数学Ｂ「統計的な推測」の指導の難しさ
～社会的な期待と数学的な立場の葛藤～

藤田　祥一
東京都立葛飾野高等学校

2023/11/30
令和5年度東京都高等学校数学教育研究会　授業研究会（第100回）

略歴

- 2017～2019　城西大学大学院理学研究科
- 2019～　　　明治大学大学院先端数理科学研究科

- 2021年4月～2023年3月　上野学園中学校・高等学校　非常勤講師
- 2023年4月～現在　東京都立葛飾野高等学校　教諭（臨時的教員）
- 日本数学教育学会 実践研究推進部高等学校部会 幹事
- 日本数式処理学会 教育分科会 運営委員

<u>統計との関わり</u>

- 修士のとき，クラスタリング分析（ウォード法，k-meansなど）を学習
- 博士のときは研究室がゲームAIの研究をしていた

今回の話題

- 新課程になって，数学Bでは「統計的な推測」をほぼ確実に指導することとなった
- これまで統計を指導したことがない・統計学を専門的に学んだことがない教員でも統計を教える
- 一方で，統計は数学とは考え方が異なる面も多く，指導しづらいという声も聞く

⇓

数学Bを人生初めて指導する藤田が授業をする中で少々困ったことを共有して，「統計的な推測」を現在指導されている方と来年度以降指導される方とで統計教育の充実を図るための意見交換をする

新課程での統計分野の変更点

- 数学I「データの分析」
  - 「四分位範囲」「箱ひげ図」が中学2年生へ移行
  - 「外れ値」という用語が出現
  - 「仮説検定の考え方」という項目が追加

- 数学B「統計的な推測」
  - 「信頼区間」「有意水準」という用語が指導要領に明確に記載される
  - 「正規分布を用いた仮説検定の方法」という項目が追加

高等学校で統計教育が充実した理由(Chat GPT抜粋)

- 社会のデータ駆動型化
  - 統計は意思決定や問題解決に不可欠なスキルとなっている.
- STEM教育の強化
  - 高等学校で統計を教えることで生徒は数学的な思考力を養える.
- インターネットとデジタルテクノロジーの普及
  - 高校生は日常生活でデジタルデータに接する機会が多く，
それを活用できるようにするために統計が重要.
- 大学進学への準備
  - 高校での統計教育が大学での学問や研究における基盤を築く役割を果たす.
- 問題解決力の養成
  - 高等学校で統計を学ぶことで，生徒はリアルの問題に対してデータに基づいたアプローチを取る能力を養うことができる.
- 科学的思考の促進
  - 高等学校での統計教育は，科学的思考を促進し，批判的思考を養う.

平成28年12月の答申では

幼稚園、小学校、中学校、高等学校及び特別支援学校の
学習指導要領等の改善及び必要な方策等について（答申）

教科等を越えた全ての学習の基盤として育まれ活用される資質・能力

また、急速に情報化が進展する社会の中で、情報や情報手段を主体的に選択し活用していくために必要な情報活用能力、物事を多面的・多角的に吟味し見定めていく力、統計的な分析に基づき判断する力、問題を見いだし解決に向けて思考するために必要な知識やスキルなどを、各学校段階を通じて体系的に育んでいくことの重要性は高まっていると考えられる。

求められる改善点

数学の学びを社会生活で活用する場面として、統計に関する学習を充実させていくことが重要である。

学習指導要領解説では(改訂の趣旨より)

高等学校数学科においては，数学的に考える資質・能力の育成を目指す観点から，現実の世界と数学の世界における問題発見・解決の過程を学習過程に反映させることを意図して数学的活動の一層の充実を図った。

また，社会生活などの様々な場面において，必要なデータを収集して分析し，その傾向を踏まえて課題を解決したり意思決定をしたりすることが求められており，そのような資質・能力を育成するため，統計的な内容等の改善・充実を図った。

このような改訂の方向は，現在，米国等で推進が図られているSTEM教育の動きと同一の方向であると考えられる。

AI戦略2019（内閣府）では

教育改革，リテラシー教育【高等学校】

全ての高等学校卒業生が、データサイエンス・ＡＩの基礎となる理数素養や基本的情報知識を習得。また、人文学・社会科学系の知識、新たな社会の在り方や製品・サービスのデザイン等に向けた問題発見・解決学習を体験

