

【購読文献】Morgan, Stephen L. and Christopher Winship, 2015, *Counterfactuals and Causal Inference Methods and Principles for Social Research: 2nd Edition*, Cambridge University Press.

Chapter 5 Matching Estimators of Causal Effects

担当：麦山 亮太（東京大学大学院人文社会系研究科修士課程）

要約

5.1 Origins of Motivations for Matching

マッチングのテクニックは、古くは 20 世紀前半の実験研究にその起源をもつ。さらに 1980 年代初頭、Rosenbaum and Rubin による一連の研究を経て、応用・開発が進んできた。

マッチングには以下の 2 つの方法がある。本章では主に後者の方法について議論する。

- (1) 準実験的な方法：処置群 (treatment group) のもつ属性とマッチするような統制群 (control group) を母集団からサンプリングすることによって、処置変数の因果効果を求める。
- (2) ノンパラメトリックな方法：すでに得られたサンプルを用いて、他の変数で層化することによって処置変数の因果効果を求める。

回帰分析とマッチングの違い：回帰分析 ($Y = \alpha + \delta D + \mathbf{X}\beta + \varepsilon^*$) は Y に対する D と \mathbf{X} の効果を同時に推定するのに対して、マッチングにおいては \mathbf{X} はたんに Y に対する D の因果効果を求める際の調整変数としてのみ用いられる。

5.2 Matching as Conditioning via Stratification

用語の確認：ATE、ATT、ATC

ATE (Average Treatment Effect)：全体における、処置の因果効果の平均値

$$E[\delta] = E[Y^1 - Y^0] = E[Y^1] - E[Y^0] \quad (2.3)$$

ATT (Average Treatment effect for the Treated)：処置群における、処置の因果効果の平均値

$$E[\delta|D = 1] = E[Y^1|D = 1] - E[Y^0|D = 1] \quad (2.7)$$

ATC (Average Treatment effect for the Controls)：統制群における、処置の因果効果の平均値

$$E[\delta|D = 0] = E[Y^1|D = 0] - E[Y^0|D = 0] \quad (2.8)$$

ATE は、ATT と ATC を、全体に占める処置群の割合 π で重みづけたものに等しい。(2.3) を変形して以下を得られる。

$$E[\delta] = \{\pi E[Y^1|D = 1] + (1 - \pi)E[Y^1|D = 0]\} - \{\pi E[Y^0|D = 1] + (1 - \pi)E[Y^0|D = 0]\} \quad (2.10)$$

ここで、 $E[Y^1|D = 0]$ と $E[Y^0|D = 1]$ は観察することができない。にもかかわらず、ATT を求めたいときはどうすればよいのか？というのが、因果推論の基本的な問題関心。

5.2.1 Estimating Causal Effects by Stratification

互いに性質は違うものの、その違いは共変量 S (割付変数、assignment/selection variables) によって完全に捉えられるような処置群と統制群があるとすると。このとき、 S は「完全な層化 (“perfect stratification”）」を可能にする、という*1。

因果効果を求める際の条件である Assumption 1 および 2 は以下のものであった。

$$\text{Assumption 1 : } E[Y^1|D = 1] = E[Y^1|D = 0] \quad (2.15)$$

$$\text{Assumption 2 : } E[Y^0|D = 1] = E[Y^0|D = 0] \quad (2.16)$$

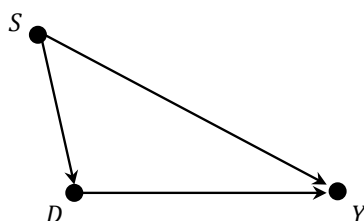
S による完全な層化が成り立つとき、これらの条件を以下のように書きなおすことができる。

$$\text{Assumption 1-S : } E[Y^1|D = 1, S] = E[Y^1|D = 0, S] \quad (5.1)$$

$$\text{Assumption 2-S : } E[Y^0|D = 1, S] = E[Y^0|D = 0, S] \quad (5.2)$$

このような S によって層化することはすなわち、Figure 4.8(b) で言うところの back-door path をブロックする、ということに等しい。

図 5.1 Figure 4.8(b)



