ベイズ統計学とその応用

鈴木雪夫・国友直人 編

$$p(\theta_i|x_j)=\frac{p(x_j,\ \theta_i)}{p(x_j)}$$

$$=\frac{p(x_j|\theta_i)\,p(\theta_i)}{\sum_{k=1}^{K}p(x_j,\ \theta_k)}$$

$$=\frac{p(x_j|\theta_i)\,p(\theta_i)}{\sum_{k=1}^{K}p(x_j|\theta_k)\,p(\theta_k)}$$

東京大学出版会

ベイズ統計学とその応用
鈴木雪夫・国友直人 編
東京大学

9 [本体3800円・税114円]
13-040108-4 C3033 P3914E

ベイズ統計学とその応用

鈴木雪夫・国友直人 編

東京大学出版会

はしがき

社会の情報化の進展とともに，あらゆる分野で統計学あるいは統計的方法にもとづく数量的分析がますます必要となっている．ちなみに編者が所属している社会科学の諸分野に限ってみても，経営学における意思決定分析，不確実性下の経済分析と計量経済分析，心理学における計量分析などをすぐに挙げることができるが，このような傾向は今後ますます盛んになることが予想されている．

他方，統計学者の立場からみるとこのような諸科学における必要に対応できるものとしては，ベイズ統計学を第一に挙げることができよう．ベイズ統計学は1950年代頃から統計学においては伝統的（しばしば非ベイズ的とも呼ばれる）立場にもとづく統計学に代わりうる一つの有力なアプローチとして注目されてきた．その後これまでに，ベイズ統計学と伝統的統計学の二つをめぐり，長い間，統計学の理論的・実際的諸問題について内外の学会を中心として論争が行なわれてきた．この間，ベイズ統計学の立場からはその理論的研究とともに多くの応用例が報告され，特に統計学における先進国である米国や英国では伝統的統計学に比肩するほどにベイズ統計学が着実に発展しているといっても過言でなかろう．わが国でも，東京大学においては1960年代始めからベイズ統計学の教育と研究が一貫して着実に進められてきたことは注目されよう．また最近わが国の統計学会でも，統計数理研究所を中心にして情報量統計学の立場からベイズ統計学についての独自の展開を見せているなどの動きも特筆に値する．このような統計学における研究の方向のなかでベイズ統計学は，一例として言及した社会科学をはじめ医学・工学を含む諸科学において必要とする理論的枠組みをそなえているとの共通認識が少なくとも専門統計学者の間ではかなり定着している．

ところが残念ながら日本では特に伝統的アプローチにもとづく統計学の教科書に比べ，ベイズ・アプローチによる統計学の教科書は非常に少ない．さ

らに諸科学における実際問題への統計学手法の適用とその理論的諸問題について，ベイズ統計学の立場から系統的に述べた書物は皆無に近い状態であろう．本書はこのような統計学における理論的動向と実際の統計分析のギャップを埋める一つの試みであり，ベイズ統計学の理論と応用を系統的に述べるとともに，ベイズ統計学をめぐるさまざまな問題を今日的視点より解説しようとするものである．

本書の作成にあたっては，まず以上に述べたような意識を共有している統計学者とその応用分野の第一線で活躍している研究者が一堂に集まり研究会議を開催することが企画された．各研究者はそれぞれがベイズ統計学とその周辺分野において日頃研究を重ねているが，中でもいわば得意とする問題が割当てられ各テーマについて報告原稿が作成された．その上で研究会議は昭和 63 年度文部省科学研究費（総合 A　63301078）の援助を得て 1988 年 9 月 5 日～7 日の 3 日間にわたり行なわれた．この研究会議は 30 名以上に及ぶ報告者・討論者を得て終始きわめて活発に「ベイズ統計学とその応用」をめぐって議論が行なわれたが，その概要については巻末にまとめておいたので参照されたい．研究会議の共通テーマはベイズ統計学であったが，各研究者が報告した内容はベイズ統計学の理論と応用についてのさまざまな問題にわたった．また報告者・討論者の研究関心および学問的背景はさまざまであるので，「ベイズ統計学と伝統的統計学」の相違についての考え方もさまざまであることは当然予想されていたのであるが，会議での報告と討論により相互の間の考え方の相違点や共通点がより明確に浮き彫りにされた．各研究者の考え方の違いはここに収録された諸論文からある程度うかがうことができよう．

本書の各章は上記研究会議における各報告者の論文に基づいている．しかしながら，研究会議の後に報告者自身が討論者・参加者のコメント等を参考にして原稿を修正あるいは加筆したので，多くの場合，最初の報告論文からは大幅に変更されている．また本書を「ベイズ統計学とその応用」についての教科書として利用可能なように，編者の一人により「ベイズ統計学の基本的考え方」に関する序章をあらたに付け加えてある．したがって，編者としてはベイズ統計学についてこれまであまり耳にしたことがない方は，まず「序章」を読み，その後に各自の問題意識に応じて各章を読まれることをすすめたい．各章で取り扱う内容を考慮して，全体を理論編と応用編に分けているので，ベイズ統計学における理論的問題に主たる関心のある読者には第 1 章～第 5 章，また社会科学を中心とするベイズ統計学の応用上の諸問題に関心のある方には第 6 章～第 10 章が参考となろう．

ここで内容を概観してみよう．第 1 章から第 3 章までは，ベイズ統計学での特有の考え方である主観確率をめぐる理論的問題を扱っている．新家論文はまず主観確率についてその歴史的経緯と考え方を解説している．松原論文は主観確率に基づくベイズ流の規範的分析と政治的行動を例として観察される実際の人間行動とのギャップに分析の焦点をあてている．赤池論文は情報量統計学の主たる提唱者として，事前分布にもとづくベイズ統計学の利用を新しい見地から提唱し，その応用例を解説している．つづいて第 4 章と第 5 章は，ベイズ統計学と伝統的統計学との理論的関連を扱った章である．竹村論文ではベイズ統計学に基づくベイズ解が，伝統的立場から考えても多くの場合に最適な統計的方法を与えていることを示している．竹内論文は伝統的統計学の立場からベイズ統計学のどの側面が評価でき，あるいは評価できないかといった論点をまとめている論争的内容である．

第 6 章と第 7 章は，ベイズ流統計分析の重要な二つの応用問題を扱っている．和合論文は動学的統計モデルを用いた予測問題をベイズ理論の立場から統一的に扱っている．また美添論文では社会科学のデータ分析でしばしば生じるやっかいな多重共線性の問題はベイズ理論を用いると自然に解けることを主張している．次に第 8 章～第 10 章では，ベイズ統計学の社会科学への応用上の問題を扱っている．国友論文は経済学におけるベイズ統計学の最近の応用について解説を加えるものである．青井論文は経営学における意思決定問題において生じる不確実性の評価の問題を例を用いて説明している．梅沢論文はベイズ統計学にもとづく意思決定理論を，経営意思決定など実際の意思決定分析に適用する際に生じるさまざまな現実的問題を整理している．

さて，以上で簡単に紹介した各章の話題は互いにある程度まで独立しているので，「序章」を読まれた後には特に順序を考えずに読むことが可能であろう．さらに編者としては，本書の各章の内容を互いに重複しないように調整し，全体として見るとベイズ統計学についての主要な問題が分かるように意

図したつもりである．このことがどれだけうまくいったかを含め，本書のさまざまな点については読者の御批判を仰ぎたい．本書を通じて，最近特に学会を中心として注目されているベイズ統計学について，その価値を読者が認識されるならば編者の喜びとするところである．

最後になるが研究会議及び本書は長年の間，東京大学経済学部および同大学院，また理学部，理学系大学院，工学系大学院においてベイズ統計学の研究と教育に尽力された鈴木雪夫教授の還暦記念をかねて企画されたことを記しておきたい．研究会議における報告・討論を通じて貴重な貢献をいただいた参加者各位，会議の実施に際して手伝っていただいた国友研究室の重田邦子さん，および本書の編集・出版についてお世話いただいた東京大学出版会の大瀬令子さんにこの場をかりて厚く御礼申し上げたい．

1989年3月

編　者

目　次

執筆者紹介（五十音順）

青井倫一　慶応義塾大学大学院経営管理研究科助教授，
経営学博士（ハーバード大学）

赤池弘次　統計数理研究所長，理学博士（東京大学）

新家健精　福島大学経済学部教授

梅沢　豊　東京大学経済学部教授

国友直人　東京大学経済学部助教授，
経済学 Ph. D.（スタンフォード大学）

鈴木雪夫　東京大学名誉教授・多摩大学教授，理学博士（東京大学）

竹内　啓　東京大学経済学部教授，経済学博士（東京大学）

竹村彰通　東京大学経済学部助教授，
統計学 Ph. D.（スタンフォード大学）

松原　望　東京大学教養学部教授，
統計学 Ph. D.（スタンフォード大学）

美添泰人　立正大学経済学経済学部教授，
統計学 Ph. D.（ハーバード大学）

和合　肇　筑波大学社会工学系講師

序　章
ベイズ統計学の考え方

鈴　木　雪　夫

統計的推論と統計的（意思）決定理論とは統計学の 2 つの理論的な核である．本序章では，ベイズ統計学における統計的推論と統計的（意思）決定理論の基本的な考え方について述べる．第 1 章以降ではベイズ統計学のいわば各論が各筆者によって展開されるのであるから，ここでベイズ統計学の基本概念を明らかにしておくことは十分に意味のあることと思う．

以下では，ベイズ統計学の立場での統計的推論および統計的（意思）決定理論を，単にベイズ推論およびベイズ意思決定理論と呼ぶことにしよう．

1.　ベイズ推論の基本的な考え方

1.1　事前分布，標本情報，事後分布

実験とか調査によって観測される確率変数を $\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n$ で表す．実験結果あるいは調査結果はこれらの確率変数の実現値 $x_1, x_2, \cdots, x_n$ にほかならない．これは標本情報ともいわれる．

確率変数 $(\tilde{x}_1, \cdots, \tilde{x}_n) \equiv \tilde{x}$ に関する数学的モデルは，実験あるいは調査の対象についての考察から導かれるのであるが，最終的には以下のように述べられるであろう．

観測される確率変数 $\tilde{x}$ のとりうる値のすべてからなる集合を $\mathcal{X}$ で表し，標本空間という．また，実験対象の状態を表す母数を θ とする．母数 θ の値は未知であるが，θ のとりうる値のすべてからなる集合を Θ で表し，母数

空間という．われわれの目的は未知母数 θ について推論することである．

さて，ベイズ推論の特徴は，推論主体（推論を行う人）が未知母数 θ についていだく不確実性を母数空間 Θ 上の確率分布として表現し，それを統計的推論に積極的に用いることである．Θ 上のこのような確率分布は，（$\tilde{x}$ を観測するという）実験の以前に評価・設定されるか，あるいは実験の以後に評価・設定されるかによって，事前（確率）分布あるいは事後（確率）分布といわれる．いずれにしても，推論主体は，それぞれの時点において利用可能な経験，知恵，知識，理論，情報等のすべてを活用して母数 θ の不確実性の程度についての総合的判断を事前分布あるいは事後分布として表現するのである．

われわれの統計的推論の数学的枠組は次のように述べられる．

(i)　θ についての事前分布が推論主体によって設定されるので，この分布に従う確率変数 $\tilde{\theta}$ を考える．$\tilde{\theta}$ の事前密度を，簡潔に，$p(\theta)$ で表す．（$\tilde{\theta}$ が離散的であれば，$p(\theta)$ は確率関数であるとする．）

(ii)　実験によって観測される確率変数 $\tilde{x}$ は，母数 θ が真のとき，$\mathcal{X}$ 上の密度関数 $p(x|\theta)$ をもつとする（$\theta \in \Theta$）．実験によって，確率変数 $\tilde{x}$ の実現値 x が得られたとき，実験結果（あるいは，標本情報）$\tilde{x}=x$ が得られたという．

(iii)　実験結果 $\tilde{x}=x$ を得たという条件の下での $\tilde{\theta}$ の条件付分布（すなわち，事後分布）の条件付密度関数（すなわち，事後密度）は，(i)，(ii)と条件付確率に関する公式

$$
\begin{aligned}
p(\theta|x) &= \frac{p(\theta,x)}{p(x)} \\
&= \frac{p(\theta)\,p(x|\theta)}{\int_\Theta p(\theta,x)\,d\theta} \\
&= \frac{p(\theta)\,p(x|\theta)}{\int_\Theta p(\theta)\,p(x|\theta)\,d\theta} \qquad (\theta \in \Theta) \qquad (1.1)
\end{aligned}
$$

によって求められる．この公式はベイズの定理と呼ばれているが，$\tilde{\theta}$ が離散的であれば，分母の積分を $p(\theta)>0$ であるような θ についての和に置き換えればよい．

ここに得られた事後密度 $p(\theta|x)$ は，推論主体が利用しうる経験，知恵，知識，理論，情報等にもとづき主体的判断によって選択した事前密度 $p(\theta)$ と実験結果 $\tilde{x}=x$ という新しい情報（標本情報）とをベイズの定理という数学的装置によって論理的に適正に結合したものである．したがって，事後分布は，実験後において，事前分布と（実験とか調査で得られた）標本情報とにもとづいた，推論主体の（母数 θ の不確実性の程度に関する）主体的な総合的判断の表現である．事後分布はこのような性質のものであるから，推論主体が自己の事後分布にもとづいて θ に関する推論を行うのはきわめて自然である．そこで，次節では事後分布から θ に関する種々のタイプの推論を導き出す仕方について述べる．

断るまでもなく，母数 θ の事前分布が推論主体の主体的判断によって選択されるのであるから，推論主体が異なれば選択される事前分布も異なって当然である．したがって，同じ標本情報（実験結果）であるにもかかわらず，異なる事前分布をもつ異なる推論主体が異なる事後分布をもつことになるのは当然のことである．それゆえ，同じ標本情報にもとづきながら，異なる推論主体が異なる推論結果を得ても当然である．この柔軟性は事前分布を活用するベイズ推論の著しい特徴である．

ベイズ推論のもう一つの著しい特徴は論理的整合性である．その意味は，推論主体が事前分布を選択，設定してしまえば，(i)事前分布と標本情報との総合である事後分布は（ベイズの定理という）論理的整合性をもつ数学的操作により導かれること，(ii)事後分布から（推定，検定などの）推論が論理的整合性をもって導かれるということである．

推論主体による事前分布の選択に関しては，主観確率の立場からの論理的整合性が与えられうるが，そこでの議論の多くは，理想化された理性的・合理的人間の主体的判断に関する，いわば規範的なものである．しかし，現実の推論主体は，そのように理想化された人間ではなく，試行錯誤しつつ学習してゆく生身の人間である．また，それは頭脳というすばらしい機能を備えており，その延長としてすぐれたコンピュータを有しているのである．このような現実の推論主体は，上記の規範的な議論をふまえた現実的な手法を用

いて，自己のもつ事前情報に対応する事前分布として適切なものを主体的に選択するのである．

さて，推論主体によって利用しうる経験，知恵，知識，理論，情報等は異なるので，同じ問題においても，推論主体が異なれば選択される事前分布が異なって当然であるが，そこにはさまざまな程度の差がみられる．両極端として，(i)ある推論主体が選択した事前分布が他のすべての人によって同意されるという（したがって，まったく客観的といえる）事前分布の場合と，(ii)それが他のすべてに同意されない（したがって，まったく主観的といえる）事前分布の場合とである．そして，現実に起こるのはこの両極端の間に位置するさまざまの場合である．

事前分布が推論主体の主体的（主観的）な要素を相対的に多分に含むものであっても，それと客観的な実験結果という標本情報との総合である事後分布においては，事前分布の主観性は標本情報の客観性により相対的に薄められる．この傾向は，標本情報の質と量の向上によっていっそう強められる．非常に極端な場合であるが，標本情報が完全情報（すなわち，未知母数の値が完全に決定できるほどに豊富な情報）であれば，すべての人は自己の事前分布の如何にかかわらずその値を真として認めねばならないであろう．

1.2 事後分布にもとづく推論

事後分布は，推論主体が実験前に措定した事前分布と実験結果（標本情報）との結合（総合）である．それは，前節でも述べたように実験前に利用可能な情報にもとづいた推論主体の（母数 θ の不確実性の程度に関する）主体的な判断の表現である事前分布と実験結果という新しい情報とを，ベイズの定理という数学的装置によって，論理的に適正に結合したものである．したがって，事後分布は，実験後において，事前分布と（実験で得られた）標本情報とにもとづいた，推論主体の（母数 θ の不確実性の程度に関する）主体的な総合的判断の表現である．

事後分布は上述のような性質のものであるから，推論主体が（自己の）事後分布にもとづいて θ に関する推論を行うのは当然である．そこで，事後分布という形で与えられる θ についての情報を，どのように推論と結びつけるかが次の問題となる．

θ が1次元であれば，$p(\theta|x)$ のグラフを画いて直観的に事後分布を提示することができる．また事後分布の特性値として，平均（期待値）$E(\tilde{\theta}|x)$，分散 $V(\tilde{\theta}|x)$ なども推論のために有用である．θ が2次元であれば，たとえば，同時密度 $p(\theta_1, \theta_2|x)$ の等密度線（等高線）の図示，あるいはコンピュータ・グラフィックスの手法による $p(\theta_1,\theta_2|x)$ の立体的画像により，事後分布を直観的に示すことができる．また，期待値ベクトル $E(\tilde{\theta}|x)$，分散行列 $V(\tilde{\theta}|x)$ も推論のために有用である．

$\theta=(\theta_1, \cdots, \theta_k)$ で $k \geqq 3$ のとき，上述のような直観的提示は困難であるが，事後密度 $p(\theta_1, \cdots, \theta_k|x)$ から $\tilde{\theta}_i$ の（周辺）事後密度 $p(\theta_i|x)\,(i=1, \cdots, k)$ あるいは $(\tilde{\theta}_i, \tilde{\theta}_j)$ の（周辺）事後密度 $p(\theta_i, \theta_j|x)\,(i, j=1, \cdots, k; i \neq j)$ を求め，これらを上述のように図形で直観的に示すことはできる．これらは事後分布 $p(\theta_1, \cdots, \theta_k|x)$ のもつ情報を推論に利用する際の大きな助けとなる．

事後密度のグラフを眺めて考えるだけで，θ についての推論がかなりできるであろう．これは推論の原点である．さて，統計学では従来，点推定，区間（領域）推定とか仮説検定といった推論の形式が用いられてきた．ベイズ推論の立場からみると，それらは次のように考えられる．

(i) 点推定

例として，$\Theta=\boldsymbol{R}$ の場合を考えてみよう．（θ が多次元の場合については，たとえば，鈴木(1)参照．）次のそれぞれが点推定の仕方として考えられよう．

(a) 事後分布の平均

$$E(\tilde{\theta}|x) = \int_R \theta p(\theta|x)\,d\theta = \mu(x) \tag{1.2}$$

を未知母数 θ の推定値として，事後分布の分散

$$V(\tilde{\theta}|x) = E[(\tilde{\theta}-\mu(x))^2|x] = \sigma^2(x) \tag{1.3}$$

あるいは標準偏差 $\sigma(x)$ で推定精度を表す．

(b) 事後分布のメディアン，すなわち

$$\int_{-\infty}^{\theta^*} p(\theta|x)\,d\theta = \frac{1}{2} \tag{1.4}$$

を満たす θ^* を未知母数 θ の推定値とする．このときの平均2乗誤差は

$$E[(\tilde{\theta}-\theta^*)^2|x] = E[(\tilde{\theta}-\mu(x)+\mu(x)-\theta^*)^2|x]$$
$$= V(\tilde{\theta}|x)+(\theta^*-\mu(x))^2 \qquad (1.5)$$

で与えられる．ただし，ここで $E(\tilde{\theta}-\mu(x)|x)=0$ を用いている．これは θ^* の推定精度を示すものである．

また，次のような形で推定精度を表現することもできる．たとえば，$\gamma=0.95$ とか 0.99 などに対して，

$$\frac{1}{2}(1-r) = \int_{-\infty}^{a_1} p(\theta|x)\,d\theta = \int_{a_2}^{\infty} p(\theta|x)\,d\theta \qquad (1.6)$$

を満たす a_1, a_2（あるいは a_2-a_1）を示すことである（したがって，a_1 は下方 $50(1-r)$% 点，a_2 は上方 $50(1-r)$% 点である）．この考えは次に述べる信用区間の考え方をとり入れたものである．事実，区間 $[a_1, a_2]$ は信用係数 r の信用区間である．精度のこのような表現は直接的で分かりやすいと思われる．

(c) 事後分布のモード，すなわち

$$\max_{\theta} p(\theta|x) = p(\theta^*|x) \qquad (1.7)$$

を満たす θ^* を未知母数 θ の推定値とする．この推定値の平均 2 乗誤差はメディアンの場合とまったく同じ式で与えられる．

また，(b)の場合と同じように，r を定めて，

$$\frac{r}{2} = \int_{a_1}^{\theta^*} p(\theta|x)\,d\theta = \int_{\theta^*}^{a_2} p(\theta|x)\,d\theta \qquad (1.8)$$

を満たす a_1, a_2（あるいは a_2-a_1）を示すことによって推定値 θ^* の精度を表現することも考えられよう．これも信用区間に近い考え方である．事実，区間 $[a_1, a_2]$ は信用係数 r の信用区間である．

(ii) 区間推定

事後分布にもとづいた

$$\Pr(\tilde{\theta}\in(a,b)|\tilde{x}=x) = \int_a^b p(\theta|x)\,d\theta = r \qquad (1.9)$$

という関係を用いて，区間 (a, b) を信用係数 r の信用区間という（非ベイズ推論の信頼係数，信頼区間とは定義が本質的に異なることに注意されたい）．

母数 θ が多次元の場合については，区間推定のほかに，領域推定という概

念も容易に考えられる．また，1 次元であっても，区間推定のほかに，集合を用いる集合推定という概念も容易に思いつくことである．これらについては，たとえば，鈴木(1)を参照されたい．

(iii) 仮説検定

母数空間 Θ のある部分集合 Θ_0 に対し，「θ が Θ_0 に属する」という言明を仮説 $H_0: \theta\in\Theta_0$ という．仮説 H_0 について，

$$P(H_0) \equiv \Pr(\tilde{\theta}\in\Theta_0) = \int_{\Theta_0} p(\theta)\,d\theta \qquad (1.10)$$

を仮説 H_0 の事前確率といい，

$$P(H_0|x) \equiv \Pr(\tilde{\theta}\in\Theta_0|\tilde{x}=x)$$
$$= \int_{\Theta_0} p(\theta|x)\,d\theta \qquad (1.11)$$

を，標本情報 $\tilde{x}=x$ を得たという条件の下での仮説 H_0 の事後確率という．

推論主体にとっては，$P(H_0|x)$ が 1 に近ければ近いほど，仮説 H_0 の成立がより確かであり，それが 0 に近ければ近いほど，仮説 H_0 の不成立がより確かとなるのである．

仮説を 1 つに限る必要はない．$\Theta\supset\Theta_i(i=,\cdots,m)$ として，m 個の仮説 $H_i: \theta\in\Theta_i(i=1,\cdots,m)$ を考え，それぞれの事後確率を (1.11) と同じように求め，それぞれの仮説の成否について推論を行うことができる（より詳しくは，たとえば，鈴木(1)参照）．

なお，事前分布の選択に際して，共役分布族という概念はきわめて有用である．また，θ に関する事前情報がきわめて稀薄な場合を表現する事前分布をどのように考えたらよいかは実際上重要であるとともに理論的にも興味深い問題である．これらについては，鈴木(1)，Box and Tiao (10), DeGroot(11)などを参照されたい．

また，推論主体は事前密度 $p(\theta)$ および条件付密度 $p(x|\theta)$ を主体的に設定し，事後密度 $p(\theta|x)$ を求め，これにもとづいて推論を行うのであるが，このとき，$p(\theta)$ とか $p(x|\theta)$ は一意的に定まるというよりはある程度の幅をもつものと考えるのが現実的である．そこで，$p(\theta)$ や $p(x|\theta)$ が少しずれたとき，それが $p(\theta|x)$ やそれにもとづく推論にどの程度の影響を与えるかを検

討する必要がある．これは推論モデルの頑健性（あるいは感度）の分析といわれるが，これについては，たとえば，鈴木(1)，Box and Tiao(10)などを参照されたい．

1.3 ベイズ推論の例

簡単な例を用いて，ベイズ推論の考え方を説明する．

A. 単純無作為標本抽出による比率に関する推論

N 個の個体からなる母集団において，性質 A（たとえば不良品）をもつ個体の個数を N_A とし，比率 $\theta = N_A/N$ を未知母数（たとえば不良率）とする．この母集団から n 個の個体をくじ引き式に無作為に抽出する．個体を1つずつ順次に無作為に抽出したとし，その順序に観測される確率変数を $\tilde{x}_1, \cdots, \tilde{x}_n$ とする．ここに

$$\tilde{x}_i = \begin{cases} 1 & (i \text{ 番目の個体が性質 A をもつ}) \\ 0 & (i \text{ 番目の個体が性質 A をもたない}) \end{cases}$$
$$(i = 1, 2, \cdots, n) \tag{1.12}$$

とおいている．

N が n に比して十分に大であれば，確率変数 $\tilde{x}_1, \cdots, \tilde{x}_n$ は近似的には独立であり，かつ同じ分布に従うことになる．このとき，

$$\Pr(\tilde{x}_i = 0|\theta) = 1-\theta, \quad Pr(\tilde{x}_i = 1|\theta) = \theta \tag{1.13}$$

であり，したがって，$\tilde{x}_i$ の確率関数は

$$p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i} I(x_i \mid \{0,1\}) \quad (i = 1, 2, \cdots, n) \tag{1.14}$$

と考えられる．ここに，$I(x_i|A)$ は集合 A の定義関数であり，$x_i \in \{0, 1\}$ のとき1，その他のときには0をとる関数である．

上の問題では，母数空間は $\Theta = \{i/N | i = 0, 1, \cdots, N\}$ であるが，N が大であれば，近似的にも実用上でも，Θ は区間 $[0, 1]$ と考えてよい．また，標本空間は $\mathcal{X} = \{(x_1, \cdots, x_n) | x_i \in \{0, 1\}, (i = 1, \cdots, n)\}$ である．

以上の考察から，N が n に比して十分に大であれば，$\tilde{\boldsymbol{x}} = (\tilde{x}_1, \cdots, \tilde{x}_n)$, $\boldsymbol{x} = (x_1, \cdots, x_n)$ とおくと，条件 $\tilde{\theta} = \theta$ の下での $\tilde{\boldsymbol{x}}$ の密度関数は

$$\begin{aligned} p(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} I(\boldsymbol{x}|\mathcal{X}) \end{aligned} \tag{1.15}$$

である．$\tilde{\theta}$ の事前密度を $p(\theta)$ とすると，標本の観測結果が，$\tilde{\boldsymbol{x}} = \boldsymbol{x} (\in \mathcal{X})$ のときの $\tilde{\theta}$ の事後密度は，(1.1) で述べたベイズの定理により，

$$\begin{aligned} p(\theta|\boldsymbol{x}) &\propto p(\theta)p(\boldsymbol{x}|\theta) \\ &\propto p(\theta)\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} \end{aligned} \tag{1.16}$$

である（ここで $\propto$ は比例の意味）．この問題では，$p(\boldsymbol{x}|\theta)$ の θ についての関数形からいわゆる共役分布はベータ分布である（共役分布については鈴木(1)参照）．そこで，いま事前分布としてベータ分布

$$\tilde{\theta} \sim \text{Beta}(\alpha_0, \nu_0 - \alpha_0) \quad (0 < \alpha_0 < \nu_0) \tag{1.17}$$

を選択したとしよう．このときの密度関数は，

$$p(\theta) = \frac{1}{B(\alpha_0, \nu_0 - \alpha_0)} \theta^{\alpha_0 - 1}(1-\theta)^{\nu_0 - \alpha_0 - 1} I(\theta | (0, 1)) \tag{1.18}$$

で与えられるので，$\sum_{i=1}^{n} x_i = t$ とおくと，(1.16) より

$$\begin{aligned} p(\theta|\boldsymbol{x}) &= p(\theta|t) \\ &\propto \theta^{\alpha_0 + t - 1}(1-\theta)^{\nu_0 - \alpha_0 + n - t - 1} I(\theta | (0, 1)) \end{aligned} \tag{1.19}$$

である．すなわち，

$$\sum_{i=1}^{n} \tilde{x}_i \equiv \tilde{t} = t \quad \in \{0, 1, \cdots, n\} \tag{1.20}$$

を観測したときの $\tilde{\theta}$ の事後分布は

$$\tilde{\theta} | (\tilde{t} = t) \sim \text{Beta}(\alpha_0 + t, \nu_0 - \alpha_0 + n - t) \tag{1.21}$$

すなわち，その事後密度は

$$p(\theta|t) = \frac{1}{B(\alpha_0 + t, \nu_0 - \alpha_0 + n - t)} \theta^{\alpha_0 + t - 1}(1-\theta)^{\nu_0 - \alpha_0 + n - t - 1} \tag{1.22}$$

であり，事後分布もベータ分布である．共役分布はこのような便利な性質を持つのである．

前節で述べたように，事後分布 $p(\theta|t)$ のグラフを見ることにより，θ に関する推論が行われうる．また，ベータ分布の性質から事後平均と事後分散は

$$E(\tilde{\theta}|t) = \frac{\alpha_0+t}{\nu_0+n} \tag{1.23}$$

$$V(\tilde{\theta}|t) = \frac{(\alpha_0+t)(\nu_0-\alpha_0+n-t)}{(\nu_0+n)^2(\nu_0+n+1)} \tag{1.24}$$

であり，また

$$\Pr(\tilde{\theta}\in(a,b)|\tilde{t}=t) = \frac{1}{B(\alpha_0+t,\ \nu_0-\alpha_0+n-t)}\int_a^b \theta^{\alpha_0+t-1}(1-\theta)^{\nu_0-\alpha_0+n-t-1}d\theta \quad (0\leqq a<b\leqq 1) \tag{1.25}$$

である（鈴木(1)参照）．これらを用いれば，前節で述べた形の推定，検定は容易に実行される．

いま，特に，$\alpha_0=1, \nu_0=2$ の場合，すなわち，事前密度 $p(\theta)$ が区間 $(0,1)$ 上の一様密度である場合を数値例として示すことにしよう．このとき，

$$E(\tilde{\theta}|t) = \frac{t+1}{n+2} \tag{1.26}$$

$$V(\tilde{\theta}|t) = \frac{(t+1)(n-t+1)}{(n+2)^2(n+3)} \tag{1.27}$$

$$\Pr(\tilde{\theta}\in(a,b)|\tilde{t}=t) = \frac{1}{B(t+1,\ n-t+1)}\int_a^b \theta^{t}(1-\theta)^{n-t}d\theta \quad (0\leqq a<b\leqq 1),\quad (t=0,1,\cdots,n) \tag{1.28}$$

である．

（i） $n=100, t=3$ のとき，

$$E(\tilde{\theta}|\tilde{t}=3) = \frac{4}{102} = 0.039,$$

$$V(\tilde{\theta}|\tilde{t}=3) = \frac{(4)(98)}{(102)^2(103)} = 3.658\times 10^{-4},$$

$$\sqrt{V(\tilde{\theta}|\tilde{t}=3)} = 0.019$$

であるから，点推定は，推定値を 0.039 とし，標準偏差 0.019 で推定精度を示す．

より理解しやすいのは区間推定である．そのために，(1.28) のタイプの積分の評価が必要である．いくつかの信用係数と信用区間の対をあげる．D はデータ（$n=100$, $\tilde{t}=3$）を表す．

$\Pr(\tilde{\theta}\in(0.05, 0.12)|D) \fallingdotseq 0.997$,　$\Pr(\tilde{\theta}\in(0.01, 0.11)|D) \fallingdotseq 0.978$,
$\Pr(\tilde{\theta}\in(0.011, 0.09)|D) \fallingdotseq 0.958$,　$\Pr(\tilde{\theta}\in(0.013, 0.08)|D) \fallingdotseq 0.920$,
$\Pr(\tilde{\theta}\in(0.013, 0.07)|D) \fallingdotseq 0.866$,

などである．仮説検定については，たとえば，

$$H_0: \theta\in(0, 0.1) = \Theta_0$$

の事後確率は，(1.28) により，

$$\Pr(\tilde{\theta}\in(0, 0.1)|D) = 0.993$$

である（信用区間と仮説検定の確率の評価は同じタイプである）．

（ii） $n=50, t=3$ のとき，

$$E(\tilde{\theta}|\tilde{t}=3) = \frac{4}{52} = 0.0769,\quad \sqrt{V(\tilde{\theta}|\tilde{t}=3)} = 0.0366$$

であるから，点推定値は 0.0769，推定精度は標準偏差 0.0366 で表す．区間推定（とか仮説検定）の例として，

$\Pr(\tilde{\theta}\in(0.01, 0.4)|D) \fallingdotseq 0.998$,　$\Pr(\tilde{\theta}\in(0.01, 0.19)|D) \fallingdotseq 0.990$,
$\Pr(\tilde{\theta}\in(0.02, 0.2)|D) \fallingdotseq 0.976$,　$\Pr(\tilde{\theta}\in(0.02, 0.15)|D) \fallingdotseq 0.939$,

などがあげられる．D はデータ（$n=50$, $\tilde{t}=3$）を表す．

（iii） $n=50, t=25$ のとき，

$$E(\tilde{\theta}|\tilde{t}=25) = \frac{26}{52} = 0.5$$

$$\sqrt{V(\tilde{\theta}|\tilde{t}=25)} = 0.069$$

また，D でデータ（$n=50, t=25$）を表すと，

$\Pr(\tilde{\theta}\in(0.3, 0.7)|D) \fallingdotseq 0.997$,　$\Pr(\tilde{\theta}\in(0.33, 0.67)|D) \fallingdotseq 0.988$,
$\Pr(\tilde{\theta}\in(0.36, 0.64)|D) \fallingdotseq 0.960$,

などである．

〔注意 1.1〕 N が n に比して十分には大でないときは，もちろん，$\tilde{x}_1,\cdots,\tilde{x}_n$ は独立とはみなされないが，$\tilde{t}=\sum_{i=1}^{n}\tilde{x}_i$ は $\tilde{\theta}=\theta$ のとき，超幾何分布に従う．すなわち

$$\Pr(\tilde{t}=t|\theta)=\frac{\binom{N\theta}{t}\binom{N(1-\theta)}{n-t}}{\binom{N}{n}} \tag{1.29}$$

である．また，このとき，

$$\Theta=\left\{\frac{i}{N}\Big|i=0,1,\cdots,N\right\} \tag{1.30}$$

である．したがって，$\tilde{\theta}$ の確率関数を $p(\theta)$ とすると，観測結果 $\tilde{t}=t$ を得たとき

$$p(\theta|t)\propto p(\theta)\binom{N\theta}{t}\binom{N(1-\theta)}{n-t} \tag{1.31}$$

である．この場合の数値例については，鈴木(1)を参照されたい．

B. 正規分布の場合

本項では，観測される確率変数が1次元正規分布 $N(\theta,\sigma^2)$（θ：未知母数，σ^2：既知）に従うという簡単な場合を例として，ベイズ推論とは如何なるものかを説明する．（分散 σ^2 も未知母数である場合とか，観測される確率変数が多次元正規分布 $N(\theta,\Sigma)$ に従い，さらに，分散行列が既知の場合および未知の場合などにおける未知母数の推論については，たとえば，鈴木(1)を参照されたい．）

母数空間を $\Theta=\boldsymbol{R}=(-\infty,\infty)$ とする．観測される確率変数 $\tilde{x}_1,\cdots,\tilde{x}_n$ は独立で

$$\tilde{x}_i|(\tilde{\theta}=\theta)\sim N(\theta,\sigma^2)\qquad(\sigma^2:\text{既知})\qquad(i=1,2,\cdots,n) \tag{1.32}$$

であるとする．すなわち，$\tilde{\theta}=\theta$ が与えられたときの $\tilde{x}_i$ の密度関数は

$$p(x_i|\theta)=(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\} \tag{1.33}$$

であり，$\tilde{x}=(\tilde{x}_1,\cdots,\tilde{x}_n)$ の密度関数は，$\tilde{\theta}=\theta$ の下では，

$$p(\boldsymbol{x}|\theta)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} \tag{1.34}$$

である．ところで，$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ とおくと，

$$\sum_{i=1}^{n}(x_i-\theta)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\theta-\bar{x})^2 \tag{1.35}$$

が成り立つので，$\tilde{\boldsymbol{x}}=\boldsymbol{x}$ を観測したとき，

$$p(\theta|\boldsymbol{x})\propto p(\theta)\exp\left\{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\right\} \tag{1.36}$$

が得られる．実は，$p(\boldsymbol{x}|\theta)$ の θ に関する関数形から分かるように，共役分布は正規分布である．いま，$\tilde{\theta}$ の事前分布として

$$\tilde{\theta}\sim N(\mu_0,\sigma_0^2) \tag{1.37}$$

を選択したとして話を進める．このとき，

$$p(\theta|\boldsymbol{x})\propto\exp\left\{-\frac{1}{2}(\sigma_0^{-2}(\theta-\mu_0)^2+n\sigma^{-2}(\theta-\bar{x})^2)\right\} \tag{1.38}$$

である．容易に示されるように

$$\begin{aligned}&\sigma_0^{-2}(\theta-\mu_0)^2+n\sigma^{-2}(\theta-\bar{x})^2\\&\quad=\sigma_1^{-2}(\theta-\mu_1)^2+n\sigma_0^{-2}\sigma^{-2}\sigma_1^2(\bar{x}-\mu_0)^2\end{aligned} \tag{1.39}$$

が成り立つ．ただし，

$$\mu_1=\frac{\sigma_0^{-2}\mu_0+n\sigma^{-2}\bar{x}}{\sigma_0^{-2}+n\sigma^{-2}},\quad\sigma_1^{-2}=\sigma_0^{-2}+n\sigma^{-2} \tag{1.40}$$

とおいている．したがって，

$$p(\theta|\boldsymbol{x})=p(\theta|\bar{x})\propto\exp\left\{-\frac{1}{2\sigma_1^2}(\theta-\mu_1)^2\right\} \tag{1.41}$$

である．すなわち，$\tilde{\theta}$ の事後分布は

$$\tilde{\theta}|(\tilde{\bar{x}}=\bar{x})\sim N(\mu_1,\sigma_1^2) \tag{1.42}$$

であり，平均と分散は

$$E(\tilde{\theta}|\bar{x})=\mu_1,\quad V(\tilde{\theta}|\bar{x})=\sigma_1^2 \tag{1.43}$$

で与えられる．また，

$$\begin{aligned}&\Pr(\tilde{\theta}\in(a,b)|\tilde{\boldsymbol{x}}=\boldsymbol{x})=\Pr(\tilde{\theta}\in(a,b)|\tilde{\bar{x}}=\bar{x})\\&\quad=\Pr\left(\frac{\tilde{\theta}-\mu_1}{\sigma_1}\in\left(\frac{a-\mu_1}{\sigma_1},\frac{b-\mu_1}{\sigma_1}\right)\Big|\tilde{\bar{x}}=\bar{x}\right)\\&\quad=\Pr\left(\tilde{z}\in\left(\frac{a-\mu_1}{\sigma_1},\frac{b-\mu_1}{\sigma_1}\right)\right)\end{aligned} \tag{1.44}$$

が成り立つ．ただし，$\tilde{z}\sim N(0,1)$ としている．ここで $N(0,1)$ は標準正規分布を表す．したがって，正規分布表が利用できることになる．

特に，$r\in(0.5,1)$ として，$z(r)$ を標準正規分布の上方 $50(1-r)\%$ 点，すなわち

$$\frac{1}{\sqrt{2\pi}}\int_{z(r)}^{\infty} e^{-\frac{x^2}{2}}dx = \frac{1-r}{2} \tag{1.45}$$

を満たすものと定義すると，正規分布表からみられるように，たとえば，

$$z(0.90) \fallingdotseq 1.645, \qquad z(0.95) \fallingdotseq 1.96,$$
$$z(0.99) \fallingdotseq 2.576, \qquad z(0.997) \fallingdotseq 2.97,$$

であるので，次の関係式

$$\Pr(\tilde{\theta} \in [\mu_1 - 1.645\sigma_1, \mu_1 + 1.645\sigma_1]|D) = 0.90,$$
$$\Pr(\tilde{\theta} \in [\mu_1 - 1.96\sigma_1, \mu_1 + 1.96\sigma_1]|D) = 0.95,$$
$$\Pr(\tilde{\theta} \in [\mu_1 - 2.576\sigma_1, \mu_1 + 2.576\sigma_1]|D) = 0.99,$$
$$\Pr(\tilde{\theta} \in [\mu_1 - 2.97\sigma_1, \mu_1 + 2.97\sigma_1]|D) = 0.997, \tag{1.46}$$

が得られる．

これらは，前節で述べたベイズ推論を行うために有用である．

〔**注意 1.2**〕 $n\sigma^{-2}$ が σ_0^{-2}に比して十分に大であるとき，

$$\mu_1 = \frac{\sigma_0^{-2}\mu_0 + n\sigma^{-2}\bar{x}}{\sigma_0^{-2} + n\sigma^{-2}} \fallingdotseq \bar{x}, \quad \sigma_1{}^2 = \frac{1}{\sigma_0^{-2} + n\sigma^{-2}} \fallingdotseq \frac{\sigma^2}{n} \tag{1.47}$$

であるので，

$$\tilde{\theta}|(\tilde{\bar{x}} = \bar{x}) \dot{\sim} N\left(\bar{x}, \frac{\sigma^2}{n}\right) \tag{1.48}$$

が成り立つ．すなわち，事後分布は近似的に $N(\bar{x}, \sigma^2/n)$ に等しい．

〔**注意 1.3**〕 上の〔注意 1.2〕の場合を例として，非ベイズ推論での区間推定のための信頼区間とベイズ推論での区間推定のための信用区間との間の違いを以下で明確にしておきたい．

(1) 信頼区間は次のように構成される．まず，信頼係数 r を定めれば，

$$\Pr\left(\theta \in [\tilde{\bar{x}} - z(r)\frac{\sigma}{\sqrt{n}}, \tilde{\bar{x}} + z(r)\frac{\sigma}{\sqrt{n}}]|\theta\right) = r \qquad (\theta \in \Theta) \tag{1.49}$$

が成り立つ．ただし，$z(r)$ は標準正規分布 $N(0, 1)$ の上方 $50(1-r)$% 点としている．そこで，確率区間 (random interval)

$$\left[\tilde{\bar{x}} - z(r)\frac{\sigma}{\sqrt{n}}, \tilde{\bar{x}} + z(r)\frac{\sigma}{\sqrt{n}}\right] \tag{1.50}$$

に注目する．ついで，観測結果 $\tilde{\bar{x}} = \bar{x}$ を得たとき，これを単に代入することによって得られる区間

$$\left[\bar{x} - z(r)\frac{\sigma}{\sqrt{n}}, \bar{x} + z(r)\frac{\sigma}{\sqrt{n}}\right] \tag{1.51}$$

を信頼係数 r の信頼区間と呼ぶのである（信頼係数と信頼区間の意味を理解するためには，もちろん，(1.49) の関係およびそこでの確率の意味を考えねばならない）．言うまでもなく，ここでは，θ の事前分布は用いられていない．また，(1.49) の左辺では，未知母数 θ が条件になっていることは特に注意されたい．

(2) 信用区間は次のように構成される．まず，$n\sigma^{-2} \gg \sigma_0^{-2}$ の場合，事後分布を (1.48) の代りに，

$$\tilde{\theta}|(\tilde{\bar{x}} = \bar{x}) \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) \tag{1.52}$$

と考えてよい．このとき，信用係数 r を定めれば，

$$\Pr\left(\tilde{\theta} \in [\bar{x} - z(r)\frac{\sigma}{\sqrt{n}}, \bar{x} + z(r)\frac{\sigma}{\sqrt{n}}|\tilde{\bar{x}} = x\right) = r \tag{1.53}$$

が成り立つ．そこで，区間

$$\left[\bar{x} - z(r)\frac{\sigma}{\sqrt{n}}, \bar{x} + z(r)\frac{\sigma}{\sqrt{n}}\right] \tag{1.54}$$

を信用係数 r の信用区間と呼ぶのである．信用区間の意味は (1.53) から明らかである．それは，「未知母数 θ が信用区間 (1.54) に属する事後確率は r である」ということである．

〔**注意 1.4**〕〔注意 **1.3**〕の例では，信頼区間 (1.51) と信用区間 (1.54) とはまったく同じ区間である．しかし，両者は異なる意味をもつ．信頼区間の意味づけは関係式 (1.49) にもとづき，信用区間の意味づけは関係式 (1.53) にもとづくということを繰り返し注意しておきたい．〔注意 1.3〕では，非ベイズ推論とベイズ推論の差を区間推定を例として示したのであるが，両者の差を点推定および仮説検定について明確に理解しておくことも重要である．特に，仮説検定の考え方は，非ベイズ推論とベイズ推論では大いに異なることに注意されたい．

以下では，本項で述べた正規分布の場合の数値例をいくつか考えてみよう．

例 1：事前分布(1.37)の μ_0, σ_0^2 および(1.32)の σ^2 を

$$\mu_0 = 5.00, \quad \sigma_0^2 = 25, \quad \sigma^2 = 1$$

とし，観測結果が

$$D : n = 50, \quad \bar{x} = 4.52$$

であったとする（上の観測結果を記号 D で表すと便利である）．このとき，(1.40) により，

$$\mu_1 = \frac{(25)^{-1}(5)+50(4.52)}{(25)^{-1}+50} = \frac{226.2}{50.04} = 4.520$$

$$\sigma_1^{-2} = 50.04, \quad \sigma_1^2 = 0.019984, \quad \sigma_1 = 0.1414$$

である．したがって，(1.42) より事後分布は

$$\tilde{\theta}|D \sim N(4.520, 0.019984) \tag{1.55}$$

である．事前分布は，(1.37) より

$$\tilde{\theta} \sim N(5.000, 25) \tag{1.56}$$

であったのだから，これに比較すれば，事後分布 (1.55) は θ についてかなり精度の高い情報を与えてくれる．

次に，事後分布にもとづく推定および検定について考える．

(i) 点 推 定

$\mu_1=4.520$ を推定値とする．精度は $\sigma_1=0.1414$ で表される．

(ii) 区間推定

信用係数 $r=0.90$ の信用区間は，(1.46) を用いて

$$(\mu_1-1.645\sigma_1, \mu_1+1.645\sigma_1) \fallingdotseq (4.287, 4.753)$$

信用係数 $r=0.95$ の信用区間は

$$(\mu_1-1.96\sigma_1, \mu_1+1.96\sigma_1) \fallingdotseq (4.243, 4.797)$$

信用係数 $r=0.99$ の信用区間は

$$(\mu_1-2.576\sigma_1, \mu_1+2.576\sigma_1) \fallingdotseq (4.156, 4.884)$$

信用係数 $r=0.997$ の信用区間は

$$(\mu_1-2.97\sigma_1, \mu_1+2.97\sigma_1) \fallingdotseq (4.100, 4.940)$$

である．

(iii) 仮説検定

$\Theta_0=[4.35, 4.75]$ とし，仮説 $H_0: \theta \in \Theta_0$ を考えてみよう．まず，H_0 の事前確率は，(1.56) より，

$$\begin{aligned}
p(H_0) &= \Pr(\tilde{\theta} \in [4.35, 4.75]) \\
&= \Pr\left(\frac{4.35-5}{5} \leqq \frac{\tilde{\theta}-5}{5} \leqq \frac{4.57}{5}\right) \\
&= \Pr(-0.13 \leqq \tilde{z} \leqq -0.05) \qquad (\tilde{z} \sim N(0,1)) \\
&= \Pr(\tilde{z} \in [0.05, 0.13]) \\
&= 0.48006-0.44828 \qquad \text{（正規分布表）} \\
&= 0.03178
\end{aligned}$$

である．次に，観測結果 D を得たときの仮説の事後確率は，(1.55) より，

$$\begin{aligned}
p(H_0|D) &= \Pr(\tilde{\theta} \in [4.35, 4.75]|D) \\
&= \Pr\left(\frac{4.35-4.52}{0.1414} \leqq \frac{\tilde{\theta}-4.52}{0.1414} \leqq \frac{4.75-4.52}{0.1414}\Big|D\right) \\
&= \Pr(-1.202 \leqq \tilde{z} \leqq 1.627) \qquad (\tilde{z} \sim N(0,1)) \\
&\fallingdotseq 0.8329
\end{aligned}$$

である．

仮説 H_0 の事前確率は

$$p(H_0) \fallingdotseq 0.0318$$

にすぎなかったのに，標本情報 D が得られたとき，その事後確率は

$$p(H_0|D) \fallingdotseq 0.833$$

である．両者の比をとると，

$$\frac{p(H_0|D)}{p(H_0)} \fallingdotseq 26.2$$

である．したがって，標本情報 D は仮説 H_0 をかなり支持するものであるといえる．しかし，$p(H_0|D)=0.833$ は，推論主体が仮説 H_0 の成立に確信を持つにいたるほど十分に 1 に近いとはいえないであろう．したがって，さらに観測を行い，新しい標本情報を得て，新しい事後確率を算出し，仮説 H_0 の成立の可能性について考えてゆくことになるであろう．

次に，別の例として，$\Theta_1=[5, +\infty]$ とし，仮説 $H_1: \theta \in \Theta_1$ を考えてみよう．明らかに，H_1 の事前確率は

$$p(H_1) = 0.5$$

である．次に，標本情報 D を得たときの仮説 H_1 の事後確率は

$$p(H_1|D) = \Pr(\tilde{\theta} \in \Theta_1|D)$$

$$= \Pr(5 \leqq \tilde{\theta}|D)$$
$$= \Pr\left(\frac{5-4.52}{0.1414} \leqq \tilde{z}\right) \qquad (\tilde{z} \sim N(0, 1))$$
$$= \Pr(0.395 \leqq \tilde{z})$$
$$\fallingdotseq 0.00034$$

である．両者の比は

$$\frac{p(H_1|D)}{p(H_1)} = 0.00068$$

である．この場合は，標本情報 D は仮説 H_1 をほとんど支持しないことが分かる．さて，$p(H_1|D)=0.00034$ をどのように考えるかは推論主体によるのであるが，推論主体は仮説 H_1 の成立を否定的に考えるであろう．

最後に，例1では $n\sigma^{-2}=50$, $\sigma_0^{-2}=0.04$ であるので，実際上は $n\sigma^{-2} \gg \sigma_0^{-2}$ と考えてよい．すなわち，〔注意1.2〕の場合に該当することを注意しておこう．

例2：事前分布の σ_0^2 だけを変えて，

$$\mu_0 = 5.00, \quad \sigma_0^2 = 0.25, \quad \sigma^2 = 1$$

とし，観測結果は前例と同じく

$$D: n = 50, \quad \bar{x} = 4.52$$

であったとしよう．このとき，

$$\mu_1 = \frac{4(5)+50(4.52)}{4+50} = 4.556$$

$$\sigma_1^{-2} = 54, \quad \sigma_1^2 = \frac{1}{54}, \quad \sigma_1 = 0.1341$$

であるので，事後分布は

$$\tilde{\theta}|D \sim N(4.556, (0.1341)^2) \tag{1.57}$$

である．

(i) 点 推 定

$\mu_1=4.556$ を推定値とし，$\sigma_1=0.1341$ で推定精度を表す．

(ii) 区間推定

信用係数 $\gamma=0.90$ の信用区間は

$$(\mu_1-1.645\sigma_1, \mu_1+1.645\sigma_1) = (4.335, 4.778)$$

信用係数 $\gamma=0.95$ の信用区間は

$$(\mu_1-1.96\sigma_1, \mu_1+1.96\sigma_1) = (4.293, 4.819)$$

信用係数 $\gamma=0.99$ の信用区間は

$$(\mu_1-2.576\sigma_1, \mu_1+2.576\sigma_1) = (4.211, 4.901)$$

信用係数 $\gamma=0.997$ の信用区間は

$$(\mu_1-2.97\sigma_1, \mu_1+2.97\sigma_1) = (4.158, 4.954)$$

である．

(iii) 仮説検定

前例と同じく，まず，$\Theta_0=[4.35, 4.75]$ とし，仮説 $H_0: \theta \in \Theta_0$ を考えてみよう．事前分布を

$$\tilde{\theta} \sim N(5, (0.5)^2) \tag{1.58}$$

としているので，仮説 H_0 の事前確率は

$$p(H_0) = \Pr(\tilde{\theta} \in [4.35, 4.75])$$
$$= \Pr\left(\frac{\tilde{\theta}-5}{0.5} \in \left[\frac{4.35-5}{0.5}, \frac{4.75-5}{0.5}\right]\right)$$
$$= \Pr(\tilde{z} \in [-1.3, -0.5])$$
$$= \Pr(\tilde{z} \in [0.5, 1.3])$$
$$= 0.2117$$

である．次に，標本情報 D を得たときの仮説 H_0 の事後確率は

$$p(H_0|D) = \Pr(\tilde{\theta} \in [4.35, 4.75]|D)$$
$$= \Pr\left(\frac{\tilde{\theta}-4.556}{0.1341} \in \left[\frac{4.35-4.556}{0.1341}, \frac{4.75-4.556}{0.1341}\right]\right)$$
$$= \Pr(\tilde{z} \in [-1.536, 1.447])$$
$$\fallingdotseq 0.864$$

である．

これらの結果についての考察は例1の場合と同様になされうる．

2. ベイズ意思決定理論の基本的な考え方

2.1 統計的意思決定問題

統計的意思決定問題を規定するためには，意思決定者が選択しうる決定の集合である決定の空間 D，可能な自然の状態の全体からなる集合である状態の空間 Θ，直積空間 $D\times\Theta$ の上で定義される効用関数 $u(d,\theta)$ がまず必要である．次に，いま，自然の状態について情報を提供しうる情報源（あるいは，調査，実験など）を1つだけ考えることにすれば，この情報源から得られうる情報（標本）の全体からなる集合を $\mathcal{X}$ で表し，これを統計的推論の場合と同じく標本空間という．

多くの典型的な意思決定問題では，ベイズ推論の場合と同じように，状態 θ が真であるときの実験結果（標本）の分布，したがって，$p(x|\theta)$，$x\in\mathcal{X}$ が想定できる（あるいは，そのように考えるのが自然である）という場合が多い．この場合，意思決定者が，実験とか観測によって標本情報を得る前に，すでに自然の状態の不確実性について一定の評価を持っているものとし，それが Θ 上の1つの分布（事前分布）として表現されるものとし，その確率関数（あるいは密度関数）を $p(\theta)$ で表すことにする．事前分布と標本分布から事後分布はベイズの定理を用いて求められる．

Θ および $\mathcal{X}$ が有限集合で，

$$\Theta=\{\theta_1,\cdots,\theta_K\},\ \mathcal{X}=\{x_1,\cdots,x_N\} \tag{2.1}$$

であれば，条件付確率の公式によれば，

$$\begin{aligned}p(\theta_i|x_j)&=\frac{p(x_j,\theta_i)}{p(x_j)}\\&=\frac{p(x_j|\theta_i)p(\theta_i)}{\sum_{k=1}^{K}p(x_j,\theta_k)}\\&=\frac{p(x_j|\theta_i)p(\theta_i)}{\sum_{k=1}^{K}p(x_j|\theta_k)p(\theta_k)}\qquad(i=1,\cdots,K:j=1,\cdots,N)\end{aligned} \tag{2.2}$$

が成り立つ．この公式はベイズの定理と呼ばれるが，Θ や $\mathcal{X}$ が可算無限集合の場合も本質的には同じである．

$\Theta=\{\theta_1,\cdots,\theta_K\}$ で $\mathcal{X}$ が，たとえば，実数の全体 $\boldsymbol{R}$ に等しく，したがって，$p(x|\theta_i)$ が θ_i が真のときの $\mathcal{X}$ 上の密度関数であれば，ベイズの定理は，

$$p(\theta_i|x)=\frac{p(x|\theta_i)p(\theta_i)}{\sum_{k=1}^{K}p(x|\theta_k)p(\theta_k)}\qquad(i=1,\cdots,K:x\in\boldsymbol{R}) \tag{2.3}$$

となる．逆に，Θ が $\boldsymbol{R}$ に等しく，$\mathcal{X}=\{x_1,\cdots,x_N\}$ であれば，

$$p(\theta|x_j)=\frac{p(x_j|\theta)p(\theta)}{\int_R p(x_j|\theta)p(\theta)\,d\theta}\qquad(i=1,\cdots,N:\theta\in\Theta) \tag{2.4}$$

となる．もちろん，ここでは，$p(\theta),p(\theta|x_j)$ は Θ 上の確率関数である．また，たとえば，$\theta,\mathcal{X}$ がともに $\boldsymbol{R}$ に等しい場合については，

$$p(\theta|x)=\frac{p(x|\theta)p(\theta)}{\int_R p(x|\theta)p(\theta)\,d\theta}\qquad((x,\theta)\in\boldsymbol{R}^2) \tag{2.5}$$

である．この場合，$p(\theta),p(x|\theta),p(\theta|x)$ はそれぞれ，$\Theta,\mathcal{X},\Theta$ 上の密度関数である（x や θ が多次元の場合への拡張も容易である）．

周辺分布と条件付分布を与えるものとして，$p(\theta),p(x|\theta)$ を考えると，

$$p(\theta)p(x|\theta)\equiv p(x,\theta) \tag{2.6}$$

は，$\mathcal{X}\times\Theta$ 上の分布を定める．この分布に従う確率変数を $(\tilde{x},\tilde{\theta})$ で表すことにする．

以上をまとめると，決定問題を規定する要素としては，

(1)　決定の空間 D

(2)　自然の状態の空間 Θ

(3)　効用関数 $u(d,\theta)$

のほかに，統計的意思決定問題の統計的な特徴として，統計的情報および不確実性を規定するものとして，

(4)　標本空間 $\mathcal{X}$

(5)　条件付密度関数 $p(x|\theta)$ および事前分布の密度関数 $p(\theta)$

が必要である．また，(5)の代りに，$\tilde{x}$ と $\tilde{\theta}$ の同時密度関数 $p(x,\theta)$ が与えられても，これから周辺密度関数 $p(\theta)=\int p(x,\theta)\,dx$，および条件付密度関数

$p(x|\theta)=p(x,\theta)/p(\theta)$ が得られるので，

(5)′　$\tilde{x}$ と $\tilde{\theta}$ の同時密度関数 $p(x,\theta)$

と(5)は同等である．さらに，(5)′から，$p(x)=\int p(x,\theta)d\theta$ によって $\tilde{x}$ の周辺密度関数が得られ，また $p(\theta|x)=p(x,\theta)/p(x)$ によって条件付密度関数が得られる．そして逆に，$p(x)$ と $p(\theta|x)$ の積として $p(x,\theta)$ が定まるのであるから，(5)′と，

(5)″　$\tilde{x}$ の密度関数 $p(x)$ と条件付密度関数 $p(\theta|x)$

とは同等である．したがって，(5)，(5)′，(5)″はすべて同等である．

さて，ある決定問題に対して上述の(1)～(5)が規定されたとして，この問題における最適決定を求める手続きを考えてみよう．まず，$D,\Theta,\mathcal{X}$ がすべて有限集合で，

$$D=\{d_1,\cdots,d_M\},\quad \Theta=\{\theta_1,\cdots,\theta_K\},\quad \mathcal{X}=\{x_1,\cdots,x_N\} \tag{2.7}$$

であるとしよう．実験とか調査の結果 $\tilde{x}=x_j$ を得たとすれば，ベイズの定理により事後確率関数 $p(\theta_k|x_j)\,(k=1,\cdots,K)$ を求め，これを用いて各決定 $d_i\ (i=1,\cdots,M)$ の条件 $\tilde{x}=x_j$ のもとでの期待効用

$$\begin{aligned} u(d_i|x_j) &= \sum_{k=1}^{K} u(d_i,\theta_k)\,p(\theta_k|x_j) \qquad (2.8)\\ &= E[u(d_i,\tilde{\theta})|\tilde{x}=x_j] \end{aligned}$$

を求め，これらを最大ならしめる決定，すなわち，

$$\max_{i=1}^{M} u(d_i|x_j) = u(d^*(x_j)|x_j) \tag{2.9}$$

によって定まる決定 $d^*(x_j)$ を最適な決定として選択すればよい（上の最適決定は x_j に依存するので $d^*(x_j)$ で表した）．

上では，$\tilde{x}=x_j$ が得られたという条件のもとでとるべき決定を求めたが，$\mathcal{X}$ の中の各実験結果について同様に考えれば，$\mathcal{X}$ 上の関数 $d^*(x_j)\,(j=1,\cdots,N)$ が定められる．これは $\mathcal{X}$ から D への1つの写像（あるいは関数）である．一般に $\mathcal{X}$ から D への（可測な）写像を決定関数という．決定関数は各実験結果に決定を対応づけるものであり，優劣の差こそあれ，1つの決定関数は1つの意思決定方式を与える．いま，$\mathcal{X}$ から D への決定関数の全体を $\mathcal{D}$ で表すと，$\mathcal{X}$ および D が有限集合であれば $\mathcal{D}$ も有限集合であるが，上に

求めた $d^*(\cdot)$ は $\mathcal{D}$ の中で，期待効用を最大ならしめるという基準によれば，最良（最適）の決定関数である．これは次のように証明できる．

$\mathcal{D}$ の任意の決定関数 $d(\cdot)$ にしたがって意思決定を行うとしたときの期待効用を $u(d(\cdot))$ で表すと，明らかに，

$$\begin{aligned} u(d(\cdot)) &= E[u(d(\tilde{x}),\tilde{\theta})]\\ &= \sum_{j=1}^{N}\sum_{k=1}^{K} u(d(x_j),\theta_k)\,p(x_j,\theta_k)\\ &= \sum_{j=1}^{N}\Bigl[\sum_{k=1}^{K} u(d(x_j),\theta_k)\,p(\theta_k|x_j)\Bigr]p(x_j) \qquad (2.10) \end{aligned}$$

と書ける．最後の式は，

$$\sum_{j=1}^{N} E[u(d(x_j),\tilde{\theta})|\tilde{x}=x_j]p(x_j) = \sum_{j=1}^{N} u(d(x_j)|x_j)\,p(x_j) \tag{2.11}$$

あるいは，2段階で期待値をとる形として，

$$EE[u(d(\tilde{x}),\tilde{\theta})|\tilde{x}] \tag{2.12}$$

と表すこともできる．さて，$d^*(\cdot)$ の定義から，

$$u(d^*(x_j)|x_j) \geqq u(d(x_j)|x_j) \qquad (x_j\in\mathcal{X},\quad d(\cdot)\in\mathcal{D}) \tag{2.13}$$

であるから，

$$\begin{aligned} u(d^*(\cdot)) &= \sum_{j=1}^{N} u(d^*(x_j)|x_j)\,p(x_j)\\ &\geqq \sum_{j=1}^{N} u(d(x_j)|x_j)\,p(x_j) = u(d(\cdot)) \qquad (d(\cdot)\in\mathcal{D}) \end{aligned} \tag{2.14}$$

が成り立つことがわかる．すなわち，$d^*(\cdot)$ の期待効用が $\mathcal{D}$ の中で最大である．

しかし，現実の問題では決定関数まで定める必要は必ずしもないことを注意しておきたい．現実に得られた実験結果（標本）を x_j とすれば，$u(d_i|x_j)$ を求め，これを最大ならしめる決定を x_j に対応する最適決定 $d^*(x_j)$ として選択すれば十分なのであり，現実に得られなかった実現結果を考慮する必要はないのである．

次に，$\Theta\subset R^m,\ \mathcal{X}\subset R^n$ の場合で，密度関数 $p(x|\theta),\ p(\theta)$ が与えられる場合を考えてみよう．この場合も，有限の場合と本質的には同じであり，ただ，和の代りに積分を用いるだけの差があるにすぎないが，E という記号を用い

れば表現もまったく同じになる．前と同じく，$\tilde{x}=x$ のとき，$p(x|\theta)$ と $p(\theta)$ からベイズの定理で $p(\theta|x)$ を求め，x を観測したときの決定 d の（条件付）期待効用

$$E[u(d,\tilde{\theta})|\tilde{x}=x]=\int_{\Theta}u(d,\theta)p(\theta|x)d\theta \tag{2.15}$$

を求める．これが $u(d|x)$ である．そこで，

$$\max_{d\in D}u(d|x)=u(d^*(x)|x) \tag{2.16}$$

を満足する $d^*(x)$ を求めれば，これが x に対する最適決定である．

これまでの記述から，決定問題の解き方に2通りあることが分かる．これらは，展開型分析および正規型分析といわれている．（次節では情報源の選択を含んだより一般な決定問題について，展開型分析が，決定樹を用いて詳しく説明されるであろう．）

(1) 展開型分析

観測結果 $\tilde{x}=x$ を得たという条件の下では，決定 d の条件付期待効用，

$$u(d|x)=E[u(d,\tilde{\theta})|\tilde{x}=x] \tag{2.17}$$

を用いて

$$\max_{d\in D}u(d|x)=u(d^*(x)|x) \tag{2.18}$$

によって，条件 $\tilde{x}=x$ に対する最適決定 $d^*(x)$ が求められる．$\mathcal{X}$ に属する各 x について，上と同じように考えれば，最適決定関数 $d^*(\cdot)$ が求められる．これが展開型分析といわれる解法である．

