経済学研究 35-4
北海道大学 1986. 3

単純なベイズ学習，母数空間の制限，及び，Bergerの見解について

園　信太郎

1. 序

ベイジアンなる語を，単に「ベイズの公式を使う」と言う意味に用いるのならば，ベイズ哲学（Bayesian philosophy）と呼ばれる方法論への極端な誤解が生じたとしても不思議ではない。例えば，Wald 流の決定理論と決定問題へのベイズ接近との方法論的混同，より一般的には，前実験的基準（pre-experimental criterion）と後実験的基準（post-experimental criterion）との概念的混同，が屢視られる。この様な混同の典型として，所謂ベイズ推定量，に基づく接近がベイズ接近の一種であると誤解する事が在るが，Box and Tiao［2］，第5章の付録5.6，304頁から315頁まで，に，この誤解に対する簡潔な分析が在る。

一方，ベイジアン内部に，統計学の基礎に関する基本的見解の相違が在る。例えば，公理或いは公準として提出される諸仮定に対する方法論的観点の相違である。Savage［5］に示されている，公準（postulate）なる概念に対する態度は，そこで提出されている諸公準を採用するか否かにかかわらず，少なくとも方法論的意味において，肯定されるべきであると，筆者は考える。例えば，数学的構造に直接連関した部分では，同書第3章第4節の42頁から43頁まで，特に43頁の最後の段落，における，統計学の基本概念としての確率に対して完全加法性，原文は countable additivity，を要請する事を拒否する論述等，は，論理的に明晰であると考える。また，同書第3章第3節，33頁から40頁まで，における，世界（world）S の「任意有限個の等確率な事象達による分割達」の存在，を演繹する際の論理の展開及び数学的形式による概念規定は，同書39頁から40頁にかけて提出される第6公準P6の論理的基盤として，簡潔且つ適確である，と，考えるのである。

ベイズの公式によって表現される事前分布から事後分布への変換を，「抽象化された単純な学習」の表現として解釈する事が，ベイジアンに共通する観点である。以下の第2節において，この「単純な学習」の定式化を母数空間の制限に基づいて模索し，第3節において，この定式化を実行する事によって，後実験的意味での，ベイズ学習の収束が容易に示される事を指摘し，さらに，これら二節の議論と Savage［5］の基本概念の1つである世界 S，世界の定義は同書9頁（第2章第3節），全事象（universal event）と世界との同一視については同書10頁（第2章第4節）を参照，との関連について，第4節において，論じる。なお，第5節では，Berger［1］に端的に現われている「事前分布の選択に関する考え方」，への簡潔な批判を試みる。

2. 母数空間の制限

事後分布の未知固定の真の母数への集積を，確率収束の概念に基づいて定式化する事は，ベ

イジアンの本来の立場からは，妥当ではない。一方，実践的なベイジアン達は多数の数値例やグラフ等によって，この様な集積が「事実上」無条件に生じる事を示している。数値実験的な諸例に基づいて，事後分布の収束が実証されたとするのも，方法論的には，一つの立場ではあるが，ここでは採らない。

母数空間 Θ 上の事前分布 $p(\theta)\lambda(d\theta)$ に対して，

$$\Theta(\varepsilon)=\{\theta\epsilon\Theta\,;\,p(\theta)\geqq\varepsilon\},\ \varepsilon>0,$$

とする場合，実践的ベイジアンは，Θ そのものよりもむしろ，$\Theta(\varepsilon)$ の型の部分空間を考察の対象としている，と仮定する事は方法論的に妥当であると筆者は考える。Θ が位相空間であり $p(\theta)$ が上半連続の場合は $\Theta(\varepsilon)$ は Θ の閉集合となるが，特に $\Theta(\varepsilon)$ が compact, 但し Hausdorff とは限らない，部分空間ならば，この $\Theta(\varepsilon)$ は，実践的な意味での「母数空間の制限 (constraint)」，を表現すると解釈する事も妥当である，と考える。

一方，未知固定の真の母数 θ_* が $\Theta-\Theta(\varepsilon)$ に含まれる場合は，$\Theta(\varepsilon)$ の部分集合の内で，何らかの意味において最も θ_* に近い集合上に，事後分布が集積しなければならない。但し，この近さを表現する為には，方法論的な解釈が或程度まで明確な概念を用いる事が要求される。また，事後分布の集積は，後実験的な収束，即ち，概収束の概念によって定式化されなければならない。

また，ベイジアンが $\Theta(\varepsilon)$ 上に位相を導入し，且つ，$\Theta(\varepsilon)$ 上に後実験的な尤度を定める場合，この尤度が，$\Theta(\varepsilon)$ 上で，正値且つ連続であると仮定する事は，妥当である様に思われる。

以上の議論に基づき，「単純な学習」の収束を定式化し，且つ，演繹する。

3.　学習の後実験的収束

この節の議論は Sono [7]，第4章，に基づく。

4つ組，$(\Theta,\ d\,;\ \theta_*,\ \Gamma)$ を考える。ここに Θ は擬距離 d によって定まる位相空間であり，θ_* は未知固定の，空間 Θ に属する，一点である。また Γ は compact, 但し Hausdorff とは限らない，な，Θ の部分空間である。$f(x|\theta)$ を，母数 $\theta\epsilon\Theta$ を持つ，可測空間 $(X,\ \boldsymbol{F})$ 上の σ 有限測度 $\mu(dx)$ に関する，確率密度とし，且つ，任意の $x\epsilon X$ に対して，$f(x|\theta)$ は Γ 上で連続であると仮定する。従って $f(x|\theta)$ は $(X\times\Gamma,\ \boldsymbol{F}\times\boldsymbol{B}(\Gamma))$，但し $\boldsymbol{B}(\Gamma)$ は Γ 上の位相的 Borel 集合族とする，上の可測関数となる。さらに，任意の $(x,\ \theta)\ \epsilon X\times(\Gamma\cup\{\theta_*\})$ に対して，$f(x|\theta)>0$ となる事を仮定する。任意の $(\theta_0,\ \delta)\epsilon\Gamma\times]0,\infty[$ に対して，集合 $\{\theta\epsilon\Gamma\,;\,d(\theta,\theta_0)<\delta\}$ を $U(\theta_0,\delta\,;\Gamma)$ と書く。Kullback-Leibler 情報量が存在するのならば，これを $I(\theta_*\,;\,\theta)$ と書く，即ち，

$$I(\theta_*\,;\,\theta)=E(\log(f(x|\theta_*)/f(x|\theta))|\theta_*),$$

但し $E(\cdot|\theta_*)$ は $f(x|\theta_*)\mu(dx)$ による期待値を表わす，とする。また，$\min\phi=+\infty$ 且つ $\log 0=-\infty$，とする。

