No.146

人間的自由と確率
―生産的実践における確率の意義に注目して―

高嶋裕一

2020 年 4 月 13 日

人間的自由と確率
——生産的実践における確率の意義に注目して——

高嶋裕一*[1]

2020 年 4 月 13 日

*[1] 岩手県立大学総合政策学部

確率の哲学について、思いつくままに挙げてみると、1) なぜ確率の見方 (哲学) に主観説と客観説の対立があるのか。2) なぜ確率論の起源が Pascal の賭けの問題にあるとされるのか。それ以前に確率学説が発達しなかったのはなぜか。3) なぜ Keynes は確率論をもってその学究生活を始めたのか。4) 近年のベイズ統計学の流行は何を意味するのか。5) 確率論と情報理論はどのような関係にあるのか。6) マルクス主義者は確率をどのように扱ってきたか、など一連の問題がある。これらの問いに答えるには、a) 一方では確率にかかる学説史をイデオロギー批判の問題として取り扱うことが枢要であり、b) 他方では確率論というイデオロギーそのものをプロレタリアートの立場から今日的に再構成することが必要である。

本稿は、後者 b) の「生産的実践における確率」について、簡単な素描をこころみたものである。そのために、生産的実践の内的論理をつかみとろうとした武谷 (1968) の〈技術論〉を手掛かりとする (第 2 章)。それに引き続く第 3 章と第 4 章で確率の存在論的側面と認識論的側面をそれぞれ記述する。第 5 章ではやはり確率の存在論的記述に属するが、測度論敵確率論の意味と確率の客観性について、物理学のトピックスを中心に論述する。最後に以上の成果を踏まえて、前者 a) の「確率学説にたいするイデオロギー批判」にかかる幾つかの論点を提示する (第 6 章)。第 7 章では冒頭に掲げた諸問題への解答を与え、今後検討されるべき研究課題を述べる。

本稿の最も強い主張は、確率の二面性 (客観説と主観説の対立) は、単純にいずれか一方が他方を打ち負かすという性格のものではなく、より高い水準での統一的理解を約束する二つの方位基点と見られるべき、というものである。またこの解明は、自由 (目的の生産) とは何か、というより実践的な課題に直結するということも付け加える。

キーワード：確率の哲学、史的唯物論、生産的実践、技術論

目次

第 1 章

問題意識

1.1　確率の哲学にかかる問題

確率の哲学について、思いつくままに挙げてみると以下のような一連の問題がある。

1). なぜ確率の見方 (哲学) に主観説と客観説の対立があるのか。「確率は測度のひとつである」と言明することがこの対立にいかなる光明も与えないのはなぜか。
2). なぜ確率論の起源が Pascal の賭けの問題にあるとされるのか。それ以前に確率学説が発達しなかったのはなぜか。
3). なぜ J.M.Keynes は確率論をもってその学究生活を始めたのか。そのことは彼の経済学説とどのような関係があるか。
4). 近年のベイズ統計学の流行は何を意味するのか。
5). 確率論と情報理論はどのような関係にあるのか。
6). マルクス主義者は確率をどのように扱ってきたか。その自由論と確率論はいかに関係するか。

これらの問いに答えるには、一方では確率にかかる学説史をイデオロギー批判の問題として取り扱うことが枢要であり、他方では確率論というイデオロギーそのものをプロレタリアートの立場から今日的に再構成することが必要である*[1]。前者のイデオロギー批判の観点では、確率の学説史の内的な構造 (学説の継承関係) を解剖し、かつ、個々の確率学説のよって立つ社会的な生産様式 (技術史との関係を含む) との関係を論じなければならない。後者においては、人間の生産的実践、社会的実践の構造を踏まえて、確率の本性をとらえることが必要であり、またそこでは人間的自由とは何か、何であるべきか、を問わねばならない。前者の問題は史的唯物論の領域に属し、後者は自然弁証法の領域に属す。

本稿は、後者の「生産的実践における確率」について、簡単な素描をこころみたものである。そのために、生産的実践の内的論理をつかみとろうとした武谷 (1968) の〈技術論〉を手掛かりとする。また、そこでの成果を踏まえて、前者の「確率学説にたいするイデオロギー批判」にかかる幾つかの論点を提示する。

1.2　確率はなぜ生産的実践にかかわりがあるか

確率を論ずるにあたって、これをただちに生産的実践 (労働) と結びつけるのは、人によっては奇妙に思われるかもしれない。確率と聞いてわれわれがまっさきに脳裏に思い浮かべるのは、ゲーム、遊び、ギャンブル、株式証券の売買などであり、あるいは災害、事故、故障、病気、保険などである。

しかしながら、われわれが無意識のうちに確率と結びつけているこれらのものは、まず第一に成功 (勝ち) と失敗 (負け) にかかわることであって、これは原初的な

*[1] 戸坂 (1934)、イデオロギーの二重性にかかわる議論を参照のこと。戸坂は文化について (特に文学作品を念頭において)、二つのモメントの区別を説いている。ひとつはその作品に固有の文学的価値であり、「普遍的な人間性に訴える高さ」によって測られるとする。もうひとつは政治的価値 (階級性) であり、社会的・歴史的存在としての価値のことであるとされる。この二つのモメントは、たんに対立的な関係にあるのではなく、ひとつの作品として統一されていることが指摘される——この点は、資本論「商品に対象化されている人間労働の二重性」の論理が想起されよう——。戸坂は同じ関係が、科学を含むイデオロギー一般にも成立すると論じている。

労働である狩猟・採集・栽培・牧畜につきもののことである、と言えないだろうか。われわれの祖先はその成否に喜び (「効用」) を見出し、あるいは恐怖に震え、落胆したであろう。リスク[*2]への感覚は喜怒哀楽と結びつき、労働の発展とともに洗練されていったに違いない。そして純粋にその感覚を楽しむものとして、偶然をともなう遊び、ゲームやギャンブルが後から発明されたのだろう。現代に生きるわれわれにとって、生産的実践はリスクの存在を感じさせないほどまでに高度化している[*3]。だから、今日では生産的実践を直接的に確率の考え方に結びつけることが奇妙に感じられるのである。ただそれは、われわれがリスクをある程度飼いならし、その存在を普段は忘れているからにすぎない。

また、第二に、天上で起きていること (天候など) と、地上での労働の成否とのあいだに相関を読み取り、またそれが実際に起きるに先んじて予見する試み (占い・迷信など) もなされたであろう。ここからも、事象のあいだの確率的な因果関係にかかる認識が生まれたと想像される。そこでも、関心の焦点となるものはこれから行われる生産的実践の成否であり、今という時がそれに適切かどうか、少しでも成功の見込みが大きいのはいつなのか、という問いであろう。このように考えるならば、科学と魔術が未分化であった古来より確率の考え方は自然のことだったと推測されよう。

ここからわれわれは次のような論点を引き出すことができる。

- 生産的実践のなかからリスクが追い出されている今日の現実。リスクのある研究開発は大学や研究機関など生産的実践の外側へと外化され、生産的実践そのものは計画的に、寸分の狂いもなく実行されている。われわれは生産的実践のただなかにおいて、失敗する機会を与えられず、したがって成功する喜びも与えられない。生産過程の外の流通、消費過程へとリスクは追いやられ、その反面われわれ自身の消費生活は動物的なもの・粗野なもの、偶然的でとるに足らないものにまで貶められている[*4]。
- 職場に就く機会にリスクが存在すること。生産的実践の場 (職場) に就くこと、離れることにつねにリスクがともなう。われわれは自分の希望する職場に就けるとはかぎらず、就けたとしてその判断の過ちにあとから気づく。さらにその職場にみずからが希望するかぎり長くいられる保障はない。また職場の側も、期待する新たな才能にめぐり合えるとはかぎらず、それどころか応募がまったくないかもしれない。
- リスクへの感覚が情動と結びつくらしいこと。一方ではリスク・テイクを好む性格、他方ではできるかぎりリスクを回避しようとする性格があり、こうした心理は脳の生理学的機能にかかわりがあるとも思われる。またそこから、確からしさの判断にかんする歪み (パラドクス)、錯覚などの存在も了解されよう[*5]。また、そうだとすればそのことが「効用」の考え方やモノの価値判断にまで影響しているとしても何ら不思議なことではない[*6]。
- 確率に関する人類の関心がどうやら古いものであると想像されること[*7]。それにもかかわらず、それが研究の対象になったのは比較的新しいとさ

[*2] ここでは「リスク」という言葉を「失敗 (成功) するかもしれないということ」と理解しておく。一般にはこの用語を否定的 (肯定的) 感情と結びつける必然性はない、とされているが、われわれはそうした通説とは異なり、この用語と感情とのつながりに十分な注意を払う。

[*3] ただし災害による物流の寸断、為替変動による原材料費の高騰など、生産的実践の外側にあるリスクは容易にその内部に入り込む。

[*4] マーケティング科学において、消費者の衝動的な行動が分析対象とされはじめている。これは近年のプロダクト・ライフサイクルの短期化が外観の重要性を増大させていること、これまでも非計画的（衝動的）な購買行動の比率は高いと目されてきたが、これを具体的に分析する手段が未発達であったことなどが背景にある。

[*5] これらの研究は今日では「実験経済学」として知られるようになった。これは「期待効用理論」に対する反措定であると言える。

[*6] 経済学の根幹が「財の希少性」にあるとされていること。これは価値法則のひとつの現れである。採取経済においては収穫が得られる見込み (確率) は、探索のために費やした時間にかかわる。資源が豊富であれば (空気のように)、その獲得に必要な労働時間は無に等しい。他方で、希少な鉱産物はその発見に膨大な手間と時間を要し、しかも時間をかけさえすればそれが確実に得られるという保証もない。また希少な産物は遠方から運ばれ、その途中の事故などにより失われる危険がある。

[*7] 賭博行為は占いから派生したものと推測され、またじっさいに賭博行為の禁止が古代ギリシャに出されていたことからもその歴史の古さを証明する。ローマ帝国が崩壊し、その知識はアラビア数学に移され、その後十字軍遠征を経てイタリア・ルネサンスにもたらされた。またその知識は当時発達しつつあった海運業でも保険計算で役立ったのではないかとも推察される。

[*8] 実はそれ以前に Cardano の著作がある。本多 (1989) を参照のこと。Gerolamo Cardano(1501-1576) は Leonardo da Vinci の友人の

れること。Laplace は、確率論の出発点を Pascal と Fermat の往復書簡 (1654 年) に求めている*8。

以上のように、生産的実践とのかかわりの中で確率の本性について考察することは、われわれが普段は忘却しているところの多くの有益な論点を提供する。逆に言えば、確率の哲学についての先に挙げた謎の多くは生産的実践という視点の喪失によってもたらされている可能性が暗示される。

以降の展開は、上の観点を可能なかぎり貫くという方針に従う。すなわち、生産的実践の論理を明らかにし、そのなかでどのように確率の本性を取り扱うべきかについて議論する。このため、次章では生産的実践＝技術の本質を明らかにした武谷三男の業績をとりあげる。

1.3　本稿の構成と記述上の留意点

本稿の構成は以下の通りである。

第 2 章では生産的実践における確率の意義を明らかにする。そのために主に武谷 (1968) より認識の三段階論と技術本質論の内容を整理し、これを踏まえて認識における実験の意義と確率を論じる。

第 3 章では、確率の存在論的把握を試みる。ラプラス (1997) に依拠しながら 3.1 節では古典的確率論を議論し、とりわけその〈充足理由律〉の疑わしさを指摘する。3.2 節では独立性と条件付き独立性を中心に事象間の関係性について具体的に論じる。そこでは Laplace の強調した母関数の方法に注目する。3.3 節では、独立性 (相関性) と因果性の違いと両者の関係を主に Pearl(2009)、Edwards(1995) によって議論する。

第 4 章では確率の認識論的側面について、予測とモデル選択に関わるトピックスを中心に取り扱う。4.1 節では KL 情報量と最尤法の関係、赤池情報量規準を最尤法の拡張として見る見方を小西・北川 (2004)、坂元・石黒・北川 (1983) などに従って整理する。4.2 節と 4.3 節では 2 種類の観測の限界を取り扱う。前者では、欠損値と潜在変数 (隠れ変数) の問題を取り上げ、その克服法としてのベイズ・モデリングとその具体的解法としての EM アルゴリズムを赤池 (1989)、ビショップ (2012) などから論じる。後者では異常値とそれにもとづく偏りの問題に注目し、その対策としてブートストラップ法と経験尤度法をそれぞれ Efron and Tibshirani(1994)、Owen(2001) をもとに整理する。

第 5 章では確率過程に関連して測度論的確率論を取り上げる。5.1 から 5.2 節では伊藤 (2007)、重川 (2016) などに依拠してブラウン運動の変数変換に関する伊藤の公式の意義を説明する。5.3 節では確率微分方程式と量子力学の関係について長澤 (2015) の理論を概説する。

第 6 章ではそれまでの展開を踏まえて確率論のイデオロギー批判を試みる。6.3 節までは主に安藤 (1997) に依拠して古典的確率論の形成までを取り扱い、6.4 節以降はギリース (2004)、傾向説については高村 (2010) に依拠して現代確率論の 4 学説のそれぞれの特徴を議論する。

第 7 章では結論を述べる。本書の冒頭で挙げたいくつかの問題についての回答を与える。本稿で述べる結論はすべて先人たちの先行研究の功績の上に立つものであり、著者のオリジナルの考えはわずかな比重を占めるのみである。しかしこのような形での統合を図った試みはこれまでなかった。

付録では本文の理解を助けるであろう断片的な考察 (ただし本文に入れるには不相応に大きく、論旨の流れを不明瞭にしかねない可能性がある) を追記した。付録 A は商品論と認識の三段階論との関わりを整理している。「資本論」に書かれてあることをなぞったものに過ぎないが、自由と計画の真の関係が誤解されている今日ではあえて記述する価値があると考える。付録 B は唯物論的目的論について、自由論の基礎付けのために整理している。付録 C は第 3 章から第 5 章に出現する計算と論証のいくつかを一括して数学的付録とした。付録 D は作図、計算の具体例を R で再現するためのプログラム・リストを載せた。

＊　＊　＊

本稿の記述上の留意点として、以下を挙げる。

- 人名について、和名以外は原則としてラテン文

息子で、イタリアの数学者・医師。三次方程式の解法発見で知られる。「さいころあそびについて」(”Liber de ludo aleae”) は死後 1663 年に出版されたのであり、その執筆はおそらく Pascal = Fermat 往復書簡より早かった。(ただ、そうであったとしても、確率研究の歴史が新しいことには変わりない。)

字を、和名については漢字を用いた。ただし、1) 著者名として引用しているばあいは、参考文献リストの表記 (版元による表現) にしたがった。2) 人格としてではなく慣用表現の一部 (例：ニュートン力学) については和名以外はカタカナで表記した。

- 参考文献からの引用につき、文章量に応じて本文中にかぎ括弧つきで示すばあいと、段落として独立させるばあいに分けた。いずれにおいても省略は記号「・・・」で示した。それらの省略はいずれも断らない限り引用元の著者によるものではなく、本稿の筆者によるものである。引用中でロジックが入り組んでおり、そのままでは読解が困難と思われる個所に適宜丸括弧を用いて論旨を補った。これにより引用の正確さを損なうことになったが、避けられないものと判断した。この補填も本稿の筆者によるものである。
- 本文中で語句を強調するために二種類の括弧「a」、〈b〉を用いた。前者は消極的あるいは便宜的な用語法であることを示すためのものであり、後者は積極的あるいは唯物論哲学に特有の用語法であることを示している。

第 2 章

生産的実践における確率

2.1　武谷〈技術論〉

ここでは武谷〈技術論〉の追体験的再構成をこころみる。以下の記述において、武谷 (1968) 所収の諸論文を次のように表記する。

[T-1] 自然弁証法、空想から科学へ―自然科学者の無遠慮な感想―, 唯物論研究,1936
[T-2] 自然の弁証法 (量子力学について)―問題の提示―, 世界文化,1936
[T-3] ニュートン力学の形成について, 科学,1942
[T-4] 技術論―迫害と戦いし知識人にささぐ―, 新生,1946
[T-5] 実験について,(媒体なし),1946
[T-6] 自然弁証法について, 学生評論,1947
[T-7] 自然の論理について, 思想,1947

2.1.1　諸前提

武谷の〈技術論〉(T-4) は 1944 年に治安維持法違反で検挙されてのち、1945 年に遺書としてつづられたとされる特高調書を補筆して 1946 年に公表されたものであるが、武谷自身の述懐によれば 1940 年の末頃にはこの立場に到達していた[*1]。まずは、技術の〈本質論〉を構築するうえで武谷の脳裏にあったであろう諸前提を列挙する。

1. 技術 (ないし技術性) は、自然と社会を媒介するものとして、社会における生産のあり方 (「人間の実践の根本」) に関わること。それゆえ、技術は自然科学的にばかりではなく、社会科学的にも取り扱われねばならないこと。
2. 人間の自己疎外からの脱却 (止揚) において、疎外された技術 (〈現象〉と〈実体〉) もその〈本質〉とともに人間に取り返されねばならないこと。それゆえ、技術の〈本質〉が、疎外されざる人間労働におけるそれとして明らかにされるべきこと。
3. 従来の技術論が、いずれも〈現象論〉あるいは〈実体論〉の水準にとどまっていること。その水準にとどまる限りでは、上の課題を果たすことができないこと。
    - 現象論的段階:1) 科学者が科学と技術を日常的に区別しないこと。2) やはり日常的には技能と技術が区別されず、勘や「行為のかたち」(三木清) のようなものとされること。「幼稚な観念論」への落ち込み。
    - 実体論的段階:資本制社会の現実 (とりわけ産業革命において労働手段の発達が果たした役割) を固定的にとらえ、そこからただちに「労働手段の体系」(相川春喜) という技術規定をあたえてしまっていること。労働対象や労働

[*1] 当時の日本は次のような時代経過をたどっていた。日独伊三国同盟の調印 (1940.9.27)、太平洋戦争の開戦 (1941.9.27)、ドイツ降伏 (1945.5.7)、広島への原爆投下 (1945.8.6)、日本降伏 (1945.8.15)。その日本降伏の熱気のなか、民主主義科学者協会が設立された (1946.11.3)。武谷は、日本の敗戦という事態を「民主主義革命」と受け止めつつ、「勤労者による経営管理」に資する〈技術論〉を公表するという意欲に燃えていたと言える。なお、武谷が何の疑問も持たずに日本の敗戦をそのまま「民主主義革命」と受け止め、またその読者にも当然のように了解された理由は、戦前の知識者の抵抗運動がコミンテルン指導の影響のもとにあったことを挙げねばならない。当時のコミンテルンは、既に国際共産主義運動の拠点としては決定的に変質しており (だからこそ、1938 年にはソ連邦を追放されていた L.Trotsky 等が、ファシズムの台頭を許したコミンテルンの腐敗を暴くために第四インターナショナルを創設した)、いわゆる「32 年テーゼ」が暗黙のうちに種々のイデオロギー闘争 (「講座派」を含む) の前提とされていた。

そのもの (マニュファクチュア) の技術性の無視。素朴な機械的唯物論への先祖返り。

従来の技術論を評価するに際して、武谷自身の〈認識の三段階論〉(本質論—実体論—現象論) が積極的に援用されている。逆にいえば、武谷がこの時点でこの理論を獲得していたことが、技術の〈実体論〉を超えて〈本質論〉の段階に進まねばならない、という指針につながった。

2.1.2 認識の三段階論

武谷の慧眼は、〈本質〉と〈現象〉を媒介するものとしての〈実体〉に着目し、これに認識方法における独自の段階という地位を与えたところにある。

> 物理学の成果は〈本質〉と〈現象〉の弁証法的な深い理解にある。方程式をつくる前にそこに何があるか、いかなる交互作用のもとにあるか、という立体的な構造、すなわちモデルを知らねばならない。たんなる現象論ではないのである。(T-2)

1936 年の時点で、武谷は当時の唯物論哲学について、1) 党派性の名の下に横行する (とくに「自然弁証法」についての) 文献解釈主義 (中世的な宗派性のにおいのする)、2) 象徴 (シンボル) 主義への逃避、3) 認識論の鏡的反映論への堕落、4) 認識と実践との関係が見失われていること (これが党派性の度外れの強調につながっている) を批判している (T-1,T-2)。

象徴主義については、一般の社会不安とないまぜになった経験主義、マッハ主義、懐疑論、不可知論などの流行、「方程式がすべてだ！」という風潮となってあらわれていた。これにたいして、武谷は数学それ自体の弁証法的な性格を対置した。

> 数学自身ですら自然の弁証法の豊富な反映なのである。それゆえにこそ悟性的思惟が及びもつかない自然の内部につきすすむことができるのである。数学的方法が反映したものはかくて実在的意味を持つ。(T-2)

また、実は上の批判そのものが三段階論を発想する「導きの糸」となっている。すなわち、〈本質〉としての運動方程式が二階の微分方程式として記述され、〈現象〉としての積分曲線 (解集合) を得るためには二つの積分定数 (物体の位置と運動量) を必要とすること、これらを〈実体〉とするならば、必然的に成立する運動方程式にたいしてそれらは偶然的なもの (初期条件、周辺条件) にとどまること、などが了解される。

また、唯物論的な認識論について、当時の主流の考え方が鏡的反映論に陥っていることを批判している。

> ・・・観測を措いては認識はないが、観測はただちに認識ではない。これを混同するところに混乱が生ずる。・・・認識は〈模写〉である。だがこの〈模写〉は鏡が物を写すような死んだ静的な反映ではなくして、認識はますます深く本質へ本質へと進む。また歴史的に言って〈模写〉と対象の合致の過程である。これは主観の恣意的な「制作」[*2]ではないのである。(T-2)

認識された内容がたんなる鏡像ではない以上、認識内容 (I.Kant の「物自体」) をわがものとすることは実践の内部においてしかありえない。逆に言えば、認識は実践の一つのモメントとされる。

> 自然科学が自然に奥深く突き進むごとに、以前の表象に無縁 (fremd) なものにぶつかる。われわれがこれを自己のものだと感ずるには繰り返しての実践を要するのである。思惟によって到達された本質的な概念は、実践を通して対象の正しい反映であることが示されるのである。(T-2)

興味深いことに、実践と〈客観的な法則性〉とのかかわりについて、既にこの時点で後の技術の本質規定に極めて似た表現が用いられている。

> 主体性とは・・・客観的な必然性を根拠としての実践なのである。(T-2)

当時の唯物論的認識論にたいする上のような批判は、〈技術論〉構築の中でも意識されている。

> ・・・技術の論理も高度な弁証法の媒介の論理を鍛えなければ把握し得ないものである。逆にまた技術の論理を把握することによって論理学が高度化される・・・(T-4)

1942 年の「ニュートン力学の形成について」(T-3) では、三段階論の内容が整序され、力学史への応用というかたちで表明される。

> 科学と技術との関連、科学と生産との関連については、われわれは現在重要なる問題に直面している。そ

[*2] おそらく三木清の「構想力の論理」への批判を念頭においた表現と思われる。

してこれについて多くのことが言われてきた。例えばニュートンのプリンキピアはニュートンの言葉やその意識に反して技術的要求のもとに生まれたという説、これに反しニュートンのプリンキピアは技術への関心に基づくものではなく、あくまで自然認識を目的としたという説がある。(T-3)

科学史研究の前提として、科学と生産的実践とのあいだの本質的関連性が示される。これは科学の〈本質〉にかかわることであり、その悟性 (形式的な思考能力) を超え出る力の源泉でもあるとされる。

科学は技術と本質的にいかなる関連にあるか。・・・科学は有力であり、そしてしばしば悟性を否定する方向にしゃにむに進むこと、これは科学が生産的実践との関連の下にあることを示している。(T-3)

われわれの実践はたんに主観的なものではなく、必ず客観的な外的自然法則性の場において行われるものである。すなわち技術的行為なのである。それゆえに実践が有効であるのはその法則性が正しくつかまれ、適用されたかぎりにおいてである。このような人間実践の歴史を通して自然の知識が人間にもたらされてきた。(T-3)

ただし、科学の〈本質〉に引き続いてただちにその〈現実〉の姿について注意があたえられる。科学はイデオロギー (文化) として内的な構成を保ち、技術的要求から相対的に独立して発展する。ここから現実の (自然) 科学史を規定する三つの契機、〈技術的地盤〉、〈自然〉の構成、〈思惟様式〉が説明される。

技術的要求という事は科学の発展にとって・・・規定的なものではなく、それは偶然的なものを科学自身の構成から見れば持つわけである。そして自然科学がこのような技術的要求の下にのみあるならば、断片的な技術的知識にとどまり、自然の本質的な解明に進まず、より進んだ人間実践において無力となるだろう。・・・これは科学が一つの文化であることによるものである。それは一つの文化として世界観の一部をなし、それ自身独立な体系をなしその道を進むのである。(T-3)

科学は・・・技術と〈自然〉自体の構成と〈思惟様式〉の三者に規定されて進むものである。そのばあいに技術との関連も本質的には〈技術的地盤〉という形をとるのである。(T-3)

〈技術的地盤〉が中世的世界観にたいするイデオロギー闘争 (当然ながら、これは階級闘争の一部をなす) のための武器を提供したことが説明される。

ガリレイの物理学の建設は・・・市民社会の世界観がアリストテレス的な中世的世界観にたいする攻撃の一翼として・・・出現したのである。そして・・・それを可能にしたのは、アリストテレス的中世的自然科学が技術からの遊離によって虚弱となっているのを、まさに〈技術的地盤〉によって武装して突くことによってであった・・・。(T-3)

〈思惟様式〉は〈生産様式〉と連関しつつも、それと単純な一対一の対応をなすものではないことにも注意があたえられる[*3]。

ギリシャ科学はギリシャ的〈思惟様式〉に、ルネサンスにおける科学の形成と発展はルネサンス的思惟に依存している。しかし〈思惟様式〉というものを一つの時代、一つの〈生産様式〉に対応させて考えるときにははなはだしく誤ることがある。生産の様式は直接に一つの〈思惟様式〉を規定するものではないからである。(T-3)

あらためて〈実体論〉が説明されるが、その理解はT-2 でのそれから格段に進歩している。すなわち、この段階の論理の特徴が「スピノザ的論理」であることが明示されている。

物理学においてはまず対象の構造、すなわちいかなる物があるかということを明らかにしなければならない。一般に認識において最初に物の概念、すなわち〈実体〉概念をわれわれは獲得するものである。そしてその物の概念と同時に、そのものに属している性質をわれわれは認めることになる。〈実体〉とその属性の概念は、我々が対象を扱う場合に有する極めて根本的な方法である。例えばスピノザの「エチカ」のように、〈実体〉を根本において、〈実体〉と発現との関連を〈実体〉の属性として見るのである。属性は〈実体〉と発現を対応として見るので、〈実体〉の働きから具体的に発現を媒介しない静的な見方である。

[*3] 因果性の論理について T-6 で興味ぶかい整理が見られる。すなわち、エジプト＝神秘的な因果性、ギリシャ (Pythagoras、Plato) ＝形而上学的な因果性 (形式論理)、ルネサンス＝科学的な因果性 (弁証法の論理)。また、実証主義については「感覚論的なものであって、対象にどんなものがあるかということは無意味である、とするもの」と否定的に評価されている。

[*4] ただしヘーゲル概念論について T-6 では Hegel 的観念論に落ち込むことを警戒して次のように注意されている。「概念自身がヘーゲルの言うように自己発展するのではない。自然自身の弁証法的構成にその根拠がある・・・。自然弁証法は自然自体の弁証法である。」最後の一

以上のことを受けて、認識の三段階論が整理される (表 2.1)。ヘーゲル概念論*4との対応がつけられ、ニュートン力学の形成を実例として説明される。

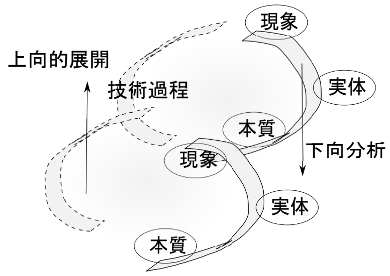


図 2.1　認識の連関

三段階をなす認識がひとつの「環」になぞらえられ、一連の「環」の連関がさらなる認識の深まりとして説明される (図 2.1)。先行する「環」の到達点である〈本質論〉は、次の「環」においては〈現象論〉として位置づけられる。

> ・・・物理学的認識は「ますますどうなる」というように一律に進むのではなく、この三つの段階の環を繰り返して進むのである。すなわち一つの環の〈本質論〉は次の環から見れば一つの〈現象論〉として、次の環が進むという具合である。(T-3)

〈現象論〉の積極的な意義づけと、それを見誤ることによる誤びゅうが説明される。これには次の二種類がある。

1. 〈現象〉にかんする知識が十分ではないにもかかわらず、性急にその〈本質〉を探ろうとして形而上学に陥る誤びゅう
2. たんに〈現象〉の記述の段階にとどまり、それを固定化する誤びゅう

〈実体論〉から〈本質論〉への移行における三つの形態が以下のように説明される。ここでは科学史において「失敗」とされるフロギストンやエーテルなどの例についても、かならずしも誤りではなく〈実体論〉の観点から肯定的な評価があたえられる。

表 2.1　認識の三段階論 (T-3)

段階	説明	ヘーゲル概念論との関係	例
現象論的段階	現象 (観察結果) の記述。個別的な知識を集める段階。	個別的判断 (Dasein の肯定的な判断)、an sich	Tycho Brahe(1546-1601)
実体論的段階	現象が起こるべき実体的な構造を知り、この構造の知識によって現象の知識が整理されて法則性を得る段階。その法則性は実体の属性として意味をもつ (Spinoza 的論理)。	特殊的判断、für sich	Kepler(1571-1630)、Galilei(1564-1642)
本質論的段階	諸実体の相互作用として法則性が認識される段階。実体的契機によって実体を含みながら、実体的な法則の見方を否定し、その認識に固有の論理的性格をあらわす。現象はこの相互作用の下における実体の運動を媒介として説明される。	普遍的判断、an und für sich	Newton (1642 - 1727)

文で、自然弁証法＝自然自体の弁証法、としてしまっているのは、後述のように武谷において〈認識論〉と〈存在論〉の区別がしばしばあいまいにされ、「因果関係のとらえかた (論理)」と「自然の構成」とが同一視されている弱点を同時にあらわしてしまっている。

1. 〈実体〉の導入が直ちに〈本質論〉に導く場合 (〈実体〉が新たな性質のものでない場合)

   例 : 海王星の発見、立体化学、物質構造論
2. 〈実体〉が全く機能的なものに解消される場合 (機能を〈実体〉として捉えていた場合)

   例 : フロギストン、エーテル
3. まったく新しい〈実体〉であって、新しい論理を要求している場合

   例 : Newton の力の概念、素粒子

なお、〈実体〉という用語について、「資本論」冒頭の商品論に典拠があることが T-7 で示されている。これについては付録 A を参照のこと。

> 私が使用した〈実体〉(Substanz) と〈本質〉(Wesen) という概念について・・・資本論・・・には価値と、価値の実体 (Substanz der Werte) をなす労働ということが語られ、両者が区別されてある。(T-7)

T-6 は戦後の 1947 年時点で、学生団体の求めに応じて架空の問答の形式で叙述されている。そこでは、その当時に公刊された三木清の手記の批判[*5]と山田坂仁への反批判 (「技術論論争」) を主題として、とくに三段階論について多く疑問を寄せられた点について、解説している。

ここではとくに 1) 予測と自然弁証法との関係、2) 自然弁証法における認識論と存在論の区別が重要と思われる。前者については、三木清の自然弁証法否定論に対応して予測というものをどのようにとらえるべきかが論じられる。またこの点は「実験について」(T-5) につながる論点[*6]も提示されている。後者については、学生たちが認識論と存在論の区別 (この論点は目的意識の形成にかかわる) を執拗に突いているが、武谷の回答は彼らの問題意識にかみ合っていない。

まず三木の主張 (自然弁証法否定論) は次のようにまとめられる。

1. 近代自然科学はその方法においては機械論的であり、その法則は機械的法則である。「数学化」とは、質を量に還元することであり、機械的な論理にほかならない。
2. 弁証法は予測を与えない。その理由は弁証法的な変化が「飛躍」「総合」であり、矛盾の統一 (新しい、予測のできぬもの) としてあらわれるからである。

これにたいして武谷は次のように答えている。

1. 機械的と見られる連続的な運動ですら、各瞬間の「飛躍」を自らのうちに持っている (Zeno の逆理、微分学)。これは弁証法の一つのあらわれと見るべきである。
2. 自然を予測することができる、ということは、自然が (必然的に作用する) 法則性をもつ、ということである。それと同時に、法則の展開において任意の初期条件を与えられるということは、偶然的な契機があることを意味する。ここに人間にとっての技術的な可能性が開示されている。

> 自然科学は質を量においてつかむ。これは社会科学においても、マルクスの資本論に示されている・・・。これが弁証法の質と量の関連なのであって、何ら質を解消してしまうということではない・・・。(T-6)

> 自然科学の弁証法ということを言う根拠は、われわれの技術論にある。われわれの技術論は、人間の実践が客観的な法則性、すなわち法則から現象が導かれる全体の構造、これを根拠として成立することを主張する・・・。そして人間の実践の可能性、その現実の発展、これが自然弁証法の観点を基礎づけて (いる)。(T-6)

この予測ということについては、唯物論的な〈目的論〉(目的意識の形成、頭脳労働の論理) とかかわりがある。認識作用は人間＝労働者にとっては、現実を説明するだけではなく、現実とは異なるもう一つの現実 (すなわち

[*5] ただし、武谷は三木清の獄中手記の一部「自然弁証法について」はかならずしも三木の真意をあらわしていないかもしれないと記している。つまり、手記は三木が思想検事にたいして自身がマルクス主義者でないことを証明 (自己弁護) するために書いたものである、と同情的に理解している。

[*6] 「条件が非常に複雑な場合、またその条件の一部しか知られていない場合、また全体の部分だけが問題になる場合、・・・予測は否定され・・・蓋然的な法則をわれわれは得ることができる。」(T-6)

[*7] 現実とそれとは異なる目標との対立を自覚することこそ、〈問題〉発見 (生産) であり、このことを指摘したのが H.Simon(1916-2001、アメリカの経営学者) であった (「意思決定の科学」(1979))。またこのような〈問題〉の捉え方は、次の Marx「経済学批判 序文」の次の言葉とも整合している。「人間が立ち向かうのは、いつも自分が解決できる課題だけである。課題そのものは、その解決の物質的諸条件がすで

目的＝目標) の可能性を見出すこと (予測) に役立てられる[*7]。頭脳内でいったん形成された意図＝目的を物質化＝対象化するところに、生産的実践の (社会を対象としたばあいには、社会的実践の) 可能性・現実性・必然性が生みだされる。

このことを了解するためには、認識作用が〈現象論〉から〈本質論〉へとさかのぼる〈下向〉の道と、〈本質論〉から〈現象論〉へと展開する〈上向〉の道との両輪からなること[*8]、哲学はこの二つの道に対応して〈認識論〉と〈存在論〉に分かれるべきことを知らなければならない。

武谷は、しかし、頭脳内での目的意識の形成とその目的の物質化＝生産的実践とを区別せずに、直接的に同一視してしまっている。またそれらを区別することを「実践における適用という実在を主観化する観念論」とまで非難している[*9]。だから学生たちが「自然自身の構成 (存在論) と認識の発展段階 (認識論) の交叉」や「帰納・演繹と三段階論とのかかわり」についていくら水を向けても、武谷には答えられないのであり、自然の弁証法性 (「自然自身の構成」) と「自然弁証法」(前者を哲学としてつかまえた内容) を同じものだとしてしまうのである[*10]。

> 認識は対象の人間の頭脳における反映であるが、実践における適用は客観的な事実の中に構成されてゆく・・・。この区別が理解できない人は実践における適用という実在を主観化する観念論に他ならない。(T-7)

なお、武谷三段階論にたいする批判については、付録 B も参照されたい。

2.1.3 技術の本質

いよいよ技術の本質が語られる (T-4)。1) 生産的実践と自由の関係、2) 自由と〈客観的法則性〉の関係、3) 生産的実践と〈目的意識性〉の関係の推論を通じて。

1. 生産的実践 (もちろん疎外されざる形態でのそれ) においてこそ、人間はみずからの自由を獲得する。このことは、生物進化との対比によってより明瞭となる。つまり進化の過程で、生物はその環境に支配されている。生物は変異と環境への適応により、その生存をかろうじて維持している。環境としての自然は生物にたいして容赦ない自然淘汰圧を加え、適者の生存のみを許す。このような生物進化において支配権を持つのは自然環境の側であって、生物自身は消極的な役割を果たすに過ぎない[*11]。

   しかし、人間は生産的実践を通じて、それまで敵対的であった環境を改変する。人間も生物の一つとして他の生物と同じ〈客観的法則性〉の支配の下にありながら、もはや環境によって淘汰されるに甘んじるのではない。人間の自然を支配する度合いは〈生産力〉と呼ばれ、その増大につれて人間は自らの生存のために必要な労働の支出を少なくし、その分だけ新たに自由を獲得する[*12]。
2. 自由をもたらすであろう生産的実践において人間が利用するものは、それまで生物としての人間を「鉄の必然性」によって盲目的に支配していた自然環境それ自体の力〈生産諸力〉である。F.Engels と V.I.Lenin によって Hegel から発見された「自由とは必然性の洞察である」[*13]という言葉は、このような解釈によってはじめて意味を

に現存しているか、または、少なくともそれができはじめているばあいに限って発生する。」

[*8] これらは従来の哲学においては〈帰納〉、〈演繹〉として知られる。Rao は、3 つの推論の論理形式として、演繹 (deduction)、帰納 (induction・・・特にこれに関連してベイズ・ルール)、アブダクション (abduction) を挙げているが、いずれも弁証法の一つの局面を切り取ったものと解釈できよう。

[*9] 武谷による三木清への批判もこの観点から見直されねばならない。三木の自然弁証法否定論にたいする武谷の批判はたしかに正当なものであるが、唯物論的な目的論の形成という問題への無理解が三木批判の視角を歪めている可能性がある。

[*10] 同じことが「法則性」と「法則」との関係についても言える。武谷はしばしば「法則性」と表現すべきところを「法則」としてしまう。

[*11] このことを人為淘汰 (農学) からの類推を通じて自然淘汰として明らかにしたのが、C.Darwin であった。

[*12] ただし、疎外された形態において〈生産力〉は容易に「破壊力」に転化する。そのことは、原子力発電所の崩壊により、あるいは戦場に打ち込まれた劣化ウラン弾などにより、放射能がまき散らされている現実を見れば明らかであろう。

[*13] この言葉を Hegel その人が発したかどうか、という不毛な議論があるが、ここではそれには立ち入らない。

持つ。武谷はここから〈客観的法則性〉というモメントが、技術の本質規定にとって必須であることを見いだす。

3. 最後に〈(目的) 意識性〉のモメントがある。これは「資本論」の労働過程の分析、「労働過程においては、人間の活動はあらかじめ目的とされた変化を労働手段によって労働対象に引き起こす。この過程は生産物の中に消え入る。」から引き出された。

こうして武谷は技術の本質規定を次のように定式化する。

> 技術とは人間実践 (生産的実践) における客観的法則性の意識的適用である。

ここで〈客観的法則性〉とは〈本質〉が〈実体〉を媒介として〈現象〉する全構造を指すとされる。ただし、生産的実践において〈客観的法則性〉が完全に認識され、明示的な法則として定立されている必要はない。実践主体が、ただ〈客観的法則性〉の存在と、それが目的を媒介しえることを承認しさえすれば、生産的実践は成立する。

2.1.4 技術と技能の区別：労働の起源と発展

技術の本質規定を受けて、技術と技能の区別が論じられる。武谷は「技術」[*14]を客観 (客体) の側にあるもの、「技能」を主観 (主体) の側にあるもの、と区分する。そこでは生産的実践における主・客の対立が措定されているのであって、客体の側 (すなわち労働対象と労働手段) の技術性を〈実体論〉的意味合いでの「技術」、主体 (すなわち労働そのもの) の技術性を同じ意味合いで「技能」と呼んでいる。「技術」と「技能」とをいずれも〈実体〉として区別していることになる。だからこそ、労働過程全体の技術性が「技術」と「技能」の統一としてつかまれ、両者の整合と相互的発展とが議論されることになる。

> 技術は客観的になるものであるのに対し、技能は主観的・心理的・個人的なるものであり、熟練によって獲得される・・・。技術はこれに反して客観的であるゆえに、組織的・社会的なものであり、知識のかたちによって個人から個人へと伝承ということが可能・・・すなわち技術は社会の進展にともない、伝承により次第に豊富化されていく・・・。(T-4)

ただし武谷には、本来は場所的・同時的に存在しなければならない「技術」・「技能」を、歴史的・異時的にずらして考えてしまう傾向がある。そのため「技能」を「技術」から見てあくまでも消極的なもの、「技術」の側から対応を要求されるものとしてしまうのである。つまり疎外された労働の現実 (ホワイトカラーとブルーカラーの分断、技術者層の固定など) をそのまま肯定的に表現する。「技術の立場」という言葉は武谷のこの「技術主義」ともいえる歪みを端的に表している。また労働そのものを労働組織と捉えていないために、「技能」の側に組織性・社会性を認めないことになる。武谷の言う「技術」と「技能」の弁証法は、疎外されざる労働過程の発展の論理として、同時に人間主体 (〈労働そのもの〉) と自然 (〈生産手段〉) 相互の技術性高度化を図る論理として読み替えなければならない。

> 労働とは技術と技能の統一において実現されうる・・・。一定の技術には一定の技能が必然的に存在して、労働を実現する・・・しかし、技術の立場というのは、つねに主観的・個人的な技能を客観的な技術に解消していく・・・。しかし、解消されることによって技能が消失するものであるかというに、けっして然らず。新たな技術には、新たな技能が要求され、これがまた再度技術に解消されながら発展していく、という弁証法的な関係をとる・・・。(T-4)

「技術」と「技能」の関係を過去にさかのぼることにより、労働の起源を論じることも可能となろう。〈生産力〉が未発達であり、客体 (生産手段) の側の技術性が乏しいという社会状態にあっては、武谷の言う「技術」の要素も見られず、〈本質〉としての技術性の大半が主体側の「技能」によって説明される。武谷が「技能」は主観的・心理的・個人的なものであると言っているのはこの意味でのみ正しい。

[*14] 技術の本質規定と区別するために、「技術」、と鉤括弧付きで表記する。このように表記した場合は、〈実体〉の属性として取り扱っているものと理解されたい。武谷においてこの区別はあいまいである。

さらにその極限、原始的労働においては、中枢神経系を備えた生物の行動とほとんど変わらないものとなるだろう。ここにおいては B.F.Skinner の発見した行動随伴性[*15]が重要な意味を持つ。原因としての行動と結果がポジティブ、またはネガティブなフィードバックで結ばれ、行動が確率的に強化、あるいは弱化される。

2.1.5　科学と技術の区別

技術の本質において、〈客観的法則性〉ははっきりとその内容を認識されたうえで、それを他者に伝えられるかたちで表現されているとは限らない。逆に言えば、認識内容を他者に伝え、かつ、その中で説得力を持たせる必要が生じて、はじめて科学 (客観的法則性の認識) と技術を区別する意味が生まれる。つまり生産的実践 (生産) が高度に組織化され、主・客両面の技術化をどのような方向性で進めるべきかを判断すること、またその論拠、を広く共同体内で共有しようとする善き意図の存在することが前提となる。

ところが、科学史の現実においては、認識内容を伝えるべき相手は必ずしも職場の同僚 (生産的実践の場を同じくするもの) ではない。むしろそれを超えて国民的統合のためのイデオロギーへと疎外されている。つまり、自然をどのように認識するべきかは生産的実践の場から遊離し、支配階級が社会 (あるいは国家) 全体に共有させようとするイデオロギーを補強する知識として、あるいは支配階級にたいする被支配階級のイデオロギー闘争の一翼を担うものとして、機能させられる。近代市民社会の成立を一つの節目として、科学者層がもっぱら軍事的要求のもとに組織され、その活動が (実験を含む) 科学研究として、生産的実践一般から独立化する。このことは言うまでもなく、〈分業〉のあらわれの一つである。

生産的実践の場から切り離された「科学」の生産過程、その内容については次の節で触れる。

[*15] 行動随伴性 (behavior contingency) とは「オペラント行動の自発頻度の変化とそれが自発された直後の環境の変化との関係」(Wikipedia) のこと。今日では行動分析学と言う心理学の一分科として体系化され、医学、労働科学などの分野に応用されている。Burrhus Frederic Skinner(1904-1990) はアメリカの心理学者で行動分析学の創始者。I.Pavlov と E.Thorndike の業績とを統合したことで知られる。

2.2 認識における実験の意義と確率

「実験について」(T-5) はとくに雑誌媒体で公表されたものではないと見受けられるが、予測について具体的に論じられているという意味で貴重である。またここでは、「科学」生産の過程が実際的に論じられており、そこでの統計的取り扱いの意味が詳細に分析される。

2.2.1 実験について

武谷は、まず〈現象論〉的段階で重要な役割を果たす「観察」から出発し、それに関与する「因子」の取り扱いに着目することから「実験」の意義を説いてゆく。ここでは〈現象論〉から〈実体論〉への移行の論理が具体的に議論されているとみてよいであろう。

最初に取り上げられるのは観察＝〈現象〉の記述であり、これには質的と量的の別があるとされる。

- 質的な記述：博物学 (動植物の分類、鉱物学、地質学)
- 量的な記述：諸惑星の運行 (エジプト、バビロニア、インド、古代中国)、地上の測量、静力学

> 自然科学はまず自然に起こっている様々な現象を記述することが必要である。これを〈経験的事実〉と呼んでいる。この記述 (観察) には質的な記述と量的な記述とがある。・・・量的な観察は特に観測及び測定と言われる。(T-5)

続けて、実験＝〈現象〉的記述を確かめること (基礎づけること、再現すること) が論じられる。ここではすでに〈実体論〉的段階への移行が示唆されている。〈因子〉が説明され、それらをコントロールしながら〈因子〉間の関係を見いだす過程が「実験」であるとされる。

> 生物の分類学も進化論が出て種の間の関係が考えられ、ダーウィン自身多くの実験によって種の変化を確かめることとなった。鉱物学も物理化学の実験によって基礎づけられた。天文学、静力学もガリレイ等の実験によってしっかりとした基礎がすえられるに至った。(T-5)

> 自然を「観察」する場合、実際の自然の現象にはいろいろな〈因子〉が複雑に入ってきているので、そのおのおのの〈因子〉がどんな役割を演ずるか、ということが問題になる。ときにはわれわれは何らかの手段を講じなくてはならない・・・。われわれは自然のままでなくて、他の条件を一定にしておいて、一つの〈因子〉だけを変えていく。・・・こうしていろいろな〈因子〉の間の関係を見いだすのである。・・・このような〈因子〉、これは概念としてつかまれるもので、これを物理(学) のばあいは〈物理量〉と呼ぶのである。(T-5)

実験過程が、〈因子〉の概念分析→実験の実施→実験法則 (実験式) の定立、としてまとめられる。ここでの実験法則[*16]は〈現象論〉的であるに過ぎない旨、注意される。

> ・・・実験を行う際にまずこの〈因子〉についての概念分析を行わなくてはならない。もしわれわれが自然現象にとって無意味な概念とかまた非常に込み入った概念、また実際には各〈因子〉の間にまたがったような概念を一つの〈因子〉として問題にして実験をしたとしたならば、何ら意味ある結果を得ることができない・・・。(T-5)

> 実験によって得られたこれらの諸〈因子〉間の関係を〈実験法則〉と言う・・・。これはもちろん自然法則なのであるが、まだ体系的につかまれていない・・・。また〈本質〉的な解明がなされていないことから、〈現象論〉的と言われる。(T-5)

実験における定性的と定量的の別が、観察と同様に議論される。しかし〈因子〉の存在が問題にされている以上、もはや〈現象論〉における質と量の区別と同一ではない。たとえば「定性的実験において量的なものが問題にならないかというと決してそうではなくて、正確な定量が初めて質の存在を決定するばあいも多く存する」として、Lavoisier によるフロギストンの否定、Planck のエネルギー量子の導入、Rutherford の原子模型などをその例としている。

- 定性的実験：〈因子〉の作用がはたして存在するかどうか、二つの〈因子〉の間の関係があるかどうか、平行的な関係か逆の関係か、ある要素が存在するかしないか、どんなもので対象ができてい

---

*16 物理学実験の現実においては、実験式はむしろ作られず、理論式にたいして観測値を重ねて示したり、観測値にたいして傾向線をひく程度で済まされることが多く、またそれで十分であるとされている。

*17 既にこれらの問いが統計的仮説として提示されていることに注意せよ。統計的仮説検定においては、ひとつひとつの問いが〈対立仮説〉と〈帰無仮説〉の組として表現されることになる。

るか (化学の定性分析)、ある二つのことの間に因果的な関係があるか[*17]。

- 定量的実験：主観的な (心理的な) 要素を取り除き、複合的な〈因子〉を分解したうえで、一つの量に対して〈単位〉を設定し、対象をこれと比較して数としてあらわす[*18]。

誤差の取り扱いが考察される。〈客観的法則性〉が、実験過程においてあらわれる現象 (結果) にはさまざまな偶然〈因子〉＝誤差が含まれることが最初に指摘される。そのために、同一の原因が同一の結果をもたらさない。

> 同一の (観測) 対象に対して同一の (観測) 結果が得られることが予期される。しかし事実はこれに反している。(T-5)

次に誤差が分類される[*19]。ここでは実験における〈観測者〉〈観測手段〉〈観測対象〉の別[*20]が暗黙のうちに措定されていることに注意する。武谷はこれらを主観と客観に分けているが、〈客観的法則性〉の作用として誤差が生じているのはいずれの側においても同じである。つまり、〈観測者〉〈観測手段〉〈観測対象〉のいずれもが物質として、〈客観的法則性〉の支配下にある。またそれら要素自体のもつ技術性の高さ (正確さや純度など) がその実験の成否なり分解能なりを左右する。

- 客体側の原因 (〈観測対象〉の側で発生する誤差)
  - 初期条件：例) ニュートン力学における〈実体〉の配置
  - 環境条件 (バックグラウンド)：例) 空気の抵抗 (Galilei)
- 主体側の原因 (〈観測者〉〈観測手段〉の側で発生する誤差)
  - 観測手段：測定機器の偶然的変動、目盛の不正確さ
  - 観測者：観測者個人による誤差・誤り

これらの誤差について、系統的なものであるか否かが吟味される。前者であれば、実験の方法を工夫することによりコントロールする[*21]。後者であれば、「・・・〈因子〉としてそれぞれ別に取り出せない〈偶然性〉として扱い、統計的方法によって処理」する。

2.2.2 本質論的段階への移行

実験の結果、〈現象〉の事実が確かめられ、再現され、〈実体論〉的な知識として記録される。しかし、これはあくまでもその実験の場での特殊の知識に過ぎず、〈本質論〉的・普遍的な認識には至っていない。〈実体論〉的知識を〈客観的法則性〉への〈本質論〉的な認識に高める方法が論じられなければならない。(T-5 においてこれについての記述はとくに見あたらない。先にまとめた T-3 の〈実体論〉から〈本質論〉への移行における三つの形態、が何らかの示唆を与えるだろう。)

〈実体論〉的知識が報告され、共同体内で共有される。学術団体の活動などがこれに寄与することになる。またこの知識は新たな〈生産的実践〉の場における予測のために活用される。そこにおいて、相異なる複数の〈実体論〉的知識はそれぞれ理論仮説としてとり扱われることになる。また新たな〈現象〉の確認に用いられた実験器具は、別の〈生産的実践〉の場で、新たな〈労働手段〉(加工技術など)、〈労働対象〉(新素材など)、〈生産物〉(新製品・すなわち生活諸手段であり生産諸手段) へと転用されるだろう。

こうして〈実体論〉的知識はさまざまな〈生産的実践〉の場で繰り返して活用されることを通じて、〈本質論〉的知識に高められる。その成果は、教育労働、出版・マスコミ労働を通じて広く共有され、新しい世界観、自然観として、〈労働そのもの〉の中に対象化される。

[*18] ここで武谷は、「定量的実験」を「観測」と混同している。そうではなく、たとえば「因子の作用が存在することを前提に、その影響の大いさを量的に把握すること」と記述すべきであった。また、上の定性的実験と平仄を合わせるならば、観測値を信頼区間付きで記録すること、影響の大いさをパラメータ (偏回帰係数など) として、その期待値とばらつき (測定の精度) を推定することなどが問われることになる。

[*19] 誤差の分類について、特殊な状況についても注意があたえられている。たとえば、〈観測者〉〈観測手段〉〈観測対象〉とのあいだに生じる相対運動の影響、量子力学の「観測問題」などである。

[*20] これは〈労働そのもの〉〈労働手段〉〈労働対象〉の区別に相応する。ただし、武谷はかならずしもその別を明りょうには意識していない。

[*21] たとえば、一つの〈因子〉だけ異なる二つの実験系を用意し、両者の結果の差を見ることでバックグラウンドの影響を排除する。

2.2.3　確率研究への端緒

誤差が客観的に存在していることを認識することは、確率研究の端緒である。これを認めるには、〈観測者〉〈観測手段〉〈観測対象〉からなる実験系が、人間の意識から独立した一つの客観的な場であること、したがって唯物論的世界観に立っていることを前提としなければならない。誤差は実験系の中で生まれ、観測結果の中に記録される。

誤差とは、ここではさしあたりは「同一の原因 (〈因子〉) が同一の (観測) 結果をもたらさないこと」(偶然性)[*22]、と〈現象論〉的に理解されるが、原因と結果のうち、何が同じで何が異なるのかについて〈実体論〉的な探求が次に要求される。誤差の原因を客体側に求めるか、主体側に求めるかの判断がまさにその作業と言えよう。

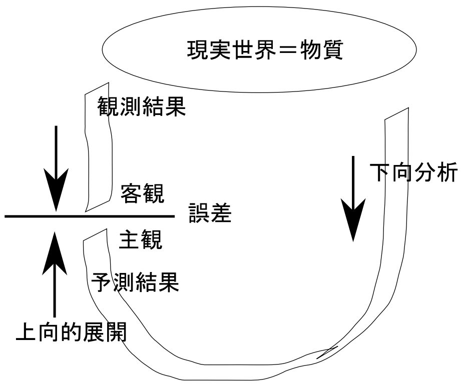


図 2.2　認識の連関と誤差

たとえばコイン投げ (ベルヌーイ試行[*23]) を一つの実験系として考えてみる。〈観測対象〉〈観測手段〉は「コイン」(表裏の区別できる)、「コインを無作為に投げ上げる手」と「それを受け止めるテーブル」である。その (読み取り) 結果として、たとえば、

表、表、裏、表、裏、裏、表、・・・

といった記号列 (データ) が記録されることになる。

誤差は、たとえば最初の結果「表」にたいして、それ以降の結果が同一であったか否かにより計られよう。「表」か「裏」のどちらを基準とすべきかは、〈観測者〉の恣意にゆだねられ、またそのこと自体は誤差研究において結果的に意味を持たない (表と裏を取り替えても同じことが後から判明する)。

そして〈観測者〉は次の一試行の結果を (誤差が一試行ごとの手の振り方、コインとテーブルの状態に依存して変わり得るがゆえに) 予測できないまでも、十分に長く繰り返された試行においては、表と裏の出現比率が一定の法則性 (大数の法則[*24]) にしたがっているらしいことを発見するわけである。また、逆にわれわれはこの事態について、誤差を支配する〈客観的法則性〉がコイン投げという実験系＝〈実体〉をつうじて、観測結果たるデータとして〈現象〉したと解釈する。

この事情は社会科学においても実は同様である。世論調査において、ある法案の賛否が問われているとする。そこでも〈観測者〉〈観測手段〉〈観測対象〉の組が考えられる。ここで〈観測対象〉となるものはたとえば「投票資格のある有権者の全体」(母集団)、〈観測手段〉としては「母集団から標本を無作為に抽出した上で、その標本にたいして調査票を用いて調査すること」となる。やはり、結果として調査票にある「賛否という項目」(刺激) にたいする標本の反応が記録されたデータが以下のように得られる。

否、否、賛、否、賛、賛、否、・・・

このばあい誤差の発生原因にはさまざまなものが考えられる。調査票のワーディングの適切さ、標本のサイズや抽出の方法の妥当性などに依存する誤差は〈観測手段〉から生じる。標本のもつ調査時までの経験・知識・信念、調査時点での身体的・精神的状態、生活状態の差異などによる誤差は〈観測対象〉の側で生じる。また、誤差の基準となるべき予測については、〈観測者〉がど

[*22] 〈差額地代〉の規定との類似性に注意せよ。

[*23] ベルヌーイ試行 (Bernoulli trial) : 状態が二つの事象のみから成るランダム実験であり、事象の出現確率が一定のもの。スイスの数学者 Jacob Bernoulli(1654-1705) の研究によって知られる。

[*24] 大数の (弱) 法則 (Weak Law of Large Numbers) : コイン投げの試行回数を限りなく増やせば、表が出る回数と裏が出る回数の比率 (相対頻度) はどちらもある一定の数に近づく。

大数の法則は数学の定理のように見え、また実際に数式による証明も付されるのであるが、実はその証明のなかにこれが誤差に関する〈客観的法則性〉の表現であることの証拠が見出される。付録 C.1 を参照のこと。

のような〈因子〉を考慮に入れたか、あるいは無視したかにも影響される。

ここでわかることは、誤差があくまでも得られた「観測結果」と期待する「予測」との差異として表示されること、すなわち〈客観〉と〈主観〉の境界面の上に存在する、ということである (図 2.2)。このことを気に留めておくことが確率の客観説と主観説との対立を解くカギとなるだろう。

2.3　小括

本節の結果をまとめると以下のようになる。

- 武谷が〈認識の三段階論〉(本質論―実体論―現象論) を獲得していたことが、技術の〈実体論〉を超えて〈本質論〉の段階に進まねばならない、という指針につながった。〈認識の三段階論〉のすぐれた点は、〈本質〉と〈現象〉を媒介するものとしての〈実体〉に着目し、これに認識方法における独自の段階という地位を与えたところにある。
- 1) 生産的実践と自由の関係、2) 自由と〈客観的法則性〉の関係、3) 生産的実践と〈目的意識性〉の関係の推論を通じて技術の本質が次のように規定された。すなわち「技術とは人間実践 (生産的実践) における客観的法則性の意識的適用」である。
- 〈現象論〉から〈実体論〉への移行の論理を具体的に議論することが、「実験」の意義の解明につながっている。〈現象論〉的段階で重要な役割をなす「観察」から出発し、実験＝〈現象〉的記述を確かめること、〈因子〉をコントロールしながら〈因子〉間の関係を見いだす過程が「実験」である。「実験」とはコントロールされた〈生産的実践〉にほかならない。
- 誤差が客観的に存在していることを認識することは、確率研究の端緒である。これを認めるには、〈観測者〉〈観測手段〉〈観測対象〉からなる実験系が、人間の意識から独立した一つの客観的な場であること、したがって唯物論的世界観に立っていることを前提とする。
- 誤差はあくまでも得られた「観測結果」と期待する「予測」との差異として表示される。すなわち〈客観〉と〈主観〉の境界面の上に存在する。このことを気に留めておくことが確率の客観説と主観説との対立を解くカギとなる。

第 3 章

確率と因果性に関する形式的理論

本節では、確率と因果性についての存在論的記述を試みる。つまり、推定、推論、予測など認識過程にかかわる議論 (これらは次節で扱うことになる) を注意深く留保する。というのも、確率の存在論的側面と認識論的側面を同時に議論することは往々にして収拾のつかない混乱を招くからである。

確率にかんする学説は 18 世紀に急速に発達を見たが、そこには既に今日の論争の種となっている様々な要素が既に萌芽のかたちで含まれていた。すなわち、(今日「ベイズの定理」として知られている) 帰納法論理とそれを含む確率の演算規則、確率を数え上げの数理[*1]の対象に帰着させる見解などである。われわれはこれらの統一物を、たとえば Laplace の体系の中に見出すことができる。そこで、ここではまず、彼の古典的確率論体系を、可能なかぎりその論述に沿ったかたちで整理することを目指す[*2]。

つづいて、独立性と条件付き独立性の概念の議論を経て、相関性と因果性の区別と後者へのアプローチを論じる。

3.1 確率変数と確率法則

P.S.Laplace は「フランスの Newton」とも称される人物であり、d'Alembert[*3]にその才能を見いだされた、フランス啓蒙思想の申し子であった。フランス革命期のエコール・ノルマル、エコール・ポリテクニークなどで教鞭をとり、ナポレオン体制後も引き続きフランス科学界で重きをなした。

「確率の哲学的試論」はエコール・ノルマルでの講義をもとに、古典的確率論を体系化したテキストとして広く読まれた。その重要な特徴は彼の言う「解析的方法」(組合せ論と差分方程式、母関数) の積極的な援用にある。Laplace はこれを確率論の基礎づけに用いるとともに、天体力学への応用、推論ルールの探求 (ベイズ的推論) にも活用した。確率についての以下の定義は彼の方法論の前提をなし、逆にその方法論を活用するためにはこのような定義こそが必要であった。

> 第一の原理は、確率の定義に他ならない。・・・確率とはすべての可能な場合の数に対する好都合な場合の数の比である。

3.1.1 Laplace の「解析的方法」

Laplace はその方法を次のように説明する。ここでは確率が数列 (関数) として表示され、それらのあいだの関係が差分方程式 (漸化式) で表現される。これを解く手段として (通常) 母関数が用いられる。

> 確率の問題を解くための最も一般的で最も直接的な方法は、問題を差分方程式に依存する形にすることである。

---

*1 これは O.Heaviside により演算子法 (ラプラス変換) として再発見されたものであり、Laplace はこれを de Moivre、Leibniz などからヒントを得てひとつの体系に仕立て上げたのであった。

*2 内井 (1997) はその解説において、Laplace の記述を「ベイズ統計学」あるいは「確率の主観説」の観点から色読している。そのために、内井はその解釈の邪魔となる母関数の方法について、その解説の中でほとんどとりあげていない。

*3 Jean Le Rond d'Alembert(1717-1783)、フランスの哲学者、数学者。Denis Diderot 等とともに「百科全書派」を組織した。

> いくつかの変数を持ち確率を表す関数が、それらの変数の値をそれぞれの差分によって増加させたときに次々に取る値を比較することによって、多くの場合その問題はこの変数の異なる値の間の極めて単純な関係を提供する。この関係は常差分方程式、あるいは、偏差分方程式と呼ばれるものである。

具体例として Pascal と Fermat の往復書簡で議論された、くだんの問題がとりあげられる。

> 等しい技量をもつ A と B が 32 ピストール*4づつ、合わせて 64 ピストールを賭ける。いずれかが 3 点を先取した時点で勝負が決せられ、掛け金は全額勝者のものとなる。ところが A が 2 点、B が 1 点を得た時点で試合は中断された。掛け金はどのように分配されるべきか。

この問題から Laplace は式 (3.1) のような漸化式を導く。ここで $p_{i,j}$ は A が勝つ確率であり、これは A と B がそれぞれあと何回勝つ必要があるか (i,j) に依存している。$p_{2,1}$ が計算できれば、分配金はこれに掛け金を合わせることにより得られる。

$$
\begin{aligned}
p_{i,j} &= \frac{1}{2}p_{i-1,j} + \frac{1}{2}p_{i,j-1} \\
& \quad p_{0,j>0} = 1,\ p_{i>0,0} = 0
\end{aligned}
\tag{3.1}
$$

式 (3.1) を解くと次が得られる (付録 C.2 を参照のこと)。

$$
\begin{aligned}
p_{i,j} = 1 - \sum_{k=1}^{i-1} \left(\frac{1}{2}\right)^{j+k} \begin{pmatrix} j+k-1 \\ k-1 \end{pmatrix} \\
- \left(\frac{1}{2}\right)^{j+i-1} \begin{pmatrix} j+i-1 \\ i-1 \end{pmatrix}
\end{aligned}
$$

この計算過程にあらわれる、

$$
\begin{aligned}
M(x,y) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p_{i,j} x^j y^i \\
&= \frac{x(2-x)}{(1-x)(2-x-y)}
\end{aligned}
\tag{3.2}
$$

が $p_{i,j}$ の (通常) 母関数*5である。

3.1.2　確率論における現象論と実体論

母関数は数列一般に適用可能な分析用具であるが、これと確率の結びつきは、個々バラバラの数値であった確率を束ねて一つの数学的実体 (確率母関数：pgf) を与えることに役立った (付録 C.3 を参照のこと)*6。これにより確率法則は個々の確率分布 (関数) として明示的に取り出されるに至った。

確率は確率分布の一部をなし、この確率分布に従う確率変数が強く意識される。すなわち、個々の観測量は確率変数にほかならないことが理解される。複数の確率変数の間の差異がその従う確率法則の違いとして識別される。さらにそれらの間の演算によって新たな確率変数が構成され、これが従う確率法則は、もとの確率変数を支配する確率分布と異なることが見出される。われわれはこの事態を確率論における〈実体論的段階〉として特徴づけることができよう。

コイン投げやサイコロ遊び、くじ引きなど、単純な事象に対する観察と実験に基づく知見は、上の〈実体論的段階〉と区別して〈現象論的段階〉とみなすことができる。Laplace のあげる確率の諸原理はこうした現象論的な観察の結果を少数の規則として整理したものであり、これが後に Kolmogorov の「公理」として結実した。それらの原理とは以下のとおりである*7。

---

*4 1 ピストール金貨は 10 リーブル (フラン) に相当する (およそ一万円)。

*5 これを通常母関数 (ordinary generating function) とわざわざ称しているのは、母関数一般にはべき級数によるもの (これが通常母関数) のほかに、指数型母関数、ポアソン母関数、ランベルト級数、ベル級数などがあるからである。

*6 ただし Laplace 自身は母関数の利用について一貫して肯定的だったというわけではないらしい。1810〜12 年には離散確率の mgf にかんする反転公式を得ていたにもかかわらず、それ以降、とりわけ 1812 年の「確率の解析的な理論」刊行以降、母関数についてはそれ以上の研究を行っていない。Seal(1949) は次のように記している。「興味深い事実がここから明らかになる。すなわち、Laplace は pgf を特定の確率変数の n 個の和の従う確率分布の陽表的な式を導出するために使ったことは一度もないということである。・・・Laplace はその作業が必要になったとき、以前 (1781 年) に打ち立てた帰納法による方法を使っている。・・・pgf の理論が 19 世紀の最初の四半世紀の終わりにはかなり発展していたにもかかわらず、それらの成果が統合され、さらなる前進が成し遂げられるまでにはなおほぼ 100 年を要した、ということは異常なことである。この原因は、Laplace の (おそらくは個人的な)pgf への嫌悪と、彼の権威が 19 世紀の数学者たちに及ぼした影響の大いさゆえであろう。」

*7 期待値に関する規則はここに含めなかった。なお pgf そのものも期待値の一種とみなせることに留意する。付録 C.3 を参照のこと。

I. 確率の定義 $p = n/N$

II. 和の法則

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

III. 積の法則

$$p(A \cap B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

IV. 独立の場合 $p(B|A) = p(B)$

V. 条件付き確率の定義

$$p(B|A) = \frac{p(A \cap B)}{p(A)}$$

VI. ベイズ・ルール
原因 $B_i \to$ 結果 A(ただし、$\Omega = \bigcup_{i=1}^{n} B_i$)

$$\begin{aligned} p(B_i|A) &= \frac{p(A \cap B_i)}{p(A)} = \frac{p(A \cap B_i)}{\displaystyle\sum_{i=1}^{n} p(A \cap B_i)} \\ &= \frac{p(B_i)p(A|B_i)}{\displaystyle\sum_{i=1}^{n} p(B_i)p(A|B_i)} \end{aligned}$$

VII. 全確率の公式

$$\begin{aligned} p(A) &= \sum_{i=1}^{n} p(A \cap B_i) \\ &= \sum_{i=1}^{n} p(B_i)p(A|B_i) \end{aligned}$$

なお、Laplace はこれらの原則について、確率の存在論的な性質と認識論的な方法論とを区別せず、両義的に取り扱っている。われわれはさしあたりはこれらを事象間のつながりを示す確率の存在論的性質 (観察の結果から認識された) として受け取っておき、認識論的な側面 (特にベイズ・ルール) については次章で議論することにしよう。

確率母関数 pgf、またその発展形であるモーメント母関数 mgf の利用は、中心極限定理 (CLT) を通じて正規分布という特別な分布の発見につながった (やはり付録 C.3 を参照のこと)。CLT は de Moivre が Bernoulli による比率の極限についての大数の法則の解明を超えて、比率の偏差の極限を解明したことに端を発する。これについて Laplace は次のように述べている。

> ド・モアブルは彼の仕事のなかで、多数の観察から得られた結果の確率についてのヤコブ・ベルヌーイの定理をとりあげた。ベルヌーイとは違って、彼は生じるはずの事象の比率が、それぞれの可能性の比率に絶えず近づくことを示すだけでは満足せず、さらにこれら二つの比率の差が与えられた範囲の中に入る確率の美しい単純な表現も与えた。

E.Borel も同じことを次のように説明している。

> n 回試行をくりかえすとき、事象の起こる「蓋然回数」は np である。この数 np は多くの場合、整数とはならぬであろう。実際に事象の起こる回数を $np + h$ とする。この数 h のことを偏差と定義する。反復試行の理論の本質的な問題は、「偏差の法則」あるいは、詳しくいうならば、この偏差 h がある限界に含まれるための確率を決定することである。

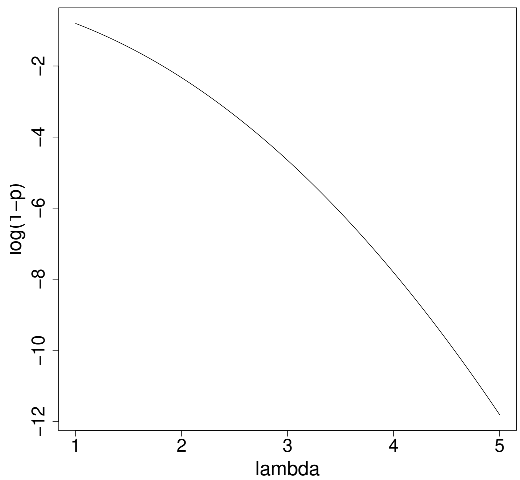


図 3.1　相対偏差 λ と反対事象の起こる確率 $1 - p$

CLT によれば、X_i の母平均、母分散をそれぞれ μ、σ^2 として、$\sum_{i=1}^{n} X_i$ をその平均と標準偏差で基準化した確率変数 T の極限は標準正規分布に従う。同じことであるが、$\sum_{i=1}^{n} X_i$ の極限は平均 $n\mu$、標準偏差 $\sqrt{n}\sigma$ の正規分布に従う。もとの X_i が成功確率を p とするベルヌーイ試行だとするならば、$n\mu = np$、$\sqrt{n}\sigma = \sqrt{np(1-p)}$ であり、このことから Borel は、上の「偏差の法則」(CLT) の意義を次のように説明するのである。

> 反復試行の理論の本質的な結果は、偏差の単位 $u (= \sqrt{2np(1-p)})$ が知られている場合、偏差が $-\lambda u$ と $+\lambda u$ との間に含まれるための確率が数 λ だけに

よって定まるということである。従って、反復試行の問題で、数 n および p がいろいろに変わる場合、それらの問題の間には一種の「相似性」が存在する。すなわち (n、p がいろいろになっても) それらの条件から u だけを計算すれば良い。結果に関係してくるのは u だけなのである。

図 3.1 は Borel に従って、絶対偏差 h が区間 $[-\lambda u, +\lambda u]$ に含まれない確率を相対偏差 λ ごとに示したものである。Borel はこれから、以下のような表現を与えている。

λ	$1-p$	意味
1.15	10^{-1}	まず確からしい
2.30	10^{-3}	非常に確からしい
3.45	10^{-6}	極度に確からしい
4.60	10^{-10}	確実である

正規分布は他の多くの分布がその極限において取る姿であり、また確率変数のとる値が離散量 (自然数) から連続量 (実数) に移行するという意味でも重要である。正規分布は多くの実証科学 (すなわち天文学、動力学等を含む) において、その観測結果を処理し、解釈するための基礎的な道具となった (最小自乗法)。またこれを一般化 (多変量化) したものは、因果関係を解明するための基盤として活用された (多変量解析)。現実の多くの観測量 (の誤差) は必ずしも正規分布に従っているとは限らないが、正規分布による近似が可能なものとみなされ[*8]、(推定・検定法の多くが正規分布を前提に開発されたがゆえに) ときとして無条件にその正規性が仮定された。

3.1.3　充足理由律

Laplace はその「確率の哲学的試論」の冒頭で、次のように充足理由律 (principium rationis sufficientis; principle of sufficient reason; 引用文中では「十分な理由の原理」) に言及している。この言及は、彼の天体力学上の業績とも相まって、彼が「絶対論者」、「宿命論者」である、との世評を固めることに貢献した。

> 現実の事象は、それに先立つ事象との間に (、) あるつながりを持っている。そのつながりはいかなる事物もそれを生み出す原因なくしては存在し得ないという自明の原理に基礎を持つ。この公理は「十分な理由の原理」という名前で知られるが、どちらでも良いような行為にまでも及ぶ。
>
> 最も自由な意思でも、決定する動機なくして行為を生み出すことができない。というのは、もし二つの立場の条件が全て正確に同じなのに一方では意志が働き他方では働かないとしたら、この意志の選択は原因を持たない結果となってしまうからである。これでは、意志はエピクロス流の盲目的な偶然となってしまう、とライプニッツ (G.W.Leibniz:1646-1716) はいう。これと反対の見解は、無差別な事柄での意志の選択の際のうつろいやすい理由を見失って、意志は動機なくして自らを決定するのだと思い込む精神の錯覚なのである。

しかし、上で Laplace は必ずしも絶対的な宿命論を語っているわけではないことに注意するべきである。そうではなく、事象と事象との間に確率論という客観的法則性 (「つながり」) が存在することを主張しているにすぎない。

また、彼はこのことを Leibniz の「充足理由律」から引き出しているが、その Leibniz は Epikouros への批判というかたちで偶然性の意義を否定している。ただし、Laplace 自身が Epikouros をどのように見ていたかはこの部分だけからは読み取れない。引用部分に見る Leibniz による Epikouros への批判は、あたかも Epikouros が「盲目的な偶然」、すなわち事象間にはなんらの因果関係も働かない、と主張したかのような誤解に立脚している。しかし、この誤解はエピクロス派に対してなされたストア派および Cicero のいわれなき批判に端を発する。また、Leibniz 存命の頃も、依然としてエピクロス主義の烙印が死罪に値する異端 (例：Giordano Bruno) を意味したことも忘れるべきではない[*9]。

Laplace が「最も自由な意志」(いわゆる「ラプラスの魔」) について語っていることは、Leibniz の自然神学

[*8] χ^2 検定などがその例である。これは多項分布が (正規分布より派生する)χ^2 分布によって近似されることに基づく。
[*9] Epikouros の哲学については、フォスター (2004) を参照のこと。
[*10] 彼は Napoléon にたいして、神について問われた際に、「帝よ、私はかかる仮説を必要としないのです」、とまで語ったとされる。ただしこの言明の証拠となるものは乏しい。Laplace が貧農の家の出自であるらしいこと、フランス革命期を含む政治的に微妙な時代を如才なく生き延びたことなどを鑑みても、それほど不用意に証拠を残しているとは思えない。ただし、彼がそう語ってもおかしくはないと世人が信じたということの方が重要かもしれない。

の内容を語ったにすぎず、Laplace 自身の見解とは区別されるべきである*10。彼は以下のように、円環状に並べた壺の比喩で自然の拡散過程を説明していることからもうかがえるように、客観的な確率過程をさまざまに議論している。

> ・・・各々非常に多数の白い玉と黒い玉を含む一連の壺が円形に並べられているとしよう。各々の壺の中の白い玉と黒い玉の比率は、最初は非常にまちまちであっても良い。例えばある壺は白い玉しか含まず、別の壺は黒い玉しか含まなくても構わない。さて、
>
> 1. ・・・第一の壺から玉を一つ取り出して第二の壺に入れ、この玉が他の玉とよく混ざるように第二の壺を揺り動かした後、
> 2. そこから玉を一つ取り出して第三の壺に入れる。
> 3. 以下同様にして、最後の壺から玉を一つ取り出して第一の壺に入れるまで続ける。
> 4. そして、以上の一連の操作を何度となく繰り返すものとする。
>
> 確率の分析によればこれらの壺における白黒の玉の比率は、最終的には全て同じになり、すべての壺に含まれる白い玉の総数と黒い玉の総数との比率に等しくなっていく。かくして、この規則的な変化のパターンにより、玉の比率における最初の不規則性は長いあいだには消滅し、最も単純な秩序に座を譲る。・・・

3.2　独立性、条件付き独立性

ここからは Laplace の著作を離れて、識別された複数の確率変数間の関係についての理論を取り扱う。

3.2.1　独立性と条件付き独立性

独立性 (independence) と条件付き独立性 (conditional independence) は事象間の因果分析の基礎をなす重要な概念である。

事象 A と B とのあいだに因果性が認められる前提として、二つの事象の間の相関関係が示されなければならない。相関関係は、どちらが原因でどちらが結果であるかを特定していないという意味では、因果関係と同じものではない。また後述のように相関関係のなかにはたんに見かけ上のものも含まれる。二つの事象間の相関関係は、それらの生じる可能性が互いに独立ではないこと、すなわち A と B との同時確率がそれぞれの事象の発生確率の積には分解できないこととして示される。このことは、前述の Laplace による確率の諸原理 III〜V のなかで既に言及されている。すなわち、「事象 A と B が独立である」($A \perp\!\!\!\perp B$) とは次の関係が成立することである。

$$\begin{array}{c} P(A \cap B) = P(A) \cdot P(B) \\ \updownarrow \\ \left\{ \begin{array}{l} P(A|B) = P(A) \\ P(B|A) = P(B) \end{array} \right. \end{array} \tag{3.3}$$

互いに独立な確率変数の共分散、または相関係数はゼロとなる。このことは mgf を使って次のようにして確かめることができる。

$$\begin{aligned} & \left. \frac{\partial^2 M_{\boldsymbol{X}}(\boldsymbol{z})}{\partial z_1 \partial z_2} - \frac{\partial M_{\boldsymbol{X}}(\boldsymbol{z})}{\partial z_1} \cdot \frac{\partial M_{\boldsymbol{X}}(\boldsymbol{z})}{\partial z_2} \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & \downarrow \text{indep} \\ & = \left. \frac{\partial M_{X_1}(z_1)}{\partial z_1} \cdot \frac{\partial M_{X_2}(z_2)}{\partial z_2} \left(1 - M_{X_1}(z_1) M_{X_2}(z_2)\right) \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & = 0 \end{aligned}$$

具体的には、たとえば次のような母共分散 (相関) 行列をもつ一般的な二変量正規分布のばあい、

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\begin{aligned} & \left. \frac{\partial^2}{\partial z_1 \partial z_2} \exp\left(\frac{1}{2} \boldsymbol{z}^t \boldsymbol{\Sigma} \boldsymbol{z}\right) \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & = \left. \exp\left(\frac{1}{2} \boldsymbol{z}^t \boldsymbol{\Sigma} \boldsymbol{z}\right) \{(z_1 + \rho z_2)(\rho z_1 + z_2) + \rho\} \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & = \rho \end{aligned}$$

$$\begin{aligned} & \left. \frac{\partial}{\partial z_1} \exp\left(\frac{1}{2} \boldsymbol{z}^t \boldsymbol{\Sigma} \boldsymbol{z}\right) \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & = \left. \exp\left(\frac{1}{2} \boldsymbol{z}^t \boldsymbol{\Sigma} \boldsymbol{z}\right) (z_1 + \rho z_2) \right|_{\boldsymbol{z}=\boldsymbol{0}} \\ & = 0 \quad (\text{同様に } z_2 \text{による偏微分も } 0) \end{aligned}$$

より、$\rho = 0$ ならば X_1 と X_2 は独立となる。

条件付き独立性は、独立性と似ているが三つ以上の事象について適用される。「事象 C が生起したという条件のもとで、事象 A と B が条件付き独立である」($A \perp\!\!\!\perp B|C$) とは、次の関係が成立することである。

$$\begin{array}{c} P(A \cap B|C) = P(A|C) \cdot P(B|C) \\ \updownarrow \\ \left\{ \begin{array}{l} P(A|C \cap B) = P(A|C) \\ P(B|C \cap A) = P(B|C) \end{array} \right. \end{array} \tag{3.4}$$

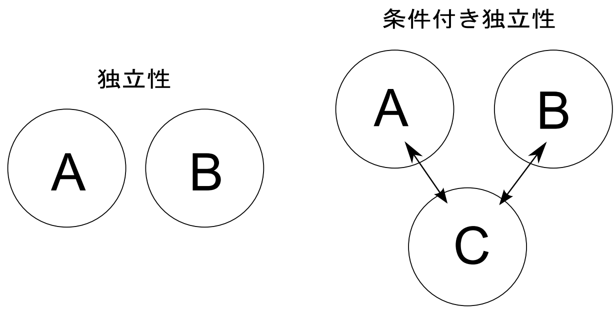


図 3.2　独立性と条件付き独立性

なお、$P(A|C \cap B) = P(A|C)$ は、A が $C \cap B$ との関連度が、A と C の関連のみに還元が可能であるという意味で「マルコフ性」を意味すると解釈できる。マルコフ性は、歴史的経緯からすれば確率過程にかかわる概

[*11] 条件付き独立性の研究のためには、多変量確率分布 (複数の確率変数をベクトルとして束ねた確率変数) を明示することが有益である。これにより同時確率、周辺確率、条件付き確率を共分散、相関、偏相関の各関係と関連させることができる。

念であるが、実質的には条件付き独立性と同等のものとして、確率過程に限定されずに用いられて良い。

独立性、条件付き独立性は、事象の間ばかりではなく、確率変数の間にも適用される[11]。以下では、量的な確率変数、質的な確率変数のそれぞれのばあいの条件付き独立性を説明する。

3.2.2 多変量正規分布における条件付き独立性

ここでは連続型の多変量確率分布の例として多変量正規分布をとりあげる (付録 C.4 を参照のこと)。

$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}$ は精度 (precision) 行列あるいは concentration 行列と呼ばれる。これが以下のような要素表示を持つとする。

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1p} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \cdots & \omega_{pp} \end{bmatrix} \tag{3.5}$$

今、確率変数 $\boldsymbol{X}$ が、

$$\begin{aligned} \boldsymbol{X} &= (\boldsymbol{X}_a, \boldsymbol{X}_b) \\ &= ((X_1, X_2)(X_3, \cdots, X_p)) \end{aligned}$$

のように二つの部分に分割されているとする。このとき、$\boldsymbol{\Omega}$ は $\boldsymbol{\Sigma}_{bb}$ にたいするシュール補行列 $\boldsymbol{C}_{bb}$ を用いて、以下のように表現される。

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{C}_{bb}^{-1} & -\boldsymbol{C}_{bb}^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \\ -\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{C}_{bb}^{-1} & \boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{C}_{bb}^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} + \boldsymbol{\Sigma}_{bb}^{-1} \end{bmatrix} \tag{3.6}$$

式 (3.5) と (3.6) を比較することにより以下を得る。

$$\boldsymbol{C}_{bb}^{-1} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \tag{3.7}$$

