

林の数量化理論

○椿 広計 (統計数理研究所) 鈴木 督久 (日経リサーチ)

Hayashi's Quantification Theory

* H. Tsubaki (The Institute of Statistical Mathematics), T. Suzuki (Nikkei Research)

Abstract— Chikio Hayashi conceived of quantification to convert qualitative variables into quantitative variables, developed the necessary quantification theory systematically and independently, and took a leadership role in solving social problems supported by quantification theory. Quantification theory and survey design based on its application have taken root in all social analysis activities in industry, government, and academia in Japan as a method for analyzing and improving qualitative variables that are routinely collected and that are factors in behavior. Quantification theory has become a representative methodology for multivariate data analysis proposed by Japan, and has also made a pioneering contribution to the international community in the construction of Data Sciences. For these reasons, the Japanese Society of Applied Statistics recommends "Hayashi's Quantification Theory" for the Kototsukuri Collection.

Index terms— Behavior-metric, Data Science, Optimal Scaling, Qualitative Variables

1 推薦対象および推薦理由の概要

応用統計学会 (Japanese Society of Applied Statistics) は、林の数量化理論、すなわち質的観測変数を量的数値に最適変換する多変量解析方法体系をコトづくりコレクションに推薦する。本報告では、数量化理論の発展と構成要素と国際的評価、社会に果たした役割を概観する。

林は、質的変数を量的変数に変換する数量化を構想し¹⁾、必要な多くの数量化方法を体系的に独自開発し、数量化理論に支援された社会課題解決のリーダーシップをとった。数量化理論とその適用を前提とした調査設計は、行動の要因となりかつ日常的に収集される質的変数を分析する方法として、わが国産官学の社会分析活動全般に定着した。数量化理論は、日本が提案した多変量データ解析の代表的な方法論となると共に、データの科学構築への国際社会に対する先駆的貢献ともなった。応用統計学会は「数量化理論」をコトづくりコレクションに次の理由で推薦する。

- 1) データ分析によって観測変数からより多くの知識価値を抽出することに資する「観測変数の数量化プロセス」というデータ科学を横断する思想の提唱により、データ科学は勿論、計量実証学術全般とその産官応用に共通する一つの研究スタイルを確立したことに対する貢献
- 2) 数量化プロセスに資する統計分析技術の体系的開発と適用プロセス事例を社会に示したことによる、人間の感性や行動に対する数量化を前提とした市場調査、世論調査、社会調査、官能検査等の社会実装によって、モノやサービスを創るためのデータに基づく情報収集活動の変革への貢献

2 推薦対象の定義および構成要素

2.1 推薦対象の定義

森本²⁾が指摘したように、林³⁾は測定量 X そのものではなく統計操作によって妥当な意味を持つ量への変換 $Y=f(X)$ を求めることがデータの科学では、極め

て重要なプロセスと考えた。特に、 p 種類の質的属性 C_i , $i=1, \dots, p$ に対して、 p 次元属性ベクトル $\mathbf{X}=(x_i)$ を考える。属性ベクトルとは測定対象が属性 C_i を持つとき、 $x_i=1$, 持たないときに $x_i=0$ とするベクトルである。林が開発したのは、属性ベクトル \mathbf{X} を一次元の連続変数 Y に変換する方法である。コトづくりコレクションの推薦対象とする「林の数量化理論」は、データ科学を貫く思想と属性ベクトルが複数存在する場合も含む多変量データ解析技法との総称である。

2.2 推薦対象の構成要素

数量化理論は、林により数量化に必要な外的基準の有無と属性により分類された。推薦対象を構成するのは、次の3つの数量化方法である。なお、この流布した命名は飽戸³⁾により、下記に示した初出文献には存在しない。なお、林は、計量的多次元尺度構成に類似の数量化IV類、多変量一対比較に関わる数量化V類も提案したが、調査に関わる産官学への社会的影響が多かったのは以下の3つと考える。

