概率

司继春

[1]上海对外经贸大学

2023年9月

概览

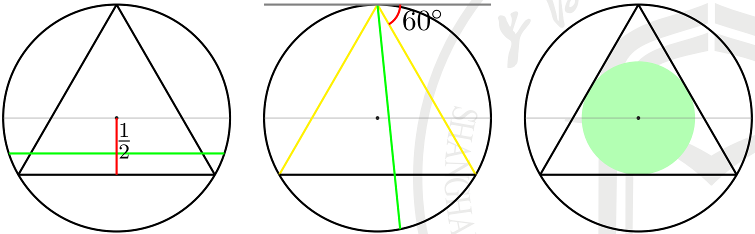
样本空间

样本空间的定义

样本空间（sample space）Ω为我们关心的随机试验的所有结果的集合，而Ω中的元素$\omega \in \Omega$称之为样本点（sample point）。

1. 随机从一堆扑克牌中抽取一张扑克，其花色的样本空间为：$\Omega_1 = \{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\}$而样本点为$\heartsuit$、$\spadesuit$、$\clubsuit$以及$\diamondsuit$。
2. 某银行一天所接待的所有客户数，其样本空间为$\Omega_2 = \mathbb{N}$，样本点为自然数。
3. 随机从人群中抽取一个人，其身高的样本空间为$\Omega_3 = \mathbb{R}^+$，而样本点为正实数。

贝特朗悖论



解法1　解法2　解法3

事件

- 我们称样本空间Ω的子集（包含Ω本身）为事件（event）。
- 如果A为Ω的一个子集，如果随机试验的结果为A中的一个样本点，我们称之为发生了事件A。
- 实际上，概率是定义在事件上的，而非样本点上的。

事件的运算法则

1. 交换律：$A \cup B = B \cup A$，$A \cap B = B \cap A$
2. 结合律：$A \cup (B \cup C) = (A \cup B) \cup C$，$A \cap (B \cap C) = (A \cap B) \cap C$
3. 分配率：$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$，$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4. 德摩根律：$(A \cup B)^c = A^c \cap B^c$，$(A \cap B)^c = A^c \cup B^c$，以及：

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

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

互斥事件

- 如果两个事件A和B满足$A \cap B = \emptyset$，我们称之为互斥事件（disjoint or exclusive）。
- 如果对于一系列事件$A_1, A_2 \ldots$，对于任意i和j，有$A_i \cap A_j = \emptyset$，则称之为两两互斥事件。
- 如果$A_1, A_2 \ldots$为两两互斥事件，且$\cup_{i=0}^{\infty} A_i = \Omega$，则$A_1, A_2 \ldots$为样本空间的一个划分（partition）。

$\sigma-$代数

对于一个一般的样本空间，有数量巨大的子集。我们希望挑出那些我们需要研究的子集，同时剔除那些性质不是十分良好的子集，这就诞生了$\sigma-$代数的概念。

$\sigma-$代数

如果样本空间Ω的一系列子集的集合$\mathscr{F}$满足:

1. $\emptyset \in \mathscr{F}$
2. 若$A \in \mathscr{F}$，则$A^c \in \mathscr{F}$
3. 若$A_1, A_2 \ldots \in \mathscr{F}$，则$\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$

我们称$\mathscr{F}$为一个$\sigma-$代数，或者$\sigma-$域。

根据德摩根律，以上第2和第3个条件要求可数个集合的交集也要在$\mathscr{F}$中:

$$\bigcap_{i=1}^{\infty} A_i = \left[\bigcup_{i=1}^{\infty} A_i^c\right]^c \in \mathscr{F}$$

$\sigma-$代数：直觉

为什么需要引入$\sigma-$代数这个概念？

- 限定了概率界定的范围：不是每个Ω的子集都可以被定义上概率（讲义例1.9）
- 与“信息”息息相关，比如，如果两个事件A和B是否发生是已知的，那么：
  - 空集必然不可能发生
  - A和B的补集是否发生也是已知的
  - A和B是否至少有一个发生也是已知的
  - A和B是否同时发生也是已知的

$\sigma-$代数

扑克牌颜色的$\sigma-$代数

对于$\Omega=\{\heartsuit,\spadesuit,\clubsuit,\diamondsuit\}$，如果我们关心抽出的扑克牌是红色还是黑色，即两个集合：$\{\clubsuit,\spadesuit\}$和$\{\heartsuit,\diamondsuit\}$，那么现构建对应的$\sigma-$代数。根据要求，首先集合中应该包含：

$$\{\emptyset,\{\clubsuit,\spadesuit\},\{\heartsuit,\diamondsuit\},\Omega_1\}\triangleq F_1$$

