第一章 概率

Monty Hall Problem

Monty Hall Problem 源自美国电视娱乐节目，曾在全美引起了相当大的争论。最后以该节目主持人的名字将该问题命名为 **Monty Hall Problem**。

例（抽门票，续）在M个人抽N张门票的问题中，若某人第k个抽，但在此之前已知前k-1个均未抽到门票，问此时他抽到的概率是否有变化？

分析 设A：“第k个人抽到门票”

B：“前k-1个人均未抽到门票”

此时A发生的概率显然为 $\dfrac{N}{M-k+1}$，即发生了变化.

直观分析：此时的概率应与$P(B)$, $P(AB)$均有关：

$$P(B)=\frac{\binom{M-k+1}{N}}{\binom{M}{N}},\qquad P(AB)=\frac{\binom{M-k}{N-1}}{\binom{M}{N}}$$

而 $$\frac{P(AB)}{P(B)}=\frac{\binom{M-k}{N-1}}{\binom{M-k+1}{N}}=\frac{N}{M-k+1}$$

此时A发生的概率已不是原来意义的概率，记为P(A|B).

(一)条件概率

定义 设 A,B 是两个事件，且 $P(B)>0$，记

$$P(A|B)\triangleq\frac{P(AB)}{P(B)}$$

称为在事件 B 发生的条件下事件 A 发生的**条件概率**.

注

① 当 $P(B)=0$ 时，条件概率 $P(A|B)$ 毫无意义

② 条件概率 $P(A|B)$ 的直观理解：

B 发生带来的“信息”对 A 的“推断”的新认识

③ “$A|B$”不是一个事件.

④ $P(A|B)\geq P(AB)$

§5 条件概率

条件概率$P(\cdot|B)$的基本性质：设$P(B)>0$，有

- ① 非负性　对于任一事件A有$P(A|B)\geq 0$
- ② 规范性　对于必然事件Ω有$P(\Omega|B)=1$
- ③ 可列可加性　设$\{A_k\}$是两两不相容事件列，则有

$$P(\bigcup_{k=1}^{\infty}A_k|B)=\sum_{k=1}^{\infty}P(A_k|B)$$

证 ③ $$P(\bigcup_{k=1}^{\infty}A_k|B)=\frac{P\{(\bigcup_{k=1}^{\infty}A_k)\cap B\}}{P(B)}=\frac{P\{\bigcup_{k=1}^{\infty}(A_k\cap B)\}}{P(B)}$$

$\because$ $\{A_k\}$ 两两不相容 $\therefore$ $\{A_k\cap B\}$ 亦两两不相容

$$\therefore\quad P(\bigcup_{k=1}^{\infty}A_k|B)=\frac{\sum_{k=1}^{\infty}P(A_k\cap B)}{P(B)}$$

$$=\sum_{k=1}^{\infty}\frac{P(A_kB)}{P(B)}=\sum_{k=1}^{\infty}P(A_k|B)$$

条件概率 $P(\cdot|B)$ 的基本性质：设 $P(B)>0$，有

- ① 非负性　对于任一事件 A 有 $P(A|B)\geq 0$
- ② 规范性　对于必然事件 Ω 有 $P(\Omega|B)=1$
- ③ 可列可加性　设 $\{A_k\}$ 是两两不相容事件列，则有

$$P\left(\bigcup_{k=1}^{\infty} A_k \mid B\right)=\sum_{k=1}^{\infty} P(A_k \mid B)$$

问　这三条性质说明了一个什么问题？

条件概率也是一种概率

分析 $\because$ B 已发生，所以样本空间变为

$$\tilde{\Omega} \quad B\cap\Omega$$

事件域

$$\tilde{\mathcal{A}} \quad B\cap\mathcal{A}=\{B\cap A \mid A\in\mathcal{A}\}\quad(\mathcal{A}\text{ 是}\Omega\text{ 的事件域})$$

对于事件 $A\in\tilde{\mathcal{A}}$ 概率为

$$P_B(A) \quad P(A|B)$$

(二)乘法定律(公式)

由条件概率$P(A|B)=\dfrac{P(AB)}{P(B)}$ $(P(B)>0)$推得**乘法公式**

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

对称地有

乘法公式的意义?

$$P(AB)=P(B|A)P(A)$$
$$=P(A|B)P(B)\quad (P(A)>0,P(B)>0)$$

例 第一个袋中有黑、白球各2只,第二个袋中有黑、白球各3只.先从第一个袋中任取一球放入第二个袋中,再从第二个袋中任取一球.求第一、二次均取到白球的概率.

