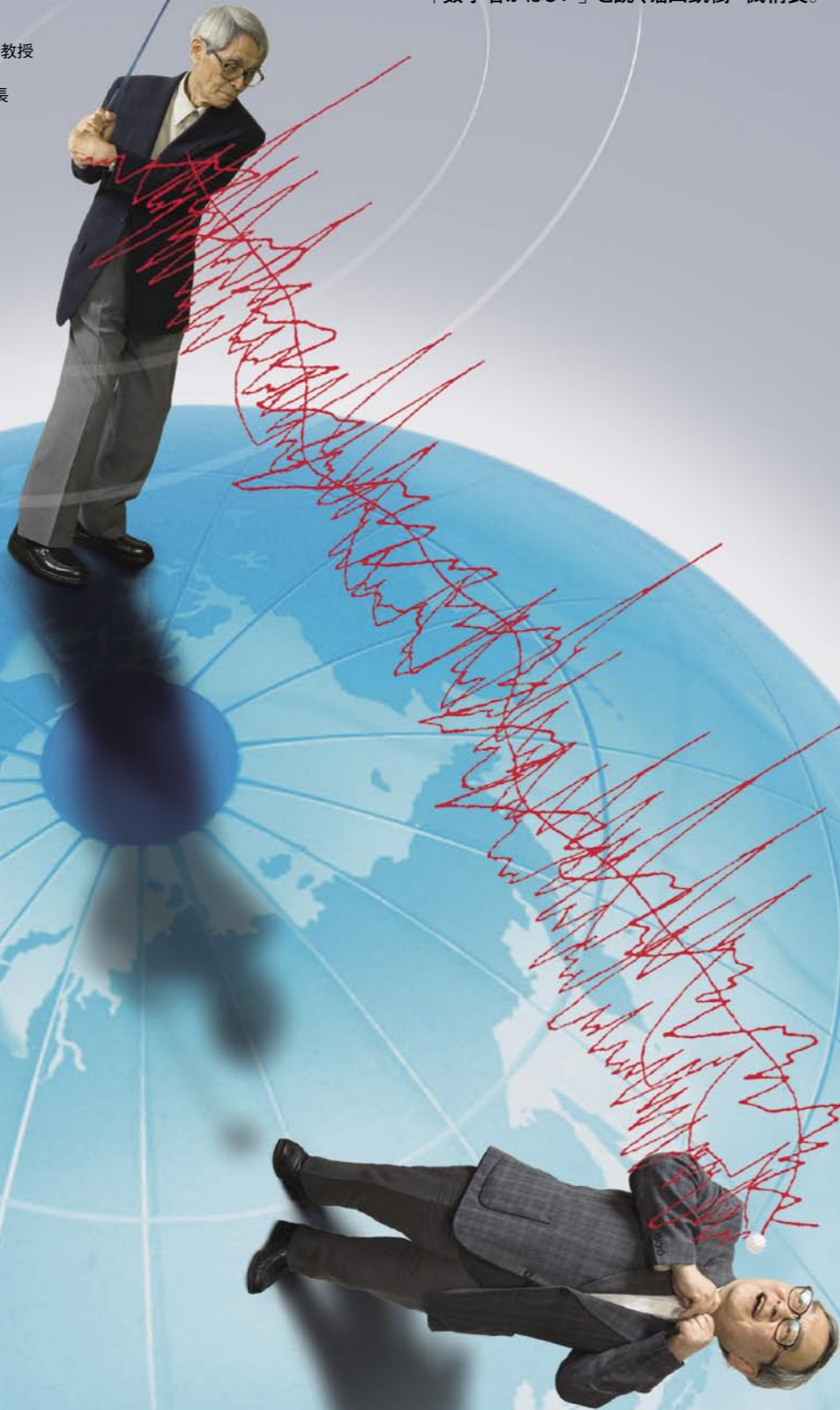


科学の目・ 統計学の目

話し手 赤池弘次 総合研究大学院大学名誉教授
聞き手 堀田凱樹 情報・システム研究機構長

赤池統計学を形づくる根源的な思考はどのようにして生まれたのだろう。現にあるものを数理で解くのが得意だった少年が、統計学におけるモデル評価の開拓者となるまでの興味深いストーリー。「数学者としては少し異端児」の話し手に対し、聞き手は、ゲノム解読以降データの洪水の中にある今の生物学には論理的思考が必須であり、「数学者がほしい」と説く堀田凱樹・機構長。



