On predictive Analysis

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/37290

予　測　的　な　分　析

平　館　道　子

Ⅰ　確率の付与と統計的モデル

経済や経営に関する統計データの分析はほとんどの場合，何らかの意味で政策や行動の選択決定に関連して行われると言ってもよいであろう。いくつかの政策や行動の中から現実に実行するものを選択する際，それらの代替的な行動の生み出す結果を相互に比較する必要があるが，一つの行動に一つの結果が対応しているのであれば，選択に際してほとんど困難はないであろう。しかしそこに不確実な要因が含まれていて単純な比較が不可能であり，不確実な結果の評価という問題に対処せざるを得ない場合が普通である。そこでその不確実な現象に関するデータがあれば，不確実性を適切に把握するために分析を行なうことになるであろう。もし問題となっている不確実事象に直接関連する統計データが存在しない時には，これまでの経験や知識，あるいは他の人の経験に学んで直感的な対処が行われるであろう。いずれにしてもその場合必要とされるのは問題の事象や量の確率的な叙述である。例えばある商品の来期の需要はどのくらいかが問題である場合，よく見られることは，予想される平均的なレベルとか，最大限あるいは最小限どのくらいという様な形で処理することであるが，包括的に対処するには可能な範囲全体について不確実さの程度を示す必要がある。しかしこの事は何らかの形で不確実性に関する判断あるいは予測といったものが加えられなければ実行できない。それは来期の需要はその期の経済状態や市場条件の様な制御できない要因に依存するであろうから，これらの要因の可能な状態を何らかの形で総合しなければ，必要とされる需要の不確実性を導出できないが，この様な総合は分析に当っている人あるいは決定を行なう人等の判断なしには行うことができないからである。問題の不確実量をy，これに影響を与える要因をθと表すことにしよう。θは実際に需要が発生する時にはある状態に定まっているであろうが，分析者にとっては未知であり，様々な状態になる可能性をもっている。このθの影響を定式的な形で表現しない場合には，これまでの経験などにもと

づいて直感的にそして暗黙的に総合して，yの変動の可能性について確率的な叙述を行うであろう。これを$p^*(y)$と記すことにしよう。以下では，この様な表現は離散的な場合も連続的な場合もあらわすことにする。

次に，θの様々な状態の各々に対して，もしその状態が実現するとしたらyの不確実性はどうであるかを考える場合，この条件つきの確率を$p(y|\theta)$と表わすと，これは分析者が制御し得ないθに依存しているから，それだけを考慮しても条件つきのことしかわからないのであり，求める条件つきでないyの不確実性について直接的な答は与えてくれないのである。必要とされる条件つきでない確率$p(y)$はθの様々な状態に対する不確実性の判断$p(\theta)$が付与された時はじめて次の様にして求めることができる。

$$p(y)=\sum_{\theta}p(y|\theta)\ p(\theta) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots(1)$$

$p(\theta)$はθの事前確率，$p(y)$は予測確率と呼ばれている。現実の問題で特定の状態を一つ想定し（θ^*），$p(y|\theta^*)$を$p(y)$と同一視しているのではないかと思われる様な事が時として見受けられる。θは制御したり選択したりできない性格のものであることに注意する必要があり，その意味では式(1)による定式的な方法による方が整合的であると言えよう。更に$p(y|\theta)$について問題によってはかなりよく根拠づけられた統計的モデルが存在する場合があり，それは分析にとって強力な武器である。定式的な方法によれば，問題は事前確率の付与に絞られる。ただここで注意すべきことは，この様な統計モデルにおいてはθは観測不可能であることが多く，問題によってはかなり抽象的な概念だという点である。従って事前確率を付与する時には，その含意を個々の問題に即して把握する必要がある。後にも述べる様に問題に関してデータを利用できる場合には,データに基づく事後確率として分析結果を集約し,それを次の分析のために蓄積しておくことができる。もし現在の問題に対してその様な形での過去の蓄積があれば，それを事前確率として用いることができ，また過去の経験がそのまま適用できないとしてもそれを基に生じた変化について明示的に考慮することが可能になるであろうから，判断に対する大きな助けとなる筈である。

