

中心極限定理のクリッカーによる教室実験が効果的であるために必要なクラスサイズの評価

樋口 三郎[†]

(受付 2017 年 6 月 30 日 ; 改訂 11 月 23 日 ; 採択 11 月 27 日)

要 旨

教室内の学習者が送信した情報を即時に収集、集約して表示する装置であるクリッカーを用いて、教室で学習者が自ら標本抽出を行って中心極限定理を実証する学習活動を提案する。この活動では、学習者が形成に関与したデータを用いて学習者の関心を喚起するが、活動が効果的であるかどうかは形成される標本のサイズに依存する。このサイズは単純な場合にはクラスサイズ(学習者数)と一致する。学習が効果的となるために必要なサイズを、質問紙調査を行い項目応答理論で分析して評価した。質問紙は、多数のヒストグラムを示し、正規分布と似ていると感じるかどうかを調べるものである。その結果、考慮した特定のタイプの教室実験については、80名のクラスにおいて、中心極限定理への信頼が高まる学習者の比率は1/2を越えない程度であることがわかった。これは十分に高い比率とは言えないので、実質的な標本サイズをクラスサイズより大きくする工夫が必要である。

キーワード：クリッカー、教室実験、中心極限定理、統計教育、項目応答理論。

1. はじめに

統計学の学習において、PPDAC すなわち Problem, Plan, Data, Analysis, Conclusion という現実の事象への統計的アプローチの手順に沿った授業内活動が重要とされている。渡辺(2013)は、扱うデータとしては、(1)学習者にとって関心のあるデータ、(2)自らデータ形成に関わったデータ、(3)意外な発見が隠されているデータ、のうちの1つ以上の要件を満たすデータを題材として採用することが望ましいとしている。

従来、学習者の関わるデータ形成に使われてきた方法のひとつとして、1回目の授業で学習者に対する質問紙調査でデータを生成し、教授者がデータを整理・取捨し2回目以降の授業で例として使用する方法がある。この方法で集める学習者の属性データは、渡辺(2013)の指針(1)、(2)に沿い、学習者が関心を持つことが期待できる。

クリッカーは教室内の学習者が送信した情報を即時に収集、集約して表示する装置である(鈴木, 2009)。クリッカーは即時の情報の送信と集約が可能であるため、統計学の学習に用いるデータの形成に用いれば、PPDACを1回の授業内で完結することができる。

確率、統計の法則は自然現象に内在するものであるため、授業内で学習者が自ら行うことができる実験が多く考案されている。Galton(1894)は中心極限定理を実証する力学的装置であるGalton Boardを考案した。何森(2008)は確率が場合の数を分母とする有理数にならない場

[†] 龍谷大学 理工学部：〒520-2194 滋賀県大津市瀬田大江町横谷 1-5

合を説明するための不等辺直方体のサイコロ「サイドタ」を用いた実践を報告している。

クリッカーを用いて、自然現象や確率モデルに基づいて生成されるデータを集約して、データを学習者が自ら生成することも可能である。そのような例として、学習者が協力して標本抽出を行い大数の法則を検証する Rogers (2003) の実践を Bruff (2009) が紹介している。また、Gunderson and McGowan (2011) は、グループ学習で PPDAC すべてのステップを実行し、Data の段階ではクリッカーでクラス全体から標本抽出を行う実践を報告している。樋口 (2014) は、クリッカーを用いて、母集団と標本、標本抽出、推定といった統計学の概念を説明する実践を報告している。

このノートでは、確率論の重要な定理である中心極限定理を実証するデータ形成を含む新しい学習活動を提案する。それが効果的であるために必要な学習者数(クラスサイズ)を、質問紙調査の結果を項目応答理論で分析することにより評価する。

2. 提案する学習活動

2.1 学習活動の対象と学習目標

この活動は、学習者がクリッカーで自ら形成したデータを用いて中心極限定理が成立する例を提示し、学習者が定理に納得し、その運用を直観的に理解することを意図してデザインされている。

