



システム科学とデータ科学

椿 広計*

Data Sciences for Systems Sciences

Hiroe TSUBAKI*

Abstract– This paper first reviews the evolution of data sciences from statistical sciences as the methodology to support the grammar of sciences. In addition, we clarify the roles of data sciences or statistical sciences that has contributed to system sciences, especially the role of approximate representation of the target system and the role of logical inferences in system performance evaluation.

Keywords– Descriptive Modeling, Logical Qualitative Methods, Optimal Design of Data, Scientific Quantitative Methods

1. はじめに

2019年10月から2020年3月まで、内閣府「数理・データサイエンス・AI教育プログラム認定会議」に参画した。大学・高専の学生50万人に対して数理・データサイエンス・AIのリテラシーレベルの教育を行うための認定システム開発を目的とした会議体である。50万人を対象とすることは、数理・データサイエンス・AIが、大学等の文理を問わず全ての学部を横断する基幹学術とされたということである。この際、AIに関わる学術を知能システムの「設計科学」[1]と見なせば、その横幹性は明らかである。特定分野の知能システム開発には、「当該分野に関わる情報（知識と現実）」、「知識や現実から価値を実現するシステムの設計」、そして「設計されたシステムの実現」が必要となる。従ってその実現には、知識ベース・データベース、情報収集環境の整備、システム科学、そして計算機工学からなる科学・工学融合プロジェクトが必要となり、知能システム設計科学は、その種の複合的科学となる。これらの構想は、2014年に（一社）データサイエンティスト協会が公表した、データサイエンティストの3つのスキルセット「1. ビジネス力（課題背景を理解した上で、ビジネス課題を整理し、解決する力）、2. データサイエンス力（情報処理、人工知能、統計学などの情報科学系の知恵を理解し、使う力）、3. データエンジニアリング力（データサイエンスを意

味のある形に使えるようにし、実装、運用できるようにする力）」とも整合的である [2].

一方、汎用的なシステムの認識科学や設計科学とはどう考えれば良いのだろうか？ 椿 [3] は、数理科学を「数学を記述言語として、現象と行動とを認識し、デザインする科学、工学」と呼んだ。また、数理科学が「現象や行動のあるべき姿を演繹的に示す規範科学」としての古典的数理科学と、「帰納的にその姿を明らかにする」計量科学ないしは統計科学に分別されて進化した歴史認識を示した。その上で、当時進行中であった演繹的アプローチと帰納的アプローチとの融合の一形態を「データ中心科学」、今日「データサイエンス」と呼んでいるものとの認識も示した。しかし、この10年急速に拡大を続けるデータサイエンスをどう整理するかが必要な段階とも考える。

一方、科学技術振興機構システム科学ユニット [4] が、システム科学の対象を数学・哲学・数理科学よりも複雑性の高い対象に対する課題認識と課題解決の設計としたことも注目に値する。システム認識科学とシステム設計科学の範囲は、古典的な統計学が対象としてきた現象よりも広い対象を扱っていると考えざるを得ない。

本解説では、先ずデータ科学の成立と発展とを振り返る。その上で、システム設計科学に寄与してきたデータ科学ないしは統計科学の役割、特に対象の近似表現における役割と、対象に対する論理的推論における役割を明確にする。

*統計数理研究所 東京都立川市緑町 10-3

*The Institute of Statistical Mathematics, 10-3 Midori-cho, Tchikawa, Tokyo

Received: 30 April 2020, Accepted: 22 May 2020.

2. データ科学の成立

2.1 科学の文法から PDCA サイクルまで

椿 [5] で示唆したように、今日のデータ科学に近い「統計科学 (Statistical Science)」を提唱したのは、Galton [6] である。Galton は、「大規模な関連する事実の集合情報を議論に適した簡潔な表現に集約する方法の体系化」としての統計科学構築を号令した。それを具現化したのが、Pearson の「科学の文法 [7]」である。Pearson の文法が、Knowledge Management の嚆矢と位置付けられるのは、データとして観測された現象に法則をどのように付与するかというプロセス（「現象の時間的順序・関連性の観測」、「現象に仮説的法則を付与」、「付与した法則の検証」）を明確にしたことである。その上で、Pearson は、その認識プロセスを支援するための数理技術として、可視化技法であるヒストグラムや、数理的方法である標準偏差、相関係数等の初等統計的方法の体系整備を開始したのである。ただし、物理学者でもあった Pearson の「文法」は、現象に関わる法則を認識するための体系、認識科学支援体系であって、現象の背後にある法則を改善するといった、設計科学支援体系、すなわち、システムを人間のために改善するといった側面は希薄である。データは観測されたデータであり、実験的に創るデータへの関心もそれほど強くはない。開発された統計的方法の体系も記述統計学と呼ばれる技法群がほとんどである。