- 現職教員のデータサイエンス・ＡＩリテラシー向上のための学習機会の提供
- ＩＴパスポート試験等を高等学校等における活用の促進
- 全ての高等学校で、データサイエンス・ＡＩの基礎となる実習授業を実施
- 意欲的な児童・生徒に対するデータサイエンス・ＡＩで問題発見・解決に挑戦する場の創出

※ AI戦略2022でも同様に目標として掲げられている

社会から求められる統計教育の充実

- 問題解決や意思決定の道具として，統計の知識を活用
- 日常生活に合わせた統計リテラシー修得

統計の内容や仕組み　＜　統計をどう使うか

数学B「統計的な推測」の教科書を見ると

- 定義や定理（証明付き）で理論ベースの作りとなっているものが多い
  - 日常生活での活用例は正直少ない
  - 前の学習指導要領と教科書の構成や題材もほぼ変わっていない（仮説検定の単元は新たに追加されたが）
  - 理論を追わないと，公式ありきの当てはめゲー・作業ゲーに陥る

※　社会的な期待とは少々違うような・・・？

数学を専門とする教員の立場からすると

- 数学を学ぶ以上，なぜその考えが正しいのか追求することが大切
  - 社会的な期待とは離れるが
- 統計にはベクトルや積分の考え方を多く含むため，その点も含めて数学では統計を扱うべき
  - 積分は数IIの最後の方で，ベクトルは数Cに行ってしまった
  - 「統計的な推測」を扱う前にはどちらも未履修・・・
- 現実的な活用については教科書や学習指導要領解説に詳しく書かれていないため，教員の努力によって教材を作らないといけない
  - 統計専門ではない教員が日常生活を題材とした意味ある教材を作れるか
  - 大学入試では日常生活が題材のものが出されそうなんだが・・・

藤田の葛藤（授業を考える上でのモチベ）

社会的な期待

統計を用いた問題解決・意思決定，活用中心

↑

↓

教科書

理論を中心とした解説と演習

相反するような状況の中で，「教科書で教える」立場に立ったとき

- 理論と活用どちらに重きをおけばいいのか？
- 理論と活用のバランスをどう取ればいいのか？
- 統計教育に貢献するために数学科として授業内でどうすればいいのか？

藤田の指導上の困り

- ここからは，藤田が指導をする中で困ったことを報告する
- 一部，藤田が自作した演習問題も紹介する
- また，都数研数学I分科会内で議論になったことも紹介する

使用教科書：数研出版　高等学校　数学B

「統計的な推測」の項目

- 確率変数と確率分布
- 確率変数の期待値と分散
- 確率変数の和と積
- 二項分布
- 正規分布
- 母集団と標本
- 標本平均の分布
- 推定　　　　　　　　　　　　（⇐　午前中ここを授業）
- 仮説検定

困り１

「aX+bの期待値，分散，標準偏差」は日常生活でいつ使うのか？

$$E(aX+b)=aE(X)+b$$
$$V(aX+b)=a^2V(X)$$
$$\sigma(aX+b)=|a|\sigma(X)$$

⇒　次のような演習問題を作ってみた

問

2つのさいころを同時に投げて、出た目の和に応じて「100円×和の値」の賞金を得られるゲームがある。

1. 出る目の和を確率変数Xとするとき、Xの確率分布を求めよ。
2. ゲームの参加費が600円のとき、利益の損得（プラス・マイナスの額）を考慮して、ゲームへの参加に価値があるか判定せよ。
3. 利益の損得（プラス・マイナスの額）の期待値が0になるためには参加費をいくらに設定すればよいか。
4. 主催者側が有利になるためには、「出た目にかける金額」と「参加費」をどのように設定すればいいか。
5. 「100円×和の値」の賞金で参加費が600円のとき、このゲームにおける利益の損得（プラス・マイナスの額）の標準偏差を求めよ。
6. 主催者に有利だが、射幸性が高いと参加者が感じるようにするためには、「出た目にかける金額」と「参加費」をどのように設定すればよいか。

困り２

「独立な2つの確率変数の積の期待値」は日常生活でいつ使うのか？

2つの確率変数X，Yが互いに独立であるとき，$E(XY) = E(X)E(Y)$

⇒　次のような演習問題を作ってみた

問

あなたは友達とあるトレーディングカードゲームを遊戯している。そこで相手は次の効果をもつ魔法カードを使用してきた。「サイコロとコインを投げて、サイコロの出た目×500ライフのダメージを、コインの表が出たら相手に、裏が出たら自分に与える。」このカードゲームではライフの上限が3000である。

