論　説

ベイズ統計学の原理と応用（その 1）

美　添　泰　人

1　はじめに

21 世紀に入ってから，ベイズ統計学に関する書物が急速に増加している。筆者がベイズ統計学を学ぶためにアメリカの大学院に留学した 1970 年代には，ベイズ統計学はアメリカでも少数派だったことに比べて，まさに隔世の感がある。一方，最近の「ベイズ統計」という名称を使っている著書や論文の中には「どの点がベイズの議論なのか」,「どこに統計学の考え方が示されているのか」など，疑問に感じる例が増えていることも事実である。たとえば［13］では根拠が明示されない事前分布を想定して，手続き的に事後分布を計算することをベイズ統計学の理論と定義している。これは，合理的な意思決定から導かれる事前分布とは無関係で，形式的には単なる罰金つき尤度に関する数値計算の技術としてとらえるもので，筆者が教育を受けたベイズ統計学とはまったく異なる視点である。

しばらく前までは，筆者もベイズ統計学の最近の動向を紹介することを考えていたものの，この 10 年ほどの膨大な量の出版物の多くを見ると，このことは現実的ではなくなり，計画を断念した。その代わりに，ベイズ統計学の基礎的な原理に立って，その応用と古典的な統計学との関係を明らかにすることに限定し，その試みのひとつとして，本稿では，特に John W. Pratt の比較的初期の論文［54］の内容を紹介する。ここで Pratt は，ベイズ統計学の視点から古典的な統計学を擁護できる部分と，批判すべき部分を指摘している。Pratt の議論が統計学者に与えた影響は大きく，［54］が報告された Royal Statistical Society で座長を務め，当時は比較的穏健なベイズ統計学の立場に立って［47］

を執筆していた Dennis V. Lindley は，その後，Pratt の所属していた Harvard University に Visitor として滞在していた間に過激ともいうべきベイズ統計学者に変身して［48］を提示したことも，その影響を示している。

ベイズ統計学の基本原理については，概要を美添［12］に記したが，本稿でも参照するため，その主要部分を補論 A として収録する。より詳しい議論と初期の応用例は美添［2］に収録されている。

Leonard Jimmy Savage を中心とするベイジアンたちは，従来，科学的とはみなされず批判の対象とされてきたベイジアンの手法を，1950 年代から 1970 年代にかけて正当化した。筆者は，Savage および同じ立場に立つ Pratt の視点を受け入れている。このうち，Savage は筆者が留学する直前に若くして亡くなったため，直接指導を受ける機会はなかったが，Savage の後を継ぐ形で指導的な立場にあった Pratt は筆者の博士論文の指導者である。

2　Pratt の議論

本節では，ベイズ統計学の立場から伝統的統計学の手法を解釈しようとした［54］の内容を，できるかぎり忠実に紹介する。［54］は §1. Introduction から §9. Miscellaneous Remarks までの 9 つの節で構成され，その後に 12 ページにわたって Discussion が収録されている。今後の参照の利便性を考慮して，以下では小節の番号および表題を［54］の番号に合わせて明示する。Pratt の証明は簡潔であり，ときとして難解である。そのため，かなりの部分で詳細な推論の過程を追記しているが，美添が補足した部分は必ずしも明示していない。いずれにせよ，本稿のうち，本質的な記述の大部分は Pratt によるものである。

2.1　非十分統計量

§1. Introduction では，この論文のねらいとして，標準的な統計的手法（Standard Inference Statements）に対するベイジアンの視点からの解釈を議論することが記されている。本稿でも，ベイズ統計学の立場に立つ人たちをベイジアンと総称するが，その中にはさまざまな違いがあり，すでに述べたとおり，Pratt は

Savage の主観確率の立場で議論を展開する。また標準的という表現については，それが特定の手法や特定の個人の実際的または理論的，哲学的な手法を意味するものではないことを注意している。それにもかかわらず，Pratt は最も広く利用されている手法で，伝統的 “orthodox”，古典的 “classical”，客観的 “objective”，頻度論的 “frequency”，または “Neyman-Pearson” などと表現される伝統的な手法またはそれをわずかに修正したものを対象としている。

§2 では非十分統計量（insufficient statistics）を扱う。標準的な分析では，もっとも単純な問題を除けば，用いられる統計量，たとえば典型的な t, F, χ^2 は十分統計量ではない。その他，母集団分布が正規分布でない場合の標本平均と分散，ノンパラメトリック問題における標本中位数，決定係数 R^2，連関係数も同様である。

全面的にベイズの手法を利用する完全ベイズ分析（full Bayesian analysis）では，典型的には十分統計量を利用する。実際，母数空間上で事前分布が常に正であれば，すべての観測値情報を集約した統計量が事後分布を定めるなら，それは十分統計量である。

以下は，この点に関する筆者の解説である。母数を θ とし，観測値 $\{x_1, \ldots, x_n\}$ の確率密度関数を $p(x_1, \ldots, x_n \mid \theta)$ と書くと，統計量 t が十分のとき，密度関数は $p(x_1, \ldots, x_n \mid \theta) = p(t \mid \theta)\, f\,(x_1, \ldots, x_n)$ と分解される。θ の事後分布では x のみを含む関数は定数だから，$p(\theta \mid x_1, \ldots, x_n) \propto p(\theta)p(x_1, \ldots, x_n \mid \theta) \propto p(\theta)\, p(t \mid \theta)$ と表されることは，θ に関して $p(x_1, \ldots, x_n \mid \theta) \propto p(t \mid \theta)$ を意味する。すなわち t は十分統計量である。なお美添による補足では，密度関数は $p(x)$, $p(\theta), p(t)$ などの記号を用いる。

非十分統計量を理解するために，完全ベイズ分析を次の 2 段階に分解する。以下で観測値を x，統計量を t，母数を θ と表すが，一般にはこれらはベクトルである。（1）統計量 t にもとづいて事後分布を計算する，（2）この結果を事前分布として，t を固定したときのすべての観測値の条件付分布を用いて事後分布を求める。Pratt の表現では第 1 段階は

$$f_1(\theta) = f(\theta \mid t) = \frac{f_0(\theta) f(t \mid \theta)}{f(t)} \tag{2.1}$$

第 2 段階は

$$f_2(\theta) = f(\theta \mid x) = \frac{f_1(\theta) f(x \mid t, \theta)}{f(x \mid t)} \tag{2.2}$$

である。f_1, f_2 の意味は自明であろう。

この第 2 段階は多くの場合極めて面倒だが，結論を大きく変えないことも多い。そのため，ベイジアンにとっては非十分統計量が与えられたときの事後分布の利用が示唆される。実際 Pratt は，後述の定理 3 および定理 6 で，この議論が成立する条件を考察している。

ここで，このような議論は，χ^2 のように，ある仮説の下で分布を容易に導出できる統計量の利用を正当化するものではないことを指摘し，χ^2 の例では，母数に関する正しくない推論および多くの情報の損失につながる可能性があると記しているが，具体的な内容は明示されていない。

非十分統計量の利用に関する簡単な例は次のとおりである。x を，平均を θ とする母集団からの大きな標本とする。また，母集団分布の形は未知だが分散 σ^2 は既知とする。ここで t を標本平均とすると，t は近似的に正規分布 $N(\theta, \sigma^2/n)$ にしたがうから，t が与えられたときの θ の事後分布は，次の式で与えられる。ただし K は母数に依存しない定数で，$\simeq$ は近似を意味する。

$$f(\theta \mid t) \simeq K f_0(\theta) \exp\{-\tfrac{1}{2} n(\theta - t)^2/\sigma^2\} \tag{2.3}$$

この表現では事前分布 $f_0(\theta)$ だけが計算に必要である。これに対してすべての情報 x を利用して事後分布 $f(\theta \mid x)$ を評価する場合には，θ と分布形を定める母数 η の同時事後分布を求め，さらに η に関する積分を適用する必要がある。η は一般に多次元で，ノンパラメトリックの場合は無限次元だから，適切な事前分布を構築することは容易ではない。一方，σ^2 を既知とできる場合は，標本平均以外のデータが θ の分布に与える影響は非常に小さいことが予想される。

§2.1 Further Conditioning in Connection with the Mean では，σ^2 が未知の場合に

は，標本平均と標本分散を用いて $t=(\bar{x}, s^2)$ とすること，さらに一般的に統計量を $t=(\bar{x}, s^2, t_1, \ldots, t_r)$ とする場合を考察している。

