



シリーズ 情報科学における確率モデル 9

Series on Stochastic Models in Informatics and Data Science

ベイズ学習と マルコフ決定過程

中井 達【著】

コロナ社



シリーズ 情報科学における確率モデル
編集委員会

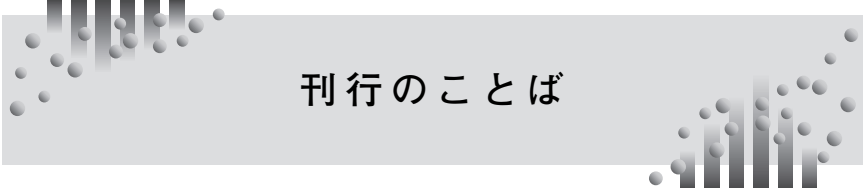
編集委員長

博士（工学） 土肥 正（広島大学）

編集委員

博士（工学） 栗田多喜夫（広島大学）

博士（工学） 岡村 寛之（広島大学）



刊行のこ と ば

われわれを取り巻く環境は、多くの場合、確定的というよりもむしろ不確実性にさらされており、自然科学、人文・社会科学、工学のあらゆる領域において不確実な現象を定量的に取り扱う必然性が生じる。「確率モデル」とは不確実な現象を数理的に記述する手段であり、古くから多くの領域において独自のモデルが考案されてきた経緯がある。情報化社会の成熟期である現在、幅広い裾野をもつ情報科学における多様な分野においてさえも、不確実性下での現象を数理的に記述し、データに基づいた定量的分析を行う必要性が増している。

一言で「確率モデル」といっても、その本質的な意味や粒度は各個別領域ごとに異なっている。統計物理学や数理生物学で現れる確率モデルでは、物理的な現象や実験的観測結果を数理的に記述する過程において不確実性を考慮し、さまざまな現象を説明するための描写をより精緻化することを目指している。一方、統計学やデータサイエンスの文脈で出現する確率モデルは、データ分析技術における数理的な仮定や確率分布関数そのものを表すことが多い。社会科学や工学の領域では、あらかじめモデルの抽象度を規定したうえで、人工物としてのシステムやそれによって派生する複雑な現象をモデルによって表現し、モデルの制御や評価を通じて現実に役立つ知見を導くことが目的となる。

昨今注目を集めている、ビッグデータ解析や人工知能開発の核となる機械学習の分野においても、確率モデルの重要性は十分に認識されていることは周知の通りである。一見して、機械学習技術は、深層学習、強化学習、サポートベクターマシンといったアルゴリズムの違いに基づいた縦串の分類と、自然言語処理、音声・画像認識、ロボット制御などの応用領域の違いによる横串の分類によって特徴づけられる。しかしながら、現実の問題を「モデリング」するためには経験とセンスが必要であるため、既存の手法やアルゴリズムをそのまま

ii 刊 行 の こ と ば


適用するだけでは不十分であることが多い。

本シリーズでは、情報科学分野で必要とされる確率・統計技法に焦点を当て、個別分野ごとに発展してきた確率モデルに関する理論的成果をオムニバス形式で俯瞰することを目指す。各分野固有の理論的な背景を深く理解しながらも、理論展開の主役はあくまでモデリングとアルゴリズムであり、確率論、統計学、最適化理論、学習理論がコア技術に相当する。このように「確率モデル」にスポットライトを当てながら、情報科学の広範な領域を深く概観するシリーズは多く見当たらず、データサイエンス、情報工学、オペレーションズ・リサーチなどの各領域に点在していた成果をモデリングの観点からあらためて整理した内容となっている。

本シリーズを構成する各書目は、おのおのの分野の第一線で活躍する研究者に執筆をお願いしており、初学者を対象とした教科書というよりも、各分野の体系を網羅的に著した専門書の色彩が強い。よって、基本的な数理的技法をマスターしたうえで、各分野における研究の最先端に上り詰めようとする意欲のある研究者や大学院生を読者として想定している。本シリーズの中に、読者の皆さんのアイデアやイマジネーションを掻き立てるような座右の書が含まれていたならば、編者にとっては存外の喜びである。

