デシジョン・メーキング序説

竹内 清

一

最近デシジョン・メーキング（意志決定）という言葉が、経営その他の分野においてもごく日常的に使われるようになってきた。ところでその場合、その語の常識的意味ではともかく、デシジョン・メーキングが科学的に必ずしも行われていないのが実情であろう。しかし一般的な傾向としては、単なる勘とか経験だけをもとにした判断を超えて、科学的なデシジョン・メーキングへの指向がさかんになってきたことは事実である。特に経営の場においてその傾向が著しいものがある。これは、最近の技術進歩の速度が、一昔前とは比べものにならないほど速くなり、それに伴ない、経済成長の高度化を背景として企業活動そのものが極めて多面化するとともに複雑化し、市場における競争も激化の度を加えてきているので、経営の技術そのものも高度化さなくては、企業存立の基礎さえ危くなり、到底企業の成長発展は期待できなくなってきているからに外ならない。

もちろん、経営の場においてだけでなく、国家的観点からもデシジョン・メーキングの科学化は必要であるとともも

に重要であり、最近の所得倍増計画等もかなり科学的に練られているものと考えられる。

個々の企業のおかれている一般的環境条件なり内部条件は、それぞれ異っているであろうが、ゴーイング・コンサーンとして企業そのものを維持発展させていくための基本的要諦は、デシジョン・メーキングを科学的に行うことである、ということができよう。ここではデシジョン・メーキングの問題をかなり一般的な形でとりあげ、その考え方に重点をおいて解明することにより、デシジョン・メーキングの体系化の一つの手がかりをつかむことにする。

最近実際的要請とも相俟って種々デシジョン・メーキングに関する文献、書物が多数出ているが、(1) ここではこれらのものも参照しながら展開を試みることにする。筆者の意図は、かなり文献的整理にも向っているのを予め断っておく。

（1）たとえば、参照したものの一部をあげると

[Ⅰ] M. H. Spencer and L. Siegelman, *Managerial Economics: Decision Making and Forward Planning*, Richard D. Irwin, Inc., 1959.

[Ⅱ] W. T. Morris, *Engineering Economy: The Analysis of Management Decisions*, Richard D. Irwin, Inc., 1960.

[Ⅲ] D. W. Miller and M.K. Starr, *Executive Decisions and Operations Research*, Prentice-Hall, Inc., 1960.

[Ⅳ] R. W. Morell, *Managerial Decision-Making: A Logical Approach*, The Bruce Pub. Co., 1960.

[Ⅴ] R. A. Howard, *Dynamic Programming and Markov Process*, The Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons, Inc., 1960.

二

デシジョン・メーキング（意志決定）の体系としては、種々のものが考えられているが、最も普通に受け容れられ

ているものは、つぎのごときものであろう。(2)

意思決定者（デシジョン・メーカー）は、当面の問題に関し、種々の可能な択一的な行動または戦略をもっており、彼は一つないしそれ以上の目標をもっている。択一的な行動として、個々の行動のあらゆる組合せをも部分集合として含むような、行動の集合を考えれば、実際的なすべての場合を網羅しうるであろう。また彼の統制し選択しうる行動と共に、実際に彼が目標を達成する度合を決定するであろう自然の状態または競争的状態――直接彼が統制しえないものである――が存在する。すなわち、意志決定者が行動を選択し、ある特定の自然の状態が起ると、彼に対して一定の事態がもたらされる。この事態は、彼に彼の目標と関連して一定の満足ないし効用をもたらすであろう。ここにおいて彼は、自分のとった行動に対してどのような成果がもたらされるか予測することが必要となる。ここにデシジョン・メーキングの過程で予測体系の果す大きな役割がある。右の成果の評価と関連しては、価値体系が役割を演ずることになる。

右に述べたことを行列の形で図式に表わすと次頁の表のようになる。S を択一的な行動、N を自然の状態、θ を成果、$V(\theta)$ を成果に関する利得ないし損失（一般的には満足とか効用）とする。$V(\theta_{ij})$ は、S_i という行動をとり、N_j という自然の状態が起きた場合の成果 θ_{ij} の利得または損失を表わす。

右のような行列の表は、支払表または精算表とよばれる。右のような支払表が与えられた場合、決定の問題は、特定の行動 S_i を選択する問題になる。それではどのようにしてそのような行動を選択すべきか、ということが問題になる。いかなる行動を選択すべきか、の規準になるのが、決定の規準であり、これは、$V(\theta)$ の比較を通して行われる。ここに意志決定者の目的ないし、目標が反映されることになり、種々の決定規準が提唱される。

以上述べたのは通常最も広く受け入れられているデシジョン・メーキングの定式化であるが、モーレル（R. W.

$S \backslash N$	N_1	$N_2 \cdots\cdots N_j \cdots\cdots N_m$
S_1	$V(\theta_{11})$	$V(\theta_{12}) \cdots\cdots V(\theta_{1j}) \cdots\cdots V(\theta_{1m})$
S_2	$V(\theta_{21})$	$V(\theta_{22}) \cdots\cdots V(\theta_{2j}) \cdots\cdots V(\theta_{2m})$
$\vdots$	$\vdots$	$\vdots \qquad \vdots \qquad \vdots$
S_i	$V(\theta_{i1})$	$V(\theta_{i2}) \cdots\cdots V(\theta_{ij}) \cdots\cdots V(\theta_{im})$
$\vdots$	$\vdots$	$\vdots \qquad \vdots \qquad \vdots$
S_n	$V(\theta_{n1})$	$V(\theta_{n2}) \cdots\cdots V(\theta_{nj}) \cdots\cdots V(\theta_{nm})$

Morell）〔VI〕は、論理学的な観点から、決定ならびにデシジョン・メーキングの問題を考える。彼は、デシジョン・メーキングに果すOR（オペレーションズ・リサーチ）的技法の役割を吟味した後、決定の定義について種々の経営学者、哲学者、心理学者等の所説を引用し、それを比較検討して、その共通の基本的な考え方としてつぎのものを把握する。すなわち、決定とは二つまたはそれ以上の選択肢のなかから一つの選択肢を選び出すことである。モーレルは、数学者や統計学者の側からする決定の定義を検討していないが、彼等の定義も基本的には右のものと同じである。これは本節前半ですでにみたところである。

　モーレルは、決定についての基本的な考え方によってつぎの含意がもたらされる、と考える。すなわち、(1)デシジョン・メーキングは有目的な行動——目標または目的によって導かれる——であるべきである。(2)デシジョン・メーキングは目的指向的であるから、選択肢の選出は、ある目的を達成する手段の選択である。(3)選択肢の何れを採択し棄却するかの理由が探究され、評価されなければならない。究極において、最終的な判断がなされなければならない。

　このように解釈して後、決定は目的間またはある目的を達成するための手段間の判断を意味するものと一応考えられなければならない。ここで判断とは、明確に断言する（肯定または否定する）知的行為を意味するものと考える。

　右のようにして、決定は目的またはある目的ないし目標を達成する手段間の判断を意味するものと一応考えられたので、また判断は明確に断言する知的行為と考えられるので、決定という語は、したがって、つぎのように正式に定義しうるであろう。すなわち、決定とは、択一的な目的、またはある目的を達成するための択一的な手段間の知的断

言（肯定または否定の）である。したがってデシジョン・メーキングは、知的活動であるとする。

決定が論理的に行われるためには、前提または命題にもとづいてなされなければならない。決定がそこから導き出される前提によって支持される決定は、論理的決定とよばれる。これは劇の終りの方で登場してくるものであり、デシジョン・メーキングのより初期の段階では、不確実性、分析、問題の意識、問題の限定、選択肢の定式化等が考慮されなければならないとする。