この逆行列、すなわち $\boldsymbol{C}_{bb}$ は以下のように計算される。これは X_1、X_2 以外のすべての変数が固定されたときの X_1 と X_2 の共分散行列である。

$$\boldsymbol{C}_{bb} = \frac{1}{|\boldsymbol{C}_{bb}^{-1}|} \begin{bmatrix} \omega_{22} & -\omega_{12} \\ -\omega_{21} & \omega_{11} \end{bmatrix} \tag{3.8}$$

これより X_1、X_2 の偏相関係数 ρ_{12} が以下のように計算される。

$$\rho_{12} = -\frac{\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}}$$

従って、元の精度行列の要素が分かっていれば、これから任意の変数対の偏相関係数を計算できる。X_1、X_2 が条件付き独立であるとは、X_1 と X_2 の偏相関係数がゼロであること ($\omega_{12} = 0$) にほかならない。実際、$\omega_{12} = \omega_{21} = 0$ のとき、同時分布の mgf は次のように分解できる。

$$M_{\boldsymbol{X}}(\boldsymbol{z}) = \frac{M_{X_1,\boldsymbol{X}_b}(z_1, \boldsymbol{z}_b) \cdot M_{X_2,\boldsymbol{X}_b}(z_2, \boldsymbol{z}_b)}{M_{\boldsymbol{X}_b}(\boldsymbol{z}_b)}$$

一変量正規分布の合成 (compound) によって多変量正規分布が生成されることについては、付録 C.4 を参照のこと。そこでは、条件付き mgf を用いて同時 mgf が計算され、また条件付き独立の変数対に対応する偏相関係数がゼロであることも示される。

3.2.3 離散型確率分布における条件付き独立性

離散型確率分布のばあいは、連続型確率分布と異なりまずデータの記述法から説明する必要がある[12]。本来は確率変数として $X_1, X_2, \cdots, X_p$ と表現すべきところを $V = \{1, 2, \cdots, p\}$ と添え字のみで表現し、これを項目 (item) と称する。次にその実現値を $\boldsymbol{i} = (i_1, i_2, \cdots, i_p)$ と表現する[13]。その個々の項目のとりうる状態をカテゴリー (category) ないし因子水準 (factor level) と呼び、$i_j = 1, 2, \cdots, k_j$ とする。データは、たとえば表 3.1 のように記述される。

表 3.1 質的データの表記

item		i_1			i_2				$\cdots$
category		1	2	3	1	2	3	4	$\cdots$
標	1	$\surd$			$\surd$				$\cdots$
本	2		$\surd$				$\surd$		$\cdots$

[12] Edwards(1995), 宮川 (1997) などを参照のこと。
[13] このような表記法は n タプル (n-tupule) と呼ばれる。

これから多元分割表を再構成し、これに対して確率分布をあてはめる。多元分割表のセル $\boldsymbol{i}$ について、その母数を $\theta_{\boldsymbol{i}}$、データを $n_{\boldsymbol{i}}$ とするとき、サンプリングの方法によってたとえば以下のような確率分布が適用される[*14]。

多項分布：

$$P(n_{\boldsymbol{i}}|\theta_{\boldsymbol{i}}) = \frac{N!}{\prod_{\boldsymbol{i}\in I} n_{\boldsymbol{i}}!} \prod_{\boldsymbol{i}\in I} \theta_{\boldsymbol{i}}^{n_{\boldsymbol{i}}}$$

ただし、$\sum_{\boldsymbol{i}\in I} \theta_{\boldsymbol{i}} = 1,\ \sum_{\boldsymbol{i}\in I} n_{\boldsymbol{i}} = N$ である。

多変量ポアソン分布：

$$P(n_{\boldsymbol{i}}|\lambda_{\boldsymbol{i}}) = \prod_{\boldsymbol{i}\in I} \left(\frac{\lambda_{\boldsymbol{i}}^{n_{\boldsymbol{i}}}}{n_{\boldsymbol{i}}!} e^{-\lambda_{\boldsymbol{i}}} \right)$$

ただし、$\lambda_{\boldsymbol{i}} / \sum_{\boldsymbol{i}\in I} \lambda_{\boldsymbol{i}} = \theta_{\boldsymbol{i}}$ である。

確率分布の母数に制約を加えることにより構造を入れたものが対数線型モデルである。制約のない飽和モデルについては次のようになる。

$$\theta_{\boldsymbol{i}} = \exp\left(\sum_{a\subseteq V} u_a(i_a) \right)$$

たとえば二元分割表のばあい、

$$\theta_{i_1,i_2} = \exp\{u_0 + u_1(i_1) + u_2(i_2) + u_{12}(i_1,i_2)\} \tag{3.9}$$

対数線型モデルへの制約の付け方には表 3.2 のようなルールがある。階層 (hiarchical) モデルがもっとも広い範囲のモデルであり、グラフィカル (graphical) モデルはその一部、さらにその一部が分解可能 (decomposable) モデルである[*15]。

表 3.2　対数線型モデルの類型

類型	説明
階層モデル	ある高次交互作用項が存在するとき、それに含まれるより低次の交互作用用 (+ 主効果) もかならず存在する。生成集合で表現可能。
グラフィカル・モデル	生成集合がクリークのみからなる。特定の 2 因子項は条件付き独立関係を表示する。
分解可能モデル	グラフが三角化されている[*16]。広義の回帰モデルに対応する。

階層モデルの制約は、最高次の交互作用項の集合 d_j(生成集合;generator と呼ばれる) により表現される[*17]。たとえば式 (3.9) のばあいは $\{(i_1,i_2)\}$ と表記される。生成集合を用いた制約式は次のようになる。

$$\theta_{\boldsymbol{i}} = \exp\left(\sum_{a\subseteq V, a\subseteq d_j} u_a(i_a) \right)$$

ここまでの準備がととのったところで、離散型分布における条件付き独立性は次のように説明される。たとえば三元分割表についての対数線型モデル $\{(i_1,i_2)(i_1,i_3)\}$ は条件付き独立関係を含んでいる。

$$\begin{aligned}
\theta_{i_1,i_2,i_3} &= \exp\left\{ \sum_{a\subseteq V, a\subseteq \{(i_1,i_2)(i_1,i_3)\}} u_a(i_a) \right\} \\
&= \exp\{u_0 + u_1(i_1) + u_2(i_2) + u_{12}(i_1,i_2)\} \\
&\quad \cdot \exp\{u'_0 + u'_1(i_1) + u_3(i_3) + u_{13}(i_1,i_3)\} \\
&= \exp\left\{ \sum_{a\subseteq V, a\subseteq \{(i_1,i_2)\}} u_a(i_a) \right\}
\end{aligned}$$

[*14] 標本数 N を決めてランダム・サンプリングするばあいは多項分布、調査時間を決めてポアソン・サンプリングするばあいはポアソン分布となる。なお、対数オッズを推定する上ではいずれを適用しても同一の推定量が得られる。これらのことについて詳しくは柳川 (1986) を参照のこと。

[*15] 坂元 (1985) は対数線型モデルについて、1) モデルの良さの統一的な評価基準の欠如、2) モデル構成の不適切さの二点において批判を行っている。これらについては、認識方法にかかわる論点として次節にて論じることとする。

[*16] 三角化 (triangulate) されているグラフとは、4 以上の長さのサイクルをもたない、あるいはクリークのみによって構成されているグラフのことである。

[*17] Goodman(1971) の記法。

$$\cdot \exp\left\{\sum_{a\subseteq V, a\subseteq\{(i_1,i_3)\}} u_a(i_a)\right\}$$

このばあいは、項目 2 と項目 3 が項目 1 が生起したという条件のもとで条件付き独立となっている。またそれはグラフィカル・モデルにおいて $u_{23}(i_2,i_3)=0$ となっている、ということでもある。

単純な (たとえばベルヌーイ分布のような) カテゴリカル分布の合成から複雑なカテゴリカル分布が構成できるのは、正規分布のばあいと同様である (付録 C.4 を参照のこと)。またそれにしたがう n 個の独立な確率ベクトルの和は多項分布にしたがう。

3.2.4 連続型確率変数と離散型確率変数の混在

ここまで複数の連続型確率変数の束の内部、ないし離散型確率変数の束の内部それぞれについての関係性 (独立性、条件付き独立性の有無) を論じてきた。しかし、両者のあいだの関係 (つまり両者が入り混じった状態) については、上と同様の統一的な取り扱いはまだ発見されていない。ここに本質的な困難が見出される[*18]。

連続型変数と離散型変数を含む同時確率分布について、まったく方向性の異なる二つのアプローチが存在する。第一は、条件付きガウス (Conditional Gaussian : CG) 分布によるものであり、第二は離散選択理論である。前者は分散分析の基礎を提供し、後者は一般線型回帰モデルの基礎を提供している。上の本質的困難とは、両者を包含する統一的なアプローチが見出されていない、ということである[*19]。

CG 分布はカテゴリカル分布と正規分布の合成から得られる、次のような同時密度関数で記述される確率分布である。母数の添え字 i はこれが別の確率変数 X_1 に依存していることを示す。

$$\begin{aligned}
&P(X_1=i, \boldsymbol{X}_2=\boldsymbol{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i,\lambda_i)\\
&=\lambda_i(2\pi)^{-\frac{q}{2}}|\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}\\
&\quad \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right\}\\
&=\exp\left(A_i+\boldsymbol{B}_i\boldsymbol{x}-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{\Omega}_i\boldsymbol{x}\right)
\end{aligned}$$

ここで $\boldsymbol{\mu}_i$、$\boldsymbol{\Sigma}_i$、λ_i はモーメント・パラメータ (moment parameter) と呼ばれるが、これらによる表記は不便な点もあるので、正準パラメータ (canonical parameter) A_i、$\boldsymbol{B}_i$、$\boldsymbol{\Omega}_i$ による表記も併用される。正準パラメータによる表記法は対数線型モデルと同様の取り扱いを可能にする[*20]。両者の変換式は以下のとおりである。

$$\begin{cases}
A_i=-\frac{q}{2}\ln 2\pi-\frac{1}{2}\ln|\boldsymbol{\Sigma}_i|-\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i+\ln\lambda_i\\
\boldsymbol{B}_i=\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\\
\boldsymbol{\Omega}_i=\boldsymbol{\Sigma}_i^{-1}
\end{cases}$$

もう一つのアプローチ、離散選択理論の代表例である多項ロジットモデルは次のガンベル分布[*21]から導出される。

$$X_k \sim \mathrm{Gumbel}(\mu_k,\theta)$$

その分布関数は、次のとおりである。

$$F(X_k=x_k|\mu_k)=\exp\left\{\exp\left(-\frac{x_k-\mu_k}{\theta}\right)\right\}$$

$k=1\cdots q$ は選択肢を示し、X_k は k を選択することによる「確率効用」をあらわす。ここで $\epsilon_k=(x_k-\mu_k)/\theta$ と変数変換することにより、$\epsilon_k\sim\mathrm{Gumbel}(0,1)$ ということがわかる。

選択肢 j と k の比較において相対的に j が選択される確率 $p_{j;k}$ は次のようになる。

$$\begin{aligned}
p_{j;k}&=Pr(X_j\geq X_k)\\
&=Pr(\epsilon_k\leq\mu_j-\mu_k+\epsilon_j)
\end{aligned}$$

[*18] 宮川 (1997) は次のように指摘している。「・・・CG 分布は、基本的に、質的変数が原因で、その組み合わせで一つの母集団をなし、量的変数は結果系の特性というモデルとみなせる。これは・・・判別分析のモデルと同じである。しかし、結果系である反応を質的変数で観測する場合もある。そのような場合に安易に CG 分布を仮定するのは危険である。例えば、Whittaker が指摘しているように、量的変数・・・が要因で、質的変数が反応の時、・・・ロジスティック回帰モデルを想定 (すると)、・・・同時密度関数は・・・CG 分布にはならない。」、「多変量の内部関連を記述する場合に、量、質が混在すると、途端に難しくなり、標準的手法が確立していないのが現状である。・・・情報損失は覚悟の上で、量的変数をカテゴライズして多次元質的データとして解析することが一案である。」

[*19] ここでは「困難」と表現したが、因果性を識別するという観点からは、この非対称性はむしろ好都合ともいえる。

[*20] なお、$\boldsymbol{\Sigma}_i$ について便宜的に i にかかわらず同一と仮定されることが多い。これを等質 (Homogeneous) モデルと呼ぶ。

[*21] ガンベル分布は三種ある安定極値分布のひとつである。極値分布とは最大値または最小値のしたがう分布のことであり、これが安定しているとは標本数 $n\to\infty$ の極限において、退化も発散もせず同一の関数型の分布を保つことを指す。

すべての $k \neq j$ について同様の比較が行われるとし、しかもそれぞれの比較は互いに影響を与え合わないものとするならば[*22]、また ϵ_j が既に確定されたものだとするならば、

$$\begin{aligned} p_j|\epsilon_j &= \prod_{k\neq j} p_{j;k} \\ &= \prod_{k\neq j} \exp\left\{e^{-(\mu_j-\mu_k+\epsilon_j)}\right\} \end{aligned}$$

実際には ϵ_j は確定されたわけではないので、ϵ_j について期待値をとると、

$$\begin{aligned} p_j &= \int_{-\infty}^{\infty} p_j|\epsilon_j f(\epsilon_j) d\epsilon_j \\ &= \int_{-\infty}^{\infty} e^{-\epsilon_j} \exp\left\{-e^{-\epsilon_j} e^{-\mu_j} \sum_k e^{\mu_k}\right\} d\epsilon_j \end{aligned}$$

ここで、$A \equiv e^{-\mu_j}\sum_k e^{\mu_k}$ とおけば (また、$-\epsilon_j + \ln A = -t$ と変数変換することにより)、

$$\begin{aligned} p_j &= \int_{-\infty}^{\infty} e^{-\epsilon_j} \exp\left\{-e^{-\epsilon_j+\ln A}\right\} d\epsilon_j \\ &= \int_{-\infty}^{\infty} e^{-t-\ln A} \exp\left\{-e^{-t}\right\} dt \\ &= e^{-\ln A} = \frac{e^{\mu_j}}{\sum_k e^{\mu_k}} \end{aligned}$$

これが多項ロジットモデルと呼ばれるものである[*23]。このモデルによれば、選択肢間の対数オッズは次のように表示される。

$$\ln \frac{p_j}{p_k} = \mu_j - \mu_k$$

ここから左辺をリンク関数として一般化し、右辺を線型回帰モデルとする一般化回帰モデルの一群が現れる。

以上の導出過程を別の方面から振り返ってみる。二つの連続型確率変数 (「確率効用」) がそれぞれ次のような Gumbel 分布に従っていたとする。

$$X_1 \sim \mathrm{Gumbel}(\mu_1, \theta), \quad X_2 \sim \mathrm{Gumbel}(\mu_2, \theta)$$

このとき、$X_1 - X_2 \sim \mathrm{Logistic}(\mu_1 - \mu_2, \theta)$ が得られる。このことは次のような計算から分かる。最後の行が Logistic 分布の mgf である。

$$\begin{aligned} M_{X_1-X_2}(z) &= M_{X_1}(z) \cdot M_{X_2}(-z) \\ &= \left\{e^{\mu_1 z}\Gamma(1-\theta z)\right\} \cdot \left\{e^{-\mu_2 z}\Gamma(1+\theta z)\right\} \\ &= e^{(\mu_1-\mu_2)z}\Gamma(1-\theta z)\Gamma(1+\theta z) \\ &= e^{(\mu_1-\mu_2)z} B(1-\theta z, 1+\theta z) \end{aligned}$$

これから、潜在変数であってけっして観測されないところの二つの「確率効用」の差 $X_1 - X_2$ が Logistic 分布に従っていることがわかる。他方で、観測されているのは二つの選択肢の選択結果 (確率) と (それに影響を与えているであろうと予想されている) 原因に相当するそれ以外の確率変数 (これが μ_1 と μ_2 の違いをもたらす) である。仮説的な「確率効用」はこれら結果と原因を媒介するものとして分析者の脳裏に描かれているのである[*24]。

[*22] この条件は i.i.a.(independence from irrelevant alternatives) の仮定と呼ばれる。この条件は強すぎるわりには非現実的に見えることから、これまで多くの批判を浴びてきた。またこの条件を回避、ないし緩和する試みも数多くなされてきた。

[*23] もっとも単純なばあいとして選択肢が二つの単項ロジットモデルを包含し、それ以外にも選択が多段階でなされるとする入れ子 (nested) ロジットモデル、選択により順序尺度が構成されるとする順序 (ordered) ロジットモデルなどのバリエーションが考案された。それらすべての総称として離散選択モデルがある。

[*24] 導出過程を振り返ってみると、効用が Gumbel 分布にしたがっていなければならない理由はかならずしも明確でないことに気が付く。確かに Gumbel 分布は安定極値分布であるという意味では順序統計量を記述するのに適当に見える。しかし、そのことが理論上の必須の条件として要請されているわけではない。事実、離散選択理論はロジットモデルに限定されるわけではなく、たとえば (Gumbel 分布の代わりに正規分布を採用した) プロビットモデルでも本質的な意味が変わるわけではない。

3.3　条件付き独立性から因果性へ

われわれは確率変数間の関係性 (あるいは関係性のなさ) を「独立性」および「条件付き独立性」という用語で叙述してきた。そこから因果性に進むには哲学的な困難が待ち受けている。というのも、独立性の概念は条件付き確率 (ベイズ・ルール) に基礎をおいており、それだけでは互いに独立ではない二つの事象のうちどちらが原因でどちらが結果であるかを区別することができないからである。Pearl(2009) の「確率論では〈原因〉という言葉を扱うことはできない」という言明はまさにこのことを指す。

3.3.1　D.Hume の議論

因果性について、批判的 (つまり経験主義的) 観点から思弁を加えた最初の人はスコットランドの哲学者 D.Hume であると言われている。そこでわれわれも彼の言説から検討をはじめることにしよう。

イギリス経験論者の系譜に属する Hume の懐疑論の功績は、因果性を神秘的な充足理由率から引きはがしたことである。彼の意図は Galilei や F.Bacon の実験的態度を人間の精神活動の領域にたいしてまで拡大して適用しようという、それ自体はまったく邪念のないものだった。その結果、彼の議論は、

1). 因果性 (causality) と関連性 (association) の違いを明らかにする一方で、
2). 因果性を意識内の現象とみなす誤びゅうをもたらした。

前者の側面は、適切な因果推論に結びつく実践的な指針を統計学者に与えるのと同時に、統計学者の議論を相関性 (correlation) の枠内に押しとどめ、因果性について語ること自体が不健全であるとする保守的な風潮を生み出した。後者は、せっかくの経験論を唯物論にではなく観念論に結びつけてしまい、別の意味での保守的なイデオロギー (「経験批判論」) の隆盛をもたらした。

後者については V.I. レーニン (1953) による以下の批判がある。Lenin は、Hume の懐疑論が一方では感覚を精霊によって説明することを否定しながらも、他方では同じく物質によって説明することをも拒否していると論じている。そして、1) 哲学上の流派を分けるものが、認識の源泉を自然の客観的法則性におくか、そうではなく観念の領域におくのかの違いであること、2) こうした区分に立つかぎり、Hume の主張は唯物論と対立するものであることを指摘している。

> ・・・ヒュームは「人性論」第 4 章第 2 節「感官にかんする懐疑論について」のなかで語っている、「われわれの知覚はわれわれの唯一の対象である」と。ヒュームは、物、精霊、等々の働きかけによって感覚を説明することを拒否すること、知覚を一方では外界へ、他方では神または未知の精霊へと帰着させることを拒否することを懐疑論と呼んでいる。

> 哲学上の流派を分かつほんとうに重要な認識論上の問題は (因果的連関についてのわれわれの記述がどの程度の正確さに達したか、またこれらの記述が正確な数学的等式で表現されることができるかどうか、という点にあるのではなく)、これらの連関についてのわれわれの認識の源泉は自然の客観的法則性であるのか、それともわれわれの心の性質、つまり、一定の先天的真理、等々を認識するところの心に内属する能力であるのか、という点にある。これこそが唯物論者フォイエルバッハ、マルクス、エンゲルスを、不可知論者 (ヒューム主義者) アヴェナリウス、マッハから終局的に分かつものである。

われわれは、哲学上の論争がたんなる真理の探究というばかりではなく、階級闘争の一翼を担うイデオロギー闘争の側面をもつことを知っている。そのかぎりで Lenin が挙げた哲学上の流派の区分には大きな意義があると考える。しかし、Lenin の議論は Hume の見解への内在的な批判ではない。Lenin が 1908 年の時点で Hume への内在的批判を志向しなかったのは、実践的唯物論のその後の発展にとって大きな損失であった。

Hume においては懐疑論にとどまっていたものが、その後の E.Mach、K.Pearson においては、自然の客観的法則性の、あるいは因果性という概念の確信的な否定にまで発展する。Lenin は両者の見解をそれぞれ以下のように引用し、その引用によって両者の見解への疑わしさ

[25] Lenin の批判の主眼は、両者の見解が「唯物論ではない」ことを明らかにすることにあり、両者への批判それ自体は目的となっていない。なぜそうなのかと言えば、Lenin にとって、ロシアの自称マルクス主義者たちがマルクス主義者であると僭称していることを暴くことが目下の目的になっていたからである。

を表明している。ただし、両者の直接の批判はここでは課題とされていない[*25]。

われわれは、なぜ Pearson が〈客観的法則性〉と〈法則〉(法則性が認識作用を通じて掴まれたもの) を取り違えるに至ったのか、をすぐ後で論じることにする。

> マッハをとりあげてみよう。「因果性と説明」というとくにこの問題にかんする章 (「熱学の原理」) にはこう書いてある、「それにもかかわらず (因果性の概念の) ヒュームの批判は依然として正しい」。カントとヒュームは因果性の問題をちがった仕方で解決している。(その他の哲学者たちにはマッハはかえりみさえもしない！) ヒュームの解釈に「われわれは賛成する。」「論理的必然性・・・以外の他の必然性、たとえば物理的必然性なるものは、まさに存在しない」。これはまさにフォイエルバッハがあのように決定的に闘争したところの見解である。

> イギリス人のカール・ピアスンは、彼に特有のはっきりした書き方で次のように表現している。「科学の法則は外界の要因であるよりはむしろ、人間の心の産物である」(「科学入門」)。「自然を人間の主君として畏敬する人々は、詩人であれ唯物論者であれ、彼らがあがめる秩序や複雑さが彼ら自身の記憶や思想とすくなくとも同じ程度に、人間の知覚能力や推理能力の産物である、ということをあまりにもしばしば忘れている」。「自然法則の包括的な性格は、人間精神の巧妙さのおかげである」。「人間は自然法則の創造者である」と第 3 章第 4 節では述べている。「人間は自然に法則をあたえるという叙述のなかには、自然は人間に法則をあたえる、というその逆の叙述のなかによりも、より多くの意味がある」。・・・因果性の問題に充てられている第 4 章第 11 節はピアスンのテーゼを定式づけている。「必然性は概念の世界のうちにあり、知覚の世界には属しない。」

Hume の議論の第一の側面、因果性と関連性 (相関性) の峻別という問題に戻ろう。Pearl(2009) はこの側面からただちに次のような現実的な課題が引き出される、と指摘している。

> Hume の考えによると、知識はすべて心に刻まれた経験に基づく・・・。そのような知識を相関であるとみなし、相関は因果関係を意味しないという事実を受け入れた場合、人間はどうやって因果的知識を得ることができるのか、という因果関係に関する最初の難問が導かれる・・・。

ところがこの課題は、Hume の時代にも、また Pearson の時代にも着手されなかった[*26]。なぜそうなったのか[*27]について、Pearl は次の二点の理由を挙げている。

1). 確率論自体が (あるいは静的な観察という手順自体が)、相関性を説明するには足りるが因果性を説明するには足りないこと。
2). K.Pearson が (因果性を不要のものとして切り捨てたために) 上のことを否定的にではなく、肯定的にとらえてしまったこと。

Pearl は Pearson の当時の心境を次のように説明している。すなわち、Pearson は F.Galton による相関性の発見にきわめて強い印象を受けた。そのことは Pearson 自身の言葉で次のように語られている。

> 私は Drake 時代の海賊のような気分だった・・・。私は、Galton が考えているものは、因果関係よりも広いカテゴリー、すなわち相関関係であり、因果関係は単なる相関関係の極端な事例にすぎないと解釈した。そして、この相関関係という新しい概念によって、心理学、人類学、医学、社会学といった多くの分野を数学的に取り扱うことができるようになった・・・。

この結果、Pearson は因果性という言葉を統計学の中から「追放」するという愚挙に及んだ。Pearson の統計学界における影響力はきわめて大きかったため、その企ては残念なことに成功してしまった[*28]。Pearl は次のように説明している。

> このように Pearson は、相関とは異なる因果関係という独立した概念の必要性を完全に否定した・・・。彼は、生涯この立場を貫き、いかなる学術論文において

[*26] R.A.Fisher による実験計画法の確立をまたねばならなかった。Pearl は次のように言っている。「・・・統計学者 Ronald Fisher が現れ、ランダム化実験を定式化するのに 25 年かか (った)・・・。ランダム化実験は、データに基づいて因果関係を検証できることが科学的に証明された唯一の方法論であり、今日に至るまで、統計学の主流において因果関係の概念を扱うことのできるたった一つの方法論で (ある)・・・。」,「アイデアを展開し、結果を共有する数学言語がないばかりに、多くの科学的発見が数世紀も遅れてしま (った)・・・Karl Pearson が因果ダイアグラムの理論を使っていれば、1901 年にランダム化実験のアイデアを思い付いたのではないか・・・。」

[*27] これらのことは唯物論陣営の観点からすれば、当時の唯物論が Marx によって到達されたはずの実践的唯物論 (「フォイエルバッハ・テーゼ」) の水準から後退してしまったこと (Lenin でさえもその影響を免れなかったこと) の結果でもある。

[*28] 明らかに彼は因果性の概念を決定論と結びつけているように見える。そのこともあって彼は因果性を古臭い観念とみなしたのであろうことがうかがえる。

> も因果関係について議論することは (なかった)・・・。彼が行った「意志」と「力」のようなアニミズム的概念に対する戦いはすさまじく、むしろ因果関係をなくそうとしていたので、決定論に対する拒否反応は相当なもの (だった)・・・。そのため、統計学において因果関係という概念が定着する機会を得る前に、彼は統計学の世界から因果関係を消滅させてしまった・・・。

ところで、Hume 自身はけっして因果性の概念を投げ捨てたのではなかった。Hume による因果性の定義を Edwards(1995) は次のように引用している。

> われわれは「原因」を、別の事態 (another) がそれに引き続いて起こる (be followed) 事態 (object) のことであり、しかも第一のものに類似したすべての事態が第二のものに類似した事態を引き起こすばあいにそのように定義する。言い換えれば、第一のものが存在しなければ、第二のものはけっして存在しない。

Edwards によれば、Hume は因果性について、(a) 知覚されるのは二つの事象の結びつき (conjuncture) にすぎず、因果性自体は直接観察されないこと、(b) また、時空の中での事象の継起性 (contiguous)、つまり原因は結果よりも前に起きなければならないこと、を強調している。

ただし、Edwards は Hume の定義が次のような二つの難点をもつことも指摘している。

1). Hume の定義の前半と後半をそのまま合わせると、原因は結果の必要十分 (neccessary and sufficient) 条件となってしまう。しかし、これは過度の単純化であり、現実には合わない。たとえば、漏電は火事の原因になりうるが、火事にはそれ以外の原因 (石油ストーブの転倒など) もあり得る。

2). Hume の定義は決定論に傾きすぎている[*29]。つまり、ある原因が起きれば、これに対応する結果もかならず生じなければならない、としている。しかし、これも現実に合わない。たとえば「喫煙は肺がんの原因になる」という医学的に十分に意味のある命題は、Hume 的な意味では取り扱いえない。

また Pearl も、Hume の定義では因果性から疑似相関を排除できない難点を指摘している。Hume は二つの事象のあいだの関係にしか言及していないので、観察されない第三の変数がこれら二つの事象に影響を与える可能性、したがって二つの事象のあいだに観察されている相関関係がたんにみかけのものにすぎない可能性を考慮していない[*30]。

因果推論の研究は、こうした Hume による定義の難点を一歩一歩克服していく試みだと言える。われわれが目指すべきことは、Lenin のように Hume の考えを不可知論として一蹴することでもなければ、Pearson のように Hume を絶対視することでもない。

3.3.2　因果推論に枠をはめるものとしての相関性

Pearson は相関関係を因果関係よりも広い意義をもつものと考えたが、このことは真実の一面を突いている。つまり、われわれは標本相関の分析を因果性探求の入り口とみなすことができる。(標本相関行列から推測される) 母相関行列が真の因果関係にかかる情報を含んでいるのだとすれば、因果推論とは母相関行列と真の因果関係との関係を明示することにほかならない。またここにおいて「シンプソンのパラドックス」についても明瞭な説明が与えられる。

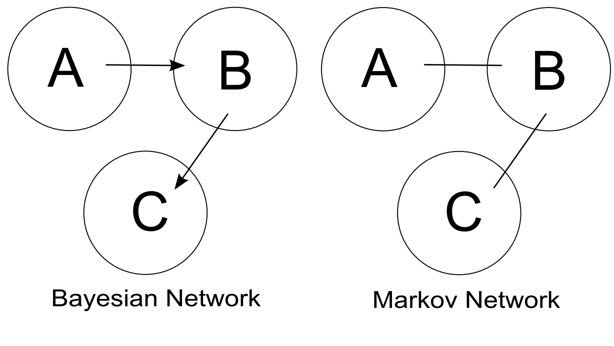


図 3.3　BN と MN

われわれは既に真の因果関係を知っているものとし

*29　言うまでもなく、Newton の影響を強く受けていた Hume からすれば、このような決定論的因果性の理解に陥ることは当時としては無理もないことだった。

*30　第三の変数を考慮するか否かによって、相関係数が変わり、その符号が逆転しさえすることは「シンプソンのパラドックス」として知られている。この現象は Pearson らによって 1899 年に、またその弟子であった G.U.Yule によって 1903 年に発見され、E.H.Simpson により 1951 年に正式に報告され、C.R.Blyth により 1972 年にその名を与えられた。

て、これがどのように現実の相関関係に反映されるのかを知りたい。そのために因果関係を表現するもの (「ベイジアン・ネットワーク」) と相関関係 (条件付き独立関係) を表現するもの (「マルコフ・ネットワーク」) をそれぞれ導入し、両者の関係を考察する。

ベイジアン・ネットワーク (Bayesian Network:BN) とは、事象 (確率変数) の同時確率がベイズ・ルールにしたがって条件付き確率の積に分解されたもののことである[*31]。たとえば、BN において事象 A、B、C のつながりは、条件付き確率 (式 (3.10))、有向グラフ (図 3.3-左) の二通りで表現される。

$$P(A \cap B \cap C) = P(C|B)P(B|A)P(A) \qquad (3.10)$$

マルコフ・ネットワーク (Markov Network:MN)[*32]とは、ここでは無向グラフによって事象間の条件付き独立性を表現したものである。たとえば、事象 A、B、C のつながりが、(任意の二つの要素についてそれ以外の要素を固定したばあいの) 条件付き独立性[*33]の有無によって観察されているものとする。

$$A \not\perp\!\!\!\perp B|C, \ \ A \perp\!\!\!\perp C|B, \ \ B \not\perp\!\!\!\perp C|A$$

これは無向グラフでは図 3.3-右のように表示される。また事象全体の同時確率は式 (3.11) のように極大クリーク[*34]ごとのポテンシャルに分解される。

$$P(A \cap B \cap C) = \phi_{AB}\phi_{BC} \qquad (3.11)$$

ここで ϕ は「ポテンシャル」と呼ばれる (これらはかならずしも確率ではない)。BN における表現と対比すれば、$P(B|A)P(A)$ を一体として ϕ_{AB} とみなしているとわかる。

ここまでの準備をもとにすれば、因果性と相関性の相互関係とは、BN と MN の対応関係にほかならないとわかる。

BN と MN の関係を、条件付き独立性を含む最小の単位である三事象の相互関係から基礎づける (表 3.3)。三つのノードと二つのリンクを含む BN は、有向線分の方向に着目して、head-to-tail、tail-to-tail、head-to-head の三種に区分できる。それぞれについて事象間のつながりを行列 L^{-1} で表現すると、積 $(L^tL)^{-1}$ は条件付き独立性の有無を示すものとなる。これをグラフとして表示したものが MN である。

BN と MN の違いは、head-to-head のばあい顕著であり、BN では a-b 間にいかなる方向の有向線分も存在しないにもかかわらず、MN では a-b 間に無向線分があらたに引かれてある。このように、1)BN に含まれる合流 (collider) の親 (parents) を無向線分で結び、2) それ以外の有向線分をすべて無向線分に置き換える操作をモラリゼーション、これによって得られる MN を BN にたいするモラルグラフと S.Lauritzen は名付けた[*35]。そして、この関係は三つより多い数の事象間においても同様に成り立つ。

BN から MN に、また MN から BN に互いに変換が可能である[*36]。BN からはモラリゼーションによって

[*31] この概念は J.Pearl により導入された。別名として、Bayes network, belief network, decision network, Bayes(ian) model, probabilistic directed graphical model などの用語も用いられる。Pearl によれば、因果推論においてグラフは、1) 実質的な仮定を表現する手段、2) 同時確率の簡便な表現、3) 観察データに基づく推論を円滑化する方法、という 3 つの役割を果たすが、ここではとりわけ 2) の役割を重視していることになる。なお、BN はかならずしもベイズ推計の立場 (すなわち推測における主観確率の承認) を前提としないことに注意せよ。

[*32] 別名として、マルコフ確率場 (Markov Random Field:MRF),probabilistic undirected graphical model とも呼ばれる。MRF は確率過程の添え字集合を、自然数から多次元ベクトルないし多様体上の点に拡大したものであり、歴史的には強磁性体にかんするイジング・モデルを一般化したものである。画像処理などの分野で今日多く活用されている。

[*33] とくに「ペアワイズ・マルコフ性」と呼ばれる。マルコフ性には、他に局所マルコフ性、大域的マルコフ性という定義を異にする概念があり、これら 3 種は相互に関連がある。詳しくは Pearl などを参照のこと。

[*34] クリークとは任意の二頂点が連結されている部分グラフのことであり、極大クリークとはそこに新たにノードを加えることによりクリークではなくなるもののことである。

[*35] 共通の子ども (child) をもつ片親同士をくっつける操作であるがゆえにこの名が与えられた。

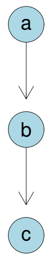
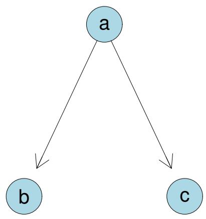
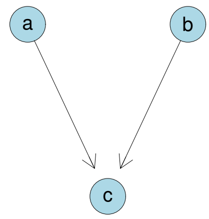
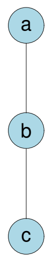
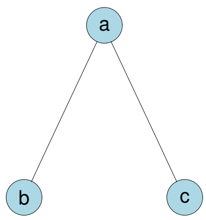
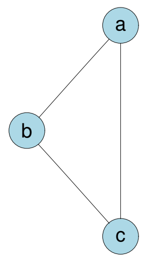
モラルグラフにより条件付き独立を判定する手順は以下のとおりである。1) 判定対象であるすべての変数 (条件付けられるものを含む) について、最小祖先集合による部分グラフをつくる、2) モラリゼーション、3) 条件付けされる変数集合を除いた部分グラフをつくる、4) 対象間が連結されていないことを確認する。

なお、モラルグラフによらず直接 BN より条件付き独立を判定する方法 (Pearl の「有向分離基準」) について付録 C.6 を参照のこと。

[*36] この考えにもとづいて、MN から (ただし、可能な限りでの)BN を得るためのさまざまなアルゴリズムが開発されている。宮川 (2004) の第 8 章を参照のこと。

MN が得られる[*37]。また MN のもとになる条件付き独立性 (ペアワイズ・マルコフ性) を示す情報、たとえば適切な母偏相関行列ないし precision 行列 (またその逆行列である共分散行列、相関行列) が知られているならば、そのコレスキー分解として BN を示すパス係数ないし偏回帰係数が得られるはずである。

表 3.3　BN と MN の対応関係

	head-to-tail	tail-to-tail	head-to-head
BN			
P	$p(a,b,c)=p(c\|b)p(b\|a)p(a)$	$p(a,b,c)=p(c\|a)p(b\|a)p(a)$	$p(a,b,c)=p(c\|a,b)p(a)p(b)$
L^{-1}	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$
$(L^tL)^{-1}$	$\begin{pmatrix} 2 & \underline{1} & \underline{0} \\ 1 & 2 & \underline{1} \\ 0 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 3 & \underline{1} & \underline{1} \\ 1 & 1 & \underline{0} \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 & \underline{1} & \underline{1} \\ 1 & 2 & \underline{1} \\ 1 & 1 & 1 \end{pmatrix}$
MN			

このように BN と MN とのあいだに形式的な対応関係があることから、因果性は結局は相関性に還元できるのではないか、との誤解が生じがちである。しかし、それは正しくない。上のような BN と MN とを形式的に対応させるには、われわれが真の同時確率分布を知っていなければならない。しかし、実際にわれわれが知り得るのは標本共分散行列、経験分布といった量的な観測結果にすぎない。当然、推定される母相関行列は分析者の想定する分布モデル族のいかんにより変わってしまう。また、一つの MN にたいして複数の BN が対応し、こ

[*37] この変換の量的確率変数、とくに多変量正規確率変数の例については付録 C.4 を参照のこと。$X_1 \sim N(0,\sigma_1^2), X_2 \sim N(X_1,\sigma_2^2), X_3 \sim N(X_2,\sigma_3^2)$ は head-to-tail に対応し、$X_1 \sim N(0,\sigma_1^2), X_2 \sim N(X_1,\sigma_2^2), X_3 \sim N(X_1,\sigma_3^2)$ は tail-to-tail に対応している (head-to-head の例も同様に作成できる)。

この関係は対数線型モデルにおいてはよりあからさまである。合流によって多項分布のパラメータが $p_{ijk} = p_i p_j p_{k|i,j}$ と分解できるとき、対数線型モデルの側ではかならず θ_{ijk} の項が必要となり、さらに階層モデルの規約にしたがって θ_{ij} の項も必要となるからである。

[*38] 量的にコレスキー分解の一意性から主張できるように思われるかもしれないが、標本相関行列の正定値性が成立せずに計算が失敗することが多い。

れをグラフの形状のみから特定することはできない[38]。

相関性 (あるいは条件付き独立性) は因果推論が正しい道筋をたどっていることを補助的にガイドするものにすぎず、相関性から因果性を機械的に導き出すことはできない。もちろん、観察された相関性と矛盾する因果関係 (BN) を採用できないという意味では、相関性は因果性の考えうる範囲を強く規定している。しかし、因果性そのものを引き出すには〈観察〉ではなく〈実験〉が、あるいは Pearl の言う〈介入〉(intervention) が必要である。

3.3.3 Pearl による因果性と相関性の区別

Pearl(2009) は因果性と相関性とがはっきりと違うことを、以下の諸点から論じている。

1). 因果性にかんする知識は「状態異常や新しい操作を行った際の結果を予測する能力」を与える。「因果モデルの獲得」は「深い理解」、「(環境を) 管理している状態」という感覚につながっている。相関性によってはこうした感覚は得られない。

Pearl によれば、「「深い感覚」とは、物体が昨日どのようにふるまっていたかだけでなく、新しい仮想的な状況の下でどのようにふるまうかもわかる、ということを意味する」[*39]。「「深い感覚」が得られると、物体を実際に制御する方法がなくても、われわれは「管理している状態にある」と感じる。・・・思慮深い推論者は、新しい操作を実際に試さなくても、その作業が結果に与える影響を予想することができる」。

2). 因果関係についての知識、つまり「ある関係が因果的である、または因果的ではないとわかった場合」、「我々の行動に大きな違いが生じる」。これを Pearl は「因果関係の利用」(われわれの用語では「客観的法則性の適用」に相応) にかかわることとして説明している。

3). 因果性は、相関性とは異なり事象間の非対称性を、したがって事象の時間的な継起にかんする非対称性をもつ。Pearl は B.Russell の言葉を引き、多くの物理法則が時間反転対称性をもつのにたいして、因果性ではそうなっていないことを指摘している。ただし、Russell は因果性を否定する立場 (「君主政治のような、過去の遺跡」) からこの違いを強調するが、Pearl は因果推論の領域を積極的に認める立場からそうしている。

4). Pearl は、因果関係にかかる分析者の直観と観測された結果から引き出される統計的結論 (条件付き独立性を含む) とが往々にして一致しないこと、人間の判断において前者が優先されがちであることを指摘している[*40]。両者が食い違うがゆえに、確率論にかかる様々のパラドックスが生み出される。

これらの論点は、唯物論哲学の観点からは次のようにとらえ返すことができよう。

1)'. 因果性の認識が、現象論とは明確に区別された本質論 (「深い理解」) の領域のあることをわれわれに教えている。Pearl の言う「管理している状態」とは、唯物論的目的論 (目的の生産) にかかわることであり、唯物論的自由論の基礎となるべきものである。

これにたいして、相関性の認識にこだわり、因果性の概念を捨て去ることは、現象論的な認識を固定化し、実体論・本質論への移行を拒否することにひとしい。そこにおいて、われわれは不自由のままであり、自由の領域を神秘的なものにとどめおかざるをえなくなる。

2)'. 因果性の認識は、意識の内部のみならず日常的な行動 (実践) にも影響を与える。あらたな因果関係の知識を獲得することは、われわれの実践の方向性を望ましいものに変え、また日々の実践の成否がその知識の正しさを確証する。この実践にはたんに対自然 (生産的実践) ばかりではなく、対社会 (社会的実践) も当然含まれる。

3)'. 因果性のもつ時間性と物理法則の無時間性との食い違いがマッハ主義者をして因果性を否定せしめた最大の論拠であった。均衡論をめぐる唯物論陣営のなかでの論争はこの難問にかかわるものとして整理するべきである[*41]。

---

*39 宮川 (2004) は、因果効果が本質的に「反事実」(counter-factural) を含むこと、それゆえに観測されないということ (「因果推論の基本的問題」) を説明している。つまり、集団についてではなく、ある個人 (たとえば自分自身) についての因果的効果を考えるならば、現在観察されている変数の状態に反する状態を考えねばならない。ゆえに因果性そのものは観測しえないものと考えなければならない。このことはヒュームによる因果性の説明にも合致するところである。しかし、それにもかかわらず、われわれは因果性が客観的に (つまり人間の意識から独立して) 存在することを認めうるし、さらに因果効果を量的に把握することさえもできる。

*40 Pearl はこのことを行動経済学の知見から引き出している。

*41 均衡論の唯物論哲学への持ち込みは次のような経過のなかに見て取れる。1890 年代のオーストリア・マルクス主義によるマッハ主義の取り込み、1900 年代のボルシェビキ内部での論争 (Lenin=Bogdanov)、その後のブハーリン派とデボーリン派の論争 (1920〜30 年代)。後述。

Pearl はこの難問について次のような抽象的な解決を与えている。すなわち、$A \to B$ と $B \to A$ という二つの因果関係が同時に存在するとして、実験においては一方の関係のみが研究対象におかれ、他方は境界条件として視野の外におかれることにより因果性が認知される (「局所管理」)。ところがこの二つを同時に視野のなかに収めてしまうならば「操作するものと操作されるものの区別がなくなり、介入が消えてしまう」、したがって「因果関係も消滅する」。

われわれは Pearl のこのような説明を弁証法論理の今日的な解明という課題に照らして再検討する必要がある。

4)'. 確率をめぐるパラドックスの発生についての Pearl の説明はひとつの興味深い論点を提示している。この論点は科学史の様々な局面とそのなかでの論争に即して確かめられねばならず、またとりわけ行動 (実験) 経済学の進展をどのように評価すべきかという課題のなかで確かめられねばならない。

Pearl は、自身がこの立場に到達したのは 1990 年から 1991 年にかけてのことであり、二つのアイディアが契機になったと言明している。すなわち、(1)1990 年に、BN の親子関係 $P(x_i|pa_i)$ を関数 (SEM) 関係 $x_i = f_i(pa_i, u_i)$ に変換することを通じて、BN の意味論を Laplace 的「決定論」によって議論できることに気が付いたこと[*42]、(2) 変数の操作によって因果ダイアグラム (BN) が受けとる変化についての P.Spirtes による説明を、〈介入〉の観点から解釈できることに気づいたこと、である。それ以前について Pearl は「著者は、統計学の同僚と同じように、条件付独立関係の中心的役割を唱えるという誤りをおかした」と自己批判している。

表 3.4　Pearl(2009) による因果的概念と統計的概念の区別

統計的概念	因果的概念
相関, 回帰, 条件付き独立, 連関, 尤度, 併合可能性, リスク化, オッズ比, 期待値, 条件付け, 周辺化, グレンジャー因果性, 強外生性	ランダム化, 影響, 効果, 交絡, 外生性, 無視可能性, 誤差, 疑似相関, パス係数, 操作変数, 介入, 説明, 反事実

Pearl は因果性と相関性を区別し、しかも前者の領域を積極的に認める立場から、これまで無頓着に使われてきた用語を因果的概念と統計的概念に区分している (表 3.4)。統計的概念とは同時確率分布から形式的に引き出される概念のことであり、因果的概念はそれにとどまることはない。

Pearl によれば、〈介入〉(これを Pearl は do オペレータによって表現している) とは、因果ダイアグラム (BN) にたいして「手術」を施すこと、すなわち、1) 対象とする確率変数に固定的な状態を割り当て、2) その確率変数に影響を与えるリンクを取り除くことである。また〈因果推論〉とは、このような「手術」が結果 (とみなされる確率変数) に与える影響を予測することである。

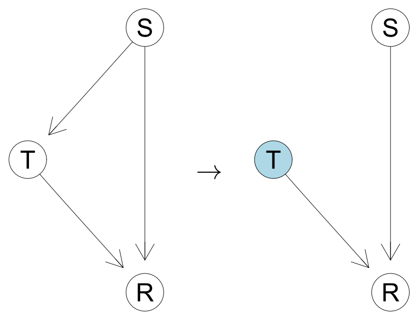


図 3.4　〈介入〉としてのランダム化実験

[*42] Pearl は BN と SEM の関係に基づいて、Laplace の思想 (「自然法則は決定論的に記述され、偶然性は常に潜在的な境界条件を知らないがゆえに現れるにすぎない」) にあらたな解釈を与えている。

BN は条件付き確率の連鎖によって表現されているという意味で本質的に確率的であり、そこには決定論的な要素はまったく含まれていない。しかし、これを SEM で書き直すことにより決定論的要素があらわれる。たとえば、$P(X_1, X_2) = P(X_1)P(X_2|X_1)$ という単純な例を考える。それぞれの確率変数が正規分布にしたがっているとして、その DGM を $X_2 \sim N(\beta X_1, \sigma_2^2)$、$X_1 \sim N(0, \sigma_1^2)$ と記述する。これを SEM で表現すると、$X_1 = u_1$, $X_2 = \beta X_1 + u_2$ となる。ここで、決定論的部分 $X_2 = \beta X_1$ と誤差部分 $u_2 \sim N(0, \sigma_2^2)$ への分割を見て取ることができる。

前者の比重は人間の知識の増大につれて大きくなり、後者は人間の無知の程度の比例して大きくなる、と Laplace は考えた。しかし、このような見かけはあくまでも BN を SEM に書き換えたことによって生じており、客観的な法則性としては何も変わっていないことに注意すべきである。

図 3.4 は Fisher のランダム化実験を〈介入〉として解釈したものである。左側は介入前の観察によって得られた BN を示している。患者の回復状態 R はある疫学的な処置 T によって影響を受けているが、同時にその患者の社会的経済的ステータス S からも影響を受けている。さらに処置 T を受けるかどうかも S により影響されているため、処置 T の回復状態 R への因果的効果はそのままでは分からない。これにたいして右側の介入後では、処置 T の割り当ては乱数によって決められ、S から T へのリンクは切断されている。

介入前後の BN の変化を do オペレータで表現すると式 (3.12) のようになる。前者と後者を比較すると、後者では t が $do(t)$ に置き換えられ、$P(t|s)$ が除かれている。

$$\begin{aligned} P(s,t,r) &= P(s)P(t|s)P(r|s,t) \\ P(s,r|do(t)) &= P(s)P(r|s,do(t)) \end{aligned} \tag{3.12}$$

これにより T の R にたいする因果効果 $P(r \mid do(t))$ がはじめて明らかになる[*43]。Pearl は因果効果を次のように定義している。

因果効果 (Causal effect)

互いに背反な集合 X と Y において、Y にたいする X の「因果効果」を $P(y \mid do(x))$ と表現する。これは X から Y 上の確率分布の空間への関数であり、$P(y \mid do(x))$ は元のモデルから X を終点とするすべての矢線 (方程式) を取り除き、それ以外の X を $X = x$ で置き換えたときの $Y = y$ の確率を与える。

因果効果 $P(r \mid do(t))$ は処置 T を実行したばあいの回復 R の確率を示しており、処置 $T = t$ が観察されたもとでの回復 $R = r$ の確率 $P(r|t)$ とは明確に異なる[*44]。

Pearl は、$P(r|do(t))$ を (もし可能ならば) 具体的に計算する方法 (すなわち周辺分布に書き直して表現する方法) を do オペレーションによる推論規則 (表 3.5) から明示した。また、計算可能であるかどうかを BN から判定する方法 (「バックドア基準」など) をも明らかにした。

表 3.5 の規則 1 は条件付き独立の定義そのものである[*45]。そして、$G_{\overline{X}}$ は X への介入により実現された因果モデルであり、そこにおいて X(および W) が Y と Z を有向分離 (したがって Y と Z は $\{X, W\}$ を条件として独立) している。このときに規則 1 は式 (3.4) に示す条件付き独立の性質 (介入によって実現される確率分布を P_x と表現すれば、$P_x(y \mid z, w) = P_x(y \mid w)$) から直接引き出される。

規則 2 は後述のバックドア基準そのものである[*46]。Z から Y への直接のパスにたいして、$Z \leftarrow W \rightarrow Y$ は「バックドア・パス」を構成する。これが $\{W\}$ によってブロックされているので、残る直接のパスがブロックされてしまえば Z と Y は独立になる。このような状況のもとで Z に介入することは、Z を条件付けることに等しい。したがって、$do(z)$ は z に置き換えが可能である。

規則 3 はやはり後述のフロントドア基準導出の際に中心的な役割を果たす。また脚注*44 で言及した $P(\text{rain} \mid do(\text{wet})) = P(\text{rain})$ は規則 3 のもっとも簡単な例となっている。

規則 2 のグラフ例では W の位置にある頂点はバックドア基準を満たすため、それが観測可能だったならば因果効果 $P(y \mid do(z))$ はバックドア調整によって識別できた。ところが、W が不幸にして観測できない (交絡因子である) ばあい、あらたな工夫が必要になる。そのため、Z と Y のあいだに両者を媒介する頂点を設定する (規則 3 のグラフ例)。もしも X が操作可能ならば、Z への操作を除去できる。ただし、そのためには Z が W の先祖であってはならない (もし先祖であるならば、W にいったん迂回して戻ってくるパスが生じてしまう)。ゆえに推論規則の条件は $G_{\overline{XZ}}$ ではなく $G_{\overline{XZ(W)}}$ となる。

---

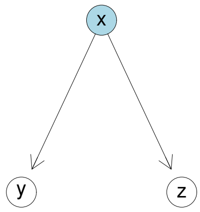
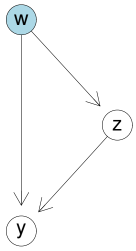
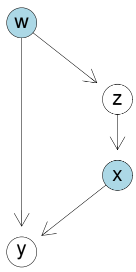
*43　$P_t(r)$、$P(r \mid \check{t})$、$P(r \mid set(t))$ などの記号も同じ意味で使われる。

*44　条件付き確率との違いについて、Pearl はさらに分かりやすい例を与えている。降雨によって地面が濡れる状況において、$P(\text{rain} \mid \text{wet})$ と $P(\text{rain} \mid do(\text{wet}))$ はあきらかに異なる。前者は地面が濡れていることが観察されたばあいの雨が降っている確率を示し、$P(\text{wet} \mid \text{rain})P(\text{rain})/P(\text{wet})$ に等しい。これにたいして後者は $P(\text{rain})$ に等しい (地面を濡らすことは降雨の確率になんら影響を与えないから)。この違いは、因果モデルに方向があることによってもたらされている。

*45　ここで X への介入は (W への条件付けとならんで) 規則の本質からみれば消極的なものであることに注意する (だからこれらは規則の前後で不変の位置にある)。

*46　グラフ例で、W は $\{Z, Y\}$ についてバックドア基準を満たし、したがって因果効果 $P(y|do(z))$ は W を用いたバックドア調整によって識別できる。

表 3.5　do オペレータを使った推論規則 (do-calculus)

	グラフ例	条件	規則
1. 観測値の挿入/削除		$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$ $G_{\overline{X}}$ は G の部分グラフで、X に直接入る矢線をすべて取り除いたもの。	$P(y \mid do(x), z, w)$ $= P(y \mid do(x), w)$
2. 行動/観測値の交換		$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$ $G_{\overline{X}\underline{Z}}$ は $G_{\overline{X}}$ の部分グラフで、Z から直接出る矢線をすべて取り除いたもの。	$P(y \mid do(x), do(z), w)$ $= P(y \mid do(x), z, w)$
3. 行動の挿入/削除		$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\overline{Z(W)}}}$ $G_{\overline{X}\overline{Z(W)}}$ は $G_{\overline{X}}$ の部分グラフで、$Z(W)$ に直接入る矢線をすべて取り除いたもの。$Z(W)$ は Z の要素のうち W の先祖を除いたもの。	$P(y \mid do(x), do(z), w)$ $= P(y \mid do(x), w)$

Pearl は do オペレータを含む因果効果の式が表 3.5 の推論規則を用いて最終的に do オペレータを含まない表現に置き換えることができるときに、その因果効果が「識別可能である」とした[*47]。以下、いくつかの因果モデルについて推論規則の使用例を説明する[*48]。

例 1:バックドア基準

図 3.4 の例では、T への介入により $P(s, r \mid do(t)) = P(r \mid s, do(t))P(s \mid do(t))$ が得られている。S は T から R へのバックドアパスをブロックする $(R \perp\!\!\!\perp T \mid S)_{G_{\underline{T}}}$ ので、規則 2 より $P(r \mid s, do(t)) = P(r \mid s, t)$ である。また S は T の子孫ではないので、規則 3 より $P(s \mid do(t)) = P(s)$ である。ゆえに、

$$P(s, r \mid do(t)) = P(r \mid s, t)P(s)$$

これを S で周辺化することにより、

$$P(r \mid do(t)) = \sum_s P(r \mid s, t)P(s)$$

つまり、因果効果 $P(r \mid do(t))$ は S が観測可能ならば識別できる。

上の関係をより一般的に表現したものが次の「バックドア基準」である。

*47 識別可能性は、計量経済学 (SEM、パス解析) の文脈においてはじめて意識された。後述。

*48 Tikka and Karvanen(2017) はこれらの推論規則から因果効果を観測値によって表現する統計分析パッケージ `causaleffect` を統計分析環境 `R` 上で実装した。

バックドア基準

「W が Z から Y への因果効果についてバックドア基準を満たす」とは、

i). W は Z の子孫ではない
ii). W は Z から Y へのバックドアパス (Z に向かう矢線を含むパスで Z と Y を結ぶもの) をすべてブロックする

またこのとき、Z から Y への因果効果は識別可能であり、その大いさは次の式で与えられる (バックドア調整)。

$$P(y \mid do(z)) = \sum_w P(y \mid w, z)P(w)$$

例 2:操作変数法

図 3.5-左は操作変数 (instrumental variable) 法の最も単純な例を因果モデルで表現したものである (U は交絡因子であり観測できない)。ここで、i)Z から X への因果効果は識別可能 (空集合でブロック) であり、ii)Z から Y への因果効果も識別可能 (同じく空集合でブロック) であるが、iii)X から Y への因果効果は識別できないことに注意する[*49]。i) については、規則 2 より $P(x \mid do(z)) = P(x \mid z)$ であり、ii) については $P(y \mid do(z)) = P(y \mid x, do(z))P(x \mid do(z)) = P(y \mid x, z)P(x \mid z) = P(y \mid z)$ である (やはり規則 2 より)。しかし、$P(y \mid do(x))$ の大いさは分からない。

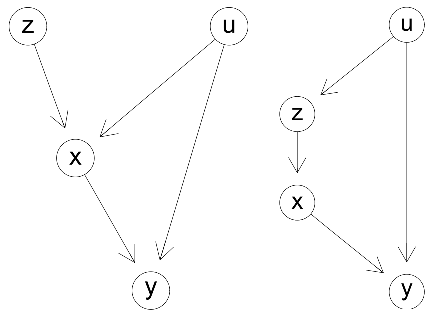


図 3.5　操作変数法と媒介変数法

例 3:フロントドア基準

図 3.5-右はフロントドア基準 (媒介変数法) を説明する因果モデルである。ここで X は Z から Y への因果効果を識別するためのバックドア基準を満たしていないが、Z から Y への有向道[*50]をすべて断ち切っていることに注意する。

求めたい因果効果は $P(y \mid do(z))$ であるが、これは X を用いて次のように分解される。

$$P(y \mid do(z)) = \sum_x P(y \mid x, do(z))P(x \mid do(z))$$

このうち、規則 2 より $P(x \mid do(z)) = P(x \mid z)$ である (第一段階のバックドア基準適用)。

また $P(y \mid x, do(z))$ については次のような変形ができる (第二段階のバックドア基準適用)。

$$\begin{aligned}
&P(y \mid x, do(z)) \\
&= P(y \mid do(x), do(z)) && \leftarrow \text{規則 2} \\
&= P(y \mid do(x)) && \leftarrow \text{規則 3} \\
&= \sum_z P(y \mid z, do(x))P(z \mid do(x)) \\
&= \sum_z P(y \mid z, x)P(z) && \leftarrow \text{規則 2 と 3}
\end{aligned}$$

すべての結果を合わせると次のようになる。

$$P(y \mid do(z)) = \sum_x P(x \mid z) \sum_{z'} P(y \mid z', x)P(z')$$

上の関係を一般化したものが次の「フロントドア基準」である。

[*49] 操作変数法では、因果モデルが線型の構造方程式で記述されているためにパス係数が識別できる。なお Pearl(2009) はそのような場合以外でも因果効果の上限と下限ならば線型計画法を用いて計算できることを示している。
[*50] 有向道とは、パスのうち head-to-tail のみで構成されているものである。

フロントドア基準

「X が Z から Y への因果効果についてフロントドア基準を満たす」とは、

i). X は Z から Y へのすべての有向道を切断する
ii). Z から X へのブロックされないバックドアパスがない
iii). X から Y へのバックドアパスはすべて Z によってブロックされる

またこのとき、Z から Y への因果効果は識別可能であり、その大いさは次の式で与えられる (フロントドア調整)。

$$\begin{aligned} &P(y \mid do(z)) \\ &\quad = \sum_x P(x \mid do(z)) \sum_{z'} P(y \mid z', x) P(z') \end{aligned}$$

前述のように Pearl は因果性と相関性の違いを認めた上で、因果性の領域 (「介入」の観点) を積極的に擁護している。それにもかかわらず、世人は因果性をできるかぎり相関性から説明しようとするし (MN → BN)、観察結果を使って因果効果を識別しようとする (do-calculus)。その動機についてあらためて考察されねばならない。

上の動機の説明として、たとえば、1) 介入それ自体が (現実的に、あるいは倫理的に) 実行不可能なばあいのあること、実行可能であってもリスクが伴うこと (予期しない結果をもたらすばあいがあること)、不可逆的な変化をもたらしうること、などの理由が挙げられる。とりわけ人命にかかわる場合、社会状態にかかわる場合についてそのように言える。また、2) 介入を行ってみる以前に、介入の結果をできるかぎり予見し、介入を必要最小限に抑えたい (リスクを避けたい) という願望がある。つまり勝算ができるだけあるところで、リスク・テイクするべきだ、という (人類に特有の) 無意識の衝動がある。

3.3.4 因果推論と実践

最後に、われわれは認識論における実践の役割という話題に移ることとする。先に、介入と実践の違いを見よう。

介入とは特定の〈因子〉を特定の値に固定することであり、その意味で実験の一部の要素である (3.2 節を参照)。実験は認識を目的とした (それに従属した) 特殊な労働であり、生産的実践の一部が外化、疎外されたものと言える。そこでの目的の主眼はあくまでも因果性 (客観的法則性) の把握あるいは因果効果の識別である。

これにたいして、労働一般 (生産的実践と社会的実践をそのうちに含む) では、意識内に既に形成された目的が外化・対象化される。ここにおいて認識過程は実践の一契機でしかない。実践の結果 (抽象的に表現すれば、生産物) は既に予見されている。この実践の結果を消費することは、外面においても (物質的生活)、内面においても (精神的生活)、実践主体を豊かにする。このうち、認識された世界観という内面において、実験における介入と実践一般とは共通性をもつ。

Lenin は認識論における〈実践〉の積極的な役割を示そうとしたが (「認識論における実践の基準」)、それに部分的にしか成功しなかった。というのも、彼は

1). 真理の問題を生物学的な「有用性」の観点と混同し、
2). その「有用性」を、実践主体による知覚内容と対象のもつ本性との照応 (あるいは一致) を後から証明するにすぎないもの、と理解してしまった

からである。これでは (自己にとって、国家にとって)「役に立たない認識は無駄」と言っているようなものであり、〈実践〉というカテゴリーが利己主義的かつ官僚主義的に歪められてしまう危険をはらむ。また、生物進化 (生物史) の領域へと論点がずらされることによって唯物論的目的論の領域 (社会史) が議論から抜け落ちてしまっている。

> 認識は、それが人間に依存しない客観的な真理を反映した場合にだけ生物学的に有用であり、人間の実践、生命の保存、種の保存に有用であることができる。・・・唯物論者にとっては、人間の実践の「効果」は、われわれの観念とわれわれが知覚する物の客観的本性との照応を証明するものである。

Lenin は「基準」という消極的な言葉を使うべきではなかった。そうではなく、Pearl 的な意味で因果推論において必要不可欠な〈介入〉という要素を〈実践〉論から論ずべきであった。

ただし、Lenin はそのすぐあとで、Fichte にたいして

Feuerbach の行った批判「観念論の根本的な欠陥はまさに、世界は客観的か主観的か、現実的か非現実的か、という問題をただ理論の立場からだけ提出し、かつ解決するという点にある」を引用しつつ、認識と実践との共進化的な発展を展望している (そして前述のとおり、この部分こそ武谷〈技術論〉の重要なヒントになったものである)。

> ・・・実践の基準は、事がらの本質上、けっして人間のなんらかの観念を完全には確証も論証もすることができない、ということをわすれてはならない。この基準はまた人間の知識が「絶対者」に転化するのを許さない程度に「不確定的」であり、同時に観念論や不可知論のあらゆる変種と仮借なく闘争する程度に確定的である。・・・マルクスの理論の道にそって進めば、われわれはますます客観的な真理に近づくであろう (けっしてこの真理を汲みつくすことはないが)。ところが、あらゆる他の道にそって進めば、われわれは混乱と虚偽以外のなにものにも到達することができない・・・。

3.4　小括

本節の結論は以下のとおりである。

- ラプラス「確率の哲学的試論」の重要な特徴は「解析的方法」の積極的な援用にある。彼はこれを確率論の基礎づけに用いるとともに、天体力学への応用、推論ルールの探求 (ベイズ的推論) にも活用した。確率の定義「すべての可能な場合の数に対する好都合な場合の数の比」はその方法論の前提をなし、逆にその方法論を活用するためにはこのような定義こそが必要であった。

  コイン投げやサイコロ遊び、くじ引きなど、単純な事象に対する観察と実験に基づく知見は、確率認識の〈現象論的段階〉とみなすことができる。Laplace のあげる確率の諸原理はこうした現象論的な観察の結果を少数の規則として整理したものであり、これが後に Kolmogorov の「公理」として結実した。

  他方、確率論における〈実体論的段階〉において、確率は確率分布の一部をなし、この確率分布に従う確率変数が強く意識される。すなわち、個々の観測量は確率変数にほかならないことが理解される。複数の確率変数の間の差異がその従う確率法則の違いとして識別される。さらにそれらの間の演算によって新たな確率変数が構成され、これが従う確率法則は、もとの確率変数を支配する確率分布と異なることが見出される。これらのことは上述の「解析的方法」が導きの糸となっている。

- Laplace は必ずしも絶対的な宿命論を語っているわけではなく、事象と事象との間に確率論という客観的法則性 (「つながり」) が存在することを主張しているにすぎない。Laplace が「最も自由な意志」について語っていることは、Leibniz の自然神学の内容を語ったにすぎず、Laplace 自身の見解とは区別されるべきである。その証拠に彼は客観的な確率過程をさまざまに議論している。

  Pearl(2009) は BN と SEM の関係に基づいて、Laplace の思想 (「自然法則は決定論的に記述され、偶然性は常に潜在的な境界条件を知らないがゆえに現れるにすぎない」) にあらたな解釈を与えている。すなわち、BN は条件付き確率の連鎖によって表現されているという意味で本質的に確率的であり、そこには決定論的な要素はまったく含まれていない。しかし、これを SEM で書き直すことにより、客観的法則性は決定論的部分と誤差部分へ分割される。前者の比重は人間の知識の増大につれて大きくなり、後者は人間の無知の程度の比例して大きくなるように見える。しかし、このような見かけはあくまでも BN を SEM に書き換えたことによって生じており、客観的な法則性としては何も変わっていない。

- 関連性 (「独立性」および「条件付き独立性」) から因果性に進むには哲学的な困難が待ち受けている。というのも、独立性の概念は条件付き確率 (ベイズ・ルール) に基礎をおいており、それだけでは互いに独立ではない二つの事象のうちどちらが原因でどちらが結果であるかを区別することができないからである。Pearl(2009) の「確率論では〈原因〉という言葉を扱うことはできない」という言明はまさにこのことを指す。

  因果効果は本質的に「反事実」(counterfactural) を含み、それゆえに観測されえない (「因果推論の基本的問題」)。このことは Hume によ

る因果性の説明にも合致する。しかし、それにもかかわらず、われわれは因果性が客観的に (つまり人間の意識から独立して) 存在することを認めうるし、さらに因果効果を量的に把握することさえもできる。

- イギリス経験論者の系譜に属する Hume の懐疑論の功績は、因果性を神秘的な充足理由率から引きはがしたことである。彼の意図は Galilei や F.Bacon の実験的態度を人間の精神活動の領域にたいしてまで拡大して適用しようという、それ自体はまったく邪念のないものだった。その結果、彼の議論は、1) 因果性 (causality) と関連性 (association) の違いを明らかにする一方で、2) 因果性を意識内の現象とみなす誤びゅうをもたらした。前者の側面は、適切な因果推論に結びつく実践的な指針を統計学者に与えるのと同時に、統計学者の議論を相関性 (correlation) の枠内に押しとどめ、因果性について語ること自体が不健全であるとする保守的な風潮を生み出した。後者は、せっかくの経験論を唯物論にではなく観念論に結びつけてしまい、別の意味での保守的なイデオロギー (「経験批判論」) の隆盛をもたらした。

  Hume 自身はけっして因果性の概念を投げ捨てたのではなかった。因果推論の研究は、Hume による定義の難点を一歩一歩克服していく試みだと言える。われわれが目指すべきことは、Lenin のように Hume を不可知論として一蹴することでもなければ、Pearson のように Hume を絶対視することでもない。

- 因果性の認識は、現象論とは明確に区別された本質論 (「深い理解」) の領域のあることをわれわれに教える。Pearl の言う「管理している状態」とは、唯物論的目的論 (目的の生産) にかかわることであり、唯物論的自由論の基礎となるべきものである。

  これにたいして、相関性の認識にこだわり、因果性の概念を捨て去ることは、現象論的な認識を固定化し、実体論・本質論への移行を拒否することにひとしい。そこにおいて、われわれは不自由のままであり、自由の領域を神秘的なものにとどめおかざるをえなくなる。

- 因果性の認識は、意識の内部のみならず日常的な行動 (実践) にも影響を与える。あらたな因果関係の知識を獲得することは、われわれの実践の方向性を望ましいものに変え、また日々の実践の成否がその知識の正しさを確証する。この実践にはたんに対自然 (生産的実践) ばかりではなく、対社会 (社会的実践) も当然含まれる。

- 因果性のもつ時間性 (歴史性) と物理法則の無時間性 (非歴史性) との食い違いがマッハ主義者をして因果性を否定せしめた最大の論拠であった。均衡論をめぐる唯物論陣営のなかでの論争はこの難問にかかわるものとして整理するべきである。

- Pearl は、因果関係にかかる分析者の直観と観測された結果から引き出される統計的結論 (条件付き独立性を含む) とが往々にして一致しないこと、人間の判断において前者が優先されがちであることを指摘している。両者が食い違うがゆえに、確率論にかかる様々のパラドックスが生み出される。確率をめぐるパラドックスの発生についての Pearl の論点は科学史の様々な局面とそのなかでの論争に即して確かめられねばならず、またとりわけ行動 (実験) 経済学の進展をどのように評価すべきかという課題のなかで確かめられねばならない。

- 一般に人は因果性をできるかぎり相関性から説明しようとするし (MN $\rightarrow$ BN)、観察結果を使って因果効果を識別しようとする (do-calculus)。その動機についてあらためて考察されねばならない。たとえば、1) 介入それ自体が (現実的に、あるいは倫理的に) 実行不可能なばあいのあること、実行可能であってもリスクが伴うこと (予期しない結果をもたらすばあいがあること)、不可逆的な変化をもたらしうること、などの理由が挙げられる。とりわけ人命にかかわる場合、社会状態にかかわる場合についてそのように言える。また、2) 介入を行ってみる以前に、介入の結果をできるかぎり予見し、介入を必要最小限に抑えたい (リスクを避けたい) という願望がある。つまり勝算ができるだけあるところで、リスク・テイクするべきだ、という (人類に特有の) 無意識の衝動がある。

第 4 章

予測と推測・制御

ここからは、実際に行われた観測の結果からどのように統計的な推論を行うべきかの方に関心を移そう。ここにおいて、ベイズ・ルールの意味合いが議論される。本節では前節で取り扱った確率の存在論的側面を後景に退け、認識論的側面を中心に議論を進める。

4.1 予測とモデル選択

ベイズ・ルールは、存在論的に理解されるかぎりでは条件付き確率の定義から代数的に引き出されるものであり、本来はそこになんら新しい意味を付加するものではない。しかし、このルールが〈情報〉というものに結び付けられるや否や (尤度)、統計的推論の新しい方法として受け止められ、その後の著しい統計分析技術の展開に寄与することになった。

4.1.1 Kullback-Leibler 情報量

Kullback(1968) は、ベイズ・ルールから情報量 (divergence) という概念を次のように導き出している。まず密度関数版のベイズ・ルールを書き下す[*1]。二番目の式は全確率の公式である。

$$f_{H|X=x}(\theta) = \frac{f_{X|H=\theta}(x) f_H(\theta)}{f_X(x)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|H=\theta}(x) f_H(\theta) d\theta$$

H_i は原因の候補 (仮説)、$X = x$ は観測結果を示す。二つの仮説 $H_0(\theta_0)$、$H_1(\theta_1)$ を比べて、どちらがよりもっともらしいかを判定する。そのために、以下のようにそれぞれの仮説に対応するベイズ・ルールの比をとり、若干の式変形をおこなう。

$$\ln \frac{f_{X|H=\theta_0}(x)}{f_{X|H=\theta_1}(x)} = \ln \frac{f_{H|X=x}(\theta_0)}{f_{H|X=x}(\theta_1)} - \ln \frac{f_H(\theta_0)}{f_H(\theta_1)}$$

ここで右辺は、対数オッズがどちらの仮説に有利であるかについての観測の前後での変化を示すことがわかる。したがって、左辺はその観測結果 ($X = x$) に含まれる情報量 (H_1 にくらべてどれだけ H_0 が有利かを判別するための) を意味する。言い換えると、左辺[*2]は観測結果 $X = x$ が H_0 にとってどれだけ「証拠としての重み」があるかを示す。

左辺の期待値をとると $I_{0;1}$ のようになる。なお、簡単のために、$f_{X|H=\theta_i}(x) \equiv f_i(x)$ としている。これが、今日 Kullback-Leibler 情報量 (divergence) と呼ばれるものである。

$$I_{0;1} = \int_{-\infty}^{\infty} f_0(x) \ln \frac{f_0(x)}{f_1(x)} dx$$

K-L 情報量は H_0 と H_1 のあいだの相違 (divergence) を示すものであり、距離 (distance) ではない。つまり、距離空間の公理をすべては満たしていない[*3]。

具体的な例として、仮説についてそれぞれの成功確率が θ_i の二項分布に従っているとした場合を以下に示す。

$$H_i : p_i(x) = \binom{n}{x} \theta_i^x (1-\theta_i)^{n-x}$$

*1 想定する確率変数は連続量に限る必要はない。実際、Kullback はより一般的なかたちで表現している。

*2 この値は尤度比 (の対数) として知られている。

*3 距離空間の公理は、i) 非負性 (ギブスの不等式)$d(x,y) \geq 0$、ii) 非退化性 $x = y \Leftrightarrow d(x,y) = 0$、iii) 対称性 $d(x,y) = d(y,x)$、iv) 三角不等式 $d(x,y) + d(y,z) \geq d(x,z)$ から成る。divergence は、このうち iii) と iv) を満たさない。

$$I_{0;1} = n\left[\theta_0 \ln \frac{\theta_0}{\theta_1} + (1-\theta_0)\ln\frac{1-\theta_0}{1-\theta_1}\right]$$

K-L 情報量は $f_0(x)$ と $f_1(x)$ のある種の近さを表す。すなわち、$f_0(x) \equiv f_1(x)$ なら $I_{0;1} = 0$ になり、$f_0(x)$ と $f_1(x)$ が遠いほど $I_{0;1}$ は大きくなる。

4.1.2　最尤法

ここまでは、H_0 と H_1 は互いに対等であり、入れ替えてもなんら差支えのないものとして取り扱ってきた。ここからは、小西・北川 (2004) にしたがい、$f_0(x)$ で示される仮説がたんなる仮説ではなく「真の状態」であると想定する。推定ないしモデル選択とは、できるだけ $f_0(x)$ に近い $f_1(x)$ を候補の中から探すことによって、$I_{0;1}$ を出来る限り小さくすることである。

ここで期待対数尤度の概念が導入される。$f_1(x)$ を評価対象となる任意の確率分布とする。これに対して $f_0(x)$ を真の確率分布として、以下のような期待値をとる。これが期待対数尤度である。

$$L_{0;1} = \int_{-\infty}^{\infty} f_0(x) \ln f_1(x) dx$$

期待対数尤度を用いて K-L 情報量 $I_{0;1}$ を以下のように書き直すことができる。

$$I_{0;1} = L_{0;0} - L_{0;1}$$

ここで推定作業において上の右辺第一項 $L_{0;0}$ はまったく寄与していないことがわかる。そのため推定においては K-L 情報量 $I_{0;1}$ を最小化するのではなく、期待対数尤度 $L_{0;1}$ を最大化しても用が足りる。

ところで、実際は真の確率分布 $f_0(x)$ は未知であり、これを用いて推定を行うことはできない。そのため真の分布の代わりに経験分布 $f_r(x)$ という別の確率分布から見た評価を考える。経験分布 $f_r(x)$ は次のような密度関数で表現される分布である[*4]。

$$f_r(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i)$$

$$L_{r;1} = \int_{-\infty}^{\infty} f_r(x) \ln f_1(x) dx = \frac{1}{n}\sum_{i=1} \ln f_1(x_i)$$

これが (標本) 平均対数尤度であり、これを最大化する推定法が最尤法[*5]である。このような近似は、大数の法則により、$n \to \infty$ で標本平均が期待値に一致することをもって正当化されている。

例：$a, b(b > a)$ を母数とする連続型一様分布

$$H_1 : p(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{それ以外} \end{cases}$$

このとき a、b の最尤推定量は $\hat{a} = \min x_i$、$\hat{b} = \max x_i$、最大対数尤度は $(\hat{b} - \hat{a})^{-1}$ である。

4.1.3　赤池情報量規準

標本平均対数尤度 $L_{r;1}$ にはバイアスが存在する。つまりこの統計量の期待値はそれが近似しようとする真の期待対数尤度 $L_{0;1}$ とは一致せずに、系統的なずれが生じる。赤池弘次[*6]は巧妙な計算によりこのバイアスを計算し、このずれが概ね〈自由パラメータ数〉に等しいことを見出した。これをあらかじめ $L_{r;1}$ から取り除いておくことにより、最尤法をモデル選択の方法論にまで拡張することができる。これが式 (4.1) の赤池情報量規準 (AIC) である[*7]。

$$\mathrm{AIC} = -2L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x)) + 2p \tag{4.1}$$

AIC 使用にかんして、坂元・石黒・北川 (1983) は次のような注意を与えている。これらの点はモデル選択にかかわる難題の多くにかかわり、これらの難題はその後のさらなる技法の発展をうながした。

[*4] ここで δ は Dirac のデルタ関数である。

[*5] 最尤法の発見は、誤差の正規性について新たな論拠を与えることになった (Gauss の逆問題)。C.5 を参照のこと。

[*6] Information theory and an extension of the maximum likelyhood principle, *Proc. 2nd International Symposium on Information Theory*, B.N.Petrov and F.Csaki eds., pp.267-281, Akademiai Kiado, Budapest, 1973. また赤池ら (2007) によれば、その前の 1971 年日本統計学会報告が初出とのことである。

[*7] $L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x))$ は最大対数尤度であり、p が自由パラメータ数である。-2 の係数は歴史的経緯により定められた。導出について付録 C.7 を参照のこと。

1). 標本数と自由パラメータ数とのあいだにトレードオフの関係があること。ただし、この関係を外れてあまりに多数のパラメータを設定したばあいには、この関係は崩壊し、AIC 最小モデルの探索に失敗する。これは、AIC 導出過程に含まれる近似計算による仮定：式 (C.26) が成立しなくなるためである。坂元らは、標本数を n として自由パラメータ数が $2\sqrt{n}$ を越えないことを推奨している。

2). AIC によっては真の自由パラメータ数は分からないこと (次数選択にかんして一致推定量を与えないこと)。これは 1) から当然引き出される結論であり、AIC に対する批判として従来から指摘されてきたものであるが、坂元らは、「予測という立場」においては「真の自由パラメータ数」という概念は無意味であるため、これは欠点ではない、と主張している。

3). 候補モデル間で AIC の差がほとんど見られない[*8]というばあいでも、それらモデルの関数型がまったく異なることがあり得ること。これは AIC(また最尤法) はあくまでも設定されたモデル族のなかでの最適モデル探索にすぎないためである。モデル族の設定が変われば、当然異なる結果が得られる (だからこそ、赤池はモデル候補のバリエーションを広げる意義を強調している)。坂元らによれば、この状況において真の分布は両候補のいずれにも近くないと解釈されるべきである。

4.1.4 モデル選択の実際

ここでは AIC を用いたモデル選択の実例を坂元 (1985) から紹介する。Ries and Smith(1963) は、1008 人の被験者にたいして洗剤の銘柄 X、M にかんする選好を調査した[*9]。このデータには選好 P のほか、銘柄 M の使用経験 U、水温 T、水の硬度 S が含まれている。坂元は、選好 P をもっともよく説明する説明変数の組み合わせを AIC を用いて判定している。

比較対象とするモデルは、表 4.1 の対数線型モデルにより記述される (Goodman の表記法)[*10]。

表 4.1 Ries and Smith データの AIC による評価

No.	Model	df	χ^2	AIC
1	$\{P\}\{STU\}$	11	32.826	10.826
2	$\{PU\}\{STU\}$	10	12.244	-7.756
3	$\{PT\}\{STU\}$	10	28.464	8.464
4	$\{PS\}\{STU\}$	9	32.430	14.430
5	$\{PTU\}\{STU\}$	8	5.656	-10.344
6	$\{PUS\}\{STU\}$	6	6.629	-5.371
7	$\{PST\}\{STU\}$	6	27.999	15.999
8	$\{PSTU\}$	0	0.000	0.000

たとえばモデル 1 に対応する対数線型モデルのパラメータ表示は次のようになる。ここで A は共通部分 $\{STU\}$ に相応する。

$$\begin{aligned}\ln\theta_{\boldsymbol{i}} &= A + a_i^P \\ A &= a_0 + a_j^S + a_k^T + a_l^U + a_{jk}^{ST} + a_{kl}^{TU} \\ &\quad + a_{lj}^{US} + a_{jkl}^{STU}\end{aligned}$$

表 4.1 では 8 つのモデルについて χ^2 統計量と AIC($= \chi^2 - 2df$) が計算されている。伝統的な (Pearson 流の) 統計学ではモデルごとに χ^2 検定が行われるが、モデル選択の立場では、1)AIC が最小となるモデル 5 が最も適切なモデルとして採用される。これは飽和モデル (モデル 8) よりも良い。このようにして、データからその標本数に応じて最適な MN が得られることになる。また、2)$\{P\}$ と $\{STU\}$ が独立であることを主張するモデル 1 と比較してモデル 5 の AIC が格段に低いことから、$\{P\}$ と $\{STU\}$ とのあいだには明らかに関連があると結論される。これは独立性についての仮説検定を不要のものとする。

[*8] これを、AIC 自体を検定するという手段で逃れようとする考えもある。たとえば、下平ら (2004) を参照のこと。下平は「せっかく AIC が抜け出した伝統的な検定の枠組みに戻ってしまうようにも見えるが、実は従来の方法はいわばモデルの正しさの検定であるのに対して、モデルの良さの検定ともいえる新しい方法はむしろ AIC の思想を取り入れたものである。」と言っている。4.3 節も参照のこと。

[*9] Yvonne M. Bishop,Y.M.,S.E.Fienberg,P.W.Holland,"Discrete Multivariate Analysis: Theory and Practice",Springer Science and Business Media, Jul 31, 2007. に数値があげられている。

[*10] なおこれらの計算は R の複数のパッケージによって再現できる。付録 C.8 を参照のこと。

ここでは $\{STU\}$ の内的な構造について不問とされていることに注意する。これは実質的に (目的変数と説明変数の別を前もって指定する) 回帰分析に等しく、そこで想定されている BN は $\{STU\} \to \{P\}$ という連鎖グラフである。

なお、ここでいくつかの統計量のあいだの関係を説明する。まず AIC と逸脱度 (deviance[11])G^2 とは $\mathrm{AIC} = -G^2 - 2df$ の関係にあることが以下の G^2 の定義からただちに分かる。

$$G^2 = 2\sum_i O_i \ln \frac{E_i}{O_i} = -2\sum_i O_i \ln \frac{O_i}{E_i}$$

また G^2 と χ^2 統計量は符号が異なるだけで実質的に同じである[12]。このことは次のような計算から分かる。

まず χ^2 の定義は次のとおりである。

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

関数 $f(x) = x\ln(x/\theta)$ を定め、これを θ のまわりでの二次までのテイラー展開により、

$$f(x) \simeq (x-\theta) + \frac{1}{2\theta}(x-\theta)^2$$

$x = O_i$、$\theta = E_i$ とおくことにより (また $\sum_i(O_i - E_i) \simeq 0$ より)、

$$G^2 \simeq -2\sum_i (O_i - E_i) - \sum_i \frac{(O_i - E_i)^2}{E_i} \simeq -\chi^2$$

