

東京電機大学

博士論文

単純因果帰納モデル $pARIs$ の合理分析および因果探索アルゴリズムへの応用

Rational analysis of the elemental causal inductive model $pARIs$ and its application to causal discovery algorithms

2023年3月

東京電機大学大学院 先端科学研究科 情報学専攻

樋口滉規

学籍番号 20UDJ01

abstract

我々を取り巻く環境は不確実性に満ちている。我々は因果推論を利用することで、不確実な未来を予測可能にし（あるいは過去を推測可能にし）、時に未来を意図した方向に操作することができる。このように因果推論は人間の認知の本質的な構成要素の一つであるが、その認知過程には未解明な点が多い。人間の因果推論は時に誤りを含むが、一方でデータが少ない状況でも適切な因果関係を学習できるなどの長所がある。その学習過程をモデル化することは、人間を理解するための科学的価値だけでなく、有用なシステムの開発に繋がるという工学的価値もある。本研究では、人間の因果推論、すなわちある事象の原因を特定する過程は、観察に基づいて無数の事象から原因でありそうな事象を絞りこむ段階と、介入（原因でありそうな事象の生起を操作すること）によって真の原因を特定する段階からなると説明する二段階仮説の枠組みを採用した。我々は、この二段階のうちの前者の段階、人間が観察に基づいて因果関係がありそうな事象を絞り込むプロセスに焦点をあてる。この段階における人間の認知過程を説明したものを因果帰納モデルと呼び、因果帰納モデル *pARIs* についてメタ分析とコンピュータ・シミュレーションによって補完する形でこのモデルの合理分析を行った。

修正を加えたメタ分析において、数十種類の因果関係誘導モデルの中で、実験データとの適合度が最も高いのは *pARIs* であった。また、シミュレーションの結果、既存の一般的な因果帰納モデルと比べても、人間の因果帰納を記述するモデルとして優れた精度を持つことが分かった。シミュレーションの結果は、*pARIs* が、特に少数のサンプルからの推定において、他のモデルよりも高い精度で母集団の共変関係を推定すること、また、その性能の優位性は母集団の共変関係が小さい場合において顕著であることを示した。これらの結果は、人間の因果帰納の優れた記述モデルが適応的な合理性を有することに加えて、母集団推定を行うにあたってどのような性質を有するのかを明らかにした。

また、共変関係の検出であった因果帰納モデルの計算上の目標を、非独立性の検出であると仮定しなおして再度合理分析を行った。母集団の非独立性のパラメータを定義するために、確率論的な独立の定義を拡張し、現実環境における事象の生起はスパースであるという仮定を置き（なお、これを稀少性仮定と呼ぶ）、この仮定における（非）独立性の指標、

DR を導出した。その後、DR を母集団パラメータに設定し、再度コンピュータ・シミュレーションによる分析を行った。結果は、 $pARIs$ が稀少性仮定の下での非独立性の指標である DR と非常に似た挙動を示し、また、母集団パラメータの優れた推定性能を有することを示した。

最後に、「単一事象間の因果関係の強さの推定」という単純化された枠組みのためのモデルであった $pARIs$ を、より複雑で現実的な因果探索の枠組みへと拡張した。ベイジアンネットワークは複数事象間の因果関係を有向非巡回グラフ (DAG) と条件付き確率表で表すグラフィカルモデルであるが、観測データからグラフ構造 (=因果構造) を推定するために用いる手法には多大な計算コストがかかるという課題がある。我々は、代表的な構造推定アルゴリズムである PC アルゴリズムにおける、条件付き独立性の検定を $pARIs$ の拡張した条件付き $pARIs$ で代替した構造推定アルゴリズムの提案を行った。提案したアルゴリズムに対して、コンピュータ・シミュレーションによる推定精度の評価を行い、結果から、PC アルゴリズムの条件付き独立性の検定のための標準的な手法であるカイ二乗検定を使用した場合と比べて、提案アルゴリズムが、特に観測データが少ない場合において優れた性能を有することを示した。

目次

abstract	i
第1章 導入	1
1.1 理論的な枠組みの設定	2
1.2 合理分析	2
1.3 人間の因果推論における二段階仮説	3
1.4 単純因果帰納	4
1.5 単純因果帰納モデル	4
第2章 メタ分析	7
2.1 結果	10
2.2 結論	10
第3章 共変動検出としての因果帰納	11
3.1 環境のモデリング	12
3.1.1 稀少性仮定	13
3.1.2 等確率性仮定	14
3.1.3 ワーキングメモリ容量	14
3.2 Hattori & Oaksford (2007) が行ったシミュレーションの問題点	15
3.2.1 問題 1: 分割表に対する暗黙的な制約	15
3.2.2 問題 2: モデルの定義可能性	16
3.3 シミュレーション Aa	17
3.3.1 方法	17
3.3.2 結果と議論	19
3.4 シミュレーション Ab	19
3.4.1 方法	20
3.4.2 結果と議論	20

3.5	シミュレーション Ac	20
	3.5.1 方法	20
	3.5.2 結果と議論	20
3.6	シミュレーション Ba	23
	3.6.1 方法	23
	3.6.2 結果と議論	23
3.7	シミュレーション Bb	24
	3.7.1 方法	24
	3.7.2 結果と議論	24
3.8	シミュレーション C	25
	3.8.1 方法	26
	3.8.2 結果と議論	26
3.9	シミュレーション D	28
	3.9.1 方法	29
	3.9.2 結果と議論	31
3.10	総合議論	33
	3.10.1 $pARIs$ vs. DFH	33
	3.10.2 シミュレーション設定の修正が結果に及ぼす影響	34
3.11	結論	35
第 4 章	非独立性の検出としての因果帰納	37
4.1	観察的因果帰納モデルの計算上の目標	37
4.2	稀少性仮定下での非独立性の尺度	38
4.3	シミュレーション	39
	4.3.1 シミュレーション Aa'	39
	4.3.2 シミュレーション Ab'	40
	4.3.3 シミュレーション Ac'	40
	4.3.4 シミュレーション Bb'	40
	4.3.5 シミュレーション C'	42
	4.3.6 シミュレーション D'	43
4.4	総合議論	45
	4.4.1 モデル値のデフォルト	45
	4.4.2 ベイジアンネットワークへの応用	46
4.5	結論	47

第 5 章	因果探索アルゴリズムへの応用	49
5.1	ベイジアン・ネットワーク	49
5.2	条件付き独立性検定	49
5.3	PC アルゴリズム	50
5.4	シミュレーション	50
	5.4.1 方法	50
	5.4.2 結果と議論	52
5.5	結論	53
第 6 章	結論	55

図の目次

2.1	メタ分析に含まれた実験のカバーストーリーの因果構造	9
3.1	定義可能性に関するモデル間の包含関係を示すハッセ図	16
3.2	等確率性仮定の下での $pARIs$ および DFH のサンプル相関係数 ϕ との 決定係数	19
3.3	等確率性仮定の下での $pARIs$ および DFH の母集団相関係数 ϕ_0 との 決定係数	21
3.4	等確率性からの乖離度合いとモデルの母集団推定性能	22
3.5	各 $P(C) = P(E)$ における生成的な母集団からサンプリングされた抑制 的な分割表の割合.	23
3.6	各モデルが定義される割合の時間発展	25
3.7	モデルの平均値と標準偏差 (SD) の時間発展	27
3.8	ϕ_0 の関数としてのモデル平均値	27
3.9	各モデルの分類性能 ROC-AUC スコアおよび PR-AUC スコア	32
4.1	等確率性仮定の下での $pARIs$ および DFH のサンプル相関係数 DR と の決定係数	39
4.2	等確率性仮定の下での $pARIs$, DFH , DR の母集団相関係数 DR_0 との 決定係数	41
4.3	等確率性からの乖離度合いとモデルの母集団推定性能	42
4.4	モデルが定義される割合の時間発展	43
4.5	モデルの平均値と標準偏差 (SD) の時間発展	44
4.6	DR_0 の関数としてのモデル平均値	44
4.7	$P(C) = P(E) = .2$ としたときの, 各 ϕ_0 の閾値における RP-AUC	45
5.1	データ生成に使用した因果グラフ	52

-
- 5.2 各アルゴリズムを用いて学習したグラフと正答グラフの間のエッジ編集距離. χ^2 検定では有意水準 $\alpha = .01$ を, $pARIs$ および $pARIs_0$ の閾値は, それぞれ全体を通して安定していた $\epsilon = .1$ と $\epsilon_0 = .5$ を用いた. . . 54

表の目次

1.1	共起情報を表す 2×2 分割表	4
2.1	メタ分析の結果. 表は, 各実験で報告された結果と各モデル値の相関係数 r , および各モデル値と参加者の回答の平均値に対する母集団相関係数の推定値 $\hat{\rho}$ とその 95% 信頼区間を示す. $\hat{\rho}$ は, 固定効果モデルに基づき各実験で報告された結果から相関係数 r を統合したものである. M は各実験の刺激数を示す. 9 つの実験から合計 78 個の刺激がある.	10
3.1	シミュレーションの対応表. 本論文のシミュレーションと Hattori and Oaksford [2007] のシミュレーションとの対応, 各シミュレーションの目標, 3.2 で述べた問題との関連を示す.	11
3.2	すべての可能なサンプルに対するモデルの定義可能性. 上の 4 行の 0 と + は, それぞれセルの値 (自然数) が 0 または 1 以上であることを示す. 下の 4 行における, — と D は, それぞれモデルが定義不可能, 定義可能であることを意味する.	16
3.3	二値分類における 2×2 混同行列	29

第1章

導入

人類は、その種が誕生して以来、厳しい自然環境の中で適応・生存することに成功し、現在の繁栄を迎えている。そのような成功の要因の一つは、人間が因果関係を効率的に学習（すなわち、帰納的に構築）し、それを活用する能力にある。学習した因果関係を活用することにより、原因の観察に基づいて未来を予測することや、結果から原因を診断すること、原因への介入により未来を操作することが可能となる。より具体的には、食物の安全性の判断、効率的な植物の栽培や狩りの方法の学習、食物の長期保存方法の発見、病気の予防・診断・治療、天候や災害の予測など、生存に不可欠であった数多くの要素は因果関係の学習に基づいている。このような因果関係を利用した生存戦略は、しばしば迷信、頑迷、陰謀論のような誤りを含むものの、データが少ない状況でも適切な因果関係を学習できるという利点がある。この点において人間の因果推論は現状の人工知能と比べても優れた性質を有している。したがって、その認知過程をモデリングを介して明らかにすることは、人間を理解するという科学的な価値だけではなく、有用なシステムのデザインに繋がる工学的な価値を有する。

本研究において、人間の因果推論を取り扱うための理論的な枠組みとして、人間の因果帰納における二段階仮説 [Hattori and Oaksford, 2007] を採用する。これは人間の因果帰納を、「共変動検出による原因候補の絞り込み」を行うヒューリスティック段階と「介入操作による原因候補の精密な検証」を行う介入段階に分解して説明するものである。前者の機能に注目したとき、それは人間が有する因果的直感に対応していると見なすことができる。Hattori and Oaksford [2007] は、人間の因果帰納におけるヒューリスティック段階のモデルとして *DFH* (Dual-Factor Heuristics) を提案し、*DFH* が実際に人間の因果帰納をよく記述することをメタ分析によって示した。また、彼らは、複数のシミュレーションを介して合理分析 [Anderson, 1990] を行い、実環境を反映した幾つかの制約を仮定することで、彼らのモデルが従来の代表的な因果帰納モデルよりも優れた性能を持つことを示した。

本章では、本研究の理論的枠組みの導入と、因果帰納モデルの紹介を行う。また、第2章において Hattori and Oaksford [2007] のメタ解析を、第3章において彼らのシミュレーションを拡張し、彼らの研究の問題点を可能な限り修正する。また、その結果、因果帰納モデル $pARIs$ が最も単純なモデルの一つでありながら、 DFH よりも二段階仮説における最初の段階（観察段階）に適したモデルであることを示す。続いて、第4章において稀少性下での独立性の尺度を定義した上でそれを母集団パラメータに設定し再度シミュレーションを行うことで $pARIs$ が稀少性仮定下の非独立性の単純かつ有効な近似であることを示す。この結果を踏まえ、第5章において、ベイジアンネットワークの枠組みにおいて $pARIs$ を応用した因果探索アルゴリズムを開発して性能の検証を行う。

1.1 理論的な枠組みの設定

1.2 合理分析

Anderson [1990] は、認知科学研究の方法論として合理分析を開発した。この方法論は、人間を始めとする「行動主体の認知システムは、進化の過程において、その主体の行動を環境に対して最適化する」という前提に基づいて認知モデルを組み立てるものである。彼は、我々人類を含む生物の認知メカニズムは進化の過程を経て、ある程度環境に対して最適化された構造を持っていると主張し、これを合理性の原理と呼んだ。^{*1} 本研究は、我々の因果帰納の仕組みもまた、我々を取り巻く環境に対して最適化されているという観点から人間の因果帰納モデルを説明するものである。したがって、合理分析において人間の因果帰納過程を説明するモデルは、それが人間の因果帰納傾向を記述し、それが現実環境に対して最適化されたものであるかということによって評価される。Anderson は、以下の6つのステップから、合理性の枠組みに基づいて理論を構築可能だと主張した。[Anderson, 1990, p.29]:

1. 認知システムの目標を正確に特定する。
2. システムが適応される環境の形式的なモデルを開発する。
3. 計算論的な制限について最小限の仮定をする。
4. 上記 1.-3. から最適な動作関数を導出する。
5. 行動関数の予測を確かめるために実証的な文献を調べる。
6. 予測が外れた場合は、上記の手順を反復する。

^{*1} なお、ここでの合理性は、認知システムが論理的に正しい推論を行うという意味でない。彼は、人間の行動は人間の目標を達成するために最適であるとする経済学の立場に鑑み、経済学の用語になぞらえて「合理性」という言葉を用いた。

ステップ2について、原理的には、環境に関する仮定が真であるかどうかは検証可能である。他方で認知システムの構造は観測できないため、ステップ3では、ほとんどすべての情報処理理論が同意するような主張（例えば、ある種の短期記憶の限界）を除き、最小限に抑えなければならない。なお、この理論は、「認知システムのメカニズム」ではなく、「メカニズムがなぜそのように計算するのか」を説明するものである。

1.3 人間の因果推論における二段階仮説

Hattori and Oaksford [2007] は、注目する結果事象に対応する原因を見つけ出すという状況において、人間の因果推論のプロセスが2つの段階からなるという二段階仮説を提唱した。彼らは、2つの段階をそれぞれ、「ヒューリスティック段階」と「介入段階」と名付けた。前者は、事象の生起に関する観察結果だけに基づいて、無数にある事象の中から注目している事象の原因でありそうな事象を絞り込む段階であり、後者は先の段階によって抽出された原因候補事象が真に原因であるかを介入操作（事象の生起を操作すること；実験）によって検証する段階である。人間が因果推論を行う現実世界において、情報処理能力、時空間的な観測範囲、ワーキングメモリの制限などが課されているのにも関わらず、考慮しなければならない事象の数は無数に存在する。それらすべての事象に対して、介入操作に基づき検証していくことは不可能であり、したがって、まずは因果関係がありそうな少数の候補事象を絞り込む必要性があるとする仮定は、我々が直面する実社会での課題を考えると妥当であると思われる。ヒューリスティック段階における目標とは、高コストだが正確な検証の段階である介入段階において取り扱うに値する原因候補事象を前もって絞り込むことである。介入操作には高いコストがかかるため、事象の絞り込みは低コストで収集することができる観測データのみを基づいて行われることが好ましい。そのような要求を満たす、絞り込みの基準となる情報の一つに共変関係が考えられる。当然、2つの事象が共変関係にあるということは、それらの間に因果関係があるということに直接意味しない。共変関係にあるが因果関係はないような例を挙げることは容易であり、例えば疑似相関のケースがそれにあたる。しかしながら、共変関係は因果関係の必要条件であるために、因果関係の蓋然性を判断するための強力な手がかりとなる。このような理由から、Hattori and Oaksford [2007] は共変動の検出によって事象の絞り込みが行われると想定している。本研究において、人間の因果推論の二段階理論における第1段階目（ヒューリスティック段階）における因果帰納モデルの計算目標（合理分析のステップ1, 1.2 参照）を、共変動の検出に基づき原因候補事象を絞り込むことであると仮定している。

表 1.1 共起情報を表す 2×2 分割表

	Effect (E)	No effect ($\neg E$)
Cause (C)	$a = N(C, E)$	$b = N(C, \neg E)$
No cause ($\neg C$)	$c = N(\neg C, E)$	$d = N(\neg C, \neg E)$

1.4 単純因果帰納

人間の因果推論に関する多くの研究が、単純因果帰納 [Griffiths and Tenenbaum, 2005] の問題に対処してきた。ここで、注目している事象を結果事象 E 、その原因であると考えられる事象を原因候補事象 C とする。単純因果帰納とは、事象 C と E の結合事象の生起頻度から、 C が E の原因である蓋然性の高さを推論する枠組みである。具体的には、事象 C と E の生起・不生起の場合を考え、各組み合わせの生起頻度を $N(C, E)$, $N(\neg C, E)$, $N(C, \neg E)$, $N(\neg C, \neg E)$ と表し、それぞれを便宜的に a, b, c, d と呼ぶ。このとき、 C と E の共起情報は表 1.1 に示すような 2×2 の分割表として表すことができる。一つの原因と一つの結果だけに注目するために他の事象を含めた因果的な構造を考慮せず、また因果関係の要件である原因の時間的な先行性を考慮しないという点で、これは単純化された枠組みである。

現在まで、人間の因果帰納を説明するために多くのモデルが提案されている。因果帰納モデルはこの分割表を用いて C と E の間の因果強度を推定する。(言い換えれば、因果帰納モデルは4つの非負整数 a, b, c, d から事象 C, E 間の因果強度への関数と見なすことができる) 次節では、本研究で使用したいくつかのモデルを概説する。

1.5 単純因果帰納モデル

2×2 の分割表における規範的な共変動のモデルが四分点相関係数であり、心理学の分野ではしばしば ϕ 係数と呼ばれる。これは統計学において2変数間の共変動の尺度であるピアソンの積率相関係数を単純因果帰納の枠組みへと特殊化したモデルである。

$$\phi = \frac{P(C, E)P(\neg C, \neg E) - P(C, \neg E)P(\neg C, E)}{\sqrt{P(C)P(\neg C)P(E)P(\neg E)}} \quad (1.1)$$

$$= \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (1.2)$$

単純因果帰納における標準的なモデルが ΔP である [Jenkins and Ward, 1965] .

$$\Delta P = P(E|C) - P(E|\neg C) \quad (1.3)$$

$$= \frac{a}{a+b} - \frac{c}{c+d} \quad (1.4)$$

$$= \frac{ad - bc}{(a+b)(c+d)} \quad (1.5)$$

このモデルは原因候補事象が不在から存在へ切り替わることで結果事象の生起確率がどれだけ上昇するか、ということの意味している。式の第一項と第二項はそれぞれ、科学研究で用いられる実験法における実験群と統制群に対応しており、このモデルは科学実験の基本的な考え方をモデル化したものと考えることができる。厳密には、原因 x の生起を所与として結果 y が生起する条件付き確率 $P(y|x)$ と、 x に関して介入した場合の y の生起確率 $P(y|do(x))$ は一致しない。これらは対象のデータがランダム化実験により得られた場合には一致するが、現実世界において我々が対処するところの観測データがそのように緻密に計画された実験から得られないことは明らかである。また、モデルが持つ合理性の割に、このモデルは因果帰納実験における人間の反応にうまく適合しないことが報告されている [Buehner et al., 2003, Hattori and Oaksford, 2007].

