＜綜合報告＞

Bayesian Decision について

関 谷 章*

§1. はじめに

最近数理統計学の分野で Bayesian と呼ばれる人々の主張が注目されている。その主張の基本的な態度は何よりも先ず彼の高名な「Bayes の仮説の確率」を何らかの形で認めようとするところにある。しかし伝統的な数理統計学ではこの様な態度をはっきりと拒否することにむしろその存立の基盤があると考えられる。両者の議論の対立を理解するには先ずこのことを確認しなければならない。更に最近，伝統的な数理統計学の内部で，R.A. Fisher 及び J. Neyman-A. Wald に代表される二つの主流があり，これらが「推論」か「決定」かという問題，およびこれと直接結びついた確率の解釈の問題をめぐって鋭く対立していることが明らかになっている。Bayesian の主張を理解するためにはこの相違についても考慮する必要があろう。ところで一般に Bayesian と呼ばれる人々の中でも L.J. Savage〔6〕,〔7〕およびそれを具体的な企業における決定問題に適用しようとする，R. Schlaifer〔8〕,〔9〕および H. Raiffa〔6〕と，H. Teffreys, I. J. Good, D.V. Lindley などでは，かなりの見解の相異がある。これは差当り「仮説の確率」の解釈の差として取上げることができるであろうが，本質的には次元を異にするとはいえ，伝統的な数理統計学における「決定」と「推論」の考え方の相違をそのまま反映していると考えられる。以下では主として Savage-Schlaifer を「決定」の立場をとる Bayesian として取上げ，彼等の見解を伝統的な数理統計学にあって「決定」の立場をとる Neyman-Wald の理論との対比においてとらえることとし，その意味で表題を〝Bayesian Decision について〟としてある。従って特に R.A. Fisher の立場と Bayesian の対比については何も触れていないことを予めお断りしておきたい。以上を念頭に置いて以下では先ず§2では一般に Bayesian と呼ばれる人々の主張の背後にどの様な基本的な態度が横たわっているかを不充分な形ではあるができるだけ明確にしようと努めた。その際，通常の統計的決定問題とは切離した形で議論してある。次いで§3ではその問題の Bayesian による取扱いを伝統的な Neyman-Wald の決定理論の直接の拡張として議論し，更に簡単な例を用いて具体的に両者の差異を明確にしたつもりである。§4は上述の立場を企業決定へ適用しようと意図する Schlaifer 及び Raiffa〔4〕の見解を主として複雑な決定問題の簡略化という側面に注目してその概要をまとめたものである。最後に Schlaifer〔9〕の例を借りて彼等の意図する企業問題への適用及び問題の簡略化がどの様な形で行なわれるかを概観した。勿論これだけでは Bayesian の議論の数多くのものが取残されて

---

* 慶応大学 1964年3月20日受理 「経営科学」 第8巻第4号

いる。例えば Bayesian の基本的な態度を更に primitive な行動上の仮定から出発して示そうとする Savage の見解，あるいは具体的な主観確率，効用函数の導出の際の具体的手続きなどには触れていない。更に「推論」の立場には触れていないので通常の統計理論における標準的な手法である有意性検定，推定問題に対する Bayesian の接近については当然省かれざるを得なかった。これらを議論するためにはここで取上げられなかった「推論」の立場をとる Bayesian の性格を明確にすることが先決であろう。

尚，全体を通じて数学的には極めて粗雑な議論が多いが，ここでは基本的な考え方の概要を述べるのが目的である以上そこまで立入ることはとうていできなかったことを御断りしたい。

§2 Bayesian の基本的な立場

ここでは所謂 Bayesian の問題設定を検討するがこれは形式的には全く従来の Neyman-Wald の定式化の枠内に止ることを予め注意しておこう。

今ある決定者が特定の決定問題に直面しているとしよう。しかもこの決定問題は次の意味で二重の不確実性を伴っている。決定者はまず決定を下す際に影響を与える環境要因をとりださなくてはならないが，この場合には決定者がどの様に環境を把握するかということが問題となるであろう。以下では通常行なわれるように，決定者をとりまき，その決定に重大な影響を与える環境が一義的に規定され，しかもそれは決定の前後で不変に止まるものとしよう。以下ではこれを自然の状態と呼ぶことにする。そして決定者にとっては，それが差当り簡単のために，$\theta_1, \theta_2, \cdots\cdots, \theta_m$ のうちのいずれか一つであることはわかっているが，それが一体どれであるのかはわからないという状況を考えよう。例えば，ある地域で当会社の新製品を購入してくれる世帯の比率というものは，上述の仮定の下ではある0と1の間の一定の数値であることはわかっていても，そのうちのどれであるかはわからない。若し許されるならば，その地域の全世帯に当ってみてそれを確認することも可能であろう。しかし今は調査費の問題，時間の問題，あるいは本来，全てを調査することが無意味となるといった様々な理由から，この様なことは不可能であって，せいぜいそのうちから標本を抽出して調査を実施し，その結果に基いて決定を下さなくてはならないと考えよう。この様な場合には決定者をとりまく自然の状態には，それが可能な値がどれであるかわからないという不確定性と，それを多少とも緩和させる役割を果すべく導入された別の意味での不確実性つまり標本抽出という手段によって生ずる不確実性とが考えられる。かかる二重の不確実性が介在する状況下で合理的な決定を下すにはどうしたらよいであろうか。ここで再び簡単のために取り得る行動は a_1, $a_2, \cdots\cdots, a_n$ だけであるとし，これらは自然の状態とは無関係であるとしよう。上の例では新製品を生産してその地域に売出す，生産しない，あるいは生産するとしても，その量はいくらか，そしてそのための費用を自己資本に頼るか，借入によるか，また借入先は銀行かなど，様々な行動が考えられるであろう。今はこれらがすべてリストにのっていてそれ以外の行動はないものとしよう。（この様な行動には広告その他需要そのものを変えてし

まう様なものは含まれない。それを取り入れるには問題の設定そのものつまり自然の状態の把握の仕方から変えなければならないであろう。）

以上の自然の状態 $\theta_1, \cdots\cdots, \theta_m$ およびとり得る可能な行動 $a_1, \cdots\cdots, a_n$ に対して決定者は自然の状態が θ_i であるときに行動 a_j をとることで，ある結果 O_{ij} を得ることになる。例えばある地域で購入者の比率が殆んど0に近い（θ_i）のであれば，新製品を銀行の借入金に頼って製造しても（a_j）莫大な損害をうける（O_{ij}）ことになるかもしれない。今はこの様な行動の結果に対して決定者が感ずる満足度を表現する共通の尺度が存在することを想定しよう（最近 Fishburn : Decision and value theory, Wiley 1964 ではこれらの間に部分的な順序関係が成立つ場合の分析を行なっている)。上の場合には O_{ij} が具体的に残された借財，スクラップ化した製品の価値減耗分などが共通の貨幣単位で測定され，この様なものがそのまま決定者の感ずる満足度をある程度表現していると考えて差支えないかもしれない。一般にはこの様な満足度は決定者の感ずる効用という主観的な尺度で測定されるとしておけば，これを効用函数 $u=u\ (O_{ij})$ あるいは $u_{ij}=u(\theta_i,\ a_j) : i=1, \cdots\cdots, m,\ \ j=1, \cdots\cdots, n,$ の形で表現することができるであろう。θ, a が共に有限の場合にはこの様な函数は簡単に右図の様な利得行列（pay-off matrix）の形に表わすことができる。

$\theta \backslash a$	$a_1,\ a_2 \cdots\cdots\cdots\cdots\cdots\cdots a_n$
θ_1	$u_{11}\ u_{12} \cdots\cdots\cdots\cdots\cdots u_{1n}$
θ_2	
$\vdots$	
θ_m	$u_{m1}\ u_{m2} \cdots\cdots\cdots\cdots u_{mn}$

以下では先ず簡単のために標本抽出による不確実性の問題は捨象して，この利得行列だけの範囲で論理の筋道を辿ることにしよう。したがって差当りは，どの θ_i が自然の状態であるかがわからないという意味での不確実性だけが問題となる。伝統的な数理統計学では（特に Neyman-Wald）この種の不確実性を確率の用語で相対的な評価を行なうことを拒否している。彼等にとってはこの種の不確実性は専ら決定を下す主体の心の状態を表現しているにすぎず，自然の状態はその様なものとは何の関係もなく，いつでもそのうちのどれか一つだけが真でそれ以外のものは誤りであるということを意味するだけである。したがってこの場合には上述の不確実性はそれが存在することは認めても伝統的な数理統計学の枠外にあるものとして触れないでおくという態度を守ることになる。これは彼等が確率というとき，いつでもそれを対象に備わる客観的な実在として把えており〝人間の心の状態〟といったむしろ客観的な対象についての知識の程度によって動かされる性格のものを出来るだけ排除して「科学」としての数理統計学を意図していることの当然の帰結と考えられる。

