解答例

STEP I 2005、第 14 問

確率・統計 ・ 混合型確率変数

STEP I 2005、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/05-s1-q14

鍵となる考え方

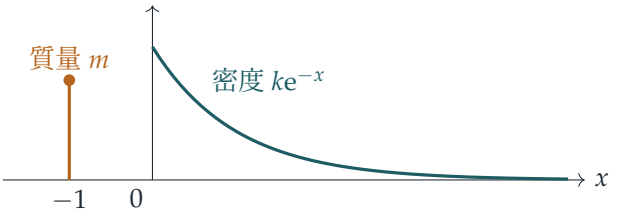
X は**混合型**の確率変数（mixed random variable）である。すなわち、1 点 -1 に集中した確率の塊と、$[0,\infty)$ 上に連続的に広がった部分とから成る。これを扱うには、あらゆる計算を 2 つの部分に分ければよい。$X=-1$ から来る離散的な項と、$[0,\infty)$ 上の確率密度関数（probability density function）に関する積分の和である。連続部分が指数分布（exponential distribution）を定数倍したものにすぎないことに気づけば、積分のほとんどはすでに知っているものになる。

解答

与えられた累積確率を微分すると、X の連続部分の確率密度関数は

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}k(1-\mathrm{e}^{-x})=k\mathrm{e}^{-x},\qquad x\geq 0.$$

となる。したがって X は次のような姿をしている。-1 に質量 m の点質量（point mass）があり、$[0,\infty)$ 上に曲線 $k\mathrm{e}^{-x}$ がある。



小問 (i)

確率の総和は 1 である。$x \to \infty$ とすると $\mathrm{P}(X \geq 0) = k(1-0) = k$ であるから、

$$m + k = 1, \qquad \text{すなわち} \qquad k = 1 - m.$$

($m \geq 0$ は確率であり、かつ $m < \frac{1}{2}$ であるから $k > \frac{1}{2} > 0$ となり、確率密度関数は確かに正である。)

小問 (ii)

2 つの部分に分けると、

$$\begin{aligned}\mathrm{E}(X) &= (-1)\cdot m + \int_0^\infty x\,(1-m)\mathrm{e}^{-x}\,\mathrm{d}x \\ &= -m + (1-m)\cdot 1 = 1 - 2m,\end{aligned}$$

となる。ここで $\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = 1$ を用いた（部分積分（integration by parts）を行うか、パラメータ 1 の指数分布の平均であると見抜けばよい)。これが求める結果である。

小問 (iii)

分散（*variance*)。同様に、$\int_0^\infty x^2\mathrm{e}^{-x}\,\mathrm{d}x = 2$（部分積分を 2 回行う）を用いると、

$$\mathrm{E}(X^2) = (-1)^2\cdot m + (1-m)\cdot 2 = 2 - m.$$

よって

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2 - m - (1-2m)^2 \\ &= 1 + 3m - 4m^2 = (1-m)(1+4m).\end{aligned}$$

考えてみよう 中央値が -1 になったり、-1 と 0 の間のどこかにあったりすることはありうるだろうか？

中央値（*median*)。ありえない。X は $(-1, 0)$ の値を決してとらず、また $\mathrm{P}(X \leq -1) = m < \frac{1}{2}$ であるから、点質量だけでは確率の半分に届かない。条件 $m < \frac{1}{2}$ が使われるのはまさにこの箇所である。したがって中央値 M は $M \geq 0$ かつ $\mathrm{P}(X \leq M) = \frac{1}{2}$ を満たす。

$$\begin{aligned}m + (1-m)(1-\mathrm{e}^{-M}) &= \tfrac{1}{2} \\ 1 - (1-m)\mathrm{e}^{-M} &= \tfrac{1}{2} \\ \mathrm{e}^{-M} &= \frac{1}{2(1-m)}.\end{aligned}$$

