【50】

氏名・(本籍)	濱田年男
学位の種類	工学博士
学位記番号	第8364号
学位授与の日付	昭和63年10月26日
学位授与の要件	学位規則第5条第2項該当
学位論文題目	バンディット問題に関する研究
論文審査委員	(主査) 教授 坂口 実 (副査) 教授 竹之内 脩　教授 稲垣 宣生　教授 田畑 吉雄

論文内容の要旨

本論文は，バンディット問題と呼ばれる統計的決定問題の一分野に属するものである。問題はアームと呼ばれる実験が複数存在し，1度にその中の1つを選んで行うと，ある分布に従う観察値を得てその値に依存した利得を得る。その分布のパラメータが未知の場合には，その事前分布が与えられており，得られた観察値を用いてそのパラメータの事後分布を得ることができる。このような実験を逐次的にn回行うものとし，目的はn回の観察値の和を最大にすることであり，そのためには各回にどの実験を行えばよいかを決定することである。この問題の特徴は，実験を行ったときの利得のみならず，得られる情報の価値をも考慮して，学習しながら最適な実験を選ぶことにある。

まず最初に問題の一般的構成を述べ，過去の文献を分類して整理することにより，この分野のこれまでの研究の中での本研究の位置付けを行う。

問題はアームの個数が2の場合と，一般にm個の場合があり，またその中にパラメータが既知のものを含むか否かに分類される。まず一様分布に従う2アームで，1アームが既知の場合における最適計画が，1つのある decision number（あるいは threshold value）により決定されることを示し，その値を求める式を導き，数値計算を行って数表を得た。これをパラメータが1個の場合，利得が割り引かれる場合，パラメータが2個の場合に分けて解析した。次に2アーム共にパラメータが未知の場合の最適計画の構造を調べた。さらに複雑なmアームの場合には，将来どのように実験を進めてデータを得ても，決して選ばれないような実験を，現在の事前知識を用いて見つけ出し，そのアームを事前に省き，問題を縮小する方法を示した。最後に，2アーム問題に関する発展問題として，実験の切り替え費用を考慮した場合，および観察値を取り損なう可能性のある場合について論じ，それらの場合の最適計画を求め

た。

以上の結果を通して，不完全な情報のもとでの学習の重要性が明らかになった。

論文の審査結果の要旨

本論文はMulti－Armed Bandit（略してMAB）とよばれる逐次実験計画問題において，主として一様実験（uniform experiment）に関する理論を研究したものである。既存の理論はベルヌイ実験に関するものだけで，一様実験に関するものは著者が最初である。

第2章では一方のarmが既知のTwo－Armed Bandit（略してTAB）問題を解いている。一様実験 $\{U_{[o,u]}, U_{[o,1]}\}$，總試行回数n，に対してPareto priorを想定するとき問題は一様確率変数の独立和に対する最適停止問題に帰し，最適政策は，一般にはnon-myopicで複雑であるが合理的なものであり，それを特徴づける閾値関数をexplicitに導いている。

第3章では両方のarmが未知のuniform TABに対して独立Pareto priorのときの解法のための1つの接近を試みている。

第4章では一般にm-armed MABについて解法の1つの接近として，既知のimmediate reward vをもつ"arm0"を仮想的に追加して（m＋1）－armed MABを考える〝最小引退年金〟の方法について考察している。

第5章では，より複雑なuniform TABモデルを考える。第5.1節はarm iからarm 3－iへの転換コース $c_i > o(i=1,2)$ がある場合，第5.2節はnが確率変数の場合である。後者の場合，逐次実験の機会が離散時刻 t＝1，2，3，…において確率的に到来し，その到来時間間隔が母数pの幾何分布確率変数であるような再生過程を考え，計画期間をNとして最適実験計画を求めよ，という問題になる。

以上本論文はuniform MABについて世界初の理論的接近を試みて，この分野への新知見を加えるものであり，博士論文として価値あるものと認める。