この様にして導出される予測確率$p(y)$に対し直感的に付与される予測確率$p^*(y)$もやはりある程度の説得力を持つものであろう。両者は一致するべきであるが，現実問題としてはとてもそれは望めない。従って両者をつき

合わせてどこに問題があるのか，付与した事前確率の含意は何であるのか等を検討することは有用であろう。

直接的に p (y) を付与する場合，過去のデータの平滑化によることがあるが，経済や経営に関するデータの場合，この様な方法が全く妥当性を欠くことが多いという点に注意しなければならない。それは過去のデータはその時々の事象の影響を受けている筈であるが，更に生産量という様な意思決定にも影響を受けている筈だからである。同じ状況が保たれている様なことはほとんどないから，過去のデータを用いようとする時にはその変化について考慮することが必要である。変化に応じた修正を可能にするためには，その時々の事象や行動がどの様な影響を与えているかを分析してそれらを分離し，把握しておかなければならない。回帰分析はこの様な場合の強力な方法の一つである。

次に問題の不確実性が解消して結果が実現した時には，それによって自分の判断の性格や良さなどについて評価することができるし，また実際その様な事後的な検討は必要とされるであろう。そのデータの，あるいはそのデータからの適切な統計量の値の確率を p (y) から計算することによって，その確率が小さなものであれば成功しなかったと感じるであろうし，またとり得たかもしれない最大の確率の値と比較する等をして，判断の仕方に対する何らかの助けや示唆が得られるであろう。この様な観点から評価を点数ルールによって行ない，極端な判断をチェックするなど確率付与の改善のための試みが行われている。またデータの分布状態と予測確率分布との適合の程度を通常のカイ二乗の様な形で評価することも考えられるが，これは点数ルールの中に包含されるであろう。

定式的に予測確率を導出する場合にはデータが得られるとベイズの定理によってθの事後確率 p ($\theta \mid y$) を求めることができ，これを次の分析のために経験として蓄積することができる。事後確率はよく知られている様に次の様にして導出される。

$$\mathrm{p}(\theta \mid y)=\frac{\mathrm{p}(\theta)\ \mathrm{p}(y \mid \theta)}{\sum_{\theta} \mathrm{p}(\theta)\ \mathrm{p}(y \mid \theta)} \quad \cdots\cdots\cdots\cdots\cdots\cdots(2)$$

θの推定値が必要であればこの事後分布の期待値あるいはもっとも確率の高い値(モード)を推定値とすることができるし，またθの値に関する確率的な叙述

が可能である。もし想定した構造に変化がないと判断されるならば，この事後確率を用いて将来のデータについての予測確率を求めることができる。将来の，あるいはまだ実現していない標本をy_fとすれば，これまでのデータyが得られた時のy_fの予測確率は次の様になる。

$$\mathrm{p}(y_f \mid y) = \sum_{\theta} \mathrm{p}(y_f \mid \theta)\,\mathrm{p}(\theta \mid y) \cdots\cdots\cdots\cdots\cdots\cdots (3)$$

これを用いてベイジアンの観点からの予測を予測域によるような様々な適切な形で行なうことになるであろう。

II　回帰モデル

ここではこれまで述べた点を線形回帰モデルによって例示しよう。不確実量をy，yに対して影響を与えると想定される要因を$Z=(z_1, z_2, \cdots\cdots, z_k)$，Zが$y$に対して与える影響は回帰係数$\theta'=(\theta_1, \cdots\cdots\cdots, \theta_k)$によって要約されると想定する。Z，$\theta$が与えられた時の$y$の確率法則$\mathrm{p}(y \mid Z, \theta)$は正規分布に従い，