Assumption 1-S および 2-S が成り立つとき、ATT は ATC および ATE を割付変数 S の周辺分布によって重みづけた平均値に一致し、したがって D の因果効果 δ を求めることができる。これを式で表現したものが以下。

$$\sum_s \{E_N[y_i|d_i = 1, s_i = s] - E_N[y_i|d_i = 0, s_i = s]\} \Pr_N[s_i = s] \xrightarrow{p} E[\delta] \quad (5.6)$$

$\Pr_N[s_i = s]$ を $\Pr_N[s_i = s|d_i = 1]$ に置き換えることで ATT を、 $\Pr_N[s_i = s|d_i = 0]$ に置き換えることで ATC を、それぞれ得ることができる。以上の議論から示されるとおり、因果効果を求めるときの層化のプロセスは、以下の2段階からなっている。

- (1) 割付変数 S を決めてその周辺確率分布 $\Pr[S]$ を求める
- (2) 割付変数 S の処置変数 D による条件つき確率分布 $\Pr[S|D]$ を求める

5.2.2 Overlap Conditions for Estimation of the ATE

ある層の集団が誰も処置を受けていない場合 (Sparseness, 処置変数と割付変数により作られるクロス表において空白セルができる場合) には、ATE は推定できないが、ATT は推定できる。

*1 「完全な」という語のより厳密な意味は以下。“individuals within groups defined by values on the variables in S are entirely indistinguishable from each other in all ways except for (1) observed treatment status and (2) differences in the potential outcomes that are independent of treatment status.” (p.143 最後から5行目を引用)

5.3 Matching as Weighting

すべての割付変数に関する情報があり、かつサンプルサイズが無限大である場合には先に述べたような層化による因果効果の推定が可能であるが、実際にはこのような理想的な状態を実現するのは困難であり、だれも処置を受けていなかったり、すべての人が処置を受けているような層が出現してしまう。そこで、推定された傾向スコア (estimated propensity score) にもとづく重み付け推定量 (weighting estimator) を用いるのがよい。

5.3.1 The Utility of Known Propensity Scores

推定された傾向スコアとは、処置への割り当てを予測する (独立) 変数群の関数として表される、処置を受ける確率である*2。真の (true) 傾向スコアが分かれば、ATT および ATC を知るために必要な情報がすべて揃う。したがって、因果効果をより精確に知るためには、真の傾向スコアに近い (推定された) 傾向スコアを求めればよい。

さきに、因果効果を求めるときの層化プロセスは $\Pr[S]$ を求めるという段階と $\Pr[S|D]$ を求める段階の2つの段階からなると述べた。傾向スコア=処置への割り当て確率は $\Pr[D|S]$ として表されるが、これがわかればベイズの定理より $\Pr[S|D]$ がわかる。したがって、因果効果を求めることができる。

ベイズの定理 (Bayes theorem)

事象 X も事象 Y も起こる確率は、事象 X の起こる確率と事象 X が起こったという条件のもとで事象 Y の起こる確率の積、および事象 Y の起こる確率と事象 Y が起こったという条件のもとで事象 X の起こる確率の積に等しい。

$$\Pr(X) \Pr(Y|X) = \Pr(Y) \Pr(X|Y) = \Pr(X \cap Y)$$

これを变形することで以下を得る。

$$\Pr(Y|X) = \frac{\Pr(Y) \Pr(X|Y)}{\Pr(X)}$$

5.3.2 Weighting with Estimated Propensity Scores to Address Sparseness

処置への割り当て確率を求める。割付変数 S がさまざまな値をとり、かつ真の傾向スコアが0から1の間に収まっていることが条件であることを鑑みると、以下のような2項ロジットモデルを用いて傾向スコアを求めることが考えられる*3。

$$\Pr[D = 1|S] = \frac{\exp(S\phi)}{1 + \exp(S\phi)} \quad (5.7)$$

2項ロジットモデルの推定によって、個人 i が処置に割り当てられる確率のオッズ $\hat{r}_i = \hat{p}_i / (1 - \hat{p}_i)$ が求まる。この値は、 $\hat{r}_i > 1$ のとき処置群に割り当てられやすいことを、 $\hat{r}_i < 1$ のとき統制群に割り当てられやすいことをそれぞれ意味している。これを用いて ATT、ATC、ATE をそれぞれ求める。