彼はこのようにして、十分な決定の段階として、不確実性の段階、分析および限定の段階、選択肢または仮説の提起の段階、立証の段階に分け、特に論理的推理の役割を強調する。最後の立証の段階では、三段論法的推論をいくつか例示している。

統計学は、ワルド（A. Wald）によって、不確実性の下におけるデシジョン・メーキングを取扱うものとして実践的に定義されてから、極めて多方面に高度の発展をしてきた。(3) しかし決定理論の本質そのものについての反省ないし吟味は極めて少ない。(4) そこでここではこの問題に対するテュキー（J. W. Tukey）の考えをみることにする。結論対決定という観点から、近代的な決定理論の意味における決定の考え方に反省を加えるとともに、彼自身の定式化を行っている。(5)

確かに彼の言うように、われわれは決定理論について読み、かつそれについて聞き、また誰もが決定を行っているわけである。しかし、決定理論の本質そのものについての論議は極めて少ない。そこで彼はこの問題に対してつぎのように考えを展開する。

実際の決定は、決定理論の取扱いにおいて長い間踏襲されてきた形――「われわれは受け入れる」――よりはむしろ「恰も……のようにここしばらく行動するよう」に決定しようという形にはるかに近いものである。ここでの「恰も…

…のように……行動する」と「ここしばらく」という制限が、二つの別々の重要な概念、すなわち、結論と決定を区別するのに役立つ概念、彼の言わんとすることを要約する概念を伝達することになる。

たとえば、技術者が直ちに二つの橋梁構築法を選択しなければならないとき、あるいはビジネスマンが当面問題となっている二つの営業政策を選択しなければならないとき、彼等は、この緊急的な状況において、選択肢Aを選択肢Bに対して秤量し、より大きな報酬をうみだすであろう選択肢を選出するように努力しなければならない。可能な行動が限定され、種々の「自然の状態」におけるそれらの結果が理解され、これらの自然の状態についての若干の証拠物が手もとにある。各場合において個人は、恰も選択肢Aからの報酬が選択肢Bからのそれよりも事実上大きくなるであろうかのように（簡単に「A＞B」と表わす）行動すべきかどうか、あるいはその逆が真である（「A＜B」）かどうかを判断しなければならない。

つぎの三つの択一的な決定、すなわち

(1)　恰も　A＞B　であるかのように現状で行動すること

(2)　恰も　A＝B　であるかのように現状で行動すること

(3)　恰も　A＜B　であるかのように現状で行動すること

という三つの択一的な決定は、合理的に述べられているように考えられる。一方

(1′)　A＞B　を受け容れること

(2′)　A＝B　を受け容れること

(3′)　A＜B　を受け容れること

という踏襲的な選択肢の叙述は、読者または研究者を誤らせるのに（無意識的に）十分適してきているように思われる。

「恰もA>Bのように行動する」という場合、「A>B」という立言の「真」または「もっともな疑いもない確実性」ということに関しての判断はしていない。「ここしばらく」と言うときは現在考察下にある特定の状況だけを指しているのである。したがってわれわれがなしたことは、AとBの相対的な価値に関しての証拠と、種々の行動（行動であって決定ではない!）の現状での確からしい結果の両方を秤量することである。最終的に、われわれは、もし本当にA>Bであったら適当であったであろう特定の行動は、当面の特殊の状況において採用するのに最も合理的なものである、と決定したのである。

ところで結論は、もし異常に強い反対の証拠が起きなければ、またそれまで、実験または観測の条件に適用できるものとして受け容れられるはずの立言であるとする。したがって、結論は、異常に強い反対の証拠が起きなければ、そして起るまで依然として受け容れられるはずであるから、すべての結論のうちのごく小部分は、やがて、くつがえされるであろうことを意味する。それゆえ、結論は、反対の証拠が十分強くなれば、将来棄却されるということが仮定されて受け容れられる。それは持続的な価値をもつが、必ずしも永続的な価値をもつものと考えられない。

結論についてのこれらの特性は、決定理論家の決定の特性とは極めて異るものであり、この相異は重要である。

(2) ここでの定式化と同一の線上にあるものとしては

[VI] I. D. J. Bross, *Design for Decision*, The Macmillan Co., 1953.
([VII] 竹内清訳「決定と計画」みすず書房, 1960.)
M, H. Spencer and L. Siegelman [I]
W. T. Morris [II]
D. W. Miller and M.K. Starr [III]
[VIII] R. N. McKean, *Efficiency in Government through Systems Analysis*, John Wiley & Sons, Inc., 1958.

（３） たとえば統計学の面での若干の文献をあげると

[IX] A. Wald, *Statistical Decision Funktions*, John Wiley & Sons, 1959.

I. D. J. Bross 〔IV〕

[X] I. J. Savage, The *Foundations of Statistics,* John Wiley & Sons, 1954.

[XI] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisionns*, John Wiley & Sons, 1954.

[XII] R. D. Luce and H. Raiffa, Games and Decisions, John Wiley & Sons, 1957.

[XIII] H. Chernoff and L. E. Moses, *Elementry Decision Theory*, John Wily & Sons, 1959.

（[XIV] 宮沢光一訳，「決定理論入門」紀伊国屋，1960.）

[XV] R. Schlaifer, *Probability and Statittics for Business Decisions*, McGraw-Hill Book Co., 1959.

（４） I. D. J. Bross [VI]は、この問題に関して極めて興味深い分析を行っている。決定理論の哲学的基礎を実用主義原則と関連して吟味し、絶対主義の立場を棄て相対主義の立場を強調する。また古典的な統計的推理においては、行動と関連し、統計的検定よりも統計的推定の重要性を考慮し、小標本理論の行きすぎに対しても反省を与えている。また統計的仮説検定において、仮説の採択または棄却に関して、抽象的段階での結論を、実際的行動の観点から批判している。ここでのテュキーの論議と関連して興味深い点である。

（５） 〔XVI〕J. W. Tukey "Conclusions vs Decision," *Technometrics*, Vol.2, No.4, pp. 423～433.（Nov. 1960.）

三

意志決定者が決定を行うのは、その結果によってある望ましさとか満足、または効用というものがえられるからに外ならない、もちろん、その決定は、条件反射的に無意識的に行われることもあろう。しかしここではそのような場合は除外する。意志決定者が二者択一的な問題に直面し、どのような行動ないし戦略を選択したら、目的が最もうまく実現されるかという観点から、決定の問題を考えることにする。

決定の問題がそもそも起きてくるのは、本質的には、決定を行った成果が、決定時点より将来において発生するも

のであり、その成果を決定時点で前もって百パーセント確実に知ることができないからに外ならない。すなわち、不確実性の問題がここでは決定的なものとなる。

ある一定の行動をとった場合、その成果は、意志決定者の統制しえない外部条件に依存するであろう。このような意志決定者の統制しえない外部条件は、一般に自然の状態とよばれるものである。もし当該意志決定者と競争的関係にある他の意志決定者の行動ないし戦略が、陽表的に対立関係におかれる場合は、競争的状態とよばれる。

不確実性は通常二つに区別される。たとえばチャーノフとモーゼス[VIII]は、つぎのように述べる。

一つは、無作為性によるものである。この型の不確実性は原理的には比較的取扱い易いものでとり、無作為性の法則が適用される場合である。

他の型の不確実性は、どのような無作為性の法則が適用されるか分らない場合に起るものである。この場合、適当な実験を行って観測をとることができうる場合には、それを推定する一つの方法であるが、実験を行うに要する費用と、間違った決定を行った場合の費用を考慮しなければならない。これは逐次決定の考えにつながるものである。