$$DFH = \sqrt{P(E|C)P(C|E)} \quad (1.6)$$

$$= \frac{P(C, E)}{\sqrt{P(C)P(E)}} \quad (1.7)$$

$$= \frac{a}{\sqrt{(a+b)(a+c)}} \quad (1.8)$$

原因から結果への予測確率と結果から原因への診断確率の幾何平均として導出されるモデルが二要因ヒューリスティックモデル (DFH : dual-factor heuristics) である。モデルの定義式は、予測確率 $P(E|C)$ と診断確率 $P(C|E)$ が両方とも高い場合に DFH は高い値を取ることを示している。従って、このモデルは、2つの確率がともに高い場合に、人間は事象間に強い関係を認識するという主張を表現したものであるといえる。このような高い予測確率の要求は、原因の生起は結果の生起確率を高めるという事実から説明できる。他方で、高い診断確率の要求は、結果の生起には原因が不可欠である（いわば「火のないところに煙は立たない」という考えを表現したものである。Hattori and Oaksford [2007] はメタ分析を行い、他の 41 のモデル（33 個のノンパラメトリックモデルと 8 個のパラメトリックモデルからなる）との比較において、 DFH が人間の因果帰納と最も高い相関を持つことを示した。彼らは、このモデルが合理的であることの根拠として、 DFH モデルが共変関係の規範的モデルである ϕ から、 $d = N(\neg C, \neg E)$ を極端な稀少性の仮定 [Oaksford and Chater, 1994] に基づいて無限大に発散させることによって導出されると

主張した [Hattori and Oaksford, 2007]. 彼らが極稀少性仮定と呼ぶこの極限操作は、結果としてモデルの定義から巨大あるいは曖昧な（後述する）の d セルを除外する。

稀少と仮定される事例の比率 ($pARIs$: proportion of Assumed to be Rare instances) は、 C と E の生起が共に稀少であるという仮定のもとで、「 C または E の生起」を所与として「 C と E が同時に生起する」条件付き確率である [Takahashi et al., 2010].

$$pARIs = P(C \wedge E | C \vee E) \quad (1.9)$$

$$= \frac{P(C, E)}{P(C \vee E)} \quad (1.10)$$

$$= \frac{P(C, E)}{P(C) + P(E) - P(C, E)} \quad (1.11)$$

$$= \frac{a}{a + b + c} \quad (1.12)$$

これは DFH と同様に、膨大であるか、あるいは曖昧な d セルを式中に含まず、したがってその影響を受けない。 $pARIs$ は変形すると以下のように表すことができる。

$$pARIs = P((E|C) \wedge (C|E)) \quad (1.13)$$

この式は、 DFH と同様に $pARIs$ も、「 C ならば E , かつ、 E ならば C 」という因果的な双条件文と密接な関係があることを示している [Kaufmann, 2009, Sanfilippo et al., 2018]. ただし、2 つモデルの双条件文との関係性は、 DFH は両条件の確率の幾何平均であるのに対し、 $pARIs$ は両条件文の連言が真である確率である点において異なる。

第 2 章

メタ分析

Hattori and Oaksford [2007] のメタ分析を再現し、前章で示したモデルの、実験データ、すなわち人間の因果帰納傾向への適合度を比較した。

彼らは過去の文献から収集した実験データを統合して再分析することで、 2×2 分割表ベースの記述モデルのうち、どれが人間の因果帰納に関して最も適切な記述モデルであるかを調査した。結果として、彼らは、42 種類のモデル（うちノンパラメトリックモデル 34 種類、パラメトリックモデル 8 種類）の中で、*DFH* が最もデータに適合することを示した。

メタ分析の対象となった実験は全て共変動する 2 つの事象の間の因果関係、相関関係、あるいは偶然性に関するものであり、以下の 5 つの基準に従って文献から選択された。

1. 参加者が人間であること。
2. 実験における因果関係が一对一の対応関係であると仮定されていること。
3. 関係の主観的な強さが（ほぼ）連続値で評価されていること。
4. 実験者があらかじめ決めた頻度でサンプルが順次与えられていること。
5. 原因が生成的（原因が効果を発生させる）であり、抑制的（原因が効果の発生を防ぐ）ではないこと。

彼らは、得られたサンプルにおいて原因が生成的であることを式 2.1 により定義した。

$$a \cdot d \geq b \cdot c \quad (2.1)$$

これは $\phi \geq 0$ および $\Delta P \geq 0$ と等価である。また、原因が抑制的であることを式 2.2 のように定義する。

$$a \cdot d < b \cdot c \quad (2.2)$$

上記の基準に基づいて彼らを選出した実験は次の 8 つであった。Anderson and Sheu [1995] の実験 1 (AS95), Buehner et al. [2003] の実験 1 と 3 (BCC03.1, BCC03.3),

Hattori and Oaksford [2007] の実験 1 と 2 (HO07.1, HO07.2), Lober and Shanks [2000] の実験 1 から 3 (LS00.1–3), White [2003] の実験 2 と 6 (W03.2, W03.6).

彼らは、各モデルと、参加者が推定した因果関係の強さの平均値との決定係数 (r^2) を算出し、各モデルのデータへの適合性を測った。なお、Hattori and Oaksford [2007] では、BCC03.1 に含まれる全 15 種類の刺激のうち、13 種類のみを扱ったようであり、 $P(E|C) = P(E|\neg C) = 0$ と $P(E|C) = P(E|\neg C) = 1$ をそれぞれ満たす 2 つの刺激を分析から除外することで、我々は元の結果をほぼ再現した。彼らが 2 つの刺激を元の解析から除外した理由は、前者については *DFH* と *Power PC* [Cheng, 1997] は分母が 0 であることによりモデル値を計算できないため (表 3.2 のケース 3)、後者については刺激が厳密に生成原因でないためと推測される。^{*1}

メタ分析の結果を解釈する上で、8 つの実験のうち、AS95 はいくつかの点で他の実験と異なっていることに注意する必要がある。第一に、刺激の数が非常に多い。メタ分析で扱った全 143 種類の刺激の半分以上に相当する 80 種類がこの実験に含まれており、全体の結果を計算する上で、この AS95 の比重が非常に大きくなっている。この実験は 40 人の参加者がそれぞれ 80 種類の刺激に対して 1 時間以上かけて取り組むものであり、他の実験よりも時間が長かった。この実験では、参加者は過去の刺激を忘れて新しい刺激に反応すべきであるのに、実験の後半になるほど古い刺激と新しい刺激を混同してしまう可能性がある。この効果は他の実験でも起こっていたかもしれないが、刺激数が圧倒的に多かった AS95 で最も顕著に現れたと思われる。

AS95 をメタアナリシスに含めない第二の理由に、AS95 で用いられたカバーストーリー構造は、他の実験と比較して特殊であった。具体的には、各実験のカバーストーリーはそれぞれ以下のような内容になっている。まず、HO07.1 は肥料の有無が植物の開花に与える影響を、HO07.2 は牛乳の飲用が腹痛の生起に与える影響を問うものであった。BCC03.1 の実験は抑制的刺激 (原因が結果の生起を抑制する場合、式 2.2)、と生成的刺激 (原因が結果の生起を促進する場合、式 2.1) の 2 つからなるが、メタ分析に含まれた後者は、あるウイルスに対する光線の曝露が突然変異に与える影響を問うものであった。また、BCC03.3 はアレルギー新薬が頭痛の有無に与える (抑制・生成の両方を含む) 影響を、LS00.1–3 は、それぞれ、三種類の動物について、ある化学物質の投与が遺伝子変異に与える影響を、W03.2 および W03.6 は、ある化学物質の摂取がアレルギー反応の生起に与える影響を問うものであった。これを総合すると、いずれも、図 2(a) のような構造になっており、いずれも原因候補と結果がそれぞれ 1 つずつ与えられた上で原因候補から結果への因果関係の強度を問うものである。一方で、AS95 のカバーストーリーは以

^{*1} 彼らのメタ分析には小さな誤りがあった。LS00.1–3 の実験 2 における刺激 $(a, b, c, d) = (28, 0, 21, 7)$ に対する参加者の反応の平均値が 34 とされていたが、実際には 65 であった。しかし、この修正は結果にはあまり影響せず、したがって彼らの議論に対しても特に影響しない。

下のような内容となっている。参加者は、架空の疾病 (C') の治療のための新薬 (C) の臨床試験を行うことを要求された。その新薬は、副作用 (E) を引き起こすことが知られているが、それは新薬が治療対象としている疾病の症状でもある。そこで、新薬の投与がこれらの副作用/症状の発生にどの程度影響を与えるかを参加者に尋ねた。このカバーストーリーの構造を図 2(b) に示した。後述するように、結果事象 E に対して二つの原因事象 (C, C') が存在するが、片方は生成的であり、もう片方は抑制的である。

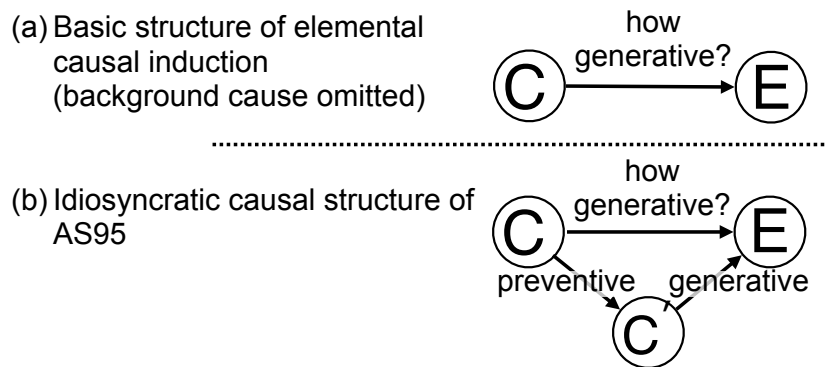


図 2.1 メタ分析に含まれた実験のカバーストーリーの因果構造

8つの実験の中で、着目すべき原因の候補とは別に、結果を引き起こす確率の高い代替的な原因の存在に明示的に言及しているものは AS95 だけである。このことは、AS95 だけが、原因と結果の間の因果の強さを推論する際に、実験参加者に対して注目する原因候補とは別の原因事象の存在を考慮することを要求したことを意味している。更に、単に原因候補以外の原因事象が存在するというだけでなく、それらの原因の間に（抑制的な）因果関係の存在が仮定されたことで、「 C は E と $\neg E$ を同時に引き起こす」という、ある種のアンチノミーのような構造となっている。つまり、「 C (薬の投与) が E (副作用) を引き起こす」という推論と、「 C (薬の投与) が $\neg E$ (病気の治癒) を引き起こし、結果として $\neg E$ (症状) を抑制する」という一見矛盾する推論が同時に成り立つのである。以上の3点から、AS95 はメタ分析に含まれる他の実験よりも複雑な推論を実験参加者に対して要求していることがわかる。最後に、2章で述べたように Hattori and Oaksford [2007] はメタ分析の対象となる実験を選択する基準を5つ設定しているが、AS95 はそのうちの1つに抵触している可能性がある。問題となる基準は2番目の「実験における因果関係は一対一対応でなければならない」であるが、前述のように、AS95 は結果事象に対して複数の原因事象が存在する可能性を明示的に提示している。以上の理由より、AS95 を他の実験と同様にメタ分析に含めることの妥当性には疑問が残るため、我々は AS95 をメタ分析の対象から除外した。

また、除外された実験を補い、分析の質を高めるために、上記の5つの条件を満たす2つの実験を追加した。追加した実験は CS06[Collins and Shanks, 2006] と PS03[Perales

and Shanks, 2003] であり, 使用したデータは Perales and Shanks [2007] の付録に記載されていたものを使用した.

2.1 結果

我々は, Hattori and Oaksford [2007] の結果を完全に再現し, *DFH* が *pARIs* を除く 34 種類のノンパラメトリックモデルの中で最も実験データに適合することを確認できたため, 上記の修正を施して再度分析を実行し, *DFH* と *pARIs* を中心に4つのモデルについて比較した. 固定効果モデルに基づき母集団の相関係数 (すべての実験に共通すると仮定される母集団パラメータ) の点推定を $\hat{\rho}$, 95% 信頼区間の限界値 ρ_L , ρ_U とし, 有効数字は3桁とした. 結果を表 2.1 に示した. *pARIs* は, *DFH* と比較して, より単純な形であるのにも関わらず, 実験データに対して同等以上に適合することを示している. *pARIs* と *DFH* は *W03.2* において共に最もデータとの適合が小さくなり, このとき *pARIs* が $r = .884$ を取るのに対して *DFH* は $r = .829$ を取る.

表 2.1 メタ分析の結果. 表は, 各実験で報告された結果と各モデル値の相関係数 r , および各モデル値と参加者の回答の平均値に対する母集団相関係数の推定値 $\hat{\rho}$ とその 95% 信頼区間を示す. $\hat{\rho}$ は, 固定効果モデルに基づき各実験で報告された結果から相関係数 r を統合したものである. M は各実験の刺激数を示す. 9つの実験から合計 78 個の刺激がある.

	BCC03.1	BCC03.3	CS06	H07.1	H07.2	LS00.1-3	PS03.1-3	W03.2	W03.6	$\hat{\rho}$	ρ_L	ρ_U
<i>pARIs</i>	0.980	0.969	0.918	0.965	0.993	0.892	0.939	0.884	0.937	0.960	0.932	0.977
<i>DFH</i>	0.974	0.955	0.926	0.964	0.982	0.892	0.963	0.829	0.892	0.952	0.919	0.972
ϕ	0.909	0.837	0.768	0.708	0.703	0.844	0.856	0.455	N/A	0.791	0.664	0.874
ΔP	0.919	0.839	0.842	0.707	0.010	0.875	0.810	0.070	N/A	0.746	0.598	0.845
M	13	6	7	12	9	11	8	8	4	(78)		

2.2 結論

メタ分析の結果は, *pARIs* が *DFH* よりも高い記述性能を有することを示した. この結果は, *pARIs* も *DFH* と同様に Hattori and Oaksford [2007] で比較した他の 33 のノンパラメトリックモデルと 8 のパラメトリックモデルよりも人間の因果帰納の優れた近似であることを意味する. 通常, 作り込んだモデルはデータへの適合度を上げることができが, 非常に単純な形で定義される *pARIs* がこのような高い適合を示すことは注目に値する. 人間が *pARIs* に似た因果的直観を持つことの適応的な利点を示すために, 次章では計算機シミュレーションに基づいて合理性の分析を行う.

第 3 章

共変動検出としての因果帰納

本章では、まず、合理分析の方法論に従い、シミュレーションに用いるパラメータやその他の設定に対して制約を設ける。その後、Hattori and Oaksford [2007] が行ったシミュレーションの設定に問題があることを指摘した。これらの問題点を可能な限り修正し、比較対象に $pARIs$ を追加した上で、再度シミュレーションを実施した。得られた結果から、 $pARIs$ は、 DFH よりも、Hattori and Oaksford [2007] が提唱する二段階仮説において、より適切なモデルであることが示された。また、これらのシミュレーション設定の変更が、元の研究の結果や考察に与える影響についても検証した。本研究と元の研究とのシミュレーションの対応関係を表 3.1 に示す。

表 3.1 シミュレーションの対応表。本論文のシミュレーションと Hattori and Oaksford [2007] のシミュレーションとの対応、各シミュレーションの目標、3.2 で述べた問題との関連を示す。

This study	HO07	Target	Problem
Sim Aa, 3.3 節 (Figure 3.2)	Sim 1 (9.2.1)	ϕ とモデルの関係	1 を修正
Sim Ab, 3.4 節 (Figure 3.3)	Sim 1 (9.2.2)	母集団 ϕ (ϕ_0) とモデルの関係	1 を修正
Sim Ac, 3.5 節 (Figure 3.4)	Sim 1 (9.2.2)	等確率性仮定下の Sim Ab	—
Sim Ba, 3.6 節 (Figure 3.5)	—	分割表に対する暗黙的な制約	1 と関連
Sim Bb, 3.7 節 (Figure 3.6)	—	モデルの定義可能性	2 と関連
Sim C, 3.8 節 (Figure 3.7)	Sim 2 (10)	モデル値の時間発展	—
Sim D, 3.9 節 (Figure 3.9)	—	モデルによる二値分類	1, 2 を修正

Hattori and Oaksford [2007] はワーキングメモリの制約、稀少性の仮定および等確率性 (3.1 節で説明する) を仮定した上で、以下のシミュレーションを行い、また、それぞれの結果から次のように主張した。まず、シミュレーション 1 によって DFH が共起の規範的モデルである ϕ と同様に振る舞うこと、また、同等の母集団パラメータ推定性能を有することを示した。その上で、 DFH が ϕ よりも計算コストの小さい (すなわち経済的

である) ため、好ましいと主張した。また、 N_W サンプルングを考慮した場合に、 DFH は同等のサンプルサイズから算出された ϕ よりも遥かに高い母集団パラメータ推定性能を示すことを明らかにした。次に、シミュレーション 2 によってモデルの平均値と標準偏差の推移を示した上で、モデルの収束の早さだけでなく、標準偏差に基づきモデルの信頼性を考慮することで、 N_W ベースの DFH が最も信頼性の高いモデルであると主張した。

我々は、彼らのシミュレーションを次に示す 4 つのグループからなる計 7 つのシミュレーションへと追加・再構成する。シミュレーション A と C では、Hattori and Oaksford [2007] で行われたシミュレーション 1 と 2 を、比較対象に $pARIs$ を追加した上で再実行する。具体的には、シミュレーション Aa では、 $pARIs$ と DFH が規範的モデルである ϕ とどの程度同様に振る舞うかをテストし、シミュレーション Ab では、 $pARIs$ 、 DFH 、 ϕ が母集団の相関係数 ϕ_0 の値をどの程度良く推測するかをテストし、また、シミュレーション Ac では、 $pARIs$ と DFH の母集団推定における等確率性の効果をテストする。ここで、母集団と標本を区別するため、母集団パラメータとしての相関係数を ϕ_0 と示し、推定モデルとしての相関係数を ϕ と示す。シミュレーション B では、シミュレーション A の解釈にも結果を与える 2 つの問題点 (3.2 節で述べる) の重要性を定量的に示す。シミュレーション C では、連続的なサンプルングにおいて、 $pARIs$ 、 DFH 、 ϕ 、 ΔP の平均値と標準偏差の推移を示し、推論過程におけるモデルの収束の速さと信頼性をテストする。最後にシミュレーション D では、より実践的な枠組みへの応用を考慮し、モデルを因果関係があるか否かを判断する二値分類器とみなし、その分類性能を評価する。これは、モデル値の収束速度とその信頼性という、重要だが同時に検証することが難しい 2 つの要素を、分類性能という 1 つの要素に統合して評価するものである。シミュレーション Ab, Ac, D はパラメータ・リカバリーの手法に基づくものである。これらのシミュレーションでは、真値 (パラメータ) を設定したのちに、そこからサンプルデータを作成する。その後で、サンプルデータのみに基づいてモデル値 (パラメータの推定値) を計算し、真値とモデル値の適合を確認することでモデルの推定精度を検証する。

3.1 環境のモデリング

1.2 節で述べたように、合理分析の枠組みを採用するにあたって、認知システムが適応する環境のモデリング、また計算的制約に関する最小限の仮定を行う必要がある。Hattori and Oaksford [2007] は合理分析を行う上で、稀少性、等確率性、ワーキングメモリの制約という 3 つの仮定を設定した。我々の合理分析は彼らの分析を踏襲し同様の仮定を設けるため、この項ではこれら 3 つの要因を順に説明し、若干の考察を加える。