坂元は Goodman の対数線型モデル (の運用法) について次のように批判している。すなわち、ある目的変数を説明しようとするとき、すべての観測変数を考慮した飽和モデルを作成する必要はない。またそれは現実的にも不可能、あるいは無駄が多い (セルの観測度数が 0 となるケース、いわゆるスパースなモデルが多くなり、しかも検討すべきモデルの数が指数関数的に増加する)。任意の二つのモデルを比較する規準があれば良い。

この批判は次のことを根拠としている。

1). 目的変数以外の変数の組 (表 4.1 では $\{STU\}$) は共通であり、AIC の差異に寄与しない。説明変数選択においては説明変数内部の構造を特定する必要はない。
2). Goodman は階層モデルすべてを考慮しているが、説明変数選択のうえでは、グラフィカル・モデルに (もっと言えば、分離可能モデルに) 評価対象を限定して良い。グラフィカル・モデル以外の階層モデルを選択したとしても、採用すべき説明変数についての結論は変わらないからである。

4.1.5 AIC の背景論理

赤池情報量規準は、赤池本人が明言しているように、推測統計学の伝統[13]にしたがって生み出されたものではなく、その伝統にたいする明確な批判を志向し、尤度すなわち情報量による統計学の書き直しを要請するものである。その意味で、この批判の源流は R.A.Fisher による K.Pearson への一連の批判[14]にまでさかのぼることができる。

赤池の功績は、尤度の利用にあたってかならずしも「真の構造」を知らなくともよいということを明確にしたことである。すなわち、尤度とは、現在のデータ (観測値) を用いて過去の時点でそのデータを生み出した「構造」を評価したものであるが、そのときに「真の構造」を知らなくとも、尤度の大きい方が「真の構造」に近い、と判断して差し支えないとしている[15]。

赤池は AIC が「オッカムのかみそり」の具体化と解される傾向を指摘した上で、本意はそこにないということも明言している。AIC を「オッカムのかみそり」に譬える者は、AIC がモデル選択すなわち多数の候補からモデルを絞り込むためのものであり、しかも自由パラ

---

[11] 逸脱度はグラフィカル・モデルのステップワイズ探索において多く活用されている。

[12] この事実は Wilks の定理とも呼ばれる。

[13] 赤池の言葉によれば、「過去の統計の平均値」が「将来に対する期待値」を説明するというのが「実用上の基本的な考え方」であり、そこにおいては「確率が統計的なデータの利用と自然に結び付く」。赤池はこのような伝統的な統計学の意義を認めつつも、それだけでは現代の産業社会の要請には応えられない、という理由でその限界を突いているのである。

[14] 竹内・大橋 (1981) を参照のこと。

[15] 赤池はこれを、海抜の高い山の山頂が「天」にヨリ近い、という言葉で譬えている。
なお、赤池は「真の構造」の存在を否定しているわけではないことに注意する。「真の構造」が存在しないとまで主張すると AIC の導出そのものが不可能になる。「真の構造」の存在は認めるが、それを直接的に観測することはできないし、またその必要もない、と言っているのである。これは交絡因子が、観測できないまでもその存在を否定されないことと同じである。

メータ数 P によって $+2p$ の「罰金」がつくことから、説明変数を「節約」するものと解釈している。しかし、赤池が重視するのは絞り込みの過程ではなくむしろそれ以前の多数のモデル候補の提案、モデル・ビルディングである。これを赤池は C.S.Peirce の「アブダクション」の過程として描写している。このような発想の根底にあるものは、「問題の要求に従って既存の理論を批判的に検討すること」であり、「何かを実現しようとする目的意識」、すなわち実践的な立場である[*16]。

AIC はまず制御理論の領域で受容され、その後統計学分野に受け入れられることになった。日本においては、1977 年の日本統計学会大会で情報量規準が共通テーマとされたことが一つの契機となっている。

AIC の導入は、伝統的統計学にたいする批判に根拠をもつ以上、統計学研究者からの反発[*17]も受けたが、それにも関わらず多くの統計分析ソフトは (BIC[*18]とならんで)AIC を出力するようになり、利用者にモデル選択のための一つの判断材料を提供するようになった。この背景には、情報量規準の理念が受け入れられたというよりは、AIC が最尤法の拡張を志向していたこと、そして尤度の利用がパラメータ推定において有力な方法論として広く採用されたこと、また最適な推定量を探索するための尤度計算に必要な繰り返し計算のコストがコンピュータの利用によって劇的に低下したことがあると思われる。尤度計算が当然のようになされたからには、そこから AIC を出力することにはそれほどの手間を要さない。またこの計算コストの低下は、以下に述べる更なる推測法の発展にも大いに寄与することになる。

[*16] 赤池が海軍兵学校で統計学の素養を学んだこと、戦後復興への貢献を志して文部省統計数理研究所に入所 (1952) し、そこで生産プロセスの制御を中心とする研究業績を重ねたことなどが、こうした発想の背景にあるのではないかと思われる。

　また入所した統計数理研究所も統計学者の戦時協力を進めるために学術研究会議の建議 (1944) を受けて設置されながらも、現場と理論との乖離に阻まれ思うような成果を得られなかったこと、戦後に GHQ の占領政策の思惑から解体を免れながらも、米国流のプラグマティズムの洗礼を受けつつ、「民主的に」改組されたこと、なども関係があるのではないかとも推察される。統計数理研究所の設置と改組の周辺については、木村 (2002)、森 (2005) を参照のこと。

[*17] AIC について最初に体系的に編まれたテキストは坂元・石黒・北川 (1983) である。これは 1966 年に刊行開始された「情報科学講座」(共立出版) の A.5「統計理論」の一冊として編まれた。その序で、同講座編集代表である北川敏男は次のように述べている。「本書には、検定論の放棄とか、標本分布論の回避とか、統計数値表の無用とか、推測統計学の築いた伝統に対する批判が、底流しているようである。編集者はこれらの意見については、著者たちの意見に直ちに全面的に同意するものでないことを、特に断っておきたい。」

[*18] BIC(Bayesian Information Criteria) とは、 Gideon E.Schwarz の提唱したモデル選択規準である。積分のラプラス近似を基礎とするもので AIC とは導出原理が異なる。また J.Rissanen の MDL(Minimum Description Length) と形式的な同一性をもつ。赤池 (1996) は AIC と BIC・MDL との混同を批判している。その意味については後述。

4.2 観測の限界とそれへの対応 (1)

存在論と認識論を隔てるものは、観測の限界という問題である。存在論が、それまでに認識のプロセスで得た内容を本質的な客観的法則性が実体を通じて現象する過程として (言い換えれば物質の自己展開の過程として) 記述することであるのにたいして、認識論では、実践の場にあり様々の物質的制約に囚われているわれわれ自身が、われわれの外なる物をいかにして認識できるかということが問われる。そして認識論においては、現実的な観測の限界ということが常に念頭に置かれなければならない。

観測の限界には様々の状況が含まれる。たとえば、不完全な観測値 (欠損値、異常値)、標本数の不足、交絡因子 (現実に観測値に影響を与えているが、直接的に観測できない要因) や潜在変数 (隠れ変数 ; 実体としても特定されていない仮想の観測値) の存在、などの問題がある。計算技術の発達は今日の統計学がこれらの問題に対処することを可能にしつつある。

以下では、ベイズ・モデリング、EM アルゴリズム、リサンプリングといった対処法を具体的に見ていく。

4.2.1 ベイズ・モデリングと EM アルゴリズム

坂元ら (1983) は AIC の利用上の注意の第一として、自由パラメータ数が標本数に比して多すぎないこと (逆に言えば、標本数が自由パラメータ数に相応して潤沢であること) の必要を挙げている。しかしこの要求は、往々にして満たされない。とりわけ時系列解析、因子分析などでこの状況は顕著である。

上の状況にたいして、パラメータそれ自体を定数ではなく潜在的確率変数とみなす、という視点の転換が可能である (式 (4.2))。このばあい、その変数 θ のしたがうべき確率分布 $h(\theta \mid \cdot)$ と (ハイパー) パラメータ ω が想定され、尤度は観測値のみならず潜在変数としてのパラメータをも包含するものに拡張される[19]。坂元 (1985)、また赤池 (1989) を参照のこと。

$$
\begin{aligned}
g(y \mid \omega) &= \int_\theta f(y, \theta \mid \omega) d\theta \\
&= \int_\theta \ell(y \mid \theta) h(\theta \mid \omega) d\theta
\end{aligned}
\tag{4.2}
$$

式 (4.2) について次のような手順でモデル・ビルディングがなされる。

1). 観測値 y を説明するモデル $\ell(y \mid \theta)$ が想定される。パラメータ θ は、「当面の情報処理の要求」を十分に満たすように「たっぷりと多く」加えられる。また「未知パラメータの数が観測データの数を超えていてもかまわない」。

2). (θ の数よりも少数の) ハイパー・パラメータ ω を使ってパラメータ θ にかんする「事前分布」$h(\theta \mid \omega)$ が想定される。これはパラメータにかかる緩やかな制約を表現する。

3). θ を積分消去した周辺分布 $g(y \mid \omega)$ について、これを最大化するハイパー・パラメータ $\hat{\omega}$ が推定される[20]。つまり、式 (4.2) は実は「拡張された尤度関数」である。

4).「事後分布」$\ell(y \mid \theta)h(\theta \mid \hat{\omega})$ を最大化する $\hat{\theta}$ が計算される[21]。

5). (必要に応じて) 最大対数尤度のバイアスが計算される。これはハイパー・パラメータにかんする自由パラメータ数に等しい。これから拡張された AIC(赤池らはこれを ABIC と命名している) が

[19] この観点にしたがえば、通常の定数としてのパラメータも退化分布 (ディラック関数) にしたがう確率変数とみなされる。つまり、ラプラスによる場合の数に依拠する確率の定義は修正を受ける。

[20] 実際にはこの計算を解析的に実行することは困難であり、ここに EM アルゴリズムが注目される理由がある。ただし、EM アルゴリズムの収束は一般には速いとはいえない。

[21] 因子分析のばあいは因子スコアを計算することに相応する。また、この計算は、この枠組みにおいては、$\hat{\omega}$ の決定に関与していないことに注意する。坂元によれば、この計算の根拠の一つは「(事後分布) が (θ について) 単峰で十分にとがっていれば (事後分布の) モードを (事後分布) の一つの代表値とみなして θ の推定値とすることは実用的であろうと期待される点にある」。逆に言えば、事後分布が複数の極値をもつばあいには計算技術上の問題が生じうる。

[22] 赤池は、「ベイズ・モデルの場合・・・モデルの比較はそれぞれのモデルの尤度の比較を通じて実現できる。ただし、いくつかのパラメータが最尤法で決定されたモデル間の比較には、情報量規準 AIC をベイズ・モデルの場合に拡張したものの利用などを考える必要がある」としている。前半の、尤度の比較だけでできるという主張は、拡張された最大対数尤度のバイアスが無視できるほど小さい、との考えと理解される。

計算され、モデル選択に使われる[22]。

そして、実はこの計算は、Dempster,Laird and Rubin(1977) の EM アルゴリズムによって実行できる[23]。EM アルゴリズムは、さまざまな特殊分析手法にかんする文献のなかに埋もれていたものを Dempster らが一般化したものである。もともとは欠損値処理の都合から開発されたものであるが、そればかりではなく、因子分析、用量-反応分析、判別分析、時系列分析など潜在変数がかかわっているとみなせる分析モデルの大半に広く適用可能な一般的枠組みを提供している。EM アルゴリズムにおいては、θ は潜在変数であり、観測値 y と併せて $x = \{y, \theta\}$ という完全データを構成する。

上の計算手順を EM アルゴリズムにしたがって書き直すと次のようになる。E ステップは上の 4) に対応し、M ステップは 3) に対応する。また E ステップと M ステップは計算が収束するまで繰り返して実行される。

a). 上の 1)、2) は基本的には変わらない。ただし初期値 $\omega^{(0)}$ は最初に与えられる。暫定的な対数尤度 $\ln g(y \mid \omega^{(0)})$ も計算される。
b). E ステップ：y と $\omega^{(k)}$ を前提として、完全データ $x = \{y, \theta\}$ の事後分布 (条件付き期待値) $Q(\omega; \omega^{(k)}) = E_\theta[f(x \mid \omega) \mid y, \omega^{(k)}]$ が計算される。
c). M ステップ：$Q(\omega; \omega^{(k)})$ を最大化する $\omega^{(k+1)}$ が計算される。
d). 収束判定：対数尤度 $\ln g(y \mid \omega^{(k)})$ に改善が見られないばあい、あるいはハイパー・パラメータ $\omega^{(k)}$ に変化が見られないばあい、繰り返し計算を打ち切る。そうでなければ、E ステップに戻る。
e). 上の 5) は基本的には変わらない。

EM アルゴリズムの収束プロセスは KL 情報量との関係から見るとその本質的な意味が分かりやすくなる。そこでは潜在変数 θ にかんする事前・事後分布が明示的に導入される。以下はビショップ (2012) の該当部分からの要約である。

θ にかんする事前分布を $q(\theta)$、事後分布を $t(\theta \mid y, \omega)$ とする。後者は次のように記述される。

$$t(\theta \mid y, \omega) = \frac{f(y, \theta \mid \omega)}{g(y \mid \omega)} = \frac{f(y, \theta \mid \omega)}{\displaystyle\int_\theta f(y, \theta \mid \omega) d\theta}$$

周辺分布の対数尤度 $\ln g(y \mid \omega)$ は以下のように $q - t$ 間の KL 情報量 $I_{q;t}$ を用いて記述できる。ここで KL 情報量の基本性質 $I_{q;t} \geq 0$ より、$\mathcal{L}(q, \omega)$ が $\ln g(y \mid \omega)$ にとっての下界をなすことが分かる。

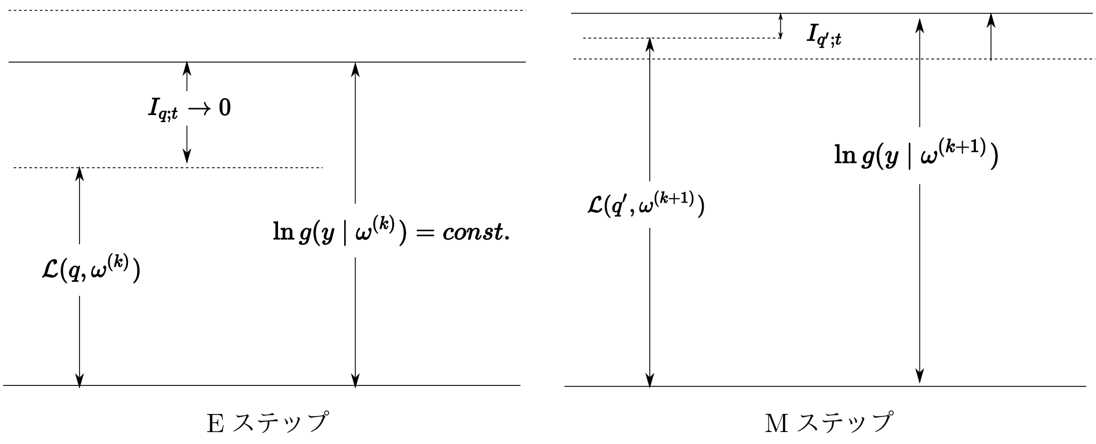


図 4.1　EM アルゴリズムの仕組み

ビショップ (2012) より記号のみ改変

[23] ただし、赤池らはこれが EM アルゴリズムであることに気が付いていない (あるいは EM アルゴリズムの存在を知らない)。またその計算方法も、繰り返し計算による EM アルゴリズムとは違い、一挙に ω について最適化するものとなっている。

$$
\begin{aligned}
\ln g(y \mid \omega) &= \int_\theta q(\theta) \ln \left\{ \frac{f(y,\theta \mid \omega)}{q(\theta)} \right\} d\theta \\
&\quad - \int_\theta q(\theta) \ln \left\{ \frac{t(\theta \mid y,\omega)}{q(\theta)} \right\} d\theta \\
&= \mathcal{L}(q,\omega) + I_{q;t}
\end{aligned}
$$

上の関係を用いて EM アルゴリズムの収束プロセスを表現したものが図 4.1 である。E ステップでは $\omega^{(k)}$ を固定し、事前分布が $q(\theta) \to q'(\theta) = t(\theta \mid y, \omega^{(k)})$ のように事後分布に一致するように更新される ($I_{q;t} \to 0$)。

E ステップの直後で、下界 $\mathcal{L}(q,\omega)$ は式 (4.3) のようになっている。ω を含まない定数項を落としたものが Q 関数である。

$$
\begin{aligned}
\mathcal{L}(q',\omega) &= \int_\theta q'(\theta) \ln \left\{ \frac{f(y,\theta \mid \omega)}{q'(\theta)} \right\} d\theta \\
&= \int_\theta t(\theta \mid y, \omega^{(k)}) \ln \left\{ \frac{f(y,\theta \mid \omega)}{t(\theta \mid y,\omega^{(k)})} \right\} d\theta \\
&= \int_\theta t(\theta \mid y, \omega^{(k)}) \ln f(y,\theta \mid \omega) d\theta \\
&\quad - \int_\theta t(\theta \mid y, \omega^{(k)}) \ln t(\theta \mid y, \omega^{(k)}) d\theta \\
&= Q(\omega, \omega^{(k)}) + const. \qquad (4.3)
\end{aligned}
$$

M ステップでは $q'(\theta)$ が固定され、ハイパーパラメータが $\omega^{(k)} \to \omega^{(k+1)}$ と更新される。$\mathcal{L}(q', \omega^{(k+1)})$ が増大する以上に $\ln g(y \mid \omega^{(k+1)})$ が増大し、対数尤度は着実に上昇する。

EM アルゴリズムの具体例 (1)

(一変量の連続変数 Y、一変量の離散変数 Z からなる)CG 分布において、離散変数が観測できなかったばあい (すなわち Z について周辺化されているばあい)、その結果は混合正規分布となる。このデータから離散変数 Z を復元したい。DGM は次のとおりである。

$$
\begin{gathered}
y_i \mid (z_i = j) \sim N(\mu_j, \sigma_j^2), i = 1, \cdots, n,\ j = 1, 2 \\
z_i \sim \mathrm{Bin}(1, \lambda = (\lambda_j))
\end{gathered}
$$

上の記法と対応させれば、$\theta \to z_i$、$\omega \to \{\lambda_j, \mu_j, \sigma_j^2\}$ のようになる。

完全データ $\{y_i, z_i\}$ にかんする対数尤度は次のようになる。(z_{ij} は z_i を二進表記したもの。ϕ は正規分布の密度関数である。)

$$
L_c = \ln \prod_{i=1}^n \sum_{j=1}^2 \lambda_j^{z_{ij}} \phi(y_i \mid \mu_j, \sigma_j^2)^{z_{ij}}
$$

まず期待対数尤度である Q 関数は次のように計算される。

$$
\begin{aligned}
Q &= E_Z[L_c] \\
&= \sum_{i=1}^n E_Z \left[\ln \sum_{j=1}^2 \lambda_j^{z_{ij}} \phi(y_i \mid \mu_j, \sigma_j^2)^{z_{ij}} \right] \\
&\geq \sum_{i=1}^n E_Z \left[\sum_{j=1}^2 z_{ij} \left\{ \ln \lambda_j + \ln \phi(y_i \mid \mu_j, \sigma_j^2) \right\} \right] \\
&\quad \uparrow \text{イェンセンの不等式} \\
&= \sum_{i=1}^n \sum_{j=1}^2 E_Z[z_{ij}] \left\{ \ln \lambda_j + \ln \phi(y_i \mid \mu_j, \sigma_j^2) \right\}
\end{aligned}
$$

上のうち、$E_Z[z_{ij}]$ は E ステップで先に計算される[*24]。

$$
\begin{aligned}
E_Z[z_{ij}] &= \sum_{j=1}^2 z_{ij} p(z_{ij} \mid y_i) \\
&= p(z_{ij} = 1 \mid y_i) \\
&= \frac{p(z_{ij} = 1) p(y_i \mid z_{ij} = 1)}{p(y_i)} \\
&= \frac{\lambda_j^{(k)} \phi(y_i \mid \mu_j^{(k)}, \sigma_j^{2(k)})}{\sum_{j=1}^2 \lambda_j^{(k)} \phi(y_i \mid \mu_j^{(k)}, \sigma_j^{2(k)})}
\end{aligned}
$$

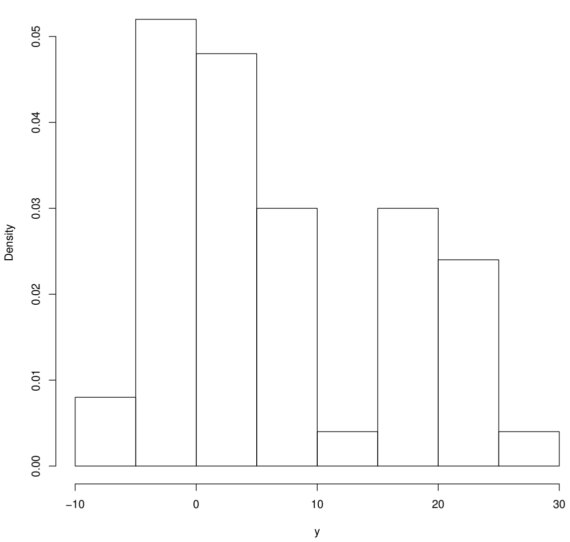
M ステップでは、$E_Z[z_{ij}]$ を前提に Q 関数が最適化される。解は次のようになる。

$$
\begin{aligned}
n_j^{(k+1)} &= \sum_{i=1}^n E_Z[z_{ij}] \\
\mu_j^{(k+1)} &= \frac{1}{n_j^{(k+1)}} \sum_{i=1}^n E_Z[z_{ij}] \cdot y_i \\
\sigma_j^{2(k+1)} &= \frac{1}{n_j^{(k+1)}} \sum_{i=1}^n E_Z[z_{ij}] \cdot \left(y_i - \mu_j^{(k+1)} \right)^2
\end{aligned}
$$

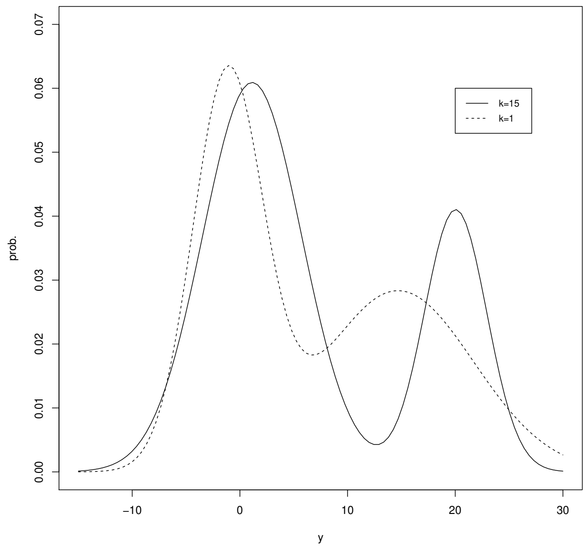
[*24] この値は負担率 (responsibilities) とも呼ばれる。

$$\lambda_j^{(k+1)} = \frac{n_j^{(k+1)}}{n}$$

図 4.2 は二種類の正規乱数 $N(0, 5^2)$ と $N(20, 3^2)$ を混合した標本をもとに EM アルゴリズムで (ハイパー) パラメータ λ_j、μ_j、σ_j を推定した例である。左図では、データ y_i のヒストグラム ($n = 100$、$z_i = 1$ の標本が 70、$z_i = 2$、右図では推定された密度関数 (点線:$k = 1$、実線: $k = 15$) が示されている。初期パラメータは y_i の大きさでソートしたものを半数づつ ($\lambda_j^{(0)} = 0.5$) 分割し、それぞれの標本平均と標本標準偏差を $\mu_j^{(0)}$、$\sigma_j^{(0)}$ とした。繰り返し計算はおおむね 15 回前後で収束している[*25]。最終的な推定値は $\lambda = (0.70, 0.30)$、$\mu = (1.14, 20.04)$、$\sigma = (4.60, 2.90)$、$L_c = -375.80$ となる。



元データのヒストグラム



推定された密度関数 (k=1,18)

図 4.2　EM アルゴリズムの具体例 (1)

EM アルゴリズムの具体例 (2)

坂元 (1985) はある世論調査[*26]からの例を引いている。図 4.3-左で、横軸は回答者の年齢 i、縦軸は「自民党単独内閣がつづく」とする回答の相対度数 $y_{i1}/y_{i\cdot}$ を示したものである。このデータにたいして、坂元は回答 y_i(y_{i0} で No の、y_{i1} で Yes の回答数をそれぞれあらわす) の発生確率 λ_i がある潜在変数 z_i のロジスティック関数で記述され、また z_i それ自体には特定のパラメトリックモデルは与えられず、年齢 i についての二階階差の自乗が正規分布にしたがうことだけを仮定して分析している[*27]。

$$\begin{aligned}
y_{i1} &\sim \mathrm{Bin}\left(y_{i\cdot}, \lambda_i\right), y_{i\cdot} = y_{i1} + y_{i0} \\
\lambda_i &= s(z_i) = \frac{\exp z_i}{1 + \exp z_i} \\
d_i^2 &\sim N(0, \sigma^2) \\
d_i &= (z_i - z_{i-1}) - (z_{i-1} - z_{i-2}) \\
i &= 1, \cdots, C
\end{aligned}$$

完全データにたいする対数尤度 L_c は次のように記述される。ここで $\omega = (\sigma^2, z_{-1}, z_0)$ がハイパー・パラメー

[*25] `R` による計算プログラムは付録 C.8 を参照のこと。

[*26] 1981 年 12 月、東京 23 区の有権者を対象とした「第 51 回東京定期調査」(統計数理研究所) より。設問は「あなたの主義・主張は別にして、将来も自民党の単独内閣はつづくと思いますか。」当時は鈴木善幸内閣であり、その前の第二次大平内閣は党内抗争が激化したことで知られる。

[*27] 前の例とは逆に、連続変量 Z が潜在変数であり、離散変量 Y が観測値となる。坂元はハウスホルダー変換を用いた最小二乗法によって解を求めているが、本稿ではこれを EM アルゴリズムを使って求める。

タである。

$$
\begin{aligned}
L_c &= \ln \ell(\boldsymbol{y} \mid \boldsymbol{z}) + \ln h(\boldsymbol{z} \mid \boldsymbol{\omega}) \\
\ell(y \mid z) &= \prod_{i=1}^{C} \frac{y_{i\cdot}!}{y_{i1}!y_{i0}!} \lambda_i^{y_{i1}} (1-\lambda_i)^{y_{i0}} \\
&= \prod_{i=1}^{C} \frac{y_{i\cdot}!}{y_{i1}!y_{i0}!} \exp(y_{i1} z_i) s(-z_i)^{y_{i\cdot}} \\
h(z \mid \omega) &= \prod_{i=1}^{C} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{d_i^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{C/2} \exp\left\{-\frac{1}{2\sigma^2} |\boldsymbol{Dz}+\boldsymbol{r}|^2\right\}
\end{aligned}
$$

$$
\boldsymbol{D} = \begin{pmatrix} 1 & & & \mathbf{0} \\ -2 & \ddots & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & & 1 & -2 & 1 \end{pmatrix}
$$

$$
\boldsymbol{r} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \\ & \mathbf{0} \end{pmatrix} \begin{pmatrix} z_{-1} \\ z_0 \end{pmatrix}
$$

ここでの問題は、EM アルゴリズムを用いて周辺対数尤度 $E_Z[L_c]$ を極大にするような ω を決定することである。しかし、λ_i が z_i のロジスティック関数で記述されているため、そのままでは z_i による期待値 (積分) を解析的に計算することができない。そこで、より単純な関数 (指数型分布族) でロジスティック関数を近似する (下界を与える) ことにより、EM アルゴリズムを実行可能にする[*28]。具体的には[*29]、

$$
s(z_i) \geq s(a_i) \exp\left\{\frac{z_i - a_i}{2} - \Lambda(a_i)(z_i^2 - a_i^2)\right\}
$$

ただし、

$$
\Lambda(a_i) = \frac{1}{2a_i}\left\{s(a_i) - \frac{1}{2}\right\}
$$

L_c について具体的に展開し、さらに z_i について平方完成すると次のようになる。

$$
\begin{aligned}
L_c &= \ln\left(\prod_{i=1}^{C} \frac{y_{i\cdot}!}{y_{i1}!y_{i0}!}\right) + \sum_{i=1}^{C} \{y_{i1} z_i + y_{i\cdot} \ln s(-z_i)\} \\
&\quad - \frac{C}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} |\boldsymbol{Dz}+\boldsymbol{r}|^2 \\
&\geq \sum_{i=1}^{C} \frac{y_{i1} - y_{i0}}{2} z_i \\
&\quad + \sum_{i=1}^{C} y_{i\cdot} \left\{\ln s(a_i) - \frac{a_i}{2} - \Lambda(a_i)(z_i^2 - a_i^2)\right\} \\
&\quad - \frac{1}{2\sigma^2} |\boldsymbol{Dz}+\boldsymbol{r}|^2 + const. \\
&= -\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_N)^t \boldsymbol{S}_N^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_N) + const.
\end{aligned}
$$

ここで、

$$
\begin{aligned}
\boldsymbol{S}_N^{-1} &= \frac{1}{\sigma^2} \boldsymbol{D}^t \boldsymbol{D} + 2\boldsymbol{\Lambda} \\
\boldsymbol{\mu}_N &= \boldsymbol{S}_N \left\{-\frac{1}{\sigma^2} \boldsymbol{D}^t \boldsymbol{r} + \boldsymbol{b}\right\}
\end{aligned}
$$

ただし、

$$
\begin{aligned}
\boldsymbol{\Lambda} &= \begin{pmatrix} y_{1\cdot}\Lambda(a_1) & & 0 \\ & \ddots & \\ 0 & & y_{C\cdot}\Lambda(a_C) \end{pmatrix} \\
\boldsymbol{b} &= \frac{1}{2} \begin{pmatrix} y_{11} - y_{10} \\ \vdots \\ y_{C1} - y_{C0} \end{pmatrix}
\end{aligned}
$$

E ステップにおいては、ハイパーパラメータにたいする適当な初期値を前提として $\boldsymbol{S}_N^{-1}$、$\boldsymbol{\mu}_N$ が計算される。これは z_i の事後分布を正規分布 $N(\boldsymbol{\mu}_N, \boldsymbol{S}_N)$ で近似することに等しい。

M ステップについて、まず a_i の更新式を考える。L_C の下界を a_i で偏微分し 0 とおくと、

$$
\begin{aligned}
\frac{\partial}{\partial a_i} E_z\left[\underline{L_c}\right] &= -y_{i\cdot}\Lambda(a_i)'\left\{E_z[z_i^2] - a_i^2\right\} = 0 \\
(a_i^2)^{\text{new}} &= E_z[z_i^2] = diag\left(\boldsymbol{S}_N + \boldsymbol{\mu}_N \boldsymbol{\mu}_N^t\right)
\end{aligned}
$$

同様に、$\boldsymbol{r}$ および σ^2 についても最適化すると、次の更新式が得られる[*30]。

[*28] このテクニックは変分ベイズ法として知られている。ビショップ (2012) を参照のこと。

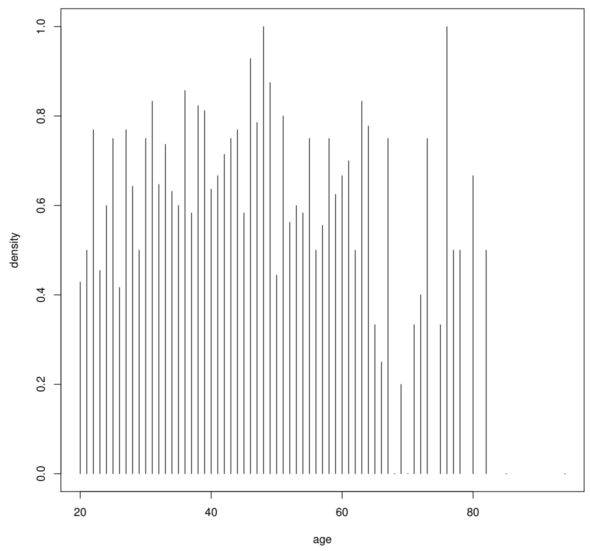
[*29] この式は $\ln s(z_i)$ について、z_i^2 を変数とする一次のテイラー展開より導出される (つまり、$s(z_i)$ を正規分布の密度関数によって近似することに等しい)。ただし、これにより z_i の近傍 a_i を新たなパラメータとして導入しなければならなくなる。

[*30] $\boldsymbol{D\mu}_N$ の i 番目の要素を $D\mu_{N(i)}$ と表現した。

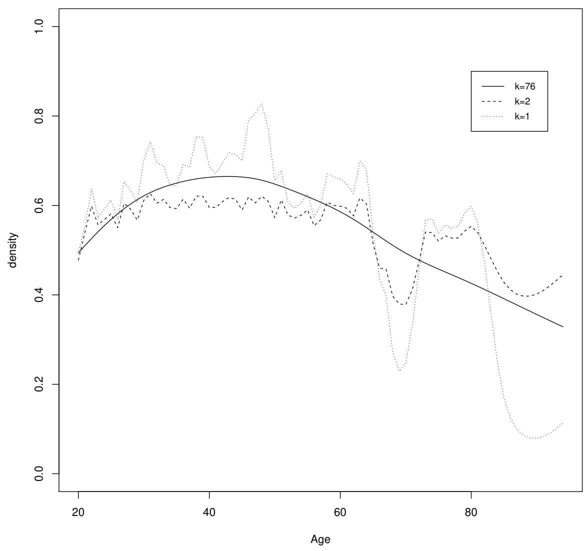
$$(\boldsymbol{r})^{\text{new}} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \\ \multicolumn{2}{c}{\mathbf{0}} \end{pmatrix} \begin{pmatrix} z_{-1} \\ z_0 \end{pmatrix}^{\text{new}} = \begin{pmatrix} -D\mu_{N(1)} \\ -D\mu_{N(2)} \\ \mathbf{0} \end{pmatrix}$$

$$(\sigma^2)^{\text{new}} = \frac{1}{C}\left\{ tr(\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{S}_N) + \sum_{i=3}^{C} D\mu_{N(i)}^2 \right\}$$

図 4.3-右は、k=1,2,81 の繰り返し計算によって得られた $E_z(\lambda_i) = s(\boldsymbol{\mu}_N^{(k)})$ を図示したものである。k=76 のとき、$\sigma^2 = 2.643 \times 10^{-4}$ となっている。これによれば、Yes の回答確率は 43 歳で最大値 0.66 となる[*31]。



元データ



λ_i の期待値 (k=1,2,81)

図 4.3　EM アルゴリズムの具体例 (2)

4.2.2　赤池によるベイズ的方法への評価

赤池 (1989)[*32]はベイズ統計学について、ベイズ・モデリングという方法そのもの (「ベイズ的方法」) と、その考えを教条的に支持する立場 (いわゆる「主観確率の立場」) を峻別しつつ、後者への批判 (とりわけその事前分布の考え方についての批判) を行っている。赤池の主張は当然のことながら伝統的な統計学の立場とは一線を画しており、むしろ「主観確率の立場」が未知パラメータの「客観性」にこだわっていることの矛盾を指摘している。赤池の主張の特異性は、ベイズ的方法の有用性を認めつつ、その正当性 (「良さ」) が自明でないことを同時に指摘している点にある。

> ・・・この方法 (ベイズ的方法) の定義自体は簡単なものであるが、なぜこの方法がよいものであるかは、それほど自明のことではない。・・・この議論の中心は、事前分布・・・の選択にある。・・・この問題をめぐる「哲学的な」議論は多数の人々を巻き込み、結局何らの決定的な解決をも見なかったのである。

[*31] R によるプログラム例を付録 C.8 に示す。坂元 (1985) では 42 歳を最大値としているが、概ね同じ結果である。

[*32] 赤池の論説がその第 3 章として収められている鈴木・国友編 (1989) の成立事情について簡単に触れておく。1988 年 9 月 5〜7 日に「ベイズ統計学とその応用」をテーマとするコンファレンスが東京大学を中心に開催され、そこでの報告と討議をもとに、教科書としても活用できるように編者による序章を加えられて一冊に編まれた。その冒頭には、ベイズ統計学と伝統的統計学をめぐる論争の未解決、それにもかかわらず前者の応用事例が米英を中心に蓄積されたこと、ひるがえって日本では前者に関わる教科書さえほとんど見られないこと、などが参加者の問題意識であったことが記されている。

赤池は、事前分布の選択をめぐってこれまでの統計学界が三潮流に分かれて以下のような論争を繰り広げたことを紹介しつつ、「これらは実際の統計的データの処理に対して広く有効な方法を与えるものとはならなかった」と断じている。

1). R.A.Fisher : fiducial probability の理論は、事前分布の選択を客観的に行おうとする試み
2). Neyman=Pearson : 第一種の誤り、第二種の誤りの概念の導入は、事前分布の導入を避ける理論の展開を試みたもの
3). 主観確率の立場 (de Finetti 等): 「事前分布はとにかく存在しているのであって、各人がその持っている情報を十分に活用すれば決定できるものである」と主張

上の論争と対比するかたちで、赤池はベイズ的方法について自身の積極的な見解を示す。すなわち、パラメータについて、それのみを単独に実在のものとして解釈することなく、事前分布の想定とともにデータからの情報抽出の手段とみなすことがベイズ的方法である、という主張である。

> 未知パラメーターを含むデータ分布が、実はデータから特定の目的に適した情報を抽出するための道具立てであること、事前分布とそれによる事後分布の計算は、データ x が与えられたときに尤度関数 $f(x \mid \theta)$ を通して得られる情報を適切に利用する仕組みを与えるものにすぎない・・・。・・・ベイズ的方法とは、データの与える情報の有効利用を目的とする人工的な仕組みであり、その有効性はデータ分布並びに事前分布の構成の仕方に依存する・・・。

赤池は客観性と主観性についてベイジアン (主観確率の立場) とは異なる見解を示している。主観確率に基づく確率定義についての赤池の見解は肯定的とも否定的とも判別できないが、未知パラメータの客観性が先験的に与えられる (尤度原理などによって) というベイジアンの主張には徹底的に反発している。そして、客観性とはデータ処理の有効性を万人が認めうるかどうかということである、と意味付けている。これは、悟性的思惟を超えて突き進む科学認識の物質性を強力に主張した武谷三男の見解に通じるところがある。

> ベイズ的方法の推進に関して主観確率の立場の人々が展開した議論は・・・何かしら客観性を求めようとする矛盾に満ちたものであった。・・・主観確率の立場の人々がベイズ的方法を論じる場合によく見られる誤りは、未知パラメーター θ が具体的な意味をも含めて先験的に与えられたものとみなして議論を進めることである。この立場に立つと、データからの情報抽出の道具として設定されるデータ分布 $f(\cdot \mid \theta)$ の主観性が全く見逃されてしまう。確率の主観性を主張する人々が、かえって未知パラメーターを何か客観的なものとみなすという混乱がそこにあるのである。

> ・・・ベイズの方法を適用して、人々が「これは役に立つデータ処理法だ」と感じるような結果が得られるように事前分布を構成すればよい・・・。ここでは客観性が要求される。すなわち、誰が見てもその意味が理解でき、納得できるような構成法であることが必要である。

さらに赤池はベイズ的方法の有効性を従来の方法にたいする相対的な優位性に求め、それが必要とされる場面 (逆に言えば、従来の方法がそこで限界にぶつかる場面) を特定していくのである。ここでは、赤池は従来の方法とベイズ的方法を連続的なつながりのあるものとして一体的に把握しようとしている (表 4.2)。両者をつなぐものが、AIC の基礎をなす KL 情報量の考え方である。

> ベイズ的方法の有効性は、何らかの公理 (尤度原理など) によって保証されるものではなく、現実の問題に対して著しい有効性を示すような適用事例の集積を通じて初めて広く一般に受け入れられるようになる・・・。

> ・・・最尤法は多くの問題に適用され、その実用上の有効性は広く認められている。この最尤法が有効に適用されない場面こそ、ベイズ的方法の適用が要求される典型的な場面を与えるであろう。このような場面としてデータの与える情報に比して未知パラメーターの数が多すぎる場合があるわけである。

表 4.2 において、従来の方法とベイズ的方法が対比されているが、後者は前者の拡張とみなされることに注意されたい。つまり、従来の方法での θ_0 とは、ベイズ的方法での θ が退化したものにすぎない。KL 情報量の構成で、従来の方法の θ_0 がベイズ的方法では θ に置き換わっているのはこのためである。

表 4.2　赤池によるベイズ的方法の理解

	従来の方法	ベイズ的方法
推定方法	$x \to \theta(x)$	$x \to p(\theta \mid x)$；推測分布
推定値の良さの評価	$f(\cdot \mid \theta_0) \Longleftrightarrow f(\cdot \mid \theta(x))$	$f(\cdot \mid \theta_0) \Longleftrightarrow g(\cdot \mid x)$ $g(y \mid x) = \int_\theta f(y \mid \theta) p(\theta \mid x) d\theta$ $g(y \mid x)$ は予測分布
KL 情報量 I の構成	$I = E_y \ln f(y \mid \theta_0)$ $-E_y \ln f(y \mid \theta(x))$	$I = E_{y\mid\theta} \ln f(y \mid \theta)$ $-E_{y\mid\theta} \ln g(y \mid x)$
最尤法との関係	直接的に I が最小化される	$E_\theta E_{x\mid\theta}[I]$ が最小化される (EM アルゴリズム)

赤池は、KL 情報量 I の最小化という観点からベイズ的方法の「良さ」を以下のように論証している。従来の方法では I が直接的に最小化される。これは以下の期待対数尤度を最大化することに等しい。

$$E_y \ln f(y \mid \theta(x)) = \int_y f(y \mid \theta_0) \ln f(y \mid \theta(x)) dy$$

同様に、$E_\theta E_{x\mid\theta}\left[E_{y\mid\theta} \ln g(y \mid x)\right]$ がベイズ的方法において最大化される。これは以下のような式変形を通じて、θ を積分消去した二つの分布 $t(x)$、$q(y \mid x)$ を用いた期待値 (E_x、$E_{y\mid x}$) によって書き直される。ここで $p(\theta)$ が事前分布である。

$$\begin{aligned}
&E_\theta E_{x\mid\theta}\left[E_{y\mid\theta} \ln g(y \mid x)\right] \\
&= \int_x \int_y \int_\theta f(y \mid \theta) \underline{f(x \mid \theta) p(\theta) d\theta} \ln g(y \mid x) dy dx \\
&\downarrow t(x) = \int_\theta f(x \mid \theta) p(\theta) d\theta \\
&= \int_x \int_y \int_\theta f(y \mid \theta) \frac{f(x \mid \theta) p(\theta)}{t(x)} d\theta \underline{\ln g(y \mid x)} dy \\
&\cdot t(x) dx \\
&\downarrow q(y \mid x) = \int_\theta f(y \mid \theta) \frac{f(x \mid \theta) p(\theta)}{t(x)} d\theta \\
&= E_x E_{y\mid x}\left[\ln g(y \mid x)\right]
\end{aligned}$$

ここで、KL 情報量の基本性質より、

$$E_{y\mid x}\left[\ln q(y \mid x)\right] - E_{y\mid x}\left[\ln g(y \mid x)\right] \geq 0$$

また等号が成立するのは、$g = q$ のときであるから、

$$\begin{aligned}
g(y \mid x) &= q(y \mid x) \\
&= \int_\theta f(y \mid \theta) \frac{f(x \mid \theta) p(\theta)}{t(x)} d\theta && (4.4) \\
&= \int_\theta f(y \mid \theta) p(\theta \mid x) d\theta && (4.5)
\end{aligned}$$

式 (4.4) と式 (4.5) を比較して、

$$p(\theta \mid x) = \frac{f(x \mid \theta) p(\theta)}{\int_\theta f(x \mid \theta) p(\theta) d\theta}$$

この $p(\theta \mid x)$ が事後分布 (推測分布) である。結果として、KL 情報量を最小化するためには、ベイズ的方法による事後分布を採用するべきである、という結論が得られる[*33]。

ここにおいて、なぜベイズ統計学が 1980 年代から隆盛を誇るようになったのか、その理由の一端が明らかになる。すなわち、コンピュータ技術の飛躍的な発展を通じて最尤法を少ない計算コストで実行できるようになったことが前提条件としてある。ついで、時系列分析など、データ量に比して推定すべきパラメータ数が多すぎる問題を解くニーズが増えたことが挙げられる。こうし

[*33] Kullback が KL 情報量の導出にベイズ・ルールを用いていた点をもって一種のトートロジーだと受け取られるかもしれないが、それは違う。KL 情報量の導出に現れるベイズ・ルールはたんなる確率の積法則の変形にすぎない。

[*34] 主観確率の問題についてはここでは保留し、性急に結論を出さないでおく。

たニーズに対応するためには、従来の方法をベイズ的方法に拡張しなければならなかった[*34]。

4.2.3　BIC について

G.Schwarz はベイズ・モデリングに対応して、「AIC に代わるモデル選択規準」として BIC を提唱した。BIC とは、ベイズ・モデリングにおいて拡張された尤度関数をラプラス近似によって (すなわち、モードを中心とする正規分布を使って) 近似したものであり、AIC のように期待対数尤度の偏りを補正するものではない。この意味で AIC と BIC は同列に扱われるべきものではなく、また赤池 (1996) もその観点から BIC の誤用を批判している[*35]。しかし、幾つかの理由から BIC の誤用[*36]は続いており、赤池の批判も無視されるかたちとなっている。その理由とは以下のようなものである。

1). AIC と BIC の外見上の類似性から、たんに「罰金項」の大きさが違うだけで両者は実質的に同じものと認識されたこと[*37]。尤度さえ計算できていれば、AIC と BIC の計算コストは変わらず、統計ソフトウェアにおいては二つがセットで出力される。そのため、分析家は両者の意味の違いを意識することなく、自身の目的にとって都合の良い方を使っている。

2). AIC が標本数の異なるデータに対して異なるモデルを最適なものとすること (次数選択における一致性の欠如) が、利点ではなく欠点と認識されたこと。BIC がこの点をクリアしているように見えたこと。また、AIC より簡素なモデルを選択する傾向にある BIC が単純さを求める分析家にとってヨリ好ましく見えたこと。

3). BIC がベイズ・モデリングを導出の基礎としていることが、新しく流行している分析手法に飛びつく研究者の好みに合致したこと (彼らはベイジアン＝主観確率の立場を無批判に受け入れている)。

ここで AIC と BIC の違いを明らかにするために、ラプラス近似を使った BIC の導出を振り返ってみる[*38]。ラプラス近似は定積分一般を近似計算するための手法である。k 次元ベクトル変数の関数 $f(\boldsymbol{x})$ について、次の定積分を計算することが求められているものとする。

$$\int f(\boldsymbol{x})d\boldsymbol{x} = \int \exp(\ln f(\boldsymbol{x}))d\boldsymbol{x}$$

$\ln f(\boldsymbol{x})$ を、モード $\boldsymbol{x}_0$ のまわりで二次までテイラー展開し[*39]、

$$\ln f(\boldsymbol{x}) \simeq \ln f(\boldsymbol{x}_0) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{H}(\boldsymbol{x}-\boldsymbol{x}_0)$$
$$\boldsymbol{H} = -\left.\frac{\partial^2}{\partial\boldsymbol{x}\partial\boldsymbol{x}^t}\ln f(\boldsymbol{x})\right|_{\boldsymbol{x}_0}$$
$$\boldsymbol{0} = \left.\frac{\partial}{\partial\boldsymbol{x}}\ln f(\boldsymbol{x})\right|_{\boldsymbol{x}_0}$$

これを積分に代入すると、

[*35] 赤池 (1996) の論点は次のようなものである。1) 事前分布に恣意的な仮定があるため、BIC がモデル規準として最適であるとする主張は誤っている。2) データの観測以前に標本数 n が未確定であるケース (たとえばポアソン・サンプリングなど) では、n 自体もパラメータとみなされるべきである。このとき、「事前分布は常に n と無関係であるべき」とする主張 (一致性:consistency を過度に強調する見解) は不適当である。

[*36] BIC の誤用とは、ベイズ・モデリングが不要の状況で (すなわち、パラメータ数にたいして十分な量の標本が確保できるとき、かつ、潜在変数を想定する必要がないとき・・・言い換えれば、AIC で十分に用が足りるときに)BIC を使うこと、あるいは期待対数尤度の近似法として (EM アルゴリズムなど) もっと良い精度の近似法が使えるときにあえて精度の低い BIC を適用すること、などである。

[*37] 赤池が、AIC はオッカムのかみそりとは関係がない、と明言していたことを想起すべきである。確かに赤池の解釈によるベイズ・モデリングは罰金と理解できないことはない。北川 (2007) によれば、そもそもは次のような時系列データの推測問題;$y_i = T_i + v_i,\quad v_i \sim N(0, \sigma^2)$ について、Whittaker(1923)、Good and Gaskins(1971) が λ^2 をトレードオフ・パラメータとして、$\min_{T_i} \sum_{i=1}^n (y_i - T_i)^2 + \lambda^2 \sum_{i=1}^n (T_i - T_{i-1})^2$ とした方法に赤池の発想の発端があるからである。λ^2 の決め方が分析者の恣意に任されていたところを、赤池はこれを最尤法 $\max \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - T_i)^2 - \frac{\lambda^2}{2\sigma^2}\sum_{i=1}^n (T_i - T_{i-1})^2\right)$ と読み替えることにより、第二項目を事前分布と解釈した。・・・しかし、AIC そのものについてはこの見方は当てはまらない。

[*38] ビショップ (2012)、小西・北川 (2004) を参照のこと。なお、赤池ら (2007) のなかで、甘利は AIC と MLD(BIC) をめぐる「不毛な論争」について、次のように言っている。すなわち、いずれの導出過程においても、階層モデル族の「正則性」が仮定されているが、問題とされるモデルは特異であって、「正則性」が成立しない。ゆえにそうしたモデルでは AIC も MLD も同様に適切でない。・・・その上で甘利は特異分布の例として混合正規分布をとりあげるのである。甘利は ABIC について軽く触れてはいるものの、この枠組みが混合正規分布に適用できることには気が付いていない。

[*39] この操作は、$f(\boldsymbol{x})$ を $N(\boldsymbol{x}_0, \boldsymbol{H}^{-1})$ の密度関数によって近似することを意味する。

$$\int f(\boldsymbol{x})d\boldsymbol{x}$$
$$\simeq f(\boldsymbol{x}_0)\int \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t\boldsymbol{H}(\boldsymbol{x}-\boldsymbol{x}_0)\right\}d\boldsymbol{x}$$
$$= f(\boldsymbol{x}_0)\cdot A$$

また
$$(2\pi)^{-k/2}|\boldsymbol{H}|^{1/2}A = 1$$

より、
$$\int f(\boldsymbol{x})d\boldsymbol{x} \simeq (2\pi)^{k/2}|\boldsymbol{H}|^{-1/2}f(\boldsymbol{x}_0)$$

これがラプラス近似である。ここで $f(\boldsymbol{x}) \to p(\boldsymbol{z} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$、$\boldsymbol{x} \to \boldsymbol{\theta}$ と置き換えることより、BIC が導出される。事前分布 $p(\boldsymbol{\theta})$ が $N(\boldsymbol{\theta}_m, \boldsymbol{V}_m)$ であると仮定し、$-\ln p(\boldsymbol{z} \mid \boldsymbol{\theta})$ のヘッセ行列を $\boldsymbol{J}$ とすれば[*40]、

$$\ln p(\boldsymbol{z}) = \ln \int p(\boldsymbol{z} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$
$$\simeq \frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{H}| + \ln p(\boldsymbol{z} \mid \boldsymbol{\theta}_0) + \ln p(\boldsymbol{\theta}_0)$$
$$= \frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{V}_m^{-1} + \boldsymbol{J}| + \ln p(\boldsymbol{z} \mid \boldsymbol{\theta}_0)$$
$$-\frac{1}{2}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_m)^t\boldsymbol{V}_m^{-1}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_m)$$

ここで $p(\boldsymbol{\theta})$ が十分な広がりをもち、したがって $|\boldsymbol{V}_m^{-1}|$ が無視できるほど小さいとすれば、

$$\ln p(\boldsymbol{z}) \simeq \frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{J}| + \ln p(\boldsymbol{z} \mid \boldsymbol{\theta}_0)$$

さらに $p(\boldsymbol{z} \mid \boldsymbol{\theta})$ が i.i.d ならば、

$$|\boldsymbol{J}| = \left|\sum_{i=1}^n \boldsymbol{J}_0\right| = |n\boldsymbol{J}_0| \simeq n^k + const.$$

したがって、

$$\ln p(\boldsymbol{z}) \simeq \ln p(\boldsymbol{z} \mid \boldsymbol{\theta}_0) - \frac{1}{2}k\ln n + const.$$

この両辺を -2 倍したものが BIC である。この導出過程からもわかるように、BIC は期待対数尤度を粗く近似したものにすぎず、バイアスを補正しているわけではない[*41]。

* * *

前述の坂元の世論調査の例について、期待対数尤度のラプラス近似という観点から振り返ってみる[*42]。L_c を z_i で偏微分することにより、

$$\nabla L_c = diag\left[y_{i1} - y_{i\cdot}s(z_i)\right] - \frac{1}{\sigma^2}\left\{\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{z} + \boldsymbol{r}^t\boldsymbol{D}\right\}$$

EM アルゴリズムにおいてはハイパー・パラメータ σ^2 が考慮されているが。BIC では $\sigma^2 \to \infty$ とされる。これより $\boldsymbol{z}$ のモード $\boldsymbol{\mu}_N$ は次のようになる。これはフィットとしては過剰である。

$$\widetilde{\mu}_{N(i)} \simeq \ln(y_{i1}/y_{i0}),\quad s(\widetilde{\mu}_{N(i)}) \simeq y_{i1}/y_{i\cdot}$$

さらに z_i で偏微分することにより、

$$\nabla^2 L_c = -diag\left[y_{i\cdot}s(z_i)\left\{1 - s(z_i)\right\}\right] - \frac{1}{\sigma^2}\boldsymbol{D}^t\boldsymbol{D} = -\widetilde{\boldsymbol{S}}_N^{-1}$$

よって、

$$\widetilde{\boldsymbol{S}}_N^{-1} \simeq diag\left(\frac{y_{i1}y_{i0}}{y_{i\cdot}}\right) + \frac{1}{\sigma^2}\boldsymbol{D}^t\boldsymbol{D}$$

本来は平滑化のために第一項を多少犠牲にして第二項を大きく (したがって σ^2 を小さく) することによって全体の最適化を図るが、BIC においては第二項は単純に無視され第一項のみが極大化される。これにより確かに対数尤度は大きくなるが、潜在変数 z_i の個数 C は自由パラメータ数としては過剰であり、バイアス評価の尺度としては使えない[*43]。

[*40] ここで $\boldsymbol{\theta}_0$ は、
$$\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(\boldsymbol{z} \mid \boldsymbol{\theta}) - \boldsymbol{V}_m^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_m) = 0$$
の解であるが、後で $|\boldsymbol{V}_m^{-1}|$ が無視できるほど小さいと仮定しているので、実質的に最尤推定量に等しい。

[*41] 赤池はベイズ・モデリングに対応する自身の方法を ABIC と名付けたが、議論をミスリードするという点においては BIC と同罪である。BIC と ABIC は近似度合いの粗さを競っているだけであって、期待対数尤度を近似するという意味では同じものである。

両者をモデル選択に用いて構わないとされるのはバイアス項が 0 とみなされているからであり、ハイパー・パラメータが存在しないか、あるいは、バイアス項が多少変化しても (拡張された) 期待対数尤度の増減に比べて無視できるほど小さいとみなされるからである。

[*42] ビショップ (2012) は近似計算としてみた BIC について、一つのモードのまわりでしか積分を評価しないので、分布の大局的な形状を考慮できず、複数のモードがあるばあいに対処できないこと、$\boldsymbol{\theta}$ が質的変数のばあい (混合正規分布など) に適用できないこと、などの難点を指摘している。

[*43] BIC では、さらに $y_{\cdot\cdot} = nC$ として、$diag(y_{i1}y_{i0}/y_{i\cdot}) \simeq \{n\lambda(1-\lambda)\}^C$ と近似される。この対数が「バイアス評価」項と誤認される。

4.3　観測の限界とそれへの対応 (2)

前節では、観測データが不完全であること (欠損値、潜在変数など) への対応を取り扱った。本節では、逆に観測データのなかに余分なものが紛れこんでいる可能性について論じる。

外れ値、ばらつき、偏り

推定作業において真の分布がわからないということは、ただちに観測値の信頼性の問題を提起する。まず、データを生み出す確率分布についてパラメータばかりではなく分布の形状さえ分からないので、観測データの中に〈外れ値〉が存在していたとしても、どの観測値がそれであるかを名指しすることができない。そればかりか、〈外れ値〉がデータの中に含まれているかどうかさえも知りえない。われわれができることは、真の分布として想定する分布にたいして、特定の観測値の外れ方が大きいか小さいかを意識するのみである。

竹内・大橋 (1981) は「外れ値のモデル」について、次の 3 つのタイプが存在しうることを指摘している。$X_1, \cdots, X_n \sim F(\mathrm{i.i.d.})$ が想定されているものとする。

I. $X_1, \cdots, X_n$ は i.i.d. であるが、F は通常想定されている正規分布よりもスソの長い分布 (例：t 分布、対数正規分布、コーシー分布など) によって記述される。「外れ値」(想定していた分布から見て) は分布のスソで生じている[*44]。

II. 分布はデータの大部分を説明する F とそれとは異質の G との混合分布である。外れ値の多くは分布 G から生じている[*45]。

III. 分布 F からのデータに「誤り」(測定ミス、転記ミス、改ざんなど) を含む観測値が紛れこんでいる。これらの誤りには分布が想定できないという意味でタイプ II とは異なる[*46]。

これらは想定される分布があるという前提のもとで Q-Q 確率プロットや各種検定法[*47]によってその有無やタイプを識別できる。逆に言えば、その想定がないかぎりはこれらを識別することは困難である[*48]。

真の分布がわからないということは、推定結果の品質、すなわち正確度 (偏り) や精度 (ばらつき、分散) についてもやはり分からない、ということをも意味する。偏りにせよ、ばらつきにせよ、真 (と想定される) 分布による期待値の計算に基礎づけれらており、想定される分布が異なればそれらの計算結果も変わるであろう。ここでは正確度 (accuracy) の高さと偏り (bias) の無さ、精度 (precision) の高さと分散 (variance) の小ささをそれぞれ等置している。前者は系統的な誤差の少なさと真値 (reference value) への近さを、後者はランダムな誤差の少なさと再現性の高さを表す。両者の関係の模式図を図 4.4 に示す。

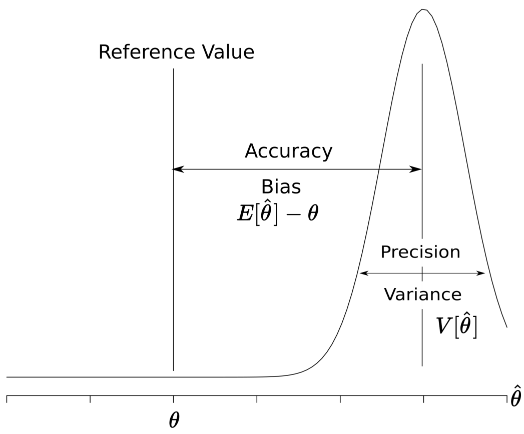


図 4.4　正確度と精度

なお、分散と偏りはトレードオフの関係にある。つま

---

*44 この「外れ値」はあくまでも想定された分布から判断されるもので、その想定の誤りに由来する。ゆえに、よりスソの長い分布から見ればかならずしも「外れ値」ではない。

*45 混合分布であるから、その混合を支配する質的潜在変数が隠されている。このばあい、前節の方法論を適用することが期待できる。

*46 もし誤りを含むデータが「特定できる」ならば、欠損値の扱いと同じになり、やはり前節の方法論が適用できる。

*47 竹内・大橋は、スミルノフ・グラブス検定、ユガミ (skewness)、トガリ (kurutosis)、ギアリー検定などの例を挙げている。

*48 ノンパラメトリック検定 (正規スコア検定、ウィルコクソンの順位和検定、メディアン検定など) が適用できるように思われるかもしれない。たしかにこれらの方法は分布を特定しないが、「ズレのモデル」に依拠しているという意味で万能のものではない。竹内・大橋 (1981) を参照のこと。

*49 したがって、偏りを安易に補正することは、分散を増大させる危険をもたらす。Efron and Tibshirani(1994) はバイアス補正が危険となりうるのは、バイアスの計測自体が標準誤差などの計測よりも難しく、ゆえに安定を欠いているからであり、バイアス補正が積極的に活用

り、分散と偏りの二乗の合計は一定の値以下にはならず、一方の減少は他方の増大を招く[*49]。このことは、以下の平均平方誤差 (mean squared error: MSE) の分解からわかる。真値を θ、その推定値を $\hat{\theta}$ とすれば、

$$\begin{aligned}
MSE(\hat{\theta}) &= E_{\hat{\theta}}\left[(\hat{\theta}-\theta)^2\right] \\
&= E_{\hat{\theta}}\left[(\hat{\theta}-E_{\hat{\theta}}[\hat{\theta}]+E_{\hat{\theta}}[\hat{\theta}]-\theta)^2\right] \\
&= E_{\hat{\theta}}\left[(\hat{\theta}-E_{\hat{\theta}}[\hat{\theta}])^2\right] \\
&\quad +2(E_{\hat{\theta}}[\hat{\theta}]-\theta)E_{\hat{\theta}}\left[\hat{\theta}-E_{\hat{\theta}}[\hat{\theta}]\right] \\
&\quad +(E_{\hat{\theta}}[\hat{\theta}]-\theta)^2 \\
&= V_{\hat{\theta}}[\hat{\theta}]+Bias_{\hat{\theta}}(\hat{\theta},\theta)^2
\end{aligned}$$

4.3.1 ブートストラップ法

上の問題は、すべて真の分布 F を想定することなしに推定結果の品質を評価することに帰着する。そのために電子計算機の使用を前提としたモンテカルロ・シミュレーションの適用が考えられる[*50]。

Efron(1979) は、疑似乱数を使った経験分布 $\hat{F}$ からの復元抽出によって推定結果の品質を評価する手法を考案し、これを「ブートストラップ」と名付けた。また、この方法を推定の品質を把握するためにそれまで提案されてきた複数の方法 (1950 年代のジャックナイフ、1970 年代クロスバリデーションなど) と関連づけた[*51]。以下、ブートストラップについて、Efron and Tibshirani(1994)、汪・田栗ら (2003) をもとに解説する。

いま、分布 F にしたがう確率変数 x について、$g(x)$ という統計量の F のもとでの期待値を計算するものと考える。この計算を経験分布 $\hat{F}$ のもとでの期待値計算に置き換える (式 (4.6))[*52]。

$$\begin{aligned}
\int g(x)dF(x) &\simeq \int g(x)d\hat{F}(x) \\
&= \frac{1}{n}\sum_{i=1}^{n} g(x_i) \qquad (4.6)\\
&\quad \{x_1,\cdots,x_n\} \sim \hat{F}
\end{aligned}$$

経験分布 $\hat{F}$ が真の分布 F の良い近似になることは、次のように CLT を使って正当化することができる[*53]。$\hat{F}(x)$ は n 個のベルヌーイ変数 $\delta(X_i \leq x)$ の和である。ベルヌーイ変数の「成功」確率は、

$$Pr\{\delta(X_i \leq x)=1\} = Pr\{X_i \leq x\} = F(x)$$

したがって、

$$\begin{aligned}
n\hat{F}(x) &\sim Bin(n, F(x)) \\
&\downarrow CLT \\
\sqrt{n}(\hat{F}(x)-F(x)) &\sim N(0, F(x)(1-F(x))) \\
&\downarrow n \to \infty \\
&Pr\left\{|\hat{F}(x)-F(x)| \leq \epsilon\right\} \to 1
\end{aligned}$$

上のような真の分布の経験分布への置き換えを、現実の世界そのものに拡張的に適用することを考えてみる[*54]。これは、実は現実の世界 (存在論;Real World) と推論のプロセス (認識論;Bootstrap World) を重ね合わせて解釈することに等しい (図 4.5)。現実の世界では未知の確率分布 F から標本 x が抽出 (観測) される。この標本をもとに統計量 $g(x)$ が計算される。推論のプロセスにおいては経験分布 $\hat{F}$ からブートストラップ標本 x^* を抽出し、それにもとづいて統計量 $g(x^*)$ を計算する。

現実の世界と推論のプロセスを媒介するものが経験分布 $\hat{F}$ である。いま得ている観測値 x は真の分布 F から得られたものであること、そしてその性質を反映してい

されるべきばあいは予測誤差の推定などの領域だと言っている。汪・田栗ら (2003) は、吟味せずにバイアスの少ない推定量を選択することが危険である理由を次のように指摘している。i) パラメータ変換によって推定量の不偏性は容易に崩れる、ii) 不偏推定量が存在しないばあいにバイアス補正を行うことは分散の増大を招く。

*50 この手法は、真の分布が既知であるばあいでも積分が解析的に困難であるならば適用されることがある。計算機を用いた疑似乱数 (これは線形合同法のようなアルゴリズムにより実装される) から一様乱数、正規乱数、その他任意の確率分布にしたがう乱数を発生させることにより実現される。

*51 これら方法の総称は「リサンプリング」と呼ばれる。

*52 このように経験分布で置き換えた統計量は「プラグイン統計量」と呼ばれる。統計量 (各種のモーメント、順序統計などを含む) の分布全体をリサンプリングを通じて再現することで、検定論 (点の比較)、区間推定 (幅) を不要にする。

*53 経験分布からのリサンプリングをとくに「ノンパラメトリック・ブートストラップ法」と呼ぶことがある。とくに断らないかぎり、ブートストラップ法とはこれのことを指す。経験分布ではなく、特定のパラメトリック分布からリサンプリングを施すばあい、前者から区別して「パラメトリック・ブートストラップ法」と呼ばれる。

*54 あたかも、ミュンハウゼンが自分の靴紐をひっぱることによって自分そのものを持ち上げるように。

ることだけは分かっている。そしてこれら観測値の束を経験分布 $\hat{F}$ としてとらえ、これが真の分布 F を十分に近似しているとみなす。そこからのリサンプリング (復元抽出) の結果から推論を行う、と視点を転換する。

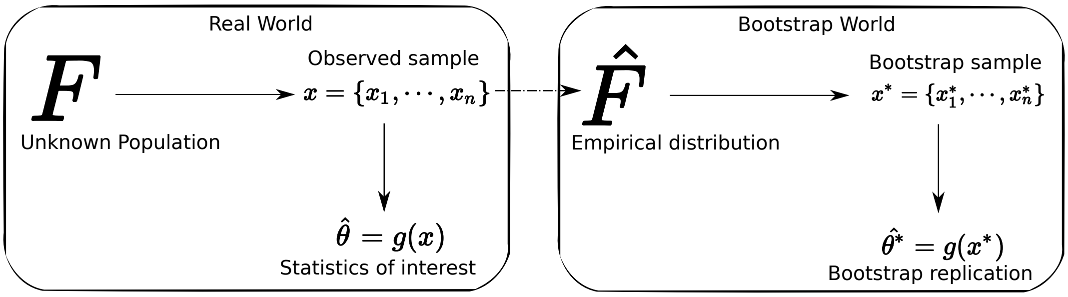


図 4.5　ブートストラップ法の模式図

出所：Efron and Tibshirani(1994)、p.87 より一部改変

ここで現実の世界と推論のプロセスの違いが、われわれにとってはむしろ有利にはたらく。現実の世界においては、理論上、無数のサンプリング (観測) が行われうると考え、またそのことを前提にして解析的な評価が行われるが、実際に行われるサンプリング (観測) はただの一回でしかない。そして、その結果として一組の観測値が得られている。ところが、推論のプロセスにおいては、経験分布からのリサンプリングをただの一回に限る必要はなく、計算機資源の許すかぎり何度でも実行することができる。そこでリサンプリングの回数を仮に B として、$B \to \infty$ ならば得られる $\hat{\theta^*}$ の分布は $\hat{\theta}$ についての (真の分布 F にもとづいて計算される) 分布に限りなく近づいていくことが期待される (さらにその上、困難な解析的計算も回避できる[*55])。

こうして得られる $\hat{\theta^*}$ の B 個の集まりは、$\hat{\theta}$ の品質 (平均、分散、偏り、信頼区間など) を評価する上で十分である。$\hat{\theta^*}$ は $\hat{\theta}$ の分布そのものの近似であるので、これをヒストグラム表示すれば全体の概観を把握でき、分散、偏り、パーセント点 (メディアン、信頼区間) などを具体的な数値として計算できる。さらに、それらの数値を用いて仮説を検定することもできる。これは真の分布が未知であることによって生ずる問題の大半を解決するものとなるだろう[*56]。

4.3.2　ブートストラップ法の具体例 (1)

表 4.3 に Efron and Tibshirani(1994) より最初の例題を示す。ある処置を施したマウスとそうでないものとで生存時間 (日数) に差があるかどうかを問うものとなっている。これによれば、処置群の平均生存日数は 86.86 日、対照群では 56.22 日となっており、その差は +30.63 日と前者の日数の方が長いように見受けられる。しかし、その精度を標準偏差として計測すると 28.93 日であり、その比はわずか 30.63/28.93 = 1.06 倍でしかない[*57]。ここで標準偏差は次の公式により求められている。

$$\widehat{se} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})}$$

問題はこの公式が、分散は異なるがいずれも正規分布にデータがしたがっている、という仮定から導き出されたものであり、その仮定が正しいかどうか、正しくないばあいにも同じ結論になるかどうかまではわからないということである。参考までに処置群と対照群を合わせて正規分布による Q-Q 確率プロットを描くと図 4.6 のようになるが、これを見ても二つの群が無差別に単一の正規分布にしたがっているのか、別々の分布にしたがっているのかは明瞭には見て取れない。

*55　Efron はメディアンの精度を計算する例でこの利点を強調している。
*56　ただし、そのように言えるのは、Real World において観測値が適切なサンプリングによって得られていることにかかっている。誤ったサンプリングや意図的な改ざんには無力である。
*57　ウェルチの検定に相当する作業をしている。R を用いて実際に検定してみると、検定統計量は 1.0587(自由度 9.6545)、p 値は 0.3155 であり、やはり二群に差があるとは明瞭には言えない。

表 4.3 マウスの生存時間

群	データ			標本数	平均	標準誤差
処置	94	197	16			
	38	99	141	7	86.86	25.24
	23					
対照	52	104	146			
	10	51	30	9	56.22	14.14
	40	27	46			
				差	30.63	28.93

出所：Efron and Tibshirani(1994)、p.11 より一部改変

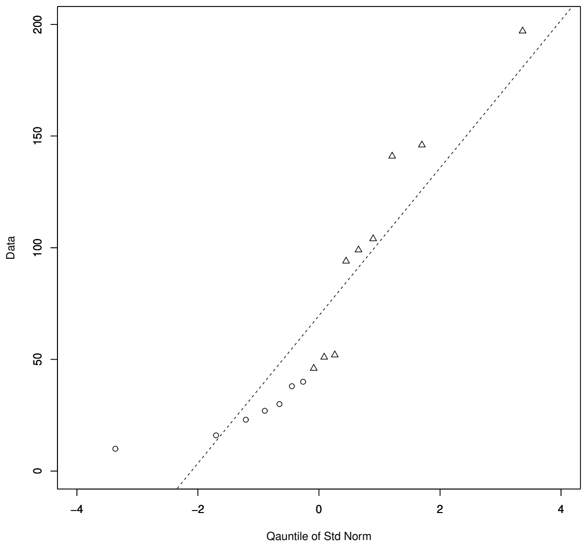


図 4.6 Q-Q 確率プロット (マウス生存時間)

次に AIC を用いたモデル選択の枠組みで同じデータを扱ってみる。具体的には坂元・石黒・北川 (1983) で紹介されている VARMOD*58を用いる。この計算の基礎となる対数尤度は式 (4.7) のとおりである。

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2 - \frac{n}{2} \tag{4.7}$$

表 4.4 VARMOD の結果

model	p	$\hat{\sigma}^2$	AIC
i_0	2	2804.734	176.431
i_1	3	2573.776	177.056

表 4.4 が VARMOD の結果である。モデル i_0 は「二群に差がない」という帰無仮説に対応、モデル i_1 が「二群に差がある」という対立仮説に対応する。これによればモデル i_0 の AIC の方が小さく、二群に差はないとするこれまでの結果と整合している。ただし、ここでは等分散が仮定されており、しかも残差が正規分布にしたがうこと、対数尤度の偏りが自由パラメータで近似できることが仮定されていることを見逃してはならない。これらの仮定が覆されたときにも同じ結論が言えるか、までは分からない。

ここからは真の分布についてパラメトリック・モデルを与えずに考察する。そのために、経験分布から B 個のブートストラップ標本をつくりだし、これに基づいて統計量の品質を評価する。

図 4.7 はブートストラップ平均を求めた結果 (それぞれの群ごとにいずれも $B = 2000$) である。対照群の平均分布にたいして処置群の分布は右側にずれており、分布全体の形状としては単一の正規分布からは外れているように見える。

ブートストラップ平均について表 4.3 と同様の計算をしてみると次のようになる。まず処置群の平均 86.96 日にたいして対照群の平均は 55.95 日、その差は +31.01 日である。他方、標準偏差は処置群で 23.37 日、対照群で 12.75 日、差の標準偏差は 26.62 日となる。比は $31.01/26.62 = 1.16$ 倍であり、多少は前より大きいがいずれにせよ大差ない。つまり正規分布の仮定がなくとも結論には大きな違いがないことになる。

*58 VARMOD はデータ全変動から各因子ごとの平均を差し引いた残差が正規分布にしたがうという仮定のもとに、最尤推定量と自由パラメータ数を求め、これから AIC を計算するというものである。比較対象のモデルのうち最も AIC の低いものが最適とされる。これは CG 分布を用いた混合グラフィカルモデルの一種 (交絡因子のない) とみなせる。付録に `R` によるプログラムを付す。

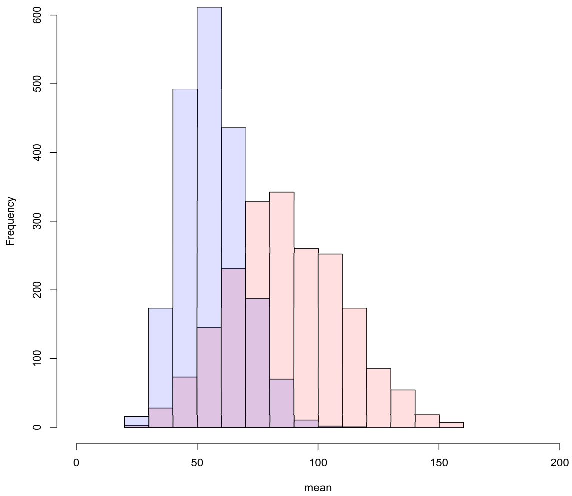


図 4.7　ブートストラップ平均 (マウス生存時間)

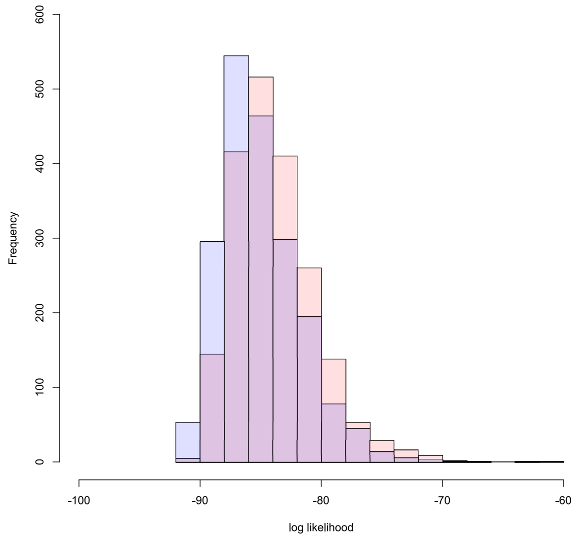


図 4.8　i_1、i_0 の対数尤度 (マウス生存時間)

同じく VARMOD でのモデル i_1 と i_0 に対応するブートストラップ対数尤度 $\ln L$[*59]を求め、そのヒストグラムを描くと図 4.8 のようになる。それぞれの分布は左右非対称であり、右側にスソが広がっている。またモデル i_1 の分布が i_0 のそれよりも全体としてわずかに右寄りに位置しているようにも見える。ただし両者はほとんど重なっており違いは目立たない。

表 4.5　対数尤度の比較と偏り

	i_1	i_0	$i_1 - i_0$
$E[\ln L]$	-83.82	-85.03	$+1.21$
$\ln L_0$	-85.53	-86.22	$+0.69$
Bias	$+1.710$	$+1.181$	$+0.529$

対数尤度 $\ln L$ の偏りについては次のように計算できる[*60]。モデル i_1 についてブートストラップ対数尤度の平均が -83.82 であるのにたいして元の標本にかんする対数尤度は -85.53 であり、その差 $+1.710$ が偏りである。同様にモデル i_0 について計算した結果と合わせると表 4.5 のようになる。$E[\ln L]$ で比較してみると、i_1 よりも i_0 の方が大きく、表 4.4 の結果は、過剰なバイアス補正によって順序が逆転したために生じたとも考えられる。いずれにせよ、二群に違いがあるかどうかを判定するためにはデータが不足していると考えるのが妥当であろう。

4.3.3　ブートストラップ法の具体例 (2)

次に量的ではなく、質的データに対してブートストラップ法を適用した例を示す。

表 4.6　国民性調査 (統計数理研究所)

X_1/X_2	1. 男に	2. 女に	計
1. 男	749	83	832
2. 女	445	636	1081
計	1194	719	1913

出所:坂元ら (1983)、p.92 より

表 4.6 は、坂元ら (1983) より、分割表分析の例題を示したものである。X_1 は被験者の性別、X_2 は「生まれ変わったとしたらどちらになりたいか」の回答を示している。

*59　処置群と対照群合わせて 16 標本より B=2000 個のブートストラップ標本を構成し、それぞれについて VARMOD を適用した。
*60　標本対数尤度の偏りをブートストラップ法で評価する手法は EIC と呼ばれている。詳しくは小西・北川 (2004) を参照のこと。

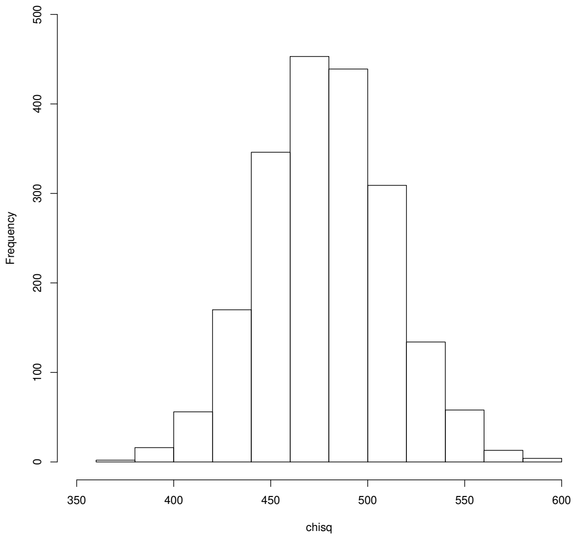


図 4.9　ブートストラップ χ^2(分割表分析)

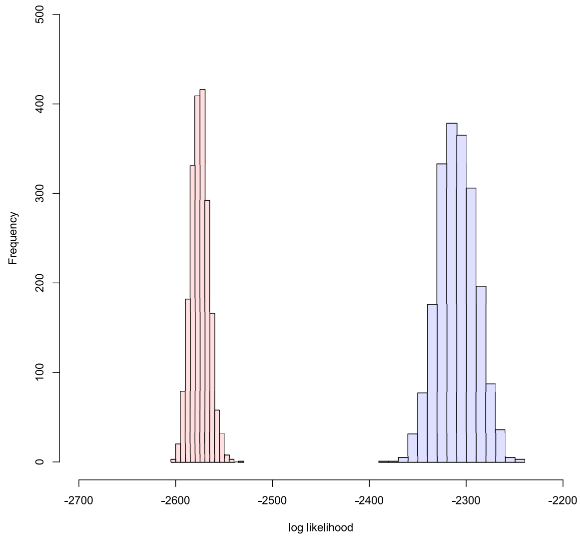


図 4.10　i_1、i_0 の対数尤度 (分割表分析)

提示されている問題は「X_1 と X_2 が独立であるかどうか」であり、「両者は独立であるとは言えない」という結論が、χ^2 検定[*61]によっても AIC による方法でも既に得られている。ブートストラップ法はこの結論の品質を明らかにする。

図 4.9 は、ブートストラップ χ^2 のヒストグラムを示したものである ($B = 2000$)。分布のレンジは 350〜800 程度と広いが、その自由度 1 に比べると十分に大きい。

表 4.7　対数尤度の比較と偏り

	i_1	i_0	$i_1 - i_0$
$E[\ln L]$	-2309.72	-2575.15	$+265.43$
$\ln L_0$	-2312.10	-2576.14	$+264.04$
Bias	$+2.383$	$+0.989$	$+1.394$

図 4.10 は、X_1 と X_2 が独立であるとする仮説 i_1 (右側) と、独立でないとする仮説 i_0 (左側) (いずれも多項分布を基礎とする) に対応するそれぞれの対数尤度のヒストグラムを重ねて示したものである。両者は互いに十分に離れており、i_1 の対数尤度の方が i_0 のそれよりも確実に大きいと判断される。表 4.7 はそれぞれの対数尤度のバイアスを示したものである。i_1 についていえば、バイアスは 2.383 であり、自由パラメーター数 3 より少ない。i_0 についても、そのバイアス 0.989 は自由パラメーター数 2 よりも小さい。しかしバイアス補正の有無にかかわらず、$i_1 > i_0$ の結論は揺るがないとみてよい。

4.3.4　ブートストラップ法の具体例 (3)

赤池のベイズ的手法 (EM アルゴリズム) の事例 (図 4.3) にブートストラップ法を適用してみる。この例での対数尤度は潜在変数を含むものに拡張されたものであった。これのバイアスを評価する。

図 4.11 はブートストラップ対数尤度のヒストグラムである (B=2000[*62])。レンジは 250〜450 の間に広がっており、右側ではなく左側にスソが広がっている。標本の対数尤度が 407.624 であるのにたいして、ブートストラップ対数尤度の平均は 395.934 であり、後者と前者の差は -11.690 である。AIC との類推で考えるとあたかも自由パラメータ数が負の値をとっているかのような

[*61] 検定結果は、$\chi^2 = 476.34$(自由度 1)、p 値は $< 2.2 \times 10^{-16}$ で、きわめて小さい有意水準で帰無仮説 (独立) は棄却される。R の `chisq.test` 関数を用いた。

[*62] ただし、一部収束計算に失敗している。成功した回数は 1876 回である。

奇妙な状態となっている。対数尤度の大きさからみれば、その正確度は (潜在変数を考慮に入れたにもかかわらず、また変分ベイズ法という近似計算を用いたにもかかわらず) 極めて高い。