前者の場合、すなわち、無作為性の法則が既知の場合の決定を、危険の下での決定とよび、後者の場合、すなわち、どのような無作為性の法則が適用されるか分らない場合の決定を、不確実性の下での決定とよぶ論者も少なくはない。(6)

一方特異の立場としてはシャックル（G. L. S. Shackle）の流れがある。(7) シャックルが、決定の問題として取扱うのは、不確実性の条件の下におけるものであるが、彼は意志決定者が、非反復的な一回限りの実験であるいくつかの行動表を心のなかにもっている場合を対象とする。したがって、既述の危険の状態の下における決定、すなわち、保険数学的に確実である成果が付与されうるような行動は当面の彼の議論から除外されることになる。(8) そのような場合の

接近法は、通常のものとかなり異っている。彼は主観的な不信、または潜在的な驚きという概念を用いて分析を進め主観的な確率を用いることに対して強く反対する。

（6） たとえば
M. H. Spencer and L. Siegelman［Ⅰ］
W. T. Morris［Ⅱ］
D. W. Miller and M. K. Starr［Ⅲ］

（7） シャックルの理論は次書によって知ることができる。
[XⅦ] G. L. S. Shackle, *Expectation in Economics*, Cambridge, 1952.
[XⅧ] G. L. S. Shackle, *Time in Econmics*, North-Holland Pub. Co., 1958.
またシャックルの理論を中心としたシンポジウムとして
[XIX] *Uncertainty and Bussiness Decisions*, ed. by C. F. Carter, G. P. Meredith and G. L. S. Shackle, Liverpool Univ. Press, 1st ed. 1945, 2nd. ed. 1957.
シャックルの理論を拡張して不確実性の下における投資決定の理論を展開したものとして
[XX] R. A. D. Egerton, *Investment Decisions under Uncertainty*, Liverpool Univ. Press, 1960.

（8） G. L. S. Shackle [XⅧ] p. 35.

四

デシジョン・メーキングの型としては、分類規準の異なるにしたがっていくつかのものが考えられるが、意志決定者が自然の状態または競争的行動についてもっている情報量の如何によって、前節でふれたように確実性の下でのデシジョン・メーキング、危険の下でのデシジョン・メーキングおよび不確実性の下でのデシジョン・メーキングという三つのものに分けられるのが普通である。

確実性の下でのデシジョン・メーキングは、意志決定者が、どの自然の状態が起るか確実に知っている場合のデシジョン・メーキングである。支払表でいえば、列が一つからなっている場合である。理論的にはその結果が最大の望ましさをもった行動を選択するのが、最適となるであろうから、問題はない。もちろん、可能な行動ないし戦略が無数に多くある場合には、実際的困難もありえようが、理論的には、決定理論本来の困難はないので、確実性の下でのデシジョン・メーキングは、決定理論そのものの対象として陽表的に取り上げられないことにもなる。(9)

危険の下でのデシジョン・メーキングは、自然の状態としては複数個のものが存在するが、意志決定者が自然の状態の各々の起る確率を知っている場合に起るものである。たとえば、保険数学の対象になるような事象を問題としている場合のデシジョン・メーキングはそうである。たとえば、現在三十歳の男子が今後二十五年生きている確率は、生命表から推定されるであろう。保険会社は、生命保険の場合、生命表をもとにして可能な自然の状態についての確率を知ることができ、ほとんど確実に損のないような保険料率というものを決めているのである。発生の確率が分らない不測の事態、たとえば、戦争とか天災のような場合については免責条項を設けているので、このような状態は考慮外におくことができるわけである。企業の場において、種々の自然の状態の確率が意志決定者に既知であるのは、先験的に知りうる少数の場合は別として、通常は過去にそれらの事象がいくらの頻度で起ったかを決定することによってえられるものである。先験的に知りうる場合とは、たとえば包装の中にラッキー・カードのごときものを入れて商品を発売するとき、各等級の割合が前もって分っているような場合である。しかしこのときとても、実際に出荷した商品が全部売り切れたとき、先験的に分っている確率が適用されるわけである。もし一定期間内に売れ残りなどあった場合には、前もって分っていた先験的確率が実現されないことになり、またラッキー・カードが当っても商品引換をしない顧客も実際にはありうるので手配しておいた賞品と実際申出のあった数の間にギャップの生ずることもあ

りうる。確率が先験的に分っていない場合、種々の自然の状態の生起確率は、過去の経験によって決定されなければならないことになろうが、実際の確率決定に当っては種々の困難が生ずることになる。通常は当該の自然の状態の相対頻度によって確率が推定されることになるが、その自然の状態が起きた一般的条件に変化があるのが普通である。たとえば、在庫の決定の問題を考える場合、需要量を自然の状態としよう。過去何期かにわたって需要量を調べれば、そのうち何単位の需要量は全体の何%かはそれぞれ分るであろう。ところで過去何期にまで遡って調べるのが妥当であろうか。単純には、相対頻度の分母にくる期間数が大きければ大きいほど、それぞれの相対頻度は安定するであろうと、推論される。ただしこれは、実験がすべて同質的条件の下で遂行される場合に妥当する立論である。すなわち、在庫の決定の問題における需要量についてみれば、需要量が実現された各期の一般的条件が同質的であるという条件の下に妥当することになる。季節的商品を対象としているときは、取り上げる需要量としては、同じ時期のもの、たとえばこれから適用しようとする時期が第一・四半期であるとすれば、各年の第一・四半期だけがまず問題になるであろう。しかしこれだけでは十分でない。何となれば経済量を対象とする場合、一般経済の循環、成長という条件を考慮しなければならないであろう。したがってこれらの条件を積極的に導入するかまたは除去して考えることが必要になろう。このように考えてくると、取り上げるべき対象期間の数が極めて限定されたものとなり、それを用いて算出した相対頻度の安定性、信頼性が問題となってくるであろう。一企業の内部にだけ限定されない決定の問題については右のごとき配慮が必要である。

たとえ企業の内部条件にだけ左右されるような自然の状態を取り扱う場合においても、相対頻度をもって確率に代える場合には、本質的には同じことがいえよう——程度の差はあるとしても。(10)

自然の状態に対して主観的確率を付与しうる場合には、一旦確率が与えられれば、後の分析は客観的な確率を用い

た場合と同様に進めることができるであろう。[11] 問題は確率をどのように定義し、実際に測定するかにあるといえよう。実際問題として主観的な確率が主として作用する場合は、つぎに述べる不確実性の下でのデシジョン・メーキングの場合であるといえよう。

危険の下での決定規準は、通常、期待効用ないし期待利得を最大にする行動または戦略を選択せよというものである。さらにこれに、期待値が等しい場合には効用ないし利得のちらばりの小さいものを選択せよ、という修正的規準も適用される。

不確実性の下でのデシジョン・メーキングは、自然の状態の各々の確率が、意志決定者に未知の場合の問題である。このような問題は、自然の状態の確率を推定する基礎が過去の経験のなかにないとき常に起るものである。たとえば戦争が起るかどうか、といったことを基にした政治的決定とか経済的決定、また新製品のマーケティングに関連した決定の問題などは、その例になろう。また新製品に関する設備投資の決定などの場合も同様の事態が起るであろう。通常このような場合には、自然の状態についての客観的な確率をうることはできないであろう。信頼度をもとにした主観的確率を付与して決定を行うことも、一部では有力な流れとなっている。[12] この面でのスッペ (P. Suppes) の示唆は、[12] 主観的確率を基礎にした、サヴェージ (L. J. Savage) 流のデシジョン・メーキングの理論に反省を与えるとともに、一つの新しい発展方向を示すものといえよう。

