

ビッグデータは「理論の終焉」をもたらすのか

佐賀 卓雄

「本来、理論家の任務は現実と一挙に融合するのではなく、一定の価値基準に照らして複雑多様な現実を方法的に整理するところにあり、従って整理された認識はいかに完璧なものでも無限に複雑多様な現実をすっぱり包みこむものでもなければ、いわんや現実の代用をするものではない。それはいわば、理論家みずからの責任において、現実から、いや現実の微細な一部から意識的にもぎとられてきたものである。従って、理論家の眼は、一方厳密な抽象の操作に注がれながら、他方自己の対象の外辺に無限の曠野をなし、その崖は薄明の中に消えてゆく現実に対するある断念と、操作の過程からこぼれ落ちてゆく素材に対するい

とおしみがそこに絶えず伴っている。」（丸山（一九六一）、六六一―六七頁、傍点は原著者）

「およそ天体の物理学であれ地上の物理学であれ…、その改革をもたらしたものは、新しい観測とか、新事実の発見とかではなく、科学者の精神の内部に起こった意識の変化なのであった。（中略）

…きわめて困難なこと、それは、従来と同じ一連のデータを用いながら、しかもそれらに別の枠組を当てはめて相互の関係を新しい体系に組みかえることであると言えよう。それはつまり、いわば新しい思考の帽子をかぶって今までとはまるつきり違った見方をして

みることである。」(バターフィールド(一九七八)、(上)、二〇頁)

「データを…与えるのは人間の認識活動である。そして、認識は、決して、客観的世界を受け取る、という行為ではなく、むしろ、いわゆる「客観的世界」なるものを造り上げるデータを、自らの手で刻みとり、選びとる行為である。そして、その場合、刻みとり選びとるための人間の概念上の道具こそ、理論である。：われわれ人間の認識は、なんらかの理論的概念枠なしには行われ得ない」(村上(一九八六)、五二―五三頁)

一

『ワイアード』誌編集長のクリス・アンダーソンは、二〇〇八年六月に「理論の終焉…データの洪水が科学的手法を時代遅れにする」という記事

を纏めている。科学者は因果関係と相関関係を厳密に峻別し、モデルによって原因と結果を結び付け、モデルのないデータはノイズに過ぎないとみなしてきたが、アンダーソンはこのアプローチは今や時代遅れであるとする。ペタバイトの情報は「相関は十分であり、モデルに頼らずとも、データを分析できる」し、「相関は因果に取って代わり、科学は守備一貫したモデルや統一的な理論、あるいは機械的な説明がまったくなくても進歩することができる」と主張している。

現実に、グーグルやアマゾンなどのプラットフォームは、膨大な顧客情報に基づいて事業戦略を構築している⁽¹⁾。また、小売企業の売上げ増加につながる顧客接点のあり方についても、ビッグデータを利用した戦略が模索されている⁽²⁾。

確かに、アマゾンは顧客とサイトとの通信記録を徹底的に分析することによって、顧客への商品

の推奨のターゲットを絞り、販売促進に役立てている。また、フィンテック・ベンチャーでは、ビッグデータの活用により伝統的な信用情報（FICOスコア）では把握しきれない情報を発見し、融資対象の拡大に結び付けている。また、シグファイグ（SigFig）は投資情報の透明性を高め、中立的な投資分析と現在の財務状況に対する評価を提供する見返りに、投資家のすべての保有口座の読み取り専用アクセスを求め、業務を拡充している。³⁾

しかし、事業分野でのこうしたビッグデータの活用による相関関係の偏重は、本当に仮説に基づく因果関係の追究を時代遅れにするのであろうか。そもそも相関関係とは複数の事象の関連を示すものであり、統計的には一からマイナス一までの相関係数によりその程度が示されるが、それらの事象の因果関係については何も語らない。どち

らが原因でどちらが結果であるか、またそれ以外の要因の影響についてはデータは何も語らないのである。

例えば、矢野（二〇一四）は、顧客と従業員にウェアラブルセンサを装着させ、両者の接触時間の長さにより売り上げに有意な影響が生じることが明らかになった。しかし、それがどのような理由から生じているかについては分からないのである。何よりも業績が優先される経営の現場では取りあえずはそれで納得するとしても、やはり何故そのような結果になったかについて知りたいし、それによつて将来生じうる様々な課題に対して取り組む手掛かりを得ることができるとであろう。