2018年11月

編集委員長 土肥 正



まえがき

多段決定過程を解決する数学的な方法として、動的計画法 (dynamic programming) がベルマン (R. Bellman) により創始され、その初期の結果が 1950 年代末には書物としてまとめられた。この動的計画法は、確率的多段決定過程に応用され、特にマルコフ決定過程についていろいろと研究されている。その中で、状態を直接に知ることができない部分観測可能なマルコフ決定過程 (POMDP) については多くの研究が進められ、1982 年にはモナハン (G. Monahan) によるサーベイ論文³³⁾†が発表されている。このような部分観測可能なマルコフ決定過程は、確率的多段決定過程を考えるうえで重要な問題であり、理論的な発展と多くの場面での応用が開発されている。この部分観測可能なマルコフ決定過程において、直接に知ることのできない状態に関する情報を解析するための基本的な方法は、ベイズの定理をもとにしたものである。本書では、ベイズの定理に基づく学習と、それをもとにした部分観測可能なマルコフ決定過程の基本的な結果と応用について著すことを目的としたものである。

マルコフ決定過程については、数多くの書籍や研究があるが、本書では動的決定過程の基本的な性質を解説し、あわせて部分観測可能なマルコフ決定過程を解析するために必要な確率的順序関係と、その性質を解説することに主眼をおいた。1 章では、本書で必要となる確率と確率過程の基本的な用語と性質について簡単に述べる。2 章では、部分観測可能なマルコフ決定過程において、ベイズの定理に基づく学習を解析するうえで基本となる確率的順序関係について解説する。このような確率的順序関係は、待ち行列理論や取替問題をはじめ数理ファイナンスなどいろいろな分野で用いられているが、本書では特に必要となる尤度比順序を中心に述べる。3 章では、多段決定過程を解析する手法であ

† 肩付き数字は、巻末の引用・参考文献の番号を表す。


動的計画法と最適性の原理を解説し、マルコフ決定過程の基本的性質についてまとめる。4章では、ジョブサーチと確率的逐次割当問題を中心に、マルコフ決定過程の一つである期待値最大化問題を解説する。この確率的逐次割当問題は、1970年代からいろいろと解析され、最近では新たなアプローチで研究されている。5章では、部分観測可能なマルコフ連鎖を中心に、ベイズの定理に基づく学習による事前分布と事後分布の関係などをまとめる。6章では、部分観測可能な2状態のマルコフ決定過程として、逐次解析や探索問題をはじめ最適停止問題などについて解説する。7章では、部分観測可能なマルコフ決定過程としてジョブサーチや確率的逐次割当問題、最適選択問題を取り上げ、その基本的な性質を解析する。

マルコフ決定過程や部分観測可能なマルコフ決定過程については、すでに多くの著書や論文があり、本書を執筆するにあたり、多くの文献を参考にさせていただいた。基本的な予備知識としては、大学1年次での線形代数、微分・積分と初歩の確率論と数理統計学を仮定しているが、必要に応じて学習していただければと思う。

本書を執筆するにあたって、執筆する機会を与えてくださり、いろいろなご助言をいただいた本シリーズ編集委員長である広島大学大学院工学研究科教授土肥正氏と、いろいろアドバイスをいただいた愛知大学名誉教授玉置光司氏に、この場を借りて心より感謝申し上げます。さらに、学校法人本山学園理事長室山義正氏にはいろいろお世話になり感謝いたします。また、本シリーズ担当委員には、原稿を丁寧に見ていただき、多くのミスを指摘していただきました。この場を借りて感謝いたします。

2022年1月

中井 達



目 次

第 1 章 確率と確率過程

1.1 確率と確率変数	1
1.1.1 確率空間	1
1.1.2 確率変数	2
1.1.3 期待値（平均）と分散	4
1.1.4 同時分布と周辺分布	5
1.1.5 独立	6
1.1.6 確率変数の和	8
1.1.7 特性関数	9
1.1.8 極限定理	10
1.2 条件付き確率と期待値	11
1.2.1 条件付き確率	11
1.2.2 条件付き期待値	12
1.3 確率分布	13
1.3.1 二項分布	13
1.3.2 ポアソン分布	14
1.3.3 一様分布	15
1.3.4 指数分布	15
1.3.5 ガンマ分布	16
1.3.6 正規分布	17
1.4 計数過程	18
1.4.1 確率過程	18
1.4.2 ポアソン過程	18
1.4.3 到着時間間隔	21
1.4.4 非斉次ポアソン過程	23