その後、現象の改善やシステムの設計のための統計科学のプロセスを産業界に対して明確に提示したのが、Deming である。Deming は、1950 年に日本で行った講義 [8] に示されたのが、デミングホイールと呼ばれる循環的プロセス「設計、製造、市場投入、市場研究」である。一方、デミングホイールは、1951 年の日本科学技術連盟の品質管理に関する講習会で「計画、実施、チェック、アクション」と抽象化された (Imai [9])。これが PDCA (Plan, Do, Check, Act) と翻訳され世界に広まった。Deming は PDCA を PDS (Study) A とすべきと、当時から批判的だったので、PDCA サイクル導入には石川馨、朝香鐵一、水野滋ら当時日本科学技術連盟に存在した SQC (Statistical Quality Control) リサーチグループが関わったと推定される。加えて、石川の顕著な貢献はチェックからアクションに移行する改善の標準シナリオ（問題解決型 QC ストーリー）を産学連携活動の中で確立し、世界標準としたことである [10]。統計科学的には、PDCA サイクルは、現象の改善やシステムの設計のためのプロセスモデル、すなわち設計科学の文法の一つに位置づけられる。実際、国際標準化機構が発行するマネジメント・システム規格は全て PDCA サイクルを用いて記述することが要求されている [11]。すなわち、PDCA という素朴なプロセスモデルは、組織目的を達成するための業務システ

ムの記述モデルともなっている。なお、PDCA サイクルの PDC についての抽象化は、品質管理学のパイオニアで Deming が強い影響をうけ、Pearson とも交流のあった Shewhart [12] が、品質管理における「仕様の決定」、「生産」、「検査」のサイクルを「達成すべき目的の設定」、「目的を達成するための行動」、「目的が達成できたか否かの検証」へと一般化したことによる [13]。

2.2 統計科学のデータ科学への改組？

「データ科学 (Data Science)」という用語出現のきっかけは、Tukey によるデータ解析 [14]、とりわけ「探索的データ解析 (Exploratory Data Analysis)」[15] の提唱である。データによって、現象の背後にあるメカニズムを検証するだけでなく、メカニズム自体を発見しようという態度が提案されたと言われている。しかし、筆者は Tukey の態度は、Pearson の科学の文法の態度を計算機科学によって現代化したものに過ぎないと考え、Pearson が法則を現象に付与するために、ヒストグラムや相関分析をはじめとして様々な記述統計学的方法を開発したのに対し、Tukey はそれを支援する可視化機能を強化した計算機環境 S 言語（今日の R 言語）の開発を指揮した。その点でも両者の態度は類似している。

竹内 [16] は、「『統計学』とは何か？」ということ語る中で、統計的決定関数 (Decision Function) などを中心とした「応用数学」と見なす立場、R. A. Fisher や当時の Bayesian が主張したように、データからの推論を行う「帰納的論理学」と見なす立場と対比して、自身は「データ解析の科学」ととらえたいと主張し、むしろそれを確立したのが R. A. Fisher だと主張した。

一方、「データサイエンス」という用語をわが国統計学コミュニティが用いるようになったのは、1996 年 3 月に神戸で開催された、データサイエンスと銘打った国際学会からである。これは、国際的にも極めて早い時期である。それを主導した林 [17] は、Tukey の探索的データ解析を評価すると共に、統計数理におけるデータ解析は「データによる現象解析」とした。その上で、データサイエンスを統計学（検証的データ解析）と探索的データ解析の単なる融合ではなく、データの意味や品質も含めた、「データの設計」、「収集」、「解析」の 3 フェーズからなる新たな概念として提唱した。新村 [18] によれば、1996 年 9 月に日本統計学会も柴田 [19] をオーガナイザーとして、「統計をデータサイエンスとして発展させよう」という趣旨のセッションを企画した。新村 [18] は、1997 年当時、データ解析ソフトウェアとネットワーク技術の共生、情報処理システムよりデータ解析に重点を移した人工知能の開発などをデータサイエンスへの途として提唱して今日までである。

片岡 [20] によれば、慶應義塾大学湘南藤沢キャンパスでは、早くも 1997 年度のカリキュラム改革で第 3 の

基礎教育分野として、データサイエンス教育を「知的創造のための思考の道楽」というキャッチフレーズのもとで開始し、筆者もその活動に協力すると共に、現時点での最高水準のデータ解析戦略体系（当時出現した第2世代人工知能を用いた）教育の必要性を説いた [21].