*2 傾向スコアを e_i とすると、 $e_i = \Pr(d_i = 1|s_i)$ のように表される。あとで述べるように、 $\Pr(d_i = 1|s_i)$ を推定するためにはロジットモデルやプロビットモデルを利用する。

*3 ロジットモデルのようなパラメトリックな方法ではなく、モデルを予め設定しないノンパラメトリック回帰によって傾向スコアを推定する方法もある。しかしその場合、説明変数が増えるにしたがって推定に必要なサンプルサイズが指数的に増加するという問題がある。これを次元の呪い (curse of dimensionality) という (星野 2009)。

ATT は、処置群における y_i の平均値から、統制群における、処置群への割り当て確率のオッズで重み付けされた個人の y_i をすべて足し合わせた値を処置群への割り当て確率のオッズの総和で割った値を引くことで求められる*4。

$$\hat{\delta}_{\text{ATT,weight}} \equiv \left(\frac{1}{n^1} \sum_{i:d_i=1} y_i \right) - \left(\frac{\sum_{i:d_i=0} \hat{r}_i y_i}{\sum_{i:d_i=0} \hat{r}_i} \right) \quad (5.12)$$

ATC は、テキストの式を少し分かりやすく書きなおすと、以下のように書ける*5*6。

$$\hat{\delta}_{\text{ATC,weight}} \equiv \left(\frac{\sum_{i:d_i=1} y_i \times \frac{1}{\hat{r}_i}}{\sum_{i:d_i=1} \frac{1}{\hat{r}_i}} \right) - \left(\frac{1}{n^0} \sum_{i:d_i=0} y_i \right) \quad (5.13)$$

ATT は、これらを処置群に入った個体の全体に占める比率で重みづけたものに等しい。

$$\hat{\delta}_{\text{ATE,weight}} \equiv \left(\frac{1}{n} \sum_i d_i \right) \left(\hat{\delta}_{\text{ATT,weight}} \right) + \left[\left(1 - \frac{1}{n} \sum_i d_i \right) \right] \left(\hat{\delta}_{\text{ATC,weight}} \right) \quad (5.14)$$

重みづけによるマッチングは、Sparseness の問題を回避することができるが、問題もある。第1に、サンプルサイズが十分に大きくなければ推定値を得ることはできない。第2に、傾向スコアを推定するためのモデルが完全に特定化されていなければ、因果効果の一致不偏推定量を得ることができない (Assumption 1-S および 2-S が成り立つ必要がある)。

5.4 Matching as a Data Analysis Algorithm

マッチング推定量を得るためのアルゴリズムは、以下の2点で異なる。(1) マッチングされるケースの数、(2) マッチングされるケースがどのように重み付けされるか。

5.4.1 Basic Variants of Matching Algorithms

Exact Matching

処置群の反事実的な値 ($E[Y^1|D=0]$) を、割付変数 S が完全に同一のケースどうしをマッチングさせることで生成する。 S が多くなるにつれて、実現可能性は顕著に低くなるため、単一ではほとんど用いられない方法。

Nearest-Neighbor, Caliper, and Radius Matching

処置群の反事実的な値を、割付変数によって求められる傾向スコアの値がもっとも近い処置群と統制群のケースどうしをマッチングさせることで生成する。傾向スコアの値がより近いグループどうしをマッチングさせることで反事実的な値をより正確に特定することができる一方で、マッチングするケースが少なすぎると結果の一般化が難しい。他方でたくさんマッチングさせると反事実的な値にはバイアスが生じる。