1.4.5	ポアソン過程の合成	24
1.5	マルコフ連鎖とマルコフ過程	25
1.5.1	マルコフ連鎖	25
1.5.2	チャップマン-コルモゴロフ方程式	27
1.5.3	離散時間マルコフ過程	28
1.6	連続時間の確率過程	30
1.6.1	連続時間マルコフ連鎖	30
1.6.2	チャップマン-コルモゴロフ方程式	30
1.6.3	コルモゴロフの方程式	32
1.6.4	出生死滅過程	35

第2章 確率的順序関係

2.1	確率順序	38
2.2	故障率関数と順序	42
2.2.1	故障率関数	42
2.2.2	故障率順序	44
2.3	尤度比順序	47
2.4	尤度比順序と TP_2	56
2.4.1	TP_2	56
2.4.2	MTP_2	62
2.4.3	シフト尤度比順序	65
2.5	関数類による順序関係	65

第3章 マルコフ決定過程

3.1	動的計画法	70
3.1.1	多段決定過程	70
3.1.2	最適方程式	72
3.1.3	最適性の原理	73

3.2 多段決定過程	77
3.2.1 確率的多段決定過程	77
3.2.2 定常政策とマルコフ決定過程	79
3.3 割引のあるマルコフ決定過程	80
3.3.1 最適方程式	80
3.3.2 最適政策	82
3.3.3 逐次近似法	85
3.3.4 政策反復法	87
3.4 最適支出問題	89
3.4.1 アウトカムと確率過程	89
3.4.2 確定的最適支出問題	89
3.5 マルコフ過程の最適支出問題	94

第4章 ジョブサーチと確率的逐次割当問題

4.1 ジョブサーチ	101
4.1.1 関数 $T_F(z)$ と $S_F(z)$	101
4.1.2 最適方程式と最適政策	104
4.1.3 リコールのあるジョブサーチ	108
4.2 確率的逐次割当問題	112
4.2.1 ハーディの補題	112
4.2.2 確率的逐次割当問題	113
4.2.3 マルコフ連鎖の確率的逐次割当問題	117
4.2.4 割引のある確率的逐次割当問題	121
4.3 ポアソン過程の確率的逐次割当問題	127
4.4 最適選択問題	130

第5章 学習と情報

5.1 ベイズの定理	139
5.1.1 ベイズの定理	139

5.1.2	事前分布と事後分布	141
5.2	共役分布族	142
5.2.1	ポアソン分布	142
5.2.2	指数分布	143
5.2.3	期待値が未知の正規分布	144
5.3	部分観測可能な2状態マルコフ連鎖	145
5.3.1	2状態マルコフ連鎖	145
5.3.2	学習プロセス	147
5.4	部分観測可能なマルコフ連鎖	150
5.4.1	可算状態のマルコフ連鎖	150
5.4.2	事前情報と事後情報	151
5.5	部分観測可能なマルコフ過程	154
5.5.1	離散時間マルコフ過程	154
5.5.2	事前情報と事後情報	156
5.5.3	正規分布に基づくモデル	158
5.6	一度に複数の値を観測する学習プロセス	159
5.6.1	独立な確率変数の場合	159
5.6.2	MTP ₂ の場合	162

第6章 部分観測可能な2状態マルコフ決定過程

6.1	逐次解析	164
6.2	探索問題	167
6.2.1	目的物が動かない場合	167
6.2.2	マルコフ連鎖の探索問題	169
6.3	部分観測可能な2状態マルコフ決定過程	172
6.3.1	部分観測可能な2状態マルコフ連鎖	172
6.3.2	部分観測可能な最適停止問題	173
6.3.3	部分観測可能な取替問題	178

第7章 部分観測可能な逐次割当問題

7.1 部分観測可能なジョブサーチ	185
7.2 部分観測可能な確率的逐次割当問題	191
7.3 部分観測可能な最適選択問題	199
お わ り に	209
引用・参考文献	211
索 引	216

