

What If 勉強会メモ：What If で出てくる算数

Tarotan (@BluesNoNo)

2020 年 5 月 30 日

このメモの目的は、以下の本（通称は”What If”. このメモでは、以下、テキストと呼ぶ）で出てくる数式の展開を平易に解説することである.

Hernán, M.A. and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/02/ci_hernanrobins_21feb20.pdf

確率の初歩を知っている人ならばある程度は分かるように心がけたのだが、現段階では、その企みは失敗している. なお、筆者の力量不足のため、確率論から見て厳密でない記述が多くある.

5 月 29 日 19:30 現在、読書会では第 4 章までが終了している

1 参考にしてほしい資料／参考にした資料

この節では、統計学での因果推論について学びたい初心者向けの動画や文献を紹介する.すでにテキストが何を言っているのか分からなくなっている人は、是非ともこれらの動画や文献に目を通してほしい.

1.1 初等確率論についての動画や教科書

初等確率論の初歩については、What If 読書会の入門講義

Sato, S. (2020). Causal Inference: What If 勉強会 疫学のための確率の基礎

<https://www.youtube.com/watch?v=efT7gPjTpHI>

で分かりやすく解説されている．とりあえず，こちらを視聴してほしい．

初等確率論のより進んだ内容に興味があるのであれば，

山本直樹（2011）慶應大学工学部 応用確率論 2011 年度

<https://www.youtube.com/watch?v=bHBTEsYC-YM>

が，初心者向けの親切的な講義となっている（上記の URL は，第 1 回の講義のもの）．

英語が苦にならない方には，

Blitzstein, J. K. (2013). Statistics 110, Harvard University

<https://www.youtube.com/watch?v=KbB0FjPg0mw>

をおすすめしたい（上記の URL は，第 1 回の講義のもの）．この講義が面白かった人は，

Blitzstein, J. K. and Hwang, J. (2015). *Introduction to Probability*. CRC Press.

という教科書もある．このメモの「アダムの公式」や「LOTUS」などの変わった名称は，この講義や教科書から借用している．

以上の動画を見れば，大学教養レベルの初等確率論はマスターできると思う（... というか，以上で挙げた講義以上の知識を私はもっていない）．

1.2 因果推論についての動画や教科書（日本語）

潜在反応モデルをまったく知らない人は，まずは，次の 2 つの動画をすすめたい．

田栗正隆（2019）潜在反応モデルと交絡調整の方法．

<https://www.youtube.com/watch?v=5hXwoJnCwGA>

篠崎智大（2020）因果推論入門：交絡とその調整

<https://www.icrweb.jp/course/view.php?id=412>

後者は，ICR 臨床研究入門のサイトであり，受講・閲覧するには，登録（無料）が必要である．

やや数学的にはマイルドで，臨床試験や疫学の例に基づく説明の方がわかりやすい方には，次の動画をすすめたい．

佐藤俊哉（2018）ランダム化ができないとき

佐藤俊哉（2018）交絡とその調整

佐藤俊哉 (2019) 回帰モデルと傾向スコア

<https://ocw.kyoto-u.ac.jp/ja/graduate-school-of-medicine-jp/12/video>

Pearl のバックドア基準に関しては、次の動画をすすめたい。

林岳彦 (2019) バックドア基準入門: ステップ・バイ・ステップでの解説.

<https://www.youtube.com/watch?v=AqifHlVi6LE>

潜在反応モデルや Pearl 流因果推論に関しては、以上の動画に比べると難易度は上がるが、すでに多くの日本語の教科書や文献がある（上記で紹介した動画の方がやさしいので、まずは動画を視聴することを個人的にはすすめたい）。

念のために断っておきたいのだが、昨今の Rubin 流や Pearl 流が流行る前にも、因果推論については日本でも話題に上がっている。ちょっと昔に遡っただけでも、たとえば、1978 年の『第 6 回 行動計量学会総会 発表論文抄録集』では「因果性をめぐる問題」というパネルディスカッションが開かれている。1986 年の雑誌『行動計量学』では「特集・因果関係の推定」という特集が組まれている。大正・明治に遡っても、きっと議論が見られるだろうが、ここでは無視する。

最近の因果推論におけるいろいろな立場を包括的に取り上げている入門書としては、

岩波データサイエンス 刊行委員会編（立森久照・林岳彦・伊庭幸人・星野崇宏 担当編集）(2016)『岩波データサイエンス Vol.3 因果推論：実世界のデータから因果を読む』岩波書店

がある。無味乾燥な説明ではなく、たとえば「モンテカルロ法と傾向スコア」など面白い見方も紹介されていて、楽しく読める。同書は、サポートページ

<https://sites.google.com/site/iwanamidatascience/vol-3/vol3-ingasuiron>

も充実している。

オーソドックスな初等統計学が好きで、Rubin 流の反実仮想モデルに入門したい場合には、次の 1 冊をすすめたい。

岩崎学 (2015)『統計的因果推論』朝倉書店

製造業関連で回帰分析を愛用している人には、次の 1 冊をすすめたい。Pearl 流の do 演算 (set 演算) や有向非巡回グラフ (DAG) が話題の中心ではあるが、前半には Rubin 流の潜在反応モデルも解説されている。

宮川雅巳 (2004) 『統計的因果推論：回帰分析の新しい枠組み』朝倉書店

Pearl が共著者となっている次の本には日本語翻訳がある。

Pearl, J., Glymour, M. and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons. (落海浩訳 2019 『入門 統計的因果推論』朝倉書店)

