千葉大学大学院　融合理工学府
令和元年度修士論文

飲食店における客席案内戦略
―マルコフ決定過程による最適化と強化学習―

令和２年３月提出

指導教員　塩田 茂雄 教授

地球環境科学専攻　都市環境システムコース
野田　脩平

Abstract

This paper discusses the seat guidance strategy at restaurants. The transition in the seat usage state is modeled by a continuous-time Markov chain. In terms of guest accommodation effect, various guidance strategies were evaluated.

When a customer comes to the store, there is volatility and it depends on the customer's guidance method. In this paper, this method is defined as "seat guidance strategy" and the efficiency of the four seat guidance strategies "*random*","*sittingcloser*","*equal random*","*equal sittingcloser*" is considered. "equal" means that a seat guidance strategy where the probability of not being able to enter the store when visiting is equal for all customers.

Numerical calculations and simulations have shown that the seat guidance strategy of "sitting closer" is efficient. On the other hand, the average seat occupancy of the seat guidance strategy of "equal" was very low.

In this model, since the transition probability matrix is determined depending on the seat guidance strategy, it is known that the optimal guidance strategy can be derived by the Markov decision process. We used Reinforcement Learning to derive an optimal guidance strategy from measured values, and examined its accuracy and efficiency. With the new method, it became possible to accurately determine the optimal guidance strategy from the measured values without information about the target.

概要

飲食店ではピーク時間帯になると，席を詰めて座るように促すことがある．しかし，実際に席を詰めることで客収容効果がどの程度改善するかは，必ずしも明らかではない．

本論文では，客の出入りのある飲食店における客席の利用状態の変化を連続時間マルコフ連鎖でモデル化し，各種客席案内戦略を客収容効果の観点で評価した．

様々な人数のグループ客が来店する飲食店の客収容効果は，「客を来店時にどの空いているテーブルに案内するか」という戦略に依存する．本論文では，この客の店内への案内方法を「客席案内戦略」と呼ぶこととし，特に，「ランダム案内」・「席詰め案内」・「公平ランダム案内」・「公平席詰め案内」の４つの客席案内戦略の効率性について考察した．ここで，公平とは来店時に入店できない確率（ブロック率）が全ての客に対して等しくなる客席案内戦略である．

グループ客（1 人客を含む）がポアソン過程に従って来店し，滞在時間が指数分布に従うとして，客席案内戦略をマルコフ決定過程により最適化する手法を検討した．また，システムに関する情報を用いずに，強化学習により観測ベースで客席案内戦略の最適化を行い，その精度や効率性について考察した．

数値計算およびシミュレーションを行った結果，席詰め案内がランダム案内と比べて，僅かな席利用数の増加がみられ，ブロック率を公平にすると席の平均利用数は大きく低下した．また，常識的な案内戦略に比べて，最適化によるゲインはさほど大きくないことが確認された．

観測時間が十分長く取れるならば，測定ベースで強化学習により客席案内戦略を改善することが可能であり，対象システムを既知としてマルコフ決定過程により解析的に客席案内戦略を最適化した場合と遜色ない結果が得られることを確認した．また，解析的なアプローチが困難な対象に対しても，強化学習は有効であることも確認された．

目次

第 1 章

序論

1.1　研究の背景

近年，SNS（ソーシャル・ネットワーク・サービス）の流行により，企業や個人店は SNS を用いることで集客やブランディングを低コストで行うことができるようになった [1]．特に，飲食業界は競合相手が多く，顧客獲得は必要であり，客数が増えることは良いことであるがデメリットもある．例えば，飲食店では、駐車待ち，座席案内待ち，食事注文待ち，食事提供待ち，お会計待ち，精算待ちなど（列形成されていないものを含め）多くの待ち行列が存在する．特に座席案内待ちと食事提供待ちがボトルネックになっており、座席数を増やすまたは、キッチンを増やしたりスタッフの能力を上げると待ちが減るといった研究結果 [2] や，お持ち帰りのみの飲食店の場合の待ち行列やスタッフ配置に関する研究結果 [3]， [4] などがある．しかし，一般的に店を拡大し，座席の追加，キッチンの拡張，従業員の教育や増員にはコストがかかる．そこで，飲食店では，客数が増えているが席数を増やせない場合，客収容の効率化が重要であると考えられる．

筆者が飲食店で働いている際にも実感しているが，来店客の案内という仕事はとても重要である．具体的には，予約の有無，人数や希望の席種を確認し，店内状況に合わせて，条件に一致したお客を随時案内していく仕事である．来店客をうまく案内しなければ，待ち時間増加による満足度の低下，店内回転率の低下などにより，主である売上 (利益) に多大な影響を及ぼす危険がある．また，増加した来店客に対応するために，フロアの拡張やテーブル・イスの増量など店内容量を増やすことはコストがかかってしまうが，案内方法の変更にはコストがかからない．お店の外で行列を成す店，発券機で呼び出しを行うお店や，完全予約制のお店など，来店客への対応は様々あるが，増加した来店客を最適に案内することは，飲食業界の課題であり，効果的でコスト効率のよい仕事である．

1.2　既存研究と本研究との関係

客席案内に関する研究として，店内のテーブル席に対してどの席から案内していくか４つの方法を提案し，シミュレーションにより待ち時間に関して評価した研究 [5]，店舗利益を最大にするためのテーブルサイズとテーブル数の組み合わせをシミュレーションにより導く研究 [6]，[7] などテーブルの組合せを工夫して客席案内を考える研究，客席状況から退店時刻を予測し、予約情報を用いて客席案内を提案するシステムの研究 [8] や，顧客の退店時刻を予測し仮想上で案内することにより，座席案内効率をあげ待ち時間を均等化する研究 [9] など，滞在時間から客席案内を考える研究，座席案内をする人をプレイヤとしたシミュレーションゲームを用いて勘や経験に基づき学習させていき案内戦略を引き出す研究 [10] などがある．

収容効率に関する研究として，通信の分野で，帯域の異なる通信を同一網で収容するときの最適な受付制御を導出する研究がある [11]， [12]．滞在時間の予測や，テーブル席の組み合わせの最適化によって様々な数の来店客を案内する研究は多くみられるが，そのほとんどは収益管理や待ち時間に関する研究であり，客収容効果がどの程度であるかは必ずしも明らかになっていない．様々な人数のグループ客が来店する飲食店の客収容効率は，「客を来店時にどの空いているテーブルに案内するか」という戦略に依存する．

本稿では，客の出入りのある飲食店をマルコフシステムとしてモデル化し，客収容効果の観点から各種の客席案内戦略の効率性を評価する．特に，グループ客（１人客を含む）がポアソン過程に従って来店し，滞在時間が指数分布に従うとして，客席案内戦略をマルコフ決定過程により最適化する手法を検討する．また，システムに関する情報を用いずに，強化学習により観測ベースで客席案内戦略を最適化する手法の有効性について評価する．

1.3　論文の構成

以下，本稿の構成内容を述べる．

第 2 章 待ち行列について述べる．
第 3 章 モデルの定式化を行う．
第 4 章 マルコフ決定過程による客席案内戦略の最適化について述べる．
第 5 章 強化学習について述べる．
第 6 章 各種案内戦略と最適戦略，および強化学習についてのシミュレーション実験を行い，評価する．
第 7 章 本稿のまとめと今後の課題について述べる．

第 2 章

待ち行列理論

本章では，本稿で必要となる待ち行列理論について説明する．本章作成にあたって [13] を参考にした．

2.1　待ち行列モデルとケンドールの記法

2.1.1　待ち行列モデル

待ち行列は私たちの生活にみられる現象である．評判のレストランに入るための待ち行列やテーマパークの人気アトラクションの待ち行列，タクシーに乗るための待ち行列など，私たちは普段から様々な種類の待ち行列を目にし，経験している．待ち行列は本質的に確率的な現象である．待ち行列理論を理解するためには確率論に関する基礎的な知識が必要になる．現実の待ち行列は様々であるが，数理的に待ち行列を分析するためには，待ち行列に関する用語を定め，特徴を抽象化して表現するモデルを構築する必要がある．

飲食店にできる待ち行列を例に用語を定義する．飲食店に人が訪れる動作を**到着**と呼び，席を**窓口**と呼ぶ．窓口を訪れる人は**客**と呼ばれる．食事の提供のように窓口で受ける仕事を**サービス**と呼ぶ．一般に店内には複数の席があるように窓口が複数あっても構わない．行列を作って待つ場所を**待ち室**と呼ぶ．現実には待ち室は有限の広さであるが，抽象的な待ち行列モデルでは無限の広さの待ち室を考えることも多い．窓口でサービスを受け終わった客は待ち行列から**退去**する．待ち室と窓口を含む全体を**待ち行列システム**と呼ぶこととする（図 2.1）．

待ち行列理論では客の到着を**到着過程** (2.2 節) でモデル化する．客が窓口で要求するサービスの量も客によって異なるため，サービス量は**確率変数**とみなす．サービス量の単位が用いられることが多いため，サービス量のことを**サービス時間**と呼び，サービス時間の**分布関数**（**サービス時間分布**）を指定する．

最後にサービス規律について述べる．**先着順サービス**は **FIFO**（First In First Out）もしくは **FCFS**（First Come First Served）と記載する．なお，複数の窓口がある場合は，先着順サービスであっても到着順序と退去順序が一致するとは限らないため，FCFS と記載する方がより正確である（図 2.2）．**後着順サービス**には幾つかの種類がある．新たに到着した客が，窓口でサー

ビス中の客のサービスを中断せずに，待ち行列の先頭に並ぶサービス規律は**非割込み後着順サービス**と呼ばれ，**LIFO-NP**（Last In First Out - Non Preemptive）もしくは **LCFS-NP**（Last Come First served - Non Preemptive）と記載する（図 2.3）．新たに到着した客が窓口でサービス中の客のサービスを一時中断して，窓口でサービスを受け始まるものもある．サービスを中断された客は行列の先頭に並びなおし，サービス再開時にはそれまで受けたサービスの続きを受ける．このサービス規律は**割込継続型後着順サービス**と呼ばれ，**LIFO-PR**（Last In First Out - Preemptive Resume）もしくは **LCFS-PR**（Last Come First Served - Preemptive Resume）と記載する．

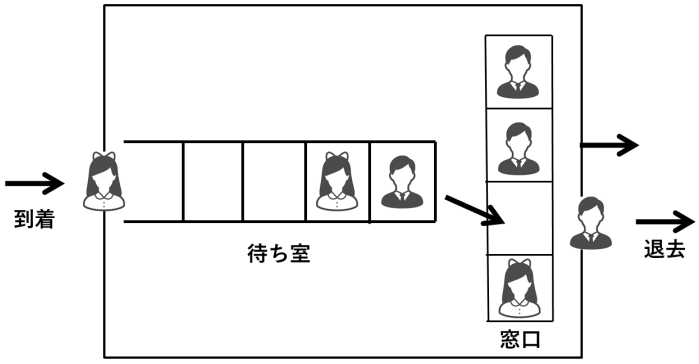

図 2.1　待ち行列システム

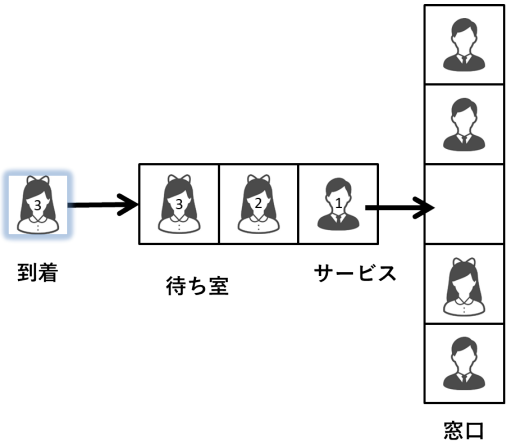

図 2.2　先着順サービスの例 FCFS.