ここでは，もっとも簡単な場合の議論を紹介する。$t=(\bar{x}, s^2)$ が与えられたときは，もし θ, σ^2 および母集団の高次モーメントに関する事前分布が σ^2 とともに緩やかに変化する場合は，$\sigma^2=s^2$ であれば，(2.3) と同じ結果を与えることが，以下のように確かめられる。新たな変数として $u_1=n^{1/2}(\bar{x}-\theta)/s$, $u_2=n^{1/2}(s^2-\sigma^2)/s^2$ を導入すると，u_1, u_2 の分布は近似的に正規分布にしたがい，期待値は 0，分散と共分散は $\mathrm{var}(u_1)=1, \mathrm{var}(u_2)=\lambda_4$ および $\mathrm{cov}(u_1, u_2)=\lambda_3$ となり，その密度関数は次式で与えられる。

$$\frac{1}{2\pi}(\lambda_4-\lambda_3^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(\lambda_4 u_1^2-2\lambda_3 u_1 u_2+u_2^2)/(\lambda_4-\lambda_3^2)\} \qquad (2.4)$$

ただし λ_3, λ_4 は n に依存しない母数であり，このとき，3 次モーメントは $\lambda_3\sigma^2$，4 次モーメントは $\lambda_4\sigma^4+\sigma^4$ である。

ここで (u_1, u_2) は，θ, σ^2 が与えられたときは $(\bar{x}, s^2)$ と 1 対 1 対応であり，また $(\bar{x}, s^2)$ が与えられたときは (θ, σ^2) と 1 対 1 対応であることに注意する。$(\bar{x}, s^2)$ が与えられたときの $(u_1, u_2, \lambda_3, \lambda_4)$ の同時分布は，近似的に (2.4) で与えられる $f(u_1, u_2 \mid \lambda_3, \lambda_4)$ と次の式の積となる。

$$K(1-n^{-1/2}u_2)f_0(\bar{x}-n^{-1/2}su_1, s^2-n^{-1/2}s^2u_2, \lambda_3, \lambda_4) \qquad (2.5)$$

ただし $f_0(\theta, \sigma^2, \lambda_3, \lambda_4)$ は同時事前分布である。

このことは次のように確かめられる。条件つき分布 $f(u_1, u_2, \lambda_3, \lambda_4 \mid \bar{x}, s^2)$ は，母数に関する事後分布 $f(\theta, \sigma^2, \lambda_3, \lambda_4 \mid \bar{x}, s^2) \propto f_0(\theta, \sigma^2, \lambda_3, \lambda_4)\, f\,(\bar{x}, s^2 \mid \theta, \sigma^2, \lambda_3, \lambda_4)$ から変数変換によって導かれる。$J=D(u_1, u_2)/D(\bar{x}, s^2)$ を Jacobian とすると

$$f(\bar{x}, s^2 \mid \theta, \sigma^2, \lambda_3, \lambda_4)=f(u_1, u_2 \mid \theta, \sigma^2, \lambda_3, \lambda_4)\frac{D(u_1, u_2)}{D(\bar{x}, s^2)}$$

であり，右辺は (2.4) 式と $J=n\sigma^2/s^5 \propto \sigma^2/s^2=(1-n^{-1/2}u_2)$ の積に等しい。また $f_0(\theta, \sigma^2, \lambda_3, \lambda_4)$ の部分は (θ, σ^2) から (u_1, u_2) への変数変換によって書き換えれば (2.5) の表現に一致する。

$(\bar{x}, s^2)$ が与えられたときには，（2.5）式から $u_2, \lambda_3, \lambda_4$ を積分によって消去すれば u_1，したがって θ の事後分布が求められる。もし事前分布 $f_0(\theta, \sigma^2, \lambda_3, \lambda_4)$ が σ^2 とともに緩やかに変化するなら，（2.5）式は近似的に次のように表される。

$$Kf_0(\bar{x} - n^{-1/2}su_1, s^2, \lambda_3, \lambda_4) \tag{2.6}$$

（2.4）と（2.6）の積を，まず σ^2 について積分すると，正規分布の性質から，次の形が導かれる。

$$f(u_1, \lambda_3, \lambda_4 \mid \bar{x}, s^2) \simeq Kf_0(\bar{x} - n^{-1/2}su_1, s^2, \lambda_3, \lambda_4)e^{-(1/2)u_1^2} \tag{2.7}$$

$f_0(\theta, \sigma^2, \lambda_3, \lambda_4)$ を (λ_3, λ_4) について積分して得られる (θ, σ^2) の周辺事前分布を f_{θ,σ^2} と表すと，次式が得られる。

$$f(u_1 \mid \bar{x}, s^2) \simeq Kf_{\theta,\sigma^2}(\bar{x} - n^{-1/2}su_1, s^2)e^{-(1/2)u_1^2} \tag{2.8}$$

さらに $\sigma^2 = s^2$ として σ^2 の周辺事前分布で割ると，（2.8）の f_{θ,σ^2} は条件付分布で置き換えられるから，θ の事後分布は次のようになる。

$$f(\theta \mid \bar{x}, s^2) \simeq Kf(\theta \mid \sigma^2 = s^2) \exp\{-\tfrac{1}{2}n(\theta - \bar{x})^2/s^2\} \tag{2.9}$$

これは，σ^2 が既知のときの（2.3）で $t = \bar{x}, \sigma^2 = s^2$ とおいた式と一致する。

なお，Pratt は次の点を注意している。上記の近似の範囲では θ の事後分布として t 分布と正規分布にはほとんど違いはない。標本が小さい場合は (λ_3, λ_4) が正規分布に近いという事前分布ないし事後分布が得られる場合に限って，θ の事後分布が t 分布となることが導かれるが，この条件は必ずしも成立しない。

また，$t = (\bar{x}, s^2)$ が与えられたときの θ の近似的な事後分布を導出するために，t の標本分布を定める母数に影響を与える $(\theta, \sigma^2, \lambda_3, \lambda_4)$ の詳細な事前分布を定める必要はなく，θ の事前分布を除けば σ^2 とともに緩やかに変化するというだけの仮定があればよいことも指摘している。

さらに多数の統計量 $t = (\bar{x}, s^2, t_3, \ldots, t_r)$ が得られたときの事後分布を考えることもできる。このときは，上記の u_1, u_2 に加えて $u_i = \sqrt{n}(t_i - \tau_i)$ $(i = 3, \ldots, r)$

が，近似的に期待値ベクトル 0，共分散行列 $\Lambda = (\lambda_{ij})$ をもつ多変量正規分布にしたがうことを仮定する。ただし τ_i, λ_{ij} は母集団特性から導かれる母数である。特に $(u_1, u_2) = (t_1, t_2)$ に対応する母数は $\lambda_{11} = 1, \lambda_{12} = \lambda_3, \lambda_{22} = \lambda_4$ である。
ここで次の定理がなりたつ。

定理 1. t が与えられたときの θ の密度関数は，以下に記す条件の下で次式で与えられる。

$$f(\theta \mid t) \simeq Kf(\theta \mid \sigma^2 = s^2, \tau_3 = t_3, \ldots, \tau_r = t_r) \exp\{-\tfrac{1}{2} n(\theta - \bar{x})^2 / s^2\} \quad (2.10)$$

ここで以下の条件を仮定する。ただし λ は Λ をベクトルとみなした表記である。

(a) $\theta, \sigma^2, \tau_3, \ldots, \tau_r$ と λ の成分は関数的に独立である。

(b) λ_{1j} は λ, θ および $\tau_{j+1}, \ldots, \tau_r$ $(2 \le j \le r)$ の関数である。

(c) λ_{ij} は λ, θ および $\tau_{i+1}, \ldots, \tau_r$ $(2 \le i \le j \le r)$ の関数である。

(d) 母数 $(\theta, \sigma^2, \tau_3, \ldots, \tau_r, \lambda)$ の同時事前分布は $\sigma^2, \tau_3, \ldots, \tau_r$ の変動に関して $n^{-1/2}$ のオーダーで近似的に一定である。

なお，母数 $\theta, \sigma^2, \tau_3, \ldots, \tau_r$ とすべての λ_{ij} との関数的な独立性を排除しない理由は，たとえば t_3, t_4 がそれぞれ $\lambda_{12}, \lambda_{22}$ のときには $\lambda_{12} = \tau_3, \lambda_{22} = \tau_4$ となるためである。

