\lambda(u) \quad (11)$$

为证明公式 (11) 定义的条件期望 $E_u(y)$ 满足以上的种种条件, 仅需证明它满足公式 (1):

$$E_{\{u \subset A\}} E_u(y) = \lim_{\lambda \rightarrow 0} E_{\{u \subset A\}} S_\lambda(u) = \lim_{\lambda \rightarrow 0} \sum_{k=-\infty}^{k=+\infty} k\lambda P_{\{u \subset A\}} \{k\lambda \leq y < (k+1)\lambda\} = E_{\{u \subset A\}}(y)$$

以上计算中, 数学期望符号和积分符号可交换, 因为当 $\lambda \rightarrow 0$ 时 $S_\lambda(u)$ 一致收敛 (converges uniformly) 于 $E_u(y)$ (由第 2 节数学期望的性质 V 可知)。数学期望符号和积分符号可交换的另一个原因是, 以下级数是收敛的 (数学期望性质 V 的直接结论) :

$$\sum_{k=-\infty}^{k=+\infty} E_{\{u \subset A\}} \{|k\lambda| P_u[k\lambda \leq y < (k+1)\lambda]\} = \sum_{k=-\infty}^{k=+\infty} |k\lambda| P_{\{u \subset A\}}[k\lambda \leq y < (k+1)\lambda]$$

公式 (11) 的等效的另一种写法是

$$E_u(y) = \int_E y P(dE) \quad (12)$$

此处需要说明, 在第四章第 1 节所述内容的意义下, 公式 (12) 并不是积分, 所以它仅仅是一个符号化 (symbolic) 的表达。

如果 x 是个随机变量, 则称 x 和 a 的函数

$$F_x^{(y)}(a) = P_x(y < a)$$

为 y 对于已知 x 的条件分布函数 (conditional distribution function of y for known x)。

$F_x^{(y)}(a)$ 几乎对所有的 a 都有定义。如果 $a < b$, 则以下不等关系几乎必定成立

$$F_x^{(y)}(a) \leq F_x^{(y)}(b)$$

由公式 (10) 和公式 (11) 可得⁴³ 以下公式几乎必然成立

$$E_x(y) = \lim_{\lambda \rightarrow 0} \sum_{k=-\infty}^{k=+\infty} k\lambda [F_x^{(y)}((k+1)\lambda) - F_x^{(y)}(k\lambda)] \quad (13)$$

这一事实可由以下公式作一符号化的表达

$$E_x(y) = \int_{-\infty}^{+\infty} adF_x^{(y)}(a) \quad (14)$$

由新获得的数学期望的定义 (公式 (10) 和公式 (11)), 很容易证明, 对任意实函数 u

$$E_u[f(u)y] = f(u)E_u(y) \quad (15)$$

因此

$$E(z_{ni}z_{nk}) = 0$$

由此可得

$$\sigma^2(s_n) = \sigma^2(z_{n1}) + \sigma^2(z_{n2}) + \cdots + \sigma^2(z_{nn}) = \beta_{n1}^2 + \beta_{n2}^2 + \cdots + \beta_{nn}^2$$

因此, 以下公式

$$\beta_{n1}^2 + \beta_{n2}^2 + \cdots + \beta_{nn}^2 \rightarrow 0 \quad n \rightarrow 0$$

为 s_n 正规稳定的充分条件。

关于数学期望概念的说明

由前文内容可知，随机变量 x 的数学期望定义为

$$E(x) = \int_E P(dE) = \int_{-\infty}^{+\infty} adF^x(a)$$

其中上式右侧的积分可解释如下

$$E(x) = \int_{-\infty}^{+\infty} adF^x(a) = \lim_{b \rightarrow -\infty, c \rightarrow +\infty} \int_b^c adF^x(a) \quad (1)$$

因此，下式可称为一种广义数学期望 (a generalized mathematical expectation)

$$E^*(x) = \lim_{b \rightarrow +\infty} \int_{-b}^{+b} a dF^x(a) \quad (2)$$

当然在这种情况下会失去数学期望的一些简单性质。而且此时

$$E(x+y) = E(x) + E(y)$$

并不总是成立。以这种方式的推广几乎是难以接受的。但如果加上一些限制性的补充条件的话，定义 (2) 就变得完全自然且实用了。

可以按如下方式讨论此问题，令

$$x_1, x_2, \dots, x_n, \dots$$

为相互独立的随机变量序列，并且这个随机变量序列与随机变量 x 有相同的分布函数，即 $F^{(x)}(a) = F^{(x_n)}(a)$ ，($n = 1, 2, \dots$)。进一步令

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

此时请问，是否存在一个常量 $E^*(x)$ ，满足对任意的 $\epsilon > 0$ ，下式都成立

$$\lim P(|s_n - E^*(x)| > \epsilon) = 0, \quad n \rightarrow +\infty \quad (3)$$

答案：如果常量 $E^*(x)$ 存在，则它可以由公式 (2) 表示。公式 (3) 成立的充要条件为：1) 公式 (2) 表示的极限存在；2) 公式 (4) 表示的关系成立。

$$P(|x| > n) = o\left(\frac{1}{n}\right) \quad (4)$$

为证明此结论，我们应用如下定理：

条件 (4) 是算数平均 s_n 稳定的充要条件。

当 s_n 稳定时令⁴⁴

$$d_n = \int_{-n}^{+n} a F^{(x)}(a)$$

如果存在公式 (1) 定义的数学期望，则条件 (4) 总是可以得到满足⁴⁵。因为此时 $E(x) = E^*(x)$ ，所以条件 (3) 实际上确实定义了数学期望概念的一种推广。对于此广义的数学期望，性质 I - VII (第四章第 2 节) 仍然成立。但 $E^*(x)$ 的存在并不能推出 $E^*|x|$ 的存在。