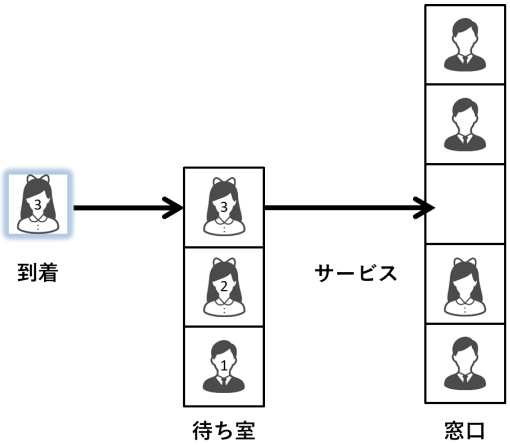

図 2.3　後着順サービスの例 LCFS.

表 2.1　ケンドールの記法.

記号	到着過程	サービス時間分布
M	ポアソン過程	指数分布
D	到着間隔が一定	サービス時間が一定
E_k	到着間隔が k 次アーラン分布	サービス時間が k 次アーラン分布
H_k	到着間隔が k 次超指数分布	サービス時間が k 次超指数分布
GI	再生過程	一般分布（互いに独立）
G	一般の到着過程（再生過程も含む）	一般分布（独立性を必ずしも仮定しない）

2.1.2　ケンドールの記法

待ち行列モデルは (1) 客の到着過程，(2) サービス時間分布 (3) 窓口数 (4) 待ち室容量（待ち室に収容可能な最大人数）により特徴づけられる．これに加えて，電話網のモデル化などでは (5) **呼源数**（電話をかける可能性のある顧客数）を与える場合もある．さらに，(6) **サービス規律**と呼ばれる，サービスの順番に関する規則（先着順など）を指定する必要がある．待ち行列理論では，これらの待ち行列モデルの特徴を以下の**ケンドールの記法**により表現する．

$$\mathrm{A}(N)/\mathrm{S}/c(K)$$

ここで，A には客の到着を表す確率過程の種類，S にはサービス時間分布の種類を表す記号が入る．また，N には呼源数，c には窓口数，K には待ち室容量を表す数字が入る．なお，N が無限大の場合は (N) の部分，K が無限大の場合は (K) の部分を省略する．
ケンドールの記法で用いる到着過程やサービス時間分布の種類とそれを表す記号を表 2.1 にまとめる．

2.2　到着過程とサービス時間分布

客がポアソン過程に従って到着し，サービス時間が指数分布に従う待ち行列モデル (M/M/c 型モデル) では，待ち行列の時間発展は現在の系内客数にのみ依存し，直近の客の到着時点や経過サービス時間などに影響されない．この性質のため M/M/c 型モデルは容易に解析できる．これは到着間隔やサービス時間を規定する指数分布の無記憶性による．また，到着間隔もしくはサービス時間分布を指数分布以外の分布に置き換えると，多くの場合，解析可能性が壊れてしまう．そのため，**アーラン分布**や**超指数分布**など，無記憶性の拡張であるマルコフ性をもった分布での解析が提案されている．本節では後の章で必要となる指数分布，ポアソン過程，アーラン分布を説明する．

2.2.1　基本概念

客の待ち行列への到着を表す確率過程を**到着過程**と呼ぶ．本節では，時間区間 A に到着する客数を $N(A)$，特に時間区間 $(a,b]$ に到着する客数を $N(a,b]$ で表すこととする．また，n 番目の客の到着時刻を T_n で表し，n 番目の客と $n+1$ 番目の客の到着間隔 $T_{n+1}-T_n$ を τ_n で表す．さらに，

$$\cdots < T_{-2} < T_{-1} < T_0 < T_1 < T_2 < \cdots$$

を満たすように，客には番号が振られているとする（図 2.4）．

客の到着時刻列 $\{T_n\}$ と $N(0,t]$ との間には密接な関係がある．例えば，n 番目の客が時刻 t までに到着している事象 $\{T_n \leqq t\}$ と，時間区間 $(0,t]$ に少なくとも n 人の客が到着している事象 $\{N(0,t] \geqq n\}$ は等価である ($\{T_n \leqq t = N(0,t] \geqq n\}$)．また，時間区間 $(0,t]$ に到着した客の到着時刻は t 以下であることなどから，$T_{N(0,t]} \leqq t$ や $T_{N(0,t]+1} > t$ が得られる．さらには，

$$T_n = \inf\{t > 0 | N(0,t] = n\}$$

が成立する．つまり，$N(A)$ の情報から，到着時刻列 $\{T_n\}$ が決定できる．

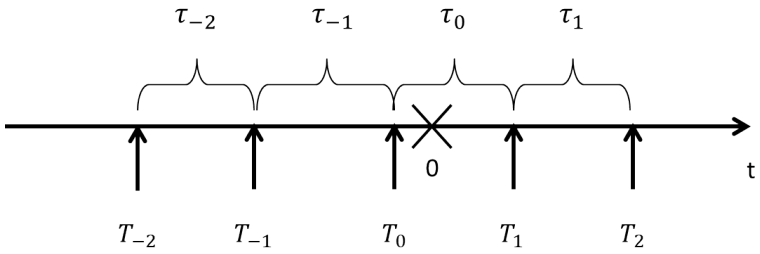

図 2.4　客番号の振り方．

一方，客が窓口で費やす時間を**サービス時間**と呼ぶ．待ち行列モデルではサービス時間を確率変数とみなし，n 番目の客のサービス時間を σ_n で表す．サービス時間の分布関数 $P(\sigma_n \leqq x)$ は通常 n によらないので，本節でもそのように仮定する．

2.2.2　指数分布

確率変数 X が以下の分布関数に従うとき，X は（パラメタ μ）の**指数分布**に従うという．

$$P(X \leqq t) = \begin{cases} 1 - \mathrm{e}^{-\mu t}, t \geqq 0 \\ 0. t < 0 \end{cases}$$

指数分布は到着間隔とサービス時間分布の分布関数として，待ち行列理論において頻繁に登場する基本的な分布関数である．確率変数 X が（パラメタ μ）の指数分布に従うならばその期待値は $1/\mu$ に等しい．

$$E[x] = \int_0^\infty P(X > t)dt = \int_0^\infty \mathrm{e}^{-\mu t}dt = 1/\mu.$$

確率変数 X がサービス時間を表すとする．時間 t が経過した時点でサービスが未終了のとき，そこからさらに時間 x が経過した時点でサービスが終了しているか確率を考える．時間 t が経過した時点でサービスが未終了である確率は $P(X > t)$ であるため，求めたい確率は $P(X \leqq x + t | X > t)$ に等しい．確率変数 X が寿命を表すならば，$P(X \leqq x + t | X > t)$ は現在の年齢が $t(\geqq 0)$ の人の残りの寿命が $x(\geqq 0)$ 以下である確率なので，これを**残余寿命分布**と呼ぶこととする．指数分布の残余寿命分布は，以下の性質を有する．

補題 2.2.1. X が期待値 $1/\mu$ の指数分布に従うならば，経過時間にかかわらず残余寿命分布も期待値 $1/\mu$ の指数分布に従う．すなわち $x \geqq 0, t \geqq 0$ のとき

$$P(X \leqq x + t | X > t) = 1 - \mathrm{e}^{-\mu x}.$$

補題 2.2.1 は，指数分布の残余寿命分布がこれまでの経過時間に依存しないという意味での**無記憶性**を表している．

2.2.3　ポアソン過程

到着間隔が指数分布に従う再生過程を**ポアソン過程**と呼ぶ．ポアソン過程は次の性質で特徴づけられる．

(1)　任意の自然数 n と任意の $0 < t_1 < t_2 < \cdots < t_n < \infty$ に対して，

$$N(0, t_1], N(t_1, t_2], \ldots, N(t_{n-1}, t_n].$$

は独立な確率変数列となる．

(2)　各 $t \geqq 0$ において，$N(t)$ は 0 または 1 に等しい．

(3)　各 $s, t \geqq 0$ に対して，$N(s, s+t]$ の分布は t のみ依存する．

ポアソン過程は，「ある時刻以降の到着客数や到着時刻は，その時刻以前の客の到着履歴とは無関係である」という意味での無記憶性を有する．条件 (1) は，数学的に無記憶性を表す．条件 (2) は，同時に二人以上の到着が生じないことを示す．この条件を満たす到着過程を**単純**な到着過程と呼ぶ．条件 (3) は，客の到着頻度が時間の経過とともに変動せず一定であることを示す．なお，条件 (1)，(2)，(3) を満たす到着過程はポアソン過程以外には存在しない．
次の補題はポアソン過程の名前の由来となる（任意区間の到着客数がポアソン分布に従うという）事実を示している．

補題 2.2.2. 到着率が λ のポアソン過程の場合，任意の整数 $n \geqq 0$ と実数 $s, t \geqq 0$ について

$$P(N(s, s+t] = n) = \frac{(\lambda t)^n}{n!} \mathrm{e}^{-\lambda t}.$$

2.2.4　k 次のアーラン分布

パラメタ μ の指数分布に従う独立な確率変数列 $H_i (i = 1, 2, \ldots)$ に対して，これらの和 $F_k = H_1 + \cdots + H_k$ と定義し，F_k の従う確率分布を見い出す．F_k の分布関数を $F_k(x) = \Pr(F_k \leqq x)$ とすれば

$$F_2(x) = \int_0^x (1 - \mathrm{e}^{-\mu(x-y)}) \mu \mathrm{e}^{-\mu y} dy = 1 - \mathrm{e}^{-\mu x} - \mathrm{e}^{-\mu x} \mu x.$$

となる．同様に，

$$F_3(x) = \int_0^x F_2(x-y) \mu \mathrm{e}^{-\mu y} dy = 1 - \mathrm{e}^{-\mu x} - \mathrm{e}^{-\mu x} \mu x - \mathrm{e}^{-\mu x} \frac{(\mu x)^2}{2}.$$

であり，帰納法によって

$$F_k(x) = 1 - \sum_{i=0}^{k-1} \mathrm{e}^{-\mu x} \frac{(\mu x)^i}{i!}, \quad k = 1, 2, \ldots. \tag{2.1}$$

となることを示すことができる．さらに $F_k(x)$ を微分することにより，F_k の密度関数 $f_k(x)$ は

$$f_k(x) = \frac{(\mu x)^{k-1}}{(k-1)!} \mathrm{e}^{-\mu x} \mu \tag{2.2}$$

で与えられることがわかる．k 個の独立なパラメタ μ をもつ指数分布に従う確率変数の和が従う確率分布を k 次の**アーラン分布**と呼ぶ．その分布関数ならびに密度関数は式 2.1 ならびに式 2.2 で与えられる．また，平均は k/μ で与えられ，分散は $(k/\mu)^2/k$ で与えられる．
k 個の独立なパラメタ μ をもつ指数分布に従う確率変数の和は，率 μ のポアソン過程において k 人の客が到着するまでの時間間隔に等しい．すなわち，時間間隔 $(0, t]$ の間に到着する客数を $A(0, t]$ とすると，

$$F_k \leqq x \Leftrightarrow A(0, t] \geqq k$$

が成立する．よって，

$$\Pr(F_k \leqq x) = \Pr(A(0, x] \geqq k) = 1 - \Pr(A(0, x] \leqq k - 1)$$

となり k 次のアーラン分布の分布関数が式 2.1 で与えられることが確認できる．

ここで，アーラン分布の変動係数は $1/\sqrt{k}$ となるため，ステージ数 k が大きくなると，変動係数は 0 に近づいていく．次数が高くなるに従って，分散が減少する．また，アーラン分布の分散は，常に同じ平均をもつ指数分布よりも小さくなる．