この定理の証明は，上記の $t = (\bar{x}, s^2)$ 場合の拡張であり，詳細は省略する。

2.2 推定

観測値 x の分布を定める指標を，ノンパラメトリックな分布族も含めて，ω と表す。母数 $\tau = \tau(\omega)$ に関して，統計量 $t = t(x)$ が

$$E(t \mid \omega) = \tau(\omega) \quad \text{for all } \omega \quad (2.11)$$

を満たすとき，t は不偏推定量と呼ばれる。これに対応するベイジアンの自然な概念は事後平均 $E(\tau \mid x)$，または統計量を利用した $E(\tau \mid t)$ である。ここで不偏推定量と事後平均について，次のような解釈がなりたつ。

定理 2. t が $\tau(\omega)$ の不偏推定量であれば，次式がなりたつ。ただし，ベイジアンではすべての統計量と母数は確率変数だから，期待値および確率分布はそ

れらの同時分布に関するものである。

$$E\{t - E(\tau \mid x)\} = 0 \tag{2.12}$$

証明　まず事前分布による不偏統計量の期待値は

$$E\{t\} = \int \left\{ \int t p(x \mid \tau) dx \right\} p(\tau) d\tau = \int \tau p(\tau) d\tau = E(\tau)$$

と事前分布の期待値と一致する。一方，事後平均の期待値は

$$\begin{aligned} E\{E(\tau \mid x)\} &= \int \left\{ \int \tau p(\tau \mid x) d\tau \right\} p(x) dx = \iint \tau p(\tau, x) d\tau dx \\ &= \int \tau \left\{ \int p(x \mid \tau) dx \right\} p(\tau) d\tau = \int \tau p(\tau) d\tau = E(\tau) \end{aligned}$$

だから，事前分布による t の期待値と一致する。　■

この定理は，いずれも x の関数である t および $E(\tau \mid x)$ について，それらの差の（事前の）期待値がゼロとなることを主張している。ところで，以上の議論は任意の統計量 y から得られる事後平均 $E(\tau \mid y)$ にも適用することができるから，どのような y による近似が優れているかが問題となる。すでに t が得られている状態では，$t = t(y)$ となるような y を考察するのが自然である。このような候補の中では $y = t$ が，次の定理の意味で，最もよい近似を与える。

定理 3.　t が y の関数となるとき，任意の凸関数 ψ に対して次の式がなりたつ。

$$E[\psi\{t - E(\tau \mid y)\}] \geq E[\psi\{t - E(\tau \mid t)\}] \tag{2.13}$$

証明　Jensen の不等式と t が y の関数という条件から次の不等式が導かれる。

$$E[\psi\{t - E(\tau \mid y) \mid t\}] \geq \psi[E\{t - E(\tau \mid y) \mid t\}] = \psi[t - E(\tau \mid t)] \tag{2.14}$$

この議論をていねいに記せば次のようになる。

$$E\{E(\tau \mid y) \mid t\} = \int \left\{ \int \tau p(\tau \mid y) d\tau \right\} p(y \mid t) dy = \iint \tau \frac{p(\tau, y)}{p(y)} \frac{p(y, t)}{p(t)} d\tau dy$$

ここで t は y の関数だから $p(\tau, y) = p(\tau, y, t)$ および $p(y) = p(y,\ t)$ となることを用いると，この式は

$$E\{E(\tau \mid y) \mid t\} = \iint \tau \frac{p(\tau, y, t)}{p(t)} d\tau dy = \int \tau \frac{p(\tau, t)}{p(t)} d\tau = E(\tau \mid t)$$

となる。(2.14) の両辺の期待値を取れば (2.13) が導かれる。 ■

以上のように，不偏推定量 t は事後平均 $E(\tau \mid x)$ または $E(\tau \mid t)$ の近似とみなされる。なお，t の誤差の平均が 0 になるという，ω あるいは τ の条件付分布に関して成立する性質は，x または t の条件付分布に関しては成立しない。不偏性とは同じ分布から発生する観測を繰り返したときの性質であり，手元の観測値を固定したときの性質ではないことは，不偏性に関して注意しなければならない点である。ただし，この定理を次のように書き換えると，多くの人が直観的に考えるように，与えられた x から求められた不偏推定量の誤差は，少なくとも事前の判断としては，近似的に 0 となる。

$$E\{E(t - \tau \mid x)\} = 0 \tag{2.15}$$

t を τ の不偏推定量とするとき，一般に t の関数 $g(t)$ は $g(\tau)$ の不偏推定量とはならない。伝統的な手法では，この性質は難点とされることがあるが，本節の議論を踏まえれば，この性質は当然といえる。すなわち，t は $E(\tau \mid x)$ の近似だから，$g(t)$ は $g\{E(\tau \mid x)\}$ の近似と考えてもよいが，これは一般には $E\{g(\tau \mid x)\}$ とは一致しない。ベイジアンの立場からは，そもそも t を用いるか $g(t)$ を用いるかは，効用関数（損失関数）など，確率モデル以外の要因に依存する。これらの外部的な要因を無視した「最適な統計的手法」は，何らかの意味で必然的に不十分になると，Pratt は批判している。

ここからは，一致性に関する議論である。標本サイズ n を明示した統計量 t_n が（弱い意味の）一致性を持つ条件は次のとおりである。

$$P(|t_n - \tau| > \varepsilon \mid \omega) \to 0 \quad (n \to \infty) \quad \text{for all } \varepsilon \text{ and } \omega \tag{2.16}$$

一致性に対応するベイジアンの性質として，Pratt は次の定理を示している。以下で x_n は大きさ n の標本を表す。

定理 4. t_n が τ の一致推定量であれば，次の式がなりたつ。

$$P\{P(|t_n - \tau| > \varepsilon \mid x_n) > \delta\} \to 0 \quad (n \to \infty) \quad \text{for all } \varepsilon \text{ and } \delta \tag{2.17}$$

証明　条件付確率の性質から，次がなりたつ。

$$E\{P(|t_n - \tau| > \varepsilon \mid x_n)\} = P(|t_n - \tau| > \varepsilon) = E\{P(|t_n - \tau| > \varepsilon \mid \omega)\} \tag{2.18}$$

この式の右辺は（2.16）と Lebesgue 収束定理によって 0 に近づくが，もし（2.17）が成立しなければ（2.18）は 0 に近づくことがない。すなわち（2.17）が成立する。　■

この定理は，大きな n に対して典型的な状況の下では τ の事後分布は t_n の近くに集中することを示している。ただし，これは特定の x_n に対して保証される性質ではない。

2.3　信頼領域

母数 $\tau = \tau(\omega)$ の信頼領域 $R = R(x)$ は次の条件をみたす領域として定義される。

$$P(\tau \in R \mid \omega) \geq 1 - \alpha \quad \text{for all } \omega \tag{2.19}$$

ここで $1 - \alpha$ は信頼係数だが Pratt は “conservative level” と呼んでいる。（2.19）の不等号を等号で置き換えたものが正確な水準 (exact confidence level)$1 - \alpha$ の信頼領域である。この概念に対応する自然なベイジアンの概念は事後確率 $P(\tau \in R \mid x)$ である。

任意の領域 $R = R(x)$ に対して，次の関係がなりたつ。

$$E\{P(\tau \in R \mid \omega)\} = P(\tau \in R) = E\{P(\tau \in R \mid x)\} \tag{2.20}$$

この式から，ただちに次の定理が導かれる（Pratt［53］）。

定理 5.　τ の信頼領域 R の水準が $1 - \alpha$ であれば，次の式がなりたつ。

$$E\{P(\tau \in R \mid x)\} \geq 1 - \alpha$$
$$E\{P(\tau \in R \mid x)\} = 1 - \alpha \qquad \text{水準が正確なとき} \tag{2.21}$$

この式は R が τ を含む事後確率（x の関数）は，期待値として正確に $1-\alpha$ となることを示している。したがって典型的な状況においては，x を観測した後でも，信頼領域が τ を含む事後確率は近似的に $1-\alpha$ とみなせる。

ところで，§3 における議論のように，この議論は任意の統計量 y に関しても考えることができる。もし y が全く母数に関する情報を含まない無意味な統計量であれば，当然，$P(\tau \in R \mid y) = P(\tau \in R) = 1-\alpha$ と正確な水準を与える。しかし実際に R が与えられる場合には，y の関数となるような R（またはその 1 対 1 変換）のうちで最もよい近似を与える y は何かという問題が考えられる。その答えは次の定理で与えられる。証明は定理 3 と同様である。

