氏名	佐藤光男（さとう みつお）
授与学位	工学博士
学位授与年月日	昭和52年3月25日
学位授与の根拠法規	学位規則第5条第1項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	**推定を伴うマルコフ決定過程に関する研究**
指導教官	東北大学教授 竹田 宏
論文審査委員	東北大学教授 竹田 宏　東北大学教授 木村 正行 東北大学教授 野口 正一　東北大学助教授 阿部 健一

論文内容要旨

第1章 緒言

マルコフ過程の適応制御問題の定式化として代表的なものの一つにベイズ的定式化があり，その最適化問題はいわゆるベルマン形方程式を解くことに帰着される。しかしながら，これまでのところ，ベルマン形方程式によって定まるベイズ最適政策の有効性—特に二元性と漸近最適性—に関する理論的解明はほとんどなされていない。

本論文では，マルコフ過程の適応制御方式を構成するという立場から，このベイズ的定式化による制御則について検討する。すなわちベイズ最適政策の有効性の問題を中心に議論を進める。更に，その考察結果に基づいて，確率的な政策を用いた新たな適応制御方式を提案する。

第2章　マルコフ過程の適応制御問題

本章では，これまでに研究がなされてきたマルコフ過程の適応制御の概要を説明し，合わせて本論文における問題の設定を行うものである。まずベイズ的定式化が説明され，次に遷移確率があるあいまいさをもって知られている場合のマルコフ決定問題において考慮されるマックスーマックス最適政策などが論じられる。

本論文で扱うマルコフ決定過程は以下の記号によって記述される。

$\Omega = \{ 1 \cdots\cdots N \}$：状態空間

$\Gamma_i = \{ 1 \cdots\cdots K_i \}$：状態 i のもとでの決定空間

P_{ij}^{k}：状態 i において決定 k を下したときに状態 j に遷移する確率

r_{ij}^{k}：上記の遷移に対して得られる利得（$|r_{ij}^{k}| < \infty$）

β：利得に対する一期間当たりの割引き率（$0 \leqq \beta < 1$）

また，政策は，$u_i \in \Gamma_i$ を状態 i において下すべく与えられた決定として，N次元ベクトル $U = (u_1, \cdots, u_N)$ で表わされる。v_i^{U} を政策Uを定常的に用いるときの初期状態 i における総期待割引き利得とするとき，v_i^{U} を最大ならしめる $U = U_{max}$ は定常最適政策と呼ばれる。

ベクトル $P_i^{k} = (P_{i1}^{k}, \cdots, P_{iN}^{k})$ とし，R_i^{k} を P_i^{k} の閉凸集合あるいは有限集合とする。更に，行列 $P = \{ P \mid P_i^{k} \in R_i^{k} \ (k = 1 \cdots K_{i;i} \ = 1 \cdots N) \}$

このとき方程式

$$\bar{v}_i = \max_{k \in \Gamma_i} \max_{P \in R} \{ \sum_j P_{ij}^{k} r_{ij}^{k} + \beta \sum_j P_{ij}^{k} \bar{v}_j \} \quad (i = 1 \cdots N)$$

によって定まる政策を R_i に基づいたマックスーマックス最適政策と呼ぶ。本章ではこの方程式の近似解法が示されている。

第3章　ベイズ最適政策による適応制御方式

本章は，ベルマン形方程式によって与えられるベイズ最適解の諸性質を論じ，それらを用いてベイズ最適政策の有効性を二元性と漸近最適性の両面から検討する。ただし，本章では，遷移確率行列は未知パラメータ $\theta \in \Theta$ に支配されているものとする。すなわち $P_{ij}^{k} = P_{ij}^{k}(\theta)$ であり，パラメータ空間 Θ は有限次元ユークリッド空間における有界閉集合あるいは有限集合であるとする。

ベイズ的定式化のもとでは，θ の事前確率密度（あるいは事前確率）$\Pi(\theta)$ が既知であることが仮定される。いま決定 k を下して状態 i から状態 j に遷移したとすると，$\Pi(\theta)$ は，ベイズ則によって，事後確率密度 $T_{ij}^{k}\Pi(\theta)$ に修正される。$v_i(\Pi)$ を初期状態 i，事前確率密度 Π に対する最大期待割引き利得とすると，$v_i(\Pi)$ は次式で与えられるベルマン形方程式を満足しなければならない。

$$v_i(\Pi) = \max_{k \in \Gamma_i} \{ \sum_j \bar{P}_{ij}^{k}(\Pi) r_{ij}^{k} + \beta \sum_j \bar{P}_{ij}^{k}(\Pi) v_j(T_{ij}^{k}\Pi) \}$$

