統計教育における
「大数の法則」「中心極限定理」の教授方法の検討

横山監[*1]・高籔学[*2]・新井一成[*3]
Email: m121628x@st.u-gakugei.ac.jp

*1: 東京学芸大学大学院
*2: 東京学芸大学
*3: 東京学芸大学

◎Key Words 統計教育, 中心極限定理, シミュレーション

1. はじめに

「大数の法則」「中心極限定理」は確率と統計の接点となる重要な定理である。経済学では金融工学や証券論の分野においてMarkowitz理論に基づくポートフォリオ構築やBlack-Scholes方程式を用いたオプション価格理論による証券投資などについて学習していく過程で大数の法則、中心極限定理は必須の概念である。これらについては、シミュレーションを用いた直感的な理解を中心とした教育内容が多く見られる。シミュレーションを用いることで大数の法則、中心極限定理の意味するところはつかめるが、理論を本質的に理解するのは難しい。

本研究では、統計教育において大数の法則、中心極限定理をいかに理解させるかについて検討する。

２では大数の法則と中心極限定理の概要をまとめ、３では大数の法則と中心極限定理を学ぶ意義について検討する。４ではそれらの教授方法についての先行研究について整理し問題提起をする。５で理解に必要な内容とそれに合わせた方法・手順について検討する。そして６において、コンピュータ利用の位置付けについて検討する。

本研究においては、社会科学の諸分野において統計的な手法を研究に用いる大学生への教育について考える。

2. 大数の法則と中心極限定理

2.1 大数の法則

大数の法則は一般的に「標本数が十分に大きければ、観察された標本平均を母集団の平均（母平均）とみなしてよい」ということを確率論で厳密に証明したものである。大数の法則には二種類あり、一般的には以下のように定義される[3]。

① 大数の弱法則

$X_1, X_2, \cdots X_n$ が独立で、同一の確率分布に従うとし

$$E(X_i)=\mu, \qquad V(X_i)=\sigma^2 \quad (i=1,2,\cdots,n)$$

とすれば

$$n \to \infty \text{のとき} \ \overline{X}_n = \frac{X_1+\cdots+X_n}{n} \xrightarrow{P} \mu$$

すなわち、$\varepsilon = 0$ を十分に小さい量として

$$n \to \infty \text{のとき} \ P(|\overline{X}_n - \mu| > \varepsilon) \to 0$$

が成立する。

② 大数の強法則

$X_1, X_2, \cdots X_n$ が独立で

$$E(X_i)=\mu, \qquad V(X_i)\le\sigma^2 \quad (i=1,2,\cdots,n)$$

とすれば、

$$n \to \infty \text{のとき} \ \overline{X}_n = \frac{X_1+\cdots+X_n}{n} \to \mu \quad \text{（確率１で）}$$

すなわち次式が成立する。

$$P(n \to \infty \text{のとき} \ \overline{X}_n \to \mu) = 1$$

2.2 中心極限定理

中心極限定理は、母集団の確率分布に関係なく標本数が十分に大きければ、標本平均の確率分布が正規分布に収束することを示している。一般的には以下のように定義される[3]。

$X_1, X_2, \cdots X_n$ が独立で、同一の確率分布に従うとし

$$E(X_i)=\mu, \qquad V(X_i)=\sigma^2 \quad (i=1,2,\cdots,n)$$

とすれば任意の $a<b$ に対して

$$\lim_{n\to\infty} P\left(a < \frac{X_1+\cdots+X_n - n\mu}{\sigma\sqrt{n}} < b\right)$$
$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

が成立する。

3. 大数の法則・中心極限定理の重要性

大数の法則、中心極限定理の重要性については、二点挙げることができる。一つにはそれらが統計的推測の理論の基礎となっていることである。

大数の法則と中心極限定理により、標本平均 $\overline{X}$ の標本分布は、標本の大きさＮが大きくなると平均 μ 、分散 σ^2/N の正規分布に漸近することが知られている。これにより、標本平均 $\overline{X}$ とわれわれが知りたい値である母集団平均 μ との確率的な対応関係が決まるので、この性質を利用して統計的仮説検定や推定の理論が展開されている。

二つ目としてはランダム・ウォーク、ブラウン運動への応用である。ランダム・ウォーク、ブラウン

運動は物理・化学・生物などの自然科学から経済・経営などの社会科学まで広く用いられている。中でも金融工学の分野において Markowitz 理論や Black-Scholes 方程式を用いたオプション価格理論では株価の「ランダム・ウォーク仮説」が前提とされている。

