〈特別寄稿〉

線型仮説の検定に関する二,三のコメント†

J. W. プラット*

1. 序

私の見るところでは,このところ統計専門家——実務家および理論家の双方——の間で,仮説検定の問題に背を向ける傾向があるというのが公平なところであろうと思われます.このような傾向は大部分(統計的な推論を行なう際にはいつでも)単にある効果が存在するかどうかだけではなく,意味のある単位でその効果の大きさも推定したいという気持から生じています.この講演での私の目的は,とくに伝統的な方法が存在しない,あるいはそれからは不満足な結果しかもたらされないという理由で,統計家が仮説検定からさらに一歩を踏み出すことをためらっているような分野で,私なりの努力を払って,このような傾向を進展せしめ,議論を喚起することにあります.不幸なことに,伝統的な方法からあるいは伝統的でない方法からも,私は万能薬といったものを何一つ提供することはできません.しかも私は,以下で述べる事柄が何一つとして理論的に新しいものであると主張しているわけではないことを明確にしておかなければなりません.以下で私は,ある実際の状況をとり上げ,通常あまり見かけられませんが,とくにベイジアンの立場では周知となっているような理論から導かれる方法で,それらを検討することにしたいと思います.

2. ある実例

議論をできる限り具体的なものとするため,はじめにある現実のデータを導入しておきたいと思います.表1は数年前ハーヴァード・ビジネス・スクールで行なわれたあるコースについての七つのクラスの最終試験の得点分布について,その平均と標準偏差を示しています.学生たちは学校当局によってランダムに近い方法で各クラスに振り分けられております.このテストはどのクラスにも共通のものであり,客観テストの一つです.そして得点もかなり正規分布に近い形で分布しています.したがって,正規分布の理論に基づく分析を行なっても満足すべき結果が得られると考えてよいでしょう.

私が以下で行なおうとしている分析は,テストの得点のようなデータよりも別のタイプのデータに対して,より適切となることもあろうと思われます.というのは,私が実際に関心を寄せているのは,いまとり上げている特殊なデータではなく,データ分析の一般的な諸問題だからです.

3. 分散分析

伝統的な方法によれば,私たちはまず表2に見られるような分散分析表を作成しなくてはなりません.私はこれを,学生たちがあたかも各クラスに100人ずついるかのように作成しました.各クラスの正確な学生数——私もよくは覚えていないのですが——,この100人という数字からは多少のずれが

† ここで,この論文を最初作成する際に援助を受けた U.S. National Science Foundation,および日本での研究を可能にしてくれた John Simon Guggenheim Foundation に感謝したい.

〔訳注〕 この論文は,昭和47年10月30日,日本OR学会統計分科会で John W. Pratt 教授が行なった講演の原稿に同教授が講演後克明に手を加えたものを邦訳したものである.今度の講演は,以前に Pratt 教授がアメリカで計量経済学会のために行なった講演に基づいて行なわれたものであるが,アメリカでの講演はきわめて好評で,さまざまな論文で引用されているにもかかわらず,いまだに公刊されるに至っていない.今度OR学会で同様な主旨で行なった講演を機にはじめて公刊されたこの論文が,わが国におけるORあるいは統計学の発展に寄与するところは大きいであろう.

* Harvard 大学および京都大学.

表1 最終試験での得点の平均と標準偏差

クラス	平均	標準偏差	n
A	103.8	29.7	95
B	93.0	30.5	95
C	88.0	28.0	95
D	88.3	28.2	95
E	104.8	29.2	95
F	99.5	32.6	95
G	96.1	31.0	95
全体	96.21		665

表2 分散分析表

	自由度	2乗和	2乗和平均	2乗和平均の期待値
クラス間	6	26,840	4,473.2	$\sigma^2+95\theta^2$
クラス内	658	589,190	895.4	σ^2
全体	664	616,030		

MSB/MSW=4.996, $P=.000052$ $F_{6,658}$
SSB/MSW=29.97, $P=.000039$ $x^2{}_6$

あると思います．この表からクラス間の2乗和の平均とクラス内2乗和の平均の比を計算しますと，その値はほぼ5になり，自由度が6および658のF分布の下ではFがこの値を越える確率はほぼ0.00005となっています[1),a)]．

したがって，これらのデータに基づけば，仮説検定からクラス間には実際に差があるという結論が導かれます．しかし私たちは，このような事実を仮説検定を行なわぬ以前に承知しているのです．というのは各クラスを異なった教師が教えており，1日のうちでも異なった時間にクラス討議が行なわれ，学生たちも当初はランダムに振り分けられていても，学期の間には相互に影響を与え合っているからです．これはいまのような状況でなぜ仮説検定に背を向ける傾向が生ずるかを例示しています．仮説検定は帰無仮説が正しいかどうかについて何事も語ってはくれません．なぜならわれわれは，はじめから差があることを知っているからです．仮説検定は十分なデータがあり，検出力が十分大きく，統計的に有意な結果を得るに十分なチャンスがあるかどうかをわれわれに語ってくれるにすぎないのです．

1) Peizer および Pratt (1968) の方法による．ちなみに，自由度658を∞と見なし，クラス間2乗和の平均とクラス内2乗和の平均の比を自由度6のx^2分布に従うものと仮定すると，x^2がこの値を越える確率は0.00005ではなく，0.00004となる．この程度の誤差は，あまり重要性を持たないとはいえ，この25%という相対誤差は，通常の方法に従ってきわめて小さな確率を近似する際に，いくつかの有意水準を引用することが無意味であることをよく示している．

4. 伝統的な点推定と信頼区間

伝統的な方法による次の段階は，おそらくそれぞれの母分散を推定することでしょう．私たちはここで，各クラスの得点の分布の母平均がランダムな変量であるかあるいは一定の母数であるかを特徴づけなければならないことに注意しておきましょう．このいずれの場合についても触れておくべきでしょうが，いまはそれを母数と見なしておくほうが，私が議論したいと思う事柄をより良く例示してくれるように思えます．そこでいまは各クラス得点の母平均がμ_i $(i=1,2,\cdots\cdots,7)$ であり，これらはそれぞれ表1で与えられる観測された平均値で自然な形で推定されるものとします．また，得点全体の平均μも，全部のクラスの学生たちの平均得点によって自然な形で推定されるものと仮定します．しかしわれわれはある尺度でどれだけクラスの間に差があるかということに興味を持っており，これらの値それ自体にはそれほど興味を惹かれません．便宜上，そして変量モデルとの両立性を考慮して，少なくとも出発点として$\theta^2=\dfrac{\sum_i(\mu_i-\mu^2)}{6}$をそのような尺度にとることにしましょう．このような前提の下で伝統的な統計家は，これを表2のそれぞれの期待値に等しいとおき，パラメータについて解くことにより，θ^2の推定値として38を，したがってθの推定値として6.1

表3 伝統的な推定値と信頼区間(正規近似による)

$\hat{\sigma}^2=\mathrm{MSW}=895.4,\ \hat{\theta}^2=(\mathrm{MSB}-\mathrm{MSW})/95=37.66,$
$\hat{\theta}=6.14$