(Pearl 単著の有名な本にも日本語翻訳があるが、少なくとも私の実力では読めない。Pearl の日本語文献に目を通したい場合は、まずは、上記の本を読むことをすすめたい。)

Pearl 流の枠組みについてさらに詳しく学びたい人には、

黒木学 (2017) 『構造的因果モデルの基礎』共立出版

をすすめたい。

Rubin 流因果推論のマーケティングやウェブ調査への応用に関心がある方には、次の 1 冊をすすめたい。

星野崇宏 (2009) 『調査観察データの統計科学：因果推論・選択バイアス・データ融合』岩波書店

ただし、難易度は高い。

経済政策や社会政策でのランダム化比較試験については、私は未読だが、ノーベル経済学賞を受賞した Esther Duflo の本が日本語に翻訳されている。

Duflo, E., Glennerster, R. and Kremer, M. (2008). *Handbook of Development Economics, Vol.4*. Elsevier B. V. (小林庸平 [監訳] 石川貴之・井上領介・名取淳 [訳] (2019) 『政策評価のための因果関係の見つけ方』日本評論社)

流行に遅れないように、目を通したい (... と言っているあいだに読めばいいのに)。

構造方程式モデルに焦点をあてて、因果推論を紹介しているものとしては、

狩野裕 (2002) 「構造方程式モデリング、因果推論、そして非正規性」甘利俊一・狩野裕・佐藤俊哉・松山裕・竹内啓・石黒真木夫『統計科学のフロンティア 5 多変量

解析の展開』岩波書店, 65-129

がある。構造方程式モデルは、たしか 1970 年後半ぐらいから、統計ソフトウェアの発展とともに、社会学および心理学で爆発的に流行した手法である。その祖先は、おそらく、社会学や遺伝学でのパス解析、心理学での因子分析、経済学での構造モデルなどが挙げられよう。

同書には、臨床試験や疫学での因果推論を紹介している

佐藤俊哉・松山裕 (2002) 「疫学・臨床研究における因果推論」 甘利俊一・狩野裕・佐藤俊哉・松山裕・竹内啓・石黒真木夫『統計科学のフロンティア 5 多変量解析の展開』岩波書店, 131-175

という論文も含まれている。

1.3 因果推論についての教科書（英語）

英語にも入門者向けの文献は、数多くある。ここでは代表的だと思われるものを取り上げる。

まず、Rubin 流の考え方や解析手順を知りたいのであれば、

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: Introduction*. Cambridge University Press

をおすすめる。ただし、同書の前書きに書いてあるように、Pearl 流の手続きをガン無視している。Pearl 流も学びたい方にはおすすめるしない。

Rubin 流を学ぶには、

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*. **100(469)**. 322-331.

という論文に Rubin の考え方のエッセンスが簡潔にまとめられているので、おすすめである。Rubin 流についての有名な解説論文としては

Holland, P. W. (1986) Statistics and Causal Inference. *Journal of the American Statistical Association*, **81(396)**. 945-960.

がある。Rubin が遅くとも 1975 年には述べていた「操作なくして因果なし」のスローガンは、この論文でも紹介されている。Holland は、Rubin よりもさらに「(実際の) 操作」に拘っていると思う。私が知る限り、Holland が一番極端な立場であり、(教育分野において) “race” や “sex” の “因果効果” を考えても仕方がないという立場をとっていると思う。この論文へのコメントで、Clark Glymour が哲学での Lewis 反事実条件などとの関連性が指摘されている。潜在反応モデルは、たびたび、「反事実仮想モデル」などとも呼ばれることもあるが、それはこのコメント以降に普及したのだろう。ちなみに、上記の Rubin(2005) を読む限り、Rubin さんは「反事実」ではなく「潜在反応 (potential outcome)」という呼び名を好んでいる。

Pearl 流の基本方針を知るには、

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Causal and Effect*. Penguin Science.

をおすすめしたい。初等統計学しか知らない人にはクセは強いと思われるが、読んでいて楽しい。なお、Pearl さん本人も Twitter アカウントをもっていて (@yudapearl)、ときどき論争している。たとえば、Lord の逆説について、Stephen Senn さん (@stephensenn) とかなり長い論争を繰り返し行っている。

社会科学向けの実践的な教科書としては

Morgan, S. L. and Winship, C. (2007:1st ed., 2015:2nd ed.). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

がある。同書は、Rubin 流と Pearl 流を融合させた実践的な入門書として、Twitter にて林岳彦さん (@takehikohayasi) が何度も推奨している。

1.4 お断り・謝辞

このメモで述べられていることは、すべて Tarotan (@BluesNoNo) に責任があるが、述べられることに独創性はなく、基本的には、ここで述べた教科書や動画で言われていることを再構成しただけである (誤読や誤解をすることで、意図しない独創性が生まれてしまっている可能性は否定できないが)。メモでは、参照元を明記していない (つまり、「この式展開は、〇〇の xx ページから引用したものである」などと断っていない)。しかし、ほぼすべてのことは、上記したような教科書や動画に基づいている (... つもりである)。