为证明新的数学期望的概念比先前的概念更具普遍性，只需给出以下示例即可。令概率密度函数 $f^{(x)}(a)$ 如下

$$f^{(x)}(a) = \frac{C}{(|a| + 2)^2 \ln(|a| + 2)}$$

其中常量 C 由下式确定

$$\int_{-\infty}^{+\infty} f^{(x)}(a) da = 1$$

很容易计算得出，此时条件 (4) 得到满足。由公式 (2) 可得

$$E^*(x) = 0$$

但积分

$$\int_{-\infty}^{+\infty} |a| dF^{(x)}(a) da = \int_{-\infty}^{+\infty} |a| f^{(x)}(a) da$$

是发散的。

强大数定律；级数收敛

以下序列

$$s_1, s_2, \dots, s_n, \dots$$

的随机变量记为 s_n 。如果存在实数序列

$$d_1, d_2, \dots, d_n, \dots$$

使得当 $n \rightarrow +\infty$ 时随机变量

$$s_n - d_n$$

几乎处处趋近于 0，则 s_n 是强稳定的。显然由强稳定性可推得普通稳定性 (ordinary)。如果令

$$d_n = E(s_n)$$

则此时的强稳定是正规强稳定 (normal)。

在切比雪夫情形 (见本章第 3 节) 下

$$s_n = \frac{x_1 + x_2 + \dots + x_n}{n}$$

其中 x_n 是相互独立的。此时以下序列的收敛是 s_n 正规稳定的一个充分条件⁴⁶

$$\sum_{n=1}^{\infty} \frac{\sigma^2(x_n)}{n^2} \tag{1}$$

从某种意义上讲，这个条件是最优的。因为对任意满足以下条件的 b_n

$$\sum_{n=1}^{\infty} \frac{b_n}{n^2} = +\infty$$

我们都可以构造一个相互独立的随机变量 x_n 的序列，满足 $\sigma^2(x_n) = b_n$ ，使得 s_n 不是强稳定的。

如果所有的 x_n 都有相同的分布函数 $F^{(x)}(a)$ ，则数学期望

$$E(x) = \int_{-\infty}^{+\infty} a dF^{(x)}(a)$$

的存在性是 s_n 强稳定的充分必要条件，并且此时的稳定性总是正规的⁴⁷。

令

$$x_1, x_2, \dots, x_n, \dots$$

为相互独立的随机变量，则级数

$$\sum_{n=1}^{\infty} x_n \tag{2}$$

收敛的概率要么等于 1，要么等于 0。特别地，当以下两个级数都收敛时，级数 (2) 收敛的概率等于 1

$$\begin{aligned} & \sum_{n=1}^{\infty} E(x_n) \\ & \sum_{n=1}^{\infty} \sigma^2(x_n) \end{aligned}$$

进一步假设

$$\begin{aligned} y_n = x_n & \quad \text{in case } |x_n| \leq 1 \\ y_n = 0 & \quad \text{in case } |x_n| > 1 \end{aligned}$$

则为级数 (1) 依概率 1 收敛的充分必要条件⁴⁸ 是, 以下级数同时收敛

$$\begin{aligned} & \sum_{n=1}^{\infty} P\{|x_n| > 1\} \\ & \sum_{n=1}^{\infty} E(y_n) \\ & \sum_{n=1}^{\infty} \sigma^2(y_n) \end{aligned}$$

附录：概率理论中的零一定律

在以上分析中, 我们注意到某些极限概率等于零或一的情况。例如, 独立随机变量级数收敛的概率只可能是 0 或 1⁴⁹。现在我们证明包括此种情况的一种更广义的定理。

定理: 令 $x_1, x_2, \dots, x_n, \dots$ 为随机变量, 令 $f(x_1, x_2, \dots, x_n, \dots)$ 为 $x_1, x_2, \dots, x_n, \dots$ 的贝尔函数 (Baire function)⁵⁰, 其中当前 n 个变量 x_1, x_2, \dots, x_n 已知时, 对于所有的 n , 以下关系

$$f(x_1, x_2, \dots, x_n, \dots) = 0$$

成立的条件概率

$$P_{x_1, x_2, \dots, x_n}\{f(x) = 0\}$$

等于绝对概率

$$P\{f(x) = 0\} \tag{1}$$

在这些条件下, 概率 (1) 等于 0 或者 1。

特別地, 当 1) 变量 x_n 相互独立, 并且 2) 当仅有有限个变量的值改变时, 函数 $f(x)$ 的值保持不变时, 此定理的前提成立。

证明: 用 A 表示事件 $f(x) = 0$ 。之后来研究由有限个随机变量 x_n 中的关系定义的所有事件构成的域 \mathcal{R} 。如果事件 B 属于域 \mathcal{R} , 则根据定理的条件可得

$$P_B(A) = P(A) \tag{2}$$

当 $P(A) = 0$ 时, 此定理显然是正确的。接下来令 $P(A) > 0$ 。则根据公式 (2) 可得以下等式

$$P_A(B) = \frac{P_B(A)P(B)}{P(A)} = P(B) \tag{3}$$

因此 $P(B)$ 和 $P_A(B)$ 是在域 \mathcal{R} 上取值相同的两个完全可加集函数。因此它们必定在域 \mathcal{R} 的波莱尔扩展 $B\mathcal{R}$ 的每一个集合上相等。因此, 特別地

$$P(A) = P_A(A) = 1$$

证毕!