$$\mathrm{p}(y \mid Z, \theta) = N(Z\theta, \sigma^2)$$

であり，σ^2は単純化のために既知であるとしよう。θに対して事前確率分布$N(m, V)$が付与されたとしよう。ここに$m'=(m_1, \cdots\cdots, m_k)$，Vは$\theta$の共分散行列である。この時Zが与えられた時の$y$の予測確率分布は式(1)の関係から（この場合和ではなく積分で求められるが）次の様に求められる。

$$\mathrm{p}(y \mid Z) = N(Zm, \sigma^2 + ZVZ') \cdots\cdots\cdots\cdots\cdots\cdots\cdots (4)$$

Zは制御可能な要因や可能でない要因を含むであろう。しかし観測可能な量であり，(4)はそれらの具体的な値が与えられた時の条件つきのものである。一般的な経済環境の様な状態を考慮したい場合の一つの方法については次の項で述べることにしよう。$\mathrm{p}(\theta)$を正規分布としたのは必ずしもこれが判断を忠実に反映しているからという訳ではない。不確実性に対する判断を解析的な関数で表現するのは近似である。更にこれがデータの分布に対して自然共役であり，事後確率分布がまた同じタイプの分布として求められるという便宜のためである。このことを認めたとして，現実の分析に際して問題になるのは特に共分散行列Vの付与である。既に述べた様にθは観測できな

いものであり，yに対する要因の影響をこの形で要約するのが適切であるとして想定された抽象的な概念である。従ってそれの平均や共分散がどれだけであるか判断することにはかなり困難を感じるのが普通であろう。ここでⅠで述べた直感的な予測確率 $p^*(y\,|Z)$ との比較検討が役に立つであろう。この問題についていくつかの示唆や研究があるが（〔2〕，〔6〕），方法論に関するもっと広範な議論が必要である。

次にこの問題に関してデータが観測されたとしてこれから得られる結果をまとめよう。これらの観測値を $y_0'=(y_1,\cdots\cdots,y_n)$, $Z_0'=(Z_1',Z_2',\cdots\cdots,Z_n')$, $Z_i=(z_{i1},z_{i2},\cdots\cdots,z_{ik})$ とすると θ の事後確率分布は次の様になる。

$$p(\theta\,|\,y_0,\ Z_0)=N(m_1,\ V_1)$$

$$m_1=(\sigma^{-2}Z_0'Z_0+V^{-1})^{-1}(\sigma^{-2}Z_0'Z_0\hat{\theta}+V^{-1}m)$$

$$=m+(Z_0'Z_0+\sigma^2V^{-1})^{-1}Z_0'(Y-Z_0m)\cdots\cdots(5)$$

ここに$\hat{\theta}$は最小２乗推定量である。

$$V_1=(\sigma^{-2}Z_0'Z_0+V^{-1})^{-1}=V-VZ_0'(Z_0VZ_0'+\sigma^2I_n)^{-1}Z_0V\cdots\cdots(6)$$

式(５)に示される様にθの事後期待値は二つの表現によって表わされている。一つは通常のθの推定値である最小２乗推定値と事前期待値mとの加重平均であり，他はmをyの予測誤差 $(y-Z_0m)$ によって修正したものとしてである。前者によればV^{-1}が$Z_0'Z_0$ に比較して無視し得るほど小さければ，θ の事後期待値は最小二乗推定値によって近似できるであろうこと，またこの時には事後の共分散V_1は$\sigma^2(Z_0'Z_0)^{-1}$で近似されることがわかる。この事は伝統的な回帰分析がベイジアン分析の一つの極限として含まれていることを示している。後者の表現からはもし次のデータが得られるとすればその時のθの期待値はm_1を用いて同様に修正したものになることがわかる。この事は経験の蓄積という点から重要である。

