統計数理 (1999)
第 47 巻 第 1 号 81-90

特集「統計基礎理論」
［原著論文］

事前分布が曖昧な場合の Bayes 検定の限界

統計数理研究所　柳　本　武　美

（受付 1998 年 8 月 14 日；改訂 1998 年 12 月 25 日）

要　旨

統計的検定に対する批判の一つに，Lindley のパラドックスがある．検定では有意水準 α を決めて，帰無仮説の下で確率 α でのみその仮説を棄却する．しかし標本サイズ n が大きいときには一致性がある方が良いとする批判である．モデル選択でも同様の議論がある．

本稿では，この批判はデータの情報量が大きい一方，事前情報の情報量が相対的に小さい場合の扱い方であることを明確にする．例えば母数の推定量の分散に比べて事前分布の分散が大きい場合である．この場合を事前分布が曖昧であると呼ぶ．事前分布が曖昧な場合には直観的に考えても事前分布を利用するメリットはない．更にこの場合が余り現実的でないこと，標本サイズに対する配慮が乏しいこと，更に事後分布の解釈が困難であることを示した．結論として科学的推論のためには事前分布が曖昧な場合の Bayes 検定が有益でないことを主張した．

キーワード：事後分布，検定の一致性，モデル選択，信頼区間，統計的検定．

1. 序

統計的検定は今日の科学的推論の中で重要な位置を占めている．その推論の枠組みは帰無仮説 H_0 と対立仮説 H_1 について極めて特殊な論理構造をもつ．近年では帰無仮説を棄却するかしないかよりも，p 値を与える，あるいは信頼区間を与える方が良いなどの意見があるが，実質的な違いはない．統計的検定は客観的な推論を企図とした，データのみに依存した推論方式を与える．

一方事前情報を積極的に利用する Bayes 推論が注目されている．データを収集する際に全く事前情報がないとは考え難い．2 つの仮説 H_0 と H_1 自身が何らかの事前情報に基づいていると考えられる．また標本サイズも，後に議論するように，事前情報に強く依存しているのが普通である．データが得られた後の情報は，事前情報及びデータのもつ情報からなるから Bayes 推論の考えは極く常識的である．しかし Bayes 推論には事前情報の信頼性という根本的な問題がある．また事後分布とか Bayes 因子のような形式的推論方式が妥当かという問題がある．その結果 Bayes 推論についての理論的研究が著しい反面，実用面への適用は限られているのが実状である．

Lindley (1957) は標準的な統計的検定と Bayes 推論に基づく検定が相矛盾する場合があることを示した．この例は後に Lindley のパラドックスと呼ばれ，多くの研究者の関心を呼んだ．Schwarz (1978) はモデルの選択に適用した．また Shafer (1982) は少し異なった立場から Lindley の主張を支持した．その後統計的検定に対応する Bayes 手法は Bayes 因子をクライテリ

アとする理論整備がなされ（例えばKass and Raftery (1995)），近年にもKass and Wasserman (1995) はBayes因子の利用の立場からLindleyのパラドックスについて言及している．これらの著者は統計的検定に批判的であるのに対して，モデル選択の立場から赤池(1984)はLindleyとSchwarzの論法の原理的な欠陥を論じ，事前分布が標本サイズに無関係に仮定していることを指摘している．

本稿の目的はLindleyのパラドックスの設定がデータの情報量が大きい一方で，事前情報が相対的に小さい場合であり，実際の適用上には非現実的で特殊な場合であることを主張することにある．本稿では実際への適用を論じることになるので，筆者のこれまでの経験と関心から議論の制約を受けるのは止むを得ない．筆者は安全性の評価とか薬効評価のような生物統計学の分野に経験と関心がある．しかし品質管理，信頼性，社会調査，更には計量経済モデル，心理学分野における統計モデルとデータ解析の知識もある．そうした意味を含め本稿では標本サイズnが小さいとは$n=5\sim 20$，例えば10，ほどほどとは$n=50\sim 200$，例えば100，非常に大きいとは例えば$n=10^6$とする．

2. Lindleyのパラドックス

確率モデル$p(x;\mu)$について帰無仮説をH_0: $\mu=0$，対立仮説をH_1: $\mu\neq 0$とする．以降では計算の簡単のために$p(x;\mu)$は平均μ，分散σ^2の正規分布とし，σ^2は既知とする．この制約は全く本質的でない．標本サイズnの標本を$x_1,\dots,x_n$とし，その標本平均を$\bar{x}$とする．これはμの十分統計量である．Lindley (1957) は，事前分布の“密度関数”$\varphi_L(\mu)$を

$$\varphi_L(\mu)=c\delta(0)+(1-c)\frac{1}{\sqrt{2\pi}\tau}e^{-\frac{1}{2\tau^2}\mu^2},\qquad 1>c>0 \tag{2.1}$$