独立同分布に従う確率変数 X_t ($t = 1, 2, \dots$) に対して、確率変数 Y_t を $Y_t = \sum_{s=1}^t X_s$ で定める。 Y_t を標準化した Z_t の分布が $t \rightarrow \infty$ の極限で標準正規分布に近づくというのが中心極限定理の主張である。

以下に述べる学習活動の学習目標は、独立同分布に従う確率変数の和の分布を中心極限定理を適用して予想できることである。対象は、確率変数の平均と分散を計算できるが、中心極限定理の主張を納得していないないために、上のような予想ができない学習者である。

2.2 クリッカーの機能の前提

ここでは、次の機能を持つクリッカーを考える。

- (1) n_{button} 個のボタンを備えており、学習者は、教授者の設定した時間帯に、選択肢 $0, 1, 2, \dots, n_{\text{button}} - 1$ の中から一つを選択してクリッカーで送信して投票することができる。
- (2) 教授者は、すべての学習者の送信の結果を、コントローラソフトウェアを用いて、投票の合計数と棒グラフとして学習者に提示することができる。

2.3 中心極限定理の実験

学習者がクラスでサイコロやコインを使ってデータを形成できる確率変数 X_t を考える。以下では、確率変数 X_t はベルヌーイ分布 $B(1, p)$ にしたがるものとする。教室には、 m 人の学習者がいるものとする。

あらゆるすべての Y_t の値をクリッカーで送信できるためには $0 \leq Y_t \leq n_{\text{button}} - 1$ である必要がある。そのため、確率変数 X_t の分布と Y_t の定義より $0 \leq Y_t \leq t$ が成立することに注意して、 $0 \leq t \leq n_{\text{button}} - 1$ の範囲に t を選んでおく。

- (1) 学習者が、 Y_t の従う分布を予想し、教授者の提示した選択肢から選んでクリッカーで投票する(図 1)。
- (2) 学習者が組 (X_1, X_2, \dots, X_t) を生成し、 Y_t を計算する。

実験

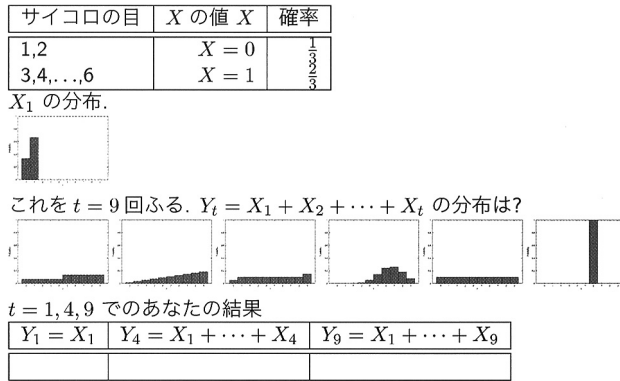


図 1. Y_t の分布を問う多肢選択問題. 正解 ($m \rightarrow \infty$ での収束先)は左から 4 番目.

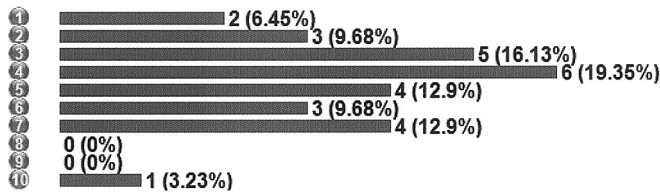


図 2. Y_9 の標本のヒストグラムのクリッカーコントロールソフトウェアによる表示. ただし選択肢 10 は $Y_9 = 0$ に対応する.

