Bayesian Theory and Statistical Models

ベイジアン理論と統計モデル

平　館　道　子

1．予測的理論

主観確率にもとづくベイジアン統計理論の中で，B. de Finetti は予測的理論（predictive theory）と呼ばれている理論を展開している（de Finetti (1964, 72, 74)）。一般にベイジアン理論の立場に立つ統計家は観測可能な変量Xのパラメータ θ を含む確率分布 $p(x \mid \theta)$（尤度関数とも呼ばれている）と θ の事前分布 $p(\theta)$ とを用いて，ベイズの定理により観測値が得られた後の事後分布 $p(\theta \mid x)$ を求め，それが θ に関する当初の不確実性の判断の，データ情報を得たのちの判断への変換を示すものとして，推測，予測，意思決定等の目的に使用する。これに対して de Finetti は，不確実性の判断は観測可能な変量に対してのみ下されるものであり，パラメータのような現実的な意味のはっきりしないものに対して下されるものではないとして，不確実性に対する判断が満足すべき条件について coherence（整合性）理論を展開している。coherence の１つの形式は次のようなものである。いま k 個の互いに素であり，すべてをつくす事象 E_1，E_2，……，E_k があり，これらはそれぞれインディケータであるとする。これらの事象に対して確率 π_i，$i=1$，……，k を付与するとき，いずれかの事象が起こったのち，賞金 $\sum_{i=1}^{k} c_i(E_i-\pi_i)$ を受け取るとする。c_i は任意に選択された定数である。ここで coherence とは，どの事象が起ろうとも賞金が常に負になるということがないように π_i を付与すべきである，ということである。この条件を満たす $\{\pi_i\}$ は［0，1］内にあり，（有限）加法定理に従うことが導びかれる。この de Finetti の定義に対してはいくつかの批判があり，strict coherence の原理（Carnap (1971)）などが述べられているが，ここでは触れない。

この理論に従って，一般に観測可能量の系列 x_1，x_2，……に対して，その任意の部分系列の同時主観確率分布を規定する確率測度Ｐが存在するものとしよう。Ｐは predictive モデルあるいは predictive distribution とも云うべきものであり，通常の数学的確率モデルとは異り，未知のパラメータを含んでいない。以下では数学的には厳密でないかもしれないが，Ｐは根底にある確率測度を意味したり，あるいはそれから導かれる分布関数を示すものとしたりして用いることにする。またｐを密度関数（あるいは離散型の場合はマス関数）とする。

いま $x_p=(x_1, \cdots, x_n)$ を既に観測された変量，$x_f=(x_{f_1}, x_{f_2}, \cdots, x_{f_m})$ をまだ観測されていないものとすると，

$$p(x_f|x_p)=\frac{p(x_p,\ x_f)}{p(x_p)} \tag{1}$$

が成立つ。この条件つき確率分布は“経験からの学習”を示すものであり，ベイジアン理論にとって鍵となるものである。もし，

$$p(x_f|x_p)=p(x_f) \tag{2}$$

であるならば，それは経験からは何も学ばないという判断を意味している。

ところで，P(x) は直接的に付与されることもあるが，一方従来のように尤度関数とパラメータの事前分布を用いて，

$$p(x)=\int_\Theta p(x\mid\theta)dQ(\theta) \tag{3}$$

のようにして求めるというアプローチもある。ここに Θはパラメータ θの空間である。

ごく単純な場合を除いて，$P(x)$ を直接に付与するということは極めて困難な仕事である。さまざまな次元の周辺分布についても納得できるようにしなければならないが，これは大変なことであり，そこで，尤度関数と事前分布の２つに分けて迂回的に $P(x)$ を求めるアプローチがしばしばとられる。ところで実際に統計分析を行う際に経験することであるが，一つの状況に対

して一つのモデルしか考えられない，というようなことはあまりなく，いくつかの代替モデルが考慮されるものであろう。そこで異るモデルと事前分布とから同じ $P(x)$ が得られるということもあり得る。それでは，もし de Finetti の coherence 理論を真剣に受けとめるならば，多くの場合に受け容れられているベルヌーイモデルや正規モデル等はどのような意味を持つのであろうか。