3.1.1 稀少性仮定

稀少性仮定 [Oaksford and Chater, 1994] とは、通常、原因と結果の基準率 $P(C)$, $P(E)$ が小さい (すなわち稀) であるとする仮定である。自然のカテゴリが疎であることは直感的に明らかであり、例えばあらゆる事物の中で鳥類に属するものは稀少であり、更に鳥類の中においてカラスは稀少である。稀少性を仮定する多くの分析は、仮説空間が自然のカテゴリに大きく依存しているという経験的な証拠に基づいて稀少性の仮定を正当化している。また、Navarro and Perfors [2011] は、事象のスパース性を仮定することの理論的な根拠は家族的類似性に基づいて説明できることを示した。

単純因果帰納において、稀少性の仮定は d セルの解釈に対して主に 2 つの問題をもたらす。一つは d セルの膨大さ、もう一つは d セルの曖昧さである。前者について、稀少性を仮定することは a , b , c セルに比べて d セルが膨大にあると仮定することに等しい。これは、原因 C と結果 E が共に稀少であれば、殆どの場合、その両方ともが生起することは極めて稀少であるという事実から説明できる。また、 d セルが曖昧であるということの説明として、ある照明器具について「スイッチを押す (C) と明かりが点く (E)」という因果条件文を考える。この例では、スイッチを押して明かりが点いた回数 $a = N(C, E)$ を数えることには何ら問題がなく、その回数は一意に定まる。また、 $b = N(\neg C, E)$ や $c = N(C, \neg E)$ を数える場合も同様である。一方で、スイッチを押さず、明かりが点かなかった場合の生起回数 $d = N(\neg C, \neg E)$ はどのように数えれば良いだろうか？前三者の場合では、 C か E のどちらかが生起した瞬間に両事象の組み合わせを考えれば十分だったが、最後の場合は、その数え方は恣意的なものとなる。一つの方法として考えられるのは、単位時間ごとに事象の生起不生起を観測しカウントすることである。しかし、その単位時間の設定は必然的に恣意的であり、そのようにして得られた回数も曖昧さを残すことになる。

DFH および $pARIs$ はともに定義式に d セルを含まないため、これらのモデルは膨大または曖昧な d セルの値を無視することができる。実際に、Hattori and Oaksford [2007] が DFH の導出に際して用いた極稀少性仮定、すなわち極限操作 $d = N(\neg C, \neg E) \rightarrow \infty$ を各モデルに対して行ったとき、 DFH 、 $pARIs$ は共にその影響を受けない。また、 DFH の導出方法から、定義上 ϕ は DFH と等しくなることがわかる。一方で ΔP はその第二項 $P(E|\neg C) = c/(c+d)$ が 0 に発散し、モデルはより意味の希薄な $P(E|C)$ になる。Hattori and Oaksford [2007] のメタ分析によると、人間の因果推論実験に対する $P(E|C)$ の記述力は他のモデルに比べても極めて低く、このことは、稀少性仮定の下で ΔP の記述的有効性が大きく損なわれることを示している。

3.1.2 等確率性仮定

過去の研究により、「もし C なら、 E である」が真であるとき、人間はその逆「もし E なら、 C 」も真であると仮定する傾向があることが示されている [Hattori and Nishida, 2009]. [Hattori and Oaksford, 2007] はこの傾向の確率的拡張を「等確率性仮定」と呼んだ。彼らの行ったシミュレーションにおいてこの仮定は、合理的な分析のために、 C と E の発生確率がほぼ等しいという母集団への制約として環境に組み込んでいる。しかし、この仮定が表現しているのは人間の認知が有する特性であるため、合理分析に際して環境のモデリングの一部とみなすのは不適切であるように思われる。人間の認知過程における事象の生起確率のデフォルト値と、実環境における事象の生起確率は異なる問題であると考える方が自然であり、彼らの議論からは、等確率性仮定が環境の特性として母集団パラメータに反映されるまでの論理の説明が欠けているように見える。しかし、確かに、原因事象の稀少性を仮定すれば、等確率性の仮定はある程度成り立つ。例えば、 $P(C), P(E)$ が共に .2 以下であると仮定した場合、両確率の差は少なくとも .2 より小さいためにある程度の等確率性が成り立つ。

3.1.3 ワーキングメモリ容量

Hattori and Oaksford [2007] において、ワーキングメモリ容量の制約はサンプルサイズ（推論に用いるインスタンスの合計）に対する制限として組み込まれた。その際、彼らはワーキングメモリ容量の制約を、人間が短期記憶に保持できる情報の数として知られるマジカルナンバー 7 ± 2 [Miller, 1956] に基づいて論じたが、より最近の研究ではマジカルナンバー 4 ± 1 が人間の認知にとってより妥当な数字であることが明らかになっている [Cowan, 2001]. Hattori and Oaksford [2007] において、制約は事象の合計頻度に対して課せられたが、これは観測した事象の列（例えば、 (a, b, c, a, \dots) ）に含まれる事象のそれぞれが 1 つのチャンクに相当するという見方に基づくものである。他方、各事象の生起頻度を一つのチャンクとする見方もありうる。その場合には、制約は同時事象の種類数 (2×2 分割表では a, b, c, d の 4 種類) に対して課せられることになる。いずれにしても、サンプルサイズに制約を課すことは、ワーキングメモリーの容量の限界としてだけでなく、データ収集の時空間的な制約という観点からも現実的な条件である。サンプルサイズは、Hattori and Oaksford [2007] と同様に、ふたつの方法でコントロールした。ひとつは、 N -サンプリングと呼ばれるサンプリング方法で、これはサンプルサイズを全ての事象の生起頻度の和 ($N = a + b + c + d$) とするものであり、もうひとつは N_W -サンプリングで、これはサンプルサイズを d セルを除く事象の生起頻度の和 ($N_W = a + b + c$)

とするものである。人間の小さなワーキングメモリ容量に対して d セルの値は膨大であることが想定されるので、 d セルが貴重な容量を圧迫することも懸念される。定義に d セルを含まない DFH や $pARIs$ は、膨大あるいは不明瞭である d セルを無視することで小さなワーキングメモリ容量を有効に活用することができる。 d セルを観察あるいは記憶せず、推論にも用いないことを前提とした N_W -サンプリングを想定することにより、モデルが使用するサンプルサイズあたりの推定性能を評価することができる。

3.2 Hattori & Oaksford (2007) が行ったシミュレーションの問題点

我々は、Hattori and Oaksford [2007] で行われたシミュレーションには大きく 2 つの問題点があることを指摘する。その後、修正可能な問題点については修正を行い、シミュレーションの比較対象に $pARIs$ を追加した上で改めてシミュレーションを行う。この新しいシミュレーションに基づき、 $pARIs$ の結果を DFH や他のモデルの結果と比較することで合理的な分析を行い、また、シミュレーション設定の変更が原論文の議論に与える影響を検証した。

3.2.1 問題 1: 分割表に対する暗黙的な制約

元のシミュレーションにおける第一の問題は、シミュレーション A において、彼らが使用するサンプルを暗黙下に生成的なものに限定していることである。シミュレーションを通して、事象の生起確率の分布は、その理論とモデルの範囲のために常に生成的 (式 2.1) なものに固定されている。しかし、仮に確率分布が生成的であったとしても、そこから生み出される分割表が必ずしも生成的であるわけではない。生成的な母集団から、偶然抑制的な分割表が得られることもありうる。使用するサンプルを恣意的に選択してしまうことは仮定した母集団に歪みを生じさせる可能性がある。ところが元のシミュレーションにおいて、使用するサンプルから抑制的 (式 2.2) なものが取り除かれている。我々は可能なシミュレーション設定を網羅的にテストすることによってこの問題を確認した。エージェントが生成的な関係にある二つの事象の間の因果関係を推論するとき、その母集団が生成的な関係にあるかどうかはエージェントにはわからない。エージェントがこの点を無視して、サンプリング結果を抑制的であるとして破棄し、生成的な分割表が得られるまでサンプリングを続けると考えるのは不自然である。したがって、生成的な確率分布から得られた抑制的な分割表は、恣意的に捨てられるのではなく、モデルが低い値をとる場合として判断されるべきである。

3.2.2 問題 2: モデルの定義可能性

もう一つの問題点が、モデルの定義可能性を考慮していない点である。モデルが定義される(モデルが特定の値を取る)場合には、表 3.2 に示すような違いがある。表は a, b, c, d がそれぞれ 0 か 1 以上 (0 or +) であるかを考慮したときに可能な 16 通りの組み合わせに対して、各モデルが定義できるか否か (D or —) を示している。発生頻度が不明瞭で

表 3.2 すべての可能なサンプルに対するモデルの定義可能性。上の 4 行の 0 と + は、それぞれセルの値 (自然数) が 0 または 1 以上であることを示す。下の 4 行における、— と D は、それぞれモデルが定義不可能、定義可能であることを意味する。

Case	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$N(C, E)$	0	0	0	0	0	0	0	0	+	+	+	+	+	+	+	+
$N(C, \neg E)$	0	0	0	0	+	+	+	+	0	0	0	0	+	+	+	+
$N(\neg C, E)$	0	0	+	+	0	0	+	+	0	0	+	+	0	0	+	+
$N(\neg C, \neg E)$	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+
ϕ	—	—	—	—	—	—	D	D	—	D	—	D	—	D	D	D
DFH	—	—	—	—	—	—	D	D	D	D	D	D	D	D	D	D
ΔP	—	—	—	—	—	D	D	D	—	D	D	D	—	D	D	D
$pARIs$	—	—	D	D	D	D	D	D	D	D	D	D	D	D	D	D

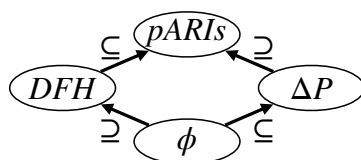


図 3.1 定義可能性に関するモデル間の包含関係を示すハッセ図

あるという枠組み外の状況を除いて、モデルの値が定義されないということはゼロ除算の発生を意味する。例えば、 $a = 0, b = 0, c = 1, d = 10$ (Table 3.2 の Case3) のとき、 DFH はゼロ除算が生じるために値が定義されないが、 $pARIs$ は値に 0 をとる (すなわち、定義可能である)。このように、定義可能な場合はモデルごとに異なり、そこには図 3.1 に示すような包含関係が存在する。表 3.2 は $pARIs$ の定義可能性は他の全てのモデルを包含しており、したがって他のモデルよりも多くの場合に定義可能であることを示している。表 3.2 は、0 ではないセルが多いほど、より多くのモデルが定義され、中でも a の生起が重要であり、 d はあまり重要ではないことを示唆している。

$pARIs$ が他のモデルよりも高い定義可能性を有するのは、 $pARIs$ が単一の分数からな

り、分母には d セルを除く全てのセルが含まれるためであると考えられる。このような定義可能性の問題は、各セルが十分に観測されていない場合が多い小サンプルからの推論を行う上で特に重要であり、また稀少性を仮定する上では更に重要となってくる。したがって、この定義可能性の問題は、モデル値の平均と標準偏差 (SD) の推移を示すシミュレーション C の特に初期段階に影響を与える。これは、定義される場合が限定的なモデルは、より多くの場合で定義されるモデルよりも SD が見かけ上小さくなるためである。異なる定義可能性を持つモデルを比較するとき、定義可能性は暗黙のうちに結果の妥当性に影響を与えているはずであるが、元のシミュレーションではモデルの定義可能性が考慮されていない。前の節で述べたように、Hattori and Oaksford [2007] のシミュレーション 1 において作り出される分割表には、生成的な分割表に限るという制約に加えて、 ϕ を定義可能な分割表に限るという条件も課されている。後者の条件は、 DFH が ϕ と同様に振る舞うことを検証するシミュレーション Aa において妥当であるが、複数のモデル間で性能の比較を行う場合は、この限りではない。表 3.2 で示すように、 DFH や ϕ が定義されない場合でも、他のモデルは定義されうるからである。つまり、複数のモデル間で性能の比較を行う場合、作り出される分割表は、比較されるモデルが全て定義できない場合を除いて破棄されるべきではない。

3.3 シミュレーション Aa

シミュレーション Aa は、 DFH および $pARIs$ が共変情報の規範的モデルである ϕ と、どの程度同様に振る舞うかをテストするものである。 DFH は極稀少性仮定に基づく ϕ の特殊化であるために、このシミュレーションには DFH が自身の元となった規範的モデルと比べてどのように振る舞うかを確認する意義があった。一方で、 $pARIs$ は ϕ の特殊形として導出されるわけではなく、 DFH の場合と比べて ϕ との適合を確認する意味が希薄である。また、モデルに要求される計算上の目標はサンプルの理解ではなく母集団推定精度にあり、これは次のシミュレーションで検証される。

3.3.1 方法

シミュレーションの手順は以下の通りである。生成する分割表の平均サンプルサイズを $\mu = \{7, 21, 56\}$ 、分散を $\sigma^2 = (\mu/7)^2$ に対して変化させ、また、 $P(C)$ 及び $P(E)$ は、稀少性仮定の下での効果を確認するために、.1 から .9 までを .1 刻みで網羅的に変化させた。以下の手順を生成される 1,000 個の分割表に対して繰り返した。[x] を自然数への四捨五入とし、ワーキングメモリ容量の制限がもたらす影響を確認するために、分割表に含

Algorithm 1 Simulation Aa

```

for Number of trials do
  for Mean sample size  $\mu$  in {7, 21, 56} do
    for  $P(C) = P(E)$  in {.1, .2, .3, ..., .9} do
      for Number of contingency tables do
         $N$ (or  $N_w$ ) is sampled by Eq. 3.1 and rounded it off.
        Determine  $P(C, E)$  by Eq. 3.2.
        Calculate  $\phi_0$  from  $P(C), P(E), P(C, E)$ .
        Calculate  $P(a), P(b), P(c), P(d)$  by Eq. 3.3.
        Produce a contingency table from probability distribution.
        Calculate the value of each model for the table.
      end for
    end for
  end for
  Calculate the coefficient of determination from all tables.
end for
Calculate the mean coefficient of determination from all trials.

```

まれる事象の数を以下のように決定した。

$$N(\text{or } N_w) \sim [\text{Norm}(\mu, \sigma^2)] \quad (3.1)$$

$P(C, E)$ を次の値域からランダムにサンプリングした。

$$P(C, E) \sim \text{Unif}(P(C)P(E), \min(P(C), P(E))) \quad (3.2)$$

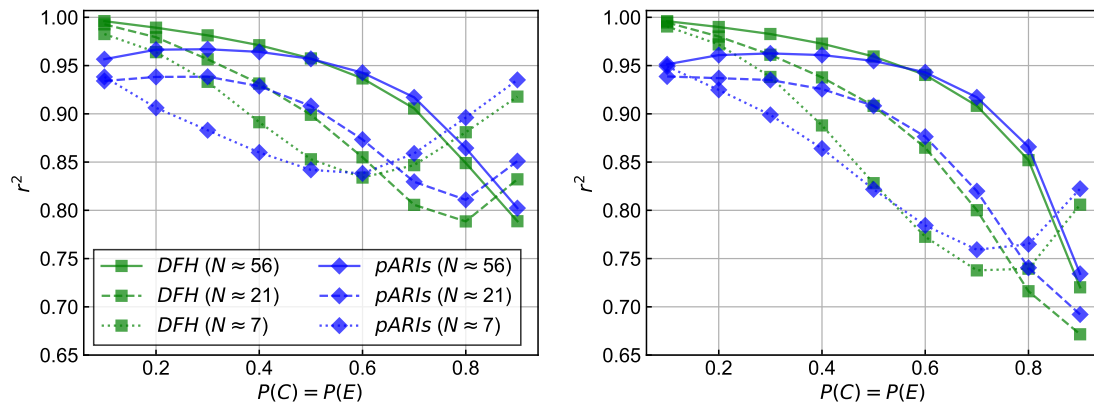
これにより、分割表の確率分布が以下のように決定された。

$$\begin{aligned}
 P(a) &= P(C, E) \\
 P(b) &= P(C) - P(C, E) \\
 P(c) &= P(E) - P(C, E) \\
 P(d) &= 1 - P(C) - P(E) + P(C, E)
 \end{aligned} \quad (3.3)$$

この確率分布に基づいて母集団の相関係数 ϕ_0 を計算し、この確率分布を元に分割表のサンプリングを行った。ここで、生成される分割表には比較するモデル全てが定義されるという条件を課し、条件を満たさないものは破棄した。先程言及したように、定義可能性はモデルごとに異なるので、シミュレーションの結果はそのことを加味して分析されるべきである。1,000個の分割表を生成した後、各分割表に対してモデル値を計算し、 ϕ との決定係数 r^2 を計算した。安定した見積もりを得るため上記の過程を100回繰り返し、平均を取った。以上の手続きを整理したものをアルゴリズム1に示した。

3.3.2 結果と議論

シミュレーションの結果を図 3.2 に示した。この図は等確率性が仮定されたときの



(a) Hattori and Oaksford [2007] と同様に抑制的なサンプル (分割表) を除いた場合

(b) 抑制的なサンプルを含めた場合

図 3.2 等確率性仮定の下での $pARIs$ および DFH のサンプル相関係数 ϕ との決定係数

DFH および $pARIs$ と ϕ の間の決定係数を示している。また、横軸は原因と結果の生起確率を示している。図 3.2(a) は抑制的なサンプルを除いた場合の結果であり、図 3.2(b) は抑制的なサンプルを含めた場合の結果である。 N_W -サンプリングについては、 d セルを無視する $pARIs$ と DFH のみ結果に含めた。結果より、抑制的な分割表を結果に含めた場合に、稀少性仮定のもとで DFH は依然として ϕ_0 と高い相関を示すことが明らかになった。他方、 $P(C) = P(E)$ が高い (すなわち、 $P(C) = P(E)$ がおおよそ .5 以上) の場合には ϕ との適合が著しく低下することが分かるが、この事実は $P(C)$ と $P(E)$ が共に小さい場合に焦点を当てている元論文の結果の解釈には影響を与えないと考えられる。また、 $pARIs$ は DFH よりもわずかに劣るものの ϕ と高い相関を示し、したがって ϕ と同様に振る舞うことが分かる。

3.4 シミュレーション Ab

リカバリーテストにおける目的はサンプル自体の理解ではなく、母集団からサンプリングされた疑似データから、元の母集団を推定することである。したがって、各モデルと ϕ_0 との適合を見ることは、前節で示したような各モデルと ϕ との適合を見ることよりも遥かに重要である。このシミュレーションは各モデルと母集団の相関係数である ϕ_0 との適合を確認する。

3.4.1 方法

各モデルと ϕ_0 の適合性を調べた以外は、シミュレーション Aa と同じ手順で行った。

3.4.2 結果と議論

図 3.3 に、各モデルの ϕ_0 に対する決定係数を示す。図 3.3(a) が生成的な分割表のみを使用した場合の結果であり、図 3.3(b) が抑制的なものも含めた場合の結果である。結果は、全体的に $pARIs$ の性能は DFH よりも優れていることを示している。まずは N -サンプリングの結果について考察する。生成的な分割表のみを含む場合に注目すると、事象の生起確率が稀であるとき、 DFH や $pARIs$ の ϕ_0 に対する適合は ϕ のそれと殆ど一致した。使用する分割表に抑制的なものも含めた場合、稀少性仮定のもとで、 $pARIs$ は DFH や母集団に用いたモデルそのものである ϕ よりも ϕ_0 優れた適合を示した。他方で DFH は ϕ よりも適合しなくなることを示した。つづいて N_W サンプリングについて考察する。生成的な分割表のみを含む場合には、 $pARIs$ が DFH よりも全ての $P(C) = P(E)$ に対して優れた適合を示し、抑制的な分割表も含めた場合にその差はより大きくなる。使用した分割表が抑制的なものを含む場合、特に $P(C) = P(E)$ が .1 または .9 に近いとき、両モデルと ϕ_0 との適合度が高くなることは、両サンプリング方法に共通である。