いずれにしろ立場に固執する限り，行動 a_j は各 $\theta_1, \cdots\cdots, \theta_m$ に対応する効用を羅列したタテベクトルとして特徴ずけられ，行動相互の比較はこの様なベクトル同志の比較としてのみ意味を与えられる。この場合，もし $a_1, \cdots\cdots, a_n$ の中でどの θ_i に対する要素も他の全ての行動の対応する要素よりも大きな値をとるものがあれば，決定者はためらうことなくその様な行動を選ぶで

あろう。しかし一般にはその様な都合の良い行動は存在せず，ある θ_i に対応する要素については一方が大であっても別の θ_j に対応する要素については逆の関係が成立っていると考えられる。この様な場合にはそれらの比較は不可能となるが通常はそれ以上の選択が個人的な行動上の原理によって行なわれるものとして処理されている。そのうちのいくつかを次にあげてみよう。

イ）手放しで楽観的な人の行動原理：この様な性格の人なら自然の状態のうちで最も自分に都合の良い状態が正しいと考え，利得表を見渡して，先ず最大の値に目を向け，その値を達成する θ が現れると考えてそれを実現させる行動をとるに違いない。これは形式的には

$$\max_{i,j} u_{ij}$$

を与える a_j をえらぶこととして表現される。

ロ）最も悲観的な人のとる行動原理（Wald のマックスミニ原理)：この様な人はイ)とは対照的に各 θ について最低限どれだけを確保し得るかをしらべて，然る後にその安全水準をなるべく大きくするような a_j をとると考えるのが合理的であろう。これは形式的には

$$\max_j \min_i u_{ij}$$

を達成する a_j をとることとして記述できる。

ハ）イ)，ロ）の中間の人の行動原理（Hurwicz の原理）：以上の二つではいずれもある自然の状態について最も都合のよいもの，あるいは最も都合の悪いものの一つだけに注目していた。ここではこれら二つの要素を同時に利用する原理について考えよう。そのために Hurwicz に従って強気の指標 α（$0 \leqq \alpha \leqq 1$）を導入する。そして

$$\max_j \{\alpha \max_i u_{ij} + (1-\alpha) \min_i u_{ij}\}$$

を与える a_j をとることにすればよい。言うまでもなくイ）は上で $\alpha=1$ の場合に，ロ）は $\alpha=0$ の場合に帰着する。

ニ）負けおしみの強い人の行動原理（Savage のミニ・マックス・リグレットの原理）：この様な人々の行動を考えるには先ず効用を次のリグレット γ_{ij} で表現しなおすことが必要である；

$$\gamma_{ij}=\max_k u_{kj}-u_{ij}.$$

これは自然の状態が θ_i であることがわかったときに行動 j をとっていたためにその人が感ずる残念さを示す数字になっているであろう。この様な利得表の変換を行なって新たに得た残念さの度合いを示す表について，今度はロ）とは逆に

$$\min_j \max_i \gamma_{ij}$$

を与える a_j をとることにすればよい。

以上の4つの行動はいずれも自然の状態に対する主体の側の不確実性を直接確率で表現することはしないで，人間の行動原理として不確実性下の決定を叙述したものであり，その中には効用

函数の与える情報と主体が自然の状態の不確実性に対して抱くであろう心の状態とが不可分の形で結びついている。しかしある場合には決定者が感ずる自然の状態に対する不確実性が，行動結果の評価によっては影響を受けないことがあるであろう。そして自然の状態が決定者をとりまく環境という決定者とは切離された客体を意味するものとして把握したことを考えれば，むしろ上述の行動原理よりは決定者の抱く自然の状態の不確実性を効用とは分離して取扱うことが望ましいかもしれない。ここではかかる性格を持つものとして二つの原理をあげておくことにしよう。

ホ）ラプラスの原理：この場合は先ず自然の状態の不確実性を主体の側の無知の表現と考え，それを全ての θ について等確率を附与することで解釈しようとする。そして行動の評価をその平均的な効用で行なうことをすすめる。これは形式的には

$$\frac{1}{m}\sum_{i=1}^{m} u_{ij}$$

を最大にする a_j をとるべきことを示唆している。

しかしこの様に主体の側の不確実性をいつでも無知という決定者の如何を問わない共通の言葉で表現しなければならない必然性はないかもしれない。その場合にはもっと一般的に次の原理を考えてよいであろう。

へ）期待効用極大化の原理：自然の状態に対する主体の側の不確実性を $\xi(\theta_i) \geqq 0$，$\sum_{i=1}^{m}\xi(\theta_i)=1$ という確率の言葉で表現し

$$\sum_{i=1}^{m}\xi(\theta_i)\ u_{ij}$$

を最大にする a_j をとる。

この場合には $\xi(\theta_i)$ は決定主体の自然の状態に対する不確実性を表現する尺度であり，これは本来ある特定の決定問題に直面した決定者の内省によって与えられるべき性格のものであると考えられる。

ここまで来るともはや確率の客観的な解釈とははっきり訣別しなくてはなるまい。たとえ $\xi(\theta_i)$ をウエイトという仮りの名で呼び，あるいはもともと θ に関する把握の仕方を変えることでホ）のラプラスの原理に溯り得たとしても，そこに導入される確率は少くとも頻度的な確率とは明確に区別されねばならない。ここでの確率は決定者をとりまく客観的な環境に属するものではなく，決定者自身が環境に対して抱く主観的な判断を表明している。

この様に不確実性下における行動の選択についてはいろいろな選択基準が提案されているが，いわゆる Bayesian とは簡単にいって上述の6つの基準のうち最後の二つの基準をとることを積極的に提唱する立場であるといって差支えないであろう。日常我々が遭遇する決定問題では自然の状態に対して決定者が何らかの情報を持っていることが多いこと。例えば新製品をある地域に売出すという場合でも，その地域がもともとその種の製品の購売層として適しているかどうかについての大雑把な判断を持っているであろう。また不良品の検査という場合には不良品の比率

として50%以上にものぼるものを考えることは実際には無意味であろう。さらに50%以下でも10%以下であることの方が殆んどであると考えてよいかもしれない。現場に居て実際に検査の仕事を行なっている人々には常識となっている不良品比率に関する情報があるのが現実であろう。この様な情報を具体的な決定を下す際に十分採り入れることが出来る定式化が望ましいと考えられる。しかし期待効用を最大にする行動を選ぶことについてはそれが直観的な基準であるということ以外にさしたる理由はない。例えば，上述の確率が最大となる θ のところで a_j を比較することも考えられるが，この両者でどちらが秀れているかは一概に評価することはできないであろう。

この様に行動の選択原理としては大別して直観的な行動上の仮説により効用と不確実性とを同時に関連させて考えるイ）〜ニ）と，先ず自然の状態に対する不確実性を効用とは分離した形でとらえて，その上で何らかの合理的と思われる行動上の仮説を導入する立場との二つに分けられる。Bayesian は基本的には後者の立場をとる。先ずこのことを確認しておこう。

ところで上記の行動原理のうち少くとも始めの4つのものについては，確率をあくまで決定の主体とは無関係なものとして把握する立場とは，少くとも表面上は抵触していない様に見える。そして，これらの基準の望ましさを，別の側面から攻めることである程度判断できるかもしれない。実際，Milnor〔3〕では専らこれらの基準を形式的な側面から10ケの公理にさかのぼって，そのうちのどれだけを仮定して上述の基準が導かれるかを調べている。

これと同じことで期待効用の極大化原理についてもそれをもっと基本的な行動上の仮定にさかのぼってそれらからの帰結として把える考え方も，例えば，Savage〔6〕で取扱われている。しかし Ramsey〔5〕における様にこれ自体を行動上の仮説として取上げる方が簡明であろう。いずれにしろ，特に「決定」の立場をとる Bayesian は自然の状態に対する主体の側の不確実性を積極的に確率の形で評価し，その上で期待効用を最大にする行動を選択するという立場をとる。ここでは期待効用の最大化原理を別の行動仮説から導出する議論には触れずもっと操作的な見地からこの原理の望ましさが，Wald によって示唆されていることを以下にのべることにしよう。

先ずこの様な直観的な選択基準によってとられる行動が最小限どの様な性格を持たねばならないかを検討しよう。通常はかかる性格を次の様な行動の許容し得る集合として定式化する。今ある二つの行動 a_j, a_k を考えるとき，すべての θ_i に対して

$$u_{ij} \geqq u_{ik}$$

が成立つとき，行動 a_j は行動 a_k と少くとも同程度に望ましい (at least as good) あるいは a_j は a_k を凌駕する (dominate) と呼ばれる。更にここである i に対して実際に $u_{ij}>u_{ik}$ が成立つとき，a_j は a_k よりも良い (better than) と呼ばれる。

