

概率理论基础

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¹。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²

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 ³；

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 $E = \{\xi_1, \xi_2, \dots, \xi_k\}$ ，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⁴

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 $P(A) = 0$ ⁵ 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. $A \dot{+} B \dot{+} \dots \dot{+} N = X$ 。	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。
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:

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

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

因为 $\bar{E} = 0$, 所以有:

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

如果事件 A, B, \dots, N 互不相容, 则由公理 V 可得:

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

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：

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

由公式(4)可得

$$P(X) = P(A_1 X) + P(A_2 X) + \dots + P(A_n X)$$

由公式(6)可得

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

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 Fréchet⁶ 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:

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

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⁷:

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

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⁸。

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⁹。在独立性的情况下, 有下方程:

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

(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 $2^n - n - 1$ ¹⁰:

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

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¹¹。

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¹²。(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¹³

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

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 $P_{A_{q_i}}(B)$ ，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¹⁴。

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¹⁵:

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

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

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

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 $1/4$)。

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

1. 见参考文献 R. von Mises [1] and [2] and S. Bernstein [1]. [↩](#)

2. 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)。 [↩](#)

3. 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。 [↩](#)

4. 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。 [↩](#)

5. 第四节公式 (3) [↩](#)

6. 见 Fréchet [1] 和 [2]. [↩](#)

7. 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。 [↩](#)

8. 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。 [↵](#)

9. 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： [↵](#)

10. 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。 [↵](#)

11. $P(A_1 \bar{A}_2) = P(A_1) - P(A_1 A_2) = P(A_1) - P(A_1)P(A_2) = P(A_1)\{1 - P(A_2)\} = P(A_1)P(\bar{A}_2)$ ，其他公式也是类似的推导方法。 [↵](#)

12. 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 [↵](#)

13. 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)。 [↵](#)

14. 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。 [↵](#)

15. 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。 [↵](#)