但し$\delta(\mu)$はDirac関数，とおいた．言い換えると，事前分布は原点0で退化した分布と平均0，分散τ^2の正規分布の重みc，$1-c$の混合分布である．この事後分布の密度は，

$$\frac{\varphi_L(\mu)\dfrac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}}{\displaystyle\int\varphi_L(\mu)\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}d\mu} \tag{2.2}$$

で定義される．この分布は，$\mu=0$のときマスを持ち

$$P_0(0)=\frac{\dfrac{c\sqrt{n}}{\sigma}e^{-\frac{n}{2\sigma^2}\bar{x}^2}}{\dfrac{c\sqrt{n}}{\sigma}e^{-\frac{n}{2\sigma^2}\bar{x}^2}+\dfrac{1-c}{\sqrt{\tau^2+\sigma^2/n}}\exp\left\{-\dfrac{1}{2(\tau^2+\sigma^2/n)}\bar{x}^2\right\}}\delta(0) \tag{2.3}$$

となる．ここで

$$\mathrm{BF}=\frac{\sqrt{n\tau^2+\sigma^2}}{\sigma}\exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}-\frac{1}{(\tau^2+\sigma^2/n)}\right)\bar{x}^2\right\} \tag{2.4}$$

とおくと，$P_0(0)=c\mathrm{BF}/(c\mathrm{BF}+(1-c))$と書ける．もし$\mu$の真の値$\mu_T$が$\mu_T=0$であれば，任意の値$M>0$に対して$\lim_{n\to\infty}\Pr(\mathrm{BF}>M)=1$となる．一方$\mu_T\neq 0$であれば，極限値は0になる．この事実は$n$が十分大きいとき，帰無仮説$H_0$が正しいとき，$H_0$の事後確率が1に近く，$H_0$が正しくないとき0に近くなると解釈される．これがLindleyの主張の骨子である．

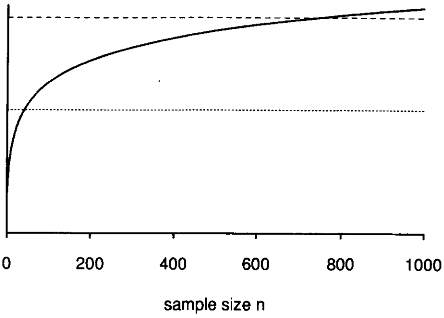

図 1. 棄却値 $\{(n+1)/n\}\log(n+1)$ の変化：$\sigma^2=\tau^2$，$c=0$，BF $=1$ の場合．図中横線は χ^2-分布の上側 5%，1%値．

Schwarz はこの事実をモデル選択として一致であると呼んでいる．一方統計的検定では

$$(2.5)\qquad n\bar{x}^2/\sigma^2 > \chi^2_{1,1-\alpha}$$

のときに H_0 を棄却する．但し α は水準，$\chi^2_{1,1-\alpha}$ は自由度 1 の χ^2-分布の $(1-\alpha)$ 点である．逆に不等式 (2.5) が成立しないときには H_0 を棄却しない．

Kass and Wasserman (1995) の主張は，その解析的な展開を除くと，次のようにまとめられる．まず事前分布は 2 通り考えて，$\varphi_0(\mu)$ は (2.1) で $c=1$ とする．これは H_0 に対応する．次に $\varphi_1(\mu)$ は (2.1) で $c=0$，$\tau^2=\sigma^2$ とする．後者の制約は「単位情報事前分布」と呼ばれ，事前分布の情報量と 1 観察値当りの情報が等しいと解釈される．そして検定を Bayes 因子 BF $=\int p(\bar{x};\mu)\,\varphi_0(\mu)\,d\mu/\int p(\bar{x};\mu)\,\varphi_1(\mu)\,d\mu$ が 1 より大きいか否かで判断する．この規準では n が大きいことを仮定していない．Bayes 因子の対数は (2.4) で $\tau^2=\sigma^2$ とおいて得られるから

$$(2.6)\qquad S=\log \mathrm{BF}=\frac{1}{2}\log(n+1)-\frac{1}{2}\frac{n}{(n+1)}\frac{n\bar{x}^2}{\sigma^2}.$$

