

目次

第 1 章	統計の誕生	4
1.1	オランダ：英国政治算術とフランス確率論の衝突	4
1.1.1	英国政治算術	4
1.1.2	伊・仏の確率論	4
1.2	オランダでの人口統計と確率論	4
1.3	フランスでの最小二乗法の誕生	5
1.3.1	天文学の大衆化	5
1.3.2	Lagrange と確率変数の和	5
1.3.3	Bernoulli と最尤法	5
1.3.4	Legendre と最小二乗法	5
1.4	Quetelet による数理科学	6
1.4.1	道徳統計学	6
1.4.2	コントの社会学	6
1.4.3	コンドルセの偉大な総合	6
1.5	ドイツでの社会統計学	6
1.6	現代誤差論	6
1.6.1	田口玄一	6
第 2 章	Ronald Fisher と農事試験, Pearson と生物測定学	7
2.1	概観	7
2.1.1	3つの学派	7
2.1.2	Fisher の業績	7
2.2	Galton までの優生学	8
2.2.1	Mendel の遺伝理論	8
2.3	最尤法	8
2.4	実験計画法	9
2.4.1	農事実験	9
2.4.2	分散分析と有意性検定	9
2.5	有意性検定	10
2.5.1	Fisher 以前の検定論	10
2.5.2	Fisher の有意性検定	11
2.6	Pearson の統計理論	11
2.6.1	記述統計学	11
2.6.2	生物測定学	11
2.6.3	モーメント法とピアソン系	11
第 3 章	Neyman-Pearson 理論	12
3.1	統計的品質管理の父	12
3.2	Neyman-Pearson の統計的仮説検定理論	12

第 4 章	前期計算機時代の統計手法	14
第 5 章	機械学習と統計科学	15
5.1	機械学習の歴史	15
5.2	類型認識問題の枠組み	15
5.2.1	決定理論からの概観	15
5.2.2	学習アルゴリズム	16
5.2.3	決定空間が有限な場合	17
5.3	教師なし学習の枠組み	17
第 6 章	ビッグデータと知識発見	18
6.1	歴史	18
6.2	形態素解析	18
第 7 章	統計的因果推論	19
7.1	歴史と発展	19
7.1.1	総覧	19
7.1.2	Pearson と Fisher	21
7.1.3	回帰模型	21
7.1.4	パス解析	22
7.1.5	操作変数法・Simon-Blalock 法	22
7.1.6	構造方程式模型	22
7.1.7	独立成分分析	23
7.1.8	因果を扱うにあたっての問題点	23
7.1.9	因果探索	23
7.2	因果推論の枠組み	23
7.2.1	Neyman-Rubin 因果模型	23
7.2.2	Pearl の因果の認知模型	24
7.2.3	Pearl の構造因果模型	24
7.2.4	構造的因果模型の問題点	24
7.2.5	因果の Bayes 確率論との比較での理解	25
第 8 章	参考文献	26
参考文献		27

豊かな統計理論の展開の歴史は、豊かな応用とともにある。

(1) Fisher の統計は Rothamsted 農業研究所におけるデータ解析問題に深く根ざしている。

(2) 1940s の Neyman-Pearson の統計理論は統計的品質管理の問題があった。

それぞれの大家をまとめる。

(1) Ronald Fisher 90-62：近視だったために夜は音で数学を学び、独自の数学への適性を身に着け、奨学金を得て Cambridge 大学で数学と物理学を学んだが、その後は第一次世界大戦もあってうやむやになる（修士はとっていないのでは？）。Pearson への反感から大学教員ではなく農事試験場のしごとを選び、36 歳にはそこで実験計画法を確立する。その後に Galton 教授職へ戻り、そこでは溜まりに溜まっていた優生学への興味を爆発させる。

(2) Karl Pearson 57-36：本当は Carl. 菊池大麓と Cambridge で学び、ドイツ留学経験も積んで（ここでマルクスに傾倒）中世ドイツ文学や法学にも打ち込んだが、最終的には数学に戻った真のアカデミアの人。Galton と仲良くなって優生学に親しみ、優生学部初代教授となる。生物測定学という全く独立の Darwin 以来の流れから、記述統計学を作る。

(3) Egon Pearson 95-80：もともとは Cambridge で天体物理学を専攻してから転向。25 年に Neyman と会い、31 年に Shewhart と会う（ここで、統計的品質管理の問題意識と出会う）。父の退官後、Fisher は優生学教授で、Egon は応用統計学部教授となる。

- (4) Jerzy Neyman 94-81：父がロシアで仕事をしていた法律家で、ウクライナの大学で Bernstein に数学を学び、祖国ポーランドへ還る。ワルシャワ大学で仕事を持ちながら、ピアソンの下で留学したときに、息子と意気投合。
- (1) Fisher と Gosset は大量生産できない農業・ビール醸造なる対象に対して、比較実験を通じて科学的結論を導く必要があった。このため、誤差の統制（分散分析）と精密分布の決定と有意性検定を作った。
- (2) Karl Pearson と Galton は大量のデータから優生学的知見を引き出したかった、どちらかといえば社会科学的営みであった。このため、誤差を最小にする fitting の技法を開発した。
- (3) Neyman と Egon Pearson は、企業意思決定に必要なサンプリング問題から発展して、契約上の品質に関する合意形成手段としての統一検定方式を開発する必要があった。そこで、決定理論の枠組みと最適性が問題になった。^{†1}

こう見ると、検定論は、最もコミュニケーションの道具、人類生態系を別の論理で飛び回るミームであり、言語の権化であるようだ。品種・肥料の有効性について有意性を下す判定が有力者の権威から統計技法の権威へと移り、また品質管理についての権威が統計技法に移った。デジタルかもその延長であるとみれる。

人々の行動指針を与えるようなミーム。人生が模倣するような権威を帯びた形式として、芸術とは異なる作品。このようなものを作るのが夢であった。これはむしろ社会彫刻のようなものであって、統計技法だけではなく、周りの言論と社会実践が肝要になってくる。端的に言えば、「立場が違う人を結ぶ形式」が好きなら、それはシミュレーションや統計の手法になるわけだ。僕はコミュニケーションのプロでなくてはならない。

^{†1} これはプログラムに関する法律などでも、採択されることになるだろう。

第 1 章

統計の誕生

人類が最初に直面した統計的問題は、次の 3 つである：

観測の誤差と確率論の問題 同一観測を繰り返しても値がばらつくが、その中でどれを「最良値」「代表値」として採用すれば良いかの問題である。答えは最小二乗法が与えるが、なぜこれが「良い」のか？ [安藤洋美, 1995] この問題設定は実は線形回帰問題と関連し、誤差分布が正規性の仮定を満たすとき（根源誤差の集まりとみなせるとき）、最小二乗推定量が最尤推定量となる。ここまでの問題に対する理論を誤差論という。

政治算術 William Petty 1623-1687 は political arithmetic を著し、労働価値説を唱えて、古典派経済学と統計学の祖と呼ばれる。

国状論

この 3 つを統合したのが、Quetelet の社会物理学である。

20 世紀のはじめに、英米の生物測定学と、ドイツの社会統計学とに分岐し、現代に至る。英国統計学は特に数学との結びつきが強く、記述論から推測論・検定論を作り、数理統計学の系譜となる。政治算術は、確率論の土壌を生かした Quetelet によって社会物理学となり、刺激を受けたドイツで国状論・社会統計学となり、計量経済学の系譜となる。

Leonard Savage 1917-1971 の『統計学の基礎』で Bayes 学派が復活し、そもそも母集団の前提がない際にも応用された。現在実社会で最も中心的なのが Bayes 学派である。

1.1 オランダ：英国政治算術とフランス確率論の衝突

イギリス経験論と大陸合理主義との二項対立がある。

英国的な統計的経験的確率と、伊・仏的な合理的先験的確率との、認識論的方法論の対立は最初からあった。「確率統計学」というときの両極端から歩み寄った。

1.1.1 英国政治算術

英国で互いに友人である 2 人、William Petty の『政治算術』(1690) と John Graunt 『死亡表に関する自然のおよび政治的諸観察』(1660) が人口統計学の源流となった。

1.1.2 伊・仏の確率論

ガリレイに源流を持ちつつ、Pascal と Fermat が完成させたのがこの確率論的な見方である。

1.2 オランダでの人口統計と確率論

ちょうど 2 つの接点に居たのがオランダで、ここで真に豊かな理論が生まれた。

Huygens の『チャンスの価格』は、完全に Pascal の「勝負の値」の系譜にある。実はデカルトの祖国以上にデカルト学の拠点であったのがオランダであり、チャンスの値の算出方法が体系化された。スピノザも、これに関して「確率書簡」などの記録があ

る。しかし Huygens は、Graunt の『諸観察』を読んで、人口統計との融合を試みた。

Johan de Witt 1625-1672 は数学と法律学を学び、28 にしてホラント州法律顧問（事実上の最高指導者）になり、財政再建を進めた。特に「償還年金と比べた終身年金の価格」を国会に提出した。これは独自の生命表と、Huygens の『チャンスの価格』をもとにしていた。

1.3 フランスでの最小二乗法の誕生

母関数の方法 (Lagrange) も、特性関数 (Laplace) も、最小二乗法の発展 (Legendre) も、フランスで起こった。確率論と結びつけて最適性を初めて指摘したのが Gauss、最尤法の意味論を読み取り、推定論を初めて展開したのが Laplace、数理統計学の応用先として推定論の道をより広くしたのが Quetelet である。

最尤法や十分統計量の命名は Fisher だが本質的には Laplace が開拓していた、Laplace の数学史の講義を拓いたのが Pearson。

19 世紀に、最小二乗法が工学者の常識となった中で、初めて根源誤差の仮定に辿り着いたのが Young。

1.3.1 天文学の大衆化

Tycho Brahe 1546-1601 は助手たちに反復同一観測させ、互いに吟味させた。Lippershey 1570-1619 が望遠鏡を発明したことより、天体観測は大衆化された。ガリレオは『天文対話』で、自然言語により誤差の扱いを議論した。