3

マルコフ決定過程

多段決定過程における動的計画法とマルコフ決定過程について考える。

3.1 動的計画法

3.1.1 多段決定過程

多段決定過程 (multistage decision process) は、あらかじめ決められた目的を達するために、段階を追って逐次に決定をとる決定問題である。この多段決定過程を考える段階の数は、有限のときもあれば無限のときもあり、この数を計画期間 (planning horizon) という。

このような多段決定過程は、四つの要素 ($S, A, r(s, a), \tau(s, a)$) により定式化される。

- (1) 状態空間 (state space) S : 状態を s で表し、状態全体の集合を S とする。
- (2) 行動空間 (action space) A : 決定を a で表し、決定全体の集合を A とする。
- (3) 利得関数 (reward function) $r(s, a)$: 状態が s のときに決定 a をとれば、利得 $r(s, a)$ が得られる。
- (4) 推移法則 (transition rule) $\tau(s, a)$: 状態が s のときに決定 a をとれば、状態は法則 $\tau(s, a)$ に従い、つぎの状態が定まる。

この多段決定過程は、それぞれの段階で状態 s を知って決定 a をとる ($s \in S, a \in A$)。この状態 s と決定 a により、この段階での利得 $r(s, a)$ が定まり、つぎの段階の状態は推移法則 $\tau(s, a)$ に従って定まる。このような決定と推移を繰り返して、利得の総和を最大にする。また、この多段決定過程の最初の状態を初

期状態 (initial state) という。ここでは、利得を考えたが、費用あるいは損失 $c(s, a)$ を考えるときは、 $-c(s, a)$ を利得と考えればよい。

以下では、離散時間の多段決定過程とし、計画期間を n とし、決定をとる時点を $\{1, 2, 3, \dots, n\}$ とする。状態空間は、有限集合あるいは可算集合とし、一般性を失うことなく整数で番号づけられ、 $S = \{1, 2, 3, \dots\}$ とする[†]。行動空間は、有限集合とする。状態が s のときに、決定 a をとったときの利得 $r(s, a)$ は非負関数とする。状態が s のときに、決定 a をとったときの推移法則 $\tau(s, a)$ は確定的でただ一つ定まる。さらに、状態空間、行動空間、利得関数、推移法則は決定の段階にかかわらず同じとする。これらの条件は、多段決定過程の基本的な性質を説明するためであり、条件が変わっても適当な仮定のもとで、基本的な性質が同じように成り立つ。以降の例 3.1 は、上記の条件から外れたものとなっている。

初期状態が i_1 のとき、この状態から始めて決定 a_1, \dots, a_n をこの順番にとれば、状態は推移法則に従って

$$i_1, i_2 = \tau(i_1, a_1), i_3 = \tau(i_2, a_2), \dots, i_{n+1} = \tau(i_n, a_n)$$

と推移する。この決定の列に対し、それぞれの期の状態と決定により利得が定まり、それぞれ

$$r(i_1, a_1), r(i_2, a_2), \dots, r(i_n, a_n)$$

となる。この多段決定過程では、 n 期目での決定をとったあと、最終の状態が i のときに、各期での利得とは別に、状態 i に応じて利得 $R(i)$ が得られ、この利得を最終利得 (terminal reward) という。

いま、 k 期目での決定 a_k を ($k = 1, 2, \dots, n$)、状態 i_k の関数として $a_k = a_k(i_k)$ と表し、この関数を決定関数 (decision function) という。これらの決定関数の列 $\{a_k(\cdot)\}_{k=1,2,\dots,n}$ を与えたとき、これを一つの政策 (policy) という。

[†] 状態空間が有限集合あるいは可算集合であるときは状態を i, j, k, \dots で表し、それ以外のときは状態を s, t, u, \dots で表す。

初期状態が i_1 のとき、 $\{a_k(\cdot)\}_{k=1,2,\dots,n}$ を一つの政策とし、この政策に従って n 期間を終えたあとの総利得 $R_n(i_1 | \{a_k(\cdot)\}_{k=1,2,\dots,n})$ は最終利得を含め

$$R_n(i_1 | \{a_k(\cdot)\}_{k=1,2,\dots,n}) = \sum_{k=1}^n r(i_k, a_k(i_k)) + R(i_{n+1}) \quad (3.1)$$