従って $S<0$ は $n\bar{x}^2/\sigma^2>\{(n+1)/n\}\log(n+1)$．この不等式を (2.4) と比較すると右辺が n に依存し，n が大きくなると $\log n$ とほぼ等しくなる．従って n が大きくなると，$\mu_T=0$ のとき $\lim\Pr(S<0)=0$，$\mu_T\neq 0$ のとき極限値は 1 になる．棄却値 $\{(n+1)/n\}\log(n+1)$ を図 1 に示した．この値は n の関数として $n\to\infty$ のとき ∞ になり，$o(n)$ である．この 2 つの性質から Bayes 検定の一致性が導かれる．しかし図 1 に示されるようにほどほどの n では帰無仮説の下での棄却確率は 0.05 と 0.01 の間になる．

以降での論点をはっきりさせるために次の事実を指摘しておく．標本サイズ n に関する漸近理論の形をとっているが，本質は $\varphi_1(\mu)$ の分散と $\bar{x}$ の分散との比 $r=n\tau^2/\sigma^2$ にある．標本サイズは $\bar{x}$ の分散を小さくする役割を果たしているだけである．実際 $\tilde{\sigma}^2=\sigma^2/n$ とおくと，$2S=\log(r+1)-\{r/(r+1)\}\,\bar{x}^2/\tilde{\sigma}^2$ となり見かけ上 n は消える．次に 2 つの棄却値 $c_t=\tilde{\sigma}^2\chi^2_{1,1-\alpha}$, $c_b=\tilde{\sigma}^2\{(r+1)\log(r+1)\}/r$ が共に τ^2 に比べて小さくなることである．論点の相違は $c_t<\bar{x}^2<c_b$ となる標本に対する検定方式がある．ところが適当なK，例えば 5，に対しては $\Pr(\mu^2>Kc_b|\varphi(\mu))$ も $\Pr(\mu^2>Kc_t|\varphi(\mu))$ も共に小さくなる．事前分布が，誤りとは言えないまでも，適切とは言えない．この場合標本の情報に比べて事前情報が相対的に曖昧，あるいは量が小さい，である．そこで以降では，$\tau^2/\tilde{\sigma}^2$ が大きい場合に，曖昧な事前分布と呼ぶ．

3. Bayes 推論と事前分布

標準的な統計的検定と Bayes 検定の間には概念的にも大きな隔たりがある．これを比較するためには比較が可能なように舞台を制約する必要がある．そのために本節では先ず現実的と考えられる状況についてその大枠を論じる．

測定誤差の分散 σ^2 を改善することは実際には難しい．前節にも述べたように標本サイズを大きくすることによって $\bar{x}$ の分散を σ^2/n にできる．一方真の平均 μ_T は未知であるが，推論にとって意義のある差 $|\mu_T|$ はある程度の値 Δ に特定できることが多い．薬の効果の同等性，生物学的同等性の分野でも相対差 0.1 等が論じられている．Spiegelhalter et al. (1994) も同様の議論をしている．真の値が 0 でなくても Δ より（ある程度）小さければ $\mu_T=0$ と見なしてしまう．実際特定の値 $\mu=0$ が推論で大きな役割を占めることはこの為である．比 Δ/σ が小さいときには $\Delta/\tilde{\sigma}$ が適当な値になるように n を大きくすることが望まれる．そうでないと統計的検定でいう検出力が小さい．いずれにしても Δ と σ^2 については統計学的には余り問題はない．

事前分布については多くの論ずべき点がある．まず事前分布として $N(m,\tau^2)$ を考えよう．事前分布としては解析者の主観あるいは信念に基づくとする考え，何らかの意味で事前の観察に基づくとする考えもある．もし真実の値 μ_T が存在すると仮定すると事前分布は大まかに μ_T に関連していると期待されよう．しかし σ^2 と Δ とは概念的に異なるだけに直接的な関連はなさそうである．そうした意味では Kass and Wasserman (1995) にある単位情報事前分布は通常いわれる事前分布というより，モデル依存型事前分布と呼べる．その結果真の平均 μ_T との関連が弱くなる．