1.3.2 Lagrange と確率変数の和

Thomas Simpson 1710-1761 (英) は Kepler 則として発見されていて、暗黙知としてもよく使われていた公式を、定積分の近似式である Simpson 式として名を残す。A Letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the Advantage of taking the Mean of a Number of Observations, in practical Astronomy (1775). にて、知識人でも「注意深くなされた 1 回の測定」神話を信じていた。そこで、Lagrange らと共に、算術平均の統計量としての分布を調べた。L'utilité de la Méthode de Prendre le Milieu. Miscellanea Taurinensia にて、母関数の方法を用いて後継した。

1.3.3 Bernoulli と最尤法

Lambert, J. H. 1728-1777 (独) は円周率の無理性を証明もした。それと Bernoulli は最尤推定の考え方を育てていた。

一方で、誤差に正規性の仮定がおけると、その分布に対応した最尤推定法が最小二乗法であることを初めて発見したのが Laplace である。

1.3.4 Legendre と最小二乗法

Nouvelles Méthodes: Pour la Détermination (1805) にて、normal equation を導く。しかし、確率論的な発想がなかった。

Robert Adrain 1775-1843 が最初に証明をしようとする。まずは誤差の従う分布がなにかを直接考察した。

Karl Gauss 1777-1855 は『天体運動論』で、正規性の仮定をおくと、算術平均が最確値であることを導く。また、先取権を巡って筆を執っている。一方でフランス学会は極めて鷹揚な態度を取る。

Laplace は著書『確率の解析的理論』の「試論」の中で、Legendre と Gauss を同等に扱っている。

1.4 Quetelet による数理科学

自然科学を回転させようとした試み

応用統計学の祖と呼ばれるのがベルギーの Quetelet である。フランスには、Laplace が統計的推論の手法を得ていた土壌があった。結局コンドルセも、次の Laplace の精神の体现であった。

政治的・道徳的科学に対しても、観察と計算とに基礎を置くこの方法を、すなわち自然科学で役に立ったこの方法を適用してみよう。

1.4.1 道徳統計学

『人間について』(1835)では「平均人の理論と社会体制の組織に関する」思想を社会物理学と呼んだ。その内実は、肉体的・知的、並びに道徳的諸性質を支配する諸法則を統計的に研究する学問であった。したがって、事実を確定する手法である統計学から派生する理論である（ちょうど生物測定学と同じ）。そして、官庁統計・人口学とも違う。

なお、Quetelet による「統計学」とは、Gottfried Achenwall(アッヘンヴァル) 1719-1772 独の言葉を借りて、「事実・状態を解明するものであって、政治社会に影響を与えるすべての要素を解明するものである」という。または歴史学者 Schözer(シュレーツァー)の言葉を借りて「歴史は動く統計であり、統計は静止した歴史である」と論じる。

1.4.2 コントの社会学

なお、コントも社会物理学を提唱したが、統計的方法、特に数学的方法を信用していなかった。事実、Quetelet は道徳統計学というべきもので、コントの現象論的な態度とは全く趣旨が違っていた。が、一般から特殊へ進むのがコントで、平均人から社会を経て人類へ進むのがケトレーである。大雑把に言えば、コントはいささか Marx の社会進化論的である。

1.4.3 コンドルセの偉大な総合

コンドルセの社会数学は、コントのような「人間精神の進歩」理論を、ケトレーのような科学的手法で、新科学として提唱した。コンドルセの地盤には、ダランベールがいた。

1.5 ドイツでの社会統計学

Adolf Wagner 1835-1917, Ernst Engel 1821-1896, Georg Mayr ら。

1.6 現代誤差論

1.6.1 田口玄一

海軍水路部天文部から統計数理研究所へ。田口メソッドは品質管理手法を開発・設計工程に取り入れた。

第2章

Ronald Fisher と農事試験，Pearson と生物測定学

Sir Ronald Aylmer Fisher 90-62 は一時期ロザムステッド農事試験場 (Rothamsted Experimental Station) の統計研究員だったが、最終的には UCL の Pearson の座 (Galton 教授職) を継いでいる。Fisher が導入した精密標本論をまとめると、次の Rao の言葉になる。

特定の標本から一般化されて得た知識は不確実なものであるが、一度その不確実性を数量化できれば、種類は異なれど、それは確かな知識となる。

また、Fisher 以降、確率モデルを想定した上でパラメータを推定する「理論モデル主導的」な枠組みが支配したが、計算資源が豊かになったことで、「データ主導的方法」が興隆した。

2.1 概観

Neyman-Pearson 学派が巨視的で、Bayes 学派が微視的であり、Fisher 学派はその中間点で、あくまで物理学実験計画法を主軸とする。巨視的には確率は相対頻度であり、微視的には確からしさである。

2.1.1 3つの学派

歴史 2.1.1. Pearson が大標本論・記述統計学と説明され、Fisher の 20s のしごとは精密標本論・推測統計学と説明される。社会からの統計学への要請で最も大きなものは検定論である（この点がのちのち Fisher が排撃した Bayes を蘇らせることとなる）。Fisher の有意性検定は農事試験のデータ解析であったが、これを 1950s の Neyman-Pearson の仕事は、大量生産を背景とした統計的品質管理が背景にあった。ここからさらに、Wald や Lehmann によって数学的に体系化されていく。この「最適性」を指導原理とした数学的体系化、Neyman-Pearson や Bayes などの社会情勢に応えた派生は、原点である Fisher の理論とは違う色を帯びていくことになる。この3派閥は、現在にもあとを引く。

- (1) Frequentist は頻度派ともいうが、Neyman-Pearson 学派のことをいう。今日では数理統計学の主流となっている。
- (2) Bayesian はベイズ派と呼ばれる。産業応用派・意思決定理論への応用である。
- (3) Fisherian は Fisher 学派のことをいう。実験家の統計学である。

議論 2.1.2 (Frequentist vs. Fisherian). Fisher にとっては、統計的推測理論はあくまで物理学実験手法の延長であり、科学的な帰納的推論の一環である。一方で、Frequentist は、現在の数理統計学同様、「最適性」を指導原理とする。

議論 2.1.3 (Frequentist vs. Bayesian). Bayes の「主観確率」の理論は一度は Fisher によって排斥されたが、1950s に再構築された。

2.1.2 Fisher の業績

- (1) 統計量の精密な標本分布について、Fisher が関わっていないものを見つける方が難しい。このとき母集団が正規分布に従うことを仮定していたが、これは主に生物データを扱っていた Fisher にとっては現実的である。
- (2) さらにその間の関係まで整理し、 χ^2 分布、 t 分布、正規分布、標本相関係数 r の分布は、すべて F -分布の特別な場合である

ことを理解した。

- (3) 大標本理論に関する論文 Fisher 22[Fisher, 1922b] にて、逆確率法を排撃し、Pearson のモーメント法に代わって最尤法を提示。^{†1} 十分性の概念も Laplace がすでに用いていたが、有効性、一致性などの観点から検討し、一般的な方法として提示した。母集団の母数と標本から得られる統計量を明確に区別。実際、農事試験データに Bayes の方法はフィットしない、純粋な科学的帰納法のみが要請される現場であるからだ。また、この論文では、統計学の目的は、次の3ステップから成る「データの縮約 (reduction of data)」であるとした。
- (i) 特定化 (specification) の問題：モデル選択。
 - (ii) 推定の問題：選択したモデルの母数を最適に推定する。
 - (iii) 分布の問題：構成した統計量の分布を求める。ここで漸近論を採用せず、精密な分布を求めたのが Fisher の特徴である。
- (4) あくまで関心は小標本にあり、情報量・十分性・尤度の概念をそちらに適用しようとした。
- (5) 実験計画法とは実験化のお家芸であったが、これが統計家の領域に引き込まれた。そして因果推論の基本となる分散分析 (ANOVA) という検定法を開発。

歴史 2.1.4 (日本での受容). Pearson までの統計学で、母集団のパラメータと標本統計量が区別されなかったように、全数調査じゃなくてランダムサンプリングを使うという数学的トリックが生じたのが Fisher からである。そこで日本で受容する際も論争が起り、数理統計学者が説得する形となった。

2.2 Galton までの優生学

2.2.1 Mendel の遺伝理論

歴史 2.2.1 (Gregor Johann Mendel 1822-84). 修道会に入会し、所属した修道院は学術研究・教育が盛んだった。司祭になってから 51 年から 2 年間 Doppler の下で数学・物理学、ウンガーから植物生理学を学び、戻ってからえんどう豆の交配実験を行った。論文 [Mendel, 1866] は数学的で抽象的な議論が理解されず、「反生物的」とさえ評された。後に井戸の水位や太陽の黒点などの気象との関係を研究し、没した時点ではむしろ気象学者としての評価のほうが高かった。

理論 2.2.2 (Hardy – Weinberg principle). すべての生物の因子型は、2 値変数 A, a の非順序対 $AA, Aa, aa \in [\{A, a\}]^2$ によって表現される。 A の遺伝子頻度を p として測られる。すると、3 元状態空間 $\{AA, Aa, aa\}$ 上の確率分布の世代毎の変化・発展が考え得る。このとき、十分大きな集団においては、親の母集団の因子型の組成がどうであろうとも、ランダムな交配は 1 世代以内に、「近似的に安定な因子型分布」を持つ [Hardy, 1908]。

要諦 2.2.3. したがって、進化＝遺伝子頻度の時間発展の研究は、どのような要因で Hardy-Weinberg の法則が破られているかに関する研究となる。

理論 2.2.4 (性染色体). 染色体のうち、性決定に関与する、雄雌異体の生物で唯一異なる染色体。男性が XY であるから、性に関係する性質は父から息子へは遺伝し得ない。

色盲などのような性に伴う遺伝因子の多くは劣性で、 a の男性すべてと aa の女性すべてに現れる。したがって、起こる確率が男性で α ならば、女性では α^2 になる。

