経営科学(日本オペレーションズ・リサーチ学会邦文機関誌) 第16巻 第6号 (1972年11月)

<招待発表>

統計的予測の問題†

——OR との関連の観点から——

竹 内 啓*

1. はじめに

統計的予測 statistical prediction とは，次のような問題を意味するものと考えよう．すなわち，あるデータ $X=(X_1, \cdots\cdots, X_n)$ にもとづいて，ある量 Y を予測する．ここで X と Y はある同時確率分布に従うが，その分布は少なくとも一部未知の要素をふくむものと考える．未知の要素を母数 θ (一般にはベクトル) で表わして，X, Y の同時確率分布を $P_\theta(X, Y)$ と表わすことにする．このとき X にもとづくどのような予測方式をえらんだらよいかを考えるのが，統計的予測理論的課題である．

ところで，Wiener 以後の確率過程論にもとづく予測理論は，同時分布が本質的に既知の場合を扱ってきた．他方，未知の母数をふくむ場合は，R. A. Fisher 以来統計的推測 statistical inference という形で，母数 θ の推定，あるいは検定という形で問題を考えてきた．そこで，上記のような意味で統計的予測が問題になる場合でも，まずデータ X から θ を推定し，次に θ を推定量 $\hat{\theta}$ でおきかえて得られる $P_{\hat{\theta}}(X, Y)$ を真の同時分布であるかのように見なして，Y の予測方式を求めるのがふつうであった．

しかし，このような考え方には，いろいろ欠陥がある．まず θ を推定する場合の推定量の基準は，予測の問題を考慮してつくられたものではない．また θ が既知のときによい予測方式が，θ の推定量 $\hat{\theta}$ が誤差をふくむ場合にも，なお適切なものであるかどうかは疑問である．すなわち，推定-予測という2段の方式において，それぞれにおいては適切な予測，あるいは推定の方式がえらばれたとしても，それを結びつけたものが，はたして全体としてよいものになっているかどうかはわからない．

さらにこのような方式には，論理的にも難点がある．すなわち，もし X と Y が独立でないなら

---

† 1972年7月20日受理．1972年5月28日，春季研究発表会招待発表．

* 東京大学経済学部．

ば，Yの予測はXに関する条件付分布を用いて行なわなければならない．ところが，そこで条件付分布が未知母数θをふくんでいるならば，それもXから推定される．そうすると同じデータXを，一方ではYの分布の条件付分布を定めるために，他方ではθを推定するために，二重に使われることになる．これは少なくとも論理的にすっきりしない点をふくんでいる．

そこで，このような2段の考え方によらず，XからYへの直接の予測として，問題を定式化する必要があるように思われる．しかしながら，このような問題を扱う一般理論は，まだ十分体系化されていないし，数理統計学の教科書においても，ほとんどふれられていない．私は数年前，そのような体系化を試みた．そうして多くの点で，母数に関する推測の場合と平行的な理論をつくることができることを示すことができた．すなわち，母数の推定，区間推定，検定に関すると同様な理論を，確率変数Yに関しても構成することができるのである．

しかしながら，他面からいえば，予測には母数に関する推測とは違った意味がある．というのは，データXは一般に一定の条件の下で現実に観測されたものであり，Yは未来において実際に観測されるべきものである．したがって，それらは何らかの意味で客観的な対象と直接結びついている．これに対して，X, Y に対して想定された確率分布 P_θ は，Fisher の言葉をかりれば，仮説的母集団 hypothetical population であり，観念的に構成されたモデルであるにすぎない．したがってそのモデルを，あるいは母集団を特徴づけるものとしての母数自体も，それだけではいわば観念的な存在にとどまっている．したがって，そのようないわば理念的な存在たる母数に対する推測と，具体的に観測されるべき量に対する予測とでは，抽象のレベルにおいて異なったところがあるといわねばならない．

このような意味で，予測は具体的な決定 decision という性格が強い．すべての推測の問題も，抽象的な形では決定問題 decision problem の一種と考えることはもちろん可能であるが，しかし抽象的に考えられた母数についての推測の場合には，推測の誤り，それから生ずる損失などを具体的に把握することはむずかしい．これに対して予測の場合には，他の種類の具体的な決定問題の場合と同様に，予測の結果，すなわち，その誤りとその大きさは明確であり，またそれから生ずるいろいろな意味の損失もはっきりしていることが多い．

さらにふつうには，母数に関する推測という形で問題が定式化されている場合でも，実際にはここにのべたような意味での予測が問題である場合が少なくないことに注意しよう．たとえば何種類かの農作物の収量を比較する場合，その母集団分布が比較されるが，そこで実際にわれわれが関心をもっているのは，母集団分布の特性値としての平均などではなく，それらの農作物と実際に畑で栽培したときの収量（あるいはその期待値）なのである．