可以验证，以上F_1已经满足$\sigma-$代数的要求，因而$\mathscr{F}^{\#}=F_1$即我们要构建的$\sigma-$代数。

$\sigma-$代数

扑克牌花色的$\sigma-$代数

对于$\Omega=\{\heartsuit,\spadesuit,\clubsuit,\diamondsuit\}$，如果我们关心单点集：$\{\heartsuit\}$,$\{\clubsuit\}$,$\{\spadesuit\}$,$\{\diamondsuit\}$的概率，那么自然地，我们也会关心四个单点集的交、并的概率。为了构建$\sigma-$代数，根据定义，我们从：

$$\{\emptyset,\{\heartsuit\},\{\clubsuit\},\{\spadesuit\},\{\diamondsuit\},\Omega_1\}\triangleq F_1$$

开始，由于$\sigma-$代数要求F_1中的元素的并也要包含在$\sigma-$代数中，从而可以继续扩充：

$$\left\{\begin{array}{cccccc}\emptyset, & \{\heartsuit\}, & \{\clubsuit\}, & \{\spadesuit\}, & \{\diamondsuit\}, & \Omega_1, \\ \{\heartsuit,\clubsuit\}, & \{\heartsuit,\spadesuit\}, & \{\heartsuit,\diamondsuit\}, & \{\clubsuit,\spadesuit\}, & \{\clubsuit,\diamondsuit\}, & \{\spadesuit,\diamondsuit\},\end{array}\right\}\triangleq F_2$$

$\sigma-$代数

扑克牌花色的$\sigma-$代数（续）

然而以上F_2仍然不达到要求，需要继续扩充：

$$\left\{\begin{array}{cccccc} \emptyset, & \{\heartsuit\}, & \{\clubsuit\}, & \{\spadesuit\}, & \{\diamondsuit\}, & \Omega_1, \\ \{\heartsuit,\clubsuit\}, & \{\heartsuit,\spadesuit\}, & \{\heartsuit,\diamondsuit\}, & \{\clubsuit,\spadesuit\}, & \{\clubsuit,\diamondsuit\}, & \{\spadesuit,\diamondsuit\}, \\ \{\heartsuit,\clubsuit,\spadesuit\}, & \{\heartsuit,\clubsuit,\diamondsuit\}, & \{\heartsuit,\spadesuit,\diamondsuit\}, & \{\clubsuit,\spadesuit,\diamondsuit\} & & \end{array}\right\} \triangleq F_3$$

经验证以上F_3已经满足$\sigma-$代数的定义，从而$\mathscr{F}^* = F_3$即我们要构建的$\sigma-$代数。

生成的$\sigma-$代数

- 值得注意的是，以上构建的$\mathscr{F}^*$同样也包含$\{\clubsuit,\spadesuit\}$和$\{\heartsuit,\diamondsuit\}$两个颜色的集合。
- 经过比较可以发现，$\mathscr{F}^\# \subset \mathscr{F}^*$。实际上，$\mathscr{F}^\#$是包含$\{\clubsuit,\spadesuit\}$和$\{\heartsuit,\diamondsuit\}$ 两个集合的最小的$\sigma-$代数。

生成的$\sigma-$代数

对于一系列的事件集合$\mathscr{C}$，我们称包含$\mathscr{C}$中所有事件的最小的$\sigma-$代数为为$\mathscr{C}$生成的$\sigma-$代数，记为$\sigma\langle\mathscr{C}\rangle$ 。

生成的$\sigma-$代数

$\sigma\langle\{\{\clubsuit,\spadesuit\},\{\heartsuit,\diamondsuit\}\}\rangle=\mathscr{F}^\#$， $\sigma\langle\{\{\heartsuit\},\{\clubsuit\},\{\spadesuit\},\{\diamondsuit\}\}\rangle=\mathscr{F}^*$