2 テキストで登場する数式展開の詳説

2.1 このメモでよく用いる公式

この項では、確率論における公式のうち、テキストの数式展開で必要と思われるものを紹介していく。

2.1.1 期待値の（ほぼ定義である）公式

「公式」と言うよりもほぼ「定義」だと思われるが、期待値は以下の計算式で求められる。

- Y が離散型の確率変数である場合

$$E[Y] = \sum_y y \Pr(Y = y)$$

ただし、このメモでは、 $\Pr(Y = y) = 0$ であるものは、上記の計算から省くものとする。つまり、

$$E[Y] = \sum_{y: \Pr(Y=y) \neq 0} y \Pr(Y = y)$$

とする。

- Y が連続型の確率変数である場合

$$E[Y] = \int_{-\infty}^{+\infty} y f_y(y) dy$$

ここで $f_y(\cdot)$ は、 Y の確率密度関数である。

私が知る由もない尊い理由により（たぶん、離散型と連続型が混じった場合に対応できるし、測度論的により厳密な議論ができるなどの理由なのだろう）、上2つの式は、次のような「スティルチェス積分」なるもので統一的に表現することもある。

$$E[Y] = \int_{-\infty}^{+\infty} y dF_y(y)$$

このスティルチェス積分は、テキスト p.7 の Technical Point 1.1 に登場し、読書会の参会者をビビらせた。

さらに、もっと私には意味不明な尊い理由により（たぶん、きちんと確率空間を定義するところから議論するとそうなるのだろう）、次のような雰囲気でも表現することもある（以下の式は、私の個人的なイメージです。正しいのかどうか、また、その意味は分かっていません）。

$$E[Y] = \int_{\omega \in \Omega} Y(\omega) dF_Y(\omega)$$

すみません.... 厳密な話は分かりません....

2.1.2 期待値の線形性

期待値には、線形性と呼ばれる性質がある。 a, b を定数、 X, Y を確率変数とすると、

$$E[aX + bY] = aE[X] + bE[Y]$$

という関係がある。私のような浅学者が注意すべきことは、「 X, Y が独立でなくても、この公式（期待値の線形性）は成立するんだ！」という点である。もちろん、独立な時でも期待値の線形性は成立するが、独立な場合に限らず、どんな時でも成立する。どんな時でも成立することの証明は、読者の練習問題とする。

2.1.3 LOTUS (蓮)

参考:https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician

以下の公式も、教科書によっては「定義」として紹介されることもある。

確率変数 Y をある関数 $g(\cdot)$ で変数変換した場合、 $g(Y)$ の期待値は次のように計算できる。

- Y が離散型の確率変数である場合

$$E[g(Y)] = \sum_{y: \Pr(Y=y) \neq 0} g(y) \Pr(Y = y)$$

- Y が連続型の確率変数である場合

$$E[g(Y)] = \int_{-\infty}^{+\infty} g(y) f_Y(y) dy$$

この公式は、“Law Of The Unconscious Statistician”（「無自覚な統計家の公式」）の頭文字をとって、“LOTUS”（「蓮」）と呼ばれることがある。証明は読者の練習問題とする。

2.1.4 アダムの公式（全期待値の公式）

参考：https://en.wikipedia.org/wiki/Law_of_total_expectation

全体平均は、「『各部分の平均』の平均」である．たとえば，全体の 40% における平均が 11.8 で，残り 60% における平均が 19.7 であれば，これら 2 つを併合したものの全体平均は

$$0.4 \times 11.8 + 0.6 \times 19.7$$

である．これと似たことを難解な記号で書くと，

$$E[Y] = E_X[E_{Y|X}[Y|X]]$$

となる．左辺は周辺期待値（周辺平均）である．また，左辺の括弧内における $E_{Y|X}[Y|X]$ は，条件付き期待値（条件付き平均）である．難しい言葉で上式を表現すると，“周辺平均は「『条件付き平均』の平均」である”と言えよう．

$g(X) = E_{Y|X}[Y|X]$ として 2.1.3 における LOTUS を適用すると， X が離散型の確率変数である場合には，

$$\begin{aligned} E[Y] &= E_X[E_{Y|X}[Y|X]] \\ &= \sum_x E_{Y|X=x}[Y|X=x] \Pr(X=x) \end{aligned}$$

となる． X が連続型の確率変数である場合には，

$$\begin{aligned} E[Y] &= E_X[E_{Y|X}[Y|X]] \\ &= \int_{-\infty}^{+\infty} E_{Y|X=x}[Y|X=x] f(x) dx \end{aligned}$$

となる．

2.1.5 独立であるときの確率や期待値

すべての x, y に関して，離散型の場合には $\Pr(X=x, Y=y) = \Pr(X=x)\Pr(Y=y)$ ，連続型の場合には $f_{(x,y)}(x,y) = f_x(x)f_y(y)$ である場合，「 X, Y は独立である」と呼ぶことにする．ここで $f_{(x,y)}(\cdot)$, $f_x(\cdot)$, $f_y(\cdot)$ は，それぞれ， (X, Y) の同時確率密度関数， X の確率密度関数， Y の確率密度関数である．

いま，1 個のサイコロと 1 枚のコインを同時に投げるとする．サイコロを投げたときに出る目を Y としよう．また，コイン投げの結果を X とし，コインの表が出たら 1，裏が

出たら 0 としよう。サイコロの目が 6 になる確率は、コインが表になろうが、裏になろうが関係ないと思われる。このことを数式で表現すると、

$$\Pr(Y = 6) = \Pr(Y = 6|X = 1) = \Pr(Y = 6|X = 0)$$

となる。より一般的には、 X, Y が独立である場合、すべての x, y に関して

$$\Pr(Y = y|X = x) = \Pr(Y = y)$$

である。「すべての x, y 」と述べたが、「 $\Pr(X = x, Y = y) > 0$ となっているすべての x, y について」と述べるべきであろう。以下、そのあたりの厳密さは無視する。

