详细解答

STEP III 1995，第 13 题

概率与统计 · 有噪声的二元信道

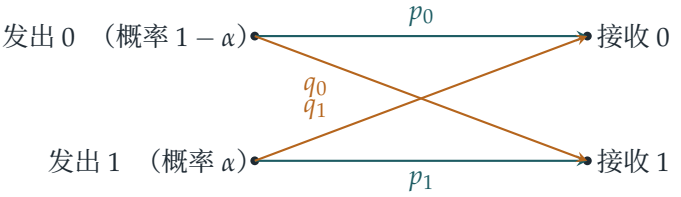
> **STEP III 1995，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/95-s3-q13

关键思路

一切都由逐个数字考虑而来。每个数字都经过一条“二元信道”（binary channel）：它以 0 或 1 发出，然后要么原样保留，要么被翻转。一旦弄清楚单一数字的情况，整段信息不过是许多独立的复本，所以概率相乘即可。至于增强站，窍门在于留意到：一个数字经过两段电缆后正确到达，当且仅当它被翻转了偶数次。

解答

全篇以 $N = 10^k$ 表示信息的长度。我们按照题目的原意，假设原信息的各个数字互相独立（independent）（每个数字是 1 的概率为 α），并假设电缆对每个数字的处理与其他数字无关。单一数字的情况如下图所示。



整段信息正确接收

设 C 为某个指定数字被正确接收的事件。这个数字要么是保留下来的 0，要么是保留下来的 1，所以

$$\mathrm{P}(C) = (1-\alpha)p_0 + \alpha p_1.$$

由于 N 个数字互相独立，

$$\mathrm{P}(\text{整段信息正确}) = \left[(1-\alpha)p_0 + \alpha p_1\right]^{10^k}.$$

概率 β

一个数字被接收为 1，要么它原本是 1 而且保留下来，要么它原本是 0 而被翻转：

$$\beta = \alpha p_1 + (1-\alpha)q_0.$$

现在

$$\begin{aligned}\beta = \alpha &\iff \alpha p_1 + (1-\alpha)q_0 = \alpha \iff (1-\alpha)q_0 = \alpha(1-p_1) = \alpha q_1 \\ &\iff q_0 = \alpha(q_0+q_1) \iff \alpha = \frac{q_0}{q_0+q_1},\end{aligned}$$

其中最后一步除以 q_0+q_1，而除非电缆完美无误（$q_0 = q_1 = 0$，此时 $\beta = \alpha$ 显然成立），否则这个数必为正数。每一步都可以逆推，所以这就是所要求的“当且仅当”（if and only if）。用文字来说：电缆不改变 1 所占的比例，恰好就是 $0 \to 1$ 的流量（即 $(1-\alpha)q_0$）与 $1 \to 0$ 的流量（即 αq_1）互相抵销之时。

全是零的信息

假设上述条件成立，而且接收到的每个数字都是 0。由于各个位置上的（发出数字，接收数字）这些数对互相独立，整段信息正确的条件概率（conditional probability）就是各个单一数字的概率 $\mathrm{P}(\text{发出0} \mid \text{接收0})$ 的乘积。

> ? **想一想** 怎样把“接收到 0”转化为一句关于发出了甚么的陈述?

用贝叶斯定理（Bayes' theorem）。一个数字被接收为 0 的概率是 $1-\beta$，而条件说明 $\beta = \alpha$，所以 $1-\beta = 1-\alpha$。因此

$$\mathrm{P}(\text{发出0} \mid \text{接收0}) = \frac{(1-\alpha)p_0}{1-\beta} = \frac{(1-\alpha)p_0}{1-\alpha} = p_0,$$