この時次の命題が成立する。

［命題］　$E(\sup\{|\log f(x|\theta)|\,;\,\theta\epsilon\Gamma\cup\{\theta_*\}\}|\theta_*)<\infty$ を仮定し，

$$\Gamma_*=\{\theta_0\epsilon\Gamma\,;\,\min_{\theta\epsilon\Gamma}I(\theta_*\,;\,\theta)=I(\theta_*\,;\,\theta_0)\},$$

とする。$P(d\theta)$ が，$0<P(\Gamma)<\infty$，且つ，任意の $(\theta_0,\ \delta)\epsilon\Gamma_*\times]0,\ \infty[$ に対して $P(U(\theta_0,\delta\,;\Gamma))>0$ を満たす，$(\Theta,\ \boldsymbol{B}(\Theta))$ 上の σ 有限な測度ならば，$F\subset\Gamma-\Gamma_*$ なる任意の閉集合と $(\theta_0,\ \delta)\epsilon\Gamma_*\times]0,\ \infty[$ なる任意の $(\theta_0,\ \delta)$ に対して，

$$\Big(\int_F P(d\theta)\prod_{k=1}^{n}f(x_k|\theta)\Big)\Big/\Big(\int_{U(\theta_0,\delta;\Gamma)}P(d\theta)\prod_{k=1}^{n}f(x_k|\theta)\Big)\xrightarrow[n\to\infty]{}0,\ P_{\theta*}^{(\infty)}a.s.,$$

但し，$P_{\theta*}^{(\infty)}=\prod_{k=1}^{\infty}f(x_k|\theta_*)\mu(dx_k)$ とする，

となる。

［証明］　$f(x^n|\theta)=\Pi_{k=1}^{n}f(x_k|\theta)$，$x^n=(x_k\,;k=1,$

$\cdots, n)$, 且つ, $r(x;\delta)=2\sup\{|\log f(x|\theta_1)-\log f(x|\theta_2)|;\ d(\theta_1,\theta_2)\leqq\delta\ \&\ (\theta_1,\theta_2)\in\Gamma\times\Gamma\}$, と置く。この時次の不等式(1)が容易に導かれる。

(1) $$\limsup_{n\to\infty}\log\left(\int_H P(d\theta)f(x^n|\theta)\Big/\int_K P(d\theta)f(x^n|\theta)\right)\leqq E(r(x;\delta)|\theta_*)-\min_{\theta\in F}I(\theta_*;\theta)+\min_{\theta\in\Gamma}I(\theta_*;\theta),\ P_{\theta_*}^{(\infty)}a.s.,$$

但し, $H\subset F\subset\Gamma-\Gamma_*$, F は Θ の閉集合, 且つ, $K\subset\Gamma$, $K\cap\Gamma_*\neq\phi$ & $P(K)>0$, 且つ, $\max(\mathrm{diam}(K),\mathrm{diam}(H))\leqq\delta$, $\delta>0$。ここに, $\mathrm{diam}(S)=\sup\{d(s_1,s_2);\ (s_1,s_2)\in S\times S\}(S\neq\phi)$, & $\mathrm{diam}(\phi)=-\infty$。

$E(r(x;\delta_0)|\theta_*)\to 0$, $\delta_0\to 0$, より, 或 $\delta_0>0$ が存在して, 任意の $\delta\in]0,\delta_0]$ に対して,「(1)の右辺」<0, となる。故に不等式(1)より, $\delta\leqq\delta_0$ ならば次の(2)が従う。

(2) $$\int_H P(d\theta)f(x^n|\theta)\Big/\int_K P(d\theta)f(x^n|\theta)\underset{n\to\infty}{\longrightarrow}0,\ P_{\theta_*}^{(\infty)}a.s..$$

F の有限被覆 $\{H_i;\ i\in I\}$ で, $\mathrm{diam}(H_i)\leqq\delta_0$, $i\in I$, を満たすものを取り, 且つ, $K=U(\theta_0,\delta;\Gamma)$, $(\theta_0,\delta)\in\Gamma_*\times]0,\delta_0]$, と置く。この時(2)及び $\int_F P(d\theta)f(x^n|\theta)\leqq\Sigma_{i\in I}\int_{H_i}P(d\theta)f(x^n|\theta)$ より結果が得られる。[証明終]

上記の議論においては, 標本空間 X に位相的な条件, 例えば Borel 空間, 可分空間, 等, が課せられていない事を注意する。一方, [命題] において仮定されている条件は, 何らかの意味において X が制限されており, その結果として,「或正数 m 及び M が存在して, 任意の $(\theta,x)\in(\Gamma\cup\{\theta_*\})\times X$ に対して, $m\leqq f(x|\theta)\leqq M$」を満たす場合には常に成立する。