$$(i = 1, \cdots, N \ ; \ \Pi \in \Lambda)$$

ここに，$\bar{P}_{ij}^{k}(\Pi)=E(P_{ij}^{k}|\Pi)$であり，$\Lambda$は確率密度関数$\Pi$から成る空間を表わす。特に$\Theta=\{\theta_1, \cdots, \theta_L\}$と表わされるときには，$\phi=(\phi_1, \cdots, \phi_L)$を$\Pi(\theta_\ell)=\phi_\ell$（$\ell=1, \cdots, L$）なる確率ベクトルとして，上式のかわりに次のベルマン形方程式を得る。

$$v_i(\phi)=\max_{k\in\Gamma_i}\{\sum_j \bar{P}_{ij}^{k}(\phi)r_{ij}^{k}+\beta\sum_j \bar{P}_{ij}^{k}(\phi)v_j(T_{ij}^{k}\phi)\}$$

$$(i=1, \cdots, N;\ \phi\in\Psi)$$

すなわちベイズ最適解は確率ベクトルの関数となる。

本章では，まず，ベイズ最適解の諸性質が論じられる。すなわち，上記方程式の解の存在性と唯一性が示され，それからその解の連続性や凸性などが明らかにされる。

次に上記方程式によって定まるベイズ最適政策の漸近最適性の問題が考察される。ベイズ最適政策はただ一つの政策に収束することが示され，従って問題はその極限政策が最適であるか否かということに帰着される。このことはパラメータ空間Θの構造と深いかかわりがある。そこで，Θの構造を規定するために，分解可能性というものが次のように定義される。

〔分解可能性〕すべての$k\in\Gamma_i$，$i\in\Omega$に対して$P_i^k(\theta)\neq P_i^k(\theta')$（$\theta\neq\theta'$）であるとき，パラメータ空間$\Theta$は分解可能であるという。

そして漸近最適性と分解可能性との関係が調べられる。その結果，Θが分解可能である場合にはベイズ最適政策の漸近最適性が保証されるが，Θが分解可能でない場合にはそれが必ずしも保証されないということが示される。更に漸近最適性が保証されない例が示される。

二元性の問題については，それがベイズ最適政策に具体的にどのような形で反映しているかということが，簡単な例において考察される。

本章は，更に，以上の考察結果に基づいて漸近最適な制御則の構成を試みる。ベイズ最適政策の欠陥は，ここで指摘されているように，ただ一つの政策が無限回用いられることにある。そこで，あらゆる政策が無限回用いられるように，政策を確率的に使用するという方法が採用される。すなわち，推定信頼度を表わす数λが定義され，ベイズ最適政策をλなる確率で使用するという制御則が提案される。そして，この制御則は，最適政策の使用確率が1に収束するという意味で漸近最適であることが示される。

第4章　ベルマン形方程式の近似解法

本章ではベルマン形方程式の一近似解法が提案される。これまでに提案されている解法がタイムステージにおける逐次近似から成り立っているのに対して，ここに提案する解法は，ΛあるいはΨの離散化近似，すなわちΛ，Ψの離散化によって前記ベルマン形方程式を有限状態マルコフ決定問題の方程式（この解法は既に確立されている）に変換するという操作から成り立つ。そし

てこの解法が実際に任意精度内で解を与えることが示される。

第5章　確率的政策による適応制御方式

これまで論じてきたベイズ的定式化のもとでは，制御問題はいずれにしてもベルマン形方程式を解くことに帰着される。ところが，従来の解法あるいは前章で提案した解法のいずれを用いても，ベルマン形方程式を解くには膨大な計算を要する。従って現在の段階ではベイズ最適政策によって実際的問題に対処することはきわめて困難である。そこで，本章は，ベルマン形方程式を解くことに伴う計算より幾らかでも負担が軽減され，かつある程度二元性が保たれるような制御則を発見的に構成することを試みる。ただし，本章では，遷移確率行列Pはパラメータϕの行列ベータ分布に従うものとする。この設定は，3章で述べた㈱が分解可能でない場合の一つの例に相当するものである。

事前確率密度$f(P|\phi)$を最大ならしめる$P=\hat{P}=\{\hat{P}_{ij}^{k}\}$を一般化最尤推定値という。本章では，まず，この推定値を含むレベルγの保証集合

$S_i^k(\gamma)=\{P_i^k \mid L_{ij}^k(\gamma)\leqq P_{ij}^k \leqq U_{ij}^k(\gamma)$ $(j=1,\cdots,N-1)$が定義される。$Pr\{P_i^k \in S_i^k(\ ) \| \phi\}\geqq\gamma$であるほか，$S_i^k(\gamma)$は下記の性質を持つ。

〔性質1〕任意のγに対して$L_{ij}^k(\gamma)\leqq\hat{P}_{ij}^k\leqq U_{ij}^k(\gamma)$である。また，$\gamma\to 0$のとき$S_i^k(\gamma)\to\hat{P}_i^k$，及び$\gamma\to 1$のとき$S_{ij}^k(\gamma)\to\{P_i^k \mid 0\leqq P_{ij}^k\leqq 1$ $(j=1,\cdots,N)\}$である。