次に将来のデータy_fの予測確率分布は，要因 Z_f が与えられるとすれば(３)から次の様になる。

$$p(y_f\,|\,y_0,\ Z_0,\ Z_f)=N(m_f,\ V_f)$$

$$m_f=Z_fm_1,\quad V_f=\sigma^2+Z_fV_1Z_f'\cdots\cdots(7)$$

III　いくつかの状態がある場合

不確実量yに対する要因Zの影響が一般的な経済環境の様な事象に依存して変化すると判断される場合，特に予測が問題になる時にはこの様な相異を総合的に考慮する必要があるであろう。事象をS_i, $i=1,\cdots\cdots,k$とすると事前確率は各々の事象の下で付与されることになる。これらを$p(\theta \mid S_i)$としよう。更に事象の事前確率$p(S_i)$が付与されたとしよう。予測確率は

$$p(y \mid Z)=\sum_i p(S_i)\sum_\theta p(y \mid \theta, S_i)\ p(\theta \mid S_i)$$

であるから，前項と同じ正規モデルの下ではこれは正規分布の混合である。$p(y \mid \theta,\ S_i)=N(Z\theta,\ \sigma^2 I_n)$，$p(\theta \mid S_i)=N(m_i,\ V_i)$であるとすれば（ここではZは$n x k$行列であるが）$p(y \mid Z)$は

$$\begin{aligned} p(y \mid Z) &= \sum_i p(S_i)\ p(y \mid Z,\ S_i) \\ &= \sum_i p(S_i)\ N(Zm_i,\ \sigma^2 I_n+ZV_iZ') \end{aligned}$$

観測値が得られた時にはθの事後確率は次の式から求められる。

$$\begin{aligned} p(\theta \mid y,\ Z) &= \frac{\sum_i p(S_i)\ p(\theta \mid S_i)\ p(y \mid Z, \theta, S_i)}{\sum_i p(S_i)\ p(y \mid Z, S_i)} \\ &= \sum_i p(S_i \mid y, Z)\ P(\theta \mid y,\ Z, S_i) \quad \cdots\cdots\cdots\cdots(8) \end{aligned}$$

これからわかる様にθの事後確率は，それぞれの状態の下での事後確率を状態の事後確率で平均したものになっている。各々の状態の下でのθの事後確率は前項と同様であり，状態の事後確率は

$$p(S_i \mid y, Z)=\frac{p(S_i)\, p(y \mid S_i, Z)}{p(y \mid Z)} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(9)$$

から求められる。

分布の混合については，例えば$p(y)$の様な予測確率は$p(y \mid \theta)$を$p(\theta)$によって混合したものと考えることができるが，ここで想定している様な状態は実際の問題では恐らく二つとか三つの様な少数であろう。ここでは典型的な状態における相異を明示的に区別することは意味があると判断されているのであるが，典型的な状態として識別されるのは恐らくそれほど多数ではないであろう。従ってその様な場合の混合された分布はそれ程取り扱いやす

いものではない。勿論数値的に確率を求め，必要な特性値を求めることは可能である。

Ⅳ　モデルの適合性の検討

既にⅠで述べた様に実現したデータの予測確率によって事後的に自分の分析が成功的かどうかを検討する方法が考えられるが，同様の考え方に立っていくつかの競合するモデルの間の比較あるいは選択というような問題を考えることができる。モデルの検討といっても，モデル $p(y \mid \theta)$ は事前確率と切り離して別個に考えるのではなく，事前の判断との関連においてその得られたデータによればモデルがどの程度確からしいと判断されるかを考えるのである。