上記の [命題] は大数の強法則の単純な系である。ベイズ分析においては, 母数空間上の事後分布が, その分析の目的に応じて, 種種の形式に変換され, 各各の変換された分布に対して, その目的に応じた解析的或いは数値的な計算が行なわれるが, 事後分布の漸近的な性質の内で仮定され且つ利用される者は, 上記の性質に他ならない。(例えば, Box and Tiao [2] は首尾一貫してこの様なベイズ分析を行なっている。特に, 同書第5章, 244頁から294頁まで, 例えば, 第5. 1. 3節, 248頁から249頁, 第5. 2. 3節, 253頁から255頁, 第5. 2. 7節, 264頁から266頁, 等において, 頻度論的な分散分析の方法論的な不整合性 (methodological inconsistency) を指摘し, 且つ, この不整合性が生じる原因達を, ベイズ分析の立場から簡潔且つ論理的に分析している。) 従って, 上記の命題は, 所謂漸近論者 (Asymptotician) 的な議論が, ベイズ哲学から見る限りでは, 統計学の基礎に関する副次的な議論である事の, 単純且つ論理的な一つの方法論的根拠であると解釈し得る。

なお, 上の [命題] において, $P(d\theta)$ を $(\Theta,\boldsymbol{B}(\Theta))$ 上の有限加法的確率測度としても, 同じ結論が成り立つ事を注意する。

4. 標本空間の制限と「世界」の位相

母数空間 Θ 及び標本空間 X を或人格が考察の対象とする場合, 当の人格は前実験的な事柄, 例えば, 予測, 頻度論的な諸基準の設定, 等, を考察する事になるのであるが, この様な情況においては, 当の人格が対象とする世界 S から $\Theta\times X$ への全射 φ, 但し φ はなんらかの方法論的解釈を有する, の存在が仮定されている。この時 $(\Theta\times X,\boldsymbol{B}\times\boldsymbol{F})$ 上の主観確率 $p(d(\theta,x))=\lambda(d\theta)\mu(dx)p(\theta,x)=\lambda(d\theta)p(\theta)\mu(dx)f(x|\theta)$ が採用されるのならば, その人格は, 前実験的な立場から, 頻度論的な統計的模型 $(f(x|\theta);(\theta,x)\in\Theta\times X)$ を仮定した事になる。さらに $p(\theta,x)$ の $\Theta\times X$ 上での有界性が仮定され, 第2節におけるが如き制限 $\Theta(\varepsilon)$ 及び $\Theta\times X$ 上の制限 $\{(\theta,x)\in\Theta\times X;\ p(\theta,x)\geqq\eta\}$,

$\eta>0$, が採用されるのならば，これらの制限によって得られる $\Theta\times X$ の部分空間を Ω とすると，或正数 m 及び M が存在して，任意の $(\theta, x)\epsilon\Omega$ に対して，$m\leqq f(x|\theta)\leqq M$, が成立する。この場合，その人格は条件付き確率 $p(d(\theta, x)|\Omega)$ を考察することになるのだが，表記の煩雑さを避ける為に，$p(d(\theta, x))$ が既に，任意の $\theta\epsilon\Theta(\varepsilon)$ に対して「$m\leqq f(x|\theta)\leqq M$, μ *a.e.*」，を満たしているものと仮定する。

この様な情況に基づいて解釈される次の［命題］を示す。

［命題］ $(\Theta, \boldsymbol{B})$ 及び $(X, \boldsymbol{F})$ を可測空間，Γ を Θ の空でない部分集合，但し $\boldsymbol{B}$ の元とは限らない，とする。また，$f(x|\theta)$ を $(\Theta\times X, \boldsymbol{B}\times\boldsymbol{F})$ 上の正値可測関数，$\mu(dx)$ を $(X, \boldsymbol{F})$ 上の測度，として，任意の $(\theta_0, \theta)\epsilon\Theta\times\Theta$ に対して，

$$J(\theta_0;\ \theta)=\int_X\mu(dx)f(x|\theta_0)(-\log)^+\left(\frac{f(x|\theta)}{f(x|\theta_0)}\right),$$

(但し，$(-\log)^+(x)=\max\{-\log x, 0\}$, $x>0$), と置く。この時，任意の $\theta\epsilon\Gamma$ に対して

$$\int_X\mu(dx)f(x|\theta)<\infty,\ \text{且つ},$$
$$0<\inf_{(\theta,\ x)\epsilon\Gamma\times X}f(x|\theta)\leqq\sup_{(\theta,\ x)\epsilon\Gamma\times X}f(x|\theta)<\infty,$$

ならば，$\{V(\theta;\ \delta);\ (\theta, \delta)\epsilon\Gamma\times]0, \infty[\}$, ここに，$V(\theta;\ \delta)=\{\theta_0\epsilon\Gamma;\ \Delta(\theta, \theta_0)<\delta\}$, 且つ，$\Delta(\theta, \theta_0)=\max\{J(\theta;\ \theta_0), J(\theta_0;\ \theta)\}$, とする，を基底とする位相が Γ 上に定義され，且つ，この位相空間は擬距離による距離づけが可能である。さらに，「$(\theta_1, \theta_2)\epsilon\Gamma\times\Gamma$, $\theta_1\neq\theta_2$, ならば $\mu(\{x\epsilon X;\ f(x|\theta_1)\neq f(x|\theta_2)\})>0$」，が満たされる場合は，この位相空間 Γ は Hausdorff 空間となり距離づけ可能となる。

［証明］ $\{V(r);\ r$ は正の有理数$\}$, 但し $V(r)=\{(\theta_1, \theta_2)\epsilon\Gamma\times\Gamma;\ \Delta(\theta_1, \theta_2)<r\}$ とする，は Γ 上の或一様構造の可算基底となる。故に，Chittenden の距離化定理，例えば柴田［6］，第IV章，第2. 3節，定理9及びその系，246頁から247頁，を参照，により結論が従う。［証明終］

第3節の［命題］においては，Θ 上に擬距離 d が与えられており，Γ は compact, $f(x|\theta)$ は $\theta\epsilon\Gamma$ 上で連続，と仮定されているが，さらにこの節の［命題］の諸仮定がすべて満たされるのならば，この節の［命題］の二番目の結論によって定まる Γ 上の Hausdorff 位相と，d によって定まる Γ 上の位相とが一致する事，を注意する。(このことは，$\Delta(\theta_0, \theta)$ が (Γ, d) 上で θ に関して連続となる事，及び，compact 空間から Hausdorff 空間への連続な全単射が同相写像となる事，より従う。) この性質は，便宜的な観点から導入される Θ 上の位相に関して，尤度関数が連続である事を要請する事の「実用性」，の，論理的根拠の一つであると解釈し得る。