- (3) クラス内の学習者が Y_t をクリッカーで送信する.
これにより, Y_t の無限母集団からサイズ m の標本 $\mathbf{y} = (y_1, \dots, y_m)$ が抽出される.
このデータ \mathbf{y} のヒストグラムを学習者に提示する (図 2).
- (4) 学習者が, データ \mathbf{y} のヒストグラム $h(\mathbf{y})$ と, 標準化されていない正規分布 $N(tp, tp(1-p))$ の確率密度関数とが「似ている」ことを確認する.
- (5) いくつかの t についてこれを繰り返し, Y_t の平均と分散が一定の法則で変化することを確かめる.
- (6) 教授者が, これが中心極限定理の帰結であることを説明する.

中心極限定理の主張は極限 $t \rightarrow \infty$ についてのものであり, 有限の t , 有限のサイズ m の 1 個の標本のヒストグラム $h(\mathbf{y})$ について何かを言っているわけではない. また, 「似ている」ことの基準も学習者に任せている.

しかし, 統計学の教科書には, これと似た方法, すなわち Y_t を標準化した Z_t の標本を擬似乱数を用いて抽出し, そのヒストグラムが t とともにどのように変化するかを図で示すことにより, 中心極限定理の意味を説明しているものが多い. このような直観的な説明は統計教育において一定の支持を得ているものと考えられる.

3. 質問紙調査と分析

3.1 信頼が高まるとき

以下では, データを生成する人数, すなわちヒストグラム $h(\mathbf{y})$ を描く標本 \mathbf{y} のサイズ m と,

そのデータから描かれたヒストグラムを見て反応する被験者の数 N とが一般には異なるとして分析する。

この活動が効果的であるかどうかは、中心極限定理の理解や運用についての事前・事後テストで判定できる。しかし、効果的であるためには、学習者が、 \mathbf{y} のヒストグラムと正規分布が似ていると感じて、中心極限定理への信頼が高まることが必要条件となる。

同一のヒストグラムを与えられたときにも、中心極限定理への信頼が高まるかどうかは、環境のノイズや学習者の短期的な内的状態に依存し、確率的に定まると考えられる。また、その確率は学習者の個性や学習歴により異なると考えられる。このような被験者 i ($i = 1, \dots, N$) の状態を表す潜在変数を ϕ_i としたとき、ヒストグラム $h(\mathbf{y})$ を見ることによって被験者 i の中心極限定理への信頼が高まる確率を $f_{\text{response}}(1|h(\mathbf{y}), \phi_i)$ とする。

標本抽出で \mathbf{y} を得る確率を $f_{\text{sample}}(\mathbf{y})$ とすると、1 回の標本抽出後に、被験者 i の中心極限定理への信頼が高まる確率は、

$$f_{\text{success}}(\phi_i) = \sum_{\mathbf{y}} f_{\text{response}}(1|h(\mathbf{y}), \phi_i) \cdot f_{\text{sample}}(\mathbf{y})$$

となる。

確率 $f_{\text{sample}}(\mathbf{y})$ は、教科書に載せる図を擬似乱数により描くときは問題にならないが、教室実験の場合には問題となる。教科書に図を載せる場合には、生成した $h(\mathbf{y})$ がたまたま正規分布と大きく異なった場合(たとえば、標本 \mathbf{y} に $(t+1)$ 種類の値が同数含まれる、1 個だけ含まれる、などの極端な \mathbf{y} を得る確率 f_{sample} も 0 ではない)、そのデータを捨てて再度生成して作図すればよい。しかし、クラス内で実験を行ってデータを生成する場合には、再実験を多くの回数くりかえすことは時間の制約から困難であり、また、学習活動の効果を低下させる可能性がある。そこで、信頼が高まる被験者の比率をあらかじめ評価しておく必要がある。

大数の法則から、 Y_i の標本サイズすなわち \mathbf{y} のデータの個数が大きいほど 2 つのヒストグラムは実際近くなるので、 m が大きいときにこの和 $f_{\text{success}}(\phi_i)$ は大きいと予想できる。確率 $f_{\text{success}}(\phi_i)$ が標本サイズ m に対してどのくらいの大きさかを知ることが以下で扱う問題である。

