

連続多段階ランダム割付試験 (SMART : Sequential Multiple Assignment Randomized Trials) の紹介

日本製薬工業協会 医薬品評価委員会 データサイエンス部会
2021年度 タスクフォース 4 / 2022年度 継続タスクフォース 4
Estimandの実装
SMART サブチーム

2022年11月

はじめに

連続多段階ランダム割付試験 (Sequential Multiple Assignment Randomized Trials, SMART) は連続する多段階のランダム割付を通して、一連の動的治療計画を立てるためのデザインである。動的治療計画とは、被験者ごとの治療開始後のデータを含む個人情報に基づきその被験者ごとにどの治療が相応しいかを同定する、治療決定ルールの系である。一般的な動的治療計画の設定は各被験者が複数の段階にまたがって介入を受ける状況を考えており、それに対する統計的な目標とはその状況における治療決定ルールの系の推定、推論を行うことにある。動的治療計画は個別化医療の一部と考えられ、個別化医療は、性別や年齢、臨床検査値、治療の途中経過時点でのレスポンス等といった患者ごとの個人情報から、その患者ごとにより良い動的治療を選択する医療である。個別化医療が考えられる分野として、健康状態が時間に応じて変化し、それにに応じて対処すべき治療が変わりうる慢性疾患の領域が挙げられる。リウマチはその一例であり、実際に American College of Rheumatology (ACR) のガイドラインでは、リウマチ患者が寛解を達成するような動的治療を推奨している (Singh et al. 2016)。

SMART の例として統合失調症患者対象の Sequential Multiple-Assignment Randomized Trials to Compare Antipsychotic Treatments (SMART-CAT) を紹介する。SMART-CAT の目的は一般的に使用されている抗精神病薬の有効性を評価することである。具体的には 1) 最初の抗精神病薬が無効であった場合、その後の最適な治療法は何か、2) 最初の試験が無効であった患者にクロザピンを使用し、他の非定型抗精神病薬と比較して優れた効果が得られるかどうか、を検討することである。本試験はまだ実施されたばかりであるため解析結果は得られていないが、最適な治療法を明確にするために SMART が用いられている。臨床開発計画や市販後の計画において、薬剤の位置づけの優位性を高める戦略を立てる際に SMART は最適な治療法を明確にするという観点では有用な方法である。

SMART は米国食品医薬品局 (Food and Drug Administration, FDA) が公表している「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020 年 12 月)」でアダプティブな介入試験として紹介されている。最近では、がん患者を対象に COVID-19 の予防薬に関する有効性の評価のため COVID-19 Prevention and Treatment in Cancer; a Sequential Multiple Assignment Randomised Trial (C-SMART) が Peter MacCallum Cancer Centre 主導で米国の国立がん感染症センター (National Centre for Infections in Cancer, NCIC) と共同で実施されており、現在注目すべき試験デザインとなっている。

SMART に関する日本語のテキストはまだ存在しておらず、まだ新しいデザインであるため、本報告書で解説を試みた。また実際に解析して実感を得られるようにサンプルデータと SAS のコードも付けた。

本報告書は Chakraborty 及び Moodie が SMART についてまとめた「Statistical Methods for Dynamic Treatment Regimes」の 2 章から 4 章にそって SMART の概要ならびに動的治療計画

の構築によく用いられる推定手法を紹介する。なお、本報告書で参考にした「Statistical Methods for Dynamic Treatment Regimes」の5章では周辺構造モデルを用いた動的治療計画の推定、6章ではG-computationを用いた動的治療計画の推定、7章ではtime-to-eventデータなど複雑なデータを扱った動的治療計画の推定、8章では動的治療計画の統計的推測の性質、9章では追加の考察と結論がまとめられている。本報告書で2章から4章を取り上げた理由としては、SMARTを学ぶ初学者がとりかかりやすい平均アウトカムに関するモデリングや代替治療との平均アウトカムの差に関するモデリングを用いた動的治療計画の推定に焦点を当てるためである。動的治療計画の推定についてより深く学びたい場合は「Statistical Methods for Dynamic Treatment Regimes」を参照いただきたい。また、本報告書では動的治療計画を立てるための推定方法に留め、得られた動的治療計画を集団全体に適用した際の平均因果効果の推定方法は扱わない。そちらに興味のある読者は、例えば、Robins and Hernán (2008) や「ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～」[日本製薬工業協会医薬品評価委員会データサイエンス部会 2021年度タスクフォース4 / 2022年度継続タスクフォース4, (2022)]を参照いただきたい。

本報告書の1章は統計専門家だけでなく臨床試験担当者向けに実例を交えて記載した。2章以降は統計専門家に向けて方法論を中心に記載した。

参考文献

1. Singh AJ, Saag GK et al. (2016). 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis, American College of Rheumatology.
2. Chakraborty B and Moodie EEM. (2013). Statistical Methods for Dynamic Treatment Regimes, Springer.
3. Robins, J. M. and Hernán, M. A. (2008). Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G, Davidian, M, Verbeke, G. (eds). Longitudinal data analysis: New York: Chapman and Hall/CRC Press.
4. 日本製薬工業協会医薬品評価委員会データサイエンス部会 2021年度タスクフォース4 / 2022年度継続タスクフォース4 因果推論による時間依存性治療の推定 サブチーム. (2022). ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～ <https://>公表後追記

目次

1. 連続多段階ランダム割付試験とは.....	6
2. 動的治療方針を与えるための強化学習	10
2.1 Q 学習.....	13
2.1.1 方法論.....	13
2.1.2 Q 学習の解析例の紹介	14
2.2 構造ネスト平均モデル	19
2.2.1 最適 blip 関数	19
2.2.2 リグレット関数.....	20
2.2.3 構造ネスト平均モデルを用いた最適な動的治療方針の推定例.....	20
2.3 G 推定.....	21
2.3.1 方法論.....	21
2.3.2 シミュレーション	24
2.4 A 学習、リグレットに基づく方法	26
2.4.1 A 学習.....	26
2.4.2 リグレットに基づく方法	26
2.4.3 弱視の治療のための眼帯療法の 2-stage 仮想試験シミュレーション	27
2.4.4 まとめ.....	29
3.統計プログラムの実装例.....	30
Appendix SMART のサンプルサイズ設計	33

略語一覧表

略語	略していない表現（英語）	略していない表現（日本語）
A 学習	Advantage learning	
CALGB	Cancer and Leukemia Group B	
DTR	Dynamic treatment regime	動的治療計画
FDA	Food and Drug Administration	米国食品医薬品局
IMOR	Iterative minimization of regrets	
RCT	Randomized control trial	ランダム化比較試験
SMART	Sequential Multiple Assignment Randomized Trials	連続多段階ランダム割付試験

1. 連続多段階ランダム割付試験とは

SMART は連続する多段階のランダム割付を通して、一連の動的治療計画を立てるためのデザインである¹。動的治療による期待されるアウトカムを関数化し、その関数を最大化する治療計画を最適な動的治療計画として提示する。多段階のランダム割付を行う理由は、動的に治療計画を推定する際に未観測交絡を回避するためである。SMART では、最初のランダム化で処置を割り当て、その後一部またはすべての患者に動的に別の処置をランダムに割り当てる。計画されたすべての治療を終えた段階で、すべてのデータの解析を行い、アウトカムに応じた最適な動的治療計画を推定する。パラメータの推定値を用いた動的治療計画は患者の背景情報やこれまでの治療やアウトカムの状態に応じて新たな治療を提案することができる。

¹ SMART について、FDA「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020 年 12 月)」ガイドラインでは“Sequential Multiple Assignment Randomized Trials (SMARTs) are designed to inform the development of adaptive interventions.”と紹介している。また、「Statistical Methods for Dynamic Treatment Regimes」(Chakraborty and Moodie 2013)は”Note that the goal of SMART design is to generate high quality data that would aid in the development and evaluation of optimal DTRs.”としており、本報告書では、これらを参考に SMART を「連続する多段階のランダム割付を通して、一連の動的治療計画を立てるためのデザイン」とした。なお、動的治療計画は「Statistical Methods for Dynamic Treatment Regimes」(Chakraborty and Moodie 2013)で以下のように示されている通り、慢性病などに対する一連の治療で、いつ、どのように治療の種類や治療の用量を変更すべきか得られたデータに基づいて決定するものとなる。

“A dynamic treatment regime (DTR) is a sequence of decision rules, one per stage of intervention, for adapting a treatment plan to the time-varying state of an individual subject. Each decision rule takes a subject’s individual characteristics and treatment history observed up to that stage as inputs, and outputs a recommended treatment at that stage; recommendations can include treatment type, dosage, and timing. DTRs are alternatively known as treatment strategies, adaptive treatment strategies, or treatment policies. Conceptually, a DTR can be viewed as a decision support system of a clinician (or more generally, any decision maker), described as a key element of the chronic cure model.”

