統計数理研究所彙報
第32巻 第2号（1984）

確率の解釈における困難について

統計数理研究所 赤 池 弘 次

（1984年6月 受付）

はじめに

統計的方法の終局的な目的は，得られるデータにもとづいて，推論に必要な確率分布を合理的に構成することにある，とみなすことができよう．したがって，確率の解釈にまつわる混乱を避けることなしには統計的方法の研究の健全な発展は望めない．

確率の考えは，サイコロ様のものを用いて将来を占う，あるいは賭けるといった形で人類の歴史の古くからあったと見られる．17世紀に至ってそれに数学的表現が与えられるようになったが，その解釈を巡っては混乱が絶えない．中でもよく知られているのが，いわゆるベイズの方法における先験的確率分布の解釈に関する，無限に続くかに見える議論である．

この先験分布の選択決定の問題も，これを単に通常の統計的モデルの選択あるいは評価の問題と考え，平均対数尤度の期待値すなわちエントロピーを評価の基準とすれば，適切な処理が可能である，とするのが筆者の立場である．この立場に立てば，統計的モデルの構成あるいは選択は，期待される平均対数尤度を大きくするような努力の継続によって実現されることになる．これが筆者のいうエントロピー最大化の立場であり，この立場に立てばベイズ流の方法を実用化する上での概念上の困難は消失し，それぞれの問題に対して適切な確率的構造を想定するという基本的な作業の重要性が一段と明瞭に認識されるわけである．

このようにして新しい考え方が展開されようとするとき，確率そのものの解釈を巡る混乱は統計的方法の発展の阻害となる．本報告では，K.R. ポッパーの〝理想的証拠のパラドックス〟，D.V. リンドレーのいわゆる〝リンドレーのパラドックス〟等を例として，確率の解釈が如何に細心の注意を要するものであるかを明らかにし，適切な統計モデルの構成の妨げとなる先入観念の除去に寄与することを試みる．

1. ポッパーの〝理想的証拠のパラドックス〟

確率が多くの哲学者の関心を惹いたことは歴史的な事実である．R. カルナップ等の論理実証主義の立場に対する批判を通じて知られるK.R. ポッパーには，「科学的発見の論理」というよく知られた著書がある．この書物の新版に附加されたノート（Popper, 1968, 406～419頁）に，「証拠の重み」（weight of evidence）の議論がある．

「証拠の重み」の概念は，プラグマチズムの祖といわれるC.S. パースによって早くに論じられているものであるが（Peirce, 1878）I.J. グッドの書物の題名に登場することで統計学の研究者には知られている．ポッパーはいわゆる主観的確率に批判的な立場をとっており，「証拠の重み」の概念の議論が，主観的確率論の内部では解決しようのないパラドックスに導くことをこのノートで論じている．

このパラドックスの一例としてポッパーが掲げるのが次の「理想的証拠のパラドックス」で

ある．zをひとつの貨幣とし，aによって「zを投げてそのn回目（未だ観測されていない）に表が出ること」という記述を表わすことにする．主観的理論に従えば，貨幣について何等の知識も持たないときこのaの先験確率を1/2と置くことができよう．すなわち

$$P(a)=1/2$$

である．今eによってひとつの統計的な証拠（statistical evidence）を表わすことにする．たとえば何千回あるいは何百万回かzを投げた時に何回表が出たかの記録である．

この証拠eが，zが正確に表裏対称に作られているという仮説に対して理想的に好意的であったとする．たとえば百万回投げた中で表の出た回数が50万回プラス・マイナス20回の範囲に入っていたというようなものである．このとき当然

$$P(a, e)=1/2$$

と考えるしかないであろう．（$P(a, e)$はポッパーの記号で，通常$P(a|e)$と書かれるものすなわちeという条件の下でのaの確率を示す．）

これはeという証拠がn回目の結果が表であると見る確率に影響を与えないことを意味する．というのは

$$P(a)=P(a, e)$$

だからである．この結果は，主観確率の理論に従うとeという理想的な証拠がaに関して全く意味の無い情報となることを示す．

以上がポッパーの「理想的証拠のパラドックス」の提示である．上記の結論は，われわれの仮説aに対する「合理的信念の度合い」を示す確率が，蓄積された証拠の知識によって全く影響されないことを示すもので，これは驚くべきことだというのである．これに続いてこのパラドックスが主観的確率論の枠内では解けないとポッパーが考える理由が示されるが，その議論はここでは不要である．

ポッパーの議論が示すように，確率の議論は客観主義と主観主義の接点に登場する．このことは，たとえばAgassi（1975）とそれに続く討論が明らかに示している．

2. ポッパーのパラドックスの解釈

ベイズ流の確率的構造の構成に慣れた読者には，ポッパーのパラドックスが実在しないものであることが明らかであろうが，ここでは確率的構造の構成に関するひとつの例としてこのパラドックスを解析してみる．実はポッパーによる言葉の不正確な取扱いからこの問題が生じているとみなすことができるのである．