すなわち、スッペは、心理学的見地から、主観的確率および効用の行動的基礎を検討し、従来のサヴェージ流のデシジョン・メーキングについての理論構成のなかで、期待効用最大化の結果について最も望ましくないことは、その静態的性格であるとする。環境が個人とどのように交互作用をするか、すなわち、どのようにして可能な自然の状態について主観的な信頼度をもつに至るのか、またはどのようにして違った可能な結果に対しての相対的な望ましさの

評価をするに至るのかについて説明が企てられていないとする。そこで彼は、最近心理学で新しい成果をあげつつある刺戟学習理論(stimulus learning theory)——これは一九五〇年心理学者のエステス(W. K. Estes)によって最初に導入されたものである——を援用して主観的確率および効用函数の拡張の必要性と方向を示唆する。刺戟学習理論はごく新しい理論であるが、スッペは最近の成果としてアトキンソン(R. C. Atkinson)との共著を出している。(14) 彼等は、そこで数学的行動理論を、ゲームの状況に密接に類似した小グループの実験に適用している。

通常不確実性の下での決定規準としては、ラプラスの規準、ワルドの規準、フールウィッツの規準、サヴェージの規準などがあげられており、(15) 種々論議されているので、ここでは取りあげないことにする。何れにせよ、どの決定規準を採用するかは、結局意志決定者の見地、態度に依存するものとなるであろう。

不確実性の下でのデシジョン・メーキングの過程として、既述のシャックル流の理論は特異な存在ではあるが、実際的な応用を考えた場合、未だ大きな影響力はもっていないということができでよう。たとえば、スペンサー・ジーゲルマン〔I〕、モーリス〔II〕、ミラー・スター〔III〕、モーレル〔IV〕等々、全然シャックルの理論を取り上げていない。

さてデシジョン・メーキングの過程を以上のように定式化した場合、それぞれの段階において、実際にはどのような問題が生ずるか、そしてそれに対してどのように接近すべきか、という問題をOR的な観点から若干考察することにしよう。

まず意志決定の第一段階において目標が定式化されなければならないが、通常、多元的な目標が設定されなければならない場合が少なくはない。目標が単一の場合は別として、目標が多元的な場合には、目標の重要度に応じて順位づけることが必要である。少なくともどれが最も重要で、他は第二義的かを決めることは必要である。多元的な目標が存在する場合には、組織においては個人と個人の目標の間、個人の目標と組織の目標の間、集団の目標と集団の目

標の間、時間的な水平線を考えれば、長期の目標と短期の目標の間の相剋ということが考えられる場合もある。もちろん、これらの目標の間に相剋のないこともある。目標間に相剋のある場合には、部分最適化の問題が生ずることになり、これは OR でも重要な領域を構成している。

ホワイト (C. M. White) は、企業の理論における多元的な目標ないし目的について論じ[16]、広く受け入れられている利潤最大化の目標は、実際には、利潤最大化の代りにか、または利潤とのある加重的組合せにおいての何れかにおいて用いられる、企業の多元的な目標の異質的な集りであるとする。そこでは、企業によって探求される主要な目標を分離、分類し、企業の現代理論における利潤の最大化の正しい位置づけを行っている。

対売上比率として表わされる純利益のある最小水準以上という制約の下での売上量の最大化という目標、また制約条件として純利益の上限と下限を設定する場合の最適化、資産の価値の最大化、短期目標としての財務比率の最大化について、リニヤー・プログラミングの手法を適用しうるよう定式化を行っている。

意志決定者に利用可能な行動または戦略、自然の状態、および競争的戦略を発見し、評価する際にも OR は重要な働きをする。目標を達成するために適当なすべての変数は、不適当なものと区別されなければならない。これらの変数が、戦略、自然の状態、および競争的戦略を形成する成分となる、これらの変数を決定するに際しては、経験と想像力、さらに問題に関する実質科学の知識、分析のための素材、技法等が十分具備されていなければならないであろう。

この外、戦略の選択や成果の決定について OR は重要な働きをするが、ここでは詳細な論議を省くことにする。[17]

(9) 確実性の下でのデシジョン・メーキングについて OR (オペレーションズ・リサーチ) との関連において次書は若干論じている。

D. W. Miller and M. K. Starr [日] chap. 5.

（10） この問題については，ブロス[VI]が，確率の妥当性と鋭さという概念を用いて，極めて示唆に富んだ議論を展開している。
I. D. J. Bross chap. 3.（竹内清訳［VII］第3章）

（11） 心理学的見地から主観的確率を用いて決定問題を取扱ったものとしては，たとえば
L. J. Savage〔X〕

（12） L. J. Savage〔X〕
その外，実験的に効用等を測定したものとして
[XXI] D. Davidson and P. Suppes, *Decision Making: An Experimental Approach*, Stanford Univ. Press, 1957.
主観的確率そのものを実験的に研究したものとしては
[XXII] J. Cohen and M. Hansel, *Risk and Gambliug: The Study of Subjective Probability*, Longmans, Green and Co.. 1959.

（13） 〔XIII〕 P. Suppes, "Some open problems in the foundation of subjective probability," in *Information and Decision Process*, ed. by R. E. Machol, McGraw-Hill Book Co., 1960, pp. 162～169.

（14） 〔XXIV〕 P. Suppes and R.C. Atkinson, *Markov Learning Models for Muliperson Interactions*, Stanford Univ. Press, 1960.

（15） たとえば
M. H. Spencer and Seigelman [I]
W. T. Morris [II]
D. W. Miller and M. K. Starr [III]
[XXV] J. Milnor, "Games against Nature," in *Decesion Processes*. ed. by R. M. Thrall, C.H. Coombs, and R. L. Davis, 1954, pp. 49～59.

（16） [XXVI] C. M. White, "Multiple Goals in the Theory of the Firm," in *Linear Programming and the Theory of the Firm*, by K. E. Boulding and W. A. Sprivey, The Macmillan Co., 1960.

（17） これらの問題に関しては次書参照。
D. W. Miller and M. K. Starr [III], chap. 6, chap. 13.

デシジョン・メーキングの過程において予測体系の果す役割は極めて大きいが、そこでは過去の経験および知識が通常主要な働きをする。しかし時には過去の経験が全然ないかまたは殆んど利用できない決定の状況もある。たとえば、新製品の市場への導入とか、新しい広告媒体の販売促進への利用等といった場合の決定には、過去の経験も余り役立ちえない場合も少なくはないであろう。しかし実際問題としては、ある行動ないし戦略が選択されなければならない。その様な場合、有力な一つの方法は、逐次決定の方法を採用することである。これは基本的には、補助情報を、それをうるためのコストまたは損失を勘案しながら、多段的に決定を改善し、所期の目標を達成しようとするものである。われわれが実際に行う決定は、本質的に逐次決定である場合が少なくはない。

五

ブロス〔VI〕が指摘しているように逐次決定は、逐次的という意味を狭義と広義の二つに分けて考えることができるであろう。(18) 狭義の逐次決定は通常の抜取検査に用いられるものに対応するものである。すなわち、一単位ずつ（または複数単位ずつ）観測がなされ、各観測後に、それまでに知られた情報にもとづいて、最終決定を行うか、または最終決定を行う前に更にもう一単位（または複数単位）観測をとるかを決定する。広義の逐次決定は、二種類の違った選択肢を設定する。すなわち、ブロスのあげている簡単な例をあげると、Aという果物屋でブドウを買うか、またはBという違った果物屋でブドウを買うかという二つの異った種類の選択肢を設定しうるであろう。AなりBなりに限定してしまって、逐次決定を行えば、その場合は狭義の逐次決定になるであろう。