3.2 信頼の高まる学習者の比率

確率 f_{response} は被験者の心理に関わる。確率 f_{response} の性質を知るために、大学生の被験者に対して次の質問紙調査を行った。以下の調査では、 $n_{\text{button}} = 10$, $t = 9$, $p = 2/3$ と固定した。

サイズ $m = 40, 60, 80, 100, 200$ の標本 \mathbf{y} を、 $p = 2/3$ の二項分布 $B(t, p)$ から擬似乱数を用いて各 $n_{\text{hist}} = 20$ 個生成し、それぞれヒストグラム $h(\mathbf{y}_j)$ を描いた ($j = 1, 2, \dots$)。質問紙(図 3)で、正規分布のヒストグラムと、ヒストグラム群 $h(\mathbf{y}_j)$ を示し、各 $h(\mathbf{y}_j)$ に対して次の質問(図 3)に、はい、いいえの 2 択で回答するように求めた。

ある人が、「この実験は確率的なものなので、毎回結果は変わるけど、理論 A によれば、実験の結果はだいたいこんな感じ(正規分布のヒストグラム)のようになるはず」と言いました。

実験を 1 回だけしたところ、下のような結果 ($h(\mathbf{y})$) になりました。

設問

この結果を見て、あなたの理論 A に対する信頼は増したでしょうか、増さなかったでしょうか。正解というものはありません。人間の感じ方についての調査です。

理論 A に対する信頼が増したときは、○をつけてください。

なお、比較対象となる正規分布のヒストグラムは、平均と分散の一致する確率密度関数を階

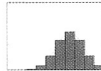
樋口三郎 2013-01-09
 使える統計!での授業方法改善のための調査にご協力お願いします。お名前を書きいただきますが、回答内容は成績には関係しません。

学籍番号	お名前
2012 年末のサイコロを 9 回振って結果をクリッカーで集める実験に参加しましたか (○つけてね)	した しなかった

状況設定

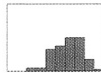
ある人が、「この実験は確率的なもので、毎回結果は変わるけど、理論 A によれば、実験の結果はだいたいこんな感じ (下の図) のようになるはず」と言いました。

理論



実験を 1 回だけしたところ、下のような結果になりました。

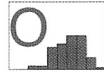
実験



設問

この結果を見て、あなたの理論 A に対する信頼は増したでしょうか、増さなかったでしょうか。正解というものはありません。人間の感じ方についての調査です。

理論 A に対する信頼が増したときは、o をこんな感じでつけてください。



理論 A に対する信頼が増さなかったときは、x をこんな感じでつけてください。



これを、裏の 64 個の結果すべてに対して行ってください。図は 64 個書いてありますが、これらと比較するのではなく、それぞれ 1 個だけ見た気持ちになって 64 回繰り返してください。

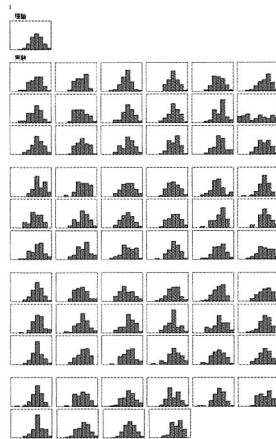


図 3. 質問紙。上(表)設問。下(裏)ヒストグラム群のランダムな配置を 4 ページ作成し、回答者にいずれか 1 ページをランダムに割り当てた。上の 1 個が「理論」、下の 64 個が「実験」。

級ごとに積分して描いた。

表 1 の 3 個の被験者集団(人数 N)に対し、それぞれ m の異なるデータセット A-E のうち複数個を選択して、それらに含まれるヒストグラムの集合をランダムな順序で 1 つの質問紙に表

表 1. 質問紙調査の対象と使用したデータセット.

