详细解答

STEP II 1991，第 15 题

概率与统计 · 期望值的线性性质

> **STEP II 1991，第 15 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/91-s2-q15

关键思路

把数对 (n_1, n_2) 看成 $m \times m$ 方格表中的一格，全部 m^2 格出现的机会均等。满足 $|n_1 - n_2| = k$ 的格子位于与主对角线平行的两条对角线上，所以很容易数出来。之后的一切都来自*期望值的线性性质*（*linearity of expectation*）：若干个差之和的期望值，等于各个期望值之和，即使相邻的项共用同一个变量、并非独立（independent）亦然。监考员（invigilator）所走的路程也是同一类的和，只是第一段路由一张固定的书桌出发，而不是由随机的一张出发。

解答

$|n_1 - n_2|$ 的分布

由于 n_1 与 n_2 独立，且各自在 $\{1, \ldots, m\}$ 上均匀分布（uniform），m^2 个有序数对 (n_1, n_2) 中每一个的概率都是 $1/m^2$。

固定 k，其中 $1 \le k \le m-1$。满足 $n_1 - n_2 = k$ 的数对是 $(k+1, 1), (k+2, 2), \ldots, (m, m-k)$，共 $m-k$ 对。由对称性，满足 $n_2 - n_1 = k$ 的数对另有 $m-k$ 对，而因为 $k \neq 0$，这两组数对并不重叠。因此

$$\mathrm{P}(|n_1 - n_2| = k) = \frac{2(m-k)}{m^2},$$

正如所求。（当 $k = 0$ 时有 m 对，所以 $\mathrm{P}(|n_1 - n_2| = 0) = 1/m$。作为验算，$\frac{1}{m} + \sum_{k=1}^{m-1} \frac{2(m-k)}{m^2} = \frac{1}{m} + \frac{m(m-1)}{m^2} = 1$。）

$|n_1 - n_2|$ 的期望值

$k = 0$ 这个值对期望值没有贡献，所以

$$\begin{aligned}\mathrm{E}|n_1 - n_2| &= \sum_{k=1}^{m-1} k \cdot \frac{2(m-k)}{m^2} = \frac{2}{m^2}\left(m\sum_{k=1}^{m-1} k - \sum_{k=1}^{m-1} k^2\right)\\ &= \frac{2}{m^2}\left(\frac{m\cdot m(m-1)}{2} - \frac{(m-1)m(2m-1)}{6}\right)\\ &= \frac{2}{m^2}\cdot\frac{m(m-1)(3m-(2m-1))}{6} = \frac{(m-1)(m+1)}{3m} = \frac{m^2-1}{3m},\end{aligned}$$

当中用了标准求和公式 $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$ 和 $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$，并取 $n = m-1$。

$m = 2$ 而有三个整数的情况

当 $m = 2$ 时，共有 $2^3 = 8$ 个机会均等的三元组 (n_1, n_2, n_3)。记 $S = |n_1 - n_2| + |n_2 - n_3|$：

(n_1, n_2, n_3)	111	112	121	122	211	212	221	222
S	0	1	2	1	1	2	1	0

总和是 8，所以 $\mathrm{E}(S) = 8/8 = 1$。题目所述的值是 $2(m^2-1)/(3m) = 2\cdot 3/6 = 1$，两者吻合。

> **想一想** $|n_1 - n_2|$ 和 $|n_2 - n_3|$ 两项都涉及 n_2。我们还可以直接把它们的期望值相加吗？

可以。期望值的线性性质 $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ 完全不需要独立性。每一对 (n_i, n_{i+1}) 单独来看，都是两个独立而均匀分布的整数，所以每一项的期望值都是 $(m^2-1)/(3m)$。一共有 $r-1$ 项，所以

$$\mathrm{E}(|n_1 - n_2| + |n_2 - n_3| + \cdots + |n_{r-1} - n_r|) = \frac{(r-1)(m^2-1)}{3m}.$$

