解答例

STEP III 2011、第 12 問

確率・統計 ・ 確率母関数による複合分布

STEP III 2011、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/11-s3-q12

鍵となる考え方

分布の平均（mean）と分散（variance）について必要なことはすべて、その確率母関数（probability generating function, pgf）の $t = 1$ における 1 階と 2 階の導関数に含まれている。そこで、合成関数（composite function）$G(H(t))$ を連鎖律（chain rule）で（2 回）微分し、$t = 1$ を代入して $H(1) = 1$ を用いる。硬貨の問題では、N は幾何分布（geometric distribution）に従い、各 X_i は公正な硬貨 1 回分の指示変数（indicator）である。したがって、どちらの確率母関数も単純であり、$G(H(t))$ は等比級数（geometric series）として展開できる有理関数となる。

解答

一般的な結果

非負整数値をとる任意の確率変数 Z で、確率母関数が $P(t) = \mathrm{E}(t^Z)$ であるものについて、$P(1) = 1$ および

$$P'(1) = \mathrm{E}(Z), \qquad P''(1) = \mathrm{E}(Z(Z-1)),$$

が成り立ち、したがって

$$\mathrm{Var}(Z) = P''(1) + P'(1) - (P'(1))^2.$$

である。$\mu = \mathrm{E}(X_i) = H'(1)$、$\sigma^2 = \mathrm{Var}(X_i)$ とおき、Y の確率母関数を $K(t) = G(H(t))$ と書く。

連鎖律により $K'(t) = G'(H(t))\,H'(t)$ である。$H(1) = 1$ であるから

$$\mathrm{E}(Y) = K'(1) = G'(1)H'(1) = \mathrm{E}(N)\,\mathrm{E}(X_i),$$

となり、これが 1 つ目の結果である。

もう一度微分すると（積の微分法則（product rule）を用い、続いて連鎖律を用いる）

$$K''(t) = G''(H(t))(H'(t))^2 + G'(H(t))\,H''(t),$$

となるので、再び $H(1) = 1$ を用いて

$$K''(1) = G''(1)\mu^2 + \mathrm{E}(N)\,H''(1).$$

を得る。上の分散の公式を N と X_i に適用すると

$$G''(1) = \mathrm{Var}(N) - \mathrm{E}(N) + (\mathrm{E}(N))^2, \qquad H''(1) = \sigma^2 - \mu + \mu^2.$$

である。

> **考えてみよう** $G''(1)$ と $H''(1)$ を書き直さずに、$K''(1)$ から直接 $\mathrm{Var}(Y)$ を計算すればよいのではないか。なぜそうしないのだろうか。

求めたい答えは階乗モーメント（factorial moments）ではなく分散で表されたものだからである。$G''(1)$ と $H''(1)$ を $\mathrm{Var}(N)$ と σ^2 を用いて表すことで、不要な項が打ち消し合う。$m = \mathrm{E}(N)$ とおくと

$$\begin{aligned}
\mathrm{Var}(Y) &= K''(1) + K'(1) - (K'(1))^2 \\
&= (\mathrm{Var}(N) - m + m^2)\mu^2 + m(\sigma^2 - \mu + \mu^2) + m\mu - m^2\mu^2 \\
&= \mathrm{Var}(N)\mu^2 + m\sigma^2 + (-m\mu^2 + m\mu^2) + (m^2\mu^2 - m^2\mu^2) + (-m\mu + m\mu) \\
&= \mathrm{Var}(N)(\mathrm{E}(X_i))^2 + \mathrm{E}(N)\,\mathrm{Var}(X_i),
\end{aligned}$$

となり、示すべき式が得られた。

硬貨の問題

N の確率母関数。 初めて表が出るのが n 回目の投げであるのは、最初の $n-1$ 回が裏で n 回目が表のときであるから、$n = 1, 2, 3, \ldots$ に対して $\mathrm{P}(N = n) = (\frac{1}{2})^n$ である（幾何分布）。よって $|t| < 2$ のとき

$$G(t) = \sum_{n=1}^{\infty} \left(\frac{t}{2}\right)^n = \frac{t/2}{1 - t/2} = \frac{t}{2 - t}.$$

各 X_i の確率母関数。 追加の投げのうち i 回目が表ならば $X_i = 1$、そうでなければ $X_i = 0$ とする。このとき $Y = X_1 + \cdots + X_N$ であり、$H(t) = \frac{1}{2} + \frac{1}{2}t = \frac{1}{2}(1 + t)$ である。追加の投げは互いに独立であり、また N を決めた投げとも独立であるから、一般的な結果の条件が満たされる。

