確率論と統計学を繋ぐ処
～中心極限定理の居場所を探る～

かわむら　としひと
河村　寿人

§1. はじめに

令和４年度からの新学習指導要領では，統計の学習内容が大幅に増え，注目度の高いところである。

ところで，統計学はその基礎知識として確率論を必要としている。確率論に進むか統計学に進むかの分岐点は，確率変数の定義以降である。その意味では，確率が「数学Ａ」の単元で，統計が「数学Ｂ」の単元であるのは問題が無いと思える。しかしながら，教育課程の編成に依っては,「数学Ａ」の確率の学習を一旦終了した後，間を置いて「数学Ｂ」の統計の学習に取り組む事にもなる。

また，次の様な教育課程も考えられる。「数学Ａ」は「数学Ⅰ」と並行またはその後に履修することが望ましいとされているが,「数学Ａ」と「数学Ｂ」の履修の順番は規定されていない。すると「数学Ⅰ」を履修した後,「数学Ａ」を履修せずに「数学Ｂ」を履修する場合も考えられる。すなわち「数学Ⅰ」で統計の初歩を学んだ後，確率を全く学ばずに「数学Ｂ」の統計へ進むという状況である。

以上から，確率と統計の学習内容を吟味する事は極めて重要であると思われる。

本論では，確率論と統計学との関連を解説し，統計学を流れ良く学ぶための留意事項と望ましい学習姿勢を述べる。

§2. 確率論の基礎知識

⑴　確率論の基礎は集合論と測度論である。ある試行における標本空間をΩとし，σ- 集合体を$\mathfrak{F}$とする。$\mathfrak{F}$上の測度Pで $P(\Omega)=1$ とした時，Pは確率測度と呼ばれ，$(\Omega,\ \mathfrak{F},\ P)$ は確率空間と呼ばれる。

⑵　確率空間を $(\Omega,\ \mathfrak{F},\ P)$ とする。任意の $A,\ B\in\mathfrak{F}$ に対して，A におけるBの条件付確率を次のように定義する。

$$P_A(B)=\begin{cases}\dfrac{P(A\cap B)}{P(A)} & P(A)\neq 0\\ P(B) & P(A)=0\end{cases}$$

条件付確率を用いて，事象の独立を定義する。条件付確率の定義から

$$P(A\cap B)=P(A)P_A(B)$$

となるが，この式で $P_A(B)=P(B)$ すなわち

$$P(A\cap B)=P(A)P(B) \qquad \cdots\cdots①$$

が成り立つとき，事象 A，B は独立であるという。注意しなくてはならないのは，１回の試行だという事である。

ここで統計学との関連が感じられる。①を２回の試行で考えると，復元抽出または非復元抽出のどちらかで考えることになる。もし，２回の試行と考えると，標本空間が変化するので，$P(B)$ とは書けない。確率分布としては，復元抽出は２項分布に該当し，非復元抽出は超幾何分布に該当する。

復元抽出の解釈は，反復試行の定義となる。ここで試行の独立が定義できる。この場合は試行回数だけの標本空間が現れるので，直積空間上での確率測度を考える。「数学Ａ」の確率の履修内容はここでほぼ終了する。

⑶　「数学Ⅰ」ではヒストグラム等初歩的な統計処理から始まり，平均値，分散，標準偏差等データの散らばりを調べる方法や相関係数で２つの変量の間の関係を調べるまでが統計の履修内容である。記述統計学の初歩を学んでいると言えるだろう。

「数学Ａ」でも平均値が扱われているが，確率変数を定義することなく説明されている。確率変数と平均値は姿を変えさせられて,「数学Ⅰ」「数学Ａ」「数学Ｂ」の３つの世界を飛び回っていると言えよう。そこで，確率変数の定義に基づく平均値の定義を行う。

§3. 確率変数以降の確率論

(1) 確率空間 $(\Omega,\ \mathfrak{F},\ P)$ において，標本空間 Ω から実数 R への関数

$$X:\Omega\to R$$
$$\omega\to x$$

が，任意の実数 $x\in R$ に対して，

$$\{\omega|X(\omega)\leqq x\}\in\mathfrak{F}$$

を満たすとき，$X(\omega)$ を確率変数と呼ぶ。

(2) 確率空間 $(\Omega,\ \mathfrak{F},\ P)$ において，確率変数 $X(\omega)$ に対する標本空間 Ω 上での積分を平均値と呼ぶ。

$$E(X)=\int_\Omega X(\omega)\,dx$$

(3) 分散とは，偏差の2乗の平均値である。平均値の回りの2次のモーメントであるとの説明も多い。

$$V(X)=E(X(\omega)-E(X(\omega))^2$$

(4) 確率空間を $(\Omega,\ \mathfrak{F},\ P)$ とする。

この確率空間上に定義された確率変数 $X(\omega)$ $(\omega\in\Omega)$ によって次の様に定義されるものをそれぞれ分布，分布関数と呼ぶ。