定理 6. 水準 $1-\alpha$ をもつ τ の信頼領域 R が y の関数であれば，任意の凸関数 ψ に対して，次の式がなりたつ。

$$E[\psi\{1-\alpha-P(\tau \in R \mid y)\}] \geq E[\psi\{1-\alpha-P(\tau \in R \mid R)\}] \tag{2.22}$$

1 次元の母数 τ について信頼区間の上限と下限を考えると，これらは τ の事後分布の分位点に対応する。Pratt は，この意味で信頼区間の累積分布（cdf）が事後分布の cdf を近似すると主張するが，その正確な意味は後に紹介する定理 7 で明らかにされる。

伝統的な手法では，信頼区間が全区間（$-\infty, \infty$）となったり空集合となるような病的な例がある。信頼区間が事後確率の近似だとすると，このような事例はベイジアンとしても好ましくはないが，Pratt は，そもそも伝統的な手法に対して根本的なベイジアンからの正当化はできないこと，病的な現象が発生することは手法の原理的な欠陥ではないと指摘している。ただし，後半の指摘に関してはもう少し検討の余地があると，筆者は考えている。

もうひとつの問題として，伝統的な手法における水準 $1-\alpha$ の決定手順が指摘されている。たとえば母数が $\tau(\theta) \leq 0$ となる区間に関心があるとき，上記の病的な状況が発生しなければ，0 が信頼区間の上限となるような水準を定めることができる。この場合の水準は観測値に依存して $1-\alpha(x)$ と表されるが，確率の頻度論的な立場では $\alpha(x)$ の意味は解釈できない。伝統的な手法では，

事前に水準 α_0 を定めて，結果として $\alpha(x)=\alpha_0$ となったときには，$\tau(\theta)\leq 0$ の水準が $1-\alpha_0$ とはいえるが，標本に依存して $\alpha_0=\alpha(x)$ となった場合には，$\tau(\theta)\leq 0$ の水準が $1-\alpha_0$ とはいえない。これらの問題は，信頼水準は根本的な信頼の尺度ではないことを示していると，Pratt は主張する。ベイジアンとしては，信頼水準は事後確率の近似と理解しておけば，根本的な問題として当惑することはない。定理 5 の意味での近似という性質から，ベイジアンでは自由に水準を定めることができる。ただし，事前に α を定める方が全体的により正確になる。

離散的な確率変数についても，伝統的な検定には 2 つの難点がある。ひとつ目の問題の例は二項分布の母数 p に関する通常の信頼区間であり，このときには被覆確率は p の不連続関数となるから，保守的な信頼区間を構成することが多い。例として $n=20$ の場合，$\alpha=5\%$ とする標準的な信頼区間が p を含まない確率は 4.2% を超えない（また p の範囲の 1/5 において 3.5% を超えない）。

ふたつ目の問題は，母集団中央値または分位点に関する信頼区間を構成する場合に生じる。たとえば $n=20$ の標本を用いて順序統計量を中央値の信頼限界とすると，水準は 1.000, 0.999, 0.994, 0.979, 0.942, 0.868, 0.748, 0.588,. . . となる。補足するとこれらの数値の根拠は説明されていないが，二項分布 $B(20, 0.5)$ の累積確率の「一部」である。そもそも，一般的に水準は離散的な数値しかとれないことは明らかである。

これらふたつの難点は，確率化信頼区間（randomized confidence interval）を利用すれば解決できるが，確率化信頼区間については Pratt は “. . . no one seriously suggests doing so in practice.” と一蹴している。

信頼区間の上限または下限と事後分布の関係として，それぞれの cdf が対応することについては，Pratt による二項分布の例を紹介する前に，もっと簡単な例を提示しよう。標本分布を $x_i\sim N(\theta,\sigma^2)$（ただし σ^2 は既知）とすると，十分統計量 $t=\bar{x}$ を用いて構成する標準的な信頼区間の上限は，z_0 を上側 α% 点として，$\bar{\theta}=t+z_0\sigma/\sqrt{n}$ で与えられる。このとき

$$P(\theta < \bar{\theta} \mid \theta) = P\{\sqrt{n}(t-\theta)/\sigma > -z_0 \mid \theta\} = P(z > -z_0) = 1-\alpha$$

が得られる。ここで $z = \sqrt{n}(t-\theta)/\sigma \sim N(0,1)$ の分布は θ に依存しないことが用いられている。一方 θ の事前分布を $N(\theta_0, \sigma_0^2)$ とすると，事後分布は $N(\theta_1, \sigma_1^2)$ となることは容易に導かれる。ただし，$1/\sigma_1^2 = 1/\sigma_0^2 + 1/(\sigma^2/n)$, $\theta_1 = \{\theta_0/\sigma_0^2 + t/(\sigma^2/n)\}\sigma_1^2$ である。特に事前分布が散漫な $p(\theta) \propto const.$ には $1/\sigma_0^2 = 0$ が対応する。このときの事後分布は $\theta \mid x \sim N(t, \sigma^2/n)$ となり，事後確率は

$$P(\theta < \bar{\theta} \mid x) = P\{\sqrt{n}(\theta-t)/\sigma < z_0 \mid x\} = P(z < z_0) = 1-\alpha$$

と信頼水準に一致する。信頼区間の下限 $\underline{\theta} = t - z_0\sigma/\sqrt{n}$ についても，散漫な事前分布から得られる事後確率が正確に対応する。これが，Pratt が指摘する cdf の対応の意味である。

二項分布に関する cdf の対応は，次の定理で正確に表現される。そこでは二項分布 $B(n,p)$ の母数 p に関する散漫な事前分布として，$f(p) = 1/p$ および $f(p) = 1/(1-p)$ が現れる。いずれも区間 $[0,1]$ における積分が発散するという意味で improper prior distribution の例である。なお，積分が存在して 1 となるときは真の事前分布（proper prior）と呼ぶ。

定理 7. 二項分布 $B(n,p)$ にしたがう観測値が与えられたとき，p の事後 cdf が信頼区間上限 cdf に一致する必要十分条件は事前分布が $f(p) = 1/(1-p)$ となること，また下限が一致する必要十分条件は事前分布が $f(p) = 1/p$ となることである。さらに $(1-p)f(p)$ が単調非増加かつ $pf(p)$ が単調非減少関数であれば，p の事後 cdf はこれらの中間にある。逆に，真の事前分布 $f(p)$ が $0 < p < 1$ で微分可能であり，任意の n と観測値に対して p の事後 cdf が信頼区間上限と下限の cdf の中間にあるならば，$(1-p)f(p)$ は単調非増加かつ $pf(p)$ は単調非減少関数である。

証明　事前分布を $f(p) = 1/(1-p)$ として，n 回の試行で r 回の成功が観測されたとき，ベータ分布と二項分布の関係から次式が成立する（詳細は補論 B を参照）。

$$P(p \le x \mid r) = \int_0^x p^r(1-p)^{n-r-1}dp/B(r+1, n-r) \\ = \sum_{r+1}^{n} \binom{n}{j} x^j(1-x)^{n-j} \tag{2.23}$$

この式は，x が p の（2.23）の水準に対する信頼上限（just conservative upper confidence limit）であることを示しているから上限に関する命題がなりたつ。下限については同様な議論，または対称性から導かれる。定理の残りの部分は次の補題から導かれる。　■

補題．任意の 1 次元母数に関する事前 cdf を F, G，その密度関数を f, g とし，対応する事後 cdf を F_1, G_1 とするとき，f/g が非減少であれば $F \le G$ かつ $F_1 \le G_1$ となる。二項分布については逆の命題も成立する。すなわち，f/g が微分可能かつ任意の n と観測値に対して $F_1 \le G_1$ であれば f/g は非減少である。

証明　f/g が非減少なら，次の不等式がなりたつから $F(x) \le G(x)$ となる。

$$\frac{F(x)}{1-F(x)} = \frac{\int_{-\infty}^{x}(f/g)g}{\int_{x}^{\infty}(f/g)g} \le \frac{\{f(x)/g(x)\}\int_{-\infty}^{x} g}{\{f(x)/g(x)\}\int_{x}^{\infty} g} = \frac{G(x)}{1-G(x)} \tag{2.24}$$