さて Bayes 推論では事前情報と観察データの 2 つの情報から推論する．不適確な事前分布，即ち $|m-\mu_T|/\tau$ が大きい，を用いるのは良くない．もし $|m-\mu_T|/\tilde{\sigma}$ が大きければ，事前分布が不適確であることがデータから検出されるので実害はない．逆に事前情報が正確 ($|m-\mu_T|/\tau$ が小さい) で，事前情報の量がデータの情報量に比べて小さくない ($1/r=\tilde{\sigma}^2/\tau^2$ が小さくない) 場合には Bayes 推論が良くなるのは常識的に考えても当然である．我々の問題は $\tilde{\sigma}^2/\tau^2$ が小さい場合で更に $m=0$ で $|\mu_T|/\tau$ も小さい場合である．詳細な議論は後に譲るとしても，直感的に考えても Bayes 推論に有利な条件ではない．実際 $\tilde{\sigma}^2/\tau^2$ が小さいことはデータの相対的情報量が大きいことを示している．情報量が小さい事前情報を敢えて加える必要はなさそうである．しかも $|\mu_T|/\tau$ が小さいことは $|\mu_T|$ を $|\mu|$ の事前分布の実現値と見なしたとき，$|\mu_T|$ が事前分布の裾（つまり端）の部分に当たることを意味している．即ち事前分布の仮定が必ずしも適確でない．以上の考察から標準的な推論と Bayes 推論を比較すると表 1 を得る．

表 1. 事前情報の相対情報量と正確さから見た標準的な推論と Bayes 推論の良さの比較.

相対情報量	正確さ		
$\tilde{\sigma}^2/\tau^2$	正確	曖昧	不適確*
小さい	標準, Bayes	(標準) 本稿での問題	標準
小さくない	Bayes	標準, Bayes	良い推論は不能

* 事前情報に大きな偏りがある場合

Lindley (1957) が仮定した (2.1) の混合型事前分布 $\varphi(\mu)$ は別の問題が生じる．この分布の Shannon のエントロピーは $-\infty$ とみなされる．事前分布の密度関数 $\varphi_0(\mu)$ は帰無仮説 H_0 が本当に正しい確率は 0 であるとする．これは科学的な推論における仮説としては極く自然である．通常 H_0 は厳密には誤りと考える．更には併合した仮説 $H_0{}^{\vee}H_1$ でさえも厳密には誤りと考える．ところが事前分布 $\varphi_L(\mu)$ では $\mu=0$ が 0 でない他の値よりも特に正しいと見なす確率を考える．このような事前分布は有限の観察からは誘導されない．むしろより本質的な事実からの誘導あるいは類推，更には場合によってひらめき，によって得られる．だから事前の観測に基づくような連続型の事前分布とは異なる．両者には概念的に定量的以上の違いがある．観測値によって H_0 が正しい事後確率か計算できることは不自然でさえある．解析的にみても $\varphi_L(\mu)$ は Dirac 関数と確率密度の重みつき和になっていて $\mu=0$ での形式的な事後確率 $P_0(0)$ が信頼おけるのか分からない．実際 Bayesian による研究でも Bayes 因子を用いる方向である．勿論 $\varphi_L(\mu)$ は新しい仮説を提示した，あるいは支持する研究者がその仮説の正しさを検証しようとするときの気分に適合している．しかしそうした主観的な気分と本当に，有限のデータから，仮説の正しさの確率が増加するかという問題は別である．

さてこの事前分布 $\varphi_L(\mu)$ についての議論は一旦おいて議論を進める．

4. 曖昧な事前分布

我々が問題としている状況は $r=\tau^2/\tilde{\sigma}^2$ が大きくて，$|\mu_T|/\tau$ が小さいことを指摘した．前節では直感的あるいは概念的に考えて Bayes 推論にとって有利な状況でないことを述べた．本節でもう少し深く考察する．

曖昧な事前分布は $\tau^2/\tilde{\sigma}^2$ が一定で n が大きいという仮定から生じている．単位事前情報では $\tau^2/\sigma^2=1$ と置く．この設定は主観的な事前情報というよりもモデル依存である．ところが標本を表現するモデルには標本サイズが含まれる．従ってもしモデルに依存する事前分布を設定するならば，標本サイズ n も考慮に入れて $r=\tau^2/\tilde{\sigma}^2=1$ と置く方が説得力がある．この主張についてデータの入手と既存の知識との関連で更に議論する．