われわれが通常ぶつかる決定の問題は、広義の逐次決定である。

逐次決定の場合には、一般的に確率過程の理論を応用することも考慮しなければならないであろう。今迄述べたデ

シジョン・メーキングの言葉で表現するならば、自然の状態についての確率と、自然の状態が時間によって推移する確率が利用可能である場合、それを利用して動態的に最適な決定を行う問題を、若干考えてみよう。ハワード（R. A. Howard）は、ダイナミックプログラミングとマルコフ過程の問題を極めて明快に論じているので、[19] ここではこれを手がかりとして話を進め、詳細な論議は別の機会にゆずることにする。以下は特に筆者の覚書的なものであることを断っておく。

自然の状態に対応するものは体系の状態とよばれる。これと状態の推移ということがマルコフ過程の基礎概念になる。ここでは離散的な単純なマルコフ過程を考えよう。すなわち、現在体系が状態 i を占めていた場合、つぎの時間間隔の間に状態 j へ推移する確率は、i と j だけの函数であり、i 以前の体系の歴史の函数ではない。すなわち、現在状態 i を占めている体系がそのつぎの推移の後に状態 j を占めるであろう条件付確率 p_{ij} の集合を規定しうるであろう。そうすると、体系はそのつぎの推移の後にある状態になければならないから

$$\sum_{j=1}^{N} p_{ij} = 1$$

ただし、右の式のなかには、体系が状態 i に留まるであろう確率 p_{ii} も含まれている。また p_{ij} は確率であるから

$$0 \leq p_{ij} \leq 1$$

つぎに、n が0のときの状態が既知であるならば、体系が n 回の推移の後に状態 i を占める確率、すなわち、状態確率 $\pi_i(n)$ はつぎのように定義される。

$$\sum_{i=1}^{N} \pi_i(n) = 1 \qquad (1)$$

$$\pi_j(n+1) = \sum_{i=1}^{N} \pi_i(n) p_{ij} \quad n = 0, 1, 2, \ldots\ldots \qquad (2)$$

右の式を行列を使って表わせばつぎのようになる。

$$\boldsymbol{\Pi}(n+1)=\boldsymbol{\Pi}(n)\boldsymbol{P} \quad n=0, 1, 2, \ldots\ldots \tag{3}$$

ただし

$$\boldsymbol{\Pi}(n)=[\pi_1(n)\pi_2(n)\ldots\pi_N(n)]$$

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \ldots & p_{1N} \\ p_{21} & p_{22} & \ldots & p_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N1} & p_{N2} & \ldots & p_{NN} \end{bmatrix}$$

(3) 式からつぎの結果がえられる。

$$\boldsymbol{\Pi}(1)=\boldsymbol{\Pi}(0)\boldsymbol{P}$$

$$\boldsymbol{\Pi}(2)=\boldsymbol{\pi}(1)\boldsymbol{P}=\boldsymbol{\Pi}(0)\boldsymbol{P}^2$$

$$\cdots\cdots\cdots\cdots$$

$$\boldsymbol{\Pi}(n)=\boldsymbol{\Pi}(0)\boldsymbol{P}^n \quad n=0, 1, 2, \ldots\ldots \tag{4}$$

完全なエルゴード的過程は、状態推移の数が大きい場合、すなわち、極限において状態の確率が体系の初期の状態に無関係なマルコフ過程であるが、この場合、π_i を、体系が多数回の推移の後に i 番目の状態を占める確率とする。$\boldsymbol{\Pi}$ を π_i を成分とする行ベクトルとすれば、これは、n が無限に近づいた場合の $\boldsymbol{\Pi}(n)$ の極限である。これは絶対状態確率または極限状態確率のベクトルとよばれる。(3) 式から

$$\boldsymbol{\Pi}=\boldsymbol{\Pi}\boldsymbol{P} \tag{5}$$

また

$$\sum_{i=1}^{N}\pi_i=1 \tag{6}$$

ハワードは、z 変換とよぶ生成函数を用いてマルコフ過程を分析する。

マルコフ過程における確率推移は幾何級数的であるから、幾何級数より速くは n とともに増加しない時間函数に対して

$$f(z)=\sum_{n=0}^{\infty}f(n)z^n \tag{7}$$

なるような変換 $f(z)$ を定義すると有用である。若干の典型的な時間函数の z 変換を示すと、つぎのようになる。

段階函数

$$f(n)=\begin{cases}1 & n=0, 1, 2, 3, \cdots\cdots \\ 0 & n<0\end{cases}$$

の z 変換は

$$f(z)=\sum_{n=0}^{\infty}f(n)z^n=1+z+z^2+z^3+\cdots\cdots=\frac{1}{1-z}$$

幾何級数 $f(n)=\alpha^n$ $(n\geqq0)$ に対しては

$$f(z)=\sum_{n=0}^{\infty}f(n)z^n=\sum_{n=0}^{\infty}(\alpha z)^n=\frac{1}{1-\alpha z}$$

ところで

$$\frac{d}{dz}f(z)=\sum_{n=0}^{\infty}n\alpha^n z^{n-1}$$

であるから

$$\sum_{n=0}^{\infty} n\alpha^n z^n = z\frac{d}{dz}f(z) = z\frac{d}{dz}\left(\frac{1}{1-\alpha z}\right) = \frac{\alpha z}{(1-\alpha z)^2}$$

したがって $f(n)=n\alpha^n$ の z 変換は $\alpha z/(1-\alpha z)^2$ となる。また変換 $f(z)$ をもつ時間函数 $f(n)$ が一単位右へ推移して $f(n+1)$ となると、その z 変換はつぎのようになる。

$$\sum_{n=0}^{\infty} f(n+1)z^n = \sum_{m=1}^{\infty} f(m)z^{m-1} = z^{-1}[f(z)-f(0)]$$

いま (3) 式に z 変換を適用して解いてみよう。ベクトル $\boldsymbol{\Pi}(n)$ のベクトル z 変換を $\boldsymbol{\Pi}(z)$ とすると次式をうる（既述の $f(n+1)$ の z 変換参照）。

$$z^{-1}[\boldsymbol{\Pi}(z)-\boldsymbol{\Pi}(0)] = \boldsymbol{\Pi}(z)\boldsymbol{P} \tag{8}$$

これからつぎの最終結果がえられる。

$$\boldsymbol{\Pi}(z) = \boldsymbol{\Pi}(0)(\boldsymbol{I}-z\boldsymbol{P})^{-1} \tag{9}$$

ただし $\boldsymbol{I}$ は単位行列。推移の問題に対する解は、初期の状態確率で $(\boldsymbol{I}-z\boldsymbol{P})^{-1}$ の行に重みをつけて合計し、結果の各要素の逆変換をとることによってえられる。

簡単な例によって z 変換の利用をみてみよう。

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix}$$

とすれば

$$(\boldsymbol{I}-z\boldsymbol{P})^{-1} = \begin{bmatrix} \dfrac{1-3/5\,z}{(1-z)(1-1/10\,z)} & \dfrac{1/2\,z}{(1-z)(1-1/10\,z)} \\ \dfrac{2/5\,z}{(1-z)(1-1/10\,z)} & \dfrac{1-1/2\,z}{(1-z)(1-1/10\,z)} \end{bmatrix}$$

$$=\begin{bmatrix} \dfrac{4/9}{1-z}+\dfrac{5/9}{1-1/10\,z} & \dfrac{5/9}{1-z}+\dfrac{-5/9}{1-1/10\,z} \\ \dfrac{4/9}{1-z}+\dfrac{-4/9}{1-1/10\,z} & \dfrac{5/9}{1-z}+\dfrac{4/9}{1-1/10\,z} \end{bmatrix}$$

$$=\frac{1}{1-z}\begin{bmatrix} 4/9 & 5/9 \\ 4/9 & 5/9 \end{bmatrix}+\frac{1}{1-1/10\,z}\begin{bmatrix} 5/9 & -5/9 \\ -4/9 & 4/9 \end{bmatrix}$$