$\hat{\delta}^2=95(6\hat{\theta}^2)/\hat{\sigma}^2=6\left(\dfrac{\mathrm{MSB}}{\mathrm{MSW}}-1\right)=23.97$ (非心パラメータ)

$\hat{V}(\hat{\theta}^2)=\dfrac{\hat{\sigma}^4}{(95)^2}\left(\dfrac{12+4\hat{\delta}^2}{36}+\dfrac{2}{658}\right)=266.5,\ \mathrm{S}_{\hat{\theta}2}=16.32$

正規近似によるθ^2の95%信頼区間

$$\hat{\theta}^2\pm1.96\,\mathrm{S}_{\hat{\theta}2}=(5.66,69.66)$$

これに対応するθの信頼区間=(2.38,8.35)

を得ます（表3を参照）．ここで彼は多少落ち着かない気持になるでしょう．というのは，θ^2 の最小分散不偏推定値が負となることがあり，したがって損失関数の名に値するどのような損失関数の下でも許容的でない（inadmissible）ことを承知しているからです．しかしながら，θ^2 の最尤推定量も彼にとってはあまり魅力のあるものとはならないでしょう．それが観測された平均の標本分散だからです．これはクラス間2乗和の平均を95で割った値です．これは負になることはありませんが，その期待値は $\theta^2+\frac{1}{95}\sigma^2$ となり，かなり大きな上方の偏りを持っています．彼はまた不偏推定量の平方根が下方に偏りをもつことを思い出して，θ の推定値についての偏りを多少心配するかもしれません．

もし信頼区間を求めなくてはならないとなれば，伝統的な統計家はとくに落ち着かない気持になることでしょう．というのはこのような単純な状況にあってすら，彼は信頼区間を求める十分な方法を知らないからです．もし彼が私のように怠惰な人間であれば，おそらく彼は θ^2 についての彼の推定値の標準偏差を推定し，正規近似がそれほど悪くないことを希望することになるでしょう．彼は表3に与えられるような形で信頼区間を求め，少なくとも下限については，おそらく正規近似がきわめて危険であろうことに気がつくでしょう．そこで彼は表4におけるような手続きに従うことになるでしょうが，私はこれ以上彼の努力の跡をたどろうとは思いません．私はこの問題に関する最終的な解答を知らないからです．

表4 非心 χ^2 分布および $\sigma^2=895.4$ を仮定して求められる伝統的な信頼限界

	下側 97.5%	50%	上側 97.5%
θ^2 に対する信頼限界	12.9	39.5	75.8
θ 〃 〃	3.6	6.3	8.7

ここで50%信頼限界はメディアン不偏推定値である（計算の際は非心 χ^2 分布を同じ平均と分散を持つ通常の χ^2 分布で近似しているiv)．

5. ベイジアンの推論

θ についてのベイジアンの推論は，私たちがたとえばSSB（級間2乗和）およびSSW（級内2乗和）のような分散分析で用いられる統計量に注意を限定する限り，直接的で単純なものです．しばらくの間これら二つの統計量以外には，データからベイジアン流に情報を引き出すことはしないことにします．いずれにせよ，私たちは θ についての伝統的な推論を行なう際にもそのようなことはしていなかったわけです．SSWは $\sigma^2\chi^2_{658}$ の分布に従い，SSBは $\sigma^2\chi'^2_6(\delta^2)$ に従います．ここで $\delta^2=95(6\theta^2)/\sigma^2$ は非心パラメータを示しています．これら二つの分布の密度関数を乗じ，その積を σ および θ の関数と見なし，これに事前分布の密度関数を乗じて正規化すると，σ と θ の事後分布の密度関数が求められます．ここで σ について積分すれば，θ の事後分布の密度関数が得られることになります．

通常の事前分布に対して，われわれは σ^2 が既知でMSW＝895であると考えて処理して，θ の事後分布を求めてもそれほど大きな誤りを犯すことにはならないでしょう．このとき $\delta^2=.637\theta^2$ となり，尤度関数の自由度は6であり，$SSB/\sigma^2=29.97$ で評価される非心パラメータ $.637\theta^2$ を持つ χ^2 分布の密度関数を θ の関数と見なしたものとなります．これを基準化すれば，事前の密度関数が区間 $(0,\infty)$ 上で一様であれば（あるいはある適当な区間内で平坦であり，その間以外であまり大きな値をとらなければ），この関数が θ の事後の密度関数となります．これは図1に示されています．図1にはその他 θ^2 および $\log\theta$ に対して一様な事前の密度関数を前提としたときの事後の密度関数も載せてあります．$\log\theta$ に対して一様という事前分布はJeffreysのいわゆる“不変性”に基づいて選ばれた分布です2)．これら三つの事後の密度関数は，θ の伝統的な方法で求めた推定値に近似した平均を持つ正規分布にきわめて近い形をしています．しかしこれらは同一の非心 χ^2 分布の仮定に従って求められた表4の信頼限界から示唆されるよりも，標準偏差はいくぶん大き目であり，歪度（skewness）は小さくなっています．表5

2) このとき尤度は $\theta\to0$ のとき0に近づかない．したがってJeffreysの選択した事前分布からは，$\theta\to0$ のとき定数 $1/\theta$ のような挙動を示し，したがって，“特異”（improper）な事後の密度関数が導かれる．おそらくわれわれはJeffreysの事前分布を，このような事態が起こらないように修正し，$\theta=0$ のところで事後の密度関数の値が小さくなるように（しかし0にはならない）するほうがよいであろう．以下の議論ではこのような修正が施されたものと仮定される．

は比較のために事後分布の分位点をあげています.

主観的なベイジアンの立場から見ると，これらの事後分布あるいは他の事後分布のうちでどれが“すぐれて”いるかは当事者の事前分布に依存しており，主観的な事柄になっています．そして主観的な事前の見解を考慮せずにより精密な推論を下そうとするどのような方法も，現実への妥当性を犠牲にしてはじめてその精度を高めていると考えられます．私はこの点に関して強い確信を持っていますが，それをここで直接的な形で議論しようとは思いません．しかしあとでいくつかこの点について触れることになりましょう．

私がこれまで述べてきた事柄の要点は次のようなことです．すなわち，伝統的な統計学ではある種の問題をきわめて厳密に解いていますが，しかしその解答は，いくつかのクラス平均が厳密に等しいかどうかという，あまり興味をそそられない問題について与えられているのです．実際に興味をそそられる問題に対して伝統的な方法で解答を与えるためには，

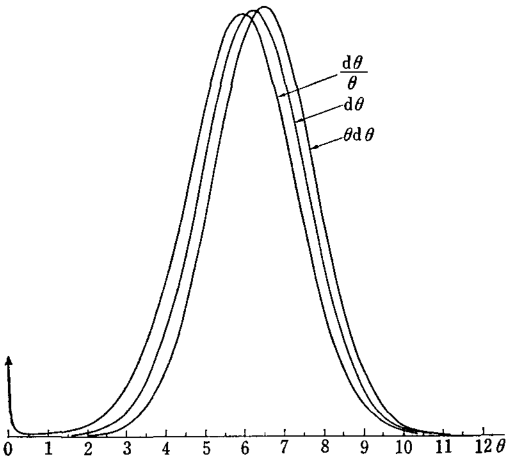

図1　さまざまな事前密度関数の下で得られる事後密度関数

表5　θ の事後分布の分位点

事前密度関数	事後分布の標準偏差	事後分布の分位点 .025	.50	.995
$\frac{d\theta}{\theta}$†	1.38	3.2	6.0	8.6
$d\theta$	1.34	3.6	6.25	8.8
$\theta d\theta$	1.30	4.0	6.5	9.0

これらの数値はグラフから求めたので多少の誤差があろう（†　注2)を見よ).