Borel $\sigma-$代数

Borel代数

若我们关心的样本空间为$\Omega = \mathbb{R}$，我们令$\mathscr{I}$为所有开区间的集合：

$$\mathscr{I} = \{(a,b) \mid -\infty < a < b < +\infty\}$$

那么包含$\mathscr{I}$的最小$\sigma-$代数：

$$\mathscr{B} = \sigma \langle \mathscr{I} \rangle$$

称为Borel$\sigma-$代数或Borel域，而$\mathscr{B}$中的元素成为Borel集。

注意由于：

$$(a,b] = \bigcap_{i=1}^{\infty} \left(a, b + \frac{1}{i}\right)$$

因而所有的左开右闭区间也都是Borel集。同理可证所有的左闭右开区间$[a,b)$、闭区间$[a,b]$及其可数并、交都为Borel集。

概率的定义

经过以上准备后，可以定义概率：

概率的公理化定义（Axioms of Probability）

（Kolmogorov axioms）给定一个样本空间Ω以及相应的$\sigma-$代数$\mathscr{F}$，函数$\mathscr{P}:\mathscr{F}\to[0,1]$若满足：

1. 对于所有的事件$A\in\mathscr{F}$，$\mathscr{P}(A)\geq 0$

2.$\mathscr{P}(\Omega)=1$

3. 若$A_1,A_2,\ldots\in\mathscr{F}$为两两互斥事件，

则$\mathscr{P}(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}\mathscr{P}(A_i)$（可数可加性或可列可加性）

那么我们称$\mathscr{P}$为概率函数或概率测度。$(\Omega,\mathscr{F},\mathscr{P})$三元组称为概率空间。

概率的定义

抛硬币

现在我们进行一项抛硬币的随机试验。记正面为H，反面为T，那么我们关心的样本空间为$\Omega=\{H,T\}$。假设硬币质地均匀，即正面和反面的概率相等，那么

$$\mathscr{P}(\{H\})=\mathscr{P}(\{T\})$$

包含$\{H\}$,$\{T\}$的最小$\sigma-$代数为

$$\mathscr{F}=\{\emptyset,\Omega,\{H\},\{T\}\}$$

根据概率的定义，

$$\mathscr{P}(\Omega)=\mathscr{P}(\{H\}\cup\{T\})=\mathscr{P}(\{H\})+\mathscr{P}(\{T\})=1$$

因而

$$\mathscr{P}(\{H\})=\mathscr{P}(\{T\})=0.5$$

离散样本空间概率的定义

离散样本空间的概率

令$\Omega=\{s_1,s_2,\ldots,s_n\}$为有限集，令$\mathscr{F}$为$\Omega$的所有子集构成的$\sigma-$代数。令$p_1,p_2,\ldots p_n$为非负实数且$\sum_{i=1}^n p_i=1$。对于任意集合$A\in\mathscr{F}$，定义

$$\mathscr{P}\left(A\right)=\sum_{\{s_i,i\in A\}} p_i$$

那么$\mathscr{P}$为$\mathscr{F}$上的概率函数。对于可数集$\Omega=\{s_1,s_2,\ldots\}$可类似构建概率函数。

泊松分布

泊松分布的定义

我们关心在一个小时之内到达某银行的客户数，客户数为可数集，样本空间为$\Omega = \{0,1,2\ldots\}$。取一个非常大的自然数n，我们可以把一个小时分解为等长的n段，即$(0,\frac{1}{n}],(\frac{1}{n},\frac{2}{n}]\ldots(\frac{n-1}{n},1]$，当$n$很大时，一个区间段内有两个客户到达的概率几乎可以忽略不计。假设每段时间客户到达的概率相等，且反比于n，不妨假设为$\frac{\lambda}{n}$，那么一小时内总的人数：

$$\mathscr{P}^*(\{k\}) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

泊松分布

泊松分布的定义

令$n \to \infty$，则

$$\frac{\binom{n}{k}}{n^k} = \frac{n!}{k! \cdot (n-k)! \cdot n^k} \to \frac{1}{k!}$$

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

因而

$$\mathscr{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} > 0$$

且$\sum_{k=0}^{\infty} \mathscr{P}(\{k\}) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$（$e^{\lambda x}$在$x = 0$处泰勒展开可得），因而根据以上定理，上式定义的概率函数$\mathscr{P}$即定义了样本空间Ω上任意$\sigma-$代数的概率函数。

$\mathbb{R}$上的概率定义

为了在$\mathbb{R}$上定义概率函数，我们首先引入分布函数（distribution function, d.f.）的概念：

分布函数

如果函数$F:\mathbb{R}\to\mathbb{R}$满足：

1. 单调性：$F(a)\leq F(b), a\leq b$
2. 右连续：$\lim_{x\to x_0^+}F(x)=F(x_0)$
3. $F(-\infty)=0, F(\infty)=1$