2.3 最尤法

Pearson のモーメント法は静的な解析であったが、最尤法は攻撃で動的で推論的である！ 22 年 [Fisher, 1922b] に系統的に提示する。

歴史 2.3.1 (赤池さんは Fisher の超克を意識していた). [赤池広次, 1976] において、

MLE は Fisher によって初めて詳細に論じられてから極めて広く用いられているが、その有効性の根拠が何にあるのかは明

^{†1} 考え方自体は、Edgeworth の "genuine inverse" や Pearson の段階ですでにあった。

らかではなく、その合理性の説明は通常直感に訴えてなされてきた。前項で論じた因子分析法には最尤法が用いられている。そこで、最尤法が何を最適化しようとしているのかが明らかになりさえすれば、先の問題は解決されることになる。

この後すぐに、漸近論的な立場から、 M -推定量のような発想「最尤法は $E[\log_e f(Y|\theta)]$ を最大にする θ の推定量である」さらに言い換えると、「KL-情報量を規準として、真の分布を最もよく近似する $f(y|\theta)$ を求めようとしている」と捉え直している。

AIC の導入は、平均対数尤度 $l(\theta)/N = \mathbb{P}_n[\log f(Y|\theta)]$ が、 $E[\log_e f(Y|\theta)]$ の推定値であるという着想の下に成立した。この見方に従えば、 $\{f(y|\theta)\}_{\theta \in \Theta}$ の族が Y の真の分布を与える $g(y)$ を含まない場合においても、最尤法が統計的モデルのパラメータ決定に有効なものでありうる事が容易にわかる。直交射影ではないが、集合 $\{f(y|\theta)\}_{\theta \in \Theta}$ の中で g に最も KL-距離の意味で近い点を選出する算譜なのだろう。つまり、赤池さんの言葉で言えば、最尤推定量 $\hat{\theta}$ はエントロピーに関して g を最もよく近似する $f(y|\theta)$ を与える θ の、観測値 y_1, \dots, y_N に基づく推定値を与える。ただし、適用条件は、平均対数尤度が $E[\log_e f(Y|\theta)]$ の推定値として有効な限りであり、またモデル $\{f(Y|\theta)\}$ も人間が勝手に自分の責任において取捨選択するものである。「Fisher が明らかにし得なかった尤度の本質的に実験的帰納的な性格と、統計理論における尤度利用の必然性とは、ここに客観的に描き出されている」。

2.4 実験計画法

農場実験での実験計画法を確立するために、精密分布を計算するなどの仕事は必要であった。なぜならば、まず暗黙知であった有意性検定をするためには、誤差のコントロールが必要になった（分散分析の開発）。特に、肥沃土の不均一性が一番大きな系統的誤差原因となるので、必然的に実験の規模は限られ、必然的に小標本理論になる。

2.4.1 農事実験

歴史 2.4.1 (高度農業). 1846 年の穀物法撤廃を契機に、英国で「高度農業」が始まり、ここから農事実験は「品種改良」「化学肥料の効果比較」が中心となった。Rothamsted を設立する Lawes J. B. は 1842 年にはじめて化学肥料を開発した。このとき農事試験を科学化しなければいけないという要請が高まり、その最初の研究が Johnston, J. F. W. (1849). *Experimental Agriculture, Being the Results of Past and Practical Agriculture* である。これはすごく自然言語で書かれた実験計画法の萌芽でもある。

歴史 2.4.2. Rothamsted では 1919 年時点で 75 年ほどの実験データが溜まっており、この分析法を考えあぐねていた。

2.4.2 分散分析と有意性検定

農事実験では統制すべき変数が多いので、そのそれぞれについて、変動への寄与を分離することが必要だった。当時 Fisher は変動分析 (analysis of variation) と呼んでいた。そしてその後、処置による変動に対して、有意性を判定することが必要。特に、精密標本分布を用いると、標本の大きさに拘わらず、正確な有意性検定を構成できる。

特に、肥沃土の不均一性が一番大きな系統的誤差原因となるので、必然的に実験の規模は限られ、必然的に小標本理論になる。

歴史 2.4.3. Fisher 18 [Fisher, 1918] で分散 (variance) を命名し、遺伝学研究のときに発明され、翌年から就職した Rothamsted で精緻化される。ここで F 分布を、当時は z 分布と命名した。唯一の先行研究は、Gosset=Student 08 [Student / Gosset, 1908] の t 分布が、当時の大標本理論に頼ることなく、標本統計量のみに基づいて検定を構成したものである。

そしてブロックデザインが大事になる。当時の実践家は誤差を最小にしようとしていたが、Fisher 以後は誤差が統計的にコントロール可能であればよい。

次に、実践家のお家芸であった有意性判定を、有意性検定で置き換えた。

2.5 有意性検定

Fisher の有意性検定以前にも, Pearson の χ^2 -適合度検定, Gosset の t 検定 (当時は z 検定) があったが, 普及したのは Fisher のものが初である. その理由は, 小標本にも適用可能であること (これは t -検定もそう), 社会的要請の2点がある. Gosset は観測値の精度評価を主な目的とした誤差論的な発想であったのに対し, 誤差が系統的に把握できている時点で OK であり, 有意性の検定に使える, という発想の転換がある. これはある意味で用途の制限であるが, これが爆発的な応用を生んだ.

2.5.1 Fisher 以前の検定論

歴史 2.5.1 (確率誤差検定 (Wood and Stratton 1910)). 基本的な問題設定は, ある正規分布を観測しており, その母平均を推定する問題である. そのために, 誤差の発生要因を統制する必要がある (これと付随する分散分析の技法をあわせて実験計画という).

標本 $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ から定まる**確率誤差 (probable error)** とは, $0.67\sigma/\sqrt{n}$ とする. すると, $\mu \pm 0.67\sigma/\sqrt{n}$ は確率 0.5 でこの範囲に入る. これを用いて検定していたので, 「帰無仮説」という概念はない. Fisher を待つ必要がある.

Edgeworth, F. Y. は経済学の文脈で, Pearson, K. は優生学の分野で検定を構成した. 前者は *Economic Journal* の創刊から編集者で, 後者は *Biometrika* を主宰.

歴史 2.5.2 (Francis Ysidro Edgeworth 1845-1926 アイルランド と 数理心理学). 発明家の祖父を持つ名家の出身. 父はユグノーの子孫で, ドイツに向かう途中, 英国博物館の階段であったスペインの難民と駆け落ち中に生んだ子供. 基本的にあらゆる学問をやっていたが (弁護士資格もある), King's College London で経済学の職に就く. 新古典派の主要人物となり, 数学的形式を個人の意思決定へと応用した最初の人物となった. 効用理論と無差別曲線 (indifference curve) でミクロ経済学, Edgeworth 展開で数理統計学に名前を残す. 81 [Edgeworth, 1881] は数学も, フランス語・ラテン語・古典ギリシャ語にまたがる文章も, アホみたいに読みにくいらしい.

マーシャルとはともに数学と倫理学を通じて経済学に達したという類似点がある. エッジワースは社会科学に数学の手法を適用した先駆者の一人である. 彼自身はその手法を「数理心理学」と名づけていた.

歴史 2.5.3 (Karl Pearson の適合度検定). Galton が確立した相関・回帰の概念から出発して, ピアソン系, モーメント法, 確率誤差論, χ^2 -適合度検定を開発した. これらは記述統計学と呼ばれたが, Fisher のように標本の大小に拘らない枠組みを作ることにはなかった.

X_1, \dots, X_n に関する理論値と観測値の間の乖離の尺度を

$$\chi^2 := \sum_{i=1}^k \frac{(m'_i - m_i)^2}{m_i}$$

とする. ただし, $k \leq n$ について m_i ($i \in [k]$) はそれぞれの値に対する頻度である. すると, χ^2 なる統計量の漸近分布は, 自由度 $k-1$ の $\chi^2(k-1)$ に従うことを示した [Pearson, 1900].

歴史 2.5.4 (William Sealy Gosset 76-37 とギネスビール). Oxford で化学と数学を学び, 学士のままギネスビール社のダブリン醸造所に就職した (農事試験同様, 数年前から統計家を積極採用していた). その後あまりの統計的困難さに, 30 歳前後で Pearson の研究室に赴いて, [Student / Gosset, 1908] を出した. ギネスビール社は機密保護の観点から社員が論文を出すことを禁止していたので, Student というペンネームで発表していた. もともと $z = \frac{t}{\sqrt{n-1}}$ を用いていたが, Fisher が自身の自由度理論に併せるために t に変えた. Fisher と Pearson の仲を取り持つ立ち回りをした.

醸造過程の共変量は互いに独立でなく, データも少ない. 誤差の正規性も仮定できないので, 当時の誤差論は適用不可能であった. [Student / Gosset, 1908] のアブストでは

繰り返すことが困難な実験は少なくない. このような場合, データ数は少なく, いくつかの化学実験, 多くの生物学的実験, そして殆どの農事試験・大規模実験が当てはまり, 従来これらはほとんど統計学の範囲外であった. また, 正規曲線の方法は大標本のみで適用可能であるが, いつ適用不可能になるかの研究がない.