まずはじめに，「主観的理論においては，このaの先験確率を1/2と置くことができよう．すなわち$P(a)=1/2$である．」という記述が問題である．証拠eが与えられる前に，何の意味で$P(a)=1/2$とおけるのであろうか．ここではC.S.パースが既に力説しているように（Peirce, 1878），前提条件を明示することなくある事柄の確率を論ずることは全く無意味であることに注意すべきである．

zという貨幣について予備的な情報が全く無いときには，a（n回目に表が出るという論述）が成立する確率についてもわれわれは確かな知識を持たない．したがって，主観確率の理論に素直に従えば，$P(a)$自体が確率変数ということになる．すなわち$P(a)$がある値xをとる確率（密度）

$$P(x)=P(P(a)=x)$$

を考えなくてはならない．このように見るとポッパーのパラドックスはたちまち解消してしまうのである．

もっともこの見方は，ある事柄の確率はかならずその場面で一意的に定まるものだとする，B. ド・フィネッテイ，L.J. サベジの流れを汲む「純粋な主観確率の立場」には矛盾する．この立場では，定まった値を持たない確率というものは存在しないとするからである．この立場に対してはポッパーのパラドックスは消失しない．

しかしパースの立場に従えば，確率はその時にわれわれが持っている知識に依存して定まるものであるから，知識の量によってはこれが一意的に定まらない状態があると考えるのが自然であろう．そこでここでは実際の形はともかく，$P(x)=P(P(a)=x)$ という $P(a)$ のとる値についての確率分布があるものと想定しよう．

理想的な証拠 e が与えられたとき，ベイズの方法に従えば，$P(x)$ は $P(x|e)$ に置き換えられる．$P(x|e)$ は $x=1/2$ で著しく高い値を取り，それ以外の値では極めて 0 に近いというのが理想的な証拠 e の特徴である．もとの $P(x)$ は $P(a)$ の値に関する不確からしさのために，当然広い範囲の x の値にわたって分布していた筈である．こうして e という証拠の知識は，われわれの $P(a)$ に関する合理的な信念の表現としての確率分布 $P(x)$ を，それとは明瞭に異なる $P(x|e)$ に変換するのである．

上記のようなパラドックスの解決は，柔軟性の無い主観主義的なベイジアンでない限りは容易に考えつくもので，前に触れた Agassi (1975) に対するコメントの中で，I.J. グッドもこのような指摘を行っている (Good, 1975)．

3. トゥベルスキによる主観確率の評価事例に対するリンドレーのコメント

さて前節の例はあまりにつまらないと思われたかもしれない．ポッパーは主観主義を批判する側の立場から前述のような議論を展開したが，実は全く同様な例を今度は主観確率の立場を主張する側の人が異なる立場を批判するために提示している．これを見ると，この問題が確率的構造の構成について極めて教育的な側面を備えていることが推測されよう．

よく知られているように，D.V. リンドレーは B. ド・フィネッティ，L.J. サベジの流れを汲むベイジアンである．彼はベイズ的方法と正統的な検定理論とが相矛盾するものであることを示すものとして「リンドレーのパラドックス」を論じた (Lindley, 1957)．このパラドックスについては後で言及するが，ここではこのパラドックスの G. シェイファーによる議論 (Shafer, 1982) に対するリンドレーのコメント (Lindley, 1982) に注目する．

このコメントの中でリンドレーは A. トゥベルスキによる主観確率評価の実験に言及する．トゥベルスキによれば，

(a) 50 箇の赤と 50 箇の緑の玉を含む壺から赤い玉を抽出する，

(b) 割合が不明な赤と緑の玉 100 箇を含む壺から赤い玉を抽出する，

というふたつの事象について，大多数の人はこのふたつが同等に確からしいと考えるが，一旦大きな賞金が赤玉の抽出に対して与えられとなると，多くの人が (a) に賭けることを好むというのである．

この現象を批判し，人はそれぞれの場合において意味のある (relevant) 確率に注目すべきだとリンドレーは言う．上記の例については，(b) の場合，赤玉の割合 θ は不明であるが，$\theta=1/2$ に関して対称な確率密度 $f(\theta)$ を考えることができよう．(a) の場合赤玉の割合は 1/2 と既知である．赤玉が出ることに賭ける時に "意味のある" 確率は (b) の場合 θ の平均値 1/2 で与えられる．これは (a) の場合の確率と一致する．したがって (b) の場合に θ に関する知識が不足しているということは，この賭の場合無意味 (irrelevant) である．

以上の理由から，賭に際して (b) よりも (a) を好むのは間違いだというのである．この議論に

従えば,前述のポッパーの理想的な証拠のパラドックスによる批判を廻避することはできない.

