高校数学における確率分布の教育について
—二項分布を中心に—

2018SS067 冨田紗羅
指導教員：小藤俊幸

1　はじめに

高等学校では，昨年度から，新学習指導要領のもとでの数学教育が実施されている．「数学 B」の「ベクトル」は「数学 C」に移行され，代わりに「数学 B」には，新たな単元「数学と社会生活」が導入された．2025（令和 7）年度からの大学入学共通テストでは,『数学 II，数学 B，数学 C』の出題範囲が『「数学 B」及び「数学 C」については，数列（数学 B），統計的な推測（数学 B），ベクトル（数学 C）及び平面上の曲線と複素数平面（数学 C）の 4 項目に対応した出題とし，4 項目のうち 3 項目の内容の問題を選択解答する。』と定められている．

「数学 B」の「統計的な推測」は，これまでのように「あまり多くの生徒が学ばない」単元ではなくなったが，実際に教育が行われた実績が乏しく，その指導には，さまざまな困難が付きまとうと考えられる．ここでは,「統計的な推測」の中でも，重要な題材の 1 つであると考えられる二項分布を取り上げて，教育における問題点を指摘する．

2　大学入学共通テストの問題例

2023 年度の大学入学共通テスト『数学 II，数学 B』で，次の趣旨の問題が出題されている（選択問題 3 の (2)）．共通テストでは，受験生の読解力を試す意図と思われる独特の出題形式が取られている．問われている数学的な内容が明確になるように問題文を全面的に書き直した．

問題　ある生産地で生産されるピーマン全体を母集団とし，この母集団におけるピーマン 1 個の重さ（単位は g) を表す確率変数を X とする．X は正規分布 $N(30.0,\ 3.6^2)$ に従うとする．k を自然数とし，ピーマンを無作為に $50+k$ 個抽出したとき，そのうちの 30g 以下のピーマンの個数を表す確率変数を U_k とする．そのとき，U_k が $25 \le U_k \le 25+k$ となる確率が

$$P(25 \le U_k \le 25+k) \ge 0.95 \tag{1}$$

となるような最小の k を求めよ．

もとの問題文の誘導に従うと，解答は以下のようになる．

解答　U_k は二項分布 $B\left(50+k,\ \dfrac{1}{2}\right)$ に従う．$50+k$ は十分に大きいので，U_k は近似的に $N\left(\dfrac{50+k}{2},\ \dfrac{50+k}{4}\right)$ に従い，$Y = \dfrac{U_k - \dfrac{50+k}{2}}{\sqrt{\dfrac{50+k}{4}}} = \dfrac{2U_k-(50+k)}{\sqrt{50+k}}$ は，近似的に標準正規分布 $N(0,\ 1)$ に従う．$25 \le U_k \le 25+k \Leftrightarrow -\dfrac{k}{\sqrt{50+k}} \le Y \le \dfrac{k}{\sqrt{50+k}}$ であるから，$\dfrac{k}{\sqrt{50+k}} \ge 1.96$（標準正規分布の上側 2.5% 点）であれば，(1) が成り立つ．簡単のため，1.96 を 2 でおきかえる（もとの問題文でそのように誘導されている）．$\dfrac{k}{\sqrt{50+k}} \ge 2$ から，$k^2 \ge 4(50+k)$，すなわち，$k^2-4k-200 \ge 0$ が得られ，$k>0$ より，$k \ge 2+2\sqrt{51}$ となる．問題文で，$\sqrt{51} \fallingdotseq 7.14$ の近似値を使うことが指示されていて，$k \ge 2+2\times 7.14 = 16.28$ となることから，求める最小の自然数は $k=17$ となる．

3　二項分布の正規分布による近似

解答は，二項分布の次の性質（[2], p. 159) に基づいている．

二項分布 $B(n,\ p)$ に従う確率変数 X は，n が大きいとき，近似的に正規分布 $N(np,\ npq)$ に従う．ただし，$q=1-p$ である．

上記の性質に対応する定理は，ド・モアブル-ラプラスの定理と呼ばれている（[3], pp. 37-41）．証明は大変難しい．前節の問題の解答を理解するためには，それ以外にも，少なくとも，次のような事柄を知っている必要があると思われる．

(1) X を二項分布 $B(n,\ p)$ に従う確率変数とするとき，その平均 $E(x)$ と分散 $V(E)$ は $E(X)=np$, $V(X)=npq$ となる．

(2) X を正規分布 $N(m,\ \sigma^2)$ に従う確率変数とするとき，

$$P(a \le X \le b) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}}dx \tag{2}$$

(3) X を正規分布 $N(m,\ \sigma^2)$ に従う確率変数とするとき，確率変数 $Z=\dfrac{X-m}{\sigma}$ は標準正規分布 $N(0,\ 1)$ に従う．

このうち，高校 2 年生の段階で，きちんとした説明が与えられるのは，**(1)** だけである．**(2)** のきちんとした説明には，高校数学の範囲外である広義積分の概念が必要になるが,『数学 II』の微分積分では，対象となる関数が，そもそも，多項式に限定されている．**(3)** の説明には，置換積分が必要であるが，これは『数学 III』の学習内容である．多くの高校 2 年生は，数学的な基礎知識が乏しいまま，問題の解法を「暗記」して，試験に臨むのではないかと思われる．

前節の問題の $P(25 \le U_k \le 25+k)$ は，U_k を正規分布で近似しないで，二項分布のままで扱うと，

$$P(25 \le U_k \le 25+k) = \sum_{j=25}^{25+k} {}_{50+k}\mathrm{C}_j \left(\frac{1}{2}\right)^{50+k} \tag{3}$$

