

題名	物の動きを読む数理—情報量規準 AIC 導入の歴史—
Title	The Use of Mathematics for Deciphering the Movement of an Object: A Historical Review of the Introduction of AIC
著者名	赤池 弘次
Author(s)	Hirotsugu Akaike
言語 Language	日本語・英語 Japanese, English
書名	稲盛財団：京都賞と助成金
Book title	Inamori Foundation: Kyoto Prize & Inamori Grants
受賞回	22
受賞年度	2006
出版者	財団法人 稲盛財団
Publisher	Inamori Foundation
発行日 Issue Date	8/1/2007
開始ページ Start page	134
終了ページ End page	161
ISBN	978-4-900663-22-0

物の動きを読む数理—情報量規準AIC導入の歴史—

赤池 弘次

はじめに

情報量規準の考え方は、従来の正統的な統計学の展開に沿う形の研究で得られたものではなく、実際的な問題処理の必要から、道のない所を手探りで進むような研究を続ける過程で到達したものです。その意味では型破りな発想に見えます。このような研究の進め方をするようになった理由を、自身の物の見方の癖と育った環境や時代の影響などと関連させて見直しながら、情報量規準AICの導入とその効果について歴史的に眺めたいと思います。

生い立ち

富士山の南山麓の田舎に生まれましたが、叔父の一人が民間航空の草分け時代の操縦士で、この叔父から貰った機関車やモーターボートの模型をひっくり返して動きの仕組みを見るのが好きでした。この辺りから、出来上がった結果を眺めるよりは、何故そのような結果が生まれるのかを探ることに興味を感じていたように見えます。結果を予測しながらあれこれ試すわけです。何事につけてもこのスタイルで進む癖が、「三つ子の魂百まで」とそのまま現在まで継続しています。

小学生の頃は、書き取りなどの暗記物や珠算のような機械的な作業は不得意で、算術の応用問題を解くことは得意でした。飛行機の操縦士の叔父の家に寄宿して学校に通っていた兄の影響もあり、また、海軍の戦闘機乗りであったもう一人の叔父の影響もあって、中学に進む頃までは飛行機に関心がありました。この叔父が数学の話聞かせてくれ、数学への関心が芽生えました。

中学では熱心な先生に英語を教えられましたが、単語の暗記は苦手で、実際の使用状況を単文の形で覚えていました。剣道と水泳の練習にも精を出して、結局当時の海軍兵学校に進学しました。

兵学校では、理工系の知識の初歩的な部分に触れることができ、確率や統計の初歩的な知識にも触れる機会が与えられました。すべての知識を有機的に利用することの有効性を体感したように思います。

The Use of Mathematics for Deciphering the Movement of an Object: A Historical Review of the Introduction of AIC

Hirotsugu Akaike

Introduction

The concept of an information criterion was formulated not through studies that followed traditional, classical statistics, but through persistent efforts to get to grips with practical problem solving, and thus it appears to be an unconventional idea. I would like to look back on why I took such an approach in relation to my way of viewing things and the influence of the environment and times I grew up in. I will also chronologically review the introduction of the AIC and its impacts.

Personal history

I was born in a rural area on the southern side of the foot of Mt. Fuji. One of my uncles, who was a pilot in the early days of commercial flight, gave me steam engine and motorboat models. As a child, I loved to turn them upside down to see how they worked. Since then, I guess, rather than observing an accomplished result, I have been more interested in exploring why such a result is produced. I like to try out different things predicting the outcome. Even today, I still maintain this habit in anything I do. "As the twig is bent, so grows the tree."

As an elementary school pupil, I was good at solving questions of applied arithmetic but not good at memorizing or rote learning, such as writing kanji characters, or calculating with an abacus. Influenced by my brother, who lived with the pilot uncle in order to go to school, and another uncle who was a naval warplane pilot, I had an interest in airplanes by the time I went to junior high school. This naval pilot told me about mathematics, in which I became interested.

My English teacher at junior high school was very enthusiastic, but I was not good at memorizing English words. Instead, I memorized simple sentences, which showed how English words were actually used. I had a passion for kendo and swimming as well, and eventually I entered what was then the Naval Academy.

At the Naval Academy, I had opportunities to learn introductory science and engineering, statistics, and probability. I believe that I learned the usefulness of applying all pieces of knowledge organically through experience.

終戦後は戦後の復興に貢献することを目指して故郷に帰りましたが、社会の価値観が音を立てて崩れるのを見て大いに悩み、あれこれ考えた末に、自他の生命の尊重が道徳の基本であると認識して心の平安を取り戻しました。旧制の第一高等学校理科に進学し将来の進路についても悩み、結局東京大学理学部数学科に進みましたが、これは私にとって、いわゆる狂瀾怒涛の時期でした。

1952年に統計学関連の数理の研究機関である当時の文部省統計数理研究所に入所し、ここで我が国の現実に根を置く研究を目指しました。10年間ほどは具体的問題の探索に時間を費やしましたが、戦後の復興を支える生産活動に直結する課題に触れる機会を得て、以来一貫して時間と共に変動する現象の解析と制御に関係する課題を追求することになりました。これは、子供の頃からの動く物の仕組みに対する興味に導かれた結果のように見えます。この課題の選択があれこれの発想を要求してAICに到達することになったわけです。

これから、AICに関連する話題に進みます。

予測の数理

確率

AICを利用して観測データの有効利用を進めるには、当面の観測データがどのような形で得られるのかを説明するモデルを使います。データを生み出す仕組みは完全には記述できないのが普通で、この不確かさを表現するために確率を利用してモデルを組み立てます。

確率の考えはごく自然なもので、不確実な事柄の処理では誰でもほとんど無意識の内に利用しています。ある結果が現れることに対する見込みの確からしさを数字で表すものが確率で、明日雨が降る確率は0.7（70パーセント）などと表現します。この場合10中7の割合で明日は雨が降ると見込んでいるわけです。この考え方自体はごく単純に見えますが、実際にある結果に対する確率を求めようとすると、いろいろ考えなくてはなりません。

Aiming to contribute to postwar reconstruction, I went home after World War II. However, I was greatly distressed to see the collapse of social values. After thinking many things, I realized that respecting my own and others' lives was the basis of morality, and this restored my inner peace. Although I attended the Science Course of the former First Higher School, I was concerned with my future career. Eventually, I went to the Department of Mathematics, in the Faculty of Science at the University of Tokyo. This period was a turbulent time for me.