最初にある仮説 H_0 を検定するときに，既に適当なデータが存在する場合，あるいは解析者とは独立にデータが得られる場合を考える．現実には非常に大きな標本サイズのデータが存在することは，少なくとも筆者の経験では，考え難いがその問題は不問にする．しかし，この場合標本 $x_1,\ldots,x_n$ は一体であって，むしろ n 次元空間のサイズ 1 の標本 $\boldsymbol{x}=(x_1,\ldots,x_n)$ が得られたと見た方が自然である．実際モデルは $x_1,\ldots,x_n$ を同時に表現するものであって，1 つの成分 x_i のみを表現するだけでは不十分である．各成分が独立に同一分布に従うことを仮定することは，モデルを表現する n 次元空間上の確率分布 $p_n(\boldsymbol{x})$ が 1 次元確率分布 $p_1(x)$ の積，$\prod p_1(x_i)$ で表されることである．独立同一性の仮定は自明ではなく，得られたデータの収集方法を吟味した後に採用される仮定である．データの収集の実際を勘案すれば，標本の 1 成分 x_i の情報量を規準にする発想は受け入れ難い．実際データの収集の妥当性とか異常値についての検討等を行った筈である．我々が 1 つの大きさ n の標本と呼ぶのはこの為である．

次にデータの解析者がデータの収集に係わる場合を考える．研究の当初では事前分布と分散のみが分かっていると想定される．そうすると $\tau^2/\sigma^2=1$ という仮定が当初の仮定としては現実的となるかも知れない．しかし次節で論じるように標本サイズ n は当然事前分布に依存するのであって，無関係に選ばれる訳ではない．

今度は上の 2 つの場合のように事前分布がモデルに依存するのではなくて，何らかの予めの情報に基づいている場合を考える．ここでは再び解析者とは独立に大きさ n の標本が得られる場合を想定する．非常に大きな n の標本が得られる場合であるから，当該標本が得られる前か

ら少なくとも関連したデータは得られ易いと思われる．そうした筆者の常識からは $r=\tau^2/\tilde{\sigma}^2$ が非常に大きくなることは考え難く，$r<1000$ のようなほどほどの値になると思われる．

しかし現実には筆者の常識が適用される状況があるかも知れない．事前の情報の多寡に拘わらず τ^2 が小さくならない場合である．平均の絶対値が大きいであろうという確信がある場合である．しかしながら我々の問題は $|\mu_T|/\tau$ が小さい場合の検定問題である．前節でも述べたように，事前分布が現実と大きくずれた時でさえ形式的な Bayes 推論が良い推論を導くとは考え難い．形式的な Bayes 推論が，現実と大きくずれた時にも，有効に働くことを確かめる研究が必要である．

5. データ収集と標本サイズ

新たにデータを収集する計画に参画した場合の標本サイズの設計を考える．我々は τ^2 と σ^2 が同程度で $|\mu_T|/\tau$ が非常に小さい場合を議論する必要がある．実際には μ_T は分からないから，2節で述べた実際に意味のある最小の値 Δ が計画にとって重要になる．もし Δ/τ が小さくなければ計画は比較的容易である．しかし Δ/τ が非常に小さいと計画は極めて複雑になる．何故なら計画担当者は事前分布の分散と観察の誤差の分散が同程度に離れていると想定している．しかし $|\mu_T|\doteqdot\Delta$ でも意味のある推論をしたいという極めて複雑な要請の下でデータを収集する．そこで場合によっては非常に大きな標本サイズを設定する．一方では初めから大きな標本サイズを設定して $|\mu_T|/\sigma$ が小さくなければ資源と労力の無駄使いになる．議論を進める前に標準的な理論における標本サイズの設定と大きな標本サイズの実際的な意味を整理する．

標準的な標本サイズの設定の理論では検出力 $1-\beta$ $(1>\beta>\alpha)$ を設定する．そして $|\mu_T|=\Delta$ のときの検出力を β にする，即ち

$$\Pr\left(n\bar{x}^2/\sigma^2>\chi^2_{1,1-\alpha}\,\middle|\,|\mu_T|=\Delta\right)=\beta$$

となるように選ぶ．これを書き改めると，

$$\chi^2_{1,\beta}(\sqrt{n}\Delta/\sigma)=\chi^2_{1,1-\alpha} \tag{5.1}$$

になる．但し $\chi^2_{1,1-\beta}(d)$ は自由度1，非心度 d の $(1-\beta)$ 点である．標本サイズの平方根と Δ/σ の積が一定になる．従って α と β を適当に固定すると，非心度 $\sqrt{n}\Delta/\sigma$ が計算される．例えば Δ/σ を1から 10^{-3} に変えると，必要標本サイズは 10^6 倍になる．