この節の［命題］の一番目の結論によって定まる Γ 上の位相を $\mathcal{T}_\Gamma$ とする。S から $\Theta\times X$ への全射 φ と，$\Theta\times X$ から Θ への射影 π, との合成写像によって，S の部分集合 $(\pi\circ\varphi)^{-1}(\Gamma)$ 上に位相 $(\pi\circ\varphi)^{-1}(\mathcal{T}_\Gamma)$ が導かれる。上記の［命題］によれば，この誘導された位相は，人格の主観確率 $p(d(\theta, x))$ によって定まるものであり，統計学の外部から与えられる，或いは，便宜的に導入される，Θ 上の位相，によって定まるものではない事を注意する。この意味で，位相 $(\pi\circ\varphi)^{-1}(\mathcal{T}_\Gamma)$, これは擬距離による距離づけが可能なのではあるが，は「内在的 (intrinsic)」である，と考えられる。

5. Berger［1］について

Berger は［1］において，繰り返し，ベイジアンの立場を支持する事を主張するのではあるが，彼の決定問題に対する態度は，事前分布及び事後分布の概念に基づく統計的推論，即ち，ベイズ推論，を基盤とする決定及び戦略，

即ち，ベイズ決定及びベイズ戦略，を支持するものと言うよりは，寧ろ，所謂ベイズ推定量，本質的にはWald流の決定関数の理論，による接近を支持するものである様に，筆者には，思われる。実際，同書においては，首尾一貫して損失関数の概念をベイズ接近の基本概念の一つとして位置づけ，さらには，ベイズ推定量に対する基準頑健性(criterion robustness)の要請を主張する。例えば，同書第4. 6. 3節，139頁から142頁，及び，第4. 7節，155頁から156頁，を参照。しかも，Wald 流の決定関数の理論においては基本概念の一つである「許容性(admissibility)」，を重視する。例えば，同書第4. 5. 2節，123頁から125頁上から14行目まで，を参照。

この様な方法論的矛盾が生じる原因の一つは，「ベイズ哲学における事前分布の概念」の論理的解釈を Berger が軽視している事に在る，と，筆者は考える。実際，以下で指摘する様に，「Box and Tiao［2］における，非報知事前分布（noninformative prior）の解釈」に対する無視，共軛事前分布達（conjugate priors）とベイズ学習に対する方法論的解釈の無視，及び，「未知固定の真の母数」を含む母数空間への「主観確率の導入」の方法論的意義の軽視，が視られる。

Berger［1］，第4. 3節，99頁，下から，19行目から15行目，は，Box and Tiao［2］が採用している非報知事前分布の概念は，その事前分布を定める規則(rule)を見る限りでは，Jeffreys の規則等の不変性基準（invariance criteria）と実質的な差がない，と言う趣旨の意見を述べる。(但し原文は丸括弧でくくってある。) しかし，Box and Tiao［2］において採用されている「事前分布の構成の方法達」は，データ移動型及び近似的データ移動型尤度(approximate data translated likelihood)の概念と「適切に変換された母数達」に関する事前独立性の仮定達(the prior independence assumptions on the suitably transformed parameters）に基づくものであり，従って Jeffreys の規則の形式的あてはめは用心深く検討されなければならない。例えば，Box and Tiao［2］，第1. 3. 6節，46頁から58頁，第5. 2. 2節，251頁から252頁，付録5. 5，303頁から304頁，付録5. 6，304頁から315頁，第10. 3. 1節，531頁から533頁，等を参照。従って，非報知事前分布をベイズ推定量を得る為の単なる入力として処理することは，ベイズ推論及びベイズ推論に基づくベイズ決定の立場からは，正当な事であるとは言い難い。例えば，Berger［1］は，第4. 5. 2節，例15，123頁から124頁，において，データ生成模型（data generating model）

$$f(x|\beta)=\beta^{-\alpha}e^{-x/\beta},\ x>0,$$

但し，$\alpha>1$ 且つ既知，$\beta>0$，とする，の未知固定なる尺度母数 β に対して，非報知事前分布 $\pi(\beta)=\beta^{-1}$ を導入し，事後分布 $\pi(\beta|x)$ の事後平均によって与えられるベイズ推定量 $\delta^0(x)=x/(\alpha-1)$ が $\alpha\to1$ の時 $+\infty$ に発散する事に基づいて，非報知事前分布 $\pi(\beta)=\beta^{-1}$ を批判しているが，Box and Tiao［2］における非報知事前分布の解釈に基づくのならば，ベイズ分析の基礎となるべき分布は $\pi(\beta|x)$ ではなく，母数 $\log\beta$ 上の事後分布 $\pi(\log\beta|x)$ であり，強て β に関する点推定を行なうのならば，$\pi(\log\beta|x)$ の mode によって与えられる β，即ち，

$$\pi(\log\hat{\beta}|x)=\max_{\beta>0}\pi(\log\beta|x),$$

によって与えられる $\hat{\beta}=x/\alpha$ が最も論理的に妥当な推定値となるので，少なくとも Box and Tiao 式の $\pi(\beta)=\beta^{-1}$ の解釈に立つ限りでは，Berger の非報知事前分布に対する批判は正当とは言い難いのである。なお同じ例において，Berger は「許容性」を重視する立場を示しているが，これはベイズ推論を基盤として決定を行なうと言う本来のベイジアンの立場からは隔たったものである。(この段落の議論は Sono［7］，第1章第2節，に基づくものである。)

Berger［1］においては，共軛事前分布の概念の方法論的有用性は否定されている。例えば，同書第4. 2. 2節，97頁下から10行目から98頁上から2行目まで，を参照。仮りに共軛事前分布の方法論的有用性を承認すると，Berger が強く，繰り返し，支持する「ベイズ推定量の基準頑健性」の概念が，少なくともベイジアンの観点からは，副次的なものとなってしまうので，（大部分のベイジアンにとっては意外である）この共軛事前分布の概念への評価も，寧ろ当然の帰結と言えるかもしれない。