- A) 数量化I類⁴⁾
予測を実施したい外的基準が存在し、間隔尺度ないしは比例尺度といった数量で与えられている場合に、その外的基準を最良予測する問題に対して、その予測に資する要因としての質的変数に数量を与える方法。分析技術的には、質的変数を入力変数とする重回帰分析に相当する。
- B) 数量化II類⁵⁾
判別 (教師付き分類) を実施したい外的基準が存在し、分類尺度ないしは順序尺度といった質的変数で与えられている場合に、その外的基準を最良判別するのに用いたい他の質的変数に数量を与える方法。分析技術的には、質的変数を入力変数とする判別分析に相当する。
- C) 数量化III類⁶⁾
外的基準が存在せず、2つ以上の質的変数が存在する場合、両者の関連性を最適化するために

質的変数に数量を与える方法。分析技術的には、質的変数に対して正準相関分析を拡張した方法に相当する。

これらの方法に関する詳細は、林⁷⁾を参照されたい。ただし、多くの多変量データ解析専門書ないしは教科書に適用事例と共に記述されている。

林の数量化理論は、人間の行動を計量化し可視化するすべての社会科学分野、あるいは医学分野に研究技法として影響を与えたのみならず、数量化を目的とした調査研究スタイルの変革を少なくとも市場調査、社会調査、官能検査などに与えた。数量化理論を構築した林の背景思想は、今日のデータの科学形成に大きな影響を与えたが、この哲学的側面については、推薦対象とすることは難しいので、3節の背景に記載する。

3 当該コトづくりの背景

3.1 数量化理論誕生の背景

林は、1946年統計数理研究所社会科学研究部に着任後、社会調査に関わる様々な活動を支援した。その中で、法務省矯正保護研修所の刑法学者西村克彦と1947年から再犯を1年以内に起こさない仮釈放の予測研究と共同研究を行った⁸⁾。これまでの調査科学でみられる質的変数に間隔尺度と考えられる5点満点あるいは7点満点のスコアを与えて分析データとする Likert scale⁹⁾等の適用に疑問を持ち、Guttman¹⁰⁾の最適尺度構成理論への関心を通じて、人間の行動の原因となる個人・家族・社会・環境に関わる属性を表現する質的変数を適切に「数量化」した数値をデータとして分析対象とすべきという、データの科学を横断する思想¹¹⁾が萌芽した。こうして、今日数量化Ⅱ類と呼ばれる仮釈放成功の判別予測に有効な質的変数に適切な得点を与える数量化理論が初めて開発された。

数量化Ⅰ類は、林が支援した「日本人の読み書き能力」という終戦直後の優れた標本調査で、日本人の読み書き能力を学歴などの質的変数で予測・説明する際に開発された¹²⁾。

数量化Ⅲ類は、どういうデザインがどのような集団に好まれるかの明確化を通じて、輸出用デザインの推奨のために、1956年に実施された「ニチレイ輸出缶詰ラベルデザイン調査」で商業デザイナー佐藤敬之輔との共同研究の中で開発された⁶⁾。

なお、数量化理論開発の経緯は林¹³⁾、日経リサーチ¹⁴⁾にも詳しい。

3.2 数量化理論の社会定着と行動計量科学の創生

林、村山¹⁵⁾が市場調査の分析と計画とを対象とした書籍を出版して以降、当時の計算機環境では実装困難と考えられた数量化理論を前提とした市場調査、社会調査が多数実施された。飽戸¹⁶⁾は、マーケティング、社会心理学、社会学分野などでは調査を行えば数量化理論に基づく分析が必須のように行われた時代もあり、毎週のように実習を伴う講習会があったことを振り返っている。こうして人間の行動や意識を数量化することが、社会にとって当たり前の活動となった。1973年に林が創設し、15年間初代理事長を務めた「日本行動

計量学会」は、その種の活動を産学横断的に繋げる役割を果たした。数量化理論の諸分野への横断的定着の経緯や日本の応用統計学発達に果たした役割は森本²⁾に詳しく紹介されている。

3.3 数量化Ⅲ類の国際化とデータ科学の創生

林の数量化理論の中でもⅢ類と呼ばれる方法は、Guttman¹⁰⁾の尺度構成と数理的には密接な関連がある。このため、計算機が発達した1970年代以降、数量化Ⅲ類と極めて類似した方法が提唱された。特に有名なのは Benzécri¹⁷⁾の対応分析(Correspondence Analysis)と西里¹⁸⁾の双対尺度法(Dual Scaling)である。今日、統計ソフトウェアでは、対応分析の名で数量化Ⅲ類が実装されていることが多い。