3.5 シミュレーション Ac

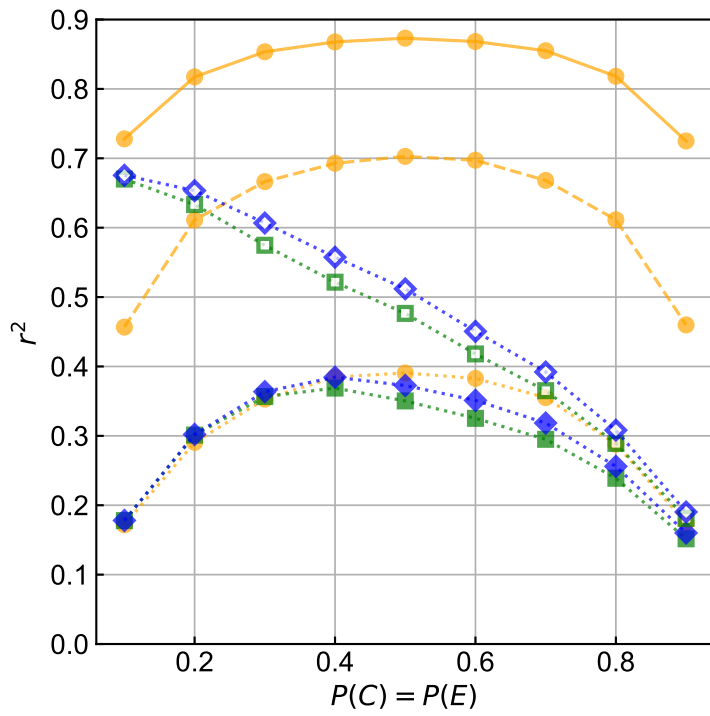
我々は母集団推定性能における等確率性の影響を調べるためにシミュレーション Ac を行った。

3.5.1 方法

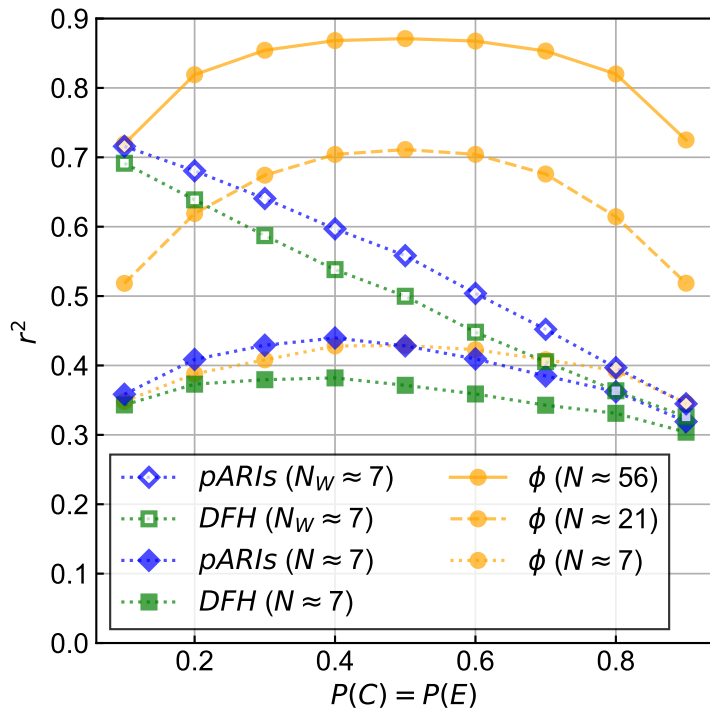
シミュレーション Ab で得られた結果から、 $N_W \approx 7$, $P(C), P(E) \leq .5$ のときの結果を抜き出し、等確率性の下での母集団推定性能を検証するためのグラフを作成した。

3.5.2 結果と議論

図 3.4 は、スモールサンプル ($N_W \approx 7$) かつ 稀少性下 ($P(C), P(E) \geq .5$) を仮定したときの、各モデルと ϕ_0 との決定係数を示している。縦軸は ϕ_0 との決定係数、横軸は原因と結果の確率の差 $P(C) - P(E)$ であり、この差が 0 のときに 2 つの事象は等確率である。原因と結果の平均確率 (m) が等しいものは線で結ばれている。例えば、 $(P(C), P(E)) = (.1, .5), (.2, .4), (.3, .3), (.4, .2), (.5, .1)$ のペアは全て $P(C)$ と $P(E)$ の



(a) Hattori and Oaksford [2007] と同様に抑制的なサンプル (分割表) を除いた場合



(b) 抑制的なサンプルを含めた場合

図 3.3 等確率性仮定の下での $pARIs$ および DFH の母集団相関係数 ϕ_0 との決定係数

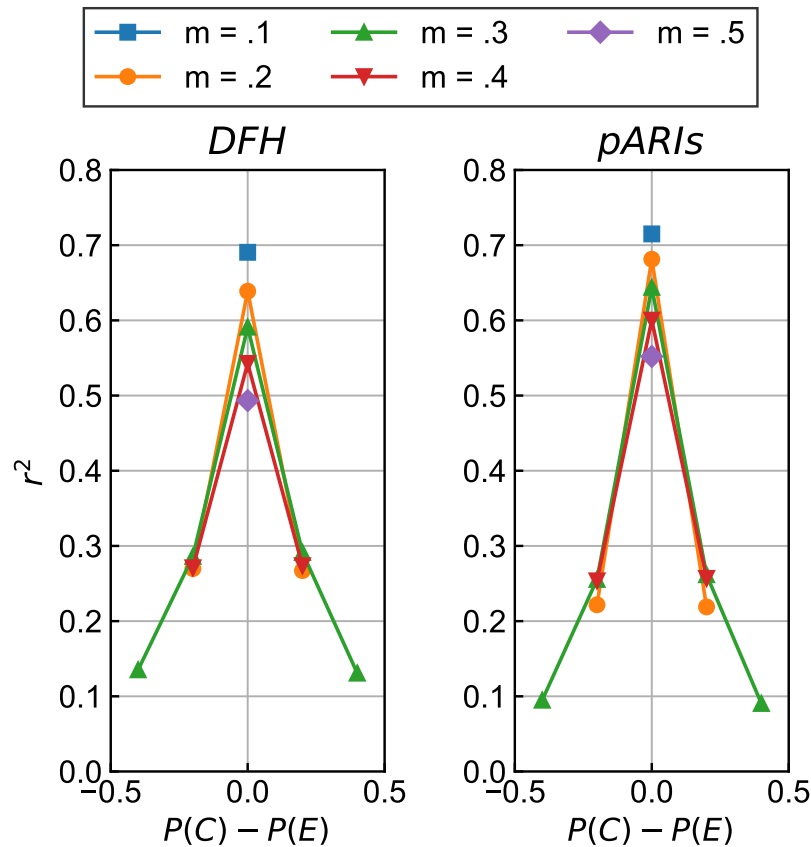


図 3.4 等確率性からの乖離度合いとモデルの母集団推定性能

平均が .3 となる場合であり、これらは $m = .3$ の線で繋がれる。また、この例において、各ペアの $P(C) - P(E)$ はそれぞれ、 $-.4, -.2, 0, .2, .4$ となる。我々は、使用する分割表を生成的なものに限定した場合と、サンプルされた全ての分割表を用いた場合の両方の場合についてシミュレーションを実行した。そして、 $pARIs$ と DFH はともに、Simulation Ab の結果と同様に、後者の場合よりも、前者の場合のほうが、全体的な適合がわずかに高いことを確認した。母集団推定において使用する分割表を制限することの影響は Simulation Ab で十分示されているうえ、どちらの場合においても両方のモデルが等確率性のもとで高い母集団推定性能を示すという主張には影響しないため、我々は、後者の結果のみを図 3.4 示した。このシミュレーションの結果は、 $pARIs$ が DFH と同様に、等確率性の仮定の下で優れた母集団推定性能を有することを示す。

3.6 シミュレーション Ba

3.2 節で述べた二つの問題点について、その重要性を定量的に検証するためにシミュレーション B を行った。シミュレーション Ba では、生成的な母集団からのサンプリングによって得られる抑制的な分割表の割合を、シミュレーション Bb では、各モデルに対して、それが定義されない場合がどの程度の割合で生じるかを定量的に分析した。

3.6.1 方法

シミュレーション A で生成された分割表について、アルゴリズム 1 に示した手順に則って、生成的 (式 2.1) な母集団から偶然サンプリングされる抑制的 (式 2.2) な分割表の割合を計算した。

3.6.2 結果と議論

シミュレーションの結果を図 3.5 に示す。結果は N -サンプリングのもとでは、事象の

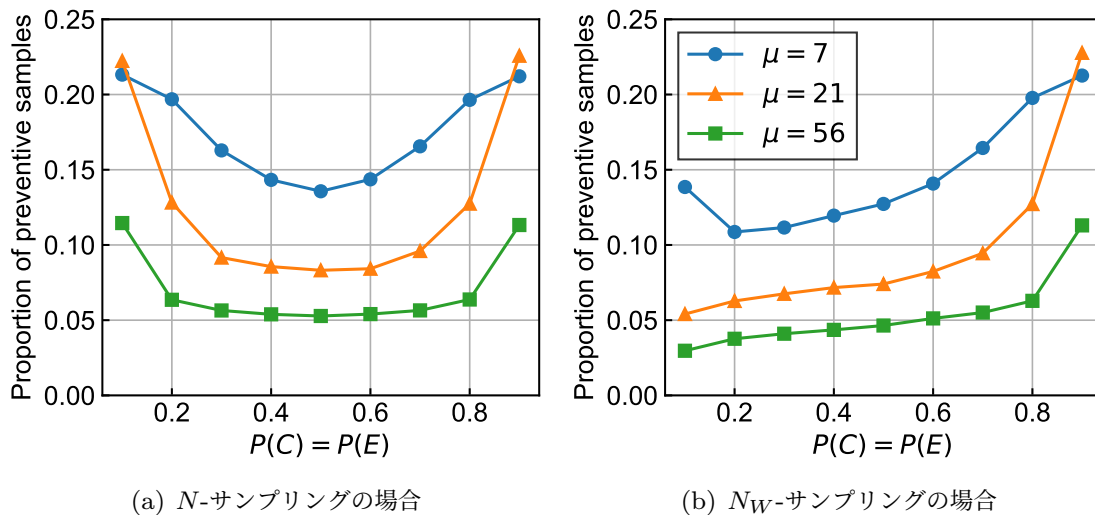


図 3.5 各 $P(C) = P(E)$ における生成的な母集団からサンプリングされた抑制的な分割表の割合。

生起確率 $P(C) = P(E)$ がより小さいかより大きいほど、あるいはサンプルサイズが小さいほど、抑制的な分割表がサンプルされる割合が大きくなり、最大で 2 割程度であることを示す。一方、 N_W サンプリングのもとでは、事象の生起確率 $P(C) = P(E)$ が高くなるほどこの割合は大きくなる傾向にあり、 N -サンプリングの場合と同様に最大で 2 割程度であることが示された。この結果は、生成的な母集団から作成される抑制的な分割表の割

Algorithm 2 Simulation Bb

```

 $P(C) = P(E) = .2$ 
for Sample size  $N = 0$  to 15 do
  for  $\phi_0$  in  $\{.0, .2, .4, \dots, 1.\}$  do
    Calculate  $P(a), P(b), P(c), P(d)$  by Eq. 3.3.
     $N$ (or  $N_W$ ) is sampled by Eq. 3.1.
    for Number of contingency tables do
      Sample an event from probability distribution.
      Calculate the value of each model for the current table.
    end for
    For each model, calculate the proportion for which a value is defined.
  end for
end for

```

合が軽視できない量であることがわかる。

3.7 シミュレーション Bb

このシミュレーションでは、観測した事象の数が一つずつ増加してゆく連続的なサンプリングにおいて、各モデルが定義される割合がどのように遷移するかを示した。

3.7.1 方法

まずはサンプルを生成するための母集団を設定した。母集団の相関係数 ϕ_0 と事象の生起確率を、それぞれ $\{.2, .8\}$ について変化させた。次に、設定した各 ϕ_0 について 100,000 個の分割表を生成した。ここで、各分割表はワーキングメモリ容量を反映し、事象の数が 15 に達するまでサンプリングを行った。各分割表における事象の数が 7 に達するまでの間、モデルが定義されかどうかを調べ、最後にその割合を求めた。以上の手続きを整理したものをアルゴリズム 2 に示した。

3.7.2 結果と議論

シミュレーションの結果を図 3.6 に示した。結果は、全ての場合で $pARIs$ の定義される割合が他のモデルと比べて大きく、その差は ϕ_0 または $P(C) = P(E)$ が小さいほど大きくなることを示している。また、 $N_W > 0$ の場合には $pARIs$ は必ず定義されることも示しており、これは、式 1.12 から明らかである。 N サンプリングかつ稀少性下では、定義される割合は概ね $pARIs > \Delta P > DFH \equiv \phi$ のような大小関係にある。また、 N サンプリングの場合かつ稀少性仮定のもとでは、 DFH の定義可能性は ϕ と同程度しかないとわかる。この結果は、図 3.1 で示し表 3.2 に詳述した包含関係と整合的である。

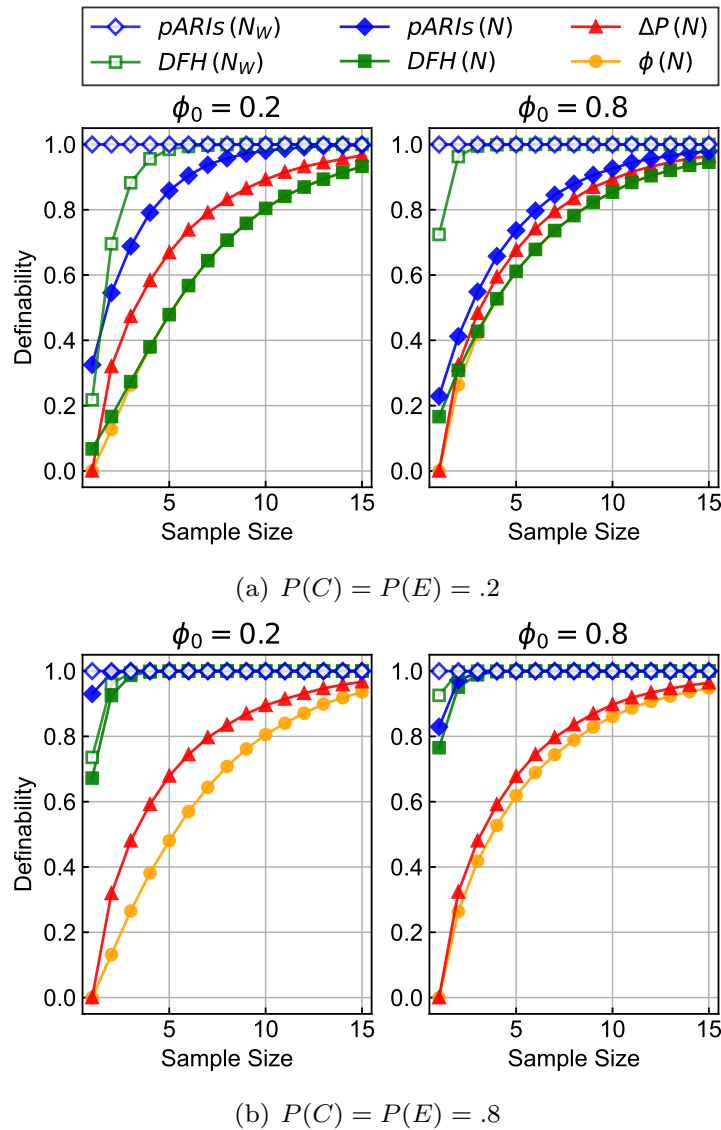


図 3.6 各モデルが定義される割合の時間発展

3.8 シミュレーション C

このシミュレーションでは、母集団の ϕ_0 を変動させた上で、分割表における事象の数が一つずつ増加してゆく連続的なサンプリングにおいて、各モデルがとる値の平均値と標準偏差がどのように推移するかを示した。現実における因果帰納はリアルタイムに事象の生起を観測する逐次サンプリングによって行われる。そのような環境下において高い精度の推定を得るためには、より多くの事象を観測する必要がある。サンプリングにかかる時間の短さ（速さ）と推定の正確さの間には一般的にトレードオフがある。つまり、より短いサンプリングで正確な推定が得られるということは、モデル値の収束が早く、より分散

Algorithm 3 Simulation C

```

Initialize  $P(C) = P(E) = .2$ .
for  $\phi_0$  in  $\{.2, .5, .8\}$  do
  Calculate  $P(a), P(b), P(c), P(d)$  by Eq. 3.3.
  for Number of contingency tables do
    for Sample size  $N = 1$  to 30 do
      Sample an event from probability distribution.
      Calculate the value of each model for the current table.
    end for
    For each model, calculate the mean and SD for each sample size.
  end for
end for

```

が小さいことを意味する。

3.8.1 方法

母集団パラメータについて、母集団の相関係数 ϕ_0 を $\{.2, .5, .8\}$ の三段階で変動させ、稀少性仮定に基づき $P(C) = P(E) = .2$ とした。設定した各 ϕ_0 について 10,000 個の分割表を作成した。各分割表における事象の数が $N(\text{or } N_w) = 30$ に達するまでの間、各分割表からモデルの値を算出し、最後に各 $N(\text{or } N_w)$ における平均と標準偏差を求めた。以上の手続きを整理したものをアルゴリズム 3 に示した。

3.8.2 結果と議論

シミュレーションの結果を図 3.7 に示す。左の 3 つの図が各モデルを 10,000 回の試行について平均したモデル値の推移であり、右の 3 つが標準偏差の推移である。それぞれ上から順に $\phi_0 = \{.2, .5, .8\}$ の場合の結果である。また、 $N = \{7, 21, 56\}$ の三段階において、 ϕ_0 を $[.0, 1.]$ の範囲で変化させたときの各モデルの平均値を図 3.8 に示す。なお、 ϕ_0 の設定を $[.0, 1.]$ の範囲で変化させた以外は 3.8.1 節で説明したものと同様の手順を用いた。黒い破線は ϕ_0 の値を示しており、これに近いほど ϕ_0 の値を忠実に近似することを意味する。平均値の推移に着目すると、 ΔP と $pARIs$ は最初から値が収束しているのに対し、 ϕ と DFH は最初に高めの値を取り、次第に値を減少させながら収束してゆく傾向にあることが分かる。 DFH のこのような挙動は、観測事象の数が 1 つしかない場合、 DFH は $a = 1$ のときにのみ定義され、その時 $DFH = 1$ であることから説明できる。また、 ϕ_0 に対する各モデルの平均値 (図 3.8) を見ると、各モデルが有するバイアスが確認できる。 DFH の平均値は常に ϕ_0 や他の平均モデル値以上であり、したがって実際の値よりも上方へ偏った値を取る傾向にある。また、 ϕ と ΔP は十分なサンプルが集まった場合 ($N \geq 21$) には ϕ_0 とほぼ一致するがスモールサンプル ($N = 7$) では微妙に上方へ偏っ

