

1 モデルランドとはどこのこと

自然な芸術作品なるものを見たことのある人がいるのかどうか知りたいものだ。自然と芸術はそれぞれ別のものであり、同じものではない。芸術を通して、われわれは自然とは異なるものの概念を表現する。

——パブロ・ピカソ「ピカソ語る」（一九二三年）

この世界はデータであふれている。人工衛星が絶えず上空から地球を監視して、アマゾンの熱帯雨林の破壊について、インド洋でのサイクロンの発生について、極地の氷床の厚さについてデータを送ってくる。ネット上での人間の行動は追跡アルゴリズムに常時監視され、人がこのウェブサイトでどれだけの時間を使ったか、どういう広告をクリックしやすいか、どういうものを購入したかが、じつに精密に捕捉されている。オリンピックに出るようなアスリートも趣味でジョギングをしている人も、毎日の運動量や体の状態をつねに計測して記録に残している。金融市場からは、毎分毎秒レベルで収集される「高頻度」価格データがとてつもない量で生みだされている。

これらはみな、現実世界リアルワールドから直接観測されることだ。だが、そうして得られたデータを解釈するた

めの枠組みがなかったら、データは意味のない数字の羅列で終わってしまうだろう。実際にデータ解釈のために使われている枠組みにはさまざまな種類があり、私はそのすべてをモデルと呼びたいと思う。モデルは純粹に概念的なものでもよく、趣味で走っている人にとっては「心拍数が上がったら、ふだんより激しく運動している」というのが一つの自分なりのモデルかもしれない。しかしデータを活用することで、概念的なモデルにとどまらない、ある変数と別の変数との関係を示すことをめざした定量的な（数値や数量で表す）モデルを作ることできる。たとえば「先週この製品をクリックして自転車用空気入れの購入候補リストに加えた人のうち、一〇パーセントがその後これを購入するにいった」というふうに。こうしたモデルなら、データ間の関係が見えてくる。これは見えている範囲での関係であり、概してそこにはなんらかの不確実性（不確かさ）が含まれる。このモデルだけでは顧客のどの一〇パーセントがその空気入れを買うのかわからない。さらに見えている範囲での関係は、見る側の都合にも影響される。その自転車用空気入れをクリックした人のうち何人がその後にあるアイロン台を購入したか、はたまたトマト缶を購入したかは検証していなくとも、その自転車用空気入れをクリックした人のうち何人がその後にライバル会社の自転車用空気入れを購入したか、はたまたその人たちが別のスポーツ用品も購入したかどうかなら、検証したくもなるかもしれない。

そもそも、それはなんのためにやるのだろう。この種の数字それ自体に関心がある人はまずいない。売り手が本当に知りたいのは、その自転車用空気入れが購入される見込みをもっと上げることができるとか、あるいは同じサイトで顧客にもっとお金を使ってもらうため、顧客の気を引く「関連商品」としてほかのアイテムの広告を出せるのかどうかだ。気象衛星の打ち上げに出資している機関

が知りたいのは、天気と気候をどうしたらもっとうまく理解し、予測して、不都合な影響の回避につなげられるかということだ。アスリートが知りたいのは、どうしたら自分の成績を上げられるかどうか、金融トレーダーが知りたいのは、どうしたら自分が勝てるかである。

二〇二〇年初頭に新型コロナウイルス感染症のパンデミックが発生し、世界中がその第一波に呑み込まれたとき、各国政府はさすがのような思いで疫学や数理モデルの専門家に助言を求めた。このあと事態はどうなるのか、どうすれば有効に対処できるのか。情報は乏しかったが、それでもただちに手を打たねばならなかった。新しいウイルスの特徴については不確実性があり、人や社会が政府による多大な介入にどう反応するかについても不確実性があり、さらに疫学面、経済面でどんな長期的な影響があるかについては並々ならぬ不確実性があった。

数理モデルの専門家はそうした状況に定量的なシナリオを持ち込んだ。つまり、ある特定の仮定のもとではどんな未来がやってくると考えられるかを提示したということだ。とくにこれらのモデルでは、重症患者の割合が一定の予想範囲にあつて、かつ何も介入がなされなかった場合には、たいへんな数の人にいつせいに入院措置が必要になるだろうとの見込みが示された。ある一人から別の一人以上に病気がうつるさまを表した最も単純な感染伝播モデルは、ある集団での指数関数的な感染増といったスケールの大きい結論を出すのにも、接触機会の削減による個人リスクの軽減といったスケールの小さい結論を出すのにも使える枠組みだ。もっと複雑なモデルなら、社会のさまざまな部門ごとに、現実に基づいた接触パターンの違いを長期的な視点で表現して、新しい情報の効果、よりよい治療の効果、ワクチン接種計画の効果などを説明することも可能になる。犠牲者数のような痛ましい指標もモ