2．パラメトリック・モデル

統計家はさまざまな統計モデルを取扱う。伝統的には，それらは実験や調査の結果のような観測可能な量を支配する確率法則を記述するものであり，パラメータに依存し，客観的な物理的実体を表現しているものと想定される。そしてパラメータは観測可能量の可能な分布の族を指定し，その中で，パラメータの未知ではあるが特定の値に対応する分布が“真”の分布である。従って統計的推測の目的は，通常，観測可能量の実現した値に照して，この未知な真の分布に関する提言を行うことである，と考えられている。観測可能量が各々の値をとる確率，あるいは確率密度は，パラメータの値がどれだけであるかに依存しており，パラメータの値が未知であるかぎり，それらの確率もまた未知である。確率は未知のパラメータの関数である。このように伝統的理論では，パラメータと観測可能量とは明確に区別されている。このような区別の現実的な根拠は，結局のところパラメータは“観測可能ではない”ということであろうが，パラメータ空間が指定されるように，可能な値の領域が想定され，そこから一つの値が実現すると考えられることからみても，この区別はそれほど明確なものではないように思われる。観測可能でないということは，パラメータはその値がある確率分布に従って実現するという想定を否定する説得的な理由にはならないであろう。また逆に，ある観測可能量の値が所与のとき，他の観測可能量の条件つき分布が意味ある形で定められることもあり，そのような場合，この観測可能量をパラメータとして取扱うこともできる。回帰モデルにおける説明変数，実験計画における共変数の場合などである。

3．ベイジアン理論

ベイジアン理論においてはこのような区別はされない。確率は不確実な量に対する個人の判断を示すものであるから，観測可能量も未知のパラメータもすべて不確実な量であり，これらすべての不確実量の同時分布の存在が前提される。未知の確率というようなものはないのである。ただ一つの相異は，観測可能量については，ベイジアンは賭を設定できるということである。

観測可能量とパラメータの同時分布は，しばしば次のように分解されて，不確実性がそれぞれ別個に判断される。

$$p(x,\ \theta)=p(x \mid \theta)p(\theta). \tag{4}$$

式(4)の右辺の最初のものが尤度関数で，確率モデルに相当し，後者が事前分布である。ベイジアン統計家にとっては尤度関数に関しては判断の根拠として物理的実体を根底に踏えていながらも，それを体現しているとは考えず，両者とも自分の不確実性に対する判断を示している，という点で区別はない。この場合，predictive 分布 $p(x)$ は式(4)から

$$p(x)=\int_{\theta}p(x \mid \theta)p(\theta)d\theta \tag{5}$$

のようにして求められることは既に述べた通りである。もし2人の統計家がそれぞれモデルと事前分布の組において異る組合せを付与したとしても，もし predictive 分布が同じであるならば，2人は観測可能量 x に関して全く同じ判断を下していることになり，どちらかのモデルが真で，他が真でない，というようなことは全くない，と云わなくてはならない。観測可能量に関する限り，どちらのモデル－事前分布の組から得たものであろうと，これをそれ以後の目的のために同等に使用できる。

ランダムな量とランダムでない量との確定的な区別をしなければ，同じ状況をいくつかのモデルによって表現できる場合が多い(Dawid(1986))。そうであるならば，あるモデルと同値なモデルを構成する際の，分析技術的な観点からの重要な配慮は，パラメータが与えられた時，各観測値が独立である，ということであろう。勿論これは本質的なことではないが。Dawid(前掲書)

の例の一つでは，離散的な時系列に関する当初の標準的モデルはエルゴード性をもたず，従って無限の観測値を得ても一致性のある推定ができないという問題をもっているが，観測可能量とパラメータの2分割をしなければ，それを全く同値な独立性モデルに変換できることが示されている。

このように一般的には，ベイジアン統計家はモデル－事前分布の組を，分析上の便宜からも選ぶ自由を持っている。しかし，これはまた云い換えれば，ベイジアンは個人としてはモデルと事前分布に分解する一意的な方法を持たないということである。それではベイジアンにとって統計モデルとは何であるのか。この問題に対して，de Finetti は不確実性に関する判断の exchangeability（交換可能性）という概念を導入して，Representation Theorem（表現定理）として知られている意味深い結果を提示している。Savage は symmetry（対称性）と呼んでいる。