1997年には Wu[22] が、林 [18] と類似のアイデア、統計的活動を「データ収集（実験計画法、標本調査論）」、「モデリングと解析」、「問題の理解と解決に基づく意思決定」の3フェイズとし、統計学をデータサイエンスに、統計家をデータサイエンティストと改名すべきという提唱を行った。Wu は、統計科学の進化の方向性を「大規模複雑データ」、「ニューラルネットワーク」、「帰納的アプローチと演繹的アプローチとの融合」、「知識のベイズ的表現」、「計算アルゴリズム」、「認知科学との相互作用」とした。Wu の示唆は、全て今日のデータサイエンス分野の主要な課題である。Wu は、ニューラルネットの予測性能が良いのが、ニューラルネットワークシステムの背後にあるシステムが複雑関係を表現できるからか、あるいは、予測誤差評価基準としての Cross Validation が統計モデルを自動的に簡潔なものとする「ケチの原理」に繋がるからかなど、今日深層学習分野でも議論される問いかけも行っている。

なお、統計科学専門家としての「統計家 (Statistician)」重要性を一般に知らせたのは Google のチーフアナリストでマイクロ経済学者 Varian[23] の “the sexy job in the next ten years will be statisticians” とした発言である。Varian は、データが無料で得られる経済環境が、データから経済価値を生成することを生業とする統計家の地位を向上させるとしている。

2.3 データ科学へのエンジニアリングスキルの導入

データ科学の専門家としての、「データサイエンティスト (Data Scientist)」の社会的重要性を、Varian 発言を修正するかのよう提示したのが、経営学者である Davenport と、まさにデータサイエンティストである Patil が提示したメッセージ [24] である。彼らが焦点を当てたのが、構造化されていない情報を結合し、構造化されたデータに加工し、データの可視化などから、企業価値の発見に貢献する専門職である。ここに至って、データサイエンティストが、統計家よりはコーディングスキルやデータベースマネジメントに秀でた専門職であることが強く印象付けられた。データサイエンティスト協会 [2] のスキルセットのエンジニアリング力をデータサイエンス力、すなわち古典的統計家のスキルセットよりは、重視したように見える。統計分析よりは科学的探究心といった力量も強調されている。実際に7週間で物理学などの学位をとった若手研究者をデータサイエンティストに速成する INSIGHT FELLOW Program も 2012 年に立ち上がり、

2020 年現在 3000 人を超える若手データサイエンティストのキャリア形成が実現している [25].

一方、米国政府が統計家などを専門職種として公的統計で調査していることは重要である。Rieley[26] は、2018 年公表の米国労働統計を用いて、数学家・統計家・オペレーションリサーチアナリスト・アクチュアリーからなる数理科学専門職が今後 10 年急増し、多くの雇用が創生されるとの見通しを発表した。特に、データサイエンスを統計学・コンピュータサイエンス・エンジニアリング・マネジメントの複合科学と位置づけ、統計家がデータサイエンティスト分野でも高い需要があるとした。

実際に、2020 年に米国労働統計局公表統計 [27] に基づき統計家と共に、データサイエンスに進出していると労働統計局 Occupational Outlook Handbook に記載された「計算機情報研究科学者 (Computer and Information Research Scientist)」などの雇用者数、10 年後の推計値 (増加率%)、年収を Table 1 に示す。Table 1 より統計家の増加率は一番高く推定されていることが分かる。

Table 1: Employment Statistics and Predicts for Data Science Related Professionals from Occupational Outlook Handbook by U.S. Bureau of Labor Statistics.