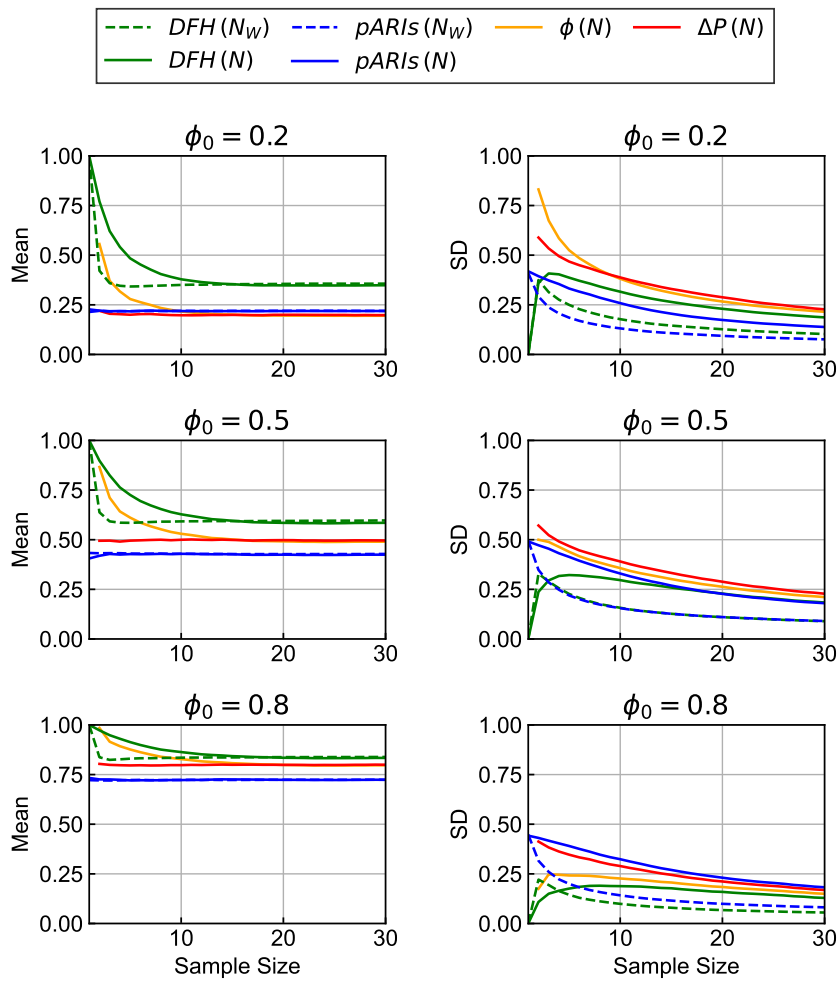


図 3.7 モデルの平均値と標準偏差 (SD) の時間発展

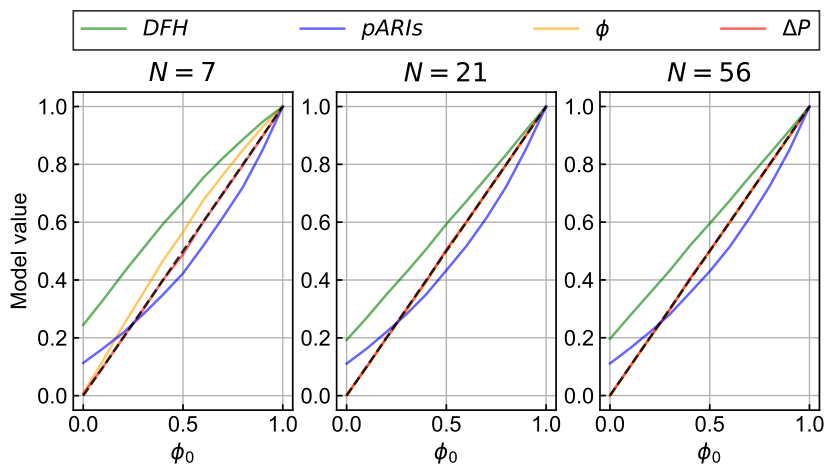


図 3.8 ϕ_0 の関数としてのモデル平均値

た値を取る。そして $pARIs$ は、 ϕ_0 が小さい場合は大きめの値を取り、それ以外の場合では小さめの値を取る。したがって $pARIs$ は小さな共変動をより過大評価し、大きい共変動を過小評価する。これは、興味深いことに $pARIs$ がプロスペクト理論 [Kahneman and Tversky, 1979] における確率加重関数のように振る舞うことを意味する。一方、標準偏差の推移に着目すると、モデル間の相対的な比較として、まず DFH の標準偏差が安定して小さいということが分かる。ただし、特に初期段階の小ささについては、必ずしもモデルの優位性を示すものではないことに留意したい。先程言及したように、 DFH は平均値の推移において事象数が1の場合に必ず1を値に取るが、したがって標準偏差の推移も必ず0から始まるからである。また、 ΔP と ϕ の標準偏差は DFH よりも常に大きく、 $pARIs$ と DFH の標準偏差の大小関係は ϕ_0 の値に応じて変化することが読み取れる。ただし、上記の比較はあくまでもモデル間の標準偏差の相対的な比較であり、 $pARIs$ の標準偏差は ϕ_0 に比例して増加している訳ではなく、むしろ全ての ϕ_0 に対して一定の推移を見せている。結果からは、平均値の収束速度に関しては最初から安定した値を取る ΔP と $pARIs$ が優れているが、標準偏差に関しては安定して小さい DFH が優れているように見える。しかし、図 3.8 を見ると、標準偏差の大小を比較することがモデルの性能を評価するにあたってあまり意味がないことがわかる。なぜなら図 3.8 は、 DFH の平均値が常に ϕ_0 や他のモデル値の平均よりも大きいことを示しており、その範囲は約 $[.2, 1.]$ と、他のモデルよりも狭い。モデルの平均的な値が狭い範囲で変動するということは、同時にモデルの標準偏差の小ささをもたらすが、それは見かけ上の小ささでしかない。いずれにしても、平均値の推移と標準偏差の算出をした上でこれらを個別に分析する限りでは、モデル値の推移と信頼性を総合して評価することはできない。我々は次節で実施するシミュレーションにより、これらを総合して評価する。

3.9 シミュレーション D

因果推論の二段階仮説において、因果帰納モデルの役割はヒューリスティック段階にある。その役割とは、無数に存在する原因候補事象の中から最も結果事象と関係がありそうな組み合わせを抽出することである。そして、ここで抽出された原因候補に対してのみ、第二段階において介入の操作（「実験的検証」）が行われて、それが真の原因であるかどうかを検証される。つまり、因果推論の二段階仮説における各モデルの役割は、原因候補事象と結果事象の間に関係がありそうか否かの二値分類（構造の推定）と考えることができる。これによって、シミュレーション C において総合的に評価することができなかったモデル値の収束速度とその信頼性という2つの要因を、分類性能という観点の元で総合的に評価することが可能になる。また、モデルが定義されない場合にはランダムに分類させることで定義可能性も含めて評価する。このシミュレーションでは、各モデルが本質的な

目的のために動作するにあたって、それらがどのような性質上の違いを検証するためのものであり、同時に、単に因果的な強度の問題である単純因果帰納を、因果関係の有無の問題である因果構造の観点から再解釈する糸口を提供するものである。

3.9.1 方法

ROC-AUC (Area Under the Curve of Receiver Operatorating Characteristic curve) は、分類性能の指標として多くの望ましい特性を示すことが示されている [Bradley, 1997]. 他方, Saito and Rehmsmeier [2015] は、不均衡データに対して分類器を評価する場合、ROC 曲線よりも RP 曲線 (Precision-Recall curve) を用いたほうがより有益であることを示した。今回のシミュレーションでは母集団パラメータ (i.e. ϕ_0 の閾値) の設定によってデータが不均衡になりうるため、分類性能の評価指標として ROC 曲線の AUC と PR 曲線の AUC の両方を用いた。

以下に、二つの曲線の計算方法を説明する上で必要な事項を確認する。共変関係があることを Positive と表し、共変関係がないことを Negative と表すと、以下の混同行列 (表 3.3) によって正解と予測の組み合わせを表すことができる。この行列から、モデル

表 3.3 二値分類における 2×2 混同行列

	Predicted Positive (PP)	Predicted Negative (PN)
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Negative (N)	False Positive (FP)	True Negative (TN)

の分類性能の評価の指標としてよく知られる 3 つの指標が計算できる。TPR は、実際に Positive であるもののうち、正しく Positive と分類した割合 ($\frac{TP}{TP+FN}$) であり、これは Recall と呼ばれる。FPR は、実際に Negative であるもののうち、誤って Positive と分類した割合 ($\frac{FP}{TN+FP}$) である。precision は、Positive と分類したもののうち、正しく Positive と分類できたものの割合 ($\frac{TP}{TP+FP}$) である。以上の概念を用いて、AUC の計算過程を説明する。まずは、母集団 ϕ_0 の閾値を設定する。続いて、生成する分割表の数だけ ϕ_0 を一様分布からサンプリングし、各 ϕ_0 に基づいて分割表を生成する。母集団 ϕ_0 とその閾値によって、各分割表が実際に Positive か、Negative か、すなわち、真に共変関係があるか否かを示すラベルが決定される。続いて、共変関係があるか否かを予測するために、全ての分割表に対して、モデル値を算出する。以上の操作により、正解ラベルとモデル値のペアが分割表の数だけ存在することになる。曲線を描くために、このペアをモデルの値の大きい順にソートし、後述するようにモデルの閾値を動かしながら縦軸を TPR、横軸を FPR とした二次元座標に点をプロットする。まず、閾値を最大のモデ

ル値よりも高い値に設定する。すると、モデル値は全て閾値以下になるので、このときモデルは全ての分割表に対して Negative と分類することになる。このとき、 $TPR = 0$, $FPR = 0$ となるので、点 ($TPR = 0, FPR = 0$) を座標にプロットする。続いて、閾値を最大のモデル値と 2 番目に大きいモデル値の間に設定し、また、このときの TPR, FPR とを求め、これを座標にプロットする。このように、降順にソート済みの各モデル値の間に閾値を設定するように、閾値をずらしながら点 (TPR, FPR) をプロットすることを繰り返す。最後に、閾値をモデルの最小値よりも低い値に設定し、点 ($TPR = 1, FPR = 1$) をプロットする。ROC 曲線は、以上のようにプロットされた全ての点を繋いだときに描かれる曲線である。曲線の下部面積が ROC-AUC であり、これが 1 に近づくほどモデルの性能が良いことを示す。ROC-AUC は最良の分類モデルに対する値は 1 であり、ランダムな分類モデルに対して .5 となる。また、PR 曲線は、同様の手順を縦軸を Precision、横軸を Recall とした座標に対して行ったときに描かれる曲線である。PR-AUC も最良の分類モデルに対する値が 1 であり、ランダムな分類モデルに対する値は全分割表における Positive の割合、すなわち $P/(P + N)$ によって決定される。両 AUC はこのような過程から計算されるため、恣意的な閾値の設定によらずモデルの性能を評価できる利点がある。

本シミュレーションにおいて、各テーブルの母集団 ϕ_0 は一様分布からサンプリングされるため、 ϕ_0 の閾値の設定によってデータは不均衡になる。^{*1} 先に述べたように、不均衡データ（正解におけるポジティブの数とネガティブの数が大きく異なるデータ）に対して、PR 曲線よりも ROC 曲線による評価のほうが有益であることが報告されているため、シミュレーションの設定によって評価に用いる曲線を変更した。ROC で用いる FPR と TPR はそれぞれ、「正解が」Positive であるデータ、「正解が」Negative であるデータを対象とした値であるために不均衡データの影響を受けやすい。したがって、 $P(C) = P(E)$ の変化に対する分類性能の変化を検証するときには ϕ_0 の閾値を .5 を固定し ROC-AUC を使用する。一方、PR 曲線は TPR を用いる点では ROC 曲線と同じだが、Positive と「予測」したデータを対象とした値である Precision を用いるために不均衡データに対して相対的にロバストである。よって、 $P(C) = P(E) = .2$ に固定し、 ϕ_0 の閾値を変化に対する AUC の変化を検証するときには PR-AUC を用いる。

シミュレーションの手順は以下の通りである。分割表を生成するために 3 つの母集団パラメータである $P(C), P(E), P(C, E)$ を設定した。 $P(C)$ 及び $P(E)$ は主に稀少性仮

^{*1} 現実世界で分類モデルが用いられる文脈において、殆どの場合、「正解が Positive である数が極端に少ない」不均衡データを対象とすることになる。例えば、スパムメールか否かの分類問題において、スパムメールはそれ以外よりも稀である。また、異常検知問題において異常の発生は稀であり、病気の診断においても同様だろう。そもそも、因果推論の二段階仮説自体が、無数にある事象のなかからごく一部の原因事象を見つけ出すことを念頭に置いている。不均衡データを対象とする分類問題は、稀少性仮定とも直接関連している可能性が高い。

Algorithm 4 Simulation D

```

for  $P(C) = P(E)$  in  $\{.2, .5, .8\}$  do
  for Mean sample size  $\mu$  in  $\{4, 8, 16, \dots, 512\}$  do
     $N$  (or  $N_W$ ) is sampled using Eq. 3.1.
    Determine  $P(C, E)$  using Eq. 3.2 and rounded off.
    Calculate  $\phi_0$  using  $P(C), P(E), P(C, E)$ .
    Calculate  $P(a), P(b), P(c), P(d)$  using Eq. 3.3.
    for Number of contingency tables do
      Produce a table (sample size  $N$ ) from probability distribution.
      Calculate the value of each model for the table.
    end for
    for Threshold of  $\phi_0$  in  $\{.2, .5, .8\}$  do
      Compute AUC scores for each model.
    end for
  end for
end for

```

定の下での効果を確認するために、 $P(C) = P(E)$ を $\{.2, .5, .8\}$ の三段階で変化させた。 $P(C, E)$ はその値域からランダムにサンプリングされた (式 3.2)。これにより、各分割表のセルの確率分布がそれぞれ 式 3.3 のように決定された。この確率分布に基づいて、分割表ごとに母集団の相関係数 ϕ_0 を計算した。シミュレーションでは、各サンプルサイズ $\mu = \{4, 8, 16, \dots, 512\}$ に対して、このような分割表を 100,000 個生成した。生成した全ての分割表に対して、次の 2 つの操作を行った。第一に、 ϕ_0 とその閾値から、事象間に真の関係性があるかどうかを設定した。ただし、 ϕ_0 の閾値は、 $\{.2, .5, .8\}$ の三段階で設定した。このとき、 ϕ_0 がその閾値よりも大きければ事象間に真の関係性が認められることになる。第二に、分割表に対応するモデルの値を算出した。モデルの二値分類は、AUC の算出過程において網羅的に検証されるモデルの閾値によって決定される。なお、モデルが定義されない場合は、モデルの二値分類は無作為に行われるもの (つまり 確率 .5 で Positive または Negative) とした。上記までの手順で得られたデータをもとに、AUC スコアを算出した。以上の手続きを整理したものをアルゴリズム 4 に示した。

3.9.2 結果と議論

シミュレーションの結果を図 3.9 に示した。各サンプルサイズに対する分類性能を確認するため、グラフの縦軸は AUC スコア、横軸はサンプルサイズの平均値とした。

図 3.9(a) は ϕ_0 の閾値を .5 と固定した場合の、各 $P(C) = P(E)$ の値に対する分類性能を示す。この場合、対象とする分割表のラベルに偏りはないため、ROC-AUC を計算した。ROC-AUC において、完全にランダムな分類をするモデルがとる値はサンプルサイズによらず .5 となるが、全てのモデルがこのベースラインよりも高い値を取る。結果は、

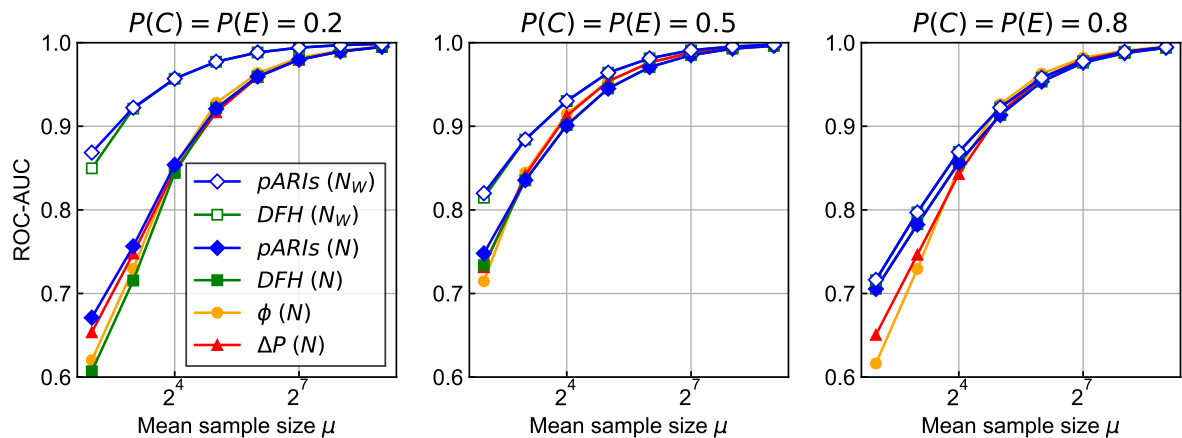
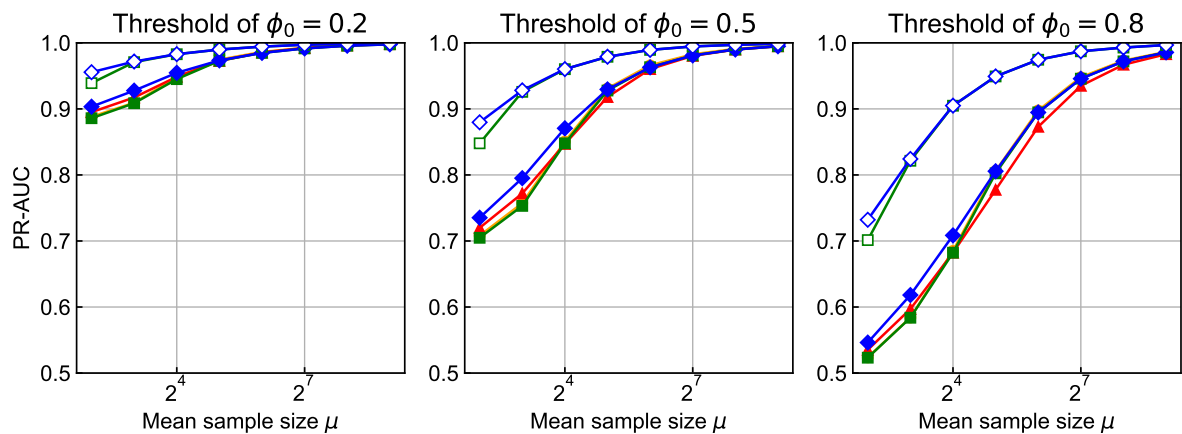
(a) ϕ_0 の閾値を .5 としたときの、各 $P(C) = P(E)$ における ROC-AUC(b) $P(C) = P(E) = .2$ としたときの、各 ϕ_0 の閾値における RP-AUC

図 3.9 各モデルの分類性能 ROC-AUC スコアおよび PR-AUC スコア

$pARIs$ が稀少性仮定下、かつスモールサンプルにおいて他のモデルよりも優れていることを示している。また、 $P(C) = P(E)$ が増加するにつれ、モデル間の差は小さくなる。

図 3.9(b) は稀少性仮定に基づいて $P(C) = P(E) = .2$ と固定した場合に、各 ϕ_0 の閾値に対する分類性能の評価を示す。 ϕ_0 は一様分布からサンプリングされるため、 ϕ_0 の閾値によっては不均衡データが分類の対象となる (例えば、 ϕ_0 の閾値を .8 としたとき、全分割表のうちの 8 割が Positive になる)。したがって、この場合は PR-AUC を計算した。完全にランダムな分類をするモデルがとる PR-AUC は、サンプルサイズに依らず、 ϕ_0 の閾値が $\{.2, .5, .8\}$ のとき、それぞれ、 $\{.8, .5, .2\}$ となる。したがって、結果より、全てのモデルがこのベースラインよりも高い AUC を示すことがわかる。また、結果は、稀少性の下で、 $pARIs$ が ϕ_0 の閾値に依らず、他のモデルよりも分類モデルとして優れていることを示している。