其他几个概率只可以取 0 或 1 的几个例子, 已由以下文献发现:

P. Lévy. See P. LÉVY, Sur un théorème de M. Khintchine, Bull. des Sci. Math. v. 55, 1931, pp. 145–160,
Theorem II.

参考文献

- [1]. BERNSTEIN, S.: On the axiomatic foundation of the theory of probability. (In Russian). Mitt. Math. Ges. Charkov, 1917, Pp. 209–274.
- [2]. — Theory of probability, 2nd edition. (In Russian). Moscow, 1927. Government publication RSFSR.
- [1]. BOREL, E.: Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. mat. Palermo Vol. 27 (1909) Pp. 247–271.
- [2]. — Principes et formules classiques, fasc. 1 du tome I du Traité des probabilités par E. BOREL et divers auteurs. Paris: Gauthier-Villars 1925.
- [3]. — Applications à l'arithmétique et à la théorie des fonctions, fasc. 1 du tome II du Traité des probabilités par E. BOREL et divers auteurs. Paris: Gauthier-Villars 1926.
- [1]. CANTELLI, F. P. : Una teoria astratta del Calcolo delle probabilità. Giorn. 1st. Ital. Attuari Vol. 3 (1932) pp. 257–265.
- [2]. — Sulla legge dei grandi numeri. Mem. Acad. Lincei Vol. 11 (1916).
- [3]. — Sulla probabilità come limite della frequenza. Rend. Accad. Lincei Vol. 26 (1917) Pp. 39–45.
- [1]. COPELAND H : The theory of probability from the point of view of admissible numbers. Ann. Math. Statist. Vol. 3 (1932) Pp. 143–156.
- [1]. DÖRGE, K. : Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. Math. Z. Vol. 32 (1930) Pp. 232–258.
- [1]. FRECHET, M.: Sur la convergence en probabilité. Metron Vol. 8 (1930) Pp. 1–48.
- [2]. — Recherches théoriques modernes, fasc. 3 du tome I du Traité des probabilités par E. Borel et divers auteurs. Paris: Gauthier-Villars.
- [1]. KOLMOGOROV, A.: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. Vol. 104 (1931) Pp. 415–458.
- [2]. — The general theory of measure and the theory of probability. (In Russian). Sbornik trudov sektii totshnykh nauk K. A., Vol. 1 (1929) pp. 8–21.
- [1]. LÉVY, P.: Calcul des probabilités. Paris: Gauthier-Villars.
- [1]. LOMNICKI, A.: Nouveaux fondements du calcul des probabilités. Fundam. Math. Vol. 4 (1923) Pp. 34–71.
- [1]. MISES, R. v. : Wahrscheinlichkeitsrechnung. Leipzig u. Wien: Fr. Deuticke 1931.
- [2]. — Grundlagen der Wahrscheinlichkeitsrechnung. Math. Z. Vol. 5 (1919) pp. 52–99.
- [3]. — Wahrscheinlichkeitsrechnung, Statistik und Wahrheit. Wien: Julius Springer 1928.
- [3']. — Probability, Statistics and Truth (translation of above). New York: The MacMillan Company 1939.
- [1]. REICHENBACH, H.: Axiomatik der Wahrscheinlichkeitsrechnung. Math. Z. Vol. 34 (1932) Pp. 568–619.
- [1]. SLUTSKY, E.: Über stochastische Asymptoten und Grenzwerte. Metron Vol. 5 (1925) Pp. 3–89.
- [2]. — On the question of the logical foundation of the theory of probability. (In Russian). Westnik Statistiki, Vol. 12 (1922), pp. 13–21.
- [1]. STEINHAUS, H.: Les probabilités dénombrables et leur rapport à la théorie de la mesure. Fundam. Math. Vol. 4 (1923) Pp. 286–310.
- [1]. TORNIER, E.: Wahrscheinlichkeitsrechnung und Zahlentheorie. J. reine angew. Math. Vol. 160 (1929) Pp. 177–198.
- [2]. — Grundlagen der Wahrscheinlichkeitsrechnung. Acta math. Vol. 60 (1933) Pp. 239–380.