注　exchangeability, Representation Theorem, predictive distribution, coherence 等についてはまだ日本語の定訳がないので，このまゝ使用する。

4．Exchangeability

あるベイジアン統計家Bが predictive distribution を付与するとしよう。式(3)を再録すると，

$$p_B(x)=\int_\Theta p_B(x \mid \theta)\, dQ_B(\theta) \tag{6}$$

のように，個人のインデックスがそれぞれ付けられることになろう。もし $p_B(x \mid \theta)$ が多くの統計家にとって受容可能なものであるならば，(6)は

$$p_B(x)=\int_\Theta p(x \mid \theta)\, dQ_B(\theta) \tag{7}$$

のように表現できるであろう。このような場合には，ベイジアンBはモデルと事前分布を区別する1つの意味のある方法を得た，ということができるだろう。Dawid はこのようなモデルを intersubjective モデルと呼んでいる (Dawid (1982))。また(7)が成立つ場合，ある意味でモデルは“客観性”を得

たと云うこともできるであろう。(7)は，モデルは“客観的”であるが，個々の統計家の判断の相異の余地は残されており，それは Q(・) で表現されることを示している。

“客観性とは”，という問題を論じることは私の手に余るが，少なくとも，現実に対する判断に大きな偏りがなく，多くの人に受け容れられること，と考えて大きな誤りはないであろう。しかし実際の分析の際には，特に社会，経済分野では，構造モデルなどに関してかなりの意見の相異があり，また参照すべき理論もいくつかある，というような状況は多い。これはモデルが個人の状況に対する認識を定式化したものと考えれば当然のことであり，また多くの人には受け容れられないものであっても，それだからこそ認識を深めるということもあり得る。またベイジアン理論の視点からは，整合的な判断を行うことが要請されるが，そのことと，それが現実と適合している，ということとは別の問題であり，常に現実との照合が求められる統計分析においては，判断を現実の観測値に照して吟味することは，客観性という点からも重要である。

predictive 分布を付与するための重要な概念として，de Finetti は exchangeability を次のように定義している。(de Finetti (1972, 74))。

無限exchangeability

不確実量の無限系列 x_1，x_2，……があり，ある確率測度Pの下で，そのすべての部分有限系列 x_{f_1}，x_{f_2}，……，x_{fn} の同時分布が，(1，…，n)上のすべての順列 π に対して

$$P(x_{f_1},\ \cdots,\ x_{fn})=P(x_{\pi_1},\ \cdots,\ x_{\pi n})$$

を満足する場合，無限系列は無限 exchangeable である。

exchangeabilityが成立つ場合，0－1変数に対して，有名な Representation Theorem が証明されている（証明については，de Finetti (1930, 1937, 1964)，Heath－Sudderth (1976) 参照)。

Representation Theorem

もし x_1, x_2, …が無限 exchangeable な 0 － 1 変量であるならば，その部分系列 x_1，…，x_n に対して同時確率関数 $p(x_1\cdots x_n)$ が次のように表現できる確

率分布Qが存在する。

$$p(x_1,\ \cdots,\ x_n)=\int_\Theta\prod_{i=1}^n\theta^{x_i}(1-\theta)^{1-x_i}dQ(\theta), \tag{8}$$

ここに　$T_n=x_1+\cdots+x_n,\qquad \theta=\lim_{n\to\infty}\frac{T_n}{n}$

$$Q(\theta)=\lim_{n\to\infty}P\left(\frac{T_n}{n}\leqq\theta\right)$$

である。

この定理のベイジアン統計家にとっての含意は極めて深いものであり，ベイジアン理論におけるパラメトリック・モデル構築に根拠を与えていると云える。すなわち，この定理は，観測可能０－１変量に対する exchangeable な判断は，a）x_i は θ が与えられる時，独立なベルヌーイ変量とみなすことができる，b）θ は１の相対頻度の極限である，c）θ には確率分布Qが存在し，それは値１の相対頻度の極限に対する不確実性の判断を示すものである，という３つの含意をもつ，ということを示している。云い換えれば，パラメータ θ が与えられる時，x_i はベルヌーイ分布からのランダム・サンプルであり，θ には事前分布が存在する，ことを意味する。これは式(7)のような表現が実現し，個人のインデックスがない $p(x|\theta)$ に対応するものは，

$$p(x\mid\theta)=\prod_i p(x_i\mid\theta)=\prod_i\theta^{x_i}(1-\theta)^{1-x_i}$$