堀田 今日、先生がどうして統計学に興味を持たれるようになったかなどについてうかがい、若い人が将来を考えたときの参考になれば、と考えています。先生はお若いころから、数学や統計学に関心をお持ちだったのですか。

得意なのは昔から応用問題

赤池 私には海軍の戦闘機乗りの叔父がいて、子供のころ、彼が数学を勉強する姿を見て、おもしろそうと思ったのです。学校の算術は、割り算はこうするというように機械的に教えるので、なじみませんでしたね。応用問題が得意で、機械的に覚えることは不得手だったのです。

堀田 実世界に即して考えるのが得意だったわけですね。

赤池 昔からそうでした。

堀田 海軍兵学校に進まれたわけですが、海兵は理工系の教育が中心だったのですか。

赤池 将校になったときにどんな仕事にも対応できるような教育を、短期間にやるのです。数カ月で微積分を終え、すぐに航空力学に近いことや魚雷を撃つ話にも応用する。熱力学ならエンジンの話になる。戦後、あらためて入学した高等学校では、何年もかけて微積分を教えていましたが、海軍兵学校では、全部の教科を対応させながら非常に速く進む。たいへん合理的で、効果的でしたね。

堀田 戦争中ですから、すぐに戦地に赴くこともあり得たのですね。

赤池 いつまで生きていられるのかな、と思ったものです。最上級生のときには、広島に落とされた原爆の閃光と爆風を経験しました。叔父の微分方程式の本を読んでいると心が休まりましたね。

堀田 そこで敗戦。価値観が大きく変わり、精神的にも困難な時代だったのではないのでしょうか。

赤池 目の前で価値観が音を立てて変わるような経験でしたから、敗戦後の数カ月は、これからどう生きたらよいかと考えました。自分と他人の生命を基本にして、それを互いに尊重することを道徳の根本にすれば、どんな状況でも生きてい

けるだろう。そう思い至って、ようやく心の平安を取り戻したのです。

堀田 その後、第一高等学校理科から東大数学科に入られた。応用問題が得意だった先生ですから、工学に興味を持たれても不思議はないと思うのですが、数学科を選ばれたのはなぜですか。

赤池 高等学校の間に、考え方が抽象的になっていったのでしょうか。海兵時代に熱力学などを聞きましたから、まず物理に関心が行き、その基礎として数学が大事だと感じました。

数学科では、きれいに完成した解析学や代数学には、自分はあまり適さない気がしました。具体的な問題に遭遇して、手探りで考えることは好きでしたが、できあがったものを覚えるのは相変わらず苦手でした。3年のセミナーでとった確率論の先生から、文部省統計数理研究所を勧められたのです。

大ざっぱでも現実に役立つモデルを

堀田 私は学生時代に教養科目のなかでいちばんおもしろかったのが統計学でした。統計学というのは、複雑なものを単純なモデルで考えて結論を出す方法で、それは脳がやっていることと同じだと思ったのです。その後、私の基本的な興味は脳に向かったのですが、その根っ子はこのあたりにあるように思います。

赤池 それは実に正しい感覚ですね。今、統計学をやっている人たちは、そういうセンスに乏しいかもしれない。最初から既成の統計学をやってしまいますから。

堀田 そのころ習った統計学では、まずモデルを作り、それに当てはまるかどうかを考える。成り立たなければ、仮説を否定する。モデルといっても、たいていは正規分布など単純なものばかり。それを少しも不思議とは思わなかったのですが、先生のご研究は、モデルの良さを客観的に測ることをめざしていると考えてよろしいのでしょうか。

