解答例

STEP III 1999、第 13 問

確率・統計 ・ 連続な事前分布をもつベイズ推論

> **STEP III 1999、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/99-s3-q13

鍵となる考え方

ランダムさは二段階になっている。まず自分の取り分の大きさ X が決まり、次に干しぶどうがどこに入るかが決まる。$X = x$ という**条件のもとで**考える（条件付け、conditioning）と、各干しぶどうはほかの干しぶどうとは独立に確率 x で自分の取り分に入るので、干しぶどうの個数は二項分布（binomial distribution）に従う。そうすれば、小問 (i) は全期待値の法則（law of total expectation）であり、小問 (ii) は事前分布（prior）が連続な場合のベイズの定理（Bayes' theorem）である。

解答

まず A を求める。確率密度関数（density）は積分すると 1 にならなければならないから、

$$\int_0^1 Ax\,\mathrm{d}x = \frac{A}{2} = 1 \quad \Longrightarrow \quad A = 2,$$

よって X は $[0,1]$ 上で密度関数 $\mathrm{f}(x) = 2x$ をもつ。

自分の取り分に入る干しぶどうの個数を N とする。「無作為に配置される」を自然な形でモデル化する。すなわち、各干しぶどうの位置はケーキ全体にわたって一様に分布し、ほかの干しぶどうの位置とも、自分が取るケーキの量とも独立であるとする。ケーキの割合 x を取るならば、各干しぶどうはほかとは独立に確率 x で自分の取り分に入るので、

$$N \mid (X = x) \sim \mathrm{B}(4, x).$$

小問 (i)

$X = x$ のとき、二項分布の平均は $\mathrm{E}(N \mid X = x) = 4x$ である。これを X について平均すると（全期待値の法則）、

$$\begin{aligned}\mathrm{E}(N) &= \int_0^1 \mathrm{E}(N \mid X = x)\,\mathrm{f}(x)\,\mathrm{d}x = \int_0^1 4x \cdot 2x\,\mathrm{d}x \\ &= 8\int_0^1 x^2\,\mathrm{d}x = \frac{8}{3}.\end{aligned}$$

同じことだが、$\mathrm{E}(X) = \frac{2}{3}$ を用いて $\mathrm{E}(N) = 4\,\mathrm{E}(X)$ としてもよい。自分の取り分に入る干しぶどうの個数の期待値は $\frac{8}{3}$ である。

小問 (ii)

求めたいのは $\mathrm{P}(X > \frac{1}{2} \mid N = 4)$ である。$X = x$ のとき、4 個の干しぶどうがすべて自分の取り分に入る確率は x^4 である。

> **考えてみよう** X が連続で $\mathrm{P}(X = x) = 0$ であるとき、ベイズの定理はどのように使えばよいだろうか？

ベイズの定理における和を、密度関数に関する積分に置き換える。取り分の大きさの微小区間 $[x, x + \mathrm{d}x]$ の確率は $2x\,\mathrm{d}x$ であり、その区間のもとで 4 個すべてが入る確率は x^4 である。これを区間全体にわたって足し合わせると、

$$\mathrm{P}(N = 4) = \int_0^1 x^4 \cdot 2x\,\mathrm{d}x = 2\int_0^1 x^5\,\mathrm{d}x = \frac{2}{6} = \frac{1}{3}.$$

同様に、$x > \frac{1}{2}$ である区間だけを残すと、

$$\mathrm{P}(X > \tfrac{1}{2} \text{ かつ} N = 4) = \int_{1/2}^1 2x^5\,\mathrm{d}x = \frac{1}{3}\left(1 - \frac{1}{64}\right) = \frac{1}{3} \cdot \frac{63}{64}.$$

したがって

$$\mathrm{P}(X > \tfrac{1}{2} \mid N = 4) = \frac{\frac{1}{3} \cdot \frac{63}{64}}{\frac{1}{3}} = \frac{63}{64}.$$

別の見方もできる。$N = 4$ のもとでの X の条件付き密度関数は $x^4 \cdot 2x$ に比例するので、$[0, 1]$ 上で $6x^5$ となる。その分布関数（distribution function）は x^6 であり、$1 - (\frac{1}{2})^6 = \frac{63}{64}$ を得る。

検算

- (ii) の答えは 1 に非常に近いが、これは納得できる。干しぶどうをすべて見つけたことは、大きな一切れを取ったことの強い証拠である。事前確率 $\mathrm{P}(X > \frac{1}{2}) = 1 - \frac{1}{4} = \frac{3}{4}$ が $\frac{63}{64} \approx 0.984$ まで引き上げられたのである。

- $\frac{8}{3}$ は 0 と 4 の間にあり、しかも 2 より大きい。密度関数 $2x$ は大きな取り分のほうに偏っているので、そうなるはずである。
- 試行回数 10^6 のモンテカルロ・シミュレーション（Monte Carlo simulation）（U を一様分布として $X = \sqrt{U}$ とおくと密度関数 $2x$ をもつので、それから 4 個の干しぶどうを配置する）では、$\mathrm{E}(N) \approx 2.667$、$\mathrm{P}(N = 4) \approx 0.333$、条件付き確率 ≈ 0.9846 となり、$\frac{63}{64} = 0.984375$ とよく一致した。

よくある間違い

- A を求め忘れる、あるいは $A = 1$ としてしまう。すると f は密度関数ではなくなり、(i) の期待値が正しく求まらない（ただし (ii) では A は約分されて消える）。
- $\mathrm{P}(X > \frac{1}{2} \mid N = 4)$ の代わりに $\mathrm{P}(N = 4 \mid X > \frac{1}{2})$ を計算してしまう。条件付けの向きが逆である。
- 積分の代わりに $(\frac{1}{2})^4$ を用いてしまう。取り分の大きさはランダムなので、ある一つの値を代入するのではなく、x^4 を密度関数について平均しなければならない。

STEP のその先へ

これはベイズ推論（Bayesian inference）の縮図である。密度関数 $2x$ は未知の割合に対する**事前分布**、二項分布は**尤度**（likelihood）、$6x^5$ は**事後分布**（posterior）である。事前分布も事後分布もベータ分布（beta distribution）であり、これは二項データに対する共役事前分布（conjugate prior）の一例である。同じ計算がラプラス（Laplace）の継起の法則（rule of succession）の基礎にもなっている。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は、ベータ・二項分布の話や連続確率変数による条件付けをたいへんわかりやすく扱っている。

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。