(2) 正規型分析

観測を行う以前の時点で一挙に，最適決定関数 $d^*(\cdot)$ を $\mathcal{D}$ の中から選択しようというのが，正規型分析といわれるものである．すなわち，決定関数 $d(\cdot)\in\mathcal{D}$ に対して，期待効用

$$u(d(\cdot))=E[u(d(\tilde{x}),\tilde{\theta})] \tag{2.19}$$

を用いて，

$$\max_{d(\cdot)\in\mathcal{D}}u(d(\cdot))=u(d^*(\cdot)) \tag{2.20}$$

によって，一挙に $d^*(\cdot)$ を求めんとするのが正規型分析である．

2.2 情報源（実験，調査）の選択を含む場合

次に情報源（実験）の選択をも含んだ意思決定問題を考えてみよう．そのため，そこで利用可能な実験の全体を E とし，$E=\{e_1,\cdots,e_L\}$ とする．L 個の利用可能な実験があるが，そのうちの1つだけを選択することが許されるものとしよう．どの実験を選択するかという決定も1つの決定であるが，ここでは D の要素に含めないで別個に扱うことにする．したがって，決定の空間 D，状態の空間 Θ は前と同様に考える．効用関数としては，前と同じく $u(d,\theta)$ を考え，実験の費用を別個に考慮することもできるが，これらをまとめて積空間 $E\times D\times\Theta$ 上に効用関数 $u(e,d,\theta)$ を考えることができる（より一般には，積空間 $E\times\mathcal{X}\times D\times\Theta$ 上の効用関数を考える場合もある．（鈴木(3)参照））．

普通，実験によって標本空間が異なるから，実験 e_l のための標本空間を $\mathcal{X}(e_l)$，あるいは，$\mathcal{X}_l$ で表すことにする．そして，自然の状態 $\theta(\in\Theta)$ が真のとき，実験 e_l の実験結果の分布の確率関数あるいは密度関数を $p(x^{(l)}|\theta,e_l)$ で表す（$l=1,\cdots,L$）．$\mathcal{X}(e_l)$，$p(x^{(l)}|\theta,e_l)$ が $\forall e_l\in E$，$\forall x^{(l)}\in\mathcal{X}(e_l)$，$\forall\theta\in\Theta$ に対して定められ，かつ，事前分布 $p(\theta)$ が想定されているものとすると，実験の選択および選択された実験の実験結果である標本に対応してとるべき決定の選択という意思決定問題を次のようにして解くことができる．

E の中の任意の実験 e_l に対して，この実験の結果と D の決定とを対応づける決定関数の全体を $\mathcal{D}(e_l)$ で表し，その要素である決定関数を $d(\cdot|e_l)$ で表す．前項でも触れたのであるが，決定問題を解くための2通りの方法がある．それらは展開型分析および正規型分析と呼ばれるものである．

A. 展開型分析

次項では，決定樹を用いて簡単な例について説明するのであるが，展開型分析では，すでに実験 e_l を選び，かつ実験結果 $\tilde{x}^{(l)}=x^{(l)}$ を得たという条件で選択すべき最適決定は何かを考える．このとき，決定 $d\in D$ の条件付期待効用を

$$u(e_l,d|x^{(l)})=E[u(e_l,d,\tilde{\theta})|\tilde{x}^{(l)}=x^{(l)}] \tag{2.21}$$

と定義するのは自然である．したがって，

$$\max_{d \in D} u(e_l, d|x^{(l)}) = u(e_l, d^*(x^{(l)}|e_l)|x^{(l)}) \qquad (2.22)$$

を満たす決定 $d^*(x^{(l)}|e_l)$ が $(e_l, x^{(l)})$ に対応する最適決定ということになる．

ついで，上の手続きを $\mathscr{X}(e_l)$ の各 $x^{(l)}$ について行うことにより，実験 e_l に対応する最適決定関数 $d^*(\cdot|e_l)$ を求める．

実験 e_l の効用は，$d^*(\cdot|e_l)$ を用いて，

$$u(e_l) = E[u(e_l, d^*(\tilde{x}^{(l)}|e_l)|\tilde{x}^{(l)})] \qquad (2.23)$$

と定義されるので，結局

$$\max_{e_l \in E} u(e_l) = u(e_{l^*})$$

によって，最適の実験 e_{l^*} を選ぶことができる．もちろん，実験 e_{l^*} を採用したとき，とるべき最適決定関数は $d^*(\cdot|e_{l^*})$ である．

展開型分析の手続きは，動的計画法でよく用いられる逆の帰納法といわれるものである．その特徴は，時間的経過を含む決定問題の動学的構造に即して素直に考える立場であるといえよう．動学的決定問題や複雑な決定問題を取扱うためには，展開型分析は非常に優れている．

B. 正規型分析

本来動学的な決定問題であっても，強引に一時点問題として解こうとするのが正規型分析である．したがって，展開型分析の逆の帰納法を用いることなく，意思決定主体は実験の選択を行う以前の時点に位置して，将来のとりうる選択（実験と決定の選択）と生起しうる実験結果のあらゆる組合せを考えて最適解を求めようとするのが正規型分析である．これは，$(e_l, d(\cdot|e_l))$ の効用を

$$u(e_l, d(\cdot|e_l)) = E[u(e_l, d(\tilde{x}^{(l)}|e_l), \tilde{\theta})] \qquad (2.24)$$

と定義して，

$$\max_{l} \max_{d(\cdot|e_l) \in \mathscr{D}(e_l)} u(e_l, d(\cdot|e_l))$$
$$= u(e_{l^*}, d^*(\cdot|e_{l^*})) \qquad (2.25)$$

によって，最適実験 e_{l^*} とそれに対応する最適決定関数 $d^*(\cdot|e_{l^*})$ が求められるというものである．

特に，動的意思決定問題を実際に解く場合には，正規型分析は非常に複雑となるので多くの場合適当ではないであろう．

C. 頑健性（感度）の検討

ベイズ推論における頑健性あるいは感度の分析についてすでに言及した．決定問題の場合には，事前分布あるいは事前密度 $p(\theta)$，尤度 $p(x|\theta)$ のほかに効用関数 $u(d, \theta)$ とか $u(e, d, \theta)$（あるいは $u(e, x, d, \theta)$）も意思決定主体によって主体的に評価，選択されるものである．これらはやはり，一意的に定まるというよりはある程度の幅をもって考えられることが多い．そこで，$p(\theta), p(x|\theta), u(e, d, \theta)$ を設定して最適解（たとえば，$e_{l^*}, d^*(\cdot|e_{l^*})$）を求めた上で，$p(\theta), p(x|\theta), u(e, d, \theta)$ のそれぞれの考えられる幅の範囲内でのずれが最適決定にもたらす影響を評価，検討することが必要である．特に，頑健性に欠ける場合（感度が鋭敏な場合）の最適解の不安定性には注意が払われねばならない．ここでは，紙幅の都合上，頑健性の検討の必要性を強調するにとどめる（French(12)には基本的な例について頑健性の詳しい説明がなされている）．

2.3　統計的意思決定問題の例

いま，ある意思決定問題に関して次のような定式化が得られたとする．すなわち，$D=\{d_1, d_2\}$, $\theta=\{\theta_1, \theta_2\}$, $E=\{e_1, e_2\}$, $\mathscr{X}(e_1)=\{x_1^{(1)}, x_2^{(1)}, x_3^{(1)}\}$, $\mathscr{X}(e_2)=\{x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}\}$ であるとし，効用関数を，

$$u(e_l, d_i, \theta_j) = u_{ij}^{(l)} \qquad (i, j, l = 1, 2) \qquad (2.26)$$

とし，確率分布については，まず，事前分布を $p(\theta_i)=p_i\ (i=1, 2)$ とし，実験 e_1 について

$$\left.\begin{aligned} p(x_i^{(1)}|\theta_1) &= p_{1i}^{(1)} \\ p(x_i^{(1)}|\theta_2) &= p_{2i}^{(1)} \end{aligned}\right\} \quad (i = 1, 2, 3) \qquad (2.27)$$

とおく．もちろん，$\sum_{i=1}^{3} p_{ji}^{(1)}=1\ (j=1, 2)$ である．同じく，実験 e_2 については，

$$p(x_i^{(2)}|\theta_j) = p_{ji}^{(2)} \qquad (i = 1, 2, 3, 4\ ;\ j = 1, 2) \qquad (2.28)$$

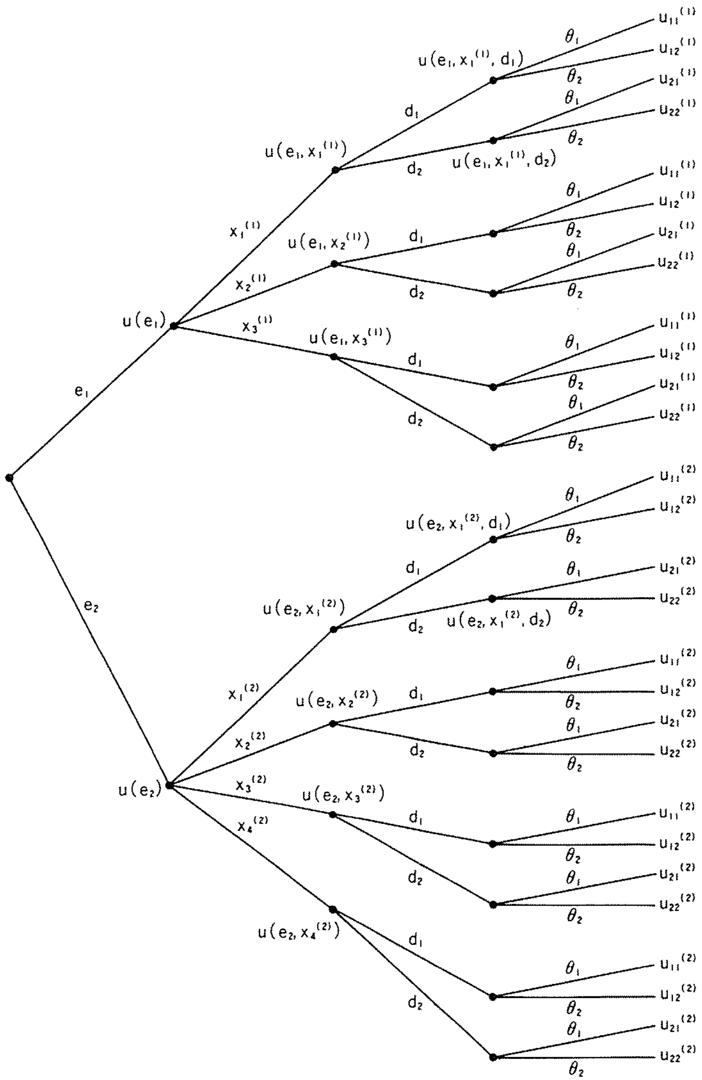


図1 決 定 樹

とする.

上のようにすべての空間が有限集合であれば決定問題を決定樹（decision tree）といわれる図に表して考えると便利である（図1）．そこでは決定問題が段階的に表される．すなわち，時間の経過に従って，まず最初に実験の選択があり，ここで e_1 と e_2 に枝分かれする．次の段階は実験を行い，実験結果による枝分かれができる．したがって，各実験を表す枝の先端から実験結果を表す枝が出る．e_1 からは3本の枝，e_2 からは4本の枝が出ている．第3の段階はわれわれの決定の選択であり，ここでさらに d_1, d_2 の2本の枝に枝分かれする．最後に，自然の状態が θ_1, θ_2 の2つの場合があるので，これに対応して2本の枝に枝分かれする．このようにしてできる枝をたどると，第1段階から第4段階までが $(e_l, x_i^{(l)}, d_j, \theta_k)$ で表されるが，第4段階の枝の先端に効用 $u(e_l, d_j, \theta_k) \equiv u_{jk}^{(l)}$ を書くことにする．このようにして描かれた図はちょうど木が横になっているように見える．したがって，この図は決定樹と呼ばれるのである．この図を用いて，いわゆる展開型分析と呼ばれる考え方を説明してみよう．これは逆の帰納法（backward induction）に従って考えるものである．

たとえば，実験 e_1 を選択し，実験結果 $x_1^{(1)}$ を得て，決定 d_1 を選択したとする．すなわち，図1において，$e_1, x_1^{(1)}, d_1$ と記した枝をたどって，枝 d_1 の先端にいるとする．この結果の評価のためには，この時点において θ_1, θ_2 が真である（事後）確率 $p(\theta_1|x_1^{(1)}, e_1)$，$p(\theta_2|x_1^{(1)}, e_1)$ を用いた効用 $u(e_1, d_1, \theta_1)$，$u(e_1, d_1, \theta_2)$ の期待値

$$\sum_{i=1}^{2} u(e_1, d_1, \theta_i)\, p(\theta_i|x_1^{(1)}, e_1) \equiv u(e_1, x_1^{(1)}, d_1) \qquad (2.29)$$

が計算されねばならない．$u(e_1, x_1^{(1)}, d_1)$ は枝 $(e_1, x_1^{(1)}, d_1)$ の効用であると考えられるものであるから，$u(e_1, x_1^{(1)}, d_1)$ を枝 d_1 の先端に記すことにしよう．ここで，もちろん，

$$p(\theta_i|x_1^{(1)}, e_1) = \frac{p(x_1^{(1)}|\theta_i, e_1)\, p(\theta_i)}{\sum_{j=1}^{2} p(x_1^{(1)}|\theta_j, e_1)\, p(\theta_j)} \qquad (i=1, 2) \qquad (2.30)$$

である．事後確率 $p(\theta_i|x_1^{(1)}, e_1)$ を枝 $(e_1, x_1^{(1)}, d_1)$ の先端から枝分かれしている枝 θ_i に付記しておくのも便利である．同様にして，実験 e_1 を選択し，

実験結果 $x_1^{(1)}$ を得たとき，決定 d_2 を選択した場合を考えて，$u(e_1, x_1^{(1)}, d_2)$ を計算し，枝 $(e_1, x_1^{(1)}, d_2)$ の先端にそれを記す．

次に，1段階もどって，実験 e_1 を選択し，実験結果 $x_1^{(1)}$ を得，まさに決定の選択を行おうとしているものとする．すなわち，枝 $(e_1, x_1^{(1)})$ の先端において，d_1, d_2 のいずれの枝をたどるべきかの岐路に立っているものとする．この選択は $(e_1, x_1^{(1)}, d_1)$ と $(e_1, x_1^{(1)}, d_2)$ の効用の比較によってなされるべきである．したがって，

$$u(e_1, x_1^{(1)}, d_1) \underset{(<)}{\geqq} u(e_1, x_1^{(1)}, d_2) \longrightarrow \underset{(d_2)}{d_1} \tag{2.31}$$

によって，この段階の最適決定が決められる．これを $d^*(x_1^{(1)}|e_1)$ で表す．そして枝 $(e_1, x_1^{(1)})$ の先端には，

$$u(e_1, x_1^{(1)}, d^*(x_1^{(1)}|e_1)) \equiv u(e_1, x_1^{(1)}) \tag{2.32}$$

で定義される $(e_1, x_1^{(1)})$ の効用 $u(e_1, x_1^{(1)})$ を付記しておく．

以上と同様にして，実験 e_1 を選択し，実験結果 $x_2^{(1)}$ を得た場合について $u(e_1, x_2^{(1)}, d_1)$，$u(e_1, x_2^{(1)}, d_2)$，$d^*(x_2^{(1)}|e_1)$ および $u(e_1, x_2^{(1)})$ を求め，また，実験結果 $x_3^{(1)}$ を得た場合については $u(e_1, x_3^{(1)}, d_1)$，$u(e_1, x_3^{(1)}, d_2)$，$d^*(x_3^{(1)}|e_1)$ および $u(e_1, x_3^{(1)})$ を求める．

さて，さらに1段階もどって，実験 e_1 を選択したが，まだ実験結果を得ていない段階に立っているものとしよう．決定樹においては，枝 e_1 の先端にあり，3つの枝 $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}$ のいずれがとられる枝となるかは不確実な状況にあるのである．しかし，この不確実性は確率 $p(x_1^{(1)}|e_1)$，$p(x_2^{(1)}|e_1)$，$p(x_3^{(1)}|e_1)$ によって表現され，また，

$$p(x_i^{(1)}|e_1) = \sum_{j=1}^{2} p(x_i^{(1)}|\theta_j, e_1)\, p(\theta_j) \qquad (i = 1, 2, 3) \tag{2.33}$$

によってこれらの確率は求められる（これらは実験 e_1 の実験結果の周辺分布を与えている）．したがって，実験 e_1 の効用（期待効用）は，

$$\sum_{i=1}^{3} u(e_1, x_i^{(1)})\, p(x_i^{(1)}|e_1) \equiv u(e_1) \tag{2.34}$$

によって定義される．

まったく類似の計算を実験 e_2 について，第4段階から1段階ずつもどりながら実行すれば，$u(e_2, x_i^{(2)}, d_j)$，最適決定 $d^*(x_i^{(2)}|e_2)$，$u(e_2, x_i^{(2)})\,(=u(e_2, x_i^{(2)}, d^*(x_i^{(2)}|e_2)))$，実験 e_2 の効用 $u(e_2)$ が求められる．

実験の選択はこのようにして得られた実験の効用の大小の比較にもとづいてなされる．すなわち，

$$u(e_1) \underset{(<)}{\geqq} u(e_2) \longrightarrow \underset{(e_2)}{e_1} \tag{2.35}$$

である．等号の場合は e_1, e_2 のいずれを選択してもよい．

以上の分析では，最適の実験が選択されると同時に，この実験を用いたときの最適決定の選択の仕方も確定されたことになる．たとえば，最適決定が e_2 であれば，可能な実験結果に対応する最適決定は $d^*(x_i^{(2)}|e_2)\,(i=1, 2, 3, 4)$ として既に得られているのである（ここで，(2)のCで述べた頑健性の検討が必要であろう（French(12)参照））．

上の例でみた逆の帰納法は多段階の決定問題あるいは逐次決定問題を解くのにきわめて有力な方法である．注意すべきことは，この方法を可能ならしめているのは θ の事前分布の想定である．もし，事前分布が想定されていなく，それを利用しないのであれば上の逆の帰納法はまったく不可能なのである．

2.4 ベイズ推論あるいはベイズ意思決定理論と主観確率

A. 主観確率

可測空間 $(\Omega, \boldsymbol{B}(\Omega))$ 上の確率測度 $P(\cdot)$ とは，

(i) $P(A) \geqq 0 \quad (A \in \boldsymbol{B}(\Omega))$ かつ $P(\Omega) = 1$

(ii) $\boldsymbol{B}(\Omega) \ni A_i \quad (i = 1, 2, \cdots)$ かつ $A_i \cap A_j = \phi \quad (i \neq j)$ であれば，

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{（完全加法性）}$$

が成り立つ．

という条件を満たすものとして定義される．いいかえれば，確率測度とは $P(\Omega)=1$ を満たす測度のことである．これは数学的な定義である（ここで問題があるとすれば，完全加法性が強すぎるので，有限加法性で置き換える

ことである).

統計学ではさらに，事象 A に対して確率 $P(A)$ をいかに定めるか，あるいは，確率をどのようなものと解釈するかがきわめて重要である.

これまで，ベイズ統計学とは何かについて述べた．そこでは，推論（あるいは意思決定）主体（以下では単に主体という）という言葉が用いられ，事前密度 $p(\theta)$ とか条件付密度（あるいは尤度）$p(x|\theta)$ は主体が評価，選択すべきものと述べられている．このように主体の役割がベイズ推論（あるいは意思決定理論）では重視される．主体が $p(\theta)$ と $p(x|\theta)$ を設定しさえすれば，事後密度 $p(\theta|x)$ が求められ，主体は $p(\theta|x)$ にもとづいてあらゆるタイプの推論を行うことができるのである.

ここでは，主体が確率をどのように考えているかによって次の2つの場合があげられよう.

(1)　主体が「確率とは相対頻度の極限である」といういわゆる頻度論者である場合をまず考えてみよう．$P(A)$ を頻度論の立場で定義できるためには，事象 A の生起，非生起を結果とする試行が独立に無限回繰返されるというベルヌーイ試行列が考えられねばならない．この立場を固執する限り，ベイズ推論あるいはベイズ意思決定理論が適用できる範囲は非常に限定される．というのは，社会的，経済的事象のみならず，自然界における事象でさえも，ほとんどの事象について頻度論的確率を考えることはできないからである．したがって，頻度論者であってベイズ統計学の立場をとるという主体にとっては，統計的推論や意思決定理論はほとんどの問題に適用できないことになる.

だから，頻度論者である主体にとって非ベイズ推論は救いと感ぜられることであろう．というのは，非ベイズ推論では $p(x|\theta)$ だけが必要であって，頻度論者にとって難物である事前密度 $p(\theta)$ の評価を必要としないからである．$p(x|\theta)$ は（原理的には同一条件で多数回の繰返しが可能な）実験に関するものと考えられ，確率の頻度論的解釈がなじみやすい場合が少なくないからである.

それでもしかし，少し厳密に考えれば，$p(x|\theta)$ でさえも，主体が（頻度論的解釈とは直結せずに）総合的判断によって選択していることが分かるであろう.

(2)　主体が確率をすべて主観確率と解釈している場合を次に考えてみよう．ここで，主観確率（あるいは個人確率）について詳しく論ずることはできないが，要するに，主体が事象 A の生起の不確実性の程度を評価し，表明したものが，その主体の主観確率 $P(A)$ であるとしておこう．ここで重要なのは，主体はいかなる事象に対しても主体的に主観確率を評価しうる能力をもつという大前提である（主観確率については，de Finetti, L. J. Savage 等は数学的（公理的）基礎の研究とともに，主観確率の理解を広めるために大きな貢献をしてきた).

この前提は，理想的な，完全に理性的な主体（理性人）は「かくあるべし」という規範的なものである．このような主体は，自己の知識・情報の状態に対応して，いかなる事象の生起に関する不確実性の程度も（主観）確率の形でごく自然に評価できるわけである．しかし，現実に生きる主体にとっては，これは決して生易しいものではないであろう．むしろ非常に難しい場合が多いのであるが，現実の主体は，自己の知識・情報の状態に対応して，自己の経験・理論，好み，さらには哲学などもとり込んだ総合的判断として，事象 A の生起の不確実性についての自己の意見を主観確率 $P(A)$ という形で表明するのである．そこでは，当然，主体は自己が評価した主観確率に責任を持つべきである.

少なくとも規範的にみたとき，主観確率の立場をとる主体にとって，ベイズ推論とベイズ意思決定理論はまさに論理的必然性をもって直結するものである．主体は常に事前密度 $p(\theta)$ と尤度 $p(x|\theta)$ を評価し，設定できるからである．もちろん，実際には，主体はこれらを主体的に定めるために努力しなければならない.

ここでぜひ補足しておくべきことがある．それは頻度論的確率と主観確率との関係についてである．事象 A に関するベルヌーイ試行列の観測結果にもとづいて頻度論的確率 $P(A|F)$ が評価されるとき，主観確率の立場をとる主体は，同じ観測結果をとり入れた自己の知識・情報の状態に対応した，自己の主観確率 $P(A|S)$ を主体的に評価する．多くの場合に，$P(A|F)$ と $P(A|S)$ とはほとんど一致することであろう．このような意味で，主観確率

は，頻度論的確率を特別な場合の主観確率として含むと考えられる．その他の確率の解釈，たとえば，等確率の概念にもとづく確率，論理確率なども，そこで利用できる情報を主体が積極的に認めた上で，問題とする事象の主観確率を主体が評価することになる．したがって，これらの確率も主観確率の特別の場合とみなされるのである．

3. 文献解題

以下ではこれからベイズ統計学を勉強しようと考えている学生・大学院生・研究者のための手引となる文献を紹介する．一般的な教科書と定評のある参考書だけを挙げておく．もちろん，すべてを網羅しているわけではないことをお断りしておかねばならない．論文については紙幅の制限のため省略した．

(1)　鈴木雪夫　1987　『統計学』，朝倉書店．

(2)　鈴木雪夫　1978　『統計解析』，筑摩書房．

(3)　鈴木雪夫　1973　『経済学のための数学入門（下）』，日本評論社．

(4)　鈴木雪夫　1975　『経済分析と確率・統計』，東洋経済新報社．

(1)〜(4)のすべては統計推論あるいは意思決定理論を一貫してベイズ統計学の立場で述べている．(3)と(4)の題名の経済学とか経済分析という言葉は本の内容とはあまり関係はない．

(5)　繁桝算男　1985　『ベイズ統計入門』，東京大学出版会．

統計学の標準的問題について詳述していると思う（題名に入門とあるが，初等という意味ではない）．理工系の学生にも適当と思われる．

(6)　松原　望　1977　『意思決定の基礎（現代人の統計 4）』，朝倉書店．

主としてベイズ統計学の立場で意思決定理論が扱われている．政治学などの例がある．

(7)　宮沢光一　1971　『経済分析と決定理論』，東洋経済新報社．

(8)　宮沢光一　1971　『情報・決定理論序説』，岩波書店．

(7)は主観確率をいかに決定するかという問題からはじめて，意思決定問題，情報の価値，多段階決定問題などが論じられる．(8)はシャノン流の情報量と情報価値との関係，適応過程・最適決定過程，最適制御過程などを論じている．

(9)　Berger, J. O.　1980　*Statistical Decision Theory : Foundations, Concepts, and Methods*, Springer-Verlag.

主としてベイズ統計学の立場で統計的推論および決定理論が述べられている．豊富な題材が意欲的にとり入れられている．なお，『経済学論集』東京大学経済学会，第 47 巻第 3 号，1981 年，には筆者による書評がある．

(10)　Box, G. E. P., and Tiao, G. C.　1973　*Bayesian Inference in Statistical Analysis*, Addison-Wesley.

2 人の著者はともに，現実の問題の統計的研究に豊富な経験をもつ．この本には理論と応用が述べられているが，特に応用の部分に分散分析，実験計画が含まれているのは特徴的である．きわめて意欲的な著作である．

(11)　DeGroot, M. H.　1970　*Optimal Statistical Decisions*, McGraw-Hill.

著者は米国におけるベイズ統計学の最有力な指導者の一人．本書は大学院レベルの教科書として定評を得ている．

(12)　French, Simon　1986　*Decision Theory : An Introduction to the Mathematics of Rationality*, Ellis Horwood Limited.

著者は広い視野と柔軟な思考力をもつ研究者であると思われる．本書では効用理論，主観確率，意思決定問題等について，平易な水準からかなり高度な水準に至るまで，粘り強く記述される．これは英国流のスタイルなのであろう．内容は非常に興味深く，得るところははなはだ多い．なかんずく，第 8 章（グループ決定と社会的選択）は力作であると思われる．

(13)　Lindley, D. V.　1965　*Introduction to Probability and Statistics from a Bayesian View-Point*, Cambridge Univ. Press.（竹内啓，新家健精訳『確率統計入門』，培風館，1969.）

(14)　Lindley, D. V.　1972　*Bayesian Statistics, A Review*, Society for Industrial and Applied Mathematics.

(13)，(14)の著者は英国におけるベイズ統計学の提唱者であり，指導的な統計学者である．(13)は理工系の学部レベル，(14)は大学院レベルである．なお(14)は，『経済学論集』東京大学経済学会，第 40 巻第 4 号，1975 年，に筆者に

よる書評が掲載されている．

(15) Markowitz, H. M. 1959 *Portfolio Selection*, John Wiley.（鈴木雪夫監訳『ポートフォリオ選択論』，東洋経済新報社，1969.）

本書の後半はベイズ意思決定理論を論じ，ポートフォリオ選択への応用が述べられ興味深いものと思う．

(16) Press, J. 1972 *Applied Multivariate Analysis*, Holt, Rinehart and Winston.

(17) Press, J. 1982 *Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference*, Krieger.

(16)，(17)はともに，主としてベイズ統計学の立場で多変量解析の手法を導く．応用例も多く興味深い．なお(16)は，『経済学論集』第 39 巻第 3 号，1973 年，に筆者による書評が掲載されている．

(18) Raiffa, H., and Schlaifer, R. 1961 *Applied Statistical Decision Theory*, Harvard Univ.

ベイズ統計学の古典ともいうべきもの．興味深い応用問題が多い．

(19) Rosenkrantz, R. D. 1977 *Inference, Method and Decision : Towards a Baysian Philosophy of Science*, D. Reidel.

科学哲学の色彩の濃い非常に興味深い著作である．ベイズ統計学の基礎を科学哲学の面から論ずる．文章はやや難しい．科学哲学あるいは科学思想に関する記述も多い．

(20) Savage, L. J. 1954 *Foundations of Statistics*, John Wiley.

まさにベイズ統計学の古典といえる．著者の強烈な個性がうかがえる．難解という定評があり，そのためか邦訳はない．しかし，ベイズ統計学を一応理解した人にはそれほど難しくはないと思う．

(21) Winkler, R. L. 1972 *An Introduction to Bayesian Inference and Decision*, Holt, Rinehart and Winston.

経済や経営の学部学生に適当なレベルのもの．説明は懇切でわかりやすい．推論と意思決定のために必要な事項はすべて含まれている．

(22) Zellner, A. 1971 *An Introduction to Bayesian Inference in Econometrics*, John Wiley.（福場庸・大沢豊共訳『ベイジアン計

量経済学入門』，培風館，1986.）

著者は計量経済学者であるが，本書は数理統計学的にも水準は高い．多変量解析が主題であるが，回帰モデルをはじめ計量経済モデルに関する統計推論がベイズ統計学の立場でかなり徹底的に論じられている．大学院レベルの教科書としても優れたものと思う．

I
理　　論

第1章
主　観　確　率

新　家　健　精

1.　主観確率の源流

主観あるいは主観的確率といっても字義通りのまったくの恣意性をもったままでは理論形成の意味がない．ここでは主観確率の整合性がどのように検討されてきたかについてみることにしよう．

1.1　確信の度合

ベイズ統計学は決定の場を背景とした立論でもあり，そこで採用される確率概念も意思決定あるいは行動選択に関わった解釈がほどこされねばならない．その意味ではT.ベイズの提唱以来，最近にいたるまで「決定」といった局面との関係で確率が論ぜられなかったのは不思議ともいえよう．

いずれにしても主観確率は，サイコロを振ったとき1の目が出る確率という場合，頻度論的解釈のように多数回試行において1の目の出る回数の割合が限りなく1/6に近づく，あるいはさらにそれを精緻化した大数法則といった内容のものではなく，古典的解釈として知られているように，1から6の目までが等可能であり，1の目はその中の1つに相当する事象で1/6を付与するといった公平なチャンスゲームを背景とする考え方にも相当しない．それは次にサイコロを振った場合の1の目の出易さに対する実験者なり統計家の確信，あるいは信念の度合（degrees of belief）といったものである．つまり確実に生起するということへの信念の度合に1を，確実に生起しないことへ0を付した場合の相対的な信念の度合の大きさといったものである．

主観確率の源流となったF. P. ラムゼイはその所論の冒頭で次のように述べている（Kyburg and Smokler[1964]所収）.

「……私には二つの方法があるように思われる．第一は確信の度合としてその所有者に感知可能なものを仮定することである．たとえば確信の度合は随伴して生起する感情の強度に左右されるであろう．これはいってみれば確信に関する感情もしくは信条とでも呼ぶべきものであろう．……しかし，この見方はきわめて不都合のように思われる．というのは感情の強さに対して数値を付与するのは容易ではないからである．そこで第二の仮定，つまり確信の度合はその原因的属性，言い替えればそれにもとづいて行動を起こしたくなるような度合といったものであるとする仮定へと進んでゆこう．……この観点は確信の度合を量的に取り扱う場合には唯一の見方のように思われる……」

彼のこの見解の中には不確実性下における人間の行動に対する深い洞察が込められているといってよい．たとえば企業家が不確実性下においてある意思決定を行おうとする場合，そこには事業展開に対する信念や情熱の強さと，成算に対する冷静な判断とが併存していることになろうが，ラムゼイは後者の冷静な判断のみを抽出しようというのである．

彼はこのような見解を背景に，現実の状況にあってはそれがかなり図式的であると断わりつつも，選択行為における主観確率を次のように導いてくる．まず彼が「価値」とも「世界」とも呼んでいる選択対象について，それが測定可能であるための公理系の中心概念として確信の度合が1/2であると信ずるに足る倫理的に中立な命題（ethically neutral proposition）pの存在を仮定する．そして選択対象αとβの差がγとδとの差に等しいことを，上述の中立命題pについて，pが生起した場合α，生起しない場合δなる選択と，pが生起する場合β，生起しない場合γなる選択とが無差別であることをもってする．この基本概念をもとにラムゼイは選択対象と実数との1対1対応，すなわち基数的効用関数を導出してくるのである．

こうしてすべての選択対象に対する価値評価が終了した後に，任意の事象Aに対する確信の度合p_rを次のようにして決定する．すなわち，αを確実に入手可能であるとする賭と，Aが生起すればβ，生起しなければδであるとする賭とがまったく公平という意味で無差別であるならば，$\alpha=p_r\beta+(1-p_r)\delta$が成立し，これから

$$p_r = \alpha-\delta/\beta-\delta$$

として決定しようというのである．ラムゼイはこうした展開をより数理的に厳密な公理系の下で行っており，これらの公理系から導き出された主観確率が条件つき確率系として，いわゆる確率の一般的性質を満たしていることを帰結している．

1.2 コヒーレンス概念

主観確率に対する信念の度合としての方向とはまったく独立にB. ド・フィネッティはコヒーレンス（coherence）と呼ばれる論理的整合性をもった概念を展開した．

今，賭に参加する局面を想定する．参加者は値$ps\,(0<p<1)$を支払って，事象Eが生起すれば賞金sを獲得し，生起しなければ何も得ないという賭に加わるものとする．このとき数pを個人のEの生起に対する確からしさの測度（the measure of the degree of probability）と呼ぶ．また参加者にとって確実に利得が保証されていたり，逆に確実に損失を蒙ることがわかっていたりする局面については，個人の不確実性に対する確率の評価にインコヒーレンス（incoherence）が含まれているといい，そうでない場合の確率付与はコヒーレント（coherent）であるということにする．

以下，事象Eは単一の事象を表すものとしよう．上述の賭の局面を事象系$E_1, E_2, \cdots, E_n$に拡張することとして，試行においてはいずれかの事象が必ず生起し，1度に2つ以上の事象は生起しない完全なクラスを想定する．この賭に対して参加者は参加料$\sum_i p_i s_i$を支払うものとする．ここで個人がこれら各事象に付与する確率測度を$p_1, p_2, \cdots, p_n$，このときの賞金をそれぞれ$s_1, s_2, \cdots, s_n$とする．すると事象$h\,(h=1, 2, \cdots, n)$が生起した場合の獲得金額は

$$g_h = s_h-\sum_i p_i s_i \qquad (h = 1, 2, \cdots, n)$$

で与えられる．いまs_hを未知数とすれば，これらn本からなる連立方程式

の係数行列式の値は

$$\begin{vmatrix} 1-p_1, & -p_2, & \cdots, & -p_n \\ -p_1, & 1-p_2, & \cdots, & -p_n \\ -p_1 & -p_2, & \cdots, & 1-p_n \end{vmatrix} = 1-(p_1+p_2+\cdots+p_n)$$

である．ここでこの値が0でないとすると，上述の連立方程式においてすべての g_h が正となるような賞金の解 s_h の可能性が示されることになる．この場合には先の意味で各 p_i の付与はインコヒーレントであり，したがって，コヒーレンスを維持するためにはこの値が0，つまり $\sum p_i=1$ が成立していなければならない．またこれらについては逆に $p_i \geqq 0\,(i=1,2,\cdots,n)$ であるとして，$\sum p_i=1$ ならば公平な賭の条件 $\sum p_i g_i=0$ を考えた場合，すべての g_h が正となり得る可能性が排除され，結局，十分であることがわかる．つまりコヒーレンスにより，完全事象のクラスについてはそれらの確率の和は1でなければならないことが示されたわけである．

条件つき確率あるいは乗法定理についてはどうであろうか．いま，E'' なる条件の下で E' の条件つき事象 $E'|E''$ について参加料 ps を支払って，これが生起した場合 s' なる賞金が獲得できる賭に参加するとしよう．この場合，E'' が生起しない，E'' が生起するとして E' が生起する，しないの計3通りの場合が考えられ，結果としての損得はそれぞれ0，$(1-p)s$，$-ps$ となる．そこでこれを基準として，E' に対しては賞金 s' が，E'' に対しては s'' が，$E'|E''$ に対しては s が与えられる3つの賭に参加することを考えよう．ただし上述のそれぞれの事象に対する参加者の主観的判断は p', p'', p である．それぞれの獲得金額は，$E'|E''$ の場合 $g_1=(1-p')s'+(1-p'')s''+(1-p)s$ であり，E' の場合 $g_2=p's'+(1-p'')s''-ps$，E'' の場合 $g_3=-p's'-p''s''$ となる．そこで再びこの賭における s', s'', s を未知数とする連立方程式体系についての係数行列式の値に注目すると，

$$\begin{vmatrix} 1-p', & 1-p'', & 1-p \\ -p', & 1-p'', & -p \\ -p', & -p'', & 0 \end{vmatrix} = p'-pp''$$

であり，したがってコヒーレンスを維持するためには前と同様に $p'=pp''$ でなければならない．この場合，これを事象に対応させると $p(E'E'')=p(E'')$

$\cdot p(E'|E'')$ となり，ここに乗法定理の導出が示される．

ド・フィネッティ（Kyburg and Smokler[1964]所収）はこの結果について次のように述べている．「これらの結論は主観的観点からの確率としてどのような事実が理解されねばならないか，そしてそれらがどのようにして導き出されるかについて，きわめて手短かに不完全な方法で要約しただけである」と．

2. 公理的アプローチによる主観確率

本節では公理系による主観確率付与の考え方について，具体的内容を重視しながら簡単に説明するとともに，前提条件や問題点などについてふれることにしよう．

2.1 主観確率の付与

いま，事象を記号 E, F, G などで表すこととし，2つの事象 E, F について $E>F$ と書いて，E は F よりも生起しやすい，$E\sim F$ は E と F との生起の大きさが同等であると読むことにする．そこでまず事象間の生起について，個人の判断の整合性を要求する次の仮定を設けよう．

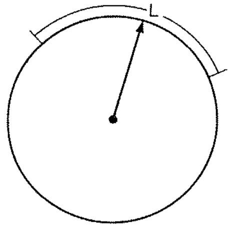
仮定1

(1) 任意の2つの事象 E, F について
　$E>F$ か，$F>E$ か，$E\sim F$ のいずれかが成立する．

(2) $E>F$ かつ $F>G$ ならば，$E>G$ である．

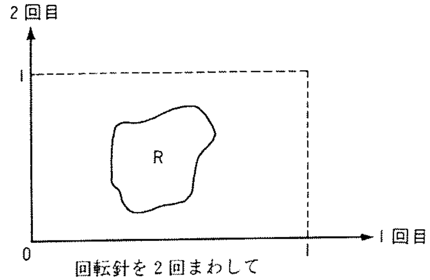
この仮定の意味するところは，(1)では任意の2つの事象の生起の大きさに関して常に比較可能であることを要求するものであり，(2)については，それが論理的矛盾に陥らないように推移性が保持されていることを要求するものである．この結果は，当面問題とする事象がいくつか与えられた場合，生起のしやすさについて順序が成立することを述べている．

さて次に以下のような標準実験装置を考えることにしよう．標準実験1（図1）は滑らかに回る回転針であり，標準実験2（図2）は標準実験を2回繰返した場合の図的表現としての装置である．



同図の長さが1の
回転針の装置

図1　標準実験1



回転針を2回まわして
得られる図的表現

図2　標準実験2

そこでこれら標準実験において事象 E に対応する周囲の長さが a，あるいは領域 R の面積が a のとき $p(E)=a$ と書いて事象 E の生起する確率は a であるということにしよう．たとえばサイコロを振って1の目の出る事象は中心角60°の弧に相当する長さとなろう．確率論に詳しい読者であるならば標準実験が幾何確率を表現していることが理解できるであろう．

問題は与えられた事象 E とその確率の大きさを標準実験を通じてどのように対応づけるかである．公理系のアプローチではここにさまざまな論理構成上の工夫がほどこされ，それに応じた厳密な展開が行われているが，ここでは決定分析における主要な決定ルール選択のための規準としての期待効用仮説を背景として，個人の効用のあり方について仮定1と並行的な次の仮定2を設定しておこう．

一般の対象物を A, B, C……とする．またここにはその確率的合成としてのくじ，すなわち確率 α で A を，確率 $(1-\alpha)$ で B が得られるといったくじをも含めて考えることにする．任意の2つの対象 A, B に対して $A \gg B$ と書いて B よりも A を選好する，$A \approx B$ を A と B は選好に関して無差別と読むことにしよう．

仮定2

(1)　任意の2つの対象 A, B について，$A \gg B$ か $B \gg A$ か $A \approx B$ のいずれかが成立する．

(2)　$A \gg B$ かつ $B \gg C$ ならば，$A \gg C$ である．

以上の事象の確からしさおよび対象の選好に関する仮定と標準実験を用い

て，事象 E の確率を次のように決定することができる．

例1　A氏がアメリカ大統領に選出される事象 E に対する個人の確率は，くじ l_1 として事象 E が生起すれば10ドルを得，事象 $\bar{E}$ が生起すれば10ドル損をする構成を考え，くじ l_2 として標準実験1で区間 L に針が止まれば10ドルを得，そうでなければ10ドル損をする，を考える．区間 L の長さを適当に調整することによって，選好関係としてくじ l_1 とくじ l_2 とが無差別，すなわち，くじ $l_1 \approx$ くじ l_2 の状態を達成することができる．このとき区間 L の長さが L' であったとするならば，$p(E)=L'$ とすればよい．

例2　少し複雑なケースを考えよう．いま c 円所有しているとして，価格 b 円の貴重品に保険をかけるか，かけないかが問題である．保険金を p 円 $(b \gg p)$ として，海外旅行などで貴重品の傷む確率 $p(E)$ が焦点である．

この問題は表1のような支払表をもつ決定問題として整理できよう．

表1

行動＼事象	E	$\bar{E}$
保険に加入する	$c-b$	c
保険に加入しない	$c-p$	$c-p$

表2

ギャンブル＼事象	R	$\bar{R}$
ギャンブルに参加する	$c-b$	c
ギャンブル不参加	$c-p$	$c-p$

これと同値な意味での標準実験との問題を構成するには次のようにすればよい．いま，$c-p$ 円所有しているとして，標準実験2でたとえば領域 R に点が落ちたら $b-p$ 円の損失を蒙り，落ちなかったら p 円獲得するという賭を設定する．するとこのギャンブルに参加する場合，点が R 内に落ちる場合には $c-p-(b-p)=c-b$，落ちなかったならば $c-p+p=c$ で，支払表は表2で与えられる．そこでこの場合 R を調整してギャンブルに参加することと，不参加であることとの選好関係が無差別であると判断される R' が得られる．このとき，$p(E)=R'$ が事象 E に対する整合的な確率の値である．

2.2　公理系アプローチの問題点

よく指摘されているように，こうした公理的アプローチは第2次大戦後のアメリカの思潮としての論理実証主義を反映するものであるが，論理的な厳

密性ゆえに多くの批判も存在している．

まず仮定1についてであるが，ここでは個人の不確実性に対する評価について弱順序と呼ばれる整合性が要求されている．論理的には一見自然にみえるかもしれないが，含意するところはきわめて厳格で，われわれの日常的営為ではこれらと矛盾することが多々ある．たとえば第(2)項で任意の2つの事象間の生起の比較可能性が要求されているが，明日の天気と次期アメリカ大統領に関する不確実性を比較したりする場がはたしてあるであろうか．またそこでは，一度下した判断を一定期間変更することなく持続するものであることが暗々裡に仮定されている．論理的展開の中に時間的配慮を組み込ませることは大変に厄介であるため，そうした問題をも派生させている．また第(2)項の仮定は推移律によって順序関係がグルグル回りにならないように首尾一貫性を主張するものであるが，多くの場合2つずつの事象を比較していくと，結果として推移律の不成立がいくつかの対の組合せに生じることが報告されている．

標準実験を通じての賭が登場するが，そうしたギャンブルを引用することについての倫理上の問題はすでにラムゼイの倫理的に中立な命題の存在で十分意識されているところである．ここで含意される幾何確率は数学的にも厳密に定義づけられる確率測度であって，実はこれとの対応が確定できれば確率法則などの成立に関して本質的には問題が生じないといってよい．たとえば，全体の流れからみれば確率で確率を定義しているという内容になっており，幾何確率がもつ客観性に判断を委ねていることになる．

仮定2は選好関係に関するもので内容的には仮定1と並行的な事柄がまず指摘できる．ただし，その背景には普遍的価値表現としての効用の可測性を含んでいる．経済学における基数的効用尺度の不成立を回避あるいは再構成するために，選好の順序概念から出発して期待効用最大化原理へと到達するプロセスは，その論理構造の見事さからいって一度はふれておくことが望ましい．例では金銭表示の値と効用とが比例的であることを仮定しているが，ここにも限界効用逓減の法則に示されるような従来からの基本的問題がある．

また本節の議論は不確実性下の決定問題において，そこに個人としての整合的な効用と主観確率とが与えられた場合，その期待効用を最大化する行動

を選択すべしとする期待効用仮説の立論にそったものである．

いずれにしても，こうした諸仮定なり仮説なりがわれわれの日常の行動においてどの程度貫徹されているかを検証する作業やそれらの諸仮定をゆるめる作業が行われているところである．そして，もしそれらが十分に正当化されるのであれば有力な理論としての座を確保できるといえよう．

3. 具体的付与をめぐって

主観確率をも含むベイズ方式の応用は後でも説明するように数多くの実例をみているが，本節では具体的な確率付与をめぐる問題点を指摘しておこう．

3.1 主観確率と情報

サイコロを振ったとき1の目の出る確率が1/6であるとする古典的解釈では，前提条件としてどの目の出る可能性も等しいことを設定している．ここでは等しい確率が明白に「真に」等しいチャンスゲームを指しているわけであるが，考えてみると現実の局面ではそうした想定は殆ど不可能に近いことがわかる．たとえば1組のトランプを混ぜた場合，トランプの積み札の一番下に赤のカードがくる確率を計算しようとするとき，その確率はカードの混ぜ合わせが公平であれば明らかに1/2といってよいであろうが，すべての赤のカードの表面がねばねばしているといった状況では黒のカードが一番下に生じやすく，賭ける場合には黒のカードということになろう．もっともこの事実がわからなければ1/2となる．このように個人あるいは主観的確率はどのような情報が与えられているかによって決定的な影響を受けることになる．この意味であらかじめ保有している情報を F として，そうした情報の下での事象 E の確率を通常の条件つき確率と同一の表現で $p(E|F)$ と書くとすると，p（一番下が赤カードである｜カードは良く混ぜ合わせられるが，赤の表面はねばねばしている）などとなる．極論するならば，単純な p（一番下のカードは赤である）といった表現は殆ど理想的状況といってよく，どのような確率も上述の意味での条件つき確率として表現されるということになる．

実際に与えられた情報によって，どのように確率付与が影響されるかにつ

いて次の例を考えてみよう．

例 徳川吉宗（第8代将軍）の誕生年について次の A_1～A_4 の事象を想定する．

A_1：彼は1650年以前に生まれている．

A_2：彼は1651～1675年の間に生まれている．

A_3：彼は1676～1700年の間に生まれている．

A_4：彼は1701年以降に生まれている．

a) 上の A_1～A_4 の事象についてどれがもっとも可能性が大きいか推定せよ．もっとも可能性の高い事象について確信の度合をパーセントで表現せよ．またもっとも低い可能性の事象は，これについても確率を付与せよ．

b) 情報として「初代徳川家康は1616年に死去している」が与えられた場合，A_1～A_4 に対する可能性はどうか．c) に行く前に a) のプロセスをもう一度繰り返せ．

c) 情報として「第5代将軍綱吉の死去は1709年である」が与えられた場合はどうか，次のステップに行く前に a) のプロセスをもう一度繰返せ．

d) 情報として「彼が将軍に即位したのは1716年である」が与えられた場合はどうか，a) のプロセスを再度繰り返せ．

（将軍吉宗の誕生年は1684年である）

これからも容易にわかるように確率の大きさは，その背後にある情報に著しく影響されている．

ところで上の例では，あらかじめ不確実性事象の空間 A を A_1～A_4 に分割しておいて，与えられた情報 B の下で，広義の条件つき確率 $p(A_i|B)$ をそれぞれ評価し，もっとも値の高い A_i を選択しようとしている．しかし，こうした事象空間 A に対する分割もいってみれば先験的な基礎情報に依存しているわけで，逆に A の分割が条件事象としての B のあり方にも影響を及ぼしているといってもよい．

3.2 確率法則の適用

追加的な情報によって確率事象に対する評価は変化するが，そうした事実を統計的推測に積極的に，しかもある意味での整合性をもった構造で接近しようとするのが本書のベイズ統計学の基本的立場である．ベイジアン・アプローチでは事前確率が情報を通じて事後確率へと移行するプロセスに注目しているが，そうした方式が常時適用可能か，とりわけ確率付与の場合に有効かというと，そうでない場合が多いことにも留意すべきであろう．

ちなみにベイズの定理あるいは公式を形式的に書くと次のようになる．

$$p(A|B) \propto p(A) \cdot p(B|A)$$

上例のように A, B の事象を仮に，

A：徳川吉宗は1651～1675年の間に生まれた．

B：吉宗の将軍即位年は1716年である．

として，これを形式的にベイズ公式に当てはめてみよう．左辺の $p(A|B)$ は明らかに B という情報が与えられた下での A の確率であるから，いま求めようとしている確率である．右辺の $p(A)$ はこの場合，事前的な A に対する確率ということになろう．このとき $p(B|A)$ はどうであろうか．字義通りに解釈すれば吉宗の誕生年が与えられた下での即位年の確率ということになる．つまり本来の $p(A|B)$ に対してちょうど逆転した内容をもった確率が入り込んでいることになる．ベイズの定理が逆確率（inverse probability）の概念を含んでいるというのはこうした事実を指している．もちろん，A, B に対して時間的因果連鎖を認めずに，並列的に取り扱うかぎりでは定理は成立している．統計的推測の立場では $p(A|B)$ が尤度と呼ばれる既知の構造としての保証が与えられている．否むしろ標本情報として確固たる情報源であるからこそ，立論が可能となっているのであって，一般的な意味での確率の推測形式としてベイズの定理が有効かどうかは長い間統計学史上の論争点となってきたところである．

一方，いま述べたように情報内容が吉宗の即位年といった明確なものではなく，情報そのものに対しても一定の評価を行わねばならない場合がある．来年度の経済状況について何人かの専門家に意見を求める場合には，それぞれから得た情報をなんらかの方式で積み上げ，総合化する必要がある．あるいは逆に直接的に問題となっている事象への確率付与を行うよりも，いくつかの条件事象に分割してそれらを総合化することも有力であろう．たとえば石油の市場価格を予想する場合，中東の政治情勢，石油開発の可能性，代替

エネルギー問題，関係諸国の経済状況など多角的な視点がとられよう．

いま，対象としている事象を A，分割された個別の条件事象を $B_1, B_2, \cdots, B_n$ とし，これらを基礎に条件つき確率 $p(A|B_i)$ を推計し，$p(B_i)$ をある種のウェイトと考えて，通常の条件つき確率の公式

$$p(A)=\sum p(A|B_i)p(B_i)$$

を採用して，$p(A)$ に対する確率付与を行うことが考えられる．

最近の研究成果によれば，構成要素としてのすべての確率 $p(A), p(A|B_i), p(B_i)$ がそれぞれ真の値と誤差項の和から構成されるといった加法的構造をもっている場合，次のことが結論されている．①条件つき事象の選択如何によっては分割方式が直接方式よりも誤差項の分散を小ならしめるという意味で効率的であることが示された．②分割事象の個数の増大については，ある一点までは誤差項の分散の縮小化に貢献するが，それを越した場合には効率は上昇しない．③分割の背景となった周辺確率の事象については，なるべく等可能が望ましい．④条件つき確率の値については，極端な値をもつようなものがないことが望ましい．⑤条件つき確率の事象や周辺確率の事象の誤差項の分散はできるかぎり小さい方が望ましい．⑥統計家は各要素の評価にそれぞれの独立性を保持するよう心がけるべきである――などである．しかしながら，こうした条件をすべて満足するような分割方式は殆ど存在しないといっても過言ではない．とりわけ意思決定問題における確率付与の局面においては，こうした分割方式を採用するために特別に専門家に依頼したりするならば，それだけ費用がかかることも考慮しなくてはならない．また先述の主観確率の付与におけるある意味での首尾一貫性が，こうした分割方式の採用によって維持されるならばともかく，そうした保証は与えられていないといってよい．言い換えればそれだけ現実面の制約を受けての立論とはなっているが，応用面の展開の中でいくばくかの示唆を与えていることも事実であろう．

3.3 経験的データの利用

観測の結果得られている相対頻度としての度数分布から，できるだけ整合性のある確率を読み取るための留意事項について述べておこう．なお，確率

変数は $\tilde{z}$ で表す．

A. 棒グラフあるいはヒストグラムの場合

表 3

需要量 $\tilde{z}$	度数	相対頻度
2	1	.063
3	3	.187
4	2	.125
5	4	.250
6	3	.187
7	2	.125
8	0	.000
9	1	.063
10+	0	.000
	16	1.000

表 4

需要量 z	目測で得られた曲線による頻度	確率
0	0	0
1	.01	.009
2	.07	.060
3	.16	.138
4	.21	.181
5	.22	.190
6	.21	.181
7	.15	.129
8	.08	.069
9	.04	.034
10	.01	.009
11+	0	0
	1.16	1.000

たとえば表3のような $\tilde{z}$ として，ある商品の需要の分布の記録が与えられているとして，これを根拠に明日の $\tilde{z}$ の分布を予測することを考えよう．実際の度数分布は3単位と5単位，7単位と9単位の間に2ヵ所に「くぼみ」が生じているが，こうした場合，何回にもわたって明日の予測を行うということを背景に，たとえば7単位と9単位にそれぞれ0.125，0.063の相対頻度が与えられていて，8単位はゼロであるからといってゼロという予測はしないであろう．通常は7単位と9単位の相対頻度の中間の値を想定する．つまり一般的には，特定された日についても，特別の事情のないかぎりは，与えられた情報は標本とみなして平滑化するのが自然であろう．

なお曲線の当てはめによる平滑化の場合には，与えられた相対頻度に対して目測で滑らかな曲線をトレースしておき，それにもとづいて表4のように縦座標を読み，各要素の和が1になるように調整した後に，それにしたがって平滑化された曲線（図3(c)）を求めるのが常套であろう．

区間表示あるいはグループ化されている場合，たとえば表5のような場合には区間は5単位分割であるので，単位幅当りの頻度を右欄のように求める．

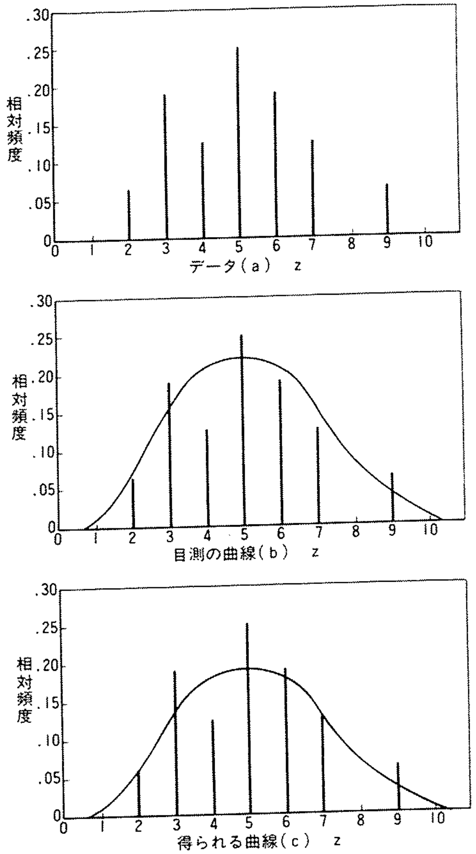


図3

平滑化の曲線については図4のように与えられた区間について曲線下の面積が区間のヒストグラムの面積に等しいように引くことが肝要である．こうした曲線によって特定の値 $\tilde{z}=z$ に関する確率は，たとえば

$$p(\tilde{z}=20)=1\times p'(20)=1\times 0.026=0.026$$
$$p(\tilde{z}=20)=1\times p'(21)=1\times 0.022=0.022$$

のように求めればよい．

表 5

需要量 z	相対頻度	単位区間当りの頻度
0-4	.051	.0102
5-9	.256	.0512
10-14	.325	.0650
15-19	.222	.0444
20-24	.094	.0188
25-29	.043	.0086
30-34	.009	.0018
	1.000	

表 6

需要量 z	度数	相対頻度
9	1	.1
11	1	.1
15	1	.1
16	1	.1
17	1	.1
20	1	.1
22	1	.1
24	1	.1
29	1	.1
35	1	.1
	10	1.0

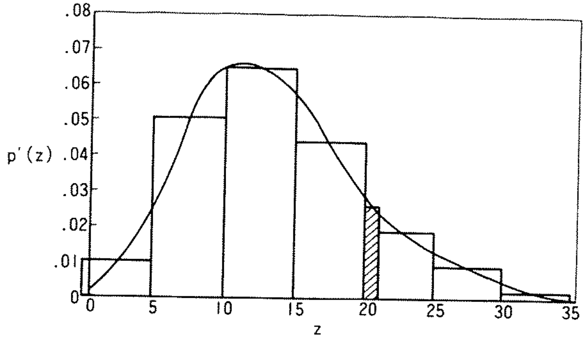


図4

B. 極端に広がりのあるデータの平滑化

たとえば表6で示されるような拡散されたデータの場合には，累積度数分布の利用が奨められる．もちろん，先述の長期的試行という想定の下で平滑化を実行することになる．したがって，表6からは $z=9$ 以下の累積度数はゼロであり，$z=3$ 以上は1となっているが，平滑化はそうした例外についても行うことになる．なお累積分布の利用ということで，分位数のとり方にも注意する必要がある．一般に分位数のとり方は，n 個の観測値がある場合，第 k 番目の観測値に対応する分位数は $k/(n+1)\times 100\%$ ということになる．表6の観測値について対応する分位数をプロットして求めた累積度数分布が

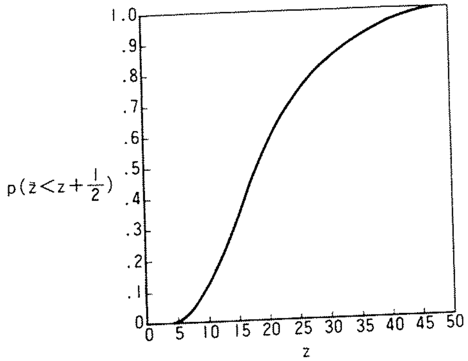


図5

図5である．さて，このようにして得た累積度数分布から確率を求めるには，

$$p(\tilde{z} < z)\cdots\left(z-\frac{1}{2}\right) \text{の点のタテ軸の値}$$

$$p(\tilde{z} \leqq z)\cdots\left(z+\frac{1}{2}\right) \text{の点のタテ軸の値}$$

ということになる．そして個々の z の確率は

$$p(\tilde{z} = z) = p(\tilde{z} < z+1) - p(\tilde{z} < z)$$

とすればよい．たとえば $p(\tilde{z}=20)=p(\tilde{z}<21)-p(\tilde{z}<20)=0.04$ と読める．この際，しばしば差の値が微小のため読み取り誤差が生じるきらいがあるが，

表 7

需要量 z	各区間の左端の累積度数	区間の確率	区間当りの確率
0-4	0	0	0
5-9	0	.108	.0216
10-14	.108	.212	.0424
15-19	.320	.234	.0468
20-24	.554	.170	.0340
25-29	.724	.104	.0208
30-34	.828	.074	.0148
35-39	.902	.054	.0108
40-44	.956	.034	.0068
45-49	.990	.010	.0020
50-54	1.000	0	0
		1.000	

そうした場合には次のステップへ進む．

(1)　平滑化された累積度数分布からグループ化された区間の確率の値を読む．

(2)　その区間の確率から各々の値の確率を計算する．

表7は図5の累積度数分布からこのようにして得た結果を記載してある．なお正規確率紙の応用も含めて，累積度数分布と度数分布の形状の関係も知っておくことが望ましい．正規確率紙において分位数が直線上に並ぶ場合は一山かつ対称で，裾野の部分はしだいにゼロになるような分布となっている．

参　考　文　献

DeGroot, M. H., 1970 *Optimal Statistical Decisions*, McGraw-Hill.

Fishburn, P. C., 1970 *Utility Theory for Decision Making*, John Wiley.

James, I. R. and M. W. Knuiman, 1987 "An Application of Bayes Methodology to the Analysis of Dairy Records from a Water Use Study," *Journal of the American Statistical Association*, Vol. 82, No. 399.

Kyburg, E. Jr. and H. E. Smokler (ed.), 1964 *Studies in Subjective Probability*, John Wiley.

Lindley, D. V., 1971 *Making Decisions*, Wiley-Interscience.

von Neumann, J. and O. Morgenstern, 1947 *Theory of Games and Economic Behavior*, Second Edition, Princeton University Press.

Raiffa, H., 1968 *Decision Analysis*, Addison-Wesley.（宮沢光一・平館道子訳『決定分析入門』，東洋経済新報社，1972）

Ravinda, H. V., Kleinnaunts, D. N. and J. S. Dyer, 1988 "The Reliability of Subjective Probabilities obtained through Decomposition, "*Management Science*, Vol. 34, No. 2.

Savage, L. J., 1954 *The Foundations of Statistics*, John Wiley.

Schlaifer, R., 1959 *Probability and Statistics for Business Decisions, An Introduction to Managerial Economics under Uncertainty*, McGraw-Hill.