In 1952, I joined the Institute of Statistical Mathematics of the Ministry of Education, where I aspired to work on studies focusing on real issues concerning Japan. After ten years of exploring solutions to specific problems, I had the opportunity to deal with an issue that was directly associated with production activities that supported postwar reconstruction efforts. Since then, I have always been dedicated to problems associated with analysis and control of phenomena that vary with time. I believe this was a result of the interest in the mechanisms of moving objects I'd had since childhood. The choice of this issue required a variety of ideas, and resulted in the creation of the AIC.

Let me move on to AIC-related topics.

Mathematics of prediction

Probability

The effective use of observational data using the AIC requires a model that explains how such data can be obtained. It is usually impossible to precisely describe the mechanism for generating data, and therefore, we create a model using probability to express this uncertainty.

The concept of probability is quite natural; everyone uses it almost unconsciously for processing uncertain things. Probability uses numbers to express the likelihood of a certain outcome occurring; for example, we say the probability of rain tomorrow is 0.7, or 70%. In this case, the odds are ten to seven that it will rain tomorrow. This seems very simple, but we must think of many factors to estimate the probability of an outcome in reality.

これに対して、サイコロを振った時に6の目が出る確率のように、確率を決める仕組みが客観的に与えられる場合もあります。これは確率の考えの出発点となった、占いや賭で利用されるランダムな動きをする道具、すなわちランダムマイザーの利用で与えられる確率です。

サイコロのそれぞれの目に賞金が出る賭では、賞金にその目が出る確率を掛けて加え合わせれば期待値が求まります。この場合の考え方の要点は

期待値 = 確率に基づいて予想される値

ということです。そこで、賭の参加者は儲けの可能性について慎重に考えます。

確率と統計の繋がり

実際に確率を使って効果的に将来に対する計画を立てようとする、確率のもつ複雑な内容に気がつきます。明日雨が降る確率が0.7と言う時は、何かの根拠に基づいてこの数字を取り上げています。たとえば、過去の記録で今日の気象状態と似た状態の日の翌日が雨であった場合の割合、というような根拠に基づいて決められます。

そこで、これまでの同様な状況の日の記録を並べてみると、過去の100日の例の中で70日が翌日雨であったというような結果が得られます。この場合、日常の言葉の使い方に従えば、過去の統計に基づく、全数に対する雨の比率 $70/100(=0.7)$ 、すなわち平均的な発生率で明日が雨になる確率が決まることになります。

これを簡単に表現すれば

過去の統計の平均値 = 将来に対する期待値

となります。実用上の基本的な考え方はこのようなもので、確率が統計的なデータの利用と自然に結びつくことが分かります。

この考え方の基礎にあるのは、サイコロを何回も投げるように、データを無限に長く記録し続ければ確率が求まるという考え方です。ところが、実際にはこれは不可能で、逆に今手許にあるデータをもとに、将来の観測値を生み出す確率的な仕組み、すなわちモデルを作り上げ、これを使って予測を実行することになります。

In contrast, the mechanism to determine probabilities is objectively given in some cases, such as the probability of getting a six when you throw a dice. This type of probability is given using a randomizer—a tool that generates random numbers used in fortune-telling and gambling—and from which the idea of probability originates.

In betting on dice, where the prize money depends the number on each side, the expected value of prize money is obtained by multiplying the amount for each number by the probability of getting each number, and totaling them. The key point here is that the expected value is based on probability. Therefore, the participants in the betting think carefully about the probabilities of winning.

The relationship between probability and statistics

When you use probability for the effective development of a future plan, you will find out how complicated probability is. When someone says that the probability of rain tomorrow is 0.7, the value is based on some evidence. For example, the evidence may be the proportion of the days on past record when the weather was like today's weather and it rained on the following day.

Looking at past records, the ratio of rain on the following day was 70 days out of 100 days, when the weather was similar to today's weather. In this case, in ordinary terms, based on statistical records from the past the ratio of rainy days to all days is 70 to 100, or 0.7. In other words, the probability of rain tomorrow is given as an average incidence.

In short, the mean value of statistical records in the past equals the expected value in the future. This represents the basic, practical idea of probability, which suggests that probability is naturally associated with the use of statistical data.

The basis of this idea is that probabilities can be derived by recording data infinitely, like throwing a dice many times. However, this is impossible in reality. Instead, we use data at hand to create a mechanism of probability that generates future observational values, namely, a model. We make predictions using this model.

このように、利用可能なデータを有効利用して、将来の値を予測する仕組みをつくるのが統計的データ処理の本質なのです。「温故知新」を具体化する形の仕事です。以下、その実際の利用の例について話を続けます。

実際問題への適用

生糸繰糸工程の統計的管理

最初の例は生糸の繰糸工程の統計的管理で、当時の農林省蚕糸試験場の島崎昭典さんとの協力で得られたものです。戦前は生糸が日本の主要な輸出品で、歴史的に見ても長い伝統のある産業です。しかし、その生産方法は経験に大きく依存していましたが、戦後の統計的工程管理手法の導入に伴い、時々刻々の観測値に基づいて工程の異常の検出を行う管理図法の導入が検討されていました。

当時は一定個数の繭の繊維を撚り合わせて一本の生糸に紡いでいました。管理の対象は一定の時間内に繊維が切れて繭が落ちる落緒の回数で、これが異常に増加すれば工程の異常と判断されます。通常管理図法の教えでは、定常状態の運転記録から平均的な落緒数と、許容される変動幅を決定し、観測値がこの巾の外に出た時点で異常と判定することになります。

ところが、繭から生糸の繊維を引き出しやすくするための煮方を決める試験繰糸の結果として、一本の繊維の長さの統計的分布のデータが得られます。これを利用すると、次々に繊維を繋ぐ繰糸の流れの中での切れ目の現れ方が確率論的に決まります。これで理想的な運転状態の特性が確認され、管理に必要な数値も理論的に求まります (Fig. 1)。こうして実用上著しい成功が得られたのです。落緒の発生という将来の事象の確率的な構造を考えることにより、試験繰糸で得られた過去の知識の有効利用が実現したのです。