標本サイズが大きいことについての緊張感が Lindley (1957) らの論調には希薄に見える．データの収集には時間と労力を必要とする．しかも独立・同一性を保つことは至難の業で，特に標本サイズが大きいときは不可能と言って良い．時間が経過するとトレンドが生じる．また多くの施設と人間の参加はブロック効果が生じる．結果として生じる非独立性，非同一性を無視することは許容できることではない．しかし，理論統計学者は大標本理論の精密な研究を行ってきた．実際，少なからぬ割合で統計理論の中心は大標本理論と信じている．その結果大標本の仮定が現実的であると思いこむ傾向があることを指摘しておく必要がある．ここで議論している状況が極めて安易に設定されているのは，理論統計学者の標本サイズについての認識が現実離れしていることに由来すると思われる．赤池 (1984) も「i.i.d.シンドローム」と呼んで同じ主張をしている．

データの収集には労力と時間がかかることを改めて再確認しておく．そうすると，事前分布の分散 τ^2 が σ^2 に近い場合を仮定しているのだから，次の標本サイズの設計が現実的である．データ収集では初めから非常に大きい標本サイズを設定しない．ほどほどの標本サイズで結論が導かれる可能性が大きいからである．事前分布が妥当でない場合をも考慮して，逐次的なデー

タ収集の計画を立てるに違いない．最初のデータが集まったときに $|\mu_T|/\tau$ が小さいらしいことが分る．そこで標本サイズを大きくすることを考える．しかし直ちに非常に大きくはしない．そうして得られたデータを再び解析すると，直面する問題の困難さに気づいて非常に大きな標本サイズでデータを収集する．このとき既に収集されたデータからの推論によって，事前分布の分散 τ^2 は小さく修正されている筈である．勿論データ収集のデータのデザインは個々の条件による．しかし $\tau^2 \doteqdot \sigma^2$ の想定のもとに直ちに非常に大きな標本サイズのデータ収集を計るような研究者は多くはないに違いない．そうすると段階を追って τ^2 が小さくなり，τ^2, $\tilde{\sigma}^2$, Δ の比が 1 に近づいていくのであって，結局 Lindley のパラドックスの仮定した状況は成立しない．

6. 信頼区間

統計的検定と信頼区間とは概念的にも密接な関係にある．実際信頼度 $1-\alpha$ とした μ の信頼区間 I に対して，$0 \notin I$ であれば，H_0 が棄却され，$0 \in I$ であれば H_0 が棄却されない．実務の場では統計的検定の結果を表現するのに有意性の有無よりも信頼区間を提示する事を薦める人が多い．また点推定は，大まかに言えば，区間幅 0 の（両側）信頼区間と見なされる．信頼区間についての検討は標準的な推測と Bayes 推測の比較に重要である．何故なら点推定に関しては検定について程には標準的な理論と Bayes 推論には大きな違いはない．

2 節で事後分布 (2.2) について $\mu = 0$ の事後確率を (2.3) で与えた．更に $\mu \neq 0$ の事後密度は

$$p_0(\mu) = \frac{\dfrac{(1-c)\sqrt{n}}{\sqrt{2\pi}\,\tau\sigma}\exp\left\{-\dfrac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \dfrac{1}{2\tau^2}\mu^2\right\}}{\dfrac{c\sqrt{n}}{\sigma}\exp\left\{-\dfrac{n}{2\sigma^2}\bar{x}^2\right\} + \dfrac{1-c}{\sqrt{\tau^2+\sigma^2/n}}\exp\left\{-\dfrac{1}{2\,(\tau^2+\sigma^2/n)}\bar{x}^2\right\}} \tag{6.1}$$

となる．従って $\bar{x} = O(1/\sqrt{n})$ あるいは $O(\sqrt{\log n/n})$ であれば，事後密度は $(1-P_0(0))\,N(\bar{x},\ \sigma^2/n)$ の密度で近似されることが分かる．事後分布は $\mu = 0$ の退化分布と $N(\bar{x},\ \sigma^2/n)$ の重み $P_0(0)$, $1-P_0(0)$ の混合分布で近似される．

特に $c = 0$ の場合の近似事後分布は $N(\bar{x}, \sigma^2/n)$ になる．この事後分布の両端 $\alpha/2$ の確率部分を除くと標準的な信頼区間 $(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n},\ \ \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$ を得られる．次に Bayes Factor (BF) による検定の基準と事後分布を考察する．BF > 1 と $P_0(0) > c$ とは同等だから直感的には妥当である．しかし $\bar{x}^2$ は $\log n/n$ のオーダーだから $N(\bar{x},\ \sigma^2/n)$ は $\mu = 0$ から大きく離れている．これを視覚的に分かり易いように図示する．状況として $\sigma^2 = \tau^2 = 1, c = 1/2$ とし，$\bar{x}$ と

