量　対　数

2026-04-25　渋谷政昭　慶応義塾大学理工学部名誉教授

量は現実・具体、数は理想・抽象である

この命題は登場のモノにより正しいこともあるが、情況により深く考えることも必要となる。ここでは、量と数の関係を再考することの重要性を強調したい。哲学論争に加わる気はなく、統計科学・データ科学の役割を説明したいというのが動機である。もちろんそのためには、実例で示すべきであり、次の課題である。

数量

名前を持つ数：数量(Numerical Value）も抽象概念となる。
学校教育（小・中・高）では、 1,…,9 を用いた自然数から出発し。各方向に拡張した十進数系が（抽象的な）数であり、現実の場面で数の役割を示して数値の有効性を経験させ十進数系を拡張する。数量は理科で扱う。負の数は中学でも扱わない。
最初に数え上げにより自然数が導入されるが、その時にも集合・要素の概念が必要である。曖昧さを残しながら、話題・十進数系が広がる。

数としての量

計量器測定を確保するのに、かつてはメートル原器という世界でただひとつのモノがつかわれた。現在国際度量衡 mi(Metric Internationale　フランス語)は下記の 7 種の有限桁数を基本単位とし、他の単位は全て基本単位のべき乗と積で定めることができ、組立単位と呼ぶ。
基本単位の中で時間（周期、振動数）だけは測定できる。また数量とその変化速度は同一とみなせる。例えば、光の真空中の速度:　c = 299,792,458 m/s　これと時間で長さの基本単位 1m が 1s に光が通過する距離として定義される。数多い物理単位が 7 個の基本単位で結びついていることはニュートン物理学の奥深さを示し、数量は同じ単位での比較を表す数であることは、当然ながら日常の生活・事業の積み重ねであり、その精度を高めるために巨額が使われていることは研究・産業の力である。
放射能、天文・宇宙など新しい分野に関する mi 内外の単位系が作られている。
百、千、… ミリ、マイクロ、… 等の副名詞も mi に定義されている。材木一本 900kg 、小さじ一杯など、単位を変換することも多い。
「参考」:　基本単位 7 種
時間：Hz ヘルツ　セシウム 133 原子の振動の周波数 1/s

長さ : 光が真空中を 1 秒間に動く距離 m
熱力学温度 : プランク定数 h in(単位は)ジュール J
電気素量 e : in クーロン C
物質量 : ボルツマン定数 k in J/K (K は絶対温度ケルビン)
アボガドロ定数 N_A in 1/mol
光強度 : ある単色の放射・視感効果度 in lm/W (lm はルーメン、W はワット)
「問題」: 電磁気学の諸単位・基本単位の間の関係をしらべよ。
「問題」: 記号 % はどのように使われているか？使用が適切か？

量の現実、感覚
長さ：定規、巻き尺、重さ：ばねばかり、圧力はかり、容量：容量ケージ、温度：体温計、気温計、等、日常の計量は容易である。 観測・実験値の制度が十分であれば数値を数とみなす。研究者は数として理論・科学を研究する。生産でも数理論で計画する。日常生活では量と数の違いは無視され「二個先輩」も使われている、学校教育の影響もあるだろう。女子短大生が溶液混合問題を理解していないと聞く。
人類が・幼児が量を意識するのは、五感による。感覚を広げ・正確にするために、前章のような努力が実行される。逆に感覚で認識できないと、理解したと確信できない。数学者の感覚は別で、自分の理解している抽象世界を独自に展開している。

自然言語、マルティメディア
郵便番号の自動読み取りと番号による分類、POS(Point Of Sales) による販売管理などパターン認識として諸分野で開発された。
印刷・手書き文字の入力、音声入力は CP 利用に便利であるため、多くの努力が払われ、少しずつ前進した。それと共に言葉の意味を処理することも始まった。自動発生では意味を含むイントネーションにより、聞きやすさが向上した。
数値以外、画像・言語・音楽・動画などの総称が Multimedia である。CT, MRI 等の医療機器もこれに含めて良いだろう。

因果法則 Causality、演繹と帰納
原因 A から対応する結果 B を生じるか否か。登場するモノにより、様相が異なる。
A が現実・主観・具体的、 B が理論・客観・抽象的、ならば演繹的 deductive、逆ならば帰納的 inductive、な推論 Inference である。
ある数の値が与えられたとき、他の数が定まるのが二つの数の関係（Relationship, Association）で、変数・関数、方程式・解であらわされる。汎用的な概念は写像 (Mapping) である。具体的な数値は有限桁であるため誤差を伴い、その評価は計算法の課題である。

重要な現代的課題は、現実的な変量を伴う場合で、以下で議論する。

確率変数・確率分布
変動する数の事象（Event）つまり値域（Domain）の部分集合、が与えられたときに、その確率が定まるとき変数を確率変数（Random Variable）、結果の確率を、その確率分布（Probability Distribution）と呼ぶ。サイコロを投げた時偶数が出る確率は 1/2, ピンキリは 1/3 のようなものである。二つの確率変数の一方の事象が他の確率分布に影響しないとき、その逆も成り立ち、両変数は独立（Independent）であるという。次々章参照。
[問題]: 種々の確率分布の定義を考え、その分布関数を求めよ。