となる。ただし、 $i_{k+1} = \tau(i_k, a_k(i_k))$ とする ($k=1, 2, 3, \dots, n$)。式 (3.1) をこの多段決定過程の目的関数 (objective function) といい、この値を最大にする政策を最適政策 (optimal policy) という。

3.1.2 最適方程式

初期状態が i_1 のとき、 n 期間にわたり最適政策に従って得られる総利得を $v_n(i_1)$ とおく。 $n=1$ のときは

$$v_1(i_1) = \max_{a \in A} (r(i_1, a) + R(\tau(i_1, a))) \quad (3.2)$$

となる。この 1 期間のモデルでの最適政策は、式 (3.2) の右辺を最大にする決定である。さらに、目的関数 (3.1) はそれぞれの期での利得を加えた関数だから、つぎの基本的な関係式が成り立つ。

定理 3.1

$$v_n(i_1) = \max_{a_1(\cdot), \dots, a_n(\cdot)} \left(\sum_{k=1}^n r(i_k, a_k(i_k)) + R(i_{n+1}) \right) \quad (3.3)$$

のとき

$$v_n(i_1) = \max_{a_1 \in A} (r(i_1, a_1) + v_{n-1}(\tau(i_1, a_1))) \quad (3.4)$$

となる ($n=1, 2, \dots$)。ただし、 $v_0(i) = R(i)$ である ($i \in S$)。

【証明】 $i_{k+1} = \tau(i_k, a_k(i_k))$ とし ($k=1, 2, 3, \dots, n$)、 $i_2 = \tau(i_1, a_1)$ に注意すれば、式 (3.3) で 1 期目での決定と、2 期目以降の決定の列に分けて

$$v_n(i_1) = \max_{a_1(\cdot), \dots, a_n(\cdot)} \left(r(i_1, a_1(i_1)) + \sum_{k=2}^n r(i_k, a_k(i_k)) + R(i_{n+1}) \right)$$

$$\begin{aligned}
 &= \max_{a_1(\cdot)} \left(r(i_1, a_1(i_1)) + \max_{a_2(\cdot), \dots, a_n(\cdot)} \left(\sum_{k=2}^n r(i_k, a_j(i_k)) + R(i_{n+1}) \right) \right) \\
 &= \max_{a_1(\cdot)} \left(r(i_1, a_1(i_1)) + \max_{a_2(\cdot), \dots, a_n(\cdot)} R_{n-1}(i_2 | \{a_k(\cdot)\}_{k=2, \dots, n}) \right) \\
 &= \max_{a_1 \in A} (r(i_1, a_1) + v_{n-1}(\tau(i_1, a_1))) \tag{3.5}
 \end{aligned}$$

となり、式 (3.4) が成り立つ。

◇

式 (3.4) をこの多段決定過程の**最適方程式** (optimality equation) という。また、式 (3.4) を上向きの**再帰方程式** (recursive equation) ともいい、 $n=1, 2, 3, \dots$ と解いて $v_1(i), v_2(i), v_3(i), \dots$ が順次求まる。また、 n 期間の多段決定過程で、式 (3.4) の右辺を最大にする a_1 の一つを $a_1^*(i_1)$ とおく。これは、初期状態が i_1 のときの n 期間モデルでの 1 期目の最適な決定となる。2 期日以降の最適な決定 a_k^* は ($k=1, 2, \dots, n$)、それぞれ

$$a_k^*(i_k^*) \quad \text{および} \quad i_k^* = \tau(i_{k-1}^*, a_{k-1}^*(i_{k-1}^*))$$

となる。ただし、 $i_1^* = i_1$ とする。すなわち、決定関数の列が $a_1^*(\cdot), \dots, a_{n-1}^*(\cdot), a_n^*(\cdot)$ のとき、 $\{i_k^*\}_{k=1, 2, \dots, n}$ は初期状態が $i_1^* = i_1$ のとき最適政策によって推移する状態の列となる。この最適方程式を用いて、最適政策を逐次に求める方法を**動的計画法** (dynamic programming) という。

