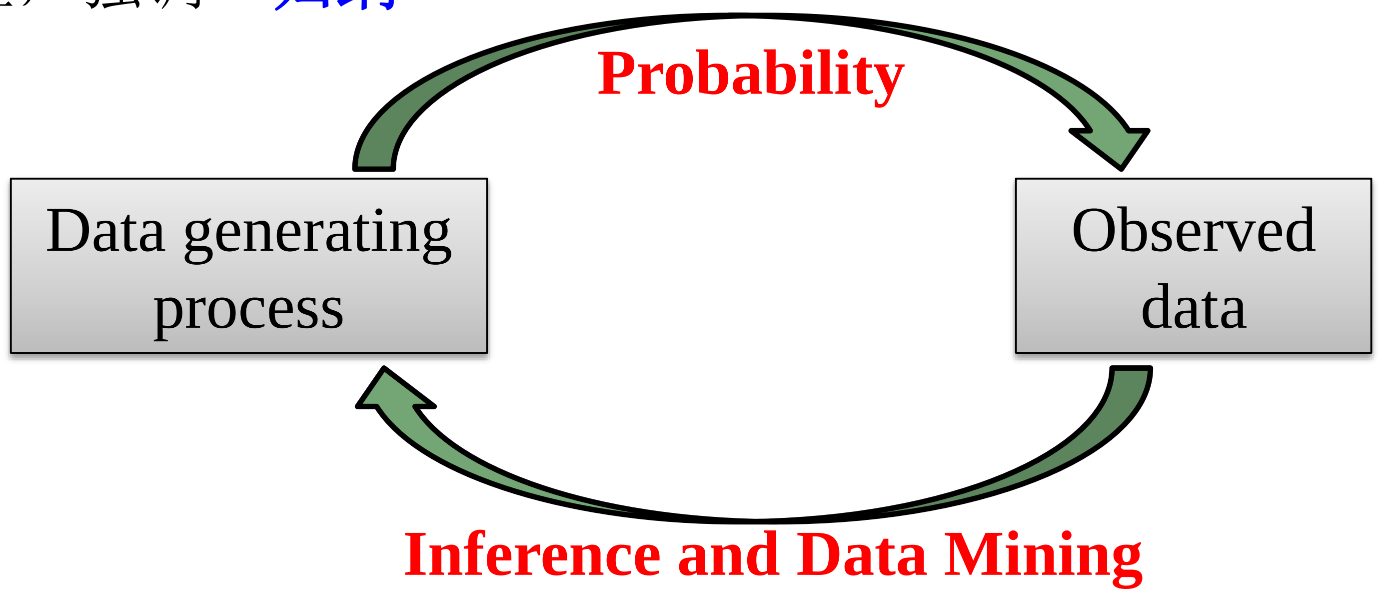
人工智能综合基础

- 概率统计

高 尉

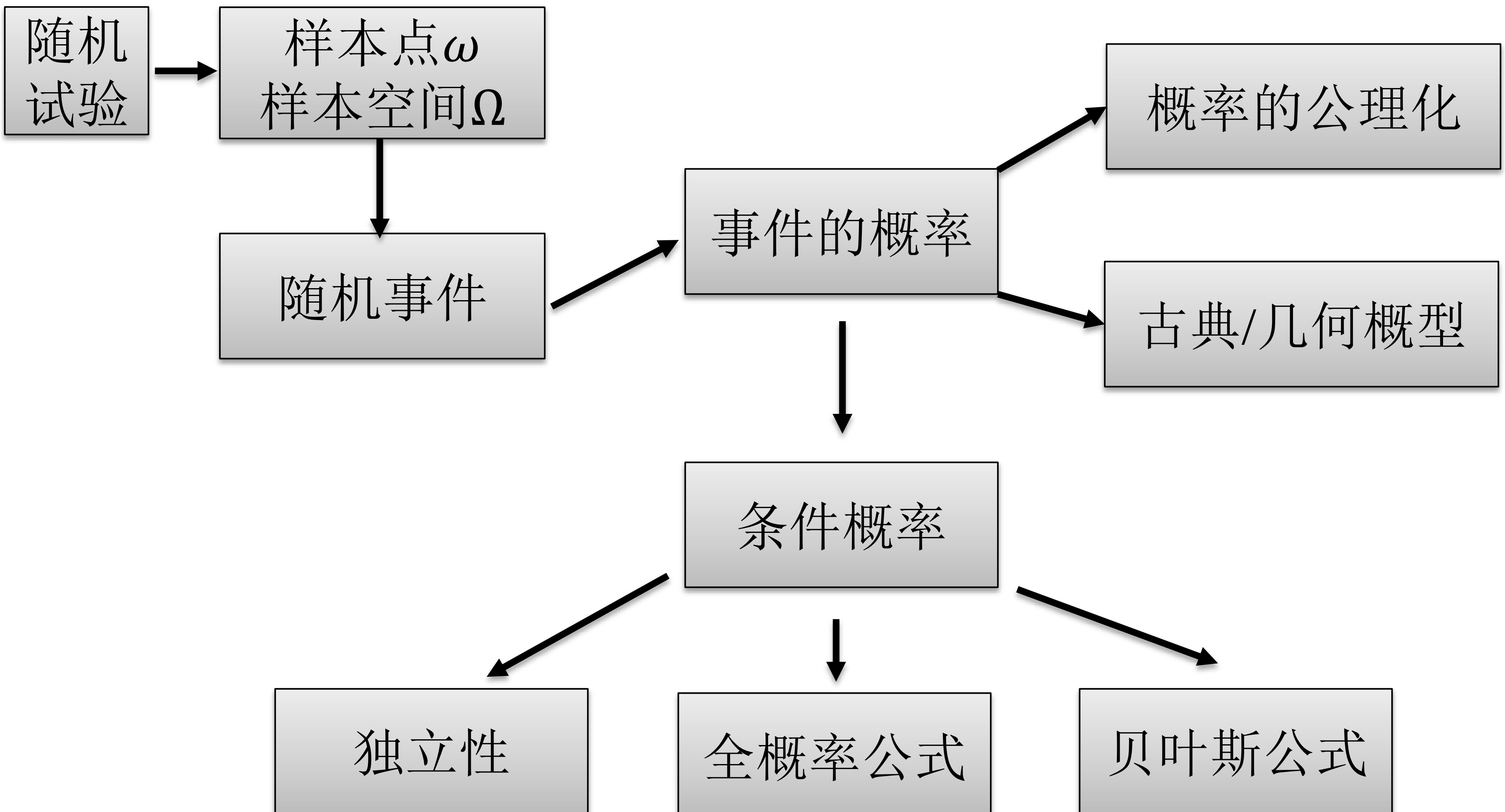
概率与统计

- **概率**：研究事件的不确定性，在给定数据生成过程中观察、研究数据的性质，强调**公理体系**、**推理**
- **统计**：收集与分析数据，根据观察的数据反思其数据生成过程，强调“**归纳**”



概率与统计重要的区别在于公理体系化

知识结构


随机
试验
样本点ω
样本空间Ω
随机事件
事件的概率
概率的公理化
古典/几何概型
条件概率
独立性
全概率公式
贝叶斯公式

必然现象与随机现象

必然现象：**在一定条件下必然发生的现象**，其特征是**条件完全决定结果**

随机现象：在一定条件下可能出现、可能不出现的现象，其特征是**条件不能完全决定结果**

随机现象的二重属性：

- 偶然性：对随机现象做一次观察，观察结果不可预知
- 必然性：对随机现象做大量观察，观察结果具有一定的规律性，即**统计规律性**

随机试验(用E表示) 具备以下三个特点的试验：

- **可重复**：可在相同的条件下重复进行
- **多结果**：结果不止一个，所有可能的结果事先已知
- **不确定**：试验前无法预测/确定哪一种结果

随机试验

- **样本点**：试验的每一种可能的结果，记为ω
- **样本空间**：试验中所有可能的结果组成的集合，记为Ω

随机事件：样本空间Ω的子集，由单个或某些样本点ω的集合，**本质是集合**，一般用字母A、B、C等。称“**随机事件$\boldsymbol{A}$发生**”当且仅当试验的结果是子集A中的元素