ユークリッド空間 R のボレルクラスを $\mathfrak{B}$ とする。

任意の $E\in\mathfrak{B}$ に対し，

$$\Phi(E)=P(X(\omega)\in E)$$

また任意の $x\in R$ に対し，

$$F(x)=P(X(\omega)\leqq x)$$

この事から分布関数は分布によって次の様に定義される。

$$F(x)=\Phi((-\infty,\ x])\qquad x\in R$$

分布が高々可算個の点で値をとるとき，その Φ を離散分布と呼ぶ。

また，積分可能な関数 $f(x)$ によって，次のように書ける時，Φ を連続分布と呼ぶ。

任意の $E\in\mathfrak{B}$ に対し，

$$\Phi(E)=\int_E f(x)\,dx$$

$f(x)$ は確率密度関数と呼ばれ，分布関数は $f(x)$ を用いて次の様に表される。

$$F(x)=\int_{-\infty}^{x} f(x)\,dx$$

統計学では，離散分布では確率関数，連続分布では確率密度関数 $f(x)$ が重要な役割を果たすが，元を辿れば確率論における分布から導かれたものである。

また前述の平均値は分布によって次のように定義される。

$$E(X(\omega))=\int_R X\Phi(dx)$$

§4. 確率分布以降の確率論

(1) 確率空間を $(\Omega,\ \mathfrak{F},\ P)$ とする。確率変数 $X(\omega)$，$Y(\omega)$ が，任意の A，$B\in\mathfrak{F}$ に対して，$P(X\in A,\ Y\in B)=P(X\in A)P(Y\in B)$ が成り立つ時，確率変数 $X(\omega)$ と $Y(\omega)$ は独立であるという。

独立な確率変数について，平均値や分散で成り立つ重要な定理があるが，ここでは収束に関する定理を述べよう。

収束には，概収束，確率収束，法則収束があるが，極限操作が必要になる。これは「数学Ⅲ」で詳しく学ぶが，「数学Ⅱ」で極限を学ぶので，その片鱗を知ることはできる。同時に関数列の概念が必要となるが，「数学B」の数列でやはりその片鱗を知ることができる。惜しむらくは「数学Ⅱ」の指数関数で，$2^{\sqrt{2}}$ の値を無限点列で定義しているのにも関わらず，その後活用されていないことである。この場で関数列を一般化しておくのは後々役に立つであろう。

(2) 収束には，概収束，確率収束，法則収束がある。概収束と確率収束は確率変数に関する収束で，法則収束は分布に関する収束である。

ここでは，大数の法則と中心極限定理について述べる。大数の法則には大数の弱法則と大数の強法則とがある。

大数の弱法則は確率収束に関するものである。大数の強法則は概収束に関するものである。

確率変数 X_1，X_2，……に対し $S_n=\sum_{k=1}^{n}X_k$ とおく。

適当な条件の下で，数列 $\{a_n\}$，$\{b_n\}$ が存在して，

$\left\{\dfrac{S_n-a_n}{b_n}\right\}$ が0へ確率収束するとき，S_n は大数の弱法則に従うという。

ここで，$\{b_n\}$ は発散する正数列である。

弱法則の例を挙げる。

独立な確率変数 X_1，X_2，……に対し，

$S_n=\sum_{k=1}^{n}X_k$ とおき，任意の $k\in N$ に対して，

$E(X_k)=\mu_k$，$V(X_k)=\sigma_k{}^2<\infty$ とする。

$\{a_n\}$ を正数列とし，

$$\frac{\sum_{k=1}^{n}{\sigma_k}^2}{{a_n}^2}\to 0 \qquad n\to\infty$$

ならば，任意の $\varepsilon>0$ に対して，

$$P\left(\left|\frac{S_n-\sum_{k=1}^{n}\mu_k}{a_n}\right|<\varepsilon\right)=1 \qquad n\to\infty$$

