シリーズ 情報科学における確率モデル　4

Series on Stochastic Models in Informatics and Data Science

# マルコフ決定過程

## ――理論とアルゴリズム――

中出　康一【著】

コロナ社

# 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

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

マルコフ決定過程は，現在の状況を表す状態を観測しながら，ある利益（費用）規範の下で最適な決定を行う確率過程です。マルコフ決定過程は Bellman の動的計画法を基に構築されています。古くからオペレーションズ リサーチの分野において理論的に発展し，また不確定な要素を含む動的な問題に対する最適な決定を求める方法としても確立しています。20 世紀末以降，人工知能分野の一つである強化学習は，このマルコフ決定過程の理論を基礎として発展してきています。

私は学生のときから現在まで，マルコフ決定過程を最適政策を求める手段として利用してきました。マルコフ決定過程の理論については，海外，特にアメリカでは発展が目覚ましく，近年でも強化学習や近似アルゴリズム等，理論ならびに応用に関する研究が進展しています。このためアメリカを中心に海外では多くの論文が掲載され，また関連本も多く出版されています。しかし，国内では，ハワードのダイナミックプログラミングの訳本は古くにはあったものの，近年でもマルコフ決定過程自体を主要なテーマとした日本語の本はほとんど見当たりませんでした。広島大学の土肥先生からこの本の出版の話をいただいたとき，浅学非才な私が引き受けることに少し躊躇（ちゅうちょ）しましたが，それでも執筆することにした一番の理由は，理論に基づきながらも，しかし関心のある研究者・学生にも理解しやすい日本語で書かれたマルコフ決定過程の本が必要であると感じたからです。また，私自身，マルコフ決定過程を利用しながらも，その幅広い理論的背景について頭の中で整理できていないと感じており，この際本に書き留めることで理解を深めたいという気持ちもありました。

この本では，マルコフ決定過程に関する理論の中でも特に重要な基本理論，また実際に問題を定式化して解き，最適決定政策を求める際必要となる計算手法

に焦点を設けています。理論の背景を理解していただくために，その基礎となる確率，マルコフ連鎖，確率過程に関しても章を設けました。さらに，部分観測可能マルコフ決定過程や，先に述べた強化学習，近似アルゴリズムについても触れています。自分自身強化学習について研究に用いてきたわけではありませんが，マルコフ決定過程と関連付けながら基礎的な内容をまとめました。

この本の内容は古典的なものが多く，あくまで基礎的理論を，読者に理解をいただけるようにつとめてきたつもりですが，十分でなければそれは著者の力量不足です。それでも，この本が読者にとって少しでもお役に立つことができれば望外の喜びです。

名古屋工業大学名誉教授の故 大野勝久先生には学生時代から名古屋工業大学の教員時代を通して，マルコフ決定過程の理論はもとより多くのことを学ばせていただきました。大野先生が亡くなられてもう 5 年半になりますが，ほんの少しでも恩返しができたでしょうか。

広島大学の土肥正先生，岡村寛之先生には執筆を勧めていただいたことに感謝いたします。最後に，コロナ社の皆様には出版に至るまで，たいへんお世話になりました。深謝いたします。

2019 年 2 月

中出　康一

# 目　　　　次

## 第 7 章　部分観測可能マルコフ決定過程



## 第 8 章　マルコフ決定過程の展開

# 1 マルコフ決定過程の概要

本章では，マルコフ決定過程の理論的基礎となる動的計画法の説明とともに，マルコフ決定過程の概要を述べ，この本の構成を示す。

## 1.1 ORと確率モデル

人間社会ではさまざまなシステムが存在する。商品を生産して販売する場合，原材料を輸入し，加工，生産して販売者まで届け，消費者が購入する。その際には輸送やお金のやりとりが必要となる。このような生産・物流・販売において，効率的な仕組みにしないとコストがかかる。コストに上乗せしてものの値段を高くすると売れなくなり，値段を低くすると利益が上がらない。また，在庫が積み上がると保管や処分に対する費用がかかり，一方で在庫を少なくすると需要を満たされず，顧客の満足度が下がってしまう。したがって，もの・かね・ひと・情報等を構成要素として，さまざまな条件を考慮しながら生産・物流・在庫に関する統合的な仕組み，すなわちシステムを構築する必要がある。

このような生産・物流・販売システム以外にも，通信，医療，大学等さまざまなシステムが存在する。これらのシステムを数理的にモデル化し，さまざまな解析や計算手法を通して効率的な運用を目指す戦術として，**オペレーションズ リサーチ**（operations research, OR）が知られている。

オペレーションズ リサーチで扱う手法は，大きく分けて決定性（deterministic）の問題と確率的（probabilistic, stochastic）問題に分けられる。

