

平成 30 年 6 月 25 日現在

機関番号：32702

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26400215

研究課題名(和文) 推移確率行列未知のマルコフ決定過程の研究

研究課題名(英文) On a study of Markov decision processes with unknown transition matrices

研究代表者

堀口 正之 (HORIGUCHI, Masayuki)

神奈川大学・理学部・教授

研究者番号：90366401

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：本課題は、意思決定過程における状態推移法則が未知である不確実環境下での最適化手法の研究である。具体的には、観測情報をもとに推移法則を区間推定してその区間に基づく最適化モデルを構成し、モデルの評価基準の区間表現を可能にするとともにその最適政策の特徴を明らかにした。また、不確実環境下での部分観測最適化問題として、ベイズ学習の研究に取り組み、最適停止問題および品質管理問題などにみられる区分確定的確率過程モデルでの最適化手法も導出した。

研究成果の概要(英文)：In this study, we consider the optimization problem of sequential decision processes with unknown transition probabilities. In this model with uncertainty, we formulate an optimization model with interval estimated transition probabilities from the information of observing the states of system. We derived the properties of optimal policies that are based on the representation of interval valued optimization criteria. We also consider Bayesian learning problems as the partially observed optimization problems with uncertain circumstances. In these models, we also deduced the optimization methods for optimal stopping problem and quality control problem in piecewise deterministic processes.

研究分野：計画数学

キーワード：マルコフ決定過程 推移法則未知 ベイズ学習

### 1. 研究開始当初の背景

意思決定過程の代表的モデルであるマルコフ決定過程(Markov Decision Processes、以下 MDP と略す)とその解析手法の動的計画法による最適解を導くアルゴリズムとして Howard の政策反復法や、修正政策反復法などが多く研究された。これらの手法には、次元の呪いの隘路の問題が存在し状態数の増加につれて計算量が爆発的に増大し実用規模の問題解析が困難である。他方、最適解の近似解法として推移確率に摂動的性質を与え、縮小写像の性質を利用し不動点問題を解析する White の Value Iteration Method が知られる。我々は、先行研究において動的モデル構造のより広い多重マルコフ連鎖で、最適解の近似解法として理論と解析アルゴリズムの研究を行った(Iki, Horiguchi, Kurano 2007)。これらの逐次決定モデルは、推移法則既知を仮定し、アルゴリズム内の反復方程式による更新手続きは、評価値が単調に増加し、更新ステップが進むほど制御政策も単調に最良へ近づくものとして表現される。これらの最適化モデルでは、既知情報のみから構成される確率制御過程下での研究がなされてきている。本研究課題では、部分観測情報モデルを含む不完全情報のもとでの適応的最適化手法の研究に取り組む。

### 2. 研究の目的

本研究では、意思決定過程において推移法則を区間として捉え、類似した推移法則モデルを含む状況下で状態推移履歴から蓄積される事前知識によるモデル情報の改定について、ベイズ統計によって区間推定し、その事後情報に基づき再構築される最適化問題から評価関数が集合値をもつ区間推定マルコフ決定過程を構成することを主要目的の一つとして取り組む。この統計的推論に基づく解析手法の数学的理論を総合的かつ統一的に構築し、モデルの構成パラメータの変化にも柔軟に対応できる頑健性ある最適化アルゴリズムの開発に取り組む。また、国際的な研究連携を継続し、不確実性下での意思決定モデルとして、停止マルコフ決定過程および区分確定的確率過程のもとでの最適化問題の研究に取り組む。

### 3. 研究の方法

(1) 不確実性下での意思決定過程における推定・学習理論について、状態観測により蓄積される経験知識や意思決定者の事前知識を用いたベイズ統計に基づく逐次学習モデル構築のための数学理論確立と、提案する区間推定 MDP における最適化アプローチが、モデル推定・意思決定に対してパラメトリックな感度解析にも頑健性・柔軟性を持つことを明らかにする。

(2) その具体的方法として以下の項目を挙げる。

- ・区間推定 MDP モデルでの推定および学習理論の研究
- ・部分観測可能なマルコフ決定過程と動的計

### 画法による解析の研究

- ・最適化アルゴリズムの開発と分析
- ・学会、研究集会、国際会議での発表と国際雑誌への投稿による研究評価
- ・海外研究協力者との共同研究課題のセミナー

### 4. 研究成果

本研究課題での研究成果を、その進展の時系列の順に以下のようにそれぞれまとめる。

(1) 推移法則未知のマルコフ決定過程における推定手法および学習理論の研究として、事前確率を区間で表現しそれをもとに事後区間推移確率を構成する区間推定型マルコフ決定モデルについて、適応学習と最適政策について考察を行った。真の推移法則の情報が未知の場合には、状態観測と逐次決定により蓄積されていく情報集合に基づいてベイズ統計による推移法則の推定とその推定に基づく最適化モデルが構成される。推移法則の事前情報に区間測度表現を用いることで事後情報も区間表現とすることができる。それによって区間型の最適方程式が導かれ評価関数は上限と下限を持つ区間に表現される。さらに、構成されたマルコフ決定モデルにおける逐次意思決定の最適政策に関しては漸近的最適政策の存在をこのモデルでも導出することができた。それらの具体的数値モデルの構成と、解析結果についてもまとめた。