$\boldsymbol{H}(n)$ を $(\boldsymbol{I}-z\boldsymbol{P})^{-1}$ の逆変換とすれば

$$\boldsymbol{H}(n)=\begin{bmatrix} 4/9 & 5/9 \\ 4/9 & 5/9 \end{bmatrix}+\left(\frac{1}{10}\right)^{n}\begin{bmatrix} 5/9 & -5/9 \\ -4/9 & 4/9 \end{bmatrix}$$

最後に (9) 式の逆変換をとることにより

$$\boldsymbol{\pi}(n)=\boldsymbol{\pi}(0)\boldsymbol{H}(n) \qquad (10)$$

$\boldsymbol{H}(n)$ がどのような形をとるかみてみよう。それは、その成分行列のなかに、少なくとも、ストカスティック行列であり、そして $1/(1-z)$ の形の $(\boldsymbol{I}-z\boldsymbol{P})^{-1}$ の項から起るものをもつであろう。すなわち、$(\boldsymbol{I}-z\boldsymbol{P})$ の行列式は、$z=1$ に対して 0 となる。もし過程が完全にエルゴード的であれば、$\boldsymbol{H}(n)$ には正確に一つのストカスティック行列が存在しこの行列の行は同じものとなり、各々過程の極限状態確率ベクトルとなるであろう。右の例では

$$\begin{bmatrix} 4/9 & 5/9 \\ 4/9 & 5/9 \end{bmatrix}$$

がそれになる。$\boldsymbol{H}(n)$ のこの部分は、n の函数ではなく、定常部分とよばれる。これを $\boldsymbol{S}$ で表わす。

$\boldsymbol{H}(n)$ の残りの項は、過程の一時的動きを表わす。この部分は n の函数であり、$\boldsymbol{T}(n)$ で表わす。この項は、α^n, $n\alpha^n$, $n^2\alpha^n$ 等の係数で乗ぜられた行列であるが、完全なエルゴード的過程では、すべての α に対して $|\alpha|>1$ であるか

ら、nが極めて大きくなるにつれて一時的成分 $T(n)$ は消失する。この成分は行に関してたして 0 にならなければならない。このような行列は微分行列とよばれる。かくして、完全なエルゴード的過程に対しては

$$H(n)=S+T(n) \tag{11}$$

マルコフ過程について一層の分析を進めるために、更に新たな概念が定義される。完全なエルゴード的過程においても、ある状態は極限状態確率 0 をもつことはありうるであろう。そのような状態は、一時的状態とよばれる。また $p_{ii}=1$ なる状態 i は trapping state とよばれる。

体系は、一時的状態を残しておいて、可能な推移によって結合される一組の状態に入りうるであろう——体系はこの一組の状態の内部で無期限に跳ぶが、この組の外部へは決して跳ばないような仕方で。このような状態の組は、マルコフ過程の再帰連鎖とよばれる。

各々のマルコフ過程は少なくとも一つの再帰連鎖をもつ。ただ一つの再帰連鎖をもつマルコフ過程は完全なエルゴード的過程である。過程が、二つまたはそれ以上の再帰連鎖をもてば、完全なエルゴード的過程の性質は成り立たなくなる。多くの再帰連鎖をもつ場合には、極限状態確率分布は、体系がどのようにして出発したかに依存するので、ストカスティック行列 S の行はもはや等しくはない。S の i 番目の行は、体系が i 番目の状態で出発した場合に存在するであろう極限状態確率分布を表わすことになる。$T(n)$ の i 番目の行は、前と同様に、i が出発の状態である場合の、状態確率の一時的成分の集合を表わす。

多重連鎖のマルコフ過程も、z変換によって容易に取扱うことができる。

ハワードは、離散的時間および連続的時間のマルコフ過程を取扱うが、前者ではZ変換が強力な分析技術を提供する。ここでは離散的な体系だけを考えることにする。離散的な体系で、どの選択肢または政策が最大の総期待報酬を生みだすであろうか、という問題に対して、もし右の最大化が過程の二、三の段階にだけわたってのものであれば、価値繰返し法が望ましい。他方過程が無限に持続するものと期待されるならば、政策繰返し法が望ましいものとなる。以下でこれをみることにしよう。

さて N 状態のマルコフ過程が、状態 i から状態 j に推移するとき、r_{ij} という報酬を生成するものとする。これは一般的には、われわれがすでに述べた効用とか望ましさとか、あるいは満足といったもので表現されるわけである。もちろん、問題によっては貨幣金額と解してもよいであろう。r_{ij} を要素とする行列 $\boldsymbol{R}$ は報酬行列とよばれる。かくして報酬は、マルコフ過程の確率的関係によって支配される確率分布をもった確率変数となる。このような状態は、ゲームの理論の適用される場になぞらえることができるであろう。すなわち、プレーヤーは、状態 i から状態 j に推移すれば、報酬 r_{ij} をうけとるというゲームを想定するこができるであろう。もちろん、r_{ij} は負にもなりうるであろうから、そのときは r_{ij} を支払わなければならなくなる。

そこでこのゲームに関して、もしいま体系が状態 i にあるとして、つぎの n 回の推移でプレーヤーがうると期待される期待報酬はいくらになるか、という問が提出される。これに対する解をうるために、体系がいま状態 i にあるとして、つぎの回の推移での期待総収入を $v_i(n)$ とする。そうすると循環的関係としてつぎの式がえられる。

$$v_i(n)=\sum_{j=1}^{N}p_{ij}[r_{ij}+v_j(n-1)] \qquad i=1,2,\ldots\ldots,N \qquad (12)$$

$$n=1,2,3,\ldots\ldots$$

そこで

$$q_i=\sum_{j=1}^{N}p_{ij}r_{ij} \qquad i=1,2,\dots\dots,N \tag{13}$$

とすれば

$$v_i(n)=q_i+\sum_{j=1}^{N}p_{ij}v_j(n-1) \qquad i=1,2,\dots\dots,N \tag{14}$$
$$n=1,2,3,\dots\dots$$

右の式で q_i は、状態 i についての直接報酬期待値とよばれる。右の式をベクトルの形で表わせば

$$\boldsymbol{V}(n)=\boldsymbol{q}+\boldsymbol{P}\boldsymbol{V}(n-1) \qquad n=1,2,3,\dots\dots \tag{15}$$

ただし $\boldsymbol{V}(n)$ は、総価値ベクトルとよばれる N 個の要素 $\boldsymbol{V}_i(n)$ をもった列ベクトルである。体系の収入期待値を決定するためには、$\boldsymbol{P}$ および $\boldsymbol{q}$ が規定されればよいことになる。

報酬をもったマルコフ過程は、z 変換によって分析されるが、総価値ベクトル $\boldsymbol{V}(n)$ の z 変換は $v(z)$ とよばれ

$$v(z)=\sum_{n=0}^{\infty}\boldsymbol{V}(n)z^n \tag{16}$$

(15) 式の変形したものに z 変換を施すと

$$z^{-1}[v(z)-\boldsymbol{V}(0)]=\frac{1}{1-z}\boldsymbol{q}+\boldsymbol{P}v(z)$$

$$\therefore v(z)=\frac{z}{1-z}(\boldsymbol{I}-z\boldsymbol{P})^{-1}\boldsymbol{q}+(\boldsymbol{I}-z\boldsymbol{P})^{-1}\boldsymbol{V}(0) \tag{17}$$