宮沢光一，1971『情報・決定理論序説』，岩波書店．

第2章
主観確率の尺度調整*

松 原 望

1. ベイズの定理の有効性

確率論におけるベイズの定理は，その分野の人なら周知のものであるが，この確率論の理論的数理がそのまま意思決定問題に適用可能か，つまり現実の意思決定の理論的記述は可能かという問題がある．その前に，初学者のため，およびベイズの定理が確率論の公理からきわめて直接に導出されることを再確認するために，その証明をする．まず，確率の公理を示す．

標本空間 Ω の任意の可測事象 E に対し実数 $P(E)$ を対応させる関数で，3つの公理

(A1) 任意の可測事象 E に対し，$P(E) \geqq 0$,

(A2) $P(\Omega)=1$,

(A3) 可測事象 $E_1, E_2, \cdots\cdots$ が互いに排反（$i \neq j$ ならば $E_i \cap E_j=\phi$）ならば，

$$P(\cup_{i=1}^{\infty} E_i)=\sum_{i=1}^{\infty} P(E_i) \tag{1.1}$$

を満足する $P(\cdot)$ を確率という（以下，「可測」を略す）．上記の（A3）は

$$\begin{aligned} &(\mathrm{i})\quad E=\cup_{i=1}^{\infty} E_i, \\ &(\mathrm{ii})\quad E_i \cap E_j=\phi \qquad (i \neq j) \end{aligned} \tag{1.2}$$

ならば（事象列 $\{E_1, E_2, \cdots\cdots\}$ は事象 E の「分割」であるという），

---

* 本論文を作成するにあたり，鈴木雪夫，国友直人，豊田敬，美添泰人，竹内啓の諸氏の御教示が大変役立ったことをここに記して，感謝の意を表したい．

$$P(E) = \sum_{i=1}^{\infty} P(E_i) \tag{1.3}$$

であること，といってよい．

条件付確率の定義は，単に[1)]

$$P(A|B) = P(A \cap B)/P(B) \qquad (P(B) > 0) \tag{1.4}$$

というものであり，ベイズの定理はこの特殊のものである．

いま，Ω の分割 $\{E_1, E_2, \cdots\cdots\}$，および任意の事象 F（ただし，$P(F)>0$）が与えられたとき，上記で $A=E_i, B=F$ とおけば

$$P(E_i|F) = P(E_i \cap F)/P(F) \tag{1.5}$$

であるが，分子については今度は $A=F, B=E_i$ とおき定義を再び用いて

$$P(E_i \cap F) = P(E_i) \cdot P(F|E_i) \tag{1.6}$$

分母については，$\{E_1 \cap F, E_2 \cap F, \cdots\cdots\}$ が F の分割であることに注意し，かつ（1.6）の表現を再び用いれば

$$\begin{aligned} P(F) &= \sum_{j=1}^{\infty} P(E_j \cap F) \\ &= \sum_{j=1}^{\infty} P(E_j) \cdot P(F|E_j) \end{aligned} \tag{1.7}$$

となる．すなわち，ベイズの定理

$$P(E_i|F) = \frac{P(E_i) \cdot P(F|E_i)}{\sum_{j=1}^{\infty} P(E_j) \cdot P(F|E_j)} \tag{1.8}$$

を得る．特別の場合として，$E_{n+1}=E_{n+2}=\cdots=\phi$ なら，この和は有限和となる．（一般に，$P(E_j)=0$ なる j は総和記号から除外するものとする．）

このベイズの定理（Bayes' theorem）をベイズの規則（Bayes' rule）という人も多い．これは，（1.8）式はある条件付確率の計算の「規則」にすぎず，「定理」というのにはあまりに簡単すぎるとする立場，あるいは過度の思弁を排する立場である（Feller[1976]）．筆者は，$E_1, E_2, \cdots\cdots$ を「原因（系）」，F を「結果」と，意味づけて考えるベイズ統計学（Byesian statistics）の基本原理を支持するので，前者の方を維持する．すなわち，本式は結果が判明しているときにそれをもたらした原因の確率を与える式として，定理と呼ぶにふさわしい深い意味をもつ有用な式である．

確率論の演繹体系，あるいはベイズ統計学の純粋な演繹体系においては，これは原理的な基礎を与えるものであるが，現実の意思決定の問題では，これからはずれる場合が多々見られるのであって，この'はずれ'がどれほどか，

また，それをどのように取り扱うかが大きな研究課題となる．なお，注意すべきことは，はずれが生じるからといって，ベイズの定理が無効無用になるのではなく，むしろ，この定理こそそのはずれをはかる標準的原点を与えるものだということは，強調しておきたい．

たとえば，Lindley[1971]は，いわゆる3囚人問題（3-prisoner problem）を出すことにより，人間の行う帰納過程が想像以上に多様であることを，具体的に呈示した．

> 3人の囚人が幽閉されているとしよう．3人の囚人の名前は，アラン（Alan），バーナード（Bernard），チャールズ（Charles）とする．アランは，翌日3人のうち2人が処刑され，1人が釈放されることを知っているが，3人のうち誰が釈放されるかについては全くわからない．このような状況において，アランが看守に対し，「3人のうち2人が処刑されるのは確実である．バーナードとチャールズのうち少なくとも1人は処刑されるのは確実である．バーナードとチャールズのうち処刑される者の名前を1人だけ教えてくれても，アランの釈放については全く情報を与えないはずだから，その名前を教えてほしい．」と言ったところ，看守は，アランの言い分を納得し，「バーナードは処刑される．」と答えた．アランはこれを聞いて，釈放される可能性があるのは，自分の他はチャーズルのみになったので，自分が釈放される可能性が増えたと喜んだとする．直観的には正しいように見えるが，この確率評価は根拠があるだろうか．

釈放されるべき人を，頭文字で，A, B, C，とし，また s を「バーナードは処刑される」という言明とすると，状況から，$P(s|A)=1/2, P(s|B)=0, P(s|C)=1$ と仮定される．事前確率は，$P(A)=P(B)=P(C)=1/3$ とする．ベイズの定理からは

$$\begin{aligned} P(A|s) &= \frac{P(A)P(s|A)}{P(A)P(s|A)+P(B)P(s|B)+P(C)P(s|C)} \\ &= \frac{(1/3)\cdot(1/2)}{(1/3)\cdot(1/2)+(1/3)\cdot 0+(1/3)\cdot 1} \\ &= 1/3 \end{aligned} \tag{1.9}$$

ゆえに，$P(A)=(A|s)$ となり，上記の確率評価は根拠がない．これは，人間

の直観と数理のいずれが正しいか，という問のようにもみえるが，数学的確率論を1つの標準として，人間の判断の「合理性」の内容の探求という捉え方の方が，生産的，発展的である．

また，「アランの釈放については全く情報を与えない」ことを前提としている（しかもアラン自身の論理）以上，確率が変化するはずがなく，依然として1/3なのであって，問題自体はありえないものだとする立場がある．これには，あらかじめ，BがAについてもたらす「情報」を，確率の変化高$P(A|B)/P(A)$と定義することを承認しておかねばならない．しかし，議論の錯綜の中で，この定義だけを特に採用することは，「論点先取の誤り」であろう．しかも，この定義は（1.4）から対称性

$$\frac{P(A|B)}{P(A)}=\frac{P(B|A)}{P(B)} \tag{1.10}$$

を結果する．'BがAについてもたらす情報は，AがBについてもたらす情報に等しい'．しかし，これは現実には，常には受け入れられない．教授の機嫌から秘書の機嫌は（蓋然的に）予測できるが，秘書の機嫌から教授の機嫌は少なくとも同程度には予測し難いのである．形而上学的議論に入らず，数理的手法で解釈できる範囲に限ろう．

2. 確率の保守傾向

Edwards[1968]は，次のような例を出している．赤玉，白玉が，それぞれ7：3，3：7で入っている2つのつぼA, Bのいずれか一方（どちらかは不知）から，玉を連続して12個復元抽出したとき，8個の赤，4個の白を得た．Aである確率w'はいくらかを問うたのに対して，かなり一貫して0.75程度という評価の回答が多い．著者も独自に同じ実験を試みたが，大略同旨の結果を得た．

しかし，ベイズの定理からは，事前確率を$w=1/2$として，共通の2項係数を略して

$$w'=\frac{(0.7)^8(0.3)^4}{(0.7)^8(0.3)^4+(0.3)^8(0.7)^4}=\frac{7^4}{7^3+3^4}=0.967 \tag{2.1}$$

となり，0.75という実際値はこれに比して，かなり保守的である．

類似の例は他にもあるが，Edwardsはこれらを，評価の保守傾向（conservatism）と呼んだ．この例では，核にあるデータ発生機の7：3，3：7という構成比を現実にどう認識したか，また，12個のデータの情報集約（aggregation）の不足が原因，などの説明が試みられる．この点は，Green他[1965]の比較的簡単な実験によっても，系統的に裏書きされており，他方，Slovic and Lichtenstein[1971]も，解説の中でかなり詳しく（第10章），この保守傾向を今後研究すべき1つの重要な論点と指摘している．

人間の行う確率情報処理（probabilistic information processing）の過程において，評価確率p'が一貫して，確率論的確率pより保守的，つまり，1の近傍では$p'<p$，0の近傍では$p'>p$であるとすれば，応用上，相当に大きい問題である．一貫してそうなるわけではなく，過程が多段階に分かれる場合は，データと仮説の間に交互作用が生じて，保守的とは反対に，超過的（excessive）となることもある（Peterson[1973]；Youssef and Peterson[1973]）．このように，現実の現象はベイズの定理で完全に記述されるほど簡単ではない[2)]．

これらは，ベイズの定理による確率変換過程を抑制（修正）することにより数理的に説明しうる．一般に$f(x|\theta_i)$, $i=1,2$を2つの分布とし，$w, 1-w$を事前確率とすれば，標本$(x_1,\cdots,x_n)$にもとづく，θ_1の事後確率w'は，容易にわかるように，

$$L_n=\textstyle\sum_j\log\frac{f(x_j|\theta_1)}{f(x_j|\theta_2)} \quad \text{（対数尤度比）} \tag{2.2}$$

として，ロジスティック曲線

$$w'(L_n)=\frac{1}{1+\gamma e^{-L_n}} \quad \left(\gamma=\frac{1-w}{w}\right) \tag{2.3}$$

で表現される．変換の抑制は因数$b(b>0$，ただし$b<1)$を入れて

$$w'(L_n)=\frac{1}{1+\gamma e^{-bL_n}} \tag{2.4}$$

のようにすれば説明され，それはL_nの代りに

$$\tilde{L}_n=\textstyle\sum_j\log\frac{(f(x_j|\theta_1))^b}{(f(x_j|\theta_2))^b} \tag{2.5}$$

を考えたことに相当する．挿入したパラメーターが1個であることには留意してほしい．すなわち，ベイズの定理に1次元の，必要にして最小限の，摂動（perturbation）を入れたのである．

これはベイズの定理の数学的「修正」であり，それ自体として格別に新しいものが生じたわけではないが，ロジスティック曲線に書かれたことは，理論的には興味深いものを示唆する．本論文はこれをヒントとして，通信モデル（communication model）を発展させる．もっとも，修正といってもこの論文のかぎりのものであることには，留意しておかねばならない．

ちなみに，意思決定理論家（decision scientist）は，人間の主観確率評価は彼の効用 $u(\cdot)$，損失 $l(\cdot)$ に密接不離に結びついていると考えるのが通常である．もし $l(\cdot)$ が w' の関数で（この仮定は統計的決定理論の通常の枠組ではない），0あるいは1に近い確率値に対する大きな損失（制裁）の項，たとえば，

$$l(w') \propto \frac{1}{w'(1-w')} \tag{2.6}$$

を含んでいれば，ある区間 $J=(\varepsilon, 1-\varepsilon)$（ただし $\varepsilon>0$）以外の値は評価に用いる値として実質上「禁止」され（出現率が極めて小），(2.4) でなく

$$w'(L_n) = \frac{1-2\varepsilon}{1+\gamma e^{-L_n}}+\varepsilon \tag{2.7}$$

という，上下限界を付する修正も可能なのである．たとえば，人間は0.99以上（あるいは0.01以下）の高い（低い）確率を主観的評価として表明することに非常に大きな抵抗を示す，と考えられるならば，$\varepsilon=0.01$ とすればよい（これも一次元の摂動である）．

ベイズの定理に対する「違反」は，他にも無数にあるといってよく，違反の程度もさまざまである．これを非難すべき違反と考えずにむしろ多様性と考えることが大切であるが，しかし，多様性をあまりに強調しすぎれば，恣意的という否定的評価を招きかねない．多様性を認めた上で，なんらかの信頼しうる法則性（確率論の法則，客観的確率等）との連絡は維持しておくことが大切である．これが本論文の主眼である．

3. 主観確率評価と尺度調整

まず，主観確率の値，その分布を具体的な形に書き出す手続を，確率の同定，引出し（elicitation），あるいは，暗号との類似から，符号化（encoding）などという（Spetzler 他[1975]）．いずれにしても，具体的な応用には不可欠の手続である．なお，Winkler[1967]は，内心の状態を「判断」（judgment）といい，表明された確率を「評価」（assessment）として，概念上の区別をし，両者の関係をモデル化している．

そこで，ベイズの定理（「修正」されたもの）を1つの規範として，この一見つかみどころのない‘主観’確率を標準的なものさしにあわせる（尺度調整 calibration）必要が出てくる．ただし，何が「標準的」かは場合によりさまざまである．そのためには，ベイズの定理を数学的恒等式と見ずに，1つの統計的模型（はっきりいえば，ある母回帰式）として考え，それに対する統計的推定，検定を考えてゆけばよい．なぜなら，よく知られているように，回帰モデルは一般的な尺度調整の問題にとって有力な手法だからである．

まず第一に，ベイズの定理に説明変数，被説明変数の概念を導入せねばならない．本来のベイズの定理は数学的恒等式であって，この区別を有しない．われわれは1つのモデルを考える必要がある．それは通信モデルである．

信号受容者（signal receiver）は「信号」（情報を含有するなんらかの事象）を受け，その「信憑性」（credibility）を評価する．ここで信憑性とは，ある事項A（製品の品質，相手の戦略など）に関する，いわゆる「確信の度合」（degree of belief）であるが，信号の語に対しては信憑性の語の方がより適合的であり，かつ広い範囲を包みうる．（もっとも，概念的にあいまいになることは避けられない．）そこで一般的に，

L＝A に対する信号の有する情報の量

p＝A に関する信号の信憑性の主観確率評価

とする．ここで情報の量は何らかの意味で量化されることを前提としている．「A に関する」とは，くわしくいえば，A に2つの状態 θ_1, θ_2 があること，p は θ_1 が真であることの確率であること，さらに重要なこととして，p は L

の単調増加関数であることなどを意味する．最後の項で，単調であることが本質的であり，増加であることの仮定は本質的ではない．

p は有界であるから，L は p について「擬 S 形」(sigmoid) と仮定することはそれほど一般性を失うものではない．むしろ，生物統計学 (biostatistics) で典型として仮定される，いわゆる量・反応関係[3] (dose-response relationship) としてこの問題を規定 (L を dose，p を response) するなら，これはむしろ標準的でさえある．本論文では，この関数形がロジスティック型

$$p(L) = \frac{1}{1+e^{-(bL+a)}} \tag{3.1}$$

となることを導出し，a, b の値を実際に推定できる場合があることを論じる．すなわち上式が，ベイズの定理の単なる数学的書き替えでなく，L が説明変数，p が被説明変数であるような実証モデルになりうることを示し，それを利用して p を L に対して (L によって) 尺度調整を行うことを考える．

それに先立ち，a, b の意味を述べる．まず

$$p(0) = 1/(1+e^{-a}) \tag{3.2}$$

であるから，$a>0, =0, <0$ に応じて，$p(0)>0.5, 0.5, <0.5$ である．そこで，a を「基底レスポンス定数」(fundamental response constant) と呼ぶ．中立的信号 (量) $L=0$ に対する本来的性向を表すのである．すなわち，$a>0$ ならば，情報がない (信号が中立的である) ときでさえも 50−50 よりも高い信憑性が引き起こされるのである．また，明らかに，a はロジスティック曲線の位置 (location) を指定するパラメーターである．さらに

$$p(-a/b) = 0.5 \tag{3.3}$$

であるが，その点における微係数は

$$p'(-a/b) = b/4 \tag{3.4}$$

である．すなわち，b は，$p=0.5$ の点で測られた，p の L に対する感度を表している．b を「レスポンス因数」(response factor) と呼ぶ．$b=1$ ならば，ベイズの定理の形になり，$b>1$ ならば，ベイズの定理よりも，大きな急な p の変化を引き起こす．この b はロジスティック曲線の尺度 (scale) を指定している．事実，$p(\cdot)$ は

$$p'(u) = b\cdot p(u)(1-p(u)) \tag{3.5}$$

なる微分方程式 (成長曲線の微分方程式) を満足する．なお統計的にはロジット変換をすれば，$p=p(L)$ は L に対して線型になり回帰モデルに帰する．

$$\log\frac{p}{1-p} = bL+a \tag{3.6}$$

そこで，まず信号発信者 (signal transmitter) の状態 θ_1, θ_2 に応じて，*i.i.d.* である，$x_1, x_2, \cdots, x_n$ が信号受容者に送られたとすると，p_0 を事前確率，p を事後確率として，

$$\frac{p}{1-p} = \frac{f(x_1|\theta_1)}{f(x_1|\theta_2)}\cdots\frac{f(x_n|\theta_1)}{f(x_n|\theta_2)}\cdot\frac{p_0}{1-p_0} \tag{3.7}$$

となる．ここで，

$$\Phi = \textstyle\sum_j \log\frac{f(x_j|\theta_1)}{f(x_j|\theta_2)} \tag{3.8}$$

とおくと，

$$\log\frac{p_R}{1-p_R} = \Phi+\log\frac{p_{0R}}{1-p_{0R}} \tag{3.9}$$

$$\log\frac{p_T}{1-p_T} = \Phi+\log\frac{p_{0T}}{1-p_{0T}} \tag{3.10}$$

となる．ここに，p_R, p_T は，それぞれ信号受容者，信号発信者における，事後確率評価をあらわす．信号発信者が，自らの発信する信号について不確実性を有するということは論理的には奇異に感ずるかもしれないが，たとえば，マーケティング広告においては，むしろこの仮定の方が自然である．信号発信者 (製造者) は自己の製品の品質や優良性について 100% の完全認識を有しているわけではない．国際政治における意思決定においても，一枚岩的 (unitary) な意思決定者というものは，現実的な仮定ではない．すなわち，相手国の行政府が支配している θ について，相手国の立法府 (議会) さえ完全な認識を有するものでなく，θ にその分布が依存する x を観測しているとする方がよい．

式 (3.9)，(3.10) を辺々引けば，

$$\log\frac{p_R}{1-p_R} = \log\frac{p_T}{1-p_T}+a, \tag{3.11}$$

ただし，

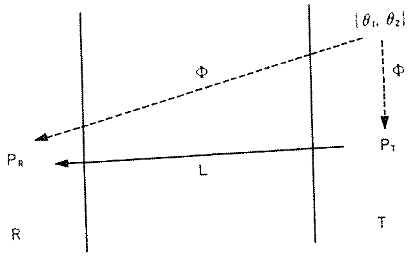


図1 信号発信者 T と信号受容者 R の関係

$$a=\log\frac{p_{0R}}{1-p_{0R}}-\log\frac{p_{0T}}{1-p_{0T}} \tag{3.12}$$

となる．ここで，(3.11) を拡張して，

$$\log\frac{p_R}{1-p_R}=b\cdot\log\frac{p_T}{1-p_T}+a \tag{3.13}$$

とおけば，両変数を関係づける母回帰式を得る．なお，

$$L=\log\frac{p_T}{1-p_T} \tag{3.14}$$

とおけば，(3.6) を得る．

現実の応用で興味ある問題は，a, b に関する統計的検定である．すなわち，T 期に渉る主観確率評価の時系列データ

$$(p_{R_i}, p_{T_i}),\quad i=1, 2, \cdots, T$$

を得れば，b, a の推定ができるが，検定としては，レスポンス因数 b については，

$H_{b1}: b=1$ （'ベイズの定理' 型）

$H_{b2}: b>1$ （積極反応型）

$H_{b3}: b<1$ （抑制反応型）

など，基底レスポンス定数 a については

$H_{a1}: a=0$ （中立型，ないしは同時型）

$H_{a2}: a>0$ （先行型）

$H_{a3}: a<0$ （遅行型）

などの行動科学的仮説が考えられる．さらに，$H_{bi}\cap H_{aj}$ の形の仮説も考えられる．なお

$$H_{b0}: b=0$$

の検定（t 検定）はいうに及ばない．万一，有意でなければ例外的事態となろう．

もし，L が説明変数なら，それに対する p の値が回帰式より定まるから，L による信号受容者の p の尺度調整が行われると考えてよい．主観確率はたしかに '不確かなもの' という見方を免れないのであるが，信号発信者からの情報にあわせて，相対的客観性を獲得するのである．しかし，1つの限界は，標本が主観確率評価値であることの現実的問題点である．いかにして標本をとるのかということであって，その値を 0.1 きざみで固定してもらっても，甲のいう 0.3 と乙の 0.3 が同じ 0.3 であるという原理的保証はない．しかし，大まかにいえば，人間の確率感覚がまったく任意であるというわけでもない．むしろ，ポーカーの手の順序づけ[4)]のように繊細な確率差を見分ける能力もある．この原理的問題について過度に懐疑的になるべきではない．

なお (3.13) 式は次のようにすれば，ベイズの定理から自然に導かれる*．

$$p(\theta_1)\equiv p_{0R},\ p(\theta_2)\equiv 1-p_{0R} \tag{3.15}$$

とし，さらに信号発信者は θ_1 に関する確率値 p_T を信号受容者に知らせるものとする．いま p_T の尤度 g は

$$\begin{aligned} p_T|\theta_1 &\sim \mathrm{Be}(\alpha, \beta)\\ p_T|\theta_2 &\sim \mathrm{Be}(\beta, \alpha)\end{aligned} \tag{3.16}$$

なるものとすると，ベイズの定理より

$$\log\frac{p(\theta_1|p_T)}{p(\theta_2|p_T)}=\log\frac{g(p_T|\theta_1)}{g(p_T|\theta_2)}+\log\frac{p(\theta_1)}{p(\theta_2)} \tag{3.17}$$

これから

$$\log\frac{p_R}{1-p_R}=\log\left(\frac{p_T}{1-p_T}\right)^b+\log\frac{p_{0R}}{1-p_{0R}} \tag{3.18}$$

$$(b=\alpha-\beta)$$

もっとも，この場合は $b<0$ もありうる．

* この示唆は豊田敬氏に負う．'p_T を相手に知らせる' という点は筆者の扱っている場合には該当しないが，b がみごとに導入される点では洗練されている．

4. 数　値　例

応用例としては，典型的にはマーケティングへのものが考えられる．また，2 国間の国際政治学的意思決定における，信号の効果の測定もある．筆者は，後者について実証研究を行ったので，数値の結果のみ述べる．

T 国を発信者，R 国を受容者とする．T 国を行政府と立法府に分かち，行政府は T，R 両国に重大な利害を有する事項について，θ_1, θ_2 の 2 つの状態（政策）のいずれかをとりうる．その立法府もこれについては重大な利害をもち，自国の行政府が θ_1, θ_2 のいずれを採用するかを関心をもって観察している．R 国を単一の受容者と考えるのが最も単純だが，実際には R 国にも，各々の異なった立場を有する複数の受容者を考えた．

ここにおける例では，共通の発信者 T に対して 3 名の受容者 R_7, R_{24}, R_6 が（実は，全体で 28 名であったが，一部を掲げる．なお表 3 参照），どう主観確率の評価をしたかを測定し（表 1），それを尺度調整する曲線を推定した．結果は表 2，図 2 のごとくである（全体については表 3，表 4 参照）．

ここで，'R_7' などは研究上で用いた匿名のための記号である．R_7 では曲線の立ち上がり，R_{24} では曲線の水平位置に注目してほしい（逆の比較も同旨）．R_6 は $b<1$ となっていることに注意するためで，形自体はごくありふれたものである．経験則であるが，b が小さいほど R^2 も低下する．これは，Winkler の「保守傾向」の語を用いるならば，評価がベイズの定理よりも保守的になるほど，誤差が大きく保守傾向自体が不安定な現象になることを表している（図 4）．

表 1

時期 i	1	2	3	4	5	6
発信者の L	−1.32	−1.28	−0.97	−0.89	−0.04	−1.65
受容者の p						
R_7	0.01	0.01	0.2	0.3	0.9	0.01
R_{24}	0.01	0.01	0.01	0.1	0.2	0.01
R_6	0.1	0.1	0.2	0.1	0.5	0.01

表 2　a, b の推定値（尺度調整曲線）および R^2

受容者 R	b	a	R^2
R_7	2.054	1.030	0.924
R_{24}	1.013	−0.589	0.810
R_6	0.995	0.074	0.757

表 3　主観確率評価

R　　i	1	2	3	4	5	6
1.	0.01	0.01	0.01	0.01	0.01	0.01
2.	0.1	0.1	0.4	0.4	0.9	0.01
3.	0.01	0.01	0.01	0.25	0.25	0.01
4.	0.1	0.01	0.01	0.01	0.55	0.01
5.	0.01	0.1	0.3	0.3	0.7	0.01
6.	0.1	0.1	0.2	0.1	0.5	0.01
7.	0.01	0.01	0.2	0.3	0.9	0.01
8.	0.25	0.3	0.4	0.5	0.7	0.01
9.	0.5	0.5	0.5	0.5	0.9	0.01
10.	0.1	0.1	0.4	0.4	0.8	0.01
11.	0.2	0.3	0.4	0.4	0.4	0.01
12.	0.2	0.2	0.3	0.2	0.6	0.2
13.	0.1	0.1	0.1	0.1	0.4	0.1
14.	0.01	0.01	0.01	0.01	0.01	0.01
15.	0.01	0.01	0.01	0.5	0.6	0.01
16.	0.05	0.1	0.07	0.1	0.15	0.01
17.	0.1	0.1	0.3	0.3	0.5	0.3
18.	0.01	0.01	0.3	0.3	0.5	0.01
19.	0.01	0.01	0.01	0.01	0.01	0.01
20.	0.01	0.01	0.01	0.01	0.01	0.01
21.	0.1	0.1	0.1	0.3	0.3	0.1
22.	0.01	0.01	0.01	0.01	0.2	0.01
23.	0.3	0.3	0.3	0.3	0.4	0.01
24.	0.01	0.01	0.01	0.1	0.2	0.01
25.	0.01	0.01	0.01	0.01	0.01	0.01
26.	0.1	0.1	0.5	0.8	0.5	0.2
27.	0.1	0.1	0.2	0.2	0.4	0.2
28.	0.01	0.01	0.1	0.15	0.3	0.01

（注）　log 0 を避けるために，0 を 0.01 に便宜上置き換えてある．これがデータ分析に若干の影響を与えているかもしれぬことはいうまでもない．

表 4 は，表 2 の範囲を拡大して表 3 の主観確率評価の全データに対して分析した結果であって，すべての信号受信者の信憑性の分析結果である．その一部を図示したものが図 3 である．

表 4　信憑性の分析

R	b	a	t	R	R^2	SNR
1.	—	—	—	—	—	—
2.	1.705	1.135	8.226	0.972	0.944	16.915
3.	1.173	−0.331	3.209	0.849	0.720	2.575
4.	1.021	−0.466	1.780	0.665	0.442	0.792
5.	1.533	0.629	4.370	0.909	0.827	4.774
6.	0.995	0.074	3.532	0.870	0.757	3.119
7.	2.054	1.030	6.997	0.961	0.924	12.241
8.	1.194	0.738	2.959	0.829	0.686	2.190
9.	1.383	1.193	2.747	0.808	0.654	1.887
10.	1.497	0.870	6.049	0.949	0.901	9.147
11.	0.865	0.273	1.903	0.689	0.475	0.905
12.	0.463	0.024	3.089	0.839	0.705	7.385
13.	0.460	−0.370	2.889	0.822	0.676	2.086
14.	—	—	—	—	—	—
15.	1.645	0.325	3.473	0.867	0.751	3.016
16.	0.610	−0.574	2.570	0.789	0.623	1.651
17.	0.436	−0.071	1.733	0.655	0.429	0.751
18.	1.493	0.354	3.613	0.875	0.765	3.264
19.	—	—	—	—	—	—
20.	—	—	—	—	—	—
21.	0.453	0.311	3.209	0.849	0.720	2.575
22.	0.824	−0.949	2.889	0.822	0.676	2.086
23.	0.781	0.165	1.700	0.648	0.419	0.723
24.	1.013	−0.589	4.128	0.900	0.810	4.260
25.	—	—	—	—	—	—
26.	0.712	0.385	1.704	0.649	0.421	0.726
27.	0.383	−0.270	2.293	0.754	0.568	1.314
28.	1.199	−0.159	4.731	0.921	0.848	5.596

（注）1）SNR $=R^2/(1-R^2)$，すなわち，信号雑音化 Signal-to-Noise Ratio. ——については，分析を行わなかった．
2）本分析については，研究協力者である筑波大学の蒲島郁夫氏に謝意を表する．

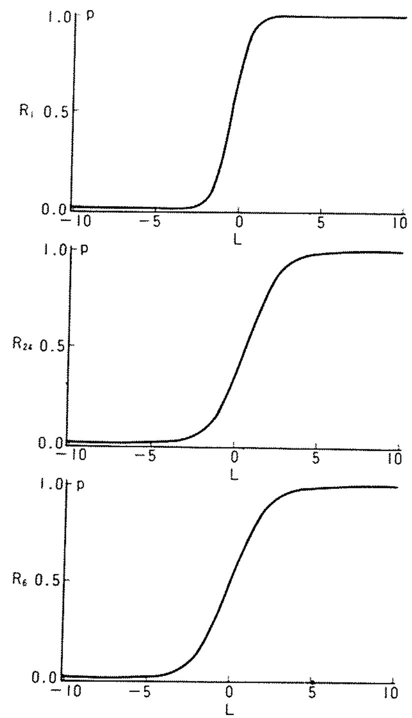


図 2　ロジスティック曲線による尺度調整

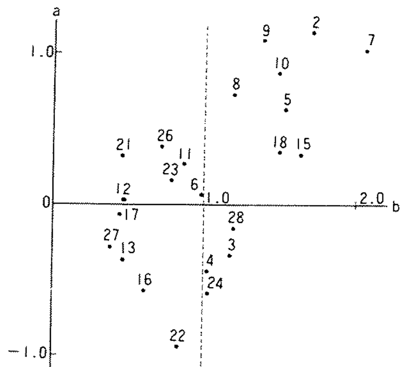


図 3　尺度調整曲線の a（基底レスポンス定数）と b（レスポンス因数）

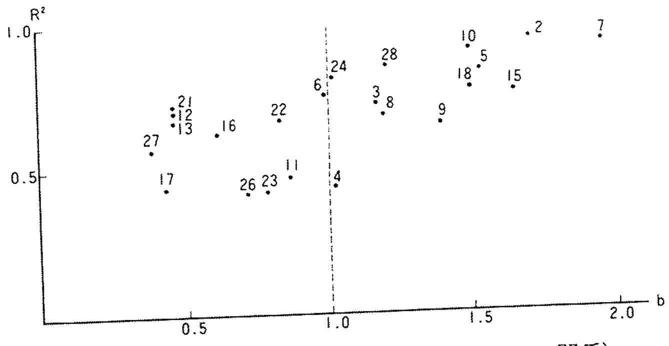


図4 確率評価の保守傾向と安定性（b と R^2 の関係）

5. 客観確率への尺度調整

ふつう，尺度調整されている calibrated とは，主観確率がある系列中で客観確率に長期的に一致すること（主観確率と客観確率の架橋），あるいは，長期的に信頼性 reliability を確保していること（Winkler[1967]，Hoerl and Fallin[1974]），あるいは，長期的に '現実のものであること' (realism) などの要請をいう．これらの性質が実現されているとき，「正しく尺度調整されている」(well-calibrated) という．正確にいうと，

現象 A（例として「雨」）の主観確率が α とされた N ケースのうち，

$N\alpha$ 回の A が実現していること

をいう．典型的な例として，気象，天候の確率予報がある（Murphy and Winkler[1977]，Lichtenstein 他[1977]，Borchering[1977]）．図5は，経験的尺度調整曲線（empirical calibration curve）と呼ばれ，たとえば，「降水確率 60%」と予想された 147 日の中で，現実に降水を観測したのが，これまた約 60%（確率予報として，完全的中）であることを示す．他の確率に対しても，的中は全般的に良く，'成績の良い' 予報者である．これは，およそ，確率予報なるものの評価として役立つ．ただし，これは，'良い' 予報者の十分条件ではない．

この尺度調整の問題も，標準的な統計手法によって取り扱うことができる．

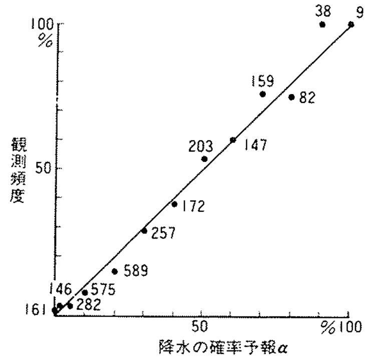


図5 降水確率予報の尺度調整（米国の例）
添付の数字は各 α に対する N. Murphy and Winkler[1977] 参照．

Cox[1958a, 1958b] は，Mosteller[1952] から示唆されて，確率を log odds に変換したのち，0-1 の二項回帰 (binary regression) の形にもち込むことによって，$H_0: P(Y_i=1) = p_i, (i=1, 2, \cdots\cdots, n)$ の仮説検定として扱った．

これは，連続パラメーター β を導入して

$$\log\{P(Y_i=1)/P(Y_i=0)\} = \beta \log\{p_i/(1-p_i)\} \qquad (5.1)$$

としたうえで，$\beta=1$ の検定に帰着する．なお，β の値で，他の傾向も検出できる．尤度関数を求めれば，十分統計量は，Y_i を

$$X_i = \begin{cases} \log(2p_i) & (Y_i=1 \text{ のとき}) \\ \log(2(1-p_i)) & (Y_i=0 \text{ のとき}) \end{cases} \qquad (5.2)$$

のように測り替えて，$X=\sum X_i$ となることがわかる．この X に対し，$\beta=1$ のもとで

$$E_1(X) = n\log 2 + \sum p_i \log p_i + \sum(1-p_i)\log(1-p_i) \qquad (5.3)$$

$$V_1(X) = \sum p_i \log p_i \{\log[p_i/(1-p_i)]\}^2 \qquad (5.4)$$

であり，さらに漸近理論を用いることができる．

例として，16 回の試行に対して，観測値 0, 1 が 8 回ずつあった．もともと，確率（予報）は，0 に対して，0.1, 0.2, 0.2, 0.4, 0.5, 0.6, 0.7，また，1 に対して，0.3, 0.3, 0.5, 0.6, 0.6, 0.8, 0.9, 0.9 であった．このとき，$X=1.106$ であるが，$E_1(X)=1.030$，$V_1(X)=0.785$ であるから観測値はこの確率指定に適合

していることがわかる．

付：3 囚人問題の発展

いま

$$P(s|A) = p, P(A|s) = x$$

とおくと，

$$x = p/(1+p)$$

となることが導かれる．アランの喜びが正当化されることは $x=1/2$ を意味するが，それは $p=1$ と同値である．そのためには，看守が，自己の言明をするにあたって，その言明が，アランによって，あいまいな状況の中で（も）バーナードが処刑されるものと，判断されることを，自ら十分了解していた，ということが前提とされねばならない．（'…provided he realizes[*that*]it means *that* Alan is judging *that* the jailer will report Bernard as the one to be executed in the ambiguous circumstance.'）たしかに，'自ら十分了解していた' とは，そう判断されたとしても '差し支えない' ということであり，人々が日常しばしば行う暗黙の通信の行為である．この場合は，看守は情報を与えているのである．だからこそ，確率が変わった．

すなわち，$P(s|A)=1/2$ は，バーナードとチャールズのどちらかという賭（lottery）であると看守が考えている[とアランが思った]ことであるのに対し，$P(s|A)=1$ は看守はどちらかを言う賭的状況の中でバーナードを必ず選ぶ（あるいは，既に選んだ）と[アランが思った]ということなのである．選んでしまえば不確実性の減少となり，たしかに情報は伝わっている．つまり，「賭的状況」そのものか「賭の選択結果」なのかという差が，本質的なのである．

さらに，もう一段，状況を複雑にしているのは，いわゆる Aumann の 'Common knowledge' 的状況である．すなわち，他人の判断に対する判断（接続詞 'that' の三重重畳）である．ベイズ的にいえば他人の主観確率に対する（関する）自分の主観確率である．この問題も既に興味深い問題として扱われているが，ここでは述べない（Aumann[1976]，Geanakoplos 他[1982]）．

注

1) **条件付確率** [この注は，後に読まれたい．]確率は，本来的に条件付であることは注意しておくべきことである．冬期に峠を越そうとしているドライバーが，'この雪でこの先の峠がはたして越せるだろうか' という確率的疑問に打ち当たったとき，これは P（不通）ではなく，P（不通｜雪）を考えているのである．さらに

（i） P（不通）は，彼にとって，何を意味するのか．

（ii） P（不通）と P（不通｜雪），あるいは P（不通，かつ，雪）と P（不通｜雪）が，実際に区別できるか．

（iii） P（不通｜雪）でなく，P（不通｜雪，かつ，彼が運転）の方が正確ではないか，いや，P（不通｜雪，かつ，彼が運転，かつ，警察が途中規制せず）がさらに正確ではないか，云々．

（iv） P（雪）は，現実に，彼にとって何の意味があるのか（その場から電話で気象庁に問い合わせるべきか）．

このような P(不通，かつ，雪)$=x$, P(不通｜雪)$=y$, P(雪)$=z$ を評価したとしても，さらに

$$x = y \cdot z$$

を満たさねばならない（これは，いわゆるコヒーレンス coherence の問題の1つである）．しかも，数学的には，未知数 3 個，方程式 1 個であるから，x, y, z にはまだ任意性が残っているのである．

すなわち，「条件付確率」そのものにも，現実上の課題がある．（以上は，一部 Lindley[1971]を参考にした．）

2) **論理と思考** '1 にかなり近いが 1 ではない確率値' という前提的概念そのものが成立している（いた）のだろうか．この，確率概念の成立は発展段階をたどるものであるという，ピアジェ的（発生史的）説明は 1 つの説明ではある．（竹内啓氏の指摘．）

主観確率を認めない統計学者はもちろん認める人々の中にも，このピアジェ的心理主義に嫌悪と反発を示す人々が多いと思われる．しかし，規範的（公理的）確率論の立場を認め，よく理解した上で，このような考え方にふれてみるのは，知的冒険としては，悪くないものである．ちなみに，なぜ「三段論法」が正しいか，あるいは，なぜそれを当然のものとして採用するかと

いう問題は，根本的には，心理問題を論ずることなしには論じ切れないであろう．

ここは，ピアジェが言ったことを述べるだけにしておこう．「論理の方が思考の鏡であって，逆ではない．」(これは繁桝算男氏の御教示による．)

3) **量・反応関係** 生物統計学における薬効検出，疫学における半数致死量 LD_{50} の測定などにおけるように，生物体（多くは，統制された集団）の化学物質等に対する反応程度について仮定される関係式．独立変数として投与・摂取される化学物質等の量，従属変数（生起確率）として理論上予期されている反応標識の生起率がとられる．ロジスティック曲線はその一例であるが，実際には，しきい値（threshold）の存在いかん等が問題にされることも多い．

4) **ポーカーの手の順序づけ** 歴史上経験的に成立した，手の強さ（出にくさ）の順序づけは，生起の確率計算によるものと一致している．次の通りである．

順位	経験的順序	確率（順列・組み合わせ計算）
(1)	ワン・ペア	4.23×10^{-1}
(2)	ツー・ペア	4.75×10^{-2}
(3)	スリーカード	2.11×10^{-2}
(4)	ストレート	3.92×10^{-3}
(5)	フラッシュ	1.97×10^{-3}
(6)	フル・ハウス	1.44×10^{-23}
(7)	フォー・カード	2.40×10^{-4}
(8)	ストレート・フラッシュ	1.39×10^{-5}
(9)	ローヤル・ストレート・フラッシュ	$1.5\ \times10^{-6}$

この確率計算の結果は，興味深いものである．(2)，(3)および(4)，(5)，(6)は有効数字のみで異なる接近した稀少確率であり，これが正しく順序づけられているのに驚く．この2群の中から，(2)および(4)を抜き出して，他も含めて配置すると（図中の矢印），ほぼ等間隔となる．これは稀少確率の尺度調整の標準として有用である．すなわち，(1)，(2)，(4)，(7)，(8)，(9)のどれかに近いかという問により，稀少確率の主観確率評価を定められるからである．この5種の確率の間隔の目安は，10のべき数の並び方から $10^{-1}=0.1$ がごく大略の数字であるが，ややよい近似として

$$\sqrt[5]{(1.5\times10^{-6})/(4.23\times10^{-1})} \fallingdotseq 0.0812$$

が考えられる．

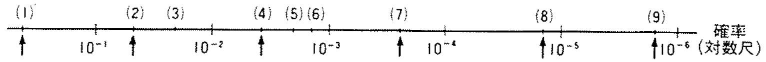


図6 ポーカーの手の生起確率（稀少確率の尺度調整）

たとえば，「明日あなたが交通事故で死ぬ確率はどれ位か．ローヤル・ストレート・フラッシュ(9)の確率より大きいでしょうか，小さいでしょうか，どう評価しますか」と聞くのである．ローヤル・ストレート・フラッシュとは，ランダムに選んだ5枚のカードが同種類で，A, K, Q, J, 10となることである．これで，個人が日常の危険（リスク）をどう評価し行動しているかがわかる．なお，この場合は，交通事故統計から頻度的（客観）確率が評価でき，およそ 3×10^{-7}，つまり(9)のほぼ1/5程度である．この程度の危険の確率が人間の日常生活の受忍（あるいは感知）限度であろう．（以上，松原『新版・意思決定の基礎』朝倉書店，1980，参照．）

参 考 文 献

Aumann, R. J., 1976 Agreeing to Disagree. *The Annals of Statistics,* 4(6), pp. 1236-1239.

Borchering, K., 1977 Calibration of Probabilities: The State of the Art. Comments. In Jungermann, H. and Zeeuw, G. de. (eds.), *Decision Making and Change in Human Affairs,* D. Reidel Publishing Company: Boston, pp. 325-329.

Cox, D. R., 1958a Two Further Applications of a Model for Binary Regression. *Biometrika,* 45, pp. 562-565.

Cox, D. R., 1958b The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society,* B, Vol. 20, No. 2, pp. 215-242.

Edwards, W., 1968 Conservatism in Human Information Processing. In Kleinmuntz, B. (ed.), *Formal Representation of Human Judgement,* New York: John Wiley, pp. 17-52.

Feller, W., 1976 *Introduction to Probability Theory and its Applications,* vol. 1, John Wiley and Sons.

Geanakoplos, J. D. and Polemarchakis, H. M., 1982 We can't disagree forever. *Journal of Economic Theory,* 28, pp. 192-200.

Green, P. E., Michael, H., and Robinson, P. J., 1965 An Experiment in Probability Estimation. *Journal of Marketing Research,* 2, pp. 266-273.

Hoerl, A. E. and Fallin, H. K., 1974 Reliability of Subjective Evaluations in a High Incentive Situation. *Journal of the Royal Statistical Society,* A, 137, pp. 221-230.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D., 1977 Calibration of Probabilities : The State of the Art. In Jungermann, H. and Zeeuw, G. de. (eds.), *Decision Making and Change in Human Affairs,* D. Reidel Publishing Company : Boston, pp. 275-324.

Lindley, D. V., 1971 *Making Decisions,* John Wiley and Sons.

Mosteller, F., 1952 Some Statistical Problems in Measuring the Subjective Response to Drugs. *Biometrics,* 8, pp. 220-226.

Murphy, A. H. and Winkler, R. L., 1977 Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Journal of Royal Statistical Society,* C, 26, pp. 41-47.

Peterson, C. R., 1973 Introduction to Special Issue on Hierarchical Inference. *Organizational Behavior and Human Performance,* 1, pp. 315-317.

Slovic, P. and Lichtenstein, S., 1971 Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgement. *Organizational Behavior and Human Performance,* 6, pp. 649-744.

Spetzler, C. L. and Steal von Holsstein, C. A. S., 1975 Probability Encoding in Decision Analysis, *Management Science,* 22, pp. 340-358.

Winkler, R. L., 1967 The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association,* 62, pp. 776-800.

Youssef, I. and Peterson, C. R., 1973 Intuitive Cascaded Inferences. *Organizational Behavior and Human Performance,* 10, pp. 349-358.

第3章
事前分布の選択とその応用

赤池 弘次

この章では，ベイズ理論の適用に際して常に問題となる事前分布の選択を取り上げ，どのように事前分布が構成され利用されるか，これまで筆者が考えたり実際に応用してみたりした事柄についてひとわたりの報告をしたい．

1. ベイズ的方法とは何か

ベイズ的方法はデータ分布 $f(\cdot|\theta)$ と未知パラメータ θ の事前分布 $p(\theta)$ とを用い，データ x が得られたときに θ の事後分布を，

$$p(\theta|x) = f(x|\theta)p(\theta)/p(x)$$

によって定義するものである．ただし

$$p(x) = \int f(x|\theta)p(\theta)d(\theta)$$

である．

この方法の定義自体は簡単なものであるが，何故この方法がよいものであるのかは，それほど自明のことではない．それどころか，T. ベイズがこの形式の推論の事例を論じた論文を遺して以来，現在にいたるまでの200年余りにわたって，この方法の妥当性に関する議論が繰返されてきているのである．

この議論の中心は，事前分布 $p(\theta)$ の選択にある．たとえば，ある貨幣を投げて表の出る確率を θ とする．n 回独立に投げて x 回表が出る確率は2項分布により

$$f(x|\theta) = {}_nC_x\theta^x(1-\theta)^{n-x}$$

と与えられる．実際データ x が観測された場合，θ の事前分布 $p(\theta)$ として何を採用すればよいであろうか．まず $p(\theta)$ として $[0,1]$ 上の一様分布をとることが考えられる．しかし，もしこの貨幣が普通の貨幣であるならば，θ が 0 とか 1 というような極端な値を取る可能性は少ないと考えられる．そこで

$$p(\theta) = 6\theta(1-\theta)$$

のような分布の方が一様分布よりもこの場合の事前分布としてはより適切ではないかと思われる．θ の真の値が 0.5 に近いことがほぼ分かっているならば，さらに $\theta=0.5$ の付近に確率が集中するような事前分布の方がよいと考えられるであろう．

この簡単な例で見ても分かるように，事前分布の選択は決して容易なものではない．それどころか，この問題をめぐる“哲学的な”議論は多数の人々を巻き込み，結局なんらの決定的な解決をも見なかったのである．

2. 迷路からの脱出

事前分布の選択の困難は，人々をベイズ的方法の実用化をめぐる迷路に誘い込んだ．伝統的な，客観性を持つ科学的方法として統計的推論の理論的展開を図った人々，たとえば R. A. フィッシャー，J. ネイマン，E. S. ピアソン等も，この迷路から逃れられなかった．

フィッシャーの fiducial probability の理論は事前分布の選択を客観的に行おうとする試みであり，ネイマン・ピアソンの検定理論における第 1 種の誤り，第 2 種の誤りの概念の導入は，事前分布の導入を避ける理論の展開を試みたものである．これらは，実際の統計的データの処理に対して広く有効な方法を与えるものとはならなかった．

これに対して，事前分布はとにかく存在しているのであって，各人がその持っている情報を十分に活用すれば決定できるものであるという主張が，いわゆる主観確率の立場の人々からなされた．この立場の人々の貢献は，確率という概念にまとわりついていた客観性（たとえば相対頻度の極限という解釈）をはぎ取り，不確実性の主観的評価の表現としての確率の本質を明らかにした点にある．

しかし，この人達も無意識の中に，主観性と客観性の混同に迷い込む．それはデータ分布に現われる未知パラメータという概念の主観性を見失うことに現われる．前述の貨幣の実験の例で表の出る確率を表す θ は，データを処理する人の意志とは無関係に，客観的にそこにあるものとの印象をわれわれに抱かせる．主観確率の立場を強く主張した B. ド・フィネッティはさらに極端に，確率論で扱うランダムなものは単に未知のものを意味し，その値は未知であるが確定しているものとした（de Finetti[1974]）．この考えに従うと，未知パラメータをランダムな変数として取扱う場合，たとえば壺の中に白玉と黒玉が入っており，何回かの取出し実験の結果にもとづいて白玉の割合を推定する場合のように，具体的客観的な意味を持つものに限ろうとする立場に導かれる．

これは，実際の統計的データ処理における未知パラメータの使い方を理解せず，単に壺からのランダムな抜き出しのような単純な例だけを考えたためである．このような限られた範囲の経験にもとづく議論の展開では，事前分布をめぐる迷路からの脱出は不可能である．

この迷路からの脱出は，未知パラメータを含むデータ分布が，実はデータから特定の目的に適した情報を抽出するための道具立てであること，事前分布とそれによる事後分布の計算は，データ x が与えられたときに，尤度関数 $f(x|\theta)$ を通じて得られる情報を適切に利用する仕組みを与えるものにすぎないこと，を理解することによってはじめて可能になる．

この見方からすれば，ベイズ的方法とは，データの与える情報の有効利用を目的とする人工的な仕組みであり，その有効性は，データ分布ならびに事前分布の構成の仕方に依存するものである．したがって，ベイズ的方法の有効性は，なんらかの公理によって保証されるものではなく，現実の問題に対して著しい有効性を示すような適用事例の集積を通じてはじめて，広く一般に受け入れられるようになるものであることが分かる．このように考えればこの迷路からの脱出は容易となる．

3. 情報抽出の仕組みとして見たベイズ的方法

ここでは，データからの情報抽出の仕組みとしてみた場合にベイズ的方法がどのように特徴づけられるかを，予測の視点から眺めてみることにする．

ベイズ的方法は，これまでの統計的方法で多く用いられてきた未知パラメータの推定法

$$x \to \theta(x)$$

にかえて，θ の上の分布を

$$x \to p(\theta|x)$$

のようにデータ x にもとづいて構成する．そこでまず一般にこのような方法の良さをどう評価するかが問題となる．

真の分布が $f(\cdot|\theta_0)$ で与えられる場合に，推定値 $\theta(x)$ の悪さを $f(\cdot|\theta_0)$ が $f(\cdot|\theta(x))$ からどれほど離れているかを示す値によって評価することにしよう．このために

$$I(f(\cdot|\theta_0); f(\cdot|\theta(x))) = E_y \log f(y|\theta_0) - E_y \log f(y|\theta(x))$$

を用いることにする．ただし E_y は y が $f(\cdot|\theta_0)$ に従って分布するものとした場合の y に関する平均値で，たとえば

$$E_y \log f(y|\theta(x)) = \int f(y|\theta_0) \log f(y|\theta(x))\, dy$$

である．$I(\ ;\ )$ は，カルバック情報量である．I の値が小さいほど $\theta(x)$ はよい推定量と見なされる．

さて $\theta(x)$ の代わりに $p(\theta|x)$ のように θ の上の分布が x の関数として与えられる場合を考えよう．このような $p(\theta|x)$ を一般的に推測分布と呼ぶことにする．推測分布 $p(\theta|x)$ の悪さを測るために，予測分布を

$$q(y|x) = \int f(y|\theta) p(\theta|x)\, d\theta$$

によって定義し，これと真の分布との離れ具合を $I(f(\cdot|\theta_0); q(\cdot|x))$ によって測ることにする．$p(\theta|x)$ の構成法の悪さの評価として，さらにこの I の x の真の分布についての期待値を考えることとし，

$$E_x I(f(\cdot|\theta_0); q(\cdot|x))$$

を用いることにする．これが小さいほど $p(\theta|x)$ の構成法はよいものとされる．

さてこの評価は θ_0 という θ のただひとつの値に対して現在の $p(\theta|x)$ の構成法を評価するものである．実際は θ_0 として種々の値の可能性があるわけであるから，このことを考慮して上記の評価量の代りに

$$E_\theta E_{x|\theta} I(f(\cdot|\theta); q(\cdot|x))$$

を考えることにする．ただしここで

$$I(f(\cdot|\theta); q(\cdot|x)) = E_{y|\theta} \log f(y|\theta) - E_{y|\theta} \log q(y|x)$$

であり，$E_{y|\theta}$ は $f(y|\theta)$ による y についての平均，$E_{x|\theta}$ は $f(x|\theta)$ による x についての平均，E_θ は θ のある分布 $p(\theta)$ に関する平均である．

ここで $p(\theta)$ が与えられた場合に，最良の推測分布 $p(\theta|x)$ を求めてみよう．このために上記の

$$E_\theta E_{x|\theta} I(f(\cdot|\theta); q(\cdot|x))$$

を最小にすることを試みる．

$$E_\theta E_{x|\theta} E_{y|\theta} \log f(y|\theta)$$

は $q(y|x)$ に無関係な定数であるからこれを無視することにすると，これは

$$E_\theta E_{x|\theta} E_{y|\theta} \log q(y|x)$$

が最大となるような $q(\theta|x)$ を求めることになる．

さてここで

$$\begin{aligned} E_\theta E_{x|\theta} E_{y|\theta} \log q(y|x) &= \iiint f(y|\theta) f(x|\theta) p(\theta)\, d\theta \log q(y|x)\, dy dx \\ &= \iiint f(y|\theta) \frac{f(x|\theta) p(\theta)}{p(x)} d\theta \log q(y|x)\, dy p(x)\, dx \end{aligned}$$

と書ける．ただし

$$p(x) = \int f(x|\theta) p(\theta)\, d\theta$$

である．上式を見やすくするために書き直すと

$$E_x E_{y|x} \log q(y|x)$$

となる．ただし E_x は $p(x)$ に関する平均，$E_{y|x}$ は

$$p(y|x) = \int f(y|\theta)\left[\frac{f(x|\theta)p(\theta)}{p(x)}\right]d\theta$$

によって与えられる y の分布 $p(x|y)$ に関する平均である．

さて

$$I(p(\cdot|x); q(\cdot|x)) = E_{x|y}\log p(y|x) - E_{y|x}\log q(y|x) \geq 0$$

であることから，

$$E_{y|x}\log q(y|x)$$

は

$$q(y|x) = p(y|x)$$

の時に最大となることが分かる．この結果から，$p(y|x)$ と $q(y|x)$ の定義を比較すると，

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{p(x)}$$

が最良の推測分布を与えるものであることが分かる．これは $p(\theta)$ を事前分布とする場合の θ の事後分布である．

上記の結果は，推測分布の悪さの評価を定め，さらに各 θ におけるパフォーマンスの総合評価の荷重関数として $p(\theta)$ を考えると，$p(\theta)$ を事前分布とするベイズの方法が最良の推測分布を与えることを示す．このように見ると，事前分布 $p(\theta)$ は単に適当な推測分布を構成するための設計基準を与えるにすぎないことが分かる．このような設計基準が実用上よいものであるかどうかは，それによって得られる推測分布の悪さをそれぞれの θ について検討する，いわゆる動作特性 (performance characteristic) を見ることによって判断するしかない．このためには，$E_{x|\theta}I(f(\cdot|\theta); p(\cdot|x))$ を θ の関数として観察すればよい (Akaike[1978])．

4. 事前分布に対する基本的な誤解

一時期ベイズ的方法の利用を主として推進したのは主観確率の立場に立つ人々であった．これらの人々は事前分布の構成は各人が持つ情報に依るべきものであることを主張し，したがって一般論として研究可能なものは主観確率の評価法の展開だけとなった．ただし特殊なものとして，事前分布に直観的な制約条件を課し，これによってその構造を定めることが試みられた．よく知られているのはド・フィネッティによる交換可能分布 (exchangeable distribution) の議論である (de Finetti[1975])．

このような議論は，ある種の客観的な規定によって事前分布を定めようとするものであり，事前分布はそれぞれの場合に各人の持つ情報に依存して決定されるべきだとする主観確率の立場の基本的態度に反するものといえる．ベイズ的方法の推進に関して主観確率の立場の人々が展開した議論は，このように常に何かしら客観性を求めようとする矛盾に満ちたものであった．

これに反し，まったく便宜主義の観点から事前分布の構成を考える立場もある．これはベイズの方法の適用に際しての解析的便宜性に注目する立場であり，与えられたデータ分布に対し，それに対応する事後分布の計算が容易に行われるような事前分布の型を考えるものである．これはベイズ的方法の本質がよく理解された後では，実用上の便利さの点から利用価値がある．しかし事前分布がなんらかの客観性をもって一意的に定まるかの印象を与える危険性には十分注意する必要がある．

主観確率の立場からは，事前分布 $p(\theta)$ はそれぞれの人が持つ未知パラメータ θ に関する事前情報を十分に活用すれば決定できる，という発言がなされる．このために，この立場では事前分布の構成法については具体的には何も論じることができない．ここでは θ が $f(\cdot|\theta)$ というデータ分布の“想定”にもとづいて生み出されるものであることが見逃されているのである．

主観確率の立場の人々がベイズ的方法を論じる場合によく見られる誤りは，未知パラメータ θ が具体的な意味をも含めて先験的に与えられたものと見なして議論を進めることである．この立場に立つと，データからの情報抽出の道具として設定されるデータ分布 $f(\cdot|\theta)$ の主観性がまったく見逃されてしまう．確率の主観性を主張する人々が，かえって未知パラメータ θ を何か客観的なものと見なすという混乱がここにあるのである．

この混乱を端的に示すものがいわゆる尤度原理によるベイズ的方法の正当化の議論である (Edwards 他 [1963])．この原理は，変数 x に対するデータ分布 $f_1(x|\theta)$ と $f_2(y|\theta)$ とがあるとき，$f_1(x|\theta)=cf_2(y|\theta)$ (c：定数) がすべ

ての θ について成立するような場合には，データ x にもとづいてなされる θ に関する推論はデータ y にもとづいてなされるものと同じであるべきことを主張する．未知パラメータ θ に関する事前分布 $p(\theta)$ が一意的に存在するとすれば，上の場合にはそれぞれの θ の事後確率分布 $p(\theta|x)$ と $p(\theta|y)$ とが相等しくなり，ベイズ的方法はこの原理の主張に反しないこととなる．

よく引合に出される，2項分布と負の2項分布の場合についてこれを見ることにしよう．ある貨幣を投げて表の出る確率を θ とする．n 回独立に投げる実験で x 回表が出る確率は2項分布

$$f_1(x|\theta) = {}_n\mathrm{C}_x\theta^x(1-\theta)^{n-x} \qquad (x = 0, 1, \cdots, n)$$

で与えられる．次に x 回表が出るまで投げ続ける実験を考え，y 回かかったとする．この時は結果の確率は負の2項分布

$$f_2(y|\theta) = {}_{y-1}\mathrm{C}_{x-1}\theta^x(1-\theta)^{y-x} \qquad (y = x, x+1, \cdots)$$

で与えられる．いま $y=n$ となった場合を考えると，

$$f_1(x|\theta) = \binom{n}{x}f_2(n|\theta)$$

となる．尤度原理は，この2つの場合に，θ に対する推論は同一のものとなるべきであるというのである．

伝統的な統計理論に従えば，たとえば θ の真の推定値として何をとるかは，その推定値に対して期待される特性にもとづいて判断される．たとえば期待値が真の値と一致する不偏性を要求すれば，上の2つの場合には異なった推定量を用いることになり，これは尤度原理に反する．したがって，不偏性などの概念を論じる伝統的な統計的理論は誤りであり，これに対して，同一の事後分布を与えるベイズの方法は，このような誤りをおかすことはない，というのが主観的ベイジアンの主張である．

B.ド・フィネッティ，L.J.サベジの主観確率の立場を継承するD.リンドレイは，A.バーンバウムによる尤度原理の議論以来，ベイズ的立場の正当化のひとつのよりどころとして，繰返しこの原理に言及している（Lindley[1975]）．事前分布の構成法を積極的に発展させるには，この議論の誤りの指摘が何より必要なのである（Akaike[1982]）．

尤度原理そのものの証明の不当性はさておき，ここではベイズ的立場が尤度原理に反しないという主張の誤りを指摘することにする．同一の尤度関数（すべての未知パラメータについてその値が一定の比例関係を保つ2つの尤度数は同一と見なされる）が同一の事後分布を与えるとの主張を正当化するためには，データ分布の構造に関係なく，未知パラメータに想定される事前分布が同一であるとの主張がまず必要である．

ここで主観確率の立場をあらためて述べてみると，合理的な行為者は不確定あるいは未知な事柄に対しては確率分布を想定すべきである，というものである．さらに，この確率分布の想定は，その行為者が事前に持つすべての情報に依拠してなされなければならないとするのである．

さてベイズ的方法の適用に際しては，事前分布 $p(\theta)$ の想定は未知パラメータ θ に対してなされるのであり，ベイズの方法の適用に際しては，データ分布 $f(\cdot|\theta)$ は既に定義されていなくてはならない．したがって，$p(\theta)$ の決定に際して利用し得る情報の中には $f(\cdot|\theta)$ が含まれているから，$p(\theta)$ がデータ分布の如何に関係なく一意的に定まるべきものとする立場は，データ分布の構造に関する知識の無視を要求するものであることが分かる．かくしてこの立場は，事前の情報をことごとく利用して確率分布は決定さるべきであるとする，主観確率の教えに真向から反するものであることが明らかとなる．

5. 事前分布の構成

事前分布の構成に関する筆者の立場は，データ分布の想定そのものがデータの特定の性質を測ること，あるいはデータの含む特定の情報を抽出することを目的としてなされるものであり，このデータ分布の含む未知パラメータの事前分布は，求める情報を適切に提供する統計的データ処理法をベイズの方法が与えるように構成されるべきである，と考えるのである．

これを要約すれば，ベイズの方法を適用して，人々がこれは役に立つデータ処理法だと感じるような結果が得られるように事前分布を構成すればよい，というわけである．ここでは客観性が要求される．すなわち，誰が見てもその意味が理解でき，納得できるような構成法であることが必要である．

さて，ひとつのデータに関してある人に固有の情報というものについては，一般的な議論は不可能である．そうなると，われわれが前提できるのはデータ分布の構造だけである．これまでの統計学の発展の過程で，数多くのデータ分布（の族）が考案されてきている．これらはいわゆる統計モデルとして知られているもので，データのさまざまな興味ある特性の測定のために開発されたものである．データ分布の特性にもとづく事前分布の構成法の議論が進められれば，これらのモデルについて必要に応じて適切な事前分布を構成することが可能となる．さらに進んで，ベイズ的方法の適用に適したデータ分布の開発も考えられるようになり，ここに伝統的な統計的方法の枠を超えるデータ処理法が実現されることになる．こういうわけであるから，伝統的な統計的方法の枠を広げる必要がどこにあるかを探れば，そこにベイズ的方法の実用化の端緒が得られるであろうことが予想される．

データ分布 $f(\cdot|\theta)$ に対してデータ x が与えられると，尤度関数 $f(x|\theta)$ によってデータの与える情報のすべてが表現される．これに表現されない情報は無視されることになる．ここでより重要な縮約した情報を求めるために，θ の“真の値”を推定することが考えられる．$f(x|\theta)$ の値を最大にする θ の値を $\theta(x)$ とし，これを真の値の推定値とするのが最尤法 (the method of maximum likelihood) である．最尤法は多くの問題に適用され，その実用上の有効性は広く認められている．この最尤法が有効に適用されない場面こそ，ベイズ的方法の適用が要求される典型的な場面を与えるであろう．このような場面として，データの与える情報に比して未知パラメータの数が多すぎる場合があるわけである．

かくして，事前分布の構成に有効な手順として，次のようなものが考えられる．

① 当面の情報処理の要求をよく表現するデータ分布を，たっぷりと多くの未知パラメータを用いて構成する（未知パラメータの数が観測データの数を超えてもかまわない）．

② この多くの未知パラメータを放任することなく，これらが満たすべきゆるやかな制約条件を表現する確率分布を考える．制約条件の強さを表すパラメータを未知パラメータとして持つ事前分布を構成する．

③ データ分布の残りの未知パラメータ，ならびに想定された事前分布の未知パラメータについては，最尤法の適用によってその値を決定する（この場合の尤度関数は，もとのデータ分布の尤度の，想定された事前分布に関する平均値で定義される），あるいは特定のパラメータに偏ったウェイトを置かない，いわゆる非報知 (non-informative) 事前分布を想定する（この分布は，場合によっては不適切 (improper) な分布，全積分が無限大になるようなもの，になることがある）．

上の①と②の段階が，ベイズ的方法の有効性を左右するものとなる．①については，ベイズ的方法の適用によってデータからどのような情報を得ようとしているのかについての十分な理解が必要である．すなわち，具体的な問題についての理解がなくては，新しいモデルの開発は不可能である．

②については，データ分布の与える尤度関数の解析的な特性の理解が必要である．当面のデータ解析の目的から見て無意味な未知パラメータの値で尤度が高い値を取る可能性がある場合に，その影響が事後分布に現われないように事前分布を設定するわけである．なおここでの事前分布は，何段かの階層をもって構成される場合もあり得る．

③の段階は，ベイズ・モデルの比較検討をどう行うかの問題を含んでいる．統計モデルの適用に際しては，常にその結果の検討が必要である．これは一般にいくつかのモデルの比較という形で実現される．ベイズ・モデルの場合，不適切事前分布が用いられていない限り，モデルの比較はそれぞれのモデルの尤度の比較を通じて実現できる．ただし，いくつかのパラメータが最尤法で決定されたモデル間の比較には，情報量規準 AIC をベイズ・モデルの場合に拡張したものの利用などを考える必要がある．

いくつかの不適切事前分布が用いられる場合には，モデルの比較に尤度の考えを適用することが困難となる．不適切事前分布は，それに正の定数を乗じたものも同一視されるからである．

6. 事前分布構成の実例とその応用

この節では，簡単なデータ分布から，適切な事前分布の想定によって，実

用上有効なデータ解析のための方法が得られることを示すことにする(Akaike[1980a]).

データ $x=(x_1, x_2, \cdots, x_n)'$ が与えられたとする.($'$ は縦横の転置を示す).以下の議論では,誤解のおそれのないかぎり,データ x の解析のために構成するデータ分布の変数も同じ x によって表現することにする.また場合によっては,この分布に従う確率変数をも同じ x によって表すことにする.

いま x のデータ分布について

$$x_i = m_i + z_i$$

で与えられる構造を考える.ただし $z=(z_1, z_2, \cdots, z_n)'$ の要素 z_i は他の要素と独立に平均 0,分散 v を持つ正規分布に従うものとする.未知パラメータ(ベクトル)θ として,$\theta=(m_1, m_2, \cdots, m_n)$ を考える.v を未知パラメータとすれば,当然未知パラメータの総数はデータの総数 n より大きい.

θ のデータ x に関する尤度は

$$f(x|\theta) = \left(\frac{1}{2\pi v}\right)^{n/2}\exp\left(-\frac{1}{2v}\sum_{i=1}^{n}(x_i-m_i)^2\right)$$

で与えられる.これに最尤法を形式的に適用すると,$m=x, v=0$ という結果が得られる.

ここで,$x_1, x_2, \cdots, x_n$ は時系列の観測データで,$m_1, m_2, \cdots, m_n$ はそのトレンド(傾向線)を示すものと考えると,実用上よく考えられるのは,時とともに滑らかに変化するようなトレンドである.このことは

$$d_i = (m_i - m_{i-1}) - (m_{i-1} - m_{i-2}) \qquad (i = 3, 4, \cdots, n)$$

が 0 に近い動きをすることを想定することに相当する.0 への近さは,観測値固有の変動を示す z の分散 v に比べて,上記の階差のバラツキが小さいということによって表現される.

そこで $(d_3, d_4, \cdots, d_n)$ に対する事前分布として d_i が互いに独立に,平均 0,分散 τ を持つ正規分布に従うものと仮定してみよう.$\tau=v/c^2 (c>0)$ として,c が 1 に比べて十分大きければ,この事前分布を想定することによって,d_i が z_i の変動に比して十分小さいような m_i で x_i の動きの中心的傾向をよく示すようなものの事後確率が高くなる.