そこで、標本平均を \bar{x} 、標本不偏標準偏差を s として

$$z = \frac{\bar{x} - \mu}{s}$$

とし、この分布と確率積分表を作った。モデルを正規 $N(\mu, \sigma)$ として μ を推定する際、2つの平均値の統計的有意性を検定できる。

しかし、Gosset はこれを小標本の場合の正規近似に用いただけで、検定を構成はしなかった。ここでも確率誤差検定同様、棄却域を設定せず、検定に2つ以上の役割を持たせており、Wald の意味で「検定」とは言えない。

2.5.2 Fisher の有意性検定

有意性検定に用いるには、Gosset の研究はほとんどそのまま使えるが、唯一、有意性査定のために特殊化する必要があった。これは、Gosset は観測値の精度評価を主な目的とした誤差論的な発想であったのに対し、誤差が系統的に把握できている時点で OK であり、有意性の検定に使える、という発想の転換がある。これはある意味で用途の制限であるが、これが爆発的な応用を生んだ。

また、Pearson の適合度検定は、事前に設定した棄却域の下で棄却するのではなく、単純に $P[\chi^2 \geq c]$ を見て判断する。

さらに、母集団と標本との確率論的な関係を厳密に扱うために自由度の概念を導入し、 t 検定として作り直した。

2.6 Pearson の統計理論

標本の理解と、特定の分布への当てはめ (fitting) を通じて、統計的手法を開発した。もちろん推測への萌芽は含まれているが、あくまで、大数観察とそこから科学的な知見を引き出そうとすることが学問的興味であった。

Pearson の検定は誤差が小さいことの確認であり（その上でマクロ生物理論を立てたいので）、Fisher の検定は系統的な誤差の中に有意なものが混じっていないかの検出である。

2.6.1 記述統計学

ドイツに留学して、Karl Marx に心酔して、革命はいただけないが、社会を科学的に改善していく精神を汲み取った。

記述の方法の発明—これは発見というより創造である。科学の進歩とはこういう記述の方法を創造することなのである」。

標本が十分に大きいときに、これ自体を母集団とみなすならば、背後に分布を仮定することも、標本と母平均との確率的関係に頭を悩ませることもない。この特権を享受した下での統計的手法を、記述統計学という。これは、全体を見渡して知恵を得たい、複雑な大規模データを要約して理解したいという傾向のある学問（マクロ生物学など）で取られる手法となる。一方で Gosset や Fisher は推論が必要になり、Edgeworth は意思決定が必要であった。

2.6.2 生物測定学

Walter Weldon 1860-1906 と創始した生物測定学は 1890s から 1920s まで唯一の統計学の高度の訓練をする組織であった。University College は世界の統計学の中心であった。

メンデル遺伝学派との論争に明け暮れてから、生物測定学派は王立協会からも、生物学会からも締め出された。

2.6.3 モーメント法とピアソン系

しかし、Pearson も非対称分布を持つ標本に出会い、これを正規分布の混合として説明しようと試みた。そのための方法としてモーメント法を開発した。4 次までの積率に注目し、正規分布の歪度 $\gamma_1 = 0$ と尖度 $\gamma_2 = 3$ と比較することで近さを測れる。母モーメントを知っていれば、標本モーメントはそれに確率収束するだろうから、必要な次数までモーメント方程式を用意すれば、その解そして一致推定量が得られる、という手法である。このとき、母分布にパラメトリックな仮定をおくことになるが、これをピアソン系と呼ぶ。左右非対称性や尖り具合によって 12 に分類されている。

こうして最後に、ピアソン系の適合度が知りたくなる。そこで適合度検定が要請された。

第 3 章

Neyman-Pearson 理論

統計的推測理論が次の発展をするのは、工業化が契機になる。農事試験などの精密なものではなく、大量生産とその誤差管理が問題になる。そこで、最適な有意性検定 (optimal test of significance) の理論的枠組が生まれた。[長屋政勝 et al., 1999]

3.1 統計的品質管理の父

歴史 3.1.1 (Walter A. Shewhart 91-67 米). California Berkeley で物理学博士を取ったのちに、AT&T の前身である Bell Telephone で通信システムの信頼性向上の仕事をした際に残した知的遺産により、統計的品質管理の父と呼ばれる。プラグマティズム哲学者 C・I・ルイス の著作に影響されて操作主義的姿勢を鮮明に表し、それが彼の統計処理に影響していた。

シューハートの上司であった George D Edwards は「シューハート博士はほんの 1 ページのメモを書いた。その 3 分の 1 は、我々が今日概略の管理図と呼ぶような単純な図だった。その図と前後の文章には、今日の我々がプロセス品質管理として知っている基本原則と考慮すべきことが全て記述されていた」と回想している。

製造工程のばらつきが正規分布の形になるならば、「特殊要因」がなくなり、「偶然要因」のみからなることになり、これが理想的な管理状態とした。管理図とは、この 2 状態の区別をするツールである。

彼はアカデミアにいる William E. Deming 00-93 米にも影響を及ぼした。デミングはシューハートの考え方に基づいて科学的推論に関して研究を展開し、PDCA サイクルも生み出した。デミングは Yale で数学と物理学の博士を取ってから Bell でインターンをしたのち、ダグラス・マッカーサー将軍の下で日本政府の国勢調査コンサルタントを務め、統計のプロセス管理手法を日本の企業経営者に教えた。その後も何度も日本に赴き、1950 年から日本の企業経営者に、Bell 研究所で Shewhart から学んだ設計/製品品質/製品検査/販売などを強化する方法を伝授していった。彼が伝授した方法は、分散分析や仮説検定といった統計学的手法の応用などである。デミングは、日本の製造業やビジネスに最も影響を与えた外国人であった。このため、以前から英雄的な捉え方をされていたが、アメリカでの認知は彼が死去したころやっと広まり始めたところであった。

3.2 Neyman-Pearson の統計的仮説検定理論

28, 33 にて、対立仮説、第 I,II 種の過誤、検出力の枠組みから Fisher の有意性検定を捉え直した。そして検定の「最適性」の枠組みをこしらえた、このメタ視点が達成点である。意思決定理論的だね。2 種類の過誤はトレードオフの形になっていることに注目すると、過誤確率の最適制御の問題になる。これについて、まず単純検定仮説・対立仮説なる対象に絞って、Neyman-Pearson の補題を立てた。

歴史 3.2.1 (Fisher の最尤法からアイデアを得て、検定論でも「最適性」を追求した。). [Neyman, Jerzy and Pearson, Egon S., 1928] では、尤度比検定が導入された。[Neyman, Jerzy and Pearson, Egon S., 1933] では、最強力検定 (most efficient test) が定義された。複合仮説の場合については一様最強力検定が最適だろうが、これが構成できる状況は稀である。33 年以降の研究は、次善策の提案が続き、38 年まで共同研究を続ける。不偏検定、一様不偏検定、相似検定 (similar test) など。

Karl Pearson と Fisher の一番の違いは、誤差への関心の高さと、最適性への関心の高さの別である。この Fisher の立場は Neyman-Pearson に確かに受け継がれた。

統計的仮設検定の問題とは、棄却域を選択する問題である。

しかし、この立場に対する Fisher の反論としては、「有意性が棄却されるか」しかないのであって、その逆は「採択する」ではないはずである。これは帰納的推論の枠組みであり、この立場からは「形式主義すぎる」ということになる。一方で Neyman-Pearson は帰納的行動の枠組みと言える。いずれにしろ、検定の用途の違いである。

背景 3.2.2 (sampling inspection). 1924 年から Bell 研究所にて、Dodge, H. F. and Romig, H. G. が抜き取り検査の研究が始められた。誤って合否を下すことは、producer's risk と consumer's risk の狭間にある。つまり、Neyman-Pearson の応用先的には、どちらかに意思決定をする必要があったのだ！

最終的に、Stanford で品質管理の集中講義が行われ、1945 年夏の時点では各企業から派遣されて出席者は 1 万人を超えたという。

ここにおいて、Neyman-Pearson の検定技法は、契約上の品質に関する合意形成手段として用いられたのである！

このように、[芝村良, 2004] の観点から見ると、統計手法に関する論争は全て応用先の要請の違いから起因したものであり、その解決は、「立場の違い自体をモデル化し、決定理論を作る」ことでアルゴリズム的に解決されてきた。

第 4 章

前期計算機時代の統計手法

第 5 章

機械学習と統計科学

機械学習の分野は数理統計学に基礎を置く「統計的機械学習」として目覚ましい発展を遂げている。「学習」は統計的には「推定」「予測」と同義であるが、特に「機械」が行うアルゴリズム的行為を表現する際に使う、concept with an attitude である。同様に、「機械」とは関数のことをいう。

5.1 機械学習の歴史

人工知能というのは、学習を実装する試みである。論理計算の分野と脳型情報処理の分野との両輪が、統計的機械学習として結実して、実質的に現在はこの分野の業績が人工知能と呼ばれている。

歴史 5.1.1 (第 1 次：記号処理と論理推論＋パーセプトロン (1960s)). 「論理的人工知能」の流れと「脳型情報処理」の流れが合流して一大潮流を作るが、パーセプトロンはすぐに限界が見つかったために収束する。

歴史 5.1.2 (第 2 次：エキスパートシステム＋誤差逆伝播法 (1980s)). 2 つの潮流は再び出会い、パーセプトロンを多層化する方法が模索された。

歴史 5.1.3 (統計的機械学習 (2000s)). 統計や凸最適化、カーネル法やベイズ推論などの成果が揃ってきた。

歴史 5.1.4 (深層学習 (2010s)). 確率的降下法、巨大深層モデルなどが提案され、次世代知能が模索されている。

5.2 類型認識問題の枠組み

決定空間の元を「仮説」または「認識器」、決定関数の空間を「学習アルゴリズム」、一様に良い決定を「Bayes 規則」という。この意味で漸近最適な決定関数を統計的一致性を持つという。統計量を構成するに当たっての計算機集約性に由来する。

5.2.1 決定理論からの概観

定義 5.2.1 (統計的決定理論との対応)。

- (1) 実験 $(\Omega := \mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B}, P(\Omega))$ から得られる標本を**訓練標本**または**観測データ**という。観測データは学習用・検証用・テスト用に分けられることもあり、これを**交差検証法**という。 \mathcal{X} を**入力空間**または**パターン空間**、 \mathcal{Y} を**出力空間**という。
- (2) 決定空間 $\mathcal{D} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y})$ の元を決定というより特に、**仮説**という。決定関数 $\delta : \Omega^n \rightarrow P(\mathcal{D})$ を (確率的) **学習アルゴリズム**ともいう。
- (3) 決定関数 $\hat{y}_n : \mathcal{X} \rightarrow \mathcal{Y} \in \mathcal{D}$ が離散値であるとき、**類型認識器**または**判別器** (classifier) ともいう。このとき、終域 $\mathcal{Y} \subset \mathbb{R}$ を**ラベルの空間**ともいう。

注 5.2.2. 判別機 $\mathcal{X} \rightarrow \mathcal{Y}$ はしばしば、関数 $\mathcal{X} \rightarrow \mathbb{R}$ と商写像 $\pi : \mathbb{R} \rightarrow \mathcal{Y}$ との合成として構成される。これを**判別／識別関数** (discrimination function)、 π を**決定領域**という。

問題 5.2.3 (classification problem).