以上の結果から、共変関係の有無に関する二値分類の文脈においても、稀少性が成り

立ち、かつスモールサンプルの場合に $pARIs$ が優れた性能を示した。これらの結果は、 $pARIs$ が稀少性仮定下やスモールサンプルからの推論のために最適化された（すなわち適応的に合理的な）モデルであることを示している。

3.10 総合議論

本研究ではメタ分析の追試とシミュレーションを行うことで DFH と $pARIs$ の合理分析を行った。この章では、メタ分析とシミュレーションによって得られた結果を元に総合的な議論を行う。

3.10.1 $pARIs$ vs. DFH

$pARIs$ と DFH のどちらが人間の因果帰納の記述モデルとして適しているかという点について、メタ分析の結果 (2.1) は、 DFH よりも $pARIs$ のほうが高い記述的性能を有することを示している。この節では、 $pARIs$ と DFH のどちらが因果推論の二段階仮説のコンセプトに適しているかを、計算コストと母集団推定性能の観点から議論する。

モデルの定義可能性は、そのモデルが多くの状況においてエージェントに情報を与えることを意味する。シミュレーション Bb の結果 (図 3.6) から、全ての ϕ_0 について、 $pARIs$ の定義される割合は他のモデルよりも大きいことが分かる。これは比較モデルの中で唯一、 $pARIs$ が単一の確率で記述されること、つまり他のモデルと比較して計算コストが小さい（つまり、経済的である）ことを反映した結果であると考えられる。因果推論の二段階仮説に基づき、因果帰納モデルの役割は膨大にある原因候補事象の中から関係がありそうな事象の抽出することであると考え、モデルは経済的であることが好ましい。各モデルの定義式 (式 1.2, 1.8, 1.12) を基本的な演算ごとに分解することで、その複雑さについて考察する。 ϕ は 4 つの加算, 6 つの乗算, それぞれ 1 つの減算と除算, 更に 1 つの平方根からなる。 DFH は 2 つの加算, それぞれ 1 つの乗算と除算, 更に 1 つの平方根からなり、確かに ϕ よりも経済的であることが分かる。ただし、 $pARIs$ は 2 つの加算と 1 つの除算からなり、 ϕ や DFH よりも経済的であることは明らかである。モデル間における定義可能性の差は、 ϕ_0 が小さいほど顕著に現れる。これは ϕ_0 が小さいほど $pARIs$ が他のモデルよりも定義可能性において優れていることを示唆する。

続いて母集団推定性能に着目すると、シミュレーション Ab (図 3.3) により、 N サンプルリング時の DFH の母集団推定性能は、使用するサンプルを生成的な分割表に限った場合には ϕ と同等であるものの、抑制的な分割表も含めると ϕ よりもわずかに劣るものとなることが明らかになった。他方で $pARIs$ の母集団推定性能は、生成的な分割表のみを使用した場合に ϕ と同等であり、抑制的な分割表も含めた場合には、稀少性仮定のもとで

ϕ よりも優れていることが示された。また、 N_W サンプリングの場合を考慮すると、同じサンプルサイズに対して DFH は ϕ よりも遥かに高い母集団推定性能を示すが、このとき $pARIs$ は全ての $P(C) = P(E)$ について DFH よりも高い母集団推定性能を有することが示された。

また、シミュレーション D の結果 (図 3.9) は、共変関係の有無に関する二値分類の文脈において、定義可能性を考慮した上で、稀少性が成り立ち、かつスモールサンプルの場合に $pARIs$ が優れた性能を有することを示した。二段階仮説におけるモデルの本来的な計算目標は、無数に存在する事象の中から介入に値する事象を絞り込むことである。このシミュレーションは、無数の事象に対して介入すべきか否かを正しく決定する性能を直接的に評価するという意味で本来の目的に即している。

Hattori and Oaksford [2007] は、彼らの提案する DFH が、3.1 節で述べた制約のもとで、規範的モデルである ϕ と同等の母集団推定性能を示し、かつ、より経済的であることを示すことで DFH の有効性を示した。本シミュレーションによって、同様の制約下において $pARIs$ が DFH や ϕ そのものよりも高い母集団推定性能を示し、更に ϕ や DFH より経済的であることが示された。この結果は、 $pARIs$ が DFH よりも因果推論の二段階仮説のコンセプトに適したモデルであることを示唆する。

3.10.2 シミュレーション設定の修正が結果に及ぼす影響

シミュレーション設定の修正が結果に及ぼす影響を考察するために、使用するサンプルに生成的な母集団から生成された抑制的な分割表を含めることで、 DFH の振る舞いがどのように変化するか考察する。まず、シミュレーション Aa の結果について考える。図 3.2 より、使用した分割表に抑制的なものを含めた場合、生成的な分割表のみを用いた場合と比べて、 C と E の生起確率が小さいときには DFH または $pARIs$ の ϕ に対する決定係数は僅かに高くなり、 C と E の生起確率が大きくなるに従って極端に低下することが分かる。したがって、稀少性仮定のもとでは、 DFH は依然として ϕ と同様に振る舞うことが分かった。続いて、シミュレーション Ab の結果について考える。図 3.3 より、 N サンプリングの場合、使用される分割表に抑制的なものを含めたとき、生成的な分割表のみを用いた場合と比べて、稀少性仮定のもとで DFH の ϕ_0 に対する適合が僅かに減少し、 ϕ よりも僅かに下回ることが分かる。これは「 DFH が ϕ と同等の母集団パラメータの推定性能を有する上で経済性の観点からも優れているため、 DFH は ϕ よりも優れたモデルである」という主張の妥当性を損ねうる結果である。

3.11 結論

本章では, Hattori and Oaksford [2007] が行ったシミュレーションの設定に問題があることを指摘した. これらの問題を可能な限り修正し, あるいは必要なシミュレーションを追加して, 再度シミュレーションを行い, 比較対象に *pARIs* を追加した上で結果を比較した. 得られた結果は, Hattori and Oaksford [2007] の提唱する二段階仮説に対して, *pARIs* が *DFH* よりも適切なモデルであることを示した. また, これらのシミュレーションの設定を変更することで, 彼らが行った合理分析の結果やその解釈に影響を与えることを示した.

第4章

非独立性の検出としての因果帰納

この章では、人間の因果的判断と高い適合を示すことが明らかになっている観察的因果帰納モデル $pARIs$ の計算論的な目標について、従来は共変動検出と見なされていたものを、稀少性仮定の下での非独立性の検出であると見なす。我々は、稀少性仮定下での非独立性の尺度を導入した上で、 $pARIs$ が非独立性の近似として振る舞うことを説明し、加えて複数のシミュレーションを通して適応的な合理性の観点から $pARIs$ の推定性能の分析を行った。

4.1 観察的因果帰納モデルの計算上の目標

1.3 節において記述した人間の因果帰納の二段階仮説において、ヒューリスティック段階における目標とは観察に基づいて結果事象の原因でありそうな事象を抽出することである。これに対応する計算上の目的の一つに、例えば前章まで仮定していた共変動の検出がある。これは、ある2事象について因果関係があればそれらは相関しているということが殆どの場合に成り立つため、相関関係にあるものを抽出すれば基本的に因果関係がある事象を見逃すことはないだろうという論理に基づくものである。Hattori and Oaksford [2007] はヒューリスティック段階の計算上の目標を共変動の検出と見なし、因果帰納モデル DFH がこれを行うとした。しかし、事象の絞り込みに適した計算上の目標は共変動検出だけに限らない。共変動検出の場合と同様の論理に基づき、非独立性の検出も事象の絞り込みに適していると言える。

相関と非独立の関係性について考えると、2事象が相関しているならばそれらは非独立であるという関係が成り立つが、その逆は成り立たない。これは相関があることは非独立であることよりも限定的な概念であることを意味している。例えば、2つの事象について、因果関係があるのに相関が見られない場合には原因事象の見逃しが発生してしまう。非独立の場合も同様の場合は存在するが、非独立であることは相関関係であることより

も包括的であるためにより見逃しが発生しづらい。この意味で非独立性の検出に基づくフィルタリングの方がよりヒューリスティック段階における事象の絞り込みという目的において優れていると考えられる。

本章は、人間の因果帰納の第一段階における計算上の目標が非独立性の検出にあると仮定した上で、これを行う観察的因果帰納モデルの合理性について分析を行う。

4.2 稀少性仮定下での非独立性の尺度

因果帰納の二段階理論によると、人間は因果帰納の最初の段階において、無数の原因候補事象の中から注目する結果事象と関係がありそうな事象の集合を抽出する。本研究では、この段階における事象の抽出が共変動検出ではなく稀少性仮定の下での非独立性の検出によって行われると仮定した上で、各モデルの振る舞いと合理性を分析する。

稀少性仮定の下での非独立性の尺度を以下の手順にしたがって定義する。原因候補事象 C と結果事象 E が独立であることの確率論上の定義は以下である。

$$P(C|E)P(E|C) - P(C)P(E) = 0 \quad (4.1)$$

$N(\Omega) = a + b + c + d$ とおき、上記の式を頻度で表すと次式のようなになる。

$$\frac{N(C, E)}{N(E)} \frac{N(C, E)}{N(C)} - \frac{N(C)}{N(\Omega)} \frac{N(E)}{N(\Omega)} = 0 \quad (4.2)$$

この等式に極稀少性仮定と呼ばれる極限操作 [Hattori and Oaksford, 2007] を行う。等式の左辺第一項は $N(\Omega)$ に含まれる $N(\neg C, \neg E)$ が無限大に発散するために 0 に収束するが、第一項は内部に $N(\neg C, \neg E)$ を含まないため、極限操作による影響を受けずに残る。

$$\lim_{N(\neg C, \neg E) \rightarrow \infty} \frac{N(C, E)}{N(E)} \frac{N(C, E)}{N(C)} = 0 \quad (4.3)$$

これを極稀少性仮定下での非独立性の尺度とみなし、本論文において以降はこれを *Dependence under Rarity (DR)* と呼ぶ。

$$DR = \frac{N(C, E)}{N(E)} \frac{N(C, E)}{N(C)} \quad (4.4)$$

DR は、 C と E が独立の場合に 0 をとり、このとき C と E の生起の仕方が完全に同じ場合には、つまり C と E は最も非独立の関係にあるとき 1 をとる。

ここで式 (1.8) および式 (4.4) より、 DR は DFH の二乗と等価である。

$$DR = DFH^2 \quad (4.5)$$

また、式 (1.13) も合わせて考えると、 DR 、 DFH 、 $pARIs$ はいずれも 2 つの条件文、「 C ならば E 」と「 E ならば C 」がともに成り立つときに高い値を取るということが分かる。

各モデルの違いは、 DR は 2 つの条件文の確率表現の積、 DFH は 2 つの条件文の幾何平均、 $pARIs$ は 2 つの条件文が同時に成り立つ確率として表現されることである。

DFH がピアソンの積率相関係数 ϕ に対して極端な稀少性仮定に基づく極限操作を行うことで導出されることはモデル導入の際に述べたことだが、ここから DR は極稀少性仮定と呼ばれる極限操作を決定係数に対して行ったものと等価であることが分かる。

$$\lim_{N(-C, -E) \rightarrow \infty} \phi^2 = P(C|E)P(E|C) \quad (4.6)$$

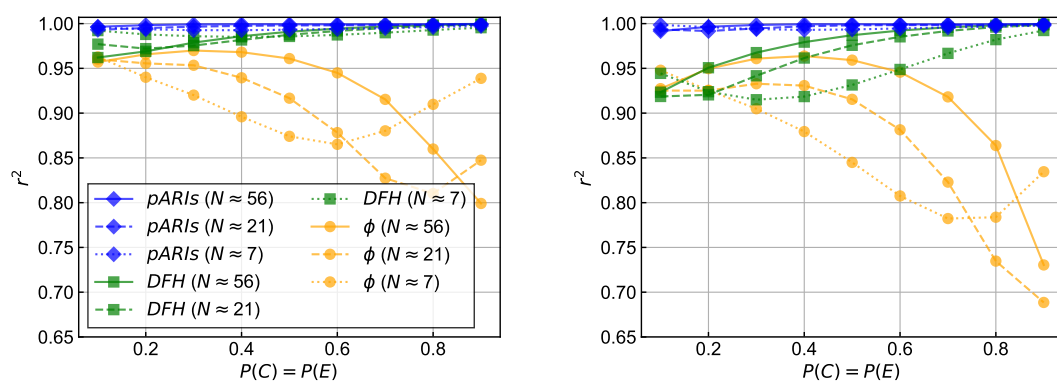
4.3 シミュレーション

手順は、母集団パラメータとして DR を用いること以外は、3 章の手順と同じであり、特筆すべき変更点がない限りは手順等の説明を省略する。また、母集団パラメータとしての DR を DR_0 と示し、推定モデルとしての DR と区別する。各シミュレーションの表記は、前章の対応するシミュレーションにプライム ($'$) を付加したものとする (例えば、前章のシミュレーション Aa に対応するものとしては、シミュレーション Aa' と表記する)。

4.3.1 シミュレーション Aa'

シミュレーション Aa' は、 ϕ , DFH , $pARIs$ が稀少性仮定下の非独立性のモデルである DR と、どの程度同様に振る舞うかをテストするものである。

シミュレーションの結果を図 4.1 に示した。この図は等確率性が仮定されたときの



(a) Hattori and Oaksford [2007] と同様に抑制的なサンプル (分割表) を除いた場合

(b) 抑制的なサンプルを含めた場合

図 4.1 等確率性仮定の下での $pARIs$ および DFH のサンプル相関係数 DR との決定係数

DFH , $pARIs$, ϕ と DR の間の決定係数を示している。図 4.1(a) は抑制的なサンプルを除いた場合の結果であり、図 4.1(b) は抑制的なサンプルを含めた場合の結果である。結

果は、 $pARIs$ は全ての場合に DR と高い適合を示すが、 DFH は稀少性の下であり DR と適合しないことを示した。また、抑制的なサンプルを含めた場合、 DFH や ϕ は適合が低下するのに対し、 $pARIs$ はいずれの場合も依然として高い適合を示す。

4.3.2 シミュレーション Ab/

このシミュレーションは各モデルと母集団 DR_0 との適合を確認する。

図 4.2 に、各モデルの DR_0 に対する決定係数を示す。図 4.2(a) が生成的な分割表のみを使用した場合の結果であり、図 4.2(b) が抑制的なものも含めた場合の結果である。シミュレーション Aa/ で示したように、 $pARIs$ は全ての場合において DR と殆ど同様の母集団推定性能を有することがわかる。まず、抑制的な分割表を除いた場合(図 4.2(a)) に注目する。 N -サンプリングにおいて強い稀少性仮定の下で(すなわち $P(C) = P(E) \lesssim .3$ において) DFH は $pARIs$ や DR と概ね同等の性能を見せるが、 $P(C) = P(E)$ が .5 付近であるときにはやや劣る。また、 N_W -サンプリングにおいて、稀少性仮定の下であっても DFH は $pARIs$, DR と比べるとやや劣る結果となっている。

続いて、抑制的な分割表を含めた場合(図 4.2(b)) に注目すると、 N -サンプリングと N_W -サンプリングの双方において DFH は $pARIs$, DR よりも低い適合を示す。

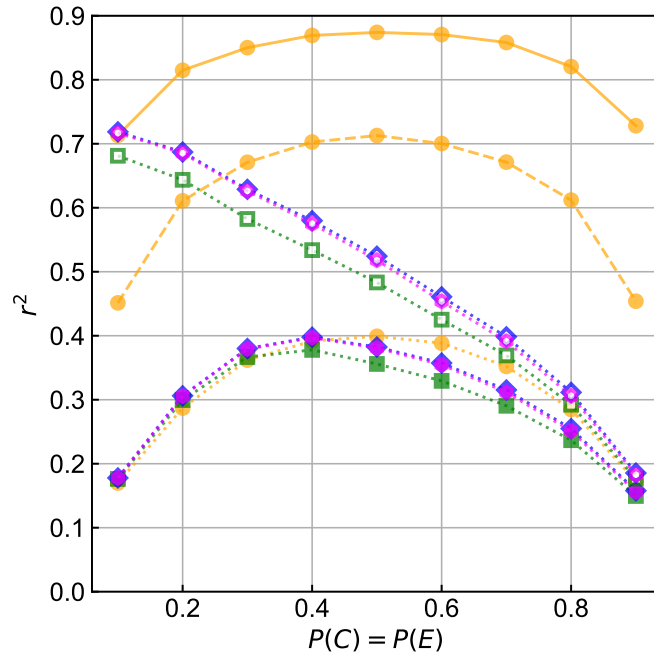
4.3.3 シミュレーション Ac/

母集団推定性能における等確率性の影響を調べるためにシミュレーション Ac/ を行った。

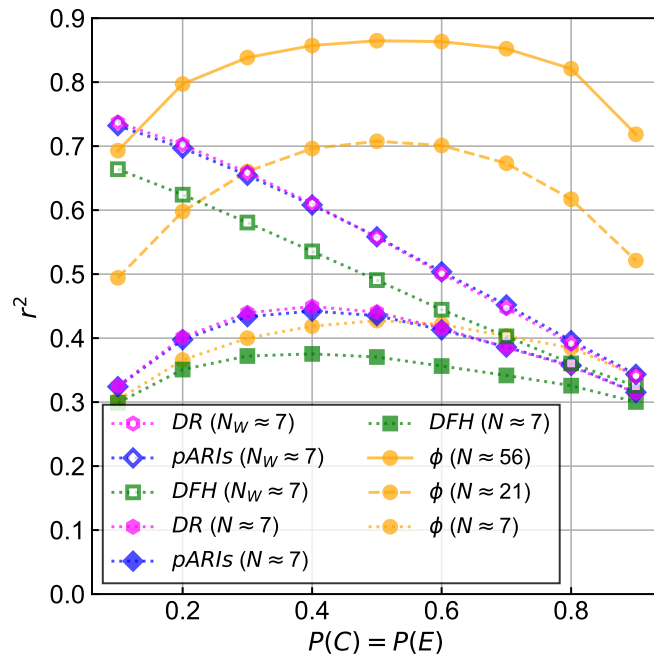
図 4.3 は、スモールサンプル ($N_W \approx 7$) かつ稀少性下 ($P(C), P(E) \geq .5$) を仮定したときの、各モデルと ϕ_0 との決定係数を示す。なお、抑制的なサンプルも含めた場合の結果のみを示した。縦軸は DR_0 との決定係数、横軸は原因と結果の確率の差 $P(C) - P(E)$ であり、この差が 0 のときに 2 つの事象は等確率である。原因と結果の平均確率 (m) が等しいものは線で結ばれている。このシミュレーションの結果は、 $pARIs$, DR , DFH は全て、等確率性の下で特に優れた母集団推定性能を有すること、また、 $pARIs$ と DR は DFH よりも僅かに等確率性の下で優れていることを示している。これは、 $pARIs$ と DR が DFH よりも等確率性に対して適応的な(すなわちより最適化された)モデルであることを意味する。

4.3.4 シミュレーション Bb/

3.2 節で述べた二つの問題点について、その重要性を定量的に検証するためにシミュレーション B/ を行った。シミュレーション Ba は、シミュレーション A において生成し