パワースペクトルの推定

これは、ランダム振動の統計的解析ということで、1960年代に多くの関心を惹いた問題ですが、不規則変動の特性を計測する方法の実用化です。当時のいすゞ自動車

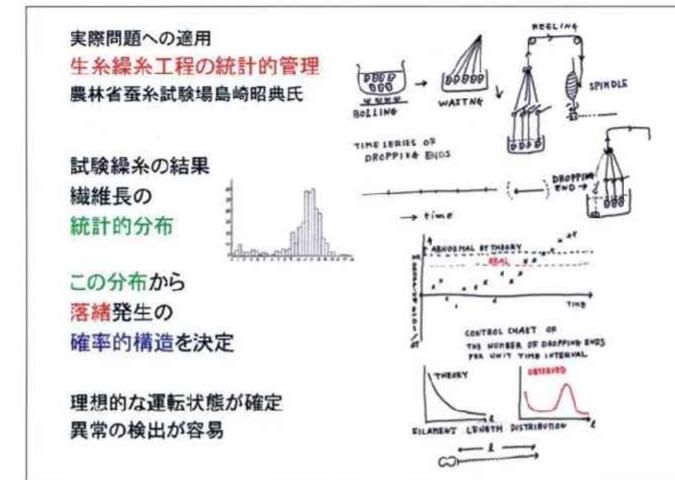


Fig. 1

In this way, creating a mechanism to forecast future values using available data is the essence of statistical data processing. It is the materialization of what Confucius said: "Study the past if you would divine the future."

Let me continue to speak about examples of the practical applications of probability.

Application to real problems

Statistical control of the silk reeling process

The first example is statistical control of the silk reeling process, achieved in cooperation with Mr. Akinori Shimazaki of the sericulture laboratory of the then Ministry of Agriculture and Forestry. Raw silk yarn was one of Japan's major exports before World War II, and sericulture is a traditional industry with a long history. Production of raw silk yarn largely depended on experience. After World War II, however, statistical process control methods were introduced, triggering the introduction of control chart methods that helped detect abnormalities in processes based on ever-changing observational values.

株式会社の兼重一郎さんと共に、この方法の実用化を進めました。

時間的に変動する時系列データを周期的な振動成分に分解すると、周波数（単位時間内の振動数）を示す横軸上に成分の強さを示すギザギザと激しく変動する系列が得られます。これはピリオドグラムと呼ばれ、その中にひときわ目立つ成分が見出されることがあります。地球物理学的な観測値に含まれる周期的成分の検出に、古くから利用されてきたものです。

ところが、この解析法では、自動車の振動のような不規則変動のデータの場合は、全面的にギザギザの動きが現れるだけです。そこで当時の統計学では実用性が疑われていました。しかし、周波数軸上のデータを局所的に平均化すると、パワースペクトルと呼ばれる滑らかな変動のパターンが現れ、光のスペクトルのようにどの辺りの周期成分が優勢であるかを見ることができます（Fig. 2）。

周波数応答関数の推定

さらに一步進めると、ランダムに変動するハンドル入力を利用して自動車の操縦性能を測るというような、システム特性の計測法が得られます。この場合のシステムの特性を表現するものが周波数応答関数です（Fig. 3）。

この計測方法では、現実の動きに類似した状況で、システムの特性を簡単に計測できます。この問題の取り扱い、当時の運輸省運輸技術研究所の山内保文さんとの協力で実用化に成功したものです。この周波数応答関数の推定法を様々な分野の研究者が実例に適用した結果が、1964年に統計数理研究所の英文報告に発表されましたが、これは当時世界的に見ても先端的な成果でした。この事例には、自動車、船舶、鉄道車両、航空機、配管系、水力発電所の水管系などへの応用結果が含まれています。

生産プロセスの最適制御

周波数応答関数の推定法を、当時秩父セメント株式会社の中川東一郎さんとの協力

In those days, raw silk yarn was made by intertwining fibers from a specific number of cocoons. What was controlled was the number of times that fibers were broken and cocoons dropped within a specific time range. When the value increased abnormally, the process was judged as defective. Ordinary control chart methods use operating records of a machine in a steady state to determine the mean value of the dropping of cocoons and a permissible range of fluctuation, judging the process as defective when the observed value goes beyond the permissible range.

However, data for the statistical distribution of length of a fiber are obtained as a result of experimental silk reeling to identify how to boil cocoons so as to facilitate reeling fibers off the cocoons. Using these data, the likelihood of a fiber breaking in the process of reeling is stochastically determined. This procedure verifies the characteristics of the ideal operation of a machine and theoretically determines the values necessary for control of the process [Fig. 1]. In this way, we achieved remarkable success in practical process control. A piece of past knowledge obtained through experimental reeling was effectively used by thinking about the stochastic structure of a future event, which was the dropping of cocoons.

Estimation of a power spectrum

This is a kind of statistical analysis of random vibration, which attracted considerable attention in the 1960s, and a practical application of a method to measure the characteristics of irregular fluctuation. I worked on putting this method into practice in cooperation with Mr. Ichiro Kaneshige, then of Isuzu Motors, Limited.

When time-series data varying with time are broken down into periodic vibration components, a sharply fluctuating zigzag line that expresses the strength of components along the x-axis indicating frequency (the number of vibrations within a unit time) is obtained. This is called a periodgram, which has an extremely conspicuous component in some cases. The periodgram has long been used for detection of periodic components of geophysical, observational values.

This analysis, however, provides only a zigzag when using data of irregular variation such as vibration of an automobile. Therefore, it was doubted whether this analysis method could be applied to the statistics of the time. Local averaging

を通じて、セメント焼成炉の自動運転の実現に応用しようとしたのですが、難しい問題が浮かび上がり、その代わりに計算機制御に適した新しい最適制御理論の適用を考えることになりました (Fig. 4、5)。

制御系の設計の基礎となるルドルフ・カルマン教授 (第1回(1985)京都賞先端技術部門受賞者) のシステム理論では、伊藤清先生 (第14回(1998)京都賞基礎科学部門受賞者) の研究成果による確率論的な表現が利用され、理論が展開されていましたが、変動の大きな実プロセスでの実用化には種々の困難がありました。これを克服するために、まずプロセスの動きの予測式を作り、これを利用して制御を実行する方法を考えました。

