详细解答

STEP I 2005，第 12 题

概率与统计 · 温氏图与界限

STEP I 2005，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/05-s1-q12

关键思路

凡是问“哪个范围的数值才不会自相矛盾?”的题目，其实都是关于一幅温氏图（Venn diagram）：图中各个区域的值未知，但必须全部非负。先用尽量少的自由未知数表示每一个区域；这样，每个“此区域 ≥ 0”的条件就变成一个界限，把所有界限合起来便得到所求范围。在第 (ii) 部分，我们用一个计数论证（把三个单项概率相加时，每名巫师被数了多少次），毋须知道个别区域的值，便可求出“恰好两件”的概率。

解答

第 (i) 部分

设 S 为哈比人抽烟斗的事件（event），H 为他戴帽子的事件。令 $t = \mathrm{P}(S \cap H)$。温氏图的四个区域是

$$\mathrm{P}(S \cap H) = t, \qquad \mathrm{P}(S \cap H') = 0.7 - t,$$
$$\mathrm{P}(S' \cap H) = 0.4 - t, \qquad \mathrm{P}(S' \cap H') = 1 - (0.7 + 0.4 - t) = t - 0.1.$$

只要所选的 t 令四者皆为非负，就得到一个真正的概率分布（probability distribution），反之亦然。真正起约束作用的条件是 $t \geq 0.1$（来自最后一个区域）和 $t \leq 0.4$（来自第三个区域）；其余条件随之成立。所以 $0.1 \leq t \leq 0.4$，而由于 $p = 0.7 - t$，

$$0.3 \leq p \leq 0.6.$$

两个端点都可以取到：当抽烟斗而又戴帽子的哈比人尽量少时（每个哈比人至少做两件事的其中一件），$p = 0.6$；当每个戴帽子的哈比人都抽烟斗时，$p = 0.3$。

第 (ii) 部分：恰好两件

设 H、C、R 分别为巫师戴帽子、披斗篷和戴戒指的事件。设 a、b 分别为巫师在这三件物品中恰好穿戴一件和恰好穿戴两件的概率。已知一件也没有的概率是 0.05，三件齐备的概率是 0.1，所以

$$0.05 + a + b + 0.1 = 1 \quad \Longrightarrow \quad a + b = 0.85.$$

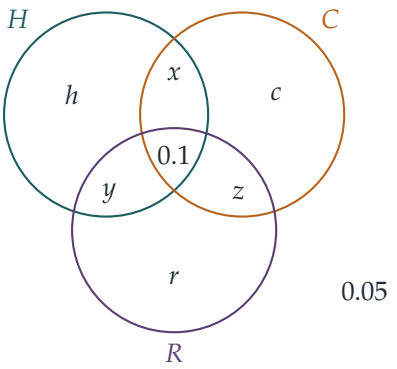
现在把三个已知概率相加：$\mathrm{P}(H) + \mathrm{P}(C) + \mathrm{P}(R) = 0.7 + 0.8 + 0.4 = 1.9$。在这个和之中，恰好穿戴 k 件物品的巫师被数了 k 次，所以

$$a + 2b + 3(0.1) = 1.9 \quad \Longrightarrow \quad a + 2b = 1.6.$$

两式相减，得 $b = 0.75$（而 $a = 0.1$）。巫师恰好穿戴其中两件物品的概率是 $\frac{3}{4}$。

第 (ii) 部分：q 的范围

如图所示，为圆内的七个区域标上记号：h、c、r 代表恰好一件，x、y、z 代表恰好两件，中央是 0.1，圆外是 0.05。



由 H、C、R 各自的总和，以及第 (ii) 部分至今所得的结果，

$$h + x + y = 0.6, \qquad c + x + z = 0.7,$$
$$r + y + z = 0.3, \qquad x + y + z = 0.75.$$

利用最后一条方程，从首三条方程中各消去一对未知数，

$$h = z - 0.15, \qquad c = y - 0.05, \qquad r = x - 0.45.$$

（验算一下：$h + c + r = 0.75 - 0.65 = 0.1 = a$。）所以整个分布完全由满足 $x + y + z = 0.75$ 的 x、y、z 决定，而它是一个真正的分布，当且仅当六个区域全部非负：

$$x \geq 0.45, \qquad y \geq 0.05, \qquad z \geq 0.15.$$

想一想 x 最大可以是多少?

由于 $x = 0.75 - y - z$，而 $y + z \geq 0.05 + 0.15 = 0.2$，所以 $x \leq 0.55$，等号在 $y = 0.05$、$z = 0.15$ 时成立。最小值 $x = 0.45$ 也可以取到，例如取 $y = 0.1$、$z = 0.2$。两者之间的每一个值同样可以取到，因为我们可以连续地改变 y 和 z。因此

$$0.45 \leq x \leq 0.55.$$

最后，披着斗篷时戴帽子但不戴戒指，正好就是区域 x，所以

$$q = \mathrm{P}(H \cap R' \mid C) = \frac{\mathrm{P}(H \cap C \cap R')}{\mathrm{P}(C)} = \frac{x}{0.8}, \qquad \text{由此得} \qquad \frac{9}{16} \leq q \leq \frac{11}{16}.$$

验算

- 在极端情况 $x = 0.55$、$y = 0.05$、$z = 0.15$ 下，我们得到 $h = 0$、$c = 0$、$r = 0.1$；八个区域之和为 1，并给出 $\mathrm{P}(H) = 0.7$、$\mathrm{P}(C) = 0.8$、$\mathrm{P}(R) = 0.4$，符合要求。
- 我们以 0.01 为步长，列举了八个区域所有与数据相符的赋值：每一个都满足 $x + y + z = 0.75$，而 q 的取值由 0.5625 至 0.6875。同样地，在 (i) 中，p 的取值由 0.3 至 0.6。

常见错误

- 在 (i) 中，忘记了“两者皆非”这个区域，因而得到下界 $t \geq 0$ 而不是 $t \geq 0.1$，结果得出错误的上限 $p \leq 0.7$。
- 在 (ii) 中，除以 1 而不是除以 $\mathrm{P}(C) = 0.8$：“已知他披着斗篷”这个条件会改变分母。
- 把 H、C、R 当作独立（independent）事件。题目从没有说它们独立，而事实上 $\mathrm{P}(H \cap C \cap R) = 0.1 \neq 0.7 \times 0.8 \times 0.4$。

STEP 之外

某个量是未知概率的线性函数，而这些概率须满足若干线性方程及非负条件，要求这个量的取值范围，其实就是一个小型的线性规划（linear programme）问题：极值总是出现在可行域（feasible region）的角点，这正是我们的极值都来自把某些区域设为零的原因。同样的想法可得出第 (i) 部分所用的弗雷歇（Fréchet）界限 $\max(0, \mathrm{P}(A) + \mathrm{P}(B) - 1) \leq \mathrm{P}(A \cap B) \leq \min(\mathrm{P}(A), \mathrm{P}(B))$。想进一步了解容斥原理（inclusion–exclusion）和条件概率（conditional probability），可阅读 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。