この様に任意の二つの行動に対する関係を定義した上で，ある行動の集合について，その集合の中のどの行動によっても凌駕されない行動をとりまとめたものを許容し得る (admissible) 行

動の集合と呼ぶことにする。明らかに上述の行動上の仮定に立った選択原理に基いてえらばれるいかなる行動も，それが最小限許容し得る行動の集合の中のものでなければならないであろう。そうでなければその様な行動は別の行動によって凌駕されてしまうことになり，別の行動をとった方がすべての θ_i に対して与えられる効用が大きいであろう。後の行動の方が直観的にも合理的な行動であって，それに凌駕される様な行動に導いた選択原理はむしろ誤ったものと考えなければならないであろう。

この様に直観的にも合理的と考えられる定義から Wald は我々のとるべき行動は最小限上述の許容される行動の集合の中から選ばなければならないと主張する。この主張は所謂 Pareto Optimality として古くから知られていたものを形式的に整えたものに他ならない。そしてこの様な考え方は伝統的な数理統計学者達によって広くうけ入れられており，この様な行動の集合を具体的な決定問題に対して導き出すまでが統計家の本来の任務であると考えられている。しかしこの仕事すらも実際には容易でないことが多い。特にとり得る行動が有限しかなくても数が多かったり，あるいは可能な行動が無限に考えられるときにはこの様な集合を直接求めることは難かしいであろう。次善の策としては，許容される行動を全て含んでいる様な集合，形式的にはある行動の集合で，それ以外の任意の行動をとれば，必ずその集合の中にそれよりも良い行動が存在する様な集合（この様な集合を完全な集合と呼ぶ）を考えてそれをなるべく完全集合という性質を保存させたままでせばめて行くことが出来れば望ましいであろう。Wald はこの様な性格を持つ集合として Bayes 解の集合を用いることを提案する。それは $\theta_1, \cdots\cdots, \theta_m$ に対して事前分布 $\xi(\theta_1), \cdots\cdots, \xi(\theta_m)$, $\xi(\theta_i) \geqq 0$，$\sum_{i=1}^{m} \xi(\theta_i)=1$ が存在すると仮定して，

$$\sum_{i=1}^{m} \xi(\theta_i)\, u_{ij}$$

を最大ならしめる a_j を $\xi(\theta_1)$, $\xi(\theta_2)$, $\cdots\cdots$, $\xi(\theta_m)$ に対する Bayes 解と呼ぶことにする。そして $\xi(\theta_i) \geqq 0$，$\sum_{i=1}^{m} \xi(\theta_i)=1$ の全てに対して Bayes 解を求めて出来る行動の集合を Bayes 解の集合と名づける。一般にはこの様な行動の集合は完全であるという性質を備えており，又 $\xi(\theta_i)>0$，$\sum \xi(\theta_i)=1$ なる $\{\xi(\theta_i)\}$ については許容される集合に含まれている。この様な Bayes 解の集合の性格からそれを許容し得る行動の集合を求めるための極めて簡明な手段として用いることができる。これが上述の期待効用最大化原理が持っている操作的な意味からする望ましい性質に他ならない。

この様に Bayesian の基本的な立場は第一に上述の意味での期待効用極大化原理をとることにより「確率」の言葉で自然に対する主体の側の不確実性を積極的に表現する点に特徴があるが，更にむしろ操作的な意味でもその立場の妥当性が確認されたことになる。この基本的な特徴から，時には Bayesian の最も独特な面として強調される重要な原則が直接導かれることを次節で取扱うことにしよう。

§3 統計的決定問題と Bayesian Decision

前節では Bayesian の基本的な考え方を述べるために「統計的」決定の問題には触れずに来た。統計的決定問題という場合には上述の決定問題を，自然の状態に関する何らかの観測を行なって得られるデータに基いて決定を下すという形に拡張したものを指している。伝統的な数理統計学ではこの場合の観測データが，どの様な観測を行なうかという実験が，定められれば，それによって特徴づけられ未知の θ に依存するある偶然機構に従って生み出されることが想定されている。そして尤度函数と呼ばれるその様な偶然機構は何らかの意味で客観的に固定されたものであって，基本的には例えば標本抽出における様に，具体的に決定者が自然に働きかけて作る偶然機構である場合を考えているが，これとは別に，自然が実験を行なったと考えられる過去のデータから比較的安定した確率法則に従うことを決定者が認めた場合であってもよい。いずれにしろこの様な偶然機構は程度の相異こそあれ，論理的に，あるいは操作的に検証が可能であるという意味で客観性を持つことが要請されている。そうでない場合でも過去の経験から当面の事象がある型の分布に従うことについての客観的な裏づけが必要とされるであろう。伝統的な統計的推論と呼ばれるものの主要な問題点はこの様な偶然機構から生み出される標本としてのデータに基いてもとの偶然機構の性格について何らかの帰納的な推論を行なうことにあった。この観点は衆知の様に Wald によって拡張され，所謂統計的決定理論として，効用函数が与えられ，観測が可能である場合に，それに基いて許容し得る決定函数の集合を求める問題として定式化されるに至った。（ただしこれに対しては R.A. Fisher の強い反対がある）

先ず上述の実験が与えられれば観測データ z の空間 Z が与えられるが，これを定義域とし行動 a の空間 A を値域とする函数 δ を定義しこれを決定函数と呼ぶことにする。可能な決定函数全体を D で表わすと，問題は D の部分集合である意味で許容される決定函数の集合を求める問題として定式化される。簡単のために Z のとり得る値が矢張り有限で s ケであるとすると，考えられる決定函数の数は s^n ケになる。§2では行動の集合について考えたことを，観測が許される場合に拡張し，s^n ケの決定函数のうちでどの決定函数が望ましい性質を持つかを考えることになる。そして§2の利得行列 u_{ij} に相当するものとして今度は

$$U(\theta_i,\ \delta_k)=\sum_{j=1}^{n} u_{ij}\ P_r\ (\delta_k(z)=a_j \mid \theta_i)$$

を定義して，これを θ_i が真であるときに決定函数 δ_k をとることによって生ずる（期待）効用と考える。ここで $P_r\ (\delta_k(z)=a_j \mid \theta_i)$ は決定函数 δ_k が与えられ，自然の状態が θ_i であるときに上述の観測を行なうときに行動 a_j をとる確率で，これは上述の偶然機構が例えば $f\ (z \mid \theta_i)$ として確定されれば，

$$\sum_{\{\delta_k(z)=a_j\}} f\ (z \mid \theta_j)$$

として計算されるものである。従って今度は§2の利得行列 (u_{ij}) の代りに上の $U(\theta_i, \delta_k)$ を要素とする $m\times s^n$ 行列が用いられる。そして考え方は全く同一であるが，δ_k 全体の集合 D についてその中の許容される決定函数の集合を求めるために上述の Bayes 解を利用することができる。即ちこの場合にも $\xi(\theta_i)\geqq 0$，$\sum_{i=1}^{m}\xi(\theta_i)=1$ なる事前分布に対して

$$\sum_{i=1}^{m}\xi(\theta_i)\,U(\theta_i, \delta_k)$$

を最大にする δ を $\{\xi(\theta_i)\}$ に対応する Bayes 決定函数と呼ぶことにする。§2と同様に $\{\xi(\theta_i)\}$ の全体に対応して Bayes 決定函数の全体が特徴づけられるがこれも論理は同一である。

ところで特定の $\{\xi(\theta_i)\}$ に対して具体的にそれに対応する Bayes 決定函数を求めるにはどの様にすればよいであろうか？　そのためには上式の内容を次の様に少しく変形してみればよい。

$$\sum_{i=1}^{m}\xi(\theta_i)\,U(\theta_i, \delta_k)$$

$$=\sum_{i=1}^{m}\xi(\theta_i)\sum_{j=1}^{n}u_{ij}\,P_r(\delta_k(z)=a_j\mid\theta_i)$$

（$U(\theta_i, \delta_k)$ の定義による書きかえ）

$$=\sum_{i=1}^{m}\xi(\theta_i)\sum_{j=1}^{n}\sum_{\{z_i\,;\,\delta_k(z)=a_j\}}f(z\mid\theta_i)\,u_{ij}$$

（$P_r(\delta_k(z)=a_j\mid\theta_i)$ の定義）

$$=\sum_{i=1}^{m}\xi(\theta_i)\sum_{z\varepsilon Z}f(z\mid\theta_i)\,u_{i\delta k(z)}$$