自己回帰モデル

関係する変数全体の時系列について、現在の値を過去の何時点かの値に適当な係数を掛けて加えたもので予測する自己回帰モデルを採用し、その構造を時系列データから推定し、その結果から各変数間の繋がりを解析するという方法を採用することになりました。この場合の予測誤差は、白色雑音と呼ばれる、単純なランダムイザーの出力として表現される構造を持ち、これを利用して実際のシステムの動きのシミュレーションが可能になります。これを利用すれば、時々刻々将来の動きを予測し、適当な制御入力を決定する制御が可能になります (Fig. 6)。

ところがここでさらに問題が発生しました。予測式に過去の何時点前までのデータを取り込むかという、いわゆるモデルの次数の決定の問題です。当時この問題についての実用的な方法は全く見当りませんでした。次数が低すぎると予測力がなくなり、高すぎるとデータの不足から推定の精度が落ちてしまいます。そこで、次数の異なるモデルの比較評価の方法を考案し、これを使って最適モデルを選ぶことで制御系の設計が可能になりました。

最適制御の実現

多変量自己回帰モデルの利用で、実際のシステムの動きの実態が簡単に把握できる

of data along the frequency axis, however, manifests a power spectrum that presents a smooth fluctuation pattern. As is the case with a light spectrum, a power spectrum indicates which frequency has a stronger periodic component [Fig. 2].

Estimation of a frequency response function

A further advanced application of probability provides a method to measure the characteristics of a system, such as measuring the steerability of an automobile by analyzing the random motion in a steering wheel. A frequency-response function represents the characteristics of a system in this case [Fig. 3].

This measurement method enables easy measurement of the characteristics of a system in a state that is similar to real movement. This measurement method was successfully put into practice in cooperation with Mr. Yasufumi Yamanouchi, the Transportation Research Institute of the then Ministry of Transport. This estimation method of a frequency-response function was one of the most advanced achievements in the world and was applied to actual cases by researchers in many fields. Results of such applications were reported in an English journal of the Institute of Statistical Mathematics in 1964. Those cases include applications to automobiles, ships, railway vehicles, aircrafts, piping systems, and water-piping systems of hydroelectric power plants.

Optimal control of production processes

I tried to apply the estimation method of a frequency-response function to the automatic operation of cement kilns in cooperation with Mr. Toichiro Nakagawa, at the then Chichibu Cement Co., but a difficult problem came up. Instead, we decided to use a new optimal control theory that is suitable for computer control [Fig. 4, 5].

Professor Rudolf Emil Kalman's system theory—the Professor was a laureate of the 1985 Kyoto Prize in Advanced Technology—served as the foundation for design of the control system. However, we had various difficulties putting Professor Kalman's system theory, which used stochastic expressions derived from the study

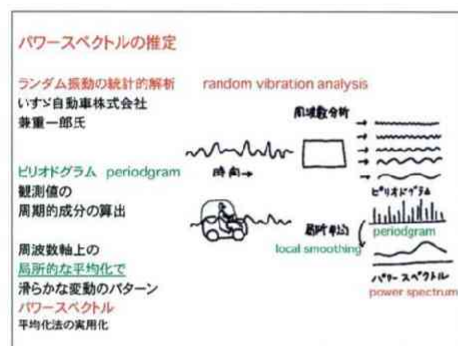


Fig. 2

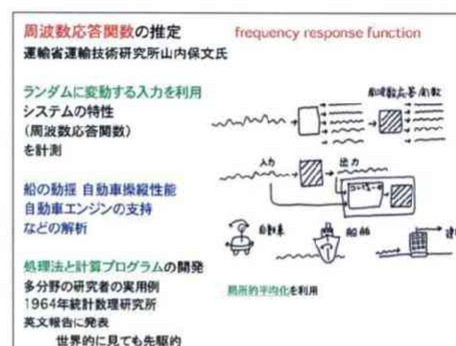


Fig. 3

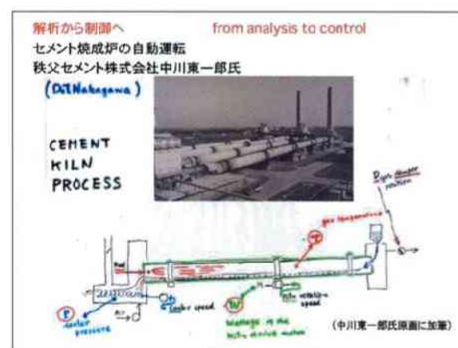


Fig. 4

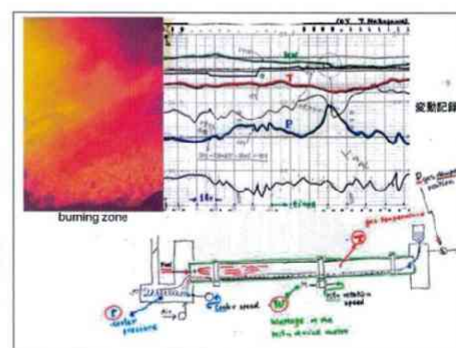


Fig. 5

時刻 n の観測値 $X(n)$: (システムの入出力を並べたベクトル)

$X(n)$ を 過去からの影響分

$A_1X(n-1)+A_2X(n-2)+\dots+A_kX(n-k)$ で予測

自己回帰(AR)モデル autoregressive model

$X(n) = (A_1X(n-1)+A_2X(n-2)+\dots+A_kX(n-k))+Z(n)$

$Z(n)$: 白色雑音

過去のデータを用い係数 A_1, A_2, \dots, A_k を決める

$Z(n)$ をランダム化の出力で置き換え
シミュレーションで制御の効果を検討

時々刻々将来の動きを予測 → 制御を実行

Fig. 6

results of Dr. Kiyosi Itô, a laureate of the 1998 Kyoto Prize in Basic Sciences, into practice in a real, largely variable process. To overcome such difficulties, we began by formulating a prediction equation for the movement of a process and then thought about how to control a process using this equation.

Autoregressive model

The autoregressive model predicts a future value by adding a current value to a sum of past values multiplied by specific coefficients in a time series of all related variables. I decided to use the following method: I estimated the structure of the autoregressive model using time-series data, and then using the results I analyzed relationships among variables. Prediction errors in this case have the structure expressed as a signal of a simple randomizer called white noise, which enables simulation of the movement of an actual system. Use of white noise makes it possible to predict future, momentary movements of a system to determine appropriate inputs for controlling the system [Fig. 6].