のようにベルヌーイモデルによるランダム・サンプルにもとづく同時分布として表現できることを述べている，と云えるであろう。

このようにして，いわゆるベルヌーイ独立試行に関する伝統的確率モデルと任意の事前分布による表現が，観測可能量に対する exchangeable な判断という概念によって根拠を与えられたのである。さらにパラメータの意味を明らかにし，事前分布とパラメータは不可分のものであることを示している。

その後の研究によって０－１変量だけでなく，一般的に実数値をとる不確実量についても，無限 exchangeability が成立つ場合には，同様な結論が得ら

れている（Chow－Teicher（1978／1988））。

すなわち，**一般 Representation 定理**によれば，

R上のすべての分布関数の空間を**F**とし，Qを**F**上の確率測度，F_n を x_1，…，x_n によって定められる経験的分布関数とすると，

$$P(x_1, \cdots, x_n)=\int_F \prod_{i=1}^{n} F(x_i)dQ(F) \tag{9}$$

となるようなQが存在する。ここに

$$Q(F)=\lim_{n\to\infty} P(F_n)$$

である。

この場合Fが無限次元のパラメータとなっているが，それがどのようなタイプのものであるかについては，さらにあるタイプの判断の仕方について規定する必要がある。ここでは，正規分布モデルを導く spherical symmetry (球面対称性）についてだけ述べることにしよう。

spherical symmetry

確率測度Pの下で観測可能量 $\boldsymbol{x}=(x_1, \cdots, x_n)'$ が spherical symmetric であるとは，$\boldsymbol{x}$と，ある直交行列Aによる変換 $A\boldsymbol{x}$ とが同一の同時確率分布をもつことである。

直交変換はベクトルの長さを保つ変換であるから，spherical symmetric な判断とは2乗和 $x_1^2+\cdots+x_n^2$ が一定の $\boldsymbol{x}$の値に対して同じ確率を付与することを意味する。順列は直交変換の特別な場合であるから，exchangeability も spherical symmetry の特別な場合である。

この場合，一般 Representation 定理のFは正規分布 N(0, σ^2), $\sigma^2=\lim_{n\to\infty}\frac{\Sigma x_i^2}{n}$ であることが導かれる（Bernardo-Smith（1994），Dawid（1977，1978））。一般の（μ，σ^2）をパラメータとする正規分布モデルを導くには，$\boldsymbol{x}$のかわりに，（$x_1-\bar{x}_n$，…，$x_n-\bar{x}_n$），$\bar{x}_n=\frac{\Sigma x_i}{n}$ が spherical symmetric である，すなわち $\sum_i (x_i-\bar{x}_n)^2$ が一定である点にすべて同じ確率が付与される（中心化球面

対称性）とすればよい。この時，

$$\mu=\lim_{n\to\infty}\bar{x}_n, \qquad \sigma^2=\lim_{n\to\infty}\frac{1}{n}\Sigma(x_i-\bar{x}_n)^2$$

であり，

$$Q(\mu,\ \sigma^2) \quad = \quad P[(\bar{x}_n<\mu)\cap(s_n^2<\sigma^2)]$$

である。$s_n^2=\frac{1}{n}\Sigma(x_i-\bar{x}_n)^2$ である（Bernardo－Smith（1994））。

このようにして正規分布が根拠づけられるわけであるが，云うまでもなく，状況によってはもっと複雑な判断を下さなければならないであろう。exchangeability から乖離した判断はさまざまにあり得るが，ここでは経済データの統計分析にもっとも使用される線形回帰モデルについて考察しよう。