ただし $\boldsymbol{I}$ は単位行列。

（17）式で $\boldsymbol{V}(0)$ が一義的に 0 であると、同式はつぎのようになる。

$$\boldsymbol{v}(z)=\frac{z}{1-z}(\boldsymbol{I}-z\boldsymbol{P})^{-1}\boldsymbol{q} \tag{18}$$

$[z/(1-z)](\boldsymbol{I}-z\boldsymbol{P})^{-1}$ の逆変換を用いて $\boldsymbol{V}(n)$ は求めることができる。

長期間持続する過程の総収入期待値は、一般にどうなるか、その漸近的な動きをみてみよう。一応（17）式に着目する。さきにみたように $(\boldsymbol{I}-z\boldsymbol{P})^{-1}$ の逆変換は $\boldsymbol{S}+\boldsymbol{T}(n)$ の形をとる。そこでその関係を利用して、つぎの関係をうる。

$$(\boldsymbol{I}-z\boldsymbol{P})^{-1}=\frac{1}{1-z}\boldsymbol{S}+\mathscr{T}(z) \tag{19}$$

ただし $\mathscr{T}(z)$ は $\boldsymbol{T}(n)$ の z 変換。（19）を（17）に代入して、つぎの結果がえられる。

$$\boldsymbol{v}(z)=\frac{z}{(1-z)^2}\boldsymbol{S}\boldsymbol{q}+\frac{z}{1-z}\mathscr{T}(z)\boldsymbol{q}+\frac{1}{1-z}\boldsymbol{S}\boldsymbol{V}(0)+\mathscr{T}(z)\boldsymbol{V}(0) \tag{20}$$

ところで、右の式の右辺の各項を吟味することにより、大きな n に対しての $\boldsymbol{V}(n)$ の漸近的な形はつぎのようになることが分る。

$$\boldsymbol{V}(n)=n\boldsymbol{S}\boldsymbol{q}+\mathscr{T}(1)\boldsymbol{q}+\boldsymbol{S}\boldsymbol{V}(0) \tag{20}$$

成分 g_i をもった列ベクトル $\boldsymbol{g}$ を、$\boldsymbol{g}=\boldsymbol{S}\boldsymbol{q}$ と定義すれば

$$\boldsymbol{V}(n)=n\boldsymbol{g}+\mathscr{T}(1)\boldsymbol{q}+\boldsymbol{S}\boldsymbol{V}(0) \tag{22}$$

g_i は直接的な報酬 q_i を、体系が i 番目の状態で出発した場合にもたらされる極限状態確率 s_{ij} によって加重し加えたものである。すなわち

$$g_i=\sum_{j=1}^{N}s_{ij}q_j$$

g_i は体系が i 番目の状態で出発し、何回も推移を行なった場合の、体系の推移当り平均収益となる。これを i 番目の状態の利益とよぶ。それは $v_i(n)$ の漸近線の傾斜でもある。完全なエルゴード的過程においては、すべての状態は同じ利益 $\boldsymbol{g}$ をもち、つぎのようになる。

$$g=\sum_{i=1}^{N}\pi_i q_i \qquad (23)$$

(22) 式の列ベクトル $\mathcal{T}(1)\boldsymbol{q}+\boldsymbol{S}\boldsymbol{V}(0)$ は、$n=0$ における $\boldsymbol{V}(n)$ の漸近線の截片を表わす。$v_i(n)$ の漸近的截片を v_i で表わすと、大きな n に対しては

$$v_i(n)=ng_i+v_i \quad i=1,2,\ldots\ldots,N \qquad (24)$$

成分 v_i をもつ列ベクトルを $\boldsymbol{V}$ で表わすと、$\boldsymbol{V}=\boldsymbol{T}(1)\boldsymbol{q}+\boldsymbol{S}\boldsymbol{V}(0)$ となる。そうすると (24) 式は大きな n に対してつぎのようになる。

$$\boldsymbol{V}(n)=n\boldsymbol{g}+\boldsymbol{V} \qquad (25)$$

体系が完全にエルゴード的であれば、すべての g_i は g に等しいが、この g を過程の利益とよぶ。したがって (24) 式はつぎのようになる。すなわち、大きな n に対して

$$v_i(n)=ng+v_i \quad i=1,2,\ldots\ldots,N \qquad (26)$$

以上でマルコフ的な逐次過程における決定の分析をするための基本的な道具だてがすんだ。そこでつぎに循環的な関係を用い、価値繰返し法によって逐次決定過程の解が求められる。

まず $d_i(n)$ を、段階 n で用いられるであろうところの、i 番目の状態における選択肢の数としよう。これは、n 番目の段階での状態 i における決定とよばれる。すべての i およびすべての n に対して $d_i(n)$ が規定されると、政策が決定される。最適政策は、各 i および n に対して総収入期待値を最大にするものとされる。

いま $v_i(n)$ を、最適政策がとられる場合の、状態 i から出発して n 段階における総収入期待値と改めて定義しなおすと、任意の n に対して、次式が導かれる。

$$v_i(n+1)=\max_k\sum_{j=1}^{N}p_{ij}{}^k[r_{ij}{}^k+v_j(n)] \quad n=0, 1, 2, \ldots\ldots \tag{27}$$

ただし上つきの添字 k は状態における選択肢を示す。すなわち、選択肢 k をとった場合、状態 i から状態 j への推移は、確率 $p_{ij}{}^k$ によって支配される。これらの推移に関連した報酬は $r_{ij}{}^k$ で示される。(27) 式は価値繰返し方程式とよばれる。(27) 式はつぎのように変形できる。

$$v_i(n+1)=\max_k\left[q_i{}^k+\sum_{j=1}^{N}p_{ij}{}^kv_j(n)\right] \tag{28}$$

ただし $q_i{}^k=\sum_{j=1}^{N}p_{ij}{}^kr_{ij}{}^k$。

(28) 式の逐次関係を用いることにより、意志決定者は、各段階で各状態においてどの選択肢を用いるべきかが分り、また過程の各段階で将来の収入の期待値が与えられる。なおこの関係を適用するためには、過程に対しての境き条件 $v_j(0)$ を規定しなければならない。

以上の逐次過程の解法が価値繰返し法とよばれるものである。この方法はいくつかの制約をもっていることが指摘

される。すなわち、通常体系は、明確に限定される終点をもたない基礎の上に働いている。終りがはるかに遠くにあるような十分に大きな n をもつまで、$v_i(0)$ を繰返し計算していくのは効率的ではないようにみえる。そのような場合には、終りに至るまでに多くの推移をなすような過程を分析する技法が必要となるわけであるが、それはつぎに展開される。

価値繰返し法は、長期持続する過程には適していないが、かなり短時間で終結するような体系には適当な方法である。

七

つぎに推移行列 $\boldsymbol{P}$ および報酬行列 $\boldsymbol{R}$ によって記述される完全なエルゴード的過程を考える。この過程は、長時間にわたって推移をなし、われわれは過程の収入に関心をもつものと想定する。体系が完全なエルゴード的過程であるから、極限状態確率 π_i は初期の状態とは独立で、体系の利益 g は、次式で表わされる。

$$g=\sum_{i=1}^{N}\pi_i q_i \tag{23}$$

ただし q_i は (13) で定義されたもので、直接収入期待値である。

報酬をもった各完全なエルゴード的過程は、(23) で与えられる利益をもつであろう。もしそのような過程をいくつかもち、長期的にはどれが最も利益があがるものであるかを知ろうとするならば、各々の利益を見出し、そして最も高い利益をもったものを選び出しうるであろう。