2.3　出生死滅過程と待ち行列モデル

客の到着過程とサービス時間がそれぞれポアソン過程と指数分布で特徴づけられる待ち行列モデルは，出生死滅過程と呼ばれる確率過程で記述される．そこでまず出生死滅過程について述べて，本稿での待ち行列モデルについて解説する．
なお，連続時間マルコフ連鎖については [14] を参照してほしい．

2.3.1　出生死滅過程

時刻 $t \geqq 0$ における，ある生物の個体数を $X(t)$ で表すとする．個体数は非負の整数であるので，$X(t)$ は集合 $\mathbb{N}_0 = \{0, 1, \ldots\}$ 上に値をとる．この生物の個体数は確率的に増減し，$\{X(t); t \geqq 0\}$ は確率過程として捉えられるとする．さらに $\{X(t); t \geqq 0\}$ は，現在の個体数 i にのみ依存し，過去の個体数とは無関係に，将来の個体数が決まると仮定する．すると，$\{X(t); t \geqq 0\}$ はマルコフ性を有すると考えられる．ここで，この生物の個体数については，一度に高々一つだけ増えるか，あるいは減るかのどちらかであるとする．この推移構造をもつ連続時間マルコフ連鎖 $\{X(t); t \geqq 0\}$ を**出生死滅過程**と呼ぶ．出生死滅過程はその推移の構造上，ある状態 $i > 0$ からは隣り合う状態，すなわち状態 $i + 1$ と状態 $i - 1$ にのみ推移が許される（図 2.5）．

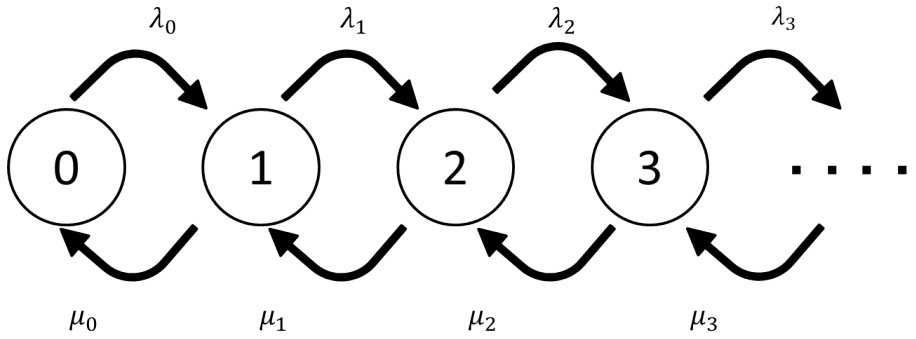

図 2.5　出生死滅過程の状態推移図．

2.3.2　M/M/c(0) モデル

本節では窓口の数は c 個とするが，待ち室は全く考えない場合を考える．客の到着過程は到着率 λ のポアソン過程に従うとする．各窓口でのサービス時間は互いに独立であり，かつ同じ平均 $1/\mu$ をもつ指数分布に従うとする．到着時点で系内客数が c である場合は到着した客はすべての窓口で客がサービスを受けている状況に遭遇する．このとき，到着客は待ち室が存在しないので，待つことができず，サービスを受けずに直ちに退去すると仮定する．その意味で，この待ち行列モデルは損失系と呼ばれる．待ち行列モデルといいながら待ち行列が形成されないのでやや奇異ではあるが，特別な待ち行列モデルと考える．また視点を変えると，受け入れられた客は待たされることなく即座にサービスを受けられることから，即時式とも呼ばれる．この待ち行列モデルを M/M/c(0) モデルという．

M/M/c(0) モデルを出生死滅過程でモデル化する．出生死滅過程の出生率は $\lambda_i = \lambda (i = 0, 1, \ldots, c-1)$ であり，$\lambda_i = 0 (i = c, c+1, \ldots)$ とする．死滅率は $\mu_i = i\mu (i = 1, 2, \ldots, c)$ である．系内客数は c を超えることができないので，定常分布は 0,1,...,c 上で定義される．
到着率が λ，サービス率が μ の M/M/c(0) モデルの定常分布 $\pi_j (j = 0, 1, \ldots, c)$ は，$a = \lambda/\mu$ とする次のように与えられる．

$$\pi_0 = \left[\sum_{j=0}^{c} \frac{a^j}{j!} \right]^{-1}, \pi_j = \frac{a^j}{j!}\pi_0, \quad j = 1, 2, \ldots, c.$$

サービス時間の平均値が $1/\mu$ に等しい一般分布とした場合でも成立することが知られている．サービス時間の分布によらず，その平均値のみに依存する性質を**不感性**（insensitivity）という．損失系である M/M/c(0) モデルにおいて，基本となる性能評価指標は，客が到着したとき c 個の窓口すべてがサービス中であり，ただちに退去する確率，すなわち損失率である．客の到着時点での系内客数分布が必要となるが，到着時点での系内客数分布は定常分布に等しい．よって，M/M/c(0) モデルの損失率は系内客数が c である定常分布 π_c に等しいことになる．すなわち，損失率を $B(c, a)$ とすると

$$B(c, a) = \frac{a^c/c!}{\sum_{j=0}^{c} a^j/j!}. \tag{2.3}$$

式 2.3 は**アーランの損失式**（アーラン B 式）と呼ばれる．一般に，現実のシステムにおいて到着過程をポアソン過程でモデル化することは正当化できることが多いが，サービス時間を指数分布でモデル化することは粗い近似となる場合が多い．アーランの損失式はサービス時間の確率分布についてはその平均値のみに依存するので，適用範囲が広く，極めて重要な式といえる．

第 3 章

モデルの定式化

3.1　モデル

K 個のカウンタ席のある飲食店に n 人客（n 人からなるグループ；$1 \leqq n \leqq N$ 本稿では簡単化のため $n = 1, 2, 3$ とする．）が到着率 λ_n で来店し，店内の滞在時間は平均 h_n とする（図 3.1）．

3.1.1　客席利用状態

n 人客は，必ず隣り合ったカウンタ席に座るものとする．カウンタ席には左から順に 1 から K までの番号をふる．更に，n 人客が客席を占める各パタンに以下で定義する番号を付与する．

$$s_n \stackrel{\text{def}}{=} \sum_{k=1}^{K-n+1} 2^{k-1} L_n^{(k)}.$$

ここで $L_n^{(k)}$ は番号 k から $k+n-1$ までの座席を 1 組の n 人客が占める場合に 1，それ以外の場合に 0 をとる変数である．客席利用状態は N 個の数字の組 $\mathbf{s} = (s_1, \ldots, s_n)$ で識別できる．

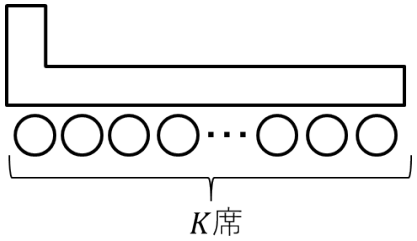

図 3.1　カウンタ席が横一列に並ぶ飲食店．

例えば，カウンタ席 $K = 6$，$n = 1, 2, 3$ の場合を考える (図 3.2).

図 3.2 のとき，$k = 1$ のカウンタ席は 1 人客，$k = 2, 3, 4$ のカウンタ席は 3 人客，$k = 5, 6$ のカウンタ席は 2 人客がそれぞれ一組ずつ座っている．
この場合 $L_1^{(1)} = 1, L_2^{(5)} = 1, L_3^{(2)} = 1$ となり，

$$s_1 = 2^{1-1} = 2^0 = 1, \quad s_2 = 2^{5-1} = 2^4 = 16, \quad s_3 = 2^{2-1} = 2^1 = 2$$

従って，この客席利用状態は $\mathbf{s} = (1, 16, 2)$ と表すことができる．

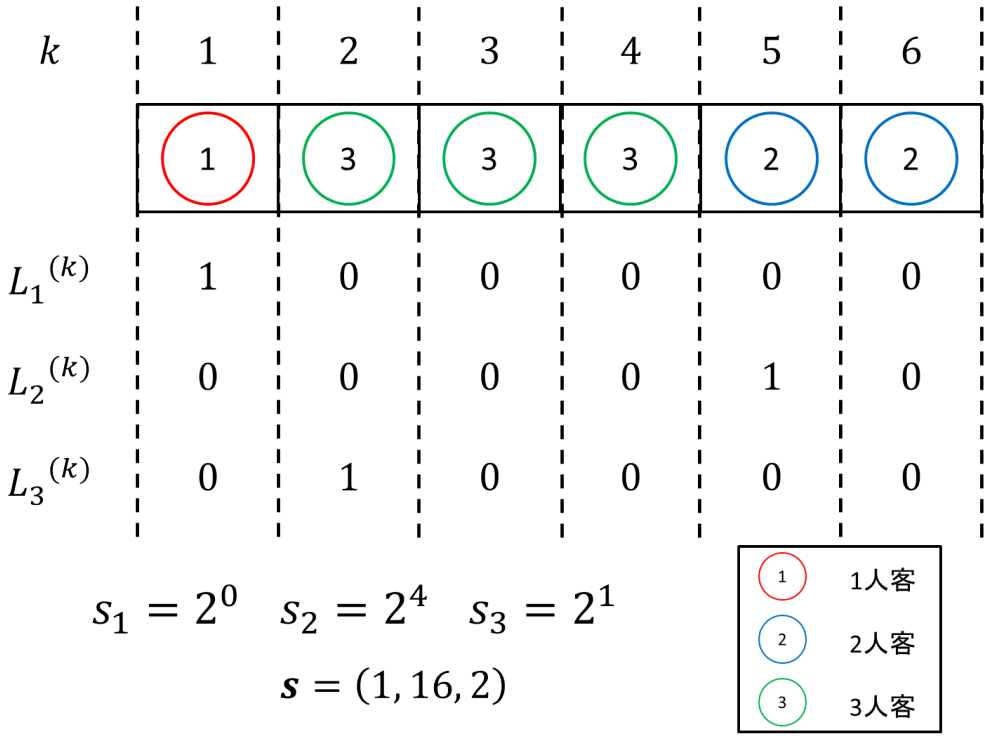

図 3.2　客席利用状態の識別．

3.1.2　案内戦略

来店時に座れる席がなかった，もしくは案内を断られた客は，席が空くことを待たずに退去するものとする．これを本稿ではブロックと呼び，客が来店時に案内できない確率をブロック率 bp_n とする．このモデルは損失系待ち行列モデルといえる．
客の来店時，もしくは退去時に客席利用状態 (s_1, s_2, s_3) の遷移が生じる．退去時の遷移先は一意に定まる．一方，客の来店時の遷移先には任意性がある．

例えば，カウンタが 3 席の店において，状態 (0,0,0)（店内に誰もいない状態）で，1 人客が来店した場合の遷移先は，
(1,0,0)（左端に案内），(2,0,0)（真ん中に案内），(4,0,0)（右端に案内），(0,0,0)（入店を断る）の 4 通り存在し，これらは客の店内への案内方法に依存する．

本稿では，この客の店内への案内を「客席案内戦略」と呼ぶこととする．客席案内戦略は来店時の客席利用状態にのみ依存することに注意する（マルコフ性を有する）．
特に次の 4 つの客席案内戦略の効率性について考察する．

1. ランダム案内 (random)
   案内可能な席の中から，ランダムに（等確率で）選んだ席に案内する（図 3.3）．
2. 席詰め案内 (sitting closer)
   案内可能な席の中で，最も左端に近い席に案内する（図 3.4）．

