

統計学指導の問題点

丸 山 求
Motomu Maruyama

1. まえがき

文系、幼児教育系の短期大学における一般教育の数学は、何をどのように扱ったらよいかさまざまな考え方や問題があるであろう。本学では統計をとり上げることにしたが、その扱い方についての問題を考えてみようというのが本稿のねらいである。

統計・確率についての専門書は、本格的なものから易しいものまで数多くあるが、程度、分量等実情に即したものはなかなか見当たらない。学生の知識・経験と能力に適したものに軽減しながら、本質的なものを見失なはないようにしたいと考えた。そのため、数式は少なして具体的なものを題材にして原理・原則を理解させ、繁雑な証明はなるべく避けて統計の精神、確率の考え方を学びたいものである。さりとて、数学である以上証明らしきものも扱いたい。そのため、とりあげた内容とその程度を述べ、次に問題点を挙げて読者諸賢の御叱正をお願いしたい次第である。

2. とりあげた内容と式

とりあげた内容と式をあげ、その視点と取り扱いの概要を述べることにする。

I. 序論

(1) 数学とロマン

集合論を創った カントール (Gorg Cantor 1845~1918) は、集合論の始まった頃の多くの人の反対に対して「数学の本質はその自由性にある。」といている。

制限なしに出来得る限り作る、また、いままでの範囲に入っていない新しい数学 (集合) をドンドン作っていくという、自由でロマンチックな精神が数学といてよいと思う。数学には無限に深いミステリーがある。……というようなことから、しばしば見られる数学は無味乾燥なつまらないものだという拒否反応を少しでも和らげと思うのである。

そのため、まづ、この項をとりあげた

(2) 資料の整理

情報社会といわれるように、われわれの生活は雑多な無秩序な情報に取り巻かれている。何らかの意味で、それらを取捨選択して整理しなければならない。ここで考える資料も同様である。その場合のいくつかの方法をあげて、それらの効用を述べ、具体的な例で考え

させることにした。

(3) 偶然と必然

偶然性ということが世の中にあるとして、それが人間とどういふかかわり合いがあるだろうかということを考えてみると、いろいろなレベルが考えられるが、一番身近なレベルは個人のレベルでのかかわり合い、言ってみれば「個人の主体性と偶然」である。自分自身は偶然の存在である自分自身にいわば縛りつけられているわけであり、それを離れてどこかもっと必然的な運命におかれたいと思っても無理である。そこに1つの矛盾がある。つまり、自分は偶然的な存在であるが、その偶然的な存在である自分が実は自分にとってはそれ以外のものでないという意味では必然的であるというわけでそこに矛盾が存在する。この矛盾を一体どういふふうにして解決するかというのが実は非常に重要な問題である。

(竹内啓)

偶然と必然の問題は考えるほどに興味深く、尽きない興行を孕んでいる。その思想を基底において、ここで扱う程度の数理的処理(統計)について学生たちに理解させたいことは次のような事柄である。

たとえば、大学1年生の女子の身長在全国平均を知りたいとき、偶然に選ばれた40人の平均を求めて全国平均との関係を理論的に求めたいというのが統計である。すなわち40人の平均を m として全国平均を $m \pm \alpha$ というある幅の中に押し込める理論が統計である。どの40人を選んだかという偶然性からのがれて必然に近づけようとする理論である。偶然性を偶然に影響されないようにする理論が統計の数理である。

(4) 偶然の科学

統計は偶然についての学問、偶然の科学と言ってよいであろう。「つゆの時期にしては珍しく日本晴れだ。」「彼は変わっている。」などの日常語について考えてみよう。前者について「つゆの時期は湿度が多く、曇り空で雨の日が多いものだ。」という通念がある。後者の場合にも人間集団のなかで、我々は常識的に言動についての尺度をもっている。これらの通念、尺度は、ある構造を経験的に構成しているといってもよいであろう。この構造ないしモデルは各人別々では議論にならない。そこで共通な客観的なモデルが必要になる。このため、多数の資料を集め1つのモデルを作っておかねばならない。

このモデルとして使われるのが確率分布である。これを基準にして統計的推測が行われるが、普通2段階で行われる。まず、第1に偶然的な変動を統計的データについての確率モデルを想定する。つまり、その統計的データの変動がなんらかの確率論的な構造に従うと想定するわけである。第2に要求水準に応じた確率の値を決め、その範囲に入る変動の幅をもって推測するのである。