〔性質2〕ℓ_i^kを状態iにおいて決定kを下す回数とする。γを一定に保ち各時刻で$S_i^k(\gamma)$を設定する場合，$\ell_i^k\to\infty$のとき$S_i^k(\gamma)\to P_i^k$である。

集合$S(\gamma)=\{P \mid P_i^k \in S_i^k(\gamma)$ $(k=1,\cdots,K_i\ ;\ i=1,\cdots,N)\}$に基づいたマックスー マックス最適政策を考えると，これは二元性を備えていることが定性的に知られる。そこで，本章は，γを一定に保ちながら$S(\gamma)$に基づいたマックスーマックス最適政策を各時刻で使用するというアルゴリズム1について検討する。その結果，アルゴリズム1はある一つの政策$\bar{U}(\gamma)$に収束し，U_{max}を最適政策，$K=\sum_i K_i$とするとき$Pr\{U_{max}=\bar{U}(\gamma)\|\phi\}$は漸近的に$\gamma^K$以上であることが示される。

次に，確率的な政策を導入しγを逐次的に改変することによってアルゴリズム1の適応機能をより強化したアルゴリズム2が提案される。アルゴリズム2のもとでは信頼度γ^K以上で最適な政策が確率γで使用される。そして最適政策の使用確率が1に収束することが示される。

本章は，また，事前確率分布の仮定を取り除き，アルゴリズム2をその漸近最適性のみが保たれるようにいくらか修正，簡略化したアルゴリズム2′を提案している。

なお，これらのアルゴリズムは行列ベータ分布の仮定に基づいたものであるが，これらはより一般的な場合にも適用できることが示される。

第6章　結　　言

本論文では，未知遷移確率の推定を伴うマルコフ決定問題に関して，適応制御方式を構成するという立場からベイズ最適政策による方式と確率的政策による方式とを論じてきた。

ベイズ的な方式の議論においては，主としてベイズ最適政策の有効性の問題を考察した。その結果，ベイズ最適政策は，ある程度すぐれた過渡特性（二元性）を備えているが，漸近特性については不備な点があるということが明らかになった。しかも，ベイズ的な方式は，ベルマン形方程式を解くことに伴う計算が膨大である。

これらベイズ最適政策に対する批判，検討に基づいて新たに提案したものが，確率的政策による方式である。本方式は，二元性なる性質を反映させる形で構成されている点で，従来より知られている確率近似法，学習強化法などとは異なる大きな特徴を持つ。

審 査 結 果 の 要 旨

マルコフ決定過程は，複雑な確率的システムの制御モデルの一つとして極めて重要であり，社会，経済システムの分野にも広く応用できるものである。近年，その理論の実用化を目ざして，遷移確率行列に含まれる未知パラメータを逐次推定しつつ制御を行う適応制御についても研究が行われるようになってきたが，現在なお解決すべき問題が多く残されている。著者は，ベイズ決定理論をマルコフ過程に適用して得られるベイズ最適政策の重要な基本的性質，すなわち漸近最適性と二元性を究明するとともにそれらの性質を具備しかつ従来より計算の容易な適応制御方式を確立した。本論文はその成果をまとめたもので全文6章よりなる。

第1章は緒言である。第2章ではベイズ決定理論，ゲーム理論に基づくマルコフ過程の適応制御方式の概要について述べ，以後の議論の展開に必要な数学的基礎を与えている。

第3章では，ベイズ論的適応制御方式における最適政策の漸近最適性と二元性について論じている。まず，パラメータ空間の構造を特徴づけるために分解可能性なる概念を新たに導入し，その性質と漸近最適性との関連に関する基本定理を導き，ついで，ベイズ最適政策がもつ二元性なる性質を明らかにしている。これらは，ベイズ最適政策の本質を明確にしたもので，興味ある知見である。

第4章では，ぼう大な計算を必要とするベイズ最適政策を算出するための新たな近似解法を与えている。本方法は，従来の方法に比べそのアルゴリズムが簡単で，計算量もかなり軽減できる。

第5章では，ベイズ論的方式とゲーム論的方式とを巧みに結合することにより，二元性を具備しかつ計算量が一層軽減されるような確率的政策による適応制御方式を提案している。本方式は高度の適応機能を有し，実用的にも極めて有用で，本研究で得られた重要な成果である。第6章は結言である。

以上要するに本論文は，未知遷移確率行列の推定を伴うマルコフ決定過程に関し，その基本的性質を究明するとともに高度の適応機能をもつ実用性に優れた適応制御方式を確立したもので，制御工学，システム工学の発展に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。