4. 知識の不足の判断への影響

さて上述のリンドレーの主張は合理的なものであろうか.これを検討するために問題の状況をくわしく書き出してみよう.ある人が(a),(b)何れかの賭を選ぶ場面を想定する.選択決定以前にはどちらの賭の方式が採用されるかが定まっていないのだから,まずそれぞれの賭が選ばれる確率 P(a), P(b) が考えられる.それぞれの方式の賭で赤玉が得られる確率が

$$P(赤|a)=P(赤|b)=1/2$$

で与えられることはリンドレーの指摘の通りである.

ここで(赤, a),(赤, b)という事象の確率を考えると

$$P(赤,\ J)=P(赤|J)\ P(J)$$

となる.$J=a$, あるいは b である.P(赤|J)=1/2であるから,(a),(b)のいずれが選ばれるかが分かっていない段階では,(赤, a)と(赤, b)のどちらを選ぶべきかを示すこれらの事象の確率は,全く P(J)の値の如何にかかっている.主観確率の理論は P(J)の値をどう定めるべきかは指示しないから,リンドレーの主張が正しいものとなるためには,P(a)=P(b)=1/2という勝手な制約を前提しなくてはならない.このような恣意的な前提を無意識の中に導入する危険が確率的構造の構築には常につきまとっているのである.

では何故(a)が(b)より好まれるかについてひとつの説明を試みてみよう.P(赤|a)=P(赤|b)=1/2 ということから,賭の結果に対する予測分布としては

$$p(x|1/2)=(1/2)^x(1/2)^{1-x}$$

を採用せざるを得ない.ただし,p(x|1/2) は P(赤|J)=1/2 という知識を前提したときの賭の結果 x(赤なら $x=1$, 緑なら $x=0$)に対する確率である.これに対して "真" の予測分布が $p(x|\theta)=\theta^x(1-\theta)^{1-x}$ であるとき,p(x|1/2)の悪さを負のボルツマンエントロピー,すなわちカルバック・ライブラー情報量

$$I\{p(\cdot|\theta);\ p(\cdot|1/2)\}=\sum_x p(x|\theta)\log\left\{\frac{p(x|\theta)}{p(x|1/2)}\right\}$$

で測ることにする.

方式(a)の場合は $\theta=1/2$ が既知であるから上記 I の値は 0 となる.方式(b)の場合には θ に対する分布(密度) $f(\theta)$ が $\theta=1/2$ に集中しない限り, $I\{p(\cdot|\theta);\ p(\cdot|1/2)\}$ の θ に関する平均は正となる.この結果は,P(赤|a)=P(赤|b)=1/2 という値を実際当面の問題の予測に用いようとする限り,方式(a)の方が方式(b)より良い結果を与えることを意味する.

確率の値は予測的に使うのが本来の姿であり,また K-L 情報量が分布のずれの自然な尺度であることを考えると,ここで得られた結果はトゥベルスキの実験結果が人間の感覚のすぐれた実態を描き出しているものであることを明らかにする.大きな賞金が関係するという現実的な条件の下では,人はより詳しく確率の予測的効用を考え,手もとにある情報がより有効に利用できると,"期待される" 方式を選んでいるのである.

5. 前節の議論に関する補足的な説明

前節ならびに前々節で論じたトゥベルスキーの事例に対するリンドレイのコメントについて,これは特別に取立てゝ議論する程のことではない,多くの人がミニマックス原理に従って

行動したというに過ぎないのではないか，とのレフェリーの指摘があった．(*b*)の場合に $p(\cdot|\theta)$ に関する知識がより不確かであるということは，ミニマックスの立場からすると無意味ではない，というのである．

このような解釈はもちろん可能である．しかし周知のように，ミニマックスの立場は，ベイズ的構造の導入を回避しようとするものである．ベイズ的構造を避ける立場から見るとき，リンドレイの議論が問題とするに値しなくなるというのは当然である．

本稿における筆者の立場は，ベイズ的構造の有効性を認めながら，そこで要求される先験分布の想定にひそむ種々の困難を明らかにしようとするものである．未知の量に対しては先験確率分布を想定するという態度を徹底して維持すれば，実は $p(\cdot|\theta)$ に関する知識の不足が問題となることを示すのが前節の議論の主旨であった．

6. リンドレーのパラドックス

H を単純仮説とし，x を観測値とする．このとき次のような状況が発生し得ることをリンドレーは例によって示した（Lindley, 1957）．

1) H に対する有意性検定で x が有意とみなされる．たとえば5% 有意．

2) 同じ x と H について x が与られたときの H の事後確率が，H の事前確率が極めて小さくとも，95% というような高い値を取る．

データにもとづく情報が 1) では H を捨てるべきだといい，2) では H は極めて信頼すべきものであるという．かくして有意性検定の水準の意味には疑問が残るというのである．