于是

$$\mathrm{P}(\text{信息正确} \mid \text{接收到的全是零}) = p_0^{10^k}.$$

一半一半的机会

当 $q_0 = q_1 = q$ 且 $\alpha = \frac{1}{2}$ 时，每个数字正确的概率是 $\frac{1}{2}(1-q) + \frac{1}{2}(1-q) = 1-q$。对于一段有 10^6 个数字的信息，我们需要

$$(1-q)^{10^6} = \tfrac{1}{2} \quad\Longrightarrow\quad 10^6 \ln(1-q) = -\ln 2.$$

由于 q 极小，$\ln(1-q)\approx -q$，得出

$$q \approx \frac{\ln 2}{10^6} \approx 6.93\times 10^{-7}.$$

加设增强站

长度减半会使 q 除以 $2^2=4$，因为 q 与长度的平方成正比。所以每半段电缆翻转一个数字的概率是 $q/4$，而且与另一半段无关（增强站只是把收到的东西原样再发出）。

> **想一想** “两半段都正确传送”是数字正确到达的唯一途径吗?

不是：一个在两半段都被翻转的数字会被翻回原状。所以

$$\mathrm{P}(\text{数字正确}) = \left(1-\frac{q}{4}\right)^2 + \left(\frac{q}{4}\right)^2 = 1-\frac{q}{2}+\frac{q^2}{8} \approx 1-\frac{q}{2},$$

这里略去了 q^2 阶的项。（在这里，双重翻转正是这类项之一，所以它对答案没有影响，但明白个中原因仍然值得。）整段信息正确的概率现在是

$$\left(1-\frac{q}{2}\right)^{10^6}.$$

在同一阶的近似下，$(1-\frac{q}{2})^2 \approx 1-q$，所以

$$\left(1-\frac{q}{2}\right)^{10^6} \approx (1-q)^{10^6/2} = \left[(1-q)^{10^6}\right]^{1/2} = \left(\tfrac{1}{2}\right)^{1/2} = \frac{1}{\sqrt{2}} \approx 0.707.$$

所以增强站把信息完全正确的机会由 $\frac{1}{2}$ 提高到约 0.71。留意我们在这里根本用不着 q 的数值：原来的条件 $(1-q)^{10^6}=\frac{1}{2}$ 已经包办了一切。同样的想法配合 $e^{-10^6 q}=\frac{1}{2}$，可得 $e^{-10^6 q/2}=1/\sqrt{2}$。

验算

- 把长度为 4 的所有发出信息和接收信息逐一准确列举（取 $p_0=0.9$，$p_1=0.8$），可以重现 $[(1-\alpha)p_0+\alpha p_1]^4$，以及在 α 满足该条件时的条件概率 $p_0^4=0.6561$。
- 准确解出 $(1-q)^{10^6}=\frac{1}{2}$ 可得 $q=6.9315\times 10^{-7}$，与 $\ln 2/10^6$ 吻合至五位有效数字。
- 取这个 q 值，增强站情况下的准确概率 $\left[(1-\frac{q}{4})^2+(\frac{q}{4})^2\right]^{10^6}$ 是 0.707107，与 $1/\sqrt{2}$ 相符。
- 合理性检查：若 $q_0=q_1$，条件给出 $\alpha=\frac{1}{2}$，正如对称性所要求。

常见错误

- 把每半段的出错概率当作 $q/2$。出错概率与长度的平方成正比，所以应是 $q/4$。

- 在“全是零”的部分只答 $(1-\alpha)^N$ 或单单一个 p_0：我们需要的是每个数字的条件概率，再取 10^k 次方。

STEP 之外

这就是信息论（information theory）中的二元信道。当 $q_0 = q_1$ 时，它就是二元对称信道（binary symmetric channel），是香农（Shannon）理论的起点——这套理论研究有噪声的信道能够传送多少信息；它也是纠错码（error-correcting codes）的起点，纠错码加入冗余，使偶尔出现的翻转可以被侦测和修复，而不仅仅是变得较为罕见。增强站的计算也是一条两步的马可夫链（Markov chain）：翻转概率的复合，就是把转移矩阵（transition matrices）相乘。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）以恰到好处的程度讲解条件概率和马可夫链。

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。