デルがあつてこそ推定できる。そうすることで、さまざまな種類の介入の効果を予測することができるようになる。そしてそれはスケールの大きな政策決定にとって明らかに重大な関心事だ。だからこそ疫学モデルはこの数年、ずっと国の政策決定の際の参考にされ、一般メディアのトップ記事で報じられてきたのである。

要するに、データ間の関係をモデル化する目的は、なんらかの全体的な目標を達成するための一助として、この先どうしたらもっと有効な手が打てるかを予測しようとすることだ。モデルに表されるものは現実世界のデータにもとづく現実世界の問題である。したがってモデル上で答えが得られるとすれば、それは現実世界に通用する答えである。この答えを見つけるために、私たちはモデルランド（模型の国）に出かけなくてはならない。

モデルランドに入ったら

現実世界とモデルランドのあいだに国境検問所はない。いつでも好きなときに越境できるし、安楽椅子にどっかり座ったままでもモデルを作りさえすれば入国できる。ふわふわした抽象的な概念モデルでもかまわないし、地に足のついたスプレッドシートでも、足どりの重い巨人のごとく複雑な数値モデルでもいい。いったんモデルランドに入ったら、そのモデルで仮定したことは文字どおり真実である。そのモデルがモデルランドでの現実なのだ。

もちろんモデルランドの内部にはさまざまな領域がある。同じ状況に対してであつても多種多様な

モデルを作ることができるから、それぞれのモデルに対応した別々の領域があるということだ。モデルランドのある地域では、この地球は平坦である。したがってクリケットの球の軌道を計算するモデルに地球の半径を含めなくていい。モデルランドの別の地域では、地球が完全な球体である。この場合、たいていの地理的な目的を満たすには十分な正確さで国や大陸の相対的な位置がわかる。モデルランドのこれら以外のところでは、地球は扁球だと見なされている。つまり極方向の直径より赤道方向の直径のほうが約四三キロメートル長い。GPS（全地球測位システム）の構築や重力場の測定といった複雑な応用をする場合なら、山や谷、高原や海溝についても知る必要があるだろう。これらはいずれも同じ地球のモデルだが、それぞれ目的とされている応用のしかたが違うのである。

モデルランドは素敵なところだ。どんな仮定であつてもモデルランドではそこそが真実だから、自分の作ったモデルをもっと深く理解できるし、そのモデルがどう機能するかをもっと詳しく知ることがもできる。予測も立てられる。モデルの構成をさまざまに変えてみたり、入力を変えて実行してみたりして、異なる状況ごとにどういう結果が出るかを見ることがもできる。モデルの何かしらの特徴を変えてみて、結果に表れる差の大小を確認することもできる。そしてモデルを作るときに用いた統計的手法の根底にある仮定もモデルランドでは真実なので、何も気にせず統計分析ができる。

とはいえ、モデルランドなら人生楽勝というわけでもない。モデルを作って「計算」しなければならぬのはここでも同じだ。モデルランドで難しいことを夢中でやっているうちに仕事人生がまるまる費やされてしまうことだってあるだろう。

モデルランドは必ずしも決定論的な——何もかもがわかるとされるような——ところではない。モ

デルランドにおいてもモデルの結果にはやはり不確実性が残りうる。その不確実性は、量や初期条件を完璧に測ることはできないところに原因があるのかもしれないし、さもなければモデルのパラメーターを定義するのが難しいとか、モデルそのものにランダムに決まる要素があるとか、あるいは「バタフライ効果」の名で知られるように、モデルの軌道がカオス的にずれていくといった理由で生じるのかもしれない。

モデルランドでは、こうした不確実性を定量化できる。少なくとも性能十分な大型コンピュータさえあれば、原理的には定量化ができる。それより小さなコンピュータでも不確実性の推定ならば可能だし、さらに詳しい情報が得られれば、それらの推定が一つの正しい答えに収束するだろう。

一方、モデルランドと現実世界の狭間には、定量化できない不確実性がある。たとえば、このモデルは構造的に正しいのだろうか。関連する変数をすべて漏れなく考慮したのだろうか。系を測定し、予測することにより、その系に影響を与えてしまっていないだろうか。