上式は Y が離散型の確率変数であるときの公式である。 Y が連続型の確率変数である場合には、確率質量関数 \Pr が、確率密度関数 f に置き換わる。

$$f(y|x) = f(y)$$

なお、紛らわしさを回避するには、上式の確率密度関数には、

$$f_{y|x}(y|x) = f_y(y)$$

と添え字を添えたほうがいいだろうが、面倒なので省いた（以下でも、そのときの気分によって、添え字をつけたり、つけなかったりする）。

これらは、逆向きでも成立する。 X が離散型の確率変数である場合には、

$$\Pr(X = x|Y = y) = \Pr(X = x)$$

が、 X が連続型の確率変数である場合には、

$$f_{x|y}(x|y) = f_x(x)$$

が成立する。

期待値にも、似た関係がある。 X, Y が独立である場合には、すべての x について

$$E[Y|X = x] = E[Y]$$

となる。ここでも、紛らわしさを回避するには、

$$E_{Y|X=x}[Y|X = x] = E_Y[Y]$$

と添え字をつけたほうがいいだろうが、面倒だから省いた。以下でも、その場の体調や雰囲気などで添え字を省くので、ごめんなさい。

なお、逆も成立する。 X, Y が独立である場合には、すべての y について

$$E_{X|Y=y} [X|Y = y] = E_X [X]$$

が成立する。

X, Y が独立であることは、 $X \perp\!\!\!\perp Y$ と記すことがある。

2.1.6 条件付き独立であるときの確率や期待値

すべての x, y, z について、離散型の場合には $\Pr(Y = y, Z = z|X = x) = \Pr(Y = y|X = x)\Pr(Z = z|X = x)$ 、連続型の場合には $f(y, z|x) = f(y|x)f(z|x)$ である場合、「 X で条件付けたもとで、 Y と Z は条件付き独立である」などと呼ぶことにする。また、記号で $Y \perp\!\!\!\perp Z|X$ と記すことにする。

$Y \perp\!\!\!\perp Z|X$ である場合、次の関係が成り立つ。

- 離散型の確率変数である場合

$$\Pr(Y = y|X = x, Z = z) = \Pr(Y = y|X = x)$$

$$\Pr(Z = Z|X = x, Y = y) = \Pr(Z = z|X = x)$$

- 連続型の確率変数である場合

$$f(y|x, z) = f(y|x)$$

$$f(z|x, y) = f(z|x)$$

期待値にも、似たような関係が成立する。

$$E[Y|X = x, Z = z] = E[Y|X = x]$$

$$E[Z|Y = x, Z = z] = E[Z|X = x]$$

2.1.7 条件付き期待値で定数のように扱えるケース

条件付き期待値において条件付けた変数は、次のように定数のように扱える。

$$E_{(X,Y)|X} [XY|X] = X E_{Y|X} [Y|X]$$

2.2 潜在反応モデルでよく仮定される前提

現場の状況によって、使われる統計モデルや前提は臨機応変に変えるべきであろう。潜在反応モデルの仲間にはさまざまな統計モデルがあり、それらの前提もいろいろある。

この章では、潜在反応モデルでよく使われる3つの前提を紹介する。それら「3種の神器」は、一致性・交換可能性・正值性の3つである。もちろん、現実においては、これら3つが成立しないとみなすのが自然な場面も多々あるだろう。しかし、それほど突拍子もない仮定でもないだろう（...と思います）。何よりも、これらたった3つの仮定をおいただけで、潜在反応の平均などが識別できるようになるのうれしい。

2.2.1 記号の説明

潜在反応モデルにおいては、観測される反応 Y 、処置 A だけではなく、潜在反応 $Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}$ が想定される。

このメモでは、基本的にこれらすべてが確率変数であるとする。特定の個体 i の潜在反応 $y_i^{a=1}, y_i^{a=2}, \dots, y_i^{a=k}$ を固定値として議論を展開していく立場もあるが、テキストでは（少なくとも4章までは）個体レベルの議論は行なっていない（もしくは、個体レベルでの議論を曖昧にしていると思う）。とりあえず、少なくとも集団レベルで見たときには、 $Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}$ は確率変数とみなせるとし、個体レベルの記述は不問にして、話を進めていくことにしよう。

A は、 k 個の離散的な値をとるとする ($A = 1, 2, \dots, k$)。 $Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}$ は、連続変数であっても、離散変数であっても、どちらでも構わないが、このメモでは、記法を単純にするため、0,1 の2値変数か、もしくは、連続変数とする。この潜在反応モデルにおいて、 $Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, A$ のなかで私たちが観測できるのは、基本的には Y, A だけだとする。

以上が最も基本となるモデルではあるが、テキストの第4章までだと、これ以外にも、「予後因子」や「効果修飾因子」と呼ばれている因子が観察されるとする。それらを、それぞれ L_1, L_2, \dots, L_d および V_1, V_2, \dots, V_e と表すことにする。

つまり、潜在反応モデルは、 $(k + d + e + 2)$ 個の確率変数を要素とする確率ベクトル $(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ に対する統計モデルである。このとき、入門書の最初で描かれるようなモデルでは、この確率ベクトルに対して、パラメトリックな確率分布を想定しない（たとえば、 $(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ が多変量正規分布に従

う」などの前提を置かない。もちろん、実践においては、特定の確率分布を想定することもある（数多くある）。潜在反応モデルに対して、入門書の最初で想定される前提は、以下で述べる一貫性・交換可能性・正値性の3つだけである。

2.2.2 一貫性

このメモでの一貫性とは、 $(k + d + e + 2)$ 個の確率変数を要素とする確率ベクトル $(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ において、「 $A = j$ であるときは $Y = Y^{a=j}$ 」となることを指す。