- (1) 有限集合 $\mathcal{Y} \subset \mathbb{R}$ に値を取る関数 $y : \mathcal{X} \rightarrow \mathcal{Y}$ を**類型 (pattern)** または**ラベリング**という. これの全体を決定空間 $\mathcal{D} \subset L(\mathcal{X}, \mathcal{Y})$ とし, 良い推定量を構成するようなアルゴリズム $\delta \in \Delta \subset L(\Omega^n; \mathcal{D})$ を考える. これを構成するという問題 $(\mathcal{X}, \mathcal{D}, \Delta, W)$ を**判別問題**という.
- (2) 分布族 $\mathcal{P} \subset P(\Omega)$ はしばしば, 真のラベル $y \in \mathcal{D}$ に基づいて設定される. このときを**決定論的パターン認識問題**という. 多くの回帰分析と同様, 殆ど場合はパターンは定かでない, 統計的存在である. このときを**統計的パターン認識問題**という.
- (3) 決定論的パターン認識の場合, 危険関数 $W : P(\Omega) \times \mathcal{D} \rightarrow \mathbb{R}_+$ は $\mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ の言葉で次のように定められる:
 - (a) \mathcal{Y} 上の分離度の指標としては, **0-1 損失** $l(\hat{y}, y) = 1_{\{\hat{y} \neq y\}}$ またはその非対称化 $l(\hat{y}, y) = l'(y)1_{\{\hat{y} \neq y\}}$ ($l' \in L(\mathcal{Y})$) が用いられる. これを**損失距離**とよぼう.
 - (b) l を所与として, 損失関数 $W : P(\Omega) \times \mathcal{D} \rightarrow \mathbb{R}_+$ を $R(P; h) := E_P[l(h(X), Y)]$ を定める. これを**予測損失 (predictive loss)** という.^{†1}
 - (c) この P を経験分布で代用して得る推定量 $\hat{R}(P; h)$ を**経験損失**という.
- (4) 一方で統計的パターン認識問題では, 訓練標本以外の類型も正しく分類する「汎化性能」が重視されるため, より特殊な損失関数が設定される.
- (5) 損失距離 l に関する損失関数 W の下限 $R^*(P) := \inf_{h \in \mathcal{D}} R(P; h)$ を **Bayes 誤差**といい, これを達成する仮説 $h \in \mathcal{D} = L(\mathcal{X}; \mathcal{Y})$, すなわち, 一様に良い決定を, **Bayes 規則**という.
- (6) 危険関数 $R : P(\Omega) \times \Delta \rightarrow \mathbb{R}$ は, 損失関数の期待値とする: $R(P; \hat{h}_n) := E_{P \otimes n}[W(P; \hat{h}_n)]$ と定める. したがって, Bayes 決定則とは誤識別した際の損失が最小になる識別である. 一方で, 正しく識別出来る確率を最大にする, などの危険関数も考え得る. これを**最大事後確率則**または**最小誤識別率則**という.
- (7) この危険関数の値 $R(P; \hat{h}_n)$ が Bayes 誤差 $R^*(P)$ に確率収束するとき, アルゴリズム \hat{h}_n は**統計的一致性を持つ**という.

例 5.2.4.

- (1) \mathcal{X} が 2 次元データであるときの 2-値判別問題の例として, ランキング問題もある.
- (2) 決定論的パターン認識問題には, 手書き文字認識などがある. この問題は一般に **domain specific** になる.
- (3) 回帰問題では, 損失関数 W とそれが定める危険関数 $R : P(\mathcal{X}) \times \Delta \rightarrow \mathbb{R}$ を $l : \mathcal{D} \rightarrow \mathbb{R}_+$ と同一視することは出来ないだろう.

□

5.2.2 学習アルゴリズム

カテゴリの事後確率分布 $p(y|x)$ を直接は推定せず, Bayes の定理に基づいてデータの生成モデル $p(x, y)$ を代わりに推定する手法を**生成モデルに基づくパターン認識**という. これと対照的に, 識別境界の推定に特化した手法として, **識別モデルに基づくパターン認識**がある.

模型 5.2.5. 学習アルゴリズムは次の 4 段階に分けて設計される.

- (1) 観測: 形態素解析, OCR.
 - (2) 前処理: パターンの正規化, 雑音消去.
 - (3) 特徴抽出: 次元を削減する. 手書き文字のどこに注目して 2 次元まで削減するかなどは計算機科学者の腕の見せどころである. または主成分分析を用いる.
 - (4) 識別: 損失が大きすぎれば, 前 3 段階をやり直す.
- (4) のみ, 学習の対象に依らない設計ができる.

例 5.2.6 (最大事後確率則を満たす識別器の構成). 現実的には Bayes 規則は複雑性故に, 最大事後確率則を用いることも多い.

^{†1} 予測損失は損失関数であるが, 危険関数の記号 R を使いがちである.

$f \in \mathcal{D}$ を $f(x) = \arg \max_{y \in \mathcal{Y}} p(y|x)$ を満たすように構成する。これは事後確率 $p(y|x)$ を推定する問題に等価であり、

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \propto p(x|y)p(y) = p(x, y)$$

より、事前分布 $p(y)$ を推定した下で (経験分布を使う)、条件付き確率 $p(x|y)$ を推定する問題に等価。このアプローチを**生成モデルに基づくパターン認識**という。 $p(x|y)$ の推定は、次のようなものがある：

- (1) Gauss モデル下での最尤推定法：決定境界は二次の超曲面になる。さらに分散の均一性を仮定するときは超平面となり、この場合の生成モデルに基づくパターン認識を**線型判別分析**という。計算が簡単で非常に実用性が高い。
- (2) $p(x|y)$ が多峰性を持つときは、少しモデルを拡張して、混合 Gauss モデル下での最尤推定法を用いる。しかしこの場合は最尤推定量の解析的な構成が困難になり、繰り返しアルゴリズムによる計算がなされる。勾配法・EM 算譜などがある。勾配法は、単純に目的関数である尤度関数を最大化する、最適化一般の数値計算算譜である。実装が非常に容易で局所最適解への収束が保証されている。大域的最適解を探すには、これを別々の初期値から何度か繰り返す (多点探索)。ただしステップ $\epsilon > 0$ の設定は繊細な注意が必要で、だんだん小さくしていく手法を焼きなまし法というが、その実装は練度が必要である。
- (3) Bayes 推定法を用いることも出来る。事後確率を最大にするパラメータ $\theta \in \Theta$ を選択する。この手法の利点は、共役な事前分布に対しては解析的な方法で Bayes 推定量が求められることである。
- (4) 一方で任意のモデル Θ と事前分布に関する Bayes 推定では、数値計算法が欠かせない。
- (5) 以上と違ってノンパラメトリックな手法には、カーネル密度推定がある。これは尤度交差検証法と併せれば、非常に汎用的かつ実用性が高い。
- (6) カーネル法に替わって、最近傍密度推定法があり、こちらはパターン認識と非常に相性が良い。

□

5.2.3 決定空間が有限な場合

問題 5.2.7. $\mathcal{Y} = 2$ の 2 値判別問題について、決定空間を $\mathcal{H} = \{h_1, \dots, h_T\} \subset L(\mathcal{X}; \mathcal{Y})$ とする。

観察 5.2.8. 経験判別誤差を最小にする仮説を出力する学習アルゴリズム $\Omega \ni S \mapsto h_S := \arg \min_{h \in \mathcal{H}} \hat{R}(h)$ が構成出来たとする。 \mathcal{H} が Bayes 規則を含むならば、 h_S の予測判別誤差は Bayes 誤差に収束する (一致性)。しかし一般にそんなことは期待できない。したがって、 \mathcal{H} をなるべく大きくしてバイアスを小さくしたいが、そうすると分散が大きくなる。よって、適切な大きさの \mathcal{H} を設定することが難しい。こうして、 \mathcal{H} が大きすぎることに對する罰則を課すことで対処する、これを**正則化**という。正則化パラメータは、交差検証法などで設定する。

□

5.3 教師なし学習の枠組み

統計的決定理論では捉えられない枠組みであり、いわば「標本が存在しない」。

例 5.3.1. クラスタリング、外れ値検出、独立成分分析などは教師なし学習問題である。

□

第 6 章

ビッグデータと知識発見

知識発見 knowledge discovery in databases とはビッグデータからあらゆる技術を用いて、通常の統計の意味とは違い、よりヒューリスティック・発見的な知識獲得を目指す分野である。

データ前処理 (1) 自然言語処理：形態素解析によりテキストデータを構造化し、機械学習に繋げる。

6.1 歴史

データマイニング、テキストマイニング、機械学習、自然言語処理、情報検索、ナレッジマネジメントなどの分野で培われてきた技術の集合体を用いて、情報爆発という危機的問題の解決に挑む分野である。

歴史 6.1.1. 起源は 1989 年である。

- (1) 1980s は関係データベースと SQL の開発で、オンデマンドかつ動的なデータ解析が可能になった。
- (2) 1990s からハードディスクが安くなったことから、汎用 PC により構成出来るようになったことで、データを長期間に渡り蓄積するという観点を実現出来るようになった。
- (3) 2000s に Kevin Ashton により、IoT によって実データが大量に収集・供給される世界の到来が一般認知された。
- (4) 2010s は Google の BigQuery や Amazon Redshift など、Data Warehouse は我々には見えなくなっている。そのおかげで、クラウドサービスとして、多くの企業がデータを蓄積することが可能になった。こうしてビッグデータ解析が席巻している (big data は The Economist 誌による)。次はなんだろうか？

2011 年にデータマイニングと推論を応用した質問応答システムである"IBM Watson"がアメリカのクイズ番組"Jeopardy!"に出場して人間に勝利する。

要諦 6.1.2. その手法の要諦は次のようなものである：

- (1) パターン抽出：特に「似たような商品を購入した人はこんなものも購入しています」という知能は、Market Basket Analysis と呼ばれる。「おむつとビール」は知識発見の例である。
- (2) クラス分類 (カテゴリの予測)、クラスタ分析 (類似度解析)、回帰分析など。

6.2 形態素解析

例 6.2.1. MeCab, Janome, GiNZA を用いる。

- (1) Python には spaCy というフレームワークがあり、これで日本語を解析するモジュールが GiNZA。

□

注 6.2.2.