事後密度関数の比 $f_1/g_1 \propto f/g$ も非減少だから，$F_1(x) \le G_1(x)$ である。二項分布に関して逆を示すために，f/g が非減少ではないと仮定する。このとき，f/g の微分係数が負となる区間 (a, b) が存在して，ここで f/g は減少関数となる。$x \in (a, b)$ として，r/n が区間 (a, b) 内のある値に近づくように $r, n \to \infty$ とすると，この区間の外では尤度関数 L は無視できるから，先ほどと同じ議論で次の不等式が導かれる。

$$\frac{F_1(x)}{1-F_1(x)} = \frac{\int_a^x fL + \mathrm{rem}}{\int_x^b fL + \mathrm{rem}} > \frac{\int_a^x gL + \mathrm{rem}}{\int_x^b gL + \mathrm{rem}} = \frac{G_1(x) + \mathrm{rem}}{1-G_1(x) + \mathrm{rem}} \tag{2.25}$$

ここで rem は剰余項である。この不等式は常に $F_1 \leq G_1$ が成立するという仮定と矛盾する。 ■

以上の論理はやや難解なので，若干の説明を補足しておく。事前分布 $f(p)$ と $g(p) = 1/(1-p)$ を考えて，$f/g = (1-p)f$ は単調非増加と仮定する。このとき，補題から，f, g に対応する分布関数について $F \geq G$ がなりたつ。さらに f から得られる事後 cdf と g から得られる事後 cdf についても $F_1 \geq G_1$ がなりたつ。このことは観測値 r から（2.23）によって定められる確率 $(1-\alpha)$ に対応する $x = x(r)$ および G から得られる事後分布の $(1-\alpha)$ 分位点より，f から得られる事後分布の $(1 - \alpha)$ 分位点の方が小さいことを意味している。信頼区間の下限についても，同様な議論がなりたつ。

2.4 その他の記述

この後につづく §5. は最尤推定に関するいくつかの指摘であるが，ほとんどは，よく知られた内容である。それにつづく §6. 近似に関する注記，§7. 仮説検定と有意水準以降の内容は，稿をあたらめて紹介したい。

補論 A. ベイズ統計学の視点

ここでは本文の理解に必要な議論と若干の応用例を紹介する。

A.1. 基礎的な構造

一般的な統計的モデルでは，標本と標本空間 $x \in X$，母数と母数空間 $\theta \in \Theta$，および x の分布の密度関数 $p(x \mid \theta)$ が与えられ，モデル $\{X, \Theta, p(x \mid \theta)\}$ にもとづく標本 $x = (x_1, \cdots, xn)$ が得られたとき，未知母数 θ に関して何らかの推論を行う。Wald（1950）による統計的決定問題の枠組みでは，可能な行動 d の集合である決定空間 $d \in D$ を導入して，損失関数 $L(d, \theta)$ が d と $\theta \in \Theta$ に対して定められ，観測値 x の関数 (decision rule) $d = \delta(x)$ をどう定めるかという問題に帰着する。損失の期待値（リスク）$R(\theta, \delta) = E[L(\delta(x), \theta) \mid \theta] = \int_X L(\delta(x), \theta)\, p(x \mid \theta)\, dx$ は未知の θ に依存するため，ミニマックス基準などによって最適な

δ が選ばれる必要がある。この枠組みは極めて広く，Lehmann（1959）は，決定問題の特別な場合として，母数の推定問題や，仮説検定の問題を導いている。

ベイズ統計では，さらに母数空間 Θ 上に設定される事前分布 $p(\theta)$ が導入され，損失関数よりも効用関数 $U(d,\theta)$ が用いられることが多い。

Savage（1954）のような正統的なベイズ統計では，事前分布とデータからベイズの定理を用いて事後分布 $p(\theta \mid x) = p(x \mid \theta)p(\theta)/p(x)$ を求める。ここで $p(x) = \int_\theta p(x \mid \theta)p(\theta)\, d\theta$ である。次に，与えられた効用関数にもとづいて，行動 d が事後期待効用

$$\int_\theta U(d,\theta)p(\theta \mid x)\, d\theta \qquad \text{(A1)}$$

を最大にするように決定される。このように，ベイズの手法においてはモデルが与えられ，事前分布 $p(\theta)$ と効用関数 $U(d,\theta)$ が定められた後は計算だけの問題となる。特に（A1）では x が固定されているから，決定関数 $d = \delta(x)$ が明示的に与えられる。

以下，ベイズ統計の論理性と，標本理論の持つ問題点に関する議論を紹介する。

Lindley（1972）などのいう整合性（coherence）は，意思決定者の合理的行動に関する公理から，事前分布 $p(\theta)$ と効用関数 $U(d,\theta)$ の存在，および事後期待効用 $E(U \mid x)$ を最大にする行動 d が選ばれる，という結論が整合的に導かれることを指す。事前分布の存在は，少くとも有限な母数空間については簡単な公理から導かれる。現実の問題はもともと離散型だから，事前分布の存在を仮定することは十分な根拠を持つ。

効用関数 U が未知，あるいはその評価に手数がかかる場合には，ひとまず事後分布を評価することがベイズ統計の基本的な課題であり，将来，適当な効用関数が与えられたときには，事後期待効用は容易に評価できる。

整合性の考え方に対して，データのもつ情報を十分に引き出すために，なるべく個人的な主観の入り込む余地のない事前分布を採用し，効用関数は特定化しないという立場がある。Jeffreys（1967）や Box and Tiao（1973）の主張が，そのようなものである。この立場では，情報を持たない事前分布（non-informative

prior）を仮定する。この考え方は，データの持つ情報を十分に引き出すことを目的とするデータ解析学派（data analysis school）とも近い。さらに，モデルに関するある種の情報を表現する「情報のある事前分布」を利用する手法は，きわめて実用的であり，筆者は［2, 3, 6, 8, 9, 10, 12］でもそのような例を取り上げている。

標本理論においては，結果的に特定の効用関数と事前分布を仮定した場合のベイズ統計の解を求めることになるのが通例である。推定問題においては，この決定方式 $d = \delta(x)$ は，ベイズ推定量（Bayes estimator）と呼ばれる。このような解は効用関数に依存するから，標本理論の手法は，主観的な判断が明示されていないだけで，実際には恣意的な性格を持っている。

上述のように，ベイズ統計の体系が公理から整合的に導かれるのに対して，標本理論は雑多な手法の寄せ集めであり，その結果として多くの論理的な矛盾が発生する。以下に指摘する例は［2］で紹介した事例の一部であるが，いずれもベイズ統計では矛盾なく解決されている。

例　不偏性　すべての θ に対して $E(T(X) \mid \theta) = \int_X T(x)p(x \mid \theta)dx = \theta$ が成立するとき，ある統計量 $T(x)$ は θ の不偏推定量と呼ばれる。ここで，積分が標本空間 X に依存しているため，この基準は標本空間の選び方に依存する。次の例は広く知られている。

成功の確率を $\theta(0 < \theta < 1)$ とする n 回のベルヌーイ試行列において，r 回の成功が観測されたものとする。n を固定してデータを観測すると，このモデルは二項分布となり，r が確率変数，$X = \{0, 1, \cdots, n\}$ となる。この場合，θ の不偏推定量は $\hat{\theta} = r/n$ で与えられる。

一方，r を固定したときは n を確率変数とする負の二項分布で $X = \{r, r+1, \cdots\}$ となる。この場合には θ の不偏推定量は $\hat{\theta}^* = (r-1)/(n-1)$ となる。

$\hat{\theta}^*$ の不偏性は周知の展開式

$$\sum_{k=0}^{\infty} \binom{k+1-1}{k} (1-\theta)^k = \theta^{-r}$$

から，次のように確かめられる。ただし以下では $k=n-r, r'=r-1$ とおく。

$$E\left[\frac{r-1}{n-1}\right]=\sum_{n=r}^{\infty}\frac{r-1}{n-1}\binom{n-1}{r-1}(1-\theta)^{n-r}\theta^{r}=\sum_{k=0}^{\infty}\frac{(k+r-2)!}{k!(r-2)!}(1-\theta)^{k}\theta^{r}$$
$$=\left[\sum_{k=0}^{\infty}\binom{k+r'-1}{k}(1-\theta)^{k}\right]\theta^{r}=\theta^{-r'}\theta^{r}=\theta$$