赤池 そうですね。数理的に美しい理論を作ることにはもちろん関心がありましたが、実際の問題では、大ざっぱなモデルで十分に役に立つ領域がたくさんあ

る。扱いやすく、しかも複雑な現象を処理できなくてはならないのです。既成の統計理論から正確な議論をしても、適用する対象の特性をうまくとらえない限り、役に立ちませんから。

具体的な問題で成果が上がる方法を追求しようとする、自然にいろいろな応用分野の人たちとつながりができてきます。10年ほどは、日本の社会に実際に存在する問題で、統計的な研究を必要とするものは何かを確定することにエネルギーを費やしました。

工業の現場からの求めに答えて

赤池 例えば、当時の重要な産業だった生糸の生産工程管理です。工程の異常を検出する管理図法を導入しようということになり、当時の農林省蚕糸試験所の方が非常に悩んでおられた。

生糸は、一定の個数の繭から引く糸を撚り合わせて1本の糸に紡ぎます。そのとき、糸が切れると繭が落ち、その切れ目を新しい繭でつないでいきます。その回数が異常に増えれば、これは工程の異常ということになるわけです。私は、その前に自動車の流れのモデリングをやっていたので、自動車の流れと生糸の切れ目の流れは同じように扱えると思いました。単位時間に来る糸の切れ目の分布を見ると、典型的なランダムな流れの特徴と考えられるポアソン分布が現れるのです。他方、試験線糸の段階で、1本の繊維の長さの統計的な分布のデータが得られます。ランダムな流れの場合とは違う長さの分布がちゃんと出てくるのです。この結果から、実際の線糸のプロセスでの切れ目の現れ方が確率論的に決まります。これを利用する線糸工程の統計的管理が導入されて、生糸の生産工程に大きい成果が出るようになりました。

堀田 私が習った統計学では、モデルの作り方はいたって拘り定規でしたね。

赤池 誰かが作ったモデルを使うのでしょうか？

堀田 無理に合わせる。そういうものかと思っていましたが、先生のお話をうかがうと、むしろ現実にあるものを上手に



堀田 凱樹 (ほった・よしき)

1938年東京生まれ。医学部に進むが、科学的であろうとするとよい臨床医にはなれないと観念し、江橋筋門下で平滑筋の電気生理学を研究。一方、脳に強い興味を持ち続け、カリフォルニア工科大学でショウジョウバエを対象に脳の遺伝学を手がける。新設の東大理学部物理学科生物物理学グループに加わり、物理学科で遺伝子実験を始めた。1997年国立遺伝学研究所長。2004年情報・システム研究機構長。複雑なものをそのまま理解する、生命科学の新たな方法論の必要を痛感している。

単純化してみると、わりに簡単で現実に対応したモデルが出てくるのですね。

赤池 生糸のあとは、自動車のサスペンションやエンジンの振動の解析、海の波の動きと船の動揺の関係などを対象にしました。周波数成分に分ける解析法を使って、長時間繰り返して実験しないで、さっと推定できるようになったのです。

ところが、セメントの炉を自動運転するときの最適制御というテーマで、この周波数解析の方法が使えないところが出てきました。なぜかという、セメント炉の場合には、温度が上がればそれを下げようとして燃料を調節するとか、状況の変化に応じてフィードバックがかかります。そのために、今までの方法は役に立たない。では、どうするか。関連する要素を全部入れて予測し、その予測のモデルを作れば、それを利用して問題が処理できることに気づきました。

モデルを作ること・評価すること

堀田 先生のお仕事では予測がキーワードですね。昔の統計学で習ったのは予測ではなかった。先生のお考えでは、予測できるようなモデルを作ることが大事だということですね。

赤池 そうなのです。自動車の振動の特性を周波数スペクトルで測る問題でも、時系列の予測式を作れば、スペクトルも決まってくる。いろいろな測定に使えるわけです。使ったのは、過去の何時点かの値に係数を掛けて加えて現在の値を予