	Employment		Entry Level Education	2019 Median Pay
	2018	Predicts 2028		
数理科学 高等教育 教員	61,100	62,900 (+2.9%)	博士	\$73,690
数学家	2,900	3,600 (+24.1%)	修士	\$105,030
統計家	44,400	58,000 (+30.6%)	修士	\$91,160
OR 分析家	109,700	137,900 (+25.7%)	学士	\$84,810
計算機科学 高等教育 教員	40,200	41,900 (+4.2%)	博士	\$85,180
計算機 情報研究 科学者	31,700	37,000 (+16.7%)	修士	\$122,840

3. データ科学のシステム科学への寄与

3.1 統計的方法からの寄与

以下では、1節で紹介したシステム科学に対してデータ科学の一部を形成する統計科学的方法・数理科学的方法がこれまでどのようにシステムの記述や利用に寄与したかについて議論したい。実際、統計的方法ないしはデータに基づく科学的方法がシステム科学に寄与する方法には異なる接近がある。なお、システム科学の問題解決プロセスにおいて、計算機科学的実装や構造化されていない情報の効率的収集と構造化とが重要なことは明らかである。しかし、筆者は、データ科学におけるその分野での専門家ではないので、その具体像を解説できない。

統計科学的接近の分類軸は、記述的 (Descriptive) 接近と推測的 (Inferential) 接近である。記述的接近は、関心のある現象に関連するデータを用いて、当該現象の近似的表現すなわちモデルを得ることを目的とする接近である。一方、推測的接近は、現象に関連するデータを用いて、当該現象について意思決定、すなわち論理的判断を行うことを目的とする接近であり、2.2節で示した帰納的論理学としての統計学の利用である。

3.2 システム科学への記述的接近とその役割

筆者が主査として取りまとめに参画した、科学技術振興機構研究開発戦略センターシステム科学ユニットモデリング分科会報告書 [28] には、統計科学的接近を超えて、一般の対象を近似するモデルに対して次のような分類軸が示されている。

- 1) 対象 (測定, システム, プロセス)
- 2) 利用目的 (予測, 制御, 意思決定等)
- 3) 近似したい状態 (理想状態, 標準状態等)
- 4) 近似したい場 (静的, 動的)
- 5) 記述様式 (直観的, 論理的, 数理的)
- 6) 特性 (主観的俊敏性, 客観的再現性)
- 7) 構造 (決定論的, 非決定論的)
- 8) 同定に利用する情報 (マクロ的, ミクロ的)

今日振り返ると、当時の議論に加えて、諸科学の原理に基づく理論的記述なのか、観測に基づき帰納的記述なのかも重要である。これは、同定に利用する先験情報の利用による分類とも考えられる。

- 9) 接近 (演繹的, 帰納的)

特に、統計的計量諸科学を支える近似モデルは、主として観測データに基づく帰納的近似である。また、近年統計的機械学習分野で論点になっているのは、モデル記述が **Black Box** か否か、すなわち解釈可能性があるかどうかである。

10) 解釈可能性 (Black Box, White Box)

分類軸 10) は、分類軸 6) の別側面とも考えられるが、意識すべきと考える。この種の分類軸を用いれば、データサイエンティストが重要視する情報の可視化は、主観的俊敏性を有する対象の帰納的かつ直観的な記述である。また、古典的計量科学モデルは、システムを対象とし、客観的再現性を求めた帰納的かつ非決定論的記述である。

更に、対象を記述し、その挙動に影響を与える様々な情報も、例えば次のように分類可能である。

- 1) 制御性を有する外生的情報・
対象内の関連性の影響を受ける内生的情報
- 2) 観測可能な情報・
実在するが観測できない潜在情報
- 3) 量的情報・質的信息
- 4) 確かな情報・不確かさを有する情報

ここで、潜在情報には特別な変数として、現在の人知では説明不可能な「誤差」情報も含めることができる。統計科学では、誤差変数を確率変数として表現するのが通常である。また、システムに含まれる未知パラメータを潜在変数と考え、確率変数として表現するのが、ベイズモデルである。これら確率分布の導入は、それらの変数が想定した確率分布に実際に従うという真実性よりは、現時点では原因不明の変動を偶然変動と仮置きする方便と考えるべきである [29], [30]。実際、誤差の推定量 (残差, residuals) が、想定した確率分布が示唆する範囲に入らない、すなわち異常変動を示した場合、その原因追究が、統計的問題発見の標準的方法である。