この例を具体的に述べれば次のようなものである．$(x_1, x_2, \cdots, x_n)$ を平均 θ，分散 σ^2（既知）の正規分布からのランダムサンプルとする．$\theta=\theta_0$（帰無仮説に対応する値）に対する事前確率を c とする．残りの確率 $1-c$ は θ_0 を含むある区間 I 上に一様に分布しているものとする．標本平均 $\bar{x}$ が区間 I の中にあるものと考えても大過ない程度に I が取られているものとする．（この表現はややあいまいである．）このとき $\theta=\theta_0$ となる事後確率は次式で与えられる．

$$\bar{c}=c\exp[-n(\bar{x}-\theta_0)^2/(2\sigma^2)]/K,$$

$$K=c\exp[-n(\bar{x}-\theta_0)^2/(2\sigma^2)]+(1-c)\int_I \exp[-n(\bar{x}-\theta)^2/(2\sigma^2)]\ dU(\theta)$$

ただし $dU(\theta)$ は I 上の一様分布を示す．I が十分広くとられていれば上記の積分は近似的に $\sigma\sqrt{2\pi/n}/U(I)$ で与えられる．（この表現もあいまいであるが一応結果だけを認めて先に進むことにする．）$U(I)$ は区間 I の長さである．

さて $\bar{x}$ が α% 有意点であったとしよう．すなわち $\bar{x}=\theta_0+\lambda_\alpha\sigma/\sqrt{n}$ で，λ_α は標準正規分布の両側検定での α% 点とする．この値を上記 $\bar{c}$ に代入すると，$\theta=\theta_0$ の事後確率として

$$\bar{c}=c\exp\left(-\frac{1}{2}\lambda_\alpha^2\right)/\left\{c\exp\left(-\frac{1}{2}\lambda_\alpha^2\right)+(1-c)\sigma\sqrt{2\pi/n}/U(I)\right\}$$

が得られる．ここで $n\to\infty$ とすれば $\bar{c}\to 1$ となることが分る．かくして $\theta=\theta_0$ の事前確率 c が（0 でない限り）どんな値であっても，n を適当に大きくとれば（大きな確率で）

1) $\bar{x}$ は θ_0 から α% 有意

2) $\theta=\theta_0$ の事後確率は $(100-\alpha)$ %

となるようにできる．α が小さい値である場合を考えれば，これが上記のパラドックスになっている．

上記の記述中に括孤で筆者が但し書きを入れたように，このパラドックスの記述は厳密さを欠く点を含んでいるが，その結論は明らかである．一定の事前分布を想定して出発し，サンプ

ルサイズ n を大きくして行く場合を考察すると，ベイズ流に見ればデータによって強く支持される仮説が検定によれば棄却されてしまうような状況が発生する，というのである．これが果してパラドックスであろうか．

筆者の結論は簡単である．本来比較すべきでないふたつのものを，あたかも同一の種類のものであるかのように比較することからこのパラドックスが生じているというのが筆者の見方である．上の検定に登場した $\alpha\%$ はいわゆるP値である．$\bar{x}$ が観測された量である以上，当然P値は $\bar{x}$ に関する確率ではない．そもそも，有意性検定の論理構成の中ではP値を確率と見做すための確率的構造は全く与えられていないのである．P値としての $\alpha\%$ は確率としての意味を持たない．これをベイズ流の事後確率 $(100-\alpha)\%$ と比較しても議論にならないのである．

この結論は多くの概念的な準備を前提して得られるものであるから，そのくわしい説明に入ることは避け，ここではリンドレーの例についてより具体的にパラドックスの内容を見ることにしよう．この例で最も注意すべき点はそこで用いられたベイズ流の確率模型の構成が極めて非ベイズ的であるということである．主観確率の理論の教える所に従えば，ある特定の場面で構築される確率的構造はその時に利用可能なすべての情報の上に組立てられねばならないとされる．したがって，これから得られるデータのサンプルサイズ n がいくらかであるかについての知識は，当然事前分布の構成に利用さるべき情報となる．これは n が全く分らない場合と対比して考えれば明らかである．上述のパラドックスは，θ に関する事前分布は n と無関係であるべきであるという恣意的な想定にもとづいて構成されたのである．

かくしてリンドレーのパラドックスの正体は明らかになる．それは古典的な検定の本質的な欠点を暴くものではなく，i.i.d. シンドロームとも言うべき，一定の条件下での同一実験の繰返しという古典的図式をわれわれが無意識の中に受入れ，それぞれのサンプルについて然るべき事前分布を想定すべしとする主観的確率分布構成上の基本原則を無視することによって生じた化け物である．この化け物はいわゆる尤度原理 (likelihood principle) と同類である（この〝原理〟の批判については Akaike (1977) を参照）．