現実には確率の変動に従うことを想定できるようにするためには、データをランダムに抽ししなければならない。このことについて次に考えよう。

(5) 偶然を作る

統計では、調査対象の集団全体（母集団）を調査するかわりに、一部の資料（標本）を抜き取って調べ、全体の実態を推論する場合がある。この標本を選び出すときに用いる方法を無作為抽出（Random Sampling）という、この抽出は無作為に、偶然に選び出されなければならない。いはば、偶然が必要となり、偶然を作らなければならない。

偶然を必要とすることは、日常生活の中でもしばしばある。宝くじの当選番号を決めたり、野球のトーナメント試合の組み合わせを決めるときなど、偶然に当たったと考えられる方法を使うことになる。一般には「くじびき」という方法が用いられている。統計では乱数表というものを用いる。乱数表には日本規格協会から出ている統計数値表というものがあり、20万個の数字がのっている。これとは別に「JIS 乱数表」というものが出されていて1000万個の数字がのせられている。

偶然（でたらめ）を作ることは、でたらめにはできない。いま正方形のなかに無作為（偶然、でたらめ）に点を打つことを考える。図2はある学生が「でたらめ」に打とうとして打ったものであり、図1は、たて、よこを座標にして乱数表の数の組から打ったものである。図2は、正方形のなかに「散らばって打つ」という作為があるように思われるのでランダムではない。図1は乱数表を信じてランダムに打たれていることになる。

II、記述的統計

(1) 度数分布

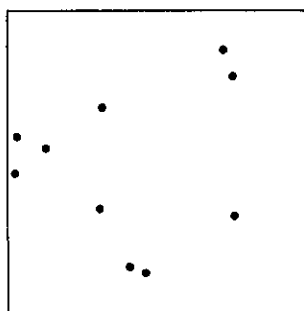
(ア) 度数分布と度数グラフ

(イ) 累積度数

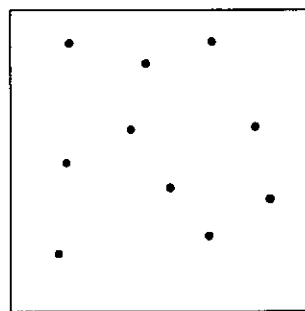
(2) 代表値と散布度

(ア) 代表値

a. 平均値



(図 1)



(図 2)

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \cdots + x_i + \cdots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad \cdots \cdots \cdots \textcircled{1}$$

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{N} \sum_{i=1}^n x_i f_i \quad \cdots \cdots \cdots \textcircled{2}$$

$$(\text{ただし、} N = \sum_{i=1}^n f_i)$$

b. メジアン（中央値、中位数）

c. モード（最頻値、並み数）

(イ) 散布度

a. 範囲

b. 四分偏差

c. 平均偏差

d. 分散と標準偏差

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 f_i \quad \text{ただし, } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i f_i \quad \dots\dots\dots ③$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 f_i} \quad \dots\dots\dots ④$$

$$\overline{ax+b} = a\bar{x} + b, \quad s^2(ax+b) = a^2 s^2(x) \quad \dots\dots\dots ⑤$$

e. チェビシエフの定理

(3) 相関係数

(ア) 相関関数

(イ) 相関係数

$$C(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad r = \frac{C(x, y)}{s_x s_y} \quad \dots\dots\dots ⑥$$

III. 確率

(1) 確率の定義と基本的な性質

(ア) 数学的確率

(イ) 加法定理

(ウ) 統計的確率

(2) 独立試行の確率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \dots\dots\dots ⑦$$

(3) 期待値

IV. 標本分布

(1) 正規分布

(ア) 正規分布の形

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (\sigma > 0) \quad \dots\dots\dots ⑧$$

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \quad \left(t = \frac{x-m}{\sigma} \text{ において, } dx = \sigma dt \right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{t^2}{2}} dt$$

$$= \frac{\sqrt{2}}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{2}} = 1 \quad \dots\dots\dots ⑨$$

この関数について、平均 $E(x)$ と標準偏差 $\sigma^2(x)$ とを求めると、

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + m) e^{-\frac{t^2}{2}} dt \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} dt + \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 0 + \frac{m}{\sqrt{2\pi}} \sqrt{2\pi} = m \quad \cdots \cdots \cdots \textcircled{10} \end{aligned}$$

$$\begin{aligned} \sigma^2(x) &= \int_{-\infty}^{\infty} (x-m)^2 f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt = \frac{\sigma}{\sqrt{2\pi}} \left\{ \left[t \left(e^{-\frac{1}{2}t^2} \right) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt \right\} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left\{ 0 + \sqrt{2\pi} \right\} = \sigma^2 \quad \cdots \cdots \cdots \textcircled{11} \end{aligned}$$