まず初めに仮想的な試験を通して SMART の具体例を紹介する。アルコール依存症患者を対象に図 1 の SMART を考える。

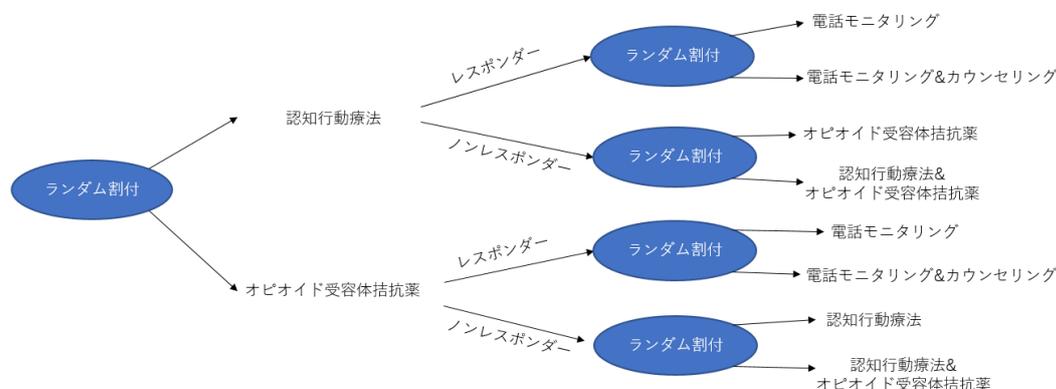


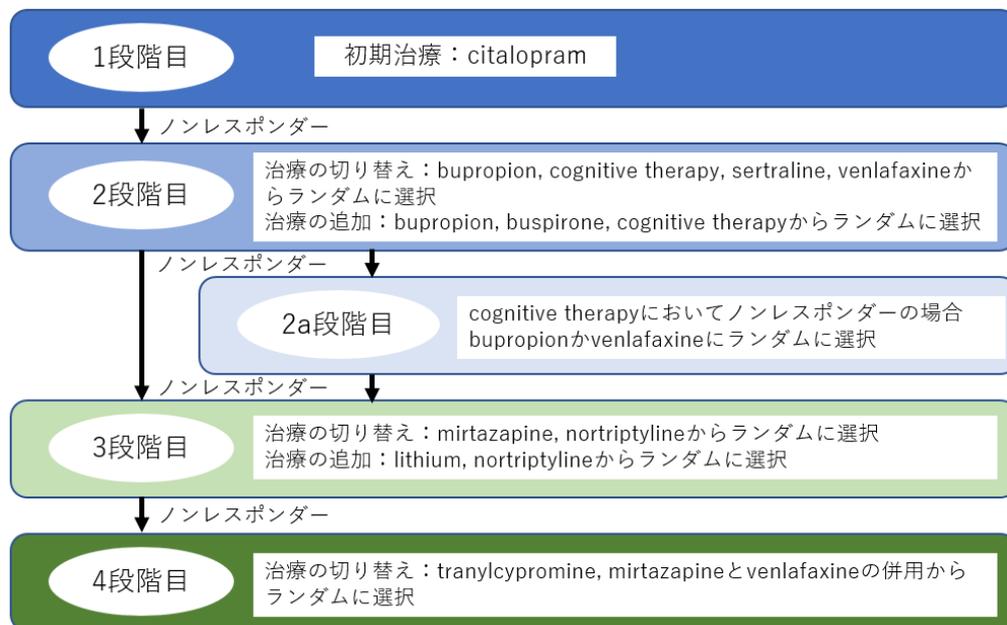
図 1 アルコール依存症患者を対象とした仮想的な SMART の例

- 本試験の目的は、患者の改善状況や処置歴から母集団に対する平均的に最適な処置を決定するための動的治療計画を構築することである。
- 1 段階目のランダム割付では、認知行動療法又はオピオイド受容体拮抗薬をランダムに割り付ける。それぞれの治療法に対してレスポナーには電話モニタリング又は電話モニタリング&カウンセリングをランダムに割り付ける。ノンレスポナーは 1 段階目で割り当てられなかった治療法又は両方の処置（認知行動療法&オピオイド受容体拮抗薬）をランダムに割り当てる。ノンレスポナーについて、より有効性が低いと考えられる治療（電話モニタリング、電話モニタリング&カウンセリング）は割り当てないため、レスポナーとノンレスポナー間の処置の直接比較は行わない。
- 一連の動的治療による期待報酬は関数 $f(O_1, Y_1, A_1, R, A_{21}, A_{22}, A_{23}) = \beta_1 + \beta_2 O_1 + \beta_3 A_1 + \beta_4 Y_1 + (1 - R)\beta_5 A_{21} + R(\beta_6 A_{22}(A_{21} + 1)/2 + \beta_7 A_{23}(1 - A_{21})/2)$ で表すものとする。ここで、 O_1 は処置前のある時点からランダム割付時点までのアルコール依存症の改善度、 A_1 は 1 段階目のランダム割付で認知行動療法が割り付けられて処置されたときに 1、オピオイド受容体拮抗薬が割り付けられて処置されたときに -1、 Y_1 は 1 段階目の処置後のアルコール依存症の改善度合いであり 0 から 100 (100 に近いほど改善とする)、 R は 1 段階目でレスポナーなら 0、ノンレスポナーなら 1、 A_{21} は 2 段階目の処置で電話モニタリングが割り付けられて処置されたときに 1、電話モニタリング&カウンセリングが割り付けられて処置されたときに -1、 A_{22} は 2 段階目のランダム割付でオピオイド受容体拮抗薬が割り付けられて処置されたときに 1、認知行動療法&オピオイド受容体拮抗薬が割り付けられて処置されたときに -1、 A_{23} は 2 段階目のランダム割付で認知行動療法が割り付けられて処置されたときに 1、認知行動療法&オピオイド受容体拮抗薬が割り付けられて処置されたときに -1 とする。

- 最適な動的治療計画を推定するために、関数 $f(O_1, Y_1, A_1, R, A_{21}, A_{22}, A_{23})$ のパラメータ $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T$ を図1の試験結果から推定する。なお、SMARTでよく用いられる推定方法は本報告書の2章に示した。また、本事例とは異なる事例での解析プログラムと解析結果を3章に示した。
- 動的治療計画の決定方法を説明する。治療は関数 $f(O_1, Y_1, A_1, R, A_{21}, A_{22}, A_{23})$ がより大きくなる処置を最適な動的治療計画として選択する。試験結果から得られたパラメータの推定値を $\hat{\beta} = (1, 0.1, 1, 0.2, -1, -1, -1)^T$ とする。例えば、認知行動療法を処置され、アルコール依存の度合いが非常に低くレスポonderである場合、候補となる処置は電話モニタリング又は電話モニタリング&カウンセリングである。1段目の処置後のアルコール依存症の改善度合いは50であったとする。電話モニタリングを処置した場合の関数は $f(10, 50, 1, 0, 1, A_{22}, A_{23}) = 12$ 及び電話モニタリング&カウンセリングを処置した場合の関数は $f(10, 50, 1, 0, -1, A_{22}, A_{23}) = 14$ で電話モニタリング&カウンセリングが最適な処置方針として選択される。

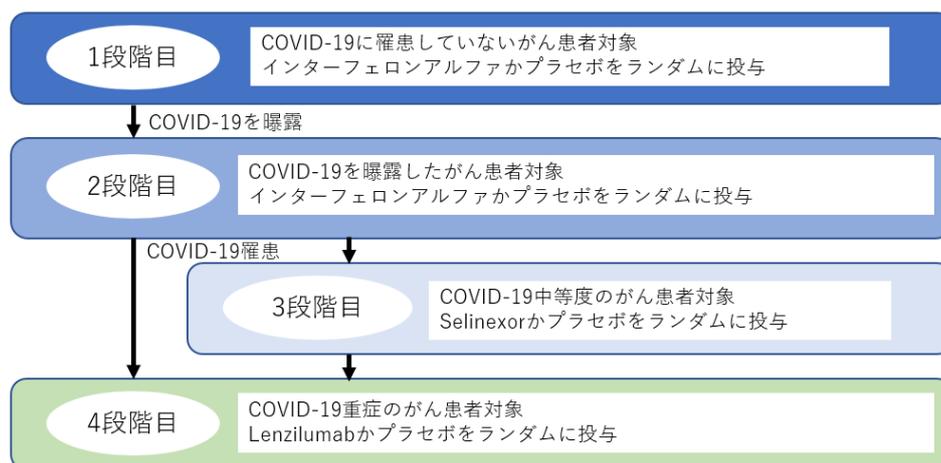
SMARTが実際に適用された試験は以下である。

- Sequenced Treatment Alternatives to Relieve Depression (STAR*D)試験はうつ病患者対象、3段階ランダム化、下図の2段階目及び3段階目において新規治療への切り替え又は薬剤の追加、4段階目において新規治療へランダムに割り付ける試験であった[Lavori et al. (2001)]

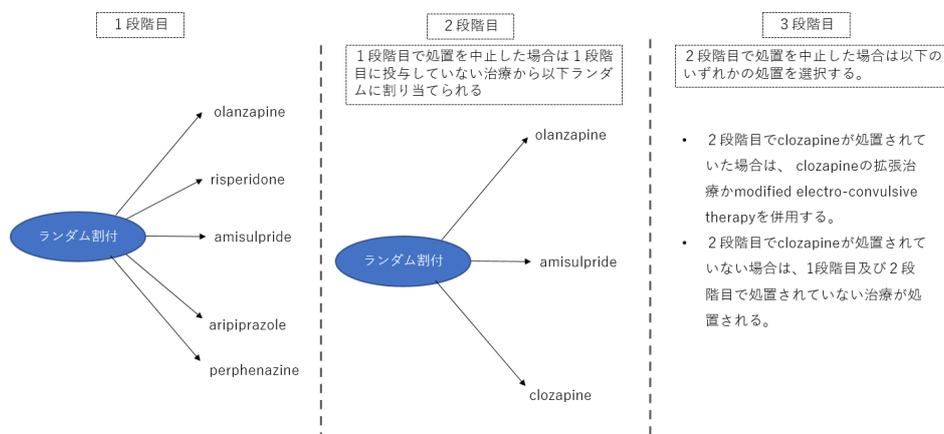


- COVID-19 Prevention and Treatment in Cancer; a Sequential Multiple Assignment Randomised Trial (C-SMART)はCOVID-19未感染またはCOVID-19陽性者との接触がない癌患者対象、4段階ランダム化、1段階目はインターフェロンアルファ又はプラ

セボをランダムに割り当てられ、その後 COVID-19 のウイルスへの曝露が確認された場合は 2 段階目としてインターフェロンアルファ又はプラセボをランダムに割り当て、中等度の COVID-19 感染症になった場合は 3 段階目としてセリネクサー又はプラセボをランダムに割り当て、重度の COVID-19 感染症になった場合は 4 段階目としてレンジルマブ又はプラセボをランダムに割り当てる試験である[ClinicalTrials.gov Identifier: NCT04534725]



- Sequential Multiple-Assignment Randomized Trials to Compare Antipsychotic Treatments (SMART-CAT)は統合失調症患者対象、3 段階ランダム化、1 段階目はオランザピン、リスペリドン、アミスルプリド、アリピプラゾール、パーフェナジンにランダムに割り当て、2 段階目は 1 段階目の処置を中止した場合、オランザピン、アミスルプリド、クロザピンのうち、1 段階目に投与していない治療からランダムに割り当て、3 段階目は 2 段階目の処置を中止した場合に 2 段階目がクロザピンの場合はクロザピンの拡張療法又は修正型電気痙攣療法を併用し、クロザピンでない場合は 1 段階目と 2 段階目で割り当てられなかった処置に割り当てる試験である。[Li et al. (2021)]



SMART で構築された動的治療計画は統計学的仮説検定や信頼区間に基づいて選択され

るわけではなく、医学的に重要だと考えられるデータを考慮することが多い。一方、動的治療計画全体や各段階での治療間の効果など、統計的仮説検定の観点から研究の興味関心に応じてサンプルサイズを計算することもできる。これについては Appendix を参照されたい。

2章の冒頭では動的治療方針を与えるための強化学習について考え、なぜ強化学習が動的治療方針を考えるうえで重要であるか説明する。2.1以降では具体的な動的治療計画の構築方法を紹介し、2.1ではQ学習を用いた推定の方法論及びQ学習の解析例の紹介とSASのサンプルプログラムの紹介、2.2では構造ネスト平均モデル、2.3ではG推定、2.4ではA学習やリグレットに基づく推定方法を紹介する。3章では、統計ソフトRのiqLearnパッケージを用いたQ学習の解析プログラムの例を紹介する。本報告書の1章は例示、2章は、Chakraborty and Moodie (2013)の要約である。

本報告書で段階とステージは同じ意味で使われるが、文章の読みやすさに応じて使い分けられている。

参考文献

5. Ertefaie A. (2020). Robust Q-Learning. JASA, 553(116), 368-381.
6. Chakraborty B and Moodie EEM. (2013). Statistical Methods for Dynamic Treatment Regimes, Springer.
7. Lavori, P. W., Rush, A. J., Wisniewski, S. R., Alpert, J., Fava, M., et al. (2001) Strengthening clinical effectiveness trials: Equipoise-stratified randomization. Biological Psychiatry, 48, 605–614.
8. Li X., Guo X., Fan X., Feng T., Wang C. et. al. (2021) Sequential Multiple-Assignment Randomized Trials to Compare Antipsychotic Treatments (SMART-CAT) in first-episode schizophrenia patients: Rationale and trial design. Schizophrenia Research, 230, 87-94.

