解答例

STEP II 2003、第 13 問

確率・統計 ・ 切断ポアソン分布

STEP II 2003、問 13

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-s2-q13

鍵となる考え方

これはポアソン分布（Poisson distribution）から値 0 を取り除き、残りの確率を総和が 1 のままになるように拡大したものである。したがって、必要な和はいずれも $k = 0$ の項が欠けたポアソン分布の和である。しかも平均（mean）と分散（variance）については、$k = 0$ の項はもともと何の寄与もしない。そこで、よく知られたポアソン分布の事実 $\sum k\, p_k = \lambda$ と $\sum k^2 p_k = \lambda + \lambda^2$ を借りてきて、単に A を掛ければよい。

解答

定数 A

通常の $\mathrm{Po}(\lambda)$ の確率を $p_k = \dfrac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$ と書く。$\sum_{k\ge 0} \lambda^k/k! = \mathrm{e}^{\lambda}$ であるから、

$$\sum_{k=1}^{\infty} p_k = \mathrm{e}^{-\lambda}(\mathrm{e}^{\lambda} - 1) = 1 - \mathrm{e}^{-\lambda}.$$

確率 $\mathrm{P}(X = k) = A p_k$ の総和は 1 でなければならないので、$A(1 - \mathrm{e}^{-\lambda}) = 1$、すなわち

$$A = (1 - \mathrm{e}^{-\lambda})^{-1},$$

となり、示すべき式が得られた。なお、$\lambda > 0$ のとき $0 < 1 - \mathrm{e}^{-\lambda} < 1$ であるから $A > 1$ である。余った確率 p_0 が $k \ge 1$ の値に分配されたわけである。

平均

$\sum k\,p_k$ の $k=0$ の項はゼロであるから、

$$\sum_{k=1}^{\infty} k\,p_k = \sum_{k=0}^{\infty} k\,p_k = \lambda,$$

これは $\mathrm{Po}(\lambda)$ の平均である。（直接計算すると $\sum_{k\geq 1} k\lambda^k \mathrm{e}^{-\lambda}/k! = \lambda \sum_{k\geq 1} \lambda^{k-1}\mathrm{e}^{-\lambda}/(k-1)! = \lambda$ となる。）したがって

$$\mu = \mathrm{E}(X) = A\lambda = \frac{\lambda}{1-\mathrm{e}^{-\lambda}}.$$

分散

同様に、$\sum k^2 p_k$ の $k=0$ の項はゼロであり、$\mathrm{Po}(\lambda)$ については $\mathrm{E}(Y^2) = \mathrm{Var}(Y) + (\mathrm{E}\,Y)^2 = \lambda + \lambda^2$ が成り立つ。この事実を引用したくなければ、$k^2 = k(k-1)+k$ と書けばよい。

$$\sum_{k\geq 2} k(k-1)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} = \lambda^2 \sum_{k\geq 2} \frac{\lambda^{k-2}\mathrm{e}^{-\lambda}}{(k-2)!} = \lambda^2, \qquad \text{よって} \qquad \sum_{k\geq 1} k^2 p_k = \lambda^2 + \lambda.$$

したがって

$$\begin{aligned} \mathrm{E}(X^2) &= A\lambda(1+\lambda) = \mu(1+\lambda), \\ \mathrm{Var}(X) &= \mathrm{E}(X^2) - \mu^2 = \mu(1+\lambda) - \mu^2 = \mu(1-\mu+\lambda), \end{aligned}$$

となり、示すべき式が得られた。

不等式

考えてみよう 狭義の不等式 $\mu < 1+\lambda$ はどこから得られるだろうか？

分散からである。X は $1,2,3,\ldots$ のすべての値を正の確率でとるので定数ではない。そして、定数でない確率変数（random variable）の分散は真に正である（$\mathrm{Var}(X) = \mathrm{E}((X-\mu)^2)$ であり、$(X-\mu)^2 > 0$ となる確率が正であるから）。よって

$$\mu(1-\mu+\lambda) > 0.$$

$\mu > 0$ であるから（正の値をとる確率変数の平均であるから）、μ で割って $1-\mu+\lambda > 0$、すなわち $\mu < 1+\lambda$ を得る。

下からの評価については、$\mu = A\lambda$ で $A > 1$ かつ $\lambda > 0$ であるから、$\mu > \lambda$ である。以上を合わせて

$$\lambda < \mu < 1+\lambda.$$

$\lambda = 100$ のときの正規近似

ここで e^{-100} は天文学的に小さい（約 4×10^{-44}）ので、実用上のどんな精度においても

$$\mu = \frac{100}{1 - \mathrm{e}^{-100}} \approx 100, \qquad \mathrm{Var}(X) = \mu(101 - \mu) \approx 100 \times 1 = 100.$$

そこで X を $Y \sim \mathrm{N}(100, 10^2)$ で近似する。X は整数値をとるので、連続性補正（continuity correction）を用いる。事象 $X = 100$ は $99.5 < Y < 100.5$ に対応する。$Z = (Y - 100)/10$ によって標準化すると、

$$\begin{aligned} \mathrm{P}(X = 100) &\approx \mathrm{P}(-0.05 < Z < 0.05) \\ &= 2\Phi(0.05) - 1 \\ &\approx 2(0.51994) - 1 = 0.0399, \end{aligned}$$

よって、小数第 2 位までで $\mathrm{P}(X = 100) \approx 0.04$ である。

検算

- $\lambda \to 0$ のとき、$\mu = \lambda/(1 - \mathrm{e}^{-\lambda}) \to 1$ であり、分散は $\to 0$ となる。ほぼすべての確率が $k = 1$ に集中するので、これは当然である。また、不等式 $\lambda < \mu < 1 + \lambda$ が μ を 1 に向けて挟み込むことをも合致する。
- 多数のランダムな λ について級数を数値的に足し合わせると、A、μ、$\mu(1 - \mu + \lambda)$ が計算機の精度で再現される。
- $\lambda = 100$ のとき、厳密な確率は $\dfrac{100^{100}\mathrm{e}^{-100}}{100!\,(1 - \mathrm{e}^{-100})} \approx 0.03986$ であり、正規近似（normal approximation）の値は 0.03988 である。どちらも四捨五入すると 0.04 になる。

よくある間違い

- 連続性補正を忘れて $\mathrm{P}(Y = 100) = 0$ と計算してしまうこと、あるいは根拠なく確率密度の値を用いること。
- $\mu \leq 1 + \lambda$ しか証明しないこと。狭義の不等式を示すには、X が定数でないという観察が必要である。
- $\mu > 0$ であることに触れずに μ で割ること。

STEP のその先へ

X は**零切断ポアソン分布**（zero-truncated Poisson distribution）であり、統計学者はこれを、個数がゼロという観測が起こりえない場合に用いる（たとえば、検問所を通過する車の乗員数である。乗員がゼロの車はない）。正規近似がうまくいくのは、Po(100) に従う変数が 100 個の独立な Po(1) 変数の和であり、中心極限定理（central limit theorem）が適用できるからである。このスケールでは切断の影響は見えない。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）では、この問題の前半と後半それぞれの背後にある母関数（generating function）と中心極限定理が扱われている。

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。