Berger は，「真の事前分布（true prior）」，同書97頁下から7行目の冒頭，を共軛事前分布で近似的に表現する事の有用性を否定或いは疑問視している。理由は，「共軛事前分布を入力として得られるベイズ推定量が，極端な値のデータに対して，不合理な推定値をもたらす。」，と要約できる。未知固定なる真の母数を含む母数空間上に主観確率を導入する際には，「真の事前分布」なる概念は方法論的な意味を持ち得ない。実際，複数の事前分布を導入する事に基づく，事後分析の遂行，が，ベイズ統計学（Bayesian statistics）の仕事である。この事以外にも，Berger の「事前分布の概念」に対する態度に，疑問を感じざるを得ない箇所が幾つか在る。例えば，同書第3. 3節，84頁，冒頭の段落，で，未知固定なる母数が「既知である統計的模型」の出力であるとの仮定に基づく，ベイズの公式の応用を，ベイズ分析に含めているが，これは頻度論的な「ベイズの公式の応用」以外の何者でもない。また，同書第4. 4. 5節，117頁から121頁，では，経験的ベイズ模型への経験的ベイズ接近に言及しているが，これも亦頻度論的分析である。しかも，同書は，経験的ベイズ模型へのベイズ接近，及び，“random effect model”へのベイズ接近，には一切言及していない。この様な例示のやり方では，ベイズ接近の柔軟性（flexibility）が不明確になるばかりか，ベイジアンへの誤解を招く恐れが在るのではなかろうか，と，筆者は考える。なお，Berger は同書第4. 6. 3節，139頁から142頁，において，共軛事前分布に代わる者として，「頑健な事前分布（robust prior）」を支持しているが，この類の事前分布，同書第4. 6. 4節，143頁に $g_n(\boldsymbol{\theta})$ なる一例が引用されている，が，“random effect model”に柔軟に対応し得るとは，少なくとも筆者には，思えない。

極端な値のデータに対しては，「標本空間の制限として表現される，事前情報の一部」（この論文の第4節を参照）に反する事後情報が得られたのであるから，事前情報に基づいて仮定された統計的模型では，その極端な値の生成を説明する事が困難である，と，判断する事が合理的である。従って，異常値を生成する統計的模型達と正常値を生成する統計的模型達とを結合して一個の統計的模型を構成する事に基づいて，ベイズ分析が遂行される。この様な分析は異常値問題（outlier problems）へのベイズ接近として知られている。（例えば，鈴木［9］を参照。）なお，Berger は同書第4. 6. 3節の最後の段落，142頁上から8行目から9行目までの丸括弧の中，で，共軛事前分布の裾（tail を直訳的に「尾」としないのは，多次元の母数空間上の山容の連想を採る為である）の部分の確率密度が速やかに減少する事を，この事前分布の欠点であるとする内容の記述をしているが，この論文の第2節で述べた様な母数空間の制限を，分布の切断（truncation）を直接に用いる事なく，簡潔且つ解析的に表現する「共軛事前密度の速やかな減少」は，この事前分布の長所の一つに他ならない。

ベイズ学習（Bayesian learning）の収束には，同書は明確な形では言及していないが，それと連関のある「高精度測定の原理（principle of precise measurement）」，例えば，宮沢［4］，第2章第5節，97頁から104頁，を参照，に，“principle of stable estimation or stable measurement”として言及している。（同書第4. 6. 2節，I，131頁下から13行目か

ら6行目，を参照。）しかし，同節同箇所のすぐ次の段落において，この事後分布の性質は，決定理論的な観点からは“uninteresting”であると評価される。（同書131頁下から5行目から1行目まで。）一方，この第4. 6. 2節，I，の表題は“Posterior Robustness”であり，この概念が同節の冒頭の段落，同書130頁下から6行目から131頁上から3行目まで，において定義されるのであるが，「ベイズ行為（Bayes action）a_0 に対する事後期待損失が，仮定される事前分布達の上で比較的安定した値をとる場合，その仮定される事前分布達に関して a_0 は“posterior robust”である」とするのがその内容である。しかし，これは「事後期待値の頑健性」以外の何者でもなく，寧ろ，「高精度測定の原理」こそ“posterior robustness”の定式化に他ならないと，筆者は考える。また，或人格が「高精度測定の原理」の不成立を仮定するのならば，尤度によって表現される「データがもたらした情報」との比較において，その人格の事前情報は報知的（informative）となり，事前分布，事後分布，及び事後密度と事前密度との比率，即ち尤度，に基づく「行為の選択」，即ち，「決定」，を行なう事が，その人格の合理的な態度となる。なお，この様な場合，仮りに「未知固定なる真の母数」から隔たった所に事前分布が集積していても，ベイズ学習による，事前密度が事後密度に変換される際の，「修正の度合」が，後実験的な尤度によって表現されている事を注意する。

「未知固定なる真の母数」を含む母数空間上に導入される主観確率に対して，その主観確率が定める事前分布と，「真の事前分布（true prior）」（同書第4. 2. 2節，97頁下から7行目），との隔りを考える事は，方法論的に意味を持たない（ベイズ学習の観点と，方法論的に矛盾する）ばかりか，Berger 自身が古典的な統計学の欠点の一つとして明示する，「定式化の不十分性（formulational inadequacy）」（同書第1. 6. 1節，III，21頁から23頁），を，より複雑な形で，ベイズ分析内部に引き起こす事になる。実際，「未知固定なる『真の母数』」式の定式化を用いずに，「母数達に関する知識の状態」を「母数空間上の確率測度」として定式化する事が，実践的であるか哲学的であるかにかかわらず，ベイズ学習の思想を重視するベイジアン達の，共通の「規範」なのである。（なお，「“true prior”からの隔り」を考慮した「ベイズ分析の為の手順（procedure）」，の，単純な一例を次節で与えるが，この様な「擬ベイジアン（pseudo-Bayesian）」としか言い様のない定式化が，方法論的な力を生じ得るとは，筆者には思えない。）