矢野はコンピュータが人間の気が付かない所で自ら仮説を立てると述べているが、そのプロセスについてはブラック・ボックスであり、人間の推論に頼らざるをえないであろう。その意味では、

やはり仮説は重要である。

一一

因果関係と相関関係の区別についてもう少し取り上げよう。⁽⁴⁾ 例えば、しばしば紹介される例として、チヨコレートの消費量とノーベル賞の授賞者数の関連がある。これは医学の専門雑誌に掲載されたもので、チヨコレートに含まれるフラボノールが認知機能を高めることに注目し、各国のデータから国民一人当たり年間四〇〇グラム多くチヨコレートを摂取するとノーベル賞受賞者の数が一人数増えると結論づけたのである。この論文に掲載されている図をみると、ヨーロッパで人口一人当たりのGDPの高い国ほどチヨコレートの消費量が多く、ノーベル賞受賞者数も多い。ただし、両者が因果関係なのか、相関関係なのかについては

言及がない。チヨコレートは贅沢品なので、所得の高い北欧諸国の消費量が多く、それらの国では教育にお金をかける余裕もあると考えられる。そのような推論ができるとすれば、この両者を結ぶ要因は他にも考えられる可能性がある（「交絡因子」の存在）。従って、この関係は因果関係の証明としては不十分で、チヨコレートの消費量の多寡が直接の原因であるかは不明である。

次に、近年関心の高まっているジェンダー多様性と企業業績の関連についてみてみよう。ノルウェーの女性取締役と企業価値の関係についての研究では、その比率の上昇が企業価値を低下させることが示されたという。このような結果になった理由として、調査期間に新しく取締役に就任した女性は、既存の取締役よりも年齢が若く、経験不足であったこと、また既存の取締役の縁故が多かったということが理由として考えられるこ

とが指摘された。

社外取締役と企業業績との関連についても同様のことが指摘されている。アメリカの研究では、取締役会の独立性とその企業の業績とは統計的にはなんの関係もないということが示されている。日本における近年の議論でも、元々、業績好調の企業が企業イメージ向上を目的として独立取締役を増やす可能性も考えられ、両者の因果関係は不明である。

また、アメリカで二〇〇一年一二月に破綻したエンロンは、同年三月の時点で取締役会は一四名で構成されていたが、会長とCEOを除く一二名が社外取締役で、監査委員長はスタンフォード大学の会計学担当教授であった。同社は前年まで優れた経営の企業として評価されていたが、実態は不正な会計処理によって利益が水増しされていたことが明らかになった。しかも、ほとんど社外取

締役で構成される取締役会はそのらの不正を見抜かず、社員の内部告発によって露見したのである。

総じて、社会や経済の現象についての因果関係の説明については厳密性に欠け、学術論文においても相関関係と因果関係が混同されている場合が散見される。

それでは、因果関係はどのようにすれば証明できるのだろうか。相関関係は、単に複数の事象が関係しているということを示すだけで、どちらが原因でどちらが結果であるかは示されない。これに対して、因果関係は事象間の関係は一方的で、逆の関係はない。厳密な因果関係が要求される医療分野では、原因と結果の両者に影響を及ぼし相関関係にみられる「まったくの偶然」、「交絡因子」、「逆の因果関係」の存在を排除するための統計的手法が確立している。「ランダム化比較試

験」である。これは原因となるようなものを与えた「介入群」と与えなかった「対象群」をランダムに割り付けて効果を測定する方法である。例えば、喫煙とガンの発症、肥満と心臓病などの関係、あるいは創薬の効果の測定などにも使われる標準的な方法である。

二

ビッグデータが科学的アプローチに及ぼす影響については、なお検討する余地がある。Hey *et al.* (2009) は、科学パラダイムは現在、「第四のパラダイム」に沿って進むようになったという。第一のパラダイムは、ギリシャ以来の自然を経験的に理解する方法 (experimental science)、第二のパラダイムは理論と実験の組合せによる方法 (theoretical science)、第三のパラダイムは現実

の複雑な現象をコンピュータでシミュレーションする方法 (computational science)、そして第四のパラダイムはビッグデータから新たな科学的な発見を見い出す研究 (データ稼動型科学 (delta-intensive science)) である。