簡単のために2つのモデルM_0，M_1があり，それぞれの下でのデータ分布，事前確率分布を $p(y \mid \theta, M_i)$，$p(\theta \mid M_i)$，モデルに対する事前確率を $p(M_i)$ $i=0$，1としよう。この時式(8)，(9)と同様にしてM_iの事後確率は次の様に求められる。

$$p(M_i \mid y)=\frac{p(M_i)\ p(y \mid M_i)}{\sum_i p(M_i)\ p(y \mid M_i)} \quad i=0,\ 1 \cdots\cdots\cdots\cdots\cdots\cdots (10)$$

$$p(y \mid M_i)=\sum_\theta p(y \mid \theta, M_i)\ p(\theta \mid M_i) \quad i=0,\ 1 \cdots\cdots\cdots\cdots (11)$$

このモデルの事後確率を妥当性の測度とすることは自然であろう。比較の方法として事後確率の比をとると，

$$\frac{p(M_0 \mid y)}{p(M_1 \mid y)}=\frac{p(M_0)\ p(y \mid M_0)}{p(M_1)\ p(y \mid M_1)} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (12)$$

事後確率の比は事前確率の比にモデルの予測確率の比を乗じたものとなっている。後者はしばしばベイズ・ファアクターと呼ばれている。もしモデルの事前確率が等しければこの事後確率の比はベイズ・ファアクターに等しいが，競合的なモデルの比較が問題になる様な場合，それらの事前の重みはそれほど大きく異なることはないであろう。その様な場合にはベイズ・ファアクターが比較の基準となる。すなわちそれぞれのモデルの予測力といったものが比較の根拠とされるのである。

これらの点の例示としてまた回帰モデルについて考えてみよう。様々なモデルを考えることが可能であることは勿論であるが，しかし本当に競合的で代替的なモデルとして考慮の対象になるのはあまりに異質なものではないで

あろう。この事は同一のデータを複数の人が分析した結果を局外者が評価しようとする場合にはあてはまらない。しかしこの場合にも同じ考え方で式(12)にもとづいて評価することができる。しかし一人の人が自分の分析に際して考える時にはある変数を説明要因として考慮すべきかどうか，要因の及ぼす影響は固定的であるかどうかという様な少くとも形式的にはむしろマイナーな相異であろう。この様な観点をJeffreysは〝one new parameter″という言葉で表現している。(〔5〕)。

ここでは次の様な問題を考えよう。yは二つのモデルM_0, M_1の下では各々

$$p(y \mid M_i,\ Z_i,\ \theta_i) = N(Z_i\theta_i,\ \sigma^2 I_n) \qquad i=0,\ 1$$

に従うと想定される。ここでもσ^2は既知であるとしよう。また$Z_1=(Z_0\ \ Z)$ $\theta_1'=(\theta_0'\ \ \theta')$であるとする。即ちZを説明変数とするかどうかという問題である。θ_iの事前確率については

$$p(\theta_1 \mid M_1) = p(\theta_0 \mid M_0)\ p(\theta) \cdots\cdots\cdots\cdots(13)$$

の様な関係が成立ち，$p(\theta_0 \mid M_0) = N(m_0,\ \sigma^2 V_0)$, $p(\theta) = N(m,\ \sigma^2 V)$ であるとしよう。したがって

$$p(\theta_1 \mid M_1) = N(m_1,\ V_1),\quad m_1' = (m_0'\ \ m')\quad V_1 = \sigma^2\begin{pmatrix} V_0 & 0 \\ 0 & V \end{pmatrix}$$

である。これらを式 (13) の様な表現におき変えられることは容易にわかる。

予測確率はこれまでの結果を用いれば次の様に求められる。

$$p(y \mid M_i) = N(Zm_i,\ \sigma^2(I_n + Z_i V_i Z_i'))\quad i=0,\ 1$$

$\sigma^2(I_n + Z_i V_i Z_i') = \Sigma_i$, $Z_i m_i = \bar{\theta}_i$とおくと，$\bar{\theta}_1 = \bar{\theta}_0 + Zm$, $\Sigma_1 = ZVZ' + \Sigma_0$であるから，ベイズ・ファアクター$B_{01}$は次の様になる。

$$B_{01} = \frac{p(y \mid M_0)}{p(y \mid M_1)} \propto exp\ \left\{-\frac{1}{2}(y-a)'\ \Sigma^{-1}(y-a)\right\},$$

$$a = Z_0 m_0 - (\Sigma_0^{-1} - \Sigma_1^{-1})^{-1}\,\Sigma_1^{-1} Zm$$

$$\Sigma = \Sigma_0^{-1} - \Sigma_1^{-1} = \Sigma_0^{-1} Z(Z'\ \Sigma_0^{-1} Z + V^{-1})^{-1} Z'\ \Sigma_0^{-1}$$

