概率论的基本概念

李孟棠，朱彬

中山大学智能工程学院

limt29@mail.sysu.edu.cn, zhub26@mail.sysu.edu.cn

2021年3月2日

主要内容

主要内容

朴素集合论：集合是由一类物体组成的整体，这些物体称为“元素”。

子集$B \subset A$（或$A \supset B$）$\Longleftrightarrow$ 陈述“$\forall \zeta \in B \implies \zeta \in A$”。

概率论通常考察某个集合S的子集，其中S被称为**空间**。

集合元素通常用希腊字母ζ表示，如$A = \{\zeta_1, \ldots, \zeta_n\}$。

元素从属关系：$\zeta_i \in A$或$\zeta_i \notin A$。

空集(empty or null set)：$\emptyset$

如果一个集合由n个元素组成，那么它共有2^n个子集。

在概率论里，我们给S的子集（事件）分配概率，并在S上定义各种函数（随机变量）。因此，必须小心区分元素ζ和单元素集合(singleton)$\{\zeta\}$。

例：用f_i表示骰子的面，他们组成集合$S=\{f_1,\dots,f_6\}$。此时$n=6$，因此S有$2^6=64$个子集

$$\emptyset,\{f_1\},\dots,\{f_1,f_2\},\dots,\{f_1,f_2,f_3\},\dots,S.$$

一般而言，集合的元素可以是任何物体。例如上面的64个子集也可以组成一个新的集合。

例：掷一枚硬币两次，所得结果构成集合$S=\{hh,ht,th,tt\}$，它有$2^4=16$个子集，例如

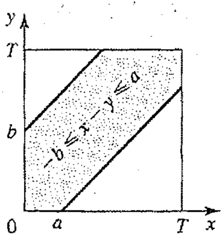
$$A=\{\text{第一次为正面}\}=\{hh,ht\}$$
$$B=\{\text{只有一次为正面}\}=\{ht,th\}$$
$$C=\{\text{至少有一次为正面}\}=\{hh,ht,th\}.$$

例：集合S由满足条件

$$0 \leq x \leq T, \quad 0 \leq y \leq T$$

的有序数对(x, y)组成，如图所示。阴影部分是S的子集

$$A = \{(x, y) \in S : -b \leq x - y \leq a\}.$$



集合运算

用维恩图(Venn diagram)表示集合S及其子集。

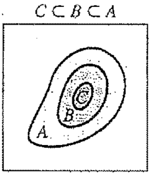


FIGURE 2-2

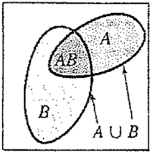


FIGURE 2-3

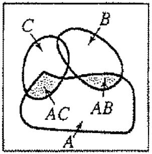


FIGURE 2-4

对于任意子集A，有$\emptyset \subset A \subset A \subset S$（方框）。

传递性(transitivity)：如果$C \subset B$且$B \subset A$，那么$C \subset A$。

集合相等：当且仅当$A \subset B$且$B \subset A$时，有$A = B$。

集合的并(union)记为$A \cup B$或$A + B$，该运算满足交换律(commutativity)和结合律(associativity)：

$$A \cup B = B \cup A, \quad (A \cup B) \cup C = A \cup (B \cup C).$$

注意到，若$B \subset A$，则$A \cup B = A$。由此可得

$$A \cup A = A, \quad A \cup \emptyset = A, \quad S \cup A = S.$$

集合的交(intersection)记为AB或$A \cap B$，该运算满足交换律、结合律和对并运算的分配律(distributivity, 注意运算优先级)：

$$AB = BA, \quad (AB)C = A(BC), \quad A(B \cup C) = AB \cup AC.$$

注意到，若$A \subset B$，则$AB = A$。由此可得

$$AA = A, \quad A\emptyset = \emptyset, \quad SA = A.$$

互斥集合：如果两个集合A和B没有公共元素，即$AB=\emptyset$，则称这两个集合互斥(mutually exclusive)或不相交(disjoint)。