これまででは、それぞれの期での利得関数は同じとしたが、利得関数が期により異なり、 k 期目の状態 i_k で決定 a_k をとったときの利得を $r_k(i_k, a_k)$ と拡張することもできる ($k=1, 2, \dots, n$)。

3.1.3 最適性の原理

初期状態が i の n 期間の多段決定過程の最適政策を $a_1^*(\cdot), a_2^*(\cdot), \dots, a_n^*(\cdot)$ とすると、式 (3.4) は

$$v_n(i_1) = r(i_1, a_1^*(i_1)) + v_{n-1}(\tau(i_1, a_1^*(i_1))) \tag{3.6}$$

となる。式 (3.6) から、初期状態が i_1 のとき最適政策によって得られる総利得

は、初期状態 i_1 で最適政策により得られる 1 期目での利得と、2 期目での状態 $\tau(i_1, a_1^*(i_1))$ を初期状態とした $n-1$ 期間の多段決定過程で最適政策により得られる総利得の和に等しい。

式 (3.1) と式 (3.4) より、 $i_2^* = \tau(i_1, a_1^*(i_1))$ となり

$$v_n(i_1) = r(i_1, a_1^*(i_1)) + R_{n-1}(i_2^* | \{a_k^*(\cdot)\}_{k=2,3,\dots,n})$$

となる。したがって、式 (3.6) より

$$v_{n-1}(\tau(i_1, a_1^*(i_1))) = R_{n-1}(i_2^* | \{a_k^*(\cdot)\}_{k=2,3,\dots,n})$$

となり、つぎのことが成り立つ。

定理 3.2 (最適性の原理 (principle of optimality)) 初期状態が i_1 のとき、 n 期間の多段決定過程の最適政策を $a_1^*(\cdot), a_2^*(\cdot), \dots, a_n^*(\cdot)$ とすれば、 $a_2^*(\cdot), a_3^*(\cdot), \dots, a_n^*(\cdot)$ は、 $i_2^* = \tau(i_1, a_1^*(i_1))$ を初期状態とする $n-1$ 期間の多段決定過程の最適政策である。

定理 3.3 最適性の原理 (定理 3.2) から定理 3.1 が導ける。

【証明】 式 (3.3) より

$$v_n(i_1) = \max_{a_1(\cdot), \dots, a_n(\cdot)} R_n(i_1 | \{a_k(\cdot)\}_{k=1,\dots,n})$$

とする。1 期目に決定 a をとり、つぎに直面する状態 $\tau(i_1, a)$ を初期状態とする $n-1$ 期間の多段決定過程では、その最適政策に従う政策は、この n 期間モデルの政策の一つとなる。よって、最初の期での任意の決定 $a \in A$ に対し