Y の確率母関数。

$$G(H(t)) = \frac{\frac{1}{2}(1+t)}{2-\frac{1}{2}(1+t)} = \frac{1+t}{3-t}.$$

平均と分散。 $p = \frac{1}{2}$ の幾何分布では $\mathrm{E}(N) = 1/p = 2$、$\mathrm{Var}(N) = (1-p)/p^2 = 2$ である（あるいは G を微分してもよい。$G'(t) = 2/(2-t)^2$、$G''(t) = 4/(2-t)^3$ より $G'(1) = 2$、$G''(1) = 4$ となり、$\mathrm{Var}(N) = 4+2-4 = 2$ である）。各 X_i はベルヌーイ分布（Bernoulli distribution）Bernoulli$(\frac{1}{2})$ に従うので、$\mathrm{E}(X_i) = \frac{1}{2}$、$\mathrm{Var}(X_i) = \frac{1}{4}$ である。したがって

$$\mathrm{E}(Y) = 2\cdot\tfrac{1}{2} = 1, \qquad \mathrm{Var}(Y) = 2\cdot\tfrac{1}{4} + 2\cdot\tfrac{1}{4} = 1.$$

Y の分布。 $|t| < 3$ のとき、次のように展開する。

$$\frac{1+t}{3-t} = \frac{1+t}{3}\cdot\frac{1}{1-t/3} = \frac{1+t}{3}\sum_{k=0}^{\infty}\frac{t^k}{3^k} = \sum_{k=0}^{\infty}\frac{t^k}{3^{k+1}} + \sum_{k=0}^{\infty}\frac{t^{k+1}}{3^{k+1}}.$$

t^0 の係数は 1 つ目の和からのみ生じる。$r \geq 1$ に対して t^r の係数は $\dfrac{1}{3^{r+1}} + \dfrac{1}{3^r} = \dfrac{4}{3^{r+1}}$ である。よって

$$\mathrm{P}(Y=0) = \frac{1}{3}, \qquad \mathrm{P}(Y=r) = \frac{4}{3^{r+1}} \quad (r = 1,2,3,\ldots).$$

検算

- 確率の総和は 1 になる：$\frac{1}{3} + 4\sum_{r\geq 1} 3^{-(r+1)} = \frac{1}{3} + 4\cdot\frac{1}{6} = 1$。
- $\mathrm{P}(Y=0)$ を直接求めると $\sum_{n\geq 1}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \sum_{n\geq 1} 4^{-n} = \frac{1}{3}$ である。同様に $\mathrm{P}(Y=1) = \sum_{n\geq 1} n\,4^{-n} = \frac{4}{9}$ であり、$4/3^2$ と一致する。
- $n \leq 200$ の範囲で $\mathrm{P}(N=n)\binom{n}{r}2^{-n}$ を厳密に足し合わせると、$r \leq 7$ について $\mathrm{P}(Y=r)$ が再現され、平均と分散はともに 1.000 となった。200 000 回のシミュレーションでは 0.996 と 0.994 が得られた。

よくある間違い

- N が 0 から始まるとしてしまう（確率母関数が $1/(2-t)$ になる）。ここでの N は表が出た投げも数えるので、$N \geq 1$ である。
- $P''(1)$ が $\mathrm{E}(Z^2)$ ではなく $\mathrm{E}(Z^2) - \mathrm{E}(Z)$ であることを忘れ、分散の公式の $+P'(1)$ を落としてしまう。
- すべての $r \geq 0$ に対して $4/3^{r+1}$ と答えてしまう。$r = 0$ の項だけは異なる。

STEP のその先へ

Y は**確率和**（random sum）（複合分布（compound distribution））であり、分散の公式は全分散の公式（law of total variance）$\mathrm{Var}(Y) = \mathrm{E}\big(\mathrm{Var}(Y \mid N)\big) + \mathrm{Var}\big(\mathrm{E}(Y \mid N)\big)$ の特別な場合である。確率母関数の合成は分岐過程（branching process）の原動力でもある。分岐過程では $G(G(\cdots G(t)))$ が第 n 世代の大きさを与え、絶滅確率は G の不動点（fixed point）となる。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、母関数、確率和、分岐過程をちょうどこの水準で扱っている。

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。