このうち、第二のパラダイムはケプラー、ニュートンなどの一七世紀の「科学革命」に象徴されるように、理論を実験データによって検証し、現実と不整合であれば仮説を修正し、再びデータと突き合わせるといふ作業の繰り返しによって近代科学を発展させた。

ところで、この過程で特徴的なことは、既存の理論とは不整合な新たな事実の発見がパラダイムの転換をもたらしたのではないということである。コペルニクスの惑星運動の地球中心説に代わる「太陽中心説」の主張、あるいはケプラーの惑星の円運動から「楕円運動」の主張は、いずれも

既に知られていたデータの再解釈によって構築されたパラダイムである。これらの転換は信仰上の呪縛構造（つまり、神の創った世界は完全であるべきで、完全な図形である円や神が創造した地球が宇宙の中心であるべきという信仰）から解放されることによって初めて可能になったのである。

確かに、ケプラーは師のティコ・ブラーエが残してくれた火星の軌道に関するデータの解釈によって楕円軌道を発見したのであるが、その探索は円運動、卵型軌道、そして最終的に楕円軌道を仮定することによって、ようやく整合的な説明が可能となったのである。また、一八世紀まで支配的であった、燃焼現象について物質からフロギストン（燃素）が逃げ出すことで起きるとする説は、燃焼後の重量が重くなっていることについて説明ができなかった。このため、苦し紛れにフロギストンは負の重量を持つという、現在からみれば

非現実的な仮定から解放されることによって、元素の結合によって説明できたのである。

このパラダイムの影響力は広範かつ強力で、特にニュートン古典力学の影響はあらゆる科学分野に及んだ。その象徴的な例は「ラプラスの悪魔」であろう。フランスの数学者、科学者であるピエール・シモン・ラプラスは、ニュートンの古典力学によって、すべての物体の現在のみならず過去、将来の動きが説明可能となったことに深く感銘し、かつての士官学校での教え子であったナポレオン皇帝に「私にはもう神は必要ありません」と述べたという。

物体の運動に関する古典力学（万有引力の法則）の影響は社会科学にも及んだ。経済学の生誕地はアダム・スミスの『諸国民の富』の公刊（一七七六年）を嚆矢とするが、リカード、マルクス、また一九世紀末の限界革命という発展の中で、い

わば「法則史観」とも呼ばれる思想が大きな影響力を持った。この例として、マルクスの「必然性」という言葉や「法則」の多用、また、収穫通減の法則、限界効用通減の法則など、枚挙に暇がない。しかし、法則や必然性というのは数学的には確率一〇〇%、あるいはそれに近い実現性を指すもので、そもそも事前には予想しがたい人間行動によって影響を受ける社会や経済の現象にはなじみにくい。

このような問題点があるにしろ、第二のパラダイムは科学を大きく発展させたことは疑いがない。それでは、第四のパラダイムであるデータ稼動型科学、あるいはデータ・サイエンスはこれらに取って代わるのであろうか。

四

第二のパラダイムである「理論的方法」は、帰納法と演繹法という区分からいえば、後者のアプローチである。すなわち、課題設定、仮定、演繹、命題、検証という一連の手続きを経て、理論は証明される。他方、帰納法は多数の事例から共通に見いだされる要因を抽出し理論を構築する。しかし、この方法では、後にそれに反する事例が一つでも発見されると、理論の座から滑り落ちることになり、科学的アプローチとしては限界があるとされる。

経済学の分野でこのアプローチを代表する現代資本市場理論を取り上げよう。これらは、現代ポートフォリオ理論、最適資本構成論、最適配当論、オプション価格理論などから構成され、一九

五二年のハリー・マーコビッツの論文「ポートフォリオ選択論」から一九七三年のフィッシュャー・ブラックとマイロン・ショールズの論文「オプション価格理論」の発表までわずか二年の間に構築されたものである。ピーター・バーンスタインはこれを「思想革命」(Capital Idea)と呼んだ。その後、いく度かノーベル経済学賞の対象になっているから、これらの業績は一般に認知されていると言って良いであろう。

この理論に共通する枠組みは演繹法であり、課題を設定し、一定の仮定に基づき命題を演繹し、それらをデータと照合し、検証する。そして、導きだされた命題が整合的でなければ、再び仮定を再検討し、演繹という作業を繰り返すのである。