如果多个集合$A_1, A_2, \ldots$满足条件：

$$\text{对每个}i\text{和}j \neq i\text{，有}A_iA_j=\emptyset,$$

则称这些集合互斥。

划分(partition)：集合S的划分$\boldsymbol{U}$是由S的互斥子集组成的集合，这些子集的并等于S。也就是说，如果

$$A_1 \cup \cdots \cup A_n = S\text{，且对于任意}i \neq j\text{有}A_iA_j=\emptyset,$$

那么称$\boldsymbol{U}=\{A_1, \ldots, A_n\}$为$S$的一个划分。

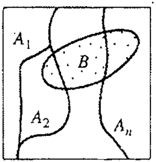


FIGURE 2-5

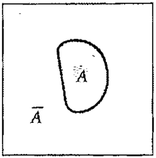


FIGURE 2-6

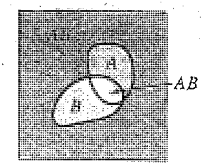


FIGURE 2-7

补集(complement)：集合A的补集$\overline{A}$由S中所有不在A中的元素组成。此定义包含如下结论：

$$A \cup \overline{A} = S, \quad A\overline{A} = \emptyset, \quad \overline{\overline{A}} = A, \quad \overline{S} = \emptyset, \quad \overline{\emptyset} = S.$$

$$B \subset A \implies \overline{B} \supset \overline{A}; \quad A = B \implies \overline{A} = \overline{B}.$$

de Morgan定律：图2-7表明

$$\overline{A \cup B} = \overline{A}\,\overline{B}, \quad \overline{AB} = \overline{A} \cup \overline{B}.$$

反复运用de Morgan定律，可得出下列结论：在集合等式中，如果我们把所有集合用它们的补集代替，所有的并用交代替，所有的交用并代替，那么等式仍然成立。

例：由集合交运算对并运算的分配律：$A(B \cup C) = AB \cup AC$。
等式左边取补：

$$\overline{A(B \cup C)} = \overline{A} \cup \overline{B \cup C} = \overline{A} \cup \overline{B}\ \overline{C}.$$

等式右边取补：

$$\overline{AB \cup AC} = (\overline{AB})(\overline{AC}) = (\overline{A} \cup \overline{B})(\overline{A} \cup \overline{C}).$$

因此，有

$$\overline{A} \cup \overline{B}\ \overline{C} = (\overline{A} \cup \overline{B})(\overline{A} \cup \overline{C}). \tag{1}$$

对偶原理
注意到，如果去掉(1)式中所有上横线，那么等式仍然成立，即

$$A \cup BC = (A \cup B)(A \cup C).$$

实际上，这是集合并关于交运算的分配律。

对偶原理可以被看做de Morgan定律的一个变体(variant)，表述为：在集合等式中，如果我们把所有的并用交代替，所有的交用并代替，所有集合S和$\emptyset$分别用$\emptyset$和S代替，那么等式仍然成立。

例如$S \cup A = S \implies \emptyset A = \emptyset$。

主要内容

概率论中的集合术语:

- 样本空间[1](sample space)S或Ω称为必然事件(certain event)。
- S的元素称为实验结果(experimental outcomes)。
- S的子集称为事件(events)。
- 空集$\emptyset$是不可能事件(impossible event)。
- 单元素集$\{\zeta_i\}$称为基本事件(elementary event)。

事件通常用斜体大写字母A, B等表示。

[1]和线性空间的概念没有联系。

在把概率论应用于物理问题时，对于实验结果的认定不总是唯一的。下面的例子展示了这种模糊性。

例：X, Y, Z三位选手对骰子实验分别有如下解读：

- X说这个实验的结果是骰子的六个面，它们组成空间$S = \{f_1, \ldots, f_6\}$。S有64个子集，其中事件{偶数}包含f_2, f_4, f_6三个结果。
- Y只想对结果的**奇偶性**下赌注，因此他认为骰子实验只有奇数和偶数两个结果，它们构成样本空间$S = \{$奇数，偶数$\}$。这里的S只有4个子集，事件{偶数}仅包含一个结果。
- Z打赌实验结果为**一点**且骰子会落在桌子的左侧。他认为骰子实验有无穷多种结果，其中每种结果都由骰子中心的坐标和六个面确定，事件{偶数}不再由一个或三个而是由无穷多个结果组成。

在今后课程中，每当提到实验，我们总假设实验结果已经被明确指定。例如骰子实验的X解读版。

相对频率解读相关术语

- 进行单独一次实验称为一次试验(trial)。
- 每次试验的结果为单个ζ_i。
- 如果事件A包含元素ζ_i，则称它在这次试验中**发生**。
- 必然事件在每次试验中都发生，不可能事件从不发生。
- 当事件A或B或两者都发生时，$A \cup B$发生。
- 当事件A和B都发生时，AB发生。
- 如果事件A和B互斥且A发生，那么B不发生。
- 如果$A \subset B$且A发生，那么B发生。
- 在每次试验中，要么A发生，要么$\overline{A}$发生。