2. 動的治療方針を与えるための強化学習

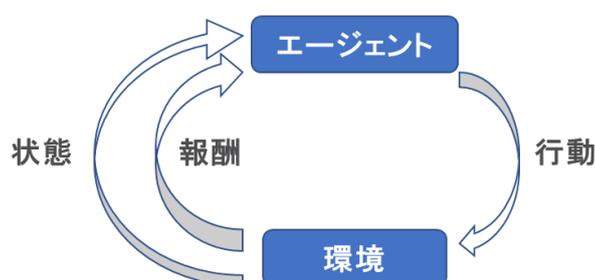
はじめに記号を以下のように定義する。次章でも同様に扱う。

- K : ステージの数
- O_j : ステージ j における背景情報 (離散変数又は連続変数)
- A_j : ステージ j の治療 (離散変数)
- Y_j : ステージ j のアウトカム (連続変数)
- 過去の情報を $\bar{O}_j \equiv (O_1, \dots, O_j)$ 、 $\bar{A}_j \equiv (A_1, \dots, A_j)$ とし、これらをまとめて、環境 (ヒストリー) $H_j \equiv (\bar{O}_j, \bar{A}_{j-1})$ と記載することで、 $Y_j = Y_j(H_j, A_j, O_{j+1}) = Y_j(\bar{O}_j, \bar{A}_j, O_{j+1})$ と書き換えることができる。ただし、 $H_1 = O_1$ である。

なお強化学習の分野では背景情報を状態、処置を行動、アウトカムを報酬と表現することが多いように見受けられるため、本節においては特別そのように呼ぶこととする。

強化学習とは、連続した多段階の意思決定ステージが存在するような意思決定問題を取り扱う機械学習の手法の1つであり、ロボット工学や自動運転などにも広く応用されている手法である。本報告書における我々の目的は、治療が多段階に亘る状況において最適な動的治療方針を推定することであり、多段階の意思決定問題と解釈できることから、強化学習の取り扱う問題と類似している。実際がんでんがん、統合失調症などさまざまな分野において、動的治療方針の開発をターゲットにして強化学習を用いている研究も存在する (Zhao et al. 2009, Guez et al. 2008, Shortreed et al. 2011)。2章冒頭では標準的な強化学習の概要を紹介し、最適な動的治療方針の推定手法の場面でよく用いられる Q 学習 (2.1節) へと進んでいく。

強化学習はエージェントと (学習対象である) 環境との相互作用で特徴づけられる。エージェントとは行動を行う主体を表す。エージェントは意思決定を行う各ステージにおいて、その時点における環境に関する情報を観測しており、それに基づいて行動をとることができる。エージェントの行動を受けて環境は新しい状態に遷移し、エージェントは即時報酬を得られるという流れの繰り返しが強化学習の基本の枠組みである。下図は1ステージ内の流れの基本図である。これはステージの数だけ連続して繰り返される。



各ステージでの環境の状態に基づくエージェントの行動は方策と呼ばれ、記号として本報告書では d を用いることとする。強化学習の目的はこのような状況において、今後獲得できる報酬の和を最大にするような方策を推定することである (なお報酬自体がしばしば確率変数であることから期待値で議論されることが多い)。方策評価のために価値関数と呼ばれる関数を定義する :

$$V^d(o_1) = E_d \left[\sum_{j=1}^K Y_j(H_j, A_j, O_{j+1}) | O_1 = o_1 \right]$$

これは特定の状態 o_1 から開始して方策 d に従って行動 A_j を選択したときの総期待報酬を表している。ステージ j まで進んだ場合、その時点までの環境 H_j は既知であることから、その情報を用いてステージ j 以降の方策 d に対する価値関数は以下のように表される：

$$V_j^d(h_j) = E_d \left[\sum_{k=j}^K Y_k(H_k, A_k, O_{k+1}) | H_j = h_j \right]$$

上述の通り、強化学習の目的は今後獲得できる報酬の和（価値関数）を最大にするような方策を推定することであり、その推定のため、環境 H_j が既知のときにステージ j 以降の総期待報酬を定義する意義がある。

なお価値関数は以下のように再帰的に表現できる：

$$V_j^d(h_j) = E_d[Y_j(H_j, A_j, O_{j+1}) + V_{j+1}^d(H_{j+1}) | H_j = h_j]$$

ステージ j とステージ $j+1$ における価値関数の漸化式は Bellman 方程式(Bellman 1957)と呼ばれる。

過去情報 h_j に対するステージ j の最適な価値関数は $V_j^{opt}(h_j) = \max_{d \in \mathcal{D}} V_j^d(h_j)$ であり、ステージ j の最適な行動とはステージ j の価値関数を最大にするものである。また最適な価値関数においても Bellman 方程式 (Bellman 1957)を満たす：

$$\begin{aligned} V_j^{opt}(h_j) &= \max_{\{a_i \in \mathcal{A}_i | i=j, j+1, \dots, K\}} E[Y_j(H_j, A_j, O_{j+1}) + V_{j+1}^d(H_{j+1}) | H_j = h_j, A_j = a_j] \\ &= \max_{a_j \in \mathcal{A}_j} E[Y_j(H_j, A_j, O_{j+1}) + V_{j+1}^{opt}(H_{j+1}) | H_j = h_j, A_j = a_j] \end{aligned}$$

Bellman 方程式を用いて最適な方策を推定することが強化学習のゴールであるが、推定の詳細については割愛する。

医薬品開発など医療分野における患者データから動的治療方針を推定するような問題と、標準的な強化学習で取り扱う問題にはいくつか大きな相違点もある。例えば強化学習が取り扱う問題では、物理法則や専門知識に基づいてデータの分布や遷移確率などが既知である場合が多い一方で、医療分野においてはしばしば遺伝構成や抱えている家族問題など未知の要因が多く存在する。また強化学習でよく仮定に用いられるマルコフ決定過程（環境の状態の遷移確率が無記憶性をもち、過去の状態に依存しないとする設定）についても、医療分野で妥当でない場合が多い（患者の状態は直近の状態と治療のみに依存し、それより前の時点と独立しているとは考えづらい）。そのため以降では強化学習を直接当てはめるのではなく、強化学習の概念や類似点をふまえた上で、最適な動的治療方針の推定によく用いられる手法を紹介していく。

参考文献

9. He W, Chen Y, Yin Z (2016). Adaptive neural network control of an uncertain robot with full-state constraints. *IEEE Transactions on Cybernetics*, 46(3), 620-629.
10. Zhao, Y, Kosorok MR, Zeng, D (2009). Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28, 3294–3315.
11. Guez A, Vincent R, Avoli M, Pineau J (2008). Adaptive treatment of epilepsy via batch-mode reinforcement learning. In *Proceedings of the innovative applications of artificial intelligence (IAAI)*, Chicago.
12. Shortreed SM, Laber E, Lizotte DJ, Stroup TS, Pineau, J, Murphy, SA (2011). Informing sequential clinical decision-making through reinforcement learning: An empirical study. *Machine Learning*, 84, 109–136.
13. Bellman RE (1957). *Dynamic programming*. Princeton: Princeton University Press.

2.1 Q 学習

Q 学習(Watkins (1989))とは、Bellman 方程式をベースにし Q 関数と呼ばれる関数を価値関数を用いた強化学習に該当する手法である。Murphy (2005)は Q 学習を最適な治療方針の決定に応用した。Q 関数は回帰モデルを用いて定義することができ直感的にわかりやすい関数である。なお、Q 学習の Q は Quality である。

2.1.1 方法論

2 段階の SMART に基づいて Q 関数を回帰モデルに近似した場合の、最適な治療方針の推定方法について説明する。背景情報並びに SMART における治療歴は (O_1, A_1, O_2, A_2) とする。 $H_1 \equiv O_1, H_2 \equiv (O_1, A_1, O_2)$ とする。各段階に治療がある SMART を考える ($A_j \in \{-1, 1\}$)。 Y_1 は中間アウトカム、 Y_2 は最終アウトカムとする。治療方針 d_j はヒストリー H_j の関数で $d_j(H_j) \in \{-1, 1\}$ と表され、ヒストリーの空間から処置の空間への写像とする。 a_j は実数値を意味する。動的治療計画を構築するための最適な Q 関数の定義を以下とする。

- $Q_2^{opt}(H_2, A_2) = E[Y_2 | H_2, A_2]$
- $Q_1^{opt}(H_1, A_1) = E \left[Y_1 + \max_{a_2} Q_2^{opt}(H_2, a_2) | H_1, A_1 \right]$

Q 学習は総報酬の最大化に興味がある。そのため、ステージ 2 では Y_2 、ステージ 1 では Y_1 に加え報酬が最大になる $\max_{a_2} Q_2^{opt}(H_2, a_2)$ を加えるため、最終治療の最適な Q 関数のパラメータから後ろ向きに推定する。各最適な治療方針は各 Q 関数を最大化することで得られる。

$$d_j^{opt}(h_j) = \arg \max_{a_j} Q_j^{opt}(h_j, a_j) \quad , j = 1, 2$$

真の Q 関数は未知なため、データから推定する必要がある。Q 関数は条件付き期待値であり、モデル化する手法は回帰モデルであることに注意する。より具体的に説明するために

Q 関数について線形回帰モデルをあてはめて考える。Q 関数を $Q_j^{opt}(H_j, A_j; \beta_j, \psi_j) = \beta_j^T H_{j0} + (\psi_j^T H_{j1}) A_j$ 、 H_{j0} を背景や中間の結果や SMART における処置歴、 $\beta_j : H_{j0}$ の回帰係数パラメータ、 H_{j1} を治療効果に影響を及ぼす背景情報とする、 $\psi_j : A_j H_{j1}$ の、治療効果に影響を及ぼすと考えられる回帰係数パラメータ。

1. ステージ 2 の Q 関数のパラメータを推定する。ここでは最小二乗法での推定例を紹介する。推定量は $(\hat{\beta}_2, \hat{\psi}_2)$ である。

- $$\operatorname{argmin}_{\beta_2, \psi_2} \frac{1}{n} \sum_{i=1}^n \left(Y_{2i} - Q_2^{opt}(H_{2i}, A_{2i}; \beta_2, \psi_2) \right)^2$$

argmin (argument of the minimum) は最小化問題について最小値を与える引数を意味する。

2. ステージ 1 の Q 関数の推定ではステージ 1 の処置選択を考慮したステージ 2 での最適な処置効果を考慮するため擬似アウトカムを生成する。

- $$\tilde{Y}_{1i} = Y_{1i} + \max_{a_2} Q_2^{opt}(H_{2i}, a_2; \hat{\beta}_2, \hat{\psi}_2), i = 1, \dots, n$$

3. ステージ 1 の Q 関数のパラメータを最小二乗法より推定する。推定量 $(\hat{\beta}_1, \hat{\psi}_1)$ である。

- $$\operatorname{argmin}_{\beta_1, \psi_1} \frac{1}{n} \sum_{i=1}^n \left(\tilde{Y}_{1i} - Q_1^{opt}(H_{1i}, A_{1i}; \beta_1, \psi_1) \right)^2$$

4. Q 学習を用いて最適治療方針が推定される。

- $$\hat{a}_j^{opt}(h_j) = \operatorname{argmax}_{a_j} Q_j^{opt}(h_j, a_j; \hat{\beta}_j, \hat{\psi}_j) \quad , j = 1, 2$$

以上の過程は 3 ステージ以上に拡張可能である。擬似アウトカムを生成する部分について、擬似アウトカムを生成するのではなく、ステージ 2 で最適な治療を受けた患者だけを用いることで 1 段階目の Q 関数を推定する方法もある。Q 学習は R パッケージ (iqLearn) で実行可能である。iqLearn の実行例は 3 章に示す。

2.1.2 Q 学習の解析例の紹介

Ertefaie, A. et al. (2016) を参考に Q 学習の例を示す。2 段階の SMART を考える。ベースライン情報を O_{11} と O_{12} とし、ステージ 1 で治療 $A_1 = 1$ 又は $A_1 = -1$ をランダムに割り当てる。 A_1 の治療後の中間アウトカムを Y_1 とし、 A_1 の治療後の共変量を O_2 とする。ステージ 2 において、 Y_1 が 5 以上の場合は治療効果が見られたとし、ステージ 1 の治療を継続する。 Y_1 が 5 未満の場合は二つの新規治療 $A_2 = 1$ と $A_2 = -1$ にランダムに割り当てる。最終アウトカムを Y_2 とする。試験デザインを図 2 にしめす。また、試験のデータを以下のようにまとめた。

