

氏名	田 畑 吉 雄 た ばた よし お
学位の種類	工 学 博 士
学位記番号	工 博 第 249 号
学位授与の日付	昭 和 46 年 3 月 23 日
学位授与の要件	学 位 規 則 第 5 条 第 1 項 該 当
研究科・専攻	工 学 研 究 科 数 理 工 学 専 攻
学位論文題目	<b>Studies on Markorian Sequential Decision Processes</b> (マルコフ逐次決定過程に関する研究)
論文調査委員	(主 査) 教 授 三 根 久 教 授 奥 川 光 太 郎 教 授 池 田 峰 夫

### 論 文 内 容 の 要 旨

最近、オペレーションズ・リサーチやシステム工学などの分野においては、決定すべき多数の要因を含んだ複雑で大規模なシステムの研究の重要性が指摘されている。このようなシステムのモデルとしてはマルコフ過程に基礎をおくものが有力であり、幾多の研究が行なわれているが、とくに、マルコフ逐次決定過程は代表的な研究接近法となっている。本論文は、マルコフ逐次決定過程に関する研究成果をまとめたものであって、緒論を含めて7章および結論からなり、前半では等間隔の時点で決定が行なわれる場合を論じ、後半では、任意の時点で決定が行なわれるモデルを扱っている。

第1章は、本論文の緒論であり、取替問題、生産在庫管理問題など確率的考察を必要とする多くの工学上の問題を解析する場合、マルコフ逐次決定過程の問題に定式化できることを示している。ついで、マルコフ決定過程の数学的構造を説明し、これに関連した研究の概要を述べている。

第2章では、まず以下の章で共通に使用する記号とモデルを説明し、在来の最適政策と定常な最適政策の求め方として、ハワードの政策反復法を紹介している。これまでに定義された最適政策が総期待利得の差にもとづいているのに対して、本論文では比にもとづいて、新しい最適政策を提案し、ここに得られたR-最適政策と在来の最適政策との関連性を論じ、これにより割引率のある場合とない場合とに対して統一的にマルコフ決定過程を取り扱うことを試みている。さらに、定常なR-最適政策の存在性について論じ、その求め方として、在来の政策反復法、数理計画法の拡張をはかっている。

第3章では、マルコフ決定過程における総期待利得が割引率に関するべき級数になることに着目し、在来の研究者が提案した各種の最適政策すなわち、 $\beta$ -最適政策、1-最適政策、平均利得最適政策、平均累積利得最適政策について統一的解釈を試みたものであり、1-最適政策、平均利得最適政策が存在するための必要十分条件について論じている。

第4章では、マルコフ決定過程の問題の線形計画による定式化を試みている。在来の結果を発展させる準備として、ベクトルと行列の直和について考察している。ついで、マルコフ決定過程の線形計画問題へ

の定式化およびその双対問題の誘導を試み、さらに分解原理を適用することによって、問題を直和形式で表現する方法を論じている。さらに、もとのマルコフ決定過程と直和形式との関係を調べ、問題が直和形式で表現された場合の最適政策を求めるための手続きの有効性を検討している。

第5章は、決定時点が任意な場合、すなわち時間パラメータが連続な連続時間マルコフ決定過程の線形計画への定式化について論じている。モデルと記号とについて説明した後、連続な問題を等時間間隔の問題に帰着させるための新しい変換法を考察し、ついで、割引率のある場合とない場合について、最適政策がその変換によって不変であることを示している。また、変換の結果得られた等時間間隔の問題に対して、線形計画による定式化を行ない、最適政策を求める手続きとその数値例とを与えている。

第6章では、連続時間マルコフ決定過程における最適政策の集合について考察している。まず、割引率のある場合に対し、 $\alpha$ -最適政策の集合の幾つかの新しい性質を論じ、離散的な場合の $\beta$ -最適政策の集合と類似の性質が連続時間問題にも存在することを確かめ、つぎに、 $\alpha$ -最適領域の決定法を論じ、 $\alpha$ -最適政策の個数の上限と領域の境界値の計算方法を与えている。さらに、割引率のない場合に対し、平均利得最適政策の集合の性質を調べ、他の最適政策の集合との関連を論じている。