$r = 3$ 的情况，就是我们刚才就 $m = 2$ 验证过的结果。

监考员

把书桌沿直线依次编号为 $1, 2, \ldots, m$，于是书桌 i 与书桌 j 相距 $d|i-j|$。设 c_j 为第 j 次召唤所来自的书桌的编号，其中 $j = 1, \ldots, N+1$。我们把召唤者设置为互相独立，且各自在 $\{1, \ldots, m\}$ 上均匀分布。这就是我们对“m 名考生中任何一人召唤的机会均等”的理解。由于考生从不同时召唤，也不会在另一次召唤进行期间召唤，监考员会逐一处理各次召唤，而当第 $j+1$ 次召唤出现时，监考员正站在书桌 c_j 旁。（若同一名考生连续召唤两次，该段路的长度为 0，而公式已经涵盖这种情况。）

因此所走的总距离是

$$D = d\,|c_1 - 1| + d\sum_{j=1}^{N} |c_{j+1} - c_j|.$$

*第一段路。*监考员由书桌 1 出发，而 c_1 是均匀分布的，所以

$$\mathrm{E}|c_1 - 1| = \frac{1}{m}\sum_{i=1}^{m}(i-1) = \frac{m-1}{2}.$$

留意这个值比 $(m^2-1)/(3m)$ 大。由这一列书桌的尽头出发，令第一段路比平均的长。

*其余 N 段路。*这里 $c_1, \ldots, c_{N+1}$ 担当 $n_1, \ldots, n_r$ 的角色，其中 $r = N+1$，所以由上面的结果可得

$$\mathrm{E}\sum_{j=1}^{N} |c_{j+1} - c_j| = \frac{N(m^2-1)}{3m}.$$

*合并。*再用一次线性性质，

$$\begin{aligned}\mathrm{E}(D) &= d\left(\frac{m-1}{2} + \frac{N(m-1)(m+1)}{3m}\right)\\ &= \frac{d(m-1)}{6m}(3m + 2N(m+1)) = \frac{d(m-1)}{6m}[2N(m+1) + 3m],\end{aligned}$$

这正是所求的结果。

验算

- 当 $m = 1$ 时只有一张书桌，所以监考员从不走动。公式给出 0，理应如此。
- 当 $N = 0$ 时只有第一次召唤，公式化简为 $d(m-1)\cdot 3m/(6m) = \frac{1}{2}d(m-1)$，即第一段路的期望值。
- 把召唤者的每一个可能序列逐一列举，可确认公式对所有 $m \le 4$ 及 $N \le 3$ 都成立。以 $m = 10$、$N = 20$、$d = 1$ 进行 2×10^5 次模拟，所得平均值为 70.50，而公式给出的是 $\frac{9}{60}(440 + 30) = 70.5$。

常见错误

- 把第一段路当作与其余各段一样处理。监考员由书桌 1 出发，而不是由随机的一张书桌出发，所以该段路的期望值是 $\frac{1}{2}(m-1)$，而不是 $(m^2-1)/(3m)$。
- 只数 $n_1 - n_2 = k$ 的情况，忘记了因 $n_2 - n_1 = k$ 而来的因子 2；或者错误地把 $k = 0$ 的情况也乘以二。
- 以为和式中各项必须互相独立，它们的期望值才可以相加。

STEP 之外

监考员的位置是一条马可夫链（*Markov chain*）：他们下一步往哪里去，只取决于他们现在身在何处。这里的链非常简单，每一步都把过去完全“忘记”。在较复杂的模型中（例如召唤较可能来自附近的书桌），我们便需要这条链的平稳分布（stationary distribution），才能求出长远而言每次召唤平均所走的路程。第一个结果的连续版本是：在 $[0,1]$ 上两个独立而均匀分布的点，平均相距 $\frac{1}{3}$，这正是 $(m^2-1)/(3m)$ 除以 m 后的极限。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖了期望值的线性性质和马可夫链。

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。