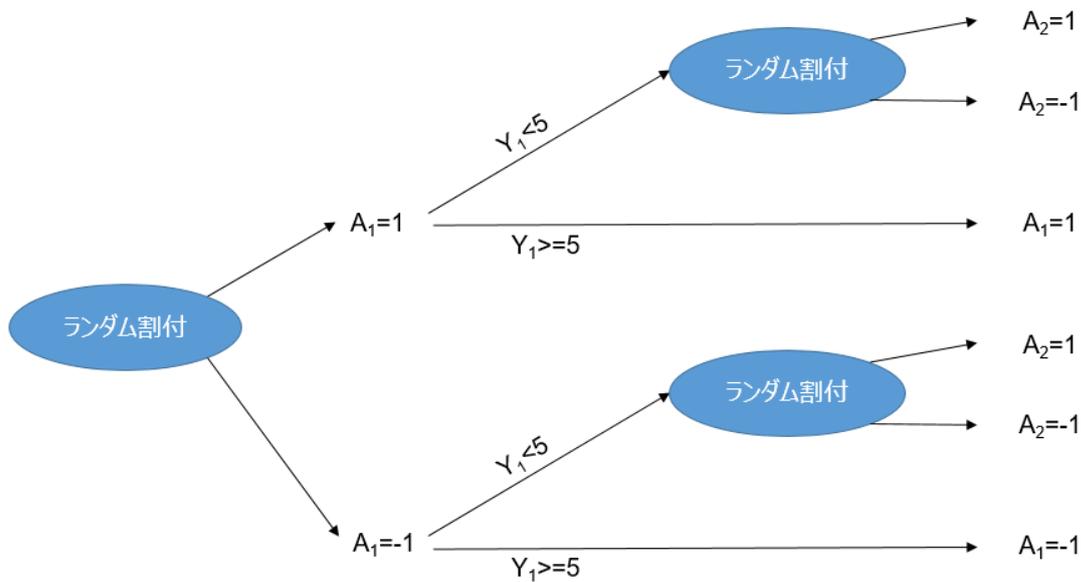


図 2 試験デザイン

- ベースライン情報： O_{11} (A_1 の治療後の共変量 O_2 に影響を与える因子)と O_{12} (A_1 の治療後の中間アウトカム Y_1 に影響を与える因子)
- ステージ1の治療： $A_1 = 1$ 又は $A_1 = -1$
- A_1 の治療後の中間アウトカム： Y_1
- A_1 の治療後の共変量： O_2
- Y_1 が5未満の場合のステージ2の治療： $A_2 = 1$ 又は $A_2 = -1$

各データを以下に従って生成する。

- O_{11} は平均0、分散が1の正規分布に従う。
- O_{12} は平均0、分散が1の正規分布に従う。
- $A_1 = 1$ と $A_1 = -1$ は0.5の確率で割り当てる。
- O_2 は $5 + 0.5A_1 + 0.5O_{11} + \varepsilon_1$ により計算する。 ε_1 は平均0、分散が1の正規分布に従う。
- Y_1 は $5 + 0.5A_1 + 0.5O_{12} + \varepsilon_2$ により計算する。 ε_2 は平均0、分散が1の正規分布に従う。
- Y_1 が5未満の場合、 $R = 1$ とし、 Y_1 が5以上の場合、 $R = 0$ とする。
- $A_2 = 1$ と $A_2 = -1$ は0.5の確率で割り当てる。
- Y_2 は $1 - 2O_{11} + 2O_{12} - O_2 + A_1 + 0.5A_1O_{11} + R(-0.8O_2 + 0.5A_2 + 0.4A_2O_2 + 0.7A_2O_{11}) + \varepsilon_3$ により計算する。 ε_3 は平均0、分散が1の正規分布に従う。
- サンプルサイズは300とする。

smart のデータセットを以下 SAS program code から1つ作成する。

```
data smart;
```

```

call streaminit(1);
do i=1 to 300;
O11=RAND('NORMAL');
O12=RAND('NORMAL');
A1=(RAND('BERNOULLI', 0.5)-0.5)*2;
A2=(RAND('BERNOULLI', 0.5)-0.5)*2;
ep1=RAND('NORMAL');
ep2=RAND('NORMAL');
ep3=RAND('NORMAL');
O2=5+0.5*A1+0.5*O11+ep1;
Y1=5+0.5*A1+0.5*O12+ep2;
if Y1>5 then R=0;else R=1;
Y2=1-2*O11+2*O12-O2+A1+0.5*A1*O11+R*(-
      0.8*O2+0.5*A2+0.4*A2*O2+0.7*A2*O11)+ep3;
output;
end;
run;

```

【Q 関数のパラメータの推定】

1. ステージ 2 の Q 関数の統計モデルを以下とする。

$$\begin{aligned}
Q_2^{opt}(H_{2i}, A_{2i}; \alpha_2, \beta_2) \\
&= \alpha_2 + \beta_{21}O_{11i} + \beta_{22}O_{12i} + \beta_{23}O_{2i} + \beta_{24}A_{1i} + \beta_{25}A_{1i}O_{11i} \\
&\quad + R_i(\beta_{26}O_{2i} + \beta_{27}A_{2i} + \beta_{28}A_{2i}O_{2i} + \beta_{29}A_{2i}O_{11i})
\end{aligned}$$

i は各被験者番号を表す添え字である。 $H_{2i} = (O_{11i}, O_{12i}, A_{1i}, O_{2i}, R_i)$

2. $i = 1, 2, \dots, 300$ に対して Y_{2i} と $Q_2^{opt}(h_{2i}, a_{2i}; \alpha_2, \beta_2)$ の最小二乗推定から推定量 $(\hat{\alpha}_2, \hat{\beta}_2)$ を得る。Q 学習では総報酬の期待値に注目し、ステージ 1 の Q 関数のパラメータを推定するため、ステージ 1 の疑似アウトカムを $Y_{1i} + \max_{a_2} Q_2^{opt}(h_{2i}, a_2; \hat{\alpha}_2, \hat{\beta}_2)$ から計算する。

$\max_{a_2} Q_2^{opt}(h_{2i}, a_2; \hat{\alpha}_2, \hat{\beta}_2)$ は、被験者 i について、 a_2 の Q 関数になっている。したがって、

$Q_2^{opt}(h_{2i}, a_{2i}; \alpha_2, \beta_2)$ の最大化は $a_{2i}(\hat{\beta}_{27} + \hat{\beta}_{28}o_{2i} + \hat{\beta}_{29}o_{11i})$ のみ考えればよい。 a_2 は 1 又は -1 をとることから $\hat{\beta}_{27} + \hat{\beta}_{28}o_{2i} + \hat{\beta}_{29}o_{11i}$ が正の場合は $a_2=1$ で $Q_2^{opt}(h_{2i}, a_{2i}; \alpha_2, \beta_2)$ が最大となり、 $\hat{\beta}_{27} + \hat{\beta}_{28}o_{2i} + \hat{\beta}_{29}o_{11i}$ が負の場合は $a_2=-1$ で $Q_2^{opt}(h_{2i}, a_{2i}; \alpha_2, \beta_2)$ が最大となる。

よって、 $\max_{a_2} Q_2^{opt}(h_{2i}, a_2; \hat{\alpha}_2, \hat{\beta}_2)$ は $\hat{\beta}_{27} + \hat{\beta}_{28}o_{2i} + \hat{\beta}_{29}o_{11i}$ の絶対値をとることで計算でき

る。したがって、ステージ 1 の疑似アウトカム \tilde{Y}_{1i} は以下より計算される。

$$\tilde{Y}_{1i} = Y_{1i} + \hat{\alpha} + \hat{\beta}_{21}o_{11i} + \hat{\beta}_{22}o_{12i} + \hat{\beta}_{23}o_{2i} + \hat{\beta}_{24}a_{1i} + \hat{\beta}_{25}a_{1i}o_{11i} \\ + r_i(\hat{\beta}_{26}o_{2i} + |\hat{\beta}_{27} + \hat{\beta}_{28}o_{2i} + \hat{\beta}_{29}o_{11i}|)$$

3. ステージ1のQ関数の統計モデルを以下とする。

$$Q_1^{opt}(H_{1i}, A_{1i}; \alpha_1, \beta_1) = \alpha_1 + \beta_{11}O_{11i} + \beta_{12}O_{12i} + \beta_{13}A_{1i} + \beta_{14}A_{1i}O_{11i}$$

$i = 1, 2, \dots, 300$ に対して \tilde{Y}_{1i} と $Q_1^{opt}(h_{1i}, A_{1i}; \alpha_1, \beta_1)$ の最小二乗推定から推定量 $(\hat{\alpha}_1, \hat{\beta}_1)$ を得る。

Q関数のパラメータの推定のためのSAS program codeを以下に示す。

/*2段階目の推定*/

```
ods output ParameterEstimates=est2;
```

```
proc glm data=smart;
```

```
    model Y2=O11 O12 O2 A1 A1*O11 R*O2 R*A2 R*A2*O2 R*A2*O11;
```

```
run;
```

/*1段階目の推定で二段階目の処置効果を用いるため横向きの変換*/

```
proc transpose data=est2 out=t_est2;
```

```
    var estimate;
```

```
run;
```

/*データのマージのためキー情報を付与*/

```
data smart; set smart;
```

```
if _n_=1 then sett_est2;
```

```
run;
```

/*1段階目のQ関数を推定するための疑似値*/

```
data smart;
```

```
    set smart;
```

```
    PO=Y1+COL1+COL2*O11+COL3*O12+COL4*O2+COL5*A1+COL6*A1*O11+R*(C \\ OL7*O2+abs(COL8+COL9*O2+COL10*O11));
```

```
run;
```

/*1段階目の推定*/

```
ods output ParameterEstimates=est1;
```

```
proc glm data=smart;
```

```
    model PO=O11 O12 A1 A1*O11;
```

run;

【データ解析の結果】

ステージ2のQ関数の推定値を以下に示す。

パラメータの係数	推定値	標準誤差
Intercept	1.547547	0.306262
O11	-2.05028	0.069839
O12	2.073804	0.063632
O2	-1.07286	0.061972
A1	1.05898	0.072901
O11*A1	0.470258	0.061135
O2*R	-0.82835	0.026439
R*A2	0.87175	0.374694
O2*R*A2	0.311777	0.070334
O11*R*A2	0.770942	0.086984

ステージ1のQ関数の推定値を以下に示す。

パラメータの係数	推定値	標準誤差
Intercept	0.287393	0.094274
O11	-2.35798	0.093764
O12	2.216814	0.092357
A1	0.535181	0.094526
O11*A1	0.599745	0.093767

【動的な治療選択の例】

治療開始時の最適処置は、 $Q_1^{opt}(h_{1i}, a_{1i}; \hat{\alpha}_1, \hat{\beta}_1)$ の処置 a_{1i} の係数部分の $0.535181 + 0.599745o_{11i}$ が正の時、 $a_1 = 1$ で、 $0.535181 + 0.599745o_{11i}$ が負の時、 $a_1 = -1$ として選択する。つまり、 $o_{11i} \geq -0.892348$ の時、処置 $a_1 = 1$ を選択し、 $o_{11i} < -0.892348$ の時、処置 $a_1 = -1$ を選択する。

初期治療投与後の最適処置は、 $r_i = 1$ において $Q_2^{opt}(h_{2i}, a_{2i}; \hat{\alpha}_2, \hat{\beta}_2)$ の処置 a_{2i} の係数部分の $0.87175 + 0.311777o_{2i} + 0.770942o_{11i}$ が正の時、 $a_2 = 1$ で、 $0.87175 + 0.311777o_{2i} + 0.770942o_{11i}$ が負の時、 $a_2 = -1$ として選択する。つまり、 $o_{2i} \geq -2.472735o_{11i} - 2.7960699$ の時、処置 $a_2 = 1$ を選択し、 $o_{2i} < -2.472735o_{11i} - 2.7960699$ の時、処置 $a_2 = -1$ を選択する。処置割り当ての意思決定は、以下図3を用いて行える。図3において赤線の右上に (o_{11i}, o_{2i})