確率論の構成、確率過程論
様々な事象に対して確率が合理的な値となるには確率をどのように定義するか、の基礎を AN Kolmogorov は Lebesgue 積分を用いて確立した。
他方 von Mieses は数の無限列で、その一部(有限・無限列)を取り出したとき意味ある確率変数ようなモノ(Mieses 列)を基礎とした。次節参照。
確率変数の Domain の中に時間・空間が含まれていると確率変数は無限列となる、これを確率過程（Random Process)と呼ぶ。

乱数・疑似乱数、モンテカルロ法・シミュレーション
0, … ,9 がデタラメに並んでいるのを、数桁に区切り並べたのが乱数表である。軍事・政治で暗号に使われたために闇の世界の話だった。CP によるネット上の通信のプライバシー保護のために乱数表が必要となり、これを計算で生成することが計算機科学の重要研究分野となった。理想的な（Mieses 列の部分系列のような ）乱数は不可能であり目標は出来るだけ近い疑似乱数で（Quasi-random Number）である。
現在は松本真・西村拓士のメルセンヌツイスター法 1998 の改良版が圧倒的に優れ、主要計算機言語の全てに組み込まれている。最近 MSExcell まで導入している。
区間（0,1）上の一様分布に従う疑似乱数から関数により任意の分布の確率変数を原理的には生成でき、その標本分布を希望の精度で求められる。これがモンテカルロ法である。任意の確率事象についても同様で、これがシミュレーション Simulation である。代表的なのがコンピューター実験 Computer Experiment である。データ同化では、シミュレーションのパラメータを実データにより補正するもので、地球科学、生態学などで利用されている。
[問題]: 確率分布に従う疑似乱数を生成し、結果を調べよ。

ブートストラップ
一確率変数の大きさ n の標本 x_1<…<x_n (分布関数が連続として）があるとする。これ

は n 個の数値上の等確率頻度である。標本の分布関数を経験分布関数 Empirical Distribution Function と呼ぶ。
今の
場合経験分布関数は各観測値で 1/n だけ上昇する階段である(分布関数が連続だから)。
この標本から n 個の度数を等確率で取り出す sampling : 取り出した度数を元に戻して with replacement 繰り返す（つまりデタラメに順列を変える）。これを m 回繰り返す。
新しい j 番目(1<=j<=m)の標本の経験分布関数は、x_1,…,x_ n のうちの、l_j(1<= l_j <= n）個の点のうちの 1 個の上で l_j_k/n: 0<= l_j_k <=l_j だけ上昇する階段である。
m 個の経験分布関数をまとめると、x_ k で(l_1_k＋…＋ l_m_k)/(m・n) だけ上昇する階段である。この結果は x_1<…<x_n だけに依存している。
この拡大した経験分布関数により、確率変数の平均・モーメントなどの計算・推定誤差を改善できる。実用的には確率変量にブートストラップを適用する。標本の情報にこれと独立な情報を加え、新しい型の標本を作ることになる。

統計モデル・確率モデル
因果関係で、確率変数から確率変数の予測を調べるのは確率論の研究である。確率変数から確率変量を予測するのはモンテカルロ法による確率論研究である。確率変量から確率変数を予測するには、原因・結果の確率変数を仮説として予想する。つまり確率モデルを前提とする。確率変量から確率変量を予想する数理統計学でも、仮説的統計モデルを前提とする。
モンテカルロ法成功の基礎となったのは中心極限定理である。確率変数の和が、変数を増すと正規分布に近づく。典型的には分散有限な独立同一分布に従う確率変数の平均値が正規分布に従う。この基本定理の一般化と、多くの場合に定理条件を近似的に満たすので、理論の保証を超えても利用されている。

統計科学・データ科学
現実・統計データ解析の第一歩は統計モデルの構築である。
そのために過去の経験・文献から目前のデータについて、種々の統計モデルを試みることになる。モデルの適合性 Fitness を判定することは難しい。指数性（指数分布の適合性）検定など単純な場合しかない。正規性検定について多くの研究があるが、満足できない。
重要なことは、現実データの理解にモデルが役立っているか、である。それには応用上重要なモデルの指標を図示して、その特徴をデータと比較する。そのための道具である、データの可視化 Visualization は、データ解析の汎用的手段である。図表は 2 次元平面への写像でしかないが、色を使った表現が可能となり、利用が広まっている。
統計モデルを構築しそれが応用目的に有効・適切かを解析・検討するのがデータ科学であり、統計科学の神髄である。

データ・モデルの特徴を見るには全ての二変量間の関係を可視化して分類する。強い関係のあるものはまとめる。観測・実験のために設定できる制御因子は適当に調節する。主要因子と独立で、制御不可能な誤差因子により、推測精度が落ちる。
このような操作を簡便化したのが、柴田里程の TRAD である。

AI
統計科学の諸場面で Generative AI は素晴らしい働きをするだろう。問題の明確化、データの収集、整理、モデルの構築、適合の検討、結果の整理、論文作成、等 AI の利用は急速に拡大しその成果が次の AI 利用に貢献するだろう。

[追記]: 田村二郎、量と数、1978 では、量が満たすべき公理、例えば長さに関する大小・加法・等分の公理、を構成しその変換により数を導いている。