- **必然事件**：试验中必定发生的事件，记为 Ω
- **不可能事件**：试验中不可能发生的事件，用 $\emptyset$ 表示

记号	概率论	集合论
Ω	样本空间, 必然事件	全集
$\emptyset$	不可能事件	空集
ω	基本事件	元素
A	随机事件	子集

事件间的关系

$A \subset B$： 若A发生必然导致B发生，称事件B包含事件A

$A = B$： 若$A \supset B$且$B \supset A$

$A \cup B$： 事件A和B至少发生一个的事件

$A \cap B = AB$： 事件A和B同时发生的事件

$\bar{A}$: 事件A不发生的事件

互斥/互不相容：若事件A和B不可能同时发生

$A - B$: A发生，而B不发生的事件

$$A - B = A - AB = A\bar{B} = (A \cup B) - B$$

事件与集合的对应关系

记号	概率论	集合论
$\bar{A}$	A的对立事件	A的补集
$A \subset B$	A发生必然导致B发生	A是B的子集
$A = B$	事件A与事件B相等	集合A与集合B相等
$A \cup B$	事件A与事件B的和	集合A与集合B的并集
$A \cap B$	事件A与B的积事件	集合A与集合B的交集
$A - B$	事件A与事件B的差	集合A与集合B的差集
$A \cap B = \emptyset$	事件A与事件B互不相容	集合A与集合B中没有相同的元素

事件的运算规律

- 幂等律：$A\cup A = A$，$A\cap A = A$
- 交换律：$A\cup B = B\cup A$，$A\cap B = B\cap A$
- 结合律：$(A\cup B)\cup C = A\cup(B\cup C)$

  $(A\cap B)\cap C = A\cap(B\cap C)$
- 分配律：$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$

  $A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$
- 对偶律：$\overline{A\cup B} = \bar{A}\bar{B}$，$\overline{AB} = \bar{A}\cup\bar{B}$

上述规律可推广到多个事件

例题

设A、B、C为任意三个随机事件，证明

$$(\overline{\mathrm{A}} \cup B)(A \cup B)(\overline{\mathrm{A}} \cup \overline{\mathrm{B}})(A \cup \overline{\mathrm{B}}) = \emptyset$$

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

设A与B同时发生，则C必发生，则（）

A） $P(C) \leq \mathrm{P}(A) + P(B) - 1$

B） $P(C) \geq \mathrm{P}(A) + P(B) - 1$

C） $P(C) = \mathrm{P}(AB)$

D） $P(C) = \mathrm{P}(A \cup B)$

概率

- 概率用于度量事件发生的可能性，是事件的固有属性
- 频率在一定程度上反映了事件发生的可能性
- 概率是恒定的，而频率在试验中具有随机性
- 若试验次数足够多，频率与概率非常接近
- 概率可以通过频率来“测量”，频率是概率的一个近似

概率的公理化定义 在随机试验的样本空间Ω上，对于每一个事件A 赋予一个实数，记为$P(A)$，若满足下列条件，称$P(A)$为事件A发生的概率：

- 非负性：$P(A) \geq 0$
- 规范性：$P(\Omega) = 1$
- 可列可加性：若$A_1, A_2, \ldots$可列个两两互不相容的事件，则

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

概率的性质

- 对不可能事件∅有 $\mathbf{P(\emptyset) = 0}$, 对必然事件Ω有 $\mathbf{P(\Omega)} = 1$【事件可以推导出概率、但反之不成立】
- **有限可加性** 若$A_1, A_2 \cdots A_n$是两两不相容事件, 则
  $$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots P(A_n)$$
- 对任意事件A有$\mathrm{P}(\overline{\mathrm{A}}) = 1 - \mathrm{P}(\mathrm{A})$
- 若$\mathrm{B} \subset \mathrm{A}$，则$\mathrm{P}(\mathrm{A} - B) = \mathrm{P}(\mathrm{A}) - P(B)$和$\mathrm{P}(\mathrm{B}) \leq \mathrm{P}(\mathrm{A})$
- 对任意事件A和B
  $$\mathrm{P}(\mathrm{A} - \mathrm{B}) = \mathrm{P}(\mathrm{A}) - \mathrm{P}(\mathrm{AB}) = \mathrm{P}(A \cup B) - P(B)$$
- 容斥原理：
  $$P(A \cup B) = P(A) + P(B) - P(AB)$$
  $$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