*4 ATT は (2.7) のように書けるのであった。(5.12) の第1項は、(2.7) の第1項 ($E[Y^1|D=1]$) と対応している。(2.7) の第2項 ($E[Y^0|D=1]$) は観察することはできないが、Assumption 2-S が成り立つのであれば、 $E[Y^0|D=0, S]$ で代用することができる。これが、(5.12) の第2項に当たる。第2項の意味は、「処置群の個体と同じ特徴を有している (処置群に入りやすい) にもかかわらず統制群に入っている個体の値をより重視して求めた重み付き平均値」として解釈することができる。

*5 ATC は (2.8) のように書けるのであった。(5.13) の第2項は、(2.8) の第2項 ($E[Y^0|D=0]$) と対応している。(2.8) の第1項 $E[Y^1|D=0]$ は観察することはできないが、Assumption 1-S が成り立つのであれば、 $E[Y^1|D=1, S]$ で代用することができる。これが、(5.13) の第1項にあたる。第1項の意味は、「統制群の個体と同じ特徴を有している (統制群に入りやすい) にもかかわらず処置群に入っている個体の値をより重視して求めた重み付き平均値」として解釈することができる。

*6 テキストの (5.13) 式の第1項の分母に n^1 が入っているが、これは不要と思われる

Interval Matching

処置群と統制群において求められた傾向スコアをもとに、それぞれの群をいくつかの区間に分割する。傾向スコアの値の近い分割されたサブグループどうしでマッチングを行う。

Kernel Matching

処置群のケースすべてを用いるが、統制群のケースは処置群のケースの傾向スコアからの距離によって重み付けする。

5.4.2 Recent Matching Routines That Seek Optimal Balance

どのマッチングの方法についても、その目的は、(傾向スコアの値に集約されるような) 処置群と統制群とで異なっている $\Pr[s_i|d_i = 1] \neq \Pr[s_i|d_i = 0]$ (5.18) 割付変数の周辺分布を、関心のある処置効果に関して調整し、マッチングされたデータにおいては、 $\Pr[s_i|d_i = 1] = \Pr[s_i|d_i = 0]$ (5.19) となるような状態を作り出すことにある。マッチングは、割付変数によって算出される傾向スコアの平均値が似通っている集団どうしでなされる。

単一のサンプルに基づく傾向スコアマッチングによって共変量の違いを完全に調整し、真の処置効果を得られるわけではない、ということには注意する必要がある。

Enhancing Routines for Balance Assessment

共変量の調整が適切に行われているかを評価するための新たな指標の開発。

Reestimating the Distance Metric

処置群と統制群の間の距離を推定するためのモデルの再考。

Optimizing the Matching Algorithm

マッチングのアルゴリズムのさらなる最適化。従来の二段階推定から同時推定へ。

Direct Coarsening of the Stratification Variables

exact matching の仮定を緩和したマッチングの方法の開発。

5.4.3 Which of These Matching Algorithms Work Best?

どのマッチングのアルゴリズムがもっとも優れているかは判断は難しい。しかし、決定的に重要な層化変数を除外していた場合、それぞれのマッチングから求められる推定値は大きく異なるものとなり、層化変数を適切に考慮していた場合は、推定値は比較的似通ったものとなる。複数の推定値を見比べながら、適切な層化変数を見つけ、共変量の調整がきちんと行われているかを見ながら分析するのがよいだろう。

5.5 Remaining Practical Issues in Matching Analysis

5.5.1 Matching When Treatment Assignment Is Nonignorable

割付変数を完全に特定できない場合でも、マッチングはデータ分析の1つの選択肢として用いてもよい。しかしこの場合真の処置効果は知ることはできないため、解釈は慎重にする必要がある。

5.5.2 Matching Only on the Region of Common Support

割付変数の分布の違いから、常に処置群のサンプル全体における処置効果を求められるとは限らない。こうしたときに、適切に処置効果を求めるサンプルを限定することが必要となる。

5.5.3 The Expected Variance of Matching Estimates

マッチングにより推定された処置効果の標準誤差の求め方については、一致した見解が得られていない。現時点では、(1) 標準誤差を求めるにあたりどのような方法を用いたのかを読者が理解できるように提示すること、(2) 単一の方法から計算された標準誤差にもとづいて結論を導かないようにすること、が望ましい。