が成り立つ。

ここで同一分布に従うという条件を加えることにより，次のようなよく見かける形に収まる。

独立で同一分布に従う確率変数 X_1，X_2，……に対し，$S_n=\sum_{k=1}^{n}X_k$ とおき，任意の $k\in N$ に対して，$E(X_k)=\mu$ とする。
任意の $\varepsilon>0$ に対して，

$$\lim_{n\to\infty}P\left(\left|\frac{S_n}{n}-\mu\right|<\varepsilon\right)=1$$

が成り立つ。この結果は，$V(X_k)=\sigma^2$ が存在しなくても成立する。

ここで，μ を母平均，σ^2 を母分散，$\overline{X_n}=\frac{S_n}{n}$ を標本平均と呼ぶ。

この証明は，「数学Ⅲ」の極限を学んでいれば難しくなく，良い練習問題であろう。

この例を大数の強法則で説明すると次のようになる。

$$P\left(\lim_{n\to\infty}\frac{S_n}{n}=\mu\right)=1 \qquad \text{a.e.}$$

これらの例は最も初歩的なものであり，条件を変えることによって異なる定理が得られる。

大数の強法則は，アンドレイ．N．コルモゴロフ(1903年～1987年 露)によって完成させられた。

大数の弱法則は，ヤコブ・ベルヌーイ(1654年～1705年 スイス)により1713年に発表されたが，統計的確率の理論的根拠となっている。例えば，硬貨を投げて表が出たら1，裏が出たら0を対応させる確率変数列 $\{X_n\}n\in N$ を考える。確率変数の和 S_n のとり得る値は表の出た回数であるから，標本平均は表の出た統計的確率で相対度数と呼ばれる。試行回数を増やせば，相対度数の値と数学的確率 $p=\frac{1}{2}$ の差が0に限りなく近づいていく確率は1である事を示している。

ただ，大数の弱法則とはあくまでも標本平均が母平均に確率収束するという事実を示しているものである事を忘れてはならない。ヤコブ・ベルヌーイの発表は，相対度数を以て確率の定義とし，それを統計的確率と呼ぶという事である。更にこの例では，数学的確率と母平均が共に $\frac{1}{2}$ で同じ値であるから余計に混同し易い。他の試行では同じ値になるとは限らない。確率論の歴史を学ぶ面では，統計的確率も忘れてはならないものであるが，現代の確率論を学ぶ際には，煩わしいものにも成りかねない。

⑶　中心極限定理

確率変数 X_1，X_2，……に対し，$S_n=\sum_{k=1}^{n}X_k$ とおく。適当な条件の下で，$\frac{S_n-E(S_n)}{\sqrt{V(S_n)}}$ の分布が標準正規分布 $N(0,\ 1)$ に収束する。これを中心極限定理という。
中心極限定理は法則収束に関するものである。
中心極限定理の最も初歩的な形は次のようなものである。独立で同一分布に従う確率変数 X_1，X_2，……に対し，$\overline{X_n}=\frac{1}{n}\sum_{k=1}^{n}X_k$ とおく。$k=1$，……，n に対して，$E(X_k)=\mu$，$V(X_k)=\sigma^2$ とする。

このとき，$\dfrac{\overline{X_n}-\mu}{\frac{\sigma}{\sqrt{n}}}$ の分布は標準正規分布 $N(0,\ 1)$ に収束する。

中心極限定理は，特性関数を用いて証明できる。「数学Ⅲ」で極限と複素数平面を学ぶので，課題学習として，その証明を行うのも一つであろう。

ところで，ここに確率論と統計学の曲がり角とも言える場面がある。

確率論では，確率変数 X_1，X_2，……の条件を変えることにより，中心極限定理のさまざまな結果が得られる。一方で統計学では，母集団の様子を調べる推測統計学の領域に入っていく。

ここで，統計学との解釈の違いを明確にしなくてはならない。

独立確率変数列 $\{X_n\}$ の各 $n\in N$ は試行の順番を表す。これは確率論での解釈であり，統計学における復元抽出の解釈である。しかしながら，推測統計学においては，確率変数 X_1，X_2，……はもはや実現値であり，X_1，……，X_n であれば，n は標本の大

きさを表している。

無限母集団からの大きさ n の標本は，独立確率変数列 $\{X_n\}$ の実現値と同じものと解釈できるが，有限母集団での非復元抽出では独立試行とはならないので，同じものと解釈できない。母集団としては無限母集団での抽出または有限母集団での復元抽出を想定しているとの説明が必要不可欠である。

推測統計学の基本的な問題は，標本から母集団の様子を推定することである。標本の大きさが十分に大きいときに，標準正規分布を利用して二項分布の確率の近似値を求めることができる。