特に $r=1$ の場合の幾何分布では，$\hat{\theta}^{*}$ は，$n=1$ のとき $\hat{\theta}^{*}=1, n\geq 2$ のとき $\hat{\theta}^{*}=0$ と定められる。この分布は指数分布族だから，これが唯一の不偏推定量である。その不自然さについては改めて指摘する必要もないであろう。

ところで，この例は，標本理論が利用する「不偏推定量」では「強い尤度原理」が必ずしも成立しないことを意味している。ベイズ統計では，事後分布による情報の集約が唯一の原理にしたがって処理されているが，標本理論では問題の性質によって，不偏性，一致性，最小分散不偏性など，多くの基準が導入されており，強い尤度原理もその一つである。ここでは，次の紹介に留め，その意義などの詳細な解説は［2］に譲る。

「弱い尤度原理」は同じ確率分布 $p(x\mid\theta)$ にしたがう 2 組の観測値 $x_1=\{x_{1i}\}$ $(i=1,\ldots,n_1), x_2=\{x_{2j}\}$ $(j=1,\ldots,n_2)$ について，それらの尤度関数が比例的 $p(x_1\mid\theta)\propto p(x_2\mid\theta)$ であれば x_1 と x_2 は同じ推論を導くことを要求する。この原理は，最小十分統計量を用いる推論と同等である。

「強い尤度原理」とは，異なった確率分布に従う確率変数 x, y についても，それらの尤度関数に $p(x_1,\ldots,x_m\mid\theta)\propto p(y_1,\ldots,y_n\mid\theta)$ という比例関係がなりたつときは両者の推論が一致することを要求するもので，弱い尤度原理よりも厳しい条件である。二項分布と負の二項分布の例では尤度関数が比例的であるが，不偏性は強い尤度原理を満たしていない。他方，ベイズ統計では，事前分布が同一なら事後分布は一致するから，強い尤度原理が成立する。

「条件性の原理」は，ある確率変数の分布が母数 θ に依存しないとき（これを補助統計量と呼ぶ），θ に関する推論では，補助統計量の値を固定した条件付き分布を用いることを意味する。例として，ある物体の特性 θ を測定する

のに，異なる器具を乱数 $z \sim U(0,1)$ で無作為に選び，$z < 1/2$ のとき $x \sim N(\theta, \sigma_1^2)$，$z > 1/2$ のとき $x \sim N(\theta, \sigma_2^2)$ とする場合，使った器具を知って推論を行うのであれば当然の原理である。

このような基本的と思われる原理について，標本理論で生じる多数の矛盾について，［2］で紹介している。結論として，個々の問題に応じて適用される原理の数を増すことによっては，標本理論の本質的な欠陥を取り除くことはできない。

例　停止規則　標本理論では，一般に弱い尤度原理と条件性の原理は認めるが，強い尤度原理は認めない。実際は，Birnbaum（1962）がこれらの原理の同等性を示しているにもかかわらず，この原理を認めない理由の例は次のとおりである。

いま $x_1, \cdots x_n$ を正規分布 $N(\theta, 1)$ からの標本として，仮説 $H_0 : \theta = 0$ を対立仮説 $H_1 : \theta \neq 0$ に対して検定する。有意水準を α とすれば，棄却域は $|\bar{x}| < c/\sqrt{n}$ で与えられる。ただし，c は標準正規分布の上側 $100\alpha/2\%$ 点である。

強い尤度原理によれば，母数に依存しない停止規則を用いる限り，どのような手順で標本を抽出しても結論は変わらない。ところが標本平均 $\bar{x}$ がその限界 $c/\sqrt{n}$ を超えるまで標本抽出を続けると，実際に $\theta = 0$ が正しいときでも，確率 1 で限界に到達する。したがって，強い尤度原理は常に誤った結論を導く。これが，強い尤度原理，およびこの原理が結論として導かれるベイズ統計を批判する例として広く用いられる。しかし，この例は逆に標本理論の欠陥を端的に示したものであることを Jeffreys［40］が示している。この話題は後に事前分布の節で取り上げる。

多くの論理的矛盾にもかかわらず，標本理論は広く用いられ，一定の成果をあげてきた。しかし Pratt［54］およびその論文に対する Lindley のコメントによれば，これは標本理論が正しいことを意味するのではなく，標本理論の解は，ある種の問題についてはベイズ統計の解の近似になっている，という偶然によるものと説明される。実際，［54］では，不偏推定量，信頼区間，仮説検定などが，適当な条件の下でベイズ統計の近似的な解として導かれることが示さ

れている。

実用上最も重要な結果は，正規線形回帰モデルにおける信頼区間がベイズ統計の最大事後密度域（highest posterior density region, HPD 域）として解釈できることであり，多くの実用的な問題において標本理論は近似的に正しい答を導いている。他方，複雑なモデルになると，ベイズ統計の柔軟性と比較して標本理論の手法を利用することは困難になる。このことは美添もいくつかの例で具体的に明らかにしている。

A.2. 事前分布

最近では，事前分布 $p(\theta)$ の利用は広く認められているようで，否定的な議論は少なくなった。主観確率に関しては，いくつかの正当化できる根拠がある。

精密測定　Savage［64］および Edwards，Lindman and Savage［27］によって安定的推定の原理（principle of stable estimation）と呼ばれ，ときに precise measurement とも呼ばれる定理によれば，任意の事前分布 $p(\theta) > 0$ に対して，それが適当に滑らかであれば，標本サイズが大きくなるにつれて事後分布は母数の真の値に集中することが示される。したがって，標本が大きいときは，事前分布の多少の差は結論に影響しない。ここで重要なのは真の値について $p(\theta) = 0$ とはならないという点である。真の $\theta = \theta_0$ について $p(\theta_0) = 0$ であれば，常に $p(\theta_0 \mid x) = 0$ となって，データの情報が事後分布に反映されない。これは自明なようだが，広い分布族を対象とするモデルが要求される場合には，$p(\theta_0) = 0$ となっていることが少なくない。たとえば頑健性の問題で，正規分布より広い分布族を扱うべき場合に正規分布のみを想定することは，他の分布の事前確率をゼロとすることを意味している。

情報のない事前分布　Bayes［18］は，二項分布の母数 θ $(0 < \theta < 1)$ に関して無知の状態を表現するのに，事前分布として一様分布 $p(\theta) = 1$ $(0 < \theta < 1)$ を採用した。これに対して，もし θ に関する無知の状態を一様分布で表わすなら，同様に無知である θ^2 に関する一様分布ではいけないのか，という問題がある。

一つの回答は Jeffreys［40］による不変事前分布であり，さらにわかりやすい形で translation invariance という名称で Box and Tiao［20］が論じているもので，母数が多次元の場合でも適用可能である。

Jeffreys の考え方は，母数 θ に関する事前分布を $\phi = \phi(\theta)$ という母数に変換したとき，ϕ に関する事前分布の表現が不変になるように事前分布の関数形を定める，というものである。その近似的な解は，$I(\theta)$ を Fisher の情報行列

$$I(\theta) = E\left[\left(\frac{\partial \log p}{\partial \theta}\right)\left(\frac{\partial \log p}{\partial \theta}\right)'\right]$$

とするとき $p(\theta) \propto |I(\theta)|^{1/2}$ で与えられる。実際に $p(\theta)$ から ϕ の事前分布 $p(\phi)$ を求めると $p(\phi) \propto |I(\phi)|^{1/2}$ となって，その関数形は不変であることが確かめられる。一般にはこの分布は Θ 上で積分すると発散するが，その場合，事前分布は improper と呼ばれる。

この事前分布の意味は，Box and Tiao［20］によれば次のように解釈される。データ y の値が変化すれば尤度関数 $p(y \mid \theta)$ も変化するが，適当に $\phi(\theta)$ を選ぶことにより ϕ に関する尤度関数の形が位置を除いて不変であるようにできたとする。このとき ϕ の事前分布として一様分布を仮定することは合理的であろう。一般には尤度関数を正確に不変とすることはできないが，近似的に不変とすることはできる。このとき ϕ の一様分布を θ に変換すれば，それが Jeffreys の不変事前分布となる。

共役事前分布　観測値が指数分布族

$$p(x \mid \theta) = a(\theta)b(x)\exp\left\{\sum_{i=1}^{m}\phi_i(\theta)T_i(x)\right\}$$