例：如果单次骰子实验的结果为f_5，那么事件$\{f_5\}$，$\{$奇数$\}$，和30个其他事件发生。

我们给每个事件A分配一个数$P(A)$，并称之为事件A的**概率**。这个数的选择需要满足如下三个条件：

1. $P(A) \geq 0$；
2. $P(S) = 1$；
3. 如果$AB = \emptyset$，则$P(A \cup B) = P(A) + P(B)$。

在概率论的发展中，所有结果都直接或间接地来自于并仅依赖于这三条公理。

实际上，概率是一个集合函数(set function)

$$P : \boldsymbol{F} \to [0, 1] \ ,$$

其中事件的集合$\boldsymbol{F} = \{A \subset S\} \subset 2^S$（幂集，power set）需满足一些额外条件。

概率的简单性质:

- 不可能事件的概率为0，即$P(\emptyset)=0$。
  证：因为$A\emptyset=\emptyset$，$A\cup\emptyset=A$，所以
  $$P(A)=P(A\cup\emptyset)=P(A)+P(\emptyset).$$
- 对于任意事件A，有$P(A)=1-P(\overline{A})\leq 1$。
  证：因为$A\cup\overline{A}=S$，$A\overline{A}=\emptyset$，所以
  $$1=P(S)=P(A+\overline{A})=P(A)+P(\overline{A}).$$
- 对于任意事件A和B，有
  $$P(A\cup B)=P(A)+P(B)-P(AB)\leq P(A)+P(B).$$
  证：把事件$A\cup B$和B分别写作两个互斥事件的并：
  $$A\cup B=A\cup\overline{A}B,\quad B=AB\cup\overline{A}B$$

概率的简单性质（续）：

- 如果事件$B \subset A$，那么

$$P(A) = P(B) + P(A\overline{B}) \geq P(B).$$

证：因为$A = B \cup A\overline{B}$，$B(A\overline{B}) = \emptyset$，故上式成立。

频率解读：通过选择概率的上述三条公理，我们后续得到的理论能够圆满地描述物理世界。因此，实际问题中的概率必须满足这些公理。对于概率的频率解读

$$P(A) \approx \frac{n_A}{n},$$

这一点容易验证。

1. 显然$P(A) \geq 0$;
2. 因为S每次都发生，所以$n_S = n$，故$P(S) = 1$;
3. 如果$AB = \emptyset$，那么$n_{A \cup B} = n_A + n_B$（由于互斥性），因此

$$P(A \cup B) \approx \frac{n_{A\cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} \approx P(A) + P(B).$$

事件相等：如果事件A和B由相同的元素组成，那么称它们相等（集合相等）。如果集合

$$(A\cup B)(\overline{AB}) = A\overline{B}\cup\overline{A}B$$

的概率为零，那么称事件A和B**以概率1相等**(equal with probability one)。注意到上面的集合由包含在事件A或B中但不在AB中的结果组成。

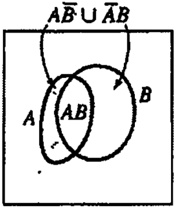


FIGURE 2-8

从定义可知（练习题），事件A和B以概率1相等当且仅当

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

如果$P(A) = P(B)$，则称事件A和B依概率相等(equal in probability)。仅给定这些信息，我们无法得出关于$P(AB)$的任何结论。实际上，事件A和B也有可能互斥。

从上一页的结论可知，如果事件N和不可能事件以概率1相等，那么$P(N) = 0$。这当然不能说明$N = \emptyset$。

事件的F类(class)

事件是被分配概率的S的子集。在很多情况下，我们通常不把S的所有子集看作事件，而只考虑一些子集构成的$\boldsymbol{F}$类。原因有两个

应用的原因：例如骰子实验的Y解读版。

数学的原因（主要）：当样本空间S包含无穷多种结果时，通常无法做到把概率分配给所有子集且满足三条公理（包括第三条的广义形式）。

事件的$\boldsymbol{F}$类不再由S的任意子集构成。我们假设：如果A和B是事件，那么$A \cup B$和AB也是事件。这带来如下定义：

一个代数(algebra)或域(field)$\boldsymbol{F}$是S子集构成的非空类，满足：

- 如果$A \in \boldsymbol{F}$，那么$\overline{A} \in \boldsymbol{F}$；
- 如果$A \in \boldsymbol{F}$且$B \in \boldsymbol{F}$，那么$A \cup B \in \boldsymbol{F}$。

