

大規模情報時代の科学的リテラシーとしての 統計思考力について

統計数理研究所 北川 源四郎 (Genshiro Kitagawa)¹
The Institute of Statistical Mathematics

1 はじめに

P.F.Drucker は著書「歴史の哲学」の中で次のように述べている。

歴史にも境界がある。…数百年に一度、際立った転換が起こる。社会は数十年をかけて、次の新しい時代のために準備する。世界観を変え、価値観を変える。社会構造を変え、政治構造を変える。技術と芸術を変え、機関を変える。やがて50年後には、新しい世界が生まれる。…この転換は2010年ないし20年まで続く。

この言葉を信じれば、我々は数百年に一度の大転換の真ただ中にあるということになる。実際、高度経済成長につづく情報化によって、戦後の人間社会はこれまで経験したことのない速さで急速に変化している。その一方で、東北地方の地震および津波に起因する大災害は、社会の在り方や個人の生き方の再検討を迫っている。

このような変化の時代において、学問の基礎・社会の基盤の役割を担う数学や数理科学とくに統計数理の役割も大きく変化していくに違いない。本稿では、社会の変化に伴って生じた、科学的方法論の変化とそれに対応するための人材育成に関して考えてみたい。

2 社会の変化と科学的研究対象の拡大

20世紀後半に始まった情報通信技術 (ICT) の急速な発展は社会の変化をもたらし、学術や科学・技術の在り方にも大きな影響を及ぼしてきた。情報社会の成立によって、情報は物質とエネルギー以上に重要な概念となり、情報の多寡が社会での成否を支配する重要な要因となったのである。

しかしながら、21世紀の現在、更なる変化の波が押し寄せている。ユビキタス社会の到来である。万人が等しく、何時でも何処でも大量の情報にアクセスできるユビキタス社会が実現すれば、情報それ自体の相対的価値は失われてしまう。ユビ

¹現在：情報・システム研究機構 (Research Organization of Information and Systems)

キタス社会においては最早、情報の量が問題ではなく、万人が共有する大量の情報から如何に目的にあった有用な知識を獲得できるかが問題となる。したがって、今後の数学や数理科学の教育を考えるにあたっては、このような現状を踏まえ、科学の方法的基礎としての社会における現代的役割を考えるとところから始める必要がある。

19世紀までの科学・技術はニュートン・デカルトパラダイムの下で、微分方程式で表現できるような確定的な物理世界を対象に発展してきた。しかしながら、19世紀中盤に提唱された進化論は自然観に大きなインパクトをもたらした。生命はもとより、我々が住む宇宙も地球も進化し発展するものであり、いわば確率的現象の一実現値と考えるべきものということになったからである。この強烈な影響のもとで、19世紀末には、K.ピアソンはあらゆる現象を科学の対象となりうることを主張し、「科学の文法」を提唱した(K. Pearson (1892))。その結果、生命現象に限らず宇宙、地球環境、経済、社会などあらゆる進化し発展するものへと、科学の適用対象が拡大されたのである。

そして、21世紀の現在、情報通信技術の飛躍的発展が人間社会を大きく変化させてつつある。今や多くの情報は、物理世界や進化世界の実験値・観測値としてではなく、むしろ人間社会が作り出した人工物として存在し、広大なサイバー世界を構築しつつある(図1)。このような社会の変化が科学的リテラシーとしての数学や統計数理に影響を及ぼさないはずはない。研究においてもまた教育においても、この現実社会の変化への対応を迫られている。

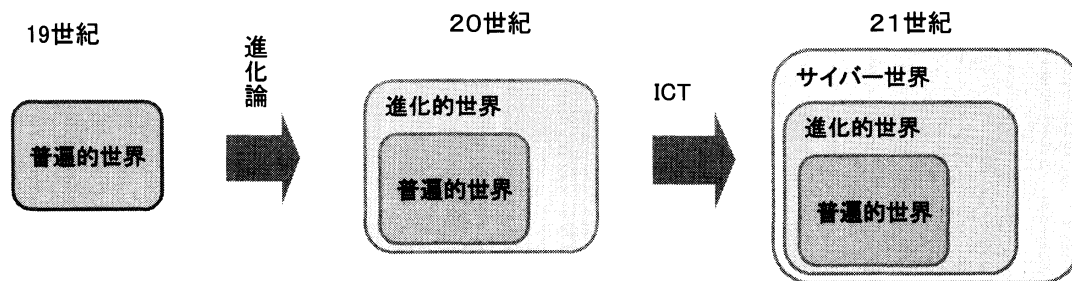


図1: 科学の対象の拡大

3 科学的方法論の発展とデータ中心科学

20世紀までの、学術研究や科学・技術は実験科学と理論科学の二つの科学的方法論を駆動力として発展してきた。理論科学はモデルを仮定する原理駆動型の演繹的方法であり、数学は科学の言語と言われてきた。一方、実験科学はデータ駆動型の