关于补充参考文献的说明

概率论的测度论方法的奠基性著作是 A. N. 柯尔莫哥洛夫的《Grundbegriffe der Wahrscheinlichkeitsrechnung》（《概率论基本概念》德文版本），本文是其英文翻译版本。可以毫不夸张地说，在过去的二十三年间，大部分概率论研究都受到这一方法的影响，而且柯尔莫哥洛夫提出的公理体系已被概率论与统计学界学者公认为正确的理论基础。

柯尔莫哥洛夫《Grundbegriffe der Wahrscheinlichkeitsrechnung》的出版开启了概率论及其研究方法的新纪元，基于柯尔莫哥洛夫提出的基本概念所产生的研究成果非常丰富。在准备柯尔莫哥洛夫专著英文译本第二版的过程中，我们认为有必要提供一份能反映概率论研究现状与方向的书目。

近几年出版了许多优秀的（概率论相关的）书籍，其中最具代表性的三部分别是Doob [12], Feller [17]，和 Loève [54]的作品。以下是涉及概率论通论及概率专题研究的其他著作：[2], [3], [6], [7], [9], [19], [23], [26], [27], [28], [34], [39], [41], [42], [47], [49], [50], [67], [70], [72]。鉴于这些著作已收录大量文献索引，此处将重点列出近年已发表的研究论文和即将发表的论文。

柯尔莫哥洛夫建立的模型可简要表述如下：在所有包含随机因素的情况下（例如，实验、观测等），都有一个与之关联的概率空间或三元组 (Ω, ξ, p) ，其中 Ω 是抽象空间（基本事件的空间）， ξ 是 Ω 的子集（事件的集合）的 σ 代数， $p(E)$ 是根据 E 、 ϵ 、 ξ 定义的一个测度（事件 E 的概率），并且满足 $p(\Omega) = 1$ 。Loś [56] 最近研究了柯尔莫哥洛夫的模型，他尝试使用抽象代数和集合的 σ 代数来代替传统意义上的代数与 σ 代数。柯尔莫哥洛夫 [44] 也尝试过在概率论中使用度量布尔代数（metric Boolean algebras）。

有许多问题，尤其是在理论物理领域，无法纳入柯尔莫哥洛夫理论，其原因在于这些问题涉及无界测度（unbounded measures）。Rényi [68] 建立了一种广义的公理化概率理论（该理论将柯尔莫哥洛夫的理论作为一个特例），这一理论允许无界测度的存在。这一理论的基础概念是事件的条件概率。Császár [10] 研究了 Rényi 的理论中的条件概率空间的测度论结构。

在另一个方向上，很多作者指出柯尔莫哥洛夫的理论过于宽泛。Gnedenko 和柯尔莫哥洛夫 [27] 提出了一种称为完美概率空间（perfect probability space）的限制性更强的概念。完美概率空间是这样一个三元组 (Ω, ξ, p) ，满足对任意实值 ξ 可测函数 g 和任意线性集 B ，若集合 $\{\omega : g(\omega) \in B\} \in \xi$ ，则存在一个波莱尔集 $D \subseteq B$ ，满足 $P\{\omega : g(\omega) \in D\} = P\{\omega : g(\omega) \in B\}$ 。最近，Blackwell [5] 提出了一个比完美概率空间更严格的概念——Lusin 空间。Lusin 空间是一个满足如下条件的数对 (Ω, ξ) ：

- ξ 是可分的；
- Ω 上的每一个实值 ξ 可测函数 g 的值域，是一个解析集（analytic set）。

显然，如果 (Ω, ξ, p) 是一个 Lusin 空间，并且 p 是 ξ 上的任意概率测度，则 (Ω, ξ, p) 是一个完美概率空间。

在第六章第 1 节，柯尔莫哥洛夫给出了马尔可夫链的定义。近些年里马尔可夫链理论和马尔可夫过程是概率领域最活跃的研究领域之一。文献 [17] 提供了该理论的一份精彩的导论。其他参考文献包括 [2], [3], [6], [12], [19], [23], [26], [34], [39], [50], [54], [67], [70], [72]。还有两篇值得关注的论文：Harris 和 Robbins [29] 关于马尔可夫链的各态便利性理论（ergodic theory），以及 Chung [8] 的关于状态数可数情况下的连续参数过程理论。Chung 提出的理论统一并扩展了 Doob（参考文献 [12]）和 Lévy [51], [52], [53] 的研究成果。

一些概率论研究工作者正在利用半群理论 [30] 来研究马尔可夫过程及其结构性质 [63]。在这一方法主要是 Yosida [80] 提出的，由一个 Banach 空间到自身的单参数（离散或连续）算子半群定义马尔可夫过程。Hille [32] 和 Kato [38] 使用半群方法对柯尔莫哥洛夫微分方程做积分，Kendall 和 Reuter [40] 研究了该理论中出现的一些病态案例。Feller [18] 和 Hille [31] 研究了连续情况下的抛物型微分方程（parabolic differential equations）。Doob [13] 使用半群理论中的鞅论（martingale theory）研究一维扩散过程。Hunt [33] 研究了李群（Lie groups）上的半群（概率）测度。

最近的研究论文致力于使用更抽象的方式研究概率。这些研究探讨具有代数结构的拓扑空间中的随机变量。文献 [14], [21], [22], [58], [59], [61] 研究了与巴拿赫空间（Banach space）中的随机变量有关的问题，文献 [4] 研究了 Orlicz 空间（广义勒贝格空间）中的类似问题。Robbins [69] 研究了任意紧致拓扑空间（compact topological group）中的随机变量。Segal [75] 研究了概率代数的结构，并且使用这一代数方法推广了柯尔莫哥洛夫关于具有任意指定联合分布的实值随机变量的存在性定理（参考第三章第 4 节）。Segal [76, 第三章第 13 节¹⁹] 还研究过非交换概率理论（non-commutative probability theory）。