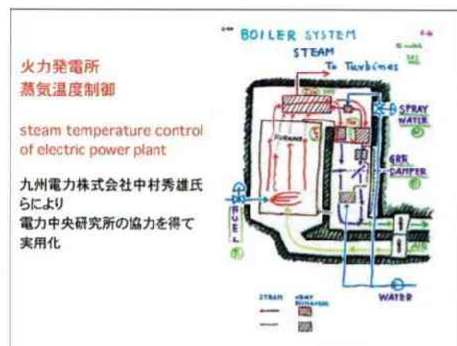


Fig. 7

ようになり、不規則な変動を示すセメント焼成炉の制御での最適制御理論の実用化に成功しました。

この方法の実用に必要な計算プログラムを公表した結果、火力発電所ボイラー温度制御、船舶の自動操舵、生体の環境維持機能（ホメオスターシス）の解析、脳波解析、経済時系列解析、原子炉雑音解析など、従来は変量間の動きの関連が確認困難であった様々な研究対象について、実測データによる構造の確認とシミュレーションによる検討が可能になり、それぞれの分野で、驚きをもって迎えられるような結果を生み出しました。

火力発電所の蒸気温度制御の実用化は、当時九州電力の中村秀雄さんが電力中央研究所の研究者の協力を得て実現しました（Fig. 7、8）。定期点検後の運転再開時の模様を見学したことがありますが、この時現場の運転責任者からこの制御を歓迎する発言がありました。後日、制御理論の大家、高橋安人先生（1912–1996）が、「光は東方から、実プロセスの最適制御は日本から」と、この制御を高く評価されたという話でも中村さんから聞いております。

この制御はその後国内外のボイラーメーカーによって利用され、国内、中国、カナダで運用中とのことです。セメント焼成炉の自動運転の実現と共に、変動の大きな現実のプロセスで最適制御理論の適用を実現した、理論と現実を繋ぐ貴重な成功例と言えます。

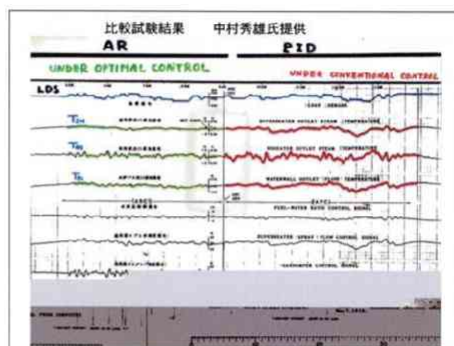


Fig. 8

I had a problem here again. It was how to determine the order of a model; in other words, how far you go back to the past to incorporate data into the prediction equation. No practical solution to this problem was found in those days. When the order is too low, the equation is less likely to predict a future. When it is too high, insufficient data causes inaccurate prediction. Therefore, I created a method to compare and evaluate models with different orders. This method enabled choosing the optimal model and, thus, designing control systems.

Making optimal control happen

The multivariable autoregressive model made it easier to grasp the movement of actual systems, helping put optimal control theory into practice in control of cement kilns with irregular variations.

As a result of publishing the calculation program necessary for practical application of this method, verification of the structure using observed data and examination by simulation were made possible and produced outstanding results in various areas where verification of the relationship of movements among variables had been difficult, such as temperature control of boilers at thermal power plants, automated steering of ships, analysis of a living organism's function for maintaining its internal environment, or homeostasis, analysis of brain waves, time-series economic analysis, and analysis of noises of nuclear reactors.

The application to steam temperature control of thermal power plants was made possible by Mr. Hideo Nakamura, then of the Kyushu Electric Power Company, Inc., with help from researchers at the Central Research Institute of Electric Power Industry [Fig. 7, 8]. When I had an opportunity to observe the restart of operation of a boiler after a regular inspection, the chief operator of the boiler welcomed the introduction of the control method. I also heard from Mr. Nakamura later that the master of control theory, Dr. Yasundo Takahashi (1912–1996), had thought highly of this control method, saying that the sun rises in the east and optimal control of actual processes comes from Japan.

Since then, this control system has been used by Japanese and overseas boiler manufacturers and is in operation in China and Canada as well as in Japan.

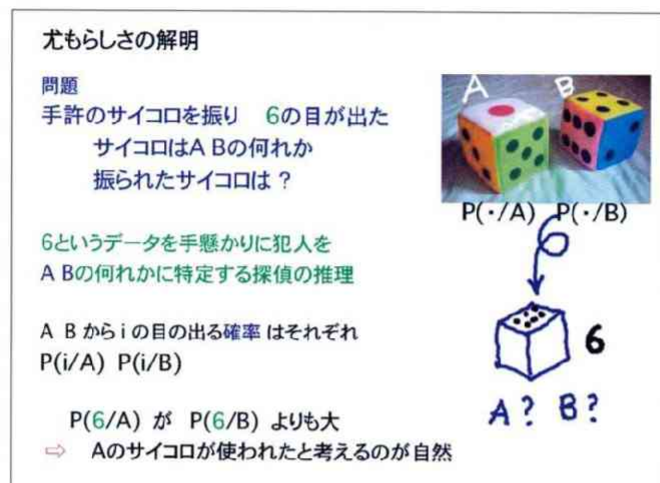


Fig. 9

実は、この場合、基礎になるモデルの次数を選ぶ際の評価値として何を取るのかは、簡単には決まらなかったのです。この問題を追究する過程で、これからお話をする尤度の概念の利用に引きつけられ、情報量規準AICに到達することになりました。

尤度（ゆうど：尤もらしさ）の解明

尤度とは？

ここで、日頃聞き慣れない言葉と思われませんが、確率的なモデルの優劣の評価を与える、尤度（ゆうど）について簡単な説明を試みます。これは非常に面白い概念で、情報量規準はこれに基づいて定義されます。

手許のサイコロを振って6の目が出たとします。ここからが問題です（Fig. 9）。A、B二つのサイコロがあり、その中の一つが振られことは確かですが、実際に振られたサイコロがどちらかは分からない、ということです。そこで、AかBかの判断が求められます。探偵が6という観測データを手懸かりに、犯人をA、Bの何れかに特定しようとする時の状況です。

This is a valuable, successful case that connected theory to reality, achieving application of optimal control theory to actual processes with wide variations, as well as the application of the automatic operation of cement kilns.

To tell the truth, it was not so easy to determine what to use as the evaluative value in choosing the order of the basic model. During the process of solving this problem, I was attracted to applying the idea of likelihood. I then came up with the information criterion, or AIC, of which I am going to speak now.