このメモではあくまでこの意味で「一貫性」と言うが、潜在反応モデルでの議論では、個体レベルについて言及し、個体レベルの潜在反応は固定値であるとみなす場合もある。つまり、ある特定の個体 i に関して、潜在反応 $y_i^{a=1}, y_i^{a=2}, \dots, y_i^{a=k}$ を固定値として、かつ、 $A = j$ のときに $y_i = y_i^{a=j}$ となることを「一貫性」と呼ぶ場合もある。

一貫性の仮定は、別の表現をすることもできる。一貫性が成立している場合、観察される反応 Y は次のようにも表される。

$$\begin{aligned} Y = & I(A = 1) \times Y^{a=1} + \\ & I(A = 2) \times Y^{a=2} + \\ & I(A = 3) \times Y^{a=3} + \\ & \dots + \\ & I(A = k) \times Y^{a=k} \end{aligned}$$

ここで $I(\cdot)$ は指示関数であり、引数の関係式が成立していれば 1, 成立していなければ 0 を返す。たとえば、 $A = 2$ が割り付けられた場合、観測される反応 Y は、

$$Y = 0 \times Y^{a=1} + 1 \times Y^{a=2} + 0 \times Y^{a=3} + \dots + 0 \times Y^{a=k}$$

となるので、 $Y = Y^{a=2}$ となる。

なお、統計学の漸近論においては、一貫性 (consistency) は、別の意味で使われている。漸近論では、「 $n \rightarrow \infty$ であるときに推定量 $\hat{\theta}$ が、真値 θ にほぼ等しくなること」を「一貫性」と呼んでいる。漸近的な一貫性をもつ推定量を、「一致推定量」などと呼ぶ。このメモの一貫性は、漸近論での一貫性とは関係ない。ちなみに、どうでもいい余談だが、David(2001)によると、漸近論の意味での“consistency”は、Fisher(1922)が初出である。

2.2.3 交換可能性：周辺ランダム化での交換可能性

ごく単純なランダム化試験を想定し、また、潜在反応モデルを仮定する。このとき、処置の確率変数 A は、他の確率変数とは独立であるとみなせよう。

$$A \perp\!\!\!\perp (Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$$

このように A と $(Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ が独立であることは、テキストの p.15 では「完全交換可能性」と呼ばれている。

完全交換可能性が成立している場合、 $j = 1, 2, \dots, k$ に対して、次のような関係も成立する。

$$\begin{aligned} \Pr(Y^{a=j} = y^{a=j}) &= \Pr(Y^{a=j} = y^{a=j} | A = 1) \\ &= \Pr(Y^{a=j} = y^{a=j} | A = 2) \\ &= \dots \\ &= \Pr(Y^{a=j} = y^{a=j} | A = k) \end{aligned}$$

$L_1, L_2, \dots, L_k, V_1, V_2, \dots, V_k$ にも、同じような関係式が成立するが、ここでは省略する。

完全交換可能性が成立している場合、平均に関しても、似たような関係が成立する。

$$\begin{aligned} E[Y^{a=j}] &= E[Y^{a=j} | A = 1] \\ &= E[Y^{a=j} | A = 2] \\ &= \dots \\ &= E[Y^{a=j} | A = k] \end{aligned}$$

このような期待値に関する交換可能性は、「平均交換可能性」と呼ばれている。完全交換可能性が成立していれば、平均交換可能性は成立する。しかし、逆は必ずしも成り立たない。

2 つ注意をしておきたい。1 つ目は、ここでの独立性は、統計モデルのなかでの独立性であるということである。標本のデータにおいては、 $y^{a=1}, y^{a=2}, \dots, y^{a=k}$ や $l_1, l_2, \dots, l_d, v_1, v_2, \dots, v_e$ には当然、アンバランスが生じるだろう。有限のデータでぴったりバランスが取れることはありえない。特に小標本であれば、アンバランスさが甚だしいものになる可能性は高くなるだろう。さらに予後因子や効果修飾因子が無数あったとしたら、有限のデータではどれかには甚だしいアンバランスが生じるだろう。

2 つ目は、「推測の問題」と「識別の問題」を区別してほしいということである。テキストの 4 章までで展開されているのは、あくまで識別の問題である。識別できなければ推定することはできないが、識別できたからといって実際の場面で良い推定量が得られるとは限らない。

話を戻そう。議論をするときに、必ずしも全部の確率変数を考慮する必要はない。たとえば、もしも、 Y, A の確率分布から第 j 番目の潜在反応における平均 $E[Y^{a=j}]$ だけを識

別したいのであれば、その他の確率変数は気にしないでいい。よって、仮定としては

$$A \perp\!\!\!\perp Y^{a=j}$$

だけが必要である。また、第4章では、ごく単純なランダム化比較試験で、効果修飾因子 V の部分集団での効果を識別する問題が説明されている。そのモデルは、1つの効果修飾因子を V と表すと、上と同じく第 j 番目の潜在反応における平均 $E[Y^{a=j}]$ だけを識別したいのであれば、

$$A \perp\!\!\!\perp (Y^{a=j}, V)$$

という仮定だけが必要である。

2.2.4 交換可能性：条件付きランダム化での交換可能性

L_1 で条件づけたランダム化試験を想定し、また、潜在反応モデルを仮定する。 L_1 は、連続変数であっても、離散変数であっても、どちらでも構わない。このとき、 $A \perp\!\!\!\perp (Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, L_2, \dots, L_d, V_1, V_2, \dots, V_e) | L_1$ を仮定するのは自然だろう。この L_1 で条件づけた時に A とその他の確率変数が独立になることを、「条件付き（完全）交換可能性」と呼ぶ。

