解答例

STEP II 2002、第 14 問

確率・統計 ・ 混合分布とベイズの定理

STEP II 2002、問 14

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/02-s2-q14

鍵となる考え方

これは二段階の試行である。まず地域が一つ選ばれ、次にその地域で事故が起こる。したがって、地域で条件付けをすればよい。全期待値の法則（law of total expectation）から平均が、全確率の法則（law of total probability）から分布が得られ、最後にベイズの定理（Bayes' theorem）を使って問いの向きを逆にする。一つの地域での事故の件数はポアソン分布（Poisson distribution）でモデル化する。問題文には平均しか与えられていないからであり、示すべき式に現れる $\mathrm{e}^{-\lambda}\lambda^k/k!$ という形がこのモデルの正しさを裏づけている。

解答

設定

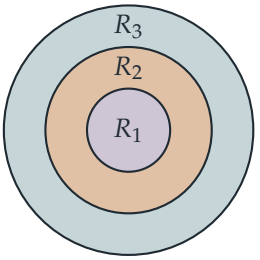
選ばれた地域の番号を S とし、R_n における事故件数の平均を $\lambda_n = 2 - \frac{n}{N}$ とする。地域 R_n は半径 $n-1$ と n の円環（annulus）であるから、その面積は

$$\pi n^2 - \pi(n-1)^2 = \pi(2n-1).$$

である。島全体の面積は πN^2 である（確認として、$\sum_{n=1}^{N}(2n-1) = N^2$ となる）。よって

$$\mathrm{P}(S=n) = \frac{2n-1}{N^2}, \qquad n = 1, \ldots, N.$$

$S = n$ のもとで、X を平均 λ_n のポアソン分布に従うものとしてモデル化する。すべての n について $\lambda_n \geq 1 > 0$ であるから、これらのポアソン分布はいずれも意味をもつ。



面積 $R_n = \pi(2n-1)$

長期的な平均

大数の法則（law of large numbers）により、長期的に見た 1 日あたりの記録件数の平均は $\mathrm{E}(X)$ である。地域で条件付けをすると

$$\begin{aligned}\mathrm{E}(X) &= \sum_{n=1}^{N} \mathrm{P}(S=n)\,\mathrm{E}(X \mid S=n) = \sum_{n=1}^{N} \frac{2n-1}{N^2}\Big(2-\frac{n}{N}\Big)\\ &= 2\sum_{n=1}^{N}\frac{2n-1}{N^2} \;-\; \frac{1}{N^3}\sum_{n=1}^{N}(2n^2-n)\\ &= 2-\frac{1}{N^3}\Big(\frac{N(N+1)(2N+1)}{3}-\frac{N(N+1)}{2}\Big).\end{aligned}$$

を得る。最初の和は、選ばれる確率の合計が 1 であるから 1 となる。括弧の中では、与えられた $\sum n^2$ の公式と $\sum n = \frac{1}{2}N(N+1)$ を用いた。$\frac{1}{6}N(N+1)$ をくくり出すと $2(2N+1)-3 = 4N-1$ が残るので、

$$\mathrm{E}(X) = 2-\frac{(N+1)(4N-1)}{6N^2} = 2-\frac{1}{6}\Big(1+\frac{1}{N}\Big)\Big(4-\frac{1}{N}\Big),$$

となる。ここでは N^2 を $N\cdot N$ として二つの因数に分けた。これが求める結果である。

X の分布

全確率の法則により、$k = 0, 1, 2, \ldots$ に対して

$$\mathrm{P}(X=k) = \sum_{n=1}^{N}\mathrm{P}(S=n)\,\mathrm{P}(X=k \mid S=n) = \sum_{n=1}^{N}\frac{2n-1}{N^2}\cdot\frac{\mathrm{e}^{-\lambda_n}\lambda_n^k}{k!}.$$

である。ここで各部分を N を用いて書き直す。

$$\lambda_n^k = \Big(\frac{2N-n}{N}\Big)^k = N^{-k}(2N-n)^k, \qquad \mathrm{e}^{-\lambda_n} = \mathrm{e}^{-2+n/N} = \mathrm{e}^{-2}\,\mathrm{e}^{n/N}.$$