椿¹⁹⁾で指摘したように、林²⁰⁾は、国際的にも「データサイエンス」という言葉を最も初期に用いて、今日のデータサイエンスの先駆けとも言える発信を行った。林は、推定や検定を重視してきた数理統計学に、Tukey²¹⁾が(探索的)データ解析という新たな概念を導入したことは評価した上で、データ解析がデータサイエンスではなく、データの設計、収集、解析の3フェイズからなる概念として提唱し、今日のデータサイエンスに繋がる体系を示唆している。数量化が単にデータ分析技法を変えただけでなく、数量化のために調査設計、すなわち社会に必要な体系的情報収集のスタイルも変革したことが、その背景にあると考えるのが良い。

4 達成された内容と生じた社会変化

数量化理論は、人間を対象とする社会科学の発展に寄与し、行動計量学という新たな学際的学術の形成に寄与しただけではなく、人間の特性とサービスや商品の特性とのこれまで暗黙知であった関係性が、数量化を意図した調査を通じて可視化されることで、新サービスや新商品の企画、あるいはこれまでのサービスなどの改善に分野横断的に寄与した。

3.1で述べたように、数量化Ⅱ類は創成した段階から、再犯を起こしにくい仮釈放の条件を明確化し、再犯率減少させるという法務省社会課題解決に用いられた⁸⁾。

1956年には、朝日新聞社広告部で、林自身が設計した新聞広告の注目率調査が行われた。ビジネスの現場で、新聞広告の料金がスペースだけでなく、場所(何面に掲載されたか)などの質的屬性で、注目率への影響力が異なることを数量化Ⅰ類で示し、今日常識化した広告取引の料金設定に決定的な影響を与えた。

数量化Ⅰ類とⅡ類との融合により大きな変革があったのは選挙予測であり、林、高倉²²⁾は、Ⅰ類で得票に影響を与える要因を探索し、Ⅱ類で当選確率を推定するというのが、その後の統計モデル進化までは、選挙予測の一つの標準技法となっていた。

森本²⁾の表8には、数量化理論が我が国の社会課題解決にどのように寄与したかが年譜として挙げられているので、ここではその一部を簡条で示すが、これは氷山の一角にすぎない。

ここで示されたような活動が、その後国内の多くの類似活動として展開されたのである。その加速に繋がったのは、朝日新聞社広告部村山孝善との総括的かつマーケティング調査における多くの事例を含む著書¹³⁾

とされている。

- 1) 人間関係と作業効率・質の関係を明らかにし作業グループを再編成(1950 国鉄保線区員調査)
- 2) 医学分野における数量化による予後の予測, 症状による病名予測を通じた診断・治療方針への影響(1953~1954), この後, 今日 AI が担っているような自動診断技術として数多くの開発事例がある。
- 3) テレビ番組の特性と視聴率の関係を数量化(1963~1967)
- 4) 音の品質判定の一方法(1965 官能的特性の改善に関する検討, この後, 食品・飲料・化粧品等の設計特性と好みの関係に関して産業界で多くの新製品起案への適用が行われた)

5 主たる貢献者

数理化理論の開発は, 林知己夫によるものである。一方その普及啓発には朝日新聞社の村山孝善やNHK放送文化研究所の杉山明子らが, 社会心理学など学術分野に対しては飽戸弘らが寄与したと考えられる。また, その普及教育については統計数理研究所に1985年まで存在した附属統計技術員養成所が役割を果たした。数理化理論を実装した計算機ソフトウェアの開発には, 矢島敬二, 大隅昇, 駒沢勉らの貢献も大きい。また, 日本行動計量学会がその創世期に数理化理論の応用と啓発に果たした役割は極めて大きい。