条件付き（完全）交換可能性が成立している場合、 $j = 1, 2, \dots, k$ に関して、次のような関係が成立する。

$$\begin{aligned} \Pr(A = j | & \begin{aligned} & L_1 = l_1, \\ & Y^{a=1} = y^{a=1}, Y^{a=2} = y^{a=2} \dots, Y^{a=k} = y^{a=k}, \\ & L_2 = l_2, \dots, L_d = l_d, \\ & V_1 = v_1, \dots, V_e = v_e \end{aligned} \\ = & \Pr(A = j | L_1 = l_1) \end{aligned}$$

つまり、条件付き（完全）交換可能性が成立していれば、 $L_1 = l_1$ および潜在反応などの変数で条件付けた A の条件付き確率は、 l_1 だけの関数となる。

2.2.5 正值性

周辺ランダム化において、すべての j について、 $\Pr(A = j) > 0$ となっていること、もしくは、条件付きランダム化において、すべての j, l_1 について、 $\Pr(A = j | L_1 = l_1) > 0$ となっていることを、「正值性がある」という。注意して欲しいのは、この前提も、あくまでモデルのなかで満たされていればいいだけであり、得られる観測値やデータに対する前提ではないという点である。

2.3 周辺ランダム化での潜在反応平均

周辺ランダム化を行った試験，すなわち，通常の単純なランダム化試験から， $E[Y^{a=j}]$ を求めたいとする．

$(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ という， $(k + d + e + 2)$ 個の確率変数を要素とする確率ベクトルがあるとする．この確率ベクトルの要素のうち，潜在反応 $Y^{a=j}$ ，観測される反応 Y ，処置 A に関し，特定の j について，以下の条件が満たされているとする．

- 一致性： $A = j$ であれば $Y = Y^{a=j}$ である
- 完全交換可能性： $A \perp\!\!\!\perp Y^{a=j}$
- 正值性： $\Pr(A = j) > 0$

2 番目の完全交換可能性は，より緩い条件である平均交換可能性 ($E[Y^{a=j}|A = j] = E[Y^{a=j}]$) であっても構わない．

このとき， $E[Y|A = j]$ は次のように変形できる．

$$\begin{aligned} E[Y|A = j] &= (\text{一致性より}) \quad E[Y^{a=j}|A = j] \\ &= (\text{交換可能性より}) \quad E[Y^{a=j}] \end{aligned}$$

ただし， $\Pr(A = j) = 0$ であれば， $A = j$ の集合が空っぽになるので， $E[Y|A = j]$ は不定である．

以上のことから，上記の 3 つの仮定（一致性・交換可能性・正值性）が成立していれば， $E[Y^{a=j}]$ を識別でき，そのとき， $E[Y^{a=j}] = E[Y|A = j]$ である．

2.4 周辺ランダム化での部分集団の潜在反応平均

この項で扱う問題は，テキストの 4 章 2 節 (pp.43-45) で扱われている．

ここでも，前項と同様，周辺ランダム化を行った試験，すなわち，通常の単純なランダム化試験を想定する．ただし，求めたいパラメータは，前項と異なり， $E[Y^{a=j}|V_1 = v_1]$ であるとする．

前項と同様， $(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ という， $(k + d + e + 2)$ 個の確率変数を要素とする確率ベクトルがあるとする．この確率ベクトルの要素のうち，潜在反応 $Y^{a=j}$ ，観測される反応 Y ，処置 A ，効果修飾因子 V_1 に関し，特定の j について，以下の条件が満たされているとする．

- 一致性： $A = j$ であれば $Y = Y^{a=j}$ である.
- 完全交換可能性： $A \perp\!\!\!\perp (Y^{a=j}, V_1)$
- 正值性： $\Pr(A = j, V_1 = v_1) > 0$

2 番目の完全交換可能性は、より緩い条件である条件付き平均交換可能性 ($E[Y^{a=j}|A = j, V_1 = v_1] = E[Y^{a=j}|V_1 = v_1]$) であっても構わない.

このとき、 $E[Y|A = j, V_1 = v_1]$ は次のように変形できる.

$$\begin{aligned} E[Y|A = j, V_1 = v_1] &= (\text{一致性より}) & E[Y^{a=j}|A = j, V_1 = v_1] \\ &= (\text{交換可能性より}) & E[Y^{a=j}|V_1 = v_1] \end{aligned}$$

ただし、 $\Pr(A = j, V_1 = v_1) = 0$ であれば、 $E[Y|A = j, V_1 = v_1]$ は不定である.

以上のことから、上記の 3 つの仮定 (一致性・交換可能性・正值性) が成立していれば、 $E[Y^{a=j}|V_1 = v_1]$ を識別でき、そのとき、 $E[Y^{a=j}|V_1 = v_1] = E[Y|A = j, V_1 = v_1]$ である.

2.5 条件付きランダム化での潜在反応平均：標準化で求める場合

この項では、条件付きランダム化を想定する. ある因子 L_1 で条件付けて、 A を割り付けるとする. 求めたいパラメータは、 $E[Y^{a=j}]$ であるとする.

前項や前々項と同様、 $(A, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L_1, L_2, \dots, L_d, V_1, V_2, \dots, V_e)$ という、 $(k + d + e + 2)$ 個の確率変数を要素とする確率ベクトルがあるとする. この確率ベクトルの要素のうち、潜在反応 $Y^{a=j}$ 、観測される反応 Y 、処置 A 、因子 L_1 に関し、特定の j について、以下の条件が満たされているとする.

- 一致性： $A = j$ であれば $Y = Y^{a=j}$ である.
- 条件付き完全交換可能性： $A \perp\!\!\!\perp Y^{a=j} | L_1$
- 正值性： $\Pr(A = j, L_1 = l_1) > 0$

2 番目の条件付き完全交換可能性は、より緩い条件である条件付き平均交換可能性 ($E[Y^{a=j}|A = j, L_1 = l_1] = E[Y^{a=j}|L_1 = l_1]$) であっても構わない.