Prohorov [66] 研究了定义在巴拿赫空间和其他函数空间上的概率分布的收敛性质。LeCam [48] 和 Parzen [64] 也研究过这些问题。

柯尔莫哥洛夫在第四章（以及文献 [12] 和 [54]）中给出了条件概率和条件数学期望的测度论定义和基本性质。使用抽象代数方法，S. T. C. Moy [60] 将条件数学期望的性质视为概率空间上所有的扩展实值可测函数（extended real-valued measurable functions）空间到自身的线性变换。在文献 [61] 中，她研究了巴拿赫空间随机变量的条件数学期望。Nakamura 和 Turamuru [62] 将数学期望视为 C^* 代数的一种给定运算。Umegaki [79] 将条件数学期望作为属于具有 W^* 代数的 L_1 可积类（integrable class）的可测算子空间到自身的映射。Umegaki 的研究发展了非可交换概率理论。以上研究用到了 Segal [74]、Dye [15] 等学者对抽象积分理论的研究成果。其他相关研究论文包括 [1], [16], [36], [45]。

Gel'fand [24] 对广义随机过程（generalized stochastic processes）的研究用到了 L. Schwartz 关于分布的理论（广义函数理论，The L. Schwartz theory of distributions）[73]。后者也被 Fortet [20] 和 Itô [35] 用来研究随机分布。

还有一些研究者致力于研究概率论中的极限定理：[27], [42], [47], [49]。此外，还有必要参考下文献 [12] 和 [54]。另外还有研究论文和综述论文：[11], [14], [25], [37], [46], [55], [57], [65], [71], [77], [78]。

补充参考文献

- [1] ALDA, V., On Conditional Expectations, Czechoslovak Math. J., Vol. 5 (1955), pp. 503–505.
- [2] ARLEY, N., On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation, Copenhagen, 1943.
- [3] BARTLETT, M. S., An Introduction to Stochastic Processes, Cambridge, 1955.
- [4] BHARUCHA-REID, A. T., On Random Elements in Orlicz Spaces, (Abstract) Bull. Amer. Math. Soc., Vol. 62 (1956). To appear.
- [5] BLACKWELL, D., On a Class of Probability Spaces, Proc. Third Berkeley Symposium on Math. Statistics and Probability, Vol. 2 (1956). To appear.
- [6] BLANC-LAPIERRE, A., and R. Fortet, Théorie des fonctions aléatoires, Paris, 1953.
- [7] BOCHNER, S., Harmonic Analysis and the Theory of Probability, Berkeley and Los Angeles, 1955.
- [8] CHUNG, K. L., Foundations of the Theory of Continuous Parameter Markov Chains, Proc. Third Berkeley Symposium on Math. Statistics and Probabiilty, Vol. 2 (1956). To appear.
- [9] CRAMÉR, H., Mathematical Methods of Statistics, Princeton, 1946.
- [10] CSÁSZÁR, A., Sur le structure des espace de probabilité conditionnelle, Acta Math. Acad. Sci. Hung., Vol. 6 (1955), pp. 337–361.
- [11] DERNAN, C., and H. ROBBINS, The Strong Law of Large Numbers when the First Moment does not Exist, Proc. Nat. Acad. Sci. U.S.A., Vol. 41 (1955), pp.586–587.
- [12] DOOB, J. L., Stochastic Processes, New York, 1953.
- [13] DOOB, J. L., Martingales and One-Dimensional Diffusion, Trans. Amer. Math. Soc., Vol. 78 (1955), pp. 168–208.
- [14] DOSS, S., Sur le théorème limite central pour des variables aléatoires dans espace de Banach, Publ. Inst. Statist. Univ. Paris, Vol. 3 (1954), pp. 143–148.
- [15] DYE, H. A., The Radon-Nikodym Theorem for Finite Rings of Operators, Trans. Amer. Math. Soc., 72 (1952), pp. 243–280.
- [16] FABIÁN, V., A Note on the Conditional Expectations, Czechoslovak Math. J., Vol. 4 (1954), pp. 187–191.
- [17] FELLER, W., An Introduction to Probability Theory and Its Applications, New York, 1950.
- [18] FELLER, W., Diffusion Processes in One Dimension, Trans. Amer. Math. Soc., Vol. 77 (1954), pp. 1–31.
- [19] FORTET, R., Calcul des probabilités, Paris, 1950.