（a_j を固定して $\delta_k(z)=a_j$ となる z について加えてから j について加える $\leftrightarrow z$ を固定して a_j を $\delta_k(z)$ で表わしておき，それをあとで z について全て加える）

$$=\sum_{z}\left\{\sum_{i=1}^{m}\left(\frac{\xi(\theta_i)f(z\mid\theta_i)}{P_\xi(z)}\right)u_{i\delta k(z)}\right\}P_\xi(z)$$

$$=\sum_{z}\left\{\sum_{i=1}^{m}\xi(\theta_i\mid z)\,u_{i\delta k(z)}\right\}P_\xi(z)$$

ここで

$$P_\xi(z)=\sum_{i=1}^{m}\xi(\theta_i)\,f(z\mid\theta_i)$$

かつ

$$\xi(\theta_i\mid z)=\frac{\xi(\theta_i)f(z\mid\theta_i)}{P_\xi(z)}$$

は所謂 Bayes の事後確率と呼ばれるものである。ここまで変形して来ると，全体としての期待値つまり $\{\xi(\theta_i)\}$ に関する Bayes 決定函数を求めるには，上式で先ず z を固定して，各 z について $\{\quad\}$ 内の値を最大にする様な行動をとることにすればよいということは明らかであろう。ところで $\{\quad\}$ 内は丁度 $\xi(\theta_i \mid z)$ を§2における $\xi(\theta_i)$ と同じものと見做せば，形式的には，各 z に対して $\xi(\theta_i \mid z)$ に対して Bayes 解となっている様な行動を対応させる函数 δ^* がこの場合求める Bayes 決定函数であるということができる。結局，このような δ^* をとることにより，任意の D の中の決定函数 δ に対して

$$\sum \xi(\theta_i)\ U(\theta_i,\ \delta) \leqq \sum \xi(\theta_i)\ U(\theta_i,\ \delta^*)$$

が成立っている。

ところで伝統的な立場では特定の $\{\xi(\theta_i)\}$ には意味が与えられていなかった。したがって上述の変形は一つの許容し得る決定函数を探すための操作的な手続きとしてしか意味を持ち得なかった。しかし Bayesian にとっては特定の $\{\xi(\theta_i)\}$ は特定の決定問題に即して決定者自身が自然の状態に対して抱いている不確実性の表現として主観的な判断によって附与さるべきものと考えられているのだから上述の手続きは次の様な具体的な意味づけが与えられる。すなわち，ある観測が行なわれて，観測値が与られた場合には所謂 Bayes の定理によって事後確率 $\xi(\theta_i \mid z)$ を計算し，それに対して Bayes 解となっている様な行動をとればよいということである。ここで注意が必要なのは，このことは本来，決定者が観測値を得る毎度にこの様な決定を下すことにすれば，長期的には全体としての効用の主観的な期待値を極大化されることになるということを意味していることである。

この様に統計的決定問題では Bayes 解が二重の働きを示している。一つは全体としての期待効用を最大にする様な決定函数を特徴づける Bayes 解であり，他はその様な決定函数を各 z に対して事後分布 $\{\xi(\theta_i \mid z)\}$ に対して Bayes 解となっている様な行動を対応させる函数として特徴づける場合のそれである。通常 Bayesian と呼ぶときにはこの後の意味での Bayes 解をそれ自体として取上げる，あるいは単に $\{\xi(\theta_i \mid z)\}$ を計算することだけをすすめる人々を指すことがある。しかし§2で述べた Bayesian の基本的な立場から見れば，これらが単独で意味を持つと考えることは余り妥当なこととは思われない。

これまではかなり抽象的なレベルでの話ばかりだったので，次に簡単な例によって上述の議論を伝統的な立場と Bayesian との相異に主眼をおいてふりかえってみよう。今自然の状態は θ_1 と θ_2 の2つだけ，とり得る行動 a_1，a_2 の2つとする。そしてこの場合通常よく行なわれる様に機会損失が左表の様に与えられていたとしよう。但し $\omega \geqq 0$ とする。更に今は観測値のとり得る値が3つだけで左表の様にその尤度が与えられているとしよう。先ずこの場合について $Z \to A$ の函数 δ で Bayes 決定函数となるものの全体を求めてみよう。そのために今 θ_1 の事前分布を ξ_1，θ_2 のそれを $\xi_2 = 1-\xi_1$ $(0 \leqq \xi_1 \leqq 1)$ としたときに，考えられる事後確率分布に対して

	a_1	a_2
θ_1	0	1
θ_2	ω	0

	z_1	z_2	z_3
θ_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
θ_2	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

Bayes 解となっている様々な行動を対応させる決定函数を求めなくてはならない。事後分布を与える公式は,

$$\xi(\theta_i \mid z_j) = \frac{\xi(\theta_i)\, f(z_j \mid \theta_i)}{\sum_{i=1}^{2} \xi(\theta_i)\, f(z_j \mid \theta_i)}$$

であるから，計算すると次表の形にまとめられる。

	z_1	z_2	z_3
$\xi(\theta_1 \mid z_j)$	$\frac{\xi_1}{2-\xi_1}$	ξ_1	$\frac{2\xi_1}{1+\xi_1}$
$\xi(\theta_2 \mid z_j)$	$\frac{2(1-\xi_1)}{2-\xi_1}$	$1-\xi_1$	$\frac{1-\xi_1}{1+\xi_1}$

したがって z_1 が観測されたときに a_1 および a_2 をとることによって生ずる事後分布の下での機会損失の期待値はそれぞれ次の様に計算される。

$$a_1 : \frac{\xi_1}{2-\xi_1} \cdot 0 + \frac{2(1-\xi_1)\omega}{2-\xi_1} = \frac{2\omega(1-\xi_1)}{2-\xi_1}$$

$$a_2 : \frac{\xi_1}{2-\xi_1} \cdot 1 + \frac{2(1-\xi_1)}{2-\xi_1} \cdot 0 = \frac{\xi_1}{2-\xi_1}$$

よって,

$$\frac{2\omega(1-\xi_1)}{2-\xi_1} \gtreqless \frac{\xi_1}{2-\xi_1} \quad \text{即ち} \quad \xi_1 \lesseqgtr \frac{2\omega}{2\omega+1}$$

にしたがって $a_1 \overset{\succ}{\underset{\prec}{=}} a_2$ となる。ここで $a_1 \overset{\succ}{\underset{\prec}{=}} a_2$ は a_2 の方が a_1 よりも期待損失が小さいという意味で a_1 より望ましいことを表わしている。

同様にして若し z_2 が観測されたときには

$$\xi_1 \gtreqless \frac{\omega}{\omega+1} \quad \text{に従って} \quad a_1 \overset{\succ}{\underset{\prec}{=}} a_2$$

z_3 が観測されたときには

$$\xi_1 \gtreqless \frac{\omega}{2+\omega} \quad \text{に従って} \quad a_1 \overset{\succ}{\underset{\prec}{=}} a_2$$

となっている。

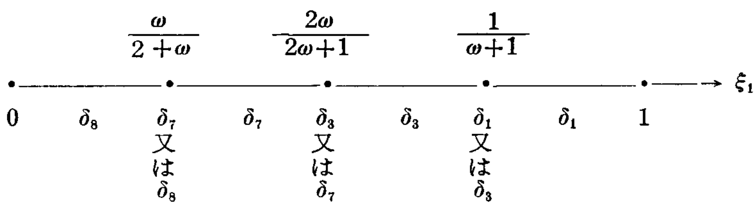
したがって，今，可能な $2^3=8$ 通りの決定函数を次の様に定めると，すなわち

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8
z_1	a_1	a_1	a_1	a_2	a_1	a_2	a_2	a_2
z_2	a_1	a_1	a_2	a_1	a_2	a_1	a_2	a_2
z_3	a_1	a_2	a_1	a_1	a_2	a_2	a_1	a_2

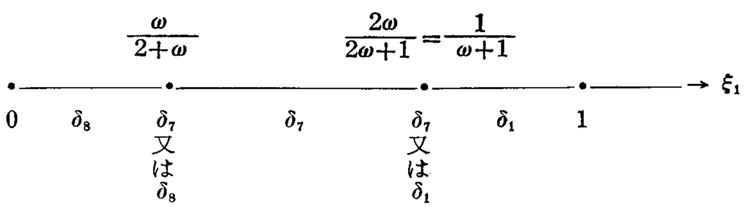
ただし例えば δ_1 は $\delta_1(z_1)=a_1$, $\delta_1(z_2)=a_1$, $\delta_1(z_3)=a_1$ を示すものとする。このとき ω の値によって先ず分類し夫々の場合に各 ξ_1 の値に対応する Bayes 決定函数を次表の様にまとめることができる。