われわれは学問的な研究という性格を持つさまざまな数学上の問題を解決しなければなりません．そしてこのような事柄は数理統計学者たちにさまざまな仕事を与えるわけです．しかし，実際家は，仮りに彼がそれらの問題を解決する能力を持っていたとしても，それに時間を費やすわけにはいきません．おそらくこのことは，なぜ OR の分野で統計学者がそうあるべきだと考えるほど統計学が使われていないかという理由でしょう（少なくともアメリカではこのように考えられています)．ベイジアンの方法はわれわれにわれわれ自身の問題を設定し，それを直截的な形で解くことを可能にします．これは OR 実務家にとって魅力ある方法でしょう．事前分布に関する厄介な問題ですら，私の見解では，真の意味でベイジアンの方法の欠陥とはなりません．というのは，この問題については感度分析を行なうことができるからです．これは図1で例示されています．

6. ベイジアンの仮説検定

統計的仮説検定についてはさまざまな議論が行なわれているので，それらをベイジアンの観点からもう少し検討しておく必要があるでしょうし，またそれが教訓的であることもありましょう．もし私たちがベイジアンの理論的枠組のなかで仮説検定についで何かいおうとすれば，われわれは事前分布の選択の問題について真剣に考えなくてはなりません．そうするために私たちは，まず θ に解釈を与えなければならないことに気がつきます．そしてこれが問題のすべてでもあります．われわれの例で θ はクラス効果の平均2乗和の平方根であり，6というクラス効果は，もしある学生が中程にいれば，彼の成績を約8％点分だけ改善する（端のほうでは0％点分の改善にしかなりませんが)ことになるでしょう．あるいは別のいい方をすれば，$\theta=6$ ということは学生間の分散の約4％，あるいは標準偏差の約2％がクラスの差異によって説明されることを意味しています（この後の解釈は θ^2 ではなく $\frac{\theta^2}{\sigma^2}$ に関して推論を行なうことを示唆しています．これは理論的にはより単純な問題ですが，やはりあとの目的のためにはあまりよい解釈とはいえません).

7. 帰無仮説 $\theta=0$ に関するベイジアンの推論

もし帰無仮説 $H_0:\theta=0$ が事前確率0を持つもの

であれば，その事後確率は常に0になってしまいます．この点はわれわれの例についても，また有意性検定が頻ぱんに利用されるきわめて数多くの他の状況においても成り立っているように思えます．しかしながら場合によっては，帰無仮説のうちにほとんどその成立が自明であり，したがって正の事前確率を付与したいと考えるようなものがあることでしょう．

いまわれわれが帰無仮説に対して実際にある正の事前確率 $P'(H_0)$ を付与したと仮定してみましょう．議論の便宜上これを帰無仮説についての勝目(odds)を用いて表現することにしておきます．このとき事前の勝目は $\dfrac{P'(H_0)}{P'(H_1)}$ となっています．ここで $P'(H_1)=1-P'(H_0)$ とします．統計量 MSB が与えられたとき，事後の勝目は

$$(1)\quad \underset{\substack{\uparrow\\ \text{事後の}\\ \text{勝目}}}{\frac{P''(H_0)}{P''(H_1)}}=\underset{\substack{\uparrow\\ \text{事前の}\\ \text{勝目}}}{\frac{P'(H_0)}{P'(H_1)}}\ \underset{\substack{\uparrow\\ \text{尤度の比}}}{\frac{f_0(\text{MSB})}{f_1(\text{MSB})}}$$

となります．ここで P'' は事後確率を表わし，f_0 は H_1 の下での MSB の尤度を，また f_1 は H_1 の下での MSB の尤度を示しています．H_1 の下では θ は未知パラメータであり，f_1 は H_1 の下での θ の事前分布に関する周辺密度関数の値となっています．

この分布は通常の方法で無知を表現するような形に選ぶことはできません．たとえば，もし θ に H_1 の下で0と R の間の一様分布が付与されたとしますと $R\to 0$ のとき $f_1(\text{MSB})\to 0$ となり，H_0 に対する事後の勝目→∞となってしまいます．直観的には，R が大きいとき，$(0,R)$ 上の一様分布は θ がきわめて大きいことに高い確率を与えることを意味しており，このことは中間的な MSB の値が起こることはほとんど考えられないことを意味しています[3]．2,3の私の同僚との話合いから，私は $(0,20)$ の一様分布よりも H_1 に対して不利な θ の密度関数，あるいは平均6，標準偏差2の正規分布よりも H_1 に対して有利な θ の密度関数を付与しようとする人はほとんどいないだろうという結論を得ました．これから(1)式に代入さるべき標本の勝目について次のような範囲が得られます．

$$(2)\quad \frac{f_0(\text{MSB})}{f_1(\text{MSB})}=\frac{f_{\chi^2}(29.97\,|\,\delta^2=0,\ \text{自由度}\ 6)}{\int p_1(\theta)f_{\chi^2}(29.97\,|\,\delta^2=.637\theta^2,\ \text{自由度}\ 6)d\theta}$$

$$=\begin{cases}.00083 & p_1\text{が平均6, 標準偏差2の正規分布のとき}\\ .00273 & p_1\text{が区間}(0,20)\text{上の一様分布のとき}\end{cases}$$

もし，$P'(H_1)$ があまり小さな値でなければ，$P''(H_1)$ は1にきわめて近い値となり，したがって $P''(H_0)$ は近似的に事後の勝目に等しくなります．とくに，もし H_0 および H_1 に等しい事前確率が付与されるならば，$P''(H_0)$ は.0008と.0027の間のある値となるでしょう．これらの値は，いずれも通常と仮説検定で得られる判定値と比較してきわめて大きなものとなっています．事実，それらは σ^2 を既知とする同じ近似方法の下で計算される判定値，$P=.00004$ の21倍から70倍にもなっています．議論の要点は P の値が帰無仮説の不可能性を測定すると仮定されていることにあります．しかしもしそれがどのような人の事後確率とも大幅に異なっているのであれば，このような解釈は疑問であり，少なくとも困難なものとなります[i)]．

8. 帰無仮説 $\theta\fallingdotseq 0$ の検定

われわれが，$\theta\fallingdotseq 0$ であることを知っているとき，帰無仮説 $\theta\leq\delta$ (ここで δ はある小さな値で，前に出た非心パラメータとは関係がない)を検定するための一つの近似として，帰無仮説 $\theta=0$ を検定してもよいという議論が成り立つかもしれません．これは誰もが誤りであることを知っている帰無仮説を検定するという偽善を緩和する役割を果たすものではありますが，不幸なことに多くの場合，すぐ上で言及

