解答例

STEP I 2003、第 12 問

確率・統計 ・ 積の期待値

> **STEP I 2003、問 12**
>
> 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/03-s1-q12

鍵となる考え方

期待値（expected value）は、等しい確率で起こるすべての結果にわたる（値 × 確率）の和である。したがって、どちらの小問も、数の順序対を列挙してその積を足し合わせることに帰着する。二重和 $\sum_r \sum_s rs$ は $(\sum r)^2$ と因数分解でき、小問 (i) はこれに尽きる。小問 (ii) で変わるのは、$r = s$ となる対がもはや起こり得ないという点だけなので、同じ二重和から対角成分を取り除けばよい。

解答

1 個目の玉の数を X、2 個目の玉の数を Y とする。

小問 (i)

1 個目の玉は袋に戻されるので、2 回目は再び玉がすべてそろった袋から取り出され、1 回目の結果には依存しない。よって、任意の $r, s \in \{1, \ldots, n\}$ に対して

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}.$$

このような順序対 (r, s) は n^2 個あり、それぞれ積 rs を与えるので、期待値の定義より

$$\mathrm{E}(XY) = \sum_{r=1}^{n} \sum_{s=1}^{n} rs\,\mathrm{P}(X = r,\ Y = s) = \frac{1}{n^2} \sum_{r=1}^{n} \sum_{s=1}^{n} rs,$$

となる。これが説明を求められていた式である。

これを簡単にするために、r は s に依存しないので内側の和の外に出せることに注意

する：

$$\sum_{r=1}^{n}\sum_{s=1}^{n} rs = \sum_{r=1}^{n} r\Big(\sum_{s=1}^{n} s\Big) = \Big(\sum_{r=1}^{n} r\Big)^2 = \tfrac{1}{4}n^2(n+1)^2.$$

したがって

$$\mathrm{E}(XY) = \frac{(n+1)^2}{4}.$$

（同じことであるが、X と Y は独立（independent）なので、$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = (\frac{n+1}{2})^2$ である。）

小問 (ii)

玉を 2 個取り出すためには $n \geq 2$ が必要である。今度は 1 個目の玉を戻さないので、起こり得る結果は $r \neq s$ を満たす順序対 (r, s) である。このような対に対して

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s \mid X = r) = \frac{1}{n} \cdot \frac{1}{n-1},$$

となる。玉 r が取り除かれた後、2 個目の玉は残りの $n-1$ 個のどれもが等しい確率で出るからである。このような対は $n(n-1)$ 個あってすべて同様に確からしく、$r = s$ となる対の確率は 0 である。よって

$$\mathrm{E}(XY) = \frac{1}{n(n-1)} \sum_{r \neq s} rs.$$

考えてみよう $r \neq s$ となる対だけにわたって rs を足し合わせるにはどうすればよいか？

小問 (i) で求めたすべての対にわたる和をとり、そこから $r \cdot r = r^2$ を与える対角項 $r = s$ を引けばよい：

$$\sum_{r \neq s} rs = \sum_{r=1}^{n}\sum_{s=1}^{n} rs - \sum_{r=1}^{n} r^2 = \tfrac{1}{4}n^2(n+1)^2 - \tfrac{1}{6}n(n+1)(2n+1).$$

共通因数 $\frac{1}{12}n(n+1)$ をくくり出すと

$$\begin{aligned}\sum_{r \neq s} rs &= \tfrac{1}{12}n(n+1)[3n(n+1) - 2(2n+1)] \\ &= \tfrac{1}{12}n(n+1)(3n^2 - n - 2) \\ &= \tfrac{1}{12}n(n+1)(n-1)(3n+2).\end{aligned}$$

$n \geq 2$ より 0 でない $n(n-1)$ で割ると

$$\mathrm{E}(XY) = \frac{(n+1)(3n+2)}{12},$$

となり、示すべき式が得られた。

検算

$n=3$ とする。復元抽出（with replacement）では、9 個の積の平均は $\frac{1}{9}(1+2+3)^2=4$ であり、$\frac{(3+1)^2}{4}=4$ と一致する。非復元抽出（without replacement）では、順序を考えない対 $\{1,2\},\{1,3\},\{2,3\}$ が積 $2,3,6$ を与え、その平均は $\frac{11}{3}$ であり、$\frac{4\cdot 11}{12}=\frac{11}{3}$ と一致する。$n=2$ のとき、非復元抽出で得られる積は 2 のみであり、公式は $\frac{3\cdot 8}{12}=2$ を与える。さらに、12 以下のすべての n について、正確な数え上げによって両方の公式を確認した。

二つの答えを比べることも、よい妥当性の確認になる：

$$\frac{(n+1)(3n+2)}{12}-\frac{(n+1)^2}{4}=-\frac{n+1}{12}<0.$$

非復元抽出のほうが答えが小さい。これは納得がいく。$n\cdot n$ のような大きな対角成分の積が失われているからである。

よくある間違い

- 小問 (ii) で n^2 で割ってしまう。起こり得る順序対は $n(n-1)$ 個しかなく、それぞれの確率は $\frac{1}{n(n-1)}$ である。
- 和には順序を考えない対を使い、個数には順序対を使う（あるいはその逆）。順序を考えない対にわたって和をとると、和と個数がともに半分になるので比は変わらないが、両者を混同すると 2 倍ずれる。
- 小問 (ii) で $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$ と書いてしまう。これには独立性が必要であり、玉を戻さない場合 X と Y は独立ではない。

STEP のその先へ

ここで求めた差は、ちょうど共分散（covariance）$\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)=-\frac{n+1}{12}$ である。これが負になるのは、1 個目の数が大きいと、2 個目の数が大きくなる可能性がわずかに下がるからである。このような共分散は、非復元抽出における和の分散を求めるのに必要となるものであり、そこから超幾何分布（hypergeometric distribution）の分散や、標本調査で用いられる「有限母集団修正」（finite population correction）へとつながっていく。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）では、まさにこの種の共分散と指示変数（indicator）の手法が展開されている。

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。