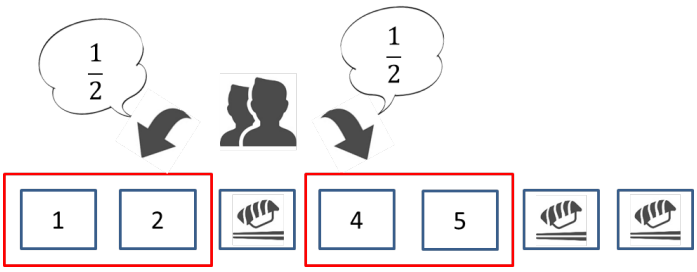

図 3.3　2 人客の場合，ランダム案内．

図 3.4　2 人客の場合，席詰め案内．

3. 公平ランダム案内 (equal random)

　3 人客の案内可能席が存在する場合のみ，案内可能な席の中から，ランダムに（等確率で）選んだ席に案内する（図 3.5）．

4. 公平席詰め案内 (equal sitting closer)

　3 人客の案内可能席が存在する場合のみ，案内可能な席の中で，最も左端に近い席に案内する（図 3.6）．

　ここで，3 と 4 は来店時に入店できない確率（ブロック率 bp_n）が 1 人客，2 人客，3 人客の全てで等しくなる（＝公平）客席案内戦略である．つまり，3 人客が入店できない状態のとき，1 人客，2 人客，3 人客の入店を断る（図 3.7）．

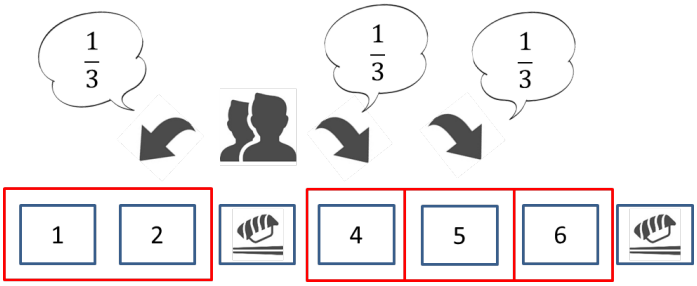

図 3.5　2 人客の場合，公平ランダム案内．

図 3.6　2 人客の場合，公平席詰め案内．

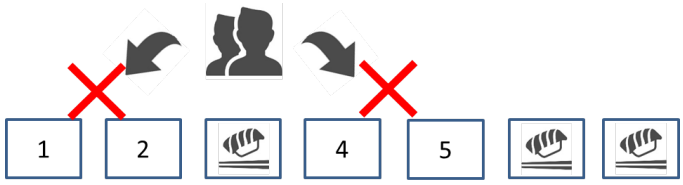

図 3.7　2 人客の場合，公平客席案内．

第 4 章

最適化

4.1　マルコフ決定過程

マルコフ決定過程（MDP）は「**状態**」，「**行動**」，「**遷移確率**」，「**報酬**」の 4 つの要素で構成される．環境（マルコフ性を有する）のとりうる状態の集合を $S = \{s_1, s_2, \ldots, s_n\}$，エージェント（学習と意思決定を行う者）がとりうる行動の集合を $A = \{a_1, a_2, \ldots, a_n\}$ と表す．環境中のある状態 $s \in S$ において，エージェントがある行動 a を実行すると，環境は確率的に状態 $s' \in S$ へ遷移する．その遷移確率を $P^a_{ss'} = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$ とする．現在の状態 s と行動 a および次の状態 s' に応じて，報酬を $R^a_{ss'} = E\{r_{t+1} | a_t = a, s_t = s, s_{t+1} = s'\}$ とする．方策 π のもとに状態 s において行動 a を選択する確率を $\pi(s, a)$ とする．

図 4.1　マルコフ決定過程モデル．

ある状態において，ある行動を行うことがどれだけ良いのかを表す**価値関数**に基づいて評価を行う．方策 π のもとでの状態 s の価値 $V^\pi(s)$ は，状態 s にいて方策 π に従ったときの期待収益である．マルコフ決定過程に対する $V^\pi(s)$ の形式的定義は次のようになる．

$$V^\pi(s) = E_\pi\{R_t|s_t = s\} = E_\pi\left\{\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t = s\right\}. \tag{4.1}$$

$E_\pi\{\}$ は，エージェントが π に従うとしたときの期待値を表す．終端状態の価値は（存在するなら）常に 0 である．関数 V^π を方策 π に対する**状態価値関数**と呼ぶ．

同様に，方策 π のもとで状態 s において行動 a をとることの価値を $Q^\pi(s,a)$ で表し，状態 s で行動 a を取り，その後に方策 π に従った期待報酬として定義する．

$$Q^\pi(s,a) = E_\pi\{R_t|s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t = s, a_t = a\right\}. \tag{4.2}$$

$Q^\pi(s,a)$ を方策 π に対する行動価値関数と呼ぶ．
任意の方策 π と状態 s に対して，s の価値と後続状態群の価値との間に以下の整合性条件が成り立つ．

$$\begin{aligned}
V^\pi(s) &= E_\pi\{R_t|s_t = s\} \\
&= E_\pi\left\{\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t = s\right\} \\
&= E_\pi\left\{r_{t+1} + \gamma\sum_{k=0}^{\infty}\gamma^k r_{t+k+2}|s_t = s\right\} \\
&= \sum_a \pi(s,a)\sum_{s'} P^a_{ss'}\left[R^a_{ss'} + \gamma E_\pi\left\{r_{t+1} + \gamma\sum_{k=0}^{\infty}\gamma^k r_{t+k+2}|s_{t+1} = s'\right\}\right] \\
&= \sum_a \pi(s,a)\sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^*(s')].
\end{aligned} \tag{4.3}$$

ただし，$\forall_s \in S, \forall_a \in A(s), \forall_{s'} \in S^+$.
式 4.3 は V^π に対する Bellman 方程式である．この方程式はある状態の価値と，その後続状態群の価値との間の関係を表すもので，開始状態の価値は，期待される次状態の価値と途中で得られる期待報酬の和に等しいということを表している．

4.1.1　方策評価

任意の方策 π に対する状態価値関数 V^π を計算する方法を**方策評価**と呼ばれる．

$$\begin{aligned} V^\pi(s) &= E_\pi\{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s\} \\ &= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^*(s')]. \end{aligned} \tag{4.4}$$

$\pi(s,a)$ は方策 π の下で，状態 s で行動 a を選択する確率を表す．$\gamma < 1$ であるか，または方策 π の下で最終的にすべての状態から終端状態に至ることが保証されるなら，V^π の存在と一意性は保証される．

環境のダイナミクスが完全に知られている場合，式 4.4 は $|S|$ 個の未知変数 $(V^\pi(s), s \in S)$ を持つ $|S|$ 個の連立 1 次方程式となる．一般に，この解は簡単に求められるが，計算に時間のかかる場合もある．

式 4.4 の V^π に対する Bellman 方程式を更新規則として用いることで，連続した近似が次のように得られる．

$$\begin{aligned} V_{k+1}(s) &= E_\pi\{r_{t+1} + \gamma V_k(s_{t+1}) | s_t = s\} \\ &= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V_k(s')]. \end{aligned} \tag{4.5}$$

ただし，$\forall_s \in S.V^\pi$ に対する Bellman 方程式から，この更新式の固定点は明らかに $V_k = V^\pi$ となる．このアルゴリズムは**反復方策評価**と呼ばれる．

反復方策評価では，以前の s の価値と即時報酬の期待値を使って新たな s の価値を計算し，それを以前の s の価値と置き換え，現在評価している方策の下で可能な 1 ステップ遷移のすべてに対してこの操作を行う（**完全バックアップ**と呼ぶ）．

4.1.2　方策改善

方策に対して価値関数を計算することで，さらに良い方策を見出す手がかりが得られる．いま任意の方策 π に対して，その価値関数 V^π が求まったとする．ここで，ある状態 s に対して，ある行動 $a \neq \pi(s)$ を選択するように方策を変更すべきかを考える．状態 s で行動 a を選択し，その後は既存の方策 π に従うとする．この価値は次のように計算される．

$$\begin{aligned} Q^\pi(s,a) &= E_\pi\{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s, a_t = a\} \\ &= \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^\pi(s')]. \end{aligned} \tag{4.6}$$

この値を $V^\pi(s)$ と比べた場合の大小が重要な判断基準となる．この値が $V^\pi(s)$ より大きい場合，つまり s で一度 a を選んで，その後 π に従うことが，常に π に従う場合よりも良いならば，状態 s に対して a を常に選ぶことが良く，したがって，この新しい方策は全体的に改善さ

れるだろうと期待できる．

ある方策とその価値関数が与えられたとする．すべての状態に，すべての可能な行動に関して，各状態で $Q^\pi(s,a)$ が最良となる行動を選択するような変更が考えられる．次式で与えられ新しい**グリーディ方策** π' を考える．

$$\begin{aligned}\pi'(s) &= \arg\max_a Q^\pi(s,a) \\ &= \arg\max_a E\{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s, a_t = a\} \\ &= \arg\max_a \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^\pi(s')].\end{aligned} \tag{4.7}$$

ここで $\arg\max_a$ は，それに続く式を最大にするような a の値を与える．このグリーティ方策は V^π によって 1 ステップ先読みを行い，短期的に最良となるであろう行動を選択する．この方策は元の方策と同等かそれ以上であることが保証される．元の方策の価値関数に従ってグリーディな行動を選択していくことで，その方策を改善するような新しい方策を作り出す過程を**方策改善**と呼ぶ．

4.1.3　方策反復

V^π を使って方策 π を改善し，より優れた方策 π' を得ることができたならば，続いて π' から $V^{\pi'}$ を計算して改善を行うことで，さらに優れた方策 π'' を得ることができる．このように方策と価値関数を次々と改善する次のような系列を考えることができる．

$$\pi_0 \xrightarrow{E} V^{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} V^{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} \cdots \xrightarrow{I} \pi_* \xrightarrow{E} V^*$$

ここで $\xrightarrow{E}$ は方策評価，$\xrightarrow{I}$ は方策改善を表す．各方策は直前の方策に対し，厳密な改善となっていることが保証される．有限マルコフ決定過程の方策は有限個であるから，この過程は有限回の繰り返しで最適価値関数に収束する．

最適方策を発見するこのような手法は**方策反復**と呼ばれている．ここで，方策評価の過程はそれ自身反復計算となっているが，その開始点は以前の方策の価値関数であることに注意する．これにより通常は方策評価の収束速度が大きく加速される．方策反復は，わずか数回の繰り返しで収束する場合がある．

4.2　マルコフ決定過程による客席案内戦略の最適化

4.2.1　定常状態確率

グループ客の到着がポアソン過程に従い，サービス時間が指数分布に従うと仮定する．時刻 t における客席利用状態 $\mathbf{s}(t) = (s_1(t), s_2(t), s_3(t))$ は有限状態連続時間マルコフ連鎖に従う．客席案内戦略 a のもとでの状態 $\mathbf{s}$ の定常状態確率を $\pi_a(\mathbf{s}) = \pi_a(s_1, s_2, s_3)$ で表し，状態を辞書式に並べた定常状態確率ベクトルを $\pi_a = (\pi_a(0,0,0), \pi_a(1,0,0), \cdots)$ により定める．客席案内戦略 a のもとでの状態 $\mathbf{s}$ から状態 $\mathbf{s}'$ への推移率を $q_a(\mathbf{s}, \mathbf{s}')$ とすると以下の大域平衡方程式が成立する．

$$0 = -\pi_a(\mathbf{s}) \sum_{\mathbf{s}' \neq \mathbf{s}} q_a(\mathbf{s}, \mathbf{s}') + \sum_{\mathbf{s}' \neq \mathbf{s}} \pi_a(\mathbf{s}') q_a(\mathbf{s}', \mathbf{s}).$$