上述两个性质给出了让$\boldsymbol{F}$成为域的最少条件。其他性质可以从定义导出：

$$\text{如果}A \in \boldsymbol{F}\text{且}B \in \boldsymbol{F}\text{，那么}AB \in \boldsymbol{F}.$$

证：从性质一可知$\overline{A} \in \boldsymbol{F}$且$\overline{B} \in \boldsymbol{F}$。然后依次利用性质二和性质一，得到

$$\overline{A} \cup \overline{B} \in \boldsymbol{F}, \quad \overline{\overline{A} \cup \overline{B}} = AB \in \boldsymbol{F}.$$

一个域包含必然事件和不可能事件，即$S \in \boldsymbol{F}$，$\emptyset \in \boldsymbol{F}$。
证：由于$\boldsymbol{F}$非空，因此$\boldsymbol{F}$至少包含一个元素A；由性质一知$\overline{A} \in \boldsymbol{F}$。由此，

$$A \cup \overline{A} = S \in \boldsymbol{F}, \quad A\overline{A} = \emptyset \in \boldsymbol{F}.$$

上述结果表明，如果一个集合可以被写作**有限个**$\boldsymbol{F}$中集合的并或交，那么它也在$\boldsymbol{F}$中。但如果考虑无穷多个集合的并或交，那么结论未必成立。

例：考虑自然数集合$\mathbb{N}$的子集A构成的一个代数，其中要么A是有限集，要么$\mathbb{N}\backslash A$是有限集。

σ-**代数**，又称σ-域[2]：假定$A_1, \ldots, A_n, \ldots$是$\boldsymbol{F}$中集合构成的无穷序列，如果它们的并和交也属于$\boldsymbol{F}$，那么称$\boldsymbol{F}$为σ-代数。

例如S的幂集是一个σ-代数。

设$\boldsymbol{C}$是S的一个子集类但不构成代数。往$\boldsymbol{C}$中加入其它S的子集，我们可以构造出以$\boldsymbol{C}$为子集的代数。另外可以证明：存在一个包含$\boldsymbol{C}$所有子集的最小σ-代数。

> 例：设S由四个元素a, b, c, d组成，而$\boldsymbol{C}$由集合$\{a\}$和$\{b\}$组成。将$\{a\}$和$\{b\}$的补以及它们的并和交加入$\boldsymbol{C}$，我们得到包含$\boldsymbol{C}$的最小代数，它有下列集合组成：
>
> $$\emptyset, \{a\}, \{b\}, \{a, b\}, \{c, d\}, \{b, c, d\}, \{a, c, d\}, S.$$

[2] Papoulis书称Borel域，但该名词现在有额外的含义。

事件：在概率论中，事件是S的某些子集，它们构成一个σ-代数。如此我们不仅可以为事件的有限并和交分配概率，而且能考察极限情形。

首先拓展公理3（有限可加性，finite additivity）：如果事件$A_1, \ldots, A_n$互斥，那么

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n).$$

但是，仅基于公理3，上式并不能被推广到极限情形。为此，我们引入一个额外条件，称为**可数可加性(countable or σ- additivity)公理**：

公理3a. 如果事件$A_1, \ldots, A_n, \ldots$互斥，那么

$$P(A_1 \cup \cdots \cup A_n \cup \cdots) = P(A_1) + \cdots + P(A_n) + \cdots .$$