これらに共通した主要な仮定は、完全資本市場、投資家の合理的行動などであるが、それに加えて明示的には言及されない場合もあるが、投資

家の予想リターンについての正規分布である。これらの仮定は、リスク管理の標準的な手法であるVAR(バリュー・アット・リスク)でも同じである。いわば、自然科学の空気抵抗ゼロの真空状態を真似てモデルを創り、そこから理論命題を導き出すのである。

もう一つ重要な仮定は効率的市場仮説である。これはあらゆる情報が証券の市場価格に迅速に織り込まれることを主張するもので、一九六五年のユージン・ファーマの論文が記念碑的労作とされる。

一九八〇年代には、これらの主張が学会の雰囲気を支配していた。また、投資実務の世界でも、一九七四年のインデックス・ファンドの開発など、投資運用は市場に打ち勝つことができないうような考え方が影響力を強めていった。

この流れを反転させたのは、一九八七年一〇月

のブラック・マンデー（株価の暴落）であった。効率的市場仮説では、あらゆる情報が取引価格に織り込まれているはずであるから、価格の大幅な変動は新たな情報が市場にもたらされることによつてしか起きないはずである。しかし、ブラック・マンデーは特に大きな情報が引き金となつて起きた訳ではなかった。このため、この事件をきっかけに、効率的市場仮説の説得力は大きく削がれることになる。さらに、これらの議論の前提になっている正規分布の仮定にも厳しい批判が向けられるようになる。

その後の分析により、株価の暴落の原因として、当時、資産運用の世界で広く採用されていたポートフォリオ・インシチュアランスという投資戦略があげられた。これは文字通りポートフォリオに保険を掛けることが目的で、株価が一定の範囲を越えて下がると、自動的に株価指数の先物を売

る仕組みになっていた（ダイナミック・ヘッジング）。理論的には、これによつて株価下落によつて生じる損失をカバーすることができる。しかし、ブラック・マンデーの際には、先物の売りが加速し、市場がそれを吸収することができなくなったため、先物価格の下落が現物市場に波及し、先物の売りにさらに拍車がかかった。まるで、カスケード（滝）のように株価の暴落をもたらしたのである。

このブラック・マンデーのメカニズムを明らかにしたSECの報告書は、株価下落の際にも大量のポジションの処理のインパクトを軽減する可能性として、「上場投資信託」（ETF）のような商品を示唆する。これがヒントとなつて、当時、振興策を模索していたアメリカン証券取引所（AMEX）、地方証券取引所によつて現在のETFが開発されるのである。⁽⁵⁾

ブラック・マンデーでは、株価は一日で二九・二%の下落を記録したが、ベノワ・マンデルブロに拠れば、正規分布の下でこのような株価の下落が生じる確率は一〇の五〇乗分の一以下であり、その仮定の非現実性が如実になった。ナシーム・ニコラス・タレブのいう「ブラック・スワン」(起こりえないこと)である。しかし、その後の東南アジア、ロシア通貨危機(一九九七―九八年)、ITバブルの崩壊(二〇〇〇年)、サブプライム問題の深刻化とリーマンショック(二〇〇八年)、フラッシュ・クラッシュ(二〇一〇年)の発生を踏まえれば、世界の金融システムは「グレート・モデレーション」(大平穩期)どころか、危機の発生の頻度が高まるとともに、危機の拡大と深刻化が進行しているといつて過言ではない。

この過程で、一九九八年には無リスク(である

はずの)裁定取引によって継続的な利益を上げるはずであった、ノーベル経済学賞受賞者を擁しドリーム・チームと呼ばれたヘッジファンドのLTCM(ロングターム・キャピタル・マネジメント)が破綻するが、その原因は市場の不安定さが強まる時には、参加者が取引を控え、流動性が枯渇するということであつた。要するに、買手が不在になり、市場そのものが消失するのである。市場が正常に機能している場合の正規分布を前提した取引戦略は破綻する宿命にある。