$$v_n(i_1) \geq r(i_1, a) + v_{n-1}(\tau(i_1, a))$$

となり

$$v_n(i_1) \geq \max_{a \in A} (r(i_1, a) + v_{n-1}(\tau(i_1, a))) \quad (3.7)$$

索引

【あ】		確率的多段決定過程	77	計数過程	19
アウトカム	89, 100	——の最適方程式	78	継続領域	177, 182
アウトプット	100	確率的逐次割当問題	113	決定関数	71
安定	109	(——)の最適政策	113, 197	検査領域	177, 182
		(——)の最適方程式	114	【こ】	
【い】		部分観測可能な——	191	行動空間	70
イェンセンの不等式	99	ポアソン過程の——	127	候補	111
一様分布	15, 60, 104, 106	マルコフ連鎖の——	117	故障率関数	42
インプット	100	割引のある——	121	故障率順序	45, 68
		確率分布	2	コルモゴロフ	
【お】		確率分布関数	3	——の後退方程式	32
凹関数	49	確率変数	2	——の前進方程式	35
凹順序	69	n 次元——	5	【さ】	
		確率母関数	10	再帰方程式	73
【か】		確率密度関数	3	最終利得	71
拡散過程	26	加重分布関数	153, 157, 186, 192, 200	再生性	
学習	141	可測関数	2	正規分布の——	18
学習プロセス	147	可測空間	1	二項分布の——	14
確定的最適支出問題	89	関数類	65	ポアソン分布の——	15
(——)の最適方程式	90	完全加法族	1	最適支出問題	89
確定的政策	79	ガンマ分布	16, 142, 143	確定的——	89
確率過程	18			マルコフ過程の——	94
離散時間の——	18	【き】		最適政策	72, 81, 85
連続時間の——	18	記憶がない	16, 22	最適性の原理	74
確率空間	2	危険愛好的な選好	69	最適選択問題	130
確率質量関数	3	危険回避的な選好	69	(——)の最適政策	134
確率順序	38	期待値	4	(——)の最適方程式	131
確率尺度	2	共分散	7	部分観測可能な——	199
確率的政策	79	共役分布族	142	最適停止問題	108
確率的増加	54	近視眼的政策	110	(——)の最適方程式	108, 174
確率的増加凹	55			部分観測可能な——	173
確率的増加凹(sp)	55	【け】		最適方程式	73
確率的増加凸	55	計画期間	70		
確率的増加凸(sp)	55				

【し】

しきい値 106, 116
 事後確率 140
 事後情報 147, 151, 156, 159, 170, 179, 191
 事後分布 141
 事象 2
 指数分布 15, 40, 42, 45, 46, 50, 103, 106, 143
 事前確率 140
 事前情報 146, 151, 156, 159, 170, 186, 191
 事前分布 141
 シフト尤度比順序 65
 周辺確率分布関数 6
 周辺確率密度関数 6
 周辺分布 6
 周辺累積分布関数 6
 縮小写像 86, 123
 出生死滅過程 35
 順序
 確率—— 38
 故障率—— 45, 68
 尤度比—— 47, 51, 67, 150, 156
 順序統計量 22, 131, 160, 200
 純々な出生過程 36
 条件付き確率 11, 139
 条件付き確率分布関数 11
 条件付き確率密度関数 12
 条件付き期待値 12
 条件付き累積分布関数 11, 12
 状態空間 18, 70
 情報過程 145, 146, 150, 155
 情報行列 146, 173
 初期状態 70
 初期分布 26
 ジョブサーチ 101
 (——の) 最適政策 106
 (——の) 最適方程式 104, 110
 部分観測可能な—— 185

リコールのある—— 110

【す】

推移確率 26
 推移確率行列 27
 推移法則 70

【せ】

正規分布 17, 40, 51, 59, 144, 158
 政策 71
 政策反復法 88
 積率母関数 9
 全確率の公式 139
 全順序 54

【た】

対称群 112
 対数正規分布 55
 大数の法則 10
 たたみ込み 8
 多段決定過程 70
 探索問題 167
 (——の) 最適政策 169
 (——の) 最適方程式 168, 171
 マルコフ連鎖の—— 169

【ち】

チェビシェフの不等式 10
 置換 112
 逐次解析 164
 ——の最適政策 166
 ——の最適方程式 165
 逐次近似法 86
 チャップマン-コルモゴロフ
 方程式 28, 32
 中心極限定理 11

【て】

定常政策 79
 定常独立増分 18
 停止領域 177, 190

【と】

同時確率分布関数 5
 同時確率密度関数 5
 同時分布 5
 同時累積分布関数 5
 動的計画法 73, 77
 特性関数 9
 独立 7
 独立増分 18
 凸関数 49, 176
 凸集合 166
 凸順序 69
 取替問題 178
 部分観測可能な—— 178
 取替領域 182

