≫ 論　説 ≪

ネイマン・ピアソンの補題についての感想

高　宮　浩　司*

概 要

数理統計学における統計的仮説検定を学んでいて気がついたことを二つ書く．第一に，検定論における基本結果であるネイマン・ピアソンの補題の直観的な理解の仕方を記す．第二に，この補題に関係する仮説検定の考え方についてざっくばらんな感想を述べる．

1　緒言

今年度 (2023 年度) 大学院のゼミで一年間学生たちとともに数理統計学を学んできたが，本稿ではそこで気がついたことのうち，とくに統計的仮説検定について二つ，ごく気楽に記しておく．

第一に，ネイマン・ピアソンの補題の理解の仕方についてである．ゼミではこの補題をいかにして直観的に理解するかということに学生たちとともに取り組んだので，その結果をここに記す．この補題は基本的な結果なのでさまざまなところで解説されているが，その直観的な意味についての説明はなぜか見当たらなかった．無論見尽くしたわけでないし，われわれが得た直観的な理解の仕方というのは当然以前から知られているものであろう．わかってしまえば当り前の話なので，わざわざ述べるほどもないと思われているのかもしれない．しかし少なくとも著者と学生の間ではこれで「分かった気」になれたし，そこまで到達するにはある程度時間がかかったので，こういう説明がもっと普及していてもよいと思った．したがって読者の勉強の助けになることがあれば望外の喜びと思い，それに期待して記しておく．

第二に述べることは，ネイマン・ピアソンの補題についての議論から敷衍した統計的仮説検定の考え方についてのごくざっくばらんな感想である．仮説検定の論理の微妙さや落し穴については，長年多くの人がいろいろなところでいろいろなことを述べているから，新しいところはないであろう．ただ思ったことを述べたまでの文字通り感想である．読者の勉強の助けになるとは思わないが，書いて悪いこともなかろうし，だれかのなにかのヒントになる可能性も皆無ではなかろうと思い，記しておく．

*新潟大学経済科学部, 950-2181 新潟市西区五十嵐 2 の町 8050 番地, takamiya@econ.niigata-u.ac.jp.

本文中，定式化，用語などの基本事項については竹村 (2020) を参照したが，必ずしも全て一致させているわけではない．上述の大学院のゼミで読んでいたのもこの書であり，大いに学ぶところがあった．元々は 1991 年に創文社から「現代経済学選書」の一巻として出たもので，そのころから数理統計学における本邦での名著として知られたものと思う．同社の解散にともなって学術図書出版社から改訂版として出された．この選書には他にもすばらしい書がいくつもあったが，いまでは手に入りづらくなっているものも多く残念である．

2　予備知識

2.1　統計的仮説検定

2.1.1　設定

ある確率空間上で定義された確率変数を考え，その確率分布があるかたちの分布でありつつも，母数 θ の値によって変わるという状況を考える．この確率分布の族 $\{F_\theta \mid \theta \in \Theta\}$ を**統計的モデル**という．

いま母数 θ はある特定の値を取っており，そのときの分布 F_θ にしたがって確率変数の実現値が決まる．しかし，観測者にとっては，統計的モデル $\{F_\theta \mid \theta \in \Theta\}$ は既知であるものの，いま実現している母数 θ の値はわからないと想定する．観測者はこの確率変数を n 回，各回独立に観測する．この観測の状況は数式ではしばしば

$$X_1,\ X_2,\ \cdots, X_n \overset{i.i.d.}{\sim} F_\theta \tag{1}$$

などと書かれる．ここで $(X_1,\ X_2,\ \cdots,\ X_n)$ ないしはその実現値 (データ)$(x_1, x_2, \cdots,\ x_n)$ を**標本**といい，n を標本の**大きさ**という．観測された $(x_1, x_2, \cdots,\ x_n)$ をもとに，観測者は θ の値が「どのようなものか」を推測しようとする[1]．これが**統計的な推測**であるが，とくにここで考えるのは**統計的仮説検定**である．それは θ の値が予め定めた特定の範囲に入ってるか否かを推測しようとする．

いま θ が取りうる値の集合 Θ を 2 つに分割し，一方を Θ_0，他方を Θ_1 とする．$(\Theta = \Theta_0 \cup \Theta_1,\ \Theta_0 \cap \Theta_1 = \emptyset.)$ 「実現している θ が Θ_0 に属すること」(あるいは Θ_0 の元の各々) を**帰無仮説**とよぶ．帰無仮説が成り立たないこと，すなわち「実現している θ が Θ_1 に属すること」(あるいは Θ_1 の元の各々) を**対立仮説**という．帰無仮説を H_0，対立仮説を H_1 と表記し，$H_0 : \theta \in \Theta_0$ と $H_1 : \theta \in \Theta_1$ というように書く．Θ_0 (もしくは Θ_1) が 1 点のみからなるとき，帰無仮説 (もしくは対立仮説) を**単純仮説**とよぶ．単純仮説でないときは**複合仮説**とよぶ．

---

[1]標本とは $(X_1,\ X_2,\ \cdots,\ X_n)$ をいうのか $(x_1, x_2, \cdots,\ x_n)$ をいうのかは案外曖昧なようである．ここではとりあえず上のように定義した．このことに限らず，確率変数とその実現値とは場面によって都合よく同一視されることが多いように感じる．

得られたデータから帰無仮説か対立仮説かのどちらが成立しているのかを推測する行為が統計的仮説検定である．その結果を通常，帰無仮説を中心に述べ，帰無仮説の成立を取ることを，それを**受容** (もしくは**採択**) するといい，帰無仮説の不成立を取ることを，それを**棄却**するという．