節の終わりではあるが，Berger［1］に対する簡潔且つ透徹した書評，鈴木［10］，が在る事を注意する。（同書評の最後の段落における指摘と主張を，筆者も亦，支持せざるを得ない。）

6. 第5節の付録

（この節の議論は Sono［7］，第3章，に基づく。）以下の2つの定義を用いる。

［定義1］ 母数空間 $\Theta=[0, 1]$ 上の事前分布の c. d. f.（累積分布関数）$P(\theta)$ が，

$$P(\theta)=\sum_{m=1}^{M}\left(P(\theta_{m-1})+\frac{P(\theta_m)-P(\theta_{m-1})}{\theta_m-\theta_{m-1}}(\theta-\theta_{m-1})\right)\cdot I_{A_m}(\theta),$$

（但し，$A_m=]\theta_{m-1}, \theta_m]$，$0=\theta_0<\cdots<\theta_{m-1}<\theta_m<\cdots<\theta_M=1$，且つ，一般に，$I_A(\theta)=$「$\theta\in A$ ならば1，$\theta\notin A$ ならば0」，とする。），と，表現されるのならば，この事前分布を，分点（θ_m; $m=0, \cdots, M$）を持つ台形事前分布（trapezoidal prior），簡略して，T. P. $(\theta_m)_{m=0}^{M}$，と呼ぶ。

［定義2］ 決定者（decision maker）の Θ 上の「真の事前分布」の c. d. f. を $P(\theta)$ とする。Θ 上の事前分布の c. d. f. $\hat{P}(\theta)$ が，T. P. $(\theta_m)_{m=0}^{M}$ であり，且つ，任意の $\theta\in\Theta$ に対して

$|\hat{P}(\theta)-P(\theta)|\leqq\delta$,

(但し，$\hat{P}(\theta_m)=P(\theta_m)$, $m=0, \cdots, M$, 且つ, $\delta\epsilon]0, 1[$, とする。)，を満たすならば，この事前分布を，誤差 δ, 分点 $(\theta_m)_{m=0}^{M}$, の主観的台形事前分布 (subjective trapezoidal prior)，簡略して，S. T. P. $(\delta, (\theta_m)_{m=0}^{M})$, と呼ぶ。

次の仮定 A1, A2, 及び, A3, に基づいて S. T. P. が構成される。

A1.　決定者の真の事前分布の c. d. f. $P(\theta)$ は狭義単調増加の Θ 上の連続関数であり $P(0)=0$ をみたす。

A2.　決定者は，2点 $\alpha, \beta, \alpha<\beta$, 及び値 $P(\alpha), P(\beta)$, が既知である場合，また，その場合に限って，$P(\gamma)=(P(\alpha)+P(\beta))/2$ をみたす点 $\gamma\epsilon]\alpha, \beta[$ を決定する事が出来る。

A3.　決定者は，2点 $\alpha, \beta, \alpha<\beta$, 値 $P(\alpha)$, $P(\beta)$, が既知である場合，また，その場合に限って，与えられた $\delta\epsilon]0, 1[$ に対して，「任意の $\theta\epsilon[\alpha, \beta]$ に対して，

$$|P(\theta)-(P(\alpha)+\frac{P(\beta)-P(\alpha)}{\beta-\alpha}(\theta-\alpha))|\leqq\delta」,$$

なる不等式の真偽を判定出来る。

［構成］ (第0段階)。T. P. $(\theta(0;m);m=0, \cdots, M(0))$ である $\hat{P}(\theta;0)$ を，$M(0)=1$, $\theta(0;0)=0$, $\theta(0;1)=1$, $\hat{p}(0;0)=0$, $\hat{p}(0;1)=1$, と置く事により，$\hat{P}(\theta;0)=\theta$, $\theta\epsilon\Theta$, と定義する。

任意の $l\epsilon\{0, 1, 2, \cdots\}$ に対して，「T. P. $(\theta(l;m);m=0, \cdots, M(l))$ である $\hat{P}(\theta;l)$ が与えられるのならば，$\hat{p}(l;m)=\hat{P}(\theta(l;m);l)$, $m=0, \cdots, M(l)$, と置き (第 $l+1$ 段階) へと進む。」

(第 $l+1$ 段階)。初めに $\theta(l+1;0)=0$, $M(l+1;0)=0$, と置く。任意の $m\epsilon\{0, \cdots, M(l)-1\}$ に対して，「不等式，

(＊)　任意の $\theta\epsilon[\theta(l;m), \theta(l;m+1)]$ に対して $|P(\theta)-\hat{P}(\theta;l)|\leqq\delta$,

が真であると判定されるのならば，

$$M(l+1;m+1)=M(l+1;m),$$
$$\theta(l+1;m+M(l+1;m)+1)=\theta(l;m+1),$$
$$\hat{p}(l+1;m+M(l+1;m)+1)=\hat{p}(l;m+1),$$

と定義し，不等式 (＊) が偽であると判定されるのならば，

$P(\theta(l;m,m+1))=(P(\theta(l;m))+P(\theta(l;m+1)))/2$, を満たす点 $\theta(l;m,m+1)$ を決定し，且つ，

$$M(l+1;m+1)=M(l+1;m)+1,$$
$$\theta(l+1;m+M(l+1;m)+1)=\theta(l;m,m+1),$$
$$\theta(l+1;m+M(l+1;m)+2)=\theta(l;m+1),$$
$$\hat{p}(l+1;m+M(l+1;m)+1)=(\hat{p}(l;m)+\hat{p}(l;m+1))/2,$$
$$\hat{p}(l+1;m+M(l+1;m)+2)=\hat{p}(l;m+1),$$