測する、という単純な線形モデルです。誤差自乗の和が最小になるように係数を決めることで処理できます。ところが予測式に何時点前までのデータを取り込めばよいかという、モデルの次数決定の問題が出てきます。次数が高すぎると推定精度が落ち、低すぎると予測力がない。検定論を使っても、次数をいくらにするかという検定はできないのです。

どうするか。同じ予測式を同じ構造の別のデータに使って検証するのが正しいわけで、私は「ファイナル・プレディクション・エラー」と名付けた評価式を作りました。使ってみると、8~9割はうまくいくのです。理論家は100%よくないと納得しませんが、現実にはこれで十分。

ところが、セメント炉に適用すると、観測値が多くて当時のコンピューターでは処理できないのです。そこで変数をぎりぎりに絞って処理する。しかし多変数の場合は、予測誤差は変数ごとに違うわけです。では、それを総合して何で評価すればよいか。予測誤差のベクトルの特性を評価するものがどうしても必要になる。それを一変数の場合の予測誤差の評価式を拡張したようなものを使ってみて、うまくいったのです。

そのうちに、一般の多変量解析の誤差のことが気になり出しました。これは、心理学でよく使う因子分析法に出てきます。因子分析では、分布は正規分布を想定していますが、その評価をある量で処理している。それは対数尤度です。確率

は、過去から将来のデータの見方を与えますが、尤度は、現在のデータを用いて、過去にこれを生みだした仕組みを評価しようとするのです。

因子分析では何を予測しているか。心理学的な調査データに基づいて、そこからどういう特徴的な因子があるかを書き出していく。因子分析は構造を見ているのです。そのうちに気がついたのが、これもモデルを利用しているということでした。モデルを決めることは、そのモデルを使って将来の問題を処理しようとしている。すなわち予測していることです。その意味でモデルを評価すればよいのだということに、朝、井の頭線の車中で気がついたのです。尤度を使えばよい。モデルの与える分布で予測していると思えば、尤度で対応する処理ができるだろう、と。**堀田** そういう形で、モデルの良さが評価できるということですね。でも、自分で考えたモデルがどの程度良いかが評価できて、より良いモデルが自動的に出てくるわけではないですね。

赤池 そうなのです。本当のモデルがわからないのに、なぜ良さを評価できるのかという基本的な問題がある。哲学的な大問題です。モデルの評価には、真の分布の対数尤度とモデルの対数尤度の差(3ページの図参照)を使います。何種類かのモデルがある場合、真の分布がわからなくても、データから決まるモデルの尤度はありますから、その対数を比較すれば、モデルの比較はできるのです。

堀田 大論争からその結論が生まれたわけですか。

赤池 論争はありませんでした。まだ皆、よくわかっていなかったのです。AICでモデルを評価できて、モデルそのものを提案してくれる人がいなければ、なんの意味もありません。

堀田 モデルはAICから出てくるものではなく、現実にある状態から、人間が考えてモデルを作る。モデルをいくつか作ったときに、予測の立場からどれが良いかは、AICで評価できるのですね。

赤池 そうです。私が今、いちばん大事にしているのは、真のモデルに肉薄する

ようなモデルをいかにして提案するかです。仮説を真実に近づけるにはどうしたらよいか。それにはまだ方法論がないのですね。しかしAICは、あまり複雑なモデルを作ってもよくないことは示します。これは、モデリングの本質的なところ

です。そこで、具体的に何がいちばん統計的かということ、言葉で表現するという事なので。言葉というのは厳密にこれを示すというのではなく、その言葉で話が通じるという、その程度のもので。非常に複雑なものから本質的な特徴をまとめて表現するという事です。つまり、ものを書くにも話すにも、統計的な努力をすることになりますね。

堀田 なるほど、そうですね。脳というのは結局、複雑な事象をモデル化して単純な形に整理し、それを言葉にしている。

赤池 しかも、他人にもわかるように。だから、意味の伝達可能性、つまり言ったことが人に伝わるのが非常に重要なわけです。統計的な概念も多分にそうですね。いろいろな検証を通じて、しだいに客観的と称するものになっていく。科学的思考というのはそういうことなので。統計的な概念は、あらゆる科学研究の基礎的素養だと思います。既知の式を組み合わせれば科学的な成果が得られると考えるのは、甚だしい誤解です。

