高中统计学 · 第壹课练习

何濯羽

2024 年 2 月 11 日

1 组合数的性质证明

请证明以下组合数的性质，其中 $n, m \in \mathbb{N}$ 且 $n \geq m$：

1) 对称性：$\mathrm{C}_n^m = \mathrm{C}_n^{n-m}$；
2) 递推性：$\mathrm{C}_n^m = \mathrm{C}_{n-1}^{m-1} + \mathrm{C}_{n-1}^m$。

2 条件概率的计算

Surflight 是一家采用高端材料制造冲浪板的公司，主要通过其官网销售冲浪板。该公司拥有浏览其官网界面的访问者的详细记录，包括他们是否通过浏览器上的广告进入其官网以及是否购买了冲浪板的信息。我们用 A 表示任一网民看到浏览器上的广告这一事件，用 B 表示任一网民购买冲浪板这一事件。有关这些事件的概率记录在下表中：

	B	B^c	总和
A	0.10	0.55	0.65
A^c	0.15	0.20	0.35
总和	0.25	0.75	

1) 请计算 $Pr(A|B)$、$Pr(A|B^c)$、$Pr(A \cup B)$。
2) 一个看见浏览器上广告的客户购买了冲浪板的概率是多少?
3) 一个购买了冲浪板的客户没有看过浏览器上广告的概率是多少?

3 反直觉的条件概率

人类免疫缺陷病毒（human immunodeficiency virus，简称 HIV），也称为“艾滋病毒”，自 1981 年在美国被识别并在之后发展成为全球大流行疾病，时至今日仍未被人类完全克服。但是，艾滋病毒检测技术已经十分完备，其准确度相当惊人：如果某人真的是 HIV 阳性，那么现代血液检测的技术有 99.9% 的把握将他的阳性检测出来；如果某人真的是 HIV 阴性，那么检测技术的精确度达到 99.99%，也就是说，只有 0.01% 的概率会将一个 HIV 阴性者误诊为阳性。

那么问题来了。已知一般在人群中 HIV 真阳性者的比例为 0.01%。假设我们随机在街头选择一个人进行 HIV 检测，检测结果为 HIV 阳性。请问，这个人真的是 HIV 阳性的概率为多少？

4 贝叶斯法则与神奇宝贝【拓展】

大木博士打算给阿呆分配一只神奇宝贝搭档，可能是皮卡丘、小火龙或者妙蛙种子。大木博士只让拱猪知道了他的决定。因为阿呆非常喜欢皮卡丘，所以他非常想知道皮卡丘是否被大木博士选中了，于是他去询问拱猪。

拱猪说：“我不能告诉你任何关于皮卡丘的信息。我只能告诉你，妙蛙种子不会是你的搭档。”已知拱猪非常诚实，不会说谎。

拱猪透露的信息对阿呆有用吗？请用数学证明。

答案

组合数的性质证明

对称性:

$$\mathsf{C}_n^{n-m} = \frac{\mathsf{A}_n^{n-m}}{\mathsf{A}_{n-m}^{n-m}} = \frac{n!}{[n-(n-m)]! \cdot (n-m)!} = \frac{n!}{m! \cdot (n-m)!} = \mathsf{C}_n^m$$

递推性:

$$\begin{aligned}
\mathsf{C}_{n-1}^{m-1} + \mathsf{C}_{n-1}^{m} &= \frac{(n-1)!}{[(n-1)-(m-1)]! \cdot (m-1)!} + \frac{(n-1)!}{[(n-1)-m]! \cdot m!} \\
&= \frac{(n-1)!}{(n-m)! \cdot (m-1)!} + \frac{(n-1)!}{(n-m-1)! \cdot m!} \\
&= \frac{n!}{(n-m)! \cdot m!} \cdot \frac{m}{n} + \frac{n!}{(n-m)! \cdot m!} \cdot \frac{n-m}{n} \\
&= \frac{n!}{(n-m)! \cdot m!} \cdot \left(\frac{m}{n} + \frac{n-m}{n}\right) \\
&= \frac{n!}{(n-m)! \cdot m!} \\
&= \mathsf{C}_n^m
\end{aligned}$$

条件概率的计算

(1)

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{0.10}{0.25} = \frac{2}{5}$$

$$Pr(A|B^c) = \frac{Pr(A \cap B^c)}{Pr(B^c)} = \frac{0.55}{0.75} = \frac{11}{15}$$

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) = 0.65 + 0.25 - 0.10 = 0.8$$

(2)

$$Pr(B|A) = \frac{Pr(B \cap A)}{Pr(A)} = \frac{0.10}{0.65} = \frac{2}{13}$$

(3)

$$Pr(A^c|B) = \frac{Pr(A^c \cap B)}{Pr(B)} = \frac{0.15}{0.25} = \frac{3}{5}$$

反直觉的条件概率

我们可以用 Y 表示“某人是真 HIV 阳性”，用 T 表示“某人检测出 HIV 阳性”。根据题意，我们有

$$Pr(Y) = 0.01\%, \quad Pr(T|Y) = 99.9\%, \quad Pr(T|Y^c) = 0.01\%$$

根据贝叶斯法则，我们有

$$\begin{aligned}
Pr(Y|T) &= \frac{Pr(T|Y) \cdot Pr(Y)}{Pr(T)} \\
&= \frac{Pr(T|Y) \cdot Pr(Y)}{Pr(T|Y) \cdot Pr(Y) + Pr(T|Y^c) \cdot Pr(Y^c)} \\
&= \frac{99.9\% \times 0.01\%}{99.9\% \times 0.01\% + 0.01\% \times (1 - 0.01\%)} \\
&\approx 49.98\%
\end{aligned}$$

这个答案或许与我们的直觉很不一致。即使是在这么惊人的检测精确度下，一个人被检测为 HIV 阳性，而他真的是 HIV 阳性的概率也接近 50%。

我们可以反过来理解这个结果：如此罕见的疾病（人群中只有万分之一的人被感染），一旦被检测为阳性，就会有大约 50% 的概率是真的被感染了。

贝叶斯法则与神奇宝贝

我们可以用 A、B、C 分别表示“皮卡丘被选中了”、“小火龙被选中了”、“妙蛙种子被选中了”这三个事件，用 $NotB$ 和 $NotC$ 分别表示“拱猪说小火龙没被选中”和“拱猪说妙蛙种子没被选中”这两个事件。

根据题意，我们有

$$Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$$

根据贝叶斯法则，我们有

$$Pr(A|NotC) = \frac{Pr(NotC|A) \cdot Pr(A)}{Pr(NotC|A) \cdot Pr(A) + Pr(NotC|B) \cdot Pr(B) + Pr(NotC|C) \cdot Pr(C)}$$

如果皮卡丘被大木博士选中，那么拱猪会说“小火龙没被选中”或者“妙蛙种子没被选中”。因此，

$$Pr(NotB|A) = Pr(NotC|A) = \frac{1}{2}$$

因为拱猪不会说谎，所以

$$Pr(NotC|B) = 1, \quad Pr(NotC|C) = 0$$

于是，

$$Pr(A|NotC) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{1}{3} = Pr(A)$$

我们证得拱猪的信息是无效信息。