ここでもし事前の期待値がすべて0で，V_0^{-1}, V^{-1}が各々$Z_0' Z_0$, $Z'Z$に比較して無視できる程度に小さく，かつZの列がZ_0の列と直交している

ならば,

$$B_{01} \propto exp\ \{-\frac{1}{2}\sigma^2(y'\ Z(Z'\ Z)^{-1}Z'\ y)\}$$

$$=exp\ \{-\frac{1}{2}\sigma^2(\hat{\theta}'\ X'\ X\hat{\theta})\}$$

となる。$\hat{\theta}$ は最小二乗推定量であり，$\hat{\theta}'\ X'\ X\ \hat{\theta}$ はよく知られている様に通常の線形回帰問題で用いられる検定統計量である。

V　予測確率

ベイジアン統計理論はデータからの証拠はデータ以外の証拠と結合して分析されるべきであり，そうして始めて意味のある分析が可能であることを提唱する。この場合一般にデータ以外の証拠はモデルに含まれるパラメータの事前確率という形で表現され，データからの証拠を結合したパラメータの事後確率が分析の要約となり，特定の目的に対して役立つことは例示した通りである。そしてこれまで主としてこの方向での理論の展開がなされて来た。事前確率に関してはしばしば論争の的になり，特に *Thomas Bayes* の論文の〝絶対的無知〟を表現する一様事前確率については様々な批判や修正が論じられている。最近になって，*Bayes*はこれまで論じられて来た様に〝絶対的無知〟をパラメータの一様事前確率によって把えていたのではなく，可能な結果に対する事前的な確率，すなわち予測確率が一様であることとして把えていたという指摘がある(〔7〕)。この様な理論的アプローチは *de Finette* によって採られ，伝統的統計理論の無作為標本に匹敵する *exchangeability* という概念が導かれている(〔4〕)。本稿でも若干の例で示した様に，分析の過程においては問題をパラメータによって要約できる局面と予測確率によって把握すべき局面とがあり，これらは表裏一体をなして分析を意味あるものにすると言うことができる。この方向での研究には Box の論文など優れた研究があり(〔1〕，〔3〕等)，また情報の評価やサンプル・サイズの決定等に関連して論じられているが，さらに広範な分野での意味を明らかにして行くことが必要であろう。

文　献

1. Box, G.E.P., (1980) 〝Sampling and Bayes' Inference in Scientific

modelling and Robustness" JRSS, A. vol 143, 383−430.

2. Dempster, A.P., Rubin, D.B., Tsutakawa, R.K. (1981) "Estimation in Covariance Components Models", JASA, vol.76, No. 374, 341−353.
3. Geisser, S., Eddy, W.F. (1979) "A Predictive Approach to Model Selection", JASA. vol. 74. No.365, 153−160.
4. de Finetti, B., (1970) "Theory of Probability" I, II, J.Wiley.
5. Jeffreys, H., (1961) "Theory of Probability" Oxford Clarendon Press.
6. Kadane, J.B., Dicky, J.M., Winkler, R. L., Smith, W.S., Peters, S.S., (1980) "Interactive Elicitation of Opinion for a Normal Linear Model", JASA, vol. 75, No.372, 845−854.
7. Stigler, S.M., (1982) "Thomas Bayes's Bayesian Inference," JRSS, A, vol. 145, part 2, 250−258.