堀田 新しい理論が次々と出てくる人生というのは、楽しいですね。

ゴルフスイングとセレンディピティ

赤池 成功したり、しくじったり、ですよ。おもしろいのはゴルフスイングの解析です。ゴルフをするときの体の動きの複雑怪奇なこと、セメント炉の比ではありませんから。人間の体には200以上の骨とそれをつなぐ筋肉がある。それなら解剖学の知識があればいいショットができるかという、できはしない。ゴルフの本も役に立たない。結局は自分で、機械的なモデルをイメージして体を動かしてみるほかないわけです。

今までの物理的なモデリングは一見理論的にやっていたましたが、たいいてい結果

が実的に解釈できないのです。なぜそれをやるとよい動きになるのかにつながらない。計測するだけなのです。私はそういうモデルを「計測モデル」と言いますが、計測モデルには構造的な内容がない。計測だけでは理解にはつながらないのです。そういう意味で、それまでの理論的な解析はあまりよくないと思っています。ときに、肺炎で3週間近く入院し、そのあとヘルペスで体中が痛くなり、ベッドに縛り付けられてしまいました。

その間、ゴルフについて考えたのです。動けないから、寝たまま手を動かす。左手は右に行き、左に引っ張ってくる。左手のスイングの間、右手は上下に動くだけ。右手は上下で、左手は左右。左右対称ではないのです。これを「革命的イメージ」と名付けました。このイメージで打つとけっこう打てる。今自分は、昔よりしっかり打てますよ。ゴルファーは実は地球にぶら下がっている、と気がついたからです。足が地球をつかんで、手はクラブをつかみ、クラブという小さい地球と大きい地球をつないでいるわけです。

ゴルフの革命的イメージにどうして到達したかという、ひたすら試みを繰り返したからです。しつこくやり抜くうちに、オーソドックスな視点の影響が消えて、今まで気づかなかったものが見えてくる。それがセレンディピティです。あらゆる可能性を繰り返し追いかけているから、あるとき、ハッと気がつくのです。研究者というのは、知識で頭の中がいっぱい

ですから、変わったものは見えない。それを打ち破るにはすごい努力が必要です。**堀田** 先生のお考えはロボットの設計にも使えそうですね。

生物学における良いモデルとは

堀田 モデリングというキーワードから思うのは、生物学の現状です。今は、1つ1つの遺伝子を操作できるようになり、ある遺伝子がどこで発現するかを細かく実験できる時代です。そこで、この遺伝子をここで働かせたら何が起るかを予測できる分子機構モデルを作ります。ある実験から可能なモデルはたくさんあって、どれが正しいかが問題です。生物学の場合には、科学的な真実が存在しますから、どのモデルがそれに近いか、どれがいちばん良く次の実験結果を予測するかなど、いろいろな意味でモデルの良さというものがあると思うのです。

また、どんな種類のモデルがいくつあるかを考えることが大事だと思うんです。たまたま思いついたモデルを報告すると、それが真実のように思われますが、全然違うモデルが実は別にあり、それも同じような予測性能を持っていることがある。そのときには、そういうモデルがいくつ残されているかを考える必要があります。そういうところに、先生が開発してこられた戦略を展開できないでしょう

か。**赤池** 観測値との対応の問題があれば利用できるでしょう。多くのモデルがある



赤池 弘次 (あかいけ・ひろつぐ)

1927年静岡県生まれ。子どものころから物の動きやしくみに興味をひかれ、統計学の研究においても時間とともに変動する現象の解析と制御を追求しつづける。1970年代にモデリングと予測の概念を打ち立て、モデルの評価規準を提案。今日の統計学の礎を築いた。1986年統計数理研究所長。1988年の総研大の創立にも尽力する。新しい統計的思考法とは、「客観的知識、経験的知識、観測データの3要素を組織的に構成し、それにもとづくモデルの提案と検証の繰り返しによって情報獲得を実現するもの」と、総研大学生セミナーで講演。これまでの功績に対し多くの賞を受けており、2006年には京都賞を受賞した。