ここまででの話では，帰無仮説と対立仮説は Θ を 2 つに分けているだけなので，どちらがどちらでもよいように思われる．しかし通常の統計的仮説検定の理論ではそうではなく，帰無仮説と対立仮説とは役割が異なり，非対称に扱われる．したがって Θ の分割の仕方が同じであっても，どちらを帰無仮説にするかで話は全く違ってくる．これについてはすぐあとに述べる．

確率変数の組である標本 $(X_1,\ X_2,\ \cdots,\ X_n)$ の可能な実現値の集合を**標本空間**と呼ぶ．これを $\mathscr{X}$ を表記しよう．すると，検定は関数 $\delta : \mathscr{X} \to \{0,\ 1\}$ と表せる．これを**検定関数**とよぶ．これは，観測値 $x = (x_1, x_2, \cdots,\ x_n)$ を得たときに，$\delta(x) = 0$ は帰無仮説を受容することを，$\delta(x) = 1$ は帰無仮説を棄却することを，それぞれ表している．

検定関数については，受容・棄却について確率化された判断までを含める場合，$\delta : \mathscr{X} \to [0,\ 1]$ と一般化できる．ここで $r = \delta(x)$ は観測値 $x = (x_1, x_2, \cdots,\ x_n)$ を得たときに，確率 r でもって帰無仮説を棄却するということを表している．

帰無仮説が成り立っているのに，これを棄却してしまうことを**第 1 種の過誤**という．帰無仮説が成り立っていないのに，これを受容してしまうことを**第 2 種の過誤**という．もちろん成り立っているのが帰無仮説なのか対立仮説なのかは最終的にもわからないことが多い．したがって上の「帰無仮説が成り立っているのに (いないのに)」というのは仮定の話である．すぐ下で説明するように，検定の良し悪しを語るときに通常問題とされるのはこれらの過誤の確率となる．

検定 δ が与えられたとき，母数 θ に対して，帰無仮説が棄却される確率 $\beta_\delta(\theta)$ を δ の母数 θ のときの**検出力**という．これを関数 $\beta_\delta : \Theta \to [0, 1]$ とみたときは**検出力関数**という．$\theta \in \Theta_0$ に対しては，検出力は第 1 種の過誤の生ずる確率を表している．$\theta \in \Theta_1$ に対しては，検出力は第 2 種の過誤の生じない確率を表している．

2.1.2 検定の考え方

「帰無仮説が棄却されることに重要性をもたせる」というのが統計的仮説検定の通常の考え方である．帰無仮説を棄却すればふつうは (データから帰無仮説の成立を完全に排除できるのでもない限り) 必ず第 1 種の過誤のリスクを負うことになる．帰無仮説の棄却にこそ重要性があるとの考えのもとでは第 1 種の過誤の起こる確率が低く抑えられていることが望ましい．これはつまり「帰無仮説はよほどのことがないかぎりは棄却されないようにしておく」ということである．そうすれば，棄却されればそれはよほどのことであるから，そのことは重要性を持つわけである．通常の考え方では，第 1 種の過誤の起こる確率を予め与えた

ある値 α より小さくなるように検定 δ を設計することが求められる．すなわち，$\sup_{\theta\in\Theta_0}\beta_\delta(\theta)\leq\alpha$ とするわけである．この α を有意水準とよび，$\sup_{\theta\in\Theta_0}\beta_\delta(\theta)$ を検定 δ のサイズという．有意水準をどの値に設定するかはいわば主観的な問題である．具体的にはどうやら習慣によるようであり，0.01 (1 パーセント) や 0.05 (5 パーセント) がよく用いられるとされている．

このように第 1 種の過誤の確率をある水準以下に抑えたうえで，第 2 種の過誤の確率もできるだけ小さくするというのが通常の考え方である．検定 δ^* が**一様最強力検定**であるとは，δ^* が，あらゆる $\theta\in\Theta_1$ について，第 2 種の過誤の確率が最小になるような検定であることをいう．すなわち，有意水準 α を満足する任意の検定 δ に対して $\forall\theta\in\Theta_1$，$\beta_\delta(\theta)\leq\beta_{\delta^*}(\theta)$ となることである．無論このような都合の良い検定がいつも存在するとは限らない．Θ_1 が 1 点のみからなるとき (対立仮説が単純仮説であるとき)，これをたんに**最強力検定**と呼ぶ．

3　ネイマン・ピアソンの補題の直観的理解

3.1　ネイマン・ピアソンの補題

以下では標本 $(X_1, X_2, \cdots, X_n)$ の確率密度関数ないしは確率質量関数を $f(\cdot\,,\theta)$ と表記する．

いま帰無仮説，対立仮説ともに単純仮説の場合を考える．すなわち，

$$H_0:\theta=\theta_0 \quad \text{vs} \quad H_1:\theta=\theta_1.$$

このとき最強力検定が特徴づける，以下の定理が成り立つ．

定理　(ネイマン・ピアソンの補題)

(i) 検定 δ が有意水準 α のもとで最強力検定であるとき，ある $c\geq 0$ が存在し，任意の $x\in\mathscr{X}$ に対して

$$\delta(x)=\begin{cases}1 & (f(x,\theta_1)-cf(x,\theta_0)>0\text{ のとき}),\\ 0 & (f(x,\theta_1)-cf(x,\theta_0)<0\text{ のとき})\end{cases} \tag{2}$$

が成り立つ．

(ii) 有意水準 α が与えられたとき，ある $0\leq r\leq 1$ と $c\geq 0$ とが存在し，任意の $x\in\mathscr{X}$ に対して

$$\delta(x)=\begin{cases}1 & (f(x,\theta_1)-cf(x,\theta_0)>0\text{ のとき}),\\ r & (f(x,\theta_1)-cf(x,\theta_0)=0\text{ のとき}),\\ 0 & (f(x,\theta_1)-cf(x,\theta_0)<0\text{ のとき}).\end{cases} \tag{3}$$