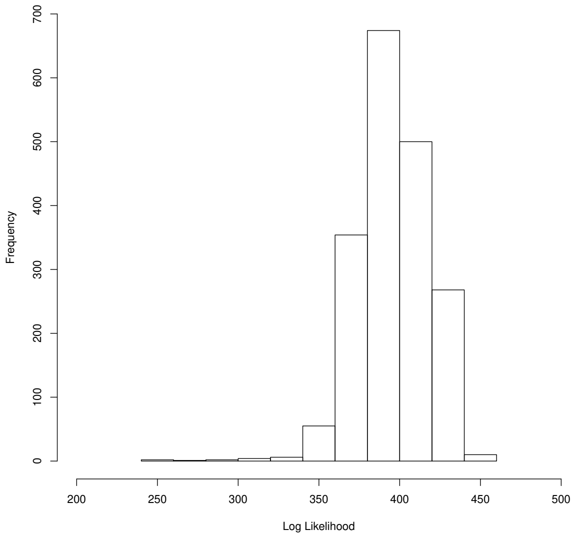


図 4.11　対数尤度 (EM アルゴリズム)

4.3.5　ブートストラップ法と尤度

ブートストラップ法はデータのサンプリングとリサンプリングに基礎を持ち、その意味では確率の頻度説 (後述) と親和性が高く、確率の主観説 (ベイジアン) とは相いれないように表面的には見える。たとえばビショップ (2001) はリサンプリングの有用性を認めたうえで、これを近似計算の一種とみなすことによってその見かけの断絶を調停しようとしている。ビショップはプラグマティックな立場を強調することで、哲学的な議論の落とし穴にはまることを警戒している。しかし、われわれは赤池の解釈に倣ってベイズ的方法を KL 情報量と尤度から引き出しているので、そこにそれほど断絶を見出さずに済んでいる。

Efron 自身はブートストラップ法をノンパラメトリック最尤法と解釈しているらしいことが、汪・田栗ら (2003) によって語られている。これは、われわれが上のいくつかの事例でブートストラップ対数尤度のヒストグラムをさんざんに計算していることとは別のことである。われわれは標本対数尤度をたんにプラグイン統計量とみなしてブートストラップ法を適用したまでであって、ブートストラップ法が最尤法の一種であるとまではまだ考えていない。そこで、あらためてブートストラップ法と尤度との関係について振り返ってみよう。

まず Efron and Tibshirani(1994) は、経験分布が経験尤度の最尤推定量として得られることを指摘している[*63]。この証明のためにはまず経験尤度とは何か、が説明されねばならない。Owen(2001) によれば、経験尤度 $L(F)$ とは、F を cdf とする分布から観測値 $\{x_1, \cdots, x_n\}$ が得られる確率のことである。そして、F が cdf であるとは、

$$F(x) = Pr(X \leq x)$$

であるから、ここから等号を外したものを考えると、

$$F(x-) = Pr(X < x)$$

より、

$$Pr(X = x_i) = F(x_i) - F(x_i-)$$

したがって、

$$\begin{aligned} L(F) &= \prod_{i=1}^{n} Pr(X = x_i) \\ &= \prod_{i=1}^{n} \{F(x_i) - F(x_i-)\} = \prod_{i=1}^{n} F(\{x_i\}) \end{aligned}$$

経験分布が経験尤度の最尤推定量であるという事実はたとえば次のように証明できる。観測値を $X = \{x_1, \cdots, x_n\}$ とする。この観測値に重複 (tie) が含まれている可能性があるので、それを取り除いたものを $\{z_1, \cdots, z_m\}$ とする。また z_j の X のなかでの出現頻度を n_j とする。z_j の出現確率を p_j とするとき、経験尤度 $L(F)$ は次のようになる。

$$L(F) = \prod_{i=1}^{n} F(\{x_i\}) = \prod_{j=1}^{m} p_j^{n_j}$$

*63　ただし、Efron and Tibshirani はこの事実の証明を与えず、練習問題として記している。Owen(2001) によれば、この定理の初出は Kiefer and Wolfowitz(1956) とのことである。

これを最大化するために次のようなラグランジェ関数を定義する。

$$G = \sum_{j=1}^{m} n_j \ln p_j - \lambda \left(\sum_{j=1}^{m} p_j - 1 \right)$$

これより、

$$\frac{\partial G}{\partial p_j} = \frac{n_j}{p_j} - \lambda = 0$$

これより、

$$\lambda = \sum_{j=1}^{m} n_j = n \Longrightarrow \widehat{p_j} = \frac{n_j}{n}$$

これは経験分布である[*64]。だから、ノンパラメトリック・ブートストラップ法とはパラメトリック・ブートストラップ法の一種であって、そこで用いられているパラメトリック分布が経験分布であるにすぎない、と言える。

4.3.6 経験尤度法による推論

ノンパラメトリック統計における尤度の取り扱い方法は多様であり、ブートストラップ法はそのなかの一つにすぎない[*65]。ノンパラメトリック尤度の取り扱いが定まらないのは、計算効率の問題ばかりでなく、その背後に確率にかんする思想の違いが潜んでいるためと思われる。

ここでは Efron らがブートストラップ法を最尤推定と結びつける根拠をなした経験尤度とそこから発展した推論方法を、主に Owen(2001) に依拠してまとめておく。

経験尤度比 $R(F)$ とは、最尤推定量である経験分布 F_n の尤度 $L(F_n)$ を基準として、検定したい仮説を制約として含む経験尤度 $L(F)$ を測ったものであり、以下のように表現される。

$$R(F) = \frac{L(F)}{L(F_n)} = \begin{cases} \displaystyle\prod_{j=1}^{m} \left(\frac{np_j}{n_j} \right)^{n_j} & tie\ \text{あり} \\ \displaystyle\prod_{i=1}^{n} np_i & tie\ \text{なし} \end{cases}$$

経験尤度比は、尤度比検定からのアナロジーにもとづいて発想された。パラメトリックモデルにおいて $H_0 : \theta = \theta_0$ という仮説 (制約) を検定するための尤度比 λ は、以下のように計算される。

$$\lambda(x) = \frac{L(\theta_0 \mid x)}{L(\theta \mid x)}$$

そのうえで、以下のように λ がある定数 c を超えるかどうかにより、仮説の棄却／採択の判断をくだすことが尤度比検定であった。ここで定数 c は有意水準 α を用いて、$Pr(\lambda \leq c \mid H_0) = \alpha$ から前もって計算される。

$$\begin{cases} \lambda > c & \to H_0\text{を棄却しない} \\ \lambda \leq c & \to H_0\text{を棄却} \end{cases}$$

また統計量 θ_0 の信頼区間は、$\{\theta \mid \lambda \geq c\}$ から計算される。このように、検定と信頼区間の推定は表裏一体のものである。

経験尤度比のばあいは、次のようになる。統計量 $\theta = g(F)$ の信頼区間は、

$$\begin{aligned} &\{g(F) \mid R(F) \geq r\} \\ &= \{\theta \mid \mathcal{R}(\theta) \geq r\} \end{aligned} \tag{4.8}$$

ここで、$\mathcal{R}(\theta)$ はプロファイル尤度と呼ばれ、次のように定義されている。

$$\mathcal{R}(\theta) = \sup \{R(F) \mid g(F) = \theta\} \tag{4.9}$$

4.3.7 経験尤度法の具体例 (1)

一変量の平均値にかんする検定と信頼区間の推定の例を挙げる。

プロファイル尤度は式 (4.9) より次のようになる。ここで p_i の代わりに w_i としているのは、これがウェイトであることを強調するためである。

$$\mathcal{R}(\mu_0) = \max \left\{ \prod_{i=1}^{n} nw_i \,\middle|\, \sum_{i=1}^{n} w_i x_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}$$

[*64] ここで Laplace の確率の定義を想起するのも興味深い。

[*65] Efron and Tibshirani は以下の互いに競合する諸方法をとりあげて比較している。経験尤度 (empirical likelihood : Owen,1988)、近似ピボット法 (approximate pivot method: Boos and Monahan,1986)、ブートストラップ部分尤度 (bootstrap partial likelihood: Davison, Hinkley and Warton,1992)、インプライド尤度 (implied likelihood: Efron,1992)。

また信頼区間はプロファイル尤度を用いて次のようになる。

$$
\begin{aligned}
&\{\mu \mid \mathcal{R}(\mu) \geq r_0\} \\
&= \left\{ \sum_{i=1}^{n} w_i x_i \,\middle|\, \prod_{i=1}^{n} n w_i \geq r_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}
\end{aligned}
$$

これらは線型計画法の定式化に見かけは似ているが、総積記号が含まれており計算上の困難を予想させる。これらの計算は、$\mu = \mu_0$ の検定のばあいは $\mathcal{R}(\mu_0)$ の計算に、r_0 にたいする信頼区間を求めるばあいは $\mathcal{R}(\mu) = r_0$ の求解に帰着される。

$\mathcal{R}(\mu)$ の計算は次のように行う。まず $x_{(1)} < \mu < x_{(n)}$ を仮定する[*66]。ラグランジェ関数を次のように定義する。

$$
G = \sum_{i=1}^{n} \ln n w_i - n\lambda \sum_{i=1}^{n} w_i (x_i - \mu) + \gamma \left(\sum_{i=1}^{n} w_i - 1 \right)
$$

これより、

$$
\frac{\partial G}{\partial w_i} = \frac{1}{w_i} - n\lambda (x_i - \mu) + \gamma = 0
$$

$$
\begin{aligned}
&\to \sum_{i=1}^{n} w_i \frac{\partial G}{\partial w_i} = n + \gamma = 0 \to \gamma = -n \\
&\to w_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda (x_i - \mu)}
\end{aligned}
$$

ここで λ は次の方程式の解である。

$$
\sum_{i=1}^{n} \frac{x_i - \mu}{1 + \lambda (x_i - \mu)} = 0
$$

この方程式を解くにはブレントの方法、ニュートン法などの数値解法を用いれば良い。その際に、解を挟み込む二点が必要になるが、たとえば $w_{(1)} = 1$、$w_{(n)} = 1$ などの特殊なウェイトを用いることにより、

$$
\frac{1 - n^{-1}}{\mu - x_{(n)}} < \lambda(\mu) < \frac{1 - n^{-1}}{\mu - x_{(1)}}
$$

などを初期値とすればよい。

μ の信頼区間を求めるには、その両端を μ_-、μ_+ として、

$$
x_{(1)} \leq \mu_- \leq \hat{x} \leq \mu_+ \leq x_{(n)}
$$

が成立することを利用して、それぞれの範囲で $\mathcal{R}(\mu) = r_0$ の解を求めればよい[*67]。

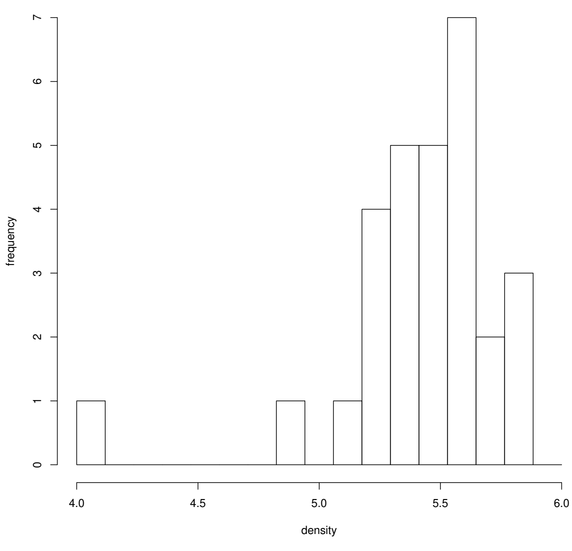


図 4.12　Cavendish(1798) の実験からの計測値

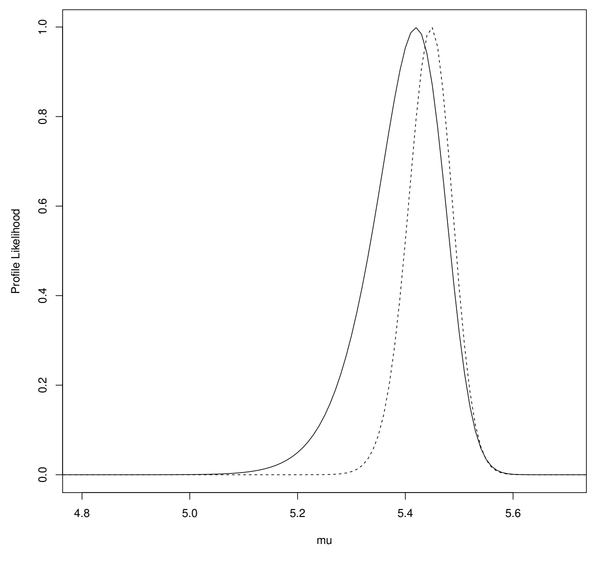


図 4.13　Cavendish(1798) 実験のプロファイル尤度

*66 これ以外は自明である。$\mu \leq x_{(1)} < x_{(n)}$ または $x_{(1)} < x_{(n)} \leq \mu$ のとき $\mathcal{R}(\mu) = 0$、$\mu = x_{(1)} = x_{(n)}$ のとき $\mathcal{R}(\mu) = 1$ である。

*67 あるいは、直接的に信頼区間についての最適化問題を解く、近似的に $\mathcal{R}(\mu) \geq \exp(-\chi^{2,1-\alpha}_{(1)}/2)$ を計算する、などの方法もある。

ここでは具体例として、Owen(2001) の p.17 にならって H.Cavendish(1798)[*68]の実験からの計測値よりプロファイル尤度を求める。ただし図 4.12 に示す 29 個のデータには疑わしい点があり、この取り扱いには注意[*69]を要する。

図 4.13 は図 4.12 のデータにもとづいて計算されたプロファイル尤度を示している[*70]。実線が Owen(2001) と同じデータ、点線が `HistData` の `Cavendish` によるものである。両者の違いは、前者に含まれる 4.02 のサンプルが後者では 5.02 になっている点のみである。前者のモードは $\hat{\mu} = 5.42$、後者は $\hat{\mu} = 5.45$ とその差はわずかであるが、分布の歪み方は大きく異なる。

4.3.8 経験尤度法の具体例 (2)

次に分散にかんする経験尤度計測の例をとりあげる。図 4.14 は日経平均株価の終値の推移を示したものである[*71]。このデータを x_i とし、$y_i = \ln(x_i/x_{i-1})$ が一日あたり収益率を示す。年間の取引日数を n として、

$$\sqrt{n\mathrm{Var}[y_i]}$$

が年あたりのボラティリティを示す。上のデータのばあい、243 日につき 0.1459087(14.59%) であった。収益率の分散がいかなる分布にしたがうのかを明らかにすることがここでの問題である。Owen(2001)、p.38 の方法にならい y_i は形式的に互いに独立であるとして計算してみる。

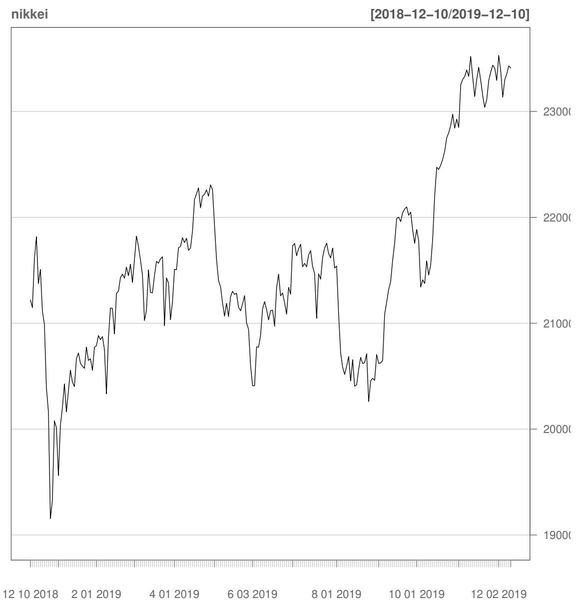


図 4.14 日経平均株価の推移 (2018/12/10〜2019/12/10)

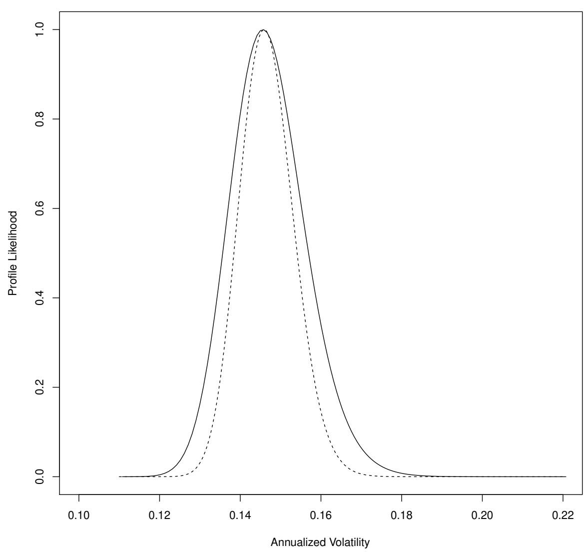


図 4.15 日経平均株価のボラティリティ

[*68] Henry Cavendish(1731-1810) はイギリスの化学者・物理学者、ロンドン王立協会の会員。水素、アルゴンの発見者であり、地球の比重を精密に測定した。多数の熱力学、電気学についての実験を行ったが、その多くは生前には公表されなかった。

[*69] 図 4.12 を見ればわかるように、ここには 4.02、4.88 といった「異常値」が含まれる。Owen はこれらについて何も注意を与えていない。Cavendish による原文、"Experiments to Determine the Density of Earth"(Philosophical Transactions of the Royal Society of London,88:469-526) の中には、4.88 は含まれるが、4.02 は存在せずその代わりに 5.02 がある。Owen は図 4.12 のもとになるデータを Stigler,S.M. "Do robust estimators work with real data?"(Annals of Statistics,5:1055-1098) から採ったとしている。しかし、`R` の `HistData` に含まれる `Cavendish` も同じ Stigler を使っているが、ここには 4.02 は含まれていない。ネットには 4.02 を含む Cavendish のデータが散見され、その多くは統計学のテキストからのものである。

Cavendish の用いた実験器具 (ねじり天秤装置) の特徴から考えて 4.02 を正常な値とみなすことは危険であり、われわれは Owen の用いたデータとは別に `HistData` からのデータも使ってプロファイル尤度を求めてみよう。

[*70] `R` の `emplik` パッケージ (Mai Zhou による) の関数 `el.test` を用いた。

[*71] `R` のパッケージ `quantmod` の `getSymbols` 関数を用いて、Yahoo よりデータを取得した。

$$\mathcal{R}(\sigma_0^2) = \max \left\{ \prod_{i=1}^n n w_i \middle| \sum_{i=1}^n w_i \sigma^2(F) = \sigma_0^2, w_i \geq 0, \sum_{i=1}^n w_i = 1 \right\}$$

ここで Owen は $\sigma^2(F)$ が平均値にかんする滑らかな関数であることに着目し、これを一次式で近似することを提案している。その方針にしたがって計算してみると、

$$\begin{aligned}
&\sum_{i=1}^n w_i \sigma^2(F) \\
&= \sum_{i=1}^n w_i y_i^2 - \left(\sum_{i=1}^n w_i y_i \right)^2 \\
&= v - \mu^2 \\
&\simeq (v_0 - \mu_0^2) + (v - v_0) - 2\mu_0(\mu - \mu_0) \\
&= \sum_{i=1}^n w_i (y_i^2 - 2\mu_0 y_i) + \mu_0^2 \\
&= \sigma_0^2 \\
&\downarrow \\
&\sum_{i=1}^n w_i (y_i^2 - 2\mu_0 y_i) \simeq \sigma_0^2 - \mu_0^2
\end{aligned}$$

そこで、さきに $\mathcal{R}(\mu)$ のモードから μ_0 を計算しておき、これをつかって上の制約式を表現すれば、平均値にかんするプロファイル尤度の計算方法がそのまま使えることになる。

他方で、伝統的な検定理論によれば、標本分散 s^2、母分散 σ^2 のとき、

$$\frac{ns^2}{\sigma^2} \sim \chi^2_{(n)}$$

が近似的に成立するので、これからも尤度比を計算できる。

図 4.15 は年あたりのボラティリティについて経験尤度比 (実線)、正規分布による近似尤度比 (点線) を示したものである。これによればあきらかに前者のほうがスソが広く、ボラティリティの分布が正規分布からははずれていることを示唆している。

4.3.9　経験尤度法の具体例 (3)

最後にマウス生存時間の例を再び取り上げ、ブートストラップ法の結果と比較してみよう。

Ying(2010) は Owen(2001) の記述にヒントを得て二群の差の検定のための経験尤度比を次のように定式化している。ここで w_a は処置群についての、w_b は対照群についてのウェイトである。両群に共通の基準値を μ_0 として、処置群の平均はそれから d だけ大きいとしている。

$$\begin{aligned}
\max_{w_a, w_b} \mathcal{L}(\mu_0, d) &= \prod_{i=1}^{n_a} (n_a w_{ai}) \cdot \prod_{j=1}^{n_b} (n_b w_{bj}) \\
s.t. \sum_{i=1}^{n_a} w_{ai} = 1, &\sum_{i=1}^{n_a} w_{ai}(y_{ai} - \mu_0 - d) = 0 \\
\sum_{j=1}^{n_b} w_{bj} = 1, &\sum_{j=1}^{n_b} w_{bj}(y_{bj} - \mu_0) = 0
\end{aligned}$$

この最適化問題の解については、w_a と w_b とで完全に切り分けて考えられることに注意する。たとえば w_a の解は次のようになる (w_b についても同様)。つまり、これまでの平均値にかんするプロファイル尤度の計算方法がそのまま適用できることを意味する。

$$\begin{cases}
w_{ai} = \dfrac{1}{n_a} \cdot \dfrac{1}{1 + \lambda_a (y_{ai} - \mu_0 - d)} \\[2ex]
\displaystyle\sum_{i=1}^{n_a} \frac{y_{ai} - \mu_0 - d}{1 + \lambda_a (y_{ai} - \mu_0 - d)} = 0
\end{cases}$$

図 4.16 はマウスの生存時間の例について、μ_0 と d を与えたときのプロファイル尤度を表示したものである。横軸が d、奥行きが μ_0 である。分布は単峰であり、d はぎりぎり負ではないだろうという程度の広がりをもっている。図 4.17 は同じデータを等高線で示している。最尤推定量は $(\hat{\mu}_0, \hat{d}) = (56, 31)$ であり、このときの尤度比は 0.9998 である。$d = 0$ の制約を入れたときの $\hat{\mu}_0$ は 87 であり、このときの尤度は 0.0307 である。両者の対数尤度の差の -2 倍 (AIC に相当) は -6.9647 である (つまり $\mu_0 = 0$ の制約を含まないモデルの方が望ましい)。

図 4.17 で興味深い点は、μ_0 と d が完全に独立しているというわけではなく、μ_0 が高ければそれに応じて d が低くなることである。μ_0 が全体として外れたところにあれば、d はそれを是正するように変化し、純粋に両群の差を表示するものではなくなると解釈されよう。

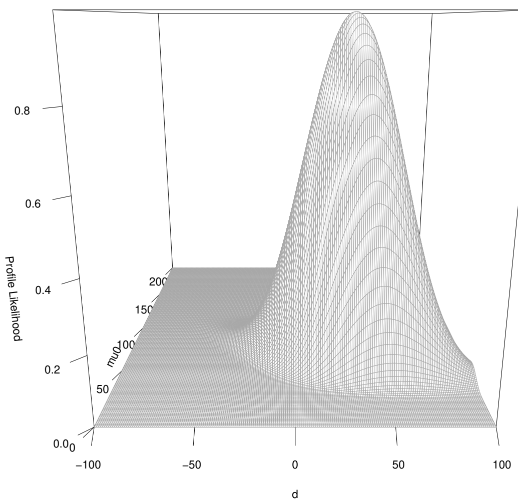


図 4.16　二群の経験尤度比 $\mathcal{L}(\mu_0, d)$

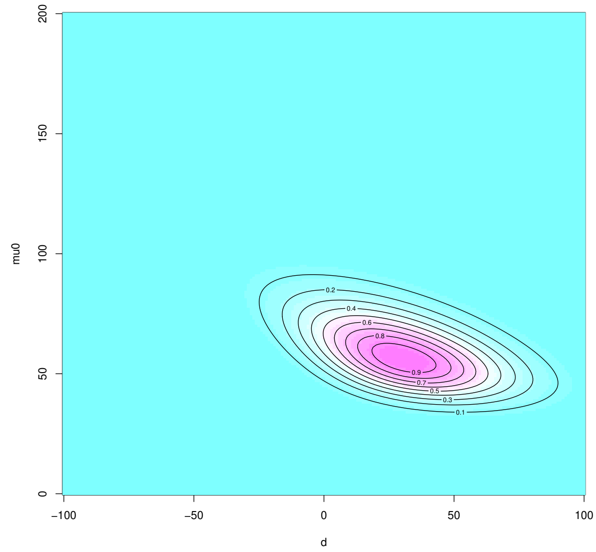


図 4.17　図 4.16 の等高線図

4.3.10　経験尤度法とブートストラップ法

経験尤度法とブートストラップ法の関係は奇妙である。両者ともに、真の分布がはっきりとパラメトリック・モデルとして記述できない領域で効力を発揮する。また両者ともに、推定量の品質 (ばらつき、偏り、信頼区間、検定) を評価することに用いられる。このように共通の目標をもつ以上、互いに競合する手法のようにも思われる。しかし、Owen は以下のように経験分布法とブートストラップ法の併用がそれ単独で使われるよりも望ましい結果をもたらす例を挙げている。

- ブートストラップ法は経験分布からのリサンプリングにもとづくが、データ以外にも追加的な情報が得られているばあいには、これより良い分布を使うことができる。具体的には、追加情報を用いて経験尤度比を構成し、そのときのウェイト w_i を使ってより洗練された分布 $\sum_{i=1}^{n} w_i \delta(x - x_i)$ からのリサンプリングを試みることができる。
- 経験尤度比の計算においてブートストラップ法を援用することも考えられる。Owen はベクトル変数についての平均の信頼区間を数値計算で求める代わりに、B 個のブートストラップ標本を用いて $B(1-\alpha)$ 番目の数値を信頼区間の推定値としている。この方法は χ^2 分布や F 分布による近似、さらに χ^2 分布のバートレット補正による近似よりも望ましい結果をもたらす、と結論している。

4.4　小括

本節の主要な結論は以下のとおりである。

- ベイズ・ルールは、存在論的に理解されるかぎりでは条件付き確率の定義から代数的に引き出されるものであり、本来はそこになんら新しい意味を付加するものではない。しかし、このルールが〈情報〉=〈尤度〉というものに結び付けられるや否や、統計的推論の新しい方法として受け止められ、その後の著しい統計分析技術の展開に寄与することになった。
- 標本平均対数尤度 $L_{r;1}$ にはバイアスが存在する。赤池弘次は巧妙な計算によりこのバイアスを計算し、このずれが概ね〈自由パラメータ数〉に等しいことを見出した。これをあらかじめ $L_{r;1}$ から取り除いておくことにより、最尤法をモデル選択の方法論にまで拡張することができる。これが赤池情報量規準 (AIC) である。

- 赤池情報量規準は推測統計学の伝統にたいする明確な批判を志向し、尤度すなわち情報量による統計学の書き直しを要請するものである。赤池の功績は、尤度の利用にあたってかならずしも「真の構造」を知らなくともよいということを明確にしたことである。
- AIC は「オッカムのかみそり」の具体化ではない。赤池が重視するのは絞り込みの過程ではなくむしろそれ以前の多数のモデル候補の提案、モデル・ビルディングである。これを赤池は C.S.Peirce の「アブダクション」の過程として描写している。
- AIC はまず制御理論の領域で受容され、その後統計学分野に受け入れられることになった。AIC は、伝統的統計学にたいする批判に根拠をもつにも関わらず多くの統計分析ソフトは AIC を出力するようになった。これは、AIC が最尤法の拡張を志向していたこと、そして尤度の利用がパラメータ推定において有力な方法論として広く採用されたこと、また最適な推定量を探索するための尤度計算に必要な繰り返し計算のコストがコンピュータの利用によって劇的に低下したためである。
- 存在論と認識論を隔てるものは、観測の限界という問題である。存在論が、それまでに認識のプロセスで得た内容を本質的な客観的法則性が実体を通じて現象する過程として記述することであるのにたいして、認識論では、実践の場にあり様々の物質的制約に囚われているわれわれ自身が、われわれの外なる物をいかにして認識できるかということが問われる。そして認識論においては、現実的な観測の限界ということが常に念頭に置かれなければならない。観測の限界には様々の状況が含まれる。たとえば、不完全な観測値、標本数の不足、交絡因子や潜在変数の存在、などの問題がある。
- 坂元ら (1983) は AIC の利用上の注意の第一として、自由パラメータ数が標本数に比して多すぎないことの必要を挙げている。しかしこの要求は、往々にして満たされない。とりわけ時系列解析、因子分析などでこの状況は顕著である。この状況にたいして、パラメータそれ自体を定数ではなく潜在的確率変数とみなす、という視点の転換が可能である。このばあい、その変数 θ のしたがうべき確率分布 $h(\theta \mid \cdot)$ と (ハイパー) パラメータ ω が想定され、尤度は観測値のみならず潜在変数としてのパラメータをも包含するものに拡張される。この計算は、Dempster,Laird and Rubin(1977) の EM アルゴリズムによって実行できる。
- 赤池 (1989) はベイズ統計学について、ベイズ・モデリングという方法そのもの (「ベイズ的方法」) と、その考えを教条的に支持する立場 (いわゆる「主観確率の立場」) を峻別しつつ、後者への批判 (とりわけその事前分布の考え方についての批判) を行っている。赤池の主張は当然のことながら伝統的な統計学の立場とは一線を画しており、むしろ「主観確率の立場」が未知パラメータの「客観性」にこだわっていることの矛盾を指摘している。赤池の主張の特異性は、ベイズ的方法の有用性を認めつつ、その正当性 (「良さ」) が自明でないことを同時に指摘している点にある。
- 赤池は、事前分布の選択をめぐるこれまでの統計学三潮流による論争について、「これらは実際の統計的データの処理に対して広く有効な方法を与えるものとはならなかった」と断じている。
  1). R.A.Fisher : fiducial probability の理論は、事前分布の選択を客観的に行おうとする試み
  2). Neyman=Pearson : 第一種の誤り、第二種の誤りの概念の導入は、事前分布の導入を避ける理論の展開を試みたもの
  3). 主観確率の立場 (de Finetti 等): 「事前分布はとにかく存在しているのであって、各人がその持っている情報を十分に活用すれば決定できるものである」と主張

  これらと対比するかたちで、赤池はベイズ的方法について自身の積極的な見解を示す。すなわち、パラメータを単独に実在するものとして解釈することなく、事前分布の想定とともにデータからの情報抽出の手段とみなすことがベイズ的方法である、という主張である。
- 赤池はベイジアン (主観確率の立場) とは異なり、方法の「客観性」をデータ処理の有効性を万人が

認めうるかどうかということに求めている。さらに赤池は方法の「有効性」を従来の方法にたいする相対的な優位性と解し、それが必要とされる場面 (逆に言えば、従来の方法がそこで限界にぶつかる場面) を特定していく。赤池は従来の方法とベイズ的方法を連続的なつながりのあるものとして一体的に把握しており、両者をつなぐものがAIC の基礎をなす KL 情報量の考え方であるとする。ベイズ的方法は従来の方法の拡張とみなされる。つまり、従来の方法での θ_0 とはベイズ的方法での θ が退化したものにすぎない。

- なぜベイズ統計学が 1980 年代から隆盛を誇るようになったのか。コンピュータ技術の飛躍的な発展を通じて最尤法を少ない計算コストで実行できるようになったことが前提条件としてある。ついで、時系列分析など、データ量に比して推定すべきパラメータ数が多すぎる問題を解くニーズが増えたことが挙げられる。こうしたニーズに対応するためには、従来の方法をベイズ的方法に拡張しなければならなかった。
- G.Schwarz はベイズ・モデリングに対応して、「AIC に代わるモデル選択規準」として BIC を提唱した。BIC とは、ラプラス近似を使ってベイズ・モデリングにおいて拡張された尤度関数を近似したものであり、AIC のように期待対数尤度の偏りを補正するものではない。ベイズ・モデリングが不要の状況で BIC を使うこと、あるいは期待対数尤度の近似法としてもっと良い精度の近似法が使えるときにあえて精度の低いラプラス近似を適用すること、などは BIC の誤用である。
- 推定作業において真の分布がわからないということは、ただちに観測値の信頼性の問題を提起する。まず、データを生み出す確率分布についてパラメータばかりではなく分布の形状さえ分からないので、観測データの中に〈外れ値〉が存在していたとしても、どの観測値がそれであるかを名指しすることができない。真の分布がわからないということは、推定結果の品質、すなわち正確度 (偏り) や精度 (ばらつき、分散) についてもやはり分からない、ということをも意味する。
- 上の問題は、すべて真の分布 F を想定することなしに推定結果の品質を評価することに帰着する。そのために電子計算機の使用を前提としたモンテカルロ・シミュレーションの適用が考えられる。Efron(1979) は、疑似乱数を使った経験分布 $\hat{F}$ からの復元抽出によって推定結果の品質を評価する手法を考案し、これを「ブートストラップ」と名付けた。また、この方法を推定の品質を把握するためにそれまで提案されてきた複数の方法と関連づけた。
- 未知の分布 F にしたがう確率変数 x について、$g(x)$ という統計量の F のもとでの期待値を計算するものと考える。この計算を経験分布 $\hat{F}$ のもとでの期待値計算に置き換える。このような真の分布の経験分布への置き換えを現実の世界そのものに拡張的に適用することを考えてみる。これは、現実の世界 (存在論;Real World) と推論のプロセス (認識論;Bootstrap World) を重ね合わせて解釈することに等しい。現実の世界では未知の確率分布 F から標本 x が抽出 (観測) される。この標本をもとに統計量 $g(x)$ が計算される。推論のプロセスにおいては経験分布 $\hat{F}$ からブートストラップ標本 x^* を抽出し、それにもとづいて統計量 $g(x^*)$ を計算する。
- 現実の世界と推論のプロセスを媒介するものが経験分布 $\hat{F}$ である。いま得ている観測値 x は真の分布 F から得られたものであること、そしてその性質を反映していることだけは分かっている。そしてこれら観測値の束を経験分布 $\hat{F}$ としてとらえ、これが真の分布 F を十分に近似しているとみなす。そこからのリサンプリング (復元抽出) の結果から推論を行う、と視点が転換される。ここで現実の世界と推論のプロセスの違いが、われわれにとってはむしろ有利にはたらく。現実の世界においては、理論上、無数のサンプリング (観測) が行われうると考え、またそのことを前提にして解析的な評価が行われるが、実際に行われるサンプリング (観測) はただの一回でしかない。ところが、推論のプロセスにおいては、経験分布からのリサンプリングをただの一回に限る必要はなく、計算機資源の許すかぎり何度でも実行することができる (さらにその上、困難な解

析的計算も回避できる)。

- こうして得られる $\hat{\theta^*}$ の B 個の集まりは、$\hat{\theta}$ の品質 (平均、分散、偏り、信頼区間など) を評価する上で十分である。$\hat{\theta^*}$ は $\hat{\theta}$ の分布そのものの近似であるので、これをヒストグラム表示すれば全体の概観を把握でき、分散、偏り、パーセント点 (メディアン、信頼区間) などを具体的な数値として計算できる。さらに、それらの数値を用いて仮説を検定することもできる (検定と信頼区間の推定は表裏一体のものである)。これは真の分布が未知であることによって生ずる問題の大半を解決するものとなる。
- ブートストラップ法はデータのサンプリングとリサンプリングに基礎を持ち、その意味では確率の頻度説と親和性が高く、確率の主観説 (ベイジアン) とは相いれないように表面的には見える。しかし、ベイズ的方法を KL 情報量と尤度との関係で考えるかぎり、そこにそれほどの断絶はない。
- Efron 自身はブートストラップ法をノンパラメトリック最尤法と解釈している。ノンパラメトリック・ブートストラップ法とはパラメトリック・ブートストラップ法の一種であって、そこで用いられているパラメトリック分布が経験分布であるにすぎない。
- ノンパラメトリック統計における尤度の取り扱い方法は多様であり、ブートストラップ法はそのなかの一つにすぎない。ノンパラメトリック尤度の取り扱いが定まらないのは、計算効率の問題ばかりでなく、その背後に確率にかんする思想の違いが潜んでいるためと思われる。
- 経験尤度法とブートストラップ法の関係は奇妙である。Efron らは経験尤度をブートストラップ法を最尤推定と結びつける根拠となしている。経験尤度法は、尤度比検定の考え方を経験尤度にもとづく推論方法にまで拡張したものである。経験尤度法とブートストラップ法は両者ともに、真の分布がはっきりとパラメトリック・モデルとして記述できない領域で効力を発揮する。また両者ともに、推定量の品質 (ばらつき、偏り、信頼区間、検定) を評価することに用いられる。このように共通の目標をもつ以上、互いに競合する手法のようにも思われる。しかし、Owen は経験分布法とブートストラップ法の併用がそれ単独で使われるよりも望ましい結果をもたらす例を挙げている。これらの方法のあいだの関係はまだ完全には解明されていない。

第 5 章

測度論的確率論と確率過程

再び存在論に戻る。ただし、ここでは観測と情報の取り扱いを十分に意識する、という意味では前々節とまったく同じではない。

5.1 測度論的確率論の直観的背景

古典的確率論から測度論的確率論への移行の論理を直観的に取り扱う。結論的に言えば、確率過程を取り扱うためにこそ Kolmogorov の公理体系が必要であった。

伊藤 (2007) は、次のような例題[*1]を挙げている。

> 甲と乙が交互にコイン投げを行い、最初に表を出したほうが勝つものとする (甲が先攻)。このとき、
>
> 1). 勝負の決まるまでの回数は平均して何回か。
> 2). 甲の勝つ確率はいくらか。

$$
\begin{aligned}
&\omega_1 = o\\
&\omega_2 = u, o\\
&\vdots\\
&\omega_i = \underbrace{u, \cdots, u}_{i-1}, o\\
&\vdots\\
&\omega_\infty = u, \cdots
\end{aligned}
$$

ここで、個々の ω_i はサンプルと呼ばれ、サンプルをすべて集めたもの $\Omega = \{\omega_1, \cdots, \omega_\infty\}$ はサンプル空間と呼ばれる。すなわち、サンプル空間は集合である。

ω_i の生じる確率は逐次的に $P(\omega_i) = 1/2^i$ として定まる。またそればかりでなく、サンプルを要素とする任意の集合 E(事象) について確率が加法的に定まる ($P(\mathsf{E}) = \sum_{\omega_i \in \mathsf{E}} P(\omega_i)$)。すなわち、確率分布は集合の関数である。

勝負の決まる回数 x はサンプルによって定まる ($x(\omega_i) = i$)。このようにサンプル空間上で定義された関数は確率変数と呼ばれる (さしあたりは確率分布と直接の関係がないことに注意する)。甲が勝つ、乙が勝つ、ということは i が奇数か、偶数かに帰着される。

第一の問題は回数 x の期待値を求めることである。次の計算により、平均して 2 回で勝敗が決まることがわかる。

$$
\begin{aligned}
E(x(\omega)) &= \sum_{\omega_i \in \Omega} x(\omega_i) P(\omega_i)\\
&= \sum_{i=1}^{\infty} x(\omega_i) P(\omega_i)\\
&= \sum_{i=1}^{\infty} \frac{i}{2^i} \equiv S\\
&\downarrow\\
S - \frac{1}{2}S &= \sum_{i=1}^{\infty} \frac{i}{2^i} - \frac{i-1}{2^i}\\
\frac{1}{2}S &= \sum_{i=1}^{\infty} \frac{1}{2^i} = 1 \to S = 2
\end{aligned}
$$

第二の問題は、事象 E として i が奇数という条件の確率を計算することである。次の計算によって甲の勝つ確率は 2/3 であると結論される。

$$
\sum_{i=0}^{\infty} P(\omega_{2i+1}) = \sum_{i=0}^{\infty} \frac{1}{2^{2i+1}}
$$

[*1] Pascal の問題との類似性に注意されたい。

$$= \frac{1}{2}\sum_{i=0}^{\infty}\frac{1}{4^i}$$
$$= \frac{1}{2}\cdot\frac{4}{3} = \frac{2}{3}$$

まずサンプル空間 Ω があり、Ω 上の確率分布 $P(\omega)$、Ω 上の確率変数 $x(\omega)$、ω にかんする条件として事象 E がある。このような取り扱いは古典的確率論とはまったく異なるものである。

Ω が有限集合、可算無限集合ならば古典的確率論で十分であったが、Ω として空間上の点、粒子の運動の軌跡などをとったばあい、これらは非可算の無限集合となり、とりわけ積分 (リーマン積分) の困難などが現れる。そのために、関数の定義域ではなく値域の分割による新しい積分の方法 (ルベーグ積分) が開発されねばならなかったし、これらを基礎とする測度論 (measure theory) 的な確率論が要求されるようになった[*2]。

5.2　確率過程

確率過程とは「時とともに変動する偶然量を抽象化した概念」である。確率過程は確率空間 $\{\Omega, \mathcal{F}, P\}$ 上で定義される、時間パラメータを含む確率変数の組のことであり、たとえば次のように記述される (長澤 (2015) を参照のこと)。

$$\{x_t(\omega) |\ t \in [a, b], P\}$$

確率空間は以下の内容から構成される。

$\omega(t)$ または ω：サンプル：　すべての時点で微分不可能なジグザグで連続のパス。ひとつひとつのパス ω は時間の関数である。

Ω：サンプル空間：　サンプル ω の全体である。

$x_t(\omega)$：確率変数：　ひとつの例として、サンプル $\omega(t)$ をたどる粒子の位置が挙げられる。位置はサンプル空間上の関数であり、時間 t のパラメータをもつ。$x_t(\omega) \equiv \omega(t)$ である。つまり、サンプルの関数である「位置」と時間の関数である「サンプル」は同一の実体であり、変量をサンプルとするか時間とするか、という見方の違いだけである。

E：事象：　サンプル空間 Ω の部分集合である。E の集まりを $\mathcal{F}$ とする[*3]。例、$\{\omega \mid x_s(\omega) = x, x_t(\omega) \in \Gamma\}$

P：確率測度：　事象の確率を測る測度である。測度一般の公理に加えて、サンプル空間全体の測度が 1 であることが要請される。例、$P(\{\omega \mid x_s(\omega) = x, x_t(\omega) \in \Gamma\})$

5.2.1　確率過程の例 (1)：ランダム・ウォーク

確率過程の例としてランダム・ウォークを挙げる。基礎となる確率変数 $\triangle_i$ としてベルヌーイ試行をとり、これの $i = 1, \cdots, n$ の和を n 分のときの位置 S_n とする。時間パラメータ i は自然数 (たとえば「分」) であり、1 分後にたとえば 1mm 動くと解釈する。

$$S_n = \sum_{i=1}^{n} \triangle_i$$

$$P(\triangle_i) = \begin{cases} 1/2 & \triangle_i = 1 \\ 1/2 & \triangle_i = -1 \end{cases}$$

mgf は次のようになる (ベルヌーイ試行の畳み込み)。

$$M_{\triangle}(z) = \frac{e^{-z} + e^{z}}{2}$$
$$M_S(z) = \left(\frac{e^{-z} + e^{z}}{2}\right)^n$$

mgf より、次のことが分かる。

$$E[\triangle_i] = 0 \quad \mathrm{Var}[\triangle_i] = 1$$
$$E[S_n] = 0 \quad \mathrm{Var}[S_n] = n$$

[*2] 伊藤 (2007) は次のように書いている。「Ω が・・・可算集合のときには各見本点の確率をその総和が 1 となるように与えたらよいのであるが、直線上の点の集合、平面上の点の集合、もっと一般に Brown 運動をする粒子のあらゆる運動のしかたの集合を Ω としてとる場合は、上のような素朴な方法では確率分布を与えることができない。」そのあとに次のような質量分布に関連する興味深い記述が続くが、これについて伊藤はそれ以上の説明をしていない。「しかしながら、確率分布の考えが質量分布とよく似ていることに注意すれば、後者の数学理論である測度論が確率分布にも適用できるであろうと想像される。・・・この考えは正しく、確率論を測度論的に構成することによって、それまで常識や直観を基礎にあいまいな推論を繰り返していた確率論が真に数学の名に値するものとなり、多くの応用上価値のある成果を生み出した。」

[*3] 厳密に言えば、ボレル集合のことである。ボレル集合とは、部分集合の可算回の和、積、差によって得られる集合のことである。

5.2.2　確率過程の例 (2)：ブラウン運動

ブラウン運動*4はランダム・ウォークの極限として構成される。ランダム・ウォーク S_n について、時間の刻みを 1 分ごとから $1/k$ 分ごとに、空間の単位を 1mm から $1/\sqrt{k}$mm に変える。このとき時間 t を $kt = n$ が成立するように定め、これを $B_k(t)$ とする。

$$B_k(t) = \frac{1}{\sqrt{k}} \sum_{i=1}^{kt} \triangle_i$$

時点 s から時点 t までの変位は次のように記述される (とくに形が変わらないことに注意)。

$$B_k(t) - B_k(s) = \frac{1}{\sqrt{k}} \sum_{i=ks}^{kt} \triangle_i$$

平均と分散は次のとおりである。

$$E[B_k(t)] = 0,\ \ \mathrm{Var}[B_k(t)] = \frac{kt}{k} = t$$

$k \to \infty$ の極限をとることによってできる確率過程 B_t は標準ブラウン運動*5と呼ばれる。CLT よりこれは正規分布に従うことがわかる (付録 C.2 を参照のこと)。

$$B_k(t) \overset{k\to\infty}{\longrightarrow} B(t) \equiv B_t \sim N(0,t)$$

mgf は次のとおりである (キュムラント母関数の二次までのテイラー展開から得られる)。

$$\begin{aligned}
\phi_{B_k}(z) &= kt \ln \frac{e^{-\frac{z}{\sqrt{k}}} + e^{\frac{z}{\sqrt{k}}}}{2} \\
&= \frac{t}{2} z^2 + \frac{\phi^{(3)}_{B_k}(\theta)}{6} \left(\frac{z}{\sqrt{k}}\right)^3 \\
&\downarrow \\
M_B(z) &= \exp\left(\frac{t}{2} z^2\right)
\end{aligned}$$

図 5.1 に二次元ブラウン運動のひとつのサンプルを示す。

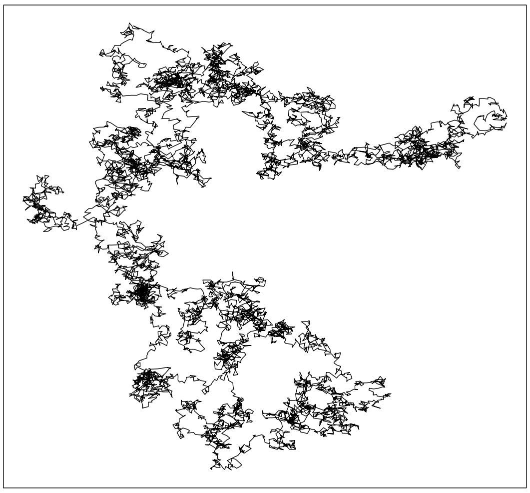


図 5.1　二次元ブラウン運動 (2000 ステップ)

ブラウン運動は次の性質をもつ。

- 増分の定常性*6と t についての連続性

$$\begin{aligned}
B_t - B_s &\sim N(0, t-s) \\
&\longrightarrow \lim_{t\to s} B_t - B_s = 0
\end{aligned}$$

また $B_0 = 0$。

- いたるところで微分不可能

$$\begin{aligned}
\mathrm{Var}[B_t] &= t \\
&\longrightarrow \mathrm{Var}[B_t - B_s] = t - s
\end{aligned}$$

$$\lim_{t\to s} E\left[\left(\frac{B_t - B_s}{t-s}\right)^2\right] = \lim_{t\to s} \frac{1}{t-s} = \infty$$

また微小変化を考えれば、$\mathrm{Var}[dB_t] = dB_t^2 = dt$*7。

*4 1827 年イギリスの植物学者 R.Brown は顕微鏡観察下での植物の花粉に含まれる微粒子の運動を発見した。1905 年、A.Einstein がこれを媒質の熱運動により説明、1908 年に J.Perrin によりこの仮説が実験的に証明された。これにより、原子・分子の実在がはじめて証明された。

長澤は Einstein の論証をもって T.Lucretius Carus (B.C.99-B.C.55) の観察が解明されたものとしているが、Lucretius が Epikouros (B.C.341-B.C.270) の思想の宣伝者であることに鑑みれば、この観察はさらに時代をさかのぼる頃から知られていたのだと言える。

*5 ウィーナー過程とも呼ばれる。

*6 増分の定常性とは増分が $t-s$ にのみ依存することである。

*7 (ブラウン運動にかんする)P.Lévy の条件 (characterization) のうちの一つである。

- 増分の独立性

$t_1 < t_2 < t_3 < t_4$ のとき、

$$B(t_4) - B(t_3) \perp\!\!\!\perp B(t_2) - B(t_1)$$

- 遷移確率が拡散方程式を満たすこと

時点 s から t の遷移確率 Q の対数は次のように記述される。ただし $y = B_t, x = B_s$ である。

$$\ln Q = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln(t-s) - \frac{(y-x)^2}{2(t-s)}$$

t と y で微分すると、

$$\begin{aligned}
\frac{\partial}{\partial t}\ln Q &= \frac{1}{Q}\cdot\frac{\partial Q}{\partial t} \\
&= -\frac{1}{2}\left\{\frac{1}{t-s} - \left(\frac{y-x}{t-s}\right)^2\right\} \\
\frac{\partial}{\partial y}\ln Q &= \frac{1}{Q}\cdot\frac{\partial Q}{\partial y} \\
&= -\frac{y-x}{t-s} \\
&\downarrow \\
-\frac{1}{t-s} &= \frac{1}{Q}\cdot\frac{\partial^2 Q}{\partial y^2} - \frac{1}{Q^2}\cdot\left(\frac{\partial Q}{\partial y}\right)^2 \\
\frac{1}{Q}\cdot\frac{\partial^2 Q}{\partial y^2} &= \frac{1}{Q^2}\cdot\left(\frac{\partial Q}{\partial y}\right)^2 - \frac{1}{t-s} \\
&= \left(\frac{y-x}{t-s}\right)^2 - \frac{1}{t-s}
\end{aligned}$$

よって、

$$\begin{aligned}
\frac{1}{Q}\left(\frac{\partial Q}{\partial t} - \frac{1}{2}\frac{\partial^2 Q}{\partial y^2}\right) &= 0 \\
\to \frac{\partial Q}{\partial t} - \frac{1}{2}\frac{\partial^2 Q}{\partial y^2} &= 0
\end{aligned}$$

標準ブラウン運動について以下のような変数変換を施したものが一般のブラウン運動である。μ は「ドリフト」と呼ばれる。ドリフトは、基礎となる状態変化の確率が 1/2 から外れることによっても生じる。

$$X_t = \mu t + \sigma B_t \sim N(\mu t, \sigma^2 t)$$

mgf は次のとおりである。

$$\begin{aligned}
M_X(z) &= e^{\mu t z} M_B(\sigma z) \\
&= \exp\left(\mu t z + \frac{\sigma^2 t}{2} z^2\right)
\end{aligned}$$

mgf からわかるように、時点 s から時点 t までの変位はやはり正規分布にしたがう。

$$X_t - X_s \sim N\left(\mu(t-s), \sigma^2(t-s)\right)$$

5.2.3　マルコフ過程

遷移確率 (条件付き確率) によって与えられる確率過程はマルコフ過程と呼ばれる。ブラウン運動はマルコフ過程の一つの例である。

時点 $s \to t$ の遷移確率 Q を次のようにあらわす。

$$\begin{aligned}
Q(t, y \mid s, x) &= Pr(X_t = y \mid X_s = x) \\
Q(t, \Gamma \mid s, x) &= Pr(X_t \in \Gamma \mid X_s = x)
\end{aligned}$$

遷移確率 Q は次のチャップマン＝コルモゴロフの等式を満たす $(s \le r \le t)$[*8]。

$$\begin{aligned}
&Q(t, \Gamma \mid s, x) \\
&= \int_R Q(t, \Gamma \mid r, y) Q(r, dy \mid s, x)
\end{aligned}$$

ブラウン運動について上の等式が成立していることはたとえば次のようにして確かめることができる。

$$\begin{aligned}
&M_{s\to r}(z)\cdot M_{r\to t}(z) \\
&= \exp\left(\frac{r-s}{2}z^2 + \frac{t-r}{2}z^2\right) \\
&= \exp\left(\frac{t-s}{2}z^2\right) = M_{s\to t}(z)
\end{aligned}$$

マルコフ過程の遷移確率が、(拡散方程式をその特殊例として含む) 放物型の偏微分方程式で特徴づけられるという事実は「コルモゴロフの定理」として知られている[*9]。$t = 0$ 時点の確率分布 Q(退化分布を含む、また位置にかかる任意の関数の期待値でもよい) が将来に向け

[*8] この等式は $s \to r$、$r \to t$ の二つの遷移確率の畳み込みとも解釈できる。なお連続ではなく離散的なマルコフ連鎖においては、この等式は行列の積として表現される。

[*9] Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung” Math. Ann. 104 (1931) 415

て時間発展する様子を示したものが以下の一行目の前進方程式であり、逆に過去から時間発展してきた確率分布が $t=0$ で特定の確率分布に収れんするさまを示すものが二行目の後退方程式である[*10]。

$$\begin{aligned}\frac{\partial Q(t,y)}{\partial t} &= -\frac{\partial K^{(1)}(y)Q(t,y)}{\partial y} + \frac{\partial^2 K^{(2)}(y)Q(t,y)}{\partial y^2} \\ -\frac{\partial Q(s,x)}{\partial s} &= K^{(1)}(x)\frac{\partial Q(s,x)}{\partial x} + K^{(2)}(x)\frac{\partial^2 Q(s,x)}{\partial x^2}\end{aligned}$$

拡散方程式は上の前進方程式について $K^{(1)}(y)=0$、$K^{(2)}(y)=1/2$ とおいたものに等しい。

5.2.4 伊藤の公式

コルモゴロフの偏微分方程式は、微粒子などの位置の存在確率の時間発展を記述するが、そのサンプルを直接にあたえるものではない。しかしブラウン運動の軌道 B_t が描けるということは、同様の確率過程の軌道についても何らかの記述が可能なのではないかとの予感をもたらす。

伊藤清[*11]は、今日「伊藤の公式」として知られる、ブラウン運動にかんする一種の変数変換の公式をしめすことで上の予想が真実であることを確認した。以下、重川 (2016) よりその内容を簡単にまとめる。

$$\begin{aligned} dB_t &= B_{t+dt} - B_t \\ &\downarrow \\ df(B_t) &= f(B_{t+dt}) - f(B_t) \\ &= f(B_t + dB_t) - f(B_t) \\ &\simeq f'(B_t)dB_t + \frac{1}{2}f''(B_t)dB_t^2 \\ &\downarrow \text{Lévy の条件} \\ &= f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt \end{aligned}$$

上が伊藤の公式の微分型表現である。軌道を得るにはその積分型が必要であり、「伊藤の確率積分」と呼ばれる規約が必要となる。

$$\begin{aligned} & f(B_T) - f(B_0) \\ &= \int_0^T f'(B_t)dB_t + \int_0^T \frac{1}{2}f''(B_t)dt \end{aligned}$$

「伊藤の確率積分」は次のように表現される。ここで、$\Delta = \{0, \cdots, t_j, \cdots, T\}$ である。

$$\begin{aligned} & \int_0^T f'(B_t)dB_t \\ &\equiv \lim_{|\Delta|\to 0} \sum_j f'(B(t_j))(B(t_{j+1}) - B(t_j)) \end{aligned}$$

上と同様の関係は (「伊藤過程」と呼ばれる) ブラウン運動を拡張した確率過程にも成立する。伊藤過程 X_t とは次の確率微分方程式[*12](下はその積分型) の解のことである。

$$\begin{aligned} dX_t &= a(t,X_t)dt + b(t,X_t)dB_t \\ X_t &= X_s + \int_s^t a(r,X_r)dr \\ &\quad + \int_s^t b(r,X_r)dB_{r-s} \end{aligned} \tag{5.1}$$

伊藤過程がマルコフ過程であることは次のようにして示すことができる ($s \to t$ を $s \to o \to t$ に分解する)。t 時点の位置は o 時点より前の情報に依存しないので、マルコフ性が成立している。

$$\begin{aligned} X_t &= X_s + \int_s^o a(r,X_r)dr + \int_o^t a(r,X_r)dr \\ &\quad + \int_s^o b(r,X_r)dB_{r-s} \\ &\quad + \int_o^t b(r,X_r)dB_{r-o} \\ X_o &= X_s + \int_s^o a(r,X_r)dr \\ &\quad + \int_s^o b(r,X_r)dB_{r-s} \\ &\downarrow \\ & X_t - X_o \\ &= \int_o^t a(r,X_r)dr + \int_o^t b(r,X_r)dB_{r-o} \end{aligned}$$

X_t が伊藤過程であるとき、すなわち式 (5.1) の解であるとき、次が成立する。

[*10] これらの導出については付録を参照のこと。
[*11] Markoff 過程ヲ定メル微分方程式,「全国紙上数学談話会」,244 号,1942
[*12] ランジュバン方程式とも呼ばれる。

$$
\begin{aligned}
& df(t, X_t) \\
& \simeq \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2}\left(\frac{\partial^2 f}{\partial t^2}(dt)^2 \right. \\
& \quad \left. + \frac{\partial^2 f}{\partial t \partial x} dt \cdot dX_t + \frac{\partial^2 f}{\partial x^2} dX_t^2 \right) \\
& \downarrow \ (dt)^2 = dt \cdot dX_t = 0 \\
& = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} dX_t^2 \\
& = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x}\left\{a(t, X_t)dt + b(t, X_t)dB_t\right\} \\
& \quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\left\{b(t, X_t)^2 dt\right\} \\
& = \left(\frac{\partial}{\partial t} + a(t, X_t)\frac{\partial}{\partial x} + \frac{1}{2} b(t, X_t)^2 \frac{\partial^2}{\partial x^2}\right) f dt \\
& \quad + b(t, X_t)\frac{\partial f}{\partial x} dB_t \\
& = \mathcal{L} f dt + b(t, X_t)\frac{\partial f}{\partial x} dB_t
\end{aligned}
$$

あるいは、

$$
\begin{aligned}
& f(t, X_t) - f(s, X_s) - \int_s^t \mathcal{L} f(r, X_r) dr \\
& = \int_s^t b(r, X_r) \frac{\partial}{\partial x} f(r, X_r) dB_r
\end{aligned} \tag{5.2}
$$

伊藤の公式 (5.2) からコルモゴロフの偏微分方程式を導出することができる。まず以下のように条件付き期待値を用いて、関数 $u(s, x)$ を定義する。

$$
\begin{aligned}
u(s, x) &= \int Q(r, z \mid s, x) f(z) dz \\
&= E_{(s,x)}\left[f(X_r)\right]
\end{aligned}
$$

同じく条件付き期待値によって以下のようにマルコフ性を表現する*[13]。

$$
E_{(s,x)}\left[f(X_r)\right] = E_{(s,x)}\left[E_{(s-h, X_{s-h})}\left[f(X_r)\right]\right]
$$

$u(s, x)$ の定義およびマルコフ性より、

$$
\begin{aligned}
u(s, x) &= E_{(s,x)}\left[f(X_r)\right] \\
&= E_{(s,x)}\left[E_{(s-h, X_{s-h})}\left[f(X_r)\right]\right] \\
&= E_{(s,x)}\left[u(s-h, X_{s-h})\right]
\end{aligned} \tag{5.3}
$$

式 (5.2) より、$f(t, X_t) \to u(t, X_t)$ として、

$$
\begin{aligned}
& u(t, X_t) - u(s, x) - \int_s^t \mathcal{L} u(r, X_r) dr \\
& = \int_s^t \sigma(r, X_r) \frac{\partial f}{\partial x}(r, X_r) dB_r \\
& \downarrow \ \text{両辺期待値をとる} \\
& E_{(s,x)}\left[u(t, X_t)\right] - u(s, x) \\
& \quad - E_{(s,x)}\left[\int_s^t \mathcal{L} u(r, X_r) dr\right] = 0 \\
& \downarrow \ t \to s - h \\
& E_{(s,x)}\left[u(s-h, X_{s-h})\right] - u(s, x) \\
& \quad - E_{(s,x)}\left[\int_s^{s-h} \mathcal{L} u(r, X_r) dr\right] = 0 \\
& \downarrow \ \text{式 (5.3) より} \\
& E_{(s,x)}\left[\int_{s-h}^{s} \mathcal{L} u(r, X_r) dr\right] = 0 \\
& \downarrow \ \text{両辺極限をとる} \\
& \lim_{h \to 0} \frac{1}{h} E_{(s,x)}\left[\int_{s-h}^{s} \mathcal{L} u(r, X_r) dr\right] \\
& \quad = \mathcal{L} u(s, x) = 0
\end{aligned}
$$

$\mathcal{L} u(s, x) = 0$ はコルモゴロフの後退方程式である*[14]。

なお伊藤の公式を幾何ブラウン運動にたいして適用したものが資産価格のブラック=ショールズ・モデル*[15]と呼ばれるものである。資産価格 S_t の変動が次の幾何ブラウン運動にしたがうものとする。

$$
\frac{dS_t}{S_t} = \mu dt + \sigma dB_t
$$

$f(x) = \ln x$ として、

$$
\begin{aligned}
df(S_t) &= d \ln S_t \\
& \downarrow \ \text{伊藤の公式} \\
&= \frac{dS_t}{S_t} - \frac{1}{2}\left(\frac{dS_t}{S_t}\right)^2 \\
&= (\mu dt + \sigma dB_t) - \frac{1}{2}(\mu dt + \sigma dB_t)^2 \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t
\end{aligned}
$$

*13 定義にしたがって展開すればわかるように、その本質はチャップマン=コルモゴロフ方程式そのものである。

*14 McCauley(2008) は後退方程式の意味を、式 (5.2) がマルチンゲールになるための条件として、マルコフ性に依拠せずに説明している。

*15 1973 年に F.Black と M.Sholes により発表されたもの。後に Sholes はこのモデルに理論的証明を与えた R.Merton とともに LTCM 社に役員として関わり自分たちの理論を実践に活用したが、1999 年に同社は経営破たんした。

よって、資産価格の変動は次のように記述される。

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right\}$$

5.3　確率微分方程式と量子力学

ブラウン運動の理論はその後の統計物理学の発展の確固たる基礎をなし、現象論的な熱力学に分子運動論という実体を与えることとなった。他方、熱輻射理論、光電効果の発見などから光量子仮説が提唱され、さらに原子中の電子運動の考察 (原子模型) から量子力学が構築された。前者の統計物理学と後者の量子力学は、その類似性にもかかわらず、またそれぞれの理論の実社会での成功にもかかわらず、直接の関係が今日においても明瞭になっているとは言い難い。

長澤 (2015) はコルモゴロフの偏微分方程式を外力のない条件下での粒子の運動方程式と解釈することを通じて両理論の直接のつながりを明らかにしようとしている。ここでは、以下できるかぎり長澤の論証にしたがいながらその論理の流れを追ってみる[*16]。

5.3.1　長澤のランダム運動理論

まず長澤が「双子の運動方程式」と呼ぶ以下の方程式系が導入される。ここに含まれる $\bar{c}$ は外力ポテンシャルとされる。

$$\begin{cases} \mathcal{L}\phi + \bar{c}(s,x)\phi = 0 \\ \mathcal{L}^o\widehat{\phi} + \bar{c}(t,y)\widehat{\phi} = 0 \end{cases}$$

ただし、$\mathcal{L}$、$\mathcal{L}^o$ は微分演算子である。

$$\begin{cases} \mathcal{L} = \dfrac{\partial}{\partial s} + \dfrac{1}{2}\sigma^2\dfrac{\partial^2}{\partial x^2} + \sigma^2 b(s,x)\dfrac{\partial}{\partial x} \\ \mathcal{L}^o = -\dfrac{\partial}{\partial t} + \dfrac{1}{2}\sigma^2\dfrac{\partial^2}{\partial y^2} - \dfrac{\partial}{\partial y}\sigma^2 b(t,y) \end{cases}$$

ϕ、$\widehat{\phi}$(「発展方程式」と呼ばれる) は偏微分方程式の基本解[*17]Ψ を用いて次のように表現される。$\widehat{\phi}_a(z)$ は入口関数、$\phi_b(z)$ は出口関数と呼ばれる。

$$\begin{cases} \phi(s,x) = \displaystyle\int \Psi(b,z,\mid s,x)\phi_b(z)dz \\ \widehat{\phi}(t,y) = \displaystyle\int dz\widehat{\phi}_a(z)\Psi(t,y,\mid a,z) \end{cases}$$

Ψ は遷移確率ではないために、そのままでは正規性条件を満たさない。ゆえに双子の方程式はマルコフ過程の性質 (チャップマン=コルモゴロフの等式) からは導出されない。その代わりに次の三つ組みの正規性条件が導入される。

$$\int dx\widehat{\phi}_a(x)\Psi(b,y\mid a,x)\phi_b(y)dy = 1$$

運動方程式の解である「速度」に相当する量 (「発展の流れドリフト」と呼ばれるもの) が定義される。そのために次の分布生成関数 $R(t,x)$、ドリフト・ポテンシャル $S(t,x)$ が導入される。

$$\begin{cases} R(t,x) = \dfrac{\ln\phi(t,x) + \ln\widehat{\phi}(t,x)}{2} \\ S(t,x) = \dfrac{\ln\phi(t,x) - \ln\widehat{\phi}(t,x)}{2} \end{cases}$$

$$\begin{cases} \phi(t,x) = e^{R(t,x)+S(t,x)} \\ \widehat{\phi}(t,x) = e^{R(t,x)-S(t,x)} \end{cases}$$

発展の流れドリフト a、$\widehat{a}$ は次のように定義される。

$$\begin{cases} a(t,x) = \sigma^2\dfrac{\partial}{\partial x}\ln\phi(t,x) \\ \widehat{a}(t,x) = \sigma^2\dfrac{\partial}{\partial x}\ln\widehat{\phi}(t,x) \end{cases}$$

以上の準備のもとで、長澤は以下の手順で双子の方程式の解軌道 X_t を導出している。

1). 推移確率 Q の導入とこれが推移確率の性質 (チャップマン=コルモゴロフの等式と正規性条件) を満たしていることを確認する。推移確率は次のように定義される。

$$Q(t,y\mid s,x) = \frac{\phi(t,y)}{\phi(s,x)}\Psi(t,y\mid s,x)$$

2). 基本解 Ψ と入口・出口関数を指定することにより、初期分布 μ_a、推移確率 Q に従う確率過程 X_t が存在することを示す。X_t がマルコフ過程であることを示す。初期分布は次のように記述さ

*16 簡単のため空間次元は便宜的に 1 次元とし、応用例で必要に応じて拡張する。

*17 カッツの汎関数とも呼ばれている。mgf との類似性に注意。遷移確率は実は mgf と同様に母関数 (generator) の意義を持っている。

れる。

$$\begin{cases} \mu_a(x_0) = \phi_a(x_0)\widehat{\phi}_a(x_0) \\ \phi_a(x_0) = \int \Psi(b, y \mid a, x_0)\phi_b(y)dy \end{cases}$$

3). Q が次の「運動学の方程式」(放物型偏微分方程式) を満たすことを示す。

$$\begin{cases} \mathcal{L}Q + a(s,x)\dfrac{\partial Q}{\partial x} = 0 \\ \mathcal{L}^o Q - \dfrac{\partial a(t,y)Q}{\partial y} = 0 \end{cases}$$

4). 運動学の方程式から次の解軌道 X_t を得る。

$$X_t = X_a + \sigma B_{t-a} + \int_a^t \left[\sigma^2 b(r, X_r) + a(r, X_r)\right] dr$$

また X_t の分布密度 $\mu_t(x)$ は次のように記述される。

$$\mu_t(x) = \phi(t,x)\widehat{\phi}(t,x) = e^{2R(t,x)}$$

1) については次のように確認される。

$$\begin{aligned} &\int Q(r, dz \mid s, x)Q(t, y \mid r, z) \\ &= \int \frac{\phi(r,z)}{\phi(s,x)}\Psi(r, dz \mid s, x)\frac{\phi(t,y)}{\phi(r,z)}\Psi(t, y \mid r, z) \\ &= \frac{\phi(t,y)}{\phi(s,x)}\int \Psi(r, dz \mid s, x)\Psi(t, y \mid r, z) \\ &= \frac{\phi(t,y)}{\phi(s,x)}\Psi(t, y \mid s, x) \\ &= Q(t, y \mid s, x) \end{aligned}$$

$$\begin{aligned} &\int Q(t, dy \mid s, x) \\ &= \frac{1}{\phi(s,x)}\int \Psi(t, dy \mid s, x)\phi(t,y) \\ &= \frac{\phi(s,x)}{\phi(s,x)} = 1 \end{aligned}$$

2) については、まず X_t の存在が次のように示される。

$$1 = \int dx_0 \widehat{\phi}_a(x_0)\Psi(b, y \mid a, x_0)\phi_b(y)dy$$

にて、区間 $[a,b]$ を次のように $a = t_0 < t_1 < \cdots < t_n < b$ と分割することにより $X_{t_1}, \cdots, X_{t_n}$ の存在を示せる。

$$\begin{aligned} &E\left[\{\omega \mid \omega(t_i) \in \Gamma_i, i = 1, \cdots, n\}\right] \\ &= \int dx_0 \widehat{\phi}_a(x_0)\Psi(t_1, x_1 \mid t_0, x_0)\delta_{\Gamma_1}(x_1)dx_1 \\ &\quad \Psi(t_2, x_2 \mid t_1, x_1)\delta_{\Gamma_2}(x_2)dx_2 \cdots \\ &\quad \cdots \Psi(b, y \mid t_n, x_n)\phi_b(y)dy \end{aligned}$$

ここで、

$$\delta_{\Gamma_i}(x_i) = \begin{cases} 1 & x_i \in \Gamma_i \\ 0 & x_i \notin \Gamma_i \end{cases}$$

X_t がマルコフ過程であることを示すには、上の式を f で書き直すことによって、

$$\begin{aligned} &E\left[\{f(X_{t_1}, \cdots, X_{t_n})\}\right] \\ &= \int dx_0 \widehat{\phi}_a(x_0)\Psi(t_1, x_1 \mid t_0, x_0)dx_1 \\ &\quad \Psi(t_2, x_2 \mid t_1, x_1)dx_2 \cdots \\ &\quad \cdots \Psi(b, y \mid t_n, x_n)\phi_b(y)dy f(x_1, \cdots, x_n) \end{aligned}$$

($f(x_1, \cdots, x_n) = \delta_{\Gamma_1}(x_1)\cdots\delta_{\Gamma_n}(x_n)$ に相当する)

さらに Q で書き直すと、

$$\begin{aligned} &E\left[\{f(X_{t_1}, \cdots, X_{t_n})\}\right] \\ &= \int dx_0 \widehat{\phi}_a(x_0)\phi_a(x_0)Q(t_1, x_1 \mid t_0, x_0)dx_1 \\ &\quad Q(t_2, x_2 \mid t_1, x_1)dx_2 \cdots \\ &\quad \cdots Q(b, y \mid t_n, x_n)dy f(x_1, \cdots, x_n) \\ &\downarrow \textstyle\int Q(b, y \mid t_n, x_n)dy = 1 \\ &= \int dx_0 \mu_a(x_0)Q(t_1, x_1 \mid t_0, x_0)dx_1 \\ &\quad Q(t_2, x_2 \mid t_1, x_1)dx_2 \cdots \\ &\quad \cdots Q(t_n, x_n \mid t_{n-1}, x_{n-1})dy f(x_1, \cdots, x_n) \end{aligned}$$

3)-第一式については、まず二回微分可能な一般的関数、ϕ、p より u を次のように構成する。

$$u(s,x) = \frac{p(s,x)}{\phi(s,x)}$$

すると、

$$\mathcal{L}u = \frac{1}{\phi}\mathcal{L}p - \frac{p}{\phi^2}\mathcal{L}\phi - \sigma^2\frac{\phi'}{\phi}u'$$

ただし、$\phi' = \partial\phi/\partial x$ である。$\overline{c}$ を任意関数として、

$$\mathcal{L}u + \sigma^2 \frac{\phi'}{\phi} u' = \frac{1}{\phi}\mathcal{L}p - \frac{p}{\phi^2}\mathcal{L}\phi$$
$$= \frac{1}{\phi}(\mathcal{L}p + \bar{c}p) - \frac{u}{\phi}(\mathcal{L}\phi + \bar{c}\phi)$$

ϕ は $\mathcal{L}\phi + \bar{c}\phi = 0$ の解として、また p を基本解 Ψ から次のように

$$p(s,x) = \Psi(t, y \mid s, x)\phi(t, y)$$

構成すれば、$\mathcal{L}p + \bar{c}p = 0$ より、

$$\mathcal{L}u + \sigma^2 \frac{\phi'}{\phi} u' = 0$$

定義より、

$$u(s,x) = \frac{p(s,x)}{\phi(s,x)} = Q(t, y \mid s, x)$$

また $a(s,x) = \sigma^2\phi'/\phi$ より、

$$\mathcal{L}Q + a(s,x)\frac{\partial Q}{\partial x} = 0$$

3)-第二式については、

$$\widehat{u}(t,y) = p(t,y)\phi(t,y)$$

$$\mathcal{L}^o\widehat{u} = -p(\mathcal{L}\phi) + \phi(\mathcal{L}^o p) + \left(\sigma^2\phi' p\right)'$$

$$\mathcal{L}^o\widehat{u} - \left(\sigma^2\phi' p\right)' = -p(\mathcal{L}\phi) + \phi(\mathcal{L}^o p)$$
$$= -p(\mathcal{L}\phi + \bar{c}\phi) + \phi(\mathcal{L}^o p + \bar{c}p)$$

ϕ は $\mathcal{L}\phi + \bar{c}\phi = 0$ の解、

$$p(t,y) = \frac{\Psi(t, y \mid s, x)}{\phi(s,x)} \to \widehat{u} = Q$$

として、$\mathcal{L}^o p + \bar{c}p = 0$。
ゆえに、

$$\mathcal{L}^o Q - \frac{\partial a(s,x)Q}{\partial y} = 0$$

4) について、まず式 (5.2) が $\mathcal{L}u = 0$ に対応していることを利用し、$f(t, X_t) \equiv X_t$ とおくことにより、また $a(t, X_t) \to b(t, X_t)$、$b(t, X_t) \to \sigma(t, X_t)$ と置き換えることにより、

$$\mathcal{L}f(t, X_t) = b(t, X_t)$$

$$\begin{cases} X_t - X_s - \int_s^t b(r, X_r)dr = \int_s^t \sigma(r, X_r)dB_{r-s} \\ \mathcal{L}u = \dfrac{\partial u}{\partial t} + b(t, X_t)\dfrac{\partial u}{\partial y} + \dfrac{1}{2}\sigma^2(t, X_t)\dfrac{\partial^2 u}{\partial y^2} = 0 \end{cases}$$

これを偏微分方程式と確率微分方程式とのあいだの読み替え表とみなす。$\mathcal{L}Q + a(s,x)Q' = 0$ に対応するものとして、

$$\begin{cases} b(t, X_t) \to \sigma^2 b(t, X_t) + a(t, X_t) \\ \sigma^2(t, X_t) \to \sigma^2 \end{cases}$$

より、

$$X_t = X_a + \sigma B_{t-a} + \int_a^t \left[\sigma^2 b(r, X_r) + a(r, X_r)\right] dr$$

5.3.2　ランダム運動の具体例 (1)：調和振動子

長澤は以下のような平面上のフックの力の下での粒子の運動を例示している[*18]。

$$c(x,y) = -\frac{1}{2}\kappa^2(x^2 + y^2)$$

運動方程式は次のようになる。

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}\sigma^2\nabla^2\phi - \frac{1}{2}\kappa^2(x^2 + y^2)\phi = 0$$

$\phi(t,x,y) = e^{\lambda t}\varphi(x,y)$ として、時間変数を消去する (定常的な運動を考えている)。

$$-\frac{1}{2}\sigma^2\nabla^2\varphi + \frac{1}{2}\kappa^2(x^2 + y^2)\varphi = \lambda\varphi$$

直交座標系から極座標系への変換を行う。

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$

$$\nabla^2\varphi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\varphi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\varphi}{\partial\theta^2}$$

変換後の運動方程式は次のようになる。

$$-\frac{1}{2}\sigma^2\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\varphi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\varphi}{\partial\theta^2}\right) + \frac{1}{2}\kappa^2 r^2\varphi = \lambda\varphi$$

変数分離のため $\varphi(r,\theta) = R(r)\Phi(\theta)$ とおく。分離のため任意定数 m を導入する。

[*18] 長澤は Pauling and Wilson(1935) の Chapter 17:3Dimensional Harmonic Oscillator(p.107 以降) を参照している。

$$\begin{cases} -\dfrac{\partial^2\Phi}{\partial\theta^2} = m^2\Phi \\ -\dfrac{1}{2}\sigma^2\dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\dfrac{\partial R}{\partial r}\right) + \left(\dfrac{1}{2}\sigma^2\dfrac{m^2}{r^2} + \dfrac{1}{2}\kappa^2 r^2\right)R = \lambda R \end{cases}$$

第一式について規格化した解は次のとおりである。

$$\Phi(\theta) = \frac{1}{\sqrt{2\pi}}e^{im\theta} = \frac{1}{\sqrt{2\pi}}(\cos m\theta + i\sin m\theta)$$

ところで $\Phi(\theta)$ は任意定数の m があるためにそのままでは一価関数にならない。つまり $\theta=0$ と $\theta=2\pi$ で異なる値をとる。$\Phi(\theta)$ が一価関数となるためには、$m = 0, \pm1, \pm2, \cdots$ でなければならない。

第二式について、$\lambda' = 2\lambda/\sigma^2$、$\kappa^2/\sigma^2 = \alpha^2$ と置くことにより、

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right) + \left(\lambda' - \alpha^2 r^2 - \frac{m^2}{r^2}\right)R = 0$$

まず r が大きいばあいの近似解を求める。解くべき方程式とその解は次のようになる。

$$\begin{cases} \dfrac{\partial^2 R}{\partial r^2} - \alpha^2 r^2 R = 0 \\ R(r) \simeq \exp\left(-\dfrac{\alpha}{2}r^2\right) \end{cases}$$

ズレを $f(r)$ とすると、

$$R(r) = \exp\left(-\frac{\alpha}{2}r^2\right)f(r)$$

これを第二式に代入することより、

$$f'' - 2\alpha r f' + \frac{1}{r}f' + (\lambda' - 2\alpha)f - \frac{m^2}{r^2}f = 0$$

ここで、$\sqrt{\alpha}r = \xi$、$F(\xi) \equiv f(x)$ と置き換える。

$$\frac{\partial^2 F}{\partial\xi^2} - 2\xi\frac{\partial F}{\partial\xi} + \frac{1}{\xi}\frac{\partial F}{\partial\xi} + \left(\frac{\lambda'}{\alpha} - 2\right)F - \frac{m^2}{\xi^2}F = 0$$

ここで[*19]、

$$F(\xi) = \xi^s\sum_{\nu=0}^{\infty}a_\nu\xi^\nu$$

とする。代入より以下の式が得られる。

$$\begin{aligned} 0 &= (s^2 - m^2)a_0\xi^{s-2} + \left\{(s+1)^2 - m^2\right\}a_1\xi^{s-1} \\ &\quad + \left[\left\{(s+2)^2 - m^2\right\}a_2 + \left\{\frac{\lambda'}{\alpha} - 2(s+1)\right\}a_0\right]\xi^s \\ &\quad + \cdots \\ &\quad + \left[\left\{(s+\nu)^2 - m^2\right\}a_\nu\right. \\ &\qquad \left. + \left\{\frac{\lambda'}{\alpha} - 2(s+\nu-1)\right\}a_{\nu-2}\right]\xi^{s+\nu-2} \\ &\quad + \cdots \end{aligned}$$

ξ が任意であることから、係数はすべてゼロでなければならない。

$$\begin{cases} (s^2 - m^2)a_0 = 0 \\ \left\{(s+1)^2 - m^2\right\}a_1 = 0 \\ \left\{(s+\nu)^2 - m^2\right\}a_\nu + \left\{\dfrac{\lambda'}{\alpha} - 2(s+\nu-1)\right\}a_{\nu-2} = 0 \end{cases}$$

第一式で $a_0 \neq 0$、$s > 0$ より、$s = |m|$。これを第二式に代入することより、$a_1 = 0$。第三式より、ν のすべての奇数項はゼロとならなければならない。また、多項式が無限大に発散しないようにするため、ある項数 $(\nu = n+2)$ 以上の偶数項 a_{n+2} がゼロとならなければならない。そのため、

$$\frac{\lambda'}{\alpha} = 2(|m| + n + 1) \quad (n = 0, 2, 4, \cdots)$$

よって、

$$\begin{cases} \lambda_{|m|+n} = \sigma\kappa(|m| + n + 1) \\ R_{m,n}(r) = F_{|m|,n}\left(\sqrt{\dfrac{\kappa}{\sigma}}r\right)\exp\left(-\dfrac{\kappa}{2\sigma}r^2\right) \end{cases}$$

$m = \pm1$、$n = 0$ のばあい

$\varphi(r,\theta)$ は定義とおり次のようになる (規格化のための定数を β とする)。

$$\begin{aligned} \varphi(r,\theta) &= R_{\pm1,1}(r)e^{\pm i\theta} \\ &= \beta r\exp\left(-\frac{\kappa}{2\sigma}r^2 \pm i\theta\right) \end{aligned}$$

ところが、時間変数を戻した $\phi(t,r,\theta)$ は定義とおりに構成しても元の運動方程式を満たすものにはならない。

[*19] $F(\xi)$ は次のような合流型超幾何関数 U とラゲール陪多項式 L の線形和である。

$$F_{|m|,n}(\xi) = c_1\xi^{|m|}U(-n/2.|m|+1, \xi^2) + c_2\xi^{|m|}L_{n/2}^{|m|}(\xi^2)$$

なお、$F_{0,0}(\xi) = c_1 + c_2 = a_0$、$F_{1,0}(\xi) = (c_1 + c_2)\xi = a_0\xi$ である。

そのため、時間について i 倍とした新たな関数 $\psi(t,r,\theta)$ を次のように構成する。また指数部を実部 R_1 と虚部 $S_{\pm1}$ に分割する。

$$\begin{aligned}\psi(t,r,\theta) &= e^{-i\lambda t}\varphi(r,\theta)\\ &= \beta e^{R_1+iS_{\pm1}}\end{aligned}$$

$$\begin{cases} R_1 = \ln r - \dfrac{\kappa}{2\sigma}r^2 \\ S_{\pm1} = -2\sigma\kappa t \pm \theta \end{cases}$$

同じ実部と虚部を使い「発展方程式」ϕ と $\widehat{\phi}$ が構成される。

$$\begin{cases} \phi(t,r,\theta) = \beta e^{R_1+S_{\pm1}} \\ \widehat{\phi}(t,r,\theta) = \beta e^{R_1-S_{\pm1}} \end{cases}$$