6 価値の観点による考察

1. 先導性: 数理化理論は, 産官学の数量化の適用を前提とする新たな調査研究とその分析結果に基づく知的改善活動を通じて, 社会価値創出に対して先導的役割を果たした。
2. 規範力: 数理化理論は, 人間社会全般に共有する質的属性をどのように数量化し, データ分析の素材とするかという新たな観点を数理最適化技法に基づき導入したものであり, あらゆる分野のコトづくりにたいして標準的方法として適用された。
3. 意味力: 行動の計量科学という新たな学際的学術分野を扱う学会としての日本行動計量学会を形成し, 多くの賛同者からなる産官学を横断する実践的行動を形成することに成功した。
4. 解決力: 数理化理論はその生誕からして社会課題解決を目指す統計数理的方法論として構想されている。しかし, 数理化理論に基づく情報の収集と分析だけで課題が解決できるわけではない。一方, 質的属性が支配的な社会課題解決に際しては, 第一段階で課題の構造を可視化するために, 数理化理論適用を目的とした調査設計と情報収集は, 課題の解決に大きな役割を果たす。

参考文献

- 1) 林知己夫, 数量化と予測に関する根本概念, 統計数理研究所彙報, 7(1), 43-64 (1959)
- 2) 森本栄一, 戦後日本の統計学の発達—数理化理論

- の形成から定着へ, 行動計量学, 32(1), 45-67, (2005)
- 3) 飽戸弘, 数理化理論—社会行動研究における適用の効用と限界について, 日本社会心理学会編, 年報社会心理学, 5, 勁草書房, (1964)
- 4) 林知己夫, 社会計量における統計数理, 数学, 3(3), 172-185, (1951)
- 5) Hayashi, C., On the Prediction of Phenomena from Qualitative Data and Quantification of Qualitative Data from the Mathematico-Statistical Point of View, *Annals of the Institute of Statistical Mathematics*, 3(2), 78-93, (1952)
- 6) 林知己夫, 数理化理論とその応用例(II), 統計数理研究所彙報, 4(2), 19-30, (1956)
- 7) 林知己夫, 数量化: 理論と方法, 1/223, 朝倉書店, (1993)
- 8) 林知己夫, 假釋放豫測に於ける一つの科学的立場, 法学私林, 法政大学法学私林協会, 48(4), 56-71 (1950)
- 9) Likert, R., A Technique for the Measurement of Attitudes, *Archives of Psychology*, 140, 1-55 (1932)
- 10) Guttman, L., A basis for scaling qualitative data, *American Sociological Review*, 9, 139-150, (1944)
- 11) 林知己夫, 統計数理的数量化の問題—定性的(質的)なるものの数量化に就いての覚書, 統計数理研究所講究録, 6(1/3), 98-119 (1950)
- 12) 読み書き能力調査委員会, 日本人の読み書き能力, 東京大学出版部, (1951)
- 13) 林知己夫, 数理化理論のできるまで, オペレーションズ・リサーチ, 31(12), 728-734, (1986)
- 14) 日経リサーチ, 数量化3類, 調査・統計用語集, <https://www.nikkei-r.co.jp/glossary/id=6531>
- 15) 林知己夫, 村山孝喜, 市場調査の計画と実際, 1/323, 日刊工業新聞社, (1964)
- 16) 飽戸弘, 多次元解析より数理化理論そして多次元尺度解析へ, 行動計量学, 45(2), 75-84, (2018)
- 17) Benzécri, J. P., *L'Analyse des Données Tome 2: L'Analyse des Correspondances*, Dunod, Paris, (1973)
- 18) Nishisato, S., *Analysis of Categorical Data: Dual Scaling and Its Applications*, 1/292, University of Toronto Press, Toronto, (1980)
- 19) 椿広計, システム科学とデータ科学, 横幹, 14(1), 64-69, (2020)
- 20) Hayashi, C., What is Data Science? : Fundamental Concepts and a Heuristic Example, Hayashi, C., Yajima, K., Bock, H., Ohsumi, N. and Tanaka, Y. eds. *Data Science, Classification and Related Methods, Proceedings of 5th Conference of the International Federation of Classification Society (IFCS-96)*, 40-51, Springer, (1996)
- 21) Tukey, J. W., The Future of Data Analysis, *the Annals of Mathematical Statistics*, 33(1), 1-67, (1962)
- 22) 林知己夫, 高倉節子, 予測に関する実証的研究—選挙予測の方法論—, 統計数理研究所彙報, 12(1), 9-86, (1964)