イ) $\omega=0$ のとき, ξ_1 の如何に関らず δ_1

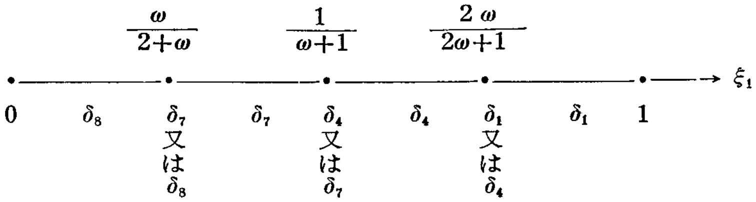
ロ) $0<\omega<\dfrac{1}{\sqrt{2}}$ のとき



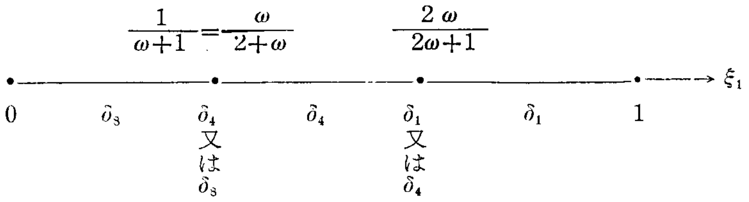
ハ) $\omega=\dfrac{1}{\sqrt{2}}$ のとき



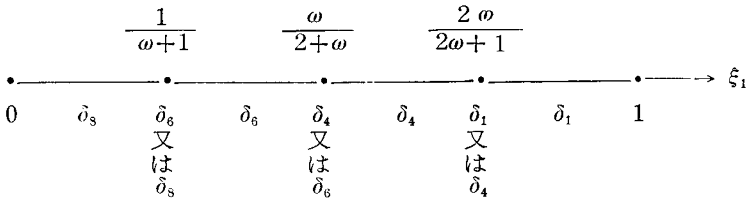
ニ) $\dfrac{1}{\sqrt{2}}<\omega<\sqrt{2}$ のとき



ホ) $\omega=\sqrt{2}$ のとき



へ) $\omega>\sqrt{2}$ のとき



この表から例えば $\omega=1$ の場合には(ニ)に現れる δ_4, δ_7, δ_8 を考えれば許容し得る決定函数を求めるには十分であるということになる。伝統的な数理統計学では先ずここまで δ を制限して，あとそのうちから実際に一つを選択する場合には §2 で述べた直観的な基準を導入して行なうのが通例である。かかる直観的な基準としてここでは先ず，効用函数を考慮しない段階での所謂 Neyman-Pearson 流の仮説検定の考え方と，Wald のマクスミニ原理（ここでは Savage のミニマックスリグレット）を取上げて，それぞれについて Bayesian の立場からどの様なことがいえるかを検討してみよう。

先ず Neyman-Pearson 流の処理には基本的には上で与えられた ω の情報が不備である場合を想定して，判定の基準として2種類の過誤に着目する。その際に決定問題に即応して θ_1, θ_2 のいずれかを仮説として取上げて非対称的な取扱いをする。今は θ_1 を仮説とする場合を想定しよう。このとき損失についての情報が不備であることは2つの行動 $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ についても，θ_1 が正しいとき望ましいと思われる行動を $\boldsymbol{a}_1$, θ_2 が正しいときに望ましい行動を $\boldsymbol{a}_2$ という莫然とした決定者の直観を表わしている，あるいは単に θ_1 が正しいときに θ_1 をとるという行動を $\boldsymbol{a}_1$, θ_2 が正しいときに θ_2 をとるという行動を $\boldsymbol{a}_2$ と考えてもよい。このとき θ_1 が正しいときに $\boldsymbol{a}_2$ をとる誤りを第1種の誤まり，θ_2 が正しいときに $\boldsymbol{a}_1$ をとる誤りを第2種の誤まりとしてある決定函数 δ の選択を，2種類の誤まりを犯す確率 $\alpha_\delta=\boldsymbol{P}_r\ (\delta(z)=\boldsymbol{a}_2 \mid \theta_1)$, $\beta_\delta=\boldsymbol{P}_r\ (\delta(z)=\boldsymbol{a}_1 \mid \theta_2)$ に基いて行なおうとする。これを今の場合に各 δ_i について計算すると次表の様に与えられる。

δ	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8
α_δ	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	1
β_δ	1	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0

ここで Neyman–Pearson は α_δ をある小さな値以下におさえておいて，その中で β_δ をなるべく小さくすることをすすめる。今かりに $\alpha_\delta \leq \frac{1}{4}$ としてみよう。この様な条件を満足する δ としては δ_1，δ_3，δ_4 がある。これらのうち β_δ が最小になっているのは δ_4 で，$\beta_{\delta_4}=\frac{1}{2}$ となっている。これが $\alpha_\delta \leq \frac{1}{4}$ のとき Neyman–Pearson のすすめる決定函数である。ところで Bayesian では δ_4 は上述のニ）ホ）およびへ）に現れている。したがって上述の δ_4 の選択は ω の如何によって様々な解釈が与えられる。特に $\omega=1$ のときには δ_4 は $\frac{1}{2}<\xi_1<\frac{2}{3}$ に対応する Bayes 決定函数となっている。このことは損失が等しいときに α_δ を比較的小さくおさえるときには，ξ_1 に始めから $\frac{1}{2}$ 以上の信頼を寄せていたことと同じであると考えることができる。何故に一方の状態を基準にとって第1種の誤りを比較的小さい恣意的な値に抑えるのか，それは損失についての情報が不足でせいぜい行ない得る最良の直観的な基準であるとしても，それを論理的に示すことは難しいであろう。それよりはむしろ損失についてのもっと信頼すべき情報を集め，更に自然の状態に対する不確実性を出来るだけ正確に把握しようと努めることが先決であろう。第一この様な基準によって重大な決定を下す人が居るだろうかという疑問が出て来るに違いない。

次に Wald の基準について調べよう。今後は損失を明確に規定して，決定函数の選択には自然の状態が θ_i である決定函数 δ を用いるときに蒙むる期待損失 $L\ (\theta_i,\ \delta)$ を利用する。これは今の場合

$$L\ (\theta_1,\ \delta_i)\ =\ P_r\ (\delta_i(z)=a_2 \mid \theta_1)\ \times 1$$

$$L\ (\theta_2,\ \delta_i)\ =\ P_r\ (\delta_i(z)=a_1 \mid \theta_2)\ \times \omega$$

となっているから，すぐ上の表で β_δ の各々に ω を乗じたものを用いればよい。ここで ω の値によってミニ・マックス解は次の様に求められる。

$$0 \leq \omega \leq \frac{1}{4}\quad \text{のとき}\quad \delta_1,\qquad \frac{1}{4} \leq \omega \leq 1\quad \text{のとき}\quad \delta_4$$

$$1 \leq \omega \leq 4\quad \text{のとき}\quad \delta_7 \qquad\qquad \omega > 4\quad \text{のとき}\quad \delta_8$$

これについても例えば $0 \leq \omega \leq \frac{1}{4}$ で δ_1 というのは，Bayesian では $\xi_1 \geq \frac{4}{5}$ と見倣すことと同等である。

この様に伝統的な数理統計学での決定函数の選択の議論はそれが許容される決定函数である限り必ず Bayesian によって ξ_1 の用語での簡明な解釈が与えられることになる。しかしこれは Bayesian の立場からは第2義的な重要さで，本来は特定の ξ_1 を決定者が指定してそれに対応する Bayes 決定函数を選ぶという点に特徴があるのである。

§4　応用のための簡略化

ここでは上述の考え方をそのままうけついで，それを企業経営における決定問題に適用することを意図した R. Schlaifer の定式化を，主として，Raiffa & Schlaifer〔4〕によって概観することにしよう。この場合の定式化の特徴は大凡そ次の3点にある。1）実験の選択を Bayesian

の立場で考慮すること，２）事前分布⟶事後分布の計算を簡略化する目的で事前分布を尤度函数とのかね合いで特定の型に制限すること，３）効用函数を実際の応用によく現れる問題を考慮して次の２つの簡約化を行なうこと，イ）効用函数を最終的実際に行動をとることによる効用と実験の効用との和で表わされる（additivity）ロ）最終効用函数は自然の状態（あるいはその単調な函数）に関して線型であること（linearity）．以下では先ずこれらを通じて問題がどの様に簡略化されるかという点に重点をおきつつ Schlaifer の定式化の特徴を検討し，最後に具体的な例を一つ取上げて実際にはいかなる手続きがふまれるかを見ることにする。