さてそれではP値の意味は何かということになるが，適当な条件の下でP値がエントロピー，あるいはK-L情報量による仮説の〝真の構造〟からの偏差の測定値と解釈できることは筆者が既に別の場所で論じてある（赤池，1982）．少なくとも上記の例の場合にはこの解釈が可能で，その意味では何等疑念の余地はない．

7. 前節の議論に関する補足的な説明

前節における筆者の主張が理解し難いものであり，この説明ではリンドレイのパラドックスは解明されたことにならないと思う，とのコメントがレフェリーから寄せられた．

レフェリーの指摘する点は次の通りである．問題となる状況はサンプルの数 n に無関係に発生するものであり，これは $n=1$ として区間 I の長さ $U(I)$ を無限に大きくして見れば容易に確認できる．本稿で筆者が〝本来比較すべきでないものを比較することからパラドックスが生じる〟と述べている点には疑問の余地があり，リンドレイは単に検定論の立場からの判断とベイズ流の立場からの判断がまったく逆になることがあるということを主張したに過ぎないのではないか．この種のパラドックスは純粋に仮説検定論の枠組内でも構成可能で，たとえば極端に離れた平均値を持つふたつの正規分布からなる仮説の間の検定を考えれば容易に類似の状況が発生することが認められる．このような状況が発生するからといって，仮説検定論が大きな打撃を受けることはなかろう．以上がレフェリーによるコメントの概要である．

レフェリーの指摘する通り，シェイファーによるリンドレイのパラドックスの提示は上記の

形でなされている（Shafer, 1982）.

筆者がここで最初に問題にしたい点は，有意性検定と仮説検定の相異である．リンドレーの例における平均値 θ_0 の検定に関するP値は，少くとも形式的には何等対立仮説の概念を必要とせずに決定される量である．その意味ではデータ分布の想定とともに一意的にデータから決定される量であり，この限りにおいて全く客観的に定義される．この客観性こそがこの場合の有意性検定に対する人々の信頼感の根源である．リンドレーのパラドックスは，この有意性検定とベイズ流の接近とが同一のデータに関して相反する結論を強く示唆する場合がありうるというものであって，これは対立仮説が与えられなくては議論が展開されない仮説検定論とは全く関係のない話である．この場合のP値はエントロピーの測定値としての解釈が可能であることは前節で言及したとおりであり，その意味する内容はサンプルサイズ n に依存して変化する．

次に指摘したい点は，リンドレーの例におけるサンプルサイズ n の果す本質的な役割りである．これを論ずる前に，ベイズ流の定式化に含まれる主観性を再確認することがまず必要である．ベイズ流の接近に関係する概念的な混乱の多くは，本来主観性がベイズ的構造の特徴であるにもかかわらず，あるひとつの構造をそれがあたかも絶対的な正当性を持つものであるかのように提示することから生じる．リンドレーのパラドックスもその一例である．

リンドレーの例の場合P値が客観性を持つことは前述の通りである．対するベイズ的構造は，区間 I の選び方で定まる θ の事前分布と $\theta=\theta_0$ の事前確率 c とによって決定される．この場合にベイズ流の接近を論じることは，I と c との選び方を論じることである．もし，P値が一定値 $\alpha_0\%$ 以下の場合に仮説 $\theta=\theta_0$ を棄却するという手続きを，対応するベイズ的構造にもとづく事後確率の大小による仮説 $\theta=\theta_0$ の棄却によって実現したければ，それに見合うように I と c とを選べばよいのである．

前節の例で $\theta=\theta_0$, $\sigma=1$ とし，$I=[-10, 10]$, $c=\frac{1}{2}$ とすれば，事後確率による $\theta=\theta_0$ の選択は，$n=1$ の場合 $\alpha_0(\%)$ として $4(\%)$ を取ることにほぼ相当する．ここで一般の n について $I=[-10/\sqrt{n},\ 10/\sqrt{n}\,]$ とすれば，事後確率による $\theta=\theta_0$ の棄却が n に無関係に一定の有意水準 $\alpha_0\%$（$\fallingdotseq 4\%$）での仮説の棄却に対応するようになることは明らかである．

この結果は，一定の有意水準での仮説の棄却という手続きが，n とともに事前分布の構造を変化させるようなベイズ流の接近で実現できることを示す．（そのような決定法が良いものであるか否かは別の問題である．）n に関せず θ の事前分布は一定のものと想定すべきである，という恣意的な制約を導入することによってはじめてリンドレーのパラドックスが構成されるものであることがこれで明らかになったであろう．

それでもなおレフェリーあるいはシェイファーの提示する形でのパラドックスは存在するではないか，という主張がなされるかもしれない．シェイファーはこの形のパラドックスに説得力を持たせるために，θ の事前分布として過去の観測値にもとづく頻度分布が与えられる場合を想定する．これならば大方の統計家がベイズ流の分析を受入れるであろうというのである．ここでベイズ的構造構成に際して主観の果すべき役割りが無視され，見掛け上の客観性が持定のベイズ的構造に与えられる．過去の観測データを現在のベイズ的構造構成にどう利用するかは，決して一意的ではない．過去の情報をどのように利用するかを定めるのは分析者自身である．たとえ θ の事前分布が与えられたとしても，$\theta=\theta_0$ の事前確率 c をどう定めるかが問題として残っている．