この $(d_3, d_4, \cdots, d_n)$ の密度関数は

$$\left(\frac{1}{2\pi\tau}\right)^{n-2}\exp\left\{-\frac{1}{2\tau}\sum_{i=3}^{n}d_i^2\right\}$$

で与えられるが,これだけでは $m=(m_1, m_2, \cdots, m_n)$ の分布は定まらない.そこで $d_1=m_1$,$d_2=m_2-m_1$ として,$d=(d_1, d_2, \cdots, d_n)$ に対して

$$p(d) = \left(\frac{1}{2\pi\tau}\right)^{n}\exp\left\{-\frac{1}{2\tau}[(ad_1)^2+(bd_2)^2+d_3^2+\cdots+d_n^2]\right\}$$

を想定すれば,これを m の分布に変換することによって,m の事前分布が得られる.これを

$$p(m) = \left(\frac{1}{2\pi\tau}\right)^{n}\exp\left\{-\frac{1}{2\tau}m'D'Dm\right\}$$

と書くと,

$$D = \begin{bmatrix} a & & & & & \\ b & -b & & & 0 & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ 0 & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}$$

と与えられる.

この事前分布は,τ, a, b, という 3 個の定数で定義されている.a, b を 1 に比べて十分小さくとれば,この分布は $d_1(=m_1), d_2(=m_2-m_1)$ の値に反応しないものと実用上見なされるようにできる.問題は τ の選び方である.外部的な要請によって τ の値が与えられる場合は問題ないが,単にゆるやかに変動するトレンドを捉えたいとする一般の応用では τ の値があらかじめ与えられることは殆どない.これがデータ解析のためにベイズ的方法が用いられる場合の典型的な状況である.

この状況では,事前分布は事前情報で定められるというベイズ統計学の説明する状況とは異なり,データの特性を把握するためにベイズ的構造が利用されているのである.したがって,τ の値の選び方にデータの特性の把握の成否がかかるわけである.一歩進んで,ここでは与えられるベイズ的構造がデータの特性をよく反映するような τ を選ぶことが要求されているのだと考えることもできる.このように考えると,問題は現在のデータの特性とよ

く対応するベイズ的構造の選択あるいは推定であると見なされる.

さて，ベイズ・モデルのデータとの適合の良否は，ベイズ・モデルの尤度によって評価できる．データ分布 $f_k(x|\theta)$ と事前分布 $p_k(\theta)$ とで与えられるベイズ・モデルの集まり $(k=1, 2, \cdots, K)$ を考える．データ x が与えられた場合，k 番目のモデルの尤度は

$$l(k) = \int f_k(x|\theta_k)\, p_k(\theta_k)\, d\theta_k$$

によって定義される．ベイズ的方法を一貫して適用しようとすれば，このベイズ・モデルの集まりの上の事前分布 $p(k)$ を考え，$l(k)$ と組合せてベイズ的推論を進めることになる．しかし，どれかひとつだけの k に対して $l(k)$ が著しく高い値をとる場合には，最尤法の考えを適用して，そのモデルだけを用いて推論を進めても十分実用的な結果が得られることになる．

前記の m の事前分布を定める τ の場合には，上記 k の代わりに τ がひとつひとつのベイズモデルを与えるわけである．この場合の τ の尤度は

$$l(\tau) = \int \left(\frac{1}{2\pi v}\right)^{n/2} \exp\left\{-\frac{1}{2v}\|x-m\|^2\right\} \left(\frac{1}{2\pi\tau}\right)^{n/2} |D| \exp\left\{-\frac{1}{2\tau}\|Dm\|^2\right\} dm$$

によって与えられる．ただし $\|\ \ \|$ はユークリッド・ノルムを表し，$\|x\|=x'x$ で与えられる．また $|D|$ は D の行列式の値を示す．

ここで $\tau=v/c^2$ と表現して $l(\tau)$ を $c(>0)$ の関数と考えれば，さらに z の分散 v も与えられるべき定数として残っていることが分かる．これは今の ベイズ・モデルの中でデータ分布に関係するものである．これを未知パラメータとすると，厳密にベイズ的方法に従うためには，m と v と c の同時分布を事前分布として与えなければならないことになる．ここでは，v に対して最尤法が有効に働く場合を想定して議論を進めることにする．そうすると，前記の尤度 $l(\tau)$ は未知パラメータ v と c との関数 $l(v, c)$ となり，

$$\operatorname*{Max}_{v,c}\ l(v, c)$$

を満たす v と c の値を求めることによって，現在のデータ x の特性をよく表現するベイズ・モデルが求まることになる．実用上は，c のひとつの値に対して $l(v, c)$ を最大にする v を求め，これに対する $l(v, c)$ の値が最大となるように c の値を数値的に探索すればよい．

このようにして v, c の値が決定されたものとすると，目的とするトレンドに関する推論は m の事後分布を利用して実現される．この事後分布は

$$p(m|x) = K \exp\left\{-\frac{1}{2v}[\|x-m\|^2+\|cDm\|^2]\right\}$$

の形で与えられる．K は m に関する全積分を 1 とする定数である．m の姿として一番確からしいものを求めるには，この $p(m|x)$ を最大にするもの，いわゆる事後分布のモードを求めてみればよい．これには

$$\operatorname*{Min}_{m}\ \{\|x-m\|^2+\|cDm\|^2\}$$

を解けばよい．これは x に D の行数だけの 0 をつけ加えて定義されるベクトル x^* を考え

$$x^* = \begin{bmatrix} x \\ 0 \end{bmatrix} \qquad F^* = \begin{bmatrix} \mathrm{I}_{n\times n} \\ cD \end{bmatrix}$$

とおけば，

$$\operatorname*{Min}_{m}\ \|x^*-F^*m\|^2$$

という最小 2 乗計算によって解が得られる．この計算の実行には，たとえばハウスホールダー変換を利用する等の優れた数値計算の手順がある．

上記の最小 2 乗計算の解を m_0 とすると

$$\|x^*-F^*m\|^2 = \|x^*-F^*m_0\|^2+\|F^*(m-m_0)\|^2$$

と書ける．この結果を利用すると前出の尤度 $l(v, c)$ が

$$l(v, c) = \int \left(\frac{1}{2\pi v}\right)^n c^n |D| \exp\left(-\frac{1}{2v}s^2\right) \exp\left\{-\frac{1}{2v}\|F^*(m-m_0)\|^2\right\} dm$$

と書けることが分かる．ただし $s^2=\|x^*-F^*m\|$ である．これは拡張した最小 2 乗問題の残差 2 乗和である．指数関数部分の積分は

$$(2\pi v)^{n/2}|F^{*\prime}F^*|^{-1/2}$$

で与えられるから，結局

$$(-2)\log l(v, c) = n\log(2\pi)+n\log v-n\log c^2$$
$$-2\log|D|+\log|F^{*\prime}F^*|+\frac{1}{v}s^2$$

が成立する．これを最小にする v の値は

$$v_0 = \frac{1}{n}s^2$$

で与えられ，

$$(-2)\log l(v_0, c) = n\log s^2 - n\log c^2 - 2\log|D| + \log|F^{*\prime}F^*| + \text{const}$$

となる．ここで const は尤度の最大化に無関係な定数である．

$l(v, c)$ が最大となるように c を選ぶのが最尤法であるが，$\tau = v/c^2$ によって c が定義されたことを考えると，$c>1$ の範囲に限ることが自然である．さもないと，x の変動分 z が m の中に吸収されるようになってしまう．このような状況を生じる事前分布は，滑らかなトレンドを求めようとする目的に適さないものと言えるからである．

これまで議論した簡単なベイズ・モデルにおいて，平均ベクトル m により一般的な線形構造を想定することにより，きわめて応用範囲の広い線形モデルに関するベイズ的方法を得ることができる．平均ベクトル m が既知のベクトル $f_1, f_2, \cdots, f_k$ の線形結合として $m = \theta_1 f_1 + \theta_2 f_2 + \cdots + \theta_k f_k$ と表されるものと想定し，$\theta = (\theta_1, \theta_2, \cdots, \theta_k)'$ を未知パラメータとすると，

$$x = F\theta + z$$

と表される．ただし $F = (f_1, f_2, \cdots, f_k)$ であり，z については前と同じく平均ベクトル 0，分散行列 $vI_{n\times n}$ の正規分布を想定する．θ に関しては，平均 θ_0，分散行列 R の正規分布を想定する．ここでは R として，適当な $p\times k$ 行列 D を用いて

$$R = \frac{c^2}{v}D'D$$

と書ける場合を考える．v, c を未知パラメータとすると，このベイズ・モデルの尤度は

$$l(v, c) = \int \left(\frac{1}{2\pi v}\right)^{n/2} \left(\frac{c^2}{2\pi v}\right)^{n/2} |D'D|^{1/2}$$
$$\exp\left\{-\frac{1}{2v}[\|x - F\theta\|^2 + c^2\|D(\theta - \theta_0)\|^2]\right\} d\theta$$

によって与えられる．

v, c の値が与えられたとして，θ の事後分布のモードを求めるには

$$x^* = \begin{bmatrix} x \\ cD\theta_0 \end{bmatrix} \qquad F^* = \begin{bmatrix} F \\ cD \end{bmatrix}$$

と置き，

$$\underset{\theta}{\text{Min}}\ \|x^* - F^*\theta\|^2$$

を解けばよい．この解を θ_0 とする．

前記の滑らかなトレンドのモデルの場合と同様にして，このモデルの尤度は，次のように与えられる．

$$(-2)\log l(v, c) = n\log 2\pi + n\log v - n\log c^2$$
$$-\log|D'D| + \log|F^{*\prime}F^*| + \frac{s^2}{v}$$

ただし $s^2 = \|x^* - F^*\theta_0\|$ である．与えられた c に対して $l(v, c)$ を最大にする v の値 v_0 は $v_0 = s^2/n$ によって与えられる．したがって

$$(-2)\log l(v_0, c) = n\log s^2 + n\log c^2 - \log|D'D| + \log|F^{*\prime}F^*| + \text{const}$$

となる．ただし const は c に無関係な定数を示す．これを最小にするような c の値を $c > c_L$（適当に決定された下限）の範囲で求め，これに対応する事前分布を利用する．

以上のような手続きによって，ベイズ型の季節調整，R. J. シラーのモデルによる分布ラグの推定等を容易に実用化することができる（Akaike[1980b, 1986]）．

7. おわりに

この章で議論したことを要約すれば，ベイズ・モデルを利用して新しい統計的推測法が容易に実用化できるということである．その際基本となるのは，データ分布の適切な構成とその尤度関数の特性の理解である．この考え方に従えば，因子分析，時系列解析などについても有効な事前分布の構成が実現できる（Akaike[1984, 1987]）．

最終的に問題となるのは，いくつかのモデルの比較である．これに対しては，対数尤度と情報量概念との対応を利用して最尤法あるいは最小 AIC 法の適用を考えればよい．

付記：この報告の準備に際しては，研究テーマ「情報量概念に基づく統計学の展開」に対して，高橋産業経済研究財団からの研究助成を得た．

参 考 文 献

Akaike, H., 1978 A new look at the Bayes procedure, *Biometrika*, 65, 53-59.

Akaike, H., 1980a Likelihood and the Bayes procedure with discussion, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.), *Bayesian Statistics*, Valencia, University Press, 143-166, 185-203.

Akaike, H., 1980b Seasonal adjustment by a Bayesian modeling, *Journal of Time Series Analysis*, 1, 1-13.

Akaike, H., 1982 On the fallacy of the likelihood principle, *Statistics & Probability Letters*, 1, 75-78.

Akaike, H., 1984 On the use of Bayesian models in time series analysis, J. Franke, W. Hardle and D. Martin (eds.), *Robust and Nonlinear Time Series Analysis*, Berlin, Springer, 1-16.

Akaike, H., 1986 The selection of smoothness priors for lag estimation, P. K. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques*, Amsterdam, Elsevier, 109-118.

Akaike, H., 1987 Factor analysis and AIC, *Psychometrika*, 52, 317-332.

Edwards, W., Lindman, H. and Savage, L. J., 1963 Bayesian statistical inference for psychological research, *Psychological Review*, 70, 450-499.

de Finetti, B., 1974, 75 *Theory of Probability*, Vols. 1, 2, London, Wiley.

Lindley, D. V., 1975 The future of statistics—A Bayesian 21st century, S. G. Ghurye (ed.), *Supplement to Advances in Probability*, 106-115.

第4章
ベイズ決定方式と許容性

竹 村 彰 通

1. は じ め に

統計的決定理論（statistical decision theory）は推定・検定等の統計的推測の方法を統一的な枠組の中で扱うための数学的理論である．統計的決定理論はワルド（A. Wald）によって創始され，ワルドの著書 *Statistical Decision Functions*（1950）は現在でもこの分野の基礎的な文献となっている．統計的決定理論の枠組は数理統計学における1つの標準的な枠組となり，その後の数理統計学の理論書は多かれ少なかれ統計的決定理論の枠組を用いている．代表的なものとしてレーマンの2冊の教科書があげられよう．数理統計学に関する論文のレベルでも統計的決定理論の諸概念は共通の前提として用いられており，統計的決定理論を数理統計学一般の研究と分離して論じるのは難しい．

このように統計的決定理論は数理統計学の研究の中に深く根をおろしているが，他方，統計的決定理論自体の研究はかなり特殊化しており研究者も限られているのが現状である．これは統計的決定理論の一般論がしばしば数学的に高度な手法を必要とすることが一因となっている．教科書について見ても統計的決定理論そのものを整理した教科書はあまり多くない．ファーガソンの教科書（Ferguson[1967]）は理論体系をわかりやすく整理した大学院レベルの教科書である．バーガーの教科書（Berger[1985]）はファーガソンの教科書より一段やさしいレベルで統計的決定理論を紹介しており，そのため数学的な証明など省略している部分がある．またバーガーは現在の統計的決

定理論研究の中心メンバーの1人であり，バーガーの教科書は統計的決定理論の最近までの成果を調べるのに便利な教科書である．本稿の話題については特に8章が参考になる．より最近の研究書としてはブラウン（Brown [1986]）があげられる．ブラウンは長い間この方面の中心的な研究者であったが彼の論文はしばしば難解である．ブラウンの本自体は高度な研究書であるが，よくまとまっておりブラウンおよび彼の周辺の人々の仕事を理解するのに非常に有用な本である．本稿もバーガーおよびブラウンの本に負うところが大きい．和書では鍋谷［1978］が統計的決定理論を中心に書かれたすぐれた教科書であり，本稿の話題についてはこの本の第6章にかなり詳しく述べられている．

統計的決定理論は統計的推測の問題を「自然」と「統計家」という2人のプレーヤーによる零和2人ゲームとしてとらえる．このゲーム理論の応用としての統計的決定理論という観点は初期には強く意識されていたが（たとえば Blackwell and Girshick[1954]），統計的決定理論がおもに零和2人ゲームの場合に限られていることもあり，ゲーム理論と統計的決定理論は別々の方向に発展して現在では両者の関連は薄くなっている．ところで零和2人ゲームという観点から見ると，ベイズ統計学における事前分布は自然の側の混合戦略として自然な形でとらえられる．これは主観確率を基礎とするベイズ統計学の考え方とは異なるが，ベイズの方法の1つの自然な動機付けを与える．以下の節ではこの観点からベイズの方法を考察してゆく．

2. 統計的決定理論の枠組とベイズの方法

この節では統計的決定理論の枠組とその中でのベイズの方法について基本的な事柄を説明する．統計家が得た観測値を X とする．ここでは簡単に X と記すがこれは必ずしも1個の観測値というわけではなく，通常は大きさ n の標本 $(X_1, \ldots, X_n)$ を簡略に X とあらわしたものである．推定の問題では観測値 X にもとづいて，未知パラメータ θ の値を $\hat{\theta}(X)$ で推定する．$\hat{\theta}(X)$ を確率変数 X の関数と見たとき θ の推定量と呼ぶ．推定量とは X のあらゆる可能な値にたいしてどのように推定するかを決めたものであり，

その意味で推定方式と呼んでもよい．検定の問題では観測値 X にもとづいて帰無仮説を受容するか棄却するかを決めるわけであるが，ここで受容を0，棄却を1で表そう．いま $\phi(X)$ を X にもとづき，受容するとき0，棄却するとき1，をとる関数とすれば1つの検定方式が定まる．$\phi(X)$ は検定関数と呼ばれるものである．

ここで推定の場合の $\hat{\theta}(X)$ および検定の場合の $\phi(X)$ を含めてこれらを統一的に $\delta(X)$ と表そう．$\delta(X)$ は決定方式（decision rule）と呼ばれ，観測値 X を得た時の統計家の行動を表す関数である．ところで統計家は誤差を含んだ観測値にもとづいて行動するから最適な行動を常にとることはできない．たとえば推定の問題においては $\hat{\theta}$ は θ からは確率的にずれてくる．そこで最適な行動からの乖離を表す損失関数（loss function）$L(\theta, d)$ を導入する．$L(\theta, d)$ は θ が真のパラメータである時に d という行動をとった場合の統計家の被る損失を表す．推定の問題では通常“2乗誤差”$L(\theta, d)=(\theta-d)^2$ を損失関数として考える．検定の問題では正しい決定を行った時には損失は0，第1種の過誤および第2種の過誤をおかした時には損失は1とする0—1損失がもっとも簡単である．ここで正しい決定とは帰無仮説が正しい時に帰無仮説を受容する，あるいは対立仮説が正しい時に帰無仮説を棄却することを意味する．また第1種の過誤とは帰無仮説が正しいときにこれを棄却することを意味し，第2種の過誤とは対立仮説が正しいときに帰無仮説を受容することを意味する．以下では明示的な記述がなければこれらの損失関数が用いられるものと仮定する．

統計的決定理論では決定方式のよさを損失の平均で評価する．もちろん，損失が平均的に小さい決定方式が望ましい決定方式である．そこで決定方式のリスク関数（risk function）$R(\theta, \delta)$ を

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$$

で定義する．ここで E_θ は真のパラメータが θ の時の X の分布に関する期待値を表している．推定で損失関数を2乗誤差にとる場合にはリスクは平均2乗誤差である．検定で0—1損失関数を考えるときにはリスクは θ が帰無仮説に属する場合には第1種の過誤の確率，また θ が対立仮説に属する場合には第2種の過誤の確率となる．

決定方式の望ましさをそのリスクで評価するとすれば，あらゆる決定方式の中でもっとも望ましい決定方式は，$R(\theta,\delta)$ をすべての θ について一様に最小にする決定方式である．しかし，通常個々の θ についてリスクを最小にする決定方式は θ ごとに異なり，すべての θ について一様にリスクを最小にする決定方式は一般には存在しない．そこで，もっと弱い最適性の基準として許容性と呼ばれる概念を考えることができる．

ある決定方式 δ が他の決定方式 δ' よりよいとは（$\delta>\delta'$）すべての θ について

$$R(\theta,\delta)\leq R(\theta,\delta') \tag{1}$$

が成立ち，かつある θ について

$$R(\theta,\delta)<R(\theta,\delta') \tag{2}$$

となることを言う（図１参照）．また(1)のみを仮定して(2)を要求しない場合には $\delta\geq\delta'$ と表し δ が δ' よりよいかまたは同等ということにする．δ が許容的（admissible）とは δ よりよい他の決定方式が存在しないことを言う．すなわち δ を一様に改善することができないならば，δ は許容的である．許容性は経済学ではパレート最適と呼ばれる概念に相当する．

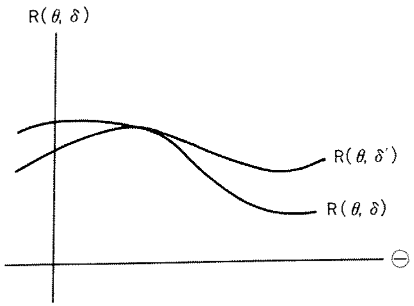


図１ $\delta>\delta'$ の例

許容性の概念は最適性の基準としては最も弱い概念で，許容的な決定方式が必ずしも合理的な決定方式とはならない．たとえば，観測値を無視して $\hat{\theta}_0(X)\equiv\theta_0$ と常に特定の θ_0 を推定値とする推定方式を考える．$\hat{\theta}_0$ は明らかに不合理な推定量であるが，たまたま $\theta=\theta_0$ ならばリスクは $R(\theta_0,\hat{\theta}_0)=0$ となる．任意の他の推定量 $\hat{\theta}(X)$ で $R(\theta_0,\hat{\theta})=0$ となるものは θ_0 の分布の

もとでのすべての可能な X の値について $\hat{\theta}(X)=\theta_0=\hat{\theta}_0$ とならなければならないから，θ_0 のもとでの可能な X の範囲が標本空間の全域となるならば，$\hat{\theta}_0$ は許容的となることが分かる．ここで標本空間とはさまざまな θ のもとで X のとり得る値をすべて含んだ集合をいう．このように許容性は弱い概念であるが，逆にその分ある決定方式が非許容的であるとすれば，その決定方式を用いる根拠はうすいものとなる．このことの有名な例はスタイン推定量の場合である．

許容性と密接な関連を持つ概念として完備類の概念がある．決定方式の集合 C が完備類（complete class）であるとは，すべての $\delta'\notin C$ に対してある $\delta\in C$ が存在して $\delta>\delta'$ となることである．完備類の定義から，完備類の外の任意の決定方式に対し完備類の中にそれよりよい決定方式が必ずあるわけであるから，完備類の外のどの決定方式も明らかに非許容的である．これを裏返して言えば許容的な決定方式はすべて完備類に含まれなければならない．すなわち次の補題が成り立つ．

補題１　任意の完備類はあらゆる許容的な決定方式を含む．

ところで完備類は小さければ小さいほどよいから完備類の中で最小のものが存在すれば便利である．完備類で集合の包含の意味で最小のものが存在すれば，それを最小完備類（minimal complete class）という．最小完備類と許容性の関係は次の補題で示される．

補題２　最小完備類はもし存在すれば許容的な決定方式の集合と一致する．

証明はそれほど難しくはないので読者にまかせる．最小完備類の存在については以下で述べる．統計的決定理論の文献では“完備類定理”（complete class theorem）としてどのような決定方式のクラスが完備類となるかを述べた定理が多く見られる．すなわち「以下の条件を満たす決定方式のクラスは完備類をなす」という形の定理である．許容性との関連で見ると上の補題からわかるように，完備類定理は許容性の必要条件を述べる定理である．こ

れに対して許容性の十分条件を述べた結果はより少ない．許容性の十分条件を求めることは必要条件を求めることより困難なようである．

以上で許容性についての基本的な概念を説明したが，次に統計的決定理論の枠組の中でのベイズ法について述べよう．前にもふれたように統計的決定理論では統計的推測の問題を零和2人ゲームとしてとらえる．プレーヤーは「自然」と統計家であり，自然は母数 θ を選び統計家は決定方式 δ を選ぶ．特定の母数 θ と特定の決定方式 δ が選ばれた時，統計家から自然への支払いが $R(\theta,\delta)$ で決まると考えよう．ところでよく知られているように，零和2人ゲームではプレーヤーがある確率分布に従ってランダムに自分の手を選ぶ“混合戦略”を考える必要が生じる．ここで自然の側の混合戦略を考えよう．自然が分布 τ に従ってランダムに母数 θ を選んでくるものとする．τ はベイズ統計学では事前分布である．この時統計家が戦略 δ をとるとすれば，τ までを考慮した平均損失は

$$r(\tau,\delta)=E_\tau(R(\theta,\delta))=E_\tau[E_\theta(L(\theta,\delta))]$$

と表される．$r(\tau,\delta)$ を事前分布 τ のもとでの決定方式 δ のベイズ・リスクという．

ところでいま統計家が有利な立場にあり事前分布 τ を知っているとしよう．このとき統計家は特定のどの θ が選ばれるかは知ることはできないが，どんな θ が選ばれやすいかは知っているわけであるから，それに応じて最適な戦略を考えることができるはずである．そこで事前分布 τ に対して最適な戦略 δ_τ を

$$r(\tau,\delta_\tau)=\min_\delta r(\tau,\delta)$$

と定義する．δ_τ を事前分布 τ に関するベイズ決定方式と呼ぶ．ベイズ統計学でよく知られているように，ベイズ決定方式はパラメータの事後分布に関してリスクを最小とすることによって定まる．これは推定の問題では事後分布の平均を求めることに帰着する．また，検定の問題では帰無仮説と対立仮説それぞれの事後確率を求め事後確率のより大きな仮説をとればよい．

このようにベイズ決定方式は統計的決定理論の観点からも自然な決定方式として動機づけられる．しかしそれ以上に重要なことは上に述べた許容性とベイズ決定方式は非常に密接な関係があるということである．それは「ベイズ及びその極限」と「許容的であること」とがほぼ同値であるという形で述べることができる．これは統計的決定理論における基本的な認識の1つである．このことを示すためにまず以下の簡単で重要な定理を証明しよう．

定理1　ある事前分布 τ についてのベイズ決定方式 δ_τ が一意的に定まれば δ_τ は許容的である．

証明　もしある δ' が $\delta'\geq\delta_\tau$ であったとする．そうするとすべての θ について $R(\theta,\delta')\leq R(\theta,\delta_\tau)$ であるから

$$r(\tau,\delta')=E_\tau(R(\theta,\delta'))\leq E_\tau(R(\theta,\delta_\tau))=r(\tau,\delta_\tau)$$

となる．しかし δ_τ はベイズ決定方式であるからここで強い不等式が成り立つことはない．したがって δ' のベイズ・リスクと δ_τ のベイズ・リスクは一致し一意性の仮定より $\delta'=\delta_\tau$ となる．このことは δ_τ が許容的であることを示している．

∎

このように証明は容易である．簡単に言えば，ベイズ決定方式は特定の事前分布に限れば最適な決定方式であるから，それより一様にリスクの小さい決定方式はないというわけである．通常ベイズ決定方式は事後分布より一意的に決まるからベイズ決定方式は許容的と考えてよい．

ところで以上の逆が成り立つかというと逆は一般には成り立たない．許容的な決定方式でもベイズ決定方式とは限らない．しかしその場合でも許容的な決定方式は一般的にベイズ決定方式の極限となることが示される．最も簡単な例は X が平均 μ，分散1の正規分布 $N(\mu,1)$ に従っている場合に μ の推定量として X 自身を用いた場合である．X は許容的であるがベイズ決定方式ではなくベイズ決定方式の極限である．このことの証明は次節で与える．ところで簡単な場合には実はベイズ決定方式の極限を考える必要がなく，定理1の逆が成立する場合もある．それは未知パラメータ θ のとり得る値が有限個である場合である．ここではこの場合について述べよう．

パラメータ θ のとり得る値の集合をパラメータ空間といい Θ で表す．ここでは Θ が有限集合 $\Theta=(\theta_1,\cdots,\theta_m)$ の場合を考える．Θ が有限の場合には“リスク集合”の考え方を用いると議論の見通しがよくなる．決定方式 δ に

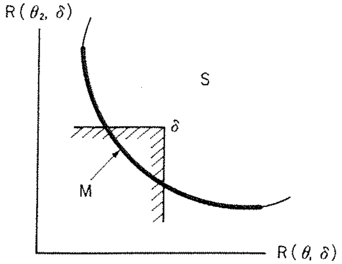


図 2

ついてすべてのパラメータの点でのリスクをベクトルにまとめて，$S_\delta=(R(\theta_1,\delta),\cdots,R(\theta_m,\delta))$ とおく．S_δ をすべての δ について考えたもの $S=\{S_\delta\}$ をリスク集合という．リスク集合 S は m 次元空間 R^m の部分集合と考える．最も簡単な場合は $m=2$ の場合（たとえば単純帰無仮説対単純対立仮説の検定問題）で，リスク集合が図示できるので大変都合がよい．ところでリスク集合は凸集合となる．これを示すには 2 つの決定方式 δ_1 と δ_2 に対して，統計家が確率 α で δ_1 を選び，確率 1-α で δ_2 を選ぶような混合戦略 $\delta_\alpha=\alpha\delta_1+(1-\alpha)\delta_2$ を考える．この時すべての θ について，$R(\theta,\delta_\alpha)=\alpha R(\theta,\delta_1)+(1-\alpha)R(\theta,\delta_2)$ となるから S_{δ_α} は S_{δ_1} と S_{δ_2} を結ぶ線分の上にくる．したがって図 2 にあるように S は凸集合である．ここで図 2 において許容的な決定方式がどこにくるかを考えよう．任意の $\delta\in S$ に対して δ よりよい決定方式は δ の左下に来ることがわかる．したがって許容的な決定方式の集合は S の原点のほうに張り出した辺（太線で示した部分 M）と

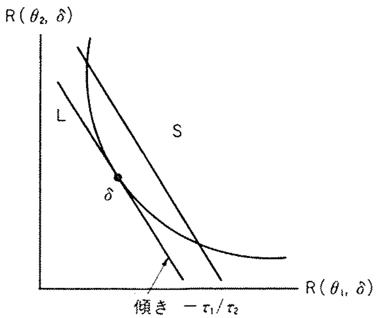


図 3

なっていることが分かる．またこの許容的な決定方式の集合が最小完備類であることも分かる．$m>2$ でも次元が高くなるだけでまったく同様の議論ができることは明らかであろう．リスク集合の考え方を用いると次の定理が証明できる．

定理 2　Θ が有限集合でリスク集合 S が閉ならば許容的な決定方式はある事前分布に対するベイズ決定方式である．

証明　m が一般でも証明の本質は変わらないので $m=2$ の場合に証明する．$\delta\in M$ を許容的とする（図 3）．δ における S の接線を L とおき L の方程式を

$$c=\tau_1R(\theta_1,\delta)+\tau_2R(\theta_2,\delta)$$

とおく．図より明らかなように接線は右下がりであるから $\tau_1\geq 0,\ \tau_2\geq 0$ でありともに 0 となることはない．そこで事前分布 τ を

$$\tau(\theta_i)=\Pr(\theta_i)=\frac{\tau_i}{\tau_1+\tau_2},\qquad i=1,2$$

とおく．事前分布 τ が与えられた時，ベイズ・リスクが一定となる決定方式は図 2 において L と平行な直線の上にある．これらの平行な直線の中で左下のものほどベイズ・リスクが小さいから S の中で最初に与えた δ がベイズ・リスクを最小とし，したがって δ が τ に関するベイズ決定方式 $\delta=\delta_\tau$ となることが分かる．

■

以上のように Θ が有限集合の場合には「ベイズ決定方式」と「許容的な決定方式」が一意性を除いて同値であることが示された．ところで定理では S が閉集合であると仮定した．これは技術的ではあるが本質的な仮定である．いまもし S が開集合で M が S に属さないとすると，許容的な決定方式は存在せず最小完備類も存在しないことが分かる．すなわち S の内部から出発して左下の方向へ進んで必ずよりよい決定方式を見つけることができるが，その限界が存在しないことになる．このことの例をあげよう．

例 1　$U(\theta,\theta+1)$ を θ と $\theta+1$ の間の一様分布とし，θ としては 2 点のみ $\theta_1=0,\ \theta_2=1$ とする．$U(\theta_i,\theta_i+1),\ i=1,2$ から 1 個の観測値 X を観測して

θ を“推定”する問題を考える．この場合もちろん X から θ_1 か θ_2 のいずれが正しいパラメータであるかわかってしまうから通常の推定の問題としては誤差なく推定できるのであるが，ここでは次のような損失関数を考える．

$$L(\theta, d) = (\theta-d)^2 \quad \text{if} \quad \theta \neq d$$
$$= 1 \quad \text{if} \quad \theta = d$$

この損失関数ではいくらでも近い値で推定するのが望ましいが，しかしぴったりと当ててしまうのは損であることになる．こうした場合にはどのように推定したらよいか困るであろう．実際容易に分かるようにこの場合はリスク集合 S は $R(\theta_1, \delta)$-軸および $R(\theta_2, \delta)$-軸を含まない第1象限（開集合）となる．したがって原点の方向にいくらでも近づくことができるが原点には到達できないことになる．

この例からもわかるようにリスク集合が開集合となるのは病的な例である．リスク集合が閉となるために必要な正則条件については次節で述べる．

3. ベイズの極限と完備類定理

前節では統計的決定理論について予備知識を前提としないでベイズ決定方式と許容的な決定方式の関係について基本的な事項を説明した．特にパラメータ空間が有限集合の時は許容性とベイズが一意性を除いて同値であることが分かった．本節では Θ が無限集合の場合について基本的な結果を整理するが，そのためある程度の統計的および数学的知識を前提として話を進める．

まず前節で問題となった S が閉であるという条件についての正則条件を整理しよう．前節ではリスク集合を Θ が有限の場合について考えたが，位相をいわゆる弱位相で考えれば Θ を有限と限る必要はない．決定方式 δ の集合を決定空間と呼び D と書く．決定方式の列 $\{\delta_n\}$ が δ に弱収束するとはすべての $\theta \in \Theta$ について

$$\lim_{n\to\infty} R(\theta, \delta_n) = R(\theta, \delta)$$

となることを言う．明らかに Θ が有限集合の場合には D が弱位相に関して閉であることと S が R^m で閉であることは同値である．いま真のパラメータが θ の時の X の密度関数を p_θ とおく．p_θ は X が連続的な場合の通常の密度関数でもよいし，離散的な場合の確率関数でもよい．より一般に p_θ は標本空間 $\mathcal{X}$ 上のある共通なシグマ有限な測度に関するラドン・ニコディムの密度関数でよい．このような場合，分布族は dominated family であるという．dominated family であることを仮定すれば，検定問題では無条件で D は弱位相に関して閉であることが示される（Lehmann[1986] の付録の weak compactness theorem を参照）．次に推定問題については損失関数に次の正則条件を仮定する．

条件 1

(i) $L(\theta, d)$ は d に関して連続である．

(ii) すべての θ について $\lim_{|d|\to\infty} L(\theta, d) = \infty$

推定問題では条件1のもとで D は弱位相について閉である．D が弱位相に関して閉じていれば最小完備類は存在し許容的な決定方式の集合に一致する．これらの証明については Brown[1986] の Appendix を参照されたい．

Θ が無限集合の場合にはベイズ決定方式のほかにベイズ決定方式の極限を考えなければならない．この問題を考えるために例として前節でもふれた正規分布の平均の推定の例をとりあげよう．X が $N(\mu, 1)$ に従っていたとする．まず次を示そう．

補題 3 X はベイズ推定量ではない．

証明 X が μ の事前分布 τ に対するベイズ推定量であるとする．すなわち $E_\tau(\mu|X) = X$ と仮定する．まず

$$E(X^2) = E((X-\mu)+\mu)^2 = E(X-\mu)^2 + 2E[(X-\mu)\mu] + E(\mu^2)$$
$$= 1 + E(\mu^2)$$

である．しかしながら $E_\tau(\mu|X) = X$ の関係を用いて X と μ の役割を入れ替えて考えるとまったく同様に

$$E(\mu^2) = 1 + E(X^2)$$

を得る．これは矛盾である．∎

次に X の許容性であるが X が許容的であることは“Blyth の方法”と呼ばれる方法で示すことができる．

補題 4　μ の推定量として X は許容的である．

証明　$\delta_0(X)=X$ が許容的でないとすると他の推定量 $\delta^*(X)$ があって

$$R(\theta, \delta^*) \le R(\theta, \delta_0) = 1 \qquad \forall\theta$$

$$R(\theta_0, \delta^*) < 1 \qquad \exists\theta_0$$

となる．有界収束定理を用いると任意の δ に対して $R(\theta, \delta)$ が θ について連続となることが示せるので，ある $\epsilon>0$ について

$$R(\theta, \delta^*) < 1-\epsilon \qquad \theta\in(\theta_0-\epsilon, \theta_0+\epsilon)$$

となる．ここで“事前測度”として

$$\tau_n(d\theta) = \sqrt{n}\tau_n^*(d\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2n}\right)d\theta$$

とおく．τ_n^* は $N(0, n)$ の密度関数であり τ_n は全積分が $\sqrt{n}$ となることを除いて正規分布の密度関数と一致する．この時

$$r(\tau_n, \delta_0)-r(\tau_n, \delta^*) = \int_{-\infty}^{\infty}(R(\theta, \delta_0)-R(\theta, \delta^*)\tau_n(d\theta)$$

$$\ge \epsilon\int_{\theta_0-\epsilon}^{\theta_0+\epsilon}\tau_n(d\theta) \to \sqrt{2/\pi}\epsilon^2 \qquad (n\to\infty) \qquad (3)$$

となる．ところで事前分布 τ_n^* についてのベイズ推定量は

$$E_{\tau_n^*}(\mu|X) = \frac{n}{n+1}X = \delta_{\tau_n^*}(X)$$

でありそのベイズリスクは容易にわかるように $r(\tau_n^*, \delta_{\tau_n^*})=n/(n+1)$ である．したがって

$$r(\tau_n, \delta^*) = \sqrt{n}r(\tau_n^*, \delta^*) \ge \sqrt{n}\frac{n}{n+1}$$

となる．これより

$$r(\tau_n, \delta_0)-r(\tau_n, \delta^*) \le \sqrt{n}\left(1-\frac{n}{n+1}\right)\to 0 \qquad (n\to\infty) \qquad (4)$$

となる．しかし(3)式と(4)式は矛盾である．∎

Blyth's の方法は単なる技術的な証明というばかりでなくベイズの極限について多くを示唆している．まず，事前分布 τ_n^* についてのベイズ推定量 $[n/(n+1)]X$ はすべての X について $n\to\infty$ の時 $X=\delta_0(X)$ に収束している．したがってこの場合は確かに $X=\delta_0(X)$ はベイズの極限であり，その意味を各 X での収束と考えることができる．以下に述べるように一般にベイズの極限をこの各点収束の意味にとることができ非常に有用である．

一方事前分布の測度としての収束を見てみると確率測度の範囲では話が終らず τ_n のように全測度が 1 とならない測度を考えなければならないことがわかる．τ_n が測度の意味で収束する先はルベーグ測度（を $\sqrt{2\pi}$ で割ったもの）になっている．そして δ_0 は事前測度としてルベーグ測度を用いたものについての一般化ベイズ推定量となっていることに注意しよう．

以上を念頭において，ここでは Brown[1986] よりベイズおよびベイズの極限が完備類をなすという意味での完備類定理を紹介しよう．まず T_0 を Θ の中の有限個の点のみに確率を与える離散事前分布の集合とし，D_0 を T_0 に属する事前分布に対するベイズ決定方式の集合とする．そして $\bar{D}_0$ を D_0 の概収束の意味での閉包とする．すなわち

$$\delta\in\bar{D}_0 \Longleftrightarrow \exists\{\delta_n\in D_0\}, \quad \delta(X) = \lim_{n\to\infty}\delta_n(X), \quad a.e.\ \mu$$

である．ただし μ は dominated family の基準測度である．検定については次の正則条件を仮定する．

条件 2

(iii)　非確率化検定が完備類をなす．

(iv)　同一のリスクを持つ検定は一致する．すなわち

$$R(\theta, \delta) = R(\theta, \delta'), \forall\theta \Longrightarrow \delta(x) = \delta'(x), a.e.\ \mu$$

この時次の定理が成り立つことが示される．

定理 3（完備類定理）　分布族が dominated family をなすとする．検定問題において正則条件 2 を仮定すれば $\bar{D}_0$ は完備類をなす．

次に推定問題については次の正則条件を考える．

条件 3

(v) 分布の台（サポート）が共通である．すなわち $\forall_{\theta\in}\Theta, \forall_{x\in}\mathcal{X}, p_\theta(x) > 0$

(vi) 任意の θ について $L(\theta, d)$ は d に関して強凸である．

この時次の定理が成り立つことが示される．

定理 4（完備類定理）（Brown[1986], Theorem 4 A. 12）分布族が dominated family をなすとする．推定問題において正則条件 1，3 を仮定すれば $\bar{D}_0$ は完備類をなす．

定理 3，4 の証明の概略は次のようにすればよい．δ を任意の決定方式とする．まず Θ が可算集合 $\{\theta_1, \theta_2, \cdots\}$ であるとしよう．ここで Θ_k を有限集合 $\{\theta_1, \cdots, \theta_k\}$ として統計的決定問題を Θ_k に制限して考える．制限された問題において許容的な決定方式は最小完備類をなすから，ある許容的な δ_k が存在して $\delta_k > \delta$ となる．ところで定理 2 により δ_k は Θ_k のみにサポートを持つある事前確率 τ_k に関するベイズ決定方式となっている．ここで k のある部分列 $\{k_j\}$ をとれば δ_{k_j} がある δ^* に概収束しもとの決定問題において $\delta^* > \delta$ となることが証明できる．これで定理が示されたことになる．また Θ が非可算集合の時は有向集合に関する極限の概念を用いて同様に議論すればよい．

この完備類定理の応用については次節で述べる．また $\bar{D}_0$ は概収束の意味での極限で定義したが標本空間において測度 0 の領域を無視すれば各点収束の意味での極限と考えてよいことに注意しよう．

以上は非常に一般的な完備類定理であるが有限の点の上のみにある事前分布とはいってもベイズ決定方式を明示的に書き下しその極限を求めることは困難なことが多い．そこで次にベイズの極限のもう 1 つの側面つまり事前分布の収束について触れておく．τ_n を有限事前測度の列とし δ_{τ_n}を τ_n に対するベイズ決定方式とする．定理 3，4 にあるように $n \to \infty$ とする時の δ_{τ_n} の極限 δ^* が完備類の元となるが，この時事前測度 τ_n もある測度 τ^* に収束し δ^* が τ^* に関する一般化ベイズ決定方式になっていると期待される．このような方針にもとづいていろいろな状況のもとで「一般化ベイズ決定方式が完備類をなす」という形の完備類定理が証明されている．一般化ベイズ決定方式は全測度が有限でない事前測度分布にたいしても有限な事前確率分布の場合と同様に事後分布に関する積分で定義される．このことを用いてさらに一般化ベイズ決定方式の満たすべき必要条件を求めることができ，許容性の必要条件が具体的な形で与えられることになる．このことの例は次節の定理 6 で論じる．

4. 許容性に関するいろいろな結果

前節では Θ が無限集合である場合の一般論について基本的な結果を整理し，またその困難も指摘した．この節ではより個別的な問題について許容性に関する最近までの結果で目につくものをいくつか取り上げる．以下で取り上げるのはおもに指数型分布族の検定・推定に関する結果でありその意味では個別的といってもかなり一般性のある結果である．より詳しくは Berger[1985] およびそこにあげられている文献を参照されたい．

まず指数型分布族の自然母数の検定について述べる．指数型分布族の自然母数の検定については Birnbaum[1955] の定理以来いくつかの一般的な定理が知られている．まず Birnbaum の定理を述べよう．ただし定理の条件は簡単な場合に限定してある．

定理 5　連続な確率変数の密度関数 $p_\theta(x)$ が m 次元の実母数 θ を自然母数とする指数型分布族をなすとする．母数空間 Θ は R^m 全域であるとする．ここで単純帰無仮説の検定問題

$$H : \theta = \theta_0 \quad vs. K : \theta \neq \theta_0$$

を考える．この時凸な受容域を持つ検定が最小完備類をなす．

このようにこの定理の設定のもとでは検定が許容的であるための必要十分条件はその検定が凸な受容域を持つということであるが，ここでは許容性の必

要条件を前節の定理3を用いて示そう．事前分布 τ が有限個の点 $\theta_0, \theta_1, \cdots, \theta_N$ のそれぞれに確率 $\tau_0, \cdots, \tau_N$ を持つとする．この時ベイズ検定は

$$\text{帰無仮説を受容} \iff \frac{\sum_{i=1}^{N} \tau_i p_{\theta_i}}{\tau_0 p_{\theta_0}} < 1$$

$$\iff \sum_{i=1}^{N} \tau_i \exp\{(\theta_i - \theta_0)' x\} < c$$

である．ところで指数関数は凸関数であるからその凸結合 $\sum_{i=1}^{N} \tau_i \exp\{(\theta_i - \theta_0)'x\}$ も凸関数である．したがって事前分布 τ に対するベイズ検定の受容域は凸集合とならなければならない．定理3によれば許容的な検定はこれらの検定の X についての各点収束の極限とならなければならないからやはり凸な受容域を持たなければならない．

このように定理3を用いれば Birnbaum の定理の必要条件は簡単に証明できる．定理の十分条件を示すには Stein の定理と呼ばれる定理を用いる必要がある．ここでは省略するが詳しくは Lehmann[1986], Theorem 6.8 を参照されたい．

Birnbaum の定理の応用例として2変量正規分布 $N((\mu_1, \mu_2)', I_2)$ の検定問題を考えよう．帰無仮説は平均ベクトルが原点にあるという仮説とする

$$H : \mu_1 = 0, \mu_2 = 0$$

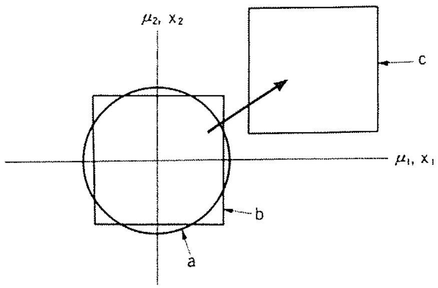


図4

この場合よく用いられるはカイ2乗検定であるがカイ2乗検定の受容域は原点を中心とする円である（図4の a）．ところでこの問題の場合カイ2乗検定のほかにも例えば

$$\max(|X_1|, |X_2|) > c \Longrightarrow \text{棄却}$$

といった検定も考えられる．受容域は図4の b である．この受容域も凸であるから検定は許容的である．ところで凸な受容域はどれでも許容的であるから図4の c のような受容域もやはり許容的でなければならない．これは不合理な感じがするがこの例も許容性が非常に弱い概念であることを示している．より具体的には b と c の受容域を比べると対立仮説が矢印の方向に無限に行ったときには確かに c の方が b より検出力が高いから c が b より一様にリスクが劣るとは言えないわけである．

この Birnbaum の定理は複合帰無仮説，傾向のある対立仮説，多変量解析の検定問題などの場合に一般化されている．複合帰無仮説については Matthes and Truax[1967]，傾向のある対立仮説については Eaton[1970], Marden[1982]，多変量解析特に多変量分散分析の検定については Stein[1956], Schwartz[1967], Anderson and Takemura[1982], Marden[1983] を参照されたい．

次に推定問題を取り上げる．まず連続な指数型分布族の自然母数の推定に関する Berger-Srinivasan[1978] の定理を紹介しよう．

定理6　連続な確率変数 X の密度関数が指数型分布族をなすとし平均2乗誤差のもとで自然母数ベクトル $\theta \in \Theta$ の推定問題を考える．Θ が閉集合ならば一般化ベイズ推定量は完備類をなす．

Θ が閉でない時には Θ の閉包を考えればよい．この定理は大変明快な形で述べられているところが便利である．

この定理の1つの応用としてスタイン推定量を取り上げよう．多変量正規分布の平均ベクトルの推定における James-Stein 推定量はよく知られているように

$$\delta^{JS}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$$

である．ただし X は多変量正規分布 $N_p(\mu, I)$ に従う確率ベクトルである．次元 p が $p \geq 3$ の時 $\delta^{JS}(X)$ は S を改善するが，一方 $\delta^{JS}(X)$ 自身も “positive part estimator”

$$\delta^{JS+}=\max\left(1-\frac{p-2}{\|X\|^2},0\right)X$$

によって改良される．ところで定理6によれば許容的な推定量は一般化ベイズ推定量であり事後分布に関する積分で表されなければならない．しかし正規分布の密度関数から分かるように積分で表される推定量は X の滑らかな関数である．δ^{JS+} は X の滑らかな関数ではないから許容的な推定量ではない．ちなみに δ^{JS+} を改良する推定量は見つかっていない．

指数型分布族の期待値母数の推定については Brown[1986] の4章に定理が与えられているが定理6ほど明快ではないので省略する．ここでは期待値母数の一般化ベイズ推定量が許容的であるための十分条件を与える Brown and Hwang[1982] の定理を紹介する．確率ベクトル X の密度関数 $p_\theta(x)$ が指数型分布族をなすとし平均二乗誤差のもとで期待値パラメータ $\psi(\theta)=E_\theta(X)$ の推定問題を考える．$\tau(\theta)$ を自然母数空間 $\Theta\subset R^p$ 上の微分可能な事前密度関数とする．ただし $\tau(\theta)$ の全積分が1であることは仮定しない．また $\nabla\tau(\theta)=(\partial/\partial\theta_1,\cdots,\partial/\partial\theta_p)\tau(\theta)$ を $\tau(\theta)$ の勾配 (gradient) とする．

定理7　以下の3つの条件のもとで $\tau(\theta)$ にもとづく $\psi(\theta)$ の一般化ベイズ推定量は許容的である．

(i)　すべての x に対して $\int|\nabla\tau(\theta)|\,p_\theta(x)\,d\theta<\infty$

(ii)　$\displaystyle\int_{\|\theta\|\ge 2}\frac{\tau(\theta)}{(\|\theta\|\log(\|\theta\|))^2}d\theta<\infty$

(iii)　$\displaystyle\int\frac{\|\nabla\tau(\theta)\|^2}{\tau(\theta)}d\theta<\infty$

証明は Blyth の方法と同様の考え方を用いればよく比較的容易である．この定理の1つの応用として再びスタイン問題をとりあげよう．$\tau(\theta)\equiv 1$ とおけば一般化ベイズ推定量として $\delta_\tau(X)=X$ を得る．$\nabla\tau(\theta)=0$ であるから(i)および(iii)は明らかに成立している．(ii)に関する条件は $\int_2^\infty r^{p-1}/(r\log r)^2dr<\infty$ となるが $p\le 2$ ならばこの条件が成立する．したがって $p\le 2$ ならば $\delta_\tau(X)=X$ は許容的である．この結果は正規分布に関してはよく

知られているが，ここでは正規分布以外の指数型分布族においても同様な結果が成り立つことがわかる．

参 考 文 献

Anderson, T. W. and A. Takemura, 1982 A new proof of admissibility of tests in multivariate analysis, *Journal of Multivariate Analysis*, 12, 457-468.

Birnbaum, A., 1955 Characterization of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests, *Ann. Math. Statist*., 26, 21-36.

Berger, J. O., 1985 *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer.

Berger, J. O. and C. Srinivasan, 1978 Generalized Bayes estimators in multivariate problems, *Ann. Statist*., 6, 783-801.

Blackwell, D. and M. A. Girshick, 1954 *Theory of Games and Statistical Decisions*, John Wiley.

Brown, L. D., 1986 *Fundamentals of Statistical Exponential Families*, Lecture notes-monograph series Vol. 9, Institute of mathematical statistics.

Brown, L. D. and J. T. Hwang, 1982 A unified admissibility proof, In *Statistical Decision Theory and Related Topics* III, S. S. Gupta nad J. Berger (eds.), Academic Press, New York.

Eaton, M. L., 1970 A complete class theorem for multidimensional one-sided alternatives, *Ann. Math. Statist*., 41, 1884-1888.

Ferguson, T. S., 1967 *Mathematical Statistics, A Decision Theoretic Approach*, Academic Press.

Lehmann, E. L., 1986 *Testing Statistical Hypotheses*, John Wiley.

Lehmann, E L., 1983 *Theory of Point Estimation*, John Wiley.

Marden, J. I., 1982 Minimal complete classes of tests of hypotheses with multivariate one-sided alternatives, *Ann. Statist*., 10. 962-970.

Marden, J. I., 1983 Admissibility of invariant tests in the general multivar-

iate analysis of variance problem, *Ann. Statist.*, 11, 1086-1099.
Matthes, T. K. and D. R. Truax, 1967 Tests of composite hypotheses for the multivariate exponential family, *Ann. Math. Statist.*, 38, 681-697.
鍋谷清治, 1978 『数理統計学』(現代の数学) 33巻, 共立出版.
Schwartz, R., 1967 Admissible tests in multivariate analysis of variance, *Ann. Math. Statist.*, 38, 698-710.
Stein, C., 1956 The admissibility of Hotelling's T^2-test, *Ann. Math. Statist.*, 27, 616-623.
Wald, A., 1950 *Statistical Decision Functions*, John Wiley.

第5章

非ベイズの立場から見たベイズ統計学

竹内 啓

1. 基本的立場

私はSavage流の主観確率論の立場は取らないので，統計学者としては"Bayesian"ではない．私の"Non-Bayesian"としての統計的推測理論に関する基本的立場は，簡単にいえば次の通りである．

1° 統計的推測の前提は，データに対して設定された確率的モデルである．それはデータの構造に対して想定された一連の仮定を表し，データの現実の構造を近似するものであると同時に，それを理想化したものである．確率的モデルにおいて「確率」とは頻度と結びついたストカスティックな構造を表すと考える．さらにデータを通じてわれわれが知りたいと思う情報は，データの確率分布を決定する未知の母数に反映されているものと考える．

2° 統計的推測理論の目的は，データから未知の母数に関して何らかの判断，あるいは結論を適切に引き出す方式を見出すことである．母数が未知であり，またデータが確率的に変動することを前提とする以上，一般に，つねに「正しい」あるいは「望ましい」結論を導くことができるような推測方式は存在しない．そこで仮定された確率モデルの下で，確率的にあるいは平均的に望ましい性質を持つ推測方式が望ましいものであると考える．

3° 2°のように考えても，なお一般に望ましい推測方式は未知の母数の値に依存することになる．そこで推測方式に関して2種類の基準を考え，まず第1の基準をみたすものの中で，第2の基準により，できる限り精密な結論を導くものを考える．第1の基準を妥当性の基準，第2のものを効率性の基

準という．検定の水準，推定量の不偏性，信頼区間の信頼係数などは妥当性の基準であり，検定の検出力，推定量の分散，信頼区間の長さなどは効率性の基準を表す．

4° 推測方式の望ましさは，データの確率分布から導かれるその確率的性質によって定められるので，そこにデータの具体的数値を代入して得られる具体的な結論には直接にはあてはめられないが，しかし確率的に望ましい推測方式から得られた結論は，やはり望ましいであろうと考える．

5° ただし，データの特定の実現値が，確率分布の下での平均的な状況から乖離していることが明らかであるような場合には，推測に当ってそのことを考慮に入れなければならない．このことから条件付推測（conditional inference）の考え方が導かれる．

以上のように考える理由については，よりくわしく論ずることもできるが，ここではその紙幅もないので立ち入らないことにしたい．

Non-Bayesian の観点からベイズ法を見れば，それは事前分布を設定して事後分布を計算することにより，2°～5° の意味で，望ましい推測方式を導く操作的方法であると考えることができる．その場合，事前分布が主観確率を表すか，なんらかの意味で客観的なものか，あるいは単なる相対的ウエイトであるかはどちらでもよいことである．そうして純粋な Bayesian の立場とは違って，事後分布だけでなく，母数のいろいろな値に対応する推測方式の確率分布を検討することが Non-Bayesian の視点からは必要になる．

2. ベイズ法の有用性

Non-Bayesian の立場からベイズ的方法を考えるとき，2 つの方向が考えられる．1 つは Non-Bayes 的枠組の中でベイズ法を用いることによって積極的な結果を得ることであり，もう 1 つは Non-Bayesian の観点からベイズ法を検討することである．

第 1 の方向からは，次のような点がとり上げられる．

1° Wald の完備類定理（complete class theorem）．

ベイズ解およびその（いろいろな意味での）極限が，完備類，あるいは本質的完備類（essentially complete class）を形成すること，したがってまた，許容解（admissible solution）はベイズ解，またはその極限に限ること，したがってベイズ解およびその極限以外は“よい”方式として取り上げる必要はないことになる．この点について詳しくは第 4 章（竹村稿）を参照されたい．

この定理の応用として，母数空間が複雑な構造を持つ場合の推測問題において，適切な推測方式を選択するための基準を，ベイズ解の構造から与えることができる．

例えば，$X_{ij}(i=1, \cdots, p; j=1, \cdots, n)$ が互いに独立に平均 μ_i，分散 σ^2 の正規分布に従うとき，母数の順序が

$$\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_p$$

という形で与えられているという条件の下での検定，推定を考える．

特にこの条件の下で，仮説

$$\mathrm{H}_0 \quad \mu_1 = \mu_2 = \cdots = \mu_p$$

を対立仮説に対して検定する問題は「順序のある対立仮説に対する検定問題」と呼ばれ，現実にもしばしば生ずる．これに対して $\bar{X}_i=\sum_j X_{ij}/n$ とおくとき，上記の定理から検定関数が許容的であるための十分条件が導かれ，それがまた（Stein の定理により）必要条件でもあることが示されるので，完全類が定められる．結果だけのべれば，σ^2 が既知のときには

$$Z_j = \sqrt{\frac{p}{j(p-j)}}(\bar{X}_1+\cdots+\bar{X}_j-j\bar{X})$$

$$j = 1, \cdots, p-1, \qquad \bar{X} = \sum \bar{X}_j/p$$

とおくとき，$Z_j(j=1, \cdots, p)$ の単調非減少関数として表され，かつ受容域が凸集合になるような検定関数の全体が完備類をつくる．

2° 母数空間や標本抽出方式が複雑な構造を持つ場合に，適切と見られる推測方式を導くこと．

たとえば $X_1, \cdots, X_n, \cdots$ が互いに独立に，密度関数 $f(x, \theta)$ を持つ分布に従う場合，単純仮説 $\theta=\theta_0$ を単純対立仮説 $\theta=\theta_1$ に対して検定する逐次検定問題を考える．2 種類の誤まりの損失，および標本観測値 1 個当りの観測コ

ストを一定にすれば，ベイズ方式は次のようなワルドの逐次確率比検定(sequential probability ratio test, 略して SPRT) 方式に一致する．すなわち 2 つの定数 A, B を与えて

$$A < \prod_{i=1}^{n} \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)} < B$$

のとき観測を続け，

$$\prod_{i=1}^{n} \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)} \leqq A$$

ならば仮説を受容し，

$$\prod_{i=1}^{n} \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)} \geqq B$$

ならば仮説を棄てることにすればよい．このようにして「望ましい」検定方式のタイプが定められるので，Non-Bayesian の観点からは，このような方式の下での 2 種の誤りそれぞれの確率と，平均観測個数とを考慮して，定数 A, B を適当に定めればよい．

別の例として次のような問題を考えよう．$X_1, X_2, \cdots, X_n, \cdots$ をベルヌーイ試行系列とする．すなわち X_i は互いに独立で

$$\Pr\{X_i = 1\} = 1 - \Pr\{x_i = 0\} = p$$

とする．そうして $X_i=1$ のとき試行は「成功」で 1 単位の「報酬」が得られるものとする．他方，1 回の試行にはコストが c かかるものとする．したがって 1 回の試行についてそれが「成功」ならば $1-c$ の利益が得られ「失敗」ならば c の損失を被ることとなる．このとき試行を続けるべきか，止めるべきかを，これまでの試行の結果にもとづいて定めることを問題とする．

この問題において p の事前分布として，α, β を母数とする事前分布を想定するのが便利である．n 回の試行ののち「成功」が x 回であったとすると，事後分布は母数が $\alpha+x$，および $\beta+n-x$ のやはりベータ分布となるからである（ベータ分布は 2 項分布のいわゆる「自然共役分布」である）．

いま最大 N 回までの試行が許されるとした場合，事前分布の母数が α, β であるとき，期待利益の最大値を $u_N(\alpha, \beta)$ と表せば，次のような関係が成り立つことがわかる．

$$u_N(\alpha, \beta) = \max\left[0, \frac{\alpha}{\alpha+\beta}\{u_{N-1}(\alpha+1, \beta)+1\} + \frac{\beta}{\alpha+\beta} u_{N-1}(\alpha, \beta+1) - c\right]$$

$$u_0(\alpha, \beta) = 0$$

となる．そうして $u_N(\alpha, \beta)>0$ ならば，最初の試行を行い，$u_N(\alpha, \beta)=0$ ならば，最初から試行しないのがよいということになる．そうして n 回の試行の後には上記の N を $N-n$ に，α, β をそれぞれ事後分布の母数 $\tilde{\alpha}, \tilde{\beta}$ に変えて上記の関係式をしらべ，$u_{N-n}(\tilde{\alpha}, \tilde{\beta})>0$ ならば試行を続けることにすればよい．

関数 $u_N(\alpha, \beta)$ は上記の関係式から再帰的に定義できるが，その形は N が大きくなればきわめて複雑になる．ただし $u_N(\alpha, \beta)>0$ となる α, β の範囲は，ある関数 Φ_N を用いて

$$\Phi_N(\alpha, \beta) > c$$

と表される．そうして Φ は N とともに単調増大で，かつ $N\to\infty$ のときある関数 Φ^* に収束することがわかる．そこで試行の回数に制限がないとすれば，

$$\Phi^*(\tilde{\alpha}, \tilde{\beta}) > c$$

である限り試行を続けるのがよいということになる．

そこで Non-Bayesian の観点から考えると，このような方式において，(1)試行が本来「有利」なとき，すなわち $p>c$ のとき，試行が無限に続けられる確率はどれくらいか，(2)試行が「不利」なとき，すなわち $p<c$ のとき試行を中止するまでの平均損失はどれくらいか，を p のいろいろな値に対して求めることが問題になる．そうしてその結果から最初の事前分布の母数 α, β の値をどのように定めるべきかが問題となるであろう．

3° 零和 2 人ゲームの基本定理により，ミニマックス解は，一般にある事前分布に対応するベイズ解になることが多い．すなわち，母数のすべての値に対応するリスクの上限（最大値）を最小にするようなミニマックス解は，最も不利な分布 (least favourable distribution, LF 分布と略記する) といわれるある事前分布に対応するベイズ解に一致する場合，あるいは事前分布の

系列に対応するベイズ解の極限になる場合が多い．このことからミニマックス解はベイズ解の範囲で求めることができる．このことを拡張すると，次のような形の問題を扱うことができる．

$X_1, X_2, \cdots, X_n, \cdots$が互いに独立に密度関数 $f(x_i, \theta, \xi_i)$ を持つ分布に従うとする．ここで θ はすべての観測値に共通であるが，局外母数 ξ_i は各 i ごとに異なるものとする．これを「無限に多くの局外母数を持つ問題」という．ここで目的は共通の母数 θ について推測を行うことであると考える．この問題は一般に複雑困難であって，漸近的にも見通しのよい結果が得られないことが多い．

ここで $\xi_i (i=1, 2, \cdots, n, \cdots)$ を母数の無限系列とする代りに，互いに独立に分布関数 G を持つ分布に従う確率変数とすると，問題はやや簡単になる．この問題は「無限に多くの偶然母数 (incidental parameter) を持つ問題」といわれる．この問題は $X_1 \cdots, X_n, \cdots$が互いに独立に，

$$f^*(x, \theta) = \int f(x, \theta, \xi)\, dG(\xi)$$

を密度関数とする分布に従うとすることと同一である．ただし一般に ξ の分布の G の形は特定化されていないから，パラメトリックなモデルにはなっていない．

さらにここで G に特定の分布形を仮定すれば，ふつうのパラメトリック・モデルになる．

そこで問題は本来「無限に多くの局外母数」をふくむ状況を「偶然母数」問題として扱ったらどうなるか．つまり θ は一応さておいて ξ_i についてベイズ的方法を採用するとどうなるか，それはミニマックス的な意味において，あるいはなんらかの別の意味で合理化されたのではないか，さらに事前分布について特定の形を LF 分布としてえらぶこともできるのではないか，ということである．

このことは直観的に次のように合理化できる．いま観測値を $X_1, \cdots, X_n$ とし，それに対応する局外母数 $\xi_1, \cdots, \xi_n$ の範囲はすべて同じであるとし，リスクは ξ_i の値にはよらないとすると，問題は標本観測値の番号 $1, \cdots, n$ の変換に関して不変になる．したがって $(\xi_1, \cdots, \xi_n)$ の LF な同時分布は対称な

分布であるとしてよい．ところでベイズ統計学の基本定理であるド・フィネッティの定理により，$(\xi_1, \cdots, \xi_n)$ の対称な同時分布は独立分布の混合として

$$\alpha G(\xi_1, \cdots, \xi_n) = \int \prod_{i=\tau}^{n} g(\xi_i, \tau)\, d\pi(\tau)$$

という形に表現できる．すなわち，これは ξ_i が未知母数 τ を含む分布，$g(\xi_i, \tau)$ を持つ「偶然母数」であると想定することに等しい．

この問題について最も簡単な例をあげよう．$X_1, \cdots, X_n$ が互いに独立に正規分布に従い，その平均は $\pm\mu$，分散は σ^2 であるとする．ただし各 X_i についてその平均が $+\mu$ であるか $-\mu$ であるかはわからないとする．もちろん $\mu \geqq 0$ としておく．このとき μ について推定あるいは検定をどのようにして行ったらよいであろうか．ここで局外母数 ξ_i を ± 1 の値をとるものとし，X_i の平均を $\xi_i\mu$ に等しいとすれば「無限に多くの局外母数を含む問題」に帰着する．

ここで μ, ξ_i, σ^2 について最尤法を直接適用すると

$$\hat{\xi}_i = \mathrm{sgn} X_i \qquad (X_i \text{ の符号})$$

$$\hat{\mu} = \sum |X_i| / n$$

$$\hat{\sigma}^2 = \sum (x_i - \hat{\mu})^2 / n$$

となるが，明らかに $\hat{\mu}, \hat{\sigma}^2$ に偏った推定量となり，一致性も持たない．

一般に「無限に多くの母数を含む問題」においては最尤法を直接適用することは必ずしも適切でないことが知られているが，これもその一例である．

ここで問題を「無限に多くの偶然母数」の問題と見なすと，

$$\Pr\{\xi_0 = 1\} = \pi = 1 - \Pr\{\xi_i = -1\}$$

とすれば，尤度関数は

$$\prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\,\sigma} \left(p e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} + q e^{-\frac{(x_i+\mu)^2}{2}} \right) \right]$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}\Sigma x_i{}^2} \prod_{i=1}^{n} \left(p e^{\frac{\mu}{\sigma^2} x_i} + q e^{-\frac{\mu}{2\sigma^2} x_i} \right)$$