§3でも注意したように前の例の様に簡単な場合でもかなり面倒な計算が必要であった，実際の問題ではとり得る行動，自然の状態，観測値のとり得る値のいずれも無数に考えられるであろう。その場合に一々上述の手続きにしたがっていたのでは実験を一つ固定した場合でも膨大な計算量を必要とするであろう。しかし一度持定の $\{\xi(\theta_i)\}$ に注目するという Bayesian の立場をとれば，許容し得る行動の集合を規定する必要は全然なかった。その場合の最も望ましい決定函数は簡潔に，各 z に対して $\{\xi(\theta_i)\}$ から事後分布を計算し，それに対して Bayes 解となっている行動を対応させる函数として特徴づけることができた。したがってこの観点から以上の実際上現れる様々な困難を考慮して問題を改めて次の様に定式化してよいであろう。

可能な自然の状態を $\theta\in\Theta$，とり得る行動を $a\in A$，可能な実験を $e\in E$，事前分布を $p(\theta)$，尤度函数を $f(z\mid\theta, e)$，効用函数を $u(e, z, a, \theta)$ と表わせば§3の定式化を Θ，Z 共に連続量として適当な条件の下で形式的に

$$\int_{\Theta}\int_{Z} p(\theta)\ f(z\mid\theta, e)\ u(e, z, a, \theta)\,dz\,d\theta$$

$$=\int_{Z}\left\{\int_{\Theta}\frac{p(\theta)f(z\mid e, \theta)}{f(z\mid e)}\,u(e, z, a, \theta)\,d\theta\right\}dz$$

$$=\int_{Z}\left\{\int_{\Theta}p(\theta\mid z, e)\ u(e, z, a, \theta)\,d\theta\right\}f(z\mid e)\,dz$$

を極大にする決定函数を求める問題として表現することができる。ここで $f(z\mid e)=\int_{\Theta}p(\theta)f(z\mid\theta, e)\,d\theta$，$p(\theta\mid z, e)$ は z，e が与えられたときの θ の事後の密度函数を示す。ここで注意が必要なのは望ましい決定函数は e が与えられれば｛　｝内だけで操作的に定められることである。§3では実験は与えられたものと仮定されていたが，今度はその上で持定の $p(\theta)$ に基いて，決定者の立場から実験そのものを比較することができる。Schlaifer はこの点を考慮して分析を次の二つの段階に区分する。ｉ）ある実験 e を固定し，その下で与えられる z に対して事後分布 $p(\theta\mid z, e)$ を計算し，それに基いて｛　｝内を最大にする行動を選択する（この段階を最終分析—terminal analysis と呼んでいる），次に ii）ｉ）を e の下でのすべての z に対してくり返し，それと $p(z\mid e)$ から，e の下で全体としての最大の期待効用を求める。更に e を変化させて全体としての最大の期待効用を最大にする e を求める（これを事前における事後分

析pre-posterior analysis と呼ぶ)。§3ではいつでも実験は与えられたものとしてその場合に最も望ましい決定函数を選ぶことだけが問題で上記 i) と ii) の前半だけが取扱われたが，ここでは更に決定者の立場から事前における実験の評価 ii) の後半に迄問題を進めている。

以上の i)， ii) の問題を実際に簡略化するために，Schlaifer は先ず事前分布 ⟶ 事後分布の計算が簡単な形で行なえるように，予め事前分布の型を持定のものに制限しようとする。しかし取扱いの簡単さということに偏する余り，決定者が自然の状態に対して抱いている様々な不確実性を近似的にも表現し得ない様に制限してしまうのでは致し方ないであろう。この様な見地から〔4〕では事前分布としては尤度函数の核 (Kernel) に随伴する (conjugate) 分布型を採用すべきことを提唱している。ここで尤度函数 $f\ (z \mid \theta)$ の核というのは，それが

$$f\ (z \mid \theta)\ =\ k\ (y \mid \theta)\ \rho\ (z)$$

の形に因数分解できるときに $k\ (y \mid \theta,\ e)$ を指す言葉である。これは y が通常の十分統計量であることに等しい。この様な核 k があるとき，事前分布の型を核 k の中で本来パラメターの役割を果した θ を確率変数と考え，逆に y をパラメターとして考えようというのが上述の分布型の規定の仕方に他ならない。すなわちある y' をパラメターとして

$$P\ (\theta \mid y)\ =\ N\ (y)\ k\ (y' \mid \theta)$$

を事前分布族としてえらぶことを主張している。ここで

$$N(y)\ =\int_{\Theta} k\ (y \mid \theta)\ d\theta$$

とする。例えば $NID\ (\theta,\ \sigma^2)$，σ^2：既知からの n ケの標本 $(z_1, \cdots\cdots, z_n)$ の尤度は

$$\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\quad e^{-\frac{n}{2\sigma^2}\ \Sigma\ (Z_i-\theta)^2}$$

であるが，この場合十分統計量は実験を表現する n も考慮すれば周知の様に

$$y=(m,\ n)=(\frac{1}{n}\sum z_i,\ n)$$

と表わされる。従ってこの場合の核は

$$k\,(y \mid \theta) = e^{-\frac{n}{2\sigma^2}\ (m-\theta)^2}$$

となる。又

$$N\,(y) = \int_{-\infty}^{\infty} e^{-\frac{n}{2\sigma^2}\ (m-\theta)^2}\quad d\theta\ =\ (2\pi)^{\frac{1}{2}}\frac{\sigma}{\sqrt{n}}$$

で，事前分布はある m'，n' をパラメターとして

$$p\ (\theta \mid m',\ n')=\frac{\sqrt{n'}}{\sqrt{2\pi}\ \sigma}\quad e^{-\frac{n'}{2\sigma^2}\ (\theta-m')^2}$$

として正規分布をとればよいことがわかる。ところでこの場合に問題の事後分布はどの様に表現

されるであろうか？　今は

$$p\ (\theta \mid y)\ \propto p\ (\theta)\ k\ (y \mid \theta)$$

で $p\ (\theta)$ としては $k\ (y' \mid \theta)$ に比例するものをえらんだのだからこれは

$$p\ (\theta \mid y)\ \propto k\ (y' \mid \theta)\ k\ (y \mid \theta)$$

として表現することができる。そしてあるゆるい条件の下ではこの右辺を同じ函数型 k を用い，新たなパラメターを y' と y を簡単なルール＊で結合した $y''=y' * y$ とした $k\ (y'' \mid \theta)$ という形で表現することができる。勿論この場合の結合のルール＊は尤度函数が異なれば異った形のものとなる。また尤度函数が指数型の分布である場合には＊が簡単に y' と y の和を表わすこともある。今上と同じ正規分布を例にとると $y'=\ (m',\ n')$，$y=\ (m,\ n)$ とすれば，

$$p\ (\theta \mid y)\ \propto e^{-\frac{n'}{2\sigma^2}(\theta-m')^2}\cdot e^{-\frac{n}{2\sigma^2}(\theta-m)^2}$$

となっている。ここで右辺を整頓して比例定数をおとせば

$$p\ (\theta \mid y)\ \propto\ e^{-\frac{n'+n}{2\sigma^2}\left(\theta-\frac{m'n'+mn}{n'+n}\right)^2}$$

と書ける。したがってこの場合には

$$y''=\ (m'',\ m'')\ =\left(\frac{m'n'+mn}{n'+n},\quad n'+n\right)$$

となっていて前述の＊は m'' については事前分布と尤度の平均 m'，m を n'，n をウエイトして加重平均することを意味し，n'' については単にそれぞれの和をとればよいというルールになっている。

尤度函数がベルヌーイ試行の与えられる場合にも全く同様な操作で，事前分布族としてはベータ函数で

$$p\ (\theta)\ \propto\ p^{r'-1}\ (1-\mathrm{p})^{n'-r'-1}$$

をとれば $y=(r,\ n)$ を得た後での事後分布は矢張りベータ分布で $r''=r'+r$，　$n''=n'+n$ として

$$p(\theta \mid y)\ \propto\ p^{r''-1}\ (1-p)^{n''-r''-1}$$

で表わすことができる。この場合には＊はどちらの要素についても単純に加え算でよい。

〔4〕では同様な議論を行なうことで以上のプロセスをこの他，尤度函数がポアッソン分布の場合，又同じ正規分布からの標本でも分散だけが未知，平均分散共に未知の場合，あるいは多変量正規分布の場合，正規回帰論の場合についても適用して興味深い結果を与えている。またこの他にも G. Kaufman〔2〕は石油採掘問題との関連で対数正規分布の場合の分布論を展開している。

この様に事前分布を尤度函数の核に随伴する様にえらぶことで，通常よく用いられる尤度函数については事後分布の計算は非常に簡約化されることになる，またものによっては既存の数表をそのまま利用できるという点でも便利である。（ここで注意が必要なのは，この様な事前分布の

型の制限は Raiffa & Schlaifer が始めて提唱したものではないということである。彼等が取扱っているもので基本的なものは既にいろいろな人によって実際に用いられていた。彼等はそれを形式的に尤度函数の核に随伴する分布型として整理し拡張したのだと考えるのが妥当であろう)。