客席案内戦略 a のもとでの推移率行列を $P_a = \{q_a(\mathbf{s}, \mathbf{s}')\}$ とすると，大域平衡方程式は $\pi_a P_a = 0.$ と書ける．なお，推移率行列の対角成分 $q_a(\mathbf{s}, \mathbf{s})$ は以下で与えられる．

$$q_a(\mathbf{s}, \mathbf{s}) = -\sum_{\mathbf{s}' \neq \mathbf{s}} q_a(\mathbf{s}, \mathbf{s}').$$

大域平衡方程式を解くことにより π_a が求まり，π_a からカウンタ席の平均利用数や n 人客のブロック率 $\mathrm{bp_n}$ を算出できる．

例えば，カウンタ席 $K = 3$，$n = 1, 2, 3$，ランダム案内（r とする）の場合を考える．
客席利用状態は状態数が 13 あり，辞書式に並べると

$$\begin{aligned}(s_1, s_2, s_3) = \{&(0,0,0), (0,0,1), (0,1,0), (0,2,0), (1,0,0), (1,2,0),\\ &(2,0,0), (3,0,0), (4,0,0), (4,1,0), (5,0,0), (6,0,0), (7,0,0)\}\end{aligned}$$

客席案内戦略 r のもとでの推移率行列 P_r は

$$P_r = \begin{bmatrix}
-1.000 & 0.143 & 0.143 & 0.143 & 0.190 & 0.000 & 0.190 & 0.000 & 0.190 & 0.000 & 0.000 & 0.000 & 0.000 \\
1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.307 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.692 & 0.000 & 0.000 & 0.000 \\
0.307 & 0.000 & 0.000 & -1.000 & 0.000 & 0.692 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.307 & 0.000 & 0.000 & 0.000 & -1.000 & 0.230 & 0.000 & 0.230 & 0.000 & 0.000 & 0.230 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.599 & 0.400 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.400 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.300 & 0.000 & 0.000 & 0.000 & 0.300 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.285 & 0.000 & 0.285 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.428 \\
0.307 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.230 & 0.230 & 0.230 & 0.000 \\
0.000 & 0.000 & 0.599 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.400 & -1.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.285 & 0.000 & 0.000 & 0.000 & 0.285 & 0.000 & -1.000 & 0.000 & 0.428 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.285 & 0.000 & 0.285 & 0.000 & 0.000 & -1.000 & 0.428 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.333 & 0.000 & 0.000 & 0.333 & 0.333 & -1.000
\end{bmatrix}$$

大域平衡方程式 $\pi_r P_r = 0.$ を解くと

$$\pi_r = \begin{bmatrix}0.134 & 0.100 & 0.065 & 0.065 & 0.072 & 0.091 & 0.069 & 0.053 & 0.072 & 0.091 & 0.054 & 0.531 & 0.080\end{bmatrix}$$

n 人客のブロック率 bp_n は，n 人客がブロックされる時の客席利用状態の定常状態確率の和で表すことができる．
1 人客がブロックされる時の客席利用状態は，$(0,0,1),(1,2,0),(4,1,0),(7,0,0)$ であるから，
1 人客のブロック率 bp_1 は

$$\mathrm{bp}_1 = \pi_r(0,0,1) + \pi_r(1,2,0) + \pi_r(4,1,0) + \pi_r(7,0,0) = 0.362.$$

同様にして，2 人客のブロック率 bp_2 は

$$\begin{aligned}\mathrm{bp}_2 &= \pi_r(0,0,1) + \pi_r(0,1,0) + \pi_r(0,2,0) + \pi_r(1,2,0) + \pi_r(2,0,0) \\ &\quad + \pi_r(3,0,0) + \pi_r(4,1,0) + \pi_r(5,0,0) + \pi_r(6,0,0) + \pi_r(7,0,0) \\ &= 0.722.\end{aligned}$$

3 人客が案内される時の客席利用状態は，$(0,0,0)$ のみであるから，余事象を考える．
3 人客のブロック率 bp_3 は

$$\mathrm{bp}_3 = 1 - \pi_r(0,0,0) = 0.866.$$

平均席利用数 AO は次の式で算出できる．

$$\mathrm{AO} = \sum_{n=1}^{3} \lambda_n h_n (1 - \mathrm{bp}_n).$$

計算すると，カウンタ席 $K = 3$，$n = 1,2,3$，ランダム案内 r の場合，平均席利用数 $\mathrm{AO} = 1.881$ となる．

本研究では，この平均席利用数を用いて，客収容効果について評価していく．

4.2.2　マルコフ決定過程による最適戦略の導出

客席案内戦略はマルコフ性を有することから，マルコフ決定過程により最適案内戦略を導出することができる [15].

例えば，目的関数を売上高に取る．客は滞在時間に比例する金額を店に支払うものとし，n 人客が滞在時間に比例して支払う金額の比例定数を α_n とする．

時刻 0 で客席利用状態 $\mathbf{s}$ から出発し，戦略 a のもとでの時間 t までの損失売上高（客の入店を断ったことによる売上高減少分）の期待値 $L_a(t;\mathbf{s})$ は t が十分大きいとき，$L_a(t;\mathbf{s}) = Q(a)t + Q(\mathbf{s},a)$ となる．ここで $Q(\mathbf{s},a)$ は状態 $\mathbf{s}$ と戦略 a に依存する定数であり，状態価値関数と呼ばれる．$Q(a)$ は単位時間あたりの平均損失量，$Q(\mathbf{s},a)$ は状態 $\mathbf{s}$ と戦略 a に依存する定数であり，状態価値関数と呼ばれる．

$Q(\mathbf{s},a)$ は次の方程式（Bellman 方程式）を満たす．

$$Q(a) = \sum_n \lambda_n \left(1 - a(n;\mathbf{s})\right) \alpha_n h_n + q_a(\mathbf{s}) \left(\sum_{\mathbf{s}' \neq \mathbf{s}} p_a(\mathbf{s},\mathbf{s}') Q(\mathbf{s}',a) - Q(\mathbf{s},a) \right). \tag{4.8}$$

$$q_a(\mathbf{s}) \stackrel{\text{def}}{=} \sum_{\mathbf{s}' \neq \mathbf{s}} q_a(\mathbf{s},\mathbf{s}') = \sum_{n=1}^{N} \lambda_n \left(a(n;\mathbf{s}) + \frac{K_n(\mathbf{s})}{h_n} \right). \tag{4.9}$$

ここで，$q_a(\mathbf{s})$ は単位時間あたりの状態 $\mathbf{s}$ からの推移回数（遷移率），$K_n(\mathbf{s})$ は客席利用状態 $(\mathbf{s})$ における n 人客の組数，$a(n;\mathbf{s})$ は客席利用状態 $\mathbf{s}$ において n 人客を入店させるときに 1，入店を断るときに 0 をとる変数（案内戦略に依存する），$p_a(\mathbf{s},\mathbf{s}')$ は状態 $\mathbf{s}$ から状態 $\mathbf{s}'$ へ推移する確率である．なお，推移率行列 P_a の成分と $q_a(\mathbf{s})$，$p_a(\mathbf{s},\mathbf{s}')$ は次の関係にある．

$$q_a(\mathbf{s}) = -q_a(\mathbf{s},\mathbf{s}), \quad p_a(\mathbf{s},\mathbf{s}') = \frac{q_a(\mathbf{s},\mathbf{s}')}{-q_a(\mathbf{s},\mathbf{s})}.$$

$Q\left((0,0,0),a\right) = 0$ とおくと，$\mathbf{Q} \stackrel{\text{def}}{=} \left(Q(a), Q\left((1,0,0),a\right), Q\left((2,0,0),a\right), \ldots\right)^{\mathsf{T}}$ は $A\mathbf{Q} = \mathbf{u}$ を満たす．状態価値関数 $Q(\mathbf{s},a)$ は，客席が埋まれば埋まるほど，小さい値をとる．これは埋まれば埋まれほど，断るリスクが高くなることを意味する．直近で客を案内する利得と，将来到着する客を断るリスクを考慮した値である．

行列 A の要素 a_{ij}，ベクトル $\mathbf{u} \stackrel{\text{def}}{=} \left(u(0,0,0), u(1,0,0), \ldots\right)^{\mathsf{T}}$ の要素は以下で定まる．

$$a_{ij} = \begin{cases} -q_{ij} & i \neq 0, \\ 1 & i = 0, \end{cases} \qquad u(\mathbf{s}) = \sum_n \lambda_n \left(1 - a(n;\mathbf{s})\right) \alpha_n h_n.$$

ここで，q_{ij} は推移率行列の要素である．式 4.8 は $Q(\mathbf{s},a)$ と $Q(a)$ に関する連立方程式を解くことで状態価値関数が求まる．

状態価値関数が求まれば，客席案内戦略が決まる．適当な初期戦略から出発して状態価値関数を求め，状態価値関数に基づいて案内戦略を更新し，更新結果に基づいて再び状態価値関数を求めることを繰り返すことにより（Howard の政策反復法），最適客席案内戦略が求まる．
以下にマルコフ決定過程による最適戦略の導出手順を示す．

ステップ 1 -初期化-

初期案内戦略 a_0 を与える．

ステップ 2 -戦略評価-

戦略 a_n の連立方程式 4.8 から状態価値関数 $Q(\mathbf{s}, a)$ を計算する．

ステップ 3 -戦略改善-

状態価値関数に基づいて案内戦略を更新 $(a_n \leftarrow a_{n+1})$

状態 $\mathbf{s}$ から状態 $\mathbf{s}'$ へ推移するとすると，更新式 $\max(\alpha_n h_n + Q(\mathbf{s}', a_n) - Q(\mathbf{s}, a_n))$ で案内戦略を更新する．ここで，$(\alpha_n h_n + Q(\mathbf{s}', a_n) - Q(\mathbf{s}, a_n)) < 0$ のとき，状態 $\mathbf{s}$ に留まることを意味し，すなわち，案内を断る戦略を採る．

更新される度ステップ 2 を繰り返す

ステップ 4 -戦略決定-

更新がないとき最適案内戦略が求まる．

（Howard の政策反復法の適用）

なお，オリジナルの Howard の政策反復法 [16] では，政策はすべて決定論的政策，すなわち各状態において，ある行動を選ぶ確率が 1 で，それ以外の行動を選ぶ確率が 0 であるものに限定され，そのため状態と行動が有限な MDP では決定論的政策の個数が有限になるため，政策反復法が有限回の繰り返しで収束し，最適政策を見出すことが証明されている．

第 5 章

強化学習

5.1 強化学習の基本概念

強化学習では，目標（数値化された報酬の最大化）を達成するために，何をすべきかを相互作用から学習する．どの行動をとればより一層の報酬に結びつくかを見つけ出す必要があり，試行錯誤的な探索と遅延報酬は強化学習の特徴である．強化学習が扱う問題は，特にマルコフ決定過程として定式化される問題に密接な関係がある．前述の Howard の政策反復法などは，現在の強化学習の理論とアルゴリズムの基礎と言われている [17].