Clarification of likelihood

What is likelihood?

I believe many of you are not familiar with this word, so I would like to provide a brief explanation of likelihood that evaluates stochastic quality of models. This is a very interesting idea, based on which an information-quantity criterion is defined.

Suppose that you throw a dice and get a six. A question then arises [Fig. 9]. There are two dice, dice A and dice B. It is certain that one of them was thrown, but it is not known which was actually thrown. In such a case, you must judge which dice was thrown (A or B). This is like a case where a private detective is trying to identify a criminal between suspects A and B using an observational value, or six.

Suppose that the probabilities of getting a six when throwing dice A and dice B are $P(6/A)$ and $P(6/B)$, respectively. If $P(6/A)$ is much larger than $P(6/B)$, it is natural to think dice A was thrown when the six came up. When generalizing this idea, $P(x/A)$ and $P(x/B)$ for observational datum “x” are called the likelihood of dice A and dice B for datum x.

In this case, we must note that datum x has already been observed and is definite. $P(x/A)$ and $P(x/B)$ are the likelihood of dice A and dice B showing a six, respectively, but they are not probabilities of getting a six.

Probability uses past knowledge and experience to predict future data, while

AとBのサイコロで6の目が出る確率を、それぞれ $P(6/A)$ 、 $P(6/B)$ とします。今6の目が出たという条件のもとで、 $P(6/A)$ が $P(6/B)$ よりも遙かに大きければ、Aのサイコロが使われたと考えるのが自然に見えます。この考えを一般化して、観測データ x に対する $P(x/A)$ 、 $P(x/B)$ の値を、それぞれサイコロA、Bのデータ x に関する“尤もらしさ”、すなわち尤度 (ゆうど) と呼びます。

この場合、 x というデータは既に観測され、確定していることに注意する必要があります。 $P(6/A)$ 、 $P(6/B)$ は、それぞれ6に関するサイコロA、Bの尤度であり、6の目が出る確率ではありません。

確率は過去の知識や経験から将来のデータの見方を与えるものですが、尤度は、現在のデータを用いて、過去の時点にこれを生み出した仕組みを評価しようとしています。一見「温故知新」を超える知識の取り扱いの議論に入っているかのように見えます。この問題の本質は、次の情報量の議論で明らかになります。

情報量規準 情報量

尤度をモデルの評価値として利用することの根拠として、情報量と呼ばれる量の利用が考えられます。情報量 $I(Q:P)$ は、モデル P から見て真の構造 Q の離れ具合を測る量です。ここで P 、 Q は観測値の現れ方を決める確率的な仕組みを表し、それぞれの仕組みからデータ x が得られる確率が、 $P(x)$ と $Q(x)$ で与えられます。

観測データ x が得られた場合、これを利用して $I(P:Q)$ の値を測るには、 P の尤度 $P(x)$ の自然対数 $\log P(x)$ を利用します。これを対数尤度と呼びます。この変換により、二つのモデルの比較では、真の構造 Q を知らなくとも、「対数尤度の大きい方が Q に近い」と評価するのが合理的である、ということが確率の立場から説明できるのです。感覚的に表現すれば、海拔の高い山が「天」に近いように、対数尤度の高いモデルが「真理」に近いと見なされることになります (Fig.10)。これが情報量規準の意味で、これにより尤度の利用の根拠が明らかになったのです。

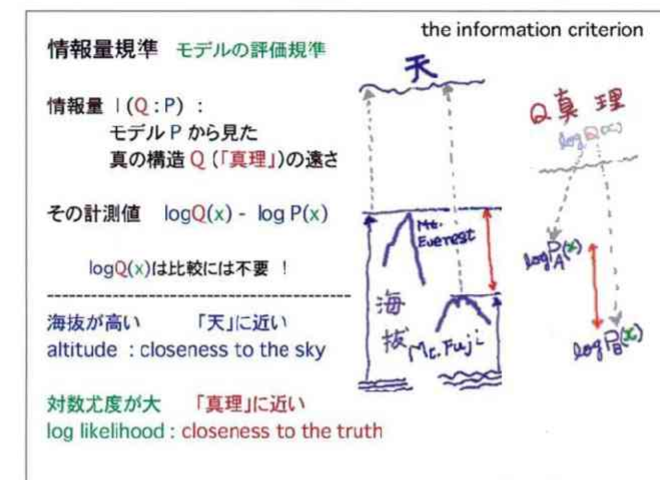


Fig. 10

likelihood uses current data to evaluate a framework that generated such data in the past. It seems that we are discussing how to handle knowledge beyond Confucius' words: "Study the past if you would divine the future." The essence of this question will be clarified in the next discussion on information quantity.

Information-quantity criterion Information quantity

One of the rationales for using likelihood as the value to evaluate a model is the use of a quantity called information quantity. Information quantity $I(Q:P)$ measures the distance from model P to real structure Q .

Here P and Q stand for stochastic mechanisms that determine how observational values are generated, and these mechanisms provide probabilities of getting datum x as $P(x)$ and $Q(x)$, respectively.

When datum x has been observed, the natural logarithm $\log P(x)$ is used to determine the value of $I(P:Q)$, where $P(x)$ is the likelihood of P . This is called log likelihood. Even though we do not know the true structure Q , this conversion explains from the perspective of probability that it is reasonable in the comparison

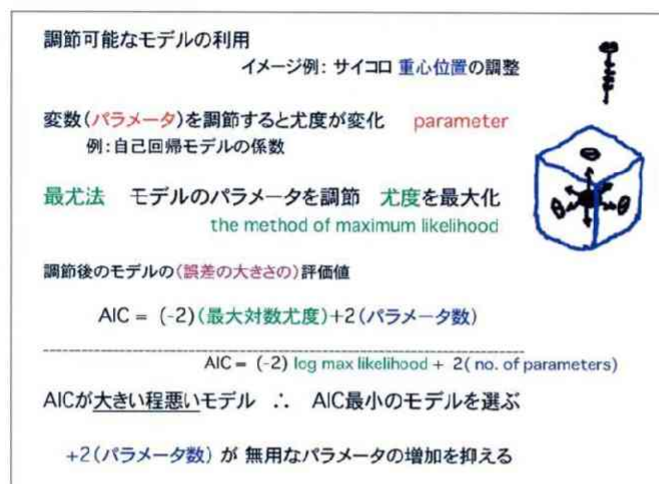


Fig. 11

パラメータを含むモデル

実際に尤度を利用して推論を展開する状況では、いくつかの調節可能な変数、すなわちパラメータを含んだモデルが利用されます。パラメータを調節するとモデルの構造が変わり、現在の観測データと同じ結果を生み出す確率、すなわち尤度を調節することができます。