解 记$A_i=\{第\,i\,次取到白球\}$,$(i=1,2)$.则

$$P(A_1)=\frac{1}{2},\quad P(A_2|A_1)=\frac{4}{7}$$

由乘法公式求得

$$P(A_1A_2)=P(A_2|A_1)P(A_1)=\frac{1}{2}\cdot\frac{4}{7}=\frac{2}{7}$$

条件概率乘法公式的说明

① 在概率论发展初期，古典概型中的加法公式

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

及乘法公式

$$P(AB)=P(A\mid B)P(B)$$

是概率论的两条基本定理，是概率论深入发展的起点

② 一般地，若$P(A_1A_2\cdots A_{n-1})>0$，则

$$P(A_1A_2\cdots A_n)=P(A_n\mid A_1A_2\cdots A_{n-1})P(A_1A_2\cdots A_{n-1})$$

$$=\cdots=P(A_n\mid A_1A_2\cdots A_{n-1})P(A_{n-1}\mid A_1A_2\cdots A_{n-2})\cdots P(A_2\mid A_1)P(A_1)$$

③ 条件概率是定义的，但条件概率的值通常是根据实际问题中的具体意义确定的

(三)全概率定律(公式)

问题 如何将一个复杂概率计算问题分解为简单概率计算之和？

样本空间的分划：

设Ω为样本空间，若事件 $B_1, B_2, \cdots, B_n$ 满足：

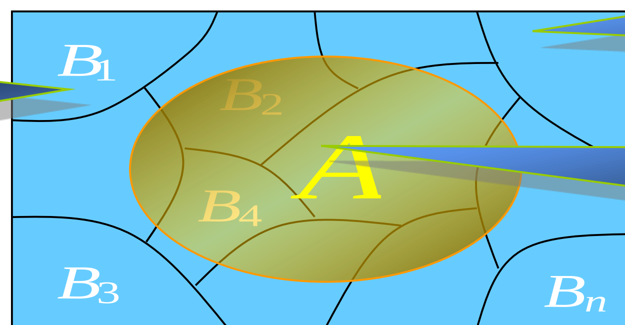
① $B_1, B_2, \cdots, B_n$ 两两不相容， 即

$$B_iB_j = \Phi \quad (i \neq j, \quad i, j = 1, 2, \cdots, n)$$

② $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$

则称 $\{B_1, B_2, \cdots, B_n\}$ 为样本空间Ω的一个**分划**.

一般要求诸 $P(B_i) > 0$



概率相当于“面积”

$P(A)$ 等于这些重叠“面积”之和

想法 将 $P(A)$ 的计算分解到 $B_1, B_2, \cdots\cdots, B_n$ 上计算然后求和.

问 怎样分解计算？

(三)全概率定律(公式)

问题 如何将一个复杂概率计算问题分解为简单概率计算之和？

样本空间的分划：

设Ω为样本空间，若事件 $B_1, B_2, \cdots, B_n$ 满足：

① $B_1, B_2, \cdots, B_n$ 两两不相容，即

$$B_iB_j = \Phi \quad (i \neq j, \quad i, j = 1, 2, \cdots, n)$$

② $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$

则称 $\{B_1, B_2, \cdots, B_n\}$ 为样本空间Ω的一个**分划**.

对任何事件 A 有

$$A = A \cap \Omega = AB_1 \cup AB_2 \cup \cdots \cup AB_n$$

$$P(A) = P(AB_1 \cup AB_2 \cup \cdots \cup AB_n)$$

$$= P(AB_1) + P(AB_2) + \cdots + P(AB_n)$$

$$= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)$$

全概率公式

(Law of Total Probability)

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

例 袋中有 a 只红球 b 只白球，先从袋中任取一球，记下颜色后放回，同时向袋中放入同颜色的球 1 只，然后再从袋中取出一球. 求第二次取到白球的概率.