从此，我们假设所有概率都满足1、2、3和3a，这些是公理化概率论的标准假设。

概率空间

至此，我们终于可以给出概率空间的完整定义：

定义

一个概率空间$(S, \boldsymbol{F}, P)$由三个要素组成，分别是：

- 样本空间S，一个任意的非空集合；
- σ-代数$\boldsymbol{F} \subset 2^S$，由$S$的子集（称为事件）组成的集合，满足：
  - $\boldsymbol{F}$包含样本空间：$S \in \boldsymbol{F}$,
  - $\boldsymbol{F}$关于集合补运算封闭：如果$A \in \boldsymbol{F}$，则$\overline{A} \in \boldsymbol{F}$，
  - $\boldsymbol{F}$关于可数并运算封闭：如果对于$i = 1, 2, \ldots$，有$A_i \in \boldsymbol{F}$，那么$(\bigcup_{i=1}^{\infty} A_i) \in \boldsymbol{F}$，
    - 由de Morgan定律可知，前两条性质的推论是$\boldsymbol{F}$关于可数交运算封闭：如果对于$i = 1, 2, \ldots$，有$A_i \in \boldsymbol{F}$，那么$(\bigcap_{i=1}^{\infty} A_i) \in \boldsymbol{F}$；
- 概率测度$P : \boldsymbol{F} \to [0, 1]$，定义在$\boldsymbol{F}$上的函数，满足：
  - 可数可加性：如果$\{A_i\}_{i=1}^{\infty} \subset \boldsymbol{F}$是可数个互斥集合，那么$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$，
  - 整个样本空间的测度等于1：$P(S) = 1$。

简而言之，概率空间是一个测度空间，且满足全空间的测度为1。

实验的公理化定义

在概率论中，一个实验由以下概念所规定:

1. 所有实验结果组成的集合S，
2. S的所有事件构成的σ-代数，
3. 这些事件的概率。

字母S不仅用来表示必然事件，还用来表示整个实验。

下面讨论实验中确定概率的方法，分为有限结果和无穷多结果情形。
可数空间 如果空间S由N个结果组成且N是一个有限数，那么所有事件的概率可以用基本事件$\{\zeta_i\}$的概率

$$P(\{\zeta_i\}) = p_i$$

来表示。由公理可知，数字p_i满足条件

$$p_i \geq 0, \quad p_1 + \cdots + p_N = 1.$$

可数空间（续）假定事件A由r个元素ζ_{k_i}组成，此时A可以写成基本事件$\{\zeta_{k_i}\}$的并。因此，

$$P(A) = P(\{\zeta_{k_1}\}) + P(\{\zeta_{k_2}\}) + \cdots + P(\{\zeta_{k_r}\}) = p_{k_1} + p_{k_2} + \cdots + p_{k_r}.$$

上述结论在S包含可数个元素$\zeta_1, \zeta_2, \ldots$时仍然成立（公理3a）。

古典定义：如果S由N个结果组成，且各基本事件的概率p_i都相等，则

$$p_i = \frac{1}{N}.$$

此时，由r个元素组成的事件A的概率

$$P(A) = \frac{r}{N}. \tag{2}$$

这个重要的特殊情形等价于概率的古典定义，但两者之间存在一个重要区别：在古典定义中，(2)式是逻辑必然性的结果；而在概率的公理化体系中，(2)式的前提仅仅是一个假设。

例：(a)在硬币实验中，空间S由结果h和t组成

$$S = \{h, t\}.$$

事件是$\emptyset, \{t\}, \{h\}, S$四个集合。如果$P(\{h\}) = p$，$P(\{t\}) = q$，那么$p + q = 1$。

(b)现在考虑掷三次硬币的实验，所有可能结果包括

$$hhh, hht, hth, htt, thh, tht, tth, ttt.$$

假设所有基本事件的概率都相等（均匀硬币）。此时每个基本事件的概率等于$1/8$。因此，得到三个正面的概率$P(\{hhh\}) = 1/8$。事件

$$\{\text{前两次为正面}\} = \{hhf, hht\}$$

的概率等于$2/8$。

实直线 如果S由不可数无穷多个元素组成，那么事件的概率一般不能用基本事件的概率来确定。例如在大多数应用中，S是n维欧式空间$\mathbb{R}^n$。

下面以实直线为例来讨论概率的确定。假定$S = \mathbb{R}$，那么可以证明：不可能为S的所有子集分配概率并使其满足所有公理（特别是可数可加性）。

为在实直线上构造一个概率空间，我们把所有区间$x_1 \le x \le x_2$和它们的可数并和交作为事件。这些事件构成了一个代数$\boldsymbol{F}$，实际上：

如此构造的$\boldsymbol{F}$是包含所有半直线$x \le x_i$的**最小**σ-代数，其中x_i为任意数.