3) θ が与えられたとき，MSB の密度関数は非心 χ^2 分布であり，これを θ の関数と考えたとき，近似的に平均6.25，標準偏差1.34の正規分布に近い形をしており，$\theta=0$ のところでの値は $\dfrac{1}{2180}\times$ 最大値となっている．したがって $(0,20)$ の一様分布を仮定するときは，

$$\frac{f_1(\text{MSB})}{f_0(\text{MSB})}=2180\int_0^{20}.05\exp\left\{-\frac{1}{2}\left(\frac{\theta-6.25}{1.34}\right)^2\right\}d\theta=366$$

となり，また平均6，標準偏差2の正規分布を仮定するときは，

$$\frac{f_1(\text{MSB})}{f_0(\text{MSB})}=2180\int_0^{\infty}(2\pi)^{-\frac{1}{2}}\frac{1}{2}\exp\left\{-\frac{(\theta-6.25)^2}{2}\left[\frac{1}{4}+\frac{1}{(1.34)^2}\right]\right\}d\theta=1200$$

を得る．

された問題を激化する傾向があるように思えます．すぐわかるように，われわれの例ではたしかにそのようになっています．

この方法が意味を持つためには，δについての"小さな"という言葉が実際的に意味を持つように定義されなければなりません．この場合，たとえば，$\theta \leq 1$ならば実際上の目的に対してθが小さいという決定を下すかもしれません．というのは，このことはあるクラスの中央値が通常全得点の中央値から約2％点だけ異なることを意味しているからです．0と1の間のθに対してMSBの密度は$\theta=0$のときの密度の1～3倍になっていますから，帰無仮説$\theta \leqq 1$の下での周辺密度f_0(MSB)は前の値の1～3倍となり，したがって勝目も前の場合の1～3倍となるでしょう．しかしそれらは，前に示唆した判定値よりもはるかに大きな値となっており，いまの場合についても，前と同じことが成り立ちます[4, b)].

9. いくつかの注意

以下の注意については，$\theta=0$という仮説から，σ^2が既知という仮定の下で統計量MSBの分布が決められるように，帰無仮説が統計量の分布を完全に決定するものと想定することにします．

A. 尤度比$f_0(\mathrm{MSB})/\max_{\theta} f(\mathrm{MSB}|\theta)$は，§7の計算に現われた標本の勝目$f_0(\mathrm{MSB})/f_1(\mathrm{MSB})$に対する一つの下限となっています．もし$H_1$の下での$\theta$の事前分布が$\theta$の最尤推定値に確率1を与えるようなものであれば，尤度比はこの標本の勝目に等しくなるでしょう．試験の得点データの例では，尤度比は$\frac{1}{1270}=.000461$となっています．この点に関する二つの意味合いは，いまの例では次のようになります．

(i) 尤度比は，H_1の下におけるθについての任意の"合理的"な事前分布から得られる標本の勝目のうちで最小の値，.00083の55％になっている．

(ii) H_1の下での任意の事前分布から得られる最小の標本の勝目は，やはり判定値の9倍になっている．

私には，最小の標本の勝目が多くの場合にいまのようにこの尤度比を乗じた値ほど小さくなることには疑いを持っています．なぜなら私ならば通常H_1にとって最も有利な"合理的"事前分布については，いまの場合よりも尤度に比較してより拡散していると考えるであろうからです．

B. I. J. Good (1950年) は，標本の勝目が常に判定値よりも大きくなるという議論を行なっております．彼の述べるところによれば (1958年, 1965年)，彼が数多くの例で，.001から.2の間の判定値に対して標本の勝目が判定値の3倍から30倍となっており，.01以上の判定値の場合は，判定値の4.5倍という場合が最も多く，.001以下の判定値の場合には約10倍という場合が多いということです．これらについて一つの興味深い例が立入った形で議論されています(Good, 1969 ; Efron, 1971 参照).

表6 さまざまな標本の勝目に対する両側t検定の判定値

標本の勝目		1	.32	.1	.032	.01
$P''(H_0)$		.5	.24	.091	.031	.01
自由度	5	.12	.03	.007	.0017	.0005
	10	.08	.02	.005	.0013	.0004
	20	.06	.015	.004	.0011	.0003
	120	.02	.006	.0016	.0005	.00014

表中の数字は判定値を示す．標本の勝目はJeffreysが彼の事前分布（コーシー分布の下で計算した値(1961, 表Ⅲ)．$P''(H_0)$は追加的な仮定$P'(H_0)=P'(H_1)=\frac{1}{2}$の下で

$$P''(H_0)=\frac{\text{標本の勝目}}{(1+\text{標本の勝目})}$$

として計算された．

4) 変化は一般にこの方向に起こると考えられる．それは帰無仮説を拡大すれば，一般に判定値が小さい状況で仮説のもっともらしさが増大するからである．これは帰無仮説$\theta \leqq \delta$の厳密な検定が行なわれたときにこのようなことが成り立つといっているわけではない．分散σ^2を既知とする近似の下では，帰無仮説$\theta \leqq 1$の検定の判定値は非心χ^2分布から計算されるが，前と同じ，.00004ではなく，.00012となっている．またf_0(MSB)の以前の値はおそらく近似的に2倍となっているであろう．したがってわれわれは，いまの場合，.00012を.0017および，.0054と比較することになろう．これは前に得た21倍から70倍という値とくらべると改善になっているが．仮説$\theta \leq 1$の厳密な検定を正当化するにはほど遠い．

Lindley (1957年) は，任意の固定された判定値に対応する標本の H_0 についての勝目が標本のサイズが大きくなるにつれ，通常，∞に近づいていくが，その近づき方は緩慢であることを示し，例をあげて説明しています．

Jeffreys(1961年)もまたこの点を例示しています．t 分布の場合，彼の事前分布の下で，表6が得られます．Jeffreys が多かれ少なかれ伝統的な事前分布を用いているという事実にもかかわらず，3倍から30倍という範囲に一致しています．

10. すべてのデータからの推論

伝統的な統計学により，われわれは多かれ少なかれ恣意的に選ばれた統計量に基づいてさまざまな推論，あるいは少なくとも有意性検定を行なうことができます．そのような推論は通常，“完全に効率的ではない”にしろ“妥当な”ものであると見なされています．おそらくベイジアンもこれと同じ自由を持つべきでしょう．しかしベイジアンにとっては，すべてのデータを用いた推論を行なうのが理論的に簡明であるため，彼がそうしようと思うことには多少抵抗があるように思えます．この場合たとえば彼は，$\mu_1, \cdots\cdots, \mu_7$ および σ に対して同時事前分布を付与し，それらの同時事後分布を求め，その後で単純に積分などの方法で θ の事後分布を導出するという手続きに従うことになるでしょう．

しかしながら実際には，とくにもっと複雑な問題では，完全な同時事前分布を慎重に付与しなければならないという仕事を回避したいと考えることもありましょう．その場合には，独立性，対称性，および散布度についてなんらかの判断を下すことで，少なくとも近似的には θ に対する事後確率，および θ に関する未知の仮説の事後確率を求めるには十分であるといったことを利用することになりましょう．