たいていの分野は、この二種類の不確実性を区別している。経済学では、定量化できる不確実性は一般に「リスク」と呼ばれ、定量化できない不確実性が「不確実性」と呼ばれる。物理学では、まぎらわしいことに正反対の用語が使われることがあり、不確実性と言えば通常は定量化できるもののことを指す。定量化できない不確実性については、深い不確実性、根本的不確実性、認識論的不確実性といった呼び方をする。

試しにしばらくモデルランドで過ごしてみよう。ここにビリヤード台があるとして、球を一個突いたとき、その球の動きがどうなるかを予測することにどんな不確実性がつきまとうかを考えてみる。

原理的に、これはエネルギーと摩擦と直線上での減速度からなる単純な計算の問題だ。しかし実際には、モデルランドにおいてさえ、一つの計算だけでは限界があることを認めざるを得ない。まず、測定上の不確実性（測定の不確かさ）がある。球が動きだす出発点を正確に定めることはできないからだ。それから初期条件の不確実性もある。球を突くときにどれだけのエネルギーが球に与えられるかを正確に定めることはできないからだ。また、境界条件の不確実性もある。球がへりに当たって跳ね返るときにどれだけのエネルギーを失うかを正確に定めることはできないからだ。そしてこれらの不確実性が合わさって、球が静止するときの正確な位置についてランダムな不確実性が生じる。原理的には、これらの定量化可能な未知数それぞれについて妥当と思われる条件の範囲を定めることはできるし、その条件を用いて妥当と思われる球の静止位置の範囲を計算することもできる。運がよければ、実際に実験を試みたときに、球はその範囲の中央近くのどこかで止まっているだろう。物理学者はこの妥当と思われる範囲のことを、計算の「エラーバー（誤差範囲）」と呼ぶ。統計学では「信頼区間」と呼ばれることもある。

深い不確実性、または根本的不確実性は、定量化できない未知数のかたちでこの構図に入ってくる。これらは予測できたはずのないものとして、計算から除外されている。たとえばビリヤード台がじつは平面でなかったのかもしれない。対戦相手が球に息を吹きかけて軌道をずらしていたのかもしれない。球がへりに当たって跳ね返る代わりにポケットに落ちてしまったのかもしれない。その場合、慎重に定めた統計的な予測結果範囲もまったく役に立たなくなってしまう。このような不確実性は、「知識」を意味するギリシャ語のエピステメー (episteme) に由来して、認識論的不確実性、また

は知識の不確かさ (epistemic uncertainty) とも呼ばれる。ここには私たちの知識の限界と予測能力の限界が反映されている。だが、これらはまったくどうしようもない問題というわけでもない。こうした不確実性をリストアップして、よく調べ、それが発生する可能性を定量的に (数値・数量の面から)、定性的に (数値・数量で表せない性質の面から) 評価することはできるし、不確実性が発生しないよう現実世界で対処しておくことだってできる。もちろん、本当に予想外の事象——いわゆる「ブラックスワン」——もありうるので、それに不意打ちを食らうこともないではない。しかし意外な事象の大半は、じつはブラックスワンではない。二〇〇八年の金融危機の到来も、一部の人は見越していた。感染症のパンデミックも、ずっと前から国家的な注意事項の最上位に挙がっていた。気候上の危機的な変化を見逃さないよう、科学調査研究の本流がつねに注視しているのは言うまでもない。私たちはこの種の事象が起こりえないと思っているのではない。ただ、そうした事象が起こりうることはわかっている、その理解を明確なたちにまとめあげて有効に対処する方法をまだ確立できていないのである。定量化できない不確実性は重要で、いたるところに存在するが、潜在的には把握できるもので、意思決定の際にも考慮すべきものだ。ただし、この不確実性をうまく使うには、私たちがモデルの限界を理解し、モデルと政治との関係も認識したうえで、モデルランドを脱出して、現実世界に即した予測を述べられるようにしていく必要がある。これが本書の前提の一つである。

リスクと不確実性の問題以外にも、モデルランドでの快適な人生を脅かすものはある。科学者はモデルのことを科学的なファクトとデータと可能性によって構成される、あくまでも科学的なものと思っていたが、モデルが倫理や政治や社会的価値観とどれほど避けがたく絡みあっているかは、もはや

明白になってきている。モデルにはリスクと責任に関するナラティブが結びついている。モデルに表現されていることにも、いないことにも倫理的判断が内在する。モデルは社会的な説得の道具でもあり、政治的な議論を闘わせる手段でもある。

