解答例

STEP III 2007、第 12 問

確率・統計 ・ ランダムな個数の和

> **STEP III 2007、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/07-s3-q12

鍵となる考え方

Y は項の個数が**確率的に決まる**和であるから、項数が一定の確率変数の和に対する規則をそのまま使うことはできない。ここでのこつは N で条件付けることである。$N = k$ であるとわかってしまえば、Y は k 個の独立な確率変数の通常の和であり、その平均と分散はわかっている。そのあとで N の分布について平均をとる。最初の小問で $\mathrm{E}(N)$ と $\mathrm{E}(N^2)$ を求めさせているのはこのためである。

解答

N の分布

数は一様ランダムに選ばれるものとする。したがって $k = 1, 2, \ldots, 2n-1$ に対して $\mathrm{P}(N = k) = \frac{1}{2n-1}$ である。$m = 2n-1$ として $\sum_{k=1}^{m} k = \frac{1}{2}m(m+1)$ と $\sum_{k=1}^{m} k^2 = \frac{1}{6}m(m+1)(2m+1)$ を用いると、

$$\mathrm{E}(N) = \frac{1}{2n-1} \cdot \frac{(2n-1)(2n)}{2} = n,$$
$$\mathrm{E}(N^2) = \frac{1}{2n-1} \cdot \frac{(2n-1)(2n)(4n-1)}{6} = \frac{n(4n-1)}{3}.$$

（平均が n であることは対称性からも明らかである。n は $1, \ldots, 2n-1$ のちょうど真ん中である。）あとで使うために、

$$\mathrm{Var}(N) = \frac{n(4n-1)}{3} - n^2 = \frac{n^2 - n}{3} = \frac{n(n-1)}{3}.$$

を求めておく。

N による条件付け

k を固定し、$N = k$ であるとする。X_i は N と独立（independent）であるから、$N = k$ で条件付けてもそれらの分布は変わらない。よって $N = k$ のもとで $Y = X_1 + \cdots + X_k$ であり、これはそれぞれ平均 μ、分散 σ^2 をもつ k 個の独立な確率変数の和である。ゆえに

$$\mathrm{E}(Y \mid N = k) = k\mu, \qquad \mathrm{Var}(Y \mid N = k) = k\sigma^2,$$

となる。2 つ目の式では X_i の独立性を用いた（分散は足し合わされる）。したがって

$$\mathrm{E}(Y^2 \mid N = k) = \mathrm{Var}(Y \mid N = k) + (\mathrm{E}(Y \mid N = k))^2 = k\sigma^2 + k^2\mu^2.$$

考えてみよう 条件付き期待値（conditional expectation）から通常の期待値に戻るにはどうすればよいか。

全期待値の法則（law of total expectation、分割定理（partition theorem）ともいう）による：$\mathrm{E}(g(Y,N)) = \sum_k \mathrm{P}(N = k)\,\mathrm{E}(g(Y,N) \mid N = k)$。各項の中では、$N$ は単なる数 k である。

Y の平均

$$\mathrm{E}(Y) = \sum_k \mathrm{P}(N = k)\, k\mu = \mu\,\mathrm{E}(N) = n\mu,$$

となり、求める結果が得られた。

共分散

$N = k$ のもとで $YN = kY$ であるから、$\mathrm{E}(YN \mid N = k) = k \cdot k\mu = k^2\mu$ である。したがって $\mathrm{E}(YN) = \mu\,\mathrm{E}(N^2)$ であり、

$$\begin{aligned}\mathrm{Cov}(Y,N) &= \mathrm{E}(YN) - \mathrm{E}(Y)\,\mathrm{E}(N) = \mu\,\mathrm{E}(N^2) - n\mu \cdot n \\ &= \mu\,\mathrm{Var}(N) = \tfrac{1}{3}n(n-1)\mu,\end{aligned}$$

となる。これが求める結果である。

Y の分散

$\mathrm{E}(Y^2 \mid N = k) = k\sigma^2 + k^2\mu^2$ を N について平均すると、

$$\mathrm{E}(Y^2) = \sigma^2\,\mathrm{E}(N) + \mu^2\,\mathrm{E}(N^2) = n\sigma^2 + \frac{n(4n-1)}{3}\mu^2.$$

ゆえに

$$\mathrm{Var}(Y) = n\sigma^2 + \frac{n(4n-1)}{3}\mu^2 - n^2\mu^2 = n\sigma^2 + \frac{n(n-1)}{3}\mu^2.$$

2 つの項には自然な意味がある。$n\sigma^2$ は試行回数を固定したときの Y のばらつきの平均であり、$\frac{1}{3}n(n-1)\mu^2 = \mu^2\,\mathrm{Var}(N)$ は試行が何回行われるかわからないことによって生じる余分なばらつきである。

検算

- $n = 1$ ならばつねに $N = 1$ であり、公式は $\mathrm{E}(Y) = \mu$、$\mathrm{Cov}(Y, N) = 0$、$\mathrm{Var}(Y) = \sigma^2$ に帰着する。単独の X_1 に対しては当然そうなるべきである。
- $\mu = 0$ ならば余分な項は消える。このとき項数は平均に影響しないので、ばらつきを増やすことも Y と相関をもつこともありえない。
- $n = 2$ とし、各 X_i は確率 $\frac{1}{2}$ ずつで 0 または 1 をとるとする。このとき $N \in \{1, 2, 3\}$ かつ $Y \mid N \sim \mathrm{B}(N, \frac{1}{2})$ であり、公式から $\mathrm{Var}(Y) = 2 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{2}{3}$ が得られる。$n = 1, \ldots, 6$ および X_i のいくつかの分布について、(N, Y) の同時分布を厳密に数え上げることで、この値とすべての結果を確認した。

よくある間違い

- N 自体が変動することを忘れて、$\mathrm{Var}(Y) = \mathrm{E}(N)\sigma^2$ だけを書いてしまう。2 つ目の項こそが要点である。
- Y と N を独立とみなして $\mathrm{Cov}(Y, N) = 0$ としてしまう。$\mu > 0$ のときは試行が多いほど和が大きくなるので、両者には正の相関がある。

STEP のその先へ

Y は**ランダム和**（random sum、複合分布（compound distribution）ともいう）であり、分散についての結果は全分散の法則（law of total variance）$\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y \mid N)) + \mathrm{Var}(\mathrm{E}(Y \mid N))$ の一例である。ここではこれが $\sigma^2\mathrm{E}(N) + \mu^2\mathrm{Var}(N)$ となる。ランダム和は保険金請求や分枝過程（branching process）のモデルとなり、確率母関数（probability generating function）を用いると $G_Y(s) = G_N(G_X(s))$ のように見事に扱える。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、ランダム和と分枝過程をこの方法で扱っている。

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。