11. パラメータの選択

パラメータ θ をクラス間の変動の尺度と考えることについては，それが神聖にして不可侵といった事柄は何一つ存在しません．単純化のために，たとえばいまわれわれが，ある2方分割表での交互作用に関心を寄せているとすれば，われわれは実際上意味を持つある閾値を上回るような交互作用があるかどうかを知りたいと思うが，それを下回る交互作用についてけ関心がないと考えるかもしれません．このことは交互作用の最大値について推論を行なうことを示唆しています．テストの得点データについてこれを考えれば，μ_i をクラス i の真の平均得点としますと，$\Delta = \max \mu_i - \min \mu_i$ についての推論を行なうことに相当するでしょう．

このようなパラメータについて伝統的方法で推論を下そうとすると，それはきわめて困難な仕事になります[5]．そして仮りに Δ について有意性検定が行なわれたとしても，それはおそらく θ についての検定の場合と同様に不適切なものとなることでしょう．ベイジアンの推論では理論的になんの問題も生じませんが，この場合でも伝統的な“特異な”事前分布を使用すれば，誤った結果がもたらされるという可能性は残されています[ii]．たとえば，もし μ_i が互いに独立で（σ とも独立な）無限大の区間上の一様分布に従っていると仮定しますと，

$$\max m_i - \min m_i > \max \mu_i - \min \mu_i$$

という事象は任意の m_i に対して $\frac{1}{2}$ よりも小さい事後確率を持ち，したがって事前確率も $\frac{1}{2}$ よりも小さくなっています（ここで m_i は観測された平均得点を示します）．しかしこの事象について任意の μ_i が与えられたときの条件付確率は $\frac{1}{2}$ より大きく，したがって（事前の）周辺確率は $\frac{1}{2}$ よりも大きくなければなりません．事実，特異な事前分布は，伝統的な統計家があれほど忌避している抽出の偏りをもたらします．任意の正常な事前分布に対して“Δ の事後の中央値 $> \Delta$ の実際の値”という事象は，ある μ_i に対しては $\frac{1}{2}$ よりも大きい条件付確率を持つでしょう．しかしこれは μ_i の事前分布に関する積分がちょうど $\frac{1}{2}$ になるという意味で，他の μ_i に対する $\frac{1}{2}$ よりも小さい条件付確率によってちょうど相殺され

5) たとえば，任意の推定量の期待値は μ_i の解析関数となるが，Δ はそうではないから，不偏推定量は存在しない．またたとえば $\Delta \leq 1$ についての検定には最も不利な検定領域，あるいは多重比較の考え方を必要とし，通常よく見られる検定領域に対してそれらは極度に「保守的」なものとなろう．そして上で言及されたような単純な相互作用についてすら，$\Delta = \max \mu_i - \min \mu_i$ よりも複雑な状況では，それらを求めることは困難なことになろう．

ることになるでしょう．しかしながら無限区間上の一様分布は，最大のμ_iがすべての他の値よりも限りなく大きくなり，また最小のμ_iが限りなく小さくなることが確実であることを意味しています．したがって最大のμ_iと最小のμ_iの選出は確実に正しく，選出の偏りは存在しないことになります．いいかえれば，m_iを最大にすればiは確実にμ_iを最大にする，等々となっています．

ところで，無限区間上で一様な事前分布を利用することから生ずる誤りは，おそらくそれほど重大なものではありませんが，いくつかの状況では，とくに数多くのパラメータが含まれており，それらのほとんどが小さな値をとるといった場合には，きわめて重大なものとなるでしょう[iii]．そしてもしそのような状況に対して，選出の偏りという前門の虎と数多くのパラメータに対する同時事前分布の直接的付与という後門の狼とを同時に処理するような簡便なベイジアンの手法が得られれば，それは素晴らしいことでありましょう．おそらく一つの満足すべき解答はμ_i（あるいはさらに複雑なモデルにおける対応するパラメータ）をランダムな効果と見なすことから得られると考えられます．これは，さまざまなμ_iが生成される母集団についてだけ推論を行なうというよりは，むしろ個々のμ_iについて推論を行ないたいと考えているという意味で，通常の変量モデルの状況とは異なるものとなりましょう．最近，この考え方に沿った研究が Zellner (1972) によって検討を加えられております．

12. 問題点の整理：推論それ自体としての検定

これまで私たちは，検定の目的が推論それ自体である場合に，いくつかのパラメータが同時に0となる，あるいは平均に小さい，といった帰無仮説の検定について考えてきました．時によっては，稀にしか起こりませんが，いくつかのパラメータが実際に0であり，この仮説が正しいかどうかについての一つの推論として検定が行なわれるという場合が起こるかもしれません．しかしもっと多く起こるのは，パラメータがけっして0にはならないが，小さいかもしれないといった可能性について推論を下したいという場合です．それは，もしこのことが正当化されれば，それらのパラメータ，すなわち，私たちが効果を測定している要因については，それ以上考慮しないですますことができるからです．

このようなことを行なう際，私たちが単に状況の記述とそれについての思考を単純化したいということもありましょう．あるいはそれらの要因を示すことによって，将来の時点における測定，そしてあるいは将来における経営上の意思決定において経費を節約したいと考えてのことかもしれません．少なくとも私たちの思考を単純化することに関する限り，数多くの可能な要因が存在するが，おそらくそれらの大部分は，実際上ほとんど重要性を持たないといった状況において最も大きな効果が現われるでしょう．

しかしこれまでの議論から，次のようなことが示されています．それはすべてそのような場合にあって，不幸なことに，いずれのタイプの帰無仮説についても，有意性検定の判定値は，少なくとも帰無仮説が正しいという事後確率がある程度の確信の下に計算しうるものである場合には，そのような事後確率とまったくかけ離れた値をとり，そのような事後確率となんら一貫した関連性を示さないということです．このことは有意性検定がせいぜいのところ，あまり信頼のおけない推論の一形式にすぎぬことを示唆しています．そしてどう見ても，この信頼がおけないという性格は，事後確率を求めるのが困難であるような状況についても持ち込まれるように思われます[6]．

点推定と信頼区間については，通常それらについての自然な形での事後分布による解釈を与えるのが有益であるばかりでなく，実際上も近似的に正しいという意味で，検定と比較してはるかにすぐれていると考えられます．

13. 予備的な検定

もっと議論しておく必要のある検定の一つの利用方法は，さらに分析を続けるための予備的な段階としての検定です．この種の検定の一例としては，モデルを正当化するために行なわれる適合度検定があ

6) 判定値が通常近似的に帰無仮説の事後確率に等しくなる一つの場合は，帰無仮説が$\mu_1-\mu_2<0$といった形の片側仮説の場合である．§8で議論した仮説$\theta\leq 1$は，もし標本のサイズが十分大きく，$\theta=0$における効果が無視しうる程度である場合にはこの形のものとなるであろう．しかし実際には，$\theta=0$での効果はこの種の仮説に対して無視しうる程度に小さくなることはほとんどないであろう．

ります（χ^2 検定のようなすぐれたゆるい仮定の下での検定を用いる場合には，たいていこの場合に相当するでしょう）. もう一つの予備的検定は異常観測値の検定です. また，今日の議論の主題にもっと近い（もっとも私は，それにそれほどとらわれていませんが……）検定としては，いくつかの交互作用あるいは共分散が0となるかどうかといった検定があります. この場合の意図は，以後の観測値の予測，あるいは，主効果またはさまざまな係数の一次式の推定を改善するために，それらがおそらく0であろうという事前の判断を利用することにあります. これまでの議論は，この種の検定の利用についてほとんど語るところがありません. 判定値について前に述べた注意はこの場合にも適用されます. しかしそれは，ある意味で的を外れたものとなっています. 実際に問題となるのは，分析から導出される推定あるいは予測に対する予備的な検定の効果なのです. 私は直接，適合度あるいは異常値について議論することはしませんが，推定あるいは予測に先き立って予備的に行なわれる，いくつかのパラメータが0であるかどうかの検定について議論しようと思います. 私の印象では，この線に沿った伝統的な統計学上の研究のほとんどは，予備的な検定に使用すべき適切な有意水準は5％ではなく50％に近い値であると結論しているように思えます. しかし私は，この点についての検討をまだ行なっていません[7].