N 個の状態があり、各状態にはいくつかの選択肢があるものとする。選択される選択肢は、その状態についての

決定とよばれる。すべての状態についての決定の集合が政策とよばれる。最適政策は利益または推移当りの平均収入を最大にする政策と定義される。（ここではすべての政策は完全なエルゴード的過程をうみ出すものと想定されるが、ハワードは後でこの仮定をゆるめている。）N 個の状態の問題では、各状態においてもたれる選択肢の数をかけ合せただけの政策があることになる。i 番目の状態に n_i 個の選択肢があるものとすれば、$n_1 \times n_2 \times \cdots\cdots \times n_N$ 個の異った政策があることになる。理論的には最大の利益をもった政策を見出すために、これらの政策の各々について利益を見出しうるであろうが、実際問題としては面倒なことになることも少なくはない。

政策繰返し法は、少数の繰返しで最適政策を見出す方法を提供するが、これは二つの部分から成っている。すなわち、価値決定オペレーションと政策改善ルーチンの二つがそれである。

まず価値決定オペレーションについて述べよう。所与の政策の下に体系を動かしているものと想定しよう。したがって報酬をもった所与のマルコフ過程を規定したものとする。もしこの過程が n 個の段階または推移に対して働くものとされれば、$v_i(n)$ を、もし体系が所与の政策の下に状態 i から出発する場合 n 回の手番において体系が稼ぎ出すであろう総報酬期待値と定義しうるであろう。

(13) 式から

$$v_i(n)=q_i+\sum_{j=1}^{N} p_{ij}v_j(n-1) \quad i=1,2,\ldots\ldots,N \quad n=1,2,\ldots\ldots \tag{13}$$

また完全なエルゴード的過程に対しては $v_i(n)$ はつぎの漸近的な形をもつこともすでにみた。すなわち、大きな n に対して

$$v_i(n)=ng+v_i \quad i=1,2,\ldots\ldots,N \tag{24}$$

ここでは、極めて多数の段階をもった体系だけを考えているので、(24) を用いてもよいことになり、つぎの結果がえられる。

$$ng+v_i=q_i+\sum_{j=1}^{N}p_{ij}[(n-1)g+v_j] \quad i=1,2,\ldots\ldots,N$$

$$ng+v_i=q_i+(n-1)g\sum_{j=1}^{N}p_{ij}+\sum_{j=1}^{N}p_{ij}v_j$$

ところで $\sum_{j=1}^{N}p_{ij}=1$ であるから、右の式はつぎのようになる。

$$g+v_i=q_i+\sum_{j=1}^{N}p_{ij}v_j \quad i=1,2,\ldots\ldots,N \tag{29}$$

右は N 個の連立一次方程式であるが、ここに未知数は N 個の v_i と 1 個の g で合計 $(N+1)$ 個存在する。したがって v_i は一般に一義的には決定されないことになる。しかしもし v_i のうちの一つ、たとえば、v_N を 0 とおけば、N 個の未知数になり、g および、0 とおいた以外の v_i について解くことが可能となる。そのようにしてえられた v_i は、(24) で定義されたものではないであろうが、それとは定数の大きさだけ異なるであろう。しかし v_i の真の値は定数項 $\sum_{i=1}^{N}v_i(0)\pi_i$ を含む ((22)式参照) ので、極めて多数の推移に対して続く過程においては、それらは実際には大した意味をもたない。$v_N=0$ とおいて (29) を解いてえられる v_i は、ここでの目的にとっては十分であろう。これらは政策の相対値とよばれる。

(29)に i 番目の状態の極限状態確率 π_i をかけ、i について加えると、次式がえられる。

$$g\sum_{i=1}^{N}\pi_i+\sum_{i=1}^{N}\pi_i v_i=\sum_{i=1}^{N}\pi_i q_i+\sum_{j=1}^{N}\sum_{i=1}^{N}\pi_i p_{ij}v_j$$

この式は、次式と同値である（（6）および（5）から）。

$$g=\sum_{i=1}^{N}\pi_i q_i \tag{23}$$

所与の政策の利益のみを求めている場合に、（23）式よりは（29）式を使う理由、また相対値のようなものを見出す理由は、つぎのように説明される。まず（29）式を使うのは、（23）式によって過程の利益は見出しうるが、よりよい政策をどのようにして見出すかということが分らないからである。ところで既述の相対値は、よりよい政策、そして終局的には最良の政策を見出す鍵になっているのである。（29）式は政策改善に必要な相対値をうみ出すのである。

つぎに政策改善ルーチンをみてみよう。もし段階nまでに最適政策をもっていれば、i番目の状態においてすべての選択肢にわたって

$$q_i^k+\sum_{j=1}^{N}p_{ij}^k v_j(n) \tag{30}$$

を最大化することによって、段階（$n+1$）でi番目の状態における最良の選択肢を見出すことができるであろう。大きなnに対しては、（25）を代入することにより、各状態において最大化されるべき試験量として次式をうる。

$$q_i^k+\sum_{j=1}^{N}p_{ij}^k(ng+v_j) \tag{31}$$

$\sum_{j=1}^{N}p_{ij}^k=1$ であるから、右の式のなかで、ng および v_j のなかの付加的な定数部分は、k とは独立なものとなる。したがって、状態iにおいて決定をする場合、i番目の状態における選択肢に関して

$$q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j \qquad (32)$$

を最大化することができる。

政策改善ルーチンはつぎのように要約される。各状態iに対して、旧い政策の下で決定された相対値を用いて、試験量

$$q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j$$

を最大にする選択肢kを見出す。このkがいまやi番目の状態における決定 d_i となる。各状態についてこの手続が遂行された場合、新しい政策が決定されてくる。新しい政策は、旧い政策よりも高い利益をもつであろう。

価値決定オペレーションと政策改善ルーチンは、その目標が、すべての可能な政策のなかで最高の利益をもつ政策を発見することである、繰返しのサイクルで結合される。

価値決定オペレーションは、政策の函数として価値をうみ出し、一方政策改善ルーチンは、価値の函数として政策をうみ出す。

繰返しのサイクルをいつ止めたらよいかの判定基準は、つぎのごときもので簡単である。すなわち、最適政策は、二つ続いた繰返しで政策が同じであった場合に、達成される。

政策繰返し法は、要約すると、つぎの性質をもつ。

（1） 逐次決定過程の解は、連立一次方程式の組を解き、ついで比較をすることに帰せられる。

（2） 繰返しのサイクルで見出された各々相続く政策は、前のものより高い利益をもつ。

（3） 繰返しのサイクルは、問題の領域内で達成可能な最高の利益をもつ政策で終結するであろう。

以上がハワードの理論の基本線である。前節で述べた政策繰返し法は、第五章において、タクシーの運用、野球、自動車の取替問題に適用が試みられており、実際への橋渡しも考慮される。第六章では、既述の完全なエルゴード的過程——各政策は単一の再帰連鎖をもったマルコフ過程であり、一義的な利益をもつ——から、複数個の多重連鎖をもった過程へ、政策繰返し法が試みられる。若干の困難さはあっても、本質的には、完全なエルゴード的過程の場合と同じ接近法である。価値決定オペレーションと政策改善ルーチンが、繰返しのサイクルのなかで統合される。

さらに第七章では、利子率で割引いた場合の逐次決定過程を、第八章では、ラプラス変換を用いて連続的時間の決定過程を取扱っているが、本質的には、いままで述べた方法と同じ行き方をとる。両者は極めて類似している。

以上のごとき確率過程の理論を実際に適用するに当っては、推移確率等を実際に推定しうるだけの十分の時系列データがえられるかどうか、また体系の選択肢を記述するのに必要なデータが利用可能かどうか、というような問題が解決されなければならない。他方このような強力な分析技法に適合するようなデータを、積極的に実査なり実験を通じて確保していくことが必要である

（18） I. D. J. Bross [VI] Chap. 8.（竹内清訳 [VII] 第八章）

（19） R. A. Howard, [V]