如果加入 c 个同色球有什么结果？

解 记 $A=\{$第 2 次取到白球$\}$

$B_1=\{$第 1 次取到白球$\}$

$B_2=\{$第 1 次取到红球$\}$

则 B_1, B_2 是 Ω 的一个分划，由全概率公式有

$$P(A)=P(A\mid B_1)P(B_1)+P(A\mid B_2)P(B_2)$$

$$=\frac{b+1}{a+b+1}\cdot\frac{b}{a+b}+\frac{b}{a+b+1}\cdot\frac{a}{a+b}$$

$$=\frac{b}{a+b}$$

第二次取到白球的概率与第一次取到白球的概率相等，与前面放入什么颜色的球无关

例 有10个袋，其中甲袋二个，每袋中有红球、白球各2个；乙袋三个，每袋中有红球3个、白球2个；丙袋五个，每袋中有红球2个、白球3个. 从十个袋中任取一袋，再从袋中任取一球，求取到白球的概率.

解 记B_1, B_2, B_3分别表示取到甲、乙、丙袋

$$A = \{ \text{取到白球} \}$$

由全概率公式有

$$P(A) = \sum_{i=1}^{3} P(A \mid B_i) P(B_i)$$

$$= \frac{2}{10} \cdot \frac{2}{4} + \frac{3}{10} \cdot \frac{2}{5} + \frac{5}{10} \cdot \frac{3}{5}$$

$$= \frac{13}{25}$$

全概率公式是概率的加权平均

从甲、乙、丙袋取到白球的概率

全概率公式应用：敏感问题调查

常常会发生被调查者拒绝回答或不真实回答的情况.

沃纳（Warner）于1965 年提出随机回答法.

思想:“愈少泄漏问题的答案实质，愈能较好合作”.

比如为被调查者设定两个问题:

Q_1：你在期末考试中作弊了吗?

Q_2：你没在期末考试中作弊吗?

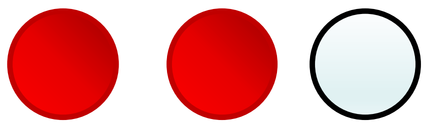
被调查者回答问题中的哪个完全由随机化的方法确定，且只有被调查者本人知道他(她)回答的是哪一个问题.

全概率公式应用：敏感问题调查

假设　　　　　　　　　　　　　　　　　　不妨令

A_1：抽中：你在期末考试中作弊了吗？　B_1：回答为“是”

A_2：抽中：你没在期末考试中作弊吗？　B_2：回答为“不是”



随机装置

被调查者在期末考试中作弊的概率 $p = 3P(B_1) - 1$

设被调查的人数为n，其中回答“是”的人数为m

则某校在期末考试中作弊人数比例的近似值 $r = \dfrac{3m}{n} - 1$

一般地，如果 $\boxed{P(A_1) = q}$，则 $r = \dfrac{1}{2q-1}[\dfrac{m}{n} - (1-q)]$

讨论？

(四) Bayes公式

设$\{B_1, B_2, \cdots, B_n\}$为样本空间$\Omega$的一个分划，且

$$P(A) > 0, P(B_i) > 0 \quad (i = 1, 2, \cdots, n)$$

则由乘法公式有

$$P(B_i \mid A)P(A) = P(A \mid B_i)P(B_i) \quad (i = 1, 2, \cdots, n)$$

由全概率公式有

$$P(A) = \sum_{j=1}^{n} P(A \mid B_j)P(B_j)$$

Bayes公式

$$\boxed{P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}}$$

$$(i = 1, 2, \cdots, n)$$

例 某工厂的一、二、三车间都生产同一产品，产量分别占总产量的15%，80%，5%，三个车间的次品率分别为2%，1%，3%，现从汇总起来的产品中任取一个，经检查是次品，判断该次品是哪个车间生产的可能性较大？

解 记 $A = \{$取到次品$\}$

$B_i = \{$取到的产品是i车间生产的$\}$，$i = 1, 2, 3$

由全概率公式有 $P(A) = \sum_{j=1}^{3} P(A \mid B_j) P(B_j) = 0.0125$

由 *Bayes* 公式有

$$P(B_1 \mid A) = \frac{P(A \mid B_1) P(B_1)}{P(A)} = \frac{0.02 \times 0.15}{0.0125} = 0.24$$

$$P(B_2 \mid A) = \frac{0.01 \times 0.80}{0.0125} = 0.64$$

$$P(B_3 \mid A) = \frac{0.03 \times 0.05}{0.0125} = 0.12$$

Bayes 推断

可见该次品是第二车间生产的可能性较大.