- [20] FORTET, R., Random Distributions with an Application to Telephone Engineering, Proc. Third Berkeley Symposium on Math. Statistics and Probability, Vol. 2 (1956). To appear.
- [21] FORTET, R., and E. MOURIER, Résultats complémentaires sur les éléments aléatoires prenant leurs valeurs dans un espace de Banach, Bull. Sci. Math. (2), Vol. 78 (1954), pp. 14–30.
- [22] FORTET, R., and E. MOURIER, Les fonctions aléatoires comme éléments aléatoires dans les espace de Banach, Stud. Math., Vol. 15 (1955), pp. 62–79.
- [23] FRECHET, M., Recherches théoriques modernes sur le calcul des probabilités. II. Méthode des fonctions arbitraires. Théorie des événements en chaîne dans d'un nombre fini d'états possibles, Paris, 1938.
- [24] GEL'FAND, I. M., Generalized Random Processes, Doklady Akad. Nauk SSSR (N.S.), Vol. 100 (1955), pp. 853–856. [In Russian.]
- [25] GIHMAN, I. L., Some Limit Theorems for Conditional Distributions, Doklady Akad. Nauk SSSR (N.S.), Vol. 91 (1953), pp. 1003–1006. [In Russian.]
- [26] GNEDENKO, B. V., Course in the Theory of Probability, Moscow-Leningrad, 1950. [In Russian.]
- [27] GNEDENKO, B. V., and A. N. KOLMOGOROV, Limit Distributions for Sums of Independent Random Variables, Translated by K. L. Chung with an appendix by J. L. Doob, Cambridge, 1954.
- [28] HALMOS, P. R., Measure Theory, New York, 1950.
- [29] HARRIS, T. E., and H. ROBBINS, Ergodic Theory of Markov Chains Admitting an Infinite Invariant Measure, Proc. Nat. Acad. Sci. U.S.A., Vol. 39 (1953), pp. 860–864.
- [30] HILLE, E., Functional Analysis and Semi-Groups, New York, 1948.
- [31] HILLE, E., On the Integration Problem for Fokker-Planck's Equation in the Theory of Stochastic Processes, Onzième congrès des math. scand. (1949), pp. 185–194.
- [32] HILLE, E., On the Integration of Kolmogoroff's Differential Equations, Proc. Nat. Acad. Sci. U.S.A., Vol. 40 (1954), pp. 20–25.
- [33] HUNT, G. A., Semi-Groups of Measures on Lie Groups, Trans. Amer. Math. Soc., Vol. 81 (1956), pp. 264–293.
- [34] ITÔ, K., Theory of Probability, Tokyo, 1953.
- [35] ITÔ, K., Stationary Random Distributions, Mem. Coll. Sci. Univ. Kyoto, Ser. A. Math., Vol. 28 (1954), pp. 209–223.
- [36] JIŘINA, M., Conditional Probabilities on Strictly Separable σ -Algebras, Czechoslovak Math. J., Vol. 4 (1954), pp. 372–380. [In Czech.]
- [37] KALLIANPUR, G., On a Limit Theorem for Dependent Random Variables, Doklady Akad. Nauk SSSR (N.S.) Vol. 101 (1955), pp. 13–16. [In Russian.]
- [38] KATO, T., On the Semi-Groups Generated by Kolmogoroff's Differential Equations, J. Math. Soc. Japan, Vol. 6 (1954), pp. 1–15.
- [39] KAWADA, Y., The Theory of Probability, Tokyo, 1952.
- [40] KENDALL, D. G., and G. E. H. REUTER, Some Pathological Markov Processes with a Denumerable Infinity of States and the Associated Semigroups of Transformations in I, Proc. Symp. on Stochastic Processes (Amsterdam), 1954. To appear.
- [41] KHINTCHINE, A., Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Berlin, 1933. [Reprint, CHELSEA PUBLISHING COMPANY.]
- [42] KHINTCHINE, A., Limit Laws of Sums of Independent Random Variables, Moscow-Leningrad, 1938. [In Russian.]
- [43] KOLMOGOROV, A., Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933. [The present work is an English translation of this.]
- [44] KOLMOGOROV, A., Algèbres de Boole métriques complètes, VI Zjazd Mat. Pols., Warsaw (1948), pp. 21–30.

