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

1. 概率的含义
   - 概率论的实际应用
   - 概率的不同定义
   - 概率与归纳

主要内容

1. 概率的含义
   - 概率论的实际应用
   - 概率的不同定义
   - 概率与归纳

概率论(Probability Theory)研究相继发生或同时发生的大量现象的平均特性，例如质量控制、机会游戏(赌博)、统计力学、湍流、噪声、出生和死亡率等等。

现象：当观测次数增加时，一些量的平均值趋于一个常数，例如抛硬币实验(experiment)。

概率论的目的就是用事件的概率来描述和预测这些平均值。

事件A的概率$P(A)$是分配给该事件的一个数字，其频率解读(frequency interpretation)如下：

如果在n次重复试验(trial)中事件A发生n_A次，那么当n足够大时，A发生的相对频率n_A/n以高度的确定性接近$P(A)$：

$$P(A) \approx n_A/n.$$

频率解读并不精确！因为“以高度的确定性”、“接近”、“足够大”的含义并不明确。

但是，这种不精确性（近似性）在描述物理现象的科学中普遍存在。

尽管如此，概率论本身建立在明确定义的公理之上，是按照逻辑规则演绎出来的一门严密学科，能够有效地处理实际问题。

用概率论解决实际问题时，必须明确区分三个步骤：观察(observation)、推理(deduction)、预测(prediction)。

第一步（观察）：通过一个不精确的过程来确定某些事件A_i的概率$P(A_i)$。例如使用相对频率n_{A_i}/n或者对称性（偏心骰子vs均匀骰子）。

第二步（推理）：假定概率满足某些公理，通过演绎推理从某些事件A_i的概率$P(A_i)$来确定其他事件B_j的概率$P(B_j)$。例如使用均匀骰子掷出偶数的概率为3/6。

第三步（预测）：基于得到的数字$P(B_j)$进行预测。例如反向使用频率解读：重复试验n次，事件B发生的次数$n_B \approx nP(B)$。

必须明确区分由实验确定的数据和由逻辑推理得到的结果。

第一步和第三步基于归纳推理(inductive reasoning)，例：为了确定掷一枚硬币得到正面的概率，掷100次还是1000次？

第二步基于演绎推理，是本门课程主要考虑的步骤。

主要内容

1. 概率的含义
   - 概率论的实际应用
   - 概率的不同定义
   - 概率与归纳

公理化定义

使用（朴素）集合论中的一些概念：

- 必然事件S，每次试验都会发生。
- 事件A和B的并$A \cup B \equiv A + B$表示A和B之一发生或两者都发生。
- 事件A和B的交$A \cap B \equiv AB$表示A和B都发生。
- 如果A和B中一个事件的发生阻止了另一个事件的发生，那么称事件A和B**互斥**。

例：骰子实验。必然事件？**偶数和小于3的并和交**？互斥事件？

概率的公理化方法仅依赖于下面三条假定(postulate)：

1. 任一事件A的概率$P(A)$是分配给该事件的一个非负实数，即$P(A) \geq 0$。
2. 必然事件的概率等于1，即$P(S) = 1$。
3. 如果两个事件A和B互斥，则$P(A \cup B) = P(A) + P(B)$。

本课程的出发点——概率的公理化定义强调了概率论的演绎特性，避免了概念模糊，为进阶课程的学习和复杂的实际应用提供了坚实的基础。

数学上严密的概率论以实分析、测度论等进阶数学课程作为先导。

本课程仅以微积分为基础来介绍概率论，试图最大程度地保持理论的严密性。

相对频率定义

事件A的概率$P(A)$被定义为如下极限(von Mises)

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n},$$

其中，n是试验次数，n_A是A发生的次数。

相对频率定义看似合理，也是概率应用的基础（第一步和第三步），但把它作为演绎理论（第二步）的基础是有问题的。原因：

- 实际中，n_A和n都是有限数字，它们的比值无法和极限划等号。
- 若采用频率定义，那么极限$P(A)$只能作为**假说**(hypothesis)，而不能通过实验确定。

一般认为Kolmogorov的公理化方法更为优越。

古典定义

事件A的概率$P(A)$可以不通过实验而先验(a priori)确定：

$$P(A) = \frac{N_A}{N},$$

其中N是可能结果的总数，N_A是支持事件A的结果数。

古典定义曾被用作概率论的基础长达几个世纪。今天，它被用来确定概率数据和作为工作假设(working hypothesis)。例：均匀骰子掷出偶数。

但必须注意到：数字N和N_A的含义有时不明确。

例：投掷两颗骰子，求结果点数之和等于7的概率p.
解：确定N和N_A.