となる有意水準 α の最強力検定 δ が存在する．□

上で，たとえば，$f(x,\theta_1)-cf(x,\theta_0)>0$ は $f(x,\theta_0)\neq 0$ のときは

$$\frac{f(x,\theta_1)}{f(x,\theta_0)}>c \tag{4}$$

と変形できる．この左辺は尤度比と呼ばれる．したがってネイマン・ピアソンの補題は尤度比の大小を受容・棄却の基準とする検定によって最強力検定を特徴付けているものとされる．

3.2　直観的説明

以下ではこの定理を証明はせず (竹村 (2020) 第 8 章ほか，一定レベル以上の統計学の教科書に書いてある)，どうしてこのような結果になるのか，その直観的な理由を説明する．

いま $\mathscr{X}$ を 1 次元とするか多次元とするかは全く本質的ではないので，イメージしやすいよう 1 次元とする．

(i) ひとまず標本 (X_1, X_2, $\cdots$, X_n) の分布が連続分布だと想像してほしい．まず $\mathscr{X}$ を多数の微小な断片 (区間) に切り刻む．これらを (Δx_i) としよう．そして検定 δ は $\mathscr{X}$ の各点ではなく，この各断片 Δx_i に対して 0 か 1 かを割り振るものと考える．各断片に対して「0」か「1」かのラベルが貼り付けられていると想像してほしい．いま有意水準 α の最強力検定を作るには，どのようにラベルを貼ったらよいのか考えてみよう．

まず最初に，全ての断片に「1」を貼り付ける．すると第 1 種の過誤の確率は 1 (100 %) であり，第 2 種の過誤の確率は 0 である．いまラベルを「1」から「0」に順次貼り替えていくこととしよう．

断片 Δx_i を 1 枚貼り替えたとき，第 1 種の過誤の確率は $f(x(i),\theta_0)\Delta x_i$ だけ減少する．一方で第 2 種の過誤の確率は $f(x(i),\theta_1)\Delta x_i$ だけ増加する．(ただしここで $x(i)$ は Δx_i の「位置」である．) この貼り替え作業は「第 2 種の過誤の確率の増加」という「費用」を払うことで,「第 1 種の過誤の確率の減少」という「財」を「購入」するプロセスと見ることができる．こうしてラベルを貼り替えていって，第 1 種の過誤の確率を有意水準 α まで小さくすれば，有意水準をクリアする検定ができあがる．これを最強力検定となるようにするのは，貼り替えの際に第 2 種の過誤の確率の増加の総計 (つまりそれは第 2 種の過誤の確率そのものである)，つまり「費用」の合計をできるだけ抑えればよい．つまり最強力検定を求めることとは「なるべく安い買物をすること」に他ならない．(したがって，これはミクロ経済学の消費者行動の理論の一部である．)

α が具体的に与えられていないから購入がいつ終るかわからない．しかし，どの時点で終ったとしても，その時点までの費用の合計が最小化されているようにするのは易しい．そのためには，財の「単価」，すなわち，財「第 1 種の過誤の

確率の減少」1 単位に対する，費用「第 2 種の過誤の確率の増加」，

$$\frac{f(x(i),\theta_1)\Delta x_i}{f(x(i),\theta_0)\Delta x_i} = \frac{f(x(i),\theta_1)}{f(x(i),\theta_0)} \tag{5}$$

の安い断片から順に，財の量が $1-\alpha$ きっかりになるまで (つまり，第 1 種の過誤の確率がきっかり α になるまで) 購入していけばいいのである．そうすると購入をすすめるにつれ単価はあがっていき，あるところでやめるということになる．この「あるところ」を c とすると

$$\frac{f(x(i),\theta_1)}{f(x(i),\theta_0)} < c \tag{6}$$

となるような断片が全て購入される，つまりそれらの断片 Δx_i にラベル「0」が貼られる，言いかえれば，$\mathscr{X}$ のその範囲では帰無仮説は受容される ($\delta(x)=0$) ということになる．

いうまでもなく，反対に

$$\frac{f(x(i),\theta_1)}{f(x(i),\theta_0)} > c \tag{7}$$

となるような断片 Δx_i にラベル「1」が貼られ，その範囲の $x \in \mathscr{X}$ に対しては帰無仮説は棄却される ($\delta(x)=1$) こととなる．これで最強力検定が得られたし，最強力検定はこのようなものしかないこともわかる．

ただし $\frac{f(x,\theta_1)}{f(x,\theta_0)} = c$ となるような x については若干の注意が必要である．このような x がただ 1 点となる場合にはその確率は 0 であるので重要ではない．しかし，そうならないときには $\delta(x)=0$ とするか $=1$ とするかは自動的には決まらないので，有意水準がちょうど α になるように恣意的に決めなければならない．1 つのやり方はこのような x に対しては特定の確率 r で 1 を選ぶようにすること ($\delta(x)=r$) である．

(ii) いま標本の分布が離散分布であるとしよう．この場合は $\mathscr{X}$ を切り刻むまでもなく，データ x の取る点は離散的で，いわば始めから断片をなしているから，それを前提に上の (i) と同じようにすればよい．

ただし，断片の刻み方は最初から与えられているので，第 1 種の過誤の確率がきっかり α になるところで購入をストップできるとは限らないが，この場合は δ を適宜確率化すればよい．

3.3　図による説明

上で説明したことを図を用いて考え直してみよう．下の図 1 のようなものは竹村 (2020) (第 5 章，8 章) をはじめ教科書にときどき載っている．図の正方形のなかの任意の点は第 1 種の過誤の確率と第 2 種の過誤の確率の組を表している．これを「確率ベクトル」と呼ぼう (これは一般的ではないここでの暫定的な呼称である)．検定 δ を 1 つ特定すれば，それに対応する確率ベクトルが決まるので，