いい例が、新型コロナウイルスのパンデミックのモデルだ。どういう人が家にこもっているべきか、マスクをつけるべきか、ワクチンを打つべきか、その他たくさん疑問にどう答えるべきかを考えるとき、モデルは決して科学だけの問題ではないことがよくわかる。私の以前の研究対象で、私にとつてはもっとなじみ深い、気候変動についても同じことが言える。ここでもやはり科学そのものは、問題全体の一部でしかない。地球の気候が二十世紀の気象パターンから少しずつ、しかし明らかにずれていっている以上、もはや温室効果——ウイルスのヒート・ヒト感染と同様の、残念ながらまぎれもない事実——に議論の余地はない。しかしながら、それに対して何をなすべきかという恐ろしく複雑な政治的問題はいまだ解決されずに残っており、これまた複雑な数理モデルを介して検討される。物理気候学のモデルと経済体制のモデルの両方が、気候変動に関してどういう目標をめざすべきなのか、対策をとる場合のコストととらない場合のコストのあいだにどういうトレードオフがあるのかを考える際の参考にされる。新型コロナウイルスの場合と同様に、どういう種類の損失と損害なら許容できるのか、さまざまな地理的共同体のそれぞれがリスクをどうかぶるのか、なんらかの対処をすることのコストを誰が引き受けるのかといった問題についての価値判断を通じて、政治がモデルに入り込んでくる。気候の問題にせよ新型コロナウイルスの問題にせよ、予防原則はあまり役に立たない。新たに出てきたリスクに用心することはできる。いままぐ被りそうな経済的打撃に用心することもできる。しかし本当に被

害がわかるのはそのあとなのだ。いきなり現れた新型コロナウイルスへの対応と、徐々に現れた気候変動リスクへの対応とのあいだには多くの類似点があり、そこから導ける結論はいくつもあるだろう。しかし何よりも先にやるべきは、科学的な数理モデルが社会とどう相互作用して、私たちの意思決定構造にどう情報を与えているかをひとつよく把握することだ。

「データだけ」を扱っているからといってモデルランドを避けて通れるわけではない。データとは測定された量のことであり、データそれ自体は何も語らない。データはモデルによって与えられる文脈と枠組みを介したときに初めて意味をなす。同様に、純粹に概念的なモデルを研究し、実際のデータを参照することなくサイコロ理論や弦理論を構築している場合でも、やはりモデルランドの回避はできない。とはいえ、正確なデータと正確な概念理解には別の利点がある。これらはモデルランドから脱出し、モデルで出した結果をふたたび現実世界に結びつけるのに役に立つ。

モデルランドから脱出する

モデルランドに入るのは簡単だが、モデルランドを出るのはそう簡単ではない。内的に整合性のある美しいモデルと、そのモデルを詳細に説明する一連の分析手法を構築してしまうと、その全体の基盤となっている最初の仮定が文字どおりの真実ではないことを認めるのが感情的に難しくなる。だからこんなにもたくさんさんのモデルに関するレポートや学術論文が、その最初の仮定を立てたまま忘れてしまっているか、あるいは通り一遍の方法でしか検証していないのだ。モデルから出力された図を

現実に観測された像の隣に並べ、両者はあらゆる重要な点でそっくりだと述べるのは、信念の表明であり、モデルの成績に対する評価ではない。そのモデルにもとづいて何を推論しようと、それは小さいモデルランドの域を出ていない。

本書の一つの目的は、もっと批判的にモデル評価のことを考えようとモデル作成者——私たちは誰もがモデル作成者だ——に呼びかけることである。自分のモデルは現実世界で何が起こりうるかを教えられるだけの十分な予測力を持ったツールであると、本当にそう言いきれぬのか。この問いに対する敗北主義的な答えは、そんなことはわかるわけがないというものだ。それはそうかもしれないが、科学と工学の歴史には、とんでもなく成功している数理モデリングの例がたつぷりとあり、私たちはそれに当たり前のように頼っていて、それに頼って失敗したこともない。私の考えでは、モデリングが今後もそのように成功していくかは、モデルを正しく理解させるための教育プログラムを用意することにかかっている。モデルはツールとしても思考とコミュニケーションの指針としても使えることを示しつつ、モデルのそうした役割自体の限界も認めてごまかさないことが肝要だろう。

したがって第一になすべきことは、モデルランドからの出口を理解して、それにもっと明確な標識を与えることだ。これを本書の次章以降でやっていきたい。端的に言えば、モデルランドからの出口は二つある。一つは定量的な、もう一つは定性的な出口である。定量的な出口から脱出するには、モデルをサンプル外データ、すなわちモデルの構築に使われなかったデータと照らし合わせればよい。ただ、これは準備するのが意外と難しい。サンプル外データは手に入れにくいし、そうでなくても集めるのに時間がかかる。しかし天気予報など、時間尺度が短い予測装置の場合には、この定量的出