なお、これらの記述様式の分類は、確定的なものではなく、折衷されるあるいは融合されるべきものである。実際、統計科学の進化の中で、システム科学に最も寄与している方法論が、学術原理に基づく演繹的近似とデータへの適合、すなわち帰納的接近とを融合させる「データ同化 (Data Assimilation)」である [31]。新型コロナウイルス感染者数の推移と対策効果予測を目的として、国内外で感染症数理モデルによるデータ同化が行われ、各国の対策決定の参考にされたことは記憶に新しい。

さて、統計科学・数理科学自体の進化系ではないが、そのシステム科学分野への高度な適用が顕著となっているのが、「確率ロボティクス」[32]である。そこでは、逐次的に統計モデルが更新され、それに基づき価値関数最適化行動が実装される。ある意味で、知能システムと行動の設計科学の典型例である。

Kitagawa[33]の非ガウス、非線形状態空間モデルに対するモンテカルロフィルタリング (粒子フィルタリング) は、データ同化や確率ロボティクス理論実装に不可欠な横断的機能を果たしている。

3.3 システム科学への推測的接近とその役割

統計学には対象の記述的近似機能以外に、有意性検定 (Pure significance testing)、仮説検定 (Hypothesis testing) や決定関数に代表される推測理論体系が整備されている。対象の不確かさを前提にした帰納的論理学機能である。有意性検定は、作業仮説を想定し、観測されたデータを基に、作業仮説からより遠くなるデータが生じる確率を有意確率 (P 値) とし、有意確率が小さければ作業仮説を棄却するといった推測形式である。単純化すれば、これは背理法の拡張である。背理法は作業仮説が論理的矛盾を生じたときに棄却する。一方、有意性検定は、論理的矛盾の代わりに、小さな確率の現象が起きている現実を作業仮説との矛盾と考える。

統計的推測が、統計的記述異なる特徴は2つある。

第1の相違点は、統計的推測が質的方法と見なせることである。Goertz and Mahoney[34] は、科学的接近を統計的な量的方法と数学的・論理的な質的方法に峻別した。「検査陽性ならば、感染している」という十分条件や「感染していれば、検査陽性である (検査陰性ならば感染していない)」といった必要条件、これらの例外 (前者は擬陽性、後者は偽陰性) 現象が、確率的にどれくらい生じるかを評価するのが、論理的方法に不確かさを加える接近である。統計的推測論は、一見定量的に見えるが、論理的方法を補強する体系と考えられる。

今後、必要な帰結を得ることを目的とし、符号列を目的達成のための入力として、帰結に一定の確率で違反や例外が生じる状況を考え、帰結に十分条件ないし必要条件としての例外が起きないことを評価する方法を確立すれば、検定や検査を拡張した方法論となりえる [35]。これら論理的方法が、目的達成のために設計されたシンボル性ないしはシグナル性プログラム [36] のみならず、システムの評価技術としても重要となろう。

第2の相違点は、論理的方法は利用するモデルの対象近似能力が低くても、無作為化という方法を用いれば、推論が可能なことである。その典型が、新医薬品の許認可制度で用いられてきた無作為比較試験 (Randomized Controlled Trial) である。通常、新薬や対照薬が患者に有効か否かは多くの治療に関わる変数や患者属性に関わる変数に依存する。しかし、それら全てを考慮した綿密な記述モデルを使う代わりに、患者を新薬群と対照群とに無作為に割り付け、新薬群と対照群とで有効性は変わらないといった帰無仮説と観測データの整合性を判断してきた。ここで、有効性への様々な変数の系統的影響は、無作為化を通じて確率変数と見なせるようになるので、統計的推測が可能となる。もちろん、治療や患者の属性と有効性との関係性を表現するモデルを精緻化することは、推論精度の劣化原因となる偶然変動を小さくするので、統計的推論の改善にも繋がる。記述的接近と推

測的接近の融合は勿論有効だし、実際にこれまでも行われてきたことである。

3.4 記述や推測を最適にするデータ設計とその役割

ここまでの記載で統計家という専門職は、記述表現や推測を行う専門職と考えるかもしれない。しかし、統計家の果たしてきた最大の役割は、対象記述や推論の精度を最適化するためのデータ収集計画を行うことである。データは分析の対象だけではなく、データ科学においては、生成ないしは創成の対象であることも重要である。