確率ベクトルはいずれかの検定によって実現可能なものとそうでないものとにわけられる．

いま標本の分布が連続分布の場合 (3.2 節 (i) の場合) を考えてほしい．図の凸レンズ状の領域の内側 (境界を含む) が実現可能な確率ベクトルの集合を表しているのだが，これを得るためには以下のように考える．まず 3.2 節の要領で $\mathscr{X}$ を微小に切り刻んだ断片 (Δx_i) を尤度比 $\frac{f(x(i),\theta_1)}{f(x(i),\theta_0)}$ の小さいものから順番に並べ，断片の全てにラベル「1」を付ける．このとき確率ベクトル (の先端) は点 A となる．これが出発である．

(i) さて第一の場合として，尤度比の小さい断片から順次ラベルを「1」から「0」へと変えていくと，確率ベクトルは点 A から動いていき最終的に全部が「0」になったときに点 B へとたどり着くが，その軌跡は図のように原点に対して凸の曲線 (下側の曲線) になる．これを「下限曲線」と呼ぼう．少し見方を変えると，$\mathscr{X}$ の微小断片ラベルの「1」から「0」への貼り替えは，点 A から点 B まで微小な曲線をつなぎ足していく作業に対応する．ここで各微小曲線の傾きの絶対値が尤度比に一致するわけである．

(ii) 第二の場合として，こんどは点 A から尤度比の大きい断片から順次「1」から「0」へと変えていくと，同様にして，原点に対して凹の曲線 (上側の曲線) が得られる．これを「上限曲線」と呼ぼう．ここで上限下限の 2 つの曲線は互いに正方形の中心について回転対称となることを注意する．

(iii) 第三の場合として，点 A から尤度比の大小にかかわらず断片を順次「1」から「0」へと変えていくと，その軌跡は上限曲線あるいは下限曲線 (どちらでもよい) を微小な曲線に切り刻んで，しかしその各微小曲線の傾きは変えずにつなぎ直したものとなる．どのようにつなげても最後は点 B にたどり着く．このような曲線は上限下限 2 つの曲線に挟まされた領域をとおる．さらに確率化された検定も考えれば，この領域 (これは凸集合であるから) の全ての確率ベクトルが実現可能であることがわかる．

以上 3 つの場合より，この上限下限 2 つの曲線とその内側の部分が実現可能な確率ベクトルの集合を表すことがわかる．このことがわかれば，有意水準 α の最強力検定を得ることはやさしい．まずその最強力検定の確率ベクトルは点 D となることは明らかである．(明らかに，これが第 1 種の過誤の確率が α 以下の実現可能な確率ベクトルのうち第 2 種の過誤の確率が最小のものである．) いま下限曲線の点 D における (接線の) 傾きを c とする．さて点 A から出発して微小曲線をつなげていって点 D にたどり着くには，ただたんに点 A から点 D まで下限曲線に沿ってすすむか，あるいは下限曲線の点 A から点 D までを微小に切り刻んで並べ替えればよいし，そうするほかない．これらの微小曲線の傾きの絶対値はいずれも c よりも小さいから，これは尤度比が c より小さい断片に「0」を付けることを意味する．反対に尤度比が c より大きい断片には「1」が付く．このように，尤度比の大小による基準で最強力検定が得られることがわかる．

標本の分布が離散分布の場合 (3.2 節 (ii) の場合)，あるいは上限 (下限) 曲線が直線部分を持つ場合などでは適宜調整が必要だが，本質的なところは上に示したものと変わらない．

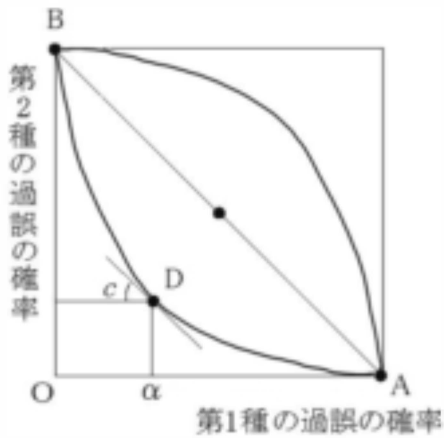

図1

4　統計的仮説検定についての感想

ネイマン・ピアソンの補題から敷衍し統計的仮説検定の考え方についてのごくざっくばらんな感想を雑駁に記す．内容はこれまでに他処で繰り返し書かれていることと多く重複するであろう．

4.1　仮説検定の考え方に対する混乱

統計的仮説検定の根幹の考え方についての説明として「珍しいことが起こったら，何か他に理由があると考えたほうがよい」ということだとするものがある．例えば，このコインには歪みがないとの仮説を立てて，実際に 10 回を投げてみたら，10 回とも表が出た．では仮説は間違っていて本当はコインが歪んでいるのではないかと考えようとなる．なんとなればもし仮説が正しければそのようなことが起こるのはとても「珍しい」からである．

このたぐいの説明や例は多くの統計学の教科書に述べられている．こういう説明が間違っているとは全然思わないが，しかしそこにはある種の混乱が見られ，そのことが仮説検定の考え方をかえって理解しづらくしているように思われる．この点を説明するためにつぎの例を考えよう．

いまコインがあり表が出る確率 p が 0.5 か 1 かのどちらかであることが分かっている．このとき $p=0.5$ を帰無仮説とする．すなわち，

$$H_0 : p = 0.5 \quad \text{vs} \quad H_1 : p = 1.$$

そのうえで，コインを 1 回だけ投げるという検定を行う．

(i) いま有意水準のことはおいておき，この検定で帰無仮説を棄却するのなら「表が出たときに棄却する」ということ以外にはありえない．しかし帰無仮説が正しかったとして表が出ることが「珍しい」わけでは全くない．あるデータが帰無仮説の棄却にふさわしいかどうかは「珍しさ」の問題ではなく，他のデータと比較したとき対立仮説に対していかに支持的であるかの問題である．