このとき、 $E[Y^{a=j}]$ は次のように変形できる.

$$\begin{aligned}
E[Y^{a=j}] &= (\text{アダムの公式より}) \\
&\quad E_{L_1} [E_{Y^{a=j}|L} [Y^{a=j}|L_1]] \\
&= (\text{条件付き交換可能性より}) \\
&\quad E_{L_1} [E_{Y^{a=j}|A=j, L_1} [Y^{a=j}|A=j, L_1]] \\
&= (\text{LOTUS より}) \\
&\quad \sum_{l_1} \Pr(L_1 = l_1) \times E_{Y^{a=j}|A=j, L_1=l_1} [Y^{a=j}|A=j, L_1 = l_1]
\end{aligned}$$

ただし、 $\Pr(A = j, L_1 = l_1) = 0$ であれば、 $E[Y|A = j, L_1 = l_1]$ は不定である。

以上のことから、上記の 3 つの仮定（一致性・交換可能性・正值性）が成立していれば、 $E[Y^{a=j}]$ を識別でき、そのとき、

$$E[Y^{a=j}] = \sum_{l_1} \Pr(L_1 = l_1) \times E_{Y^{a=j}|A=j, L_1=l_1} [Y^{a=j}|A=j, L_1 = l_1]$$

である。

2.6 条件付きランダム化での潜在反応平均：逆確率重み付き法で求める場合

この項では、テキストの Technical Point 2.3 (p.24) での式展開を丁寧に追う。テキストとは、やや異なる式展開を用いている。

前項と同様、条件付きランダム化を想定し、求めたいパラメータは、 $E[Y^{a=j}]$ であるとする。

記法を簡略化するため、 $Y^{a=1}, Y^{a=2}, \dots, Y^{a=k}, Y, L, A$ という $k+3$ 個の確率変数をもつ確率ベクトルを考える。

次の 3 条件が満たされているとする。

- 一致性： $A = j$ であれば $Y = Y^{a=j}$ である。
- 条件付き完全交換可能性： $A \perp\!\!\!\perp Y^{a=j} | L$
- 正值性： $\Pr(A = j | L = l) > 0$

ここで次の平均を求めることにする。

$$E_{L,Y,A} \left[\frac{Y \times I(A = j)}{\Pr(A|L)} \right]$$

ここで、 $\Pr(A|L)$ という記号が分かりづらい（少なくとも、私はなかなか理解できなかった）。これは、 $L = l$ で条件付けたときの $A = a$ が生じる確率 $\Pr(A = a|L = l)$ において、 a, l の部分を確率変数 A, L とみなしたものである。よって $\Pr(A = A|L = L)$ とでも記したほうがいいのかもしいが、それは記号として間違っているから、テキストにならって $\Pr(A|L)$ とだけここでは記した。

テキストの p.32 における Technical Point 3.1 より、正值性 $e_j(L) = \Pr(A = j|L) > 0$ が成立しているならば、

$$E_{L,Y,A} \left[\frac{Y \times I(A = j)}{\Pr(A|L)} \right] = E_{L,Y,A} \left[\frac{Y \times I(A = j)}{e_j(L)} \right]$$

である。上式においては分母が微妙に違っている点に注意。よって、以下では

$$E_{L,Y,A} \left[\frac{Y \times I(A = j)}{e_j(L)} \right] = E_{L,Y^{a=1},Y^{a=2},\dots,Y^{a=k},A} \left[\frac{Y \times I(A = j)}{e_j(L)} \right]$$

を求めることにする。

ここで、上式における分母の $e_j(L)$ は、 L だけの関数になっていることに注意してほしい（ A は消えている点に注意してほしい）。

2.1.4 の「アダムの公式」より、上式は

$$E_{L,Y^{a=1},Y^{a=2},\dots,Y^{a=k}} \left[E_{A|(L,Y^{a=1},Y^{a=2},\dots,Y^{a=k})} \left[\frac{Y \times I(A = j)}{e_j(L)} \middle| L, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k} \right] \right]$$

と変形できる。

この式に、 $Y = I(A = 1) \times Y^{a=1} + I(A = 2) \times Y^{a=2} + \dots + I(A = k) \times Y^{a=k}$ を代入すると、 $I(A = j) \times I(A = j) = I(A = j)$ であり、 $j \neq j'$ のときは $I(A = j) \times I(A = j') = 0$ であるので、他の箇所はそっくりそのままのまま、上式の Y だけを $Y^{a=j}$ に置き換えることができ、

$$E_{L,Y^{a=1},Y^{a=2},\dots,Y^{a=k}} \left[E_{A|(L,Y^{a=1},Y^{a=2},\dots,Y^{a=k})} \left[\frac{Y^{a=j} \times I(A = j)}{e_j(L)} \middle| L, Y^{a=1}, Y^{a=2}, \dots, Y^{a=k} \right] \right]$$

となる。潜在反応に関しては $Y^{a=j}$ しか登場しないので、 j 以外の潜在反応は周辺化できて、上式は、

$$E_{L,Y^{a=j}} \left[E_{A|(L,Y^{a=j})} \left[\frac{Y^{a=j} \times I(A = j)}{e_j(L)} \middle| L, Y^{a=j} \right] \right]$$