被験者集団 (年度-学年)	被験者数 N	データセットと標本サイズ m					
		A	B	C1	C2	D	E
		40	60	80	80	100	200
1(2012-1)	65	○	○	○			
2(2013-1)	83	○	○		○		
3(2014-3)	52				○	○	○

示して回答を求めた. 選択したデータセットの個数を n_d とすると, 表示するヒストグラムの個数は $n_d \times n_{\text{hist}}$ である. 今回は $n_d = 3$ としたため, 2×60 個のヒストグラムとなる. さらに, 回答状況の確認のため, 一様分布のサンプルや, 正規分布のサンプルなどのヒストグラム 4 個を加えて 64 個とした(図 3).

3 個の被験者集団はいずれも 4 年制大学生からなるが, 学年や特徴が異なる. 集団 1 は 2012 年度, 2 は 2013 年度の, 理系文系学部両方からなる統計学科目の主に 1 年生の受講生からなる. 集団 3 は 2014 年度の理系学部の科学計算科目の主に 3 年生の受講生からなる. 一方, データセット C1, C2 はともに $m = 80$ だが, 異なるヒストグラム群である.

3.3 項目応答理論によるモデル化

各被験者集団に対する質問紙調査により, $N \times (n_d \times n_{\text{hist}})$ の 2 値の反応パターンが得られる. この反応を, テスト理論の項目応答理論(豊田, 2014)を用いてモデル化する. ヒストグラム h をアイテム(問題), 被験者をテスト受験者に対応させ, 信頼が高まったという回答を正解と考える. 2 パラメタロジスティックモデルを採用し, 学習者を記述する実パラメタ $\theta(\phi_i)$, ヒストグラムを記述する実パラメタ $a(h), b(h)$ で確率 f_{response} が次のように書けることを仮定する.

$$(3.1) \quad f_{\text{response}}(1|h, \phi_i) = \frac{1}{1 + e^{-a(h)(\theta(\phi_i) - b(h))}}.$$

パラメタ θ は被験者がヒストグラムが正規分布に「似ている」と考える傾向を表す特徴量, b は正規分布との似ていなさを表すヒストグラム h の特徴量, a は被験者によって判断の分かれる程度を表すヒストグラムの特徴量と解釈できる.

3.4 パラメタの推定方法

被験者集団 1, 2, 3 の反応パターンに対して, 統計ソフトウェア R の `irt` や `ltm` パッケージを利用して分析を行った. それぞれ, 周辺尤度最大化によりパラメタ $\theta_i, a(h_j), b(h_j)$ を推定した ($i = 1, \dots, N, j = 1, \dots, n_d \times n_{\text{hist}}$). 被験者集団 1, 2 はデータセット A, B, 被験者集団 2, 3 はデータセット C2 を共通項目として持つため, 集団 2 を基準として等化を行い, 共通の尺度による $\theta_i, a(h_j), b(h_j)$ を与えた. 方法としては共通項目法の mean-sigma 法を用いた.

3.5 ブートストラップ法による信頼が高まる比率の推定

パラメタ θ_i が標準正規分布にしたがうと仮定される被験者集団 2 に, パラメタ $a(h), b(h)$ を持つヒストグラム h を提示したとき, 信頼が高まる被験者の比率は積分

$$r(h) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} \frac{1}{1 + e^{-a(h)(\theta - b(h))}} d\theta$$

で与えられる. データセットに属する n_{hist} 個のヒストグラム h の $r(h)$ を母集団とみなして,

ブートストラップ法により、この比率 r の信頼区間を求めた(標本サイズ $m = 80$ の2個のデータセット C1, C2 は併合して $n_{\text{hist}} = 40$ として扱った)。

3.6 結果

推定結果 $b(h_j)$ を、標本サイズ m ごとに $b(h_j)$ の大きさにソートして j を再付番し、プロットした結果を図4に示す。ただし、横軸は $\rho(j) = j/n_{\text{hist}}$ である。パラメタ $b(h)$ の大きさは、 $m = 60, 80, 100$ では差がほとんどないが、 $m = 40, m = 200$ まで見ると、 m が増加すると $b(h)$ が減少する傾向があることがわかる。