第 7 章

統計的因果推論

因果推論は従来哲学の分野であった。この分野も近年急速に統計化しつつある。

- (1) 確率空間を観測可能な形で広げる行為が、その確率的現象をより詳しく知る行為だとすれば、まず独立試行を繰り返すことによって直積を取ることが出来る。
- (2) これが確率統計学の最初の歴史となった。続いて、反実仮想をすることで、確率空間の直和を取ることも出来る！

7.1 歴史と発展

統計的因果推論の要諦

統計的推定・検定は、統計的実験 $(\Omega, \mathcal{F}, (P_\theta))$ の直積の列上で展開される漸近論であった。しかし、多くの分野で関心の対象となるような、データの生成過程に関する問いは、この枠組みでは十分な回答を得られない。データの生成過程に対する模型として、部分的解決を与えていたのが構造方程式模型である。これに対して、反実仮想確率変数 Z を通じて、統計的実験 $(\Omega, \mathcal{F}, (P_\theta))$ の直和を考える営みが、統計的因果推論である [清水昌平, 2010]。

7.1.1 総覧

以前はどのように対処されていたのだろうか？因果分析手法をまとめた [Asher, 1983] では因子分析や Simon の方法を挙げており、同時に、簡単な方法のみが注目され、Wright の逐次パス解析の方法などの利用が進まない状況に警鐘を鳴らしている。1980 年に付された役者後書では、[Asher, 1983] は因果分析最良の教科書の一つだと言っているが、文中では構造方程式模型に当たるものが「多重指標模型」と呼ばれており、[Jöreskog, 1970] の結果を「彼の確認的因子分析法を因果分析に利用したユニークなもので、LISREL としてプログラム化されたとも聞いている」としている。

歴史 7.1.1.

- (1) Pearson を初めとした Galton のグループが相関関係 1880s を研究し、その途中で線形回帰の手法を用いた。最も、手法自体は、天体観測から真の軌道の模型を fitting する仮定で、Legendre (1805) や Gauss (1809) によって、Gauss-Markov の定理の原型まで得られていた。回帰模型の正規性の仮定を弱めたのは [Fisher, 1922a] である。
- (2) 心理学の分野にて、[Spearman, 1904] が因子分析を始めた。
- (3) 遺伝学の文脈でモルモットの研究を通じて、[Wright, 1918], [Wright,] がパス解析を展開した。
- (4) [Neyman et al., 1990] は潜在結果の概念を定義したが、1990 年まで英訳は出版されなかった。
- (5) 生物統計にて、[Cox, 1958b] は Z への介入が有効なのは他の独立変数に影響を受けないときに限ることに継承を鳴らした。因果分析に関する知識の欠如が自覚され始める。60 年代に、統計学としては無視されていたパス解析が再発見される。
- (6) 社会学にて、[Jöreskog, 1970] は一連の因果模型の前身を構造方程式模型として枠組みを作ったが、一度定型化した手続きに落ちると、因果についての議論は覆い隠されるようになった。特にパス分析の部分を同時方程式模型と呼んで経済学が取り入れたが、因果の議論は不十分なままでひたすらに研究が生産された。
- (7) 数学者 Samuel Karlin の批評を受けて、Wright は 60 年前の自身の論文を要約するような論文を発表した。

- (8) 元々心理学を専攻していた Rubin は [Rubin, 1974] にて、因果を扱うための語彙として「潜在結果」を定義した。
- (9) 心理学雑誌にて [Baron and Kenny, 1986] が線型方程式系の中で中間変数の存在を検出する手法を提出し、33 番目の引用数を誇る。
- (10) 同年に Greenland と Robins は交換可能性によって交絡を取り扱う手法を導入した。すなわち、**交絡がない**とは、処置群が処置を受けなかった場合の効果が対照群のそれに等しいことをいう。

歴史 7.1.2 (主な研究者).

- (1) Judea Pearl 36- イスラエル
 - (a) 「因果推論のための解析手法を開発し、人工知能研究に本質的な貢献をした」として 2011 年 Turing 賞を受賞した計算機科学者。
 - (b) Tel Aviv の生まれであり、電気工学学士を取ると渡米し、引き続き電気工学修士 (New Jerzy 工科大学), 物理学修士 (Rutgers 大学), 電気工学博士 (現 New York 大学)。
 - (c) 王手電機メーカーの RCA から始まり、民間で記憶デバイスの開発に携わるが、「半導体により仕事が奪われた」のちに UCLA 工学科へ (1970 年)。研究テーマがここで人工知能に移る。
 - (d) Bayesian Network の手法の基本的な提案者であり、これをはじめとして人工知能界限に確率統計学的手法を取り入れた第一世代である。このテーマにおける著書は [Pearl, 1984], [Pearl, 1988]。
 - (e) 計算機が因果律を理解するための形式・認知模型としての興味もあった (計算機科学者であると同時に認知科学者でもある) ことが、因果推論の極めて一般的な枠組みを提案することができた背景の一つであろう。
 - (f) 現在 UCLA で Cognitive Systems Laboratory の所長を務める。
- (2) Donald B. Rubin 43- 米 Harvard 大学 John L. Loeb 教授職 (統計学)。
 - (a) ワシントンで生まれ、Princeton 大学で心理学学士, Harvard 大学で計算機科学修士, 統計学博士 (Cochran の下)。
 - (b) 1983 以来, Harvard 大学教授。そもそも欠測データ [Rubin and Little, 2019], Bayes 統計 [Gelman, Andrew, Carlin, John F. 2013] の専門家でもある。
 - (c) Imbens と [Imbens and Rubin, 2015] を共著。
- (3) Paul R. Rosenbaum 54- 米 Pennsylvania 大学 Wharton 校 Robert G. Putzel 教授職 (統計学)。
 - (a) 統計学博士 (Harvard 大学, 1980) ののち, 1986 以降 Pennsylvania 大学 Wharton 校に勤務。
 - (b) Rubin との共同研究 [Rosenbaum and Rubin, 1983] で傾向スコアを提案。
- (4) David R. Cox 24-22 英 Imperial College London で数学科長を務める: 学士 (数学, St John's College), 博士 (統計, Leeds)。
- (5) Steffen Lauritzen 47- 英 Oxford : Pearl の構造的因果模型をグラフィカルモデルから受容した論文 [Lauritzen, 2000] を書いている。

歴史 7.1.3 (疫学分野).

- (1) James Robins 米 Harvard 公共衛生学校 Mitchell L. and Robin LaFoley Dong 教授職 (疫学)。

歴史 7.1.4 (計量経済学分野).

- (1) Guido Imbens 63- 米 Stanford : Angrist と自然実験に対する LATE の枠組みを開発 [Imbens and Angrist, 1994] し, その可能性と限界についての基本的な貢献をした。
- (2) Joshua Angrist 60- 米 MIT : 擬実験研究 (quasi-experimental) の枠組みを用いた政策提言など。
- (3) Scott Cunningham
- (4) Charles Manski

歴史 7.1.5 (日本の主な研究者).

因果探索

- (1) 清水昌平@滋賀大学: 大阪大学 基礎工学研究科 博士。
- (2) 狩野裕@大阪大学 基礎工学研究科 数理科学領域。

疫学

- (1) 佐藤俊哉@京都大学 医学研究科 社会健康医学系専攻 医療統計学 教授：東京大学 医学系研究科 保健学専攻 博士。
- (2) 松山裕@東京大学 医学系研究科 公共健康医学専攻 生物統計学分野 教授：同上の博士（保健学）。

経済学

- (1) 竹内啓 33-：東大経済学部卒業後そのまま助手採用。経済学部教授時代のゼミで、佐和隆光、竹村彰通、国友直人を教える。退官後は明治大学。
- (2) 佐和隆光 42-：修士を得ると助手採用の後に京都大学経済学部助教授へ。退官後は立命館大学、滋賀大学学長。
- (3) 美添泰人 46-：博士を得た後に Ph. D. (Harvard) も取り、立正大学、Carnegie-Mellon 大学、青山大学で教授。
- (4) 国友直人 50- @東京大学：修士を得ると Stanford 統計学科へ。帰国後は経済学部教授。
- (5) 竹村彰通 52- @滋賀大学：芸大ピアノ科ののちに東大経済学部へ。博士は Stanford 統計学科にて。日本に戻ると、東大経済学部・情報理工学系研究科を得て、滋賀大学データサイエンス教育研究センターへ。

数学

- (1) 林知己夫 18-：学士（数学，東京帝国大学）の後に陸軍航技候補生として水戸陸軍飛行学校に入校。4年後に統計数理研究所へ。国立大学共同利用機関への改組転換と総合研究大学院大学の創設準備に力した。John Pelzel の提案による日本語ローマ字化計画を阻止した社会調査のサンプリングを実施した。
- (2) 赤池弘次 27-：学士（数学，東京大学）の後に文部省統計数理研究所，そのまま勤め上げる。自動制御理論で扱った時系列解析・因子分析の知見を基に [Akaike, 1974] を発表。
- (3) 鈴木雪夫 29-：学士（数学，東京大学）の後に文部省統計数理研究所，初代統計学会理事長，経済学部助教。定年後は多摩大学へ。
- (4) 赤平昌文 45-：博士（数学，早稲田大学）の後に筑波で勤め上げる。
- (5) 北川源四郎 48-：修士（数学，東京大学）の後に博士を中退して統計数理研究所研究員。その後ずっと統計数理研究所。
- (6) 岩崎学 52- @横浜市立大学：修士（数学，東京理科大学）ののちに防衛大学校，成蹊大学を経て横浜市立大学。

7.1.2 Pearson と Fisher

統計学者は、「因果」の用語こそ表向きに使わなかったものの、ずっと関心を持ってきた [Cox and Wermuth, 2004]。Yule (1900) が因果と相関の違いを特に時系列解析の分野で声高に警鐘を鳴らした。Fisher はランダム化によって因果推論は可能であるとした [Fisher, 1926], [Fisher, 1935]。これが RCT である。Fisher と協業していた [Cochran and Chambers, 1965] は、より因果についての分析を推し進めるため、Wright のパス解析を発展させるべきだと論じた。後にその弟子が Neyman-Rubin 因果モデルを作ることになる。