これら発展方程式が満たす運動方程式は以下のようになる。

$$\begin{cases} \dfrac{\partial\phi}{\partial t} + \dfrac{1}{2}\sigma^2\nabla^2\phi + \left\{\dfrac{1}{2}\kappa^2 r^2 + \widetilde{V}(r)\right\}\phi = 0 \\ -\dfrac{\partial\widehat{\phi}}{\partial t} + \dfrac{1}{2}\sigma^2\nabla^2\widehat{\phi} + \left\{\dfrac{1}{2}\kappa^2 r^2 + \widetilde{V}(r)\right\}\widehat{\phi} = 0 \end{cases}$$

運動方程式は最初のものとは一致せず、以下の追加項が生じる。長澤はこれを外力に由来しないという意味で「自己ポテンシャル」と名付けている。

$$\widetilde{V}(r) = \frac{\sigma^2}{r^2} - 4\sigma\kappa$$

粒子の存在確率 μ は二つの発展方程式の積より次のように示される。この式を図示すると図 5.2 のようになる。原点付近では密度がゼロとなり原点から等距離離れたところに二つのモードが現れている。

$$\begin{aligned}\mu(t,r,\theta) &= \phi(t,r,\theta)\widehat{\phi}(t,r,\theta)\\ &= \beta^2 e^{2R_1}\\ &= \beta^2 r^2 \exp\left(-\frac{\kappa}{\sigma}r^2\right)\end{aligned}$$

「発展の流れドリフト」は次のようになる。

$$\begin{cases} a_r(r,\theta) = \sigma^2 \dfrac{1}{\phi}\dfrac{\partial\phi}{\partial r} = \dfrac{\sigma^2}{r} - \sigma\kappa r \\ a_\theta(r,\theta) = \sigma^2 \dfrac{1}{\phi}\dfrac{1}{r}\dfrac{\partial\phi}{\partial\theta} = \pm\dfrac{\sigma^2}{r} \end{cases}$$

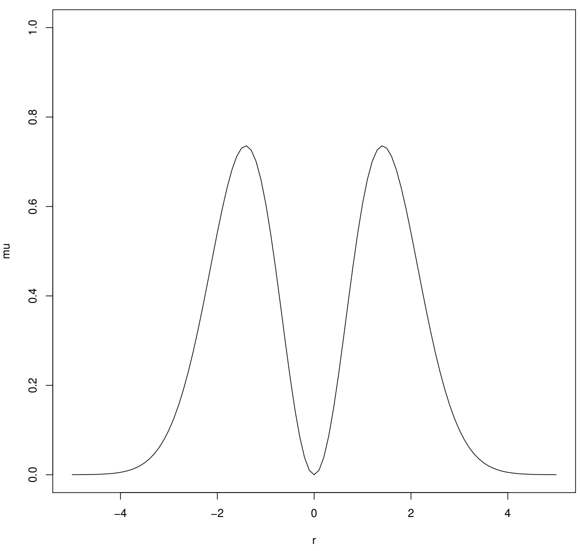


図 5.2　$\mu(t,r,\theta)$

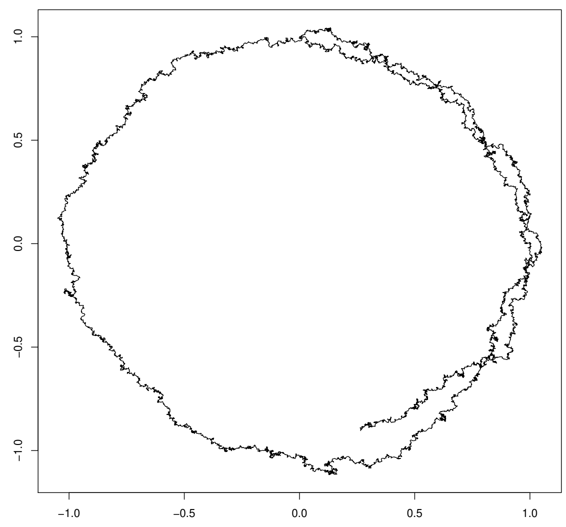


図 5.3　フックの力の下での運動 ($m=+1$、$n=0$)

a_r を見ると半径方向のドリフトは $r = \pm\sqrt{\sigma/\kappa}$ で均衡することがわかる。これは図 5.2 の二つのモードに対応している。a_θ は $m = \pm 1$ に対応してそれぞれ固有の回転方向をもっている。$m = +1$ で反時計まわり、$m = -1$ で時計まわりのドリフトがある。

運動学の方程式は次のようになる。

$$\frac{\partial \boldsymbol{u}}{\partial t} + \frac{1}{2}\sigma^2\nabla^2\boldsymbol{u} + \boldsymbol{a}\nabla\boldsymbol{u} = 0$$

これに対応する確率過程は次のようになる[*20]。

$$d\boldsymbol{X}_t = \boldsymbol{a}dt + \sigma d\boldsymbol{B}_t$$

すなわち、

$$\begin{pmatrix} dr_t \\ d\theta_t \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma^2}{r_t} - \sigma\kappa r_t \\ \pm\dfrac{\sigma^2}{r_t} \end{pmatrix} dt + \sigma \begin{pmatrix} dB_t^r \\ dB_t^\theta \end{pmatrix}$$

図 5.3 はシミュレーションの結果である ($\sqrt{\sigma/\kappa} = 1$ に調整している。10000 ステップ)。粒子はいびつであるがほぼ円軌道を描いて運動している。

5.3.3 シュレディンガー方程式

先の事例において発展関数 ϕ がそのままでは構成されず、虚数を含む関数 ψ(長澤はこれを「複素発展関数」と呼んでいる) を迂回しなければならなかった。この「不具合」についてあらためて考察される。長澤が以下の命題 A と B が、C の条件の下で互いに同等であることを示している。

A:シュレディンガー方程式

複素発展関数 $\psi(t,x) = e^{R(t,x)+iS(t,x)}$ は次のシュレディンガー方程式を満たす。

$$i\frac{\partial\psi}{\partial t} + \frac{1}{2}\sigma^2(\nabla + ib(t,x))^2\psi - V(t,x)\psi = 0$$

B:運動方程式

発展関数 ϕ、$\widehat{\phi}$ は次の運動方程式を満たす。

$$\begin{cases} \dfrac{\partial\phi}{\partial t} + \dfrac{1}{2}\sigma^2(\nabla + b(t,x))^2\phi + c(t,x)\phi = 0 \\ -\dfrac{\partial\widehat{\phi}}{\partial t} + \dfrac{1}{2}\sigma^2(\nabla - b(t,x))^2\widehat{\phi} + c(t,x)\widehat{\phi} = 0 \end{cases}$$

C:自己ポテンシャル

自己ポテンシャルを、

$$\widetilde{V}(t,x) = 2\frac{\partial S}{\partial t} + \sigma^2(\nabla S)^2 + 2\sigma^2 b\cdot\nabla S$$

とすると次の関係式が成立する。

$$V(t,x) + c(t,x) + \sigma^2 b^2 + \widetilde{V}(t,x) = 0$$

A の成立から B の成立が次のように導かれる。まず、ψ をシュレディンガー方程式に代入することにより、

$$\psi(-V + R^* + iS^*) = 0 \rightarrow R^* - V = S^* = 0$$

が示される。ここで、

$$R^* = -\frac{\partial S}{\partial t} + \frac{1}{2}\sigma^2\left\{\nabla^2 R + (\nabla R)^2 - (\nabla S)^2 - 2b\nabla S - b^2\right\}$$

$$S^* = \frac{\partial R}{\partial t} + \frac{1}{2}\sigma^2\left\{\nabla^2 S + 2(\nabla R)(\nabla S) + 2b(\nabla R) + \nabla b\right\}$$

ϕ を運動方程式の左辺に代入することにより、

$$\phi\left(R^* + S^* + c + \sigma^2 b^2 + \widetilde{V}\right) \tag{5.4}$$

$R^* = V$、$S^* = 0$ より、

$$\phi\left(V + c + \sigma^2 b^2 + \widetilde{V}\right)$$

$V + c + \sigma^2 b^2 + \widetilde{V} = 0$ ならば全体はゼロとなり、命題 B が成立する ($\widehat{\phi}$ についても同様)[*21]。

逆に B が成立するならば、$\mu_t = \phi_t\widehat{\phi}_t$ が運動学の方程式 $\mathcal{L}^o\mu - \nabla(a\mu) = 0$ を満たすことから、

[*20] 長澤は一度この式を展開して r と θ の二つの方程式に分割している。その計算過程で、論旨には影響しないものの直交座標系の計算と極座標系の計算を混在させてしまっているように見受けられる。

[*21] なお、先の事例のばあい、$b(t,r,\theta) = 0, V(t,r,\theta) = \frac{1}{2}\kappa^2 r^2, \widetilde{V}(t,r,\theta) = \frac{\sigma^2}{r^2} - 4\sigma\kappa, -c(t,r,\theta) = \frac{1}{2}\kappa^2 r^2 + \frac{\sigma^2}{r^2} - 4\sigma\kappa$ であり、$V + c + \widetilde{V} = 0$ は成立している。

[*22] 任意関数 $u = p\widehat{\phi}$ に $\mathcal{L}$ を作用させ、$\mathcal{L}u - \sigma^2 p'\widehat{\phi}' = \widehat{\phi}\mathcal{L}p + p\mathcal{L}\widehat{\phi}$ を得る。$\mathcal{L} + \mathcal{L}^o = \sigma^2(\nabla^2 - (\nabla b))$ を $\mathcal{L}\widehat{\phi}$ に適用し、若干の式変形を行えば得られる。

$$-\frac{\partial\mu}{\partial t}+\frac{1}{2}\sigma^2\nabla^2\mu-\nabla\cdot\left\{(\sigma^2 b+a)\mu\right\}=0$$

同様に $\mathcal{L}\mu-\sigma^2\mu\nabla b-\nabla(\widehat{a}\mu)=0$ より[*22]、

$$\frac{\partial\mu}{\partial t}+\frac{1}{2}\sigma^2\nabla^2\mu-\nabla\cdot\left\{(-\sigma^2 b+\widehat{a})\mu\right\}=0$$

二式の両辺を差し引きすることにより、

$$\frac{\partial\mu}{\partial t}+\nabla\cdot\left\{\left(\sigma^2 b+\frac{a-\widehat{a}}{2}\right)\mu\right\}=0$$

$\mu=e^{2R}$、$(a-\widehat{a})/2=\sigma^2\nabla S$ を代入し、

$$\frac{\partial R}{\partial t}+\frac{1}{2}\sigma^2\nabla(b+\nabla S)+\sigma^2(b+\nabla S)\nabla R=S^*=0$$

もし条件 C が成立しているならば、式 (5.4) において、$R^*-V=0$。

ゆえにシュレディンガー方程式

$$\psi(-V+R^*+iS^*)=0$$

より命題 A が成立する。

長澤は、命題 A と B の同等性の意義を次のように説明している。

1). 生成関数 (R,S) を双子の運動方程式の解として求めることは現実的にはほとんど不可能であること。(R,S) はシュレディンガー方程式の解から求めなければならない。
2). 生成関数 (R,S) から発展の流れドリフト a、$\widehat{a}$ を経て、確率過程 X_t を求めることは容易であること。
3). シュレディンガー方程式のみでは、a、$\widehat{a}$ が何を意味するのか解釈が困難であること。とりわけ複素発展関数の重ね合わせ(「量子もつれ」)をめぐって混乱した議論が行われてきた。

長澤は複素発展関数 ψ_1 と ψ_2 の重ね合わせを「絡み合った複素発展関数」と呼び、次のように生成関数 (R^*,S^*) による表示を与えている。これによれば、(R^*,S^*) が元の複素発展関数 ψ_1、ψ_2 の生成関数と異なることは明白である。

$$\begin{aligned}\psi^*(t,x)&=\alpha\psi_1(t,x)+\beta\psi_2(t,x)\\&=e^{R^*(t,x)+iS^*(t,x)}\end{aligned}$$

ただし、

$$\int|\psi^*(t,x)|^2dx=1$$

ここから長澤は「絡み合った発展関数」を次のように定義し、これに対応する (元の確率過程とは異なる) 新たな確率過程が存在することを示している。

$$\begin{cases}\phi^*(t,x)=e^{R^*(t,x)+S^*(t,x)}\\ \widehat{\phi}^*(t,x)=e^{R^*(t,x)-S^*(t,x)}\end{cases}$$

上の新たな発展関数より「絡み合ったドリフト」$a^*(t,x)=\sigma^2\phi^{*\prime}/\phi^*=\sigma^2(\nabla R^*+\nabla S^*)$ が導かれ、したがって対応する確率微分方程式も次のように導出される。

$$X_t^*=X_s^*+\sigma B_{t-s}+\int_s^t a^*(r,X_r)dr$$

新たな確率過程を辿る粒子の分布密度 μ^* は次のように計算される。

$$\begin{aligned}\mu^*(t,x)&=\phi^*\widehat{\phi}^*=\psi^*\overline{\psi^*}=|\psi^*|^2\\&=|\alpha\psi_1+\beta\psi_2|^2\\&=|\alpha|^2e^{2R_1}+|\beta|^2e^{2R_2}\\&\quad+2e^{R_1+R_2}\Re(\alpha\overline{\beta})\cos(S_1-S_2)\\&\quad-2e^{R_1+R_2}\Im(\alpha\overline{\beta})\sin(S_1-S_2)\end{aligned}$$

5.3.4 ランダム運動の具体例 (2):自由粒子の運動

「絡み合った発展関数」の意味、またとりわけ「二重スリットの実験」の意味を明らかにするために、その準備として二次元の外力のない自由粒子の運動を考察する。

多くのばあい外力のないシュレディンガー方程式の解として定常解[*23]が紹介されるが、ランダム運動を考察する際にはこれは適切ではない。運動量の時間変化が明

[*23] 定常解については、Pauling and Wilson(1935) Chapter 13 "The Free Particle",p.90 以降を参照。定常解は W.Heisenberg の不確定性原理との関係で言及されることがある。

[*24] 新谷はこれを「対数微分法」と呼んでいる。なお、以下では $\hbar=h/(2\pi)=1$、$1/m=\sigma^2$ とする。また最初は一次元で定式化する。

示されなければならないので、ここでは非定常解を議論する。また、非定常解を得るため、新谷 (2010) によって提案された方法[*24]を採用する。

$$i\frac{\partial}{\partial t}\psi + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}\psi - V\psi = 0$$

ここで次の変数変換を行う。

$$T = \frac{1}{i}\frac{\partial}{\partial x}\ln\psi$$

結果は次のようになる[*25]。

$$\frac{\partial T}{\partial t} + \sigma^2 T\frac{\partial T}{\partial x} - \frac{i}{2}\sigma^2\frac{\partial^2 T}{\partial x^2} - \frac{\partial V}{\partial x} = 0$$

これ以降外力をゼロとする ($V = 0$)。また初期状態[*26]として次を仮定する。

$$\psi(0,x) = \beta\exp\left(-\frac{x^2}{2a^2} + i\kappa x\right)$$

この条件の下で次の解が得られる[*27]。

$$T(t,x) = \frac{A}{i}\left(-\frac{x}{a^2} + i\kappa\right), \text{ただし } A = \left(1 + \frac{i\sigma^2}{a^2}t\right)^{-1}$$

これをもとに戻せば最終的に次の解が得られる。

$$\psi(t,x) = \beta\sqrt{A}\exp\left\{A\left(-\frac{x^2}{2a^2} + i\kappa x - \frac{i}{2}\kappa^2\sigma^2 t\right)\right\}$$

粒子の分布密度は次のようになる。

$$\begin{aligned}\mu(t,x) &= |\psi|^2 \\ &= |\beta|^2\sqrt{|A|^2}\exp\left\{-\frac{|A|^2}{a^2}\left(x - \kappa\sigma^2 t\right)^2\right\}\end{aligned}$$

生成関数 (R,S) は次のようになる[*28]。

$$\begin{cases} R(t,x) &= \dfrac{1}{2}\left(\ln\psi + \ln\overline{\psi}\right) = \dfrac{1}{2}\ln\mu(t,x) \\ &= \dfrac{1}{4}\ln|A|^2 - \dfrac{|A|^2}{2a^2}(x-\kappa\sigma^2 t)^2 + const. \\ S(t,x) &= -\dfrac{i}{2}\left(\ln\psi - \ln\overline{\psi}\right) \\ &= -\dfrac{i}{4}(\ln A - \ln\overline{A}) \\ &\quad + \dfrac{|A|^2}{2}\left(\dfrac{\sigma^2 t}{4a^4}x^2 + 2\kappa x - \kappa^2\sigma^2 t\right) \end{cases}$$

以降、二次元に拡張する。「発展の流れドリフト」は次のようになる。

$$\begin{aligned} a(t,\boldsymbol{x}) &= \sigma^2(\nabla R + \nabla S) \\ &= \sigma^2|A|^2\left\{\frac{1}{a^2}\left(\frac{\sigma^2 t}{2a^2} - 1\right)\boldsymbol{x} + \left(1 + \frac{\sigma^2 t}{a^2}\right)\boldsymbol{\kappa}\right\} \\ &= \alpha\boldsymbol{x} + \gamma\boldsymbol{\kappa} = \alpha\begin{pmatrix} x \\ y \end{pmatrix} + \gamma\begin{pmatrix} \kappa_x \\ \kappa_y \end{pmatrix} \end{aligned}$$

対応する確率過程は次のとおりである ($\kappa_y = 0$ とする)。

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} \alpha_t X_t + \kappa_x\gamma_t \\ \alpha_t Y_t \end{pmatrix} dt + \sigma\begin{pmatrix} dB_t^x \\ dB_t^y \end{pmatrix}$$

$$\begin{cases} \alpha_t = \dfrac{\sigma^2|A|^2}{a^2}\left(\dfrac{\sigma^2 t}{2a^2} - 1\right) \\ \gamma_t = \sigma^2|A|^2\left(1 + \dfrac{\sigma^2 t}{a^2}\right) \end{cases}$$

なお、$a \to \infty$ のばあいは実は定常解と同一となる。逆に言えば、定常解とは非定常解において $a \to \infty$ の極限をとったものである。対応する確率過程は次のように極めて単純なものとなる。

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} \sigma^2\kappa_x \\ 0 \end{pmatrix} dt + \sigma\begin{pmatrix} dB_t^x \\ dB_t^y \end{pmatrix}$$

この定常状態は、初期状態において位置が空間上の一点 (すなわち退化分布) に収束し、確定していること (すなわち零点振動の欠如) を認めながらも、その他の経路においてはランダム運動を認める、という矛盾したものである。われわれは、定常状態を現実には決して到達しえない特殊な極限状態であると考えなければならない。

長澤は外力のないランダム運動について定常解を念頭において考察しており、この状態が特殊なものであるとの認識をもっていない。このことが不確定原理についての彼の解釈に混乱を持ち込んでいる危惧は否定できない。

[*25] 最初に $\psi = \exp(\mathcal{S})$ と置いて $\mathcal{S}$ についての微分方程式を得る。さらにこれを微分したうえで $\partial\mathcal{S}/\partial x = iT$ と置き換えることにより得られる。この方程式について、新谷はニュートン力学との類推、すなわち第一項から順次、運動量の時間変化、ドリフト、量子論的ポテンシャル、外力ポテンシャルと解釈し、確率過程との関係に言及している。

[*26] 新谷はこれを「波束」、つまり複数の粒子の集合体と捉えているが、ランダム運動の立場からするならば、単一の粒子を記述する入口関数ないし出口関数であり、$t = 0$ での粒子の位置の分布を反映したものと捉えるべきであろう。

[*27] 計算過程は付録に示す。

[*28] 定常解のばあいは $R(t,x) = 0$ となる。

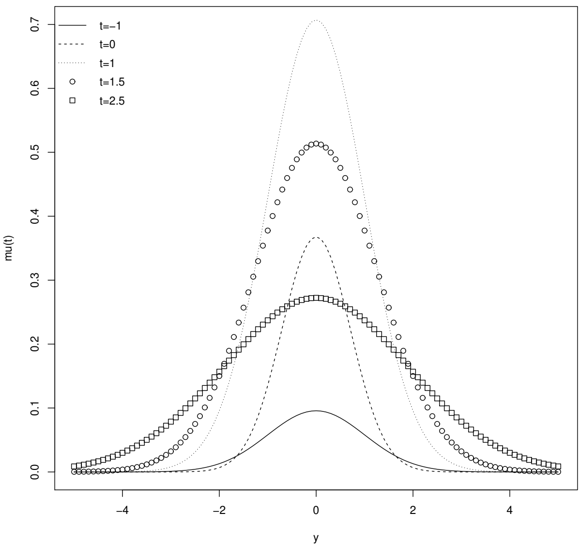


図 5.4　自由粒子の $\mu(t, x_0, y_0)$

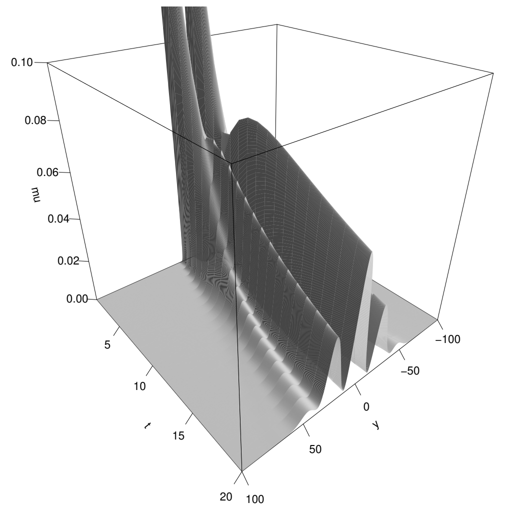


図 5.5　二重スリット実験における μ^*

図 5.4 は空間上の一点 (x_0, y_0) での μ の時間変化を横軸を y_0 として示したものである。空間の尺度は $t = 1$ で x_0 を通過するように調整している。これによれば、$t < 1$ の間、確率密度は $y_0 = 0$ を中心に正規分布の密度関数の形をとりながら立ち上がり、$t = 1$ で最高点に達し、$t > 1$ でばらつきを大きくしながら x 軸に向って倒れこんでいく[*29]。

5.3.5　ランダム運動の具体例 (3)：二重スリット実験

自由粒子の運動を基礎としていわゆる「二重スリット実験」[*30]の説明ができる。スリットの位置を $\boldsymbol{x}_0^{\pm} = (x_0, \pm y_0)$、スクリーンの位置を $\boldsymbol{x}_1 = (x_1, y_1)$ とする $(x_1 > x_0)$。$\boldsymbol{\kappa} = (\kappa_x, 0)$ の方向のドリフトをもち原点を出発した粒子は $t = 0$ で $\boldsymbol{x}_0^{+}$、$\boldsymbol{x}_0^{-}$ のいずれかのスリットを通過して、$t = t$ でより遠方のスクリーン上の点 $\boldsymbol{x}_1$ に到達する。

スリットと x 軸との距離が等距離であるので、二つのスリットを通過する確率は等しい。したがって、絡み合った発展関数の混合割合は $\alpha = \beta = 1/\sqrt{2}$ でなければならない。このとき点 $\boldsymbol{x}_1$ 上の粒子の分布密度 μ^* は次のように計算される ((R^+, S^+)、(R^-, S^-) はそれぞれのスリットを通過した粒子の生成関数)。

$$
\begin{cases}
R^{\pm} &= \dfrac{1}{4}\ln|A|^2 - \dfrac{|A|^2}{2a^2}|\boldsymbol{x}_1 - \boldsymbol{x}_0^{\pm} - \boldsymbol{\kappa}\sigma^2 t|^2 + const. \\
S^{\pm} &= -\dfrac{i}{4}(\ln A - \ln \overline{A}) \\
& \quad + \dfrac{|A|^2}{2}\left(\dfrac{\sigma^2 t}{4a^4}|\boldsymbol{x}_1 - \boldsymbol{x}_0^{\pm}|^2\right. \\
& \quad \left. +2\boldsymbol{\kappa}\cdot(\boldsymbol{x}_1 - \boldsymbol{x}_0^{\pm}) - |\boldsymbol{\kappa}|^2\sigma^2 t\right)
\end{cases}
$$

$$
\begin{aligned}
\mu^*(t, \boldsymbol{x}_1) &= |\alpha\psi^+ + \beta\psi^-|^2 \\
&= \frac{1}{2}\left(e^{2R^+} + e^{2R^-}\right. \\
&\quad \left. + e^{R^+ + R^-}\cos\left(S^+ - S^-\right)\right) \\
&= e^{R_0}\left\{\frac{e^r + e^{-r}}{2} + \cos(S^+ - S^-)\right\}
\end{aligned}
$$

[*29] この図で $t < -1$ よりも過去を考察すると、原点からどのような遠方の (x_0, y_0) をとったとしても、到達可能性を示す確率がゼロではないことがわかる。これは「薄気味の悪い遠隔作用」であり明らかに特殊相対性理論に反する。この事態が生じているのは、既に初期状態において正規分布の密度関数を採用しているからであり、シュレディンガー方程式がこれを許容しているからにすぎない。なお、これは「量子もつれ」とはさしあたり何の関係もないことに注意する。

[*30] 二重スリット実験 (Double-slit experiment) とは、光の波動性を示したヤング (Thomas Young:1773-1829) の実験を、光の代わりに電子などに置き換えたものであり、電子の「粒子性と波動性の二重性」を実証するものとして言及されることが多い。1961 年の C.Jönsson、1974 年の P.G.Merli ら、1989 年の外村彰の実験などが知られる。

$$
\begin{cases}
R_0 &= \frac{1}{2}\ln|A|^2 - \frac{|A|^2}{2a^2}\left\{(x_1 - x_0 - \sigma^2 t\kappa_x)^2 \right. \\
& \left. \quad + y_1^2 + y_0^2\right\} + const. \\
r &= \frac{|A|^2}{a^2} y_0 y_1 \\
S^+ - S^- &= \frac{|A|^2\sigma^2 t}{4a^4}\boldsymbol{x}_1 \cdot (\boldsymbol{x}_0^- - \boldsymbol{x}_0^+) \\
&= -\frac{|A|^2\sigma^2 t}{2a^4} y_0 y_1
\end{cases}
$$

図 5.5 はスクリーン上の μ^* の時間変化を $t = 1, \cdots, 20$ まで追跡したものである[*31]。$t = 1$ 時点でスリット位置に対応する二つのモードが立ち上がっており、これらが時間経過に伴って低くなっていく。その間と両端に新たなモードが出現し、全体として 5 本の縞模様が浮かび上がる。

長澤の言うように、この結果は粒子と波動の二重性を示す「波の干渉縞」ではない。粒子は単一のサンプル・パスに沿ってランダム運動をしており、そのうち二重スリットを通過するものだけをカウントした統計的結果として紋様があらわれているにすぎないからである。

5.4 小括

本節の主要な結論は以下のとおりである。

- 古典的確率論から測度論的確率論 (Kolmogorov の公理体系) への移行は確率過程を取り扱うためにこそ必要であった。まずサンプル空間 Ω があり、Ω 上の確率分布 $P(\omega)$、Ω 上の確率変数 $x(\omega)$、ω にかんする条件として事象 E がある。このような取り扱いは古典的確率論とはまったく異なる。Ω が有限集合、可算無限集合ならば古典的確率論で十分であったが、Ω として空間上の点、粒子の運動の軌跡などをとったばあい、これらは非可算の無限集合となり、とりわけ積分 (リーマン積分) の困難などが現れる。そのために、関数の定義域ではなく値域の分割による新しい積分の方法 (ルベーグ積分) が開発されねばならなかったし、これらを基礎とする測度論 (measure theory) 的な確率論が要求された。ここで確率過程とは「時とともに変動する偶然量を抽象化した概念」のことである。
- 遷移確率 (条件付き確率) によって与えられる確率過程はマルコフ過程と呼ばれる。ブラウン運動はマルコフ過程の一つの例である。マルコフ過程の遷移確率が、(拡散方程式をその特殊例として含む) 放物型の偏微分方程式で特徴づけられるという事実は「コルモゴロフの定理」として知られている。
- コルモゴロフの偏微分方程式は、微粒子などの位置の存在確率の時間発展を記述するが、そのサンプルを直接にあたえるものではない。しかしブラウン運動の軌道 B_t が描けるということは、同様の確率過程の軌道についても何らかの記述が可能なのではないかとの予感をもたらす。伊藤清は、今日「伊藤の公式」として知られる、ブラウン運動にかんする一種の変数変換の公式をしめすことで上の予想が真実であることを確認した。
- ブラウン運動の理論はその後の統計物理学の発展の確固たる基礎をなし、現象論的な熱力学に分子運動論という実体を与えることとなった。他方、熱輻射理論、光電効果の発見などから光量子仮説が提唱され、さらに原子中の電子運動の考察 (原子模型) から量子力学が構築された。前者の統計物理学と後者の量子力学は、その類似性にもかかわらず、またそれぞれの理論の実社会での成功にもかかわらず、直接の関係が今日においても明瞭になっているとは言い難い。長澤 (2015) はコルモゴロフの偏微分方程式を外力のない条件下での粒子の運動方程式と解釈することを通じて両理論の直接のつながりを明らかにすることに成功した。
- 長澤の理論において、1) 生成関数 (R, S) を双子の運動方程式の解として求めることは現実的にはほとんど不可能である。(R, S) はシュレディンガー方程式の解から求めなければならない。2) 生成関数 (R, S) から発展の流れドリフト a、$\widehat{a}$ を経て、確率過程 X_t を求めることは容易である。3) シュレディンガー方程式のみでは、a、$\widehat{a}$ が何

[*31] スリット通過時点を $t = 0$ として、$t = 1$ でスクリーンに到達するものとしている。

を意味するのか解釈が困難である。とりわけ複素発展関数[*32]の重ね合わせ (「量子もつれ」) をめぐって混乱した議論が行われてきた。長澤の言うように、二重スリット実験の結果は粒子と波動の二重性を示す「波の干渉縞」ではない。粒子は単一のサンプル・パスに沿ってランダム運動をしており、そのうち二重スリットを通過するものだけをカウントした統計的結果として紋様があらわれているにすぎないからである。

- 長澤は外力のないランダム運動について定常解を念頭において考察しており、この状態が特殊なものであるとの認識をもっていない。しかし、この定常状態は、初期状態において位置が空間上の一点 (すなわち退化分布) に収束し、確定していること (すなわち零点振動の欠如) を認めながらも、その他の経路においてはランダム運動を認める、という矛盾したものである。われわれは、定常状態を現実には決して到達しえない特殊な極限状態であると考えなければならない。

[*32] 複素発展関数とは、シュレディンガー方程式の解のことであり、「波動関数」とも呼ばれている。長澤はシュレディンガー方程式は波を表現するものではないのだから、その解を「波動関数」と呼ぶのはおかしいとしている。この指摘は正しい。

第 6 章

確率学説にたいするイデオロギー批判

6.1　イデオロギー批判の方法

確率学説はイデオロギーのひとつであり、歴史のなかで人々が確率について抱いてきた観念のかたまりである。だからわれわれは個々の学説を、イデオロギー批判の方法にしたがって評価するべきである。個々の学説はそれら特有の歴史的・社会的な被規定性を帯びており、超歴史的な真理というものは存在しない。仮にわれわれが「真理性」の高みから個々の学説を断罪するという態度をとったとしたら、それはわれわれ自身が自己の社会的被規定性を忘れてしまっている証左にほかならない。むしろ、上の被規定性をもつ各学説は、その時代の人類の到達地点と可能性を示すものとして敬意が払われるべきである。

第 2 章で言及したように、イデオロギーとしての (とりわけ自然を対象とする) 学説は、〈技術的地盤〉、〈自然の構成〉、〈思惟様式〉の三つの契機・モメントをもつ。われわれは、個々の学説がこれら三つの契機からどのように組み立てられているかを注意深く吟味しなければならないし、そればかりでなくその学説を提示した諸個人がどのように創意を働かせたかをも明らかにしなければならない。各学説を論じた諸個人が、その時代・社会の制約を帯びた〈歴史的人間〉なのだとしても、その創意すべてが社会体制に還元されてしまうものではなく、そこに個性の存在を認めないわけにはいかないのである。

確率学説の批判に際して、われわれ自身はもちろん実践的唯物論の見地に立つものとしてその任にあたる。しかし、それは政治的・党派的な、いわゆる「為にする」批評であってはならない。われわれの見地それ自体がどのような発達過程を経て形作られてきたかをこれら学説の連鎖のなかに見いだし、唯物論という見地の歴史的正統性を証明するためにこそ、こうした批評が行われねばならない。

実践の物質性を自己の立脚点となす唯物論者からすれば、確率が主観的側面と客観的側面の二面性をもつこと、前者が後者の反映であることは自明のことわりのように思える[*1]。しかしすべての人間がそのように考えてきたわけではないし、また今日でもそうではないことを肝に銘じるべきである。むしろこの事態こそ解かれなければならない謎であり、本章もこの謎の解明に充てられる。

6.2　統計記録

われわれはまず統計記録 (データ) と確率学説を区別することから始めよう。前者は後者の成立を促した〈技術的基盤〉のひとつであるにせよ、後者そのものではない。また統計記録は既に古代社会に存在したが、そのことが確率学説の発達にただちにつながったわけではなかった。

共同体に「財産」とみなされるものが出現するやいなや、したがって労働が集約され、余剰生産物が出現するやいなや、それを記録する必要性が生じたと思われる。

[*1] このような理解が可能となるのは、われわれの哲学が存在論と認識論の根底的な区別をヘーゲル弁証法 (すなわち、大陸合理論とイギリス経験論の遺産) から受け継いでいるからである。この区別を見失えば、われわれは確率の「ヤヌス的二面性」の混乱に容易に絡めとられるであろう。

ラオ (2010) は最初期の統計記録について「数を数える技術が完全なものになる以前・・・家畜や他の財産の帳簿をつけるために行われた」と推測している。そしてその必要は、古代の王国 (夏・周などの古代中国王朝、古代インド王朝、旧約聖書の頃のユダヤ王国、ローマ帝国など) では既に明白なものとなっていた。ただしこの〈技術的地盤〉がただちには確率学説というイデオロギーには結びつかなかった。その当時の〈思惟様式〉がそれを許さなかったからである。

確率と統計の結びつきが今日知られるような姿を現したのは近代国家 (したがって政治＝経済学) の出現とほぼ同時期、つまり重商主義者、W.Petty[*2]が「租税効能論」(1662)、「政治算術」(1690) を刊行し、その友人である J.Graunt[*3]が「ロンドン死亡表」を著して以来のことである[*4]。Petty と同じく重商主義者と目されるプロイセン王国の G.Achenwall[*5]は、statistik(統計学) という新語をつくり出した。この言葉はラオによればラテン語の status(国家) に由来し、その後広くヨーロッパ中に流布することになった (ラオは 1885 年の国際統計協会の設立に至るまでのこの間のいきさつを簡潔に説明している)。

6.3 確率の二面性

確率の二面性 (可能性の度合いとしての/信頼性の度合いとしての) は古くから知られていた[*6]。しかし、両者がいかなる関係にあるのか、またどちらが重視されるべきかは、その時代ごとの〈思惟様式〉や、その当時に解決を迫られた理論的な課題に対応して変化した。ここでは簡単な歴史的素描を試みる。

6.3.1 古代

安藤 (1997) によれば、古代ギリシャ語にすでに確率のそれぞれの側面を言い表す言葉が用意されていた。確率の客観的側面に対応して「テュケー」($\tau\acute{\upsilon}\kappa\eta$:偶運)、「アウトマトン」($\alpha\acute{\upsilon}\tau\acute{o}\mu\breve{\alpha}\tau o\nu$:自己偶発) という言葉があり、主観的側面に対応するものとして「エイコス」($\epsilon\acute{\iota}\kappa\acute{o}\varsigma$:蓋然性)、「エンドクソス」($\acute{\epsilon}\nu\delta o\xi o\varsigma$:通念) がある。これらはいずれも Aristotle[*7]によって論じられた。

Aristotle は上のチュケーについてその存在を認めていたものの、科学の対象としなかったと安藤は論じている。この理由は、テュケーがもし本質的原因というものをもつならば、それは必然ということになってしまい、その語の本来の意味に反するからである[*8]。Aristotle の論理学はこの矛盾に対処する術をもたず、それ以上の考究は不可能であった。この矛盾を適切に取り扱うには何らかの弁証法論理を必要とする。この「禁令」はやがて中世キリスト教神学に持ち込まれ、確率論成立への最大の障害になる。

他の学派はかならずしも Aristotle と同じようには考えなかったが、やはり確率計算を生み出すことはできなかった。Epikouros の哲学[*9]は運動学として偶然の存在を論じ、Lucretius はそれを継承し宣伝に努めた。しか

*2 Sir William Petty(1623–1687): イングランドの医師、オクスフォード大学解剖学教授、アイルランド派遣軍軍医などを務める。〈価値法則〉は彼の「租税効能論」の中ではじめて著述される

*3 John Graunt(1620–1674): ロンドンの商人。「政治算術」の発案者。W.Petty とともに Francis Bacon の影響を受ける。

*4 この頃死亡統計に注目が集まった理由は、ロンドンにおけるペストの大流行 (1665) にその一端がある。黒死病の蔓延はケンブリッジ大学を休校に追い込み、その間疎開を強いられた Newton はそこで彼の重力理論を作り上げた。また彼の友人の E.Halley は死亡の規則性を見いだし、その後の生命保険計算の発展に寄与したとされる。

ただし、ギリース (2004) によれば、これらの努力はただちに現実社会での成功を見なかったようである。たとえば de Moivre「生涯年金の論考」(1725) などそのころに数学者の著したマニュアル類には読者の便宜を周到に図った豊富な数表が付されたが、当時の保険数理士には活用してもらえなかった。というのも当時の保険数理士がたんなる事務員以上のものではなかったからである。事態が変わるのは早くとも 1762 年のエクイタブル生命 (The Equitable Life Assurance Society) の設立を待たねばならなかった。

*5 Gottfried Achenwall(1719–1772)：ゲッティンゲン大学哲学教授、「ヨーロッパ諸国の構造概要」で知られる。

*6 確率の二面性は、I.Hacking(1975)、D.Garber and S.Zabel(1979)、L.Daston(1988) などにより言及されている。安藤は、この二面性が確率の古典的な定義の出現にたいする大きな障害になったと論じている。

*7 Aristoteles(BC384-BC322)

*8 大災害の原因をチュケー ($T\upsilon\xi\eta$,Tyche)、すなわち運命の女神の意思としてしまうことにたとえられよう。

*9 彼は Democritus と Plato の自然哲学の統合を志向した。とりわけ原子の「偏奇」理論が、Democritus の原子論との決定的な違いである。

*10 Alexander of Aphrodisias(生年不詳,AD200 頃)。彼は Aristotle の注釈者として名高いが、「運命について」は彼独自の論考であり、ストア派の決定論を反駁して人間の自由意志を擁護したものであった。

し、安藤は「原因がわからずに存在する偏奇」はかえって確率を定義する試みを遠ざけてしまう、と論じている。Aphrodisias の Alexander[*10]は「運命について」で因果関係を論じるなかで古典的確率論の定義に極めて近いところまで至ったが、それ以上には進むことができなかった。安藤はこれに二つの理由があったと推察している。1) 頻度の安定性を直観的に把握させるゲームに対する哲学者の軽蔑[*11]、2) 記数法とくに分数概念の不備、である。

確率の主観的側面 (信頼性の度合い) については、客観的な側面と切り離されて論理学・修辞学として論じられた。ただし、この方面からも確率計算の解明には至らなかった。排反事象についてはよく知られていたが、個別の事象が確実に真であるか確実に偽であるばあいのみが取り扱われた。蓋然性には心理状態が影響するものであるから、そこから確率値の計算に入ることは困難であったし、またその必要性も認められなかった。

安藤によれば、偶然事象について古代でもっとも現代の考え方に近づいた人は Cicero[*12]であった。Cicero の「占いについて」は、1) サイコロ投げで稀な事象が起こる外的な原因を探すことはできない、2) 勝負を長く続ければ稀な事象でもいずれは起きる、3) 予言・占いを神意 (必然) と解釈すべきでなく、そのような態度は意志薄弱のしるしとみなすべき、と論じている。

6.3.2 中世

中世社会[*13]はローマ帝国の衰退とゲルマン諸族による領土の分割統治を経て確立されていった。これに先立ってキリスト教がローマ帝国の国教となり、中心地ローマにおいて神学体系が発展し、多くの知識人層を引き付けた。その中には Hippo の聖 Augustinus[*14]もいた。そうした教義はローマ帝国の崩壊後もキリスト教会の権威のうちに長く中世世界を支配した。

安藤は Augustinus の影響について次のように論じている。Aristotle が論理的な困難に直面して偶運を科学の対象から除外するにとどまったのにたいして、Augustinus は偶運の存在そのものを否定し、確率の客観的側面の追及を根絶やしにしてしまった。

> ・・・時の経過と供にキリスト教の伝道が広がるにつれ、アウグスティヌスの「すべては神の摂理に従う」(「神国論」V 巻 11 章) ので、「random なものは何もないし、チャンス (偶運) なるものは存在しない」(「八十三問題集」) という考え方は社会に浸透した。そして賭博は異教徒のものとされた。

> ・・・ローマ帝国においてキリスト教が勢力を得るとともに、アウグスティヌスらの神学により、キケロ風の偶然現象に対する感覚は急速にしぼんでいく。・・・アウグスティヌスの神学は中世ヨーロッパに計り知れぬ影響を及ぼした。それは千年以上にわたりヨーロッパの最も重要で絶対的な神学体系だった。

他方、確率の主観的側面は、客観的側面と対照的にキリスト教会において「決議論」(causistry) として隆盛をみた。決議論とは、「個別の倫理問題を解決する方法を書き記したもの」(安藤) のことである。決議論では、まず律法などにより定められた所与の一般的道徳規則があり、これを典型的な事例 (cases) について、いつどのように適用されるべきかが論じられる。次いでそのほかの事例にたいしても類推を用いて考察が行われるが、その際に典型的な事例との類似点と相違点が整理される[*15]。ただし、後のイエズス会に投げかけられたように、詭弁に堕する傾向を多分に秘めていた。

[*11] 安藤は Plato がその当時流行していたアストラガルスのゲームを「子供の遊び」と表現したことを紹介している。ギリースは、F.N.David(1962) の説に倣い、アストラガルスが羊のくるぶしの骨をほとんどそのまま使った不細工な道具であったことから、確率計算がその当時生まれなかったとしている。しかし、逆に言えば精密な加工を施す必要がなかったことを疑問に思うべきである。当時は精密な加工技術がなかったわけではなく、実際精密なサイコロも出土している。勝敗にこだわる大人のギャンブラーが存在しなかったのはなぜか (賭博の禁止と商業・利殖の制限)、が問われるべきであった。

[*12] Marcus Tullius Cicero(BC106-BC43)

[*13] 西欧において古代と中世を分かつものは、ゴート族によるローマ陥落 (410)、「西ローマ」正帝の廃位 (480)、フランク王国の建国 (508 以降) の頃である。

[*14] Aurelius Augustinus(354-430)。世界史を地の国 (世俗国家) と神の国 (キリスト教共同体) とに弁別して叙述した「神の国」の著述で知られる。

[*15] 決議論はまず〈一般〉-〈特殊〉-〈個別〉の判断の関係を論じたものとみることができる。加えて決議論の主眼は、義務・戒律と人間的自由との関係を具体的事案に即して議論するためのものでもあった。それゆえ、われわれは決議論を弁証法 (〈本質〉-〈実体〉-〈現象〉、〈可能性〉-〈現実性〉-〈必然性〉) の萌芽的なものと見ることができよう。

[*16] Charlemagne(742-814) はフランク王、Charles the Great とも呼ばれる。西欧世界の統一者として知られる。

Charlemagne[*16]の戴冠 (800) により、皇帝権 (王権) は神によって忖度されたものとの考えが広まった。カロリング朝宮廷に知識人が呼び込まれたが、特に有名な人物としては York の Alcuim(735?-804) がいる。彼は信徒の教育にラテン語や数論を取り入れた。この頃の文化的雰囲気はのちに「カロリング朝ルネサンス」と呼ばれ、後の Aristotle 哲学受容など文芸復興を準備するものとなった。

ヨーロッパ中世社会において商業は一般に衰退し、地中海 (東方) 貿易はイスラム商人の手に握られた。しかし、ビザンツ (東ローマ) 帝国がセルジューク朝トルコにおびやかされたことをきっかけとして、ローマ教皇 Urbanus II は十字軍を呼びかけ (クレルモン公会議:1095)、これ以降数度にわたる遠征がなされた。十字軍は東方貿易をイタリア諸都市の商人の手に取り戻し、同時に古代ギリシャ・ローマの文化的遺産が西欧にもたらされた。これはやがて「12 世紀ルネサンス」、また 14 世紀の「イタリア・ルネサンス」の文化運動につながることとなる。キリスト教神学においても、Aristotle 哲学が再発見され神学体系との統合が試みられた[*17]。ことに Thomas Aquinas はイスラーム学者と Aristotle 哲学の正統な継承者の地位を争うため、教会から異端を疑われるほどに Avicenna[*18]や Averroes[*19]などの文献を精力的に研究した。彼の活動が偶然性の哲学をキリスト教神学の内部で議論できる土壌をつくりあげた。

6.3.3 近世

確率の客観的側面を追及することはなしくずしに解禁された。たとえば 1638 年には欧州最初の公設カジノ (”Il Ridotto”) が設置されている。ヴェネツィア大評議会がこの決定を行ったのは、既に私設カジノが多数設置されており、もはやギャンブルを禁じるよりはそれをコントロールすることを重視したからと伝えられている。Caldano(1576 の死の以前)、Galilei(1622) も賭けの問題を取り扱っており、ギャンブラーにとっては確率を理解する切実な動機があった。

Pascal の思想

Pascal-Fermat の往復書簡 (1654) について、また Pascal の思想をどのようにみるべきかについて、ここでは事実関係を積み上げげることを糸口に答えてみたい。

- Blaise Pascal は 1623 年クレルモンに生まれた。そこは奇しくも十字軍の呼びかけが行われた公会議の開催地である。後年、彼の論敵となるイエズス会 (1540 年設立) は、パリにクレルモン学院という名のコレージュを 1563 年に建てている。
- Blaise の父 Etienne は徴税官であり、法服貴族 (Nobless de robe) の一人であった。自然科学をたしなみ、1631 年に一家でパリに転居しそこでサロンを運営したらしい。そのころ Blaise は円錐曲線論などで早熟の天才を示し、Descartes[*20]らを驚嘆させた。一家は 1640 年にルーアンに移住したが、これには財産上の理由[*21]があった。
- Blaise たちは父の捻挫を治療した医師たちを介して、一家をあげて Jansen 主義に入信した (1646)。Jansen 主義は、Jansen の盟友である”Abbé de Saint-Cyran”がパリに持ち込んで 1640 年に出版した著書「Augustinus」を通じてフランスの貴族階級のあいだで流行していた。Jansen 主義者はポール・ロワイヤル修道院を拠点に活動し、Saint-Cyran の死去 (1643) 以降は Antoine Arnauld[*22]によって指導されていた。Jansen 主義は (彼らの自己認識とは食い違うが) イエズス会からは Calvin 主義の一変種とみなされ迫害を受けていた。
- Blaise は 1647 年に Torricelli の実験を再現する

[*17] Albertus Magnus(1193-1280)、Thomas Aquinas(1225-1274)、Ockham の William(1288-1348) など。

[*18] ibn Sīnā(980-1037)。「第二の Aristotle」とも称される哲学者、医師。

[*19] ibn rušd(1126-1198)。Aristotle の注釈者、医師。

[*20] René Descartes(1596-1650)

[*21] Etiennne が官位を売ってこれを公債に投じたが、30 年戦争遂行のため Richelieu が公債を債務不履行としたため財産が激減し、Richelieu を批判したことがわざわいして一時期パリに居られなくなった。

[*22] Antoine Arnauld:1612-1694。神学者。「ポール・ロワイヤル文法」(1660)、「ポール・ロワイヤル論理学」(1662) の共著者の一人。Descartes 主義者と目される。ギリースは、「ポール・ロワイヤル論理学」の第 4 章が不確実性のもとでの推論と確率についての議論に充てられている、と述べている。

ことで、真空の存在に反対する Aristotle の哲学を (したがってまた Descartes の哲学を) 乗り越える論拠を手に入れた。

- Etienne の死去を契機に、妹 Jacqueline がポール・ロワイヤル修道院に隠遁 (1651)。Blaise は反対したが妹を翻意させることができなかった。
- 同じころ、Blaise はロアネーズ公 (Artus Gouffier)、Chevalier de Méré[*23]らと親交を結ぶ。Méré がサロンに持ち込んだ問題が Pascal と Fermat[*24]の往復書簡 (1654) の主題になった。
- 1654 に Blaise は「恩寵の火」の経験を機に回心し、1655 からポール・ロワイヤル・デ・シャンでの隠棲生活に入る。1656〜1657 に「田舎の友への手紙（プロバンシアル）」と題する 18 の偽名による文書を公表した[*25]。内容は弛緩したイエズス会の道徳律 (また決議論) を批判するものであり[*26]、当時の読者層から好評をもって受け入れられ、また後にはフランス語文章の見本とされた。
- 1662 年に Blaise は死去、ポール・ロワイヤル修道院は Clemens XI により 1708 年に廃止を布告され、1710 年に取り壊された。

これら一連の事実をより広い歴史的視角から見ると次のようになるだろう。まずプロテスタンティズムにたいするカトリシズムの防波堤としてのイエズス会があり、プロテスタントの側でも対抗するイデオロギー闘争が組織化された。これら闘争は中等教育のための学校運営というかたちをとっていた。一方はイエズス会によるコレージュであり、他方はポール・ロワイヤルのような修道院である。

イデオロギー闘争の焦点は、(相対的な意味で) 他国より文化的に遅れたフランスにおいて、封建的な伝統文化とブルジョアジーの新興文化のいずれに加担するかという問題として提起された。イエズス会は教皇の権威をより重視し、異端を摘発することを目的とした。ポール・ロワイヤル派はより世俗的な権威 (皇帝など) に依拠し、また異端の疑いを避けるために Hippo の Augustinus など古典に範を求めながらも、イタリアやイギリス、北方諸都市の先進的な文化を取り入れることを重視した。

こうしたイデオロギー闘争は、古い封建社会から新しい資本主義社会への過渡期を決する階級闘争の一環でもあり、この過程から近代国家が形作られた。フランスにおいては当時の Louis XIII 治下のブルボン王朝が絶対王政として統治されていた。「絶対王政」とは、その名に反して絶対的な権威をもたず、諸階級の勢力均衡のもとで維持されるものだった。絶対主義は封建社会の終末に対応する政治体制であったが、Etienne が身をもって示したように経済の主力は貨幣経済に移っており、身分は金で買われ、30 年戦争[*27]は公債でまかなわれた。

Blaise は Descartes の影響下から出発し、イタリアの実験精神に学びつつ独自の哲学を構築した。当初から幾何学と無限小解析に天才を発揮させつつも、de Méré、Fermat との交流から組み合わせ数学の思考方法と客観

[*23] 本名 Antoine Gomboud(1607-1684)。作家。貴族ではなく、作品に登場させた自己の分身を「Chevalier(騎士)」と名付けた。それが後に本人にたいしても「Méré で教育を受けた」「Chevalier」と呼ばれることになった。

[*24] Pierre de Fermat(1607-1665) は法律家であり、オルレアン大学で法学の学位をえた。当時トゥールーズ議会の勅選委員 (終身)。

[*25] 石川 (2009) に偽名にかんする興味深い考察がある。

[*26] 安藤は「プロバンシアル」(第六の手紙) をとりあげ、イエズス会が会士のいい加減な行動を正当化する蓋然説 (probabilisme) への Pascal の批判を紹介している。それによれば、蓋然説の例として、「ミサをあげるように頼まれて金を受け取った司祭がさらに別の人からもお金をもらって頼まれたときにミサは一回でよいのか」という問題が例示される。蓋然説によれば、肯定的回答、否定的回答のいずれもある程度の蓋然性をもつので、どちらに従っても良心を痛める必要はない、とされる。

「蓋然的命題」の考え自体は Aristotle の「エンドクソス」を源流としている。Ockham の William によれば「すべての人々によって真であるか、大多数の人々にとって真であるか、あるいは最も賢明な人々にとって真であるか、いずれかである命題」のことである。

[*27] ボヘミア (ベーメン) でのプロテスタントの反乱を起点とし、1618 年から 1648 年に戦われた戦争。当初は神聖ローマ帝国内での局所的な争乱であったものが、周辺諸国を巻き込み、ヨーロッパ全土で戦われた。ヴェストファーレン条約により終結を見た。

[*28] ギリースは Pascal から Fermat にあてた第一の手紙を引用しているが、その中で Pascal は de Méré のことを「非常にできる人」ではあるが「残念ながら幾何学者ではない」と評している。そして de Méré による (サイコロの出る目の確率にかんする) 出題の動機が「数の理論の欠陥」を彼が証明したがっていたことにあったと Fermat に伝えている。Pascal はおそらく de Méré が「幾何学」に詳しくないためにそのような間違いを犯したと信じた。しかし、少なくとも経験的な確率の見積もりについては de Méré は間違っておらず、Fermat の (焼失した) 返信では de Méré の直観を裏付ける正しい計算が与えられている (らしい)。de Méré の技りょうについて、ギリースは David(1962) からの引用で次のように説明している。「シュバリエ・ド・メレはとても根気強い賭博者だったので、0.4914 と 0.5 という確率を経験的に区別することができた。ちょうどガリレオにアドヴァイスを求めた賭博者が 0.0108 の差を見たように、0.0086 の差を見たの

的確率の考え方を学び、これを自家薬籠中のものとした[*28]。「プロバンシアル」ではいい加減な決議論をもてあそぶイエズス会を皮肉り、「パンセ」では彼らにたいして客観的確率論を「神への賭け」(今日の用語でいえば、期待効用理論) としてたたきつけた。このとき、Pascal の頭脳内で確率の客観的側面と主観的側面とが並立したのだと言えよう。

これ以降、確率の二面性は機械的唯物論の欠陥のゆえに後景に追いやられることになる。存在、即、認識をあらわす鏡的反映論 (模写説、素朴実在論)[*29]が幅を利かせた結果、確率の客観的側面のみが当面の追及対象となり、主観的側面の方はあとから自然についてくるものとみなされた。ただし、確率の主観的側面は一時的に見えなくなっただけで、消え去ったわけではない。それは Bayes の規則の帰納法論理への応用[*30]、サンクト・ペテルブルグの逆理[*31]のような謎として現代に至るまで引き継がれている。

古典的確率論の形成

ギリースは Laplace(1814) の与えた古典的確率の定義が彼自身のオリジナルのものではなく、先人たちの成果を集約したものであったことを指摘している[*32]。また安藤は最初にその定義を与えた者が de Moivre「運の測定について」(1771) であったと述べている。それ以前には、Jakob Bernoulli にせよ、Christiaan Huygens にせよ、確率を明示的に定義することは避けて確率計算を行っていた。またギリースは古典確率論の実質的な内容が Bernoulli の「推測論」(1713) に見られること、またそこに展開されたアイディアが Gottfried Leibniz との書簡のなかで固められていったことを指摘している。

安藤は、Jakob Bernoulli が大数の弱法則を論証しようとした動機 (また彼が確率の明示的な定義を与えなかった理由) は、「同様に確からしい」と先験的に判断する論理が相当にあやふやであったため、これを事後的な「大量観察」に置き換えようとしたためだった、と論じている。そして大数の弱法則については苦心してその証明に成功したものの、肝心の確率の定義には復帰できなかった。これにたいして、de Moivre は最初から確率研究の対象を偶然ゲームに絞ったために、確率の定義を明示し得たし、またそれを出発点にして大数の法則をさらに精密にした中心極限定理の証明に至ることもできた[*33]。

である。」

[*29] Plato のイデア論を源流とし、John Locke(1632-1704) の「タブラ・ラサ」の議論を典型とする認識の理解。K.Marx の実践的唯物論によって乗り越えられたはずであったが、現実にはそうは理解されていない。

[*30] ギリースは R.Price によって 1763 年に出版された T.Bayes の論考が、D.Hume の帰納法への懐疑にたいする回答であった、と述べている。

[*31] 1738 年に Daniel Bernoulli(1700-1782;Jakob の甥) により公表された。なお、この逆理にたいしては、のちに Karl Menger(1902-1985) によって効用理論による解釈 (この解決には疑問の余地が大きい) が与えられた。Karl は限界効用理論創始者の一人 Carl Menger(1840-1921) の息子で、ウィーン学団にも所属していた。

[*32] ギリースは、Laplace の著作が有名である理由は彼が先人たちとは異なり、ラテン語ではなくフランス語で著述したから、としている。

[*33] この成果は、最初に分数による定義から始めて、後から実数値への完備化を図ったものと解釈でき、より現代的な確率理解の可能性を内包するものであったと言える。

6.4　現代

現代[*34]において確率をめぐる思想は、古典確率論の成立を受けて出現したさまざまの技術領域と科学理論によって発展の素材を提供されながらも、その時々の「科学の危機」(マッハ主義をめぐる、集合論と数学の基礎をめぐる、量子力学をめぐる) に応じて揺れ動いてきた。加えて、金融資本の成立と資本主義の帝国主義段階への変転が確率の主観的側面の問題への取り組みを避けられないものにした。確率の二面性はそれまで背景に隠されていただけで、けっして無くなったものではなかったが、ここではことあるごとに表層に現れ、確率にかかわる難問を提起した。それらの難問にたいするわれわれ現代人の応答が現代の確率思想を形成した。

ギリースは、現代の確率思想を 1) 論理説 (Keynes)、2) 主観説 (de Finetti)、3) 頻度説 (R.von Mises)、4) 傾向説 (K.Popper) に分類し、前二者が確率の主観的側面にかかわり、後二者が客観的側面にかかわるものとしている[*35]。その上で、ギリースは確率の多元主義的な見方を支持し、社会科学においては主観説が、自然科学においては傾向説がより親和性が高いと結論している。ギリースの結論には賛同できないが[*36]、この分類と特徴づけについては大いに利用できる。

古典的確率論の崩壊の起点は、Jakob Bernoulli によって名付けられた「不充足理由律」[*37](Keynes の用語では「無差別の原理」) である。それは以下のように表現される (ギリースより)。

> 無差別の原理は、もしいくつかの選択肢のうち、他よりもある一つだとする理由が知られていなければ、そのような知識の状態に対応して。これらの選択肢のそれぞれは同じように確からしいと主張できることである。
>
> Keynes(1921,p.42)

古典的確率の定義はこの原理に立脚するが、この原理はさまざまな逆理を生み出し、同じ状況にたいしていくらでも異なる確率値を与えるものだった。Jakob Bernoulli は前述のようにこの原理を避けるために確率を無定義のままにしておき、大量観察に置き換えようとした。Keynes の論理説は「不充足理由律」を擁護し、確率論を論理学 (「蓋然説」) に純化しようとする試みであったし、von Mises の頻度説は Jakob Bernoulli の試みを引継ぎ、完遂させようとするものだった。このようなことが必要になった理由は、ひとえに確率論の対象を偶然ゲームに限定した de Moivre による応急措置がそれでは収まらなくなり、狭い意味における統計学 (Achenwall)、(心理学を含む) 生物統計学 (”Biometrika”)、物理学の新分野 (流体力学、熱力学) などと確率論の適用範囲が次々に拡がったことにある。

以下では、ギリースの分類にしたがって各学説を紹介し、その上でそれぞれについて論評したい。

[*34] われわれがここで「現代」と言っているのは、われわれの生きている今という時間とそれ以前の時間を明確に区切った上で、より新しい時間について便宜的に与えた名称である。どのような事件を区切りとするのかはわれわれの自己認識に依存する。プロレタリアートにとってそれは言うまでもなく、K.Marx によって実践的唯物論 (「フォイエルバッハ・テーゼ」) が発見された 1845 年のことである。この事件以降、階級社会のつらなりである人類前史と、そこからの脱却をめざす運動が現に存在する過渡期社会とを明確に区別する道が拓かれた。

[*35] ギリース自身はそれぞれ「認識論的解釈」、「客観的解釈」という用語を使っている。彼がこうした用語法を使った理由は次のとおりである。前者について、以前は「主観的解釈」という用語を用いたが、これでは狭い意味で主観説を指しているのか、広い意味で論理説と主観説を合わせたものを指しているのか紛らわしいために「認識論的解釈」に変えた。後者について、以前「科学的解釈」という用語法だったが、論理説や主観説が「非科学的」であるという誤解を避けたいので「客観的解釈」に変えた、と説明している。彼はこの用語法があくまでも便宜的なもので、改善の余地があることを認めている。おそらく最大の難点は、〈客観/主観〉、〈存在論/認識論〉の二分法がこの用語法では渾然となってしまい、ギリース自身がそこから言葉の迷宮の中に迷い込んでしまっていることだろう。それゆえに、本文ではギリースの用語法を採用しない。

[*36] ギリースの結論には、彼特有の論理がある。ギリースには、操作主義的な取り扱いが「現代の科学哲学」には原則としてなじまない、という感覚がある。このために、自然科学の領域においては、マッハ主義の最大の特徴ともいえる操作主義から出来るかぎり身を遠ざけようとして客観的側面の中でも傾向説を支持している。これに反して社会科学の領域においては、操作主義の効用を積極的に認めており、そこで持ち出されるのが社会科学の「規範的な性質」であり、そこに関与する人間の「合理的な意思決定」である。自然科学の対象的実体たる物質と社会科学の対象である人間とがこの点でまったく異なるので、後者においては操作主義的な理論がより望ましく、それゆえに主観説を支持しているのである。

ギリースの論点は表面的には整合性がとれているように見えるが、われわれの観点とは一致しない。確かに自然の領域 (自然弁証法) と社会の領域 (史的唯物論) とは運動の基体が根本的に異なるという理解は正しいが、その違いを哲学一般の領域 (確率の見方) にまで持ち込むことは、確率の認識論的側面と存在論的側面をあたかも別々の実体の性格であるかのように実体の側に押し付ける発想でしかない。

[*37] Leibniz の「充足理由律」(3.1.3 節を参照) との類似性に留意のこと。

6.4.1 頻度説

頻度説についてはまずウィーン学団[*38]とマッハ主義との関係に言及しなければならない。頻度説の最大の理論家 Richard von Mises[*39]はウィーン学団のインナー・サークルに所属し、彼の主著「確率・統計・真実」(1928) はウィーン学団の Moritz Schlick[*40]、Philipp Frank[*41]によって編まれた選集「科学的世界構想に関するモノグラフ」の一巻として刊行されたものだった。またもう一人の頻度説提唱者 Hans Reichenbach[*42]もウィーン学団とかかわりがあった。ウィーン学団は当初「Ernst Mach 協会」と名乗っており、Mach の思想[*43](のちに「論理実証主義」と呼ばれることになる) に共鳴するウィーン大学を拠点とする一群の科学者、哲学者たちによって組織されていた。もちろん von Mises も自分の確率学説がマッハ主義に依拠していることを公言している。

当時のオーストリアはどのような文化状況におかれていたか。1914 年のサラエボ事件を機に生じた最初の世界大戦 (これには Richard も飛行士、航空機設計士として従軍している) の前後が焦点となる。それ以前より、1871 年のドイツ帝国成立以後、英独二強国を中心に軍事同盟が組まれ各国の軍備拡張が進行していた。ドイツ帝国とロシア帝国に挟まれたオーストリアも 1878 年のボスニア・ヘルツェゴビナ占領以来、ロシアを後ろ盾とするセルビアと対立していた。民族主義が猖獗をきわめ、それまで積み上げられてきた国際主義を風前の灯とした。これがやがて第一次世界大戦への導火線となった。

ウィーンでもこの世界情勢と無縁ではいられなかった。オーストリアの政治的脆弱性[*44]はかえって自由で国際主義的な多元主義文化をもたらし、多くの文化人を引き付け、その様は「世紀末ウィーン」とも表現された。これが 1895 年以降には一変し、反ユダヤ主義的市長が誕生し、ウィーンでも反ユダヤ主義と排外主義の度が強まった[*45]。一転して第一次世界大戦後の戦間期には「赤いウィーン」と称される進歩主義的な雰囲気の中で短い期間ではあるがさまざまな社会改革がなされた。

[*38] ウィーン学団は 1907 年に物理学者 Philipp Frank、数学者 Hans Hahn(1879-1934)、経済学者 Otto Neurath(1882-1945) らにより、科学哲学と認識論を議論するためのサークルとして組織された。第一次世界大戦による中断の後、戦間期には Moritz Schlick らを世話役として活動したが、ナチス台頭に伴い離散した。

[*39] Richard von Mises(1883-1953) は経済学者 Ludvig(1881-1973) の弟であり、ウィーン工科大学で流体力学、航空工学を学び、彼自身も航空機の設計家だった。ベルリン大学で教えたが、ナチスの政権獲得にともないアメリカに亡命した。またフォン・ミーゼス分布 (円周正規分布とも呼ばれる) は方位統計学 (風向など) で基本となる分布である。

[*40] Moritz Schlick(1882-1936)。哲学者、ウィーン学団指導者の一人。その反形而上学的主張がナチス心酔者の怒りをかい、射殺された。

[*41] Philipp Frank(184-1966)。物理学者、L.Boltzman の弟子。ナチスの政権掌握とともにアメリカに亡命、ハーバード大学にてウィーン学団を引き継ぐ活動を行った。

[*42] Hans Reichenbach(1891-1953) はベルリン学派 (経験哲学協会:1928 創設) の組織者のひとりであり、R. von Mises もこれに参加していた。ベルリン学派はウィーン学派と協働して雑誌"Erkenntnis"を刊行しており、Reichenbach は Rudolph Carnap(1891-1970) とともに共同編集者を務めた。

[*43] マッハ主義については周知のように V.Lenin による批判がある。彼はマッハ主義を「カント哲学・マイナス・"もの自体"」として、あるいは「ヒューム・バークリ主義」として、あるいは観念論として描き出している。

当時の Lenin による批判の目的は、その当時「物心二元論を克服する」と称して流行した「要素一元論」に抗して〈唯物論/観念論〉の違いを明確に区別し、ロシア社会民主党に蔓延しているブハーリン＝ボグダノフ派の思想を (唯物論を僭称しながらも) 実質的に観念論に落ち込んでいるものとして暴き出すことにあった。それは彼の当面の目的にはかなったが、Hume、Mach の思想を内在的に批判するものではなかった。つまり、David Hume が人文哲学にベイコン主義的実験精神を持ち込んだ最初の人であったことを見逃すものであり、Mach がそのキャリアの初期を実験物理学者として出発したことをほとんど無視するものであった。

Ernst Mach(1838-1916) は彼の力学史の研究に入る前に、当時として最先端技術であった写真術を駆使して超音速と衝撃波の研究を行ったが、これは Christian Andreas Doppler(1803-1853) による音波と光波の周波数にたいする観測者と振動源との相対運動の効果 (いわゆる「ドップラー効果」) の研究の延長線上にあるとみなせる。こうした研究テーマは (その当時ドイツ・オーストリアで勃興しつつあった) 心理的・生理的現象に物理学的方法論を適用する科学研究の新潮流 (E.Weber、G.Fechner、W.Wundt など) に適うものであったし、(Mach 自身は否定するものの)A.Einstein の特殊・一般相対性理論への道を拓くものだった。

Lenin はボグダノフ派の安直な弁証法理解を大いに批判すべきではあったが、それと本来のマッハ主義批判 (すなわち、マッハ主義を自然弁証法の展開のなかで内在的に批判すること) とはまた別のことである。

[*44] オーストリアは 1859 年のイタリア統一によりミラノを失い、1866 年の普墺戦争の敗北とそれに引き続くアウグスライヒ (オーストリア＝ハンガリー二重帝国) により、他国のような政治的・経済的力を失ったゆえに文化的な力に頼らざるを得なかった。

[*45] この状況は、ロシア 10 月革命によって変わるはずであったが、ドイツ革命は挫折し、ワイマール共和制という妥協になって終わり、のちのヒトラーの政権掌握を許すことになった。

その成果はナチス台頭により吹き飛ばされた。このような激変する文化環境のなかでウィーン学団は活動していた。

頻度説の最大の特徴は、1) 確率論が他の力学分野と並んで (数理) 科学のひとつであるとされていること[*46]、2) マッハ主義をその方法論として採用していることである。これら二つの特徴が頻度説をその他の現代的な確率学説とも古典的確率論とも区別するメルクマールとなる。

第一の点、経験科学としての確率論という側面は次のように説明される。確率論は「力学のような数理科学であり、観察できる現象を扱う」、「幾何学の主題が空間現象の研究であるように、確率論は大量の現象と繰り返し起こる事象を対象とする」。「力学において質量がその測定者から独立であるように、確率はそれを評価する個人から独立している」。このような理解は、von Mises の数理科学にたいする (当時としては主流の) 見方に由来している。すなわち、まず経験 (観測) があり、そこから実験的・経験的命題が引き出される。さらにそこから数学的概念が抽象される。このような抽象化・理論化は二重に行われる。第一は現象に含まれる実体そのものが数学的概念に移される。確率論のばあいは「経験的集合 (Kollektiv)」が数学的集合の概念に移される。第二に、実験的命題が数学的公理に移される。この二つの抽象化の結果が数理科学的な理論であり、この理論の具体的な現実への適用が現象の説明・予測とされる[*47]。

第二の点、マッハ主義ないし操作主義という側面は次のように説明される。Mach が「経験から打ち立てられる命題」から質量と力の定義を引き出すように[*48]、von Mises は次の二つの命題を経験的法則として提示する。第一は「相対的頻度の安定化」であり、第二が「系列のランダムさ」あるいは「相対的頻度に影響を与えるようなギャンブルの必勝戦略の不在」である。von Mises は古典確率論が第一の法則 (「大数の法則」) しか知らなかったために不完全なものにとどまったと考えている[*49]。そして第二の法則を熱力学からの類推において、すなわちエネルギー保存則と第一種永久運動の非存在の関係にたとえている。また第一の法則に関連して「繰り返しの相対的頻度は確率の測定であり、それはちょうど温度を〈測定〉するときの水銀柱の長さのようなもの」と捉えている。

その上で von Mises は第一の法則から限界頻度 (相対頻度の極限) において確率を次のように定義している。ここで C は Kollectiv であり、A は C についての任意の性質 (条件・場合・標本) である[*50]。この定義の興味深い点は、ギリースによれば、(論理説と同様に) すべての確率が条件付き確率として定義されていることである。

$$P(A \mid C) = \lim_{n\to\infty} \frac{m(A)}{n}$$

頻度説へはさまざまな批判が投げかけられ、ギリースはそれらを網羅的に紹介している。これらのなかで頻度説にとってもっとも深刻なものは 4) の批判であり、これは最近まで未解決だった。

1). 理論の適用範囲が経験的集合 (Kollektiv) を定義できる領域に限られており、一度しか起きない現象に適用できないこと。この批判は、経済現象への適用を視野に入れる論理説・主観説の立場からも、量子力学への適用を視野に入れる傾向説の論者からも同様になされる。von Mises は「まず集合を、しかるのちに確率を」という標語で頻度説の精神を表している[*51]が、彼自身はこれを必ずしも欠点であるとは考えていなかった。偶然ゲー

[*46] 竹内・大橋 (1981) も R.A.Fisher の業績紹介に関連して、確率論が論理学というよりは経験科学であると主張している。

[*47] ここで実践的唯物論における認識論と比較・対照しておきたい。まず、いずれにおいても上向の過程と下向の過程が区別できる。理論の現実の適用とは抽象的なものから具体的なものに移る上向の過程であり、実践の側から見ればこれは目的の生産である。それ以前は具体的なものから抽象的なものに至る下向の過程である。von Mises の理解では、現象、実体、本質の区別があいまいとなっている。マッハ主義、操作主義が批判されるのはこの区別のあいまいさに起因しているが、抽象化・理論化の過程を論理的に整理するという問題意識そのものは正当である。また細部において、現象の観察から実験法則を定立することも正しい。

[*48] たとえば慣性質量はニュートン力学の第三法則 (作用・反作用の法則) から定義される。

[*49] von Mises は頻度説の先駆者としてフランスの Antoine Augustin Cournot(1801-1877)、イギリスの John Venn(1834-1923)、ドイツの Georg Ferdinand Helm(1851-1923) を挙げている。これに加えてギリースはイギリスの Robert Leslie Ellis(1817-1859) の名も挙げている。

[*50] たとえば C は一つのサイコロを 4 回投げること、A は出目が 6 であることに対応する。$m(A)$ はその頻度・度数である。

[*51] また先の温度計の類推 (分子一つに対しては〈温度〉概念が適用できない) から考えてもこれは von Mises からすれば当然である。

ムの領域に限られていた古典確率論に比べれば、当時の自然科学研究に使うためにはこれでも十分に広かった。

2). 頻度説は帰納法論理とほとんどかかわりをもたないこと。これは論理説・主観説の立場からの批判であり、彼らが重視する命題論理と確率の関係に何も回答を与えていない。そもそも批判者らは確率論を論理学として捉えており、von Mises のように科学として捉える立場に立っていない[*52]。

3). 頻度説のとるマッハ主義・操作主義という方法論への批判がある。

   i). Cramér(1946) は「(頻度説による確率定義は) 経験的要素と理論的要素の混合物を含んでおり、現代の公理論的な数学の立場と相いれない」。ギリースも「理論的な概念を観察されるものから直接的に定義するべきだとは信じられておらず・・・未定義のままで、もっと間接的な仕方[*53]で実験に結び付けられるべき」と論じている。またすべての概念が操作的に定義されねばならないとすると、定義が際限なく後退してしまうことになる。

   ii). 確率の定義にあたって無限の系列による操作を採用している点。確かに定義は「操作」的に行われるが、(可算) 無限回の操作は現実には実行できない。von Mises はこの点を「数学的概念への抽象化」という言葉でごまかしている[*54]。

   iii). 上の ii) に関連して、理論上の確率と現実の頻度との関係を直接的に対応することへの批判。de Finetti は確率論と物理学 (実験科学) は直接には同一視されないと主張し、次のように述べている[*55]。「もし理論が完全に操作的であれば、何が起こるかを確実かつ正確に主張し、予見することができるが、確率計算においては理論自身がすべての頻度の可能性を認めざるを得ない。」

4). von Mises のプログラムが未完成であること。とりわけ第二の法則 (ランダムさの法則) とギャンブルの必勝法の不在との関係が証明されていないこと。またのちの研究により、「ランダムさの公理の定式化」に深刻な数学上の問題が含まれていることが指摘されたこと[*56]。

von Mises の頻度説が確率論の展開にもたらした功績には、a)Kolmogorov の公理系に必要とされる道具立ての多く (とりわけ「性質空間」の概念) を準備したこと、b) マッハ主義をあくまでも唯物論の観点から擁護し、救い上げる努力をしたこと、c) 頻度説の最大の難関がランダムさの数学的な定義にあることを明るみに出したこと、などが挙げられる。

a). 第一の点に関連して〈標本空間〉Ω の概念が生み出されたが、von Mises 自身はこれを「性質空間」

[*52] またこの批判は von Mises が存在論的な理論として確率を定義しようとしているにもかかわらず、無自覚のままに認識論的な局面に議論の舞台を移そうとしている。

[*53] 唯物弁証法ではこれを〈媒介性〉という名で表現している。すなわち、本質的なものは実体を媒介にして現象している。ヘーゲル弁証法のように、本質が一足飛びに現象へと疎外・外化されるのではない。なお脚注*47 も参照のこと。

[*54] 有限回の操作ということに着目すれば、(母数としての) 確率は仮説的な地位にとどまらざるをえず、その代わりに推定論・検定論を扱わざるを得ない。これは認識論上の課題であり、確率論という存在論的な理論体系からは逸脱する。これは、そのことが十分に意識されているならば必ずしも欠陥ではない (ちょうど「資本論」において、〈競争〉の取り扱いが価値論の展開のなかでは叙述留保されているように)。

[*55] この批判と Aristotle がチュケーを科学の対象ではないとした論理と対照することは興味深い視点をもたらすだろう。de Finetti は無意識に Aristotle と同じような誤りに落ち込んでいるように見える。

[*56] たとえば宮部 (2011) を参照のこと。これによれば、ランダムネスを定義する von Mises の試みが〈Kollectiv〉の定義 (その任意の部分列が同一の確率値に収束する) に結実したが、肝心の部分列を構成する操作主義的手続きを von Mises は与えなかった。Alonzo Church(1903-1995) は帰納的関数の概念を使って〈Kollectiv〉の数学的定義を完成させたが、Jean Ville(1939) によりこれが Aleksandr Khinchin(1924) の重複対数の法則 (これは Émile Borel(1909) の大数の強法則の精密版であった) を満たさないことが見いだされた。その後、Per Martin-Löf(1966) が「テスト」概念を用いて Martin-Löf ランダムネスを定義し、これがランダム列に期待される多くの性質を備えていることが発見された。それらの性質の中には Ville によって〈Kollectiv〉の代わりにゲームの必勝戦略の不在を示すものとして導入された〈マルチンゲール〉も含まれていた。そしてシェイファー・ウォフク (2006) は〈マルチンゲール〉を基礎として逆にランダム性を定義することに成功した。

シェイファー・ウォフクは彼らのアプローチを「新主観主義」と命名している。彼らがなぜ von Mises という入り口から出発して de Finetti の出口に至ったのかは興味深い研究テーマである。

[*57] ギリースは〈標本空間〉という用語は「性質空間」とくらべて改悪であるとしている。なぜなら、それは〈標本化〉(サンプリング) とはなんの関係もないからである。

と呼んでいた[*57]。ギリースによれば、性質空間とは「(繰り返し起こる事象や大量現象に関して) アプリオリに可能とみなされる一連の性質」を指す。たとえば $\Omega = \{$ コイン投げの「表」,「裏」$\}$ が挙げられる。Ω は「基本的」性質であり、ここから「事象」すなわち「構成的」性質が組み立てられる。たとえば Ω がサイコロの出る目であるとして、その部分集合 $\Omega' = \{2, 4, 6\}$ は「偶数」という性質をなす。

b). 第二の点に関連して von Mises は確かに唯物論の立場を堅持しており、そのことは「確率が、それを評価する個人から独立している」という言明からも明らかである。彼は (Lenin とは異なり) マッハ主義と唯物論とは互いに両立するものとみなしているが、彼には弁証法の考え方がないためにこの二つを調和させることができず、他の科学哲学者からの批判を浴びることになった。

c). 第三について、von Mises は〈Kollectiv〉という概念を提示した。これは「特定の観察しうる性質 (色、数、その他) において異なる、一連の均等な事象やプロセスを示すもの」としている。今日の用語では〈データ〉であり、抽象的に二進数の有限数列と同一視されている。von Mises が努力したのは、〈Kollectiv〉がランダム性を示す条件・定義、さらにはそれとギャンブルの必勝戦略の不在が同じであることを証明することであり、それにはついに成功しなかった。

頻度説の最終的な評価として次のことが言える。この理論は〈不充足理由律〉を大量観察に置き換える Jakob Bernoulli の試みを引継ぎ、先験的な公理ではなく観測に重点を置くという視点の転換をもたらした。そのことによって偶然ゲーム以外の諸科学に確率論を拡大適用する根拠を与えた。観測値〈Kollectiv〉におけるランダム性とは何か、についてはたしかに von Mises の時代には片付かず、情報理論/計算理論の登場を待たねばならなかった。また von Mises の示唆した熱力学との類似性 (賭けにおける必勝戦略の不在) も、後代のゲーム理論に基づくマルチンゲール戦略の研究を必要とした。それでも、頻度説の枢要な部分は傾向説へと引き継がれ、客観的に存在する偶然性の仮説が量子力学の解釈問題を通じて鍛えられることになった。

6.4.2 傾向説

確率の客観性を主張し、かつ、頻度説をとらない見解は一括して「傾向説」と呼ばれる。この説を最初に唱えた人物は Karl Raimund Popper(1902-1994) であるが、その外にも複数の論者が様々なバージョンの傾向説を提案した。しかし、ここではそれらを網羅的に紹介することはしない。われわれにとって重要であるのは、頻度説がなぜ否定され、それに代わるものとして傾向説がなぜ提案されねばならなかったかを明らかにすることである。そのために、ここでは Popper の説を中心に検討しなければならない。

Popper はウィーン学団の随伴者ではあったが、論理実証主義 (Logical positivism) とは一線を画し、自身の立場を反証主義 (Falsificationism)[*58]と称していた。確率について当初は頻度説の見方をとっていたが、1957 年にはじめて傾向説を明らかにした。

Popper が傾向説に転じた動機は、量子力学の解釈において (頻度説の否定するところの) 単一事象の客観確率 (singular probabilities)[*59]が必要と考えたからで

*58 「反証主義」は統計的検定理論の背後にある基礎的な考え方であり、「論理実証主義」よりはより確実 (保守的) な態度と親和性がある。これが認識論の領域の議論であって、存在論ではないということに注意が払われるべきである (Popper は確率の客観性という点では一貫して存在論として語っている)。イギリス経験論の伝統的枠組みでは認識論と存在論の区別がはっきりせず、Popper の議論もこのような弱点を共有している。彼の哲学がこのような弱点を抱えているので、ヘーゲル哲学とマルクス哲学 (両者とも存在論の領域があることを明示している) をともに「歴史主義」の烙印を押して否定してしまう。(友人である Hayek らにそそのかされて執筆した)「開かれた社会とその敵」(1945) の第 2 巻は上の哲学上の難点と当時のナチスによるオーストリア併合という政治情勢に付随してあらわれたドイツ哲学への嫌悪感の混合物とみなせる。

なお、イギリス経験論が「明晰さ」という自覚されざるイデオロギー的特質を備えること、これがキリスト教神学に由来することを大陸系の哲学者たち (たとえばアドルノ (1963)) が暴き出している。「明晰さ」に価値を置きすぎるならば、Hegel の言わんとするところを捉えそこない、「たわごと」にしか見えなくなる。

*59 ギリースは Popper の弟子ではあったが、単一確率の必要性を否定し、なおかつ傾向説を支持している。Popper があくまでも量子力学という存在論的理論の解釈を問題にしていることにたいして、ギリースは確率論への「反証ルール」の貫徹という認識論的側面を重視したからである。

あった。ギリースはどういうわけかそのことを詳細には論じていない。そこで、以下は高村 (2010) の説明に沿って Popper の論旨を追ってみる。

高村の基本的主張はこうである。1)Popper が「確率の傾向性解釈」を考えた動機は量子力学のパラドクス (「コペンハーゲン解釈」) の解決および客観的実在の復権を意図してのことである。それにもかかわらず、当初も今も Popper の提案によって量子力学の難点が解消されたとは一般に受け止められていない。これは「コペンハーゲン解釈」をもたらす原因となった「二重スリットの実験」[*60]について、Popper がなんら合理的な解釈を与えなかったからである。2)Edward Nelson(1932-2014) の「確率力学」が実は Popper 構想の埋められていないパズルの最後のピースである。ただし、「確率力学」においても確率は主観的なものと解釈されているために、パラドクスを最終的に解決するものとは受け止められていない。これを客観的なものと解釈替えする「客観的確率過程解釈」[*61]が必要である。これによって Popper 構想は完成され、その本来の意義 (「傾向性の場」としての確率) が明らかにされる。このような高村の主張は、正しい。

まず Popper の主張が次のように整理される。

1). Niels Bohr(1885-1962)、Werner Heisenberg(1901-1976) らによる「コペンハーゲン解釈」、すなわち微視的物質が (粒子であると同時に波であるという) 二元論的存在であるという見解は、客観的実在が保証されないという意味で「量子論の泥沼」である。

2). 上の混乱は i) 確率の主観主義的解釈と ii) 量子力学の完全説[*62]に根拠をもつ。

　i). 量子論においても、確率概念を人間の無知の産物とする (古典確率論から引き継いだ) 誤った考え方が組み込まれている。この誤りは〈不確定性関係〉についての Heisenberg の初期の解釈において最高潮に達した。量子論における確率は (客観的な零点振動としてではなく) 観測精度の上限として、観測者たる人間主観の側に押し付けられた。

　ii). その反面、量子論はニュートン力学と同様の決定論的性格を与えられた。Popper によれば量子論は「物理学のたどり着いた終着点」、「乗り越えられることのない最終的な物理学の革命」を僭称している。本来は本質的に確率的な理論である量子論が、確率が観測者の側に押し付けられたが故に、それ自身は完全な性格を与えられている。当時の量子力学の成功体験がこうした感覚をますます強めている。

3). Popper は量子論における問題、とりわけ「波束の収縮」[*63]を説明するために、確率の「傾向性解釈」を打ち出している。その内容は、「単一試行における客観的な確率的性質」を「素直に認め」、「なおかつそれとは別に実際の頻度としての確率も (理論値としての傾向性が正しいかどうかをテストするための道具として) 認める」というものである。ここで以下の注意が与えられる。

　i). 傾向性解釈は「単一試行における確率とは何か」という哲学的な課題に答えるものではない。

　ii). 重み (傾向性の場) としての確率は実在する「実際の物理的属性」である。これはピンボール・マシンにたとえられる。ピンボールからピンを一本引き抜くことは、ボールの運動経

[*60] 二重スリットの実験において、スクリーン上に映し出される「干渉模様」について、「コペンハーゲン解釈」は次のように説明する。すなわち、「この干渉現象があることから一個の電子は一回の実験において、何か波のような大域性を持った存在として両方のスリットを通過し・・・一方で検出フィルムは常に局在した点を示すことから、電子の粒子性も否定できない。ここに波と粒子の二重性という概念が生じる」と。

[*61] 5.3 節で紹介した長澤の理論がまさしくこれである。

[*62] 高村は量子論の完全性・不完全性が現代でも議論の的になっているのだから、Popper がこのように量子論の不完全性を断言するのは理解できない、と言っている。高村は今日の視点でその当時の Popper の感覚を否定しているようにも見える。

Popper は量子力学の不完全性を強調しなければならない強い動機を少なくとも二つもっていた。第一に、二重スリット実験において干渉縞 (量子もつれ) が起きる理由を説明する「義務」をもつのは自分 (科学哲学者) ではなく、量子論 (物理学者) の側だと考えていた。第二に Popper の言う「傾向性の場」としての実在的確率の姿を目に見えるかたちで示すことも量子論 (物理学者) の仕事と考えていた節がある。

[*63] 「シュレディンガーの猫」のたとえとしても知られている。

路 (確率過程) と最終的な状態にたいする確率分布全体に影響を与える。

4). Popper はその上で量子論における「波束の収縮」、二重スリット実験の「干渉模様」(量子もつれ) について次のような説明を与える。

i).「波束の収縮」とは、ピンボールで特定のピンに当たるボールに注目したときの、元の確率分布から新しい確率分布への「遷移」(視点の転換) のことであり、これは確率論一般のあたりまえの帰結である。けっして「超高速作用」のような神秘的なものではない。「波束の収縮」なるものは、実験環境によって決まる確率分布 (傾向の場) と粒子自体とを混同した結果に過ぎない。

ii). 傾向説は、「粒子であるはずの電子がどちらか一方のスリットのみを通るとしたら、なぜその粒子はもう一方のスリットの開閉を知っている・・・のだろうか」という (誤った) 問いにたいして次のように答える。すなわち、「傾向性を決めているのは実験配置全体であり、スリットの開閉に影響を及ぼすのは傾向性なのであって、粒子それ自体ではない」。「スリットの開閉を〈知っている〉のは粒子ではなく、傾向性の場である」と。

以上の Popper の立論にたいして高村は次の難点を投げかける。すなわち、「量子論に主観性をもたらしたのは確率の主観主義的解釈などではない」、「粒子自体と傾向性の場の混同には別の理由があるのであり、ポパーの指摘するように単なる勘違いや誘惑などではない」。この理由こそが「二重スリット実験」であった。これが説明されないかぎり「コペンハーゲン解釈」は無傷のままである。

ここまでの議論は結局のところ、1)(量子力学を補足し)「二重スリット実験」の干渉縞を説明する理論とは何か、2)「傾向性の場」とは具体的に何を意味するか、という二つの問題に帰着される。この二つはそれぞれ別のことを言っているが、Popper はそのいずれについても解答を与えることができなかった。高村は、前者について E.Nelson の「確率力学」がその解答であるとしている。この理論は 5.3 節の長澤理論と実質的に同じものである[*64]。後者について高村は初期と終期の確率分布 (長澤の用語では「入口関数」と「出口関数」) に関連して「最小作用則」こそが「傾向性の場」であると答えている。高村のこの見解については判断を留保しておきたい[*65]。

傾向説の意義は、客観的確率を二重に捉えたところにある。つまり、頻度としてのそれと傾向としてのそれであり、両者は (すくなくともそれがあるがままのものとしては) 互いに別のものである。前者は試行 (実験) のたびにその都度観測され、記録される観測値を集積したものとしての〈Kollectiv〉、すなわち〈標本〉のなかから生まれる。後者は、存在していることは知られているが、それをマッハ的操作主義によっては明示し得ないものであり、言い換えれば〈母数〉のことである。〈母数〉の大雑把な見積もりは人間の経験や論理 (それ自体経験から抽出されたもの) に基づいて、恣意的に構成されなければ

[*64] ただし重大な違いはある。高村は Nelson の理論が主観的確率解釈をとっているために、そのままでは Popper の議論に結び付かないとしている。すなわち「確率を主観的に解釈することが、やはり干渉現象など起きないだろうという概念的ネックにつながり、量子力学を再現する確率過程が存在しているにもかかわらず、それが真の物理的記述として受け入れられず、未だにコペンハーゲン解釈がまかり通っている・・・」と結論している。

他方で長澤はどういうわけか自説と Nelson 理論との差異を強調する方向に傾いている。これは長澤が Nelson の最初の論文しか見ておらず、その後の展開を見落としているためである。長澤は特に以下の点で Nelson との違いを強調している。つまり、1)Nelson(1966) が Schrödinger(1931,1932) を引用せず、ニュートン力学からの導出にこだわっていることに否定的であること、2)Nelson 理論が従来の量子力学の枠から出ていない (「確率量子化」という) 量子化の方法の一変種にすぎないと見ていること、Nelson がシュレディンガー方程式の解である複素発展関数をそのまま「波動関数」と表現していることから、彼がコペンハーゲン解釈を捨て去っていないと考えていること、3) 主に計算上の問題として、Nelson(1966) が「出口関数」の概念を導入していないこと、Nelson の運動方程式が非線型であるため扱いずらいこと、相対論的な拡張が難しいと考えていること、などを指摘している。長澤は高村の引用している Nelson(1985) などを見ておらず、枝葉末節にこだわることによって二つの理論の本質的な同一性を認めようとしていないのである。

[*65] たしかに「最小作用則」は「傾向性の場」のひとつの言い換え (より精密な) と言えるかもしれない。しかし、それで問題が解決したということにはならない。長澤は「傾向性の場」(長澤の言葉では「ランダム仮説」) について、「真空が電子のブラウン運動を生み出す性質をもっている」という以上のことは言えない、それは理論の展開において「本質的に避けることはできない」が「ひとつの仮説である」としている。武谷の三段階論を援用すれば、「傾向性の場」とは実体的な把握ということであり、それが何であるかの本質的な把握は依然として (拡散係数などの実測を踏まえた) 将来の研究に委ねられていると見なければならない。

ばならない。これがベイジアンの言う事前確率である。事前確率の構成法について何か公理のようなものがあるわけではなく、これを先験的に想定することは、赤池が不毛な議論として切って捨てたところのものである。

上の二重化された客観的確率は、推定・検定作業を経て最終的に同一視される。日常的な人間行動においては、当然のことながら相対頻度としての確率と〈母数〉としての確率との間の差異は検証されることがない。事前確率は初期の心理的歪み (パラドクス) を内包したまま放置される。しかし、経験科学においては、事前確率は未知の確率分布として (退化しているばあいは未知数として) 扱われ、〈Kollective〉=〈標本〉と対照され、その存在範囲が推測的に絞り込まれる。

傾向としての確率は、その存在が客観的であるとは言え、それを確定的に表現することは出来ず、主観においてしか示すことができない矛盾したものである。しかしこの矛盾を認めないならば、Aristotle 的な禁止則に再びはまり込んでしまうことになる。Hic Rhodus, hic salta!