(イ) 標準正規分布

$$\begin{aligned} N(m, \sigma^2) \text{ に従う } f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \text{ を } T = \frac{x-m}{\sigma} \text{ によって } T \text{ に変換すれば} \\ P(T \leq t) &= P\left(X \leq \frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \quad \left(t = \frac{x-m}{\sigma} \text{ で変換して}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}t^2} dx \quad \therefore g(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \quad \cdots \cdots \cdots \textcircled{12} \end{aligned}$$

これは $N(0, 1)$ に従う。これを標準正規分布という。

ここで正規分布表の引き方、扱い方を練習させ、偏差値なども扱う。

(2) 標本分布

母集団、標本、無作為抽出、乱数表などを扱う。

V. 統計的推定

(1) 推定の考え方

推定の基本的な考え方を述べて次の2つの定理を扱う。

(定理) 母集団 Ω が正規分布 $N(m, \sigma^2)$ に従うとき、復元抽出法による大きさ n の標本平均 \bar{X} は正規分布 $N(m, \frac{\sigma^2}{n})$ に従う。

(定理) 母集団 Ω の母平均が m 、母分散が σ^2 であるとする。復元抽出法による大きさ n の標本平均 \bar{X} の確率分布は、 n を十分大きくすれば、正規分布 $N(m, \frac{\sigma^2}{n})$ によって近似される。

(2) 区間推定

$$\text{母平均} \quad P\left(\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95 \quad \cdots \cdots \cdots (13)$$

$$\text{母平均} \quad P\left(\bar{X} - 2.58 \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{V} + 2.58 \times \frac{\sigma}{\sqrt{n}}\right) = 0.99 \quad \cdots \cdots \cdots (14)$$

$$\text{母比率} \quad P\left(\frac{x}{n} - 1.96 \times \frac{\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)}}{n} \leq P \leq \frac{x}{n} + 1.96 \times \frac{\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)}}{n}\right) \quad \cdots \cdots \cdots (15)$$

VI. 統計的検定

(1) 検定の考え方

$$\left| \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq 1.96 \quad \cdots \cdots \cdots (16)$$

(2) 両側検定と片側検定

$$\left| \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq 1.65 \quad \cdots \cdots \cdots (17)$$

VII. 実際研究の例

ここで、学生が夏休み等を使って実際に出来る程度の実験研究の例を挙げて、学生の参考にした。必ずしも推定や検定の手法を用いなくても、資料を集める着想がよい場合、さまざまな興味ある結果を得たり、予測したりすることができる。

テーマの例を挙げると……長野県人の県民性、新聞投書欄に現れた世相、新聞の気象欄の気温の末位の数字、姓の分布、目測の精度、太宰治の用語の頻度、松本市の保育園の分布、幼児の将来望む職業、音楽に対する十代の意識……

60～70の例を挙げて参考にし、学生自らユニークなテーマを発想するようにした。

3. 指導上の問題点

教育過程の概要と指導内容を述べてきたが、勿論このほかにそれぞれの内容についての「問」や「練習問題」で演習を行って理解を深めるようにした。そして、2単位……90分、14～15時間で一通り終わるように考えた。

その中で主な数式の取り扱い、定理（または原理）の証明については、幾つかの問題点、困難点、あるいは迷いを感じざるを得なかった。これらの諸点について順次述べてみたい。このとき学生基礎知識を考えないわけにはいかない。高等学校おいてどこまで数学を履修してきているかであるが、微積分を一通り学習してきていることは期待したけれども、数学1だけしか履修してきていないものが30～40%あることも考えなくてはならない。

式①、②についてはよく理解するが、 Σ については不明確なものがあることに注意しなければならない。一連の操作をまとめて表す記号であることに気づかせる。式③、④についても同様である。式⑤については $\overline{a \cdot x + b}$ と $a \cdot \overline{x} + b$ のちがいについても案外説明が必要のよう

である。式⑥については Σ の操作について具体的に数をいれて話している。式⑦については $P(A|B)$ の意味になれさせる。

式⑧は、 m 、 σ によってグラフの形がきまることに注意して天下り。式⑨の証明はぜひ扱いたいところで、この証明は初めて式⑧が確率密度関数であることが保証され意味が明らかになる。しかし、この証明には一つ難所がある。次の積分である。

$$\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$

これは、ふつう次のようにして計算するが、これは学生には無理である。

$0 \leq x \leq a$, $0 \leq y \leq a$ の正方形の領域を I , 原点を中心とする半径 a , $\sqrt{2}a$ の四分円をそれぞれ D_1 ,

D_2 とする。このとき

このとき

$$e^{-(x^2+y^2)} > 0$$

よって

$$\iint_{D_1} e^{-(x^2+y^2)} dx dy < \iint_I e^{-(x^2+y^2)} dx dy < \iint_{D_2} e^{-(x^2+y^2)} dx dy$$