(a) Hattori and Oaksford [2007] と同様に抑制的なサンプル (分割表) を除いた場合



(b) 抑制的なサンプルを含めた場合

図 4.2 等確率性仮定の下での $pARIs$, DFH , DR の母集団相関係数 DR_0 との決定係数

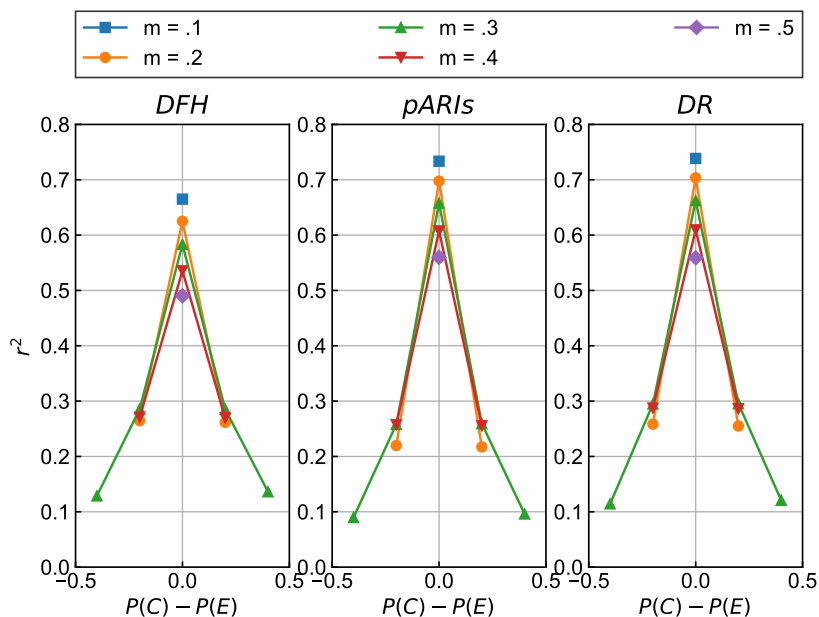


図 4.3 等確率性からの乖離度合いとモデルの母集団推定性能

た分割表について抑制的な分割表の割合を計算するものであった。シミュレーション A またはシミュレーション A' における分割表の生成過程は $P(C)$, $P(E)$ と式 (3.2) に基づいてサンプリングされる $P(C, E)$ のみに依存し、推定対象とするパラメータ (ϕ_0 または DR_0) には影響されないため、このシミュレーションについては割愛した。

シミュレーション Bb' では、観測した事象の数が一つずつ増加してゆく連続的なサンプリングにおいて、各モデルが定義される割合がどのように遷移するかを示した。なお、母集団パラメータである DR_0 は稀少性仮定を前提とした非独立性の尺度であるので、ここでは $P(C) = P(E) = .2$ の場合のみを実施した。

シミュレーションの結果を図 4.4 に示した。DR と DFH の定義式から明らかなように、DR は DFH の二乗であり、両モデルの定義可能性は完全に一致している。したがって、pARIs は DR_0 と極めて高い適合を見せる上で、DR よりも定義可能性において優れていることが示された。

4.3.5 シミュレーション C'

このシミュレーションでは、母集団の DR_0 を変動させた上で、分割表における事象の数が一つずつ増加してゆく連続的なサンプリングにおいて、各モデルがとる値の平均値と標準偏差がどのように推移するかを示した。

シミュレーションの結果を図 4.5 に示す。左の 3 つの図が各モデルを 10,000 回の試行

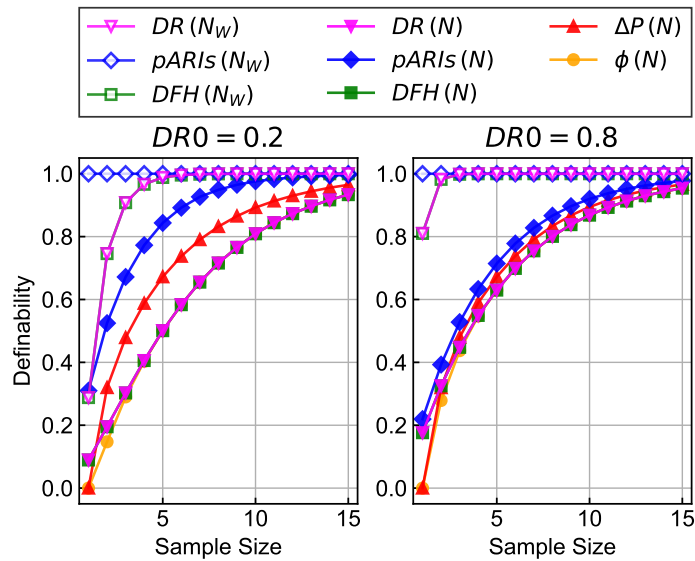


図 4.4 モデルが定義される割合の時間発展

について平均したモデル値の推移であり、右の3つが標準偏差の推移である。それぞれ上から順に $DR_0 = \{.2, .5, .8\}$ の場合の結果である。また、 $N = \{7, 21, 56\}$ の三段階において、 DR_0 を $[.0, 1.]$ の範囲で $.1$ 刻みで変化させたときに各モデルの平均値を図 4.6 に示す。黒い破線は DR_0 の値を示しており、これに近いほど DR_0 の値を忠実に近似することを意味する。

平均値の推移について、 DR は DFH と同様に最初に 1 をとり、そこから次第に収束に向かうことがわかる。他方、SD の遷移については、 DR は全体的に $pARIs$ と DFH の中間に位置する傾向にあることが読み取れる。

また、 ϕ_0 に対する各モデルの平均値 (図 4.6) を見ると、各モデルが有するバイアスが確認できる。結果は、 DR はサンプルサイズが増加するに従って DR_0 と完全に一致していくこと、また、スモールサンプルにおいて、 $pARIs$ は母集団のモデルそのものである DR よりもバイアスが小さくなることを示している。また、 DFH は常に最も上に偏った値を取る。

4.3.6 シミュレーション D'

シミュレーション D' では、モデルの二値分類性能を定量的に示した。シミュレーション D においては、稀少性 ($P(C) = P(E) = .2$) の下で ϕ_0 の閾値を変動させたときの PR-AUC (図 3.9(a)) と、 ϕ_0 の閾値を $.5$ とした上で $P(C) = P(E)$ を変動させたときの ROC-AUC (図 3.9(b)) の両方を計算した。本シミュレーションにおいては、母集団パラ

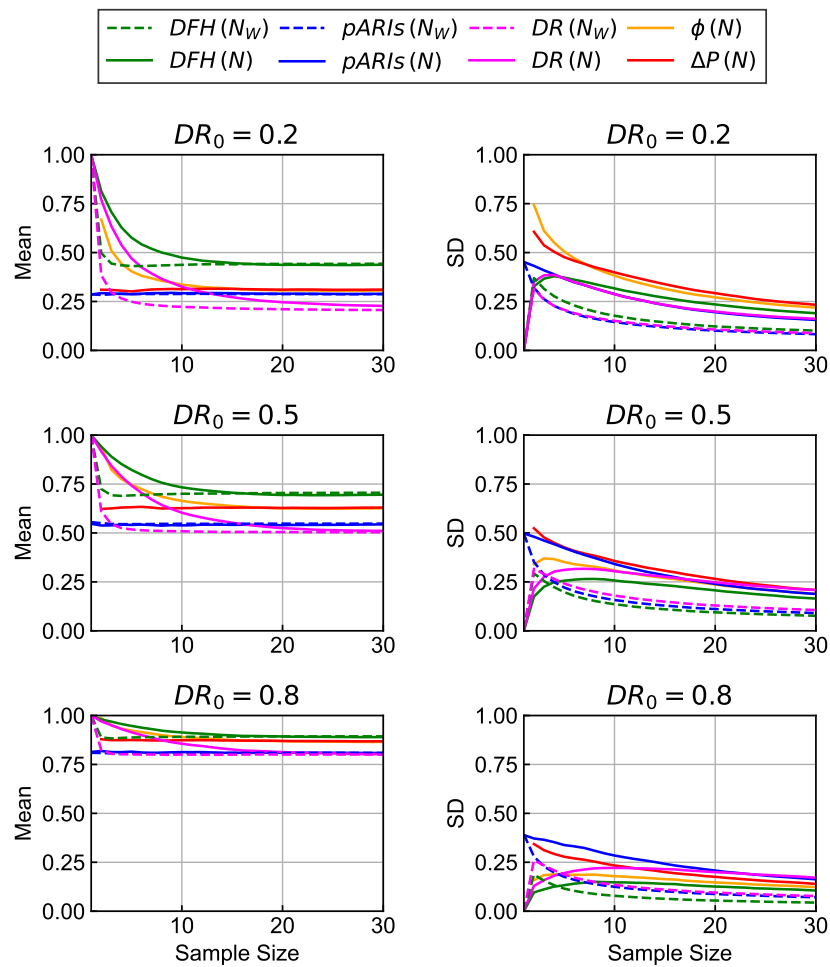
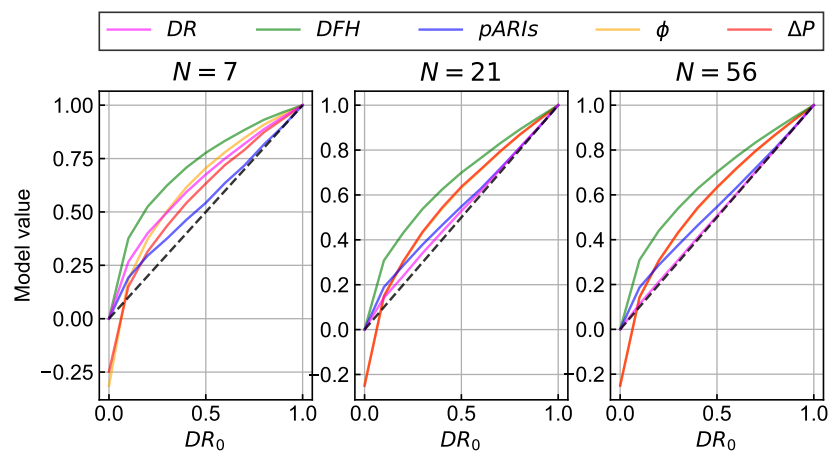


図 4.5 モデルの平均値と標準偏差 (SD) の時間発展

図 4.6 DR_0 の関数としてのモデル平均値

メータである DR_0 は稀少性仮定を前提とした非独立性の尺度であるので、前者の場合のみを実施した。

シミュレーションの結果を図 4.7 に示した。各サンプルサイズに対する分類性能を確認するため、グラフの縦軸は AUC スコア、横軸はサンプルサイズの平均値とした。

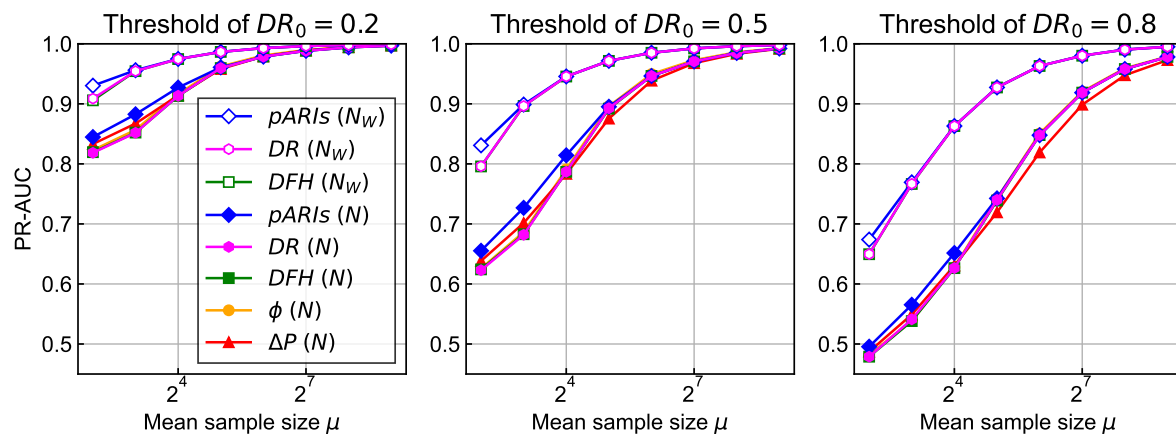


図 4.7 $P(C) = P(E) = .2$ としたときの、各 ϕ_0 の閾値における RP-AUC

図 4.7 は稀少性仮定に基づいて $P(C) = P(E) = .2$ と固定した場合に、各 ϕ_0 の閾値に対する分類性能の評価を示す。なお、完全にランダムな分類をするモデルがとる PR-AUC は、サンプルサイズに依らず、 DR_0 の閾値が $\{.2, .5, .8\}$ のとき、それぞれ、 $\{.8, .5, .2\}$ となる。したがって、全てのモデルがこのベースラインよりも高い AUC を示すことがわかる。また、結果は、 $pARIs$ が特にスモールサンプルにおいて、 DR_0 の閾値に依らず、母集団のモデル自体である DR を含む他のモデルよりも分類モデルとして優れていることを示している。この結果から、非独立性の有無に関する二値分類の文脈においても、稀少性が成り立ち、かつスモールサンプルの場合に $pARIs$ が優れた性能を有することが明らかになった。

4.4 総合議論

4.4.1 モデル値のデフォルト

シミュレーション Bb1 ではモデルの定義可能性についての分析を行い、他のモデルと比べて、 $pARIs$ がより多くの場合に定義されることを示した。この結果は全てのモデル間の性能比較を行うことを趣旨としたシミュレーションの結果に影響を及ぼす可能性がある。シミュレーション A1 では、分析に用いるサンプルをモデル間で統一するために、定義されないモデルが存在するような分割表がサンプリングされた場合にはそれを除外して再サンプリングを行っているが、モデル間の性能比較という観点においてこれは公正では

ない。例えば $pARIs$ は定義されているのにも関わらず、 ϕ が定義されないという理由によってある分割表が棄却されるということが起こりうる。シミュレーション D1において、我々はモデルが定義されない場合にはモデルはランダムに二値分類を行うと仮定することでモデルの定義可能性の問題を考慮した。これを解決するもうひとつ方法として考えられるのは、モデルが定義されない場合のデフォルトの値を考慮することである。このデフォルト値の問題は、次節で述べる因果探索 (ベイジアンネットワークの構造学習) とも深い関連を持つ。例えばモデルが定義されない場合に、モデルのデフォルト値を 1 とするならば、2 事象の間には関係がある (あるいは、「ないとはいえない」と推定することになり、0 とするならばそれらが無関係であると推定することになる。どちらのデフォルト値が妥当であるかという判断は、原因候補の抽出が行われる方法によって異なる。例えばデフォルト値が 1 であるということは予め全ての事象に関係性を認め、その後の観察によって関係性がないと判断されたものから除外していくという引き算的過程を想定しており、一方、デフォルト値が 0 であるということは、その逆で、予めすべての事象は無関係であると考え、その後の観察によって関係性があると判断された 2 事象を選定していく足し算的過程を想定していると考えられる。前者はいわば判断の保留であり、実際は関係があるのに無関係だと判断される Type I error/False positive (表 3.3) を最小化することが期待される。後者はその逆で、実際は関係がないのに関係があると判断される Type II error/False negative (表 3.3) を最小化することが期待される。真の原因の検出という目的をより確実に達成するためには、より見逃しの少ない前者の方が優れていると考えられるが、計算コストという観点においては後者の方が優れていると考えられる。

4.4.2 ベイジアンネットワークへの応用

人間が行動する環境の因果的構造は、無数の変数を持つ複雑な因果ネットワークであると想定されるが、本研究で用いた単純因果帰納の枠組みは、実験的な研究のしやすさのために極度に単純化されている。人間が因果関係の知識を獲得してそれを活用して扱う実際のタスクにおいては、計算・記憶リソースの制約や時間的制約 (即応性の要求) など、より現実的な合理性が求められる。ベイジアンネットワーク [Pearl, 2000] という確率推論モデルは複数の事象の間の因果的な関係の強さと構造をグラフ構造と条件付き確率表を用いて表現する。そのグラフの構造学習の 1 つに厳密解探索法と呼ばれる手法が存在するが、この手法による構造推定は NP 困難であり、ノード数の増加に伴い計算コストが膨大になる問題がある。そこで、より実用的な手法として条件付き独立性検定を用いた構造学習が用いられる。この手法の最も代表的な PC アルゴリズム [Spirtes and Glymour, 1991] は、まず無向完全グラフを仮定し、その後無関係な辺を除外していく。これは前節で述べた、デフォルト値を 1 とした引き算的過程に相当していると考えられる。無向完

全グラフを仮定した後、PC アルゴリズムは、条件付き独立性検定のフェーズにおいて無向グラフを作成したのち、オリエンテーションフェーズにおいて辺の方向を決定づける。(各フェーズは二段階仮説におけるヒューリスティック段階と介入段階と意味するところが似ているようであり、正確な対応関係については更なる考察が必要である。)

ここで、シミュレーション D1 で行われたモデルの二値分類 (すなわち、2 事象が非独立であるか否か) が、ベイジアンネットワークの構造学習におけるエッジの有無 (2 事象が条件付き独立か否か) の判断に対応していると考え、シミュレーション D1 の結果は、特にサンプルが小さいときに $pARIs$ がベイジアンネットワークの独立性検定において有効な近似となる可能性を示唆している。また、条件付き独立性検定のフェーズで問われるのは 2 事象が独立か否かであり、その方向性は考慮されない。 $pARIs$ は ϕ や DFH と同様に関係の方向性を考慮しない、無向なモデルであり、 ΔP のような有向なモデルよりも、独立関係の判定に適している。^{*1} 以上の点から、PC アルゴリズムにおける条件付き独立性の検定の代わりに、人間の因果帰納と高い適合を示し、稀少性仮定の下での日独立性の尺度の優れた近似でもある $pARIs$ を用いることで、より儉約的かつ人間的な構造学習が実現できる可能性がある。

4.5 結論

本研究では、人間の因果帰納の二段階理論における第一段階の計算論的な目標として、従来は共変動検出と見なされていたものを稀少性仮定の下での非独立性の検出であると見なした上で、観察的因果帰納モデル $pARIs$ について適応的合理性の観点から分析を行った。結果から、 $pARIs$ が従来のモデルよりも多様なサンプルに対して、稀少性仮定の下での非独立性の優れた近似として振る舞うことを示した。人間の因果的判断と高い適合を示すことが明らかになっている $pARIs$ がこのように振る舞う事実は、人間の因果帰納の最初のプロセスが共変動の検出よりも非独立性の検出に基づいている可能性を示唆する。

^{*1} モデルが無向であるということは C と E を入れ替えてもモデルがとる値が変わらないことを意味する。ここで、 C と E の交換は 2×2 分割表においては b と c の交換を意味するが (表 1.1 参照)、これらのモデルは b と c を交換してもその定義上その値を変えない。 ΔP のように有向なモデルは、常に両方向に対して検定を行わなければならない、事前に関係がありそうな事象を抽出をするという目的のために用いるには計算コストが大きく効率が悪いと考えられる。

第 5 章

因果探索アルゴリズムへの応用

5.1 ベイジアン・ネットワーク

ベイジアン・ネットワークは事象間の因果関係を、有向非巡回グラフ (DAG) と条件付き確率分布を用いて表現したグラフィカルモデルである。本研究は観測データからネットワークの構造を学習するプロセスに関するものであり、特に制約ベース法と呼ばれるアプローチに焦点を当てる。制約ベース法は事象間の共変関係に基づいて因果構造を学習するアプローチであり、2 変数の独立性の検定を行う段階と因果関係の方向付けを行う段階からなる。

5.2 条件付き独立性検定

条件付き独立性の検定に際して、一般には、条件付き独立であることを帰無仮説とした古典的な統計的検定が用いられる。これに用いられる代表的な統計量には χ^2 統計量や G^2 統計量がある。帰無仮説の元での期待頻度を Expected, 実際に観測した頻度を Observed とすると以下のように定義される。

$$\chi^2 = \sum_{x,y,z} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \quad (5.1)$$

$$G^2 = 2 \sum_{x,y,z} \text{Observed} \cdot \log \frac{\text{Observed}}{\text{Expected}} \quad (5.2)$$