1. このカードはチート（禁止）カードにするべきか論理的に評価せよ。

2. このカードがチートカードだとされるとき、どのようにビーフアップ（上方修正）またはナーフ（下方修正）を加えればよいか。アイデアを考えよ（そう考えた根拠も）。ただし、サイコロとコインを使うという設定は変えてはいけない。（対戦時の不公平感をなくすように！）

3. 最初の問題文の設定において、サイコロの目に応じたダメージ量をいくら調整してもダメージの期待値は一定になるがそれはなぜか説明せよ。

数I分科会での議論１

「独立な2つの確率変数の和の分散」の公式

$$\mathrm{V}(X+Y)=V(X)+V(Y)$$

確率変数が独立でない（従属）の場合は和の分散はどうなるの？

⇒　共分散が出てくる（確率変数の話でも）

$E(X)=m,\quad \mathrm{E}(Y)=n$ とする.

$$\begin{aligned} V(X+Y) &= E[\{(X+Y)-(m+n)\}^2] \\ &= E[\{(X-m)+(Y-n)\}^2] \\ &= E[(x-m)^2+2(X-m)(Y-n)+(Y-n)^2] \\ &= E[(x-m)^2]+\mathrm{E}[(Y-n)^2]+2E[(X-m)(Y-n)] \\ &= V(X)+V(Y)+2E[(X-m)(Y-n)] \end{aligned}$$

$E[(X-m)(Y-n)]$のことを共分散$Cov[X,Y]$と定義される量.

ちなみに，　$E[(X-m)(Y-n)]=E(XY)-E(X)E(Y)$となる.

困り３

二項分布の時点で，仮説検定の導入に繋がるような教材は作れないか？

⇒　次のような演習問題を作ってみた

問

あるファーストフード店Mでは、新商品のスイーツを販売していた。その商品のキャッチコピーとして、「パティシエの試食者15人に3択（おいしい・どちらでもない・おいしくない）でアンケートを取ったところ、10人がおいしいと答えました。」と書かれていた。

1. 15人全員がアンケートを適当（ランダム）に答えたと仮定する。アンケートにおいしいと答えた人数をXとするとき、このアンケート結果はどのような確率分布になるか。
2. (1)の確率分布において、P(10≦X)となる確率を求めよ。
3. (2)の結果を参考にして、「パティシエ15人が適当に答えた」とアンケートの結果を考えることはできるか。また、その理由を答えよ。
4. 今までの議論を踏まえて、あなたは新商品のスイーツがおいしいと思うか。（理由も含めて。）

困り４

確率密度関数がなんなのか上手く説明出来ない・・・

確率＝面積をイメージさせるのも難しい

原因：そもそも定積分をやっていない

藤田のアプローチ

離散型確率変数では，確率分布が書けた.

連続型だと確率分布が無限に続く（項目が無限にかける）.
しかし，確率変数の値に対応した確率は必ず1つあって，その確率たちをみると規則性があって関数で表現できそう.
その関数が確率密度関数.

関数の値が確率に直結するので，一定の区間の関数の値を図に書いていくと無数の線が重なる.

だから，確率は面積になる.

困り５・数Ⅰ分科会での議論２

正規分布の式に $Z=\frac{X-m}{\sigma_2}$ を代入してもZの確率密度関数にそのまま代入しても，$f(z)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$ にならない．　（$\frac{1}{\sigma}$が残る）

このことを高2の生徒へどう教えるか．

数Ⅰ分科会での話

- 置換積分
- $V(aX)=a^2V(X)$ によって，$\sigma=1$であることを示す．
- 分散の定積分定義 $V(x)=\int_\alpha^\beta (x-m)^2 f(x)dx$ を説明
- mを引くことは平均を0にすること，
  σで割ることは標準偏差を1にすること

藤田のアプローチ

「補整」という表現を用いて説明．Mで引くと分布曲線全体が平行移動する（平均も０になる）．σで割ると曲線が拡大・縮小する．

困り 6

標準正規分布で $P(-2 \leqq Z)$や$P(1.23 \leqq Z \leqq 2)$ などの確率を求めるとき，なぜ引き算や足し算が必要になるのか理由が分からない