と表される．この式の近似値は，C 言語や C++言語の多くのコンパイラでサポートされている数学関数 lgamma（ガンマ関数の対数を与える関数）を用いると，容易に計算することができる．$k = 0, 1, \ldots, 17$ について (3) の値を計算すると，下の表のようになる．

k	0	1	2	3	4	5
P	0.112	0.220	0.322	0.417	0.503	0.581

k	6	7	8	9	10	11
P	0.650	0.711	0.763	0.807	0.845	0.876

k	12	13	14	15	16	17
P	0.902	0.923	0.940	0.954	0.964	0.973

したがって，(1) が成り立つ最小の自然数は $k = 15$ である．前節の解答では，上側 2.5% 点 1.96 を 2 でおきかえて，最小の k を求めているが，これを 1.96 に戻すと，k の条件は $k \geq 15.91$ となり，答えは $k = 16$ に変わる．いずれにしても，正規分布による近似を利用した解は「真の解」$k = 15$ に近い値になっている．この問題は，二項分布の正規分布による近似の有効性を示すために活用できるのではないかと思われる．

4　復元抽出と非復元抽出

大学入学共通テストの問題には，留意すべき点がもう 1 つある．問題に現れるピーマンの抽出は，非復元抽出と考えられる点である．『数学 B』の教科書には，以下のように述べられている（[2], p. 138).

> 母集団の中から標本を抽出するのに，毎回もとに戻しながら次のものを 1 個ずつ取り出すことを**復元抽出**という．これに対して，取り出したものをもとに戻さずに続けて抽出することを**非復元抽出**という．
>
> − 中略 −
>
> 非復元抽出の場合でも，母集団の大きさが標本の大きさ n に比べて十分大きい場合には，非復元抽出と復元抽出との違いは小さくなる．
>
> したがって，非復元抽出による標本も近似的に復元抽出による標本とみなすことができ，X_1，$X_2, \cdots\cdots, X_n$ は，それぞれが母集団分布に従う互いに独立な確率変数と考えてよい．

§2 の解答では，このことに基づいて，「U_k は二項分布 $B\left(50 + k,\ \dfrac{1}{2}\right)$ に従う」としている．しかし，非復元抽出による標本が従うのは，厳密には，二項分布ではなく，**超幾何分布**である（[4], pp. 49-52）．『数学 B』の学習内容には「確率変数の独立」も含まれていることから（[2], pp. 134-149)，このことにも触れる必要があるのではないだろうか．

大きさ N（個）の母集団に，ある性質をもつ要素が K 個，もたない要素が L 個含まれる（$N = K + L$）とする．n を $n \leq K, n \leq L$ を満たす自然数とし，母集団から大きさ n（個）の標本を無作為抽出したとき，そのうち，ある性質をもつ要素の個数を U とすると，$j = 0, 1, \ldots, n$ に対して，$P(U = j) = \dfrac{{}_K\mathrm{C}_j \cdot {}_L\mathrm{C}_{n-j}}{{}_N\mathrm{C}_n}$ が成り立つ．このとき，U は超幾何分布に従うという．

この式の値も（K や L が極端に大きくなければ）数学関数 lgamma を使って計算ができる．$k = 12$（$n = 50 + k = 62$）の場合について，母集団の大きさを $K = L = 80, 100, \ldots$ のように変えて，(**??**) によって，$P(25 \leq U \leq 25 + k)$ を計算すると，下の表のようになる．

K	80	100	250	500	750
P	0.9655	0.9536	0.9229	0.9125	0.9090

K	1000	2500	5000	7500	10000
P	0.9072	0.9041	0.9031	0.9027	0.9026

母集団の大きさが小さければ，$k = 12$ でも　(1) が成り立つ．母集団の大きさが大きくなるにつれて，$P(25 \leq U \leq 25 + k)$ は，(3) を使って求めた確率に近づいて行く．

5　大学入試の傾向

2015 年〜2022 年のセンター試験，共通テストの本試験，追試験過去問の内容を調べた．調べた結果，下の表のようになった．内容は，

A: 二項分布の正規分布による近似
B: 中心極限定理　　C: 復元抽出
D:1 次式で表される確率密度関数

である．

内容	A	B	C	D
問題数	10	2	1	1

このように、8 年間でほとんど毎年のように正規分布による近似の問題が出題されている．他の数学の問題にも共通するかもしれないが，共通テストの「確率分布と統計的な推測」では，計算力はそこまで求められなく，推定などの仕組みを理解しているかなどを問われることが多い．

二項分布の基本的な概念を理解し，実際の問題に対して適切に応用する能力を身につけなければならない．

6　おわりに

センター試験，共通テスト等で問われている事項は全て教科書で扱われている基本的な公式，考え方で解答できてしまうレベルである．「数学 B」の「確率分布と統計的な推測」は学校の授業では扱われない場合が多い．対策としては，まず，教科書を丁寧に読んで，用語，公式，考え方を理解すること．基本事項を正しく覚えて理解を深めておくことが，特にこの分野で肝心なことであると考える．

参考文献

[1] 大学入学共通テスト『数学 II，数学 B』, 2023.

[2] 坪井俊 他 12 名：『数学 B』，数研出版，2012.

[3] Ya.G. シナイ (森真訳)：『シナイ確率論入門コース』，シュプリンガー・フェアラーク東京，1995.

[4] 鈴木義一郎：『例解標本調査』，実教出版，1981.