7.1.3 回帰模型

回帰模型はただの径数的な確率模型ではなく、変量間の回帰関係を分析する、一歩踏み込んだデータ解析法である。

- (1) 回帰関係は新しい観測の「予測」に役に立つ。変量 x が大きいとわかっているなら、 y も大きい確率が高い、というのが Bayes 推定になる。
- (2) 一方で、「制御」「操作」「介入」「干渉」の予測や評価には、このままでは一切の知識を与えない。

論点 1. Box (1966) は化学工学の文脈で回帰分析の結果を因果の言葉で翻訳する際には注意すべき点が多い行為だという警鐘を鳴らした [Cox and Wermuth, 2004]。

7.1.4 パス解析

米国の遺伝学者である Wright は [Wright, 1918], [Wright,] でパス解析を展開し、因果の模型とグラフの使用の先駆けとなった。これは計量経済学ではその双方向性(潜在変数を持たない点)を強調した名前「同時方程式模型」を持って継承された [豊田秀樹, 2007]。特に Keynes 経済学におけるマクロな経済計画の発想で模型として採用された [甘利俊一 et al., 2002]。

歴史 7.1.6 (Sewall Wright 89-88 米). Haldane, Wright, Fisher らは集団遺伝学 (population genetics) の創始者 3 人衆である。Newton 以来の性淘汰の理論を蘇らせた点も新しい。彼らには、Maynard Smith, Hamilton らが続く。パス解析の一連の論文 [Wright, 1918], [Wright,] は、モルモットの体表面の模様の遺伝法則を調べる際に、介入を記述するためであった。

7.1.5 操作変数法・Simon-Blalock 法

擬似相関への対処法として Herbert Simon による因果関係検証法が社会学で知られていた。彼はもちろん当時暗黙理には構造方程式模型の枠組みを持っていたと考えて良いだろうが [Asher, 1983]、後に実際に計量経済学の発展に結びついた。

例 7.1.7. X_1, X_2 の間に $X_1 \rightarrow X_2$ の方向に相関があったときに

$$\begin{cases} X_1 = p_{1u}R_u, \\ X_2 = p_{21}X_1 + p_{2v}R_v \end{cases}, \quad R_u \perp\!\!\!\perp R_v$$

なる模型を想定できるとする。すなわち、交絡因子は存在しないとする。このとき、 p_{21} の値を推定し、非零であったならば因果関係があると言えるだろう。「上述の種々の仮定の下では、相関は因果を意味する」という基準を Simon の方法という。[Simon, 1957] による。同様に地道な方法で 3 変数以上の場合に拡張できるかもしれない。これを実際に行ったのが Blalock であった。本人も「もっと強力な方法を使うべきだ」と主張したが、この方法は盲目的に使用され続けた。□

7.1.6 構造方程式模型

主に経済学、そして実験・観察データを扱うビジネス・社会・行動科学を渡る多くの実践では、構造方程式模型／同次方程式模型という、線型な因果模型を持っていた [Hoyer et al., 2008]。数理的には「推定方程式 (estimating equation) 模型」と同等であるが、歴史と文脈が違う。

問題 7.1.8 (因果関係の暗黙的取り扱い). 問題点は次のように指摘できる (竹内啓による前文 [甘利俊一 et al., 2002])。

ベクトル変数の各成分はいずれも形式的に対等なものとして扱われ、それらの変数の間には論理的な前後関係、あるいは因果関係の存在が仮定されていなかった。因果関係が想定される場合には、原因を表す変量は外生変数、または独立変数として、モデルの外で決定されるものとして扱われ、確率的に変動する変量はそれらによって決定される内生変数、または従属変数とされたのである。そうして内生変数の外生変数に対する回帰関係が、因果関係を数量的に表現するものとしてモデル化されたのである [甘利俊一 et al., 2002]

しかし、潜在変数の存在によって、交絡因子を模型内部に取り入れることが出来る点は極めて柔軟性の高い模型である。

問題 7.1.9 (非線形性の扱い). 模型から線形性の仮定を取り払うと、構造方程式から誘導形を導くことが解析的には不可能になる。その場合、解の一意性と連続性が破れる。

7.1.7 独立成分分析

上述した相関分析法では対処できない問題の一つに、「変量の独立性」がある [甘利俊一 et al., 2002]. 無相関は独立性を含まないためである.

7.1.8 因果を扱うにあたっての問題点

注 7.1.10. 特に従来手法から敢行された変化は

- (1) 実証的でない、理論的・主観的な仮定を置かなければ因果分析には着手できないということ. 特に, Neyman-Rubin 模型では下添字で表される.
- (2) 確率の解析に新たな語彙を追加する必要があること. 例えば, [Pearl, 2009a] は, [Cox, 1958a] における仮定「共変量は処置によって影響を受けない」ことを, 単に自然言語によって説明しているのみで数式による定式をしていない点を指摘している.

の 2 点で, 新規参入の心理的障害になっているであろうということ.

7.1.9 因果探索

7.2 因果推論の枠組み

歴史 7.2.1. [Rubin, 1974] が教育心理学の雑誌にて, RCT をモデルケースにして因果を取り扱う枠組みを提案した. これは [Neyman et al., 1990] の思想的直系に当たり, 実験を中心に扱う分野に広まった. その後, Pearl は構造方程式模型の拡張として, その枠組みを大きく一般化した.

7.2.1 Neyman-Rubin 因果模型

確率空間を割当変数 $Z: \Omega \rightarrow 2$ に沿った直和によって拡張し, 新たな統計量として潜在結果 (potential outcome) を定義する. その差を「因果効果」として推定することで, 因果についての知識を深めることが出来る. これを「因果関係である」と断定するには, この模型の他に共変量の統制が取れていることを議論せねばならない (これには Pearl による手法などが必要になる). すなわち, 因果関係そのものを模型内で十分に扱うわけではない [Pearl, 2009a].

模型 7.2.2 (Neyman-Rubin potential-outcome framework). Neyman-Rubin 因果模型では, 次のデータを所与とする:

- (1) $(\Omega, \mathcal{F}, (P_\theta))$ を統計的実験, $u \in [n]$ を実験単位 (experimental unit) とする.
- (2) その上の確率変数 $Z \in L(\Omega; 2)$ を割り当てとする.
- (3) すると, Z に沿った可測空間の直和 $(\Omega \oplus \Omega, \mathcal{F} \oplus \mathcal{F})$ が考えられ, 同様に確率測度も押し出されて実験 $(\Omega \oplus \Omega, \mathcal{F} \oplus \mathcal{F}, (P_*))$ を得る. Z は自然にこの実験上の確率変数とみなせる.
- (4) $u \in [n]$ で添字付けられた Z -可測な確率変数の族 $\{Y(u)\}_{u \in [n]} \subset L_Z(\Omega \otimes \Omega)$ を単位ごとの結果 (unit-based response variable) とする. すなわち, Z の実現値 $z \in 2$ で条件付けた $Y_z(u) := E[Y(u)|Z = z]$ は確定的な実数である. よって, 実数列 $\{Y_z(u)\}_{z \in 2, u \in [n]} \subset \mathbb{R}$ に等価である.

このデータを通じて, 潜在結果とは, Z -可測な確率変数 $Y \in L_Z(\Omega \otimes \Omega)$ であって,

$$Y := zY_1 + (1 - z)Y_0, \quad z \in 2$$

で定まるものとする. 実験により得られるデータは $Y(u)$

注 7.2.3.

- (1) 上述のデータ, 特に $Y(u)$ を挙動で

$$Y = xY_1 + (1 - x)Y_0$$

で定義するよりも正確な定義を与える基礎理論を与えるのが Pearl の構造因果模型である [Pearl, 2009a].

- (2) 直和確率空間 $\Omega \oplus \Omega$ 上の確率測度 P_* を Pearl は P^* と表して "super" probability function と呼んでいる.

定義 7.2.4. (1)

7.2.2 Pearl の因果の認知模型

例えば [Pearl, 2009b] には「因果の本性、確率との関係、反事実条件、単称因果と般因果、決定論、法則とメカニズム、因果と時間の矢などといった哲学における伝統的なトピックが、随所で顔を出す」ことが指摘されている [大塚淳, 2012].

論点 2. 次の主張は正しいか? (竹内啓による前文 [甘利俊一 et al., 2002])

つまり因果性の存在や, その方向を積極的に確立するということは不可能である. このことは統計的方法にのみ関わることでなく, そもそも本質的に「因果性」は経験的事実だけでは言うことが出来ないものである [甘利俊一 et al., 2002]

7.2.3 Pearl の構造因果模型

構造方程式模型, Neyman-Rubin 模型, そしてグラフィカルモデル [Pearl, 1988] の総合である.

歴史 7.2.5. 社会科学による継承 [Morgan and Winship, 2014], 疫学による継承 [Greenland et al., 1999] と, 統計による受容 [Cox and Wermuth, 2004]. 特に統計の立場からは, [Lauritzen, 2000] はグラフィカルモデルの観点から捉え直した. [Lindley, 2002] は [Pearl, 2009b] をレビューして重要な指摘をした. [Dawid, 2002] は

定義 7.2.6 (intervent). まず, 確率測度の変換として, 介入という操作を定義する. ただし, その確率測度の変換の仕方は条件付き確率のそれに酷似するはずであるから, 記法を借用して次のように定める:

- (1) $P[Y = y | \text{Do}\{Z = z\}]$ によって, 事象 $\{Z = z\}$ が被験者 $u \in [z]$ 上に一様に課す介入を行った際の, 事象 $\{Y = y\}$ の条件付き期待値を表す [Pearl, 1995].

しかし, 最も大事な点は介入と条件付けは違うという点である.

例 7.2.7. 交絡の定義と検出は「介入」の語彙がないと達成出来ない.

Hence the definition must be false. Therefore, to the bitter disappointment of generations of epidemiologist and social science researchers, confounding bias cannot be detected or corrected by statistical methods alone [Pearl, 2009a]

□

7.2.4 構造的因果模型の問題点

問題 7.2.8. いつ第一レベルの因果が結論付けられるかの特徴付けを初めて (?) 与えたことになるが, これが満たされているかのチェックは現実的でないほど難しくなり得る [Lindley, 2002].