となるので，この値を最大にするような μ, σ^2, p の推定値とすることができる．このような推定量は ξ_i が実際に確率変数でなく，ベルヌーイ系列になっていない場合でも $\Sigma \xi_i / n \longrightarrow p$ となる場合には一致推定量になることが示される．

4° 大標本理論において，事前分布がすべての母数の値に対して正，かつ滑らかな密度関数を持つとすると，標本の大きさ n が大きくなるとき，事後分布は最尤推定量を平均とし，フィッシャー情報量の逆数を分散とする正規分布に近づくことが示される．

より詳しくいえば次のようになる．θ を実母数，$f(x,\theta)$ を観測値に対する確率密度，$\pi(\theta)$ を事前分布の密度とし，

(1) $f(x,\theta)>0$ となる x の範囲は θ と無関係

(2) $f(x,\theta)$ は θ に関して連続2回微分可能

(3) $\pi(\theta)>0 \quad \forall\theta$

(4) $\pi(\theta)$ は θ に関して連続2回微分可能

と仮定する．

そうすると事後密度を $\bar{\pi}(\theta)$ と表すと

$$\log\bar{\pi}(\theta)=\text{const}+\sum_{i=1}^{n}\log f(x_i,\theta)+\log\pi(\theta)$$

となるから，$\hat{\theta}$ を最尤推定量，すなわち

$$\sum\frac{\partial}{\partial\theta}\log f(X_i,\hat{\theta})=0$$

とすれば

$$\log\bar{\pi}(\theta)=\text{const}+\frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\hat{\theta})(\theta-\hat{\theta})^2$$
$$+\frac{\partial}{\partial\theta}\log\pi(\hat{\theta})(\theta-\hat{\theta})$$
$$+\frac{1}{2}\frac{\partial^2}{\partial\theta^2}\log\pi(\hat{\theta})(\theta-\hat{\theta})^2+o\|\theta-\hat{\theta}\|^2$$

と展開される．そこで n が大きくなるとき

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\hat{\theta})\longrightarrow -I(\hat{\theta})=-E\left(\frac{\partial}{\partial\theta}\log f(X_i,\theta)\right)^2\Big|\theta=\hat{\theta}$$

となることに注意して，$\sqrt{n}(\theta-\hat{\theta})=\xi$ とおきかえれば

$$\log\bar{\pi}(\theta)=\text{const}-\frac{1}{2}I(\hat{\theta})\xi^2+\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log\pi(\hat{\theta})\xi+0(1)$$

となるから，$\sqrt{n}(\theta-\hat{\theta})$ の事後分布が漸近的に平均0，分散 $1/I(\hat{\theta})$ の正規分布に従うことがわかる．このことは，漸近的には事後分布が，事前分布が滑らかな密度を持つ限り，事前分布と無関係になり，かつ最尤推定量にのみ依存することを意味し，それはまた最尤推定量が漸近十分統計量になることを意味している．

もう1つ高次の項まで求めると，$J(\hat{\theta})$ と $K(\hat{\theta})$ を適当にとり

$$\log\bar{\pi}(\theta)=c_1-\frac{I(\hat{\theta})}{2}\xi^2+\frac{Z_2(\hat{\theta})}{2\sqrt{n}}\xi^2+\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log\pi(\hat{\theta})\xi$$
$$-\frac{3J(\hat{\theta})+K(\hat{\theta})}{6\sqrt{n}}\xi^3+o\left(\frac{1}{\sqrt{n}}\right)$$

となるから，事後分布を $1/\sqrt{n}$ のオーダーまで求めるためには，最尤推定量 $\hat{\theta}$ のほかに，

$$Z_2(\hat{\theta})=\frac{1}{\sqrt{n}}\left\{\sum\frac{\partial^2}{\partial\theta^2}\log f(X,\hat{\theta})+I(\hat{\theta})\right\}$$

を用いなければならないことになる．

しかしながら，ここで

$$\log\bar{\bar{\pi}}(\theta)=c_2-\frac{I(\hat{\theta})}{2}\xi^2+\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log\pi(\hat{\theta})\xi$$
$$-\frac{3J(\hat{\theta})+K(\hat{\theta})}{6\sqrt{n}}\xi^3+o\left(\frac{1}{\sqrt{n}}\right)$$

とおくと，$\bar{\pi}$ と $\bar{\bar{\pi}}$ の分布の間の Kullback 情報量は

$$\int\bar{\Pi}(\theta)\log\frac{\bar{\pi}(\theta)}{\bar{\bar{\pi}}(\theta)}d\theta=\int\left(c_1-c_2+\frac{Z^2(\hat{\theta})}{2\sqrt{n}}\right)\xi^2d\theta+o\left(\frac{1}{\sqrt{n}}\right)$$
$$=\left(c_1-c_2+\frac{Z_2(\hat{\theta})}{2\sqrt{n}}\right)\frac{1}{I(\hat{\theta})}+o\left(\frac{1}{\sqrt{n}}\right)$$

となるが，ここで c_1, c_2 は $\bar{\pi}(\theta), \bar{\Pi}(\theta)$ がともに密度関数になるように定められた定数であることを考慮すれば，

$$c_1-c_2+\frac{Z_2(\hat{\theta})}{2\sqrt{n}}=o\left(\frac{1}{\sqrt{n}}\right)$$

となる．したがって $\bar{\bar{\pi}}(\theta)$ は分布としては $\bar{\pi}(\theta)$ を $1/\sqrt{n}$ のオーダーまで近似しているものと考えることができる．すなわち $\hat{\theta}$ のみを用いても，事後分布をある意味で $1/\sqrt{n}$ のオーダーまで近似できることになる．このことから，$\hat{\theta}$ は $1/\sqrt{n}$ のオーダーまでの高次漸近十分統計量になっていくと考えることができる．

同様の考え方をもう一段進めれば，$\hat{\theta}$ と $Z_2(\hat{\theta})$ の2つの推定量を組み合せることによって $1/n$ のオーダーまでの高次漸近十分性が得られることがわかる．

このような考え方を拡張すれば非正的な場合の漸近十分性の議論を展開することができるであろう．この問題について詳しくは，Akahira and Takeuchi[1981] が参考となろう．

3. ベイズ法の評価

逆に Non-Bayesian の観点からベイズ法，特に事前分布の設定の方法を評価することができる．それについてもいくつかの例題が考えられる．

1° 観測値の分布に指数型分布が想定される場合，事前分布としていわゆる共役分布が用いられることが多い．

すなわち $X_1, \cdots, X_n$ が互いに独立に

$$f(x, \theta) = k(x)\exp\left(\sum_{i=1}^{p} c_i(\theta)\, t_i(x) + c_0(\theta)\right)$$

という形の密度を持つ分布に従うとき，θ に対して想定された

$$g(\theta, \xi) = l(\xi)\exp\left(\sum_{i=1}^{p}\xi_i c_i(\theta) + \xi_0 c_0(\theta)\right)$$

という形の密度を共役密度という．ただし，$\xi = (\xi_0, \xi_1, \cdots, \xi_p)$ はこの分布の $p+1$ 次元母数である．共役密度を持つ事前分布を想定すると，事後分布も同じ型の分布になり，その母数は

$$\xi + T = (\xi_0 + n, \xi_1 + T_1, \cdots, \xi_p + T_b)$$

$$\text{ただし}\ T_1 = \sum_{j=1}^{n} t_1(x_j), \cdots, T_p = \sum_{j=1}^{n} t_p(x_j)$$

となる．

共役分布を想定すると，事後分布の計算が容易になるので，きわめて便利であり，また $p+1$ 個の母数を定めることができるので，ある程度広い範囲の事前分布を考慮することができる．しかし逆に事後分布の型が一定であるということは，事事上不都合を生ずることもある．

たとえば，$X_1, \cdots, X_n$ が互いに独立に平均 μ，分散 σ^2 の正規分布に従うとき，X_i の密度関数は

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma}x - \frac{\mu^2}{2\sigma^2}\right]$$

と表されるから，これの共役分布は3つの母数 ξ, τ, ν を用いて

$$g(\mu, \sigma \mid \xi, \tau, \nu) = \frac{c}{\sigma^{\nu}}\exp\left[-\frac{1}{2\sigma^2}\tau + \frac{\mu}{\sigma^2}\xi - \frac{\mu^2}{2\sigma^2}\nu\right]$$

$$= \frac{c'}{\sigma}\exp - \frac{\nu}{2\sigma^2}\left(\mu - \frac{\xi}{\nu}\right)^2 \cdot \frac{c''}{\sigma^{\nu-1}}\exp - \frac{1}{2\sigma^2}\left(\tau - \frac{\xi^2}{\nu}\right)$$

と表されるから，$1/\sigma^2$ の事前分布は自由度 $\nu-1$ の χ^2 分布の定数倍に，σ^2 が与えられたときの μ の分布は平均 ξ/ν，分散 σ^2/ν の正規分布になる．したがって，$\mu - \xi/\nu$ の周辺分布は自由度 $\nu-1$ の t 分布の定数倍になる．

このことから $X_1, \cdots, X_n$ が与えられたときの事後分布は，母数 ξ, τ, ν がそれぞれ，

$$\xi + \sum X_i,\ \tau + \sum X_i^2,\ \nu + n$$

に変った形になることがわかる．したがって μ の事後分布を求めると

$$\sqrt{\nu+n}\left(\mu - \frac{\xi + \sum X_i}{\nu+n}\right) \Big/ V$$

が自由度 $\nu+n-1$ の t 分布の定数倍になることがわかり，そうしてその尺度は，

$$V = \left[\left(\tau + \sum X_i^2 - \frac{1}{\nu+n}(\xi + \sum X_i)^2\right) \Big/ (\nu+n-1)\right]^{\frac{1}{2}}$$

に比例することがわかる．

そこでたとえば μ を推定するとすれば，対称な損失問題を用いればベイズ推定量は，

$$\hat{\mu} = \frac{\xi + \sum X_i}{\nu + n}$$

となり，結局これは事前分布の平均 $\bar{\xi} = \xi/\nu$ と標本平均 $\bar{X} = \sum X/n$ の一定比重の加重平均となる．そうしてその平均2乗誤差は，

$$\frac{n}{\nu+n}\sigma^2 + \frac{\nu^2}{(\nu+n)^2}(\mu - \bar{\xi})^2$$

である．

この値は μ が $\bar{\xi}$ から離れればいくらでも大きくなる．つまり事前分布の

平均の想定が著しく誤っていればリスクは大きくなり得る.

そこでかりに事前分布の想定が真の値から離れていてもリスクがあまり大きくならないようにすることが要請されるかもしれない.

別の方向から考えると次のような場合が想定される. いま $X_1, \cdots, X_n$ が観測されると, μ の値はほぼ信頼区間,

$$\bar{X} = t_\alpha S$$
$$\bar{X} = \sum X_i/n,\ S^2 = \sum(X_i-\bar{X})^2/(n-1)$$

の中に入っていると思われる. したがって事前分布の平均 ξ がこの区間のはるか外にあれば, その想定は「誤まり」であり, したがってまた上記のベイズ推定量も適当でないと思われるであろう.

このとき事前分布自体に対してそのまた「信頼度」を考慮することも考えられるが, しかしSavageも主張したように, このような方法はいたずらに論理を複雑にするだけで, 望ましくないであろう.

むしろ難点は事後分布における事前情報と標本からの情報との比 $\nu : n$ がつねに一定であるところから生じていると思われる. したがって事前情報と標本情報との事後の比が標本の値によって変化するように, つまり ξ が信頼区間の外に出るような場合には, 事後平均が ξ より $\bar{x}$ に近くなるような事前分布を想定すれば上記の難点はさけられるであろう.

このような事前分布としては, 正規分布より「スソの長い」分布をえらべばよい.

σ^2 未知のままで考えると面倒だから σ^2 を既知とし, その値を1としよう. このとき標本の確率密度関数は,

$$\prod f(x_i, \mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum x_i^2 + \mu\sum x_i - \frac{n}{2}\mu^2\right\}$$

となるから, 共役密度は2つの母数 ξ, ν を用いて,

$$g(\mu \mid \xi, \nu) = c\exp\left\{\mu\xi\nu - \frac{\nu}{2}\mu^2\right\}$$

$$= c'\exp - \frac{\nu}{2}(\mu-\xi)^2$$

となり, μ の事前分布として平均 ξ, 分散 $1/\nu$ の正規分布を想定することになる. このとき μ の事後分布が, 平均 $(\xi+\sum X_i)/(\nu+n)$, 分散 $1/(\nu+n)$ の正規分布になることは容易にわかるが, $\bar{X}$ が ξ から大きく離れた場合, これは先に述べた理由であまり好ましくない. そこで事前分布としてたとえばコーシー分布,

$$g(\mu \mid \xi, \tau) = \frac{1}{\pi\tau(1+(\mu-\xi)^2/\tau^2)}$$

を想定すれば, 事後密度は,

$$f(\mu \mid x_1, \cdots, x_n, \xi, \tau) = c\exp\left(-\frac{n(\mu-\bar{x})^2}{2}\right)\cdot\frac{1}{\tau^2+(\mu-\xi)^2}$$

と表される.

この具体的な形を求めることは難しいが, $\bar{x}$ と ξ とが離れると事後分布の形が変ることは明らかである. そして $\bar{x}$ と ξ が離れれば事後分布は $\bar{x}$ を平均とする正規分布に近づく. また μ の推定値として事後平均をとるとすると, それを具体的な形に表すことは難しいので, 事後分布のモードを求めると, それは,

$$\frac{\partial}{\partial\mu}\log f(\mu \mid x_1, \cdots, x_n, \xi, \tau) = 0$$

$$= n(\mu-\bar{x}) - \frac{2(\mu-\xi)}{\tau^2+(\mu-\xi)^2}$$

の解 $\hat{\mu}$ として与えられる. ところがすべての μ に対して

$$\left|\frac{2(\mu-\xi)}{\tau^2+(\mu-\xi)^2}\right| < \frac{1}{\tau}$$

であるから,

$$|\hat{\mu}-\bar{x}| < 1/n\tau$$

となる. すなわち事後モードは標本平均から $1/n\tau$ 以上離れることはない. したがってこれを μ の推定量として平均2乗誤差を求めると

$$|\hat{\mu}-\mu| \leqq |\bar{X}-\mu| + |\hat{\mu}-\bar{X}|$$
$$< |\bar{X}-\mu| + 1/n\tau$$

であるから,

$$E(\hat{\mu}-\mu)^2 < E(\bar{X}-\mu)^2 + \frac{2}{n\tau}E|\bar{X}-\mu|\,\tau\frac{1}{n^2\tau^2}$$

$$= \frac{1}{n} + \frac{2\sqrt{2}}{n\tau\sqrt{n\pi}} + \frac{1}{n^2\tau^2}$$

となって，これは有界であることがわかる．

このように「スソの長い」事前分布を用いることによって，ベイズ法を用いる場合，時として生ずる「不都合な」状況を避けることができる．

一般の指数型分布においても，事前分布において母数 ξ_0 をさらにある分布によって平均した混合共役分布

$$g^*(\theta \mid \xi) = \int g(\theta \mid \xi_1, \cdots, \xi_p, \xi_0)\, dG(\xi_0)$$

を用いることによって，同様の結果を得ることができるであろう．

2° ベイズ法において，事前分布としてしばしば「情報のない」あるいは「漠然とした情報」(vague information) を表すものとして一様分布や，あるいはその他の improper な分布が用いられることがある．位置母数の場合は，ルベーグ測度 $d\mu$，尺度母数の場合は $d\sigma/\sigma$，そしてより一般の場合には情報量行列を $I(\theta)$ として $I(\theta)^{-\frac{1}{2}}d\theta$ を確率要素とする事前分布を用いることが多い．そしてこれらの密度は全積分が無限大になるから improper 分布である．ところが improper な事前分布を用いると，ベイズ解が非許容的 (inadmissible) になることがある．proper な事前分布に対応するベイズ解はすでに述べたように原則として許容的 (admissible) であるから，このことは improper な事前分布を用いることの危険性を表している．

このことをはじめて印象的な形で示したのは C. Stein である．すなわち $X_{ij}(i=1, \cdots, p\,;\, j=1, \cdots, n)$ が互いに独立に平均 θ_i，分散 1 の正規分布に従うとき，θ_i の事前分布に一様分布を想定すると，θ_i の事後分布は互いに独立，平均 $\bar{X}_i$，分散 $1/n$ の正規分布になる．したがって対称な損失関数，たとえば 2 乗誤差の和，

$$\sum(\hat{\theta}_i - \theta_i)^2$$

を仮定すると，ベイズ解は $\hat{\theta}_i = \bar{X}_i$ となる (これは最小 2 乗分散不偏推定量である)．

ところが $p=1, 2$ の場合にはこの解は許容的であることが証明されるが，$p \geqq 3$ の場合には，これが非許容的であることが Stein によって示されたのである．すなわち

$$\hat{\theta}_i^* = \left(1 - \frac{p-2}{n\sum \bar{X}_i^2}\right)\bar{X}_i$$

とおくと，つねに，

$$E(\sum(\hat{\theta}_i^* - \theta_i)^2) < \frac{p}{n} = E(\sum(\bar{X}_i - \theta_i)^2)$$

となることが証明される．しかし p が大きくなると $E(\sum(\hat{\theta}_i^* - \theta_i)^2)$ は p/n よりかなり小さくなり得る．

Stein の例は多くの問題について拡張され，ほとんどすべての例について，母数の次元がある値以上になると improper な事前分布に対応するベイズ推定量は非許容的であることが示されている．ただしその境界となる次元は場合によって異なる．

許容性はきわめて弱い条件であるから，非許容的な解は望ましくないことは明らかであろう．しかしベイズ統計学の論理の内部からはどのような場合に improper な事前分布に対応するベイズ解が非許容的になるかを示すことはできない．

3° データに対して想定される確率分布が，現実に厳密にあてはまるという保証はない．確率が想定された形からある程度離れた場合でも，著しく妥当性を失ったり効率が悪くなったりしないような推測方式を求めるのが「ロバストな推測」(robust inference) の問題である．この問題は，ふつう Non-Bayensian 的な観点から論ぜられている．

ベイズ統計学の観点からロバスト推測の問題を論ずるには，抽象的には，考えられるすべての確率分布のクラスについて事前分布を想定し，母数 θ の事後周辺分布を計算すればよいということになるかもしれない．しかし考えられる確率分布のクラスの全体は有限次元の母数で表されないこと，仮にその点について数学的に解決できたとしても，現実に事前分布を具体的に設定することは著しく困難であろう．それよりも現実的な方法は，確率分布のタイプを表す少数次元の母数を導入してそこにふつうのベイズ法を適用することであろう．しかし Non-Bayesian の立場からすれば，さらにそのような方法で得られた方式が，そのような母数をもつクラスによっては正確には表さ

れないような分布の下で，どのような性質を持つかをしらべることが必要である．

たとえば，$X_1, \cdots, X_n$ が互いに θ を平均とする対称で連続分布に従うとしよう．このとき最もふつうに用いられ，かつ数学的に好都合なのは正規分布の仮定である．しかし実際にはデータが厳密に正規分布に従うとは考えられない場合も少なくない．もちろん正規分布を想定すれば，$\hat{\theta}=\bar{X}=\sum X_i/n$ が「最もよい」推定量であると考えられるが，しかし分布が正規分布からある程度大きく離れ，特に「スソが長く」なると，これはあまりよい推定量ではなくなり，分散が大きくなる．正規分布からある程度離れた分布の下でも分散があまり大きくならないような「ロバスト推定量」に関してはすでに主としてNon-Bayesianの立場から多くの研究がなされているが，Bayesianの観点からは，例えば次のような方法が考えられる．すなわち X_i の分布として t 分布のクラス

$$f(x\,|\,\theta, \sigma, \nu)=\frac{c(\nu)}{\sigma}\left(1+\frac{(x-\theta)^2}{\nu\sigma}\right)^{-\frac{\nu+1}{2}}$$

を想定する．そうしてここで θ, σ, ν，の事前同次密度として，

$$\frac{h(\nu)}{\sigma}d\theta d\sigma d\nu$$

という形を想定すれば，θ, σ, ν の事後密度は，

$$f(\theta, \sigma, \nu\,|\,x_1, \cdots, x_n)=\frac{c(\nu)h(\nu)}{\sigma^{n+1}}\prod\left(1+\frac{(x_i-\theta)^2}{\nu\sigma}\right)^{-\frac{\nu+1}{2}}$$

となる．これを σ, ν で積分すると θ の事後密度が得られるので，それの平均あるいはモードを用いて θ を推定することができる．

ここで $h(\nu)$ にどのような形を想定したらよいかが1つの問題であるが，自由度 ν の t 分布と $\nu+1$ の t 分布の間の「距離」は ν が増大すれば小さくなることを考慮し，また ν の値が著しく小さくなる可能性は小さいとして

$$h(\nu)=\frac{1}{\nu+1}$$

とするのが1つの合理的な仮定であるように思われる．

θ の事後分布を解析的に表現することはできないが，データが与えられたときに数値的に計算することは可能である．このような方法によってデータの分布が必ずしも t 分布でなくてもある程度スソが長い対称でなめらかな分布である場合には，かなりよい性質を持つ推定量が得られることが期待される．モンテカルロ実験などにより，そのことを確めてみることが望ましい．

分布型のクラスとしては t 分布以外のものも考えられよう．Box-Tiao [1973] は，

$$f(x\,|\,\theta, \sigma, \alpha)=c\exp-\frac{|x-\theta|^{\alpha}}{\sigma}$$

という形を想定しているが，これは密度関数がなめらかでないこと，十分スソが長い分布を表現し得ないこと，の2つの理由により，あまり適当とは思われない．

どのような分布のクラスを想定するのがよいかは，それを前提して得られるベイズ推定量の性質によって定めることができる．

4. おわりに

Non-Bayesianの観点から見れば，ベイズ法は特に複雑な構造を持つ推測問題において，「よい」推測方法を導く有効な手段であると思われる．しかしその「よさ」は主観確率の論理的一貫性によってではなく，得られた推測方法のいろいろな条件の下での確率分布によって基礎づけられなければならないというのがNon-Bayesianの基本的な立場であり，それはおそらくBayesianには受け入れられないであろう．

しかし現実の問題においては，そのような「哲学的な」立場の相違はBayesianとNon-Bayesianとの協力を妨げる重大な障害とはならないであろう．Savageも *The Foundations of Statistics* の序文の中でいっているように，基礎的方法論における立場の相違は，実際の応用の場ではそれほど大きな違いをもたらさないのであって，常識というものが判断の基礎となるからである．その意味で実際的な問題の中で，ここに述べたようなタイプの問題について具体的な手法を導くためにベイズ的方法を操作的に用いることについて，BayesianとNon-Bayesianが協力することは有益であると思われる．この一文がそのような研究を盛んにする1つのきっかけともなれば幸い

ある.

参 考 文 献

Akahira, M. and K. Takeuchi, 1981 *Asymptotic Efficiency of Statistical Estimators*, Springer-Verlag.

Box, G. E. and G. C. Tiao, 1973 *Bayesian Inference in Statistical Analysis*, Addison-Wesley.

Savage, L. J., 1954 *The Foundations of Statistics*, John-Wiley.

鈴木雪夫, 1987 『統計学』, 朝倉書店.

竹内 啓, 1975 「ベイズ統計学の“非正統的”検討」,『経済学論集』第 41 巻第 1 号.

II 応 用

第6章
ベイズ予測*

和 合 肇

1. はじめに

予測とは，不確実な将来についての1つの見方であり，予測を行う主な理由の1つは，賢明な行動がとれることにある．経済予測の場合を考えてみよう．適切な経済計画の遂行のためには，来年の経済成長はどのぐらいになるだろうか，あるいはインフレ率は何パーセントぐらいになるだろうかということを知りたい．企業ではそれによって新規の工場建設に対する設備投資を行うべきか否かを決定し，将来の顧客の需要，販売コスト，利益について計画をたてる．また，多品種少量生産のもとでの在庫管理では，個々の商品の毎月の販売高の予測を的確に行う必要がある．このような予測を行う方法としては，単純な指数平滑法から，時系列モデル (Box-Jenkins 法)，回帰モデル，そして複雑な計量経済モデルによる方法などが知られている．しかし，これらの方法による予測では，実際はただ1つの予測値だけしか考慮の対象にしていない．これは非常に断定的で危険が多い結果をもたらす場合が多い．われわれが必要とするのは，むしろ可能性や確率を表す確率分布による予測であろう．ベイズ予測は，このような予測を可能にする方法であるといえよう．

予測を行う場合に利用可能な情報源は，大ざっぱに言えば，過去の歴史的

* 本稿を作成するにあたって，統計学コンファレンスにおいてコメントを頂いた文部省統計数理研究所の北川源四郎氏に深く感謝致します．

な出来事と，将来のことに関する知識とに分けられる．過去の歴史は単にGNPや物価の動きとかある製品の販売高や価格のような記録されたデータだけでなく，すべての過去の経験が含まれる．これらはすべて一緒になって予測システムを形成する．すなわち，これは経験と分析にもとづいてルーティン的に処理したデータと情報を予測に取り入れた統計モデルである．このようなモデルによる予測は次のような例外的な事態がない限り用いられる．1つは将来起きるかも知れない出来事について，その専門知識にもとづいた見通しに関するものである．たとえば，半導体の輸出規制の問題とか，突然の円高や，株暴落の可能性，第3次オイルショックによるエネルギー危機の可能性，等々についての予想に関しては，分析者がそのインパクトをどの様に考えるかによっている．もう1つはフィードバック情報に関するもので，予測パフォーマンスが不適切であるために生ずる．これは通常，統計的モニタリング，すなわち最近時点でのモデルの予測能力を連続的に評価することによって発見することができる．完全な予測システムは，このように統計モデルによって処理した情報と，主観的な情報を結び付けたものでなくてはならない．そして，例外的な事態が生じたときには，その原因を調べてモデルの関連した部分を，新しい情報をもっと早く取り入れることができるようなシステムに改良する必要がある．あるいは，その変化について特定の対立仮説を想定する多段階モデルの考え方（Harrison and Stevens[1971, 76] 参照）をシステムに含める方法もある．多段階モデルやベイズ予測の考え方は，それが完全に自動的な予測システムを提案するものであると思われている．しかし，ベイズ予測はまったく新しい考え方というわけではなく，モデル化したデータ処理と主観的な情報を結び付けることの重要性を強調した予測のシステムである．

以下で述べるベイズ予測の方法は，動的線形モデルとその各要素についてのモデルビルディングにもとづいて行われるもので，Harrison and Stevens[1976] によって提唱され，West, Harrison and Migon[1985] や West[1986] でさらに展開されている．システムを状態空間（state space）表現を用いて表し，状態変数を未知の係数と考えたとき，この係数は時間で変化するので，このモデルのことを DLM（dynamic linear model）という．この係数とその共分散行列の推定には，カルマンフィルター方程式が用いられる．ベイズ予測の目的は，将来観測値に対する予測分布を導くことにある．いままでに状態空間表現を用いたいろいろなモデルが提案されている．たとえば，Harvey[1984] や Harvey and Todd[1983] の一変量状態空間「構造」モデル（univariate state space structural model），Kitagawa[1981], Gersch and Kitagawa[1983], Kitagawa and Gersch[1984, 1985] などはこのモデルに非常に似通ったモデルである．ここで用いられる状態空間モデルによる予測は，ベイズ予測モデルをはじめ，伝統的な回帰モデルや ARMA 時系列モデル，そして時変係数モデルなどをその特別な場合として含んでいる．このモデルについてはその特徴は一部の専門家を除いてあまり知られていないので，Harrison[1967] や Harrison and Stevens[1971, 1976] に沿って第2節で状態空間モデルについて簡単に紹介し，第3節では予測に用いる構造時系列モデルの各成分のモデル化の方法を述べ，第4節では主要結果と情報割引の考え方を示し，外生的な出来事（event）によるモデルの影響と修正について説明する．第5節で典型的な分析例について述べた後，最後に第6節で多変量データを扱うモデルへの一般化を考える．

2. ベイズ予測と動的線形モデル（DLM）

A. 動的線形モデル

ベイズ予測システムの基本要素は次のようにまとめることができる．

(1) 構造モデルによって分析対象をはっきりと定義し，パラメータによってモデルの性質を意味づけし，解釈し，他の分析者とコミュニケートすることを可能にする．

(2) 任意の時間 t で利用できる情報と知識 D_t の下で，パラメータについての確率的な情報を用いる．

(3) どの様にシステムを考え，どの様にパラメータが確率的に変化するのかについて記述する逐次的モデルを用いる．

(4) 予測は確率分布の形をとる．

動的線形モデル（DLM）とは，(1)ある確率過程（以下，簡単化のためプロセスと呼ぶ）の観測値がパラメータにどのように確率的に依存しているか，(2)プロセスパラメータが，その固有の動学プロセスと確率ショックまたは攪乱項の両方から時間が経過するにつれてどのように変動するか，を表す方程式体系である．モデルは，離散型，等時間間隔で表される．

このモデルで用いられる状態空間モデルとは，観測される T 個のデータ y_t を，時間と共に変化する n 個の状態変数 θ_t と結び付ける方程式体系である．このモデルは各時間 t に対して 4 つのパラメータ $\{F, G, V, W\}_t$ で定義される．n 個の状態を記述した $n\times l$ プロセスベクトル θ_t と事後情報のもとで，連続的な確率関係は次の 2 本の確率逐次方程式の形で書くことができる：

$$\text{観測方程式}：y_t = F'_t\theta_t + v_t \quad v_t \sim N[0, V_t] \tag{2.1}$$

$$\text{システム方程式}：\theta_t = G_t\theta_{t-1} + w_t \quad w_t \sim N[0, W_t] \tag{2.2}$$

y_t は将来観測値を表し，F_t は n 個の独立変数ベクトルで t 期では既知，G_t は既知の $n\times n$ システム行列，v_t は独立な正規確率変数ベクトルで平均ゼロ，分散は未知の V_t, w_t は独立な正規確率変数ベクトルで平均はゼロ，分散共分散行列は $W_t(=V_tB$, B は任意の正値定符号行列）と仮定する．D_{t-1} は $t-1$ 期までに用いられたすべての知識を表す．

古典的な線形回帰モデルの場合は，$\theta_t=\theta$ であり，V_t が時間に依存しない単純な静学的な場合の DLM になり，システム方程式 $\theta_t=\theta_{t-1}$ は必要なくなる．このプロセスパラメータ θ_t は従属変数と独立変数との関係を表し，静学的な線形回帰モデルのパラメータ的な解釈ができる．F_t が固定されている場合はモデルは時系列の状態空間表現になり，このときはパラメータはプロセス水準とかプロセス成長として解釈することもできる．この DLM はいくつかのモデルの線形結合として表すこともでき，Lindley and Smith [1972] の階層（hierarchical）モデルもまた DLM として表すことができる．

B. 推　定

初期時点 $t=0$ では，パラメータベクトル θ_0 に関する情報は，平均 m_0 で分散共分散行列 c_0 の正規分布で表される．最初の観測値以前の θ_0 の分布を $N(m_0, c_0)$ とすると，t 期における θ_t の事後分布も正規分布をする．

$$(\theta_t|D_t) \sim N(m_t, c_t) \tag{2.3}$$

ここで，$m_t=E[\theta_t|D_t]$ と $c_t=V[\theta_t|D_t]$ の値は逐次的に得られる．y_t のプロセスを DLM の形で表すと，Kalman[1960] の結果を用いてパラメータベクトルの情報を更新する．この結果を予測への応用に便利な形で表すと次のようになる（Anderson and Moore[1979] または Harvey[1981] 第 4 章を参照）．

$$E[\theta_t|D_{t-1}] = G_tE[\theta_{t-1}|D_{t-1}] \tag{2.4a}$$
$$V[\theta_t|D_{t-1}] = G_tV[\theta_{t-1}|D_{t-1}]G'_t + W_t \tag{2.4b}$$
$$E[y_t|D_{t-1}] = F'_tE[\theta_t|D_{t-1}] \tag{2.4c}$$
$$V[y_t|D_{t-1}] = F'_tV[\theta_t|D_{t-1}]F_t + V_t \tag{2.4d}$$
$$A_t = V[\theta_t|D_{t-1}]F_tV[y_t|D_{t-1}]^{-1} \tag{2.4e}$$
$$e_t = y_t - E[y_t|D_{t-1}] \tag{2.4f}$$
$$E[\theta_t|D_t] = E[\theta_t|D_{t-1}] + A_te_t \tag{2.4g}$$
$$V[\theta_t|D_t] = (I - A_tF'_t)V[\theta_t|D_{t-1}] \tag{2.4h}$$

このアルゴリズムではこの逐次性が重要である．実際の予測では現在の事後分布は，最も最近の値 (y_t, F_t)，事後分布 $(\theta_{t-1}|D_{t-1})$，現在の観測値ノイズ分散 V_t と攪乱項分散 W_t から計算される．ここで，$E[y_t|D_{t-1}]$ と $V[y_t|D_{t-1}]$ は $t-1$ 期までの情報の下での y_t の 1 期先予測の期待値と分散であり，したがって e_t は条件付の 1 期先予測誤差になる．A_t はカルマンゲインと呼ばれ，多くのシステムでの平滑化定数ときわめて似たものであるが，A_t は一般には一定ではなく，$n\times T$ 行列である．

カルマンフィルターでは各時点において利用可能な情報の下で θ_t の最小平均二乗線形推定量（MMSE）を得る．しかしベイズ分析ではカルマンフィルターはパラメータの事前分布を更新して事後分布を得る方法として考えている．したがって，m_t は MMSE としてではなく事後分布の平均として解釈される．こう考えることによって，標本期間の最初で事前情報が利用できる場合にはそれを予測に組み入れることが可能になる．そのような情報が利用できない場合には，散漫な事前情報（diffuse prior）を用いればよい．これは線形回帰モデルで non-informative な事前分布を使う場合と同じである．

C. 予 測

$t=1, 2, \cdots$ のすべての時点におけるパラメータベクトル θ_t の事後分布を得た後で，将来の観測値 $y_{t+k}(k=1, 2, \cdots)$ の分布を求める．予測は分布で表され，現在のパラメータの不確実性や，将来の観測値のノイズ分散 V_{t+k}，そして攪乱項分散 W_{t+k}，から導かれる．分布情報から，どんな予測がこの状態では適当かが導かれる．したがって，もしある方向での誤差の結果が，逆方向での同程度の誤差より重大なら，予測はこれらを考慮にいれて偏ったものとすることもできる．

(2.1) と (2.2) の DLM の式から t 期での将来観測変数は，

$$y_{t+k} = F'_{t+k}\theta_{t+k} + v_{t+k} \tag{2.5}$$

$$\theta_{t+k} = G_t\theta_{t+k-1} + w_{t+k} \tag{2.6}$$

と書ける．この式より確率ベクトル v_{t+k}, w_{t+k} の分散を既知とすると，y_{t+k} の将来値の予測にはパラメータ θ_{t+k} と独立変数ベクトル F_{t+k} の予測が必要になる．

$$m_{k|t} = E(\theta_{t+k}|D_t) \tag{2.7}$$

$$C_{k|t} = V(\theta_{t+k}|D_t) \tag{2.8}$$

ここで，$m_{0|t}=m_t=E(\theta_t|D_t)$ と $c_{0|t}=c_t=V(\theta_t|D_t)$ は，カルマンフィルター・アルゴリズムより既知である．(2.6)式より，

$$m_{k|t} = G_t m_{k-1|t} \tag{2.9}$$

$$C_{k|t} = G_t C_{k-1|t} G'_t + W_{t+k} \tag{2.10}$$

そこで，将来パラメータの平均と分散は，既知の初期値 $m_{0|t}$ と $C_{0|t}$ を用いて $k=1, 2, \cdots$ と順次に計算できる．k 期先の予測値の平均と分散は，

$$y_{k\cdot t} = E(y_{t+k}|D_t) \tag{2.11}$$

$$Q_{k\cdot t} = V(y_{t+k}|D_t) \tag{2.12}$$

であるが，これを計算するには 2 つの場合に分けて考える必要がある．F_{t+k} が t 期で既知の場合とそうでない場合である．後者の場合（予測するときの外生変数が確率的であるような場合），(2.5) は未知パラメータに関して非線形になる．

D. 平 滑 化

すべての観測値が得られた後で，t 期以後に得られた観測値を考慮するともっとよい推定量が得られる．このような推定量を計算する方法を平滑化という．これにはいくつかの方法があるが，固定区間平滑化（fixed interval smoothing）の方法は，最後に得られたカルマンフィルター推定量 m_T と C_T から出発し後向きに平滑化方程式を作用させる（詳しくは Anderson and Moore[1979] または Harvey[1981] 第 4 章を参照）．$m_{t|T}$ と $C_{t|T}$ をそれぞれ平滑化推定量とその共分散行列を表すとすると，平滑化は次のように行われる．

$$m_{t|T} = m_t + C_t^*(m_{t+1|T} - G_{t+1}m_t) \tag{2.13}$$

$$C_{t|T} = C_t + C_t^*(C_{t+1|T} - C_{t+1|t})C_t^{*\prime} \tag{2.14}$$

ここで，

$$C_t^* = C_t G'_{t+1} C_{t+1|t}^{-1} \qquad t = T-1, \cdots, 1 \tag{2.15}$$

また $m_{T|T}=m_T$ で $C_{T|T}=C_T$ である．このとき得られた平滑化推定量 $m_{t|T}$ は，最後の観測値を含む T 期までのすべての情報にもとづく最適推定量になる．経済学での応用では，それらは観測されない要素（unobserved component）を表しており，平滑化とは信号除去（signal extraction）を行う方法と考えられる．Engle and Watson[1981, 1987] などの研究でこの平滑化された変数についてさまざまな経済学的解釈がなされている．また時系列データがいくつかの要素から構成されているとき，その 1 つである季節要素を除いて季節調整済み系列を作ることにも使われており，最近その応用がさまざまな分野で行われている．

3. トレンド，季節性，回帰成分の DLM

大きな線形モデルは，いくつかの単純な線形モデルの線形結合と考えることができる．モデルビルディングにおいては，モデルを各構成部分ごとに作ることは実用的と理論的観点の両面から通常利点が多い．このときには，m 個の成分からなるパラメータベクトル $\theta_t=(\theta'_{1\cdot t}, \theta'_{2\cdot t}, \cdots \theta'_{m\cdot t})'$ において

$(\theta_{i\cdot t}|\theta_{t-1})=(\theta_{i\cdot t}|\theta_{i\cdot t-1})$ となる．m 個の成分モデルを考えると，各モデルはそれぞれ $\{F_{it}, G_{it}, 0, W_{it}\}$ とノイズ部分のモデル $\{0, 0, V_t, 0\}$ で表現される．各成分 DLM は次のような $\{F, G, V, W\}_t$ で表される．

$$F_t = (F'_{1t}, F'_{2t}, \cdots, F'_{mt})'$$
$$G_t = \text{ブロック対角}\,(G_{1t}, G_{2t}, \cdots, G_{mt})$$
$$W_t = \text{ブロック対角}\,(W_{1t}, W_{2t}, \cdots, W_{mt})$$

このような構造にするとモデルビルディング，意味づけ，操作性が簡単になるという利点がある．大きなモデルは，最初に解釈しやすい個々の成分を別々にモデル化し，最後にそれらを組み合わせて作ることがしばしば行われる．この定式化によって，他の成分に対する影響を考えずに，特定の成分の影響だけを取り入れることができる．個別の成分モデルに関しては，(1)全体のモデル式は個々の独立な成分式として表せるが，$(\theta_t|D_t)$ は各周辺分布 $(\theta_{it}|D_t)$ からは決められない．これは，観測値系列 y_t を通じて全ての構成部分に関する情報が得られるためである．そして(2)分解の程度に関して注意が必要である．1つの成分をもっと細かく分解することができる場合がある．たとえば，13 期の季節要素は理論的には 6 つの別々の直交する周期部分に分解することができる．しかしながらこれはあまり望ましいことではない．むしろ，すべての季節性については単一の成分モデルとして扱ったほうがよい．逆に，ある回帰関係が固定的成分と変動成分からなっている場合には，回帰成分についてもさらに分解したほうがよい場合がある．

実際の時系列分析や予測でよく用いられるのは，多項式トレンドモデル，季節モデル，回帰モデルを含んだダイナミックモデルである．これら 3 つの成分モデルの例を以下で挙げておこう．次のようなスカラーの時系列，$y_t, t=1, 2, \cdots,$ を考える．

$$y_t = \mu_{1t} + \mu_{2t} + \mu_{3t} + v_t \qquad (3.1)$$

ここで μ_{1t} はトレンド項，μ_{2t} は季節項，そして μ_{3t} は回帰項で，v_t の観測ノイズがある．これらの各項を以下でモデル化する．

A. 線形成長モデル（多項式トレンドモデル）成分

このモデルは，単純なトレンドモデルに勾配項を加えたもので，水準 μ と

勾配 α のパラメータがランダム・ウォークにしたがって時間と共にゆっくりと変化し，次のように表される：

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim NID(0, \sigma^2) \qquad (3.2)$$
$$\mu_t = \mu_{t-1} + \alpha_t + \eta_t \qquad \eta_t \sim NID(0, \sigma_\eta^2) \qquad (3.3)$$
$$\alpha_t = \alpha_{t-1} + \zeta_t \qquad \zeta_t \sim NID(0, \sigma_\zeta^2) \qquad (3.4)$$

このとき，μ_{1t} に対する線形成長の成分 DLM は，次のように表される．

$$\{F_1, G_1, 0, W_1\} = \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}, 0, W_{1t}\right\} \qquad (3.5)$$

ここで，パラメータ・ベクトル，$\theta_{1t}=(\mu_t, \alpha_t)$ は，

$$\mu_{1t} = \mu_t = F_1'\theta_{1t} \qquad (3.6)$$
$$\theta_{1t} = G_{1t}\theta_{1\cdot t-1} + w_{1t} \qquad (3.7)$$

と書ける．μ_t は t 期でのトレンド水準を表し，α_t は $t-1$ 期と t 期の間での成長の増分を表す．

B. 加法的季節モデル成分

季節性を示すプロセスは経済，産業，社会などのデータではごく普通にみられるものであり，予測には重要な因子である．季節性を表すモデルにはいくつかある．

(a) 単純な季節効果成分

季節性の影響は，1 年間での季節サイクル P の各期間に対して加法的な季節項ベクトルを対応させたモデルで表される．s_t をある特定の t に対する期（月）とするとモデルは，

$$y_t = \gamma_{s(t)\cdot t} + \varepsilon_t \qquad (3.8)$$
$$\gamma_{i\cdot t} = \gamma_{i\cdot t-1} + \omega_t \qquad i = 1, \cdots, P \qquad (3.9)$$

となる．確率的な季節モデルでは，季節効果の和はゼロになるとの制約は，

$$\textstyle\sum_{i=0}^{P}\gamma_{i\cdot t} = 0 \qquad t = 1, \cdots, T \qquad (3.10)$$

と表される．単純な形の季節効果モデルを DLM では，

$$y_t = (0, \cdots, 1, \cdots, 0)\,\gamma_t + \varepsilon_t \qquad (3.11)$$
$$\gamma_t = \gamma_{t-1} + \omega_t \qquad \omega_t \sim NID(0, \sigma_\omega^2) \qquad (3.12)$$

と表す．これは多項式モデルと比較すると，F 行列は s_t 番目の成分だけが 1

ではかは 0，そして時間とともに変化する．

(b) フーリエ型季節性

季節性を扱うためには季節効果パラメータ $(\gamma_{1t}, \cdots, \gamma_{pt})$ への変換が必要であるが，操作性とモデル化という観点からは季節効果をいくつかの調和波あるいはコサイン波を組合せた項でモデル化するのが望ましい．このようにモデル化することの利点は，(1)簡単に季節制約を扱える，(2)季節要素をさらにいくつかの副要素に分解することが可能，(3)一定数の調和系列で簡潔に季節性を表現することができる，という点にある．この場合，もしそれが適切なものなら予測精度が上がるが，不適切なら季節パターンの形は誤ったものになる．奇数期間，$p=2q+1$ (q は整数)，の任意の季節効果パターンの DLM$\{F_2, G_2, 0, W_2\}$ 表現は，$\lambda=2\pi/p$ として q 個の部分調和 DLM$\{(1,0)', H_i, 0, W_i^*\}, j=1, \cdots, q$，を組合せて表すこともできる．ここで，

$$H_j = \begin{pmatrix} \cos(j\lambda) & \sin(j\lambda) \\ -\sin(j\lambda) & \cos(j\lambda) \end{pmatrix} \tag{3.13}$$

したがって，
$$F_2 = [(1,0), (1,0), \cdots, (1,0)]'$$
$$G_2 = \text{ブロック対角}[H_1, H_2, \cdots, H_a]$$
$$W_2 = \text{ブロック対角}[W_1^*, W_2^*, \cdots, W_a^*]$$

フーリエ表現ではパラメータベクトル θ_2 は P-1 より少ない項でモデル化することもできる．

C. 回帰モデル成分

説明変数が X_t の値をとる 1 つの時系列を考えると，その成分 DLM$\{F_{3t}, G_{3t}, 0, W_{3t}\}$ は $\{X_{3t}, 1, 0, W_{3t}\}$ となり，パラメータ θ_{3t} をつかうと，

$$\mu_{3t} = X_{3t}\theta_{3t} \tag{3.14}$$
$$\theta_{3t} = \theta_{3t-1} + w_{3t} \tag{3.15}$$

となる．一般に r 個の説明変数，$X_{it}\,(i=1, \cdots, r)$，があるとき，各々は 1 つの成分 DLM と，r 個の成分を組合せて作られている全体の回帰 DLM によって表される．計量経済分析では，線形回帰モデルの係数を過去のデータに当てはめて推定する．このモデルからたとえある関係が得られたとしても，その時点での係数の値が時間とともに将来も変化しないと保証できる理由はない．もし短い期間から推定された関係ならば，何年か後にはまったく変化してしまうかも知れない．従来のモデルでは，この状況はとらえることが難しいが，動学モデルでは「動く」関係をある程度表すことができる．ここで用いられるカルマンフィルターの計算方法は，新しいデータが利用可能になったときに最初から全部の計算をやり直したり，逆行列の計算をする必要がないので，回帰係数の推定値の修正を経済的に行えるという利点がある．この方法は以前 Plackett[1950] によって指摘されていたが，その実用上の重要性はほとんど見過ごされている．

4. DLM の予測と更新

本節では，実際の予測に必要な主要結果と更新関係を，DLM$\{F_t, G, V, W_t\}$ の場合，$(G_t = G, V_t = \sigma^2)$ について述べる．より詳細な説明や，多変量，非正規，非線形，その他の時間に依存するモデルへの一般化については参考文献を参照してほしい．実用上の重要な概念は割引情報の考え方である．これは各成分 DLM と一緒に用いると特に便利である．

以下で用いる Harrison-Stevens のモデルでは，構造方程式の攪乱項の分散は事前に既知であると仮定されている．また B 項で述べる情報割引という定式化が，ちょうど Holt-Winters のモデルにおける平滑化定数と同じ形で導入されている．これはかなり恣意的な方法であるが，非常に短いデータを用いる場合には有効な方法である．また，Harrison-Stevens の推定方法はあまり効率的ではないので，以下では，未知の平均と分散をもつ正規分布に対する同時共役事前分布に基づく，より効率的な方法 (DeGroot[1970] を参照) を用いている．

A. 主 要 結 果

2. A のモデル (2. 1) と (2. 2) の仮定のもとで，DLM$\{F_t, G, V, W_t\}$ は次のように表される．

$$y_t = F_t'\theta_t + v_t \tag{4.1}$$
$$\theta_t = G\,\theta_{t-1} + w_t. \tag{4.2}$$

ここで，$\mathrm{var}(v_t)=V_t=\sigma^2, \mathrm{var}(w_t)=W_t=\sigma^2 H_w$ とすると，初期時点において σ^2 が既知で

$$\theta_0 \sim N(m_0, \sigma^2 C_0) \tag{4.3}$$

と仮定すると，$t-1$ 時点ではその時点までのデータ D_{t-1} が与えられたときの θ_{t-1} の分布は

$$(\theta_{t-1}|D_{t-1}) \sim N(m_{t-1}, \sigma^2 C_{t-1}) \tag{4.4}$$

と表される．(2.4) の 8 ステップからなるカルマンフィルターをベイズ的に解釈した結果は，次のようにまとめることができる．

y_t が観測される以前では，

$$a_t = G m_{t-1} \tag{4.5a}$$

$$R_t = G C_{t-1} G' + W_t \tag{4.5b}$$

パラメータベクトル θ_t の事前分布は $(\theta_t|\sigma^2, D_{t-1}) \sim N(a_t, \sigma^2 R_t)$．ここで

$$f_t = F_t' a_t \tag{4.5c}$$

$$Q_t = F_t' R_t F_t + 1 \tag{4.5d}$$

とおけば，y_t に対する 1 期先予測分布は $(y_t|\sigma^2, D_{t-1}) \sim N(f_t, \sigma^2 Q_t)$．また

$$A_t = R_t F_t Q_t^{-1} \tag{4.5e}$$

とする．y_t が観測された以後では，1 期先予測誤差は，

$$e_t = y_t - f_t \tag{4.5f}$$

となるので，θ_t の事後パラメータベクトルは，次のように逐次的に更新される．

$$m_t = a_t + A_t e_t \tag{4.5g}$$

$$C_t = (I_t - A_t F_t') R_t = R_t - A_t Q_t A_t'. \tag{4.5h}$$

また事後分布は，$(\theta_t|\sigma^2, D_t) \sim N(m_t, \sigma^2 C_t)$ となる．

さらに σ^2 が未知の場合の σ^2 の事前分布 $(\sigma^2|D_0)$ は，次のような逆ガンマ分布 $IG(n_0/2, 2/n_0 S_0)$ で，

$$p(\sigma^2|D_0) \sim \sigma^{-(n_0+1)} \exp(-n_0 S_0/2\sigma^2)$$

と仮定しよう．ここで $X \sim IG(\alpha, \beta)$ とは，$Y=X^{-1}$ がパラメータ α と β のガンマ分布に従う意味である（鈴木[1987]参照）．この場合には逐次更新は，標準的なベイズ分析を利用して計算すれば，次のように表すことができる．

$$n_t = n_{t-1} + 1 = n_0 + t \tag{4.5i}$$

$$n_t S_t = n_{t-1} S_{t-1} + e_t^2/Q_t. \tag{4.5j}$$

したがって，分散は，$(\sigma|D_t) \sim IG(n_t/2, 2/n_t S_t)$ と分布する．ここで，$w_t=0$ で $G=I$ とすると，この関係は重回帰モデルにおける自然共役事前分析を逐次的に表わしたものと考えることができる（たとえば Zellner[1971] を参照）．

(4.5b)，(4.5d)，(4.5h) において σ^2 が未知の場合には，周辺事前分布，1 期予測分布，周辺事後分布は標準的なベイズ分析によって導くことができ，多変量 t 密度の形をしていることがわかる．なお将来予測分布と回顧的予測に対する平滑化された分布は次のようになる．

$k=1, 2, \cdots$ に対する予測分布は次式で与えられる．パラメータ分布は，

$$(\theta_{t+k}|D_t) \sim T_{n_t}[a_t(k), R_t(k)] \tag{4.6}$$

y_t の分布は，

$$(y_{t+k}|D_t) \sim T_{n_t}[f_t(k), Q_t(k)] \tag{4.7}$$

ここで，$a_t(k) = G a_t(k-1)$, ただし $a_t(0) = m_t$,

$$R_t(k) = G R_t(k-1) G' + W_{t+k},\ R_t(0) = C_t,$$

$$f_t(k) = F'_{t+k} a_t(k),\ Q_t(k) = F'_{t+k} R'_t(k) F_{t+k} + S_t$$

である．回顧的予測に対する平滑化された分布は，$k=1, 2, \cdots$ に対して，

$$(\theta_{t-k}|D_t) \sim T_{n_t}[a_t(-k), R_t(-k)] \tag{4.8}$$

ここで，$a_t(-k)$ と $R_t(-k)$ は次式のように逐次的に求められる．

$$a_t(-k) = m_{t-k} + C_{t-k} G' R_{t-k+1}^{-1} [a_t(-k+1) - a_{t-k+1}] \tag{4.9}$$

$$R_t(-k) = \{C_{t-k} - C_{t-k} G' R_{t-k+1}^{-1} \cdot [R_{t-k+1} - R_t(-k+1)] \cdot R_{t-k+1}^{-1} G C_{t-k}\} S_t / S_{t-k}. \tag{4.10}$$

B. 情報割引 (Information Discounting)

m 個の部分 DLM に対しては，$W_t=$ ブロック対角 $(W_{1t}, W_{2t}, \cdots W_{mt})$ であり，各 W_{it} は，Ameen and Harrison[1985] による割引法を用いて決められる．各パラメータ θ_{it} で表される第 i 成分に 1 つの割引率 $0 \leqq \delta_i \leqq 1$ が対応する．各成分の事後周辺分布は，

$$(\theta_{i,t-1}|D_{t-1}) \sim T_{n_{t-1}}[m_{i,t-1}, C_{i,t-1}] \tag{4.11}$$

であり，対応する周辺事前分布は，

$$(\theta_{it}|D_{t-1})\sim T_{n_{t-1}}[a_{it}, R_{it}]. \tag{4.12}$$

ここで，$a_i=G_i m_{t-1}$，$R_{it}=G_iC_{t-1}G_i'+W_{it}$, $W_{it}=(\delta_i^{-1}-1)G_iC_{t-1}G_i'$ とおいた．

したがって，$R_{it}=G_iC_{t-1}G_i'/\delta_i$ となる．そこで，θ_t に対する事前分布は，

$$(\theta_t|D_{t-1})\sim T_{n_{t-1}}[a_t, R_t] \tag{4.13}$$

となり，ここで $a_t=(a'_{1t},\cdots,a'_{mt})'$，そして $R_t=GC_{t-1}G'+W_t$ である．簡単には，将来何期か先の予測では，システム分散 W_{t+k} の将来値は，$W_{t+k}=W_{t+1}$ とおかれる場合が多い．この方法に関する一般的考え方は，D_{t-1} の下でパラメータベクトル θ_{t+k} に伴う精度は k とともに減少し，θ_t の精度は θ_{t-1} の δ の割合になるというものである．この意味で精度は δ の割合で割引される．$0\leqq\delta_i\leqq1$ と $G_i=1$ での回帰成分を考えると，$\delta_i=1$ ならば $W_{it}=0$ となり，将来の $t+k$ に対する $\theta_{i.t+k}$ についての情報は $\theta_{i.t-1}$ と同じになる．したがって，θ_{it} の値は期間中一定であり，i 番目の成分に関するシステム方程式は決定論的になる．他方，$\delta_i=0$ の場合は $\theta_{i.t-1}$ についての知識は将来値 $\theta_{i.t+k}$ についての情報をまったく含んでいないので，それに伴う成分システムはいらなくなる．これらの極端な値の中間では，D_{t-1} は δ_i が小さいほど i 番目の成分ベクトルについての情報を含んでいないことを意味する．したがって，割引率 δ_i は i 番目の成分ベクトルに関する値 $[m_{it}, C_{it}]$ が各期間でどのぐらい意味を持ち，永続性があるかということを表している．モデル化に際して割引率を1に近くするということは，決定論的なモデルの方を不確実なモデルより好むという意味から一般的である．しかしながら，適切でない決定論的モデルは，モデルがデータから乖離してもそれに適応しないので，予測は惨めな結果になることがある．

DLM において，モデル $\{F_t, G, V, W_t\}$ がシステム分散行列 W_t に関して定式化される場合，なにか目安がないと実際に適用するときに問題が生ずる．W_t は説明変数の計測単位に対して不変ではないのでこの値の決め方は明確ではない．そこでこのような割引という概念を導入し，それを成分モデルと結びつけることによって，単純な割引・指数加重回帰概念にもとづいたモデル作成が可能になる．この考え方は，最近の観測値をより重要視するという意味からも直感的に受け入れやすい方法に見える．モデルがトレンド・季節・回帰の各 DLM という3つの成分からなる場合では，別々の割引率を3つの各成分に関係づけることができる．通常，観測値系列の変動は，制御可能であるような独立変数にすべて帰することができる方がよい．したがって，大きく変動する回帰成分よりも安定的な回帰成分のあるモデルを使うほうが望ましい．またトレンド季節モデルでは，季節成分はトレンドモデルよりずっと安定的であるから，季節に対する割引係数はトレンドに対する割引係数よりも高い値をとる場合が多くみられる．

C. インターベンション分析

政策変更等の出来事についての外部情報にもとづくインターベンションがあると，系列にそれがはっきりと表れ，モデル式が実際と大きく乖離してしまう．実際には重要であるがモデルの構造に含まれていない事態が発生したとき，それをうまく取り込んで連続的に予測できることが必要である．インターベンションは，それが前もって分かっている場合でも分からない場合でも，次の形をとる．第1は外れ値（outlier）で，ある観測値がモデルの1期先予測値から著しく離れているが，他はモデルとそうは違っていないような場合である．もう1つは構造変化で，モデルのあるパラメータが突然変化したために起こる．たとえば水準が落ちたり，季節パターンの山と谷がずれたり，観測誤差の分散が増大したためにデータに表れる形である．インターベンションを処理するには次の2つの方法がある．第1にモデルを拡張して θ_t にいくらか加えるか，あるいは標準的な DLM にその乖離を説明する部分モデルを加え，F と G をそれにしたがって修正する．第2に，外れ値や特定の変化を現在のモデルで説明できるものとして考える．原因はなんであれ，大きく外れた観測値は標準的な方法で処理すると分析に悪影響を及ぼす場合がある．時間 t を越えて予測するとき，外れ値を扱うもっともよい方法はそれを無視することである．したがって更新分析は D_{t-1} のもとでの現在の事前分布を，y_t を無視して D_t の下での事後分布として考えるように修正される．すなわち，$m_t=a_t$, $C_t=R_t$, $n_t=n_{t-1}$, $S_t=S_{t-1}$ とする．構造変化に関しては，季節パターンや回帰項は，外生的な影響に対して安定的であるので，突然変動した水準や成長についての情報をさまざまな方法で直接モデルに含

めるようにする．ベイズ予測の特徴は，このような例外的な事態に対する主観的情報をごく自然な方法で取り入れることができる点にある．

外れ値を除いたり分散が大きくなる場合をモデルに含めるという考え方は，それを発見し，識別するモデル監視の方法と結びつけると，例外的な事態を自動的に調整する方法として使うことができる．この方法は，①主観的な予想インターベンションが適切な外部情報がなかったために作動しなかったときは自動的に無視し，②短期予測への応用では，多数の系列を平行的に分析できる単純で頑健なモデルの採用が可能で，③外れ値や構造変化点がモデル推定に影響を及ぼすような時系列分析に役立つものである．もう1つの方法は多段階モデルにもとづいたもので同じ様な能力があるが，比較的小さなモデルでさえもかなり計算費用がかかるという欠点がある．ルーティン的なモデル推定，データ平滑化，短期予測には単純なものの方がよい．自動化の方法には，外れ値や構造変化を識別することによって予期せざる事態を発見する，外れ値を捨て変化に適応することによって予期せざる事態を処理する，という2つの要素がある．前者に対しては，モデルの予測パフォーマンスを適当な比較的一般的な対立モデルと較べて一連のベイズファクターを使って評価する．しかし，乖離が生じたときそれは構造変化の始まりによるものかも知れない．その場合に，その時点で小さな割引ファクターを適用することによって外れ値を落とすと，トレンド，季節性，そして回帰に対する不確実性が増大してしまう．大きなベイズファクターは構造変化が最近起こったことを示しているが，それは大きな予測誤差によるためかも知れない．そこで，そのような変化は，次のような累積ベイズファクターによって調べることができる．

$$L_t(k) = H_t H_{t-1} \cdots H_{t-k+1} = H_t L_{t-1}(k-1) \qquad (k = 1, 2, \cdots, t) \tag{4.14}$$

これは，モデルの予測パフォーマンスを対立モデルのそれと，それ以前の連続する k 期の観測値について較べてみたものである（詳細は West[1986] を参照）．

5. DLM 分析の応用例

前節までに述べられた DLM モデルによるベイズ予測の方法を，いくつかの例によって示すことにしよう．実際の分析では，モデルの定式化と $t=0$ における初期事前分布，そして割引率の値を決めることが必要になる．$t=1, \cdots, T$ と表される等時間間隔で観測された一変量時系列データをモデル化する．各時点 t で，y_t はレベル (μ_t) と観測誤差あるいはノイズ項 (v_t) という2つの成分の和として表される．

$$y_t = \mu_t + v_t \tag{5.1}$$

レベルはさらに，局所的になめらかな多項式トレンド (p_t)，季節性 (s_t)，回帰項 (r_t) の3つの成分からなっている．したがって，加法モデルでは，

$$\mu_t = p_t + s_t + r_t \tag{5.2}$$

あるいは，乗法モデルでは，

$$\mu_t = (p_t + r_t) s_t \tag{5.3}$$

の形をとる．時系列は観測誤差，トレンド，季節，回帰の各成分に分解される．誤差項は平均ゼロの正規確率変量で，レベルに関して純粋に確率的で予測できない変動を表している．この誤差項は過去の情報 D_t とは独立である．レベルに対する変動の程度は誤差分散によってコントロールできる．

$$V[v_t] = c\mu_t^a \tag{5.4}$$

ここで，a は観測誤差分散のスケールパラメータで，a は分散のべき乗部分を表す指標である．$a=0$ なら一定分散であり，$a>0$ なら系列のレベルとともに確率誤差の分散が増加する．トレンド項は多項式の次数で表し，常に存在し，通常1次あるいは2次の多項式が用いられる．周期が P である季節性をモデル化すると，s_t は t 期での季節因子を表す．加法モデルでは，P 個の季節因子はレベルからの季節的な偏差を表し，すべての t で P 個の季節因子は識別のために和がゼロとされる．乗法モデルでは季節因子は季節乗数であり，和が P となる制約が課される．回帰項については，r_t が独立変数に対する線形回帰になり，β_t を t 期での回帰係数とすると $r_t = \beta_t X_t$ となる．回帰係数にはきわめて小さな確率変動を認め，そのために添え字 t がつ

けられている．したがって，モデル関係は固定的ではなく，モデルに定式化の誤りを認めることになる．

t 期における過去の情報 D_t はモデルの各成分を表すダイナミックなパラメータに対する事後分布で要約される．これらはモデルの各部分に対してはステューデント-t 分布であり，c に対しては基準化された逆カイ2乗分布（より一般的には逆ガンマ分布）とする．時間が $t+1, t+2, \cdots$ とたつにつれて各要素は変化し，新しいデータが得られると情報セットを更新しながらこれらの事後分布を更新する．これらの次々と得られる各成分の事後点推定値（各時点 t に対する事後モード）は，パラメータに対する推定された時間経路を表している．事後分散はこれらの推定値における不確実性の尺度になる．同じような軌跡は，多項式トレンド，回帰効果，季節性の逐次推定値があれば時系列の各成分に対して同様に得られる．たとえば，回帰係数 β_t の軌跡は $(\beta_t|D_t), t=1, \cdots, T$ の連続する事後ステューデント-t 分布で決められ，このモード，スケール，自由度は時間がたつに連れて連続的に更新される．モデル評価にとって重要なのは，1期先予測誤差が作られる1期先予測分布の系列と，それに伴うステューデント-t 予測確率密度である．

回顧的な評価は，系列の最後 $t=T$ ですべての利用できるデータを使ってから，モデルパラメータの平滑化された推定値とすべての情報 D_T のもとで，各時点 t での平滑化分布から導かれる成分をもとに行う．したがって，回帰係数の軌跡の回顧的評価はステューデント-t 分布 $(\beta_t|D_T)$ をもとにしている．最後に，予測値はいつの時点でも求められる．t 期での将来観測値 y_{t+k}, $(k>0)$ の予測分布は，D_t で表されるモード，スケール，自由度を持つ標準ステューデント-t の形をしている．

データを処理する以前の D_0 という条件の下で，事前分布の初期値をモデルの各成分に与える．事前分布は，事前モードあるいは事前平均と不確実性の尺度の推定値を各パラメータに対して与えればよい．これらの事前分布は，自由度 n_0 のステューデント-t 分布と，分散スケールパラメータ c に対する逆カイ2乗，あるいは一般的には逆ガンマ事前分布である．不確実性の尺度はモデルパラメータの標準偏差とする．ただし，これは，事前分布が基本的に正規の時は，これらの値は n_0 が非常に大きくない限り実際の標準偏差ではない．その他の場合はステューデント-t 事前分布のスケールパラメータを与えることになり，実際の標準偏差は $\sqrt{n_0/(n_0-2)}$ を掛けて得られる．観測スケール c に対する逆カイ2乗事前分布の初期値は，点推定値と自由度 n_0 を与えると決まる．