アブラーム・ド・モアブル(1667 年～1754 年 仏)は，1733 年に二項分布の確率を正規分布で近似できるド・モアブル=ラプラスの極限定理と呼ばれる定理を発表した。これは中心極限定理の特別な場合である。以下に述べる。

X は二項分布 $B(n,\ p)$ に従う確率変数とする。標本の大きさが十分に大きく，平均値と分散も十分に大きいとき，$\dfrac{X-np}{\sqrt{np(1-p)}}$ の確率の近似値を標準正規分布 $N(0,\ 1)$ を利用して求めることができる。ピエール=シモン・ラプラス(1749 年～1827 年 仏)やカール・フリードリッヒ・ガウス(1777 年～1855 年 独)もこの定理を発見している。

ところで，中心極限定理は確率変数の和 S_n を用いて書かれている場合と，標本平均 $\overline{X_n}$ を用いて書かれている場合がある事は注意した方が良い。

先述の最も初歩的な形と述べたもので，$\dfrac{\overline{X_n}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$

ではなく，$\dfrac{S_n-n\mu}{\sqrt{n}\,\sigma}$ で書かれている事も多い。例えば，ド・モアブル=ラプラスの極限定理で X をベルヌーイ試行列 $\{X_n\}n\in N$ の和 $\sum_{k=1}^{n} X_k$ と考えて，X の確率分布は正規分布に収束すると述べることができる。これは，確率変数の和で述べている一例とも言える。

§5. まとめと考察

§1でも述べたが,「数学Ａ」は「数学Ⅰ」と並行またはその後に履修することが望ましいとされていても,「数学Ａ」と「数学Ｂ」の履修の順番は規定されていない。仮に「数学Ａ」を履修せずに「数学Ｂ」を履修する場合を考えてみる。すなわち，確率を全く知らない人が統計を勉強しようとした場合を想定する。

その人にとっては，中心極限定理は，母集団の推測を行う便利な道具に過ぎない。測度論や集合論や解析学に無縁な世界に入っていく。

応用統計学の領域は幅広い。推測統計学だけではなく，多変量解析等の手法は経済学や社会工学において重要な役割を果たしている。また応用確率論においてはオペレーションズリサーチ，確率過程論の応用分野としては時系列解析が挙げられよう。

また，新学習指導要領で統計の学習内容が大幅に増えた背景には，ビッグデータ問題の扱いには統計学が必要不可欠であるとの認識がある。これを否定する人はいないだろう。統計学に寄せられる期待は予想以上に大きいと言える。

ところで，統計学の文献を解読する為の確率論の知識がどこまで必要かはその場面に依って異なるであろう。先述した事だが，例えば確率変数の解釈にしても確率論と統計学で明確な違いがある。ところがその一方で，確率論におけるその理解が確実でなくとも応用面において必ずしも障壁になるとは限らない。そのような理由で，確率論の基礎を学ばずして，統計学の応用分野へと向かう傾向があるのではないだろうか。

しかしながら，その知識や理解が必要であるか否かという観点だけで確率論曳いては数学を学ぶ意欲が減退してしまうのは寂しい事である。
それぞれの専門領域を修める事も尊いが，確率論と統計学の関連とその差異を学ぶ事も尊い。
統計学を学ぶ姿勢においても，また使う立場においても，確率論の基礎を学ぶ事を避けて欲しくない。
そして，そのような姿勢を育む事こそが，数学教育に携わる者に課せられた使命ではないだろうか。

《参考文献》

〔１〕「確率論」鶴見茂　至文堂
〔２〕「確率」本間鶴千代　筑摩書房
〔３〕「工科系のための統計概論」
Ｉ・ガットマン／Ｓ・Ｓ・ウィルクス　共著
石井恵一／堀素夫　共訳　培風館
〔４〕「数理統計学」竹内啓　東洋経済新報社

〔５〕「統計的品質管理の基礎」
高木金地　産業図書株式会社
〔６〕「現代経営科学全集　５　確率論とその応用　Ⅰ　上」
W・フェラー　河田龍夫　監訳
ト部舜一　矢部真　池守昌幸　大平担　阿部俊一　訳　紀伊國屋書店
〔７〕「R による極値統計学」
西郷達彦　有本彰雄　オーム社
〔８〕「数学セミナー」1995　6　日本評論社
〔９〕「数学セミナー」1998　11　日本評論社
〔10〕「高等学校学習指導要領解説　数学編　理数編」平成 30 年 7 月　文部科学省
〔11〕「新編　数学Ⅰ」数研出版
〔12〕「新編　数学Ａ」数研出版
〔13〕「新編　数学Ⅱ」数研出版

（静岡県立磐田北高等学校）