また現実に与えられるデータについて，考えられるモデルは実は一通りでない場合が多い．このとき二つのモデル，あるいはそれを表現した二つの確率分布 $P_\theta\{X, Y\}$，$P_{\theta'}'\{X, Y\}$ を比較することは，推測という観点からは困難である．これに対して，同じYに対する予測方式を検討することによって，二つのモデルのいわば“機能”を比較することが可能になり，また二つのモデ

ルが予測という観点からはまったく，あるいはほとんど同等であるような場合も明らかになる．

2. 予測概念の拡張

このような統計的予測の問題が，OR にも関係が深いものであることは明らかであろう．いやむしろ数理統計的方法が直接 OR に有効であるのは，ほとんどすべて予測に関係した場合であるといってもよいであろう．

しかし他方，OR の観点からすれば，上記のように定式化された予測問題は，なお不十分あるいは中途半端なものであるといわれるかもしれない．というのは，予測される量 Y の分布は未知の母数 θ にのみ依存し，“統計家”のとる行動には依存しないものとされている．しかし予測は，当然に未来に向かっての行動をふくむ．そうしてその結果は，当然 Y の分布にも影響を与えることになるであろう．すなわち，X が与えられたときの Y の条件付分布は，未知の母数 θ とともに，とられる行動 action Z にも依存し $P_\theta\{Y|X, Z\}$ と表わされるものと考えるほうが，より一般的であり，また実際的であろう．このような場合にわれわれの目的は，Y を望ましい水準に近づけるために，適切な Z の水準をデータ X にもとづいて定めること，あるいは二つの Z_1, Z_2 に対応する Y の値の分布を比較して，どちらが望ましいかを判定することなどであろう．私は最近，予測の理論をこのような方向へ一般化することを試みた．

もちろんもっと一般的に考えれば，データの観測・予測・行動決定という過程をすべて一体化して，情報を得つつ行動していくという形を考えることもできるであろう．このような形で考えれば，それはもはや予測というよりも，最も広い意味での決定理論 decision theory の問題と考えねばならない．そうしてそれは，いわゆる制御 control の問題をもふくむことになるだろう．

しかし，理論の定式化を，無限定に一般化するなら，その内容の空白化を招くだけである．統計的予測の理論を，予測の理論という形で定式化することの効用は，それによって実質的な内容をふくむ理論がつくられ，現実に有効な結論が得られるという点にある．そのなかには，OR に実際に役立つようなものも少なくないと私は思う．

以下若干の論点をえらび出して，このことをいわば例示的に説明しよう．

3. 予測の数学的定式化

最初に統計的予測の形式的な定式化を考えよう．いま可能な予測の集合を D とし，X にもとづいて D の中の要素をえらぶことが“予測”であると考える．D の要素 d は Y の値と関連して，その“正しさ”あるいは“誤り”の程度が決定される．したがって，その誤りの程度を表わす損失 loss を d と Y の値 y との2変数実数値関数 $W(y, d)$ で与えることができると考える．ここに $W \geqq 0$ と仮定する．予測方式は，X に D の要素を対応させる写像であると考えることができる．それを予測関数と呼ぶ．予測関数 π が与えられれば，母数が θ に等しいときの期待損失を

$$(3.1)\quad E_\theta\{W(Y, \pi(X))\} = \int W(y, \pi(x)) P_\theta(dx, dy)$$

で与えることができる．これをリスク risk と呼び，$r(\theta,\pi)$ と表わすことにする．そうすると，$r(\theta,\pi)$ をすべての θ に対してなるべく小さくするように π をえらぶことが，統計的予測の問題であると考えられる．

このような定式化は，Wald の統計的決定関数 statistical decision function の定式化とほとんど同じである．実際 Y と X とが独立ならば

$$W^*(\theta,d)=E_\theta\{W(Y,d)\}$$

とおくと，

$$r(\theta,\pi)=E_\theta\{W^*(\theta,\pi(X))\}$$

と表わされるから，これは W^* を損失あるいは加重 weight として定義した決定関数の定式化とまったく一致する．したがって Y と X とが独立ならば，このように定義した予測の理論は，一般的抽象的に考える限り，決定関数の理論とまったく異なるところはないであろう．

