概率论与数理统计

任课教师: 胡家信

清华大学数学科学学系

2021 年 9 月 13 日

任课教师信息

- 任课教师: 胡家信
- 邮件：hujiaxin@tsinghua.edu.cn
- 办公室: 理科楼 3 楼A-320
  Tel: 627 87513.
- 请用邮件联系，勿用手机
- 助教的姓名、邮件将公布在网络学堂上
- 上课时准备纸、笔（有时使用雨课堂，需要手机）。

注意事项

- 总成绩: 作业**(35%)**+终考**(60%)**+讨论、软件**(5%)**
- 作业：★ 下次课前一天的17:00前交作业（一般原则）
  (9月19日周日17:00为第1次交作业时间截止日期）
  ★作业一律交**电子版到网络学堂（扫描）**，
  第1页左上角醒目**学号、姓名**
  ★未按时递交者，成绩0分（特殊情况须老师批准）
- 习题课：第7、9、11、13、15周
  时间：19:20-20:55（周二、周三、周四），内容一样，地点待定。
- 助教答疑：微信群随时答疑。线下答疑的时间、地点（待安排）。

注意事项（继续）

- 教师办公时间：每周五16：00-17：00（不用预约）。
  其它时间请同学们用邮件预约（不用手机）。
- 请随时浏览“网络学堂”通告。
- 请将联系方式（包括邮件、手机）放在“网络学堂”。
- 课代表（两名），告知联系地址（手机、学号、院系）。
- 按清华规定，2周不上课，予以退学。特殊情况，须书面请假。本课程2次不上课，视为自动退课。
- 助教将对作业进行查重。

问题答复

- **问：是否可以补交作业？**
  答：不可以。
  （疾病等特殊原因除外，但须向老师说明）。

- **问：作业抄袭如何处理？**
  答：零容忍。

- **问：我交了作业但网络学堂显示未交，怎么办？**
  答：只要按时交作业，这种情况不会出现。

参考书（注意版权，请勿外传）

1. K. Chung and F. AitSahlia: **Elementary probability theory**, 2003 (第1页-253页)（概率部分）（已放在网络学堂）.
2. M. Spiegel，J. Schiller and R. Srinivasan, **Probability and Statistics** (第151页-410页)（统计部分）（将放在网络学堂）.
3. 熟悉**"Excel"**软件中，概率统计内置函数及应用，也可用**R**软件或**Python**软件等

本课程预备知识要求：熟练掌握微积分。

几点建议

- 预习课本(非常重要)。
- 做好笔记(非常重要)。
- 独立做完教材每一道习题(非常重要).
- 独立证明教材重要定理(非常重要).
- 勤于思考(非常重要).
- 不缺堂(缺堂两次，视为自动退课).

- “学而不思则罔，思而不学则殆”（孔子）
- “由厚变薄，由薄变厚”（华罗庚）
- “内事不决问info（清华网络学堂），外事不决问google”（胡家信）

两个著名例子

1 赌金分配问题：Fermat先生和Pascal先生坐在巴黎咖啡馆，玩一种简单的游戏，抛硬币。若正面，Fermat 得1 分；若反面，Pascal 得1分。先得10分的人赢全部赌注160法郎（每人各出80法郎）。但奇怪的事情发生了，Fermat的朋友病了，他要在8 : 7领先的形势下赶回Toulouse（图卢兹,法国南部）。**问：全部赌注160法郎如何分配?**

- Pierre de Fermat，1601年8月17日–1665年1月12日，法国人，终年64岁。律师，数学家，已婚，五个子女。
- Blaise Pascal，1623年6月19日–1662年8月19日，法国人，终年39岁，死于疾病，从未结婚。三岁失母，父为收税员。

赌金分配问题

解法1（Luca Pacioli（帕乔利），1494年）：按每个人所赢的局数分配赌金（过去）。

Fermat获得$160 \times \frac{8}{15} \simeq 85.3333\cdots$法郎，

Pascal获得$160 \times \frac{7}{15} \simeq 74.6666\cdots$法郎。

该方法的缺陷在哪里？

- Luca Pacioli (1445-1517)，出生意大利，终年72岁。

赌金分配问题（继续）

缺陷：若赌局在一次后中断，即某个选手（如Fermat）赢一次，而另一个选手（如Pascal）未赢的情况下中断。此时，全部赌金归一个选手获得。但这显然不符合直觉（公平），特别是赌局很多的情况下。

Niccolò Tartaglia(塔尔塔利亚，16世纪中期)发现这个缺陷，并提出一个解决问题的办法，但他不自信。事实上，Tartaglia所提出的解决方法也有问题，如100次中，比分为$65:55$和比分为$99:89$的两种不同情形，所分赌资一样。