4. 先行研究の整理と問題提起

4.1 先行研究の整理

中心極限定理の教授方法については、コンピュータでのシミュレーションをとおして、理解することに重点が置かれている研究が多い。

小林(2008)[(1)]では二項分布が正規分布に収束する過程をシミュレーションをとおして理解させる実践を行い、学生が確率分布（主に正規分布）の特徴を理解することができたとしている。一方で、中心極限定理の理解については十分に出来なかったと評価している。

また米澤(1995)[(4)]では、正規母集団を乱数を用いて作成し、そこから抽出したサンプルの平均の分布を描き、正規分布に収束する様子をシミュレーションをとおして視覚的に理解させようと試みている。

4.2 問題提起

特定の分布が中心極限定理によって正規分布に収束する様子はシミュレーションによって視覚的・直感的に理解できるが、“どうして正規分布とは形が違う一様分布や二項分布などの分布が試行回数Ｎを十分に大きくすることによって正規分布に近似していくのか”についてはシミュレーションだけでは理解できない。そこには独立な確率変数の和の分布がどのように求められるかという、確率の計算方法の理解が必要になる。次章では、その内容を中心としてどのように教えていくべきか述べる。

5. 教えるべき内容と方法・手順

中心極限定理を理解する上では、独立な確率変数の和（合成積）がどのように求められるかということが重要である。それは独立な確率変数の和の計算が単純に確率を足し合わせていくものではなく、確率の積をとるという確率の独特の計算方法のためである。中心極限定理を教える際には、ド・モアブル－ラプラスの定理や収束の種類などを教えていくことも重要であるが、本稿では独立な確率変数の和に焦点を絞って検討する。以下で概要を整理する。[(3)]

離散型確率変数 X 、Y が独立であるとし、その確率分布を $g(x), h(y)$ とする。和 $X+Y$ の確率分布 $k(z)$ は確率 $P(X+Y=z)$ を考えれば得られる。$X+Y=z$ となるのは $X=x$ 、$Y=z-x$ の形の加えて z になるすべての組み合わせであるから、それを確率の積で表して

$$k(z)=\sum g(x)h(z-x)$$

となる。

関数 g , h から k を作る数学操作をたたみこみという。このたたみこみについては、サイコロの例で説明されることが多い。

このたたみこみがＮ個の確率変数に対していかに計算され、正規分布に収束するかについては、特性関数を用いた証明まで学習者は理解する必要がある。

方法・手順としては、まず確率変数が２つの場合の和の計算方法について数式の意味を正確に理解させる。その際にたたみこみという数学操作について説明する必要がある。

次に、簡単な例で確率変数の和の計算を手作業で行わせる。サイコロが２個、３個、４個のときの出目の和の確率分布をすべて計算させ、グラフを作らせる。また、確率変数に１か-1 だけをとる単純ランダム・ウォークの樹形図を作成させ、その和の確率分布のグラフを作らせる。これらの作業により、一様分布や二項分布の独立な確率変数の和がどのように計算され、その確率分布が正規分布に近づいていくのかを学習者は理解できる。

最後に、特性関数を用いた数理的な証明などを紹介しつつ、中心極限定理が成り立つための条件などを確認し、中心極限定理の限界について触れるのが良い。

6. コンピュータ利用の位置づけ

大数の法則、中心極限定理を理解する際には、確率変数の和の分布について理解することが重要である。その際にたたみこみという数学操作の理解が必要である。中心極限定理についてたたみこみを中心に据えて教える場合、シミュレーションは以下の二つの部分において補助的に用いると有効である。一つは、導入部分での問題提起としての利用である。コンピュータ上でシミュレートすることで確率変数の和の分布が正規分布に収束する様子が分かるが、「なぜ和の確率分布がこのように正規分布に近づいていくのか」という問いかけし、学習者に問題意識を持たせることができる。二つ目としては、N を十分大きくしたときの和の分布は実際の手計算ではできないため、コンピュータに頼らざるをえない。数理的な理解の上で補助的にコンピュータによるシミュレーションで理解の確認を行うことが重要である。

7. おわりに

本稿においては、大数の法則と中心極限定理を教えるにあたっての教育内容についてとその方法・手順について検討した。その中で、コンピュータの利用はあくまでも補助的なものとして位置付けられ、理解のための中心的な方法とはならないと考えられる。

参考文献

(1) 小林文美子：“統計教育の教育内容 - 区間推定について - ”数学教育論文発表会論文集,第 41 巻,pp.489-494（2008）
(2) 清水良一：“中心極限定理”,教育出版（1976）
(3) 松原望：“入門確率過程”,東京図書（2003）
(4) 米澤忠幸：“コンピュータを利用した統計学の基礎教育”,数学教育研究,第 25 号,pp.67-71（1995）