解答例

STEP II 1991、第 15 問

確率・統計 ・ 期待値の線形性

> **STEP II 1991、問 15**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/91-s2-q15

鍵となる考え方

組 (n_1, n_2) を $m \times m$ の格子のマス目と考え、m^2 個のマス目がすべて同様に確からしいとする。$|n_1 - n_2| = k$ となるマス目は主対角線に平行な 2 本の対角線上に並ぶので、数えるのは容易である。その後はすべて**期待値の線形性**（linearity of expectation）から導かれる。すなわち、差の和の期待値は各期待値の和に等しい。隣り合う項が変数を共有していて独立でなくても、このことは成り立つ。試験監督者の歩く距離も同じ種類の和であるが、最初の移動だけはランダムな机ではなく決まった机から出発する点が異なる。

解答

$|n_1 - n_2|$ の分布

n_1 と n_2 は独立（independent）で、それぞれ $\{1, \ldots, m\}$ 上の一様分布（uniform distribution）に従うので、m^2 個の順序対 (n_1, n_2) はそれぞれ確率 $1/m^2$ をもつ。

$1 \le k \le m - 1$ を満たす k を固定する。$n_1 - n_2 = k$ となる組は $(k+1, 1), (k+2, 2), \ldots, (m, m-k)$ であり、$m - k$ 個ある。対称性から $n_2 - n_1 = k$ となる組も $m - k$ 個あり、$k \neq 0$ なので両者は重ならない。したがって

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

となり、示すべき式が得られた。（$k = 0$ のときは組が m 個あるので、$\mathrm{P}(|n_1 - n_2| = 0) = 1/m$ である。確認として、$\frac{1}{m} + \sum_{k=1}^{m-1} \frac{2(m-k)}{m^2} = \frac{1}{m} + \frac{m(m-1)}{m^2} = 1$ となる。）

$|n_1 - n_2|$ の期待値

値 $k = 0$ は期待値に何も寄与しないので、

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

となる。ここで、よく知られた和の公式 $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$ と $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$ を $n = m-1$ として用いた。

$m = 2$ で整数が 3 個の場合

$m = 2$ のとき、同様に確からしい三つ組 (n_1, n_2, n_3) は $2^3 = 8$ 個ある。$S = |n_1 - n_2| + |n_2 - n_3|$ とおくと、次のようになる。

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

合計は 8 なので、$\mathrm{E}(S) = 8/8 = 1$ である。主張されている値は $2(m^2-1)/(3m) = 2\cdot 3/6 = 1$ であり、一致する。

> **考えてみよう** 項 $|n_1 - n_2|$ と $|n_2 - n_3|$ はどちらも n_2 を含んでいる。それでも単純に期待値を足し合わせてよいのだろうか。

よいのである。期待値の線形性 $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ は独立性をまったく必要としない。各組 (n_i, n_{i+1}) は単独で見れば 2 つの独立な一様整数からなるので、各項の期待値は $(m^2-1)/(3m)$ である。項は $r-1$ 個あるので、

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

$r = 3$ の場合が、先ほど $m = 2$ で確かめた結果である。

試験監督者

机に一列に沿って $1, 2, \ldots, m$ と番号を付けると、机 i と机 j の距離は $d|i-j|$ である。$j = 1, \ldots, N+1$ に対し、j 番目に呼んだ机の番号を c_j とする。呼ぶ人は互いに独立で、それぞれ $\{1, \ldots, m\}$ 上で一様であるとモデル化する。これが「m 人の受験者のうち誰から呼ばれることも同様に確からしい」の解釈である。受験者が同時に呼んだり、別の呼び出しの対応中に呼んだりすることはないので、試験監督者は呼び出しに順に対

応し、呼び出し $j+1$ が来たときには机 c_j に立っている。（同じ受験者が続けて 2 回呼んだ場合、その移動の長さは 0 であり、式はこの場合も含んでいる。）

したがって、歩いた距離の合計は

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

最初の移動。 試験監督者は机 1 から出発し、c_1 は一様なので、

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

これは $(m^2-1)/(3m)$ より大きいことに注意する。列の端から出発するので、最初の移動は平均より長くなるのである。

残りの N 回の移動。 ここでは $c_1, \ldots, c_{N+1}$ が $r = N+1$ としたときの $n_1, \ldots, n_r$ の役割を果たすので、上の結果より

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

まとめる。 再び線形性により、

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

となり、これが求める結果である。

検算

- $m = 1$ のときは机が 1 つしかないので、試験監督者はまったく動かない。式は 0 を与え、そうなるべきである。
- $N = 0$ のときは最初の呼び出ししかなく、式は $d(m-1)\cdot 3m/(6m) = \frac{1}{2}d(m-1)$ となり、最初の移動の期待値に一致する。
- 呼ぶ人のあらゆる列を厳密に数え上げると、すべての $m \le 4$ と $N \le 3$ について式が正しいことが確かめられる。$m = 10$、$N = 20$、$d = 1$ として 2×10^5 回のシミュレーションを行うと平均は 70.50 となり、式の値 $\frac{9}{60}(440 + 30) = 70.5$ と一致した。

よくある間違い

- 最初の移動を他の移動と同じように扱ってしまう。試験監督者はランダムな机ではなく机 1 から出発するので、その移動の期待値は $(m^2-1)/(3m)$ ではなく $\frac{1}{2}(m-1)$ である。
- $n_1 - n_2 = k$ だけを数えて $n_2 - n_1 = k$ の分の因子 2 を忘れる、あるいは $k = 0$ の場合を誤って 2 倍してしまう。
- 期待値を足し合わせるには和の各項が独立でなければならないと考えてしまう。

STEP のその先へ

試験監督者の位置は**マルコフ連鎖**（Markov chain）をなす。すなわち、次にどこへ行くかは現在どこにいるかだけで決まる。ここでの連鎖は非常に単純で、各ステップは過去を完全に「忘れる」。より豊かなモデル（たとえば近くの机からの呼び出しほど起こりやすいもの）では、呼び出し 1 回あたりの長期的な平均移動距離を求めるのに、連鎖の定常分布（stationary distribution）が必要になる。最初の結果の連続版は、$[0,1]$ 上の独立な一様な 2 点の距離の平均が $\frac{1}{3}$ であるというものであり、これは $(m^2-1)/(3m)$ を m で割ったものの極限である。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、期待値の線形性とマルコフ連鎖の両方を扱っている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。