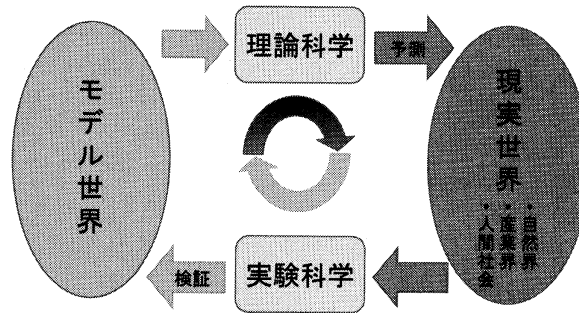


図 2: 理論科学と実験科学がつくる知識発展のサイクル

演繹的方法論であり、統計学は科学の文法とも言われその一翼を担ってきた。この二つの方法論はむしろ意識的に分離され、その緊張感のなかで予測と検証のサイクルに基づく学問発展の手続きが確立していた(図 2)。

この枠組みに変化が生じたのは 20 世紀の終盤である。この時期、複雑な非線形方程式で表現される流体などの非線形現象、人間の行動や心理が関与しゲーム的要素を含んだ経済現象、多くの要素が複雑に相互関連する大規模な人工システムや生命現象などに対しては、従来解析的方法では実用的な解を与えることができなかった。しかし、ちょうどこの時期、計算機と計算技術が飛躍的に発展し、これまで歯がたたなかった複雑なモデルに対しても、その挙動を直接シミュレーションする計算科学が確立した。伝統的な二つの科学的方法論がいわば研究者の知識とひらめきに依存した職人芸の科学的方法論と考えるならば、新たに発展した計算科学は計算機に代表される ICT の発展が可能にした演繹的方法論という位置付けになる。

しかしながら、演繹的方法だけの発展は片手落ちといわざるを得ない。計算科学だけでは、現実世界からのフィードバックの仕組みが欠落しているのである。近年、測定機器、計算機、インターネットおよびデータベースなどの情報通信技術の発展によって、多くの学術分野や社会において大量・大規模なデータが蓄積しつつある。生命科学におけるゲノムデータやマイクロアレイデータ、マーケティングにおける POS データ、ファイナンスにおける高頻度データ、交通情報、環境データ、防災データ、天文データなど大量・大規模データの例は枚挙に遑がない。第 4 の科学とも呼ばれるデータ中心科学が、このような大量・大規模データを組織的に活用する新しい科学的方法論として確立しようとしている(図 3)。

この中で、科学の対象の変化に対応してモデルの意味が大きく変化したことは注目に値する。古典的な科学研究においては、普遍の真理の探究が主要な目的であり、そのためのモデルは普遍の真理を表現あるいは近似したものであり、それを発見し、推定し、あるいは検証することが主要な課題であった。しかし、知識社会におけるモデルは、知識獲得・知識発展あるいはシステム設計のためのいわば道具であり、モ