が含まれるとき $a_2 = 1$ が選択され、赤線の左下に (o_{11i}, o_{2i}) が含まれるとき $a_2 = -1$ が選択される。

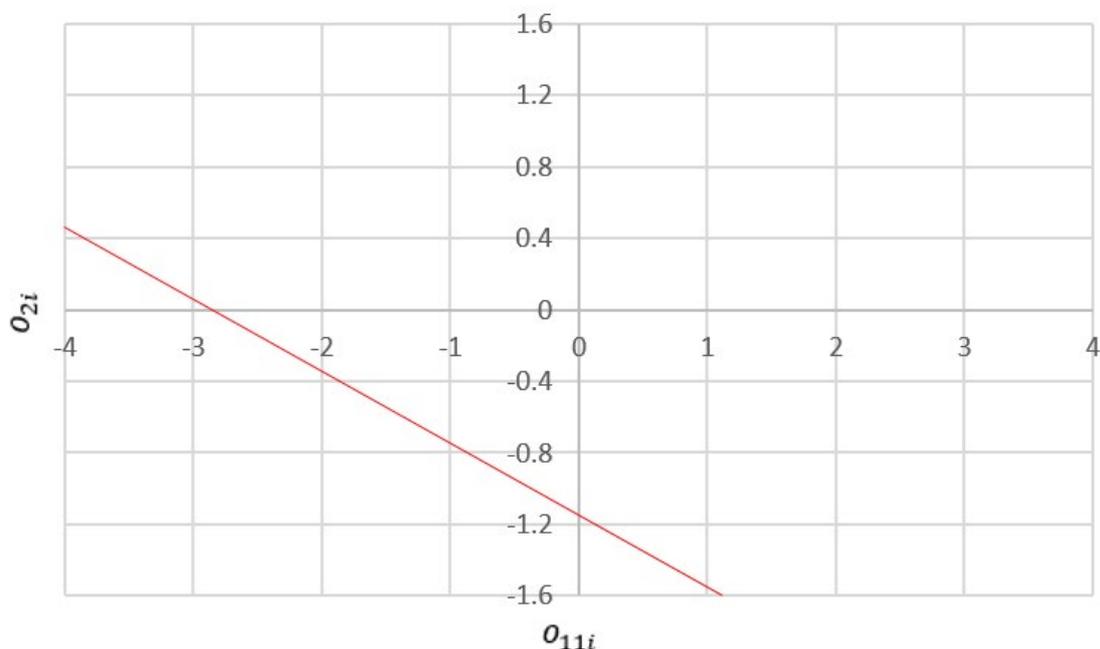


図 3 処置割り当ての意思決定図

参考文献

14. Watkins CJCH. (1989). Learning from delayed rewards. Doctoral Thesis. University of Cambridge; England.
15. Murphy SA. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. 24:455–1481.
16. Moodie EEM., Dean N and Sun YR. (2014) Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences* 6:223–243.

2.2 構造ネスト平均モデル

前節の Q 学習ではアウトカムの条件付き期待値に直接回帰モデルを当てはめたが、本節ではアウトカムの条件付き期待値の差分に対してモデリングを行う構造ネスト平均モデルについて紹介する。なお構造ネスト平均モデルの推定方法の詳細については 2022 年 9 月に公開された“ICHE9(R1)の理解に役立つ因果推論～時間依存性治療編～”の 3 章を参考されたい。

2.2.1 最適 blip 関数

治療 \bar{a}_j を受け、 $j + 1$ 時点以降は最適な治療を受けるあるアウトカムと、そこからステージ j のみ治療を変更（つまり a_j を reference レジメン d_j^{ref} に変更）したアウトカムとの差分の期

待値である以下を最適 blip 関数と定義する：

$$\gamma_j(h_j, a_j) = E[Y(\bar{a}_j, \underline{d}_{j+1}^{opt}) - Y(\bar{a}_{j-1}, \underline{d}_j^{ref}, \underline{d}_{j+1}^{opt}) | H_j = h_j]$$

2.2.2 リグレット関数

最適 blip 関数の特殊なケースであり、reference レジメンの部分に最適治療を代入し、マイナスをかけることでリグレット関数は定義される：

$$\mu_j(h_j, a_j) = E[Y(\bar{a}_{j-1}, \underline{d}_j^{opt}) - Y(\bar{a}_j, \underline{d}_{j+1}^{opt}) | H_j = h_j]$$

なお最適 blip 関数とリグレット関数は reference レジメンの行動を書きかえただけになることから以下の関係が成立する：

$$\mu_j(h_j, a_j) = \max_{a_j} \gamma_j(h_j, a_j) - \gamma_j(h_j, a_j)$$

$$\gamma_j(h_j, a_j) = \mu_j(h_j, \underline{d}_j^{ref}) - \mu_j(h_j, a_j)$$

リグレット関数を活用した最適な動的治療方針の推定については後の節にて紹介する。

2.2.3 構造ネスト平均モデルを用いた最適な動的治療方針の推定例

本節では 2 ステージで各ステージにおける reference レジメンが $\underline{d}_j^{ref} = -1$ (reference レジメンを $A_j = -1$ に定めたという意味) とした簡易的な例を用いて、構造ネスト平均モデルによる最適な動的治療方針の推定を行う。

なお、 O_1 の分布は特定しないが、以下については仮定を置く：

$$O_2 = \beta_{10} + \beta_{11} O_1 + (\psi_{10} + \psi_{11} O_1)(A_1 + 1)/2 + \varepsilon_1$$

$$Y(A_1, A_2) = \beta_{20} + \beta_{21} O_2 + (\psi_{20} + \psi_{21} O_2 + \psi_{22}(A_1 + 1)/2)(A_2 + 1)/2 + \varepsilon_2$$

ここで、 ε_1 および ε_2 は平均が 0 の確率変数である。分散については指定しない。

構造ネスト平均モデルにより最適な動的治療方針を推定する場合にも Q 関数の場合と同様に、最終治療ステージの最適 blip 関数のパラメータ推定から後ろ向きに推定を行う。

ステージ 2 では最適 blip 関数は、

$$E[Y(A_1, A_2) - Y(A_1, -1) | H_2] = (\psi_{20} + \psi_{21} O_2 + \psi_{22}(A_1 + 1)/2)(A_2 + 1)/2$$

となることから、 $\underline{d}_2^{opt} = \text{sign}(\psi_{20} + \psi_{21} O_2 + \psi_{22}(A_1 + 1)/2)$ が得られる。もし符号関数の中が 0 になる場合は、 $\underline{d}_2^{opt} = 0$ とする。この考えは一般に“治療 = 0”は標準治療やプラセボを置くことが多いため、治療効果がない場合には試験治療より標準治療やプラセボを考えるほうがよいことに基づく。

次にステージ 1 では最適 blip 関数は、

$$E[Y(A_1, A_2^{opt}(\bar{O}_2, A_1)) - Y(-1, A_2^{opt}(\bar{O}_2, -1)) | H_1]$$

となる (A_2^{opt} は引数に基づくステージ 2 での最適な治療を表す) ことから、

$$d_1^{opt} = \operatorname{argmax}_{A_1 \in \{-1, 1\}} E[(\beta_{21}\psi_{10} + \beta_{21}\psi_{11}O_1)(A_1 + 1)/2 + (\varepsilon' + c_1 + c_2)(\operatorname{sign}(\varepsilon' + c_1 + c_2) + 1)/2 | H_1]$$

が得られる。ただし、

$$\begin{aligned}\varepsilon' &= \psi_{21}\varepsilon_1 \\ c_1 &= \psi_{20} + \psi_{21}\beta_{10} + \psi_{21}\beta_{11}O_1 \\ c_2 &= (\psi_{22} + \psi_{21}\psi_{10} + \psi_{21}\psi_{11}O_1)(A_1 + 1)/2\end{aligned}$$

ここで d_1^{opt} に着目すると、 ε_1 という背景情報に関わる確率変数が含まれており、正規分布などの仮定が必要になる。このように構造ネスト平均モデルは背景情報やアウトカムの分布に仮定が必要となる手法である。この制約を緩めた推定手法として、次節の G 推定やリグレット関数に基づく推定が考えられる。

参考文献

17. He W, Chen Y, Yin Z. (2016). Adaptive neural network control of an uncertain robot with full-state constraints. IEEE Transactions on Cybernetics, 46(3), 620-629.

2.3 G 推定

Robins(2004)にて、最適 blip 関数もしくはリグレット関数のパラメータ ψ の G 推定による推定方法が提案されている。また、ある特定の場合には Q 学習の結果と同等となることが示されている(Chakraborty et al. 2010)。本 2.3 節では最適 blip 関数を用いる。2.3.1 節で、最適 blip 関数を用いた G 推定の概要を紹介し、2.3.2 節で、単純な状況でのシミュレーション例に基づき、G 推定と ordinary least squares (OLS) の解析モデリングの違いと各推定量の特徴を紹介する。

2.3.1 方法論

本節では、最適動的治療計画 (Dynamic treatment regime; DTR) の推定に対する G 推定の概要を紹介する。まず初めに、最適 blip 関数 γ を用いた G 推定方程式を説明し、次いで、二重頑健性、推定方程式が明示解を持つ条件、Q 学習と同等の結果を与える条件、について補足する。

G 推定は、共変量 L による条件付き交換可能性 $Y^a \perp\!\!\!\perp A | L$ for all a の下、推定関数 $U(\psi, \alpha) = G_i(\psi)\{A_i - E[A_i | L_i; \alpha]\}$ を用いた推定方程式 $0 = E[U(\psi, \alpha)]$ により、因果パラメータの推定値 $\hat{\psi}$ を得る推定方法である。 A_i がランダム化されている場合には条件付き交換可能性が成立す

る。ここで $G_i(\psi)$ は、 $G_i(\psi) = Y_i - \gamma(L_i, A_i; \psi)$ として、例えば reference レジメンが無治療だった場合の潜在結果 Y^0 などを因果パラメータ ψ を用いてモデル化する。つまり、条件付き交換可能性が成立し、 $G_i(\psi)$ のモデル化が正しい場合には、因果パラメータ ψ の真値にて、推定方程式が成立する。この関係を用いて因果パラメータの推定値 $\hat{\psi}$ を得るのである。なお、ここの $\gamma(L_i, A_i; \psi)$ は一般的に最適 blip 関数に限定されるものではなく、任意のモデルを用いることが出来る。

最適 DTRs の推定においても、同様の条件付き交換可能性の下、実際受けた治療の任意の関数 $S_j(A_j)$ を用い、推定関数として次の式

$$U(\psi, \alpha) = \sum_{j=1}^K G_j(\psi) \{S_j(A_j) - E[S_j(A_j)|H_j; \alpha]\} \quad (2.3.1 - 1)$$

を設定することで因果パラメータの推定値を得ることが出来る。因果効果をバイアス無く推定するためには、 $G_j(\psi)$ 及び $E[S_j(A_j)|H_j; \alpha]$ が正しくモデル化され、複数の最適な治療が存在しないことが必要である。DTR の推測では、上記に加え、blip 関数の形、治療や中間アウトカム分布も考慮しなくてはならない。詳しくは、Chakraborty and Moodie (2013)の 8.2 節を参照されたい。以下、 $G_j(\psi)$ 及び $S_j(A_j)$ について順に説明する。