に従う場合は，事前分布を

$$p(\theta) \propto \{a(\theta)\}^{\alpha_0}\exp\left\{\sum_{i=1}^{m}\alpha_i\phi_i(\theta)\right\}$$

とおくのが便利である。ここで $\alpha_0 \cdots, \alpha_m$ は超母数（hyper parameter）と呼ばれ

る。この形の分布は，事後分布の関数形が事前分布と全く変らず，超母数だけが変化する。このような事前分布を，データの分布に対して共役（conjugate）という。

通常は，母数に関する情報は正確な関数形で与えられるわけではなく，平均やちらばりというもっと弱い形で与えられるものだから，数学的に扱いやすい形で近似するのが合理的である。上述の精密測定の原理によって，この近似を安心して採用することができる。

ところで共役事前分布は，情報のない状態から出発して，いくつかのデータを観測した後の状態と考えることもできる。過去のデータによって得られた事後分布は，今後の分析に際しては事前分布として用いられることになるが，情報のない事前分布が $\alpha_0 = \cdots = \alpha_m = 0$ で近似されるならば，データを観測した後の事後分布は，十分統計量を超母数とする共役分布となることが容易に導かれる。実用的な指数分布族とそれに共役な事前分布の形は，ベイズ統計がほとんど受け入れられていなかった時期から Raiffa and Schlaifer［57］にまとめられている。なおこの本は Harvard Graduate School of Business Administration（通称 Harvard Business School）の教材として利用されたもので，その後のベイズ統計学に関する教育的な内容は Pratt, Raiffa and Schlaifer［56］に整理されている。

A.3.　情報の利用

母数に関して何らかの先験的情報を持っている場合は，実際にも少なくない。このようなときに標本理論を用いると，そのような情報の確からしさを表現することが難しく，結局，母数に関するある制約条件を完全に正しいものとして扱うか，あるいは全く無視するかの選択を迫られる。その結果，いずれにしても十分なデータ分析は妨げられる。このような例に，線形モデルにおける線形制約の問題，連立方程式体系における識別条件の問題，Behrens-Fisher 問題，分散分析における変量模型の問題などがある。

以上の例については，ベイズ統計の手法によって，従来十分に解明されていなかった本質的な問題を明らかにすることができる。それは，与えられた先験

情報を母数に関する制約式についての確信の度合を表現する事前分布で表現することによって可能となる。Box and Tiao［20］が Behrens-Fisher 問題や分散分析の変量模型を例として指摘するとおり，事後分布 $p(\theta \mid x) \propto p(x \mid \theta)\, p(\theta)$ を通じて，データ情報 $p(x \mid \theta)$ と，先験的な情報 $p(\theta)$ の両者がどのように結論に影響しているかを明確にできることが，ベイズ統計の大きな長所である。

補論 B.　二項分布とベータ分布の関係

この節は美添［12］に記した内容の引用である。ベイズ統計における共役な分布という関係があるこれらの分布には，よく知られた関係がある。そのためにテイラーの公式の積分表示を利用する。

$$f(b) = f(a) + \sum_{r=1}^{m} \frac{1}{r!} f^{(r)}(a)(b-a)^r + \frac{1}{m!} \int_a^b (b-t)^m f^{(m+1)}(t)\, dt \tag{2.26}$$

区間 $(0, x)$ に対して，この公式を $f(x) = (1+x)^n$ に適用すると

$$(1+x)^n = 1 + n\,x + \frac{n^{(2)}}{2!} x^2 + \cdots + \frac{n^{(m)}}{m!} x^m + R = \sum_{r=0}^{m} \binom{n}{r} x^r + R \tag{B1}$$

を得る。ここで

$$R = \frac{n^{(m+1)}}{m!} \int_0^x (1+t)^{n-m-1} (x-t)^m\, dt$$

であり，

$$n^{(r)} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

という記法を用いている。ところで，ガンマ関数 $\Gamma(n+1) = n!$ とベータ関数の関係 $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ を用いると

$$\frac{n^{(m+1)}}{m!} = \frac{n!}{m!(n-m-1)!} = \frac{1}{B(m+1, n-m)}$$

が導かれる。ここで，簡単のためにこの定数を B^{-1} と表す。一方，$Y \sim B(n, \theta)$

に対して

$$\Pr(Y \le m) = \sum_{r=0}^{m} \binom{n}{r} \theta^r (1-\theta)^{n-r} = (1-\theta)^n \sum_{r=0}^{m} \binom{n}{r} \left(\frac{\theta}{1-\theta}\right)^r$$

だから，$x = \theta/(1-\theta)$ とおくと $1 + x = 1/(1-\theta)$ となり，上式の右辺を（B1）式と比較することによって次式が得られる。

$$\Pr(Y \le m) = (1-\theta)^n \left\{ (1+x)^n - B^{-1} \int_0^x (1+t)^{n-m-1} (x-t)^m \, dt \right\}$$

$$= 1 - B^{-1} (1-\theta)^n \int_0^x (1+t)^{n-m-1} (x-t)^m \, dt$$

さらに $x - t = u/(1-\theta)$ とおくと，$u = (1-\theta)(x-t) = (1-\theta)x - (1-\theta)t = \theta - (1-\theta)\, t$, $t = (\theta - u)\,/(1-\theta)$, $dt = -du/(1-\theta)$, および $1 + t = (1-u)/(1-\theta)$ となるから，

$$\Pr(Y \le m) = 1 - \int_0^\theta B^{-1} u^m (1-u)^{n-m-1} \, du$$

が得られる。この式の被積分関数はベータ分布 $\mathrm{Be}(m+1, n-m)$ の密度関数そのものである。以上をまとめて，次の関係が得られる。

命題 1　$p(y \mid \theta)$ を $B(n, \theta)$ の密度関数，$g(u \mid p, q)$ を $u \sim \mathrm{Be}(p, q)$ の密度関数とするとき，

$$\sum_{y=0}^{m} p(y \mid \theta) = \int_\theta^1 g(u \mid m+1, n-m) \, du \tag{B2}$$

定理 7 の証明にある（2.23）式は，この式を次のように変形し，記号を書き替えれば導かれる。

$$\sum_{j=m+1}^{n} \binom{n}{j} \theta^j (1-\theta)^{n-j} = \int_0^\theta \frac{u^m (1-u)^{n-m-1}}{B(m+1, n-m)} du$$

下限に対応する式は（B2）を直接適用して次の形に表したものである。

$$\sum_{j=0}^{m} \binom{n}{j} \theta^j (1-\theta)^{n-j} = \int_\theta^1 \frac{u^m (1-u)^{n-m-1}}{B(m+1, n-m)} du$$

謝辞

本研究は JSPS 科研費 JP18H00837 の助成を受けた。

This work was supported by JSPS KAKENHI Grant Number JP18H00837.

以下の文献にはベイズ統計学の基礎として，今回は取り上げなかったものも収録している。

参考文献

［1］ 後藤智弘・後藤文廣・美添泰人「可変回帰係数モデルを用いた消費者行動の分析」『青山経済論集』第 63 巻第 3 号，pp. 45–74, 2011

［2］ 美添泰人「ベイズの手法による統計分析―部分的なサーベイと今後の展望」，東京大学出版会，竹内啓（編）『計量経済学の新展開』第 6 章，1983

［3］ 美添泰人「ベイズの手法による 2 項回帰モデルの推定」東京大学出版会，鈴木雪夫・竹内啓（編）『社会科学の計量分析』第 3 章，1987

［4］ 美添泰人「ベイジアン多変量解析（その 1）」立正大学『経済学季報』第 38 巻，1988

［5］ 美添泰人「ベイジアン多変量解析（その 2）」立正大学『経済学季報』第 38 巻，1988

［6］ 美添泰人「多重共線性へのベイズ・アプローチ」東京大学出版会，鈴木雪夫・國友直人（編）『ベイズ統計学とその応用』第 7 章，1989

［7］ 美添泰人「効用についてのパラドックス―ベイズ統計からの解答―」立正大学『経済学季報』第 39 巻，1990

［8］ 美添泰人「有限母集団からの標本抽出―ベイジアン統計学からの解釈―」『創価経営論集』第 18 巻第 3 号，1994

［9］ 美添泰人「ベイズの手法による分布ラグモデルと季節変動の分析」一橋大学『経済研究』第 45 巻第 2 号，1994

［10］ 美添泰人「小地域統計の推定手法と応用」一橋大学『経済研究』第 52 巻第 3 号，pp. 231–238, 2001

［11］ 美添泰人「経済と統計の間で」『日本統計学会誌』第 39 巻シリーズ J 第 2 号，pp. 161–179, 2010

［12］ 美添泰人「統計的推論の原理とベイズの理論」『青山経済論集』第 63 巻第 3 号，pp. 113–143, 2011

［13］ 渡辺澄夫『ベイズ統計の理論と方法』コロナ社，2012

［14］ Akaike,H., “Likelihood and the Bayes procedure”, *Bayesian Statistics*, proceedings of the first international meetings held in Valencia (Spain).(Bernardo, J.M.*et al*., eds), Valencia, University Press, 1980.