強化学習の問題を解くことは，最終的に多くの報酬獲得を達成するような方策を見つけることを意味する．有限マルコフ決定過程に対しては，最適方策を正確に定義することができる．すべての $s \in S$ に対して，$V^{\pi}(s) \geqq V^{\pi'}$ であるなら，そのときに限り $\pi \geqq \pi'$ である．他の方策よりも良いか，それに等しい方策が常に少なくとも 1 つ存在し，これが 1 つの**最適方策**である．最適方策は 1 つ以上存在するかもしれないが，全ての最適方策を π^* と記す．

環境が有限のマルコフ決定過程であることを前提とする．状態と行動の集合 $S, A(s)(s \in S)$ が有限で，そのダイナミクスが遷移確率集合 $P^a_{ss'} = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$ および即時報酬の期待値 $R^a_{ss'} = E\{r_{t+1} | a_t = a, s_t = s, s_{t+1} = s'\}$ によって決定されるものとする．ただし $\forall_s \in S, \forall_a \in A(s), \forall_{s'} \in S^+$（エピソード的な問題の場合，$S^+$ は S に終端状態を加えたもの）．強化学習では，価値関数を利用して良い方策の探索を計画し組み立てる考え方が中心となる．具体的には，以下の Bellman 最適方程式を満足させる最適な価値関数 V^* または Q^* が見つかれば，最適方策は容易に得られる．

$$\begin{aligned} V^*(s) &= \max_a E\{r_{t+1} + \gamma V^*(s_{t+1}) | s_t = s, a_t = a\} \\ &= \max_a \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V^*(s')]. \end{aligned} \tag{5.1}$$

$$
\begin{aligned}
Q^*(s,a) &= E\{r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') | s_t = s, a_t = a\} \\
&= \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma \max_{a'} Q^*(s', a')].
\end{aligned} \tag{5.2}
$$

ただし，$\forall_s \in S, \forall_a \in A(s), \forall_{s'} \in S^+$

5.2　本稿での強化学習

Bellman 方程式 4.8 は以下のように表すことができる．

$$
Q(a) = \frac{1}{T(\mathbf{s})}\left(r(\mathbf{s},a) + \sum_{\mathbf{s}' \neq \mathbf{s}} p_a(\mathbf{s},\mathbf{s}')Q(\mathbf{s}',a) - Q(\mathbf{s},a)\right). \tag{5.3}
$$

$$
r(\mathbf{s},a) \stackrel{\text{def}}{=} T(\mathbf{s})\left(\sum_n \lambda_n \left(1 - a(n;\mathbf{s})\right)\alpha_n h_n\right).
$$

ここで，$T(\mathbf{s}) \stackrel{\text{def}}{=} (q_a(a))^{-1}$ は状態 $\mathbf{s}$ から次の状態に遷移するまでの平均滞在時間，$p_a(\mathbf{s},\mathbf{s}')$ は状態 $\mathbf{s}$ から $\mathbf{s}'$ に推移する確率，$r(\mathbf{s},a)$ は状態 $\mathbf{s}$ の 1 回の滞在あたりの売上高の平均損失である．
また，式 5.3 は次式のようにも表せる．

$$
Q(a)T(\mathbf{s}) = \left(r(\mathbf{s},a) + \sum_{\mathbf{s}' \neq \mathbf{s}} p_a(\mathbf{s},\mathbf{s}')Q(\mathbf{s}',a) - Q(\mathbf{s},a)\right). \tag{5.4}
$$

式 5.4 の左辺 $Q(a)T(\mathbf{s})$ は状態 $\mathbf{s}$ に滞在している間に生じる損失量を表す．
式 5.4 の右辺において，$r(\mathbf{s},a)$ は状態 $\mathbf{s}$ の 1 回の滞在あたりの売上高の平均損失，$\sum_{\mathbf{s}' \neq \mathbf{s}} p_a(\mathbf{s},\mathbf{s}')Q(\mathbf{s}',a)$ は遷移先 $\mathbf{s}'$ での相対損失量，$Q(\mathbf{s},a)$ は状態 $\mathbf{s}$ の相対損失量を表す．
この左辺と右辺が釣り合うというのが Bellman 方程式である．
ここで，入店できなかった n 人客の単位時間あたりの平均組数 $\bar{B}_n(\mathbf{s})$ や，状態 $\mathbf{s}$ から次の状態に遷移するまで入店できなかった n 人客の平均組数 $B_n(\mathbf{s},a)$ を測定することにより，$r(\mathbf{s},a)$, $Q(a)$ は次式で推定できる．

$$
Q(a) = \sum_{n=1}^{3} \bar{B}_n(\mathbf{s})\alpha_n h_n, \quad r(\mathbf{s},a) = \sum_{n=1}^{3} B_n(\mathbf{s},a)\alpha_n h_n.
$$

本稿では

- 各状態の滞在時間（どのくらい長く滞在するか）
- 各状態からの遷移先の状態と遷移確率（どこにどのような確率で遷移するか）
- 各状態で生じる入店できない（入店を断る）客数

を測定し，式 5.3 を信じて状態価値関数 $Q(\mathbf{s},a)$ を求めて戦略を決定する方法である．
対象システムに関する事前知識を用いずに，観測ベースで必要な情報を取得し，状態価値関数と案内戦略をオンラインで更新することが可能である．

第 6 章

数値実験

6.1　種々の客席案内戦略と最適案内戦略の比較

本節では，各客席案内戦略（ 1. ランダム案内，2. 席詰め案内，3. 公平ランダム案内，4. 公平席詰め案内）のもとでの平均客席利用数を最適戦略と比較して示す．

図 6.1 において，上段は現在の客席状況（テーブル内の数字は顧客の番号）を表し，
下段は時刻別の客席状況（横軸が時間（分），縦軸がテーブル番号）を表す．

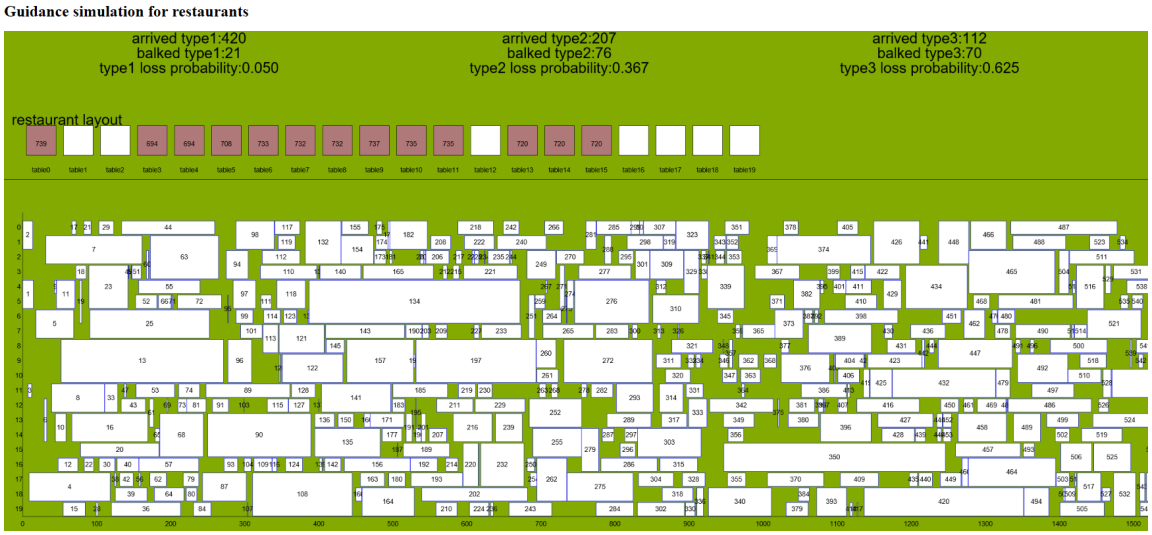

図 6.1　$K = 20$ 席，ランダム案内の場合のシミュレーション例．

表 6.1　カウンタ席数と状態数の関係.

テーブル席数	3	4	5	6	7	8	9	10
状態数	13	33	84	214	545	1388	3535	9003

6.1.1　状態数

本項では，客席案内戦略をマルコフ決定過程（もしくは強化学習）で求める際に必要となる状態数について述べる．状態は 3.1.1 項で説明した方法により定義する．表 6.1 は，カウンタ席数と状態数の関係を示したものである．なお，客は 1 人客，2 人客，3 人客の 3 種類のみとする．

表 6.1 より，3 席では状態数 13 であるが，10 席になると 9000 を超えており，状態数が客席数に対して指数関数的に増大しているのが確認できる．(この現象は次元の呪いと呼ばれる [18]).

カウンタ席数が 10 を超えると状態数が 1 万を超える．したがって，マルコフ決定過程により最適戦略を導出することが困難となる．そのため，以下，カウンタ席数は 10 で数値評価を行うが，参考として，カウンタ席数が 20 の場合についても各種客案内戦略の下での平均席利用数をシミュレーションにより評価する．

6.1.2　パラメタ

シミュレーション実験でのパラメタは以下の通りである．

サービス時間は指数分布に従い，平均サービス時間 [時] は $(h_1, h_2, h_3) = (1/2, 3/4, 1)$

到着過程はポアソン過程に従い，平均到着率 [組/時] は 10 席のとき $(\lambda_1, \lambda_2, \lambda_3) = (6, 3, 3/2)$, 20 席のとき $(\lambda_1, \lambda_2, \lambda_3) = (12, 6, 3)$ とした．

比例定数 $\alpha_n = n$ と仮定した．制御の目的は席利用数最大化になる．

シミュレーションの測定時間 $T = 10^2$ で行った．

6.1.3　対照実験

シミュレーション結果を評価する指標として**移動席組合せ**という案内方法での結果を用いた．移動席組合せ案内は，店内の客の座席位置を変えて新しい客を案内できるスペースが生じ得るならば，常に客の入店を許可する案内方法とする（図 6.2）．
移動席組合せ案内は，最大限客詰めして達し得る平均利用席数の上限値の目安となる．

なお，移動席組合せ案内の定常状態確率 π_a は正規化定数を C として以下のように積形式で求められ，カウンタ $K = 20$ 席の場合でも数値評価が可能である．

$$\pi_a(\mathbf{s}) = \pi_a(s_1, s_2, s_3) = \frac{1}{C} \frac{\prod_{i=1}^{3} \frac{(\lambda_i h_i)^{s_i}}{s_i!}}{\sum_{\mathbf{s}} \prod_{i=1}^{3} \frac{(\lambda_i h_i)^{s_i}}{s_i!}}$$

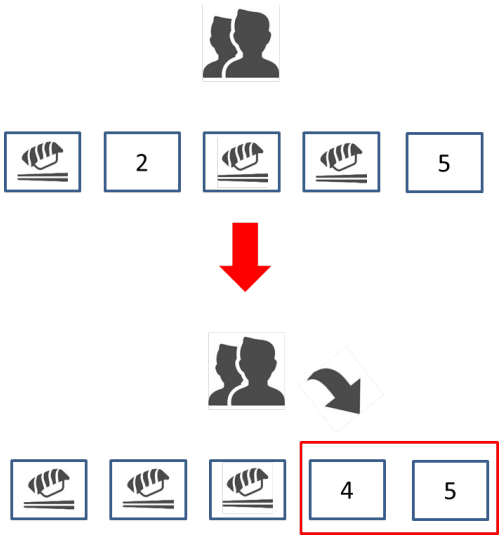

図 6.2　2 人客の場合，移動席組合せ案内

6.1.4　結果

カウンタ席 $K = 20$ 席の場合，各種案内戦略はシミュレーションの測定時間 $T = 10^2$ で 100 回の平均値を結果とした．

左軸（棒グラフ）は各客のブロック率を表し，右軸（折れ線グラフ）はカウンタ席の平均席利用数を表している．

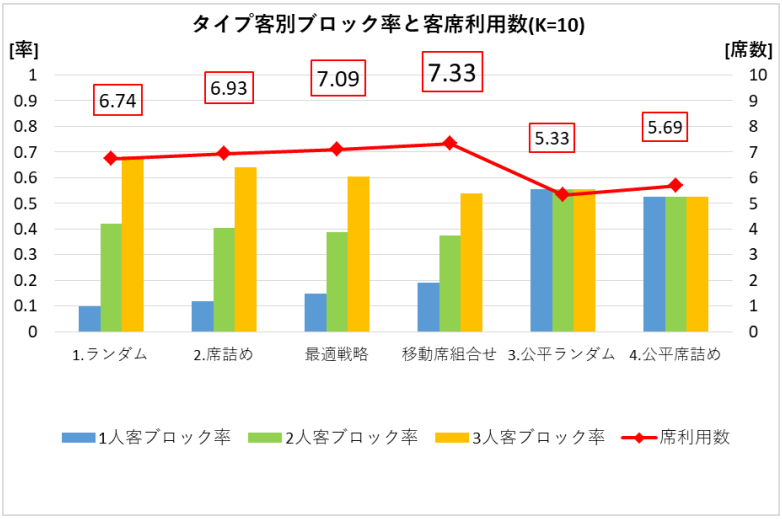

図 6.3　数値計算の結果.