$G_j(\psi)$ は「j-1 時点までの治療歴 h_j (実際に観察された治療)で、j 時点以降は最適治療を受ける場合の潜在結果」である。 $G_j(\psi)$ は具体的には以下のように blip 関数 $\gamma(h_k, a_j; \psi)$ を用いて定義される。

$$\begin{aligned} G_j(\psi) &= Y + \sum_{k=j}^K [\gamma_k(h_k, \underline{d}_k^{opt}; \psi) - \gamma_k(h_k, a_k; \psi)] \\ &= Y + \sum_{k=j}^K E[\{Y(\bar{a}_{k-1}, \underline{d}_k^{opt}) - Y(\bar{a}_{k-1}, 0_k, \underline{d}_{k+1}^{opt})\} - \{Y(\bar{a}_k, \underline{d}_{k+1}^{opt}) - Y(\bar{a}_{k-1}, 0_k, \underline{d}_{k+1}^{opt})\} | H_k = h_k] \\ &= Y + \sum_{k=j}^K E[Y(\bar{a}_{k-1}, \underline{d}_k^{opt}) - Y(\bar{a}_k, \underline{d}_{k+1}^{opt}) | H_k = h_k] \end{aligned}$$

K=2 の場合を考える。j=2 の場合、 $G_j(\psi)$ で j=2 とし、期待値を取ると、 $E[Y(a_1, d_2^{opt})]$ となり、「時点 1 で観察治療を受け、時点 2 で最適治療を受ける潜在結果の期待値」が得られる。また、j=1 の場合、 $G_j(\psi)$ で j=1 とし、期待値を取ると、 $E[Y(d_1^{opt})]$ となり、「時点 1 及び 2 で最適治療を受ける潜在結果の期待値」が得られる。

$S_j(A_j)$ は、実際受けた治療の任意の関数である。 $S_j(H_j, A_j)$ として、治療との交互作用を考慮した変数 H_j を含めることも出来る。K=2 の場合、 $S_j(A_j)$ として例えば以下のものが考えられる。(注：以下の式中の「 \cdot 」は積を表す。Chakraborty and Moodie (2013)の表記に従った。)

$$S_1(A_1) = (A_1 + 1)/2 \cdot (1, 0_1)^T$$

$$S_2(A_2) = (A_2 + 1)/2 \cdot (1, O_1, A_1, O_1 A_1)^T$$

$S_j(A_j)$ の指定は任意ではあるものの、推定効率に影響を与える。効率を高めるための $S_j(A_j)$ の形は通常複雑である。また、 $E[S_j(A_j)]$ がデザイン上既知であったとしても、推定値を用いた方がより効率が高い。なお、推定関数 $U(\psi, \alpha)$ を構成するために $E[S_j(A_j)]$ をモデル化する際、2 値の治療の場合には、一般的には logistic 回帰が用いられることが多い。

二重頑健性を持たせるためには、 $E[G_j(\psi)|H_j; \zeta]$ を追加でモデリングすることにより、以下の推定方程式が利用できる。

$$U(\psi, \zeta(\psi), \alpha) = \sum_{j=1}^K \{G_j(\psi) - E[G_j(\psi)|H_j; \zeta]\} \{S_j(A_j) - E[S_j(A_j)|H_j; \alpha]\} \quad (2.3.1 - 2)$$

このとき、 $E[S_j(A_j)]$ か $E[G_j(\psi)|H_j; \zeta]$ かどちらか一方が正しくモデリング出来れば一致推定量を与えることができ、この性質を二重頑健性という。なお、 $E[S_j(A_j)]$ の議論と同様に、 $E[G_j(\psi)|H_j; \zeta]$ の真のモデル・パラメータが既知であったとしても推定値を使う方が効率が良い。

推定方程式が明示的に解けるためには、blip 関数が ψ に関して線型で、パラメータが stage 間で共有されている必要がある（注：Chakraborty and Moodie (2013)では「共有されていない必要がある」と記載されているが、誤記であると思われる）。明示的に解けない場合には、 $G_j(\psi)$ に修正を施した $G_{mod,j}(\psi)$ を用いた再帰計算が提案されている。詳細は、Chakraborty and Moodie (2013)の 4.3.2 節を参照されたい

G 推定の特珠な場合には、Q 学習と同等の結果を与える。Q 関数に線形モデルを用いる場合、以下の条件で代数的に同等となる。

- (i) Q_1^{opt} と Q_2^{opt} のパラメータが共有されていない
- (ii) A_j の $H_j = (\bar{O}_j, \bar{A}_{j-1})$ での条件付き平均がゼロ
- (iii) Q_1^{opt} のモデルで用いる共変量が、 Q_2^{opt} のモデルで用いる共変量に含まれる。

なお、同等であることの証明は Chakraborty and Moodie (2013)の 4.3.4 節を参照されたい。

参考文献

18. Robins JM. (2004). Optimal structural nested models for optimal sequential decisions. In Lin and Heagerty (Eds.), Proceedings of the second Seattle symposium on biostatistics (pp. 189-326). NewYork: Springer
19. Charkraborty B, Murphy SA, Strecher V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. Statistical Methods in Medical Research, 19, 317-343.

2.3.2 シミュレーション

本節では、1 段階の動的治療に G 推定と OLS を適用するシミュレーション[Chakraborty and Moodie, 2013])を基にした各モデルの性能評価を紹介する。

1 段階の動的治療を想定する。O は背景情報（予後因子）、A は予後因子 O に依存する動的な曝露変数、Y は予後因子及び曝露変数に依存する結果変数である。各変数 O、A、Y のデータ発生モデルは以下の通りである。なお、Y の発生シナリオは 2 パターン設定した。

- $O \sim \text{Uniform}(-0.5, 3)$
- A : $\text{expit}(-2 + 1.8O)$ の確率で 1 をとり、 $1 - \text{expit}(-2 + 1.8O)$ の確率で -1 をとる
- Y (シナリオ 1) : $Y(A) = -1.4 + 0.8O + (5 + 2O)(A + 1)/2 + \epsilon$
- Y (シナリオ 2) : $Y(A) = -1.4O^3 + e^3 + (5 + 2O)(A + 1)/2 + \epsilon$

両シナリオとも真の因果パラメータは、 $\psi_0 = 5$ 、 $\psi_1 = 2$ である。因果パラメータに直接的な関係のない背景情報 O に依存する結果変数のデータ発生モデルは、シナリオ 1 では線型、シナリオ 2 では非線型である。また、シミュレーション回数は 1000 回である。

比較する解析モデルは以下の通り。

G 推定は、線形の最適 blip 関数を用い、 $S(A) = (A + 1)/2(1, O)^T$ とし、二重頑健性無しもしくは有りの 2 つの解析モデル「G 推定(非 DR)」及び「G 推定(DR)」を考えた。ここで、DR とは二重頑健性を指す英語 doubly-robust を示す略語として利用した。また、DR では、 $E[G_j(\psi)|H_j; \zeta]$ を H_j の線形関数としてモデル化した。

また、OLS の解析モデルは、最適 blip 関数から以下の関係式を考え、

$$E[Y|O = o, A = a] - E[Y|O = o, A = -1] = (\psi_0 + \psi_1 o)(a + 1)/2$$

次のモデル式を想定した。

$$E[Y|O = o, A = a] = (\psi_0 + \psi_1 o)(a + 1)/2 + b(o)$$

ここで、 $b(o)$ は、未治療患者での結果 Y の推定値であり、セミパラメトリックである G 推定ではモデリングされない部分である。OLS では、この $b(o)$ に線型もしくは spline を当てはめる 2 つの解析モデル「OLS ($b(o)$ 線型)」及び「OLS ($b(o)$ spline)」を考えた。

シミュレーション結果は表 2.3.2-1 となった。シナリオ 1 は、データ発生モデルにて動的治療の因果パラメータ及び局外パラメータについて線型であり、4 つの全ての解析モデルがパラメトリックもしくはセミパラメトリックに正しく特定されている状況である。バイアス ($\hat{\psi}$ 列の推定値 $\hat{\psi}_0$ および $\hat{\psi}_1$ と各行の真値との差) は全ての解析モデルで小さかった。ばらつき (SE 列) は局外パラメータをより特定している解析モデルで小さく、パラメトリックモデルである OLS($b(o)$ 線型)で最小であり、次に、G 推定(DR)が続いた。これにより、二重頑健性を得るために追加でモデリングされた $E[G_j(\psi)|H_j; \zeta]$ の利用による効率上昇が確認さ

れた。G 推定(非 DR)と OLS(b(o)spline)のばらつきは、他の 2 つのモデルよりも相対的に大きく、同程度であった。

シナリオ 2 は、データ発生モデルにて動的治療の因果パラメータが線型で全ての解析モデルで正しく特定されているものの、局外パラメータはデータ発生モデルで非線型であり、解析モデルによっては誤特定を生じている状況である。バイアスは、OLS(b(o)線型)で顕著に大きく、OLS(b(o)spline)でも大きかった。G 推定(非 DR)が最小のバイアスだったものの、ばらつきは最大だった。G 推定(DR)では、ばらつきがかなり小さくなったものの、バイアスは標本数が小さい場合に増大した。

以下に本シミュレーション結果の考察を与える。特にシナリオ 2 のような状況で解析する可能性を念頭に置くと G 推定の利用が望ましく思われる。しかし、2.3.1 節で紹介したように、一般的に G 推定の推定方程式が明示的に解けるとは限らず、その場合、再帰的手順により、推定はより複雑なものとなる。また、簡便な OLS を用いた場合には、b(o)に spline を用いた場合でもバイアスが大きいシナリオも確認されている。ここで述べた OLS での局外パラメータの誤特定の問題については、次節で示す A 学習により対応が可能である。本節と同様に G 推定との比較検討も含め、確認されたい。

表 2.3.2-1.シミュレーション結果 (Chakraborty and Moodie, 2013 より引用)

n	ψ	G 推定(非 DR)		G 推定(DR)		OLS(b(o)線型)		OLS(b(o)spline)	
		$\hat{\psi}$	SE	$\hat{\psi}$	SE	$\hat{\psi}$	SE	$\hat{\psi}$	SE
【シナリオ 1】 $Y(A) = -1.4 + 0.80 + (5 + 20)(A + 1)/2 + \epsilon$									
50	$\psi_0 = 5$	5.183	1.071	5.020	0.777	5.029	0.599	5.062	1.004
	$\psi_1 = 2$	1.864	0.823	1.978	0.563	1.975	0.412	1.938	0.778
100	$\psi_0 = 5$	5.040	0.602	4.984	0.476	0.381	0.381	4.996	0.598
	$\psi_1 = 2$	1.963	0.464	2.000	0.355	2.007	0.263	1.993	0.480
1000	$\psi_0 = 5$	5.008	0.176	5.001	0.150	5.000	0.123	4.998	0.172
	$\psi_1 = 2$	1.993	0.136	1.998	0.111	1.999	0.084	2.001	0.130
【シナリオ 2】 $Y(A) = -1.40^3 + e^3 + (5 + 20)(A + 1)/2 + \epsilon$									
50	$\psi_0 = 5$	4.626	3.846	5.655	1.917	11.358	2.389	7.449	2.018
	$\psi_1 = 2$	2.167	3.494	1.452	1.501	-3.134	1.585	-0.018	1.695
100	$\psi_0 = 5$	4.940	1.893	5.318	1.187	10.982	1.541	6.817	1.208
	$\psi_1 = 2$	1.944	1.980	1.680	0.990	-2.907	1.098	0.523	1.086
1000	$\psi_0 = 5$	4.981	0.481	5.011	0.355	10.726	0.476	6.319	0.286
	$\psi_1 = 2$	2.011	0.550	1.990	0.295	-2.654	0.337	1.001	0.244