G^2 統計量は尤度比 χ^2 統計量とも呼ばれる。 N を サンプルサイズとすると、 G^2 統計量は $1 - pARIs$ の情報量的解釈である相互情報量 (MI) と以下のような線形関係にある。

$$G^2 = 2 \cdot N \cdot \text{MI}(X, Y) \quad (5.3)$$

第4章において、単純因果帰納の枠組みにおいて稀少性を仮定したとき、 $pARIs$ は非独立性の優れた近似として振る舞うことが示された。 $pARIs$ を条件付き独立性の指標として拡張するうえで、条件となる変数集合 Z を所与としたときの二変数 X, Y に対する $pARIs$ の値を以下のように定義する。

$$pARIs_Z = \min_z \frac{P(x, y|z)}{P(x|z) + P(y|z) - P(x, y|z)} \quad (5.4)$$

条件付き独立性の判定に際して、任意の閾値 ϵ を用いて、 $pARIs < \epsilon$ のとき2変数を条件付き独立と見なすこととする。

5.3 PC アルゴリズム

古典的な条件付き独立性検定に基づく最も代表的な構造学習アルゴリズムが PC アルゴリズム [Spirtes and Glymour, 1991] である。このアルゴリズムは、条件付き独立性検定の計算量がノード数に対して指数的に増大する問題を抱えていた IC アルゴリズム [Verma and Pearl, 1990] を発展させたものである。PC アルゴリズムの手続きをアルゴリズム 5.3, 5.3 に示す。PC アルゴリズムは、まず、すべてのノードを無向辺で結んだ完全グラフから出発し、変数間の条件付き独立性の検定を行い、有意に独立な事象間のエッジを取り除いていく。条件部は空集合より始まり (これを0次の条件付き検定と呼ぶ)、1次、2次と条件付ける変数の数を順次増やしながら行う。その後は有向非巡回グラフ (DAG) の制約にしたがい、可能なエッジに対して方向づけを行う。

5.4 シミュレーション

PC アルゴリズムにおける条件付き独立性の検定の際に $pARIs$ を使用することの有効性をシミュレーションを介して検証した。シミュレーションは、 χ^2 統計量、条件付き $pARIs$ 、条件付けを行わない $pARIs$ (これは0次の条件付けを行うことに等しい。以降 $pARIs_0$ と称する) の3種類に対して行った。なお、 $pARIs_0$ を比較対象に加えたのは、単純因果帰納の枠組みにおいて人間の観察的因果帰納と高い適合を示す $pARIs$ が複数の事象間の因果関係を扱う因果探索の枠組みでどの程度有効なのかを検証するためである。以下にシミュレーションの手順と結果を示す。

5.4.1 方法

提案した指標の精度と性質の評価を、人工データ生成とパラメータ・リカバリーの確認の組からなる計算機シミュレーションによって行った。データの生成には、図 5.1 に示す

Algorithm 5 PC algorithm (条件付き独立性検定の段階)**Require:** N ノード変数, データ X **Ensure:** ネットワーク構造 \mathcal{G} ノード集合 $C \leftarrow \emptyset$ 初期グラフ $\mathcal{G} \leftarrow$ 無向完全グラフ K_N **while** $t < N - 2$ **do** $t \leftarrow 0$ **for** 各ノード X_i **do****for** X_i 以外の各ノード X_j **do****for** X_i, X_j の共通隣接ノード集合中の t 個のノードの組み合わせ **do** $C \leftarrow$ 上記ノードの組み合わせ**if** $CI(X_i, X_j|C) = 1$ **then** \mathcal{G} からリンク (X_i, X_j) を削除: ただし, $CI(X_i, X_j|C)$ は, C を所与として X_i と X_j が条件付き独立のとき 1 となりそれ以外で 0 を返す条件付き独立テスト関数**end if****end for****end for****end for** $t \leftarrow t + 1$ **end while****Algorithm 6** PC algorithm (方向づけの段階)**for** \mathcal{G} 中の $X - Y - Z$ を形成する全ての 3 ノード X, Y, Z **do****if** X と Z が隣接せず, かつ $CI(X, Z|Y) = 0$ **then** $X \rightarrow Y \leftarrow Z$ **end if****end for****while** 方向づけされていない辺がない **do****if** $X \rightarrow Y - Z$ かつ X と Z が隣接していない **then** $X \rightarrow Z \leftarrow Y$ **end if****if** $X \rightarrow Y \rightarrow Z$ かつ X と Z が隣接している **then** $X \rightarrow Z$ **end if****if** $X \rightarrow Y \leftarrow Z$ かつ X と Z が隣接していない, その上, $X - W - Z$ かつ $W - Y$ が存在する **then** $W \leftarrow Y$ **end if****end while**

ように, 条件付き独立性を考慮する上で重要な 3 種類の構造 (連鎖, 分岐, 合流) のすべてを含む, 最もシンプルな因果グラフを使用した。

各変数は $pARIs$ による計算を可能とするために生起不生起の値を取る 2 値変数とし, 分析を平易にするために生起確率と条件付き確率は全変数で等しいものとし, 大中小の三段階で変化させた。具体的には, 各変数の生起確率を $\{.2, .5, .8\}$ と変化させ, 各変数の条件付き確率は単純因果帰納の枠組みにおける因果性の規範的指標である

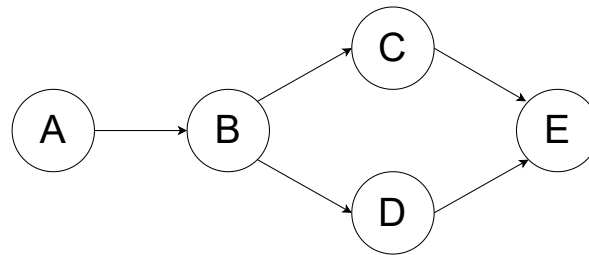


図 5.1 データ生成に使用した因果グラフ

$\Delta P = P(E|C) - P(E|\neg C)$ が $\{.2, .5, .8\}$ の値をとるように変化させた。上記のように設定した母集団の因果モデルから、各サンプルサイズ $N = \{10, 50, 100, 500\}$ の観測データを生成し、それに対して各アルゴリズムによる構造学習を行った。このとき、独立性検定の精度に焦点を当てるためにエッジの方向付けは行わず、無向辺の構造学習を行った。なお、 χ^2 統計量による検定の有意水準には $\alpha = .01$ を用い、 $pARIs$ および $pARIs_0$ の閾値は、それぞれ全体を通して安定していた $\epsilon = .1$ と $\epsilon_0 = .5$ を用いた。アルゴリズムによる構造学習の精度は、母集団である図 5.1 と等価にするために必要な最小のエッジの変更点の数であるエッジ編集距離を用いて評価した。これは値が小さいほど精度が高いことを意味する。結果は上記の試行を 300 回繰り返した平均距離である。

5.4.2 結果と議論

結果を図 5.2 に示した。ヒートマップは平均エッジ編集距離を表し、横軸は設定した ΔP の値を、縦軸は各事象の生起確率を表している。結果から、サンプルサイズが小さい場合に $pARIs$ および $pARIs_0$ は χ^2 統計量よりも母集団とのエッジ編集距離が近くなることが読み取れる。また、その傾向は $N = 10, 50$ の場合に顕著であり、十分なサンプルが揃う $N = 100, 500$ の場合に逆転する。 $N = 10$ における χ^2 統計量の結果は、統計的な観点から独立性を判断するには十分なサンプルが得られていないことを示しているとも解釈でき、一概に $pARIs$ のほうが優れていると結論づけることはできない。しかし、結果は少なくとも $pARIs$ による条件付き独立性の検定が特にスモールサンプルからの推論において有用でありうることを示している。また、十分なサンプルが得られた場合にも、特に稀少性のもとで平均エッジ編集距離 0.2 付近の精度で母集団を推定できることを示している。他方、エッジの有無の判断へ単純因果帰納の枠組みをそのまま適用した $pARIs_0$ は、稀少性のもとで、スモールサンプルかつ因果強度が大きい場合には条件付きの $pARIs$ と同等以上の振る舞いを示すことが読み取れる。このことは、単純かつ人間の因果的直感と高い適合を示す指標が、複数の事象間の因果探索という枠組みにおいても一定の有効性を持つことを示している。

今回のシミュレーションにおいて、 $pARIs$ と $pARIs_0$ の閾値を全体を通して安定した値に固定したが、実際は各確率分布に合わせて最適なものを個別に設定することも可能である。閾値を変化させた場合の挙動の変化や優れた閾値の候補の計算方法など、閾値に関するより詳細な分析が必要である。また、事象ごとに生起確率や条件付き確率が異なるような因果グラフにおける性能の分析についても本研究では未検討であるが、重要な課題である。

5.5 結論

本論文では、ベイジアン・ネットワークの構造学習の際に行われる条件付き独立性検定を、人間の因果的直感と高い相関を示すことが明らかになっている指標であり、稀少性仮定のもとでは非独立性の尺度に対する優れた近似として振る舞うことが明らかになっている $pARIs$ で代替するアルゴリズムを提案し、計算機シミュレーションに基づく性能分析を行った。結果から、提案アルゴリズムは条件付き独立性の検定として一定の有効性を持ち、特にスモールサンプルにおいて有用である可能性が示された。

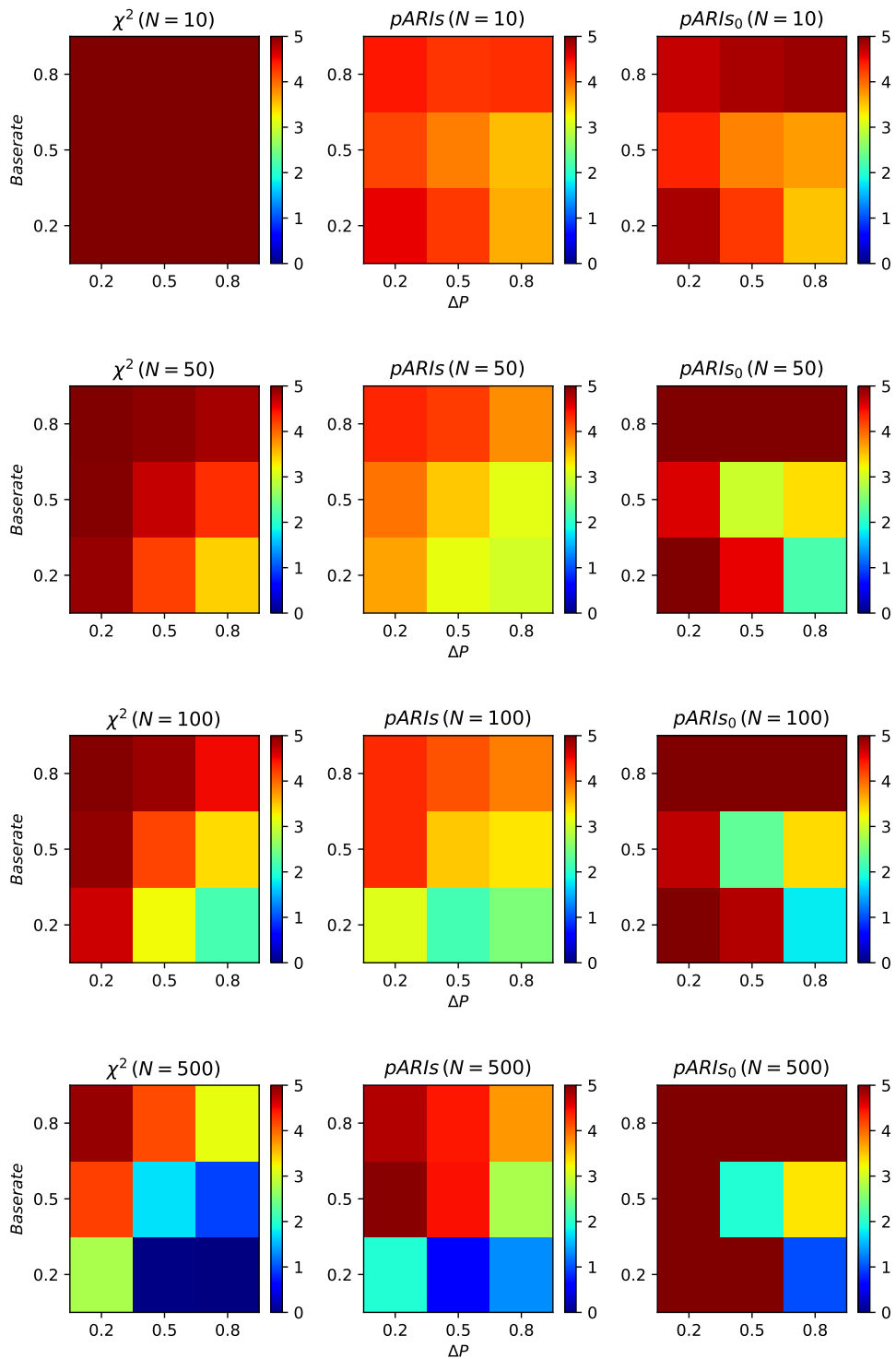


図 5.2 各アルゴリズムを用いて学習したグラフと正答グラフの間のエッジ編集距離。 χ^2 検定では有意水準 $\alpha = .01$ を, $pARIs$ および $pARIs_0$ の閾値は, それぞれ全体を通して安定していた $\epsilon = .1$ と $\epsilon_0 = .5$ を用いた。

第6章

結論

本研究では、Hattori and Oaksford [2007] が行ったメタ分析とシミュレーションの設定に問題があることを指摘した。これらの問題を可能な限り修正し、あるいは必要なシミュレーションを追加した上で、比較対象に $pARIs$ を追加して再度シミュレーションを行った。メタ分析の結果は、 $pARIs$ は DFH よりも高い記述性能を有することを示した。また、シミュレーションの結果は、 $pARIs$ が DFH よりも優れた母集団推定性能を有する上に、経済性の観点からも優れていることを示した。得られた結果から、Hattori and Oaksford [2007] の提唱する二段階仮説では、 $pARIs$ が DFH よりも適切なモデルであることが示された。また、これらのシミュレーションの設定を変更することで、元の研究の結果や考察に影響を与えることを示しました。

また、因果帰納モデルの計算上の目標を、非独立性の検出であると仮定しなおして再度分析を行った。その後、稀少性仮定の下での非独立性の尺度 DR を母集団パラメータとして設定し、再度コンピュータ・シミュレーションによる分析を行った。結果は、 $pARIs$ が稀少性仮定の下での非独立性の指標である DR と非常に似た挙動を示し、また、母集団パラメータの優れた推定性能を有することを示した。これらの結果は人間の因果帰納が非独立性の検出に基づいている可能性を示唆すると同時に、 $pARIs$ の適用範囲を「単一事象間の因果関係の強さ」から「複数事象間の因果関係の構造」の問題へと拡張するための足がかりとなった。

最後に、因果ベイジアンネットワークにおける構造推定のための新たなアルゴリズムを開発し、性能の検証を行った。結果は、単純因果帰納における $pARIs$ の性質と同様に、稀少性仮定のもとで、また、サンプルサイズが小さい場合に有効であることを示した。このアルゴリズムによる因果構造の推定が人間の構造推定とどの程度適合するかという点に関して、認知実験に基づく検証が今後の課題である。

謝辞

本研究を進めるにあたり、ご協力いただいた皆様に深く感謝いたします。特に、高橋達二教授には、長くご指導いただき、論文執筆にとどまらない多岐にわたるご助言をいただきました。心より感謝申し上げます。

参考文献

- J. R. Anderson. *The adaptive character of thought*. Erlbaum, Hillsdale, NJ, 1990.
- J. R. Anderson and C. F. Sheu. Causal inferences as perceptual judgments. *Memory and Cognition*, 23(4):510–524, 1995. ISSN 0090502X. doi: 10.3758/BF03197251.
- A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997. ISSN 0031-3203. doi: [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2). URL <https://www.sciencedirect.com/science/article/pii/S0031320396001422>.
- M. J. Buehner, P. W. Cheng, and D. Clifford. From Covariation to Causation: A Test of the Assumption of Causal Power. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(6):1119–1140, 2003. ISSN 02787393. doi: 10.1037/0278-7393.29.6.1119.
- P. W. Cheng. From Covariation to Causation: A Causal Power Theory. *Psychological Review*, 104(2):367–405, 1997. ISSN 0033295X. doi: 10.1037/0033-295X.104.2.367.
- D. J. Collins and D. R. Shanks. Short article: Conformity to the power pc theory of causal induction depends on the type of probe question. *Quarterly Journal of Experimental Psychology*, 59(2):225–232, 2006. doi: 10.1080/17470210500370457. URL <https://doi.org/10.1080/17470210500370457>. PMID: 16618631.
- N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001. ISSN 0140525X. doi: 10.1017/S0140525X01003922.
- T. L. Griffiths and J. B. Tenenbaum. Elemental Causal Induction. *Cognitive Psychology*, *(617):**, 2005.
- M. Hattori and Y. Nishida. Why does the base rate appear to be ignored? the equiprobability hypothesis. *Psychon. Bull. Rev.*, 16(6):1065–1070, Dec. 2009.
- M. Hattori and M. Oaksford. Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive*

- Science*, 31(5):765–814, 2007. ISSN 03640213. doi: 10.1080/03640210701530755.
- H. M. Jenkins and W. C. Ward. Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79(1):1–17, 1965. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.
- D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1914185>.
- S. Kaufmann. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38(1):1–53, 2009. doi: 10.1007/s10992-008-9088-0.
- K. Lober and D. R. Shanks. Is Causal Induction Based on Causal Power? Critique of Cheng (1997). *Psychological Review*, 107(1):195–212, 2000. ISSN 0033295X. doi: 10.1037/0033-295X.107.1.195.
- G. A. Miller. The magical number seven plus or minus two: some limits on our capacity for processing information. Technical report, 1956.
- D. J. Navarro and A. F. Perfors. Hypothesis Generation, Sparse Categories, and the Positive Test Strategy. *Psychological Review*, 118(1):120–134, 2011. ISSN 0033295X. doi: 10.1037/a0021110.
- M. Oaksford and N. Chater. A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101(4):608–631, 1994. ISSN 0033295X. doi: 10.1037/0033-295X.101.4.608.
- J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
- J. C. Perales and D. R. Shanks. Normative and descriptive accounts of the influence of power and contingency on causal judgement. *The Quarterly Journal of Experimental Psychology Section A*, 56(6):977–1007, 2003. doi: 10.1080/02724980244000738. URL <https://doi.org/10.1080/02724980244000738>.
- J. C. Perales and D. R. Shanks. Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin and Review*, 14(4):577–596, 2007. ISSN 10699384. doi: 10.3758/BF03196807.
- T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3):1–21, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0118432.
- G. Sanfilippo, N. Pfeifer, D. E. Over, and A. Gilio. Probabilistic inferences from conjoined to iterated conditionals. *International Journal of Approximate Reasoning*, 93:103–118, 2018. ISSN

- 0888-613X. doi: <https://doi.org/10.1016/j.ijar.2017.10.027>. URL <https://www.sciencedirect.com/science/article/pii/S0888613X17306643>.
- P. Spirtes and C. Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review - SOC SCI COMPUT REV*, 9:62–72, 04 1991. doi: 10.1177/089443939100900106.
- T. Takahashi, Y. Kohno, and K. Oyo. Causal induction heuristics as proportion of assumed-to-be rare Instances (pARIs). In *Proceedings of the 7th International Conference on Cognitive Science (ICCS2010)*, pages 361–362, 2010.
- T. Verma and J. Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, UAI '90*, page 255–270, USA, 1990. Elsevier Science Inc. ISBN 0444892648.
- P. A. White. Making Causal Judgments from the Proportion of Confirming Instances: The pCI Rule. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(4):710–727, 2003. ISSN 02787393. doi: 10.1037/0278-7393.29.4.710.