- Niccolò Tartaglia (1500-1557)，出生意大利，终年57岁。

赌金分配问题（Fermat的分法，1654年）

解法2（Fermat的分法。1654年8月24日（顺治十一年），Fermat给Pascal书信)：按每个人所需要赢的局数分配赌金（未来）。 首先注意赌4局可以定胜负：因为Fermat要赢2分，Pascal要赢3分，所以共需要$2+3-1=4$次就可决定胜负：这是因为在4局中，若Fermat得2分或2分以上，则Fermat 赢；若Fermat得0分或1分，则Pascal至少可得3分，那么Pascal赢。结果有16种情况：

1. FFFF（Fermat赢4次）
2. FFFP，FFPF，FPFF，PFFF（Fermat赢3次）
3. FFPP，FPFP，FPPF，PFFP，PFPF，PPFF（Fermat赢2次）
4. FPPP，PFPP，PPFP，PPPF（Fermat赢1次）
5. PPPP（Fermat全输）

共1+4+6+4+1=16种情况。

在16种情况中Fermat至少赢2分（包括2分）的情况有

$$1+4+6=11\text{（种)}.$$

故Fermat赢的概率为（假定色子是均匀的）

$$\frac{11}{16}.$$

他应得赌金为

$$160\times\frac{11}{16}=110\text{（法郎)}.$$

解毕。

Fermat的分法总结

1. 需赌4局定胜负；
2. 4局共有16种情形;
3. Fermat胜出情形有11种，故得全部赌金的$\frac{11}{16}$，即110法郎。或者，Pascal胜出情形有5种，故得全部赌金的$\frac{5}{16}$，即50法郎。

一个聪明人的分法（但结果一样）

一个**聪明人**反对Fermat的分法：只要胜者定出(*最多赌4次，但也可能2次或3次就可决定胜负*), 不必考虑上面所有情况。实际上，只需下列情况，Fermat就可赢(注意Fermat赢2次即可得全部赌注)：

- FF，FPF，FPPF （Fermat第一次**赢**的所有情形）
- PFF，PFPF，PPFF （Fermat第一次**输**的所有情形）

共6种情况。故Fermat赢的**机会**（**概率**）为

$$\frac{1}{2}\times\frac{1}{2}+\frac{1}{8}+\frac{1}{16}+\frac{1}{8}+\frac{1}{16}+\frac{1}{16}=\frac{11}{16}.$$

他应得赌金为

$$160\times\frac{11}{16}=110（法郎).$$

结果一样，但方法不同。

情况变化呢？

课堂讨论：**Fermat**走的时候结果为6 : 9，情况如何？

情况变化呢？

Fermat走的时候结果为6 : 9，情况如何？

答案：需赌4局决胜负（因为：$4+1-1=4$）。4局总共有16种情况，其中Fermat赢的情况只有1种，他赢得的机会（概率）为$\frac{1}{16}$，Fermat 分得赌注为$\frac{1}{16}\times 160=10$法郎。

想一想，结果是否合理？（Pascal只需赢一次，而Fermat需赢4次！）

课后思考：利用聪明人的解法，结果是否一样？

Pascal的推广

课堂讨论：若**Fermat**需r局赢得赌资，而**Pascal**需s局赢得赌资，情况如何？

Pascal的推广（继续）

若**Fermat**需r局赢得赌资，而**Pascal**需s局赢得赌资，情况如何？

答案：赌资按下列比例分配

$$\frac{\sum_{k=r}^{r+s-1}\binom{r+s-1}{k}}{\sum_{k=s}^{r+s-1}\binom{r+s-1}{k}}=\frac{\binom{r+s-1}{r}+\binom{r+s-1}{r+1}+\cdots+\binom{r+s-1}{r+s-1}}{\binom{r+s-1}{s}+\binom{r+s-1}{s+1}+\cdots+\binom{r+s-1}{s+r-1}}.$$

想一想，为什么？留以后（第三次课）讨论!

如$r=2,\ s=3$,上述比为

$$\frac{\sum_{k=2}^{4}\binom{4}{k}}{\sum_{k=3}^{4}\binom{4}{k}}=\frac{\binom{4}{2}+\binom{4}{3}+\binom{4}{4}}{\binom{4}{3}+\binom{4}{4}}=\frac{6+4+1}{4+1}=\frac{11}{5}.$$

2 太阳从东边升起的概率：设太阳连续从东边升起n次，问再次从东边升起的概率是多少?

太阳从东边升起的概率

Laplace的答案是：$\frac{n+1}{n+2}$.

小于1!

该问题留以后讨论!