Bayes 公式的实际意义

假定$B_1, B_2, \cdots, B_n$为导致试验结果的“**原因**”

称 $P(B_i)\ (i = 1, 2, \cdots, n)$ 为**先验概率**

若试验产生事件 A, 则要探讨事件发生的“原因”

$$P(B_i \mid A) \quad (i = 1, 2, \cdots, n)$$

称$P(B_i \mid A)$为**后验概率**,称 $P(A \mid B_i)$为**原因概率**

① 后验概率可以通过 Bayes 公式进行计算

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)} \quad (i = 1, 2, \cdots, n)$$

② **Bayes** 方法广泛应用于网络、分类、诊断、估计、检验、判别、推理等方面

实例：计算机自动辅助诊断系统

假定$B_1, B_2, \cdots, B_n$为各种“疾病”.
应用统计方法确定先验概率：

$$P(B_i) \quad (i = 1, 2, \cdots, n)$$

应用医学知识确定原因概率：

$$P(A \mid B_i) \quad (i = 1, 2, \cdots, n)$$

对人进行观察与检查，可以确定某个指标 A，如体温、脉搏、血液中转氨酶含量等.

应用 **Bayes** 公式，计算机可计算出后验概率

$$P(B_i \mid A) \quad (i = 1, 2, \cdots, n)$$

对应于较大 $P(B_i \mid A)$的“疾病”B_i可提供给医生作进一步的临床诊断.

§5 条件概率

例（测谎试验） + / -: 测谎仪显示受测者撒谎/未撒谎

T / L: 受测者说的是真话/假话

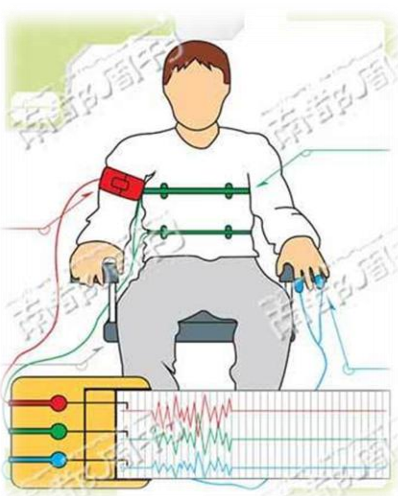
已知 $P(+|L)=0.88, P(-|T)=0.86$，又设人群中 $P(T)=0.99$，现在若有一人在测谎中显示“+”性，问测谎仪出错的概率是多少？

解 由 Bayes 公式, 此人说真话的概率为

$$P(T|+)=\frac{P(+|T)P(T)}{P(+|T)P(T)+P(+|L)P(L)}$$

$$=\frac{0.14\times 0.99}{0.14\times 0.99+0.88\times 0.01}$$

$$=0.94$$



可见，在大量群体中使用测谎仪有一定程度的危险性

例 用某种诊断法诊断癌症, 记

$$A=\{\text{判断被检验者患有癌症}\}$$

$$C=\{\text{被检验者患有癌症}\}$$

已知 $P(A\mid C)=0.95, P(\bar{A}\mid\bar{C})=0.90$, 又设人群中$P(C)=0.0004$. 现在若有一人被诊断患有癌症，问此人真正患有癌症的可能性有多大?

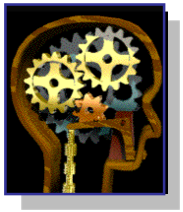
解 由 **Bayes** 公式, 此人真正患有癌症的概率为

$$P(C\mid A)=\frac{P(A\mid C)P(C)}{P(A\mid C)P(C)+P(A\mid\bar{C})P(\bar{C})}$$

$$=\frac{0.95\times0.0004}{0.95\times0.0004+0.1\times0.9996}$$

$$=0.0038$$

可见，虽然检验法相当可靠，但被诊断患有癌症而真正患有癌症的可能性并不大



课后作业

P22：46, 53, 54, 63

补充题：

请用本节所讨论的工具给出**Monty Hall**问题（即：三扇门问题）的解答。

Important Conclusion:

When dealing with conditional probabilities:
don't trust your intuition, do computations!

2. 据以往资料表明,某一 3 口之家,患某种传染病的概率有以下规律:

$$P\{孩子得病\} = 0.6,$$

$$P\{母亲得病|孩子得病\} = 0.5,$$

$$P\{父亲得病|母亲及孩子得病\} = 0.4.$$

求母亲及孩子得病但父亲未得病的概率.

3. 对以往数据分析结果表明,当机器调整得良好时，产品的合格率为 0.98；而当机器发生某种故障时，产品的合格率为 0.55．每天早上机器开动时，机器调整良好的概率为 0.95．试求:已知某日早上的第一件产品是合格品时，机器调整得良好的概率.