場合、AICのような評価値はそれぞれのモデルにいちおう付けられるわけです。その集団を見て判断する。

もしそういうモデルの集団が尤度で評価できる性質のものなら、それで比較すればよい。データをモデルに合わせるといふ操作が入っていれば、AICで比較すればよいということでしょう。しかし、最も重要なことは新しいモデルの提案です。

堀田 生物学では実験は無限にはできないので、次に何の実験をすべきかを考えます。普通は、自分の興味に応じて次の実験をするわけですが、どんな可能性が残っていて、それをどういう順序で実験

していくかを考えるときに、何か基準がいろいろある。それは、先生のおられることと似ていると思うのです。

赤池 私は、最近の情報空間のモデリングにはまったく疎いのですが、生物学のモデリングは、それに近いのではないのでしょうか。いろいろな情報があって、それをどうクラス分けして整理すると、全部を探し当てられるか。これは今いちばん新しい領域です。検索のプログラムなどもこれですね。昔はどうにもならなかったものが、今はある程度まとめて拾ってこられます。ですから、技術的な問題が大きいと思います。具体的に問題

の内容を提示して、情報研究者と協力すれば、成果が出るのではないのでしょうか。
堀田 生物学は今やゲノムの時代で、遺伝子を全部数え上げることができるようになり、原理的には1つ1つの遺伝子について実験ができる時代です。昔は、一部を見て全体を理解するのが科学だと思っていたのですが、ゲノムは逆で、何かわからないが全部を見てしまう。しかし、そのすべてについて実験することはできない。どういう学問をしたらよいのか、生物学者の悩みの種なのです。

赤池 データが無数にある。そのときに、いちばん大事なものは、何が欲しいかですね。目的意識が明確ならば、それによって見方が決まってくるわけです。ですから、モデリングでは、それをどうとらえるかという知的な部分が肝腎です。

4 研究所が合同して新しい成果を

赤池 サイエンスをやるときには、目的意識をはっきりさせないと、何も出てきません。あれに使おう、と思っていけば、いろいろな可能性が見えてきます。その体制があるかどうか問題ですね。

堀田 理論的思考に強い人が必要で、今の生物学にはそれが欠けていると思うのです。ゲノムの時代の生物学は、下手をすると「分子生物学」になってしまう。元来、博物学を脱したのが生物学の革命だったはずなのに、1つ上のレベルの博物学になりかねない。これをどうしたらよいか。生物学者だけでは解決できないと思いますね。

統数研、遺伝研、極地研、情報研が一緒になって情報・システム研究機構ができたきっかけは、法人化という外からの圧力ですが、仕方がないから一緒になりました、というのではなく、積極的に何かを生み出したい。その何かのイメージとして、例えば生物学のこんな事情があるのです。

赤池 いちばん心配なのは、今の世の中はすぐ目に見える業績をあげないと、生きにくいシステムになっています。それで果たして、新しい分野を切り開く余裕があるのかどうかということですね。と

新しい科学の「かたち」

樋口知之

総合研究大学院大学教授 統計科学専攻／
情報・システム研究機構 統計数理研究所副所長

ICT革命が進むなかで、科学の方法論が変容しつつある。自然科学においてはシミュレーションの役割が増大してきた。シミュレーションは通常、研究対象の基本方程式群を計算機に実装するために数理モデルに変換した、いわゆる「シミュレーションモデル」の開発から始まる。それが時間的に変化するものであれば、初期条件、境界条件などを与えて計算を進め、得られた結果から科学知を発見していく。これは、演繹的推論と呼ぶフォワード思考（順問題解決）であり、ほとんどの研究領域における王道とされてきた。一方、統計科学においては、研究対象の理解のために、現象を支配している関係式や経験則を、観測や計測データから推定していく。帰納的推論と呼ぶバックワード思考（逆問題解決）がそこでは王道である。帰納的推論を行うには、対象そのものを基礎方程式群などの積み上げによって実体的にモデル化するのではなく、対象の機能自体を模倣する「数理モデル」を構築する。従来、このようなアプローチの活用は限られていたが、複雑な対象から大