例题

设A与B满足P(A)=P(B)=1/2，P(A∪B)=1则(　　)

A)　$A \cup B = \Omega$　　　　B)$AB = \emptyset$

C)　$\mathrm{P}(\bar{A} \cup \bar{B}) = 1$　　　　D)　$\mathrm{P}(A - B) = 0$

设A和B满足$P(AB) = \mathrm{P}(\bar{A}\bar{B})$, 且$P(A) = p$, 求$P(B)$

设$P(A) = P(B) = P(C) = 1/4$, $P(AB) = 0$, $P(AC) = P(BC) = 1/8$, 求A, B, C全不发生的概率

古典概型

古典概型：试验结果只有有限种可能$\Omega = \{\omega_1, \omega_2, \dots \omega_n\}$，每种结果发生的可能性相同$\mathrm{P}(\omega_i) = \mathrm{P}(\omega_j)$

若事件A包含k个基本事件，则事件A发生的概率为：

$$P(A) = k/n = |A|/|\Omega|$$

计数的两条基本原理：**加法原理、乘法原理**

排列、组合（组合计数十二路了解）

以前的经典例题：抽签、生日悖论、Matching问题

例题

一个盒子中装有n个球，编号为$1,2,\ldots,n$, 有放回的取出k个球，求取出的球中最大编号为m的概率。

几何概型

在一个测度有限的区域Ω内等可能性投点, 落入Ω内的任意子区域A的可能性与A的测度成正比, 与A的位置与形状无关, 这样的概率模型称之为**几何概型**. 事件A发生的概率为

$$P(A)=\frac{A\text{的测度}}{\Omega\text{的测度}}=\frac{\mu(A)}{\mu(\Omega)}$$

例：假设一乘客到达汽车站的时间是任意的，客车间隔一段时间发班, 请规划最长的间隔发车时间, 才能确保乘客候车等待时间不超过20分钟的概率大于80%.

条件概率

条件概率：设A和B为同一样本空间下的随机事件, 且$P(A) > 0$,

$$P(B|A) = P(AB)/P(A)$$

为事件A发生的条件下事件B发生的概率, 简称 **条件概率**

两种计算方法：1) 定义　　2) 空间缩减法

条件概率是概率，具有概率的所有性质：

- 非负性、规范性、可列可加性、容斥原理
- $P(B_1 - B_2|A) = P(B_1|A) - P(B_1B_2|A)$
- 乘法公式：$P(AB) = P(A)P(B|A) = P(B)P(A|B)$

全概率公式和贝叶斯公式

若事件$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 对任意事件B有

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

称之为 **全概率公式**

贝叶斯公式: 设$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 且事件B满足$P(B) > 0$. 对任意$1 \leq i \leq n$有

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

小概率原理: 若事件A在一次试验中发生的概率非常小, 但经过多次独立地重复试验, 事件A的发生是必然的

例题

事件A、B、C中A与C互不相容，$P(AB) = 1/2$，$P(C) = 1/3$，求$P(AB|\bar{C})$

例题

第一个袋子中有5个红球和3个白球，第二个袋子中有4个红球和3个白球，从第一个袋子中取3个球放入第二个袋子后，从第二个袋子任取一球，求：1）此球为红球的概率

2）若已知第二个袋子取出红球，求从第一个袋子取出3个球无红球的概率

例题

犯人a, b, c均被判为死刑, 法官随机赦免其中一人, 看守知道谁被赦免但不会说. 犯人a问看守: b和c谁会被执行死刑? 看守的策略: i) 若赦免b, 则说c; ii) 若赦免c, 则说b; iii) 若赦免a, 则以1/2的概率说b或c. 看守回答a: 犯人b会被执行死刑, 求在此信息下三人被赦免的概率?