しかし Y と X とが独立でないならば，抽象的な理論のレベルにおいても，異なる面が生ずる．

その一つの例として，十分統計量 sufficient statistic の概念を考えよう．いま $T=t(X)$ が Y の予測に関して十分であるということは，X にもとづく任意の予測方式 π に対して，それと同等な T のみの関数として表わされる π^* が存在することである．すなわち，すべての θ に対して

$$r(\theta,\pi^*)=r(\theta,\pi) \tag{3.2}$$

となると定義しよう．ここで理論上の便宜のためにランダム予測方式をもふくめて考えると，任意の D, W に対して (3.2) がつねになり立つための必要十分条件は，$T=t$ が与えられたときの Y の条件付分布が X とは独立であり，また X の条件付分布は θ と独立であることが示される．同時密度関数 $f(x,y,\theta)$ が存在する場合には，このことは，

$$f(x,y,\theta)=h(x)g(t(x),y,\theta) \tag{3.3}$$

と表わされることに等しい．これをいいかえると，$T=t(X)$ はふつうの意味で十分統計量であり，またさらに，Y の条件付分布は T の値にのみ依存するということを意味する．すなわち T は θ に関する情報をすべてふくみ，また Y と X との相互関連をもすべて表現しているということになる．このとき，T が Y の予測に関して十分 Y-prediction sufficient であるということにする．

より具体的な定式化においては，予測方式を決定あるいは推測の方式と区別して定式化することが，X と Y とが独立である場合にも必要になる．

ところで推測，あるいは決定の場合と同じく，予測においてもベイズ式の接近法 Bayesian approach と，非ベイズ式の接近法の二つが考えられる．前者についてはさらに二つの方法が考えられる．一つはリスクの事前分布系についての平均

$$r(\xi,\pi)=\int r(\theta,\pi)d\xi_\theta$$

を最小にするような π をえらぶことであり，もう一つは，X が与えられたとき Y の条件付分布を事前分布 ξ と同時分布 $P_\theta(X,Y)$ とから計算して，その条件付分布に関して

$$W(y,d)P_\xi(dy|X=x)$$

を最小にするように $\pi(x)=d$ を定めることである．この二つは，考え方としては異なる面をふ

くんでいるが，結果的には一致する．

ベイズ式の接近法によれば，結局 X, Y の同時分布が与えられた場合と同じことになるから，問題は容易になる．しかし，そこには事前分布の選択をどのようにすべきかの問題がつねにつきまとうので，私としては，すべての場合にそれで十分であるとは思われない．理論的興味も，実際的重要性も，非ベイズ的，あるいは Neyman-Pearson 流の接近法のほうが大きいように思われる．以下においては，もっぱら後者についてのべる．

4. 点 予 測

予測方式のなかで最も簡単なものは，点予測 point prediction と呼ばれる．これはYが実数の場合，XからYの値そのものを予測しようとする場合である．$\gamma(X)$ を予測量 predictor とするとき，

$$E_\theta(Y-\gamma(X))=0 \quad \forall\theta$$

ならば，γは不偏 unbiased であると呼ばれる．不偏な予測量のなかで，予測誤差の分散

$$E_\theta(Y-\gamma(X))^2$$

を最小にするものが最もよい予測量であると見なされる．

最小分散不偏予測量を求める問題は，最小分散不偏推定量を求める問題と非常によく似た性質をもっている．実際XとYとが独立ならば，

$$E_\theta(Y)=g(\theta)$$

とおくとき

$$E_\theta(Y-\gamma(X))^2=V_\theta(Y)+E(\gamma(X)-g(\theta))^2$$

となるから，問題は $g(\theta)$ の不偏推定に帰着する．

しかし，具体的な問題に即して考えると，予測方式を予測誤差の２乗の期待値で評価することには疑問があるかもしれない．たとえば需要量の予測においては，過大な測定から生ずる損失と，同じ大きさだけ過小な予測から生ずる損失とは同じではないであろう．このような場合には，不偏な予測量は必ずしもよいものではないであろう．

5. 区 間 予 測

予測の第２の形式としては，区間予測 interval prediction がある．これはXから $l(X), u(X)$ を計算して，Yが $l(X)$ と $u(X)$ の間にあるという形で予測を行なうものである．このとき

$$P_\theta\{l(X)\leqq Y\leqq u(X)\}\geqq 1-\alpha \qquad \forall\theta$$

となるならば，$(l(X), u(X))$は信頼係数 $1-\alpha$ の予測区間 prediction interval であるという．もっと一般的には，任意のYに対して，Xに対応してYの値域の部分集合 $S(X)$ を対応させ