モデルの枠内では、 $b(h) = 0$ であるヒストグラムに対しては、等化の基準となる被験者集団2の中央値 $\theta = 0$ の被験者が確率 $1/2$ で「信頼が高まる」と答える。より小さい $b(h)$ ではこの比率はより大きくなる。この比率が $1/2$ 以上となるヒストグラムが得られる確率は、 $b(h) = 0$ となる $\rho(j)$ で与えられ、 $m = 40$ で 0.1 、 $m = 80$ で 0.4 、 $m = 200$ で 0.7 程度と読み取れる。

ブートストラップ法により求めた比率 r の m 依存性を図5に示す。ここでも m が増加すると比率が増加する傾向が観察できる。学習者が m 人のクラスで実際に活動を実行するかどうかは、比率の信頼区間の下限の値を許容するかどうか、という保守的な立場で行うことになる。

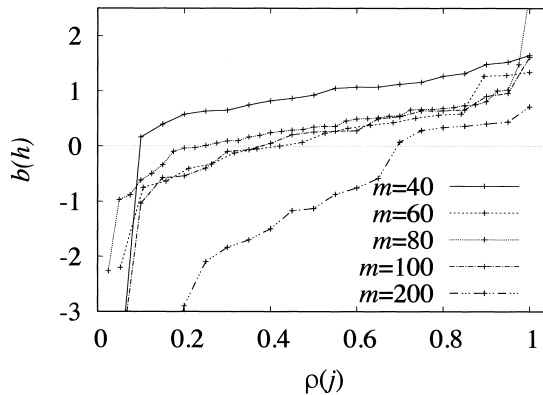


図4. 困難度母数 $b(h)$ の推定の結果.

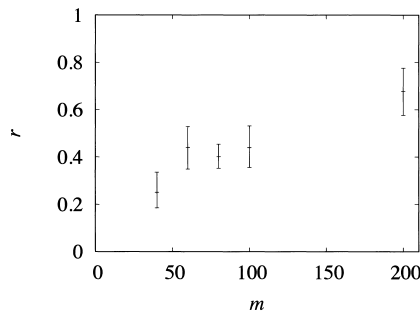


図5. ブートストラップ法で評価した、 N_{vote} ごとの、信頼が深まる学習者の比率. 信頼係数 0.95 の信頼区間.

4. 考察と結論

図4, 5より, m が大きくなるとともに, 中心極限定理への信頼が増す学習者の比率が高い傾向が読み取れる. これは, 第3.1節で述べたとおり, m が大きいと, 得るヒストグラムと正規分布からの標本のヒストグラムとが近くなるためと考えられる. 80人のクラスで1人1回標本抽出する $m = N$ の場合, 「信頼が高まる」学習者が半数を超えないことを示しており, 授業中の活動としては, 影響を与えられる学習者の比率が低い. 提案した活動が実用可能であるためには, m を大きくとる必要がある. 具体的な値は, 要求する「信頼が高まる」確率ごとに図4, 5から読み取れる.

小さなクラスでこの活動を効果的にする方法として, クラスの学習者数よりも m を大きくとるような活動の改良が考えられる. 具体的には, 1人の学習者が複数個のサンプルを形成し送信する, 他のクラスや過去の年度のクラスで形成されたデータと併合する, などの方法をとれば, 学習者自らが生成に関わったデータでありながらクラスの規模に制限されずに m を大きくすることができる. また, あらかじめ擬似乱数などで形成しておいたデータと併合することも考えられる.