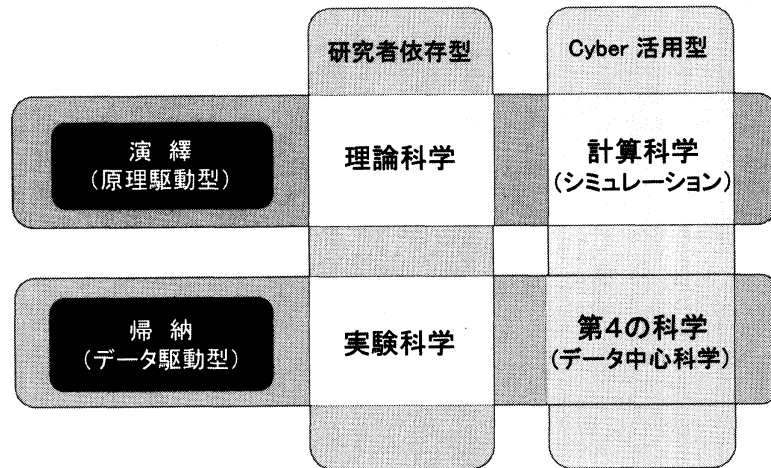


図 3: データ中心科学の位置づけ

デルは発見するものではなく構築すべきものとなる。その場合、モデルは対象に関する理論、経験的知識、現在のデータさらにはモデリングの目的をも考慮して構築すべきものであり、いったんモデルができれば、情報抽出、予測・シミュレーション、制御や意思決定は原理的には演繹的に実現可能となる。

モデリングにおいて、本質的な問題は普遍的な知識と個別的情報の統合である。現実のモデリングにおいて情報統合の問題は、事前情報とデータの持つ情報の統合、時間発展とデータ更新、シミュレーションモデルとデータの持つ情報の統合のように様々な形で現れる。これを少し別の観点から見ると以下のようになる。

3.1 能動的モデリング

対象に関するあらゆる取得可能な情報 (対象に関する理論、経験的知識、観測データ) およびモデリングの目的などの様々な異種情報 (知識、データ) の統合が必要であり、これはベイズモデリングの形で実現できる。一旦モデルが得られると、そのモデルを通して、情報抽出、知識獲得、予測、シミュレーション、制御、意思決定が可能となる。このようにモデルを通して知識が得られ、新しく得られた知識はさらにモデルを改良する。したがって、P.F. Drucker が “Knowledge is being applied to knowledge itself” と表現したように、自然に知識発展のスパイラルが構築できる。このように発展し進化するモデルを中心とした知識発展のプロセスは能動的モデリングともいわれる。

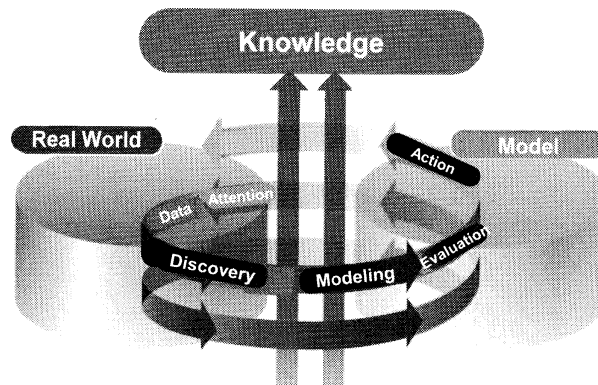


図 4: 知識社会におけるモデリング

3.2 サービス科学の確立

20 世紀の後半から社会は急速にサービス化している。現在では、日本の産業に占めるサービス産業の割合は 70%以上、欧米では 80%といわれている。また、日本が得意とするものづくりにおいてもサービス化が進んでいる。この中で特徴的なことは、効率を目指した社会から一人ひとりの満足度を重視した社会への転換、いければ生産者の視点から消費者の視点への転換である。

これを可能とし、また逆にそれを要求する背景に情報科学技術の飛躍的發展を背景とする大量・大規模データの出現がある。例えば、個人のゲノム情報の獲得は、人間全体集団における効果や副作用を考慮して開発されてきた従来の創薬法に代わって、個人のゲノム情報に基づいて一人ひとりの特性に対応した医療や創薬の実現が現実のものになりつつある。また、マーケティングや情報提供サービスにおいても、個人の購買履歴や検索履歴が自動的に記録できるようになったことから、個人の傾向や希望を推定し、より適切に情報提供ができるようになっている。

この個別化 (personalization) を実現するにあたっては、人間集団に対する一般知識をマクロ的に獲得してきた従来の方法に加えて、個々人に関するゲノム情報や購買履歴などの個別情報を加味していくことが不可欠である。これはまさに一般的知識と個別情報の統合という問題に他ならない。

3.3 データ同化

シミュレーションの分野においては、仮定したモデルの情報と現実の観測データの情報を統合しながら、より現実 に即した予測を行うデータ同化 (data assimilation)