と定義する。」終わりに，

$$M(l+1)=M(l)+M(l+1;M(l)),$$

と置き，且つ，

$$\hat{P}(\theta;l+1)=\sum_{m=1}^{M(l+1)}(\hat{p}(l+1;m-1)+\frac{\hat{p}(l+1;m)-\hat{p}(l+1;m-1)}{\theta(l+1;m)-\theta(l+1;m-1)}\cdot(\theta-\theta(l+1;m-1)))I_{A(l+1;m)}(\theta),$$
$$A(l+1;m)=]\theta(l+1;m-1), \theta(l+1;m)],$$
$$\hat{p}(l+1;0)=0,$$

と定義する。

もし $M(l+1;M(l))=0$ ならば，手順を停止して $\hat{P}(\theta)=\hat{P}(\theta;l)$, $\theta\epsilon\Theta$, と定義する。これが S. T. P. $(\delta, (\theta(l;m);m=0, \cdots, M(l)))$ である。［構成終］

以下の3つの性質，P1，P2，P3，が容易に確認される。

P1．上記の構成の手順は，仮定 A1，A2，A3，に基づいて，決定者によって主観的に遂行される。

P2．$M(l+1; M(l))=0$ の場合，またその場合に限って，任意の $\theta\epsilon\Theta$ に対して $\hat{P}(\theta; l)=\hat{P}(\theta; l+1)$ が成立する。

P3．$l(\delta)$ を $2^{-l}\leqq\delta$ をみたす最小の非負整数とするならば，$M(l(\delta)+1; M(l(\delta)))=0$ となる。

従って，上記の手順はよく定義されており，且つ，有限段階で停止する。

以下の2例において，S. T. P. を単純なベイズ決定問題に応用し，且つ，「真のベイズ危険」と近似的ベイズ危険との間の誤差を粗く評価する。

［例1］ 有限標本空間 X とその上の密度 $f(x|\theta)$, $(\theta, x)\epsilon\Theta\times X$, を考える。$l(\theta|x)=f(x|\theta)$ と置き，且つ，尤度を単峰形と仮定して，最尤推定量を $\hat{\theta}(x)$ と書く。$A=[0, 1]$ を行為空間 (action space),

$$L(\theta, a)=K^{+}|\theta-a|\cdot I_{[\theta, 1]}(a)+K^{-}|\theta-a|\cdot I_{[0, \theta]}(a),\ K^{+}, K^{-}>0,$$

を損失関数，とする。$P(0)=\hat{P}(0)=0$ を満たす任意の連続な c. d. f.' s, $P(\theta)$ 及び $\hat{P}(\theta)$，に対して，部分積分法により，以下の不等式を得る。

$$\begin{aligned}
&|\Sigma_{x\epsilon X}\int_{\Theta} L(\theta, a(x))l(\theta|x)d(\hat{P}(\theta)-P(\theta))|\\
&\leqq\Sigma_{x\epsilon X}\int_0^1|\hat{P}(\theta)-P(\theta)|\cdot(L(\theta, a(x))\cdot|dl(\theta|x)|+l(\theta|x)\cdot|dL(\theta, a(x))|)\\
&\leqq\delta K\Sigma_{x\epsilon X}(2\cdot l(\hat{\theta}(x)|x)-l(0|x)-l(1|x)+\int_0^1 l(\theta|x)d\theta)\\
&\leqq\delta K(2\cdot\Sigma_{x\epsilon X}\ l(\hat{\theta}(x)|x)-1),
\end{aligned}$$

但し，$\delta=\max_{\theta\epsilon\Theta}|\hat{P}(\theta)-P(\theta)|$，$K=\max(K^{+}, K^{-})$，且つ，$a(\cdot)$ を任意の決定関数とする。

故に，次の粗い評価 (♯) を得る。

$$(\sharp)\quad |r(\hat{P})-r(P)|\leqq\delta K(2\Sigma_{x\epsilon X}l(\hat{\theta}(x)|x)-1).$$

(但し，一般に，事前分布 Q に対するベイズ危険を $r(Q)$ と書く事にする。) $P(\theta)$ を「真の事前分布」とする場合，近似的事後分布 $\hat{P}(\theta|x)$ の $K^{+}/(K^{-}+K^{+})$ 分位点が，x に対する近似的なベイズ決定となる。

仮定 A1，A2，A3，の下で，決定者が (♯) の左辺が $\varepsilon>0$ を超えない事を要求するのならば，$\delta\leqq\varepsilon/(K(2\Sigma_{x\epsilon X}l(\hat{\theta}(x)|x)-1))$ なる δ に対して，上記の［構成］の手順に従って，S. T. P. $\hat{P}$ を構成する事で決定者の要求を満たす事が可能である。この T. P. $(\theta_m)_{m=0}^{M}$ に対する事後分布の c. d. f.，$\hat{P}(\theta|x)$，は次の様に表現される。

$\theta\epsilon[\theta_{m-1}, \theta_m[$ ならば，

$$\hat{P}(\theta|x)=\frac{A_m S(\theta)+\Sigma_{k=1}^{m-1}(A_k-A_{k+1})S(\theta_k)}{A_M+\Sigma_{k=1}^{M-1}(A_k-A_{k+1})S(\theta_k)},$$

$m=1, \cdots, M$. 但し，$A_k=\dfrac{\hat{P}(\theta_k)-\hat{P}(\theta_{k-1})}{\theta_k-\theta_{k-1}}$,

$k=1, \cdots, M$, 且つ,

$$S(\theta)=\frac{\int_0^{\theta}l(\theta|x)d\theta}{\int_0^1 l(\theta|x)d\theta},$$

とする。

［例2］ (S. T. P. を多次元の母数空間に対して直接あてはめる事は出来ないが，特殊な場合にはそれが可能となる。) 母数 $\theta=(\theta_i)_{i=1}^{p}\epsilon\Theta^p$ を持つ，有限標本空間 X 上の密度 $f(x|\theta)$ を考える。$l(\theta|x)=f(x|\theta)$ と置き，$l(\theta|x)=\Pi_{i=1}^{p} l(\theta_i|x)$，(但し，各 $l(\theta_i|x)$ は $(\theta_i, x)\epsilon\Theta\times X$ のみに依存する非負の関数であり，唯一の極大点 $\hat{\theta}_i(x)$ を持つものとする)，と分解する事が可能である，と，仮定する。(この様な場合には，事前独立性の仮定が役に立つかもしれない。)