簡単で具体的なイメージとしては、サイコロの重心部分に重い鉄球があると考え、その位置を前後、左右、上下に移動させる、三つの調節用のねじ、すなわちパラメータを考えます。これを動かせば、それぞれ鉄球の近づく面が下向きになる確率が高まるでしょう。このような方法で、確率分布を調節できるモデルが得られます (Fig. 11)。

AIC

いくつかのパラメータを持つモデルを用い、与えられた観測データに関する尤度を最大化するようにパラメータを調節する最尤法を適用すると、このモデルの情報量規

of two models to judge that the one with a higher log likelihood is more likely to be Q. A model with higher log likelihood is deemed to be closer to the true one, and an analogy for this is a mountain with a higher altitude being closer to the sky [Fig. 10]. This is the meaning of an information-quantity criterion, which clarified the rationale for the use of likelihood.

Models that include parameters

In an actual situation where we develop inferences using likelihood, we use models that include some adjustable variables, or parameters. When parameters are adjusted, the structure of a model changes. Consequently, the probability of getting the same conclusion as the current observed value, in short, the likelihood, can be adjusted.

A simple and concrete image of this is given by a dice with a heavy iron ball in its barycenter. Suppose that the position of the ball can be adjusted back and forth, from side to side, and up and down using three adjustment screws, or parameters. If you move these screws, it will increase the probability of the dice landing on the side that is closer to the iron ball. In this way, you can obtain a model that can adjust probability distribution [Fig. 11].

AIC

If you use a model with some parameters and apply the maximum likelihood method in which parameters are adjusted to maximize the likelihood for a given observational value, the model can be assessed using an information-quantity criterion, as shown in this equation:

$$AIC = (-2) (\text{the maximum log likelihood}) + 2 (\text{the number of parameters})$$

A subtle feature of the AIC is to use the log likelihood instead of the likelihood itself.

The larger the AIC is, the worse the model is deemed, due to the coefficient of -2 to the maximum log likelihood. The second term, or “2 (the number of parameters),” is obtained by adding the corrective value for an increment of the likelihood due to adjustment of parameters to the evaluation of errors in the prediction using the applied model. This term prevents an increase in unnecessary

準による評価は

$$AIC = (-2)(\text{最大対数尤度}) + 2(\text{パラメータ数})$$

で与えられます。尤度そのものではなく対数尤度を使うところが、情報量規準の微妙な点です。

最大対数尤度に掛かる (-2) という係数のために、AICが大きいほど悪いモデルと見なされます。第2項の 2 (パラメータ数) は、パラメータの調節による尤度の上昇分の補正と、あてはめられたモデルによる予測の誤差分の評価を加え合わせたものです。この項は無用なパラメータの数の増加を抑えます。

AICの発表

AICについての最初の発表は、1971年の日本統計学会での報告です。これに次いで、同じ年にアルメニア共和国で開催された第2回情報理論国際シンポジウムで詳しい内容を発表しました。これが1973年に公刊され、その内容の解説を1974年にアメリカ電気電子学会制御システム部会紀要に発表しました。

当初、応用分野からは積極的な反応がありましたが、正統的な統計学の分野からの反応は消極的でした。しかし、1992年になると、1973年の論文が現代の統計学への顕著な理論的貢献のいくつかを纏めた書物に収録されました。AICがゆっくりと受け入れられて行った様子が窺われます。

現在では、哲学者もAICに対して関心を示すようになりました。論理学では「オッカムの剃刀」と呼ばれる原理、すなわち不要な要素は使わないという原理が知られています。AICの 2 (パラメータ数) の項は、無用なパラメータ数の増加を抑えます。その意味で、AICをこの原理の具体化と捉えて議論を展開している人もいます。

しかし、AICの利用の前提はモデルの提案です。モデルの構成には、客観的知識、経験的知識、観測データなどを有機的に利用し、仮説の提案と検証を限りなく続ける努力が要求されます。このような仮説の提案の過程を通じて真理を目指す推論の展開を議論したC. S. パースは、これをアブダクション、あるいはリトロダクションと呼

parameters.

Publication of AIC

The AIC was first reported at the meeting of the Japan Statistical Society in 1971, and later in the same year, I presented its details at the 2nd International Symposium on Information Theory held in the Republic of Armenia. The presentation was published in 1973. In 1974, I contributed a review of the article to the bulletin of the Control Systems Society of the Institute of Electrical and Electronics Engineers, Inc., U.S.A.

There was positive response from the applied fields, but muted reaction from the classical statistics field. However, in 1992, the 1973 article was published in a book that put together some remarkable, theoretical contributions to contemporary statistics. This shows that the AIC had gradually been accepted.

At present, philosophers are also interested in the AIC. In logic, there is a principle called “Occam’s razor,” that does not use unnecessary assumptions. The term “2 (the number of parameters)” of the AIC prevents an increase in unnecessary parameters. In this context, some people have developed an argument considering the AIC to be the materialization of this principle.

However, the precondition for the use of the AIC is to propose a model. Modeling requires effort to repeat hypothesizing and hypothesis testing infinitely using objective knowledge, empirical knowledge, and observational data. C. S. Peirce proposed the idea of inference aimed at achieving truth through such a hypothesizing process and emphasized this type of inference, naming it abduction, or retrodution. The AIC helps carry out this process.