原因

- 面積＝確率が定着していない
- 正規分布表は正の一部分の確率しか載っていないことが理解できていない
- 0を境にして対称であることがイメージできていない

⇒個別に丁寧に説明すれば分かってくれる

※ 何か定着させる良い指導法があれば教えてください．

困り７

標本平均の分布で，「抽出された標本が近似的に正規分布に従う」という表現の意味が伝わりにくい．

原因：抽出された標本の平均に，ばらつきがあるというイメージが
　　　教科書だけでは伝えづらい．

藤田のアプローチ

極端な固体を抽出すると標本平均が母平均と大幅にずれることを解説

※　特別な教材などは作っていません．
　　良い指導法があれば教えてください．

数I分科会での議論３

$E(X)=m, V(X)=\sigma^2$ のとき標本平均を$\bar{X}$とするとき

$\mathrm{E}(\bar{X})=m, V(\bar{X})=\left(\frac{\sigma}{\sqrt{n}}\right)^2$ となる.

- 母集団20万人から標本100人を抽出したとき，母標準偏差よりも標本標準偏差が狭くなることは感覚的にわかる.
- 標本が100人よりも標本200人の方が，標本標準偏差が狭くなることが分かりにくいらしい.
- 標本が100人と標本19万人では標本平均は母平均mに近くなるかどうか.

残り時間を用いて，**周囲**の方と「統計的な推測」の指導について
提示した話題などにご参考に議論してみましょう！

困り８

正規分布曲線 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-m)^2}{2\sigma^2})$ は高2の生徒へ
どう教えればいい？

そもそも，この確率密度関数をガウスはどう見つけた？
（実験的にという感じに書かれていたネット記事を見たことはあるが）

※　藤田はこういう関数があるよの紹介で終わりました．
　　良い紹介方法があれば教えてください．

困り９

仮説検定において,
「帰無仮説の（言葉的な意味での）立て方」
「帰無仮説が棄却されないときの言い回し」
を伝えるのが難しい.
（数Ｉ「仮説検定の考え方」の指導経験から）

原因：背理法のように「〜ではない」という結論ではないから

※　藤田は「〜とは言えない」が帰無仮説だと思ってミスりました.
　　仮説を立てるのに良い指導法があればお教えください.

<u>困り１０</u>

二項分布における期待値・分散について上手い導出がないか？
なるべく予備知識がいらず高２の生徒でも分かる範囲で.

<u>藤田のアプローチ</u>

離散型確率変数の期待値・分散の定義に従って導出.
シグマ計算の足す範囲にとても注意が必要.
（次のページから導出）

二項分布における期待値・分散の導出

X	0	1	……	r	……	n	計
P	${}_nC_0q^n$	${}_nC_1pq^{n-1}$	……	${}_nC_rp^rq^{n-r}$	……	${}_nC_np^n$	1

期待値の定義より、

$$E(X)=\cancel{0\cdot {}_nC_0 q^n}+1\cdot{}_nC_1 p q^{n-1}+\cdots+n\cdot{}_nC_n p^n$$

$$=1\cdot{}_nC_1 p q^{n-1}+2\cdot{}_nC_2\cdot p^2q^{n-2}+\cdots+n\cdot{}_nC_n p^n$$

Σにまとめる。

$$=\sum_{k=1}^{n} k\cdot{}_nC_k\cdot p^k\cdot q^{n-k}$$

${}_nC_k$を分数で表す。

$$=\sum_{k=1}^{n} k\cdot\frac{n!}{(n-k)!\,k!}\,p^k\,q^{n-k}$$

$\frac{k}{k!}=\frac{1}{(k-1)!}$ 約分

$$=\sum_{k=1}^{n}\frac{n!}{(n-k)!\,(k-1)!}\,p^k\,q^{n-k}$$

$n-k$
$=n-k+1-1$
$=n-1-(k-1)$

$n!$
$=n\cdot(n-1)\cdots2\cdot1$
$=n\cdot(n-1)!$

pも同様

$$=\sum_{k=1}^{n}\frac{n\cdot(n-1)!}{\{(n-1)-(k-1)\}!\,(k-1)!}\cdot p\cdot p^{k-1}\cdot q^{n-k}$$

$$\sum_{k=1}^{n}\frac{n\cdot(n-1)!}{\{(n-1)-(k-1)\}!\,(k-1)!}\cdot p\cdot p^{k-1}\cdot p^{n-k}$$