此外，$\boldsymbol{F}$包含了所有开区间、闭区间、单点集$\{x_i\}$和所有在实际问题中“有用”的集合。

问：为什么不能把$\boldsymbol{F}$取做幂集$2^{\mathbb{R}}$？

问：为什么不能把$\boldsymbol{F}$取做幂集$2^{\mathbb{R}}$？
答：可以证明，存在某些实直线的子集，它们不能被表示为可数个区间的并和交(Vitali set)。不过多数应用不关心这些集合。

为完成对概率空间$(S, \boldsymbol{F}, P)$的指定，通常只需要对事件$\{x \leq x_i\}$分配概率即可。其他事件的概率可以通过公理来确定。

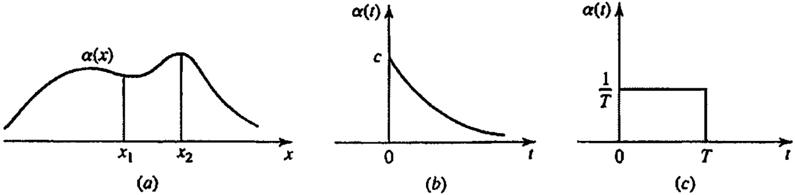


FIGURE 2-9

例（概率密度函数，probability density function, pdf）：如图所示，假定函数$\alpha(x)$满足

$$\int_{-\infty}^{\infty} \alpha(x)dx = 1, \quad \alpha(x) \geq 0.$$

例（续）：我们用积分来定义事件$\{x \le x_i\}$的概率

$$P(\{x \le x_i\}) = \int_{-\infty}^{x_i} \alpha(x)dx.$$

这规定了S中所有事件的概率，例如事件$\{x_1 < x \le x_2\}$的概率

$$P(\{x_1 < x \le x_2\}) = \int_{x_1}^{x_2} \alpha(x)dx. \tag{3}$$

原因：显然事件$\{x \le x_1\}$和$\{x_1 < x \le x_2\}$互斥且它们并为$\{x \le x_2\}$，所以

$$P(\{x \le x_1\}) + P(\{x_1 < x \le x_2\}) = P(\{x \le x_2\}).$$

对积分区间进行操作即可得到(3)式。

注意，如果函数$\alpha(x)$有界，则当$x_1 \to x_2$时，(3)中的积分趋向于0。由此得出结论：对于任意x_2，基本事件$\{x_2\}$的概率为0。尽管如此，所有基本事件的并的概率等于1。这一点和可数可加性公理并不矛盾，因为S中元素的总数不可数。

例：在$t=0$时选中一块放射性物质，观察其辐射粒子的时刻$t\geq 0$。此过程定义了一个实验，其结果包含正实轴上的所有点。把S取做整个时间轴，并令负半轴事件的概率为0，则该实验可以被看作实直线上实验的特例。

类似前一个例子，设这里的函数$\alpha(t)$具有如下形式：

$$\alpha(t)=ce^{-ct}U(t)\text{，其中}U(t)=\begin{cases}1 & \text{如果}t\geq 0\\ 0 & \text{如果}t<0\end{cases}.$$

代入(3)后计算得到该放射性物质在时间区间$(0,t_0]$内辐射出粒子的概率等于

$$c\int_0^{t_0}e^{-ct}dt=1-e^{-ct_0}.$$

例：一次电话呼叫在时间区间$[0, T]$内**随机**发生。这意味着呼叫在区间$0 \le t \le t_0$内发生的概率等于t_0/T。所以这个实验的结果是区间$[0, T]$上的所有点，且事件$\{$呼叫将在区间$[t_1, t_2]$内发生$\}$的概率

$$P(t_1 \le t \le t_2) = \frac{t_2 - t_1}{T}.$$

这还是式(3)的一个特殊情形，此时

$$\alpha(t) = \begin{cases} 1/T & \text{如果} 0 \le t \le T \\ 0 & \text{其它} t \end{cases}.$$