则称F为分布函数。

分布函数

- 令

$$\delta_t(x) = \begin{cases} 0 & x < t \\ 1 & x \geq t \end{cases}$$

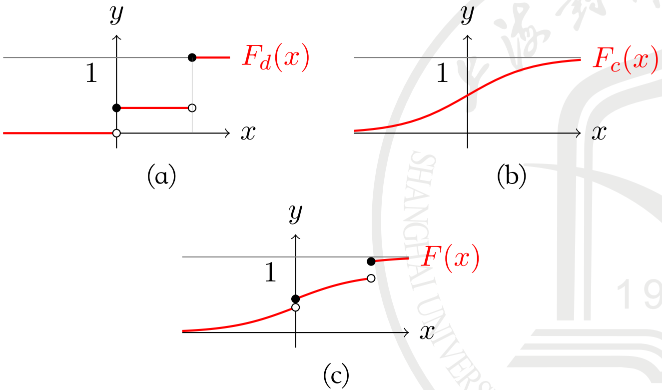
若$\{a_j\}$为可数集，$b_j > 0, \sum_j b_j = 1$，
则$F(x) = \sum_j b_j \delta_{a_j}(x)$为分布函数，我们称之为**离散型分布函数**（**discrete d.f.**）

- 处处连续的分布函数称为**连续型分布函数**（**continuous d.f.**）

分布函数的分解

每个分布函数都可以写为一个离散型分布函数和一个连续型分布函数的凸组合，且该分解唯一。

分布函数



$\mathbb{R}$上的概率定义

可以使用分布函数定义$\mathbb{R}$上的概率函数。定义：

$$P((-\infty, x]) = F(x)$$

则对于任意的$-\infty < a < b < +\infty$，有：

$$\begin{cases} P((a,b]) = F(b) - F(a) \\ P((a,b)) = F(b-) - F(a) \\ P([a,b)) = F(b-) - F(a-) \\ P([a,b]) = F(b) - F(a-) \end{cases}$$

其中$F(b-) = \lim_{x \to b^-} F(x)$。据此可以定义所有Borel集的概率，从而产生概率空间$(\mathbb{R}, \mathscr{B}, P)$。

概率的性质

1. $\mathscr{P}(A) \le 1$
2. $\mathscr{P}(A^c) = 1 - P(A)$
3. $\mathscr{P}(\emptyset) = 0$
4. $\mathscr{P}(A \cup B) + \mathscr{P}(A \cap B) = \mathscr{P}(A) + \mathscr{P}(B)$
5. $A \subset B \Rightarrow \mathscr{P}(A) = \mathscr{P}(B) - \mathscr{P}(B \backslash A) \le \mathscr{P}(B)$
6. $\mathscr{P}(\bigcup_i A_i) \le \sum_i \mathscr{P}(A_i)$
7. 如果$C_1, C_2, \ldots$为样本空间Ω的一个划分，那么

$$\mathscr{P}(A) = \sum_{i=1}^{\infty} \mathscr{P}(A \cap C_i)$$

条件概率

条件概率

如果A和B为Ω中的两个事件，且$\mathscr{P}(B) > 0$，那么给定B，事件A发生的条件概率为:

$$\mathscr{P}(A|B) = \frac{\mathscr{P}(A \cap B)}{\mathscr{P}(B)}$$

全概率公式

全概率公式

如果$C_1, C_2, ...$为样本空间Ω的一个划分，那么：

$$\mathscr{P}(A) = \sum_{i=1}^{\infty} \mathscr{P}(A|C_i) \cdot P(C_i)$$

特别的，对于任意事件B，有

$$\mathscr{P}(A) = \mathscr{P}(A|B) \cdot \mathscr{P}(B) + \mathscr{P}(A|B^c) \cdot \mathscr{P}(B^c)$$

贝叶斯法则

如果事件A和B都有正的概率，那么我们可以同时定义$\mathscr{P}(A|B)$以及$\mathscr{P}(B|A)$，根据定义：

$$\mathscr{P}(A\cap B)=\mathscr{P}(A|B)\,\mathscr{P}(B)=\mathscr{P}(B|A)\,\mathscr{P}(A)$$

从而：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(A\cap B)}{\mathscr{P}(B)}=\frac{\mathscr{P}(B|A)\cdot\mathscr{P}(A)}{\mathscr{P}(B)}$$