5.5.4 Matching Estimators for Many-Valued Causes

多値の処置変数の効果を推定することもできる。当該変数が名義尺度の場合、異なる2値どうしでマッチングを行う。順序尺度の場合、共変量の分布が近くなるような適当な値の幅をとってマッチングを行う。

5.6 Conclusions

省略。

コメント

因果効果の異質性に関して | 因果推論の枠組みにおいては、興味のある処置変数の効果は母集団（もしくは処置群・統制群）で一定であると仮定されている。しかしこの仮定は社会学の立場からはしばしば受け入れがたい。例えば Brand and Xie によれば、大学教育が賃金に与える効果には異質性があり、大学教育を受ける見込みがもっとも低い層においてその便益はもっとも高いという Negative selection が存在するという (Brand and Xie 2010)。因果推論は確かに強力なツールだが、これを社会学の関心に近い方法で活用するやり方についても考えながら、活用していければ良いと思う。

参考文献リスト

傾向スコア法について

- 林岳彦, 2012, 「統計的因果推論（傾向スコア）の勉強会資料をアップしてみた」Take a Risk: 林岳彦の研究メモ (2016年1月13日最終閲覧, <http://takehiko-i-hayashi.hatenablog.com/entry/20120427/1335475881>),
→因果推論がやろうとしていることをビジュアルに分かりやすく説明している。おすすめ。
- 星野崇宏・岡田謙介, 2006, 「傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について」『保健医療科学』55(3): 230-43.
→上で触れた林 (2012) の元ネタとなっている論文。
- Morgan, Stephen L. and David J. Harding, 2006, "Matching Estimators of Causal Effects," *Sociological Methods & Research*, 35(1): 3-60.
→本章とほぼ同じ内容が書かれている。

経験的な研究への適用例

- Morgan, Stephen L., 2001, “Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning,” *Sociology of Education*, 74(4): 341–74.
→本文の例でたびたび用いられている Morgan (2001) はこの論文を指す。
- Brand, Jennie E. and Yu Xie, 2010, “Who Benefits Most from College?: Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education,” *American Sociological Review*, 75(2): 273–302.
→社会学における因果推論を用いた経験的な研究としてはたぶんもっとも有名な論文。傾向スコアを用いた分析の例としても参考になるが、それ以上に、社会学にとって因果推論をどのように扱うべきかを考えるうえで興味深い。
- 中澤渉, 2013, 「通塾が進路選択に及ぼす因果効果の異質性——傾向スコア・マッチングの応用」『教育社会学研究』92: 151–74.
→日本語で読める文献だとこれが数少ない適用例のうちの1つ。

因果推論一般

- Gangl, Markus, 2010, “Causal Inference in Sociological Research,” *Annual Review of Sociology*, 36: 21–47.
→因果推論の基礎的な方法について一通り網羅している。社会学において因果推論はどのような意義を持つかが書かれている。
- 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。
→数式展開が難しくついていけなくなると評判の（個人の感想です）一冊。細かいことが知りたいときの参考書として使える。
- 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。 →未読。
- Pearl, Judea, 2010, “The Foundations of Causal Inference,” *Sociological Methodology*, 40(1):75–149.
→未読。

社会学において因果推論をどのように考えるか

- Elwert, Felix and Christopher Winship, 2014, “Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable,” *Annual Review of Sociology*, 40: 31–53.
→ Chapter 4 で出てきた Collider variable について、これを統制することがなぜ問題となるのか、具体的な事例をあげながら説明している論文。分かりやすい。
- 石田浩, 2012, 「社会科学における因果推論の可能性」『理論と方法』27(1): 1–18.
→数理社会学会の会長講演の書き起こし。
- Xie, Yu, 2007, “Otis Dudley Duncan’s Legacy: The Demographic Approach to Quantitative Reasoning in Social Science,” *Research in Social Stratification and Mobility*, 25(2):141 – 56.
→ Duncan のパス解析との関連で量的調査の分析から因果関係を捉えるアプローチについて考察を展開している。