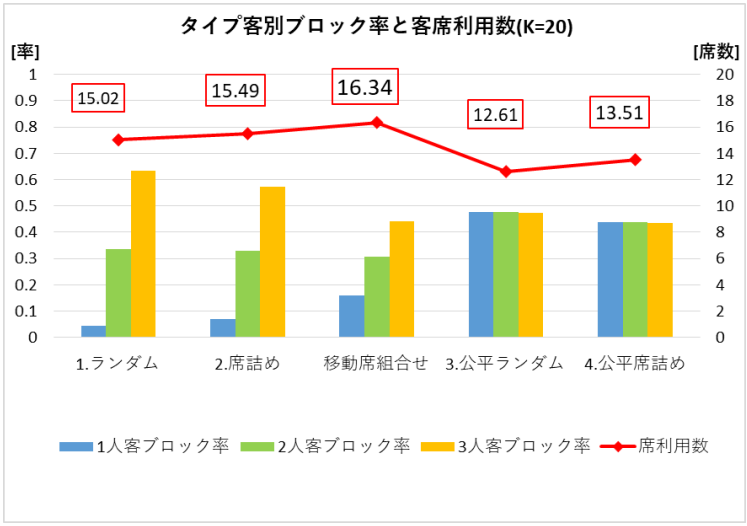

図 6.4　シミュレーションの結果.

表 6.2　状態と状態価値関数.

状態	(96,0,0)	(96,0,1)	(96,0,2)	(96,0,4)	(96,0,128)
状態価値関数	-0.35087	-1.87679	-2.05563	-1.96904	-1.87507

最適戦略の戦略改善について，$K = 10, \alpha_3 = 3, h_3 = 1$ での結果に基づいて述べる.
状態 (96,0,0) での案内戦略を考える (図 6.5).（状態については 3.1.1 項参照.）状態 (96,0,0) 時に 3 人客が来たときの推移先は，案内する場合 (96,0,1)，(96,0,2)，(96,0,4)，(96,0,128)，案内を断る場合 (96,0,0) の 5 通りある.
それぞれの状態価値関数を表 6.2 に示す.
状態 $(96, 0, 0)$ から状態 $(96, 0, 1)$ への推移の場合 (図 6.6),

$$3 + Q((96, 0, 1), a) = 1.12321 > Q((96, 0, 0), a) = -0.35087$$

状態 $(96, 0, 0)$ から状態 $(96, 0, 2)$ への推移の場合 (図 6.7),

$$3 + Q((96, 0, 2), a) = 0.94437 > Q((96, 0, 0), a) = -0.35087$$

状態 $(96, 0, 0)$ から状態 $(96, 0, 4)$ への推移の場合 (図 6.8),

$$3 + Q((96, 0, 4), a) = 1.03096 > Q((96, 0, 0), a) = -0.35087$$

状態 $(96, 0, 0)$ から状態 $(96, 0, 128)$ への推移の場合 (図 6.9),

$$3 + Q((96, 0, 128), a) = 1.12493 > Q((96, 0, 0), a) = -0.35087$$

戦略更新では $(\alpha_n h_n + Q(\mathbf{s}', a_n) - Q(\mathbf{s}, a_n))$ が最大となる戦略を採用するので，この場合，状態 $(96, 0, 0)$ から状態 $(96, 0, 128)$ に推移する案内を採用し，状態 $(96, 0, 0)$ 時の案内戦略を更新する.

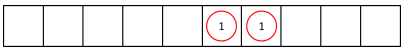

図 6.5　状態 $(96, 0, 0)$.

図 6.6　状態 $(96, 0, 0)$ から状態 $(96, 0, 1)$.

図 6.7　状態 $(96, 0, 0)$ から状態 $(96, 0, 2)$.

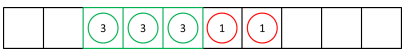

図 6.8　状態 $(96, 0, 0)$ から状態 $(96, 0, 4)$.

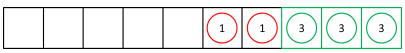

図 6.9　状態 $(96, 0, 0)$ から状態 $(96, 0, 128)$.

最適戦略の案内傾向について，$K = 10, \alpha_n = n$ での結果に基づいて述べる．

傾向 1

案内可能な席の中から，連続した空席の数が最大になるように案内する（図 6.10）．

傾向 2

案内可能な席の中から，連続した空席の数が最大になるような席が複数ある場合，左端または右端に案内する（図 6.11）．

傾向 3

案内可能な席の中から，退席可能性の低い客（1 人客と 2 人客がいる場合，2 人客）に近い席に案内する（図 6.12）．（将来的に連続した空席の数が多くなりうるように案内する．）

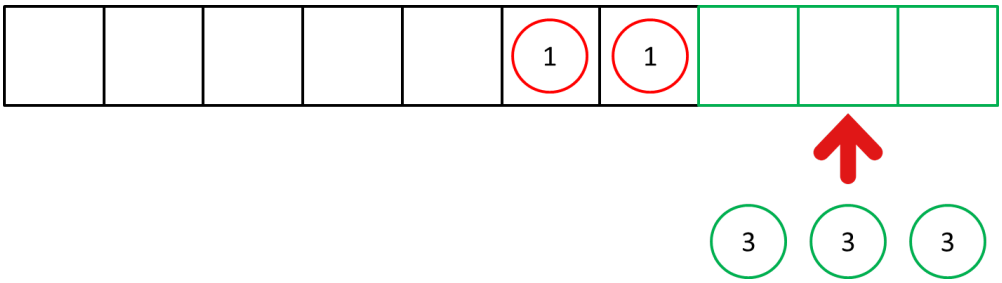

図 6.10　傾向 1，状態 $(96, 0, 0)$ で，3 人客来店時．

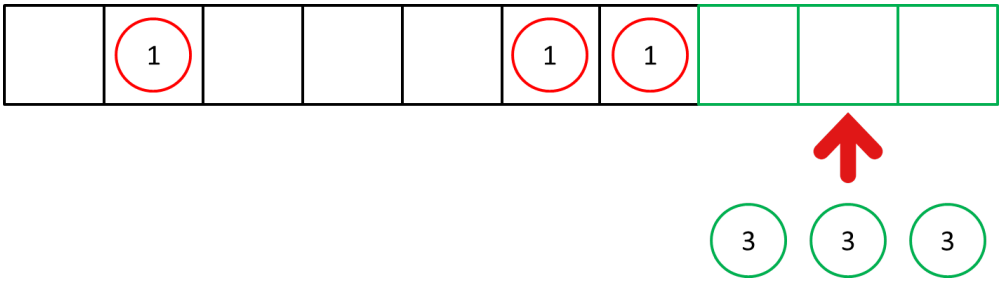

図 6.11　傾向 2，状態 $(98, 0, 0)$ で，3 人客来店時．

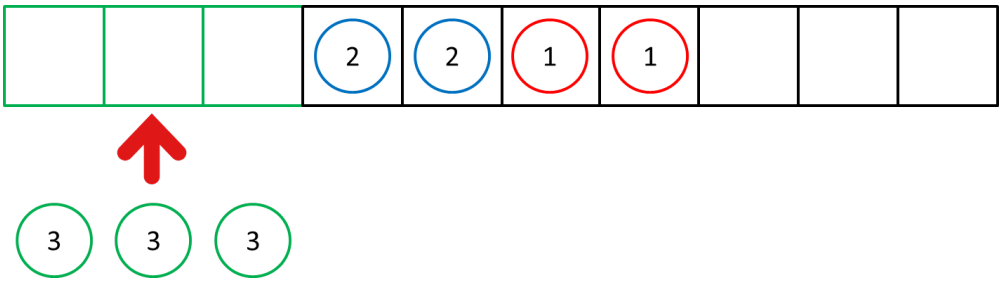

図 6.12　傾向 3，状態 $(96, 24, 0)$ で，3 人客来店時．

6.1.5　考察

図 6.3 より，ランダム案内，席詰め案内，マルコフ決定過程に基づく最適案内による客席利用数は，それぞれ 6.74，6.93，7.09 とその差は比較的に小さいことが読み取れる．また，移動席組合せの結果と比較しても大きな差はない．ただ，いずれの客席案内戦略においても，1 人客と 3 人客のブロック率には大きな差が生じる．一方で，ブロック率を公平にすると，席の平均利用数は大きく低下し，カウンタ席数の約半分が空いてしまっていることになる．

図 6.3，図 6.4 より，どの案内方法でも，席数が多くなるにつれて，席利用率（＝平均席利用数/カウンタ席数）が高いことが確認できる．これは席が増えることによって，客席案内の組合せが増えるためと考えられる．

最適戦略の傾向より，ランダム案内や席詰め案内では空席情報によって案内席を決定しているのに対して，最適戦略では，空席情報に加えて着席している客の情報（$\alpha_n = n$ の場合はサービス時間）によって案内席を決定していることによって，平均席利用数が高くなることが考えられる．

各種案内戦略の効率性について，移動席組合せ案内の平均席利用数は最大となるが，ランダム案内や席詰め案内との差はさほど顕著ではなく，（空いている席に座れる限り客を案内するというような）常識的な案内戦略で十分効果が得られることが明らかになった．
一方で客の案内する機会を公平にすると，平均席利用数は大きく低下するため，効率が悪いことがわかった．

6.2　強化学習による最適化

強化学習の有効性を確認するため，オンラインで測定するシミュレーションを行なった．カウンタ $K = 10$ 席のケースについて，到着間隔やサービス時間の分布を変え，平均席利用数や状態価値関数の相対誤差（＝｜状態価値関数の厳密解 - 強化学習による推定値｜/状態価値関数の厳密解）について評価した．

6.2.1　パラメタ

シミュレーション実験でのパラメタは以下の通りである．
平均サービス時間は，指数分布，k 次のアーラン分布，到着過程は，ポアソン過程，k 次のアーラン分布に従う再生過程で行った．平均サービス時間 [時] は $(h_1, h_2, h_3) = (1/2, 3/4, 1)$，平均到着率 [組/時] は $(\lambda_1, \lambda_2, \lambda_3) = (6, 3, 3/2)$ とした．
比例定数 $\alpha_n = n$ と仮定した．制御の目的は席利用数最大化になる．
初期案内戦略はランダム案内とした．