一方で、現実にはべき乗分布に近似していると、クレジット・デフォルト・スワップやブラック・スワン・プロテクションという暴落に備えた運用戦略が登場している。⁶⁾

このように、実証的に市場効率性が確認できる分野では市場そのものに投資するインデックス運用が支配的戦略になる一方では、テール・リスク

が問題となる非効率的な市場では、そのリスクに対応した運用戦略が採られているのが現実である。

五

そこで、第四のパラダイムであるデータ稼動型科学との関係を考えよう。依然として第二のパラダイムの理論的方法（演繹法）は様々な現象を分析する上で科学的アプローチであると考えられるが、データ稼動型科学（データ・サイエンス）はこれに取って代わることができるであろうか。

演繹法の分析アプローチは、課題設定、仮定、演繹、命題、検証という手順を踏む。このプロセスの最初の段階は研究の問題意識そのものであるから、研究者の思考過程に裏付けられたオリジナリティなもので、他では代位できない属人的なもので

ある。作業の枠組みが決まれば、一定のプログラムを組んでコンピュータを走らせることが可能で、人間よりも作業効率が高いであろう。そして、検証についても、それこそビッグデータから必要な情報を引き出すことができよう。コペルニクスやケプラーの時代の科学者のように、理論的な命題を導き出す時に、信仰が障害になることもないであろう。

このようにみえてくると、研究の発端の問題意識と課題設定はコンピュータでは代位できず、しかも重要なことは、それによって取捨選択されるデータも左右されるということである。

従って、事実（データ）はそれ自体は何も語らない。それどころか、分析の対象として取捨選択されたデータは一定の視点から選ばれたもので、決して現実そのものではない。そこには、シグナルもノイズも含まれてはいるものの、現実の一部

を切り取ったものである。ここに理論のパラドキシカルな性格がある。

つまり、理論に整合的だと思われるデータが取捨選択され、しばしばそれとは不整合な都合の悪い事実は無視ないし軽視されるということが起きる。事実（データ）は理論依存的であり、価値判断から自由にはなりえないのである。これが科学革命が理論の修正を迫る新しい事実の発見によってではなく、その理論の拠って立つ準拠枠の変更によって引き起こされた理由である。すなわち、パラダイム・シフトというのは分析の準拠枠の転換をとともなうのである。

AIの発展により、シグナルのみならずノイズも含めてあらゆるデータがすくい上げられようと、科学と現実へのアプローチの本質には基本的な変化はないのである。

(注)

- (1) ワイガンド (二〇一七)、第七章一節参照
- (2) 矢野 (二〇一四)、第五章
- (3) ワイガンド (二〇一七)、三〇六頁
- (4) 以下の例は、中室、津川 (二〇一七) による。
- (5) ETF開発の経緯については、ウイグルスワース (二〇二四)、第一章、を参照。
- (6) この分野の最近までの動向については、パターンソン (二〇二四) を参照。

(引用・参考文献)

- 武村彰通 (二〇一八)、『データサイエンス入門』岩波新書
- 中室牧子、津川友介 (二〇一七)、『原因と結果の経済学』ダイヤモンド社
- 丸山真男 (一九六二)、『日本の思想』岩波新書
- 村上陽一郎 (一九八六)、『近代科学を越えて』講談社学術文庫
- 矢野和男 (二〇一四)、『データの見えざる手』草思社
- Anderson C. (2008), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired*, Jun 23
- Bernstein P.L. (2007), *Capital Ideas Evolving*, 山口勝業訳 (二〇一七)

- 〇〇九)、『アルファを求める男たち』東洋経済新報社
- Butterfield H. (1957), *The Origins of Modern Science, 1300-1800*: 渡辺正雄訳 (一九七八)、『近代科学の誕生』(上)(下)、『講談社学術文庫
- Hanson N.R. (1958), *Patterns of Discovery*: 村上陽一郎訳 (一九八六)、『科学的発見のパターン』講談社学術文庫
- Hey T., Tansley S. and Tolle K. (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*
- Patterson S. (2024), *Chaos Kings: How Wall Street Traders Make Billions in the New Age of Crisis*: 月谷真紀訳 (二〇二四)、『カオスの帝王』東洋経済新報社
- Weigend A. (2017), *Data for the People: How to Make Our Post-Privacy Economy Work for You*: 土方奈美訳 (二〇一七)、『アマゾンノミクス データ・サイエンティストは、どう考える』文藝春秋
- Wiggleworth R. (2021), *Trillions: 貫井佳子訳 (二〇二四)、『インデックス・ファンド革命』日本経済新聞社*

(次)が たかお・当研究所名誉研究員)