議論を単純化するために，少し単純化しすぎるきらいはありますが，観測値は互いに独立でそれぞれ $N(\mu, \sigma^2)$ に従って分布しており，私たちは $H_0: \mu=0$ の予備的な検定をしようとしていると仮定します. これはもちろん通常よくみられる状況ではありません. しかしおそらくこの仮定から，もっと複雑な状況で行なわれる予備的な検定について，いろいろと知ることができることでしょう. また信頼区間や将来の観測値に対する予測および許容区間も重要ですが，少なくとも最初のうちは μ の推定に考察を限定することにしましょう.

$\mu=0$ を水準 α で検定する一つの予備的な検定で

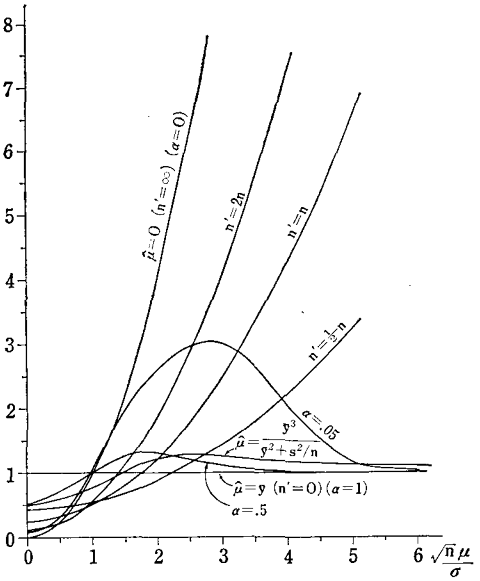

図2 （$n=9$ の場合の）さまざまな α に対する予備的 t 検定から得られる推定量および James R. Thompson のいわゆる“縮約”推定量（Shrunken Estimator），$\bar{y}^3/(\bar{y}^2+\frac{s^2}{n})$ について，さまざまな n' に対する事後平均の平均2乗誤差. 後者の二つについては Thompson (1968) から収録した.

ある t 検定では，$|\bar{Y}|>t_\alpha s/\sqrt{n}$ であるかあるいは $|\bar{Y}|<t_\alpha s/\sqrt{n}$ であるかに従って μ を $\bar{Y}$ と推定するか，あるいは0と推定することになります. ここで t_α は t 分布の両側 α％点，s は標本標準偏差，そして n は標本のサイズを示します. この推定量の形は奇妙なものです——それは観測値の不連続関数となっています. そしてもちろん，それは不偏推定量でも，最尤推定量でも，最小2乗推定量でもありません.

これに対応するベイジアンの方法は，まず (μ, σ) の事前分布を付与し，μ の事後分布を求め，その事後分布の平均で μ を推定することになりましょう. この場合も，無限区間で一様な事前分布は適切なものとはいえないでしょう. なぜなら，私たちが $\mu=0$ を検定しようと考えているという事実は，私たちが $\mu=0$ の近傍に正の確率を付与することを意味しているからです. もし μ の事前分布として平均が0となっている H. Raiffa および R. Schlaifer のいわゆる自然な形で（十分統計量に）随伴する natural conjugate）分布をとりますと，μ の事後分布の平均は $\bar{Y}\cdot n/(n+n')$ となります. ここで n' は μ に関する

7) 通常の線型仮説については，50％の水準で棄却するということは $F\geqq 1$ に対して棄却することとほとんど同等である. そしてこのことは，周知のようにまたしばしば見直されていることだが，より小さい残差分散推定値を選ぶこと，あるいはなるべく大きな修正ずみの R^2 を選ぶことと同等である.

事前の情報を観測値の数に換算したものとして解釈される，事前分布パラメータを示しています．

図2は，さまざまなn'/nに対するこのような事後分布の平均およびさまざまな水準αの予備的な検定から得られる推定量の平均2乗誤差のグラフを示しています．もちろん，事後分布の平均は，ウエイトが事前分布の密度であるような平均2乗誤差の加重平均を最小にします．

私たちは，予備的な検定にはどのような水準αを用いるのが最良であるかという疑問を持つことでしょう．もし私たちがこの問題に答える際に，ベイジアンの観点を採用するのであれば，あるいは事前分布の密度に等しいウエイトでの加重平均を使用するのであれば，これまでの事前分布に対しては$n'<n$のとき$\alpha=1$，$n'>n$のとき$\alpha=0$が最良であるという結論に導かれます．すなわち，これは極端な$\alpha=0$，あるいは1という水準以外のなんらかの水準αで予備的な検定を行なうよりは，μを$\bar{Y}$あるいは常に0で推定するほうがよいことを意味しています．

実際はもっと多くのことがいえます．この問題に対する最も純粋な形での伝統的な接近法と考えられるものでは，あらかじめモデル$N(\mu,\sigma^2)$および推定値$\bar{Y}$を選ぶか，あるいはモデル$N(0,\sigma^2)$および推定値0を選ぶかのいずれかであることに注意しましょう[8]．予備的検定はこのような選択を，伝統的な統計学上の用語に固執しながら，データに依存して行なう一つの方法です．しかしベイジアンの観点からは，たとえその選択が事後的に行なわれるとしても，$n'<n$ならば$\bar{Y}$が常にすぐれており，$n'>n$ならば0が常にすぐれていることになります（このことは，事後分布の平均が$n'<n$ならば$\bar{Y}$のほうに近く，$n'>n$ならば0に近くなるという事実から導かれます）．

これらの結果は，事前分布の密度関数の形に依存しています．さらに，それらは予備的な検定がいまの問題について誤った方向に目を向けていることを示唆しています．私たちは実際に，事前の判断を持とうと努力するでしょうし，またおそらく私たちは，この問題に直面し，それを予備的検定という裏口に隠しておくよりは，むしろ積極的に事前分布を付与すべきでしょう．

さまざまな欠陥にもかかわらず，予備的検定から導かれる推定値は直観に訴える一つの特徴を持っています．それは$|\bar{Y}|$がきわめて大きいとき，これは私たちの事前の見解がすべて誤っていることを示唆し，推定値は$\bar{Y}$に近くなるということです．推定値をちょうど$\bar{Y}$にすることはおそらく少し行過ぎでしょう．しかし事後分布の平均$\bar{Y}n/(n+n')$は反対の方向に行き過ぎるかもしれません．しかしながら，これは事前分布の使用に対する反論にはならず，上で用いた事前分布の型に対する反論となっているにすぎません．