量かつ多面的なデータが得られるようになった今日では、その出番は比較にならないほど増えてきている。

その代表例が近年におけるロボティクスの著しい発展である。かりに、悲しい顔を見たら相手を元気づけるようなロボットを作りたいとする。演繹的な方法論では、目の生理機能から始めて、信号伝達、脳での信号処理の理解、そして運動方程式にもとづいたロボット制御まで、すべての素過程を積み上げていくことになり、目的の達成まではなかなか至らない。そこで、機能のモデル化を優先し、入力データと出力データの関係を近似した数理モデルを構築するという手法をとっている。

このようなアプローチの重要性に最も早く気づいたのが赤池先生であった。「我々が追求する真理は、現在の知識に依存するという意味で相対的な、対象の一つの近似を与えるモデルによって表現されるようなものに過ぎない」（『時系列解析の実際II』朝倉書店、1995）と看破している。ベスト（真理）を求める一元的価値観が西洋的思想とするなら、ベストをイメージしつつもベターを志向し続ける、多元的価値観を重んじる東洋的思想といえよう。

一元的価値観、大量消費戦略が社会・環境問題において綻びを見せている現在、未来の地球を担う若者にはぜひ赤池思想を学び、真理の探究にとどまらない新しい科学を開拓してほしい。

んでもない人が出てこない。

堀田 それが難しい。合理化だ、人件費削減だ、という流れに対応しながら本質を見失わないようにしたいものです。

赤池 違う分野が接触するのは非常によいことですが、それだけでは

堀田 完全に同じ研究をしているところが一緒になっても、縮小するだけ。全然違うことをしているけれど接点があるものが協力し合うのがよいのではないかと、こういう組み合わせになったのです。なんとか先生方がつくられた伝統をさらに発展させて、法人化が新しいものを生み出すようにしたいと思います。

情報・システム研究機構では今、ライフサイエンスの統合データベース作りという活動をスタートさせました。ライフサイエンスのデータベースは、ゲノム、タンパク質、代謝、病気などについて独立に作られてきて、酵素や遺伝子の呼び方までそれぞれ違うのです。まずは辞書作りが必要です。そして、多様なデータベースをつなげるプラットフォームを作

り、そこに皆で知識をはり付けていけば、大きな百科事典ができるはずですが、これも機構だから可能な仕事で、幸い、4つの研究所があるので、あと4～5年経つと何か見えてくるかもしれません。

異端児がいられる空間が必要

堀田 私は、数学者にぜひ、生物学に入ってきてほしいと思っています。先生は数学者の中では少し異端児、と申し上げてもよいのではないのでしょうか。普通の数学者は、現場のデータを眺めて研究するというセンスはあまりありませんから。

赤池 そうでしょうね。

堀田 金融などには、数学の専門家がかなり入っていますが、生物学は尻込みされる。先日、東大数学科にセミナーに行つて、「ぜひ天才に来てほしい」と宣伝したのです。

赤池 普通の優等生では無理でしょうね。

堀田 数学ができるだけではなく、生物を見て、複雑なままにまず理解し、そのうえで生物学者が考えつかないことを考

えつく。そういう人が欲しいのですが、すぐには出てこないかもしれない。

赤池 わかっている問題の解き方ばかりを教える教育が多すぎますからね。どんなことでも考えれば解けるという経験をさせないといけません。さらに、とんでもないことをやっても大丈夫という空間を確保してやらないと。