概率质量(probability masses) 事件A的概率$P(A)$可以被解读为它在维恩图中形状的质量。各种等式都可以被如此解读，例如

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

下面的两个例子将展示系统地计算复杂事件概率的方法：把它们写成简单互斥事件的并。

例：一个盒子里装有m个白球和n个黑球，每次随机抓出一个且不再放回。求抓k次得到白球的概率。
解：令W_k表示事件

$$W_k = \{\text{抓}k\text{次得到白球}\}.$$

该事件能按以下互斥的方式发生：第一次就抓到白球；第一次抓到黑球，第二次抓到白球；前两次抓到黑球，第三次抓到白球；如此等等。定义事件

$$X_i = \{\text{抓到}i\text{个黑球后抓到一个白球}\} \quad i = 0, 1, 2, \ldots, n.$$

那么有$W_k = X_0 \cup X_1 \cup \cdots \cup X_{k-1}$。概率的有限可加性表明

$$P(W_k) = \sum_{i=0}^{k-1} P(X_i).$$

例（续）：现有

$$
\begin{aligned}
P(X_0) &= \frac{m}{m+n} \\
P(X_1) &= \frac{n}{m+n}\cdot\frac{m}{m+n-1} \\
&\vdots \\
P(X_{k-1}) &= \frac{n(n-1)\cdots(n-k+2)m}{(m+n)(m+n-1)\cdots(m+n-k+1)}.
\end{aligned}
$$

因此

$$
\begin{aligned}
P(W_k) = \frac{m}{m+n}\Bigg(1 &+ \frac{n}{m+n-1} + \frac{n(n-1)}{(m+n-1)(m+n-2)} + \cdots \\
&+ \frac{n(n-1)\cdots(n-k+2)}{(m+n-1)(m+n-2)\cdots(m+n-k+1)}\Bigg)
\end{aligned}
$$

摸到第$n+1$次，我们必得到白球，所以$P(W_{n+1})=1$。上面的结果表明

$$
\begin{aligned}
&1 + \frac{n}{m+n-1} + \frac{n(n-1)}{(m+n-1)(m+n-2)} + \cdots \\
&+ \frac{n(n-1)\cdots 2\cdot 1}{(m+n-1)(m+n-2)\cdots(m+1)m} = \frac{m+n}{n}.
\end{aligned} \tag{4}
$$

例：两位玩家依次从装有m个白球和n个黑球的盒子里抓球，每次抓一个。假定先抓到白球的玩家获胜。问：先手玩家获胜的概率是多少？

解：不妨设A为先手。A能以下面这些方式获胜：A第一次就抓到白球；A和B各抓一个黑球，然后A抓到一个白球；A和B前两次都抓到黑球，然后A抓到白球；如此等等。设

$$X_k = \{A\text{和}B\text{前}k\text{次都抓到黑球，然后}A\text{抓到白球}\} \quad k = 0, 1, 2, \ldots$$

这些X_k是互斥事件，并且

$$\{A\text{获胜}\} = X_0 \cup X_1 \cup X_2 \cup \cdots.$$

因此

$$\begin{aligned} P_A := P(\{A\text{获胜}\}) &= P(X_0 \cup X_1 \cup X_2 \cup \cdots) \\ &= P(X_0) + P(X_1) + P(X_2) + \cdots \end{aligned}$$

其中

$$\begin{aligned} P(X_0) &= \frac{m}{m+n} \\ P(X_1) &= \frac{n}{n+m} \cdot \frac{n-1}{m+n-1} \cdot \frac{m}{m+n-2} \\ P(X_2) &= \frac{n(n-1)(n-2)(n-3)m}{(m+n)(m+n-1)(m+n-2)(m+n-3)(m+n-4)} \\ &\vdots \end{aligned}$$

例（续）：因此，

$$
\begin{aligned}
P_A = & \frac{m}{m+n}\left(1+\frac{n(n-1)}{(m+n-1)(m+n-2)}\right. \\
& \left.+\frac{n(n-1)(n-2)(n-3)}{(m+n-1)(m+n-2)(m+n-3)(m+n-4)}+\cdots\right)
\end{aligned}
$$

上面的和式只包含有限项，当其中一项为零时，求和即终止。类似可得

$$
\begin{aligned}
Q_B & := P(\{B\text{获胜}\}) \\
& = \frac{m}{m+n}\left(\frac{n}{m+n-1}+\frac{n(n-1)(n-2)}{(m+n-1)(m+n-2)(m+n-3)}+\cdots\right).
\end{aligned}
$$