(2) システムの状態が直接には観測できず間接的な情報からシステムの状態を推定し逐次意思決定を行うモデルである部分観測可能なマルコフ決定過程における問題として、2つの視点から研究に取り組んだ。品質管理モデルにおける多変量ベイズ管理図については、システムの継続運用と検査(停止)実行の2決定の問題を扱いモデルパラメータに関して閾値をもつ最適政策の導出とその管理モデルの考察を行った。

(3) システム情報の部分観測による推定を考慮した多段決定モデルにおいて、逐次支出問題の定式化及び最適化手法の構築に取り組んだ。具体的には、ベイズ学習や確率的凸性に関する最適性の条件や多段支出問題における評価関数の性質を、確率順序および離散凸解析における最適化手法の本モデルへの適用の考察も行い、最適政策の導出に取り組んだ。また、それぞれモデルが未知パラメータを含む場合の Bandit Problem に関する最適化問題の考察にも取り組み、パラメータ比較による最適政策の閾値の研究を行った。ベイズ最適停止問題におけるベータ分布族での有限試行停止規則として、試行回数に残数に応じた最適行動および評価関数の性質の研究にも取り組んだ。

(4) 実際的なモデルとして、品質管理におけるマルコフ決定モデルに関して、ベイズアプローチによる故障発生分布のパラメータに対する学習理論と適応政策の最適性の研究に取り組む推定制御による適応手法を明らかにした。さらに、システムの特性的パラ

メータが区分的に変化し得る一般的なマルコフモデルにおける最適制御の理論研究やシステム状態を間接的な観測情報から推定し逐次意思決定を行うモデルである部分観測可能なマルコフ決定過程における問題として、学習プロセスと逐次決定に対するベイズ的解析手法の研究へも取り組み、既存研究で知られている連続状態をもつマルコフ決定過程への同値変形が可能であるかどうかを考察し、本モデルでは、故障取り換えを繰り返す無限期間問題として同値変形が困難であり、最適停止問題の1つのサイクルにおける最適化問題を考察することで、部分観測情報に基づく最適政策の存在を明らかにした。

(5) ベイズ手法による推移法則の区間推定を行うとき、推移確率行列の各成分はそれぞれ閉区間として構成され、超平面上の凸集合によって構成される。凸解析での位相によって、区間型推移確率行列の演算が与えられ、最適化問題としては評価関数が集合値写像としての閉区間表現される区間型マルコフ決定過程として構成される。逐次観測のもとで構成される区間型確率行列から最適方程式が得られ、その最適解としての閉区間内の上限値と下限値の表現が導出される。いくつかの統計的評価関数形式における区間型確率行列の導出と、その区間行列下での最適化モデルの構築および最適政策のロバスト性の研究を行った。

(6) 我々の品質管理モデルでは、2状態または3状態の推移確率過程として、ひとたびシステムの劣化が発生すれば状態推移が区分的にジャンプのある確率制御過程として構成できることが知られていて、一般的な確率制御モデルと同様に最適方程式によって最適化問題が解析される。抽象空間での区分的確率過程での最適停止問題の研究にも取り組み、連続時間確率過程下での最適方程式の導出および最適解の特徴について明らかにした。

(7) 区間型マルコフ決定過程の適応的最適解の漸近的性質の考察として、評価関数値の収束の速さに関する研究に取り組んだ。さらに、品質管理問題として乱数発生シミュレーションによる分析にも取り組み、このほか、若手研究者とマルコフモデルの応用分野(生物統計およびレベニューマネジメント)に関してそれぞれ研究セミナー及び討議を実施した。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計10件)

[1] 佐々木稔、堀口正之、蔵野正美, 多変量ベイズ管理図の適応手法, 京都大学数理解析研究所講究録「不確実性の下での数理的意思決定の理論と応用」, 査読無, vol. 1912, pp.181-192,

[2] 中井 達, 確率的凸性と部分観測可能なマルコフ決定過程について, 京都大学数理解析研究所講究録「不確実性の下での数理的意思決定の理論と応用」, 査読無, vol. 1912, 193--201, 2014.8.

[3] 堀口正之, 多変量ベイズ管理図の適応手法(II), 京都大学数理解析研究所講究録1939「不確実性の下での数理モデルとその周辺」, 査読無, 152-161, 2015

[4] M. Horiguchi, "Bayesian Inference in Markov Decision Processes," Modern Trends in Controlled Stochastic Processes: Theory and Applications, Vol. 2 (A.B. Piunovskiy ed.), Luniver Press, 査読無, Vol. 2, 2015, pp. 177-189.