因数 e^{-2}、N^{-2}、N^{-k}、$1/k!$ は n によらないので、和の外に出せる。

$$\mathrm{P}(X=k) = \frac{\mathrm{e}^{-2}N^{-k-2}}{k!}\sum_{n=1}^{N}(2n-1)(2N-n)^k\,\mathrm{e}^{n/N},$$

これで示された。

$N=3$、事故 2 件の場合

求めたいのは $\mathrm{P}(S=2 \mid X=2)$ である。ベイズの定理により

$$\mathrm{P}(S=2 \mid X=2)=\frac{\mathrm{P}(S=2)\,\mathrm{P}(X=2 \mid S=2)}{\mathrm{P}(X=2)}.$$

考えてみよう $\mathrm{P}(X=2)$ をすべて計算する必要があるだろうか？

その必要はない。分子は $\mathrm{P}(X=2)$ を表す和の $n=2$ の項にほかならない。したがって共通因数 $\mathrm{e}^{-2}N^{-k-2}/k!$ は約分され、求める確率は $\sum(2n-1)(2N-n)^k\mathrm{e}^{n/N}$ の $n=2$ の項を和全体で割ったものになる。

$N=3$、$k=2$ のとき、各項 $(2n-1)(6-n)^2\mathrm{e}^{n/3}$ は

$$n=1:\ 1\cdot 25\,\mathrm{e}^{1/3},\qquad n=2:\ 3\cdot 16\,\mathrm{e}^{2/3}=48\,\mathrm{e}^{2/3},\qquad n=3:\ 5\cdot 9\,\mathrm{e}=45\,\mathrm{e}.$$

である。よって、分母と分子を $\mathrm{e}^{2/3}$ で割ると

$$\mathrm{P}(S=2 \mid X=2)=\frac{48\,\mathrm{e}^{2/3}}{25\,\mathrm{e}^{1/3}+48\,\mathrm{e}^{2/3}+45\,\mathrm{e}}=\frac{48}{48+45\,\mathrm{e}^{1/3}+25\,\mathrm{e}^{-1/3}},$$

となり、これが求める結果である。

検算

- $N=1$ のときは平均 1 の地域が一つだけであり、公式は $2-\frac{1}{6}\cdot 2\cdot 3=1$ を与える。$N\to\infty$ のとき平均は $2-\frac{4}{6}=\frac{4}{3}$ に近づく。これは $\int_0^1 2x(2-x)\,\mathrm{d}x=\frac{4}{3}$ と一致する。ここで、一様に選んだ点の相対半径 x の密度は $2x$ である。
- $N=3$ のとき平均は $\frac{32}{27}\approx 1.185$ である。4×10^5 日分のシミュレーション（円板内に一様な点を選び、次にポアソン分布に従う件数を生成する）では平均 1.182 が得られた。また $\mathrm{P}(X=2)\approx 0.209$（厳密値 0.2094）が得られ、事故 2 件の日のうち R_2 で起きたものの割合は 0.372 となった。最終的な答えの値は 0.3729 である。
- 分布の公式を k について足し合わせると、試したすべての N について全確率が 1 となる。

よくある間違い

- 地域を面積 $2n-1$ に比例する確率ではなく、等しい確率 $1/N$ で選んでしまう。
- ポアソン分布の**平均**を平均し、X がその平均をもつポアソン分布に従うと主張してしまう。ポアソン分布の混合はポアソン分布ではない。このように単純に平均操作を通り抜けられるのは期待値だけである。

STEP のその先へ

変数 X は**混合分布**（mixture distribution）に従う。地域は隠れ変数（hidden variable）の役割を果たし、最後の小問はそれについてのベイズ推論（Bayesian inference）を行っている。上の二つ目の間違いは全分散の法則（law of total variance）$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X \mid S)) + \mathrm{Var}(\mathrm{E}(X \mid S))$ によって説明できる。この法則から、混合分布は同じ平均をもつ単一のポアソン分布よりも常に散らばりが大きいことがわかる。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、条件付き期待値（conditional expectation）とこうした混合分布をよく扱っている。

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。