但是必然有一位玩家获胜，所以

$$P_A + Q_B = 1.$$

可以验证，上式和(4)相同，原因在于两个例题密切相关。

测度的用处？**积分理论。**

抛开概率测度，我们讨论实直线上的勒贝格(Lebesgue)测度μ。

μ赋予$\mathbb{R}$子集“长度”的意义，希望它满足性质：

- $\mu\big((a,b)\big)=b-a.$
- 平移不变性(translation invariance)：对于任意的子集$A\subset\mathbb{R}$和任意点$x\in\mathbb{R}$，有$\mu(x+A)=\mu(A)$,其中集合

$$x+A:=\{x+y:y\in A\}.$$

- σ-可加性：如果集合$A_1,\ldots,A_n,\ldots$互斥，那么

$$\mu(A_1\cup\cdots\cup A_n\cup\cdots)=\mu(A_1)+\cdots+\mu(A_n)+\cdots.$$

可以证明：如果要求μ定义在$\mathbb{R}$的所有子集上，那么上述三条性质无法同时满足。

Banach-Tarski悖论（定理）：给定一个三维空间中的固体球，那么可以把它分解为有限个不相交子集，然后再把这些子集以另一种方式重组得到两个与原球完全相同的球。

实际上，重组过程只涉及碎片的平移、旋转等全等变换（不改变形状）；最少只需要5片即可完成重组。

定理的加强版本——"豌豆和太阳悖论(pea and the Sun paradox)"：一粒豌豆可以被切开然后重组成太阳。

为什么称为悖论？因为该定理的结论和基本的几何常识相矛盾。

主要内容

1. 概率的公理
   - 集合论
   - 概率空间
   - 条件概率

定义：假设事件M发生，在此情况下事件A的**条件概率**$P(A|M)$定义为

$$P(A|M)=\frac{P(AM)}{P(M)},$$

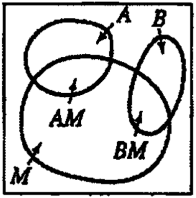
其中我们假设$P(M)$非零。

从定义容易导出如下性质：

如果$M\in A$，那么$P(A|M)=1$，

原因是$AM=M$。类似地，

如果$A\in M$，那么$P(A|M)=\frac{P(A)}{P(M)}\geq P(A)$.



频率解读：使用n_A, n_M和n_{AM}分别表示事件A, M和AM发生的次数，那么有

$$P(A) = \frac{n_A}{n}, \quad P(M) = \frac{n_M}{n}, \quad P(AM) = \frac{n_{AM}}{n}.$$

因此，

$$P(A|M) = \frac{P(AM)}{P(M)} \approx \frac{n_{AM}/n}{n_M/n} = \frac{n_A M}{n_M}.$$

以上结果可以表达为：如果我们去掉事件M不发生的所有试验，那么M发生的试验构成所有试验的一个子序列，然后$P(A|M)$等于事件A在该子序列中发生的相对频率n_{AM}/n_M。

基本性质：对于给定的M，条件概率满足概率的公理。

1. 第一条公理显然被满足：因为$P(AM) \geq 0$且$P(M) > 0$，所以

$$P(A|M) \geq 0.$$

2. 第二条公理成立：因为$M \subset S$，所以

$$P(S|M) = 1.$$

3. 为了证明第三条公理，观察到：如果事件A和B互斥，那么事件AM和BM也互斥。因此

$$P(A \cup B|M) = \frac{P\left((A \cup B)M\right)}{P(M)} = \frac{P(AM) + P(BM)}{P(M)}$$
$$= P(A|M) + P(B|M).$$

从这些结果可知，关于概率的所有结论也适用于条件概率。这一点的重要性见于后面的贝叶斯定理(Bayes' theorem)。

例：在公平骰子实验中，假设事件{偶数}已经发生，确定事件f_2发生的条件概率。其中，

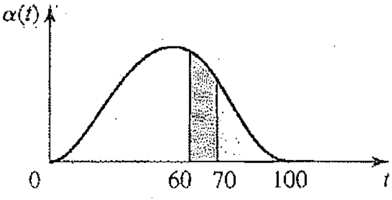
$$A = \{f_2\}, \quad M = \{\text{偶数}\} = \{f_2, f_4, f_6\},$$

因此$P(A) = 1/6$，$P(M) = 3/6$。又因为$AM = A$，由条件概率定义知

$$P(f_2|\{\text{偶数}\}) = \frac{P(\{f_2\})}{P(\{\text{偶数}\})} = \frac{1}{3}.$$

上面的数字等于以偶数为结果的试验子序列中事件{两点}发生的相对频率。

例：



The End