モデルパラメータが時間とともに確率的に変化する率が割引率で，モデルの各要素ごとに決められる．レベル，成長，季節，回帰パラメータにおける確率変化の程度は，各時点でのこれらの値について得られる情報に直接関係している．すなわち，観測値間での確率的な動きは各パラメータに対する事後分散の増加，すなわち情報の減少に等しくなる．したがって0.95という割引率は約5%の情報の減少，あるいは分散で測った不確実性の5%増大を意味する．この割引率が1というのは，対応する部分に時間を通じて確率変動がない，あるいはその部分のパラメータがスタティックであることを意味している．以下で，このような数値の例とその計算結果をグラフ表示したものを示す．

ケースI

データは，食料品の輸入数量指数（外国貿易概況）と円ベースでの食料品輸入価格指数（物価指数月報）の系列である．これらのデータは標準的なベイズ分析をどのように行うかを示すための例で，1973年第1四半期から1987年第4四半期までの四半期データである．散漫な事前分布を用いた分析で特定の事前分布を考える前の参考となるような予測を行う．プロパーな情報量のある事前分布を用いると，短い時系列データからも予測可能であり，短期予測パフォーマンスが向上する．

最初に各系列のデータ分析を行う．各系列について，平均，分散，モードなどの要約統計量並びにヒストグラムなどからデータの分布状況を見る．さらに自己相関関係をチェックすることも必要である．必要ならデータ変換を行う．データがどのような動きをしているのかを見るために系列をプロットする（図1）．データプロットを見ると食料品（数量）にはっきりとした季節性が認められる．両変数間の散布図（図2）を見ると負の関係がみられる．そこで，この例での食料品に対するモデルは次のような構造をしていると考

える．一定の（ステディ）トレンド成分，加法的季節要素の成分，そして価格変数に対する回帰成分である．このモデルの加法的季節モデル要素は周期が4で，調和項は2つである．最後にモデルの各要素についての割引率を決める．この例では，レベル（水準）0.90；季節 0.97；回帰 0.95；分散 0.99 としている．モデルのパラメータに対して情報がない，すなわち散漫な事前分布を基に行われた分析は図3である．観測値は順に処理され，各時点で1期先予測分布が計算される．モデルに関する事後分布の要約とその要約統計量は表1にあり，最後の53個の観測値をもとに計算される．

次にこのモデルのパラメータに適切な事前分布として次の値を与えてみる．

レベル（水準）：平均 70，標準偏差 10

季節ファクター：平均 −10，2，4，4 で標準偏差はすべて 2

価格変数の回帰係数：平均 −0.20，で標準偏差は 0.2

分散：推定値 0.15，自由度 5

ここで，標準偏差を大きくとると散漫な事前分布の場合に近いものになり，不確実性の範囲が比較的広いものになる．標準偏差は名目的な正規事前分布で定義されるにもかかわらず，事前分布は実際はステューデント-t 分布であり，したがって正規性で考えるよりいくらか散漫になる．この事前分布にもとづいて当てはめられたモデルの事後分布の要約とその要約統計量は表2の通りであり，その1期先予測分布は図4である．散漫な事前分布によるものより予測パフォーマンスはいくらか改良されていることが分かる．次の図5から図6は，モデルの各要素の1期ずつのデータを順に処理したオンライン推定値を表しており，その時点までに処理したデータをもとに計算した各時点での推定値である．図5は説明変数を除いたトレンドで，モデルの推定された1次多項式の成分（各観測値での定数項）のプロットである．図6は，説明変数を含めたトレンドで，推定された各観測値での定数項プラス価格変数の回帰項，すなわちこのモデルのオンラインでの計算された季節調整済み系列のプロットになる．図7は季節成分の推定値で，推定された季節パターンとその標準偏差を表している．図8はその季節要素で，データから推定された最後の12ヵ月の季節ファクターとその標準偏差である．図9は価格変数に対する係数の推定された時間をおっての軌跡を表した係数プロットであり，各係数の安定性や反応の変化が分かる．この例の場合は価格変数の係数が，最近では僅かながら小さくなる傾向にある．さらに，分散スケールパラメータの推定値や，1期先予測誤差，そして回帰変数の従属変数に対する毎期毎期の影響を評価することによってモデルを改良することができる．

次はオンライン推定値の代わりに，平滑化推定値を用いたグラフである．平滑化推定値は，すべてのデータをもとにしたデータ系列を回顧的に改訂した点推定値であり，オンライン推定値に較べて滑らかになる．図10は，当てはめられた値とステューデント-t 分布にもとづく90％確率限界である．次は，図5と図6に対応した平滑化された説明変数を除いたトレンド（図11）と説明変数を含めたトレンド（図12）である．平滑化された全体の季節パターンと，季節ファクターも同様にして得られる．図13は，1期先予測ステューデント-t 分布のスケールパラメータで標準化された1期先予測誤差のプロットである．この例では，最近時点での当てはまりがあまりよくない．

1987年第4四半期までの数値を基に，3年先（1980年第4四半期まで）を予測して得られた予測値が図14である．モデルは回帰項を含んでいるので，将来3年間の価格変数の値を与える必要がある．この例では価格変数はこの期間変化しないとして計算してある．将来予測分析では，この価格変数の将来値の値をいろいろ変えて分析する．

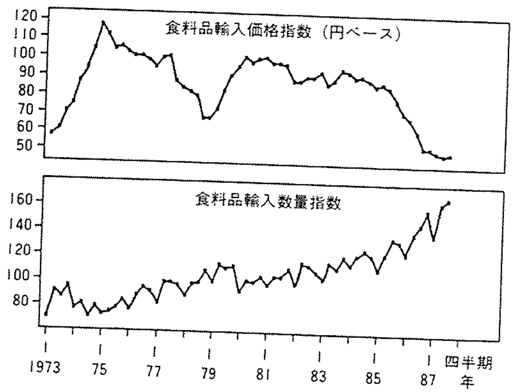


図1 原系列のプロット

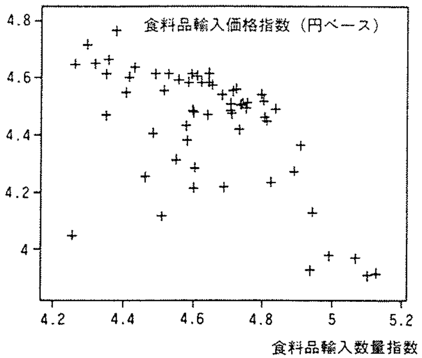


図2 散 布 図

成 分	平 均	標準偏差
レベル（水準）	169.326	6.497
季節要素	-10.964	2.161
	2.766	2.112
	4.697	2.083
	3.501	2.252
価格変数	-0.211	0.117
平均2乗誤差		46.40
平均絶対誤差		5.430
対数予測密度		-2173

表1 事後分布の要約表(散漫な事前分布)

成 分	平 均	標準偏差
レベル（水準）	176.673	8.829
季節要素	-12.797	1.433
	2.973	1.361
	4.285	1.416
	5.540	1.524
価格変数	-0.248	0.173
平均2乗誤差		38.05
平均絶対誤差		4.899
対数予測密度		-2982

表2 事後分布の要約表(報知的事前分布)

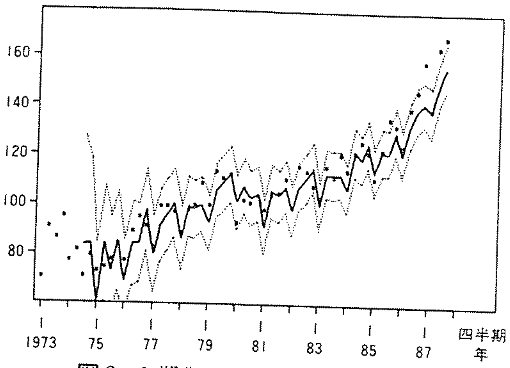


図3 1期先予測(散漫な事前分布)

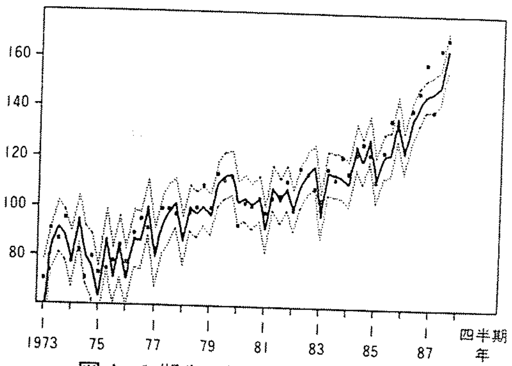


図4 1期先予測(報知的事前分布)

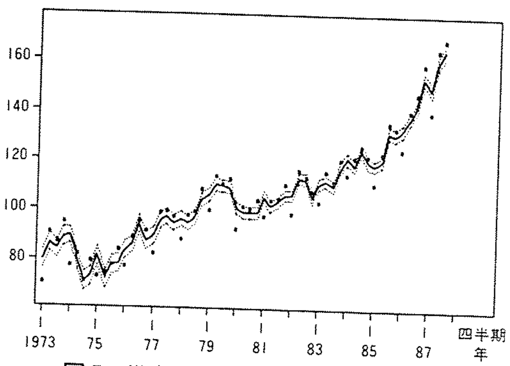


図5 推定されたオンライン・レベル

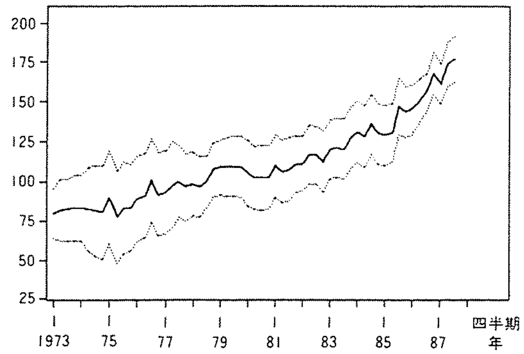


図 6 推定されたオンライン・回帰項とレベル

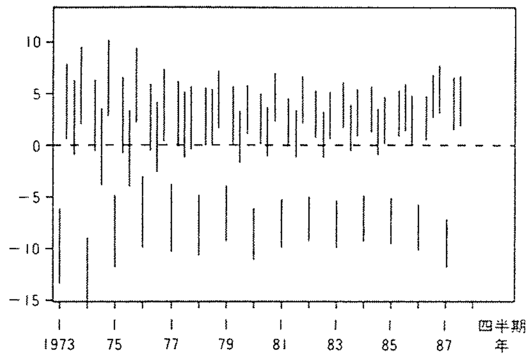


図 7 推定されたオンライン・季節項

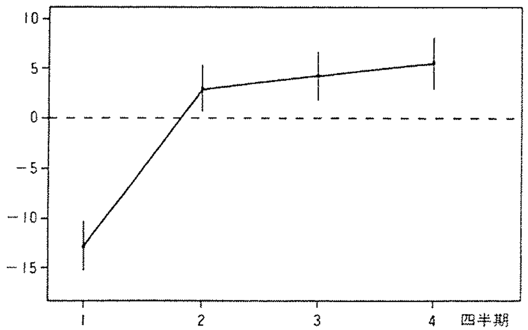


図 8 推定された季節ファクター

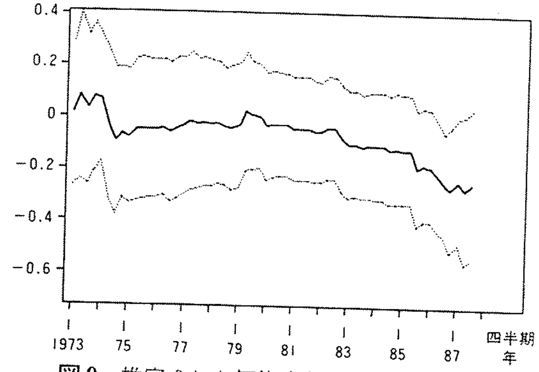


図 9 推定された価格変数の係数プロット

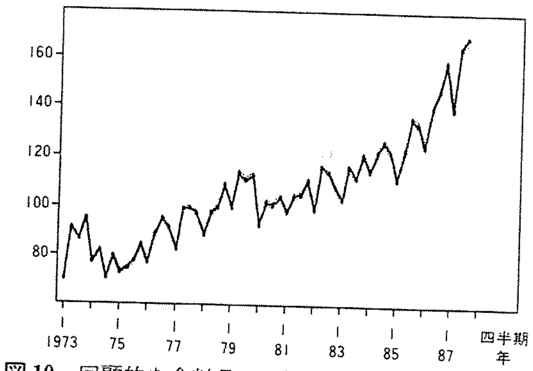


図 10 回顧的な食料品の 1 期先予測と 90% 信頼区間

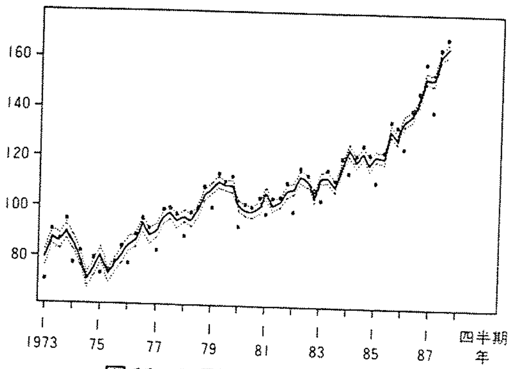


図 11 回顧的な平滑化レベル

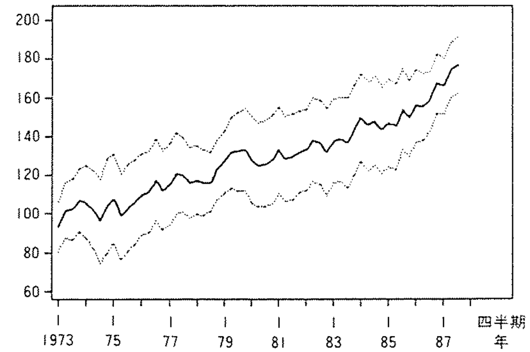


図 12 回顧的な平滑化回帰項とレベル

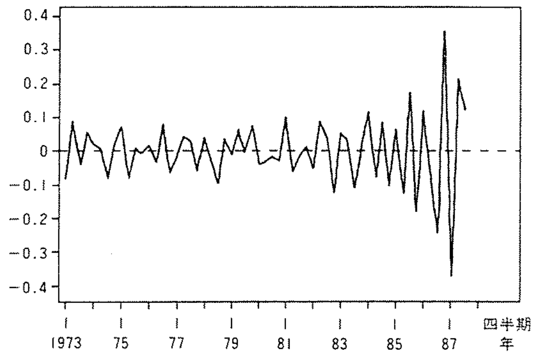


図 13 回顧的な平滑化標準化残差

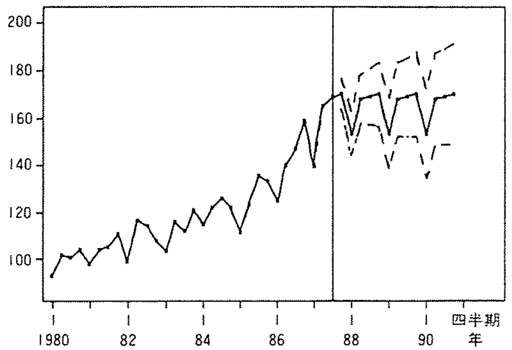


図 14 食料品データの将来予測

ケース II

次の例は，インターベンション分析の例である．図 15 のような販売データを用いる．データは 1955 年 1 月から 1960 年 12 月までのある商品の月次販売高である．最初は前もって決められた時点でのインターベンションを行う例である．データのプロットより，このデータには 3 つの局面が含まれており，各局面でははっきりとしたレベルの変化があることが分かる．この例では，予測する人にはさらに次の付加的な事前情報があったと仮定しよう．

- 1955 年 12 月に，記録ミスがあった．
- 1957 年 1 月にはレベルは上昇したが成長率は変わっていないように見える．
- 1958 年 1 月には 1 年前と同様な動きがあるようだが，今度は不確実性が増している．

モデルは，トレンドの部分を線形（レベルと成長）とする．初期の事前分布は，レベル平均が 600，で標準偏差が 100；成長平均が 0 で，標準偏差が 10；分散推定値が 1.5% で自由度は 5 である．割引ファクターは，レベルが 0.90 で分散を 0.98 としている．ここで次のようにしてこの情報をモデルに含める．

(1) モデルを定式化する段階で，3 つの予め分かっているインターベンションの時点を，1955/12，1957/1，1958/1 としモデルに含める．

(2) モデルを当てはめていく段階で各々のインターベンションの時点の手前で分析を止めインターベンションを確認する．このデータを分析に用いるか，外れ値としてモデルから外すかを決める．

(3) 最初の 1955/12 の時点では，この観測値を欠測値とし分析を進める．1 期先予測をプロットした図 17 では，この点は欠測値となっている．

(4) 2 番目のインターベンション時点，1957/1，ではこれを分析に含め，次の事前分布を設定する．レベル平均が 900 で，標準偏差が 20；成長の平均が 0 で，標準偏差が 5 である．

(5) 最後のインターベンション時点，1958/1，では事前情報を次のように変更する．レベル平均が 940，で標準偏差が 50；成長平均が −2 で，標

準偏差が7である．

図16はこのインターベンションを無視してモデルを当てはめたときの予測値であり，図17は上記のようにして計算された予測に対するインターベンションの結果である．図16と比較すると，外れ値は無視されレベル変化ははっきりと処理されていることがわかる．回顧分析では変化点でのインターベンションに対する反応がよりはっきり表れているので，その方を示すことにする．図18と図19は平滑化されたトレンドと成長で，レベルではインターベンション時点で2回上昇変化があり，成長率は1957/1で大きく変化しているのが分かる．図20はその標準化残差で，点線で示された上側と下側限界はそのt分布からの，ゼロの回りで対称な90%確率区間を表している．

次に，自動的な予測監視とインターベンション機能を組み込んだ分析の例である．この例では，同じ事前情報を用いている．主な違いは，割引ファクターに関するもので，モデル監視機能が指示したときに新しいデータをすばやく採り入れるかどうかを決める．自動監視の場合の割引ファクターは，レベルを0.10，分散を0.90としてある．図21はこの分析による1期先予測で，図の横軸にモデル監視機能からのしるしが付けられてある．図22は自動モデル監視機能を働かした場合での回顧分析であり，図23と図24は前の分析での図19と図20に対応するものである．このモデルでの1期先回顧予測誤差は図25に，そして1年先までの予測値は図26にある．

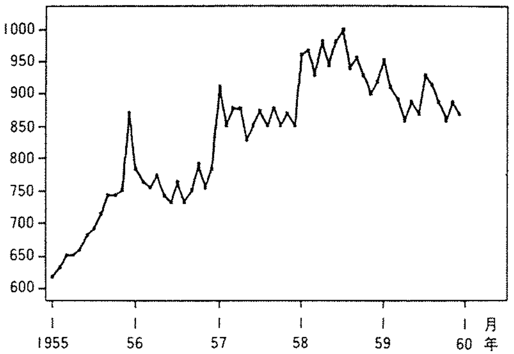


図15　販売データのプロット

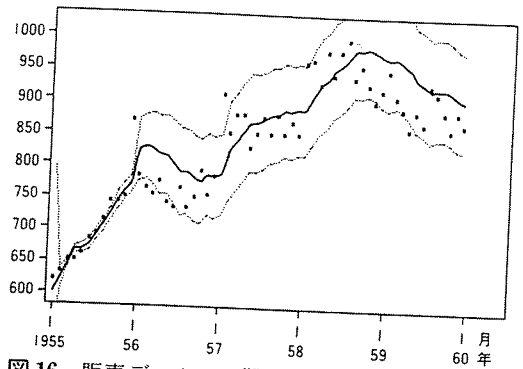


図16　販売データの1期先予測と90%信頼区間

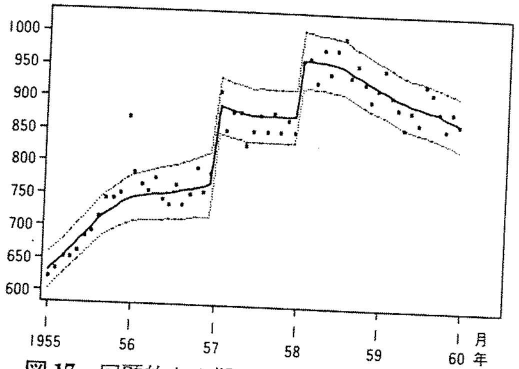


図17　回顧的な1期先予測と90%信頼区間

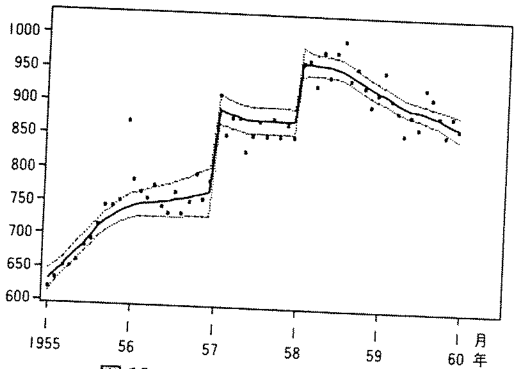


図18　回顧的な平滑化レベル

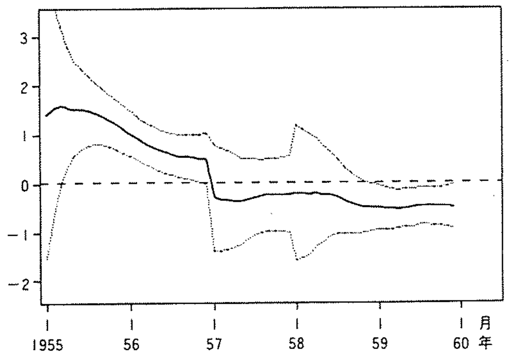


図 19 回顧的な平滑化成長率

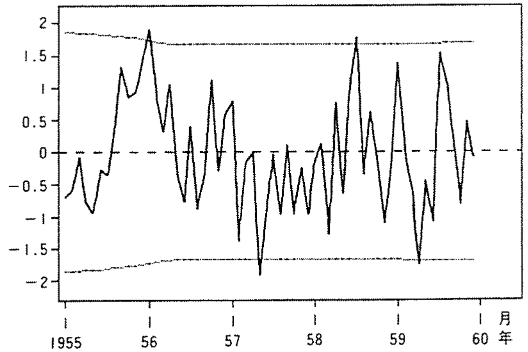


図 20 回顧的な平滑化標準化残差

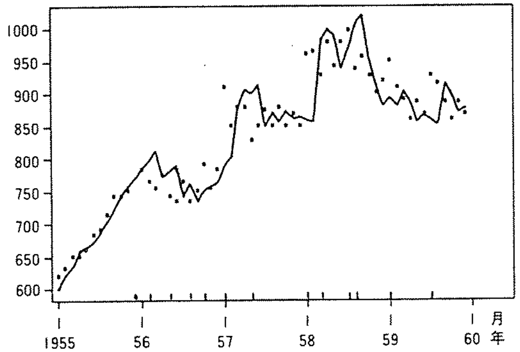


図 21 自動的予測監視による 1 期先予測

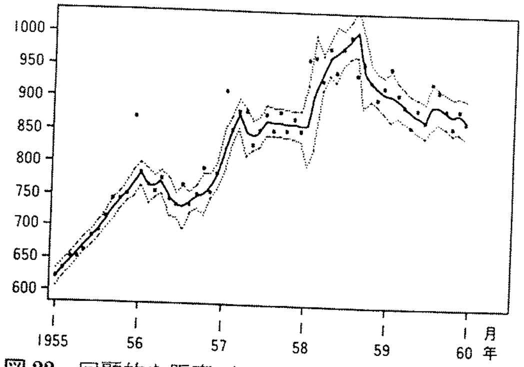


図 22 回顧的な販売データの 1 期先予測(自動的)

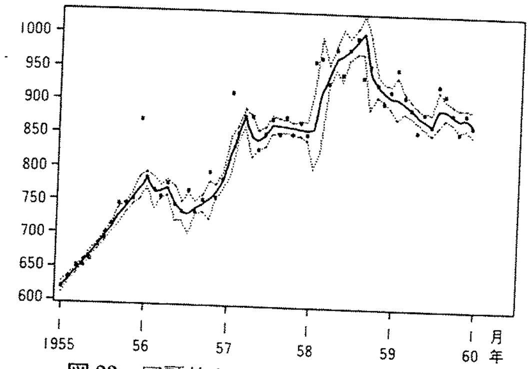


図 23 回顧的な平滑化レベル(自動的)

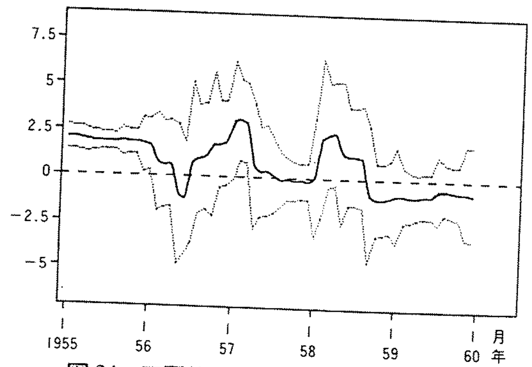


図 24 回顧的な平滑化成長率(自動的)

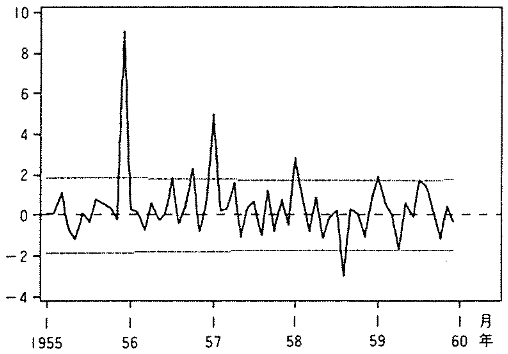


図 25　回顧的な平滑化標準化残差(自動的)

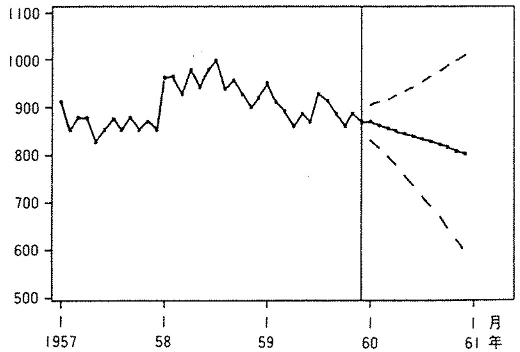


図 26　販売データの将来予測(自動的)

6. 多変量モデルへの一般化

前節までに述べた単変量 DLM を多変量時系列データを取扱うモデルに拡張し，一般化することができる．多変量時系列モデルは，Tiao and Box [1981], Chan and Wallis[1978], Jenkins and Alavi[1981] そして Abraham [1980] などで取扱われているが，いずれも単変量 ARIMA モデルを一般化したものである．これは，構造モデルがトレンドや季節性といった直接に解釈可能な成分で定式化されているのに対して，先験的な制約が ARIMA モデルにおかれるので誘導型と見なされる．基本的な構造時系列モデルは，そのまま多変量の場合に一般化することができる．以下で，結果だけを簡単に示すことにする．

モデルは，前と同じように k 個の変数について次のように表す．

$$y'_t = F'_t\theta_t + \varepsilon'_t \tag{6.1}$$

$$\theta_t = G_t\theta_{t-1} + u_t \qquad t=1, \cdots, T \tag{6.2}$$

ここで y_t は k 個の従属変数についての $k\times 1$ ベクトルで，F_t は $p\times 1$ 説明変数ベクトル，θ_t は $p\times k$ の係数行列で，G_t は $p\times p$ の既知の定数行列である．ε_t は $k\times 1$ 観測値ノイズベクトルで，u_t は $p\times k$ のシステム攪乱項行列である．誤差分布に関して次の仮定をおくことにする．

$$\varepsilon_t \sim NID(0, v_t\Sigma) \tag{6.3}$$

$$\mathrm{vec}[u_t] \sim NID(0, H_t), \qquad H_t = \Sigma\otimes W_t \text{ で } \varepsilon_t \text{ とは独立} \tag{6.4}$$

ここで，$\otimes$ はクロネッカー積を表す．係数の初期時点の事前分布は，

$$p(\mathrm{vec}[\theta_0]|\Sigma, D_0) \sim N(\mathrm{vec}[m_0], \Sigma\otimes C_0) \tag{6.5}$$

Σ の $k(k+1)/2$ 個の異なる要素は，次のような逆ウィシャート分布をすると考える．

$$p(\Sigma|D_0) \propto \det(\Sigma)^{-\eta/2}\exp(-tr[\Sigma^{-1}S_0]/2) \tag{6.6}$$

ここで，$\eta=n_0+k$ であり，S_t は $k\times k$ 対称行列である．これらの仮定の下で単変量の場合と同様にして多変量カルマンフィルター逐次計算は次のように表される：

(i)　t 期で y ベクトルを観測する以前

$$a_t = G_tm_{t-1}, \tag{6.7a}$$

$$R_t = G_tC_{t-1}G'_t + W_t. \tag{6.7b}$$

したがって，$(\mathrm{vec}[\theta_t]|\Sigma, D_{t-1}) \sim N(\mathrm{vec}[a_t], \Sigma\otimes R_t)$

$$f_t = F'_ta_t, \tag{6.7c}$$

$$Q_t = v_tI + F'_tR_tF_t. \tag{6.7d}$$

したがって，$(y_t|\Sigma, D_{t-1}) \sim N(f_t, \Sigma\otimes Q_t)$

$$A_t = R_tF_tQ_t^{-1}. \tag{6.7e}$$

(ii)　t 期で y ベクトルを観測した以後

$$e_t = y_t - f_t, \tag{6.7f}$$
$$m_t = a_t + A_t e_t', \tag{6.7g}$$
$$C_t = (I - A_t F_t') R_t. \tag{6.7h}$$

したがって，$(\mathrm{vec}[\theta_t]|\Omega, D_t) \sim N(\mathrm{vec}[m_t], \sum \otimes C_t)$

$$n_t = n_{t-1} + 1 \tag{6.7i}$$
$$S_t = S_{t-1} + e_t e_t'/Q_t \tag{6.7j}$$

したがって，$(\sum|D_t) \sim IW(S_t, n_t)$ と分布する.

v_0 と行列 θ_0, C_0, S_0 と $W_t, t=1, \cdots, T$ のパラメータを決める必要がある．$\sum$ の $k(k+1)/2$ 個のユニークなパラメータに対する多変量散漫事前分布は，v_0 をゼロ，S_0 をゼロ行列とすると得られる．多変量構造モデルにおける個々の時系列データは，同じ形の k 本の単変量構造モデルを考えればよい．k 本の単変量モデルは，これらの方程式の観測誤差，e_{it}，が時間に関して独立ならば多変量システムとして書くことができる．これらの誤差間の同時点相関は，$\sum$ の非対角要素に表れる．

単変量の場合と同様に周辺事前分布，予測密度，事後密度を導出することができる（Zellner[1971]ch. 8 参照）．周辺事前分布は，

$$p(\theta_t|D_{t-1}) \propto (\det[S_{t-1} + (\theta_t - a_t)' R_t (\theta_t - a_t)])^{-\gamma/2} \tag{6.8a}$$

ここで，$\gamma = n_{t-1} + p$，すなわち，一般化多変量ステューデント-t 分布をする．

1 期先予測密度は，

$$p(y_t|D_{t-1}) \propto [Q_t + (y_t - f_t)' S_t^{-1} (y_t - f_t)]^{-(\eta+k)/2} \tag{6.8b}$$

ここで，$\eta = n_{t-1} - k + 1$，すなわち，多変量ステューデント-t 分布をする．

事後密度は，

$$p(\theta_t|D_t) \propto (\det[S_t + (\theta_t - m_t)' C_t (\theta_t - m_t)])^{-\tau/2} \tag{6.8c}$$

ここで，$\tau = n_t + 1$ である．

任意の t に対して $W_t \neq 0$ の時，(6.1) と (6.2) のモデルは係数が時間で変化するものになる．これはモデルを新しい状態に対してある意味で適応するようにする 1 つの方法である．これを示すために，多変量での (6.7g) 式を見てみよう．これらの方程式ではパラメータの事後分布の平均は，1 期先予測誤差の関数を事前分布の平均に加えることによって得られる．事前分布の平均がこれらの誤差によって修正される大きさは，しばしばカルマンゲインと呼ばれる A_t に依存している．(6.7c) と (6.7e) における f_t と A_t の定義を用いて，事後平均 m_t を次のように表すことができる．

$$m_t = (R_t^{-1} + F_t F_t')^{-1} (F_t F_t' + R_t^{-1} a_t) \tag{6.9}$$

この式は，事前精度が増加するにつれて，すなわち R_t^{-1} の対角要素が大きくなるにつれて，事前平均 a_t は m_t を求めるさいにデータを支配 (dominate) し，極限では $m_t = a_t$ となることを表している．逆に言えば，事前精度が低下するにつれて，t 期において観測される新しいデータはより重要になる．(6.7b) より W_t が大きくなれば R_t^{-1} は小さくなることがわかる．したがってゼロでない W_t は，古典的な固定パラメータの線形モデルの場合における重要性に比べて過去のデータの重要性を割り引く効果を持っているといえる．前述したように，Ameen and Harrison[1985] ではこの割引法を (6.2) 式の u_t の共分散行列と結び付けた W_t 行列を割引要素行列と置き換えたモデルを用いている．このモデルは W_t を次のように定義すると，単変量の場合と同様に，いわゆる正規割引(normal discount) ベイズモデルの形になる．

$$W_t = \Delta_t^{-\frac{1}{2}} C_{t-1} \Delta_t^{-\frac{1}{2}} - C_{t-1} \tag{6.10}$$

ここで，Δ_t は $0 \leqq \Delta_{iit} \leqq 1$ のような割引要素からなる対角行列である．(6.10) 式を (6.7b) に代入すると，

$$R_t = \Delta_t^{-\frac{1}{2}} C_{t-1} \Delta_t^{-\frac{1}{2}} \tag{6.11}$$

となり，$\Delta_t = I$ のとき $W_t = 0$ である．応用にさいしては通常 W_t を直接与える代わりに，すべての t に対して $\Delta_t = \Delta$ とおかれる．

7. おわりに

ここで述べられたベイズ予測モデルは，「基本的構造モデル」を基に定式化されている．このモデルを使ってある変数についての将来分析と回顧分析を行うことによって，意思決定と予測が関連する「もしこうしたら，どうなるか」という分析と「何が起こったのか」という分析をすることができる．このような回顧的な分析を行う意味は，単にある出来事の影響を記述するだけでなしに，それから学びとることによって似たような将来の状況に対する対

処の仕方を身につけることができることにある．この学習という側面は，この種の予測システムでは非常に重要であり，モデルを改良し適切なものへ変えることも目的の１つであるといえる．そしていろいろな人とコミュニケーションをするためには，結果を数値情報だけでなくグラフィック表示することも重要である．多変量モデルへの一般化の方法についての概要は述べたが，応用はまだあまりされていない．多変量時系列モデルの応用では，多くの研究で利用されている Sims[1980] によって提唱された VAR モデルが，経済予測の分野でその実用性が認められている．このモデルにもとづいた Litterman[1981] によるベイズ VAR モデルは，制約を穏やかに導入することによって予測に有効なモデルを提供しようとする方法である．VAR モデルは，一種の誘導型方程式ともいうべきものであるので，個々のパラメータについての経済学的解釈は明確ではない．したがって，ある意味では，Litterman[1980, 1984] や Doan, Litterman and Sims[1984] で開発されたベイズ VAR の方法は，前節までに扱った構造状態空間モデルと競争的なモデルとして見ることができるが，潜在的にはマクロ経済予測という目的には利用価値があるように思える (Litterman[1986] を参照)．このベイズ VAR モデルを状態空間モデルの形に書き直して用いると，Litterman 他によって採用された VAR の方法をもっと明瞭にし，拡張し，ある意味では単純化することができる．しかしながら，このベイズ VAR にはいくつかの問題点がある．紙数の関係でその詳細を述べることはできないが，たとえば係数行列への事前分布の与え方や，パラメータの推定そしてモデル選択の問題があげられるが，それに関しては，Litterman[1980] や Doan, Litterman and Sims [1984] などでいくつかの方法が提案されている．また，モデルをもっと一般的で実用的にするためには，高速で有効な数値計算の方法とグラフィック表示によるコミュニケーションの方法の開発が必要となろう．

参 考 文 献

Abraham, B., 1980 "Intervention Analysis and Multiple Time Series", *Biometrika*, 67, 73-78.

Ameen, J.R.M. and Harrison, P. J., 1985 "Normal Discount Bayesian

Models," In J. M. Bernardo, M.H. DeGroot, D.V. Lindley and A. F. M. Smith, (eds.) *Bayesian Statistics 2*, North-Holland, Amsterdom.

Anderson, B. D. O. and Moore, J. B., 1979 *Optimal Filtering*, New Jersey, Prentice-Hall.

Box, G. E. P. and Jenkins, G.M., 1976 *Time Series Analysis : Forecasting and Control*, Holden-Day, San Francisco.

Chan, W. Y. T. and Wallis, K. F., 1978 "Multiple Time Series Modelling: Another Look at the Mink-Muscrat Interaction", *Applied Statistics*, 27, 168-175.

DeGroot, M. H., 1970 *Optimal Statistical Decisions*, McGraw-Hill.

Doan, T., Litterman, R. and Sims, C., 1984 "Forecasting and Conditional Projection Using Realistic Prior Distributions (with Discussion)", *Econometric Reviews*, 3-1, 1-144.

Engle, R. F. and Watson, M., 1981 "A One Factor Multivariate Time Series of Metropolitan Wage Rate", *Journal of American Statistical Association*, 76, 774-781.

Engle, R. F. and Watson, M., 1987 "The Kalman Filter: Application to Forecasting and Rational-Expectations Models", Truman F. Bewley (ed.), *Advances in Econometrics : Fifth World Congress*, 245-284.

Gersch, W. and Kitagawa, G., 1983 "The Prediction of Time Series with Trend and Seasonalities", *Journal of Business and Economic Statistics*, 1, 253-264.

Harrison, P. J., 1967 "Exponential Smoothing and Short-term Sales Forecasting", *Management Sciense*, 13, 102-139.

Harrison, P. J. and Stevens, G. F., 1971 "A Bayesian Approach to Short Term Forecasting", *Operation Research Quarterly*, 22, 341-362.

Harrison, P. J. and Stevens, G. F., 1976 "Bayesian Forecasting (with discussion)", *Journal of the Royal Statistical Society Ser. B.*, 38, 205-247.

Harvey, A. C., 1981 *Time Series Models*, Phillip Allan (国友直人・山本拓訳『時系列モデル入門』東京大学出版会，1985 年).

Harvey, A. C., 1984 "A Unified View of Statistical Forecasting Procedures", *Journal of Forecasting*, 3, 245-275.

Harvey, A. C. and Todd, P.H.J., 1983 "Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study", *Journal of Business and Economic Statistics*, 1-4, 299-315.

Jenkins, G. M. and Alavi, A. S., 1981 "Some Aspects of Modelling and Forecasting Multivariate Time Series", *Journal of Time Series Analysis*, 2, 1-48.

Kalman, R. E., 1960 "A New Approach to Linear Filtering and Prediction Problems", Transaction of ASME, Series D, *Journal of Basic Engineering*, 82, 35-45.

Kitagawa, G., 1981 "A Nonstationary Time Series Model and its Fitting by Recursive Filter", *Journal of Time Series Analysis*, 2, 103-116.

Kitagawa, G. and Gersch, W., 1984 "A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality", *Journal of American Statistical Association*, 79, 378-389.

Kitagawa, G. and Gersch, W., 1985 "A Smoothness Priors Time Varying AR Coeffiicient Modeling of Nonstationary Covariance Time Series", *IEEE Transactions on Automatic Control*, AC-30, 1, 48-56.

Lindley, D. V. and Smith, A. F. M., 1972 "Bayes' Estimates for the Linear Model", *Journal of the Royal Statistical Society Ser. B.*, 34, 1-41.

Litterman, R., 1980 "A Bayesian Procedure for Forecasting with Vector Autoregressions", Working Paper MIT.

Litterman, R., 1984 "Specifying Vector Autoregressions for Macroeconomic Forecasting", Research Department Staff Report 92, Federal Bank of Minneapolis.

Litterman, R., 1986 "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience", *Journal of Business and Economic Statistics*, 4, 25-38.

Plackett, R. L., 1950 "Some Theorems in Least Squares", *Biometrika*, 37, 149-157.

Sims, C., 1980 "Macroeconomics and Reality", *Econometrica*, 48, 1-48.

鈴木雪夫，1987 『統計学』，朝倉書店.

Tiao, G. C. and Box, G. E. P., 1983 "Modelling Multiple Time Series with Applications", *Journal of American Statistical Association*, 76, 802-816.

West, M., 1986 "Bayesian Model Monitoring", *Journal of the Royal Statistical Society Ser. B.*, 48, 70-78.

West, M., Harrison, P. J. and Migon, H. S., 1985 "Dynamic Generalized Linear Models and Bayesian Forecasting (with Discussion) ", *Journal of American Statistical Association*, 80, 389, 73-97.

West, M. and Harrison, P. J., 1986 "Monitoring and Adaptation in Bayesian Forecasting Models", *Journal of American Statistical Association*, 81, 395, 741-750.

Zellner, A., 1971 *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons. (福場庸・大澤豊訳『ベイジアン計量経済学入門』培風館，1984 年).

第7章
多重共線性へのベイズ・アプローチ

美 添 泰 人

1. 目　　的

一般に多重共線性（multicollinearity, 最近は単に collinearity と呼ぶことも多い）とは，回帰分析において説明変数の間に近似的な線形関係が成立し，そのために回帰係数の推定の精度が低下する問題を指す．

多重共線性については，特に経済分析で問題にされることが多い．その最大の理由は，経済理論によって，または政策シミュレーションの目的によって，回帰式に含まれるべき説明変数があらかじめ与えられていることが多いということであろう．

統計学の立場からは，多重共線性については，(1)その検出法，(2)影響の大きさの測定，(3)最小2乗法に代わる（混合推定，あるいはベイズなどの）推定法，(4)共線関係の少ない変数選択の方法，(5)望ましいデータ収集の設計，などの問題が考えられる．ところで通常の経済分析では，モデルおよびデータは与えられたものとみなすほかなく，最後の2つの問題は除外しなければならない．

また，よく知られているように，単に予測だけが目的なら，多重共線性は本質的な困難ではなく，主成分回帰などを用いることによっても解決できる．そのような手法については，たとえば Chatterjee and Price[1977]などを参考にされたい．

ここでは，予測の問題ではなく，経済構造の解明を目的とした場合，上記の問題について，もう一度見直してみたい．われわれの用いる手法は「ベイ

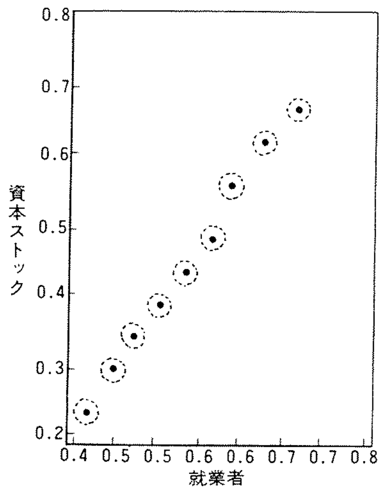


図1 就業者と資本ストックのプロット

ジアン」であるが，視点は測定誤差を中心とする「ロバストネス」である．

図1は，表1に掲げる生産関数の推定に用いられるデータのうち，説明変数である就業者 (X_1) と資本ストック (X_2) をプロットしたものである．説明変数間の相関係数は 0.9972 と比較的高いが，計量経済分析で扱うデータとしては多重共線性による影響はそれほど大きいものではない．もとのデータの有効数字は2ないし4桁与えられているが，コブ・ダグラス型の生産関数を推定する問題で，表に掲げる数値は対数変換したものである．したがって，データの信頼性を考えれば，仮にきわめて保守的な立場をとっても，変数の

表1 Leser[1974]のデータ

X_1	X_2	Y
.423	.242	.832
.447	.300	.863
.471	.345	.908
.503	.387	.908
.535	.443	.987
.564	.491	1.017
.593	.568	1.045
.631	.628	1.093
.669	.669	1.114

真の値は図の観測値の回りに示した小さな円の中のどこかにあるとみなすのが妥当であろう．しかし，誤差の範囲と考える程度のデータの変更が，回帰係数の推定値に与える影響は予想以上に大きい．

表2には，表1のデータにもとづいて計算した，最小2乗法による生産関数の推定結果と，X_2 の6期目を 0.564 から 0.554 へと変更したデータによる結果を示してある．このようにわずかな結果の変更が最小2乗法に与える影響は，きわめて大きい．

われわれの欲しいのは，この例のように説明変数のデータに関する小さな

表2 最小2乗法による回帰分析の結果

(a) オリジナルデータ

```
Dependent Variable:
  PRODUCT
  Mean:              0.97411111     Std. Deviation:            0.10125271

Regression Statistics:
  NOBS:                       9     Multiple Correlation:      0.99197769
  Std Error:         0.01477978     R-Squared:                 0.98401973
  Dataset Type:            DATA     Adjusted R-Squared:        0.97869298

Analysis of Variance:
  Source     DF   Sum of Squares    Mean Square      F Value    Prob > F
  Model       2          0.08071        0.04035    184.73151     0.00000
  Error       6          0.00131        0.00022
  Total       8          0.08202

Parameter Estimates:
Variable           Coefficient  Std. Error  Tolerance   Std.Coeff.     T  Prob
CONSTANT              0.590627  0.23301309          .   0.00000000  2.53 0.044
LABOR                 0.274171  0.82677853 0.00563247   0.22803352  0.33 0.751
CAPITAL               0.521844  0.46942305 0.00563247   0.76443963  1.11 0.309
```

(b) 変更したデータ

```
Dependent Variable:
  PRODUCT
  Mean:              0.97411111     Std. Deviation:            0.10125271

Regression Statistics:
  NOBS:                       9     Multiple Correlation:      0.99196022
  Std Error:         0.01479580     R-Squared:                 0.98398508
  Dataset Type:            DATA     Adjusted R-Squared:        0.97864678

Analysis of Variance:
  Source     DF   Sum of Squares    Mean Square      F Value    Prob > F
  Model       2          0.08070        0.04035    184.32536     0.00000
  Error       6          0.00131        0.00022
  Total       8          0.08202

Parameter Estimates:
Variable           Coefficient  Std. Error  Tolerance   Std.Coeff.     T  Prob
CONSTANT              0.741191  0.23686642          .   0.00000000  3.13 0.020
LABOR                -0.261552  0.84088291 0.00550000  -0.21668444 -0.31 0.766
CAPITAL               0.824584  0.47555747 0.00550000   1.20791778  1.73 0.134
```

誤差を前提としたときの，回帰モデルについての情報である．多重共線性についてのこれまでの分析では，説明変数にあるはずの測定誤差については，ほとんど考慮に入れられていなかった．唯一の例外は Beaton, Rubin and Barone[1976] であり，彼らは，多重共線関係にあるデータについては，最小2乗法による推定結果は丸め誤差程度の小さなデータの変更によって大きく変化することを，Longley[1967] のデータによる数値実験で示している．

次節で述べるように，われわれの考えているモデルは，測定誤差のモデル (measurement error model) と呼ぶこともできるし，事実，基本的な考え方は説明変数のデータにも集計や測定の誤差があることを認めることから出発している．しかし，Fuller[1987] などで通常扱われている測定誤差のモデルは，われわれが用いるモデルとは考え方が異なっている．また Beaton, Rubin and Barone[1976] は，ここでの測定誤差と同じ考えを示しているが，彼らの扱った問題は，最小2乗法にのみ注目した議論である．

2. モ デ ル

われわれのモデルは以下のように表現される．通常の線形回帰モデル

$$y_t = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad t = 1, \cdots, T \tag{1}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_T) \tag{2}$$

を考える．ここで $\boldsymbol{x}_t$ と $\boldsymbol{\beta}$ は p 次元のベクトルである．$\tau = \sigma^{-2}$ という表現も併用することにする．τ は，ベイジアンでよく用いる**精度** (precision) による表現である．定数項 α をわざわざ入れたのは，各変数に含まれる誤差が与える影響は，定数項とそれ以外の係数とでは異なるからである．結論からさきにいえば，各変数を偏差で表現して，定数項を除いたモデル

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \tag{3}$$

を用いても，$\boldsymbol{\beta}$ に関する推定問題には同じ解答が得られるが，われわれは最小2乗法を用いるわけではないから，このことは自明ではない．

ここで，われわれが得ることのできる観測値は(1)式に現われる真のデータではなく，測定誤差を含んだものであることを認めよう．一般には，各変数には粗誤差 (gross error) と丸め誤差 (rounding error) が含まれている

が，ここでは，主として丸め誤差およびデータ収集上の精度から経済データに反映されるすべての誤差を想定している．粗誤差がなくても，多重共線関係に近い状態にあるときには最小2乗法などが不安定であることを示すのが目的である．

さて，誤差のない観測値を $\tilde{\boldsymbol{x}}_t, \tilde{y}_t$ と表し，実際の観測値を $\boldsymbol{x}_t, y_t$ と表すことにする．$\tilde{\boldsymbol{x}}_t$ が与えられたときの $\boldsymbol{x}_t$ の分布は

$$\boldsymbol{x}_t | \tilde{\boldsymbol{x}}_t \sim N(\tilde{\boldsymbol{x}}_t, \boldsymbol{D}) \tag{4}$$

$\tilde{y}_t$ が与えられたときの y_t の分布は

$$y_t | \tilde{y}_t \sim N(0, d_y)$$

と想定する．ただし $\boldsymbol{D}$ は p 次元の分散共分散行列である．いま考えている測定誤差の性質から，$\boldsymbol{D}$ は対角行列であるとしてさしつかえない．経済データのある種のものについては $\boldsymbol{D}$ の大きさに見当をつけられるものもかなりある．

ここで，実は $\tilde{y}_t$ と y_t は区別する必要がない．このことは直感的にも明らかであるが，次のようにして示すことができる．正しいモデルは(1)式の $\boldsymbol{x}_t, y_t$ を，真の観測値 $\tilde{\boldsymbol{x}}_t, \tilde{y}_t$ で置き換えたものであるが，それを

$$y_t = \alpha + \tilde{\boldsymbol{x}}_t'\boldsymbol{\beta} + (\varepsilon_t + y_t - \tilde{y}_t) \tag{5}$$

と書き換えれば，y に関する測定誤差は通常の誤差項に含めて考えてもモデルは変化しない．ただし $y_t - \tilde{y}_t$ は，従属変数の測定誤差に対応する．したがって，はじめから測定誤差は ε に含まれているものとする．

以上から，われわれの考えているモデルは次の(6)式と，さきの(2)式および(4)式で完全に表現される．

$$y_t = \alpha + \tilde{\boldsymbol{x}}_t'\boldsymbol{\beta} + \varepsilon_t \tag{6}$$

以下では次のような行列による表現を用いる．$\boldsymbol{y}$ を従属変数の T 次元ベクトル，$\boldsymbol{X}, \tilde{\boldsymbol{X}}$ を観測値および真の値の $T \times p$ 行列，$\boldsymbol{\varepsilon}$ を T 次元ベクトルとする．また，$\mathbf{1}$ を定数項を表す1を成分とする T 次元のベクトルとする．そうすると，モデルは

$$\boldsymbol{y} = \mathbf{1}\alpha + \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{7}$$

と書ける．また，$\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau$ が与えられたときの $\boldsymbol{y}$ の分布，および $\tilde{\boldsymbol{X}}$ が与えられたときの $\boldsymbol{X}$ の分布は次のようになる．

$$p(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}}) \propto \tau^{T/2} \exp -\frac{\tau}{2}(\boldsymbol{y}-\mathbf{1}\alpha-\tilde{\boldsymbol{X}}\boldsymbol{\beta})'(\boldsymbol{y}-\mathbf{1}\alpha-\tilde{\boldsymbol{X}}\boldsymbol{\beta}) \tag{8}$$

$$\begin{aligned} p(\boldsymbol{X}|\tilde{\boldsymbol{X}}) &\propto \exp -\frac{1}{2}\sum_t (\boldsymbol{x}_t - \tilde{\boldsymbol{x}}_t)'\boldsymbol{D}^{-1}(\boldsymbol{x}_t - \tilde{\boldsymbol{x}}_t) \\ &= \exp -\frac{1}{2}\mathrm{tr}\boldsymbol{D}^{-1}(\tilde{\boldsymbol{X}}-\boldsymbol{X})'(\tilde{\boldsymbol{X}}-\boldsymbol{X}) \end{aligned} \tag{9}$$

さらに，パラメータについては次の事前分布を想定する．

$$p(\alpha, \boldsymbol{\beta}, \tau) = p(\alpha, \boldsymbol{\beta})p(\tau)$$

$\alpha, \boldsymbol{\beta}$ については，通常の情報のない（散漫な）事前分布（diffuse prior）

$$p(\alpha, \boldsymbol{\beta}) \propto \text{const.} \tag{10}$$

を仮定する．なお，情報のない事前分布の型については DeGroot[1970] や Box and Tiao[1973] などに詳しい解説がある．

τ についても情報のない事前分布を仮定することが必要である．しかし，その形として何を選んだらいいかについてはまだ結論は出せていない．その候補としては

$$p(\tau) \propto \tau^{-1}$$

と，Jeffreys のタイプの不変事前分布を選ぶことがとりあえず考えられる．しかし

$$p(\tau) \propto \text{const.}$$

を選ぶのは 4 節に示すような難点がある．とりあえず，本稿では τ の分布は特定化しないまま，τ についての条件付きの議論を示す．

$\tilde{\boldsymbol{X}}$ についても同じく散漫な分布

$$p(\tilde{\boldsymbol{X}}) \propto \text{const.} \tag{11}$$

を仮定する．特に $\tilde{\boldsymbol{X}}$ は $\alpha, \boldsymbol{\beta}, \tau$ から独立であること，すなわち

$$p(\tilde{\boldsymbol{X}} \mid \alpha, \boldsymbol{\beta}, \tau) = p(\tilde{\boldsymbol{X}}) \tag{12}$$

が仮定されているが，この仮定は通常の回帰分析でも用いられている補助統計量の仮定であり，説明変数は回帰モデルのパラメータについての情報を持っていないことを意味している．

なお，実際の経済分析では $\tilde{\boldsymbol{X}}$ についてまったく情報を持たないということはほとんどない．そのような場合には，$\tilde{\boldsymbol{X}}$ の事前分布としては，平均 $\boldsymbol{C}$

（$T \times p$ 行列）を持つ，互いに独立な正規分布を想定することができる．ここで，事前分布における $\tilde{\boldsymbol{X}}$ は互いに独立としても，$\boldsymbol{C}$ を通じてデータとしての $\tilde{\boldsymbol{X}}$ には時系列的な相関を導入することができるから，この仮定は見かけほど制約の強いものではなく，十分実用的なものである．以下で(11)の事前分布を仮定するのは，まったく情報がない場合にどのような結論が導かれるかを明らかにしたいからであって，実際の応用では情報のある事前分布を用いることを否定するものではない．その際の変更はごくわずかである．また，$\tilde{\boldsymbol{X}}$ についての事前情報は，$\boldsymbol{X}$ を観測する精度に比べてかなり小さいはずで，(11)式の事前分布を用いることによっても情報の損失はほとんどないと考えられる．

3. 事 後 分 布

$\boldsymbol{X}, \boldsymbol{y}$ が与えられたときの $\alpha, \boldsymbol{\beta}, \tau$ の事後分布 $p(\alpha, \boldsymbol{\beta}, \tau|\boldsymbol{X}, \boldsymbol{y})$ を導こう．そのためにまず $\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}}$ の事後分布 $p(\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}}|\boldsymbol{X}, \boldsymbol{y})$ を導いてこれを $\tilde{\boldsymbol{X}}$ について積分する，という方法をとる．

前節の仮定に加えて，さらに $\boldsymbol{y}$ の分布が $\boldsymbol{X}$ と独立であること，すなわち

$$p(\boldsymbol{y}|\boldsymbol{X}, \tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) = p(\boldsymbol{y} \mid \tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) \tag{13}$$

となることを認めれば，観測値 (X, y) の分布は

$$\begin{aligned} p(\boldsymbol{X}, \boldsymbol{y}|\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) &= p(\boldsymbol{y}|\boldsymbol{X}, \tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau)p(\boldsymbol{X}|\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) \\ &= p(\boldsymbol{y}|\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau)p(\boldsymbol{X}|\tilde{\boldsymbol{X}}) \end{aligned} \tag{14}$$

となることが導かれる．一方，事前分布は

$$p(\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) = p(\tilde{\boldsymbol{X}})p(\alpha, \boldsymbol{\beta})p(\tau) \propto p(\tau) \tag{15}$$

だから，$\boldsymbol{X}, \boldsymbol{y}$ を観測したときの $\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}}$ の事後分布は，ベイズの定理によって次のように求めることができる．

$$\begin{aligned} p(\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}}|\boldsymbol{X}, \boldsymbol{y}) &\propto p(\boldsymbol{X}, \boldsymbol{y} \mid \tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau)p(\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau) \\ &\propto p(\boldsymbol{y}|\tilde{\boldsymbol{X}}, \alpha, \boldsymbol{\beta}, \tau)p(\boldsymbol{X} \mid \tilde{\boldsymbol{X}})p(\tau) \\ &\propto \tau^{T/2}p(\tau)\exp -\frac{1}{2}[\mathrm{tr}\boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{U} + \tau(\boldsymbol{r}-\boldsymbol{U}\boldsymbol{\beta})'(\boldsymbol{r}-\boldsymbol{U}\boldsymbol{\beta})] \end{aligned} \tag{16}$$

ただし，$T \times p$ の行列 $\boldsymbol{U}$ および T 次元のベクトル $\boldsymbol{r}$ は，次のように定義さ

れる．

$$U = \tilde{X} - X \tag{17}$$

$$r = y - 1\alpha - X\beta \tag{18}$$

したがって，(16)式では

$$\begin{aligned} y - 1\alpha - \tilde{X}\beta &= y - 1\alpha - (X+U)\beta \\ &= r - U\beta \end{aligned} \tag{19}$$

と表されることを用いている．r は通常の残差に相当し，U は説明変数の測定誤差（の符号を変えたもの）に相当する．

(16) 式を変形して整理すれば，X, y が与えられたときの α, β, τ および $\tilde{X}$ の事後分布は次のようになる．

$$p(\alpha, \beta, \tau, \tilde{X} \mid X, y) \propto \tau^{T/2} p(\tau) \exp -\frac{1}{2} \mathrm{tr}(U-M)\Delta(U-M)' \cdot$$

$$\exp -\frac{\tau}{2}(1-\tau\beta'\Delta^{-1}\beta) r'r \tag{20}$$

ただし Δ は $p\times p$ の正値定符号（対称）行列，M は $T\times p$ の行列で，次の式で定義される．

$$\Delta = D^{-1} + \tau\beta\beta' \tag{21}$$

$$M = \tau r\beta'\Delta^{-1} \tag{22}$$

さらに(20)式の事後分布から $\tilde{X}$ を積分して追い出すことによって，α, β, τ の事後分布を次のように求めることができる．

命題 1 α, β, τ の事後分布は次式で与えられる[1]．

$$p(\alpha, \beta, \tau | X, y) \propto p(\tau)\left(\frac{\tau}{1+\tau k}\right)^{\frac{T}{2}} \exp\left[-\frac{1}{2}\frac{\tau}{1+\tau k} r'r\right] \tag{23}$$

ただし $k=k(\beta)$ は

$$k = \beta' D\beta \tag{24}$$

で定義される．

この事後分布から，β のモードを求めることができる．この節ではさしあたり τ を固定しておく．(23)式の対数を

$$\log p = \log p(\tau) + \frac{T}{2}\log\frac{\tau}{1+\tau k} - \frac{1}{2}\frac{\tau}{1+\tau k} r'r \tag{25}$$

と表す．まず(25)式を α で微分してゼロとおくと，

$$\frac{\partial \log p}{\partial \alpha} = -\frac{1}{2}\frac{\tau}{1+\tau k}(-2)1'(y - 1\alpha - X\beta) = 0 \tag{26}$$

が得られる．これを解いて

$$\alpha = T^{-1}1'(y - X\beta) = \bar{y} - \bar{x}'\beta \tag{27}$$

が求められる．ここで $\bar{x}$ は説明変数の平均（p 次元ベクトル）で，$\bar{y}$ は従属変数の平均（スカラー）である．この結果は通常の最小 2 乗法における，定数項とその他の係数の間の関係と同じである．

この α を(18)式に代入して整理すると

$$\begin{aligned} r &= (y - 1\bar{y}) - (X - 1\bar{x}')\beta \\ &= Py - PX\beta = P(y - X\beta) \end{aligned} \tag{28}$$

が得られる．ここで P は

$$P = I - 1(1/T)1' \tag{29}$$

で定義される射影行列であり，PX および Py は，平均からの偏差をとったデータを用いることに対応している．さきに(1)式のモデルについてふれたときに，「各変数を偏差で表現して α を含まないモデルをはじめから想定しても同じ結論が得られる」といったのは，以上の性質による．

次に(25)式を β で微分してゼロとおくと，次のようになる．

$$\begin{aligned} \frac{\partial \log p}{\partial \beta} = &-\frac{T}{2}\frac{1}{1+\tau k}2\tau D\beta - \frac{1}{2}\frac{-\tau}{(1+\tau k)^2}(2\tau D\beta) r'r \\ &-\frac{1}{2}\frac{\tau}{1+\tau k}2(-X)'(y - 1\alpha - X\beta) = 0 \end{aligned} \tag{30}$$

この式に(27)式の α を代入して整理すると，次の結果を得る．

命題 2 X, y が与えられたときの β のモードは，次の方程式の解として定められる．

$$\left[\left(T - \frac{\tau r'r}{1+\tau k}\right)D + X'PX\right]\beta = X'Py \tag{31}$$

ただし $k=k(\beta)$ は命題 1 で定義した β の関数である．

この形は $\boldsymbol{D}$ のウェイトに $\boldsymbol{\beta}$ が含まれているが，形としては**リッジ推定量**そのものである．ただし $\boldsymbol{D}$ の係数は無条件ではプラスにはならない．

事後分布の精度として，分散を求めるのは容易ではない．そこで次の形の尺度を導入することにする．とりあえず精度と呼ぶが，標本理論における情報量と同じ概念であり，密度関数の対数の2回微分の符号を変えたものである．この尺度によってモードの回りにおける分布の集中度を測ることができる．もしも分布が正規分布であれば，この尺度は分散と一致する．正規近似を考えれば，分布のすその形状が適当なものであれば，これを分散に代用することができよう．また外れ値の存在を考慮すれば，分布の中心の形状にのみ依存するこの尺度は，かえってロバストな尺度ともいえる．

$$I(\boldsymbol{\beta}) = -\frac{\partial^2 \log p}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\bigg|_{\text{mode}} \tag{32}$$

2回微分を実際に計算して，$\boldsymbol{\beta}$ のモードで評価すれば次の表現が得られる[2)]．

$$I(\boldsymbol{\beta}) = \frac{\tau}{1+\tau k}\left(T-\frac{\tau\boldsymbol{r}'\boldsymbol{r}}{1+\tau k}\right)\boldsymbol{D}+\frac{2\tau^2 T}{(1+\tau k)^2}\boldsymbol{D\beta\beta'D}+\frac{\tau}{1+\tau k}\boldsymbol{X'PX} \tag{33}$$

この式のはじめの2つの項が，測定誤差を考慮したことによる影響であり，最後の項が観測値からの影響である．測定誤差がないとき，すなわち $\boldsymbol{D}=\boldsymbol{0}$ のときには(33)式は

$$I(\boldsymbol{\beta}) = \tau\boldsymbol{X'PX} \tag{34}$$

となり，通常の回帰分析における最尤推定量の分散の逆行列と一致する．

$\boldsymbol{D}\neq\boldsymbol{0}$ のときは第3項からの精度は低下するが，第2項の影響は逆に精度を高める方向である．第1項はつねにプラスとは限らない．われわれのモデルでは測定誤差があるから，$\boldsymbol{\beta}$ のモードは一致性を持たないが，仮に $\boldsymbol{\beta}$ と τ のモードに両方とも一致性があれば

$$\boldsymbol{r}'\boldsymbol{r}/T-\tau^{-1}\to 0 \tag{35}$$

となるから，(33)式の第1項のかっこの中は

$$T\left(1-\frac{\tau\boldsymbol{r}'\boldsymbol{r}/T}{1+\tau k}\right) = T\left(1-\frac{1}{1+\tau k}\right) = \frac{T\tau k}{1+\tau k} \tag{36}$$

と，必ず正になる．したがって，この項が精度に与える影響は，近似的な一致性があればプラスだが，一般にはマイナスとなる可能性もある．

全体として，直感的には $\boldsymbol{D}$ を考慮にいれた場合は精度が低下すると考えられるかもしれないが，以上の議論によればこのことは明らかではない．回帰モデルでは $\tilde{\boldsymbol{X}}$ は補助統計量だから，$\boldsymbol{D}$ を考慮にいれても $\boldsymbol{\beta}$ の（推定値そのものではなく）精度についてはそれほどの差が出ない方が，むしろ自然とも考えられる．

4. 誤差分散の扱い

前節では $\tau(=\sigma^{-2})$ については事前分布 $p(\tau)$ の形を定めなかったが，(23)式の事後分布でわかるように，$p(\tau)$ は τ の事後モードに影響し，(31)式を通じて $\boldsymbol{\beta}$ の事後モードにも影響を与える．情報のない事前分布としてどのような分布を用いたらいいかを考えよう．