- **集合:**不同的东西放在一起的总称，或叫“空间"，“样品空间"等等。
- “**样品空间**": Ω (希腊24个字母中的最后一个字母)，其元素用$\omega, \omega_1, \omega_2$等等表示。
- 非常有用的集合：**空集**，$\emptyset$.
- 符号：$A \subset \Omega, A \subseteq \Omega$.

  $A = B \Leftrightarrow A \subseteq B, B \subseteq A$.

  例：偶数集$= \{x \in \mathbb{R} : \sin(\pi x/2) = 0\}$.

集合的运算

- 余集(complement): $A^c = \Omega \setminus A = \{\omega \in \Omega : \omega \notin A\}$.
- 交换律(commutative law):

$$A \cup B = B \cup A, A \cap B = B \cap A.$$

- 结合律(associative law):

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

集合的运算（续）

- 分配律(distributive law):

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

- De Morgan定律:

$$(A \cup B)^c = A^c \cap B^c,$$
$$(A \cap B)^c = A^c \cup B^c.$$

- 差集:

$$A \backslash B = \{\omega \in A : \omega \notin B\}.$$

集合的运算（续）

- 对称差(symmetric difference):

$$A \Delta B = (A \backslash B) \cup (B \backslash A).$$

- 可数交,可数并:

$$\bigcup_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{对某个} i\},$$

$$\bigcap_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{对所有} i\}.$$

- 记号: $A + B = A \cup B$ 若$A \cap B = \emptyset$.

集合的运算（续）

- **De Morgan定律**:

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B) \text{（验证）}$$

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c.$$

- 两个有用的集合:

$$\bigcup_{n=1}^{\infty}\left(\bigcap_{m=n}^{\infty} A_m\right) \text{(注意：} B_n = \bigcap_{m=n}^{\infty} A_m \text{关于} n \text{递增)}$$

$$\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right) \text{(注意：} C_n = \bigcup_{m=n}^{\infty} A_m \text{关于} n \text{递减)}.$$

例子：设$A_m = \left[0, \frac{1}{m}\right]$，则上述两个集合？

集合的运算（续）

- 特征函数(indicator):

$$I_A(\omega) = \begin{cases} 1 & 若 \omega \in A, \\ 0 & 若 \omega \notin A. \end{cases}$$

- 记号$\vee, \wedge$

$$a \vee b = \max\{a, b\},$$
$$a \wedge b = \min\{a, b\}.$$

例子： $2 \vee 10 = 10$，$2 \wedge (-1) = -1$.

命题

命题： 证明

$$I_{A\cap B} = I_A \times I_B = I_A \wedge I_B, \tag{1}$$

$$I_{A\cup B} = I_A \vee I_B, \tag{2}$$

$$I_{A\cup B} + I_{A\cap B} = I_A + I_B. \tag{3}$$

证明: 对第一个方程需证明

$$I_{A\cap B}(\omega) = I_A(\omega) \times I_B(\omega), \text{对每个}\omega \in \Omega\text{成立。}$$

直接验证即可（略）。 **课堂练习：** 证明(2)

现证方程(3)。事实上，利用$I_A = 1 - I_{A^c}$,

$$I_{A\cup B} = 1 - I_{A^c\cap B^c} = 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B)$$
$$= I_A + I_B - I_A I_B = I_A + I_B - I_{A\cap B}.$$

课堂练习

命题：证明

$$I_{A\Delta B} = I_A + I_B - 2I_{AB}, \tag{4}$$
$$= I_A + I_B \pmod 2, \tag{5}$$
$$A\Delta(B\Delta C) = (A\Delta B)\Delta C. \tag{6}$$

- 对方程（4）直接验证。
- （5）显然，这里：$a = b \bmod 2$是指差$a - b$是一个被2整除的数，即$(I_A + I_B - 2I_{AB}) - (I_A + I_B) = -2I_{AB}$能被2整除.
- 对（6），利用（5）验证下列等式成立

$$I_{A\Delta(B\Delta C)} = I_{(A\Delta B)\Delta C}.$$

(若两个集合的特征函数相同，则该两个集合相同。)

有限样本空间上概率的定义

定义：有限样本空间Ω（即集合Ω包含有限个点）上的概率P是一个**集函数**（定义在集合上的函数），满足下面三个条件:

- （非负性）$P(A) \geq 0$, 对每个$A \subset \Omega$;
- （正规化）$P(\Omega) = 1$;
- （有限可加性）对任意两个不相交的$A, B \subset \Omega$,均有

$$P\left(A+B\right) = P\left(A\right) + P(B).$$

例：设样本空间Ω如下:

$$\Omega = \{1,2,3,4,5,6\}.$$

定义概率P如下: $P(\{k\}) = \frac{1}{6}$对任意$1 \le k \le 6$.