$$P_\theta\{Y\in S(X)\}\geqq 1-\alpha \qquad \forall\theta$$

となるようにすることが考えられる．このような S を，信頼係数 $1-\alpha$ の予測域 prediction region であるという．

区間予測の理論は，あまりくわしく知られていないが，応用上の重要性は大きいといわねばならない．またその理論もある程度までは統一的に展開することが可能である．

また場合によっては，片側信頼区間，すなわち

(5.1)　$P_\theta\{Y \leqq u(X)\} \geqq 1-\alpha$

となるような u を求めることが必要である．

たとえば，Y を今後 t という長さの期間に生ずるであろう故障部品の数，X を s の長さの期間に発生した故障部品の数とするとき，X にもとづいて Y の数の限界をおさえておくということは，実際にも重要な意味をもつであろう．そこで (3.1) をみたすような，$u(X)$ を求める．いま故障発生率が一定であるとすると，X, Y は互いに独立にそれぞれ $s\lambda, t\lambda$ を母数とするポアソン分布に従うと考えられる．ここに λ は未知母数である．このとき λ の値に無関係に (5.1) をなり立たせるような方式が次のようにして求められる．

いま $X+Y=T$ が与えられたものとすると，Y の条件付分布は2項分布 $B(T,p)$．ただし，$p=t/(s+t)$ となり，λ とは無関係になる．したがって，個々の $T=t$ の値に対して

$$P\{Y \leqq c(t) \mid T=t\} \geqq 1-\alpha$$

となるような $c(t)$ は，2 項分布表から定められる．ところで，すべての T についてこのように $c(T)$ を定めておけば，すべての λ に対して

$$P_\lambda\{Y \leqq c(T)\} \geqq 1-\alpha$$

となる．ところで，$T=X+Y$ であるから

$$P_\lambda\{Y \leqq c(X+Y)\} \geqq 1-\alpha$$

となり，{　} 内を Y に関して解けば，

$$P_\lambda\{Y \leqq u(X)\} \geqq 1-\alpha$$

という形になることがわかる．

6.　二者択一予測

二者択一予測 dichotomous prediction というのはあまりよいことばではないが，ある Y に関する事象が起こるか起こらないか，すなわち，ある A に対して $Y \in A$ となるか否かを判定しようという場合をさすものとしよう．これは推測の場合の仮説検定に対応するものであり，現実にもこのような形での予測が求められる場合は少なくないと思われるが，その理論はまだほとんどできあがっていない．いま $\phi(X)$ を，X の値に対応して $Y \in A$ と予測するならば 0, $Y \notin A$ とするならば1という値をとる関数とし，また $\chi_A(y)$ を A の特性関数（すなわち $y \in A$ のとき $\chi_A(y)=1$, $y \notin A$ のとき $\chi_A(y)=0$）とすると，2 種類の誤差の確率がそれぞれ

$$\alpha_\theta = E_\theta\{\phi(X)\chi_A(Y)\}$$

$$\beta_\theta = E_\theta\{(1-\phi(X))(1-\chi_A(Y))\}$$

が与えられる．このとき $\alpha_\theta, \beta_\theta$ をともになるべく小さくする，あるいは α_θ を一定の小さい値以下におさえた上で β_θ も小さくする，というような基準をとることが考えられる．

この問題をもう少し拡張すると，Yの値域をいくつかの部分空間 $A_1, \cdots\cdots, A_k$ に分割して，Yがそのなかのどこにふくまれるであろうかと予測する問題が考えられる．この問題は多重決定問題の類推で，多重予測 multiple prediction の問題と呼ぶことにしよう．いまXが観測されるとき，$Y\in A_i$ と予測する確率を $\phi_i(X), i=1, \cdots\cdots, k$ とする（確率的予測を考えないとすれば，ϕ_i のうちいずれか一つだけが 1，他は 0 とすればよい)．そうすると，現実に $Y\in A_j$ となるのに $Y\in A_i$ と予測する確率が

$$\pi_{ij}=E_\theta\{\phi_i(X)\chi_{A_i}(Y)\}$$

で定義される．そこで，π_{ii} をなるべく大きくするように ϕ_i を定めることが問題となる．

7. 確率予測

前記の問題をある意味で拡張して，“$Y\in A$ となる確率はpである”という形で予測する場合を考えよう．これは現に，アメリカでは天気予報で行なわれている方法である．この問題は，ベイズ式で考えれば単純な問題であって，単にYの事後確率を計算すればよいことになる．そうでない場合には，これはYの条件付確率 $P_\theta\{Y\in A|X\}$ の推定問題とも見なされるかもしれない．しかし，この問題は，単なる推定問題とは違ったとり扱いを必要とする．まず条件付確率の不偏推定を求めたりすることは，この場合にはあまり意味がない．というのは，ある事象の真の確率が1％であって小さいが，もしそのことが起こったら重大な結果が生ずるような場合，その確率を0％とするのと2％と推定するのとでは，結果は対称的にはならないからである．