ここでちょっと信頼区間の問題にもどって，予備的検定が分散の推定値に下方の偏りを与えることに注意しておきましょう[9]．しかしこの問題についても，私は最終的な解答を知りません．しかし，伝統的な方法に真に忠実な信頼区間を求めるのは途方もない問題のように思えます．予測についての問題もほとんど同様です．なぜなら，どの方法でもy_{n+1}の点予測値は対応するμの推定値と同一であり，予測の分散の推定値は単に$\hat{V}(\hat{\mu})+\hat{\sigma}^2$となるにすぎないからです．

問題が推定を行なうことであり，予備的検定で検定されるよりも，帰無仮説が誤りであることがわかっている以上，単純に無視しえないパラメータのすべてを推定すべきであり，予備的な検定はまったく行なうべきではないと議論したい気持に強く駆られます．不幸なことに，rのパラメータを含む通常の線型モデルの下で，もしn個の観測値から伝統的な方法により推定が行なわれるならば，平均の推定値の分散の観測値についての平均は$\sigma^2\frac{r}{n}$となり，予測子のそれは$\sigma^2(1+\frac{r}{n})$となります[10]．近似的にいえば，σ^2はrが増大するにつれて減少します（厳密にいえば，このモデルはrの一つの値についてだけ成り立ちます）．しかしながら，明らかにrをできるだけ小さくしておくことにはそれだけの根拠があります．これはとくに予測の場合には明瞭です．平均についての予測では，rの変化とともに推定される

8) もう少し複雑な状況では，われわれは，たとえば，通常の方法でデータから交互作用を推定するか，あるいはそれが0であると仮定するかのいずれかをあらかじめ選択することになろう．

9) とくに検定の基準が$F>1$であるときにそうである．以前の脚注を参照のこと．

10) 私は，この結果がCuthbert DanielおよびGeoffrey Watsonによっていると考えているが，適切な参考文献については知らない．

量の意味が変化します．

私には，やはりベイジアンの道がこのディレンマから逃れる最善の方法を与えてくれるように思えますが，私はここでもう一度，伝統的な一様の事前密度関数が基本的な問題については何も役立たないことを強調しておきたいと思います．それはこのような密度関数からは伝統的な方法とまったく同一の推定値および予測値が得られるからです．しかしながら，この話のはじめの部分で，一様な事前密度関数の場合ですら，ベイジアンの方法は非線型パラメータに対する推定および信頼区間の設定に役立つことが示されています．そして私にとって，ベイジアンの方法は，常に意味のある事柄を無意味な事柄から区別するのに役立っています．

注

a) 各クラスの得点の平均 $\mu_1, \cdots\cdots, \mu_7$，分散 σ^2

$$\theta^2=\frac{1}{6}\sum_{i=1}^{7}(\mu_i-\mu)^2,\quad \text{ここで}\ \mu=\frac{1}{7}\sum_{i=1}^{7}\mu_i$$

$\mathrm{SSW}\sim\sigma^2\chi^2_{658}$

$\mathrm{SSB}\sim\sigma^2\chi^2_6(\delta^2)$，ここで $\delta^2=95(6\theta^2)/\sigma^2$ は非心パラメータ

$\mathrm{SSW}=589,190, \mathrm{SSB}=26,840$（表 2 を見よ）

ここで $\sigma^2=\mathrm{MSW}=895$ とおけば，

$\delta^2=.637\theta^2, \mathrm{SSB}/\sigma^2=29.97$

尤度 $\propto f_{\chi}{}^2(29.97\,|\,\delta^2=.637\theta^2$, 自由度 6)

b) $H_0:\theta=0, P=.00004$

$$\frac{f_0(\mathrm{MSB})}{f_1(\mathrm{MSB})}=\frac{f_{\chi}{}^2(29.97\,|\,{}^2\delta=0,\ \text{自由度}\ 6)}{\int p_1(\theta)f_{\chi}{}^2(29.97\,|\,\delta^2=.637\theta^2,\ \text{自由度}6)}$$

$$\fallingdotseq\begin{cases}.00083,\ P_1=N(6,4)\text{のとき}\\ .00273,\ P_1=R(0,20)\text{のとき}\\ \qquad\qquad (R\text{は一様分布})\end{cases}$$

$$\text{尤度比}:\frac{f_0(\mathrm{MSB})}{\max f(\mathrm{MSB}\,|\,\theta)}=.00046=.55\times.00083$$

$H_0;\ \theta\leqq1, P=.00012$

$$\frac{f_0(\mathrm{MSB})}{f_1(\mathrm{MSB})}\fallingdotseq\begin{cases}.00166,\ P_1=N(6,4)\text{のとき}\\ .00546,\ P_1=R(0,20)\text{のとき}\end{cases}$$

i) この点を例示するため，次のような例をとり上げて議論することにしよう．いま $x_1, \cdots\cdots, x_n$ は互いに独立にすべて平均 μ，分散 σ^2 の正規分布に従うものとし，仮説 $H_0:\mu=0$ を検定する問題を考えていると仮定しよう．対立仮説に対する事前分布の密度関数を $g(\mu)$ で示し，$g(\mu)$ は連続かつ有界と仮定する．このとき H_1 に対する事後の尤度は

$$f_1=\left(\frac{1}{\sqrt{2\pi}}\right)^n\int\exp\left[-\frac{\sum(x_i-\theta)^2}{2}\right]\cdot g(\theta)d\theta$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\Sigma(x_i-\bar{x})^2/2}\int e^{-n(\theta-\bar{x})^2/2}g(\theta)d\theta$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\Sigma-(x_i-\bar{x})^2/2}\frac{1}{\sqrt{n}}$$

$$\times\int e^{-\frac{t^2}{2}}\cdot g\left(\bar{x}+\frac{t}{\sqrt{n}}\right)dt$$

$$\fallingdotseq\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} e^{-(\Sigma x_i-\bar{x})^2/2}\frac{1}{\sqrt{n}}g(\bar{x}).$$

そして H_0 に対する尤度は

$$f_0=\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp{}^{-[\Sigma x_i{}^2/2]}$$

となる．

したがって n が大であるとき，これらの尤度の比は近似的に

$$\frac{f_1}{f_0}\fallingdotseq\frac{1}{\sqrt{2\pi n}}e^{-n\bar{x}^2/2}\cdot g(\bar{x})=\frac{1}{\sqrt{n}}g(\bar{x})\phi(\sqrt{n}\bar{x})$$

ここで ϕ は標準正規分布の密度関数を示す．他方，伝統的な判定値 P_0 は

$$P_0=\int_{|\bar{x}|}^{\infty}\frac{1}{\sqrt{2\pi n}}e^{-nt^2/2}dt$$

$$=\int_{|\sqrt{n}\bar{x}|}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

である．したがって P_0 と事後の勝目の比は

$$\frac{P_0\cdot P'(H_1)f_1}{P'(H_0)f_0}\fallingdotseq\frac{P'(H_1)}{P'(\mathrm{H}_0)}\left(\frac{1-\Phi(\sqrt{n}|\bar{x}|)}{\phi(\sqrt{n}|\bar{x}|)}\right)\frac{g(\bar{x})}{\sqrt{n}}$$