ここで更に c も与えられたとしよう．かくしてベイズ的構造が完全に規定され，そこでリンドレイのパラドックスに対応する状況が発生した場合を想定しよう．この場合にあっても，P値

の意味の客観性は失われない．$\theta=\theta_0$ からの偏差が著しく有意であるとP値にもとづいて判断することが可能ならば，これに反する結論を示唆するベイズ的構造は，それ程の有意性を示す偏差をも検出する能力を持たないような事前情報にもとづく構造であったと判断されるだけのことである．

上記の議論に説得性が無いと感じられるならば，以下の議論が問題点を更に明らかにするであろう．まずリンドレーのパラドックスを尤もらしく見せている条件のひとつとして，"H(仮説)の事前確率(c)が極めて小さくても" という記述があることに注意しよう．この条件下で，パラドックスを生成するために n を大にするという場面が想定されたのである．これに対して，n を固定したまま事前確率 c を極めて小にする場合に発生するもうひとつのパラドックスにここでは注目しよう．

前節の議論において，$\theta=\theta_0$ の事後確率 $\bar{c}$ に注目し，$\bar{x}=\theta_0$ という場合を考える．これは $\bar{x}$ が $\theta=\theta_0$ で与えられる仮説を支える証拠として最も強力な場合と考えられよう．このとき

$$\bar{c}=c/K$$

$$K=c+(1-c)\int_I \exp\left[-n(\theta-\theta_0)^2/(2\sigma^2)\right]dU(\theta)$$

である．I が θ_0 を中心とする $\sigma/\sqrt{n}$ に比し十分大きな長さを持つ区間であれば，$K=c+(1-c)\sqrt{2\pi/n}\sigma/U(I)$ とみなすことができることから，c を十分小さくとれば $K\fallingdotseq\sqrt{2\pi/n}\sigma/U(I)\gg c$，したがって

$$\bar{c}\ll 1$$

が成立することが分かる．

この結果は，たとえ最強の証拠が得られた場合でも $\theta=\theta_0$ という仮説は容認されないことを示す．シェイファーが例として論じている法廷での証拠の解釈の場合について考えると，上記のベイズ的構造を認める限り，$\theta=\theta_0$ が棄却されることは証拠を見る迄もなく決っている．このような構造が法廷での証拠の議論の根拠として採用される筈がないことは明らかであろう．

この例は前述の場合のほぼ逆の状況を与えるものであり，ベイズ的構造はそれが想定されたというだけでは何等の権威も無いものであることを一段と明らかに示してくれる．パラドックスが発生する場合にはベイズ的構造そのものの妥当性も疑われなくてはならないのである．

以上の議論で，n の如何にかかわらずパラドックスが解消され得るものであることは明らかにされたと思う．しかしリンドレーの本来のパラドックスは n の変化に対する事前分布の固定化にもとづいて発生したものであり，この構造を採用しない場合にはもとのパラドックスの最も教育的な側面が失われるといえよう．特にこれを次節の議論と結ぶ環は消失してしまうのである．

8. BICによるAIC批判の解明

前節に登場した確率解釈上の怪物は，単に知的な興味をそそるという程度のものではなく，実は統計理論の研究そのものに明瞭な影を落しているのである．その実例として，いわゆるBICによるAICの批判について論じてみよう．

情報量規準AICについては，これを最小にするモデルを採用するという最小AIC法が多くの適用例を見出しているが，この最小AIC法について，その "不一致性" がしばしば議論の対象とされてきた．ここでいう不一致性とは，たとえば多項式のあてはめに最小AIC法を適用した時に得られる多項式の次数の場合，データ数を如何に大きくとっても真の次数を超える確率が0に近づかないというものである．この現象は，AICに先立つFPE (final prediction

error）の導入に際して筆者が指摘したものである（Akaike, 1970).

初めに明らかにしておかなくてはならないことは，次数に関する不一致性はモデルそのものに関する不一致性を意味しないということである．多項式の次数の上限が有限に定められていれば，データ数の増大とともに最高次数のモデルでも正しい値に収れんして行くことは明らかである．このことから，実用上は不一致性よりも推定の有効性の方が遙かに重要な問題であることがわかる．

さて定義によればAICは

$$\text{AIC}=(-2)(\text{最大対数尤度})+2(\text{パラメータ数})$$

によって与えられる．G. シュワルツ（1978）は未知パラメータについて連続な事前分布を想定し，データが一定の分布からの独立な N 回の観測値によって構成される場合についてパラメータ数の異るモデルの事後確率を考察した．その結果をAICに対応する形で表現すれば

$$\text{BIC}=(-2)(\text{最大対数尤度})+\log N(\text{パラメータ数})$$

となる．$(-1/2)$ BICがモデルの対数尤度の漸近的な表現を与える．この結果はH. ジェフレイス（1948）の K 統計量に対応するものであり，筆者自身も特定の条件下で類似の結果を得ているAkaike (1977).