通常用いられるのは Jeffreys による不変事前分布であり，

$$p(\tau)\propto\tau^{-1} \tag{37}$$

とするものである．ここではもう少し一般にして，

$$p(\tau)\propto\tau^{-q} \tag{38}$$

という形の事前分布を考える．この事前分布を用いたときは，対数事後分布は次のようになる．

$$\log p = \text{const.} - q\log\tau+\frac{T}{2}\log\frac{\tau}{1+\tau k}-\frac{1}{2}\frac{\tau}{1+\tau k}\boldsymbol{r}'\boldsymbol{r} \tag{39}$$

この分布は前節の(25)式を特別な場合として含むが，$\alpha, \boldsymbol{\beta}$ のモードを与える式(27)と(31)には変更がない．一方，τ のモードは(39)式について $\partial\log p/\partial\tau=0$ を解いて求められるから，その値を用いて $\alpha, \boldsymbol{\beta}$ のモードも確定する．

ところで実際に $\partial\log p/\partial\tau=0$ を計算すると，

$$\frac{\boldsymbol{r}'\boldsymbol{r}}{(1+\tau k)^2} = \frac{T}{\tau(1+\tau k)}-2q\frac{1}{\tau} \tag{40}$$

が得られる．これは τ に関する2次方程式であり，不変事前分布による事後モードの計算はあまり簡単にはならない．この式は，$q=0$ の場合に限って1次式になる．したがって簡単化のためには，事前分布として2節で示したよ

うな

$$p(\tau) \propto \text{const.} \tag{41}$$

を用いることが考えられるかもしれないが，これは不適当な選択である．なぜならば，このとき(40)式から，τ のモードは次の式で与えられる．

$$\tau = \max((T^{-1}\boldsymbol{r}'\boldsymbol{r} - \boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta})^{-1}, 0) \tag{42}$$

そうすると，$\tau=0$ とならない限り，(31) 式における $\boldsymbol{D}$ のウェイトは常にゼロになってしまう．

以上のような問題点に注意する必要はあるものの，適当な事前分布を導入することにより形式的な分析は可能となる．しかし，われわれのモデルの誤差分散については，次のような意味でさらに注意が必要である．

はじめに指摘したように，従属変数の誤差には，測定誤差とモデルの誤差の 2 つが混在している．測定誤差については，説明変数の場合と同じようにある程度その大きさに見当をつけられる場合もある．そしてこれを超える部分がモデルの誤差に相当する．

しかし，もしも説明変数の測定誤差の分散を過大に想定すると，モデルの誤差分散が正になるという条件が成立しなくなる．(42)式が示しているように，$\boldsymbol{D}$ の影響だけ誤差分散は小さくなる．したがって事前分布を利用するときには，測定誤差の分散 $\boldsymbol{D}$ の設定に気をつけなければならない．実際いくつかの数値例では，$\boldsymbol{D}$ を大きくとると σ^2 のモードがゼロになってしまう．これは通常の測定誤差のモデルでも見られる現象である．

τ に関する事前分布を用いない最も簡単な方法として，従属変数の誤差分散 σ^2（あるいはその逆数である τ）を最小 2 乗法などで得られた $\boldsymbol{\beta}$ の推定値から一応推定しておき，その後の分析を条件付き分布にもとづいて行うことが考えられる．ただし，この方法でも $\boldsymbol{D}$ を大きく設定すると誤差分散はゼロとなるという問題は避けられない．

5. 数 値 例

ここでは Leser[1974]の例題 3.2 のデータを用いて実験を行ってみよう．データは先の表 1 に示すとおりである．Y は工業生産高，X は就業者数と資本ストックであり，いずれも 10 を底とする対数である．このデータでは説明変数間の相関係数が 0.99718 と比較的高く，共線関係に近い．通常の最小 2 乗法による推定結果は表 2 である．ここで，オリジナルデータと，1 節で示したように変更したデータとに，われわれの手法を適用してみよう．

表 3 オリジナルデータの分析

D	β (s. e.)		R $r'r$	w
0.1000	.1327	.2341	.99198	
	(.143)	(.093)	.0451378	2.766 E 0000
0.0100	.2699	.4765	.99198	
	(.359)	(.215)	.0217117	1.763 E 0000
0.0001	.2742	.5218	.99198	
	(2.722)	(1.545)	.0213107	3.890 E¯013
0.0000	.2742	.5218	.99198	
	(2.722)	(1.545)	.0213107	3.872 E¯013

（注） s. e. は $I(\beta)$ から計算した標準誤差，$r'r$ は残差平方和，w は D の係数．

表 4 変更したデータの分析

D	β (s. e.)		R $r'r$	w
0.1000	.1310	.2343	.99113	
	(.143)	(.092)	.0448366	2.779 E 0000
0.0100	.2668	.4744	.99113	
	(.370)	(.218)	.0217859	1.648 E 0000
0.0001	.4077	.4388	.99116	
	(3.266)	(1.827)	.0214444	1.325 E¯012
0.0000	.4077	.4388	.99116	
	(3.281)	(1.835)	.0214444	1.329 E¯012

（注） s. e. は $I(\beta)$ から計算した標準誤差，$r'r$ は残差平方和，w は D の係数．

われわれの手法を用いた結果は表 3（オリジナルデータ）と表 4（変更したデータ）に掲げてある．表で D としたのは測定誤差の分散行列 $\boldsymbol{D}$ の対角成分で，この実験では X_1 も X_2 も同じ大きさの測定誤差を想定している．従属変数の誤差分散 σ^2 については，条件付きの分析を行った．すなわち，測定誤差を無視した最尤法によって求めた誤差分散を固定して，条件付きの事後分布による推論を行った．

D の大きさを変えると，$D=0$ から $D=0.0001$ くらいまでは最小2乗法とパラメータの値はほとんど変わらない結果が得られている．表中の R は Y と Y の推定値の相関係数で，最小2乗法との比較のために掲げたものである．最小2乗法では重相関係数そのものである．これでみると $D=0.01$ とかなり大きい誤差を想定してもあてはまりはほとんど悪くならないように見える．ただし残差平方和でみると，$D=0.001$ あたりを境にして D を大きくするにつれて大きな残差がでてくる．もちろん残差平方和を基準にすれば最小2乗法がいちばんいい結果となるわけで，これだけではあまり意味がない．

ところで表3と表4とで，小さい D の値に対してはそれぞれ最小2乗法の推定値に近い $\boldsymbol{\beta}$ を与えているが，$D=0.001$ より大きい D に対してはどちらのデータでもほとんど同じモードを与えるようになる．これは，われわれの本来のアプローチを考えれば当然予想された結果である．結局，説明変数についての測定誤差を前提とする事後モードは，説明変数のデータの若干の変更に対してロバストな推定値を求めていることになる．なお，これらの表で D が大きくなるにつれ $I(\beta)$ が増大しているのは直感と逆であるが，今のところその原因は解明できていない．ここでは今後の課題として指摘しておくにとどめたい．

6. まとめと関連する問題

通常の測定誤差のモデルでは，説明変数 $\tilde{X}$ を確率変数とみなし，その平均，分散を考える．この分散と y の分散（誤差分散）との比が与えられて初めてモデルは識別可能になって，この条件のもとで回帰係数の推定問題が論じられる．農業試験や，生物実験のように，繰返し同じような状況でデータ解析が行える場合には，このようなアプローチも意味がある．しかし，冒頭で述べたように，われわれが一般に対象とする経済データでは，同じ条件は期待できない．われわれがなすべきことは，説明変数にも測定誤差があることを認めたとき，そして繰返し実験が不可能なときに，変数間の関係をできるだけ安定的に推定することである．この点で，経済モデルにおける測定誤差の問題は，通常の問題と性質を異にしている．われわれのモデルにおける

$\boldsymbol{\beta}$ の事後モードは，この問題に対するベイジアンからの解答である．

通常のモデルとのもっとも大きな違いは，$\boldsymbol{D}$ を既知とすることにあるが，説明変数の真の値 $\tilde{X}$ について事後分布を積分して，周辺分布を求める方法についても，ノンベイジアンのアプローチとの差とみることもできよう．

ベイジアンだと尤度関数という概念を持ち込む必要はなく，単に条件付き分布を考えればすむので，理論的に無理がない．たとえば $\tilde{X}$ は確率変数か，パラメータかなどという議論は不必要であり，周辺尤度関数などという概念も不要である．このような議論については本稿では省略するが，一般的なベイジアンの方法については Box and Tiao[1973] などを参照されたい．美添[1983] にも関連する問題の解説がある．

この稿のもう1つの目的であった多重共線性の影響の程度を検出する問題は，測定誤差の導入が推定値にどれだけ大きな影響を与えるかで，その大きさを測ることができよう．基本的には(31)式における $\boldsymbol{\beta}$ の係数行列と，OLS で用いる $\boldsymbol{\beta}$ の係数行列 $\boldsymbol{X'PX}$ との差または比が多重共線性の大きさを表しているといえる．この点については，まだなすべきことが多い．

ところで，測定誤差のあるデータにおける最小2乗法推定値の挙動を扱った Beaton, Rubin and Barone[1976] も，測定誤差についてはわれわれと同じ考えを示している．彼らは，与えられたデータに測定誤差を加えた（類似の）データが無数に得られるとしたときの最小2乗法による回帰係数の推定値は，現実のデータから得られる（ある重みを用いた）リッジ推定値となることを示した．彼らの目的は，単に計算の精度を上げるだけでは多重共線性の問題は解決しないことを示すことにあるが，データの誤差に対するロバストネスの視点という基本的な考え方はわれわれと同じである．ただ最小2乗法の挙動のみに着目した点，および標本理論の範囲での議論である点が異なっている．

われわれのモデルでも，$\boldsymbol{\beta}$ の事後分布ではなく，測定誤差のないデータから得られるはずの $\boldsymbol{\beta}$ の OLS 推定量 $\boldsymbol{b}$

$$\boldsymbol{b} = (\tilde{\boldsymbol{X}}'\boldsymbol{P}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{P}\tilde{\boldsymbol{y}} \tag{46}$$

の分布を直接扱うことができる．$\boldsymbol{X}, \boldsymbol{y}$ を与えたとき，この $\boldsymbol{b}$ の分布は（$\tilde{\boldsymbol{X}}$ の関数だから）理論的には求めることができるが簡単ではない．しかし，お

なじ $(\boldsymbol{X}, \boldsymbol{y})$ の組を繰返すような状況のもとで $T\to\infty$ とすることはできる．前と同じように $\boldsymbol{U}=\tilde{\boldsymbol{X}}-\boldsymbol{X}$ と書くと，$(T\to\infty$ のとき$)$

$$\begin{aligned}\boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}+\boldsymbol{X}'\boldsymbol{P}\boldsymbol{U}+\boldsymbol{U}'\boldsymbol{P}\boldsymbol{X}+\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U})^{-1}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y}+\boldsymbol{U}'\boldsymbol{P}\boldsymbol{y})\\ &= [T^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}+T^{-1}\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U}+O_p(T^{-\frac{1}{2}})]^{-1}[T^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y}+O_p(T^{-\frac{1}{2}})]\\ &= (T^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}+\boldsymbol{D})^{-1}T^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y}+O_p(T^{-\frac{1}{2}}) \qquad (47)\end{aligned}$$

となる．ここで $\boldsymbol{X}$ の分散行列などは，同一のデータの繰返しだから収束の問題は生じない．この式は $\boldsymbol{b}$ がリッジ推定で近似的に得られることを示しており，Beaton らの結果と形式的に一致する．われわれの方法は，有限個の観測値のもとでの議論である点と，直接 $\boldsymbol{\beta}$ のモードを扱っている点で，彼らの考え方とは異なっている．

注

1)　(23) 式の導出

本文の(20)式は(16)式を $\boldsymbol{U}$ に関する2次形式に整理することによって得られる．(20)式について

$$\int p(\alpha, \boldsymbol{\beta}, \tau, \tilde{\boldsymbol{X}} \mid \boldsymbol{X}, \boldsymbol{y})\, d\tilde{\boldsymbol{X}}$$

を評価するのに $\boldsymbol{W}=\boldsymbol{U}-\boldsymbol{M}=\tilde{\boldsymbol{X}}-\boldsymbol{X}-\boldsymbol{M}$ と変換すると，この変換のヤコビアンは定数である．ここで周知の積分公式

$$\int \exp\left[-\frac{1}{2}\mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{W}'\right]d\boldsymbol{W} \propto |\boldsymbol{\Delta}|^{-T/2}$$

を用いると

$$p(\alpha, \boldsymbol{\beta}, \tau \mid \boldsymbol{X}, \boldsymbol{y}) \propto p(\tau)\cdot\tau^{1/2}|\boldsymbol{\Delta}|^{-T/2}\exp\left[-\frac{\tau}{2}(1-\tau\boldsymbol{\beta}'\boldsymbol{\Delta}^{-1}\boldsymbol{\beta})\boldsymbol{r}'\boldsymbol{r}\right]$$

が得られる．ここで

$$|\boldsymbol{\Delta}| = |\boldsymbol{D}^{-1}+\tau\boldsymbol{\beta}\boldsymbol{\beta}'| = |\boldsymbol{D}^{-1}|(1+\tau\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta})$$

$(|\boldsymbol{D}|\neq 0$ のとき$)$ と，

$$\boldsymbol{\Delta}^{-1} = \boldsymbol{D}-\frac{\tau}{1+\tau\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta}}\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{D}$$

という2つの公式を使い，さらに

$$1-\tau\boldsymbol{\beta}'\boldsymbol{\Delta}^{-1}\boldsymbol{\beta} = (1+\tau\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta})^{-1}$$

という関係を用いると(23)式が導かれる．

2)　(33)式の導出

$$\boldsymbol{I}(\boldsymbol{\beta}) = \frac{\partial^2 \log p}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}$$

を実際に計算すると次のようになる．ただし簡単のために $v=\tau/(1+\tau k)$ とおく．

$$\begin{aligned}-\frac{\partial^2 \log p}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} &= v-(T-v\boldsymbol{r}'\boldsymbol{r})\boldsymbol{D}+(v^3\boldsymbol{r}'\boldsymbol{r}-2v^2T)\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{D}\\ &\quad +2v^2(\boldsymbol{X}'\boldsymbol{r}\boldsymbol{\beta}'\boldsymbol{D}+\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{r}'\boldsymbol{X})+v\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}\end{aligned}$$

となる．ここでモードでは

$$\boldsymbol{X}'\boldsymbol{r}\boldsymbol{\beta}'\boldsymbol{D} = \boldsymbol{D}\boldsymbol{\beta}\boldsymbol{r}'\boldsymbol{X} = (T-v\boldsymbol{r}'\boldsymbol{r})\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{D}$$

となることを用いると (33) 式が得られる．

参考文献

Beaton, A. E., D. B. Rubin and J. L. Barone, 1976 "The Acceptability of Regression Solutions: Another Look at Computational Accuracy," *JASA*, 71 : 158-168.

Belsley, D. A., E. Kuh and R. E. Welsch, 1980 *Regression Diagnostics*, Wiley.

Box, G. E. P. and G. C. Tiao, 1973 *Bayesian Inference in Statistical Analysis*, Addison-Wesley.

Chatterjee, S. and B. Price, 1977 *Regression Analysis by Example*, Wiley. (佐和隆光・加納悟訳『回帰分析の実際』，新曜社，1981年)

DeGroot, M. G., 1970 *Optimal Statistical Decisions*, McGraw-Hill.

Fuller, W. A., 1987 *Measurement Error Models*, Wiley.

Leser, C. E. V., 1974 *Econometric Techniques and Problems*, Charles Griffin. (佐和・前川訳『初等計量経済学』，東洋経済新報社，1977)

Longley, J. W., 1967 "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User," *JASA*, 62 : 819-841.

美添泰人，1983 「ベイズの手法による統計分析」，竹内啓編『計量経済学の新展開』第6章，東京大学出版会．

第8章
経済学におけるベイズ分析の発展

国 友 直 人

1. は じ め に

経済学において不確実性と情報の経済分析の発展はここ数十年にわたり目ざましいものがある．その発展を支えている原動力の中でも中心的考え方の1つがベイズ分析の考え方であるといっても過言ではないであろう．よく知られているようにベイズ決定理論の骨組は1940年代から1950年代にかけて形成されたが，現在では経済学におけるベイズ分析は多岐におよんでいる．本節では経済学におけるベイズ分析の発展を4つの主要なトピックにしぼって展望してみたい．あらかじめ本章の内容を要約するならば以下の通りである．第2節では最近における期待効用理論の再検討の議論を例を用いて考える．次に第3節ではゲーム理論におけるベイズ的方法について不完備情報ゲームの例を用いて説明する．第4節では計量経済分析におけるベイズ分析について簡単に触れる．第5節と6節では情報の経済学におけるベイズ分析の例としてベイズ学習過程と経済均衡の問題を考察する．より具体的には，合理的期待均衡の達成可能性について簡単なMuth[1961]の市場均衡モデルを用いて考察する．第2節～第4節は最近の研究動向に関する簡単な例を用いた部分的サーベイである．それに対して第5節～第6節ではこれまで知られている合理的期待均衡についての収束結果を若干拡張した命題を報告していることをあらかじめお断りしておくことにしよう．

2. 期待効用理論の再検討

不確実性が存在するとき人々の行動原理として指標（効用関数）を公理的アプローチから導き出す試みが von-Neumann and Morgenstern[1944] によって提唱されたことはよく知られている．その効用関数はフォンノイマン・モルゲンシュテルン効用関数と呼ばれているがその後の不確実性下の経済分析において1つの重要な分析用具を提供したということができよう．公理的アプローチによる効用関数の導出についてはたとえば DeGroot[1970] の第6章が要領のよい要約となっている．さて，古典的な応用例としてはたとえば Friedman and Savage[1948] を挙げることができるが，彼らは巧みにフォンノイマン・モルゲンシュテルン効用関数によって保険と賭博の同時的な存在を説明した．この例は期待効用理論を用いた古典的な経済分析として注目されていて，経済学における金融論の教科書においてしばしば取り上げられている．以来最近にいたるまで不確実性の下での経済分析，とりわけ金融論，マクロ経済学，ファイナンス（財務）などの各分野においてはフォンノイマン・モルゲンシュテルン効用関数は基本的な役割を果たしている．また，Savage[1954] は主観確率にもとづく統計学の基礎理論においてフォンノイマン・モルゲンシュテルン効用関数を導入して以来，ベイズ統計学の基本的な分析用具となっていることはよく知られていることであろう（序章を参照）．

さて，フォンノイマン・モルゲンシュテルン効用関数の現実妥当性については古くからいろいろな論議があるが，次に示すフランスのノーベル賞経済学者アレ（Allais）の反例が有名である．

例1　2つの選択問題を考える．第1の選択は

$$\begin{cases} \text{a1}: 100\text{万円（確率 }1\text{）} \\ \text{a2}: 500\text{万円（確率 }.10\text{），}100\text{万円（確率 }.89\text{），}0\text{万円（確率 }.01\text{）} \end{cases}$$

の中からどちらかを選ぶ．第2の選択は

$$\begin{cases} \text{a3}: 500\text{万円（確率 }.10\text{），}0\text{万円（確率 }.90\text{）} \\ \text{a4}: 100\text{万円（確率 }.11\text{），}0\text{万円（確率 }.89\text{）} \end{cases}$$

の中からどちらかを選ぶ．心理実験では大多数の参加者が第1の選択では a1，第2の選択では a3 を選択することが報告されている．この時，フォンノイマン・モルゲンシュテルン効用関数を $U(x)$ とすれば，第1の選択では

$$.11U(100) > .1U(500) + .01U(0) \qquad (1)$$

第2の選択では反対の不等号を意味するのでフォンノイマン・モルゲンシュテルン効用関数にもとづく選好は矛盾が生じる．

この例は当初きわめて人工的な特殊例と考えられたが，その後，特に実証的立場から米国の心理学者が多くの心理実験を行った結果，きわめて重要な現象の1例として理解されるようになってきた．極端な例としては実験室におけるネズミを使った実験でも同様な現象が観察されている（Machina[1987]）．このように，最近では実験的データにもとづいて人間の行動説明原理としてのフォンノイマン・モルゲンシュテルン効用関数に対する批判がいろいろな形で投げかけられている．ここで Machina[1987] の説明を用いるならば，この種の問題は次のように整理することができよう．いま，x を3つの実数 $x_1<x_2<x_3$ 上で定義されている確率変数で各々の確率を $p(x_i)$，フォンノイマン・モルゲンシュテルン効用関数の値を $U(x_i)$ を $U(x_1)<U(x_2)<U(x_3)$ としておくと

$$E(x) = \sum_{i=1}^{3} U(x_i)p(x_i) \qquad (2)$$

で定義される期待効用は図1で表される．

ここで期待効用 $E(x)$ を一定とする組合せは $p_2=1-p_1-p_3$ より (p_1, p_3) 平面上で直線で表現される．そして例1における a_1～a_4 の組は図1のように表すことができる．ここで左上方がより高い期待効用を示しているのでもし例1のように人々が行動すると考えれば矛盾が起きることになる．この図から2つの選択子（$x_1=0$, $x_2=100$ 万円）の他に第3番目の要素（すなわち $x_3=500$ 万円）の存在のため起こることが想像されよう．事実，この場合フォンノイマン・モルゲンシュテルン効用関数の構成に際して仮定されている

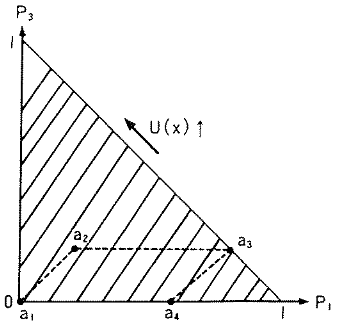


図 1　等期待効用線

第 3 選択にたいする独立性公理（independence axiom）を満たしていないのである．このことは次のように考えればより理解されやすいであろう．いま，2 つの値 0 万円と 500 万円を確率 1/11 と 10/11 でとる宝くじを a^* としよう．また E を確率 .11 の事象，E^c を E の余事象としよう．このとき例 1 における最初の選択は a_1（E の下でも E^c の下でも 100 万円）対 a_2（E の下で a^*, E^c の下で 100 万円）であり，第 2 の選択は a_3（E の下で a^*，E^c の下で 0 万円）対 a_4（E の下で 100 万円，E^c の下で 0 万円）と考えることができよう．ここで，もし事象 E^c が存在しなければ a_1 と a_4 は無差別であり，a_2 と a_4 は無差別であるから，2 つの選択問題における矛盾は第 3 の事象 E^c の存在によって生じることがわかる．

さて，以上のような批判に答える形で人々の不確実性のもとでの経済行動原理としてのフォンノイマン・モルゲンシュテルン効用関数を修正するさまざまな試みが行われている．その代表的ないくつかの試みについて Machina[1987] が展望しているが，スタンフォード大学の心理学者 Tversky やベル研究所の P. C. Fishburn を中心に精力的に研究が行われている．多くの研究の中でも理論的に最も注目に値するように思われる研究は Machina [1982] であろう．彼は独立性公理を仮定しない効用汎関数（utility functional）理論を提唱している．効用汎関数の場合には（2）式よりも一般的に

$$E(x)=\sum_{i=1}^{3}U(x_i,p)p(x_i) \qquad (3)$$

と書くが，ここで選好についての滑らかさを要求すると汎関数 $U(\cdot,p)$ は

p にたいして滑らかな汎関数となることを要求することになる．Machina [1982] は図 1 の矛盾を回避するような性質を持つ汎関数を提案している．すなわち，もしフォンノイマン・モルゲンシュテルン効用関数のように線形性を取り除いて図 2 のように期待効用の等高線を “うちわ” のようにすれば矛盾が発生しなくなる．

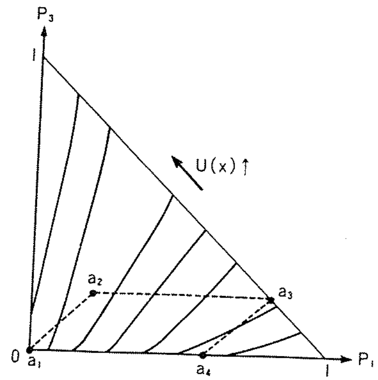


図 2　等期待効用線

ところで Machina[1987] で展望している研究とは別方向ではあるが，より伝統的ベイズ流の考え方の修正と考えられる Cyert and DeGroot[1987] による適合的効用関数（adaptive utility function）の定式化の試みもまた注目に値しよう．この場合にはフォンノイマン・モルゲンシュテルン効用関数にもとづく期待効用仮説を否定しないが，代わりに選好が時間とともに少しずつ変化することを考慮して，効用関数が確率的に変化する母数 θ を含む $U(x,\theta)$ と表現される．

3. ゲーム理論とベイズ分析

第 2 にベイズ分析の重要な分野は経済学と結びついたゲーム理論を挙げることができる．von-Neumann and Morgenstern[1944] によって開発されたゲーム理論は当初，主としてミニ・マックス原理に代表されるような非ベイズ決定理論にもとづく解を用いていた．ところが Harsanyi[1967] がベイズ・ルールにもとづくベイズ均衡の考え方をゲーム理論に導入して以来ベイ

ズ解が有力な分析手段として注目されてきている．今井・小林［1982］はこの問題について簡単な例を作っているので彼らの作った不完備情報ゲームをさらに簡明に説明しよう．

例 2　いま太郎は胴元とすべて異なる数字からなるカード・ゲームを楽しんでいるとしよう．胴元はカードの中味を見ながら自分と太郎に 1 枚ずつカードを配る．胴元は両方のカードの中味をみることができるとして 1 万円ないし，2 万円の賭け金を太郎に宣言する．他方，カードの中味を知らない太郎は勝負にでるかおりるかを決定することができるものとする．勝負が成立するとき強いカードの所有者が掛金を総て手にいれる．太郎が勝負からおりた場合にはカードを表に返して，胴元が強ければ胴元は自分の賭け金を回収するが，もし太郎が強ければ胴元に 1 万円を“罰金”として払うルールを考えてみよう．このゲームの利得表は表 1 のように与えられていることがわかる．ここでもし太郎が 2 枚のカードの中味を知ることができればゲームはまったくトリビアルなものとなる．このゲームが面白い側面を持つとすれば胴元がカードの中味を知ることができ，太郎が知ることのできないという情報の非対称性と不完備性にあるといってよい．

さて，このゲームでは胴元は自分のカードが強いときに 2 万円，自分のカードが弱いときに 1 万円賭けることが考えられよう．ところがこの場合太郎は“胴元がカードの中味を知って行動している”ことを知っていれば胴元の行動を通じてカードの中味についての情報を得ることができる．もし胴元が 2 万円賭けていれば胴元のカードが太郎のカードよりも強いことがわかるわけであり，したがって太郎はそのことも考慮して行動する，すなわちこのときには勝負からおりればよいことになる．

表 1　利得表

(i)　(胴元のカード)＞(太郎のカード)

胴元の戦略 ＼ 太郎の戦略	おりる	勝負
1 万円	0,0	1,−1
2 万円	0,0	2,−2

(ii)　(太郎のカード)＞(胴元のカード)

胴元の戦略 ＼ 太郎の戦略	おりる	勝負
1 万円	1,−1	−1,1
2 万円	1,−1	−2,2

このゲームには均衡解があるであろうか？　この場合のベイズ均衡であるハーサーニ・ナッシュ解についての詳しい説明は今井・小林［1982］を参照していただくこととして，以下でその答を要約しておこう．胴元にとっての均衡戦略は，(i)自分のカードが強いとき 2 万円を賭ける．(ii)自分のカードが弱いときには 1/3 の確率で 1 万円，2/3 の確率で 2 万円を賭ける，というものである．これにたいして太郎の均衡戦略は，(i)胴元の掛金が 1 万円ならば勝負にでる．(ii)胴元の掛金が 2 万円ならば 1/3 の確率で勝負をおり，2/3 の確率で勝負する，といったものであり，この均衡戦略を図 3 に示しておいた．

まず太郎の行動を所与として胴元の行動を考えよう．胴元が太郎より強いカードを持つとき A_1 にいる．ここで 1 万円賭ければ太郎に B_1 か B_3 の選択を迫ることになるが太郎の均衡戦略ではこのとき勝負にでる．その結果胴元は 1 万円の利得をうる．他方，2 万円を賭けるならば太郎は B_2 か B_4 の選択を迫られ，確率 2/3 で勝負することになる．この場合胴元の利得は (2/3)×2

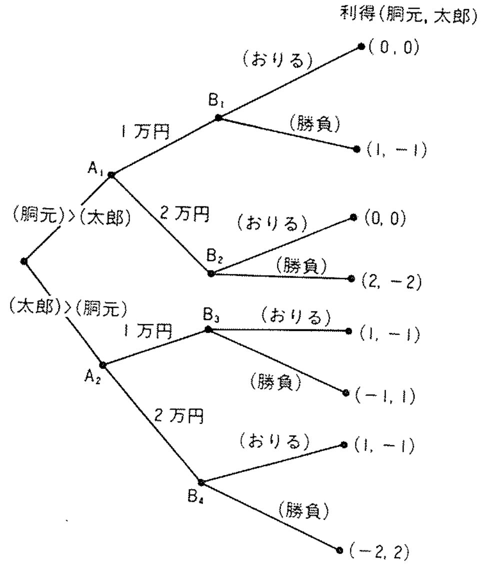


図 3　ゲーム樹

万円＝4/3 万円となるので結局，胴元は 2 万円賭けるのが最適となる．以下，同様に胴元と太郎の最適戦略を考えればよい．たとえば 2 万円胴元が賭けているときの太郎の選択 B_2 と B_4 を考えよう．胴元のカードが太郎のカードよりも強いことを事象 D，胴元が 2 万円賭けることを事象 E する．ここで $P(D)=1/2$, $P(E|D)=1$, $P(E|D^c)=2/3$ であることに注意しよう．もし太郎がベイズの公式を知っているとすれば，事象 E の下で D の確率を計算するであろう．この値は

$$P(D|E)=\frac{P(D\cap E)}{P(D\cap E)+P(D^c\cap E)}$$
$$=\frac{1/2}{1/2+(1/2)(2/3)}=3/5 \qquad (4)$$

と計算できる．したがってこの事後確率を用いると勝負からおりるときの太郎の利得は (0 万円)(3/5)＋(−1 万円)(2/5)＝−2/5 万円，勝負に参加するときの利得は (−2 万円)(3/5)＋(2 万円)(2/5)＝−2/5 万円の無差別であるから太郎の均衡戦略は最適となっている．

以上の例が示すように情報が人々の間で非対称であったり，不完備である状況はいろいろな市場において広くみられる．有名な例としては資金貸出における銀行と借り手の問題や中古車市場における売り手と買い手の問題（“レモン” の経済学）などである．このように情報が関与する経済現象の分析においてベイズ均衡の役割が大変重要視されるようになっているが，その理由の一端を上の例は示しているように思われる．さらに付け加えるならば，最近のミクロ経済学では単に 1 度のゲームではなく連続するゲームの中で人々の決定・行動を記述し，経済均衡を考えることに関心が集中している．このような状況でゲームに参加する人々が他の参加者やゲームのルールなどについての情報が不完全な場合，ベイズ・ルールは人々の行動についての 1 つのきわめて自然な定式化を与えると考えられよう．ここではこの分野の著名な仕事としてたとえば Kreps and Wilson[1982] などを挙げておくことにとどめておこう．

4. ベイズ計量経済分析

経済学におけるベイズ分析の第 3 の分野としては計量経済学にあらわれる統計モデルに対するベイズ・アプローチを挙げることができる．このアプローチはしばしばベイズ流計量経済学 (bayesian econometrics) と呼ばれているが最近の理論的発展について Zellner[1985] が概観している．たとえば計量経済分析で最も基本的な同時方程式モデル (simultaneous equation models) にもとづく計量経済モデルは

$$B(L)y_t=\Gamma(L)z_t+u_t \qquad (5)$$

と表現することができる．ここで y_t は $m\times 1$ 内生変数ベクトル，z_t は $n\times 1$ 外生変数ベクトル，u_t は $m\times 1$ 誤差項ベクトルである．またラグ作用素 L を用いると，係数行列は $B(L)=B_0-B_1L-\cdots-B_pL^p$, $\Gamma(L)=\Gamma_0-\Gamma_1L-\cdots-\Gamma_qL^q$ と表せる．ここでさらに，(i)行列 B_0 が正則であること，(ii)外生変数 z_t に時系列モデル

$$z_t=\Phi(L)z_t+v_t \qquad (6)$$

を仮定しよう．ただし $\Phi(L)=\Phi_1L-\cdots-\Phi_sL^s$ とする．このとき $(m+n)\times 1$ 変数ベクトル $x_t=(y_t, z_t)$ に対して，多変量時系列モデル

$$x_t=A(L)x_t+w_t \qquad (7)$$

を得ることができよう．ここで v_t は誤差項である．また $r=\max(p,q,s)$ とすれば係数行列は

$$A(L)=A_1L-\cdots-A_rL^r \qquad (8)$$

と書ける．この時系列モデルは状態空間 (state space) 表現を持つが，ここでこのように時系列モデルで表現すると統計モデルの母数空間の次元が大きくなる．たとえば内生変数 $m=10$，外生変数 $n=10$，$p=2$，$q=4$，$s=3$ とすれば時系列モデル (7) における係数母数は $(m+n)^2\times r=1600$，共分散行列の母数は $(m+n)(m+n+1)/2=210$ より総計 1810 個の母数が必要となる．これに対してたとえばマクロ・モデルの観測期間は四半期データで 20 年とすれば $4\times 20=80$ である．したがって，このようなモデルに対する通常のベイズ・モデルの実用的利用はたやすくないことが予想されよう．この問題を

克服する試みとして超母数（ハイパー・パラメータ）を用いる方法が有力と思われるが，その詳しい内容については本書第 3 章が参考になろう．

5. ベイズ学習過程と経済均衡

経済学におけるベイズ分析の重要な分野として，最後に情報の経済学の分野を挙げておこう．この分野に関連して特筆すべきものは，邦文の宮沢 [1971] であり，経済・経営行動における情報と意思決定，さらに動学的学習過程についての 1970 年頃までの研究を要領よく要約している．その後の不確実性の経済学についての発展の一端を知るには，Arrow[1965]，邦文の酒井 [1982]，細江 [1987] などが参考となろう．

さて，以下では情報の経済分析の例としてベイズ学習過程と経済均衡について最近特に重要と思われる予想を含んだ経済モデルの例を考えてみよう．ここである市場を構成する経済主体に関心のある経済状態が確率過程 $\{y_t\}$ であらわされているとしよう．時間変数 t は離散的であって y_t は実数値をとるものとして，さらに時点 t において，この経済主体にとって利用可能な情報を I_t としてみよう．このとき経済主体は，さまざまな理由から将来時点 $t+h(h\geqq 1)$ における経済状態 y_{t+h} の予想を行ってさまざまな経済行動を行っていると考えられる．

たとえば畜産農家にとって豚肉の将来価格の予想はこれからの生産計画を決定するに際して重要な役割を演じていると農業の経済分析では考える．もし将来における予想卸売価格が低ければ農家の生産拡大の意欲は低いであろうし，予想価格が非常に高ければ資金を調達して投資を行い，生産能力を拡張しようと計画するであろう．また株式市場では株式の売り手と買い手にとって株価の将来予想とともに将来の企業の収益見通し（期待）が経済の基礎的条件として重要な役割を演じていると考えられている．ある特定な企業が将来有望な新商品の開発に成功したのとニュースによってその企業の株価が急上昇することがよくみうけられる．以上の 2 つの例は人々の多面的な経済行動の中でほんの些細な例にしかすぎない．現代の経済活動，たとえば生産，消費，投資，貯蓄，金融，労働等々，あるゆる経済分野において人々はなん

らかの将来に関する計算（予想）をもとにして行動していると考えられよう．そこで，このように将来のことを予想しながらいろいろな活動を行っている経済主体からなる経済を分析しようとする経済学においては，経済主体が利用可能な情報をもとに将来についてどのように期待形成を行っているかが 1 つの重要な問題と考えられている．

この期待形成をめぐる問題は単に経済の現実を説明するだけにはとどまらず，経済の政策問題を考える際にも重要となる．事実，この問題が期待形成をめぐって経済学が議論する最大の理由といっても過言ではない．たとえば来年はじめから間接税についての 1 年限りの増税が本年中に発表されたとしよう．このとき，この増税の経済効果はどのように考えたらよいであろうか．この場合，人々が増税が“1 年限り”であることを予想して行動するであろうことを考慮せずに経済分析を行い，経済政策の効果，たとえば税収予想や景気に対する影響などを予測する場合には相当ミスリーディングな結論が導かれうることが想像されるであろう．

そこで以下では，経済学における著名な例をもとに予想形成方式についてベイズ・アプローチによって人々が学習するときどのような条件のもとで期待形成方式が正当化されうるかを考察することにしよう．

例 3　Muth[1961] は農産物市場におけるクモの巣モデルを考えている．ここで Muth のモデルを若干拡張して需要関数と供給関数をそれぞれ

$$\begin{cases} Q_t^d = \gamma p_t + \delta_1' z_{1t} + v_{1t} \\ Q_t^s = \eta\,{}_{t-1}p_t^e + \delta_2' z_{2t} + v_{2t} \end{cases} \tag{9}$$

と表すことにしよう．ここで Q_t と p_t はそれぞれ時刻 t における数量と価格を示す．γ と η はスカラー母数，δ_1' と δ_2' はベクトル母数を表し，z_{1t}，z_{2t} は外生変数，v_{1t} と v_{2t} は誤差項をそれぞれ示している．${}_{t-1}p_t^e$ は $t-1$ 期までの情報にもとづく生産者の価格の予想値であり，生産者はこの予想にもとづいて供給を決定する状況を想定している．市場で数量について需要量 Q_t^d が供給量 Q_t^s に一致する均衡条件 $Q_t^d = Q_t^s$ を与えれば $\gamma \neq 0$ のとき

$$p_t - \left(\frac{\eta}{\gamma}\right){}_{t-1}p_t^e = \frac{1}{\gamma}\{\delta_2' z_{2t} - \delta_1' z_{1t} + v_{2t} - v_{1t}\} \tag{10}$$

と書き直すことができる．

ところで例3の経済モデルにおいて $\alpha=\eta/\gamma$, $y_t=p_t$, $u_t=\{\delta_2' z_{2t}-\delta_1' z_{1t}+v_{2t}-v_{1t}\}/\gamma$ とおけば，

$$y_t-\alpha_{t-1}y_t^e=u_t \qquad (11)$$

と書き直すことができよう．ただし α は定数，${}_{t-1}y_t^e$ は $t-1$ までの情報を用いた時に y_t について人々が予想する値である．$\{u_t\}$ を便宜上，政策変数あるいは外生変数と呼ぶが，さらにこの政策変数がしたがう確率過程を政策過程と呼ぶことにしよう．

ここで，方程式(11)によって記述される経済モデルにおける人々の期待形成を考えよう．人々は必ずしもこの経済システム(11)について情報が完全でなく，利用可能な情報から $\{y_t\}$ に関する予想を形成しているものとする．この不完全情報の状況を定式化して，人々は変数 $\{y_t\}$ を観測し，さらにこの経済システム（すなわち需要関数と供給関数）についての情報（説明変数）z_t ($p\times1$ ベクトル）も利用可能としよう．初期時点において y_t の分布についての母数 θ（$p\times1$ ベクトル）について人々は事前分布

$$\theta\sim N\left(\theta_0, \frac{S_0}{\omega^2}\right) \qquad (12)$$

を想定しているとしよう．ただし，ここで簡単化のために比 S_0/ω^2 は既知として $N(\cdot)$ は正規分布を表すものとしよう．

さて次に人々の学習過程を考えよう．以上の設定の下では，人々が $\{y_t\}$ について観測可能な $p\times1$ 説明変数ベクトル z_t を持つ線形モデルを仮定しベイズ・ルールにもとづいて母数を推定していると考えるのは自然であろう．ここで誤差項の平均ゼロ，分散 ω^2 とおけば時刻 $t-1$ における θ の事後分布の平均は

$$\theta_{t-1}=S_{t-1}^{-1}\left\{S_0\theta_0+\sum_{s=1}^{t-1}z_s y_s\right\} \qquad (13)$$

となる[1]．ただしここで記号

$$S_{t-1}=S_0+\sum_{s=1}^{t-1}z_s z_s'$$

を用いた．ここでより一般的に ω^2 を未知として，(12)の事前分布として自然共役分布の正規・ガンマ分布を仮定することが考えられる．このとき事後分布は多変量 t 分布となるが，事後平均は(13)に一致することがわかる（鈴木[1987]第6章を参照）．したがって，以下の議論は本質的に変わらないことになる．

この情報にもとづいて人々が $\{y_t\}$ の予測量を計算するものとすれば予測分布の平均値

$${}_{t-1}y_t^e=z_t'\theta_{t-1} \qquad (14)$$

を考えることが自然であろう．

さらに分析の簡単化のために経済モデル(11)における政策変数と人々が利用可能な説明変数 $\{z_t\}$ について次のように仮定しよう．すなわち $t-1$ 期までの情報を条件にするとき

$$u_t=z_t'\beta+\varepsilon_t \qquad (15)$$

と書ける場合を考えよう．ここで過去の情報を条件とするときの u_t の条件付期待値が $E_{t-1}(u_t)=z_t'\beta$ となることを仮定することを意味している．すると誤差 ε_t は

$$\varepsilon_t=u_t-E_{t-1}(u_t) \qquad (16)$$

となるので

$$\begin{cases}\text{(i)} & E_{t-1}(\varepsilon_t)=0\\ \text{(ii)} & E_{t-1}(\varepsilon_t^2)=\sigma^2\end{cases}$$

が成立する．ここで以上の仮定のうち，条件(i)は(7)より条件付期待値の定義から明らかであろう．条件(ii)はたとえば $\{u_t\}$ と $\{z_t\}$ に同時正規分布を仮定すれば満たされる．いずれにしても以下の議論はあまり制約的とはならないことに注意しよう．例3においては

$$\begin{cases}\varepsilon_t=-(1/\gamma)v_{1t}+(1/\gamma)v_{2t}\\ \beta=(-\delta_1'/\gamma, \delta_2'/\gamma)\\ z_t'=(z_{1t}', z_{2t}')\end{cases} \qquad (17)$$

とすれば以上のすべての条件を満足していることが分かる．

ここで以上で説明した経済モデルの構造と学習過程の定式化は Fourgeand, Gourieroux, and Pradel[1986] および Bray and Savin[1986] を若干

一般化していることをあらかじめ断わっておこう．また以上の分析には，きわめて特殊な学習過程を扱っているBray[1983]など，これまでに研究されている経済予想における学習過程についての多くの分析を特殊な場合として含んでいる．

6. 合理的期待への収束問題

人々の予想形成方式についてこれまで経済分析によく用いられる仮説は伝統的には外挿的（extraporative）期待形成仮説や適合的（adaptive）期待形成仮説であった．これら2つの仮説は“後向き”（backward looking）の期待形成方式と言われる．これに対して最近では最も有力な仮説として用いられているのが“前向き”（forward looking）な期待形成方式の1つと考えられる合理的期待（rational expectation）形成仮説である．これら異なる期待仮説にもとづいた計量経済分析の特質については国友[1988]が概観しているのでそちらに譲るとして，以下では合理的期待仮説のみに注目しよう．合理的期待仮説を用いれば第3節の経済モデル(11)において

$$ {}_{t-1}y_t^e = E_{t-1}(y_t) \tag{18} $$

と表すことができる．ここで $E_{t-1}(y_t)$ は $t-1$ 時刻における利用可能な情報が与えられた y_t の条件付期待値である．この期待値の計算に際して用いる確率分布は期待形成に際しての人々の主観分布であるが通常の分析では確率過程 $\{y_t\}$ の分布を用いて計算するので人々の主観分布は同一であってしかもそれが客観分布に一致することを仮定していることになる．一見して分かるようにこの仮定はかなり強いものであり，合理的期待仮説を用いた合理的期待均衡の分析の有効性については，理論的にも実証的にもいまだ評価は分かれている．

ここで合理的期待の下では(18)を(11)に代入すると

$$ y_t = \alpha E_{t-1}(y_t) + u_t \tag{19} $$

となる．ここで両辺の条件付期待値をとれば $\alpha \neq 1$ のとき

$$ E_{t-1}(y_t) = \frac{1}{1-\alpha} E_{t-1}(u_t) \tag{20} $$

となる．したがって，$E_{t-1}(y_t) = \alpha E_{t-1}(y_t) + \beta' z_t$ より合理的期待仮説の下での解は

$$ y_t = \left(\frac{1}{1-\alpha}\beta\right)' z_t + u_t \tag{21} $$

と書くことができる．ここで方程式(21)は母数係数が一定の回帰モデルの形になっていることに注意しておこう．

さて，本節では主観分布をもって学習しながら行動している人々から成る経済が合理的期待均衡へ収束するか否かを簡単な経済モデルを用いて考えよう．ここで人々がベイズ・ルールにしたがって学習して行動していると仮定すると時刻 t において

$$ y_t = \alpha \theta_{t-1}' z_t + u_t \tag{22} $$

となる．したがって(19)を用いると経済の状態方程式は

$$ y_t = (\beta + \alpha \theta_{t-1})' z_t + \varepsilon_t \tag{23} $$

と書き直すことができる．次に時刻 t において新しい情報をもとにして θ_{t-1} から θ_t へと更新することを考えよう．ここで

$$ \begin{aligned} \theta_t &= S_t^{-1}\left\{S_0\theta_0 + \sum_{s=1}^{t} z_s y_s\right\} \\ &= S_t^{-1}\{S_{t-1}\theta_{t-1} + z_t y_t\} \end{aligned} \tag{24} $$

となることに注意しよう．(24)を用いて(23)を最終項に代入すれば

$$ \theta_t = S_t^{-1}\{S_{t-1}\theta_{t-1} + z_t z_t'(\beta + \alpha\theta_{t-1})\} + v_t \tag{25} $$

を得る．ただし，

$$ v_t = S_t^{-1} z_t \varepsilon_t \tag{26} $$

とおいた．ここで $z_t z_t' = S_t - S_{t-1}$ となることに注意すれば，事後平均が $\theta_t = \theta_{t-1} = \theta$ となる定常状態では，$\alpha \neq 1$ のとき $\theta = \beta/(1-\alpha)$ となる．そこで変数（$p \times 1$ ベクトル）を θ の回りで定義しなおして

$$ c_t = \theta_t - \frac{\beta}{1-\alpha} \tag{27} $$

とおこう．方程式(16)を整理すると確率過程 $\{c_t\}$ についての1次差分方程式

$$ c_t = \beta_t c_{t-1} + v_t \tag{28} $$

を得る．ここで係数行列は

$$\beta_t = I_p - (1-\alpha)S_t^{-1}z_t z_t' \tag{29}$$

となることに注意しておこう．もし $\theta=\beta/(1-\alpha)$ となって事後平均が一定となる場合には $c_t=0$ となることに注目しよう．すなわちわれわれはここで解の収束問題を時刻 t の経過とともに $c_t\to 0$ となるか否かの数学的問題へと変換したのである．このとき以下に述べる2つの結果が成立する．

命題1： $p=1$ かつ $0<\alpha<1$ のとき次の条件を仮定する．(i) $t\to+\infty$ につれて $S_t\to+\infty$ (a. s.)，(ii)ある正整数 q が存在して $E(S_q^{-1})<+\infty$．このとき $t\to+\infty$ につれて

$$c_t \to 0 \text{ (a. s.)} \tag{30}$$

となる[2]．

命題2： $p\geqq 1$ かつ $1/2<\alpha<1$ のとき次の条件を仮定する．(i) $t\to+\infty$ につれて $S_t\to\infty$ (a. s.)，(ii)ある正整数 q が存在して $E(S_q^{-1})<+\infty$，(iii) S_T の最大固有値を μ_T，最小固有値を λ_T とするとき任意の t に対して $\mu_t/\lambda_t<M<+\infty$ となる M が存在する，(iv) $t\to+\infty$ につれて $\|S_t^{-1}z_t z_t'\|\to 0$ となる．このとき $t\to+\infty$ につれて

$$c_t \to 0 \text{ (a.s.)} \tag{31}$$

となる．

ただし，ここで任意の行列 A に対して $\|A\|$ は行列ノルム $\sqrt{\lambda_{\max}(A'A)}$ を表している．

ここで念のため，a. s. とは概収束（almost sure convergence）の意味であり，概収束ならば確率収束（convergence in probability）することに注意しておこう．命題1の証明は念のため注2)で示しておいたが，命題2の証明は長くなるので省略した．以上で用いた経済モデルの定式化はほぼ Bray and Savin[1986] と Fourgeand ら[1986] に沿っているが，命題1と命題2は彼らの得た結果における仮定を相当弱めているばかりではなく，より簡明と思われる．また注2)で示した命題1の証明はきわめて自然であるように思われる．Bray and Savin[1986] は説明変数 $\{z_t\}$ が互いに独立で同一の分布にしたがう仮定をおいているが，一見して明らかなように，この仮定は応用上はきわめて限定的といえるであろう．

さて以上の議論では，母数 θ についての事後分布の平均を用いて収束の問題を考えてきた．事後分布の分散共分散行列については容易にわかるように（(注1)参照)，

$$\left(\frac{1}{\omega^2}S_{t-1}\right)^{-1} \tag{32}$$

となるので $t\to\infty$ につれて発散する．したがって命題1および命題2のいずれの仮定のもとでも事後分布は事後平均に収束することがわかる．

次に命題1と命題2の意味を考えてみよう．合理的期待均衡に対する批判としてしばしば期待形成における条件付分布の計算において，人々があたかも将来の“正しい”確率分布を知っているかのごとくに扱う仮定の非現実性が取り上げられている．それに対して本稿の枠組では，人々は経済についての“真”のモデルを知らないときに，それぞれ利用可能な情報を用いて経済の“真”のモデルについて，ベイズ・ルールによる主観確率を用いて学習している状況を考えている．ここでの主張は，このような状況では一般的に多くの場合十分に時間をかけて学習してゆけば，極限としては合理的期待均衡が達成されることを示している．むろん第2節で挙げた Muth の市場モデルは通常の安定条件を満たす場合 $0<\alpha<1$ となるので命題1を適用できよう．またさらに $0<\alpha<1/2$ の場合には命題2を適用することができるであろう．

ところで以上の命題では，一定の条件の下で均衡の収束を保証するものであるが，実際には収束のスピードが問題となろう．本節で述べた命題に関連して，収束スピードやその他多くの興味ある問題についての詳しい結果は他の機会に譲ることにして，図4で1つのシミュレーションの例を示すことにとどめよう．

ここで外生変数はランダム・ウォーク・モデル $z_t=z_{t-1}+\varepsilon_t$ から発生させ，経済モデル(11)における母数は $\alpha=.1$ とおいている．この場合 c_t は $t\to\infty$ につれて少しずつ均衡値へ収束しているようであるが，そのスピードは非常に遅く，$t=200$ においても完全に収束しているとは言いがたい．このようなパターンが数値実験においてしばしば生じることは大変興味深い．

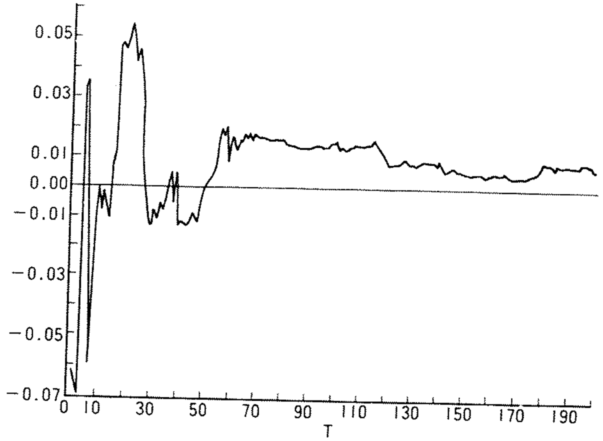


図4 シミュレーション

7. おわりに

以上で説明したように現代の経済学においてベイズ分析は理論と応用の両面において多岐に及んでいて，必要不可欠な道具となっていることが分かる．またこの方面での残されている研究テーマは数多くあり，今後の発展が期待される．また同時に，この分野はベイズ統計学と経済学の双方にとってその共同の発展が望まれる分野であろう．

（なお，注において，第5節と第6節で述べた命題の数学的証明を与える．また，念のため証明に用いた確率論でよく知られている若干の命題[3]も説明した．）

注

1) **事後平均・分散の導出** この事後分布については専門家の間ではよく知られているが念のため説明する．詳しくはたとえば鈴木［1987］第7章を参照されたい．線形回帰モデルを

$$y_s = X_s\beta + \varepsilon_s \tag{33}$$

と書こう．ここで

$$y_s = (y_1, \cdots, y_s)' : s\times 1 \text{ ベクトル},\quad X_s = (z_1, \cdots, z_s)' : s\times p \text{ 行列}$$
$$\beta = (\beta_1, \cdots, \beta_p)' : p\times 1 \text{ ベクトル},\quad \varepsilon_s = (\varepsilon_1, \cdots, \varepsilon_s)' : s\times 1 \text{ ベクトル}$$

とおいた．$\varepsilon_s \sim N(O, \Sigma_s)$，$\Sigma_s = \omega^2 I$ とすると変換 $\{\varepsilon_s\}\to\{y_s\}$ は1対1である．したがって y_s の尤度関数は平均 $X_s\beta$ 分散 Σ_s の正規分布 $N(X_s\beta, \Sigma_s)$ であるから

$$f(y_s|\beta) = (2\pi)^{-s/2}|\Sigma_s|^{-1/2}\exp -\{y_s - X_s\beta\}' \times\Sigma_s^{-1}\{y_s - X_s\beta\}/2 \tag{34}$$

と書ける．いま β の事前分布を $N(\beta_0, \Sigma_0)$ とすれば時刻 s における事後分布は

$$\xi(\beta|y_s) = n(X_s\beta, \Sigma_s)\times n(\beta_0, \Sigma_0) \propto \exp -Q/2 \tag{35}$$

ただし $n(a, b)$ は平均 a，分散共分散行列 b の正規分布の密度関数，Q は2次形式の和であり，

$$\begin{aligned} Q &= \{\beta - \beta_0\}'\Sigma_0^{-1}\{\beta - \beta_0\} + \{y_s - X_s\beta\}'\Sigma_s^{-1}\{y_s - X_s\beta\} \\ &= \{\beta - \beta_s\}'\Sigma_s^{-1}\{\beta - \beta_s\} + y_s'\Sigma_0^{-1}y_s - \beta_s'\Sigma_s^{-1}\beta_s \\ &\quad + \beta_0'\Sigma_0^{-1}\beta_0 \end{aligned} \tag{36}$$

となる．ここで

$$\beta_s = \Sigma_s^{-1}(\Sigma_0^{-1} + X_s'\Sigma_0^{-1}y_s) \tag{37}$$
$$\Sigma_s^{-1} = \Sigma_0^{-1} + X_s'\Sigma_0^{-1}X_s \tag{38}$$

である．したがって時刻 s における β の事後分布は $N(\beta_s, \Sigma_s)$ で与えられる．次に $s=1, \cdots, t$ について繰返しここで $\Sigma_t = \omega^2 I$ であることに注意すれば，求める事後分布を得る．

2) **命題1の証明** 時刻 s 期において利用可能な変数から生成される σ－集合体を F_s とする．ここで説明変数 z_s については $z_{s+1}\varepsilon F_s$ と考えておくことにしよう．

ここで(28)より代入を k 回繰り返せば

$$c_t = v_t + \beta_t v_{t-1} + \cdots + \left(\prod_{i=0}^{k-1}\beta_{t-i}\right)v_{t-k} + \left(\prod_{i=0}^{k}\beta_{t-i}\right)c_{t-k-1} \tag{39}$$

と書き直すことができる．ここである正整数 $q\,(0<q<t)$ を用いて

$$x_t = \sum_{s=q}^{t}\gamma_s v_s \tag{40}$$

を定義しよう．ただし係数 $\{\gamma_s\}$ は $\gamma_t = 1$,

$$\gamma_s = \prod_{i=0}^{t-1-s} \beta_{t-i} \quad (s \leqq t-1) \tag{41}$$

としよう．このとき $E(v_s|F_{s-1})=0$(a. s.) であるから $\{x_t\}$ は離散マルチンゲールになっていることに注意しよう．ここで Kronecker の補題を用いると

$$\begin{aligned} E(x_t - x_q)^2 &= E\sum_{s,r=q}^{t} \{\gamma_s \gamma_r v_s v_r\} \\ &= \sigma^2 E\left\{\sum_{s=q}^{t} \gamma_s^2 S_s^{-2} z_s^2\right\} \\ &\leqq \sigma^2 E\left\{\sum_{s=q}^{t} S_s^{-2} z_s^2\right\} \\ &\leqq \sigma^2 E\{S_q^{-1}\} \end{aligned} \tag{42}$$

である．ここで $0<|\gamma_s|<1$ に注意しておこう．命題1の仮定より $E\{S_q^{-1}\}$ は有界であるからマルチンゲール $\{x_t\}$ は2次モーメントが有界となっている．したがってマルチンゲール収束定理（以下の補助命題(3)を参照）を用いると $t\to\infty$ につれて x_t は確率1(a. s.) で収束する．

さらに係数 $\{\beta_s\}$ より

$$b_t = \frac{1}{\prod_{k=1}^{t} \beta_k} \tag{43}$$

を定義しよう．ここでもし $\prod_{k=1}^{t} \beta_k \to 0$ を示すことができれば明らかに $b_t \to +\infty$(a. s.) である．すると

$$x_t = \frac{1}{b_t}\sum_{s=1}^{t} b_s v_s \tag{44}$$

と書き直すことにすれば Kronecker の補題（以下の補助命題(2)を参照）を用いると $t\to\infty$ につれて $x_t \to 0$(a. s.) が示される．

ところで $0<\alpha<1$ のとき(27)より

$$\log b_t = -\sum_{s=1}^{t} \log\{1-(1-\alpha)S_s^{-1}z_s^2\} \geqq (1-\alpha)\sum_{s=1}^{t} S_s^{-1}z_s^2 \tag{45}$$

である．したがって命題1の条件 $S_t\to+\infty$(a. s.) から再び Kronecker の補題を用いると $t\to\infty$ のとき $\log b_t \to +\infty$(a. s.) となるので命題1が証明された．

3) **補助命題**

(1) **Kroneker の補題**（Breiman[1968] 第3章参照）

実数列 $\{x_t\}$ が条件

$$\sum_{s=1}^{n} x_s \to x \quad (\text{有限}) \tag{46}$$

$$b_n \uparrow \infty \tag{47}$$

を満たすものとしよう．このとき $n\to\infty$ につれて

$$\frac{1}{b_n}\sum_{s=1}^{n} b_s x_s \to 0 \tag{48}$$

となる．

(2) **マルチンゲール収束定理**（Breiman[1968] 第5章参照）

$x_1, x_2, \cdots$ を離散マルチンゲール，すなわち $E_{t-1}(x_t)=x_{t-1}$ を満足する確率過程として，条件

$$\varlimsup_{n\to\infty} E|x_n| < \infty \tag{49}$$

を満足するものとしよう．このとき

$$x_s \to x \quad (\text{a. s.})$$

となる確率変数 x が存在してかつ $E|x|<\infty$ となる．

参 考 文 献

Arrow, K. J., 1965 *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjo Jahnsson Saatio.

Bray, M., 1983 "Convergence to Rational Expectations Equilibrium", ch. 6, in *Individual Forecasting and Aggregate Outcomes*, Cambridge University Press.

Bray, M. M. and Savin, N. E., 1986 "Rational Expectations Equilibria, Learning, and Model Specification", *Econometrica*, 54, 1129-1160.

Breiman, L., 1968 *Probability*, Addison-Wisley.

Cyert, R. M. and DeGroot, M. H., 1987 *Bayesian Analysis and Uncertainty in Economic Theory*, Rowman and Littlefield.

DeGroot, M. H., 1970 *Optimal Statistical Decisions*, McGraw-Hill.

Fourgeand, C. C. Gourieroux, and J. Pradel, 1986 "Learning Procedures and Convergence to Rationality", *Econometrica*, 54, 845-868.

Friedman, M. and Savage, L., 1948 "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, Vol. LVI, No. 4, 279-304.

Harsanyi, J., 1967 "Games with Incomplete Information Played by 'Bayesian' Players I-III," *Management Science*, Vol. 14.
Kreps, D. M. and R. Wilson, 1982 "*Sequential Equilibria,*" *Econometrica*, 50, 863-894.
Machina, M. J., 1982 "Expected Utility: Analysis without the Independence Axiom," *Econometrica*, 50, 277-324.
Machina, M. J., 1987 "Choice under Uncertainty: Problems Solved and Unsolved," *Economic Perspectives*, vol. 1.
Muth, J. F., 1961 "Rational Expectations and the Theory of Price Movements," *Econometrica*, Vol. 29, 315-335.
Savage, L. J., 1954 *The Foundations of Statistics*, John-Wiley.
von-Neumann, J. and Morgenstern, O., 1944 *Theory of Games and Economic Behavior*, Princeton University Press.
Zellner, A., 1985 "Bayesian Econometrics," *Econometrica*, 53, 253-270.
今井晴雄・小林孝雄, 1982 「ゲームの理論と経済学」,『経済セミナー』4月～12月号.
細江守紀, 1987 『不確実性と情報の経済分析』, 九州大学出版会.
国友直人, 1988 「予想と時系列(1), (2)」,『経済学論集』(東京大学経済学会), 54巻2, 3号.
酒井泰弘, 1982 『不確実性の経済学』, 有斐閣.
鈴木雪夫, 1987 『統計学』, 朝倉書店.
宮沢光一, 1971 『情報・決定理論序説』, 岩波書店.