7.2.5 因果の Bayes 確率論との比較での理解

[Cox and Wermuth, 2004] による, 相関関係に対して因果関係を認めることが出来るための必要十分条件へ向けた理解と消化を試みる. 推定のためのモデルと, 因果分析のためのモデルは違う, という点も強調している.

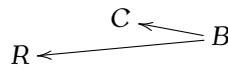
記法 7.2.9.

- (1) C を反応 R に対する因果の候補 (candidate cause) とする.
- (2) I を中間変数 (intermediating variable), B を背景変数 (background variable) とする.

ただしこの 4 変数は, 結合分布関数 f_{RICB} が

$$f_{RICB} = f_{R|ICB}f_{I|CB}f_{C|B}f_B$$

を満たすように取る. $i \in I$ について積分すれば, 上の式から I を無視しても問題ない. ここには, 種々の因果関係が想定出来る. ただし, 辺がないことは条件付き独立性を意味する:



は $R \perp\!\!\!\perp C | B$ を意味する.

仮定 7.2.10. さらに, 次の 2 条件を仮定する.

- (1) $f_{R|C}, f_{R|B}$ は介入 $\text{Do}\{C = c\}$ によって変わらない.
- (2) $f_{I|C}, f_{I|B}$ は介入 $\text{Do}\{C = c\}$ によって変わらない.

深刻な介入をすると, 経済システム内のあらゆる関係は影響を受けて変化してしまうため, これらの前提が破られてしまうだろう, という点が Lucas 批判 [Lucas Jr, 1976] の本質であった.

観察 7.2.11 (従来の Bayes 法による回帰関係の推定). 従来の分析法を考える. 単に $C = c$ と条件付け, 背景変数 B も消去することを考えると,

$$f_{R|C} := \int f_{R|CB}f_{B|C}dB, \quad (f_{B|C} := f_{CB}/f_C).$$

これを通じて C の R への回帰関係について良い推定が行える. □

定義 7.2.12 (介入を通じた因果効果の推定). B の分布を変えない介入 $\text{Do}\{C = c\}$ を行った際, $f_{B|C} = f_B$ であるから,

$$f_{R||C} := \int f_{R|CB}f_BdB.$$

ただし右辺の記法は [Lauritzen, 2000] による.

要諦 7.2.13. B を周辺化した際の, Pearl の意味での**因果効果**の定義である. したがって, 相関関係と因果関係の違いとは, $f_{R|C}$ と $f_{R||C}$ の違いにほかならない.

論点 3. 残る論点は, 観測されていない共変量の扱いである. 特に, 感度分析によって, 仮定の違反が大きな影響を与えないと分かるならば, その方面での解決もあり得るかもしれない, と最後に結んでいる [Cox and Wermuth, 2004]. もしかしたら統計の真意は, 「不確実性・不確定性を取り扱える」という点で光るのかもしれない. 譲歩しているように見えて, 何一つ譲歩していないのである.

第 8 章

参考文献

参考文献

- [Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transaction on Automatic Control, 19(6):716–723.
- [Asher, 1983] Asher, H. B. (1983). Causal Modelling, volume 3 of Quantitative Applications in the Social Sciences. SAGE Publications, Inc.
- [Baron and Kenny, 1986] Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6):1173–1182. 2014 年時点で 33 番目の引用数を誇る論文である。
- [Cochran and Chambers, 1965] Cochran, W. G. and Chambers, S. P. (1965). The planning of observational studies of human populations. Journal of the Royal Statistical Society. Series A (General), 128(2):234–266.
- [Cox, 1958a] Cox, D. (1958a). The Planning of Experiments. John Wiley and Sons, NY.
- [Cox, 1958b] Cox, D. R. (1958b). The regression analysis of binary sequences. Journal of the Royal Statistical Society. Series B (Methodological), 20(2):215–242.
- [Cox and Wermuth, 2004] Cox, D. R. and Wermuth, N. (2004). Causality: A statistical view. International Statistical Review, 72(3):285–305.
- [Dawid, 2002] Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. International Statistical Review, 70(2):161–189.
- [Edgeworth, 1881] Edgeworth, F. Y. (1881). Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences. London: C. Kegan Paul and Co.
- [Fisher, 1918] Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh, 52:399–433.
- [Fisher, 1922a] Fisher, R. A. (1922a). The goodness of fit of regression formulae, and the distribution of regression coefficients. Journal of the Royal Statistical Society, 85(4):597–612.
- [Fisher, 1922b] Fisher, R. A. (1922b). On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222:594–604.
- [Fisher, 1926] Fisher, R. A. (1926). The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain, 33:503–513.
- [Fisher, 1935] Fisher, R. A. (1935). The Design of Experiments. Edinburgh: Oliver and Boyd.
- [Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, D. B., Vehtari, A., and Rubin, Donald B., 2014] Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, D. B., Vehtari, A., and Rubin, Donald B. (2014). Bayesian Data Analysis. Texts in statistical science. Boca Raton : CRC Press, 3 edition.
- [Greenland et al., 1999] Greenland, S., Pearl, J., and Robins, M. (1999). Causal diagrams for epidemiologic research. Epidemiology, 10(1):37–48.
- [Hardy, 1908] Hardy, G. H. (1908). Mendelian proportions in a mixed population. Science, 28(706):49–50.
- [Hoyer et al., 2008] Hoyer, P. O., Shimizu, S., Kerminen, A. J., and Palviainen, M. (2008). Estimation of causal effects using linear non-gaussian causal models with hidden variables. International Journal of Approximate Reasoning, 49(2):362–378. Special Section on Probabilistic Rough Sets and Special Section on PGM’ 06.
- [Imbens and Angrist, 1994] Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. Econometrica, 62(2):467–475.

- [Imbens and Rubin, 2015] Imbens, G. W. and Rubin, D. B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
- [Jöreskog, 1970] Jöreskog, K. G. (1970). A general method for analysis of covariance structures. Biometrika, 57(2):239–251.
- [Lauritzen, 2000] Lauritzen, S. (2000). Complex stochastic systems, chapter Causal inference from graphical models, pages 63–107. London: Chapman and Hall.
- [Lindley, 2002] Lindley, D. V. (2002). Seeing and doing: the concept of causation. International Statistical Review, 70(2):191–214.
- [Lucas Jr, 1976] Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy, 1:19–46.
- [Mendel, 1866] Mendel, G. J. (1866). Versuche über pflanzen-hybriden ("experiments on plant hybridization"). Proceedings of the Natural History Society of Brunn, 4:3–47. For the English translation, see: Druery, C. T.; Bateson, William (1901).
- [Morgan and Winship, 2014] Morgan, S. L. and Winship, C. (2014). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Analytical Methods for Social Research. Cambridge University Press, 2 edition.
- [Neyman et al., 1990] Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. Statistical Science, 5(4):465 – 472. 元々ポーランド語で書かれており, 1990年にDabrowskaとSpeedが翻訳した.
- [Neyman, Jerzy and Pearson, Egon S., 1928] Neyman, Jerzy and Pearson, Egon S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. Biometrika, 20A(1):175–240.
- [Neyman, Jerzy and Pearson, Egon S., 1933] Neyman, Jerzy and Pearson, Egon S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231:289–337.
- [Pearl, 1984] Pearl, J. (1984). Heuristics. Addison-Wesley.
- [Pearl, 1988] Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann.
- [Pearl, 1995] Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4):669–688.
- [Pearl, 2009a] Pearl, J. (2009a). Causal inference in statistics: An overview. Statistics Surveys, 3:96–146.
- [Pearl, 2009b] Pearl, J. (2009b). Causality: Models, Reasoning and Inference. Cambridge University Press, 2 edition.
- [Pearson, 1900] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302):157–175.
- [Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1):41–55.
- [Rubin, 1974] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5):688–701.
- [Rubin and Little, 2019] Rubin, D. B. and Little, R. J. A. (2019). Statistical Analysis with Missing Data. Wiley Series in Probability and Statistics. Wiley, 3 edition.
- [Simon, 1957] Simon, H. (1957). Models of man; social and rational. Wiley.
- [Spearman, 1904] Spearman, C. (1904). 'general intelligence,' objectively determined and measured. he American Journal of Psychology, 15(2):201–293.
- [Student / Gosset, 1908] Student / Gosset, W. S. (1908). The probability error of a mean. Biometrika, 6(1):1–25.
- [Wright,] Wright, S. Correlation and causation.
- [Wright, 1918] Wright, S. (1918). On the nature of size factors. Genetics, 3(4):367.
- [大塚淳, 2012] 大塚淳 (2012). 因果と実在 judea pearl, causality, 第二版 書評. 科学基礎論研究, 39(2):59–65.
- [安藤洋美, 1995] 安藤洋美 (1995). 最小二乗法の歴史. 現代数学社.
- [清水昌平, 2010] 清水昌平 (2010). 構造方程式モデルによるデータ生成過程の学習 - 特に非ガウス性の利用. Web上に公開されているスライド. 第13回情報論的学習理論ワークショップ (IBIS 2010).

- [甘利俊一 et al, 2002] 甘利俊一, 狩野裕, 佐藤俊哉, 松山裕, 竹内啓, and 石黒真木夫 (2002). 多変量解析の展開：隠れた構造と因果を推理する, volume 5 of 統計科学のフロンティア. 岩波書店.
- [芝村良, 2004] 芝村良 (2004). R. A. フィッシャーの統計理論：推測統計学の形成とその社会的背景. 九州大学出版会.
- [豊田秀樹, 2007] 豊田秀樹 (2007). 共分散構造分析 [理論編]. 統計ライブラリー. 朝倉書店.
- [赤池広次, 1976] 赤池広次 (1976). 情報量規準 aic とは何か. In 数理科学 153 号, pages 5–11.
- [長屋政勝 et al, 1999] 長屋政勝, 金子治平, and 上藤一郎 (編著) (1999). 統計と統計理論の社会的形成, volume 1 of 統計と社会経済分析. 北海道大学図書刊行会.