1. 考虑所有可能的点数之和$2, 3, \ldots, 12$，共11种。$p = 1/11$，错！
2. 考虑所有可能的点数对，而不区分两颗骰子。共21种结果，其中支持事件A的有$(1, 6), (2, 5), (3, 4)$，所以$p = 3/21$，还是错！
3. 上述答案错误的原因是之前考虑的结果不**等可能**(equally likely)。"正确"做法应该是在**区分第一和第二颗骰子**的前提下考虑所有可能的点数对。共36种结果，其中支持事件A的有$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$，所以$p = 6/36$。

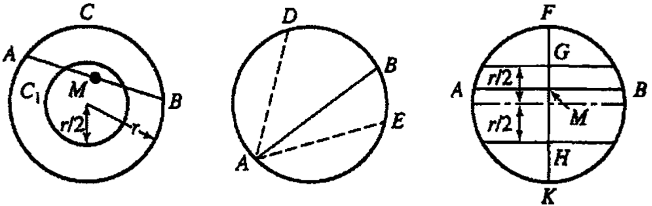
上面例子表明我们需要对之前的古典定义做出改进：如果所有结果**等可能**，那么一个事件的概率等于支持它的结果数与总结果数的比值。

但是，这种改进并不足以消除古典定义的问题。

对古典概率定义的批判：

1. 术语“等可能”实际上意味着“等或然”(equally probable)，因此定义使用的正是有待定义的概念。这通常带来确定N和N_A的困难。
2. 该定义适用的问题类型有限。例如，该定义不适用于偏心骰子：假如掷出4的概率为0.2，那么该数字无法从定义导出。
3. 古典定义似乎是不考虑经验的逻辑必然结果，然而情况并非如此。我们承认某些结果为等可能，正是因为人类的集体经验。例：均匀骰子和新生儿性别。
4. 如果可能的结果有无穷多个，那么在利用古典定义时，我们必须使用长度、面积或者其他度量来确定比值N_A/N。这容易导致概念的模糊。

例（Bertrand悖论）：给定一个半径为r的圆周C，AB是一条“随机选出”的弦，确定其长度ℓ大于内接等边三角形边长$r\sqrt{3}$的概率p。



(a) (b) (c)

解：图(a)，弦AB的中点位于半径为$r/2$的同心圆C_1内。由此$p = 1/4$。
图(b)，假设弦AB的端点A固定，由对称性可知这不影响p。那么支持$\ell > r\sqrt{3}$的端点B位于120°弧DBE上。因此$p = 1/3$。
图(c)，假设弦AB垂直于FK，同上可知此限制不影响p。若AB中点M位于GH上，那么$\ell > r\sqrt{3}$。所以$p = 1/2$。

同一个问题有三个不同的解！实际上，这些答案对应了三个不同的实验。上述例子表明了古典概率定义本身的模糊性。

古典概率定义的价值：确定概率数据和作为工作假设。

- 在许多应用场合，基于长期经验，人们做出了有N个等可能结果的合理假设。于是古典定义以一种“不证自明”的方式被接受。

> 例如“从装有m个黑球和n个白球的盒子里随机摸出一个球，它为白球的概率等于$n/(m+n)$”，或“如果一次电话呼叫在时间区间$(0, T)$里随机发生，则在区间(t_1, t_2)发生呼叫的概率等于$(t_2 - t_1)/T$”。
>
> 这种结论是正确而有用的，但它们的有效性取决于“随机”的含义。在电话里例子中，“未知概率等于$(t_2 - t_1)/T$”并非由呼叫的“随机性”导致，而是**等价**于后者。两者都从过去的电话呼叫记录而非先验推理得出。