The effect of the introduction of an information-quantity criterion

The theory of the maximum likelihood method has been discussed as a way of using observational data to forecast values of parameters in the true structure. But in practice, the maximum likelihood method can only be applied to artificial, experimental cases. For practical use, the maximum likelihood method is applied to an artificial

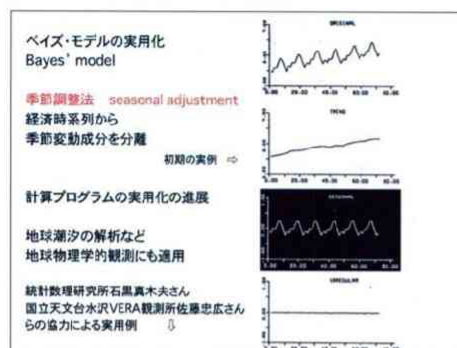


Fig. 12

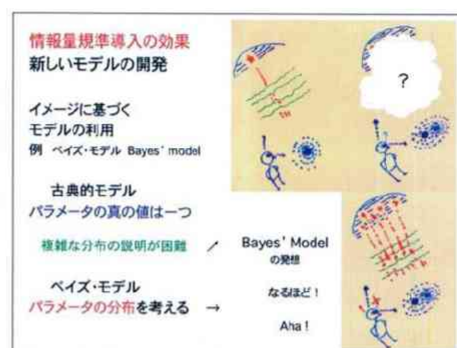


Fig. 13

んで重視しています。AICはこの過程の実行を助けるものです。

情報量規準導入の効果

最尤法は、真の構造の中のパラメータの値を、観測値を用いて推定する方法として、これまで理論が展開されてきました。しかし、現実の適用では、これは人工的な実験の場合以外にはあり得ません。実用場面で最尤法が適用されるのは、データからの有効な情報の抽出を目指した人工的な構造物、すなわちモデルなのです。このように考えると、伝統的な統計学が強調した主観性の排除が、実は柔軟な発想によるモデルの展開とその利用を制約し、有効なデータ処理法の実現を妨げてきたことが明らかになります。

情報量規準の導入がもたらす自由な発想により、モデル構成の自由度が高まり、心理的なイメージに基づくモデルをも含め、利用できるモデルの種類が急激に拡大されました。その効果は、ベイズ・モデルと呼ばれるモデルにより、経済時系列から季節変動の成分の影響を分離する、季節調整法の実現の例で確認できます (Fig.12)。ベイズ・モデルの発想の様子を図に示します (Fig.13)。

このモデルは、地球潮汐の解析などの地球物理学的観測値の有効な処理法の開発にも利用されています。統計数理研究所の石黒真木夫さんと国立天文台水沢VERA観測所の佐藤忠広さん他の皆さんが纏め上げた計算プログラムによる出力例をご覧に入れ

structure that is aimed at extracting useful information from data; in other words, a model. Thinking in these terms, we may realize that elimination of subjectivity, as emphasized by classical statistics, in fact restricts the flexible thinking needed to develop and use a model, and as a result prevents the realization of a useful data-processing method.

The introduction of an information criterion brought about flexible thinking, which increased the degree of freedom in creating a model. Consequently, usable models including those based on psychological images rapidly increased in kind. This effect can be confirmed in the example of the seasonal adjustment method that uses Bayes model to separate impacts of seasonally varying components from an economic time series [Fig. 12]. This figure illustrates the concept of Bayes model [Fig. 13].

This model has also been used in the development of useful processing of geophysical observation data obtained from analysis of earth tides and other natural phenomena. Let me show you an example of outputs from a calculation program compiled by Mr. Makio Ishiguro from the Institute of Statistical Mathematics, and Mr. Tadahiro Sato and others from the Mizusawa VERA Observatory of National Astronomical Observatory of Japan [Fig. 14].

Conclusion

In closing, let me talk about effective use of knowledge. There is a well-known story with respect to the practical use of the airplane. Scientists used their scientific knowledge to prove the airplane could not be put into practical use. In contrast, the Wright brothers used their knowledge to fly in the sky and put the airplane into practical use. This is the case with the AIC, too. I realized the limitations of data processing based on the existing theory. I examined the existing theory, according to the requirements of problems, and then achieved the AIC.

You might have an impression that my story about the AIC started with how I had a connection to the airplane in my childhood, and ended as a comparison of studies on the AIC with studies on the airplane. However, in closing, what I want to emphasize is the importance of having a sense of purpose, or trying to achieve

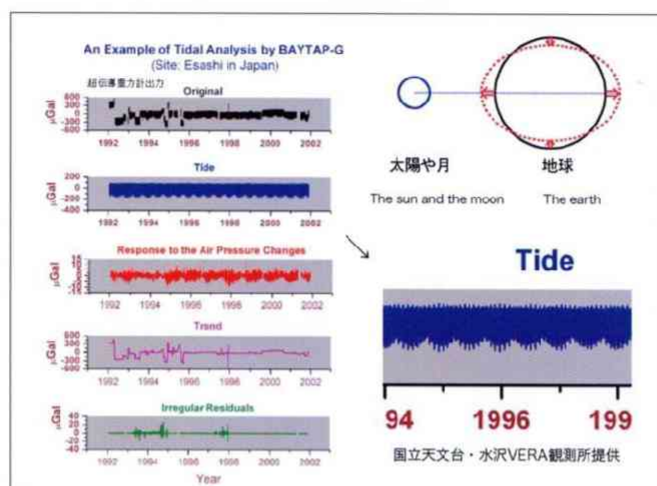


Fig. 14

ます (Fig. 14)。

終わりに

話の終わりに、知識の有効利用について触れたいと思います。飛行機の実用化に際しては、科学者が科学的知識を利用して飛行機は実用にならないことを示し、ライト兄弟は逆に空を飛ぶことを目指して知識を活用し、飛行機の実用化に成功したことが知られています。AICの場合も、既存の理論に基づくデータ処理法に限界を感じ、問題の要求に従って既存の理論を検討することで結果に到達しました。

私のAICの話は、子供の頃の飛行機との関わりに始まり、飛行機の研究との類比で終わる感じがしますが、ここで最後に強調したいことは、何かを実現しようとする目的意識の重要性です。セメント焼成炉あるいは火力発電所ボイラーの制御の実現では、関係する研究者の粘りと迫力には目覚ましいものがありました。この点についての理解と関心の深まりを願って、私の話を終わりたいと思います。

something. In achieving control of cement kilns and boilers at thermal power plants, the researchers were remarkably persistent and energetic. Hoping for deeper understanding and interest in this regard, I would like to end my talk.

稲盛財団 2006——第22回京都賞と助成金

発 行 2007年8月1日

制 作 財団法人 稲盛財団

〒600-8411 京都市下京区烏丸通四条下ル水銀屋町620番地

Phone: 075-353-7272 Fax: 075-353-7270

E-mail admin@inamori-for.jp URL <http://www.inamori-for.jp/>

ISBN4-900663-22-0 C0000