6.2.2　測定方法

ステップ 1 -測定-

$T(\mathbf{s}), p_a(\mathbf{s}, \mathbf{s}'), r(\mathbf{s}, a), Q(a)$ を T 時間測定

ステップ 2 -戦略評価-

測定値から式 5.3 を解き，状態価値関数 $Q(\mathbf{s}, a)$ を計算する．

ステップ 3 -戦略改善-

状態価値関数に基づいて案内戦略を更新
更新される度ステップ 1 を繰り返す．

状態価値関数と客席案内戦略は T 時間ごとに一斉に更新することとした．（ブートストラップという）．本稿では，更新を 4 回繰り返す（図 6.13）．

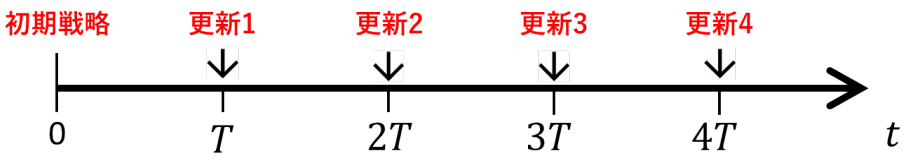

図 6.13　シミュレーション実験：測定の流れ．

表 6.3　到着過程がポアソン過程，サービス時間が指数分布．

T	平均席利用数								状態価値関数相対誤差の平均値
	初期戦略	更新 1 回	更新 2 回	更新 3 回	更新 4 回	ランダム	席詰め	最適値	
10^8	6.745	7.088	7.094	7.094	7.094	6.743	6.928	7.093	0.005
10^7	6.745	7.087	7.086	7.088	7.088	6.743	6.928	7.093	0.039
10^6	6.744	7.078	7.005	7.044	6.963	6.743	6.928	7.093	0.085
10^5	6.759	6.883	6.685	6.736	6.660	6.743	6.928	7.093	0.435

6.2.3　結果・考察

シミュレーションの結果を表 6.3 に示す．

表 6.3 で，ランダム，席詰め，最適値は，それぞれ，ランダム案内時の大域平衡方程式を解いた場合の結果，席詰め案内時の大域平衡方程式を解いた場合の結果，システムパラメタを既知としてマルコフ決定過程を解き最適戦略を求めた場合の結果である．また，状態価値関数相対誤差の平均値は，最適戦略を求めた場合と強化学習で求めた場合の更新 1 回目の各状態の状態価値関数の相対誤差を計算し，その平均をとったものである．

表 6.3 に示したように測定時間 T が十分長ければ，強化学習は最適値とほぼ遜色ない性能を示す．また，状態価値関数相対誤差も十分小さい．

一方で，測定時間 T が短くなるにつれて平均席利用数は減少し，$T = 10^5$ では，強化学習の適用により（ランダム案内や席詰め案内のような）常識的な戦略よりも性能がかえって劣化する現象がみられた．

測定時間 T を長くすると，システムパラメタを既知としてマルコフ決定過程を解き最適戦略を求めた場合の最適値と同等の結果を強化学習によって得ることが可能といえる．一方で，観測時間が短い場合は強化学習は必ずしも有効でない．これは測定されるデータが，強化学習エージェントの現客席案内戦略に強く依存していることが要因であると考えられる．

表 6.4　到着過程：到着間隔が k 次のアーラン分布に従う再生過程，サービス時間：指数分布，$T = 10^7$.

到着間隔	平均席利用数						
	初期戦略	更新 1 回	更新 2 回	更新 3 回	更新 4 回	ランダム	席詰め
2 次のアーラン分布	6.946	7.341	7.341	7.338	7.339	6.946	7.157
3 次のアーラン分布	7.013	7.434	7.433	7.434	7.433	7.013	7.238
4 次のアーラン分布	7.046	7.485	7.480	7.482	7.480	7.046	7.283

表 6.5　到着過程：ポアソン過程，サービス時間：k 次のアーラン分布，$T = 10^7$.

サービス時間	平均席利用数						
	初期戦略	更新 1 回	更新 2 回	更新 3 回	更新 4 回	ランダム	席詰め
2 次のアーラン分布	6.743	7.083	7.007	7.050	6.974	6.743	6.937
3 次のアーラン分布	6.737	7.076	6.963	7.021	6.933	6.737	6.945
4 次のアーラン分布	6.735	7.076	6.964	7.000	6.892	6.735	6.947

測定時間 $T = 10^7$ として，到着間隔または，サービス時間を変えた場合について強化学習の有効性を評価する．k 次のアーラン分布などの相型分布では，相の変数が加わり，さらに状態数が爆発するので，マルコフ決定過程により解析的に最適案内戦略を求めることは困難である．強化学習による最適化およびランダム案内戦略，席詰め案内戦略についてシミュレーション実験を行い，その結果を表 6.4，表 6.5 に示す．

どの分布でも（ランダム案内や席詰め案内のような）常識的な戦略よりも平均席利用数が向上している．更新するごとに平均席利用数が増えており，前更新時の値と比べて，同等かそれ以上になっている．このことから T を長くすると強化学習による最適化は有効であるといえる．また，k が増えるほど平均席利用数は増えている．これにより到着間隔の分散が小さくなるほど，最適戦略のもとでの収容効率が高くなることがわかる．

待ち行列モデルでは到着間隔のばらつきが小さくなるほど，待ち時間が減ることは計算が可能である [19].

どの分布でもランダム案内戦略よりも平均席利用数が向上している．ただ，更新ごとに平均席利用数は変化し，不安定である．

また，k が増えるほど，強化学習で得られた平均席利用数は減っている．これによりサービス時間分布の分散が小さくなるほど，最適戦略のもとでの収容効率が落ちていることがわかる．一方で，席詰め案内戦略では k が増えるほど，平均席利用数がわずかではあるが向上していることが確認できる．サービス時間分布の分散と，客席案内戦略による平均席利用数の関係については明らかになっていないため，今後の課題である．

これらのケースから，解析的なアプローチが困難であっても，強化学習により平均席利用数を向上させることが可能といえる．

第 7 章

結論

飲食店を対象として，種々の客席案内戦略の有効性について評価した．客がポアソン過程に従って来店し，滞在時間が指数分布に従う場合，マルコフ決定過程により客席案内戦略を最適化することが可能である．しかし，常識的な案内戦略に比べて，最適化によるゲインはさほど大きくないことが確認された．観測時間が十分長く取れるならば，測定ベースで強化学習により客席案内戦略を改善することが可能であり，対象システムを既知としてマルコフ決定過程により解析的に客席案内戦略を最適化した場合と遜色ない結果が得られることを確認した．また，解析的なアプローチが困難な対象に対しても，強化学習は有効であることも確認された．

一方，測定時間 T が十分に長く取れない対象に対しては，強化学習は必ずしも有効ではなく，常識的な案内戦略よりも性能がかえって劣化することも判明した．

本研究では，客席利用状態の状態数の爆発が本質的な問題となる．例えば，客席が 10 を超えると状態数は 9000 を超え，状態数が客席数に対して指数関数的に増大する．このため大規模な飲食店の場合，状態価値関数を求めるための逆行列計算に膨大な時間がかかり，解析的な最適化はもちろん強化学習の straightforward な適用も困難である．対応策として，客席利用数を丸め込んでマクロに定義し，状態数を削減することが考えられる．一方，状態数の削減による計算時間の削減と性能劣化も生じうる．状態数の削減による計算時間の削減と性能劣化のトレードオフの評価は今後の課題である．

参考文献

[1] 伊藤嘉浩, 高橋優音,”日本企業における SNS を用いたマーケティング戦略: 有効な活用とマネジメント,” 山形大学紀要 *(*社会科学*)*,vol. 45,no. 1,pp. 91-127,2014.

[2] FUNG,Kwok-Kwan,”It is not how long it is, but how you make it long–waiting lines in a multi - step service process, ” *System Dynamics Review*,vol. 17,no. 4,pp. 333-340,2001.

[3] 宇都宮陽一, 奥田隆史,”多段待ち行列モデルとなる店舗サービスのスタッフ配置に関する解析,” 研究報告数理モデル化と問題解決 *(MPS)*,vol. 113,no. 14,pp. 1-5,2017.

[4] 宇都宮陽一, 奥田隆史,”多段待ち行列モデルを使った店舗サービスにおける待ち時間の評価,” 研究報告数理モデル化と問題解決 *(MPS)*,vol. 115,no. 5,pp. 1-5,2017.

[5] Hwang.J,”Restaurant table management to reduce customer waiting times,” *Journal of Food-service Business Research*,vol. 11,no. 4,pp. 334-351,2008.

[6] Thompson.G.,”Optimizing restaurant-table configurations: Specifying combinable tables,” *Cornell Hotel and Restaurant Administration Quarterly*,vol. 44,no. 1,pp. 53-60,2003.

[7] Kimes.S, and Thompson.G,”An evaluation of heuristic methods for determining the best table mix in full-service restaurant,” *Journal of Operations management*,vol. 23,no. 6,pp. 599-617,2005.

[8] 松永和也, 佐々木龍太, 竹渕瑛一, 速水治夫, ”時間・空間的進行状況を把握する飲食店卓管理システムの提案,” *In* ワークショップ *2016 (GN Workshop 2016)* 論文集 ,vol. 2016, pp. 1-7,2016.

[9] 野中朋美, 清水香那, 水山元,”顧客の予測退店時刻を考慮した飲食店における動的な座席割当てシステム,” 日本機械学会論文集,vol. 82,no. 842,2016.

[10] Mizuyama.H,Yoshida.A, and Nonaka.T,”A Serious Game for Eliciting Tacit Strategies for Dynamic Table Assignment in a Restaurant,” *ISAGA(International Simulation and Gaming Association)*2016.

[11] 三宅功,”異速度通信混在回線の最適回線留保制御,” 電子情報通信学会論文誌 *B*,vol. J71-B,no. 12,pp. 1419-1424,1988.

[12] 白田純子, 宮田高道, 山岡克式,”多元トラヒックの対等性を考慮したフロー受付制御の特性解析 (二元モデル),” 電子情報通信学会技術研究報告会 *IN* 情報ネットワーク,vol. 108,no. 92,pp. 11-16,2008.

[13] 塩田茂雄, 河西憲一, 豊泉洋, 会田雅樹, 待ち行列理論の基礎と応用. 川島幸乃助監修, 共立

出版,2014.

[14] 滝根哲哉,”連続時間マルコフ連鎖と即時系に対する通信トラヒック理論,” 通信トラヒック工学講義資料,2019.

[15] Henk. C. Tijms, *Stochastic Models : An Algorithmic Approach*.New York:Wiley,1994.

[16] Ronald. A. Howard,*Dynamic Programming and Markov Processes*.The MIT Press,1960.

[17] 森村哲郎, 強化学習. 講談社,2014.

[18] 中出康一,”マルコフ決定過程における近似ＤＰアルゴリズム,” オペレーションズ・リサーチ,vol. 58,no. 9,pp. 545-551,2013.

[19] 高橋幸雄. ”やさしい持ち行列 (2)-等間隔運転は待ちを減らす,” オペレーションズ・リサーチ*:* 経営の科学, vol. 40,no. 12,pp. 716-721,1995.

謝辞

ご多忙な中，塩田茂雄教授には多大なご迷惑をおかけしてしまいました．本稿についてご指導していただき，誠にありがとうございました．また，研究室の同期や後輩相談に乗ってくださった方々に感謝の意を表します．ありがとうございました．同じフロアの丸山研究室および関口研究室の同期には，研究以外のところで大変お世話になりました．ありがとうございました．食事や洗濯物など普段の生活を支えてくださった親に感謝の意を表します．ありがとうございました．研究環境に恵まれていたことを改めて実感しております．

研究成果

野田脩平，塩田 茂雄，”飲食店における客席案内の最適戦略，”
待ち行列シンポジウム「確率モデルとその応用」，2019 年 1 月．

野田脩平，塩田茂雄，” 強化学習による客席案内戦略のオンライン構築，”
待ち行列シンポジウム「確率モデルとその応用」，2020 年 1 月．

野田脩平，塩田 茂雄，”飲食店における客席案内の最適戦略，”
（名古屋大学，早稲田大学，電気通信大学，広島市立大学，千葉大学，国立情報学研究所合同）
第 2 回学生技術交流集会，2019 年 10 月．