固定された P_0 に対して，したがって固定された $\sqrt{n}|\bar{x}|$ に対して，これは $n^{-1/2}$ に比例しており，また固定された n と小さな P_1 に対してこれは近似的に

$$\frac{P(H_1)}{P(H_0)}\cdot\frac{g(\bar{x})}{\sqrt{n}}\cdot\frac{1}{\sqrt{n}|\bar{x}|}=\frac{P(H_1)}{P(H_0)}\cdot\frac{g(\bar{x})}{n|\bar{x}|}$$

に等しくなる．この値は $\bar{x}\to\infty$ のとき 0 に近づく．

ii) 事前分布の密度関数が特異なものでなければこのようなことは起こらない．一般に，x を観測値 θ をパラメータとする．任意の x に対して $S(x)$ をパラメータ空間におけるある“信頼領域”とする．そして

$$1-\alpha(\theta)=P^x(S(x)\ni\theta\,|\,\theta)$$

と定義すると，これを $S(x)$ の信頼係数と呼ぶことができる．このとき $S(x)$ の事後確率は

$$P(S\,|\,x)=\tilde{P}^{\theta}(S(x)\ni\theta\,|\,x)$$

と定義される．ここで$\tilde{P}^{\theta}$は，あるθの（特異でない）事前分布に関する，xが与えられたときのθの事後確率を示している．

このとき

$$E_0(1-\alpha(\theta))=E_x(P(S|x)=\tilde{P}(S(x)\ni\theta)$$

ここで$\tilde{P}$はxおよびθの同時確率を示す．したがって確率の支持平面がパラメータ空間全体を覆うことを前提とすれば

$$\inf_{\theta}(1-\alpha(\theta))\geqq\sup P(S|x)$$

iii)　この点はC. Steinの有名な例と密接に関連している．いま$x_{ij}(i=1,\cdots,k;j=1,\cdots,n)$が互いに独立して，各$i$について平均$\mu^i$，分散$\sigma^2$の正規分布に従っているとする．ここで$\mu_i$の同時推定を考え，$\sum_{i=1}^{k}E(\hat{\mu}_i-\mu_i)^2$を最小にしたいものとしよう．不偏推定量$\hat{\mu}_i=\bar{x}_i=\frac{1}{n}\sum_{\sigma=1}^{n}x_{ij}$; $i=1,\cdots,k$は事前分布の一様な密度関数$\prod_{i=1}^{k}d\mu_i$に対する事後分布の平均になっている．Steinは，もし$k\geqq3$なら$\hat{\mu}_i$が許容しえない推定量であることを示した．なぜならもし

$$\hat{\mu}^k{}_i=\left(1-\frac{n(k-2)}{\sum_{i=1}^{k}x_i{}^2}\cdot\frac{fs^2}{f+2}\right)\bar{x}_i$$

とおくと，すべてのμ_iに対して

$$\sum_{i=1}^{k}E(\hat{\mu}^k{}_i-\mu_i)^2<\sum_{i=1}^{k}(E\hat{\mu}_i-\mu_i)^2=\frac{R\sigma^2}{n}$$

となるからである．ここで

$$S^2=\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2/f_1$$

$f=k(n-1)$である．われわれは$\hat{\mu}^k{}_i$を

$$\hat{\mu}^k{}_i=\frac{1}{1+\frac{k-2}{k}\frac{f}{f+2}\frac{1}{F}}\bar{x}_i$$

でおきかえてもよい．ここで$F=n\sum_{i=1}^{k}\bar{x}^2{}_i/S^2$であるが，これは仮説$\mu_i=0$に対する$F$統計量である．この場合，平均2乗誤差の増大は無視しえない（竹内著「計量経済学の研究」を参照）．

iv)　非心χ^2分布の平均は$f+\delta^2$であり，分散は$2f+4\delta^2$である．もしCをある定数とし，χ^2を通常の自由度f'のχ^2分布として$c\chi'^2$で近似しようと思えば

$$cf'=f+\delta^2$$
$$c^2f'=f+2\delta^2$$

とする方法が考えられるが，このとき

$$c=\frac{f+2\delta^2}{f+\delta^2}$$

$$f'=\frac{(f+\delta^2)^2}{(f+\delta^2)}$$

を得る．この近似に基づく信頼下限は（δ^2に関する）方程式

$$c\chi^2{}_{\frac{\alpha}{2}}(f')=\chi^2{}_0$$

を解くことによって求められる．ここで$\chi^2{}_0$は観測値である．また同様にして信頼上限は，方程式

$$c\chi^2{}_{1-\frac{\alpha}{2}}(f')=\chi^2{}_0$$

から求められる．これらの方程式はくり返し法によって解くことができる．もっと単純な近似は，通常のχ（χ^2の平方根）を平均

$$\sqrt{f+\delta^2-\frac{f+2\delta^2}{2(f+\delta^2)}}\text{および分散}\frac{(f+2\delta^2)}{2(f+2\delta^2)}\text{の}$$

正規分布に近似することによって求められる．

この近似から

$$\chi^2{}_0=\sqrt{f+\delta^2-\frac{f+2\delta^2}{2(f+\delta^2)}}=k\frac{1}{2}\sqrt{\frac{f+2\delta^2}{2(f+\delta^2)}}\text{を}$$

解いてδ^2に対する信頼区間が求められる．いまの例ではθ^2に対する信頼区間は$\theta^2=\sigma^2\delta^2/(95\times6)$として得られ，$\theta$に対する95％信頼区間は(6.9, 9.1)として求められる．

〔注の記号についての注意〕

記号　1), 2), ……, 10)は，Pratt氏の原稿にあった注である．

記号　a), b)は，昭和47年10月30日，OR学会統計研究部会での講演のレジュメから再録したものである．

記号　i), ii), iii), iv)は，竹内　啓氏による訳注である．

参　考　文　献

Berkson, Joseph, *JASA*, **33**(1938), 526-536.

Efron Bradley, *JASA*, **66**(1971), 552-568(討論を含む).

Good, I. J., *Probability and the Weighing of Evidence* (1950).

———, *JASA*, **53**(1958), 799-813.

———, *IRSS, Series(B)*, **27**(1965), 196-197.

———, *JASA*, **64**(1969), 23–66(討論を含む).
Kruksal, William H., "Significance, Test of" の項目, International Encyclopedia of Social Science(1968).
Jeffreys, Harold, *Theory of Probabilty*(1961 およびそれ以前の版).
Lindley, Dennis V., *Brometika*, **44**(1957), 187–192.
Peizer, David B. and Pratt, John W., *JASA*, **63** (1963)
Pratt, John W., *JRSS, Series* (*B*), **27**(1965), 166–203 (討論を含む).
Raiffa, Howard and Schlaifer, Robert, *Applied Statistical Decision Theory*(1961).
Scheffé, Henry, *The Analysis of Variance*(1959).
Thompson, James R., *JASA*, **63**(1968).
Zellner, Arnold, *Technical Report*(1972).