このような場合，“$Y\in A$ となる確率がpである”とのべることは，“$P\{Y\in A\}=p$ であるものと仮定して行動せよ”ということを意味していると考えることができる．ということは，Aが起こることに対する賭けの率が，$p/(1-p)$ 以上ならば，$Y\in A$ というほうに賭け，それ以下ならば $Y\notin A$ にかけることを意味している．そこでこのようにして行動したとき，平均利益がなるべく大きくなるように，あるいは平均リグレット（pがわかっているとしたときに得られる平均利益と，予測値にもとづいて行動したときの平均利益との差）を最小にするように，pをXにもとづいて決定することが考えられる．

8. 政策予測

最初にのべたように，予測の問題のなかには，予測される量Yの確率分布が，ある行動によっても影響される場合もふくめて考える必要がある．このような問題のなかで最も自然な問題は，Yの値がなるべく望ましい水準に近づくように，行動をとる問題である．

いまYが実数 $Z_1, \cdots\cdots, Z_p$ が制御可能な変数として，その間に

$$Y=\beta_1Z_1+\cdots\cdots+\beta_pZ_p+u$$

という線形回帰関係がなり立っているとしよう．ここに $\beta_1, \cdots\cdots, \beta_p$ は未知の回帰係数であり，uは誤差または攪乱を表わす項である．uは平均 0，分散 σ^2 の正規分布に従うものとする．

いま，$Z_1, \cdots\cdots, Z_p$ およびYに関して，u組のデータが与えられ，それから最小2乗推定量 $\hat{\beta}_1,$

……, $\hat{\beta}_p$ および σ^2 の不偏推定量 $\hat{\sigma}^2$ が計算されたものとしよう. このとき Y の値をなるべくある定数 c に近づけるように $Z_1, \cdots\cdots, Z_p$ を定める問題を考える. すなわち, 母数の推定量にもとづいて $\hat{Z}_1, \cdots\cdots, \hat{Z}_p$ を求め, それに対応する Y の値を

$$\hat{Y}=\beta_1\hat{Z}_1+\cdots\cdots+\beta_r\hat{Z}_r+u$$

とするとき,

$$E(\hat{Y}-c)^2$$

をなるべく小さくするようにする問題である. この問題は, Fisher, Zellner がベイズ式の接近法で解いたほかあまり論ぜられていないが, 非ベイズ的な考え方からは, Z を次の形で定めればよいことがわかる. $\underline{\hat{Z}}, \underline{\hat{\beta}}$ をそれぞれベクトルで表わすと, その解は

$$\underline{\hat{Z}}=\frac{cM}{\underline{\hat{\beta}}'M\underline{\hat{\beta}}+k\hat{\sigma}^2}\underline{\hat{\beta}}$$

ただし, ここで M はデータにおける Z の積率行列となる. k は正の定数であるが, だいたい $k=\max(5-p, 0)$ とするのがよい.

同様の問題としては, Y の値をなるべく大きくするように制御変数 (あるいは政策変数) の値を定めることも考えられる.

9. むすび

以上は, 統計的予測のいろいろな形の問題のなかから, いわばランダムにとりあげたものにすぎない. 今後なおなされるべきことは多い. 今後の課題について, 若干の展望をのべておこう.

1) 普及の面：予測の問題に必要な理論のうちで, すでに開発されている部分も少なくない. しかしそのなかのかなりの部分は, 数理統計学の教科書や論文のなかに散在しており, 用語や概念も統一されていないので, それを体系的に整理し, OR ワーカーの目にもふれやすい形にする必要がある.

2) 計算の面：ある種の問題, とくに区間予測に必要な数値計算, あるいは数表の整備が望まれる. また計算機プログラムの開発が必要な部分も少なくない.

3) 理論の面：点予測以外の予測形式の理論は, まだきわめて不十分である. "よい予測方式" の基準とその求め方をもっとくわしく分析する必要がある.

4) モデルの面：モデル選択の問題を, 予測との関連でより深く研究する必要がある. とくに分布型の問題, 回帰式の関数形の問題等について立ち入った研究が望ましい.

5) 実際家と理論家との交流：現実に必要とされる予測のあり方については, すでにのべたものよりもっといろいろな種類の状況が考えられ, また, そこに要求されるものも複雑なものがあるであろう. それを適切な形で定式化し, 答を与えるためには, 実際家と理論家との深いレベルでの交流が望まれる.

文 献

以上にのべたことの理論的内容はついては, 竹内啓, 計量経済学の研究, 東洋経済新報社, 1972 の第Ⅲ部にあつかわれている.