(ii) つぎに有意水準のことを考えよう．当り前のことだが (その割には言われないと気が付かないことかもしれないが)，有意水準をクリアするだけなら，データを全く見ないでたんに所望の確率を設定したクジを引いて帰無仮説を棄却すればそれでよい．棄却の確率を低く設定すれば，それは当然「珍しい」ことである．しかしその「珍しさ」とはそのときのデータの内容が帰無仮説の棄却にふさわしいか否かとは全く関係がない．

(iii) 上の 2 つの話を合わせれば「表が出たらくじを引いて有意水準をクリアする確率で帰無仮説を棄却する」といういかにも妥当な検定が出てくる．帰無仮説が棄却される確率が有意水準ぴったりになるようにくじの確率を調整してやれば，この検定が最強力検定になる．

要するに「データに対して適切に帰無仮説の受容・棄却を決めること」と「有意水準を低くおさえること」とは本来全く別の要請であって，これらを混同してはならない．無論状況によってはこれらの要請の含意が重なってくることはある．むしろそれが通常であることから，この区別はわかりづらくなっている．「「珍しいこと」が起こったら帰無仮説を棄却する」というよくある説明は常識的でわかりやすくはあるが，これら 2 つの要請を区別しないことで理解を混乱させているというのが，わたくしの感想である．

4.2 尤度比基準への直接的アプローチ

ネイマン・ピアソンの補題のメッセージは「尤度比の大小によって帰無仮説の受容か棄却かを決めよ」ということである．それを具現化しているのが最強力検定の概念なわけであるが，これによってかえって尤度比を使うことの意味がわかりづらくなっているきらいがある．上ではそのわかりづらさを購買行動とのアナロジーで解決したわけであるが，ここでは代替的なアプローチを考えてみよう．以下では，データを見て帰無仮説の受容・棄却を判断する際の「自然な態度」とはどんなものかを考え，そこからネイマン・ピアソンの補題で述べられている尤度比を用いた基準を導く．

前節で議論したとおり，あるデータに対して帰無仮説を棄却すべきかは (それが帰無仮説のもとで「珍しい」かではなく) 他のデータに比較してより対立仮説に対して支持的であるかで決まるべきである．この考えを定式化することから出発しよう．

以下では簡単化のために，標本空間 $\mathscr{X}$ を有限とし，任意の $x \in \mathscr{X}$ に対して $f(x, \theta_0)$，$f(x, \theta_1) > 0$ とする．まず，いま検定 δ において，あるデータ $x \in \mathscr{X}$ に対して帰無仮説を棄却したとする ($\delta(x) = 1$)．このとき別のデータ x' を拾い上げてみて，x とその尤度を比較したら，$f(x', \theta_1) > f(x, \theta_1)$ かつ $f(x', \theta_0) < f(x, \theta_0)$ であった．これをもって x のもとでよりも x' のもとでのほうが対立仮説 θ_1 はより尤もらしいとするのは全く無理がない．ならば x' に対しても帰無仮説を棄却する ($\delta(x') = 1$) のが自然であろう．この考え方は δ についての以下の条件にまとめることができる．

単調性　いかなる 2 つのデータ x, $x' \in \mathscr{X}$ に対しても，もし $[f(x', \theta_1) \geq f(x, \theta_1)$ かつ $f(x', \theta_0) \leq f(x, \theta_0)]$ であるなら，$\delta(x') \geq \delta(x)$ が成り立つ．

つぎに，同等性について考えよう．2 つの別々のデータが受容・棄却の判断において同じく扱われるべきなのはどのような状況だろうか．2 つのデータ x, $x' \in \mathscr{X}$ を考える．いまある実数 $\alpha > 0$ が存在し $f(x', \theta_0) = \alpha f(x, \theta_0)$ かつ $f(x', \theta_1) = \alpha f(x, \theta_1)$ が成り立つとしよう．このとき x と x' とに対しては受容・棄却の判断を同じくするのが妥当であろう．なぜならば，いま母数 θ の条件付き確率を考えれば，$\Pr\,[\theta_0 \mid x] = \Pr\,[\theta_0 \mid x']$ かつ $\Pr\,[\theta_1 \mid x] = \Pr\,[\theta_1 \mid x']$ が成り立つからである．つまり x と x' とは母数の成立の確からしさに対して全く同等な情報を持つわけである．

ここで母数 θ の確率を考えているのは明らかにベイズ主義的である．統計的仮説検定を含む古典統計の考え方では，どの母数が成立しているかはあくまで「未知」なのであって，そこに確率を付与できるものではない (付与すべきではない) とされる．したがってこれは明らかな逸脱であるので，上の議論を許容するか否かは議論のあるところであろう．ただし，上の 2 つの確率等式を得るには Θ 上の事前分布を知っている必要はない．必要なのは確率を付与できるという事実であって，その確率自体は未知で全然よいわけである．確率が付与できるがそれを知らない (知れるほどの情報を持たない) ということと，未知ゆえに確率が付与できないということとはどのように異なるのかは興味ある哲学的論題であるが，ここで立ち入ることはしない．

上述の単調性にいま議論した同等性の条件を取り入れることで以下の条件を定義することができる．この条件は受容・棄却の判断における「自然な態度」を規定するものとみなせる．

広単調性　いかなる 2 つのデータ x, $x' \in \mathscr{X}$ に対しても，もし，ある 2 つの正の

実数 α，β に対して，$[\alpha f(x',\theta_1) \geq \beta f(x,\theta_1)$ かつ $\alpha f(x',\theta_0) \leq \beta f(x,\theta_0)]$ が成り立つのならば，$\delta(x') \geq \delta(x)$ が成り立つ．