第7章では、第2章の結果の連続時間マルコフ決定過程に対する拡張を試み、 $R$ -最適政策の定義とその性質について、類似な結果がえられることを示している。

結論では、以上の研究結果を要約しており、オペレーションズ・リサーチ、統計学などとの関連性、工学上の応用分野および今後研究すべき問題点を論じている。

## 論文審査の結果の要旨

計画工学、システム工学などにおいては、構成が大規模かつ複雑で、確率的な挙動を示すシステムの解析および構成法の確立がつよく望まれている。本論文はこのようなシステムに対する有力な技法であるマルコフ逐次決定過程に関し、意志決定のための最適政策の基本的性質を解明し、在来の理論の統一を行ない、最適政策を求めるためのアルゴリズムを提案したものである。

まずマルコフ決定過程で最も重要な最適政策に対して新たに $R$ -最適政策なる概念を導入し、これが在来の最適政策を含むより広い概念であることを示している。従来、マルコフ決定過程においては割引率のある場合とない場合とでは、形式的にも全く異なった定義によって最適政策が定められていたが、新しい定義によって、統一的にしかも同一形式で論じること成功している。さらに、この定義はこれまで求めることができなかつた最適政策を定められるだけでなく、直観的な概念を十分反映しているという利点も持っている。また、在来の政策反復法および数値計画法のいずれを用いても、方法を少し修正するだけで定常な $R$ -最適政策を簡単に求められることを示している。

ついで、マルコフ決定過程における在来の種々の最適政策、すなわち $\beta$ -最適、 $1$ -最適、平均利得最適および平均累積利得最適の政策が、総期待利得を割引率に関するべき級数と見なすことにより、級数論の見地から統一的に解釈できることを明らかにしている。在来の研究では、上述の最適政策が存在するものとして議論されていたのに対し、級数論を適用することによって、それらが存在するための必要十分条件を導いている。この方法と結果とは、マルコフ決定過程における最適政策を理論的に厳密に考察する場

合に有力な根拠を与えるものである。

さらに、マルコフ決定過程の線形計画問題への定式化について考察している。在来の結果を利用してその双対問題を作り、分解原理を適用することによってもとのマルコフ決定過程がベクトルと行列の直和形式で表現できることを示している。直和形式による表現は、実際上の大規模な問題を解く場合に、計算機の記憶容量の節減に有効である。

論文の後半では、時間パラメータが連続な連続時間マルコフ決定過程を考察している。この種のモデルは、取替理論、待合せ理論、制御理論などの分野によく現われるものである。

まず、連続時間マルコフ決定過程の線形計画への定式化を行なっている。その方法として、新しく連続問題から等時間間隔問題への変換法を確立し、最適政策がこの変換のもとで不変になることを確証した後、変換の結果得られた等時間間隔問題を線形計画に定式化することを提案している。在来の方法がラプラス・スチルチェス変換を用いた複雑なものであったのに比べ、この方法は数学的手法としてはるかに簡単で有効であり、連続時間マルコフ決定過程の線形計画への定式化を、初めて明確な形で与えることに成功している。

つぎに、連続時間マルコフ決定過程において最適政策の集合の構造を考察している。割引率のある場合に、 $\alpha$ -最適政策の集合の性質と構造のいくつかを解明し、 $\alpha$ -最適領域の個数の上限とその決定方法とを与えている。連続的な場合にも等時間間隔問題と類似の性質があることを確かめ、さらに、連続時間問題特有の性質として  $R$ -最適領域の個数の上限を与えられることを示している。また、割引率のない場合には、平均利得最適政策の集合の性質を明らかにしている。

また、等時間間隔問題で与えた  $R$ -最適政策が、連続時間問題にも同一の形式で定義でき、類似の結果がえられることを示している。

以上、要するに本論文は、マルコフ決定過程について詳細な検討を行ない、最適政策の数学的な構造を理論的に解明し、実際問題への適用に際し、最適政策を求めるための有力なアルゴリズムを確立し、それと数理計画との関連性を明らかにしたものであって、實際上、学術上寄与するところが大きい。

よって、本論文は工学博士の学位論文として価値あるものと認める。