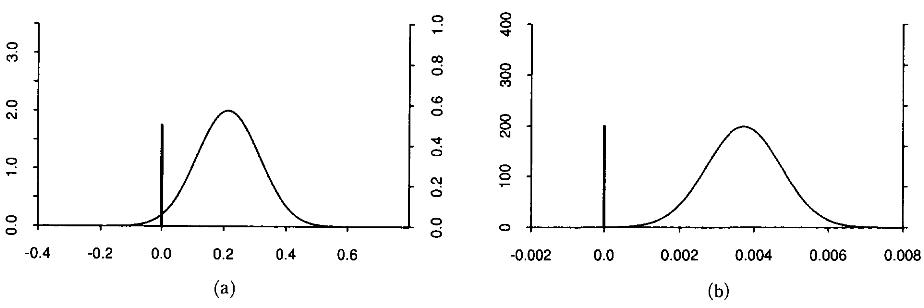

図 2．平均 μ の事後密度 $p_0(\mu)$ と事後確率 $P_0(0)$；$\sigma^2 = \tau^2 = 1$，$c = 1/2$，BF $= 1$ の場合．縦軸の左右の目盛りは各々事後密度，事後確率に対応．(a) $n = 10^2$ の場合．(b) $n = 10^6$ の場合．

して $\mathrm{BF}=1$ となる場合を $n=10^2, 10^6$ について図2に示した．確かに $P_0(0)=1/2$ であって，H_0 を採択するか否かの限界値である．しかし $p_0(\mu)$ の分布は標本サイズによって大きく依存している．標本サイズが 10^2 の場合はともかく，10^6 の場合は $p_0(\mu)$ はほぼ $\mu>0$ に分布している．後者の場合に $p_0(\mu)$ のみで判断すれば帰無仮説を棄却する証拠になる．Bayes 推測における事後分布の重要性を照らすと解釈に苦慮するように思われる．

7. 議論と見解

我々が Bayes 検定に批判的であった点は 1）曖昧な事前分布を用いること，2）H_0 が正しい確率 c を仮定すること，3）データ収集の実際に照らして稀な場合と考えられること，4）事後分布との対応がないこと，を掲げた．これらの考察の上で本節では3点について論じる．

7.1 モデル選択

本稿では検定との比較を行ったが，Bayes 検定はむしろモデル選択の規準と見なされる．実際一致性の概念として帰無仮説が正しく選ばれる確率を要請するのはモデル選択の立場であって，統計的検定の立場では意味がない．しかし概念的な違いを除けば上で論じた内容は適用可能である．

モデル選択を論じる際により明確にされる重要な点がある．それは大きな標本サイズを考えることによって H_0 と H_1 の2つのモデルの距離を大きくしている事実である．実際 $\mu\neq 0$ を任意に選んだとき $N(0, \sigma^2/n)$ と $N(\mu, \sigma^2/n)$ の自然な距離は $\sqrt{n}\,|\mu|/\sigma$ である．平均と分散を固定して n を大きくするとこの距離は大きくなる．さてモデル選択の代表的な場合である独立な正規誤差項をもつ多項式回帰モデル

$$y_i=\sum_{j=0}^{k} a_j x_i^j+\epsilon_i, \qquad i=1,\ldots, n$$

の次数を決定する問題を与える．ここで M_j, $a_{j+1}=\cdots=a_k=0, j=0,\ldots,k-1$ を考える．各々のモデル M_ℓ から各々 $\boldsymbol{a}_\ell\neq\boldsymbol{0}$ をとり，n を大きくすると，各々のモデル間の距離が同時に n のオーダーで大きくなる．すべてのモデル間の距離が同時に大きく離れるとする仮定は，理論的には大変に好都合であっても，実際のモデル選択の問題では現実離れした強い仮定である．

7.2 平滑化法

近年 Bayes 推測が注目を浴びている手法として平滑化法がある．実際に（経験）Bayes 推測が有用である．平滑化法では

$$\begin{aligned}&\boldsymbol{y}\sim N(\boldsymbol{\mu}, \sigma^2 I_n), \qquad \epsilon_i\sim N_i(0, \sigma^2)\\ &\boldsymbol{\mu}\sim N(\boldsymbol{0}, \tau^2 D_n^-)\end{aligned}$$