連続な周辺事前分布達，$P_i(\theta_i)$ 及び $\hat{P}_i(\theta_i)$，$i=1, \cdots, p$，と，損失関数 $L(\theta, a)=\Sigma_{i=1}^{p} L_i(\theta_i, a_i)$，但し，$a=(a_i)_{i=1}^{p}\epsilon[0, 1]^p$ 且つ

$L_i(\theta_i, a)=K_i^+(a_i-\theta_i)I_{[\theta_i, 1]}(a_i)+K_i^-(\theta_i-a_i) I_{[0, \theta_i]}(a_i)$，$K_i^+$，$K_i^->0$，$i=1, \cdots, p$，とする，を考える。この時，部分積分法，及び，

$\Pi_{i=1}^{p}\hat{P}_i(\theta_i)-\Pi_{i=1}^{p}P_i(\theta_i)=\Sigma_{j=1}^{p}(\Pi_{i=1}^{j-1}\hat{P}_i(\theta_i)) (\hat{P}_j(\theta_j)-P_j(\theta_j)))\Pi_{i=j+1}^{p}P_i(\theta_i))$，を用いると，［例１］の不等式（♯）に対応する次の結果が得られる。

$$|r(\Pi_{i=1}^{p}P_i)-r(\Pi_{i=1}^{p}\hat{P}_i)|\leqq (\Sigma_{i=1}^{p}K_i)(\Sigma_{i=1}^{p}\delta_i)(2(\Sigma_{x\epsilon X}l(\hat{\theta}(x)|x)-1)+\frac{\Sigma_{i=1}^{p}\ \delta_iK_i}{(\Sigma_{i=1}^{p}\delta_i)(\Sigma_{i=1}^{p}K_i)}),$$

但し，$K_i=\max(K_i^+,\ K_i^-)$，$\delta_i=\max(|\hat{P}_i(\theta_i)-P_i(\theta_i)|;\ \theta_i\epsilon\Theta)$，$i=1, \cdots, p$，且つ $\hat{\theta}(x)=(\hat{\theta}_i(x))_{i=1}^{p}$ とする。この不等式に基づいて［例１］と同様な議論を行なう事が出来る。

7.　註

註1.　第４節の［証明］の中で言及した，「Chittenden の距離化定理」なる呼称は，ケリー (Kelley)［３］，第６章，「13　距離化定理」(190頁)，「14　註」(191頁)，及び，同章の章末問題の内の「Ｉ」の註 (212頁)，より借りたものである。

註2.　第５節で引用した，(Berger［１］，第4. 5. 2節，例15，123頁から124頁，における）事後密度 $\pi(\beta|x)$ を最大にする β を $\hat{\beta}_0$ とすると，$\hat{\beta}_0=x/(\alpha+1)$ となるが，同書同例で示されている様に，平均自乗誤差をベイズ危険とした場合の，$\delta_c(x)=cx$ なる型のベイズ決定の内の「最適」なものが $\hat{\beta}_0$ と一致する。一方，この論文の第５節では，$\hat{\beta}=\frac{x}{\alpha}$，なる推定値（形式的には最尤推定値と一致する）が Box-Tiao 流の客観的ベイズ推論から支持され得る事を注意したが，この $\hat{\beta}$ は事後密度の比率 $\pi(\log\beta|x)/\pi(\log\hat{\beta}|x)$ に基づく客観的ベイズ推論の為に導入される概念であり，所謂「数学的最適化」とは直接的には連関の無い，ベイズ統計学的概念である。

註3.　Box and Tiao［２］の参考文献の表の中で，578頁冒頭から４番目まで Savage, L. J., に関する文献が引用されているが，第１番目，第３番目，及び第４番目，の文献の表題の各各に現われる“Foundation”なる単語は“Foundations”と修正すべきものである。また Savage［５］の付録2，269頁上から４行目（但し，３行目は空）に太文字で“**Schwartz inequality**”と在るがこれは太文字で“**Schwarz inequality**”とすべきものである。

註4.　第４節で前実験的模型の選択に触れたが，（より重要な）後実験的模型の選択に対しては，園［８］に簡単な論述が在る。

参考文献

［１］Berger, J. O., Statistical Decision Theory (Foundations, Concepts, and Methods), Springer-Verlag, New York, 1980.

［２］Box, G. E. P. and Tiao, G. C., Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Massachusetts, 1973.

［３］ケリー (Kelley, J. L.), 児玉之宏 訳，『位相空間論 (General Topology)』，吉岡書店，京都，1968.

［４］宮沢光一，『情報・決定理論序説』，岩波書店，東京，1971．(この本の奥付では，「宮沢」の上に「みやざわ」なる振り仮名が在るが，例えば，同書507頁，参考文献［8］，598頁，参考文献［2］，には“Miyasawa, K.”と在るので，「みやさわ」と仮名を振るべきである，と，鈴木教授（東京大学）から注意されたので付記しておく。)

［５］Savage, L. J., The Foundations of Statistics, 2nd revised edition, Dover, New York, 1972.

［６］柴田敏男，『集合と位相空間』，共立出版，東京，1972.

［７］Sono, S., “On Bayesian inference and Bayes decision problems,” 理学博士学位論文（東京大学，昭和60年５月），1985.

［８］園信太郎，「模型の選択に対するベイズ方法論的演繹について」，経済学研究（北海道大学）第35巻第２号，21(197)-27(203)，９月，1985.

［９］鈴木雪夫，「異常値問題について」，『経済学論集』，第46巻第3号，2-15，10月，1980.

［10］鈴木雪夫，上掲 Berger［1］の書評，『経済学論集』，第47巻第3号，111-115，10月，1981.

(1985年，10月15日（火），脱稿)