ゆえに

$$M = \ln\left(2(1-m)\right) = \ln(2-2m).$$

整合性の確認として、$0 \le m < \frac{1}{2}$ から $1 < 2-2m \le 2$ が得られるので $M > 0$ であり、確かにそうなっているべきとおりである。分布は $[0,\infty)$ 上で連続であるから、この M は $\mathrm{P}(X \ge M) = \frac{1}{2}$ も満たし、中央値は一意に定まる。

小問 (iv)

$X = -1$ のとき $|X|^{1/2} = 1$ であり、$X \ge 0$ のとき $|X|^{1/2} = \sqrt{X}$ である。よって

$$\mathrm{E}(|X|^{1/2}) = 1 \cdot m + (1-m)\int_0^\infty \sqrt{x}\,\mathrm{e}^{-x}\,\mathrm{d}x.$$

考えてみよう $\int_0^\infty \sqrt{x}\,\mathrm{e}^{-x}\,\mathrm{d}x$ を、与えられた積分にどうやって変形すればよいだろうか？

与えられた積分には e^{-y^2} が含まれているので、$x = y^2$（$y \ge 0$）と置換することが示唆される。このとき $\sqrt{x} = y$、$\mathrm{d}x = 2y\,\mathrm{d}y$ であり、積分範囲の 0 と ∞ は変わらない。

$$\int_0^\infty \sqrt{x}\,\mathrm{e}^{-x}\,\mathrm{d}x = \int_0^\infty y\,\mathrm{e}^{-y^2}\,2y\,\mathrm{d}y = 2\int_0^\infty y^2\mathrm{e}^{-y^2}\,\mathrm{d}y = 2\cdot\tfrac{1}{4}\sqrt{\pi} = \tfrac{1}{2}\sqrt{\pi}.$$

よって

$$\mathrm{E}(|X|^{1/2}) = m + \tfrac{1}{2}\sqrt{\pi}\,(1-m).$$

検算

- $m = 0$ のとき、X はちょうどパラメータ 1 の指数分布に従う。上の公式から平均 1、分散 1、中央値 $\ln 2$、$\mathrm{E}(\sqrt{X}) = \frac{1}{2}\sqrt{\pi} = \Gamma(\frac{3}{2})$ が得られ、いずれもよく知られた値に一致する。
- $m \to \frac{1}{2}$ のとき中央値は $\ln 1 = 0$ に近づく。これは理にかなっている。確率の半分が -1 にあるので、中央値は連続部分の下端まで滑り落ちるのである。
- $m = 0.3$ として 4×10^5 個の標本でシミュレーションを行ったところ、平均 0.402（公式では 0.4）、分散 1.551（公式では 1.54）、中央値 0.336（公式では $\ln 1.4 \approx 0.336$）、$\mathrm{E}(|X|^{1/2}) \approx 0.922$（公式では 0.920）となった。

よくある間違い

- $\mathrm{E}(X^2)$ や $\mathrm{E}(|X|^{1/2})$ で点質量を忘れること。$(-1)^2 = |-1|^{1/2} = 1$ であるから、その寄与は $-m$ ではなく $+m$ である。

- 中央値を求めるのに $(1-m)(1-\mathrm{e}^{-M}) = \frac{1}{2}$ を解いてしまうこと。これは -1 にすでに蓄積されている確率 m を無視している。
- 変数変換の際に $\mathrm{d}x$ を $2y\,\mathrm{d}y$ に直し忘れること。そうすると $\frac{1}{2}\sqrt{\pi}$ ではなく $\frac{1}{4}\sqrt{\pi}$ が得られてしまう。

STEP のその先へ

点質量と確率密度が混ざった分布は、実際の場面でよく現れる（保険金請求額は、ある確率で 0 となり、それ以外のときは連続的な値をとる）。大学では、期待値を分布関数（distribution function）に関する積分 $\mathrm{E}(g(X)) = \int g(x)\,\mathrm{d}F(x)$ として書くことで、これらを統一的に扱う。この書き方は離散型・連続型・混合型の場合を一度にまとめて扱える。小問 (iv) の積分はガンマ関数（Gamma function）の値 $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$ である。次の一歩としては、Grimmett と Welsh による *Probability: An Introduction*（確率論入門）がよい。

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。