计算:

$$P(\{1,2,3\}) = ?$$

$$P(\{1,2,3,4,5\}) = ?$$

进而计算Ω所有子集(或事件)的概率.

注意符号：$P(k) := P(\{k\})$.

定义(σ代数的概念) 设X是一个集合, $\mathcal{F}$是X的子集构成的集合，如果$\mathcal{F}$关于补运算和可数并运算是封闭的：

$$1)\text{若}A\in\mathcal{F}\text{，则}A^c\in\mathcal{F}\text{；}\quad 2)\text{若每个}A_i\in\mathcal{F}\text{，则}\bigcup_{i=1}^{\infty}A_i\in\mathcal{F},$$

则称$\mathcal{F}$是X上的一个σ代数。

定义(测度的概念)测度P是定义域为$\mathcal{F}$的一个集函数，满足条件:

- （非负性）$P(A)\geq 0$, 对每个$A\in\mathcal{F}$;
- （可数可加性）对任意两两互不相交的$A_n\in\mathcal{F}$，均有

$$P\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n).$$

利用可数可加性知：$P(\emptyset)=0$（空集的零测性）。**概率是一个满足$P(\Omega)=1$的测度**，有时也称**概率测度**P。

概率的计算

例：一枚色子掷6次，问每次出现的数字都不同的概率是多少?

例：一枚色子掷6次，问每次出现的数字都不同的概率是多少?

解：掷6次,共有6^6中不同的结果(考虑顺序).
每次出现不同的数字的情况有6!种.
所求事件的概率为$\frac{6!}{6^6} \approx 0.015432$.

概率的性质

设P是有限样本空间的概率，则下列性质成立：

- 对任意$A, B \subset \Omega$，均有
$$P(A \cup B) + P(A \cap B) = P(A) + P(B).$$
- 若$A \subset B$，则$P(A) \leq P(B)$;
- 对每个$A \subset \Omega$，有$0 \leq P(A) \leq 1$;
- 对两两不相交的$A_1, A_2, \cdots, A_n \subset \Omega$，均有
$$P(A_1 + A_2 + \cdots + A_n) = \sum_{k=1}^{n} P(A_k).$$
- 对任意$A_1, A_2, \cdots, A_n \subset \Omega$，均有
$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{k=1}^{n} P(A_k).$$

证明：仅证明第一个性质：对任意$A, B \subset \Omega$，均有

$$P\left(A \cup B\right) + P\left(A \cap B\right) = P(A) + P(B).$$

事实上，$A \cup B = A + B \setminus (A \cap B)$，所以

$$\begin{aligned} P(A \cup B) &= P(A + B \setminus (A \cap B)) \\ &= P(A) + P(B \setminus (A \cap B)) \quad \text{（有限可加性）} \\ &= P(A) + P(B) - P(A \cap B) \quad \text{（有限可加性）}, \end{aligned}$$

从而，

$$P(A \cup B) + P\left(A \cap B\right) = P(A) + P(B).$$

证毕（其余性质同理证明）。

课堂练习

证明：若$A \subset B$，则$P(A) \leq P(B)$。

可数样本空间上概率的定义

定义：设Ω是一个**可数**集合，那么概率P是一个集函数, 满足下面三个条件:

- （非负性）$P(A) \geq 0$, 对每个$A \in \mathcal{F}$;
- （正规化）$P(\Omega) = 1$;
- （可数可加性）对任意两两互不相交的$A_n \subset \Omega$，均有

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P\left(A_n\right).$$

注意：此处$\mathcal{F}$是Ω的所有子集（包括空集）构成的集合。

比较有限样本空间：概率定义中前两个性质一样，第三个性质变为**可数**可加性（不是**有限**可加性）。注意，**可数**可加性意味**有限**可加性。

可数样本空间上概率的构造

例：设$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n, \cdots\}$是一个无限集合。令$\{p_i\}_{i=1}^{\infty}$是一个数列, 满足每个$p_i \geq 0$，且

$$\sum_{i=1}^{\infty} p_i = 1 \quad （如 p_i = 2^{-i}），$$

构造集函数P如下:

$$P(\{\omega_i\}) = p_i，任意 i \geq 1;$$
$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

证明P是一个**概率**，即满足上面三个条件.

证明（略）

第一次作业

第一次作业(钟开来书)：

P. 17-19：第1，3，8，9，12题，第15，18，19，20，21题。

P. 42-43：第6，7，8，9，10题。

作业：(1)在作业第1页左上角醒目**学号、姓名**。

(2)将作业扫描成**单个、PDF**文件（不要压缩、不要Word格式）

下节预习内容：独立、抽样、分配。