【に】

二項分布 13, 50
 2状態マルコフ決定過程
 部分観測可能な—— 172
 2状態マルコフ連鎖 145
 部分観測可能な—— 146, 172, 178

【の】

ノルム 85

【は】

ハーディの補題 112
 半順序 53

【ひ】

秘書問題 111
 非斉次ポアソン過程 23
 標準正規分布 17
 標準偏差 5
 標本空間 2

【ふ】

不動点定理 86

部分観測可能な	平均値関数	23		
——確率的逐次割当問題	ベイズの定理	140		【む】
191	ベルヌーイ分布	13	無作為標本	142
——最適選択問題				【も】
199	【ほ】			
——最適停止問題	ポアソン過程	19, 36, 61, 127	目的関数	72
173	ポアソン過程の確率的逐次		【ゆ】	
——ジョブサーチ	割当問題	127	尤度関数	141
185	(——の) 最適政策	129	尤度比	47
——取替問題	(——の) 最適方程式	128	尤度比順序	47, 51,
178	ポアソン分布	14, 50, 58, 142	67, 150, 156	
——2状態マルコフ	ボレル集合	2	ユール過程	36
決定過程	ボレル集合族	2		
172				
——2状態マルコフ連鎖	【ま】			【り】
146, 172, 178	マルコフ過程	26	リコール	101, 110
——マルコフ過程	部分観測可能な——	154, 185	離散確率変数	2
154, 185	離散時間——	26, 28, 59, 185	離散時間マルコフ過程	26, 28, 59, 185
——マルコフ決定過程	マルコフ過程の最適支出		利得関数	70
164	問題	94		【る】
——マルコフ連鎖	(——の) 最適方程式	94	累積分布関数	3
145, 150, 191	マルコフ決定過程	79	ルベージュスティルチェス	
	——の最適方程式	81	積分	4
	部分観測可能な——	164		【れ】
	割引のある——	80	連続確率変数	2
	マルコフ性	25	連続時間マルコフ連鎖	26, 30, 61
	マルコフ連鎖	26, 56, 117, 145, 172		
	部分観測可能な——	145, 150, 191		【わ】
	連続時間——	26, 30, 61	ワイブル分布	43, 45
部分観測可能な確率的逐次	マルコフ連鎖の確率的逐次		割引係数	80
割当問題	割当問題	117	割引のある確率的逐次割当	
(——の) 最適政策	(——の) 最適政策	118	問題	121
(——の) 最適方程式	(——の) 最適方程式	118	(——の) 最適政策	122
192			(——の) 最適方程式	121
部分観測可能な最適選択				
問題				
(——の) 最適政策				
(——の) 最適方程式				
203				
200				
部分観測可能な最適停止				
問題				
(——の) 最適政策				
(——の) 最適方程式				
173				
177				
174				
部分観測可能なジョブ				
サーチ				
(——の) 最適政策				
(——の) 最適方程式				
185				
189				
186				
部分観測可能な取替問題				
(——の) 最適政策				
(——の) 最適方程式				
178				
182				
180				
分位点				
117				
分散				
5				
【へ】				
平均				
4				

[A]		[O]		[U]
α -最適政策	81	OLA-政策	110	$V_F(u)$ 186, 192
[D]		[S]		[V]
DFR 確率変数	42, 49	$S_F(z)$	101	$U_F(u)$ 186, 192
[I]		[T]		
IFR 確率変数	42, 49	$T_F(z)$	101	
[M]		TP_r	56, 57, 59, 61	
MTP ₂	63, 163	TP_2	56, 58, 59,	
			61, 62, 148, 150, 156	

— 著者略歴 —

1975年 京都大学理学部卒業
1978年 大阪大学大学院基礎工学研究科博士前期課程修了（数理系専攻）
1981年 大阪大学大学院基礎工学研究科博士後期課程退学（数理系専攻）
1981年 大阪府立大学助手
1985年 工学博士（大阪大学）
1987年 神戸大学助教授
1991年 九州大学助教授
1996年 九州大学教授
2008年 千葉大学教授
2018年 千葉大学名誉教授
2020年 学校法人本山学園岡山医療専門職大学理事
現在に至る

ベイズ学習とマルコフ決定過程

Markov Decision Process Based on Bayes' Theorem

© Toru Nakai 2022

2022年3月25日 初版第1刷発行

検印省略

著者 ^{なか} ^い ^{とおる}
中井 達
発行者 株式会社 コロナ社
代表者 牛来真也
印刷所 三美印刷株式会社
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社
CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844 · 電話 (03) 3941-3131(代)

ホームページ <https://www.coronasha.co.jp>

ISBN 978-4-339-02839-3 C3355 Printed in Japan

(新井)



JCCOPY <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構（電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jccopy.or.jp）の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。