決定性の問題として代表的なものは，線形計画法（linear programming）を

始めとする数理計画問題（mathematical programming problem）であり，古くから研究がなされている。数理計画問題は，対象となるシステムについて制御できるものを変数として与え，与えられた種々の制約の下で，費用を最小にする（あるいは利益を最大にする）問題である。例えば，ある製品を製造し，店舗で販売を行うとしよう。生産を行うために必要となる原料の量は限られており，1 日に生産できる量には上限がある。物流には輸送量の上限がある。さらに，数日後の商品の需要が事前にわかっている場合には，前工程を含め過不足なく生産し販売することが必要である。このようなさまざまな条件を考慮して生産・輸送等の費用を考慮して適切に製品を生産し，物流を通して各店舗に運ばなければならない。最小化すべき費用は，例えば原料費，生産・在庫費，物流費の総和であり，目的関数と呼ばれる。

目的関数，制約条件をともに変数を用いて式で表現する。すべて 1 次式で表され，かつ変数のとり得る値が実数であるとき，**線形計画問題**（linear programming problem）と呼ぶ。また，一部の式が 2 次式等の非線形の式で表現されると，この問題は非線形計画問題となる。

これらの変数の一部，あるいはすべてが整数制約で表される場合（混合）**整数計画問題**（mixed integer programming problem）と呼ばれる。品物の個数や作業者への作業の割当てとに関する変数は整数の値しかとれないので，整数計画問題になる。

決定性の問題として代表的なものの一つにスケジューリング問題がある。最も基本的なものは，いくつかの対象物が与えられたとき，目的関数を最小化するように順序付けする問題である。その代表例が巡回セールスマン問題である。近年，決定性問題として AHP, DEA 等の意思決定問題も多く扱われている。

一方，つねに一定の値をとるとはいえない要素が含まれているとき，確率分布を用いて対象とするシステムをモデル化することが多い。**確率モデル**（stochastic model）として代表的なものには待ち行列理論，在庫理論と信頼性理論がある。

待ち行列理論は 100 年あまり前の電話回線網の分析から始まり，理論的発展がなされ，特に通信分野を中心とした応用が発展している。在庫理論は生産・

販売における部品・完成品の管理を対象としている。商品の日々の需要は確率的である一方，店舗からの発注には発注から到着までにかかる時間（発注リードタイム）が必要なため，日々の実需要や需要予測をみながら適切に発注する必要がある。信頼性理論は工程等の故障の分析と保全政策の決定等が該当する。例えば，機械で製品の加工を続けるとき，適切な時期に点検を行い，修理する，あるいは部品を交換するなどの作業をしないと工程の維持費用が高価になる。

これらのモデルにおいて，電話の呼の発生時期や製品の需要，機械部品の寿命は確定していないため，確率的な要素として考える必要がある。近年はファイナンスなど金融の分野にも OR が適用されている。株価や為替が確定的でないことは明らかであろう。

決定性の問題にも確率的な問題にも適用できる手法として，動的計画法がある。次節ではこの動的計画法を取り上げる。

## 1.2 動 的 計 画 法

**動的計画法**（dynamic programming）はベルマン[8)]（R. Bellman）によって提案された，最適経路問題，在庫管理等多くの最適化問題に適用可能な手法である。これはつぎの二つの原理に基づいている。

・**不変埋込みの原理**（principle of invariant imbedding）

解くべき原問題に対し，その原問題を含むような部分問題群を考える（すなわち，原問題を部分問題群に埋め込む)。その部分問題群に属する問題間の関係により，各問題を順に解いていく。部分問題群の問題をすべて解くことにより，原問題を解いたことになる。

この問題間の関係を生み出す基本原理がつぎの最適性の原理である。

・**最適性の原理**（principle of optimality）

逐次的に定める決定から成る政策のうち最適となる政策はつぎの性質を持つ。

「最初の状態と決定が何であれ，残りの決定列は最初の決定により生じた状態に関して最適政策を構成する。」

この二つの原理が成り立つ問題は動的計画法を適用することができ，逐次的に問題を解くことができる。例として，つぎの最短経路問題を考えよう。

---

**例 1.1（最短経路問題）**　$n$ 個の地点（$N = \{1, 2, \cdots, n\}$）があり，地点 $i \in N$ から地点 $j \in N$ へ直接移動するための所要時間は $t_{ij}$ であるとする。ここで $t_{ij} > 0$ である。ただし，直接移動する路（みち）がない場合は $t_{ij} = \infty$ とする。このとき，地点 1 から各地点に最短時間で到達する経路を求めよ。

---

各地点を節点（node）として，地点 $i$ から地点 $j$ への直接移動する路がある場合，有向枝（arc, directed edge）を節点 $i$ から $j$ に引く。これを枝 $(i, j)$ とする。これにより地点と路をグラフ $(N, A)$ で表現することができる。ここで $N$ は節点の集合，$A$ は枝の集合である。例を**図 1.1** に示す。