シュワルツは彼の論文で，上記の結果は最小AIC法の最適性に関する如何なる証明をも否定するものであろうと述べている．事前分布がどんなものであっても適当に滑らかにひろがってさえいれば，それによる事後確率の対数の (-2) 倍は漸近的にBICの形になる．したがって，これに一致しない形のAICはモデルの選択における最適性を持ち得ない，というのである．この議論は一見説得力を持つ．事実この結果からAICは駄目であると言う人や，数値実験によって最小BIC法と最小AIC法とを比較し，BICによる選択がすぐれている等と言う人が各所に現われるようになった．果してシュワルツの論理は正しいものであろうか．

前節迄の議論の本質が理解されていれば，シュワルツの議論の限界はたちまち明らかとなる．一定事前分布とi.i.d. 型のモデル構成が上記の結論に導いているのである．AICの定義には N は登場しない．より詳しく解釈すれば $N=1$ を想定しているものといえる．AICの導入に際して漸近的な議論を用いたのは，最大対数尤度の差の平均値の評価を正当化するためである．この評価が正当化される場合には，i.i.d. 型であろうとなかろうと同じように取扱える．多項式のあてはめ，因子分析等の場合等がその例である．$N=1$ とすればAICの2の代りに $\log N$ あるいは $\log\log N$ 等を考えることの無意味さは明らかである．

問題を更にはっきりさせるために，最小BIC法が極めて恣意的なものであることを示すことにしよう．L 次元円型正規分布を考える．i 番目の変量の平均を m_i とし，この分布から N 個の独立な観測値 $(x_{1n}, x_{2n}, \cdots, x_{Ln})$ $(n=1, 2, \cdots, N)$ が得られたとする．データ分布の尤度が

$$\left(\frac{1}{2\pi}\right)^{LN/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{L}\sum_{n=1}^{N}(x_{in}-m_i)^2\right)$$

で与えられるとする．今 m_i について

$$\left(\frac{1}{2\pi}\right)^{(L-j)/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{j}m_i^2\right)\prod_{k=j+1}^{L}\delta(m_k)$$

のような事前分布密度を想定し，これで与えられるモデルを j 番目のモデル $(j=0, 1, \cdots, L)$ とする．ただし $\delta(m_k)$ は $m_k=0$ に集中した分布を示す．

i 番目のモデルの尤度は上記2式の積の m_k $(k=1, 2, \cdots, L)$ に関する積分によって与えられ，事後確率はこれと事前確率 α_j との積に比例する．j 番目のベイズモデルの対数尤度を求めこれを（-2）倍すると，すべてのモデルに共通な項を無視すれば

$$N\sum_{k=j+1}^{L}\bar{x}_k^2+j\log N+\frac{N}{N+1}\sum_{i=1}^{j}\bar{x}_i^2+j\log\left(1+\frac{1}{N}\right)$$

となる．BICはこの式のはじめの2項の和で与えられる．それはこの第1項が，$m_1,\cdots,m_j$ を最尤法で推定し残りを $m_{j+1}=\cdots=m_L=0$ とおく場合の最大対数尤度に一致するからである．したがって，上記の値の代りにBICをモデルの比較に使うということは，後の2項の和を無視することに相当する．これは丁度それに見合うように

$$2\log\alpha_j=\frac{N}{N+1}\sum_{i=1}^{j}\bar{x}_i^2+j\log\left(1+\frac{1}{N}\right)+\text{const.}$$

で与えられる事前確率 α_j を j 番目のモデルに対して与えることに相当する．

この結果はデータから与えられる情報に依存する事前確率を想定することとなり，そのような選択の最適性は明らかではない．N が大となる場合を考えれば，BICによる比較は

$$\log\alpha_j=\frac{1}{2}\sum_{i=1}^{j}m_i^2+\text{const.}$$

で与えられる事前確率にもとづく事後確率による比較を近似していることがわかる．N が大となればこの項の影響は相対的に無視されるようになるというのが漸近理論の立場であるが，現実にBICを適用するのは常に有限の N についてであり，ひとつのサンプルしたがって N が固定された状態でBICを適用することは，結局上記の事前確率を想定してベイズの方法を適用することに一致する．