統計科学では、実験計画や調査計画と呼ばれる収集情報量の最適化計画 (Optimal design) が用いられた。データの収集にコストがかかる環境下では、この種の数理的最適化技法を用いれば、最小コストで必要な意思決定が行える。実験計画分野では、直交計画や一様計画、調査計画では、層化抽出や標本配分など様々な方法が開発されてきた。

シミュレーションによる仮想データ生成が容易になった今日、逐次実験計画法の数理は、統計的機械学習分野で強化学習へと変貌しつつある。しかし、複雑な入出力システムを効率的に記述する一様計画などが、強化学習分野でも大きな役割を果たす可能性は大きい [37]。

4. おわりに

データ科学から創生される経済価値・知識価値は、データ科学単独で実現するものではない。システム科学の中で利用されるあるいはシステム科学と融合することで、その意義が明確になる。逆に、データ科学は統計学を含む多様な横断的学術の複合体である。従って、システム科学適用の経済価値を最適化し、システムによるイノベーションを加速するデータ科学の理論・方法・プロセスの探究こそ、データ科学自体を進化させる原動力と考えることもできる。

謝辞: 本解説の執筆を勧めて下さった木村英紀先生に感謝申し上げます。

参考文献

- [1] 日本学術会議学術の在り方常置委員会: 報告 新しい学術の在り方—真の society for society を求めて、日本学術会議, <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-19-t1032-11.pdf> (2005).
- [2] (一社) データサイエンティスト協会: データサイエンティスト協会, データサイエンティストのミッション, スキルセット, 定義, スキルレベルを發