6.4.3　論理説

論理説の著名な提唱者は John Maynard Keynes (1884-1946) である[*66]。ギリースによれば、その著作「確率論」(1921) は「ケンブリッジ・アポスルズ」[*67]の先輩であった George Edward Moore (1873-1958) の「プリンキピア・エチカ」(1903) と Bertrand Russell (1872-1970) と Alfred North Whitehead (1861-1947) の「プリンキピア・マセマティカ」(1903) の知的影響下で 1904 年頃から構想されたものであるらしい。ケンブリッジ大学は John McTaggard らに指導されたイギリス観念論、とりわけ新ヘーゲル主義の支配下にあった。この思想は、物質に対する精神の優位性、人格主義、自由主義を標榜していた。Moore、Russell はイギリス観念論の学徒として出発するも、やがてこれに反旗を翻し、新実在論と (体系化ではなく) 分析的方法を重視する「分析哲学」を提唱し、イギリス経験論の伝統への復帰を目論んだ。

その当時のイギリスの政治・経済的状況はどうであったか。イギリス資本主義はその自由主義的段階から帝国主義の段階への既に変貌を遂げていた。通商問題が英国政界を揺さぶり、Chamberlain[*68]の関税改革が自由貿易派の自由放任主義とぶつかりあっていた。労働者階級の力が強まり、ブルジョアジーといえども社会政策の重要性を無視しえなかった。それを財政的に支える植民地政策はますます新たな財政上の困難をもたらしており、政治情勢もそれにつれて揺れ動いた。

Keynes の確率の論理説提出の動機は、ギリースによれば Moore への反ばくを意図したものだった。Moore の考えでは、「善」[*69]を求めての練りに練った行動も所詮は目先のことであり、長い目で見れば不確実性のために台無しにされる。だから、倫理に関しては常識的な道徳律に委ねるのが正しく、またそれ以上のことはできない。Keynes はこの議論が間違っていると感じていた[*70]。というのも、彼は (アポスルズのメンバーのような) 理性的な人間ならば、ある行為がたとえ慣習的道徳に反していたとしても、それが「善」であるかどうかを判断できる、と信じていたからである。Keynes は Moore が (実在論の立場から) 確率の頻度説に足をすくわれたがゆえに間違った結論を導いたとして、頻度説によらない独自の確率論を構築する必要に駆られた、と推察できる。

Keynes は Moore に反ばくするために次のような論理を組み立てた。

i). 短期的にもたらされる「善」の大いさを 行為 A > 行為 B とする。
ii). 長期的には、無差別の原理から 行為 A > 行為 B

*66 ギリースによれば確率の論理説の先駆者は William Ernest Johnson(1858-1931) であった。彼はケンブリッジ大学で論理学を教え、効用理論の確立に貢献した。John Maynard の父 John Neville と同僚であり、教え子には John Maynard のほか、Harold Jeffreys(1891-1989)、Frank Ramsey(1903-1930) などの主観説の立役者がいる。そのような意味で、論理説と主観説とは学説上、密接な系譜関係があると言えよう。

*67 ケンブリッジ大学で 1820 年に創設された秘密会であり、メンバーに多数の著名人を抱える。

*68 Joseph Chamberlain(1836-1914) : 当時、植民地大臣。

*69 しかも「善」は定義できないものと Moore は考えていたので、その大いさを測ることも Moore にとっては疑わしいものだったはずである。

*70 またこの議論が正しいならば、後に Keynes が主張することになる経済理論の主要な部分が否定されることになってしまう。

も 行為 A ＜ 行為 B もいずれも同様に確からしい。

iii). したがって、長期的効果は無視してさしつかえなく、短期的効果のみに着目して行為 A を選択すれば良い。

ギリースはこの議論について次のような興味深い指摘を行っている[*71]。1)Pascal の「神への賭け」との類似性、2)「善」を「投資収益」に置き換えても議論が違和感なく成立すること。

Keynes は D.Hume の因果推論を拡張する方向で議論を進めた。次の三つの命題、

$$\begin{cases} e & n \text{ 羽のカラスは黒かった} \\ h & \text{すべてのカラスは黒い} \\ d & \text{次に観察されるカラスは黒いだろう} \end{cases}$$

の間の関係を見たとき、Hume は $e \to h$ も $e \to d$ も誤った推論であると論じている。それはたしかに正しいが、Keynes は「部分的 (確率的) に」ならば $e \to h$ と $e \to d$ は間違っていないだろうと考えた。つまり、Hume が言っているのは (3.1.2 節の Borel の意味で)「確実な」$(1-p=10^{-10})$ 推論のことであり、「まず確からしい」$(1-p=10^{-1})$、「非常に確からしい」$(1-p=10^{-3})$ という条件 (有意水準) であれば許容されるべきだと主張したのである[*72]。Keynes の考えは今日では因果推論 (Causal Inference) と呼ばれる理論分野の嚆矢と評価すべきものであり、具体的にはベイジアン・ネットワーク (3.3 節) の考え方を述べたものと解釈できる。

Keynes の議論からは確率の論理説の以下の特色が自然に引き出される。

1). 命題間の関係としての確率は、条件付き確率として与えられ、その計算には次のベイズ・ルールが適用される。

$$\begin{aligned} P(h_1 \mid e) &= \frac{P(h_1 \cap e)}{P(e)} \\ &= \frac{P(h_1 \cap e)}{\sum_i P(h_i \cap e)} \\ &= \frac{P(e \mid h_1)P(h_1)}{\sum_i P(e \mid h_i)P(h_i)} \end{aligned}$$

2). 確率は命題間の関係であるので、一次元の尺度ではなくネットワーク (今日の用語では「因果グラフ」) として表現される。命題自体の確率には計算 (識別) できないものもあり、完全に順序付けできるとも限らない[*73]。

3). ただし、確率が先験的に与えられるべき命題はある[*74]。これは「無差別の原理」から計算されるべきである。この原理が多くのパラドクスを生み出すことは Keynes も承知しており、そのため選択肢の数が有限でそれ以上分割できないばあいにのみ適用できるとしている。

Keynes の論理説は、Hume の因果推論の議論を確率

[*71] これらの指摘が興味深いのは、Keynes が人間 (ないし投資家) の行動を暗黙の対象にしていることを明らかにしていることである。つまり、人間は現実にリスク評価を行って行動していること、その行為そのものは物質的であること、そのリスク評価が合理的であるかどうかを計る尺度があると考えていること、などである。ここでの人間とは、資本家的合理性を備えた人間であり、超歴史的な人間でもなければ、実験家 (観測者) としての人間でもないことも明らかである。またその資本家的合理性も、産業資本家のそれと銀行資本家とでは異なる。前者のばあいは巨額の設備投資のために価値を生まない建設期間の存在、実現された莫大な生産能力にたいする原材料と労働力の枯渇、減価償却期間を待たずに技術が陳腐化する「道徳的摩滅」、キャッシュフローの不足から新規の需要に応じられないことによる埋没費用などがリスクとなる。後者にとっては、手元の資金を利子生み資本として活用できる有利な投資先の発見、期待された利潤率に届かず元本割れやデフォルトになるリスクなどがある。

プロレタリアートにとってのリスク評価は資本家とはまったく異なる。彼が参照するのは資本ではなく常に自分の労働の可能支出量と必要支出量であり、それぞれの不確実性がリスクの源泉となっている。

[*72] カラスの羽の色 (烏羽色:からすば色) は実際には紫がかった黒であり、完全な黒ではない (色コードには#180614 が割りあてられている)。また実際の羽の色もすべてのカラス属で黒一色というわけではない。カラスの羽の黒はメラニン色素に由来し、その色素ができなければアルビノになり、またそうした個体は実際に発見されている。

しかし、現実には「カラスは黒い」という観念が強く支配している。アルビノのカラスの発見例が希少である理由にはなんらかの淘汰圧の存在 (親鳥が飼育を放棄する、天敵に捕食されやすい、など) が考えられる。また、発見者が実際に白いカラスに遭遇したとしてもカラスではなく「白くて見慣れない鳥」としか認識されない可能性がある。このように見ると、観測対象、観測者のそれぞれの側で「カラスは黒い」という信念を強化する客観的なメカニズムがあり、命題 h が提出される根拠になる。

[*73] ギリースはこの議論に賛否両論があったことを指摘している。de Finetti は否定的であり、Jochen Runde(1994) は賛同している。de Finnetti の批判は、Keynes が事象の確率そのものとベイジアン・ネットワーク (あるいは条件付き確率) を混同してみずから混乱に陥っていたことを指摘するものであったが、その指摘は Keynes の問題意識とかみ合っていない、という意味で超越的である。

[*74] Ramsey は Keynes が先験的に確率を与えられると考えていることについて、それは錯覚にすぎない、と攻撃している。Ramsey が主観説を提出した動機はここにある。Ramsey もアポスルズの一人であり、Ludwig Wittgenstein(1889-1951) の友人だった。

分野に正しく拡大しようとする意図に支えられたものであり、その基本的な考え方は今日では Pearl などの研究に引き継がれ、応用技術としてもエキスパート・システムの実装などに使われている。ベイジアン・ネットワークの研究では〈エビデンス〉の意義は消えてしまうどころか、より一層重視されており、その意味で von Mises の頻度説は因果推論の不可欠な要素として論理説と一体化している。Keynes の説で強い批判を受けた部分は、先験的な確率判断を可能だとする〈無差別の原理〉、〈不充足理由律〉である。これは因果推論のなかでしぶとく生き残り続けた最後の観念論的要素と呼ばれるべきであろう[*75]。

Keynes の動機の深いところでは、Moore 的レッセフェールへの拒否感があった。これは資本制経済の自由主義段階から帝国主義段階への移行、そこでの金融資本の内部における産業資本家と銀行資本家の対立を根拠とするものである。Keynes はこの対立において、産業資本家の立場で論陣を張ったものと評価されよう。またこれは当時のドイツ、オーストリアで高まった (生産の計画化を基礎とする) 経済の計画化への衝動に対する肯定的反応というべきである。これが後の (長期均衡ではなく) 短期の経済運営を重視するケインズ経済学をもたらした。

*　*　*

なお、ケインズ経済学 (それとともに経済計画) はブレトンウッズ体制の崩壊とともに批判を浴びて、今日の合理的期待形成学派に圧倒されたが、その間の事情について菊地 (1982,1996) に基づいて以下にまとめる。

計量経済学からの貢献と挫折

菊池 (1982) によれば、計量経済学は「1929 年恐慌の落とし子」であって、1870 年代の限界革命の抽象性を批判して 1910 年代から開始された経済法則の統計的研究を引き継ぐものであった。大恐慌は緊急の経済対策の必要性をもたらし、計量経済学を体系的に研究するための計量経済学会の創立 (1930) をもたらした。これ以降、理論上ではケインズ経済学、実証上では計量経済学の分業体制が出来上がり、経済学の主流を占めることになる。1930 年代は計量経済学の「疾風怒涛の時代」(W. Leontief) であり、各種の (ただし単一の) 方程式モデルの推計が試みられた。

1940 年代には計量経済学が内部に抱える深刻な二つの問題が浮き彫りにされた。1) 識別問題と 2) 多重共線性問題である。

1). 識別問題 (identification problem):「需給曲線のように複数の関係式の交点で現実の統計データを説明しようとすると、仮に適合度の高い回帰式を導いたとしても、それだけでは、それが需要曲線であるのか、供給曲線であるのか判断できない」こと[*76]。

2). 多重共線性問題 (multi-colinearity problem):「多元回帰モデルにおいて、説明変数の全部または一部に高い相関があると、回帰パラメータが不安定になり、推定結果の信頼度が著しく低下する」こと。R.Frisch によって見出された。

上の問題は T.Haavelmo[*77]の同時連立方程式アプローチによって、両者の同時解決というかたちで一応の決着をみた。それによれば、多重共線性の発生は説明変数間に別の関係式が合流したことによる。だから、合流した関係式を分離し、複数の関係式の同時推定を行えば識別問題も解決する。そのために以下のような新しい枠組みが導入された。a) 内生変数と外生変数の区別、b) 構造方程式 (structural equation) と誘導型方程式 (reduced form; 前者を内生変数について解いた

[*75] 〈不充足理由律〉はベイジアンたちの意識のなかで事前分布の設定法として生き残った。すなわち、活用できる事前の情報がないばあいには一様分布を事前分布として使う、というものである。

[*76] これは Henry Ludwell Moore(1869-1958) が回帰法を用いて正の傾斜をもつ需要曲線を導いたことに、R.A. Lehfeldt が疑問を呈したことを契機とする、とされている。

[*77] Trygve Magnus Haavelmo(1911-1999) は、ノルウェーの経済学者であり、Ragnar Frisch(1895-1973) の高弟、コウルズ委員会に所属 (1939-1947) した。

[*78] ただし当時の計算手段の制約から、現実的には誘導型方程式に二段階最小二乗法を適用し、変数変換により構造パラメータを求めなければならなかった。

[*79] 経済学研究のためのコウルズ委員会 (The Cowles Commission for Research in Economics) またはコウルズ財団 (Cowles Foundation) とも呼ばれる。1932 年に設立。1939 年から 1955 年までシカゴ大学に本部を置き、その後 (シカゴ大学の内部紛争の結果) イエール

もの) の区別、c) 誤差項としての確率変数の導入 (パラメータ同時推定の手段)、d) 推定方法としての最尤法の採用*78。その後コウルズ委員会*79により操作変数法が確立され、Haavelmo の解決法を実証に移すことが可能になった。

計量経済モデルは 1940 年代まではたんに現実の説明手段として使われるのみで、その予測能力には期待がさほど置かれなかった。ところが、1950 年代になると現実経済を予測する手段として見られるようになり、モデル評価手段*80が開発された。予測の外れは外生変数の内生変数への取り込み、新たな外生変数の追加といったモデルの大型化をもたらした。このモデルの大型化により、せっかく 1940 年に完成された推定方法は使えなくなった。菊地 (1996) はこれを以下のように説明している。

1). モデルの大型化により同時方程式アプローチは無視され、パラメータ推定は OLS で済まされるようになった。
   i). OLS の前提である誤差にかんする仮定への関心から「回帰診断」*81が行われた。また誤差項の系列相関の取り扱いは時系列モデルへの関心を高めた*82。「定常性」*83を満たすためにデータの階差がとられたが、これは結果的にもとのデータとは似ても似つかない数値に依拠したモデル構築につながった。
   ii). データは統計調査の結果であるから誤差項の仮定の経験的証明はそもそも困難であった。そうした仮定は主観 (恣意) として取り扱い、これをデータによって補正することにより現実に近づけるベイズ・アプローチに期待がかかった。
2). こうした過程において、経済理論はなんら積極的な役割を果たさなかった。モデルのデータへの形式的な適合のみが重視された。その意味で同時方程式アプローチは時系列解析にまるごと置き換えられても何も不都合は生じなかった。
3). 結果として、回帰分析、時系列解析、ベイズ・アプローチが「それぞれの適用限界を補うかたちで」それぞれの位置を占めた。これは混乱と折衷以外のなにものでもない。

われわれは以上のような計量経済学の展開＝崩壊過程を別様にも解釈できる。

- 予測の外れは経済体制自体の変化 (ドルショックを契機とする固定相場制度から変動相場制度への移行) がそれとして意識されず、過去からの延長線上でそのままモデルが構築された結果と見て取れる。経済は安定成長ではなく成長の限界と金融的要素の比重の増大を基調とするものに変わった。ミニ・バブルの膨張と破裂、金融危機が慢性化した。利子率は金融自由化により低く抑え込まれた。
- 本来は上の変化を識別すべき「理論」はその役割を果たさなかった。むしろこのような体制変化は経済計画の否定*84、財政支出ではなく金融政策の優位 (マネタリズム)、「期待」なるものを基礎とする理論構築、小さい政府を目指すイデオロギーの隆盛をもたらした。計量経済学を指導すべき経済理論 (ケインズ経済学) はその地位を追わ

大学に移り現在に至る。「理論と測定」をモットーとし、多くのノーベル経済学賞受賞者を輩出した (Haavelmo は自身の受賞を喜ばなかった)。

*80 平均平方誤差、不一致係数などである。

*81 系列相関、分散不均一性、正規性など、また回帰式全体の適合度検定などが行われる。

*82 ラグ付き内生変数から多変量自己回帰 (VAR) モデルへと拡張された。Lucas(1978) は VAR を前提とする立場から従来モデルがパラメータをア・プリオリにゼロと仮定していることを批判した。

*83 定常性が満たされているかどうかは、ユニット・ルートの検定により確かめられる。

*84 Robert Lucas(1937) はマクロ経済モデルにもとづいた経済計画の意義を否定したが、それは政府が経済変数を目標とするや否や、他の「経済主体」がその裏をかいて利益を得ることを理由としている。その現実的な基礎は、戦後経済秩序を内側から蚕食した「ホットマネー」の存在である。これがヨーロッパにおいてポンド危機を生じさせ、ついにはドル・ショック＝金とドルの兌換停止に追い込んだものの正体である。

Finn Kydland(1943-) と Edward Prescott(1940-) のリアル・ビジネス・サイクルの理論は景気循環を実物ベースで議論するというものであり、「経済主体」の「期待」＝価格要素を経済モデルから排除しようというものである。その議論は貨幣の中立性、すなわち〈貨幣数量説〉を前提としており、その結論は景気循環はそれ自体合理的なものであるから、政府はなんら対策を行う必要がない、という観念的なものであった。

れた。

- 取り残された計量経済学はモデルの大型化によって生き残るわけにはいかなかった。変数の数の増大はモデル評価尺度それ自体の劣化をもたらすからである。そのため、何らかの事前分布＝緩い制約の導入によって、表向きは変数を増やしつつも実質的に変数の数が抑制された。これが (赤池の解釈による) ベイズ・アプローチの意義である。

ここからわれわれは未来の計量経済学の姿について次のような教訓を得るだろう。

- 計量経済学は、現実との接点をもつかぎり歴史科学であることを運命づけられる。つまり理論的な要素と歴史的な要素の結合物でなければならない。このばあいの歴史性とは、ドイツ歴史学派のようなものではなく、本質的な経済法則を取り扱う原理論とそれが特定の年代の国際情勢において発現する段階論との区別のことである。
- 計量経済学はそれ自体がイデオロギーとして理論からの制約を免れないこと。とりわけ、経済計画・政策の理論的な取り扱いが問題となる。プロレタリアートにとっては、価値法則が廃絶されたあとの過渡期社会の姿を明らかにすることが重要であり、〈擬制的価格〉＝労働支出時間のもとでの「架空の」(搾取率の意味合いを失った)〈剰余「価値」率〉(必要労働時間と総労働時間の比) の計測と予測・制御が重要となる。これは過去のスターリニスト官僚制国家の指令的経済計画とは異なるものとして想像されなければならない。

6.4.4　主観説

主観説を提唱したのは Frank Ramsey と Bruno de Finetti[*85]である。両者は 1930 年代に互いに独立に操作的主観確率の概念に到達した。ここで「操作的」と表現されるのは彼らの確率定義が「賭け」という観測手段に基礎づけられているからである。彼らの説が今日一定の支持を受けているのは、賭けの価格付けの整合性から確率の公理を導くという方針が論理的に一貫したものと受け止められているからである。

操作的主観確率の特質として以下のことが挙げられる。

1). Ramsey、de Finetti のいずれも従来の確率論のテキストに見られた〈充足理由律〉をパラドクスを呼び込む混乱の源泉として排除した。ここから、Keynes の主張する、主観的でありながらも単一の秩序 (合理性) をもつ確率 (論理説) を批判し、確率を評価する主体ごとに異なる確率 (主観説) の導入が必要になった。

2).「単一の合理性」の代わりに「主体ごとの整合性」(coherence または consistency) が重視された。この整合性を保証するものが、主体の受け取る損・得の観念である[*86]。

3). 主観確率は、観測者から独立した観測対象としての人間の脳髄の中にある。つまり、ここでの主観とは観測者の主観のことではなく、観測対象の主観のことである[*87]。この確率は直接に見ることはできないが、観測対象の行動として外に現れる (行動主義)。観測者は「賭け」を観測手段として用いることによってこれを知ることができる。

「賭け」はたとえば次のように構成される (新家 (1989) を参照した)。

[*85] Bruno de Finetti(1906-1985) はオーストリア生まれのイタリアの統計学者。大学卒業後、アクチュアリーとして活躍しながら研究を行い、その後アカデミズムに転じた。

[*86] 損・得を知る経済的人間が観測対象であることが前提とされている。なお、Ramsey はこれとは別に恐怖などの感情と主観確率の関係を検討したが、あたりまえの事象 (確率の大きい事象) にたいしてわれわれが恐怖を抱くことはない、という理由でその追及をあきらめてしまった。しかし、近代市民社会 (と市場経済) が成立する以前の社会において、主観確率の整合性を保証するものは何か、という問いかけは今後も追及するに値する問題である。

[*87] ベイジアンは、この観測者から見て「客観的に存在する」主観確率を観測者 (統計学者) の認識「主観」と誤解・混同しているが、この混同は興味深い視点を提示している。観測者 A は観測対象 B の抱く主観確率を計測しようとしているが、その主観確率の内容は、客観的な事象 E の生起確率である。図式的に表現すれば、$A \to (B \to E)$ である。だから上の誤解は $A \to B$ を $B \to E$ と (無意識のうちに) 同一視してしまっていることを意味している。

i). 参加者 B は a 円を支払ってくじを購入する。このくじは、

$$\begin{cases} \text{事象 } E \text{ が生起すれば、賞金 } s \text{ 円が払い戻される。} \\ \text{生起しなければ何も得られない。} \end{cases}$$

ii). このとき $p = a/s$ は参加者 B の事象 E の生起についての「主観確率」=「確からしさの測度」と呼ばれる[*88]。

iii). 整合性 (coherence) : 次の条件を満たさない主観確率は整合的ではない (このばあい、確実に利得・損失がある)。

$$\begin{cases} p \geq 0 \quad (s > 0, a \geq 0 \text{ より}) \\ p \leq 1 \quad (s \geq a \text{ より}) \end{cases}$$

整合性概念から確率の公理はたとえば次のように導かれる。

E_1, E_2 のいずれかの事象がかならず生起するが、両者が同時に生起することはないものとする。それぞれについての主観確率、賞金は p_1, p_2、s_1, s_2 であるとする。E_j $(j = 1, 2)$ が生起したときの参加者 B の獲得金額 g_j は次のとおりである。

$$g_j = s_j - (p_1 s_1 + p_2 s_2), \quad j = 1, 2$$

したがって、

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 - p_1 & -p_2 \\ -p_1 & 1 - p_2 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

s_j が未知数であるとしてこの解が存在するには次の条件が必要となる。

$$\begin{vmatrix} 1 - p_1 & -p_2 \\ -p_1 & 1 - p_2 \end{vmatrix} = 1 - (p_1 + p_2) \neq 0$$

このばあい、s_j について実際に連立方程式を解くことによって、A は任意の g_j をもたらすようなくじ s_j をつくることができる。つまり B が一方的に損をしたり得をしたりする賭けをデザインできる。これは B の主観確率が整合性をもたないことを意味するので、その可能性を排除するには $p_1 + p_2 = 1$ でなければならない。

確率の乗法法則についても次のような手順で導出できる。まず二つの事象 E_1, E_2 から次のような特別なくじ $\{E_1 \mid E_2\}$ を構成する[*89]。

$$\begin{cases} E_2\text{が生起する} \begin{cases} E_1\text{が生起する} \cdots g = (1 - p_0)s_0 \\ E_1\text{が生起しない} \cdots g = -p_0 s_0 \end{cases} \\ E_2\text{が生起しない} \cdots g = 0 \end{cases}$$

このくじの他に普通のくじ $\{E_1\}, \{E_2\}$ も考える。それぞれ主観確率は p_1, p_2、払戻額は s_1, s_2 とする。三つのくじを合わせて考えると、

$$\begin{cases} E_2\text{が生起する} \begin{cases} E_1\text{が生起する} \\ \cdots g = (1 - p_1)s_1 \\ \quad +(1 - p_2)s_2 + (1 - p_0)s_0 \\ E_1\text{が生起しない} \\ \cdots g = -p_1 s_1 + (1 - p_2)s_2 - p_0 s_0 \end{cases} \\ E_2\text{が生起しない} \cdots g = -p_1 s_1 - p_2 s_2 \end{cases}$$

これを行列表現すると次のようになる (便宜的に各行を g_0, g_1, g_2 と名付ける)。

$$\begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 - p_1 & 1 - p_2 & 1 - p_0 \\ -p_1 & 1 - p_2 & -p_0 \\ -p_1 & -p_2 & 0 \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ s_2 \end{pmatrix}$$

やはり整合性を維持するには次の条件が必要となる。

$$\begin{vmatrix} 1 - p_1 & 1 - p_2 & 1 - p_0 \\ -p_1 & 1 - p_2 & -p_0 \\ -p_1 & -p_2 & 0 \end{vmatrix} = p_1 - p_0 p_2 = 0$$

したがって、$p_0 = p_1/p_2$ である。これを事象に対応付ければ、次の乗法公式が得られる。

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

* * *

主観説の意義は、(外的な事象にかんする) 客観的な確率から (人間主体の意識内部で) 確率感覚がどのように形成されるか、そのメカニズムの本性を明らかにしたことである。唯物論哲学の世界観を前提とするかぎり、客観的な確率の存在が第一に認められなければならない。しかし、そのことと人間 (あるいは生物) がその生存の

[*88] つまり、これはレバレッジの逆数である。

[*89] くじは E_2 が生起した時点ではじめて購入される。

[*90] ばらつきのある個人的なリスク感覚が間主観的な確率論理に「収束」するには、損・得を評価する単一の「ものさし」が必要である。この意味でギリースが効用ではなく貨幣の意義を強調したのは正しい。効用のままではリスク感覚の個人性は是正されないままに終わり、社会全般に通用するものにはならないからである。また、実際のその個々のリスク評価を試し、是正する「場」=市場が必要である。

この論理をさらに商品市場に直面する商人資本の論理に応用することが可能である。ここでは〈一般的利潤率〉が市場価格と生産価格の相克から、均衡過程を通じて形成される。高嶋 (2019) を参照のこと。

ために一定の危険にたいする感覚を備えることとはまた別のことである。

この感覚は Keynes の合理的な (すなわち間主観的な) 確率論理にまで一足飛びに成長するわけではない[*90]。それ以前に個別の整合性をもったリスク感覚があり、さらにそれ以前の整合性のない (一方的な利得と損失を許容する) 原初的なリスク感覚 (反射) がある。こうした感覚の進化は、生命史においては自然選択を通じた生態の進化として現れ、原始社会においては労働の社会進化として発達した。またこの感覚は子供の知覚の発達として日々再生産されている。今日の株式市場に見られるトレーダーの行動はこの発達が究極の水準に到達した姿のひとつである[*91]。

このように考えるかぎり、主観説はその他の学説 (頻度説、傾向説など) にたいして排反する主張をなすものではない。そもそもこれらすべての確率学説は確率現象について互いに異なる局面を理論化しようとするものであり、それら相互の関係と適用可能領域を誤らないかぎりは、すべて正しいものと認められるべきである。

* * *

本章の最後に (ギリースが言及することのなかった) 情報と確率の関係に少しだけ触れる。

情報理論

〈情報量〉とは、主観確率の (ベイズ法則による) 更新に際しての「エビデンス」の重さを評価したものである (4.1.1 節を参照のこと)。「エビデンス」の存在いかんで客観確率が変わることはあり得ないので、これはさしあたりは主観確率を主題とするものだとわかる[*92]。

情報は Ramsey が否定したところの (確率にまつわる) 感情の大いさにかかわる。つまり主観確率の大きさそのものではなく、その更新の変化の度合いにかかわる。

情報概念は尤度理論の発展の中で徐々に形成されてきた。F.Gauss による最小二乗法の導出にその最初期の例がみられる。情報が大々的に使われるようになったのは、Fisher[*93]の"Theory of statistical estimation"(1925) 以来であり、彼はそこでフィッシャー情報量 (未知パラメータにたいしてデータの持つ情報) の概念を提唱した。

その後の情報理論は通信工学の展開、さらには通信ディジタル化と結びついたコンピュータ工学の発展とともに高度化した。たとえば 1920 年代のベル研究所での熱雑音の研究[*94]は 1948 年に Shannon[*95]の通信理論に集大成された。同年に Wiener[*96]がサイバネティクスを提唱し、通信工学と制御工学の統一を図った。アメリカ国家安全保障局 (National Security Agency:NSA) で暗号の研究を行った Kullback と Leibler[*97]は 1951 年に K-L 情報量 (divergence) の概念を生み出し、統計学を情報理論によって記述する基礎を作り上げた。赤池は K-L 情報量から 1971 年にモデル選択規準という発想を生み出した。

6.5 小括

本章の主要な結論は以下のとおりである。

- 確率の二面性 (可能性の度合いとしての/信頼性の度合いとしての) は古くから知られていた。しかし、両者がいかなる関係にあるのか、またどちらが重視されるべきかは、その時代ごとの〈思惟様式〉や、その当時に解決を迫られた理論的な課題に対応して変化した。

  i). 古代：安藤 (1997) によれば、古代ギリシャ

[*91] ただし、この感覚は統計学者や自然科学者のような科学的分析に立脚した理性的予測なのではない。あくまでも個人的・直観的な感覚 (「アニマル・スピリット」) にとどまるという意味で、人間疎外の一つの現れである。

[*92] 情報量はエントロピー概念 (ボルツマン・エントロピーなど) と深い関係があることから、情報量が客観的側面をまったくもたないとは言い切れないが、そのことについては本稿では留保しておく。

[*93] Ronald Aylmer Fisher(1890-1962) イギリスの遺伝学者、統計学者。ケンブリッジ大学生時代に Keynes らと優生学研究会を組織する。ロザムステッド農事試験場に勤務していたころに分散分析、各種の検定理論、実験計画法を考案し、集団遺伝学の研究を行った。

[*94] John Bertrand Johnson(1887-1970) と Harry Nyqvist(1889-1976) による。

[*95] Claude Shannon(1916-2001)

[*96] Norbert Wiener(1894-1964) はアメリカの数学者。アバディーン性能試験場で弾道学の研究を行い、射撃制御 (と哲学) の関心からサイバネティクスの発想に至る。ブラウン運動 (ウィーナー過程) の研究でも知られる。

[*97] Solomon Kullback(1907-1994) と Richard Leibler(1914-2003) はともに NSA の職員だった。

語にすでに確率のそれぞれの側面を言い表す言葉が用意されており、これらはいずれもAristotle によって論じられた。Aristotle は確率の客観的側面「テュケー」($\tau\acute{\upsilon}\kappa\eta$:偶運) の存在を認めていたものの、科学の対象としなかった。この理由は、テュケーがもし本質的原因というものをもつならば、それは必然ということになってしまい、その語の本来の意味に反するからである。

ii). 中世：(Aristotle が論理的な困難に直面して偶運を科学の対象から除外するにとどまったのにたいして)Augustinus は偶運の存在そのものを否定し、確率の客観的側面の追及を根絶やしにしてしまった。他方、確率の主観的側面は、客観的側面と対照的にキリスト教会において「決議論」(causistry) として隆盛をみた。決議論の主眼は、義務・戒律と人間的自由との関係を具体的事案に即して議論するためのものでもあった。

iii). 近世：確率の客観的側面を追及することはギャンブルの普及によってなしくずしに解禁された。Caldano(1576 の死の以前)、Galilei(1622) も賭けの問題を取り扱っており、ギャンブラーにとっては確率を理解する切実な動機があった。

フランスにおける 1650 年代のイデオロギー闘争の焦点は、封建的な伝統文化とブルジョアジーの新興文化のいずれに加担するかという問題として提起された。イエズス会は教皇の権威をより重視し、異端を摘発することを目的とした。ポール・ロワイヤル派はより世俗的な権威 (皇帝など) に依拠し、また異端の疑いを避けるために Hippo の Augustinus など古典に範を求めながらも、イタリアやイギリス、北方諸都市の先進的な文化を取り入れることを重視した。Blaise Pascal はDescartes の影響下から出発し、イタリアの実験精神に学びつつ独自の哲学を構築した。de Méré、Fermat との交流から学んだ組み合わせ数学の思考方法と客観的確率の考え方から「プロバンシアル」ではいい加減な決議論をもてあそぶイエズス会を批判し、「パンセ」では彼らにたいして客観的確率論を「神への賭け」(今日の用語でいえば、期待効用理論) としてたたきつけた。このとき、Pascal の頭脳内で確率の客観的側面と主観的側面とが並立した。

これ以降、確率の二面性は機械的唯物論の欠陥のゆえに後景に追いやられた。存在、即、認識をあらわす鏡的反映論が幅を利かせた結果、確率の客観的側面のみが当面の追及対象となり、主観的側面の方はあとから自然についてくるものとみなされた。ただし、確率の主観的側面は一時的に見えなくなっただけで、消え去ったわけではなかった。

iv). 古典確率論の形成：Laplace(1814) の与えた古典的確率の定義は彼自身のオリジナルのものではなく、先人たちの成果、とりわけde Moivre「運の測定について」(1771) に由来するものであった。それ以前には、Jakob Bernoulli にせよ、Christiaan Huygens にせよ、確率を明示的に定義することは避けて確率計算を行っていた。

Jakob Bernoulli が大数の弱法則を論証しようとした動機 (また彼が確率の明示的な定義を与えなかった理由) は、〈不充足理由律〉が相当に疑わしかったため、これを事後的な「大量観察」に置き換えようとしたためだった。そして大数の弱法則については苦心してその証明に成功したものの、肝心の確率の定義には復帰できなかった。これにたいして、de Moivre は最初から確率研究の対象を偶然ゲームに絞ったために、確率の定義を明示し得たし、またそれを出発点にして大数の法則をさらに精密にした中心極限定理の証明に至ることもできた。

- 現代：確率をめぐる思想は、古典確率論の成立を受けて出現したさまざまの技術領域と科学理論によって発展の素材を提供されながらも、その時々の「科学の危機」に応じて揺れ動いてきた。加えて、金融資本の成立と資本主義の帝国主義段階への変転が確率の主観的側面の問題への取り組みを

避けられないものにした。確率の二面性はそれまで背景に隠されていただけで、けっして無くなったものではなかったが、ここではことあるごとに表層に現れ、確率にかかわる難問を提起した。

古典的確率論の崩壊の起点は、Jakob Bernoulli によって名付けられた「不充足理由律」(Keynes の用語では「無差別の原理」) である。この原理はさまざまな逆理を生み出し、同じ状況にたいしていくらでも異なる確率値を与えるものだった。Keynes の論理説は「不充足理由律」を擁護し、確率論を論理学 (「蓋然説」) に純化しようとする試みであったし、von Mises の頻度説は Jakob Bernoulli の試みを引継ぎ、完遂させようとするものだった。このようなことが必要になった理由は、ひとえに確率論の対象を偶然ゲームに限定した de Moivre による応急措置がそれでは収まらなくなり、その適用範囲が次々に拡がったことにある。

おのおのの学説は確率現象について互いに異なる局面を理論化しようとするものであり、互いに排反する主張をなすものではない。それら相互の関係と適用可能領域を誤らないかぎりは、すべて正しいものと認められるべきである。

i). 頻度説 : von Mises の理論は〈不充足理由律〉を大量観察に置き換える Jakob Bernoulli の試みを引継ぎ、先験的な公理ではなく観測に重点を置くという視点の転換をもたらした。そのことによって偶然ゲーム以外の諸科学に確率論を拡大適用する根拠を与えた。頻度説の枢要な部分は傾向説へと引き継がれ、客観的に存在する偶然性の仮説が量子力学の解釈問題を通じて鍛えられることになった。

ii). 傾向説 : Popper の理論の意義は、客観的確率を二重に捉えたところにある。つまり、頻度としてのそれと傾向としてのそれであり、両者はさしあたり互いに別のものである。二重化された客観的確率は、推定・検定作業を経て最終的に同一視される。傾向としての確率は、その存在が客観的であるとは言え、それを確定的に表現することは出来ず、主観においてしか示すことができない矛盾したものである。しかしこの矛盾を認めないならば、Aristotle 的な禁止則に再びはまり込んでしまうことになる。

iii). 論理説 : Keynes の論理説は、Hume の因果推論の議論を確率分野に正しく拡大しようとする意図に支えられたものであり、その基本的な考え方は今日ではベイジアン・ネットワークの研究などに引き継がれている。そこでは〈エビデンス〉の意義は消えてしまうどころか、より一層重視されており、その意味で von Mises の頻度説は因果推論の不可欠な要素として論理説と一体化している。Keynes の説で強い批判を受けた部分は、先験的な確率判断を可能と見る〈無差別の原理〉、〈不充足理由律〉である。これは因果推論のなかでしぶとく生き残り続けた最後の観念論的要素である。

iv). 主観説 : de Finnetti の主観説の意義は、(外的な事象にかんする) 客観的な確率から (人間主体の意識内部で) 確率感覚がどのように形成されるか、そのメカニズムの本性を明らかにしたことである。唯物論哲学の世界観を前提とするかぎり、客観的な確率の存在が第一に認められなければならない。しかし、そのことと人間 (あるいは生物) がその生存のために一定の危険にたいする感覚を備えることとはまた別のことである。

第 7 章

結論

結論として、冒頭に掲げた幾つかの問いについての解答を与える。

> **Q1** なぜ確率の見方 (哲学) に主観説と客観説の対立があるのか。「確率は測度のひとつである」と言明することがこの対立にいかなる光明も与えないのはなぜか。

確率の二面性 (主観説と客観説の対立) は古代より認識されていたものであった (6.3 節)。それは、根源的には誤差、すなわち生産的実践 (実験) における予測 (主観) と観測結果 (客観) の対立に起因するものであり、下向分析の深化に伴い縮小されるものの、けっして汲みつくされることがない (2.2 節)。

確率の客観性は今日では確率の傾向説として量子力学の原理 (「ランダム仮説」) として定式化されている (5.3 節) が、それ以上の解明は将来の研究の進展に委ねられている。他方、確率の主観性は Hume 以来の因果分析の延長線上にあり、ベイジアン・ネットワークの研究として、あるいは主観確率の形成をめぐる行動分析学的な研究としてなお進行中である (3.3 節、6.4.4 節)。

こうして確率の二面性は (単純な対立としてではなく) 哲学における存在論と認識論の二つのモメントに対応するものとして、より高い水準での確率の統一的理解を約束するものとして、今もなおわれわれの眼前に置かれている。

* * *

Kolmogorov の測度論的確率論の意義は二つある。第一は古典的確率論によっては取り扱いが困難であった確率過程を、数学的に明晰なやり方 (公理論的方法) で取り扱うためのものであった (5.1 節)。第二に、確率の客観的側面の問いに答えるための von Mises の頻度説から操作的 (マッハ主義的) 定義の困難[*1]を取り除く試みであった。操作的定義を取り除いたあとでは、確率にかんする諸学説の対立から中立性を保つことに成功し、各学説から平等に頼りにされることになった (6.4.1 節)。

> **Q2** なぜ確率論の起源が Pascal の賭けの問題にあるとされるのか。それ以前に確率学説が発達しなかったのはなぜか。

Blaise Pascal が確率論の発展に果たした役割は、その見かけよりも深い。中世から近世への世界観の転換に際してのイエズス会とポール・ロワイヤル派のイデオロギー闘争において彼は後者に加担し、古来からの決議論に代えて期待効用理論を武器として戦った。またこれが可能になった前提として、ギャンブルの数理に取り組んだことにより確率の客観的側面について十分に深い理解を得ていたことが大きい (6.3.3 節)。

他方で、古代から中世にかけて確率学説が発達しなかった理由はギャンブルの禁止 (とそれによる公平な賭けを保証する「偏りのない」ギャンブル用具の未発達) に加えて、論理的な困難が Aristotle の哲学の中に既にあらわれていたことに求められる (6.3 節)。すなわち「チュケー」(偶有) は科学の対象になり得ず、なり得るとすると矛盾をもたらしてしまうこと、したがって矛盾を考察の対象とする哲学＝弁証法論理の不在が確率研究

[*1] ただし誤りと証明されたものではない。

の桎梏となっていた。この桎梏を取り除くには、認識主観の上でなく対象的存在 (ギャンブル) において偶有が日常的に観察されねばならなかった。

Q3 なぜ J.M.Keynes は確率論をもってその学究生活を始めたのか。そのことは彼の経済学説とどのような関係があるか。

Keynes はケンブリッジ・アポスルズの一人として Moore 的レッセフェール論と対決する必要性を感じていた。G.E.Moore の道徳哲学はその根本に不確実な事態の判断に際して (理性にたいする) 常識の優位を唱えており、これを批判するためには確率の論理説を構築する必要があった。彼の確率学説はその深いところで、当時の (そして今の) 経済的現実 (帝国主義) への直観に結び付いていた (6.4.3 節)。

＊　＊　＊

Keynes の経済学説は「1929 年恐慌の落とし子」であり、計量経済学会の創立 (1930) と結びつきこれ以降の経済イデオロギーの支配的潮流となった。その (「市場の失敗」への認識に支えられた) 理論は (東側の計画経済に対抗して) 経済の計画化を志向し、さらに行政など他分野の計画化をも促した。これらはすべて Keynes の最初の直観 (レッセフェールの否定) に沿ったものだった。

これらのイデオロギーはブレトン・ウッズ体制の終焉とともに二重の意味で瓦解した。第一には、戦後復興とともに膨張した「ホット・マネー」が為替管理を内側から食い破り、財政ではなく金融に力点を置く経済運営 (マネタリズム) に頼らざるを得なくなったことがある[*2]。安定した経済計画の前提であった安定した為替相場は放棄せざるを得なくなり、変動相場制に移行した。第二には、激変する経済環境が時系列分析への関心を生み出し、計量マクロモデルの巨大化 (経済変数の倍増) と推定精度の劣悪化がもたらされた。計量経済学は観測の限界をコントロールすることに失敗し、現実を説明するのではなく、「期待」なるものによって政策を正当化する規範的な理論にその席を譲り渡した。

Q4 近年のベイズ統計学の流行は何を意味するのか。

ベイズ統計学の流行には二つの見方があり得る。

第一はより実践的な見地からのものであり、観測の限界にかかわるものである。大量の潜在変数の存在 (逆に言えば、観測された変数の不在) が推測理論の武器である統計指標 (尤度) の機能を阻害する。そのため潜在変数 (パラメータ) を放任することなく、その自由度に緩やかな制約をかけなければならない。この制約がいわゆる「事前分布」をなす。事前分布は (ベイジアンが誤解するように) なんらかの先験的な公理によって決まるのではなく、分析者の恣意によって決めなければならない。その恣意 (アブダクション) の優劣は (機能を回復した) 尤度によって示されなければならない。このことを明確に指摘したのが赤池 (1989) であった (4.2 節)。

第二は確率の主観的な側面 (論理説と主観説) の研究の進展に伴って、確率の論理 (間主観的な確率評価) と感覚 (主観的個別的な確率評価) の形成が体系的に論じられるようになったことと関わる。理論の適用限界さえ誤らなければ、この流行はけっして間違ったものではない (6.4.4 節)。しかし、ベイジアン的教条主義に陥るならば、さまざまな誤り (BIC のような) の源泉となる危険がある (4.2.3 節)。

Q5 確率論と情報理論はどのような関係にあるのか。

KL 情報量 (divergence) は主観確率の更新式 (ベイズ公式) における「エビデンス」の重さにかかわる。F.Ramsey は確率と感情の関係性はないと (誤って) 結

[*2] 高嶋 (2019) を参照のこと。

[*3] 大いなる不確実性の存在は人間の身体を確実に蝕み、これが大いなる情動的反応 (不安、怒りなど) を引き起こす。そのことをわれわれは 2019 年 12 月ごろから始まった新型コロナウィルス (COVID19) 禍で体験している。

物資供給網の寸断、事業計画見直し (中止、無期限の延期、条件付き実施など)、行動制約 (集会や往来の自粛要請、休業要請から強制力のある外出禁止令まで)、感染の恐怖、無症状感染者への疑心暗鬼、政府報道・方針への不信などがストレッサとして諸個人を取り囲む。上の様々な状況は客観的な確率 (リスク) にしたがうが、我々自身はそれを知ることができない。考慮すべき行動選択肢は爆発的に増え、脳内のワーキングメモリーの疲弊、いわゆる「意思決定疲れ」を引き起こす。主観的確率評価は混乱して整合性を保てない。間主観的リスク評価は一致を見ない。あるものはリスクを過小に、あるものは過大に評価し、互いの不信感を増長させる。

論したが、実際には両者は情報量の大いさを介して結びついている (4.1 節)[*3]。KL 情報量は距離空間の公理を満たさないが、距離類似の性質をもち、確率分布を「点」とする一種の「幾何学」を構成する。これにより、統計的推測の各種のテクニックを統一した枠組みで理解できるようになった。

Q6 マルクス主義者は確率をどのように扱ってきたか。その自由論と確率論はいかに関係するか。

K.Marx 本人は (Hegel の思想的影響下にありながらも)、その学位論文で Epikouros の哲学 (とりわけその自由論と無神論) を共感をもって論じた (3.1.3 節)。当然、客観的に存在する確率については十分に認識をもっていたと考えられる。実際その片鱗が「資本論」第三巻の市場競争論、差額地代論の記述から読み取れる。これらの理論をさらに発展させる努力が続けられたならば、ボリシェビキの一部 (N.Bukharin、A.Bogdanov ら) の陥った「均衡論」的誤びゅうに陥ることはなかっただろう[*4]。

V.I.Lenin はその「唯物論と経験批判論」(これは Bogdanov らの批判に充てられた) のなかで、K.Pearson らが相関性を強調するあまり因果性を否定しさった姿勢を的確に批判している。しかし、それとともに因果性の哲学に実験家の精神で臨んだ D.Hume についても不可知論として片づけてしまった。これはその後の唯物論哲学が機械論に堕するひとつのきっかけをなすものだった (3.3 節)。

武谷三男は認識の三段階論を具体的に論じる過程で、とりわけ現象論から実体論への移行を議論するなかで、実験の意義について詳述した。武谷は自身も物理学者だったので、自らの実験家としての実践を省みることで、因子、偶然性、本質的な法則性の相互の関係を明確に記述できた (2.2 節)。

＊　＊　＊

マルクス主義者 (唯物論哲学) にとって自由論は最も困難な研究課題であったし、今もなおそうである。Marx は「フォイエルバッハ・テーゼ」で哲学の使命が世界の解釈ではなく、世界の創造にあることを宣言した。われわれはここに唯物論的目的論の形成の端緒を見ることができる。自由への希求は現実的自己の「不自由」の認識、すなわち労働の〈自己疎外〉の認識を前提とする。だから唯物論的目的論のためには「経済学・哲学草稿」に示された労働の〈自己疎外〉論が参照されるべきである。また、革命＝〈自己疎外〉からの脱却を果たしつつある人間が自己の労働をどのように意識的に発達させるかについては、必要労働と剰余労働の関係を必然と自由との関係に読み替える「ゴータ綱領批判」の視座が必要である (A.3 節)。

しかし、第二インターナショナル、初期コミンテルンの理論家のいずれも現実の階級闘争の提起する諸課題への対処に忙殺され、唯物論的目的論を構築する余力をもたなかった。戦時共産主義は自由ではなく必然の領域を強調することを強いた。計画経済は「資本家的無秩序」への反措定とされ、むしろ「自由」を束縛するものとして観念された。これに「均衡論」的誤びゅうが重なった。階級闘争は社会的実践＝自由の獲得という問題としてではなく、社会体制の変動として「外から」解釈され、その説明に物理学からのアナロジーが使われた。それが「生産関係」の交代を説明する「唯物論的」な、「物質的」な説明とされ、そのばかばかしさを指摘する者にたいして「観念論者」のレッテルが貼られた。こうしたこと一切が、Stalin と Hitler の協定によって完成された (A.4 節)。

武谷の技術本質論は Lenin の記述をもとに〈客観的法則性〉と〈目的意識性〉、すなわち必然と自由の関係を唯物論的に明らかにすることにはじめて成功した (2.1.3 節)。しかし、この功績は今日ほとんどかえりみられることなく埋もれてしまっている。これには武谷三段階論それ自体の欠陥も寄与している。このことを明確に示したのが黒田寛一であった (B.1 節)。

＊　＊　＊

こうしたストレスから身を守るための身体的反応としてまず情報 (エビデンス) を求める行動があらわれる。これによって主観確率をより信頼できるものに更新しようとしているのだ。これが叶えられない (例えば「積極的な疫学調査」からの限定的な検査方針) となると、さらなるストレスが生じる。やがてストレスが制御できなくなり、自暴自棄の行動、他者への攻撃、感覚の遮断、うつ、などが生じる。

[*4] しかし、「金融資本論」の著者 R.Hilfading が〈価値形態論〉をほとんど理解しなかったために、この道は閉ざされていた。やはり高嶋 (2019) を参照のこと。

今後の研究課題として次のことが挙げられる。

- 武谷の認識の三段階論のさらなる展開が必要である。とりわけ、1) 実体論から本質論への移行の論理、2) 上に関連して、架空の実体 (フロギストン、エーテルなど) の止揚の論理を熱力学、相対性理論の形成から探ること、3) エネルゲティク、マッハ主義の内在的批判。
- 量子力学にかんする長澤の理論の検証と是正 (Nelson の確率力学との統合)。演算子法＝作用素代数の本質的理解。不確定性原理の解釈、相対論との関わり、保存則と対称性。
- 確率をめぐるパラドクスの解明と行動経済学、行動分析学との関係。
- 確率変数について、質的 (離散型) と量的 (連続型) の混在状況の理解。空間の分割。
- EM アルゴリズム (赤池の解釈によるベイズ・モデリング) の体系的適用 (とりわけ因子分析、LDA など)。甘利の特異分布族との関係。
- ブートストラップ法と経験尤度法の関係。

参考文献

[1] Dempster,A.P., N.M.Laird and D.B.Rubin, Maximum Likelihood from Incomplete Data viaq the EM Algorithm, Journal of the Royal Statistical Society, Series B(Methodlogical), Vol.39, No.1, pp.1-38, 1977

[2] Edwards D. , “Introduction to Graphical Modelling”, Springer Texts in Statistics, Springer-Verlag New York, Inc., 1995

[3] Efron,B. and R.J.Tibshirani, ”An Introduction to the Bootstrap,” Chapman & Hall CRC, 1994

[4] Hayek,F.A.,”The Road to Serfdom : with The Intellectuals and Socialism”, The Institute of Economic Affairs, 1945

[5] Højsgaard S., D.Edwards and S.Lauritzen, “Graphical Models with R,” Springer,2012

[6] Kullback,S.,”Information Theory and Statistics”, Dover Publications,Inc.,1968

[7] McCauley,J.L., Fokker-Planck and Chapman-Kolmogorov equations for Itoh processes with finite memory, (in Skjeltorp,A.T. and A.V.Belushkin,”Evolution from Cellular to Social Scales”,Sprimger),99-110,2008

[8] Ng E. W. and M. Geller, A Table of Integrals of the Error Functions, Journal of Research of the National Bureau of Standards - B, Mathematical Sciences, Vol.738, No.1, 1968

[9] Owen, A. B., ”Empirical Likelihood,” Chapman & Hall CRC, 2001

[10] Pauling,L. and E.B.Wilson,”Introduction to Quantum Mechanics with Applications to Chemistry”,McGraw-Hill Book Co.Inc.,New York,1935

[11] Pearl,J., Causal diagrams for empirical research (with Discussions), Biomtrika, 82, 4, pp.669-710, 1995

[12] Pearl,J.,「統計的因果推論 ―モデル・推論・推測―」(黒木学訳), 共立出版,2009

[13] Seal, H.L.,The historical development of the use of generating functions in probability theory, Bulletin / Association of Swiss Actuaries,49,1949

[14] Shachter,R.D.,Bayes-Ball: The Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams),Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence,2013

[15] Tikka,S., and J.Karavanen, Identifying Causal Effects with the R Package `causaleffect`, the Journal of Statistical Software, 2017

[16] Tipping M.E. and C.M.Bishop, Probablistic Principal Component Analysis, Technocal Report NCRG/97/010, Neural Computing Research Group, Dept. of Computer and Applied Mathematics, Aston University,1997

[17] Trotsky,L., J.Dewey and G.Novack,”Their Morals and Ours – Marxist vs. Liberal views on Morality”, 5th Edition, Pathfinder Press, 1973

[18] Ying Y. , Empirical Likelihood Inference for Two-Sample Problems, A thesis presented to the University of Waterloo in fulfillment of the thesis requirement for the degree of Master of Mathematics in Statistics, Waterloo, Ontario, Canada, 2010

[19] 赤池弘次, 事前分布の選択とその応用,(鈴木雪夫・国友直人編,「ベイズ統計学とその応用」, 東京大学出版会,1989)

[20] 赤池弘次, 甘利俊一, 北川源四郎, 樺島祥介, 下平英寿,「赤池情報量規準 AIC – モデリング・予測・知識発見」, 共立出版,2007

[21] 赤池弘次,AIC と MDL と BIC、オペレーションズ・リサーチ, 日本オペレーションズ・リサーチ学会、1996

[22] アドルノ, T.W.,「三つのヘーゲル研究」(渡辺祐邦訳), 筑摩書房,1963

[23] 安藤洋美, 古典確率論の歴史の諸問題 (数学史の研究), 数理解析研究所講究録 (京都大学),1019.40-60,1997

[24] 石川知宏, 狂愚の国のソロモン—パスカルの偽名について—, 人文学報 (首都大学東京/東京都立大学),420 巻き,2009

[25] 伊藤清,「確率過程」, 岩波書店,2007

[26] 岩崎允胤, 均衡論について, 哲学 (日本哲学会), 第 8 号,1958

[27] 内井惣七, ラプラスと確率の哲学 (解説),([69] の解説),1997

[28] 汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功,「計算統計 I—確率計算の新しい手法 (統計科学のフロンティア 11)」, 岩波書店,2003

[29] 菊地進, 同時方程式モデルとその計算方法の展開について : 手法の開発からモデルの大型化まで, 立教経済学研究,Vol.36,No.2,1982

[30] 菊地進, 計量経済モデルの大型化の一帰結, 立教経済学研究,Vol.49,No.4,1996

[31] 北川源四郎, 情報量規準と統計的モデリング,(20 の第 II 篇第 2 章),2007

[32] 木村洋, 第二次世界大戦前に於ける日本人数学者の戦時研究 (数学史の研究), 数理解析研究所講究録 (京都大学),1257.260-274,2002

[33] ギリース,D.,「確率の哲学理論」(中山智香子訳), 日本経済評論社,2004

[34] 黒田寛一,「スターリン批判以後　下 (1995 年〜1962 年)」, 現代思潮社,1977

[35] 黒田寛一,「ヘーゲルとマルクス　技術論と史的唯物論・序説」, 現代思潮社,1968

[36] 小西貞則, 北川源四郎、「情報量規準」, 朝倉書店,2004

[37] 坂元慶行, 石黒真木夫, 北川源四郎,「情報量統計学」, 共立出版,1983

[38] 坂元慶行,「カテゴリカルデータのモデル分析」, 共立出版,1985

[39] シェイファー,G., ウォフク,V.,「ゲームとしての確率とファイナンス」(竹内啓・公文雅之訳), 岩波書店,2006

[40] 重川一郎, 伊藤清先生の業績紹介, 同窓会設立報告号 (京都大学理学研究科・理学部数学教室),57-69,2016

[41] 下平英寿, 伊藤秀一, 久保川達也, 竹内啓,「モデル選択 予測・検定・推定の交差点」, 岩波書店,2004

[42] 新家健精, 主観確率,(鈴木雪夫・国友直人編,「ベイズ統計学とその応用」, 東京大学出版会,1989)

[43] 新谷明雲, 非定常シュレーディンガー方程式の対数微分法による求積—簡単なポテンシャルの中での波束の運動, 山口県立大学学術情報 (共通教育機構紀要), 第 3 号,2010

[44] 武谷三男,「弁証法の諸問題——武谷三男著作集 1」, 勁草書房,1968

[45] 武谷三男,「物理学入門　力と運動」, ちくま学芸文庫,2014

[46] 高嶋裕一, デフレ経済の解明——貨幣数量説批判を手掛かりとして——, 岩手県立大学総合政策学会 Working Papers Series, No.139,2019

[47] 高嶋裕一, 民族問題と労働の自己疎外——史的唯物論における民族・試論——, 岩手県立大学総合政策学会 Working Papers Series, No.122,2017

[48] 高嶋裕一, 共分散選択モデルに対する赤池情報量規準の有限修正, 岩手県立大学総合政策学会 Working Papers Series, No.95,2014

[49] 高村友也, 現代量子力学から見たポパーの傾向性解釈, イギリス哲学研究, 第 33 号,2010

[50] 竹内啓, 大橋靖雄,「統計的推測—2 標本問題」, 日本評論社,1981

[51] 戸坂潤,「現代哲学講話」, 白揚社 (青空文庫),1934

[52] 長澤正雄,「マルコフ過程論による新しい量子理論」, 創英社/三省堂書店,2015

[53] 日本品質管理学会テクノメトリックス研究会編,

「グラフィカルモデリングの実際」, 日科技連, 1999

[54] ビショップ,C.M.,「パターン認識と機械学習・ベイズ理論による統計的予測 (下)」(元田浩, 栗田多喜男, 樋口和之, 松本裕治, 村田昇監訳), 丸善出版,2012

[55] フォスター,J.B.,「マルクスのエコロジー」(渡辺景子訳), こぶし書房,2004

[56] ボレル,E.,「確率と確実性」(彌永昌吉, 高橋禮司訳), 文庫クセジュ (白水社),1952

[57] 本多修郎,「近代数学の発酵とヘーゲル弁証法」, 現代数学社,1989

[58] 三木清,「三木清全集 第一巻」, 岩波書店,1966

[59] 三木清,「三木清全集 第十八巻」, 岩波書店,1968

[60] 三木清,「三木清全集 第十九巻」, 岩波書店,1968

[61] 宮川雅巳, 「統計的因果推論—回帰分析の新しい枠組み—」, 朝倉書店, 2004

[62] 宮川雅巳,「グラフィカルモデリング」, 朝倉書店,1997

[63] 宮部賢志, ランダムネスの一般化, 数理解析研究所講究録 (京都大学),1729.84-94,2011

[64] 森庄栄, 戦後日本の統計学の発達—数量化理論の形成から定着へ—, 行動計量学, 第 32 巻第 1 号 (通巻 62 号),45-67,2005

[65] 山本哲朗, 「行列解析の基礎—Advanced 線形代数」, サイエンス社, 2010

[66] 山本哲朗, 「行列解析ノート—珠玉の定理と精選問題—」, サイエンス社, 2013

[67] 柳川堯,「離散多変量データの解析」, 共立出版,1986

[68] ラオ,C.R.,「統計学とは何か ; 偶然を生かす」(藤越康祝, 柳井晴夫, 田栗正章訳), ちくま学芸文庫,2010

[69] ラプラス,P.-S.,「確率の哲学的試論」(内井惣七訳), 岩波文庫,1997

[70] リウ,C.L.,「組合わせ数学入門 I」(伊理正夫, 伊理由美訳), 共立全書,1972

[71] レーニン,V.I.,「唯物論と経験批判論」(寺沢恒信訳), 大月書店,1953

付録 A

商品論における価値の〈実体〉

武谷は〈実体〉という用語の典拠の一つが「資本論」冒頭の商品論であると説明している。ここでは、武谷の三段階論と商品価値の下向分析とのかかわり (図 A.1) を整理しておく[*1]。

A.1　商品価値の下向分析

まず、資本制社会の富が「巨大なる商品集積」であること、個々の商品を〈現象論〉的に見れば、a) 有用物としての具体的な商品体であること (ただし、それは交換価値の「素材的な担い手」であって、その有用性は他人のためのものにすぎず、生産者はそれを自分で使うことはできない)、b) ほかの様々な商品と交換され、その交換比率が偶然的なものに見えること (ただし、商品が互いにどんなものにでも交換されるという事実は、そこに本質的なものが隠されていることを暗示する)、が指摘されている。

次に、商品が〈実体論〉的に、それに「対象化された労働」の観点から把握される。それが有用性を持つということは、それが指物労働、建築労働、紡織労働などの〈具体的な有用労働〉(Work) の産物であることを示す。ここで労働は「互いに異なる」という質的な側面からのみ理解されている。この質的側面が捨象され、〈抽象的な人間労働〉(Labor) が見出され、これが価値の〈実体〉とされる。ここには質の要素がまったくなく、純粋に量的にのみ (すなわちその継続時間としてのみ) 取り上げられている。ここに「質と量の弁証法的な関係」が表現されている。

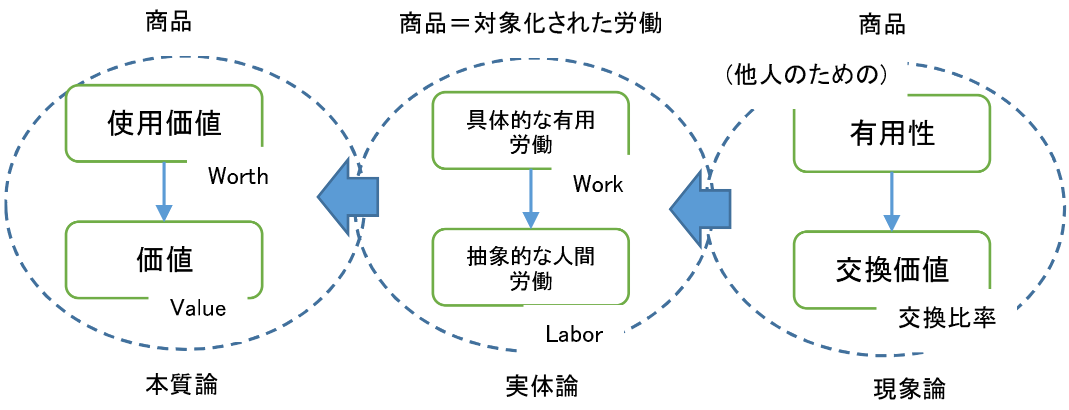


図 A.1　商品価値の下向分析

*1 高嶋 (2019) も参照のこと。

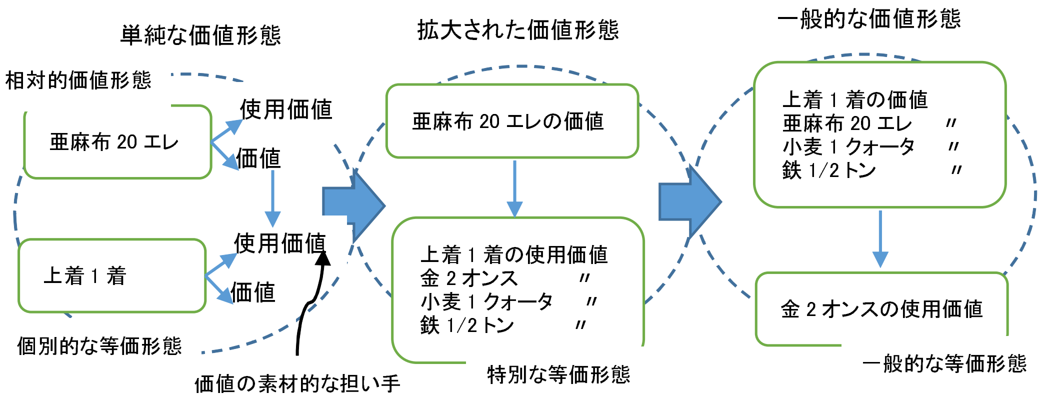


図 A.2　上向的展開＝〈価値形態論〉

最後に〈本質論〉的段階において、商品は〈使用価値〉(Worth) と〈価値〉(Value) が統一されたものとして把握される。商品は等価交換を通じて互いに交換されており、商品にとって大事なものは〈価値〉である。本来人間にとって大事なものは〈使用価値〉のほうであるが、商品にとってそれはどうでも良い。ここに資本制社会の倒錯が如実に示されている。

〈価値形態論〉(図 A.2) は、上の下向分析とは反対に、抽象的なものから出発して「巨大なる商品集積」の全体像にまで復帰すること (上向的展開) が意図される。〈価値形態〉とは、一つの商品〈価値〉が別の商品体の〈使用価値〉によってどのように表現されるかの論理を示す。最初に単純な価値形態 (個別的な等価形態)、次に拡大された価値形態 (特別な等価形態)、最後に一般的な価値形態 (一般的等価形態) が説明される。この一般的等価こそが〈貨幣〉である。

なお〈価値形態論〉の叙述の順序は必ずしも歴史的現実の過程をそのまま示すものではない。つまり共同体と共同体との間に偶然的に行われた余剰生産物の交換が、交易の拡大とともに〈貨幣〉を生み出した、というわけではない。というのも、単純な価値形態といえども最初から資本制商品の倒錯した性格を備えているのであって、決して先資本制社会[*2]における余剰生産物一般の性質に解消されるものではないからである。

A.2　〈価値法則〉の廃絶と自由

プロレタリアートは〈価値法則〉を認識したとしても、それを適用 (「利用」) することはできない。いわんや、〈価値法則〉を残存させながら、みずからの自由を得ることはできない。けだし、〈価値法則〉それ自体が彼の不自由を、「隷属」のくびきを意味するからである。これをそのままにして、いくら自然の〈客観的法則性〉への認識を深めたとしても、彼が奴隷の身であることには変わりがない。それどころか、〈生産力〉の増大は、資本の〈生産力〉の、自然 (〈生産諸手段〉としての大地と彼自身の人間的自然) への「破壊力」の増大として〈現象〉することになる。

〈価値法則〉がプロレタリアートのくびきであることの証拠の一端は図 A.1 にすでにあらわれている。つまり価値の〈実体〉をつくりだすものが彼の疎外された労働であること、自己の労働を自分のためには使えず、他人のために支出しなければならないこと、またその立場はたんなる〈抽象的な人間労働〉の提供者にまで零落し、そこには彼の個性、彼の人間性の入り込む余地はまったく残されていないこと、などが一目瞭然となっている。

このようなプロレタリアートの隷属がなぜ生じるのかと言えば、それは彼が〈生産諸手段〉から疎外され、みずからの生存のためにはみずからの労働時間を切り売りするしかないからである。だからプロレタリアートが彼

[*2] そもそもそこには、その使用 (消費) が「抽象的な人間労働」の支出となってあらわれる特別な商品たる〈労働力〉(Labor)、すなわちプロレタリアートが存在していない。

自身の自由を勝ち取るためには、〈生産諸手段〉の私有制をやめさせることが絶対の条件となる。その結果として、(階級廃絶とともに国家が「死滅」するのと同じように)〈価値法則〉そのものが「死滅」する。この意味で、プロレタリアートが〈価値法則〉を「利用」することはありえない。

A.3　いつわりの「自由主義」

資本制社会は、別名「自由主義」社会とよばれる。しかし、そこでの「自由」は、疎外された労働の現実が放置されている以上、みせかけのものでしかない。つまり、われわれは本当の意味では「自由」ではない。

この社会が「自由主義」であると呼ばれるゆえんは、われわれがすべて法のもとに「平等」であり、市場においては「自由」な商品所有者として互いに相対するからである。商品・富は私有財産として蓄積され、われわれは (もし持ち合わせているならば) これを自分の意志で処分する「自由」をもつ、とされる。

しかし、この「自由」は形式的なものでしかない。商品所有者としては万人は商品の奴隷である*3。また特にプロレタリアートには、生産諸手段から切り離されているという「自由」(不自由) と、自己の労働力を商品として売り渡す「自由」(不自由) があたえられている。

この社会はたしかに諸階級を封建的な生産諸関係から解放したのだとは言えようが、その出自においては絶対主義権力の手を借りたこと、その行く末においてはファシズム*4を生み出さざるを得ない宿命にあることを忘れてはならない。

これら形式的な「自由」にたいして、われわれはプロレタリア民主主義 (またその基礎としてのプロレタリア自由主義) を対置しなければならない*5。プロレタリアにとって、自由とは共同管理された生産諸手段へのアクセスの自由であり、言い換えれば「職場選択の自由」にほかならない。ただし (しかも)、個人が生産諸手段を排他的に占有することは許されない。職場は官僚的計画によって選ばれるのではなく、みずからの自由意志によって選ぶ。国籍・人種・性別・出自 (祖先)・居住地などによってはその選択はさまたげられない。「思想・信条・宗教」によってさえ差別されない。それらは、個人的な趣味 (「私事」) であり、生産的実践の根幹にかかわらない限り、それらは放っておかれる*6。

「職場選択の自由」は、その基礎に〈労働証書制度〉が敷かれていることが前提になる。つまり、〈価値法則〉は「死滅」しており、そのために自己増殖する〈価値〉たる〈資本〉は存在しない。生産物はもはや商品となることはなく、〈労働証書制〉のもとでの〈等量労働交換〉*7が社会的な秩序を内発的に形成する (だから官僚的計画は不要となる)。

ただし、こうした〈共産主義の第一段階〉に至るまでのあいだに〈過渡期社会〉を通らなければならない。ここではプロレタリア民主主義 (あるいは「コミューン型国家の原則」) が適用される。プロレタリアートのみに公務員＝議員への選挙権・被選挙権が与えられる*8。生産諸手段を管理する公務員の給与は他の労働者並みとされ、またいかなる特権も与えられない。公務員の仕事がどれだけ共同体にとって重要であったとしても、〈過渡期社会〉＝〈労働者国家〉が存続するあいだの臨時的なものであり、「国家の死滅」とともにいずれ消滅することが予告されている。

*3 商品所有者は「商品に欠けている商品体の具体的なるものにたいする感覚」を補い、商品の価格を代弁するものにすぎない。そして、商品のたましいは〈価値〉であり、自己増殖する〈価値〉が〈資本〉であった。

*4 固定資本更新のタイミングのずれは、資本制社会を必然的に帝国主義の段階に引き上げる。資本の過剰を資本制社会の内部では処理できず、外的な政治的な力に頼らざるを得ない。幼年期の資本主義 (重商主義) が絶対主義権力に依拠して生誕した (政治的力→経済的力) のと対照的に、その老年期においては経済的力が政治的力 (帝国主義) に転化する。帝国主義戦争は帝国主義政策の延長線上にある。こうした状況は先鋭な階級闘争を惹起せずにはおかないが、それを抑え込む、あるいは未然に防ぐためにファシズム統治体制が用意される。

*5 これに関して高嶋 (2017) も参考のこと。

*6 ただし、事が共同体の運命にかかわるものならば、それは放任されない。意見表明する権利があるどころか、むしろそれを公言せずに黙っていることの方が罪深い。

*7 〈等量労働交換〉は〈等価交換〉と同じものではない。まず労働の質は考慮されず、純粋に労働量 (継続時間) のみによって計られる。また死んだ労働時間たる〈不変資本〉はそもそも存在せず、生産物の中に移転されない。

*8 「プロレタリアート独裁」:形式的にはこれは普通選挙となんら変わらない。ただ選挙権を得るために「生産諸手段としての」私有財産を放棄することが求められているだけである。また、立法機能と行政機能は分離されない。ブルジョア民主主義の「三権分立」はフランス革命当時の階級状況を反映したものにすぎない。

A.4 「自由主義者」のいつわり

マルクス主義者は果敢に「自由主義」イデオロギーの欺瞞性をあばき立ててきたが、これに対抗して各種の異議申し立てが行われてきた。ここではその代表的なものとして Hayek(1945) を取り上げる。

Hayek はウィーン大学で Ludwig von Mises[*9]に学び、ナチスの政権掌握にともないアメリカに亡命、シカゴ大学に移った。またモンペルラン協会[*10]の中心的創設者として、今日の「新自由主義」のイデオロギー的支柱を用意した。

Hayek の主張は、次の諸点からなる。

1). ドイツのナチス (国家社会主義) とロシアを中心とするソビエト連邦とを同じ「社会主義」に属するものとして (P.Drucker[*11]の発言に同調しつつ) 同一視したこと。これはドイツの二つの政権 (ヴァイマル共和政:1919-1933 とナチス・ドイツ:1933-1945) を連続のものと見る視点に支えられている。両者の類似性・親近性は、かつての「社会主義者」が容易にファシストに転向したという事実をもって「証明」されている。その上で、「民主的な社会主義」とは形容矛盾であり、達成不可能なユートピアだと断じている。

2).「計画」と「競争」を対立するものと捉え、前者を悪、後者を善となしていること。その理由は「計画」が大衆にたいする指令 (とその「自由」の剥奪) を意味するからである。「計画」を強制するためには強大な権力 (独裁) を必要とし、またそれを必然とする。私企業においても「計画」は存在するが、それは同じ悪であるとしても大衆への影響は小さく、またそれは「競争」のためのものであるから容認できる。国家の「計画」はこれにたいし、大衆への影響が破滅的に大きく、しかも「競争」を排除する。

3).「競争」がなぜ善なのかと言えば、それは「競争」が私有財産の (集中ではなく) 分散されている状態を示すからであり、そして私有財産こそが「自由」を保障するものだからとされている。私有財産の集中 (集産体制) は「計画化」と同義であり、個人の持つ権力を国家に譲り渡すこと＝放棄することに等しい。

4). プロレタリアート独裁を専制政治と同一視していること。権力は制限されなければならない、ということを至上の命題として、「計画」が単一目的に仕えるからという理由 (戦時体制のアナロジー) でこれを否定する。

5). ファシズム (と「社会主義」) が力を持ったのはなぜか。「無教養」で「単純」な大衆はシンプルなスローガンを好むためである。大衆は肯定的 (建設的) な綱領 (プログラム) ではなく、否定的 (攻撃的) な綱領を好む。その典型が内部の敵を名指しするプロパガンダであり、ドイツではユダヤ人が、ロシアでは富農 (「クラーク」) がスケープゴートとされた。人間の個性は変わることはなく、(F.Knight[*12]に倣い) 善意の人間よりは悪意ある人間が計画家として上に立つ可能性 (確率) が大きいと断じる。そうした人間のために大衆を奉仕させるためには、情報操作が不可欠であり、そこには真実は認められない。

以上のような Hayek の主張の反プロレタリア性を指弾し、断罪することは容易である[*13]。しかしそれ以上に重要であるのは、なぜこのような主張がいまもなお力を持ち、一定の支持を得ているのか、その秘密を明るみに出し、内在的に批判することにある。Hayek の所説の流布は、マルクス主義哲学がスターリニズムへと変質し、このような批判を正面からおこなうだけの生命力を喪失したことの反映である。

[*9] Ludwig Heinrich Edler von Mises(1881-1973) はオーストリア帝国財務省の官吏であり、ウィーン大学教授でその後亡命、ニューヨーク大学に移る。Carl Menger の系列に連なる経済学者。工学者 Richard von Mises の兄。

[*10] 1947 年に経済学者を中心に創設された団体。そのメンバーはノーベル記念経済科学賞 (いわゆる「ノーベル経済学賞」) 委員会 (Committee for the Prize in Economic Sciences in Memory of Alfred Nobel) の委員にも選任され、同賞の性格にも影響を及ぼした。

[*11] Peter Ferdinand Drucker(1909-2005)、ウィーン出身の経営学者。ロンドンで Keynes に学び、その後アメリカに移住。