一方, 図4, 5からは, $m = 60, 80, 100$ で信頼が増す学習者の比率が大きく変わらず, m に關する単調性があるかどうかは確かではない. この振る舞いの理由は明らかではないが, 第2.3節で述べた通り, ヒストグラムが「似ている」ことの基準を学習者に任せたことに起因するかもしれない. 学習者のヒストグラムの近さの基準は, 大数の法則の漸近性と対応する幾何学的意味での曲線の距離だけでなく, 形状の特徴, 例えば曲線が単峰か多峰か, サポートの個数は何個かなども影響されている可能性がある. このとき, m が小さく単純な形のヒストグラムがより「似ている」と感じる効果が想定できる. これが, 大数の法則による増加の効果を, 特定の m の範囲で打ち消しているのかもしれない. この振る舞いの理由を明らかにするには, より詳細な実験が必要と考えられる.

この学習活動では, 有限の m の標本を1つとり, そのヒストグラムと $m = \infty$ のヒストグラムとを目視で比較して, 法則に対する信頼を高めることを期待している. Galton Boardの使用でもこのような方法がとられるが, 統計学において標本の法則への適合を論ずる方法ではない. より進んだ学習者であれば, 与えられたヒストグラムに対してコルモゴロフ-スミルノフ検定やシャピロ-ウィルク検定を行い, データが法則に従うという帰無仮説を棄却するかどうか判断するだろう. より学習が進んだときには, このような見方で活動を振り返ることが望ましい.

ここで得た必要な m の具体的評価は, t, p , ヒストグラムの描き方などに依存する. これらを改善することで, 信頼が高まる比率を高めることができる可能性もある. 本ノートで述べた, 投票でなく擬似乱数を用いて提示する標本を生成し, 質問紙調査と項目応答理論で解析する方法は, 異なる t, p や, 他の設定での教室実験にも広く適用可能である.

ここでは効果的であるために必要な m の大きさを調べたが, 実際に学習目標を達成するのに効果的にはたらくことを示すには, 学習目標に関する学力の事前・事後テストで検証を行うことが必要であり, これが将来の課題である.

参 考 文 献

- Bruff, D. (2009). *Teaching with Classroom Response System*, Wiley, San Francisco.
- Galton, F. (1894). *Natural Inheritance*, Macmillan, London.
- Gunderson, B. K. and McGowan, H. M. (2011). Using clickers in a statistics classroom, *Teaching Mathematics with Classroom Voting with and without Clickers* (eds. K. Cline and H. Zullo), MAA Notes 79, 53–60, The Mathematical Association of America, Washington, DC.
- 樋口三郎 (2014). 統計教育でのクリッカー・教室応答システムの利用, 統計教育実践研究 第6巻, 統計数理研究所共同研究レポート, No.313, 17–18, 統計数理研究所.
- 何森 仁 (2008). サイドタ, ドタドタ! — デタラメの法則を知るために振って振って振りまくる!, 数学セミナー, 47(12), 51–55.
- Rogers, R. (2003). Using personal response system to engage students and enhance learning, *Making Statistics More Effective in Schools and Business Conference*, Georgetown University, Washington, DC.
- 鈴木久男 (2009). クイズで授業を楽しもう, 『学生と変える大学教育』(清水亮, 橋本勝, 松本美奈 編), 166–183, ナカニシヤ出版, 京都.
- 豊田秀樹 (2014). 『項目反応理論 [入門編]』, 朝倉書店, 東京.
- 渡辺美智子 (2013). 知識基盤社会における統計教育の新しい枠組み: 科学的探究・問題解決・意思決定に至る統計思考力, 日本統計学会誌, J42(2), 253–271.

An Estimation of Minimum Class Size for an Effective Classroom Experiment on the Central Limit Theorem Using Clickers

Saburo Higuchi

Faculty of Science and Technology, Ryukoku University

We examined a classroom activity in which learners perform sampling and validate the central limit theorem using clickers or a classroom response system. In this activity, learners are motivated by a dataset that they generate themselves. The effectiveness of the activity, however, depends on the sample size, which is equal to the number of learners in the class. The required size is estimated from questionnaire data by making use of item response theory. It turns out that in classes that consist of 80 learners, no more than half could convince themselves that the theorem is reliable. Therefore, it is important to seek an activity design for which a larger sample is available.