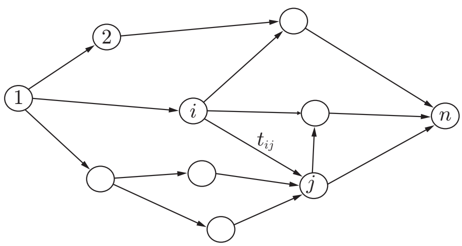


**図 1.1**　最短経路問題

地点 1 から地点 $j$ への最短時間を $f_j$ とする。また高々 $k$ ステップで $j$ に到達する経路のうち最短時間となる路の所要時間を $f_j^k$ とする。そのような路が存在しなければ $f_j^k = \infty$ である。このとき, 解くべき原問題は各 $j = 2, 3, \cdots, n$ について $f = \{f_j^\infty\}$ を求める問題である。また, 高々 $k$ ステップで各地点に到達する最短時間を求める部分問題を $f^k = \{f_j^k; j = 2, 3, \cdots, n\}$ $(k = 1, 2, \cdots)$ とする。

すべての枝 $(i, j) \in A$ について $t_{ij} > 0$ であるとき，じつは高々 $k = n - 1$ まで求めれば十分である。なぜなら，$k \geqq n$ のとき，$k$ ステップで地点 1 から到達する経路について，少なくとも 2 回同じ点を通ることになる（最初の地点 1 を含め $k + 1 (\geqq n + 1)$ 個の節点が経路上に存在し，節点は $n$ 個しか存在しないため

である)。一方，$t_{ij} > 0$ であることから，その同じ 2 点間を通る部分経路を取り去った経路は，元の $k$ ステップの経路より真に短い時間で $j$ に到達する。したがって，$n$ ステップ以上の経路を考えることは必要ない。すなわち，部分問題群 $\{f^k; k = 1, 2, \cdots, n-1\}$ を求めることにより，原問題 $f$ を求めることができる。

この部分問題群について調べる。$f_j^1$ はただちにつぎの式で求められる。

$$f_j^1 = t_{1j}, \quad j = 2, 3, \cdots, n$$

問題群 $f^{k-1}$ と $f^k$ の間にはつぎの関係がある。$f_j^k$ を達成する高々 $k$ ステップの点 $j$ への最適経路について，その最後の枝が $(i, j)$ であるとする。この最適路の 1 から $i$ までの経路は，高々 $k-1$ ステップの 1 から $i$ への路の中で最短時間で到達できる路になっている（もしそれより短い時間で行く経路が存在すると，その路と枝 $(i, j)$ から成る路の方が，先に示した高々 $k$ ステップの路のうちの最適経路より短くなり矛盾する)。このことから，高々 $k-1$ ステップの各地点 $i$ への最短経路 $f_i^{k-1}$ と，直接の路 $t_{ij}$ の和を求めていき，その中で最小となる値を達成する路の所要時間が $f_j^k$ となることがわかる。

この議論より，つぎの漸化式を導くことができる。min は最小値を表す。

$$f_j^k = \min_{i=2,3,\cdots,n} \{f_i^{k-1} + t_{ij}\},\ j = 2, 3, \cdots, n,\ k = 2, 3, \cdots, n-1 \quad (1.1)$$

式 (1.1) を用いて $k = 2, 3, \cdots$ の順に $f_j^k$ を求めたとき，$f_j^{n-1}$ が 1 から $j$ への最短経路の長さとなる。なお，より効率的に最適経路を求める方法としてダイクストラ法が知られている（例えば柳浦，茨木[5)†]を見よ)。

動的計画法による定式化に関する他の例として，つぎの有限時間最適制御問題を考える。

---

**例 1.2**（有限時間最適制御問題）　時刻 $0, 1, 2, \cdots, T$ の有限期間から成り，時刻 $T$ を除く各時点 $t \in \{0, 1, \cdots, T-1\}$ において，状態 $x_t$ を観測し，決定 $y_t \in A_t(x_t)$ を行う。ここで，$A_t(x_t)$ は時刻 $t$ において状態 $x_t$ のとき

† 肩付き数字は，巻末の引用・参考文献番号を表す。

# 索　　　　引

—— 著 者 略 歴 ——

1986年　京都大学工学部数理工学科卒業
1988年　京都大学大学院工学研究科修士課程修了（数理工学専攻）
1988年　名古屋工業大学助手
1997年　博士（工学）（名古屋工業大学）
1997年　名古屋工業大学講師
2001年　名古屋工業大学助教授
2006年　名古屋工業大学教授
　　　　現在に至る


**マルコフ決定過程 ——理論とアルゴリズム——**
Markov Decision Processes ——Theory and Algorithms——



2019 年 4 月 5 日　初版第 1 刷発行

検印省略

著　　者　中出康一（なかで こういち）
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　http://www.coronasha.co.jp

ISBN 978–4–339–02834–8　C3355　Printed in Japan　（横尾）