以上关系式我们称之为贝叶斯法则（**Bayes' Rule**）。更进一步，根据全概率公式：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(B|A)\cdot P(A)}{\mathscr{P}(B)}=\frac{\mathscr{P}(B|A)\cdot\mathscr{P}(A)}{\mathscr{P}(B|A)\cdot\mathscr{P}(A)+\mathscr{P}(B|A^c)\cdot\mathscr{P}(A^c)}$$

贝叶斯法则

双胞胎

如果在所有双胞胎中，同卵双胞胎的比率是1/3，异卵双胞胎的比率是2/3。生物学理论告诉我们，同卵双胞胎性别一定相同，而异卵双胞胎性别相同的概率为1/2。如果观察到双胞胎的性别相同，那么其为同卵双胞胎的概率为：

$$\mathscr{P}(\text{同卵双胞胎}|\text{性别相同}) = \frac{\mathscr{P}(\text{性别相同}|\text{同卵双胞胎})\,\mathscr{P}(\text{同卵双胞胎})}{\mathscr{P}(\text{性别相同})}$$

而根据全概率公式，在双胞胎中，性别相同的概率：

$$\begin{aligned}\mathscr{P}(\text{性别相同}) &= \mathscr{P}(\text{性别相同}|\text{同卵双胞胎})\,\mathscr{P}(\text{同卵双胞胎})\\ &+ \mathscr{P}(\text{性别相同}|\text{异卵双胞胎})\,\mathscr{P}(\text{异卵双胞胎})\\ &= 1\times\frac{1}{3}+\frac{1}{2}\times\frac{2}{3}=\frac{2}{3}\end{aligned}$$

因而如果观察到性别相同，那么该双胞胎为同卵双胞胎的概率为：

$$\mathscr{P}(\text{同卵双胞胎}|\text{性别相同}) = \frac{1\times\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

贝叶斯公式

贝叶斯公式

如果$B_1, B_2, \ldots$为样本空间Ω的一个划分，令$A \in \mathscr{B}$，那么：

$$\mathscr{P}(B_i|A) = \frac{\mathscr{P}(A|B_i)\,\mathscr{P}(B_i)}{\sum_{j=1}^{\infty} \mathscr{P}(A|B_j)\,\mathscr{P}(B_j)}$$

独立性

统计独立性

如果两个事件A和B满足：

$$\mathscr{P}(A\cap B)=\mathscr{P}(A)\cdot\mathscr{P}(B)$$

那么我们称事件A和B为独立事件。

如果假设$\mathscr{P}(B)\neq 0$，且事件A和事件B独立，那么根据条件概率的定义：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(A\cap B)}{\mathscr{P}(B)}=\frac{\mathscr{P}(A)\cdot\mathscr{P}(B)}{\mathscr{P}(B)}=\mathscr{P}(A)$$

因而如果条件概率存在，那么事件A和事件B独立意味着两个事件之间不能相互预测。

独立性

三个事件的独立性

现在抛掷两枚骰子，我们关心如下三个事件：

$$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),\}$$
$$B = \{\text{两枚骰子之和介于7和10之间}\}$$
$$C = \{\text{两枚骰子之和为2,7或者8}\}$$

可以计算得到：$\mathscr{P}(A) = \frac{1}{6}$，$\mathscr{P}(B) = \frac{1}{2}$，$\mathscr{P}(C) = \frac{1}{3}$，而：

$$\begin{aligned}\mathscr{P}(A \cap B \cap C) &= \mathscr{P}(\{(4,4)\}) \\ &= \frac{1}{36} = \mathscr{P}(A) \cdot \mathscr{P}(B) \cdot \mathscr{P}(C)\end{aligned}$$

然而：$\mathscr{P}(B \cap C) = \frac{11}{36} \neq \mathscr{P}(B) \cdot \mathscr{P}(C)$，因而事件B和C并不独立。

统计独立性

独立性

我们称一系列事件$A_1, A_2, ...A_n$为相互独立的（**mutually independent or jointly independent**），如果对于任意的子列$A_{i_1}, A_{i_2}, ...A_{i_k}$，有：

$$\mathscr{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathscr{P}\left(A_{i_j}\right)$$

其中$\{i_1, ..., i_k\} \subset \{1, 2, ..., n\}, 1 \leq k \leq n$。

作业

- 练习1.3，1.5，1.10