第9章

経営意思決定における不確実性の評価*

青 井 倫 一

1. はじめに

マネジメントにとって経営上の意思決定には大なり小なり不確実性の処理が伴う．市場動向の判断，技術進歩の速度と方向性，投資効果の測定，競争他社の戦略等に伴う不確実性の処理により意思決定が大きく左右される場合も多いであろう．従来，不確実性の下の意思決定分析では次のようなアプローチ（デシジョン・ツリー・アプローチ）を進めてきていた．すなわち意思決定分析を2つの小さな，しかし容易でない意思決定に分解することである．この分解された2つの意思決定は，不確実な事象に関する確率の付与と選好度の決定である．分解して，それぞれ確率付与と選好度の決定をした後，これらを統合することにより本来の意思決定分析を完成する（Raiffa[1968]）．しかし，上に述べた方法のなかで不確実性の評価，不確実な事象に対する確率の付与という意思決定は企業のマネジメントにとり複雑化する環境の下ではますます困難なものとなってきている．したがって対象領域に関する専門家（ここでは意思決定者より相対的にある分野の知識のある人々を専門家ということにする．意思決定者の部下も，友人も，世間でいうその道のプロもここでは専門家である）に情報提供，場合によっては確率情報の提供，すなわち不確実性の評価という意思決定の代行を委託する，せざるをえない場合も多くなっているといえる．

ところで不確実性の評価を専門家という「他人」に委託した場合，意思決定者は不確実性の処理ということから解放されるのかといえば，話はそれほ

ど単純ではなく，専門家を利用した場合，新たな問題・課題を抱えることになる．というのは専門家による不確実性に関する評価の報告そのものは専門家による意思決定そのものであるので，今度は専門家の意思決定そのものの評価をしなければならなくなる．しかも確率情報そのものの妥当性は白黒の世界ではないので，結果が生起しても簡単には検証できないのである．または衆知を集めようと複数の専門家に同一の不確実性の評価を委託した場合には，その報告が異なったり，似ていたりしている時，報告の統合をどうすべきかの問題が残る（Aoi[1981]）．

ここでは経営上の意思決定によくみられる，不確実性評価を専門家という他人に委託した時生じる諸問題の検討を行うことにしよう．といってもすべての問題を検討することもできないので，大きく4つの観点からこの状況を見ることにする．これら4つの観点を，ここでは石油掘削の意思決定におけるマネジメント（意思決定者）と地質学者（石油の出る確率評価の専門家）という構図で例示的に表現することにしよう（Grayson[1960]）．

マネジメントはある鉱区で石油掘削のための投資をすべきかどうか考えている．投資決定に大きな影響を与えるひとつの要因は，石油が出る可能性がどの程度かということである．石油の出る確率の大小により投資の決定は大きく影響を受けるのは明白である．

しかしマネジメントは石油の出る可能性を示す地質のことは分からないので，石油の出る確率の付与は専門家である地質学者に頼らなければならない状況であった．まずマネジメントが考えるのは，地質学者はその分野の専門家であることは疑わないのであるが，はたして，地質学者は彼自身の知識を確率情報に“適切に”表現できるのであろうかという確率表現上の認識の問題，次に認識上の問題はなくても動機上の問題，すなわち地質学者は正直に自分の確率情報を報告しているのだろうかという問題である．地質学者は自分のなんらかの利害を考慮して，それに都合のよいある決定をマネジメントに選択させようとして偏りをかけた報告をしているのではなかろうか．

上の例は地質学者が1人の場合だが，複数の地質学者がいて，それぞれ異なる，対立するような報告をするとしたら，マネジメントはどうすべきか，また複数の地質学者の報告には重複するデータが存在していることが分かっている時，マネジメントは地質学者による複数の，したがって情報量の多い，しかし重複を含む報告をどう評価すべきなのかという問題もまた衆知をあつめようとすると生じるであろう．上に挙げたような問題についてこれから考えていくことにしよう．

2. 認識上の問題

地質学者は地質の専門家であるが，必ずしも確率表現の“専門家”ではない．したがってマネジメントの受け取る確率情報は，「適切な地質学上の判断を不適切に確率表現」したものになっていることはマネジメントにとって好ましいことではない．そこで専門家，そしてマネジメントも含めて自分自身の持つ不確実な事象に関する知識・判断をいかに容易に確率の形に変換するかの方法の開発が望まれることとなる．理論的研究とは異なり，実際の意思決定で種々の専門家に依存せざるを得ないマネジメントにとってこの知識の確率表現の実用的方法の進展は望まれることである．

古典的にはこの問題は気象学者に統計学を学ばせるのと，統計学者に気象学を学ばせるのとどちらが適策かという問題であって従来の回答は前者だとされている．対象問題に対する判断（Judgement）と，それを確率表現に変換する評価（Assessment）に分けたモデルで人間の情報処理プロセスを把握しようとするアプローチがこの分野を開拓している．不確実性を伴う対象への認識上の諸問題を扱ったものとして Kahneman, Slovic and Tversky [1982]がある．米国では気象の確率表現の適切度の向上のための研究はかつてなされたが，今日主観確率の表現方法についての研究が今まで以上に望まれることとなっている（Winkler[1960]）．

3. 動機上の問題——正直な報告へのインセンティブ——

地質学者は地質上の知識を適切に石油のでる確率の付与（Pとする）をするとしよう．しかし彼の報告する確率q（石油の出る確率）ははたして正直な報告，すなわち$q=P$なのであろうか．どのような報酬制度（地質学者の

報告値，この場合 q と現実に生起した事象に依存する）の下で地質学者は正直に行動することにより，彼の期待報酬を極大化できるのであろうか．もし可能ならば，そのような報酬制度を設けることにより地質学者の動機的なバイアスを除きたいとマネジメントは願うであろう[1]．

このような状況に対応するために考えられたのがプロパー・スコアリング・システムである．ここではプロパー・スコアリング・システムの代表例のブライヤーズ・スコアについて説明して，スコアリング・システムを理解した上でプロパー・スコアリング・システムの一般形，およびそれへの批判について展開していこう（McCarthy[1956], Savage[1971]）．

ところでここでは世界は 2 項的世界（E および $\bar{E}$）を仮定し，また地質学者たち専門家はリスク中立的，すなわち与えられる報酬（利得）の期待値を極大化する行動をとるものとする．

いま，専門家は事象 E の生起について確率付与 P，したがって E の生起確率の付与は $(1-P)$ としていて，マネジメントには E の生起確率を q と報告するとする（q は必ずしも P と一致しない）．ここでブライヤーズ・スコアとは上のような状況で，もし事象 E が生起したら $R1$，事象 $\bar{E}$ が生起すれば $R2$ の報酬を与える制度である．

$$R1 = 2q-q^2,\quad R2 = 1-q^2 \tag{1}$$

すぐに理解できるように専門家は（1）の下での次式で表現される自分の期待報酬 V を極大にするように q を決定するので，専門家にとって最適な q の値は P，すなわち専門家自身の付与した確率となる．

$$V = P\times R1+(1-P)\times R2$$

ブライヤーズ・スコアは正直な報告の期待報酬が最大になるようなシステムなのである．ブライヤーズ・スコアをその 1 例として含む，正直な報告を専門家の最適行動として誘導する報酬制度をプロパー・スコアリング・システムといい，その一般形は 2 項的世界では下のように表現される．

$$R1(q,E) = f(q)+F(q)-qf(q)$$
$$R2(q,\bar{E}) = E(q)-qf(q) \tag{2}$$

ここで $R1(q,E)$ は事象 E の生起確率を q と報告してその後事象 E が生起した時に専門家に与えられる報酬であり，$R2(q,\bar{E})$ は同じような報告をした時に事象 $\bar{E}$ が生起した場合に与えられる報酬である．ここで $f(q)$ は q の増加関数であり，$F'=f$ である．

（2）式で示されるような報酬制度のもとでは，専門家にとっては期待報酬 $R(q,P)$ を極大化する q は常に P であることは容易に示されよう．

$$R(q,P) = P\times R1(q,E)+(1-P)\times R2(q,\bar{E})$$
$$= Pf(q)+F(q)-qf(q)$$

ここで $F(q)=q^2-q+1$ の時はブライヤーズ・スコアになり，F がエントロピー関数の時は log スコアとして知られているものになる．

このようなプロパー・スコアリング・システムにおいては，専門家は報酬（スコア）以外には無関心と仮定されている．しかしそのような仮定は非現実的であるという批判が存在する．そのような批判に対して否定はしないが次のような考え方をしてみよう．

いま，専門家は P と評価した事象 E について q と報告すれば，社会（すなわち意思決定主体であるマネジメント以外を指す）より，事象 E が生起すれば $f(q)$，事象 $\bar{E}$ が生起すれば $g(q)$ の利得を得るものとする．したがって専門家の報告 q は，次に示される専門家の期待利得 V を極大にするが必ずしも P とは限らない．

$$V = P\times f(q)+(1-P)\times g(q) \tag{3}$$

ここでマネジメントは適切なインセンティブ U を専門家に与えることにより正直な報告，$q\equiv P$ となるように U を構成できるかという問題を考えることにする（Pratt and Zeckhauser[1980]）．

しかし U は報告された q のみに依存するものとし（何の事象が生起するかには依存しない），マネジメントは f, g について知っているとする．したがって専門家の利得は，E が生起した時 $R1$，$\bar{E}$ が生起した時 $R2$ となり，それぞれ

$$R1 = f(q)+U(q),\quad R2 = g(q)+U(q) \tag{4}$$

と表現される．ここで $U(t)$ を

$$U(t) \equiv F(t)-G(t)-tf(t)+tg(t)-g(t)$$

とすると，（4）の $R1, R2$ はそれぞれ

$$R1 = (1-q)[f(q)-g(q)]+F(q)-G(q)$$

$$R2 = -q[f(q)-g(q)]+F(q)-G(q) \tag{5}$$

となる（ここで $F'=f, G'=g$）．ここで $H(t)$ を

$$H(t) \equiv F(t)-G(t)$$

と定義すれば(5)の表現は

$$R1 = (1-q)H'+H$$
$$R2 = -qH'+H \tag{6}$$

となる．いま f, g が，$0 \leqq t \leqq 1$ の t に対して $f'(t)-g'(t)>0$ ならば $H''(t)>0$ となり(6)の $R1, R2$ の表現は $[0,\ 1]$ で凸関数である H から導出されたプロパー・スコアリング・システムと同等の報酬システムとなっている．専門家に対する社会からの報酬が既知である限り，$[f'(t)-g'(t)>0]$ という条件を満たせば適当な q のみに依存するインセンティブ U を使用することにより，正直な報告を誘因でき，その全体としての専門家への報酬はプロパー・スコアリング・システムと同等の形を取ることがここで示されたのである[2)]．

4. 異なる意見の統合

地質学者を複数利用するとどういう状況が生じるであろうか．A という地質学者は .7 で石油は出るというし，B という地質学者は石油のでる確率は .5 であるといったとする．マネジメントはどうすべきであろうか．このように情報を求めて専門家を複数求めれば，次には専門家の意見をどう統合するかの問題が生まれてくる．新薬の認可の権限をもつ行政当局は，副作用に関して相対立する専門家たちの意見で動きがとれなくなることもある．銀行での資金運用担当者は，行内の複数のエコノミストに短期金利の動向を聞いてまわって，異なる見解に運用計画が立てられなくなるかもしれない．専門家たちが集まれば一般に情報は増加するのであろうが意思決定者にとっては情報の統合という面倒な作業が出現することになる．

専門家による意見を収束させる1つのアプローチはデルファイ法であろう．しかしデルファイ法は収束のプロセスをすべて専門家に依存する方法であり，時間に制約があるような意思決定には向いていないといえよう．専門家の意見・情報の荷重平均，もしくは単純平均を取ることも検討・実験されていて，そして荷重を過去の実績で決定することも考えられてはいるが荷重決定の恣意性に問題があるとの批判は強い．しかし時間制約という面の強い意思決定においては意思決定者サイドより積極的に制御できるような専門家の意見統合メカニズムの開発が望まれている[3)]．

5. 重複部分を含む情報の統合

第3節においては対立するような専門家の意見，情報を統合する方法の必要性を指摘したが，複数の専門家を利用する際，彼らの間に重複する情報がある可能性は少ない．意思決定者であるマネジメントは原情報ではなく，加工情報（確率情報や統計情報）の形で専門家の意見を受けとる方が多いため，この重複する情報により“過度”に影響を受けることは望ましいことではない．そこで部分的に重複する専門家の意見・情報をどのように統合するかという問題に直面することになる．

ここではこの重複情報の問題を非常に簡単なモデルで考える．現実はしばしば複雑でこのモデルでカバーできない点も多いが，それでも重複情報の持つ興味深い結果が得られるのである．

モデルとしては，精度（precision）が既知（ここでは1としても一般性を失わない）で，平均 Θ が未知である正規分布からのランダム・サンプリングを想定する．意思決定者の決定問題は Θ に依存しているとする．そして Θ の事前分布に関しては平均 μ，精度 $\tau(-\infty<\mu<\infty, \tau>0)$ の正規分布とする．したがってランダム標本 $X_i=x_i(i=1, 2, \cdots, n)$ の時，Θ の事後分布は平均 μ'，精度 $\tau+n$ の正規分布となる．

$$\mu' = \frac{\tau\mu+n\bar{X}}{\tau+n}, \quad \bar{X} = \sum_{i=1}^{n} x_i/n \tag{7}$$

ここで次のような状況を考える．2人の専門家（意思決定者とは他人という意味）が存在し，それぞれ共通の正規分布より，標本数，n_1, n_2 のランダム標本を観測して，標本平均，標本数，それぞれ $Z_i, n_i(i=1, 2)$ を報告してくる．ところが彼らの観測した標本に m 個だけ重複していることが分かった

がどの標本かは識別できない．このような状況で，意思決定者としては2人の報告をどう統合すべきであろうか．Z_i, $n_i(i=1,2)$ および m 個の重複標本の存在という条件下で，Θ の事後分布はどのようになるのかという問題である（Zeckhauser[1971]）．

この時，Θ の尤度関数は同じ正規分布からの標本平均Z，標本数 N のランダム標本と同等になることが示されていて，Z, N はそれぞれ

$$\begin{aligned} Z &= [Z_1(n_1-m)+Z_2(n_2-m)]/(n_1+n_2-2m) \\ N &= n_1+n_2-m-\eta \\ \eta &= 1\Big/\left(\frac{1}{n_1-m}+\frac{1}{n_2-m}+\frac{1}{m}\right) \end{aligned} \tag{8}$$

ここで統合された独立の標本数は (n_1+n_2-m) であるが，どの標本が重複しているか識別できないための情報の損失を，独立な標本数で表現した指標が η と考えられる．

さて専門家が2人から3人になった場合どうであろうか．2人のケースと異なり重複のパターンが変わってくる．3人の中で2人ずつに重複する標本も存在し，また，3人全員に重複する標本もあり得る．このような時の Θ の尤度関数を考えてみよう．n_i, Z_i をそれぞれ専門家 $i(i=1,2,3)$ の標本数，および標本平均とする．そして m_{ij} を専門家 i, j 間で重複している標本数，そして，m_0 を3人すべてに共通している重複標本数とすると，このような $(n_i, Z_i, m_{ij}, m_0\ i \neq j; i, j=1,2,3)$ の報告を受けた意思決定者にとり，Θ の尤度関数は標本数 N，標本平均 Z のランダム標本と同等になる．ここで N, Z はそれぞれ

$$\begin{aligned} N &= (n_1, n_2, n_3)\Sigma^{-1}\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} \\ \Sigma &= \begin{bmatrix} n_1 & m_{12} & m_{13} \\ m_{12} & n_2 & m_{23} \\ m_{13} & m_{23} & n_3 \end{bmatrix} \\ Z &= (n_1Z_1, n_2Z_2, n_3Z_3)\Sigma^{-1}\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}\Big/N \end{aligned} \tag{9}$$

である[4]．

たとえば，意思決定者が3種類のケースからの情報の選択に直面したとしよう．A ケースは3人の専門家がぞれぞれ10個の標本を有しているが，かれらの2人ずつには2個の重複標本が存在することが分かっている．B ケースでは2人の専門家が12個の標本を有しているが2個の重複標本が存在する．C ケースでは1人の専門家が22個の標本を有している．どのケースを選択すべきであろうか．(8)(9)によれば A, B ケースはそれぞれ213/7および204/7相当のランダム標本と同等の情報量ということになり，まず C ケースを選択すべきということになる．

以上の簡単なモデルでさえ重複情報のある場合の“衆知を集める”ことへの過信の修正の必要性を示しているといえる．また重複の評価の容易さという視点に立つと(9)式の N は次のようにも表現できるので

$$N = (\sqrt{n_1}, \sqrt{n_2}, \sqrt{n_3})R^{-1}\begin{pmatrix} \sqrt{n_1} \\ \sqrt{n_2} \\ \sqrt{n_3} \end{pmatrix} \quad (R:\text{相関行列}) \tag{10}$$

もし意思決定者にとって専門家間の重複標本の数よりも R，すなわち専門家の“相関の程度”の方が評価しやすい場合は(10)を専門家の選択に用いる方がよいであろう．

不確実性の下の意思決定分析は個人の意思決定における“直観”をいかにレベルアップすべきかという課題に取り組んできたといえる．そのために最初に述べたようなデシジョン・ツリー・アプローチ，すなわち“divide and conquer”戦略が有効であることが実証的に示されてきている．しかし環境の複雑度が増加する状況において，意思決定者ことに企業のマネジメントにとっては気象の専門家にも，法律の専門家にも，そして技術予測の専門家にもなることは不可能にちかいであろう．そこで不確実性の評価，事象への確率付与の意思決定を他人に委ねざるを得なくなってきている．そして委ねら

れる他人をここでは専門家として呼んできたのである.

そして意思決定のプロセスの一部を他人に委ねることにより不確実な対象についての判断（judgement）という責務は免れたものの，専門家という他人により提供された情報をどう処理するかという新たな問題に直面することになる．環境の多様性の増加により，まさにマネジメントの固有の仕事である“他人を通じて（through others)”のプロセスを意思決定プロセスにも適用せざるを得ない段階にきているといえよう．従来の意思決定研究が個人の意思決定を軸に展開されてきたが，“衆知を集める”意思決定論，すなわちチームワークにもとづく意思決定論の理論的および実際的な研究の分野に対して，より多くの研究資源を配分すべきではなかろうか.

注

* 本コンファレンスにおける参加者，とくに大塚英作氏のコメントに感謝する.

1) 意思決定者（マネジメント）と専門家（地質学者）との間に利害関係がなければ任意の報告になるのであろうといわれている．すなわち，正直である必要も正直でない必要も専門家には存在しないからである．しかし，専門家には意思決定者以外の人々（ここでは社会という）との利害関係が存在すると考えられるのでその点を考慮しなければならない.

2) 正直な報告へのインセンティブ・システムにはプロパー・スコアリング・システム以外にも，種々の方法が考えられている．専門家に対する意思決定者の賭の形をとるベッティング・システムもGrayson[1960]やAoi[1981]で研究されている．またボーナス制度やベッティング・システム，そしてスコアリング・システムの変形の有効性の比較分析が大塚英作（横浜国大）によってなされている．また，このようなインセンティブ・システムが一方では専門家の“うで”をみがくインセンティブにもなることも留意してほしい.

3) デルファイ法の定式化の試みとして，専門家たちが他の専門家の意見を参考にするプロセスにマルコフ・プロセスを応用したDeGroot[1974]がある．また専門家の意見をデータとして処理するベイジアン的接近を試みたのがMorris[1977]である.

4) 正規分布からのサンプリングの時は m_0，すなわち3人で共有している標本数という情報は m_{ij}，すなわちそれぞれ2人の間で共有している標本数が分かっていれば必要ではない．しかし他の分布，たとえば2項分布からのサンプリングではそうはいえない（Aoi[1981]).

参考文献

Aoi, M., 1981 “Reporting about Uncertainties by Several Experts”, Doctor Thesis, Harvard Business School.

DeGroot, M. H., 1974 “Reaching a Consensus,” *Journal of the American Statistical Association*, 69.

Grayson, C. J., 1960 *Decision under Uncertainty : Drilling Decision by Oil and Gas Operators*, Harvard Business School.

Kahneman D., P. Slovic and A. Tversky, 1982 *Judgment under uncertainty : Heuristic and Biases*, Cambridge University Press.

McCarthy, J., 1956 “Measures of the Value of Information,” *Proceedings of the National Academy of Sciences*.

Morris, R. A., 1977 “Combining Expert Judgement : A Bayesian Approach,” *Management Science*, 23.

Pratt, J .W. and R. Zeckhauser, 1980 “Incentive-Based Decentralization : Expected Externality Payments Induce Efficient Behavior in Groups,” Discussion Paper 83 B, Kennedy School of Government, Harvard University.

Raiffa, H., 1968 *Decision Analysis*, Addison-Wesley.

Savage, L. J., 1971 “Elicitaion of Personal Probabilities and Expectations,” *Journal of the American Statistical Association*, 66.

Winkler, R. L., 1960 “Scoring Rules and the Evaluation of Probability Assessors,” *Journal of the American Statistical Association*, 64.

Zeckhauser, R. J., 1971 “Combining Overlapping Information,” *Journal of the American Statistical Association*, 66.

第10章
意思決定モデルの構造上の諸問題

梅沢　豊

1. はじめに

ベイズ意思決定理論は，そのモデルの構造が明確で比較的に単純であるわりには，そして，不確実性下の合理的選択に関する期待効用最大化原理等にみられるごとくミクロ経済学の基礎理論の一翼を構成するほどの一般性を認められているわりには，現実の企業経営における種々の意思決定にはあまり適用されていない．このことは，つとにSimon[1955]によってきびしく指摘されているとおりである．その理由としてどのようなことが考えられるであろうか．意思決定問題を展開型で分析するためには，(1)問題を決定樹に表し，(2)最終結果には効用を付与し，(3)偶然分岐点から出ている分肢（自然の状態あるいは観測結果に対応する）には確率を付与し，(4)期待値計算の逆向過程をたどる，という4つの操作が必要になるが，(2)については，最終結果が多属性である場合にどう扱うか等の問題があり，また(3)についても，(1)とのからみで自然の状態の集合をどのように分割するか，確率付与を自然の状態と観測結果のそれぞれの組合せに対する同時確率で行うか，それとも周辺確率と条件つき確率に別けて行うか，別ける場合どちらを周辺確率にするか等，やっかいな問題が多い．Raiffa[1968], Kahneman, Slovic and Tversky[1982]等は，これらの諸問題に対処する方策にも考察を加えているが，これらが非常に大きな困難を伴う問題であることに変わりはない．また，Kahneman and Tversky[1979]は，かなり高い教養を持った人々の集団でさえ，"The Certainty Effect"（確実性効果），"The Isolation Effect"（分離効果）

などと呼ばれる，通常の合理的選択の公理系で仮定されている推移性，代替可能性の公準などとははっきりと矛盾する選択行動を示すことを報告している．いずれにせよ，意思決定者が合理的選択の公理系と整合的な効用および確率の付与を行うことは至難の業であり，それゆえにこそ，現実の意思決定分析において，これら主観的数値の付与を行うことを要求される当該意思決定者の心理的抵抗感はかなり強いものとなるのである．

以上は，一応この理論の枠組を認めたうえで，そこで仮定されている合理的選択の公理系に則して分析を進めようとする際に遭遇する困難性について述べたものであるが，以下本章の第2節以降では，第1節で定義する一般的なベイズ意思決定理論の枠組自体に，このモデルの現実への適用可能性を著しく制約している原因がありはしないかという点を，具体的実例の検討を通じて考察してゆくことにする．なお，第3節の例は Magee[1964] に，第4節の例は Holloway[1979] にそれぞれ拠っている．

2. ベイズ意思決定モデルの基本構造

ベイズ法定理論において，不確実性下の意思決定問題は一般に次のようにモデル化されている．A を意思決定者が選択できる代替案 a の集合，Θ を代替案選択の段階で意思決定者にはどれが実際に生ずるか確実にはわかっていない自然の状態 θ の集合とする．O を，意思決定者が代替案 a を選択し，自然の状態が θ であった（あるいは，θ となった）ときに得られる最終結果（利得）$r(\theta, a)$ の集合とする．r は直積 $\Theta\times A$ から O への写像であって利得関数という．任意の $\theta\in\Theta, a\in A$ について，$r(\theta, a)$ の効用を $u(\theta, a)$ で表す．効用関数 u は直積 $\Theta\times A$ から実数空間 R への写像である．自然の状態 θ の確からしさについての意思決定者の判断を，Θ 上の確率分布 $p(\theta)$ で表す．これを事前分布という．確率分布 $p(\theta)$ に従う確率変数を $\tilde{\theta}$ で表す．また意思決定者は代替案選択に際して，任意の $\theta\in\Theta$ に対し $\tilde{\theta}=\theta$ を条件とする条件つき分布 $p(x|\theta), x\in X$ に従う確率変数 $\tilde{x}$ の値を情報として入手することもできる．X を標本空間という．

以上のように，不確実性下の意思決定問題を構成する基本的要素および関係は，(1)代替案の集合 A，(2)自然の状態の集合 Θ，(3)利得関数 $r(\theta, a)$, $\theta\in\Theta, a\in A$，(4)効用関数 $u(\theta, a)$, $\theta\in\Theta, a\in A$，(5)標本空間 X，(6)Θ 上の確率分布（事前分布）$p(\theta), \theta\in\Theta$，(7)$\theta$ を条件とする X 上の条件つき確率分布 $p(x|\theta), x\in X, \theta\in\Theta$，の7つである．

任意の観測結果 $x\in X$ に，ある代替案 $a\in A$ を対応させる方式，つまり X から A への写像を決定関数，決定方式，戦略などといい，d で表す．d の集合を D で表す．観測結果 $x\in X$ に対して，決定関数 $d\in D$ を用いて代替案選択を行ったときの期待効用を $U(d|x)$ で表す．次式を満足する決定関数 $d^*\in D$ が存在すると仮定する．

$$
\begin{aligned}
U(d|x) &= E[u(\tilde{\theta}, d(x))|\tilde{x}=x] \\
&\leqq E[u(\tilde{\theta}, d^*(x))|\tilde{x}=x] \\
&= U(d^*|x).
\end{aligned}
\tag{1}
$$

観測結果 $x\in X$ に対しては代替案 $d(x)$ を選択する場合に得られる最終結果の期待効用が存在するとし，これを $U(d)$ で表せば，上式により

$$
\begin{aligned}
U(d) &= E[u(\tilde{\theta}, d(\tilde{x}))] \\
&= E[E\{u(\tilde{\theta}, d(x))|\tilde{x}=x\}] \\
&= E[U(d|\tilde{x})] \\
&\leqq E[U(d^*|\tilde{x})] \\
&= U(d^*)
\end{aligned}
\tag{2}
$$

が成立する．d^* を最適決定関数（あるいは，最適決定方式，最適戦略）という．

自然の状態の集合 Θ および代替案の集合 A が次に示すように有限集合であるとする．

$$\Theta=\{\theta_1, \theta_2, \cdots, \theta_m\}, \quad A=\{a_1, a_2, \cdots, a_n\}$$

代替案 a_j を選択し，自然の状態が θ_i であったときにもたらされる最終結果 $r(\theta_i, a_j)$ を簡単のために o_{ij} で表せば，利得関数 r は表1のようなマトリックスで表示される．これを利得表という．

表1　利　得　表

自然の状態	代替案			
	a_1	a_2	$\cdots$	a_n
θ_1	o_{11}	o_{12}	$\cdots$	o_{1n}
θ_2	o_{21}	o_{22}	$\cdots$	o_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
θ_m	o_{m1}	o_{m2}	$\cdots$	o_{mn}

3. 確率付与について

仕入量の決定問題　A君は明日の大学祭でクラブの仲間達と屋台でホットドッグを販売する計画をたてている．材料はパン，ソーセージ，マスタード，ケチャップ等で，ホットドッグ1本分の材料費はしめて130円になる．明日はこれを1本250円で売ることにしている．したがって順調にゆけば，売上げ1本当り120円の利益があがる勘定になっている．ただし，これらの材料はいずれもなまものであるから，売れ残ったらその分の材料費はすべて損失になるので，A君は，もし売残りが出そうになった場合には，1本100円で安売りして，とにかく仕入れた分はすべて売りつくしてしまうことにしている．

このようなわけで，品切れを起こさぬよう，また売残りも生じぬよう，うまい仕入れをしたいと考えているが，なにしろ屋外での営業なので，客足が多いか少ないか，したがって沢山売れるか売れないかは，明日の天気次第である．A君たちは，昨年までの同様の経験から，ホットドッグの売上げは，天気が晴（θ_1）ならば500本，曇（θ_2）ならば300本，雨（θ_3）ならば100本となることを知っている．したがって，ホットドッグ何本分の材料を仕入れるかの意思決定は，当然のことながら明日の天気をどう予測するかにかかっているのである．

今朝，A君が家を出る前に朝刊の天気予報を見たら，明日は「曇所により一時雨後晴」となっていた．昨日から降りはじめた雨が今日も午前中いっぱい降り続き，昼すぎにようやくやみはしたが，その後も厚い雨雲がたれこめていた．明日は休日でほとんどの商店はお休みになるので，材料は今日中に

仕入れておく必要があるが，空模様の推移を見守るため，A君は午後も材料仕入れに出掛けるのを見合わせていた．しかし夕方になり，商店の閉店時刻も近づいてきたので，とうとう1人の後輩を伴って買出しに出かけることにした．商店街の入口までやってきたA君は，角の電話ボックスから177番にダイヤルしてみた．受話器から流れてきた最新の天気予報は，「まだ気圧は不安定な状態が続いているが，天気の回復はいくぶん早まる見込み」というものであった．ふと西の空に目をやると，この天気予報を裏付けるかのように厚い雨雲の端の部分が夕焼けであかね色に輝いているのが見えた．A君はこの最新の天気予報（x で表す）を聞き，夕焼け（y で表す）を見たあとで，明日の天気が晴，曇，雨になる確率は，それぞれ0.6，0.3，0.1であると判断した．つまり，これらの情報を得たのちに自然の状態 $\theta_1, \theta_2, \theta_3$ に関してA君がもつ条件付き確率（事後確率ともいう）は，それぞれ，$p(\theta_1|x\cdot y)=0.6$, $p(\theta_2|x\cdot y)=0.3$, $p(\theta_3|x\cdot y)=0.1$ であった．

さて，A君は何人分の材料を仕入れるべきであろうか．

まずこの問題の簡単な分析から開始しよう．a 人分の材料を仕入れるとして，$a=100, 300, 500$ それぞれの場合の利得表が表2に示してある．500人分仕入れて天気が曇のときは，すなわち θ_2 で $a=500$ のときは，単価250円で300本が売れて，120(円)×300＝36,000(円) の粗利益があがるが，売残りの200人分は単価100円分で処分しなければならないので，1本当り30円の損失が発生し，この損失は総額で30(円)×200＝6,000(円) となる．したがって，この場合の純利益は36,000(円)－6,000(円)＝30,000(円) になる．また，$a=100$ のときは，天気のいかんにかかわらず仕入れた分はすべて単価250

表2　仕入れ問題の利得表

明日の天気	仕入れ量（人分）			
	100	300	500	事後確率
晴（θ_1）	1.2 万円	3.6 万円	6 万円	0.6
曇（θ_2）	1.2	3.6	3.0	0.3
雨（θ_3）	1.2	0.6	0	0.1
期待利益	1.2	3.3	4.5	

表3 $100 \leqq a \leqq 300$ の場合の期待純利益

天 気	純 利 益	事 後 確 率
晴 (θ_1)	$120a$	0.6
曇 (θ_2)	$120a$	0.3
雨 (θ_3)	$15{,}000-30a$	0.1
期待純利益	$105a+1{,}500$	

表4 $300 \leqq a \leqq 500$ の場合の期待純利益

天 気	純 利 益	事 後 確 率
晴 (θ_1)	$120a$	0.6
曇 (θ_2)	$45{,}000-30a$	0.3
雨 (θ_3)	$15{,}000-30a$	0.1
期待純利益	$60a+15{,}000$	

表5 $a \geqq 500$ の場合の期待純利益

天 気	純 利 益	事 後 確 率
晴 (θ_1)	$75{,}000-30a$	0.6
曇 (θ_2)	$45{,}000-30a$	0.3
雨 (θ_3)	$15{,}000-30a$	0.1
期待純利益	$60{,}000-30a$	

円で売りさばけるから，利益は120(円)×100=12,000(円) になる．

事後確率に関してこれらの純利益の期待値をとると，表のように $a=500$ のときの 45,000 (円) が3つの代替案の中では最高である．期待純利益最大化の基準に従うかぎり，500人分仕入れる，というのが最適であることが分かる．

次に，一般の a に対する期待純利益を計算してみよう．$a \leqq 100$ の場合には，天気に関係なく完売となるから，純利益は $120a$ となる．$100 \leqq a \leqq 300$ の場合には，晴 (θ_1)，曇 (θ_2) のときには完売となるが，雨 (θ_3) のときには100本しか売れなくて，($a-100$) 人分の売残りが発生するから，純利益は120(円)×100−30(円)($a-100$)=(15,000−30a)(円) となり，期待値をとると表3のように ($105a+1{,}500$)(円) となる．さらに $300 \leqq a \leqq 500$ の場合には，完売となるのは晴 (θ_1) のときだけで，曇 (θ_2) のときには ($a-300$) 人分の売残りが発生するから，純利益は120(円)×300−30(円)($a-300$)=(45,000−30a)(円)，雨 (θ_3) のときには，同様にして，純利益は120(円)×100−30(円)×($a-100$)=(15,000−30a)(円) となる．期待純利益は，表4のように，($60a+15{,}000$)(円) となる．最後に，$a \geqq 500$ の場合には，天気のいかんにかかわらず売残りが発生することになるが，計算上は上述の場合と大体同じである．期待純利益は表5に示すように，(60,000−30a)(円) となる．

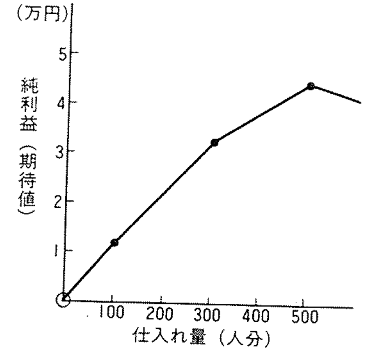


図1 仕入れ量と期待純利益のグラフ

以上をまとめてグラフに表したのが図1である．このグラフから，この場合の最適仕入量は，やはり500人分であることが分かる．

以上の分析結果から明らかなように，この仕入問題は，大学祭という特殊な舞台設定ではありながら，現実の企業の販売活動において日常的に生じる，需要が不確実な場合の在庫量（あるいは生産量）決定の問題の一般的特徴を十分よく備えている．さて，この問題を通常の決定理論の枠組に沿って分析しようとすると，どのようになるであろうか．意思決定者すなわちA君は，明日の天気が晴 (θ_1)，曇 (θ_2)，雨 (θ_3) となる確からしさについての判断確率を付与する．さらに，夕方の天気予報が「天気の回復は早まる見込み」であり (x)，西の空が夕焼けに輝く (y) ということが生じる（標本が得られる），自然の状態 $\theta_i (i=1, 2, 3)$ を条件とする条件つき確率 $p(x \cdot y|\theta_i)$ が必要になる．しかし，標本すなわち事象 $x \cdot y$ の生起は今夕であり，一方自然の状態すなわち事象 θ_i の生起は明日であるから，θ_i を条件とする事象 $x \cdot y$ の条件つき確率 $p(x \cdot y|\theta_i)$ を云々することは，事象発生の時間的前後関係あるい

は因果関係にてらして無理がある．前述したように，確率分布 $p(x\cdot y|\theta_i)$ のパラメータとして性格づけられる θ_i は，本来，事象としての $x\cdot y$ を規定する要因でなければならない．しかるに，上述のように，ここでは原因と結果の順序が完全に逆転してしまっているのである．

ある種の問題においては，標本 x がとられる時点ですでに自然の状態は確定していて，標本分布がこのすでに確定している自然の状態に依拠して定まる（自然の状態が標本分布のパラメータとして機能する）という関係が，実態的にも確かに成立している．しかし，この仕入問題にみられるように，真の自然の状態 θ_i が実態的にも標本分布を規定するという関係が成立していない例が，むしろ一般的である．そもそも意思決定を1つの過程として時間的経過を追ってとらえる場合，意思決定者が最終的に1つの代替案を選択しその結果が生じてこの過程がひとまず終結するまでの諸段階で，そこまでに入手した，その問題がらみの状況推移についての情報にもとづいて，意思決定者がその問題にかかわる不確実性に関する判断を修正してゆく過程が存在していることはひろく認められているとおりである（Simon[1977], Ch. 2)．このようにダイナミックな過程として把握される意思決定過程の中に前節で定式化した1段階の意思決定問題を位置づけるとすれば，それは，最終的に代替案を選択する直前に最後の標本すなわち情報を得た（あるいはまさに得ようとしている）意思決定者の目から見た代替案選択問題である，ということができよう．意思決定者はそこまで到達する過程で各種の情報を得ているであろうが，それらはすべて意思決定者の判断で合理的に処理されて，現在意思決定者が持っている Θ 上の事前（主観的）確率分布 $p(\theta)$ に体化されていると想定する．この体化の過程は理論のうえではブラックボックスになっていて，どのような処理が実際に行われているかについては陽表的には扱われない．意思決定問題は，最後の情報を得る段階以降，真の自然の状態と選択された代替案とによって最終結果が決められる時点までの過程を対象としており，なかでも情報にもとづいて“最適”代替案を選択する方式の導出を主たる課題としている．

このように一般に意思決定問題として定式化される状況・局面は，もともとは時間の経過を伴ったダイナミックな意思決定過程の一部分であって，情報の入手と自然の状態の生起の間には一定の時間の経過が存在している場合が多い．しかも，当然のことながら，先に情報の入手があり，おくれて自然の状態が定まってくるという場合がほとんどである．つまり真の自然の状態が標本分布を規定するという関係が厳密な意味で成立するような意思決定過程は，一般の意思決定過程のうちではほんの一部にすぎないのである．これが，現実の不確実性下の意思決定の多くが統計的決定理論のモデルどおりには分析されえない，第1の理由である．

前節の(2)から明らかなように，決定関数 $d\in D$ の評価測度である期待効用 $U(d)$ を計算するためには，直積 $\Theta\times X$ 上の確率分布 $p(\theta, x)$ が必要である．この同時分布は伝統的な決定理論においては，事前分布 $p(\theta)$ と標本分布 $p(x|\theta)$ から

$$p(\theta, x)=p(\theta)\cdot p(x|\theta) \qquad (3)$$

によって導出されるのが通常である．上述のように，事前分布 $p(\theta)$ が形成されるメカニズムには理論は立入らない．これは，単に意思決定者がその時点まで得た経験や情報等あらゆるものが意思決定者の思考に作用して体化したものとされている．この意味で，$p(\theta)$, $\theta\in\Theta$ を，主観的確率分布ないしは判断確率分布としばしば呼称する．一方，標本分布 $p(x|\theta)$ については，過去からのデータの集積があり，頻度論的な確率分布の付与が可能であるとか，理論モデルから導出される理論分布が存在する等の理由により，$p(\theta)$ に比べてこちらの方がより客観的な確率分布であると主張される．

さて，決定関数 $d\in D$ に対して $U(d)$ を計算する必要が生じるのは，2つの決定方程式 $d, d'\in D$ を比較する場合か，最適決定方式 d^* を求める場合である．単に，ある標本観測値 x が得られたとして，このときとるべき最適代替案 $d^*(x)$ を求めるためだけなら，前節の(1)式から明らかなように，必要なのは x が与えられたという条件つきの Θ 上の確率分布 $p(\theta|x)$, $\theta\in\Theta$ のみである．この条件つき分布は，もちろん，(3)の同時分布が与えられれば容易に導出できる．しかし，(3)の同時分布を導く過程でもしも主観分布 $p(\theta)$ を用いているのなら，わざわざこのようなまわり道をしなくても，標本観測値 x を得てから，じっくりと考えて，この x をも体化させた主観的確率分布 $p(\theta|x)$ を付与すればよいという論理も十分に成立ちうるのである．と

いうのは，上述のように，そこに定式化されている意思決定問題は一連の意思決定過程のうちの最終局面を取出したものにすぎないのであって，意思決定者はそこに到達するまでにも，Θ 上の事前確率分布 $p(\theta)$ に体化されたところの数多くの経験や情報を得る過程を経てきているわけであって，$p(\theta)$ を適切に計算する能力を意思決定者に対して仮定するのであれば，この局面においても“主観的確率分布” $p(\theta|x)$ を適切に計算できると仮定してもいっこうに不都合はないからである．

逆に，もし(3)の同時分布から計算によって導出された条件つき確率分布 $p(\theta|x)$ が，標本観測値 x を実際に得たあとに意思決定者が“主観的確率分布” $p'(\theta|x)$ を付与したとして，この事後的な“主観的確率分布” $p'(\theta|x)$ と一致しない場合には，いずれをとるべきであろうか．答は明白に後者である．両者が一致しない理由はいろいろとあろう．たとえば，(3)の右辺の $p(\theta)$ または $p(x|\theta)$ が適切に付与されていなかったということかもしれない．また，$p(\theta)$ あるいは $p(x|\theta)$ を付与したときと，現に標本観測値 x を得たときでは，意思決定者の不確実性に関する判断が整合性を欠いたということかもしれない．しかし，これらいずれの場合でも，現時点での判断に対して以前の判断を優先させるべき理由はどこにも見当たらない．

このように考えてくると，結局，実際に得られた標本観測値 x に対して最適代替案 $d^*(x)$ だけを求めればよいときに，(3)の同時分布 $p(\theta, x)$ から計算によって導出される条件つき分布 $p(\theta|x)$ を用いるのが適当である場合というのは，唯一，(3)の右辺の $p(\theta)$ および $p(x|\theta)$ が適切に付与され，かつ，標本観測値 x を得た段階で事後的な“主観的確率分布” $p(\theta|x)$ を付与することが意思決定者に不可能な場合にかぎられる，ということができるのである．

決定関数の間の比較とか最適決定関数の導出のためには，上述のように，直積 $\Theta\times X$ 上の同時確率分布 $p(\theta, x)$ が必要になる．この同時確率は(3)以外に

$$p(\theta, x) = p(x)\cdot p(\theta|x) \tag{4}$$

によっても求められる．鈴木[1978]（第3章）がはっきりと述べているように，意思決定モデルの構成要素として，(i)Θ 上の事前確率分布 $p(\theta)$ と条件 θ が与えられたときの X 上の条件つき確率分布 $p(x|\theta)$ の組合せ，(ii)X 上の確率分布 $p(x)$ と条件 x が与えられたときの Θ 上の条件つき確率分布 $p(\theta|x)$ の組合せ，(iii)直積 $\Theta\times X$ 上の同時確率分布 $p(\theta, x)$，の3つは同等である．(i)からは(3)により，(ii)からは(4)によって，それぞれ(iii)の同時確率分布が求められるからである．伝統的な決定理論のモデルがみずからを(i)の構成のみに限定してしまっていることの問題性はすでに指摘したとおりであって，(ii)および(iii)の形での確率付与も可能となるようなモデル構築をしておくことが論理的にも，またモデルの適用可能性を拡げるという実践的必要性からも，肝要である．

4. 利得関数の特定について

ガンマ化学の設備投資問題 ガンマ化学会社の経営首脳陣は，自社の研究開発部門が開発したある新製品を商品化するにあたり，いかなる規模の設備投資を行うべきかを検討していた．製品のライフサイクルを一応10年間と想定し，最初の2年間を導入期と定めた．この製品に対する需要は，さしあたり，次の3つのケースのいずれかであろうと考えていた．

I. 導入当初から需要は高水準で，10年間それが持続する．

II. 導入期には需要は高水準である．しかし顧客がこの製品にそれほど満足せず，あとの8年間需要は低迷する．

III. 初めから終りまで需要は低水準である．

このような状況のもとでガンマ化学の経営陣は，この製品を製造するために建設すべきプラントを大規模なものにするか，それとも小規模なプラントにするかの決定を迫られていた．ただし，小プラントを建設した場合には導入期における需要の動向を見たうえで2年後に大規模プラントへ拡張することもできるが，ひとたび大プラントを建設してしまった場合にはのちにプラントの規模を縮小することは技術的に不可能である．したがって選択可能な代替案は次の3つのいずれかであった．

a：大プラントを建設する．

b：小プラントを建設しておいて，導入期の需要が高水準ならば2年後に

拡張するし，低水準ならば拡張しない．

c：小プラントを建設し，のちの拡張は行わない．

さて，a, b, c のうちのどの代替案を選択すべきかを検討する過程で，この企業の首脳部は，将来の需要に関して，第1のケースの可能性は60%，第2，第3のケースの可能性はそれぞれ10%，30%であると判断した．

さらに，各代替案が選択された場合にそれぞれのケースにおいて得られる年間の利益が企画部門によって見積られた．当初から大プラントを建設した場合（代替案 a），需要が高水準なら（ケースI），利益は年間10億円，導入期に高水準で以後は低迷するなら（ケースII），あとの8年間は利益は年間1億円に減少し，初めからずっと需要が低水準なら（ケースIII），利益は10年間を通じて年間1億円となることが見込まれた．代替案 b をとった場合には，導入期にはケースI，IIともに供給力が不十分であるために，年間4.5億円しか利益があがらず，設備拡張後の3年目以降も中途増設による生産設備の非能率性ばかりでなく他社の参入にもわざわいされて，年間利益はケースIで7億円，またケースIIでは0.5億円になると算定された．代替案 c をとった場合の年間利益は，導入期にはケースI，IIとも4.5億円，のちの8年間はケースIで3億円，ケースIIで4億円になるであろうと推定された．この場合，需要の大きいケースIの方が需要の小さいケースIIよりも利益が逆に減少すると推定されたのは，他社の参入によって競争が激化することが考慮に入れられたためである．ケースIIIの場合には，10年間を通じて利益は年間4億円と見積もられた．なお，プラント建設に要する費用は，大プラント，小プラントそれぞれで30億円，13億円であり，2年後に小プラントを大プラントに拡張する場合には，さらに22億円の追加投資が必要であると算定された．

一方，代替案 a, b, c のいずれを選択するかを決定するまえに，市場調査部門に対してこの製品の需要予測を行わせるかどうかも検討された．この調査部門が行う予測の信憑性については，従来のほぼ同様なケースにおいてこの部門の行った需要予測の当り外れの実績を調べた結果，もしケースIが真であるならば，調査部門がこの製品は有望であるという調査結果を出す確率は0.7であり，ケースII，ケースIIIの場合には，それぞれこの確率が0.5，0.05

表6　ガンマ化学の設備投資問題の利得表

自然の状態	代替案	
	大プラント建設	小プラント建設
（I）HH	10/10億円	4.5/7　億円（拡張） 4.5/3　〃
（II）HL	10/1　〃	4.5/0.5　〃　（拡張） 4.5/4　〃
（III）LL	1/1　〃	4/4　〃
	−30　〃（建設費）	−13　〃　（建設費） −22　〃　（2年後の拡張費）

（注）1．H は需要が高水準，L は低水準であることを表す．HL は，導入期の2年間は需要が高水準であるが，あとの8年間は低迷することを表す．以下同様．

2．表の数字は，/（斜線）の左側がはじめの2年間の平均年間利益を，右側があとの8年間の平均利益を表す．

3．小プラント建設の欄は，上段が2年後に設備を拡張した場合の利益を，下段が拡張しない場合の利益を表す．

であることがわかった．市場調査部門に需要予測を行わせるための費用は1億円である．

この問題では，表6からも明らかなように，最終結果であるところの，この製品の製造販売から発生する年々の利益は，設備の規模と需要の水準の2要因によって完全に決まると仮定されている．世の中の化学製品の種類はおびただしいから，このようにきわめて単純な仮定が妥当する製品ももちろん存在はしているであろう．しかし，通常の状況はもっといろいろと複雑である．

利益は，まず，製品価格（したがって，売上高）に大きく依存する．ガンマ化学がこの製品について，ある程度の市場支配力を持っていて価格の安定を目ざしたとしても，競合他社の価格政策次第で，価格が大きく変動する可能性もありえよう．また，原油価格等のエネルギー価格をも含む原材料価格一般の変動も，費用の側面から最終利益に大きく影響する．ガンマ化学はこれらの原材料に関しては，需要家サイドにいるわけで受け身の立場であるから，価格に対して強力な支配力を持つとは考えられず，このコスト面からの収益への影響もかなり大幅なものになるであろう．

次に，10年というこの製品のライフサイクルの短縮を余儀なくされるような事態も，これだけ技術革新の激しい現今の環境下では十分に起こりえよう．数年をまたずに，性能面あるいは価格面でこの製品を凌ぐ競合新製品が出現するかもしれないし，この製品の川下に位置する産業に大きな変動が生じてこの製品に対する需要が消滅するかもしれない．かなり基礎的資材と目されている製品の分野でも，現在，このようなことは日常茶飯に生じている．

さらに，この製品の市場が海外にまで及んでいるか，もっぱら国内であるかにかかわらず，昨今では為替レートの変動も大きな不確実要因となってきた．海外へ輸出していれば，円ベースでの取引でないかぎり最終的には獲得した外貨を円に交換しなければならないであろうし，国内が主たる市場である場合には，輸入品との競争という側面でやはり為替レートの影響を受けざるを得ないからである．前述の原材料価格にも為替レートは直接的影響を及ぼすから，コスト面を通じての為替レートの収益への影響も無視しえないであろう．状況をなおいっそう複雑にしているのは，閉じた因果関係の存在，たとえば諸般の事情で価格が変動したとき，需要の価格弾力性のメカニズムその他を通じて需要水準それ自体にも変化が生じ，それがこの製品を供給している各企業の収益構造を変化させて製品価格のさらなる変更を引き起こす，といった因果連環の存在である．

ともあれ，以上のようなガンマ化学の設備投資問題についての検討からも明らかになったように，経営の実際の意思決定問題を第1節の冒頭で提示した伝統的な決定理論の枠組で定式化しようとすると，自然の状態 θ としてはいく種類もの要因をモデルに含めなければならず，しかもそれらのうちのいくつかの要因の間には複雑な相互作用が働いているというような有様であるから，それら何種類かの不確実要因と代替案の組 (θ, a) のそれぞれに対して最終結果を対応させる利得関数 $r(\theta, a)$ を特定するなどということはまさに至難の業なのである．以下，この点につきもう少し一般的に考察を行ってみよう．

企業経営における重要な意思決定問題のいくつかを，計画期間が長期，中期，短期のものにそれぞれ区分けして例示すれば，次のようになるであろう．長期の計画期間を持つものとしては，業態の確定とか研究開発戦略策定の問題などがあり，これより計画期間がやや短い，中期的なものには，この節で検討したような設備投資の規模やタイミング決定の問題，経営資源・労働力の確保にかかわる人事・労務問題，有利な資金調達を目ざす財務管理の問題，流通経路の開拓・整備や製品系列の充実等を目的とするマーケティング戦略策定の問題などがある．そして最後にごく短期的な問題としては，生産計画策定，販売促進・広告宣伝計画策定，運転資金管理などの問題がある．これらはいずれも，まさに典型的な不確実性下の問題であるが，どの1つをとっても，関連する要因の数はきわめて多く，それらのうちでどれが意思決定者にとって制御可能な要因（つまり代替案の範疇に含められるべきもの）であり，どれが制御不可能な要因（つまり自然の状態の範疇に含められるべきもの）であるのかさえなかなか判然とはしないほど諸要因が複雑に絡みあっている．Simon[1957] が主張しているように，これらのうちで組織全体にかかわりを持つような広範な問題においては，そこでの目標さえ明確化しえないことが多く，その場合には全体的な目標に対し代替的な機能をするものとして下位目標を使わざるを得なくなる．しかし，そのような下位目標に対してさえ，問題を構成している多くの要因のそれぞれがどのように関連しているか，すなわち，この下位目標にてらしてそれぞれの最終結果を測ることにしたときの利得関数 r の特定は，決して簡単ではない．伝統的決定理論のモデルの仮定，すなわち，「最終結果を規定する要因としては θ と a の2つがあり，これら2つが定まれば最終結果は $r(\theta, a)$ に定まる（つまり，関数 r が特定されている）．θ の生起には不確実性が伴うが，θ が定まれば，あとは意思決定者がいずれかの代替案 a を選択するのに対応して最終結果 $r(\theta, a)$ が確実に定まる（つまり，すべての不確実性は θ の一点に凝縮してしまっている)」という仮定どおりに上記の経営問題を定式化することはまず困難なのである．

ところで，経営の意思決定問題に通常ともなっているような不確実性の源泉としては，大別して，(1)一般の気象，災害（地震・風水害・火山爆発）などの天然・自然現象，(2)人間が全知全能でありえないこと，(3)合理的な競争相手の存在，(4)標本抽出，くじの当り外れのような物理的ランダムネス，などが挙げられる．これらのうち，(3)は経済学一般あるいはゲーム理論に固有

の分析対象であり，(4)は確率論あるいは統計学に固有の分析対象である．したがって，意思決定の問題と直接的なかかわりを有するのは(1)および(2)に起因する不確実性であるが，特に最近になって，これらのうちでも比較的に一般性のある不確実性に関しては，意思決定者が“望めば”それらの不確実性から生じるリスクをヘッジしうるような，保険とか先物の市場が急速に整備されてきた．したがって，それらを不確実要因として扱う必要はほとんどなくなったといっても過言ではなかろう．結局最後に残された不確実性は，(2)の一部の，意思決定者自身の問題把握能力の不足に起因するもの，すなわち意思決定者が不確実性下の意思決定問題らしきものに直面していることはわかっているにもかかわらず，その問題の構造がいかなるものであり，代替案はなになにで自然の状態はなになにで，したがって利得関数はこうである，というように問題を明確に定式化できないことにもとづく不確実性なのである．この結論が本節で行った具体例の検討から導出される含意と整合していることについては，まったく議論の余地がないであろう．

5. 代替案の探索

制御用コンピュータの受注・製造問題　今年の6月7日，ある自動制御機械メーカーの製造部門の管理者A氏は，取引先の化学会社からプロセス制御用の専用小型コンピュータの注文に応ずるか否かの引合いを受けた．その引合いは概略以下の3条件から成立っていた．(i)注文台数はこの時点では明らかではないが，40台（自然の状態 θ_1）か20台（自然の状態 θ_2）のいずれかであって，それは来年1月に確定する．(ii)価格は1台につき1000万円である．(iii)納期は来年3月であるが，この注文に応ずるか否かは，1週間以内に回答しなければならない．

A氏は部下達といろいろ検討した結果，この小型コンピュータを製造するには，在来の標準機種を与えられた仕様（スペック）に合うよう部分的に改良する（製法1），仕様に合わせて設計から製造まですべて新規に行う（製法2），外注に出す，という3通りの方法があることがわかった．製法1が成功すれば費用的には最も安上がりであるが，在来機種の部分的改良ではは たして仕様通りの性能を出せるか否かは，実際に改良作業を行ってみないとわからない．この改良作業を試みた場合，その性能試験の結果は9月に判明する．もし改良作業が失敗しても，それから製法2に切りかえる余裕は，時間的には十分ある．もっともこの場合には，製法1における改良作業に投ぜられた資金はすべて損失になる．製法1，2のいずれをとるにせよ，3月の納期に間に合わせるためには，1月に注文台数が40あるいは20のいずれかに決まる以前に，製造台数を確定して製造を開始しなければならない．第3の方法をとる場合の外注先は，大手のある電算機専業メーカーである．いま直ちに外注契約を結べば，比較的安い費用で契約できるが，9月以降の契約になるとこの特典は消える．ただ，いずれの場合にも，この外注先は十分な生産能力を備えているので，来年の1月までに注文台数を伝えれば，3月の納期に間に合わせてもらえることになっている．

製法1の改良作業に従事する可能性のある技術者達は，その改良作業が成功する確率は0.5であろうと考えている．A氏は発注元の化学会社との接触を通じて，注文が40台になる確立は0.4であろうとの判断を持った（θ_1 の判断確率は0.4）．この部門の技術および経理のスタッフ達は，諸費用を以下のように算定している．

〔製法1〕

改良作業費用	20,000千円
1台当り製造原価	4,000千円

〔製法2〕

1台当りの製造原価	6,000千円
1台当りの外注費用	
1982年8月以前の契約	7,000千円
〃 9月以降の契約	9,000千円

なお，以下の3点を仮定する．

1. 40台の注文が来たにもかかわらず自社内で20台しか製造していなかった場合には，不足分の20台は1台当り9,000千円で外注できる．
2. 20台分の注文しか来なかったのに，40台製造していた場合には，余分の20台は1台当り2,000千円で処分できる．

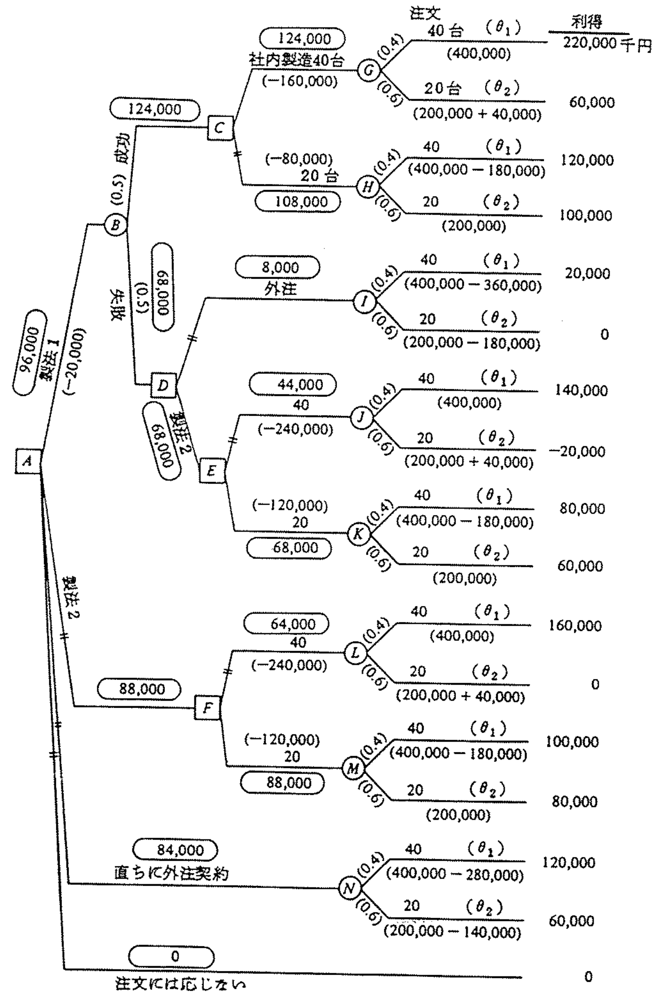


図 2　制御用コンピュータの受注・製造問題の決定樹

3. A 氏の効用は問題の金額の範囲内では線形である.

図 2 はこの問題を展開型で分析した決定樹である. この問題は, 注文された製品を, 内製するか外注に出すか, あるいは注文そのものを断わるか, という企業の営業・製造部門にまたがる 1 つの典型的な経営意思決定問題である. この問題自体は事後確率を計算する必要のない, 確率の観点からはきわめて単純な決定問題であるが, 第 1 節のモデルの(1)の代替案の集合 A がこの場合にいかなるものになっているかという観点でみると, 事情はにわかに一変する.

この問題では, 一番有利な, 自社内での汎用品の改良作業だけで要求仕様に合った製品を製造するという製法 1 が選択可能な代替案なのかどうかということは, 6～7 月段階では完全にはわからない. 別の表現を用いるならば, 一定の費用を投じて情報を収集することによって, はじめてこの代替案が選択可能な代替案の集合 A に属しているかどうかがわかるのである. ところがベイズ意思決定モデルでは不確実性は自然の状態のみにかかわっていて, 代替案の集合 A ははじめから確定していることになっている. 第 1 節の意思決定モデルが現実の経営意思決定問題の多くに対してあてはまりがよくない理由の 1 つがここにある.

しかし, 本稿で検討している 1 節のモデルはスタティックな (静学的, 1 段階の) モデルであって, ダイナミックな (動学的, 多段階の) モデルへこれを拡張しさえすれば, 特定の代替案が選択可能なものとなるのか否かといったここでの問題は, 少なくとも抽象的なレベルではモデルの内部に取込むことが可能になる. Wald[1947, 1950] は, 自然の状態にかかわる不確実性を緩和するための情報収集活動に限って多段階化を行っているにすぎないので, ここでの選択可能な代替案の集会云々の問題にはあまり関係がない. 一方 Bellman[1957], Bellman and Dreyfus[1962], Fel'Dbaum[1965], Aoki[1967], Howard[1971] 等は, 一般的なコントロールの問題を多段階決定問題として定式化し, それを解くためのダイナミック・プログラミング (以下 DP と略記する) と呼ばれる計画法を開発した. DP は「状態の変化を規定する関係式と, 下した意思決定の結果に対する評価関数とがある条件を満たす場合には, 最適決定方式は, 最初の状態および決定がどうであっても, その決定方式に従って定められる残存期間に対する決定が, 最初の決定の結果出てきた状態に関して最適になっていなければならないという特性を持つ」という最適性の原理を唯一の根本原理としている. この考え方は, すでに Wald の定式化の中でも使われていたものであって, とり立てて新しいもの

ではなかったが，この原理をきわめて広い分野の問題の分析に積極的に援用したDPの手法は，大容量の高速計算機の発達と相まって，多段階決定問題を実際に解くのにかなりの威力を発揮してきた．

しかし，以上はすべて確実性下のコントロールの問題に関していえることであって，不確実性下のコントロールの問題に関しては，宮沢[1970, 1971]，梅沢[1970]等による一般化への試みがなされてはいるものの，最適性の原理そのものがはたして一般的に成立するか否かさえ，厳密には明らかにされていない．それは，不確実性が入ってきた場合におけるコントロールの問題の，決定理論的な構造そのものがまだ一般性を持つような形ではモデル化されていないことに起因している．以上のような事情で，代替案の集合Aがこの制御用コンピュータの受注・製造問題にみられるような不確実性を伴う問題の分析へのDPの適用については，あまり明るい展望を持ちえないのである．

一般に，そのときまでに気がついていた代替案以外にもっとよい代替案がないものかと，いろいろな情報収集活動を行うことを探索という．この制御用コンピュータの問題のように，本来スタティックなモデルでは対応しきれないような問題で，代替案の集合Aに不確実性が伴っているようなものは，通常の意思決定問題の中に多数存在しているわけであるが，Simon[1955]は普通の人間が第1節のモデルでそれらを扱うのは無理であるとし，探索を内生変数化することによって一種のダイナミックなモデルを作り出した．これが限定的合理性を仮定したSimonの行動モデルである．

6. おわりに

本章では，ベイズ意思決定理論が現実の経営意思決定問題の分析にそれほど応用されていないという現状認識に立って，その広範な応用を困難にしている原因はなにかを実例の検討を通じて考察した．その結果，ベイズ意思決定モデルの現実の意思決定問題への適用可能性を著しく制約している原因として，(1)一方で情報（標本）の確率分布のパラメータとして機能しつつ，他方でいずれの代替案に対しても，どの最終結果が生ずるかを決める要因として機能している自然の状態という要素の存在を仮定していること，(2)モデルが静学的で代替案探索の局面が捨象されていること，が挙げられるとの結論に到達した．

このような結論が適切なものであるとすれば，次には自然の状態の存在を仮定しないモデルを構築し，さらにモデルを動学化するという課題に取組む必要があるであろうが，紙数の制限もあるので，これらの課題に対する検討は機会をあらためて行うこととする．

引用文献

Aoki, M., 1967 *Optimization of Stochastic Systems*, Academic Press, New York.

Bellman, R. E., 1957 *Dynamic Programming*, Princeton University Press, Princeton.

Bellman, R. E. and S. E. Dreyfus, 1962 *Applied Dynamic Programming*, Princeton Unversity Press, Princeton.

de Finetti, B., 1937 *Foresight : Its Logical Laws, Its Subjective Sources*. In H. R. Kyburg, Jr. and H. E. Smokler, *Studies in Subjective Probability*, John Wiley, New York, 1964.

Fel'Dbaum, A. A., 1965 *Opimal Control Systems*, Academic Press, New York.

Holloway, C. A., 1979 *Decision Making under Uncertainty : Models and Choices*, Prentice-Hall, Englewood Cliffs.

Howard, R. A., 1971 *Dynamic Probabilistic Systems*, Vol. I, II, John Wiley, New York.

Kahneman, D. and A. Tversky, 1979 "Prospect Theory : An Analysis of Decisions under Risk", *Econometrica*, 47, 263-291.

———, P. Slovic and A. Tversky (ed.), 1982 *Judgement under Uncertainty : Heuristics and Biases*, Cambridge University Press, Cambridge.

Magee, J. F., 1964 "Decision Tress for Decision Making", *Harvard Business Review*, 42, 4, 126-138.

宮沢光一, 1970 「制御問題における1つの一般理論」『経済学論集』(東京大学) 36巻1号.
宮沢光一, 1971 『情報・決定理論序説』岩波書店.
Raiffa, H., 1968 *Decision Analysis*, Addison-Wesley, Reading.
Simon, H. A., 1955 "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics*, 69, February.
———, 1957 *Administrative Behavior*, Macmillan, New York.
———, 1977 *The New Science of Management Decisions*, Prentice Hall.
鈴木雪夫, 1978 『統計解析』, 筑摩書房, 東京.
梅沢 豊, 1970 「不確実性下における計画化と統制の過程」『経済学論集』(東京大学) 36巻2号.
Wald, A., 1947 *Sequential Analysis*, John Wiley, New York.
———, 1950 *Statistical Decision Functions*, John Wiley, New York.

コンファレンス・プログラム

日程：1988年9月5日～7日（於・箱根）
＊印は本書収録論文の執筆者

第1日／9月5日 午後（座長 楠岡成雄・京都大学数理解析研究所）
テーマ1：「主観確率の分解と応用について」
報告者：＊新家健精（福島大学経済学部）
討論者：広松 毅（東京大学教養学部）／新居玄武（学習院大学経済学部）
テーマ2：「主観分布の Elicitation」
報告者：＊松原 望（東京大学教養学部）
討論者：豊田 敬（法政大学経営学部）
特別講演：「Bayesian の一人としての回顧」
（宮沢光一 東京大学名誉教授・創価大学）

第2日／9月6日 午前（座長 早川毅・一橋大学経済学部）
テーマ3：「許容性」
報告者：＊竹村彰通（東京大学経済学部）
討論者：野上佳子（筑波大学社会工学系）／園信太郎（北海道大学経済学部）
テーマ4：「局外母数と推測」
報告者：＊竹内 啓（東京大学経済学部）
討論者：佃 良彦（東北大学経済学部）
午後（座長 中塚利直・東京都立大学経済学部）
テーマ5：「事前分布の選択と応用」
報告者：＊赤池弘次（統計数理研究所）
討論者：細谷雄三（東北大学経済学部）
テーマ6：「多重共線性へのベイズ・アプローチ」
報告者：＊美添泰人（立正大学経済学部）
討論者：加納 悟（横浜国立大学経済学部）／斯波恒正（富山大学経済学部）
テーマ7：「ベイズ予測：理論と応用」
報告者：＊和合 肇（筑波大学社会工学系）
討論者：北川源四郎（統計数理研究所）

第3日／9月7日　午前（座長 森棟公夫・京都大学経済研究所）
テーマ8：「学習過程と経済均衡」
報告者：＊国友直人（東京大学経済学部）
討論者：小林正人（京都大学経済学部）
テーマ9：「経営意思決定における情報」
報告者：＊青井倫一（慶応義塾大学大学院経営管理研究科）
討論者：大塚英作（横浜国立大学経営学部）
テーマ10：「最適化と満足化」
報告者：＊梅沢　豊（東京大学経済学部）
討論者：鈴木　武（法政大学経営学部）
テーマ11：「生存可能システムにおける動的線形モデル」
報告者：平館道子（金沢大学経済学部）／関谷　章（慶応義塾大学大学院経営管理研究科）／新家健精（福島大学経済学部）
報告者：舟岡史雄（信州大学経済学部）
総　　括：＊鈴木雪夫（東京大学名誉教授・多摩大学）

ベイズ統計学とその応用

1989年6月30日　初　版

［検印廃止］

編　者　鈴木雪夫（すずきゆきお）
国友直人（くにともなおと）©

発行所　財団法人　東京大学出版会
代 表 者　菅野卓雄
113 東京都文京区本郷 7-3-1 東大構内
電話 03（811）8814・振替東京 6-59964

印刷所　株式会社精興社
製本所　矢嶋製本株式会社

ISBN 4-13-040108-4