［15］ Almon. S. “The Distributed Lags between Capital Appropriations and Expenditures,” Econometrica, 1965.

［16］ Aykac, A. and C. Brumat (eds.) *New Developments in the Applications of Bayesian*

Methods, North-Holland, 1977.
［17］ Basu, D. “Recovery of ancillary information,” Sankhya, A. 26, 1964.
［18］ Bayes, Thomas “An Essay towards solving a Problem in the Doctrine of Chances”. *Philosophical Transactions*. 53: 370–418, 1763.（https://www.jstor. org/stable/105741 から入手可能）
［19］ Birnbaum, A. “On the foundations of statistical inference,” JASA,3 2, 1962.
［20］ Box. G. E. P. and G. C. Tiao *Bayesian Inference in Statistical Analysis*, Addison-Wesley, 1973.
［21］ Cox, D.R. “The Choice Between Alternative Ancillary Statistics,” JRSS, B . 30, 1970.
［22］ DeGroot, M. H. *Optimal Statistical Decisions*, McGraw-Hill, 1970.
［23］ Dempster, A . P. “A Generalization of Bayesian Inference”. JRSS, B, 1968.
［24］ Dreze, J. H. “Bayesian Limited Information Analysis of the Simultaneous Equations Model,” Econometrica, 44, 1976.
［25］ Dreze, J.H. and J.-A.Morales “Bayesian Full Information Analysis of Simultaneous Equations,” JASA, 1976, reprinted in Zellner［74］.
［26］ Edwards, A.W.F. *Likelihood*, Cambridge University Press, 1972.
［27］ Edwards, W., H. Lindman and L. J. Savage “Bayesian statistical inference for psychological research”, Psychol. Rev.70, 1963, reprinted in Savage［66］.
［28］ Ferguson, T. S. *Mathematical statistics : a decision theoretic approach*, Academic Prcss, 1967.
［29］ Fieller, E. C. “Some Problems in Interval Estimation,” JRSS, B, 1954.
［30］ Fienberg, S. E. and A. Zellner (eds.) *Studies in Bayesian Econometrics and Statistics*, North-Holland, 1975.
［31］ Fisher, I. “Note on a Short-cut Method for Calculatini Distributed Lags,” lnternational Statist. Bulletin, 1935.
［32］ Fisher. R. A. *Statistical Methods and Scientific Inference*, 3rd ed. Oliver and Lloyd, 1973.
［33］ Fraser D. A. S. *Probability and Statistics,Theory and Applications*, Duxbury Press, 1976.
［34］ Good, I. J. *Probability and the weighting of evidende*, Charles Griffin, 1950.
［35］ Good, I. J. *The Estimation of Probabilities*, MIT Press, 1965.
［36］ Hacking, I. *Logic of Statistical Inference*, Cambridge University Press, 1965.
［37］ Hacking, I. *The Emergence of Probability*, Cambridge University Press, 1975.
［38］ Harkema, R. *Simultaneous Equations: A Bayesian Approach*, Rotterdam University Press, 1971.
［39］ lshiguro, M. and H. Akaike “Trading-day Adjustment for the Bayesian Seasonal Adjustment Program BAYSEA,” Research Memorandom. No. 189, The Institute of Statistical Mathcmatics, 1980.
［40］ Jeffreys, H. *Theory of Probability*, 3rd ed.(corrected), Clarendon Press, 1967.
［41］ Kadane. J. B. “The role of identification in Bayesian theory,” in Fienberg and Zellner［30］.
［42］ Kalbfleisch, J. D. and D. A. Sprott “Application of likelihood methods to models involving large numbers of parameters,” JRSS, B. 32, 1970.

［43］ Kloek, T. and H. K. van Dijk “Bayesian Estimates of Equation System Parameters: An Application of Integral by Monte-Carlo,” Econometrica, 1978, reprinted in Zellner［74］.
［44］ Koyck, L. M. *Distributed Lags and Investment Analysis*, North-Holland, 1954.
［45］ Lehmann, E. L. “Significance level and power,” AMS, 29, 1958.
［46］ Lehmann, E. L. *Testing Statistical Hypotheses*, Wiley, 1959.
［47］ Lindley, D. V. *Introduction to Probability and Statistics from a Bayesian View point*, Cambridge University Press, 1965.
［48］ Lindley, D. V. *Bayesian Statistics, a review*, SIAM Monograph, 1972.
［49］ Lindley, D.V. and A.F.M. Smith “Bayes estimatesf or the linear model (with discussion),” JRSS, B.3 4, 1972.
［50］ Maddala, G.S. “Weak Priors and Sharp Posteriors in Simultaneous Equation Models,” Econometrica, 44, 1976.
［51］ Morales, J.A . *Bayesian Full lnformalion Structural Analysis*, Springer, 1971.
［52］ Mosteler, F. and J. W. Tukey *Data Analysis and Regression*, Addison-Wesley, 1977.
［53］ Pratt, J. W. “Length of confidence intervals,” JASA, 56, 1961.
［54］ Pratt, J. W. “Bayesian interpretation of standard inference statements,” JRSS, B, 27, 1965.
［55］ Pratt. J. W., H. Raiffa and R. Schlaifer “The foundations of decision under uncertainty: an elementary exposition,” JASA, 63, 1964.
［56］ Pratt. J. W., H. Raiffa and R. Schlaifer *Introduction to Stalislical Decision Theory*, MIT Press, 1995.
［57］ Raiffa, H. and R. Schlaifer *Applied Stalislical Decision Theory*, MIT Press, 1961.
［58］ Ramsey. F. P. “Truth and Probability,” 1931, reprinted in Kyberg and Smokier (eds.), Studies in Subjective Probability, Wiley (1964).
［59］ Richard.J. F. *Posterior and Prediclive Designs for Simultaneous Equation Models*, Springer, 1973.
［60］ Rothenberg, T. J. *Efficient Estimation wilh A Priori Informalion*, Yale University Press, 1973.
［61］ Rothenberg, T.J. “Bayesian Analysis of Simultaneous Equation Models,” in Fienberg and Zelner［30］.
［62］ Savage, L. J. *The Foundations of Statistics*, Wiley, 1954.
［63］ Savage. L. J. “The foundations of statistics. Reconsidered,” Proc. Fourth Berkeley Sympos.,1961, reprinted in Savage［66］.
［64］ Savage. L. J. “Bayesian Statistics,” in Recent Developments in Information and Decision Processes, (R. E. Manchol and P. Gray, eds.), Macmillan, 1962. reprinted in Savage［66］.
［65］ Savage, L. J. “Elicitation of Personal Probabilities and Expectations,” JASA, 1971, reprinted in Savage［66］.
［66］ Savage, L. J. *The Writings of Leonard Jimmie Savage — A Memorial Selection*, ASA/IMS, 1981.
［67］ Shiller, R. J. “A distributed lag estimator derived from smoothness priors, ” Econometrica, 4l, 1973, reprinted in Fienberg and Zellner［30］.
［68］ Theil. H. and A.S. Goldberger “On Pure and Mixed Statistical Estimation in Economics,”

IER, 2, 1960.
［69］ Tukey, J. W.“The future of data analysis,” AMS, 33, 1962.
［70］ Tukey, J. W. *Exploratory Data Analysis*, Addison-Wesley, 1977.
［71］ Wald, A. *Statistical Decision Functions*, Wiley, 1950.
［72］ Yoshizoe, Y. “A Bayesian method of estimating mortality rates in small areas”, *Bulletin of the International Statistical Institute*, 2001.
［73］ Zellner, A. *An Introduction to Bayesian Inference in Econometrics*, Wiley, 1971.
［74］ Zellner, A . (Ed.) *Bayesian Analysis in Econometrics and Statistics*, North-Holand, 1980.