12.24 中心極限定理

ワークシート

問題

STEP III 2008、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/08-s3-q12

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

考えてみよう　密度 $\frac{1}{2}e^{-|x|}$ は対称である。そのグラフはどのような形か。平均（mean）はいくらか。

段階 1 積率母関数

扱う内容：グラフの概形と積率母関数

考えてみよう $\mathrm{E}(e^{\theta X})$ を求める積分を 0 で分けよ。なぜ $|\theta| < 1$ が必要なのだろうか？

考えてみよう $\frac{1}{1-\theta^2}$ を級数（series）に展開して、平均（mean）と分散（variance）を読み取れ。

段階 2　カエルの変位

扱う内容：$Y/\sqrt{2n}$ の積率母関数

> **手を止めて、やってみよう**　各跳躍は独立（independent）であるから、その和 Y の積率母関数（moment generating function）はどうなるか？また、Y を $\sqrt{2n}$ で割ると θ はどう変わるか？

段階 3　極限

扱う内容：極限

> **考えてみよう**　対数をとり、小さい u に対して $\ln(1-u) = -u - \frac{1}{2}u^2 - \cdots$ を用いよ。極限の積率母関数をもつのはどの分布か？

段階 4 カエルの見込み

扱う内容：跳躍の最小回数

考えてみよう n が大きいとき、Y は近似的に $\mathrm{N}(0, 2n)$ に従う。標準正規分布（standard normal distribution）に従う Z について $\mathrm{P}(|Z| \geq 1.96) = 0.05$ である。これを用いて n を求めよ。

背後にある数学

一般の定理

同じ議論は、平均が 0 で有限の分散 σ^2 をもつ任意の分布に対して成り立つ（積率母関数の存在も仮定するが、この仮定は外すことができる）。このとき $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$ であり、

$$\left(\mathrm{M}\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\theta^2/2}.$$

となる。極限で生き残るのは分散だけであり、もとの分布の形は忘れ去られる。正規分布が普遍的である理由はここにある。大学では、この定理は常に存在する**特性関数**（characteristic function）$\mathrm{E}(e^{i\theta X})$ を用いて証明され、分布の収束はレヴィ（Lévy）の連続性定理によって厳密に述べられる。

収束の速さは？

中心極限定理は極限で何が起こるかを述べるが、どれだけ速く収束するかは述べない。**ベリー・エッセーンの定理**（Berry–Esseen theorem）は、誤差を定数と $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$ の積で上から評価する。カエルの分布のような対称な分布では、数十回跳べば近似は良好になる。歪んだ分布や、分布の裾の極端な部分では、n が大きくても近似が悪いことがある。正規近似は非常にまれな事象の確率を過小評価するのであり、この事実は保険や金融に影響を及ぼす。

ド・モアブルからリンデベルグまで

ド・モアブル（de Moivre）は 1733 年に二項分布の正規近似を発見し（Section 12.23）、ラプラス（Laplace）は 1810 年にそれをより一般の確率変数の和へと拡張して、測定誤差は多くの小さな効果の和であるという考え方を打ち出した。厳密な証明が得られたのはずっと後のことである。1901 年にアレクサンドル・リャプノフ（Aleksandr Lyapunov）が、そして 1922 年にヤール・リンデベルグ（Jarl Lindeberg）が証明を与え、リンデベルグは定理が成り立つための本質的に最も弱い条件を見いだした。「中心極限定理」という名前は 1920 年にジョージ・ポリア（George Pólya）によって付けられたもので、この定理が確率論の中心にあるという意味である。

次の一歩

1. **2 回跳んだ場合の厳密な計算。** $Y = X_1 + X_2$ の密度が $\frac{1}{4}(1 + |y|)e^{-|y|}$ であることを示し、その分散が 4 であることを確かめよ。

2. **二項分布から正規分布へ。** $X \sim \mathrm{B}(1, \frac{1}{2})$ を平均が 0 になるように平行移動したもの（値 $\pm\frac{1}{2}$ をとる）について、$\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$ を求め、そのような確率変数 n 個の和を標準化したものの積率母関数が $e^{\theta^2/2}$ に近づくことを示せ。

3. **カエル、厳密に。** コンピュータを使って（あるいは検算として $n = 2$ のときの厳密な密度を使って）、$n = 82$ のとき正規近似がどれほど正確かを調べよ。

4. **なぜコーシー分布ではだめなのか。** コーシー分布（Section 12.22）は積率母関数をもたない。議論のどの段階が破綻するか。また、n 個のコーシー分布に従う確率変数の平均はどうなるか。

さらに学ぶために

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press）。母関数と極限定理の章。
- **Hans Fischer, *A History of the Central Limit Theorem***（Springer）。中心極限定理の歴史。ラプラスからリンデベルグまで。
- **David Spiegelhalter, *The Art of Statistics***（Pelican）。統計学の技法。正規分布がいたるところに現れる理由について。

詳しい解説付きの解答は第 12.24 節にある：

https://beyondstep.keithmcnulty.org/ja/ch12/clt.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com