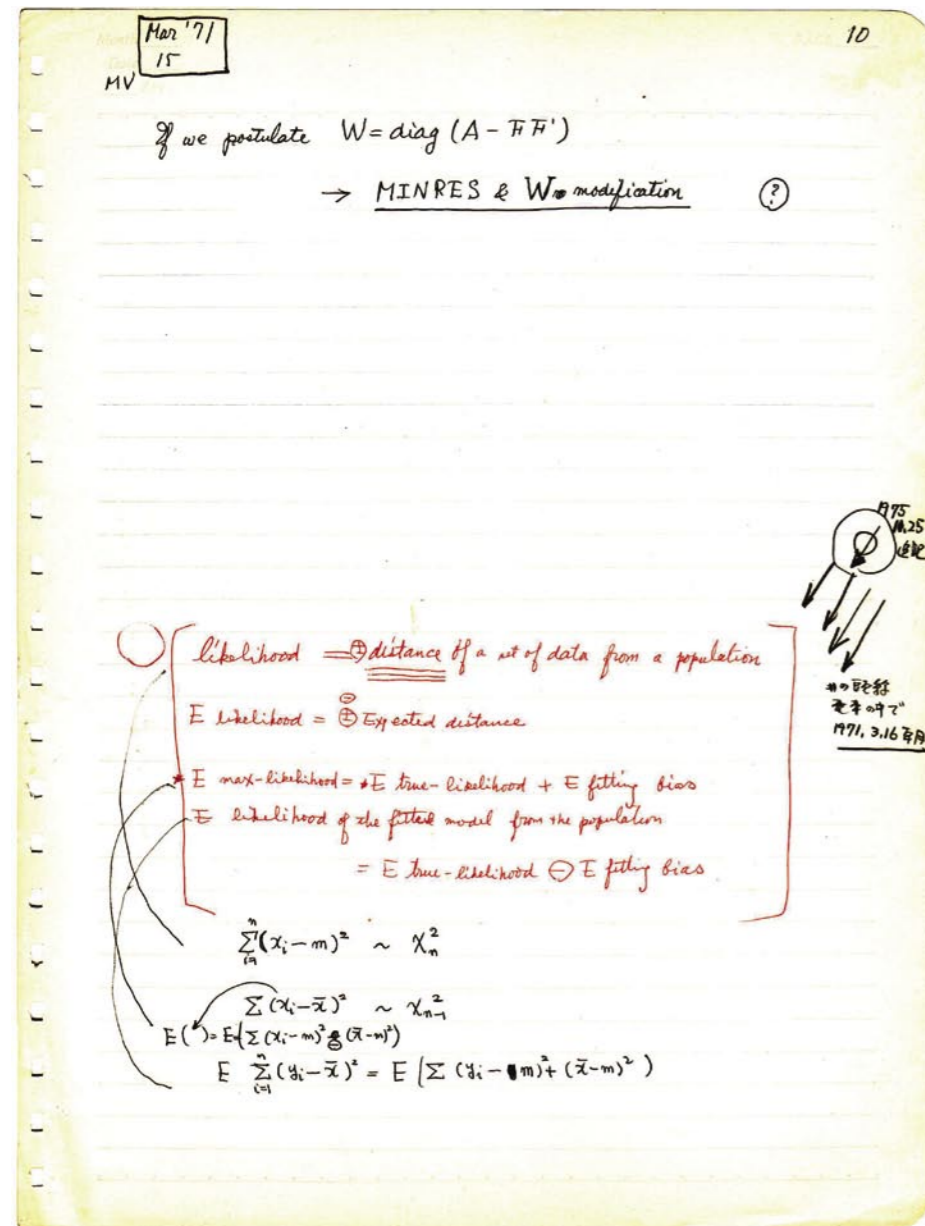
堀田 昔なら、ラボ全体としてはきちんとスタンダードな仕事をしていても、その中に異端児がいて、それが次の時代をつくっていくような余裕があったのですが、今はそういう余裕がだんだんなくなっていますね。

赤池 楽しくて仕方がないから研究するというでないとね。

堀田 そのとおりですね。そういう学問をなんとか続けていきたいものです。本日はおもしろく、示唆に富むお話をありがとうございました。

（2007年7月6日、つくば市にて収録）

（構成 古郡 悦子）



対数尤度からAICへの着想を書いた赤池博士のメモ