[5] 中井 達, マルコフ決定過程における学習プロセスと決定について, 京都大学数理解析研究所講究録「不確実性の下での数理モデルとその周辺」, 査読無, vol. 1939, 79--87, 2015.4.

[6] Dufour, F., Horiguchi, M. and Piunovskiy, A.B. "Optimal Impulsive Control of Piecewise Deterministic Markov Processes", Stochastics, 査読有, Volume 88, 2016 - Issue 7, Pages 1073-1098. DOI: <http://dx.doi.org/10.1080/17442508.2016.1197925>

[7] Masayuki HORIGUCHI, Interval Bayesian Method for Markov Decision Processes with Unknown Transition Matrices. In: Proceedings of the 2016 International Conference on Management and Operations Research, Qun Zhang, Zhenling Liu, Kuinam J. Kim Ed., 査読有, Apple Academic Press, pp. 284--287, 2016.

[8] 中井 達, 決定回数が未知の多段決定問題について, 京都大学数理解析研究所講究録「不確実・不確定性の下での数理意思決定モデルとその周辺」, vol. 1990, 査読無, 222--229, 2016.4.

[9] 中井 達, 確率的逐次割り当て問題について, 京都大学数理解析研究所講究録「確率的環境下における数理モデルの理論と応用」, vol. 2044, 査読無, 207--217, 2017.9.

[10] 堀口正之, 区間型マルコフ決定モデルについて, 京都大学数理解析研究所講究録 No.2044「確率的環境下における数理モデルの理論と応用」, 査読無, pp.199-206, 2017.9

[学会発表](計13件)

[1] 堀口正之, Adaptive Markov Control Processes について, 日本オペレーションズ・リサーチ学会研究部会確率モデルとその応用(第3回), 2014年9月, 放送大学 千葉学習センター,

[2] 堀口正之, 推移確率行列未知のマルコフ決定過程について, 日本オペレーションズ・リサーチ学会常設研究部会「待ち行列研究部会」(第252回), 2015年2月, 東京工業大学

[3] T. Nakai, A partially observable Markov decision process under stochastic convexity as an optimal maintenance problem, 20th Conference of the International Federation of Operational Research Societies, Barcelona, Spain, 2014.7.

[4] 中井 達, 不完備情報マルコフ過程での決定問題と確率的凸性について, 日本オペレーションズ・リサーチ学会 2014 年度秋期研究発表会 (北海道・北海道科学大学), 2014.8.

[5] M. Horiguchi and A.B. Piunovskiy, Optimal Stopping problem in uncertain Markov Decision Processes, 日本数学会 2016 年度年会, 2016 年 03 月 18 日, 筑波大学

[6] 中井 達, 不完備情報マルコフ過程での逐次支出問題について, 日本オペレーションズ・リサーチ学会 2015 年度春期研究発表会 (東京・東京理科大学), 2015.3.

[7] T. Nakai, A partially observable Markov process under stochastic convexity with supermodular function as an optimal investment problem, 10th Triennial Conference of the Association of Asia-Pacific Operational Research Societies (APORS 2015)., Kuching, Malaysia, 2015.8.

[8] T. Nakai, Optimization Problem on Partially Observable Markov Process under Stochastic Convexity with Supermodular Function, 7th European Congress of Mathematics (The quadrennial Congress of the European Mathematical Society), Berlin, Germany, 2016.7.

[9] 堀口正之, On a multivariate Bayesian control problem in Markov decision processes, 日本数学会, 2017 年 3 月, 首都大学東京

[10] 堀口正之, A Prior Detection Procedure on a Sequential Sampling Problem, 日本数学会, 2016 年 9 月, 関西大学

[11] 中井 達, 決定回数が未知の最適支出問題について, 日本オペレーションズ・リサーチ学会 2017 年度春期研究発表会 (沖縄・沖縄県市町村自治会館), 2017.3.

[12] M. Horiguchi, "Bayesian control chart with unknown parameter," 21st Conference of International Federation Operational Research Societies (IFORS2017), 17-21 July 2017, Québec City Convention Centre, Québec City Canada.

[13] 堀口正之, "Adaptive approach in a multivariate Bayesian control chart", 日本数学会 2018 年年会、東大 2018 年 3 月 [図書] (計 0 件)

[産業財産権]

出願状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
出願年月日 :  
国内外の別 :

取得状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
取得年月日 :  
国内外の別 :

[その他]  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

堀口正之 (HORIGUCHI, Masayuki)  
神奈川大学・理学部・教授  
研究者番号 : 9 0 3 6 6 4 0 1

### (2) 研究分担者

中井達 (NAKAI, TORU)  
千葉大学・教育学部・教授  
研究者番号 : 2 0 1 4 5 8 0 8

### (3) 連携研究者

安田正實 (YASUDA, Masami)  
千葉大学・理学研究科・名誉教授  
研究者番号 : 0 0 0 4 1 2 4 4

### (4) 研究協力者

Piunovskiy, Alexey ( )  
Dufour, Francois ( )