2.4 A 学習、リグレットに基づく方法

2.4.1 A 学習

A 学習 (Advantage learning) は、アドバンテージ (リグレットの負の値の関数)、リグレット、または同様の量のモデル化に焦点を当てることにより、最適な動的レジメを特定する方法である。リグレットは Q 関数における差であった。最適な DTR を特定するために必要なのは差のみであるため、完全な Q 関数ではなく、Q 関数の処置 A に関する項の差のみのモデリングに焦点を当てた DTR 推定へのアプローチが、A 学習である。Q 学習と比較して、処置に関係しないパラメータを含まないことにより、すなわち、推定手順の完全なデータ分布への依存を減らすことにより、モデルの誤特定に対する頑健性が得られる。例えば、Q 関数と未治療のアウトカムの差、 $\tilde{\mu}_j = Q_j^{opt}(H_j, A_j) - Q_j^{opt}(H_j, -1)$ (-1 は、プラセボや通常のケアなどの標準治療または無治療を示す) などである。Q 関数は、アドバンテージ関数と価値関数の合計として記述できるため、A 学習には、Q 関数の一部、つまり Q 関数の意思決定または調整部分に特化した交互作用項のみのモデルが必要であることがわかる。

A 学習の手順は、各段階 j で、以下の推定方程式を解く。

$$\{Y - C(H_j, A_j; \Psi_j) - \theta(H_K)\} \{S_j(A_j) - E[S_j(A_j)|H_j; \alpha]\} = 0$$

ここで、 $C(H_j, A_j; \Psi_j)$ は Q 関数の差である。Q 関数の線形回帰モデル、 $\beta_j^T H_{j0} + (\Psi_j^T H_{j1})(A_j + 1)/2$ 、については、 $C(H_j, A_j; \Psi_j) = (\Psi_j^T H_{j1})(A_j + 1)/2$ となり、最適 blip 関数と同じである。より効率的な G 推定と同様に、上記の推定方程式を使用した A 学習は二重頑健となる。すなわち、DTR パラメータ Ψ の一致推定量を取得するには、 $S_j(A_j) - E[S_j(A_j)|H_j; \alpha]$ 、 $Y - C(H_j, A_j; \Psi_j) - \theta(H_K)$ 、のいずれかが平均的にゼロであるだけで十分である (すなわち、治療またはアウトカムのいずれかを適切にモデル化するだけで十分である)。実際、 $\theta(H_K)$ の特定の選択では、A 学習と G 推定は同じである。

Schulte et al (2012) は、A 学習と Q 学習の詳細と 2 つの方法の比較について提供し、一方の方法が他方より優れている領域を特定することを目的とした幾つかのシミュレーションを行っている。シミュレーションでは、DTR パラメータ Ψ の推定効率を評価するにあたって、モンテカルロ平均二乗誤差に注目している。2-stage 問題で、すべてのモデルが正しく指定されている場合、Q 学習は A 学習と比較して、2nd stage のパラメータを推定するのにほぼ 2 倍効率的であるのに対し、1st stage での効率アップははるかに控えめで、また、治療モデルが誤特定された場合でも、Q 学習は A 学習よりも効率的だった。ただし、Q 関数が誤特定された場合、発生するバイアスが、Q 学習による効率アップを示すパラメータの値を明らかに上回る場合があり、平均二乗誤差の観点から A 学習が好ましいものとされる。

2.4.2 リグレットに基づく方法

Murphy (2003) は、全ての c と Ψ について、以下の式を満たす $(\hat{\Psi}, \hat{c})$ を求めることで、最適レジメのパラメータ Ψ を推定できることを明らかにした。

$$\sum_{j=1}^K \mathbb{P}_n \left[Y + \hat{c} + \sum_{k=1}^K \mu_k(H_k, A_k; \hat{\Psi}) - \sum_a \mu_j(H_j, a; \hat{\Psi}) p_j(a|H_j; \hat{\alpha}) \right]^2$$

$$\leq \sum_{j=1}^K \mathbb{P}_n \left[Y + c + \sum_{k \neq j} \mu_k(H_k, A_k; \Psi) + \mu_j(H_j, A_j; \Psi) - \sum_a \mu_j(H_j, a; \Psi) p_j(a|H_j; \hat{\alpha}) \right]^2 \quad (2.4.2-1)$$

治療確率、すなわち傾向スコアのパラメータ α は、G推定と同じ方法で推定できる。スカラー量 c は、 $\hat{c} = -\mathbb{P}_n(Y)$ 、すなわち、アウトカムの標本平均の負の値であり、処理の効果が無い特別な場合を除いて、簡単に解釈できない。実際、 $c(\hat{c})$ は式(2.4.2-1)から省略でき、推定には必要ないが、手順の安定性が大幅に向上する (Murphy 2003)。治療の割り当て確率、 $p_j(A_j = 1|H_j; \alpha)$ 、が交絡変数に関して正しく特定されている場合、推定量 $\hat{\Psi}$ は、 Ψ の一致推定量となる。式(2.4.2-1)の解を見つける反復法を、iterative minimization of regrets (IMOR) と呼ぶ。Rosthøj et al. (2006) は、抗凝固薬の最適用量の推定に用いられる方法のシミュレーションに基づく証明を提示し、広範なシミュレーションを通じて収束特性を研究した。シミュレーション研究では、サンプルが小さい場合 (例えば 300)、データにかなりのノイズがある場合、研究者が探索アルゴリズムの適切な初期値を提示できない場合、IMOR が収束しない可能性があることが示唆された。治療モデルの誤特定は、深刻な収束の問題を引き起こす可能性があり、IMOR が二重頑健な手順ではないことを示している。

G推定とIMORの相対的な効率に関して、次の点を指摘することができる：

(i) 治療効果がないという帰無仮説の下では、IMORは、 $E[G_{mod,j}(\Psi)|H_j]$ が一定であると想定される式(2.3.1-2)を使用したG推定の特殊なケースである。

(ii) 正則条件下では、(交絡因子に関する)傾向スコアと予想される反事実モデルの両方が正しく指定されている場合、式(2.3.1-2)からの推定は、特定の関数 $S(A_j)$ を使用するG推定のクラスの中で最も効率的である。(Robins 2004、定理 3.3 (ii)、3.4)

(iii) 式(2.3.1-2)は、 $\Psi = 0$ の近傍での推定方程式が微分可能でなく、帰無仮説の下での正則条件を満たしていない。ただし、条件は一定のblip関数に当てはまる。 $\gamma_j(h_j, a_j) = a_j \Psi_j$ (j に依存する可能性があるが h_j には依存しない可能性がある)は、治療の交互作用を想定していない。

したがって、帰無仮説が成り立ち、一定のblipモデル (治療効果がないという帰無仮説の下で自明に正しく指定される) を推定する場合、 $E[G_{mod,j}(\Psi)|H_j = h_j] = E[Y|H_j = h_j]$ が H_j に依存し、式(2.3.1-2)のG推定で正しく指定されているとき、G推定はIMORより効率的である。 $E[G_{mod,j}(\Psi)|H_j = h_j]$ が一定の場合、IMORと式(2.3.1-2)は効率的な推定量を生成する。

2.4.3 弱視の治療のための眼帯療法の 2-stage 仮想試験シミュレーション

36~96 か月の子供が、弱視の治療を受けたと仮定し、12週間にわたって眼帯を着用、8週時に検査を行うとする。アウトカムは、子供の視力と眼帯を着用することによって耐えられる心理的ストレスの測定を組み込んだ utility 関数とする。

変数の分布は次のように定義する：

Age: $O_1 \sim Uniform(36,96)$

8 週時視力: $O_2 = O_1 + (1 - 0.8O_1)(A_1 + 1)/2 + \epsilon$

最適治療下アウトカム: $Y(A_1^{opt}, A_2^{opt}) = |\delta| + O_1/400 - O_2/30$

ここで、 $\epsilon \sim \mathcal{N}(0,0.05)$ と $\delta \sim \mathcal{N}(0,0.12)$ は互いに独立で、 \bar{A}_2 、 \bar{O}_2 、 Y も独立とする。

処理 A_j は、確率 p_j で値 1 を取り、 $p_1 = \text{expit}(2 - 0.03O_1)$ 、 $p_2 = \text{expit}(-0.1 + 0.5O_2)$ とする。

最適な blip は、

$$\gamma_1(H_1, A_1) = E[Y(A_1, A_2^{opt}) - Y(0, A_2^{opt}) | H_1, A_1] = (18 - 0.3O_1)(A_1 + 1)/2$$

$$\gamma_2(H_2, A_2) = E[Y(A_1, A_2) - Y(A_1, 0) | H_2, A_2] = (4 + 0.5O_2 + A_1)(A_2 + 1)/2$$

対応するリグレットは、

$$\mu_1(H_1, A_1) = |18 - 0.3O_1|(A_1 - A_1^{opt})^2/4$$

$$\mu_2(H_2, A_2) = |4 + 0.5O_2 + A_1|(A_2 - A_2^{opt})^2/4$$

とし、additive local rank preservation の仮定の下で観察結果として、 $Y(A_1, A_2) = Y(A_1^{opt}, A_2^{opt}) - \mu_1(H_1, A_1) - \mu_2(H_2, A_2)$

とする。

1,000 回のシミュレーションでは、バイアスに関して G 推定と IMOR の両方が良好に機能する (表 2.4.3-1)。予想される反事実モデルが誤特定されている：最適な blip が線形である場合、予想される反事実、 $E[G_{mod,j}(\Psi) | H_j, A_{j-1}]$ 、は互いに素で区分線形となる (Moodie et al. 2007)。これらのシミュレーションでは、 $E[G_{mod,j}(\Psi) | H_j, A_{j-1}]$ のモデルが正しくない G 推定も、IMOR も、6 つの決定ルールパラメータすべてにおいて、ばらつき (SE 列) が他方と比べて極端に小さいということは無く、効率の点で他方に対する優位性は認められていない。

表 2.4.3-1 サンプルサイズ 500 の 1,000 データセットにおける 2-stage の G 推定と IMOR の比較：弱視の治療のための眼帯療法の仮想試験

Ψ	G 推定				IMOR	
	Eq. (2.3.1-1)		Eq. (2.3.1-2)			
	$\hat{\Psi}$	SE	$\hat{\Psi}$	SE	$\hat{\Psi}$	SE
$n = 500$						
$\Psi_{10} = 18.0$	17.845	2.554	17.984	0.717	17.996	0.561
$\Psi_{11} = -0.3$	-0.298	0.038	-0.300	0.009	-0.300	0.008
$\Psi_{20} = 3.0$	3.059	2.122	3.000	0.652	3.023	0.416
$\Psi_{21} = 0.5$	0.456	3.001	0.507	0.857	0.475	0.635
$\Psi_{22} = 2.0$	1.952	2.750	2.006	0.728	1.986	0.613
$\Psi_{23} = 0.0$	0.016	4.630	-0.031	1.017	0.010	0.985
$n = 1,000$						
$\Psi_{10} = 18.0$	17.914	1.734	17.998	0.470	18.019	0.381

$\Psi_{11} = -0.3$	-0.299	0.026	-0.300	0.006	-0.300	0.006
$\Psi_{20} = 3.0$	3.031	1.463	2.994	0.447	3.014	0.281
$\Psi_{21} = 0.5$	0.469	2.068	0.506	0.582	0.483	0.432
$\Psi_{22} = 2.0$	1.968	1.886	2.002	0.508	1.987	0.418
$\Psi_{23} = 0.0$	0.028	3.171	-0.008	0.700	0.017	0.674
$n = 2,000$						
$\Psi_{10} = 18.0$	17.916	1.197	17.997	0.327	18.006	0.269
$\Psi_{11} = -0.3$	-0.299	0.018	-0.300	0.004	-0.300	0.004
$\Psi_{20} = 3.0$	3.085	1.012	3.018	0.314	3.018	0.193
$\Psi_{21} = 0.5$	0.386	1.432	0.473	0.412	0.473	0.297
$\Psi_{22} = 2.0$	1.902	1.302	1.976	0.355	1.976	0.288
$\Psi_{23} = 0.0$	0.144	2.196	0.030	0.487	0.038	0.463