線形回帰モデル

観測可能不確実量 $\boldsymbol{x}=(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n)$ があるとしよう。ここに $\boldsymbol{x}_i=(x_{i1},\ \cdots,\ x_{il})$ である。$\boldsymbol{x}$ の不確実性は他の変量 $\boldsymbol{z}=(\boldsymbol{z}_1,\ \cdots,\ \boldsymbol{z}_n)$，$\boldsymbol{z}_i=(z_{i1},\ \cdots,\ z_{ip})$ の観測値に依存していると判断される。$\boldsymbol{z}$ は共変量，説明変数などと呼ばれている。この依存性を考慮して $\boldsymbol{x}$ の同時確率分布を

$$p(\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_n\,|\,\boldsymbol{z}_1,\ \cdots,\ \boldsymbol{z}_n)$$

と書く。これがさらに次のように書けると判断されるとする。すなわち，$\boldsymbol{x}$ は変量の各グループ $\boldsymbol{x}_i$ 内において無限 exchangeable であり，かつ，

$$p(\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_n\,|\,\boldsymbol{z}_1,\ \cdots,\ \boldsymbol{z}_n)=\prod_{i=1}^{n}p(\boldsymbol{x}_i\,|\,\boldsymbol{z}_i) \tag{10}$$

と書くことができる。これを部分的 exchangeability と呼ぶことにしよう。さて $\boldsymbol{z}$ への依存性であるが，$p(\boldsymbol{x}_i\,|\,\boldsymbol{z}_i)$ における $\boldsymbol{x}_i$ のグループ内の平均 $E(\boldsymbol{x}_i)$ が $\boldsymbol{z}_i$ の一次関数として表現され，この関係は $\boldsymbol{x}_j$, $\boldsymbol{z}_j$ 等のグループにおいても保たれると判断される。

$\boldsymbol{x}$ 全体については式(10)が成立ち，各 $\boldsymbol{x}_i$ 内では中心化された spherical symmetry が成立つ，と判断される。このような判断の下で predictive distribution は次のように表現される。

$$p(x_1,\ \cdots x_n \mid z_1,\ \cdots,\ z_n)=\int \prod_{i=1}^{n} \prod_{j=1}^{\ell} N(x_{ij} \mid \mu_i(z_i),\ \sigma^2{}_i(z_i))dQ(\boldsymbol{\mu},\ \boldsymbol{\sigma}) \quad (11)$$

ここに $N(\cdot \mid \mu,\ \sigma^2)$ は正規分布の密度関数を，$(\boldsymbol{\mu},\ \boldsymbol{\sigma})=(\mu_1(z_1),\ \cdots,\ \mu_n(z_n),\ \sigma_1(z_1),\ \cdots,\ \sigma_n(z_n))$ をあらわす。さらに，

$$\mu_i(z_i)=\lim_{\ell\to\infty}\bar{x}_{i(\ell)},\quad \sigma_i^2(z_i)=\lim_{\ell\to\infty}s_{i(\ell)}^2$$

である。さらに z への線形的依存性という判断から

$$\mu_i(z_i)=\mu(z_i)=z'_i\boldsymbol{\beta}$$

のように表現されるであろう。すなわち

$$E(x_{ik})=E_\theta[E_x(x_{ik} \mid \theta)]$$

であるから，$E_x(x_{ik} \mid \theta)=\mu_i$　とすれば，

$$E(x_{ik})=E_\theta(\mu_i)=\mu_{0i}$$

である。ここに μ_{0i} は μ_i の事前平均である。したがって，$\mu_{0i}=\mu_0(z_i)$ と書けば線形性の判断は $\mu_{0i}=z'_i\boldsymbol{b}_0$ のように書ける。これはあるパラメータ $\boldsymbol{\beta}$ について，

$$\mu=z'\boldsymbol{\beta}$$

のように書けることを意味する。$\boldsymbol{b}_0$ はしたがって $E_\theta(\boldsymbol{\beta})$，すなわち $\boldsymbol{\beta}$ の事前平均であり，統計家が付与した定数である。Representation 定理から，　$\mu=\lim_{\ell\to\infty}\bar{x}_{\ell}$ であるから，$\boldsymbol{\beta}$ は z の変化に対する $\bar{x}$ の変化，すなわち変化率の極限と考えられる。