の方法が確立しつつある。従来のシミュレーションは理論モデルを仮定し、与えられた初期値および境界条件のもとで計算により時間発展させるものであった。しかし、現在では急速に発展したセンシング技術の発展を背景にした大量データを活用することにより、より現実に即したシミュレーションや個別的現象に対応したシミュレーションが実現可能となっている。

このようなモデルにもとづく演繹（前向き計算）とデータに基づく機能（後ろ向き計算）を統合するデータ同化の方法はある意味、演繹的方法と帰納的方法の統合という画期的な試みともいえる。この背景には、真理の探究を目指した認識科学から、何らかの目的を実現するための設計科学への転換があることに注目する必要がある。

このように、最近の科学・技術研究の発展は領域間の融合に止まらず、モデルとデータの統合、方法論の融合など、様々な側面での融合に発展しつつある。

4 現代の科学的研究に求められる人物像

このように見てくると、今後の知識社会に必要な人物像や数理科学とくに統計数理が教育において果すべき役割が見えてくる。

まず、物理世界から進化世界・サイバー世界への発展は、問題設定における重要な転換を意味している。進化世界は確率的現象のひとつの実現に過ぎないし、ましてや多くの社会現象やサイバー世界では真のモデルなど始めから存在しない。このような状況では、モデルはある現象に関して理解、知識獲得、予測、制御あるいは合理的判断をするために導入する、ものの見方と考えるべきである。モデルは発見あるいは推定すべきものではなく、現象やデータにもとづいて積極的に構築すべきものとなる。真理なき世界では出来上がったモデルではなく、モデリングのプロセスが重要なのである。

ただし、これはデータにだけ基づいてモデルを推定するという古典的な推論を意味しているわけではない。モデリングにおいては、当該分野の理論、これまでの知識、そして観測データさらには研究の目的にも依存して、柔軟にモデリングを行う必要がある。このような異種の情報の統合は、統計数理や情報学においてベイズ推論を研究の表舞台に引き出している。

また、サービス科学に代表される近年の研究においては、一般原理の探究だけでなく個への対応が重視されるようになっている。例として示したように、医療や創薬における個人のゲノム情報の利用、シミュレーションにおける位置情報、個人情報あるいは環境情報等を取り込むデータ同化などがある。

これらの新しい課題は、モデリング、数理、データ解析、計算技術を備えた新しいタイプの人材を必要としており、このような人材の育成が急務となっている。海

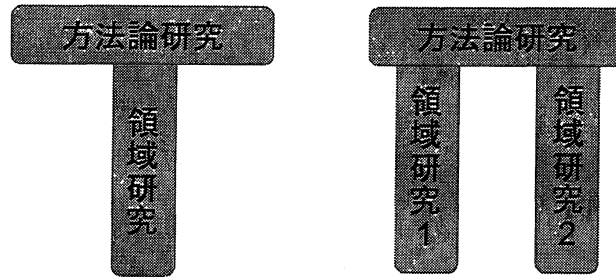


図 5: これからの研究を推進する T-型およびΠ-型人材

外では既にモデラーと呼ばれるこのような人材の供給が始まっている。

今後はモデルにもとづく理論的解析以上にモデリングが重要になる以上、モデリングに不可欠な知識と技術そしてマインドを備えた人材の育成が急務である。モデリングは現実の現象に基づきモデルを構築していくものである以上、現実の問題から離れることはできない。したがって、関連領域の（縦型の）知識と経験およびモデリング、数理解析、データ解析、情報処理などの方法論（横型）知識との二つを兼ね備えていることが必要である。実際には、複数の個別領域間の交流を想定すると、複数の個別領域に関する知識を持った人材が望ましいことは言を俟たない。いわゆる T 型あるいは Π 型の人材である。

統計数理の分野においては、現実の知識とデータにもとづいて、ものごとの本質を捉える統計思考力を持ち、さらに方法論の研究者の強みを発揮して、異分野間交流のハブや研究コーディネータの役割を果たし得る人材の育成を目指すことが必要となっている。このような統計思考力を備えた人材の、今後果たすべき役割は大きい。

参考文献

- P.F. ドラッカー (2003), 歴史の哲学, 上田惇生訳, ダイヤモンド社 (本書は P.F. Drucker の著作集として編集されたもので, Post-Capitalist Society (1993) (和訳: ポスト資本主義社会) を抜粋したものとされている).
- K. Pearson (1892), *The Grammar of Science*, Dover Publications, 2004 edition.