検定を非確率的なものに限定したとき，上の条件から「尤度の大小による帰無仮説の受容・棄却」が帰結することが示せる．

定理　もし $\mathscr{X}$ 上の非確率的な検定 δ が 広単調性 をみたすとすると，ある実数 $c > 0$ が存在し，任意の $x \in \mathscr{X}$ に対して

$$\delta(x) = \begin{cases} 1 & (f(x,\theta_1) - cf(x,\theta_0) > 0 \text{ のとき}), \\ 0 & (f(x,\theta_1) - cf(x,\theta_0) < 0 \text{ のとき}) \end{cases} \tag{8}$$

が成り立つ．

証明　いま $\sigma(x)$ でもって $\frac{1}{f(x,\theta_0)+f(x,\theta_1)}$ を表す．任意の 2 つの x'，$x'' \in \mathscr{X}$ を取る．このとき $\sigma(x')f(x',\theta_1) \geq \sigma(x'')f(x'',\theta_1)$ ならば必ず $\sigma(x')f(x',\theta_0) \leq \sigma(x'')f(x'',\theta_0)$ となることを確認するのは容易である．したがって検定の非確率性と広単調性とにより，$\mathscr{X}$ の元 x を $\sigma(x)f(x,\theta_1)$ の大きい順に並べれば，ある元 $x^\star$ とそれより上では全て $\delta(x) = 1$，それより下では $\delta(x) = 0$ となる．いまこの順序で $x^\star$ のすぐ下のものを $x^{\star\star}$ とする．そうして $\frac{f(x^{\star\star},\theta_1)}{f(x^{\star\star},\theta_0)} < c < \frac{f(x^\star,\theta_1)}{f(x^\star,\theta_0)}$ をみたす c を任意に取れば定理の主張のとおりの式が成り立つ．□

はじめに述べたとおりこの導出は最強力検定の概念とは関係がない．それどころか 2.1.2 節に説明した統計的仮説検定の通常の考え方すらも前提にしていない．要になるのはベイズ主義的な同等性であり，それが尤度を比でもって比較する根拠を与えている．

4.3　棄却の含意

上に述べたとおり，統計的仮説検定による立証では，帰無仮説が棄却されることが重要性を持っている．帰無仮説が棄却されるということは対立仮説が支持されるということとされ，それによってなにかを立証するというやり方が取られる．このロジックもまた少なからぬ混乱の原因となっているように見受けられる．

対立仮説は帰無仮説の否定であるから，帰無仮説が棄却されれば対立仮説が支持されるのは当然と思われるかもしれないが，そうではない．問題は対立仮説は帰無仮説の論理的な否定であるが，帰無仮説がデータをもって棄却されるときには，それは論理的に否定されるわけではないということである．ならば，対立仮説を支持する根拠はどこにあるのか．管見のかぎり，統計的仮説検定の説明ではその根拠を欠いているように見える．(ここで問題にしていることは「統計的な判断は完全に確実なものではない」という一般的な注意よりもはるかに微妙なことである．いま問題にしていることは，そういう不確実な判断を受入れたとし

て，その含意を確実な判断のそれとどこまで同じように扱えるのか，あるいはどのように扱ったらよいのかである．）このことを理解するために以下の例を考えよう．

いまコインがあり表が出る確率 p が 0.5 か 0.501 かのどちらかであることが分かっている．(このようなコインを実際に作るのは不可能であろう．しかし計算機シミュレーションで同じ状況を設定することはかんたんであり，したがって全然非現実的な状況ではない．）このとき $p=0.5$ を帰無仮説としてコインを 10 回投げるという検定をする．すなわち，

(A) $H_0 : p = 0.5 \quad \text{vs} \quad H_1 : p = 0.501.$

さて，いまコインを投げてみたら，驚くことに，10 回とも表だった．帰無仮説が正しいとするとそうなる確率は $0.5^{10} = \frac{1}{1024}$ であり，有意水準が 5 %でも 1 %でも帰無仮説 $H_0 : p = 0.5$ は棄却されるのが妥当であろう．しかしこれをもって対立仮説 $H_1 : p = 0.501$ が支持されたとするのは全くバカげている．なぜならこれは対立仮説のもとでも同じ程度に起こりそうもないできごとだからである．むしろこの場合，常識的に正しい判断はそもそも「表が出る確率 p が 0.5 か 0.501 かのどちらか」という設定そのものを疑うことであろう．

ではもし設定が以下のようであったらどうであろうか．

(B) $H_0 : p = 0.5 \quad \text{vs} \quad H_1 : p = 0.9.$

この場合は「10 回とも表」というデータに対して H_0 を棄却し，したがって H_1 を支持するのは全く妥当であろう．

では $H_1 : p = 0.502$ ならどうか．$H_1 : p = 0.503$ ならば．(A) から出発してこのように確率を上げていけばいつかは (B) になるはずだが，その途中のどこかで (徐々にでも) 対立仮説を支持することが「莫迦げている」から「妥当である」にはなるはずである．このこと自体は常識的な判断の範囲と思うし (H_1 が正しいとしたときの確率と比較するなどできる)，もっときちんとした理論もきっとだれかが作っているであろう．しかし，通常の統計的仮説検定の説明を受け入れるかぎりは，そのような判断を行う論理は仮説検定の論理とは整合しないように思われる．

4.4 科学的と客観性

巷では数理統計学の方法は客観的な判断の権化のように扱われて（思われて）いるが，それは適切ではない．昨今の科学の実践では，しばしば特定の統計的方法の形式がみたされていることがあるレベルの客観的な立証がなされたことと同等とみなされ，結果として驚くほど多くの不適切な結論が出されている．これは現代の科学が直前している深刻な危機的状況の大きな部分をなしている．