[*12] Frank Hyneman Knight(1885-1972) は経済学者でシカゴ学派の創始者の一人。リスクと不確実性を区別したことで知られる。

[*13] Hayek は一人のマルクス主義者の発言さえも引用せず、またマルクス主義者のとった行動一つさえも特定することなく、それを非難するという不誠実さを見せている。

A). Hayek の言うナチスとソ連邦の現象的類似性は、スターリニスト党のジグザグの結果によって生じた。これを決定的にしたものが 1939 年の独ソ不可侵条約 (Molotov = Ribbentrop 協定：ファシストとの同盟) であった。またそれ以前でも、右翼的政策 (NEP)[*14]から左翼的政策 (農業集団化)[*15]への転換があり、そのたびに反対派が脱落し、これが「転向者」の豊富な供給源をなした。スターリニスト党が目的とするものは、労働疎外の廃絶ではなく、「一国社会主義」＝ナショナリズムの鼓舞である。彼らは国家フォンドを官僚として占有し、そのために国家フォンドを至高の存在とする態度＝物神崇拝に陥った。労働疎外の廃絶 (価値法則の廃棄) は彼岸化された。これは目的 (疎外された労働の廃止) と手段 (私有財産の廃止) の取り違えを意味する。

B). スターリニストは、ファシストとの同盟を許容してしまうほど、ファシズム統治形態の本性を理解しなかった。ファシズムは帝国主義の積極型 (ドイツ) と消極型 (イギリス) の対立、その結果としての第一次世界大戦の戦後状況から生み出された。ロシア十月革命は第一次世界大戦にたいする反戦運動から直接的に成長した。ところがドイツ革命は挫折し、ヴァイマル体制という妥協となって終わった。ドイツ帝国主義 (またその先兵であるフライコール) は、第一次世界大戦でのドイツの敗北を「背後からの一刺し」によるものとしてプロレタリアートへの階級的憎悪をあらわにした。これがファシズムである。

C). スターリニストは、国家フォンドの占有に官僚的な利害を見いだすがゆえに、プロレタリア国際主義 (国家の死滅) を彼岸化する。彼らが生産力を強化するための「計画」に明け暮れるのは、国家フォンドを防衛するための戦争 (「総力戦」) に必要だからである。そこではロシア十月革命を準備した反戦思想、すなわち「革命的敗北主義」(Lenin) も「平和のための闘争」(Trotsky) も忘れ去られている。

D). Hayek の言う「計画」と「競争」との対立は、スターリニストが疎外された労働の廃絶をネグレクトした結果であり、両者の弁証法的連関[*16]を理解せず「計画」と「競争」の現象論的・実体論的な把握を固定したことの反映である。Hayek は「競争」の側から「計画」を攻撃し、「社会主義者」は「計画」の側から「競争」を非難する[*17]。両者はたんにお互いの鏡像をたたき合っているにすぎない。

「計画」と「競争」との対立は次のように止揚される。まず、職場は生産諸手段と分かちがたく結びついており、そこに個人が自由にアクセスできることが前提となる。生産諸手段は誰のものでもあり、同時に誰のものでもない。職場は生産諸手段そのものであり、個人にとってそこで生き、活動する作業環境であり、同時に生産物＝作品でもある。個人は、生産諸手段から排除されないのだから、そもそも職場内で競争する動機をもたない。生産活動の目標 (「計画」) は職場内の討議をもとに作られ (頭脳労働)、その結果として共有されるのであって、国家官僚から指令されるべきものではない。その達成のための各人の創意工夫を「競争」と呼べばそう呼べないことはないが、その成功も失敗も職場において共有される。各人の取り分は支出された労働時間 (労働量) でしか区別されない。

職場間の「競争」については、地代論が参照される。等量の労働にたいして、異なった生産物の結果が生じる。しかしこの差は、質的なものとしてそのまま受け入れられ、量には還元されない。これは、価値法則が廃絶されているのだから、質の違いが特別利潤＝地代には転化しないからであり、一般的利潤率 (そもそも存在しない) の形成にも寄与しないからである。ここでは、量をめぐっての「競争」はなく、強いて言えば質をめぐる

[*14] 内戦と戦時共産主義 (1918-1921) からの経済の立て直しのため、食糧現物税が導入され、余剰農産物の自由販売が容認された。また一部企業が合法的に再私有化された。これが「資本主義的要素」の容認 (国家独占資本主義) と受け止められた。「農民よ、富め」をスローガンとするブハーリン派が台頭した。

[*15] 工業化による工業製品と農産物との「価格差」の結果として、穀物調達危機 (1927-1928) が起きた。第一次五ヵ年計画 (1928) が立案され、コルホーズへの農業集団化が開始され、1934 まで続いた。

[*16] これは Hayek の言う私企業における「計画」やいわゆる「競争政策」とは違う。

[*17] Engels による「生産の社会的性格と取得の資本主義的性格の矛盾」にその誤びゅうの根源を見いだすことができる。

「競争」がある (農産物での産地ブランドなどから類推されよう)。

職場間の「計画」は、労働量をどのような職場＝活動に振り向けるべきか、にかかわる。戦時共産主義という例外的状況を除けば (あるいはそこにおいてさえも)、やはり指令ではなく、プロレタリア民主主義が貫徹されねばならない。原則として、〈労働証書制度〉にもとづく〈等量労働交換〉があり、これを意識的に実現するものとして計画がある。この計画は何から何まで決めなければならない、というものではなく、もっぱら Marx が「ゴータ綱領批判」に記したような各種の控除にかかわる。またその控除分の全労働量に占める比率 (すなわち計画の範囲と規模) は、生産力の増大とともに小さくなる。

外的な強制・指令が無用のものである以上、「競争」と「計画」の対立は、労働の質と量の弁証法的な発展過程のなかに解消されていくだろう。労働の質としては、労働者の内面の能力 (芸術性と創意工夫) の進歩として、労働の量としては、必要労働の縮小 (自然を飼いならしたという意味での〈生産力〉の増大) と真の意味での自由の拡大として。

付録 B

唯物論的目的論

B.1 黒田 (1977) による武谷三段階論批判

武谷の三段階論に対する批判は多数存在するが、その理論の積極的な意義を認めたうえでその創造的な発展を志向したという意味でおそらく最も誠実な批判は黒田 (1977) のそれであろう。

> ・・・あらゆる分野において、生きた現実的諸課題に答えることができないような唯物論哲学は、そのことによって、もはや唯物論哲学であることをやめてしまう。そのような意味で武谷は「哲学の有効性」[*1]の問題を提起したのであった。・・・解釈主義とドグマティズムの泥沼にはまりこみ続けてきた現代唯物論の現状にたいして、それは極めて実践的な意義を担っていた。(と同時にそれ〔注 :「哲学の有効性」というスローガン〕は・・・武谷理論の限界を示すものでもあった。)
>
> 武谷三段階論の実践的性格にもかかわらず、その理論戦線への浸透力が限られたのは一体なぜなのだろうか?・・・その提出の仕方それ自体における欠陥、すなわちその提起の形態の直観性と理論としての不完全な完結性が三段階論を普遍化するためにはそのブレーキとして働いた・・・。
>
> 具体的に言うならば、
>
> - 武谷論理学は、これまで「唯物弁証法」とみなされてきたところのもの (多分に俗流化された弁証法) へのたんなる反措定として提出され・・・、したがって (それを受け止める側の理論的低水準を前提としたばあい)、この両者の結びつきについてのイメージさえも与えられない・・・。
> - しかもそれは簡潔にまとめられたことによって、それ自体の完結性をあらかじめ備えてしまっている。だから当然武谷三段階論に共鳴する人々は、その内容を深める方向において自己の研究を推し進めるのではなく、・・・既にできあがったもの (不完全なる完全) としてとらえ、その適用ならぬ当てはめに転落していく・・・。
>
> 武谷論理学のこのような直観性と不完全な完結性を打ち破っていくためには、1) 一方では、絶えず個別科学の具体的研究においてその有効性と実践性を、その限界をつきだしてゆく努力が要求されるとともに、2) 他方では、マルクス弁証法研究の発展系列にそれを理論として正しく位置づけ、そして発展させ、充実化させていくための研究が必然的に要請されるのである。・・・後者の側面・・・について・・・マルクス弁証法の革新的なもの (資本論の弁証法) に踏まえつつ、三段階論を主体的に展開すること以外にはありえない。

これを成し遂げるためには、しかし、武谷の「科学主義」あるいは「技術主義」の傾向が大きな阻害要因として働かざるを得ない。

・・・「哲学」に関する武谷の理解の仕方の浅さ、そのプラグマティックな本性はもっぱら「哲学の有効性」なるものを基準としつつ、唯物論哲学からその主体的側面を追放する傾向を生み出す。言い換えれば、マルクス主義哲学がプロレタリアート自己解放の精神的武器であり、現実的ヒューマニズムに貫かれた革命的世界観であるという、この本質的で革新的な問題を疎外する傾向が発生するとともに、同時に、この疎外された領域の問題は、人間による自然の支配という技術的問題としてのみとらえ返されるにいたる。人間の人間的な問題、(人間主体の技術的側面ではなく、それに媒介された) 人間主体の内面性の問題への実存的関わりの問題は、その「科学主義」(「哲学の有効性」を基準とした) とその「技術主義」(人間による対象の技術的

[*1] 武谷「哲学はいかにして有効さを取戻しうるか」(1942)

支配を中軸におく) のゆえに彼岸の問題として追放され、そして追放されたこの問題は「思想の科学」という客体主義へ委ねられている・・・。

・・・こうして武谷三段階論の次のような諸弱点が、無自覚のままに残される結果となった・・・。

1. 現象論→実体論→本質論への人間認識の高まりの内的構造の究明の欠如、あるいは実体論的認識・・・の固定化傾向・・・。[*2]
2. 本質論そのものの内的な立体的構造の究明が無視されていること (このことは物理学における分析対象の対象的性格や、歴史性がその抽象的形態としての時間性において把握されること、などと密接に結びついている)
3. 人間認識の段階的深まりが直線化され (「ツッコミ認識論」)、その螺旋的本質が無視ないし軽視されていること、つまり下向・上向の論理からの切断において三段階論が提出されていること[*3](このことは、人間主体の問題がその技術的な側面に矮小化され、哲学的人間論が欠如し、真に場所的な立場の深みが掘り下げられていない、その科学主義的傾向の必然的な帰結である)。
4. 行為的現在における人間認識の深化の論理的過程とその歴史性の問題とが正しく統一的に展開される道が塞がれていること、つまり論理的把握と歴史的把握との完全に現実的統一の論理が三段階論からは完全に消失してしまっていること。
5. (上の諸欠陥が) 歴史的過去において形成された諸理論体系の学的発展 (研究結果の歴史的連鎖) にかの三段階論を外的に投影する客観主義的操作によって隠蔽されている。いい換えれば、人間認識の歴史的発展の構造を、現在的認識の論理的発展の構造からのアナロジーとして理解しようとする客観主義と歴史主義。

・・・このゆえに武谷自身が自己の三段階論をマルクス弁証法の発展系列に位置づけようともせず、またそれをなし得ないでいる・・・。そしてまた三段階論の絶賛者が実際にはそれを無視しかつ適用しえないという珍奇な現象 (三浦つとむのばあい) が発生するとともに、三段階論のたんななる当てはめの限界に関する自覚を、ただちにそれの非有効性と等値するような堕落 (東大自弁研の一部の会員) もまた不可避となるのである。

本稿の筆者は、以上の黒田による批判に全面的に賛意を示すものである。

B.2　唯物論的目的論＝認識論の確立

黒田 (1977) は「三段階論をマルクス弁証法の発展系列に位置づけ」る試みのひとつとして、黒田 (1968) の第 3 章第 2 節の III[*4]、第 3 章第 3 節[*5]を挙げている。以下ではその該当部分を取り出して整理してみる。

まず〈武谷技術論〉の主体的把握[*6]とはどういうことかが説明される。それは、唯物論的目的論を確立することであり、「意識的」ということそのものの論理を掘り下げることである。

〈武谷技術論〉の・・・哲学的ないし認識論的意義は、(観念論哲学における実践哲学、あるいは) 観念論的目的論にたいして、唯物論的目的論 (人間労働の合目的性の唯物論的論理構造) を解明する「実践論」・・・認識論および史的唯物論の基礎理論として提起されたところ、に存する。

〈生産的実践〉において、「意識的適用」とは、実践主体が認識し、判断し、決意し、もって実現するということなのであるから、〈客観的法則性〉の「意識的適用」ということの存在論的意味ないし可能根拠 (を明らかにすること) が・・・「認識論の基礎としての技術論」の存在論的基礎付けが・・・〈武谷技術論〉の認識論的＝存在論的把握が、同時になされなければならない・・・。

・・・「意識的」とは「主体の適応行動の社会的規定性」(田中吉六) であるが、この「意識的」ということそのものの論理、〈客観的法則性〉をいかに認識し、かつ、それがいかにして「実践への決意」へ転化し、物質的世界で現実化されるか、という「意識的一適用」の論理過程そのものを究明すること、これこそが〈武谷

[*2] 「実験について」(T-5) は現象論から実体論への移行の論理を具体的に追及したものとみなせる。またガリレイ論文 (「ガリレイの動力学について」,1946)、ケプラー論文 (「ヨハネス・ケプラー」,1947) についてもそのように言える。しかし、実体論から本質論への移行の論理は十分に展開されないままに残された。それは武谷の主たる関心が本質論としての量子力学の完成に集中し、結局それが果たされないままに終わったことに関係があるかもしれない。武谷はガリレイ論文、ケプラー論文に続いて「ニュートン論文」に着手すべきであった。なお、武谷 (2014) の第 5 章以降はこの素描とみなせる。これとともに、1951 年時点で燃素、熱素、エーテルを取り扱うと予告されていた下巻がついに刊行されなかったことは残念であり、この課題は依然として未解決のままわれわれの前に残されている。

[*3] 下向分析に偏り、上向の論理が空白のまま残されたこと、これが「実践の物質性」の問題に上滑りさせられたこと、などにかかわる。

[*4] 第 3 章「ヘーゲル目的論と技術論」第 2 節 III「唯物論的自覚の問題と技術論」

[*5] 同第 3 節「技術論と自覚の論理—ヘーゲル概念論の「武谷＝梯」的転倒—」

[*6] その中において〈武谷三段階論〉も統一的に把握される。技術論は「認識論の基礎」なのだから。

技術論〉の主体的把握だということである。

観念論的目的論 (またそこから生み出される実存哲学における「無」) の誤びゅうは、自由判断という意識の事実を「存在の根源」へと置き換えてしまうことから生み出される。そこから、唯物論的目的論を確立するヒントが、外なるものに「人間的自由の根源」を認める、という点にあることが指摘される。

> 人間的自由の根源 (自由判断) を「空虚な自己反省の方向に実在化」するとともに、この意識の事実を存在の根源に置き換えることによって、逆に自然史の発生の問題をもそこから基礎づけるのが観念論の常套手段である。したがって、かかる観念論的倒錯から脱却するためには、・・・「それ (人間的自由の根源) を外なるものに転ぜしめ」なければならない。

> ・・・内なるものは外なるものの反映であり、外なるものが内に映しだされるのである。しかるに内なるものが外にあるものから切り離されて、無底なる自己といった思惟の抽象・・・が固定化され、絶対化されるとき、非合理主義としての実存哲学における「無」が生みだされる・・・。

「人間的自由の根源」あるいは「決意のパトス性」は、外なるもの、すなわち (自己のなかに認められる)「物質の実体的力」に由来することが明らかにされる。また、〈物質〉とは、観念論哲学者の誤解するように静的な「たんなる客体的な自然物質」のことではなく、「自己運動性」という「自由」をもつ〈弁証法的物質〉[*7](弁証法性をもった物質) を指す。人間それ自身のもつ「物質性」とは〈弁証法的物質〉のもつ諸特性を人間もまた引き継いでいるということを意味する。また人間の思惟が反省 (下向) と構成 (上向) の統一をなすということそれ自体も「物質の自己運動性」のなせるわざであることが指摘される。

> ・・・自己みずからを不断に対象化しつつ、自己自身を越えてゆくというこの自己の無底性は、根源的な物質そのものの自由なる実体性、宇宙的質料の無限性にもとづくのであり、「物質の実体的力」が「決意のパトス性」として意識されるのである。(人間主体の超越の無限性、すなわち) 意思の自由は、自己のうちに自己を超えたものを潜在させている「自己運動する根源的物質」そのものの無限性 (ないし普遍性) を自覚することにおいてはじめて現実的な意義を持つのであり、それこそが・・・唯物論的自覚である。

> 物質はたんに客体的な自然的物質ではない。主体性原理をもち (したがって能動的実体と受動的実体との対立・矛盾・闘争において交互媒介的に)・・・自己形成をとげる〈弁証法的物質〉である。かかる物質の・・・自己運動を内在化したところの・・・意思的自覚、これこそが唯物論的自覚である。・・・意思 (・自覚) の立場は、人間が (客観的な物質過程に自己否定的に即することによって) 自己に矛盾した超越的な感性的対象を (その超越性において) 内在化したところの物質と思惟との矛盾的自己同一、すなわち「生きた直観」の内容としての〈歴史的自然〉を内に直観し、かつ、それを合理化し、対象化してゆく立場に他ならない。外的世界の論理が、内部から衝動的に精神を動かすのであって、ここにおいて・・・人間をして実践へ駆り立てる原動力となるのである。

> ・・・反省することが構成することであり、構成することが反省することである、というこの思惟の弁証法は・・・根源的な物質の自由なる実体性すなわち自己運動性にもとづく、その所産である。

B.2.1　現象的目的：自由の可能性

現象論的段階 (感性的＝物質的直観) にあっては、対象的自然とそれに向かい合う人間的自然とが、主客の対立をはらみつつ、そのまま合わせて意識内に内在化 (反映) される (〈歴史的自然〉)。原初的な「目的」は衝動性となってあらわれる。この衝動は反射としてただちに表出されはするが、これはただちに対象的世界で否定され、かえって不自由なる自己をあらわにする。ここにおける自由は可能的＝偶然的なものに過ぎない。

> ・・・感性的＝物質的直観においては、認識対象である直接的＝媒介的な対象的自然と、同時にその根底に実存する根源的な物質が、その超越性において内在化されるのであり、この直観内容が〈歴史的自然〉にほかならない。これは、いまだ反省の対象となっていない・・・ものとして、・・・直接的に主観的衝動としてあらわれる。

> ・・・〈歴史的自然〉を自己の根底にみる・・・すなわち「意志の立場」にある人間は、だから、物質の自己運動の一契機として自由な実体となりうる可能性に

[*7] Lenin による規定「人間の意識の外にあり、自己運動する物質」を参照。武谷の言う「自然そのものの構成」は、同じものを存在論の側ではなく認識論の側から静的に表現したものと言える。

ある自己である。しかし、たんなる可能性にあるものとして、同時に不自由なる自己であり、偶然的な自己である。

衝動としての目的は、思惟作用により個別的判断へと純化される。自己意識が発生する。これは人類史の起点においてなされた歴史的事実を個々の人間意識の事実として再生することにほかならない。主観と客観とが分裂し、また現実の自己と観念上の自己とが分裂し、〈歴史的自然〉をその物質的基礎〈客観的法則性〉からとらえ返そうとする認識の出発点が形成される。

・・・物質的世界における主客の対立は、外 (＝客体) と内 (＝主体) との対立をとおして、内そのものにおける客観性と主観性との対立として映しだされることによって、実践主体である自己は、現実の自己と観念的自己とへ・・・分裂する・・・。

・・・いまだ反省の対象となっていない物質的直観内容の即自態は、思惟作用によって内から構成され、主観的な衝動の恣意的で偶然的な要素がとり払われて、客観的なものへ純化され、高揚されて、それ自身個別的なものとなる・・・。〈歴史的自然〉は、それが「生きた直観」の内容であるかぎり、それの物質的基礎＝認識対象である特定の歴史的社会的な現実、感性的自然の科学認識に・・・媒介されなければならない・・・。

物質的判断＝〈原始判断〉は、自然史的過程の一発展段階における、物質と人間の意識との〈原始分割〉(自然の弁証法性の必然的結果としての「社会の起源」) に照応した、「認識の起源」である。「社会の起源」における自然史的事実が「認識の起源」における意識的事実として再生産されるのである。・・・

B.2.2　実体的目的：自由の現実性

実体論的段階 (反省) においては、原初的目的であった衝動は「理性的目的」にまで高められる。結果としての諸現象の世界 (対象的自然・・・これには対象とされた自己をも含む) から偶然性 (諸実体) の認識を媒介として、「原因」の探求が思惟の内部で目指される (認識過程)。因果関係 (原因→結果の関係) が論理的に明らかにされる。これは歴史的 (時間的) には現在から過去にさかのぼることでもある[*8]。

・・・それ (目的) は、直接的には意欲や衝動や感性的欲望としてあらわれる。・・・しかし、たとえ主観的なものに過ぎないとしても、主客の適応矛盾を発条として形成されたかぎり・・・(客体を自己に適合させようとする) 意志の立場にあるものとして、目的の「直接性」であるということができる。だから問題は、このような・・・衝動や感性的欲望をば (そこにおける主観的な、恣意的な要素を純化して)・・・(客体を手段的に駆使しつつ、物質的世界で有効的に実現されうる)「理性的目的」へ高めることにある・・・。・・・「客観的世界から規定的内容と充実とを取り出す」(ヘーゲル) ことによって、(外的必然性の認識を媒介として開示される・・・)「実在的可能性としての目的」が獲得されなければならない。・・・唯物論における目的とは、認識論的には対象的自然の科学的認識によって構成される「実在的可能性」(であり、存在論的には根源的な物質に内在する宇宙的内容の自覚) である。

・・・ここでの問題は目的意識の形成にあるのであるから、逆に感性的対象の認識活動が「同時に」いかなる概念活動となってあらわれるかをみるのでなければならない。・・・主体に対立する直接的な感性的対象 (＝客体) は「個別化の形態」にある種々の現象形態 (・・・諸現象の世界＝結果) にすぎないがゆえに・・・そこに実在するところの普遍的本質を・・・科学的に分析 (結果→原因) しなければならない。それは・・・偶然性 (・・・諸実体間の交互作用) を媒介として現象した (必然的結果としての) 直接的現実性に対する「可能性としての原因」を探求する (現在→ 過去) ための「認識過程」であり・・・分析的思惟活動である。

つかみとられた本質 (「法則」) から、現在の現象世界への復帰が目指される (思惟の中での「存在過程」の再生)。ここでもやはり偶然性 (実体) の存在が忘れられてはならない[*9]。直面する現実が偶然性を媒介としていることがこの過程で了解されるからこそ、現実の対象的世界とは異なるもう一つの現実 (「理性的目的」) が見いだされることになる。

・・・ (このような認識に媒介された思惟活動の帰結として開示された) 本質的な関係＝法則を・・・逆に起点として、それが偶然的諸条件・・・を媒介として特殊的に実存する個別的諸現象形態をとらえる (原因→結果) という綜合的思惟活動によって、〈客観的法則性〉が概念的に把握される・・・。これが認識過程によって開示された「存在過程」であり・・・ (「具体的

[*8] もちろん思惟の中においてである。

[*9] 黒田の記述は目的の本質論的取り扱いに性急に進もうとするあまり、実体論と本質論の記述が入り混じっているので注意が必要である。本質論に相当すると思われる箇所 (主に存在論的な解釈にかかわる部分) は引用中でも括弧に入れて区別した。

なもの自体」の) 思惟による再生産過程 (過去→現在) である。

原因と結果の関係が目的と手段の関係に逆転される。思惟の中での「存在過程」の再生は、時間的には過去→現在という流れを示すが、得られた「理性的目的」は現在の現象世界を「原因」としたところの「帰結」として、未来の領域にある。それはいまだ実現されたものではない (当然、その目的の実現にも失敗するかもしれない) という意味で、主客の適応矛盾の観念的な解決 (止揚) でしかない。しかし「理性的目的」とそこにいたる過程についての構想を手にした人間は、もはやこれまでの対象的世界に蹂躙されるばかりの存在ではなく、意志的な自由を獲得した立場にある。ここにおける自由は、特殊的にして現実的な自由である。

> ・・・このような存在過程の概念的把握は、直接的現実性——これは存在論的には、(可能性としての) 原因が、種々の特殊的現実を媒介として、必然的に現象した結果である——を基礎として・・・実践的立場に合致した事態への「過程」および「帰結」を構想し、(この因果の連鎖において構成されたところの) 変革的に止揚された客体、とそれへの手段、を含むこの構想を「意識的目的」とするための思惟過程である。・・・それは、(存在論的には物質の必然的自己展開として意義をもつところの〈歴史的自然〉そのものの自己展開であり、) 認識論的には主客の直接的統一として、物質的世界における主客の適応矛盾の実践的な解決、だがいまだなお現実化されていない観念的止揚である。(かかる「理性的目的」を獲得した認識主体は、同時に、現実の自己を否定的に超越した意志自覚にある「自由なる自己」である。・・・かかる人間主体は、認識論的には外的＝対象的自然を有効的に、技術的に変革しうる「理性的目的」を構成し、自己に否定的に対立した対象的現実を自己の目的にしたがって「空無」なものとして措定し、もって物質的世界を発展させせようとする意志自覚にある・・・。)

合目的性と因果性の関係が説明される。目的→手段の過程は「将来からの限定」として意識される。これは原因→結果の時間経過とは逆であり、神秘的なものに思えるが、(思惟の内部ではなく) 現実の存在過程からすれば、いずれも原因→結果の方向性を持つものであり、何ら神秘的な要素を含まない。

> ・・・結果として生ずべき事態を構想し、それをあらかじめ目的として先取し、それを結果せしむべき原因を手段とするのが (因果的必然性と逆の過程をたどる) 合目的性である。この目的→手段の過程は、客観的時間の系列に対して逆の方向をたどるがゆえに、「将来からの限定」として意識される。(・・・認識論的には結果→原因であるが、同時に存在論的には原因→結果たる目的―手段の体系であればこそ、客観的時間の系列を逆の方向をたどる合目的性が現実に成立しうる・・・。)
>
> ・・・「原因→結果の因果的必然性にたいして、合目的性においては目的によって手段が制約される、という意味で時間的に逆の方向をたどる (結果→原因)」と言われるのは、結果としてあらわれるものがあらかじめ目的として設定され、(それを結果させる) 原因が目的実現の手段となっている、とわれわれが反省するがゆえでなければならない。・・・それゆえ、目的関係は因果関係の反面にほかならず、なんら神秘的要素をもつものではない。

B.2.3 本質的目的：自由の必然性

本質論的段階においては、思惟 (観念的活動・頭脳労働) の世界から現実的実践 (対象的＝感性的活動・肉体労働) の世界に舞台が移される。この肉体労働は頭脳労働の延長として行われるものであり、この事態は「行動の理性的推論」としての意義をもつ。

> ・・・概念活動を媒介として〈歴史的自然〉と (観念の上で) 合一した、物質的自覚にある超越的自己[*10]、物質を自己の内容＝本質として自覚した歴史的自己は、物質の自己運動過程にあるものとして、〈根源的な物質〉の主体的自己形成の創造的尖端たる意義を持つ・・・。このような物質的自覚にある人間は・・・しかし現実の自己において・・・真に自由なる自然の主体ではない。現実に自由を獲得したのではない。なぜなら・・・ (主体をして自覚という自己矛盾的な観念活動をなさしめた物質的基礎である) 彼に否定的に対立した客体、まさにこの客体が、いぜん物質的に変革されることなく外に残されているからである。主体はその物質的自覚を客観的世界へ実現し、客体を否定し、変革的に止揚し、もって自己の意志自由を現実的なものとして確認しないかぎり、たんに可能的に自然の主体でしかないからである。・・・
>
> ・・・かくして、今や意志自覚・意識的目的を物質

[*10] 超越的自己とは自己に対立する対象とともに自己をも否定する自己のことである。つまり現在あるままの自己を否定し、将来のあるべき姿を志向する自己である。

> 的世界へ実現し、もって自己否定的に対立した客体をば「自己にとっての存在」たらしめるということの論理が・・・物質的自覚、目的意識の形成の論理過程の発展として展開されなければならない。
>
> ・・・ところで意志・目的の実現はまさにそれが物質的世界への実現なのであるから、もはや観念的活動の次元を超えているのであって、それは対象的＝感性的活動の、物質的労働の問題でなければならない。・・・(ヘーゲルにおけるように目的そのものが客体へのりうつるのではない。)・・・我々の出発点は、意識から独立した客観的実在の物質的諸条件そのものが提起する課題の実践的克服にあるのだから、つねに課題の解決は、課題を提起した物質的諸条件におけるその変革的展開 (という客観的過程) によってのみ成立する。「さらに進んで実践が唯一の解決手段であるがごとき課題において終局する」(マルクス) のでなければならない。・・・それは物質的判断の自覚過程の発展として (行動の)「理性的推論」として意義を持つ。

対象的＝感性的活動の論理が、「主体の客体化」をきっかけとした「客体の主体化」であることが明らかにされる。「主体の客体化」とは、人間がみずからの肉体を「客体的な一つの力」とせしめることであり、これは同時に一つの自然物を自己の延長 (道具) とみなすこと、すなわち「客体の主体化」でもある。ここにおいて、主体と客体との関係は流動化し、一つの対象的生産物となって「帰結」する。この帰結が観念的目的の対象化として意義をもつことが確認されて、はじめて対象的＝感性的活動が合目的的活動＝技術的実践たるの実を示すことになる。

> 技術的実践の基底にあるところの人間活動一般、すなわち対象的＝感性的活動は、人間主体が根源的に・・・自然によって・・・創造された客体であるにもかかわらず、同時にそれが能動性を持つこと (・・・自然に働きかけてそれを変化させ、・・・自己に対立・矛盾した存在を自己に適合した存在へ転化させうるという・・・存在性) の動的表現にほかならない。それは・・・自己をも「一つの自然力として」(マルクス)・・・手段的に働きかけること、すなわち「主体の客体化」である。・・・主体が「客体的な一つの力」として自己を手段的に駆使し、対象＝客体へ働きかける・・・ということは、しかしながら、同時に「客体の主体化」・・・である。・・・「主体の客体化」にもとづく「客体の主体化」という主客の媒介運動、これがすなわち対象的活動の唯物論的規定にほかならない。・・・その直接的結果は対象的生産物の創造・・・である。かくして直接的対象は、人間実践の歴史的成果として「人間化された自然」として (媒介されて) 実在するとともに、人間もまた (労働の歴史的所産として) 技術性を獲得し・・・自然を自由に変革し、変化させてゆく自然の主体となったのである。
>
> ・・・主体の客体化に基づく客体の主体化 (が) 目的の支配のもとにある客観的過程 (であること)・・・主体の客体化がつねに目的の対象化＝現実化をともない、目的に規定された主体が自己を「客体的な一つの力」として手段的に駆使することによって、同時に客体を自己の目的の手段 (道具) として機能させること、これが合目的的活動＝技術的実践にほかならない。・・・意識的目的・自覚的意志の物質的世界への実現とは、かかる合目的的活動である。

生産物において目的の実現を認めること (つまり生産物の消費[*11]) が、人間主体をしてこの生産物を、〈根源的物質〉の自己展開の成果であると解釈せしめる。ここにおいて、設定された「理性的目的」ははじめてその普遍性・必然性を明らかにするものと言える。同時に、人間主体においても、現実の自己と観念的自己とへの分裂・対立が解消 (止揚) され、一つの自由なる主体のもとに統合される。これらすべてのことが、Marx の言う「自然主義＝人間主義」の実現である。

[*11] 目的の実現を確認することは、その目的が対象化された生産物を消費することでなければならない。だから疎外された労働においては、労働者はその生産物を奪われるがゆえに、自己の目的を見失うことになる。

表 B.1　唯物論的目的論の構造

段階	説明	目的	主客の適応矛盾
現象的目的	感性的＝物質的直観において、外的な対象的世界が主客の対立をはらみつつ思惟の内部に反映される (〈歴史的自然〉)。	目的は衝動性、感性的欲望となってあらわれる。	現実的自己と観念的自己、主観と客観の分裂。不自由なる自己。
実体的目的	下向分析を通じて〈歴史的自然〉が反省され、因果関係 (〈客観的法則性〉) がつかまれる。上向的構成を通じて、因果関係は目的―手段関係に反転される。	衝動は理性的目的に昇華される (目的の生産)。	主客の適応矛盾は観念的に止揚される。
本質的目的	対象的＝感性的活動において目的が対象的世界の中に実現される。「主体の客体化」をきっかけとした「客体の主体化」、これによる主客の対立関係の流動化。感性的活動は合目的的活動たるの実を示す。	生産物における目的の実現とその確認 (消費)。	主客の適応矛盾は実践的に解決される。自由の必然性。あるいは自由と必然性の合一。

> ・・・技術的実践による感性的自然の変革の直接的結果は、言うまでもなく対象的生産物の実践的創出であり、「対象的自然の加工・創造」である。それは存在論的には、根源的な物質に自己矛盾的に内在化している宇宙的内容の、技術的実践を媒介とした、部分的現実化・・・としての意義を持つ。・・・目的・意志の実現の結果 (の直接性) を・・・このように解釈することは、しかしすでに対象的活動の結果の認識に媒介されている・・・。労働の結果の対象的認識・・・を通じて、目的の実現を、自己の設定した「理性的目的」の有効性を、検証し、意志の自由を確認すること、その客体的表現にほかならない。(なぜなら客体はそれ自身意識するのではないからである*12。)・・・実践に先立って意識の中に構想された「理性的目的」が対象化＝実現され、客体の中に物質化されているか否かを、・・・変革された対象的現実＝客体において認識し、それを媒介として主体的に自覚するのである。もしもそこに目的の意識的適用の成功を、すなわち、実現された目的を、(あるいは自己の表現を) 見いだしえるならば・・・(実践の以前には自己に否定的に対立していた) 客体が人間的現実性へ止揚された (と言える) のである。
>
> ・・・対象的活動を通じて (すなわち過程とその結果において)、主体は自己を自由なる自然の主体として自覚し、ここにおいて人間は現実的人間性を獲得し・・・たのである。・・・客体を自己の表現としてみる人間主体・・・は、それゆえに、現実的自己と超越的自我 (自己否定の立場にある) 自己とへの自己分裂を止揚した、自由なる主体である。たんなる観念的自由ではなく、その物質的基礎をもった現実的自由を獲得したのである。

最後に理論的思惟活動がつねに目的意識的活動であることが振り返られる。この根本が忘れ去られるとき、理論それ自体が疎外され、みずからの欠陥を訂正する生命力を失い、またそのことの隠ぺいを図り、もって虚偽のイデオロギーへと堕落する。

> ・・・理論的思惟は本質上つねに必ず目的意識的・・・である。いい換えれば、われわれはつねに外的現実に触発された実践的意志の立場を出発点とするのであって、理論的認識や思惟はその媒介的契機に過ぎない・・・。(この根本的な立脚点を喪失するとき、唯物論は俗流化され、階級性の強調をもって理論的欠陥を隠蔽するという機械論に必然的にはまり込まざるをえなかった・・・)

以上のことをすべてまとめると表 B.1 のようになるだろう。

---

*12　「・・・意義をもつ」という黒田の用語法は、直接には人間が対象的世界を解釈することを意味する。しかしこの人間がたんなる人間ではなく、〈弁証法的物質〉から由来される「人間的自然」であることを念頭におくかぎりで、この解釈は主観的なものにとどまらず、「唯物論的自覚」と言いうる。

B.3　三木清「手記」の読み方

三木清の「手記」のなかに見える自然弁証法否定論にたいして武谷は「自然科学者の立場」から十分な批判を与えている。ここでは、三木がなにゆえにこのような手記を書くに至ったのかを別の方面から考えてみたい。

「手記」の書きぶりからは三木は 1945 年豊多摩刑務所での自己の死をつゆほどにも予感していなかったように見える。むしろ自分の著作を引用しつつ得々と自己の哲学的立場を思想検事にたいして説明し、あまつさえ将来の出版計画にさえ言及している。「自分はマルクス主義者ではない」と信念に反して偽証する必要もなく、ある程度は本心から「マルクス主義者たち」と対峙しているようにも見える。

そこで以下では、三木が本心からこの「手記」を書いたことを前提に、疑問文を連ねるかたちで「手記」を読み解く。

1). 三木はなぜマルクス主義批判に際して、なにをおいても宗教と自然弁証法の問題を取り上げなければならなかったのか。「マルクス主義の根本思想に対して私自身の立場から、世界的文献の間を伍して、極めて特色ある独自なる解釈を与えたと信ずる。この点自ら密かに誇りを感じているところである。」と三木が言っているのは何を意味するのか。

2). 三木が自己の立場を「人間学的存在論」と称しているのは何を意味するのか。またそれが弁証法の適用領域を限定すべし (自然弁証法の否定) という主張につながるのはなぜなのか。

3). 三木の処女作である「パスカルにおける人間の研究」について「パスカルに対する私の解釈を通じて私の宗教的感情が流れている・・・」と説明する三木の心情はいかなるものか。

4). 〈宗教の死滅〉について三木があれほどまでに反発するのはなぜなのか。三木は「マルクス主義者」の宗教否定のうちに (不安な魂を嘲笑する) 不真面目な Montaigne[*13]の姿を見て、これに Pascal とともに怒りをあらわにしているのではないか。三木の怒りは科学主義的 (あるいは実証主義的) 進歩観を振りかざす者たちに向けられているように見える。

5). 三木が「宗教は・・・人間の本性のうちにその自然的な基礎を持つ」という時、(Marx が最初にその身を置いた) ヘーゲル左派がキリスト教批判から出発したということになぜ言及しなかったのだろうか。「マルクス主義者たち」は「宗教家を研究していない」という時、Marx その人が宗教的自己疎外のことを全く考えなかったとなぜ信じるのか。

6). 三木が「近代自然科学がその方法において「機械論的」であるから「予見」や「予測」が可能だ」といっているときに、かつて自分が書いたこと[*14]を忘れているのはなぜか。

7).「弁証法的自然観はヘーゲルがその自然哲学の中で叙述したところである。そして自然哲学はヘーゲルの体系の中における弱点と従来みなされてきた。この思弁的な自然哲学こそヘーゲルの哲学をその当時没落せしめるに至った最も重大な原因の

[*13] Michel Eyquem de Montaigne(1533-1592)。「パスカルにおける人間の研究」の第二章「賭」で、三木は次のように書いている。「彼自身の証言によれば、「エセエ」の著者は、「恐れなく、悔いなく、救済に対して無頓着なる心を吹き込む」に反して、「パンセ」は「この無限にして部分なき存在に祈るために、先にも後にも跪くひとりの人間によって作られている」。「モンテエニュはふざけている」。」

[*14] 「偶然と必然—クルノーを憶う—」(1932) より、「・・・従来普通の見解に従えば、因果律は事物の機械的な必然性の原理であり、然るに機械論と弁証法とは同一であり得ない・・・この点に関してヘーゲルの思想を顧みるならば、彼は因果性の範疇は交互作用の範疇に移行せねばならぬと考えた。・・・けれども問題はそれで尽きておらず、しかもなおそのもう一つ手前に横たわっているように見える。もしも因果の連結が機械的な必然性を表すものとすれば、因果性の観念は弁証法と相いれないものでなければならぬ。因果性の観念が何らかの仕方で弁証法と結びつき得るためには、因果性そのものが従来考えられてきたのとは反対に機械的必然的な規律性とは異なる意味のものでなければならぬであろう。この点について最近の物理学の発展 (量子力学) は我々にとって興味ある結果をもたらしている・・・それによれば、自然法則はなんら機械的な、いわゆる宿命的な必然性を表すのではなく、その根本において既に蓋然的である。・・・私はただこの問題に関心する読者に向ってひとりの優れたプロバビリストに対する注意を新たに喚起するにとどめよう。・・・クルノー (Antoine Augustin Cournot:1801-1877) によれば、偶然は我々の知識の状態に依存するもの、すなわち主観的なものではなく、客観的実在性を有する。・・・」

[*15] Abram Moiseevich Deborin(1881-1963)。Georgi Plekhanov の弟子として出発し、Stalin が台頭するまでは「マルクス主義の旗の下で」の編集長を務めた。

一つであった」とまで三木は書いている。そのような三木がなぜ Deborin[*15]にたいして中途半端な支持しか与えなかったのだろうか。Deborin の試み[*16]を自己の哲学とは関係ないものとして、冷笑をもって傍観者的に見ているのはなぜなのか。

8).「現実存在」〈実存〉の弁証法について、・・・「単なるバラの花」ではなく「愛らしいバラの花」について語るのは、バラの育種についてその技術的実践のことを語っているのではないか。それは自然の領域であると「同時に」社会の領域の問題でもあるのではないか。その事実を語ることがなぜ自然の弁証法性の否定にまで行き着かなければならないのか[*17]。

B.4 Bukharin の均衡論

Bukharin が均衡論を提唱し、この理論がある時点までボリシェヴィキ党内部で共有 (許容) されていた事実の真の意味を今日の時点で了解することは難しい[*18]。ここでは岩崎 (1958) をヒントに Bukharin の均衡論についていくつかの論点を列挙したい。

1). Bukharin がオーストリア・マルクス主義に対抗して均衡論を提示した動機は次のようなものであったとされる。「弁証法の物質的根底を明らかにすること、すなわち変動する物質の形態にヘーゲルの弁証法的公式が対応するところのものを発見すること」。岩崎はこれを「皮相なもの」と評している。

> 彼のヘーゲル理解はこの著作の全体を通じてこの程度の皮相さを出ず、ヘーゲル弁証法の合理的核心を見失った上で、その公式に対応するところの現実として均衡のかく乱と回復、「可動的均衡」を持ち出す・・・

Bukharin がこのような Hegel 理解の水準にとどまったのはなぜか (逆に言えば、Bukharin は Hegel その人の自然哲学にはさほど興味を持っていなかった可能性がある。)

2). 岩崎は Bukharin が力学的概念をもって哲学を論じていることを非難している。哲学と自然科学との関係についての Bukharin の理解[*19]と岩崎のそれとは一致していない。岩崎も Bukharin もそれぞれの立場から「均衡」を解釈することに熱中しており、その解釈の優劣を競っているだけのようにも見える。

> 力学的運動は、あらゆる運動の基礎的形態であるがゆえに均衡の概念をここから一般化することができ、またこれによってアドラーの修正と戦い、マルクス主義の弁証法を発展させることができるというのが、彼の機械論的な考えである。しかし、こうして形成された均衡概念には特殊科学の内容が押し込まれており、もはや哲学的カテゴリーとは言えない・・・

3). 岩崎にとって Bukharin の態度が許しがたいのは、次のようにカトリックの理論家[*20]から揶揄されても仕方のない体たらくをボリシェビキ党が示した実例になってしまっていることにある。

> ブハーリンの均衡論は、いかにそれがつつましやかな試みであろうとも、とにかくマルクス主義をヘーゲル弁証法の障害から救おうとする試みであったように思われる。ヘーゲルの「定立」、「反定立」、「総合」に代わるに、「安定均衡」、「肯定的印を持つ不安定均衡」、「否定的印を持つ不安定均衡」の状態をもってした。ブハーリンの代替物に対しては、微笑を持ってこれを迎えることができるであろう。
> —ピオヴェザーナ「現代ソヴェト史的唯物論」

[*16] 「デボーリンのごときはかくのごときヘーゲルの自然哲学を新しい形で復活せしめようとしているかのごとくに見える。」

[*17] 「愛らしいバラの花」を感じるわれわれの感性そのものが自然 (対象的自然ではなく人間的自然) そのものに由来するのだが、三木の「人間学的存在論」はそのことを認めることができない。

[*18] この理論がたとえ「真正」なものではないとしても、彼らがそこにある種の希望的なビジョン (革命の「上げ潮」期の) を見出していたのではないかと想像できる。Bukharin は「共産主義の ABC」、「過渡期経済学」、「史的唯物論」といったボリシェヴィキ党公認の教科書の執筆者である。たとえ筆が滑ったのだとしても、社会全体が相転移のように変動し、それを誰にも止めることができないという壮大なビジョンは党員への宣伝として大目に見られたのであろう。Bukharin その人は場の雰囲気に流される人であり、ブレストリトフスク条約への反対論から一国社会主義 (ネップ) へと、極端から極端へとその態度は揺れ動いた。

[*19] おそらくこの点は Engels の哲学観 (科学の進展につれて哲学自体が消滅する) が影響していると思われる。また過渡期経済学 (広義の経済学) に関連して、Bukharin は資本主義経済がなくなるのだから経済学それ自体も消滅すると理解していたことも想起すべきである。

[*20] Gino Kirill Piovesana S.J. (1917-1996)

4). 岩崎は Bukharin がそれほどまでに「均衡」を重視する (「均衡の絶対化」) 理由が、現象の絶対化 (マッハ主義) から来ると見ている。

> この理論では、均衡の概念が絶対化され最も基本的なカテゴリーとされている点を、指摘しておきたい。言うまでもなく、マッハ主義は、物質の運動と矛盾という唯物弁証法の見地に真っ向から反対するのであるが、均衡理論も、現象を見て実体を見ず、現象間の量的依存関係を見て本質的な矛盾の統一を見ず、均衡を経済学の最も基本的なカテゴリーとみなすのである。

5). これに対して岩崎は「均衡」についてのマルクス主義的な見方を Deborin[*21]を引用することをもって対置している。

> 唯物弁証法の光の下では、特定の傾向の運動法則の制限としての、「否定」としての均衡そのものが、資本主義の運動法則を照らしだすことができる。「均衡」は、この場合には、なぜそれが不断に存在しないか、なぜこの均衡が最も純粋な偶然であるか、ということを証明するためにのみ用いられる。

B.5　目的と手段の弁証法

Trotsky(1973) は社会的実践 (階級闘争) の領域で道徳 (目的の生産) の問題について答えようとしている[*22]。その内容を簡単にまとめると次のようになる (また Dewey による批評について脚注の中で触れる)。

- 諸階級はそれぞれ独自の歴史的任務 (目的) をもって闘争の場にあらわれる。超階級的 (全人類的) 道徳を主張することは、階級闘争の現実に目をふさぐことでもある。モラリストは大衆運動から切り離されて (疎外されて) いるために、保守的かつ独善的な (つまり臆病な) 精神に凝り固まる。
- 異なった目的をもった諸階級が似たような手段を採用することはあり得る (例；軍隊組織の規則)。その意味では、階級を超えた手段の類似性とその洗練は人間性の進化と言えないこともない。
- (ボリシェビキへの非難に使われる)「手段は目的を正当化する」という標語はもともとプロテスタントから 16 世紀前半のイエズス会の行動指針に向って投げかけられたものだった。しかし、イエズス会が主張したのは、手段それ自体は目的とは独立している (例：銃はそれ自体善でも悪でもない。狂犬を撃つのは善であり、銃による殺人は悪である) ということに過ぎない。イエズス会は集権化された独自の軍事組織をもち、その英雄的な精神と行動で敵ばかりではなく味方からも恐れられた。ただしボリシェビズムとの類似性は表層のものであり、彼らは「反動」に奉仕していた[*23]。
- 「最大多数の最大幸福」を標語とする功利主義者は「公共の福祉」と「最高の善」という二つの目的が予定調和的に一致すると信じており、そこでは手段は (あたかも商品がその値札を運ぶように) それ自身のモラルを運ぶ。道徳の (Spencer 的) 進化は「直接的な快楽」と「将来の持続的な、より高度の快楽」との比較の問題にすぎない。彼らは目的の創造という問題 (何をなすべきか、なさざるべきか) には答えられない。そのため、歴史の変革期にはこれに目を閉ざし宙づりになる。
- 現代の資本制社会においては、支配階級 (資本家) はそれ自身の目的を共同利害として社会に押し付け (国民国家の創造)、それに反する手段をすべて罪とみなす[*24]。社会は功利主義者の希望に反し

[*21] この引用は価値法則についての、市場価格変動の役割についての Marx の説明に対応するものである。また岩崎は「均衡」をめぐる混乱が「再生産表式」論争に関わっていると論じている。すなわち「「資本論」第 2 巻の再生産表式論をめぐって試みられたマルクス主義の諸々の修正 (ツガンやブハーリンらの修正) は、資本主義の矛盾を、表式の不均衡と同一視し、均衡を矛盾の欠如とみなす点で、共通点を持っていた」。

[*22] 原稿そのものは 1938 年に書かれた。その献辞には Leon Sedov(1906-1938) の名が見える。彼は Trotsky の息子でありトロツキスト運動の協力者だったが、Stalin の手下により暗殺された。

[*23] スターリニズムの出現は、従来の単純な「進歩」と「反動」という図式化を許さなくなった。Trotsky がこのことに気づくのはこの文章が書かれた一年後の 1939 年のことである。

なお Dewey は多くの「マルクス主義者」がスターリニストの所業 (手段) を「社会主義体制」という「目的」を維持するために正当化してきた、と指摘し、Trotsky がその地平をはるかに超え出た水準で議論を展開していることを正しく見抜いている。

[*24] この記述は明らかに Marx(と Engels) の「ドイツ・イデオロギー」を下敷きにしている。

て「少数の、消え去りつつある者のための幸福」を目指す。このような体制は本来は強制力なくしては一瞬たりとも維持できず、そのために道徳というセメントが必要になる。その生産を引き受けるのが小ブルジョアの理論家である。

- 「万人に対する義務」は内容としては空疎で形式的なもの (Kunt の定言命法;kategorischer Imperativ) である。このことは、人が社会全体にではなく階級により直接的な連帯を感じるという事実から容易に予想できる。「万人の義務」(宗教、哲学、常識) は敵対する階級、ブルジョワジーの利益を代表しており階級闘争の事実を覆い隠す欺瞞のための必要不可欠の要素である。この欺瞞を暴露することがプロレタリア革命の最初の仕事となる。
- 第二次世界大戦以前に見られた帝国主義経済の持続的成長は階級協調と民主主義の雰囲気 (自由、公正、人道、進歩) を生んだ。恐慌はこの流れを断ち切り、階級対立は先鋭化した。民主主義の安全弁は次々に破壊された。凶悪犯罪、中傷、賄賂、強欲、殺人が社会を支配した。これらは帝国主義の衰退の兆候である。これらの兆候の全てが統合されたものが「ファシズム」である。
- 手段は目的によってのみ正当化できる。しかしその次には目的そのものが正当化されなければならない。プロレタリアートの歴史的使命の表現であるマルクス主義的世界観からすれば、次のように言える。目的は、それが人間の自然にたいする「支配」の増大*25につながり、人間の人間にたいする「支配」の廃絶につながるかぎりにおいて正当化できる*26。この目的は革命を通してのみ達成される。プロレタリアート解放という目的は革命的な性格を帯び、宗教的ドグマばかりでなく、支配階級の憲法であるあらゆる種類の観念的フェティシズムと対立する*27。
- 社会の発展法則、すなわち階級闘争*28というすべての法に優先される法から行為の規則が引き出される。モラリストは反論するであろう。「それでは今までと同じではないか。資本家との闘争においても、嘘、でっち上げ、裏切り、殺人などあらゆる手段が許容されてしまうのではないか」。われわれはそれにこう答える。許容され、また義務付けられるものは次のような手段に限られる。すなわち、革命的プロレタリアを団結させ、抑圧への敵意で心を満たし、形式的な道徳と民主的なみせかけを侮蔑し、彼自身の歴史的使命感を吹き込み、勇気そして闘争における自己犠牲の精神を高揚させるような手段、これらのみである。
- したがって、正確に言えばすべての手段が正当化されるわけではない。われわれが「目的は手段を正当化する」というとき、革命の大いなる目的は次のような手段をにべもなくはねつけるだろう。すなわち、労働者階級の内紛を招くこと、参加者抜きで幸福をもたらすこと、大衆の自己または組

*25 これは誤解されているような自然改造のことではなく、人間がどの程度自然諸力の法則性 (必然性) を認識し、それを自己の目的と調和させるか、ということに関わり、逆に言えば総労働のうち必要労働がどの程度に抑えられ、自由の領域がどの程度広がっているかに関わる。

*26 Dewey は Trotsky のこの論理を次のように批判している。すなわち、ここでは「目的」という語が二重の意味で使われている。第一は他のもので正当化される必要のない究極の「大目標」であり、第二はそのための (手段と位置付けられる)「諸目標」である。前者はある種の絶対的道徳、永遠の真理を意味し、Dewey はそれを独善的なものとして拒絶している。そして後者の小目標こそが、手段と目的の相互関係による点検を受けたもの (”permissible”) として人類の解放に役立つのではないか、と問うている。なお、脚注*28 も参照のこと。

*27 おそらく三木清はこの理屈に反発するだろう。しかし、Trotsky の主張は、革命家にとって宗教的ドグマがその革命的行動を押しとどめる理由にはなり得ない、ということであって、各人がどのような「宗教的感情」をもつべきかを論じているのではない。宗教は「私事」であるのだから、それを他者に押し売りさえしなければ、容認されるべきだろう。

*28 Trotsky は「法則」という表現を使っているが、そのような表現は人間の存在からあたかも独立した自然の法則性であるかのような誤解をまねくだろう。この「法則」は実際には、社会を対象とする目的意識的な実践＝「技術的」実践である。だから「革命はアートである」と表現されるのであり、階級闘争はその時点での政治・経済的な情勢 (社会の法則性) 認識と目的意識性を統合したものであり、純粋な客観的法則性ではない。

Dewey は Trotsky が”law”を客観的法則性の意味で使っていると誤解しているし、Trotsky 自身もその点をあいまいにしている。Dewey の批判に応えるには、階級闘争＝社会的実践が目的意識的に遂行されること、その意味でそれが客観的な法則性とは違うことを明言しなければならない。そこには「手段」すなわち、各国で現在を生き、各地の伝統的風土・文化を伝え、それぞれの未来の展望を描いている労働者階級 (それはあたかも一人の人間のように多様性をもち、自由を謳歌し、規律を重んじる) が存在している。目的は手段を (自己) 規定し、手段は目的を作り出す。このような注釈を加えた上で、Trotsky の結論「すべての手段が正当化されるわけではない」はきわめて正しいものとして認められるべきである。

織への信頼を低下させこれを「指導者」への崇拝におき替えることなどである。

- 弁証法的唯物論は、手段と目的の二元論を知らない。目的は歴史の運動から自然に立ち上る。手段は目的と組織的に結びつく*29。即自的目的はより高次の目的のための手段となる。
- 労働者階級の解放は労働者階級自身の仕事である。ゆえに、大衆を騙すこと、敗北を勝利として喧伝すること、友人を敵として扱うこと、労働者の指導者たちを賄賂で籠絡し、伝説を捏造し、偽りの裁判を行うこと、一言で言えばスターリニストの所業は犯罪以外の何物でもない。これらの「手段」は唯一の「目的」、すなわち Starlin 一派の支配を延命させることにのみ奉仕している。第四インターナショナルがスターリニズムと生死をかけた闘争をしているのはこのためである。

*29 政策体系として知られる。この目的と手段の網の目は、同時に因果関係の網の目でもあり、選び取る小目標 (中間地点) の違いが最終的に到達する大目標 (終点) を変える。これを説明するのに Trotsky は Ferdinand Lassalle(1825-1864) の戯曲の登場人物の台詞を引用している。

付録 C

数学的付録

C.1　大数の法則の証明に見られる〈客観的法則性〉

「大数の弱法則」の証明は次の手順でなされる。

i). マルコフの不等式を示すこと。

$$Pr(|X| \geq a) \leq \frac{E(|X|)}{a} \tag{C.1}$$

ii). チェビシェフの不等式を得ること。

$$Pr(|X-\mu| \geq a) \leq \frac{E(X-\mu)^2}{a^2} \tag{C.2}$$

iii). 大数の法則を得ること。

$$\lim_{n\to\infty} Pr(|Y_n-\mu| \geq \epsilon) = 0 \tag{C.3}$$

ここで X は誤差、Y_n が n 回の試行による相対頻度である。この流れを成立させるために必要な条件は、i) 確率分布関数の一般的な性質 (そして「誤差がこの確率分布関数に従っている」という事実)、ii) 期待値の存在と iii) 試行間の独立性 (無相関性) である。これらの条件が満たされない試行においては、「大数の法則」は成立しない。これらの事実のなかに誤差のしたがう〈客観的法則性〉の存在が認められる。

i) マルコフの不等式は次のように示される $(a>0)$[*1]。

$$\begin{aligned}
E(|X|) &= \int_{-\infty}^{\infty} |x|dF(x) \\
&= \int_{a}^{\infty} |x|dF(x) + \int_{-a}^{a} |x|dF(x) \\
&\quad + \int_{-\infty}^{-a} |x|dF(x) \\
&\geq \int_{a}^{\infty} |x|dF(x) + \int_{-\infty}^{-a} |x|dF(x) \\
&\geq a\int_{a}^{\infty} dF(x) + a\int_{-\infty}^{-a} dF(x) \\
&= aPr(X \geq a) + aPr(X \leq -a) \\
&= aPr(|X| \geq a)
\end{aligned}$$

ii) のチェビシェフの不等式は式 (C.1) で $X \to (X-\mu)^2$、$a \to a^2$ の置き換えで得られる。なお、$\mu = E(X)$ である。

iii) は、次のように示される。$(\sigma^2 = E(X_i-\mu)^2)$

$$Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$$

$$E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

$$\begin{aligned}
E(Y_n-\mu)^2 &= \frac{1}{n^2}E\left(\sum_i (X_i-\mu)\right)^2 \\
&= \frac{1}{n^2}E\left(\sum_i (X_i-\mu)^2 \right. \\
&\quad \left. + 2\sum_{i\neq j}(X_i-\mu)(X_j-\mu)\right) \\
&= \frac{1}{n^2}\sum_i E(X_i-\mu)^2 = \frac{\sigma^2}{n}
\end{aligned}$$

これらより、式 (C.2) で、$X \to Y_n$、$a \to \epsilon$ と置き換えることにより、

[*1] ここでの積分は離散分布、連続分布の双方に適用するためにリーマン＝スティルチェス積分で表記している。$F(x)$ は確率分布関数。

$$Pr(|Y_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

両辺で $n \to \infty$ の極限をとり、

$$\lim_{n \to \infty} Pr(|Y_n - \mu| \geq \epsilon) = 0$$

ベルヌーイ試行 (コイン投げ) における大数の法則は、次のように説明される。

誤差 X_i は予測 (=「表」) が当たるか ($X_i = 0$)、外れるか ($X_i = 1$) で示される。n 回の試行による外れの相対頻度は $Y_n = \sum_i X_i/n$ で表現される。裏の出る確率が μ(表の出る確率が $1 - \mu$) ならば、相対頻度は n を増やすにつれて $Y_n \to \mu$ に近づいていくだろう。

さらに言えば、コインについてどちらの面を「表」と名付けようと有利さにおいて事態が変わらないように見受けられる場合 (「同様に確からしい」)、$\mu = 1 - \mu$ より、$\mu = 0.5$ を結論しても良いかもしれない[*2]。あるいは、$\mu \neq 0.5$ という仮説を設定し、検定することも考えられる。いずれにせよ、最終的な判断結果はコインの客観的な性質にも依存しており、主観のみによって決まるのではない。

[*2] だからこそ、コイン投げをゲームの手段として用いようとする動機が生じる。ただし、これはコインが細工されていないことを前提としてである。

C.2 Pascal の問題の母関数による解法

ここでは Pascal の問題の母関数による解法を解説する。

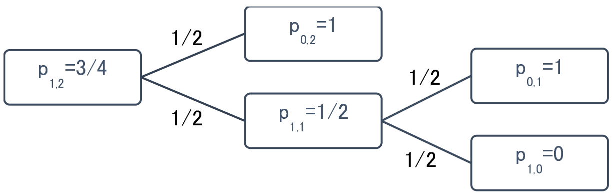


図 C.1 Pascal の問題

Pascal は次のように考えた。$p_{1,2}$ の状態から出発し、もし A が勝てば $p_{0,2}=1$ で勝負がつく、B が勝てば $p_{1,1}$ の状態に移る。さらにそこから A が勝てば $p_{0,1}=1$、B が勝てば $p_{1,0}=0$ で必ず勝敗がつく。その状態から逆算して、$p_{1,1}=1/2$、$p_{1,2}=3/4$ と決まる。そこで Pascal は、A にたいして賭け金の 3/4 倍である 48 ピストールを分配すべきと回答した。

式 (3.1) は上の関係を一般化したものである。これは偏差分方程式である。これの一般項を Laplace は母関数を用いて解いている。

まず解の見通しをたてるために、$i=1$ と固定した場合について考察する。

$$p_{1,j}=\frac{1}{2}p_{0,j}+\frac{1}{2}p_{1,j-1}$$

両辺に x^j を乗じて、$j=1,\cdots,\infty$ で総和をとる。ここで $p_{0,j>0}=1$ に注意。

$$\begin{aligned}\sum_{j=1}^{\infty}p_{1,j}x^j&=\frac{1}{2}\sum_{j=1}^{\infty}x^j+\frac{1}{2}\sum_{j=1}^{\infty}p_{1,j-1}x^j\\&=\frac{1}{2}\sum_{j=1}^{\infty}x^j+\frac{x}{2}\sum_{j=1}^{\infty}p_{1,j}x^j \qquad \text{(C.4)}\end{aligned}$$

ここで、$A_1(x)=\sum_{j=0}^{\infty}p_{1,j}x^j$ とおく。これが $p_{1,j}$ の (通常) 母関数[*3]と呼ばれるものである。

式 (C.4) を整理すると、

$$A_1(x)=\frac{x}{2(1-x)}+\frac{x}{2}A_1(x)$$

より、

$$\begin{aligned}A_1(x)&=\frac{x}{(1-x)(2-x)}=\frac{1}{1-x}-\frac{2}{2-x}\\&=\sum_{j=0}^{\infty}x^j-\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j x^j\\&=\sum_{j=0}^{\infty}\left\{1-\left(\frac{1}{2}\right)^j\right\}x^j=\sum_{j=0}^{\infty}p_{1,j}x^j \qquad \text{(C.5)}\end{aligned}$$

辺々比較することにより、

$$p_{1,j}=1-\left(\frac{1}{2}\right)^j$$

次に、任意の i について同様の計算を行うと次の結果が得られる。

$$p_{i,j}=\frac{1}{2}p_{i-1,j}+\frac{1}{2}p_{i,j-1}$$

$$A_i(x)=\frac{1}{2-x}A_{i-1}(x),\quad A_i(x)=\sum_{j=0}^{\infty}p_{i,j}x^j$$

ここで、

$$A_0(x)=\frac{x}{1-x}$$

また $A_i(x)$ のすべての母関数は次のとおりである。

$$M(x,y)=\sum_{i=0}^{\infty}A_i(x)y^i=\frac{x(2-x)}{(1-x)(2-x-y)}$$

$A_i(x)$ の一般項 (すなわち $M(x,y)$ の一般項) は、次のようにして得られる (D は形式的微分演算子)。

$$\begin{aligned}A_i(x)&=\frac{x}{(1-x)(2-x)^i}\\&=\frac{1}{1-x}-\sum_{k=1}^{i-1}\frac{1}{(2-x)^k}-\frac{2}{(2-x)^i}\\&=\frac{1}{1-x}-\sum_{k=1}^{i-1}\frac{D^{k-1}}{(k-1)!}\left(\frac{1}{2-x}\right)\\&\quad-\frac{2D^{i-1}}{(i-1)!}\left(\frac{1}{2-x}\right)\end{aligned}$$

[*3] 形式的ベキ級数とも呼ばれる。このばあい、x は変数ではなく標識にすぎず、かならずしも収束は考慮されない。ちなみに $x=z^{-1}$ としたものは Z 変換と呼ばれ、ディジタル信号の解析に用いられる。

$$= \sum_{j=0}^{\infty} x^j \left\{ 1 - \sum_{k=1}^{i-1} \left(\frac{1}{2}\right)^{j+k} \binom{j+k-1}{k-1} - \left(\frac{1}{2}\right)^{j+i-1} \binom{j+i-1}{i-1} \right\} \tag{C.6}$$

ゆえに、

$$p_{i,j} = 1 - \sum_{k=1}^{i-1} \left(\frac{1}{2}\right)^{j+k} \binom{j+k-1}{k-1} - \left(\frac{1}{2}\right)^{j+i-1} \binom{j+i-1}{i-1} \tag{C.7}$$

C.3　母関数の諸性質

C.3.1　確率母関数について

確率母関数 (probability generating function:pgf) は (一変量のばあい) 以下のように定義される。

$$G_X(z) = \sum_{i=0}^{\infty} p_i z^i$$

ここで i は (離散的) 確率変数であり、i に対応する事象が存在しないときは $p_i = 0$ と規約する[*4]。z には一般に複素数が入るが、収束を議論しないばあいは実数とみなしても良いし、形式的記号として取り扱っても良い[*5]。

$z = 1$ を代入することにより、全確率の式 $G_X(1) = 1$ を得る。これは pgf の基本的な性質の一つである。

pgf の具体例：ベルヌーイ試行 (コイン投げ) で表の出る事象を $i = 1$、裏を $i = 0$ と規約する[*6]。また表の出る確率を p とする (不正のないコインならば $p = 1/2$)。

$$G_X(z) = q + pz \ \ (\text{ただし、} p + q = 1)$$

特殊化：確実性[*7]

$$G_X(z) = 1$$

一つの事象 $i = 0$ のまま変化しない。つまりそれ以外の事象は発生しない ($p_{i \neq 0} = 0$)。

一般化 1：離散型一様分布

複数の事象について、すべての母数が同一 ($p_i = 1/n,\ n = b + 1 - a$) のものは離散型一様分布と呼ばれる。たとえば、一つのサイコロの出る目の数は $a = 1$、$b = 6$ の一様分布に従う。

$$G_X(z) = \sum_{i=a}^{b} \frac{1}{n} z^i = \frac{1}{n} \left(\frac{z^a - z^{b+1}}{1 - z} \right)$$

一般化 2(多変量化[*8])：カテゴリカル分布

$$G_X(z_{i=0,1,\cdots,n-1}) = \sum_{i=0}^{n-1} p_i z_i$$
$$(\text{ただし、} z_0 = 1\text{、} \sum_{i=0}^{n-1} p_i = 1)$$

たとえば、一人の調査回答者の、n 個の選択肢をもつ単一回答の設問にたいする反応はカテゴリカル分布に従う。$n = 2$ のばあいはベルヌーイ試行に相当する。なお、独立変量の個数は n ではなく、母数の制約[*9]によって $n - 1$ となっていることに注意する。

pgf から確率を引き出すには、直接的にベキ級数に展開して係数を見る方法、または以下のテイラー展開による方法がある。

$$p_k = \frac{1}{k!} \cdot \frac{\partial^k}{\partial z^k} G_X(z) \bigg|_{z \to 0} = \frac{1}{k!} G_X^{(k)}(0)$$

上のテイラー展開から pgf と確率 (分布) とが一対一対応すること (pgf の一意性) がわかる。この事実は、未

[*4] p_i の代わりに $Pr(X = i)$ と表記することもある。i は基礎的な事象に便宜的 (名目的に) 数値を割り当てたものであり、さしあたりはこれらを入れ替えても事象の名称が変化しただけであり、何も変わらないことに注意する。ただし、後々にはこの確率変数の尺度を名目のものとしてばかりでなく、変数間で順序比較、演算が可能なものと解釈し直す。

[*5] 1820 年に A.Cauchy が無限級数の収束を論じた後に、Laplace は自著の中の級数の収束を確認するためにしばし面会謝絶としたというほほえましいエピソードが伝わっている。級数の収束性が Cauchy 以前には軽んじられていたのであり、Laplace もそれに無自覚であったと解釈するのが自然であろう。ただし、今日では級数の収束を問題にせずこれを一種の代数構造とみなす形式的べき級数の理論も発展しており、その意味では収束問題を過剰に意識する必要はないのではとも思われる。

[*6] ここでは確率とパラメータ (母数) が区別されていないことに注意する。

[*7] 退化分布 (degenerate distribution)、または Heaviside の階段関数とも呼ばれる。

[*8] 多変量確率変数はベクトル値をもち、同時確率と周辺確率の区別、変量間の共分散 (したがって相関) を明示的に分析できるという意味で、因果関係を理解する上で重要である。なお、カテゴリカル分布は複数のベルヌーイ試行から、それらの混合分布 (mixture distribution) として導出できる。

[*9] Birch(1963) の制約と呼ばれる。

知の確率分布を pgf の構成から明らかにするばあいに役立てられる。

互いに独立な確率変数の和の pgf

$$G_X(z) = \sum_{i=0}^{\infty} p_i z^i,\ \ G_Y(z) = \sum_{j=0}^{\infty} q_j z^j$$

とするとき、両者の積は、

$$G_X(z)G_Y(z) = \sum_{k=0}^{\infty} r_k z^k,\ \text{ただし}\ r_k = \sum_{i=0}^{k} p_i q_{k-i}$$

ここで、

$$\begin{aligned} r_0 &= p_0 q_0 \\ r_1 &= p_0 q_1 + p_1 q_0 \\ r_2 &= p_0 q_2 + p_1 q_1 + p_2 q_0 \\ &\ \vdots \end{aligned}$$

のように、r の添え字 $= p$ の添え字 $+ q$ の添え字 を満たしていることがわかる。これから、確率変数 i、j の和である新しい確率変数 k の従う確率分布の pgf は、もとの pgf の積によって表示されることがわかる。

例：同一のカテゴリカル分布に従う m 個の確率変数の和は多項分布に従う ($n = 2$ のばあいは二項分布)。たとえば、m 人の回答者による調査結果はこれによって解釈される。その pgf は、

$$G_X(z_{i=0,1,\cdots,n-1}) = \left(\sum_{i=0}^{n-1} p_i z_i\right)^m$$
$$(\text{ただし、}z_0 = 1\text{、}\sum_{i=0}^{n-1} p_i = 1)$$

調査の結果としてデータ $(k_0, k_1, \cdots, k_{n-1})$ が得られる確率は、下の多項定理による展開の各項に相応する。

$$\begin{aligned} 1 = G_X(1) &= \left(\sum_{i=0}^{n-1} p_i\right)^m \\ &= \sum_{k_0+k_1+\cdots+k_{n-1}=m} \frac{m!}{k_0!k_1!\cdots k_{n-1}!} p_0^{k_0} p_1^{k_1} \cdots p_{n-1}^{k_{n-1}} \end{aligned}$$

確率変数の和 (確率分布の畳み込み)$X+Y$ は、pgf を特殊な期待値ととらえることにより、次のように簡略に整理される。

$$\begin{aligned} G_X(z) &= \sum_{i=0}^{\infty} p_i z^i = E\left(z^X\right) \\ &\to G_X(z)G_Y(z) = E(z^X)E(z^Y) = E(z^{X+Y}) \\ &= G_{X+Y}(z) \end{aligned}$$

同様に、確率変数 X のスカラー倍 tX の pgf は次のように表示される。

$$G_X(z^t) = E\{(z^t)^X\} = E(z^{tX}) = G_{tX}(z)$$

応用例として、確率変数の差 $X - Y$ の pgf は次のように表示される。

$$\begin{aligned} G_X(z)G_Y(z^{-1}) &= E(z^X)E(z^{-Y}) = E(z^{X-Y}) \\ &= G_{X-Y}(z) \end{aligned}$$

pgf の微分に $z = 0$ を代入したものがもとの確率に対応する ($G_X^{(k)}(0) = k!p_k$) とすれば、$z = 1$ を代入したものは、階乗モーメント (factorial moment) に相応する。

$$G_X^{(k)}(1) = E\left(\frac{X!}{(X-k)!}\right)$$

これより、各モーメントが逐次的に定まる。例えば、

$$\begin{aligned} E(X) &= G_X^{(1)}(1) \\ \mathrm{Var}(X) &= G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2 \end{aligned}$$

例：二項分布の平均と分散

$$\begin{aligned} G_X(z) &= (q + pz)^n \\ G_X^{(1)}(z) &= np(q + pz)^{n-1} \\ &\to G_X^{(1)}(1) = np \\ G_X^{(2)}(z) &= n(n-1)p^2(q + pz)^{n-2} \\ &\to G_X^{(2)}(1) = n(n-1)p^2 \end{aligned}$$

より、

$$E(X) = np,\ \ \mathrm{Var}(X) = np(1-p)$$

例 2：母平均 μ、母分散 σ^2 をもつ同一の一般的確率変数の和の平均と分散

ここで X_i の母平均と母分散が μ、σ^2 であるということは、次のことを意味する。

$$\begin{aligned} G_{X_i}^{(1)}(1) &= \mu \\ G_{X_i}^{(2)}(1) &= \sigma^2 - \mu + \mu^2 \end{aligned}$$

ここで、$S=\sum_{i=1}^{n} X_i$ とする。

$$
\begin{aligned}
G_S(z) &= [G_{X_i}(z)]^n \\
G_S^{(1)}(z) &= n\,[G_{X_i}(z)]^{n-1}\,G_{X_i}^{(1)}(z) \\
&\quad \to G_S^{(1)}(1) = n\mu \\
G_S^{(2)}(z) &= n(n-1)\,[G_{X_i}(z)]^{n-2}\left[G_{X_i}^{(1)}(z)\right]^2 \\
&\quad +n\,[G_{X_i}(z)]^{n-1}\,G_{X_i}^{(2)}(z) \\
&\quad \to G_S^{(2)}(1) = n\sigma^2 + n^2\mu^2 - n\mu
\end{aligned}
$$

よって、

$$E(S) = n\mu,\ \ \mathrm{Var}(S) = n\sigma^2$$

C.3.2　積率母関数について

積率母関数 (moment generating function: mgf) $M_X(z)$ とは、確率母関数 $G_X(z)$ において、$z \to e^z (= \sum_{k=0}^{\infty} \frac{1}{k!} z^k)$ としたもののことである。mgf は確率変数間の演算にかんして pgf と同様の性質を保ち、しかも連続的変量にも適用できるという利点をもつ[*10]。

mgf の定義をあらためて記述すると次のとおりとなる[*11]。ここで $f(x)$ は確率密度関数である。

$$
\begin{aligned}
M_X(z) &= G_X(e^z) \\
&= \int_{-\infty}^{\infty} dP(x) e^{zx} = \begin{cases} \displaystyle\int_{-\infty}^{\infty} f(x) e^{zx} dx \\ \displaystyle\sum_{x=0}^{\infty} p_x e^{xz} \end{cases}
\end{aligned}
$$

n 個の互いに独立な (mgf の判明している) 確率変数について、その総和を基準化した確率変数を構成するものとする。この確率変数の従う確率分布は、$n \to \infty$ で、ある分布 (標準正規分布) に近づいていく。この事実は中心極限定理 (central limit theorem:CLT)[*12] と呼ばれるが、これが成立することを mgf を使って証明できる。

まず構成要素となるもとの確率変数を $X_i (i = 1, \cdots, n)$、その母平均、母分散をそれぞれ μ、σ^2 とする。最終的に構成される確率変数 T は以下のように記述される。また T は、ひとつひとつの X_i を基準化した T_i の和にも分解できることも容易に見て取れる。ここで、$E(T_i) = 0,\ \mathrm{Var}(T_i) = 1$ である。

$$
\begin{aligned}
T_i &= \frac{X_i - \mu}{\sigma} \\
T &= \frac{\displaystyle\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_i
\end{aligned}
$$

T_i の mgf と各モーメントは次のように計算される。

$$
\begin{aligned}
M_{T_i}(z) &= E(e^{zT_i}) \\
&\quad \to M_{T_i}(0) = 1 \\
M_{T_i}^{(1)}(z) &= E(T_i e^{zT_i}) \\
&\quad \to M_{T_i}^{(1)}(0) = E(T_i) = 0 \\
M_{T_i}^{(2)}(z) &= E(T_i^2 e^{zT_i}) \\
&\quad \to M_{T_i}^{(2)}(0) = E(T_i^2) \\
&\quad = \mathrm{Var}(T_i) + [E(T_i)]^2 = 1
\end{aligned}
$$

ここで mgf の自然対数をとった関数 $\psi_{T_i}(z)$ を定義する。これはキュムラント母関数 (cumulant generating function: cgf) と呼ばれる。T_i の cgf と各キュムラントを計算すると次のようになる。

$$
\begin{aligned}
\psi_{T_i}(z) &= \ln M_{T_i}(z) \\
&\quad \to \psi_{T_i}(0) = 0 \\
\psi_{T_i}^{(1)}(z) &= \frac{M_{T_i}^{(1)}(z)}{M_{T_i}(z)} \\
&\quad \to \psi_{T_i}^{(1)}(0) = 0
\end{aligned}
$$

[*10] pgf が通常母関数を基礎としているのにたいして、mgf は指数型母関数にもとづく。また mgf は両側ラプラス変換と実質的に同じもの (z の符号を変えたもの) であるが、これは pgf が z 変換に関係するのと似ている。

[*11] 添え字は、数列ではなく関数であることを強調し、虚数単位との誤解を避けるために i ではなく x に変更する。添え字の動く範囲は p_x の定義される範囲に合わせる。

[*12] CLT は大数の法則を含む。また CLT はもとの確率変数が mgf をもつ、という条件よりも弱い条件 (リンドベルグ条件) のもとで成立する。逆に、その成立条件があるということは、あらゆる確率変数の和が正規分布に近づくというわけではないこと、大数の法則と同様に CLT も自然の客観的法則性を定理・法則として定立したものであることを意味している。いみじくも H.Cramer は次のように語ったと言われるが (竹内・大橋 (1981))、このばあいは数学者の発言の方に分がある。「誰もが誤差法則 (誤差分布の正規性) を信じている。実験家は数学的定理であると思っているからであり、数学家は実験的事実と思っているからである。—H.Cramer,1946」

$$\psi_{T_i}^{(2)}(z) = \frac{M_{T_i}^{(2)}(z)\cdot M_{T_i}(z) - \left[M_{T_i}^{(1)}(z)\right]^2}{\left[M_{T_i}(z)\right]^2}$$
$$\to \psi_{T_i}^{(2)}(0) = 1$$

$\psi_{T_i}(z)$ を 3 次まで原点周りで展開することにより、cgf の近似式が次のように得られる[*13]。

$$\begin{aligned}\psi_{T_i}(z) &= \psi(0)_{T_i} + \frac{\psi_{T_i}^{(1)}(0)}{1!}z + \frac{\psi_{T_i}^{(2)}(0)}{2!}z^2 + \frac{\psi_{T_i}^{(3)}(\theta)}{3!}z^3 \\ &= \frac{1}{2}z^2 + \frac{\psi_{T_i}^{(3)}(\theta)}{6}z^3 \end{aligned} \tag{C.8}$$

T の mgf は、T_i の mgf を用いて次のように計算される。最終行にあらわれるものが、標準正規分布の mgf である。

$$\begin{aligned}M_T(z) &= \left[M_{T_i}\left(\frac{z}{\sqrt{n}}\right)\right]^n \\ &= \exp\left\{n\psi_{T_i}\left(\frac{z}{\sqrt{n}}\right)\right\} \\ &= \exp\left\{n\left[\frac{1}{2}\left(\frac{z}{\sqrt{n}}\right)^2 + \frac{\psi_{T_i}^{(3)}(\theta)}{6}\left(\frac{z}{\sqrt{n}}\right)^3\right]\right\} \\ &= \exp\left\{\frac{1}{2}z^2 + \frac{\psi_{T_i}^{(3)}(\theta)}{6}\frac{z^3}{\sqrt{n}}\right\} \\ &\overset{n\to\infty}{\longrightarrow} \exp\left(\frac{1}{2}z^2\right)\end{aligned}$$

正規分布の mgf から以下のようにその確率密度関数を得ることができる。ここで式 (C.9) は両側ラプラス変換である[*14]。下で C_0 は積分定数であり、全確率が 1 となるように決定される ($\sqrt{2\pi}e^{C_0} = 1$)。

$$\begin{array}{ccc} M_T(z) = \exp\left(\frac{1}{2}z^2\right) & & \\ \downarrow \quad z = -s & & \\ \ln\phi(s) = \frac{1}{2}s^2 \to & & \frac{d}{ds}\phi(s) = s\phi(s) \\ & & \updownarrow \quad (C.9) \\ \frac{d}{dx}\ln f_0(x) = -x \leftarrow & & -xf_0(x) = \frac{d}{dx}f_0(x) \\ \downarrow & & \\ f_0(x) = \exp\left(-\frac{x^2}{2} + C_0\right) & & = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\end{array}$$

$$\begin{aligned}\frac{d}{ds}B\left[f(x)\right](s) &= B\left[-xf(x)\right](s) \\ B\left[f'(x)\right](s) &= sB\left[f(x)\right](s)\end{aligned} \tag{C.9}$$

標準正規分布 $N(0,1)$ の密度関数 f_0 から、母数 μ、σ^2 をもつ一般の正規分布 $N(\mu,\sigma^2)$ の密度関数も以下のように得られる。

$$\begin{aligned}f(x) &= f_0\left(\frac{x-\mu}{\sigma}\right)\cdot\left|\left(\frac{x-\mu}{\sigma}\right)'\right| \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\end{aligned}$$

mgf は多変量 (同時) 確率分布に容易に拡大される。joint mgf は次のように定義される。

$$M_{\boldsymbol{X}}(\boldsymbol{z}) = E\left[\exp\left\{\boldsymbol{z}^t\boldsymbol{X}\right\}\right]$$

これにたいして、一変量を除き他の変量をすべて積分消去 (integrate out) した周辺確率分布には次の marginal mgf が対応する。また、これは joint mgf において、z_i 以外の要素をすべて 0 としたものに等しい。

$$\begin{aligned}M_{X_i}(z_i) &= E\left[\exp\left\{z_iX_i\right\}\right] \\ &= M_{\boldsymbol{X}}(0,\cdots,0,z_i,0,\cdots,0)\end{aligned}$$

確率変数ベクトルの各要素が互いに独立であるとき、joint mgf は marginal mgf の積に以下のように分解される (独立性判定の因数分解規準)。

$$\begin{aligned}M_{\boldsymbol{X}}(\boldsymbol{z}) &= E\left[\exp\left\{\boldsymbol{z}^t\boldsymbol{X}\right\}\right] \\ &= E\left[\prod_i \exp\left\{z_iX_i\right\}\right] \\ &\downarrow \text{indep} \\ &= \prod_i E\left[\exp\left\{z_iX_i\right\}\right] \\ &= \prod_i M_{X_i}(z_i)\end{aligned}$$

[*13] 4 次以降の項はいずれ消去される。

[*14] mgf とラプラス変換の違いに対応して z を $-s$ に置き換えている (この場合は結果は変わらない)。

C.4　多変量確率分布について

C.4.1　多変量正規分布

多変量正規分布は、p 個の標準正規分布に従う確率変数 ϵ_i から構成できる。

$$\begin{aligned} P(\boldsymbol{\epsilon}) &= \prod_{i=1}^{p} \int_{-\infty}^{\epsilon_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\epsilon_i^2\right\} d\epsilon_i \\ &= \int_D (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\epsilon}^t\boldsymbol{\epsilon}\right\} d^p\boldsymbol{\epsilon} \end{aligned}$$

x_i を ϵ_i の線型変換によって生じる確率変数とする。ここで、$\boldsymbol{\mu}$、$\boldsymbol{\Sigma}$ はそれぞれ母平均ベクトル、母共分散行列; 正定値対称行列である。また、偏回帰係数 $\boldsymbol{A}$ は $\boldsymbol{\Sigma}^{-1}$ のコレスキー分解となっている[*15]。$\boldsymbol{\Sigma}^{-1}$ は precision 行列、または concentration 行列と呼ばれる。これを基準化したものが偏相関行列である。

$$\begin{aligned} \boldsymbol{\epsilon} &= \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu}) \\ \boldsymbol{\epsilon}^t\boldsymbol{\epsilon} &= (\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{A}^t\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu}) \\ &= (\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ d^p\boldsymbol{\epsilon} &= |\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, d^p\boldsymbol{x} \end{aligned}$$