2.4.4 まとめ

リグレットに基づく推定方法は、多くの場合 G 推定に類似しており（場合によっては同一）、リグレットを blip 関数に、またはその逆に変換できるようにするための単純な変換ができる。

構造ネスト平均モデル、G 推定、リグレットに基づく方法は、ランダム化されていないデータに適用可能で、モデル仕様への依存を減らし、同時に精度を向上させるために、これらのいくつかを二重頑健にすることができる。

すべての方法は、いくつかの仮定の妥当性に依存しており、そのうちのいくつかは検証不可だが、モデル診断と検討下の健康状態についての専門知識を使用して少なくともインフォーマルに評価できる。

実用上、Q 学習か G 推定やリグレットに基づく推定を用いるのかは、処置に関係しないデータも統計モデルで考慮したいかどうかになる。事前に調整したい共変量が定まっている場合は、Q 学習を用いる方がよく、情報があまりない場合はむやみにモデルにデータを入れず G 推定やリグレットに基づく推定を用いるとよいと考える。

参考文献

- Moodie EEM, Richardson TS, Stephens DA. (2007) Demystifying optimal dynamic treatment regimes. *Biometrics*, 63, 447–455.
- Murphy S. A. (2003) Optimal dynamic treatment regimes (with Discussion). *Journal of the Royal Statistical Society, Series B*, 65, 331–366.
- Robins JM (2004) Optimal structural nested models for optimal sequential decisions. In D. Y. Lin & P. Heagerty (Eds.), *Proceedings of the second Seattle symposium on biostatistics* (pp. 189–326). New York: Springer

23. Rosthøj S, Fullwood C, Henderson R, Stewart S. (2006) Estimation of optimal dynamic anticoagulation regimes from observational data: A regret-based approach. *Statistics in Medicine*, 25, 4197–4215.
24. Schulte PJ, Tsiatis AA, Laber EB, Davidian M (2012) Q- and A-learning methods for estimating optimal dynamic treatment regimes. *arXiv*, 1202.4177v1.

3.統計プログラムの実装例

R の `iqLearn` パッケージ²を用いた解析例を紹介する。サンプルデータセットである `bmiData` は思春期の肥満に対するミール・リプレースメント・シェイクの効果を研究した2段階 SMART である。背景情報は性別、人種、親の BMI、ベースライン BMI がある。1 段階目の処置後の中間アウトカムは 4 か月後の BMI (`month4BMI`) である。A1 は1段階目に `meal replacement(MR)` 又は `conventional diet(CD)` にランダムに割り当てられた処置で、A2 は2段階目に MR 又は CD にランダムに割り当てられた処置である。主要アウトカムは 12 か月後の BMI (`month12BMI`) である。

表 3 `bmiData` の一部データ

ID	gender	race	parent_BMI	baseline_BMI	month4_BMI	month12_BMI	A1	A2
1	0	1	31.59683	35.84005	34.22717	34.27263	CD	MR
2	1	0	30.17564	37.30396	36.38014	36.38401	CD	MR
3	1	0	30.27918	36.83889	34.42168	34.41447	MR	CD
4	1	0	27.49256	36.70679	32.52011	32.52397	CD	CD
5	1	1	26.4235	34.84207	33.72922	33.73546	CD	CD
6	0	0	29.3097	36.6864	32.06622	32.15977	MR	MR
7	0	1	25.9005	35.42631	32.54064	32.51531	CD	MR

以下に Q 学習による線形モデルを用いた治療方針を推定するためのサンプルプログラムを示す。

[サンプルプログラム]

```
install.packages("iqLearn")
library(iqLearn)
## 2段階の BMI data データを読み込む
data(bmiData)
bmiData$A1[which(bmiData$A1=="MR")] = 1
bmiData$A1[which(bmiData$A1=="CD")] = -1
```

² 以下からファイルをダウンロードし、ローカルからインストールする必要がある。

<https://cran.r-project.org/src/contrib/Archive/iqLearn/>

```

bmiData$A2[which (bmiData$A2=="MR")] = 1
bmiData$A2[which (bmiData$A2=="CD")] = -1
## ダミー変数化
bmiData$A1 = as.numeric (bmiData$A1)
bmiData$A2 = as.numeric (bmiData$A2)
## y のベースラインからの変化率を計算
y = -(bmiData[,6] - bmiData[,4])/bmiData[,4]
## 2 段階目の Q 関数のパラメータを推定
fitIQ2 = learnIQ2 (y ~ gender + parent_BMI + month4_BMI + A2*(parent_BMI +
month4_BMI),data=bmiData,"A2",c("parent_BMI","month4_BMI"))
## 2 段階目の Q 関数のパラメータの推定結果を表示
summary (fitIQ2)

```

2 段階目の治療方針に関するパラメータの推定結果

	推定値	標準誤差
intercept	0.4128845	0.0398789
gender	-0.0064891	0.0089924
parent_BMI	-0.0015509	0.0010236
month4_BMI	-0.0082067	0.0013992
A2	-0.0738709	0.0397545
parent_BMI×A2	0.0020223	0.0010201
month4_BMI × A2	0.0002816	0.0013982

治療方針は推定値を代入した Q 関数を最大化する処置でもあることから、A2=1 と A2=-1 のそれぞれを代入した場合の最大の Q 関数を持つ処置を最適な処置として選ぶ。

```

## 新しい患者(gender=1, parent_BMI=30, month4_BMI=45)の最適な処置を考える
h2 = c (1, 30, 45)
optIQ2 = IQ2 (fitIQ2, h2)
optIQ2$q2opt
## A2=1 の Q 関数の値は-0.009962029、A2=-1 の Q 関数の値は-0.008904261 となり、A2=-1
で Q 関数が最大となるため A2=-1 が最適な処置である。

```

```

## 一段階の Q 関数の推定
fitIQ1main = learnIQ1main (~ gender + race + parent_BMI + baseline_BMI+ A1*(gender +

```

```
parent_BMI), data=bmiData, "A1", c("gender","parent_BMI"), fitIQ2)
summary (fitIQ1main)
```

1 段階目の治療方針に関するパラメータの推定結果

	推定値	標準誤差
intercept	0.4029745	0.0125558
gender	-0.0062882	0.0024014
race	-0.0014183	0.0024233
parent_BMI	-0.0037081	0.0002276
baseline_BMI	-0.0054769	0.0003475
A1	0.0505355	0.0072522
gender×A1	0.0018455	0.0024083
parent_BMI×A1	-0.0016380	0.0002115

新しい患者(gender=1,race=1,parent_BMI=30, baseline_BMI=35)の最終アウトカムまでの処置効果の改善を見据えた 1 段階目の最適な処置を考える

h1 = c (1, 1, 30, 35)

A1=1

```
fitIQ1main$alphaHat0[1]+fitIQ1main$alphaHat0[2:5]*%*%h1+A1*(fitIQ1main$alphaHat1[1]+h1[c(1,3)]*%*%fitIQ1main$alphaHat1[2:3])
```

A1=-1

```
fitIQ1main$alphaHat0[1]+fitIQ1main$alphaHat0[2:5]*%*%h1+A1*(fitIQ1main$alphaHat1[1]+h1[c(1,3)]*%*%fitIQ1main$alphaHat1[2:3])
```

A1=1 の Q 関数の値は 0.09557631、A1=-1 の Q 関数の値は 0.08909254 となり、A1=1 で Q 関数が最大となるため A1=1 が最適な処置である。

Appendix SMART のサンプルサイズ設計

1章のアルコール依存症患者対象の SMART の例を用いてサンプルサイズ設計法を考える。本試験で最も興味があるのは「二次治療を周辺化した場合の平均における最適な一次治療とは何か？」であるとする。つまり、認知行動療法を最初に割り付けられた集団の平均アウトカムとオピオイド受容体拮抗薬を最初に割り付けられた集団の平均アウトカムに一番興味があるとする。

Cohen(1988)より標準化された一次治療の治療効果の大きさ δ を以下のように定義する。

$$\delta = \frac{E(Y|A_1 = \text{オピオイド受容体拮抗薬}) - E(Y|A_1 = \text{認知行動療法})}{\sqrt{(\text{Var}(Y|A_1 = \text{オピオイド受容体拮抗薬}) + \text{Var}(Y|A_1 = \text{認知行動療法})) / 2}}$$

各群の割付確率は 0.5 とする。サンプルサイズ n は $n = 4(z_{\alpha/2} + z_{\beta})^2 \delta^{-2}$ より得られる。 $z_{\alpha/2}$ は標準正規分布の $(1 - \alpha/2)$ 分位点、 z_{β} は標準正規分布の $(1 - \beta)$ 分位点である。

その他の例も考える。一番興味があるのは初期治療のレスポonderにおける電話モニタリングと電話モニタリング&カウンセリングのどちらの方が最良の治療なのか、であるとする。標準化された治療効果の大きさ δ を以下のように定義する。

$$\delta = \frac{E(Y|\text{レスポonder}, A_2 = \text{電話モニタリング}) - E(Y|\text{レスポonder}, A_2 = \text{電話モニタリング\&カウンセリング})}{\sqrt{(\text{Var}(Y|\text{レスポonder}, A_2 = \text{電話モニタリング}) + \text{Var}(Y|\text{レスポonder}, A_2 = \text{電話モニタリング\&カウンセリング})) / 2}}$$

全体のレスポonderの割合を γ とする。各群の割付確率は 0.5 とする。サンプルサイズ n は $n = 4(z_{\alpha/2} + z_{\beta})^2 \delta^{-2} \gamma^{-1}$ より得られる。 $z_{\alpha/2}$ は標準正規分布の $(1 - \alpha/2)$ 分位点、 z_{β} は標準正規分布の $(1 - \beta)$ 分位点である。

動的処置計画全体の差に興味がある場合を考える。二つの治療方針を \bar{d} 及び \bar{d}' とする。治療の割付がレジメン \bar{d} と一致するサブグループと治療の割付がレジメン \bar{d}' と一致するサブグループを比較することを考える。標準化された治療効果の大きさ δ を以下のように定義する。

$$\delta = \frac{(\mu_{\bar{d}} - \mu_{\bar{d}'})}{\sqrt{(\sigma_{\bar{d}}^2 + \sigma_{\bar{d}'}^2) / 2}}$$

$\mu_{\bar{d}}$ はレジメン \bar{d} の下での平均主要アウトカム、 $\sigma_{\bar{d}}^2$ は分散である。各群の割付確率は 0.5 とする。サンプルサイズ n は $n = 8(z_{\alpha/2} + z_{\beta})^2 \delta^{-2}$ より得られる。

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2021 年度タスクフォース 4
/ 2022 年度 継続タスクフォース 4

Estimand SMART チームメンバー、担当

小島 将裕	協和キリン株式会社、1 章、2.1.1 項（共著）、2.1.2 項、3 章、チームリーダー
小谷 基	小野薬品工業株式会社、2.1.1 項起草、2.1.2 項起草（2021 年 9 月 16 日まで）
中西 豊支	興和株式会社、2.1.1 項（共著）
竹ノ内 一雅	日本イーライリリー株式会社、2.4 節
寺田 亘	アステラス製薬株式会社、2 章冒頭、2.2 節、全体の確認
土川 克	ゼリア新薬工業株式会社、2.3 節
尾崎 凌斗	中外製薬株式会社、報告書レビュー
作井 将	武田薬品工業株式会社、報告書レビュー

タスクフォースリーダー兼推進委員

土川 克	ゼリア新薬工業株式会社
------	-------------

担当副部長

土屋 悟	住友ファーマ株式会社
菅波 秀規	興和株式会社

以上