- 表, <http://www.datascientist.or.jp/files/news/2014-12-10.pdf> (2014).
- [3] 椿 広計: 数理科学の機能, 横幹, Vol.8, No.1, pp. 32-35 (2014).
- [4] 科学技術振興機構研究開発戦略研究センター: 研究開発の俯瞰報告書, システム科学技術分野, CRDS-FY2012-FR-07, <http://www.jst.go.jp/crds/pdf/2012/FR/CRDS-FY2012-FR-07.pdf> (2013).
- [5] 椿 広計: 統計科学の横断性と設計科学への寄与, 横幹, Vol.1, No.1, pp. 22-28 (2007).
- [6] F. Galton: *Inquiries of Human Faculty and Its Development*, Macmillan (1883).
- [7] K. Pearson: *The Grammar of Science*, Adam and Charles Black (1892).
- [8] E. Deming: *Elementary Principles of the Statistical Quality Control* 日本科学技術連盟 (1950). 日本科学技術連盟訳: 統計的品質管理の基礎理論と応用: デミング博士講義録, 日本科学技術連盟 (1952).
- [9] M. Imai: *Kaizen: The Key to Japan's Competitive Success*, Random House (1986).
- [10] 光藤義郎, 小原好一, 椿 広計: QC サークル活動 - 国際化した小集団改善活動 -, 第9回横幹連合コンファレンス, https://www.jstage.jst.go.jp/article/oukan/2018/0/2018_B-4-1/_pdf/-char/ja (2018).
- [11] ISO: Annex L (normative) Proposals for management system standards, ISO/IEC Directives part 1, 10th version, ISO (2019).
- [12] W. A. Shewhart: *Statistical method from the viewpoint of quality control*, The Graduate School of the Department of Agriculture (1939). 坂本平八 訳: 品質管理の基礎概念 - 品質管理の観点からみた統計的方法, 岩波書店 (1960).
- [13] 椿 広計: 第2章日本の本質管理と統計科学, 国友直人, 山本拓 編, 統計と日本社会 データサイエンス時代の展開, 東京大学出版会 (2019).
- [14] J. W. Tukey: *The Future of Data Analysis*, the *Annals of Mathematical Statistics*, Vol.33, No.1, pp. 1-67 (1962).
- [15] J. W. Tukey: *Exploratory Data Analysis*, Pearson (1977).
- [16] 竹内 啓: 付録1 R. A. Fisher の統計学, 竹内 啓, 大橋靖雄: 統計的推測 - 2 標本問題, 入門現代の数学 [11], 日本評論社 (1981).
- [17] C. Hayashi: *What is Data Science? Fundamental Concepts and a Heuristic Example*, C. Hayashi, K. Yajima, H. Bock, N. Ohsumi and Y. Tanaka eds., *Data Science, Classification and Related Methods*, Proceedings of 5th Conference of the International Federation of Classification Societies (IFCS-96), pp. 40-51, Springer (1996).
- [18] 新村秀一: データ解析からデータ・サイエンスへ - 情報技術 (統計ソフト・WWW・AI) との共生により統計知識を国民の知的共有罪に -, 統計数理, Vol.45, No.1, pp. 23-40 (1997).
- [19] 柴田里程: データサイエンスのすすめ, 日本統計学会誌, Vol.30, pp. 327-332 (2000).
- [20] 片岡正昭: データサイエンス教育の理念と成果 - 慶應 SFC の新たな挑戦, 第1部データサイエンス登場, 特集データサイエンス, KEIO SFC REVIEW, Vol.4, pp. 7-13 (1999).
- [21] 椿 広計: データサイエンスの社会人教育, 第1部データサイエンス登場, 特集データサイエンス, KEIO SFC REVIEW, Vol.4, pp. 38-43 (1999).
- [22] C. F. J. Wu: *Statistics = Data Science?*, in 7th series of P. C. Mahalanobis Memorial Lectures (1997).
- [23] H. Varian: *Hal Varian on how the Web challenges managers, Our Insights*, Jan 2009, McKinsey & Company (2009), <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/hal-varian-on-how-the-web-challenges-managers>
- [24] T. H. Davenport, and D. J. Patil: *Data Scientist: The Sexiest Job of the 21st Century*, *Harvard Business Review*, Vol.90, No.10, pp. 70-76 (2012).
- [25] *Insight Your bridge to a thriving career*, <https://blog.insightdatascience.com/>
- [26] M. Rieley: *Big data adds up to the opportunities in math careers, Beyond the Numbers: Employment and Unemployment*, Vol.7, No.8, <https://www.bls.gov/opub/btn/volume-7/pdf/big-data-adds-up.pdf>
- [27] *Occupational Outlook Handbook*, U.S. Bureau of Labor Statistics, <https://www.bls.gov/ooh/home.htm>
- [28] (独) 科学技術振興機構研究開発戦略センターシステム科学ユニット: 平成22年度システム科学技術委員会モデリング分科会報告書, CRDS-FY2010-XR-20, 科学技術振興機構 (2011), <https://www.jst.go.jp/crds/pdf/2010/XR/CRDS-FY2010-XR-20.pdf>
- [29] 竹内 啓: 偶然とは何か - その積極的意味, 岩波新書, No.1269, 岩波書店 (2010).
- [30] 椿 広計: データサイエンス入門 (5) 方便としての確率モデル, 統計, Vol.70, No.11, pp. 69-75 (2019).
- [31] 樋口知之 編著: データ同化入門 - 次世代のシミュレーション技術 -, シリーズ予測と発見の科学 6, 朝倉書店 (2011).
- [32] S. Thrun, W. Burgard and D. Fox: *Probabilistic Robotics*, The MIT Press (2005). 上田隆一 訳: 確率ロボティクス, マイナビ出版 (2015).
- [33] G. Kitagawa: *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*. *Journal of Computational and Graphical Statistics*, Vol.5, pp. 1-25 (1996).
- [34] G. Goertz and J. Mahoney: *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton University Press (2012). 西川 賢, 今井真士 訳: 社会学のパラダイム論争: 2つの文化の物語, 勁草書房 (2015).
- [35] 椿 広計: 統計科学の文法と論理的推論, 統計数理, Vol.67, No.2, pp. 165-180 (2019).
- [36] 吉田民人: 近代科学の情報論的展開 プログラム科学論, 吉田民人論集編集委員会編, 勁草書房 (2013), 横断型基幹科学技術研究団体連合 (略称 横幹連合) ホームページ <https://www.trafst.jp/index.html>
- [37] 山田 秀: コンピュータシミュレーションのための実験計画法 - 一様実験計画と過飽和実験計画, 品質, Vol.34, No.3, pp. 21-29 (2004).

椿 広計



1956年10月20日生。82年東京大学大学院工学研究科修士課程計数工学専攻修了。同助手、慶應義塾大学講師、筑波大学大学院ビジネス科学研究科助教授・教授、統計数理研究所データ科学研究系教授、(独)統計センター理事長を経て、2019年統計数理研究所長、横幹連合副会長現在に至る。応用統計学などの研究・実務に従事。工学博士、日本品質管理学会などの会員。