- [45] KOLMOGOROV, A., A Theorem on the Convergence of Conditional Mathematical Expectations and Some of Its Applications, *Comptes Rendus du Premier Congrès des Mathématiciens Hongrois* (1952), pp. 367–386. [In Russian and Hungarian.]
- [46] KOLMOGOROV, A., Some Work of Recent Years in the Field of Limit Theorems in the Theory of Probability, *Vestnik Moskov. Univ. Ser. Fiz.-Mat. Estest. Nauk*, Vol. 8 (1953), pp. 29–38.
- [47] KUNISAWA, K., *Limit Theorems in Probability Theory*, Tokyo, 1949.
- [48] LECAM, L., Convergence in Distribution of Stochastic Processes, *Univ. California Publ. Statistics*. To appear.
- [49] LEVY, P., *Théorie de l'addition des variables aléatoires*, Paris, 1937.
- [50] LEVY, P., *Processus stochastiques et mouvement Brownien*, Paris, 1948.
- [51] LEVY, P., Systèmes markoviens et stationnaires. Cas dénombrable, *Ann. Sci. Ecole Norm. Sup.*, Vol. 68 (1951), pp. 327–401.
- [52] LEVY, P., Complément à l'étude des processus de Markoff, *Ann. Sci. Ecole Norm. Sup.*, Vol. 69 (1952), pp. 203–212.
- [53] LEVY, P., Processus markoviens et stationnaires du cinquième type (infinité dénombrable d'états possibles, paramètre continu), *C. R. Acad. Sci. Paris*, Vol. 236 (1953), pp. 1630–1632.
- [54] LOÈVE, M., *Probability Theory*, New York, 1955.
- [55] LOÈVE, M., Variational Terms and the Central Limit Problem, *Proc. Third Berkeley Symposium on Math. Statistics and Probability*, Vol. 2 (1956). To appear.
- [56] LOS, J., On the Axiomatic Treatment of Probability, *Colloq Math.*, Vol. 3 (1955), pp. 125–137.
- [57] MARSAGLIA, G., Iterated Limits and the Central Limit Theorem for Dependent Variables, *Proc. Amer. Math. Soc.*, Vol. 5 (1954), pp. 987–991.
- [58] MOURIER, E., Eléments aléatoires dans un espace de Banach, *Ann. Inst. H. Poincaré*, Vol. 13 (1953), pp. 161–244.
- [59] MOURIER, E., L-Random Elements and L-Random Elements in Banach Spaces, *Proc. Berkeley Symposium on Math. Statistics and Probability*, Vol. 2 (1956). To appear.
- [60] MOY, S. T. C., Characterizations of Conditional Expectation as a Transformation on Function Spaces, *Pacific J. Math.*, Vol. 4 (1954), pp. 47–63.
- [61] MOY, S. T. C., Conditional Expectations of Banach Space Valued Random Variables and Their Properties, (Abstract) *Bull. Amer. Math. Soc.*, Vol. 62 (1956). To appear.
- [62] NAKAMURA, M. and T. TURUMARU, Expectation in an Operator Algebra, *Tôhoku Math. J.*, Vol. 6 (1954), pp. 182–188.
- [63] NEVEU, J., Etude des semi-groups de Markoff, (Thesis) Paris, 1955.
- [64] PARZEN, E., Convergence in Distribution and Fourier-Stieltjes Transforms of Random Functions, (Abstract) *Ann. Math. Statistics*, Vol. 26 (1955), p. 771.
- [65] PARZEN, E., A Central Limit Theorem for Multilinear Stochastic Processes, (Abstract) *Ann. Math. Statistics*, Vol. 27 (1956), p. 206.
- [66] PROHOROV, Yu. V., Probability Distributions in Functional Spaces, *Uspehi Matem. Nauk* (N.S.), Vol. 8 (1953), pp. 165–167. [In Russian.]
- [67] RENYI, A., *The Calculus of Probabilities*, Budapest, 1954. [In Hungarian.]
- [68] RENYI, A., On a New Axiomatic Theory of Probability, *Acta Math. Acad. Sci. Hung.*, Vol. 6 (1955), pp. 285–335.
- [69] ROBBINS, H., On the Equidistribution of Sums of Independent Random Variables, *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 786–799.
- [70] ROMANOVSKI, V. I., Discrete Markov Chains, Moscow-Leningrad, 1949. [In Russian.]

- [71] ROSENBLATT, M., A Central Limit Theorem and a Strong Mixing Condition, Proc. Nat. Acad. Sci. U.S.A., Vol. 42 (1956), pp. 43–47.
- [72] SARYMSAKOV, T. A., Elements of the Theory of Markov Processes, Moscow, 1954. [In Russian.]
- [73] SCHWARTZ, L., Théorie des distributions, Paris, 1950–51.
- [74] SEGAL, I. E., A Non-Commutative Extension of Abstract Integration, Ann. Math. Vol. 57 (1953), pp. 401–457.
- [75] SEGAL, I. E., Abstract Probability Spaces and a Theorem of Kolmogoroff, Amer. J. Math., Vol. 76 (1954), pp. 177–181.
- [76] SEGAL, I. E., A Mathematical Approach to Elementary Particles and Their Fields, University of Chicago, 1955. [Mimeographed Lecture Notes.]
- [77] TAKANO, K., On Some Limit Theorems of Probability Distributions, Ann. Inst. Statist. Math., Tokyo, Vol. 6 (1954), pp. 37–113.
- [78] TSURUMI, S., On the Strong Law of Large Numbers, Tôhoku Math. J., Vol. 7 (1955), pp. 166–170.
- [79] UMEGAKI, H., Conditional Expectation in an Operator Algebra, Tôhoku Math. J., Vol. 6 (1954), pp. 177–181.
- [80] YOSIDA, K., Operator Theoretical Treatment of the Markoff's Process, Proc. Imp. Acad. Japan, Vol. 14 (1938), pp. 363–367.

