详细解答

STEP I 2010，第 12 题

概率与统计 · 以尾和求期望值

STEP I 2010，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/10-s1-q12

关键思路

尾和公式（tail-sum formula）的来源，是把 $n\,\mathrm{P}(X=n)$ 写成 $\mathrm{P}(X=n)$ 相加 n 次，然后把各项按另一种次序重新归并。就企鹅这个问题而言，事件 $X \geq n$ 比 $X = n$ 容易描述得多：开了 $n-1$ 个盒之后仍须再开，当且仅当这 $n-1$ 个盒里全是同一种企鹅。因此尾概率（tail probability）十分简单，而刚证明的公式就把 $\mathrm{E}(X)$ 化成两个等比级数（geometric series）。

解答

定义与尾和公式

由于 X 只取 $1, 2, 3, \ldots$ 这些值，

$$\mathrm{E}(X) = \sum_{m=1}^{\infty} m\,\mathrm{P}(X=m),$$

前提是这个级数收敛（converge）。

对每个 $n \geq 1$，事件 $X \geq n$ 是各事件 $X = m$（$m \geq n$）的不相交并集（disjoint union），所以 $\mathrm{P}(X \geq n) = \sum_{m \geq n} \mathrm{P}(X=m)$。因此

$$\begin{aligned}\sum_{n=1}^{\infty} \mathrm{P}(X \geq n) &= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \mathrm{P}(X=m) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{m} \mathrm{P}(X=m) \\ &= \sum_{m=1}^{\infty} m\,\mathrm{P}(X=m) = \mathrm{E}(X).\end{aligned}$$

想一想 为甚么可以对调两个求和的次序?

这个二重和是把 $\mathrm{P}(X=m)$ 对所有满足 $1 \le n \le m$ 的数对 (n,m) 相加。第一行是固定 n，对 m 求和；第二行是固定 m，对 $n=1,\ldots,m$ 这 m 个值求和。每一项都是非负的，而对于非负项级数，总和与求和的次序无关（两边都可能是无穷大，若然则两边同为无穷大）。不妨把各项想像成排成一个三角形，第 m 行有 m 个 $\mathrm{P}(X=m)$：逐行相加得 $\mathrm{E}(X)$，逐列相加得尾和。公式得证。

尾概率

按题目原意，假设不同盒子的内容互相独立（independent）。设 D_k 为第 k 个盒内是企鹅爸爸的事件，则 $\mathrm{P}(D_k)=p$，$\mathrm{P}(D_k')=q$。

取 $n \ge 2$。$X \ge n$ 成立，当且仅当首 $n-1$ 个盒尚未凑齐两种企鹅，即它们全是爸爸或全是妈妈。这两种情况互不相交，故由独立性得

$$\mathrm{P}(X \ge n) = p^{n-1} + q^{n-1} \qquad (n \ge 2).$$

特别地，取 $n=4$，

$$\mathrm{P}(X \ge 4) = p^3 + q^3,$$

正是所求。此外，$X \ge 2$ 必然成立，因为一个盒不可能同时装有两种企鹅，所以 $\mathrm{P}(X \ge 1) = \mathrm{P}(X \ge 2) = 1$。（上面的公式在 $n=2$ 时给出 $p+q=1$，与此相符；但在 $n=1$ 时却给出 2，所以 $n=1$ 这一项必须另行处理。）

期望值

由尾和公式，

$$\begin{aligned}\mathrm{E}(X) &= \mathrm{P}(X \ge 1) + \sum_{n=2}^{\infty} (p^{n-1} + q^{n-1}) \\ &= 1 + \sum_{k=1}^{\infty} p^k + \sum_{k=1}^{\infty} q^k \\ &= 1 + \frac{p}{1-p} + \frac{q}{1-q},\end{aligned}$$

其中两个等比级数都收敛，因为 $0<p<1$ 及 $0<q<1$（这里正是用到 $p \ne 0$、$q \ne 0$ 的地方）。由于 $1-p=q$ 及 $1-q=p$，

$$\mathrm{E}(X) = 1 + \frac{p}{q} + \frac{q}{p} = 1 + \frac{p^2+q^2}{pq}.$$

而 $p^2+q^2 = (p+q)^2 - 2pq = 1-2pq$，所以

$$\mathrm{E}(X) = 1 + \frac{1-2pq}{pq} = 1 + \frac{1}{pq} - 2 = \frac{1}{pq} - 1,$$

正是所求。

最后一部分

对正数 p 和 q 应用算术平均—几何平均不等式（AM–GM inequality），

$$\sqrt{pq} \le \frac{p+q}{2} = \frac{1}{2}, \qquad \text{所以} \qquad 0 < pq \le \frac{1}{4}.$$

想一想 取倒数之后，不等号朝哪个方向?

由于 $pq > 0$，取倒数会令不等号反向：$\dfrac{1}{pq} \ge 4$。因此

$$\mathrm{E}(X) = \frac{1}{pq} - 1 \ge 4 - 1 = 3,$$

等号成立当且仅当 $p = q = \frac{1}{2}$。（等价地，$pq = p(1-p) = \frac{1}{4} - (p - \frac{1}{2})^2 \le \frac{1}{4}$。）

验算

- 取 $p = q = \frac{1}{2}$：$X - 1$ 是等到第一个与第 1 盒不同的盒子所需的次数，是一个平均值为 2 的几何变量（geometric variable），所以 $\mathrm{E}(X) = 3$，与 $\frac{1}{1/4} - 1$ 吻合。
- 当 $p \to 0$ 时，$\mathrm{E}(X) \to \infty$：企鹅爸爸十分罕有，所以要等很久。这与 $\frac{1}{pq} - 1$ 趋向无穷大相符。
- 直接计算，对 $n \ge 2$ 有 $\mathrm{P}(X = n) = p^{n-1}q + q^{n-1}p$。取 $p = \frac{1}{3}$，以数值方法把 $n\,\mathrm{P}(X = n)$ 相加得 3.5；另外模拟 200 000 名收集者（取 $p = 0.3$），所得平均值为 3.761，而准确值是 $\frac{1}{0.21} - 1 \approx 3.762$。

常见错误

- 在 $n = 1$ 时也使用 $p^{n-1} + q^{n-1}$，结果得出一个等于 2 的“概率”，期望值亦因而多了 1。
- 写成 $\mathrm{P}(X \ge 4) = p^4 + q^4$：只要首三个盒相同，就需要开第四个盒。
- 在最后一部分忘记了正数之间的不等式取倒数后会反向。

STEP 之外

尾和公式有一个连续版本的孪生公式：对于非负随机变量（random variable），$\mathrm{E}(X) = \int_0^\infty \mathrm{P}(X > x)\,\mathrm{d}x$,证明方法同样是对调次序(在一般情况下,这一步由托内利定理(Tonelli's

theorem）保证成立）。企鹅问题是赠券收集问题（coupon collector's problem）中只有两种类型的情况；若有 N 种机会均等的类型，所需盒数的期望值是 $N(1+\frac{1}{2}+\cdots+\frac{1}{N})$，约为 $N\ln N$。这两个想法都见于 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门），以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）。

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。