以上の結果を整理すると，BICを利用してモデル選択をすることは，ある特定の事前確率を各モデルに想定した場合に対してのみ漸近的に最適性が保証されるものであり，しかもこの事前確率は未知のパラメータに依存して定まるようなものであるということになる．このような方法と，AICによるモデルの選択とが一致しないからといって，AICに関するすべての最適性の議論が否定されるものではありえないことは明らかであろう．事実上記の問題についての最小AIC法のMin Max性の証明が筆者によって与えられている(Akaike, 1978)．かくしてBICによるAICの批判と呼ばれるものは，固定した事前確率の想定と古典的なi.i.d.図式との結合にもとづく産物であることが明らかとなった．

おわりに

確率の概念が人間の感覚を通じて形を得たものであることは間違いない．したがって新しい確率的構造を現実の問題に対して構築しようとする時には，従来の固定観念を脱却して問題の本質を見抜く努力が要求される．本稿で取上げた事例は，確率的な構造を構成する際にわれわれが如何に無意識の中に勝手な論理の枠組みを前提して議論を展開するものであるかを示している．この傾向が，確率を客観的に捉えようとする人と主観的に捉えようとする人とを区別することなく現われるものであることをもこれらの事例は示している．

さて確率的構造の設定は結局全く主観的・個人的なものに止まるのであろうか．答はもちろん否である．観測データとの対比を通じて客観性獲得への努力は無限に続けられる．その一場面において提案される種々の確率的構造については，その対数尤度によって客観的評価が与えられる．

有効な統計モデルの開発の成否が統計学の将来の発展を左右するものであることの認識が深まりつつある現在，確率の柔軟な解釈が有効なモデル構成への重要な鍵を与えるものであると考え，確率の解釈に伴う困難をいくつかのパラドックスと呼ばれるものを例として論じたのが本稿である．

われわれの確率の解釈にひそむあいまいさに鋭く問いかけるこれらのパラドックスを書きし

るし興味ある議論を展開してくれた人々に対する感謝を記して本稿を終りたい.

謝　　辞

第一稿を精読し種々示唆に富むコメントを与えられたレフェリーに深く感謝する. 本研究は科学研究費補助金（課題番号 58450058）による研究の一部である.

参 考 文 献

Agassi, J. (1975). Subjectivism : from infantile to chronic illness, *Synthese,* **30**, 3-14.

Akaike, H. (1970). Statistical predictor identification, *Ann. Inst. Statist. Math.*, **22**, 203-217.

Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics* (ed. P.R. Krishnaiah), North-Holland, Amsterdam, 27-41.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, **30, A,** 9-14.

赤池弘次（1982）統計とエントロピー, 数学セミナー, **21** 巻 12 号, 2-12.

Good, I.J. (1975). Comments on Joseph Agassi, *Synthese,* **30**, 31.

Jeffreys, H. (1948). *Theory of Probability,* 2nd edition, Oxford University Press.

Lindley, D.V. (1957). A statistical paradox, *Biometrika,* **44**, 187-192.

Lindley, D.V. (1982). Comment on "Lindley's paradox" by Glenn Shafer, *J. Amer. Statist. Ass.*, **77**, 334-336.

Peirce, C.S. (1878). On the doctorine of chances, with later reflections, *Philosophical Writings of Peirce*, (ed. J. Buchler), Dover, New York, 1955, 157-173.

Peirce, C.S. (1878). The probability of induction, *Philosophical Writings of Peirce*, (ed. J. Buchler,), Dover, New York, 1955, 174-189.

Popper, K.R. (1968). *The Logic of Scientific Discovery,* Second Harper Torchbook Edition, Harper and Row, New York.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics,* **6**, 461-464.

Shafer, G. (1982). Lindley's paradox, *J. Amer. Statist. Ass.*, **77**, 325-334.

On the Difficulty in the Interpretation of Probabilities

Hirotugu Akaike
(The Institute of Statistical Mathematics)

The danger of introducing arbitrary constraints in developing a probabilistic argument is demonstrated with examples. These include the discussion of the pradox of ideal evidence by K.R. Popper, the comment by D.V. Lindley on the inconsistent human behavior in the assessment of uncertainty reported by A. Tversky, the so-called Lindley's paradox of the conventional significance test and Bayesian evaluation of posterior probability, and the disproof of the optimality of AIC by G. Schwarz.

It is argued that the first paradox can be solved by accepting the uncertainty in the choice of the probability of an event, although this is unacceptable to strict Bayesians such as B. de Finetti or L.J. Savage. With the aid of the concept of Boltzmann entropy, or the Kullback-Leibler information, an interpretation of Tversky's finding is developed to show the rationality of the seemingly inconsistent behavior in the assessment of uncertainty. Lindley's paradox and Schwarz's disproof of the optimality of AIC are both explained to be due to the adoption of the conventional i.i.d. scheme and a fixed arbitrary prior distribution.

The paper concludes with the appreciation of efforts of the authors of these thought provoking examples that have contributed significantly to the advancement of the interpretation of probabilities.