- 在一些应用中，无法用足够多次的重复试验来确定各种事件的概率。因此，我们只好假定某些结果等可能，并把古典定义作为**工作假设**来确定概率。如果假设的逻辑后承(logical consequence)与实验结果相符，那么就采纳此假设，否则就拒绝该假设。

> 统计力学例子

例：给定n个粒子和m个盒子且$m > n$，随机地把每个粒子放进一个盒子。预先选定n个盒子，求出每个盒子里有且仅有一个粒子的概率p。

特别地，取$n = 2$和$m = 6$，则问题对应于掷两个骰子。提前选定面（盒子）3和4。

此问题的解取决于对可能结果和支持结果的选择。考虑如下三种知名情况：

Maxwell-Boltzmann statistics

假设每个粒子均区别于其他粒子，并将这n个粒子放进m个盒子各种可能的排列作为结果，那么

$$p = \frac{n!}{m^n}.$$

Bose-Einstein statistics

假设粒子之间不可区分，也就是说，把它们的不同排列视作同一个结果，那么

$$p = \frac{(m-1)!n!}{(n+m-1)!}.$$

Fermi-Dirac statistics

假设粒子之间不可区分，同时每个盒子最多只能容纳一个粒子，那么

$$p = \frac{n!(m-n)!}{m!}.$$

例（续）：当$n=2$和$m=6$时，$p_{\text{M-B}}=2/36$，$p_{\text{B-E}}=1/21$，$p_{\text{F-D}}=1/15$。

评论：在统计力学发展早期，人们倾向于认为只有第一种解才合乎逻辑。但在缺乏实验证据的情况下，这三种模型都只是**假设**。物理学家只接受与实验结果相符合的假设。

古典概率定义的价值（续）

- 假设实验1中事件A的概率$P(A)$和实验2中事件B的概率$P(B)$已知。一般而言，我们无法仅用这些信息确定事件A和B都发生的概率$P(AB)$。但是，若两个实验相互**独立**，则有

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

  在许多情况下，我们推断实验1的结果不会影响实验2的结果，由此先验地确定这种独立性。例如同时进行硬币和骰子实验。

- 如果把古典概率定义作为**假设**，那么它可以成为一种演绎理论的基础而不需要其他假设。在该理论中，公理化定义中的三条假定变成了（容易证明的）定理。但是，这种理论远远没有概率的公理化方法强大。

主要内容

1. 概率的含义
   - 概率论的实际应用
   - 概率的不同定义
   - 概率与归纳

在概率论的应用中，我们会遇到如下问题：假设在一个给定实验中事件A的概率$P(A)$已知，对于该事件在未来**单次**试验中是否会发生，我们能做出何种结论？

分两种情形给出两种截然不同的解答：

1. 假设$P(A)$明显不同于0或1，例如$P(A)=0.6$。在这种情况下，数字0.6仅给我们“关于事件A会发生的一些信心”（概率的**主观**解读，无法被实验验证）。在未来的单次试验中，事件A可能发生或不发生。但即便它不发生，这也不足以构成我们对假设$P(A)=0.6$的质疑。
2. 假设$P(A)=0.999$。现在我们几乎确定事件A在下一次试验中一定会发生（**客观**结论）。如果事件A在下一次试验中不发生，那么我们必须严重怀疑（即使不马上拒绝）$P(A)=0.999$的假设。

这两种情形区分了科学的“软”结论和“硬”结论。概率论所提供的解析工具能够把“主观”论述转化为“客观”论述。

例：后续课程内容表明，$P(A) = 0.6$这一信息能够导出如下结论：如果该试验重复1000次，那么“几乎可以确定”事件A发生的次数介于550和650之间。
考虑把原实验重复1000次作为一个新实验，那么在新实验中，事件

$$A_1 = \{A\text{发生的次数介于550和650之间}\}$$

的概率等于0.999。因此我们推断事件A_1几乎一定会发生。

评论：无论是“主观”还是“客观”结论都依赖于归纳性推理，两者的显著区别仅在于数量而非性质。“客观”结论并非必然而是一种推断(inference)，毕竟任何基于过去经验对未来事件的预测都不是逻辑必然性。这一点并不限于概率论，而适用于所有的科学。

The End