两事件的独立性

事件A和B独立 $\Leftrightarrow P(AB)=P(A)P(B) \Leftrightarrow P(B|A)=P(B)$

A和B独立 $\Leftrightarrow$ A和$\bar{B}$独立$\Leftrightarrow$ $\bar{A}$和B独立$\Leftrightarrow$ $\bar{A}$和$\bar{B}$独立

概率为0或1的事件与任何事件独立

概念: 若事件A, B, C的相互独立与A, B, C的两两独立

性质: 若事件$P(B)\in(0,1)$，则

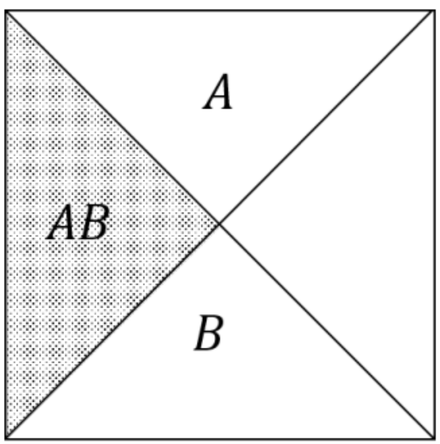
A和B独立 $\Leftrightarrow P(A|B)=P(A|\bar{B}) \Leftrightarrow P(A|B)+P(\bar{A}|\bar{B})=1$

独立与互斥的关系

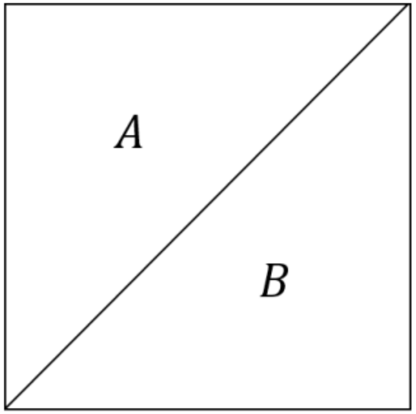
A与B相互独立： $P(AB) = P(A)P(B)$, 独立性与概率相关,
反映事件的概率属性

A与B互不相容：$AB = \emptyset$, 与事件运算关系相关, 与概率无关

独立与互斥反映事件不同的性质, 无必然联系



A与B独立, 但并不互斥



A与B互斥, 但并不独立

例题

事件A、B、C两两独立，且C与$A-B$独立, 则A、B、C相互独立.

案例分析：验证大矩阵乘法是否相等

给定矩阵$A, B, C \in \{0,1\}^{n\times n}$ $(n \geq 10000000)$, 验证$AB = C$?

独立随机产生一个向量$r \in \{0,1\}^n$, 判断

$$A(Br) = Cr?$$

计算$A(Br)$ 和Cr的复杂度均为$O(n^2)$. 若$A(Br) \neq Cr$则直接有$AB \neq C$; 若$A(Br) = Cr$并不能得出$AB = C$.

将上述过程独立进行K次, 可以证明以较大的概率有$AB = C$成立, 该过程被称为Freivalds算法

Freivalds算法分析

该算法的计算复杂度为$O(Kn^2)$, 若K比较小则显著降低了计算复杂度.

若返回No, 则必然有$AB \neq C$
若返回Yes, 然而并不一定有$AB = C$成立, 下面研究成立的概率.

设$D = AB - C \neq 0$, 则D中必存在一些元素不为0, 不妨令$d_{11} \neq 0$. 对任意一轮循环, 不妨设随机向量$r = \{r_1, r_2, \cdots, r_n\}$, 根据返回Yes可知$Dr = 0$, 进一步可得向量$Dr$的第一个元素等于0, 即

$$\sum_{j=1}^{n} d_{1j} r_j = 0 \quad \Longrightarrow \quad r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j$$

Freivalds算法分析

$$\sum_{j=1}^{n} d_{1j} r_j = 0 \quad \Longrightarrow \quad r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j$$

无论$r_2, \ldots, r_n$取何值, 等式$\sum_{j=1}^{n} d_{1j} r_j = 0$是否成立由$r_1$的值决定.

根据$P(r_1 = 0) = P(r_1 = 1) = 1/2$可知$\sum_{j=1}^{n} d_{1j} r_j = 0$成立的概率不超过1/2

在K轮独立循环中, 等式$\sum_{j=1}^{n} d_{1j} r_j = 0$成立的概率不超过$\frac{1}{2^K}$

取$K = \log_2 n$, 则算法Freivalds计算复杂度为$O(n^2 \log n)$, 若算法返回No, 则$AB \neq C$; 若返回Yes, 则有

$$P(AB = C) > 1 - 1/n$$

即至少以$1 - 1/n$的概率有$AB = C$成立