となる。

上式の内側にある条件付き期待値において、 $L, Y^{a=j}$ で条件付けられている点に注目する。

また、条件付き交換可能性より、 $E_{A|(L, Y^{a=j})} [I(A = j)|L, Y^{a=j}] = E_{A|(L, Y^{a=j})} [I(A = j)|L]$ であり、0, 1 の 2 値変数の期待値は 1 が生じる確率と等しいので、 $E_{A|(L, Y^{a=j})} [I(A = j)|L] = \Pr(A = j|L) = e_j(L)$ である。つまり、 $E_{A|(L, Y^{a=j})} [I(A = j)|L, Y^{a=j}] = e_j(L)$ である。

よって、条件付き期待値の部分は

$$\begin{aligned} E_{A|(L, Y^{a=j})} \left[\frac{Y^{a=j} \times I(A = j)}{e_j(L)} \middle| L, Y^{a=j} \right] &= \frac{Y^{a=j} \times 1}{e_j(L)} \times e_j(L) + \frac{Y^{a=j} \times 0}{e_j(L)} \times (1 - e_j(L)) \\ &= Y^{a=j} + 0 \\ &= Y^{a=j} \end{aligned}$$

となる。以下のように 2.1.7 の公式を用いても、同じ結果を得られる。

$$\begin{aligned} E_{A|(L, Y^{a=j})} \left[\frac{Y^{a=j} \times I(A = j)}{e_j(L)} \middle| L, Y^{a=j} \right] &= \frac{Y^{a=j}}{e_j(L)} E_{A|L, Y^{a=j}} [I(A = j)|L, Y^{a=j}] \\ &= \frac{Y^{a=j}}{e_j(L)} e_j(L) \\ &= Y^{a=j} \end{aligned}$$

以上のことから、最終的には、

$$\begin{aligned} E_{L, Y, A} \left[\frac{Y \times I(A = j)}{e_j(L)} \right] &= E_{L, Y^{a=j}} [Y^{a=j}] \\ &= E_{Y^{a=j}} [Y^{a=j}] \end{aligned}$$

となる。つまり観測される反応と逆確率を掛けたものの期待値は、潜在反応の期待値と等しい。なお、最後の式変形は、期待値の計算に L は含まれていないので周辺化した。

逆確率重み付け法で得られる

$$E_{L, Y, A} \left[\frac{Y \times I(A = j)}{e_j(L)} \right]$$

も、前項の標準化で得られる

$$\sum_{l_1} \Pr(L_1 = l_1) \times E_{Y^{a=j}|A=j, L_1=l_1} [Y^{a=j}|A = j, L_1 = l_1]$$

も、いずれも $E_{Y^{a=j}} [Y^{a=j}]$ と等しいので、これら 2 つの方法で得られる結果は（基本的には）等しい。

2.6.1 A が連続変数である場合の逆確率重み付き法

前節では A は離散変数であったが、この節では A を連続変数とする。その場合、すべての実数 b に関して、 $\Pr(A = b|L) = 0$ つまり $\Pr(I(A = b) = 0|L) = 1$ となるので、前節のような関係が成立しなくなる。そこで、微小な Δb を導入し、前節における指示関数の部分を $I(b < A \leq b + \Delta b)$ 、また、分母の確率を $e_b(l)\Delta b = f(b|l)\Delta b$ としよう。ここで $f(b|l)$ は、条件付き分布の確率密度関数である。

$(b, b + \Delta b]$ の区間で Y が一定だとすると、前節の結果をほぼそのまま用いることができ、

$$E_{L,Y,A} \left[\frac{Y \times I(b < A \leq b + \Delta b)}{e_j(L)\Delta b} \right] = E_{Y^{a=j}} [Y^{a=j}]$$

となる。

この式は、分母に Δb が登場し、この値によって結果が変わってくるので、使い勝手が悪いような気がする（個人の感想です）。

ここで、 $E_{L,Y,A} \left[\frac{I(b < A \leq b + \Delta b)}{e_j(L)\Delta b} \right]$ という期待値を考えると、前節とほぼ同じ式展開で解けば、これは 1 である。これで上式を割ると、

$$\frac{E_{L,Y,A} \left[\frac{Y \times I(b < A \leq b + \Delta b)}{e_j(L)\Delta b} \right]}{E_{L,Y,A} \left[\frac{I(b < A \leq b + \Delta b)}{e_j(L)\Delta b} \right]} = E_{Y^{a=j}} [Y^{a=j}]$$

となり、分母と分子の Δb が消去されて、

$$\frac{E_{L,Y,A} \left[\frac{Y \times I(b < A \leq b + \Delta b)}{e_j(L)} \right]}{E_{L,Y,A} \left[\frac{I(b < A \leq b + \Delta b)}{e_j(L)} \right]} = E_{Y^{a=j}} [Y^{a=j}]$$

となる。この最後の式を使う場合にも、どれぐらいの幅で A をグループ化するかという問題が残る。ある程度、 Δb を小さくしなければいけないが、小さいと少数の観測しか推定に使われなくなってしまう。 A が連続変数の場合、単純な逆確率重み付き法を使いこな

すのは難しそうだ（何かしらの工夫が必要. テキストのこの後で発展させたモデルが出てくる?）.

2.7 条件付きランダム化での ATT：標準化で求める場合

██████████ TODO ██████████

2.8 条件付きランダム化での ATT：逆確率重み付け法で求める場合

██████████ TODO ██████████

3 文献

David, H. A. (2001). First(?) Occurrence of Common Terms in Statistics and Probability. David, H.A. and Edwards, A.W.F.(ed.) *Annotated Readings in the History of Statistics*, Springer, 209-246.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 309-368.

4 更新履歷

- 第 0 版：2020 年 5 月 30 日 PoC 版