と表現される．行列 D_n は適当な行列で，例えば $\boldsymbol{y}'D_n\boldsymbol{y}=\sum(y_i-2y_{i+1}+y_{i+2})^2$，であり D_n^- は D_n の（一般化）逆行列である．平滑化法はその性能が良く，前小節の多項式回帰モデルにおける変数選択は急速にその適用例を減少させている．その理由の1つは多項式回帰モデルに格別の現実性がない事実である．ところで平滑法のモデルでは n が増加すると共に $\boldsymbol{\mu}$ の次元が増加する．2つの分散 σ^2，τ^2 を与えて n を大きくしても4節で言う曖昧な事前情報にならない．実際 $\boldsymbol{\mu}$ の十分統計量は $\boldsymbol{y}$ そのものであり，観察モデルと事前情報モデルの Shannon のエントロピーはともに n のオーダーで大きくなる．実際の平滑化では σ^2 と τ^2 もデータから推定する経験 Bayes 法が用いられるが，有用な Bayes 推測は事前分布が曖昧な場合でないという3節の

主張と符合している.

7.3 Bayes 検定が受け入れられる場合

本稿での議論は限られた内容であり，Bayes 検定を完全に否定した訳ではない．しかし上記の批判が適用されないとすれば，次の要件が満たされる必要がある.

先ず帰無仮説 H_0 が本当に正しいことが信じ得る場合である．これは観測に基づかないでより基本的な命題から演繹される知見である．しかもわずかな仮説からずれ Δ を検出したい場合である．次に，少なくとも解析者には，大量のデータが極めて短期間に得られる場合である．最後に事後分布については，分布全体ではなく，帰無仮説の正しい確率に特別な関心がある場合である.

筆者の理解する限りでは上のような状況は考えつかない．しかし敢えて考えてみれば高速通信システムで大量に2通りに送られたデータが同一の発信源であるかを調べるような場合には起こるかもしれない．しかしそうした場合でも統計的検定のような，即ち帰無仮説が正しいと見なすことがないという禁欲的な，推論方式が適合するかは保証の限りではない.

一般論から言って，異なる計量的な推論方式は一方が正しくて他方が誤っているという場合は多くない．むしろ委細に調べれば，双方がそれなりに妥当で適用範囲が違う場合が多い．しかし変動のあるデータから科学的認識を深めようとする問題に関する限り，Lindley パラドックスの提示した統計的検定の批判は含蓄に富むものではない．そしてその適用範囲は極めて限られていると思われる.

謝　辞

査読者からは投稿原稿の誤り，不適切な表現に関して多くの適切なコメントを頂いたことに感謝します.

参 考 文 献

赤池弘次 (1984)．確率の解釈における困難について，統数研彙報，**32**, 117-128.

Kass, R. E. and Raftery A. E. (1995). Bayes factors, *J. Amer. Statist. Assoc.*, **90**, 773-795.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion (with discussions), *J. Amer. Statist. Assoc.*, **90**, 928-934.

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

Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461-464.

Shafer, G. (1982). Lindley's paradox (with discussions), *J. Amer. Statist. Assoc.*, **71**, 325-351.

Spiegelhalter, D. J., Freedman, L. S. and Parmar, M. K. B. (1994). Bayesian approaches to randomized trials, *J. Roy. Statist. Soc. Ser. A*, **157**, 357-416.

Limitations on the Use of Bayesian Test under a Vague Prior Distribution

Takemi Yanagimoto

(The Institute of Statistical Mathematics)

One of critiques on the statistical test procedure is raised in association with Lindley's paradox. In the test procedure the null hypothesis is rejected with the probability α, which is a prefixed value named the significance level. The critique asserts that such a probability should tend to zero as the sample size tends to infinity. This requirement is called consistency of a test. The same controversy is found also in the model selection problem.

In this article we make it clear that the problem appears when the amount of information of data at hand is large while that of a prior distribution is relatively small. Then the variance of an estimator becomes much less than that of a prior distribution. Such a prior distribution is called vague. Emphasized here are: 1) Such a prior distribution is not realistic in practical scientific reasonings, 2) Careful considerations on a sample size are not taken account into, and 3) Difficulties in interpreting the posterior distribution arise. We conclude that Bayesian test will not be useful in scientific reasonings.

Key words: Confidence interval, consistency of test, model selection, posterior distribution, statistical test.