最終的に得られる分布関数は以下のとおりである。

$$\begin{aligned} p(\boldsymbol{x}) = \int_D & (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \\ & \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} d^p\boldsymbol{x} \end{aligned}$$

多変量正規分布の mgf は次のとおりである。

$$M_{\boldsymbol{X}}(\boldsymbol{z}) = \exp\left\{\boldsymbol{\mu}^t\boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^t\boldsymbol{\Sigma}\boldsymbol{z}\right\}$$

多変量正規分布の条件付き確率分布についてまとめる[*16]。$\boldsymbol{X}$ が式 (C.10) のように二つの部分に分けられるとする。ただし、それぞれの変数の数を $p_a + p_b = p$ とする。

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_a \\ \boldsymbol{X}_b \end{bmatrix} \tag{C.10}$$

同様にパラメータについても以下のように分割されるものとする。

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix},\ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}$$

$\boldsymbol{X}_b$ が与えられた時の $\boldsymbol{X}_a$ の条件付き分布の密度関数 $f_{a|b}$ の対数は、ベイズ・ルールを用いて次のように表現される。

$$\ln f_{a|b} = \ln f - \ln f_b$$

ただし、

$$\ln f = -\frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

$$\ln f_b = -\frac{p_b}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_{bb}| - \frac{1}{2}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)^t\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)$$

逆に言えば、元の同時分布 f は条件付き確率分布を用いて以下のように分解される。

$$\ln f = \ln f_{a\cap b} = \ln f_{a|b} + \ln f_b \tag{C.11}$$

$\ln f_{a|b}$(つまり、$\boldsymbol{X}_b$ が与えられた時の $\boldsymbol{X}_a$ の条件付き分布) を計算すると、次のような平均 $\boldsymbol{\lambda}$、共分散 $\boldsymbol{C}_{bb}$ の多変量正規分布となる。ここで、$\boldsymbol{C}_{bb}$ は $\boldsymbol{\Sigma}_{bb}$ のシュール補行列 (Schur complement) である[*17]。

$$\ln f_{a|b} = -\frac{p_a}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{C}_{bb}| - \frac{1}{2}(\boldsymbol{x}_a-\boldsymbol{\lambda})^t\boldsymbol{C}_{bb}^{-1}(\boldsymbol{x}_a-\boldsymbol{\lambda}) \tag{C.12}$$

$$\begin{aligned} \boldsymbol{C}_{bb} &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} \\ \boldsymbol{\lambda} &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b) \end{aligned} \tag{C.13}$$

回帰モデル (Regression model) は、上の条件付き確率分布の特殊ケースである[*18]。式 (C.10) のような分解で、$\boldsymbol{X}_a$ を被説明 (従属) 変数 (多くのばあい $p_a = 1$)、

[*15] 推定作業においては、標本共分散行列から母共分散行列、さらには偏回帰係数を得ることがひとつの目標となる。ここで $\boldsymbol{\Sigma}^{-1}$ は因果関係の方向が特定されていないという意味で相関関係 (無向グラフ) を示し、後者の $\boldsymbol{A}$ が本来の因果関係 (有向グラフ) を示す。前者に対応する推定法がグラフィカル・モデリング (または共分散選択)、後者のそれが共分散構造分析 (SEM) であるが、両者は実質的に同じものである。

[*16] 竹内・大橋 (1981) を参照のこと。

[*17] シュール補行列の性質について、詳しくは山本哲朗 (2010) を参照のこと。

[*18] Edwards(2000) を参照のこと。

$\boldsymbol{X}_b$ を説明 (独立) 変数と考え、以下のような表現を考える。

$$\underset{p_a\times 1}{\boldsymbol{X}_a} = \underset{p_a\times 1}{\boldsymbol{A}} + \underset{p_a\times p_b}{\boldsymbol{B}} \underset{p_b\times 1}{\boldsymbol{x}_b} + \underset{p_a\times 1}{\boldsymbol{\epsilon}} \tag{C.14}$$

ここで $\boldsymbol{\epsilon}$ は $N(\boldsymbol{O},\boldsymbol{\Psi})$ に従う残差ベクトルである。

式 (C.14) と式 (C.13) を比較することにより、以下を得る。

$$\begin{cases} \boldsymbol{B} = \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \\ \boldsymbol{A} = \boldsymbol{\mu}_a - \boldsymbol{B}\boldsymbol{\mu}_b \\ \boldsymbol{\Psi} = \boldsymbol{C}_{bb} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{B}\boldsymbol{\Sigma}_{ba} \end{cases} \tag{C.15}$$

ここで、幾つかの単純な一変量正規分布の合成 (compound) により構造をもった多変量正規分布が出現することを示す。そのためには、条件付き期待値にかんする次の事実を用いる[*19]。

$$E[g(Y)] = E\left[E[g(Y)|X]\right]$$

まず二つの確率変数が次のような Data Generating Model(Process); DGM に従っているとする。

$$X_1 \sim N(0,\sigma_1^2),\ \ X_2 \sim N(X_1,\sigma_2^2)$$

$\boldsymbol{X} = (X_1, X_2)$ 全体の mgf は次のように計算される。

$$\begin{aligned}
M_{\boldsymbol{X}}(\boldsymbol{z}) &= E_{\boldsymbol{X}}\left[e^{\boldsymbol{z}^t\boldsymbol{X}}\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}[e^{z_2X_2}|X_1]\right] \\
&= E_{X_1}\left[e^{z_1X_1}M_{X_2|X_1}(z_2)\right] \\
&= E_{X_1}\left[e^{z_1X_1}\exp\left\{z_2X_1 + \frac{1}{2}\sigma_2^2z_2^2\right\}\right] \\
&= E_{X_1}\left[e^{(z_1+z_2)X_1}\right]\exp\left(\frac{1}{2}\sigma_2^2z_2^2\right) \\
&= M_{X_1}(z_1+z_2)\exp\left(\frac{1}{2}\sigma_2^2z_2^2\right) \\
&= \exp\left\{\frac{1}{2}\sigma_1^2(z_1+z_2)^2\right\}\exp\left(\frac{1}{2}\sigma_2^2z_2^2\right) \\
&= \exp\left\{\frac{1}{2}\boldsymbol{z}^t\boldsymbol{L}^t\boldsymbol{L}\boldsymbol{z}\right\}
\end{aligned}$$

ただし、

$$\boldsymbol{L} = \begin{pmatrix} \sigma_1 & \sigma_1 \\ 0 & \sigma_2 \end{pmatrix},\ \ \boldsymbol{L}^t\boldsymbol{L} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 \end{pmatrix}$$

同様のことは三つ以上の確率変数のあいだにも成立する。たとえば次の DGM を想定する。

$$X_1 \sim N(0,\sigma_1^2),\ \ X_2 \sim N(X_1,\sigma_2^2),\ \ X_3 \sim N(X_2,\sigma_3^2)$$

$$\begin{aligned}
M_{\boldsymbol{X}}(\boldsymbol{z}) &= E_{\boldsymbol{X}}\left[e^{\boldsymbol{z}^t\boldsymbol{X}}\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{z_2X_2}\right.\right. \\
&\qquad \left.\left.E_{X_3|X_1,X_2}\left[e^{z_3X_3}|X_1,X_2\right]\right]\right] \\
&\downarrow \quad X_1 \text{ と } X_3 \text{ は条件付き独立} \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{z_2X_2}M_{X_3|X_2}(z_3)\right]\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{z_2X_2}\right.\right. \\
&\qquad \left.\left.\exp\left\{X_2z_3 + \frac{1}{2}\sigma_3^2z_3^2\right\}\right]\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{(z_2+z_3)X_2}\right]\right] \\
&\qquad \exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= E_{X_1}\left[e^{z_1X_1}M_{X_2|X_1}(z_2+z_3)\right]\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= E_{X_1}\left[e^{z_1X_1}\exp\left\{X_1(z_2+z_3)\right.\right. \\
&\qquad \left.\left.+\frac{1}{2}\sigma_2^2(z_2+z_3)^2\right\}\right]\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= E_{X_1}\left[e^{(z_1+z_2+z_3)X_1}\right] \\
&\qquad \exp\left\{\frac{1}{2}\sigma_2^2(z_2+z_3)^2\right\}\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= M_{X_1}(z_1+z_2+z_3) \\
&\qquad \exp\left\{\frac{1}{2}\sigma_2^2(z_2+z_3)^2\right\}\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= \exp\left\{\frac{1}{2}\sigma_1^2(z_1+z_2+z_3)^2\right\} \\
&\qquad \exp\left\{\frac{1}{2}\sigma_2^2(z_2+z_3)^2\right\}\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= \exp\left\{\frac{1}{2}\boldsymbol{z}^t\boldsymbol{L}_a^t\boldsymbol{L}_a\boldsymbol{z}\right\}
\end{aligned}$$

ただし、

$$\boldsymbol{L}_a = \begin{pmatrix} \sigma_1 & \sigma_1 & \sigma_1 \\ 0 & \sigma_2 & \sigma_2 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

$$\boldsymbol{L}_a^t\boldsymbol{L}_a = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2+\sigma_3^2 \end{pmatrix}$$

また、precision 行列は次のようになる。X_1 と X_3 のあいだの偏相関係数が 0 となっていることに留意する。

[*19] これは全確率の公式そのものでもある。$g(Y)$ は確率変数 Y にかんする任意の関数である。

$$\left(\boldsymbol{L}_a^t\boldsymbol{L}_a\right)^{-1} = \begin{pmatrix} \sigma_1^{-2}+\sigma_2^{-2} & -\sigma_2^{-2} & 0 \\ -\sigma_2^{-2} & \sigma_2^{-2}+\sigma_3^{-2} & -\sigma_3^{-2} \\ 0 & -\sigma_3^{-2} & \sigma_3^{-2} \end{pmatrix}$$

上と似ているが、X_3 が X_2 ではなく X_1 に影響されているばあいは次のようになる。

$$X_1 \sim N(0,\sigma_1^2),\ \ X_2 \sim N(X_1,\sigma_2^2),\ \ X_3 \sim N(X_1,\sigma_3^2)$$

$$\begin{aligned}
M_{\boldsymbol{X}}(\boldsymbol{z}) &= E_{\boldsymbol{X}}\left[e^{\boldsymbol{z}^t\boldsymbol{X}}\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2,X_3|X_1}\left[e^{z_2X_2+z_3X_3}|X_1\right]\right] \\
&\downarrow\ X_2 \text{ と } X_3 \text{ は条件付き独立} \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{z_2X_2}|X_1\right]\right. \\
&\qquad \left.E_{X_3|X_1}\left[e^{z_3X_3}|X_1\right]\right] \\
&= E_{X_1}\left[e^{z_1X_1}M_{X_2|X_1}(z_2)M_{X_3|X_1}(z_3)\right] \\
&= E_{X_1}\left[e^{z_1X_1}\exp\left\{X_1z_2+\frac{1}{2}\sigma_2^2z_2^2\right\}\right. \\
&\qquad \left.\exp\left\{X_1z_3+\frac{1}{2}\sigma_3^2z_3^2\right\}\right] \\
&= E_{X_1}\left[e^{(z_1+z_2+z_3)X_1}\right] \\
&\qquad \exp\left(\frac{1}{2}\sigma_2^2z_2^2\right)\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= M_{X_1}(z_1+z_2+z_3) \\
&\qquad \exp\left(\frac{1}{2}\sigma_2^2z_2^2\right)\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= \exp\left\{\frac{1}{2}\sigma_1^2(z_1+z_2+z_3)^2\right\} \\
&\qquad \exp\left(\frac{1}{2}\sigma_2^2z_2^2\right)\exp\left(\frac{1}{2}\sigma_3^2z_3^2\right) \\
&= \exp\left\{\frac{1}{2}\boldsymbol{z}^t\boldsymbol{L}_b^t\boldsymbol{L}_b\boldsymbol{z}\right\}
\end{aligned}$$

ただし、

$$\boldsymbol{L}_b = \begin{pmatrix} \sigma_1 & \sigma_1 & \sigma_1 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

$$\boldsymbol{L}_b^t\boldsymbol{L}_b = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma_3^2 \end{pmatrix}$$

また、precision 行列は次のように、X_2 と X_3 のあいだの偏相関係数が 0 となっている。

$$\left(\boldsymbol{L}_b^t\boldsymbol{L}_b\right)^{-1} = \begin{pmatrix} \sigma_1^{-2}+\sigma_2^{-2}+\sigma_3^{-2} & -\sigma_2^{-2} & -\sigma_3^{-2} \\ -\sigma_2^{-2} & \sigma_2^{-2} & 0 \\ -\sigma_3^{-2} & 0 & \sigma_3^{-2} \end{pmatrix}$$

C.4.2 多項分布

p 次元確率ベクトル $\boldsymbol{X}$ が多項分布に従っていることを $\boldsymbol{X} \sim \mathrm{Multi}(N,\boldsymbol{\theta})$ と表現する。ただし、$\boldsymbol{\theta}^t\boldsymbol{e}_p = 1, \boldsymbol{X}^t\boldsymbol{e}_p = N$ である[*20]。$N=1$ のばあいを特にカテゴリカル分布と呼ぶ。

多項分布の mgf は次のとおりである。

$$M_{\boldsymbol{X}}(\boldsymbol{z}) = \left\{\boldsymbol{\theta}^t\exp(\boldsymbol{z})\right\}^N$$

また、

$$\begin{aligned}
E(\boldsymbol{X}) &= \nabla M_{\boldsymbol{X}}\big|_{\boldsymbol{z}\to\boldsymbol{0}} \\
&= N\boldsymbol{\theta} \\
Var(\boldsymbol{X}) &= \nabla^2 M_{\boldsymbol{X}} - (\nabla M_{\boldsymbol{X}})(\nabla M_{\boldsymbol{X}})^t\big|_{\boldsymbol{z}\to\boldsymbol{0}} \\
&= N\left\{\mathrm{diag}(\boldsymbol{\theta}) - \boldsymbol{\theta}\boldsymbol{\theta}^t\right\}
\end{aligned}$$

多項分布とポアソン分布の関わりについて次のことを示す。

1). ポアソン分布の和 N はポアソン分布であること。
2). (多変量) ポアソン分布に N が与えられたときの条件付き分布は多項分布となること、同じことであるが、多項分布において N が定数ではなくポアソン分布にしたがうばあいの同時分布は (多変量) ポアソン分布となること。

1) について：

$X_{\boldsymbol{i}}$ がポアソン分布に従っているとき：$X_{\boldsymbol{i}} \sim P_o(\lambda_{\boldsymbol{i}})$、その mgf は次のようになる。

$$M_{X_{\boldsymbol{i}}}(z_{\boldsymbol{i}}) = \exp\left\{\lambda_{\boldsymbol{i}}(e^{z_{\boldsymbol{i}}}-1)\right\}$$

$N = \sum_{\boldsymbol{i}\in I} X_{\boldsymbol{i}}$ の mgf は、

$$\begin{aligned}
M_N(z) &= \prod_{\boldsymbol{i}\in I} M_{X_{\boldsymbol{i}}}(z) \\
&= \exp\left\{\left(\sum_{\boldsymbol{i}\in I}\lambda_{\boldsymbol{i}}\right)(e^z-1)\right\}
\end{aligned}$$

[*20] $\boldsymbol{e}_p$ はすべての要素が 1 である p 次元ベクトルを示す。

したがって、$N \sim P_o\left(\sum_{\boldsymbol{i}\in I}\lambda_{\boldsymbol{i}}\right)$

2) について：

$$
\begin{aligned}
M_{\boldsymbol{X},N}(\boldsymbol{z}) &= E_N\left[E_{\boldsymbol{X}|N}\left[e^{\boldsymbol{X}\boldsymbol{z}}|N\right]\right]\\
&= E_N\left[M_{\boldsymbol{X}|N}(\boldsymbol{z})\right]\\
&= E_N\left[\left(\sum_{\boldsymbol{i}\in I}\theta_{\boldsymbol{i}}e^{z_{\boldsymbol{i}}}\right)^N\right]\\
&= M_N\left(\ln\sum_{\boldsymbol{i}\in I}\theta_{\boldsymbol{i}}e^{z_{\boldsymbol{i}}}\right)\\
&= \exp\left\{\lambda\left(\sum_{\boldsymbol{i}\in I}\theta_{\boldsymbol{i}}e^{z_{\boldsymbol{i}}}-1\right)\right\}\\
&\downarrow \sum_{\boldsymbol{i}\in I}\theta_{\boldsymbol{i}}=1,\ \ \lambda\theta_{\boldsymbol{i}}=\lambda_{\boldsymbol{i}}\\
&= \exp\left\{\sum_{\boldsymbol{i}\in I}\lambda_{\boldsymbol{i}}\left(e^{z_{\boldsymbol{i}}}-1\right)\right\}\\
&= \prod_{\boldsymbol{i}\in I}M_{X_{\boldsymbol{i}}}(z_{\boldsymbol{i}})
\end{aligned}
$$

多項分布のうち部分要素が固定されたばあいの条件付き分布はやはり多項分布となる。たとえばセルの頻度を(x_1,x_2,x_3)とする三項分布のうちひとつx_3が固定された場合は次のようになる。

$$
\begin{aligned}
&\sum_{(x_1,x_2,x_3)}\frac{n!}{x_1!x_2!x_3!}p_1^{x_1}p_2^{x_2}p_3^{x_3}\\
&=\sum_{(x_3)}\frac{1}{x_3!}p_3^{x_3}\sum_{(x_1,x_2)}\frac{(n-x_3)!}{x_1!x_2!}\frac{n!}{(n-x_3)!}p_1^{x_1}p_2^{x_2}\\
&=\sum_{(x_3)}\frac{n!}{x_3!(n-x_3)!}p_3^{x_3}(1-p_3)^{n-x_3}\\
&\quad\sum_{(x_1,x_2)}\frac{(n-x_3)!}{x_1!x_2!}p_1^{x_1}p_2^{x_2}\frac{1}{(1-p_3)^{n-x_3}}\\
&=\sum_{(x_3)}\frac{n!}{x_3!(n-x_3)!}p_3^{x_3}(1-p_3)^{n-x_3}\\
&\quad\sum_{(x_1,x_2)}\frac{(n-x_3)!}{x_1!x_2!}\left(\frac{p_1}{p_1+p_2}\right)^{x_1}\left(\frac{p_2}{p_1+p_2}\right)^{x_2}
\end{aligned}
$$

複数のカテゴリカル分布の合成からカテゴリカル分布が生じることを示す。まずは、二つのベルヌーイ分布の合成が三項のカテゴリカル分布となることを示す。$\boldsymbol{X}_1=(X_1,1-X_1)$は母数が$\boldsymbol{\lambda}=(\lambda,1-\lambda)$のベルヌーイ分布にしたがう確率変数であり、$\boldsymbol{X}_2$は$X_1=1$のときに母数が$\boldsymbol{\mu}=(\mu,1-\mu)$のベルヌーイ分布、$X_1=0$のときに退化分布にしたがう確率変数とする[*21]。

$$\boldsymbol{X}_1\sim\mathrm{Bin}(1,\boldsymbol{\lambda}),\ \boldsymbol{X}_2\sim\mathrm{Bin}(X_1,\boldsymbol{\mu})$$

$\boldsymbol{X}_1$と$\boldsymbol{X}_2$の同時分布のmgfは次のようになる。

$$
\begin{aligned}
&M_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{z})=E_{\boldsymbol{X}_1,\boldsymbol{X}_2}\left[e^{\boldsymbol{z}_1^t\boldsymbol{X}_1+\boldsymbol{z}_2^t\boldsymbol{X}_2}\right]\\
&=E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t\boldsymbol{X}_1}E_{\boldsymbol{X}_2|\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_2^t\boldsymbol{X}_2}|\boldsymbol{X}_1\right]\right]\\
&=E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t\boldsymbol{X}_1}M_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{z}_2)\right]\\
&=E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t\boldsymbol{X}_1}\left(\sum_{j=1}^{2}\mu_je^{z_{2j}}\right)^{X_1}(1)^{1-X_1}\right]\\
&=E_{\boldsymbol{X}_1}\left[\exp\boldsymbol{X}_1\begin{pmatrix}z_{11}+\ln\sum_{j=1}^{2}\mu_je^{z_{2j}}\\ z_{12}+0\end{pmatrix}\right]\\
&=M_{\boldsymbol{X}_1}\begin{pmatrix}z_{11}+\ln\sum_{j=1}^{2}\mu_je^{z_{2j}}\\ z_{12}+0\end{pmatrix}\\
&=\lambda_1e^{z_{11}}\sum_{j=1}^{2}\mu_je^{z_{2j}}+\lambda_2e^{z_{12}}\\
&\downarrow(z_{i1},z_{i2})\equiv(z_i,0)\\
&=\lambda e^{z_1}\left\{\mu e^{z_2}+(1-\mu)\right\}+(1-\lambda)\\
&=\lambda\mu e^{z_1+z_2}+\lambda(1-\mu)e^{z_1}+(1-\lambda)
\end{aligned}
$$

これは$(z_1=z_2=\mathrm{Yes})$、$(z_1=\mathrm{Yes},z_2=\mathrm{No})$、$(z_1=\mathrm{No})$の三項からなるカテゴリカル分布である。

同様にして、p_1次のカテゴリカル分布と$p_2\times p_1$次のカテゴリカル分布の合成分布はやはりカテゴリカル分布となる。

$$\boldsymbol{X}_1\sim\mathrm{Multi}(1,\boldsymbol{\lambda}),\ \boldsymbol{X}_2\sim\mathrm{Multi}(\boldsymbol{X}_1,\boldsymbol{\mu})$$

$$
\begin{aligned}
&M_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{z})\\
&=E_{\boldsymbol{X}_1,\boldsymbol{X}_2}\left[e^{\boldsymbol{z}_1^t\boldsymbol{X}_1+\boldsymbol{z}_2^t\boldsymbol{X}_2}\right]
\end{aligned}
$$

[*21] ここでは多変量分布への拡大を意図してあえてベクトル表記している。

$$
\begin{aligned}
&= E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t \boldsymbol{X}_1} E_{\boldsymbol{X}_2|\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_2^t \boldsymbol{X}_2}|\boldsymbol{X}_1\right]\right] \\
&= E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t \boldsymbol{X}_1} M_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{z}_2)\right] \\
&= E_{\boldsymbol{X}_1}\left[e^{\boldsymbol{z}_1^t \boldsymbol{X}_1} \prod_{i=1}^{p_1}\left(\sum_{j=1}^{p_2}\mu_{ij}e^{z_{2ij}}\right)^{X_{1i}}\right] \\
&= E_{\boldsymbol{X}_1}\left[\exp \boldsymbol{X}_1\left(\boldsymbol{z}_1 + \ln \boldsymbol{\mu} e^{\boldsymbol{z}_2}\right)\right] \\
&= M_{\boldsymbol{X}_1}\left(\boldsymbol{z}_1 + \ln \boldsymbol{\mu} e^{\boldsymbol{z}_2}\right) \\
&= \sum_{i=1}^{p_1}\left(\lambda_i e^{z_{1i}} \sum_{j=1}^{p_2}\mu_{ij}e^{z_{2ij}}\right) \\
&= \sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\left(\lambda_i \mu_{ij} e^{z_{1i}+z_{2ij}}\right)
\end{aligned}
$$

C.4.3 CG 分布

カテゴリカル分布と正規分布の合成は CG (Conditional Gaussian) 分布と呼ばれる。これは離散型と連続型の確率変数双方を含む確率分布のひとつの例であり、分散分析、数量化理論 I 類 (または対応分析) の基礎を提供する。またカテゴリカル分布側が積分消去された周辺分布は混合正規分布となる。CG 分布の mgf は次のとおりである[*22]。

$$X_1 \sim \mathrm{Bin}(1,\lambda),\ \ X_2 \sim N(\mu X_1, \sigma^2)$$

$$
\begin{aligned}
&M_{X_1,X_2}(z_1,z_2) \\
&= E_{X_1,X_2}\left[e^{z_1X_1+z_2X_2}\right] \\
&= E_{X_1}\left[e^{z_1X_1}E_{X_2|X_1}\left[e^{z_2X_2}|X_1\right]\right] \\
&= E_{X_1}\left[e^{z_1X_1}M_{X_2|X_1}(z_2)\right] \\
&= E_{X_1}\left[e^{z_1X_1}\exp\left(\mu X_1 z_2 + \frac{1}{2}\sigma^2 z_2^2\right)\right] \\
&= E_{X_1}\left[\exp\left\{(z_1+\mu z_2)X_1\right\}\right]\exp\left(\frac{1}{2}\sigma^2 z_2^2\right) \\
&= M_{X_1}(z_1+\mu z_2)\cdot \exp\left(\frac{1}{2}\sigma^2 z_2^2\right) \\
&= \left\{\lambda e^{z_1+\mu z_2} + (1-\lambda)\right\}\exp\left(\frac{1}{2}\sigma^2 z_2^2\right) \\
&= \lambda \exp\left(z_1 + \mu z_2 + \frac{1}{2}\sigma^2 z_2^2\right) \\
&\quad +(1-\lambda)\exp\left(\frac{1}{2}\sigma^2 z_2^2\right)
\end{aligned}
$$

C.4.4 離散選択理論

離散選択理論のうち、たとえばロジット・モデルについて、次のようなモデルからデータが生成されるものとする。たとえば、θ はある消費者が一定の価格のもとである商品にたいしてもつ期待、X_1 は実際の商品の品質、X_2 はその商品を購入するかどうかの選択結果を表現する。

$$
\begin{aligned}
&X_1 \sim \mathrm{Logistic}(0,1) \\
&X_2 \sim \mathrm{Bin}(1,\lambda),\ \ \ln\frac{\lambda}{1-\lambda} = X_1 - \theta
\end{aligned}
$$

このモデルの同時確率分布にたいする mgf は次のように計算される。

$$
\begin{aligned}
&M_{X_1,X_2}(z_1,z_2) \\
&= E_{X_1}\left[e^{z_1X_1}M_{X_2|X_1}(z_2)\right] \\
&= E_{X_1}\left[e^{z_1X_1}\left\{\lambda e^{z_2} + (1-\lambda)\right\}\right] \\
&= (e^{z_2}-1)E_{X_1}\left[e^{z_1X_1}\lambda\right] + M_{X_1}(z_1)
\end{aligned}
$$

ここで、$M_{X_1}(z_1) = B(1+z_1, 1-z_1)$ である。

$E_{X_1}\left[e^{z_1X_1}\lambda\right]$ は次のように計算される。

$$
\begin{aligned}
&E_{X_1}\left[e^{z_1X_1}\lambda\right] \\
&= \int_{-\infty}^{\infty} e^{z_1X_1}\left(\frac{1}{1+e^{\theta-X_1}}\right)\left(\frac{1}{1+e^{-X_1}}\right)' dX_1 \\
&= \int_{-\infty}^{\infty} e^{z_1X_1}\frac{e^{-X_1}}{(1+e^{\theta-X_1})(1+e^{-X_1})^2}dX_1 \\
&= \int_{-\infty}^{\infty} e^{z_1X_1}\left(\frac{A}{1+e^{\theta-X_1}} + \frac{B}{1+e^{-X_1}}\right. \\
&\quad \left.+\frac{C}{(1+e^{-X_1})^2}\right)dX_1
\end{aligned}
$$

ここで、

$$A = -\frac{e^\theta}{(e^\theta-1)^2},\ B = \frac{1}{(e^\theta-1)^2},\ C = \frac{1}{e^\theta-1}$$

項別に変数変換しながら積分し、最終的に次を得る。

$$M_{X_1,X_2}(z_1,z_2)$$

[*22] 一見してわかるように、CG モデルは離散型変量→連続型変量への因果関係を記述できるが、その逆方向の因果関係には適用できない。本文を参照のこと。

$$= B(1+z_1, 1-z_1)$$
$$\left[1+\frac{e^{z_2}-1}{z_1}\left\{\frac{e^{(z_1+1)\theta}}{(e^\theta-1)^2}-\frac{z_1+1}{e^\theta-1}\right\}\right]$$

X_2 の期待値は mgf を用いて次のように計算される。

$$E_{X_1,X_2}\left[X_2\right]$$
$$= \left.\frac{\partial}{\partial z_2}M_{X_1,X_2}(z_1,z_2)\right|_{z_1\to 0, z_2\to 0}$$
$$= \frac{e^\theta(\theta-1)+1}{(e^\theta-1)^2}$$

これはロジスティック曲線ではないが、θ の増大とともに減少するシグモイド曲線であり、高すぎる期待に商品が応えられないときに、その購入頻度が急速に減少する様子を表現する。

プロビット・モデルについても同様の計算ができる。

$$X_1 \sim N(0,\sigma^2)$$
$$X_2 \sim \mathrm{Bin}(1,\lambda),\quad \lambda = \Phi(X_1-\theta)$$

$$M_{X_1,X_2}(z_1,z_2)$$
$$= (e^{z_2}-1)E_{X_1}\left[e^{z_1X_1}\lambda\right] + M_{X_1}(z_1)$$

ここで、$M_{X_1}(z_1)$ は、次のとおりである。

$$M_{X_1}(z_1) = \exp\left(\frac{1}{2}\sigma^2 z_1^2\right)$$

また $N(0,\sigma^2)$ の分布関数を $\Phi(x)$ とすれば、

$$E_{X_1}\left[e^{z_1X_1}\lambda\right]$$
$$= \int_{-\infty}^{\infty} e^{z_1X_1}\Phi(X_1-\theta)\Phi'(X_1)dX_1$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{1}{2}\sigma^2 z_1^2\right)$$
$$\int_{-\infty}^{\infty}\Phi(y)\exp\left\{-\frac{(y+\theta-\sigma^2 z_1)^2}{2\sigma^2}\right\}dy$$

上で $y = X_1-\theta$ と変数変換している。

Ng and Geller(1968) より次の関係式を導き出せる[*23]。

$$\int_{-\infty}^{\infty}\Phi(y)\exp\left\{-\frac{(y+A)^2}{2\sigma^2}\right\}dy$$
$$= \sqrt{2\pi}\sigma\left\{1-\Phi\left(\frac{A}{\sqrt{2}}\right)\right\}$$

$A = \theta-\sigma^2 z_1$ とおいて、最終的に次の mgf を得る。これは確かに CG 分布のそれとは一致しない。

$$M_{X_1,X_2}(z_1,z_2)$$
$$= \exp\left(\frac{1}{2}\sigma^2 z_1^2\right)$$
$$\left[1+(e^{z_2}-1)\left\{1-\Phi\left(\frac{\theta-\sigma^2 z_1}{\sqrt{2}}\right)\right\}\right]$$

X_2 の期待値は mgf より次のように計算される。

$$E_{X_1,X_2}\left[X_2\right] = 1-\Phi\left(\frac{\theta}{\sqrt{2}}\right)$$

C.5 Gauss の逆問題

誤差分布の正規性の論拠のひとつに、算術平均が最尤推定量となる確率分布は正規分布でなければならないことがあげられる (Gauss の逆問題)。竹内・大橋 (1981) を参考に、簡単にその内容をまとめる。

観測値 $x_i, (i=1,\cdots,n)$ が $x_i = \mu+\epsilon_i$ と表示されているものとして、その標本対数尤度は次のようになる。

$$L_{r;1}(\mu) = \sum_{i=1}^{n}\ln f(x_i-\mu)$$

これが算術平均 $\overline{x}$ で極小値をとるので $(g(x) \equiv (\partial/\partial x)\ln f(x)$ とすると)、

$$\left.\frac{\partial L_{r;1}}{\partial\mu}\right|_{\mu\to\overline{x}} = \sum_{i=1}^{n} g(x_i-\overline{x}) = 0 \qquad \text{(C.16)}$$

ここで、$\boldsymbol{x} = (a,a,\cdots,a,x_n)$ という特殊例をとると、

$$x_i - \overline{x} = \begin{cases} \dfrac{a-x_n}{n} = -y & (i=1,\cdots,n-1) \\ (1-n)\dfrac{a-x_n}{n} = (n-1)y & (i=n) \end{cases}$$

したがって、これを式 (C.16) に代入して、次の関数等式を得る。

$$\frac{g\left\{(n-1)y\right\}}{(n-1)y} = \frac{g(-y)}{-y}$$

[*23] $\int_{-\infty}^{\infty}\mathrm{erf}(x)e^{-(ax+b)^2}dx = -(\sqrt{\pi}/a)\mathrm{erf}(b/\sqrt{a^2+1})$ より。

ここで、$h(x) \equiv g(x)/x$ とすると、

$$h\left\{(n-1)y\right\} = h(-y)$$

さらに次の特殊例を考えると、結局 $h(x)$ は定数関数でなければならないとわかる。

$$\left\{\begin{array}{ll} n = 2 & h(y) = h(-y) \\ y = \dfrac{1}{n-1} & h(1) = h\left(-\dfrac{1}{n-1}\right) \end{array}\right\} \to h(x) = c$$

したがって、次のように $f(x)$ は正規分布の密度関数であることが判明する。

$$g(x) = cx \to f(x) = \exp\left(c_0 + c_1 x^2\right)$$

C.6　有向分離基準

ここでは有向分離基準について、その定義と使用例を解説する。有向分離基準 (d-separate) は、Pearl(1988) の与えた BN にもとづく条件付き独立の判定基準である。

Lauritzen ら (1990) のモラルグラフによる方法は、有向グラフを無向グラフに変換して判定するため、着実でかつ視覚的にすっきりした印象を与える。しかしその反面、判定対象に応じてグラフを変換する手間がかかる。

これにたいして有向分離基準は、頂点間を結ぶパスに着目するものである。これは、情報の流れ (かならずしも有向道：矢線の方向にしたがったものとは限らない) としての解釈の容易さ (「ブロックする」という用語法に代表される)、所与の BN を書き換えずに済むことなどの利点を備える。またバックドア基準などより進んだ分析用具はこの基準を基礎として開発されたため、それらを理解する上でも有向分離基準への理解を深めておくことは重要である。

有向分離基準は次のように定義される。

有向分離基準

パス P が次の条件のいずれかを満たすとき、「パス P は頂点集合 Z によって有向分離 (ブロック) される」という。

i). パス P は、ある頂点 m が Z に含まれるような連鎖 (chain)$i \to m \to j$、あるいは分岐 (fork)$i \leftarrow m \to j$ を含む。

ii). パス P は、頂点 m もその子孫も Z に含まれないような、合流 (collider)$i \to m \leftarrow j$ を含む。

頂点集合 Z が (X の任意の頂点と Y の任意の頂点をむすぶ) すべてのパスをブロックするとき、「集合 Z は X と Y を有向分離する」という。

有向分離基準はこのように明瞭に定義されているにもかかわらず、入り組んだ印象を与え、実際の利用に困難をもたらす。つまり、分析者はパス P と頂点集合 Z を同時に視野に入れつつ、条件 i) と ii) を交互に確認するという作業を強いられる。

表 C.1　Bayes Ball algorithm の 10 ルール

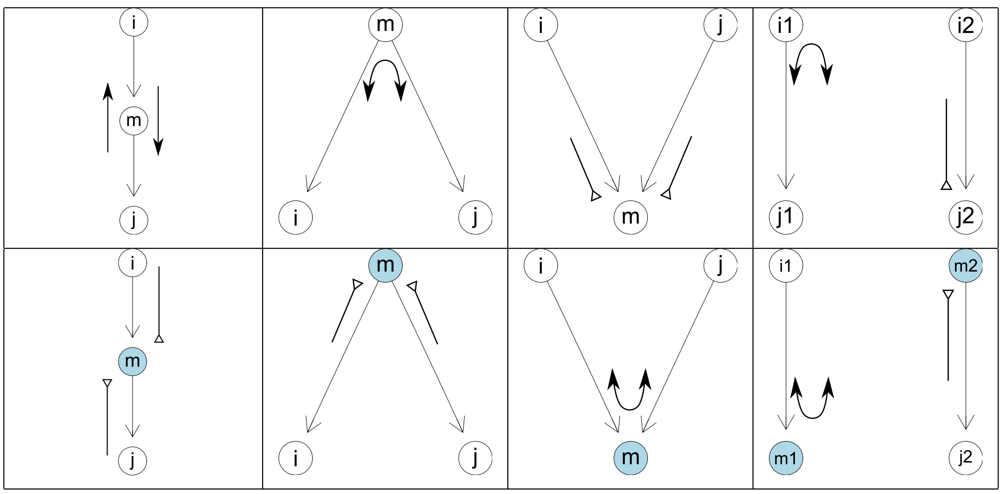


表 C.2　有向分離の例 (1)

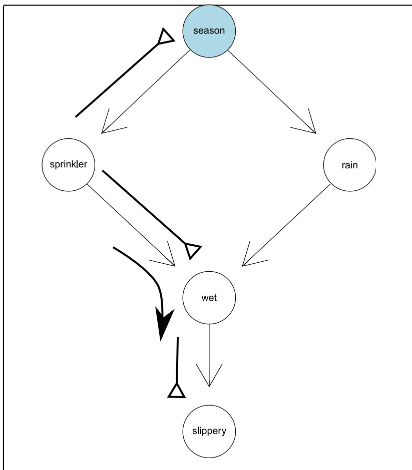
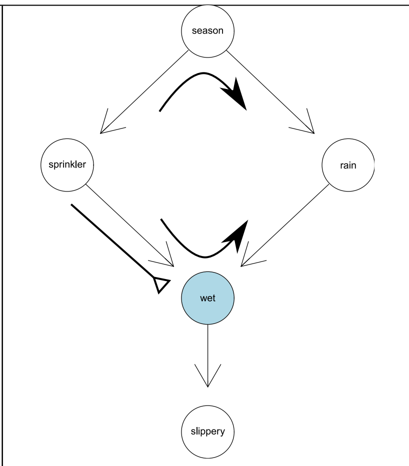
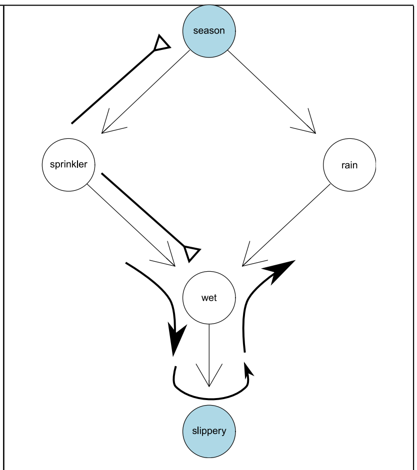
		
すべてのパスはブロックされる。season が観測されていれば、sprinkler と rain は独立。	有向分離されない。active なパスが二つある。wet が観測されていれば、sprinkler を知ることにより rain を知ることができる。	有向分離されない。slippery が観測されていることは実質的に wet が観測されていることに等しい。

表 C.3　有向分離の例 (2)

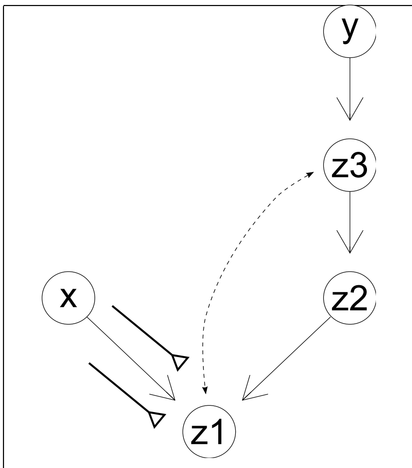
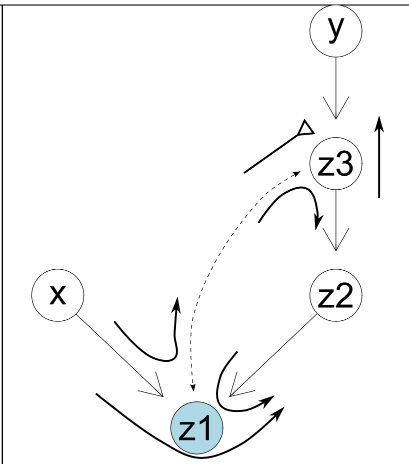
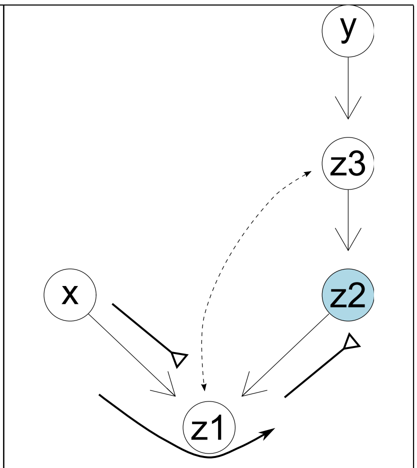
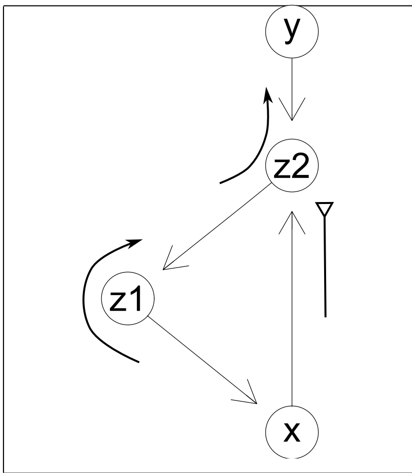
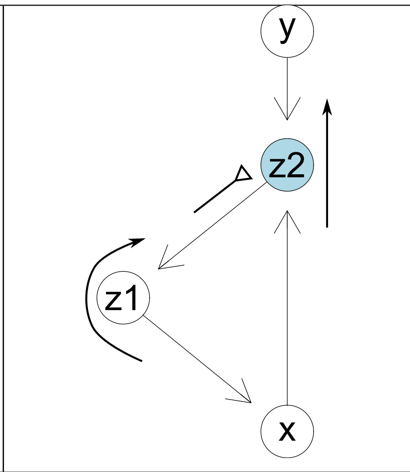
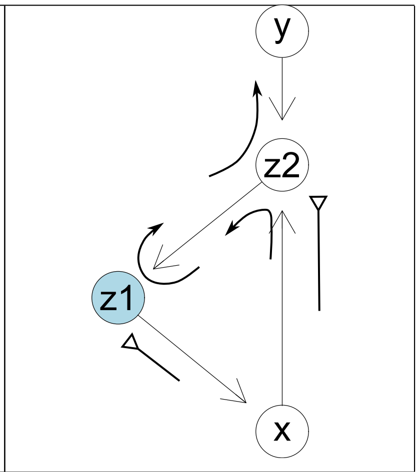
		
Z_i のいずれも観測されないとき (観測される変数は空集合)、X と Y は有向分離される。パスは二つともブロックされている。	Z_1 が観測されているばあい、X と Y は有向分離されない。active なパスが二つある。	Z_2 が観測されているばあい、X と Y は有向分離される。パスは二つともブロックされている。

表 C.4　有向分離の例 (3)

<table>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_i のいずれも観測されないとき (観測される変数は空集合)、X と Y は有向分離されない。$X \to Z_1 \to Z_2 \to Y$ のパスが active。</td>
<td>Z_2 が観測されているばあい、X と Y は有向分離されない。$X \to Z_2 \to Y$ のパスが active。</td>
<td>Z_1 が観測されているばあいも、X と Y は有向分離されない。やはり $X \to Z_2 \to Y$ のパスが active。</td>
</tr>
</table>

R.D.Shachter(2013) はこの困難を取り除くために、「Bayes Ball」algorithm なる工夫を提案した。これは、頂点集合 X と Y が有向分離されているか否かを、X と Y をむすぶ active なパスが存在するかどうかによって判別するというものである。active なパスがひとつもないばあいに X と Y は有向分離しているとし、ひとつでも active なパスがあれば有向分離されていないと判定する。ここで、active なパスとはその始点 X から終点 Y にかけて一度も「stop」に出会わないパスであり、「stop」があるかどうかは表 C.1 に示す 10 の規則 (これらは条件 i)、ii) の言い換えになっている) から判定する。

表 C.2〜表 C.4 は Pearl(2009) が示した事例を Bayes Ball ルールを用いてやや詳しく解説したものである。

表 C.2 では、季節 (season)、スプリンクラーの作動状態 (sprinkler)、降雨 (rain)、歩道が濡れているか (wet)、歩道が滑りやすいか (slippery) の 5 つの確率変数から構成される BN において、sprinkler と rain のあいだの条件付き独立関係を判定するために、sprinkler から rain に向って active なパスがあるかどうかを調べている。どの変数が観測されているかによって、その判定結果は異なっている。左では sprinkler $\perp\!\!\!\perp$ rain | season、中では sprinkler $\not\perp\!\!\!\perp$ rain | wet、右では sprinkler $\not\perp\!\!\!\perp$ rain | season,slippery である。

表 C.3 では、X と Y のあいだの条件付き独立が調べられている。Z_1 と Z_3 のあいだの両側矢線 (点線) によって、これらの変数について観測されない交絡因子があることが示されている。やはり観測される (条件付けられる) 変数によって判定結果は異なり、左では $X \perp\!\!\!\perp Y$、中では $X \not\perp\!\!\!\perp Y \mid Z_1$(active なパスが二つあることに注意)、右では $X \perp\!\!\!\perp Y \mid Z_2$ となる。

表 C.4 では、どの変数が観測されるかによらず、すべてのケースで (たとえ Z_1 と Z_2 の双方が観測されたとしても)X と Y は (条件付き) 独立とならない。

有向分離は条件付き独立性の判定のために使われるのであるから、この情報を用いて無関係な変数 (観測値) を除去することにより、同時確率、周辺確率をより簡単に表示することができる。たとえば式 (3.4) において、$(A \perp\!\!\!\perp B \mid C)_G$ ならば、$P(A \mid C, B) = P(A \mid C)$ のように観測値 B を除去できる。表 3.5 の規則 1(観測値の挿入/削除) は、この関係と実質的に同じである。

表 C.5　有向分離の使用例

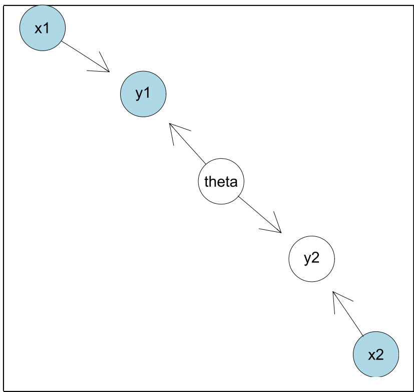
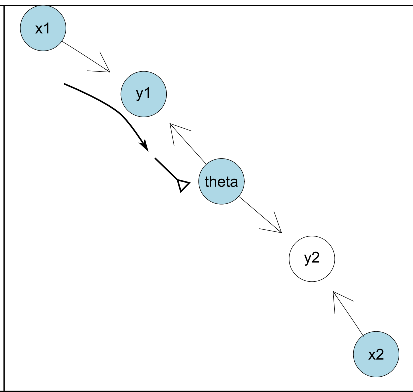
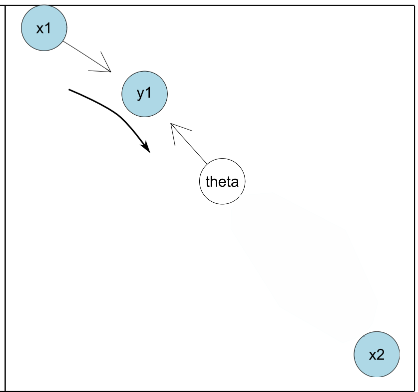
		
X_1, Y_1, X_2 が観測されており、Y_2, θ を推測したいものとする。$P(y_2, \theta \mid x_1, y_1, x_2) = P(y_2 \mid \theta, \underline{\cdots})P(\theta \mid \underline{\cdots})$	集合 $\{X_1, Y_1\}$ は θ により Y_2 と有向分離されている。$P(y_2 \mid \theta, \underline{\cdots}) = P(y_2 \mid \theta, x_2)$	X_2 はパスが存在しないために θ と有向分離されている。他方 $\{X_1, Y_1\}$ は θ と有向分離されない。$P(\theta \mid \underline{\cdots}) = P(\theta \mid x_1, y_1)$

表 C.5 に観測値の削除の例を示す[*24]。左では BN の全体像が示され、観測値/非観測値の別が明らかにされている (ここで $\underline{\cdots}$ は観測値の集合 x_1, y_1, x_2 を指す)。これが後の計算のために積の法則により分解される。

中では、分解された要素 $P(y_2 \mid \theta, \underline{\cdots})$ について、y_2 と無関係な観測値 (ここでは $\{x_1, y_1\}$) が調べられる。右ではもうひとつの要素 $P(\theta \mid \underline{\cdots})$ について同じことが行われる。

全体を合わせると結局次のようになる。

$$
\begin{aligned}
&P(y_2, \theta \mid x_1, y_1, x_2) \\
&= P(y_2 \mid \theta, x_2)P(\theta \mid x_1, y_1)
\end{aligned}
$$

ここで θ で周辺化することにより、最終的に次の結果を得る。

$$
\begin{aligned}
&P(y_2 \mid x_1, y_1, x_2) \\
&= \sum_{\theta} P(y_2 \mid \theta, x_2)P(\theta \mid x_1, y_1)P(\theta)
\end{aligned}
$$

[*24] ベイズ統計を使った機械学習の典型例では、θ は事前分布の役割を果たすものと考えられている。しかし確率変数 θ は主観的なものと考える必要はないことに注意すべきである。θ は (なんらかの事情により) 観測され得ないとされているにすぎず、客観的な確率変数であることには変わりがない。

C.7 AIC の導出

以下、北川・小西 (2006) にしたがって AIC の導出過程を整理する。

まずは対数尤度 $L_{r;1}$、$L_{0;1}$ のそれぞれを母数にかんする二次までのテイラー展開で近似する。そのために、$f_1(x)$ をパラメトリックモデルで表現する。これは様々な f_1 の候補を対象にするために、いくつかの未知量 (パラメータ) を置いて可変にすることである。これによって、f_1 は点ではなく、ある曲面を表すものになる。

$$f_1(\boldsymbol{x}|\boldsymbol{\theta}) \equiv g_\theta(\boldsymbol{x}) \equiv h_x(\boldsymbol{\theta})$$

次に $L_{0;1}$ が極大値をとるときのパラメータ $\hat{\boldsymbol{\theta}}_0$ を考える。このパラメータを代入したときの f_1 が f_0 に最も近い、ということが出来る。そのときの条件は以下のようになる。

$$\frac{\partial L_{0;1}(h_x(\boldsymbol{\theta}))}{\partial \theta_i} = 0 \tag{C.17}$$

ここで次の注意点に気がつく。

1). $L_{0;1}$ はパラメータ空間の上で滑らかに変化する必要がある。
2). 極大値をとったとしても、そのとき $f_0 \equiv f_1$ になっているとは限らない。あくまでも一番近い、ということに過ぎない。

次に、$L_{0;1}$ 関数の二次近似 (極大値のまわりの二次までのテイラー展開) を考える。ここで $\hat{\boldsymbol{\theta}}_0$ は式 (C.17) の解である[*25]。

$$L_{0;1}(h_x(\boldsymbol{\theta})) = L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_0)) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)^t \boldsymbol{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0) \tag{C.18}$$

ここで $\boldsymbol{H}$ は $L_{0;1}$ のヘッセ行列である。

$$\boldsymbol{H} = \left(\left. \frac{\partial^2 L_{0;1}(h_x(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j} \right|_{\theta=\hat{\theta}_0} \right)$$

次は、真の分布 f_0 ではなくて、経験分布 f_r から見た評価を考える。式 (C.19) は f_r から見た f_1 の平均対数尤度関数の二次近似である。$\hat{\boldsymbol{\theta}}_x$ は f_r に関する $L_{r;1}$ 関数の極値を示す。

$$L_{r;1}(h_x(\boldsymbol{\theta})) = L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x)) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_x)^t \boldsymbol{H}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_x) \tag{C.19}$$

式 (C.19) の $\boldsymbol{\theta}$ に $\boldsymbol{\theta}_0$ を代入、式 (C.18) の $\boldsymbol{\theta}$ に $\boldsymbol{\theta}_x$ をそれぞれ代入すると以下のようになる。

$$L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_x)) = L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_0)) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t \boldsymbol{H}(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0) \tag{C.20}$$

$$L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_0)) = L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x)) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_x)^t \boldsymbol{H}'(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_x) \tag{C.21}$$

次に、上の二つの式について、真の分布 f_0 を使って期待値を計算する。

式 (C.20) について、左辺は変化なし、右辺第一項も変化なしであるが、右辺第二項は以下のようになる。

$$\frac{1}{2}\mathrm{tr}\left[\boldsymbol{H} E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right]$$

次に式 (C.21) について、左辺は、次のように変化する。

$$L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_0)) \to L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_0))$$

右辺第一項は次のように変化する。

$$L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x)) \to E_0[L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x))]$$

右辺第二項は、

$$\frac{1}{2}\mathrm{tr}\left[\boldsymbol{H} E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right]$$

と、式 (C.20) の右辺第二項と同じになる。

整理すると、

$$\begin{aligned} &L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_x)) \\ &= L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_0)) \\ &\quad + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{H} E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right] \end{aligned} \tag{C.22}$$

$$\begin{aligned} &L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_0)) \\ &= E_0[L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x))] \\ &\quad + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{H} E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right] \end{aligned} \tag{C.23}$$

[*25] $\hat{\boldsymbol{\theta}}_0$ は極大値であるから、一次の項は式 (C.17) よりゼロになっていることに注意する。

式 (C.22) と式 (C.23) を辺々足すと、

$$L_{0;1}(h_x(\hat{\boldsymbol{\theta}}_x)) = E_0[L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x))] + \mathrm{tr}\left[\boldsymbol{H}E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right] \tag{C.24}$$

結果的に、経験分布から求めた平均対数尤度（いわゆる対数尤度）にはバイアスがある（大きく見積もり過ぎている）こと、そのバイアスの大きさは、

$$\mathrm{tr}\left[\boldsymbol{H}E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t\right] \tag{C.25}$$

であることが分かる。

次に、最尤推定量の性質（漸近分布のパラメータ）を用いて、式 (C.25) を変形する。

最尤推定量 $\boldsymbol{\theta}_x$ の従う分布はサンプル数が無限大のとき、正規分布に近づいていく。その平均は $\boldsymbol{\theta}_0$ になり、分散（の n 倍）は次のようなものになる。

$$E_0(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_x - \hat{\boldsymbol{\theta}}_0)^t \to \boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1} \tag{C.26}$$

ここで、$\boldsymbol{G}$ はフィッシャー情報量と呼ばれる次のような行列である。

$$\boldsymbol{G} = \left(\left.\frac{\partial L_{0;1}(h_x(\boldsymbol{\theta}))}{\partial \theta_i}\right|_{\theta=\hat{\theta}_0} \left.\frac{\partial L_{0;1}(h_x(\boldsymbol{\theta}))}{\partial \theta_j}\right|_{\theta=\hat{\theta}_0} \right)$$

式 (C.26) を式 (C.25) に代入すると、以下のようになる。

$$\mathrm{tr}\left[\boldsymbol{G}\boldsymbol{H}^{-1}\right] \tag{C.27}$$

一般には $\boldsymbol{G} \neq \boldsymbol{H}$ であるが、もし $\boldsymbol{G} = \boldsymbol{H}$ である場合、トレースの中は単位行列になる。

$$\mathrm{tr}\left[\boldsymbol{I}_p\right] = p$$

従って、バイアスは (自由) パラメータ数になる。これをあらかじめ対数尤度から差し引いておいたものが AIC と呼ばれるものである[*26]。

$$\mathrm{AIC} = -2L_{r;1}(h_x(\hat{\boldsymbol{\theta}}_x)) + 2p$$

C.8 Kolmogorov の偏微分方程式

C.8.1 前進方程式の導出

$$A = \int_R f(y) \frac{\partial Q(t, y \mid x)}{\partial t} dy$$

ただし、次の取り決めをする。

$$\begin{aligned}
Q(t, y \mid x) &\equiv Q(t, y \mid 0, x) \\
Q(t + \Delta t, y \mid t, z) &\equiv Q(\Delta t, y \mid 0, z) \quad \text{(時間の同質性)} \\
\frac{\partial Q(t, y \mid x)}{\partial t} &\equiv \lim_{\Delta t \to 0} \frac{1}{\Delta t} \{Q(t + \Delta t, y \mid x) \\
&\qquad - Q(t, y \mid x)\}
\end{aligned}$$

$$\begin{aligned}
A &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ \int_R f(y) Q(t + \Delta t, y \mid x) dy \right. \\
&\qquad \left. - \int_R f(y) Q(t, y \mid x) dy \right\} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \{B - C\}
\end{aligned}$$

B についてチャップマン=コルモゴロフの等式を適用する。また時間の同質性を適用する。

$$\begin{aligned}
& Q(t + \Delta t, y \mid x) \\
&= \int_R Q(\Delta t, y \mid z) Q(t, z \mid x) dz \\
B &= \int_R f(y) \int_R Q(\Delta t, y \mid z) Q(t, z \mid x) dz dy \\
& \downarrow \text{積分入れ替え} \\
&= \int_R \int_R f(y) Q(\Delta t, y \mid z) Q(t, z \mid x) dy dz \\
&= \int_R Q(t, z \mid x) \int_R f(y) Q(\Delta t, y \mid z) dy dz
\end{aligned}$$

C について、y を z に変更する。また、次の正規性条件を使う。

$$1 = \int_R Q(\Delta t, y \mid z) dy$$

$$C = \int_R f(z) Q(t, z \mid x) dz$$

[*26] もし $\boldsymbol{G} \neq \boldsymbol{H}$ ならば、$\boldsymbol{G}$,$\boldsymbol{H}$ の計算で真の分布 f_0 を経験分布 f_r で置き換えて式 (C.27) を計算する。これは竹内の情報量規準 TIC と呼ばれる。また真の分布が分かっていると仮定して、f_0 を明示して式 (C.25) を直接計算することも考えられるが、これは「有限修正」と呼ばれる。小西・北川 (2004) を参照のこと。また、高嶋 (2014) はこれをグラフィカル・モデルについて計算している。

$$= \int_R f(z)Q(t,z \mid x) \int_R Q(\Delta t, y \mid z) dy dz$$
$$= \int_R Q(t,z \mid x) \int_R f(z)Q(\Delta t, y \mid z) dy dz$$

ここまでの結果を合わせて、

$$\begin{aligned} &B - C \\ &= \int_R Q(t,z \mid x) \int_R \{f(y) - f(z)\} Q(\Delta t, y \mid z) dy dz \\ &= \int_R Q(t,z \mid x) G dz \end{aligned}$$

G について、$f(y)$ を z のまわりでテイラー展開する。

$$G = \int_R Q(\Delta t, y \mid z) \sum_{n=1}^{\infty} f^{(n)}(z) \frac{(y-z)^n}{n!} dy$$

次のジャンプ・モーメント $K^{(n)}(z)$ を定義する。

$$K^{(n)}(z) = \frac{1}{n!} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_R (y-z)^n Q(\Delta t, y \mid z) dy$$

よって[*27]、

$$\begin{aligned} A &= \int_R Q(t,z \mid x) \sum_{n=1}^{\infty} K^{(n)}(z) f^{(n)}(z) dz \\ &= \int_R f(z) \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial z}\right)^n \left[K^{(n)}(z) Q(t,z \mid x)\right] \end{aligned}$$

A について y を z に変更する。

$$\begin{aligned} 0 = \int_R f(z) \Bigg\{ &\frac{\partial Q(t,z \mid x)}{\partial t} \\ &- \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial z}\right)^n \left[K^{(n)}(z) Q(t,z \mid x)\right] \Bigg\} dz \end{aligned}$$

$f(z)$ は任意なので、

$$\frac{\partial Q(t,z \mid x)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial z}\right)^n \left[K^{(n)}(z) Q(t,z \mid x)\right]$$

初期条件として $t=0$ で x が退化分布を仮定する。z を y に変更する $(Q(t,z \mid x) \to Q(y,t))$[*28]。

$$\frac{\partial Q(t,y)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial y}\right)^n \left[K^{(n)}(y) Q(t,y)\right]$$

$K^{(n)}(x) = 0, n > 2$ を仮定する。

$$\frac{\partial Q(t,y)}{\partial t} = -\frac{\partial K^{(1)}(y) Q(t,y)}{\partial y} + \frac{\partial^2 K^{(2)}(y) Q(t,y)}{\partial y^2}$$

これは Kolmogorov の前進方程式[*29](forward equation) と呼ばれる。

C.8.2 後退方程式の導出

以下を取り決める。

$$\begin{aligned} u(s,x) &\equiv \int_R Q(y \mid s,x) f(y) dy \\ Q(y \mid s,x) &\equiv Q(0,y \mid s,x) \\ -\frac{\partial}{\partial s} u(s,x) &\equiv \lim_{\Delta s \to 0} \frac{1}{\Delta s} \{u(s-\Delta s,x) - u(s,x)\} \end{aligned}$$

チャップマン＝コルモゴロフの式より、

$$\begin{aligned} &Q(y \mid s - \Delta s, x) \\ &= \int_R Q(y \mid s,z) Q(z \mid -\Delta s, x) dz \\ &u(s - \Delta s, x) \\ &= \int_R f(y) \int_R Q(y \mid s,z) Q(z \mid -\Delta s, x) dz dy \\ &= \int_R Q(z \mid -\Delta s, x) \int_R f(y) Q(y \mid s,z) dy dz \\ &= \int_R Q(z \mid -\Delta s, x) u(s,z) dz \\ &\downarrow z \text{ を } y \text{ に変更} \\ &= \int_R Q(y \mid -\Delta s, x) u(s,y) dy \end{aligned}$$

また、

$$\begin{aligned} u(s,x) &= u(s,x) \int_R Q(y \mid -\Delta s, x) dy \\ &= \int_R Q(y \mid -\Delta s, x) u(s,x) dy \end{aligned}$$

合わせて、

*27 偏微分演算子についてエルミート共役をとっている。

*28 この式は、Kramers=Moyal expansion とも呼ばれる。

*29 フォッカー=プランク方程式とも呼ばれる。

$$
\begin{aligned}
&u(s-\Delta s,x)-u(s,x)\\
&=\int_R Q(y\mid -\Delta s,x)\left\{u(s,y)-u(s,x)\right\}dy\\
&u(s,y)-u(s,x)\\
&=\sum_{n=1}^{\infty}u^{(n)}(x)\frac{(y-x)^n}{n!}
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\partial}{\partial s}u(s,x)\\
&=\sum_{n=1}^{\infty}u^{(n)}(x)\lim_{\Delta s\to 0}\frac{1}{\Delta s}\int_R\frac{(y-x)^n}{n!}Q(y\mid -\Delta s,x)dy\\
&=\sum_{n=1}^{\infty}K^{(n)}(x)\cdot\left(\frac{\partial}{\partial x}\right)^{(n)}u(s,x)\\
&=\int_R f(y)\left\{\sum_{n=1}^{\infty}K^{(n)}(x)\cdot\left(\frac{\partial}{\partial x}\right)^{(n)}Q(y\mid s,x)\right\}dy
\end{aligned}
$$

よって、

$$
-\frac{\partial}{\partial s}Q(y\mid s,x)=\sum_{n=1}^{\infty}K^{(n)}(x)\cdot\left(\frac{\partial}{\partial x}\right)^{(n)}Q(y\mid s,x)
$$

$K^{(n)}(x)=0, n>2$ を仮定する。やはり、y が退化分布に従うとする ($Q(y\mid s,x)\to Q(x,s)$)。

$$
-\frac{\partial Q(s,x)}{\partial s}=K^{(1)}(x)\frac{\partial Q(s,x)}{\partial x}+K^{(2)}(x)\frac{\partial^2 Q(s,x)}{\partial x^2}
$$

これは Kolmogorov の後退方程式 (backward equation) と呼ばれる。

C.9　シュレディンガー方程式の解

C.9.1　自由粒子の定常解

$\psi(t,x)=\Psi(x)\varphi(t)$ として時間変数を分離する。代入して両辺を ψ で割ることにより、

$$
\frac{1}{\varphi}\left(i\frac{\partial\varphi}{\partial t}\right)=-\frac{1}{\Psi}\left(\frac{1}{2}\sigma^2\nabla^2\Psi-V\Psi\right)=W
$$

W は t、x のいずれにも依存しない定数である。よって、

$$
\begin{cases}
i\dfrac{\partial\varphi}{\partial t}=W\varphi\\[2ex]
-\dfrac{1}{2}\sigma^2\nabla^2\Psi+V\Psi=W\Psi
\end{cases}
$$

第一式を解くことにより、$\varphi(t)=e^{-iWt}$ となる。W は運動エネルギーを表現するが、事例 (1) とは異なり離散的量子数でなく連続量となる。

第二式において外力をゼロとする ($V=0$) ことにより、また $\Psi=X(x)Y(y)Z(z)$ として、

$$
\frac{1}{2}\sigma^2\left(\frac{1}{X}\frac{\partial^2 X}{\partial x^2}+\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2}+\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2}\right)+W=0
$$

$(1/2)\sigma^2(1/X)(\partial^2X/\partial x^2)=-W_x$ とすれば、

$$
\frac{1}{2}\sigma^2\frac{\partial^2 X}{\partial x^2}+W_xX=0
$$

これを解いて

$$
X(x)=\beta_x\exp\left\{i\frac{\sqrt{2W_x}}{\sigma}(x-x_0)\right\}
$$

$W_x+W_y+W_z=W$ であり、$W_x,W_y,W_z\geq 0$ である。仮に $W_y=W_z=0, W_x=W$ とすれば次の一般解を得る[*30]。

$$
\psi(t,x)=\beta\exp\left\{i\frac{\sqrt{2W}}{\sigma}(x-x_0)\right\}e^{-iWt}
$$

ここで $\sqrt{2W}/\sigma=\kappa$、$W=\sigma^2\kappa^2/2$ とする。一般に三次元の解は次のようになる。

$$
\psi(t,\boldsymbol{x})=\beta\exp\left\{i\left(\boldsymbol{\kappa}\cdot\boldsymbol{x}-\frac{1}{2}\sigma^2|\boldsymbol{\kappa}|^2t\right)\right\}
$$

C.9.2　自由粒子の非定常解

$$
T(0,x)=\frac{1}{i}\frac{\partial}{\partial x}\ln\psi(0,x)=\frac{1}{i}\left(-\frac{x}{a^2}+i\kappa\right)=T_0
$$

ここで $T(t,x)=T_0A_t, A_0=1$ とする。

$$
T_0A_t'=-\frac{i\sigma^2}{a^2}T_0A_t^2
$$

[*30] これは空間上を前進する解である。逆に後退する解 $\psi(t,x)=\beta\exp\left\{-i\frac{\sqrt{2W}}{\sigma}(x-x_0)-iWt\right\}$ も同様にシュレディンガー方程式を満たす。

$$\downarrow$$

$$A_t = \left(1 + \frac{i\sigma^2}{a^2}t\right)^{-1}$$

$$\downarrow$$

$$T(t,x) = \frac{A_t}{i}\left(-\frac{x}{a^2} + i\kappa\right)$$

これより、

$$\frac{\partial \mathcal{S}}{\partial x} = iT = A_t\left(-\frac{x}{a^2} + i\kappa\right)$$

$$\downarrow$$

$$\mathcal{S} = A_t\left(-\frac{x^2}{2a^2} + i\kappa x\right) + g(t) = \ln\psi - \ln\beta$$

$$\downarrow \text{ ただし } g(0) = 0$$

$$\begin{aligned} i\frac{\partial}{\partial t}\ln\psi &= ig' + i\left(-\frac{x^2}{2a^2} + i\kappa x\right)A_t' \\ &= ig' + \frac{\sigma^2}{a^2}A_t^2\left(-\frac{x^2}{2a^2} + i\kappa x\right) \end{aligned}$$

$$\begin{aligned} i\frac{\partial}{\partial t}\ln\psi &= -\frac{1}{2}\sigma^2\left\{\left(\frac{\partial \mathcal{S}}{\partial x}\right)^2 + \frac{\partial^2 \mathcal{S}}{\partial x^2}\right\} \\ &= -\frac{1}{2}\sigma^2\left\{A_t^2\left(-\frac{x}{a^2} + i\kappa\right)^2 - \frac{A_t}{a^2}\right\} \end{aligned}$$

ゆえに、

$$ig' = \frac{1}{2}\sigma^2\kappa^2A_t^2 + \frac{1}{2a^2}\sigma^2A_t$$

$$\downarrow$$

$$\begin{aligned} g &= -\frac{1}{2}a^2\kappa^2(1 - A_t) + \frac{1}{2}\ln A_t \\ &= -\frac{i}{2}A_t\sigma^2\kappa^2 t + \frac{1}{2}\ln A_t \end{aligned}$$

$$\downarrow$$

$$\ln\psi = \ln\beta + \frac{1}{2}\ln A_t + A_t\left(-\frac{x^2}{2a^2} + i\kappa x - \frac{i}{2}\sigma^2\kappa^2 t\right)$$

付録 D

R による計算例

D.1　対数線型モデルの AIC 計算例

```
> summary(RSdat)
 p      s      t      u           freq
 M:12   H:8    H:12   N:12   Min.   :19.00
 X:12   M:8    L:12   Y:12   1st Qu.:29.00
        S:8                  Median :42.50
                             Mean   :42.00
                             3rd Qu.:52.25
                             Max.   :68.00
> library(MASS)
> res19 <- loglm(freq ~ t + u + s + t*u + t*s
+ u*s + t*u*s + p , data=RSdat)
> aice1 <- res19$lrt - 2 * res19$df
>
> library(igraph)
> library(gRbase)
> library(gRim)
> library(devtools)
> library(Rgraphviz)
>
> dm1 <- xtabs(freq ~ ., data=RSdat)
> gres19 <- dmod(~ t + u + s + t*u + t*s + u*s
+ t*u*s + p , data=dm1,fit=TRUE)
> aice2 <- gres19$fitinfo$dev - 2
* as.numeric(gres19$fitinfo$dimension[4])
>
```

D.2　EM アルゴリズムの計算例 (1)

```
> y1 <- rnorm(70)*5 + 0
> x1 <- cbind(y1,rep(1,70))
> y2 <- rnorm(30)*3 + 20
> x2 <- cbind(y2,rep(2,30))
> x <- rbind(x1,x2)
> x <- x[sample(100),]
> hist(x[,1],freq=F,main="",xlab="y")
>
> la <- c(0.5,0.5)
> yu <- sort(x[,1])
> mu <- c(mean(yu[1:50]),mean(yu[51:100]))
> sigma  <- c(sd(yu[1:50]),sd(yu[51:100]))
>
> w <- c(0,la,mu,sigma,0)
> pz <- estepfunc(x[,1],w)
> oldlc <- lcfunc(x[,1],pz,w)
> w <- c(0,la,mu,sigma,oldlc)
> res <- w
>
> for(i in 0:100){
>  pz <- estepfunc(x[,1],w)
>  w <- c(i+1,mstepfunc(x[,1],pz),lc)
>  lc <- lcfunc(x[,1],pz,w)
>  res <- rbind(res,w)
>  if((lc-oldlc)^2 < 10^(-4)) break
>  oldlc <- lc
> }
```

```
>
> # LogLikelihood
> lcfunc <- function(y,pz,w){
>  k <- (length(w)-1)/3
>  lcz <- c()
>  for (y0 in y) {
>   lc0 <- rep(0,k)
>   for(i in 1:k){
>    lc0[i] <- log(w[1+i])
+log(dnorm(y0,w[1+k+i],w[1+2*k+i]))
>   }
>   lcz <- rbind(lcz,lc0)
>  }
>  return(sum(pz * lcz))
> }
>
> # E-step
> estepfunc <- function(y,w){
>  retval <- c()
>  k <- (length(w)-1)/3
>  for (y0 in y) {
>   z0 <- rep(0,k)
>   zz <- 0
>   for(i in 1:k){
>    z0[i] <- w[1+i]*dnorm(y0,w[1+k+i],
     w[1+2*k+i])
>    zz <- zz + z0[i]
>   }
>   z0 <- z0 / zz
>   retval <- rbind(retval,z0)
>  }
>  return(retval)
> }
>
> # M-step
> mstepfunc <- function(y,pz){
>  n <- length(y)
>  n0 <- colSums(pz)
>  mu0 <- colSums(pz*y)/n0
>  sigma0 <- sqrt(colSums(pz*(y^2))/n0
     - (mu0^2))
>  la0 <- n0/n
>  return(c(la0,mu0,sigma0))
> }
```

D.3　EM アルゴリズムの計算例 (2)

```
> # sakamotodat
>
> sakamotodat <- read.csv("sakamoto.csv",
  header = TRUE,na.strings="NA")
>
> ym <- 1/(1+sakamotodat[,3]/sakamotodat[,2])
>
> # param. initialize
> a <- numeric(75)
> sigma2 <- 1
> z0 <- numeric(2)
> C <- 75
>
> t0 <- a
> t0[1:3] <- c(1,-2,1)
> D <- toeplitz(t0)
> D[upper.tri(D)] <- 0
>
> # r,b,Lamb
>
> r <- a
> r[1] <- z0[1] - 2 * z0[2]
> r[2] <- z0[2]
>
> b <- sakamotodat[,"Yes"] - sakamotodat[,"No"]
> b <- b/2
>
> Lamb <- diag((a + 1)/8 *
  (sakamotodat[,"Yes"] + sakamotodat[,"No"]))
>
> # DNinv muN
>
> SNinv <- (t(D) %*% D)/sigma2 + Lamb * 2
> SN <- solve(SNinv)
```

```
> muN <- SN %*% (b - (t(D) %*% r) / sigma2)
>
> l0 <- 1/(1+exp(-muN))
>
> # param. renewal
> anew <- diag(SN + muN %*% t(muN))
> rnew <- - D %*% muN
> p2 <- rnew[-(1:2)]
> sigma2new <- (t(p2) %*% p2
  + sum(diag(t(D) %*% D %*% SN)))/C
> rnew[-(1:2)] <- 0
>
> # 2nd itteration
> a <- anew
> r <- rnew
> sigma2 <- as.numeric(sigma2new)
>
> Lamb <- diag((a + 1)/8
  * (sakamotodat[,"Yes"] + sakamotodat[,"No"]))
> SNinv <- (t(D) %*% D)/sigma2 + Lamb * 2
> SN <- solve(SNinv)
> muN <- SN %*% (b - (t(D) %*% r) / sigma2)
>
> l0 <- cbind(l0,1/(1+exp(-muN)))
>
> # itteration using 'til sigma2 converge
>
> sigma2old <- sigma2
>
> for (i in 1:500) {
>  anew <- diag(SN + muN %*% t(muN))
>  rnew <- - D %*% muN
>  p2 <- rnew[-(1:2)]
>  sigma2new <- (t(p2) %*% p2
   + sum(diag(t(D) %*% D %*% SN)))/C
>  rnew[-(1:2)] <- 0
>
>  a <- anew
>  r <- rnew
>  sigma2 <- as.numeric(sigma2new)
>
>  Lamb <- diag((a + 1)/8
   * (sakamotodat[,"Yes"] + sakamotodat[,"No"]))
>  SNinv <- (t(D) %*% D)/sigma2 + Lamb * 2
>  SN <- solve(SNinv)
>  muN <- SN %*% (b - (t(D) %*% r) / sigma2)
>
>  if((sigma2 - sigma2old)^2 < 10^{-10}) break
>  sigma2old <- sigma2
> }
>
> l0 <- cbind(l0,1/(1+exp(-muN)))
>
> (sigma2 - sigma2old)^2
> sakamotodat[which.max(l0[,3]),"Age"]
> max(l0[,3])
```

D.4 VARMOD プログラム

```
># ----------------------------------------------
># VARMOD : AIC for Variance Analysis Models
># ----------------------------------------------
>varmod <- function(df0,a=0,b=0){
> if((a>b && b!=0) || min(a,b)<0 || max(a,b) >
  ncol(df0)-1) {return(NULL)}
> m0 <- mean(df0[,1])
> n0 <- nrow(df0)
> v0 <- sum(df0[,1]^2)
> w0 <- v0 - m0*sum(df0[,1])
> w1 <- w2 <- 0
> s0 <- (w0 - w1 - w2)/n0
> npar <- 2
> aic <- n0 * log(2*pi)+n0 * log(s0) + n0
  + 2*npar
> if(a == 0){
>  modelstr <- "i(0)"
>  return (list(model=modelstr,npar=npar,s=s0,
    aic=aic))
> }
>
> m1 <- c()
```

```
> y1 <- c()
> nc1 <- max(df0[,a+1])
> for (i in 1:nc1){
>  m1 <- c(m1,mean(df0[df0[,a+1]==i,1])-m0)
>  y1 <- c(y1,sum(df0[df0[,a+1]==i,1]))
> }
> w1 <- sum(m1 * y1)
> s0 <- (w0 - w1 - w2)/n0
> npar <- npar + nc1-1
> aic <- n0 * log(2*pi)+n0 * log(s0) + n0
  + 2*npar
> if(b == 0){
>  modelstr <- paste0("i(",a,")")
>  return (list(model=modelstr,npar=npar,s=s0,
    aic=aic))
> }
>
> m2 <- c()
> y2 <- c()
> nc2 <- max(df0[,b+1])
> for (i in 1:nc2){
>  m2 <- c(m2,mean(df0[df0[,b+1]==i,1])-m0)
>  y2 <- c(y2,sum(df0[df0[,b+1]==i,1]))
> }
> w2 <- sum(m2 * y2)
> s0 <- (w0 - w1 - w2)/n0
> npar <- npar + nc2-1
> aic <- n0 * log(2*pi)+n0 * log(s0) + n0
  + 2*npar
>
> modelstr <- paste0("i(",a,",",b,")")
> return (list(model=modelstr,npar=npar,s=s0,
  aic=aic))
>}
>
>y<-c(268,233,254,281,240,249,231,314,256,
 250,280,291,265,250,248,271)
>x1<-c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)
>x2<-c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4)
>sk2 <- data.frame(y,x1,x2)
>
>res <- unlist(varmod(sk2))
>res <- rbind(res,unlist(varmod(sk2,1)))
>res <- rbind(res,unlist(varmod(sk2,2)))
>res <- rbind(res,unlist(varmod(sk2,1,2)))
>res[order(res[,4]),]
```

D.5　長澤理論のシミュレーション

```
># --------------------------------------------
># Fig5.3 : 2-D Harmonic Oscilator
># --------------------------------------------
>
> library(mvtnorm)
> n = 10000
> sigma = matrix(c(0.01, 0, 0, 0.01), ncol = 2)
> rand = rmvnorm(n, mean = c(0, 0), sigma)
> dr0 = rand[, 1]
> ds0 = rand[, 2]
> r0 <- numeric(n)
> s0 <- numeric(n)
> sig <- 0.03
> r0[1] <- 1
> for(i in 2:n){
>  r0[i] <- r0[i-1] + (sig^2)/r0[i-1]
    - (sig^2)*r0[i-1] + sig*dr0[i-1]
>  s0[i] <- s0[i-1] + (sig^2)/r0[i-1]
    + sig*ds0[i-1]
> }
> plot(r0*sin(s0),r0*cos(s0),type="l",xlab="",
   ylab="")
>
># --------------------------------------------
># Fig5.4 : Mu of Free Particles
># --------------------------------------------
>
> mut <- function(t,y){
>  x0 <- 1
>  s <- 1
>  b2 <- 1
>  A <- b2 / sqrt(1+t^2)
```

```
>  C <- 1/(s^2)/(1+t^2)
>  D <- ((x0)^2) * ((1-t)^2) + y^2
>  retval <- A * exp(-C*D)
>  return(retval)
> }
>
> curve(mut(-1,x),-5,5,ylim=c(0,0.7),lty=1,
    ann=FALSE,axes = FALSE)
> par(new=TRUE)
> curve(mut(0,x),-5,5,ylim=c(0,0.7),lty=2,
    ann=FALSE,axes = FALSE)
> par(new=TRUE)
> curve(mut(1,x),-5,5,ylim=c(0,0.7),lty=3,
    ann=FALSE,axes = FALSE)
> par(new=TRUE)
> curve(mut(1.5,x),-5,5,ylim=c(0,0.7),type="p",
    ann=FALSE,axes = FALSE)
> par(new=TRUE)
> curve(mut(2.5,x),-5,5,xlab="y",ylab="mu(t)",
    ylim=c(0,0.7),type="p",pch=0)
> labels <- c("t=-1","t=0","t=1","t=1.5",
    "t=2.5")
> pchs <- c(NA_integer_,NA_integer_,
    NA_integer_,1,0)
> ltys <- c(1,2,3,0,0)
> legend("topleft", legend = labels,
    pch = pchs, lty = ltys,bty="n")
>
># ------------------------------------------
># Fig5.5 : Double Slit Experiment
># ------------------------------------------
>
> mut2 <- function(t,y){
>  A2 <- 1 / (1+t^2)
>  y0 <- 10
>  spm <- -A2 * t /2 * y0 * y
>  r <- A2 * y0 *y
>  R0 <- 0.5*log(A2) - A2 /2
     * ((1-t)^2 + y0^2 + y^2)
>  retval <- exp(R0)*(0.5*exp(r)
     +0.5*exp(-r)+cos(spm))
>  return(retval)
> }
> sgr_t <- seq(1,20,by=1)
> sgr_y <- seq(-100,100,by=.1)
> syd <- c()
> for(t0 in sgr_t){
>  for(d0 in sgr_y){
>   syd <- c(syd,mut2(t0,d0))
>  }
> }
> sz <- matrix(syd,ncol=length(sgr_t))
>
> persp(sgr_y,sgr_t,sz, theta = 140, phi = 30,
    zlim=c(0,0.1),ticktype="detailed",lwd=0.1,
    ltheta = -120, shade = 0.4,col="lightgrey",
    border=NA,xlab="y",ylab="t",zlab="mu")
>
```

Human Freedom and the Probability Theory

– An Essay on the Philosophy of Probability in Productive Practices –

Yuichi Takashima 13 Apr 2020

There are a number of problems with the philosophy of probability that come to mind as follows. 1) Why are there tremendous debates between subjectivism and objectivism in the viewpoint (the philosophy) of the plobability theory? 2) Why is the origin of probability theory attributed to the issue of a betting problem, which was argued by Pascal? Why did the theory of probability not develop before then? 3) Why did Keynes start his academic life with the research for his original probability theory? 4) What meanings does the recent popularity of Bayesian statistics have? 5) What is the relationship between probability theory and information theory? 6) How Marxists have treated the probability theory in their philosophy? and so on.

To answer these questions, it is essential for us a) to treat the history of the theory of probability as a matter of criticism of ideology, and also b) to reconstruct the ideology of probability itself from the perspective of proletariat.

This paper is a brief sketch of the latter, b) "philosophy of probability in productive practice." For that end, we see "the theory of technology" by Taketani(1968) as the clue to grasp the internal logic of productive practice (chapter 2). Subsequent chapters 3 and 4 describe the ontological and epistemological aspects of probability, respectively. Chapter 5 also belongs to the ontological description of probability, but discusses the meaning of measure-based probability theory and the objectivism of probability, focusing on the topics in quantum physics. Finally, based on the above results, we present some issues concerning a) the "criticism of ideology on probability theory" (chapter 6). Chapter 7 gives answers to the problems listed at the beginning, and describes further research topics that should be considered in the future.

The strongest argument in this paper is that the dual nature of probabilities (the debates between objectivism and subjectivism) does not simply mean that one defeats the other, but should be seen as two azimuth base points which promise an unified understanding in higher level. It also adds that this elucidation is directly related to more practical issues; what is freedom ("creation of purposes").

keywods : the philosophy of probability, historical materialism, productive practice, the theory of technology