統計的仮説検定だけを取ってみても，当然ながら主観的な要素は多くある．(有意水準の主観性については2.1.2節で注意した．) たとえば，統計的仮説検定の理論では既知の部分と未知の部分とを明確にわけており，統計的モデル $\{F_\theta \mid \theta \in \Theta\}$(とデータの観測のされ方) は既知であり，実現している母数が未知とされる．したがって，そこでテストされるのは母数に関しての仮説であり，統計的モデルそのものではない．しかし実際の応用においては，理論通りの統計的モデルが成り立っておりしかもそれが既知であるということは，人工的に設定された状況を除いては，ないであろう．未知のものに対して既知を装うならば，それは主観的な選択にならざるをえない．

統計的仮説検定は主観的判断を土台にした局所的な客観性を実現しているに過ぎない．と言って，そのこと自体は良くも悪くもない．そういうものだと分かって適切に使えばよいだけのことなのであるが，悪意なのか誤解なのか，濫用されることが多いのは問題視されるとおりである．

統計的仮説検定 (および，それに代表される統計的方法) の濫用の大方は科学者の不誠実によるものだと思うが，しかし，より深い原因として，その根底にあるニヒリスティックな科学観と，そのような科学観と妥当な科学的思考との不整合とがあると思う．仮説検定の理論には「科学的な判断とはなにか」ということに対してのある種の態度や姿勢が込められている．しかし，それはある時代のあるグループの人たちの「科学哲学」であって，それが必ずしも科学的思考の実情と合致しているわけではないし，科学の理想的な思考形式を体現しているわけでもない．わたくしの見たところ，それは本来の科学のあり方に比較して過度に殺伐としており，しかもそれが科学に還流することにより科学を殺伐としたものに変えていっている．

統計的仮説検定の考え方はカール・ポパーの反証主義に通じている．反証主義は科学を科学理論の生き残りの消去プロセスと特徴づける．そこでは，提案された科学理論は，その含意が観測や実験などによる「事実」と照合され，整合しなければ失格となり，捨てられる．これがすなわち反証である．事実との照合は客観的で論理的な手続きであり，そのような手続きによって捨てられないような曖昧さや流動性を持つ理論は科学理論としての資格がないとされる．すなわち，科学理論は反証可能でなければならない．ポパーによれば，科学の進展は反証とそれを受けた新たな理論の提案というプロセスの繰り返しであり，現時点での科学は当該生き残っているものが生き残っているだけである．

しかしこのような考え方は初期の科学哲学者の理想化された思い込みに過ぎない．反証主義は当時の科学哲学の展開からすれば非常に重要な貢献であったのは間違いないだろうが，今日日これを額面通りに受け取っている科学哲学者は非常に少ないであろう．しかし反証主義は今日においても科学の中心的なテーゼであるかのように説明されることがある．実際そのように考えてる研究者にもよく会った．反証主義は非常にわかりやすく魅力的なところがあり，このような考え

方が広まってしつこく残り続けたことが，とくに科学者自身の科学に対する理解を歪めたように思われる．

実際の科学はポパーの言うようにはすすまない．科学哲学の書を開けばそのような例が挙げられているであろうが，ここでは，わたくしの好きな例として，ダークエネルギーについて取り上げよう．現代の科学では宇宙にはダークエネルギーと呼ばれる正体不明の巨大なエネルギーが存在するとされている．その根拠はIa型超新星爆発の観測データである．そのデータから宇宙の膨張の速さについての算定ができ，それによれば宇宙の膨張は加速している．それは観測から推定される宇宙のエネルギーの総量とは整合しないが，ダークエネルギーが存在しているとすれば整合するわけである．

しかし超新星爆発の観測データがダークエネルギーの存在を意味していると必ずしも考える必要はない．たとえば観測データそのものの解釈が適当ではなかった可能性もあり，実際にそういう研究もあった．あるいは，いっそ超新星爆発の観測データをビッグバン理論への反証だと解釈することもできる．そうした複数の可能性のあるなかで，理論的にも全く正体のわからないダークエネルギーなるものを想定し，前にすすむというのが実際に行われていることである．これはポパーが非科学的と断じる流動的な（「あとづけの」、「アドホックな」）説明に他ならない．むしろポパーの見解に忠実ならば，これをもってビッグバン理論は御破算にすべきであろう．

明らかに科学者たちは，こうした既存の理解と整合しない観測事実に直面して，既存の理論を捨てること，既存の理論を保持したうえで新しい要素で辻褄を合わせることを含む複数の進むべき道を天秤にかけている．(科学哲学を知っている人はデュエム・クワインのテーゼを想起するだろう．）論理的に可能な多数の選択肢のなかから，ダークエネルギーの存在によってビッグバン理論を維持するのが適切だと考えているのである．そう考える理由をかれらは説明できるであろうし，当然それは科学的で合理的な説明であろう．しかしそれは主観的な判断である．客観的には上述したようなさまざまな可能性があり，そのなかから，いわばダークエネルギーに賭け，そこに資源も努力も投入しようというのである．これはたとえ事実にもとづいた合理的判断であるとしても，他にもそう判断できる選択肢がある以上，それのみで説明しうるものではない．科学者たちは明らかに科学が「いかにあるべきか」についての主観的な価値判断による選択を行っているのである[2]．

このような主観的判断をしないかぎり，科学は進歩しようもない．主観的な判断は重要でも適切でもありうるのであって，主観的だから程度が低く価値が低い，客観的なのが程度も価値も高いというのは現代社会にはびこっている病的な思い込みである．その根底には人間の価値を信じず，それよりむしろ機械的な手

[2]全然余談だが，わたくし個人はダークエネルギーなどなく，ビッグバン理論が間違っているのだと思う．

続きや形式あるいは機械そのものに信をおく深刻なニヒリズムがある．

人間の行う判断はつねに主観的な要素を含むのであって科学的な判断も例外ではない．それどころか，主観性は健全な判断に必要なものであり，科学においても同様である．健全な判断には健全な主観が必要なのであり，判断の主観的な部分は主観的と認めたうえで，その適切性を検討すべきである．念を押しておくが，わたくしは科学が非論理的で主観的な思い込みにもとづいたものでよいと言っているのではない．論理的でならねばならぬところも，客観的でならねばならぬところもあり，そういうところはそうでなければならない．しかし主観的な部分，非論理的な部分はそう認め，さらにそうであることの必要性と意義とを認識せねばならないと言っているのである．