他方

$$\iint_I e^{-(x^2+y^2)} dx dy = \int_0^a \left\{ \int_0^a e^{-x^2} dx \right\} e^{-y^2} dy = \left(\int_0^a e^{-x^2} dx \right)^2$$

$$\iint_{D_1} e^{-(x^2+y^2)} dx dy = \int_0^{\frac{\pi}{2}} \left\{ \int_0^a e^{-r^2} r dr \right\} d\theta = \frac{\pi}{4} (1 - e^{-a^2})$$

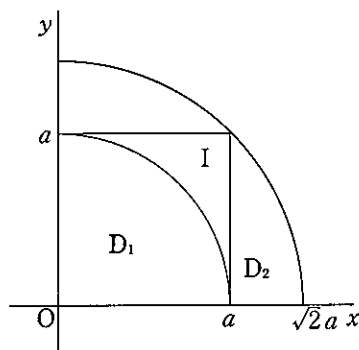
$$\iint_{D_2} e^{-(x^2+y^2)} dx dy = \int_0^{\frac{\pi}{2}} \left\{ \int_0^{\sqrt{2}a} e^{-r^2} r dr \right\} d\theta = \frac{\pi}{4} (1 - e^{-2a^2})$$

$$\therefore \frac{\pi}{4} (1 - e^{-a^2}) < \left(\int_0^a e^{-x^2} dx \right)^2 < \frac{\pi}{4} (1 - e^{-2a^2})$$

ここで、 $a \rightarrow \infty$ とすると、 $e^{-a^2} \rightarrow 0$ となるから

$$\left(\int_0^{\infty} e^{-x^2} dx \right)^2 = \frac{\pi}{4}$$

$$\therefore \int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$



式⑩、⑪についても同様である。部分積分はなんとかなるから積分 $\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$ を

天下りにして何とか切り抜けている。いっそんな証明はすべて省略して結果だけを認めさせ先へいくのも方法であろう。このとき何か釈然としないものが残る…いかにすべきかというのが本稿中心の問題である。

次には母集団分布（正規）と標本分布との関係の定理であるが、実験でイメージをつくり納得させる手があるが大変な時間がかかる。結果を話して意味を理解させ使っていくより方法がない。

中心極限定理も同様にお話しておく。

⑬、⑭、⑮の不等式は理解できる。⑯、⑰も同様である。

以上のようなが、これだけでは味気ないのでVI. 実験研究の例を示し、夏休み等を利用して個人またはグループで身近なテーマでレポート出させることにしている。興味ある、奇抜なものがしばしば出てくるし、熱心に根気よく取り組むのも予想以上であった。このレポートについては以前報告した。（松本短期大紀要第2号）

問題点を個々に挙げてきたが、概観して次にまとめて考察する。

4. あとがき

一般教育における数学はいわばハイライトで扱いたい。かなり程度の高い内容を易しくトータルとして理解できればそれに越したことはない。登山にたとえれば、アプローチの長い平凡なところバスで走り、急な難所で体力に自身がなかったら、できればリフトかケーブルを利用する。途中の景色のよい所では十分時間をとって鑑賞する。頂上で360°眺めて空間的な関係を明らかにする。下りはゆっくり歩いて体力つける。そうすれば登山家しか登れなかった山も、老若男女が登れる庶民のものとなる。スイスアルプスのユングフラウの3400mの所まで登山電車で登り、アルプスの雄姿を楽しめるのである。山登り専門家は、岩登りを試み、ヒマラヤに挑戦するのに、はかりしれない興味があるであろう。

数学に話を戻せば、複素関数のコーシー・リーマンの定理はいまや庶民のものとなった。小標本理論の糸口をつくったのは、1908年スチューデントの筆名で発表されたゴゼットの論文である。彼の考えは、後にフィッシャーによって理論化され、その後多くの人たちによって研究され、開発され今日に至り庶民のものとして理解され利用されている。短大の女子学生には、どこでバスを使い、どこでケーブルを利用したら数学を愉しむことができるか、この稿のねらいは、このへんにあったのである。十分な研究にはならなかったが、今後日々問題として追求していきたい。

参考文献

- 竹内外史著 数学の世界観 紀伊國屋書店
竹内 啓編 無限と有限 東京大学出版部

- 竹内 啓著 偶然と必然 東京大学出版部
" 著 偶然の科学 "
- 丸山 求著 応用数学 1 大日本図書株式会社
" 微分積分学 "
" ゆらぎ 銀河書房
" 稿 松本短期大紀要第 2 号 「女子短大における確率・統計」
" 著 女子短大生のための確率・統計 電算印刷