nとpはkに無関係

${}_nC_k=\frac{n!}{(n-k)!\,k!}$ のnを$n-1$に，kを$k-1$におきかえ

$$=n\cdot p\cdot\sum_{k=1}^{n}{}_{n-1}C_{k-1}\cdot p^{k-1}\cdot q^{n-k}$$

← $k\to(k+1)$ にかわる。

$\sum_{k=1}^{n}\Rightarrow\sum_{k=0}^{n-1}$ と数え方を変えたい。

$k-1$をkにおきかえ、

$$=np\sum_{k=0}^{n-1}{}_{n-1}C_k\cdot p^k\cdot q^{n-1-k}$$

$(p+q)^n=\sum_{k=0}^{n}{}_nC_k\,p^kq^{n-k}$ のnを$n-1$におきかえ。

$$=np\,(p+q)^{n-1}$$

$p=1-q$より
$p+q=1$

$$=np\cdot 1^{n-1}$$

$$=np$$

//

$V(X) = E(X^2) - \{E(X)\}^2$ より.

$E(X) = np$ であるので.

X	0	1	……	r	……	n	計
P	${}_nC_0q^n$	${}_nC_1pq^{n-1}$	……	${}_nC_rp^rq^{n-r}$	……	${}_nC_np^n$	1

$$E(X^2) = 0^2\cdot {}_nC_0\, q^n + 1^2\cdot {}_nC_1\, p\, q^{n-1} + \cdots + n^2\cdot {}_nC_n\, p^n$$

$$= 1^2\cdot {}_nC_1\, p\, q^{n-1} + 2^2\cdot {}_nC_2\cdot p^2 q^{n-2} + \cdots + n^2\cdot {}_nC_n\, p^n$$

$$= \sum_{k=1}^{n} k^2\cdot {}_nC_k\, p^k q^{n-k}$$

$$= \sum_{k=1}^{n} k^2\, {}_nC_k\, p^k q^{n-k} + \underbrace{\left(\sum_{k=1}^{n} k\, {}_nC_k\, p^k q^{n-k} - \sum_{k=1}^{n} k\, {}_nC_k\, p^k q^{n-k}\right)}_{}$$

あえてつける.

$$= \underbrace{\left\{\sum_{k=1}^{n} k^2\, {}_nC_k\, p^k q^{n-k} - \sum_{k=1}^{n} k\, {}_nC_k\, p^k q^{n-k}\right\}}_{} + \sum_{k=1}^{n} k\, {}_nC_k\, p^k q^{n-k}$$

順序変え

$$= \sum_{k=1}^{n} k(k-1)\, {}_nC_k\, p^k q^{n-k} + \underbrace{\sum_{k=1}^{n} k\, {}_nC_k\, p^k q^{n-k}}_{E(X)}$$

$$= \sum_{k=1}^{n} k(k-1)\frac{n!}{(n-k)!\,k!}\, p^k q^{n-k} + np$$

$k(k-1)$で約分

$$= \sum_{k=2}^{n} \frac{n!}{(n-k)!\,(k-2)!}\, p^k q^{n-k} + np$$

Σの始まりが変わる.
($(-1)!$がないから)

$$= \sum_{k=2}^{n} \frac{n\cdot(n-1)\cdot(n-2)!}{\{(n-2)-(k-2)\}!\,(k-2)!}\cdot p^2\cdot p^{k-2} q^{n-k} + np$$

式変形

n, pはΣに無関係

$$= n\cdot(n-1)\cdot p^2 \sum_{k=2}^{n} {}_{n-2}C_{k-2}\, p^{k-2} q^{n-k} + np$$

$\sum_{k=2}^{n} \Rightarrow \sum_{k=0}^{n-2}$ と数え方を変えたい.
$k-2$をkにおきかえ.

$$= n\cdot(n-1)\, p^2 \underbrace{\sum_{k=0}^{n-2} {}_{n-2}C_k\, p^k q^{n-2-k}}_{} + np$$

$(p+q)^n = \sum_{k=0}^{n} {}_nC_k\, p^k q^{n-k}$ の n を $n-2$ におきかえ.

$$= n\cdot(n-1)\cdot p^2\cdot \underbrace{(p+q)}_{1}{}^{n-2} + np$$

$$= n^2p^2 - np^2 + np$$

したがって、

$$V(X) = E(X^2) - \{E(X)\}^2$$

$$= n^2p^2 - np^2 + np - (np)^2$$

$$= -np^2 + np$$

$$= np(1-p)$$

$$= npq$$

//