1. 原始德文版本发表于1933年。 [2](#)
2. 暂不确定此人是谁。 [2](#)
3. 原文脚注1, Zur Statistik der kontinuierlichen Systeme und des zeitlichen Verlaufes der physikalischen Vorgänge. 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.* [2](#)
4. 见参考文献 R. von Mises [1] and [2] and S. Bernstein [1]. [2](#)
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)。 [2](#)
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。 [2](#)
7. 只对概率论的纯数学发展感兴趣的读者可以忽略此节。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, 作者们在很大程度上使用了冯·米泽斯的工作, [1] 第21–27页。 [2](#)
8. 第四节公式 (3) [2](#)
9. 见 Fréchet [1] 和 [2]. [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, $\xi_{q_1 q_2 \dots q_n}$. 令每一个元素对应的基本概率为 $p_{q_1 q_2 \dots q_n}$, 则 $A_{q_n}^{(i)}$ 表示当 $q_i = q$ 时的所有 $\xi_{q_1 q_2 \dots q_n}$ 组成的集合。 [2](#)
11. 事实上在独立的情况下, 可以只选取 $r_1 + r_2 + \dots + r_n$ 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 $r_1 + r_2 + \dots + r_n - n$ 个自由度。 [2](#)
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 n mutually exclusive and exhaustive events or sets.) 假如公式 (2) 正确, 则有: [2](#)
13. 参考 S. N. Bernstein [1] pp. 47–57. 不过读者应该很容易采用数学归纳法自己证明这一推论。 [2](#)
14. $P(A_1 \bar{A}_2) = P(A_1) - P(A_1 A_2) = P(A_1) - P(A_1)P(A_2) = P(A_1)\{1 - P(A_2)\} = P(A_1)P(\bar{A}_2)$, 其他公式也是类似的推导方法。 [2](#)
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 [2](#)
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)。 [2](#)
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。 [2](#)
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。 [2](#)
19. See, for example, O. NIKODYM, Sur une généralisation des intégrales de M. J. Radon, Fund. Math. v. 15, 1930, p. 136. [2](#)[2](#)[2](#)[2](#)
20. 参考 HAUSDORFF, Mengenlehre, 1927, p. 85. [2](#)
21. CARATHÉODORY, Vorlesungen über reelle Funktionen, pp. 237–258. (New York, Chelsea Publishing Company). [2](#)
22. 此处原文, 区间表示方式, 中间为 \vdash 。 [2](#)
23. Cf., for example, LEBESGUE, Leçons sur l'intégration, 1928, p. 152–156. [2](#)
24. 参考前一个注释中的内容。 [2](#)
25. For a definition of Borel sets in R see HAUSDORFF, Mengenlehre, 1927, pp. 177–181. [2](#)
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. [2](#)

27. a_i 也可以取无限值 $\pm\infty$ 。 [↗](#)
28. 参见第 IV 章第 3 节。 [↗](#)
29. 参考HAUSDORFF, Mengenlehre, 1927, p. 23。 [↗](#)
30. 由以上内容可知, 波莱尔圆柱集是可以通过公式 (1) 中的关系定义的波莱尔集合。现假定集合 A 和 B 是两个由如下关系定义的波莱尔圆柱集: [↗](#)
31. 这一概念来源于 Bernoulli, 它的完整的、一般性的分析由 E. E. Slutsky 提出 (见参考文献 [1])。 [↗](#)
32. 事实上 $F_n(a)$ 至多只有可数个间断点 (见 LEBESGUE, Leçons sur l'intégration, 1928, p. 50)。因此连续点是处处稠密的 (everywhere dense), 并且函数 $F(a)$ 在间断点的值由它左连续点的函数值的极限决定。 [↗](#)
33. 和第三章第 5 节一样, 本章及之后的章节的研究内容都是在波莱尔概率域上开展的。 [↗](#)
34. FRÉCHET, Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, Bull. Soc. Math. France v. 43, 1915, p. 248. [↗](#)
35. 假设 y 是一个随机变量, 按照积分理论的术语来说, y 对于 \mathcal{E} 是可测的。 [↗](#)
36. 参考 V. GLIVENKO, Sur les valeurs probables de fonctions, Rend. Accad. Lincei v. 8, 1928, pp. 480-483. [↗](#)
37. 因此当 $x \neq 0$ 时, $f(x) > 0$ 。 [↗](#)
38. 参考A. KOLMOGOROV and M. LEONTOVICH, Zur Berechnung der mittleren Brownschen Fläche, Physik. Zeitschr. d. Sovietunion, v. 4, 1933. [↗](#)
39. O. NIKODYM, Sur une généralisation des intégrales de M. J. Radon, Fund. Math. v. 15, 1930 p. 168 (Theorem III). [↗](#)
40. 以下公式, 原文等式左边为 $P(Bu^{-1}(A_n))$, 我认为多了个 n 。 [↗](#)
41. Lebesgue, I. c., 1928, pp. 301-302. [↗](#)
42. 这种情况下我们仅考虑 λ 的可数序列的值 (a Countable sequence of values), 这样对于所有的这些 λ , 所有的概率 $P_u\{k\lambda \leq y < (k+1)\lambda\}$ 就都有定义了。 [↗](#)
43. Cf. footnote 3. [↗](#)
44. Cf. A. KOLMOGOROV, Bemerkungen zu meiner Arbeit, "Über die Summen zufälliger Größen." Math. Ann. v. 102, 1929, pp. 484-488, Theorem XII. [↗](#)
45. Ibid, Theorem XIII. (同上, 定理 XIII。) [↗](#)
46. Cf. A. KOLMOGOROV, Sur la loi forte des grandes nombres, C. R. Acad. Sci. Paris v. 191, 1930, pp. 910-911. [↗](#)
47. The proof of this statement has not yet been published. [↗](#)
48. Cf. A. KHINTCHINE and A. KOLMOGOROV, On the Convergence of Series, Rec. Math. Soc. Moscow, v. 32, 1925, p. 668-677. [↗](#)
49. 参考第四章第 5 节。在强大数定律条件下, 以下概率也是相同的结果 (只可能等于 0 或 1) $P\{s_n - d_n \rightarrow 0\}$ 。至少当随机变量 x_n 相互独立时是如此。 [↗](#)
50. 贝尔函数是指那些从多项式出发, 通过对函数序列逐次取极限而得到的函数。 [↗](#)