次に効用函数についての簡略化について上記3)の条件を検討しよう。先ず3)イ)を導入することは

$$u(e,\ z,\ a,\ \theta)=u_s(e,\ z)+u_t(a,\ \theta)$$

と表わされることを意味している。ここで $u_s(e,\ z)$ は実験の効用，実際にはマイナス費用を示し，$u_t(a,\ \theta)$ は実験を行なった結果最終的に a という行動をとることによって得られる効用を意味している。この仮定の下では上述の目的函数は

$$\int_Z u_s(e,\ z)\ f(z\mid e)\ dz+\int_Z\left\{\int_\Theta p(\theta\mid z,\ e)\ u_t(a,\theta)d\theta\right\}f(z\mid e)\ dz$$

と書ける。したがって§3での議論は丁度この第2項だけを取扱っていたことになる。この仮定の下では実験の効用はこれまでの分析と別に行なって e の評価の時にはそれを最終効用についての全体としての期待値に加えてやればよい。これは極めて簡明な実験の取扱いである。次に3)ロ)の仮定は

$$u_t(a,\ \theta)=K_a+k_a\,\omega$$

とかけることを意味している。ここで $\omega=\omega(\theta)$ は θ の単調な変換を示す。この仮定の下では上式の｛　｝内は

$$\int_\Theta (K_a+k_a\,\omega)\ p(\theta\mid z,\ e)\ d\theta$$

$$=K_a+k_a\int_\Theta \omega(\theta)\ p(\theta\mid z,\ e)\ d\theta$$

$$\overset{def}{=}K_a+k_a\,\overline{\omega}$$

となってこの値を最大にする a に簡単に $\max u_t(a,\ \overline{\omega})$ を与える a として表現される。これは具体的には効用函数の事後の期待値の比較を事後分布の下での ω の平均のところだけ比較して済ますことができることを意味しており，最終分析が極めて簡約化されていることがわかるであろう。更にこの様な $\overline{\omega}$ は z の値によって様々な値をとり，その意味で一つの確率変数と考えて，その分布を導出しておけば，事前における事後分析の際に全体としての期待値を計算する手続きが簡単なものとなるであろう。〔4〕ではこの様に $\overline{\omega}$ を z の函数として一つの確率変数と考えたときの $\overline{\omega}$ の分布を事前における事後平均の分布 (prior distribution of posterior mean) と名づけて，その一般的な性質を議論してから，尤度函数の核に随伴する事前分布を用いる場合にそれぞれの分布を導出してある。例えば先程の NID $(\theta,\ \sigma^2)$, σ^2：既知からの n ケの標本の場合には $\omega=\theta$ なら $\overline{\omega}$ は矢張り正規分布でその平均値は事前分布の平均に等しく，分散は事前分

布の分散から事後分布の分散を差引いたものに等しくなっている。いずれにしろこの様な事実はあくまでも問題の整理簡略化という専ら実用主義的な観点から取扱われていることは注意を要する。

以上の簡約化の他に〔4〕では効用を§2で述べた機会損失の用語に飜訳して分析を進めている。§3で触れたようにその場合も論理は同一で期待効用の最大化問題を期待損失の最小化問題に変えるだけでよい。只，期待損失の計算が，積分範囲が限定されることから多少複雑になることに注意しなければならない。〔4〕ではこの点についても上述のいくつかの簡略化の仮定と併せて詳しい分析を行なっている。

以上で極く大雑把ではあるが Schlaifer による実際問題への適用という観点からする問題の簡略化の基本的なものを概観したことになる。これらを更に様々な問題に拡張することは措いて，最後に尤もらしい例を一つ上げて実際にはどの様な手続きで分析が行なわれるかを見ておくことにする。

今簡単のために固定した読者層（50,000人）を持つある専門書の出版会社で，予約制で，新たな書籍を出版するかどうかを決定したいとする。若し，出版する（a_1 で表わす）ことになれば，自然の状態を全読者層のうちでその書籍を購入する人の比率 p で表わすと，p の大きさに依存して，ある場合には利益を，又ある場合には損失をもたらすであろう。他方出版しない（a_2で表わす）ことにすれば p が何であっても現状が維持され，新たな利益も得ない代りに損失も全然うけないで済むであろう。話をもっと具体的にするために，今出版すると決った場合に考えられる経費として，固定費（版権の取得，初版の組版代など）が80万円，変動費（紙代，インク代など）が1冊当り200円と仮定しよう，更に今の書籍はある叢書の1冊として計画されているので定価は1冊当り600円でこれは動かせないものと考えよう。このとき自然の状態が p で a_1 という行動をとった場合の効用を収益で評価することは合理的であろう。これを $u(a_1, p)$ で表わすと，

$$u(a_1, p) = \text{売上げ} - \text{経費} = 50,000p \times 0.06 - (80 + 50,000p \times 0.02)\ (\text{万円})$$
$$= 2000p - 80$$

となる。他方 a_2 についても形式的に全ての p について，

$$u(a_2, p) = 0$$

と書けるであろう。先ずこの問題について損益分岐点を求めてみると，$p=0.04$ となる。これを目途にして考えると，$p>0.04$ なら a_1 という行動をとることにより，何程かの収益をあげることができるが，その場合逆に a_2 という行動をとれば，みすみす a_1 で得られたであろう利益を見逃したことになる。そこでこの問題を次の様な機会損失によって定式化するのが合理的であるかもしれない。そしてこの方が実際の企業家の行動基準により合致しているであろう。今は期待損失を ℓ で表わすと，

$$\ell(a_1, p) = \begin{cases} u(a_2, p) - u(a_1, p) = 80 - 2,000p & p \leq 0.04 \\ u(a_1, p) - u(a_1, p) = 0 & p > 0.04 \end{cases}$$

$$\ell\ (a_2,\ p)\ =\ \begin{cases} u(a_2,\ p)-u(a_2,\ p)=0 & p\leq 0.04 \\ u(a_1,\ p)-u(a_2,\ p)=2,000p-80 & p>0.04 \end{cases}$$

となっている。この様な状況下で a_1, a_2 のいずれを選択するかを決定するのに, Bayesian は先ず, p に関する情報を主観的確率で表現することをすすめる。そしてその様な確率の附与は, 具体的には自社の過去の経験から同種の書籍についての予約申し込みの状態から判断してもよいであろうし, あるいは全く人工的な賭を行なうことでそれを抽出してもよいであろう。今これが差当り次表の数値で与えられたものとしよう。

p	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
$p_r(p)$	0.04	0.08	0.15	0.22	0.23	0.17	0.09	0.02

勿論これをそのまま利用して直接それぞれの行動の期待損失を

$$\sum p_r(p)\ \ell(a_1,\ p)=8.6$$

$$\sum p_r(p)\ \ell(a_2,\ p)=18.4$$

と計算して, この結果から a_1 の方即ち出版するのが望ましいと結論することもできるであろう。

ところで今, この書籍の発行について50,000人のうちからランダムに50人を選んで調査したところ, そのうち予約申込みが2人あることがわかったとしよう。上の問題をこの様な調査結果に基いて解くとしたらどの様に考えたらよいであろうか。それには先ず公式

$$p_r(p\mid 2)=\frac{p_r(p)\ {}_{50}C_2\,p^2(1-p)^{50-2}}{\sum_{p=0.01}^{0.08} p_r(p)\ {}_{50}C_2\,p^2(1-p)^{50-2}}$$

から事後確率を計算すると

p	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
$p_r(p\mid 2)$	0.013	0.063	0.163	0.259	0.255	0.164	0.071	0.012

となる。これから a_1 をとるときの事後の期待損失を計算すると

$$\sum_{p=0.01}^{0.04}(80-2000p)\ p_r(p\mid 2)+\sum_{p=0.05}^{0.08}0\times p_r(p\mid 2)=6.56$$

となる。同様に a_2 をとるときの事後の期待損失は

$$\sum_{p=0.01}^{0.04}0\times p_r(p\mid 2)+\sum_{p=0.05}^{0.08}(2000p-80)\ p_r(p\mid 2)=16.88$$

である。したがって $z=2$ のときは a_1 の方が事後の期待損失が小さいから, §3の教えるところにより, a_1 をとることが望ましい。同様に $z=3$ のときの期待損失は a_1 については 3.38, a_2 に

ついては28.90となってやはりa_1が望ましい。そしてこれより大きなzについてはいずれもa_1が望ましいことが確かめられる。逆に$z=0$のときにはa_1をとることによる期待損失は22.92, a_2のそれは5.82となって出版しないa_2という行動の方が望ましく，$z=1$のときにもa_1に対して12.54, a_2に対して11.00となって矢張りa_2が望ましい行動になっている。以上から，結局全体としての期待損失を最小にする決定函数は$z=0, 1$に対してはa_2を，又$z=2, 3, \cdots\cdots, 50$に対してはa_1を対応させる函数として与えられることがわかった。