さらに

$$\sigma_i^2(z_i)=\sigma^2(z_i)=\sigma^2$$

のような判断が加われば，通常の線形回帰モデルが得られる。

その他実験計画や構造に関する，正規モデルについても同様な形で導出することができ，また他のタイプの確率分布モデルについても，判断に関するある種の不変性を導入することによって導くことができる。

5．おわりに

Representation 定理は，統計モデル，パラメータ，ランダム・サンプルという統計理論を構成する基本的な要素に対して，主観確率の視点から新しい光をあて，観測可能な変量の不確実性を判断することと，統計理論の展開の中で樹立されて来た推測の方法論との関係を明らかにしている。de Finetti はパラメータのようなミステリアスな性格のものに対して不確実性の判断をするものではなく，あくまでも観測可能な量に対して行うものである，と云っているが，定理においてパラメータの大数法則的な性格を明らかにし，またパラメトリックな推測の位置づけを私たちに教えている。predictive な視点からはパラメトリックな推測は副産物であるかもしれないが，それは観測可能量に関する推測の極限的な形であり，それ自身固有な位置をもつと云えるであろう。本質的に，あるいは近似的にも，無限系列 x_1，…が想定できないような状況については，現在のところこのような根拠づけは得られていない。しかし，有限系列 x_1，…，x_n が任意の順列において

$$P(x_1,\cdots,x_n)=P(x_{\pi 1},\cdots,x_{\pi n})$$

である場合，x_1，…，x_n がＰの下で有限 exchangeable である，と定義すれば，例えば，０－１変量の有限母集団の場合，$T_N=x_1+\cdots+x_N$ とする時，尤度関数は T_N をパラメータとする超幾何分布で表現できるだろうから，T_N の正確な事後分布が得られる。この場合パラメータの意味は明確である。伝統的統計理論のようにランダムな量とそうでないものとの厳格な区別をしなければ，同値な複数のモデルを構成できることが示すように，主観確率論の

視点からはモデルはユニークなものではない。しかし，不確実性に対する判断を確率という形式で表現するという立場から考えれば，そのモデルにおいて不確実量が，それがパラメータであれ，どのような意味内容をもつものか，認識できなければならない。モデルとは自分の認識を定式化する有効な方法であると私は思うが，また他の統計家たちとの間で会話を成り立たせるための有力な手段でもある。判断の仕方は云うまでもなくさまざまにあり得るが，このような点から exchangeability はベイジアンモデル構築にとって必須の意味を持っていると云えるであろう。

文　献

Bernardo, J, M., A, F, M, Smith,(1994), "Bayesian Theory", Wiley.

Carnap, R. (1971),. "A Basic System of Inductive Logic(part I)", Studies in Subjective Logic and Probability, eds. R. Carnap, Berkeley, UC press.

Chow, Y. S., Teicher, H. (1978/88), "Probability Theory". Berlin, Springer, 2 nd ed. in 1988.

Dawid, A. P., (1977), "Invariant Distribution and Analysis of Variance Model", Biometrika **64**.

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory", J. Roy. Statist. Soc. B **41**.

Dawid, A. P. (1982), "Intersubjective Statistical Models", Exchangeability in Probability and Statistics, eds. G. Koch and F. Spizzichino, North-Holland.

Dawid, A. P. (1986), "A Bayesian View of Statistical Modelling", Bayesian Inference and Decision Techniques, eds. P. K. Goel and A. Zellner.

de Finetti, B. (1964), "Foresights : Its Logical Laws, Its subjective Sources". H. E. Kyberg, H. E. Smokler(eds.) Studies in Subjective Probability, Wiley.

de Finetti, B. (1972), "Probability, Induction and Statistics", Wiley.

de Finetti, B. (1974), "Theory of Probability", Wiley.

Heath, D. L., Sudderth, W. D.(1972), "On a theorem of de Finetti, oddsmaking and game theory". Ann. Math. Statist. **43**.

Heath, D. L., Sudderth, W. D. (1976) "de Finetti's Theorem for Exchangeable Random Variables", Amer. Statist. **30**.

Smith, A. F. M. (1981), "On a Random Sequences with Centered Symmetry", J. Roy. Statist. Soc. B **43**.