客観性と形式性とは緊密に結びついており，通常客観的な判断は形式的，機械的な方法をもって実現されるものと考えられている．反対に主観的な判断は形式的である必要はないし，ふつう形式化できるものでもない．したがって主観的な判断が存在していること，それが必要なことへの認識の欠如が，形式性，機械化への盲従とそれを盾に取った自己正当化につながっていくのは自然なことであり，それこそが現にいま科学の実践で起きていることである．

恣意的な論証が一定の形式が満たされていることでありもしない客観性を装い，そのうえで客観的であることを盾に取り，議論を封殺し判断を押し通す．あるいは逆に元々実現できないような (あるいは必要もないような) 形式性，客観性を要求することで，それらがみたされていないとして他人の議論や結果を退ける．特定の形式をみたしていないことで「客観的ではないから科学的ではない」と断する．などなど，主観性を認識せずに客観性にうったえることによって，間違った結論が押し通され，本来有意義で生産的でありうる多くの科学的議論が非生産的な方向に導かれている．

科学の進歩を，ひいては人類の進歩を牽引するのは人間の主体的な選択そのものであったし，またそうであるべきである．ポパーの描くように，人間は理論を作りはするが，それらは論理的で客観的な手続きよって裁定され，生き残るべきものが生き残るだけであるという科学観は，あたかも人間の行いを機械が裁くかのごときものであり，人類が自らの行くべき道を自らの意志で積極的に選び取るという理想も希望も欠如している．ポパーはかれの謳うような裁定を自由な議論によって行う「開かれた社会」を提唱した．それは聞こえは美しいが，その開放性は予め決まっている勝者を効率的にあぶり出すためのものにすぎない．そしてその裁定は定義により勝者に有利なように準備されたものである．このような世界観は病的なニヒリズムの極致にわたくしの目には映る．

このようなニヒリズムの先に待っているものは，人間が AI に裁定され服従して生きるような社会であろう．そして，その AI の背後にいるのは人間である．「開かれた社会」で「議論」に勝った連中である．それはわたくしの目にはディストピア以外の何物でもないが，それを望んでいるとしか思えない人たちが多く

いる以上，そのような社会は目の前かもしれない．

統計的仮説検定に話を戻る．ところでネイマンとピアソンの原論文 [2] を読むと，かれらの統計的仮説検定の理論に対する理解は現在一般的に (通俗的に) それが扱われている仕方とはかなり異なるように思われる．かれらの明言するところでは，統計的仮説検定は個々の仮説の真偽については (確率的にすらも) なにも主張しない．それは１つの「行動のルール」であって「それにしたがっていれば長期的にはそんなに間違うことはない」(“in the long run of experience, we shall not be too often wrong.’) という行動パターンを与えているにすぎない．これはこの理論の内容のあらわすところそのままであり，全く穏当で妥当な理解であろう．しかし，ひとたび形式化，機械化されてしまえば，それは本来の限界を飛び越して使われるようになってしまう．AI と同様である．その背景には上で批判したようなポパー的でニヒリスティックな科学観が強くあらわれており，ネイマンやピアソンもその世界観のなかにいたように，わたくしには思える．そしてそれは現代において現代の仕方で科学を侵食している．

最後に，主観性について少し異なった論点を付け加えておく．科学における主観的判断は集団的な判断でもある．もちろん個々の科学者がまず個人的に主観的判断を行うことが土台にあるが，それが科学的に意味のある判断になるのは集団レベルになったときである．べつの言い方をすれば，科学においてはただ一人の人間が「正しきこと」を知るものとなることはあり得ない．いや何度もあったではないかと思う人もいるかもしれないが，理論なり事実なりが「正しきこと」となるのは他の大勢がそれを認めたときであって，そうならない限りはそれはたとえ「正しく」ても「正しきこと」とはならないのである．かように科学はその本質において集団的な営みである．したがって科学的判断の十全な特徴付けはその集団性を考慮していなくてはならない．ここに詳しく論ずることはしないが，ゲーム理論と社会選択理論の知見によれば，集団的な選択と判断には個人的なそれらとは別次元のさまざまな問題や困難が存在する．科学はある部分では社会選択の一種である．十全な「科学哲学」は集団的主観にもとづいた社会選択の説明を含んでいなくてはならない．(トマス・クーンの有名なパラダイム論はその方向での説明の初期のものであろう．) そしてそれは統計的推測の理論にもある程度は同様かもしれない．少なくともそれが科学的論証において現行謳歌している特権的な地位に見合うものであるためには．

現代の科学はインチキが横行する危機的な状況にある．その少なからぬ部分が統計的方法の濫用に関係している．なかんずく統計的仮説検定はそのチャンピオンであろう．科学の実践において，権威付けられた機械的な手続きを盾にとって「客観的で論理的」などというありもしないファンタジーを押し通していれば人類の未来は明るくない．数理統計学それ自体はすばらしい学問であると思う．それは見事なアイディアに溢れた厳密で有用な応用数学である．しかしそれは使い方次第であり，それを使う者は実直な理解と何よりも健全な精神とを持たねばな

らない．以上が今回わたくしが数理統計学を学生たちとともに学んでの感想である．

謝辞

小林佑衣さんと毛利玲雄君 (ともに新潟大学大学院学生) とは大学院のゼミで数理統計学について，またそれ以外のさまざまな知的話題について話し合うことができ非常に楽しくまた有意義であった．篤く感謝申し上げる．

参考文献

[1] 竹村彰通 (2020) 「新装改訂版 現代数理統計学」 学術図書出版社．

[2] Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231: 289–337.