ところで，§3で形式的に示されている様に，これが実際に任意の決定函数の中で全体としての期待損失を最小にしていることをチェックしておくことは興味深いかもしれない。ここでは許容し得る決定函数として

$$\delta_i(z)=\begin{cases} a_2 & z=0,\ 1,\ 2, \cdots\cdots,\ i \text{ のとき,} \\ a_1 & \text{その他で} \end{cases}$$

の形のものを考えよう。これらについて，ていねいに，

$$L_{\delta_i}=\sum_{p} p_r(p) \sum_{i=1}^{2} p\left(\delta_i(z)=a_j \mid p\right) \ell\left(a_j,\ p\right)$$

を計算してみると，始めの4つのものについては結果が次表の様に与えられる。

δ_i	δ_0	δ_1	δ_2	δ_3
L_{δ_i}	0.16	5.79	6.54	11.71

残りのものについてはδ_3の値よりも大きくなっていることは容易に確かめられるから上で求めた決定函数が実際に最も望ましいものであることがわかる。また当然の事乍ら前の方法で得られた全体としての最小の期待損失を計算してみると5.79となって一致している。

ところでこの様にして求めた全体としての期待損失は，実験を行なわないで決定を下したときの期待損失の値よりも確かに小さくなっている。したがってこの差から実験の費用を差引いたものが上の実験を行なうことによって与えられる利益の増分と考えることができる。しかしこれはあくまで決定者が事前に評価を行なう実験の評価であることを注意する必要がある。更に上の手続きを例えばnをいろいろの値に変えてそれぞれについて実験が与える利益の増分を計算することで，それが最も大きな実験を選ぶことができる。

ところで持定の実験が与えられた場合の terminal analysis でも上の手続きをふむことは一般に容易ではない。今の様に事前分布を簡単な形で与えても実際の計算はかなり面倒である。更にnの決定ともなれば，pのとり得る値が多くなれば計算量は膨大なものになるであろう。この点を考慮して事前分布が尤度函数とのかね合いで制限されることになる。上例では事前分布をベータ分布から選んで

$$p_r(p)=\frac{1}{B(6,\ 120-6)}\,p^{6-1}\ (1-p)^{(120-6)-1}$$

とすると，かなりよく前の分布を近似しているであろう。この近似が許されれば事後分布は一々計算しなくても，矢張りベータ分布で

$$p_r(p\mid z)=\frac{1}{B(6+z,\ 120-6+z)}\,p^{6+z-1}\ (1-p)^{(120-6+z)-1}$$

として与えられる。従って上述の範囲の計算にはすべて既存の数表を利用することができる（例えば期待損失の計算には不完全ベータ函数表が用いられる）ので実験の比較の困難も大巾に減少することになる。

§5 結　び

これまで述べて来たことをふりかえってみよう。所謂 Bayesian にとっての共通の地盤は自然の状態に対して決定者が抱く不確実性を事前確率として表現し，それを積極的に決定問題に導入するという点にある。その場合の確率の性格は伝統的な数理統計学で支配的であった頻度概念とは明確に区別されなければならない。頻度説そのものは，決定者がとりまかれている客観的対象そのものに，あるいは sampling という客観的操作に固有の実体であるという意味で，決定者という主体が対象を把握する際に多かれ少なかれ不可避的に介入せざるを得ない様な不確実性とは，はっきりと異っていると考えられる。従ってこれまで述べて来た「全体としての期待値」を極大にするということは，あくまで特定の決定者が自己の責任において極大にすべき性質のものであって，それはいかなる意味においても「真の」目的と考えることはできない量である。したがってそれを目的として求められる最適な決定函数，あるいはそれに基づく具体的な行動は，いずれも決定者自身の責任においてとられるべき性質のものであって，その様なものとしてのみ意味のある行動と考えなくてはならない。勿論同じ問題であっても決定者が異れば，それぞれが自然の状態に対して抱く不確実性の程度も異り，恐らくその結果異なった行動を最適なものとしてとるであろう。この様な性格を望ましくないと考える人々は上述の伝統的な数理統計学を信奉する人々以外にも数多くみられる。R.A. Fisher はそもそも「決定」を拒否する立場から当然としても，Bayesian と普通に呼ばれている人々の中にも，例えば H. Jeffreys, D. V. Lindley, I.J. Good といった人々の様に「合理的な事前分布」を求めている人々が居る。これらの人々はいずれも異なった決定者が異る事前分布を持ち得るという見解を排除し，自然の状態に対する不確実性を主体の側の「客観的な無知」として把握しようと試みている。そしてこれらの人々にとって特徴的なことは，いずれも多かれ少なかれ具体的な決定の問題とは切離した形で「合理的な事前分布を求めようとしている態度である。これは丁度，伝統的な数理統計学にあって「決定」の立場と，「推論」の立場の間に鋭い対立があることと軌を一にするものであり，それがそのまま Bayesian の中に持込まれていると考えてよいであろう。その意味でこれらの人々を「推論」

の立場をとる Bayesian と呼んでこれまで述べて来た「決定」の立場をとる Bayesian と区別するのが妥当であろう。実際これらの人々は Savage, Schlaifer 等からは Pseudo-Bayesian と呼ばれて区別されている。しかし最近 Savage はこれらの人々の見解も「決定」の立場をとる Bayesian の特別な場合として解釈し得ることを主張している。

この様な事前分布についての解釈の相異はあるにしても，現実の決定問題はいつでも決定者に即断を求めている。その場合に不確実な自然の状態に対して彼が抱いている直観をできるだけ明確な形で表面に出し，それに基いて決定を行なうのは大いに意味があることであろう。しかも次第に複雑化の様相を呈している決定問題をできる限り簡略化し，見通しをよくすることの有用性は量り知ることのできないものがあろう。期待効用の極大化原理は単に直観的な意味を持つに止らず，操作的にも極めて重要な意味を持っていると考えられる。観測データは単に Bayes の定理に基く事後分布としてまとめておけばよいのであり，各段階ではそれをあたかも現在得られる自然の状態に対する最良の guess であるかの様に行動すればよいのである（§3）。しかも実際によく見られる効用函数の性質についての仮定と相俟って特定の観測についてはその様な行動自体も大巾に簡略化される可能性が認められている（§4）。勿論 Bayesian の基本的な態度に対しては様々の議論があり，これらは今後も恐らく続けられるであろうし又続けられなくてはならない問題である。しかし乍ら，この問題を別としても，現代の次第に複雑化しつつある決定問題に対して，問題の大巾な簡略化，とるべき行動を定める簡明なルールを与える Bayesian の良さを改めて見直す必要があるのではなかろうか？

最後にこの草稿に幾度も目を通され数々の有益な助言を与えて下さった東大の竹内啓助教授と，同大学院の新家健精，佐伯道子を始めとする諸兄姉，また，様々な専門分野から持に企業における実際問題の難かしさを教えて下さっている慶大ビジネス・スクールの諸先生と同僚の方々に深く感謝したい，しかもそれらを十分に生かすことができなかった私の非力を深くお詫びしたい。

参考文献（上で直接引用したものだけに止める。Bayesian に関連する様々な文献については下記〔6〕および〔7〕の巻末の文献目録を参照されたい）

〔1〕 H. Jeffreys: Theory of Probability, 3rd ed. Oxford Univ. Press, 1961.

〔2〕 G.M. Kaufman: Statistical Decision and Related Techniques in Oil and Gas Exploration, Prentice-Hall, 1963.

〔3〕 J. Milnor: "Games against Nature" in R.M. Thrall (ed.) Decision process 49～59. John Wiley, 1954.（なお同論文は M. Shubik (ed.) Game Theory and Related Approaches to Social Behavior, John Wiley 1964 にも再録されている）

〔4〕 H. Raiffa & R. Schlaifer : Applied Statistical Decision Theory, Division of Research, Harvard Business School, 1961.

〔5〕 F. Ramsey: "Truth and Probability" (1926) in The Foundations of Mathematics and Other Logical Essays, Kegan Paul, 1931, (この論文は H.E. Kybury(ed.) Studies in Subjective Probability, John Wiley, 1964 に集録されている)

〔6〕 L.J. Savage: Foundations of Statistics, John Wiley, 1954.

〔7〕 L.J. Savage and Others : The Foundations of Statistical Inference, Methuen, 1962.

〔8〕 R. Schlaifer: Probability and Statistics for Business Decisions, McGraw-Hill, 1959.

〔9〕 R. Schlaifer: Introduction to Statistics for Business Decisions, McGraw-Hill, 1961.